Tropical Cyclone Physics

Outline

1. Introduction

Part I: The Setting

2. Radiation, Convection, and Radiative-Convective Equilibrium
- 2.1. Radiative transfer
- 2.2. Radiative equilibrium
- 2.3. Dry convection
- 2.4. Dry convective-radiative equilibrium
- 2.5. Introduction to Moist RCE

3. Moist Convection and Moist Radiative-Convective Equilibrium
- 3.1. Convective clouds and mesoscale organization
- 3.2. Moist convective stability and the tropical tropospheric lapse rate
- 3.3. Moist RCE
  - 3.3.1. General features of moist RCE
  - 3.3.2. Behavior of global perturbations to RCE
  - 3.3.3. Behavior of local perturbations to RCE
- 3.4. Stability of RCE and the aggregation of tropical convection
- 3.5. Tropical squall lines

4. Tropical Boundary Layers
- 4.1. Some general principles
- 4.2. Semi-infinite thermal and mechanical turbulent layers
- 4.3. Aerodynamic flux formulae
- 4.4. Clear boundary layers
- 4.5. Stratocumulus-topped boundary layers
- 4.6. Transitioning boundary layers
- 4.7. Trade cumulus boundary layers
- 4.8. Summary

5. Large-scale Circulations in the Tropical Atmosphere
- 5.1. Overview
  - 5.1.1. Observations
  - 5.1.2. Analysis techniques
  - 5.1.3. Large-scale climatology of the tropics
  - 5.1.4. Zonal mean, time-mean climatology
- 5.2. The Hadley Circulation
  - 5.2.1. The equatorially symmetric Hadley circulation
  - 5.2.2. Equatorially asymmetric thermally direct circulations
  - 5.2.3. The time-dependent Hadley Circulation
- 5.3. Monsoons
  - 5.3.1. Zonally symmetric theory and the effects of land

Part II: Tropical Cyclones

8. Overview of Tropical Cyclones

8.1. Definition
8.2. A note on terminology and metrics
8.3. Climatology
  8.3.1. Frequency
  8.3.2. Geographic distribution
  8.3.3. Life cycle
  8.3.4. Seasonality
  8.3.5. Intensity
  8.3.6. Size
  8.3.7. Rain
  8.3.8. Extratropical and tropical transitions
  8.3.9. Polar lows, medicanes, and agukabams
8.4. Structure
  8.4.1. Overview
  8.4.2. Kinematic and thermodynamic structure
  8.4.3. The eye and eyewall
  8.4.4. Spiral rainbands and concentric eyewalls
  8.4.5. Associated phenomena
    8.4.5.1. Waves
    8.4.5.2. Storm surges
    8.4.5.3. Freshwater flooding
    8.4.5.4. Tornadoes

9. Axisymmetric Steady-State Theory

9.1. Tropical cyclones as Carnot engines
  9.1.1. The differential Carnot cycle
  9.1.2. Boundary layer sources of enthalpy and angular momentum
  9.1.3. Potential intensity
9.2. Tropical cyclones as quasi-balanced, slantwise-neutral vortices
  9.2.1. Slantwise convection and slantwise neutrality
  9.2.2. Analogy to quasi-balanced Eady model
  9.2.3. Thermal wind balance and characteristic surfaces
  9.2.4. Theory of the outflow layer
  9.2.5. Putting it all together: A closed model of a slantwise neutral tropical cyclone
  9.2.6. Dry and moist tropical cyclones
9.3. Axisymmetric, steady structure theory
  9.3.1. Outer wind field
  9.3.2. Merged inner and outer wind fields
  9.3.3. Thermodynamic balance in the outer boundary layer and equilibrium cyclone size
9.4. The tropical cyclone eye
9.5. The tropical cyclone boundary layer
  9.5.1. Boundary layer equations
  9.5.2. The slab boundary layer
  9.5.3. Vertically resolved tropical cyclone boundary layers

10. Axisymmetric Spin-up and Transients

10.1. Elements of a simple model of inner core spin-up
- 10.1.1. Some approximate solutions for rates of intensification
- 10.1.2. Eyewall frontogenesis and eye formation
- 10.1.3. Full numerical solutions of the simple time-dependent model

10.2. Full-physics axisymmetric models

10.3. Secondary eyewalls and eyewall replacement cycles

10.4. Decay over land and cold water

10.5. Observed statistics of intensification and dissipation

11. Asymmetric Features: Theory and Numerical Experiments

11.1. Spontaneous symmetry breaking
- 11.1.1. Eyewall mesovortices
  - 11.1.1.1. Rossby waves and associated instabilities
  - 11.1.1.2. Inertial instabilities
- 11.1.2. Asymmetric boundary layer structures
- 11.1.3. Spiral rainbands
  - 11.1.3.1. Vortex modes
  - 11.1.3.2. Squall line in a tropical cyclone wind field

11.2. Asymmetries owing to storm translation

11.3. Asymmetries owing to beta

12. Interaction of Tropical Cyclones with their Atmospheric Environment

12.1. Movement of tropical cyclones
- 12.1.1. Tropical cyclone motion in flows with vertical shear
- 12.1.2. Effects of anticyclonic outflow on tropical cyclone motion
- 12.1.3. Effects of environmental potential vorticity gradients
- 12.1.4. Thermodynamic considerations

12.2. Effects of environmental wind shear on structure and intensity
- 12.2.1. Dynamics of shear-vortex interaction
- 12.2.2. Thermodynamics of shear-vortex interaction

12.3. Interaction with environmental potential vorticity
- 12.3.1. Radiation of Rossby waves and other wave types from tropical cyclones
- 12.3.2. Interaction of tropical cyclones with localized environmental potential vorticity anomalies

12.4. Extratropical and tropical transitions

12.5. Polar lows and medicanes

13. Interaction of Tropical Cyclones with the Ocean and Land Surface

13.1. Ocean surface waves

13.2. Storm surges

13.3. Sea spray

13.4. Near-inertial ocean current and temperature response

13.5. Non-baroclinic re-intensification over land

14. The Genesis Problem

14.1. The local view
- 14.1.1. Case study: The development of Hurricane Ivan of 2004
- 14.1.2. Case study: The non-development of Tropical Storm Gaston of 2010
- 14.1.3. Ivan and Gaston in the context of other observational studies and field programs
- 14.1.4. Inferences from numerical simulations
- 14.1.5. Synopsis of the local view of tropical cyclogenesis

14.2. The global view
- 14.2.1. Genesis indices
- 14.2.2. “TC-World” Experiments
- 14.2.3. Random seeding experiments

14.3. Synthesis

15. Climate and Tropical Cyclones

15.1. Modulation of tropical cyclone activity by ENSO

15.2. Modulation of tropical cyclone activity by tropical intraseasonal variability

15.3. Paleotempestology
- 15.3.1. Tempestites
- 15.3.2. Overwash deposits and blue holes
- 15.3.3. Beach deposits
- 15.3.4. Geochemical proxies
- 15.3.5. Historical data mining
- 15.3.6. Perspective on paleotempestology

15.4. Theoretical considerations
- 15.4.1. Potential intensity
- 15.4.2. Normalized mid-tropospheric moisture
- 15.4.3. Thermodynamic component of a genesis potential index
- 15.4.4. Wind shear
- 15.4.5. The importance of the spatial distribution of tropical SST

15.5. Historical trends in environmental parameters

15.6. Aerosol forcing and the late 20th century hurricane drought

15.7. Greenhouse gas forcing: global warming
- 15.7.1. Behavior of TC-relevant environmental parameters
- 15.7.2. Direct numerical simulations
- 15.7.3. Downscaled numerical simulations

15.8. Possible feedback of TCs on climate
- 15.8.1. Feedbacks on clouds and water vapor
- 15.8.2. Feedbacks on ocean heat transport

Tropical Cyclone Physics

1. Introduction

The climate of the tropics is at once the most benign and the most dangerous on our planet. Most of the time the weather, as Christopher Columbus described it, is like May in Seville. Yet this salubrious climate, the placid subject of travel agency brochures and dream source of snowbound northerners, begets the most violent storms on earth. How and why does this happen? What physics determines the genesis, intensification, structure, and ultimate power of tropical cyclones? What role do they play in Earth's climate? This book addresses these and other fundamental questions.

Beside its intrinsic interest as a natural phenomenon, the tropical cyclone is a nearly perfect laboratory for geophysical fluid dynamics and thermodynamics, encompassing such concepts as balanced flow, potential vorticity conservation and inversion, dry and moist convection, boundary layer physics, frontogenesis, inertial, symmetric, and Kelvin-Helmholtz instabilities, Rossby waves, Carnot cycles, irreversible entropy production, interaction of longwave and shortwave radiation with clouds and water vapor, and upper ocean dynamics and mixing processes. As such, the content of this book is designed for the student with a firm background in classical physics, integral and differential equations, and geophysical fluid dynamics at the first-year graduate level. Some preparation in the physics of radiative and convective energy transfer will also be helpful.

To reach a viable understanding of tropical cyclones, it is first necessary to set them in the context of the tropical atmosphere-ocean system as a whole. Therefore, we first turn to the subjects of radiation and convection, which together form the two main elements of radiative-convective equilibrium, which we take to be a useful and natural starting point in describing the tropical atmosphere and ocean surface. Progressing from one spatial dimension to two, we inquire whether and to what extent the existence of meridional gradients of solar radiation perturbs our one-dimensional picture, leading us to an exposition of large-scale tropical circulations. The existence of continents breaks the symmetry of the response of ocean currents to wind stresses imposed by the Hadley circulation, creating strong east-west gradients in ocean surface temperature, which in turn lead to zonal atmospheric circulations. So, too, do continents provide the setting for seasonal monsoonal circulations. Besides these phenomena, there is a rich variety of more time-dependent organized structures, from squall lines and cloud clusters to equatorially trapped and African easterly waves, to the massive and mysterious planetary wave known as the Madden-Julian Oscillation. In addition to being of inherent interest to us, all these circulations play a role in the formation and behavior of tropical cyclones.

Then we arrive at our true destination: a comprehensive description of the physics of tropical cyclones as best we know them today. After reviewing the observed characteristics of tropical cyclones, we begin with perhaps the best known problem, the dynamics and energetics of the steady, circularly symmetric tropical cyclone vortex. While clearly an idealization, it has proven a useful one even though it is unlikely that the outflow region at the top of the storm is ever very symmetric or steady in time. We then tackle the time-dependent problem of the intensification of tropical cyclones, retaining the idealization of

circular symmetry. Finally relaxing this assumption, we examine the nature of departures from circular symmetric, including spiral rainbands and outflow jets, and time-dependent phenomena such as secondary eyewall cycles.

Tropical cyclones are powered by turbulent enthalpy fluxes from the sea and retarded by turbulent dissipation in the atmospheric boundary layer owing ultimately to frictional stress with the underlying surface. But these exchanges are complicated by the complex physics of the air-sea interface at the extraordinary wind speeds of tropical cyclones. The ocean is far from a passive source of heat and sink of momentum; it reacts strongly to the passage of tropical cyclones. Surface waves carry off a surprising fraction of the energy generated by the cyclone, and the storm-generated surface stresses can resonate strongly with near-inertial currents in the upper ocean. The turbulent breakdown of such currents leads to mixing to the sea surface of colder waters in the seasonal thermocline, producing strong negative feedback on the storm itself. The role of ocean feedback and the importance of the air-sea interface are the central subjects of Chapter 13, as are the hydrodynamics of storm surges, tsunami-like phenomena generated by wind stresses but strongly affected by local bathymetry and astronomical tides. Surges are the main cause of loss of life in tropical cyclones globally.

The large-scale atmospheric environment also exerts strong influences on tropical cyclones. Vertical shear of the horizontal wind interacts in complex ways with tropical cyclones, stretching and separating the vortex's potential vorticity anomaly, leading to such phenomena as precession and re-alignment. Chapter 12 describes these interactions as well as the thermodynamic consequence of injecting large quantities of low entropy air from the environment into the vortex core, an effect akin to throwing buckets of water on a fire. But as tropical cyclones move into higher latitudes, they often interact with extratropical potential vorticity anomalies and large-scale horizontal temperature gradients, leading to fascinating and not fully understood evolutions, which can include re-intensification of the storm long after it has left its feeding grounds deep in the tropics. We also take the opportunity to describe tropical-cyclone like phenomena outside the tropics, including certain classes of polar lows that paradoxically favor arctic oceans in winter, medicanes…hurricane-like storms that form over the Mediterranean Sea… and agukabams, which develop over land under special circumstances.

Of all the mysteries that confront the tropical cyclone physicist, the nature of the processes by which these storms are generated remains the most enigmatic. In Chapter 9 we delve into this problem, examining in detail the sequence of events leading to the genesis of individual storms but also looking at the physics by which tropical storms emerge spontaneously from the chaos of radiative-convective equilibrium as simulated by contemporary numerical models that explicitly simulate cumulus clouds. This may provide clues about what controls the frequency of tropical cyclones on our planet in our current climate.

Which leads to the last topic of our treatise: How does tropical cyclone activity respond to climate change, and are these magnificent and terrifying storms incidental to our climate or integral elements of the climate machine, providing important feedbacks that stabilize climate or accelerate its change? Here we leave our reader with a perplexing unsolved mystery, but one which he or she should by then be well equipped to tackle.

2. The Setting: Overview of the Tropical Atmosphere

The tropical atmosphere is one of the last frontiers of meteorology. Formally, it is the portion of the atmosphere that lies between the Tropics of Cancer and Capricorn (23.87^o north and south, respectively), comprising 40% of the mass of the atmosphere. But in a more colloquial sense, the tropical atmosphere is loosely defined as the part of the atmosphere in which radiation, moist convection, and thermally direct overturning circulations dominate the physics. For our purposes, we will take this to include places in which shallow convection prevails beneath a region in which net radiative cooling approximately balances warming associated with large-scale subsidence. Such regions are often referred to as "sub-tropical". It is important, though, to recognize that there are no real geographic boundaries between sub-tropical and tropical regimes; indeed, in many places there is an alternation between periods of deep and shallow convection.

In coming to a conceptual understanding of complex fluid systems, it is often helpful to begin by identifying simple, often stationary, nonlinear solutions of the system and ask whether those systems are stable. If not, what is the nature of the instabilities that develop, and what kind of statistical equilibrium do they help establish?

For example, in conceptualizing the behavior of synoptic and planetary scales at middle and high latitudes, it is helpful to begin with a simple, zonally symmetric, stably stratified atmosphere with a meridional temperature gradient in thermal wind balance with a zonal wind. In the absence of friction, topography, and diabatic processes, such a state is an exact solution of the governing equations. In elementary treatments of geophysical fluid dynamics and meteorology, the processes that establish and/or maintain such states are seldom dwelled upon, and many students think of such a state simply as the time mean state of the system[1].

Disturbances to such a state are usually idealized as being inviscid and adiabatic and thus contain a set of invariant quantities, the most important of which is the potential vorticity. If the time scales of the disturbances are sufficiently long compared to the inverse of the Coriolis parameter, then the disturbances themselves are approximately in geostrophic (and hydrostatic) balance, and the potential vorticity can be inverted, subject to certain boundary conditions, to obtain the balanced flow and thermodynamic variables. The conservation and invertibility of potential vorticity offer a compact way of understanding synoptic and planetary scales phenomena in middle and high latitudes. One can then go on to ask how these disturbances affect the mean state. Frictional and diabatic processes are usually regarded as of secondary importance.

By contrast, diabatic processes are of first-order importance to much of what happens in the tropics, and thus the tools we are accustomed to using at higher latitudes are of less utility in the tropics. In particular, it is seldom useful to regard the dynamics of tropical disturbances as adiabatic perturbations on a stably stratified background state. Even so, the concept of balance

[1] Formally, the time mean, zonal mean state of an eddy-containing atmosphere cannot be in strict thermal wind balance because of Reynold stresses in the meridional momentum equation.

applies to a surprisingly large spectrum of tropical phenomena, though in some cases the balance is nonlinear and the geostrophic approximation is poor.

Where the middle latitudes are characterized by strong horizontal temperature gradients, the tropics are relatively homogeneous, because strong temperature gradients cannot usually be maintained in the presence of small values of the Coriolis parameter. (But strong vortices can and do have strong local temperature gradients, an example of nonlinear balance.) Consequently, a natural starting point for the tropics is the state of radiative-convective equilibrium (RCE), in which there is a statistical equilibrium between radiation and deep moist convection, and horizontal energy transport is neglected. In RCE, diabatic processes are not only non-negligible, they are central. And, unlike the case of a zonal wind in thermal wind balance, RCE is a *statistical* equilibrium state, definable only in terms of a space-time or ensemble average.

Perturbations to RCE cannot logically be considered to be dry adiabatic, and in many cases it is not consistent to neglect associated perturbations in the radiation fields. For this reason, the middle latitude toolbox based on conservation of dry adiabatic invariants is not usually applicable to the tropics, and the paired principle of potential vorticity invertibility, based on geostrophic balance, can only be applied to certain large-scale tropical circulations and, in modified form, to strongly-rotating local disturbances such as tropical cyclones.

But if we must discard or modify our cherished middle latitude tools, we have one or two new ones that may serve to pave a pathway to conceptual understanding of many tropical phenomena. An important one, that we shall make extensive use of throughout this book, is the principle of *convective criticality*, which holds that the deep-convecting tropical atmosphere is in a state that is nearly neutral to deep moist convection. The lack of strong horizontal gradients in the properties of the tropical atmosphere and underlying surface makes it difficult to build up large amounts of convective inhibition or the convective available potential energy (CAPE) that often goes with it, so that deep convection is much more nearly in a state of statistical equilibrium with its large-scale environment, much as dry boundary layer convection is regarded as being in equilibrium with whatever is forcing it.

In its simplest form, convective criticality implies that the (virtual) temperature profile of the deep-convective tropical atmosphere lies along some suitably defined moist adiabat. To the extent this is true, and if the motions may be considered hydrostatic, we will show that the vertical structure of the pressure field is pre-determined, placing strong constraints on the dynamics of motions that are strongly coupled to deep convection. In the limiting case of small amplitude perturbations with a rigid lid at the tropopause, the linear equations reduce to the shallow water equations.

In locally balanced flows, such as tropical cyclones, the principle of convective criticality can be generalized to a statement that the vertical temperature profile along a vortex line (or angular momentum surface of a balanced, axisymmetric vortex) of the balanced flow lies along some suitably defined moist adiabat. If such a moist adiabat can be defined by the constancy of some saturated moist entropy variable, then it follows that a potential vorticity based on that variable is zero, since the dot product of the vorticity with the gradient of that entropy vanishes. Thus, to some level of approximation, the deep-convective portions of the tropical atmosphere may be characterized as having *zero saturation potential vorticity*. This quantity is constant not because it is conserved (it isn't), but because convection forces it to vanish. Yet, under suitable balance

conditions, it is fully invertible. In this case, as in the Eady problem of baroclinic instability, all the dynamics collapses to time-dependent boundary conditions.

Thus, when we make the intellectual journey from the middle latitudes to the tropics, we must abandon some of our traditional assumptions and tools, but on the other hand we gain some new ones, and other important concepts from higher latitudes do travel, in altered form, with us. To begin this journey we must familiarize ourselves with, and indeed learn to love, the physics of radiative transfer and water so neglected in most text books on middle latitude dynamics.

2.1 Radiative transfer

Understanding how electromagnetic radiation interacts with the atmosphere, the clouds within it, and the surface is critical for understanding much of what goes on in the tropics. This text is meant to provide an overview but can be no substitute for a full treatment of the subject; it assumes that the student has some working familiarity with the physics of radiative transfer. Those wishing a more comprehensive treatment are encouraged to read the excellent text by Pierrehumbert (2010).

Let's begin by reviewing the earth's radiation budget, as determined from satellite and surface measurements of radiation and turbulent fluxes. Figure 2.1 summarizes in very broad form the fluxes of energy through the global atmosphere.

Averaged over a year and over the whole surface area of the planet, about $342\,Wm^{-2}$ of solar energy enters the "top of the atmosphere" (TOA), which can be thought of as the level below which 99.9% of the mass of the atmosphere lies. Of this, about $30\,Wm^{-2}$ (a little less than 9%) is reflected by the surface and another $77\,Wm^{-2}$ (around 22%) is backscattered to space from clouds and, to a lesser extent, aerosols and the gaseous constituents of the atmosphere itself. This yields a net planetary reflectivity (albedo) of around $107\,Wm^{-2}$ or 30%.

About $67\,Wm^{-2}$, or 20%, of the incoming solar radiation is absorbed by the atmosphere and clouds and aerosols within it. Most of the gaseous absorption in the troposphere is by water vapor, which is somewhat more abundant in the tropical atmosphere than at higher latitudes, so this percentage is higher in the tropics. This helps create a diurnal cycle in tropical weather even over deep ocean waters whose surface temperature may not have a strong diurnal cycle.

Only a little less than half ($168\,Wm^{-2}$) of the TOA solar radiation is absorbed by the surface, on average. The surface radiates nearly as a blackbody with a characteristic temperature, in the tropics, of around 300 Kelvins. At this temperature, most of the $460\,Wm^{-2}$ of emitted radiation (as well as the $390\,Wm^{-2}$ for the globe as a whole) is in the infrared bands. Much of this is absorbed in the lower atmosphere by greenhouse gases and clouds, which also re-emit infrared radiation. In the global mean, $324\,Wm^{-2}$ are re-radiated to and absorbed the surface; this number is quite a bit larger in the tropics. This back-radiation from the atmosphere constitutes what is known in common parlance as the "greenhouse effect". Thus, again in the global mean, only about $66\,Wm^{-2}$ of the $168\,Wm^{-2}$ of incoming solar radiation at the surface is lost in the form of infrared radiation, the rest is transmitted to the atmosphere by turbulent motions in the

boundary layer. Over tropical oceans, the fraction lost by infrared radiation is appreciably less, thanks to the greater water vapor content (and thus greenhouse effect) of the atmosphere, and most of the incoming solar radiation is balanced by evaporation over the oceans and a combination of evapotranspiration and sensible enthalpy fluxes over land.

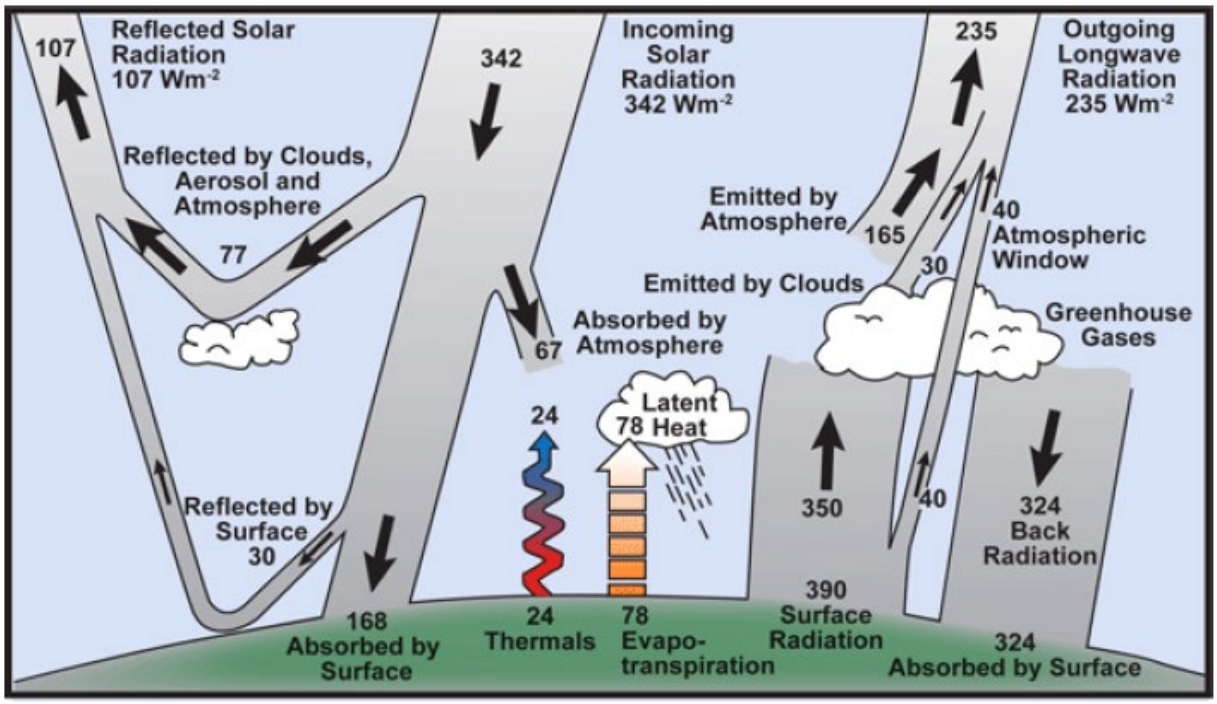


Figure 2.1: Broad overview of the earth's energy budget.

Ultimately, the earth system radiates $235\,Wm^{-2}$ back to space, in the form of mostly infrared radiation, to balance the incoming solar radiation less the fraction reflected back to space.

In its most fundamental form, radiation can be characterized by its wavelength, energy intensity at that wavelength, and its direction. In particular, the *radiant intensity,* I_λ, is the flux of energy per unit solid angle per unit wavelength of the radiation, and its relationship to the radiant flux density through a horizontal surface is illustrated in Figure 2.2. The monochromatic flux density (radiant energy flux per unit horizontal area per unit wavelength), F_λ, is given by

$$F_\lambda = \int_\Omega I_\lambda \cos\theta d\Omega, \tag{2.1}$$

where $d\Omega$ is the incremental solid angle. The total flux density (energy flux per unit horizontal area), F, is then just given by its integral over wavelength:

$$F = \int_0^\infty F_\lambda d\lambda. \tag{2.2}$$

As a beam of photons of a given wavelength travels through the atmosphere, it may be partially absorbed, scattered, and added to by emissions from greenhouse gases, aerosols, and clouds. The fundamental equation governing the change of radiant intensity per unit distance, s , is

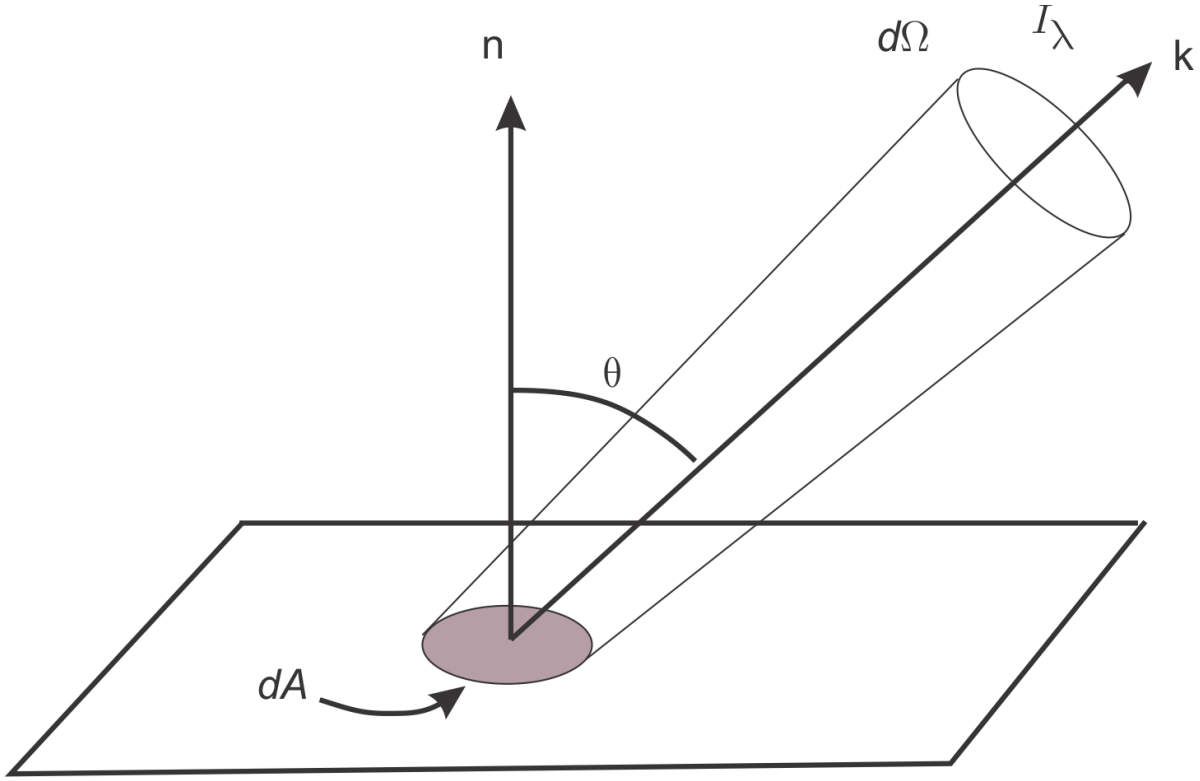


Figure 2.2: Relationship between the monochromatic flux density through a horizontal surface, the monochromatic radiant intensity, I_λ , the incremental solid angle, $d\Omega$, and the zenith angle, θ .

$$\frac{dI_\lambda}{ds} = -\left(\kappa_\lambda + \sigma_\lambda\right) I_\lambda + \varepsilon_\lambda + \beta_\lambda, \tag{2.3}$$

where κ_λ is the *absorption coefficient,* σ_λ *is* the *scattering coefficient* (both with dimensions of inverse length), ε_λ is the *emission coefficient*, and β_λ is the scattering source function. Scattering by the gaseous constituents of the atmosphere is well approximated by Rayleigh scattering, in which the size of the scatterers (molecules of the constituents of air) is very small compared to the wavelength of the radiation. This affects primarily the shortwave end of the visible spectrum along with the ultraviolet and is responsible for the blue color of the sky. Scattering by clouds is more nearly in the Mie regime, where the particle sizes are comparable to the wavelengths being scattered, which includes much of the visible and slightly longer

components of the solar spectrum. Scattering of terrestrial (infrared) radiation in our atmosphere is negligible.

Provided local thermodynamic equilibrium (LTE) conditions apply, *Kirchoff's Law*, which relates emissions to blackbody radiation is valid. This can be written

$$\varepsilon_\lambda = \kappa_\lambda B_\lambda(T), \tag{2.4}$$

where $B_\lambda(T)$ is the blackbody intensity, given by *Planck's Law*:

$$B_\lambda(T) = \frac{2hc^2}{\lambda^5\left[e^{hc/\lambda kT} - 1\right]}, \tag{2.5}$$

in which h is the Planck constant, c is the speed of light, k is Boltzmann's constant, and T is temperature. Local thermodynamic equilibrium is a good approximation for all the regions in which appreciable quantities of solar and infrared radiation are absorbed and emitted in our atmosphere. Using (2.5) allows us to write (2.3) as

$$\frac{dI_\lambda}{ds} = -\left(\kappa_\lambda + \sigma_\lambda\right) I_\lambda + \kappa_\lambda B_\lambda(T) + \beta_\lambda. \tag{2.6}$$

If we can neglect scattering, (2.6) further simplifies to

$$\frac{dI_\lambda}{\kappa_\lambda ds} = -I_\lambda + B_\lambda(T). \tag{2.7}$$

The form of (2.7) shows that the natural scale over which radiant intensity changes is a dimensionless quantity called the optical depth, τ_λ , defined so that

$$d\tau_\lambda \equiv \kappa_\lambda ds. \tag{2.8}$$

In the absence of emissions, the radiant intensity decays by a factor of $1/e$ over an optical depth of unity. The physical distance Δs over which this happens can be shown to be proportional to the mean free path of photons of wavelength λ . Thus optical depth, rather than actual distance, is the natural coordinate for radiative transfer. Using (2.8) we can write (2.7) in the form

$$\frac{dI_\lambda}{d\tau_\lambda} = -I_\lambda + B_\lambda(T). \tag{2.9}$$

We can make one more assumption that further simplifies matters. In a *plane-parallel atmosphere*, we assume that temperature and the concentration of absorbers are locally horizontally homogeneous, so that the variations in radiant intensity can be well approximated by vertical variations. In that case, since $ds = dz/\cos(\theta)$ (see Figure 2.2), we can write (2.7) as

$$\cos(\theta)\frac{dI_\lambda}{\kappa_\lambda dz} = -I_\lambda + B_\lambda(T), \tag{2.10}$$

where κ_λ and T are assumed to be functions of altitude (z) alone. In this case, it makes better sense to define the optical depth from

$$d\tau_\lambda \equiv \kappa_\lambda dz. \tag{2.11}$$

We will use this second definition and write the equation of radiative transfer for a plane-parallel atmosphere in the absence of scattering as

$$\cos(\theta)\frac{dI_\lambda}{d\tau_\lambda} = -I_\lambda + B_\lambda(T). \tag{2.12}$$

Defining $\mu \equiv \sec(\theta)$, we can write (2.12) as

$$\frac{d}{d\tau_\lambda}\left(I_\lambda e^{\mu\tau_\lambda}\right) = \mu B_\lambda e^{\mu\tau_\lambda}, \tag{2.13}$$

which can be integrated, starting from an optical depth of zero, to

$$I_\lambda = I_\lambda(0)e^{-\mu\tau_\lambda} + e^{-\mu\tau_\lambda}\int_0^{\tau_\lambda} \mu B_\lambda(T)e^{\mu\tau'}d\tau'. \tag{2.14}$$

Thus in a non-scattering, plane-parallel atmosphere, the monochromatic radiant intensity at any point is just that at the boundary, attenuated according to the optical depth times the secant of the zenith angle and added to by blackbody radiation from any emitters along the path, weighted by a function of the zenith angle, temperature, and the optical depth at that wavelength. To get the total radiant flux density through a horizontal plane at that wavelength, we would integrate the result using (2.1) and to get the flux over all wavelengths we would use (2.2). The rate of radiative heating is just the vertical derivative of the flux.

One very important point about radiative transfer is that it is non-local. Perturbing the temperature or concentration of absorbers/emitters at any level in the atmosphere affects the amount of radiation passing through (and potentially partially absorbed by) every other level. In general, radiative heating cannot be modeled as a Fickian diffusion or Newtonian relaxation. The non-local character of radiation plays an important role in various tropical weather systems, as we will describe later in this chapter.

Apart from scattering, the physics clearly is wrapped up in the absorption coefficient, κ_λ, which also dictates what fraction of the blackbody radiation is emitted, according to Kirchoff's Law (2.4). These physics are too extensive and complex to review here in any comprehensive way, but broadly, most of the absorption in our atmosphere is owing to transitions in the quantized vibrational, rotational, and combined rotational-transitional states of molecules. These quantum transitions result in absorption/emission at particular wavelengths, but these absorption/emission lines are widened by Doppler broadening, owing to the random molecular motion relative to the radiation source, and pressure broadening due to collisional effects.

Molecular oxygen (O_2) and nitrogen (N_2) together comprise almost 99% of the mole fraction of the dry atmosphere, with argon (Ar) adding another 0.93%. Oxygen and nitrogen are simple homonuclear, diatomic molecules with no electric dipole moment and they therefore have very few degrees of freedom with which to interact with electromagnetic radiation. Likewise, argon

has no interaction with radiation. Essentially all the absorption and emission in the atmosphere are owing to clouds, aerosols, and trace amounts of more complex molecules, notably water vapor

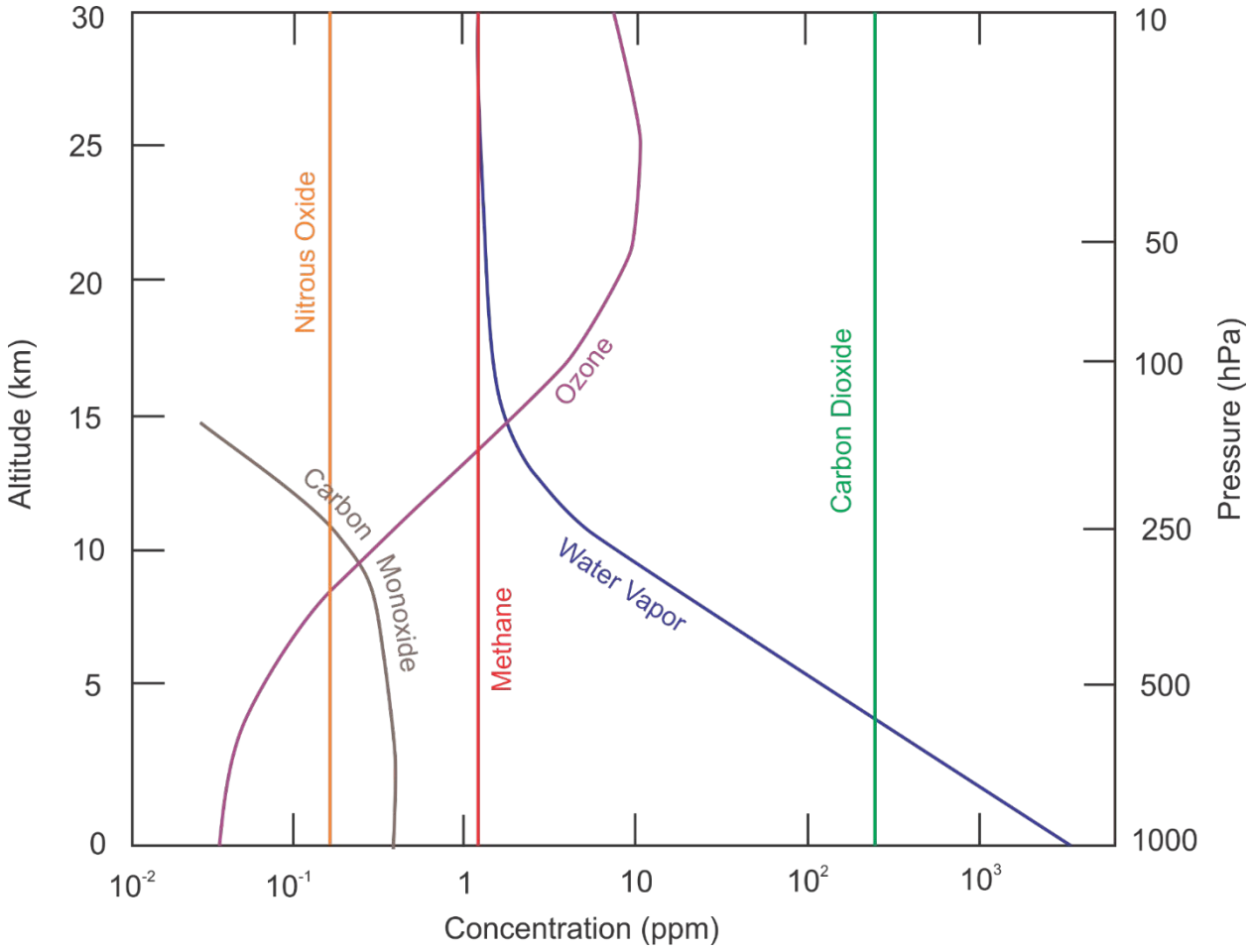


Figure 2.3: Mass concentrations (ppm) of some of the most important greenhouse gases in the atmosphere, as a function of altitude.

(H_2O), which is far and away the most abundant of these trace gases in the troposphere, but is highly variable in space and time, carbon dioxide (CO_2), methane (CH_4), ozone (O_3), carbon monoxide(CO), and nitrous oxide (N_2O). The time-mean, global-mean vertical distributions of these gases are shown in Figure 2.3, but note that the long-lived species (nitrous oxide, methane and carbon dioxide) are quite well homogenized in three-dimensional space and time. Ozone is formed in the middle and upper stratosphere by photodissociation of molecular oxygen by the high energy ultraviolet portion of the solar spectrum followed by recombination of some of the atomic oxygen with ordinary diatomic molecular oxygen, forming O_3. Ozone is unstable and itself breaks down into molecular and atomic oxygen by absorbing more ultraviolet radiation. These reactions absorb most of the ultraviolet portion of the solar spectrum, protecting life at the surface. Once formed, O_3 also absorbs and re-emits infrared radiation and is therefore a greenhouse gas.

Quantitatively, water vapor is the most important greenhouse gas owing to its relatively great abundance. Near the surface in the tropics, it can constitute up to 5% of the mole fraction of the

atmosphere. Its variability and the variability of it condensed phase in the form of clouds, play a major role in tropical meteorology.

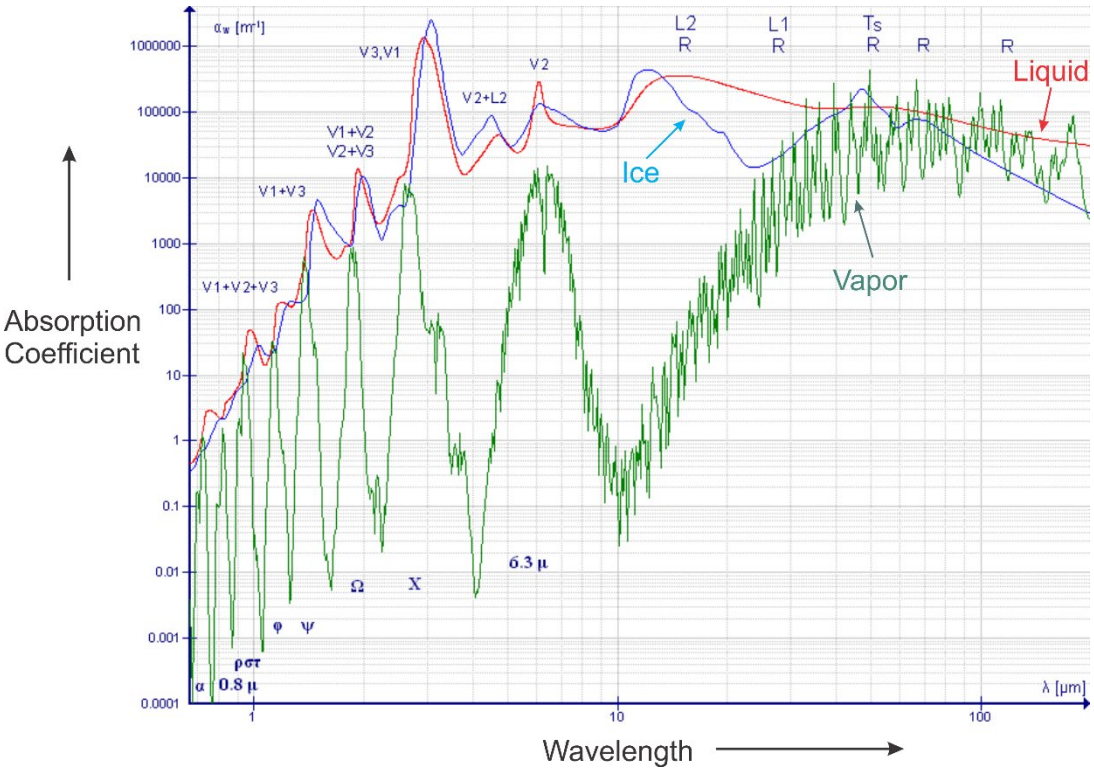


Figure 2.4 Absorption coefficient of water in its liquid (red), ice (blue), and vapor (green) phases in an important part of the terrestrial infrared spectrum, between $667\ nm$ *and* $200\ \mu m$ *. Note that both axes are logarithmic.*

Next to water vapor, CO_2 is the most abundant greenhouse gas, and its very long residence time in the atmosphere ensures that it is well mixed. CO_2 and the other long-lived greenhouse gases play an important role in radiative transfer in the tropics and elsewhere, but unlike H_2O, their variability plays no important role in tropical weather and short-term climate variations.

Although the greenhouse gases constitute a very small fraction of the mass of the atmosphere, they collectively have a large effect on radiative transfer. An example of the complexity of atmospheric absorption is illustrated in Figure 2.4, which shows the absorption coefficient of water in all three of its phases, for wavelengths between 667 nm and 200 μm . Pressure and Doppler broadening lead to an appreciable overlap in many of the absorption bands[2].

Let's examine some of the more prominent absorption/emission bands in our atmosphere. We begin by comparing the Planck (blackbody) curves of radiant flux density as a function of wavelength, for terrestrial and solar radiation, in the top panel of Figure 2.5. Because the effective emission temperatures of the two bodies are very different, there is hardly any overlap in their Planck functions and we can refer to solar and terrestrial radiation separately. (Note that the solar intensity has been reduced by a very large factor to make it easier to compare the shape of the distributions.)

[2] The reader is encouraged to explore the myriad absorption/emission bands relevant to our atmosphere through David Archer's web-based program at http://climatemodels.uchicago.edu/modtran .

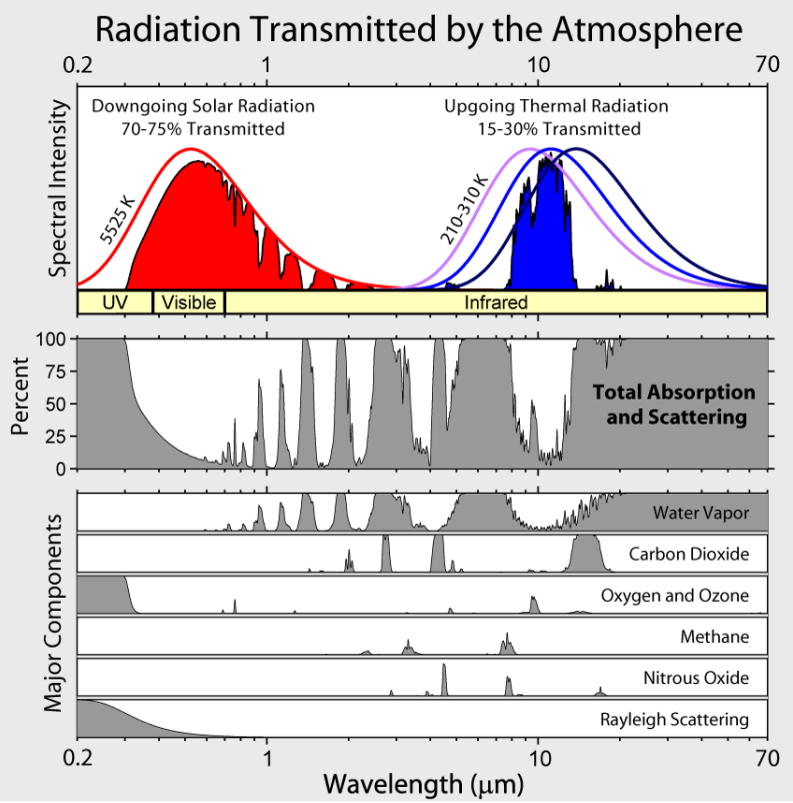


*Figure 2.5: Overview of absorption and scattering in the atmosphere. Top: Scaled blackbody curves for the sun (red) and the earth (black, blue, and magenta curves corresponding to different temperatures). The red and blue shaded areas represent the transmission of radiation. Middle: Fractional absorption and scattering of solar radiation passing downward toward the surface (left) and upward from the surface *right). Bottom panel: Breakdown of absorption among the greenhouse gases, and Rayleigh scattering by the gaseous constituents of the atmosphere.*

The color shading in the upper panels of Figure 2.5 shows how much radiant energy is actually transmitted through the cloud-free atmosphere at each wavelength. Most of the visible part of the solar spectrum is transmitted, which is presumably why our eyes evolved to detect light in this part of the spectrum. But at the infrared end of the solar spectrum, there is quite a bit of absorption, mostly by water vapor (see middle and bottom panels of Figure 2.5). Bear in mind that water vapor is highly variable in space and time. In the tropics, heating due to absorption of the infrared portion of the solar spectrum by water vapor can be quite large. Almost all the ultraviolet end of the spectrum is scattered and/or absorbed, the latter through the aforementioned photochemical reactions involved in ozone creation and destruction.

By contrast, much of the terrestrial spectrum is absorbed, with prominent absorption bands in water vapor (see Figure 2.4), carbon dioxide, methane, and nitrous oxide. There is a prominent window centered near $10\ \mu m$, which is at the peak of the blackbody curve corresponding to about 260 K.

The absorption spectra shown in Figure 2.5 do not include the effects of clouds. Clouds strongly absorb in the infrared and scatter solar radiation; the former acts to warm the surface while the

latter cools it. For a given infrared optical depth, high clouds are more effective at warming the surface than low clouds, because their effective emission temperature is lower.

2.2 Radiative equilibrium

Radiative equilibrium is achieved when each sample of the atmosphere, and the surface beneath it, emits as much radiation as it receives. Although the tropical troposphere is not near a state of radiative equilibrium, it is nevertheless of interest to calculate it since in a very loose sense, it is the state to which radiative processes are driving the actual atmosphere. (This is only very roughly true because the presence of other processes, such as convection and large-scale circulation, changes the distributions of water vapor and clouds and so alter the radiative processes.)

Before undertaking a comprehensive calculation of radiative equilibrium characteristic of the tropical atmosphere, it is instructive to obtain a set of very approximate solutions that demonstrate some important, general features of the equilibrium state. Let's begin by making things as simple as possible. Beginning with the radiative transfer equation (2.14) in a non-scattering, plane-parallel atmosphere, we will consider a single isothermal layer of gas with temperature T at rest over a solid or liquid surface. We are going to make a radical (and obviously bad) approximation that the optical depth is independent of wavelength λ. This is called the *gray atmosphere* approximation. We will refer to this wavelength-independent optical depth as just τ_g, where the subscript "g" stands for "gray". Thus we can write (2.14) as

$$I_\lambda = I_\lambda(0)e^{-\mu\tau_g} + e^{-\mu\tau_g}\int_0^{\tau_g} \mu B_\lambda(T)e^{\mu\tau'}d\tau', \tag{2.15}$$

and since temperature is constant, we can take the blackbody function outside the integral and perform the integration, giving

$$I_\lambda = I_\lambda(0)e^{-\mu\tau_g} + B_\lambda(T)\left(1-e^{-\mu\tau_g}\right). \tag{2.16}$$

We next substitute this into (2.1), using $d\Omega = \sin(\theta)d\theta d\phi,$ where ϕ is azimuth. Carrying out the integration in (2.1) then gives

$$F_\lambda = \left(1-\varepsilon\right)F_\lambda(0) + \varepsilon\pi B_\lambda(T), \tag{2.17}$$

where ε is an *emissivity*, given by

$$\varepsilon \equiv 1 - e^{-\bar{\mu}\tau_g}, \tag{2.18}$$

where $\bar{\mu}$ is defined such that

$$e^{-\bar{\mu}\tau_g} \equiv 2\int_0^{\pi/2} \cos(\theta)\sin(\theta)e^{-\mu\tau_g}d\theta.$$

Finally, we can integrate (2.17) over all wavelengths to get the net flux per unit horizontal area, using Planck's Law (2.5) for the blackbody intensity:

$$F = \left(1-\varepsilon\right)F(0) + \varepsilon\sigma T^4, \tag{2.19}$$

where σ is the *Stefan-Boltzmann constant*, given to three significant figures by

$$\sigma = \frac{2\pi^5 k^4}{15c^2h^3} = 5.67 \times 10^{-8} Wm^{-2}K^{-4}. \tag{2.20}$$

Thus, according to (2.19), the flux out one side of an isothermal, gray atmosphere is the incoming flux attenuated by $1-\varepsilon$ and added to by ε multiplied by the wavelength-integrated blackbody radiation, which is proportional to the 4th power of the (absolute) temperature.

Now let's apply this to an isothermal atmosphere that is entirely transparent to solar radiation and completely opaque to infrared radiation $(\varepsilon = 1)$, overlying a surface whose shortwave and longwave emissivity is also unity (Figure 2.6).

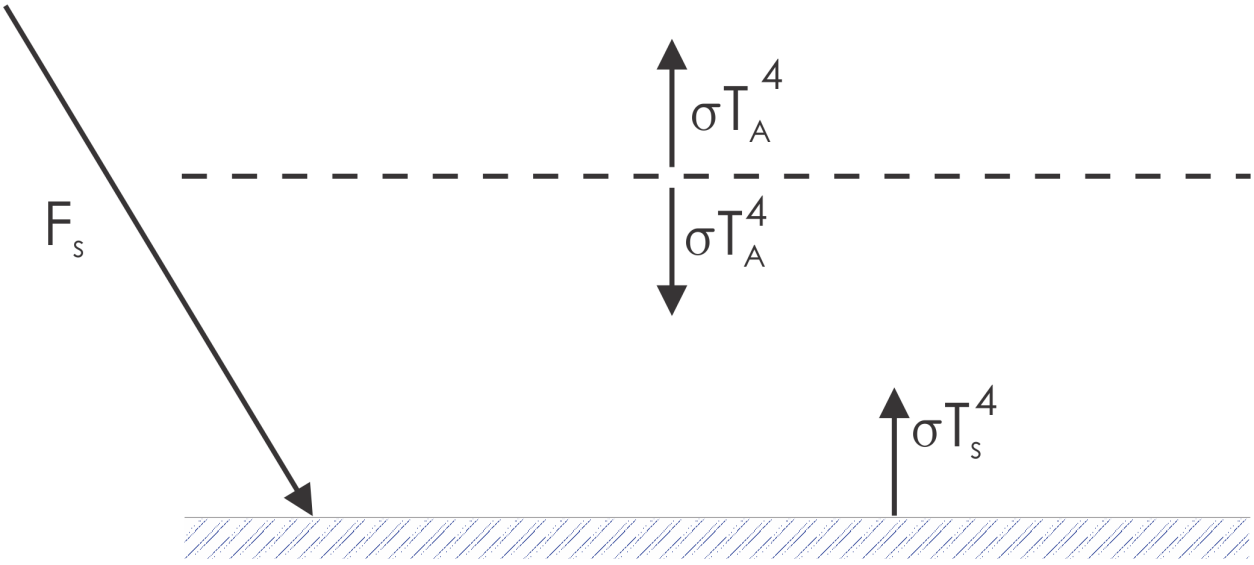


Figure 2.6: Plane-parallel, isothermal, non-scattering gray atmosphere transparent to incoming solar radiation, F_s, and opaque $(\varepsilon = 1)$ to infrared radiation. Atmospheric temperature is T_A and surface temperature is T_s.

First note that in the absence of an atmosphere, the total radiative flux from the surface would have to equal the incoming solar flux: $\sigma T_s^4 = F_s$. For global mean, annual mean insolation, after subtracting the portion reflected to space, this would yield a surface temperature of 255 K, which we will henceforth refer to as the *effective emission temperature* (T_e) of Earth. We can therefore represent the solar flux as σT_e^4. With the single opaque layer of isothermal gas shown in Figure 2.6, energy balance at the top of the atmosphere requires that $\sigma T_A^4 = F_s = \sigma T_e^4$, or $T_A = T_e$: The atmospheric temperature in this case is just the effective emission temperature of the planet. At the surface, however, energy balance requires that

$$\sigma T_s^4 = F_s + \sigma T_A^4 = 2\sigma T_e^4. \tag{2.21}$$

That is, the surface must be warm enough that its blackbody emission balances solar radiation *and* infrared back-radiation from the atmosphere. In this simple model, the surface receives the same quantity of back-radiation from the atmosphere as it does directly from the sun. This back-

radiation is conventionally called the *greenhouse effect* (though it has little to do with how an actual greenhouse works).

We next look very crudely at how temperature in an IR-opaque gray atmopshere varies with altitude. This can be illustrated by extending the single layer in Figure 2.6 to two layers, in Figure 2.7

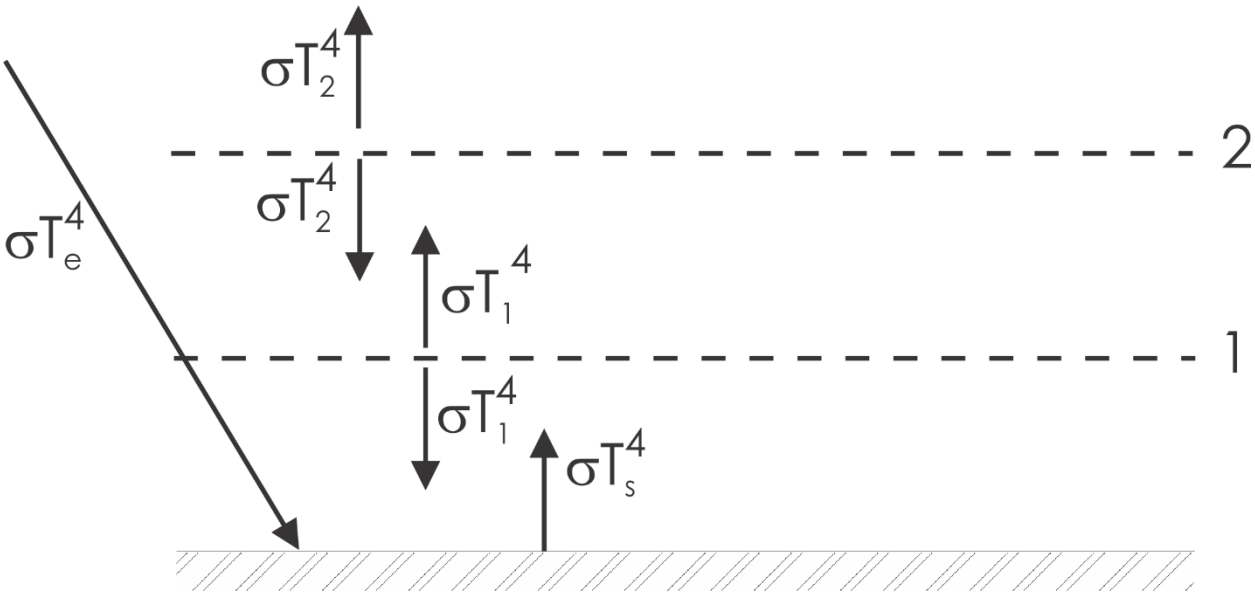


Figure 2.7: Extension of the one-layer model (Figure 2.6) to two layers. Incoming solar flux is expressed here as σT_e^4.

As before, energy balance at the top requires that $T_2 = T_e$. Energy balance in the middle layer requires that

$$2\sigma T_1^4 = \sigma T_2^4 + \sigma T_s^4, \tag{2.22}$$

while the surface energy balance can be written

$$\sigma T_s^4 = \sigma T_1^4 + \sigma T_e^4. \tag{2.23}$$

Solving (2.22) and (2.23) yields

$$T_s = 3^{1/4} T_e, \quad T_1 = 2^{1/4} T_e. \tag{2.24}$$

It is straightforward to generalize this result to N layers (each one having an emissivity of unity), in which case the temperature of the n^{th} layer, counting down from the top, is just $n^{1/4} T_e$, and the surface temperature is $(N+1)^{1/4} T_e$. (Note here that adding opaque layers is somewhat like adding to the mass of greenhouse gases; it is not equivalent to dividing a fixed atmosphere into ever thinner layers, which would have progressively smaller emissivities.)

The top layer receives infrared radiation from the layer below it, but nothing from above, whereas the bottom layer receives radiation from both the surface and the top layer, so it is hardly surprising that its temperature is higher than that of the top layer. In general, *in radiative equilibrium in a gray atmosphere with uniform emissivity, temperature decreases with altitude*.

In this problem, the radiative transfer is effectively "local" in the sense that each layer only influences its neighbor, because each layer absorbs all the radiation incident upon it. But if we relax the assumption of unitary emissivities, then some radiation from each layer will penetrate through to be partially absorbed by all other layers. This makes the problem algebraically more complex.

We look at two more cases in which we add two layers of very *low* emissivity, as illustrated in Figure 2.8.

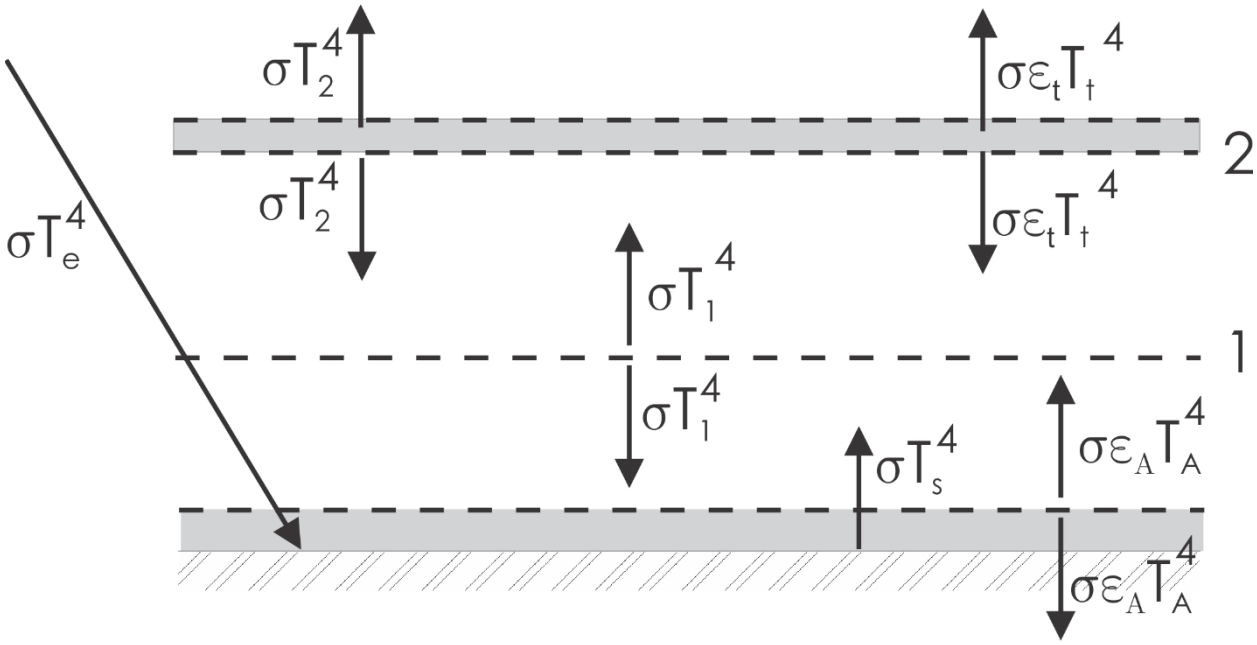


Figure 2.8: Two-layer model as in Figure 2.7, but with two additional layers of vanishingly small emissivities: one at the top, and once adjacent to the surface.

The infrared emissivities of the new top and bottom layers are ε_t and ε_A,respectively. If we consider the limit in which they vanish, then these new layers will have no effect on the temperature of the main layers and the surface, given by (2.24) (together with $T_2 = T_e$). On the other hand, the energy balance of the top layer itself requires that

$$2\varepsilon_t \sigma T_t^4 = \varepsilon_t \sigma T_2^4, \tag{2.25}$$

that is, the energy emitted up and down from this new layer must balance the partial absorption by this layer of energy upwelling from the second main layer. Thus $T_t = 2^{-1/4} T_e$. Therefore, *parts of an atmosphere in radiative equilibrium can have temperatures less that the effective emission temperature.* (In fact, this is usually the case.)

Likewise, for the thin layer next to the surface, the energy balance is

$$2\varepsilon_A \sigma T_A^4 = \varepsilon_A \sigma T_1^4 + \varepsilon_A \sigma T_s^4, \tag{2.26}$$

or $T_A^4 = \frac{1}{2}\left(T_s^4 + T_1^4\right)$. The thin layer of air in contact with the surface has a temperature intermediate between that of the surface and that of the first main layer of gas. The essential point here, and one that is important for understanding a variety of tropical weather phenomena, including tropical cyclones, is that in radiative equilibrium, *a discontinuity in emissivity (as between a gas and a solid or liquid surface) entails a discontinuity in temperature*. Thus, *for a gas overlying a surface, radiative equilibrium implies thermodynamic disequilibrium*. We will return to this important point time and again, but for now note that for a variety of physical and biological systems, thermodynamic disequilibrium is the basis for the development of self-organizing structures, even in otherwise chaotic background states (Nicolis and Prigogine, 1977).

We can obtain analytic radiative equilibrium solutions for a non-scattering, plane-parallel gray atmosphere that are continuous in the vertical for some special cases. We begin by re-writing (2.12) for a gray atmosphere that is transparent to solar radiation as

$$\cos(\theta)\frac{dI_\lambda}{d\tau_g} = -I_\lambda + B_\lambda(T). \tag{2.27}$$

Because τ_g is independent of wavelength, we substitute (2.27) into (2.1) and (2.2) and integrate over solid angle in a single hemisphere, and wavelength to obtain

$$\overline{\cos(\theta)}\frac{\partial F}{\partial \tau_g} = -F + \sigma T^4, \tag{2.28}$$

where $\overline{\cos(\theta)}$ is a suitably defined mean cosine of the zenith angle. In deriving (2.28) we have not specified whether the flux is upward or downward, and it helpful to divide the flux at each level into upward and downward traveling components:

$$F = U - D. \tag{2.29}$$

Defining $\tau_v \equiv \tau_g / \overline{\cos(\theta)},$ we write separate equations for the up and down components:

$$\frac{\partial U}{\partial \tau_v} = -U + \sigma T^4, \tag{2.30}$$

and

$$-\frac{\partial D}{\partial \tau_v} = -D + \sigma T^4. \tag{2.31}$$

Radiative equilibrium requires that there be no convergence of the net flux, $U - D$, through each level. Thus adding (2.30) to (2.31) gives

$$\frac{\partial}{\partial \tau_v}(U-D)=0=-U-D+2\sigma T^4, \tag{2.32}$$

while subtracting (2.31) from (2.30) yields

$$\frac{\partial}{\partial \tau_v}(U+D)=-U+D. \tag{2.33}$$

From (2.32) the net flux, $U-D$ must be independent of optical depth, and at the top of the atmosphere $D=0$ and the upward flux, U_∞, must equal the net absorbed solar flux (after subtracting the fraction reflected), S, so $U-D=S$. The solutions to (2.33) that meet the boundary conditions at the top of the atmosphere, $U_\infty=S, D_\infty=0$, are then

$$\sigma T^4=\frac{S}{2}(1+\tau_{v\infty}-\tau_v), \tag{2.34}$$

$$D=\frac{S}{2}(\tau_{v\infty}-\tau_v), \tag{2.35}$$

and

$$U=S\left[1+\frac{1}{2}(\tau_{v\infty}-\tau_v)\right]. \tag{2.36}$$

Here $\tau_{v\infty}$ is the infrared optical depth of the whole atmosphere. Note that as τ_v approaches $\tau_{v\infty}$ at the top of the atmosphere, $\sigma T^4 \to \frac{S}{2}$, which is the same solution (2.25) we found for the layer of small emissivity at the top of the layer model.

Note also that the downward infrared flux at $\tau_v=0$ is $S\tau_{v\infty}/2$. So for optically thin atmospheres, there is hardly any downwelling infrared flux at the surface, while the flux can exceed the solar flux for optically thick atmospheres. If we insert, at z=0, a rigid or liquid surface of unitary emissivity, then the energy balance at the surface is

$$\sigma T_s^4=S+D(0)=S\left(1+\frac{\tau_{v\infty}}{2}\right). \tag{2.37}$$

Comparing to the air temperature at $\tau_v=0$, from (2.34), we see that the difference between the surface and air temperatures is given by

$$\sigma\left(T_s^4-T^4(0)\right)=\frac{S}{2}. \tag{2.38}$$

Curiously, the difference in the fourth powers of the air and ground temperatures is one-half the incoming solar flux (divided by the Stefan-Boltzmann constant), regardless of infrared opacity. If

we let $T(0) = T_s - \Delta T$ and expand $T^4(0)$ to order one in ΔT, assuming that $\Delta T / T_s << 1,$ then it is approximately true that

$$T_s - T(0) \simeq \frac{S}{8\sigma T_s^3}, \tag{2.39}$$

So the ground-air temperature difference in radiative equilibrium drops off rapidly with surface temperature (and thus with infrared opacity).

We can plot these solutions as a function of altitude or pressure, given a particular distribution of optical depth with altitude. Suppose, for example, that optical depth from the surface upwards increases linearly with the difference between actual and surface pressures, according to

$$\tau_v = \tau_{v\infty}\left(\frac{p_0 - p}{p_0}\right), \tag{2.40}$$

where p_0 is the surface pressure. The solutions (2.34) and (2.39) are plotted in Figure 2.9 for particular choices of $\tau_{v\infty}$ and p_0. Note that the plot is logarithmic in pressure, which is approximately linear in altitude in a hydrostatic atmosphere.

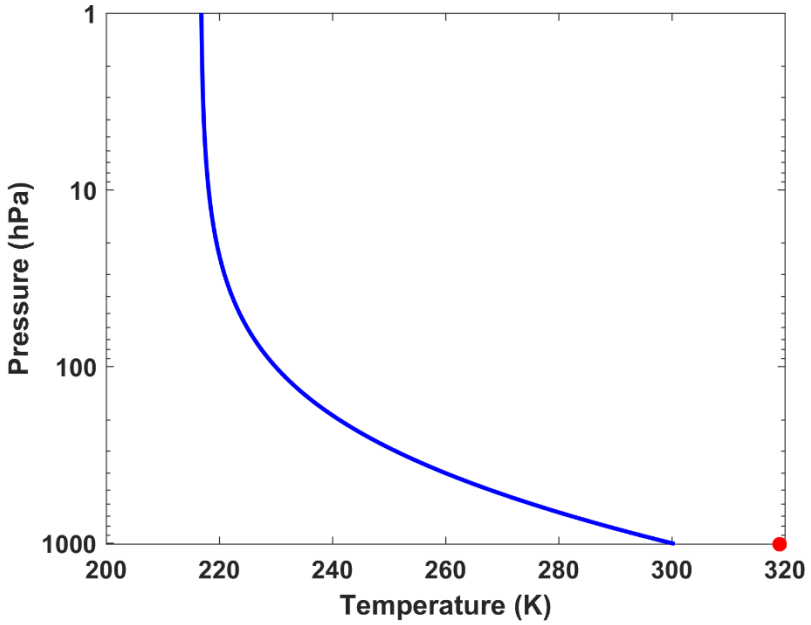


Figure 2.9: Radiative equilibrium temperature of a non-scattering, plane-parallel, gray atmosphere that does not absorb solar radiation, as given by (2.34) and (2.39) when the optical depth varies with pressure according to (2.40) with $S = 250\ Wm^{-2}$, $\tau_{v\infty} = 2.7$, *and* $p_0 = 1010\ hPa$. *The red dot shows the solution for the surface temperature.*

As with the discrete layer models, temperature decreases with altitude, but since the infrared optical thickness becomes small in the upper atmosphere, the temperature becomes nearly isothermal. As we shall see presently, the actual temperature of the tropical atmosphere increases in the stratosphere (above the level where the pressure is approximately 100 hPa), and this is due to absorption of high energy ultraviolet solar radiation in the middle and upper

stratosphere. As noted previously, radiative equilibrium entails a profound thermodynamic disequilibrium between the surface and the atmosphere just above it, a fact of great consequence for tropical meteorology and climate.

While the approximation of a gray atmosphere is very helpful in revealing some of the most essential aspects of radiative transfer and radiative equilibrium, it is quantitatively a poor approximation to the real atmosphere, with its very complex set of absorption/emission lines. Fortunately, the spectra for all the important greenhouse gases in the atmosphere have been calculated for a range of temperatures and pressures and are included in a comprehensive database known as HITRAN (*hi*gh-resolution *tran*smission molecular absorption database; see www.hitran.org). But a comprehensive, line-by-line treatment of radiative transfer is too expensive to use in most atmospheric models and so we turn to band-averaged codes which are much faster and which have been carefully compared to both observations and HITRAN data.

One such band-averaged radiation code is included as part of a single-column radiative-convective model, freely available at https://zenodo.org/doi/10.5281/zenodo.10795221 . This is coupled to a comprehensive representation of convection (about which more in due course), but for now we turn off the convection and look at the radiative equilibrium solution representative of the tropical atmosphere. This is problematic, because moist convection lofts water from its source at the surface into the free atmosphere; without it, there is no control on atmospheric water vapor, the most important greenhouse gas. Removing water vapor altogether would result in very low temperatures, but retaining values characteristic of the tropical atmosphere results in strong supersaturations in a pure radiative equilibrium calculation. We here choose the intermediate route of holding the relative humidity (the ratio of the vapor pressure to its saturation value) to be a fixed function of altitude, characteristic of the tropical atmosphere.

Figure 2.10 shows a calculation of pure radiative equilibrium using observed constant concentrations of carbon dioxide, methane, and nitrous oxide, CFC11, and CFC12, a fixed profile of ozone representing a global average, and water vapor calculated from the model predicted temperature and a fixed vertical profile of relative humidity characteristic of the tropical atmosphere. This calculation does not include the effect of clouds on radiative transfer and uses instead a representative fixed albedo of 0.25 and a fixed, annual mean insolation at 23 degrees north latitude. The lower boundary is a thin slab of water whose temperature is also calculated from radiant energy balance; the equilibrium water temperature is shown by the red dot in Figure 2.10. The dashed blue line in Figure 2.10 represents the mean temperature profile over more than 800 balloon soundings from the equatorial North Pacific island of Kapingamorangi, taken as part of the TOGA COARE field campaign.

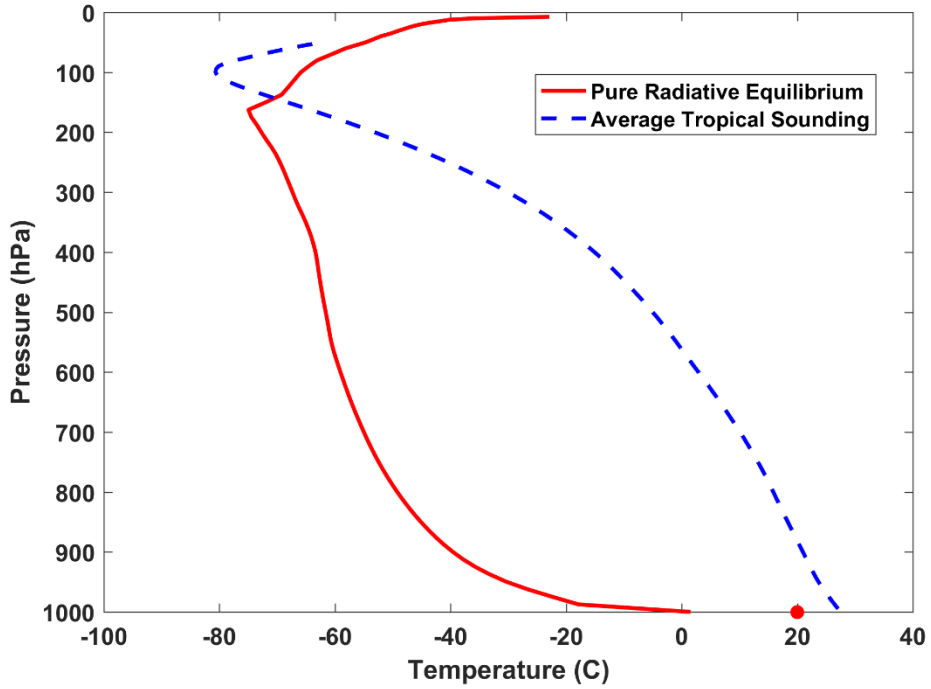


Figure 2.10: Calculation of pure radiative equilibrium (red curve) for the tropical atmosphere, under the assumption that the relative humidity is fixed at a representative tropical profile. The calculated temperature of the underlying water slab is shown by the red dot. (See text for further details.) The dashed blue line represents the average of over 800 rawinsonde temperature profiles from the equatorial North Pacific island of Kapingamarangi.

The radiative equilibrium shares some features with the highly idealized analytic solution shown in Figure 2.9. Both have a temperature jump of around 20 K between the surface and the atmosphere just above it, and both show a rapid fall of temperature with altitude, tapering off toward a constant temperature higher up. But the band-averaged solution then shows some notable departures. Above 400 hPa, the temperature lapse rate increases again, with temperature reaching a sharp minimum at around 175 hPa. This is likely owing to a layer of relative high humidity in the upper troposphere which, in the moist convective atmosphere, results from outflow from deep convective clouds. According to (2.34), if the infrared optical depth increases more rapidly, because of a layer of higher concentration of absorbers, the temperature should decrease more rapidly. In stark contrast to the simple solution, temperature increases rapidly above this tropopause, and this is owing to absorption of solar radiation by ozone in the stratosphere, which was neglected in the simple gray model.

The radiative equilibrium solution shown in Figure 2.10 bears little resemblance to the temperature profiles seen in the real tropical atmosphere (dashed blue curve in Figure 2.10). The whole troposphere is substantially colder than reality, and the curvature of the temperature profile is mostly opposite to that seen in the real soundings. The relatively low surface temperature and the steep decline of temperature with altitude imply, though the Clausius-Clapeyron equation coupled with our assumption of fixed relative humidity, that the water vapor concentration is much lower than in the real atmosphere. Since this is the most important greenhouse gas, the whole system is much colder than observed.

Why is the radiative equilibrium solution such a poor representation of the tropical atmosphere? Far and away the most important reason for this is that the radiative equilibrium solution is strongly unstable to convection, which is as important as radiation in transferring energy through the tropical atmosphere and which also serves to loft the most important greenhouse substance – water, in all three of its phases. Thus we turn to the important subject of convection.

2.3 Dry convection

A dynamical system in equilibrium is unstable if there are any small disturbances, added to the equilibrium state, that amplify with time. An unstable system will generally evolve to a state, which may be time-dependent, that differs from the original equilibrium. By contrast, small perturbations added to a stable equilibrium will either oscillate about that equilibrium or decay (or both), leaving the original equilibrium state essentially unaltered. A meta-stable equilibrium is stable to infinitesimal perturbations but unstable to a sufficiently large disturbance.

A simple system that can exhibit each of these stability states is a frictionless marble on a wavy surface in a constant gravitational field. The marble sitting at the bottom of a depression, when displaced in any direction, will roll back towards the bottom and will oscillate around the lowest point. Naturally, this equilibrium is stable. A marble on a perfectly flat portion of the surface is neutrally stable; when displaced it will not accelerate in any direction. When the marble is balanced on the top of a hill, it is in equilibrium, but any displacement will cause it to accelerate toward a depression, so that equilibrium is unstable. Now imagine a marble at rest on a saddle point: The topography goes uphill in, say, the positive and negative x direction, but downhill in the positive or negative y direction. (Think of the shape of a horse saddle.) If displaced exactly in the x direction, it will accelerate back toward the saddle point, but in any other direction it will not return to its original position. This is also an unstable equilibrium because there is at least one perturbation that amplifies. Finally, think of a hill with a small depression at its top. A marble in the base of that depression is stable because if pushed very slightly in any direction it will accelerate back towards it initial state. On the other hand, if pushed hard enough, it will roll over the top of the rim and accelerate away from the hill. This equilibrium state is metastable.

In thinking about the marble, you probably made some tacit assumptions about the system without necessarily being conscious of them. You almost certainly assumed that the mass of the marble is invariant….it did not accrete or lose mass on the way. You probably also assumed that after the initial displacement, the sum of the marble's kinetic and potential energies is conserved; there are no demons around giving the marble a shove. Most stability problems assume the dynamical system has certain invariants, like mass and total energy.

Now consider a horizontally homogeneous gas in hydrostatic equilibrium at rest over a rigid or solid surface, also in a state of radiative equilibrium. How can we tell if this system is stable?

As with the marble on a wavy surface, we will test the system for stability by vertically displacing a sample adiabatically and reversibly, so that it conserves its entropy. If the sample displaced upward obtains positive buoyancy, it will accelerate upward and thus the state is unstable. (Likewise, if the parcel is displaced downward and obtains a negative buoyancy, the system is unstable.) The equations of motion can be used to show that buoyancy is proportional to the perturbation of specific volume (inverse density) along a surface of constant pressure.

Thus, if the specific volume of a sample of air displaced upward is larger than that of its environment at the same pressure level, it will accelerate upward, in the same direction as the original displacement, and the state will be unstable. But in a gas, specific volume is not conserved and so this is not a simple matter of that sample taking its original specific volume with it. Consequently, we have to identify an adiabatic invariant of the air that we can use to deduce the specific volume. Here we use the specific entropy, s, of the sample as a quantity conserved under reversible adiabatic displacements from which one can deduce specific volume.

To relate specific volume to entropy we use one of many Maxwell relations from the field of thermodynamics. Not all students are familiar with these, so we will derive the relevant relation here. We begin with a statement of the first law of thermodynamics:

$$dk = Tds + \alpha dp, \tag{2.41}$$

where k is the specific enthalpy, T is temperature, s is specific entropy, α is specific volume, and p is pressure. From (2.41) it is evident that

$$\left(\frac{\partial k}{\partial s}\right)_p = T, \tag{2.42}$$

and

$$\left(\frac{\partial k}{\partial p}\right)_s = \alpha, \tag{2.43}$$

where the subscripts on the partial derivatives denote quantities being held constant. Now in partial differentiation, the order of differentiation is immaterial, so

$$\left(\frac{\partial}{\partial s}\right)_p \left(\frac{\partial k}{\partial p}\right)_s = \left(\frac{\partial}{\partial p}\right)_s \left(\frac{\partial k}{\partial s}\right)_p. \tag{2.44}$$

Using (2.42) and (2.43) in (2.44) gives

$$\left(\frac{\partial \alpha}{\partial s}\right)_p = \left(\frac{\partial T}{\partial p}\right)_s. \tag{2.45}$$

This is the desired Maxwell relation and we will make extensive use of it through the rest of this book. It is valid for a homogeneous gas, in which any state variable can be written as a function of any other two state variables. So, using the chain law, we can write

$$d\alpha = \left(\frac{\partial \alpha}{\partial s}\right)_p ds + \left(\frac{\partial \alpha}{\partial p}\right)_s dp. \tag{2.46}$$

Since buoyancy is proportional to perturbations of specific volume at constant pressure, we can use (2.45) and (2.46) to write an expression for perturbations of specific volume at constant pressure:

$$\alpha' = \left(\frac{\partial T}{\partial p}\right)_s s', \tag{2.47}$$

where the primes denote perturbation of the sample's properties from those of its environment. Finally, buoyancy itself is just the specific volume perturbation multiplied by the acceleration of gravity and divided by the mean state specific volume:

$$B = g\frac{\alpha'}{\bar{\alpha}} = \frac{g}{\bar{\alpha}}\left(\frac{\partial T}{\partial p}\right)_s s', \tag{2.48}$$

where g is the acceleration of gravity. Since the mean state is in hydrostatic balance, it is more compact to write (2.48) as

$$B = \Gamma_d s', \tag{2.49}$$

where Γ_d is the dry adiabatic lapse rate, defined

$$\Gamma_d \equiv -\left(\frac{\partial T}{\partial z}\right)_s. \tag{2.50}$$

It is a simple exercise to derive from the first law of thermodynamics that, for a homogeneous gas, $\Gamma_d = g / c_p$, where c_p is the heat capacity at constant pressure. In our atmosphere, the dry adiabatic lapse rate Γ_d, is nearly $1\,K/100\,m$.

The relationship between buoyancy and entropy, given by (2.49), is extraordinarily simple and powerful. A homogeneous atmosphere in hydrostatic balance will be unstable if its entropy decreases with altitude. Under these circumstances, a sample displaced upward reversibly and adiabatically will carry its entropy with it, and since the environmental entropy decreases upward, it will have an entropy larger than that of its environment and accelerate upward, in the same direction as the original displacement. Conversely, a gas at rest with entropy increasing upward is stable and therefore supports oscillations, which in meteorology are referred to as internal gravity waves.

Figure 2.11 shows the vertical distribution of the specific entropy of the radiative equilibrium temperature profile shown in Figure 2.10. From the first law of thermodynamics, the entropy may be expressed

$$s_d = c_p \ln(T / T_0) - R_d \ln(p / p_0), \tag{2.51}$$

where T_0 is an arbitrary reference temperature and p_0 is an arbitrary reference pressure. (Note that entropy is arbitrary to within an additive constant; in calculating entropy here we have normalized temperature and pressure by representative values, and one effect of this is to make the entropy negative at certain altitudes.) We have added the subscript *d* to the entropy here to denote the entropy of dry air. Also shown by the red dot in Figure 2.11 is the specific entropy of air at the same temperature as the underlying layer of liquid water.

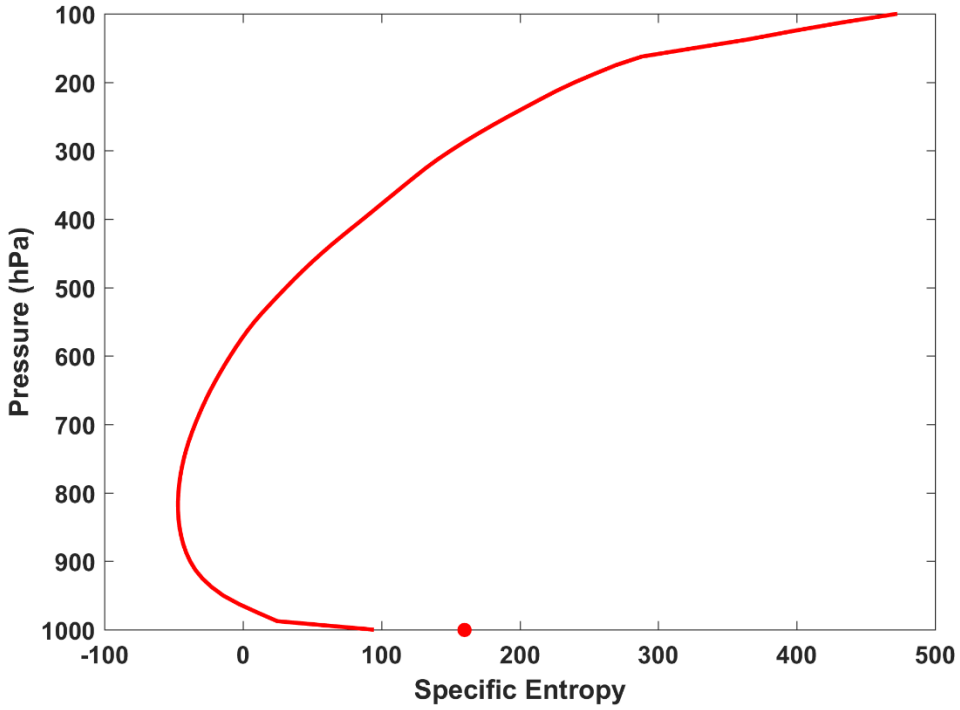


Figure 2.11: Vertical distribution of specific entropy corresponding to the radiative equilibrium temperature profile shown in Figure 2.10. The specific entropy of air in thermodynamic equilibrium with the underlying surface is shown by the red dot. Note that the value of the specific entropy is arbitrary to within an additive constant.

The profile is clearly unstable. A sample of air lifted upward from the surface will have a positive entropy perturbation up to about 400 hPa, while a sample that has been brought to the temperature of the underlying water (red dot) will be buoyant all the way up to about 300 hPa. Thus we would expect virtually the whole troposphere, starting from a state of radiative equilibrium, to convect.

What form does the convection take, and what is the nature of the new equilibrium it establishes? In our atmosphere, the effective Reynolds Number of convective flows is extremely large, so that convection is always fully turbulent – molecular diffusion plays no direct role.

As one might expect, turbulent convection consists of a spectrum of turbulent eddies with a strong correlation between vertical velocity and specific volume, so that that enthalpy is turbulently transported upward, heating the fluid aloft and cooling it at lower levels.

In chaotic systems such as convection, we usually have to give up on deterministic models of individual convective elements and turn instead to statistical descriptions of the flows. These are often the most useful descriptions for the purpose of representing the ensemble effects of the turbulence on larger scales of motion that we do wish to predict more deterministically.

Among the first questions we might raise about the nature of dry convective turbulence, is what determines characteristic scales for vertical velocity, plume radius, and buoyancy? How does the ensemble- or time-average temperature vary with altitude? How does the convection affect the temperature jump between the lower boundary and the atmosphere just above it? In confronting highly complex fluid flows to obtain answers to even simple questions such as

these, it is often useful to contemplate systems with an extremely limited set of control parameters. If they are sufficiently limited, then inferences can be made on strictly dimensional grounds, and tested against observations or laboratory experiments.

A classic example of dimensional reasoning is the case of homogeneous, isotropic three-dimensional turbulence in an infinite incompressible fluid. Kinetic energy is injected at large scales and cascades downscale ultimately to such small scales that it is dissipated by molecular diffusion. The critical hypothesis here is that above these smallest scales, molecular diffusion plays no significant role, and the only external parameter that matters is the rate of kinetic energy injection per unit mass. We must also assume that we are examining scales small enough that they are not directly affected by the large scale at which energy is presumed to be injected.

This injection rate, call it I, has the dimensions of kinetic energy per unit time per unit mass, or L^2t^{-3}, where L represents length and t represents time. Suppose we ask how the spectral density of kinetic energy per unit mass, E_k, depends on wavenumber. That quantity has dimensions L^3t^{-2}. If E_k depends only on I, then dimensionally there is only one choice: $E_k \sim I^{2/3}k^{-5/3}$, where k is the wavenumber. Thus the spectral kinetic energy density per unit mass must fall off as the 5/3 power of the wavenumber. The dimensionless constant of proportionality must be determined by experiment.

This hypothesis, by the Russian mathematician Andrei Kolmogorov in 1941, has been tested experimentally in high Reynolds Number homogeneous flows in statistical equilibrium and found to be very well verified.

If we consider scales so small that molecular viscosity, ν (dimensions L^2t^{-1}), plays a role, then we can form a length scale, l, from I and ν: $\nu^{3/4}I^{-1/4}$. This is called the *Kolomogorov length* and can be interpreted as an upper bound on the length scale of eddies influenced directly by molecular viscosity. Scales larger than this, but smaller than the scales at which energy is injected, are referred to as the inertial subrange and are the scales to which the Kolomogorov scaling applies.

It is possible to formulate an equally simple problem for convection[3]. The setup is illustrated in Figure 2.12.

[3] This formulation has been attributed to the German fluid dynamicist Ludwig Prandtl.

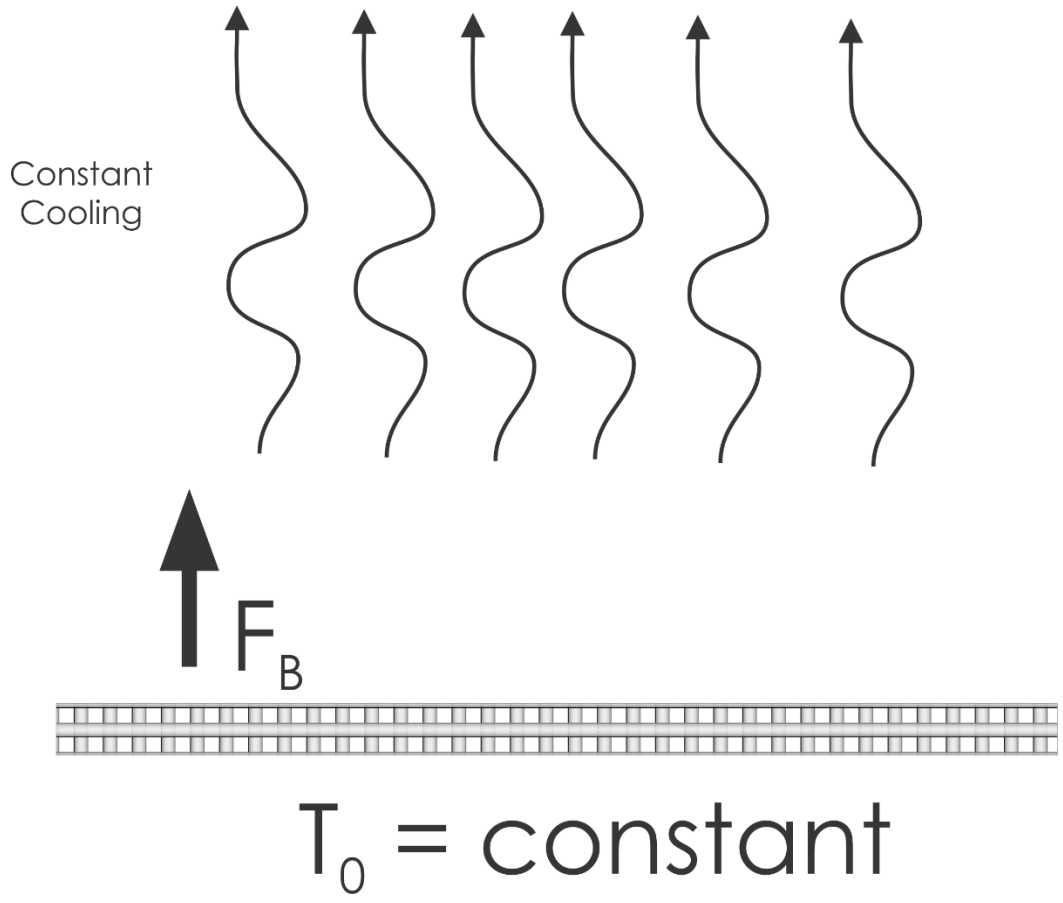


Figure 2.12: Semi-infinite, Boussinesq fluid subject to a constant rate of cooling, overlying a rigid plate held at fixed temperature T_0.

Imagine a semi-infinite, Boussinesq fluid in a constant gravitational field and cooled at some constant rate per unit mass, overlying a rigid plate at fixed temperature, T_0. A Boussinesq fluid is one that is nearly incompressible, so that density fluctuations only matter in the buoyancy term of the equations of motion. Otherwise, the fluid may be considered incompressible and, in particular, obeys the incompressible version of the mass continuity equation.

The fluid initially cools owing to the imposed cooling, but is heated from below by contact with the surface. This heating destabilizes the fluid and produces convection. After a very long time, the convergence of the upward flux of heat by the presumably turbulent convection comes into statistical equilibrium with the imposed cooling, and the system reaches a statistical equilibrium state. For overall energy balance, the heat flux F through the lower boundary must equal the vertically integrated imposed cooling rate, $\dot{Q}$:

$$F = \int_0^\infty \dot{Q} dz. \tag{2.52}$$

The surface heating produces temperature fluctuations, but according to the Boussinesq approximation, these only influence the system through the buoyancy. Thus the important external parameters is not the heat flux but the buoyancy flux:

$$F_B = g\beta F / c_p, \tag{2.53}$$

where β is the coefficient of thermal expansion of the fluid. Note in (2.52) that the imposed cooling per unit mass is infinitesimal, such that its integral through the infinite depth of the fluid is a finite constant. We might imagine that very close to the surface, the turbulent convective eddies are small, with their length scale constrained by their distance from the surface. We will assume that the boundary is thermally rough, with a typical thermal roughness scale z_0^T, a scale we will consider to be large compared to the scales at which molecular diffusion is important.

Above the roughness scale z_0^T, we shall assume, in analogy to Kolomorogov's hypothesis, that molecular diffusion plays no direct role. If that is the case, then the imposed cooling, as translated to the surface buoyancy flux per unit area (dimensions L^2t^{-3}), given by (2.53) is the only control parameter in the problem. If that is indeed the case, we can make several deductions on purely dimensional grounds:

1. The characteristic size of the turbulent eddies scales with the altitude z above the surface.
2. The characteristic buoyancy of the eddies, $g\beta|T'|$, scales as $F_B^{2/3} z^{-1/3}$.
3. The characteristic velocity of the eddies scales as $\left(zF_B\right)^{1/3}$.

While the characteristic velocity increases slowly with altitude, the characteristic buoyancy actually decreases. Note that these solutions are only valid for $z >> z_0$.

We can also make some inferences about the ensemble mean temperature gradient. On dimensional grounds,

$$g\beta \frac{d\overline{T}}{dz} = -c_1 F_B^{2/3} z^{-4/3}, \tag{2.54}$$

where c_1 is just a constant. This can be integrated in the vertical to give

$$\overline{T} = \overline{T}_0 - \frac{3c_1}{g\beta} F_B^{2/3} \left[\left(z_0^T\right)^{-1/3} - z^{-1/3} \right], \tag{2.55}$$

where we have used $\overline{T} = \overline{T}_0$ at $z = z_0^T$. The mean temperature falls off with height and approaches an asymptotic value given by

$$\overline{T}_\infty = \overline{T}_0 - \frac{3c_1}{g\beta} F_B^{2/3} \left(z_0^T\right)^{-1/3}. \tag{2.56}$$

While this simple convection problem has an elegant dimensional solution, it is difficult to test it in the laboratory. Quite apart from not being able to achieve a good approximation to an infinite Boussinesq fluid, cooling a laboratory fluid uniformly through its depth is problematic. Fortunately, in a Boussinesq system, the equations are equivalent to a system in which there is

no cooling but the boundary temperature increases linearly in time. To the author's knowledge, this has not actually been tried.

Alternatively, we can make measurements of the actual atmosphere under circumstances that approximate the idealized problem. Boundary layers over deserts can extend upward to 5 km, and one might hope that the lower portions of these, away from the boundary layer top, will not be too affected by the existence of a top. Also, air is not nearly incompressible, being close to an ideal gas. Fortunately, the Boussinesq approximation can be extended by a suitable change of the temperature variable to potential temperature, θ, defined

$$\theta = T\left(\frac{p_0}{p}\right)^{R/c_p}, \tag{2.57}$$

where p_0 is a reference pressure and R is the gas constant for (dry) air. Note that, by comparing (2.51) to (2.57), *the entropy is proportional to the natural logarithm of the potential temperature*. In this case, θ (or entropy) rather than T is the adiabatic invariant of the system, and buoyancy is defined

$$B = g\frac{\theta'}{\overline{\theta}}, \tag{2.58}$$

where the prime denotes a fluctuation away from the mean state, denoted by the overbar. The Boussinesq approximation can be shown to be valid in this case for convective layer depths much smaller than the density scale height of the atmosphere, which is roughly 8 km. Thus boundary layers extending to 5 km over deserts do not satisfy the Boussinesq approximation very well.

A far more serious problem, however, is that close to the surface, most of the turbulence kinetic energy is generated mechanically by wind flowing over terrain rather than by convection. We shall return to this problem in Chapter 4. For now we will look at observations in a location with very strong surface heating and weak ambient winds. Figure 2.13 shows measurements of potential temperature over a desert during the middle of the day, made using an instrumented model airplane. The red dots show potential temperature in air moving upward while the blue dots are for downward moving air. The magenta curve is a solution of a relation formed by combining (2.55) and (2.56) and using potential rather than absolute temperature:

$$\overline{\theta} = \theta_0 - \left(\theta_0 - \overline{\theta}_\infty\right)\left(1 - \left(\frac{z_0^T}{z}\right)^{1/3}\right). \tag{2.59}$$

We plot the solution for $T_0 = 330\ K,\ \overline{T}_\infty = 316.2\ K,$ and $z_0^T = 0.01\ m.$ To be sure, this is a weak test of the theory, but both the observations and the theory show that, except in a very thin layer near the surface, *dry convection produces a mean state in which the potential temperature is nearly constant with height*. (When there is significant variability of water vapor concentration, as in the tropical atmosphere, we must modify this statement to account for the effect of water vapor on specific volume. In Chapter 3 we show that accounting for this, in the absence of phase change of water, convection produces a mean state in which the *virtual* potential temperature is constant with height.)

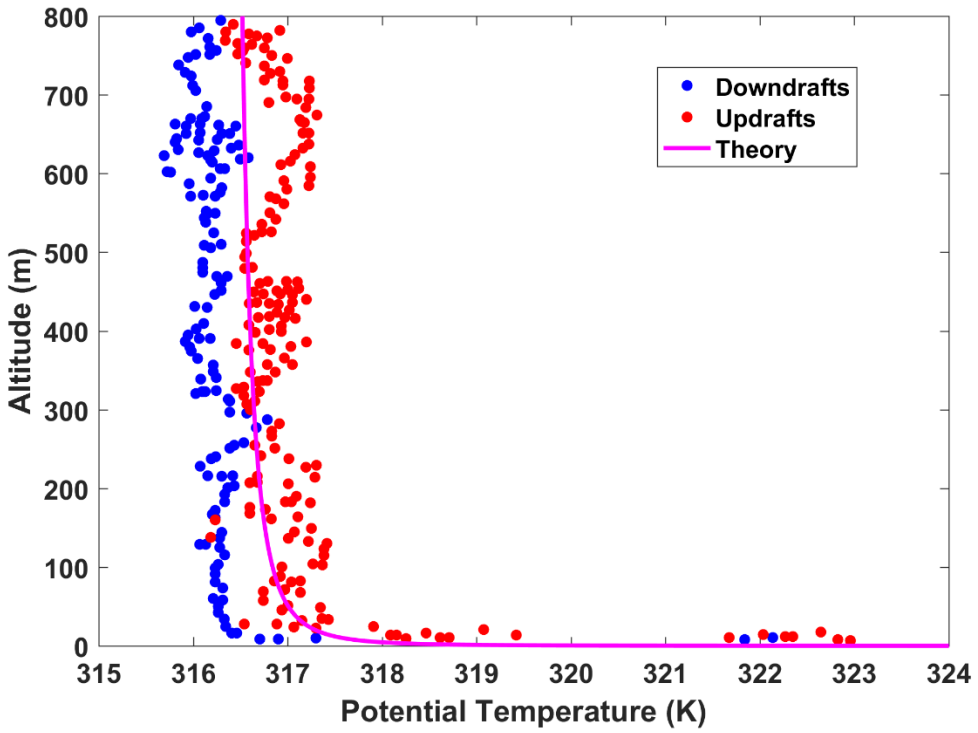


Figure 2.13: Model aircraft measurements of potential temperature during a sunny day over a desert near Albuquerque, New Mexico, on 3 August 1993. Red dots show upward moving air; blue dots show downward moving air. Magenta curve shows the theoretical solution (2.55) with $T_0 = 330\ K$, $z_0^T = 0.01\ m$, *and* c_1 *chosen so that* $\overline{T}_\infty = 316.2\ K$.

While this elegant and simple problem illustrates some of the most fundamental features of dry convection, real tropical boundary layers are strongly influenced by the presence of mean winds, which in practice drive much of the turbulence near the surface. We shall discuss this in much more detail in Chapter 3. We shall see that the general statement that (virtual) potential temperature tends to be constant with altitude in convecting boundary layers remains true even in the presence of mean wind. For now we make use of this observation to illustrate dry convective-radiative equilibrium.

2.4 Dry convective-radiative equilibrium

We next turn to a model of dry convective turbulence in statistical equilibrium with radiative forcing. Fundamentally, we have to add convective heat transfer to radiative transfer, making sure also to account for convective transfer from the surface to the atmosphere.

One way to think about this is to recognize, from Figure 2.11, that radiation tries to drive the atmosphere to a state that is unstable to dry convection in much of the troposphere. Convection, on the other hand, transports enthalpy upward, trying to restabilize the atmosphere. This tug-of-war results in a statistical equilibrium state in which the sum of the radiative (including solar) and convective fluxes vanishes at each altitude.

But this is far from an equal competition, because in our atmosphere, the time scale for perturbations to relax back to radiative equilibrium is much larger than the time it takes convection to erase entropy gradients.

Let's try to develop characteristic time scales for relaxation to radiative equilibrium and to convective neutrality.

For radiation, we will return to our simple gray-body model given by (2.29) – (2.31). We will start with the equilibrium solution given by (2.34) – (2.36) and add infinitesimal temperature perturbations to it, denoted here by primes. The change in enthalpy at a given altitude is just the convergence of the net flux of energy:

$$\rho c_p \frac{\partial T'}{\partial t} = -\frac{\partial F}{\partial z} = -\frac{\partial \tau_v}{\partial z}\frac{\partial F}{\partial \tau_v} = -\frac{\partial \tau_v}{\partial z}\left(-U' - D' + 8\sigma \overline{T}^3 T'\right), \tag{2.60}$$

where $\overline{T}$ is the radiative equilibrium temperature (which is a function of τ_v). In deriving the right side of (2.60) we used (2.29) and (2.32), linearizing about the equilibrium state and assuming that the absorption coefficient is not a function of temperature. Now eliminating U' and D' using the linearized versions of (2.30) and (2.31) yields

$$\left(\frac{\partial^2}{\partial \tau_v^2} - 1\right)\left(\frac{\partial p}{\partial \tau_v}\frac{c_p}{g}\frac{\partial T'}{\partial t}\right) = \frac{\partial^2}{\partial \tau_v^2}\left(8\sigma \overline{T}^3 T'\right). \tag{2.61}$$

Here we have also used hydrostatic balance: $\frac{\partial \tau_v}{\partial z} = -\rho g \frac{\partial \tau_v}{\partial p}$. We are not going to actually solve (2.61) but rather use it to get an estimate of the radiative relaxation time scale. Suppose the temperature perturbations oscillate quasi-sinusoidally over a characteristic optical depth $\Delta\tau_v$. We also neglect the time dependence of the lower boundary condition, equivalent to assuming that the surface temperature is fixed. Then it follows from (2.61) that a characteristic time scale for radiative relaxation is

$$t_{rad} \cong \frac{c_p\left(1 + \left(\Delta\tau_v\right)^2\right)}{8\sigma g \overline{T}^3 \frac{-\partial \tau_v}{\partial p}}. \tag{2.62}$$

Clearly the shortest time scales will be associated with temperature perturbations that vary rapidly with altitude, so that $\Delta\tau_v << 1$. For these, and calculating $\partial\tau_v / \partial p$ from (2.40), we get

$$t_{rad} \cong \frac{c_p p_0}{8\sigma g \overline{T}^3 \tau_{v\infty}}. \tag{2.63}$$

For our atmosphere, this time scale is around 5 days. At the opposite limit, when the perturbations vary over long optical depths, the time scale would be around 35 days. (Later we will show that the characteristic radiative relaxation time scale in a moist atmosphere is somewhat longer, and if the surface temperature is also calculated, longer still.)

By contrast, a characteristic convective adjustment time scale is just the time it takes a turbulent plume to traverse the depth, H, of an unstable layer. This is just H divided by a characteristic vertical velocity scale. For the simple convection problem we solved earlier, this vertical velocity scale is $\left(F_b H\right)^{1/3}$, so the characteristic time scale is $\left(H^2 / F_B\right)^{1/3}$. For typical values of the buoyancy flux and depth, this is a few hours, roughly two orders-of-magnitude faster than the radiative relaxation time. Thus in our atmosphere, in the competition between radiation and convection, the convection "wins" and the equilibrium state is much closer to convective neutrality than to radiative equilibrium. We expect radiative-dry convective equilibrium to be characterized by nearly constant (virtual) potential temperature.

For this reason, the simplest conceivable way to account for convective heat transfer within the atmosphere is simply to adjust the profile of potential temperature to a constant wherever radiation forces it to decrease with altitude. We assume no external energy sources operate during the adjustment, so that the total energy content of the column is invariant during the adjustment. We assume that any kinetic energy generated by the convection is locally dissipated back into enthalpy, thus the mass-weighted vertical integral of the enthalpy must remain constant during the adjustment. So if the layer between pressure levels p_1 and p_2 has decreasing potential temperature, θ, then it would be adjusted so that

$$\theta_{new} = c + \frac{1}{p_2 - p_1}\int_{p_1}^{p_2} \theta_{old} dp', \tag{2.64}$$

where c is a constant determined so that enthalpy is conserved during the adjustment:

$$\int_{p_1}^{p_2} c_p T_{new} dp' = \int_{p_1}^{p_2} c_p T_{old} dp'. \tag{2.65}$$

The adjustment will of course increase the potential temperature at the top of the initially unstable layer, and decrease it at the bottom. In so doing, it may cause the potential temperature to decrease between the top of the initially unstable layer and the air just above it, and/or between the air just below the layer and the base of the layer. So these will have to be adjusted as well. In general, this *dry adiabatic adjustment* is an iterative process.

The thermodynamic disequilibrium between the surface and the air in contact with it, a general feature of radiative equilibrium, will give rise to a convective heat flux from the surface to the atmosphere, further altering the equilibrium state. In the case of pure convection, if we know the surface temperature and the potential temperature away from the surface, we can turn (2.56) around algebraically and model the surface buoyancy flux (which is proportional to the surface heat flux in a dry atmosphere) as

$$F_B = \left[\frac{g}{3c_1\theta_s}\left(\theta_s - \overline{\theta}_a\right)\right]^{3/2}\left(z_0^T\right)^{1/2}, \tag{2.66}$$

where θ_s is the surface potential temperature, $\overline{\theta}_a$ is the ensemble mean potential temperature some distance above the surface, and we have made use of the fact that the coefficient of thermal expansion of an ideal gas is its inverse temperature. But as discussed in the next chapter, turbulence near the surface is dominated by shear production when even a weak wind

is present, and in cases where shear production dominates, the heat flux from the surface is better modelled as

$$F_s = C_k \rho \, |\, \overline{\mathbf{V}} \,| \left(T_s - \overline{T}_a \right), \tag{2.67}$$

where C_k is a dimensionless transfer coefficient that depends on surface roughness but is of order 10^{-3}, ρ is the near-surface air density, $|\, \overline{\mathbf{V}} \,|$ is the ensemble mean speed of the horizontal wind a short distance above the surface, T_s is the surface temperature, and $\overline{T}_a$ is the ensemble mean air temperature a short distance above the surface.

To illustrate the profound effect of convection on the equilibrium state, we will apply dry convective adjustment to the same radiative model used to produce Figures 2.11 and 2.12, and use (2.67) with a fixed wind speed of $5\, ms^{-1}$ to calculate surface sensible heat flux into the atmosphere. The fixed relative humidity profile is the same as was used in the calculation of pure radiative equilibrium. The result is shown by the magenta curve in Figure 2.14 and compared with both a mean tropical sounding (dashed blue curve) and the pure radiative equilibrium solution (solid blue curve). The magenta and blue dots at the bottom show the calculated sea surface temperatures in the two cases.

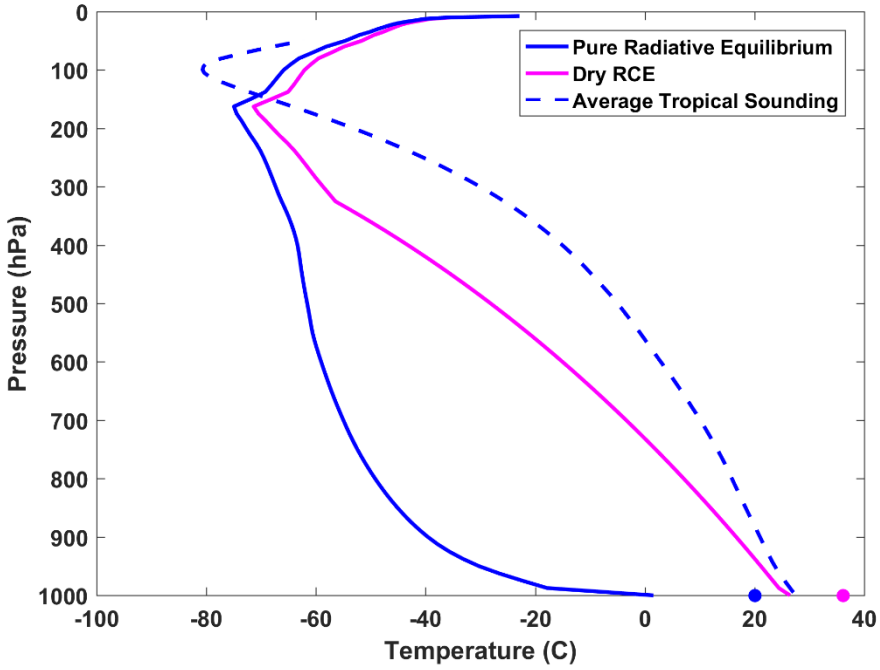


Figure 2.14: Radiative-dry convective equilibrium (magenta curve) calculated using dry convective adjustment and surface fluxes given by (2.67). This is compared to the pure radiative equilibrium solution (solid blue) and mean tropical temperature profile (dashed blue). The magenta and blue dots at the bottom show the calculated sea surface temperatures for the adjusted and pure radiative cases, respectively.

The temperature profile is dry adiabatic up to about 300 hPa and the whole profile is considerably warmer, and closer to the tropical mean sounding, than the pure radiative equilibrium solution. The surface temperature is close to 36° C, considerably warmer than the pure radiative equilibrium solution and much warmer than observed sea surface temperatures in the tropics, which peak at around 30° C. The tropopause and the lower stratosphere are a bit warmer than the pure radiative equilibrium solution.

The addition of a turbulent heat flux from the surface to the atmosphere cools the surface and heats the atmosphere; this is one reason that the adjusted profile is warmer that the pure radiative equilibrium solution. This effect is amplified by water vapor feedback: The warmer state, under the assumption of constant relative humidity, has more water vapor – an important greenhouse gas. Thus the system warms further.

Note that heat transfer from the surface reduces but does not eliminate the temperature jump between the surface and the atmosphere.

Figure 2.15 shows the rates of radiative and dry convective heating in equilibrium, expressed as degrees of temperature change per day. In equilibrium, the two must sum to zero and so they are mirror images of each other. The strong turbulent heat flux from the surface to the atmosphere causes so much warming that the whole troposphere up to about 300 hPa cools radiatively, at a remarkably constant rate of roughly $1.5\ K\ day^{-1}$, though a bit larger close to the surface.

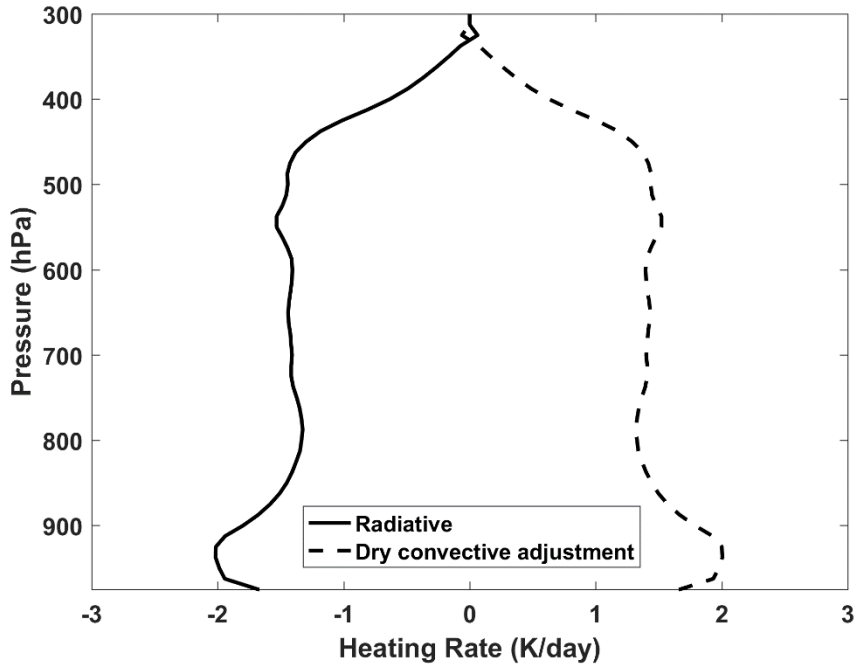


Figure 2.15: Radiative (solid) and convective (dashed) heating rates in the radiative-dry convective equilibrium solution corresponding to Figure 2.14.

In reality, dry convection would rapidly homogenize the mass concentration of water (specific humidity) through the convecting layer, likely producing clouds in the upper part of the layer and thereby radically altering the radiative transfer. Nonetheless, the equilibrium state illustrated in Figures 2.14 and 2.15 serves to contrast the pure radiative equilibrium with a convectively adjusted equilibrium and sets the stage for the real problem, moist convection.

2.5 Moist RCE, Part I

The dynamics and physics of the tropical atmosphere are made vastly richer by phase changes of water. Most of these phase changes occur within cumulus and cumulonimbus clouds, whose presence is much of what makes tropical skyscapes so beautiful. Water plays a dual role: latent heat released and absorbed when water changes phase makes a first-order contribution to atmospheric thermodynamics, and at the same time, water in all its phases has a strong effect on radiative transfer. As we shall see, this makes even relatively simple problems like radiative-convective equilibrium daunting, but far richer and more interesting than their dry counterpart. Our currently incomplete understanding of moist convection makes tropical meteorology still something of a frontier subject in atmospheric science and climate.

The subject of moist convection is far too broad and deep to treat in anything like a comprehensive way in a book that attempts to cover the landscape of tropical meteorology; all we can do here is review some essential elements. In my years of teaching tropical meteorology at the graduate level I have found that dislodging cherished but flawed concepts absorbed at the undergraduate level more challenging than conveying wholly new material.

One false conception that is particularly difficult to dislodge is the idea that large-scale tropical circulations are driven by latent heat release. This idea is widely expressed in textbooks and in research literature. In the field of tropical cyclones this is expressed in the notion that such storms are driven by latent heat release within the deep cumulonimbi that surround the cyclone center. This idea is just as false as the statement that elevators are driven by counterweights, and for the same essential reason. It fails to recognize that convection in the one case, and the transmission of tension through steel cables in the other, are fast processes, and that the system dynamics are rate-limited by slower processes….radiation and surface enthalpy fluxes in the tropical atmosphere and the electric motor in the case of the elevator. We will begin by illustrating this point with a simple example. But first, we review some essential points about the thermodynamics and microphysics of moist convection:

1. Condensation and evaporation of cloud condensate are nearly reversible processes. The very small supersaturations needed for heterogeneous nucleation ensure that, except in rare circumstances, supersaturations are very small and water vapor and cloud condensate may be considered to be in thermodynamic equilibrium with each other.

2. For this reason, condensation and evaporation can be considered internal processes that can best be handled by a redefinition of the adiabatic invariants of the system. They should not be treated in the same way as external heat sources and sinks. (This is not the case, however, for evaporation of precipitation, which is not usually a thermodynamic equilibrium process.)

3. The exotic nature of moist convection is therefore not owing to phase change per se but to key irreversible processes: the formation, fallout, and partial or total re-evaporation of precipitation, and mixing across sharp gradients of water concentration.

To illustrate these points, we consider a variation on the very simple convection problem we considered in section 2.3. In this variation, we take the atmosphere to be saturated with respect to water vapor, and filled with cloud, such that the total water mass concentration per unit mass of dry air, r_t (also called the total water *mixing ratio*) is constant. We do not allow any precipitation or flux of water through the lower boundary, so that r_t is a locally conserved variable. As in the original problem, we cool the atmosphere and keep the surface temperature fixed, so that the whole system is turbulently convecting, and this turbulent convection thoroughly mixes the total water concentration, which consequently is constant. Temperature will be decreasing with height, and thus, according to the Clausius-Clapeyron equation, the saturation vapor pressure decreases with height. Thus as we ascend, more and more of the total water concentration is in the form of cloud, and less and less in the form of vapor. This is illustrated by the shading in Figure 2.16.

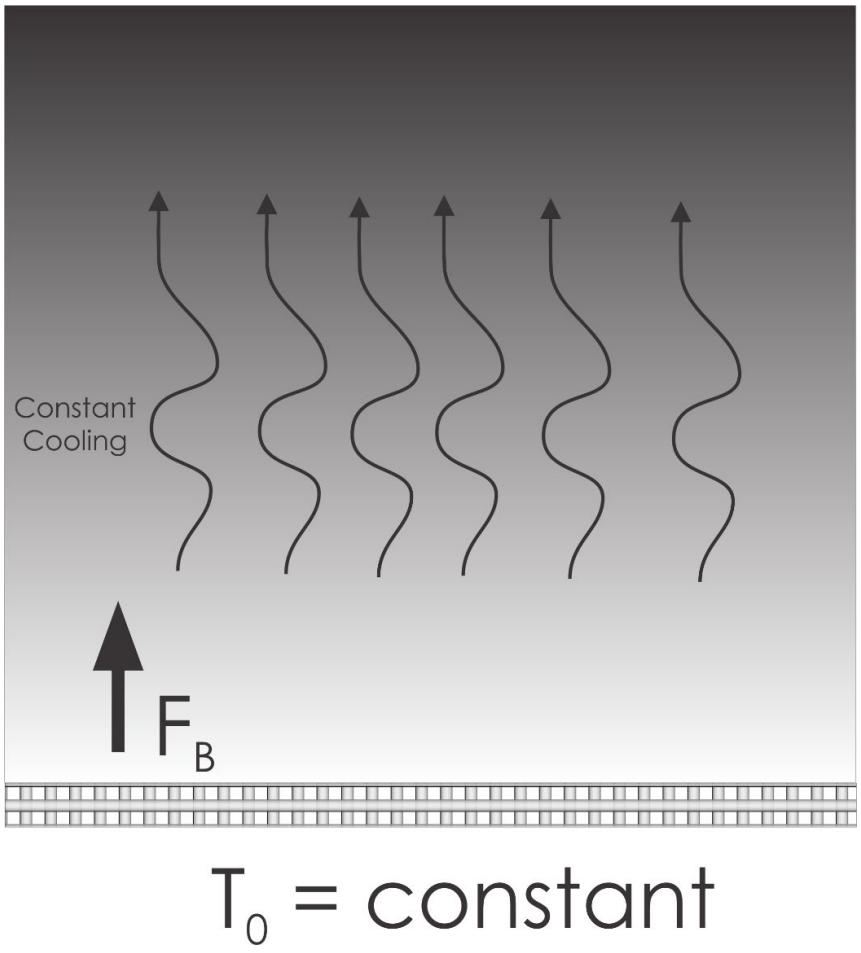


Figure 2.16: As in Figure 2.12 but for a water-saturated atmosphere with constant total water concentration, but whose cloud water content (shading) increases with altitude.

For this inhomogeneous gas, any state variable may be expressed as a function of any other *three* state variables (as opposed to two for homogeneous systems). Thus, if we like, we can express the specific volume as a function of the system entropy, pressure, and total water mixing ratio:

$$\alpha = \alpha\left(s, p, r_t\right), \tag{2.68}$$

where here the entropy, s, is the entropy of a mixture of dry air, water vapor, and liquid water. (We ignore the ice phase here.) This may be written (e.g. see Emanuel, 1994) as

$$s = \left(c_{pd} + r_t c_l\right)\ln(\frac{T}{T_0}) - R_d \ln(\frac{p_d}{p_0}) + \frac{L_v r}{T} - rR_v \ln(\mathcal{H}) + c_l r_t, \tag{2.69}$$

where c_{pd} is the heat capacity at constant pressure of dry air, c_l is the heat capacity of liquid water, r is the mixing ratio (mass of water vapor per unit mass of dry air), r_t is the total water mixing ratio (mass per unit mass of dry air), R_d is the gas constant for dry air, p_d is the partial pressure of dry air, L_v is the latent heat of vaporization, R_v is the gas constant for water vapor, and $\mathcal{H}$ is the relative humidity, defined as the ratio of actual to saturation vapor pressure. As we may add arbitrary constants to the definition of entropy, we have included reference temperature and pressure, T_0 and p_0, and also added the very last term which is a function of another conserved variable; this makes what follows more compact. In our problem, the air is always saturated – the relative humidity is 1 – and so the second to last term in (2.69) is absent.

Using the chain rule, we can write perturbations to the specific volume, at constant pressure, as

$$\alpha' = \left(\frac{\partial \alpha}{\partial s}\right)_{p,r_t} s', \tag{2.70}$$

remembering that the total water concentration is constant in the problem. Now, as in the dry case, we can develop a Maxwell relation from the definition of specific enthalpy, which for a mixture of dry air, water vapor, and liquid water may be written (Emanuel, 1994):

$$k = \left(c_{pd} + r_t c_l\right)T + L_v r. \tag{2.71}$$

By differentiating (2.69) and (2.71), we can show that

$$dk = Tds - c_l T \ln(\frac{T}{T_0})dr_t + \alpha\left(1 + r_T\right)dp. \tag{2.72}$$

By applying to (2.72) the following identity:

$$\left(\frac{\partial}{\partial s}\right)_{p,r_t}\left(\frac{\partial k}{\partial p}\right)_{s,r_t} = \left(\frac{\partial}{\partial p}\right)_{s,r_t}\left(\frac{\partial k}{\partial s}\right)_{p,r_t}, \tag{2.73}$$

we arrive at the desired Maxwell relation:

$$\left(\frac{\partial \alpha}{\partial s}\right)_{p,r_t} = \frac{1}{1+r_t}\left(\frac{\partial T}{\partial p}\right)_{s,r_t}, \tag{2.74}$$

which is practically identical to the dry case (2.45) except for the different definition of entropy and the factor $1+r_t$, which is close to unity in our atmosphere. From (2.70), the buoyancy, defined by (2.48), is then

$$B = \frac{\Gamma_m}{1+r_t} s', \tag{2.75}$$

where $\Gamma_m = -\left(\frac{\partial T}{\partial z}\right)_{s,r_t}$ is the *moist adiabatic lapse rate*. This is very similar to the dry case (2.49) except for the different definition of entropy, the $1+r_t$ factor, and the moist rather than dry adiabatic lapse rate.

It is important to note that the moist adiabatic lapse rate varies with altitude, in contrast to the dry adiabatic lapse rate, which is a constant, assuming that the acceleration of gravity does not vary much over the depth of the convecting layer. Thus, according to (2.75), for a given entropy fluctuation, the buoyancy will vary depending on the altitude and it is no longer the case that the buoyancy flux is related to the heat flux by a fixed constant. Consequently, it would appear that the dimensional reasoning we used to deduce how various quantities vary with altitude in the dry case will not work for the present case, because there is an additional non-constant parameter, Γ_m.

To see the problem, let's look at the equation for the time mean, horizontally averaged turbulence kinetic energy. This is formed by representing all the quantities (velocities, buoyancy, and pressure) by the sum of a time mean, horizontally averaged component plus a fluctuating part. Symbolically, for any variable v,

$$v = \overline{v}(z) + v'(x,y,z,t). \tag{2.76}$$

We apply this decomposition to all the variables, substitute them into the Boussinesq equations, and form from them an equation for the kinetic energy. We then take the time and horizontal averages of this equation. Neglecting the molecular diffusion terms, as we did in the dry case, the result is

$$\frac{\partial}{\partial z}\left[\overline{w'^3} + \alpha_0 \overline{p'w'}\right] = \overline{w'B'} = \frac{\Gamma_m}{1+r_t}\overline{w's'}, \tag{2.77}$$

where, as a result of the Boussinesq approximation, α_0 is a representative (constant) value of the specific volume and the overbars represent time and horizontal means. The far right side of (2.77) uses (2.75) for the buoyancy. Now to within the Boussinesq approximation, the convective heat flux, F, is just $T_0\overline{w's'}$, where T_0 is a representative temperature. Thus we may write (2.77) as

$$\frac{\partial}{\partial z}\left[\overline{w'^3}+\alpha_0\overline{p'w'}\right]=\frac{\Gamma_m F}{\left(1+r_t\right)T_0}. \tag{2.78}$$

The problem is that the right side of (2.78) varies with altitude and so we no longer have a one-parameter problem. However, we can reduce this back to a problem involving a single constant parameter by defining a new vertical coordinate, μ, that is linear in absolute temperature, rather than altitude, but has the same dimensions as altitude:

$$\mu \equiv \frac{T_0-\overline{T}}{\Gamma_d}. \tag{2.79}$$

Here $\overline{T}$ is the absolute temperature along a moist adiabat (that is, a curve of constant entropy). We have from this that

$$\frac{\partial}{\partial z}=\frac{\partial}{\partial \mu}\frac{\partial \mu}{\partial z}=\frac{\Gamma_m}{\Gamma_d}\frac{\partial}{\partial \mu}. \tag{2.80}$$

Applying (2.80) to (2.78) transforms the latter to

$$\frac{\partial}{\partial \mu}\left[\overline{w'^3}+\alpha_0\overline{p'w'}\right]=\frac{\Gamma_d F}{\left(1+r_t\right)T_0}\equiv F_B. \tag{2.81}$$

Now the right had side is a constant, which we will call F_B, and we are back to a one parameter problem. All the conclusions that we reached in the dry case apply to this case as well, with a very slightly different definition of F_B and with the alternative vertical coordinate defined by (2.79). Among these conclusions are

4. The characteristic size of the turbulent eddies scales with the modified altitude μ above the surface.
5. The characteristic buoyancy of the eddies scales as $F_B^{2/3}\mu^{-1/3}$.
6. The characteristic velocity of the eddies scales as $\left(\mu F_B\right)^{1/3}$.

Note from (2.80) that quantities vary more slowly with z than they do with μ, because $\Gamma_m<\Gamma_d$. For a surface temperature of 30 °C, the variations with z would be only about 1/3 as fast as variations with μ, but traveling upward into lower temperatures, the moist adiabatic lapse rate increases and asymptotically approaches the dry adiabatic lapse rate. Thus, for low surface temperatures or in the upper troposphere, μ is essentially equivalent to z.

It can be seen from this simple example that the phase change of water does not materially change convection. Moist convection behaves pretty much like dry convection. One can always revert to a description of this system in terms of dry entropy (or potential temperature), in which case the thermodynamic equation would have large sources and sinks owing to condensation and evaporation. One could say that thermals are driven upward by latent heat release and downward by absorption of latent heat, but this would be a clumsy way of describing it. In

general, dynamical systems are described most compactly and intuitively by using the most conserved variables available.

What makes real moist convection *qualitatively* different from dry convection is not latent heat release per se, but rather the strongly irreversible processes of the fall of precipitation and its re-evaporation into unsaturated air.

Suppose now that, in our simple convecting system, a demon removes condensed water as soon as it forms, so that while upward moving air is generally saturated, it contains a very small amount of condensed water (just enough to make the condensate visible as cloud). A slight downward displacement, by warming the sample, renders it unsaturated. Thus, broadly speaking, most upward moving air is saturated and experiences latent heat release, while most downward moving air is unsaturated. The form the convection takes is illustrated by Figure 2.17.

Figure 2.17: Configuration of clouds in an atmosphere in which a uniform cooling is imposed above a liquid water surface held at fixed temperature. A demon removes condensed water as soon as it forms, leaving just enough to render the saturated air visible as cloud.

The upward-moving saturated air, were it not for mixing, would follow a *pseudo-adiabatic* lapse rate of temperature. This is a defined by a process in which displaced air remains saturated, but any condensate that forms is removed. Condensate therefore does not contribute to the effective buoyancy of the air, nor does it contribute to it heat capacity. Since at each step the system's state is defined uniquely by its temperature and pressure (with its water vapor content determined by Clausius-Clapeyron), it is possible to define a pseudo-entropy that is exactly conserved in a pseudo-adiabatic process. Although it is not possible to derive an exact closed-form expression for the pseudo entropy, an empirical expression valid over the observed range of tropospheric temperature and pressure was developed by Bolton (1980). He chose to express it in terms of a temperature-like variable called the pseudo *equivalent potential temperature*, θ_{ep} , whose logarithm is proportional to the pseudo entropy:

$$\theta_{ep} = T\left(\frac{1000}{p}\right)^{0.2854(1-0.28r)} \times \exp\left[r(1+0.81r)\left(\frac{3376}{T^*} - 2.54\right)\right], \tag{2.82}$$

where r is the mixing ratio and T^* is the *saturation temperature*; that is, the temperature at which air becomes saturated when displaced adiabatically. This is accurate to within 0.3 K for tropospheric conditions. Note that θ_{ep} is always larger than θ and that $\lim_{r\to 0}\theta_{ep} = \theta$; the pseudo-equivalent potential temperature asymptotically approaches ordinary potential temperature as the mixing ratio vanishes.

Since pseudo-entropy is a function of temperature and pressure alone when the air is saturated (and thus the mixing ratio, r, is at is saturated value, which is itself a function of temperature and pressure alone), lines of constant saturation pseudo-entropy, usually referred to as *moist adiabats*, can be graphed on a thermodynamic diagram, as shown in Figure 2.18.

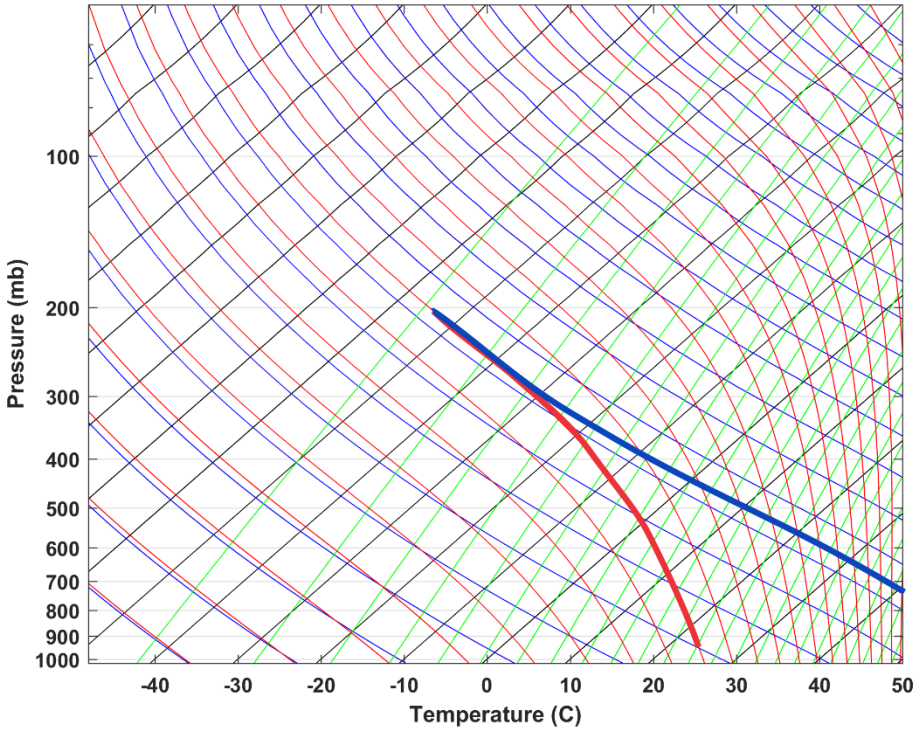


Figure 2.18: A typical atmospheric thermodynamic diagram. This particular format is known as a Skew-T-log-p diagram, because isotherms run diagonally rather than vertically (so that typical atmospheric temperature profiles are more erect). Pressure forms the ordinate, at equal intervals of the logarithm of the pressure. The blue curves are dry adiabats (curves of constant θ) and the red curves are pseudo- moist adiabats (curves of constant θ_{ep}). Curves of constant saturation mixing ratio are shown in green. The heavy red lines shows an example of saturated, pseudo-adiabatic ascent from 950 hPa and 25 °C to 200 hPa while the heavy blue curve illustrates dry adiabatic descent.

The particular thermodynamic diagram illustrated in Figure 2.18 is called a *Skew-T* diagram because isotherms run diagonally rather than vertically; this is done so that typical atmospheric temperature profiles are more vertical. Pressure forms the ordinate, at equal intervals of the logarithm of the pressure. The blue curves are dry adiabats (curves of constant θ) and the red curves are pseudo- moist adiabats (curves of constant θ_{ep}). Curves of constant saturation mixing ratio are shown in green.

Suppose we consider a sample of air that is saturated at 25 °C and 950 hPa and lift it by a pseudo-adiabatic process to 200 hPa. This is shown by the heavy red line in Figure 2.18. Now take that air and increase its pressure without adding or subtracting energy. Because it never has any condensed water in it, there is no evaporation, and the sample descends dry adiabatically, as shown by the thick blue curve in Figure 2.18.

Note that the descending air is warmer than the ascending air, so that the rising air is negatively buoyant with respect to the descending air. This process would convert kinetic to potential energy and so would not happen spontaneously. (Were the air to descend all the way back down to 950 hPa, its temperature would be 88 °C; the difference between this and the starting temperature of the sample reflects the latent heat added to the sample during its ascent.)

For this reason, there is a problem with the picture of moist RCE shown by Figure 2.17. For the descending air to have a temperature equal to or lower than that of the ascending stream, it must lose energy, and in reality it does so by radiative cooling. Observations, which we shall review in the next section, show that there is in fact very little difference between the temperature of the ascending and descending streams, and that the temperature profile in the clouds is not far from moist adiabatic. The thermodynamic balance in the descending air is then

$$c_p \rho w_e \frac{T}{\theta}\frac{d\theta}{dz} = \dot{Q}_{rad}, \tag{2.83}$$

where w_e is the vertical velocity and $\dot{Q}_{rad}$ is the rate of radiative heating in the unsaturated air. As we shall show, $\dot{Q}_{rad}$ is almost always negative, consistent with descent in between clouds. We expect that the potential temperature stratification, $d\theta / dz$, is determined by the condition that the temperature profile be nearly moist adiabatic. Thus (2.83) really determines the descent rate in the clear air, w_e, given the rate of radiative cooling.

In RCE, there can be no net mass flux through any level, otherwise the air density would be changing over time. Suppose that updrafts have a characteristic upward velocity w_c and cover a fractional area σ. Then, ignoring density differences between the upward and downward moving air, the flux of mass through any level, which must vanish, is given by

$$\sigma \rho w_c + (1-\sigma)\rho w_e = 0. \tag{2.84}$$

For convenience, we define M to be the net upward convective mass flux per unit area:

$$M \equiv \sigma \rho w_c = -(1-\sigma)\rho w_e. \tag{2.85}$$

Observations show that $\sigma << 1,$ so that $M \cong -\rho w_e.$ Since the clear-air velocity is determined by the rate of radiative cooling and the potential temperature stratification along a moist adiabat, it follows that the net upward convective mass flux in moist RCE is very nearly determined by these two quantities.

Note that where the convective mass flux, as given by (2.85) together with (2.83), decreases with altitude, mass must flow from the convective updrafts into the environment. Therefore, the water budget in the sinking air in between clouds (which occupies the majority of the volume of the system since $\sigma << 1$) is given by

$$\rho w_e \frac{dr_e}{dz} = \left(r^* - r_e \right) \mathrm{MAX}\left(-\frac{dM}{dz}, 0 \right), \tag{2.86}$$

where r_e is the mixing ratio in the clear air and $r*$ is the saturation mixing ratio in the cloud. If the profile of radiative cooling and the system temperature are known, then (2.83) can be solved for w_e and thus M , using (2.85), and then (2.86) can be solved for the mixing ratio of the clear air, r_e.

But note that this would require iterating through (2.83), (2.85), and (2.86) because the radiative heating rate depends strongly on the clear-air water vapor content. This illustrates the strongly nonlinear nature of moist RCE – radiative cooling drives the convection, which in turn determines the moisture (and, indirectly, the clouds) of the free troposphere and thereby strongly affects the radiative cooling.

In the case of radiative cooling that is constant with altitude, (2.83) shows that w_e and thus M would increase with altitude, since the potential temperature stratification decreases upward along moist adiabats. In that case, the right side of (2.86) would be zero except for a delta function at the tropopause. Thus the mixing ratio of the whole troposphere above the subcloud layer would be equal to the saturation mixing ratio of the tropopause. Given that tropical tropopause temperatures are around -70 ^{o}C, this would be miniscule, and the relative humidity would thus be close to zero except very close to the tropopause. This is very far from what is observed in the tropics, or what is produced by numerical simulations of idealized moist RCE states, about which more in due course.

Since in the real world, radiative cooling profiles are roughly constant with altitude in the troposphere, the excessive dryness of solutions to (2.86) is likely not owing to the lack of convective detrainment but almost certainly to our extreme assumption that all condensed water is removed from the system. This strong, highly artificial sink of water is responsible for the excessive dryness of the solution to (2.86). We must modify our simple picture of RCE to include the microphysics governing precipitation formation, fall, and re-evaporation.

So we revise the conceptual picture of RCE of Figure 2.17 to include formation, fallout, and partial re-evaporation of precipitation, as shown in Figure 2.19. Saturated air ascends with mass flux M_u and re-evaporation of falling precipitation drives downward motion of mostly unsaturated air with mass flux M_d . We here redefine the “clear air” as the slowly subsiding air outside both the clouds and the unsaturated downdrafts. If we define the total convective mass

flux as M and the "clear air" as excluding both the cloud and rain areas, then (2.83) and (2.85) apply as with the pseudo-adiabatic case.

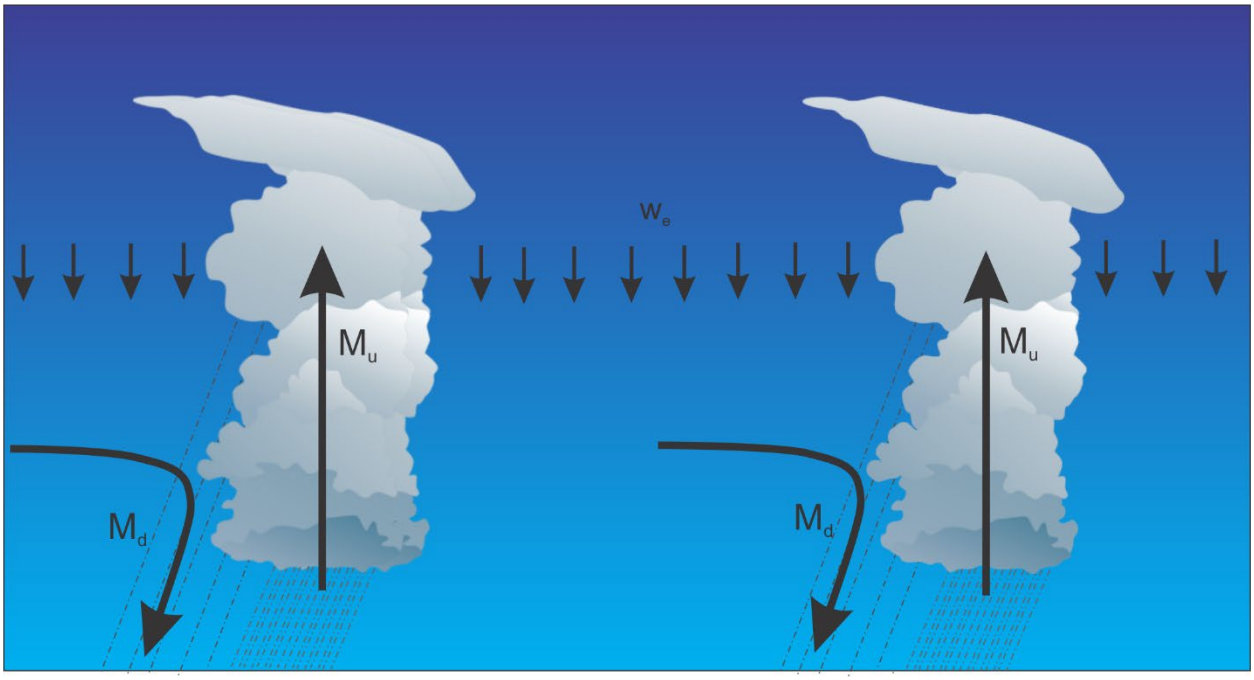


Figure 2.19: Moist RCE with precipitation. Broadly, saturated air ascends with mass flux M_u and re-evaporation of falling precipitation drives downward motion of mostly unsaturated air with mass flux M_d. We here redefine the "clear air" as the slowly subsiding air outside both the clouds and the unsaturated downdrafts.

Rather than focus on the water budget of just the clear air, we will look at the flux of water through each level across the whole system. That flux must vanish, since in equilibrium water cannot accumulate above any level. This condition is given by

$$M\left(r_t - r_e\right) - P = 0, \tag{2.87}$$

where r_t is the total water mixing ratio (including cloud water) inside the clouds, and P is the total downward precipitation flux through the level in question. Rearranging this and using (2.85) for M under the approximation that $\sigma << 1$ gives

$$r_e = r_t - \frac{P}{\rho w_e}, \tag{2.88}$$

with ρw_e given by (2.83). This shows that the humidity of the clear air is governed by the microphysics that determine how much cloud water exists at a given level (which sets the amount by which r_t exceeds r^*) and how much precipitation is formed. More cloud water yields a moister environment, while more precipitation dries the environment.

We can make this more intuitive with a simple example. Suppose that a fixed fraction ϵ_p of water condensed at any level is converted to precipitation, and that another fixed fraction, ϵ_r of

the precipitation evaporates before reaching the level in question. The precipitation flux through any level z is then given by

$$P = -\left(1-\epsilon_r\right)\int_z^{z_t} \epsilon_p M \frac{dr^*}{dz} dz, \tag{2.89}$$

where z_t is the altitude of the clouds tops. The integrand in (2.89) is just the precipitation formation efficiency multiplied by the rate of condensation. Again using (2.85) with $\sigma << 1$ and substituting (2.89) into (2.88) gives

$$r_e = r_t + \frac{\left(1-\epsilon_r\right)}{\rho w_e}\int_z^{z_t} \epsilon_p \rho w_e \frac{dr^*}{dz} dz. \tag{2.90}$$

One interesting special case of (2.90) is the case of ρw_e and ϵ_p constant with height, in which case (2.90) reduces to

$$r_e = r_t - \left(1-\epsilon_r\right)\epsilon_p \left(r^* - r(z_t)\right). \tag{2.91}$$

Note that we can recover the pseudo-adiabatic limit by setting $\epsilon_r = 0$, $\epsilon_p = 1$, and $r_t = r^*$. In that case, $r_e = r(z_t)$ as before.

The relation (2.91) shows that the smaller the efficiency ϵ_p with which precipitation is formed, and the greater the re-evaporation of precipitation ϵ_r, the moister the environment. Note that smaller precipitation efficiency will also elevate the cloud water content, and thus the total water mixing ratio r_t.

Thus, unequivocally, the humidity of the clear air in RCE is set by cloud microphysical and radiative processes. The rate of radiative cooling and the ambient potential temperature stratification along a moist adiabat, which together determine ρw_e, enter in a more subtle way. Note that if we simply multiply the radiative cooling everywhere by a constant factor, that factor cancels in (2.90) and there is no effect on the environmental mixing ratio. (However, the system temperature will be different, thus altering the relative humidity.) From (2.90) we can conclude that, all other things being equal, if the radiative subsidence ρw_e increases in magnitude with altitude above the level in question, the air will be drier, whereas if it decreases with altitude, the air will be moister. Always remember that the water vapor content strongly influences the radiative cooling rate, so that ρw_e depends on water. Moreover, the vapor content outside convective clouds is important in the formation of stratiform clouds, which themselves strongly affect radiative transfer in both the visible and infrared, so the actual problem of moist RCE is a strongly two-way process, with the radiation and surface fluxes driving the convection and the latter lofting water from its source of the surface and thereby influencing radiative transfer by altering the vapor and cloud water profiles outside the convective clouds themselves.

We will return to a detailed treatment of moist RCE in the following section. But next we turn to the topic of moist convection itself.

References

Bolton, D., 1980: The computation of equivalent potential temperature. *Mon. Wea. Rev.*, **108,** 1046-1053.

Emanuel, K. A., 1994: *Atmospheric convection.* Oxford Univ. Press, New York, 580 pp., translator.

Nicolis, G., and I. Prigogine, 1977: *Self-organization in nonequilibrium systems.* John Wiley & Sons, New York, 490 pp., translator.

Pierrehumbert, R. T., 2010: *Principles of planetary climate.* Cambridge University Press, New York, 652 pp., translator, 978-0-521-86556-2.

3 Moist Convection and Moist Radiative-Convective Equilibrium

3.1 Convective clouds and mesoscale organization

We have seen that the irreversible formation, fallout, and partial or total re-evaporation of precipitation completely transforms the nature of moist convection so that ascending air is generally saturated and contains condensed water, while the air in between clouds is mostly unsaturated and would descend dry adiabatically were it not for radiative cooling. Air also descends in unsaturated downdrafts cooled by the evaporation of precipitation falling through them, as illustrated in Figure 2.19. The dry stratification of the atmosphere and the relatively small rates of radiative cooling keep the subsidence velocity given by (2.83) small…less than $1\,cm\,s^{-1}$. For reasons we will return to, the convective updraft velocities are about two orders-of-magnitude stronger, so that the fractional area covered by convective updrafts implied by (2.84) is of the order of 10^{-2}. The small fraction of the sky covered by convective clouds in the tropics can be seen in the photograph in Figure 3.1

Figure 3.1: Spectrum of tropical cumulus clouds photographed from the International Space Station.

The rich tapestry of clouds in Figure 3.1 range from the barely visible shallow, non-precipitating clouds to cumulonimbi that span the depth of the troposphere, crowned by icy stratiform anvil clouds that spread out at or near the tropopause. These cumuli are responsible for most of the vertical transport of water, aerosols, and trace gases, and almost all of the non-radiative energy transport in the tropical atmosphere above a thin, sub-cloud layer. Their small size – ranging from a few hundred meters to perhaps 10 km – render them unresolvable by past and present global models, posing an immense challenge to representing their effects in such models.

The isolation of many of the convective towers in Figure 3.1 (though there are some clusters) is associated with large gradients of water concentration along their lateral boundaries. The plumes themselves are fully turbulent, and the turbulent eddies mix air from just outside the clouds into the clouds themselves, a process known as *entrainment*. This introduces large inhomogeneity in the thermodynamic properties within the cloud boundaries, which strongly affects buoyancy as well as the cloud microphysical properties that control the formation, fallout, and partial re-evaporation of precipitation. Mixing of cloudy air with unsaturated air can lead to mixtures that are colder than either of the two original samples, an outcome that is not possible in dry convection, and leads to downdrafts within the cloud. In fact, the net latent heating integrated over the volume and life cycle of a non-precipitating cumulus cloud must be zero. The net convective heating, from (2.83) combined with (2.85) and assuming that $\sigma << 1,$ is given by

$$c_p M \frac{T}{\theta}\frac{d\theta}{dz},$$

and since the stratification $d\theta / dz$ is always non-negative, M must have both positive and negative values. In reality, the net convective mass flux is downward near the tops of non-precipitating and weakly precipitating cumulus clouds, and positive elsewhere.

Thus, entrainment is an essential aspect of moist convection and greatly enriches the problem of understanding how convection operates. The irreversible mixing across cloud boundaries, coupled with the fallout of precipitation, is the dominant source of irreversible entropy production in the tropical atmosphere and perhaps for the whole planet (Pauluis and Held, 2002).

Early attempts to model the entrainment process were based on similarity theory of Schmidt (1941) as extended by Morton et al. (1956), which was developed to describe convection originating from sustained or instantaneous point sources of buoyancy and found to provide excellent descriptions of laboratory plumes and real convection from limited sources, such as volcanic plumes. Simple dimensional reasoning yields an expression for the upward increase in plume mass flux:

$$\frac{dM}{dz} = \alpha \frac{M}{z}, \tag{3.1}$$

where α was estimated to be $\frac{5}{3}$ from laboratory experiments. Similarly, the plumes were both predicted and observed to be conical in shape, with $r = \frac{6}{5}\alpha z.$ The entrained air dilutes the buoyancy of the plume and also directly slows down the updraft as the entrained air must be accelerated to the plume velocity.

Unfortunately, cumulus clouds are not observed to be very similar to entraining plumes, as they do not originate from highly localized sources and, unlike classical plumes, can experience buoyancy reversal from entrainment. They are not observed to have conical shapes and the air within them is far more inhomogeneous than classical plumes (Paluch, 1979). Nevertheless, moist convection continues to be modeled as modified convective plumes, using a variant of (3.1):

$$\frac{dM}{dz} = \lambda M, \tag{3.2}$$

where λ is called the *entrainment parameter* and is usually assumed to be constant, with an estimated value of around $0.15\ km^{-1}$ (Singh and O'Gorman, 2015).

Shallow convective clouds moisten their environments, leaving behind anomalously moist air that is favorable for the development of subsequent clouds, which will be less dilute than their predecessors. It is thus possible for a sequence of plumes to reach successively greater altitudes. The general picture of a turbulent plume ascending through a pristine, unperturbed environment may be too simple.

The dynamics of moist convective clouds are inextricably woven with the cloud microphysical processes that determine the spectrum of cloud drop sizes, the conversion of cloud water to precipitation, and the fall and partial or total re-evaporation of precipitation. Such processes are highly complex and fascinating in their own right and are the subject of an entire sub-discipline of atmospheric science; no short summary could do justice to the subject. Broadly, the initial spectrum of cloud drop sizes is largely determined by the size distribution of the aerosol particles that serve as cloud condensation nuclei. Virtually all the condensation that takes place is heterogeneous – occurring on the surfaces of liquid or solid aerosol particles rather than by spontaneous clumping of water molecules. There are usually enough cloud condensation nuclei that no particles grow by condensation alone to sizes sufficient to have terminal velocities comparable to those of the air motion. Thus other processes must be involved to convert small cloud water droplets or ice crystals to precipitation-size particles.

Two quite different cloud microphysical processes produce precipitation. When ice crystals coexist with (supercooled) liquid water droplets, a rapid flux of vapor from the liquid droplets to the ice crystals occurs, since the saturation vapor pressure over ice is less than that over liquid water at the same temperature. This is known as the Wegener-Bergeron-Findeisen process, named after the scientists[1] who proposed that it could produce precipitation. Rain can also form by stochastic coalescence of liquid water droplets owing to their differential fall speeds and, to some extent, as a result of small-scale turbulent motions of the air in which they are embedded. Greater rates of collision occur with broader drop size distributions, which in turn occur with broader cloud condensation nuclei size distributions.

Observed distributions of liquid water drop sizes in precipitating cumulus clouds are strongly bimodal, with largely separate populations of cloud droplets, whose terminal velocities are small compared to air motions, and raindrops that are falling at appreciable speeds.

Maritime air masses tend to have broad aerosol size distributions, owing in part to the presence of large salt particles resulting from evaporation of spray droplets, and so cumuli developing

[1] Alfred Wegener, the German meteorologist credited with founding the theory of continental drift, was also the first to propose the thermodynamic hypothesis for precipitation formation, in 1911, and eleven years later, the Norwegian scientist Tor Bergeron discovered the process at work in the deposition of ice on cold trees immersed in fog whose temperature was above freezing. The hypothesis was further refined and extended in the 1930s by the German meteorologist Walter Findeisen.

over tropical oceans tend to precipitate rather easily compared to their terrestrial counterparts, as can be seen in Figure 3.2.

Thus, cloud microphysical processes have a strong bearing on the amount of latent heat release in clouds, the formation and concentration of cloud ice, and by modifying cloud drop size distributions, on the optical properties of clouds. Those who assume that the details of such processes are of secondary concern for large-scale tropical phenomena – including tropical cyclones – do so at their peril. We have already seen that the efficiency with which precipitation forms, falls out, and re-evaporates largely determines the water vapor content of RCE states, and this is just the tip of the iceberg.

Very shallow tropical cumuli do not live long enough to form precipitation size-particles, and the re-evaporation of cloud water in turn limits the vertical development of the clouds. Under conditions of strong large-scale subsidence over the ocean, shallow cumuli and/or stratocumuli – which in contrast to cumuli cover large fractions of the sky – dominate the skyscape.

But, especially over the oceans, cumuli do not have to become very deep before they precipitate. Figure 3.2 shows an example of a shallow but precipitating cumulus cloud near the Caribbean island of Barbados. A cloud of similar dimensions over a continental interior would probably not produce rain or snow.

Figure 3.2: Precipitating cumulus mediocris over the tropical Atlantic near Barbados. The cloud is a few kilometers in horizontal and vertical size.

Although precipitation from shallow cumuli over tropical oceans is usually quite light, evaporation is enough to cause downdrafts and cold pools in the subcloud layer and these make beautiful patterns in space-based images, an example of which is shown in Figure 3.3. The black, nearly cloud-free patches represent cold pools produced by evaporation of rain from shallow cumuli. These spread out in the subcloud layer and often produces arcs of cumuli at their edges. The clouds that produced the cold pools in most places have vanished before the time of the image, reminding us that while the pattern of clouds pay persist for many days, each individual cloud has a lifetime of a few tens of minutes.

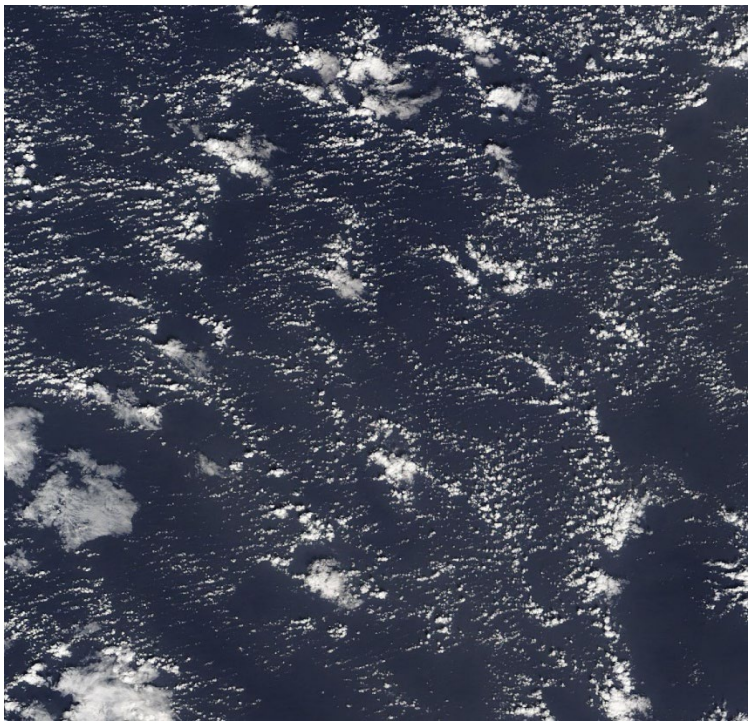


Figure 3.3: Patterns in shallow cumuli over the tropical Atlantic as observed by NASA's MODIS/Aqua on December 19th, 2013.

Deep cumulonimbi can span the whole depth of the troposphere and occasionally overshoot into the lower stratosphere, where they are important agents for injecting water and other tropospheric tracers into the stratosphere. Figure 3.4 shows an example of a tropical cumulonimbus as seen from the International Space Station.

Figure 3.4: Cumulonimbus over western equatorial Africa as captured from the International Space Station on February 5, 2008.

Deep convection is often organized into clusters and squall lines, or arcs, as can be seen in figure 3.5. Evaporation of falling precipitation drives strong downdrafts, which spread out as density currents in the boundary layer. In the presence of low-level shear, deep convection redevelops on the downshear side of the spreading cold pools by a mechanism elucidated by Thorpe et al. (1982) and Rotunno et al. (1988). Potentially buoyant air approaches the squall line from the downshear side, carrying horizontal vorticity associates with the ambient low-level wind shear. The vorticity vector is approximately parallel to the edge of the cold pool. When the air encounters the edge of the cold pool, the baroclinic generation of vorticity tends to decrease this vorticity, so that the air forced up over the cold pool emerges from the boundary layer with little vorticity, favoring an upright updraft. A schematic cross-section through a squall line is shown in Figure 3.6.

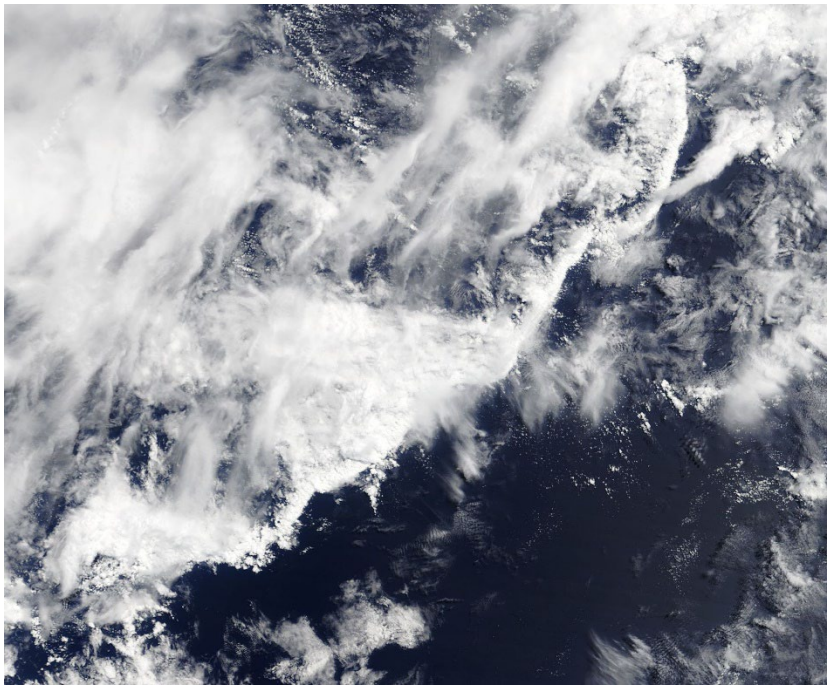


Figure 3.5: A tropical squall line over the western tropical North Pacific on August 14th, 2018. The entire system spans about 400 km from southwest to northeast.

Low level air approaches from the right, or downshear side of the system: Its vorticity is directed into the page. As it encounters the leading edge of the cold pool, it suffers a decrease in vorticity and is forced upward over the cold pool, then rising buoyantly through deep cumulonimbi. Individual deep convective cells form this way and are swept back (to the left) by the squall line-relative flow, decaying as they go. Cloud particles are also swept leftward near the top of the squall line; these settle downward, melting as they cross the freezing level. The cooling produced by melting and evaporation drive a mesoscale downdraft and rear inflow. The cells that make up the squall line have typical lifetimes of 30-45 minutes while the squall line itself may last many hours.

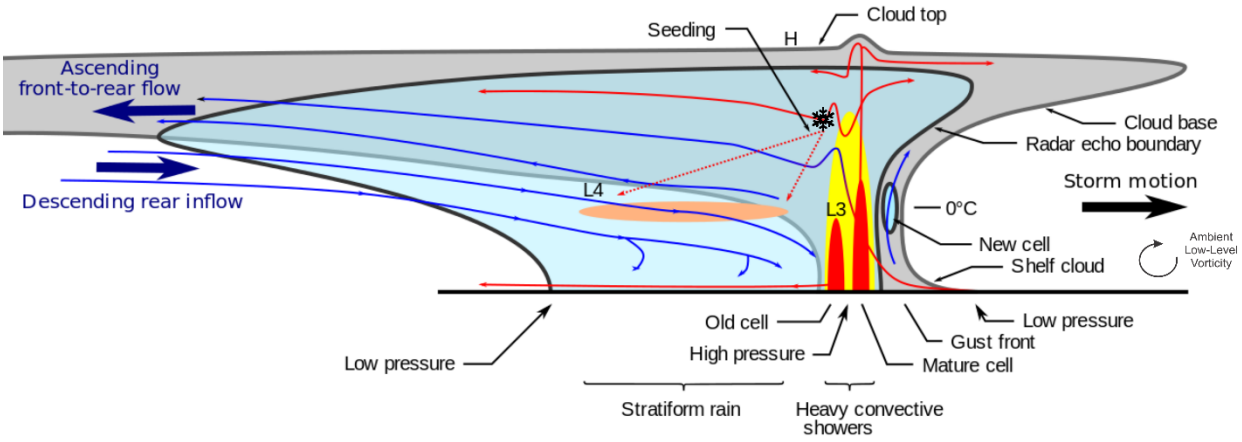


Figure 3.6: Schematic cross-section through a squall line. The cross-section spans about 100 km in the horizontal and 15 km in the vertical.

Squall lines represent a mode of organization brought about by the interaction of deep convection with environmental shear flow. Deep convection can also self-organize by modulating radiative and surface fluxes, forming clusters of deep convective clouds. If they occur sufficiently far from the equator they can begin to rotate, developing into tropical cyclones. We will have much more to say about this, but for now Figure 3.7 shows an example of such a cluster. The cluster contains around 10 individual cumulonimbi at this time, and once can see shallow cumuli forming at the leading edge of a gust front on the near side. The atmosphere around the cluster is unusually free of convection.

Figure 3.7: Cluster of thunderstorms over the floodplain of the Brahmaputra River near the India-Bangladesh border, July 31st 1985, photographed from the space shuttle Challenger.

Overall, the depth of cumulus clouds in the tropics has a tri-modal distribution, as illustrated in Figure 3.8. There are many shallow cumuli, some congesti (reaching to the middle troposphere around the freezing level), and a few deep cumulonimbi spanning the depth of the troposphere. The spectrum is, of course, continuous but does seem to exhibit these three peaks; one can see a suggestion of this in Figure 3.1. Here there are many shallow cumuli in the foreground, interspersed with some cumulus congestus clouds. Deep cumulonimbi dominate the background. Although the deep cumulonimbi are few and far in between, they are responsible for most of the rainfall in the warmer parts of the tropics.

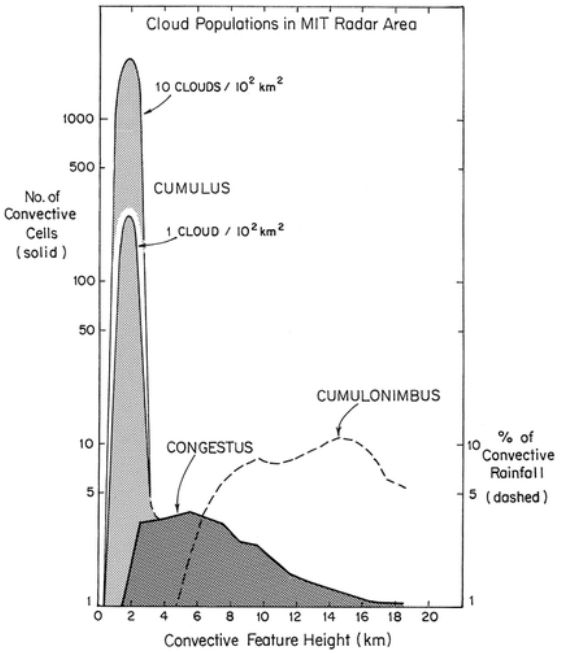


Figure 3.8: Average number of clouds over a region of the tropical South Pacific (solid curves) and percent of convective rainfall (dashed curve) as a function of convective feature height.

Convective rainfall has distinct diurnal cycles. Over the sea, rainfall tends to peak in the hours just before sunrise, while over land the peak occurs in the late afternoon or early evening. Surface fluxes over the sea do not usually show strong diurnal variations owing to the large heat capacity of the ocean's mixed layer, but during the day, direct absorption of sunlight in the atmosphere, mostly by water vapor, leads to a reduction in the net radiative cooling rate and therefore of the net rate of destabilization to convection. Thus convective forcing is maximum during the night. Over land, on the other hand, the very low heat capacity and thermal conductivity of most soils yields surface fluxes that are nearly in balance with the net (solar and infrared) radiative forcing of the surface, which peaks around noon, local time. Thus the rate of destabilization peaks in the middle of the day, and this peak is far stronger than the oceanic nocturnal maximum of destabilization.

As we will see in the next subsection, the time scale over which convection removes instability is not very short compared to a day, so the response of the convection lags the forcing, explaining the ocean convective peak around sunrise and the land peak near sunset. The much stronger diurnal forcing over the land coupled with the lag of the convective response allows somewhat more instability to build up, yielding stronger convection. This is evident in a global map of lightning detected from space presented in Figure 3.9. Lightning is almost exclusively a

terrestrial phenomenon, with very little activity over the sea. Strong peaks are evident over equatorial Africa and the mountainous regions of northeastern Columbia and far western Venezuela.

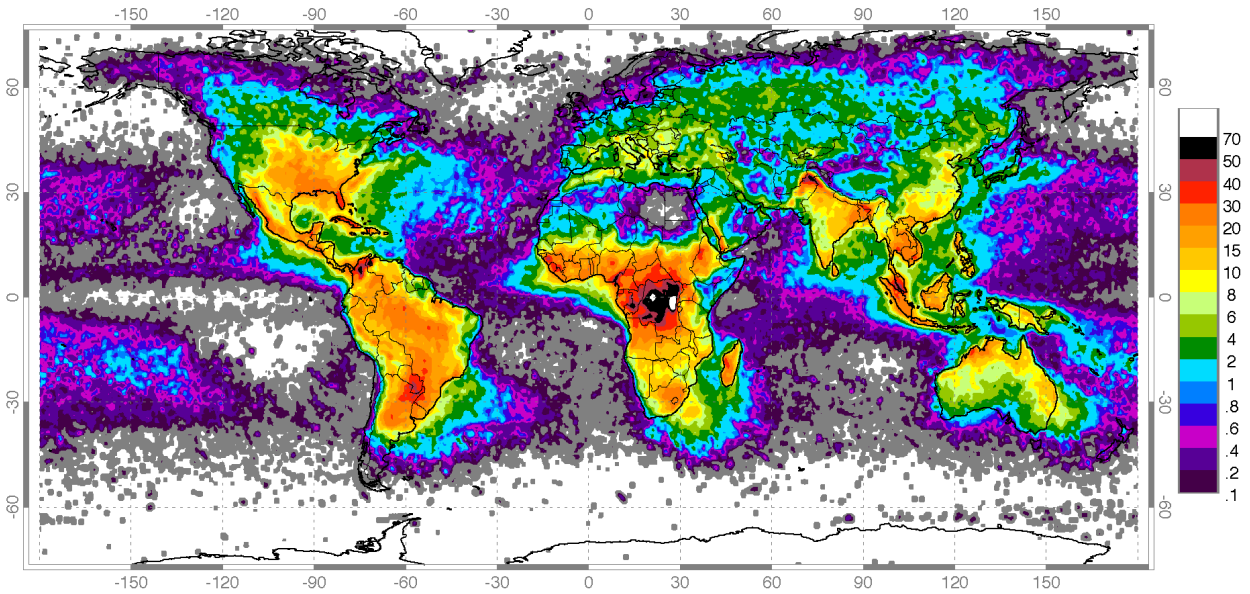


Figure 3.9: Global lightning flash rate, in flashes km^{-2} yr^{-1}. Data obtained from April 1995 to February 2003 from NASA's Optical Transient Detector and from January 1998 to February 2003 from NASA's Lightning Imaging Sensor.

3.2 Moist convective stability and the tropical tropospheric lapse rate

Entrainment and the irreversible fallout and partial re-evaporation of precipitation complicates the assessment of stability of an atmosphere to moist convection. The presence of saturated and unsaturated air at the same pressure level means that the buoyancy of saturated air, with respect to the surrounding unsaturated air, cannot be expressed as a function of pressure and just one other variable; a third variable is required. For this reason, the stability of an unsaturated atmosphere to moist convection cannot be expressed in terms of the vertical gradient of a single variable, a fact that turns out to have profound consequences for the atmosphere.

Here it proves easier to work with a different entropy variable called the *saturation entropy*, denoted by the symbol $s*$. It is the entropy air would have were it saturated at the same temperature and pressure. From (2.69) its definition is then

$$s* \equiv \left(c_{pd} + r_t c_l\right)\ln(\frac{T}{T_0}) - R_d \ln(\frac{p_d}{p_0}) + \frac{L_v r*}{T} + c_l r_t, \tag{3.3}$$

where $r*$ is the saturation mixing ratio. It is important to note that $s*$ is not conserved, except when the air is saturated, in which case it is identical to the ordinary entropy, s. But $s*$ is a state variable, which is important to what follows. The last term in (3.3) is unnecessary, but makes further derivations a little more compact. As entropy is arbitrary to within and additive

constant, which can be a conserved variable, (3.3) is still conserved in reversible processes in saturated air.

Using the chain rule, we can express small fluctuations of specific volume at constant pressure as

$$\alpha' = \left(\frac{\partial \alpha}{\partial s^*}\right)_{p,r_t} s^{*\prime} + \left(\frac{\partial \alpha}{\partial r_t}\right)_{p,s^*} r_t', \tag{3.4}$$

where we have chosen total water mixing ratio, r_t, to be the third variable. Using (2.69) and (2.71) we can write

$$dk = Tds^* - c_l T \ln\left(\frac{T}{T_0}\right) dr_t + \alpha\left(1 + r_T\right) dp, \tag{3.5}$$

where, in the last term, we have assumed that the total water content is equal to the vapor content in unsaturated air. By using (2.73), and a similar identity:

$$\left(\frac{\partial}{\partial r_t}\right)_{p,s^*}\left(\frac{\partial k}{\partial p}\right)_{s^*,r_t} = \left(\frac{\partial}{\partial p}\right)_{s^*,r_t}\left(\frac{\partial k}{\partial r_t}\right)_{p,s^*}, \tag{3.6}$$

and using (3.5), we have

$$\left(\frac{\partial \alpha}{\partial s^*}\right)_{p,r_t} = \frac{1}{1+r_t}\left(\frac{\partial T}{\partial p}\right)_{s^*,r_t},$$

which is the same as (2.74) but in terms of saturation entropy, and

$$\left(\frac{\partial \alpha}{\partial r_t}\right)_{p,s^*} = \frac{1}{1+r_t}\left[-\alpha - c_l\left(\frac{\partial T}{\partial p}\right)_{s^*,r_t}\left(1 + \ln\left(\frac{T}{T_0}\right)\right)\right]. \tag{3.7}$$

Using these and the hydrostatic equation in (3.4) we can write the perturbation buoyancy as

$$\begin{aligned} B = \frac{\alpha'}{\alpha} &= \frac{1}{\alpha}\left(\frac{\partial \alpha}{\partial s^*}\right)_{p,r_t} s^{*\prime} + \frac{1}{\alpha}\left(\frac{\partial \alpha}{\partial r_t}\right)_{p,s^*} r_t' \\ &= \frac{1}{1+r_t}\left[\Gamma_m s^{*\prime} - g r_t' - \left(\Gamma_m c_l\left(\ln\left(\frac{T}{T_0}\right) + 1\right)\right) r_t'\right]. \end{aligned} \tag{3.8}$$

At first glance, it would appear that buoyancy depends on the arbitrary constant T_0, but remember that T_0 also appears in the definition of s^* and the two terms cancel in (3.8).

If the relation (3.4) did not do the trick, (3.8) makes it clear that to calculate buoyancy one needs to know not just the difference between the saturation entropies of the parcel to be lifted and its environment, but the difference between the water contents of the parcel and environment as

well. *Knowing the vertical profile of* s *or* s^* *does not suffice to determine the stability of an atmosphere to a moist process*.

Even if one neglects the condensed water in clouds, the contribution of water vapor to density is significant enough to make the contribution of r_t' important in (3.8). Thus it is not possible to make quantitatively accurate estimates of the buoyancy of parcels lifted by some adiabatic process using two-dimensional diagrams such as skew-T-log-p charts (e.g. Figure 2.18).

Fortunately, the easy numerical computation of thermodynamic processes renders thermodynamic diagrams unnecessary, though they are still useful for plotting the results of such computations.

As an illustration, let's travel to the beautiful island of Majuro, part of the Marshall Islands of the tropical North Pacific. Rawinsondes (weather balloons) are launched twice daily from this small coral atoll, as they are from many stations around the world[2]. Instruments suspended from the ascending balloons measure pressure, temperature, relative humidity, and GPS location, from which the balloon drift and thus the horizontal winds can be inferred.

Figure 3.10 is a variant on the classical skew-T-log-p chart in which isopleths of constant density temperature, T_ρ, are plotted instead of isotherms.

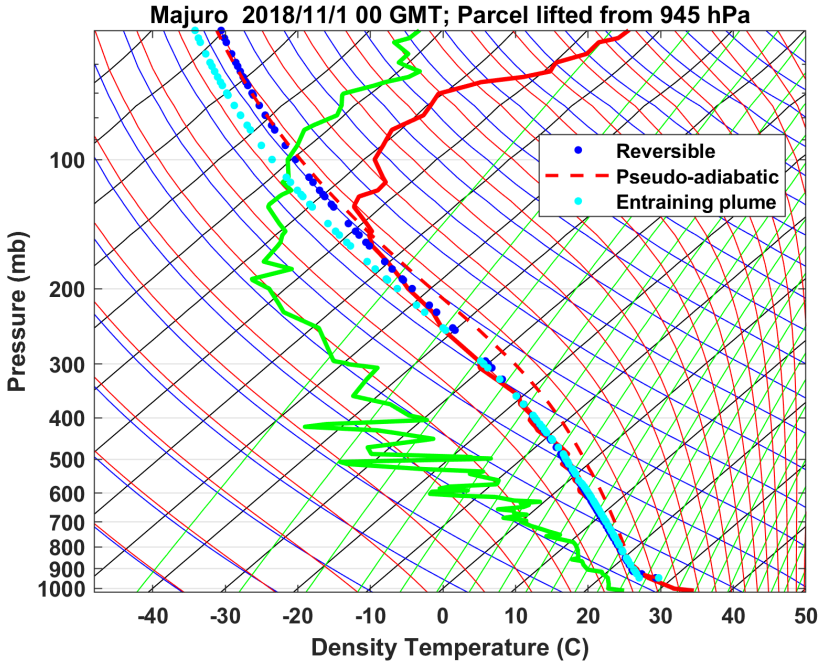


Figure 3.10: Skew-T_ρ – log-p diagram similar to the thermodynamic diagram in Figure 2.18, except that the density temperature replaces the actual temperature. The thin solid red curves are virtual pseudo-adiabats, showing the virtual temperature of air displaced pseudo-adiabatically. The thick, solid red and green curves show the virtual temperature and dew point temperature recorded by the rawinsonde launched from Majuro at 00 GMT, November 1st 2018. See text for a description of the other curves.

[2] For a map of active rawinsonde stations, see the web site maintained by the University of Wyoming: http://weather.uwyo.edu/upperair/sounding.html

The *density temperature* is defined as the actual temperature a perfectly dry sample of air would have if it had exactly the same density as the actual, moist sample (which may also contain suspended condensed water). It is given by

$$T_\rho \equiv T\left[\frac{1+r/\epsilon}{1+r_t}\right], \tag{3.9}$$

where r is the mixing ratio, r_t is the total water content, and ϵ is the ratio of the molecular weight of water to a suitably defined weighted mean molecular weight of the other constituents of the atmosphere. (It has a value of about 0.622 in our atmosphere.) If there is no condensed water in a sample, its density temperature is the same as its *virtual temperature*, T_v:

$$T_v \equiv T\left[\frac{1+r/\epsilon}{1+r}\right]. \tag{3.10}$$

The thick, solid red and green curves in Figure 3.10 show the virtual temperature and dew point temperature recorded by the rawinsonde launched from Majuro at 00 GMT, November 1st 2018. The thick red dashed curve shows the density temperature of a sample of air lifted pseudo-adiabatically from 945 hPa (one of the sample points in the sounding) and the dark blue dots show the density temperature of a sample lifted reversibly (i.e. not allowing any condensed water to fall out) and without freezing, from the same starting pressure. The light blue dots in the figure show the virtual temperature of a sample lifted from the same level, but allowing for both entrainment of undisturbed environmental air, and latent heat of fusion when water vapor condenses directly into ice crystals (which are then removed, after entrainment, according to the pseudo-adiabatic process.) The entrainment constant used here (see equation 2.92) is 0.18, about an order-of-magnitude smaller than deduced from laboratory experiments (Morton et al., 1956).

Once can see in Figure 3.10 that air lifted from about 945 hPa (roughly 500 m over the sea surface) through the depth of the troposphere by a pseudo-adiabatic process has appreciable buoyancy up to about 180 hPa, but the same sample lifted by either a reversible process (in which the condensed water contributes to the effective density of the lifted air) or as an entraining plume with freezing has little buoyancy. This is made more apparent in Figure 3.11, which simply graphs the difference between the lifted sample and environmental density temperatures of this sounding, for the three aforementioned processes.

We mention here two other ways of characterizing the stability of soundings. The first begins by defining a new quantity, the *buoyancy entropy*, or s_B. This is simply defined as the reversible entropy that a hypothetical saturated sample with no condensed water would have to have at some reference pressure such that, when lifted reversibly (with no freezing), it would have the same density temperature as the air at a given pressure. That is, at each level in the sounding, we find that value of reversible entropy at a reference pressure, p_0, that, when lifted to the given level, has the same density temperature as the virtual temperature at that level. We then define the *buoyancy potential temperature*, θ_B, as

$$\theta_B \equiv 300\,\mathrm{e}^{s_B/c_p}, \tag{3.11}$$

where c_p is the heat capacity of dry air at constant pressure, and the factor 300 is an arbitrary constant. It is important to note that with these quantities, *the reference pressure matters*! It is not arbitrary. It should be chosen as a level typical of where air ascending in cumulus clouds first becomes saturated when lifted from the subcloud layer. In what follows, we choose it to be the lifted condensation level of air at the lowest level in the soundings.

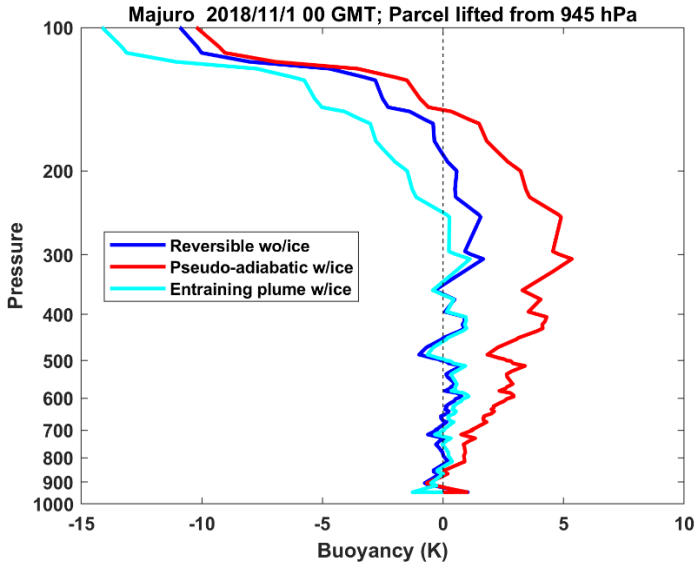


Figure 3.11: Difference between the density temperatures of a sample lifted from 945 hPa by three processes, and the environmental density temperature for the same sounding displayed in Figure 3.10.

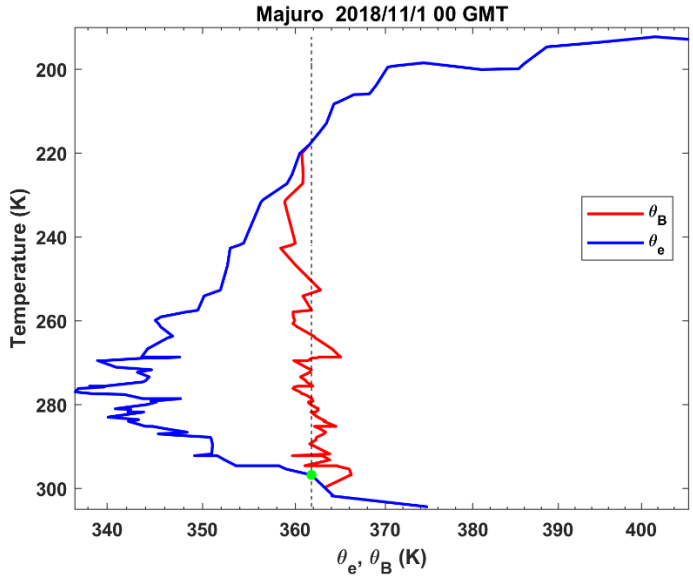


Figure 3.12: The buoyancy potential temperature, θ_B (red), of the sounding shown in Figure 3.10, compared to the equivalent potential temperature (θ_e, blue). The green dot indicates the 945 hPa level. Note that the ordinate here is absolute temperature and the abscissa is in equal increments of the logarithm of potential temperature.

Note that θ_B is conserved under reversible adiabatic displacements.

Figure 3.12 shows, in red, the buoyancy potential temperature graphed against absolute temperature and compared to the equivalent potential temperature, θ_e, defined as

$$\theta_e \equiv 300 e^{s/c_p}, \tag{3.12}$$

where the entropy, s, is given by (2.69). This is for the same sounding as that used in Figures 3.10-3.11. A sample lifted from about 945 hPa in this case has roughly the same θ_B as its environment, rendering it nearly neutrally buoyant, consistent with the blue curve in Figure 3.11. We have plotted θ_B against absolute temperature rather than pressure, and the abscissa in Figure 3.12 is in equal increments of $\ln(\theta_B)$. The reason for doing this will become apparent shortly.

A more comprehensive evaluation of stability can be accomplished using a buoyancy diagram of the type shown in Figure 3.13. Here we simply calculate the difference between density temperature of samples lifted from all levels of the sounding to all levels, and contour the difference. The ordinate is the logarithm of the pressure to which the sample is lifted, and the abscissa is the pressure from which the parcel is lifted. Figure 3.13 is for reversible adiabatic displacements (dry adiabatc to the lifted condensation level, and moist adiabatic thereafter) but the concept can be applied to any parcel lifting process.

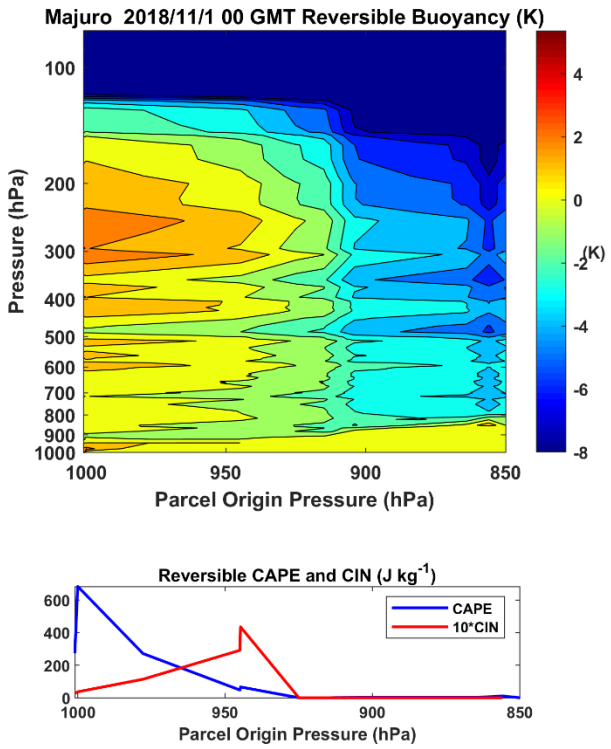


Figure 3.13: Top: Buoyancy matrix, showing the difference between the density temperatures of parcels lifted by a reversible, ice-free moist adiabatic process and that of their environment, using the same sounding as in Figures 3.10-3.12. The abscissa shows the origin pressures of the lifted parcels and the ordinate the pressures to which the

parcels are lifted. Bottom: The CAPE and CIN of parcels lifted to their levels of neutral buoyancy (see text for explanation).

The stability of vertical profiles to moist processes can be compactly summarized by the two vertically integrated quantities: The *Convective Available Potential Energy* ($CAPE_i$), and the *Convective Inhibition* (CIN_i). $CAPE_i$ is the work per unit mass done by a parcel lifted from its origin level *i* to *its level of neutral buoyancy*, LNB_i. The latter is the highest level in the sounding at which the lifted parcel's buoyancy vanishes. If we follow the parcel from its origin level i up to LNB_i and then back down in the adjacent undisturbed environment, the work done is given by

$$CAPE_i = -\oint \alpha dp = \int_{LNB_i}^{i} R_d \left(T_{\rho p} - T_{\rho e}\right) d\ln(p), \tag{3.13}$$

where $T_{\rho p}$ is the density temperature of the lifted parcel and $T_{\rho e}$ that of its environment. The $CAPE_i$ of the Majuro sounding is shown as a function of the parcel origin level i in the bottom panel of Figure 3.13 (blue curve). The blue curve in the bottom panel is simply the vertical integral of the quantity contoured in the top panel of Figure 3.13. The $CAPE_i$ of the parcel originating at 945 hPa is also proportional to the area between the colored curves and the vertical dashed black line in Figure 3.11.

The convective inhibition of a parcel i, CIN_i, is defined as the work needed to lift a parcel from its origin pressure to its *Level of Free Convection*, LFC_i., which is usually defined as the lowest level at which a lifted parcel achieves positive buoyancy. The convective inhibition is given by

$$CIN_i = -\int_{LFC_i}^{i} R_d \left(T_{\rho p} - T_{\rho e}\right) d\ln(p). \tag{3.14}$$

Note the negative sign. By definition, the integrand of (3.14) is negative, so CIN_i is defined as a positive quantity. But here we define CIN_i a bit differently as the integral from the origin level up the level of neutral buoyancy, LNB_i, but only including nonpositive values of $T_{\rho p} - T_{\rho e}$. This is because, in the tropics, there may be multiple levels at which the sign of the buoyancy reverses, causing some ambiguity in defining the convective inhibition. But when a lifted parcel has appreciable positive buoyancy through most of its ascent, the two definitions usually yield the same result. The CIN_i of the Majuro sounding is shown by the red curve in the bottom panel of Figure 3.13; here it has been multiplied by 10 to give it more visibility.

It is also possible to describe stability in terms of the difference between suitably defined entropy of a lifted parcel and that of its environment. We can relate the work done on a parcel, given by (3.13), to entropy and other quantities using (2.72):

$$\begin{aligned} CAPE_i = -\oint \alpha dp &= \oint \left[Tds - c_l T \ln(\frac{T}{T_0}) dr_t + \alpha r_T dp - dk \right] \\ &\cong \oint \left[Tds - c_l T \ln(\frac{T}{T_0}) dr_T + gz dr_T \right]. \end{aligned} \tag{3.15}$$

Variations in enthalpy, k, drop out of the integral as they constitute perfect differentials, and in the small term involving pressure work on water, we have used the hydrostatic approximation:

$$\alpha r_T dp \simeq -g r_T dz = -d(g r_T z) + gz dr_T .$$

Remember that T_0 also appears in our definition of entropy, (2.69), which cancels with the middle term in (3.15). For a particular thermodynamic process, we can always define a value of T_0 that minimizes the middle term in (3.15) and we will therefore not concern ourselves with that term here except to choose a suitable value in calculating the entropy such that this term is small.

The last term in (3.15) represents the work needed to lift water.

In defining the buoyancy potential temperature, θ_B, we took into account the density temperatures of the lifted parcel and its environment. Thus, to a good approximation,

$$CAPE_i \simeq \int_{LNB_i}^{i} c_p \left(\ln \theta_{Bp} - \ln \theta_{Be} \right) dT, \tag{3.16}$$

which is proportional to the area enclosed by the red curve and the vertical dashed line in Figure 3.12, for a parcel lifted from 945 hPa. This is why we plotted the sounding in $\ln \theta_B - T$ coordinates.

Diagrams of the type displayed in Figures 3.10-3.13 can easily be made for any rawinsonde station over the last few decades using software available at https://zenodo.org/doi/10.5281/zenodo.10795293 .

The parcel origin level and the nature of the lifting process are both terribly important for defining $CAPE_i$ and CIN_i in the tropics, where buoyancies usually have small magnitude. Many research papers and textbooks are sloppy about this, referring to CAPE and CIN as though they are unique properties of a sounding, without stating the parcel origin level and without defining the lifting process. (In these cases, CAPE and CIN usually refer to pseudo-adiabatic lifting from the lowest level in the sounding.) None of the lifting processes described in this chapter perfectly capture the likely buoyancy of convective clouds, and it must always be remembered that the air within clouds, and indeed the air entering the bases of the clouds, are fully turbulent and have strongly inhomogeneous thermodynamic properties.

Figures 3.10 – 3.13 show that this particular sounding is unstable to air lifted from near the surface by any of the three processes considered. As particularly evident in Figure 3.13, the parcel buoyancy decreases as its origin height increases, with an abrupt transition to low buoyancy near 940 hPa – about 600 m above the ocean surface. As demonstrated by Xu and Emanuel (1989), the tropical atmosphere over the warmest parts of the tropics is nearly neutral to reversible ascent from just below the top of the subcloud layer (see Chapter 4 for a discussion of tropical boundary layers), to within the instrumental accuracy of rawinsonde measurements. Why this should be so remains something of a mystery. Perhaps, as hinted in Figures 3.10 and 3.11, this is because a reversible adiabat happens to lie close to the temperature of an entraining plume in which the latent heat of fusion is accounted for. Yet the latent heat of fusion, when included, must affect not only the parcel temperature but, ultimately the temperature of the environment in moist radiative-convective equilibrium (see the next section of this chapter). Since no precipitation reaches the surface in the ice phase in the tropics (except on high mountains) there can be no net release of latent heat of fusion in the column. Moreover, since freezing generally occurs at a lower temperature than melting, thanks to supercooling of cloud water, the latent heat of fusion subtracts from the work done in any closed thermodynamic cycle, so in principle latent heat of fusion should make convection weaker. This

contradicts conclusions based on considering the latent heat of fusion's effects only on lifted parcel temperature and not, ultimately, on the environment.

The condensed water content of clouds is almost everywhere much lower than that obtained by adiabatic ascent from cloud base. Yet clouds are highly inhomogeneous, and in measurements taken by slow-flying gliders, a few samples do seem to have nearly adiabatic water content (Paluch, 1979; Figure 4). It is also possible that samples that escape mixing also have a hard time forming precipitation, so that the majority of samples having sub-adiabatic water content have also been diluted by mixing, which usually decreases their buoyancy.

Whatever the reason, it remains an empirical fact that the tropical troposphere is nearly neutral to reversible ascent from near the top of the subcloud layer; that is, it has nearly constant θ_B. We will have many occasions to make use of this property through the rest of this book.

But there are several circumstances that lead to appreciable values of $CAPE_i$, defined in any reasonable way. The strong diurnal variation of surfaces fluxes on land destabilizes the atmosphere faster than convection can develop and remove the instability, so that CAPE accumulates during the day, typically reaching peak values in the late afternoon. Figure 3.14 shows a buoyancy matrix and $CAPE_i$ diagram for a sounding at Tampa, Florida at 00 GMT on September 1st, 2017. This is about an hour before sunset in Tampa at this time of year.

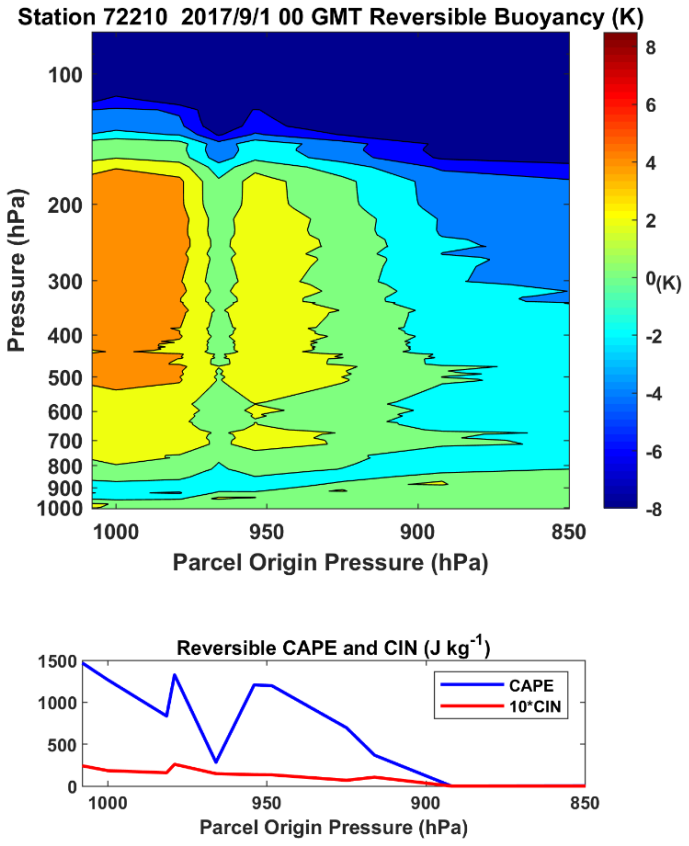


Figure 3.14: Buoyancy matrix and CAPE and CIN diagram as in in Figure 3.13, but for Tampa, Florida (USA) at 00 GMT on September 1st, 2017.

Inspection of the sounding plotted on a skew-T-log-p diagram as in Figure 3.10 shows that the temperature lapse rate in the free troposphere nearly follows the density temperature of a reversible adiabat, as in most maritime tropical soundings. But the entropy of the boundary layer is high so that large values of CAPE develop for air through the whole depth of the subcloud layer (bottom panel of Figure 3.14). Averaged over this layer, CAPE values are almost an order-of-magnitude higher than is typical for maritime soundings such as the Majuro profile in Figures 3.10-3.13. Were all of the potential energy per unit mass, quantified by CAPE, converted to kinetic energy, $50\,ms^{-1}$ updrafts would result.

Although Tampa is near the west coast of Florida, air flow during the summer is generally from the east, so that afternoon and evening thunderstorms develop over the peninsula and drift westward toward Tampa. The region around Tampa has the highest annual lightning discharge rate in the U.S. The accumulation of CAPE during daylight hours over land is almost certainly the reason for the high lightning rates seen in Figure 3.9. The charge separation that leads to lightning depends on significant lofting of large ice particles, which can only occur in sufficiently strong updrafts.

Even larger values of CAPE can build up when deep, nearly dry-adiabatic layers spawned over dry soils such as found in deserts are advected over moist air and/or moist soils. This is common over the U.S. plains during spring, when strong solar heating of the dry soils of the desert southwest and northern Mexico heats the still-chilly springtime air from below, yielding deep, dry-convecting layers. These are then advected by the strong westerly springtime flow over the moister soils to the east, and as water evaporates into them, the surface air is chilled and moistened. Daytime growth of a boundary layer under the desert air can result in very large values of CAPE, as illustrated by the Norman, Oklahoma sounding at 00 GMT on May 5th, 2007 (Figures 3.15 and 3.16). This sounding was made about 2 hours before sunset. Reversible CAPE values are as high as $4000\,J\,Kg^{-1}$, much larger than the unstable Tampa sounding, and the unstable air occupies a deep layer. Unlike tropical soundings, CIN values are also large and prevent any release of the instability, permitting large values of CAPE to accumulate with no convection at all. Note also in Figure 3.15 that the lapse rate of temperature in the free troposphere is somewhat steeper than a reversible adiabat, again in contrast to tropical soundings, including the Tampa sounding.

The Norman sounding illustrates a rather special circumstance under which instability builds up, contained by a potential barrier represented by large values of CIN, and may subsequently be released explosively, resulting in severe thunderstorms and attendant phenomena such as tornadoes and hail. These conditions develop when dry soils are found upwind of and close to wet soils, and where there is sufficiently strong airflow from the dry toward the moist regions. They are rare in the tropics, and we will not further concern ourselves with such conditions.

Over tropical oceans, convection – whether deep or shallow – may be considered a *quasi-equilibrium process* (Arakawa and Schubert, 1974) wherein convection releases instability at the rate it is generated by larger-scale processes such as radiative cooling of the atmosphere and surface enthalpy fluxes, much like dry convection over a heated surface or steadily boiling water in a pot. It should not be thought of as a store-release mechanism as in the Norman case illustrated above. Even over land, where diurnal surface heating is too fast for the convection to keep up with the generation of instability, one can describe the physics as a lagged-equilibrium process, where although the convective response lags the diurnally varying forcing, over longer

time scales the convection can be regarded as being in equilibrium. This is the view of convection we will adhere to throughout the remainder of this book.

3.3 Moist Radiative-Convective Equilibrium

3.3.1 General features of moist RCE

We are now in a position to extend our discussion of radiative-convective equilibrium from Chapter 2 to the case in which the convection is moist and precipitating. Moist RCE is considerably more complicated and endlessly more interesting than the dry case. While, as we have just seen, moist convection tends to establish characteristic profiles of (virtual) temperature, as in the case of dry convection, it most certainly does not mix conserved variables though the depth of the convective layer, simply because the turbulent patches of convection occupy a small factional volume of the convecting layer, thanks to the irreversible fall-out of condensed water. Thus the central challenge of understanding moist RCE, and indeed of representing cumulus convection in models in which it cannot be resolved, is getting the vertical distribution of moisture and stratiform clouds right. Since water vapor is the most important greenhouse gas in our atmosphere, and its presence above the subcloud layer in almost entirely owing to lofting of water by convection, *moist RCE is a strongly two-way process*, with radiation driving convection and convection determining the distribution of the most important greenhouse gas. Convection also determines the distribution of associated stratiform or quasi-stratiform clouds such as ice anvils and boundary-layer clouds, which greatly affect the distribution of radiative cooling.

In recent years, it has been possible to simulate moist RCE using numerical models that permit but do not come close to resolving moist convective clouds. Before turning to such models, we aim for some conceptual understanding of moist RCE using very simple models.

We begin with a simple RCE state in which the radiative cooling and ocean temperature are specified rather than calculated; that is, we first tackle the one-way problem. The structure of a very simple analytical model is shown in Figure 3.17. Deep, moist convective updrafts carry an upward mass flux per unit area M_u out of the boundary layer and through the density-weighted depth, H, of the troposphere. Downdrafts driven by evaporation of falling precipitation carry a mass flux M_d, and the air outside convective systems slowly subsides through the troposphere at a velocity w_e. Conservation of mass requires that

$$M_u = M_d + \rho w_e, \tag{3.17}$$

where ρ is a characteristic density of the lower troposphere and all the variables in (3.17) have been defined to be positive or zero. The troposphere is cooling at a constant rate given by Q_{rad}, and there is a surface enthalpy flux (dominated by a latent heat flux) given by F_s. Conservation of energy in the whole system requires that

$$F_s = Q_{rad} H. \tag{3.18}$$

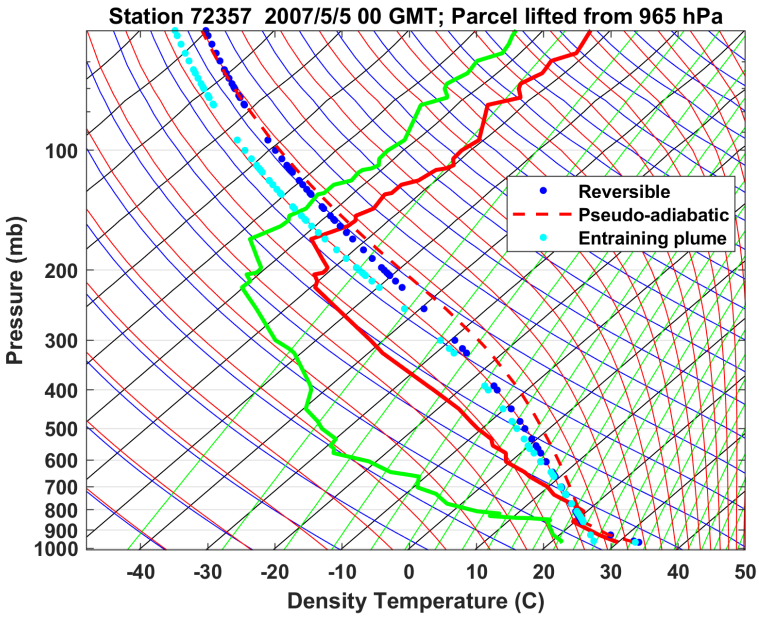


Figure 3.15: As in Figure 3.10, but for Norman, Oklahoma at 00 GMT on May 5th, 2007. The dashed red curve and the light- and dark-blue dots show, respectively, pseudo adiabatic, reversible and entraining plume ascents from 965 hPa (which is near the surface).

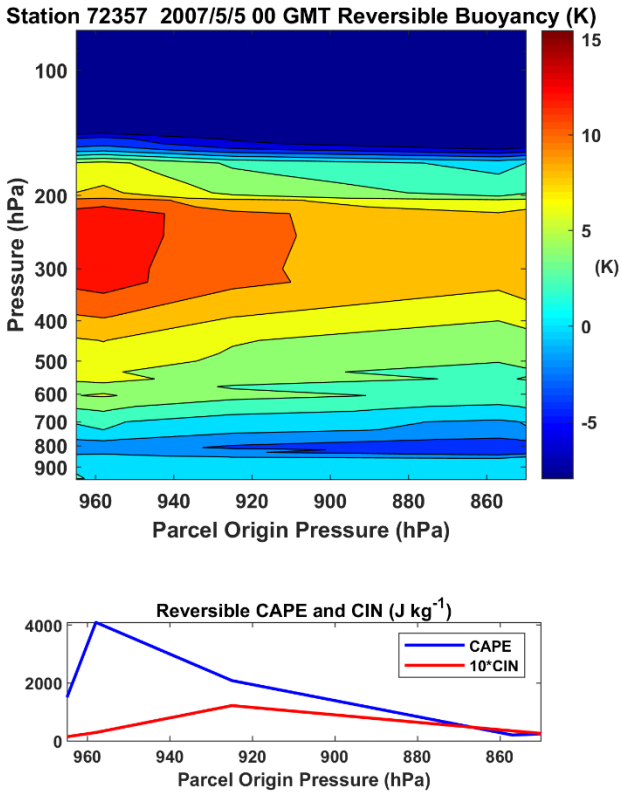


Figure 3.16: As in Figure 3.13, but for Norman, Oklahoma at 00 GMT on May 5th, 2007

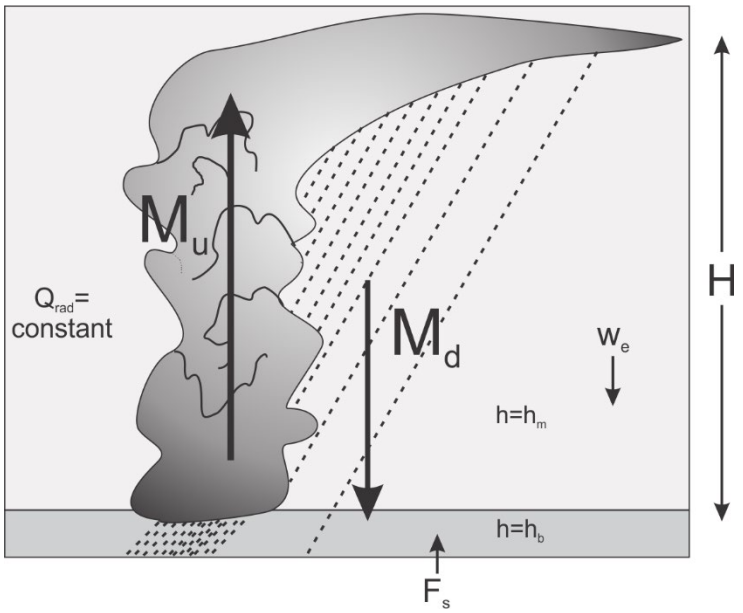


Figure 3.17: Very simple model of moist radiative-convective equilibrium. See text for full description.

We will take the thin subcloud layer to have a moist static energy of h_b and characterize the moist static energy just above the subcloud layer as having the value h_m.

Given the radiative cooling rate Q_{rad} we wish to find the convective mass fluxes and the moist static energies of this very simple system.

We begin with the energy balance of the subcloud layer. The flux of moist static energy from the middle troposphere must balance the flux from the surface:

$$\begin{aligned} F_s &= (M_d + \rho w_e)(h_b - h_m) \\ &= M_u (h_b - h_m), \end{aligned} \tag{3.19}$$

where we have used (3.17) for the second line of (3.19). This can be written alternatively as

$$M_u = \frac{F_s}{h_b - h_m}, \tag{3.20}$$

showing that the convective updraft mass flux increases with increasing surface enthalpy flux and decreasing difference between the moist static energies of the subcloud layer and lower free troposphere.

Next we consider the energy balance in the clear, subsiding air. Here, the radiative cooling must be balanced by the vertical advection of dry static energy, h_d. This may be expressed by

$$\rho w_e \frac{dh_d}{dz} = Q_{rad}. \tag{3.21}$$

Henceforth we will use the symbol S in place of dh_d/dz to denote the dray static stability. As we have seen earlier in this chapter, the virtual temperature lapse rate in the tropics falls nearly along a reversible adiabat, so S is really just a function of the temperature of the system.

Using (3.17) to eliminate w_e gives

$$M_u - M_d = \frac{Q_{rad}}{S}. \tag{3.22}$$

Next we provide a crude representation of the microphysics that drive deep downdrafts by relating the downdraft mass flux to the updraft mass flux via

$$M_d = \left(1 - \epsilon_p\right) M_u, \tag{3.23}$$

where ϵ_p is a bulk precipitation efficiency: the fraction of condensed water that ultimately falls to the ground. If that fraction is zero, there can be no net latent heat release and the downdraft mass flux would equal the updraft mass flux, making it impossible to satisfy (3.22). Also, no precipitation reaching the ground implies zero evaporation from the surface, so this is not a viable limit of the system. At the other extreme, a precipitation efficiency of 1 means that no condensed water re-evaporates and so there can be no deep downdraft.

Substitution of (3.23) into (3.22) gives

$$M_u = \frac{Q_{rad}}{\epsilon_p S}. \tag{3.24}$$

Thus for a given cooling rate the convective updraft mass flux increases with increasing radiative cooling, decreasing precipitation efficiency, and decreasing static stability (i.e. decreasing system temperature). Through (3.23), this is also true of the downdraft mass flux.

Comparing (3.24) to (3.20) implies that

$$h_b - h_m = \frac{\epsilon_p S F_s}{Q_{rad}} = \epsilon_p SH, \tag{3.25}$$

where we have made use of (3.18). Thus the decrease of moist static energy from the subcloud layer to the lower free troposphere increases with static stability (i.e. system temperature) and precipitation efficiency.

Since convecting atmospheres have nearly moist adiabatic lapse rates, the saturation moist static energy, h^* above the boundary layer will be nearly equal to the actual moist static energy of the subcloud layer, so we can write (3.25) as

$$h_b - h_m \simeq h^* - h_m = L_v q^* \left(1 - \mathcal{H}\right) = \epsilon_p SH, \tag{3.26}$$

where q^* is a characteristic value of saturation specific humidity just above the subcloud layer and $\mathcal{H}$ is the relative humidity of air entering the top of the subcloud layer. Since SH is the

difference between the dry static energy along a moist adiabat, going from the lower troposphere to near the tropopause, and since this difference is nearly equal to $L_v q^*$, we have that

$$\mathscr{H} \cong 1 - \epsilon_p, \tag{3.27}$$

consistent with what we found in a more detailed analysis of the RCE water budget in Chapter 2 (see, e.g., 2.91). The more efficient the formation and fallout of precipitation, the drier the free troposphere.

Although the details of the vertical profile of humidity in moist RCE depend on many processes, such as entrainment and detrainment, the basic dependence also on cloud microphysics is unequivocal.

We can complete the specification of the system by representing the surface enthalpy flux by a classical aerodynamic surface exchange formulation:

$$F_s = QH = \rho_s C_k |\mathbf{V}_s| \left(h_0^* - h_b \right), \tag{3.28}$$

where ρ_s is the surface air density, C_k is a nondimensional surface exchange coefficient, $|\mathbf{V}_s|$ is a characteristic surface wind speed, and h_0^* is the saturation moist static energy at sea surface temperature. (For the present purpose, we assume that this is a known property of the system). Turning this around, we have

$$h_b = h_0^* - \frac{QH}{\rho_s C_k |\mathbf{V}|}, \tag{3.29}$$

which yields the subcloud layer moist static energy in RCE as a function of the sea surface temperature, the integrated radiative cooling of the atmosphere, and the surface wind speed. Note that smaller surface winds result in lower moist static energy of the subcloud layer.

Using (3.29) and (3.25) gives an expression for the moist static energy just above the subcloud layer:

$$h_m = h_0^* - \epsilon_p SH - \frac{QH}{\rho_s C_k |\mathbf{V}|}. \tag{3.30}$$

Thus if we are given the radiative cooling of the troposphere, the surface temperature, and the surface wind speed, along with the precipitation efficiency and the system temperature, we can determine the convective mass fluxes and the moist static energies. Note in particular the important role of surface winds, which in a single-column model must be specified. This is because, except in conditions of very light wind, the surface turbulent flux of enthalpy is largely determined by wind rather than by convective instability. We will return to this important point in Chapter 4.

Simple models like this one serve to illuminate some of the basic elements of moist RCE. For example, in Chapter 4 we will develop a variant of the relationship (3.20) between cloud base mass flux and surface enthalpy fluxes as a more general way of representing convection in time varying systems with large-scale circulations. But the really interesting physics, largely absent in the case of dry convection, involve the two-way interaction of convection and water vapor (and

clouds), wherein the convection controls the water profile which in turn largely determines the radiative cooling profile, which drives the convection. This interaction is sufficiently complex, depending strongly on turbulent and microphysical processes in clouds, that moist RCE is not really very well understood even today.

Superficially, there appear to be some nice parallels with dry convection. We observe that the temperature profile of moist convecting atmospheres lies along a reversible moist adiabat, and we could just adjust steeper profiles back to an adiabat, the way we often handle dry convection.

But which value of reversible entropy should we adjust to? In the dry problem, we insist that energy be conserved, so that the dry static energy of the adjusted state equals that of the initial state. This produces a unique solution. In the moist problem, we must also insist that static energy (moist static energy in this case) be conserved in the adjustment, but we know from observations that the profiles of moist static energy in convecting atmospheres are nowhere near constant in altitude. (A typical profile of moist static energy will look very much like the profile of θ_e, an example of which is shown in Figure 3.12.) Thus we have an integral constraint on the adjusted moist static energy profile but that is not sufficient to tell us what adiabat to adjust to. In moist RCE, the convective effects on temperature and moisture are inextricably bound.

One of earliest attempts to deal with moist convection in moist RCE states was that of Manabe and Strickler (1964), who simply held the specific humidity of the atmosphere to be unaffected by convective adjustment. They merely set any lapse rates in excess of a critical rate of $6.5\ K\ Km^{-1}$ back to the critical rate, enforcing that the gain in the mass-weighted vertical integral of the dry static energy equal the vertically integrated radiative cooling. There was no feedback of convection to moisture in this early simulation. Three years later, they allowed for a limited feedback by assuming that the relative rather than absolute humidity remains constant (Manabe and Wetherald, 1967).

When general circulation models were first developed, they often suffered from a catastrophic numerical instability that occurred when the model atmosphere became unstable to moist convection. Mostly to avoid this problem, Manabe et al. (1965) developed what amounts to an internal convective adjustment of explicitly simulated clouds for use in general circulation models (GCMs). Whenever the relative humidity equals or exceeds 100% and the lapse rate is steeper than $6.5\ K\ Km^{-1}$, the lapse rate is adjusted back to the critical value assuming that the mass-weighted vertical integral of moist static energy is conserved and that the air remains saturated at all levels adjusted, while any condensed water is immediately removed. The assumption of 100% relative humidity within the region of adjustment suffices to close the problem.

This is a well-defined and unique adjustment, but makes an extreme assumption about condensed water. Consequently, it produces unrealistic profiles of relative humidity, as we will see presently. But it did what it was designed to do, which is to prevent GCMs from blowing up. While the absolute humidity was predicted, it was not used in the radiative transfer calculations in the GCM simulations described in the 1965 paper, so they could not be said to be two-way interactive.

After the seminal work of Manabe and his colleagues, work on convection focused for some time on the problem of relating the intensity of convection to large-scale variables; this was often referred to as the "closure problem". In the 1960s and 70s, there was a move to relate the intensity of moist convection to the supply of water rather than of energy. Schemes based on the convergence of moisture in the boundary layer enjoyed a brief period of popularity, but suffered from certain pathologies. For example, large quantities of CAPE could accumulate in regions where the imposed criterion of positive moisture convergence was not satisfied. When moisture convergence ensued, CAPE was released suddenly, resulting in a strong positive feedback and grid-scale noise, sometimes referred to as "grid-point storms". The moisture-convergence schemes got into trouble because they largely disregarded the simple fact that convection is driven by thermal instability.

The central role of energy in controlling convection was re-affirmed by the "quasi-equilibrium" hypothesis introduced by Arakawa and Schubert (1974). This hypothesis holds that convection consumes available potential energy at the rate that it is supplied by larger-scale processes such as surface fluxes and radiative cooling, much as the rate of boiling in a pot of water is controlled by the heat input from the stove, rather than the temperature of the water, once the latter is at the boiling point.

The work of Arakawa and Schubert (1974) inaugurated a period of a few decades in which representations of moist convection were either built explicitly on the quasi-equilibrium hypothesis or implicitly, by assuming that convection restores thermal stability or neutrality on short time scales. This allowed researchers to focus their attention on the much more challenging problem of convective control of moisture.

As we have seen earlier in this chapter, moist convection keeps the temperature profile close to a suitably defined adiabat, but there is no simple constraint on the shape of the water profile. Moreover, it takes substantially longer for the moisture field to return to its RCE profile after being perturbed than it does the virtual temperature field. In essence, the adjustment time for virtual temperature is dictated by the time it takes deep internal buoyancy waves to travel between clouds (Bretherton and Smolarkiewicz, 1989). A typical phase speed of a mode whose vertical half-wavelength spans the troposphere is around $35\, ms^{-1},$ and if the cloud spacing is around $100\, km,$ this gives an adjustment time on the order of an hour. But moisture responds on a much longer time scale…basically, the time it takes for turbulent diffusion to spread moisture anomalies from clouds to their environment; internal waves act to eliminate buoyancy perturbations, not perturbations of passive tracers.

We can estimate a time scale for moisture adjustment by writing a budget equation for the moist static energy, $h_e,$ of air in between deep convective clouds:

$$\frac{\partial h_e}{\partial t} = -w_e \frac{\partial h_e}{\partial z} + \frac{Q_{rad}}{\rho} - \nabla \bullet \overline{\mathbf{V}' h_e'}, \tag{3.31}$$

where w_e is the vertical velocity in the clear air, Q_{rad} is the radiative heating rate, and the last term in (3.13) represents the eddy diffusion of moist static energy as the ensemble mean of eddy correlations of velocity and moist static energy. To get a time scale, we will look at a level in the atmosphere at which h_e has its minimum value, so that that the first term on the right of (3.13) is small there. We will suppose that variations in the radiative cooling rate are of

secondary importance in setting the adjustment time scale. Finally, we will suppose that the fluctuating part of the moist static energy scales as the difference between the cloud moist static energy, h^*, and the clear air value, while a characteristic inter-cloud spacing is denoted by L. Thus, from a scaling point of view, we have, from (3.31)

$$\frac{h_e'}{\tau_m} \sim \frac{V}{L}(h^{*\prime} - h_e'), \tag{3.32}$$

where V is a characteristic horizontal velocity in between clouds and τ_m is the moisture adjustment time scale we seek. From mass continuity in the air in between clouds,

$$\frac{V}{L} \sim \frac{w_e}{H},$$

where H is roughly half the depth of the troposphere. From this and (3.32),

$$\tau_m \sim \frac{H}{w_e}, \tag{3.33}$$

that is, the characteristic time scale for moisture adjustment is the time it takes air subsiding in between clouds to traverse half the troposphere. If we use (3.21) for the vertical velocity scale in the clear air, we get

$$\tau_m \sim \frac{\frac{\partial h_d}{\partial z} H \rho}{Q_{rad}} \sim \frac{L_v q^* \rho}{Q_{rad}}, \tag{3.34}$$

where in the second part we have made use of the fact that the total change in dry static energy along a moist adiabat is roughly equal to $L_v q^*$, where q^* is a characteristic value of the saturation mixing ratio in the lower troposphere. This shows that the adjustment time increases with temperature, since q^* varies according to Clausius-Clapeyron, while Q_{rad} increases somewhat more slowly with temperature.

If we use typical tropical values for the quantities in (3.34) ($q^* \simeq 0.02$ and a radiative cooling rate of $1\,K\,day^{-1}$) this time scale is around 35 days, about 400 times longer than the virtual temperature adjustment time scale. Later in this chapter, we will see that these numbers are consistent with those characterizing adjustment to equilibrium in a fairly comprehensive single-column model. In particular, the moisture time scale, as it is longer, is the rate-limiting factor in adjusting to RCE in systems in which the surface temperature is specified.

The great disparity between the buoyancy and moisture adjustment time scales has profound consequences for tropical dynamics and for practical aspects of representing moist convection in models. Moist importantly, some of the errors in the representation of convection may not be apparent on time scales much shorter than the moisture adjustment time, so that tests of convective schemes against observations require long time series of the latter.

Moist convective adjustment is fundamentally nonlocal. That is, introducing a perturbation to RCE at some particular level does not usually result in a relaxation at that level alone. For example, suppose a positive temperature perturbation is applied to the RCE state at some

instant in time. If it is large enough, convection will be temporarily prevented from reaching levels above that level, so that the convective heating will be temporarily shut down not just at the level in which the temperature perturbation was introduced, but all levels above that. Moreover, if this affects moist convective downdrafts, levels below the perturbation level will also be affected. The adjustment cannot be represented as a simple relaxation back to the RCE profile.

For virtual temperature, this may not matter much in practice because the total adjustment is fast and, integrated over the whole period of relaxation, the effect is simply to damp the original perturbation.

But for moisture, the non-locality of the adjustment may have repercussions over a significant period of time, affecting the evolution of the system on time scales that are not short compared to, e.g., atmospheric circulation systems.

For this reason, representations of moist convection that simply relax temperature and moisture profiles back to empirical equilibrium profiles have largely been abandoned in favor of approaches in which bulk measures of convective activity are based either explicitly on the quasi-equilibrium hypothesis, as in the original work of Arakawa and Schubert (1974) and its successors, or implicitly as in so-called “CAPE-based” mass flux schemes. In such schemes, convective mass fluxes are formulated so as to consume available potential energy quite rapidly (but not usually instantaneously), keeping CAPE from building up and thereby effectively, if somewhat gently, enforcing quasi-equilibrium. But the convective mass fluxes are then coupled with representations of cloud physical and turbulent processes to calculate water fluxes by the convection, and these are then used to calculate convective tendencies of water at each model level. In RCE, these tendencies will necessarily be zero above the subcloud layer, since there are no non-convective sources of water in the cloud-bearing layer.

Reasonably formulated mass-flux convective schemes keep the tropical temperature lapse rate near some moist adiabat or a profile based on neutrality to an entraining plume. But the convective moistening profile inevitably depends on myriad assumptions about how entrainment and cloud physics operate in clouds (and, indeed, how these to processes interact), so that a typical convective scheme has many parameters very few of which, if any, can lay any claim to universality. Moist convective schemes are Gothic affairs and it is hardly surprising that they are usually the first sub-grid scale representations to be blamed for what goes wrong with models, or that the community as a whole is trying to get away from them by explicitly simulating at least deep moist convective clouds, even in climate models. Yet some of the basic challenges remain with explicit convection including how to deal with the all-important (and usually sub-grid-scale) turbulence and associated entrainment and how to represent cloud microphysical processes, which will always be unresolved. A somewhat less explored issue is how deep moist convection controls stratiform clouds such as cirrus anvils and boundary layer clouds, all of which have strong influences on radiative transfer.

Among the many challenges of representing moist convection in models is dealing with clouds and water vapor in the upper troposphere. In the high portion of deep convective clouds, the condensed water content exceeds the saturation vapor content, often by a large factor, yet not all convective schemes detrain condensed water into the environment, and very few pay much attention to making good estimates of the amount of ice water in the tops of deep cumulonimbi. Although the absolute water vapor content of the upper troposphere is small, it is nonetheless a

very important absorber and emitter of radiation, as are high clouds, which also can have large albedos. We will see later in this chapter that high clouds are important in the phenomenon of self-aggregation of deep convection, so that getting them wrong may have serious consequences for weather and climate. In particular, the interaction of convectively produced high clouds with radiation is a key factor in low frequency variability in the equatorial region and in the genesis and intensification of tropical cyclones.

Owing to the large number of ad-hoc parameters it is vital to test them against observations. These tests can also help calibrate the parameters. One way to do this is to carry out field experiments designed to collect enough data over a sufficient period of time to supply a single-column model with everything it needs, except for convection, to predict the evolution of the vertical profile of humidity over time. This, of course, is much easier to say than to do.

Consider first the budget equation for the conservation of moist static energy, h, excluding a term proportional to the time dependence of the pressure field:

$$\frac{\partial}{\partial t}\left(\rho h\right) + \nabla \bullet \left(\rho \mathbf{V} h\right) = -\frac{\partial F_r}{\partial z} - \frac{\partial F_t}{\partial z}, \tag{3.35}$$

where $\mathbf{V}$ is the three-dimensional velocity vector, F_r is the radiative energy flux, and F_t is the total turbulent enthalpy flux, including that by moist and dry convection. These two fluxes are approximated as being purely in the vertical. To test the ability of a representation of moist convection to predict the evolution of the moist static energy, it is necessary to measure that evolution together with horizontal and vertical advection of moist static energy, radiative fluxes, and turbulent fluxes in the subcloud layer, including from the surface. And these must be done over periods of at least tens of days to capture the evolution time scales of moist static energy.

Making these measurements is quite challenging. Estimating the small but critically important vertical velocity is especially difficult. One strategy is to carefully measure horizontal velocities over a closed circle or polygon and use mass continuity to estimate the vertical velocity. One example of an effort to do this occurred during a field experiment, the Tropical Ocean Global Atmosphere (TOGA) Coupled Ocean Atmosphere Response Experiment (COARE), conducted in 1992-1993 in the tropical South Pacific. Figure 3.18 shows a polygon of 4 vertices at which rawinsondes were launched at 6-hour intervals.

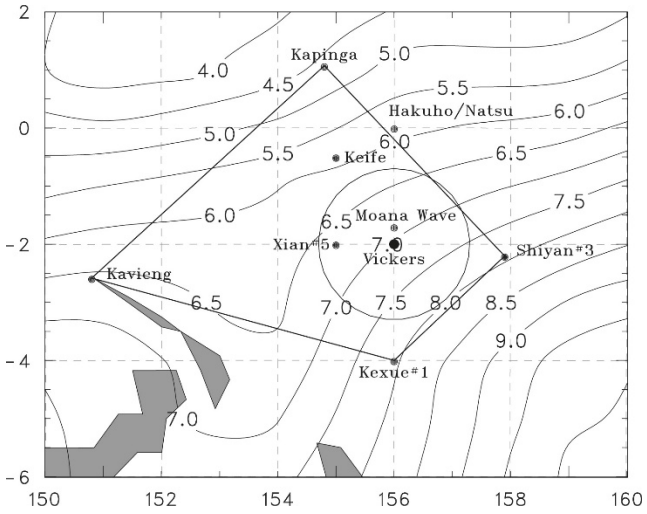


Figure 3.18: TOGA COARE Intensive Flux Array (IFA), which operated from November 1st 1992 to February 28th 1993. The contours are of mean rainfall (mm day^{-1}) over the 4-month period.

Unfortunately, there are several gaps in soundings from the two southeastern-most stations. In the end, estimating vertical velocity from mass continuity applied to horizontal wind and density estimates at only 4 stations is prone to sampling and instrumental errors.

In addition to rawinsonde measurements of temperature, moisture and horizontal winds, radiometers on moored buoys measured surface radiative fluxes, and top-of-the-atmosphere radiative fluxes were measured by satellites. These radiometric measurements do not capture the vertical structure of the radiative fluxes, and either an assumption must be made about their vertical structure, or radiative fluxes must be calculated from a radiative code. Surface sensible and latent heat fluxes were measured from moored buoys.

The lack of energetic consistency of the measurements can be demonstrated by integrating (3.35) through the depth of the troposphere, assuming that vertical velocity vanishes at the tropopause and using mass continuity:

$$\frac{\partial}{\partial t}\left(\widetilde{\rho h}\right)+\widetilde{\rho \mathbf{V}_2 \bullet \nabla h}+\widetilde{\rho w \frac{\partial h}{\partial z}}=F_{R0}-F_{TOA}+F_s, \tag{3.36}$$

where the symbol $\sim$ indicates a vertically integrated quantity, F_{R0} and F_{TOA} are the net upward radiative fluxes at the surface and top-of-the-atmosphere, respectively, and F_s is the turbulent flux of enthalpy from the surface. There is no contribution to (3.36) from moist convection because convection cannot change the column mean enthalpy, it can only redistribute it in the vertical.

We can now integrate (3.36) in time using observational estimates of the horizontal and vertical advections, radiative fluxes, and surface enthalpy fluxes. When this is done for the whole 4-month period of the TOGA COARE IFA measurements, the predicted value of the vertically averaged enthalpy drifts away from its observed value, with an accumulated error at the end of the period amounting to the equivalent of 25 K if that enthalpy error were owing to temperature only (Emanuel and Živkovic-Rothman, 1999). This is a spectacular error and must be accounted for before using the measurements to test the ability of the convection scheme to predict the time evolution of temperature and moisture. Given estimates of the magnitudes of likely errors in

vertical velocities and other quantities in (3.36), a variational approach can be developed to optimally correct for the errors (Zhang and Lin, 1997).

Integrating the moist static energy equation (3.35) for many weeks using a suitably adjusted set of measurements of the advective, radiative and surface flux terms and a convective representation for the last term in the equation provides a fairly rigorous test of the ability of the scheme to predict the evolution of moist static energy. It is an easy test to fail. Since almost all variations of moist static energy above the boundary layer in the tropics are due to variations of moisture, this effectively tests the scheme's ability to simulate vertical moisture fluxes. The scheme must also be able to predict the evolution of the temperature field, though this is much more difficult to evaluate as instrumental and sampling errors of temperature are comparable to typical real temperature variations in the deep tropics.

Ideally, a scheme should be able to pass similar tests forced by field measurements in very different meteorological conditions, but not many such field measurements exist. Obtaining the field measurements and then optimizing and evaluating convection schemes against such measurements is arduous and time-consuming. For this reason, not many convective schemes have been rigorously tested against field measurements and instead are optimized to improve the performance of the numerical weather prediction and/or climate models in which they are embedded. But this is problematic, as the optimization may result in a partial compensation for other model errors.

In this text we will use the representation of convection by Emanuel and Živkovic-Rothman (1999). This scheme regulates the flux of mass though cloud base according to the buoyancy of air lifted reversibly and adiabatically from the subcloud layer to the first model level above the lifted condensation level. More buoyancy yields a greater mass flux, resulting in faster stabilization of the atmosphere. The effect is to drive convection toward a form of energy quasi-equilibrium, where convection removes available potential energy at about the rate it is generated by larger-scale processes such as radiative cooling and surface enthalpy fluxes.

The mass ascending through cloud base in then partitioned into a number of branches that may be as large as the number of model levels between cloud base and cloud top. Air in each branch first ascends reversibly and adiabatically to a particular model level, where a fraction of the condensed water is removed. This fraction depends on the water content itself as well as temperature. Once the condensed cloud water has been adjusted, the sample mixes with the unperturbed environment, resulting in an integer number of samples of varying fractions of environmental air mixed in. Each mixture then ascends or descends, reversibly and adiabatically, to its level of neutral buoyancy, where it is detrained into the environment.

The removed condensate is added to a single, unsaturated downdraft driven by evaporation of the falling condensate; this unsaturated downdraft transports energy and water and can affect the thermal properties of the subcloud layer.

The convection scheme is also coupled to a representation of stratiform clouds, such as anvil clouds, that are a direct result of moist convection in the atmosphere (Bony and Emanuel, 2001). Downdrafts produced by the scheme also alter surface fluxes primarily by changing the mean wind speed at the surface.

The free parameters of the scheme include the specification of the fractions of condensed water converted to precipitation as functions of temperature and water content, the rate of fall and re-

evaporation of precipitation, and the fraction of precipitation that falls though environmental air, as opposed to cloud. All of these parameters affect the behavior of the scheme and the evolution of atmospheric water vapor in models that employ the scheme. The values of the parameters have been adjusted to optimize the evolution of the vertical profile of relative humidity in a single-column model driven by TOGA-COARE IFA data, as described earlier in this section.

To provide a relatively straightforward simulation of moist RCE, we design a “single-column model” (sometimes referred to as an “SCM”) that consists of the aforementioned convection scheme coupled to the radiative transfer scheme of Morcrette (1991). The radiative transfer scheme is able to handle clouds and aerosols, and the cloud properties are represented according to Bony and Emanuel (2001). Concentrations of the important long-lived greenhouse gases and aerosols are prescribed as constants or as functions of height, but clouds and water vapor are calculated interactively. Dry convection is represented by a simple dry-adiabatic adjustment, and conserved tracers and water concentration are made uniform with height by any such adjustments. If water condenses explicitly during the adjustment, the condensed water is assumed to precipitate with no re-evaporation. Surface fluxes are calculated using simple aerodynamic flux formulae, with a background wind supplied as an external parameter, augmented by downdrafts produced by the moist convection scheme.

Finally, the surface itself is represented as a single slab of water, although the evaporation rate can be artificially altered to loosely simulate land surfaces with restricted water availability. The slab’s temperature can be either specified or calculated from surface energy balance.

The SCM is integrated forward in time starting from a specified sounding whose vertical structure also specifies the vertical structure of the SCM. Although the model rarely achieves a strictly time-independent state, it does eventually come into a state of statistical equilibrium in which there are no trends in the thermodynamic properties of the system.

The SCM described here can be run on ordinary laptop or desktop computers and is available at https://zenodo.org/doi/10.5281/zenodo.1079522 .

The final RCE state is not steady but rather fluctuates around a stable equilibrium state, with near balances between precipitation and surface evaporation, and between incoming solar and outgoing longwave radiation.

The temperature profile of the equilibrium state is compared to that of pure radiative equilibrium and radiative-dry-convective equilibrium in Figure 3.19. Note that the moist RCE solution is very close to an average observed vertical profile of temperature in the tropics, except in the lower stratosphere. In this case, the surface air temperature is only a few degrees cooler than the sea surface temperature, but the relative humidity of near surface air is only around 80%, so there is still a profound thermodynamic disequilibrium between the sea and the air, showing up mostly in the latent heat term.

The state shown in Figure 3.19 is still not a fully interactive calculation as the relative humidity has been fixed at an observed profile. Figure 3.20 shows the relative humidity in a fully interactive calculation, in which the precipitation efficiencies are dependent on condensed water content and temperature. This is compared to the humidity profile of three other simulations, in which the surface temperature has been fixed at the control value but the precipitation

efficiencies have been set to either 0.1 or 1.0 everywhere, and in the last case the evaporation of precipitation has been turned off.

In the control experiment, the precipitation efficiencies and other parameters have been set so as to optimize a comparison between predicted and observed relative humidity, as a function of altitude, during the 4 months of TOGA COARE IFA measurements (Emanuel and Živkovic-Rothman, 1999). When the precipitation efficiency is set to 0.1 everywhere, the entire upper troposphere becomes saturated owing to the high condensed water content of air detrained from deep convective clouds. The lower troposphere actually becomes drier than the control, at least in part because the control precipitation efficiencies are less than 0.1 there. Thus there is more condensed water detrained from shallow clouds in the control experiment.

When, on the other hand, the precipitation efficiencies are set to 1 everywhere, the whole troposphere becomes substantially drier, except in a thin layer just above the subcloud layer. When in addition, the evaporation of precipitation is turned off, a dry limit is reached that is determined by the balance between drying by subsidence between clouds and detrainment of just-saturated air from the clouds.

While the imposed changes in precipitation efficiencies here are deliberately extreme, they illustrate that the humidity of the atmosphere in RCE is sensitive to cloud microphysical processes.

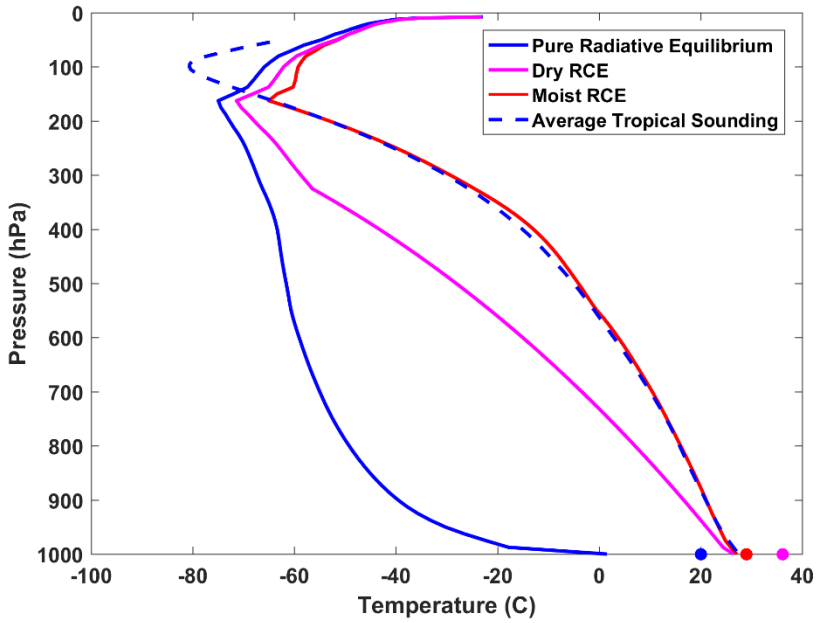


Figure 3.19: As in Figure 2.14 but also showing a fixed relative humidity, moist RCE solution from the single-column model in red.

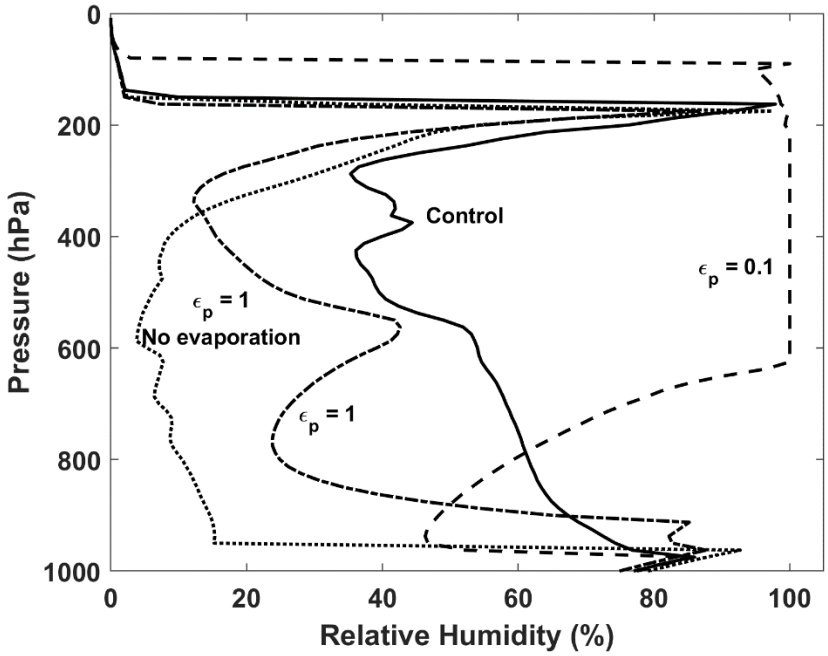


Figure 3.20: Profiles of relative humidity (%) in moist RCE for the standard formulation of parcel precipitation efficiencies (solid), for $\epsilon_p = 0.1$ (dashed), $\epsilon_p = 1$ (dash-dot), and $\epsilon_p = 1$ with no evaporation of precipitation (dotted).

Table 1 shows the external conditions for the control experiment, driven by annual average solar radiation, and omits all clouds, while Table 2 lists values of the some of the important parameters averaged over the last 100 days of the experiment.

Table 1: Control run conditions

Latitude (degrees)	Surface albedo	Mean Surface Wind Speed (m/s)	CO_2 (ppm)	CH_4 (ppm)	N_2O (ppb)	CFC-11 (ppt)	CFC-12 (ppt)
23.0	0.248	5.0	360.0	1.72	310.0	280.0	484.0

Table 2: Value of key time-averaged quantities, control run

Surface T (C)	Precipitation (mm/day)	Evaporation (mm/day)	TOA net shortwave (W/m^2)	TOA net longwave (W/m^2)	Surface net shortwave (W/m^2)	Surface net longwave (W/m^2)	Net surface radiative (W/m^2)	Surface sensible (W/m^2)	Surface latent (W/m^2)	Net surface turbulent flux (W/m^2)
27.0	4.215	4.222	-296.71	296.89	-217.13	85.10	**-132.03**	10.81	121.46	**132.27**

Note the near balances between surface evaporation and precipitation, top-of-the-atmosphere (TOA) net shortwave (incoming minus reflected shortwave) and outgoing longwave radiation, and net surface radiative and turbulent fluxes. In principle, there should be exact balances, but this is almost never seen in models because of both physical and numerical errors in the computations. For example, heating by the frictional dissipation of falling precipitation in the model amounts to about $1.4\,Wm^2$, almost an order-of-magnitude greater than the energy imbalance at the top of the atmosphere. Thus omitting this seemingly insignificant energy conversion would greatly increase the magnitude of the TOA energy imbalance. Similarly, cooling of the air by precipitation falling through a temperature gradient cools the system by around $0.6\,Wm^2$. The convection scheme internally advects air upward and downward, and while real advection is thermodynamically reversible, finite difference schemes may not exactly conserve both energy and entropy, so some of the excess upward longwave radiation at the top of the atmosphere may be owing to numerical generation of entropy in the convection scheme. Getting near balance in TOA radiation requires great attention to apparently small thermodynamic details.

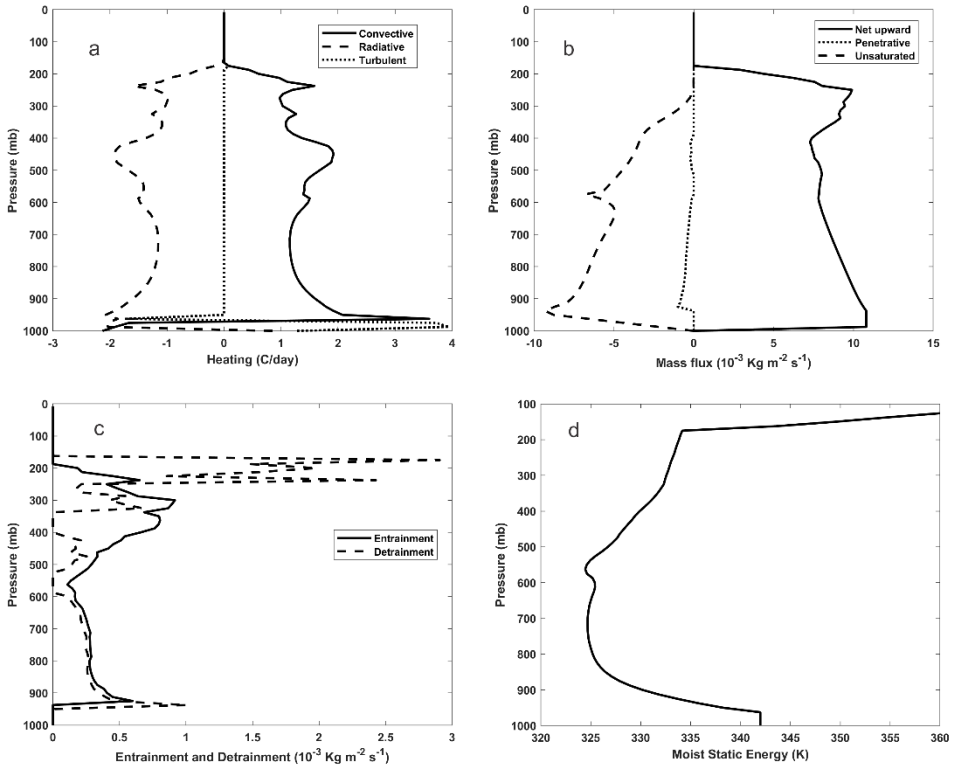


Figure 3.21: Vertical profiles of control moist radiative-convective equilibrium solution. a) Heating rates (expressed in K day^{-1}): Net convective (solid), radiative (dashed) and boundary layer convective adjustment (dotted); b) Convective mass fluxes: Upward (solid), saturated in-cloud downdrafts (dotted) and unsaturated downdrafts driven by evaporation of precipitation (dashed); c) Convective entrainment (solid) and detrainment (dashed); and d) moist static energy (divided by heat capacity). These profiles have been averaged over the last 10 days of the simulation.

Vertical profiles of various quantities, averaged over the final 10 days of the control simulations, are shown in Fig. 3.21. Convection heats the troposphere at about 1.5 K day^{-1}, with only fairly weak variations with height, and radiation cools the atmosphere at the same rate. In the subcloud layer, dry convection heats the boundary layer at about 3.5 K day-1, compensating both radiative cooling and cooling by evaporation of rain, the latter of which is included in the definition of convective heating. Fig. 3.21b shows the vertical profiles of convective mass fluxes, partitioned into updrafts, saturated downdrafts driven by the evaporation of cloud water, and unsaturated downdrafts driven by the evaporation of falling precipitation. Saturated downdrafts play a minor role in RCE but, as we shall see in the next chapter, are very important in shallow convective boundary layers.

The convection scheme used here has a spectrum of updrafts and downdrafts, so the net mass entrainment/detrainment is not a simple vertical derivative of the net mass fluxes. For example, air at some level may both detrain from a shallow cloud and be entrained into a deep cloud. The profiles of entrainment and detrainment for this RCE state are shown in Figure 3.21c. Note that profiles of both are quite similar below 600 hPa, but diverge above that level, with strong detrainment near the tropopause, as one would expect.

The equilibrium moist static energy profile (Fig 3.21d) shows a broad minimum in the middle troposphere and a well-mixed subcloud layer. The profile is qualitatively very similar to that of equivalent potential temperature (θ_e) seen in tropical soundings (e.g. Fig. 3.12). Convection mixes high moist static energy air in the subcloud layer into the free troposphere, coupling the strong surface source with the deep troposphere, while radiative cooling provides the balancing sink.

3.3.2 Behavior of global perturbations to RCE

The statistical equilibrium state of moist RCE is a useful zero-order approximation to the actual state of the tropical atmosphere, and thus a good starting point for understanding the meteorology of the tropics. Naturally, the real atmosphere is evolving in time and there can be substantial flows of energy by large-scale circulations, pushing the state away from RCE. As a first step to understanding the effects of time dependence and horizontal flows of energy and water, we can ask how the RCE state responds to global (that is, horizontally homogeneous) perturbations in the boundary and/or initial conditions, focusing on the characteristic time scales and structural aspects of the response.

In essence, we can characterize the response by three time scales: The time it takes for the troposphere to relax back to RCE holding the surface temperature fixed, the time it takes the surface to relax back to RCE holding the whole atmosphere fixed, and the relaxation time scale for the coupled system. This last is a function of the first two time scales, but defies intuition in being much longer than either the ocean-only or atmosphere-only time scales.

We begin by performing some simple experiments with the single-column model described earlier in this chapter. These experiments all begin with a moist RCE state achieved by running the model into equilibrium with a slab ocean surface whose temperature is calculated from surface energy balance. In the first experiment, we fix the surface temperature and subtract 3 K from all the air temperatures below the 100 hPa level, then run the model until equilibrium is regained.

As can be seen in Fig. 3.22, the recovery of precipitation and temperature in the troposphere has a time scale of around 10 days. Above the RCE tropopause, located near the 175 hPa level, the relaxation time scale is somewhat longer. Note that the relaxation cannot be described as a smooth exponential. This may be partially owing to the nonlinearities of radiative and convective transfer and also to the particular way the convection scheme is designed.

Fig. 3.23 shows the results of holding atmospheric temperature and water vapor fixed at their RCE profiles and perturbing the initial sea temperature by -3 K. In this case, the surface is a 1m thick slab of water. Recovery to RCE is smooth and rapid, occurring with a relaxation time scale of around 1-2 days. Here it should be noted that the recovery time is proportional to the thickness of the slab of water. Thus the recovery of a 50 m slab, more representative of the ocean's surface mixed layer, would have a relaxation time scale of around 50-100 days.

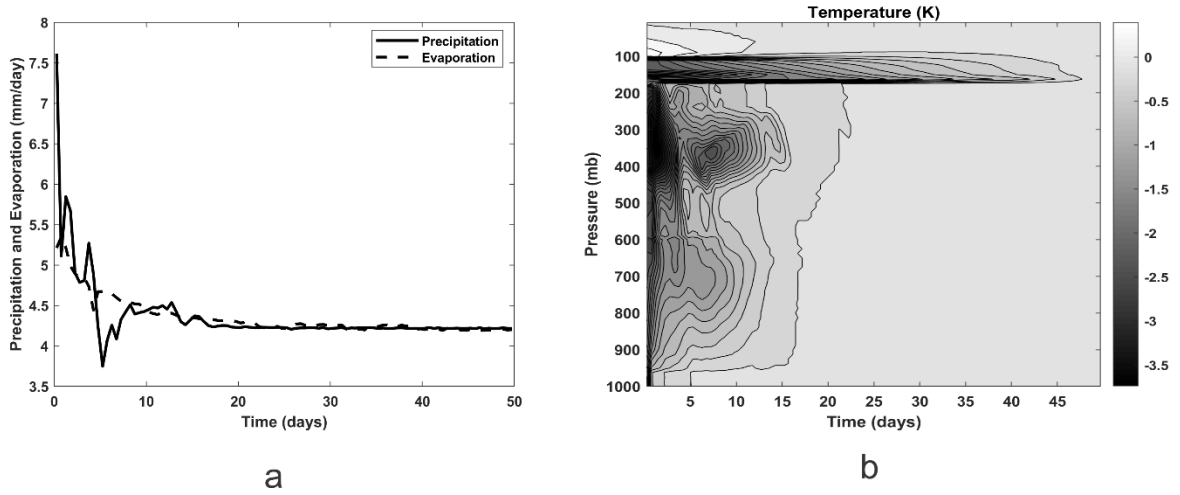


Figure 3.22: Recovery from 3K cooling below 100 hPa to a moist RCE state. The surface temperature is fixed at its RCE value. a) Surface precipitation (solid) and evaporation (dashed), in mm/day. b) Evolution of the departure of temperature (K) from RCE.

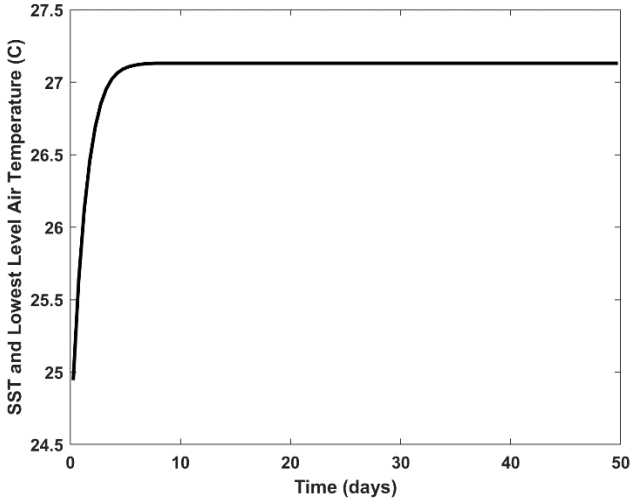


Figure 3.23: Recovery of a 1m thick ocean slab to RCE from an initial -3 K anomaly. In this experiment, the properties of the atmosphere are held constant at their RCE values.

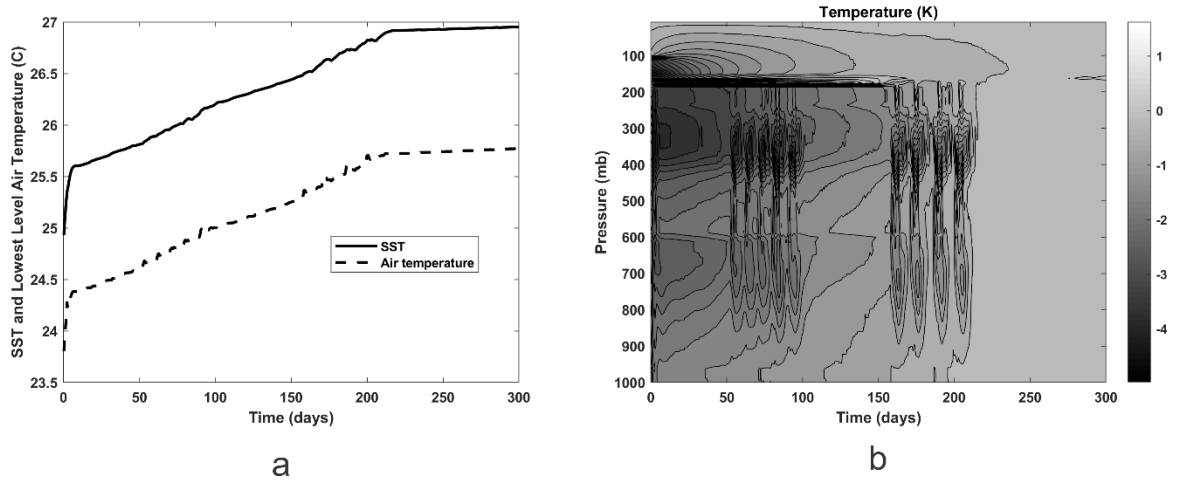


Figure 3.24: Recovery of the coupled ocean-atmosphere system to RCE from a -3K perturbation to the atmosphere below the 100 hPa level and a -2 K perturbation to the temperature of a 1m thick slab of water. a) Water temperature (solid) and surface air temperature (dashed), in C. b) Time-pressure section of the perturbation of temperature (K) from the RCE state.

The relaxation of the coupled system back to RCE is shown in Fig 3.24. The time scale of the relaxation is around 200 days, *much* longer than either the atmospheric relaxation time of 10 days or the ocean time of 5 days.

Why is the coupled time scale so much longer than either of the uncoupled scales? To see this, we will construct a very simple model of the coupled system, illustrated in Figure 3.25 (Cronin and Emanuel 2013). A troposphere with a uniform value of saturation most static energy, h^*, overlies a slab of water of depth Δz_{ml} and a turbulent enthalpy flux F_c transports energy from the ocean to the atmosphere. At the top of the atmosphere the longwave radiative flux is F_{rad}. The following simplifying assumptions are made to obtain analytic solutions:

- The lower troposphere is virtually opaque to infrared radiation, so that all the energy flux from the water to the air is carried by turbulence.
- The relative humidity profile of the atmosphere is fixed, so that changes in the radiative flux may be expressed in terms of changes in h^* alone.
- Changes in the absorption of solar radiation by the atmosphere may be ignored.

With these three assumptions, we may write down simple equations for the time evolution of fluctuations in atmospheric temperature, as represented by h^* and water temperature, as represented by h_0^*:

$$\frac{\Delta p}{g}\frac{dh^{*\prime}}{dt} = -\frac{dF_{rad}}{dh^*}h^{*\prime} + F_c', \tag{3.37}$$

$$c_l \rho_l \Delta z_{ML} \frac{dT_s}{dh_0^*}\frac{dh_0^{*\prime}}{dt} = -F_c'. \tag{3.38}$$

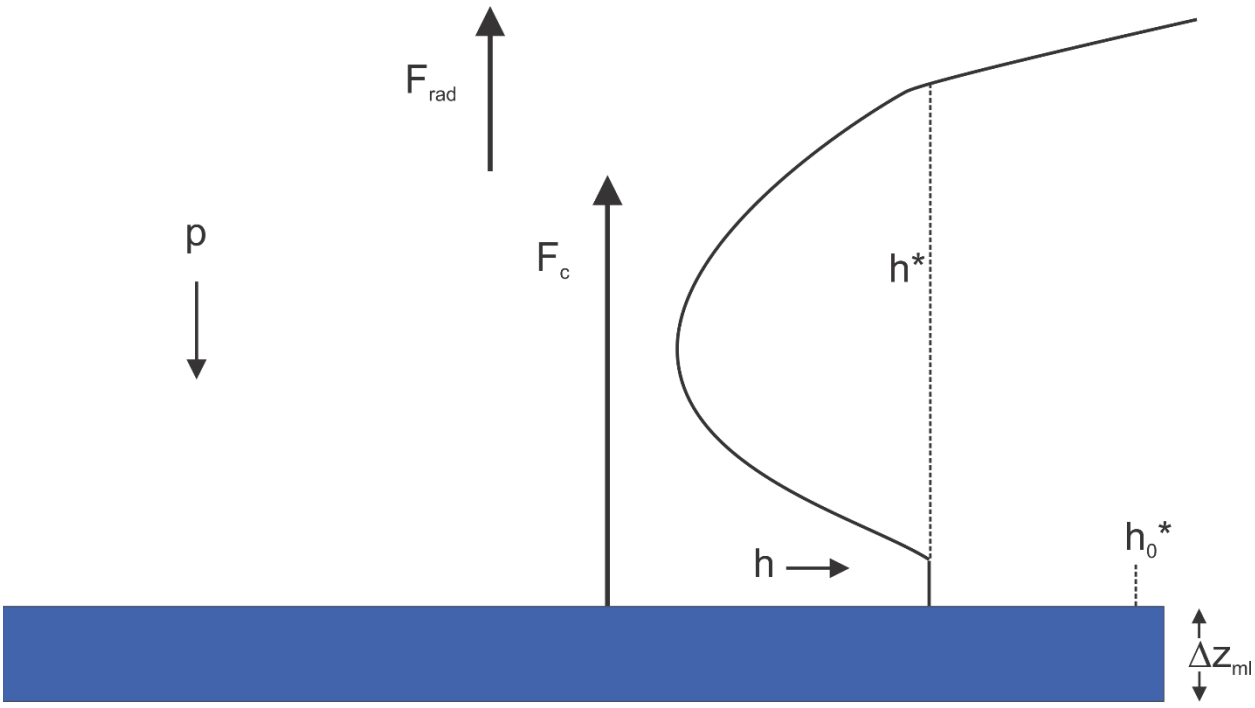


Figure 3.25: Simple model of the coupled system. A troposphere with a moist adiabatic lapse rates rests on a slab of water of uniform temperature and depth Δz_{ml}. The saturation moist static energy corresponding to the water temperature is h_0^*. The temperature of the troposphere is characterized by a vertically uniform saturation moist static energy h^*. The turbulent flux of enthalpy from the water to the atmosphere is F_c and the outgoing longwave radiative flux is F_{rad}.

Here Δp is the pressure depth of the troposphere, g is the acceleration of gravity, and c_l and ρ are the heat capacity and density of liquid water.

Next we represent the surface turbulent enthalpy flux by a classical aerodynamic flux formula:

$$F_c' = \rho_s C_k |\mathbf{V}_s| \left(h_0^{*\prime} - h^{*\prime} \right), \tag{3.39}$$

where ρ_s is the surface air density, C_k is the dimensionless coefficient of enthalpy exchange, and $|\mathbf{V}_s|$ is the surface wind speed, which must be externally specified here as in the single column model. In the last term in (3.39) we have replaced the boundary layer moist static energy by h^* as the two are equivalent if the atmosphere is neutral to moist convection.

The coefficient $\frac{dF_{rad}}{dh^*}$ that appears in (3.37) may be expressed in terms of the traditional climate sensitivity parameter $\lambda \equiv \frac{\partial T_s}{\partial F_{rad}}$, where T_s is the surface air temperature:

$$\frac{dF_{rad}}{dh^*} = \frac{dT_s}{dh^*} \lambda^{-1}. \tag{3.40}$$

The coefficient $\frac{dT_s}{dh_0^*}$ that appears in (3.38) may be derived from the definition of h^* and the Clausius-Clapeyron equation:

$$\frac{dT_s}{dh_0^*} = \left(c_p + \frac{L_v^2 q_s^*}{R_v T_s^2} \right)^{-1} \tag{3.41}$$

We also use this expression for $\frac{dT_s}{dh^*}$ that appears in (3.40) since the surface air and water temperatures are very close to each other in the tropics.

Using (3.39) - (3.41) allows us to express (3.37)-(3.38) as

$$\tau_a \frac{dh^{*\prime}}{dt} = \chi h_0^{*\prime} - h^{*\prime}, \tag{3.42}$$

and

$$\tau_o \frac{dh_o^{*\prime}}{dt} = h^{*\prime} - h_o^{*\prime}, \tag{3.43}$$

where τ_a and τ_o are atmospheric and ocean relaxation time scales given by

$$\tau_a \equiv \frac{\Delta p}{g} \left(c_p + \frac{L_v^2 q_s^*}{R_v T_s^2} \right) \frac{\lambda}{1 + \lambda c}, \tag{3.44}$$

and

$$\tau_o \equiv \frac{c_l \rho_l \Delta z_{ml}}{c}, \tag{3.45}$$

with

$$c \equiv \rho_s C_k \mid \mathbf{V}_s \mid \left(c_p + \frac{L_v^2 q_s^*}{R_v T_s^2} \right), \tag{3.46}$$

and

$$\chi \equiv \frac{\lambda c}{1 + \lambda c}. \tag{3.47}$$

The form of equations (3.42) and (3.43) makes it clear that τ_a governs the rate at which the atmosphere relaxes to a fixed ocean, while τ_o governs the rate at which the ocean relaxes to a fixed atmosphere. These equations admit exponentially decaying solutions of the form $e^{-t/\tau}$, where

$$\tau = \frac{2\tau_o}{1 + \tau_o / \tau_a \pm \sqrt{\left(1 + \tau_o / \tau_a\right)^2 - 4\left(1 - \chi\right)\tau_o / \tau_a}}. \tag{3.48}$$

Clearly, the longer time scale will be associated with taking the negative root in the denominator of (3.48). From (3.47), if $\lambda c >> 1,$ then the last term in the denominator of (3.48) will be relatively small compared to the first term in the discriminant and we can approximate (3.48) as

$$\tau \cong \frac{\tau_a + \tau_o}{1 - \chi} = \left(\tau_a + \tau_o\right)\left(1 + \lambda c\right). \tag{3.49}$$

Consequently, if $\lambda c >> 1,$ the coupled time scale given by (3.49) will be large compared to the sum of the time scales of the two individual components of the system.

For example, for an RCE ocean temperature of 27° C, a surface wind speed of $5\, ms^{-1}$, a tropospheric pressure depth of 800 hPa, a water depth of 1 m, and a climate sensitivity (λ) of $0.5\, K/(Wm^{-2})$, the value of λc is about 16. The atmospheric relaxation time scale is around 17 days, the ocean time scale is about 1.5 days, but the coupled time scale is about 325 days, or almost a year. Even in the limit of vanishing water depth, the coupled time scale is about 300 days. For a more realistic ocean mixed layer depth of 50 m, the ocean relaxation time scale is about 70 days and the coupled time scale is about 560 days.

So coupling the atmosphere to a surface—even a land surface with vanishing heat capacity—increases the relaxation time scale of the system by at least an order-of-magnitude.

One way to think about this is to recognize that the case of fixed water temperature is the same thing as taking the depth of the ocean slab to be infinite. In essence, the atmosphere relaxes back to the ocean on its atmospheric relaxation time scale. When, at the opposite limit, the depth of the water approaches zero, there can be no perturbation heat flux from the lower boundary (since the lower troposphere is considered opaque in the IR and the solar flux is constant) so the atmosphere must relax purely by internal radiative relaxation: The balance in

(3.37) is then between the relaxation on the left side and the perturbation radiative flux, as represented by the first term in the right side. The time scale is now $\lambda \frac{\Delta p}{g}\left(c_p + \frac{L_v^2 q_s^*}{R_v T_s^2} \right)$, which is the climate sensitivity multiplied by the effective heat capacity of the atmosphere. (At constant relative humidity, the energy needed to heat a sample of air by 1 K includes the energy needed to increase the latent heat content of the sample. With a temperature of 300 K, the latent heat term is more than 5 times the sensible heat contribution to the heat capacity.) For the parameters listed above, this is about 300 days.

In contemplating how the RCE state reacts to perturbations, it is important to keep in mind that both convection and radiation are nonlocal: perturbations at any level can have substantial effects at any other level. In the case of radiation, this is essentially instantaneous and extends through the whole atmosphere. Convection has a response time ranging from hours to days, and so the effect is not instantaneous but is nonlocal through the depth of the convecting layer. One should not think of the reaction to perturbations as constituting anything like a simple Newtonian relaxation.

To illustrate this, we perturb the RCE state of the single-column model by reducing the specific humidity by 70% at 800 hPa, decaying to zero 100 hPa above and below that level, with the ocean temperature fixed. Figure 3.26 shows the response of the model to this rather large perturbation.

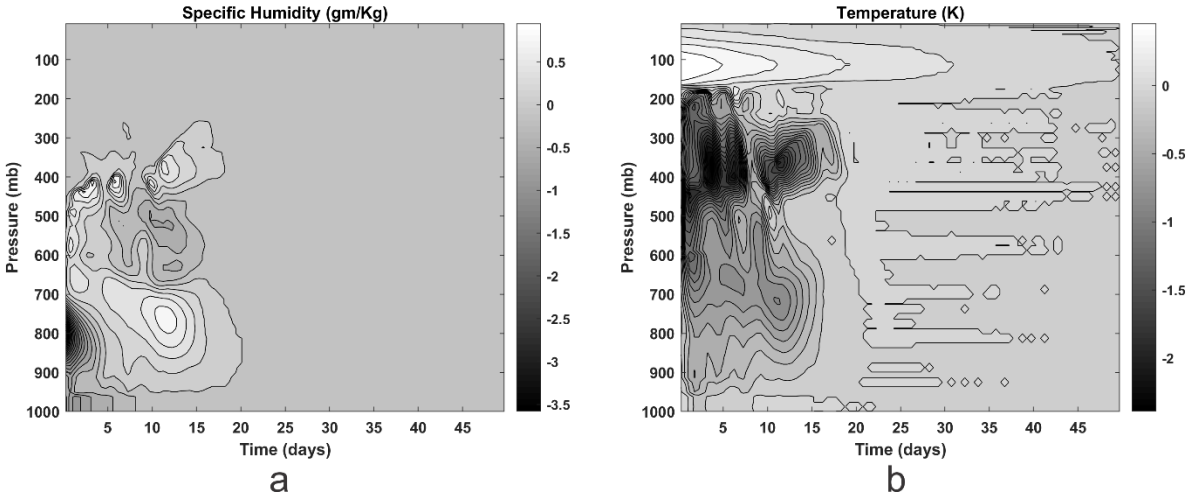


Figure 3.26: Response of the single-column RCE state to a large negative anomaly of water vapor centered at 800 hPa, holding the surface temperature fixed. a) Perturbation specific humidity (g Kg^{-1}); b) Perturbation temperature (K).

Several aspects of the response are particularly interesting. First, far from being Newtonian, the initial water vapor anomaly in the lower troposphere reverses sign on the way to recovery. And, even though the temperature was not perturbed, there is a strong response nonetheless. The initial dry perturbation greatly reduces convective heating, so the column cools, particularly in the upper troposphere. The initial recovery of convection is fairly shallow, leading to excessive moistening of the lower troposphere by day 10, while the upper troposphere continues to cool in the absence of deep convection.

The important point is that the response of RCE to global perturbations is both nonlocal in altitude and non-Newtonian in time. And, as we saw earlier, the time scale to return to RCE is greatly lengthened by coupling to an interactive surface.

3.3.3 Behavior of local perturbations to RCE

Imagine beginning with a moist RCE state and then artificially adding heat to (or subtracting it froml features) a localized patch of the underlying ocean. This can happen, for example, when ocean currents converge heat into or away from area. Alternatively, the surface winds may change in a particular place, increasing or decreasing the surface enthalpy flux without changing the ocean temperature very much on short time scales.

If the patch of ocean in question is very small, then advection of unperturbed RCE air over the patch would no doubt minimize any response to it. At the other extreme, if the patch is very large, we are effectively back to the global RCE perturbation discussed in the previous section.

In between, at intermediate scales, we would intuitively expect to find more convection and rising air over a warm patch, and sinking air and less convection, or perhaps no convection, over a cold patch.

In either case, the density temperature over the anomalously warm or cold surface cannot depart much from that of the surrounding atmosphere, because of it did, internal waves would quickly eliminate any differences. If the Coriolis parameter is nonzero and the surface patch is not small compared to an internal deformation radius, then rotational dynamics will eventually allow density temperature gradients to build up while maintaining approximate geostrophic and hydrostatic balance. Moreover, frictional stresses in the boundary layer can potentially balance pressure gradients, so we would not be surprised to find temperature gradients there.

Yet above the boundary layer, and for surface patches that are much smaller than the deformation radius (which is large in the tropics), we expect that for sufficiently slow systems the density temperature will not vary much on isobaric surfaces and we can use this fact to estimate the response of the atmosphere to the surface perturbations. This approximation, introduced by Sobel and Bretherton (2000), has become known as the *Weak Temperature Gradient* (or “WTG”) approximation and it is a very powerful means of understanding quasi-steady circulations in the tropics.

When combined with a method of representing moist convection, WTG allows us to calculate the vertical motion of the tropical atmosphere. The general idea is illustrated in Fig. 3.27.

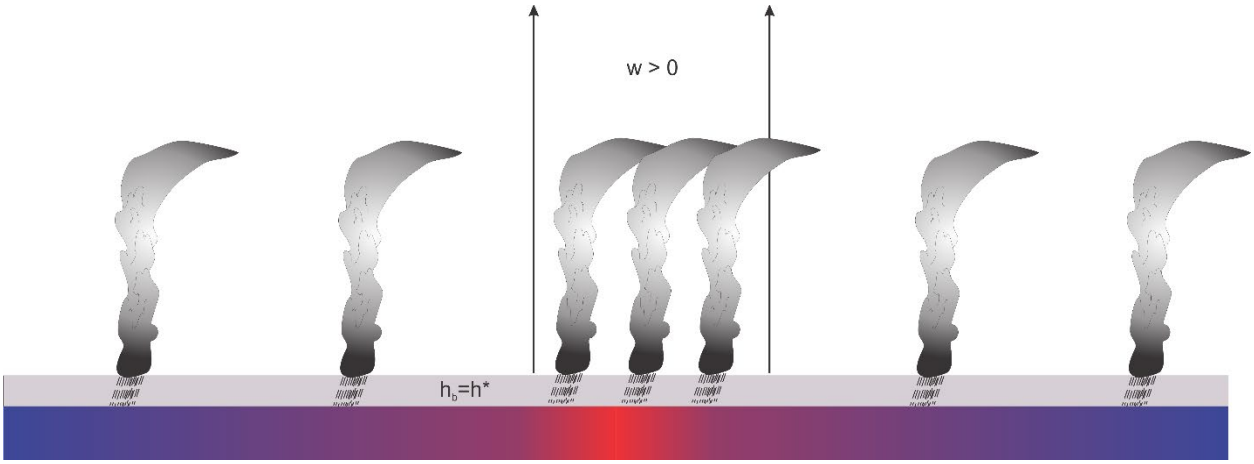


Figure 3.27: Response of an RCE atmosphere to a positive ocean temperature anomaly in the center of the figure. Colored bar at bottom denotes ocean temperature. The gray bar above that represents the subcloud layer, with a moist static energy equal to the uniform value of the saturation moist static energy, h, of the free troposphere. The excess surface enthalpy flux over the warm water causes extra convection, which leads to excess heating of the column. In the WTG approximation, this is offset by the adiabatic cooling associated with large-scale ascent of magnitude w.*

Over the warmer water, assuming that the surface wind speed is not smaller, there will be larger enthalpy flux from the sea. This will drive excess moist convection, which will cause excess heating of the column unless the radiative cooling also increases over the warm patch. Assuming that this is not the case, to obey the WTG approximation, there will have to be just enough large-scale ascent (arrows in Fig. 3.27) so that the adiabatic cooling associate with that ascent just cancels the excess convective heating. Therefore, application of WTG yields the large-scale ascent.

WTG, coupled with the principle of moist convective neutrality, implies that the subcloud layer moist static energy, h_b is equal to the uniform value h^* of the saturation moist static energy of the troposphere. Neglecting the direct effects of water substance on density, h^* will be constant in the vertical, up to the tropopause, because the troposphere will have a moist adiabatic lapse rate. It will also be constant in the horizontal, according to WTG. Thus the whole tropospheric temperature can be characterized by a single, constant value of h^* [3].

If the domain is horizontally infinite, then the value of h^* is that of the RCE state. But if the domain has finite dimensions, then h^* will have to be recalculated such that there is gentle, large-scale descent outside the warm patch of sufficient strength to make the domain-average vertical velocity vanish. We will see how to do this presently.

Suppose that, instead, we insert a patch of water in the middle of the domain that is colder than that of the RCE state; this is illustrated in Fig. 3.28.

[3] This does not account for the direct effects of water substance on density.

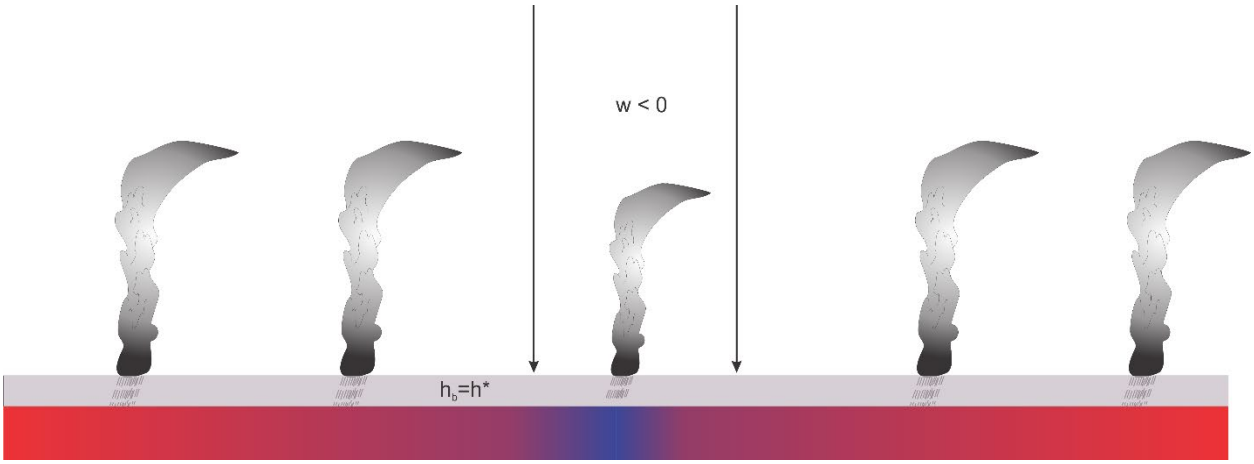


Figure 3.28: As in Fig. 3.27 but with colder water in the middle of the domain.

Now there is descent over the colder water, owing to deficient surface enthalpy fluxes there. If the water is not too cold, the deep convective mass flux will decrease but not vanish, and one can still assume that the atmosphere is convectively neutral, with subcloud layer moist static energy equal to the h^* of the free troposphere. But if the patch is cold enough, relative to the surrounding RCE state, the deep convection shuts off, though shallow convection may remain (Fig. 3.29), and the large-scale descent rate is limited by the balance between subsidence warming and radiative cooling. In this case, convective neutrality cannot apply and the subcloud layer becomes thermodynamically decoupled from the free troposphere. Now we have to solve a rate equation to find the value of the subcloud layer moist static energy.

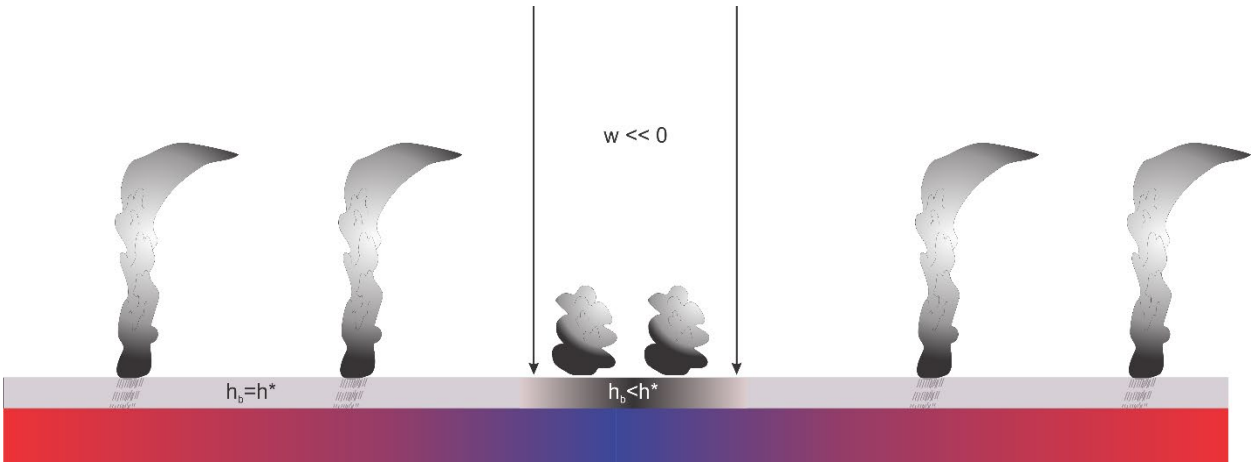


Figure 3.29: As in Fig. 3.28 but with colder water in the center. Deep convection has vanished over the cold water and the large-scale descent rate is that for which subsidence warming balances radiative cooling.

Once we have calculated the vertical velocity everywhere, we can use mass continuity together with the vorticity equation to calculate the entire velocity field. If there is no background vorticity, then no relative vorticity will develop and the resulting horizontal flow will be irrotational. But the full momentum equations can be inverted to find the distribution of pressure in the system, and through the hydrostatic equation, perturbations to the density temperature. The original WTG can, in principle, be modified to include these perturbations and the calculation repeated. For

typical tropical circulations like the Walker circulation, these temperature perturbations amount to a few tenths of kelvins.

If the system has some background vorticity, as is usually the case off the equator, then larger temperature perturbations can build up, particularly if the scale of the surface perturbations is not small compared to a deformation radius. But the same reasoning applies… once the vertical motion has been calculated, the irrotational flow can be calculated from mass continuity and the nondivergent part can then be calculated from the vorticity equation. The momentum equations can then be inverted to find the pressure and thus the density temperature perturbations, and a correction can be made to the original WTG state and the calculation repeated.

The important point here is that for systems for which WTG is a good approximation, dynamics take a back seat to the thermodynamic processes driving the system and the horizontal flows can be estimated from the distribution of vertical velocity. Naturally, the whole system must be dynamically and thermodynamically self-consistent, but the easiest way to think about systems like these is to think about what distribution of large-scale vertical velocity is consistent with the thermodynamic forcing, and then what distribution of horizontal velocity is consistent with that vertical velocity via mass continuity and the momentum equations.

The response of the RCE state to local perturbations can be quantified by marrying the WTG principle with a representation of moist convection. For the latter, we introduce the principle of boundary layer quasi-equilibrium (Emanuel, 1993). This simply maintains that convective activity must be sufficient to remove moist static energy from the subcloud layer at the rate at which surface fluxes and horizontal advection supply it, so that there is no accumulation of moist static energy in the relatively thin subcloud layer.

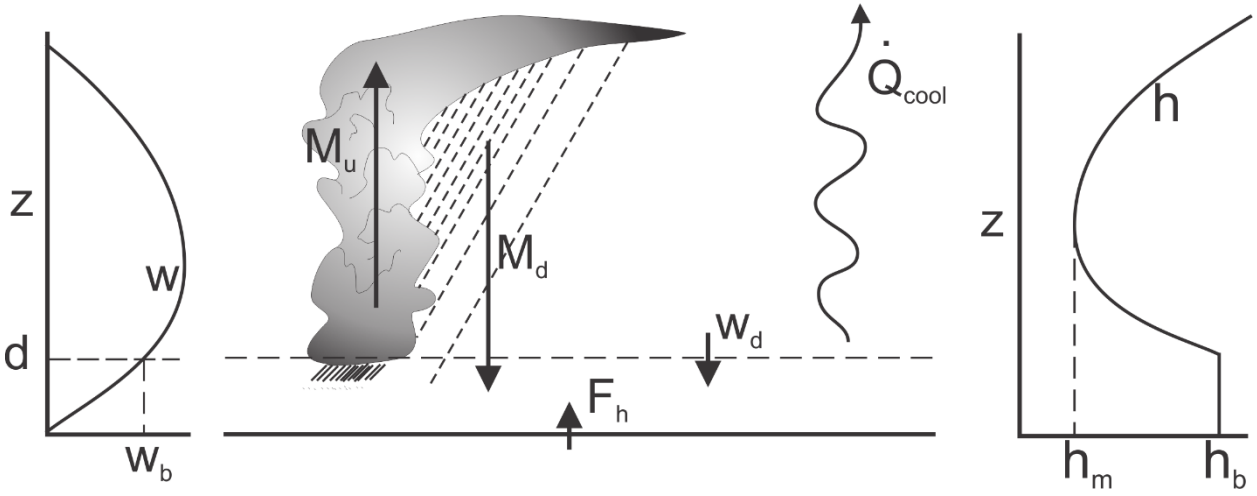


Figure 3.30: Sketch of the principles of boundary layer quasi-equilibrium. The subcloud layer has depth d and moist static energy h_b. A turbulent enthalpy flux F_h transfers enthalpy from the surface to the atmosphere and energy is removed from the troposphere by radiative cooling whose magnitude just above the top of the subcloud layer is $\dot{Q}_{cool}$. A large-scale vertical velocity w has a magnitude w_b just above the top of the subcloud layer. Convective updrafts transport mass upward at a rate M_u while convective downdrafts transport mass downward at a rate M_d and subsidence in the clear air between clouds transports mass into the boundary layer at a velocity w_d. The convective and clear sky downdrafts are assumed to transport moist static energy with a value of h_m into the subcloud layer.

The principles of boundary layer quasi-equilibrium (BLQE) are illustrated in Fig. 3.30. The subcloud layer has depth d and moist static energy h_b . A turbulent enthalpy flux F_h transfers

enthalpy from the surface to the atmosphere and energy is removed from the troposphere by radiative cooling whose magnitude just above the top of the subcloud layer is $\dot{Q}_{cool}$. A large-scale vertical velocity w has a magnitude w_b just above the top of the subcloud layer. Convective updrafts transport mass upward at a rate M_u while convective downdrafts transport mass downward at a rate M_d and subsidence in the clear air between clouds transports mass into the boundary layer at a velocity w_d . The convective and clear sky downdrafts are assumed to transport moist static energy with a value of h_m into the subcloud layer. Note that, as before, we define $\dot{Q}_{cool}$ to be positive when there is radiative cooling, and M_d and w_d to be positive downward.

In regions of deep convection, conservation of energy in the subcloud layer can be written

$$d\frac{\partial h_b}{\partial t} = d\mathbf{V}_h \bullet \nabla h_b + \alpha F_h - \left(\alpha M_d + (1-\sigma) w_d\right)\left(h_b - h_m\right) - \alpha \dot{Q}_{cool} d, \tag{3.50}$$

where α is the specific volume, σ is the fractional area covered by deep convection $\mathbf{V}_h$ is the large-scale horizontal velocity in the subcloud layer (assumed constant with height within the layer, as is h_b itself), and F_h is the surface enthalpy flux.

At the same time, conservation of mass in the subcloud layer dictates that

$$\alpha\left(M_u - M_d\right) - (1-\sigma) w_d = w_b \tag{3.51}$$

Eliminating the quantity $\alpha M_d + (1-\sigma) w_d$ between (3.50) and (3.51) yields

$$d\frac{\partial h_b}{\partial t} = d\mathbf{V}_h \bullet \nabla h_b + \alpha F_h - \left(\alpha M_u - w_b\right)\left(h_b - h_m\right) - \alpha \dot{Q}_{cool} d. \tag{3.52}$$

The BLQE approximation is just the limit of taking d to be small in (3.52). This can be regarded as a constraint on the convective updraft mass flux:

$$M_u = \rho w_b + \frac{F_h}{h_b - h_m}. \tag{3.53}$$

Note that for positive surface fluxes, $M_u > \rho w_b$, since we take $h_m < h_b$.

Just how good is the BLQE approximation? We can get some feeling for that by comparing the magnitude of the time tendency term in (3.52) with one of the other terms on the right side. For example, for constant surface fluxes in RCE $\left(w_b = 0\right)$, perturbing the boundary layer moist static energy, h_b, in (3.52) gives

$$\rho d\frac{\partial h_b'}{\partial t} = -\overline{M_u} h_b', \tag{3.54}$$

where $\overline{M_u}$ is the mean convective updraft mass flux in RCE, which from (3.24) is given by $\overline{M_u} = \dot{Q}_{cool} / \epsilon_p \mathbf{S}$. It is then evident from (3.54) that the relaxation time scale is given by

$$\tau \sim \frac{\epsilon_p \rho \mathbf{S} d}{\dot{Q}_{cool}}. \tag{3.55}$$

For a subcloud layer depth of 1 km, $\epsilon_p = 0.5,$ a temperature lapse rate of $6K\ km^{-1},$ and a radiative cooling of $2K\ day^{-1},$ this works out to be about 1 day. So this would not be a good approximation for dealing with the response of a deep convecting system to the diurnal variation of solar radiation, a point we shall return to shortly. But for steady responses to imposed steady forcing, or forcing varying on time scales more than a few days, (3.53) should work well.

The WTG approximation states that above the boundary layer, the density temperature is constant in horizontal space and time. In the simple model we are building here, this is approximated by constant dry static energy above the boundary layer. Since there are no horizontal gradients, there is no horizontal advection and we can write

$$\frac{\partial s_d}{\partial t} = 0 = -w_d \frac{ds_d}{dz} - \alpha \dot{Q}_{cool} = -w_d \mathbf{S} - \alpha \dot{Q}_{cool}, \tag{3.56}$$

where $s_d \cong c_p T + gz$ is the dry static energy and $\mathbf{S} \equiv ds_d / dz$ is the dry static stability just above the top of the subcloud layer and is also constant under WTG.

Thus from (3.56) and (3.51) we have

$$\rho w_d \mathbf{S} \cong \left(M_u - M_d - \rho w_b \right) \mathbf{S} = \dot{Q}_{cool}, \tag{3.57}$$

where we have made the approximation $\sigma << 1;$ alternatively, one could replace $\mathbf{S}$ by $(1-\sigma)\mathbf{S}.$

The downdraft mass flux, $M_d,$ is driven by evaporation of falling precipitation. We represent it here as proportional to the updraft mass flux via (3.23), which we repeat here:

$$M_d = \left(1 - \epsilon_p\right) M_u, \tag{3.58}$$

where ϵ_p is a precipitation efficiency, defined so that $0 \le \epsilon_p < 1.$ When $\epsilon_p = 1,$ all the precipitation that forms reaches the ground and none re-evaporates, so there can be no deep downdrafts. At the other extreme, when $\epsilon_p = 0,$ there is no net latent heat release and the downdraft must be as strong as the updraft. Equation (3.58) just linearly interpolates between these two physical limits.

If we use this formulation for the convective downdrafts in (3.57), then that equation and (3.53) constitute a pair of equations for the updraft mass flux M_u and the large-scale vertical velocity w_b that can be solved together for these two quantities, resulting in

$$\rho w_b = \frac{1}{1 - \epsilon_p} \left[\frac{\epsilon_p F_h}{h_b - h_m} - \frac{\dot{Q}_{cool}}{\mathbf{S}} \right], \tag{3.59}$$

and

$$M_u = \frac{1}{1-\epsilon_p}\left[\frac{F_h}{h_b - h_m} - \frac{\dot{Q}_{cool}}{\mathbf{S}}\right]. \tag{3.60}$$

These should be valid for sufficiently slow variations of surface fluxes and radiative cooling so long as the updraft mass flux given by (3.60) turns out to be nonnegative. It should also be remarked that, according to WTG, the saturation moist static energy h^* of the free troposphere is a constant, and since we assume convective neutrality, $h_b \left(= h^*\right)$ must also be constant.

Remember that these solutions pertain to quantities just above the subcloud layer, and are calculated using surface fluxes, radiative cooling that is also evaluated just above the subcloud layer, and moist static energy h_m that represents some weighted average of the moist static energy just above the subcloud layer and that characterizing convective downdrafts. We interpret this here as a value broadly representative of the lower to middle troposphere. As we saw in our very simple model of RCE, that value is ultimately set by precipitation efficiency and static stability. Here we will develop a simplified rate equation for h_m by assuming it is proportional to a deep tropospheric mean moist static energy:

$$h_m \cong \langle h \rangle \equiv \frac{1}{p_s - p_t}\int_{p_t}^{p_s} h dp, \tag{3.61}$$

where p_s and p_t are the pressure at the surface and tropopause, respectively. Henceforth we will use angle brackets, as in (3.61), to denote a mass-weighted vertical average.

To derive a rate equation for $h_m \left(= \langle h \rangle\right),$ we begin with the general conservation equation for moist static energy:

$$\rho \frac{\partial h}{\partial t} = -\rho \mathbf{V}_h \bullet \nabla h - \rho w \frac{\partial h}{\partial z} - \frac{\partial F_{rad}}{\partial z} - \frac{\partial F_c}{\partial z}, \tag{3.62}$$

where F_{rad} and F_c are the radiative and turbulent fluxes. If we divide (3.62) through by the density ρ and make use of the hydrostatic equation, (3.62) becomes

$$\frac{\partial h}{\partial t} = -\mathbf{V}_h \bullet \nabla h + \rho g w \frac{\partial h}{\partial p} + g\frac{\partial F_{rad}}{\partial p} + g\frac{\partial F_c}{\partial p}. \tag{3.63}$$

Apply the angle bracket operator to this equation yields

$$\frac{\partial h_m}{\partial t} = -\langle \mathbf{V}_h \bullet \nabla h \rangle - G w_b - \langle \alpha \dot{Q}_{cool} \rangle + \frac{g}{p_s - p_t} F_h, \tag{3.64}$$

where

$$G \equiv \frac{1}{w_b}\left\langle -\rho g w \frac{\partial h}{\partial p} \right\rangle. \tag{3.65}$$

The quantity G is referred to as the *gross moist stability* (Neelin and Held, 1987). This quantity plays an important role in the physics of the tropical atmosphere, and its sign and magnitude are

determined by the joint profile of vertical velocity and moist static energy. Since the gradient of the latter changes sign within the deep troposphere, it is not immediate apparent what the sign of G is, and this quantity will at any rate vary in magnitude with conditions. For the remainder of this chapter we shall assume that G is positive, meaning that upward motion reduces the mean moist static energy.

In deriving (3.65) we have assumed that the convective flux of moist static energy vanishes at the tropopause.

The horizontal advection (first term on left of (3.64)) is not necessarily small. Since we assume no horizontal gradients of temperature above the boundary layer, this term consists of horizontal advection of moisture. (There may also be a contribution from temperature advection in the boundary layer, where WTG does not apply.) In the tropics, there is nearly always at least an eddy flux of water vapor and this can have a substantial effect on tropical physics through (3.64.) Nevertheless, for the present purpose of a simple conceptual framework, we neglect it in most of the following analysis.

If the equation for the updraft mass flux, (3.60), predicts a negative value, this indicates that deep convection does not occur and we must set $M_u = 0.$ The free tropospheric energy balance then shows that

$$\rho w_b = \rho w_d = -\frac{\dot{Q}_{cool}}{\mathbf{S}} \quad \text{if } M_u = 0. \tag{3.66}$$

The steady-state form of (3.52) must then be solved for the subcloud layer moist static energy, which is now decoupled from the saturation moist static energy of the troposphere.

$$d\mathbf{V}_h \bullet \nabla h_b + \alpha F_h + w_d \left(h_b - h_m \right) - \alpha \dot{Q}_{cool} d = 0, \tag{3.67}$$

with w_d given by (3.66). But to solve this, one needs the value of moist static energy in the free troposphere, h_m. Although (3.64) is still technically valid, evaluating the gross moist stability G in this case would be difficult as the distribution of moist static energy is no longer being controlled by deep moist convection. Referring to Fig. 3.29, the only source of water above the boundary layer in this case would be advection from nearby regions of deep convection. But we can place a firm lower bound on h_m : the value it would have if the concentration of water were zero:

$$\left(h_m \right)_{min} = h^* - L_v q^*, \tag{3.68}$$

and clearly this would reach a minimum value just above the top of the subcloud layer, where the q^* of the free troposphere is a maximum. The value of q^* would correspond to the temperature of the free troposphere in the regions of deep convection, according to WTG. If the surface fluxes and radiative cooling of the subcloud layer are known, then (3.68) can be used with (3.67) to solve for the subcloud layer moist static energy.

To summarize, our simple system consists of (3.59), (3.60) and (3.64), if M_u turns out to be positive, and (3.66)-(3.68) otherwise. This assumes that the surface enthalpy flux, radiative cooling, and precipitation efficiency can be specified or determined as functions of the other variables.

It is worth nothing that the only time dependence in this system comes in through the deep tropospheric moist static energy budget, (3.64). Even for a fixed ocean temperature, this is a long time scale, equivalent to the RCE atmospheric relaxation scale given by (3.44) – around 20 days. This is well more than the boundary layer relaxation time scale of around 1 day given by (3.55). We will see just how important this time scale is later in this chapter.

So we return to the problem illustrated in Figures 3.27-3.29: The response of a horizontally infinite RCE atmosphere to a localized patch of anomalous surface enthalpy flux. To keep things simple, we hold the radiative cooling, gross moist stability, and precipitation efficiency constant and ignore horizontal advection in (3.64) and (3.67). We seek steady solutions and thus set the time derivative term in (3.64) to zero; this together with (3.59) and (3.60) constitute our system.

There is a circumstance in which such a solution is not viable. According to the steady form of (3.64), the solution for w_b will always be finite for finite radiative cooling and surface flux, but (3.59) indicates that w_b blows up in the limit of unitary precipitation efficiency $\left(\epsilon_p = 1\right)$ unless the quantity in brackets on the right side of (3.59) vanishes too. But this cannot happen for sufficiently large surface flux because h_b is fixed and there is a definite lower bound to the value of h_m. Thus, under this circumstance, there is an upper bound to the surface enthalpy flux, F_h, for the solution to be viable.

To see which circumstances lead to a bound on the surface flux, we can look at the solution for the column relative humidity (CRH) in the limit of large surface fluxes. Here we define CRH as

$$CRH \equiv 1 - \frac{h^* - h_m}{\mathbf{S}H}. \tag{3.69}$$

With this definition, when the atmosphere is saturated $\left(h^* = h_m\right)$, $CRH = 1,$ and when the moist static energy of the troposphere has a value corresponding to vanishing specific humidity, $CRH = 0.$ Solving for $h^* - h_m$ using (3.59) and the steady form of (3.64) in the limit of large surface fluxes gives

$$\lim_{F_h \to \infty} \left(CRH\right) = 1 - \frac{G}{\mathbf{S}} \frac{\epsilon_p}{1 - \epsilon_p}. \tag{3.70}$$

The fact that this limiting value of column relative humidity should not be less than zero restricts the last term in (3.70) to be less than 1. This will not be possible for value of ϵ_p sufficiently close to 1.

In reality, it is unlikely that the last term on the right side of (3.70) will exceed unity. Precipitation efficiency is likely to be large only in a sufficiently humid atmosphere, for which the value of G will be small (G vanishes in a saturated atmosphere with a moist adiabatic lapse rate). In fact, we might parameterize G as equal to $\mathbf{S}\left(1 - \epsilon_p\right)$.

But (3.70) does show that *CRH* remains less than 1 even in the limit of large surface fluxes as long as G and ϵ_p are positive.

With that said, Fig. 3.31 shows solutions to the system for $\epsilon_p = 0.5,\ \alpha\dot{Q}_{cool}$ equivalent to $1\,K\,day^{-1},\ p_s - p_t = 800\,hPa,\ \mathbf{S} = 3.5\,ms^{-2}$ and $G = 0.2\mathbf{S}$. For these values of the parameters, the RCE surface enthalpy flux is $94\,W\,m^{-2}$.

Deep convective mass fluxes vanish when the surface enthalpy flux falls below about $80\,W\,m^{-2}$, and the vertical velocity in the absence of convection has a magnitude of around $3\,mm\,s^{-1}$. Otherwise, both the mass fluxes and vertical velocity increase linearly with the surface enthalpy flux. The degree of subsaturation, $h^* - h_m$, has its maximum value in the region devoid of deep convection, and this value corresponds (by assumption) to the complete absence of water vapor. Otherwise, the subsaturation diminishes with increasing enthalpy flux and therefore with increase convective mass flux, as expected. The column relative humidity is zero, by assumption, in the absence of dep convection and increases with enthalpy flux, approaching the asymptotic value given by (3.70) at large forcing.

Precipitation increases rapidly with column humidity (Fig. 3.31d), with the slope asymptoting to infinity at large forcing, because of the bounded value of column humidity. A similar relationship is observed in nature (Fig 3.32). This observed relationship is widely interpreted as indicating a direct sensitivity of convective plumes to water vapor, but the model we are using here has no direct dependence of convection on free tropospheric water vapor. Thus the data in Fig 3.32 should not be interpreted as necessarily implying any such dependence. Nevertheless, it is undoubtedly the case that moist convection is sensitive to moisture, and in our simple model, this would enter through a dependence of the precipitation efficiency $\left(\epsilon_p\right)$ on humidity. Allowing ϵ_p to increase with humidity would yield an even greater sensitivity of precipitation to column humidity, according to (3.59) and (3.60).

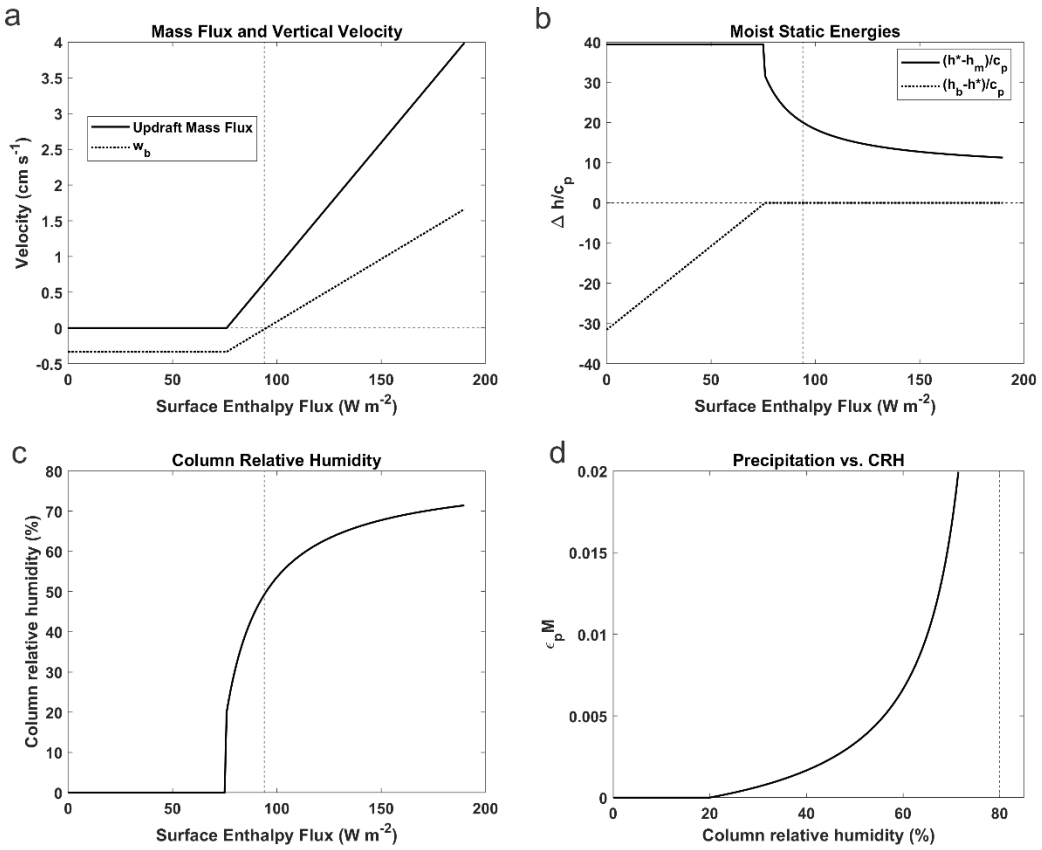


Figure 3.31: Response of a horizontally infinite RCE atmosphere to a local patch of anomalous surface enthalpy flux. The large-scale vertical velocity and convective updraft mass flux (normalized by air density) are shown in a). The differences between the actual and saturation moist static energies of the free troposphere and between the boundary layer moist static energy and the saturation moist static energy of the free troposphere, both normalized by heat capacity, are shown in b). The column relative humidity is shown in c) and a proxy for precipitation (updraft mass flux multiplied by precipitation efficiency) is graphed against column humidity in d). The RCE value of the surface enthalpy flux in indicated in a)-c) by the vertical dashed line. In d) the vertical dashed line indicates the asymptotic value of column relative humidity as the surface enthalpy flux gets very large. See text for values of the parameters used here.

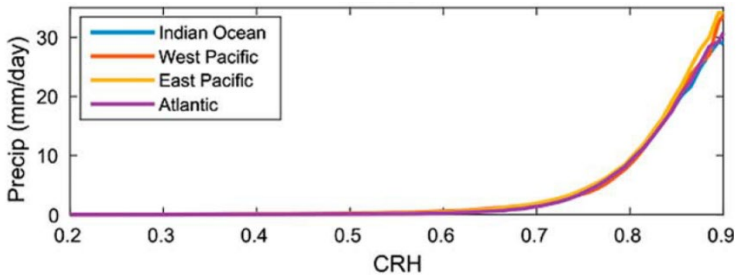


Figure 3.32: Observed mean precipitation in 1% bins of column relative humidity for tropical ocean basins using version 7 SSM/I satellite-derived products over the period 1998–2001.

We can also examine the response of RCE to local perturbations by modifying the single-column model described earlier in this chapter. This follows the concept described in detail in Sobel and Bretherton (2000). We first run the single-column model into a state of RCE, as usual. We then restart the model from its RCE state but do not permit the temperature above a specified pressure level to change, and specify a higher or lower sea surface temperature than the RCE state. At each time step, the sum of the convective and radiative heating is calculated at each level. To maintain energy balance at each level, a vertical velocity is calculated such that the adiabatic cooling or warming associated with that velocity is just sufficient to offset the sum of the convective and radiative heating. That vertical velocity is then used to advect water vapor in the vertical direction. As with the RCE calculation, the integration is run until the variables reach statistically stable values. (This feature is built into the model available at https://zenodo.org/doi/10.5281/zenodo.10795221.)

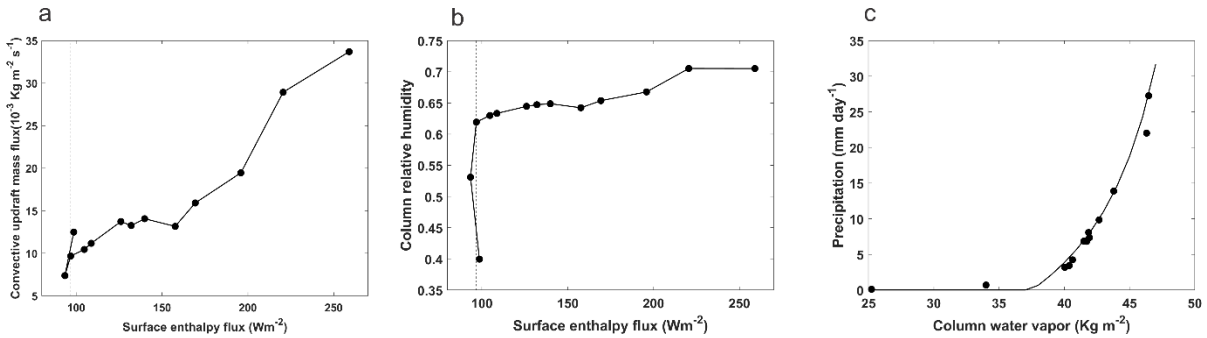


Figure 3.33: Response of the single-column model run with temperature fixed at its RCE value and vertical motion calculated to enforce local thermodynamic balance. Sea surface temperature is varied from -0.5 K to 3.5 K relative to its RCE value. a) shows the convective updraft mass flux as a function of the surface enthalpy flux. The variation of column relative humidity with surface flux is shown in b) and the variation of precipitation with column humidity is shown in c). The curve in c) is a fit of the toy model to the data. The RCE value of the surface enthalpy flux is indicated by the vertical dashed line in a) and b).

An example of the response of the single-column model to sea surface temperature anomalies ranging from -0.5 K to 3.5 K, relative to its RCE value, is shown in Fig. 3.33. When the SST anomaly is -0.5 K only small trade cumuli are present and the net vertical motion is downward. Unlike in the analytical model just described, the column does not completely dry out, because the single-column model assumes that convergence into the column is from an adjacent column in RCE and carries with it the water vapor content of the RCE column.

When the SST anomaly is -0.25 K, the dominant mode of convection is cumulus congestus that extend up to 600 hPa and which do weakly precipitate.

As expected, the net vertical motion is upward for all positive SST anomalies and the convective mass flux increases nearly monotonically with surface fluxes (Fig. 3.33a). However, the column moistens only very slowly (Fig. 3.33b). Precipitation climbs steeply with column humidity (Fig. 3.33c) and a fit of the analytical model (using constants that best fit the single-column model output) works well.

Surprisingly, upward and downward motion may arise spontaneously in the single-column model run in WTG mode even when no surface temperature anomaly is imposed, suggesting a fundamental instability of the RCE state. This is examined in the next section. But before

proceeding, we note two important limitations of applying and/or interpreting tropical atmospheric physics using this framework.

The first is a caution against regarding either the surface enthalpy flux or the SST as legitimate externally specifiable parameters. This works on time scales short compared to the adjustment time scale of the ocean surface mixed layer (given by (3.45), around 70 days for typical mixed layer depths), but for longer time scales the ocean-atmosphere system must be regarded as coupled. In the absence of enthalpy advection within the ocean, the surface enthalpy flux must equal the net (shortwave plus longwave) surface radiative flux. In the tropics, where the high infrared opacity of the lower atmosphere prevents much net longwave flux, solar flux will be the main influence on the net surface radiative flux.

Married with the WTG framework discussed in this section, this coupling to radiation can give rise to interesting phenomena, such as climate hysteresis. Referring to Figure (3.29), if a patch of seawater of dimensions less than the local radius of deformation were to become cool enough, not only would deep convection cease, but stratocumulus clouds would form near the top of the boundary layer (see Chapter 4). This would greatly reduce the solar flux and, owing to the requirement of energy balance in the ocean mixed layer, thereby reduce the surface turbulent enthalpy flux. This would reinforce the cold SST anomaly, possibly leading to two stable solutions: One with no SST anomaly and one with low SST and stratocumuli. (But note that there would be a dynamical response to the cold anomaly within the ocean itself, possibly mitigating this effect.)

The second caution has already been pointed out: The framework only works if the subcloud layer can be assumed to be in energy balance. As the relaxation time scale for subcloud layer moist entropy is about 1 day (see (3.55) and subsequent discussion), the diurnal cycle is too fast for BLQE to apply. Nevertheless, we can learn something about the diurnal cycle by gently relaxing BLQE and applying a periodic cycle of surface fluxes driven by the diurnal cycle of insolation. That is, instead of rigorously enforcing BLQE we relax the actual convective updraft mass flux, M_u, back towards it equilibrium value according to

$$\frac{\partial M_u}{\partial t} = \frac{M_e - M_u}{\tau}, \tag{3.71}$$

where

$$M_e \equiv \rho w_b + \frac{F_h}{h_b - h_m}. \tag{3.72}$$

is the BLQE mass fluxes and τ is a boundary layer response time scale given by (3.55). This just relaxes the convective mass flux towards its equilibrium value over the time scale τ. We assume that the land area has a length scale much less than the deformation radius, so that WTG applies, so that, combining (3.57) with (3.58) we have

$$\epsilon_p M_u = \rho w_b + \frac{\dot{Q}_{cool}}{\mathbf{S}}. \tag{3.73}$$

We can eliminate ρw_b between (3.72) and (3.73) and write (3.71) as

$$\tau \frac{\partial M_u}{\partial t} + \left(1-\epsilon_p\right) M_u = \frac{F_h}{h_b - h_m} - \frac{\dot{Q}_{cool}}{\mathbf{S}}. \tag{3.74}$$

To get an analytic solution, we will make a few simplifying assumptions. First, since the time scale for adjustments of the free tropospheric moist static energy, h_m, is on the order of 20 days, we will assume that it remains constant over a diurnal cycle. Second, we assume that the radiative cooling rate, $\dot{Q}_{cool}$, is constant, and the static stability $\mathbf{S}$ is anyway constant under WTG. Finally, we will specify the surface enthalpy flux as a constant needed to balance the radiative cooling plus a specified small perturbation varying on a diurnal time scale and likewise assume that M_u and h_b consist of a mean (RCE) part plus a small perturbation. Then linearizing (3.74) gives

$$\tau \frac{\partial M_u'}{\partial t} + \left(1-\epsilon_p\right) M_u' = \frac{F_h' - h_b' \left(\dot{Q}_{cool} / \epsilon_p \mathbf{S}\right)}{\overline{h_b} - \overline{h_m}}, \tag{3.75}$$

where the primes denote the perturbation quantities, the overbars denote RCE values, and we have made use of $\overline{F_h} = \overline{M_u}\left(\overline{h_b} - \overline{h_m}\right)$ from (3.72) and $\overline{M_u} = \dfrac{\dot{Q}_{cool}}{\epsilon_p \mathbf{S}}$ from (3.73).

Likewise, we linearize (3.52) about the RCE state, ignoring horizontal advection and radiative cooling in the boundary layer, resulting in

$$\rho d \frac{\partial h_b'}{\partial t} + \frac{\dot{Q}_{cool}}{\epsilon_p \mathbf{S}} h_b' = F_h' - \left(\overline{h_b} - \overline{h_m}\right)\left(1-\epsilon_p\right) M_u'. \tag{3.76}$$

Thus (3.75) and (3.76) are a coupled linear system for the perturbation convective updraft mass flux and boundary layer moist static energy, given specified diurnal surface enthalpy flux F_h'. This is only strictly valid for a diurnal cycle that is a small perturbation to the RCE state, which is not usually the case, but we might learn something in spite of this restriction.

Notice first that if we multiply (3.75) by $\overline{h_b} - \overline{h_m}$ and subtract (3.76) from the result, making use of (3.55) for τ, we get

$$h_b' = \frac{\epsilon_p \mathbf{S}}{\dot{Q}_{cool}} \left(\overline{h}_b - \overline{h}_m\right) M_u'. \tag{3.77}$$

Thus the convective mass flux tracks the boundary layer moist static energy. Loosely speaking, the convection peaks during the hottest part of the day.

If we specify a simple, periodic forcing that peaks at noon local time $\left(t = 0\right)$:

$$F_h' = F_0 \cos(\omega t), \tag{3.78}$$

where ω is the diurnal frequency $2\pi/(1\,day)$, the solution of (3.75) and (3.76) is

$$M_u' = \frac{F_0}{\left(2-\epsilon_p\right)\left(\overline{h_b}-\overline{h_m}\right)}\cos(\omega t-\theta), \tag{3.79}$$

where θ is the phase of the response, given by

$$\tan(\theta) = \frac{\omega\tau}{2-\epsilon_p}, \tag{3.80}$$

with τ given by (3.55). The solution for the boundary layer moist static energy is then given by (3.77).

According to (3.79), the diurnal cycle of moist convection will be stronger when the free troposphere is moist and when there is large precipitation efficiency. The convective available potential energy (CAPE) scales with h_b', given by (3.77) and (3.79), will also be larger in a more humid atmosphere with larger precipitation efficiency and in an atmosphere with large dry static stability and smaller radiative cooling rate. Given a typical value of the boundary layer relaxation time scale τ of about a day, the phase lag given by (3.80) is around 5 hours, depending also on the precipitation efficiency.

From this line of reasoning, we expect strong diurnal cycles of convection in moist atmospheres over land areas whose dimensions are small compared to the local radius of deformation but not so small that horizontal advection in the boundary layer damps the diurnal response of the land by importing cooler air from over the sea. (A sea breeze of $5\,ms^{-1}$ would cross a peninsula of 100 km width in about 5 hours). We expect high CAPE (large h_b') to be associated with lightning, and a glance at the climatological lightning map (Figure 3.9) shows large activity in places like Florida and Cuba (dimensions small compared to the local deformation radius but not much smaller than 100 km or so) and equatorial locations like central Africa.

One takeaway from this is that the diurnal cycle, being too fast for BLQE, allows CAPE to build up and then be released, giving strong convection late in the afternoon. But for the remainder of this book we will concern ourselves mostly with phenomena on time scales much longer than a day for which convective quasi-equilibrium is a good approximation. We next turn to the interesting question of whether RCE states (without diurnal cycles) are actually stable.

3.4 Stability of RCE and the aggregation of tropical convection

Throughout this chapter we have emphasized the value of treating RCE as a starting point for understanding the tropical atmosphere. But is the RCE state stable? Given that RCE is a statistical equilibrium, we can ask whether the statistics remain invariant in time and space or whether, to the contrary, perturbations to that state can amplify and transform it into a different state with large horizontal and/or time variations in the convective and radiative statistics.

It turns out that under some circumstances, yet to be fully delineated, statistically homogenous RCE states break down spontaneously into clusters whose horizontal scales are many times the typical spacing between cumulus clouds in RCE. With a large enough Coriolis parameter, these clusters take the form of tropical cyclones. *Self-aggregation* of moist convection, as the phenomenon has come to be known, was first discovered in non-rotating cloud-permitting numerical simulations of RCE, first in two dimensions (Held et al., 1993) and a few years later in three dimensions (Tompkins and Craig, 1998). A good review of self-aggregation in numerical models is provided by Wing et al. (2017).

Figure 3.34 shows what non-rotating self-aggregation looks like in a cloud-permitting numerical simulation. This simulation was performed with using a cloud-permitting model, the System for Atmospheric Modeling (SAM; Khairoutdinov and Randall, 2003), run with a horizontal grid spacing of 3 km and a stretched vertical grid, with periodic boundary conditions in both horizontal dimensions. A uniform sea surface temperature (SST) of 305 K was imposed, and a full radiation scheme was used but with time-invariant insolation. The model state was initialized with vertical profiles from a previous integration of the model in a domain too small to permit self-aggregation. (For details see Wing and Emanuel (2014).)

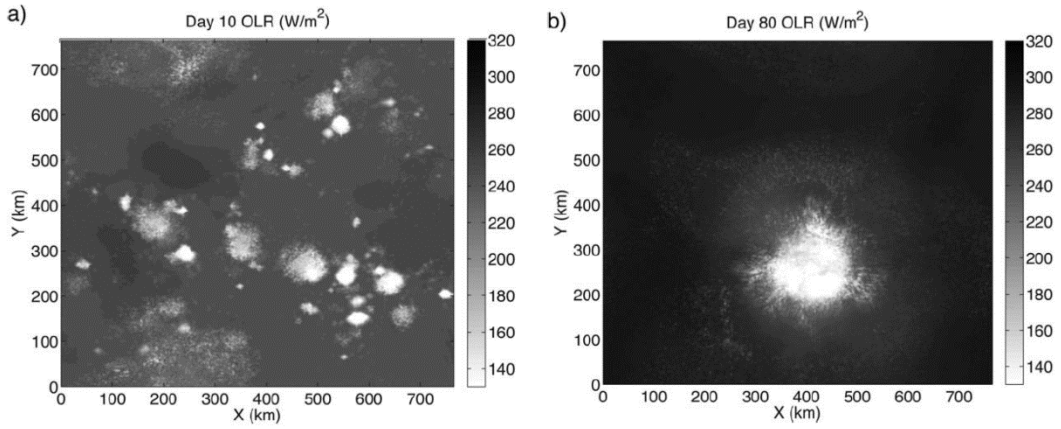


Figure 3.34: Snapshots of outgoing longwave radiation (OLR) in a cloud-permitting simulation of moist convection over an ocean surface of constant temperature at 10 days into the simulation (a) and at 80 days (b). From Wing and Emanuel (2014).

Initially, the convection is fairly randomly distributed in space and time (Figure 3.34a). If the numerical domain is large enough and certain other conditions are satisfied, the convection begins to clump together, and over a period of a few tens of days collapses to a single cluster, as shown in Figure 3.34b. The cluster typically occupied a few tens of percent of the domain, independent of domain size if the domain is square and not too large. Virtually all the rain occurs within the cluster, and the air outside the cluster is extremely dry above the boundary layer, as shown in Figure 3.35 taken from a similar, previously published simulation (Bretherton et al., 2005).

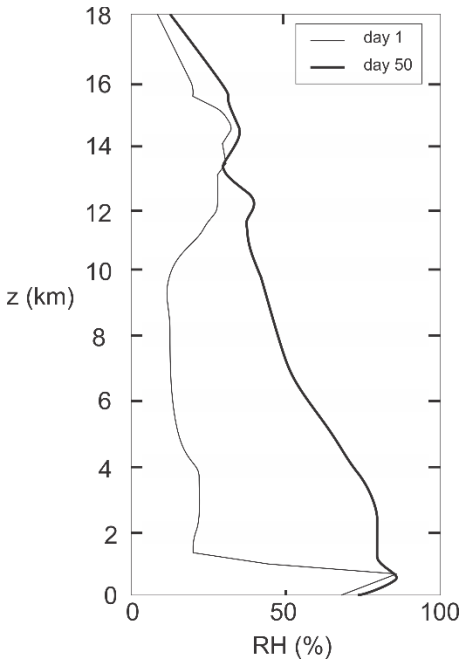


Figure 3.35: Vertical profile of horizontally domain averaged relative humidity at 1 day (thin) and 50 days (thick) into an integration of a cloud-permitting model (Bretherton et al., 2005).

Although the cluster shown here is roughly circular, in other simulations the aggregation takes the form of wavy lines ((e.g. Tompkins and Craig, 1998; Wing and Cronin, 2016). As of this writing, there is no generally accepted theory for the scale of clusters in non-rotating aggregated convection. In the case of rotating self-aggregation, there are well defined scales and these will be discussed in Chapter 14 on tropical cyclogenesis.

As mentioned earlier in this chapter, clusters of deep convective clouds are often observed, both within and outside the tropics (see, e.g., Figure 3.7), but it is difficult to determine whether they arise spontaneously or are associated with pre-existing disturbances.

Why should deep moist convection spontaneously aggregate? To get some insight into this important question, we first note again that in the aggregated state the column-integrated moisture content outside the cluster is relatively small (see Figure 3.35) whereas within the cluster is nearly saturated (not shown). Thus the horizontal variance of the column-integrated moisture is larger than that of the non-aggregated state, where the humidity outside of clouds is much larger.

Given that temperature perturbations in non-rotating convective states are very small, these large moisture excursions will strongly dominate horizontal fluctuations of the moist static energy. Therefore, the horizontal variance of column moist static energy is larger in the aggregated than in the disaggregated state. Column moist static energy has the very useful property that it is *not directly affected by convection*, which is an energy transfer process and not a source or sink of moist static energy. Thus the only way to change the variance of column moist static energy is by the external energy sources of radiative heating and turbulent surface fluxes, and also by large-scale advection. To see this, we begin, following a development closely parallel to (3.61)-(3.65), by writing the equation for conservation of moist static energy in pressure coordinates, assuming that the cluster-scale flow is hydrostatic:

$$\frac{\partial h}{\partial t} = g\frac{\partial F_r}{\partial p} + g\frac{\partial F_t}{\partial p} - \omega\frac{\partial h}{\partial p} - \mathbf{V_h}\bullet\nabla_h h, \tag{3.81}$$

where h is the moist static energy, F_r and F_t are respectively the upward radiative and turbulent fluxes (both assumed to be in the vertical direction only), ω is the total time derivative of pressure, and $\mathbf{V_h}$ is the horizontal velocity vector.

We next define a column integral operator acting on any scalar a:

$$\tilde{a} \equiv \frac{1}{g}\int_{p_{tr}}^{p_0} a dp', \tag{3.82}$$

where p_0 is the surface pressure and p_{tr} is the pressure at the tropopause. We assume that the large-scale vertical pressure velocity ω and the turbulent fluxes F_t both vanish at the tropopause and apply the integral operator (3.82) to (3.31) to arrive at

$$\frac{\partial \hat{h}}{\partial t} = F_{r0} - F_{rt} + F_{t0} - \widetilde{\omega\frac{\partial h}{\partial p}} - \widetilde{\mathbf{V_h}\bullet\nabla_h h}, \tag{3.83}$$

where F_{r0} and F_{rt} are the upward radiative fluxes at the surface and tropopause, respectively, and F_{t0} is the surface turbulent enthalpy flux (equivalent to the moist static energy flux). The last two terms in (3.83) represents the advective export of moist static energy from the column. Since horizontal gradients are often (but not always!) weak in the tropics, we usually neglect the last term in (3.83). We recognize G as a slightly different form of the gross moist stability defined by (3.65):

$$G_i \equiv -\frac{1}{\tilde{\omega}}\widetilde{\omega\frac{\partial h}{\partial p}} = (p_0 - p_{tr})G, \tag{3.84}$$

where G was defined by (3.65). The gross moist stability is an $\omega-$ weighted vertical integral of the vertical gradient of moist static energy. Since h almost always has a minimum value somewhere in the middle troposphere, the sign of G_i is not *a priori* obvious. It can be, in general, a function of horizontal space and time.

In the long-time average RCE state, there is no time tendency or large-scale advection of the column moist static energy, so we have the classical balance $F_{r0} - F_{rt} + F_{t0} = 0.$ We consider fluctuations around this equilibrium state and denote these by primes. With these definitions, neglecting the last term in (3.83), multiplying (3.83) through by $\tilde{h}'$, and averaging over area, yields an equation for the evolution of the variance of the column moist static energy:

$$\frac{1}{2}\frac{\partial\overline{\tilde{h}'^2}}{\partial t} = \overline{\tilde{h}'F_{r0}'} - \overline{\tilde{h}'F_{rt}'} + \overline{\tilde{h}'F_{t0}'} + \overline{G_i\tilde{\omega}'\tilde{h}'}, \tag{3.85}$$

where for this purpose we have neglected time variations of G_i, and the overbars represent horizontal averages over distances that span many clouds.

Self-aggregation of convection is observed (in models) to be accompanied by a strong increase in the variance of column moist static energy, so during aggregation the left side of (3.85) will be positive. Therefore, the right side must also be positive, and this requires fluctuations of the column moist static energy to be positively correlated with one or more of the following:

1. Fluctuations of the net radiative heating of the column, as given by the sum of the first two terms on the right side of (3.85),
2. Fluctuations of the surface enthalpy flux,
3. Fluctuations of $\widetilde{\omega}$, if G_i is positive.

That is, perturbations to the RCE state may amplify if net tropospheric radiative heating increases with column moisture (positive $\tilde{h}'$) or if there are greater surface fluxes where $\tilde{h}'$ is elevated (and/or less surface flux where $\tilde{h}'$ is depressed). The last term in (3.85) would be positive if there were a net downward motion (positive $\widetilde{\omega}$) where the column is anomalously moist and/or upward motion where it is dry. This is not usually the case in the kind of aggregation illustrated in Figure 3.34, where the moist columns are ascending and the dry air is descending, but can transiently make a positive contribution. Thus in this case, and in most cases examined to date, the last term in (3.85) is usually a damping term and acts against the aggregation. But there may be places and times where the gross moist stability G_i is negative, and this would indicate a direct feedback between moistening by convection and large-scale ascent.

An important result of this analysis is that the perturbations to column moist static energy may grow only if there are interactions involving radiation and/or surface enthalpy fluxes; barring negative G_i, direct interactions between convection and larger-scale circulations *per se* are stable and cannot cause circulation systems to amplify, at least as measured by the variance of column moist static energy. Indeed, most of the studies undertaken with three-dimensional cloud-permitting models (e.g. those reviewed by Wing et al. (2017)) show no self-aggregation when both surface fluxes and radiative heating are horizontally homogenized at each level.

It should be noted that this does not prohibit the growth of disturbances in non WTG-conditions in which column temperature and column water vapor perturbations may both grow but cancel out in their contributions to moist static energy.

Figure 3.36 shows the evolution over time of the values of terms like those on the right side of (3.85), averaged over the domain, for a simulation like that used to construct Figure 3.34. All the terms except for the last term of (3.85) have been calculated directly from the model output, but the last term (the advective term) is difficult to calculate from output that is sparse in time, so instead it has been estimated as a budget residual of (3.85).

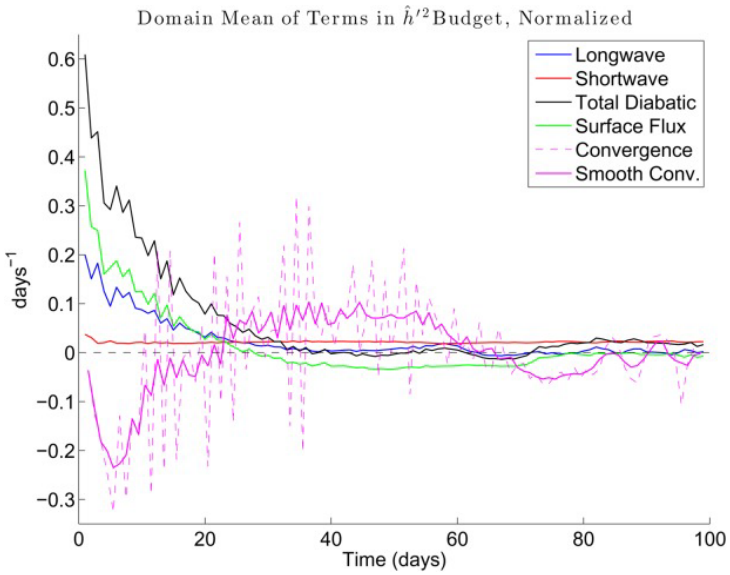


Figure 3.36: Domain averages of the budget terms in the column moist static energy variance equation (3.85), as a function of time from initialization with an RCE state. The first two terms on the right of (3.85) are combined into a radiative term and then split between shortwave (solar) and longwave (terrestrial) components, indicated respectively by the blue and red curves. The term involving surface enthalpy fluxes (third term on right of (3.85)) is shown in green, and the sum of all the diabatic terms is on black. The advective term (last term in (3.85)) is in magenta; it has been derived as a budget residual and the time-smoothed values are indicated by the solid magenta line.

Initially, all the budget terms except the advective term are positive, indicating positive feedbacks between the amplifying circulation and radiation and convection. Among the larger positive feedbacks is that from surface enthalpy fluxes. Column moist static energy is elevated in the emerging convective cluster, and downdrafts associated with the enhanced deep convection create gusts of wind at the surface, which enhance the surface fluxes. (Note that this cannot happen over land, as the land surface has essentially no heat capacity…the first and third terms on the right of (3.85) must cancel to insure no net heat exchange with the soil. In that case, the remaining diabatic term is approximately equal to fluctuations in the net top-of-the-atmosphere radiation.)

The longwave radiative term in of the same order of magnitude as the surface flux term, and more detailed analysis of the simulations reveals that this is dominated by water vapor perturbations early in the simulation. Most of this feedback is manifest in the simulations by developing dry patches emitting more infrared radiation from the surface and from boundary layer water vapor, leading to net radiative cooling of the column. Later in the simulations, longwave trapping by the thick, high anvil ice clouds associated with the deep convective cluster dominate the feedback.

There is some direct absorption of sunlight in the atmosphere, mostly by water vapor bands (see Figure 2.5). As there is more moisture in the convective cluster, there can be more absorption of sunlight there (unless there is too much additional reflection by the clouds) and less absorption in the dry patches, so this is usually a positive feedback, as evident in Figure 3.36. (This effect might be expected to be stronger at higher temperatures, owing to the greater concentrations of water vapor.)

But note that around day 30, the sign of the various contributions reverses, with the diabatic terms all putting a break on the increase in variance and the advective term now driving the increase.

Under what circumstances does aggregation occur, and what drives it? To get some insight into these questions, we first consider the effects of water vapor feedback all by itself, and put aside for the moment the important feedbacks between clouds and radiation and surface fluxes and convection.

We begin by inquiring about the response of radiation to drying of a patch of air in the RCE state. The reduction of water vapor, an important greenhouse gas, causes a decrease in both the absorption and emission of infrared radiation by the dry sample. But because, in RCE, most of the troposphere is warmer than it would be in radiative equilibrium (see Figure 3.19), reducing the emission is the larger effect and the sample warms. At the same time, since the sample is emitting less infrared radiation, there is less of it to absorb by greenhouses gases above and below the level of our sample. Thus those samples cool.

This effect is illustrated in Figure 3.37, which shows the response throughout the troposphere to the imposed drying at a particular level (abscissa) in RCE states simulated with the single-column mode (SCM) described earlier in this chapter and available through the course website.

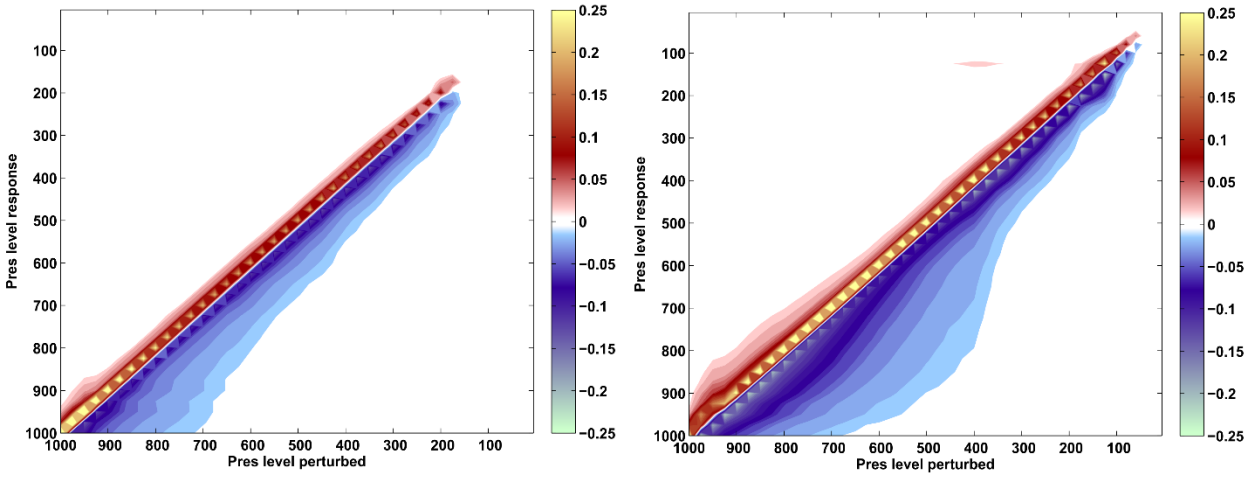


Figure 3.37: Perturbation heating rate as a function of pressure (ordinate) and the pressure level of a 20% negative perturbation in specific humidity (abscissa) to and RCE state, for SSTs of (left) 25°C and (right) 40°C. Plotted is the logarithm of 1 + the actual heating rate, for positive heating rates, and minus the logarithm of 1 minus the heating rate, for negative heating rates. Figure and caption from Emanuel et al. (2014).

The left side of the figure is for an RCE state with a fixed SST of 25°C while the right side is for 40°C. The diagonal red streak shows that drying the parcel reduces the radiative cooling of the parcel itself, while the blue shows that air below the dried parcel level is anomalously cooled owing to the reduced downward infrared flux from the dried sample. (Note that the contours are of the natural logarithm of the cooling or heating rate.) The air above the sample level experiences anomalous heating owing to the greater upward infrared flux from below the sample level. Both the warming and the cooling are larger in magnitude in the warmer atmosphere, again because of its larger infrared optical thickness.

It is also instructive to look at the response of the troposphere to an instantaneous reduction of specific humidity by 20% through the whole column, again starting from the RCE state as simulated by the SCM. This is shown in Figure 3.38.

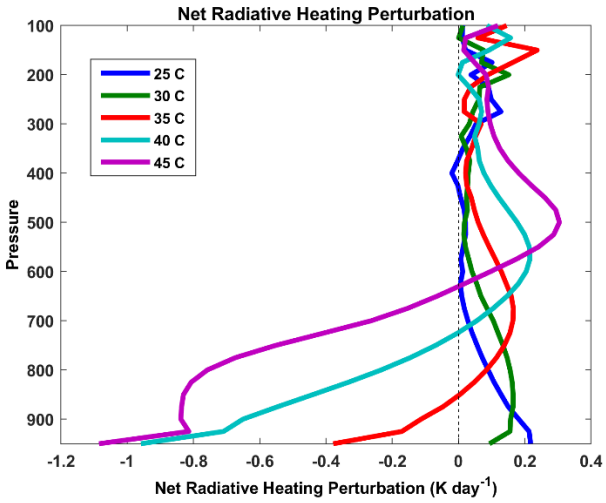


Figure 3.38: Perturbation net radiative heating rates in response to an instantaneous reduction of specific humidity of 20% through the whole troposphere, from the RCE states with SSTs ranging from 25 to 45°C. From Emanuel et al. (2014).

Note that for the lower SSTs, the column-integrated response to the drying is net heating, whereas this reverses for sufficiently high SSTs. At these higher SSTs, the negative response of column-heating to column drying is destabilizing, according to (3.85): negative values of $\tilde{h}'$ would be associated with negative values of the column radiative cooling (first two terms on the right of (3.85)). This is does not lead to an exact criterion for instability owing to water vapor feedback alone, as we do not know, *a priori*, what the vertical structure of the drying is.

At high temperatures, the lower troposphere becomes nearly opaque in the infrared, and local perturbations to the humidity have little effect. But drying of the middle and/or upper troposphere opens up an IR window and so the lower troposphere can radiative more heat to space; thus it cools. (See also Figure 3.37.)

We can do a more legitimate test for stability by using our SCM in WTG mode, as described in section 3.3c. After running the SCM into RCE, we continue the integration but fix the temperature above the boundary layer…in this case, above 850 hPa, and calculate at each model level the vertical velocity necessary to hold the temperature constant. That vertical velocity is then used to vertically advect water vapor. In section 3.3c we specified perturbations

to the RCE SST but here we hold the SST at its RCE value and add low amplitude noise. Details are described in Emanuel et al. (2014).

At relatively cool sea surface temperatures, the noise gradually dies away, but for SSTs above a critical value, somewhere in the range of 30-35 C, the single column transitions to one of two non-RCE equilibrium states: one with mean ascent and the other with mean descent. The vertical profiles of vertical velocity and relative humidity of these two states are compared in Figure 3.39, and the RCE relative humidity profile is also shown for comparison.

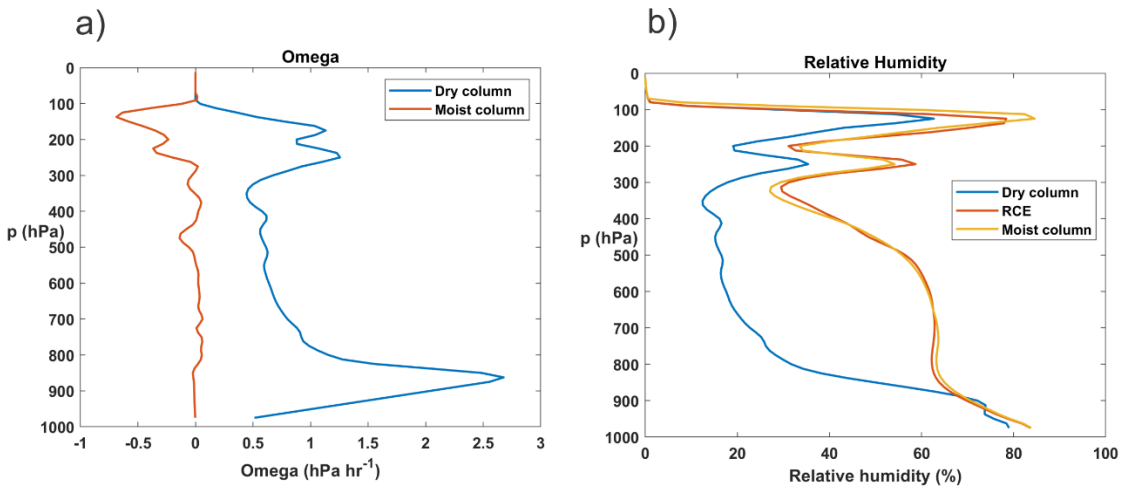


Figure 3.39: Vertical profiles of ω (a) and relative humidity (b) in the WTG single-column integrations starting from an RCE state with an RCE SST of 35.6 C. Blue curves show the subsiding column while red curves are for the ascending column. The RCE relative humidity profile is shown by the yellow curve in (b).

The asymmetry between the descending and ascending WTG solutions is striking, with much stronger departures from RCE in the descending column. Mass conservation would imply that the descending columns should occupy a small fractional area of the total domain, and this is indeed the case early in the self-aggregation process, at least in some cloud-permitting models (Wing et al., 2017). This has the effect of cooling the horizontal mean temperature above the boundary layer (essentially shifting the domain-mean RCE state itself), gradually favoring stronger ascent in the ascending columns, whose fractional area then shrinks. Eventually a balance is reached whereby the additional infrared emission in the dry regions balances the reduced emission from the convective clusters.

The net drying of the troposphere when aggregation occurs may represent an important stabilization mechanism for tropical climate. If some other forcing (e.g. addition of CO_2 to the atmosphere) causes a warming, and if that warming enhances aggregation, the consequent drying of the atmosphere would tend to damp the response to the original forcing.

It follows from the preceding that a necessary condition for the instability of RCE driven entirely by radiation is that perturbations of column water vapor and/or clouds give rise to column radiative heating perturbations of the same sign (from 3.85). However, that this is not a sufficient condition for instability can be demonstrated using the WTG framework.

We begin with an RCE state and assume that any self-aggregation that occurs evolves slowly enough that the WTG approximation is always valid. (Given the long time scale of aggregation relative to the time scales of internal waves, this should be an excellent approximation.) We

assume that the temperature above the boundary layer remains constant in time, and that unless deep convection vanishes in regions of downward motion, the sounding remains neutral so that $h_b = h^* = \text{constant}$. We further assume that the surface fluxes remain constant in space and time so as to focus on strictly radiative destabilization of the RCE state, and neglect any large-scale horizontal advection of moist static energy.

As a very crude representation of the feedback between radiation and clouds and water vapor, we specify the bulk radiative cooling of the system as a function of the convective updraft mass flux:

$$\dot{Q}_{cool} = Q_0 - \gamma\left(M_u - M_{eq}\right), \tag{3.86}$$

where Q_0 is the RCE state radiative cooling rate, M_{eq} is the RCE convective updraft mass flux, and γ is a constant parameter. If γ is positive, then the rate of radiative cooling diminishes as the convective updraft mass flux rises above its RCE value, under the assumption that more convection is associated with more high clouds and column moisture. Then (3.60) becomes

$$M_u = \frac{1}{1-\epsilon_p}\left[\frac{F_h}{h_b - h_m} - \frac{Q_0 - \gamma\left(M_u - M_{eq}\right)}{\mathbf{S}}\right], \tag{3.87}$$

or making use of the relation (from 3.60 and 3.59 with $w_b = 0$) that $M_{eq} = Q_0 / \epsilon_p \mathbf{S}$,

$$M_u = \frac{1}{1-\epsilon_p - \frac{\gamma}{\mathbf{S}}}\left[\frac{F_h}{h_b - h_m} - \frac{Q_0}{\mathbf{S}}\left(1 + \frac{\gamma}{\epsilon_p \mathbf{S}}\right)\right]. \tag{3.88}$$

The WTG equation for the aggregate-scale vertical velocity, 3.59, becomes

$$\rho w_b = \frac{1}{1-\epsilon_p}\left[\frac{\epsilon_p F_h}{h_b - h_m} - \frac{Q_0 - \gamma\left(M_u - \frac{Q_0}{\epsilon_p \mathbf{S}}\right)}{\mathbf{S}}\right]. \tag{3.89}$$

Finally, with the aforementioned approximation of neglecting horizontal advection, conservation of column-average moist static energy (3.64) takes the form

$$\rho \frac{dh_m}{dt} = -G\rho w_b + \gamma\left(M_u - \frac{Q_0}{\epsilon_p \mathbf{S}}\right), \tag{3.90}$$

where we have made use of the balance between surface fluxes and radiative cooling in RCE:

$$Q_0 = \frac{\rho g F_h}{p_s - p_t}.$$

The set (3.88)-(3.90) constitutes a close system for the variables M_u, ρw_b, and h_m, and all the parameters that appear $\left(G, \gamma, \epsilon_p, Q_0, \mathbf{S}\right)$ are considered constants. As long as deep convection does not vanish in the descending branch, the assumptions of moist adiabatic lapse rates and WTG imply that the boundary layer moist static energy h_b is constant too.

It is convenient to express this system as a single ordinary differential equation for column-average moist static energy, h_m, by eliminating M_u and ρw_b from (3.88)-(3.90). The notation is simplified by defining a new variable D that measures the saturation deficit:

$$D \equiv h_b - h_m,$$

remembering that h_b is constant, and let

$$D = D_0\left(1 + D'\right), \tag{3.91}$$

where D_0 is the RCE value of D and D' is its nondimensional departure from the RCE state. (From the RCE version of (3.59), (3.60) and (3.64), $D_0 = \dfrac{\epsilon_p \mathbf{S}\left(p_s - p_t\right)}{\rho g}$.) With these notational simplifications, (3.90) becomes, after substitutions from (3.88) and (3.89)

$$\frac{dD'}{dt} = \chi \frac{D'}{1 + D'}, \tag{3.92}$$

where

$$\chi \equiv \frac{Q_0}{\rho D_0\left(1 - \epsilon_p - \gamma / \mathbf{S}\right)}\left(\frac{\gamma}{\epsilon_p \mathbf{S}}\left(1 - \frac{G}{\mathbf{S}}\right) - \frac{G}{\mathbf{S}}\right). \tag{3.93}$$

It is clear from (3.92) and (3.93) that disturbances will amplify only if $\chi > 0$ and therefore

$$\gamma > \frac{\epsilon_p G}{1 - \dfrac{G}{\mathbf{S}}}.$$

Thus the radiative feedbacks from convective clouds and moisture must not only be positive but be large enough to meet this criterion. The gross moist stability, represented by G, inhibits instability, and smaller values of G promote amplification.

Although (3.92) is nonlinear, it is easily integrated, yielding

$$D' - D'_i + \ln\left(\frac{D'}{D'_i}\right) = \chi t, \tag{3.94}$$

where D'_i is the initial value of D'. But note that the solutions cannot be continued to arbitrarily large time because there are physical bounds on the magnitude of D'. There are two separate upper bounds on D': First, it cannot become so large that the total water vapor content becomes negative, and second, we must insure that M_u does not become negative in the dry branch of the circulation. From (3.88), M_u crosses zero when

$$D' = \frac{1-\epsilon_p - \gamma/\mathbf{S}}{\epsilon_p + \gamma/\mathbf{S}}.$$

Second, it is not usually observed that the radiative effects of aggregated clouds and water vapor are powerful enough to change the sign of the net column radiative cooling, so in the moist columns we enforce $\dot{Q}_{cool} = Q_0 - \gamma\left(M_u - M_{eq}\right) \geq 0$. From (3.88), net radiation vanishes when

$$D' \equiv D'_Q = -\frac{\epsilon_p}{1-\epsilon_p}\frac{1-\epsilon_p-\gamma/\mathbf{S}}{\epsilon_p+\gamma/\mathbf{S}}.$$

So after the time when D' in the moist columns exceeds this value, we assume that the net radiative cooling in these columns vanishes. Applying this condition to (3.87), (3.89) and (3.90) gives

$$\frac{dD'}{dt} = \frac{Q_0}{\rho D_0}\left[\frac{G}{\left(1-\epsilon_p\right)\mathbf{S}}\frac{1}{1+D'} - 1\right],$$

whose solution is

$$G'\ln\left[\frac{1+D'_Q-G'}{1+D'-G'}\right] + D'_Q - D' = \frac{Q_0}{\rho D_0}\left(t-t_Q\right) \quad \text{for } t > t_Q, \tag{3.95}$$

where $G' \equiv G/\left[\left(1-\epsilon_p\right)\mathbf{S}\right]$ and t_Q is the time when $D' = D'_Q$.

Figure 3.40 shows a solution set of (3.94) and (3.95) for both the positive and negative branches (determined by the sign of D_i) in the case that the positive branch of D' for particular values of the parameters and for which the initial values D_i are ± 0.01. The solution is expressed in terms of a column-relative humidity-like variable defined as $1-\epsilon_p\left(1+D'\right)$ and the aggregate-scale vertical velocity w_b, suitably normalized, is also shown in the figure. The two branches clearly follow different trajectories after the disturbances attain noticeable amplitude, and the upward motion attains a larger magnitude than the downward motion, indicating that the ascent branch occupies a smaller fractional area than does the descending air.

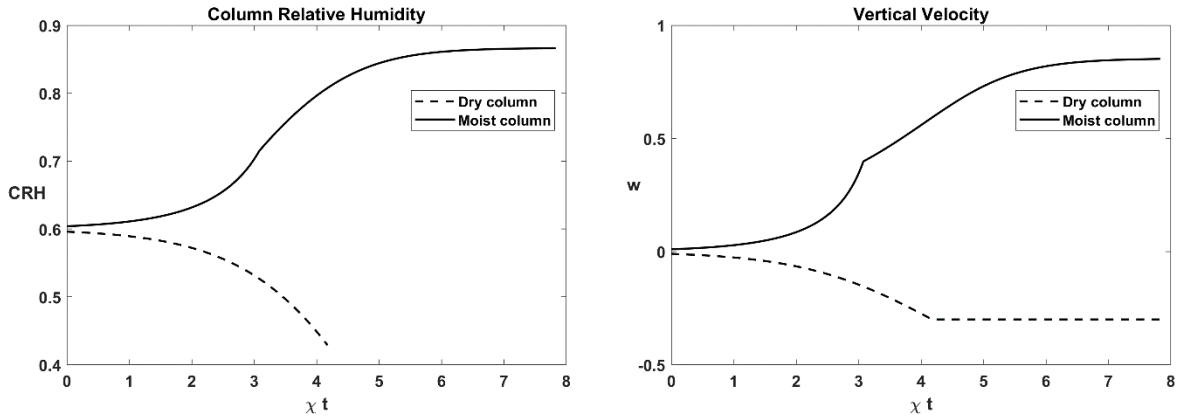


Figure 3.40: Solutions (3.94) and (3.95) for a particular set of parameters. The left panel shows a column-relative-humidity variable defined as $1-\epsilon_p\left(1+D'\right)$, *and the right panel shows a normalized aggregate-scale vertical velocity. The solid line is for moist ascending column while the dashed line is for descending columns.*

The reader is invited to explore the parameter dependencies of these solutions using the MATLAB script *aggplot.m* that can be downloaded from https://zenodo.org/doi/10.5281/zenodo.10800835 .

The destabilization of the RCE state by feedbacks involving radiation and surface fluxes is an important source of intraseasonal variability and is also important in driving tropical cyclones. We will have occasion to discuss self-aggregation again in our coming treatments of intraseasonal variability and tropical cyclones.

3.5 Tropical squall lines

On shorter space and time scales, deep convection is often organized into lines and arcs, as illustrated in Figures 3.5 and 3.6. In this case, the organization is brought about by the interaction of cold pools, produced by evaporating rain, with ambient wind shear.

The interaction of dry convection with background shear flows strongly favors convective rolls aligned with the shear vector (e.g. Kuo, 1963) and this is observed to be the case in laboratory experiments (e.g. Graham and Walker, 1933). But strong squall lines are observed to be oriented more nearly perpendicular to the low-level shear vector (Barnes and Sieckman, 1984; LeMone and G. M. Barnes, 1984), and this tends to be true outside the tropics as well (Bluestein and Jain, 1985; Wyss and Emanuel, 1988). Tropical squall lines seem to be strongest when the shear is confined to low levels or is associated with a low-level jet with reversed shear above the jet (LeMone and G. M. Barnes, 1984).

An essential process in deep moist convection is the development of cold pools of air in the boundary layer, resulting from the evaporation of rain falling into the subcloud layer or through unsaturated air above the subcloud layer. The cold, dense air flows outward as a density current. In environments of low shear, such as classical RCE, the outflowing cold air eventually cuts off the supply of potentially unstable boundary layer air and the convective cell dies, as first documented by Byers and Braham (1948) based on results of one of the first field experiments designed to understand how deep convection works. The evolution of a typical deep convective cell in a low-shear environment is shown in Figure 3.41.

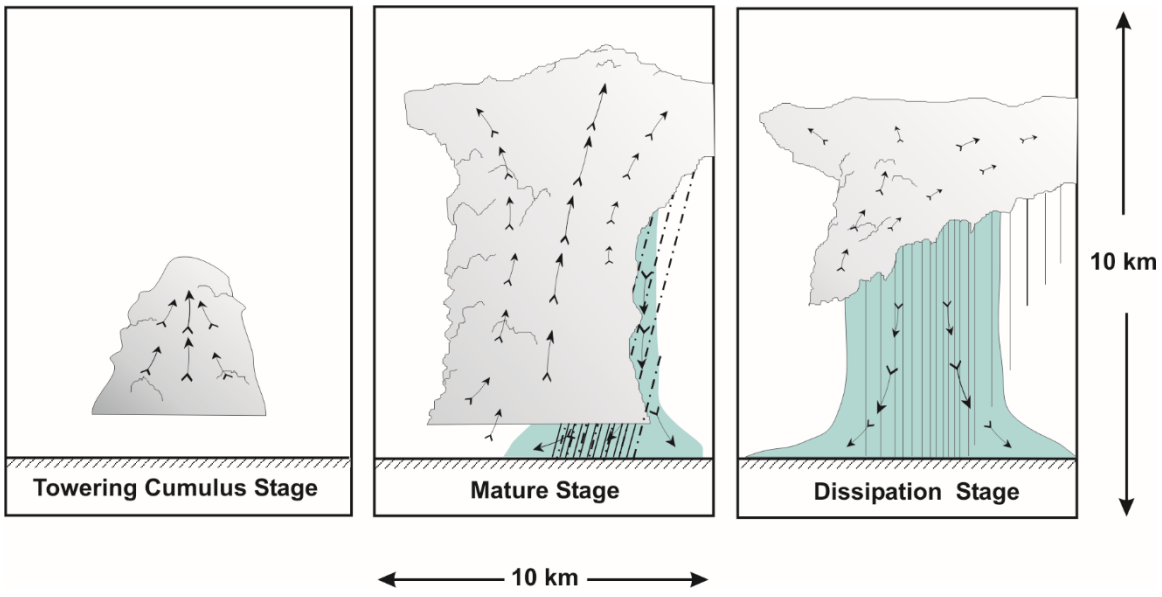


Figure 3.41: Evolution of a convective cell in a low-shear environment. In the earlies stages (left), an updraft develops in response to instability, but cloud droplets have not yet aggregated into precipitation. Later (center). precipitation forms and falls out, some if it evaporating in the subcloud layer and in unsaturated air above the subcloud layer. The cold, dense air begins (blue shading) to flow outward as a density current. Finally, the outflowing cold air cuts off the supply of potentially buoyant air at low levels, and the cell dissipates (right). The whole sequence may take 45 minutes to an hour.

However, if enough low-level shear is present, then the cold pool on the downshear side of the system will be inhibited from outrunning the deep convection aloft, and a quasi-two-dimensional squall line can thus persists somewhat longer, as theorized by Thorpe et al. (1982) and Rotunno et al. (1988) and illustrated in Figure 3.42. Following Rotunno et al. (1988), we show that there is an optimal magnitude of the low level shear for strong, persistent convection.

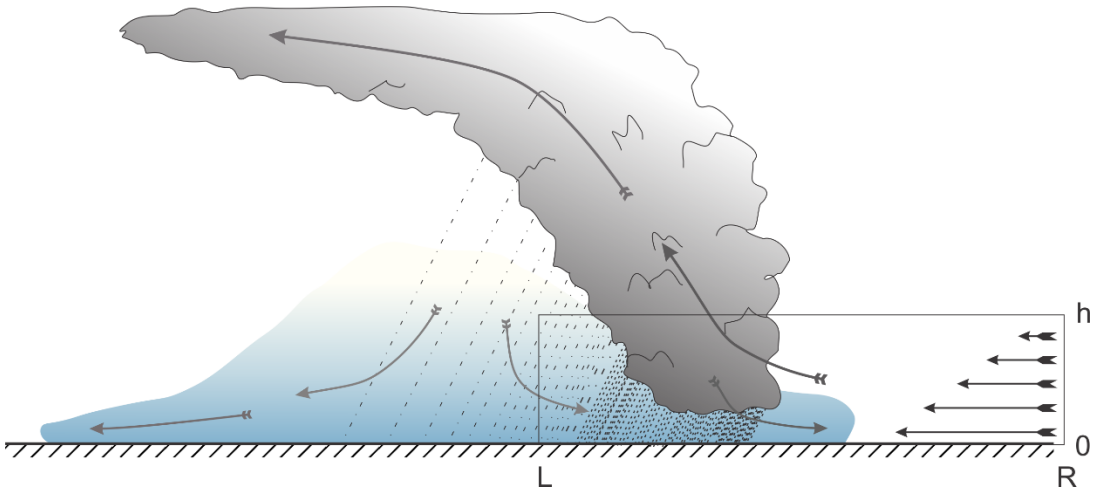


Figure 3.42: Schematic cross-section through a tropical squall line, in a frame of reference moving with the system from left to right. Gray arrows denote system-relative airflow and the straight black arrows at right show the ambient low-level shear flow relative to the squall line. The blue shading depicts downdraft air cooled by evaporation of falling precipitation and, at higher levels, melting of ice. The rectangular box in the lower right is a control volume described in the text.

We begin by defining η as the component of vorticity into the page, referring to Figure 3.42, and then write the conservation equation for η assuming that the flow is two-dimensional and Boussinesq at low levels:

$$\frac{\partial \eta}{\partial t} + \frac{\partial}{\partial x}(u\eta) + \frac{\partial}{\partial z}(w\eta) = -\frac{\partial B}{\partial x}, \tag{3.96}$$

where u and w are the components of velocity in the x and z directions, respectively, and B is the buoyancy, defined

$$B \equiv g\frac{T_v'}{\overline{T_v}}$$

in which T_v' is the departure of the virtual temperature from its unperturbed value ahead of the squall line, $\overline{T_v}$. By definition, then, $B = 0$ in the air out ahead of the squall line, on the right side of Figure 3.42. We now integrate (3.96) over the control volume indicated by the rectangle in the lower right of Figure 3.42. In doing this integral, we assume that the flow is steady, that there is no horizontal flow u on the left side of the box (denoted by L in the figure), and that the vertical velocity w vanishes at the surface. With these assumptions, integrating (3.95) over the area of the box yields

$$\int_0^h (u\eta)_R \, dz + \int_L^R (w\eta)_h \, dx = \int_0^h B_L dz. \tag{3.97}$$

If we further assume that the flow is strictly horizontal along the right edge of the box, then $\eta = \frac{\partial u}{\partial z}$ there and we can integrate the first term in (3.97):

$$\frac{1}{2}\left[u_0^2 - u_h^2\right]_L - \int_L^R (w\eta)_h \, dx = -\int_0^h B_L dz. \tag{3.98}$$

(Recall that the buoyancy, B_L , of the cold pool is negative, so the right side of (3.98) is positive.) Rotunno, Weiss, and Klemp (1988) contend that for strong, long-lived convection, the updraft must not emerge from the boundary layer with much vorticity η, else the updraft would be strongly tilted away from the vertical. If we further assume that there is no squall line-relative ambient flow in the cloud layer, that optimal condition is

$$u_{0L}^2 = -2\int_0^h B_L dz. \tag{3.99}$$

Thus the colder the cold pool, the greater the difference in the square velocity between the boundary layer and the cloud layer is needed to meet the optimal condition (3.99).

If the flow is superoptimal from this point of view, we might expect the squall lines to form at an angle to the shear vector such that the cross-line flow is optimal. This is because shear along a two-dimensional disturbance has little or no effect on it, so there is no penalty to pay for rotating the orientation of the convection off an axis perpendicular to the low-level shear.

It is possible to impose a background wind profile, including shear, on a cloud-permitting RCE simulation, by relaxing the horizontally averaged flow at each level toward a target wind speed. (A nice fringe benefit of doing so is the ability to deduce the convective contribution to the vertical flux of horizontal momentum by keeping track of how much momentum has to be added to each layer to keep the mean wind field close to the target background wind.)

A series of numerical experiments was performed by Robe and Emanuel (2001), running a cloud-permitting model into an RCE state while driving the domain-average winds towards prescribed wind profiles. They ran the model on a 180 x 180 km domain with periodic boundary conditions and a horizontal grid spacing of 2 km, and forced the convection by imposing a constant radiative cooling rate of $5.4\ K\ day^{-1}$ from the surface to 13 km altitude[4]. Figure 3.43 shows snapshots of the updraft fields at 250 m altitude during the RCE states of three experiments, for three different prescribed background wind profiles. The left panel shows the classical zero-mean-wind case; the middle panel shows a simulation in which a low-level shear of $6\ ms^{-1}$ was imposed over the lowest 2 km, and the right panel shows the same but with a net shear of $10\ ms^{-1}$. In both the latter two cases, the shear vector points from right to left (opposite to the direction of the shear shown in Figure 3.42). In these simulations, the negative buoyancy of the cold pool increased with increasing imposed shear, so the right-hand side of (3.99) is not independent of the specified background flow.

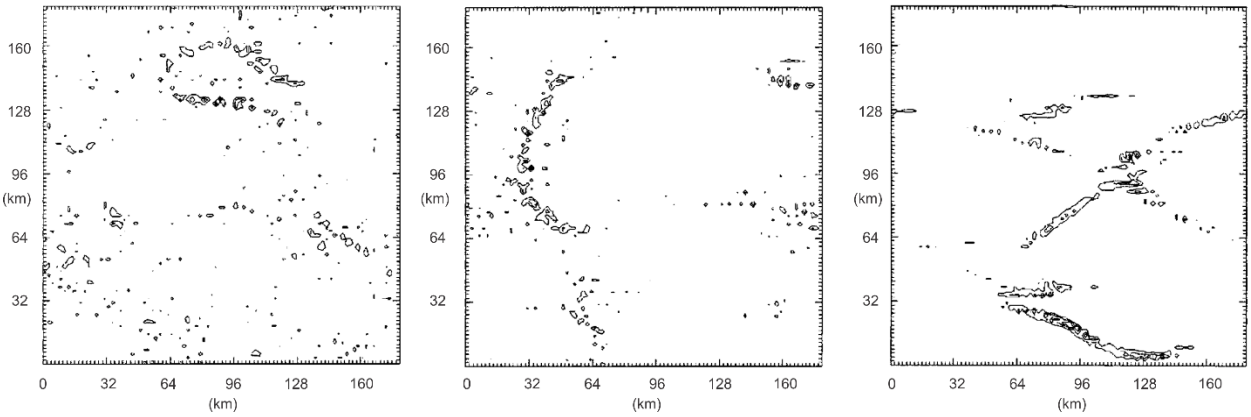


Figure 3.43: Snapshots of the updraft velocities at 250 m altitude from three simulations using a cloud-permitting model run into RCE states in a doubly-periodic domain. Left: No imposed background wind. Center and Right: $6\ ms^{-1}$ *and* $10\ ms^{-1}$*, respectively, of imposed background shear over the altitude range 0-2 km and directed right to left.*

In the classical RCE case with no imposed background shear (left panel of Figure 3.43), the convection is more or less random in space and time. In the simulation with moderate imposed low-level shear flow (center panel), the convection takes the form of arcs oriented across the shear vector and propagating downshear (right to left). The simulation with strong low-level shear (right panel) shows convection aligned at angles to the imposed shear, and also propagating downshear.

[4] The constant radiative cooling and the relatively small domain size prevented self-aggregation from occurring, though the phenomenon was relatively unknown at the time. The rate of imposed cooling was unrealistically large but hastened the approach to RCE.

Other experiments not shown here also exhibit impressive squall-line organization when the shear vector rotates, or even reverses with altitude. If mono-directional shear extends through too deep a layer, the convection becomes organized along rather than across the shear vector.

These experiments also showed that the convective momentum flux is broadly down-gradient, but is not closely related to the local shear. The non-local nature of the convective momentum flux is consistent with results derived from observations of tropical squall lines by LeMone et al. (1984).

We shall return to the subject of the spatial and temporal organization of moist convection in our discussion of boundary layer convection in the next chapter.

References

Arakawa, A., and W. H. Schubert, 1974: Interaction of a cumulus cloud ensemble with the large-scale environment, part I. *J. Atmos. Sci.*, **31,** 674-701.

Barnes, G. M., and a. K. Sieckman, 1984: The environment of fast- and slow-moving tropical mesoscale convective cloud lines. *Mon. Wea. Rev.*, 112.

Bluestein, H. B., and a. M. H. Jain, 1985: Formation of mesoscale lines of precipitation: Severe squall lines in Oklahoma during the spring. *J. Atmos. Sci.*, **42,** 1711-1732.

Bony, S., and K. A. Emanuel, 2001: A parameterization of the cloudiness associated with cumulus convection: Evaluation using toga coare data. *J. Atmos. Sci.*, **58,** 3158-3183.

Bretherton, C. S., and a. P. K. Smolarkiewicz, 1989: Gravity waves, compensating subsidence and detrainment around cumulus clouds. *J. Atmos.. Sci.*, **46,** 740-759.

Bretherton, C. S., P. N. Blossey, and M. F. Khairoutdinov, 2005: An energy-balance analysis of deep convective self-aggregation above uniform sst. *J. Atmos. Sci.*, **62,** 4273-4292.

Byers, H. R., and R. R. Braham, Jr., 1948: Thunderstorm structure and circulation. *J. Meteor.*, **5,** 71-86.

Cronin, T., and K. A. Emanuel 2013: The climate time scale in the approach to radiative-convective equilibrium. *J. Adv. Model. Earth Sys.*, **5**, doi:10.1002/jame.20049.

Emanuel, K., A. A. Wing, and E. M. Vincent, 2014: Radiative-convective instability. *J. Adv. Model. Earth Sys.*, **5,** 1-14, doi::10.1002/2013MS000269.

Emanuel, K. A., 1993: The effect of convective response time on WISHE modes. *J. Atmos.. Sci.*, **50,** 1763-1775.

Emanuel, K. A., and M. Živkovic-Rothman, 1999: Development and evaluation of a convection scheme for use in climate models. *J. Atmos. Sci.*, **56,** 1766-1782.

Graham, A., and G. T. Walker, 1933: Shear patterns in an unstable layer of air. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, **232,** 285-296, doi:doi:10.1098/rsta.1934.0008.

Held, I. M., R. S. Hemler, and V. Ramaswamy, 1993: Radiative-convective equilibrium with explicit two-dimensional moist convection. *J. Atmos. Sci.*, **50,** 3909-3927, doi:10.1175/1520-0469(1993)050<3909:Rcewet>2.0.Co;2.

Khairoutdinov, M. F., and D. A. Randall, 2003: Cloud resolving modeling of the arm summer 1997 iop: Model formulation, results, uncertainties and sensitivities. *J. Atmos. Sci.*, **60,** 607-625.

Kuo, H. L., 1963: Perturbations of plane Couette flow in stratified fluid and origin of cloud streets. *The Physics of Fluids*, **6,** 195-211, doi:10.1063/1.1706719.

LeMone, M. A., G. M. Barnes, and E. J. Zipser, 1984: Momentum flux by lines of cumulonimbus over the tropical oceans. *J. Atmos.. Sci.*, **41,** 1914-1932.

LeMone, M. A., and E. J. S. G. M. Barnes, and E. J. Zipser, 1984: The tilt of the leading edge of mesoscale tropical convective lines. *Mon. Wea. Rev.*, **112,** 510-519.

Manabe, S., and R. F. Strickler, 1964: On the thermal equilibrium of the atmosphere with convective adjustment. *J. Atmos. Sci.*, **21,** 361-385.

Manabe, S., J. Smagorinsky, and R. F. Strickler, 1965: Simulated climatology of a general circulation model with a hydrologic cycle. *Mon. Wea. Rev.*, **93,** 769-798.

Manabe, S., and R. T. Wetherald, 1967: Thermal equilibrium of the atmosphere with a given distribution of relative humidity. *J. Atmos. Sci.*, **24,** 241-259, doi:10.1175/1520-0469(1967)024<0241:teotaw>2.0.co;2.

Morcrette, J.-J., 1991: Radiation and cloud radiative properties in the european centre for medium-range weather forecasts forecasting system. *J. Geophys. Res.*, **96,** 9121-9132.

Morton, B. R., G. I. Taylor, and J. S. Turner, 1956: Turbulent gravitational convection from maintained and instantaneous sources. *Proc. Roy. Soc. London*, **A234,** 1-23.

Neelin, J. D., and I. M. Held, 1987: Modeling tropical convergence based on the moist static energy budget. *Mon. Wea. Rev.*, **115,** 3-12.

Paluch, I. R., 1979: The entrainment mechanism in colorado cumuli. *J. Atmos. Sci.*, **36,** 2462-2478.

Pauluis, O., and I. M. Held, 2002: Entropy budget of an atmosphere in radiative-convective equilibrium. Part I: Maximum work and frictional dissipation. *J. Atmos. Sci.*, **59,** 125-139.

Robe, F. R., and K. Emanuel, 2001: The effect of vertical wind shear on radiative-convective equilibrium states. *J. Atmos. Sci.*, **58,** 1427-1445.

Rotunno, R., J. B. Klemp, and M. L. Weisman, 1988: A theory for strong, long-lived squall lines. *J. Atmos. Sci.*, **45,** 463-485.

Schmidt, W., 1941: Turbulent propagation of a stream of heated air. *Z. angew. Math. Mech.*, **21,** 265-351.

Singh, M. S., and P. A. O'Gorman, 2015: Increases in moist-convective updraught velocities with warming in radiative-convective equilibrium. *Quart. J. Roy. Meteor. Soc.*, **141,** 2828-2838, doi:10.1002/qj.2567.

Sobel, A. H., and C. S. Bretherton, 2000: Modeling tropical precipitation in a single column. *J. Climate*, **13,** 4378-4392, doi: 10.1175/1520-0442(2000)013<4378:Mtpias>2.0.Co;2.

Thorpe, A. J., M. J. Miller, and M. W. Moncrieff, 1982: Two-dimensional convection in non-constant shear: A model of mid-latitude squall lines. *Quart. J. Roy. Meteor. Soc.*, **108,** 739-762.

Tompkins, A. M., and G. C. Craig, 1998: Radiative–convective equilibrium in a three-dimensional cloud-ensemble model. *Quart. J. Roy. Meteor. Soc.*, **124,** 2073-2097, doi:doi:10.1002/qj.49712455013.

Wing, A. A., and K. A. Emanuel, 2014: Physical mechanisms controlling self-aggregation of convection in idealized numerical modeling simulations. *J. Adv. Model. Earth Sys.*, **6,** 75-90, doi::10.1002/2013MS000270.

Wing, A. A., and T. W. Cronin, 2016: Self-aggregation of convection in long channel geometry. *Quart. J. Roy. Meteor. Soc.*, **142,** 1-15, doi:10.1002/qj.2628.

Wing, A. A., K. Emanuel, C. E. Holloway, and C. Muller, 2017: Convective self-aggregation in numerical simulations: A review. *Surv. Geophys.*, doi: 10.1007/s10712-017-9408-4.

Wyss, J., and a. K. A. Emanuel, 1988: The pre-storm environment of midlatitude prefrontal squall lines. *Mon. Wea. Rev.*, **116,** 790-794.

Xu, K.-M., and K. A. Emanuel, 1989: Is the tropical atmosphere conditionally unstable? *Mon. Wea. Rev.*, **117,** 1471-1479.

Zhang, M. H., and J. L. Lin, 1997: Constrained variational analysis of sounding data based on column-integrated budgets of mass, heat, moisture, and momentum: Approach and application to arm measurements. *J. Atmos. Sci.*, **54,** 1503-1524, doi:10.1175/1520-0469(1997)054<1503:cvaosd>2.0.co;2.

4 Tropical Boundary Layers

The atmospheric boundary layer is the layer of air adjacent to the surface that is influenced by the surface on time scales of minutes to hours. It is usually dominated by turbulent transport, with the exception of calm nights over land in which the cooling from below, in the absence of mean wind, suppresses any turbulence. The turbulence may be generated by shear instabilities as air passes over rough surfaces, and/or by convection when there is a buoyancy source at the surface. Turbulent boundary layers often have well defined upper boundaries above which the turbulence ceases abruptly, and air from the overlying quiescent atmosphere is turbulently entrained into the boundary layer.

Boundary layer turbulence is an essential means by which the atmosphere and the underlying surface communicate, and the rates of transfer of heat, moisture, momentum, aerosols, and trace gases between the surface and the atmosphere are controlled by the turbulence. It is therefore essential to understand boundary layers for understanding virtually everything that happens in the tropical atmosphere. For example, when we come to discuss tropical cyclones, we will see that their intensity and structure is mostly controlled by enthalpy and momentum fluxes between the atmosphere and ocean.

Over land, the character of the boundary layer is strongly influenced by the diurnal cycle of solar radiation, but even over the sea there are subtle diurnal variations owing to direct absorption of sunlight in the atmosphere and by clouds, and if there is little surface wind, by variations in the skin temperature of the ocean.

Water plays several important roles in boundary layer physics. Because its molecular weight is less than a suitably defined mean molecular weight of the other constituents of air, moist air is less dense than dry air at the same pressure and temperature. Consequently, water evaporating into the atmosphere from the surface is a source of buoyancy and can be the dominant source over warm oceans. Much of the boundary layer over tropical oceans contains cloud, and the phase changes of water – including condensation and evaporation of cloud droplets and the fall and evaporation of rain – can have important effects on the structure and turbulent properties of the boundary layer. Moreover, the radiative effects of clouds are substantial, particularly in the case of stratocumulus-topped boundary layers, whether the cloud cover can be close to 100%.

4.1 Some general principles

For the purposes of this book, we will consider only turbulent boundary layers and pass over the problem of laminar boundary layers, which can sometimes be found over land or cold water under very low wind conditions. We will consider all of the earth's surface to be thermodynamically and mechanically "rough", in the sense that fluxes to and from the surface are not rate-limited by molecular diffusivities. Therefore, none of the scaling laws we will review here depend on molecular diffusivities. This makes atmospheric boundary layers fundamentally different from other boundary layers found in literature on, e.g. Raleigh convection, pipe flow, aerodynamics, and other applications.

In the case of clear or stratocumulus-topped boundary layers, the whole layer may be considered to be filled with turbulence, but in trade cumulus boundary layers the turbulence is

mostly confined to the cumulus clouds and the cloud-free air in between may be largely free of small-scale turbulence. Even so, we will consider the cloud-bearing layer to be part of the overall boundary layer. In this case, the cumuli are important elements of the turbulent exchange between the surface and the atmosphere.

In regions experiencing deep convection, by the same reasoning, one might consider the whole troposphere to be a boundary layer, as convective turbulence is, as we have seen in previous chapters, and essential turbulent process in transferring energy, water, aerosols, and other trace constituents through the atmosphere. Nevertheless, by convention, we do not consider the whole layer containing deep convection to be a boundary layer.

It is important to be precise about the kinds of questions we ask about boundary layers. We might be interested in them for their own sake – for example, the nature of the turbulence itself, and its organization into geometric forms. On the other hand, our primary interest might be in how the boundary layer affects the free troposphere above it; for example, how it mediates the turbulent and radiative fluxes into the free troposphere from below. In the climate arena, most of the planet's albedo arises from clouds, and much of this is from boundary layer clouds, so here it is clearly essential to understand and be able to predict where and when cloud-topped boundary layers are present.

In thinking about the boundary layer as a system, it is important to be clear about which are the externally specifiable conditions and which are internally controlled conditions. For example, in the Prandtl convection problem explored in Chapter 2, section 2.3, the external conditions are the cooling rate applied to the fluid and the temperature and roughness of the surface; everything else, such as the turbulent velocities and time-mean vertical temperature profile, are internally determined. One can express the surface heat flux as a function of the mean turbulent flow speed, the surface roughness, and the temperature gradient across some specified altitude range, and it becomes tempting to think that the surface flux is "caused" by the flow and the temperature gradient, but this would be misleading. As we shall see, surface fluxes in models are often parameterized as functions of a near-surface flow speed and gradients in quantities between the surface and some specified altitude. In reality, both the gradients and flow speeds are functions of more nearly external conditions such as internal cooling rates and surface roughness. Calculations of surface fluxes using parameterizations that involve gradients and flow speeds may have to be iterative, as both are functions of externally specifiable conditions.

In what follows we treat boundary layers of increasing complexity, yet aiming for simplicity and conceptual clarity.

4.2 Semi-infinite thermal and mechanical turbulent layers

Although clearly unrealistic, semi-infinite Boussinesq fluids in a constant gravitational field and bounded below by a rough, rigid surface are useful idealizations because they are controlled by a very limited number of parameters.

We have already derived, on dimensional grounds, velocity and buoyancy scales for a convective boundary layer driven by internal cooling and with no mechanical forcing (see Chapter 2). Here we follow the same strategy in deriving scales for a boundary layer driven

entirely by mechanical forcing, with no thermal forcing; here all the turbulence is driven by shearing instabilities arising from flow deceleration near the mechanically rough lower boundary.

In the Prandtl convective system we specified a finite vertically integrated cooling rate that, in equilibrium, must also equal the total heat flux through the lower boundary. Here we do the mechanical equivalent by specifying a vertically integrated momentum flux per unit mass, M, that has dimensions L^2t^{-2} that can be thought of as supplied by a vertically uniform, infinitesimal horizontal pressure gradient. In statistical equilibrium, the vertically integrated momentum sink must be balanced by a momentum flux through the surface, or equivalently, a surface stress. By convention, this surface stress per unit mass is denoted as u_*^2, where u_* is known as the *friction velocity.*

To avoid singularities, we introduce a mechanical roughness length z_0 below which molecular diffusion of momentum becomes important. Once again, the size of the turbulent eddies must scale with altitude, z. The velocity of the eddies scales as u_*, and the ensemble average wind shear must scale as $u_* z^{-1}$:

$$\frac{d\bar{U}}{dz} = \frac{1}{k} u_* z^{-1}, \tag{4.1}$$

where the constant factor k is called the von Kármán constant. Integrating (4.1) with altitude and taking $\bar{U} = 0$ at $z = z_0$ gives

$$\bar{U} = \frac{1}{k} u_* \ln\left(\frac{z}{z_0}\right). \tag{4.2}$$

The logarithmic increase of mean winds with attitude in mechanically forced boundary layers has been experimentally verified in many settings. The von Kármán constant is based on fits of observed wind profiles to (4.2) and is a number close to 0.4.

What happens when there is both mechanical and thermodynamic forcing? We can begin by looking at the ratio of convective to mechanical turbulent velocity scales:

$$R = \frac{\left(zF_B\right)^{1/3}}{u_*}. \tag{4.3}$$

This ratio increases with height, so we might expect that at low levels shear driven turbulence dominates while at high levels the turbulence is driven mostly by convection. From (4.3) the altitude L at which the two velocity scales are roughly equal is given by

$$L = \frac{u_*^3}{kF_B}. \tag{4.4}$$

This length scale is known as the Obukhov length[1] after the Russian scientist who helped develop the theory of mechanically forced boundary layers that also have buoyancy effects. The von Kármán constant is included by convention. For our purposes, L can be regarded as the altitude above which convective turbulence dominates. In boundary layers that are both mechanically and convectively forced, we expect turbulence quantities to also scale with the nondimensional altitude z/L. Typical values of L range from 1 to 50 m, which is usually a reasonably small fraction of the total boundary layer depth in the tropics. Over tropical oceans, the top of the boundary layer is typically a few hundred meters, so it is usually the case that the bulk of the boundary layer is convectively driven.

The presence of both mechanical and convective forcing effectively introduces a non-dimensional parameter into the problem; namely, the ratio of the altitude to the Obukhov length L. This requires us to modify (4.1), which had been developed for the strictly mechanical boundary layer, to

$$\frac{d\overline{U}}{dz} = \frac{1}{k} u_* z^{-1} \Psi_M \left(\frac{z}{L} \right), \tag{4.5}$$

where Ψ_M is some universal function. Likewise, we must modify the strictly thermally forced boundary layer profile given by (2.54) to

$$g\beta \frac{d\overline{T}}{dz} = -c_1 F_B^{2/3} z^{-4/3} \Psi_T \left(\frac{z}{L} \right), \tag{4.6}$$

where Ψ_T is another universal function. These relations are the cornerstone of turbulent boundary layer theory, known as Monin-Obukhov similarity theory (Monin and Obukhov, 1954). The functions Ψ_M and Ψ_T must be determined empirically, through field measurements, but to reduce appropriately to the purely mechanical and purely thermal limits, they must obey $\Psi_M \left(L \to \infty \right) = 1$ and $\Psi_T \left(L \to 0 \right) = 1$. Moreover, we expect that near the thermal limit, momentum may behave more like a passive tracer while near the mechanical limit, temperature may act like a passive tracer; this hypothesis places more limits on the functional forms. Candidate functions derived from field experimental data (e.g. Businger and J. C. Wyngaard, 1971; Dyer, 1974) that meet these limits are

$$\Psi_M = \left(1 + 11 \left(\frac{z}{L} \right) \right)^{-1/3} \tag{4.7}$$

and

[1] Note that when boundary layers are thermally stable, F_B is negative. But by convention, the Monin-Obukov length is defined to be negative when F_B is positive. Since we will not be dealing with stable boundary layers here, we buck that convection and define L to be positive for convecting boundary layers.

$$\Psi_T = \left(1 + \frac{1}{13}\left(\frac{L}{z}\right)\right)^{-1/3}. \tag{4.8}$$

These relations are only valid for unstable boundary layers for which L (as we have defined it here) is positive.

We have thus far treated the atmosphere as a homogeneous gas. But particularly in the lower tropical atmosphere, water vapor variability makes an important contribution to the variability of specific volume. This can be seen in the definition of the virtual temperature given by (3.10) and repeated here:

$$T_v \equiv T\left(\frac{1 + r/\varepsilon}{1 + r}\right), \tag{4.9}$$

where r is the mass concentration of water vapor (mass of water per unit mass of dry air) often referred to simply as the *mixing ratio*, and $\varepsilon \equiv R_d / R_v \cong 0.622$ is the ratio of the molecular weight of water to a suitably defined average molecular weight of dry air. Since $\varepsilon < 1$, the specific volume of moist air is greater than that of dry air at the same temperature and pressure, and thus the virtual temperature is greater than the actual temperature. A balloon filled with moist air will rise through dry air at the same temperature. Unfortunately, this is not a strong enough effect to be of use to balloonists, but it is important over tropical oceans: Evaporation of surface liquid water into air increases is specific humidity, making it lighter and thus destabilizing the boundary layer. *Much of the buoyancy flux in the boundary layer over tropical oceans is owing to evaporation rather than sensible heat flux.*

Because r is conserved in the absence of evaporation and condensation, we can define a conserved variable called the *virtual potential temperature*:

$$\theta_v = T_v\left(\frac{p_0}{p}\right)^{R_d/c_p}. \tag{4.10}$$

(Technically, the constants that appears in the exponent of (4.10) should be mixing ratio-weighted combinations of the gas constants and heat capacities of dry air and water vapor, but by a happy accident, the ratio of these weighted parameters is nearly equal to the ratio of their dry values.) Likewise, buoyancy fluctuations can be written

$$B = g\frac{\theta_v'}{\overline{\theta_v}}. \tag{4.11}$$

All the solutions for the convective boundary layer discussed in Chapter 2, section 2.3, should be re-written replacing actual by virtual temperatures and potential temperatures.

4.3 Aerodynamic flux formulae

Specification of the bulk heat sink and momentum source of the fluid is equivalent to specifying the surface fluxes of heat and momentum only if the system is in equilibrium. If the system is not in equilibrium, or if one only knows the wind speed and temperature, it is better to try to relate the surface fluxes to quantities like the surface temperature, atmospheric temperature, and wind speed.

This is particularly straightforward if we can measure quantities at an altitude (or choose a model level) that is much larger than the roughness length but much smaller than L. This will be usually be possible except where background winds are very light and/or the surface buoyancy forcing is very large (for example, over a desert in daytime). Standard measuring altitudes for surface meteorological stations around the world is, for example, 10 m. If these conditions are met, we can assume we are in the mechanical limit of turbulence generation and write a relation for the surface stress τ by algebraically inverting (4.2):

$$\tau \equiv \rho u^{*2} = \frac{k^2}{\left(\ln\left(\frac{z_m}{z_0}\right)\right)^2}\rho\overline{U_m}^2 \equiv C_D\rho\overline{U_m}^2, \tag{4.12}$$

where the subscript m denotes quantities at the measurement altitude and

$$C_D \equiv \frac{k^2}{\left(\ln\left(\frac{z_m}{z_0}\right)\right)^2} \tag{4.13}$$

is a non-dimensional number known as the *drag coefficient*, which is clearly a function of the roughness length and the measurement altitude. This can easily be generalized to flows in an arbitrary direction:

$$\boldsymbol{\tau} = C_D\rho\,|\,\mathbf{V}_m\,|\,\mathbf{V}_m, \tag{4.14}$$

where $\mathbf{V}_m$ is the horizontal velocity vector at the measurement altitude. We can also integrate (4.6) in the vertical and then invert it, using the empirical function Ψ_T given by (4.8) in the limit of large L/z, to relate the surface buoyancy flux to the difference in virtual potential temperature between the surface itself and the measurement altitude z_m:

$$F_B = C_T\,|\,\mathbf{V}_m\,|\,g\beta\left(\theta_{v0} - \theta_{vm}\right), \tag{4.15}$$

where θ_{v0} is the virtual potential temperature of the surface and C_T is the surface exchange coefficient for heat and is related to the constants we have previously defined:

$$C_T = \frac{\sqrt{C_D}}{c_1(13k)^{1/3}\ln\left(z_m/z_0\right)}. \tag{4.16}$$

As with the drag coefficient, this nondimensional coefficient is a function of the measurement altitude and the surface roughness length.

Fluxes of other passive tracers should obey relations similar to (4.15), as long as the measurement altitude is much larger than the surface roughness length but much smaller than L.

The relations (4.14) and (4.15), and similar relations for passive tracers, are collectively known as aerodynamic flux formulae. They are straightforward to apply to field measurements to estimate surface fluxes, but in models it must always be remembered that quantities like $\mathbf{V}_m$ and θ_{vm} are themselves functions of the surface fluxes, so the fluxes must be calculated iteratively. This is a consequence of the fact that such quantities are not true externals of turbulent boundary layers. In addition, over the ocean, the surface roughness depends on the presence and amplitude of surface waves and capillary waves, and will also vary with the direction of the wind relative to the orientation of the surface waves. In spite of these complications, the aerodynamic flux formulae enjoy widespread use, and when we come to discuss tropical cyclone physics, we will see that they depend crucially on turbulent exchanges of enthalpy and momentum between the boundary layer and the surface.

4.4 Clear boundary layers

The atmosphere sufficiently far away from the surface is nearly always stably stratified to vertical displacements that do not involve phase change of water. In the tropics, this is mostly because deep moist convection, even if it is distant, keeps the virtual temperature lapse rate close to neutral to moist displacements (see Chapter 3, section 3.2), and such a lapse rate is stable to reversible vertical displacements in which water does not change phase.

When stably stratified fluid flows over the earth's surface, the resulting shear causes a turbulent boundary layer to form; likewise, a positive surface buoyancy flux will result in convection that forms a turbulent layer.

As the turbulent eddies impinge on the stably stratified fluid above the turbulent layer, they engulf some of that fluid, which then becomes part of the turbulent layer. This process is called *entrainment*, similar to the lateral entrainment that occurs with deep moist convection (Chapter 3).

As air is entrained into the boundary layer from above, the layer grows, and the rate of growth can be described by an *entrainment velocity*, w_e. Over land, the growth of the boundary layer is often limited by the duration of solar heating of the surface; after the sun sets the source of turbulent energy vanishes (unless there is enough background wind) and the boundary layer collapses. Over the sea, the depth of the boundary layer is usually limited by radiatively induced subsidence in the overlying air. If extensive clouds form at the top of the boundary layer, though, phase changes of water and strong radiative cooling from the tops of the clouds can dramatically alter the properties and growth of the layer. Here we first consider boundary layers that, for whatever reason, are dry enough that water does not condense. We also assume that,

as is typical over the oceans, there is a positive buoyancy flux that is strong enough that the Obukhov length is less than the overall depth of the boundary layer.

The presence of another important length scale, the overall boundary layer depth, formally obviates the dimensional arguments that underlie the Monin-Obukhov scaling. In particular, the largest turbulent eddies span the whole boundary layer, so that, for example convection may have a larger effect on the surface winds than would be the case with the Monin-Obukhov scaling. Nevertheless, observations show that the scaling relations work quite well up to about the depth corresponding to the Obukhov length, L. This layer, in which thermodynamic variables and momentum may vary relatively rapidly, is known as the *surface layer*. Above that, the boundary layer turbulence is mostly generated by buoyancy, and the conserved buoyancy variable θ_v, is nearly constant with height. Other conserved variables, such as the water vapor mixing ratio r may have weak but detectable mean gradients, however, and therefore we prefer to refer to this layer as the *convective layer* rather than the more conventional term *mixed layer*. Near the top of the boundary layer there are usually sharp gradients in momentum and conserved variables; in this layer, called the *entrainment zone*, quiescent air from above the boundary layer has a strong influence. At the upper boundary of the entrainment zone, the transition from turbulence to quiescence may be quite abrupt. The overall structure of the clear boundary layer is illustrated by Figure 4.1, which also shows typical physical altitudes of the various transitions over tropical oceans.

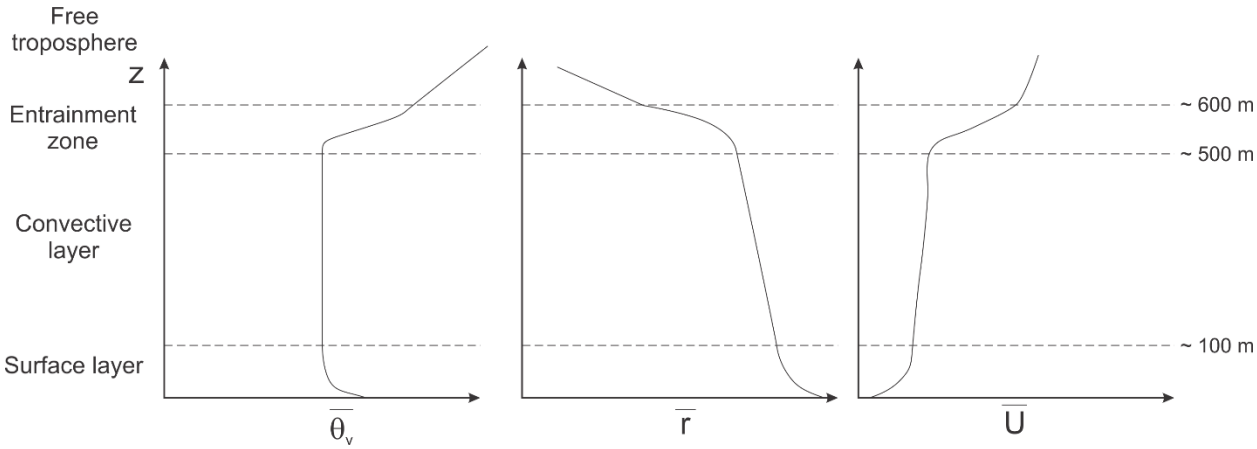


Figure 4.1: Schematic vertical structure of clear boundary layers over the ocean, showing mean virtual potential temperature (left), water vapor mixing ratio (middle) and wind speed (right). Typical altitudes of the transitions are shown at far right.

An important semi-empirical formulation of the entrainment velocity was developed by Lilly (1968), who assumed that some fraction of the turbulent kinetic energy generated by the surface buoyancy flux is used to increase the potential energy of entrained air. Put another way, the work done against buoyancy in entraining higher virtual potential air from above the boundary layer is provided by part of the positive buoyancy flux from the surface:

$$\overline{w_t' B_t'} = -a\overline{w_0' B_0'}, \tag{4.17}$$

where the term on the left represents the turbulent buoyancy flux at the top of the layer and the term on the right is the surface buoyancy flux multiplied by a constant a. From field measurements, a typical value of a is 0.2.

In turn, we can model the entrainment buoyancy flux as the product of an entrainment velocity, w_e, and the jump $\Delta\overline{B}$ in mean buoyancy across the entrainment zone:

$$\overline{w_t' B_t'} = -w_e \Delta\overline{B}. \tag{4.18}$$

Combining (4.17) with (4.18) yields a formulation for the entrainment velocity:

$$w_e = \frac{a\overline{w_0' B_0'}}{\Delta\overline{B}} = \frac{a\overline{w_0' \theta_{v0}'}}{\Delta\overline{\theta_v}}. \tag{4.19}$$

The second relation on the right of (4.19) results from expressing buoyancy in terms of virtual potential temperature.

The entrainment velocity together with the free tropospheric clear-sky subsidence rate determine the growth of the boundary layer depth:

$$\frac{\partial h}{\partial t} = w_e - w_{rad}, \tag{4.20}$$

in which h is the boundary layer depth, and w_{rad} is the clear-sky subsidence velocity above the boundary layer, defined positive downward. This latter quantity is related to the radiative cooling rate $\dot{Q}_{cool}$ by

$$w_{rad} = \frac{\dot{Q}_{cool}}{S}, \tag{4.21}$$

where S is the vertical gradient of dry static stability.

For a set surface buoyancy flux, the entrainment velocity decreases as the boundary layer depth grows. Referring to the left panel of Figure 4.1, note that the jump in $\overline{\theta_v}$ across the entrainment zone increases as the boundary layer depth increases, because the vertical gradient of $\overline{\theta_v}$ in the free troposphere is positive. Thus, from (4.19), w_e decreases. Eventually it will become as small as the radiative subsidence velocity and the boundary layer will cease growing.

Note that the surface virtual potential temperature flux in this conceptual model is positive while the turbulent flux of virtual potential temperature into the boundary layer from the top is negative; both of these act to increase the mean θ_v of the boundary layer. In equilibrium, this must be balanced by internal sinks, including radiative cooling and, under some conditions, horizontal advection of cooler air.

Also note from the right panel of Figure 4.1 that the mean wind usually has a shear across the entrainment zone. If this is such that the Richardson Number, defined

$$Ri \equiv \frac{\frac{g}{\theta_v}\frac{\partial \theta_v}{\partial z}}{\left|\frac{\partial \mathbf{V}}{\partial z}\right|^2} \tag{4.22}$$

is smaller than a critical value, then instabilities driven by the local wind shear can amplify and enhance the turbulent entrainment into the top of the boundary layer, causing it to grow faster and/or become deeper.

To be self-consistent, the solution for clear-sky boundary layers must be everywhere sub-saturated with respect to water vapor, otherwise clouds will develop and can dramatically alter the properties of the boundary layer, as explored in the next sections.

4.5 Stratocumulus-topped boundary layers

When air near the top of the boundary layer is moist enough to approach 100% relative humidity, clouds form. If the cloudy air is stable to reversible displacements above the top of the boundary layer, the cloudy air is trapped in the boundary layer, and depending on the particular circumstances, the clouds may form a nearly solid layer of stratocumulus or may break down into more isolated shallow cumulus clouds. Here we briefly review stratocumulus-topped boundary layers. Even though the clouds are relatively thin, they cover extensive areas of subtropical (and also arctic) oceans and play a large role in climate by contributing much of the planet's albedo. An example of boundary layer stratocumuli seen from above is provided in Figure 4.2.

The general structure of the cloud-topped boundary layer is similar to that of clear boundary layers (see Figure 4.3), but there are important differences. First, the relevant conserved variables are the total water mixing ratio r_T, which is conserved provided there is no precipitation, and the *liquid water density potential temperature*, defined

$$\theta_{\rho l} \equiv T_\rho \left(\frac{p_0}{p}\right)^{R_d/c_p} \exp\left(\frac{-L_v r_l}{c_p T}\right), \tag{4.23}$$

Figure 4.2: Boundary layer stratocumuli off Mexico's Baja California coast, on September 12th 2018, as seen from satellite. The northwesterly flow past Isla Guadaloupe is creating a vortex street to the southeast of the island.

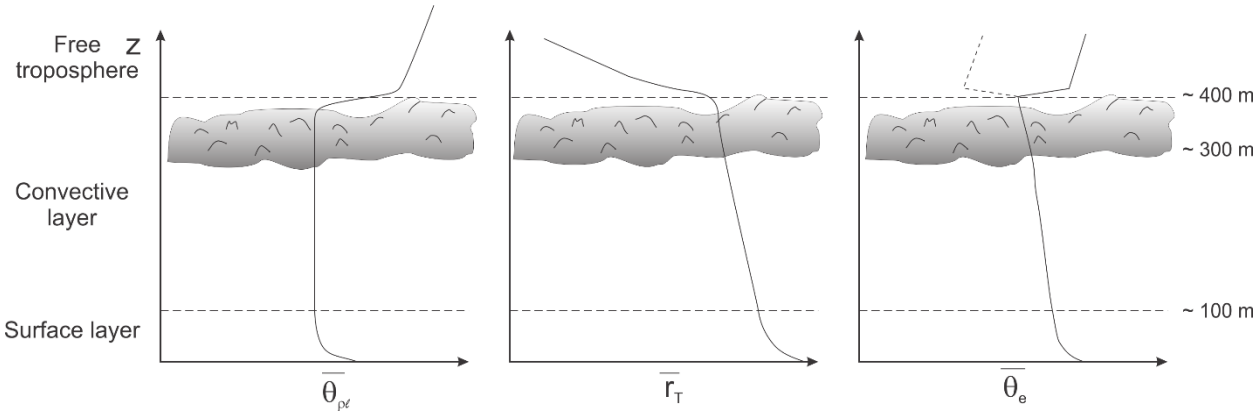


Figure 4.3: As in Figure 4.1 but for stratocumulus-topped boundary layers. The left panel shows the liquid water virtual potential temperature, the center panel shows the total water mixing ratio, and the right panel shows the equivalent potential temperature. The dashed line in the right panel shows an alternative θ_e profile above the boundary layer.

where T_ρ is the density temperature defined by (3.9) and r_l is the liquid water mixing ratio. Note that (4.23) reduces to the virtual potential temperature in the absence of liquid water. Also note that in Figure 3.2 we have omitted the entrainment layer, which tends to be very thin in stratocumulus-topped boundary layers.

Another conserved variable, absent radiative heating, is the equivalent potential temperature, θ_e, a typical profile of which is shown in the right panel of Figure 4.3. For reasons we will discuss shortly, two alternative profiles of θ_e are shown above the boundary layer.

The tops of stratocumulus layers are usually marked by abrupt changes in conserved variables and in momentum, but these variables are continuous through cloud base. The stratocumulus, if they are thick enough, may produce drizzle, which can play a role in transitioning the layer toward a trade cumulus boundary layer, as will be discussed in the next section.

In addition to the presence of clouds and possible drizzle, there are two aspects of cloudy boundary layers that distinguish them from other kinds of boundary layers. Most importantly, the large infrared opacity of the clouds gives rise to a strong jump in infrared flux, producing a spike in radiative cooling at the cloud top. This cooling at the layer top is an important source of convective turbulence in the boundary layer. The clouds absorb solar radiation during the day, and this heating may offset the infrared cooling enough to thin and/or cause breaks in the clouds.

Moreover, if the overlying atmosphere in sufficiently dry and not too warm, then the jump in θ_e across the top of the layer may be negative, a possibility illustrated in Figure 4.3. When this happens, air mixed into the boundary layer from above will have a lower value of θ_e than the surrounding cloudy air, and if enough liquid water evaporates into the mixture to saturate it, it will have a negative temperature anomaly and may be negatively buoyant. (In contrast, air entrained into clear boundary layers will always be positively buoyant.) This can cause the cloud top to be unstable; this is known as *cloud top entrainment instability* and enhances the production of turbulence in the layer.

Advective drying of the free troposphere above stratocumulus-topped boundary layers will enhance the aforementioned instability and also the jump in infrared flux across clouds top, both of which will increase the turbulence kinetic energy of the boundary layer. On the other hand, the jump in buoyancy across the boundary layer top is likely to increase, and these act in opposite directions on the entrainment velocity (see (4.19)). The drying will also be associated with a reduction in the radiative cooling of the free troposphere, reducing the clear-sky subsidence velocity. From (4.20) the layer may either deepen or become shallower, depending on the relative importance of the processes. All other things being equal, a deeper boundary layer would be more likely to have clouds at its top, but on the other hand the air being entrained from the free troposphere may be dryer, working in the opposite direction. This illustrates the complexity of the response of stratocumulus layers to changing environmental conditions.

On time scales of a few years and greater, stratocumulus-topped boundary layers, by shielding the surface from sunlight and thereby keeping it cold, may be self-preserving.

Over warmer surfaces, stratocumulus-topped boundary layers may transition toward layers containing scattered cumulus clouds. The nature of this transition is explored in the next section.

4.6 Transitioning boundary layers

Stratocumulus-topped boundary layers generally occur over cold water, often resulting from upwelling in the ocean. As boundary layer air moves westward and equatorward following the general flow of the trade winds, it encounters increasing sea surface temperatures, resulting in stronger enthalpy and buoyancy fluxes from the surface. The cloud-topped boundary layer grows deeper, entrainment increases, and the higher temperatures and deeper boundary layer yield higher liquid water concentrations in the clouds, increasing the probability and intensity of drizzle. The higher θ_e of the boundary layer makes more likely a negative jump in θ_e across the boundary layer top.

The depletion of net water concentration by drizzle and further depletion by increased entrainment from above lead to drier downdrafts in the cloud layer, and consequently to higher lifted condensation levels in the downdrafts. At levels between the LCLs of updrafts and downdrafts, the air is likely to have a moist adiabatic lapse, which is stable to the downward-moving unsaturated air. The asymmetry between nearly neutral stability to saturated updrafts and strong stability to unsaturated downdrafts concentrates the updrafts and broadens the downdrafts.

The result of this asymmetry is the development of cumuliform clouds in the layer between the LCLs of the upward moving air from the surface layer and the downward moving air from the top of the boundary layer, as illustrated on Figure 4.4. The stratocumulus layer becomes broken and partially decoupled from the surface, with the cumulus clouds being the agents of coupling. As illustrated in the figure, there are now four distinct layers: The surface layer, the subcloud layer (which includes the surface layer), the convecting layer, and the stratocumulus layer.

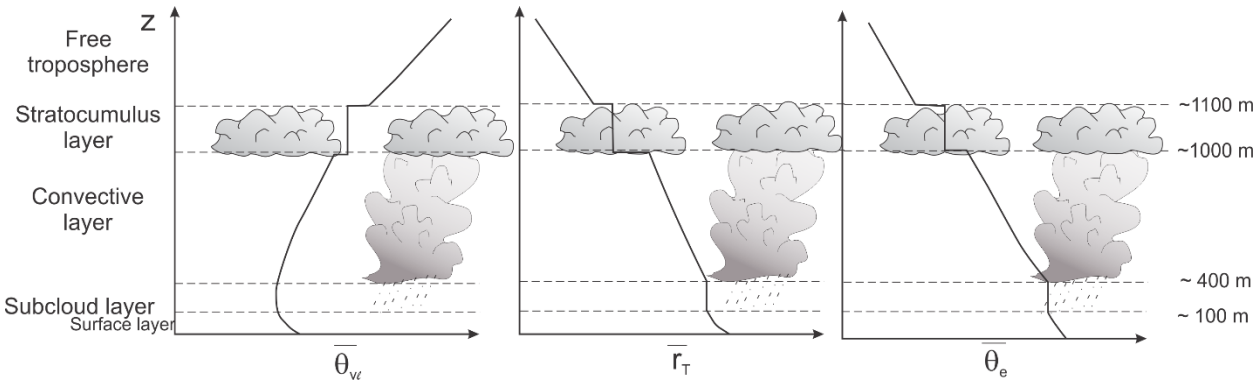


Figure 4.4: Vertical structure of a stratocumulus-to-cumulus transitioning boundary layer. Liquid water virtual potential temperature is shown at left, the total water mixing ratio in middle, and the equivalent potential temperature at right. The subcloud layer consists of a surface layer and an unsaturated convecting layer driven by the surface buoyancy flux. Cumulus clouds extend though the convecting layer and penetrate the stratocumulus layer, which is internally well mixed.

The subcloud layer above the surface layer is generally well mixed in the vertical, as is the stratocumlus layer, but the convecting layer is usually stably stratified to dry convetion and has slowly decreasing total water concentration with height. There may or may not be small jumps in thermodynamic properties across the base of the starocumulus layer.

In this configuration, the stratocumuus layer is supplied with water by the cumulus clouds, which extend vertically into the layer and may ocassionaly overshoot a small distance into the free

troposphere. The elevated stratocumulus layer loses water mostly by entraining dry air from above but sometimes also by drizzle, and the cumuli many be lightly precipitating. Radiative cooling of the stratocumulus tops is an important source of turbulence in the elevated cloud layer.

Figure 4.5 is a photo of a transitioning boundary layer taken from the surface.

Figure 4.5: Showing a transitioning boundary layer, with shallow cumuli penetrating a stratocumulus layer that is otherwise decoupled form the surface.

Transitioning boundary layers are widespread in the subtropics, and the fractional area covered by clouds can vary widely from around 10% to nearly 100%, so are they also important in determining local and plantary albedo.

4.7 Trade cumulus boundary layers

Over warmer water but away from regions of deep moist convection, transitioning boundary layers give way to yet deeper boundary layers populated by shallow cumulus clouds unaccompanied by extensive stratocumulus layers near their tops. Since these clouds are frequent in the trade wind[2] belts, they are often referred to as “trade cumuli”. They are sometimes weakly precipitating, and anyone who has spent time in the tropics will be familiar with the sudden but brief showers they may produce. Figure 3.2 shows a nice example of a trade cumulus cloud.

[2] The “trade winds” are belts of easterly winds found through much of the tropics and are so called because they were used in the Age of Sail to facilitate maritime trade mostly between Europe and the East and West Indies. We will explore their underlying dynamics in Chapter 5.

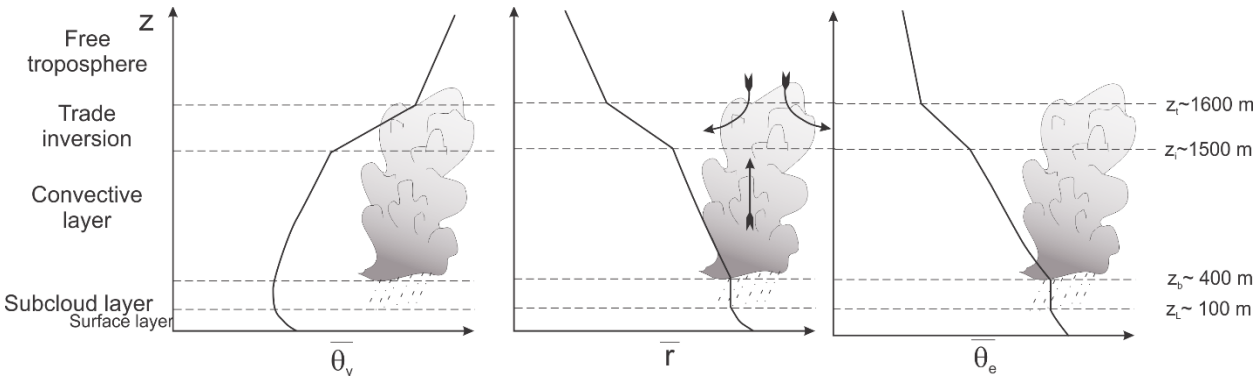


Figure 4.6: Vertical thermodynamic structure of a trade cumulus boundary layer, showing mean virtual potential temperature (left), water vapor mixing ratio (center) and equivalent potential temperature (right). The middle panel also shows typical air trajectories in the cumulus clouds.

The thermodynamic structure of trade cumulus boundary layers is illustrated in Figure 4.6. The layer bearing cumulus clouds is stably stratified and is capped by a "trade inversion" in which the virtual potential temperature increases sharply, giving way to the more modest increase with height typical of the tropical free troposphere. Likewise, water vapor mixing ratio and θ_e decrease with altitude in the convective layer and often more sharply in the inversion.

Entrainment from the free troposphere is an essential process in trade cumulus layers, as in other boundary layers, but much of the entrainment occurs into the tops of the cumuli, even though they occupy a relatively small fractional area. As the cumuli poke upward into the inversion layer, they turbulently entrain air from the free troposphere. This air is typically very dry, and since the liquid water virtual potential temperature (defined by (4.23)) is smaller in the clouds than in the free troposphere (unless the clouds are strongly precipitating), any mixture that contains little or no liquid water will have a density temperature that is smaller than that of the cloud, and will thus be negatively buoyant. This cloud top entrainment instability results in fairly large entrainment into the tops of the clouds, and the mixtures typically detrain into the inversion. Over the lifetime of the cloud, there is a net downward cumulus mass flux in the trade inversion, while there is a net upward flux in the convecting layer, as illustrated in the middle panel of Figure 4.6.

For a strictly non-precipitating cloud, there can be no net latent heat release over the lifetime of the cloud, and this, together with the virtual potential temperature structure of the trade cumulus boundary layer, constrains the relative magnitudes of the downward mass flux at the clouds top and the upward mass flux in the cloud layer. Let the former be defined as M_d (defined positive downward) and the latter as M_u. Then

$$\int_{z_b}^{z_i} \rho M_u \frac{\partial \theta_v}{\partial z} = \int_{z_i}^{z_t} \rho M_d \frac{\partial \theta_v}{\partial z}, \tag{4.24}$$

where z_b, z_i, and z_t are the respective altitude of cloud base, the base of the inversion, and the top of the inversion (see Figure 4.6). The absence of latent heating also means that the radiative cooling of the cloud layer must be compensated by warming by large-scale subsidence. This shows that there must be substantial net downward motion in the inversion

layer; indeed, the cloud-top cooling maintains the strong virtual potential temperature gradients across the inversion.

But trade cumuli often *do* precipitate. It is important to recognize, from (4.23), that *precipitation is a source of liquid water density potential temperature*. When the cumuli precipitate, they gain $\theta_{\rho l}$ and this decreases the degree of cloud top detrainment instability because the mixtures will not be as negatively buoyant, owing to less condensed water available for evaporation. With lowered entrainment, the depth of the boundary layer will decrease (see 4.20), and as the clouds become shallower, they become less likely to precipitate. Thus there may be a self-regulating mechanism in play that keeps the cumuli, on average, just barely precipitating, and holds the depth of the layer such that the clouds are just barely deep enough to develop precipitation.

The evaporation of rain falling from trade cumuli cools the air and the cold pools spread out as density currents. The interaction of these cold pools with surrounding air, with the surface, with ambient wind shear and with convection can result in intricate and varied forms of mesoscale organization, as is apparent in Figure 3.3. The physics governing this organization is only now beginning to be understood.

While the clouds in trade cumulus boundary layers occupy a small fractional area relative to stratocumulus and transitioning boundary layers, they still contribute non-negligibly to planetary albedo as they occupy large areas of the tropics and sub-tropics, and both the clouds and the humidity of these layers have important influences on radiative transfer.

4.8 Summary

In the tropics, boundary layers are defined, as elsewhere, as layers in which turbulent exchange between the atmosphere and underlying surface is important, and the turbulence can include convective clouds. These may occupy between 5% and 20% of the horizontal area of the atmosphere and may contain much of its net water content. They are important in both radiative and turbulent exchanges. Boundary layers may be clear, topped with stratocumulus clouds, or populated with trade cumuli, but large regions of the subtropics are in a state of transition between stratocumulus and trade cumulus regimes. These transitions are important to understand and predict and they greatly affect albedo. The currently poor ability of global climate models to simulate stratocumulus layers and transitioning boundary layers is considered a major (and perhaps *the* major) source of uncertainty in climate projections.

Figure 4.7 shows a satellite-derived snapshot of clouds over the eastern and central North Pacific with an attempt to visually delineate the various kinds of boundary layers discussed in this chapter. It is not always easy to identify cloud types from satellite imagery; for example, it may be difficult or impossible to see cumulus growing upward into a stratocumulus layer. Nevertheless, the diagram shows that the regions occupied by deep convection (and these are liberal estimates) cover a small region compared to the areas in which most clouds are in the boundary layer. Note also that even the trade cumulus regimes are noticeable brighter than the clear regions, indicating that the albedo contributed by the cumuli is not negligible.

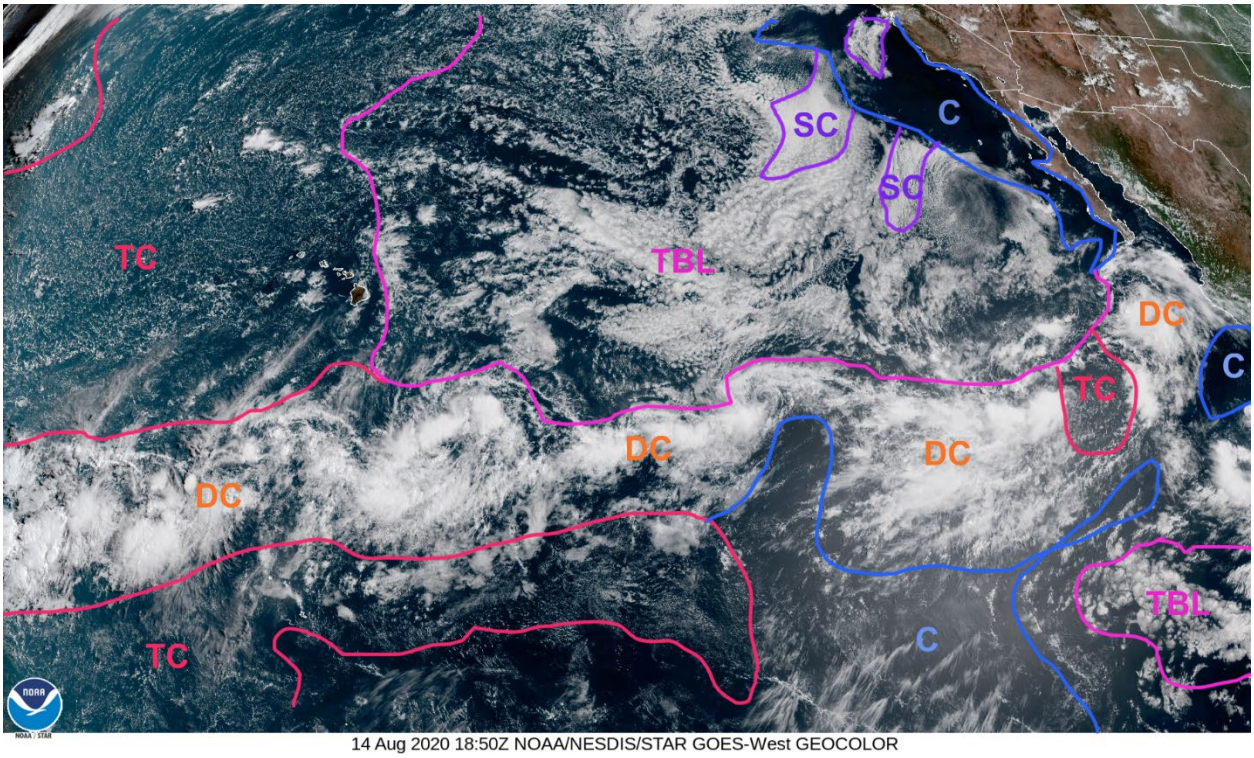


Figure 4.7: Visible satellite imagery from 18:50 GMT on 14 August 2020, showing clouds over the eastern and central subtropical North Pacific. Colored overlays are subjective estimates of the various regimes: "C" for "clear", "SC" for "stratocumulus-topped boundary layers", "TBL" for "transitioning boundary layers", "TC" for "trade cumulus boundary layers", and "DC" for "deep convection".

A schematic cross-section showing the progression of regimes as one travels from cold water upwelling regions near coastlines, bounding the eastern sides of oceans, down into the deep tropics is presented in Figure 4.8.

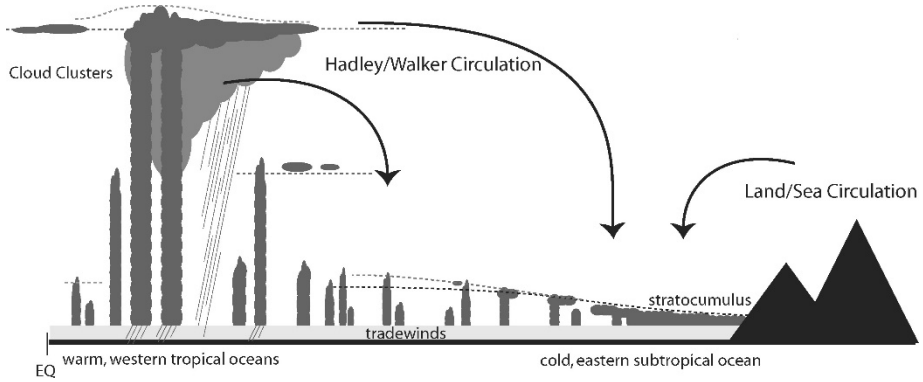


Figure 4.8: Showing the transition of boundary layer regimes from stratocumulus layers over cold, subtropical eastern oceans (right) to deep convective regimes over warm tropical oceans (left).

Travelling from cold ocean water on the right to very warm tropical oceans on the left, the boundary layer deepens from a few hundred meters to a few kilometers and transitions from a completely overcast stratocumulus-topped layer to a trade cumulus boundary layer and finally to a relatively narrow deep convective regime.

Boundary layers are interesting but complex features of the tropical atmosphere and much of what we will discuss in forthcoming chapters depends on knowledge of their behavior. We have only skimmed the surface of their physics here, omitting important contributions such as planetary rotation and horizontal advection, and they are prominent subjects of ongoing research. The reader is encouraged to consult more thorough treatment of boundary layers such as the classic text by Stull (1988) and a recent and comprehensive review (LeMone et al., 2019).

References

Businger, J. A., and Y. I. J. C. Wyngaard, and E. F. Bradley, 1971: Flux profile relationships in the atmospheric surface layer. *J. Atmos.. Sci.*, **28,** 181-189.

Dyer, A. J., 1974: A review of flux-profile relations. *Bound. Layer Meteor.*, **1,** 363-372.

LeMone, M. A., and Coauthors, 2019: 100 years of progress in boundary layer meteorology. *Meteorological Monographs*, **59,** 9.1-9.85, doi:10.1175/amsmonographs-d-18-0013.1.

Lilly, D. K., 1968: Models of cloud-topped mixed layers under a strong inversion. *Quart. J. Roy. Meteor. Soc.*, **94,** 292-309.

Monin, A. S., and A. M. Obukhov, 1954: Basic laws of turbulent mixing in the surface layer of the atmosphere. *Tr. Akad. Nauk. SSSR Geophiz. Inst.*, **24,** 163-187.

Stull, R. B., 1988: *An introduction to boundary layer meteorology .* Kluwer Acad. Publ., Boston, 666 pp. pp., translator.

5 Large-scale Circulations in the Tropical Atmosphere

Up until now we have treated the tropical atmosphere mostly as a one-dimensional entity, with each vertical column in a state of radiative-convective equilibrium. But the tropical atmosphere is in constant motion, and the modulation of clouds and precipitation, as well as transport of heat, moisture and momentum by large-scale circulations is an important aspect of the physics of the tropical atmosphere. Here, after a brief overview, we tackle the dynamics of large-scale tropical circulations.

5.1 Overview of large-scale tropical circulations

Compared to the highly variable weather of middle and high latitudes, tropical weather is relatively steady and mostly benign, with the stunning exception of tropical cyclones. The steadiness of tropical winds in particular is owing, in part, to the absence of appreciable baroclinic instability, the primary driver of synoptic scale variability at higher latitudes. In many regions within the tropics, trade winds blow from a nearly constant direction at almost constant speed, as discovered by 15th-17th century mariners, who established trade routes mostly between Europe, the Americas, and the Far East. They quickly discovered how to use the trade winds to optimize sea voyages and in so doing recorded observations of wind and weather, enabling the British scientist Edmond Halley (1656-1742), who had predicted the return of the comet later named after him, to construct the first know weather map (Halley, 1686), reproduced here as Figure 5.1. This is an example of a composite map that is based on observations made at widely disparate times, in contrast to synoptic maps made using nearly simultaneous observations that only became possible toward the end of the 19th century.

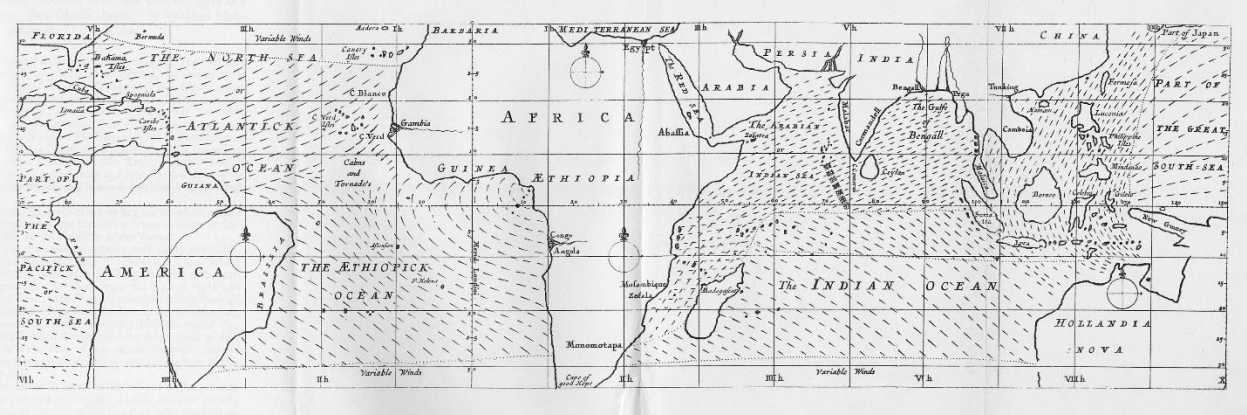


Figure 5.1: Edmond Halley's 1686 map of the trade winds, depicted as dashed lines.

Halley's paper contains detailed descriptions of tropical airflow, and his map clearly shows the northeasterly and southeasterly trades of the two hemispheres, converging along a near-equatorial discontinuity known today as the Intertropical Convergence Zone (ITCZ). Comparison with monthly mean winds based on far more plentiful and recent data (Figure 5.5) attests to the perspicacity of Halley's analysis.

Before reviewing modern analyses of tropical airflow and weather, we pause to consider the observational basis of such analyses.

5.1.1. Observations

Up through the early 20th century, observations of the atmosphere consisted mostly of measurements made at the surface by both professional and amateur observers, and by mariners, as recorded in ship's logs. Some inferences about conditions above the surface were made by recording weather on mountains and by observing the motion of clouds, supplemented by the odd manned balloon flight beginning in 1783. Late in the 19th century, observations using instrumented kites became routine in some regions, but these were limited by weather conditions and in any event could only extend a few kilometers above the surface. With the widespread advent of aviation in the 1930s, temperature and pressure could be measured at altitude, though flight level winds could only be inferred from the drift of aircraft between navigational fixes. Like kites, these observations were limited in altitude and by weather. Yet the discovery that winds on synoptic and larger scales were nearly geostrophic, particularly above the boundary layer, enabled early 20th century meteorologists to deduce winds aloft from temperature and pressure measurements closer to the surface.

Beginning around 1909, scientists began to experiment with balloon-borne measurements using instruments that had to be retrieved from the balloons after their descent. By the late 1920s, primitive radios were used to transmit weather data to surface receivers, obviating the need to recover the balloons, and by the early 1930s the "radiosonde" observations were reaching the lower stratosphere. In 1937 the U.S. Weather Bureau (later renamed the National Weather Service) established a network of radiosonde stations launching balloons at regular intervals, and this was expanded in the 1940s owing to the war-related demands for better upper-air observations. The 1940s also saw the development of radio-direction finding techniques, replacing the manual tracking of sondes using theodolites, resulting in the radio-wind sonde, or "rawinsonde". Sonde networks were established throughout the world, and radio-direction finding was replaced by more advanced techniques, culminating in the GPS sondes in use today. Figure 5.2 shows the locations of sonde stations as of 2020.

Sondes provide measurements of wind, temperature, pressure, and relative humidity at very high vertical resolution, but they are snapshots and may not be very representative of regional conditions, particularly if they are sampling strong local features such as cumulus clouds or mountain waves. Sampling over oceanic regions and even some land regions, such as central Africa, is sparse. Yet, sondes themselves are relatively inexpensive, and automation has reduced the need for human labor, so they remain an economical means of sampling the atmosphere.

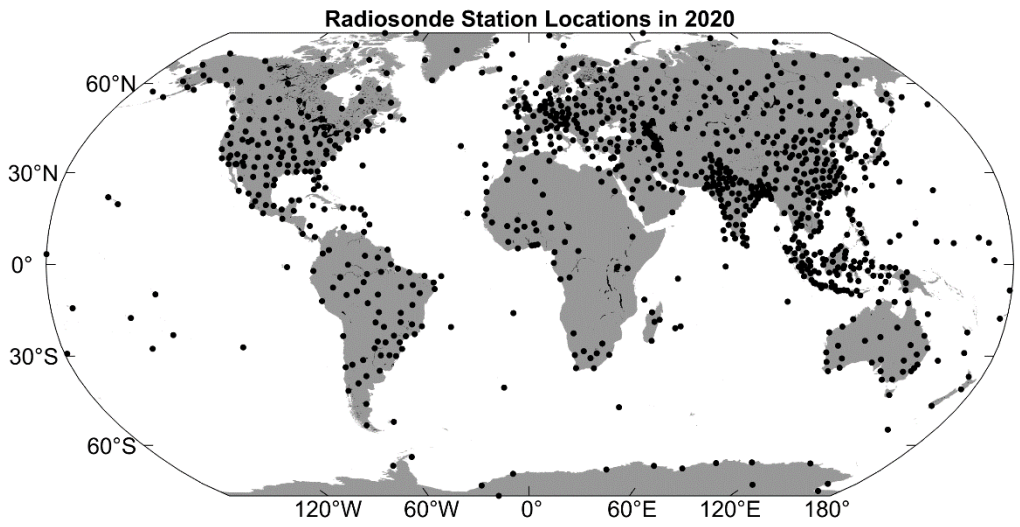


Figure 5.2: Locations of sonde stations in 2020. Sondes are generally launched twice per day, at 00 and 12 GMT.

Beginning in the 1960s, earth-orbiting satellites have been used as platforms for instruments designed to measure the atmosphere and oceans. Up though the late 1970s, almost all of these measurements were in the form of passive visible and infrared radiometric measurements – passive in the sense that the instruments were not transmitting signals but measuring reflected, scattered, and emitted radiation at various wavelengths. These passive measurements remain the backbone of space-based observations, allowing researcher and forecasters to detect and track features revealed by visible imagery, during the day, and infrared imagery at all hours. Because different greenhouse gases, clouds, and the surface itself emit and absorb radiation differently according to the wavelength of the radiation and temperature, multi-spectral infrared measurements contain information about the vertical structure of temperature, clouds, and greenhouse gases like water vapor, and information about surface temperature.

In 1978, NASA launched Seasat, which carried an active microwave scatterometer. This instrument sends pulses of radiation of wavelength comparable to the wavelength of capillary waves on water surfaces. By measuring the amplitude of the backscattered radiation, an estimate of capillary wave amplitude can be made, and this amplitude is directly proportional to wind stress on the water surface. By comparing amplitudes of the same foot print seen from two or more angles during the spacecraft's orbit, one can estimate the orientation of the capillary waves and thus the direction of the surface stress vector. Unfortunately, radiation at this wavelength is heavily absorbed by precipitation, so surface stress cannot be accurately estimated in regions of heavy rain.

The era of sea surface altimetry commenced in 1992 with the launch of TOPEX/Poseidon, a joint U.S.-France venture to monitor sea level and seasonal variations in oceanic circulation. Radar altimeters measure the time between transmission of a pulse of radiation and the reception of that pulse after reflection from the surface. State-of-the-art altimeters measure the surface with an accuracy of 4-5 cm. Knowledge of the slope of the sea surface relative to the geoid allows one to calculate geostrophic velocities.

Beginning in 1995, a suite of low-earth-orbiting satellites has been deployed to measure transmissions from GPS satellites as they rise and set relative to the low-orbit platforms. In

essence, the low-orbit satellites measure the bending of the GPS signal transmitted through the atmosphere, which depends on the refractive index, which in turn is a function of temperature, pressure, and water vapor. This “radio occultation” technique yields vertical profiles of temperature and water vapor, and the 24 GPS satellites together with the low-earth-orbiting satellites yield many soundings each day. Figure 5.3 shows the location of radio occultation soundings over a single 24-hour period.

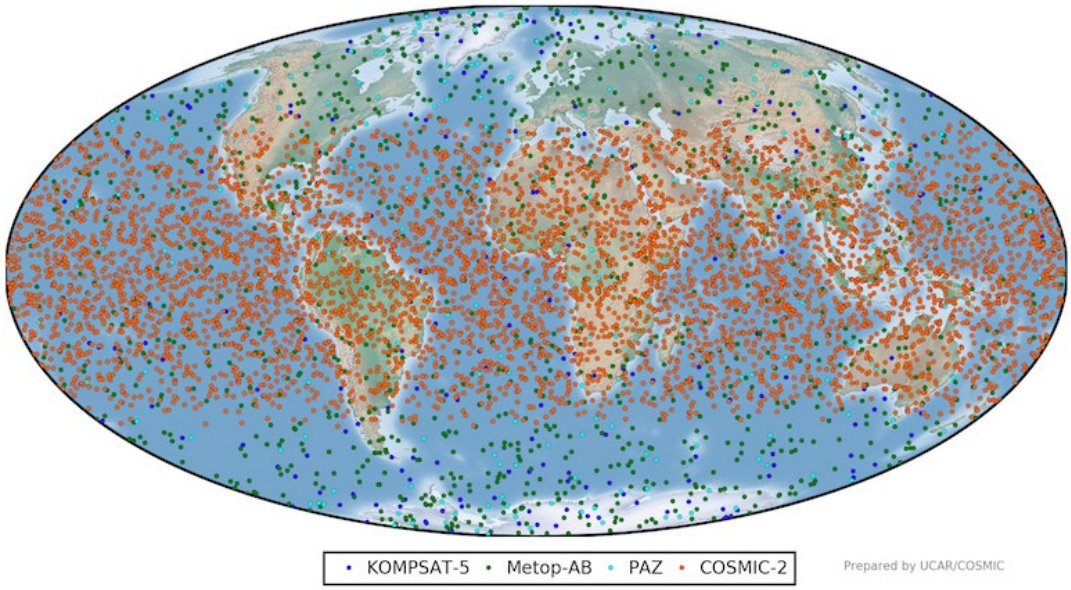


Figure 5.3: Location of radio occultation soundings over a 24-hour period. Note the especially dense coverage in the tropics.

From 1997 to 2015, NASA operated the Tropical Rainfall Measurement Mission (TRMM) research satellite to provide active radar measurements from space. The scanning radar transmitted pulsed microwave radiation downward and measured radiation backscattered from precipitation in the atmosphere. Some examples of TRMM measurements are provided in the next subsection.

Many commercial airlines as well as package delivery aircraft are equipped with sensors that measure temperature and humidity, and winds are deduced from the airspeed and GPS positions. This data is now routinely collected in real time and provides vertical profiles at airports and flight-level conditions around the world, focused in air traffic corridors.

On the ocean side, measurements of surface temperature and salinity have been made from ships for over a century, and in recent decades these have been augmented by space-borne passive radiometric measurements. Until fairly recently, properties of the subsurface ocean could only be deduced from expensive and laborious measurements made from ships. These have now been augmented by direct measurements using robotic sounders known as ARGO floats. The floats are mini-submarines, that descend from the surface to depths of 1-2 km, (in recent years a few can go as deep as 6 km), and then ascend back to the surface, making measurements of temperature and salinity along the way. These are then transmitted to satellites. The floats can be programmed to linger at a prescribed depth so that the ocean current velocity at that depth can be detected via the float’s net drift. Figure 5.4 shows the positions of ARGO floats on a particular day in January, 2021.

The ARGO floats, with the previously discussed sea surface altimetry, temperature, and salinity measured from satellite, augmenting traditional measurements from ships, constitute the basis of routine ocean measurement today.

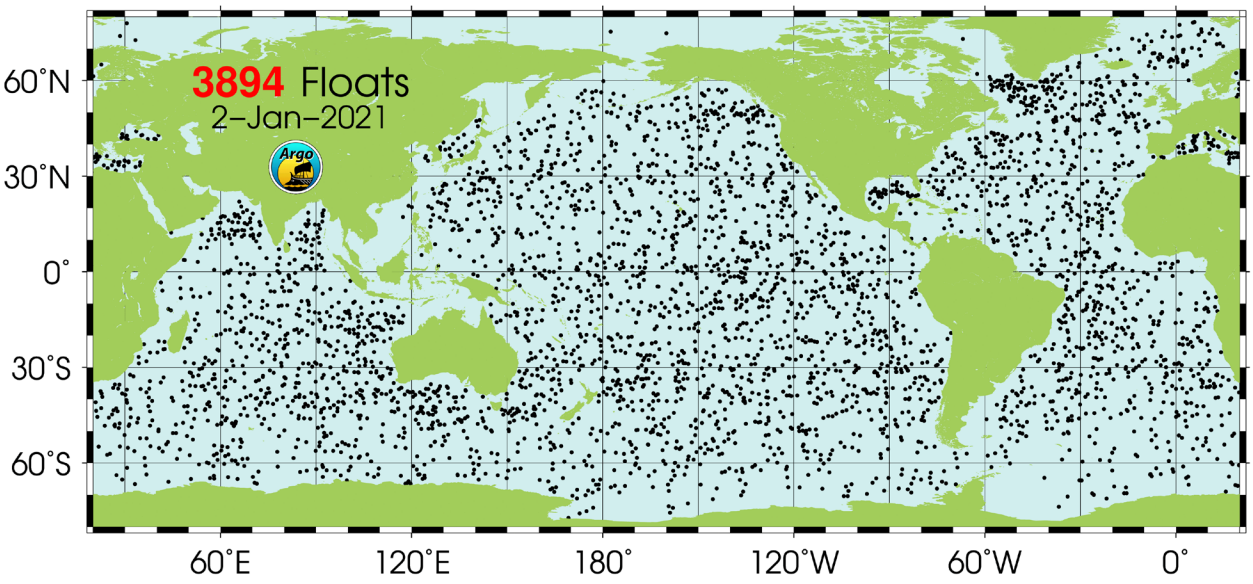


Figure 5.4: Locations of ARGO floats on January 2nd, 2021.

5.1.2 Analysis techniques

The aforementioned measurements form the basis of contemporary state estimates of the atmosphere and oceans. Optimal state estimation is a large and challenging enterprise, given the great disparity in measurement type and the spatial and temporal heterogeneity of the data. These state estimates are used as the initial conditions for numerical weather forecasts and also as the basis of long-term climatologies.

The topic of optimal state estimation is rich, complex, and mathematically challenging and cannot be done justice in a few paragraphs. Broadly, the approach taken today is fundamentally Bayesian, with a prior state estimate given by a previous numerical forecast, updated using various kind of filters that account for measurement and sampling errors and errors in the model used to provide the prior state estimate. Different numerical weather prediction centers use different types of filters, and what constitutes an optimal filter (defined as that which leads to the best subsequent numerical predictions) is a subject of ongoing research. An important advantage of contemporary state estimation is its ability to deal with highly disparate types of data. For example, satellite-based radiances are directly assimilated into numerical weather prediction models, rather than first being inverted to estimate vertical temperature profiles. When new data sources arise, it is fairly straightforward to incorporate these into these data assimilation schemes. Much of the improvement in the quality of weather forecasts in recent decades can be attributed to the advent of high-performance data assimilation schemes.

These data assimilation schemes can also be used to create optimal estimates of past states of the atmosphere and ocean; this enterprise is known as “reanalysis”. A good numerical weather prediction model is used to assimilate past observations over time to create a sequence of

atmospheric states, or “reanalyses”. More recently, the same strategy has been used to provide optimal state estimates of the ocean.

Most of these reanalyses, providing many variables on regular spherical grids, are freely and publicly available and are widely used in climate, atmospheric, and oceanic research. Their ready availability and ease of use has tempted many to regard them as representing the true state of the atmosphere and ocean, but it is important to recognize their limitations. For analyzing instantaneous states, the main limitation is the diversity of observation types and the often enormous spatial and temporal inhomogeneity of the observations. At every point in space and time, the reanalyses are influenced by model error as well as measurement and sampling errors. In locations and at times when measurements are sparse, the re-analyzed state may more represent the model physics than the real world.

Another, perhaps more serious problem arises when one tries to use reanalysis data to detect temporal trends in atmospheric or oceanic properties. The evolution over time of the sources of assimilated data introduce spurious trends in reanalyses. For example, it is well known that there are spurious jumps in reanalysis variables across the late 1970s owing to the introduction of satellite-based observations. One way around this particular problem is to assimilate only surface pressure and temperature observations, that have not evolved quite so radically, at least over the 20th and 21st centuries; this is the strategy used by NOAA’s 20th century reanalysis project (Compo et al., 2008). Such reanalyses have fewer problems with spurious trends but the instantaneous states are not as well represented as in conventional reanalyses, owing to their omission of upper-air data.

In what follows, we make heavy use of both direct satellite measurements and reanalysis data, and in interpreting these climatologies one should be aware of the limitations of both the measurements and the analysis techniques.

5.1.3 Large-scale climatology of the tropics

We begin where Edmond Halley left off, with a description of the large-scale, monthly mean surface airflow in the tropics. Figure 5.5 shows the monthly mean surface winds in January, April, July and October, compiled from reanalysis data on a 2.5-degree grid, from 1981 to 2010. Comparison with Halley’s map (Figure 5.1) attests to the quality of his 1686 analysis. Across the Pacific and Atlantic oceans, winds in the tropics are dominated by northeasterlies in the northern hemisphere and southeasterlies in the southern hemisphere. The lines of discontinuity in Halley’s analysis do not show up clearly in the monthly mean winds; we might speculate that Halley’s inference was based on the logs of individual ships which often did encounter abrupt wind shifts near that ITZC. But the position of such shifts, which move around, would naturally be smoothed out in monthly mean data.

The tropical South Indian ocean is dominated by southeasterly trade winds at all time of the year, but the North Indian Ocean winds are strongly seasonal, with northeasterlies in the winter and strong southwesterlies in summer. This is a signature of the Asian monsoon, about which more in section 5.3. Elsewhere, one can notice a gentle shifting of the trade winds towards the summer hemisphere and a strengthening of the winds in the winter hemisphere.

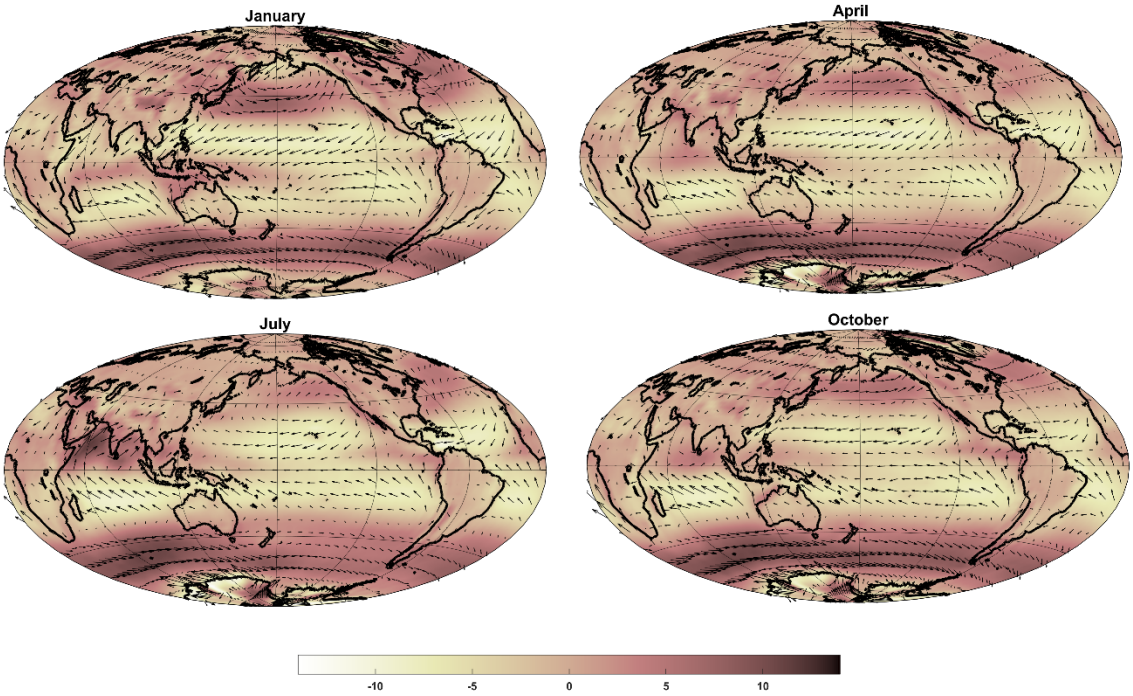


Figure 5.5: Monthly mean surface winds from NCAR/NCEP reanalyses averaged over the period 1979-2019. The shading indicates the wind speed (m/s) according to the scale at bottom, and the arrow directions and lengths correspond to the wind direction and speed, respectively.

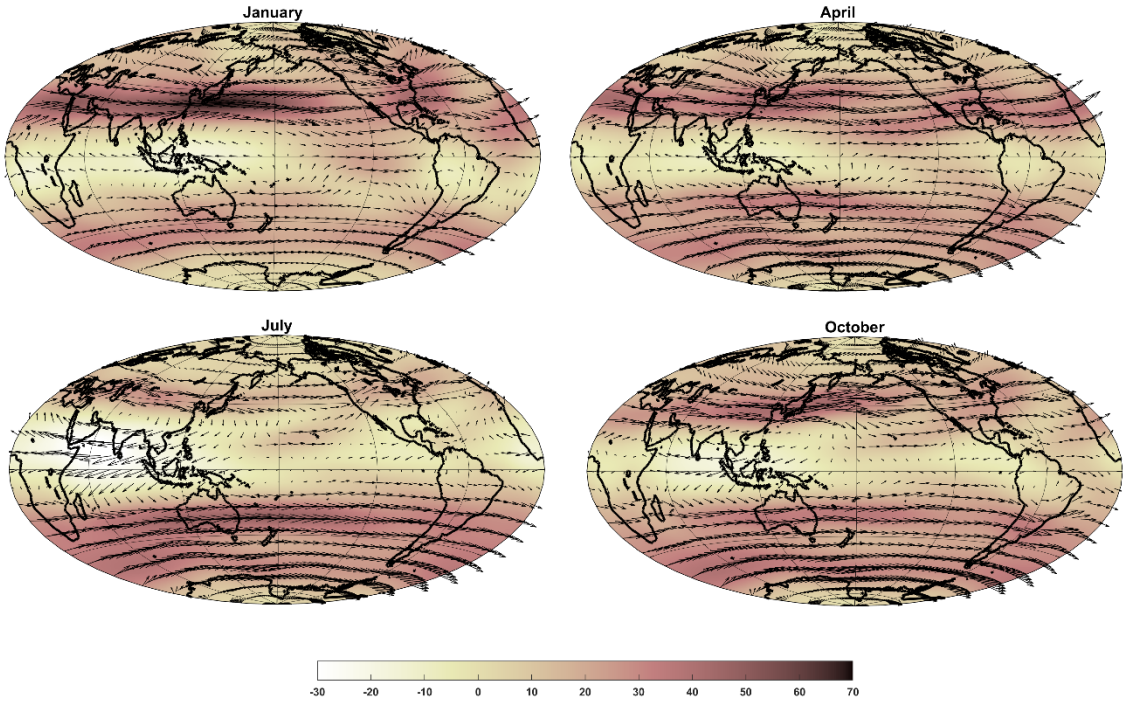


Figure 5.6: Monthly mean near-tropopause (150 hPa) winds from NCAR/NCEP reanalyses averaged over the period 1979-2019. The shading indicates the wind speed (m/s) according to the scale at bottom, and the arrow directions and lengths correspond to the wind direction and speed, respectively. Note that the wind scale differs from that in Figure 5.5.

Figure 5.6 shows monthly mean wind direction and speed at the 150 hPa level, which is near the tropopause in the tropics. The tropical region is bounded by strong westerly jets that extend from the subtropics well into middle latitudes. The jets are stronger and closer to the equator in winter. In the deep tropics, near-tropopause winds are generally light and from the east, but with some notable exceptions. In January, there is an extensive region of westerly wind in the eastern tropical Pacific, and this is bridged to the northern hemisphere subtropical and mid-latitude westerlies. This is important because it provides a conduit for nearly stationary Rossby waves to propagate into the tropics from higher latitudes in the northern hemisphere. This patch of westerlies retreats into the central subtropical North Pacific in July and migrates back toward the eastern equatorial Pacific during the boreal autumn.

Monthly mean virtual temperature at 600 hPa is displayed in Figure 5.7. Consistent with the weak temperature approximation (WTG), spatial as well as seasonal variations of virtual temperature are very small – generally less than 1 K – equatorward of 15 degrees. A most striking feature of these maps is the large positive virtual temperature anomaly over the Tibetan plateau in July.

Sea surface temperature (Figure 5.8) varies spatially somewhat more than does 600 hPa virtual temperature equatorward of 15 degrees. The warmest water is in the Indo-Pacific “warm pool”, whose latitude oscillates gently with the seasons. By contrast, the eastern Atlantic and South Pacific are quite cool even though the 600 hPa virtual temperature is not appreciably cooler than at other longitudes. As we shall explore later in this chapter, the anomalously cool water in these regions is owing to ocean upwelling and lateral heat transport, augmented by reduced absorbed solar radiation owing to stratocumulus clouds in these regions.

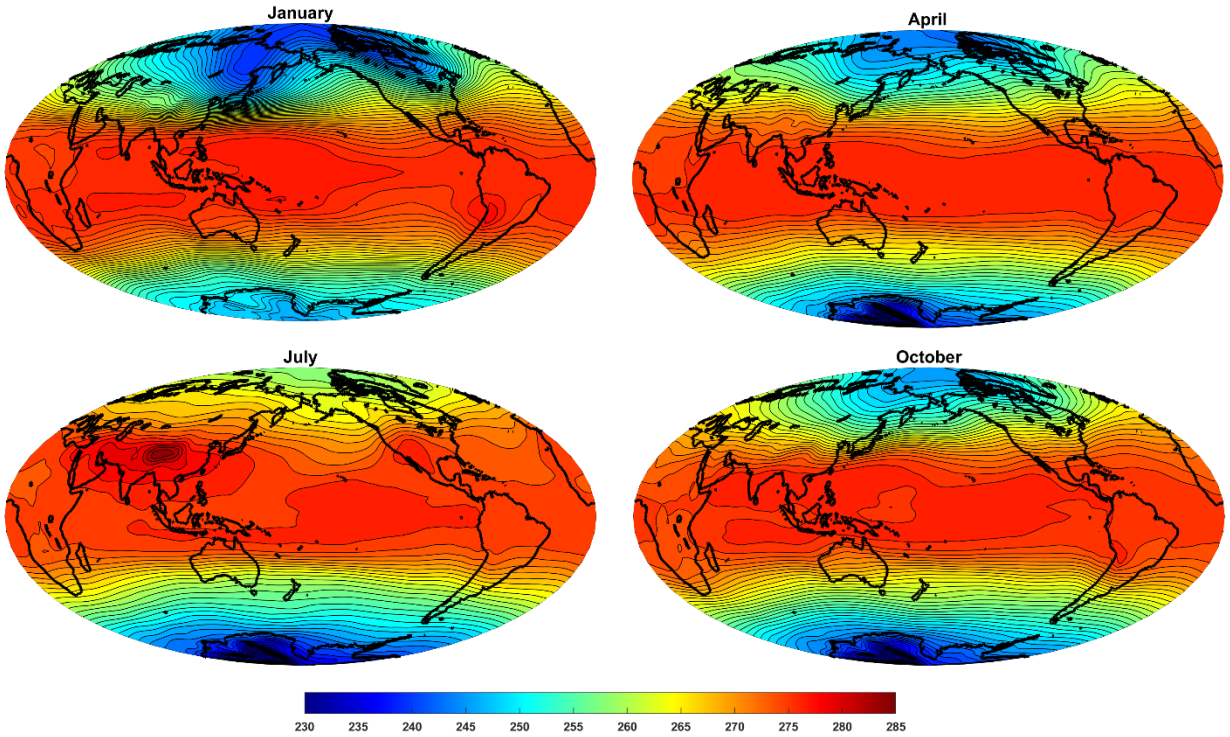


Figure 5.7: Monthly mean virtual temperature (K) at 600 hPa in January, April, July and October, constructed from NCAR/NCEP reanalysis data over the period 1949-2019. Contour interval is 1.25 K.

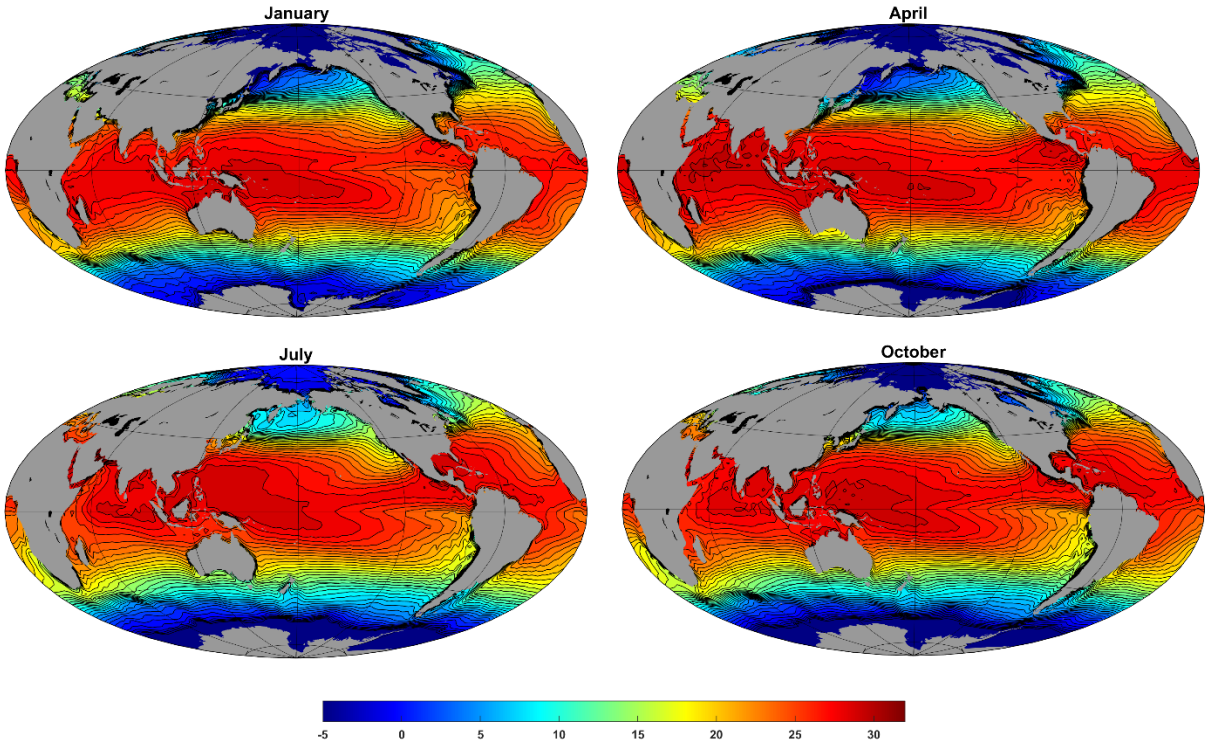


Figure 5.8: Monthly mean sea surface temperature (C) in January, April, July and October, constructed from NCAR/NCEP reanalysis data over the period 1949-2019. Contour interval is 0.84 C.

The distribution of monthly mean precipitation in January, April, July and October is shown in Figure 5.9, constructed mostly from satellite-derived precipitation measurements, particularly those made using the active radar onboard the TRMM satellite.

Focusing on the tropics, one of the first things one notices is the strong concentration of precipitation into an ITCZ that spans the Pacific and Atlantic oceans but becomes somewhat amorphous in the far western Pacific. This ITCZ nicely corresponds to where the trade winds converge (Figure 5.5). Another interesting feature is a wide band of enhanced precipitation extending southeastward from the far western equatorial Pacific. This is known as the “South Pacific Convergence Zone (SPCZ)”. The Asian monsoon is evident as a strong maximum of precipitation over the Western Ghat mountains of India, northern India, and the Bay of Bengal in July, coupled with a near absence of precipitation in those same regions in January. Less prominent monsoonal rainfall is also evident over extreme northern Australia and the Pacific coast of Mexico.

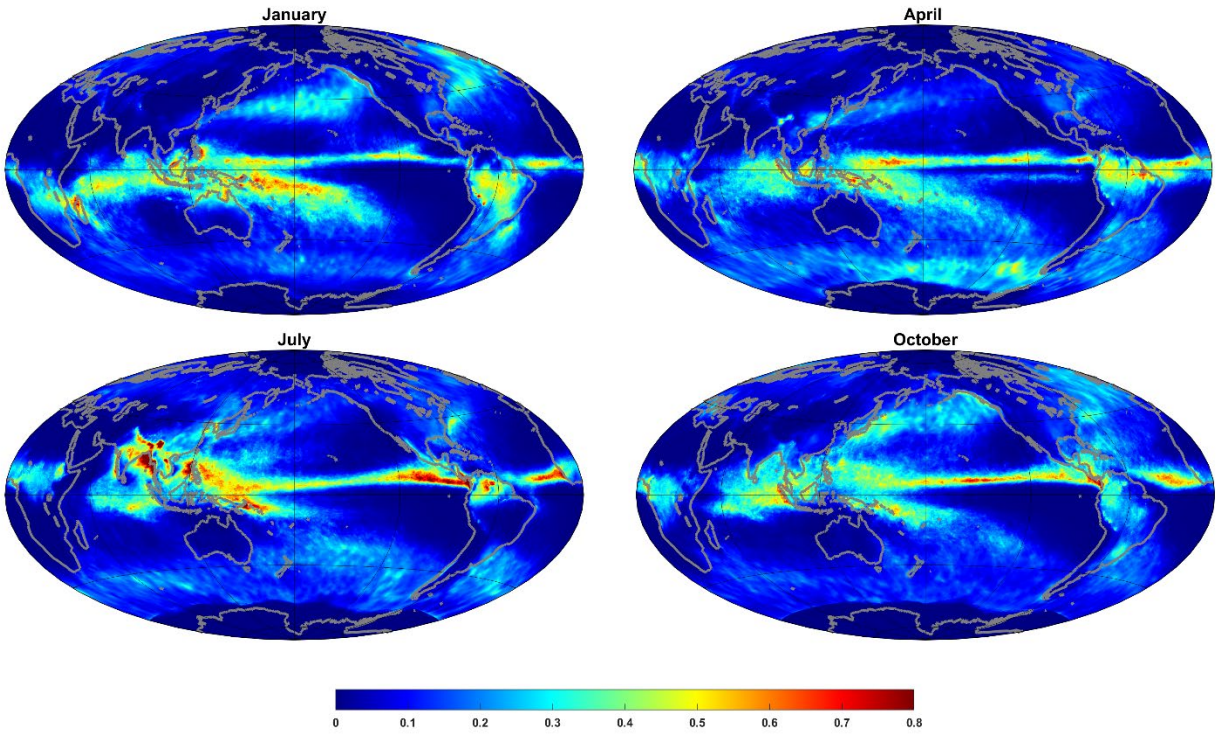


Figure 5.9: Monthly mean rainfall rate (mm/hr) in January, April, July, and October, from NASA's IMERG data based on a constellation of satellites augmented where possible with rain gauge data, 2002-2019. Measurements from the Tropical Rainfall Measurement Mission (TRMM) active precipitation radar are an important component of these synthesized data.

There are also prominent regions of heavy precipitation over equatorial Africa and South America that oscillate northward and southward with the seasons.

In general, the precipitation tends to follow the regions of highest sea surface temperature, but the relationship between SST and precipitation is far from one-to-one.

Also conspicuous are the large 'deserts' (regions of little or no precipitation) over the eastern subtropical oceans. The eastern Atlantic deserts extend well eastward over land in the northern hemisphere at all times of the year, and inland over southern Africa in southern hemisphere winter. The "maritime continent" of Indonesia is also a significant locus of heavy precipitation.

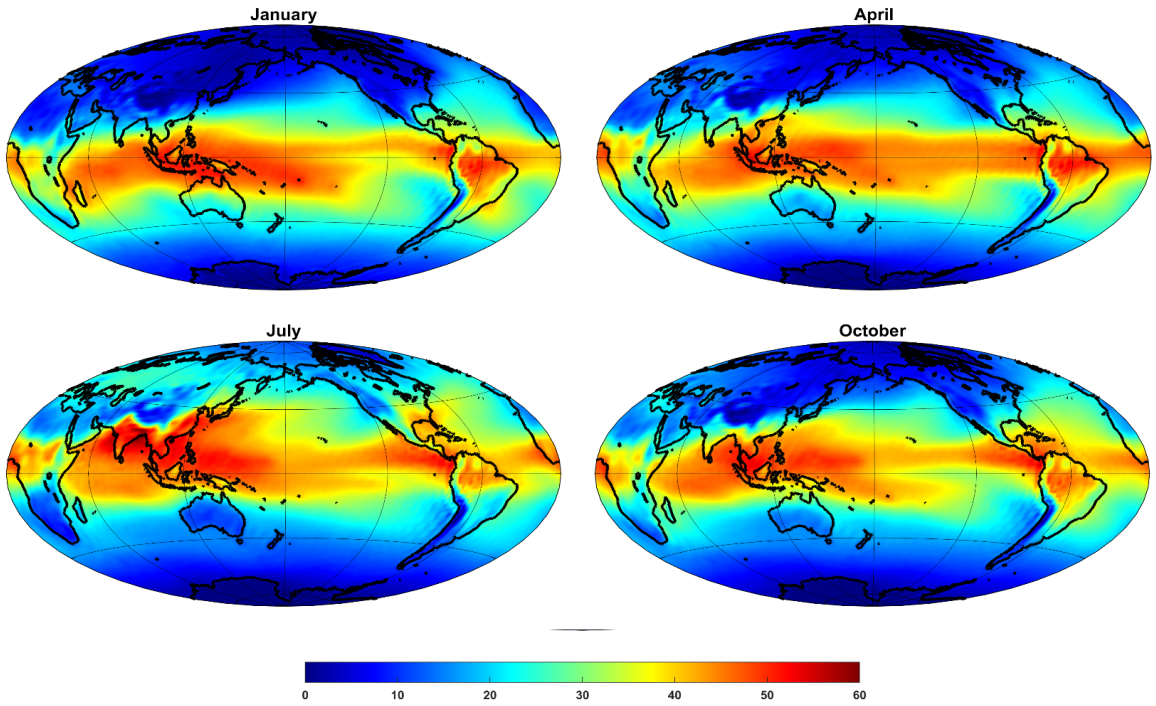


Figure 5.10: Mass-weighted, vertically integrated atmospheric water vapor content (kg m^{-2}) in January, April, July, and October, from NCAR/NCEP reanalyses, 1979-2019.

Most of the water vapor on our planet is within the tropics, as illustrated in Figure 5.10 which shows global monthly mean mass-weighted, vertically integrated water vapor averaged over three-month periods. Note that there is a reasonable correspondence between column water vapor and precipitation, but in many places, rainfall is far more concentrated than is the water vapor. As noted in Chapter 3, section 3.3.3c, there is a strong, observed statistical relationship between column water vapor and precipitation in the tropics (Figure 3.32) but the reader is reminded that one should be cautious about inferring causality from this statistical relationship; indeed, one can reproduce the general character of this curve by varying the surface enthalpy flux in a simple single-column model with boundary layer quasi-equilibrium and the weak temperature gradient approximation. In this case, the 'causal' agent is the varying surface fluxes, with both column water vapor and precipitation varying in response. The nonlinearity of the precipitation-column water vapor response helps explain why the precipitation is more focused than column water vapor in the climatology displayed in Figure 5.9.

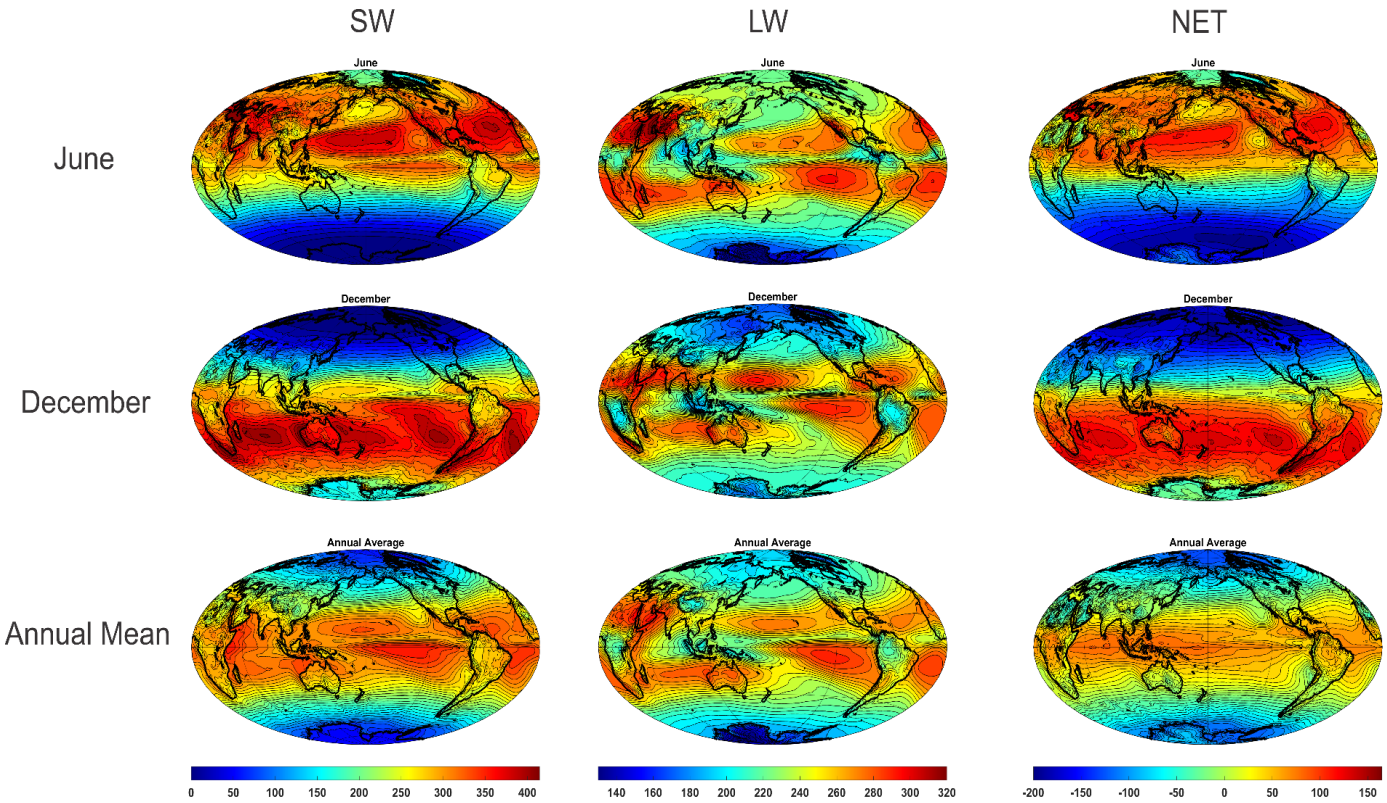


Figure 5.11: Top-of the-atmosphere (TOA) radiative fluxes (W m^{-2}): Net downward shortwave flux (left), net upward longwave flux (middle), and total net flux (right), for June (top), December (middle) and annual average (bottom). Data is from NASA's Clouds and the Earth's Radiant Energy System (CERES) program for the period 2001-2019.

Earth-orbiting and geostationary satellites have been measuring total outgoing longwave radiation, incoming solar radiation, and reflected solar radiation since the early 1980s, and much can be learned by examining these radiative fluxes. Figure 5.11 represents a compilation of total incoming solar radiation (incoming minus reflected shortwave) outgoing longwave radiation, and net incoming radiation (total incoming solar minus outgoing longwave). The total incoming solar radiation (left panels) reflects not just the latitudinal and seasonal distribution of top-of-the-atmosphere insolation, but albedo, which is mostly owing to clouds, and at high latitudes, snow, land ice, and sea ice. Once can clearly see, in the annual mean, the tropical maximum diminished by clouds in the ITCZ, over the tropical continents, and over cold ocean water in the subtropics just off the west coasts of continents.

The outgoing longwave radiation has peaks in broad subtropical belts where there are few high clouds and relatively small column water vapor. (These subtropical maxima are sometimes referred to as the planet's "radiator fins".) In the annual mean, there is some spatial correlation between outgoing and net incoming radiation in the tropics: where there are few clouds, there is little reflection of shortwave and longwave photons originate at higher temperatures lower in the atmosphere. But note that the seasonal variation of net incoming shortwave radiation greatly exceeds that of longwave radiation, accounting for most of the seasonal swing in net radiation. A strong north-south imbalance of net radiation occurs during the extreme seasons, and is

compensated partially by inter-hemispheric transport and partially by energy storage in the atmosphere, cryosphere and oceans as they heat up and cool off.

One curiosity of the distribution of annual mean net incoming radiation (lower right panel of Figure 5.11) is that it is somewhat more zonally symmetric than annual means of either the net shortwave or longwave radiative fluxes. This implies a degree of compensation between departures from the zonal means of shortwave and longwave fluxes. Indeed, detailed comparisons of clear sky and cloudy radiative fluxes shows that the large attendant anomalies of shortwave and longwave TOA radiation almost perfectly cancel. On the other hand, there clearly exists a net surplus of incoming radiation in the zonal, annual mean and this must be compensated by material fluxes of enthalpy away from the tropics by the atmosphere and/or oceans. The distribution of annual mean net radiative fluxes suggests that the annual mean enthalpy flux by the combined ocean-atmosphere system will also be approximately zonally symmetric.

Viewing monthly mean maps of various quantities, as we have been doing, can have the undesirable side effect of lulling one into thinking of the tropical atmosphere as having small variability on time scales of weeks or less. While day-to-day variability of winds, temperature, and pressures are indeed small compared to variability at higher latitudes, figures 5.12 and 5.13 remind us of this large sub-monthly variability in clouds and precipitation. While one can still identify long-term mean features such as the ITCZ, there is clearly much more going on. We will explore the nature of shorter-term variability in Chapter 7.

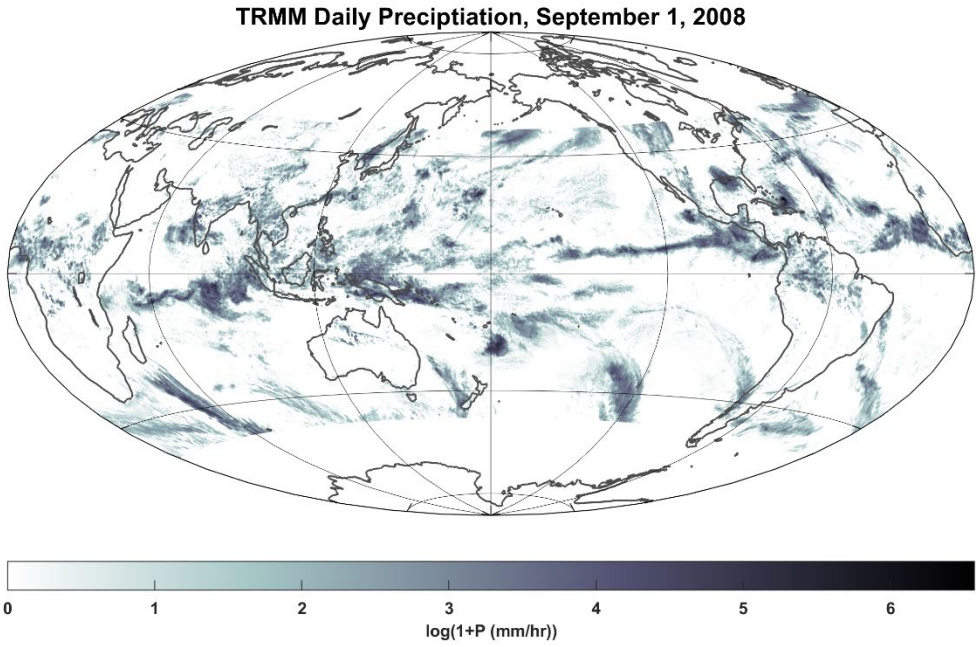


Figure 5.12: Distribution of daily precipitation measured by the NAS TRMM active precipitation radar on September 1, 2008, expressed as the natural logarithm of 1 + the precipitation rate in mm hr^{-1}.

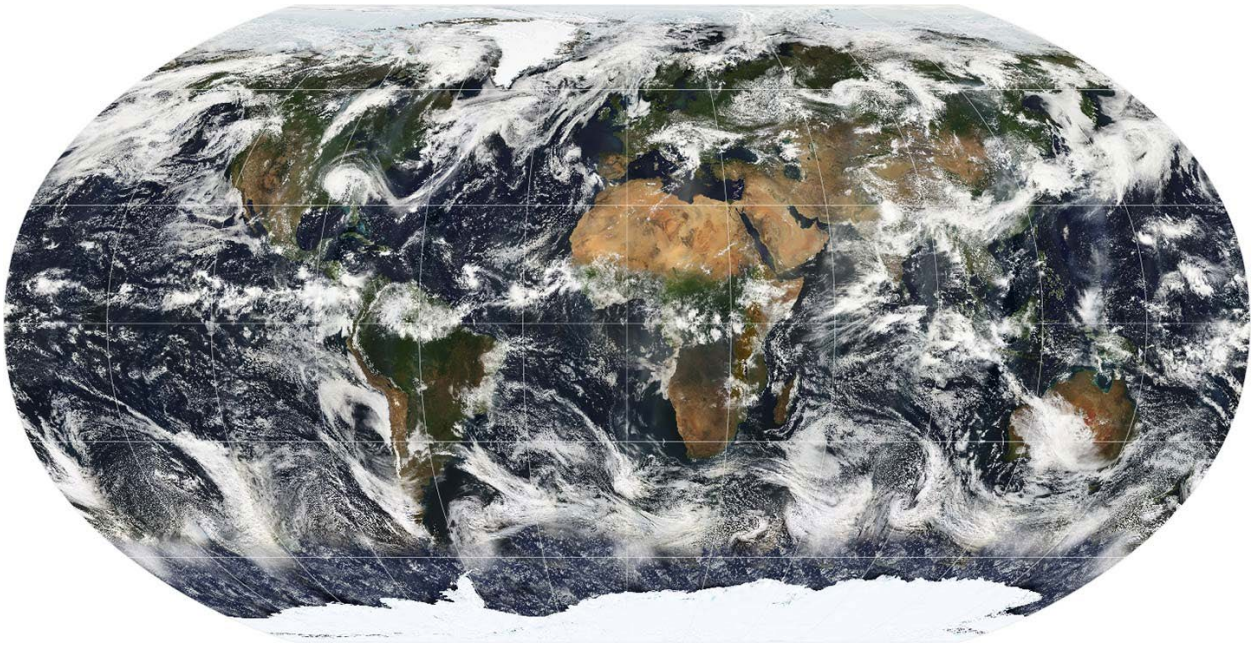


Figure 5.13: Global cloud cover on July 11th 2005 based largely on observations by the MODIS sensor on the NASA Terra satellite

5.1.4: Zonal mean, time-mean climatology

The near zonal symmetry of the annual mean net incoming radiation (Figure 5.11, lower right panel) implies that heat is exported by the atmosphere and oceans nearly symmetrically, though there is no requirement that transport in each medium be symmetric. Measurements show that the oceans and atmosphere export roughly equal amount of heat to higher latitudes, but here we focus on the zonally averaged atmospheric circulation. Figure 5.14 shows the mass streamfunction corresponding to the zonally averaged meridional flow. The streamfunction is defined so that the hydrostatic form of the mass continuity equation in pressure co-ordinates is satisfied:

$$\frac{1}{a\cos(\theta)}\frac{\partial}{\partial\theta}\big(v\cos(\theta)\big)+\frac{\partial\omega}{\partial p}=0, \tag{5.1}$$

where a is the radius of the earth, θ is latitude, p is pressure, v is the meridional wind component, and ω is the pressure velocity, dp/dt. We define a mass streamfunction, ψ, such that

$$\begin{aligned}\frac{\partial\psi}{\partial p}&=v\cos(\theta),\\ \frac{1}{a}\frac{\partial\psi}{\partial\theta}&=-\omega\cos(\theta).\end{aligned} \tag{5.2}$$

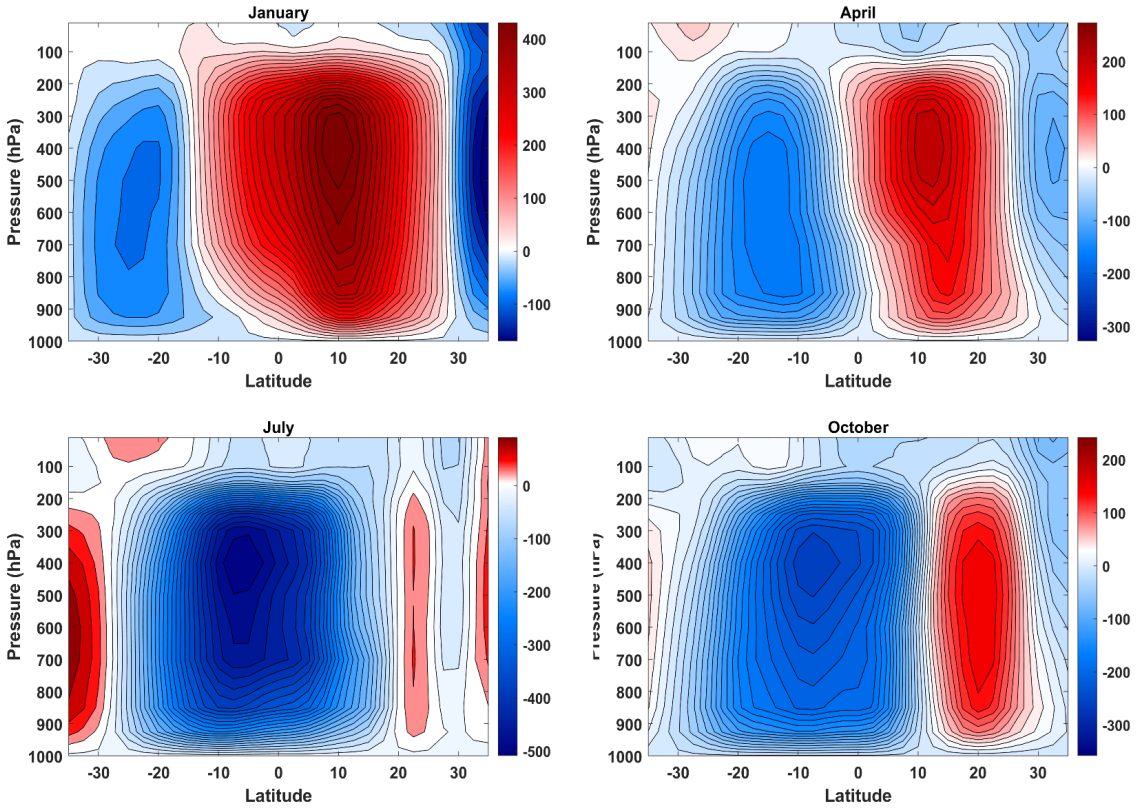


Figure 5.14: Zonal mean, monthly mean mass streamfunction (0.1 Kg m^{-1} s^{-1}) for January, April, July and October, calculated from NCAR/NCEP reanalysis data over the period 1979-2019. Circulation is clockwise around positive centers and counterclockwise around negative center. The contour interval in each plot is the same but the color scale differs for each plot, through it spans the same total interval.

Substitution of (5.2) into (5.1) verifies that mass continuity is satisfied. The mass flux per unit zonal length is the same between any two streamlines. Here we calculate the streamfunction by integrating in pressure the top part of 5.2 using meridional winds from NCAR/NCEP reanalyses, taking ψ to be zeros at 1000 hPa.

In January there is strong descent in the northern hemisphere between about 10 degrees and 40 degrees north latitude, and ascent between about 25 degrees south and 10 degrees north, associated with a strong clockwise circulation cell centered around 10 degrees north and 400 hPa. A far weaker counterclockwise cell is centered around 25 degrees south. In July, the strongest overturning cell is in the southern hemisphere, with ascent from about 5 degrees south to 22 degrees north, strong descent in the southern hemisphere from 5 degrees south to about 35 degrees south, and a vanishingly weak clockwise cell in the northern hemisphere.

In the near-equinoctial months of April and October, the circulation cells are much weaker (note the different color scales) and somewhat more symmetric about the equator, although in October the ascent distinctly favors the northern hemisphere.

These overturning circulations are referred to as the “Hadley Circulation”, named after the British meteorologist George Hadley (1685-1768) who was the first to give a qualitatively correct account of their origin (Hadley, 1735). The individual overturning cells are likewise referred to as “Hadley cells”. As can be seen by comparing Figure 5.14 with the mid-tropospheric

distribution of virtual temperature (Figure 5.7), these cells are thermally direct, with less dense air rising and denser air sinking.

The generally poleward net transport of heat by the Hadley Circulation can be visualized by superimposing the mass streamfunction on the distribution of moist entropy (given by (2.69), omitting the last term and taking $T_0 = 250\ K$ and $p_0 = 1000\ hPa$), also calculated from reanalysis data. This superposition is shown in Figure 5.15.

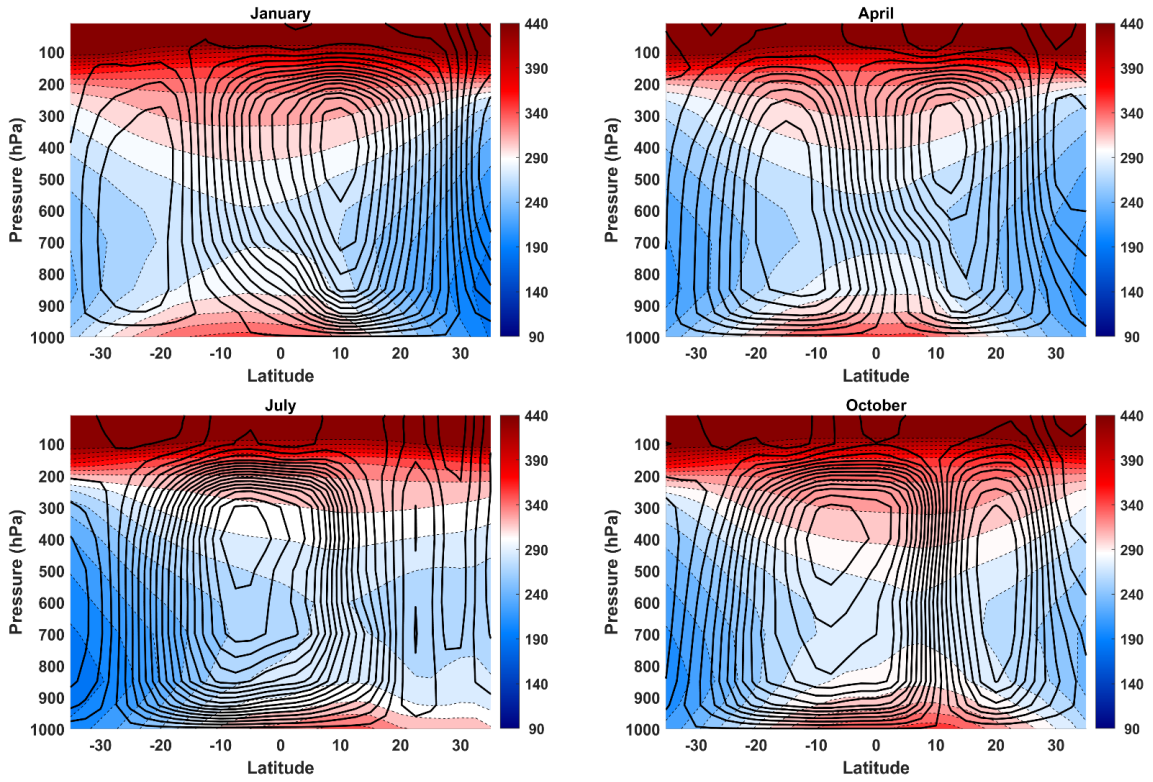


Figure 5.15: Zonal mean, monthly mean mass streamfunction (0.1 Kg m^{-1} s^{-1}) (black contours) superimposed on zonally averaged, monthly mean most entropy given by eq. (2.69), (shading), both from NCAR/NCEP reanalysis data, 1979-2019. Refer to Figure 5.14 for mass streamfunction scale and sign. The moist entropy is capped at 440 J Kg^{-1} K^{-1} to avoid having the plot dominated by large values in the stratosphere.

The vertically averaged, mass-weighted northward flux of entropy is given by the vertical average of the product of the moist entropy and the meridional wind in Figure 5.15. (Note that the meridional energy flux is more nearly given by the flux of moist static energy, not entropy, but the two are qualitatively similar.) It is clear from Figure 5.15 that this vertically averaged meridional entropy flux is generally away from the locus of maximum ascent since the entropy in the top branches of the cells is systematically greater than that in the bottom branches.

The monthly mean zonal wind is shown in Figure 5.16. At the surface, easterlies prevail throughout the tropics, except for a small region around 3-10 degrees north latitude in July. The latitudinal range of tropical easterlies shrinks with altitude, although easterlies prevail near the equatorial tropopause in all seasons. Subtropical jets near the tropopause are strongest in the winter hemisphere and fairly symmetric about the equator near the equinoxes.

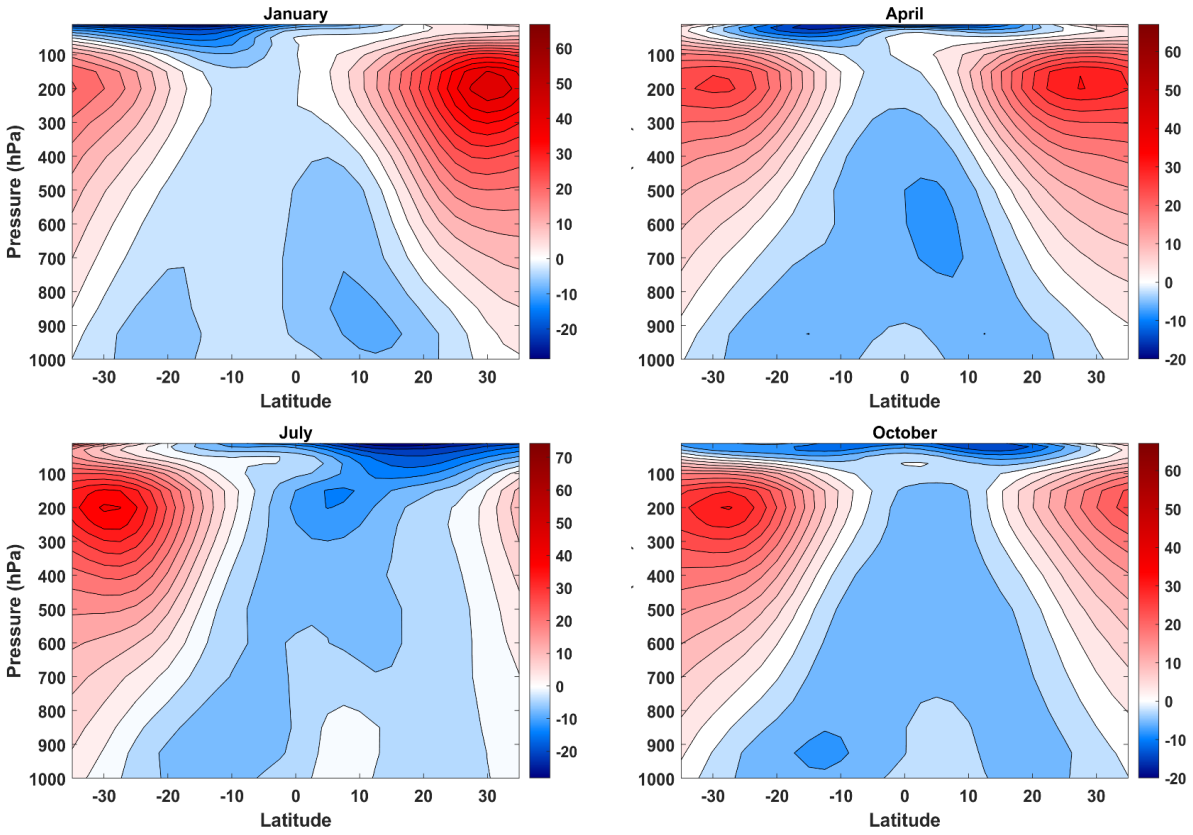


Figure 5.16: Zonal mean, monthly mean zonal wind for January, April, July and October compiled from NCAR/NCEP reanalysis data, 1979-2019.

Of some interest for the dynamics of the Hadley circulation, addressed in the next section, is a quasi-conserved quantity strongly related to the zonal wind: the absolute angular momentum per unit mass:

$$M = a\cos(\theta)\left[\Omega a\cos\left(\theta\right)+u\right], \tag{5.3}$$

where a and Ω are the mean radius and angular rotation rate of the earth, θ is latitude, and u is the zonal velocity. In (5.3) the local distance from the center of the earth has been approximated by the mean planetary radius (i.e. variations with altitude have been neglected along with the departures of the surface from a perfect sphere). The equations of motion on a sphere demonstrate that M is materially conserved in zonally invariant motion in the absence of frictional torque.

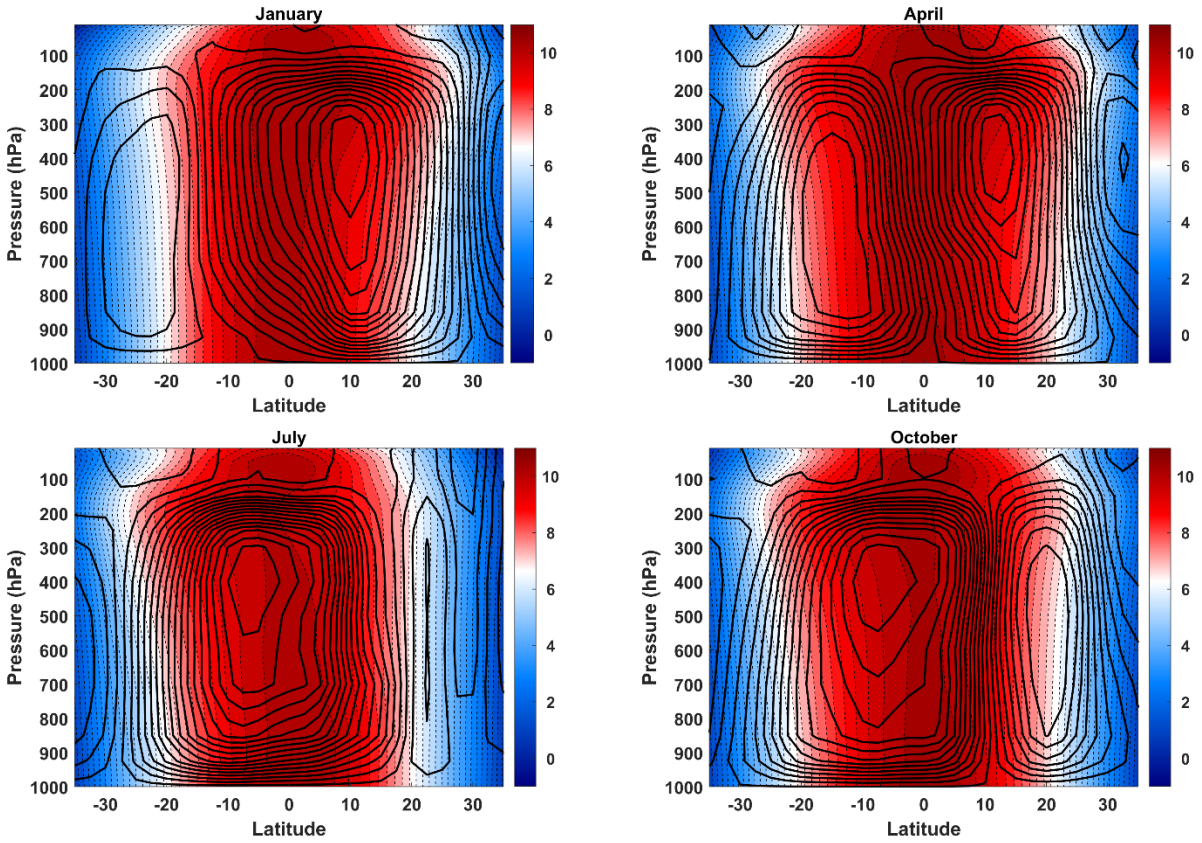


Figure 5.17: Zonal mean, monthly mean absolute angular momentum per unit mass (shading; in 10^8 m^2s^{-1}) and mass streamfunction (black curves) for January, April, July and October compiled from NCAR/NCEP reanalysis data, 1979-2019. Consult Figure 5.14 for mass streamfunction sign and magnitude.

Figure 5.17 shows the long-term mean of monthly values of absolute angular momentum with the mass streamfunction from Figure 5.14 superimposed. While the ascending and descending branches of the Hadley Circulation are more or less along curves of constant M, there is strong flow across M contours in the boundary layer and also in the return branches in the upper troposphere. The flow across M contours in the boundary layer mostly reflects the frictional torque surface drag exerts on the zonal boundary layer flow. Aloft, the cross-M flow partially reflects strong eddy fluxes of angular momentum out of the tropics, accomplished mostly by synoptic-scale eddies, and partially reflects the non-steadiness of the flow owing to seasonal variations (Boos and Emanuel, 2008). We will have more to say about the role of non-stationarity and eddy fluxes in the next section.

The Hadley Circulation is thermally direct, as can be seen in Figure 5.18 which shows the distribution of a temperature-like variable with the mass streamfunction imposed. The temperature variable is defined in terms of the saturation entropy, which is just the moist entropy of air saturated at a given pressure and temperature, and is given, following (2.69), by

$$s^* = \left(c_{pd} + r^* c_l\right)\ln\left(\frac{T}{T_0}\right) - R_d \ln\left(\frac{p_d}{p_0}\right) + \frac{L_v r^*}{T}, \tag{5.4}$$

where the symbols have their usual meanings (see Chapter 2) and r^* is the saturation mixing ratio. Because r^* is a function of temperature and pressure alone, s^* is a function of

temperature only on isobaric surfaces, so it measures buoyancy, but neglecting the direct effect of water on density. The shaded quantity in Figure 5.18 is actually the difference between s^* and it value at the same pressure at the latitude where the lowest level actual entropy (Figure 5.15) reaches its maximum value. Furthermore, we replace s^* by the actual entropy, s, at levels below 850 hPa. By design, this quantity vanishes at all altitudes at the latitude of maximum boundary layer entropy, and generally decreases away from this latitude, in the troposphere.

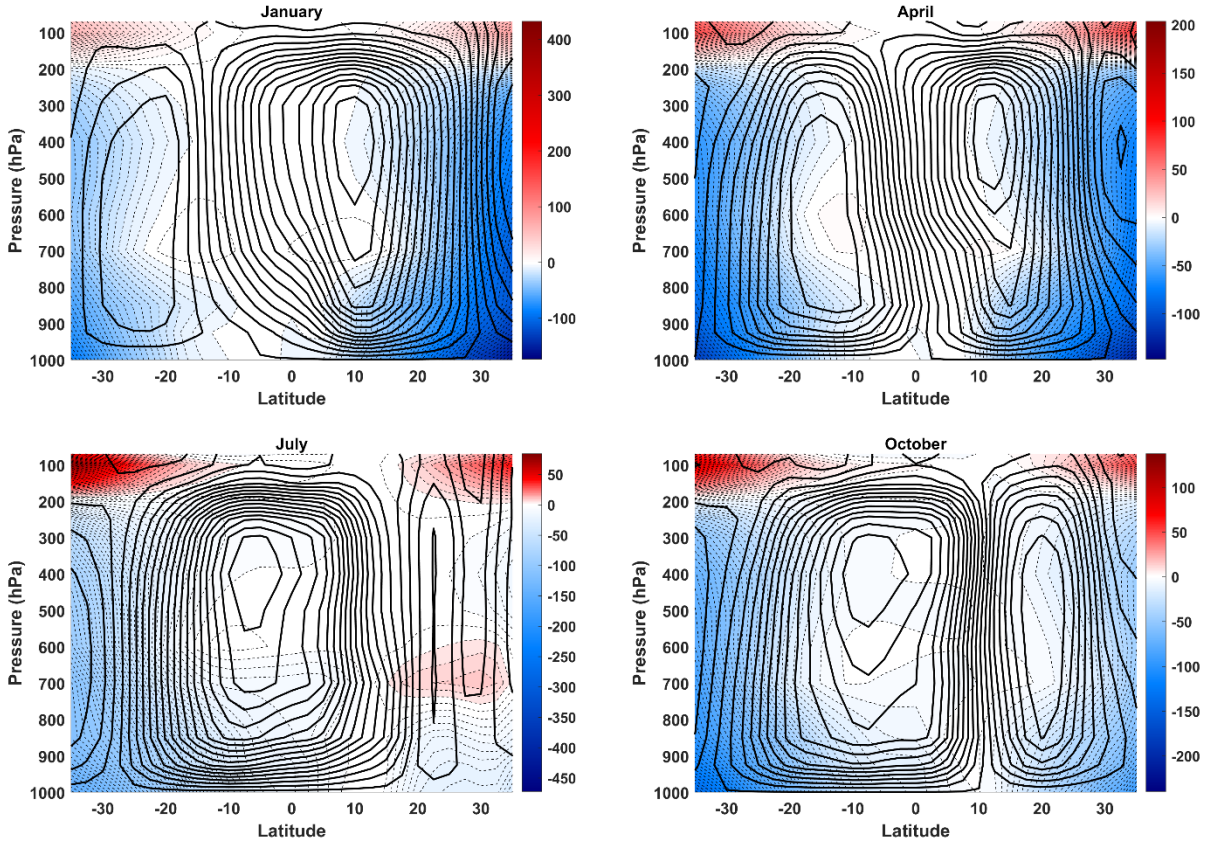


Figure 5.18: Zonal mean, monthly saturation entropy (shaded) and mass streamfunction (black curves) for January, April, July and October compiled from NCAR/NCEP reanalysis data, 1979-2019. The shaded variable is actually the difference between the actual saturation entropy and its value at the same pressure at the latitude at which the boundary layer entropy has its maximum value. (See text for more details, and consult Figure 5.14 for mass streamfunction sign and magnitude.)

As expected, this variable is nearly constant in the deep tropical troposphere, within the ascending branches of the Hadley Circulation. The descending branches are colder, demonstrating that the circulation is thermally direct, converting potential energy into kinetic energy. However, as we will discuss in the next section, this fact does not prove that the Hadley Circulation is driven by temperature gradients; indeed, the causal pathway for the Hadley flow is still a matter of debate among tropical researchers.

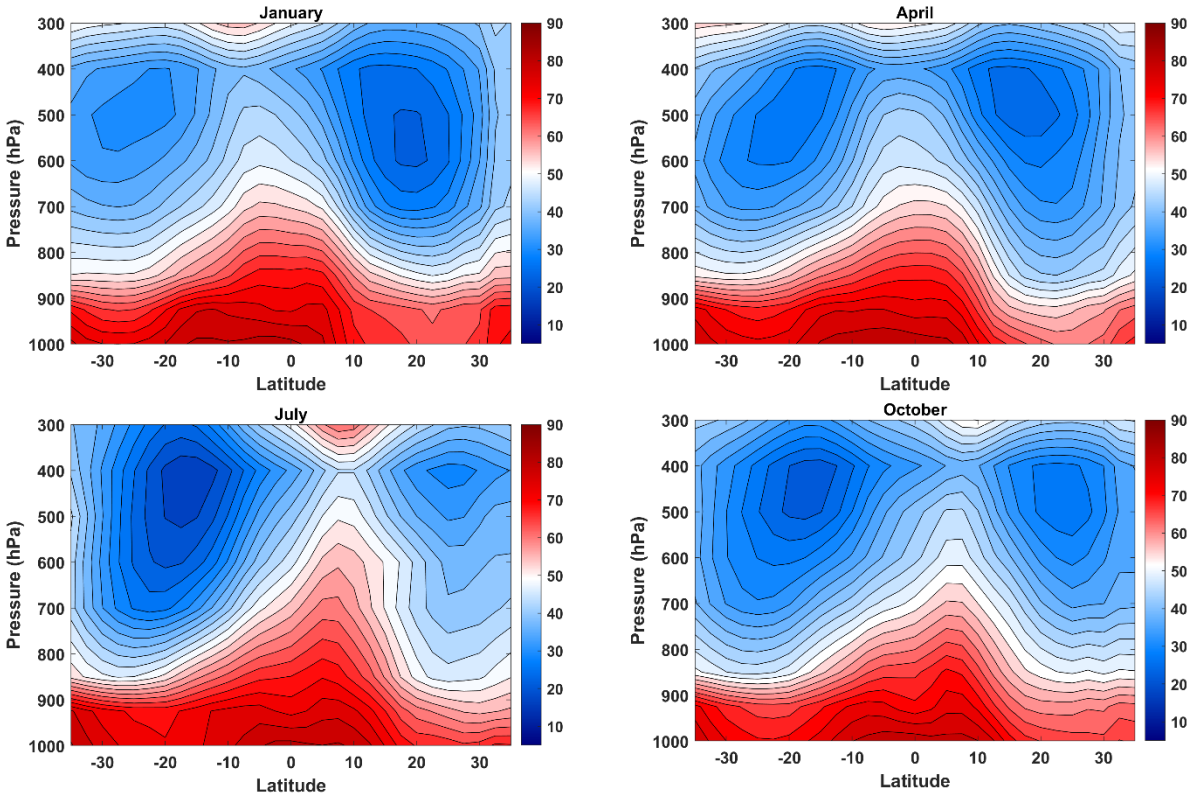


Figure 5.19: Zonal mean, monthly relative humidity (%) for January, April, July and October compiled from NCAR/NCEP reanalysis data, 1979-2019. Note that in contrast to the previous figures, 300 hPa marks the top of these diagrams as NCAR/NCEP reanalyses do not provide water vapor above this level.

Finally, Figure 5.19 shows the zonal mean, monthly mean relative humidity averaged over the same period (1979-2019). Relative humidity peaks in the near-equatorial boundary layer and decreases rapidly upward, reaching a minimum around the 400 hPa level. (This is consistent with single-column RCE calculations, as shown for example in Figure 3.20, and these simulations strongly suggest that relative humidity increases above 400 hPa and often reaches saturation at the tropopause.) The latitude of maximum humidity oscillates north and south with the seasons. Relative humidity has minima in the subtropics that are more extreme in the winter hemisphere.

Unlike temperature, which is dynamically constrained to vary slowly with time and space in the tropics, water vapor is a passive tracer in the vast majority of tropical air that is unsaturated, and typically varies at all time and space scales. Zonally and temporally averaging these distributions yields, as one would expect, the smooth distributions shown in Figure 5.19, but the reader should not be fooled into thinking that these are representative of typical water vapor distributions. Indeed, a snapshot of water vapor at 600 hPa (Figure 5.20) serves to remind us of the enormous space-time inhomogeneity of water vapor.

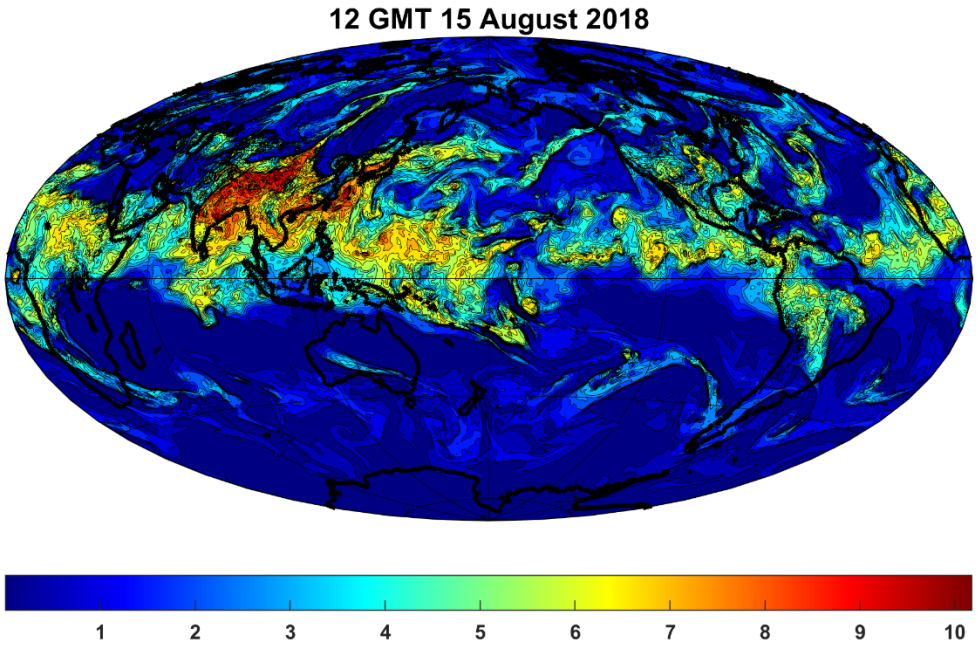


Figure 5.20: Specific humidity (g Kg^{-1}) at the 600 hPa level at 12 GMT on 15 August, 2018, from ERA-5 reanalysis data.

5.2 The Hadley Circulation

The relative zonal symmetry of the radiative imbalance, time-mean winds, temperature, and pressure in the tropical atmosphere suggests that, to first order, we might approximate the Hadley Circulation as a two-dimensional, steady overturning, symmetric about the equator, as though there were a perpetual equinox. We will start from this radical simplification and later introduce progressively more realistic physics.

To begin with, it is not at all obvious that there should be a thermally direct, Hadley-like circulation on a rotating planet with no seasons. To see this point, consider this counter-factual hypothesis: Each column of the tropical atmosphere is individually in an RCE state. Owing to diminishing insolation with latitude, the temperature of this state decreases poleward, giving rise to meridional pressure gradients. These pressure gradients are everywhere in geostrophic balance with a zonal wind. For energetic consistency, we demand that these pressure gradients and zonal winds vanish at the surface and that the interior of the atmosphere is inviscid, so that there is no dissipation of mechanical energy. The zonal wind blows along contours of entropy and angular momentum, so that these quantities are not advected and the RCE state is maintained.

If we assume geostrophic and hydrostatic balance, and that the vertical lapse rate is everywhere moist adiabatic, approximated here by constant s^* as given by (5.4), then we can derive an analytic solution for the zonal winds as a function of the meridional distribution of boundary layer moist entropy, s_b.

We begin with the expression of zonal geostrophic balance in pressure coordinates on a sphere:

$$\left(\frac{\partial \phi}{a \partial \theta}\right)_p = -2\Omega \sin\theta u - \frac{u^2}{a}\tan\theta, \tag{5.5}$$

where ϕ is the geopotential, u is the zonal wind, Ω is the angular velocity of the earth's rotation, a is the mean radius of the earth, and θ is latitude. Using the absolute angular momentum as defined by (5.3) in place of the zonal velocity, (5.5) can be written instead

$$\left(\frac{\partial \phi}{\partial \theta}\right)_p = -\sin\theta\left[\frac{M^2 - \Omega^2 a^4 \cos^4\theta}{a^2 \cos^3\theta}\right]. \tag{5.6}$$

The hydrostatic equation in pressure coordinates is

$$\frac{\partial \phi}{\partial p} = -\alpha, \tag{5.7}$$

where α is the specific volume. Using cross-differentiation to eliminate ϕ between (5.6) and (5.7) we can derive a thermal wind equation:

$$\frac{1}{a^2}\frac{\tan\theta}{\cos^2\theta}\frac{\partial M^2}{\partial p} = \left(\frac{\partial \alpha}{\partial \theta}\right)_p = \left(\frac{\partial \alpha}{\partial s^*}\right)_p \left(\frac{ds^*}{d\theta}\right), \tag{5.8}$$

since, to a reasonable approximation, α is just a function of s^* and p, and we recognize that in a steady, zonally symmetric flow with moist adiabatic lapse rates, s^* is only a function of θ. We next turn to one of Maxwell's relations derived in Chapter 2 (see equation 2.74), neglecting the water vapor mixing ratio in comparison to unity:

$$\left(\frac{\partial \alpha}{\partial s^*}\right)_p \cong \left(\frac{\partial T}{\partial p}\right)_{s^*}. \tag{5.9}$$

Substituting this into (5.8) yields

$$\frac{1}{a^2}\frac{\tan\theta}{\cos^2\theta}\frac{\partial M^2}{\partial p} = \left(\frac{\partial T}{\partial p}\right)_{s^*}\left(\frac{ds^*}{d\theta}\right). \tag{5.10}$$

We can simplify the form of (5.10) by transforming the meridional coordinate. Defining

$$y \equiv \sec^2(\theta), \tag{5.11}$$

(5.10) becomes

$$\frac{\partial M^2}{\partial p} = 2a^2\left(\frac{\partial T}{\partial p}\right)_{s^*}\left(\frac{ds^*}{dy}\right). \tag{5.12}$$

Because s^* does not vary with pressure along a moist adiabat, (5.12) can be directly integrated once in pressure. Applying the boundary condition $u = 0$ $\left(M = \Omega a^2 y^{-1}\right)$ at the surface, integrating (5.12) yields

$$M^2 = \Omega^2 a^4 y^{-2} - 2a^2\left(T_s - T\right)\left(\frac{ds^*}{dy}\right), \tag{5.13}$$

where T_s is the surface temperature, which may be a function of latitude. We see here that *in a moist adiabatic, hydrostatic atmosphere, absolute temperature along a moist adiabat is the natural vertical coordinate*, rather than either altitude or pressure. We will make use of this finding throughout the rest of this book.

In terms of zonal velocity, (5.13) may be written

$$u^2 + 2\Omega a y^{-\frac{1}{2}} u + 2y\left(T_s - T\right)\left(\frac{ds^*}{dy}\right) = 0. \tag{5.14}$$

Since $2\Omega a$ is about $930\ ms^{-1}$, the first term in (5.12) is considerably smaller in magnitude than the second term, and so we write to a good approximation

$$u \simeq -\frac{\left(T_s - T\right)}{\Omega a} y^{\frac{3}{2}}\left(\frac{ds^*}{dy}\right). \tag{5.15}$$

Because y does not change monotonically across the equator, it is helpful also to express (5.15) in terms of latitude, using the definition (5.11):

$$u \simeq -\frac{\left(T_s - T\right)}{2\Omega a \sin\theta}\frac{ds*}{d\theta}. \tag{5.16}$$

5.2.1. The equatorially symmetric Hadley Circulation

We first consider the case of a perpetual equinox, when the RCE temperature peaks at the equator (ignoring the ellipticity of the earth's orbit around the sun). In that case, $s*$ is symmetric about the equator, and so the last factor on the right-hand side of (5.16) vanishes there, but then so does $\sin(\theta)$. For the solution to be viable, the gradient of $s*$ must approach zero more rapidly than $\sin(\theta)$ as one approaches the equator. It also follows that this solution fails for any distribution of $s*$ that is not symmetric about the equator close to the equator.

We see through (5.16) that this RCE-everywhere solution has zonal winds increasing with altitude (that is, with decreasing absolute temperature) in proportion to the magnitude of the $s*$ gradient, but because of the $\sin(\theta)$ term, the strongest zonal winds will be found equatorward of the latitude of the largest magnitude of the $s*$ gradient.

On a rotating aquaplanet (no topography) with an inviscid atmosphere above the boundary layer and perpetual equinox insolation, RCE at every latitude together with the solution of (5.14) constitutes a complete, nonlinear, steady solution of the governing equations in a moist adiabatic troposphere. There is no meridional or vertical velocity and thus no Hadley cells. So why do we find Hadley cells in nature?

The answer to this question is complex, but there is one contributing factor that is readily apparent. Returning to the expression (5.13) for the angular momentum, we first note that the last term on the right has its maximum magnitude at the tropopause, where we define the absolute temperature to be T_t. (Above the tropopause, the stratosphere is not moist adiabatic and not linked to the surface through deep convection.) And the larger the equatorward $s*$ gradient, the larger the magnitude of the absolute angular momentum at the tropopause. But M cannot be arbitrarily large. Since it is materially conserved in zonally symmetric flows, obeying a diffusion equation of the form

$$\frac{dM}{dt} = \nabla \bullet \nu \nabla M,$$

where ν is a diffusion coefficient, any interior extremum of M must eventually diffuse away. Thus, in a steady state, all extrema of M must occur on boundaries. This is known as Hide's theorem named after Raymond Hide (Hide, 1969). For this reason, in a steady, zonally symmetric flow, M is bounded above by its resting value at the equator, Ωa^2. That is, from (5.13) we require that

$$-\left(T_s - T_t\right)\frac{ds*}{dy} \le \frac{1}{2}a^2\Omega^2\left(1 - \frac{1}{y^2}\right), \tag{5.17}$$

Therefore, the geostrophically balanced RCE solution is only viable when (5.17) is satisfied. But what if the meridional gradients of RCE temperature (s^*) are strong enough to violate (5.17)? We can interpret this as the problem that there is not enough angular momentum in the system to allow for enough zonal wind (at the tropopause) to satisfy geostrophic balance. The pressure gradients there are too large, and thus air will have to accelerate meridionally down the pressure gradient (i.e. toward the poles). To satisfy mass continuity, there will have to be a return flow at the surface. This is one mechanism for creating a Hadley Circulation, as first demonstrated by Held and Hou (1980).

One way to avoid violating Hide's Theorem is to relax the assumption that the surface zonal wind is zero. But surface drag acting on a zonal wind will lead to an Ekman drift, establishing a circulation in the meridional-vertical direction. This circulation, by advecting heat and moisture, will upset the RCE state and this will in turn change the geostrophic zonal wind. Nevertheless, we will begin by calculating the state as though this feedback is absent and later consider its effect.

Following this idea, we assume that the poleward branches of the Hadley circulation conserve angular momentum, and that the value of M at the tropopause is that of air at rest at the equator, Ωa^2. Next, we integrate (5.12) *downward* from the tropopause to find the zonal wind throughout the troposphere. The result is

$$u = \frac{y^{\frac{3}{2}}}{2\Omega a}\left[\Omega^2 a^2\left(1-y^{-2}\right)+2\left(T_s-T\right)\frac{ds^*}{dy}\right]. \tag{5.18}$$

Comparing this with the criterion for violating Hide's theorem, (5.17), we can express (5.18) as

$$u = \frac{y^{\frac{3}{2}}}{\Omega a}\left[\left(T-T_t\right)\frac{ds^*}{dy}-\left(T_s-T_t\right)\left(\frac{ds^*}{dy}\right)_c\right], \tag{5.19}$$

in which the last term is, from (5.17) the critical value of the meridional saturation entropy gradient that corresponds to constant angular momentum at the tropopause. At the tropopause, the first term in brackets vanishes and, since ds^*/dy is negative the zonal wind will be positive. The magnitude of the westerly wind decreases downward from the tropopause and at some absolute temperature changes sign provided the magnitude of ds^*/dy exceeds its critical value (which we have assumed in deciding to integrate (5.12) down from the tropopause). At the surface, the zonal geostrophic wind is easterly, given by

$$u_s = \frac{y^{\frac{3}{2}}}{\Omega a}\left(T_s-T_t\right)\left[\frac{ds^*}{dy}-\left(\frac{ds^*}{dy}\right)_c\right]. \tag{5.20}$$

This expression can tell us something about how close the actual meridional gradient of s^* is to its critical value versus its RCE value. To answer this question, we need to know what the RCE gradient is. Unfortunately, it is not possible to calculate it since, as it turns out, the RCE near the equator produces a runaway greenhouse and there is no solution. But suppose, for the sake of argument, the RCE gradient at 15 degrees latitude were twice its critical value. Then,

from (5.20) and using the critical gradient given by taking the equality in (5.17), the surface zonal wind would be about $-35\ ms^{-1}$, far in excess of what is observed. We can conclude from this that the observed s^* gradient around the time of the equinoxes (when the Hadley circulation is approximately symmetric around the equator) is much closer to its critical value than it is to its RCE value (assuming it could be computed). This also means than we cannot use (5.20) with the RCE gradient of s^* to estimate the surface zonal wind and then, using Ekman balance, the meridional wind in the boundary layer. So then, how can we go about estimating the Hadley Circulation?

One way of doing this was suggested by Held and Hou (1980): Represent the net column cooling and heating rates as linear functions of their departure from RCE, and use those rates to calculate the vertical velocity and then use mass continuity to calculate the meridional flow. In essence this uses the energy budget to calculate the circulation. We will here employ a variant of Held and Hou's method based on the boundary layer quasi-equilibrium framework we developed in Chapter 3 (see (3.59)-(3.66)). This framework was developed for WTG conditions but will work for steady flows provided that horizontal moist static energy advection in the boundary layer is accounted for in addition to surface enthalpy fluxes. The modified framework consists of expressions for the net vertical velocity:

$$\rho w_b = \frac{1}{1-\epsilon_p}\left[\frac{\epsilon_p\left(F_h + adv\right)}{h_b - h_m} - \frac{\dot{Q}_{cool}}{\mathbf{S}}\right], \tag{5.21}$$

convective updraft mass flux:

$$M_u = \frac{1}{1-\epsilon_p}\left[\frac{F_h + adv}{h_b - h_m} - \frac{\dot{Q}_{cool}}{\mathbf{S}}\right], \tag{5.22}$$

and mid-tropospheric moist static energy:

$$\frac{\partial h_m}{\partial t} = -\left\langle \mathbf{V}_h \bullet \nabla h \right\rangle - G w_b - \left\langle \alpha \dot{Q}_{cool} \right\rangle + \frac{g}{p_s - p_t} F_h. \tag{5.23}$$

Here the term adv is the vertically integrated horizontal advection of subcloud layer moist static energy, which in deep convecting regions is considered equal to the tropospheric saturation moist static energy h^* :

$$adv \equiv -\frac{\rho dv}{a}\frac{dh^*}{d\theta} = -\frac{\rho dv \overline{T}}{a}\sqrt{y^3 - y^2}\,\frac{ds^*}{dy}, \tag{5.24}$$

where v is the meridional velocity, d is the depth of the subcloud layer, and $\overline{T}$ is a representative temperature of the troposphere. We use (5.17) for the critical value of the meridional s^* gradient.

In regions where deep convection is absent, we set $M_u = 0$ and $\rho w_b = -\dot{Q}_{cool} / \mathbf{S}$, and we must solve a separate equation for the subcloud layer moist static energy, h_b . Here we will use these equations, with some simplifications, to estimate the vertical circulation and convective mass

flux associated with a steady, equatorially symmetric Hadley circulation under the approximation that the saturation entropy distribution with latitude is near its critical value.

First, as we are considering steady circulations, we set the left side of (5.23) to zero. We also ignore vertically integrated horizontal advection of moist static energy and assume that the tropospheric mean of the radiative cooling rate, $\left\langle \alpha \dot{Q}_{cool} \right\rangle$, is equal to the radiative cooling rate just above the top of the boundary layer that appears in (5.21) and (5.22). Then in regions of deep convection, (5.23) becomes

$$\rho w_b = \frac{1}{G}\left[\frac{\rho g F_h}{p_s - p_t} - \dot{Q}_{cool} \right] \quad (M_u > 0). \tag{5.25}$$

The large-scale vertical velocity is just the difference between the column enthalpy increase owing to surface fluxes, and its decrease by radiative cooling. Elsewhere, subsidence warming is assumed to balance radiative cooling:

$$\rho w_b = -\frac{\dot{Q}_{cool}}{\mathbf{S}}, \quad (M_u = 0) \tag{5.26}$$

For this system we will take the surface enthalpy flux, F_h, to be a specified function of latitude. If there is no advection of heat in the ocean itself, and the ocean is in thermal equilibrium, then the turbulent heat flux from the sea must equal the net radiative flux into the sea. But because of the high moisture content of the tropical lower atmosphere, the net infrared flux into the ocean is small and so F_h is dominated by the solar flux, which we specify as a function of latitude here. As a proxy for the surface enthalpy flux, we will define an RCE radiative cooling rate as that which balances the surface enthalpy flux:

$$\dot{Q}_{RCE} \equiv \frac{\rho g F_h}{p_s - p_t}. \tag{5.27}$$

So we consider $\dot{Q}_{RCE}$ so be a specified function of latitude. With this notation, we can write (5.26) and (5.27) in the form

$$\rho w_b = \begin{cases} \dfrac{1}{G}\left[\dot{Q}_{RCE} - \dot{Q}_{cool} \right] & (M_u > 0) \\ -\dfrac{\dot{Q}_{cool}}{\mathbf{S}} & (M_u = 0) \end{cases} \tag{5.28}$$

The vertical motion in the convecting region is just the difference between the actual cooling rate and the RCE cooling rate, divided by the gross moist stability, G. We will approximate the actual radiative cooling rate as a function of the critical saturation entropy distribution given by the equality in (5.17). Note that mass conservation demands that the integral of (5.28) over latitude vanishes:

$$\int_0^{\theta_{max}} \rho w_b \cos(\theta) d\theta = 0, \tag{5.29}$$

where θ_{max} is the poleward boundary of the Hadley Circulation.

Given the distributions of the actual and RCE radiative cooling rates, (5.29) can be used to determine the latitudinal extent of the Hadley circulation.

We also need to ensure energetic consistency. For that to be the case, there must be no net TOA (shortwave plus longwave) radiation, and in our formulation, this would mean that

$$\int_0^{\theta_{max}} \left[\dot{Q}_{RCE} - \dot{Q}_{cool}\right] \cos(\theta) d\theta = 0. \tag{5.30}$$

Note that if G is constant, this relationship is automatically satisfied if there is deep convection everywhere, according to (5.28) and (5.29). But in nature that is not the case, and the condition (5.30) must be used to determine the boundary between large-scale ascent and descent.

Finally, in determining $\dot{Q}_{cool}$ from the assumed critical distribution of temperature, there is a free constant that comes from integrating (5.17). We can choose this constant so that the actual and RCE values of the radiative cooling are equal at the poleward boundary of the Hadley circulation. Therefore, if we assume that variations in $\dot{Q}_{cool}$ are proportional to variations in s^* and integrate (5.17) in y we have

$$\dot{Q}_{cool} = \dot{Q}_{c0} + \gamma\left(2 - y - 1/y\right), \tag{5.31}$$

where $\dot{Q}_{c0}$ is the actual cooling rate in the equator and γ is related to the proportionality between variations in radiative cooling and variations in s^*.

Once we specify the solar flux as a function of latitude, which yields $\dot{Q}_{RCE}$ in our framework, and given values of G and $\mathbf{S}$, we can solve (5.29) varying $\dot{Q}_{c0}$ and the latitude separating ascent from descent so as to satisfy mass and energy conservation as expressed by (5.29) and (5.30).

Note that to solve for the large-scale ascent in this system, we have not needed to calculate the convective updraft mass flux or the moist static energy deficit of the middle troposphere, $h_b - h_m$, but we have found it necessary to calculate the latitude separating deep convection from no deep convection. But once we have calculated the ascent together with the separation latitude and the constant $\dot{Q}_{c0}$, we can then use (5.21) and (5.22) to calculate the mass flux and moist static energy deficit in the ascent region. (By definition, the mass flux vanishes elsewhere, and we lose the ability of calculate the moist static energy deficit where deep convection is absent. Technically, if the tropopause is very cold, the moist static energy of the descent region would be nearly as low as the dry static energy in this zonally symmetric model, but in nature, zonally asymmetric eddies transport some water into the descending branches of the Hadley circulation.)

Comparing (5.28) to (5.21) in regions where deep convection does not vanish yields an expression for the difference between the boundary layer and lower troposphere moist static energies:

$$h_b - h_m = \frac{G\epsilon_p\left(p_s - p_t\right)}{\rho g}\frac{\dot{Q}_{RCE} + adv}{\left(1-\epsilon_p\right)\left(\dot{Q}_{RCE} + adv - \dot{Q}_{cool}\right) + \frac{G}{\mathbf{S}}\dot{Q}_{cool}}. \tag{5.32}$$

By substituting (5.32) into (5.22), we get an expression for the convective updraft mass flux:

$$\epsilon_p G M_u = \dot{Q}_{RCE} + adv - \dot{Q}_{cool} + \frac{G}{\mathbf{S}}\dot{Q}_{cool}. \tag{5.33}$$

Figure 5.21 shows a particular solution of the system given by (5.28), and (5.31) - (5.33) satisfying the constraints (5.29) and (5.30). Here we have taken the RCE insolation $\dot{Q}_{RCE}$ to depend on $\cos^2(\theta)$, $\gamma = 1$, $\epsilon_p = 0.6$, $\mathbf{S} = 5\ ms^{-2}$, and $G = 0.03\mathbf{S}$.

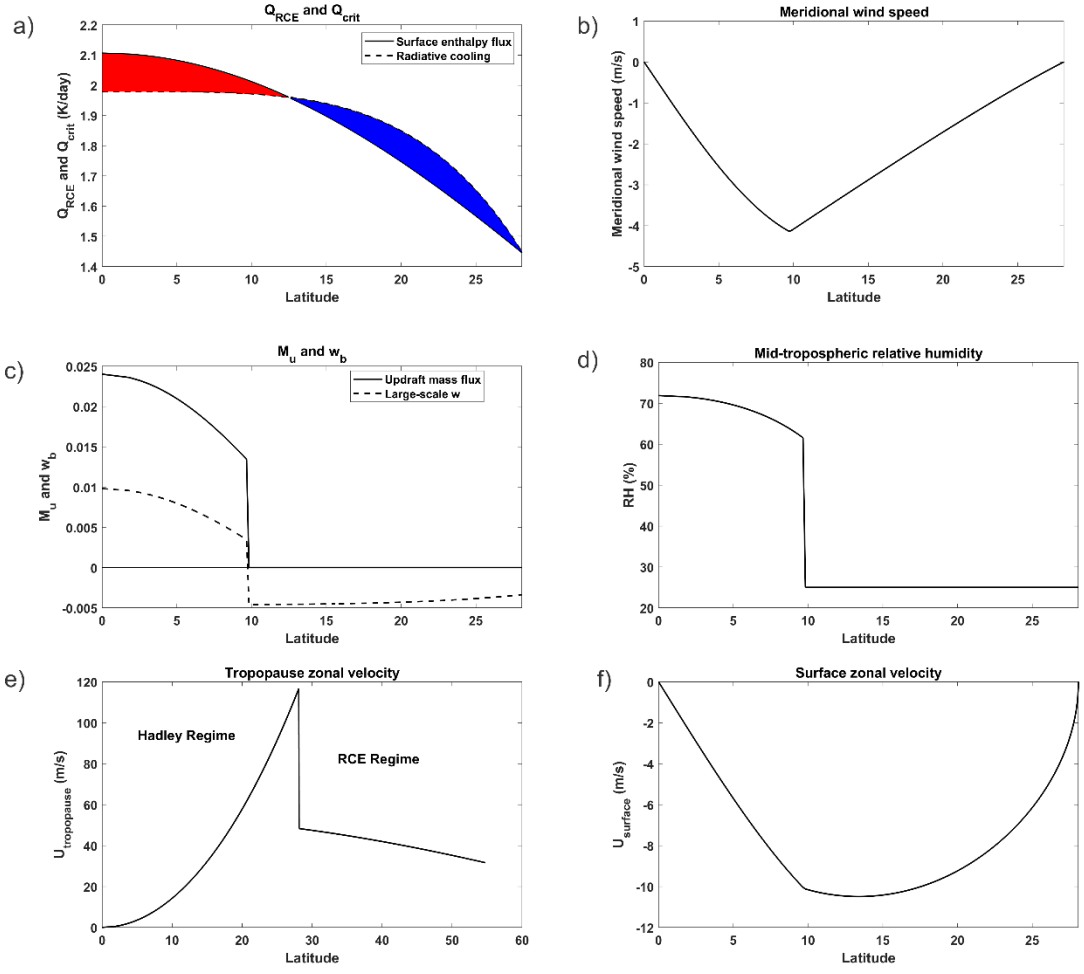


Figure 5.21: A solution for the zonally symmetric, equatorially symmetric Hadley circulation for prescribed surface enthalpy flux and radiative cooling rate proportional to the critical saturation entropy distribution. a: Surface enthalpy flux (solid) and radiative cooling (dashed); red shading indicates net heating and blue shading shows net cooling. b: Surface meridional velocity. c: Convective updraft mass flux (solid) and large-scale vertical velocity (dashed). d: Mid-tropospheric relative humidity deduced from difference between boundary layer and free troposphere moist static energies. The humidity has arbitrarily been set to 25% in the region free of deep convection. e: Zonal wind at the tropopause. Note expanded latitude scale. f: Surface zonal wind.

The ascent extends to about 9 degrees latitude, but the surface enthalpy flux exceeds the radiative cooling up to about 18 degrees. In this region, the surface fluxes are being used to increase the subcloud layer moist static energy in its journey toward the equator, rather than being used to heat the atmosphere. The meridional wind speed reaches it maximum magnitude at the edge of the region of ascent and has been calculated from the large-scale vertical velocity using mass continuity. The tropospheric relative humidity in the ascent zone has been estimated from the quantity $h_b - h_m$ given by (5.32) and set arbitrarily to 25% in the descending branch of the circulation.

While the solution shown here satisfies conservation of mass and thermodynamic energy, it does not satisfy conservation of angular momentum. As we have assumed that the actual horizontal temperature (saturation entropy) gradient is just critical – that is, having constant angular momentum along the tropopause – the poleward flow near the tropopause conserves angular momentum, but the downflow must at some point cross angular momentum surfaces. Here it should be pointed out that the angular momentum surfaces cannot "bend over" to follow the flow downward in the subtropics. Were they to do so, the poleward gradient of angular momentum would be locally positive, which is highly unstable to small-scale inertial instability. Given this circumstance, how do Hadley circulations resolve this conundrum?

Figure 5.21e shows the zonal wind at the tropopause in this solution. Within the Hadley regime, by assumption, the temperature gradient is critical and the angular momentum is constant at its resting value at the equator. Poleward of the Hadley regime, the zonal wind is that which is consistent with the RCE temperature gradient. The zonal wind jumps cyclonically across the poleward edge of the Hadley circulation.

In nature, even a small viscosity would give rise to a flux of angular velocity across the jump in the zonal wind, transporting angular momentum from the Hadley circulation to the RCE state on its poleward flanks. This would give rise to surface easterlies in the Hadley regime and westerlies in the RCE state; note that the surface frictional sink of angular momentum at high latitudes would necessarily balance its surface frictional source in the Hadley regime.

In fact, we can estimate the flux of angular momentum to higher latitudes by "reverse engineering" the Ekman flow in the Hadley regime. In the steady state, the advection of angular momentum by the equatorward flow near the surface must balance the source of angular momentum owing to surface stress arising from a zonal wind:

$$\int_0^d \rho v \frac{\partial M}{a \partial \theta} dz = -a \cos(\theta) \tau_x, \tag{5.34}$$

where τ_x is the zonal surface stress and the integral is over the depth of the friction layer. If we assume a neutrally stratified boundary layer, then the surface stress can be well approximated by a simple aerodynamic drag formula:

$$\tau_x = \rho C_D |\mathbf{V}| u \tag{5.35}$$

where C_D is a nondimensional drag coefficient and $|\mathbf{V}|$ is the magnitude of the surface wind. If we further assume that the magnitude of the surface wind is dominated by its zonal component, that the latter is negative (easterlies), that density does not vary much over the depth of the

boundary layer, and that the angular momentum near the surface is dominated by its planetary component, then (5.34) and (5.35) can be solved for the surface zonal wind:

$$u = -\sqrt{\frac{2\Omega d}{C_D}|v|\sqrt{1-1/y}}, \tag{5.36}$$

with the meridional wind v given by mass continuity applied to the large-scale vertical velocity given by (5.28). This estimate of the surface zonal wind is shown in Figure 5.21f.

The required leakage of angular momentum across the poleward boundary of the Hadley circulation causes westerly winds to develop at the surface the RCE regime. The surface frictional torque exerted by these westerlies must, in equilibrium, balance the torque of the opposite sign in the Hadley regime. From the Ekman relation (5.34), this would be associated with poleward winds near the surface, giving rise to a thermally indirect circulation known as a Ferrel Cell, after the American meteorologist William Ferrel, who in 1856 published an essay theorizing its existence. Portions of Ferrel Cells can be discerned near the edges of the panels in Figure 5.14. Ferrel Cells are thermally indirect… they are refrigerators driven mechanically by the angular momentum flux from the Hadley zones.

While zonally symmetric models (e.g. that of Held and Hou, 1980) indeed produce Hadley circulations that strongly resemble the solution shown in Figure 5.21, in reality the flux of angular momentum from the Hadley cell to middle latitudes is provided by strongly three-dimensional baroclinic eddies. The flux can be visualized by noting from quasi-geostrophic theory that the acceleration of the zonal mean zonal wind can be related to the zonal mean of the meridional eddy flux of the quasi-geostrophic form of potential vorticity. But here we instead look at the distribution of Ertel's potential vorticity, which is materially conserved in the absence of friction and heating:

$$PV = \alpha\left(2\Omega\sin(\theta)\mathbf{k} + \nabla\times\mathbf{V}\right)\bullet\nabla s_d, \tag{5.37}$$

where α is the specific volume, $\mathbf{k}$ is the unit vector in the vertical, $\mathbf{V}$ is the three-dimensional vector velocity, and s_d is the entropy of dry air, usually represented by the potential temperature. Figure 5.22 shows a snapshot in time of the spatial distribution of potential temperature on a surface of constant PV that closely coincides with the tropopause. (This surface is usually referred to as the "dynamic tropopause".)

Extremely strong horizontal gradients of PV (or entropy on a PV surface) mark the instantaneous poleward boundary of the Hadley regime. The swirls at higher latitudes are Rossby waves associated with baroclinic instability. Chunks of high PV (or entropy on a PV surface) occasionally break off into the Hadley zone, an irreversible process known as "Rossby wave breaking", while low PV from the tropics is similarly ingested into middle and high latitudes. This wave breaking is the main agent of angular momentum flux from the tropics to higher latitudes. It is important to note that this process would occur even if the RCE temperature distribution in the tropics were subcritical and would force a Hadley-like circulation. This raises the question, still debated, about whether the actual Hadley cells are really driven by the violation of Hide's theorem or by the eddy angular momentum fluxes from the tropics to

middle latitudes. These two ideas are not mutually exclusive and it is likely that both processes contribute to the observed circulation.

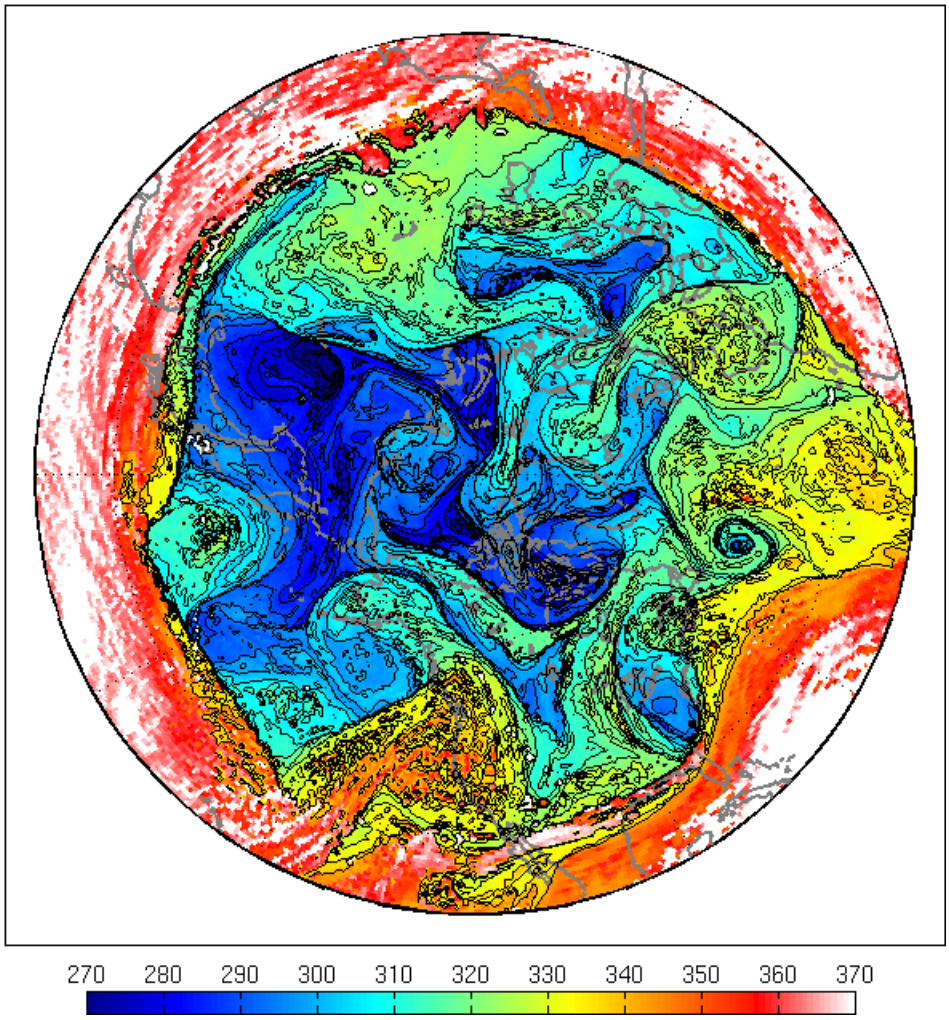


Figure 5.22: Distribution of potential temperature (K) on a surface of constant PV (the "dynamic tropopause') at 18 UTC on January 16th, 2021. The North Pole is at the center of the figure.

The divergence of the long-term average, zonal mean eddy flux of angular momentum is superimposed on the Hadley streamfunction in Figure 5.23. This is a direct measure of the sink of angular momentum by eddies and, in effect, shows the long-term zonal mean statistical effects of Rossby wave breaking. In all seasons, there are strong sinks of angular momentum in the poleward flow in the upper troposphere. As mentioned above, these sinks are necessary to close the angular momentum budget of the equatorially symmetric Hadley system. But, as described in the next section, they are not strictly necessary in the case of cross-equatorial circulations.

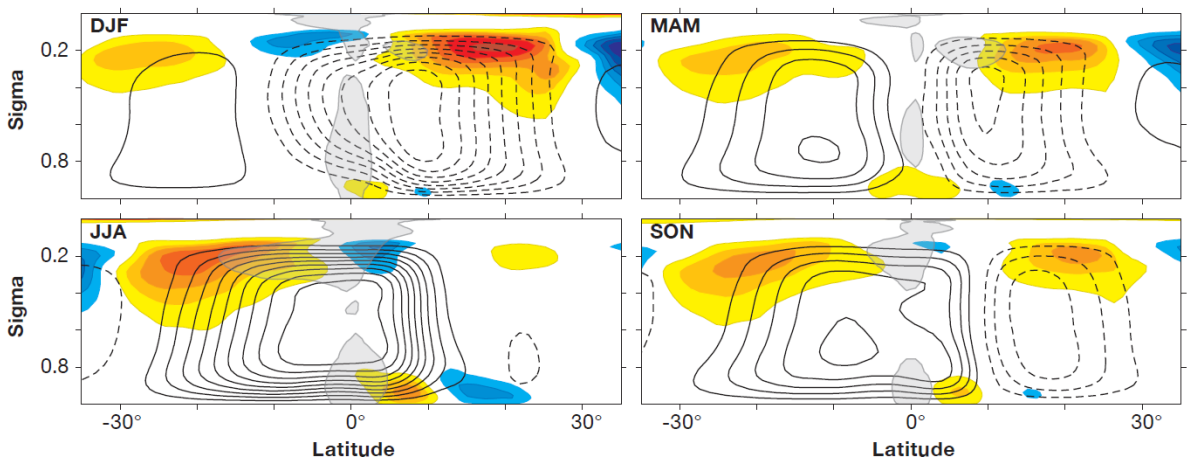


Figure 5.23: Zonal mean, seasonal mean divergence of the eddy flux of angular momentum (shading) and the mass streamfunction (black contours) for December-February, March-May, June-August, and September-November, based on European Center for Medium Range Forecasts reanalysis data average over 1980-2001. Dashed contours indicate clockwise circulation.

5.2.2. Equatorially asymmetric thermally direct circulations

When an RCE temperature gradient exists at the equator, as for example during solstices, a geostrophically balanced zonal wind is not possible and a circulation must ensue. If we continue to suppose that the circulation drives the actual temperature (saturation entropy) gradient back towards it critical value, then because of the symmetry of the thermal wind relation (5.12) and the y coordinate (5.11), it is clear that the gradient of s^* must vanish on the equator. RCE cannot be achieved there, and the gradient of actual subcloud layer moist entropy across the equator cannot match the zero gradient of s^*. Therefore, the subcloud layer must be convectively stable on the cold side of the equator.

For the sake of simplicity, let's continue to suppose that the RCE state is steady and zonally symmetric, but let the peak RCE temperature (s^*) peak off the equator. In what follows, we assume that the peak is in the northern hemisphere, but any solution will work if the peak is instead in the southern hemisphere just by swapping the sign of the latitude and the meridional wind speed.

Poleward of the latitude of maximum s^*, which we will refer to as θ_m, the situation is much like the equatorially symmetric case. If the gradient of s^* exceeds a critical value, a thermally direct circulation must ensue; the only difference from the equatorially symmetric case is that the maximum value of the angular momentum at the tropopause is the resting value of M at θ_m, since the circulation cannot access high value of M equatorward of θ_m. We do not here take the trouble to derive such a solution but it is a very straightforward extension of the development presented in section 5.2.1.

On the equatorward side of θ_m, the situation is quite different from the classical Hadley problem considered in the previous section. The RCE state will have a poleward gradient of s^*, with

easterly rather than westerly winds aloft. As long as M never decreases toward the equator, the RCE solution is viable; that is, we require that

$$\frac{dM_t^2}{dy} \leq 0, \tag{5.38}$$

where M_t is the angular momentum at the tropopause[1]. If we use (5.13) for M_t, then (5.38) can be written

$$1 + \frac{T_s - T_t}{\Omega^2 a^2} y^3 \frac{d^2 s^*}{dy^2} \geq 0. \tag{5.39}$$

(In deriving 5.39 we have neglected gradients in the difference between surface and tropopause temperature.) That is, the negative curvature of the distribution of s^* must not exceed a latitude-dependent critical value. If it does, a circulation in the vertical plane must ensue.

We could go about constructing the solution the same way we did in the equatorially symmetric case, but problems arise when angular momentum surfaces originating at the surface off the equator intersect the equator at some altitude below the tropopause. For this reason, and to illustrate a different approach to the problem, we proceed with a slightly different method based on an alternative assumption about the vertical distribution of s^*.

Instead of assuming moist adiabatic lapse rates (constant s^*) in the vertical, we assume instead that s^* is constant along M surfaces. This is actually a more general statement of convective neutrality that includes the centrifugal as well as the gravitational contribution to buoyancy (Emanuel, 1994) and is discussed in more detail in Chapter 9, section 9.2.1. Since, in any case, buoyancy gradients are weak, and since for simplicity we are here neglecting the contribution of water substance to density, it might be difficult to measure the difference between the two distributions in nature.

We begin by going back to the thermal wind equation (5.12). Since we are assuming that s^* is a function of M only we can write that equation as

$$\frac{\partial M^2}{\partial p} = 2a^2 \left(\frac{\partial T}{\partial p} \right)_{s^*} \left(\frac{ds^*}{dM^2} \right) \left(\frac{\partial M^2}{\partial y} \right). \tag{5.40}$$

By dividing through by $\partial M^2 / \partial y,$ we find an equation for the slope, in the $y - T$ plane, of angular momentum (and therefore s^*) surfaces:

$$\left(\frac{\partial y}{\partial T} \right)_M = -\frac{\partial M^2 / \partial T}{\partial M^2 / \partial y} = -2a^2 \frac{ds^*}{dM}. \tag{5.41}$$

[1] Recall that the variable *y* here is the square of the secant of the latitude.

Since s^* is only a function of M, its derivative with respect to M is also constant on surfaces of constant M, and so the slope of M surfaces is constant…they are just straight lines in the $y-T$ plane. We can integrate 5.41 upward from the surface to arrive at

$$y = y_s + 2a^2\left(T_s - T\right)\frac{ds^*}{dM^2}, \tag{5.42}$$

where the subscript indicates the (latitude-dependent) value at the surface. By using the definition of y given by (5.11), we can express this in terms of latitude:

$$\cos^2(\theta) = \frac{\cos^2(\theta_s)}{1 + 2a^2\cos^2(\theta_s)\left(T_s - T\right)ds^*/dM^2}. \tag{5.43}$$

Now if we know the RCE distribution of s^* with latitude at the surface, and assume that the zonal wind is very small at the surface, we can easily calculate ds^*/dM^2. As a particular example, suppose that

$$2a^2\frac{ds^*}{dM^2} = -\beta\left[\frac{\cos^2(\theta_s) - \cos^2(\theta_{sm})}{1 - \cos^2(\theta_{sm})}\right], \tag{5.44}$$

where β is a constant. This is a distribution in which s^* has its maximum value at latitude θ_{sm}. (The subscript *s* has been retained in (5.44) to remind us that this is along the surface.) Substitution of (5.44) in (5.43) would give us an explicit relationship for the shape of M and s^* in the latitude – temperature plane.

If all the M surfaces reach the tropopause before they reach the equator, then this "radiative-"slantwise-convective"[2] regime (hereafter RSCE) is viable, at least in the summer hemisphere. (Note that this will not be the case if there is a gradient of s^* across the equator.) In the above example given by (5.44) this will be the case if β is small enough. But if the surfaces reach the equator before they reach the tropopause (which will happen here when β exceeds a critical value), then for continuity across the equator the M (and s^*) surfaces must reconnect with their counterparts in the winter hemisphere, forming closes loops around a point at the surface and at the equator.

With the latitude of M surfaces given by (5.43) and using the definition of M, (5.3), we have

$$u = \frac{M - \Omega a^2\cos^2(\theta)}{a\cos(\theta)} = \Omega a\frac{\left[\cos^2(\theta_s) - \cos^2(\theta)\right]}{\cos(\theta)}, \tag{5.45}$$

where once again we have assumed that $u = 0$ at the surface. Substituting (5.43) into this expression, and specifying the RCE distribution of s^* (for example, relation (5.44)) gives an explicit expression for the distribution of zonal wind in the latitude – temperature plane.

[2] In which moist convection is assumed to establish moist adiabatic lapse rates along angular momentum surfaces

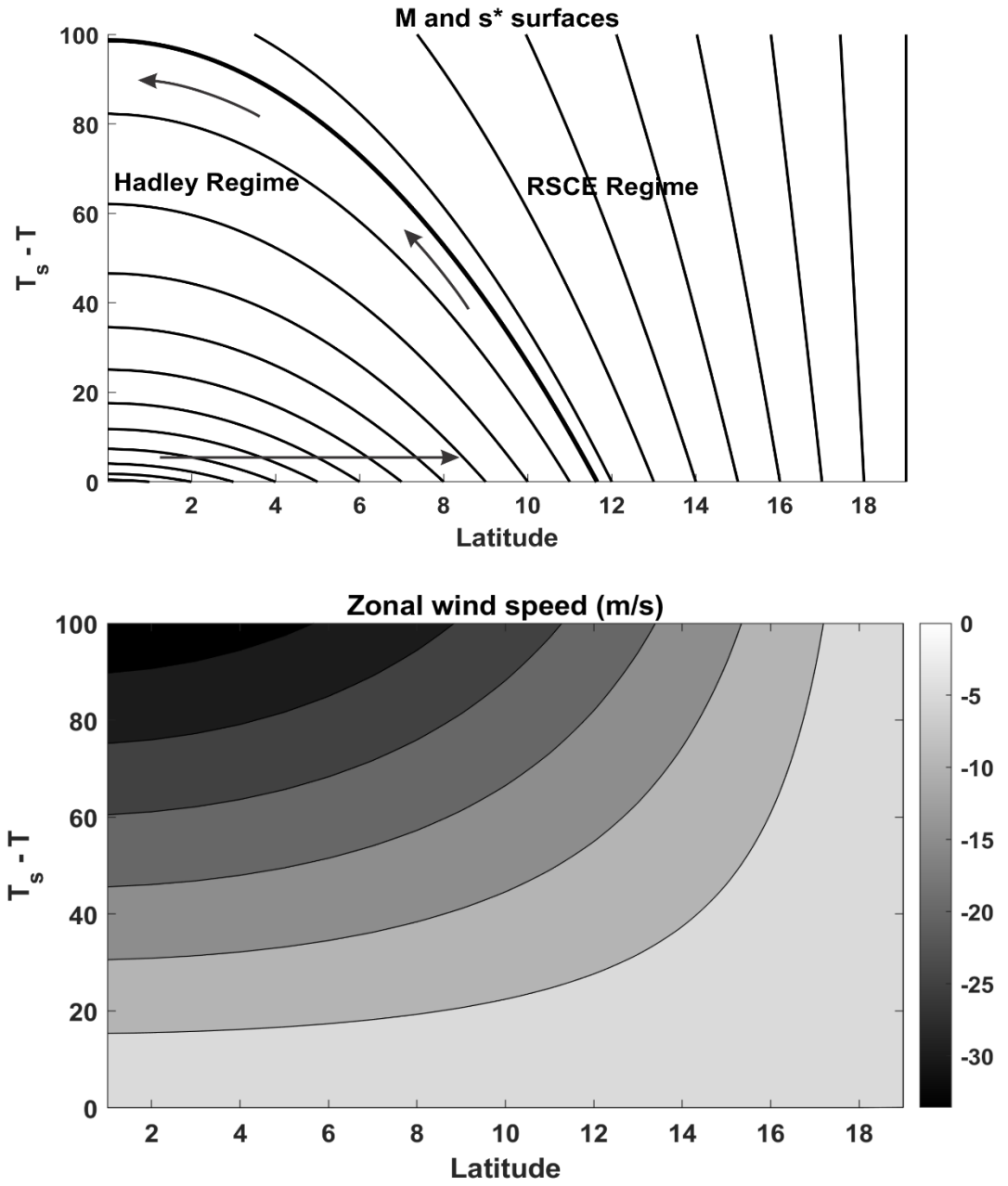


Figure 5.24: Top: Configuration of M and s surfaces as a function of latitude and the difference between the surface and local temperature (a proxy for altitude) as solutions for (5.43) – (5.45) with* $\beta = 0.0007$ *and* $\theta_{sm} = 19^{o}$. *The bold curve shows the transition between Hadley and RCE regimes. "RSCE" stands for "radiative-slantwise-convective equilibrium". Bottom: Zonal wind distribution.*

The configuration of M surfaces (which are also surfaces of constant s^*), given by (5.43), and the associated zonal wind (given by (5.45)) are shown in Figure 5.24 for the chosen distribution of s^* at the surface given by (5.44), and with $\beta = 0.0007$ and $\theta_{sm} = 19^o$. These parameter values give a supercritical distribution of s^*, so that some M surfaces intersect the equator before they reach the tropopause. Poleward of the bold curve in the top panel of Figure 5.24, the RSCE regime is viable, but equatorward of this curve the M (and s^*) surfaces intersect the equator. This leaves hanging the question of what happens on the cold side of the equator.

We have already seen that the thermal wind relation demands that the density (or s^*) distribution across the equator is locally symmetric and that, coupled with the fact that the RSCE temperature increases northward across the equator in our example, implies that the boundary layer moist entropy is too small for deep convection in the winter hemisphere. In the absence of deep convection, there must be mean subsidence to balance the radiative cooling of the atmosphere, and this implies a mass circulation with descent in the winter hemisphere and ascent in the summer hemisphere. Knowledge of the radiative cooling rate in the winter hemisphere is sufficient to calculate the mass streamfunction. In Figure 5.24 we indicate the mean ascent in the summer hemisphere. If we assume that the surface zonal winds are not very strong in the winter hemisphere, then the poleward extent of the Hadley regime in that hemisphere should match that of the summer hemisphere. Thus, above the subcloud layer, the winter hemisphere should be roughly mirror symmetric around the equator; the difference is that owing to the diminished radiative forcing in the winter hemisphere, its subcloud layer is relatively cool and stable to convection, so that the radiative cooling is everywhere balanced by subsidence.

While the assumption of relative weak surface winds requires M surfaces to re-enter the winter hemisphere boundary layer at roughly the mirror image latitude that they originate in the summer hemisphere, they may take a different path to get there. Moreover, the absence of deep convection in the winter hemisphere removes the constraint that M and s^* surfaces should be congruent. Together with the requirement that the winter hemisphere Hadley regime should blend smoothly into an RCE (or RSCE) regime in the winter hemisphere, and that the latter will be colder than its counterpart in the summer hemisphere, this suggests that the winter hemisphere temperature distribution will begin to depart from its mirror image in the northern hemisphere toward the poleward boundary of the Hadley regime.

Another consideration is that the circulation itself affects the RSCE solution of the summer hemisphere by advecting moist entropy poleward in that hemisphere. This effect was found to be quantitatively important by Privé and Plumb (2007a). The advection will reduce the boundary layer moist entropy (and therefore the free troposphere s^*) in the region of steep poleward gradients of the original RSCE solution, concentrating the gradients (and therefore the large-scale ascent) into a narrower band of latitude further poleward.

These effects are evident in aquaplanet simulations of solstitial Hadley circulations using full-up primitive equation models in moist atmospheres. An example of such a simulation, using a zonally symmetric model (Pauluis, 2004), is shown in Figure 5.25.

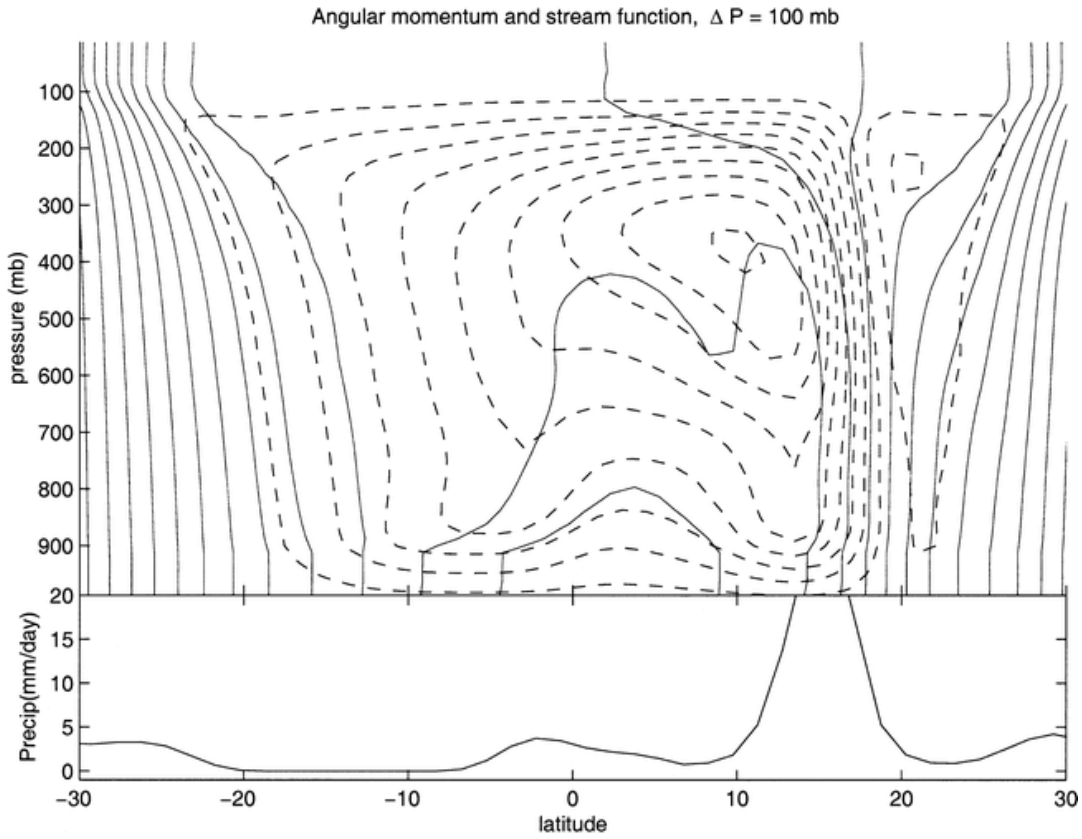


Figure 5.25: Streamfunction (dashed) and absolute angular momentum (solid) contours from a zonally-symmetric simulations of northern hemisphere summer conditions using an aquaplanet model. Precipitation is shown in the bottom panel.

There are several notable features of this simulation. First, the circulation above the boundary layer is approximately angular momentum conserving, as we assumed in our simple model and in contrast to the equatorially symmetric Hadley circulation. But, as seen in Figure 5.17, the real solstitial circulations are far from angular momentum-conserving, as indicated also by the large convergence of eddy angular momentum fluxes in the poleward branches of real circulations (Figure 5.23). Also, the ascent is indeed more concentrated than the descent, and the latter is mostly sloped the opposite way as it would be with a purely equatorially symmetric angular momentum distribution as in our simple model, consistent with the idea that the winter hemisphere temperature distribution should meld gracefully into the RSCE solution at higher latitudes.

But there is another feature of the solution in Figure 5.25 that we had not anticipated: the peculiar "jump" of the circulation across the equator, with low-level ascent in the near equatorial winter hemisphere (also evident in a minor maximum of precipitation there) and descent in the summer hemisphere. Why should that happen?

As first demonstrated by Pauluis (2004), the jump is nature's way around a dynamical conundrum, which can be seen through the lens of the meridional momentum equation. If we go back to (5.5) and relax the balance assumption, then we can write it as

$$\frac{v}{a}\frac{\partial v}{\partial \theta}+\left(\frac{\partial \phi}{a\partial \theta}\right)_p - g\frac{\partial \tau_\theta}{\partial p} = -2\Omega \sin\theta u - \frac{u^2}{a}\tan\theta, \tag{5.46}$$

where τ_θ is the meridional stress and we have neglected the vertical advection of v. The third term on the left of (5.46) is the meridional acceleration owing to the vertical convergence of meridional stress; we have neglected the horizontal convergence of stress as is traditional in boundary layer theory. If we also neglect the last term in (5.46) compared to the penultimate term, eliminate ϕ between (5.46) and the hydrostatic equation (5.7), and use Maxwell's relation as before, we get

$$\frac{\partial}{\partial p}\frac{1}{2}\left(\frac{\partial v^2}{\partial \theta}\right)-\left(\frac{\partial T}{\partial p}\right)_{s*}\frac{\partial s*}{\partial \theta}-g\frac{\partial^2 \tau_\theta}{\partial p^2}=-2\Omega a\sin(\theta)\frac{\partial u}{\partial p} \tag{5.47}$$

Finally, we integrate (5.47) across the depth of a mechanical boundary layer whose top is defined as the altitude at which τ_θ and its vertical derivative vanish. We also assume vanishing surface zonal wind as before. The result is

$$\frac{1}{2}\frac{\partial}{\partial \theta}\left[v_s^2-v_b^2\right]-\left(T_s-T_b\right)\frac{d\overline{s*}}{d\theta}-ga\frac{\partial \tau_\theta}{\partial p}=2\Omega a\sin(\theta)u_b, \tag{5.48}$$

where the subscript b denotes evaluation at the top of the boundary layer, and $\overline{s*}$ is the vertically averaged value of boundary layer saturation entropy. Re-ordering the terms in (5.48) and using the aerodynamic flux formula for the surface stress gives

$$\frac{1}{2}\frac{\partial}{\partial \theta}\left[v_s^2-v_b^2\right]+\frac{ga\rho}{\Delta p}C_D\,|\mathbf{V}|\,v-\left(T_s-T_b\right)\frac{d\overline{s*}}{d\theta}=2\Omega a\sin(\theta)u_b, \tag{5.49}$$

in which Δp is the pressure depth of the mechanical boundary layer.

Even though there can be no gradient of free tropospheric $s*$ across the equator, there can be a temperature gradient in the boundary layer if the boundary layer is locally decoupled from the free troposphere.

The first term in (5.49) is likely to be small at the equator owing to symmetry and relatively small wind shear across the boundary layer, while the last term in (5.49) vanishes at the equator. Thus, to a good approximation, on the equator

$$ga\rho C_D\,|\mathbf{V}|\,v=\Delta p\left(T_s-T_b\right)\frac{d\overline{s*}}{d\theta}. \tag{5.50}$$

If the boundary layer temperature gradient is small, or the pressure depth of the boundary layer is small (or both), the meridional flow will be very small at the equator. Essentially, in the absence of Coriolis torques, there is not enough meridional pressure gradient in the boundary layer to drive meridional flow against friction. For this reason, the flow turns upward on the cold side of the equator and flows toward the summer hemisphere in an angular-momentum-conserving loop, descending back into the boundary layer where it can resume its poleward flow across M surfaces, thanks to torque acting on surface westerlies.

Secondary maxima in precipitation on the opposite side of the equator from the main ITZC can be observed in the precipitation fields displayed in Figure 5.9, for example in the South Indian Ocean in July and in the eastern tropical South Pacific in April.

Finally, given the boundary layer meridional flow, the surface zonal wind can be diagnosed from the Ekman layer relations given by (5.34) and (5.35). This will yield easterlies in the winter hemisphere and westerlies in the summer hemisphere Hadley regimes, with very weak zonal winds near the equator.

5.2.3 The time-dependent Hadley Circulation

In Chapter 3 we showed that the time scale over which single columns with coupled slab oceans relax to their RCE state is of order 1-2 years, obviously not short compared to a seasonal cycle. For this reason, we cannot expect the Hadley circulation ever to be in an equilibrium state in the real world. There are several consequences of not being in equilibrium. First, storage of heat in the oceans becomes important, and the wind-dependence of turbulent heat fluxes from the ocean becomes a significant factor in determining the moist entropy of the boundary layer and thus the free tropospheric temperature. Also, storage of angular momentum in the atmosphere can be significant compared to advection, so flow across M surfaces does not necessarily indicate non-conservation.

Some essential features of the seasonal variation of the Hadley circulation were examined by Boos and Emanuel (2008) using a zonally symmetric moist GCM with interactive radiation and parameterized moist convection. They specified a sea surface temperature distribution in latitude that was constant everywhere except between 10° and 40° north latitude, where the distribution was given by

$$T_s = T_0 + \Delta T \cos^2\left(\frac{\pi}{2}\frac{\theta-\theta_0}{\Delta\theta}\right)\cos\left(\frac{2\pi t}{\tau}\right), \tag{5.51}$$

where T_0 is the constant background temperature, ΔT is the amplitude of the anomaly, θ_0 is 25°, and $\Delta\theta$ is 15°. The anomaly oscillates over a year, with $\tau = 365\ days$. (Note that there is no counterpart anomaly in the southern hemisphere.) The results for a simulation with $\Delta T = 3\ K$ are summarized in Figure 5.26. This is a phase diagram depicting the meridional mass flux in the boundary layer (ordinate) against the time-varying amplitude of the imposed sea surface temperature anomaly (abscissa), with the phase moving clockwise around the closed curves. The light gray curve is for a simulation in which the surface enthalpy flux is independent of wind speed, while the solid black curve is for identical conditions but with the full wind-dependence of the surface fluxes included.

If the system were moving through a sequence of steady states, the phase diagram would be a single fixed curve rather than a loop and the system would just migrate back and forth along the same curve. The “thickness” of the loop is a measure of how far out of equilibrium the system is.

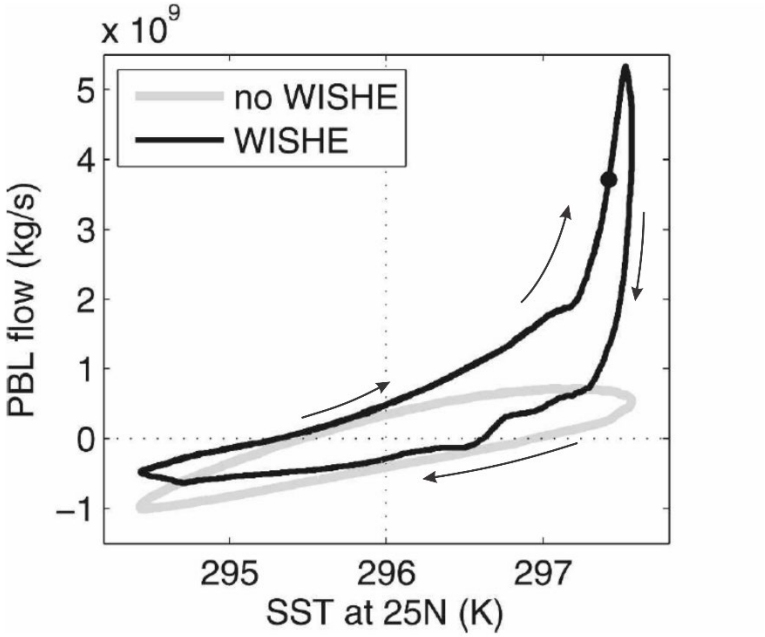


Figure 5.26: Phase diagram of meridional flow in the boundary layer (PBL flow) versus the time-varying amplitude of the imposed sea surface temperature anomaly, whose peak is at 25º N. The curved arrows denote increasing time, so the phase moves clockwise around the closed curves. The light gray curve is for a simulation in which the surface enthalpy flux is independent of wind speed while the solid black curve is for identical conditions but with the full wind-dependence of the surface fluxes included.

In springtime equinoctial conditions, the flow is toward the warming hemisphere while at the fall equinox the flow is in the opposite direction.

Given the shape, latitude, and magnitude of the imposed sea surface temperature anomaly at its summer peak, the gradient of sea surface temperature, were it matched by the gradient of s^* in the free troposphere, would be in this case insufficient for angular momentum surfaces to "breakthrough" to the equator, and indeed the simulation with wind-independent surface enthalpy fluxes is weak. (Note, however, that the time dependence all by itself requires a meridional circulation to adjust the atmosphere toward the time-varying equilibrium state.) However, when the enthalpy fluxes are allowed to depend on wind, as they do in nature, a strong feedback ensues in which stronger poleward winds (and their zonal counterpart) give rise to stronger surface fluxes, which, together with meridional entropy advection in the boundary layer, steepen the boundary layer moist entropy gradient to the point of super-criticality, and a much stronger flow ensues. This feedback is often referred to as "Wind-dependent Surface Heat Exchange (WISHE)" feedback and will see that it plays an important role in many tropical weather systems.

While simulations like these with specified (though time-varying) sea surface temperature illustrate important processes such as WISHE, they do not satisfy the energy budget of the underlying ocean, implying very large and unrealistic heat fluxes within the ocean. The real system is strongly coupled, with the thermal inertia of the ocean mixed layer and lateral heat fluxes in the ocean playing important roles. Real Hadley circulations are coupled ocean-atmosphere phenomena.

You can further explore the nature of Hadley circulations, including thermodynamic coupling to a slab ocean, by running a zonally symmetric primitive equation model, using the package HadleyPak, available at https://zenodo.org/doi/10.5281/zenodo.10815377 . Here you may explore simulated Hadley circulations under a variety of conditions, including time-dependent and time-independent solar forcing, the effect of the interaction of radiation with clouds, coupling with a slab ocean, and other options. Figure 5.27 shows a time-latitude series of rainfall using an annual cycle of radiation and an ocean mixed layer depth of *25 m*. With this relatively small mixed layer depth, summertime precipitation extends all the way to about 30^{o} latitude, while a near-equatorial maximum can be found at most times of the year, owing to the previously discussed jumping across the equator. This gives a triple peak to the annually averaged rainfall.

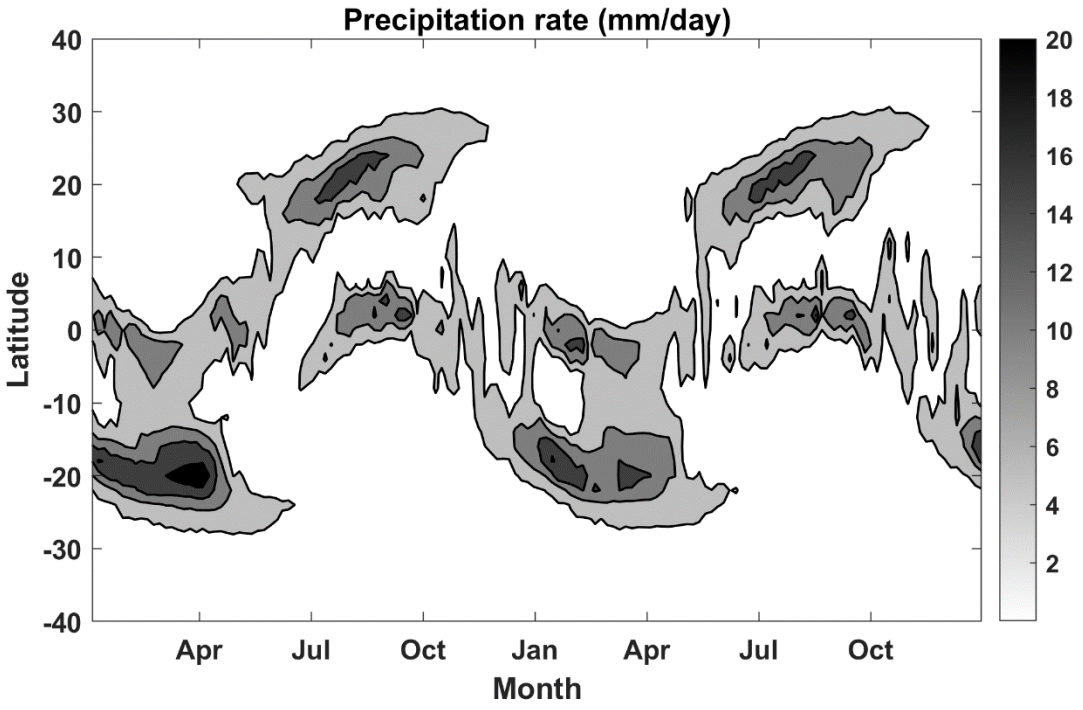


Figure 5.27: Precipitation as a function of latitude and time over two full annual cycles beginning on January 1, as output from a zonally symmetric primitive equation model coupled to a slab ocean of 25 m depth.

5.3 Monsoons

Monsoons are seasonally varying thermally direct circulations driven by land-sea contrasts, aided in some cases by the elevation of land surfaces. (In popular parlance, the term 'monsoon' often refers to the heavy rains associated with the summer monsoon.) They favor subtropical continents located adjacent to oceans; here the seasonal forcing is large but the Coriolis parameter is relatively small, favoring unbalanced circulations. Besides their inherent scientific interest, monsoons and their variability strongly influence the socioeconomics of the regions they affect.

Figure 5.28 shows an index of monsoon activity, defined as the difference between July and January monthly mean precipitation, shown only where magnitude of the difference exceeds *5 mm day^{-1}* and where the summer precipitation exceeds 55% of the sum of the summer and winter precipitation. In the northern hemisphere, strong monsoons are found in a region containing India, southeast Asia, and the Bay of Bengal, in sub-Saharan Africa and over the adjacent tropical North Atlantic, and in Mexico and the adjacent eastern North Pacific.

Monsoon conditions in the southern hemisphere are found in the Amazon basin, across northern Australia and surrounding waters and across subtropical southern Africa, including Madagascar.

In the rest of this section we focus on the Indian monsoon, as it is arguably the most studied of the monsoons, and many of its dynamical aspects carry over to other monsoon systems.

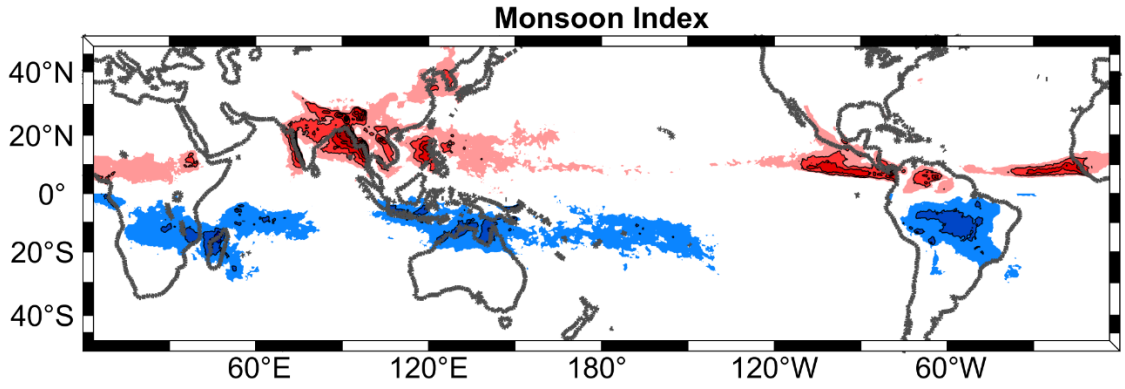


Figure 5.28: July minus January precipitation (shading) from NASA's IMERG data based on a constellation of satellites augmented where possible with rain gauge data, 2002-2019. The quantity is plotted only where the absolute value of the difference exceeds 5 mm day^{-1} and where the summer precipitation exceeds 55% of the summer+winter precipitation.

Monthly mean rainfall over the Indian subcontinent is shown in Figure 5.29 for January and July. Clearly, almost all of India's rain occurs during the summer monsoon, with dry conditions prevailing in the winter. During the summer monsoon, rainfall is concentrated along the western slopes of the western Ghat Mountatins that run along India's west coast, reflecting the forced upslope flow in the strong southwesterly winds that prevail in the summer monsoon (Figure 5.30; see also Figure 5.5 for a larger scale map of surface winds). Note also the pronounced rain shadow in the lee of the Ghats.

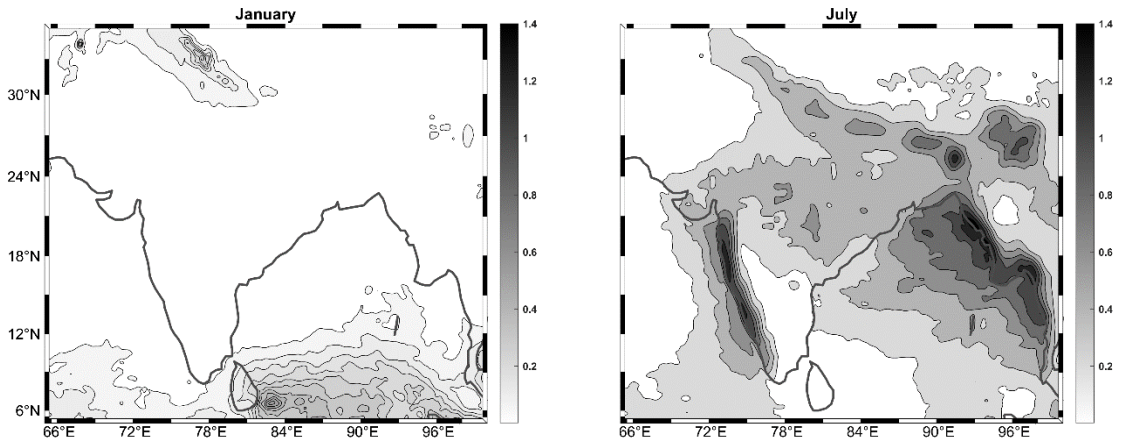


Figure 5.29: Monthly mean rainfall rate (mm/hr) in January (left) and July (right) from NASA's IMERG data based on a constellation of satellites augmented where possible with rain gauge data, 2002-2019.

Summer monsoon rainfall increases eastward across the Bay of Bengal reaching a maximum along the western coast of Myanmar. This reflects increasing boundary layer moist entropy as the strong southwesterly winds cross the Bay. Another rainfall maximum is found along the foothills of the Himalaya, forced in part by upslope flow.

Surface winds also dramatically differ between winter and summer (Figure 5.30), with offshore flow in January and remarkably strong southwesterly onshore flow in July. Note also the strong apparent convergence of the flow in western India and coastal Myanmar in July.

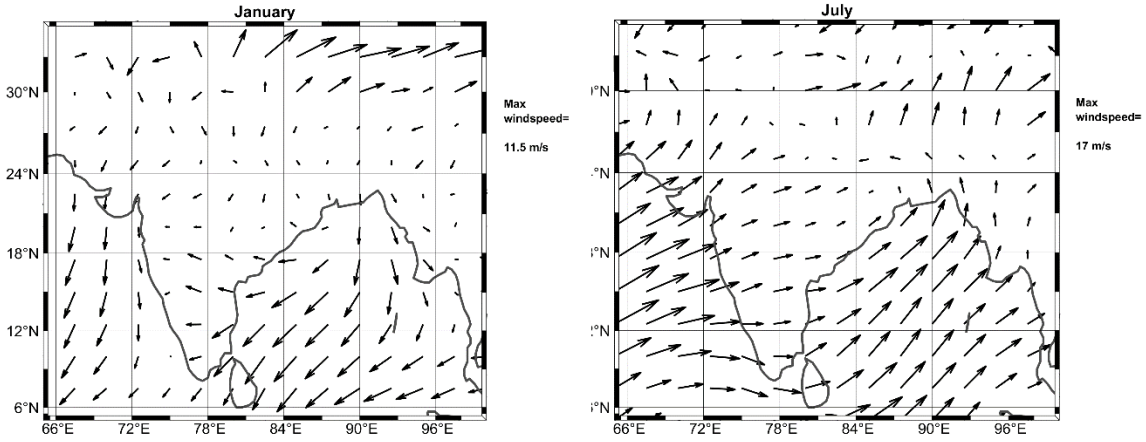


Figure 5.30: 1000 hPa winds from NCAR/NCEP reanalyses averaged over 1979-2019 in January (left) and July (right).

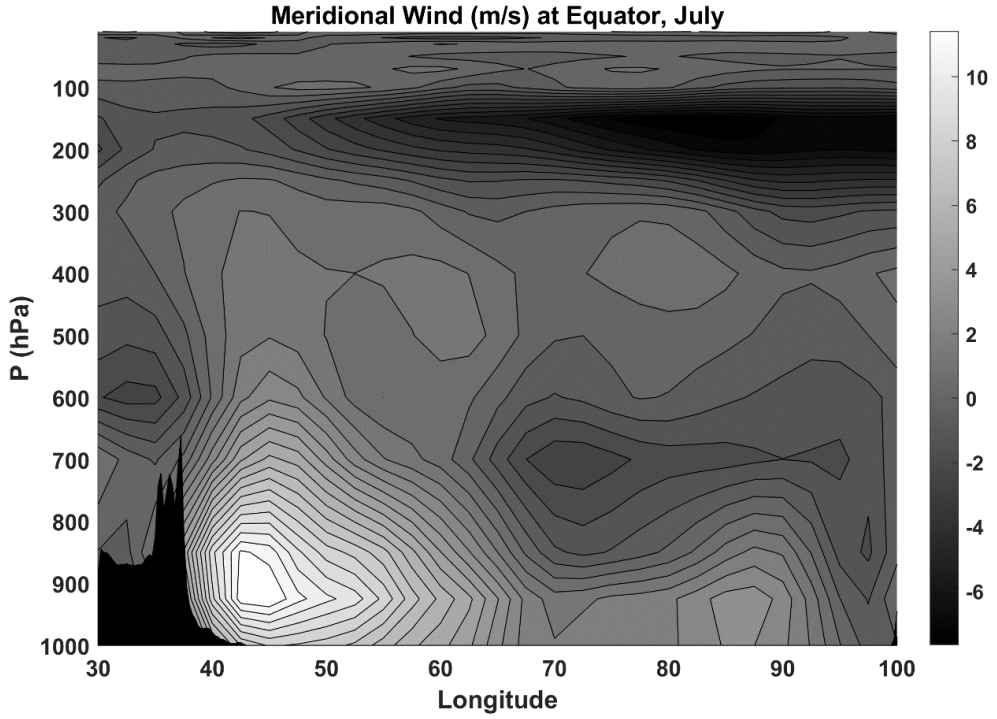


Figure 5.31: Cross-section of meridional wind (m/s) across the equatorial Indian Ocean in July, from NCAR/NCEP reanalysis data, 1979-2019. The topography of equatorial Somalia and Kenya is shown at lower left.

As can be seen in Figure 5.5, the Indian summer monsoon circulation extends well into the southern hemisphere. The meridional flow is strongly concentrated into a low-level jet hugging the mountains of Somalia and Kenya. This is often referred to as the Somali Jet or the Findlater Jet (Findlater, 1969) and apparently results from the same western boundary current intensification physics as are responsible for strong ocean currents such as the Gulf Stream and Kuroshio Current (Anderson, 1976). North of the equator the strong surface flow parallel to the East African coastline drives powerful upwelling in the Arabian Sea, leading to a dramatic cooling of ocean waters in the western part of the Sea. This cooling has strong effects on marine biology and stabilizes the atmospheric boundary layer, with possible feedbacks on the monsoon circulation.

Figure 5.32 shows meridional cross-sections of flow in the latitude-pressure plane in January and July, averaged over longitudes 45W – 95W. Note that the flow vectors are not streamlines as the flow is not non-divergent in this plane. The January circulation resembles a simple cross-equatorial overturning, with relatively cool, dry air sinking over the southern slopes of the Himalayas, flowing southward across the equator, and ascending near 10^{o}S. Note that the descending branch slopes down toward the south, conforming to the northward-leaning angular momentum surfaces (not shown) in the northern hemisphere.

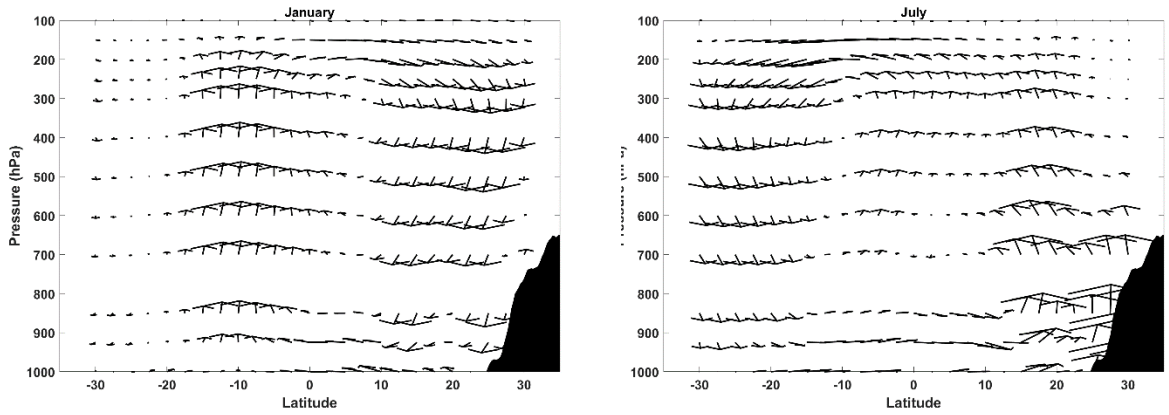


Figure 5.32: Cross-sections of circulation in January (left) and July (right), from 30°S to 30°N and averaged between 45°E and 95°E, from NCAR/NCEP reanalysis data, 1979-2019. The arrows are based on reanalysis meridional and pressure (omega) velocities. Topography shown at lower right.

The July circulation is somewhat more complex. There are three areas of ascent: Shallow ascent along the slopes of the Himalaya, deep ascent centered near 18°N and another area of deep ascent just south of the equator. Descent extends poleward of about 12°S. The circulation shares some features with model solutions sown in Figures 5.25 and 5.27, with the return branch jumping over the equator for reasons explained in section 5.2. The ascent south of the equator is associated with a secondary peak in precipitation, as shown in Figure 5.33, even more prominent than that in the zonally symmetric simulation in Figure 5.25; in fact, the secondary peak rainfall is fully 75% of the primary peak.

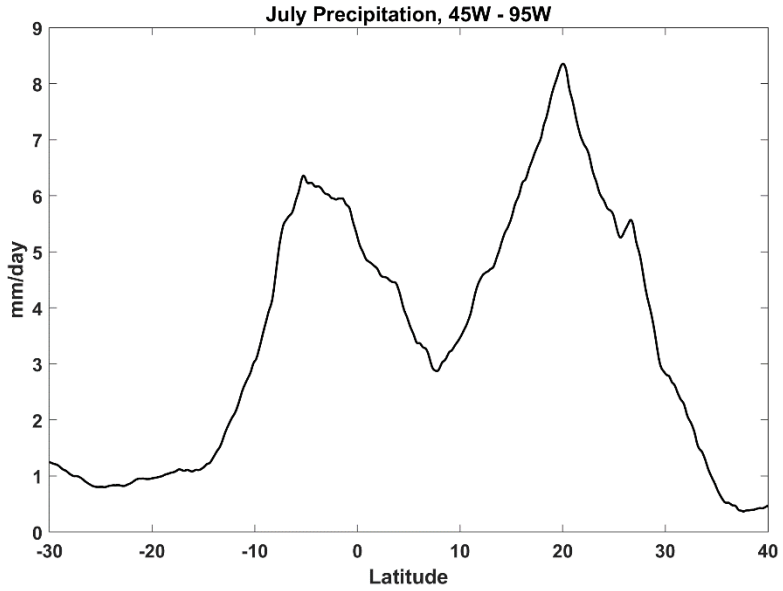


Figure 5.33: July average precipitation, zonally averaged from 45°W to 95°W using NASA's IMERG data based on a constellation of satellites augmented where possible with rain gauge data, 2002-2019.

5.3.1. Zonally symmetric theory and the effects of land

The Indian monsoon is hardly zonally symmetric, occupying about 100 degrees of longitude, but zonally symmetric theory is nevertheless a useful starting point. In essence, the zonally symmetric problem is similar to the seasonally varying Hadley Circulation covered in the last section, but with one wrinkle: the surface north of some latitude is land rather than sea.

The introduction of a land surface changes the problem in interesting ways. First, the effective heat capacity of land is much smaller than that of the ocean. On seasonal time scales, heat transfer to and from the ocean surface is transmitted by ocean turbulence through the mixed layer, but in the tropics, away from areas of deep water formation, little heat is exchanged with the deep ocean. Therefore, the heat capacity of the sea approximately corresponds to that of the mixed layer. Heat is diffused very slowly through most land surfaces, by contrast, although on seasonal time scales one can see temperature variations down to a meter or so. In some cases, wet soil can diffuse heat much faster than when the same soil is dry, so the first rains of the monsoon can tap into heat stored in the soil.

Focusing on the different heat capacity of land (in isolation from other effects), the theory for relaxation to RCE states developed in Chapter 3 predicts a relaxation time scale given by (3.48), as a function of the climate sensitivity of the system, the individual heat capacities of the ocean and atmosphere, temperature, and surface wind speed. For the typical values of these parameters given in Chapter 3, an ocean of depth 1 m yields an effective relaxation time scale of about 150 days, whereas one of 50 m depth has a time scale of around 750 days – 5 times longer. But note that the RCE relaxation time scale over a surface with heat capacity characteristic of wet land is still not short compared to a year. Thus, one cannot assume that the system over land is close to an RCE state even neglecting horizontal heat transport.

Specific characteristics of the land surface, including albedo and moisture availability, will also have potentially strong effects on the RCE state over land, changing the net radiative forcing of the monsoon. All other things being equal, a dry land surface will have a drier RCE state, with less greenhouse effect of water vapor and clouds, but more surface solar radiation owing to fewer clouds. The subcloud layer will be deeper, with its dry adiabatic lapse rate yielding a higher surface temperature. The higher temperatures of the lower atmosphere coupled with lower temperatures aloft also change the character of the monsoon forcing.

A comparison of RCE calculations with moist and dry surfaces is shown in Figure 5.34, using the single-column model discussed in Chapter 3 and available at https://zenodo.org/doi/10.5281/zenodo.10795221 . In these calculations, the dry surface is identical to the moist surface except that the surface evaporation rate from the aerodynamic flux formula is reduced by 80%. The panel on the left shows the two simulations with no clouds while the right panel compares two simulations with interactive clouds. Without clouds, the reduction of water vapor when the surface is dry cools the whole troposphere, although the surface temperature actually increases by 1 K and the lowest 50 hPa is almost as warm as the moist surface case. With interactive clouds, the effect on atmospheric temperature of the reduction of water vapor is essentially cancelled by a large reduction in low clouds (not shown), so the column temperatures are nearly the same except in lower troposphere, where the deeper dry-adiabatic layer yields a warmer profile.

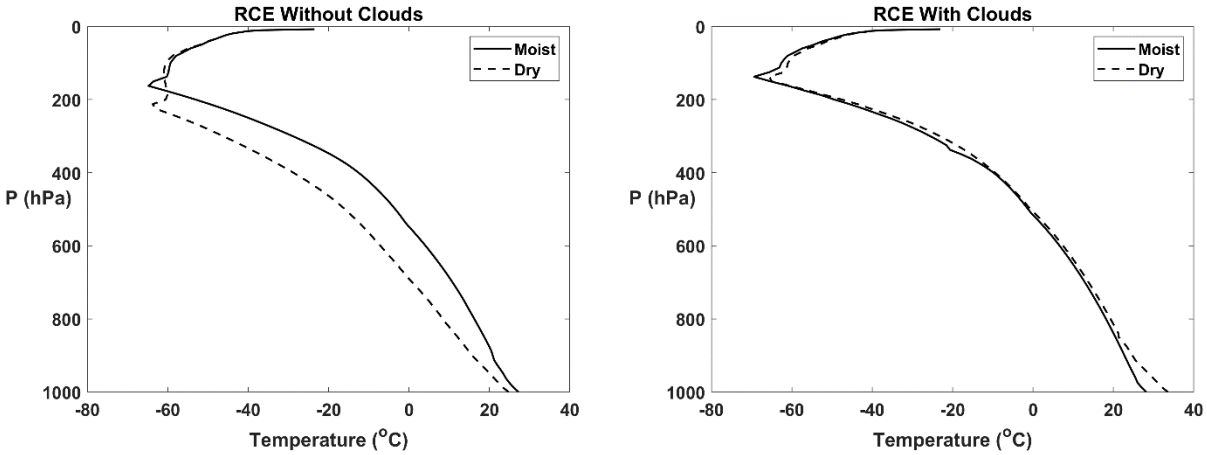


Figure 5.34: RCE calculations of the vertical temperature profile made with the single-column model with a slab ocean surface (solid) and a land surface (dashed) identical to the ocean except that the surface evaporation rate is reduced by 80%. Panel at left compares two simulations without clouds while right panel shows simulations with interactive clouds.

While this shows that interactive clouds can strongly affect the sensitivity of RCE states to surface water availability, the details of the calculations shown here are sensitive to many assumptions that have to be made about how clouds work in a single column, and so the details are likely to be wrong.

Generally, land surfaces have higher albedo than water, and this acts to diminish the RCE state temperature and thus the monsoon forcing. For example, it has long been recognized that Indian monsoon rainfall tends to be weaker in years with larger than normal snowfall in Eurasia (e.g. Blanford, 1884).

Other key elements at work on land are mountains and plateaus. Mountain ranges can block the lateral advection of quantities like entropy and angular momentum, and elevated plateaus often have different RCE states owing in part to their elevation itself.

To understand this last point, consider what happens to RCE surface temperature as the altitude of the land surface increases. We will assume that this happens over a large enough area that the new RCE state would not be immediately obliterated by lateral advection from surrounding areas, but not so large that the surface pressure would fail to drop (as would be the case in the absurd limit of the whole surface of the globe expanding upward.)

First, consider the optically thin limit of an atmosphere with a very small infrared optical depth. In that case, reducing the surface pressure (i.e. elevating the land mass) would not result in any appreciable change in the infrared opacity of the atmosphere and the surface temperature would remain nearly unchanged. In this case, if we compare the RCE temperature over an elevated surface to the RCE temperature over an adjacent surface at sea level, the temperature over the elevated surface will be larger *at the same pressure level*. At the opposite extreme, in an atmosphere with very large infrared optical depth, reducing surface pressure (i.e. elevating the surface) will reduce the infrared optical depth simply because the mass of infrared absorbers above the surface is smaller. Therefore, elevating a surface in this case will result in a decrease in surface temperature. In general, if the conditions are not close to the limit of a runaway greenhouse, the decrease in surface temperature with elevation will not be so rapid as

along a moist adiabat, and it will continue to be the case that the temperature over the elevated surface will be greater than that over sea level, *at the same pressure (Molnar and Emanuel, 1999)* .

These inferences are confirmed by running RCE calculations over surfaces at various elevations, using the single-column model described in Chapter 3. For these simulations we hold the surface properties other than elevation constant. Figure 5.35 compares the RCE temperature profile over a water surface at 1000 hPa with one over a water surface at 600 hPa. At all pressure levels in the troposphere from 600 hPa up the temperature over the elevated surface is higher than that over sea level at the same pressure.

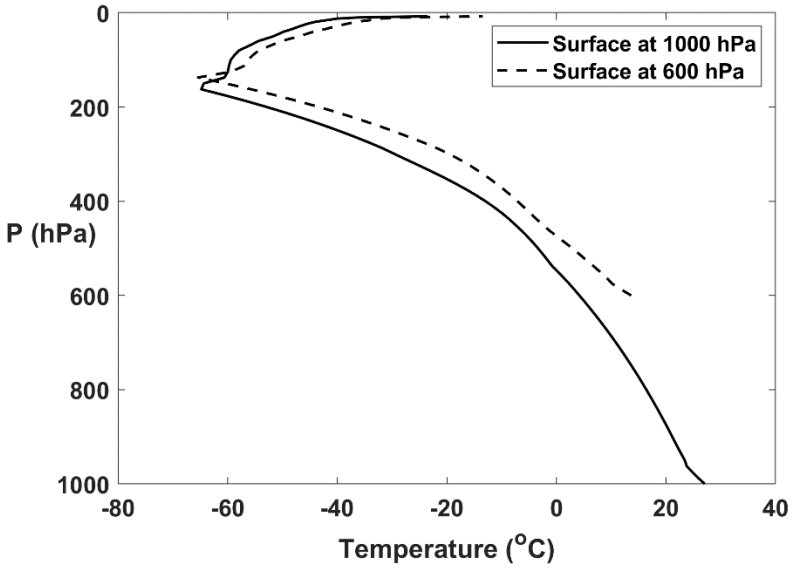


Figure 5.35: RCE temperature profiles over a water surface at 1000 hPa (solid) and at 600 hPa (dashed), calculated using the single-column model described in Chapter 3.

The Tibetan Plateau is a large area of terrain with mean surface pressures between 600 and 700 hPa, and although its albedo is somewhat higher than that of India and surrounding ocean, it is has nevertheless long been regarded as an important forcing of the Indian monsoon. The actual temperatures in the upper troposphere over the plateau are higher than elsewhere, giving rise to an extensive upper tropospheric anticyclone known as the *Tibetan High*. The July mean 200 hPa heights are shown in Figure 5.36. The highest heights are over the southern portion of the plateau, but they extend considerably southwestward of the highest topography. The westward extension of the anticyclone over the Middle East is a dynamical consequence of the forcing, as explained by Rodwell and Hoskins (2001).

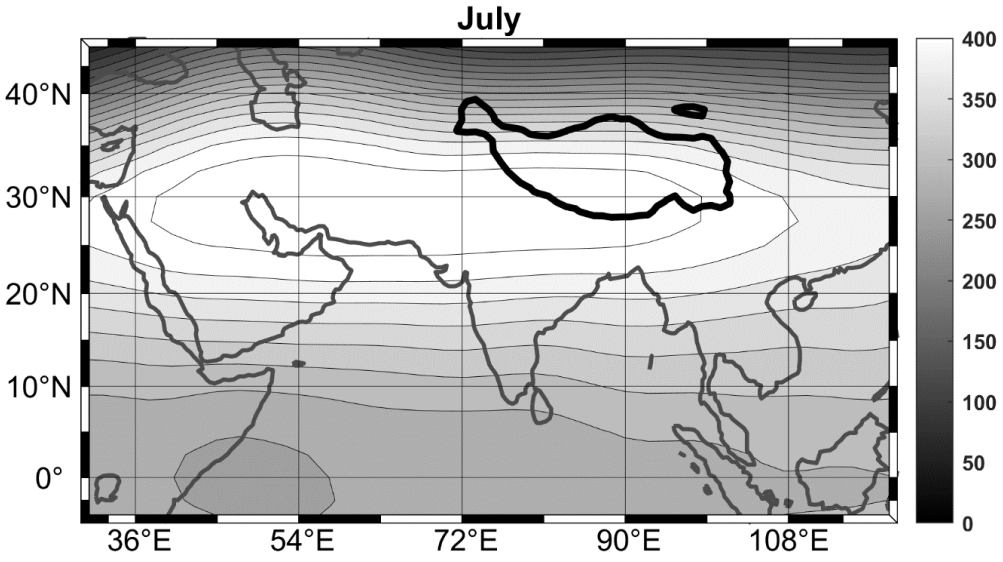


Figure 5.36: Mean 200 hPa height anomaly (m) in July, from NCAR/NCEP reanalysis data averaged over 1979-2019. The bold black curve is the 4000 m topographic contour.

The thermodynamic influence of the regional topography can be seen in the July mean fields of surface potential and equivalent potential temperature in Figure 5.37. The dry entropy (potential temperature) has a decided maximum over the plateau itself, while the moist entropy (equivalent potential temperature) has a broad maximum covering the northern Bay of Bengal, much of northern India and Bangladesh, and the southern part of the Tibetan Plateau. Note also high moist entropy values in the far northern Arabian Sea.

To the extent we can regard surface entropy as a reflection of the thermodynamic forcing of the monsoon, it paints an interesting and complex picture. Referring back to Figure 5.33 and the right panel of Figure 5.32, we see that the strongest ascent and largest precipitation rates are around 20^{o}N, along the southern periphery of the largest values of surface moist entropy, from the right panel of Figure 5.37. But there is a decided but somewhat shallower secondary maximum of ascent at 30^{o}N, along the southern periphery of the region of highest surface dry entropy, as depicted in the left panel of Figure 5.37.

This is precisely what we would expect given that over the central Tibetan Plateau, the dry entropy is a maximum but the moist entropy is not, reflecting the combined effect of elevation and surface entropy and wetness. The dry summertime soils of the Plateau produce deep subcloud layers with dry adiabatic lapse rates in the daytime; this together with the surface elevation produce a maximum in potential temperature at low levels. But owing to the relatively high albedo, the surface moist entropy is not as large as that found further south over the lowlands and Bay of Bengal, thus the upper troposphere over the Plateau would be cooler if the column is convectively neutral. Consequently, on the southern edge of the Plateau the low-level meridional temperature gradient would be positive, but the upper tropospheric temperature gradient would be negative.

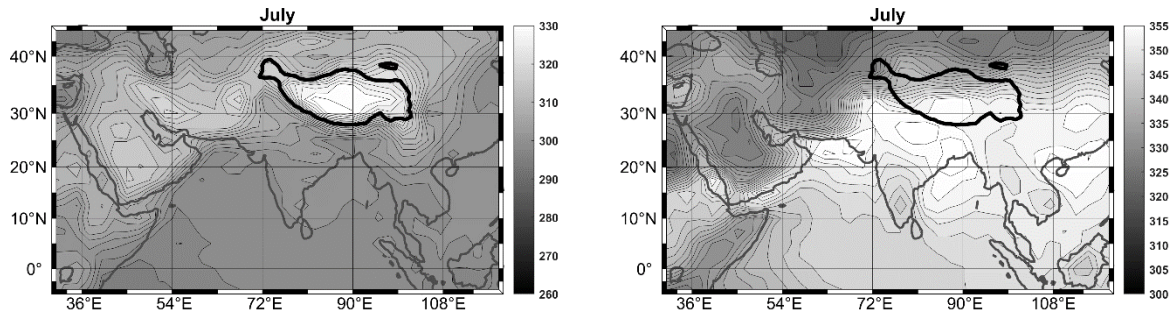


Figure 5.37: July mean surface potential temperature (left) and equivalent potential temperature (right) from NCAR/NCEP reanalyses, 1979-2019. The bold curve is the 4000 m topographic contour

Note from the right panel of Figure 5.32 that the monsoon circulation crosses the equator, even though the surface moist entropy gradient is very weak or non-existent there (right panel of Figure 5.37). We also notice that our steady, zonally symmetric theoretical solution for radiative-slantwise-moist-convective equilibrium (RSCE) given by (5.43) does not require a gradient of the RSCE entropy across the equator, only a sufficiently large poleward gradient somewhere in the system, with "sufficiently large" being a function of latitude. To illustrate this, we modify the trial function of $ds*/dM^2$ given by (5.44) to have a maximum value well off the equator:

$$2a^2\frac{ds*}{dM^2}=-2\beta\left[\frac{\cos^2(\theta_s)-\cos^2(\theta_{sm})}{1-\cos^2(\theta_{sm})}\right]\left[\frac{\sin(\theta_s)}{\sin(\theta_{sm})}\right]^2 \tag{5.52}$$

This function is plotted in Figure 5.38 for a case in which the moist entropy peaks at 26° latitude. The moist entropy gradient vanishes on the equator and peaks around 25°N.

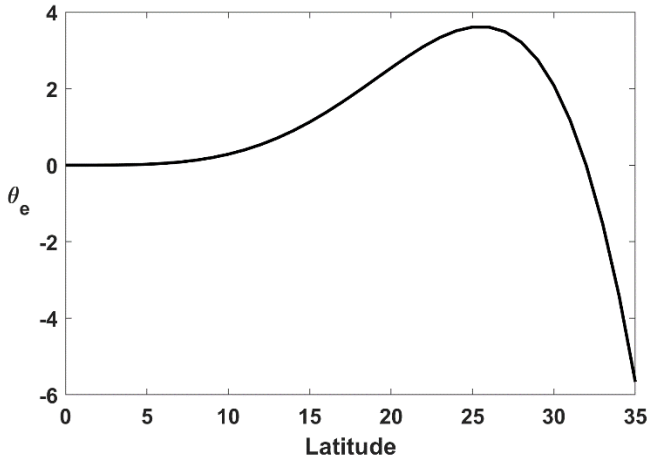


Figure 5.38: Equivalent potential temperature (K) relative to its equatorial value, as given by (5.52) with $\beta = 0.0015$ *and* $\theta_{sm} = 26^o$.

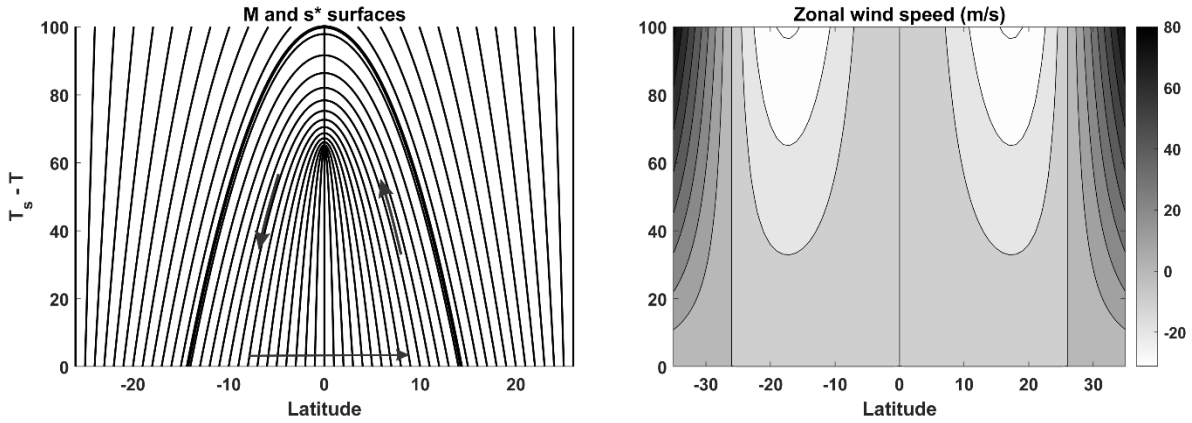


Figure 5.39: Left: Shape of $s*$ and M surfaces in the solution to (5.43) with the distribution of moist entropy given by (5.52) with $\beta = 0.0015$ and $\theta_{sm} = 26^{o}$. Bold line shows boundary of monsoon regime and arrows show sense of circulation. Right: The corresponding zonal wind speed (m/s). Note the different latitude scales of the two figures.

As before, all of the moist entropy and angular momentum surfaces within the monsoon regime, shown by the bold curve in the figure, cross the equator, but the altitude at which they do so decreases toward a finite limit as their latitude of origin approaches the equator. Here again we expect ascent north of the equator and descent in the southern hemisphere, with the possibility of an angular momentum-conserving jump in the northward flow across the equator. Note that the strongest ascent will tend to coincide more nearly with the latitude of strongest poleward entropy gradient, not the latitude at which the entropy itself has a maximum. The zonal wind distribution shows easterlies everywhere above the surface equatorward of the latitude of maximum RSCE surface moist entropy, and westerlies poleward of that latitude.

The main feature of the RSCE solution is that angular-momentum-conserving cross-equatorial flow does not require an RSCE entropy gradient at the equator itself, and so thermal forcing by land-sea contrast well off the equator can cause a strong thermally direct response.

A more realistic simulation of an idealized, zonally symmetric monsoon can be made using HadleyPak, available at https://zenodo.org/doi/10.5281/zenodo.10815377 . This is a zonally symmetric primitive equation model using the same radiation and convection schemes as used in the single-column model available at https://zenodo.org/doi/10.5281/zenodo.10795221 . For these simulations, we used a latitudinal resolution of 2 degrees, spanning 40°S to 40°N and having land poleward of 20°N. The ocean consists of a slab 35 m in depth and the land surface has a primitive soil moisture scheme that accounts for precipitation, evaporation, and runoff. The model does not include topography, however, and so cannot be used to explore the effects of elevated land.

Some fields produced by a simulation spanning two years are shown in Figure 5.40. The time-latitude display of precipitation in the left panel of the figure shows some interesting features. In contrast to the simulation shown in Figure 5.27, which differs only in having an ocean surface throughout the domain, precipitation over the land begins just a few weeks after the spring equinox and migrates northward in response to solar forcing over land. Just offshore, the onset of monsoon precipitation is delayed by the downward motion associated directly with the land-sea low-level temperature contrast. This is essentially a sea breeze circulation with descent just offshore.

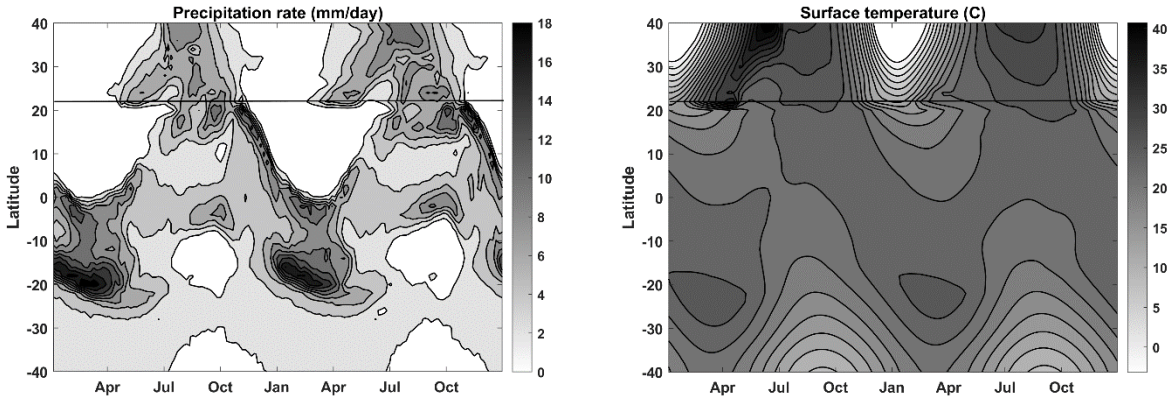


Figure 5.40: Time-latitude section of precipitation (left, in mm/day) and surface temperature (right, ^{o}C) for a simulation forced by seasonally varying TOA insolation. The horizontal line marks the boundary between land to the north and sea to the south.

But in September, as the monsoon recedes, land breezes cause a local maximum in ascent and precipitation just offshore. In October, a narrow band of heavy rainfall forms just offshore and moves southward, reaching a southern extremity in late winter just south of the equator. A separate precipitation maximum occurs in the southern hemisphere summer in response to insolation forcing over the ocean.

The evolution over the two-year period of surface temperature is shown on the right side of Figure 5.40. The pattern is naturally dominated by the large annual cycle over land, particularly at high latitudes. Note that the temperature extreme over land occurs about 1 month after the solstices, whereas over the ocean the lag is closer to 3 months.

Simulations such as these demonstrate the rich complexity of monsoon physics even before we have considered departures from zonal symmetry and other aspects of monsoon flows. In the next section we extend these results by relaxing the requirement of zonal symmetry.

5.3.2 The effects of zonal asymmetry

Zonally asymmetric thermally direct circulations differ in several fundamental ways from the zonally symmetric case we have focused on. First, we need to generalize Hide's theorem to the case of three-dimensional flows, and second, we have to consider zonal advection of entropy and vorticity.

In the zonally symmetric case, Hide's theorem forbids angular momentum from increasing toward the poles. This is equivalent to the statement that the absolute vorticity at the tropopause must never have the opposite sign as the Coriolis parameters. At a level at which the vertical velocity vanishes, it is not possible to change the sign of the absolute vorticity from its background value – the tilting term in the vorticity equation vanishes and no amount of stretching can flip the sign of the absolute vorticity. It is possible, though, to advect absolute vorticity across the equator, giving rise to absolute vorticity with the opposite sign of the Coriolis parameter. This configuration is, however, strongly unstable to inertial instability, whose effect will be to eliminate the wrong sign vorticity by mixing with adjacent regions.

Here, as in the zonally symmetric case, we seek an RCE solution that is also in dynamical equilibrium, and generalize Hide's Theorem to state that the solution must not have absolute vorticity of the opposite sign from the Coriolis parameter. A general solution that is in gradient balance is difficult to construct, but we can build an approximate one to the extent that geostrophic balance is a good approximation.

We begin with the vector form of the thermal wind equation derived by assuming geostrophic and hydrostatic balance:

$$\frac{\partial \mathbf{v}}{\partial p} = -\frac{1}{2\Omega \sin(\theta)} \mathbf{k} \times \boldsymbol{\nabla}_p \alpha, \tag{5.53}$$

where $\mathbf{v}$ is the horizontal velocity vector on isobaric surfaces and $\mathbf{k}$ is the unit vector in the vertical. Applying Maxwell's relation (5.9) and integrating (5.53) from the surface to the tropopause, assuming as before vanishing surface wind and a moist adiabatic lapse rate in RCE, yields

$$\mathbf{v}_t = \frac{T_s - T_t}{2\Omega \sin(\theta)} \mathbf{k} \times \boldsymbol{\nabla}_p s^*, \tag{5.54}$$

where $\mathbf{v}_t$ is the horizontal velocity vector at the tropopause. The condition that the absolute vorticity at the tropopause have the same sign as the Coriolis parameter may be stated as

$$\sin(\theta)\left[2\Omega \sin(\theta) + \mathbf{k} \bullet \boldsymbol{\nabla} \times \mathbf{v}_t\right] \geq 0. \tag{5.55}$$

Applying this to (5.54), and substituting the boundary layer moist entropy for s^* (that is, once again assuming convective neutrality) gives the condition for the dynamical viability of the geostrophically balanced RCE state:

$$\sin(\theta)\left[2\Omega \sin(\theta) + \frac{1}{2\Omega} \boldsymbol{\nabla} \bullet \left(\frac{T_s - T_t}{\sin(\theta)} \boldsymbol{\nabla} s_b \right)\right] \geq 0, \tag{5.56}$$

where s_b is the moist entropy of the boundary layer and we have used a vector identity. Here the divergence and gradient operators are those appropriate to spherical coordinates. If (5.56) is satisfied everywhere, then the geostrophically balanced RCE state may be dynamically viable if the geostrophic assumption is well satisfied by the flow. Note from the thermal wind equation (5.54) that the geostrophic flow will not advect s^*, which would invalidate the steady-state assumption. On the other hand, the geostrophic flow may advect itself, depending on the shape of the s_b field, and if we revert to quasi-geostrophic dynamics, the advection of pseudo-potential vorticity may not vanish, invalidating the assumption of a steady state. (This would never be a problem, of course, in zonally symmetric configurations.)

It should be remarked, however, that potential vorticity (and its quasi-geostrophic counterpart) is not conserved in RCE states owing to the strong diabatic effects of radiation and convection. Thus if (5.56) is satisfied, the state may be approximately stationary if the advection of potential vorticity is weak compared to the diabatic tendencies to restore the potential vorticity distribution of the RCE state. But often the interaction of localized RCE temperature anomalies with

background potential vorticity gradients causes significant transient behavior which may lead to substantial remote effects. A case in point is the Tibetan High, which advects the background meridional gradient of potential vorticity, causing the anticyclone to stretch westward and then break of into blobs of negative potential vorticity anomaly, which move westward, owing partially to the “beta drift’ effect, and which cause large-scale subsidence to the west of Tibet (Rodwell and Hoskins, 2001). An example is provided in Figure 5.41. The downstream subsidence is thought to be an important contribution to the extreme dryness of the Middle East and Sahara Desert.

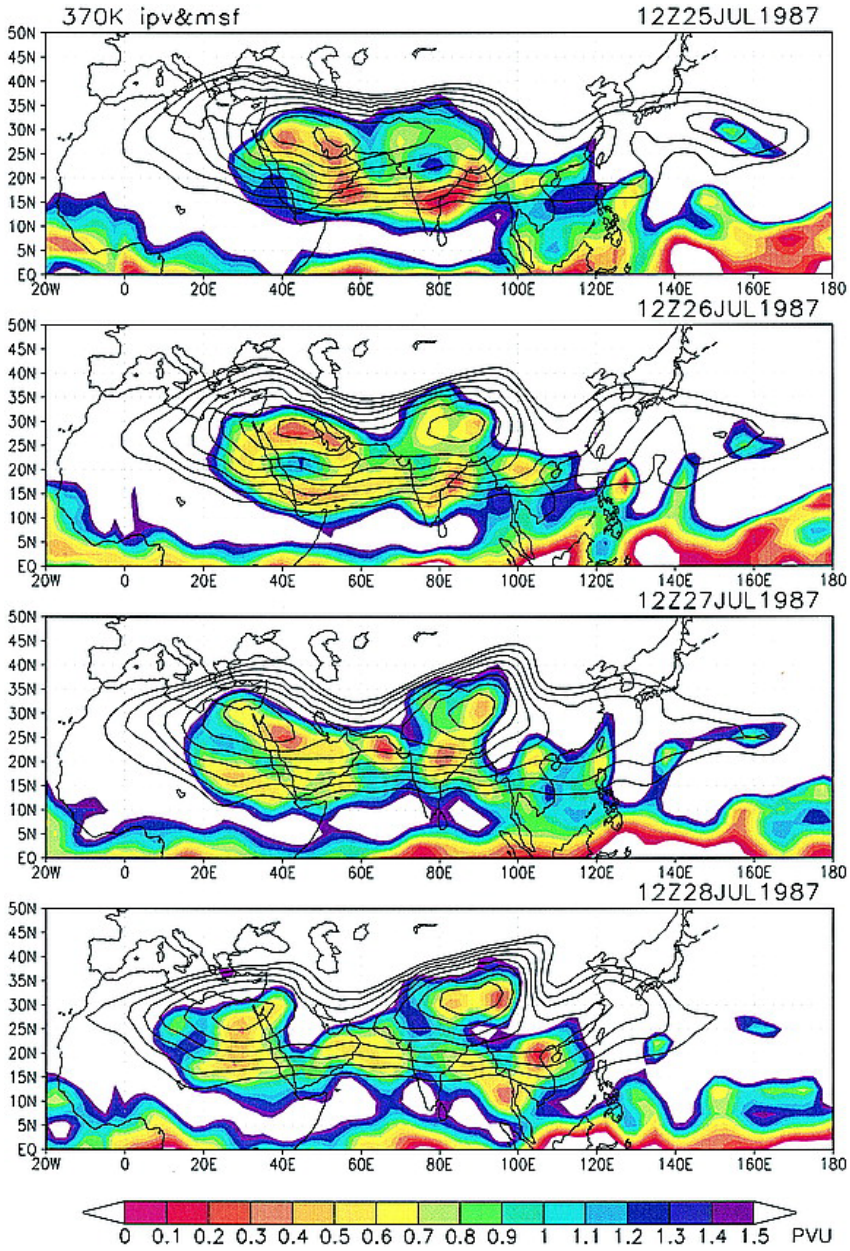


Figure 5.41: Time sequence of potential vorticity (shading) and Montgomery streamfunction (solid black contours) on the 370 K potential temperature surface at 24-hr intervals from July 25 to July 28, 1987.

If (5.56) is violated, a thermally direct circulation must result in any case. The second term in (5.56) is essentially a Laplacian operator, so if the negative curvature of the RCE boundary layer moist entropy distribution is sufficiently strong, a circulation must result. As with the Hadley circulation, a component of flow along isopleths of s_b must develop along with the component up the gradient of s_b to satisfy Ekman balance. In general, this will result in weak cyclone flow at low levels and, of course, a strong anticyclone aloft.

One consequence of the direct circulation itself is that is does modify the distribution of boundary layer entropy through advection, and if the surface energy budget is not balanced (as, for example, when there is appreciable seasonality) the wind itself modifies the surface fluxes through shear-driven turbulence, as represented by aerodynamic flux formulas.

In some cases, advection of low-entropy may shut off deep convection entirely and radically change the nature of the monsoon.

A nice illustration of this is provided by idealized simulations carried out by Privé and Plumb (2007b). They used a general circulation model with prescribed, zonally symmetric sea surface temperature and a land mass covering longitudes 0^o – 180^o and latitudes north of 16^o. Water availability over land was represented by a simple bucket scheme. They first performed a simulation in which the ocean temperature was uniformly warm at all latitudes. This produced a monsoon rainfall distribution (Figure 5.42) not unlike the summer rainfall of our zonally symmetric model (figure 5.40), with strong precipitation just inland of the coast and a weak maximum just south of the equator in association with the previously discussing “jump” of the northward flow over the equator. However, when they prescribed a latitudinal sea surface temperature gradient more typical of northern hemisphere summer, the continent became very dry (middle panel of Figure 5.42). Following earlier work of Chou et al. (2001), they showed that the dryness of the continent in this case is owing to the lateral advection of low moist entropy air from cooler ocean regions. The third panel shows the results of an experiment in which thin north-south walls were placed along the eastern and western flanks of the continent, preventing intrusion of this low entropy maritime air. This has the effect of increasing precipitation, particularly over the eastern and southern portions of the continent.

This shows that advection of moist entropy can have very significant effects on monsoons. Another interesting case of this is the role that the Himalaya play in the Indian monsoon. Boos and Kuang (2010) used a full-physics general circulation model to explore what would happen to the Indian monsoon if the Tibetan Plateau and Himalaya Mountains were removed (i.e. their elevation set to zero), and it the Plateau were removed, leaving the mountain range itself intact. The model used prescribed time-varying sea surface temperatures over a 5-year period and the June-August fields were taken to represent the monsoon season. Some results are shown in Figure 5.43. Removing just the plateau has little effect on monsoon precipitation, even though the change in surface elevation (unaccompanied, in this experiment, by any other changes in land characteristics) has a large effect on the local RCE thermodynamic profile, while removing the Himalayas greatly reduced monsoon rainfall. The Himalayas are too narrow to have much effect on RCE, but are effective in blocking the flow of low entropy air from the north into the lowlands and ocean waters south of the mountains. When the mountains are removed, the flow of low entropy air essentially quenches the production of entropy by seasonal solar forcing, and the monsoon is greatly reduced.

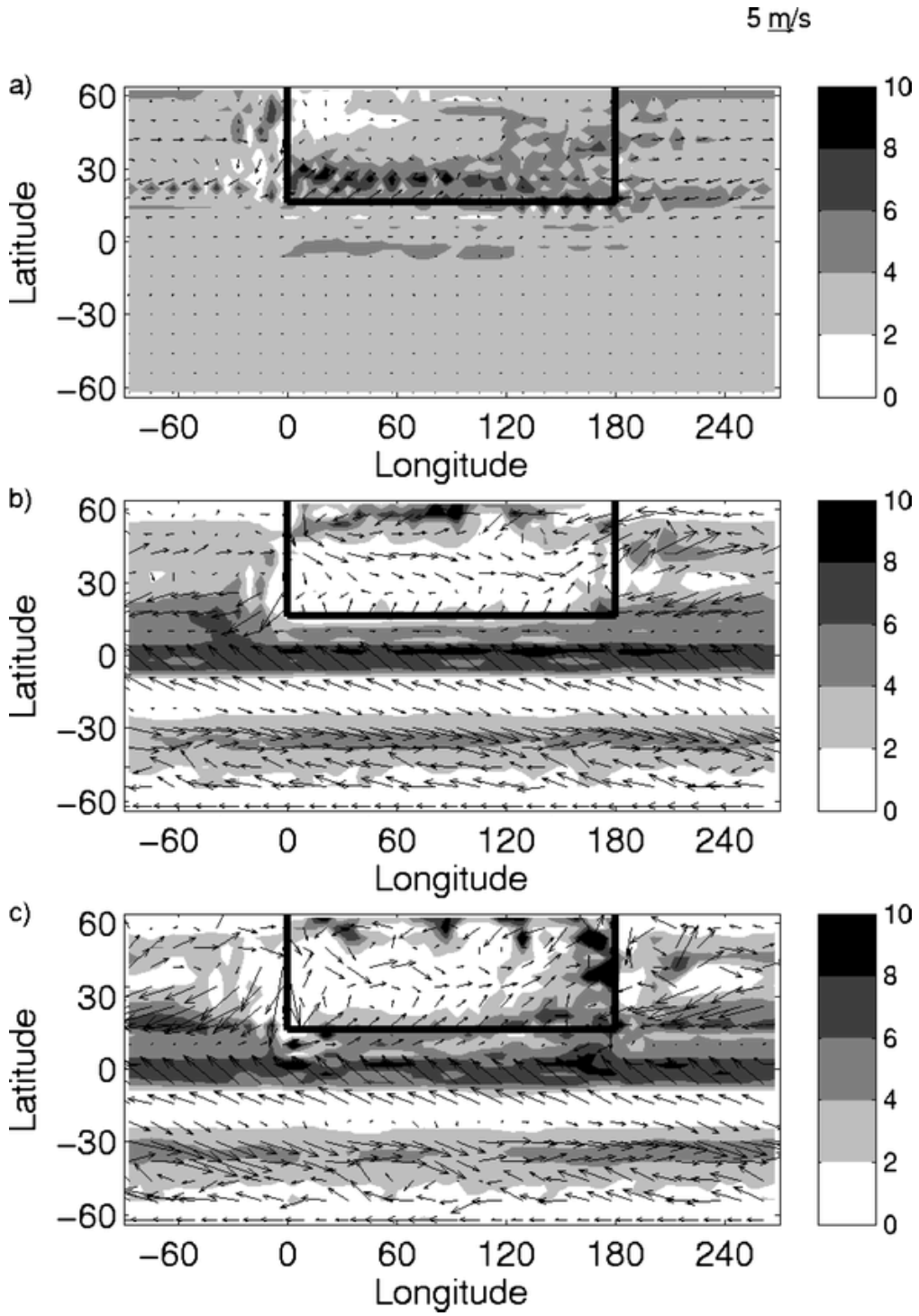


Figure 5.42: 1000-hPa winds (arrows, m s^{-1}) and precipitation (shading, mm day^{-1}) with zonally uniform sea surface temperature (SST) and land within the box depicted by the bold rectangle. (a) Case with uniform warm ocean; (b) case with summer-like SST; (c) case with summer-like SST and thin walls at eastern and western coastlines.

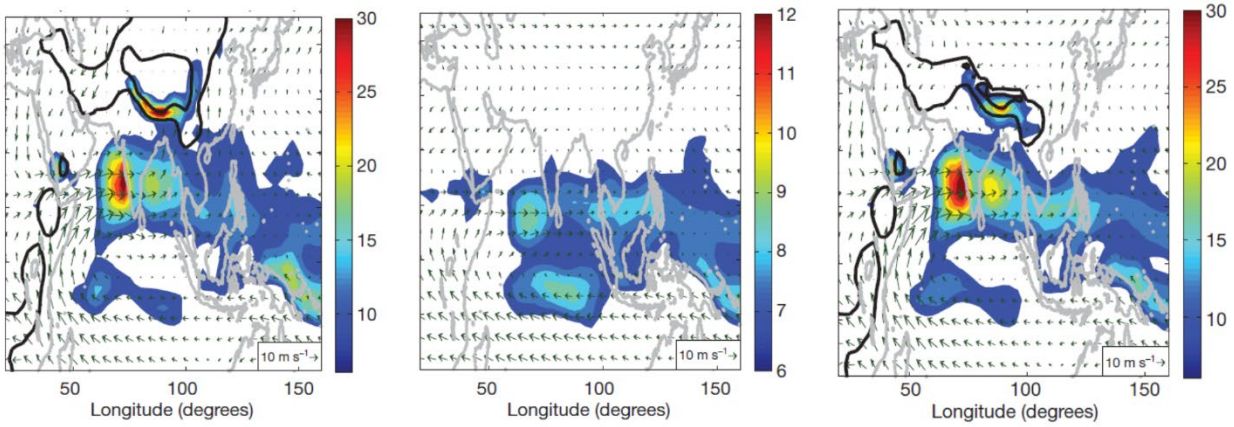


Figure 5.43: Precipitation (shading) and wind (arrows) from general circulation model experiments with prescribed, time-varying sea surface temperatures. Results are averaged over 5 years and over the months of June-August. Thick black contours surround surface pressures lower than 900 hPa and 700 hPa. Left: Control simulation; Middle: Tibetan Plateau and Himalayas removed; Right: Himalayas retained but Plateau removed.

5.3.3 Intraseasonal variability of the monsoon

The seasonal march of solar radiation might be expected, at first glance, to produce a comparably steady evolution of wind and rain accompanying the monsoon. But actual records of monsoon precipitation show interesting higher frequency modulation. Figure 5.44 displays rainfall over western India (13°-21° N, 71°-78° E) each day from May 1st through September 30th of 2011. The monsoon rains begin abruptly around June 1st and are interrupted by breaks lasting 10-20 days. These "active-break" cycles are characteristic of monsoon precipitation and are also strongly correlated with monsoon circulation.

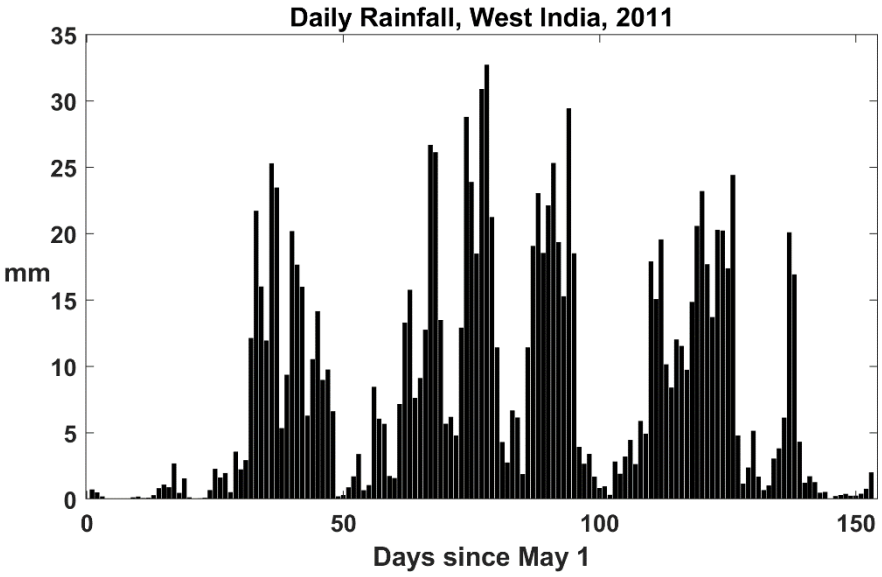


Figure 5.44: Daily rainfall (mm) over western India (13°-21° N, 71°-78° E) from May 1st through September 30th, 2011. Made using NASA's IMERG data based on a constellation of satellites.

Figure 5.45 shows a smoothed version of the west India rainfall record together with a scaled record of the kinetic energy of the Somali Jet as measured by the square of the meridional wind at 925 hPa[3]. The correlation between the two records is highly statistically significant; this correlation has been recognized since the earliest quantitative analyses of the Indian monsoon (Findlater, 1969).

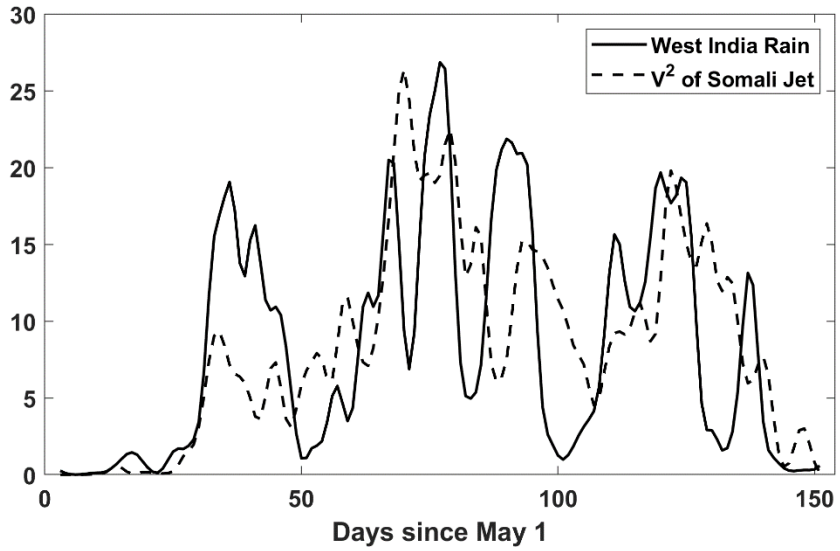


Figure 5.45: Daily precipitation (mm) as in Figure 5.44 (solid) and a measure[3] of the kinetic energy of the Somali Jet (dashed) during 2011. Both quantities have been smoothed with a 1-3-4-3-1 filter.

What is the source of this variability? Part of it may be owing to the effect of wind-induced surface heat exchange (WISHE; Boos and Emanuel, 2008). As late spring insolation heats up the north Indian Ocean, the light surface winds and associated small turbulent heat fluxes from the ocean are not sufficient to produce a positive boundary layer moist entropy gradient that is supercritical in the sense that angular momentum surfaces intersect the equator. As the critical gradient is approached, westerly surface winds begin to increase, increasing the enthalpy flux from the sea in a positive-feedback loop that rapidly leads to cross-equatorial flow. Boos and Emanuel (2008) also showed that WISHE drives the strongest positive meridional gradients of moist entropy closer to the equator, which also drives the system to earlier supercriticality. Depending on the magnitude of the insolation forcing, the system may never achieve supercriticality in the absence of WISHE, as shown in Figure 5.26.

Some aspects of this behavior can be simulated using the zonally symmetric primitive equation model described earlier in this chapter (HadleyPak, available at https://zenodo.org/doi/10.5281/zenodo.10815377). Figure 5.46 shows a time series of precipitation at 20° N from the simulation shown in Figure 5.40 and compares it to a second simulation identical to the first but in which the wind used to calculate the surface fluxes is fixed in time, eliminating the WISHE effect.

---

[3] Specifically, this function is $MAX\left[(V-6),0\right]^2$, where V is the daily average meridional wind in ms^{-1}, averaged between 7.5° S and 7.5° N, and 42° E and 53° W. Both records have been smoothed with a 1-3-4-3-1 filter.

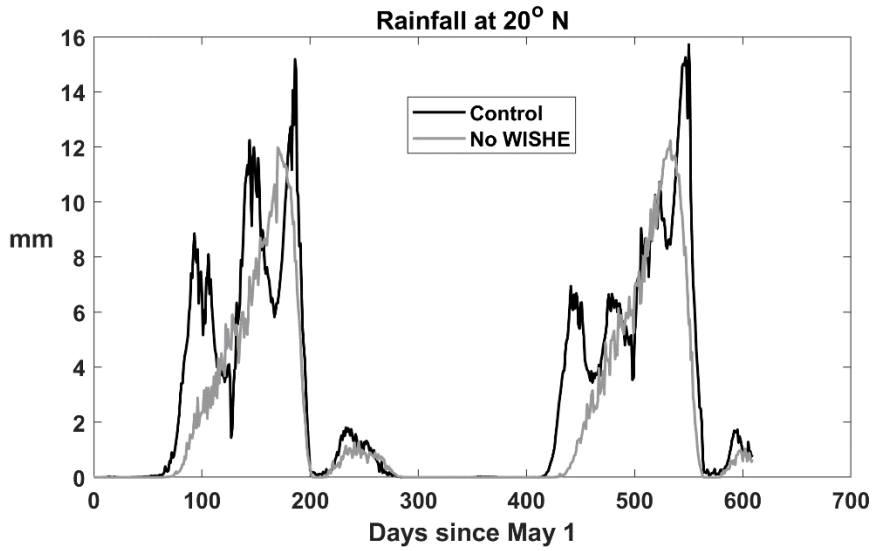


Figure 5.46: Time series of daily precipitation (mm) at 20° N from two simulations using the zonally symmetric primitive equation model used to create Figure 5.40. The control run (black) is that used in Figure 5.40 while a simulation with fixed wind speed in the turbulent surface fluxes is shown in gray. The time series begins on May 1 and extends until January 1 of the second year.

In the no-WISHE simulation, the precipitation at the coast gradually ramps up beginning in late July and reaches a peak at the end of October, collapsing abruptly around 1 November, presumably as the boundary layer moist entropy distribution becomes subcritical. The same behavior can be seen during the second monsoon season in the simulation.

Rainfall begins sooner (in mid-June by the second season) and increases far more rapidly in the control simulation in which WISHE is active. There is much more variability in monsoonal precipitation than in the no-WISHE simulation, but far less than in observations (Figure 5.44), though the frequency of variability is similar to that observed. Note also that the evolution of rainfall in the second monsoon season is, in its details, quite different from the first, suggesting that WISHE may introduce chaotic dynamics into the system, compromising the fundamental predictability limits of the onset and subsequent variability of the rainfall.

Other processes that might contribute to the transient behavior of the monsoon include land-atmosphere interactions, cloud-radiation processes, and three-dimensional effects such as the quasi-periodic shedding of anticyclones from the Tibetan High, as shown in Figure 5.41. Detailed understanding of the sudden onset of monsoon rains and active-break cycles in the monsoon circulation and rainfall awaits further research.

5.3.4 Monsoon depressions

Another interesting transient affecting the Indian summer monsoon is the monsoon depression. These are cyclonic, warm core depressions that typically form over the Bay of Bengal and move northwestward over northern India, often traveling well inland before dissipating. Figure 5.47 shows sea level pressure and winds just above the surface in a particular monsoon depression.

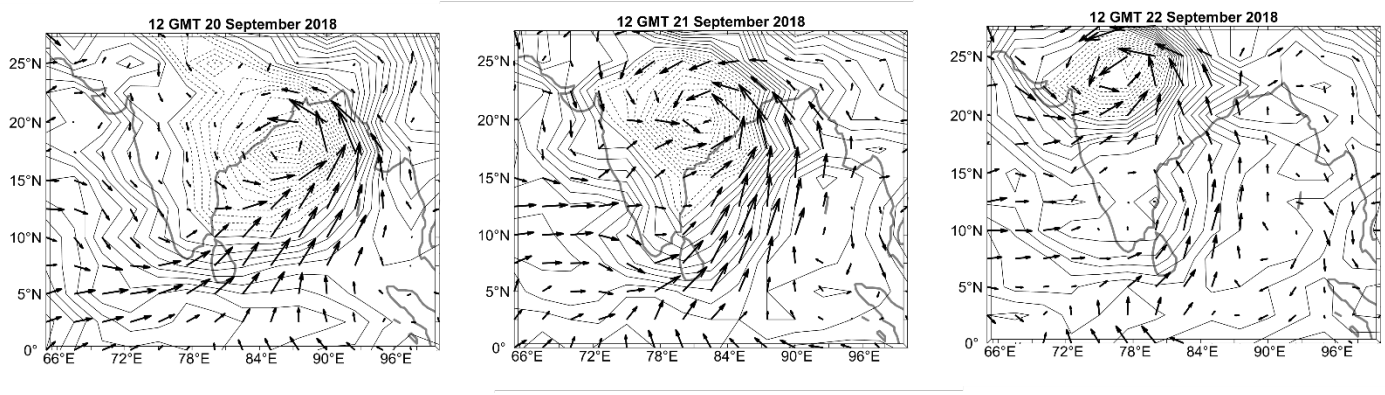


Figure 5.47: Sea level pressure (contours) and near-surface winds (arrows) associated with a monsoon depression moving from the far northern Bay of Bengal on 20 September 2018 (left) to northeastern India on 21 September (center) and continuing northwestward into north central India on September 22 (right). Sea level pressure contours are at 0.5 hPa intervals and are dashed for pressures less than 1007 hPa.

A closed cyclonic circulation developed over the northern Bay of Bengal and Bangladesh in the two days preceding September 20th, 2018 (left panel of Figure 5.47) and on the 20th the well-developed cyclone began moving inland, toward the northwest, only slowly losing intensity. A distinct cyclone is still quite evident on September 22nd (right panel of Figure 5.47).

These disturbances typically amplify while over the Bay of Bengal, reaching maximum intensity at the coastline, and then decay as they move inland. Although the amplitudes of the pressure and wind perturbations are not shown in the figure, they are typically around 3-10 hPa and 10-15 ms^{-1}, respectively[4]. They usually last 4-6 days and 3-6 of them may develop in a typical summer (Hunt et al., 2016). They strongly modulate monsoon rainfall, with the heaviest rain usually found southwest of the depression center.

The average rate of genesis of monsoon depressions is shown in Figure 5.48, based on reanalysis data for the period 1979-2003. Most of these depressions develop in the northern Bay of Bengal and over northeastern India and Bangladesh, with a few developing over northeastern India and the far eastern Arabian Sea. The majority of the systems track northward or northwestward.

A nice overview of the structure of monsoon depressions is provided by Hunt et al. (2016), who composited various quantities with respect to the position of monsoon depression centers. Figure 5.49 shows a composite distribution of TRMM-derived precipitation with respect to the depression center, with each of the contributing precipitation fields located with respect to the direction of propagation of the system. The composite maximum rainfall is a few hundred kilometers to the left of and slightly behind the storm with respect to its direction of movement, though lighter rain extends quite far to the right of, and ahead of the system.

[4] The India Meteorology Department defines a monsoon depression as having peak winds between 8.5 ms^{-1} and 16.5 ms^{-1}. Weaker cyclonic disturbances are simply known as “lows” and stronger ones are classified as tropical cyclones.

The location of the maximum precipitation downshear of the vortex center is consistent with expectations based on quasi-balanced theory (Murthy and Boos, 2020). Basically, this theory can be understood in terms of the necessary balance in either the thermodynamic equation or the vorticity equation in a coordinate system in which the system is stationary.

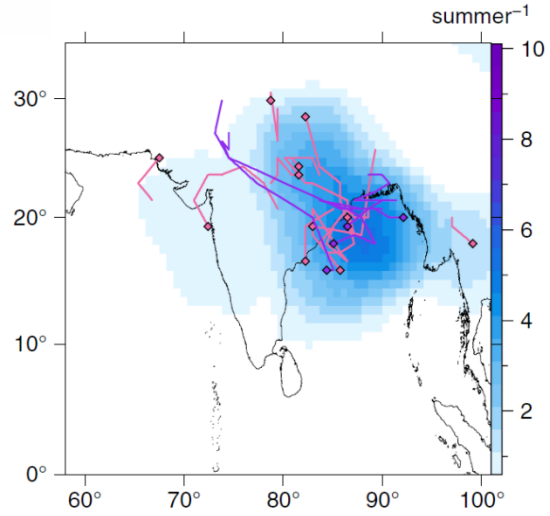


Figure 5.48: Shading shows the number of sub-tropical cyclone low pressures centers generated per summer within 500 km of each point, based in ERA-Interim reanalyses from 1979 to 2003. For the particular summer of 2003, genesis points are shown by filled diamonds while the lines show the tracks of monsoon lows (pink) and monsoon depressions (purple).

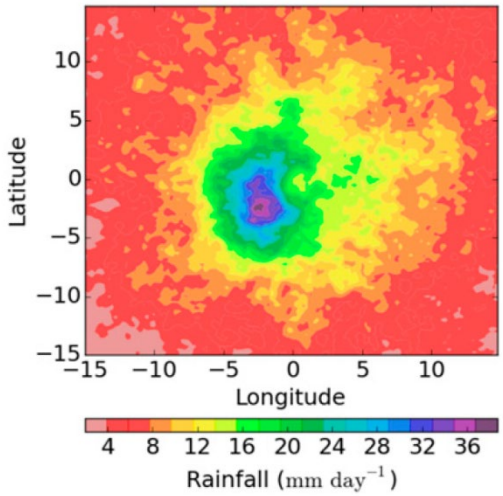


Figure 5.49: Composite precipitation (mm day^{-1}) from TRMM for 34 depressions during 1998–2013. The composite is performed with respect to the center, located arbitrarily at a latitude and longitude of 0, and rotated so that the direction of propagation is always toward the north in this diagram.

From the vorticity perspective, in a steady system, the advection of the positive vorticity of the system by the background flow must be balanced by the stretching term in the quasi-balanced form of the vorticity equation:

$$f\frac{\partial \omega}{\partial p} \cong \mathbf{V}\bullet\nabla\eta, \tag{5.57}$$

in which f is the Coriolis parameter, ω is the pressure velocity, $\mathbf{V}$ is the horizontal velocity vector, and η is the vertical component of the absolute vorticity. If we approximate $\mathbf{V}$ by the background flow, and η by the approximately circular vorticity anomaly of the disturbance, then in the moving reference frame of the storm, downshear from the vortex center, the right-hand side of (5.57) will be negative at upper levels, where the storm relative flow is downshear, and positive at low levels, where it is upshear. This would be consistent, through (5.57), with ascent at middle levels, downshear of the vortex center, and conversely, descent upshear of the center.

Besides this forcing of vertical velocity through quasi-balanced interaction with the background shear, the Ekman flow in the boundary layer will be convergent where the flow is cyclonic, contributing to the general ascent around the system. Cloud-radiation interactions may also contribute to general ascent.

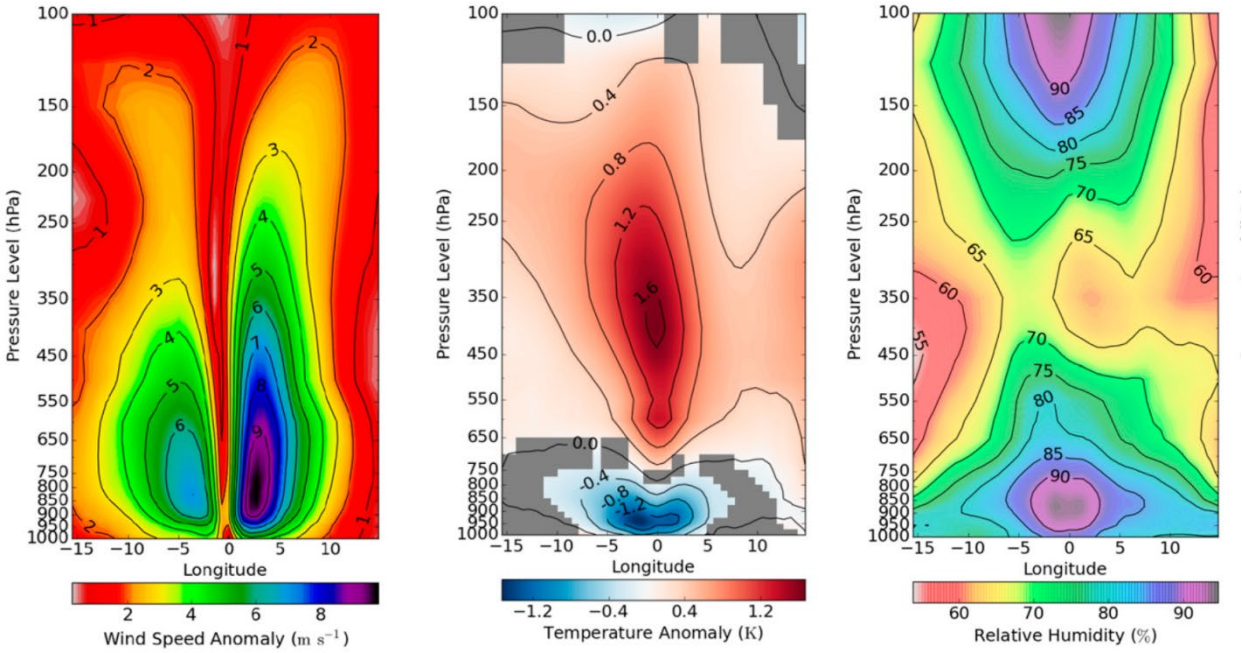


Figure 5.50: Composite cross-sections of tangential wind speed (ms^{-1}, left), temperature anomaly (K, center) and relative humidity (%, right) from ERA-Interim data, 1979-2014. The composite is made by rotating the fields into a reference frame having a common center and with the depression moving into the page. Shading grayed-out where anomaly does not significantly differ from the background at the 95% level.

Composite vertical cross-sections through monsoon depressions are shown in Figure 5.50. The cyclonic wind slopes outward with height, with a broad maximum between 950 and 650 hPa. The vortex is warm core through the whole troposphere above 700 hPa, but cold core in the lower troposphere. The temperature is depressed in the core below the 800 hPa level, but given the positive humidity anomaly that more or less coincides with the warm core, the virtual temperature depression is less. (A rough calculation gives a virtual temperature depression that is about half the temperature depression.) The whole core of the cyclone is anomalously humid, especially in the lowermost and upper most troposphere, and note that the mid-level anomaly is displaced to the left of center relative to the direction of motion, probably reflecting the upward motion inferred to exist on the downshear side of the vortex. (A vertical cross-section of vertical

motion, not shown here, also exhibits enhanced upward motion on the left side of the vortex with respect to its direction of movement.)

The dynamics of monsoon depression remain a subject of vigorous research. Although the structure and distribution of precipitation around monsoon depressions resembles that of nascent tropical cyclones (se Chapter 9), unlike the latter, monsoon depressions develop under conditions of moderate vertical wind shear and sometimes develop over land. Comparing climatological low-level flow (Figure 5.3) and upper tropospheric flow inferred geostrophically from the 200 hPa height field (Figure 5.36), the vertical shear vector points to the west-southwest over the Bay of Bengal and northern India in summer. This is also consistent, in a convectively adjusted environment, with the generally northward gradient of boundary layer moist entropy (right-hand panel of Figure 5.37).

We first note the presence of a climatological maximum of low-level moist static energy over the northern Bay of Bengal and northeast India (Figure 5.37). If we can assume that the broad region is convectively adjusted and that the Hide condition is locally violated, we would expect the climatological low-level flow to be cyclonic in this region, and indeed Figure 5.30 shows cyclonic shear and curvature over this region in summer. The high moist entropy and cyclonic low-level flow would be expected to favor deep convection. It is possible that cyclonic disturbances periodically form in this favorable environment from some combination of WISHE and cloud-radiation feedbacks, as is the case with tropical cyclones (Chapter 14), and once closed contours of vorticity develop, the cyclones migrate northwestward owing to some combination of ‘beta drift’ (Chapter 12) and Eady-like propagation along contours of the time-mean surface moist static energy. Indeed, in the example shown in Figure 5.47, cyclonic vorticity built up over the far northern Bay of Bengal and Bangladesh for several days before the disturbance began to migrate northwestward.

Most extant literature on the dynamics of monsoon depressions hypothesize that they result from dynamical instabilities of the sheared monsoonal flow in a moist convecting environment. Diaz and Boos (2021) approached the problem using a convection-permitting model with idealized, initially zonally uniform distributions of wind and moist static energy. The initial zonal wind has both easterly shear and, at low levels, cyclonic horizontal shear. The easterly vertical shear is associated with a northward-directed boundary layer moist entropy gradient (and associated temperature gradient aloft). They showed that this flow is unstable to disturbances that behave very much like observed monsoon depressions, and that the growth of such disturbances increases with both the horizontal shear and the magnitude of the boundary layer moist entropy gradient. The absence of either the moist entropy gradient or the absolute vorticity maximum associated with the background horizontal shear of the zonal wind prevents growth of the depressions. This study omitted feedbacks associated with surface fluxes and radiative interactions. It should be remarked that the basic state satisfies the Charney- Stern necessary conditions for quasi-balanced dynamical instability (Charney and Stern, 1962) in two ways: First, the horizontally sheared flow possess a maximum of absolute vorticity, allowing barotropic instability, and second, flow-relative eastward-propagating Rossby waves north of the base state vorticity maximum can interact with flow-relative westward propagating Eady edge waves on the surface moist entropy gradient; the Doppler shifting of these two kind of waves by the vertically sheared mean flow can potentially phase lock them, leading to unstable growth.

The challenge of understanding monsoon depressions is enhanced by the myriad possible mechanisms involving dynamical instabilities, moist convection, and feedbacks involving surface

fluxes and radiation. While it has been possible to simulate these depressions with both idealized and forecast-oriented models, a clear, widely accepted conceptual picture of their physics remains elusive.

5.4 The Walker Circulation

The “Walker Circulation” is an east-west overturning circulation spanning the troposphere of most of the near-equatorial Pacific region, with ascent and anomalously large rainfall in the far west and descent and dry conditions in the east. It was named, in 1969, by Jacob Bjerknes in honor of Gilbert Walker, a British applied mathematician who in 1904 became director-general of observatories in India where he conducted studies of the interannual variability of the monsoon. Figure 5.51 shows and east-west cross-section of the annual mean circulation along the equatorial Pacific, as well as the annual mean precipitation and distribution of sea surface temperature.

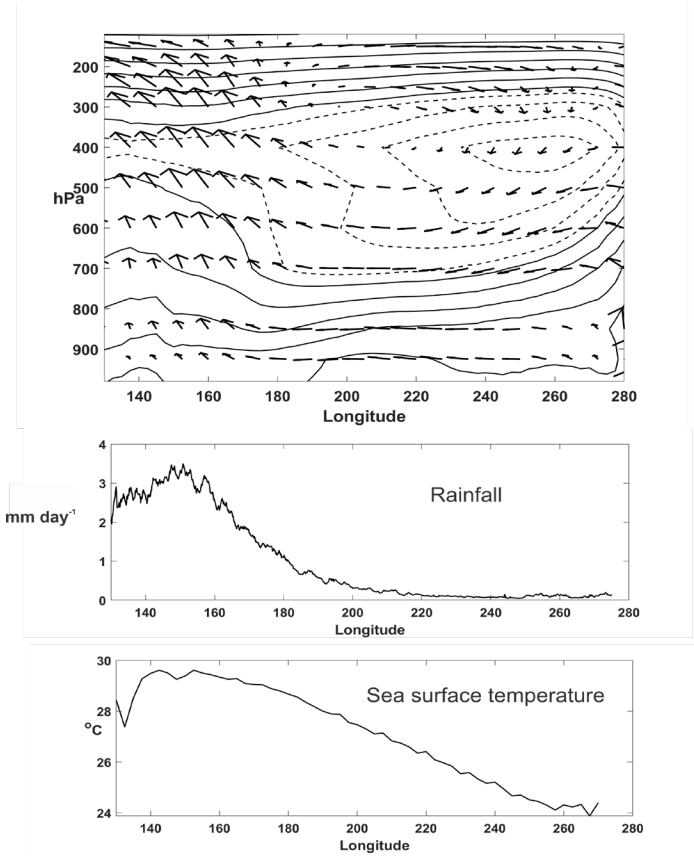


Figure 5.51: Top: Mean circulation (arrows) in the longitude-pressure plane, averaged from 7.5°S to 7.5°N and over each year from 1979-2019, using NCAR/NCEP reanalysis data. The contours show relative humidity (%) in increments of 7.5%, with values below 50% in dashed. Middle: Annual mean precipitation along the equator, from NASA’s IMERG data, 2002-2019. Bottom: Annual mean sea surface temperature on the equator, from NCAR/NCEP reanalyses, 1979-2019.

A key feature of the equatorial Pacific is the strong east-west sea surface temperature gradient… a change of about 5°C across the basin. One finds somewhat concentrated, strong ascent in the western Pacific, and weaker descent east of about 140°E. The heaviest rainfall is west of the dateline, and although both rainfall and sea surface temperature decrease eastward, one is not a simple function of the other.

The Walker circulation is properly understood as a coupled air-sea phenomenon. It would not exist on an aquaplanet (and indeed simulations with coupled global models run in aquaplanet configurations show no analogs) but owes its existence to the presence of continental boundaries at both ends of the basin. The basic physics can begin to be understood with the aid of Figure 5.52. Easterly winds drive ocean waters westward near the equator and also, through Ekman transport, drive upwelling of cold water. In the near-equatorial ocean, the wind stress is balanced by ocean pressure gradients related to an upward, eastward slope of the thermocline. Coastal upwelling off South America and Ekman-induced upwelling maintain cold surface temperatures in the east and a westward-directed surface temperature gradient. The resulting westward gradient in surface turbulent enthalpy fluxes to the atmosphere drives upward motion over the western equatorial Pacific, compensated in part by descent, balanced by radiative cooling, in the east.

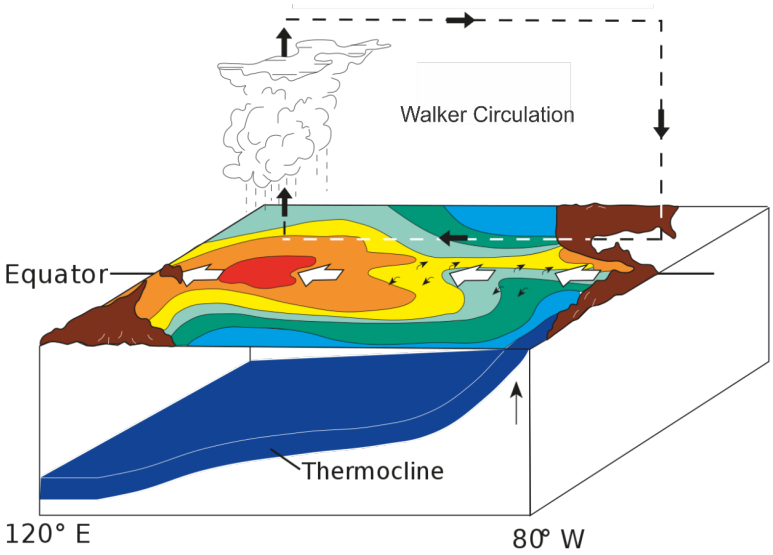


Figure 5.52: Schematic of the Walker circulation in the equatorial Pacific. Shading indicates sea surface temperature. Easterly winds near the equator drive warm water westward, with the surface wind stressed balanced by ocean pressure gradients associated with the upward eastward slope of the thermocline. The large surface heat fluxes on the western side of the basin drive upward motion, with descent in the east balanced by radiative cooling. Small black arrows represent divergence of the upper ocean flow as an Ekman response to westward directed surface wind stress.

Ocean transects along the equator (Figure 5.53) clearly show the warmer water in the west and the upward and eastward sloping thermocline of concentrated temperature gradients. Also note that the surface temperature is approximately a function of the thermocline depth. We can crudely approximate this distribution by dividing it into two fluid layers: An inert lower layer of water superimposed by a slightly less dense layer. The hydrostatic approximation allows us to treat these layers using the shallow water equation. For the upper layer, the zonal equation of motion on the equator itself is

$$\frac{du}{dt} = -g' \frac{\partial h}{\partial x} + \frac{\tau_x}{\rho h}, \tag{5.58}$$

where u is the zonal velocity, h is the depth of the upper layer, and τ_x is the zonal wind stress. Here g' is the reduced gravity:

$$g' \equiv g \frac{\Delta \rho}{\rho}, \tag{5.59}$$

where $\Delta \rho$ is the (slight) difference between the densities of the lower and upper layers. In a steady state, if we ignore the advection of zonal velocity, there is a direct relationship between the wind stress and the slope of the thermocline:

$$\frac{1}{2} g' \frac{\partial h^2}{\partial x} = \frac{\tau_x}{\rho}. \tag{5.60}$$

With surface easterlies producing a negative zonal wind stress, the thermocline shallows toward the east.

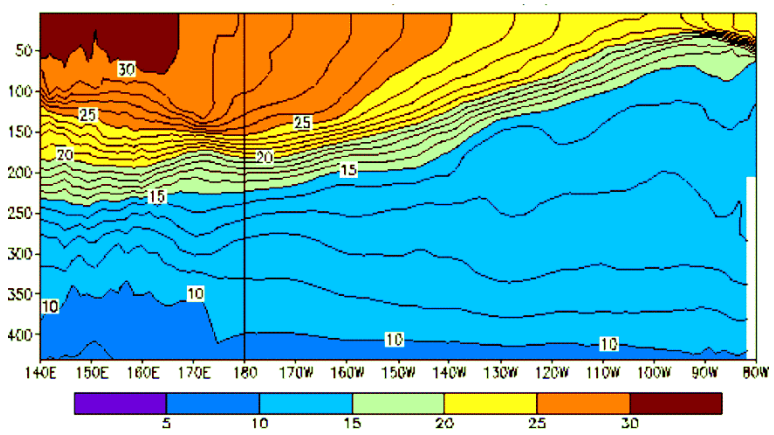


Figure 5.53: Cross-section of ocean temperature (°C) along the equator in December, 2001. Depth (m) shown at left.

5.4.1: Simple model of the coupled Walker Circulation

A first-order description of the coupled Walker circulation can be developed by coupling the simple relations for vertical motion and convection in the tropical atmosphere, developed in Chapter 3, with ocean surface fluxes dependent on surface wind speed and ocean temperature as determined from the ocean dynamics. Here we will just empirically relate sea surface

temperature to perturbations in the upper layer depth of the ocean using analyses such as shown in Figure 5.53, and then calculate the ocean depth perturbation from (5.60).

First, we assume that there is a convectively coupled region in the west, and a decoupled boundary layer in the east. We will look for steady-state solutions, and begin by re-writing the budget equation for vertically integrated moist static energy (3.64) for the steady case, and ignoring horizontal advection:

$$Gw_b = -\left\langle \alpha \dot{Q}_{cool} \right\rangle + \frac{g}{p_s - p_t} F_h, \tag{5.61}$$

where G is the gross moist stability, $\left\langle \alpha \dot{Q}_{cool} \right\rangle$ is the mass-weighted vertical average of the radiative cooling per unit mass, p_s and p_t are the surface and tropopause pressures, and F_h is the surface enthalpy flux, which we here represent using the aerodynamic flux formula for neutral stability:

$$F_h = \rho C_k |\mathbf{V}| \left(h_0^* - h^* \right), \tag{5.62}$$

where ρ is the surface air density, C_k is the nondimensional surface enthalpy flux coefficient, $|\mathbf{V}|$ is the surface wind speed, h_0^* is the saturation moist static energy of the sea surface, and h^* is the saturation moist static energy of the free troposphere, assumed to be constant vertically (moist adiabatic assumption) and horizontally (WTG). (We assume, according to convective neutrality, that the boundary layer moist static energy equals h^*). We define a scale height $H \equiv \left(p_s - p_t \right) / \rho g$, and for notational convenience, denote $\left\langle \alpha \dot{Q}_{cool} \right\rangle$ as just $\dot{Q}$. Then using these and (5.62) we can re-write (5.61) as

$$Gw_b = -\dot{Q} + H^{-1} C_k |\mathbf{V}| \left(h_0^* - h^* \right). \tag{5.63}$$

Next, to make the problem analytically tractable, we linearize the last term in (5.63) around a basic state that is everywhere in RCE (for which $w_b = 0$) and in which there is a background easterly wind of magnitude U_0:

$$Gw_b' = H^{-1} C_k \left[-u' \left(\overline{h_0^*} - h^* \right) + U_0 h_0^{*\prime} \right], \tag{5.64}$$

where the primes denote perturbations from the base state, and the overbar represents a base state value. We have assumed that $\dot{Q}$ is constant (and so ignore cloud and water vapor feedbacks in radiation) and also ignore the nonlinear correlation between perturbation ocean temperature and perturbation wind. We note that in the RCE state

$$C_k U_0 \left(\overline{h_0^*} - h^* \right) = \dot{Q} H$$

and for notational convenience use this in (5.64):

$$Gw_b' = \left[-u' \frac{\dot{Q}}{U_0} + H^{-1} C_k U_0 h_0^{*\prime} \right]. \tag{5.65}$$

This expression indicates that perturbation upward motion can arise from negative perturbation surface winds – which enhance the background easterlies and thus increase the surface enthalpy flux – or from positive ocean temperature perturbations, which likewise increase the surface flux.

Next, we relate perturbations in ocean surface temperature $\left(T_s'\right)$ to perturbations in ocean mixed layer depth by simply assuming a linear dependence, based on Figure 5.53:

$$T_s' = \left(\frac{\partial T_s}{\partial d} \right) d', \tag{5.66}$$

where d' is the perturbation mixed layer depth and we estimate the coefficient $\left(\frac{\partial T_s}{\partial d} \right)$ empirically. From the definition of moist static energy, and using the Clausius-Clapeyron equation, we derive

$$h_0^{*\prime} = \left(c_p + \frac{L_v^2 q_0^*}{R_v T_s^2} \right) T_s' = \left(c_p + \frac{L_v^2 q_0^*}{R_v T_s^2} \right) \left(\frac{\partial T_s}{\partial d} \right) d', \tag{5.67}$$

where we have made use of (5.66). Using (5.67), we can express vertical motion as a function of perturbations to the surface zonal wind and to the ocean mixed layer depth.

To close the system, we approximate the two-dimensional mass continuity equation along the equator as

$$\frac{\partial u'}{\partial x} = -2 \frac{w_b'}{H}, \tag{5.68}$$

and we linearize the ocean force balance equation (5.60) as

$$g' D \frac{\partial d'}{\partial x} = \frac{\tau_x}{\rho} = 2 C_D U_0 u', \tag{5.69}$$

where D is the mean mixed layer depth and C_D is the drag coefficient. The right side of (5.69) results from linearizing the aerodynamic drag formula about the mean easterly wind. Negative perturbations of the surface zonal wind (enhanced easterlies) give rise to shoaling mixed layer depth toward the east.

Equation (5.65) is valid in the deep convectively coupled region in the west, while (5.68) and (5.69) are valid everywhere. We first note that there are as many as three possible solutions. First the RCE solution, for which all the perturbation variables vanish, is a stationary state of the system. Without a time-dependent ocean, and possibly atmosphere, it is not clear whether this RCE state is stable. As we will see, for some combinations of parameters, this is the only viable solution. Second, we may have state in which there is deep convection everywhere, but

stronger in the west and with a Walker-like circulation. Finally, as we envisioned in setting up this simple model, the circulation may be strong enough to shut down deep convection in the eastern portion of the basin. We will begin by searching for this third solution.

In the eastern portion of the basin, where we assume a decoupled atmospheric boundary layer and no deep convection, there must be a balance between subsidence warming and radiative cooling. This is expressed by (3.66), which in the current notation can be written

$$w_b' = -\frac{\dot{Q}}{\mathbf{S}}, \tag{5.70}$$

where recall that $\mathbf{S}$ is the base state dry static stability. We can then use mass continuity (5.68) to derive the surface zonal wind. Integrating (5.68) using (5.70) and taking $u' = 0$ at $x = L$, where L is the value of x at the eastern boundary of the basin, we get

$$u' = -2\frac{\dot{Q}}{\mathbf{S}}(L - x). \tag{5.71}$$

Then we can solve (5.69) for the perturbation ocean mixed layer depth, though in so doing we will acquire an integration constant. We will return to this later.

Before proceeding, we can simplify the notation a great deal by nondimensionalizing the dependent variables and the x coordinate. We choose the following normalizations:

$$\begin{aligned} x &\rightarrow ax \\ u' &\rightarrow U_0 u \\ w_b' &\rightarrow \frac{HU_0}{a} w \\ d' &\rightarrow \frac{C_D U_0^2 a}{g' D} d \end{aligned} \tag{5.72}$$

in which a is the radius of the Earth. With the substitutions, the equations for conservation of mass and ocean dynamic balance (5.68 and 5.69) can be written simply

$$\frac{\partial u}{\partial x} = -2w, \tag{5.73}$$

and

$$\frac{\partial d}{\partial x} = u. \tag{5.74}$$

In the convective region, the energy balance equation (5.65) (with the additional relations (5.66) and (5.67)) can be written

$$w = -\chi u + \Gamma d, \tag{5.75}$$

where the nondimensional constant parameters are given by

$$\chi \equiv \frac{a\dot{Q}}{GHU_0} \tag{5.76}$$

and

$$\Gamma \equiv \frac{2C_k C_D a^2 U_0^2 \left(c_p + \frac{L_v^2 q_0^*}{R_v T_s^2} \right) \left(\frac{\partial T_s}{\partial d} \right)}{g' G H^2 D}. \tag{5.77}$$

While 5.75 is valid in the deep convecting regime, in the decoupled regime the nondimensional equivalent of (5.70) is

$$w = -w_d \equiv \frac{-\dot{Q}a}{U_0 H \mathbf{S}}, \tag{5.78}$$

where we have introduced w_d for notational convenience.

First, in the deep convecting regime, we can find both the convective updraft mass flux and the difference between the boundary layer and free tropospheric moist static energy though the WTG relations (3.59) and (3.60). Eliminating the moist static energy difference between these two relations, and normalizing the convective updraft mass flux the same was we normalized w_b in (5.72), we have simply

$$M = \frac{w + w_d}{\epsilon_p}, \tag{5.79}$$

where recall that ϵ_p is the precipitation efficiency. For the deep convective solution to be viable, M must be non-negative. Given M or w we can derive the moist static energy deficit through either (3.59) or (3.60).

Thus our relatively simple system consists of (5.73) and (5.74), and either (5.75) if there is deep convection, or (5.78) if there is not.

Let's first look for solutions in the deep convecting regime. If we eliminate w and d in (5.73)-(5.75), the result is a second order equation for u:

$$\frac{1}{2}\frac{\partial^2 u}{\partial x^2} - \chi \frac{\partial u}{\partial x} + \Gamma u = 0. \tag{5.80}$$

The solution to this that satisfies the imposed boundary condition $u = 0$ at $x = 0$ is

$$u = -Ae^{\chi x} \sin(cx), \tag{5.81}$$

where

$$c \equiv \sqrt{2\Gamma - \chi^2} \tag{5.82}$$

and A is an integration constant. Note that the argument of the square root in (5.82) may be negative for some values of the dimensional parameters, in which case we flip the sign of the argument and replace the sine function in (5.81) by a hyperbolic sine.

The corresponding solutions for w and d (when c is a real number) are

$$w = \frac{1}{2} A e^{\chi x} \left[\chi \sin(cx) + c \cos(cx) \right], \tag{5.83}$$

and using (5.75)

$$d = \frac{1}{2\Gamma} A e^{\chi x} \left[c \cos(cx) - \chi \sin(cx) \right]. \tag{5.84}$$

An interesting situation arises if c is large enough that $cL > \pi,$ where L is the nondimensional basin length. In that case, according to (5.81), u returns to zero somewhere west of the eastern basin boundary, and the deep convective regime may be viable everywhere. In that case, there is no way to determine the integration constant A in (5.81), except that it is bounded above by two separate conditions: First, the convective updraft mass flux M as given by (5.79) must never become negative, else the system becomes convectively decoupled in places, and second, we must insist that $u < 1$ everywhere, else the total zonal flow (background plus perturbation) would change sign, violating our linearization of the absolute wind speed in the aerodynamic flux formulae. Even if take a value of A small enough to prevent these violations, there is no guarantee that the solution is stable. We will return to this issue later in the chapter.

Now we turn to the case where there is assumed to be a deep convective regime in the west and an uncoupled regime in the east. In the second case, $w = -w_d$ and integration of the mass continuity equation, subject to $u = 0$ on $x = L$ yields the nondimensional equivalent of (5.71):

$$u = -2 w_d (L - x) \tag{5.85}$$

Using this, we can integrate (5.74) to find the perturbation mixed layer depth:

$$d = d_0 - w_d x (2L - x), \tag{5.86}$$

where d_0 is an integration constant.

We next need to match the solutions in the deep convecting and uncoupled regimes. We note that we have two integration constants, A and d_0, and also we need to determine the position, in $x,$ of the transition between the two regimes. Altogether, we require three additional stipulations to close the problem. Two of them are straightforward: mass continuity demands that we match the zonal velocity perturbations u at the regime transition, nor can we entertain an infinite acceleration in the ocean that would result from a discontinuity of d, so we must match the perturbation mixed layer depths across the transition. As a third condition, we require the conservation of overall mass of the ocean mixed layer…fluid can rearrange itself in x but the overall mass of the mixed layer should be conserved, so that

$$\int_0^L d\,(dx) = 0 \tag{5.87}$$

Figure 5.54 shows a solution of this system for a particular choice of χ, Γ, w_d and L. Deep convection is confined to roughly the western quarter of the basin, and the thermocline is deepest at the western boundary. The convectively decoupled region spans most of the domain, with easterly winds maximizing at the transition.

The dependence of the transition longitude on Γ with the other nondimensional parameters fixed at the values used in Figure 5.54 is shown in Figure 5.55. No viable solution exists for Γ less than about 40% of the control value used in Figure 5.54. From the definition of Γ given by (5.77), this implies that there will be no Walker circulation if, for example, the exchange coefficients are too small or the sensitivity of ocean surface temperature to the thermocline depth is too small; in that case the RCE state is the only viable steady solution. In our linearization, we assumed a background wind speed in the aerodynamic flux formulae for the surface drag and enthalpy flux. Had we kept the fully nonlinear wind speed in these flux formulae, we likely would have found two solutions: a resting state (for which $\Gamma = 0$ and χ is very large) and a Walker state, in which Γ and χ assume moderate values.

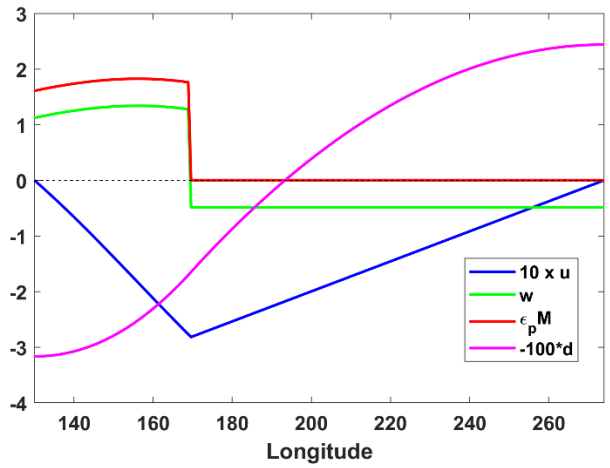


Figure 5:54: Solution of the simple Walker model for χ=2.4 , Γ=35, w_d=0.5, and L=0.4. The blue curve shows zonal velocity perturbation (multiplied by 10), the green curve shows large-scale vertical velocity, the red curve shows the product of the precipitation efficiency and the convective updraft mass flux, and the magenta curve shows the ocean mixed layer depth perturbation multiplied by 100. The mixed layer depth has been flipped so that negative values are positive perturbations to the depth. All graphed quantities are nondimensional. The western boundary is at 130° and the eastern boundary at 274°.

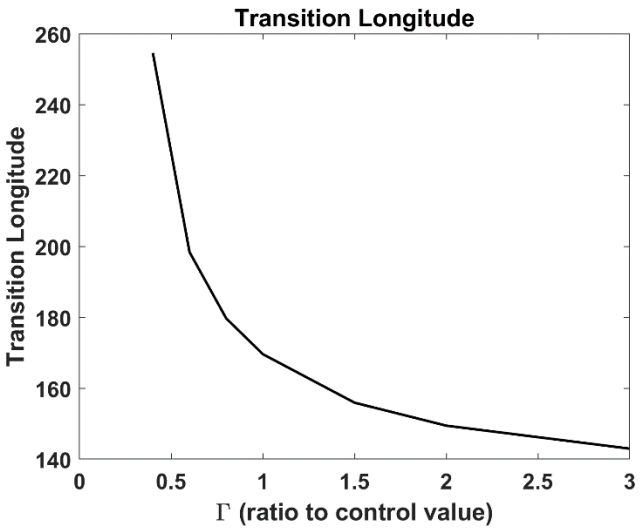


Figure 5.55: Transition longitude as a function of the ratio of Γ *to its control value as used in Figure 5.54. No solutions exist for* $\Gamma / \Gamma_c \lessapprox 0.4$*, where* $\Gamma_c = 2.4$*. The western boundary is at 130º and the eastern boundary at 274º.*

The reader can explore the parameter space of this simple model more thoroughly using the MATLAB script *Walker.m* available at https://zenodo.org/doi/10.5281/zenodo.10825247 .

This simple model does not include potentially important feedbacks involving radiation. Qualitatively, we might expect the presence of high clouds and high column water vapor in the west to provide additional column heating there, while low clouds and a dry troposphere would lead to greater column cooling in the east; as a result, the radiative feedbacks might well enhance the Walker Circulation.

One final point here: We might ask whether we could have found solutions for a mean state with westerly winds. The answer is almost certainly no, because westerlies near the equator would drive an Ekman convergence there, making the ocean mixed layer deeper and preventing any upwelling that would cool the surface. There are two asymmetries at work here: Upwelling cools the surface, but downwelling does not warm it, and easterlies cause upwelling but westerlies cause downwelling. This second asymmetry is, of course, related to the sign of rotation of the Earth.

While formulating a simple view of the Walker Circulation is instructive, we have not addressed what happens off the equator or the important question of whether such a solution is stable and what kind of free oscillations are possible. We will address these issues in the next chapter, as part of our review of ENSO.

References

Anderson, D. L. T., 1976: The low-level jet as a western boundary current. *Mon. Wea. Rev.*, **104,** 907-921, doi:10.1175/1520-0493(1976)104<0907:Tlljaa>2.0.Co;2.

Blanford, H. F., 1884: II. On the connexion of the Himalaya snowfall with dry winds and seasons of drought in India. *Proceedings of the Royal Society of London*, **37,** 3-22, doi:10.1098/rspl.1884.0003.

Boos, W. R., and K. A. Emanuel, 2008: Wind–evaporation feedback and abrupt seasonal transitions of weak, axisymmetric Hadley circulations. *J. Atmos. Sci.*, **65,** 2194-2214, doi:10.1175/2007jas2608.1.

Boos, W. R., and Z. Kuang, 2010: Dominant control of the South Asian monsoon by orographic insulation versus plateau heating. *Nature*, **463,** 218-222, doi:10.1038/nature08707.

Charney, J. G., and M. E. Stern, 1962: On the stability of internal baroclinic jets in a rotating atmosphere. *J. Atmos. Sci.*, **19,** 159-172, doi:10.1175/1520-0469(1962)019<0159:otsoib>2.0.co;2.

Chou, C., J. D. Neelin, and H. Su, 2001: Ocean-atmosphere-land feedbacks in an idealized monsoon. *Quart. J. Roy. Meteor. Soc.*, **127,** 1869-1891, doi:https://doi.org/10.1002/qj.49712757602.

Compo, G. P., J. S. Whitaker, and P. D. Sardeshmukh, 2008: The 20th century reanalysis project. *Third WCRP International Conference on Reanalysis*.

Diaz, M., and W. R. Boos, 2021: Evolution of idealized vortices in monsoon-like shears: Application to monsoon depressions. *J. Atmos. Sci.*, doi:10.1175/jas-d-20-0286.1.

Emanuel, K. A., 1994: *Atmospheric convection.* Oxford Univ. Press, New York, 580 pp.

Findlater, J., 1969: A major low-level air current near the Indian ocean during the northern summer. *Quart. J. Roy. Meteor. Soc.*, **95,** 362-380, doi:https://doi.org/10.1002/qj.49709540409.

Hadley, G., 1735: Concerning the cause of the general trade winds. *Phil. Trans.;*, **29,** 58-62.

Halley, E., 1686: An historical account of the trade winds, and monsoons, observable in the seas between and near the tropicks, with an attempt to assign the phisical cause of the said winds. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, **16,** 153-168.

Held, I. M., and A. Y. Hou, 1980: Nonlinear axially symmetric circulations in a nearly inviscid atmosphere. *J. Atmos.. Sci.*, **37,** 515-533.

Hide, R., 1969: Dynamics of the atmospheres of the major planets with an appendix on the viscous boundary layer at the rigid bounding surface of an electrically-conducting rotating fluid in the presence of a magnetic field. *Journal of Atmospheric Sciences*, **26,** 841-853, doi:10.1175/1520-0469(1969)026<0841:Dotaot>2.0.Co;2.

Hunt, K. M. R., A. G. Turner, P. M. Inness, D. E. Parker, and R. C. Levine, 2016: On the structure and dynamics of Indian monsoon depressions. *Mon. Wea. Rev.*, **144,** 3391-3416, doi:10.1175/mwr-d-15-0138.1.

Molnar, P., and K. Emanuel, 1999: Temperature profiles in radiative-convective equilibrium above surfaces at different heights. J. Geophys. Res., 104, 24265–24484.

Murthy, V. S., and W. R. Boos, 2020: Quasigeostrophic controls on precipitating ascent in monsoon depressions. *J. Atmos. Sci.*, **77,** 1213-1232, doi:10.1175/jas-d-19-0202.1.

Pauluis, O., 2004: Boundary layer dynamics and cross-equatorial Hadley circulation. *J. Atmos. Sci.*, **61,** 1161-1173, doi:10.1175/1520-0469(2004)061<1161:Bldach>2.0.Co;2.

Privé, N. C., and R. A. Plumb, 2007a: Monsoon dynamics with interactive forcing. Part I: Axisymmetric studies. *J. Atmos. Sci.*, **64,** 1417-1430, doi:10.1175/jas3916.1.

——, 2007b: Monsoon dynamics with interactive forcing. Part II: Impact of eddies and asymmetric geometries. *J. Atmos. Sci.*, **64,** 1431-1442, doi:10.1175/jas3917.1.

Rodwell, M. J., and B. J. Hoskins, 2001: Subtropical anticyclones and summer monsoons. *J. Climate*, **14,** 3192-3211, doi:10.1175/1520-0442(2001)014<3192:Saasm>2.0.Co;2.

6 El Niño – Southern Oscillation (ENSO)

6.1 Introduction

At irregular intervals of two to seven years, the cold tongue of water along the equator in the eastern Pacific region warms up by as much as 3 K, accompanied by vast changes in atmospheric circulation and the distribution of rainfall over the tropical Pacific and Indian Oceans. These events were well known to pre-Columbian residents of Pacific coastal South America, and Peruvian fishermen referred to such episodes as El Niño events as they tend to occur around Christmas ('El Niño', literally "the boy', referring to the Christ child), bringing large changes to local fish and seabird populations. In contemporary parlance, the opposite phase of the system, perhaps better characterized as the absence of El Niño, is referred to as 'La Niña'.

In the mid-1920s, Gilbert Walker (after whom the Walker Circulation was named) noticed a strong anti-correlation between seasonally varying sea surface pressures at Darwin, Australia, and Tahiti, in the central tropical South Pacific, and coined the term "Southern Oscillation" to describe this see-saw of pressures. Somewhat later it was recognized that the Southern Oscillation and the dramatic warming of the eastern equatorial Pacific are part of the same phenomenon, which is now referred to as "El Niño-Southern Oscillation" (ENSO). We use the terms El Niño and La Niña to describe different phases of ENSO.

ENSO is an enormously broad and intellectually extraordinarily rich subject that deserves a textbook of its own; here we provide only a rudimentary overview. The reader is referred to Webster (2020) for a comprehensive review of the scientific understanding of ENSO. Wang et al. (2017) also provide a nice review of the subject.

Examples of the distributions of sea surface temperature anomalies in two different types of El Niño episode are shown in Figure 6.1. These patterns represent the two leading EOF's of ENSO variability (Takahashi et al., 2011) and most El Niño events can be well described as a weighted sum of these two patterns. The eastern Pacific pattern peaks just off the coast of Peru, while the second pattern peaks in the central equatorial Pacific.

The strong sea surface temperature anomalies are associated with equally pronounced circulation anomalies. Figure 6.2 is a schematic diagram showing distributions of surface pressures and winds, superimposed on sea surface temperature anomalies, for the two ENSO phases. El Niño events are associated with strong westerly wind anomalies (and the actual wind can become westerly over large portions of the central and western equatorial Pacific), reduced pressure in the east and increased pressure in the west. As one might expect, the westerly low-level wind anomalies are associated with anomalous ascent in the eastern equatorial Pacific, and descent in the west, and this is reflected in the pattern on anomalous precipitation (Figure 6.3). The precipitation anomalies can be quite large and persist for many months, leading to severe drought in Indonesia, Australia, southeast Asia, and as far away as India, and flooding in equatorial South America.

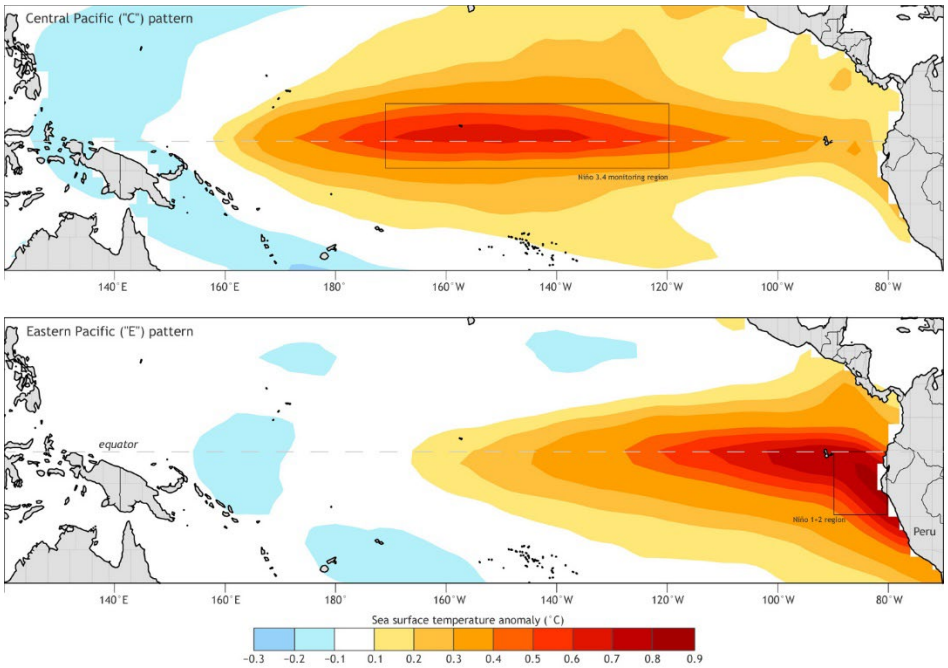


Figure 6.1: Composite sea surface temperature anomalies (K) for two different classifications of El Niño. Top: Central Pacific pattern; Bottom: Eastern Pacific pattern. These represent the two leading empirical orthogonal functions of the ENSO variability, and most El Niño events can be described as a weighted sum of these. The thin black boxes represent conventional ENSO monitoring regions.

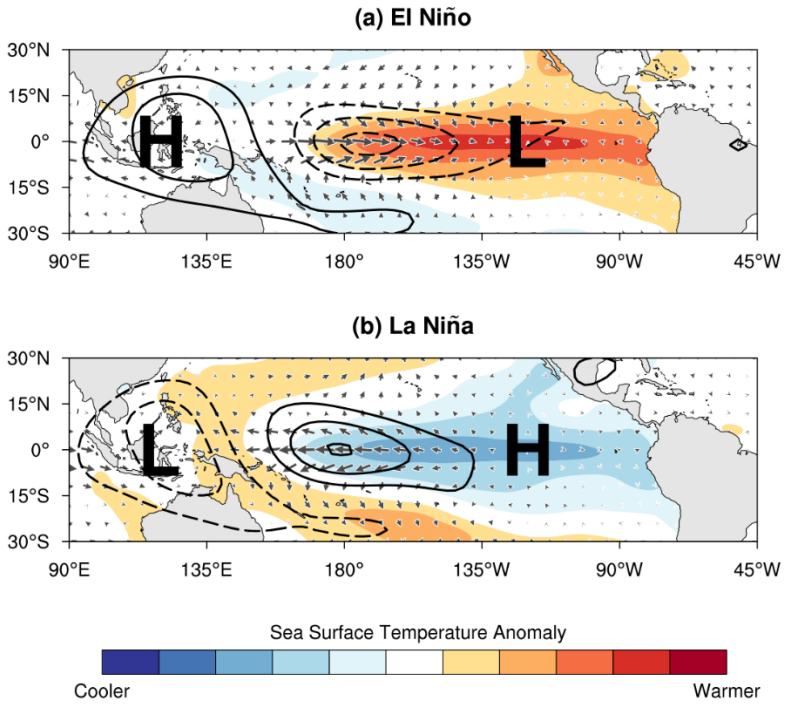


Figure 6.2: Schematic diagram showing the distributions of anomalous surface winds and pressures and associated sea surface temperature anomalies for the two phases of ENSO. The schematic is based on the composite anomalies for November-December drawn from 11 warm events and 11 cold events during 1980-2016.

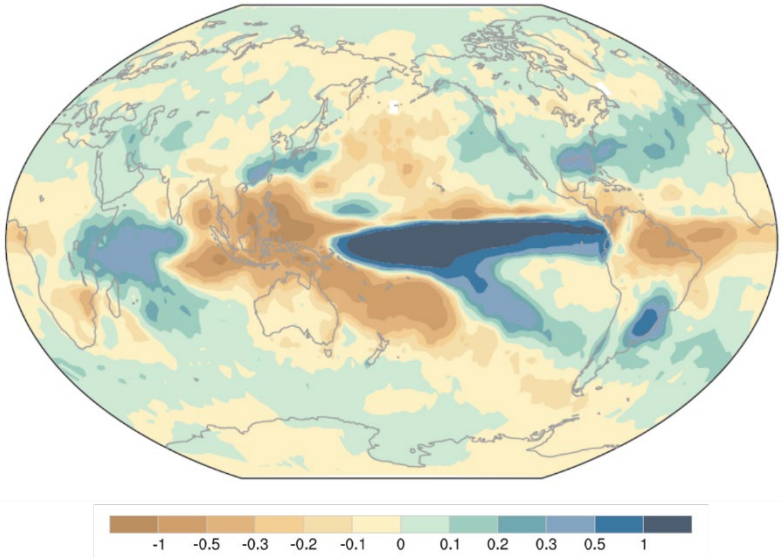


Figure 6.3: Monthly precipitation anomalies regressed onto the first principle component of ENSO time series of sea surface temperature. The precipitation is based on the satellite-based Global Precipitation Climatology Project for the period 1979–2011. Units are mm day-1 per standard deviation of the principle component time series of sea surface temperature.

ENSO is a strongly coupled phenomena, accompanied by strong perturbations in the upper equatorial Pacific Ocean. During El Niño events, the thermocline sinks in the east and rises in the west in association with positive sea surface temperature anomalies in the east and negative anomalies in the west (Figure 6.4). This represents a strong relaxation of the normal upward slope of the thermocline toward the east and massive eastward movement of the warm pool waters of the western equatorial Pacific.

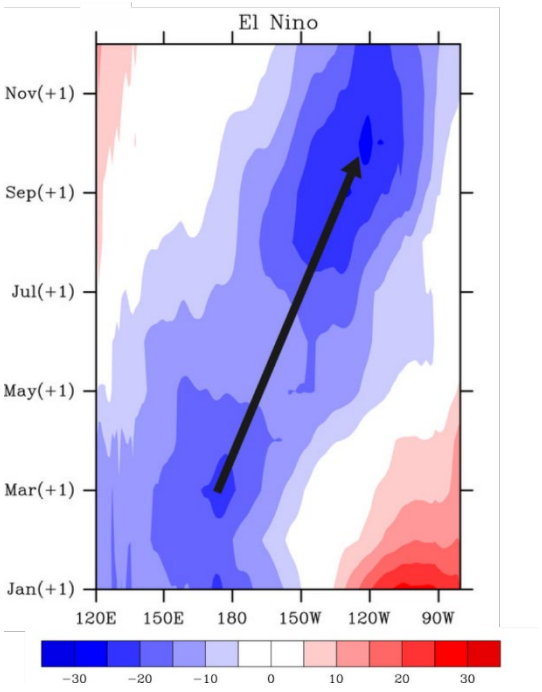


Figure 6.4: Composite time evolution of thermocline depth during an El Niño event. Evolution is from January (bottom) to November (top). Deeper thermoclines are indicated by blue.

The time evolution of ENSO is best illustrated by time series of various ENSO indices, which are based in whole or in part on near-equatorial sea surface temperature anomalies. One such index spanning the period 1980-2020 is reproduced in Figure 6.5. This multivariate index is the leading combined empirical orthogonal function of five different variables: sea level pressure, sea surface temperature, zonal and meridional components of the surface wind, and outgoing longwave radiation, over the tropical Pacific basin (30°S-30°N and 100°E-70°W). Positive values are associated with El Niño events. Exceptionally strong El Niños are evident in 1983, 1997 and 2016. The events occur irregularly every 2-7 years or so.

While not obvious in Figure 6.5, the evolution of El Niño events is strongly tied to the seasonal cycle. El Niño and La Niña episodes typically last 9-12 months. They both tend to develop during the Boreal spring (March-June), reach peak intensity during the late autumn or winter (November-February), and then weaken during the following Boreal spring or early summer (March-June).

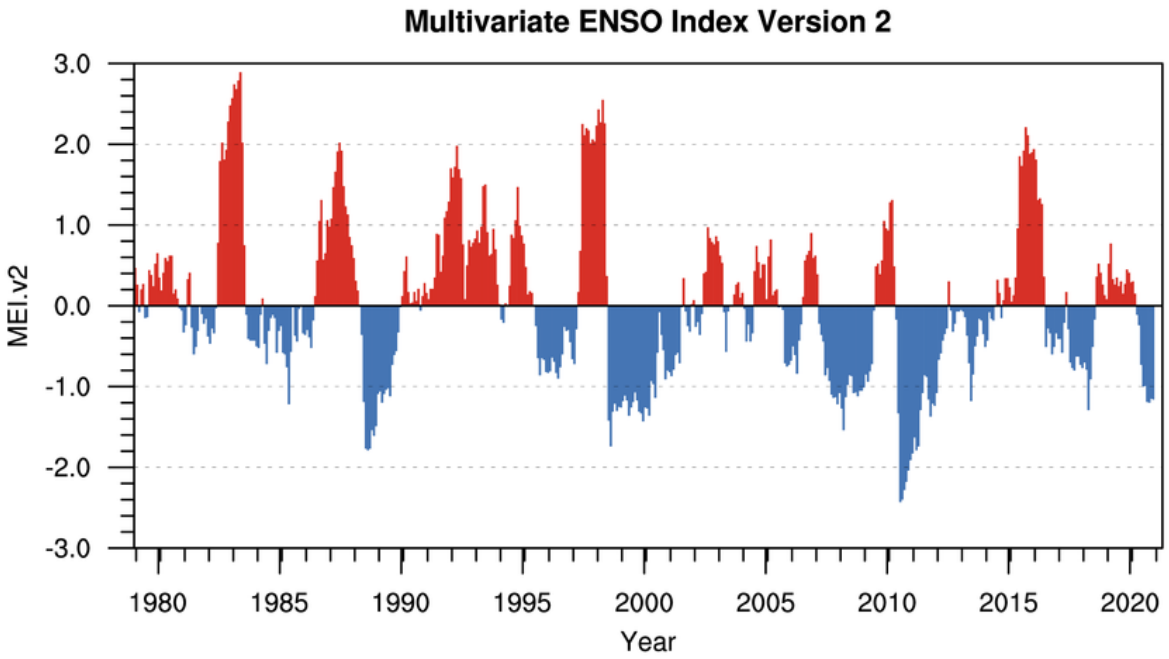


Figure 6.5: Evolution of a multivariate ENSO index from 1980 to 2020. Positive values are associated with El Niño events.

While ENSO is usually measured and characterized in terms of sea surface temperature anomalies in the equatorial Pacific, it is associated with strong short-term climate perturbations around the world. Many of these effects are summarize in Figure 6.6. Some of these weather patterns depend on the particular character of El Niño events, especially whether the positive sea surface temperature anomalies peak in the central versus eastern Pacific (Figure 6.1). Besides some of the characteristics illustrated in Figure 6.6, ENSO strongly modulates tropical cyclone activity in many regions; for example, El Niño tends to suppress North Atlantic tropical cyclone activity, which, conversely, is enhanced during La Niña conditions.

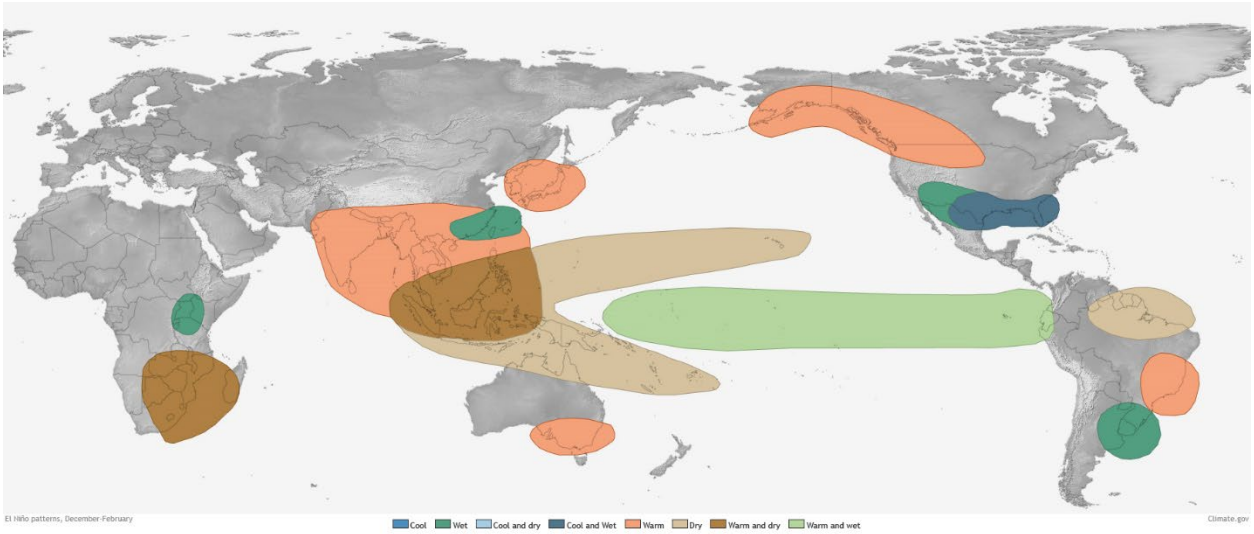


Figure 6.6: Short term climate variations during Boreal winter, associated with El Niño events. Light brown: dry conditions; lime green: warm and wet, ochre: warm, medium green: wet; dark brown: warm and dry, blue-green: cool and wet.

As has already been mentioned, ENSO has profound effects on marine life and is accompanied by strong shifts in everything from microbial populations to fish and seabirds. The very strong El Niño event of 1982-1983 led to large coral mortality in the tropical eastern Pacific. Consequently, in addition to the inherent intellectual interest in ENSO, it has serious consequences for biological activity and human society.

6.2 Oscillations of the equatorial ocean

The physics of ENSO involve the interaction of the tropical atmosphere and ocean, both of which support a rich spectrum of wave modes. In general, the atmospheric waves are fast compared to their oceanic counterparts, which introduces some interesting subtleties in the interactions between the two media. We begin with a discussion of linear wave modes on a resting equatorial ocean, continuing to approximate it as a two-fluid system with an inert deep ocean superimposed by a less dense layer, and treating the system using the shallow water approximation as in section 5.4.1. The shallow-water equations are phrased on an "equatorial $\beta-$plane", linearizing the sinusoidal latitudinal dependence of the Coriolis parameter about its vanishing value at the equator: $f \simeq \beta y$, where y is the distance north of the equator. These equations are

$$\frac{du}{dt} = -g' \frac{\partial h}{\partial x} + \beta y v + \frac{\tau_x}{h}, \tag{6.1}$$

$$\frac{dv}{dt} = -g' \frac{\partial h}{\partial y} - \beta y u + \frac{\tau_y}{h}, \tag{6.2}$$

$$\frac{dh}{dt} + h\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = 0, \tag{6.3}$$

where, as before, u and v are the zonal and meridional velocities, h is the fluid depth, g' is the reduce gravitational acceleration, given by (5.59), and τ_x and τ_y are the zonal and meridional surface wind stresses.

We begin by linearizing these three equations about a resting state in which the wind stresses and velocities are zero and the fluid depth is D. Denoting perturbations to this state with primes, we replace the variables with

$$\begin{aligned} u &\to u', \\ v &\to v', \\ \tau_x &\to \tau_x' \\ \tau_y &\to \tau_y' \\ h &\to D + h'. \end{aligned} \tag{6.4}$$

Substituting these into (6.1) – (6.3) and dropping quantities that are quadratic in the prime variables (but remember that the prime modifying g means something else) gives

$$\frac{\partial u'}{\partial t} = -g'\frac{\partial h'}{\partial x} + \beta y v' + \frac{\tau_x'}{D}, \tag{6.5}$$

$$\frac{\partial v'}{\partial t} = -g'\frac{\partial h'}{\partial y} - \beta y u' + \frac{\tau_y'}{D}, \tag{6.6}$$

and

$$\frac{\partial h'}{\partial t} + D\left(\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y}\right) = 0. \tag{6.7}$$

First consider the case where there is no surface wind stress $(\tau = 0)$. In that case, there is no forcing of this system, but it does admit neutral, oscillatory modes. Of these, let's first consider a special solution to (6.5) – (6.7) for which $v' = 0$ everywhere. In that case, we can see that (6.5) and (6.7) constitute a closed system, while (6.6) restricts the meridional variation of the mode.

Taking $\tau = 0$ and eliminating u' between (6.5) and (6.7) yields

$$\frac{\partial^2 h'}{\partial t^2} - g' D\frac{\partial^2 h'}{\partial x^2} = 0. \tag{6.8}$$

This is clearly a wave equation that admits separable solutions of the form

$$h' = G(y)F(x \pm ct), \tag{6.9}$$

where $c \equiv \sqrt{g'D}$, F is any reasonably behaved function, and $G(y)$ is a function of y that we will determine using (6.6). Substituting (6.9) into (6.6) with $v' = \tau_y' = 0$ gives an equation for $G(y)$:

$$\frac{1}{G}\frac{dG}{dy} = \pm\frac{\beta}{c}y, \tag{6.10}$$

whose solution is

$$G = Ae^{\pm\frac{\beta}{2c}y^2}, \tag{6.11}$$

where A is an arbitrary integration constant. Here, as a boundary conditions, we must insist that the solution be well behaved as $y \to \pm\infty$, so we must take the negative choice in (6.11), which also means we must take the negative root in (6.9). Therefore, we have as a viable solution

$$h' = Ae^{\frac{-\beta}{2c}y^2}F\left(x-ct\right), \tag{6.12}$$

with $c = \sqrt{g'D}$ with c positive and F any reasonably well-behaved function. This solution describes eastward-propagating, non-dispersive waves that decay exponentially away from the equator on a decay scale L_y given by

$$L_y \equiv \sqrt{\frac{2g'D}{c\beta}} = \sqrt{\frac{2c}{\beta}}. \tag{6.13}$$

This scale is called the *equatorial deformation radius*. Taking $g' = 0.06\ ms^{-2}$ and $D = 100\,m$ (see Figure 5.53), the eastward propagation speed c is about $2.5\ ms^{-1}$ and the equatorial deformation radius is about $650\,km$.

This special solution to (6.5) – (6.7) describes phenomena known as *equatorially trapped Kelvin waves*, named after William Thomson, Lord Kelvin (Thomson, 1880). These are the main means by which adjustments to localized forcing are propagated eastward in the equatorial ocean[1]. From (6.5) or (6.7) it is clear that the layer depth perturbations in zonal velocity perturbations are out of phase in these neutral waves, with positive depth perturbations associated with negative zonal velocity. The structure of an equatorial Kelvin wave is shown in Figure 6.7; here a sinusoidal zonal structure has been assumed.

[1] The waves described here are more precisely defined as first baroclinic mode waves. There are also barotropic Kelvin waves that fill the depth of the ocean and for which the restoring force is gravity acting on perturbations of the sea surface elevation. These have a phase speed of about $150\ ms^{-1}$, and while they can be excited by earthquakes, they are not important in ENSO dynamics.

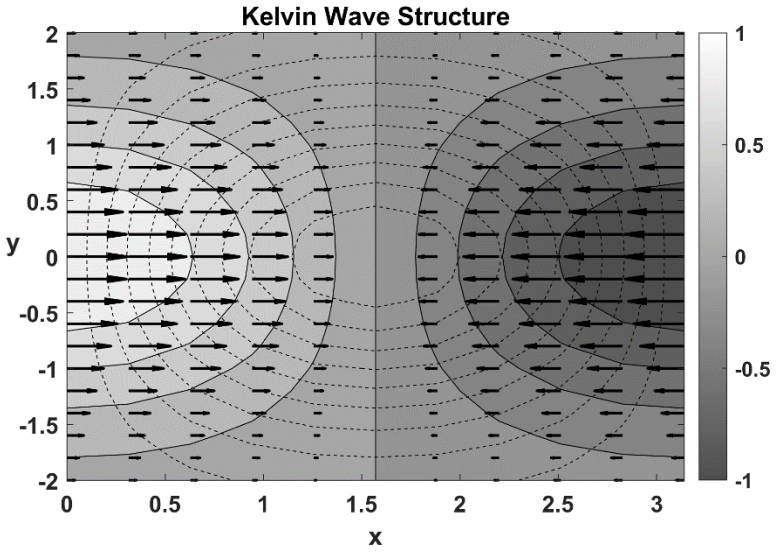


Figure 6.7: Structure of an equatorial Kelvin wave. All quantities are nondimensional, with the x axis in radians, spanning one-half wavelength. Shading shows thermocline depth perturbations, arrows show zonal velocity, and dashed contours show downwelling velocity. The wave is propagating eastward.

The general solutions to the unforced versions of (6.5) – (6.7) were first described by Taro Matsuno (Matsuno, 1966). Since the coefficients are independent of time and longitude, there is a class of solutions that are oscillatory in x and t and which we insist be well behaved (not blow up) when $y \to \pm\infty$. Thus we look for viable solutions of (6.5) – (6.7) with $\tau' = 0$ of the form

$$[u', v', h'] \to [U(y), V(y), H(y)] e^{ikx - i\omega t}, \tag{6.14}$$

where k is a zonal wavenumber and ω is the wave frequency. If we substitute (6.14) into (6.5) – (6.7) and eliminate $U(y)$ and $H(y)$ in favor of $V(y)$, we obtain a second-order ordinary differential equation:

$$\frac{d^2V}{dy^2} + V\left[\frac{\omega^2}{c^2} - k^2 - \frac{\beta k}{\omega} - \frac{\beta^2 y^2}{c^2}\right] = 0, \tag{6.15}$$

with $c \equiv \sqrt{g'D}$, as before. It is helpful to write this more simply by nondimensionalizing the frequency, wavenumber, and y according to

$$\begin{aligned} \omega &\to \sqrt{c\beta}\,\omega, \\ k &\to \sqrt{\beta/c}\,k, \\ y &\to \sqrt{c/\beta}\,y, \end{aligned} \tag{6.16}$$

whereupon (6.15) becomes

$$\frac{d^2V}{dy^2}+V\left[\omega^2-k^2-\frac{k}{\omega}-y^2\right]=0, \tag{6.17}$$

This is Schrödinger's equation for a quantum harmonic oscillator, which has well behaved solutions only if the wave frequency takes on certain discrete values given by the solution of

$$\omega^2-k^2-\frac{k}{\omega}=(2n+1), \tag{6.18}$$

where $n=0,1,2....$ and the eigenfunctions are parabolic cylinder functions:

$$V\sim e^{-\frac{1}{2}y^2}\left[1,2y\,,4y^2-2,...\right] \tag{6.19}$$

where the polynomial in brackets are the first three Hermite polynomials[2]. The n^{th} Hermite polynomial corresponds to the integer n in the dispersion relation (6.16).

Note that (6.18) is invariant to changing the signs of both ω and k as is the phase speed, ω/k. By convention, we allow negative values of k but not ω.

For $n=0,$ the cubic dispersion relation (6.18) can be factored:

$$(\omega+k)\left[\omega^2-\omega k-1\right]=0. \tag{6.20}$$

The particular root $\omega=-k$ is unphysical and corresponds to solutions that are not well behaved for large y, leaving two roots:

$$\omega=\frac{1}{2}\left[k\pm\sqrt{k^2+4}\right]. \tag{6.21}$$

Since by convention we allow negative k (corresponding to westward phase propagation) but not negative $\omega,$ we select the positive root of (6.21).

In the limit of large negative *k*, (6.21) becomes $\lim_{k\to-\infty}\omega=\frac{1}{|k|}$, or re-dimensionalizing using (6.16), $\omega\sim\frac{\beta}{|k|}$. In this limit, the solutions are equatorially trapped, westward propagating Rossby waves (remember that this is a limit in which k is negative). When k is large and positive, (6.21) becomes $\lim_{k\to\infty}\omega=k,$ or in dimensional terms, $\omega=ck.$ This is the dispersion relation for Kelvin waves, in which buoyancy is the restoring force. In between these limits, the index of refraction depends on both c and $\beta,$ consequently, these $n=0$ solutions are called *mixed Rossby-gravity waves*, or alternatively *Yanai waves*, named after the tropical

[2] A general expression for the n^{th} Hermite polynomial is $H_n(y)=(-1)^n e^{y^2}\frac{d^n}{dy^n}e^{-y^2}$.

meteorologist Michio Yanai (Yanai and Maruyama, 1966) who discovered them in stratospheric data.

For $n > 0$ solutions to (6.20) fall into two classes: High frequency, equatorially trapped internal gravity-inertia waves called *Poincare waves*, and low frequency, strictly-westward propagating Rossby–like waves. For the low-frequency modes, we ignore the first term in (6.18) and re-dimensionalize the result using (6.16) to get

$$\omega \simeq \frac{-k\beta}{k^2 + (2n+1)\frac{\beta}{c}}. \tag{6.22}$$

In the longwave limit, they behave like westward-propagating inertia-gravity waves, but in the high negative (westward) zonal wavenumber limit they are equatorially trapped Rossby waves. The phase speeds of these low frequency modes are always negative but their group velocity can be positive for sufficiently short waves.

In the high frequency limit, we ignore the third term in (6.18) and in that case, the dimensional solution is

$$\omega^2 \simeq c^2k^2 + (2n+1)\beta c. \tag{6.23}$$

These are the eastward and westward propagating Poincare waves.

The full dispersion diagram illustrating the solutions to (6.18) are shown in Figure 6.8. Note that the Poincare waves and the Yanai wave have nonzero frequency when $k = 0,$ indicating zonally symmetric standing oscillations.

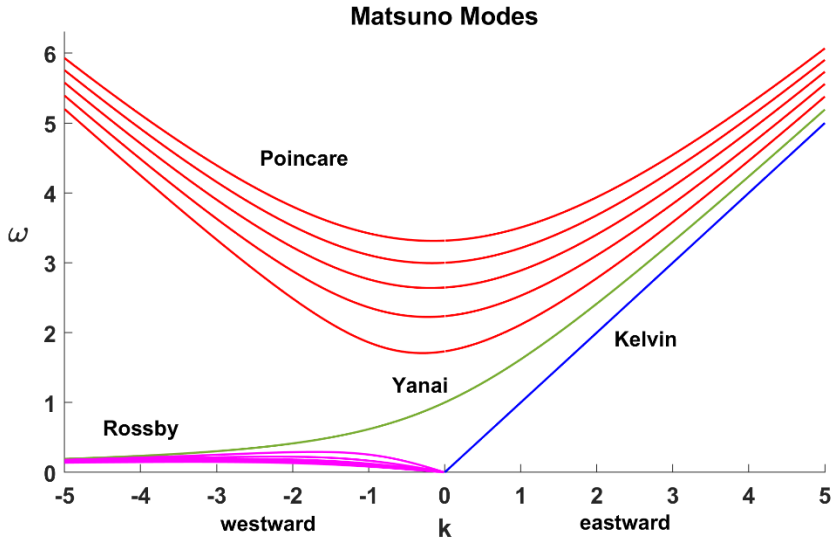


Figure 6.8: Nondimensional frequency as a function of nondimensional zonal wavenumber; solutions to (6.18). The Kelvin wave is shown in blue, the Yanai wave in green, Equatorial Rossby waves in magenta, and Poincare waves in red.

An example of an eigenfunction, for the $k = -1$ Yanai wave, is shown in Figure 6.9. This mode has a relatively complex structure, with vanishing depth, zonal velocity and vertical velocity at the equator, and maximum values at about one equatorial deformation radius north and south of the equator. Cyclonic and anticyclonic gyres alternate, with gyre centers at the equator. The vertical velocity is in quadrature with the depth perturbations. The interested reader can plot eigenfunctions of other modes using the script *Matsuno.m* available at https://zenodo.org/doi/10.5281/zenodo.10843630 .

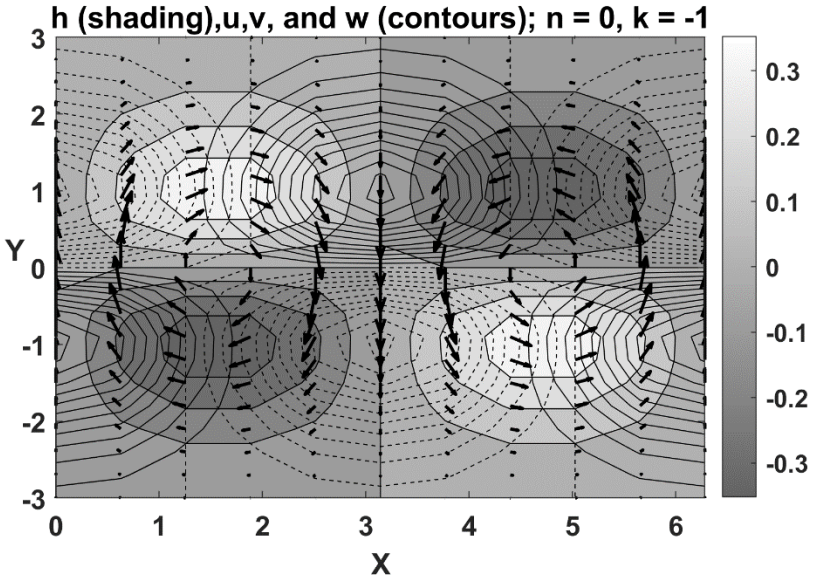


Figure 6.9: Eigenfunction for the Yanai wave (n=0) with k=-1. The shading shows the nondimensional fluid depth perturbation, the arrows shows the horizontal perturbation velocity vectors, and the thin black contours show the upward motion, with dashed contours indicating negative values.

Note that if $c = 2.5\ ms^{-1}$, and using the equatorial value of β, the dimensional wavelength associated with wavenumber 1 is about 2,800 km and a nondimensional frequency of 1 corresponds to a period of about 14 days. Consulting Figure 6.8 and noting that El Niño evolves over months, the main kinds of waves involved are low frequency Rossby and Kelvin waves. We will return to the role of these waves in ENSO evolution in section 6.4, but on our way to thinking about the fully coupled problem we must first look at the atmospheric response to surface anomalies on the scale of the equatorial Pacific.

6.3 The atmospheric response to large-scale equatorial surface flux anomalies

ENSO evolves slowly enough that, to a first approximation, the atmospheric response to ENSO-related sea surface temperature anomalies can be considered steady on atmospheric dynamical time scales. On the other hand, the very large spatial scales and slow time scales of ENSO imply that the weak-temperature -gradient approximation cannot be applied here as we have, for example, in considering the Walker Circulation on the equator (Chapter 5). We must include the effect of Coriolis accelerations, but as we did in examining equatorial ocean

dynamics, we will approximate its variation as linear, continuing to use the equatorial β-plane approximation.

The principal simplifying assumption we will make about the atmosphere is that in regions experiencing deep convection, the lapse rate of temperature is moist adiabatic. As noted in Chapter 3, this is well satisfied, especially if one assumes reversible moist adiabatic ascent from the top of the subcloud layer. But here we will ignore the direct effects of water substance on density and assume that the saturation entropy is constant with altitude above cloud base, and equal to the actual moist entropy of air in the subcloud layer. As a reminder, the saturation entropy is given by

$$s^* = c_p \ln\left(T/T_0\right) - R_d \ln\left(p/p_0\right) + \frac{L_v q^*}{T}, \tag{6.24}$$

where T_0 and p_0 are reference temperature and pressure, respectively. We assume that while s^* varies in horizontal space and in time, it is always uniform above cloud base in the troposphere.

6.3.1 Dynamics

Now consider linear perturbations to a resting tropical atmosphere in RCE, denoting departures from the resting state with primes. First, we write the hydrostatic equation in pressure coordinates as

$$\frac{\partial \phi'}{\partial p} = -\alpha', \tag{6.25}$$

where, as usual, ϕ is the geopotential and α is the specific volume.

Using the chain rule and one of Maxwell's relations (Chapter 2), this can be written

$$\alpha' = \left(\frac{\partial \alpha}{\partial s^*}\right)_p s^{*\prime} = \frac{1}{1 + r_t}\left(\frac{\partial T}{\partial p}\right)_{s^*} s^{*\prime}. \tag{6.26}$$

This shows that in a hydrostatic, most adiabatic atmosphere, specific volume perturbations are constrained to have moist adiabatic vertical profiles. Substituting (6.26) into (6.25) gives

$$\frac{\partial \phi'}{\partial p} \simeq -\left(\frac{\partial T}{\partial p}\right)_{s^*} s^{*\prime}. \tag{6.27}$$

Because we are assuming that $s^{*\prime}$ is not a function of pressure, we can integrate (6.27) from the top of the subcloud layer to the tropopause to get

$$\phi' = \phi_{bt}'(x, y, t) + \left(\overline{T} - T(p)\right) s^{*\prime}(x, y, t), \tag{6.28}$$

where we have divided the integration constant into two parts: a height-independent mean temperature, $\overline{T}$, that multiplies the perturbation saturation entropy, and a height-independent

geopotential, ϕ_{bt}. The former is just the pressure-weighted mean temperature of the troposphere above cloud base, and the latter we call the barotropic component of the geopotential. Note from the definition of $\overline{T}$ that the pressure-mean geopotential perturbation consists only of its barotropic component. The second term on the right of (6.28) describes what has come to be called the *first baroclinic mode* of tropical dynamics.

The relation (6.28) is a very powerful constraint on the dynamics of deep convecting regimes. It filters all vertical modes except the barotropic and first baroclinic modes, and reduces the full primitive equations to a barotropic vorticity equation and equations that take the mathematical form of shallow water equations, as shown presently.

The hydrostatic horizontal momentum equations linearized about a state of rest on an equatorial best plane are

$$\frac{\partial u'}{\partial t} = -\frac{\partial \phi'}{\partial x} + \beta y v' + g\frac{\partial \tau_x'}{\partial p}, \tag{6.29}$$

and

$$\frac{\partial v'}{\partial t} = -\frac{\partial \phi'}{\partial y} - \beta y u' + g\frac{\partial \tau_y'}{\partial p}, \tag{6.30}$$

where τ_x and τ_y are the horizontal components of wind stress. We specialize these to the cloud layer by assuming that the stress terms vanish above the subcloud layer, and by substituting (6.28) for the geopotential terms:

$$\frac{\partial u'}{\partial t} = -\frac{\partial \phi_{bt}'}{\partial x} - \left(\overline{T} - T\right)\frac{\partial s^{*\prime}}{\partial x} + \beta y v', \tag{6.31}$$

and

$$\frac{\partial v'}{\partial t} = -\frac{\partial \phi_{bt}'}{\partial y} - \left(\overline{T} - T\right)\frac{\partial s^{*\prime}}{\partial y} - \beta y u', \tag{6.32}$$

where $\overline{T}$ is now a constant representing the basic state, and T is a function of pressure only. Inspection of these two equations shows that the solutions for the horizontal velocities can be broken down the same way as geopotential, namely

$$(u', v') = (u_{bt}', v_{bt}') + \frac{\overline{T} - T}{\overline{T} - T_{PBL}}\left(u_{bc}', v_{bc}'\right), \tag{6.33}$$

where the subscript *bt* and *bc* refer to the barotropic and baroclinic components, respectively, and T_{PBL} is the base state temperature at the top of the subcloud layer. Note that the primed variables in (6.33) are functions of x, y, and t only; all the vertical structure is in the $\overline{T} - T$ term. Substitution of the forms (6.33) into (6.31) and (6.32) gives two pairs of equations for the baroclinic and barotropic components. The set for the baroclinic component is

$$\frac{\partial u_{bc}'}{\partial t} = -\left(\overline{T} - T_{PBL}\right)\frac{\partial s^{*}'}{\partial x} + \beta y v_{bc}', \tag{6.34}$$

and

$$\frac{\partial v_{bc}'}{\partial t} = -\left(\overline{T} - T_{PBL}\right)\frac{\partial s^{*}'}{\partial y} - \beta y u_{bc}'. \tag{6.35}$$

Clearly, we need an equation for the tropospheric temperature, as represented by s^{*}' to close the baroclinic system. If this equation depends, among other things, on the barotropic component of the pressure and/or velocity, then we will also need to solve for the barotropic component even if our interest were to lie solely with the baroclinic component.

The equations for the barotropic component are decoupled from temperature:

$$\frac{\partial u_{bt}'}{\partial t} = -\frac{\partial \phi_{bt}'}{\partial x} + \beta y v_{bt}', \tag{6.36}$$

and

$$\frac{\partial v_{bt}'}{\partial t} = -\frac{\partial \phi_{bt}'}{\partial y} - \beta y u_{bt}'. \tag{6.37}$$

We can form a vorticity equation by eliminating ϕ_{bt}' from (6.36) and (6.37):

$$\frac{\partial}{\partial t}\left(\frac{\partial v_{bt}'}{\partial x} - \frac{\partial u_{bt}'}{\partial y}\right) = -\beta y\left(\frac{\partial u_{bt}'}{\partial x} + \frac{\partial v_{bt}'}{\partial y}\right) - \beta v_{bt}'. \tag{6.38}$$

Now we can write the full equation of mass continuity as

$$\frac{\partial \omega'}{\partial p} = -\left(\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y}\right) = -\left(\frac{\partial u_{bt}'}{\partial x} + \frac{\partial v_{bt}'}{\partial y}\right) - \frac{\overline{T} - T}{\overline{T} - T_{PBL}}\left(\frac{\partial u_{bc}'}{\partial x} + \frac{\partial v_{bc}'}{\partial y}\right), \tag{6.39}$$

where we have substituted (6.33) breaking down the velocities into baroclinic and barotropic components. We next integrate (6.39) down from the tropopause to an arbitrary pressure p :

$$\omega' = \omega_t' - \left(p - p_t\right)\left(\frac{\partial u_{bt}'}{\partial x} + \frac{\partial v_{bt}'}{\partial y}\right) - J(T)\left(\frac{\partial u_{bc}'}{\partial x} + \frac{\partial v_{bc}'}{\partial y}\right), \tag{6.40}$$

where ω_t' is the value of ω' at the tropopause, whose pressure is p_t, and

$$J(T) \equiv \int_{p_t}^{p} \frac{\overline{T} - T}{\overline{T} - T_{PBL}} dp. \tag{6.41}$$

Evaluating (6.40) at the top of the subcloud layer gives

$$\omega'_{PBL} = \omega'_t - \Delta P\left(\frac{\partial u_{bt}'}{\partial x} + \frac{\partial v_{bt}'}{\partial y}\right), \tag{6.42}$$

where $\Delta P \equiv p_{PBL} - p_t$ is the pressure depth of the cloud layer. The last term in (6.40) vanishes because, by definition, $J(T)$ vanishes at the top of the subcloud layer.

Note that the barotropic vorticity equation (6.38) together with (6.42) constitutes a close system if both ω'_t and ω'_{PBL} are known. We have to determine these two variables by matching what is going on in the cloud layer with equations expressing the dynamics of the stratosphere above, and equations describing the behavior of the subcloud layer below. In particular, if our tropospheric system supports waves, such as Kelvin and Rossby waves, we would expect them to couple to some extent with wave modes in the stratosphere. The latter has to be treated as a continuously stratified fluid (as its name implies!) but open-ended at the top, so that we may expect upward wave energy propagation. The leakage of wave energy from the troposphere may damp disturbances there; a point we shall return to in Chapter 7[3]. But for the very slowly evolving ENSO, leakage of energy into the stratosphere will be very small, and so for the present purpose we take the tropopause to act as a rigid lid:

$$\omega'_t = 0. \tag{6.43}$$

This leaves us with the problem if solving for the vertical motion at the top of the subcloud layer, ω'_{PBL}. To do so, we must write down linearized dynamics for the subcloud layer.

Over tropical oceans, the subcloud layer has nearly constant virtual potential temperature, and although the value of that temperature will vary with conditions, we will ignore these here and assume that the horizontal pressure gradients do not vary with depth within the layer, so that, in effect, the pressure gradient in the layer is imposed from above. We shall also assume that the frictional deceleration of the horizontal flow owing to surface friction is vertically uniform. These two conditions imply that the wind components themselves are independent of height. With these assumptions, the linearized momentum equations in the subcloud layer are

$$\frac{\partial u_{sc}'}{\partial t} = -\frac{\partial \phi_{bt}'}{\partial x} - \left(\overline{T} - T_{PBL}\right)\frac{\partial s^{*\prime}}{\partial x} + \beta y v_{sc}' + g\frac{\tau_{xs}'}{\Delta p}, \tag{6.44}$$

and

$$\frac{\partial v_{sc}'}{\partial t} = -\frac{\partial \phi_{bt}'}{\partial y} - \left(\overline{T} - T_{PBL}\right)\frac{\partial s^{*\prime}}{\partial y} - \beta y u_{sc}' + g\frac{\tau_{ys}'}{\Delta p}, \tag{6.45}$$

[3] If there is enough wind shear in the stratosphere, as for example accompanies the Quasi-Biennial Oscillation, there may be altitudes at which the background wind speed matches the speed of waves emanating from the troposphere. In that case, some wave energy may backscatter from these “critical levels” back down into the troposphere, providing a possible feedback of stratospheric dynamics on tropospheric waves.

where T_{PBL} is the base state temperature at the top of the subcloud layer, τ_{xs} and τ_{ys} are the surface wind stresses, Δp is the pressure depth of the subcloud layer, and the subscript sc refers to subcloud layer quantities. We can simplify these greatly by changing variables:

$$\begin{aligned} u_{sc}' &\rightarrow u_{bt}' + u_{bc}' + u_f', \\ v_{sc}' &\rightarrow v_{bt}' + v_{bc}' + v_f', \end{aligned} \tag{6.46}$$

where the new variables u_f' and v_f' represent departures of the subcloud layer velocities from the sum of the barotropic and baroclinic velocities just above the subcloud layer. Substitution of (6.46) into (6.44) and (6.45), and making use of (6.34) – (6.37) yields

$$\frac{\partial u_f'}{\partial t} = \beta y v_f' + g\frac{\tau_{xs}'}{\Delta p}, \tag{6.47}$$

and

$$\frac{\partial v_f'}{\partial t} = -\beta y u_f' + g\frac{\tau_{ys}'}{\Delta p}. \tag{6.48}$$

In the absence of any surface stresses, these describe pure inertial oscillations on an equatorial β plane and we assume that in this limit they vanish. But in general, surface stresses depend on the *full* surface winds, as expressed by (6.46), and so (6.47) and (6.48) couple the boundary layer winds to both the barotropic and baroclinic wind components. In particular, *surface friction fundamentally couples the barotropic and baroclinic components*.

Mass continuity applied to the boundary layer flow and integrated over the depth of the subcloud layer yields

$$\omega_{PBL}' = \Delta p\left(\frac{\partial u_f'}{\partial x} + \frac{\partial v_f'}{\partial y} + \frac{\partial u_{bt}'}{\partial x} + \frac{\partial v_{bt}'}{\partial y} + \frac{\partial u_{bc}'}{\partial x} + \frac{\partial v_{bc}'}{\partial y}\right). \tag{6.49}$$

If we eliminate ω_{PBL}' between (6.49) and (6.42), with $\omega_t' = 0,$ the result is

$$\left(\frac{\partial u_{bt}'}{\partial x} + \frac{\partial v_{bt}'}{\partial y}\right) = \frac{-\Delta p}{p_0 - p_t}\left(\frac{\partial u_f'}{\partial x} + \frac{\partial v_f'}{\partial y} + \frac{\partial u_{bc}'}{\partial x} + \frac{\partial v_{bc}'}{\partial y}\right). \tag{6.50}$$

This determines the barotropic divergence as a function of the divergence of the baroclinic and frictional components. This shows that the barotropic divergence is small compared to the sum of the baroclinic and frictional divergences. If we substitute (6.50) back into (6.49) we get

$$\omega_{PBL}' = \Delta p\left(1 - \frac{\Delta p}{p_0 - p_t}\right)\left(\frac{\partial u_f'}{\partial x} + \frac{\partial v_f'}{\partial y} + \frac{\partial u_{bc}'}{\partial x} + \frac{\partial v_{bc}'}{\partial y}\right). \tag{6.51}$$

Since boundary layers over tropical oceans are thin compared to the depth of the troposphere, the first bracketed term on the right side of (6.51) may be approximated by unity.

Now, except for needing a relation for the tropospheric temperature $s^{*\prime}$ together with the specification of surface wind stresses in terms of perturbation surface winds, the relations (6.34) – (6.35), (6.38), (6.47), (6.48), and (6.51) would constitute a closed system for the variables $u_{bc}', v_{bc}', u_f', v_f', \omega'_{PBL}$ and $s^{*\prime}$(which, as well see presently, depends on ω'_{PBL}).

We next turn our attention to the thermodynamics that determine the perturbation saturation entropy, $s^{*\prime}$.

6.3.2 Thermodynamics

The basic thermodynamic framework we will use here is very similar to that developed in Chapter 3 but extended to non-steady conditions. Referring back to Figure 3.17, we can approximate the saturation entropy of the atmosphere by its value outside of deep convective clouds, since the latter cover a small fractional area and have density temperatures not very different from the clear air. The temperature tendency in the clear air reflects two competing processes: subsidence, which warms the column, and radiative cooling. For convenience, we repeat (3.56) here:

$$\frac{\partial h_d}{\partial t} = 0 = w_d \frac{d\overline{h}_d}{dz} - \alpha \dot{Q}_{cool} = w_d \mathbf{S} - \alpha \dot{Q}_{cool}, \tag{6.52}$$

where h_d is the dry static energy, $\mathbf{S}$ is the dry static stability along the base state moist adiabat, $\mathbf{S} \equiv d\overline{h}_d / dz,$ and w_d is the vertical velocity between clouds, defined positive downward. We also repeat the equation for conservation of mass, (3.51):

$$w = \alpha \left(M_u - M_d \right) - w_d, \tag{6.53}$$

where w is the ensemble-averaged (large-scale) vertical velocity, α is specific volume, and M_u and M_d are the convective updraft and downdraft mass fluxes. In (6.53) we have neglected the fractional area σ covered by convection, compared to unity. We also repeat our formulation (3.58) relating convective downdraft mass flux to convective updraft mass flux through a precipitation efficiency, ϵ_p:

$$M_d = \left(1 - \epsilon_p\right) M_u. \tag{6.54}$$

Substituting (6.53) into (6.52) and using (6.54) gives

$$\frac{\partial h_d}{\partial t} = \mathbf{S}\left(\alpha \epsilon_p M_u - w\right) - \alpha \dot{Q}_{cool}. \tag{6.55}$$

Air is warmed by convective updrafts, (through their forcing of environmental subsidence) and cooled by radiation and by adiabatic cooling associated with large-scale ascent.

Now both the dry static energy, h_d and the saturation entropy, s^*, are functions of pressure and temperature. Thus, on isobaric surfaces, there is a one-to-one relationship between

fluctuations of the two quantities. First, we relate isobaric fluctuations of dry static energy to fluctuations of dry air entropy:

$$\delta h_d = T \delta s_d . \tag{6.56}$$

Now we relate isobaric fluctuations of specific volume to fluctuations of dry air entropy and to fluctuations of saturated entropy using (2.47) and (2.74), respectively:

$$\delta \alpha = \left(\frac{\partial T}{\partial p} \right)_{s_d} \delta s_d = \frac{1}{1 + r_t} \left(\frac{\partial T}{\partial p} \right)_{s*} \delta s^* . \tag{6.57}$$

Neglecting the total water concentration r_t compared to unity, and using the definitions of dry and moist adiabatic lapse rates, Γ_d and Γ_m, we can write (6.57) as

$$\delta s^* = \frac{\Gamma_d}{\Gamma_m} \delta s_d . \tag{6.58}$$

Using (6.56) and (6.58), we can write the thermodynamic equation (6.55) as

$$T \frac{\Gamma_m}{\Gamma_d} \frac{\partial s^*}{\partial t} = \mathbf{S} \left(\alpha \epsilon_p M_u - w \right) - \alpha \dot{Q}_{cool} . \tag{6.59}$$

This will be the form of the thermodynamic equation we will use in deep convecting tropical atmospheres. Remember that we are continuing to assume that the temperature lapse rate is always moist adiabatic, and so s^* is constant with height. We choose to evaluate (6.59) just above the boundary layer in what follows.

As we did in Chapter 3, we will close for the convective updraft mass flux using the boundary layer quasi-equilibrium hypothesis, (3.53):

$$M_u = \rho w_{PBL} + \frac{F_h}{h_b - h_m} , \tag{6.60}$$

where recall that F_h is the surface enthalpy flux and h_b and h_m are the subcloud layer and free troposphere moist static energies. Substituting (6.60) into (6.59) evaluated just above the top of the boundary layer gives

$$\left(T \frac{\Gamma_m}{\Gamma_d} \right)_{PBL} \frac{\partial s^*}{\partial t} = \mathbf{S} \left(\frac{\alpha \epsilon_p F_h}{h_b - h_m} - \left(1 - \epsilon_p \right) w_{PBL} \right) - \left(\alpha \dot{Q}_{cool} \right)_{PBL} . \tag{6.61}$$

We note that in our base state of RCE, there is no time tendency of temperature and $w = 0$, so from (6.61)

$$\mathbf{S} \epsilon_p \overline{\alpha F}_h = \left(\overline{h_b} - \overline{h_m} \right) \left(\overline{\alpha \dot{Q}_{cool}} \right)_{PBL} , \tag{6.62}$$

where the overbar represented the base state. If we make the further approximation that the base state (RCE) radiative cooling rate per unit mass is relatively constant with height in the

troposphere (not usually a bad approximate; see Figure 3.21a), then overall energy balance requires that

$$\overline{\alpha F}_h = H\left(\overline{\alpha \dot{Q}_{cool}}\right)_{PBL}, \tag{6.63}$$

where $H = \frac{1}{g}\int_{p_t}^{p_0} \alpha dp$ is the scale height of the troposphere. We next linearize (6.61) around the basic RCE state, making use of (6.62) and (6.63) and ignoring fluctuations in precipitation efficiency and in the basic state stratification, $\mathbf{S}$, yielding

$$\left(T\frac{\Gamma_m}{\Gamma_d}\right)_{PBL}\frac{\partial s^{*\prime}}{\partial t} = \frac{\left(\alpha F_h{}'\right)}{H} - \frac{\overline{\alpha \dot{Q}_{cool}}}{\epsilon_p H \mathbf{S}}\left(h_b{}' - h_m{}'\right) - \mathbf{S}\left(1-\epsilon_p\right)w'_{PBL} - \left(\alpha \dot{Q}_{cool}\right)'_{PBL}. \tag{6.64}$$

Because we assume moist convective neutrality, fluctuations in subcloud layer moist entropy, $s_b{}'$ must equal fluctuations in the saturation entropy in the cloud layer, $s^{*\prime}$. Moreover, fluctuations of moist static energy, $h_b{}'$, are, to a good approximation, related to fluctuations of entropy by $h_b{}' \cong T s_b{}'$, so that we may write (6.64) as

$$\left(T\frac{\Gamma_m}{\Gamma_d}\right)_{PBL}\frac{\partial s^{*\prime}}{\partial t} = \frac{\left(\alpha F_h{}'\right)}{H} - \frac{\overline{\alpha \dot{Q}_{cool}}}{\epsilon_p H \mathbf{S}}\left(T s^{*\prime} - h_m{}'\right) - \mathbf{S}\left(1-\epsilon_p\right)w'_{PBL} - \left(\alpha \dot{Q}_{cool}\right)'_{PBL}. \tag{6.65}$$

This will be the linearized form of the tropospheric thermodynamic equation we will use here and in the next chapter, so it is worth describing the terms in some detail:

Term in (6.65)	Description
$\frac{\left(\alpha F_h{}'\right)}{H}$	Perturbations in the surface enthalpy flux lead, through deep convection, to heating or cooling of the cloud layer
$-\frac{\overline{\alpha \dot{Q}_{cool}}}{\epsilon_p H \mathbf{S}}\left(T s^{*\prime} - h_m{}'\right)$	A positive perturbation in boundary entropy, also proportional to $s^{*\prime}$, reduces the boundary layer quasi-equilibrium convective mass flux, cooling the troposphere. Likewise, a positive perturbation in mid-tropospheric moist static energy increases the mass flux, warming the troposphere.
$-\mathbf{S}\left(1-\epsilon_p\right)w'_{PBL}$	$\mathbf{S}\left(1-\epsilon_p\right)$ is the effective stratification of a moist convecting atmosphere. It is the dry stratification modified by moist convection according to how efficiently the convection precipitates.
$-\left(\alpha \dot{Q}_{cool}\right)'_{PBL}$	Perturbations in the radiative cooling rate directly cause change in tropospheric temperature.

Aside from needing to specify the perturbation surface fluxes and radiative cooling in terms of known variables, we also need an expression for the perturbation moist static energy of the cloud layer, and for that we turn to (3.64):

$$\frac{\partial h_m}{\partial t} = -\left\langle \mathbf{V}_h \bullet \nabla h \right\rangle - G w_{PBL} - \left\langle \alpha \dot{Q}_{cool} \right\rangle + \frac{g}{p_0 - p_t} F_h. \tag{6.66}$$

Recall that the angle brackets denote averaging over the troposphere, and that G is the gross moist stability. Linearizing this about the RCE state gives

$$\frac{\partial h_m'}{\partial t} = -G w'_{PBL} - \left\langle \alpha \dot{Q}_{cool} \right\rangle' + \frac{g}{p_0 - p_t} F_h'. \tag{6.67}$$

We represent surface enthalpy fluxes in terms of the aerodynamic flux formula (3.28) assuming neutral stability of the surface layer:

$$\alpha F_h = C_k \, | \mathbf{V}_s | \left(h_0^* - h_b \right), \tag{6.68}$$

where C_k is a nondimensional surface exchange coefficient, $| \mathbf{V}_s |$ is the surface wind speed, and h_0^* is the saturation moist static energy at sea surface temperature. In linearizing this about the RCE state, we assume that, in keeping with average conditions in the tropics, there is background easterly surface wind. In that case, the perturbation enthalpy flux is given by

$$\alpha F_h' = C_k \, | \bar{U} | \left(h_0^{*\prime} - h_b' \right) - C_k u_{sc}' \left(\overline{h_0^*} - \overline{h_b} \right), \tag{6.69}$$

where $| \bar{U} |$ is the magnitude of the background easterlies and u_{sc}' is the perturbation background zonal flow in the subcloud layer. The first term on the right side of (6.69) represents the effect of perturbations in sea surface temperature and boundary layer moist static energy on surface fluxes, while the second term indicates that easterly wind perturbations will, by adding to the background easterlies, enhance the surface enthalpy flux.

Aside from the radiative cooling perturbations, we now have a closed linear system consisting of (6.34), (6.35), (6.47), (6.48), (6.51), (6.65), (6.67) and (6.69). This is quite a notationally unwieldy system, but we can make it less cumbersome both by introducing some further small approximations and by nondimensionalizing the dependent and independent variables.

First, as indicated by (6.50) and (6.38), the barotropic velocities will be much smaller than the baroclinic and frictional components and we hereafter neglect them. We also approximate the first term in brackets in (6.51) as unity and use the hydrostatic equation to re-express it in terms of the vertical velocity at the top of the boundary layer:

$$w_{PBL}' \cong -d \left(\frac{\partial u_f'}{\partial x} + \frac{\partial v_f'}{\partial y} + \frac{\partial u_{bc}'}{\partial x} + \frac{\partial v_{bc}'}{\partial y} \right), \tag{6.70}$$

where d is the physical depth of the subcloud layer.

In equations (6.47) and (6.48) we represent the surface stress terms using neutral aerodynamic flux formulae, linearized around an easterly zonal mean wind to arrive at

$$\frac{\partial u_f'}{\partial t} = \beta y v_f' - \frac{2 C_d \, | \bar{U} |}{d} \left(u_f' + u'_{bc} \right), \tag{6.71}$$

and

$$\frac{\partial v_f{}'}{\partial t} = -\beta y u_f{}' - \frac{C_d |\bar{U}|}{d}\left(v_f{}' + v'_{bc}\right). \tag{6.72}$$

The factor of 2 in the drag term in (6.71) results because perturbations in the zonal wind also contribute linearly to the net surface wind speed.

By adding equations (6.34) and (6.35) to (6.71) and (6.72), we notice that now the frictional and baroclinic velocities only appear in combination, so we can replace their sum by a single new variable.

In (6.67) we approximate the vertically averaged perturbation radiative cooling, $\left\langle \alpha \dot{Q}_{cool} \right\rangle'$, by its value just above the top of the subcloud layer, $\left(\alpha \dot{Q}_{cool}\right)'_{PBL}$.

Finally, we replace the dimensional independent and dependent variables by nondimensional counterparts as follows:

$$\begin{aligned}
&x \to ax, \qquad y \to L_y y, \qquad t \to \frac{a}{\beta L_y{}^2} t, \\
&u'_{bc} + u'_f \to \frac{aC_k |\bar{U}|}{H} u, \\
&v'_{bc} + v'_f \to \frac{L_y C_k |\bar{U}|}{H} v, \\
&w'_{PBL} \to \frac{dC_k |\bar{U}|}{H} w, \\
&\left(s^{*\prime}, s_0{}^{*\prime}\right) \to \frac{\beta L_y{}^2}{T_{PBL} - \bar{T}} \frac{aC_k |\bar{U}|}{H} \left(s, s_0\right), \\
&h'_m \to \frac{\epsilon_p \left(1-\epsilon_p\right) \mathbf{S}^2 H d}{\overline{h_0{}^*} - \overline{h^*}} h_m, \\
&\left(\alpha \dot{Q}_{cool}\right)'_{PBL} \to \frac{C_k |\bar{U}| \left(1-\epsilon_p\right) \mathbf{S} d}{H} Q,
\end{aligned} \tag{6.73}$$

where

$$L_y \equiv \left(\frac{\Gamma_d}{\Gamma_m} \frac{T_{PBL} - \bar{T}}{T_{PBL}} \frac{\mathbf{S} d \left(1-\epsilon_p\right)}{\beta^2} \right)^{\frac{1}{4}}. \tag{6.74}$$

With these substitutions, the linear perturbation equations take the form

$$\frac{\partial u}{\partial t} = \frac{\partial s}{\partial x} + yv - 2Du, \tag{6.75}$$

$$\frac{\partial v}{\partial t} = \delta\left(\frac{\partial s}{\partial y} - yu\right) - Dv, \tag{6.76}$$

$$w = -\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right), \tag{6.77}$$

$$\frac{\partial s}{\partial t} = \chi\left(s_0 - s\right) - \chi_2 s - \alpha u - w + h_m - Q, \tag{6.78}$$

and

$$\gamma\frac{\partial h_m}{\partial t} = \chi\left(s_0 - s\right) - \alpha u - G_m w - Q, \tag{6.79}$$

where the nondimensional coefficients are defined as follows:

$$\begin{aligned}
\delta &\equiv \frac{a^2}{L_y{}^2}, \\
D &\equiv \frac{C_k\,|\bar{U}\,|\,a}{d\beta L_y{}^2}, \\
\chi &\equiv \frac{C_k\,|\bar{U}\,|\,a}{H}\sqrt{\frac{T_{PBL}}{T_{PBL}-\bar{T}}\frac{\Gamma_d}{\Gamma_m}\frac{1}{\mathbf{S}d\left(1-\epsilon_p\right)}}, \\
\chi_2 &\equiv \chi\left(\frac{\overline{h^*{}_0} - \overline{h^*}}{\epsilon_p \mathbf{S} H}\right), \\
\alpha &\equiv \frac{C_k a}{H}\frac{\overline{h^*{}_0} - \overline{h^*}}{\mathbf{S}d\left(1-\epsilon_p\right)}, \\
G_m &\equiv \frac{G}{\mathbf{S}\left(1-\epsilon_p\right)}, \\
\gamma &\equiv \frac{\epsilon_p \mathbf{S} H^2 \beta L_y{}^2}{C_k\,|\bar{U}\,|\,a\left(\overline{h^*{}_0} - \overline{h^*}\right)}.
\end{aligned} \tag{6.80}$$

The set (6.75)-(6.81) constitutes our model of linear perturbations to deep convecting regions on an equatorial $\beta-$plane. In the next section we will use these to estimate the response of the convecting equatorial atmosphere to sea surface temperature perturbations.

6.3.3 Tropical atmospheric response to slowly evolving sea surface temperature anomalies

ENSO evolves on time scales that are long compared to the atmospheric dynamic response to changing sea surface temperature. The time scale used in (6.73) to nondimensionalize time has a characteristic value of a few days, for example. Thus, we can look for the steady response to prescribed sea surface temperature anomalies (represented in (6.80) and (6.81) by the saturation entropy of the sea surface) by setting the time derivatives in (6.75) – (6.81) to zero. To make the solutions more analytically tractable, we will focus on solutions with no surface drag or radiative forcing by setting D and Q to zero:

$$\frac{\partial s}{\partial x} = -yv, \tag{6.81}$$

$$\frac{\partial s}{\partial y} = yu, \tag{6.82}$$

$$w = -\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right), \tag{6.83}$$

$$\alpha u + G_m w + \chi s = \chi s_0. \tag{6.84}$$

This is a complete set, and (6.78) becomes a simply diagnostic relation for the mid-tropospheric moist static energy perturbation, h_m. Using cross-differentiation to eliminate all the variables in favor of the saturation entropy gives a single partial differential equation:

$$\tilde{\alpha} y \frac{\partial s}{\partial y} - \tilde{G} \frac{\partial s}{\partial x} + y^2 s = y^2 s_0, \tag{6.85}$$

where $\tilde{\alpha} \equiv \alpha / \chi$ and $\tilde{G} \equiv G_m / \chi$. The right side of (6.85) contains the sea surface saturation entropy distribution, which we will prescribe here. Remember that $\tilde{G}$ measures the gross moist stability, and $\tilde{\alpha}$ is a measure of the strength of the WISHE effect.

As an example, we represent the distribution of s_0 as the separable product of a function of y and a function of x, the latter of which we can take to be a discrete Fourier series, with the periodicity given by the circumference of the earth. Therrfore, we specify

$$s_0 = \sum_{n=1:\infty} \mathrm{Re}\left[F_n(y) e^{inx}\right]. \tag{6.86}$$

Here n is the zonal wavenumber; remember that we have normalized x by the radius of the Earth. Likewise, we represent the solution as

$$s = \sum_{n=1:\infty} \mathrm{Re}\left[J_n(y) e^{inx}\right], \tag{6.87}$$

where, from substituting (6.87) into (6.86) the functions J_n must satisfy

$$\tilde{\alpha} y \frac{dJ_n}{dy} + \left(y^2 - \tilde{G} i n\right) J_n = y^2 F_n(y). \tag{6.88}$$

Before proceeding, we note from the form of this equation that $J = 0$ at the equator, $y = 0$, thus there are no temperature perturbations on the equator. We also note that in the special case of no WISHE $(\alpha = 0)$, (6.88) has the simple solution

$$J_n = \frac{y^2 F_n(y)}{y^2 - \tilde{G} i n}. \tag{6.89}$$

Otherwise, using variation of parameters, the general solution of (6.88) that satisfies the boundary condition at $y = 0$ is

$$J_n = \frac{1}{\tilde{\alpha}} y^{in\tilde{G}/\tilde{\alpha}} e^{-y^2/2\tilde{\alpha}} \int_0^y F_n(u) u^{1-in\tilde{G}/\tilde{\alpha}} e^{u^2/2\tilde{\alpha}} du. \tag{6.90}$$

Whether this function is well behaved for large y will depend on the function $F(y)$. The full solution for s is then found by substituting J into (6.87). Given the solution for s, we can then find u, v, and w from (6.81), (6.82), and (6.84).

Here we evaluate the integral in (6.90) numerically for a sea surface saturation entropy perturbation with a single sinusoidal zonal wavenumber of 2 and a distribution in y given by

$$F_n = e^{-1.5y^4}.$$

This distribution is quite flat near the equator but has sharp shoulders at higher latitudes, and is symmetric about the equator. The top row of Figure 6.10 shows the fields for the case in which $\tilde{G} = 0.8$, $n = 2$, and $\alpha = 0$, for which the solution for s is given by (6.89). The vertical motion pattern in this case (left panel) strongly resembles the sea surface temperature distribution, but the temperature distribution peaks off the equator and more or less zonally in quadrature with the surface temperature. The wind distributions show cyclones and anticyclones around the warm and cold atmospheric temperature anomalies, respectively. It is important to note that the equatorial zonal wind is in quadrature with the sea surface temperature anomalies.

When the WISHE effect is added ($\alpha = 1$; second row in Figure 6.10), the whole pattern shifts eastward and extends to higher latitudes. In this case, the equatorial zonal winds have their maximum magnitudes near the peaks in the surface temperature distribution, with strong westerly surface winds over the warmest water. This is because the maximum surface enthalpy flux is shifted eastward toward the maximum surface easterly winds, where the net (mean plus perturbation) surface wind speed is maximum. With $\alpha = 1$, this has a strong effect on the surface enthalpy flux.

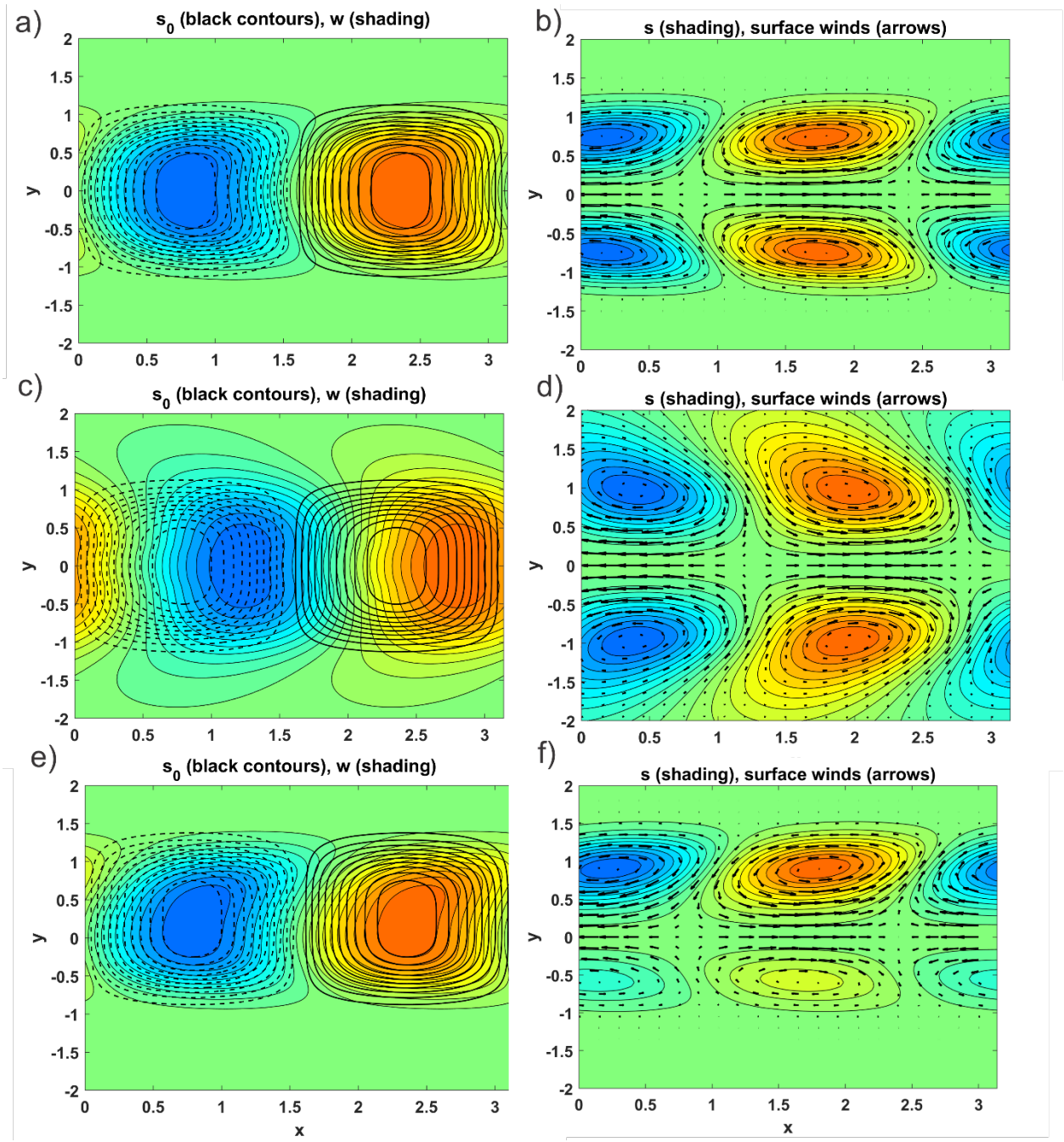


Figure 6.10: Steady, frictionless response of the equatorial atmosphere to sea surface saturation entropy anomalies as described in the text. In each row, the left-hand panel shows the specified distribution of s_0 (black contours, with dashed showing negative values) and vertical velocity w (shading; warm colors positive); the left-hand panel shows saturation entropy s (shading) and horizontal winds (arrows). For all plots, $\tilde{G}=0.8$ and $n=2$, and for the top and bottom rows, $\alpha=0$ while for the middle row $\alpha=1$. The third row is for conditions identical to the first except that the s_0 anomaly pattern has maximum amplitude at $y=0.25$.

The last row in Figure 6.10 is for conditions identical to those used to construct the first row, except that the sea surface saturation enthalpy pattern has been shifted northward so as to have maximum amplitude at $y = 0.25$. Even this relatively small northward displacement dramatically shifts the temperature and horizontal winds response into the northern hemisphere, even though the vertical motion pattern is hardly affected.

The reader is encouraged to explore these solutions using the code *Gill.m*, which can be found at https://zenodo.org/doi/10.5281/zenodo.10844820 .

One problem with this solution is that, when $D = 0$ and $\alpha > 0$, the basic state turns out to be absolutely unstable. For this reason, the solutions shown here must be regarded, at best, as the steady part of a time-evolving pattern. We will explore the linear instability of this system in detail in Chapter 7.

Gill (1980) found solutions for the system presented here for the case $\alpha = 0$ and for small values of D. He showed that the homogeneous solution can be expressed in terms of parabolic cylinder functions and found analytic solutions in the case where the forcing also has the form of the first or the second parabolic cylinder function. In both cases he sought solutions in which the forcing was confined to a limited range of x rather than being periodic in x as we have specified here.

To include surface drag, we will here solve the system (6.75) – (6.79), for the case $Q = 0$, by numerically marching forward in time until a steady state is attained. Given today's computational speeds, this is quite fast and allows one the flexibility of specifying any reasonable distribution of the sea surface temperature. But here we continue to take the sea surface temperature and the forced solutions to be sinusoidal in x, so as before we replace all the dependent variables by complex functions of y multiplied by e^{inx}. The code used to generate these solutions, *GillD.m*, is available at https://zenodo.org/doi/10.5281/zenodo.10845301 .

Results for the same function of y as we used in the zero drag case are shown in Figure 6.11, which is in the same format as Figure 6.10, but also includes contours of the normalized moist static energy perturbation, for a fairly strongly damped case with $n = 2$, $D = 0.2$, $\alpha = 0.5$, $G_m = 0.8$, $\chi = 1$, $\chi_2 = 0.5$, and $\delta = 10$. Surface drag strongly changes the solution. The vertical velocity, which was strongly shifted eastward from the sea surface entropy pattern in the zero drag case with WISHE, is shifted back to more nearly coincide with the sea surface entropy. Now there are large atmosphere temperature (saturation entropy) perturbations on the equator. One can also notice flow down the pressure gradient in the off-equator gyres. Recall that these are the sum of the baroclinic and frictional velocities; we could now solve the nondimensional equivalents of (6.71) and (6.72) for the frictional and baroclinic velocities separately. Note, however, that one or both of these would have a singularity at the equator which cancels in their sum.

Armed with knowledge of the basic response of the atmosphere to sea surface temperature anomalies, we are now in a position to tackle the coupled ENSO system.

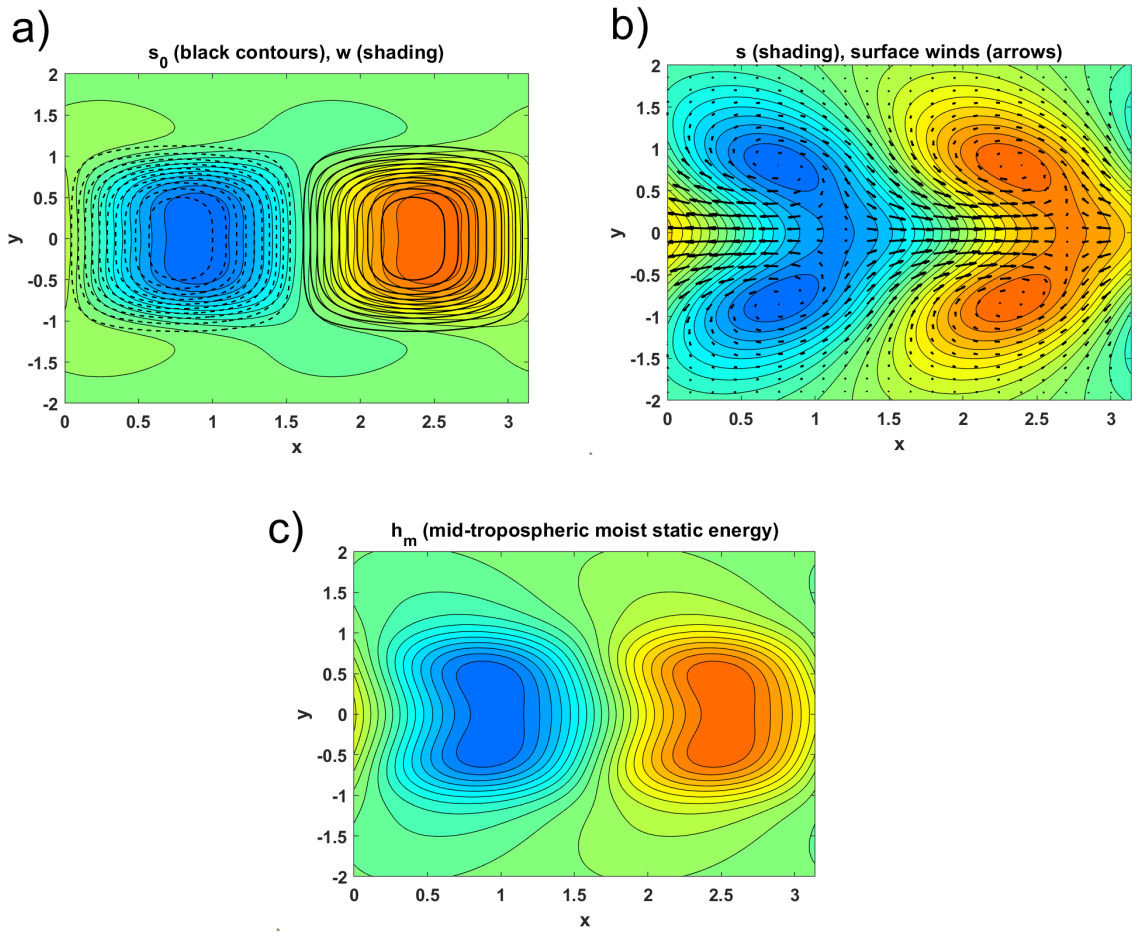


Figure 6.11: Panels a) and b) are the same as in Figure 6.10 but with surface drag included: n=2, $D = 0.2$, $\alpha = 0.5$, G_m=0.8, χ=1, χ_2=0.5, *and* δ=10. *Panel c) is for the same parameters but shows the mid-level moist static energy perturbation.*

6.3.3 Dynamics of the fully coupled ENSO system

ENSO involves the complex interplay of equatorial waves in the ocean and atmosphere. The real system relies in part on partial reflection of equatorial ocean waves off continental boundaries as well as leakage of the some of the wave energy into coastally trapped waves. (See Federov and Brown (2009) for an excellent review of equatorial wave dynamics.) Here we provide a mostly qualitative overview of the fundamental dynamics of ENSO. The reader is referred to the review paper by Wang et al. (2017) and the textbook by Webster (2020) for a more comprehensive overview.

At this writing, there are two views of ENSO dynamics. The first regards it as a fundamental instability of the Walker system resulting in growing and decaying oscillations. The second holds that the Walker system admits neutral and/or weakly damped modes that are stochastically excited by higher frequency disturbances of the equatorial region, such as the Madden-Julian Oscillation (see Chapter 7). In both cases, a mechanism first proposed by Bjerknes (1969) serves to amplify positive sea surface temperature perturbations in the central equatorial Pacific. With some small modifications, we describe this mechanism here.

Suppose a positive sea surface temperature anomaly is introduced into the central equatorial Pacific. Then, referring to either Figure 6.10 (middle row) or Figure 6.11, ascent will develop over the warm anomaly, but shifted eastward by the WISHE effect, given that the background surface flow in the Walker state is easterly. This will induce westerly surface wind anomalies to the west of the positive sea surface temperature anomaly, overlapping with the sea surface temperature anomaly if the WISHE effect is present. Note that this is fully consistent with the observations summarized in Figure 6.2. The relaxation of the surface easterlies over and to the west of the warm ocean anomaly leaves the ocean pressure gradient, associated with the sloping thermocline in the Walker state, unbalanced, and so water begins to accelerate toward the east, advecting the warm pool water with it. This will both amplify the original anomaly and propagate it eastward, more or less at the speed of ocean Kelvin waves. Note that this is more likely to start in April, when, as part of the normal seasonal cycle of the central Pacific, the equatorial easterlies are weakest. (This may help explain why observed El Niño events are phase locked to the seasonal cycle.)

Once the eastern equatorial Pacific has warmed up and zonal gradients across the Pacific are weak, it is less clear how an El Niño event dissipates. (Indeed, there is paleoclimatological evidence that “permanent” El Niño states existed at certain times in the past, such as the Pliocene period (3-5 million years ago), but the evidence is controversial.) We point out here that if the equatorial winds averaged across the Pacific become westerly, as they are often observed to during El Niño episodes, the WISHE effect changes sign and the westerlies relax near the location of peak anomalous sea surface temperature. This is illustrated in Figure 6.12 which shows solutions for the same conditions as in Figure 6.11 (*a* and *b*) except that we assume background westerly winds $(\alpha = -1.5)$. The ascent is now shifted decidedly west of the warm sea surface temperature anomaly and the winds have also shifted westward. Easterly wind anomalies east of the warm ocean anomaly should enhance ocean upwelling there, acting to restore cold conditions. Note also the cyclonic gyres north and south of the position of the warm ocean anomaly. These also exist when mean easterlies are present (Figure 6.11) but are more zonally aligned with the ocean temperature with mean westerlies.

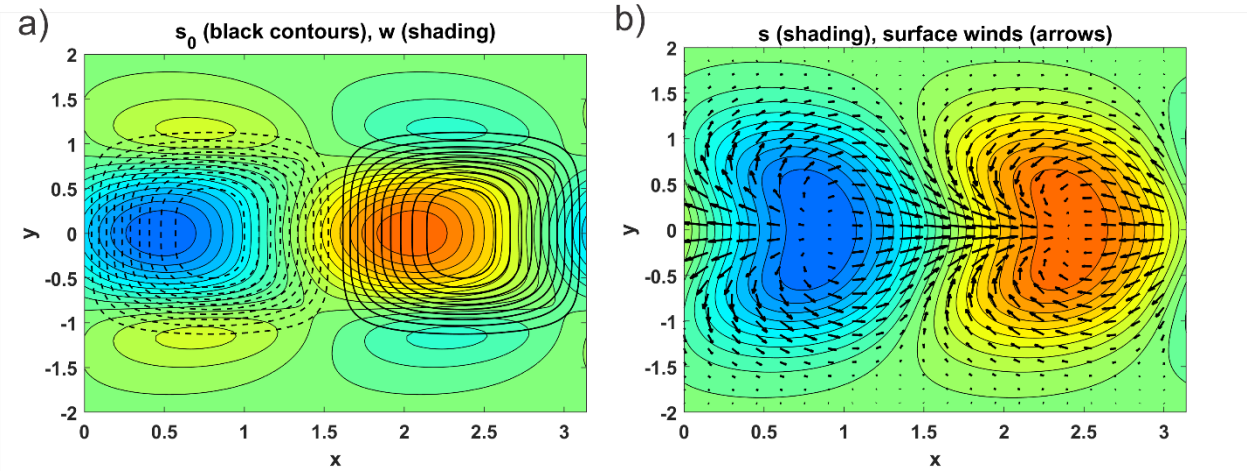


Figure 6.12: Same as Figure 6.11 a and b, except that $\alpha = -1.5$.

If we were to assume that the ENSO anomalies are slowly varying relative to the time scales of oceanic equatorial Kelvin and Rossby waves (not necessarily a very good assumption!) we could make certain deductions about the off-equatorial ocean currents from the steady forms of the shallow water equations, (6.5) and (6.6). Taking the derivative of (6.6) in x and subtracting from it the derivative of (6.5) in y gives

$$y\left(\frac{\partial u'}{\partial x}+\frac{\partial v'}{\partial y}\right)+v=\frac{1}{D\beta}\left(\frac{\partial \tau_y'}{\partial x}-\frac{\partial \tau_x'}{\partial y}\right), \tag{6.91}$$

where the velocities pertain to the ocean and D is here the thermocline depth. Recall that τ_x' and τ_y' are the meridional and zonal components of the surface wind stress per unit mass. But the steady form of (6.7) requires the divergence of the velocities to be zero, so the first term in (6.91) vanishes and we are left with

$$v=\frac{1}{D\beta}\left(\frac{\partial \tau_y'}{\partial x}-\frac{\partial \tau_x'}{\partial y}\right). \tag{6.92}$$

This is known as *Sverdrup balance*, named after its discoverer, the Norwegian oceanographer Harald Sverdrup. Thus, where the atmospheric surface winds are cyclonic, the steady response of the ocean is poleward flow, and where they are anticyclonic, the flow is equatorward. Applying this to Figure 6.12b, we have water diverging away from the equator at the location of the warm sea surface temperature anomaly (black contours in Figure 6.12a), implying equatorial upwelling, which would damp down the anomaly. These physics are a key part of the "Recharge Oscillator" theory of ENSO, formulated by Wyrtki (1975) and Wyrtki (1985) and modified by Jin (1996).

Another body of theory, developed by Suarez and Schopf (1988) and Battisti and Hirst (1989), postulates that the development of the eastern Pacific warm anomaly during El Niño excites a westward-propagating oceanic Rossby wave that partially reflects of the western boundary, returning eastward as a Kelvin wave that lifts the thermocline in the eastern Pacific, reversing

the warm anomaly. This is known as the “delayed oscillator model” of ENSO. A variant on this idea, also involving partial reflection of eastward-propagating Kelvin waves off the eastern boundary as westward propagating Rossby waves, was proposed by Picaut et al. (1997).

The mechanisms described above are by no means mutually exclusive, and the physics behind them are contained in reduced ENSO models, such as the model developed by Cane and Zebiak (1985), which has proven modestly successful in forecasting El Niño events (Cane et al., 1986). It would probably be premature to claim that we have a reasonably complete understanding of ENSO physics at the time of this writing, and further progress is likely to depend on better understanding of tropical atmospheric physics and dynamics, ocean dynamics, and air-sea coupling.

References

Battisti, D. S., and A. C. Hirst, 1989: Interannual variability in a tropical atmosphere–ocean model: Influence of the basic state, ocean geometry and nonlinearity. *Journal of Atmospheric Sciences*, **46,** 1687-1712, doi:10.1175/1520-0469(1989)046<1687:Iviata>2.0.Co;2.

Bjerknes, J., 1969: Atmospheric teleconnections from the equatorial pacific. *Mon. Wea. Rev.*, **97,** 163-172, doi:10.1175/1520-0493(1969)097<0163:Atftep>2.3.Co;2.

Cane, M. A., and S. E. Zebiak, 1985: A theory for El Niño and the Southern Oscillation. Science, 228, 1085–1087, https://doi.org/10.1126/science.228.4703.1085.

Cane, M. A., S. E. Zebiak, and S. C. Dolan, 1986: Experimental forecasts of El Niño. Nature, 321, 827–832, https://doi.org/10.1038/321827a0.

Federov, A. V., and J. N. Brown, 2009: Equatorial waves. *Encyclopedia of ocean sciences*, J. Steele, Ed., Academic, San Diego 3679=3695 pp.

Gill, A. E., 1980: Some simple solutions for heat-induced tropical circulation. *Quart. J. Roy. Meteor. Soc.*, **106,** 447-462, doi:https://doi.org/10.1002/qj.49710644905.

Jin, F.-F., 1996: Tropical ocean-atmosphere interaction, the Pacific cold tongue, and the El Niño-southern oscillation. Science, 274, 76–78, https://doi.org/10.1126/science.274.5284.76.

Matsuno, T., 1966: Quasi-geostrophic motions in the equatorial area. *J. Meteor. Soc. Japan*, **44,** 25-42.

Picaut, J., F. Masia, and Y. du Penhoat, 1997: An advective-reflective conceptual model for the oscillatory nature of the ENSO. Science, 277, 663–666, https://doi.org/10.1126/science.277.5326.663.

Suarez, M. J., and P. S. Schopf, 1988: A delayed action oscillator for ENSO. *Journal of Atmospheric Sciences*, **45,** 3283-3287, doi:10.1175/1520-0469(1988)045<3283:Adaofe>2.0.Co;2.

Takahashi, K., A. Montecinos, K. Goubanova, and B. Dewitte, 2011: ENSO regimes: Reinterpreting the canonical and modoki El Niño. *Geophys. Res. Lett.*, **38**, doi:https://doi.org/10.1029/2011GL047364.

Thomson, W., 1880: On gravitational oscillations of rotating water. *Proceedings of the Royal Society of Edinburgh*, **10,** 92-100, doi:10.1017/S0370164600043467.

Wang, C., C. Deser, J.-Y. Yu, P. DiNezio, and A. Clement, 2017: El Niño and Southern Oscillation (ENSO): A review. *Coral reefs of the eastern tropical Pacific: Persistence and loss in a dynamic environment*, P. W. Glynn, D. P. Manzello, and I. C. Enochs, Eds., Springer Netherlands, Dordrecht 85-106 pp.

Webster, P. J., 2020: *Dynamics of the tropical atmosphere and oceans.* John Wiley and Sons, Oxford, U.K., 501 pp.

Yanai, M., and T. Maruyama, 1966: A stratosphere wave propagating over the equatorial Pacific ocean. *J. Meteor. Soc. Jap.*, **44,** 291-294.

7 Tropical Weather on Time Scales of Days to Months

7.1 Introduction

Compared to the weather in middle and high latitudes, tropical weather on subseasonal time scales is subtler and has so far proven less predictable. We have already seen, in Chapter 5, that the Indian monsoon exhibits striking variability on times scales on the order of 10 days, with active and break cycles associated with profound changes in rainfall and wind. In this chapter, we will explore observations of tropical weather variability on time scales of days to months and review what is currently known about the mechanisms involved. We begin with an exploration of weather associated with a rich spectrum of equatorially trapped disturbances and end with an exposition of African easterly waves, which are sometimes associated with North Atlantic tropical cyclones.

7.2 Disturbances of the equatorial waveguide

In much of the equatorial region, rain showers can occur at any time of day and year. Nevertheless, the probabilities of rain are modulated by a fascinating variety of disturbances with time scales ranging from a few days to a few months. While much of this modulation can be characterized as random noise with a red spectrum in space and time, some of it appears to be an excitation of equatorial waves of the kind we discussed in connection with ENSO ocean dynamics, and yet some of it, while clearly organized, is not part of the Matsuno spectrum of neutral linear modes.

Figure 7.1 displays two time-longitude sections (often referred to as Hovmöller[1] plots) of satellite-derived precipitation and outgoing longwave radiation for the first five months of 2005. The horizontal axis of these plots spans the equatorial Pacific. In the precipitation plot (left panel), the eye naturally picks out crisscrossing diagonal lines, and the phases speeds corresponding to two of these sets of lines are indicated. There is a prominent set of disturbances traveling east at about 12 ms^{-1}, and another traveling west at around 4 ms^{-1}. The latter may simply be disturbances drifting west with the trade winds. But other sets are discernible, including disturbances propagating westward much faster. The sharp transition at around 97^{o} longitude marks the western boundary of the chain of islands that constitute Indonesia, showing the strong effect of land-sea contrasts on convective precipitation.

Curiously, the OLR plot (right panel of Figure 7.1), though it is for exactly the same period of time and span of longitude, shows somewhat different patterns, including a broad envelope (marked by an ellipse in the figure) traveling east, against the trade winds, at about 4 ms^{-1}. Some of the faster eastward moving disturbances we saw in the precipitation fields are also evident in the OLR.

[1] Named after the Danish meteorologist Ernest Hovmöller, who introduced them in 1949

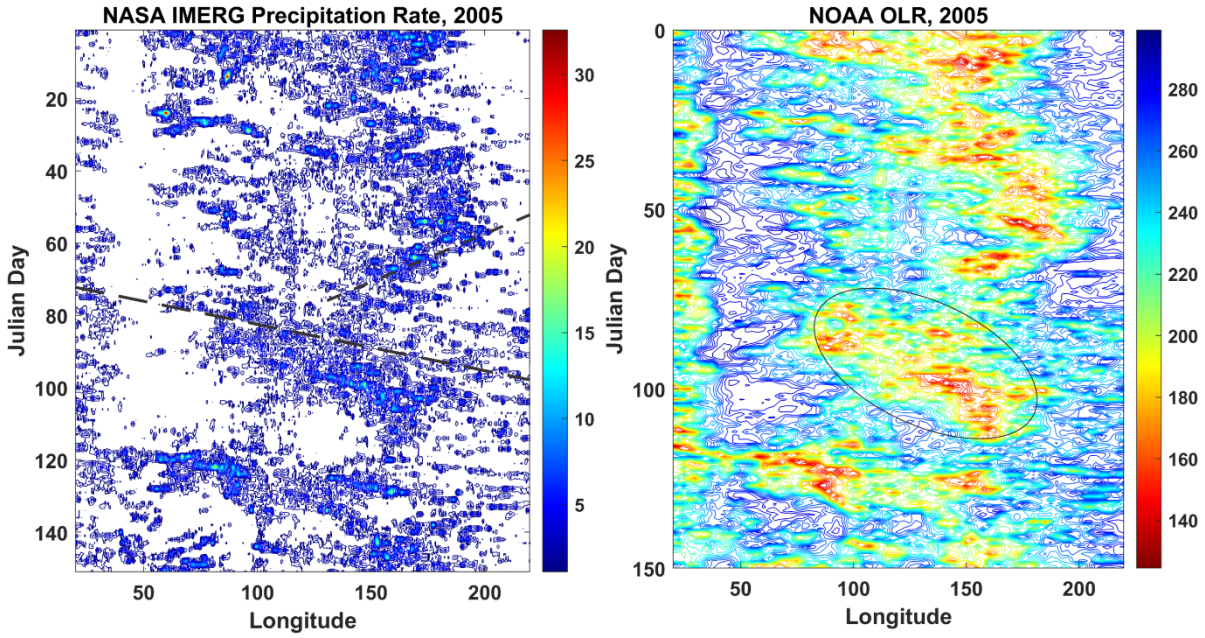


Figure 7.1: Hovmöller diagrams spanning the period January 1 – May 31, 2005 (vertical axis) and longitudes from 20°E to 220°E (horizontal axis). Left: Precipitation with longitudinal resolution of 0.1 degree; right: Outgoing Longwave Radiation (OLR) with horizontal resolution of 2.5 degrees. Both data sets have a temporal resolution of 1 day. In the left panel, the upper and lower dashed lines have slopes of -5 ms^{-1} and 10 ms^{-1}, respectively, and in the right panel the ellipse has a slope of 4 ms^{-1}

An observer on the equator at 50° east longitude would experience some rain nearly every day, but greatly enhanced rain every third or fourth day associated with the fast, eastward-propagating disturbances. Note also that there is much more rain after about day 70 (mid-March) than before.

The variability evident in Hovmöller plots such as those shown in Figure 7.1 can be further quantified by spectral decomposition of the fields, as pioneered by Hayashi (1982) and further refined by Wheeler and Kiladis (1999, 2020). The first step in their procedure is to take a long record of observations, typically satellite-derived OLR or brightness temperature, and remove the first three annual harmonics at each grid point. They then break the record into segments of 96 days each, with this time scale being approximately the largest period oscillation they expect to find in the de-seasonalized series. Then, for each segment, the time series is detrended and the field ϕ is divided into equatorially symmetric and anti-symmetric parts:

$$\phi_{sym} = \frac{1}{2}\left(\phi(\theta) + \phi(-\theta)\right),$$
$$\phi_{antisym} = \frac{1}{2}\left(\phi(\theta) - \phi(-\theta)\right), \tag{7.1}$$

where θ is the latitude. For each component, a fast Fourier transform is applied in the time and longitude dimensions at each latitude, and the resulting spectral power is summed over a band of latitude centered at the equator and averaged over all the 96-day segments of the record.

Figure 7.2 shows the base-10 logarithm of the average of the symmetric and anti-symmetric spectral power (which, from (7.1), is just half the total spectral power) of the Cloud Archive User Data Service (CLAUS) record of brightness temperature (10.5-12.5-micron radiances) from a collection of operational meteorological satellites, from 1984 to 2008. These have been averaged from 15°S to 15°N latitude. Negative zonal wavenumber denotes westward propagation.

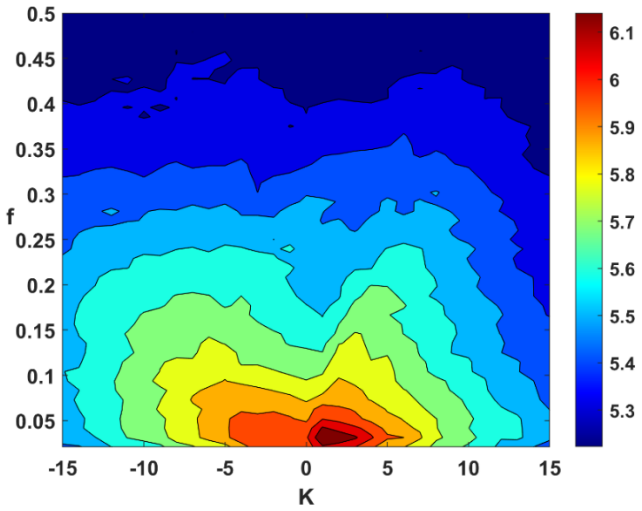


Figure 7.2: Base 10 logarithm of the spectral power of 10.5-12.5-micron radiances from CLAUS data over the period 1984-2008. This is the mean of the symmetric and anti-symmetric components, which is in turn ½ of the total power. The power has been summed over all latitudes between 20°S and 20°N in the data set. The frequency, on the y-axis, is in cycles per day.

The concentration of spectral power at low frequencies and wavenumbers is evident, as are significant departures from a simple red noise spectrum, some of which seem to loosely follow the Matsuno dispersion curves described in Chapter 6. Note in particular the concentration of very low frequency spectral energy in low zonal wavenumber eastward-propagating disturbances.

We might hypothesize that this power spectrum in wavenumber-frequency space is the sum of some kind of red noise that is random in space and time, and regular oscillations propagating along the equator. Wheeler and Kiladis (1999) estimated the background noise spectrum by smoothing the raw total power spectrum (that includes both symmetric and anti-symmetric components) by applying several times in succession a simple 1-2-1 smoother in wavenumber and frequency. Here we depart from that method and define the background noise as that part of the total power spectrum that projects roughly equally onto symmetric and anti-symmetric variability; that is, we will regard the background as ½ the total power spectral density, as illustrated in Figure 7.2. Then the spectral power representing the symmetric and antisymmetric variability is obtained by subtracting this background from, respectively, the symmetric and anti-symmetric power spectra, and setting to zero any negative values that result. The result of applying this procedure to the CLAUS data is illustrated in Figure 7.3.

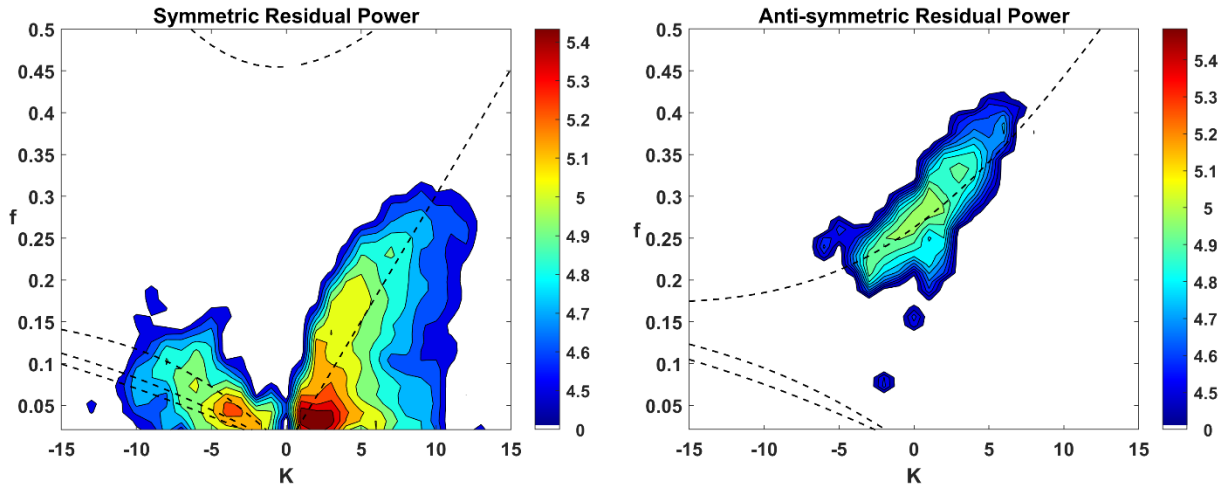


Figure 7.3: Base 10 logarithm of the residual power of the symmetric (left) and anti-symmetric (right) components of the CLAUS brightness temperature record. Fields have been smoothed by applying a 1-2-1 filter 7 times in wavenumber only. Contours are omitted for values less than 4.4. The dashed lines show the Matsuno dispersion curves for n= -1 to +5 assuming a shallow water phase speed of 16 ms^{-1}, with all curves Doppler shifted assuming a background easterly wind of 2 ms^{-1}. The frequency (f) is in cycles per day.

To deal with the strict quantization of the data in wavenumber and frequency, we have applied a 1-2-1 filter seven times, in wavenumber only. Figure 7.3 also shows the Matsuno dispersion relations, taking $c = 16\ ms^{-1}$ and Doppler shifting the curves assuming a background easterly wind of $2\ ms^{-1}$.

The most prominent signal in the symmetric part of the residual spectrum is the MJO, seen here as a spectral peak centered at the (quantized) frequency of one cycle per 48 days and extending from eastward zonal wavenumber 1 to 3[2]. The residual spectral power also extends along Kelvin wave dispersion lines. There is also quite a bit of energy along the n=1,3, and 5 equatorial Rossby wave curves, especially around westward zonal wavenumber 4 and (quantized) frequency of once in 24 days.

The only prominent signal in the anti-symmetric residual power spectrum corresponds to the dispersion curve of an n = 0 mixed Rossby-gravity/eastward inertio-gravity mode, with most of the activity in zonal wavenumber between -4 and 8.

Given the signals evident in the left panels of Figure 7.3, one can begin to examine the particular characteristics of variability in different sectors of wavenumber-frequency space by filtering within those sectors and regressing other fields, such as winds and temperatures from reanalyses against the filtered fields. An example of this is presented in Figure 7.4, which starts with a similar wavenumber-frequency decomposition of the same brightness temperature data as used to construct Figure 7.3 except that the background has been defined by smoothing the total power in wavenumber-frequency space, following Wheeler and Kiladis (1999). We use this figure to define sectors corresponding to the most prominent modes of variability and proceed to examine the characteristics of equatorial variability in many of these sectors.

[2] Take care to remember that the zonal wavenumber is quantized in integers and the frequency in integer multiples of once in 96 days. The applied smoothing and contouring routine smooth out this quantization in the graphs.

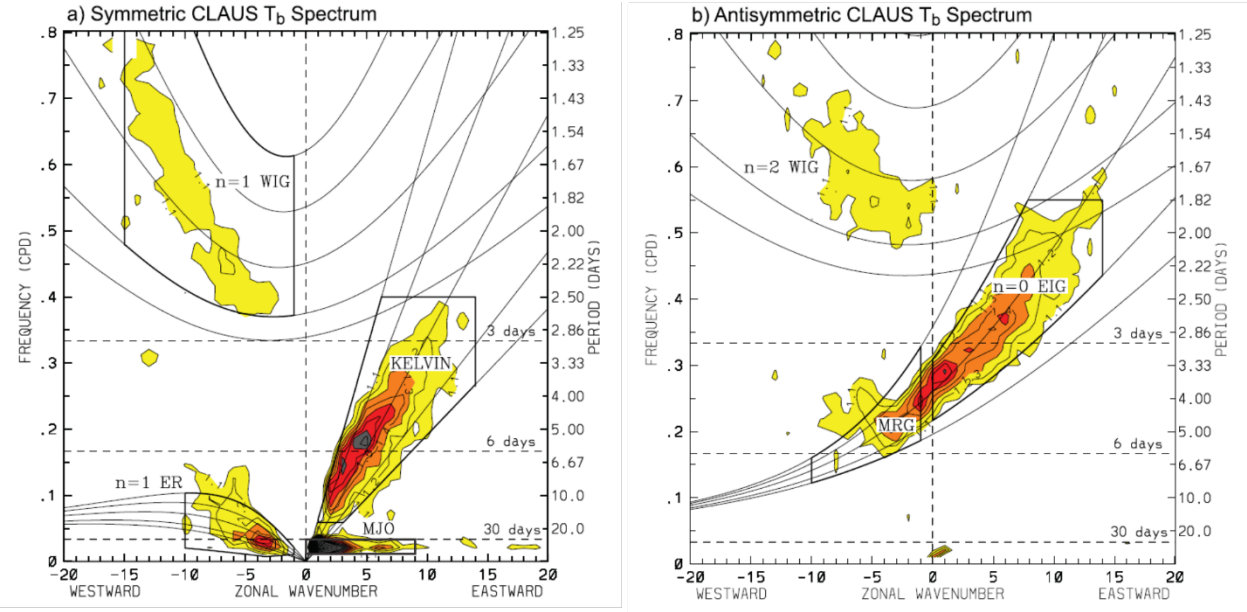


Figure 7.4: Similar to Figure 7.3 (right panels) but showing sectors used in wavenumber-frequency filtering. For this plot, the background spectrum has been created by smoothing the total (symmetric plus anti-symmetric) power, as in Wheeler and Kiladis (1999).

7.2.1: The Madden-Julian Oscillation

In terms of spectral power of many variables, including outgoing longwave radiation, the Madden-Julian Oscillation ("MJO") is the most prominent mode of subseasonal variability in the tropics. It was first discovered by Roland Madden and Paul Julian (Madden and Julian 1971, 1972) in their examination of zonal winds and surface pressure in time series of near-equatorial rawinsonde data. The variability consists of disturbances dominated by wavenumber 1 traveling eastward, with peak amplitudes near the equator. In fields related to deep convection, such as OLR and precipitation, the MJO is seen to develop in the western equatorial Indian Ocean and move eastward, reaching maximum amplitude near or just east of Indonesia, and then decaying eastward, losing coherence near the dateline. However, the MJO can often be traced around the globe in near-tropopause variables such as zonal wind and velocity potential. A sketch of the structure of the MJO is shown in Figure 7.5.

The composite structure of an MJO event does not resemble that of any of the classical Matsuno modes, nor does its frequency dependence on wavenumber fall along any of the Matsuno dispersion curves (see Figure 7.4). Figure 7.6 shows a composite of the MJO structure made by regressing MJO frequency- and wavenumber-filtered fields over the period 1979-1989 onto OLR at 0°N and 125°E. The data were first projected onto spherical harmonics and truncated at "extended" rhomboidal 8, whereby 16 meridional modes are retained for zonal wavenumbers 0- 8. The fields are OLR itself, 200 hPa winds from operational analyses, and satellite microwave sounding unit channel 2 temperature (MSUT), which has a peak sensitivity at mid-tropospheric levels. The regressions are performed using fields lagged with the respect to the OLR at time 0.

A similar regression analysis is shown in Figure 7.7, but also includes fields at 850 hPa.

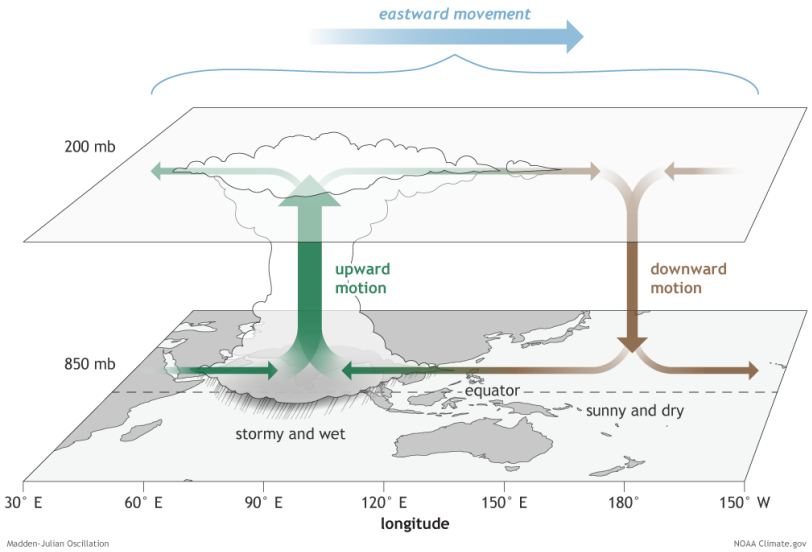


Figure 7.5: Sketch of the vertical circulation, zonal winds, and convection associated with the Madden-Julian Oscillation, at a time when the maximum ascent and rainfall is over the Indian Ocean. The horizontal winds shown are perturbations from the long-term mean. The MJO moves eastward with a period of 30-60 days.

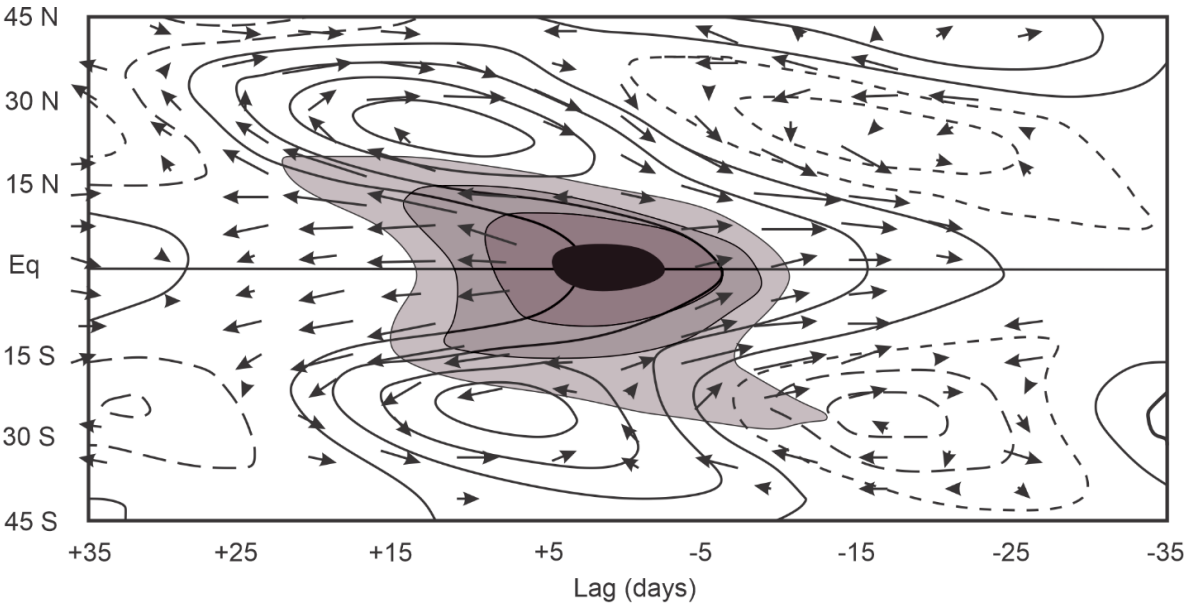


Figure 7.6: Various fields bandpass filtered for eastward-propagating zonal wavenumber 3 and periods of 35-95 days and linearly regressed onto OLR at 0ºN and 125ºE with the time lags indicated on the abscissa: OLR (shading), 200 hPa winds from operational analyses (arrows), and satellite microwave sounding unit channel 2 temperature (MSUT) (black solid and dashed contours). MSUT which has a peak sensitivity at mid-tropospheric levels. The maximum wind speed is $2.7\ ms^{-1}$ *and the temperature contour interval is 0.05 K. Data from 1979-1989. The data have been subject to some smoothing before compositing (see text for more detail).*

The pattern in Figures 7.6 and 7.7 more nearly resembles the Gill pattern shown in Figure 6.11 than it does any of the Matsuno modes, as though it were a response to an imposed sea surface temperature anomaly. The MJO ascent region, more or less coinciding with low values of OLR, is preceded by surface easterlies and upper tropospheric westerlies, an enhancement

of the background Walker circulation. Cyclonic gyres are found behind (west of) and poleward of the OLR minimum at low levels, with corresponding anticyclones in the upper troposphere. At upper levels there are also prominent cyclonic gyres straddling the equator to the east of the OLR minimum. In Figure 7.6 one can see a kind of chevron-shaped middle tropospheric temperature pattern, with a maximum along the equator just east of the OLR minimum and maxima in the anticyclonic gyres poleward and westward of the OLR minimum.

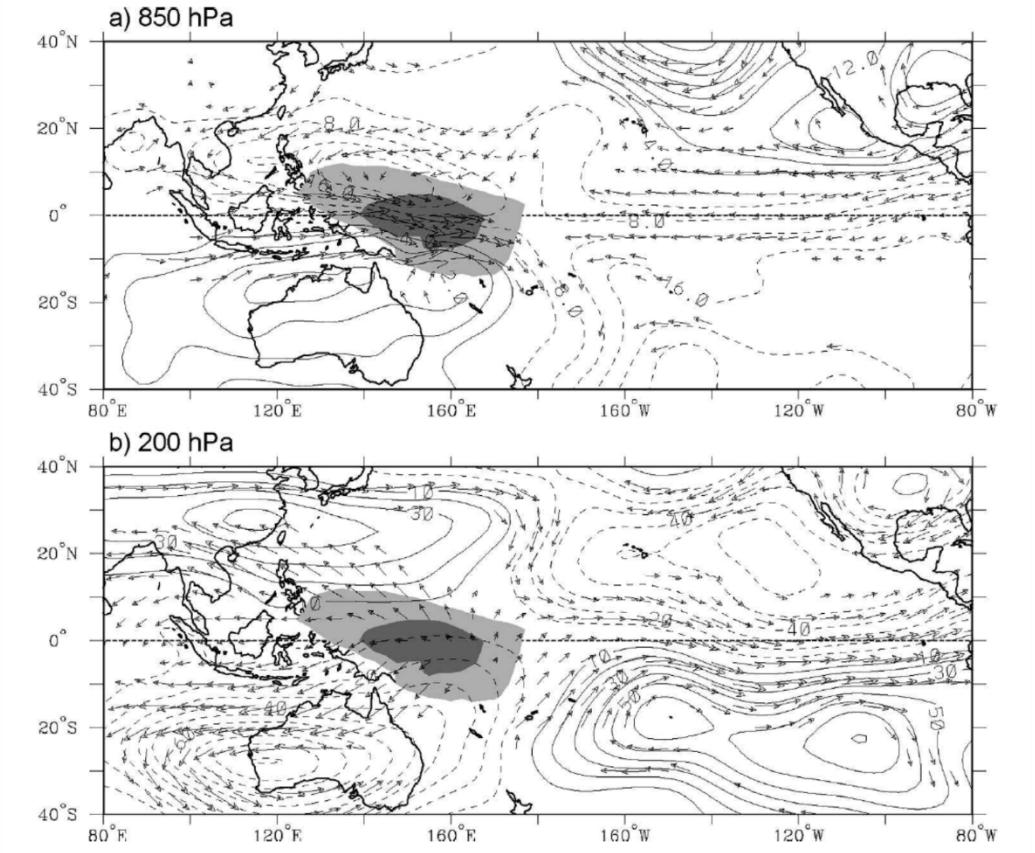


Figure 7.7: Anomalous outgoing longwave radiation (OLR; shading) and anomalous winds from the European Center for Medium Range Weather Forecasts 15-year reanalysis (ERA-15, arrows) regressed against MJO-filtered OLR at the equator and 155°E at 850 hPa (top) and 200 hPa (bottom). Contours show streamfunction. Largest wind in top panel is about 2 ms^{-1} and about 5 ms^{-1} in bottom panel.

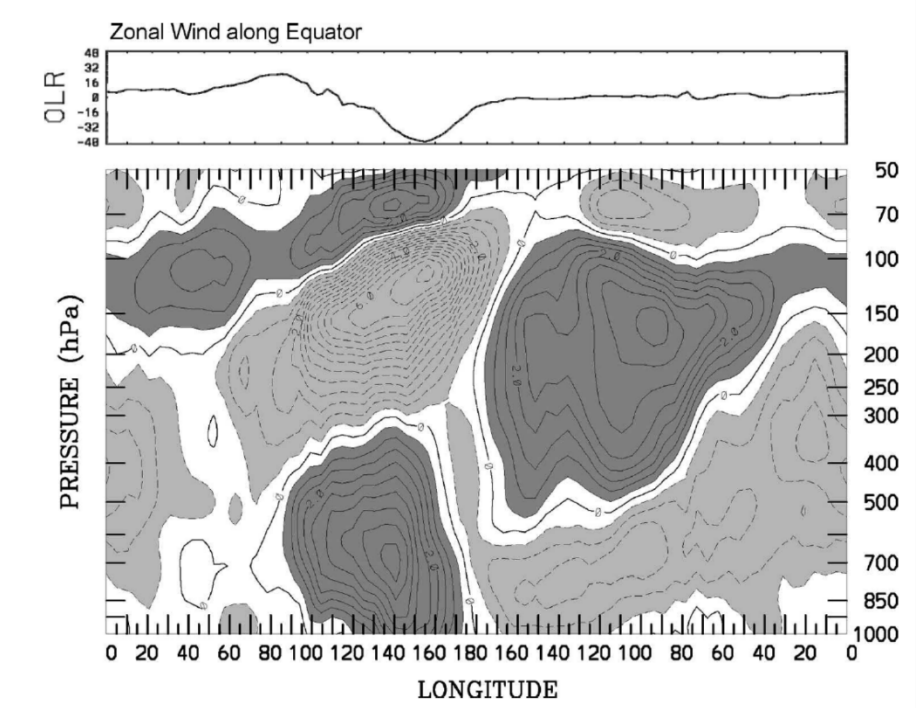


Figure 7.8: As in Figure 7.7 but showing the vertical structure of zonal winds regressed against MJO-filtered OLR at the equator and 155°E. Contour interval is 0.5 ms^{-1} and the associated OLR along the equator is shown in the top panel.

The vertical structure of zonal winds associated with the MJO is evident in Figure 7.8. To a first approximation, winds in the troposphere show first mode baroclinic structure, with opposite signs in the lower and upper troposphere. There is some indication also of a lower stratospheric response of the opposite sign as the upper troposphere.

The MJO has a strong annual cycle, at least as measured by outgoing longwave radiation. Figure 7.9 shows the annual cycle of the OLR power from a global 11-year daily record averaged over eastward propagating zonal wavenumbers 1-3 and periods of 35-95 days on individual days of the year. The power peaks in late boreal winter to early boreal spring in the southern hemisphere, and shows a minimum in mid boreal summer in the northern hemisphere. Note that the globally averaged ITCZ latitude varies in a similar way.

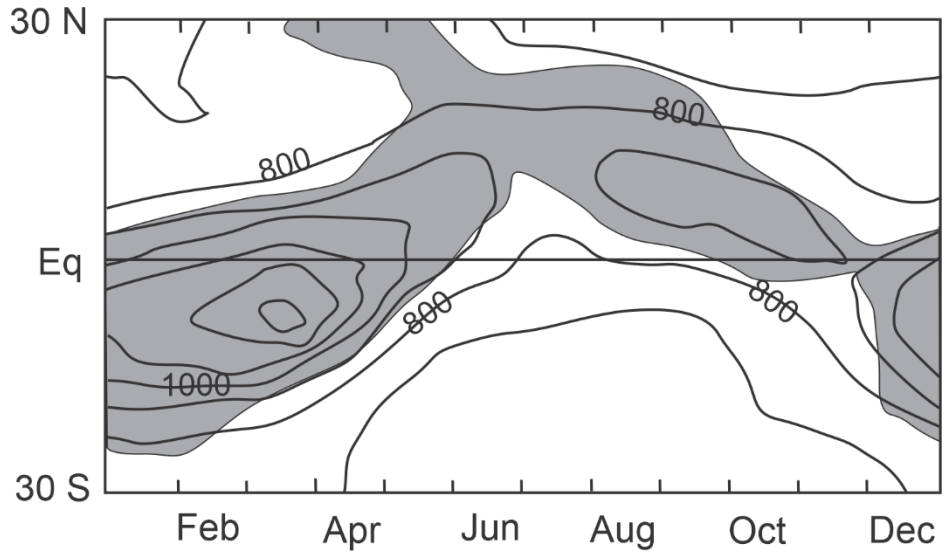


Figure 7.9: Outgoing longwave radiation spectral power averaged over eastward-propagating wavenumbers 1-3 and periods of 35-95 days on individual days of the year (open contours) from a global 11-year daily record. Regions with statistical significance greater than 95% are shaded.

7.2.2: Convectively coupled equatorial waves

Some idea of the structure of the non-MJO equatorial disturbances whose variance seems to follow the classical Matsuno dispersion curves (figure 7.4) can be gleaned by regressing radiosonde- and reanalysis-derived variables against satellite brightness temperature filtered in the wavenumber-frequency domain according to the regions outlined in black in Figure 7.4 (Kiladis et al. 2009). In particular, the temporal progression of atmospheric variables as a function of pressure can be obtained by regressing against the satellite data with a time lag.

Figure 7.10 and 7.11 show the vertical structure of wind, temperature, and specific humidity for the Kelvin wave, westward mixed Rossby-gravity wave (MRG), eastward mixed Rossby-gravity wave (EIG), and n=1 equatorial Rossby wave, respectively, using rawinsonde data at the island of Majuro (7.1^{o}N, 171.4^{o}E) regressed against satellite brightness data with the same temporal, resolution at the grid point closest to Majuro.

A striking feature of all but the equatorial Rossby wave is the vertical tilt of the wave in the direction opposite to its propagation; e.g. the Kelvin wave tilts westward with height. The relationships between the 850 hPa and 150 hPa zonal wind anomalies are close to what one would expect if the first baroclinic mode were dominant, but the existence of significant anomalies in the middle troposphere is not consistent with first baroclinic mode structure. We will return to this issue when we discuss theoretical explanation for these wave modes, but in short, many theories now incorporate the barotropic mode together with the first baroclinic mode, and/or allow for non-moist adiabatic behavior such as the second baroclinic mode. Either of the barotropic or the second baroclinic modes, in tandem with the first baroclinic mode, can potentially explain the wave tilt seen in Figures 7.10 and 7.11.

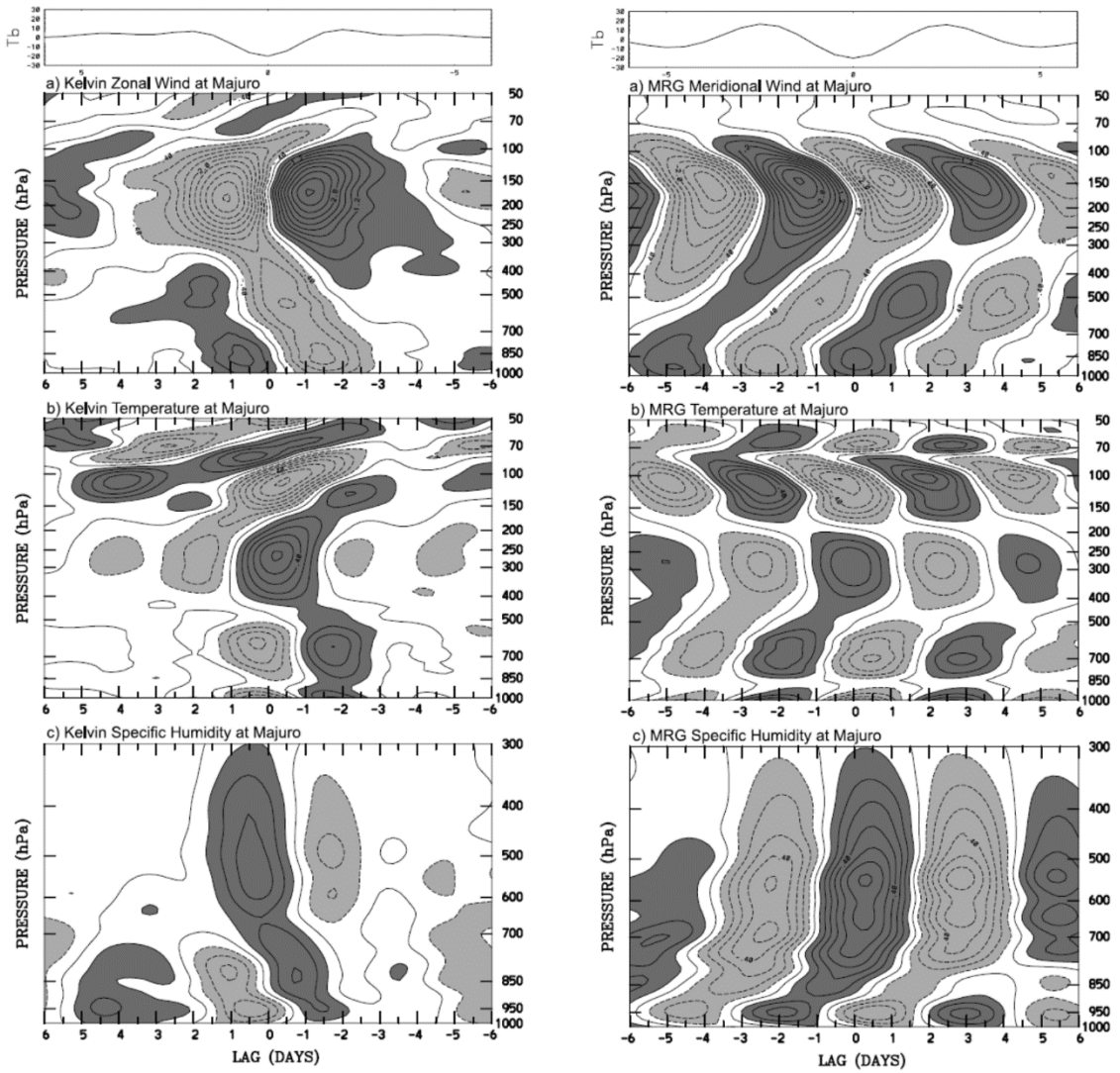


Figure 7.10: Time-pressure sections of anomalous zonal wind (top), temperature (middle), and specific humidity (bottom) from rawinsonde data at Majuro regressed against satellite brightness data filtered in the wavenumber-frequency domains shown in Figure 7.4 for the Kelvin wave (left) and the westward mixed Rossby-gravity wave (MRG; right). Negative lag indicates rawinsonde fields leading the satellite brightness temperature. All fields have been scaled to a brightness temperature anomaly of -20 K. Contour interval is 0.4 m s^{-1} for wind, 0.1 K for temperature, and 0.1 g kg^{-1} for specific humidity, with negative contours dashed and dark shading representing positive perturbations. The corresponding brightness temperature is shown at top, in K.

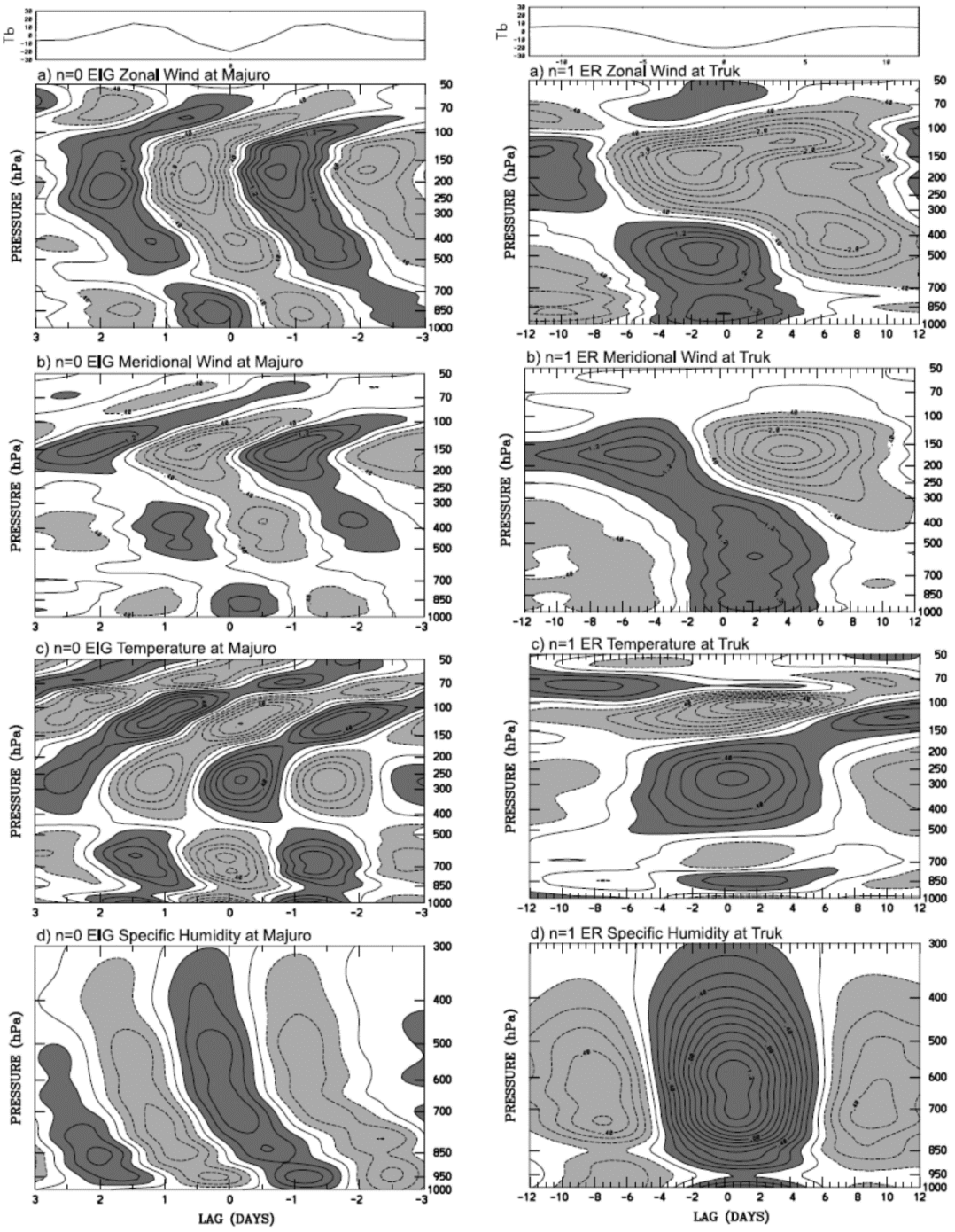


Figure 7.11: Same as Figure 7.10 except showing the eastward mixed Rossby-gravity wave (EIG; left) and the n=1 equatorial Rossby wave (ER; right), except that the second row shows the meridional wind anomaly (same contour interval as used for the zonal winds).

Consistent with the possible presence of a second baroclinic mode, the vertical structures of temperature anomalies for all but the equatorial Rossby wave appear to have two maxima within the troposphere as well as one or more maxima in the lower stratosphere, though the amplitude of the temperature anomalies is not very large compared to expected rawinsonde sampling and instrumental error, so one should exercise some caution in interpreting them.

The vertical structure of the equatorial Rossby mode (right column of Figure 7.11) is much more consistent with dominance by the first baroclinic mode. At first blush, it looks like the temperature structure may have higher order mode contributions, but it must be remembered that the dynamically essential variable here is virtual temperature, not actual temperature. In round numbers, 5 contours of specific humidity are worth one contour of temperature, in terms of its effect on virtual temperature. Consequently, the single negative contour of temperature near 700 hPa (right side of Figure 7.11, third row, near a time lag of 0-2 days) is easily outweighed by the ~10 contours of specific humidity at the same time lag and pressure in terms of its effect on virtual temperature. Thus, it is likely that, consistent with the dominance of the first baroclinic mode in zonal winds, there is a single maximum in the vertical structure of virtual temperature anomalies.

Close inspection of Figures 7.10 and 7.11 show that in the lower stratosphere (above 150 hPa), for all but the equatorial Rossby mode, the phase lines are tilted in the direction of wave propagation. This indicates upward wave group velocity in the lower stratosphere, consistent with upward radiation of wave energy. The observations therefore indicate that approximating the tropopause as a rigid lid may not work well for these faster wave modes. In the case of the equatorial Rossby mode, however, the signal is muted and ambiguous above the tropopause, so these low frequency modes may not be losing much energy to the stratosphere.

7.2.3: Equatorial subseasonal variability in a convection-permitting global model

How well do current global climate models simulate subseasonal tropical variability? Most climate models have horizontal resolutions insufficient to permit explicit simulation of moist convection, and therefore have to represent moist (and dry) convection using parameterizations. A thorough analysis of two most recent generations (5 and 6) of models that participated in the Climate Model Intercomparison Project (CMIP)[3] show that their simulation of convectively coupled waves is highly inconsistent among models, and only one of the many different models was able to simulate an eastward-propagating Madden-Julian Oscillation (Hung et al. 2013; Bartana et al. 2023). It is widely believed that deficiencies in cumulus parameterizations are responsible for the relatively poor performance of climate models in simulating this kind of variability.

In recent years, advances in numerical modeling and in computing allow one, for the first time, to simulate global climate on the time scale of several years with horizontal resolution that permits explicit simulation of moist convection, although the convective clouds are poorly resolved. These models simulate tropical variability at many scales and reproduce much of the observed subseasonal variability near the equator. Here we analyze the output of one such model, an aquaplanet global version of the System for Atmospheric Modeling (SAM,

[3] https://www.wcrp-climate.org/wgcm-cmip

Khairoutdinov and Randall 2003), with equatorial horizontal grid spacing of about 20 km and a zonally invariant sea surface temperature distribution given by

$$SST = 273 + 27\left(1 - \sin^6(\theta)\right),$$

where θ is the latitude in radians and SST is in kelvins. This distribution is displayed in Figure 7.12.

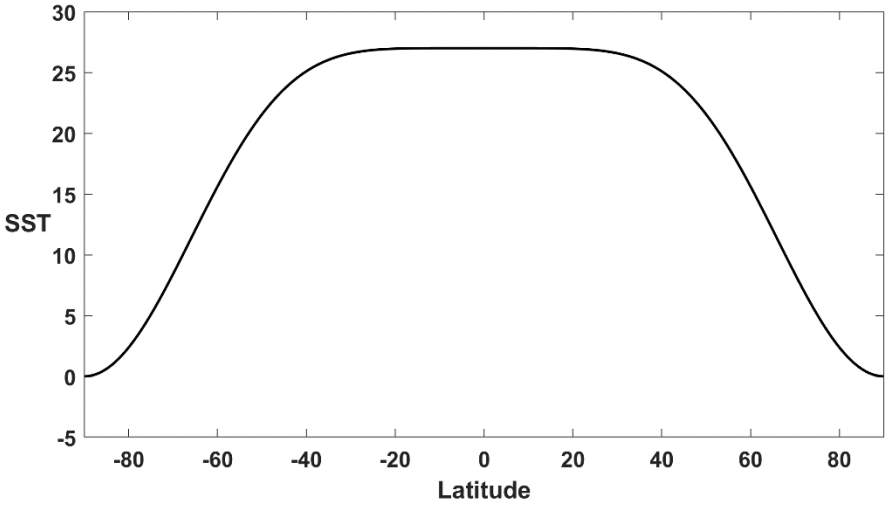


Figure 7.12: Distribution of sea surface temperature used in the convection-permitting simulations on an aquaplanet.

The tropical SST is much flatter than observed, and does not vary in time. The intent is to minimize the influence on the tropical atmosphere of baroclinic wave activity from higher latitudes. Experience with convection-permitting aquaplanet models shows that if SST has no meridional gradients at all, the climate becomes dominated by tropical cyclones at high latitudes. The SST distribution shown in Figure 7.12 is a compromise designed to minimize the influence of baroclinic instability and high latitude tropical cyclones.

This global version of SAM has an insolation of $400\,Wm^{-2}$, constant in space and time, and has state-of-the-art representations of subgrid-scale turbulence and cloud microphysics, including the ice phase. The model is run for 360 days and we analyses the last 300 days of the simulation.

Figure 7.13 shows a snapshot of outgoing longwave radiation and 850 hPa winds at day 170 of the simulation. Gentle mean easterlies are present in the tropics, though clearly modulated at zonal wavenumber 1. Tropical cyclone-like disturbances are evident at around 30^o latitude in both hemispheres; reference to Figure 7.12 shows that these are over very warm water but near the shoulders of the SST distribution, where the Coriolis parameter is relatively large. (Chavas and Reed (2019) showed that tropical cyclogenesis is strongly suppressed within an equatorial deformation radius of the equator.) At higher latitudes, baroclinic waves prevail, and the mean surface winds are westerly. Note also the relative lack of low OLR in the subtropics, between about 20^o and 25^o latitude, between the equatorial belt and the region of tropical cyclones. Cross-sections (not shown here) reveal these are the descent branches of Hadley-like circulations.

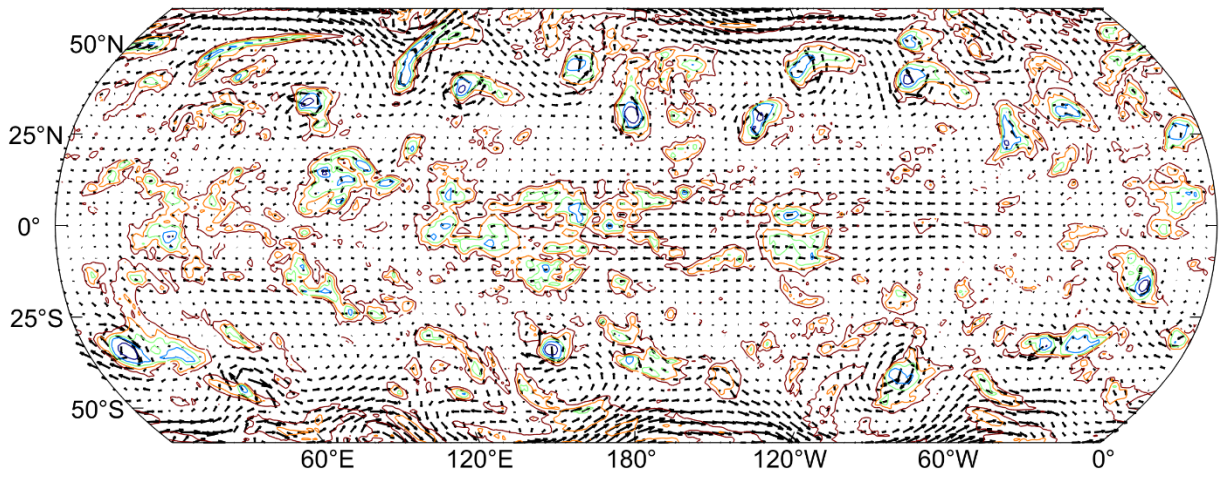


Figure 7.13: Outgoing longwave radiation (contoured only for values less than 280 W m^{-2}) and 850 hPa winds (arrows) on day 170 of a 360-day global SAM simulation.

A time-longitude section of outgoing longwave radiation averaged between 10^{o} S and 10^{o} N for the last 300 days of the simulation is shown in Figure 7.14. Aside from interpolation to a uniform 1-degree grid, no filtering has been applied.

The most prominent signal obvious to the naked eye is a wavenumber 1 disturbance propagating eastward with a period of around 40 days; this is the model's analog to an MJO. But there is also organized convection at small scales and higher frequencies, including a prominent westward-propagating mode that takes roughly 100 days to traverse the globe.

Spectral power density of precipitation averaged between 10^{o} S and 10^{o} N for the last 300 days of the simulation is shown in Figure 7.15, for the equatorially symmetric and anti-symmetric components. These plots have been made following the procedure described in Wheeler and Kiladis (1999) except that there is no need here to remove an annual cycle; and, as described before, the subtracted background consists of the average of the symmetric and anti-symmetric components. To deal with the quantization in zonal wavenumber, the plots have been smoothed in wavenumber only (not frequency) using a 1-2-1 smoother applied successively 7 times.

The equatorially symmetric precipitation power spectra exhibit a prominent peak at eastward zonal wavenumber 1 and frequencies ranging from once in 96 days to once on 24 days. (Note that the procedure for frequency decomposition operates on multiple 96-day subsets of the 300-day record, so that frequencies are quantized in increments of 1/96.) This peak, corresponding to the model's rendition of an MJO, is by far the largest spectral peak in the dataset. There is also a broad band of power extending along the convectively coupled Kelvin wave line corresponding to an intrinsic phase speed of about $12\ ms^{-1}$, and some power in westward-moving disturbances possibly corresponding to $n=1$ equatorial Rossby waves.

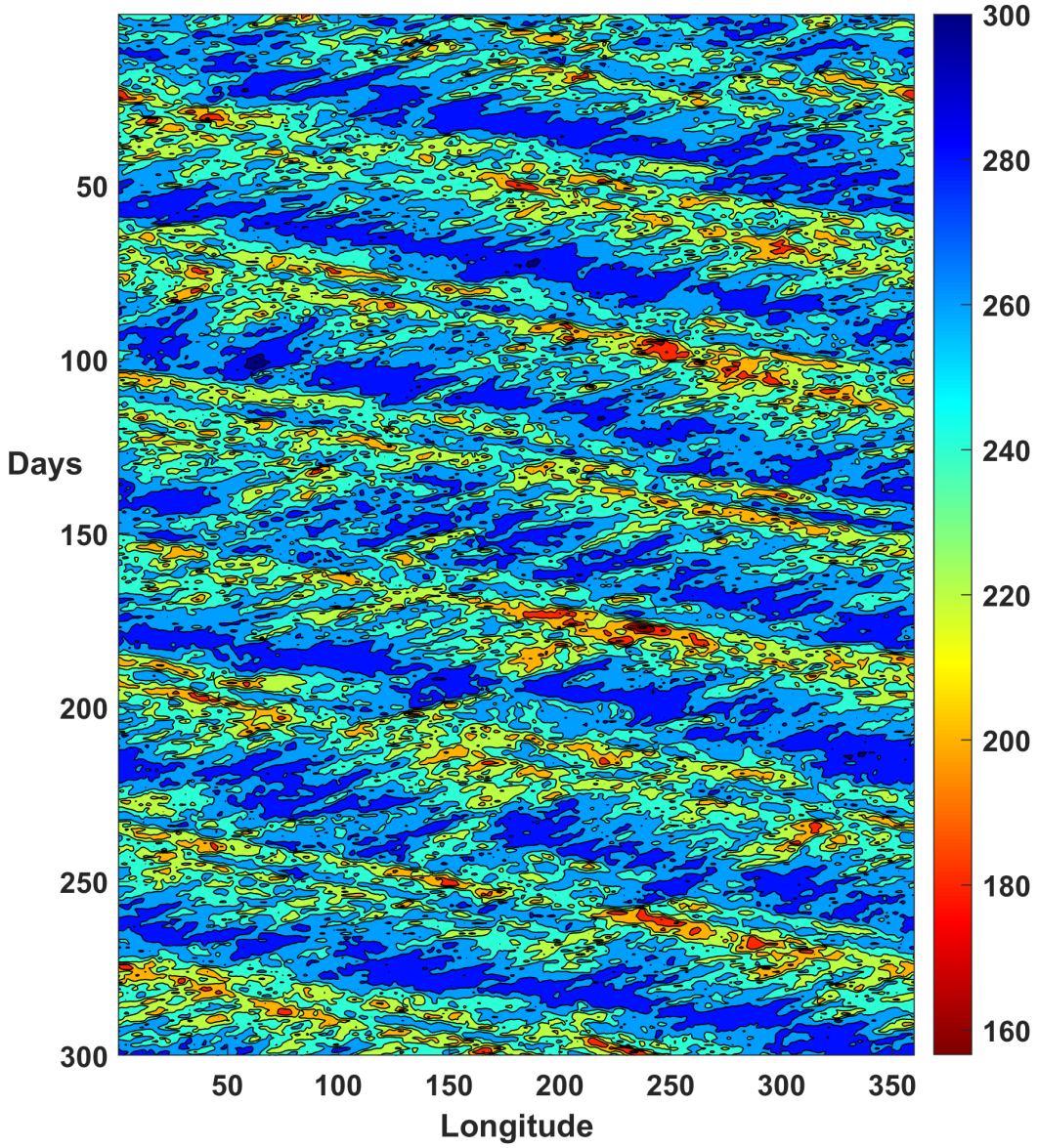


Figure 7.14: Outgoing longwave radiation (W m^{-2}) averaged between 10^{o} S and 10^{o} N during the last 300 days of the simulation. Model output has been interpolated to a 1-degree grid.

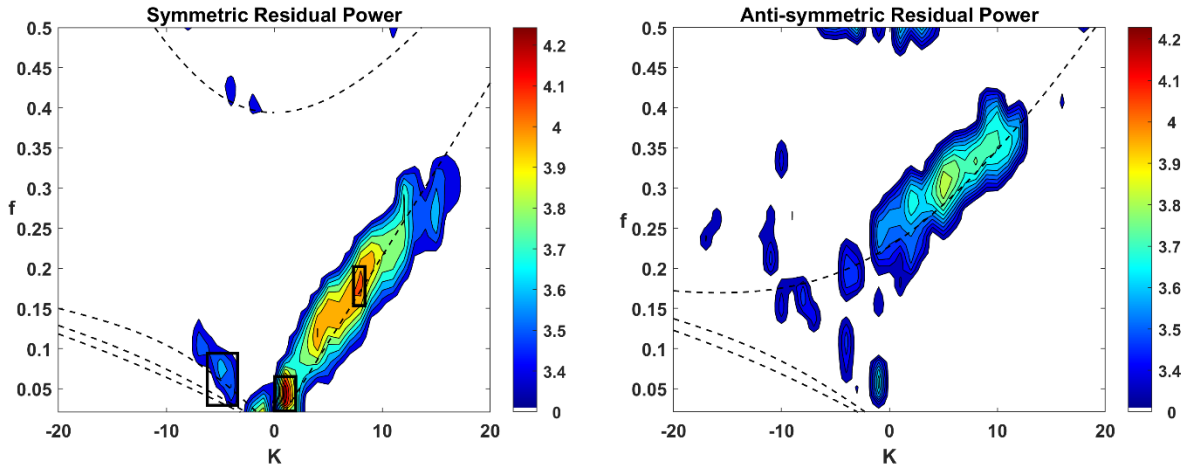


Figure 7.15: The base 10 logarithm of the spectral power density of precipitation from the last 300 days of the global SAM simulation. Equatorially symmetric (left) and antisymmetric (right) components. Power has been smoothed in wavenumber only by applying a 1-2-1 smoother 7 times. Values less than 3.3 have not been contoured. The dashed lines show the Matsuno dispersion curves for n= -1 to +5 assuming a shallow water phase speed of 12 ms^{-1}, with all curves Doppler shifted assuming a background easterly wind of 2 ms^{-1}. The solid boxes in the left-hand plot show the filtering domains for the n=1 equatorial Rossby wave, MJO and Kelvin waves, from left to right.

The equatorially anti-symmetric power spectral density plot shows a broad maximum along the Doppler-shifted mixed Rossby-gravity wave dispersion curve corresponding to the non-rotating shallow water intrinsic phase speed of $12\ ms^{-1}$ and extending from eastward zonal wavenumber 0 through about 12. Unlike the observed spectrum (Figure 7.3 and 7.4, right panels) there is little power in westward-propagating mixed Rossby-gravity waves. The peak signal strength is not quite large enough to obtain good reconstructions of spatial variability.

To examine the spatial structure of the SAM equatorial variability, we first choose a base variable that will serve as wave tracer and filter to include specified wavenumber-frequency combinations as indicated by the boxes in the left panel of Figure 7.15. The next step is to subdivide the longitude-time series at each latitude into space-time subdomains whose longitudinal extent is the wavelength of a single specified wavenumber and whose time extent corresponds to the period of the lowest specified frequency. For each subdomain, we calculate the frequency-wavenumber Fourier components and filter out all but the desired range of frequencies and single wavenumber. We then recombine the remaining frequencies and wavenumber into time-longitude sections, averaging them between specified latitude bounds corresponding approximately to expected meridional scales of the disturbances (e.g. 15° in the case of the MJO mode). At each time, we find the longitude of the maximum or minimum of the base variables…this longitude is not necessarily continuous in time.

We similarly filter all the other variables of interest and recombine the filtered single wavenumber and frequency range into time-longitude sections at each latitude for each space-time subdomain. We then average these in time but use longitude relative to that of the maximum or minimum value of the base variable, as described in the previous paragraph[4]. Finally, we average the results over all the space-time subdomains. The disadvantage of this method is that it only includes a single wavenumber and the lowest specified frequency and its

[4] This procedure differs from the time lag analysis employed by Wheeler and Kiladis (2000) and here takes advantage of the zonal statistical homogeneity of the SAM aquaplanet fields.

harmonics up to a specified upper bound; it excludes frequencies higher than the specified lower bound but not part of the set of harmonics of the lower bound. It must be borne in mind that the structures that emerge from this procedure may not be very representative of what one might see in a snapshot of the full fields, which are comprised of many wavenumbers and frequencies.

To create a composite vertical structure on the equator we apply the same procedure, but substituting pressure for latitude.

As mentioned previously, this procedure failed to detect any coherent structures anywhere in the equatorially anti-symmetric part of the power spectrum.

7.2.3.1 The MJO

For the MJO we used OLR as the base variable. The horizontal structure of the MJO mode is summarized in Figure 7.16. Following Wheeler set al. (2000), this composite includes both equatorially symmetric and anti-symmetric components. Near the equator, the wind and surface pressure fields have the same relative relationship as in a classical Kelvin wave, with the strongest easterlies aligned with the lowest surface pressure and divergence dominated by its zonal component.

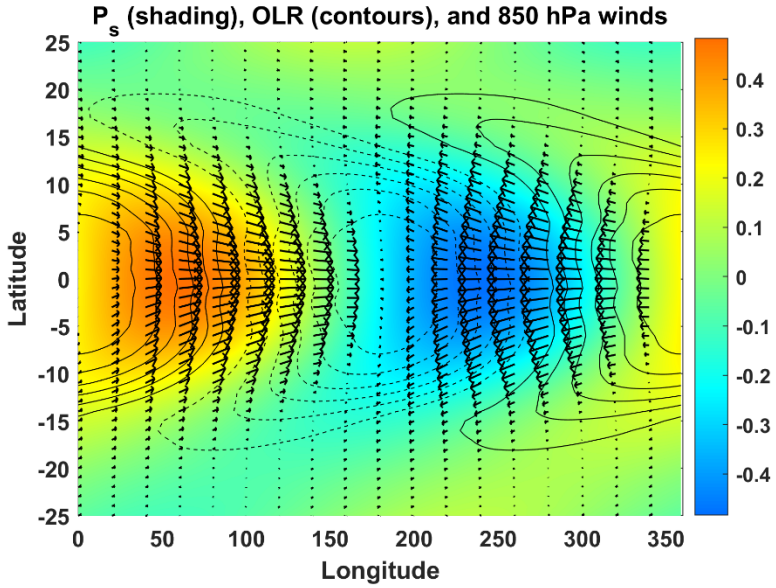


Figure 7.16: Composite structure of the MJO mode in the SAM aquaplanet simulation. Surface pressure (hPa) shaded, OLR contoured, with dashed lines denoting negative anomalies, and 850 hPa winds denoted by arrows. The maximum wind speed anomaly is 2.3 ms^{-1} and the maximum OLR anomaly is 10.0 $W\ m^{-2}$. This composite includes both equatorially symmetric and antisymmetric components, though it is dominated by the former.

Unlike a pure Kelvin wave, the meridional velocity off the equator is non-zero and shows meridional convergence west of and divergence east of the OLR minimum.

The physics driving this mode is illuminated through the budget of vertically integrated moist static energy, as illustrated in Figure 7.17.

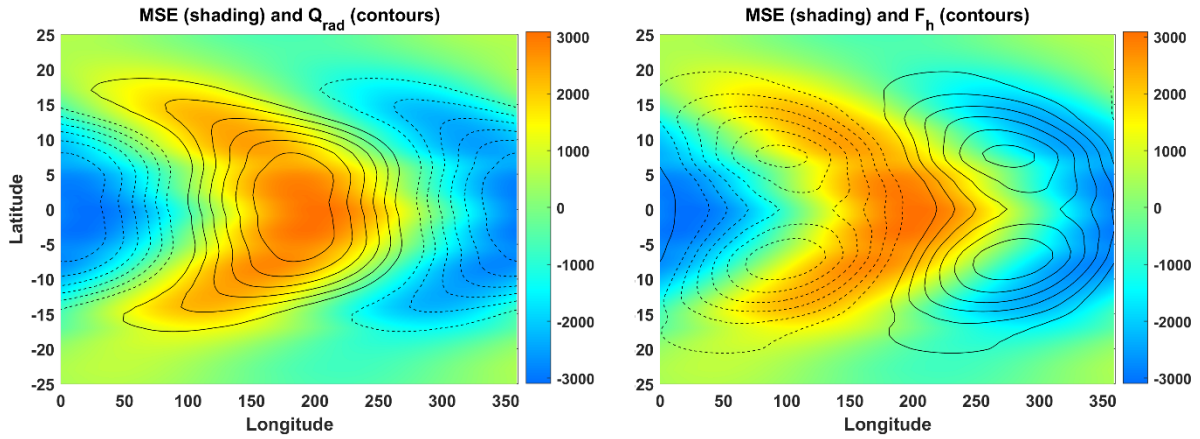


Figure 7.17: Composite structure of column-average moist static energy (J Kg^{-1}; shading) and column-integrated radiative heating (contours in left panel) and surface enthalpy flux (right panel) for the MJO mode. The maximum absolute value of the column-integrated heating is 9.3 W m^{-2} and the maximum absolute value of the surface enthalpy flux is 8.7 W m^{-2}.

The budget equation for the moist static energy, vertically integrated through the troposphere, may be written

$$\frac{\partial \langle h_m \rangle}{\partial t} + c\frac{\partial \langle h_m \rangle}{\partial x} = -\langle \mathbf{V} \bullet \nabla h_m \rangle - \left\langle \omega \frac{\partial h_m}{\partial p} \right\rangle + F_0 + R, \tag{7.2}$$

where the angle brackets denote a mass-weighted vertical integral through the depth of the troposphere, h_m is the moist static energy, c is the zonal phase speed of the disturbance, the first term on the left of (7.2) is the vertically integrated horizontal advection of moist static energy, the second term is the integrated vertical advection, the third term is the surface enthalpy flux, and the last term is the column-integrated radiative heating, which is just the difference between the surface and tropopause values of the net radiative flux.

In the long run, and in a coordinate system moving zonally with the disturbance, we expect the left side of (7.2) to vanish if the system is in a steady state. Then the right side of (7.2) should sum to zero.

Figure 7.17 shows the vertically averaged moist static energy (shading) and the radiative (left panel) and surface flux (right panel) terms in (7.2). Khairoutdinov and Emanuel (2018) showed in a similar simulation that the surface flux and radiative terms in (7.2) were very nearly balanced by the vertical advection term, so we do not show that here. The net radiative heating term is strongly correlated with the vertically averaged moist static energy itself, so that this mode is being strongly amplified by radiation. The surface flux variability, much of which is owing to fluctuating surface wind speed, is nearly in quadrature with the moist static energy, so that WISHE is driving the wave eastward. But close inspection of the figure reveals a small negative correlation between moist static energy and the surface enthalpy flux, so the latter is exerting a weak damping effect in the wave.

In short, the MJO in this model is being driven by radiation (most of whose variability is driven by clouds) and propagated eastward by WISHE.

The vertical structure of the MJO mode, subject to the same filtering, is shown in figure 7.18. The zonal wind, vertical velocity, and virtual temperature have tropospheric structure consistent with a dominant first baroclinic mode structure, as predicted in an atmosphere constrained to have a moist adiabatic virtual temperature lapse rate. The slight forward tilt of the zonal wind structure is consistent with the presence of a barotropic component excited by some combination of upward energy radiation into the stratosphere and surface friction. The forward phase tilt of the structures of all four variables in the stratosphere is owing to upward radiation of wave energy by Kelvin waves, a process difficult (but not impossible) to include in simple linear models (Lin and Emanuel 2022). The moist static energy has a pronounced eastward tilt, with anomalies in the shallow convective boundary layer (below about 900 hPa) of the opposite sign as free tropospheric anomalies. Given that deep convection is likely mostly in phase with the large-scale vertical velocity, it may be the case that the deep convection is simultaneously increasing the moist static energy of the free troposphere and reducing it in the boundary layer through downdrafts. Note, as predicted by convective neutrality, that the boundary layer moist static energy is in phase with the free tropospheric virtual temperature.

7.2.3.2 Kelvin waves

We apply the same procedure to examine the structure of Kelvin waves, focusing on zonal wavenumber 9 and frequencies from 0.15 to 0.21 d^{-1}, but in contrast to the MJO we consider only equatorially symmetric perturbations. Figure 7.19 shows the horizontal structure of this relatively low frequency Kelvin component. The structure of the wind and pressure field is that of a classical Kelvin wave, with the zonal winds very nearly in phase with the surface pressure. There is no obvious meridional wind component, though the procedure does pick up maximum absolute values of about 0.05 ms^{-1}, perhaps more a measure of how adept the procedure is than of any real meridional wind contribution. Note that the OLR minimum is displaced west of the maximum convergence of 850 hPa zonal winds.

The phase relationships between the column mean moist static energy and column radiative heating and surface enthalpy fluxes are shown in Figure 7.20. It is clear that both WISHE and radiation act in concert to amplify this Kelvin-like mode, with WISHE helping to drive it eastward while the radiation anomalies retard this motion. Curiously, when we apply the same diagnostics to lower frequency Kelvin-like modes (not shown here), the wave driving shifts away from WISHE and toward radiation, and conversely, at higher frequencies WISHE begins to dominate the wave driving. This demonstrates that the relative importance of different contributions to the column moist static energy budget may change along what appears to be a continuum in the power spectral density.

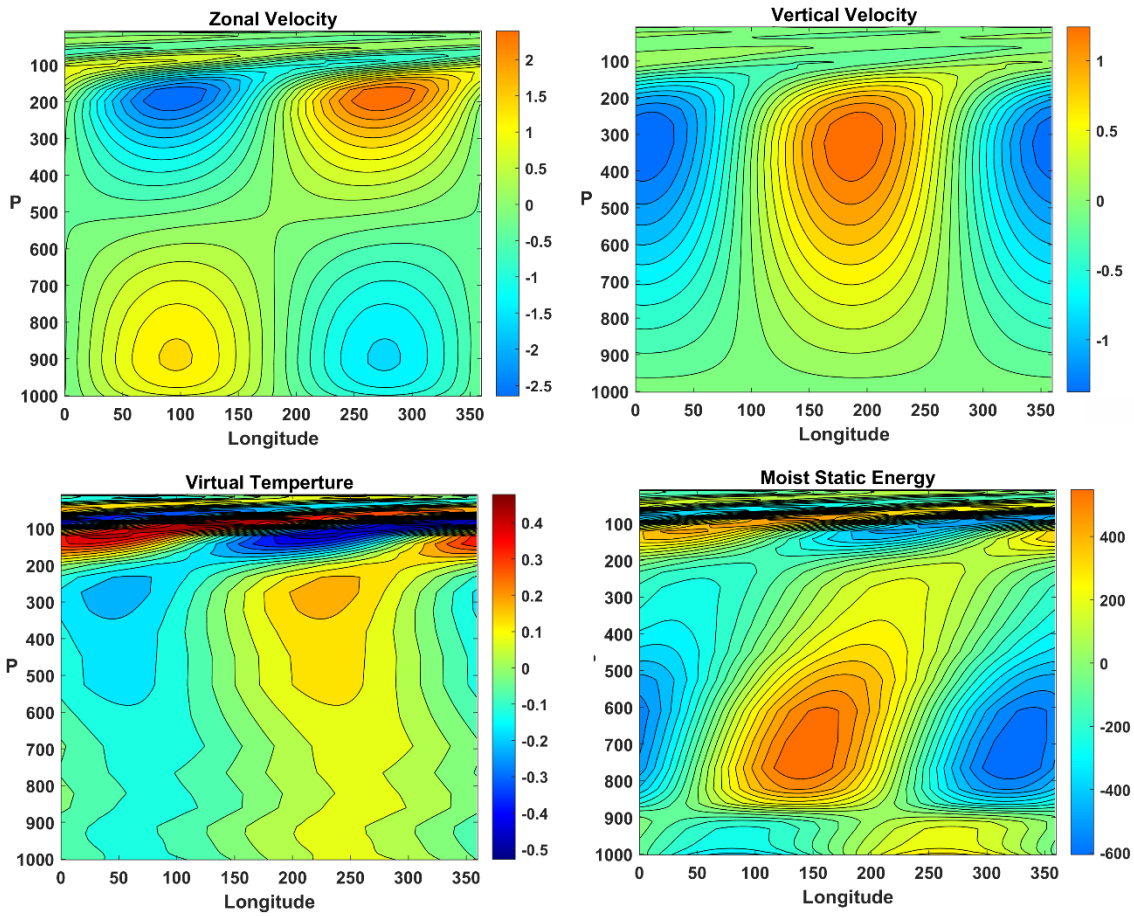


Figure 7.18: Vertical structure of the MJO mode averaged between 15°N and 15°S: Upper left: zonal wind (ms^{-1}); upper right: vertical velocity ($mm\ s^{-1}$); lower left: virtual temperature (K) and lower right: moist static energy ($J\ Kg^{-1}$).

The vertical structure of the $k = 9$ Kelvin mode is shown in Figure 7.21, where we have averaged the fields between 5°S and 5°N. It differs in several remarkable ways from the vertical structure of the MJO mode shown in Figure 7.18; most particularly, the virtual temperature strongly departs from first baroclinic mode structure and exhibits two prominent peaks in the troposphere and a third near the tropopause. Comparison of the virtual temperature with the boundary layer moist static energy suggests that while convective neutrality may hold for the lower half of the troposphere, it breaks down dramatically in the upper half, which is more or less in quadrature with the lower half. Both the zonal and vertical velocity fields tilt substantially westward with height, although they are otherwise consistent with first baroclinic mode structure, with two extrema of zonal wind in the troposphere and a single extremum of vertical velocity. As with the MJO, there are prominent indicators of upward propagation of Kelvin wave energy in the stratosphere.

Note that the vertical structure of this Kelvin-like mode strongly resembles the observed vertical structure of equatorial Kelvin waves, as shown on the left side of Figure 7.10, including the two temperature extrema in the troposphere and the westward-sloping zonal wind structure.

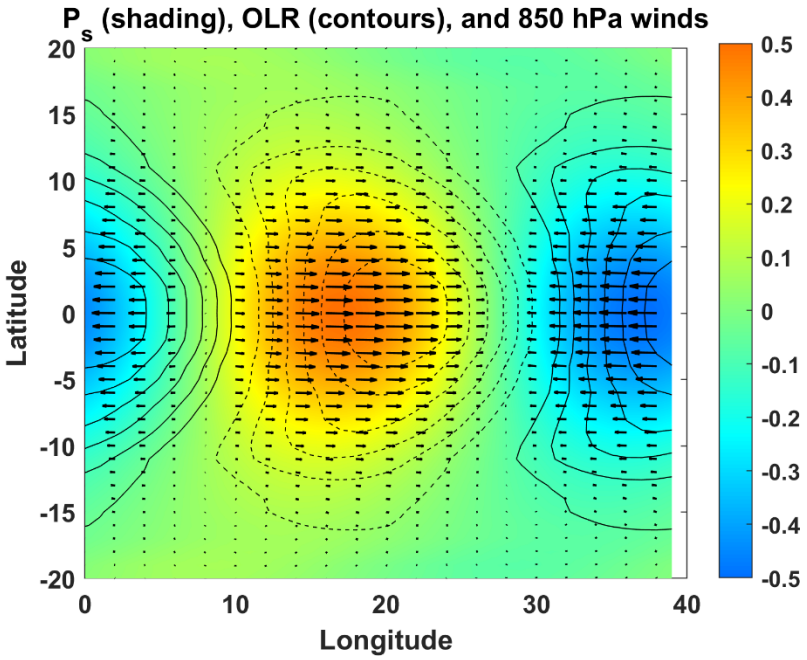


Figure 7.19: Composite structure of the Kelvin mode in the SAM aquaplanet simulation. Surface pressure (hPa) shaded, OLR contoured, with dashed lines denoting negative anomalies, and 850 hPa winds denoted by arrows. The maximum wind speed anomaly is 1.5 ms^{-1} and the maximum OLR anomaly is 17.3 W m^{-2}.

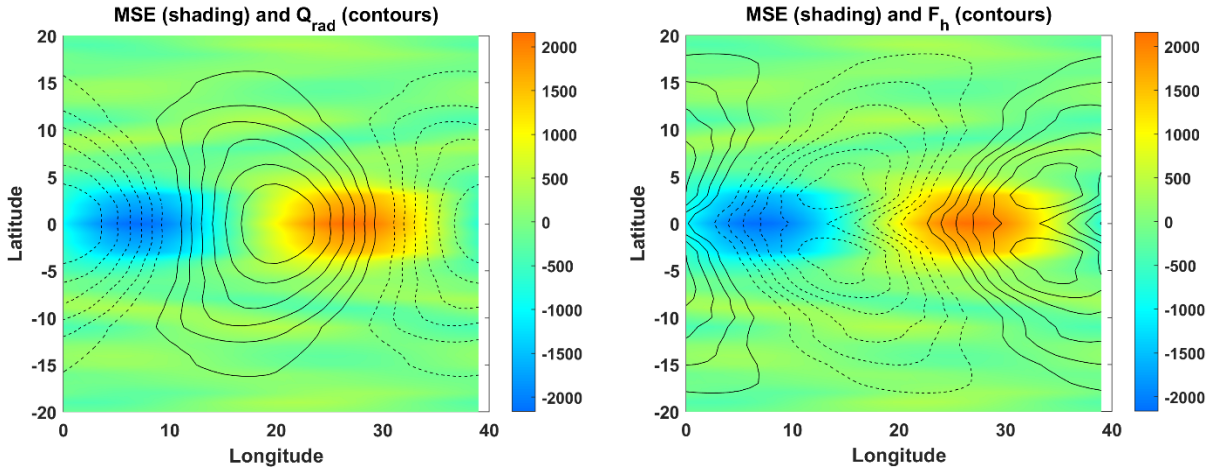


Figure 7.20: Composite structure of column-average moist static energy (J Kg^{-1}; shading) and column-integrated radiative heating (contours in left panel) and surface enthalpy flux (right panel) for the Kelvin mode. The maximum absolute value of the column-integrated heating is 17.7 W m^{-2} and the maximum absolute value of the surface enthalpy flux is 6.5 W m^{-2}.

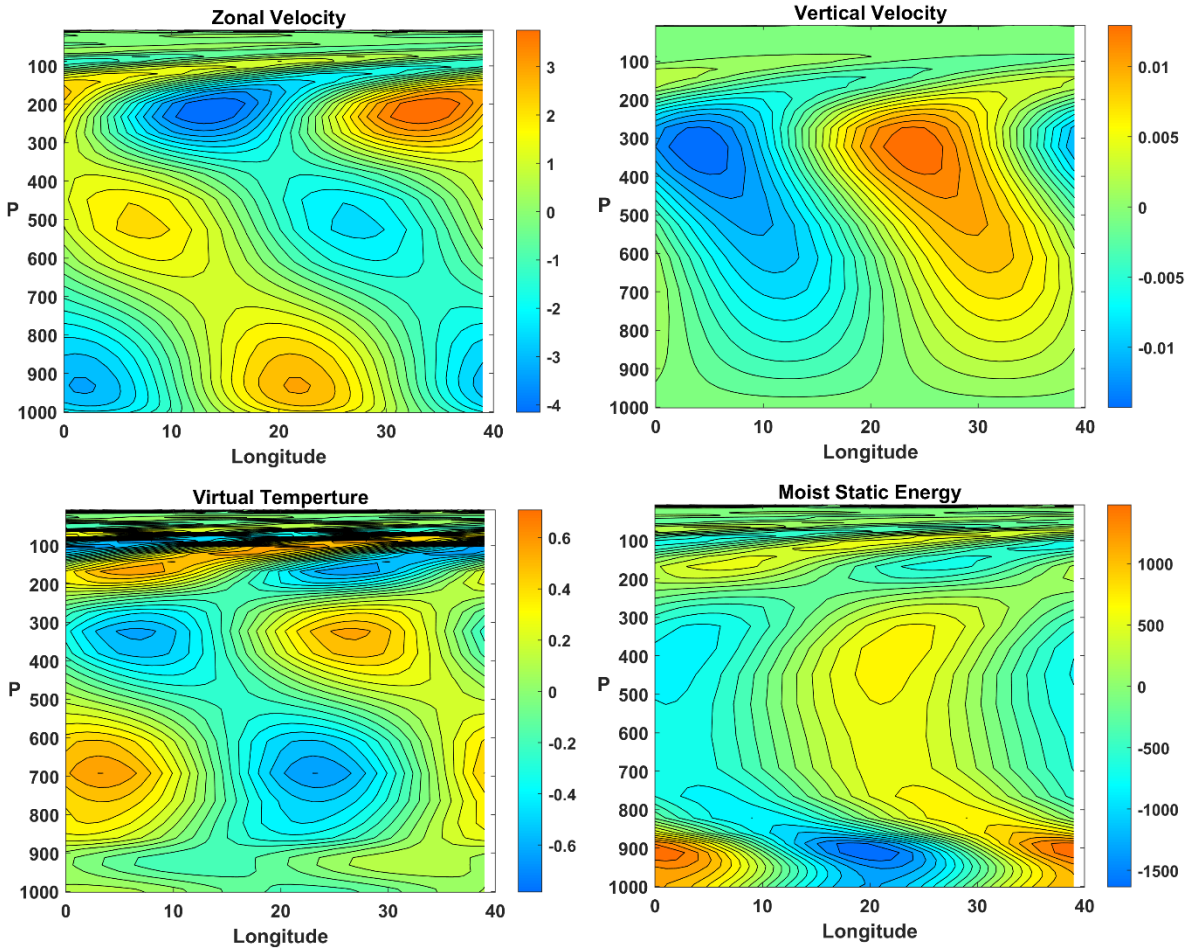


Figure 7.21: Vertical structure of the Kelvin mode averaged between 5ºN and 5ºS: Upper left: zonal wind (ms^{-1}); upper right: vertical velocity ($mm\ s^{-1}$); lower left: virtual temperature (K) and lower right: moist static energy ($J\ Kg^{-1}$).

7.2.3.3 Westward-propagating symmetric mode

Finally, we look at the structure of the variability in a region of spectral space ordinarily associated with an $n=1$ equatorial Rossby wave; this region is enclosed by the left-most box in Figure 7.15. Figure 7.22 summarizes the structure of this mode at wavenumber $k=-5$ and frequencies between 0.03 and 0.09 cycles per day. The structure of this disturbance does not resemble any single classical Matsuno mode. The phase relationship between the column mean moist static energy and the radiation and surface enthalpy fluxes (Figure 7.23) shows that this mode, like the MJO mode, is amplified by radiation and propagated (westward, in this case) by WISHE, which also slightly damps the wave. Unlike either the MJO or Kelvin modes, most of the strong forcing of the column moist static energy occurs well off the equator, even though the winds are strongest along the equator. The vertical structure of the disturbance (Figure 7.24) is close to the first baroclinic mode, but the free troposphere virtual temperature perturbations are not in phase with the lowest level moist static energy, so the waves are not maintaining strict moist neutrality, at least not along the equator. The lack of correspondence between the structure of this disturbance and that of any Matsuno mode cautions us against trying to fit all

equatorial variability into the simple framework of linear neutral, shallow water equatorial modes. If the effects of radiation and surface fluxes are strong enough, even a loose correspondence of the amplifying modes with neutral modes may vanish, leaving us without much vocabulary to label what we see in this model, and perhaps in observations as well.

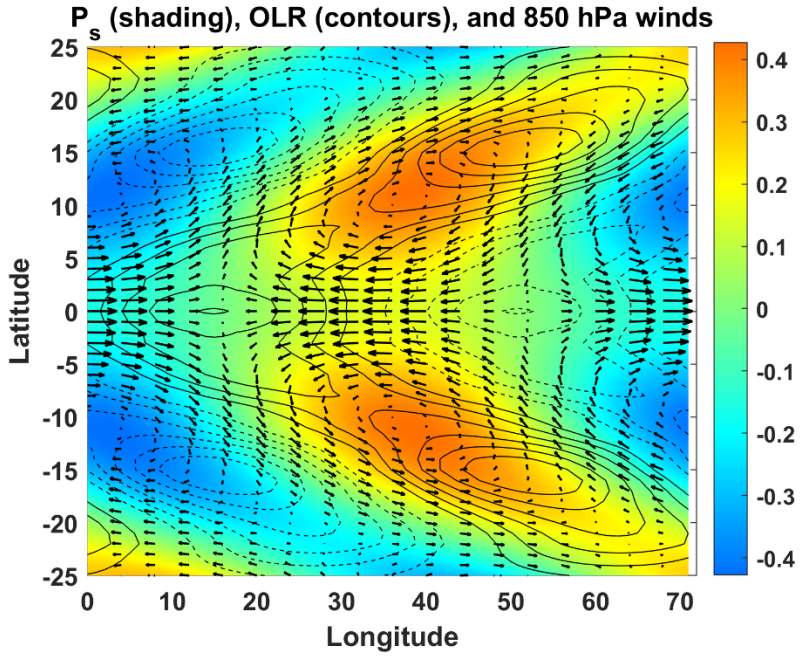


Figure 7.22: Composite structure of a low frequency westward-propagating mode in the SAM aquaplanet simulation. Surface pressure (hPa) shaded, OLR contoured, with dashed lines denoting negative anomalies, and 850 hPa winds denoted by arrows. The maximum wind speed anomaly is 2.9 ms^{-1} and the maximum OLR anomaly is 10.9 $W\ m^{-2}$.

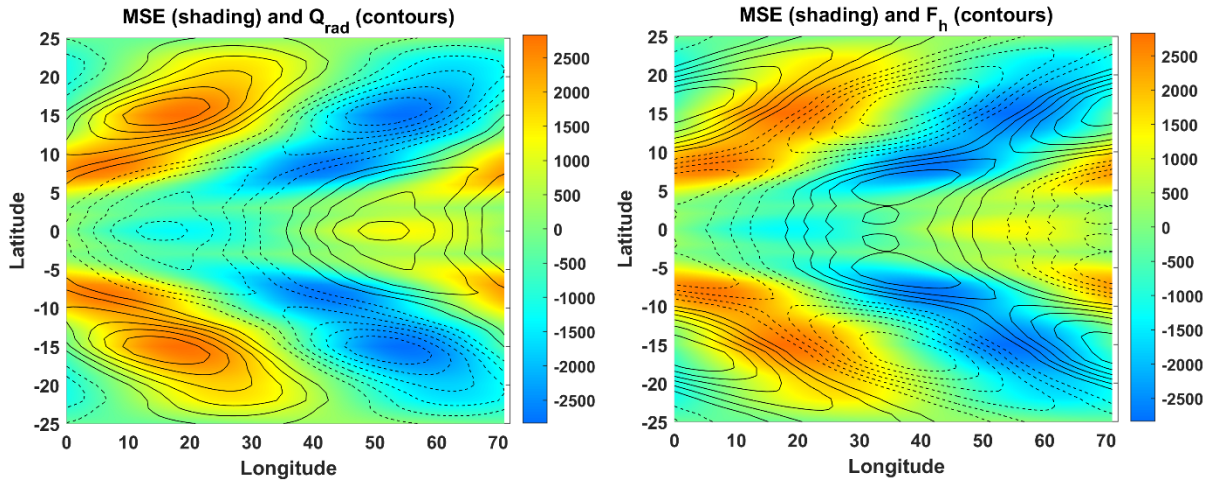


Figure 7.23: Composite structure of column-average moist static energy (J Kg^{-1}; shading) and column-integrated radiative heating (contours in left panel) and surface enthalpy flux (right panel) for the low-frequency, westward-propagating mode. The maximum absolute value of the column-integrated heating is 9.9 $W\ m^{-2}$ and the maximum absolute value of the surface enthalpy flux is 8.3 $W\ m^{-2}$.

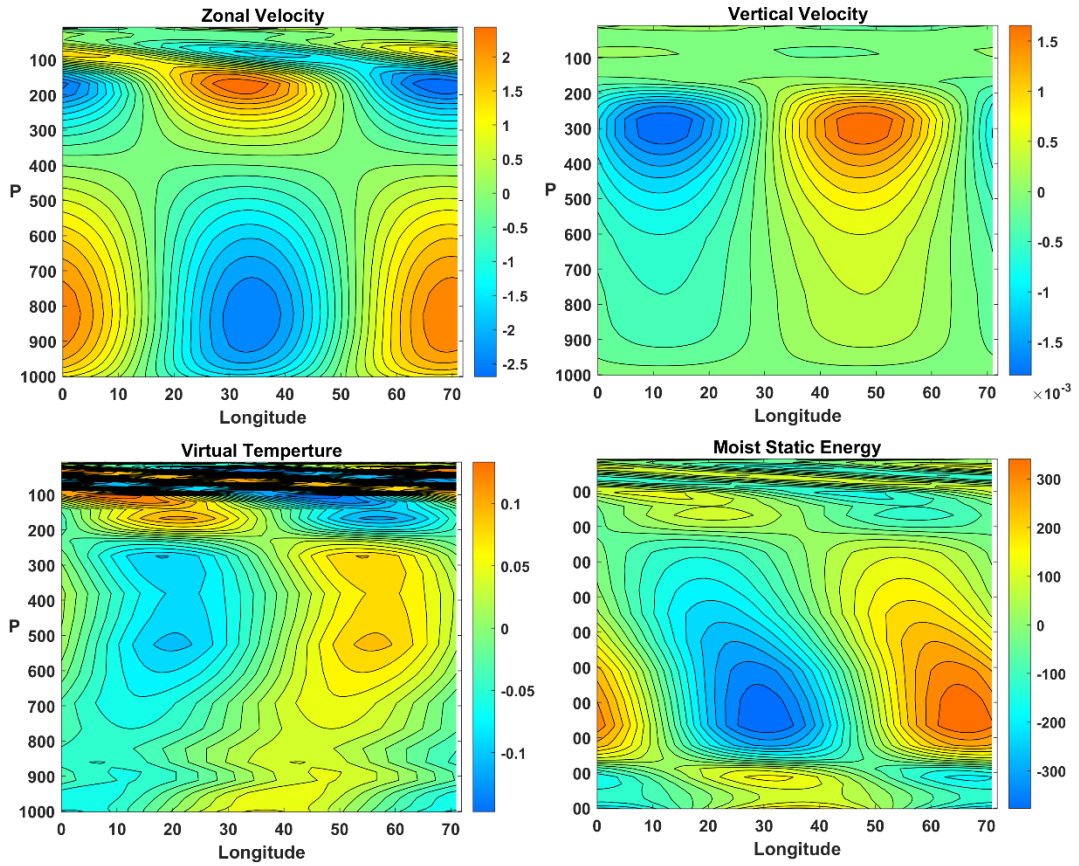


Figure 7.24: Vertical structure of the low-frequency, westward-propagating mode averaged between 8ºN and 8ºS: Upper left: zonal wind (ms^{-1}); upper right: vertical velocity (mm s^{-1}); lower left: virtual temperature (K) and lower right: moist static energy (J Kg^{-1}).

7.2.4 Theory

In Chapter 6 we showed that the variation of the Coriolis parameter across the equator has the effect of trapping various kinds of oceanic waves near the equator, and the very similar dynamics governing the moist adiabatic troposphere will likewise contain solutions in the form of equatorially trapped waves. We can explore these by adapting the linear atmospheric model developed in Chapter 6 and expressed there by equations (6.75) – (6.79). Given the rich variability produced by the SAM simulations with a fixed distribution of SST, described in the previous section, one can hope to produce some realistic variability in a linear model with fixed SST, so we omit perturbations to the surface saturation entropy, s_0^*. We drop the drag terms in the momentum equations. While these are not necessarily small, they do not appear essential to the problem and increase the mathematical complexity of the system.

The SAM results presented in the previous subsection demonstrate that variability of radiation is an essential element of tropical intraseasonal variability. We expect that most of this variability owes its existence to variations in tropospheric humidity and clouds; the tiny magnitude of temperature perturbations (e.g. a few tenths of a degree in the SAM simulations) make that an unlikely explanation for the large variations in radiative cooling rates.

To make progress in relating variations in radiative cooling to the dependent variables of our linear system, we make use of an interesting relationship between OLR and moist entropy discovered by Sandrine Bony (Bony and Emanuel 2005) as illustrated in Figure 7.25, which shows a time series of satellite-measured OLR and compares it to the quantity $\ln\left(\frac{\theta_{eb}}{\theta_{em}}\right)$, where θ_{eb} is the equivalent potential temperature of the subcloud layer (975 hPa) and θ_{em} is the same quantity averaged between 975 hPa and 300 hPa.

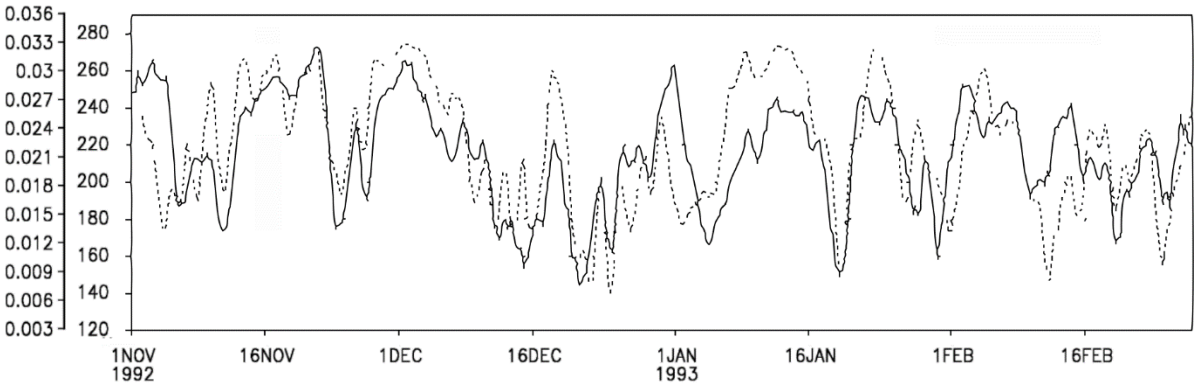


Figure 7.25: Time series of OLR (W m^{-2}; dotted curve) and the logarithm of the ratio of 975 hPa θ_e to the mean θ_e between 975 hPa and 300 hPa (solid). The OLR scale is on the inner side of the y-axis label, while the scale of the log of the ratios of θ_e is on the outer side of the y-axis label. From satellite and rawinsonde data collected during TOGA COARE.

The data were collected as part of the TOGA-COARE experiment in the western tropical Pacific in 1992-1993, with the equivalent potential temperature taken from an array of 4 rawinsondes (see Figure 3.18). If we assume that the modulation of net radiative cooling of the troposphere closely follows the modulation of OLR (i.e., neglecting variations of the net surface longwave flux and variations in the absorption of solar radiation in the troposphere), and we also assume that the variations in the ratio of subcloud layer θ_e to mean tropospheric θ_e are dominated by the latter[5], we can represent fluctuations in radiative cooling by the simple formula

$$Q = -Ch_m, \tag{7.3}$$

where C is a nondimensional constant that represents the magnitude of the modulation of radiative cooling, and h_m is the (nondimensional) perturbation column-average moist static energy of the troposphere.

Finally, to qualitatively mimic the effect of upward radiation of wave energy into the stratosphere (a process we will discuss later), we add a wavelength-dependent damping term to (6.79), which governs the evolution of column-mean moist static energy.

[5] We could incorporate the effect of fluctuating subcloud layer θ_e, but that would only have the effect of slightly changing the value of the parameter χ in the equations.

With this parameterization and the aforementioned assumptions, the linear model represented by equations (6.75) – (6.79) becomes

$$\frac{\partial u}{\partial t} = \frac{\partial s}{\partial x} + yv, \tag{7.4}$$

$$\frac{\partial v}{\partial t} = \delta\left(\frac{\partial s}{\partial y} - yu\right), \tag{7.5}$$

$$w = -\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right), \tag{7.6}$$

$$\frac{\partial s}{\partial t} = -\left(\chi + \chi_2\right)s - \alpha u - w + \left(1 + C\right)h_m, \tag{7.7}$$

and

$$\gamma\frac{\partial h_m}{\partial t} = -\chi s - \alpha u - G_m w + Ch_m + d\frac{\partial^2 h_m}{\partial x^2}. \tag{7.8}$$

Here the new nondimensional parameter d governs the wavelength-dependent damping of moist static energy.

Recall that a fundamental assumption underlying this linear model of the equatorial troposphere is that the lapse rate of temperature is always moist adiabatic; this assumption filters out all but a barotropic mode and the first baroclinic mode. In this case, having assumed a rigid lid at the tropopause and no surface drag, only the first baroclinic mode survives. This mode has a single virtual temperature extremum in the troposphere, and while this is consistent with the MJO in observations and in the SAM model (see Figure 7.18), it is not consistent with observations or SAM simulations of the higher frequency modes such as Kelvin waves (Figures 7.10 and 7.21). For the present, we press on to examine solutions to this linear model, as a baseline, and later return to the issue of the higher frequency modes.

We begin by eliminating, through cross-differentiation, all the variables in (7.4) – (7.8) in favor of the meridional velocity, v, and look for modal solutions of the form

$$v = \text{Real}\left[V(y)e^{ikx+\sigma t}\right], \tag{7.9}$$

where $V(y)$ is a complex function, k is the zonal wavenumber $\left(k = 1, 2, 3....\right)$ and σ is a complex growth rate. This yields an ordinary differential equation for $V(y)$:

$$\sigma a_3\frac{d^2V}{dy^2} - a_2 y\frac{dV}{dy} + \left[ika_3 - a_2 - \frac{\sigma}{\delta}\left(a_1\sigma + ika_2 + k^2a_3\right) - a_1y^2\right]V = 0, \tag{7.10}$$

where

$$
\begin{aligned}
a_1 &\equiv \chi\left(1+C\right)+\left(\chi+\chi_2+\sigma\right)\left(\gamma\sigma+dk^2-C\right),\\
a_2 &\equiv \alpha\left(\gamma\sigma+dk^2+1\right),\\
a_3 &\equiv \gamma\sigma+dk^2-C+G\left(1+C\right).
\end{aligned}
\tag{7.11}
$$

As can be verified by substitution, equation (7.10) has solutions of the form

$$
V_n = H_n(y)e^{-by^2}, \tag{7.12}
$$

where the complex exponential coefficient b is given by

$$
b = \frac{-a_2 \pm \sqrt{a_2^2 + 4\sigma a_1 a_3}}{4\sigma a_3}, \tag{7.13}
$$

and $H_n(y)$ are Hermite polynomials whose first 4 terms are

$$
\begin{aligned}
H_0 &= 1,\\
H_1 &= y,\\
H_2 &= \frac{1}{2}y^2 + c_1,\\
H_3 &= \frac{1}{3}y^3 + c_2 y,
\end{aligned}
\tag{7.14}
$$

with constants c_1 and c_2 given by

$$
\begin{aligned}
c_1 &= \frac{\sigma a_3}{\left(2b\sigma - ik\right)a_3 + a_2 + \dfrac{\sigma}{\delta}\left(a_1\sigma + ika_2 + k^2 a_3\right)},\\
c_2 &= \frac{2\sigma a_3}{\left(6b\sigma - ik\right)a_3 + 2a_2 + \dfrac{\sigma}{\delta}\left(a_1\sigma + ika_2 + k^2 a_3\right)}.
\end{aligned}
\tag{7.15}
$$

For these solutions, the complex growth rate σ must satisfy the dispersion relation

$$
\frac{1}{2}a_2 - ika_3 + \frac{\sigma}{\delta}\left(a_1\sigma + ika_2 + k^2 a_3\right) \pm \left(n+\frac{1}{2}\right)\sqrt{a_2^2 + 4\sigma a_1 a_3} = 0, \tag{7.16}
$$

in which $n = 0,1,2,3....$, corresponding to the order of the Hermite polynomials in (7.14). The choice of sign in the square root in (7.16) must match that of (7.13). To satisfy the boundary conditions at $y = \pm\infty,$ the real part of b must be positive; any roots of (7.16) that do not satisfy this condition are discarded.

When (7.16) is expanded, it becomes apparent that it is an 8th order polynomial equation for the complex growth rate σ. We use a polynomial equation solver in MATLAB to solve this for those roots that obey the boundary conditions in y and check, by substitution, that they indeed solve

the dispersion relation and the boundary conditions. Routines for finding these eigenvalues and plotting the associated eigenfunctions, for specified choices of the governing parameters, are available at https://zenodo.org/doi/10.5281/zenodo.10888602 . Here we first consider some special cases before exploring the general solutions.

We first note that setting the damping parameters d, χ and χ_2, the WISHE term α, and the cloud-radiation term C all to zero reduces our system to a slightly modified form of the Matsuno equations for neutral oscillations on an equatorial β plane. (Setting h_m to zero in (7.7) decouples (7.8) from the rest of the system, which then takes the exact from of the Matsuno equations.) It can be demonstrated without solving the equations that their solutions are all neutral modes, which were described and discussed in Chapter 6. It can further be shown that including positive values of d, χ and χ_2 in any combination leads to solutions that decay exponentially in time. Amplifying solutions, on the other hand, require positive values of the WISHE parameter α and/or the cloud-radiation parameter C.

7.2.4.1 WISHE modes

We first explore amplifying disturbances that result from sufficiently large values of α while omitting the cloud-radiation terms. As a specific example, we take $\alpha = 1.5,\ \gamma = 1,\ G_m = 0.1,$ $C = 0,\ \chi = 1,\ \chi_2 = 0.2,\ d = 0.02,$ and $\delta = 30.$ The dispersion solutions for growing modes only are shown in Figure 7.26.

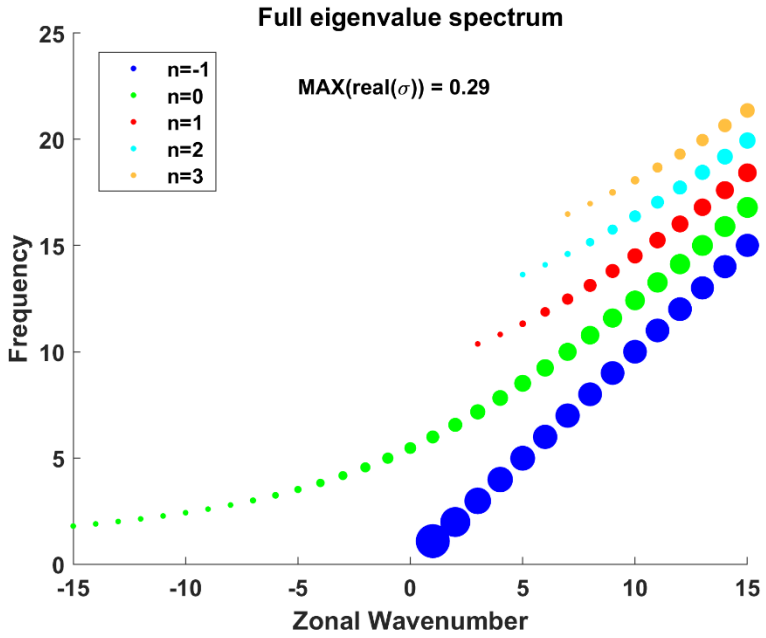


Figure 7.26: Amplifying solutions to the dispersion relation for WISHE-only wave driving and parameter values described in the text. The diameters of the circles are proportional to the exponential growth rate, with the value of the fastest growing mode indicated near the top of the figure. Solutions for the first 5 meridional mode numbers are displayed, including the Kelvin-like mode as n=-1, according to the colors described in the legend. The frequency here is nondimensional (Chapter 6, eq. 6.73).

The $v = 0$ Kelvin-like mode grows most rapidly, especially at wavenumber 1. The other Matsuno-like modes also grow, but only the mixed Rossby-gravity mode shows growing, westward-propagating solutions and these do not grow as fast as their eastward-propagating counterparts. (When α is set to a negative value, there are growing, westward-propagating Poincare waves, but no other growing solutions.) Note also that, except for the Kelvin-like mode, growth rates increase with wavenumber, implying that disturbances with shorter zonal wavelengths should dominate.

The physics underlying the wave growth and propagation are straightforward to understand, and is illustrated in Figure 7.27. The structure is, to a close approximation, that of a classic equatorial Kelvin waves, with the strongest easterlies closely aligned with the minimum surface pressure, and ascent and descent nearly in quadrature with the temperature field (which is the inverse of the surface pressure field) but note the slight displacement of the w field such that it is slightly positively correlated with temperature, so that potential energy is being converted to kinetic energy. The surface enthalpy flux (right panel) is almost in quadrature with the column moist static energy, but here too there is a slight positive correlation between enthalpy flux and MSE, which is consistent with wave growth. Note also that the surface fluxes are shifted eastward from the maximum surface winds (see left panel)…this is because the saturation entropy is damping the surface fluxes, in phase with the entropy. The WISHE effect is responsible for both the wave growth and, in part, its eastward propagation. (The rest of the propagation can be regarded as the contribution (not shown in the figure) of vertical advection of moist static energy (third term on the right of (7.8)) to the net tendency of column moist static energy.) We are looking at the propagation through a thermodynamic lens here. Were we to adopt a more dynamical view, we would see that the off-equatorial, shear-related vorticity is propagating eastward owing to wave divergence acting on the planetary vorticity (e.g. Matthews 2021).

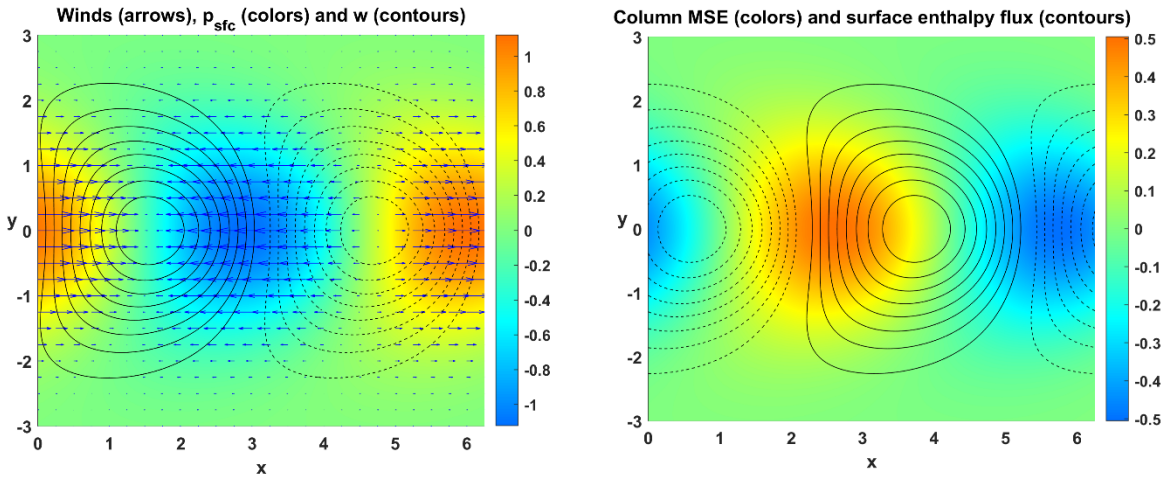


Figure 7.27: Eigenfunctions of the zonal wavenumber 1 Kelvin-like mode excited by WISHE. In the left panel, the shading shows the surface pressure, the vertical velocity is in black contours with dashes indicating negative values, and zonal wind is indicated by arrows. The right panel shows the column moist static energy (shading) and the surface enthalpy flux (black contours). The latitude variable is nondimensional, and the x axis spans one full zonal wavelength. The eigenfunctions also depend on the value of the precipitation efficiency, taken to be 0.8 here.

7.2.4.2 Effect of cloud-radiation interactions

Modulation of radiative cooling rates by clouds and water vapor have a large effect on low-frequency variability in the equatorial waveguide. This effect works in tandem with WISHE, turning on the cloud radiation term while turning off WISHE (not shown here) yields no viable growing solutions.

Figure 7.28 shows the dispersion relationship using the same parameters as in the previous subsection, but setting $C = 1$. Thus both WISHE and cloud radiation are active. The right panel of Figure 7.28 zooms in on the lower frequencies and wavenumbers

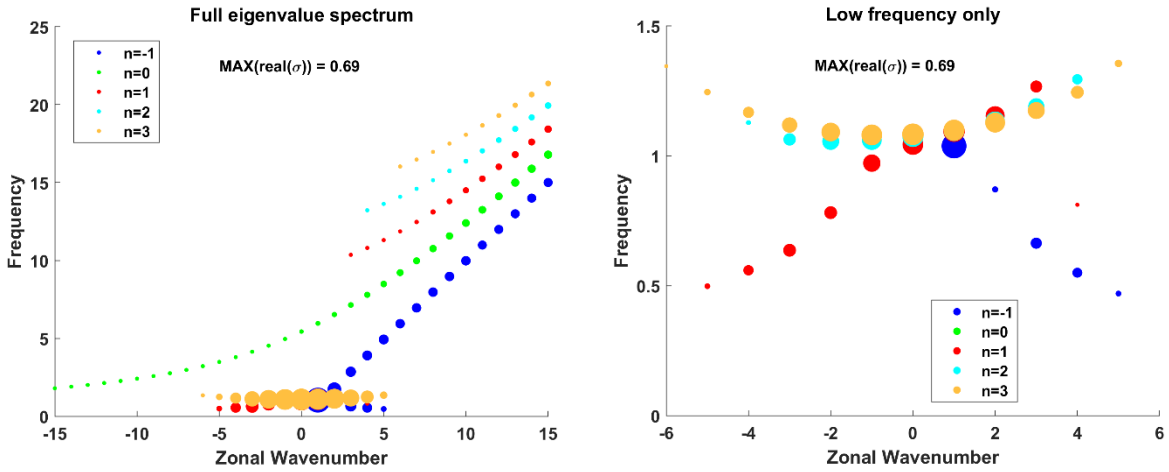


Figure 7.28: Amplifying solutions to the dispersion relation when both WISHE and cloud radiation interactions are active, and parameter values as before but with C=1. The right panel zooms in on lower frequencies and wavenumbers. The diameters of the circles are proportional to the exponential growth rate, with the value of the fastest growing mode indicated near the top of the figure. Solutions for the first 5 meridional mode numbers are displayed, including the Kelvin-like mode as n=-1, according to the colors described in the legend. The frequency here is nondimensional (Chapter 6, eq. 6.73).

Note that the maximum growth rate is more than twice what it was in the WISHE-only solution. The higher frequency WISHE modes remain intact, but the very lowest frequency Kelvin-like mode is greatly amplified compared to its WISHE-only counterpart. But the most interesting feature is a new array of very low frequency disturbances spanning zonal wavenumbers between about -5 and 5. (Setting the diffusion parameter d to zero allows for growing low frequency solutions across the whole span of zonal wavenumber.) The right panel of Figure 7.28 zooms in on these. It is important to notice that these have no counterparts among the classical Matsuno modes, so we will avoid labeling them accordingly and refer to them simply as cloud-radiation modes. These can be regarded as the manifestation of self-aggregation of moist convection in the equatorial waveguide.

Figure 7.29 shows the structure of the $n = -1, \quad k = 1$ mode and the relationship between the column moist static energy and the surface enthalpy flux. (Note that in this parameterization, the radiative heating anomaly is just a constant multiplied by the column moist static energy itself.) The structure is similar, but not identical, to the structure of the WISHE-only mode shown in Figure 7.27, and the peak radiative heating anomaly is 88% of the peak surface enthalpy flux anomaly.

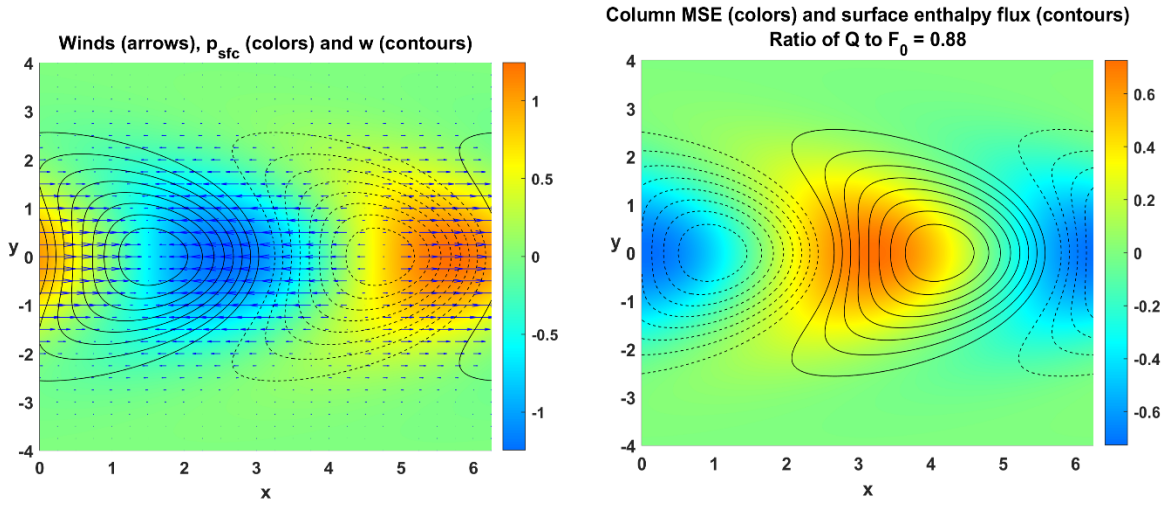


Figure 7.29: Eigenfunctions of the $n=-1, k=1$ mode with both WISHE and cloud radiation. In the left panel, the shading shows the surface pressure, the vertical velocity is in black contours with dashes indicating negative values, and zonal wind is indicated by arrows. The right panel shows the column moist static energy (shading) and the surface enthalpy flux (black contours); the cloud radiative heating anomaly is exactly in phase with the moist static energy itself. The latitude variable is nondimensional, and the x axis spans one full zonal wavelength.

The eigenfunctions shown in Figure 7.29 may be compared to the structures that emerge from a filtered analysis of the SAM aquaplanet modes shown in Figure 7.16 and 7.17. In particular, the phase relationship between wind and surface pressure (which in Figure 7.29 is just the saturation entropy with the sign flipped) is very similar to the SAM mode, while the left panel of Figure 7.17 shows that, as is assumed in the linear model, the radiative heating is very well aligned with the column moist static energy. In both SAM and the linear model, the surface enthalpy flux leads the column moist static energy anomalies, but in the linear model they are positively correlated with the column MSE, whereas in SAM the correlation is negative. Thus, WISHE is damping the SAM modes but acting to amplify the linear mode shown here. (It is possible that the linear correlation would be of the opposite sign for other combinations of parameter values.)

With some parameter combinations, the $n=-1, \quad k=1$ mode does not have the largest growth rate, and some interesting structures can appear with appreciable growth and $n>-1$. For example, Figure 7.30 shows aspects of the structure of the $n=1$ mode at zonal wavenumber 4, for the same parameters as used in Figure 7.28 and 7.29 except $C=1.5$ and $\chi=1.5$. Along the equator, the structure is approximately that of a Kelvin wave but with divergence of the meridional wind contributing negatively to the vertical velocity anomaly. Off the equator there are prominent cyclones and anticyclones manifested in both the surface pressure and wind fields.

The linear theory of equatorial waves suggests that while the WISHE effect can destabilize some of the classical Matsuno modes, cloud-radiation interactions can produce a new suite of unstable modes that have no counterparts among the Matsuno modes, including low frequency disturbances with gyres similar to equatorial Rossby waves but propagating eastward. Clearly, the modulation of radiative cooling rates by clouds and moisture is a critically important component of the physics governing many tropical phenomena.

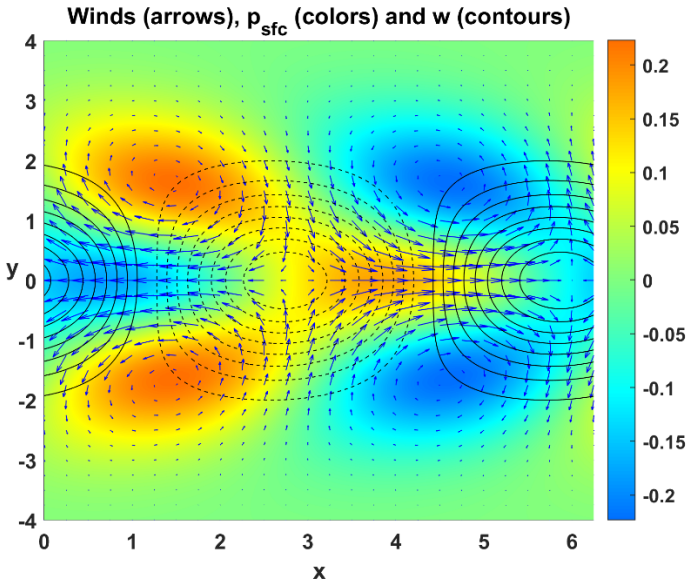


Figure 7.30: Eigenfunctions of the $n = 1, k = 4$ mode with both WISHE and cloud radiation. The shading shows the surface pressure, the vertical velocity is in black contours with dashes indicating negative values, and wind is indicated by arrows.

7.2.4.3 Effect of coupling to the stratosphere

The equatorial stratosphere, like the ocean and the troposphere, supports a spectrum of Matsuno modes, but unlike the troposphere, these are not coupled to convection or clouds or forced by surface fluxes or cloud-radiation interactions. Tropospheric disturbances can excite some of these wave modes, which carry the energy upward (Holton 1972). The disturbances may take the form of stochastic deep convective noise (Salby and Garcia 1987; Garcia and Salby 1987) or more organized tropospheric wave activity that couples to stratospheric modes with the same frequency and zonal wavelength. We focus on the latter here.

For disturbances that behave like internal gravity waves, like Kelvin and Poincare waves, the vertical group velocity increases with the wave frequency, so for these modes the stratosphere acts as a reddening filter, allowing the higher frequency modes to escape upward more rapidly than the lower frequency modes. In linear theory, this damps the growth rates of otherwise growing modes in proportion to their wavenumber (Yano and Emanuel 1991; Lin and Emanuel 2022).

For the more Rossby-like disturbances, such as equatorial Rossby waves and the westward-propagating mixed-Rossby gravity mode, the relationship between upward group velocity and wave frequency is more complex, but in general these disturbances have slower upward energy propagation and are thus better trapped in the troposphere; this is also true of the MJO (Lin and Emanuel 2022).

Another effect of coupling to the stratosphere is to excite the barotropic component in the troposphere. The response of the tropopause effectively couples the barotropic and baroclinic modes, allowing energy to be transferred from the latter to the former. For reasonable values of the stratification of the stratosphere, the barotropic component is considerably weaker than the

baroclinic wave and the superposition of the two modes gives a small eastward tilt with height of the eastward-propagating modes (Lin and Emanuel, 2022).

7.2.4.4 Effect of surface drag

Including the drag terms in linear systems like (7.4) – (7.8) also couples the baroclinic and barotropic modes and damps the growth rates. The surface drag operates directly on the boundary layer and this step-like vertical structure projects roughly equally on the barotropic and baroclinic modes. Unlike coupling to the stratosphere, surface drag damps disturbances more or less interpedently of their frequency, and so does not redden the overall spectrum. Solutions of (7.4) – (7.8) with surface drag were obtained by Lin and Emanuel (2022) and show, as expected, that the growth rates are diminished across the whole spectrum of unstable waves, but for reasonable values of the surface drag coefficient, not enough to eliminate the disturbances. As with the stratosphere coupling, the barotropic mode excited by surface drag is somewhat weaker than the baroclinic mode and leads to a small eastward tilt with height of the combined barotropic-baroclinic structure.

7.2.4.5 Excitation of higher order baroclinic modes

The vertical structure of temperature in observed Kelvin (Figure 7.10b) and Mixed Rossby-gravity (Figure 7.11c) waves, and the structure of virtual temperature in the SAM model Kelvin waves (Figure 7.21, lower left) show clear departures from the first baroclinic mode structure that would obtain if the troposphere always remained moist adiabatic. The magnitude of these higher frequency virtual temperature anomalies in the SAM simulations is around 0.5 K, a departure from moist adiabatic conditions that would be hard to detect in observations.

Khairoutdinov and Emanuel (2018) performed an experiment with an earlier version of SAM, zonally homogenizing radiative cooling rates and the wind speed that enters the formulation of surface sensible and latent heat fluxes. While several of the prominent equatorial modes, like the MJO and planetary Rossby waves, disappeared, there were still strong spectral peaks corresponding to Kelvin-like modes. While it is possible that these are weakly damped modes excited by convective noise and/or tropical cyclones and baroclinic waves at higher latitudes, one cannot rule out that these are locally generated within the tropics.

If the WISHE and cloud-radiation terms are set to zero in (7.4) – (7.8), these equations will admit growing solutions, provided the damping terms in (7.7) and (7.8) are small enough. In particular, it can be shown than for zero damping and asymptotically small $G_m / \gamma,$ the complex growth rate is $-ik + G_m / 3\gamma.$ It is not clear whether the energetics of such growing modes are internally consistent, and in any case, for reasonable values of $G_m, \gamma,$ and the damping parameters χ and $\chi_2,$ all the $v = 0$ modes have negative growth rates.

The tilted vertical temperature structure of both observed and SAM-simulated higher frequency disturbances suggests that variations in the temperature lapse rate and/or coupling between the boundary layer and the free troposphere may play an important role in these disturbances. Several theoretical explanations have been proposed. Mapes (2000) presented a simple model

of tropical atmospheric perturbations around and RCE state in which a second vertical mode is active and excited by variations in convective inhibition (CIN) brought about by stratiform rainfall. He termed the resulting instability “stratiform instability”, and the simple model is successful in re-producing the observed vertical structure of higher frequency tropical disturbances, such a Kelvin waves. Another nice aspect of this model is that the energetics of the stratiform instability are internally consistent and involve the build-up and subsequent release of moist available potential energy, proximately controlled by variations in CIN. No modulation of surface enthalpy fluxes or radiative cooling is required. Temporal variations of CAPE in the mode are consistent with observed variations, with a build-up CAPE prior to convective onset followed by a rapid diminution.

Building on this work, Kuang (2008) proposed that while the lower tropospheric virtual temperature is tightly coupled to the boundary layer moist static energy, as assumed in the first-baroclinic-mode theory described earlier in this subsection, the upper tropospheric virtual temperature is strongly influenced by lower tropospheric humidity through its effect on convective cloud buoyancy and thus depth. He showed that the stratiform-type mode identified by Mapes could equally well be excited by variations in cloud depth brought about by variations in lower tropospheric moisture as by variations in CIN, and also obtained theoretical results in good agreement with observed Kelvin waves.

While much progress has been made in defining the characteristics of equatorially trapped disturbances by increasingly refined observational technology and analysis techniques, and in theoretical understanding of the origins of these modes, a fair number of different mechanisms remain in play and more work needs to be done to improve both the understand and numerical simulation of convectively coupled equatorial disturbances.

7.3 African easterly waves

Away from the equator, the trade winds are often perturbed by disturbances that are not related to equatorially trapped waves. These are often of “synoptic” scale, with wavelengths on the order of thousands of kilometers, and generally, though not universally, travel from east to west with the Trades. Since they appear as waves in surface isobars and in ground-relative streamlines, they are often described and thought of as waves, but though we will adhere to tradition and continue to refer to these as waves, most of them are almost certainly not waves in the dynamical sense, as they are not dispersive and have closed contours of conserved variables like potential vorticity. Some of these disturbances eventually become tropical cyclones.

As early as the 1940s, forecasters concerned with hurricanes in the tropical Atlantic, including the Caribbean and Gulf of Mexico, noticed westward propagating wavy disturbances in the summertime easterly low-level flow, which they referred to as “easterly waves” and regarded as precursors to hurricanes. They noticed that the waves not only modulate low-level winds but also rainfall, with rain usually suppressed west of the trough axis and enhanced to the east, though later it was noticed that the opposite distribution sometime happens, especially in the far eastern North Atlantic. An example of an early analysis of such a wave is presented in Figure 7.31.

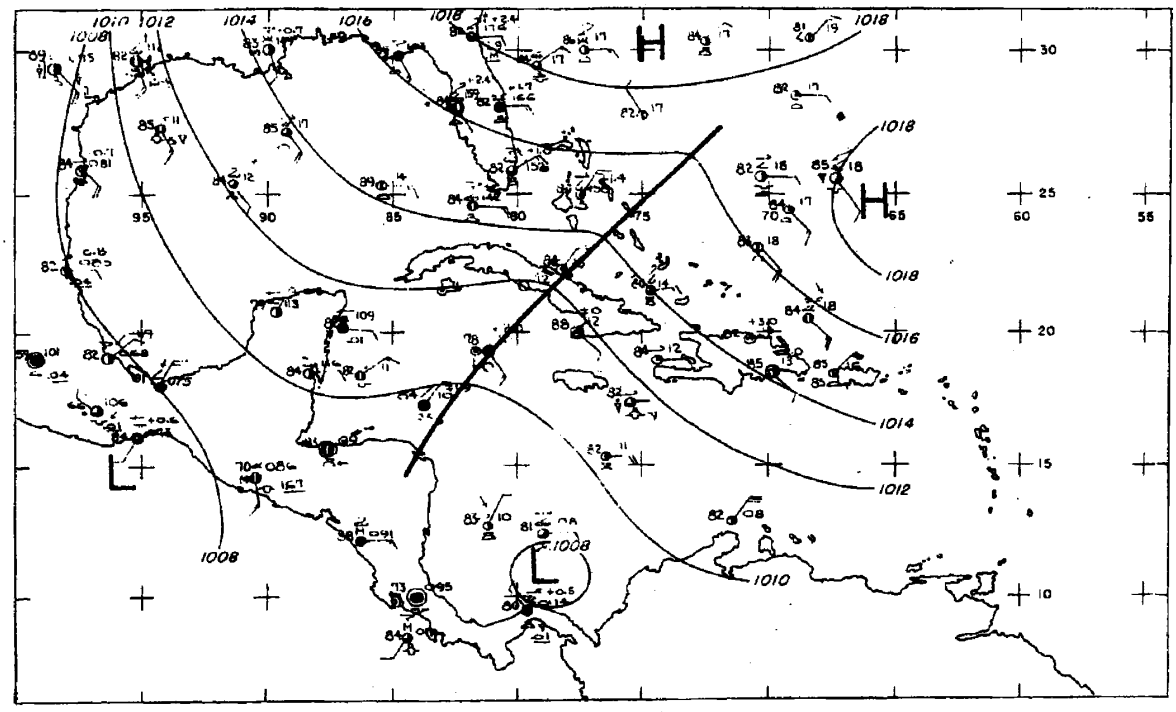


Figure 7.31: Surface weather map for September 9, 1941 at 00:30 GMT. The solid black curve shows the trough axis extending from the Bahamas to Central America. It had moved over Puerto Rico two days earlier. The circle and barbs depict weather conditions, including wind direction and speed, in standard meteorological notation. The trough is progressing northwestward.

The surface pressure and station winds indicate a general southeasterly flow, typical for the tropical North Atlantic in summer, but a prominent trough in the pressure field extends from Central America through Cuba and the Bahamas. This trough had been observed over Puerto Rico two days earlier and continued northwestward into the Gulf of Mexico.

7.3.1 Basic structure

As the radiosonde network expanded to cover parts of Africa and islands in the Caribbean, and as more surface observation sites were established, it became evident that easterly waves generally originate over central and eastern sub-Saharan Africa, and today the waves first documented in the Caribbean and Gulf of Mexico are known as African easterly waves, or "AEWs", which also distinguishes them from westward moving disturbances in the western North Pacific and elsewhere.

Today, a suite of satellite-borne instruments, and contemporary data assimilation techniques allow us to construct detailed analyses of the structure and behavior of AEWs. Figure 7.32 shows the surface structure of an AEW over the central tropical North Atlantic, as depicted by sea-surface scatterometer data (left panel) and ERA5 reanalysis data (right panel). The reanalysis, which assimilates the scatterometer data among other observations, is in good agreement with the scatterometer winds. The westward-moving AEW is marked by an obvious perturbation in the surface winds, with a pronounced relative vorticity maximum (not shown here) extending from 16°N to 20°N along 42°W. Both the surface wind speed and the rainfall show pronounced maxima north of the vorticity center, and there is a secondary maximum in rainfall east of the vortex center.

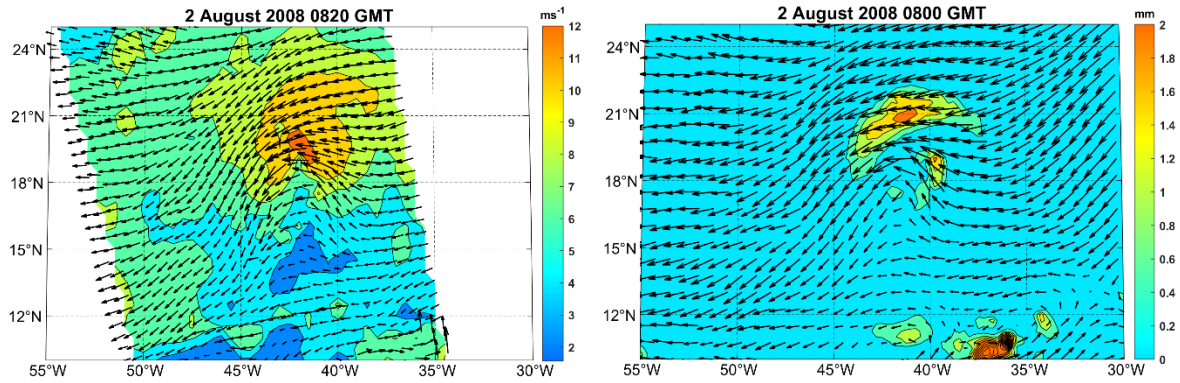


Figure 7.32: An African easterly wave (AEW) over the central tropical North Atlantic on 2 August, 2008, as revealed by sea surface scatterometer winds (left) and in ERA5 reanalysis data (right). The arrows in both panels shows the 10 m wind speed and direction, while the shading in the left panel shows the 10 m wind speed (ms^{-1}) and in the right panel shows the rainfall (mm) over the past hour.

A time-longitude section of outgoing longwave radiation and precipitation, averaged between 7.5ºN and 22.5ºN and extending from western Africa through the far western tropical North Atlantic, is shown in Figure 7.33. The time series spans all of August and September, 2004. A fairly steady progression of AEWS is evident, with periods of 3-5 days and westward migration speeds comparable to those of the low-level summertime Trades.

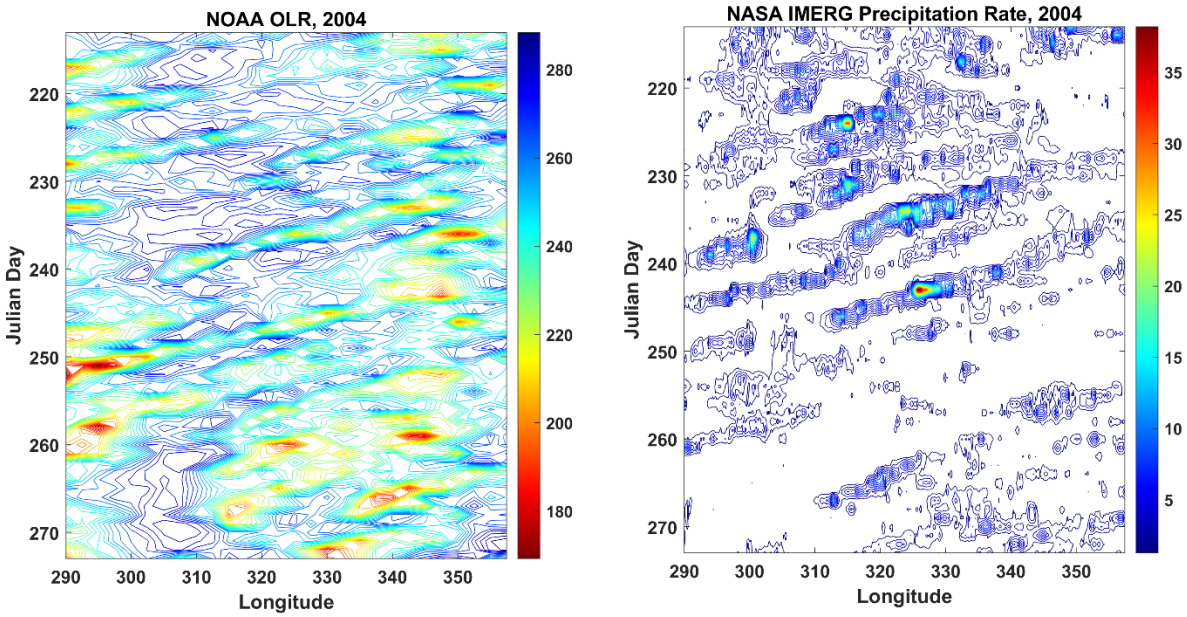


Figure 7.33: Time-longitude sections of OLR (left) and precipitation (right) averaged between 7.5ºN and 22.5ºN for August and September, 2004. The sections span the width of the tropical North Atlantic from western Africa eastward to the U.S. coastline.

The vertical structure of AEWs is illustrated in Figure 7.34, constructed using time-lagged analysis of rawinsonde data from Dakar, in Senegal on the extreme western coast of Africa, with the compositing done relative to the time of minimum AEW-filtered OLR in satellite imagery at 12.5^{o}N, 7.5 ^{o}W (Kiladis et al. 2006).

The peak lower tropospheric easterly wave winds occur between 850 and 500 hPa, and there is a node in the temperature anomalies around 600 hPa; as we shall see, this also corresponds to the altitude of the peak flow in an easterly jet that is present in the Sahel region in summer. As the disturbance center, defined by the OLR minimum, approaches, surface winds are northerly, switching to southerly about a day before the OLR minimum. But note that in the middle troposphere, around 400-700 hPa, the OLR minimum more nearly coincides with the wind shift from northerlies to southerlies. The lower tropospheric meridional winds are strongly correlated with temperature, with high temperatures associated with northerly flow from the Sahara and lower temperatures coinciding with southerly flow from the Gulf of Guinea. The peak cyclonic vorticity in the middle troposphere is 'cold core' in the lower troposphere below 600 hPa and warm core in the upper troposphere. In spite of the vertical shears of the meridional winds, there is clearly a strong barotropic component to easterly waves, and like the higher frequency equatorial modes, the vertical temperature structure shows a strong influence of the second baroclinic mode, violating the principle of strict moist adiabaticity.

In contrast with the other fields, the specific humidity anomalies are strongly coherent in the vertical, with little indication of any phase shifts with altitude. This suggests that modulation of deep convection may be the principal direct cause of the humidity anomalies.

7.3.2 Climatology

African easterly wave activity has a strong seasonal cycle, with peak activity in summer. Figure 7.35 shows the eddy kinetic energy over northern Africa and the tropical North Atlantic averaged over the months of July-September 2010-2019 at 700 and 925 hPa. The eddy kinetic energy is defined as the mean square departures of hourly winds at 00 GMT every day from their average values over the period of record. At 700 hPa, there is an east-west oriented local maximum of eddy kinetic energy between 6^{o}N and 12^{o}N extending from the Ethiopian highlands around 30^{o}E to the west coast of the continent, and then continuing into the northeastern Caribbean. This reflects mostly AEW activity at this level. But closer to the surface (bottom panel of Figure 7.35), the largest values of eddy kinetic energy are found much further north, from 18^{o}N to 24^{o}N, in the western Sahel and Sahara. This locus of activity extends southwestward, eventually overlapping with the 700 hPa activity maximum but petering out in the western tropical North Atlantic. This demonstrates the complexity of AEWs, but complexity that has come to be fairly well understood, as will be explained presently.

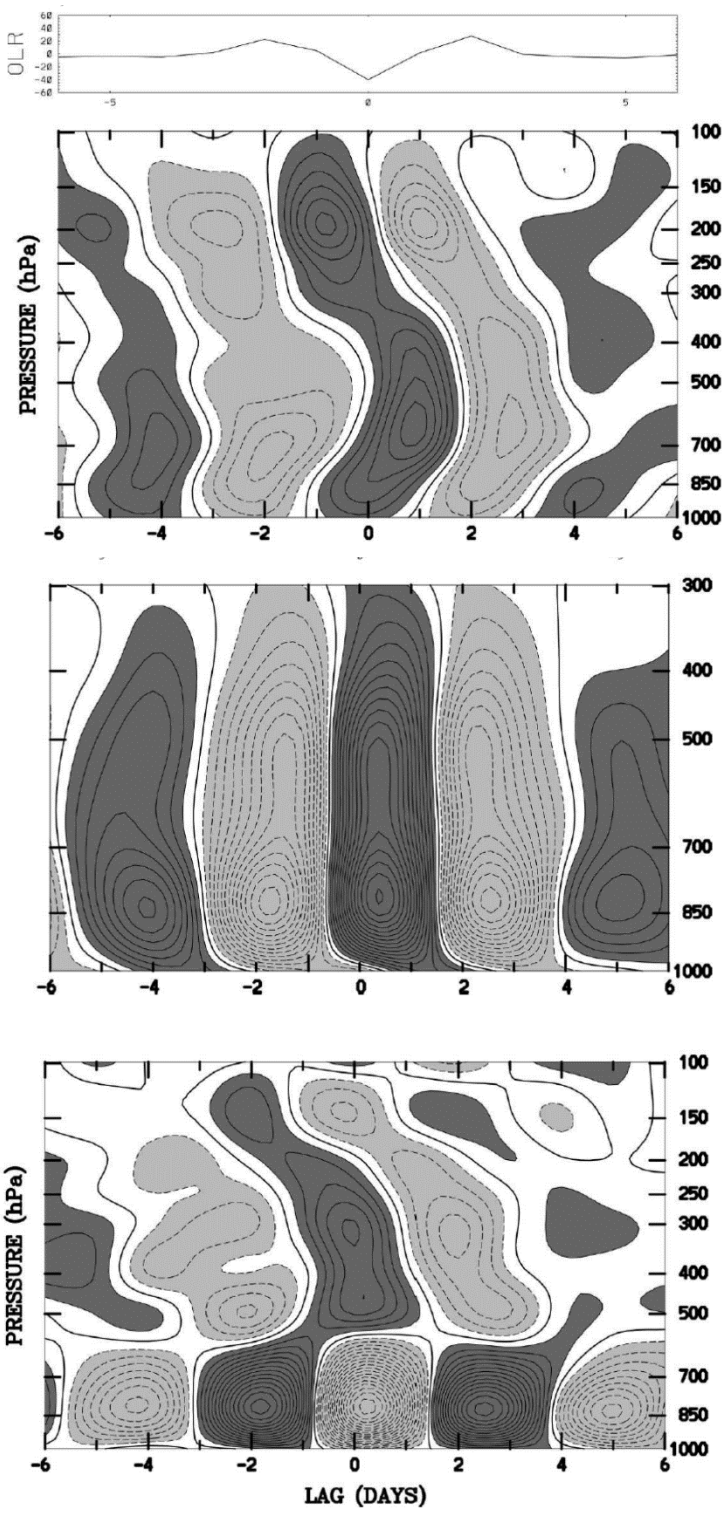


Figure 7.34: Time-lagged composite of radiosonde data at Dakar, with time relative to the minimum OLR at 12.5°N, 7.5 °W. OLR anomaly shown at top. Meridional winds (top), specific humidity (middle) and temperature (bottom), all scaled to a minimum OLR anomaly of 40 Wm^{-2}. Negative values shown by dashed contours. For the meridional wind, the contour interval is 0.5 ms^{-1} with dark (light) shading denoting anomalies greater than (less than) 0.5 ms^{-1}; for specific humidity the contour interval is 0.1 g Kg^{-1}, and for temperature the interval is 0.1 K. Note that the specific humidity plot extends upwards only to 300 hPa.

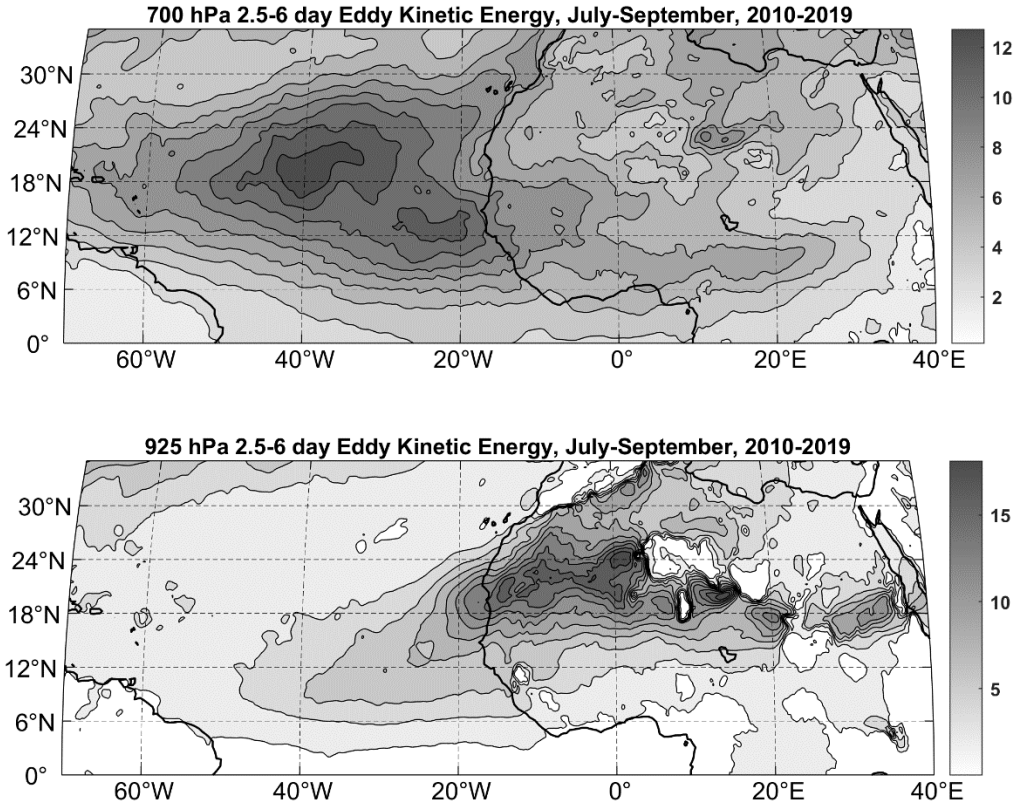


Figure 7.35: Sum of the variances of 2.5-6 day filtered zonal and meridional velocities at 700 hPa (top) and 925 hPa (bottom) averaged over July-September, 2010-2019. The variances represent the departures of hourly 00 GMT velocities from their long-term July-September mean values. From ERA5 reanalyses.

7.3.3 Time-mean background conditions

To understand the dynamics of AEWs, it is first necessary to describe the background summertime flow over northern Africa and the adjacent North Atlantic. The August mean rainfall and 10-m winds in the region are shown in Figure 7.36. A prominent belt of rainfall extends westward from eastern sub-Saharan Africa across the tropical North Atlantic, occupying latitudes from 3°N to 15°N. Over the ocean, this rain belt is marked by the convergence of southerly flow from the southern hemisphere with northeasterly Trades to the north. This circulation may be regarded as part of the solstitial Hadley cell. Over Africa, the southerlies extend much further north into the southern margins of the Sahara, marking the African summer monsoon circulation, driven in part by intense heating over the land. But note that the rain belt does not extend further north over Africa than it does over the ocean. The rain belt over land also coincides with the axis of maximum eddy activity at 700 hPa shown in Figure 7.35. A strong "Azores High" anticyclone dominates the airflow over the North Atlantic at this time of year, and the northeasterlies to the east of the anticyclone center penetrate inland over coastal Africa. A weak cyclonic circulation occupies the northern Sahara, while northerlies off the Mediterranean are found over Libya, Egypt and the Sudan.

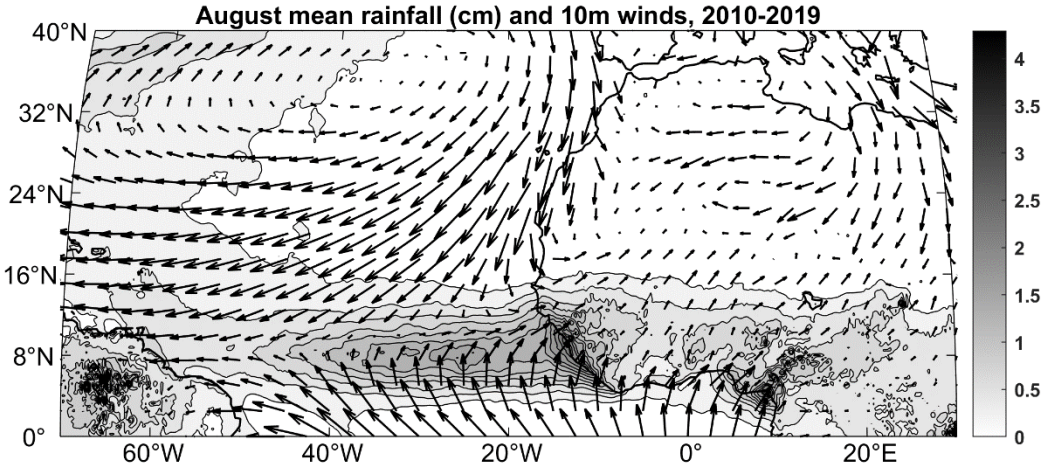


Figure 7.36: August mean rainfall (cm; shaded) and 10-m wind vectors averaged over the period 2010-2019. From ERA5 reanalyses.

Figure 7.37 shows the August mean zonal wind at 700 hPa, averaged over the period 2010-2019 (top) and the 1000 hPa potential temperature and 700 hPa Ertel potential vorticity (bottom). A strong, narrow easterly jet, whose axis lies along 16ºN, extends from far eastern tropical North Africa westward across all of the tropical North Atlantic, reaching peak intensity near the western African coast. The rain belt lies on the cyclonic (southern) flank of this African easterly jet.

The bottom panel of Figure 7.37 shows the distributions of 1000 hPa potential temperature (dashed contours) and 700 hPa Ertel potential vorticity (shading). This diagram will be central to our discussion of AEW dynamics. There is a strong ‘reverse” gradient of low-level potential temperature over sub-Saharan Africa, with temperature increasing northward. From a thermal wind balance perspective, this coincides with the strong easterly zonal wind shear below about 700 hPa. The very dry soils of the Sahara convert all of the absorbed sunlight into sensible heat, leading to very high surface temperatures and dry adiabatic lapse rates that extend as high as 600 hPa in places. Thus, the potential temperature stratification over the Sahara is essentially zero, leading to vanishingly small values of potential vorticity over the desert, clearly visible in the lower panel of Figure 7.37. This in turn leads to an elongated local maximum of potential vorticity that nearly coincides with the rain belt and extends westward to roughly 30ºW. This local potential vorticity maximum is central to the dynamics of AEWs.

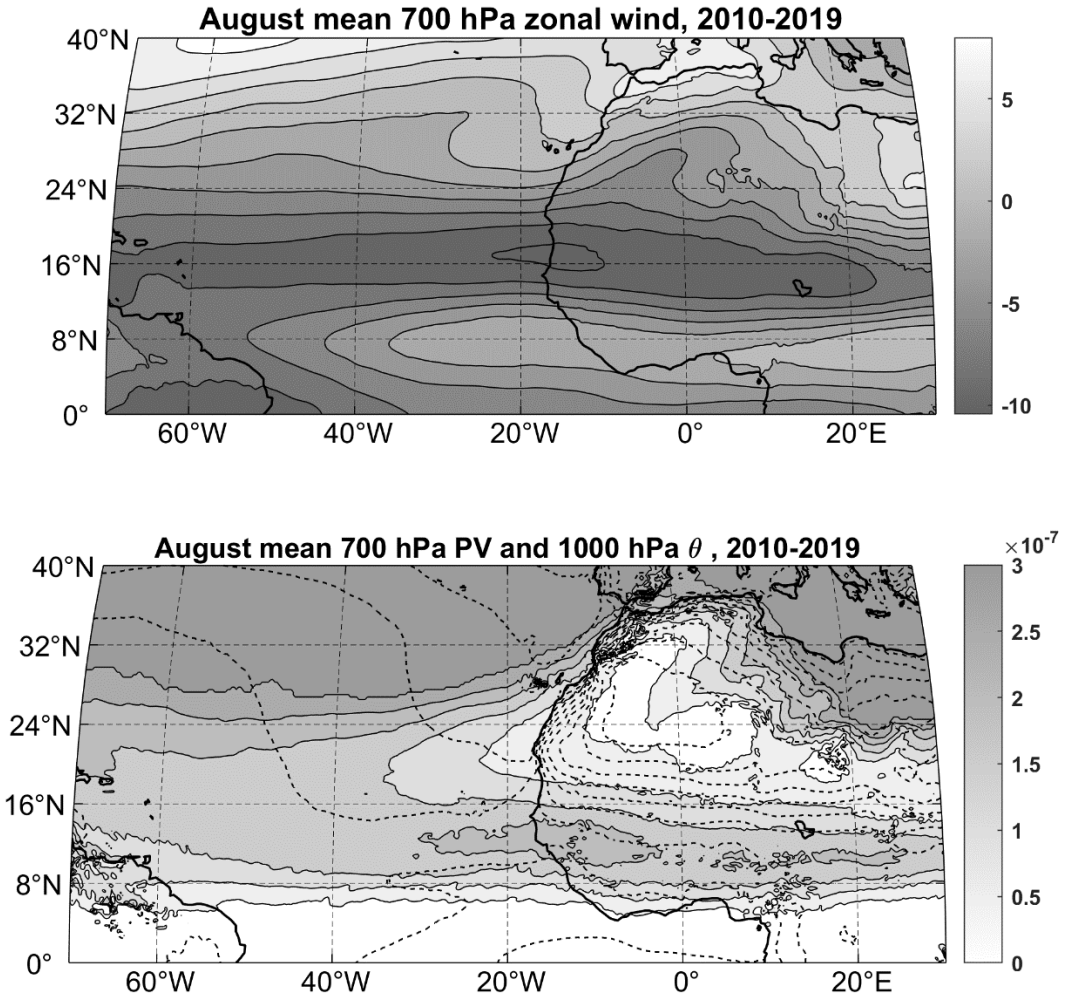


Figure 7.37: August mean 700 hPa zonal wind (ms^{-1}; top) and 700 hPa Ertel potential vorticity ($K\ Kg^{-1}\ m^2\ s^{-1}$; shading) and 1000 hPa potential temperature (K; dashed contours) at bottom. Potential temperature ranges from a minimum of 292 K over the North Atlantic off Portugal to a maximum of 314 K over the northeastern Sahara. Quantities are from ERA5 reanalyses and have been averaged over 2010-2019.

Cross-sections through the August mean climatological state, averaged from 10°W to 20°E and over the period 2010-2019, are shown in Figure 7.38. The African easterly jet is prominent, centered near 15°N and 600 hPa, with another maximum of easterly flow around the tropopause between the equator and 15°N. The mid-latitude westerly jet is in evidence near the northern boundary of the plot. The flow is broadly in thermal wind balance with the potential temperature distribution, which also reveals the region of very low stratification over the Sahara. The latter also shows up as a prominent minimum in the potential vorticity distribution (bottom panel), except very close to the surface where nocturnal cooling contributes to strong stratification in the diurnal average. Note also the strong surface potential temperature gradients between 10°N and 22°N.

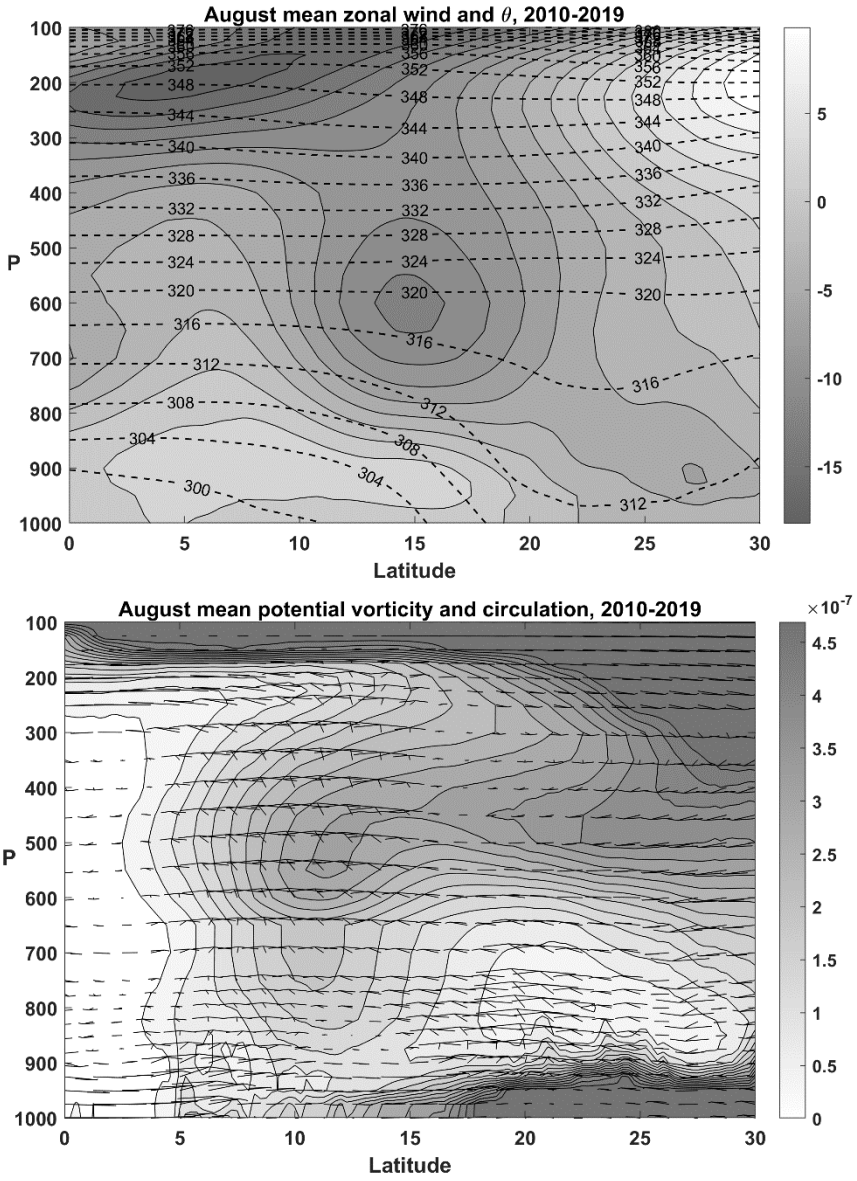


Figure 7.38: Latitude-pressure cross-sections for August mean conditions averaged between 10°W and 20°E and over the period 2010-2019. Top: Zonal wind (ms^{-1}; shading) and potential temperature (K, dashed contours). Bottom: potential vorticity ($K\ Kg^{-1}\ m^2\ s^{-1}$; shading) and vectors showing meridional wind and pressure velocity.

The potential vorticity distribution shows a prominent maximum between 5°N and 15°N, from about 900 hPa up to 500 hPa. Note the reversed potential vorticity gradient between 12°N and 18°N in the lower troposphere, while in the high troposphere the gradient is 'normal'; i.e., potential vorticity increases northward. This implies that the intrinsic (flow-relative) phase speed of Rossby waves in this range of latitude is eastward in the lower troposphere and westward in the high troposphere.

Also shown in the bottom panel of Figure 7.38 is the time-mean circulation in the latitude-pressure plane. Strong, northward monsoonal flow is evident below 900 hPa from the equator all the way up to about 20°N, with strong convergence in the rain belt around 5°N to 15°N and also between 20°N and 25°N. This leads to strong upward motion in the rain belt, especially in

the upper troposphere, and to ascent up to about 700 hPa over the Sahara. The latter ascent region is not associated with rainfall, owing to the extreme dryness of the air in this region. Much of the ascending air returns southward in the 850-700 hPa layer; this shallow, dry circulation may be thought of as a local monsoonal flow responding to the steep temperature gradient resulting from large surface sensible heat fluxes from the desert floor.

7.3.4 AEW dynamics

African easterly waves are thought to result from a barotropic/baroclinic instability of the summertime mean state of the atmosphere over the Sahel. This instability arises from the interaction of sets of Rossby waves that have opposite intrinsic phase speeds but are Doppler shifted by the mean flow in such a way that they may lock phases and amplify each other. The basic dynamics of barotropic and baroclinic instabilities is nicely summarized in Hoskins et al. (1985) and Hoskins and James (2014). Here we review an important result, the "Charney-Stern Theorem" (Charney and Stern 1962), which sets necessary (but no sufficient) conditions for steady, zonal flows to be unstable to normal mode disturbances whose intrinsic time scales are sufficiently slow that they may be considered quasi-geostrophic.

Quasi-geostrophy is founded on the basic principle of conservation of a quantity, the pseudo-potential vorticity, q_p, that is conserved following the geostrophic (not actual) wind. For quasi-geostrophic flows in the absence of friction of heating[6], this conservation principle may be written in pressure coordinates:

$$\left(\frac{\partial}{\partial t}+\mathbf{V}_g\bullet\nabla\right)q_p=0, \tag{7.17}$$

where $\mathbf{V}_g$ is the geostrophic wind and q_p is a function of a single variable, the geopotential φ:

$$q_p=\frac{1}{f_0}\nabla^2\varphi+\beta y+f_0\frac{\partial}{\partial p}\left(\frac{1}{S}\frac{\partial\varphi}{\partial p}\right), \tag{7.18}$$

where f_0 is a constant Coriolis parameter evaluated at a central latitude, β is a constant latitudinal gradient of the Coriolis parameter at the central latitude, and S is a background static stability that can be, at most, a function of pressure:

$$S(p)\equiv-\frac{\overline{\alpha}}{\overline{\theta}}\frac{d\overline{\theta}}{dp}, \tag{7.19}$$

in which the overbars refer to the mean state and α is specific volume. Since the geostrophic wind (in (7.17)) is also just a function of φ, (7.17) is a closed, nonlinear expression provided appropriate boundary conditions are supplied.

[6] Real AEWs are strongly affected by latent heating, but it is not clear that this heating is essential to their existence. Indeed, disturbances strongly resembling AEWs have been found in dry numerical simulations (Hall et al. 2006). Here, as a first look at the problem, we will neglect latent hearting.

The Charney-Stern Theorem addresses the stability of a steady, specified, balanced background flow that varies only in pressure and latitude to infinitesimal perturbations that may vary in all three spatial dimensions and time. Denoting these perturbations by primes and the background state by overbars, we express all the variables as

$$\begin{aligned} \varphi &= \overline{\varphi}(y,p) + \varphi'(x,y,p,t), \\ \mathbf{V}_g &= \overline{u}(y,p)\hat{i} + \mathbf{V}_g'(x,y,p,t), \\ q_p &= \overline{q}_p(y,p) + q_p'(x,y,p,t). \end{aligned} \tag{7.20}$$

Substituting (7.21) into (7.17), using (7.18), and dropping terms that are quadratic in the perturbation variables gives

$$\left(\frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}\right)\left[\frac{1}{f_0}\nabla^2\varphi' + f_0\frac{\partial}{\partial p}\left(\frac{1}{S}\frac{\partial\varphi'}{\partial p}\right)\right] + v_g'\frac{\partial\overline{q}_p}{\partial y} = 0. \tag{7.21}$$

Here $v_g' = \frac{1}{f_0}\frac{\partial\varphi'}{\partial x}$ is the meridional component of the perturbation geostrophic wind. Note the importance of the meridional gradient of the background pseudo-potential vorticity, that appears in the last term in (7.21).

Charney and Stern were interested in modal solutions[7] to (7.21), of the form

$$\varphi' = \Phi(y,p)e^{ik(x-ct)}, \tag{7.22}$$

where Φ is a complex function of y and p, and c is a complex phase speed. (If c has a positive imaginary part, this corresponds to solutions that grow exponentially in time.) Substituting (7.22) into (7.21) gives

$$\frac{\partial^2\Phi}{\partial y^2} + f_0^2\frac{\partial}{\partial p}\left(\frac{1}{S}\frac{\partial\Phi}{\partial p}\right) + \Phi\left(\frac{\partial\overline{q}_p/\partial y}{\overline{u}-c} - k^2\right) = 0. \tag{7.23}$$

Given suitable boundary conditions together with self-consistent distributions of $\partial\overline{q}_p/\partial y$, $\overline{u}$, and $S(p)$, (7.23) can, in principle, be solved for the eigenvalues c and associated eigenfunctions. But Charney and Stern sought instead to derive certain constraints on any normal mode instabilities that might develop in general background states. They did this through a procedure in which (7.23) is first multiplied through by the complex conjugate of Φ, Φ^*, and then integrated over a domain that is infinite in y but bounded above and below by rigid plates as specified pressures p_0 and p_1, which might here be thought of as the surface and the tropopause:

[7] Note that in this problem, the normal modes are not 'complete'; i.e., a reasonably smooth disturbance of arbitrary shape cannot in general be expressed as a weighted sum of the normal modes. The parts of any initial perturbation that do not project onto normal modes may nevertheless grow rapidly but transiently and are the subject of a large body of published literature (e.g. Farrell 1984).

$$\int_{-\infty}^{\infty}\int_{p_1}^{p_0}\left[\Phi^*\frac{\partial^2\Phi}{\partial y^2}+f_0^2\Phi^*\frac{\partial}{\partial p}\left(\frac{1}{S}\frac{\partial\Phi}{\partial p}\right)+|\,\Phi\,|^2\left(\frac{\partial\overline{q}_p}{\partial y}\frac{1}{\overline{u}-c}-k^2\right)\right]dpdy=0. \tag{7.24}$$

Next, we integrate the first two terms by parts. In doing so, we assume that the disturbances decay toward $y=\pm\infty$, but we do not assume that disturbances vanish on the top and bottom boundaries. The result is

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{p_1}^{p_0}\left\{\left|\frac{\partial\Phi}{\partial y}\right|^2+\left(k^2-\frac{\partial\overline{q}_p/\partial y}{\overline{u}-c}\right)|\,\Phi\,|^2+\frac{f_0^2}{S}\left|\frac{\partial\Phi}{\partial p}\right|^2\right\}dydp\\ +f_0^2\int_{-\infty}^{\infty}\left[\frac{\Phi^*}{S}\frac{\partial\Phi}{\partial p}\right]\Big|_{p_1}^{p_0}dy=0.\end{aligned} \tag{7.25}$$

The last term is evaluated at the upper and lower boundaries. If there are any mean state potential temperature gradients on either boundary, they can potentially be advected by the perturbation flow at the boundaries, giving rise to temperature perturbations, and, through the hydrostatic condition, to geopotential perturbations. Thus, we must use dynamical boundary conditions on potential temperature θ:

$$\left(\frac{\partial}{\partial t}+\mathbf{V}_g\bullet\nabla\right)\theta=0\quad\text{on}\ \ p=p_0,p_1 \tag{7.26}$$

There is no vertical advection on the boundaries as they are considered rigid and horizontal. Linearizing as before, using (7.20), and making use of the hydrostatic equation

$$\frac{\partial\varphi'}{\partial p}=-\frac{\overline{\alpha}}{\overline{\theta}}\theta'$$

and substituting (7.22) allows to write the linearized version on (7.26) as

$$\frac{\partial\Phi}{\partial p}-\frac{\overline{\alpha}}{f_0\overline{\theta}(\overline{u}-c)}\frac{\partial\overline{\theta}}{\partial y}\Phi=0\quad\text{on}\ \ p=p_0,p_1. \tag{7.27}$$

We substitute this into the last term in (7.25) to arrive at

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{p_1}^{p_0}\left(\left|\frac{\partial\Phi}{\partial y}\right|^2+\left(k^2-\frac{\partial\overline{q}_p/\partial y}{\overline{u}-c}\right)|\,\Phi\,|^2+\frac{f_0^2}{S}\left|\frac{\partial\Phi}{\partial p}\right|^2\right)dpdy\\ -\int_{-\infty}^{\infty}\left(\frac{\overline{\alpha}f_0\dfrac{\partial\overline{\theta}}{\partial y}}{\overline{\theta}S(\overline{u}-c)}\right)|\,\Phi\,|^2\Big|_{p_1}^{p_0}dy=0.\end{aligned} \tag{7.28}$$

We can simplify the notation in (7.28) by defining a new quantity Q_y:

$$Q_y \equiv \frac{\partial q_p}{\partial y} + \frac{\overline{\alpha} f_0}{\overline{\theta} S} \frac{\partial \overline{\theta}}{\partial y} \left[\delta(p_0 - p) - \delta(p - p_1) \right], \tag{7.29}$$

where δ is the Dirac delta function. With this definition, (7.28) may be written more compactly as

$$\int_{-\infty}^{\infty} \int_{p_1}^{p_0} \left(\left| \frac{\partial \Phi}{\partial y} \right|^2 + \left(k^2 - \frac{Q_y}{\overline{u} - c} \right) |\Phi|^2 + \frac{f_0^2}{S} \left| \frac{\partial \Phi}{\partial p} \right|^2 \right) dp dy = 0 \tag{7.30}$$

The quantity Q_y is a generalized pseudo potential vorticity gradient, with potential temperature gradients on the lower and upper boundaries acting like delta functions of positive and negative potential vorticity gradients, respectively.

Because c is, in general, a complex number, the real and imaginary parts of (7.30) must be individually satisfied. In particular, the imaginary part of (7.30) is

$$c_i \int_{-\infty}^{\infty} \int_{p_1}^{p_0} \left(\frac{Q_y}{|\overline{u} - c|^2} |\Phi|^2 \right) dp dy = 0, \tag{7.31}$$

where c_i is the imaginary part of the phase speed; positive for exponentially growing solutions. For (7.31) to be satisfied, either the solutions cannot be growing or decaying exponentially $(c_i = 0)$, or the effective gradient of pseudo potential vorticity, Q_y, must change sign at least once within the domain. Referring to its definition, (7.29), one or more of the following must be true:

1. The meridional gradient of pseudo-potential vorticity, $\partial \overline{q}_p / \partial y$, changes sign in the domain;
2. The meridional temperature gradient, $\partial \overline{\theta} / \partial y$, changes sign along one or both boundaries;
3. The meridional temperature gradient, $\partial \overline{\theta} / \partial y$, at the lower boundary (p_0) has the same sign as $\partial \overline{\theta} / \partial y$ at the upper boundary (p_1) and/or the opposite sign of the interior pseudo potential vorticity gradient, $\partial \overline{q}_p / \partial y$;
4. The meridional temperature gradient, $\partial \overline{\theta} / \partial y$, at the upper boundary (p_1) has the same sign as either or both the meridional temperature gradient at the lower boundary (p_0) and the interior potential vorticity gradient.

Equation (7.31) forms the core of the Charney-Stern theorem. In the classical Charney problem of baroclinic instability, (Charney 1947), $\partial \overline{q}_p / \partial y$ is taken to be a positive constant in the troposphere, while $\partial \overline{\theta} / \partial y$ is negative at the surface and the atmosphere is unbounded in the vertical. Thus Q_y, as given by (7.29), changes sign in the domain and the necessary (but not

sufficient) condition for instability is satisfied. In fact, the linear stability analysis does turn up unstable modes with the structure and scale of observed baroclinic waves. In the also-classical Eady linear stability analysis (Eady 1949), $\partial \overline{q}_p / \partial y$ is taken to be zero everywhere, and $\partial \overline{\theta} / \partial y$ is constant with the same value on both boundaries. Once again, Q_y changes sign within the domain and Eady's analysis does show a rich spectrum of unstable modes, with reasonable structure and length and time scales.

Taking the real part of (7.30) also provides a useful constraint:

$$\int_{-\infty}^{\infty}\int_{p_1}^{p_0}\left(\left|\frac{\partial\Phi}{\partial y}\right|^2+\left(k^2-\frac{(\overline{u}-c_r)Q_y}{|\overline{u}-c|^2}\right)|\Phi|^2+\frac{f_0^2}{S}\left|\frac{\partial\Phi}{\partial p}\right|^2\right)dpdy=0, \tag{7.32}$$

where c_r is the real part of the phase speed. Whether the solutions are growing, decaying, or neutral, (7.32) must be satisfied. Now for growing (or decaying) solutions $\left(c_i \neq 0\right)$, we are free to add any constant multiple of the double integral in (7.31) to (7.32), since the former vanishes. We choose the multiplier to be c_r, and subtracting the result from (7.32) gives

$$\int_{-\infty}^{\infty}\int_{p_1}^{p_0}\left(\left|\frac{\partial\Phi}{\partial y}\right|^2+\left(k^2-\frac{\overline{u}Q_y}{|\overline{u}-c|^2}\right)|\Phi|^2+\frac{f_0^2}{S}\left|\frac{\partial\Phi}{\partial p}\right|^2\right)dpdy=0. \tag{7.33}$$

Consequently, another constraint that exponentially growing solutions must satisfy is that $\overline{u}$ and Q_y must be positively correlated. This is a variation on a theorem by Ragnar Fjørtoft (Fjørtoft 1950).

There is a nice, straightforward interpretation of the Charney-Stern and Fjørtoft theorems. It stems from the recognition that Q_y serves as a refractive index for Rossby waves. The intrinsic (flow-relative) phase speed of such waves is the opposite sign of Q_y, so that on earth, where Q_y is usually positive in the free troposphere, Rossby waves propagate westward relative to the flow. Therefore, the requirement, from (7.31), that Q_y change sign is a requirement that the intrinsic phase speed of Rossby waves have both signs (we might refer to this as having to have two different 'flavors' of Rossby wave interacting). The condition (7.33) that $\overline{u}$ and Q_y be positively correlated can be interpreted to mean that the Doppler shifting of the two flavors of Rossby wave must be such as to make it possible for them to phase lock: the westward flavor waves $\left(Q_y > 0\right)$ must be Doppler shifted eastward, and/or the eastward flavor waves $\left(Q_y < 0\right)$ must be Doppler shifted westward.

In the classical Charney baroclinic instability problem, the two flavors consist of intrinsic westward-propagating internal Rossby waves in the troposphere, where $Q_y > 0$ and intrinsic eastward propagating waves on the surface delta function. (These latter waves decay exponentially away from the surface and are known as "Eady edge waves".) The increasing westerlies with altitude Doppler shift the internal Rossby waves eastward so that they potentially phase lock with the surface Eady edge wave, and in fact do phase lock in the unstable models.

In the classical Eady baroclinic instability problem, there are no internal Rossby waves, since Q_y is zero in the free troposphere, but there are two flavors of Eady edge wave: A surface wave with intrinsic easterly phase speed, and a tropopause edge wave with intrinsic westerly phase speed. The increasing westerly wind with altitude Doppler shifts the edge waves so that they can phase lock, and do so in the unstable modes first found by Eady.

Armed with the Charney-Stern and Fjørtoft theorems, we can now analyze whether the mean state over North Africa in summer may be subject to baroclinic/barotropic instability.

First, the bottom panels of Figures 7.37 and 7.38 show that the internal meridional gradient of potential vorticity (assumed to be like that of the pseudo-potential vorticity) changes sign across the axis of the mid-level easterly jet. Moreover, it is clear from the top panels of these two figures that the strong easterlies north of the axis of maximum potential vorticity can Doppler shift the intrinsic eastward propagating internal Rossby waves westward relative to the intrinsically westward propagating internal Rossby waves south of the axis of maximum potential vorticity. The potentially unstable interaction of these two flavors of internal Rossby waves on either side of the axis of maximum potential vorticity would normally be called barotropic instability as it need not involve temperature perturbations[8]. It is believed that this interaction is the main source of wave energy in the genesis region of AEWs in eastern and central sub-Saharan Africa, as indicated by the strip of high eddy kinetic energy at 700 hPa top part of Figure 7.35) near the axis of maximum potential vorticity.

Next, examine the bottom panel of Figure 7.38. Note that between 10^{o}N and 20^{o}N, the interior meridional potential vorticity gradient switches from negative to positive going from the lower to the upper troposphere. (There is a second reversal below 900 hPa, but that is strictly nocturnal, resulting from surface radiative cooling at night, and a night is too short to engender instabilities like these, with time scales of several days.) So we have two flavors of Rossby wave: intrinsic eastward propagation in the lower troposphere, and westward in the upper troposphere. But examination of the zonal flow (upper panel of Figure .38) shows that what little Doppler shifting exists is in the wrong direction, acting to increase the difference in actual phase speeds, so we do not expect this to be a source of wave growth.

Finally, examine the bottom panel of Figure 7.37. Note that across all of Africa, but not west of the coastline, the 700 hPa potential vorticity gradient is negative and the surface potential temperature gradient is positive roughly between 10^{o}N and 24^{o}N. This means that there are intrinsic eastward-propagating internal Rossby waves in the lower troposphere, and intrinsic westward-propagating Eady edge waves at the surface. This time, the strong easterly low-level shear Doppler shifts the two flavors of Rossby waves so that they can potentially phase lock and amplify each other…both the Charney-Stern and Fjørtoft necessary conditions are satisfied. (This is the classical Charney problem upside down!) Now refer back to the bottom panel of Figure 7.35: This unstable interaction is probably the source of all the low-level eddy kinetic energy between 10^{o}N and 24^{o}N in western Africa as well as some of the eddy kinetic energy at 700 hPa west of the Greenwich meridian (top panel of Figure 7.35).

[8] The Charney-Stern and Fjørtoft theorems are silent about any distinction between barotropic and baroclinic instability. Indeed, while the energetics shift with the relative orientations of the two flavors of Rossby wave, the dynamics are the same, and thus there is little reason to make a distinction between barotropic and baroclinic instability. Here we treat them as one and the same.

Therefore, it appears that there are two main sources of African easterly wave activity: unstable interaction of two flavors of internal Rossby wave across the axis of maximum potential vorticity in the lower to middle troposphere, extending across sub-Saharan Africa, and interaction of the northern member of this pair with surface Eady edge waves living on the strong northward gradient of potential temperature at the surface. This is perhaps why there are two distinct wave trains over Africa, which merge near the western coastline.

7.3.5. Case study with dynamical interpretation

Let's examine a particular AEW that developed during late July and early August, 2008, tracing back in time from its state as depicted in Figure 7.32. The left-hand columns of Figures 7.39-7.40 show the 700 hPa winds and potential vorticity (PV), while the right-hand columns show 1000 hPa winds and potential temperature. These charts have been created from hourly ERA5 reanalysis data at 00 GMT over the period 27 July – 1 August 2008.

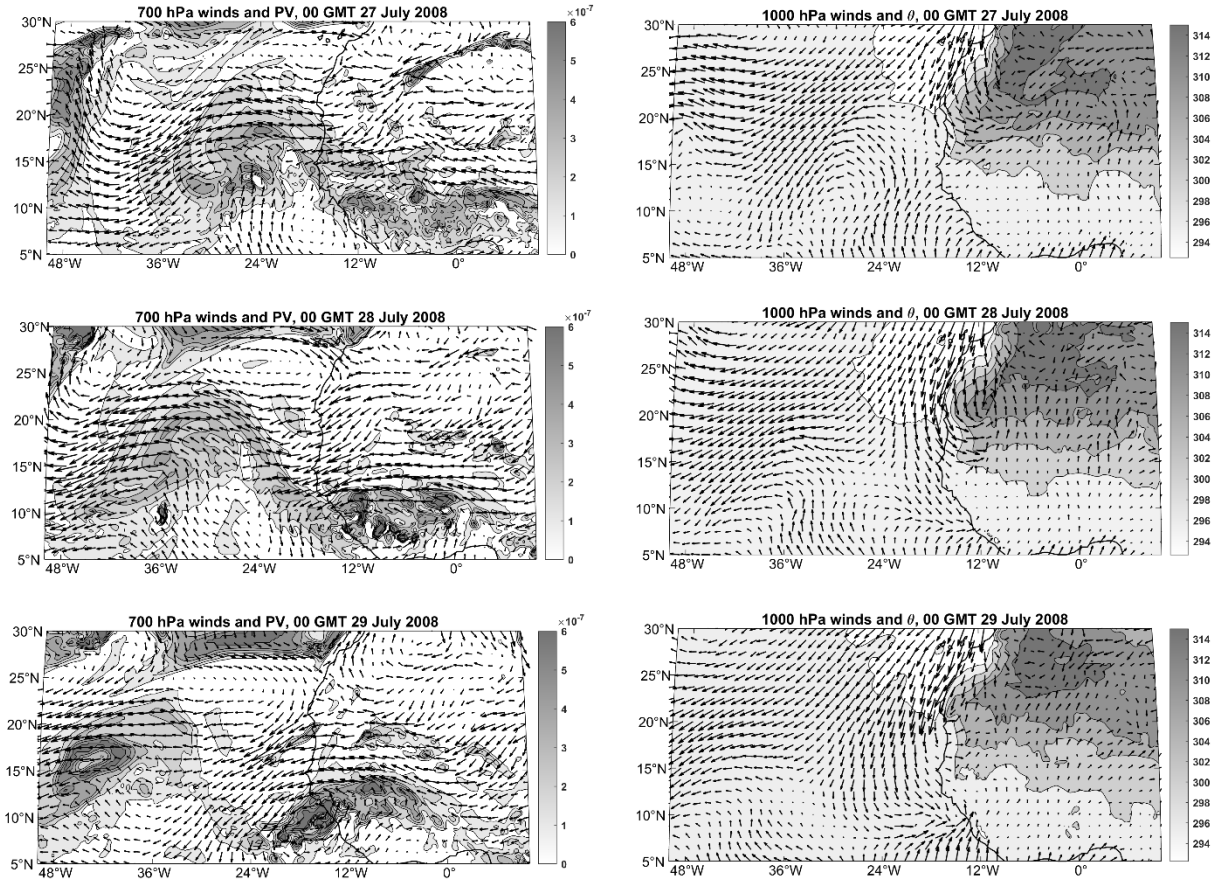


Figure 7.39: 700 hPa winds and potential vorticity (K Kg^{-1} m^2 s^{-1}; left column) and 1000 hPa winds and potential temperature (K; right column) at 00 GMT on July 27 – 29 2008, from ERA5 reanalyses.

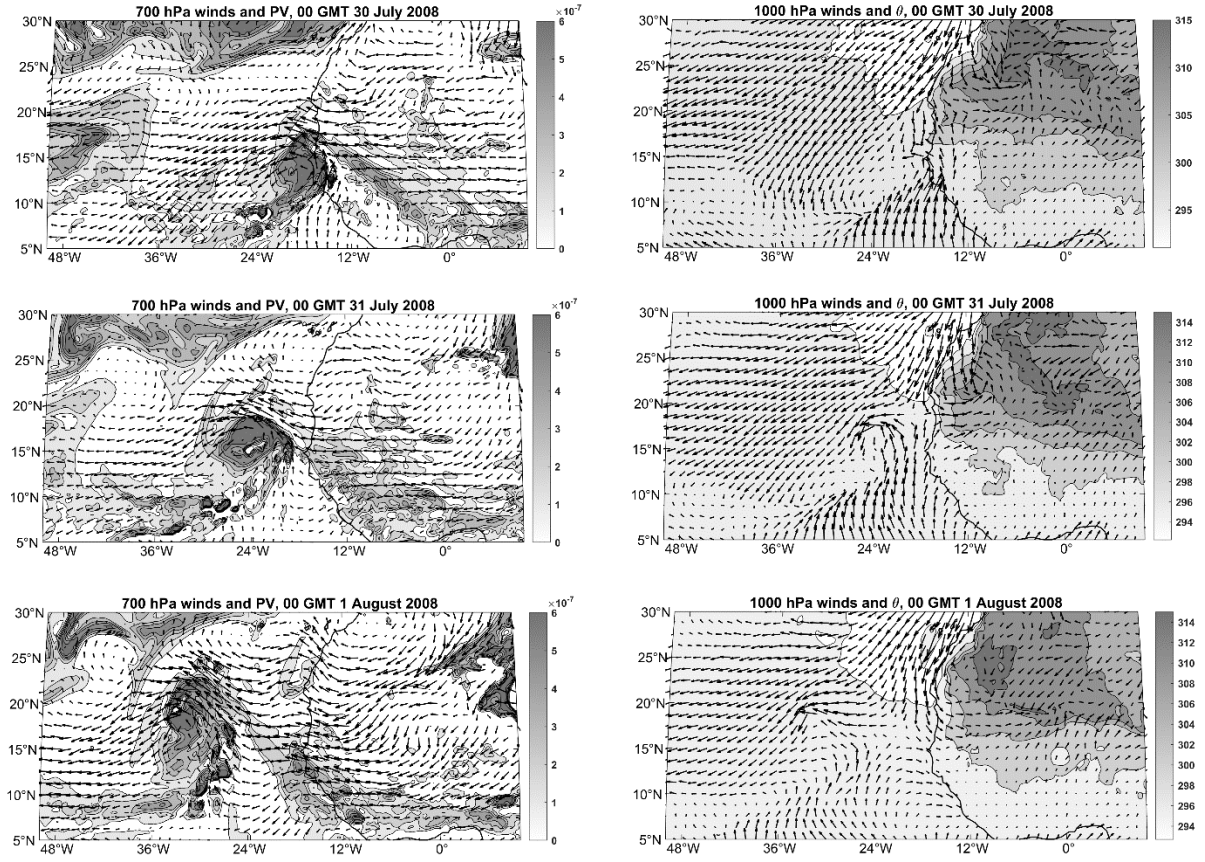


Figure 7.40: 700 hPa winds and potential vorticity (K Kg^{-1} m^2 s^{-1}; left column) and 1000 hPa winds and potential temperature (K; right column) at 00 GMT on July 30 – August 1 2008, from ERA5 reanalyses.

On July 27th (top row of Figure 7.39) one can see two AEWs at 700 hPa over the Atlantic, with strong signatures in the winds at both 700 hPa and 1000 hPa (right column). But focus on the western coastal region of Africa and note the high values of PV there, in keeping with the long-term summer climatology (bottom panel of Figure 7.37). Also notice the anomalously high PV around 8^{o}W and 10^{o}N – 15^{o}N, as well as some fragments of high PV to the northeast; these are associated with a pronounced cyclonic circulation extending to 20^{o}N. This circulation projects down to the surface (upper left panel of Figure 7.39) where it is exciting an Eady edge wave on the pronounced northerly potential temperature gradient at 1000 hPa. But note that there is little or no circulation at 1000 hPa underneath the 700 hPa high potential vorticity strip near the coast.

By the next day (middle row of Figure 7.39), a strong circulation is beginning to consolidate around 10^{o}W and there continues to be a strong cyclonic circulation around 10^{o}W and 17^{o}N. associated with the fragmented positive PV anomalies which have also moved westward. To its west, at the surface, the Eady edge wave has intensified and moved southwest from its location a day earlier. On the 29th of July (bottom panel of Figure 7.39) the PV at 700 hPa has further consolidated into a more nearly circular mass centered just offshore at around 14^{o}W, 10^{o}N, and the cyclonic circulation around it, obvious at 700 hPa, extends down to the surface where there

is now a prominent cyclonic circulation. Meanwhile, the Eady edge wave to the north is beginning to dissipate, having run out of 1000 hPa temperature gradient at the coast.

By the 30th of July (top row of Figure 7.40) the 700 hPa PV is gathered into a quite circular mass with a strong manifestation at the surface. The larger peak values of PV may be owing to diabatic process such as radiative heating under the thick deep clouds associated with this AEW. There is now a strong manifestation at the surface, with a pronounced cyclonic trough extending southwest from the African coast. The absence now of any appreciable surface potential temperature anomalies suggests that the edge wave dynamics are no longer playing any role.

The rest of Figure 7.40 documents the westward migration of the AEW, whose maximum surface manifestation is around July 31st. It is important to notice that only in its early evolution can the AEW be considered a to involve wave dynamics. Certainly by the 31st of July the AEW is far better characterized as an isolated maximum of PV in the lower to middle troposphere, just being carried westward with the Trades. In no dynamical sense is it a wave at all. Perhaps AEWs should instead be referred to as African Easterly Vortices (AEVs). This is also why the disturbances can be easily tracked across the Atlantic; were they Rossby waves they would be dispersive and not generally trackable beyond a wavelength or two.

7.4 Easterly waves over the eastern North Pacific

Westward-propagating disturbances with dimensions and time scales comparable to those of African easterly waves are observed over the tropical eastern North Pacific, west of Central America and Mexico. An example of such a wave is shown in Figure 7.41.

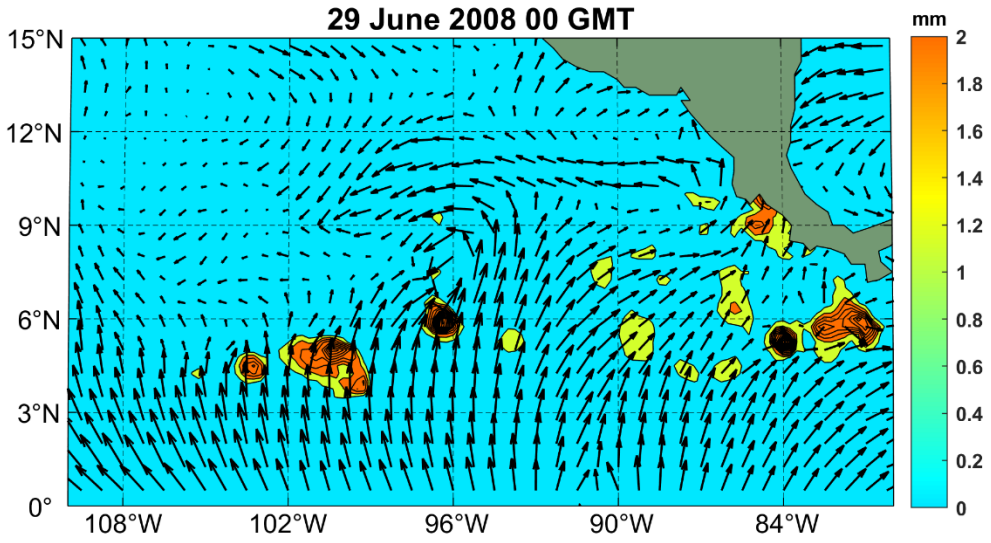


Figure 7.41: 10 m winds (arrows) and hourly precipitation amount (mm, shading) at 00 GMT on 29 June, 2008, from ERA-5 reanalyses.

The surface circulation is centered near 8°N, 96°W, and there are broken patches of rainfall along the eastern North Pacific ITCZ. At this time, there is not much rainfall associated with the wave in the easterlies. We shall examine the evolution of this particular event in greater detail later in this section.

Time-longitude sections (Figure 7.42) show a regular progression westward-propagating OLR and precipitation anomalies, similar to AEWs (Figure 7.33). The disturbances move at around 6 ms^{-1} and with a period of a few days, corresponding to a wavelength of 1,000-3,000 km. But, as is the case with AEWs, by the time these perturbations are well offshore, they are closed vortices (closed PV contours) and do not behave dynamically like waves.

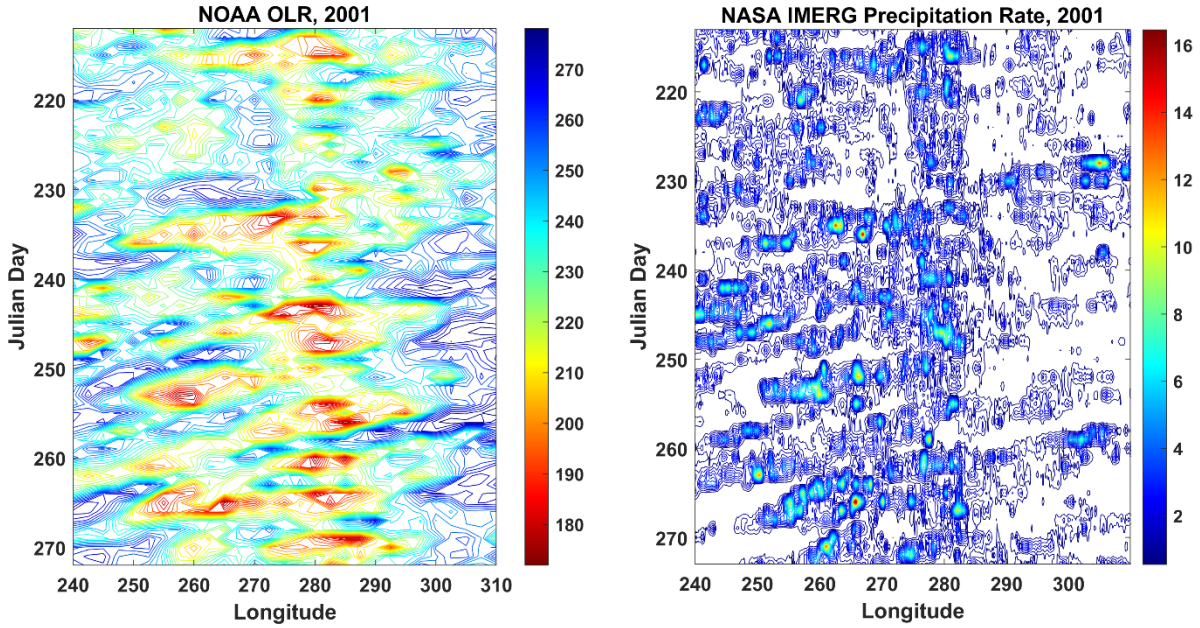


Figure 7.42: Time longitude sections of OLR (left) and precipitation (right) averaged between 5°N and 15°N for August and September, 2001. The sections span the tropical eastern North Pacific from 120°W to 60°W.

The climatological track density of easterly waves over both the North Atlantic and eastern North Pacific is shown in Figure 7.43. The tracks were computed from frequency-filtered vorticity at 850, 700, and 600 hPa from ERA-Interim reanalyses from 1998-2013, applying criterion for direction of motion, vorticity magnitude, and other criteria as described in details in Dominguez et al. (2020). Although the belt of high activity is continuous across the tropical North Atlantic, Central America, and the eastern North Pacific, there is a clear maximum of track density in the far eastern North Pacific west of Central America.

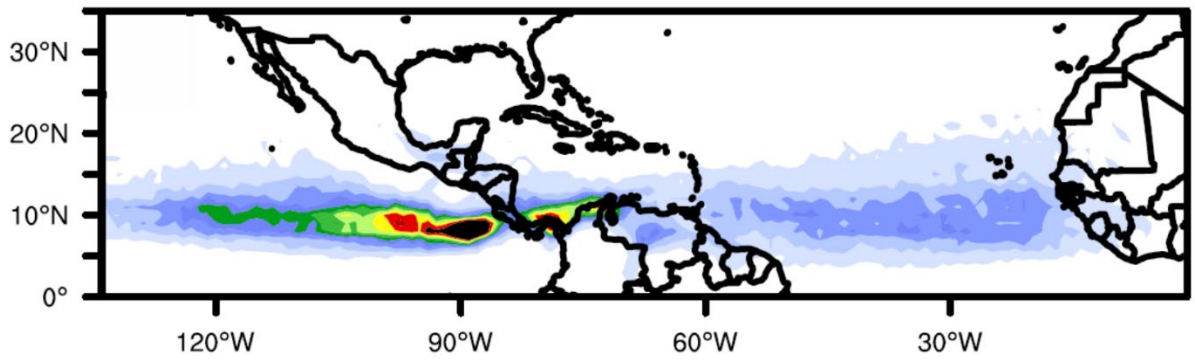


Figure 7.43: Track density of easterly waves, from reanalysis data. See text for description of tracking method.

The climatological distribution of near-surface winds and precipitation during the month of July is displayed in Figure 7.44. Low-level conditions differ dramatically between the Caribbean region and the eastern North Pacific. Strong low-level easterlies dominate the Caribbean Sea in July, with almost no rain, while the eastern Pacific has a prominent ITCZ extending from 5°N to 10°N. The airflow within and south of the ITCZ is from the south, with very light surface winds to the north, giving way to northeasterly Trades in the western portion of the domain. Note also the high rainfall over the Andes and just west of Lake Maracaibo in northern Venezuela; there is also intense convective rainfall over Panama and Costa Rica.

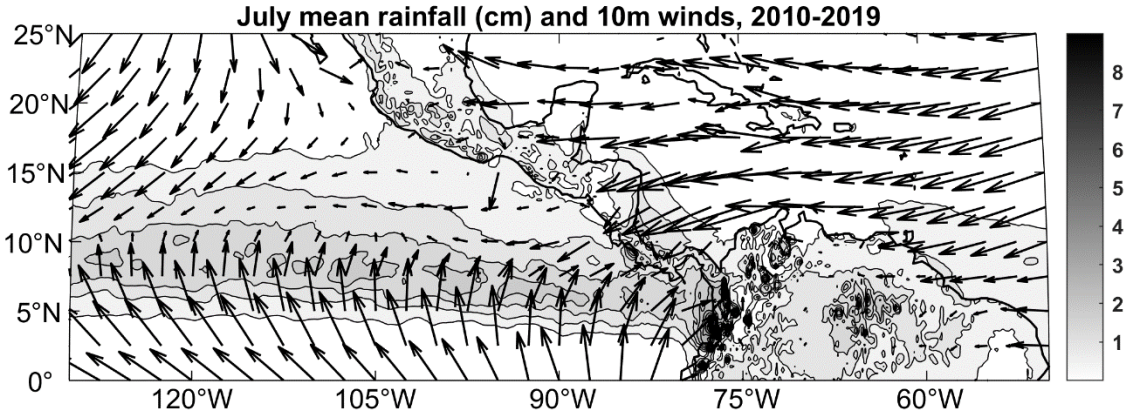


Figure 7.44: July mean surface (10m) winds (arrows) and rainfall (cm; shading) averaged over 2010-2019 from ERA-5 reanalyses.

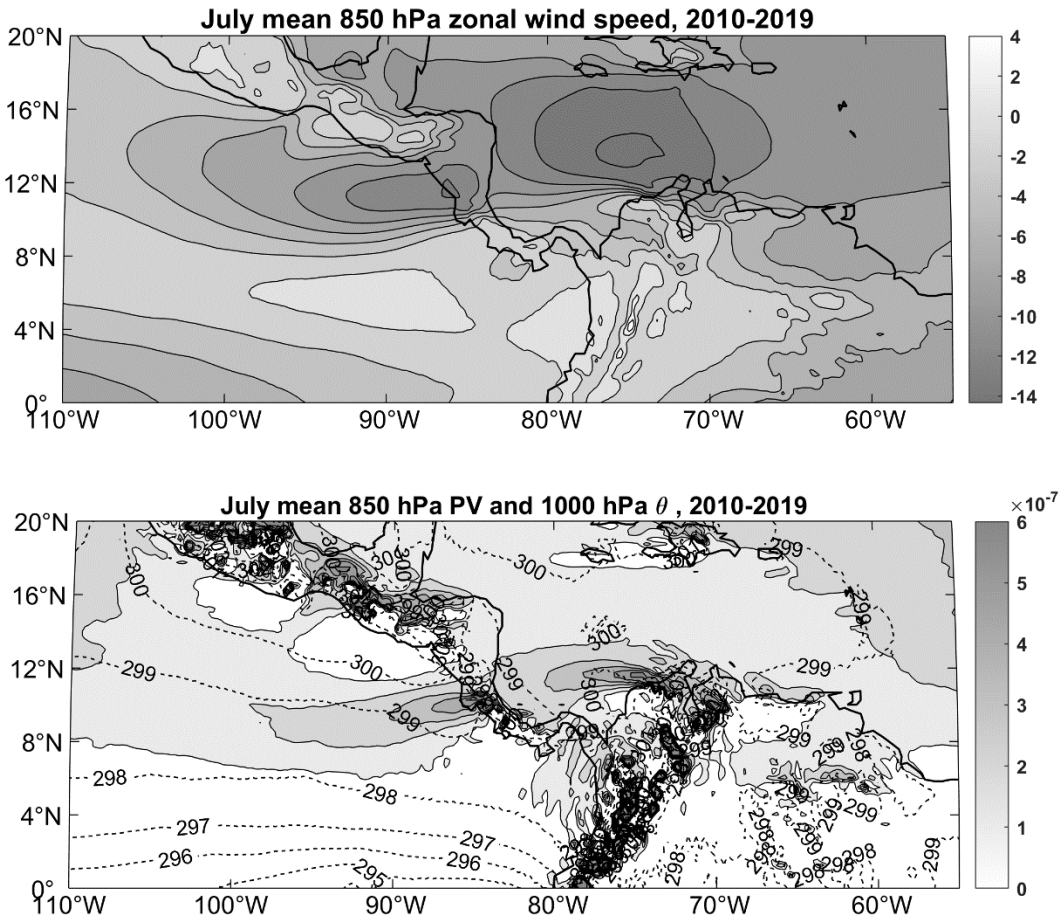


Figure 7.45: July mean 850 hPa zonal wind (ms^{-1}; top), and 850 hPa Ertel potential vorticity ($K\ Kg^{-1}\ m^2\ s^{-1}$; shading) and 1000 hPa potential temperature (K; dashed contours) at bottom. Quantities are from ERA5 reanalyses and have been averaged over 2010-2019.

July-mean zonal winds and potential vorticity at 850 hPa are shown in Figure 7.45 along with 1000 hPa potential temperature. There is a strong low-level easterly jet over the southern Caribbean Sea, reaching peak values north of Venezuela, around 76°W, and a secondary maximum over and west of Nicaragua. These jets are associated with two prominent plumes of high potential vorticity emanating from just west of Lake Maracaibo in Venezuela and from a region near the border between Nicaragua and Costa Rica. Both are regions of large convective rainfall and very high lightning density, suggesting that the high PV plumes may be generated by latent heat release associated with the strong convective storms in those regions. The region near the Catatumbo River, west of Lake Maracaibo, has the highest density of lightning anywhere on the planet (Albrecht et al. 2016). The nearly continuous lightning display is known as the "Relámpago del Catatumbo" and has been used as a navigation beacon by fishermen on Lake Maracaibo[9]. The more westerly plume also originates a in a region of high convective rainfall in summer.

[9] This highly persistent electrical storm was mentioned by the Spanish poet Lope de Vega in his 1597 epic poem *La Dragoneta*, describing the failed attack of Francis Drake on the Maracaibo region (Bürgesser et al. 2012).

While there is hardly any low-level temperature gradient over the Caribbean in summer, there is a strong, northward-directed temperature gradient over all of the easterly wave-prone region of the eastern North Pacific. This is partly owing to the cold water near the equator cause by the climatological equatorial upwelling in that region.

July-mean cross-sections of zonal wind, absolute vorticity, and circulation in the meridional plane, averaged between 70^{o}W and 80^{o}W and averaged between 2010 and 2019, are shown in Figure 7.46.

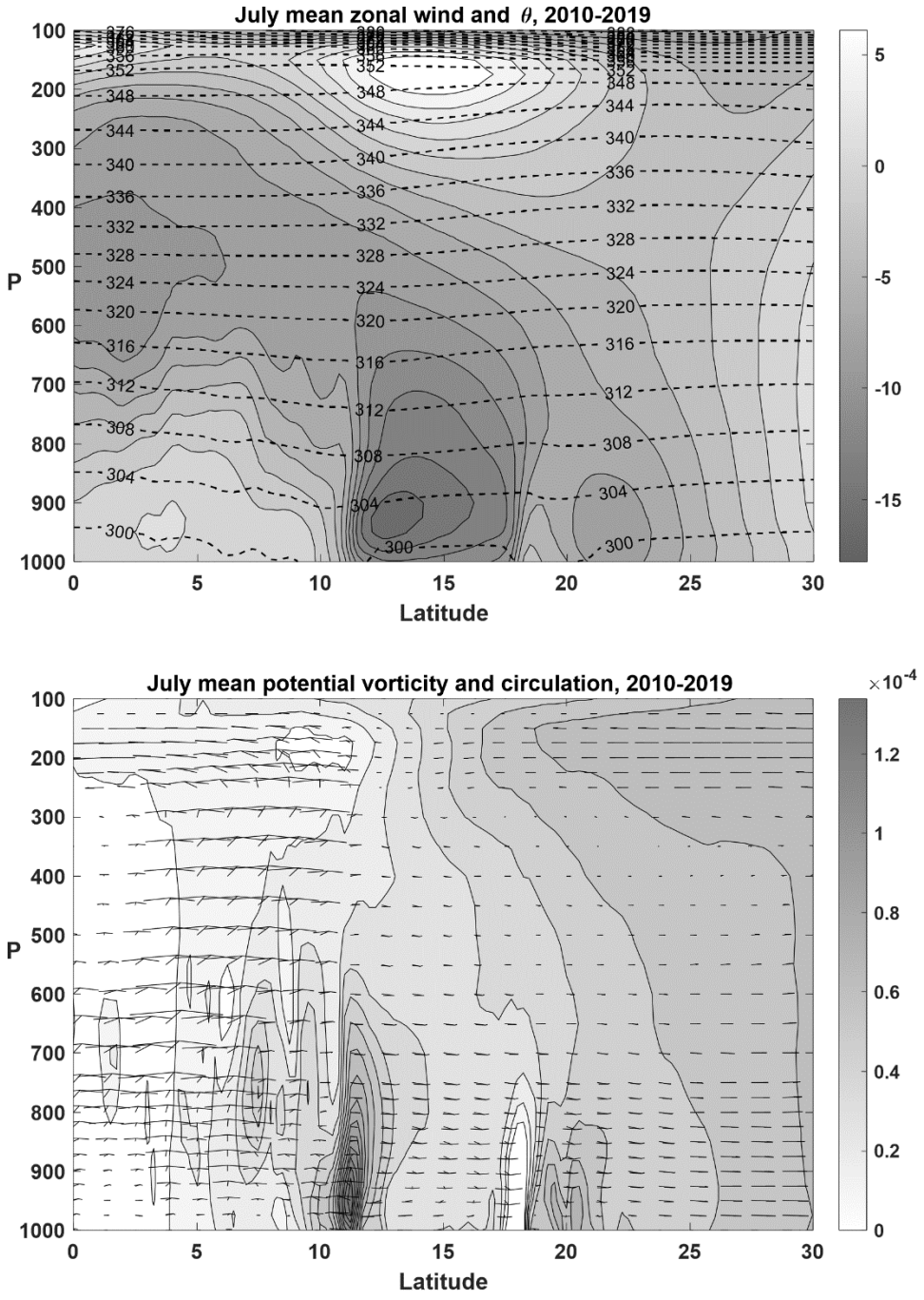


Figure 7.46: Latitude-pressure cross-sections for July mean conditions averaged between 80°W and 70°W and over the period 2010-2019. Top: Zonal wind (ms^{-1}; shading) and potential temperature (K, dashed contours). Bottom: absolute vorticity (s^{-1}; shading) and vectors showing meridional wind and pressure velocity.

(Here we display absolute vorticity rather than PV, since, owing to the high topography of the northern Andes, the PV field is very noisy over the continent.)

The low-level Caribbean Jet is strongly concentrated between the surface and 800 hPa – somewhat lower than the African easterly jet – and its associated cyclonic vorticity is also highly concentrated in latitude, as is also evident in Figure 7.45. There is strong ascent, particularly at

high levels, over South America, between 3ºN and 10ºN; elsewhere, the meridional circulation in this plane is quite weak, but with descending motion in the lower troposphere north of the jet.

From a dynamic standpoint, the two strips of high potential vorticity air in the lower troposphere, evident in Figure 7.45, are highly unstable to internal baroclinic/barotropic instability, owing to the sign reversal of the PV gradient coupled with the strong easterly flow north of the strips. This is very similar to the conditions over central sub-Saharan Africa in summer, except that the jet is much lower and more concentrated in latitude. Indeed, several studies of the energetics of eastern North Pacific easterly waves shows that, initially, they are powered mostly by barotropic conversion of the kinetic energy of the easterly jet into eddy kinetic energy associated with the easterly wave disturbances (e.g. Rydbeck, A.V. and Maloney, E. D. 2014).

The westernmost of the two high PV plumes evident in Figure 7.45 overlies a region of monotonic, northward-directed potential temperature gradient at the surface, thus supporting westward-propagating Eady edge waves at the surface. This is similar to the conditions over the western part of sub-Saharan Africa in summer. At the same time, the decreasing PV north of the PV maximum supports eastward (flow-relative) Rossby waves, and the strong easterlies aloft can Doppler shift those waves to phase lock with the surface Eady edge wave. This, in addition to the internal jet instability, can lead to wave growth. This circumstance is quite different from the conditions over the far western North Atlantic, off Africa, where there is little surface air temperature gradient to support Eady edge waves.

As with the AEWs, the dynamics of eastern North Pacific easterly waves is complicated by moist processes, and the very warm sea surface temperatures of the region coupled with low tropopause temperature, favor deep convective clusters whose cold, high clouds can trap infrared radiation and promote self-aggregation of convection. As the disturbances move westward, wind-induced surface fluxes (WISHE) may also begin to contribute to the development of the disturbances as they morph into tropical cyclones.

A case study of particular easterly wave system is displayed in Figures 7.47 – 7.49. Figures 7.47-7.48 show the 850 hPa PV and winds (left columns) and 1000 hPa potential temperatures and winds (right columns). Figure 7.49 shows the Cloud Archive User Data Service (CLAUS) brightness temperatures (10.5-12.5-micron radiances) corresponding to the dates and times in Figures 7.47 – 7.48. Note that the map frame shifts from one date to the next so as to keep the system being analyzed more or less in the center of the domain.

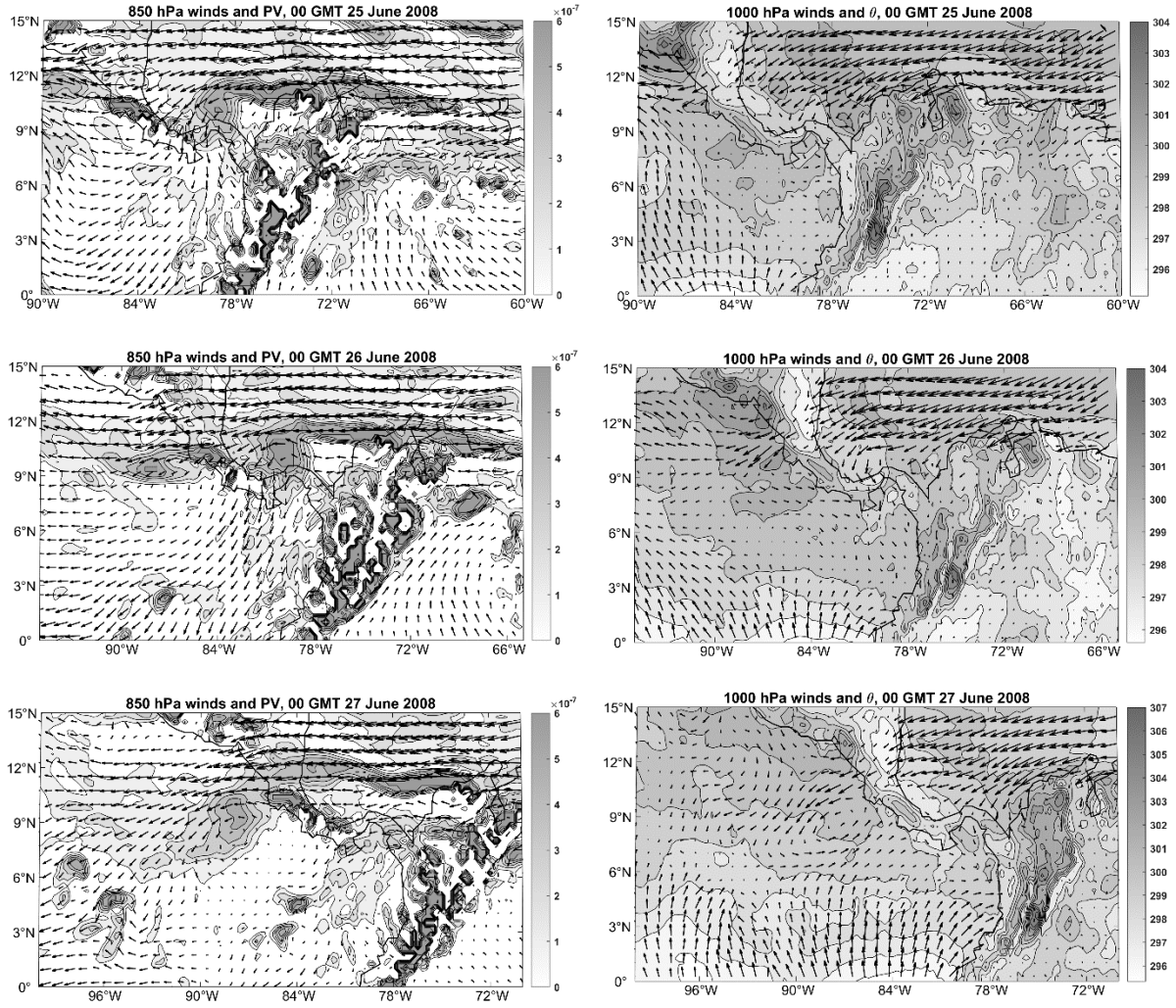


Figure 7.47: 850 hPa winds and potential vorticity (K Kg^{-1} m^2 s^{-1}; left column) and 1000 hPa winds and potential temperature (K; right column) at 00 GMT on June 25 – 27 2008, from ERA5 reanalyses. Note that map frame moves (more or less) with the easterly wave system being analyzed.

On June 25th (top row of Figure 7.47), a thin strip of high PV at 850 hPa is just beginning to roll up at its western end, just north of Panama, accompanied by a strong turning of the flow from easterly to northerly. This is likely a result of the mostly barotropic instability of the Low-Level Caribbean Jet. Note also a second mass of high PV along the Costa Rica-Nicaragua border, around 10ºN, 95ºW. The flow at 1000 hPa is not greatly different from climatology (Figure 7.44).

The roll up of the western end of the PV plume continues on June 26th, while the second plume has stretched westward into the eastern Pacific and is beginning to roll up, with a circulation beginning to develop underneath it at 1000 hPa.

By June 27th, the roll up of the western PV streamer has progressed, with a well-defined maximum near 10ºN, 89ºW, accompanied by a closed circulation at 1000 hPa. The surface circulation also has a weak maximum of potential temperature owing to advection by the northerly flow that preceded the system. At this time there is little by way of thick, high clouds associated with the surface circulation or PV aloft (Figure 7.49).

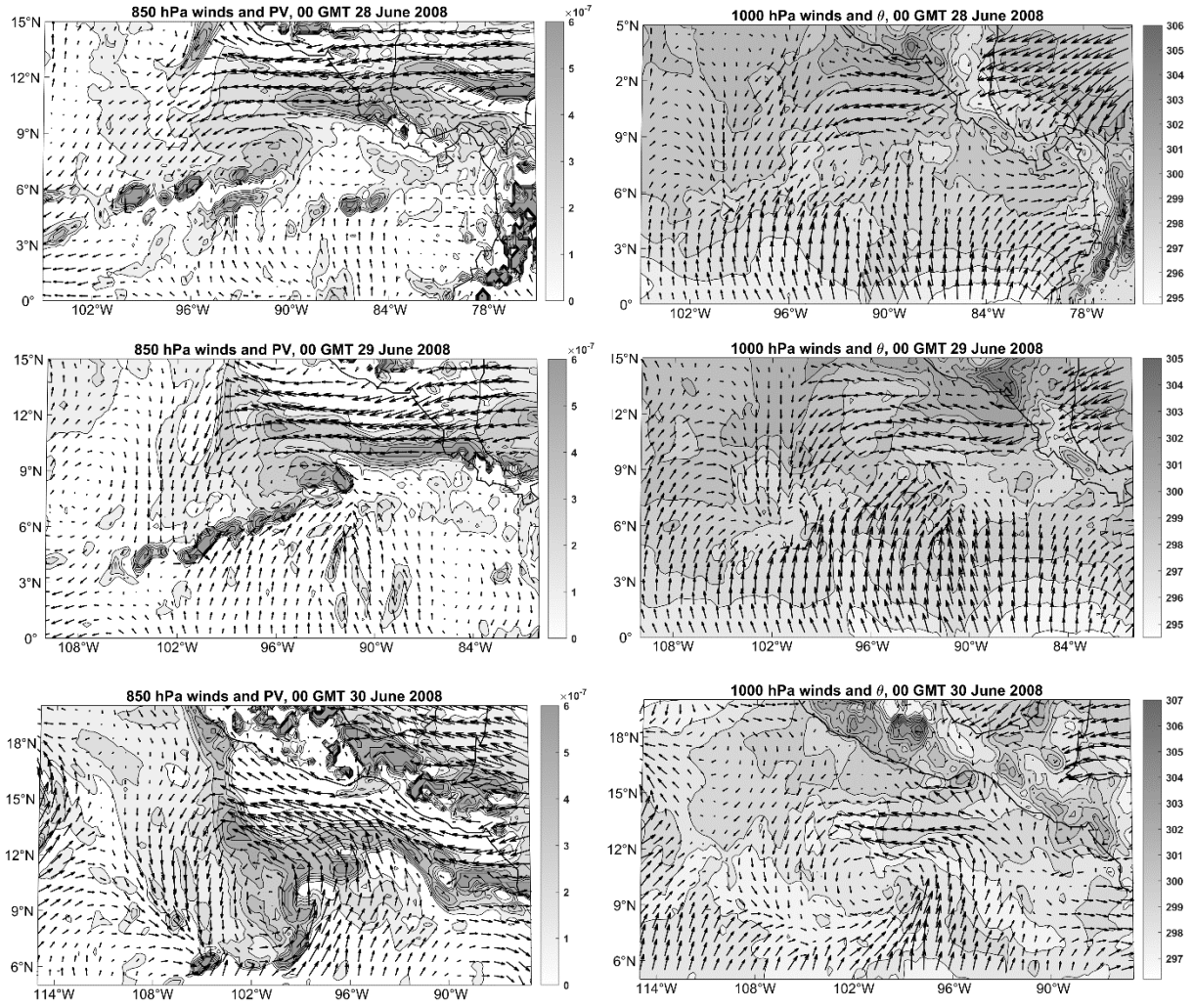


Figure 7.48: 850 hPa winds and potential vorticity (K Kg^{-1} m^2 s^{-1}; left column) and 1000 hPa winds and potential temperature (K; right column) at 00 GMT on June 28 – 30 2008, from ERA5 reanalyses. Note that map frame moves (more or less) with the easterly wave system being analyzed.

On June 28th the system is continuing to develop while moving west-northwestward. It has still not detached from the prominent PV streamer emanating from Central America and it appears to be ingesting high PV air associated with convection along the eastern Pacific ITCZ (high PV blobs west-southwest of the system in the upper left panel of Figure 7.48). The circulation at 1000 hPa is broadening and beginning to interact with the southerly airflow south of the ITZC, and the brightness temperature shows a cluster of thick high cloud co-located with the cyclonic system (Figure 7.49).

During the following two days, June 29th and 20th, the system remains attached to the strong PV streamer emanating from Central America and in fact does not become detached from it until July 2nd (not shown here). The apparent convectively produced, localized masses of high PV air continue to be ingested into the system, which intensifies while its canopy of thick, high cloud expands. This system went on to become Tropical Storm Douglas on July 2nd 2008.

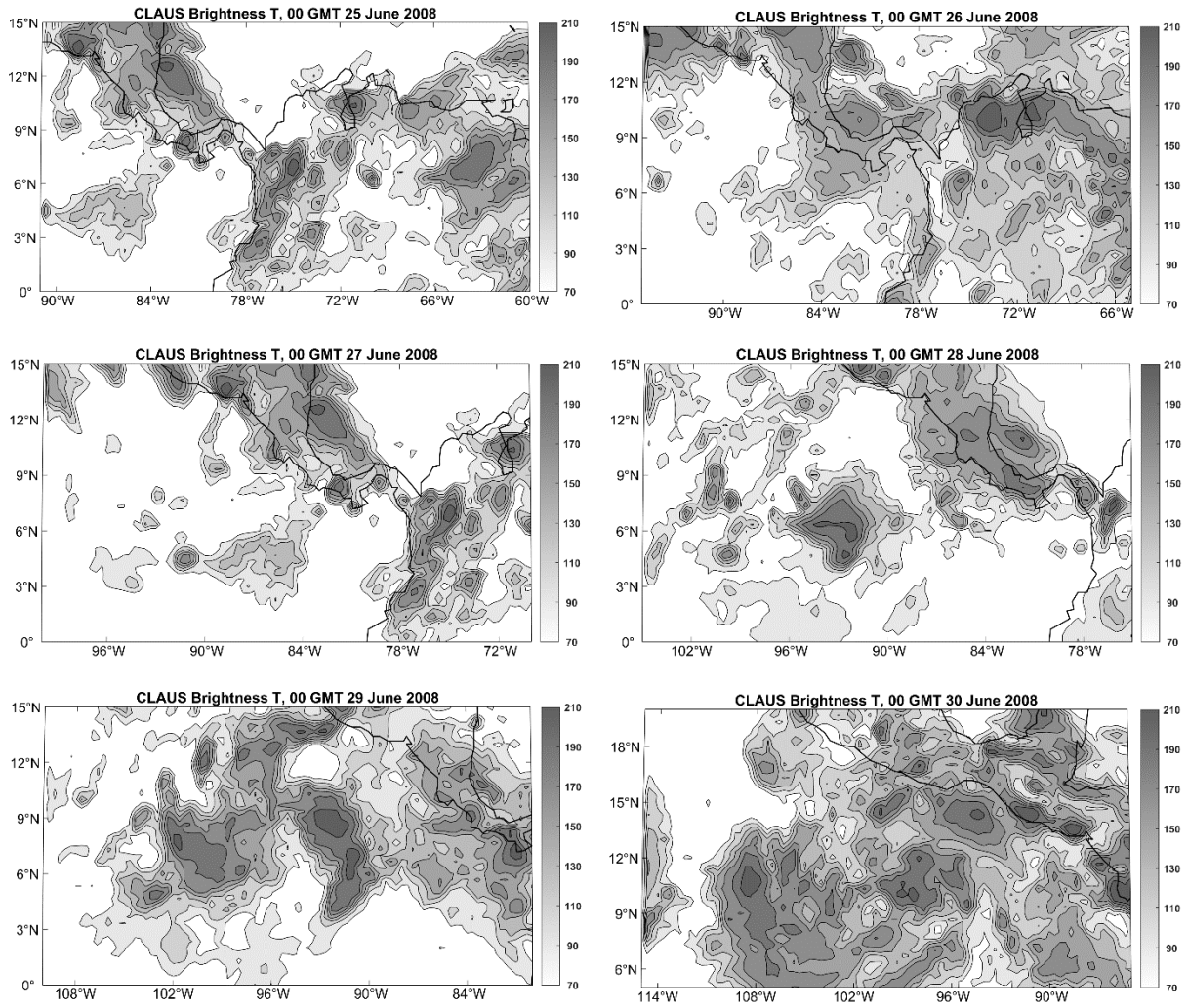


Figure 7.49: Cloud Archive User Data Service (CLAUS) brightness temperature (10.5-12.5-micron radiances) at 00 GMT on June 25 – 30 2008, read from upper left to lower right. Map frames are identical, at each date, to those used in Figures 7.47 – 7.48.

The evolution of this eastern North Pacific easterly wave is complex, with two prominent, low-level PV streamers, the surface potential temperature gradient over the e astern Pacific, the ITCZ, and convectively produced PV anomalies all appearing to play some role. As with AEWs, wave dynamics do not appear to be very important after the initial roll-up of the PV streamers, although, unlike the case of AEWs, Eady edge wave dynamics may play some role through June 29th. Perhaps more so than in the case of AEWs, diabatic processes seem influential from very early in the development of the disturbances, and the relative roles of barotropic/baroclinic instability and diabatic PV generation my depend on the particular phase of low frequency, equatorially trapped disturbances (Rydbeck, A.V. and Maloney, E. D. 2014).

The physics underlying intraseasonal and synoptic-scale variability in the tropics remains a fascinating and consequential research topic, involving processes peculiar to the tropics as well as variants of classical, mid-latitude dynamics.

References

Albrecht, R. I., S. J. Goodman, D. E. Buechler, R. J. Blakeslee, and H. J. Christian, 2016: Where are the lightning hotspots on earth? *Bull. Am. Meteorol. Soc.*, **97**, 2051–2068, https://doi.org/10.1175/BAMS-D-14-00193.1.

Bartana, H., C. I. Garfinkel, O. Shamir, and J. Rao, 2023: Projected future changes in equatorial wave spectrum in CMIP6. *Clim. Dyn.*, **60**, 3277–3289, https://doi.org/10.1007/s00382-022-06510-y.

Bony, S., and K. Emanuel, 2005: On the role of moist processes in tropical intraseasonal variability: cloud-radiation and moisture-convection feedbacks. *J Atmos Sci*, **62**, 2770–2789.

Bürgesser, R. E., M. G. Nicora, and E. E. Ávila, 2012: Characterization of the lightning activity of "Relámpago del Catatumbo." *J. Atmospheric Sol.-Terr. Phys.*, **77**, 241–247, https://doi.org/10.1016/j.jastp.2012.01.013.

Charney, J. G., 1947: The dynamics of long waves in a westerly baroclinic current. *J Meteor*, **4**, 135–163.

——, and M. E. Stern, 1962: On the Stability of Internal Baroclinic Jets in a Rotating Atmosphere. *J. Atmospheric Sci.*, **19**, 159–172, https://doi.org/10.1175/1520-0469(1962)019<0159:otsoib>2.0.co;2.

Chavas, D. R., and K. A. Reed, 2019: Dynamical aquaplanet experiments with uniform thermal forcing: System dynamics and implications for tropical cyclone genesis and size. *J. Atmospheric Sci.*, **76**, 2257–2274, https://doi.org/10.1175/JAS-D-19-0001.1.

Dominguez, C., J. M. Done, and C. L. Bruyère, 2020: Easterly wave contributions to seasonal rainfall over the tropical Americas in observations and a regional climate model. *Clim. Dyn.*, **54**, 191–209, https://doi.org/10.1007/s00382-019-04996-7.

Eady, E. T., 1949: Long waves and cyclone waves. *Tellus*, **1**, 33–52.

Farrell, B., 1984: Modal and non-modal baroclinic waves. *J. Atmospheric Sci.*, **41**, 668–673, https://doi.org/10.1175/1520-0469(1984)041<0668:MANMBW>2.0.CO;2.

Fjørtoft, R., 1950: Application of integral theorems in deriving criteria of stability for laminar flows and for the baroclinic circular vortex. *Geofys Publ*, **17**, 1–52.

Garcia, R. R., and M. L. Salby, 1987: Transient response to localized episodic heating in the tropics. Part II: Far-field behavior. *J. Atmospheric Sci.*, **44**, 499–532, https://doi.org/10.1175/1520-0469(1987)044<0499:TRTLEH>2.0.CO;2.

Hall, N. M. J., G. N. Kiladis, and C. D. Thorncroft, 2006: Three-dimensional structure and dynamics of African easterly waves. Part II: dynamical modes. *J. Atmospheric Sci.*, **63**, 2231–2245, https://doi.org/10.1175/JAS3742.1.

Hayashi, Y., 1982: Space-time spectral analysis and its applications to atmospheric waves. *J. Meteorol. Soc. Jpn. Ser II*, **60**, 156–171, https://doi.org/10.2151/jmsj1965.60.1_156.

Holton, J. R., 1972: Waves in the equatorial stratosphere generated by tropospheric heat sources. *J. Atmospheric Sci.*, **29**, 368–375, https://doi.org/10.1175/1520-0469(1972)029<0368:WITESG>2.0.CO;2.

Hoskins, B., and I. James, 2014: *Fluid dynamics of the mid-latitude atmosphere*. Wiley-Blackwell, 432 pp.

Hoskins, B. J., M. E. McIntyre, and A. W. Robertson, 1985: On the use and significance of isentropic potential vorticity maps. *Quart J Roy Meteor Soc*, **111**, 877–946.

Hung, M.-P., J.-L. Lin, W. Wang, D. Kim, T. Shinoda, and S. J. Weaver, 2013: MJO and Convectively Coupled Equatorial Waves Simulated by CMIP5 Climate Models. *J. Clim.*, **26**, 6185–6214, https://doi.org/10.1175/JCLI-D-12-00541.1.

Khairoutdinov, M. F., and D. A. Randall, 2003: Cloud resolving modeling of the ARM summer 1997 IOP: Model formulation, results, uncertainties and sensitivities. *J Atmos Sci*, **60**, 607–625.

Khairoutdinov, M. F., and K. Emanuel, 2018: Intraseasonal Variability in a Cloud-Permitting Near-Global Equatorial Aquaplanet Model. *J. Atmospheric Sci.*, **75**, 4337–4355, https://doi.org/10.1175/jas-d-18-0152.1.

Kiladis, G. N., C. D. Thorncroft, and N. M. J. Hall, 2006: Three-Dimensional Structure and Dynamics of African Easterly Waves. Part I: Observations. *J. Atmospheric Sci.*, **63**, 2212–2230, https://doi.org/10.1175/JAS3741.1.

——, M. C. Wheeler, P. T. Haertel, K. H. Straub, and P. E. Roundy, 2009: Convectively coupled equatorial waves. *Rev. Geophys.*, **47**, https://doi.org/10.1029/2008RG000266.

Kuang, Z., 2008: A Moisture-Stratiform Instability for Convectively Coupled Waves. *J. Atmospheric Sci.*, **65**, 834–854, https://doi.org/10.1175/2007JAS2444.1.

Lin, J., and K. Emanuel, 2022: On the effect of surface friction and upward radiation of energy on equatorial waves. *J. Atmospheric Sci.*, **79**, 837–857, https://doi.org/10.1175/JAS-D-21-0199.1.

Madden, R., and P. R. Julian, 1971: Detection of a 40-50 day oscillation in the zonal wind in the tropical Pacific. *J Atmos Sci*, **28**, 702–708.

——, and ——, 1972: Description of global circulation cells in the tropics with a 40-50 day period. *J Atmos Sci*, **29**, 1109–1123.

Mapes, B. E., 2000: Convective inhibition, subgrid-scale triggering energy, and stratiform instability in a toy tropical wave model. *J Atmos Sci*, **57**, 1515–1535.

Matthews, A. J., 2021: Dynamical propagation and growth mechanisms for convectively coupled equatorial Kelvin waves over the Indian Ocean. *Q. J. R. Meteorol. Soc.*, **147**, 4310–4336, https://doi.org/10.1002/qj.4179.

Rydbeck, A.V. and Maloney, E. D., 2014: Energetics of East Pacific Easterly Waves during Intraseasonal Events. *J Clim*, **27**, 7603–7621.

Salby, M. L., and R. R. Garcia, 1987: Transient response to localized episodic heating in the tropics. Part I: Excitation and short-time near-field behavior. *J. Atmospheric Sci.*, **44**, 458–498, https://doi.org/10.1175/1520-0469(1987)044<0458:TRTLEH>2.0.CO;2.

Wheeler, M., and G. N. Kiladis, 1999: Convectively coupled equatorial waves: Analysis of clouds and temperature in the wavenumber-frequency domain. *J Atmos Sci*, **56**, 374–399.

Wheeler, M., G. N. Kiladis, and P. J. Webster, 2000: Large-scale dynamical fields associated with convectively coupled equatorial waves. *J Atmos Sci*, **57**, 613–640.

Yano, J.-I., and and K. A. Emanuel, 1991: An improved WISHE model of the equatorial atmosphere and its coupling with the stratosphere. *J Atmos Sci*, **48**, 377-389.

Part II

Tropical Cyclones

8. Overview of Tropical Cyclones

8.1. Definition

According to the American Meteorological Society Glossary, *tropical cyclone* is “the general term for a cyclone that originates over the tropical oceans”. But for our purposes, this definition is at once too broad and too restrictive. Not all cyclones that form over tropical oceans operate by the same physical mechanism. For example, in winter, some baroclinic cyclones form in the tropics. On the other hand, certain types of polar low and medicanes (nearly circular cyclones that occasionally develop over the Mediterranean Sea) operate on essentially the same mechanism as conventional tropical cyclones yet they develop outside the tropics. And “agukabams” whose appearance in satellite imagery is nearly identical to classical tropical cyclones, develop over land. In models, tropical cyclone-like vortices develop in dry adiabatic atmospheres free of any water vapor; therefore, moist convection plays no role in these cyclones.

What unites these various phenomena is that they are all powered by surface enthalpy fluxes. This book is about the physics of such storms, and thus for our purposes, the term *tropical cyclone* will be taken to apply to *any cyclonic vortex powered primarily by enthalpy fluxes from the surface*. As with any formal definition, there are drawbacks. As mentioned above, the modifier “tropical” is too restrictive; we will consider surface flux-powered storms outside the tropics. These could even include yet-to-be-discovered storms on extra-solar planets. Some larger dust devils may turn out to be driven by surface enthalpy fluxes and yet such vortices seem very far removed from the conventional definition of tropical cyclone. And one can always nitpick. What do we mean by “vortex”? There are other phenomena, such as convectively coupled equatorial waves, which might be driven by surface enthalpy fluxes, but these do not take the form of closed vortices (i.e. having closed contours of vorticity). What do we mean by “primarily”? Classical tropical cyclones often transition into baroclinic storms powered by pre-existing horizontal density gradients; at what point do they cease being “tropical cyclones”? We must recognize that there is a continuum between cyclones driven purely by surface enthalpy fluxes and those driven entirely by baroclinic conversion of background available potential energy. Bearing such caveats in mind, we forge on.

8.2. A note on terminology and metrics

Scientists interested in tropical cyclones confront a bewildering zoo of names, measurement conventions, and units owing to the different histories of the regions affected by these storms.

Distances and length scales are variously reported in conventional MKS units but also in nautical miles (1 nm = 1.852 km), statute miles (1.609 km) and, in the case of altitude, feet (1 ft = 0.3048 m). Velocities are often reported in nautical miles per hour, or “knots” (1 kt = 0.5144 ms^{-1}). Millibars (mb) are often used to quantify pressure (1 mb = 1 hectopascal or hPa). Throughout this text we will generally adhere to the conventional MKS system of units but will make exceptions where it seems desirable.

Tropical cyclones are classified according to their peak wind speeds. Unfortunately, even the definition of “wind speed” differs from one region to the next. U.S. agencies use a 1-minute average, whereas elsewhere the average is taken over 10 minutes. (“Surface winds” usually refer to winds measured at the conventional altitude of 10 m.) Because boundary layer winds vary at high frequency, 1-minute peaks are larger than 10-minute peaks. In the absence of detailed information about boundary layer turbulence, a conversion factor of 0.88 is usually applied. In the definitions that follow, we use 1-minute averages.

We will use the word “intensity” synonymously with “wind speed”, but central surface pressure is often used as a measure of storm intensity as well. Tropical cyclone “size” refers to the horizontal scale of the storm, and here several metrics are relevant. The most commonly used size metric is the radius of maximum winds (RMW); that is, the average distance from the storm center of the peak surface winds. It is also common to refer to the radius of gale force (17 ms^{-1} or 34 kt) winds, the radius of the outermost closed isobar (ROCI) of surface pressure, and the radius of vanishing wind (RVW); i.e., some measure of the radius of the outermost detectable influence of the tropical cyclone on surface winds. The diameter of the eye visible in satellite or radar imagery is sometimes recorded; this should not be confused with the radius of maximum winds.

There are three main intensity-based classifications of tropical cyclones[1]. The first and weakest is the *tropical depression*, defined to have winds less than 17 ms^{-1}. There is considerable ambiguity is identifying a system as a tropical depression, wrapped up in the problem of defining *genesis* (see Chapter 14). In particular, storms identified as tropical depressions may not even be tropical cyclones as we have defined them in section 8.1 above; that is, they may not at this stage be powered primarily by surface enthalpy fluxes. Indeed, tropical depression-strength cyclones occasionally develop over land and in these cases may not powered by surface fluxes (but see section 8.3.9 on agukabams).

The term *tropical storm* is generally used to describe tropical cyclones with peak 1-minute winds between 17 and 32 ms^{-1}. Many different names are applied to storms with wind speeds higher than 32 ms^{-1}, and there are several classification schemes that further subdivide these more intense storms. For example, the term *hurricane* (derived from the various names, such as Hurácan and Hunraken, of the god of evil of pre-Columbian inhabitants of the Caribbean and Mexico) is used in the North (and South) Atlantic and eastern North Pacific, while *typhoon* (derived from the Chinese *jufeng* and/or the Greek earth demon *typhon*) is used in the western North Pacific. Sub-divisions of hurricane-intensity tropical cyclones include the Safir-Simpson scale, a 5-category subdivision used by the U.S. and several other nations[2].

For the purposes of this text, we use 1-minute average surface winds and MKS units to describe storm intensity and occasionally use the terms *tropical depression*, *tropical storm,* and *hurricane* as ways of classifying storms by intensity.

Well-developed tropical cyclones are highly coherent structures that can be tracked for long distances and time, so it is natural to give them names. In the North Atlantic region prior to 1950, significant storms were named informally, often after the places, ships, or even individuals

[1] A comprehensive summary of the classification schemes used by various government agencies worldwide may be found at https://en.wikipedia.org/wiki/Tropical_cyclone .
[2] See "Saffir–Simpson scale." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 11 Mar. 2024. Web. 4 Apr. 2024.

they affected, or after the Saint's day on which their effects were felt. For example, a deadly hurricane that affected the Caribbean and Florida in 1928 is known in the continental U.S. today as the *Okeechobee Hurricane* (named after a large southern Florida lake whose bordering towns were demolished by the storm) and as the *San Felipe Hurricane* in Puerto Rico. In Australia, Clement Wragge began naming storms in the late 19th century, ultimately using the names of unpopular politicians, a potentially limitless resource. Three storms were observed simultaneously during the 1950 North Atlantic hurricane season, causing much confusion and leading U.S. forecasters to begin using the Army/Navy phonetic alphabet (*Able, Baker, Charlie*) to identify individual tropical cyclones, replacing that in 1952 with the NATO phonetic alphabet (*Alpha, Bravo, Charlie*, etc.). From 1953 through 1978 North Atlantic tropical cyclones were given exclusively female names, after which alternating male and female names were applied. Today, all tropical cyclones of tropical storm strength or greater are named from pre-determined lists drawn up by one of eleven tropical cyclone warning centers, depending on the region affected.

8.3. Climatology[3]

8.3.1. Frequency

On average, some 90 tropical cyclones of at least tropical storm strength form on our planet each year[4]. Since about 1980, by which time nearly all tropical cyclones were detected using satellites, there has been no significant trend in this number, and its year-to-year standard deviation is about 9.5, almost exactly the variability that would occur in a Poisson process (for which the standard deviation is the square root of the mean).

8.3.2. Geographic distribution

The tracks of all tropical cyclones in the historical record from 1851 to 2010 are displayed in Figure 8.1. Figure 8.2 shows the number of tracks passing through each 3-degree latitude-longitude square each year, averaged over all the tropical cyclones in from 1980 to 2014.

The greatest concentration of tropical cyclones is in the western North Pacific region. These storms typically form south of 20° N, anywhere from the South China Sea eastward as far as the dateline, and move westward and northward. Those that do not strike land or otherwise dissipate often recurve northward and eastward, penetrating middle and high latitudes. Tropical cyclones form in the Bay of Bengal and in the Arabian Sea, mostly during the Boreal spring and fall. (Strong wind shear associated with the Asian summer monsoon suppresses formation during the summer; see section 12.2) The deadliest tropical cyclone in history struck East Pakistan (now Bangladesh) in November of 1970, killing between 300,000 and 500,000 people.

[3] Here, and elsewhere in this book, we rely on the International Best Track Archive for Climate Stewardship (IBTrACS) database (Knapp et al., 2010). The reader is urged to consult that reference for a detailed discussion of the historical record, which is sometimes referred to in literature as the "best track" record. In particular, outside the North Atlantic region, even the frequency of historical tropical cyclones is uncertain before about 1980. Data are available at https://www.ncei.noaa.gov/products/international-best-track-archive.

[4] These do not include some storms, such as polar lows, that develop under our extended definition of "tropical cyclone".

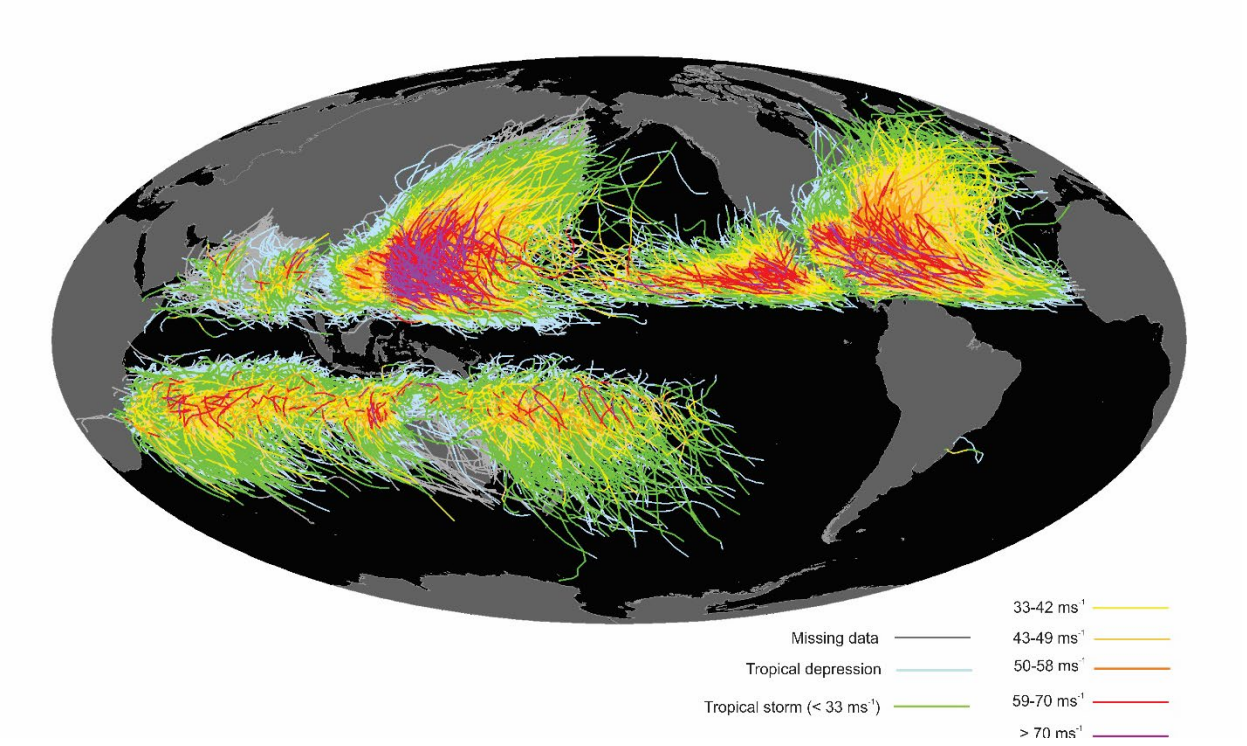


Figure 8.1: Tracks of all tropical cyclones in the historical record from 1851 to 2010. The tracks are colored according to the maximum wind at 10 m altitude, on the scale at lower right.

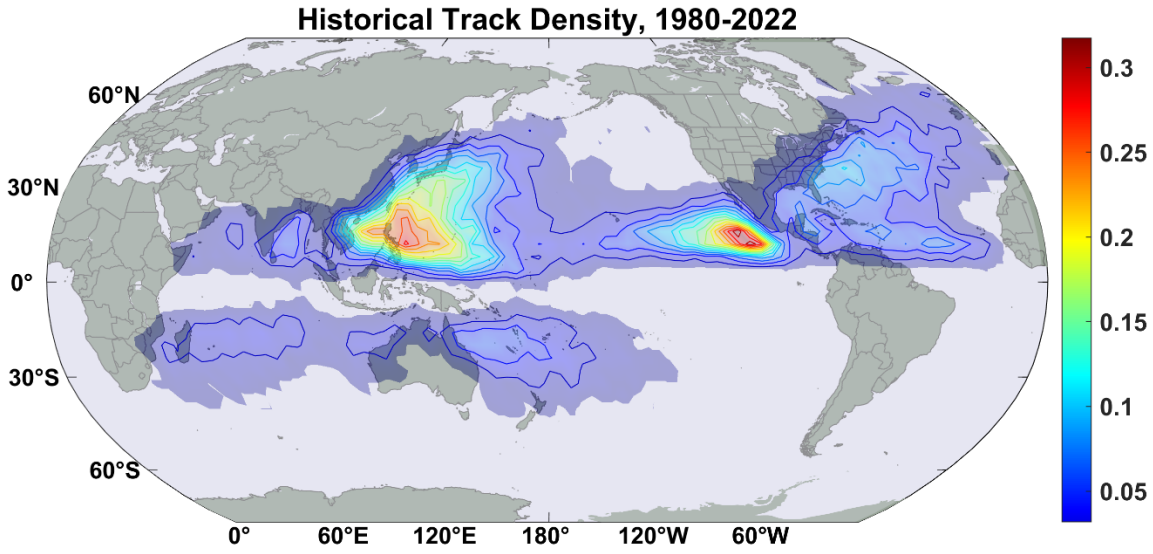


Figure 8.2: Number of historical tropical cyclone events per 3 x 3-degree latitude square per year averaged from 1980 to 2022. Color scale at right.

Another active region is the eastern North Pacific. Storms here typically develop off the west coast of Mexico and usually travel westward, rarely affecting land, but on occasion they recurve northward and eastward, striking the Mexican coast as far north as Baja California and may even reach southern California, bringing unusually heavy rains to that region. Some eastern Pacific tropical cyclones travel far enough west to affect the Hawaiian Islands.

In the North Atlantic, tropical cyclones develop anywhere from the west African coast westward to the Caribbean and the Gulf of Mexico and as far north as 30^{o} N off the coasts of the Carolinas. Once developed, they typically move westward and northward, and many eventually recurve northward and eastward, sometimes travelling far northward into the high latitudes of the North Atlantic. As shown in Figure 8.2, the density of tracks in the North Atlantic is low compared to other cyclone-prone regions, and only about 12% of all tropical cyclones develop there. Yet much of the literature focuses on North Atlantic storms because, on average, they are far better observed than storms in other regions, thanks to a vigorous program of aircraft reconnaissance dating back to the mid-1940s. There is much more media coverage of North Atlantic storms because their destructiveness is far out of proportion to their number, owing to the very high concentration of wealth along the U.S. east and Gulf coasts, coupled with poor risk management policies.

There are no tropical cyclones within a few degrees latitude of the equator, because long-lived cyclonic vortices require a non-zero Coriolis parameter (the projection onto the local vertical of twice the planetary angular rotation vector) and the large meridional gradient of the Coriolis parameter strongly inhibits coherent vortex development. There is an active belt of tropical cyclones in the southern hemisphere, stretching from Mozambique across Madagascar and the Southern Indian Ocean, northern Australia, and into the South Pacific as far east as French Polynesia.

In the historical record from 1851 to 2024 there has been only one hurricane-strength tropical cyclone in the South Atlantic, Hurricane *Catarina* of March, 2004; this event did considerable damage in Brazil. It is possible that a few hurricane-strength events occurred in the South Atlantic between 1851 and the dawn of the satellite era and were not detected by ships or land-based observations.

8.3.3. Life Cycle

Most, if not all, tropical cyclones originate in preexisting disturbances of largely independent origins. (At least, that is the reigning dogma in tropical meteorology. We shall see in Chapter 14 that tropical cyclones can develop spontaneously in computational models under certain circumstances.) For this reason, it is difficult, on a purely phenomenological basis, to decide when a tropical cyclone has formed. While there is no generally accepted set of criteria for when a tropical cyclone has developed, most operational forecast centers will declare that a tropical depression has formed based on the existence of a closed surface wind circulation around a well-defined center, assuming that the system is otherwise broadly consistent with expectations based on structure and climatology.

Figure 8.3 shows the origin points of all tropical cyclones in the historical record between 1980 and 2022. There are three main genesis regions: the North Atlantic including the Caribbean and the Gulf of Mexico, extending westward across Mexico to the eastern tropical North Pacific. A second region, not altogether separated from the first, spans from the central North Pacific across the far western North Pacific and southeast Asia to the Bay of Bengal and the Arabian

Sea. In the Southern Hemisphere, tropical cyclones originate from just east of French Polynesia westward across the South Pacific, spanning the region around northern Australia, including the Coral and Timor Seas and the Gulf of Carpentaria, and across the whole of the South Indian Ocean across Madagascar to Mozambique, southern Tanzania, and the northeast coast of South Africa.

As mentioned above, tropical cyclones do not develop near the equator because they require a sufficiently large absolute value of the Coriolis parameter. One exceptional storm was Cyclone *Agni* that developed at around 1.5° N in the Indian Ocean in November of 2004, traveling as far south as 0.7° N. Close inspection of Figure 8.3 shows three tropical cyclones developing north of 40° N. These include Tropical Storm *Grace* of 2009, which developed near 45° N. This storm was an example of an extratropical cyclone that eventually travelled far enough south to become a tropical cyclone.

As we will see in Chapter 14, tropical cyclones develop where the thermodynamic potential is high and where the absolute value of the Coriolis parameter is sufficiently large. We will examine the physics of genesis in Chapter 14.

Once formed, tropical cyclones will continue to develop unless and until environmental conditions become adverse. An obvious case is when they travel over land, losing the upper ocean reservoir of heat that sustains them. (But see 8.3.9 on agukabams later in this chapter.) They may also travel into regions whose thermodynamic potential (Chapter 14) is insufficient, causing them to decay and eventually dissipate. Vertical shear of the horizontal wind is highly deleterious to tropical cyclones (Chapter 12) and may lead to dissipation even under otherwise very favorable conditions. Finally, tropical cyclones may undergo a transition to baroclinic systems, a process known as *extratropical transition* (Chapter 12), and they may even re-intensify in their new incarnations, generally becoming physically larger and sometimes very destructive.

Under idealized conditions in numerical simulations, tropical cyclones may persist indefinitely (see Chapter 14). But in the real world, only a few tropical cyclones intensify right up to their thermodynamic limiting intensity (Chapter 9). The vast majority lead lives of quiet desperation, struggling against environmental adversity, occasionally triumphing but finally succumbing to any of a variety of fatal influences.

Figure 8.4 shows a histogram of the lifetimes (times between first and last points in the historical track database) of all tropical cyclone globally between 1980 and 2022. The average storm lifetime is a little more than one week, and the median lifetime is 6.8 days. The longest lasting tropical cyclone in the record through 2022 was Cyclone *Pancho-Helinda*[5] of 1997, which lasted more than 32 days as it traversed virtually the entire South Indian Ocean. Very recently, Cyclone Freddy broke this record, traversing most of the South Indian Ocean for 37 days during February and March, 2023.

The last track point in historical databases is just as arbitrary as the first. In most cases, tropical cyclones are labeled as dissipated when their peak surface winds drop below some threshold, typically 25 knots. But the track may end when the storm is considered to have transitioned to a

[5] The storm's initial name was *Pancho* but it was renamed *Helinda* when it passed into the southwestern Indian Ocean.

primarily baroclinic system, even though its maximum surface winds may be well in excess of the usual threshold.

Figure 8.5 shows the last track points of all tropical cyclones in the historical record from 1980 through 2022. Termination points are as far poleward as the Labrador Sea and as far inland as Ontario and central Australia.

8.3.4. Seasonality

Tropical cyclones form preferably in the summer and early fall, as shown in Figure 8.6. An exception to this rule is the North Indian Ocean, which exhibits two peaks, one in May and the other in October and November. As we will discuss in Chapter 12, tropical cyclones are strongly suppressed by vertical shear of the horizontal wind, which becomes large during the Asian summer monsoon.

8.3.5. Intensity

Analysis of the climatology of tropical cyclones is compromised by the relatively poor quality of measurements of storm intensity. Reliable estimates based on aircraft and aircraft-deployed dropsondes are only available in the North Atlantic from roughly the late 1950s to the present, and in the western North Pacific from the late 1950s to 1987, when aircraft reconnaissance was discontinued there. Aircraft reconnaissance extends back to about 1945 in both basins, but quantitative intensity estimates were based on central pressure, if dropsonde measurements were available, or extrapolation of flight level pressure data to the surface using flight level temperature and a standard temperature profile (such as a moist adiabat). Direct wind speed estimates were made by visual inspection of the sea surface, but these necessarily have a strong subjective component. By the late 1950s or early 1960s, forward-scanning Doppler radars were installed on some aircraft, allowing for a quantitative estimate of the aircraft's ground speed. By around 1968, implementation of inertial navigation systems on reconnaissance aircraft further increased the accuracy of flight level winds. By 1980, earth-orbiting satellites were detecting most tropical cyclones around the globe, and methods were developed to estimate storm intensity from satellite-based measurements.

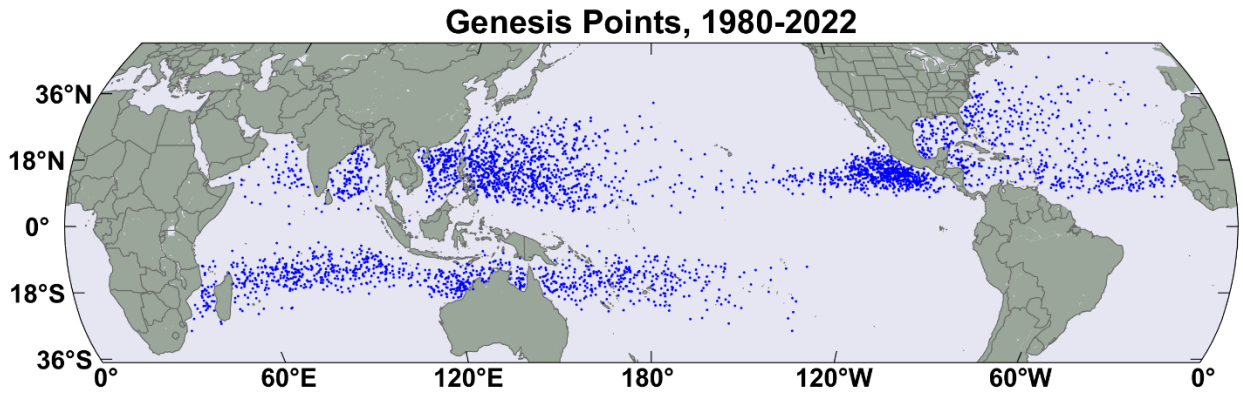


Figure 8.3: Points of origin of all tropical cyclones in the historical record between 1980 and 2022 This does not include polar lows, medicanes, or agukabams.

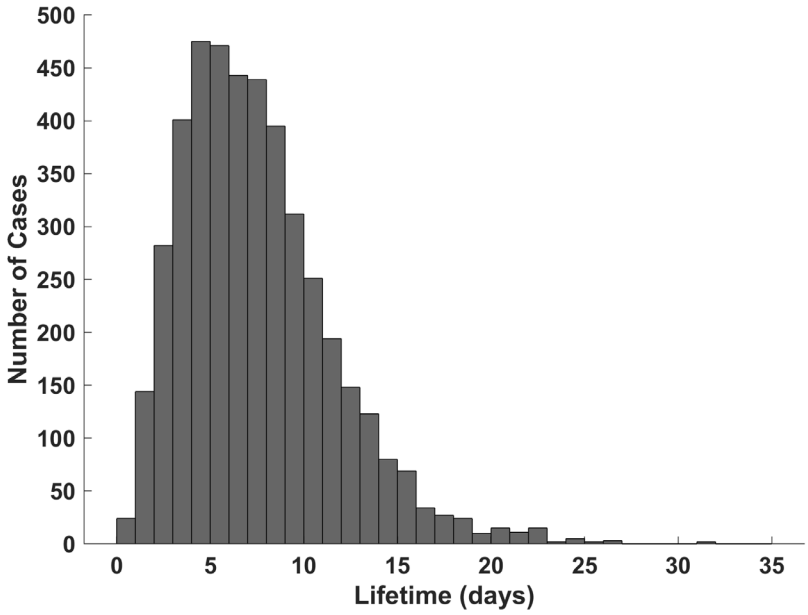


Figure 8.4: Histogram of the lifetimes of tropical cyclones globally, from 1980-2022. The number of cases are binned into intervals of 1 day.

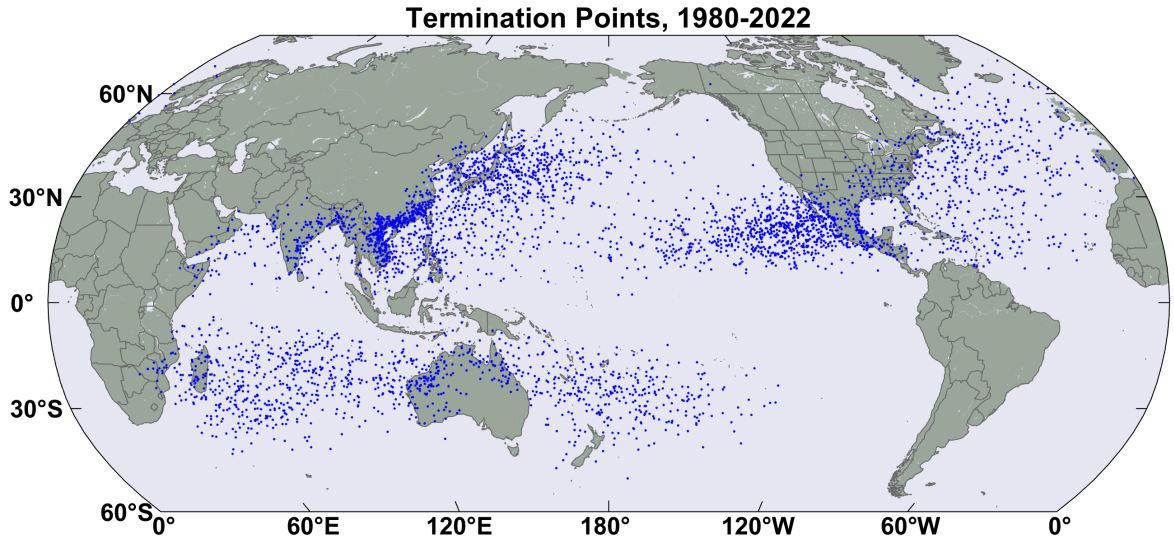


Figure 8.5: Termination points of all tropical cyclones in the historical record from 1980 to 2022. This does not include polar lows or medicanes, though some agukabams may be included.

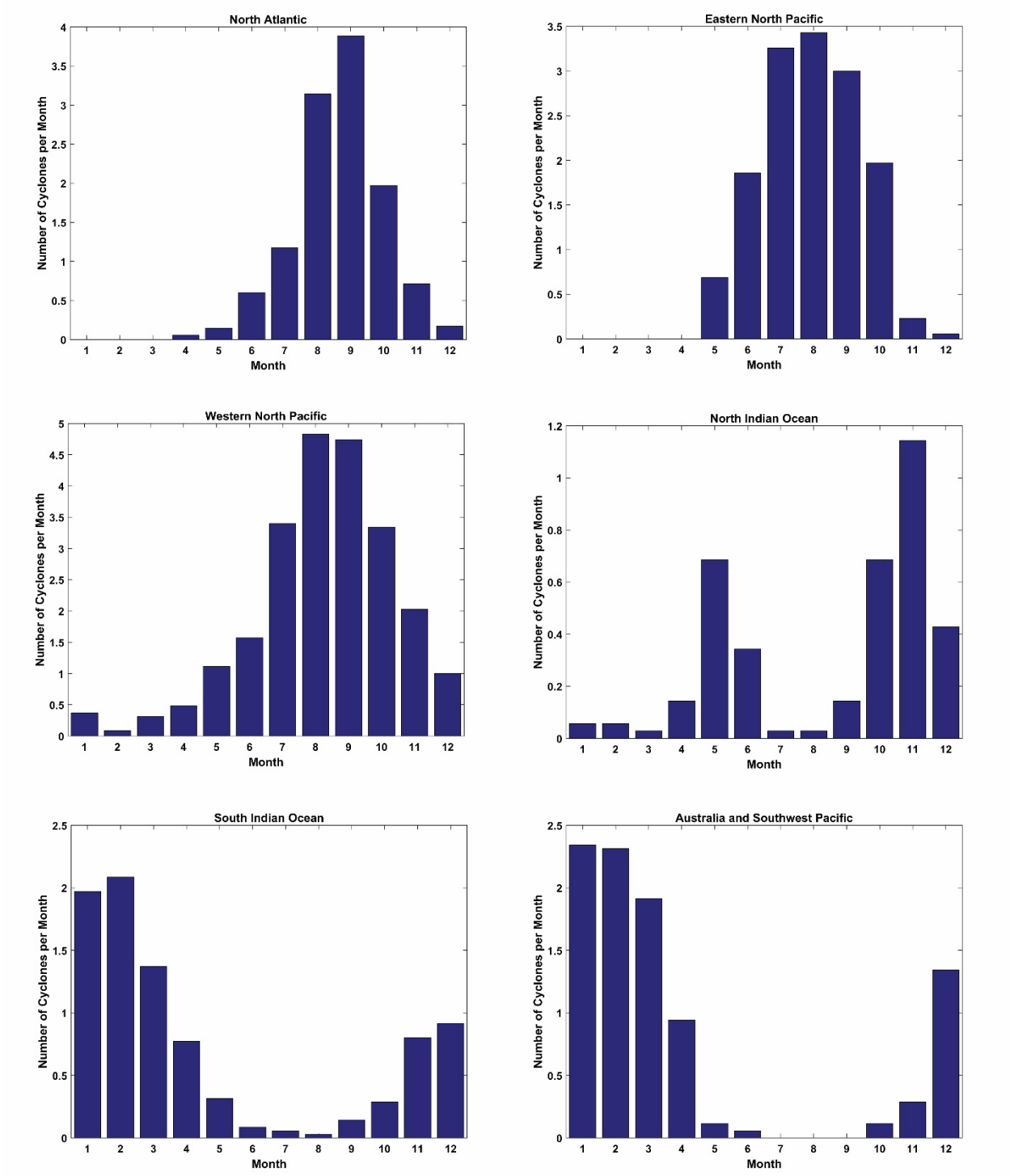


Figure 8.6: Number of tropical cyclones per month, averaged over the period 1980-2014, in six ocean basins. From left-to-right and top-to-bottom: North Atlantic, eastern North Pacific, western North Pacific, north Indian, south Indian, and Australia and the southwestern Pacific.

One useful intensity measure is the maximum surface wind estimated for each storm during its lifetime. This quantity is known as its *Lifetime Maximum Intensity (LMI)*. Figure 8.7 shows the probability densities of LMI for tropical cyclones in the North Atlantic and western North Pacific regions over the period 1945-1987, during which there was aircraft surveillance of at least some of the events.

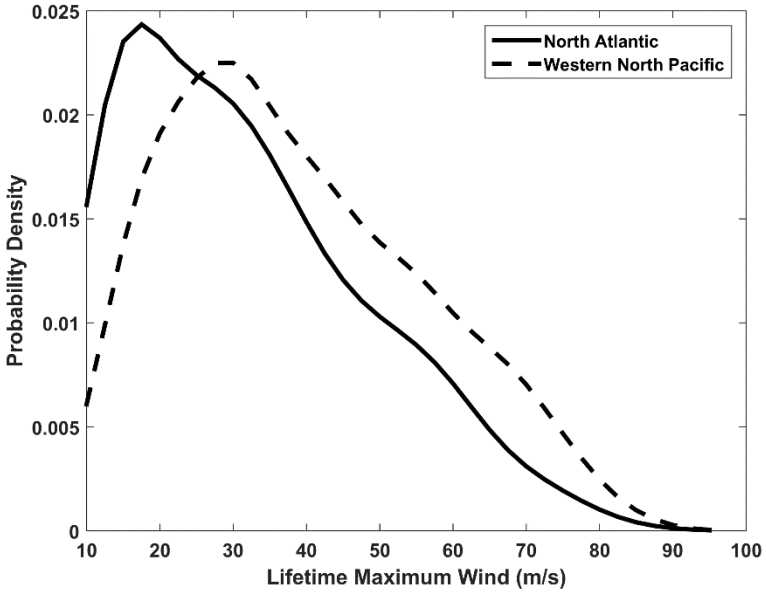


Figure 8.7: Probability densities of lifetime maximum intensities (LMIs) of tropical cyclones in the North Atlantic (solid) and western North Pacific (dashed) from 1945 to 1987. The probabilities are calculated on intervals of 2.5 ms^{-1}.

For storms of hurricane strength (> 32 ms^{-1}), the probability density decreases nearly linearly with LMI. Note that the probability distribution is bounded; there are no recorded cases of wind speeds in excess of 100 ms^{-1}. The world record tropical cyclone wind speed of 95 ms^{-1} was set by eastern North Pacific Hurricane *Patricia* of 2015. In chapter 15 we will show that normalizing the LMIs by potential intensity at the time and place of maximum intensity results in a near universal probability distribution.

Figure 8.8 shows the probability densities of tropical cyclones in the eastern and western North Pacific in the period 1988-2014, during which almost all the intensity estimates were made using satellite-based techniques (Appendix 1). In contrast to the probabilities during the aircraft reconnaissance era (Figure 8.7), the distributions are noticeably bimodal. As of this writing, it is not known whether this difference is owing to climate change between the earlier and later periods or to differences between aircraft-based and satellite-based intensity inferences[6].

A map of the average annual maximum tropical cyclone surface wind over each 3-degree latitude-longitude square is shown in Figure 8.9. The most intense storms are in the far eastern and far western North Pacific.

[6] Velden et al. (2006) present evidence that when eyes are just forming, they are often covered with a cirrus canopy that hides the eye in satellite imagery, leading to underestimation of storm intensity using the then-current Dvorak technique for estimating tropical cyclone intensity from satellite imagery. This could lead to the dip in the probability distribution at weak hurricane intensities, seen in Figure 8.8.

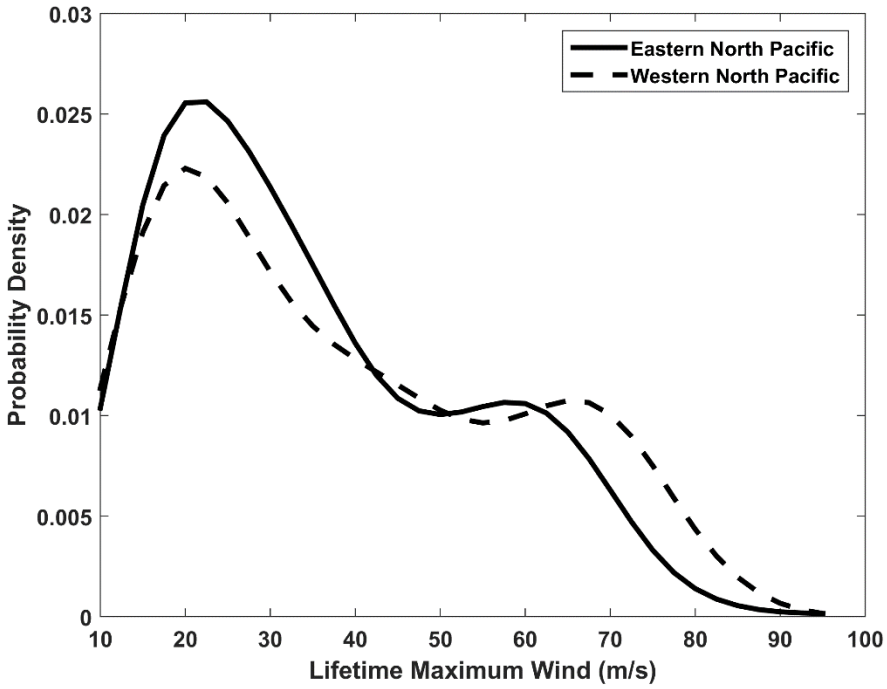


Figure 8.8: Probability densities of lifetime maximum intensities (LMIs) of tropical cyclones in the eastern (solid) and western (dashed) North Pacific from 1988 to 2014. The probabilities are calculated on intervals of 2.5 ms^{-1}.

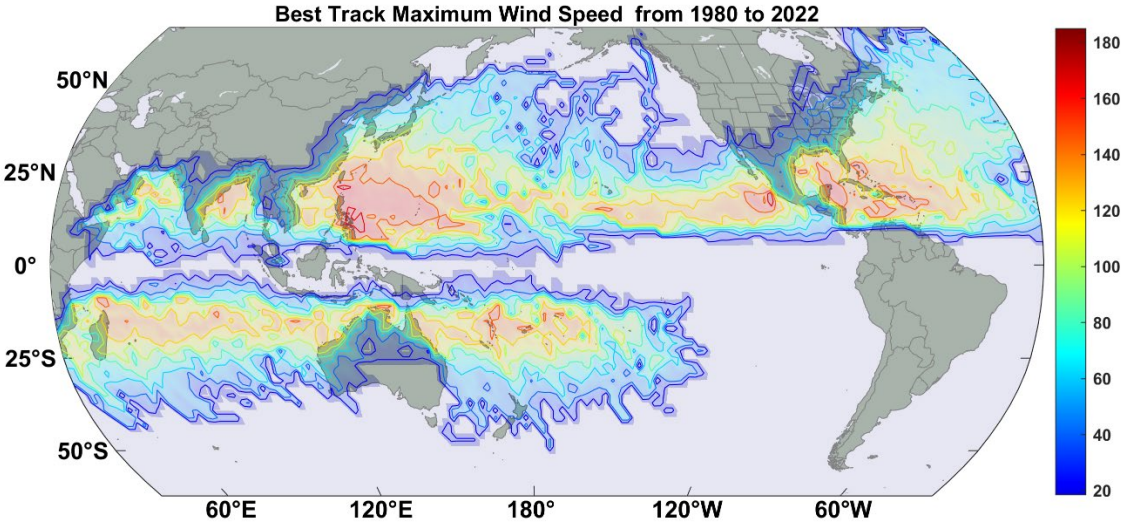


Figure 8.9: Maximum tropical cyclone wind speed (ms^{-1}) in each 3 x 3-degree latitude square, over the period 1980-2022.

Any serious student of tropical cyclone intensity must be aware of the substantial inhomogeneities that plague the historical tropical cyclone record, brought about by different techniques for estimating intensity and evolving standards and techniques. Historical intensity data should not be accepted at face value. We shall have more to say about climatological tropical cyclone intensity in Chapter 15.

8.3.6. Size

Tropical cyclone diameters span a wide range. Atlantic tropical storm *Marco* of 2008 had gale force (17 ms^{-1}) winds extending outward to only about 18 km; its radius of vanishing winds was likely less than 40 km. Yet the eye of Typhoon *Carmen* of 1960 was measured by radar to have a radius of 160 km, so that four entire *Marcos* could fit within *Carmen's* eye.

Climatologies of tropical cyclone size metrics are compromised by a comparative dearth of recorded observations. Historical data for the North Atlantic and eastern North Pacific now contains information about the radii of gale force (17 ms^{-1}), 25 ms^{-1} and hurricane force (33 ms^{-1}) winds going back to 2004, but this information is not available for every storm at every observation time. Similarly, size information is available in historical tropical cyclone data provided by the U.S. Joint Typhoon Warning Center (JTWC) going back to 2001.

One measure of size is the radius of maximum winds (RMW), which refers to the azimuthally averaged radius at which surface winds reach their peak values. (As discussed later in this chapter, there can be secondary and even tertiary wind maxima at various radii outward or inward from the primary RMW.) The RMW can vary greatly within the lifetime of an individual storm, and from one storm to the next; for a given storm, it generally contracts as the storm intensifies, but secondary eyewalls can form at greater radii and contract inward, choking the original eyewall and replacing it at a greater radius (see section 8.4.4). As this metric is rather sparsely reported in standard historical tropical cyclone data sets, we here use global satellite-based estimates over the period 1983-2005, as reported by Kossin et al. (2007). Figure 8.10 shows the probability density of the RMW from this data set, which contains 11,937 individual fixes. The mean RMW is around 65 km, and most storms have RMWs between 20 and 100 km.

Another useful and interesting size metric is the radius of vanishing wind (RVW), defined as azimuthally averaged distance from the storm center at which there is no detectable influence of the tropical cyclone on surface winds. This metric is entirely absent from conventional historical data sets because it is very difficult to determine from aircraft data or from satellite visible or infrared imagery. One way to estimate the RVW is to fit a theoretically derived outer wind profile to observations at intermediate radii and use that to extrapolate out to the RVW (Dean et al., 2009). An even better method is to take advantage of winds determined from satellite-based sea surface scatterometry. A scatterometer send pulses of microwave radiation downward at an angle to the sea surface and measures the power of the back-scattered radiation. The pulses have a wavelength such that they are Bragg-scattered from capillary-gravity waves on the sea surface; these are usually nearly in equilibrium with the surface wind stress. By measuring the backscattered power from several different viewing angles as the satellite orbits, the orientation of the waves can be determined, giving the direction of the wind stress; the magnitude of the backscattered radiation is proportional to the amplitude of the stress. Empirical and/or theoretical relationships are used to deduce standard 1-minute winds at 10 m altitude form the surface stress. An example of a surface wind field associated with a western North Pacific typhoon is shown in Figure 8.11. Heavy rain also strongly backscatters scatterometer pulses,

introducing errors, and so the technique does not work well in the inner core regions of tropical cyclones, where heavy rain is the norm. But the technique works well in the outer regions, which are usually free of heavy rain.

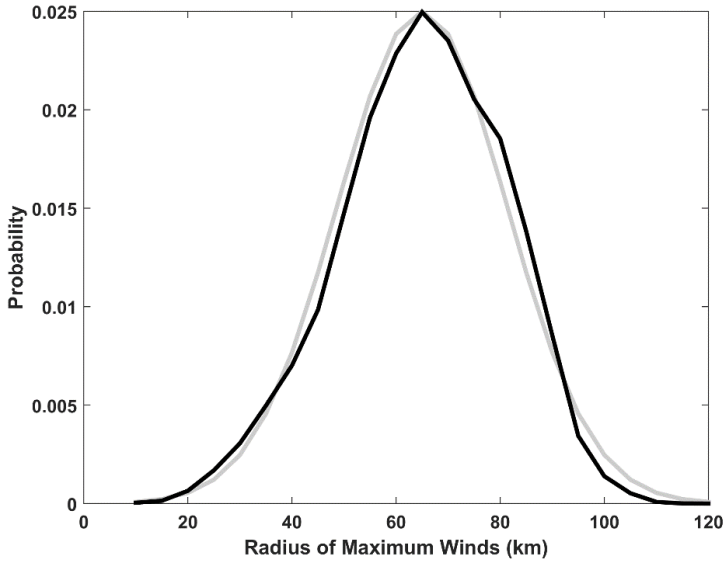


Figure 8.10: Probability density (black curve) of the radii of maximum winds (km) estimated from satellite infrared imagery by Kossin et al. (2007). The probabilities are binned in 5 km intervals of RMW. The gray curve is a best fit Gaussian distribution.

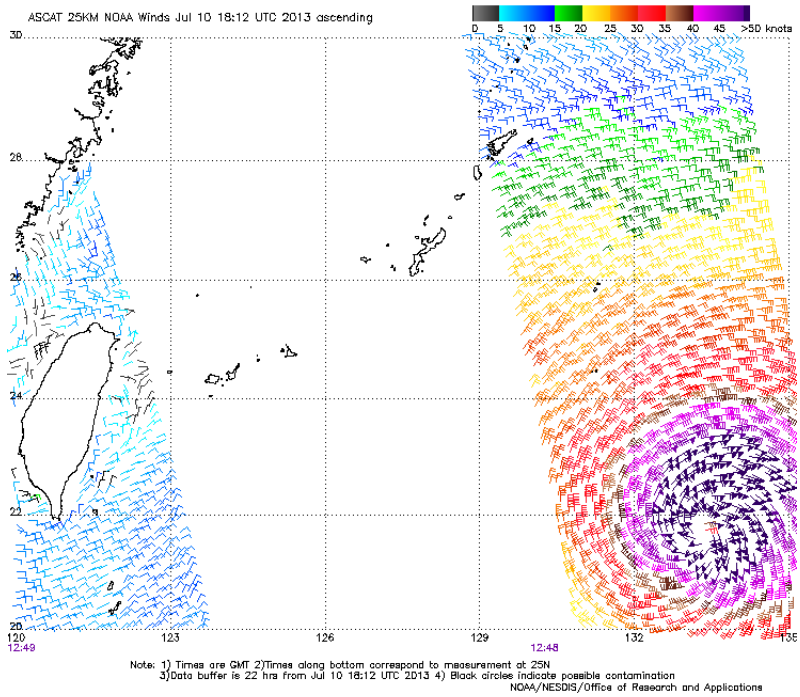


Figure 8.11: Surface winds associated with Typhoon Soulik on July 10th, 2013 determined from by Eumetsat's ASCAT (Advanced Scatterometer) instrument on board the Metop-A satellite. The barbs point into the wind, and the colors indicate the wind speeds according to the scale at upper right.

Here we rely on the work of Chavas and Emanuel (2010), who used QuickSCAT scatterometer data from 1999 to 2008 to determine the radius at which the azimuthally averaged wind falls to 12 ms^{-1}. They then used the same outer wind model as employed in Dean et al. (2009) to estimate the RVW. They found that the frequency distribution of the RVW is very nearly lognormal (Figure 8.12) with a global median corresponding to an RVW of 423 km. They also found that, in contrast to the RMW, the RVW typically varies little over the life of a given storm but can vary greatly from one storm to the next.

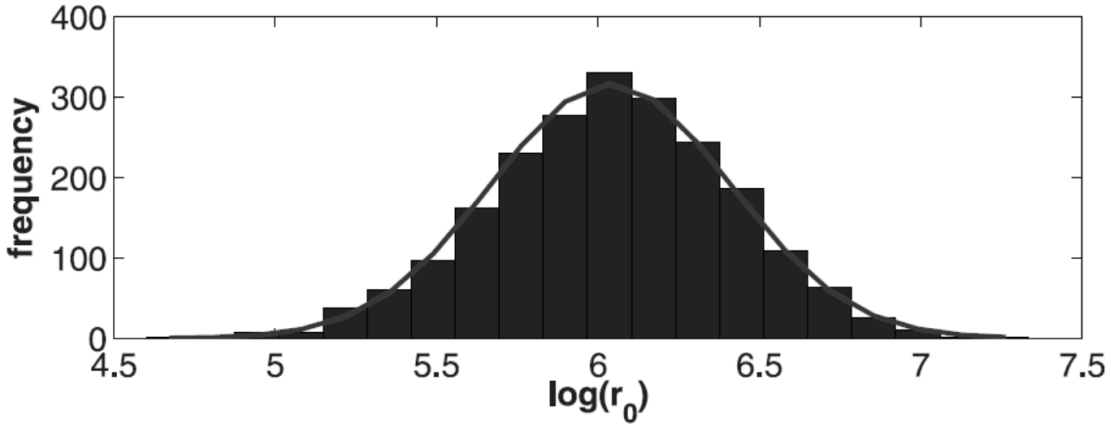


Figure 8.12: Frequency distribution of the natural logarithm of the radius (km) of vanishing wind (RVW, here indicated as r_0) from 2154 scatterometer observations of tropical cyclones as reported in Chavas and Emanuel (2010). The frequencies were calculated in bins of width 0.1 in the natural logarithm of the RVW.

8.3.7. Rain

Tropical cyclones produce extremely heavy rain. The worlds' record rainfalls at 12, 24, 72, and 96 hours were all caused by the interaction of tropical cyclones with the high relief of the island of La Réunion in the South Indian Ocean. But rain is difficult to measure, because the correlation lengths scales of precipitation are small, particularly in convective events, and sampling by limited arrays of rain gauges is thus highly problematic. At sea, where tropical cyclones spend most of their lives, there are no gauges aside from those on the odd island station. Land-based radars, which measure the power of backscattered microwave radiation, are useful for detecting tropical cyclone rainfall over land and coastal waters, though deducing actual rainfall rates from radar reflectivity is not straightforward and subject to some uncertainty.

From late November 1997, to mid-April, 2015, the National Aeronautics and Space Administration (NASA) and the Japan Aerospace Exploration Agency conducted the Tropical Rainfall Measuring Mission (TRMM) to quantify precipitation in the tropics. It carried a microwave precipitation radar, allowing active measurement of precipitation from space. Figure 8.13 shows a global climatology of tropical precipitation over the period 1998-2006 as well as rainfall attributed to tropical cyclones. This was made by combining TRMM precipitation radar measurements with conventional passive infrared from a variety of satellites (Huffman and co-authors, 2007). These measurements show that tropical cyclones contribute roughly 7% of the total precipitation in the tropics, though there are a few locations in which they contribute more than half the total. The general pattern of tropical cyclone rainfall resembles that of tropical cyclone tracks (Figure 8.2), as one might expect.

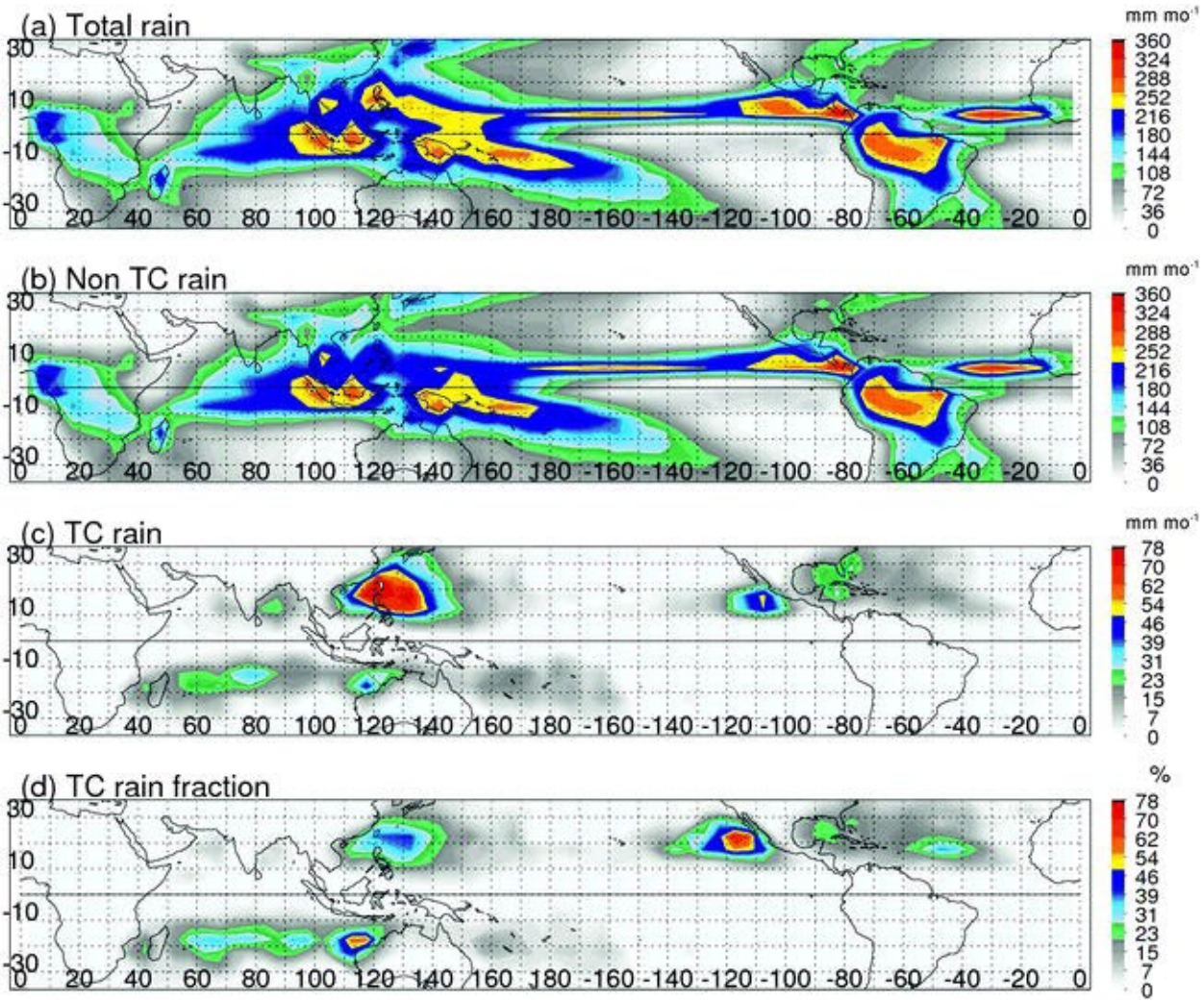


Figure 8.13: Tropical climatology of precipitation (mm per month) from a combination of TRMM precipitation radar and conventional infrared measurements from other satellites over the period 1998-2006. Panel (a) shows the total rainfall, while the non-tropical cyclone, tropical cyclone, and fraction (%) of tropical cyclone rainfall are shown in panels b,c, and d. From Jiang and Zipser (2010).

Although tropical cyclones do not dominate tropical mean precipitation, they do contribute disproportionately to rainfall extremes on time scales of several days and are a major source of freshwater flooding in many tropical cyclone-prone locations.

8.3.8: Extratropical and tropical transitions

Tropical cyclones are powered by surface enthalpy fluxes. By contrast, extratropical cyclones are powered by the release of available potential energy when horizontal density contrasts are reduced by ascent of relatively light air and descent of relatively heavy air through the process of baroclinic instability. This process lowers the center of gravity of the atmosphere and thereby converts potential into kinetic energy. The energy sources of tropical and extratropical cyclones are therefore very different.

Yet in many cases, storms that start their lives as pure tropical cyclones begin at some point to interact with temperature gradients at higher latitudes and may eventually transform into purely

baroclinic systems. It is also possible for the transition to occur in reverse, with extratropical cyclones gradually transitioning to tropical cyclones. In their hybrid states, such cyclones tap energy from both surface fluxes and the available potential energy associated with larger-scale horizontal density gradients on isobaric surfaces. Some systems, like Atlantic Hurricane Sandy of 2012, transition back and forth several times between primarily baroclinically and mostly surface flux-driven cyclones.

A case in point in Atlantic Hurricane Floyd of 1999, which began in the central tropical Atlantic and moved northwestward toward the U.S. east coast, making landfall briefly in the Carolinas before turning northeastward, moving along the New England coast and then over Nova Scotia and Newfoundland. Satellite imagery at four times during Floyd's life is shown in Figure 8.14.

i

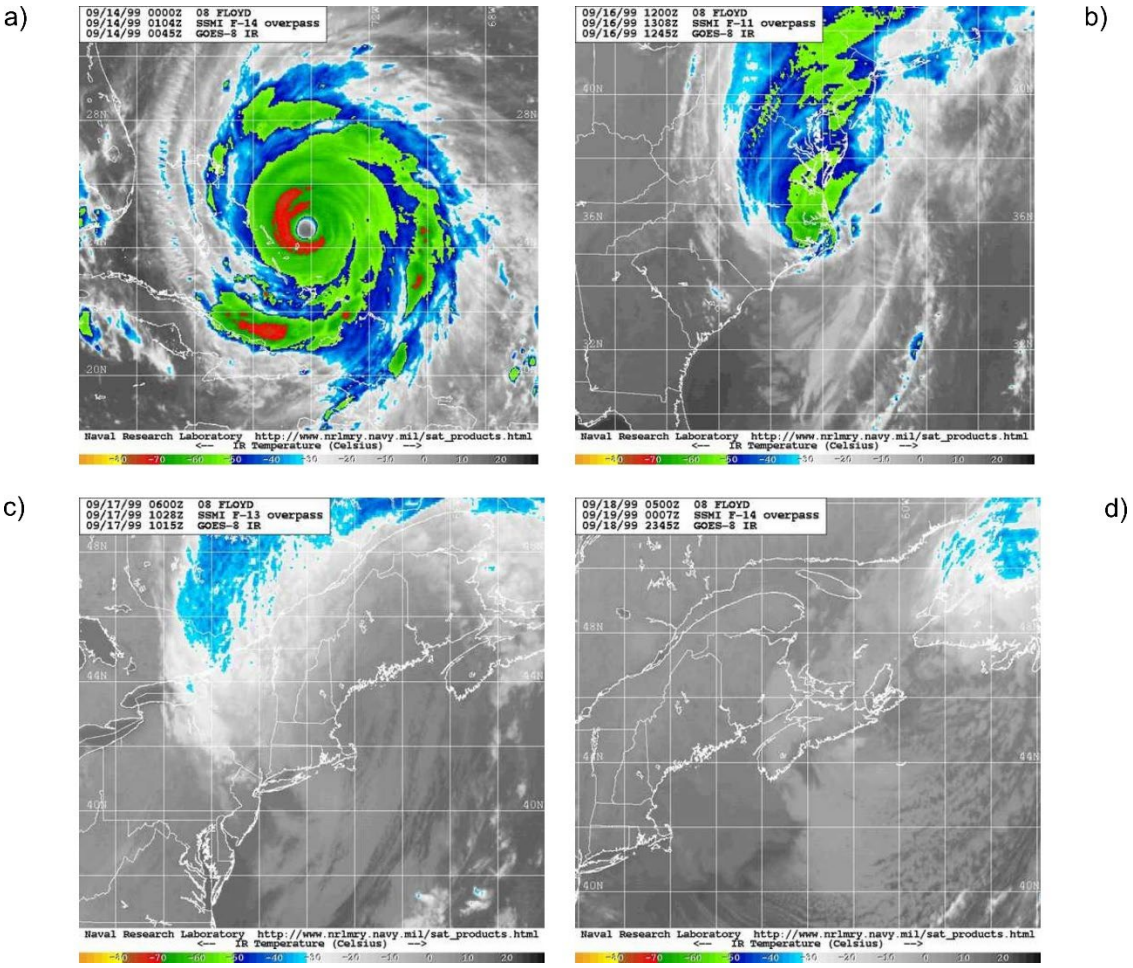


Figure 8.14: Infrared imagery at four times during the life of Atlantic Hurricane Floyd of 1999: a) 00:45 GMT September 14, b) 12:45 GMT September 16, c) 10:15 GMT September 17, and d) 23:45 GMT September 18. In b) the surface cyclone center was near the North Carolina-Virginia border at the coast, in c) it was on the coast near the New Hampshire-Maine border, and in d) it was over the south coast of Newfoundland.

Figure 8.15 shows Ertel potential vorticity (PV) on the 325 K isentropic surface, and 1000-500 hPa thickness at four times very nearly corresponding to those of Figure 8.14. On September 14th, Floyd is over the Bahamas and is naturally associated with very large PV. Note that the thickness gradient is simply the nearly axisymmetric field of Floyd itself. Note the large mid-latitude trough, signified by high PV over south-central Canada and the Great Lakes. By 12 GMT on 16 September, Floyd is beginning to embed itself in a large-scale thickness gradient over the eastern U.S., and its PV signature is coalescing with the mid-latitude trough, which is extending southwestward as a breaking Rossby wave. At this time, the storm center is making landfall and thus losing support from surface enthalpy fluxes. Yet the superposition of the remaining low-level PV of the tropical cyclone with the PV of the mid-latitude trough is a process that generates kinetic energy. (Using the quasi-geostrophic equations, it can be shown than bringing like-signed PV anomalies into closer proximity increases the kinetic energy of the geostrophic flow.) By 12 GMT on September 17th the merger is complete, and aside from plentiful moisture, the system is hardly distinguishable form an ordinary baroclinic cyclone. The PV anomaly is absorbed back into the main hemispheric gradient by 19 September. (Meanwhile, Hurricane Gert has moved into the field of view in the southeastern part of the map.)

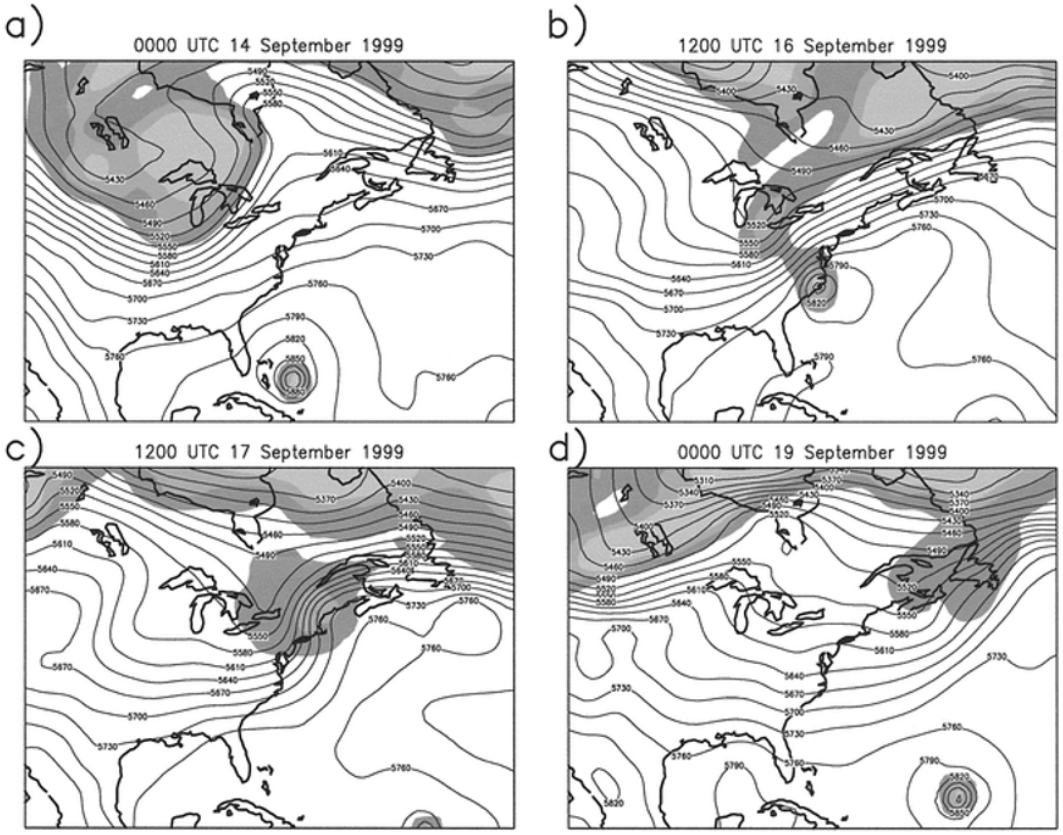


Figure 8.15: Ertel's potential vorticity (shaded at 1,2, and 3 potential vorticity units) on the 325 K isentropic surface, and 1000–5000-hPa thickness (contoured every 30 m) at (a) 0000 UTC 14 September, (b) 1200 UTC 16 September, (c) 1200 UTC 17 September, and (d) 0000 UTC 19 September. Analyses are from NOGAPS 1° × 1° initialization fields.

It would be nice to calculate the ratio of mass-integrated energy production by surface enthalpy fluxes and by release of large-scale available potential energy, but such a calculation would be cumbersome and likely sensitive to biases in the observations. A good alternative, developed by Bob Hart at Florida State University, describes the evolution of the system in “cyclone phase space”, a plane in which the two axes consist of a measure of storm symmetry and a measure of the relative temperature of the core (warm core vs cold core). The evolution of Hurricane Floyd in cyclone phase space is depicted in Figure 8.16.

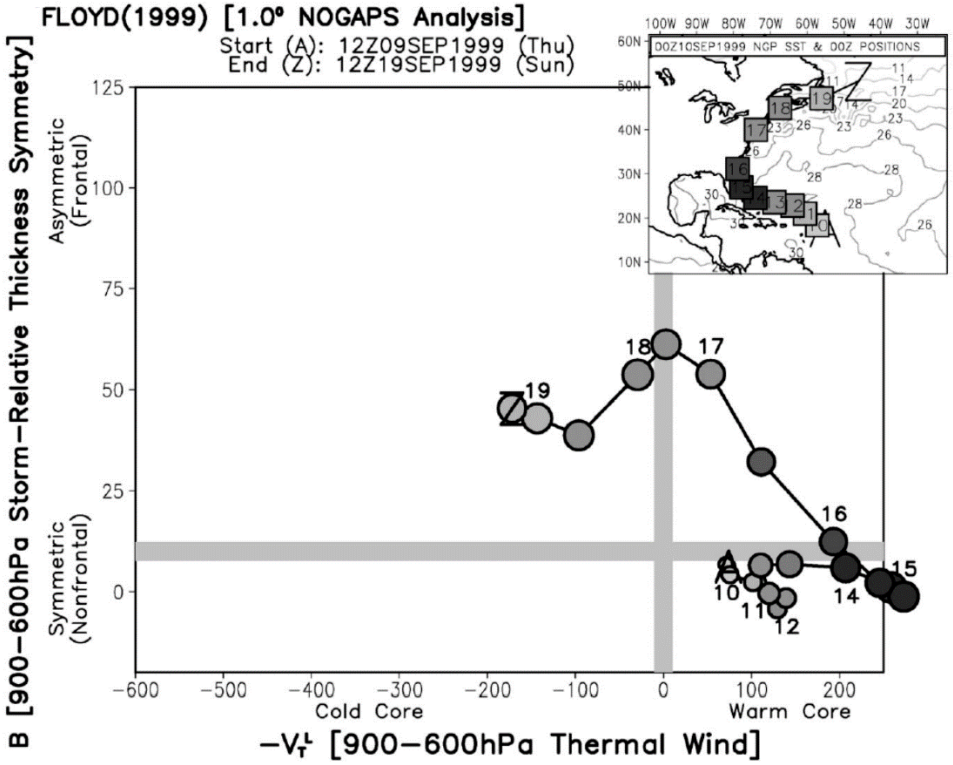


Figure 8.16: Cyclone phase diagram showing the evolution of Hurricane Floyd, whose track (from A to Z) is shown in the inset at upper right. The phase diagram has the magnitude and sign of the environmental thermal wind between 900 and 600 hPa on the abscissa (with negative values indicating a cold core) and a measure of the asymmetry of the low-level thermal field on the ordinate. The trajectory of the storm in this phase space extends from point A (12 GMT on September 9th1999, when Floyd was northeast of the lesser Antilles) to point Z (12 GMT on September 19th). The shading of each marker indicates cyclone MSLP intensity (white, 1010 hPa; black, ,970 hPa) and the size of the circular marker within the phase space indicates the relative size (mean radius) of the 925-hPa gale force (17 ms^{-1}) wind field.

Floyd begins as a pure tropical cyclone, with a symmetric warm core, and remains that way through approximately September 16th, when it was about to make landfall in the Carolinas. On the 16th Floyd rapidly loses axisymmetry, and after the 17th becomes cold core, like a classical baroclinic cyclone.

We shall discuss the dynamics of extratropical transition in more detail in Chapter 12.

8.3.9. Polar Lows, Medicanes, and Agukabams

Thermodynamic conditions over warm tropical oceans are nearly always conducive to the development of tropical cyclones, as demonstrated in Chapter 15. The strong thermodynamic disequilibrium between the atmosphere and the surface, necessary to sustain tropical cyclones, is usually absent outside the tropics, but there are occasions when deep masses of cold air move over relatively warm water and thermodynamic conditions become temporarily favorable for flux-driven vortices to develop.

Wintertime outbreaks of deep, cold air over relatively warm water occur regularly in the far North Atlantic east of Greenland and over the Norwegian and Barents Seas, and in the northern reaches of the North Pacific region, including the Sea of Japan. Polar lows have also been detected in the Southern Ocean near Antarctica. The more severe polar lows are usually accompanied by deep, cut-off cyclones in the middle and upper troposphere, which provide low-shear environments near their centers. The strong thermodynamic disequilibrium between the deep, cold arctic air and the relatively warm sea surface in these regions (where they are free of sea ice) provide favorable conditions for tropical cyclone-like development. Since these regions can be highly baroclinic as well, it is not always easy to distinguish surface flux-driven cyclones from conventional baroclinic cyclones, and indeed there is likely a continuum between these two end-member cyclone types. In models, and to some extent in reanalysis data sets, it is possible to quantify baroclinic and surface-flux driven energy sources, but the quantity of observations in these regions is generally insufficient for quantitative analysis and inferences about the physical mechanism are based mostly on the appearance of the vortices in satellite imagery and on assessment of the thermodynamic potential for flux-driven cyclones from nearby soundings, if they exist, or operational analyses or reanalyses.

A visible satellite image of a polar low over the Sea of Japan in December, 2009 is shown in Figure 8.17a. Note the strong semi-circular ring of deep convection surrounding a clear eye, and the general environment of moist convection as cold air swept out over the relatively warm ocean from the Eurasian land mass. As with most polar low cases, this one was located nearly under a deep cutoff cyclone in the upper troposphere (Figure 8.17b), signifying a broad, deep pool of anomalously cold air.

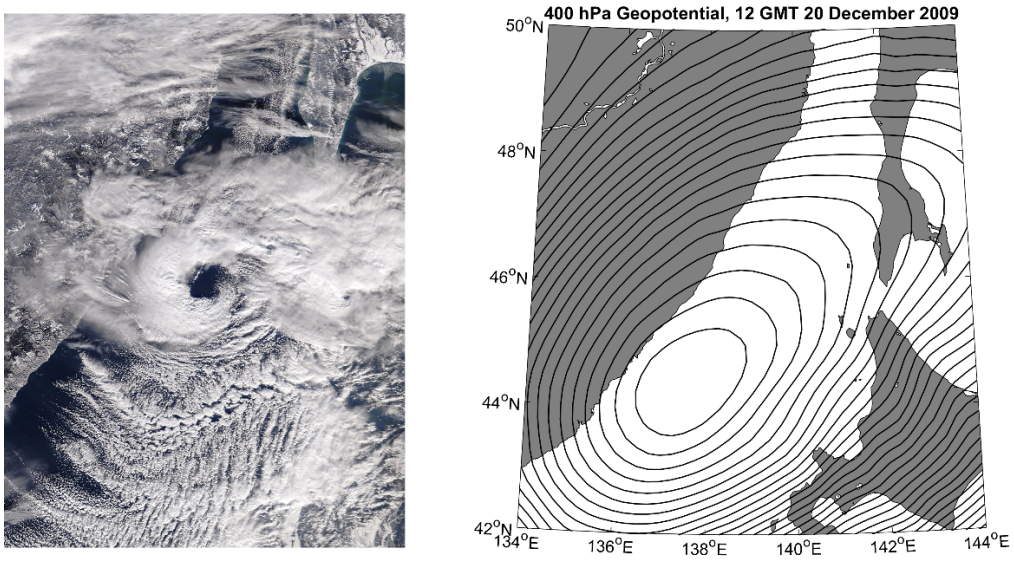


Figure 8.17: a) Terra satellite image at 02:00 GMT 20 December 2009 showing a polar low over the northern Sea of Japan. b) 400 hPa geopotential at 12:00 GMT showing a minimum over the northern Sea of Japan

Figure 8.18a shows the density of polar lows identified in numerical simulations made using a regional model driven by boundary and initial conditions provided by a reanalysis data set (Zhan and von Storch, 2008). The polar lows were identified objectively using criteria based on the sea level pressure perturbation and the difference between the ocean temperature and the temperature at 500 hPa, the latter insuring that the thermodynamic environment could support flux-driven vortices. Similarly, Figure 8.18b shows the origin points and tracks of polar lows over the Sea of Japan identified in a high resolution reanalysis (Yanase and co-authors, 2016). In both cases, the activity peaks in mid-winter, when the air-sea temperature contrast is greatest during cold-air outbreaks.

A phenomenon similar to polar lows has been observed over the Mediterranean Sea during the colder months. These are relatively rare, with an annual probability of around 1.5, and they develop under circumstances nearly identical to those of conventional polar lows…usually under deep, cut-off cyclone aloft. As with polar lows, the enhanced air-sea temperature contrast and low wind shear near the center of such cut-off cyclones provides a conducive environment for flux-driven vortices. Likewise, medicanes are most common during the winter months when the contrast between the air temperature in cold air outbreaks and the sea surface temperature is greatest. The incidence of medicanes is small enough that climatologies made over the course of the satellite era tend to be noisy (Cavicchia et al., 2014), and so in Figure 8.19 we present a climatology produced by downscaling ERA-40 reanalyses over the period 1981-2000 (Romero and Emanuel, 2013). Medicanes favor the central Mediterranean but can occur throughout the region and in the Black Sea as well.

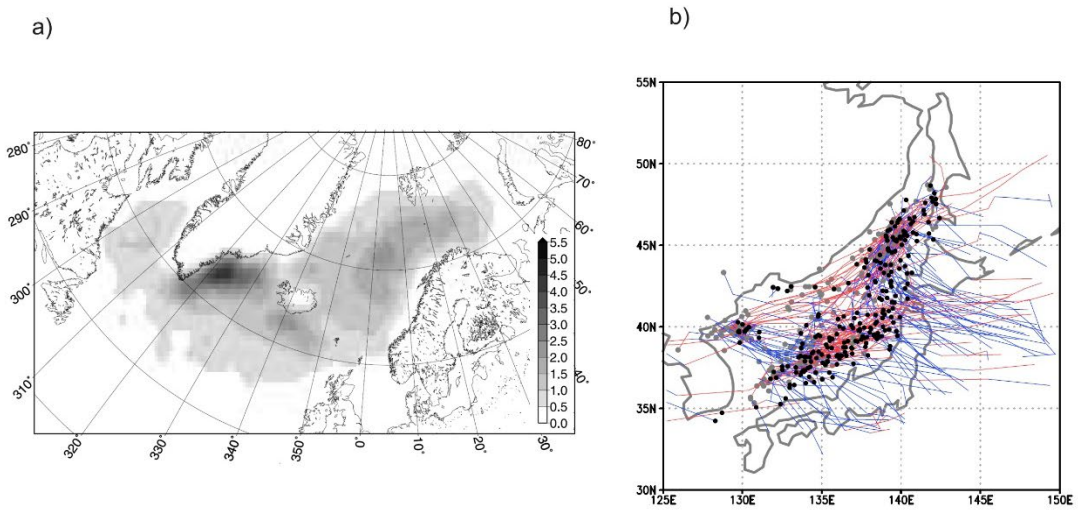


Figure 8.18: a) Polar lows per 250 km^2 in the North Atlantic, downscaled from NCAR/NCEP reanalyses over the period 1948-2006. b) Origin points and tracks of polar lows over the Sea of Japan identified in JRA-55 reanalyses over the period 1979-2015.

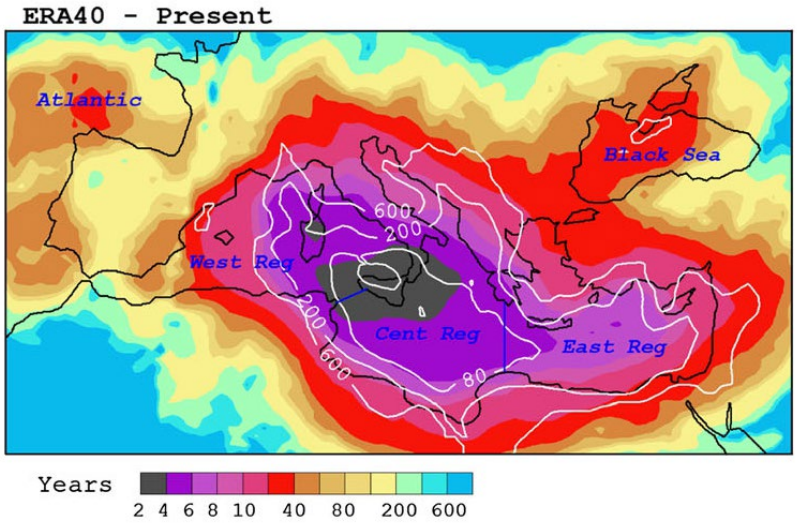


Figure 8.19: Return periods for the present climate of medicane surface winds above 34 kt (color) and 60 kt (white contours, for the values shown in the color scale).

Even rarer than medicanes are *agukabams* (a word constructed from the Australian aboriginal Ayapathu word roots “agu,” meaning land, and “kabam,” meaning storm), also known as *landphoons*. These are tropical cyclone-like storms that develop over land in the absence of appreciable baroclinic forcing (active during extratropical transition) and may be powered by surface enthalpy fluxes made possible when hot sand is wetted by precursor rains, temporarily increasing the thermal conductivity of the soil to the point that appreciable heat can escape to the atmosphere (Emanuel et al., 2008). This appears to happen most frequently over the deserts of northern Australia, where relatively dark soils can absorb much sunlight in summer. It

is possible that some cases of inland intensification elsewhere are also owing to unusually large heat fluxes from the soil; there is some indication that Tropical Storm Erin of 2007 re-intensified over Oklahoma (see Chapter 13, section 13.5) because of enhanced surface enthalpy fluxes (Evans et al., 2011). As of this writing, no formal climatology of agukabams has been published. Figure 8.20 shows a TRMM image of informally named Agukabam Abigail over northern Australia in March, 2001. The storm had been over land for almost 4 days when this image was made, and yet it still has the structure of a developed tropical cyclone (Emanuel et al., 2008). Figure 8.21 shows the track and evolution of an agukabam in northern Australia in January of 2006 (Tang et al., 2016). Observations of the storm make it clear that it intensified while over land.

We will have more to say about polar lows, medicanes, and agukabams in Chapters 12 and 13.

8.4. Structure

8.4.1: Overview

Tropical cyclones are complex bounded vortices, with cyclonic flow at low levels decaying with altitude and becoming anticyclonic near their tops, which are generally close to the tropopause. A broad overview of the structure of clouds and airflow in a fully developed northern hemisphere tropical cyclone is displayed in Figure 8.22. A central core, called the *eye*, generally free of clouds and becoming broader with altitude, is surrounded by an outward slopping *eyewall*, a dense concentration of cumulonimbus convection underneath which the strongest winds and heaviest rains of the storm are found. Outside the eyewall, deep convection is organized into bands which may take the form of *concentric eyewalls* (also referred to as *secondary eyewalls*) – nearly annular bands of deep precipitating cumulonimbi more or less centered at the storm center – or *spiral rainbands*, which are clearly evident in satellite and radar imagery. For example, Figure 8.23 is a satellite image of Hurricane *Ivan* of 2004 as it approached the Gulf Coast of the U.S. in 2004. Several spiral bands are visible outside the primary eyewall cloud, and the eye itself is prominent. Figure 8.21 has been constructed from reflectivity data from the ELDORA (ELectra DOppler RAdar) mounted on the Naval Research Laboratory Lockheed P-3 Orion aircraft. This is an X-band radar, emitting pulses of radiation of wavelength 8.2 cm that backscatter from precipitation particles. The reflectivity, corrected for distance from the radar, is a rough measure of the concentration of rain. The eye, indicated by white (no radar returns), is at the center of this image and is surrounded by an almost-closed reflectivity maximum (the eyewall). Outside the primary eyewall is a pronounced reflectivity minimum, sometimes referred to as the *moat*. In this particular case, a very pronounced secondary eyewall is present outside the moat, made visible by a nearly closed maximum of radar reflectivity. Spiral bands of radar reflectivity begin on the north and south sides of the secondary eyewall and spiral outward from there.

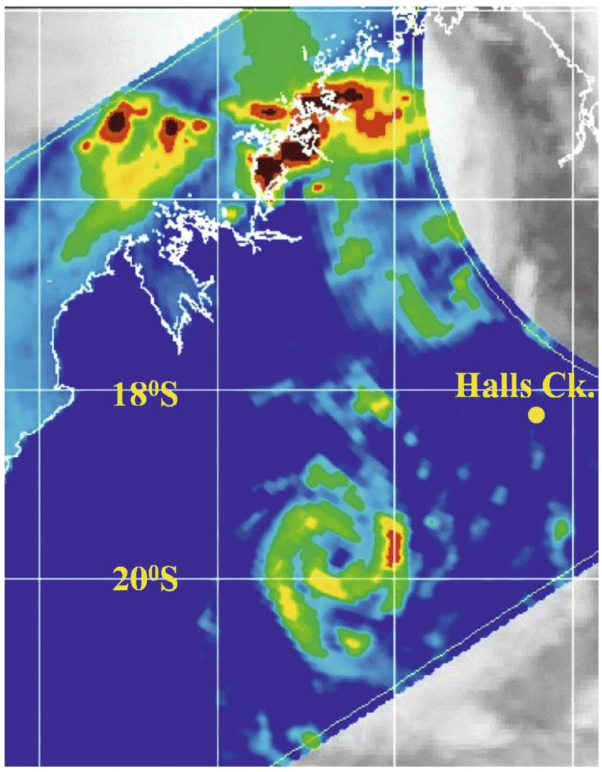


Figure 8.20: A Tropical Rainfall Measuring Mission (TRMM) 85-GHz (horizontally polarized) microwave image of Agukabam Abigail in northwestern Australia, at 1903 UTC 2 March 2001. From Emanuel et al. (2008).

It has become conventional to divide the tropical cyclone into three regions: The eye, the core, and the outer region, as illustrated in Figure 8.22. The eye is the cloud-free region at the center of well-developed tropical cyclones. The core includes the primary eyewall and the secondary eyewall (if there is one), while the outer region includes the region may have shallow convection and extends out to the outer limits of the storm's surface circulation. We will see later, in Chapter 9, that based on tropical cyclone physics, a case can be made for dividing the core into two regions: An inner core, which is nearly saturated on the mesoscale, and an outer core in which the mesoscale is unsaturated but deep convection is present.

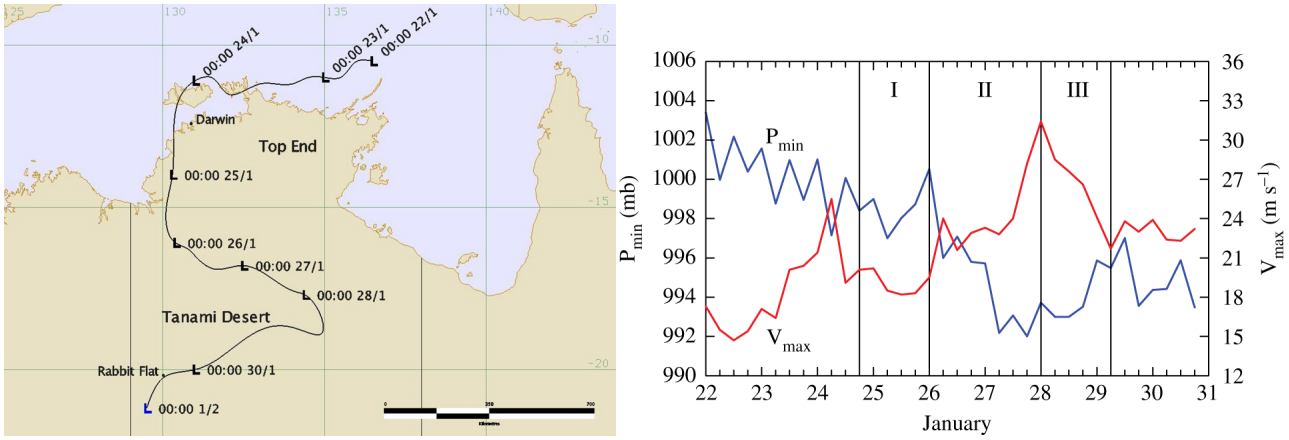


Figure 8.21: a) Track through northern Australia of agukbam of January, 2006. b) Time series of the minimum sea-level pressure, P_{min}, and maximum horizontal total wind speed at the height of 850 mb, V_{max}, from European Center for Medium Range Forecasts analyses. From Tang et al. (2016).

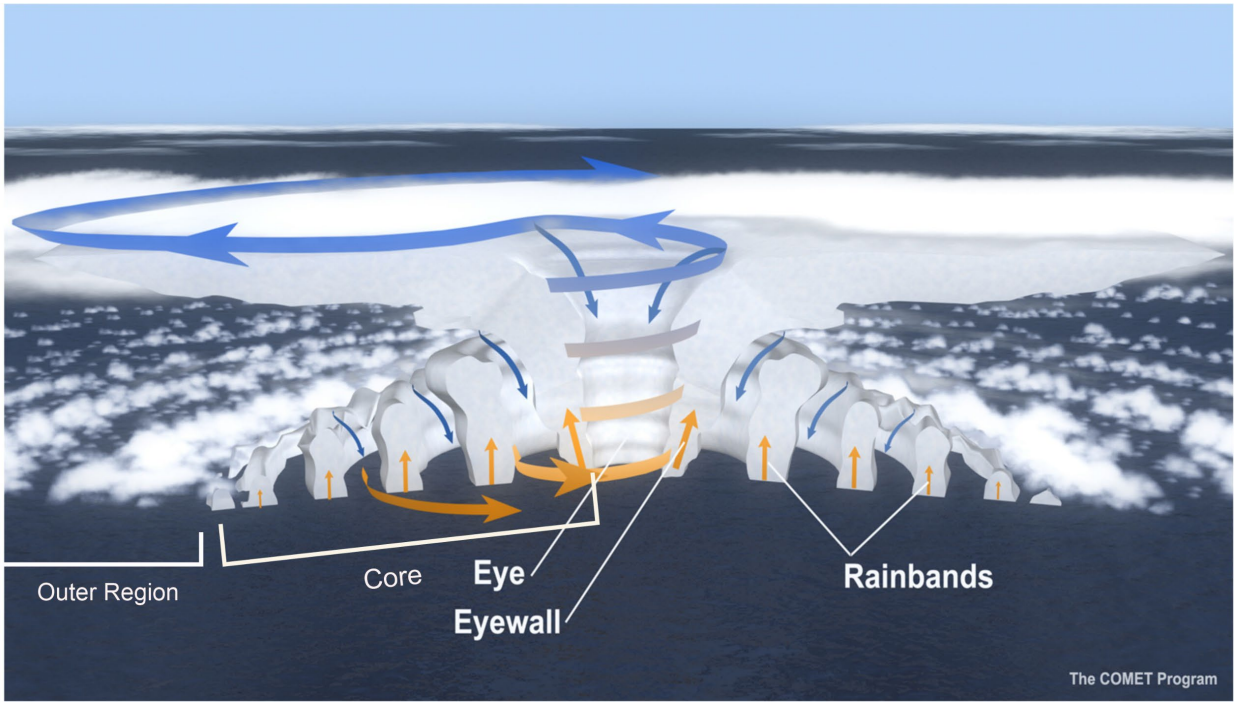


Figure 8.22: Overview of clouds and airflow in a northern hemisphere tropical cyclone. The diagram represents a cross section of roughly 500 km in diameter and 18 km in altitude.

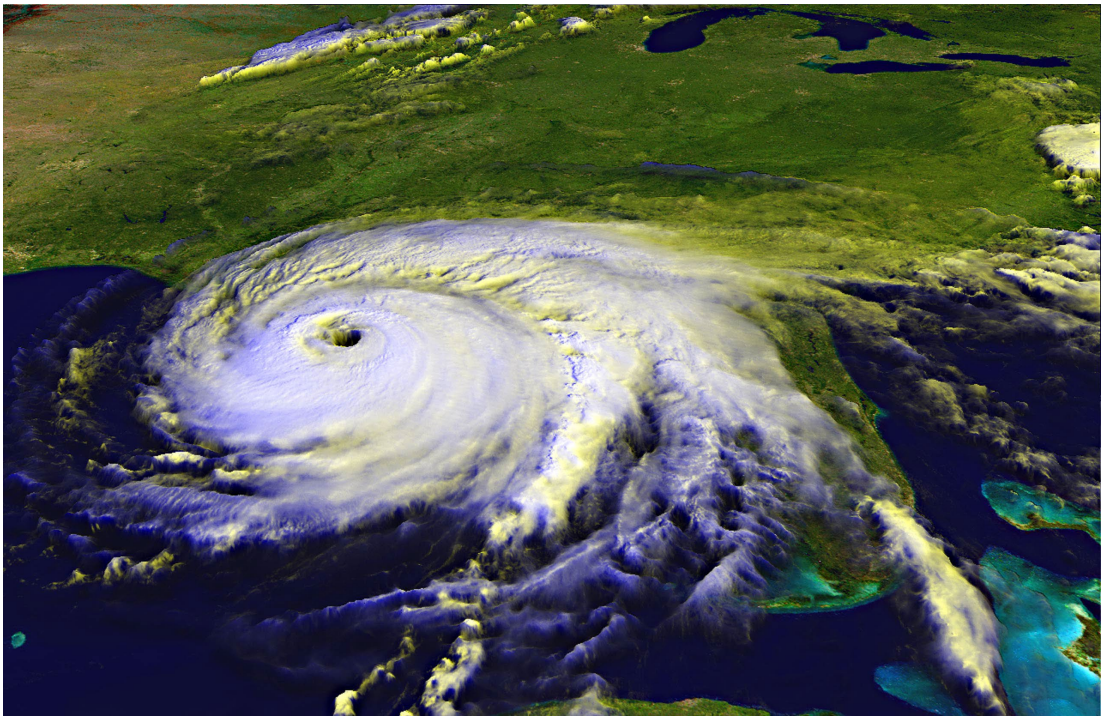


Figure 8.23: Hurricane Ivan in the northern Gulf of Mexico on 16 September 2004.

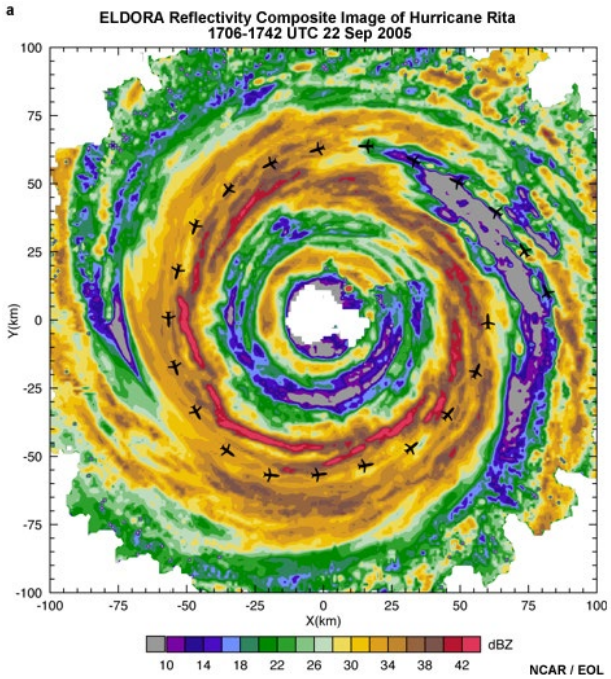


Figure 8.24: Radar reflectivity composite image of Hurricane Rita taken by an airborne radar, 22 Sep 2005, during the Hurricane Rainband and Intensity Change Experiment (RAINEX). The flight track is marked by airplane-shaped markers.

Not all tropical cyclones are as circularly symmetric as the cases shown in Figures 8.23 and 8.24. Figure 8.25 displays a satellite image of Subtropical Storm Ana on May 8th, 2015. It is hard to define visually where the center of the storm is, and most of the deep convection is located south of what appears to be the circulation center. The surface winds and pressure can also be strongly asymmetric. We will see in Chapter 12 that the interaction between localized potential vorticity anomalies and ambient vertical shear of the horizontal wind is often the source of asymmetries, and most tropical cyclones exhibit varying degrees of asymmetry at different times during their lives.

Figure 8.26 displays a vertical cross-section of reflectivity from a radar aboard an airplane located in the eye of the storm. The eye itself is mostly free of radar returns, except within a few hundred meters of the sea surface, where light precipitation from shallow clouds and perhaps sea spray provide some reflection. The eye is surrounded by very deep, outward-sloping intense radar echoes comprising the eyewall. Allowing for the vertical exaggeration of the graphic, the eyewall slope is roughly 1:1. Stratiform rain from dense clouds in the upper tropospheric outflow is particularly evident on the right side of the diagram, and precipitation from shallow convective clouds can be seen outside the eyewall.

Tropical cyclones usually develop visible eyes when their peak winds reach 30-40 ms^{-1} and lose them during rapid dissipation or when the peak surface winds become sufficiently weak.

A view of the eye of a hurricane from the vantage of a hurricane reconnaissance aircraft is shown in Figure 8.27. The deep blue sky of the stratosphere is visible above the spectacular, outward sloping eyewall, and low cumulus and stratocumulus clouds can be seen in the base of the eye. The whole view resembles that of an enormous white coliseum. The diameters of the visible eyes of well-developed hurricanes span two order-of-magnitude, from a few kilometers to several hundred kilometers.

The strongly cyclonic low-level tropical cyclone flow is associated with an intense, localized surface pressure minimum as shown in Figure 8.28. The pressure gradients across the eyewall can be extreme; for example, a gradient of 2.5 hPa/km was measured in Typhoon *Haiyan* just after it made landfall in the Philippines in 2018.

8.4.2 Kinematic and thermodynamic structure

The airflow and thermodynamic properties of tropical cyclone have been measured by research aircraft beginning in the mid-1940s. These aircraft usually deploy instruments on parachutes known as dropwindsondes; today these carry GPS chips and are tracked, yielding estimates of the horizontal wind, while sensors on the sondes transmit information about temperature, humidity, and pressure back to the aircraft. Today, highly accurate measurements of airflow in the core of the storm can be made using airborne Doppler radar, which measures the velocity of the reflectors towards and away from the aircraft. By viewing the same reflectors from different angles as the aircraft flies past, two components of the velocity can be obtained and, after correcting for the fall speed of the reflectors, the third wind component may be deduced from mass continuity.

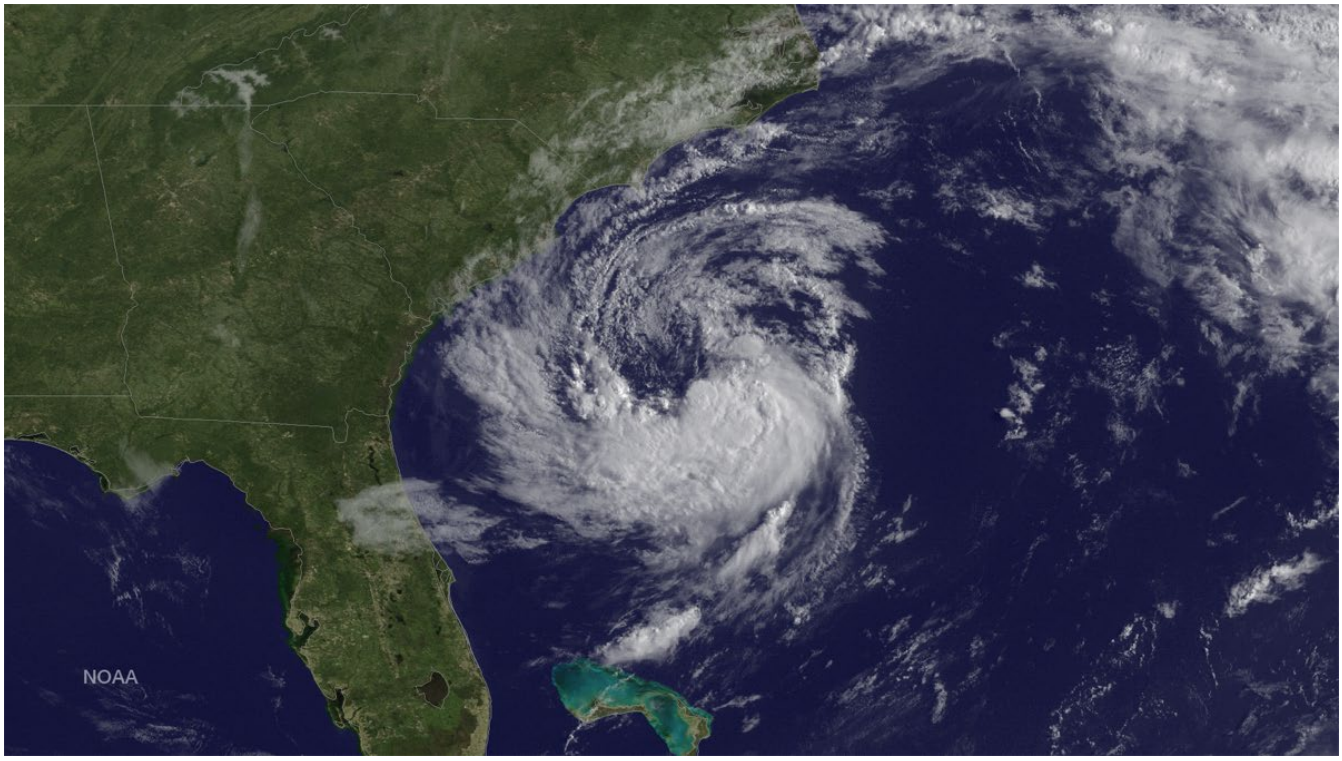


Figure 8.25: Subtropical Storm Ana developing off the U.S. southeast coast on May 8, 2015, as observed by the GOES east satellite.

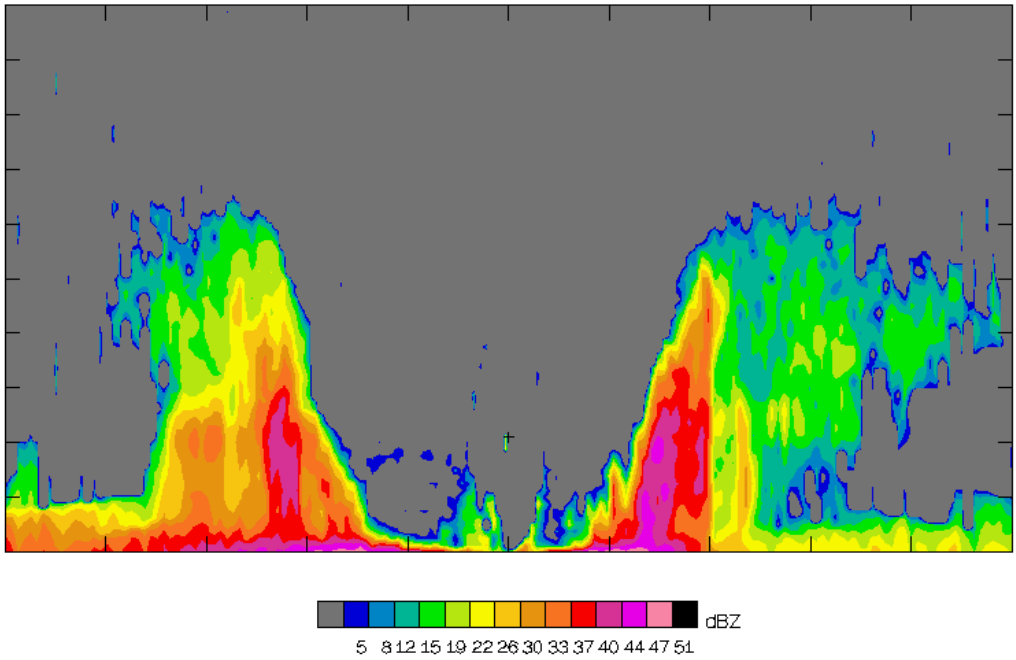


Figure 8.26: Vertical cross section of radar reflectivity in Hurricane Floyd of 1999, made from hurricane reconnaissance aircraft in the eye, located at the "+" sign. The diagram spans 20 km in height and is 120 km across.

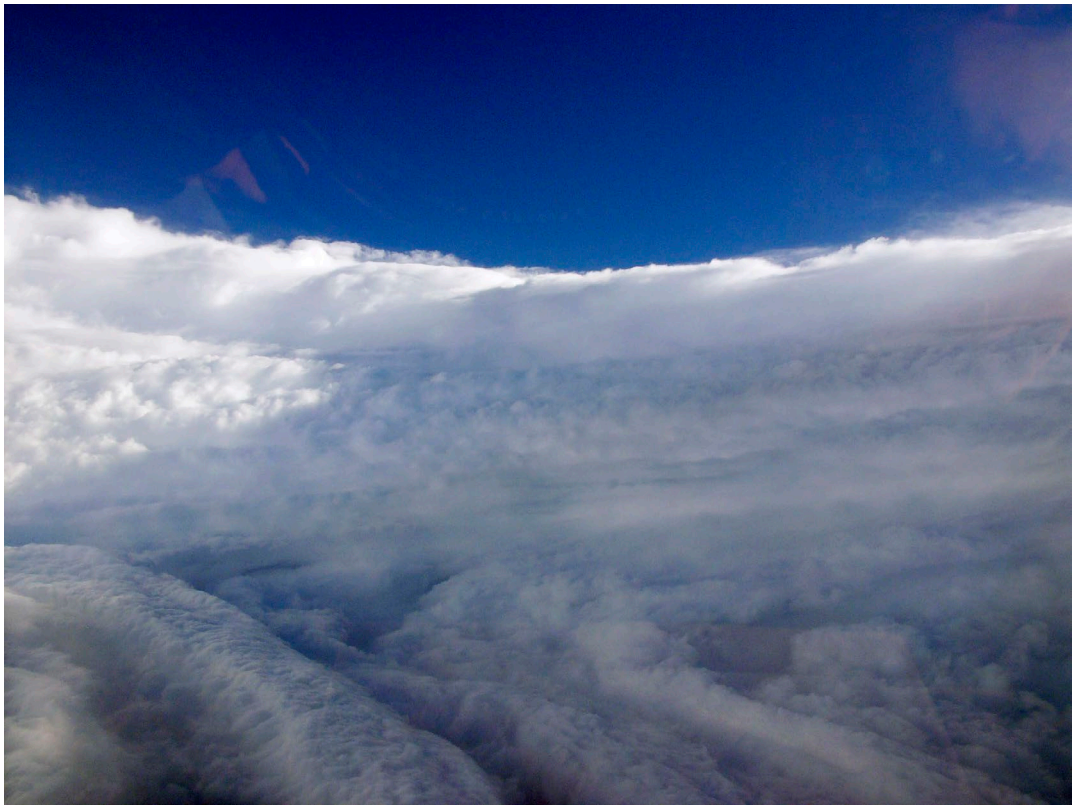


Figure 8.27: View of the eye of Hurricane Katrina on August 28th, 2005, as seen from a NOAA WP-3D hurricane reconnaissance aircraft.

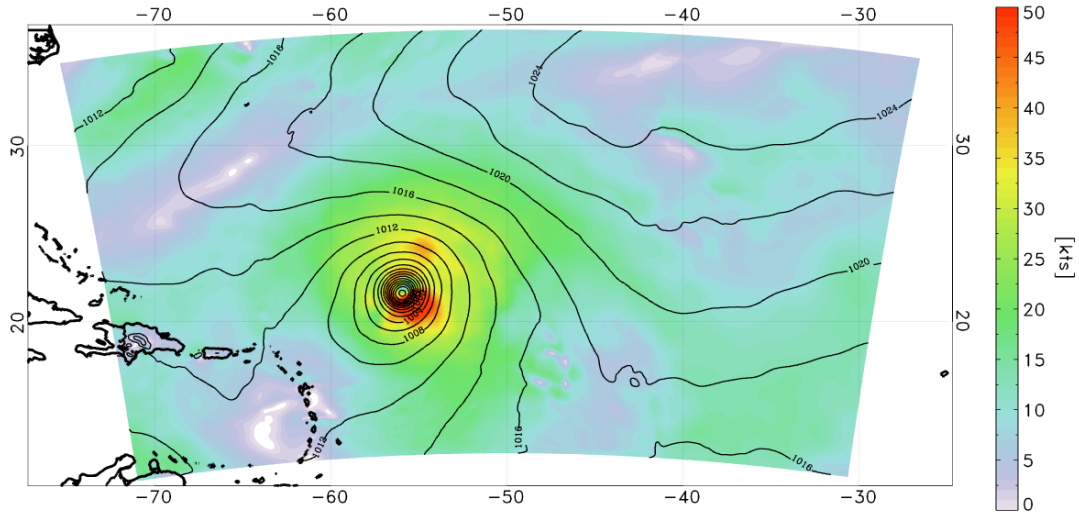


Figure 8.28: Surface pressure distribution associated with a numerically simulated tropical cyclone in the North Atlantic.

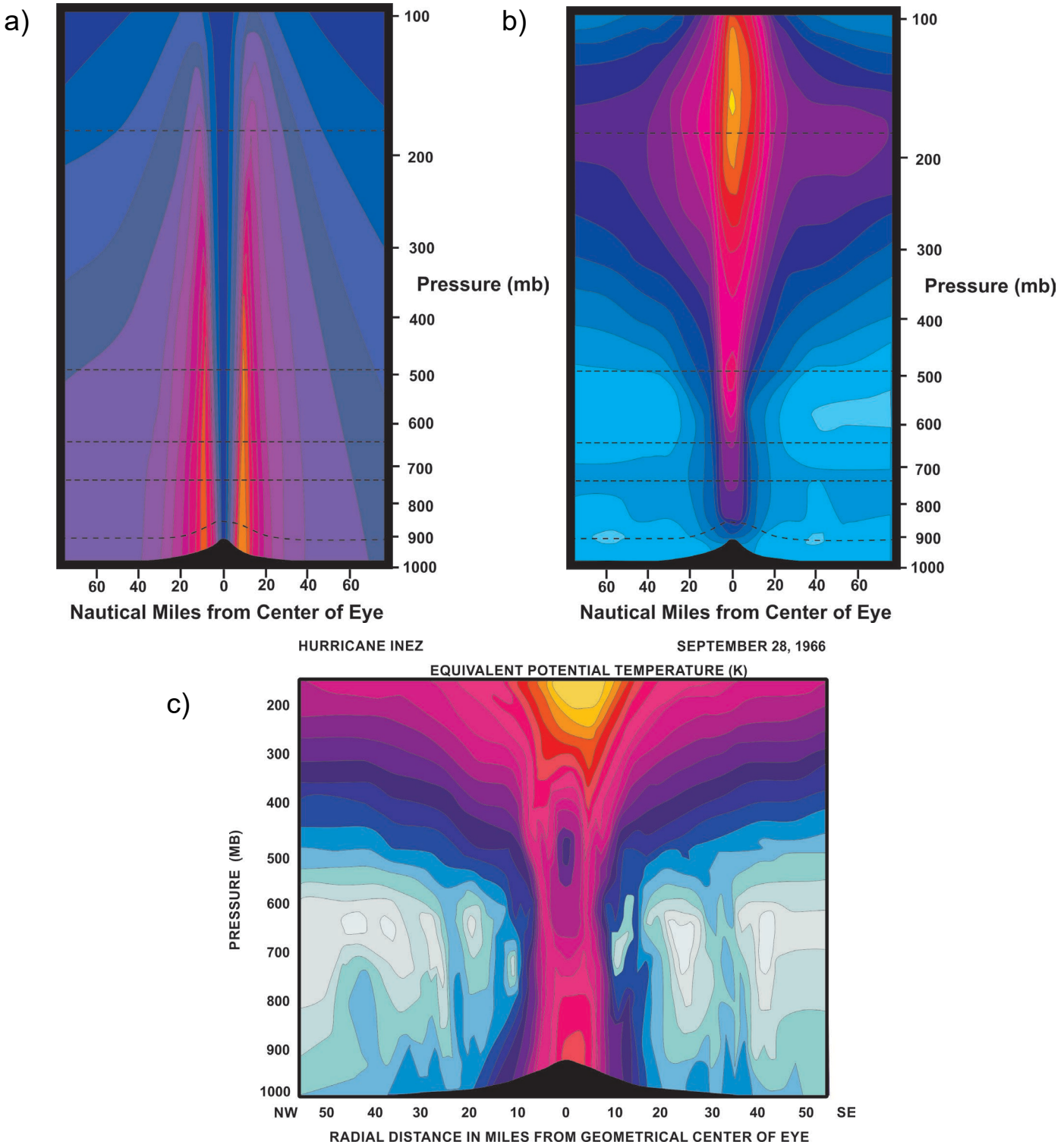


Figure 8.29: Distribution of various quantities in Atlantic Hurricane Inez of 1966. a) azimuthal velocity, from 10 knots to 130 knots with a contour interval of 10 knots; b) temperature anomaly from environment at the same altitude, from 0 K to 16 K with a contour interval of 1K; c) equivalent potential temperature from 336 K to 376 K with a contour interval of 2 K. Cross-sections constructed from data collected by aircraft along the dashed lines in a) and b) and from dropsondes deployed therefrom.

Cross-sections of wind, temperature, and equivalent potential temperature through an Atlantic hurricane are displayed in Figure 8.29. These have been made from measurements aboard a research aircraft that traversed five levels, shown by the dashed lines in Figures 8.29 a) and b), supplemented with measurements made from dropwindsondes deployed from the aircraft.

The strongest winds are quite close to the surface (we will have more to say about this and other aspects of the boundary layer in Chapter 9) and decay upwards toward the tropopause. The azimuthal wind increases somewhat linearly from the central axis outwards to the radius of maximum winds, and close inspection of Figure 8.29a reveals that the RMW slopes outward with height.

Winds decay gracefully away from the RMW, at first roughly following a $r^{-1/2}$ power law, but further out the decay rate somewhat faster and the storm winds vanish at a finite radius (Chapter 9).

There are hardly any variations in air temperature along the surface (Figure 8.29b), though more recent and detailed measurements of the boundary layers of tropical cyclones tend to show some cooling near the RMW, perhaps owing to evaporation of sea spray and rain, direct sensible enthalpy transfer by rain falling through the vertical air temperature gradient, and some cooling due to mixing of the upper ocean (Chapter 13). But aloft the tropical cyclone has a surprisingly hot core, with a peak anomaly of 16 K over the undisturbed environment in the case of Hurricane Inez. The highest temperature anomaly is usually located on the central axis near the level of the ambient tropopause. We will see in Chapter 9 that the warmth of the eyewall directly reflects the enthalpy gained from the ocean by inflowing boundary layer air, while the extraordinary temperatures of the eye itself arise from mechanically forced subsidence, reflecting a small reverse heat engine operating within the eye. The warm-core character of tropical cyclones implies, through hydrostatic balance, that the very low surface pressure anomaly associated with the storm decays with altitude; in fact, a broad anticyclone is usually present close the storm top away from the center.

The distribution of equivalent potential temperature (θ_e) in Hurricane *Inez* is shown in Figure 8.29c. This is a quantity whose logarithm is proportional to the entropy of a mixture of dry air, water vapor, and condensed phase water, all assumed to be in thermodynamic equilibrium with each other (Chapter 9). It is conserved for reversible transformations regardless of phase changes of water; its main source in a tropical cyclone is enthalpy transfer from the surface to the atmosphere, which is the driver of the tropical cyclone heat engine. This source is reflected in the strong radially inward gradient of θ_e near the storm center. On the scale of Figure 8.28c it is reasonable to suppose that θ_e is conserved above the boundary layer, but ultimately, the excess entropy flowing out of the top of the storm is lost by infrared radiation to space. Air flowing radially inward turns upward at the eyewall and, supposing the system to be in a steady state, would flow along the contours of θ_e in Figure 8.29c, gradually turning outward near the storm top.

The impressive high-θ_e chimney in Figure 8.29c is a direct signature of the physics of the storm, a giant machine sucking heat out of the warm tropical surface and expelling it into the cold reaches of the upper troposphere. We will have more to say about this in Chapter 9.

The radial and vertical velocity components in tropical cyclones are generally noisier and thus more difficult to measure than the azimuthal flow, though the radial flow in the boundary layer can be quite strong. For this reason, we display in Figure 8.30 time-averaged radial and vertical velocities from a numerical simulation using a nonhydrostatic (convection-permitting) axisymmetric model (Rotunno and Emanuel, 1987). The model is run on a grid with 8.75 km radial and 312.5 m vertical grid spacing, and the quantities shown in the figure have been averaged over one day of a simulation that has reached a quasi-steady state.

At the most basic level, the secondary circulation of a steady-state tropical cyclone consists of radial inflow, mostly confined to a turbulent boundary layer of a few hundred meters to perhaps 2 km in thickness, intensifying inward to a maximum outside the eyewall and then rapidly decaying to approximately zero just inside the eyewall as the bulk of the flow turns upward into a strong eyewall updraft. In the exceptionally intense vortex shown in Figure 8.30, the time-mean eyewall updraft reaches a peak speed of about 8 ms^{-1} at around 11 km altitude, then decreasing upward as the flow turns radially outward. The outflow in this simulated storm reaches a peak magnitude of about 20 ms^{-1} not far outside the core of the storm. The outflow layer, near the level of the ambient tropopause, progressively thins with radius. In an axisymmetric framework such as used here, the outflow takes much longer to become quasi-steady, if it does so at all, because there are no explicit three-dimensional eddies to add angular momentum to it, to balance the persistent outward advection of the low angular momentum originating in the storm's core.

The simple in-up-and out circulation is augmented by a curious secondary inflow just below the primary outflow layer, visible in Figure 8.30a. This begins outward of 800 km at an altitude of about 14 km, and slopes inward and downward, reaching a peak magnitude of about 5 ms^{-1} just outside the eyewall, at an altitude of about 8 km. The axis of this elevated inflow layer more or less follows a surface of constant absolute angular momentum (Figure 8.28) and may be driven by the evaporation of precipitation falling from dense stratiform cloud in the outflow layer above it, or may be a form of internal Ekman layer driven by the downward turbulent diffusion of low angular momentum air in the outflow region.

Note in Figure 8.30 the very sharp horizontal gradients of both the radial and vertical velocities along the inner edge of the eyewall updraft. In many places, most of the total change from the small values within the eye itself to the maximum value in the eyewall region occurs over a single model grid interval. As we will show in Chapter 10, the lower eyewall is strongly frontogenetical. Note also in Figure 8.30b that the eyewall updraft slopes outward, and that in this finite-difference model, the outward slope is accomplished by a series of outward steps at various altitudes.

Physical systems are most concisely described using conserved variables. We have already introduced one essential conserved thermodynamic variable: the equivalent potential temperature (θ_e; Figure 8.29c). For axisymmetric flows, a very useful conserved dynamical variable is the absolute angular momentum per unit mass, M:

$$M \equiv rV + \Omega \sin\theta r^2, \tag{1}$$

Where r is the radius, V the azimuthal component of the velocity, Ω is the planetary angular rotation rate, and θ the latitude.

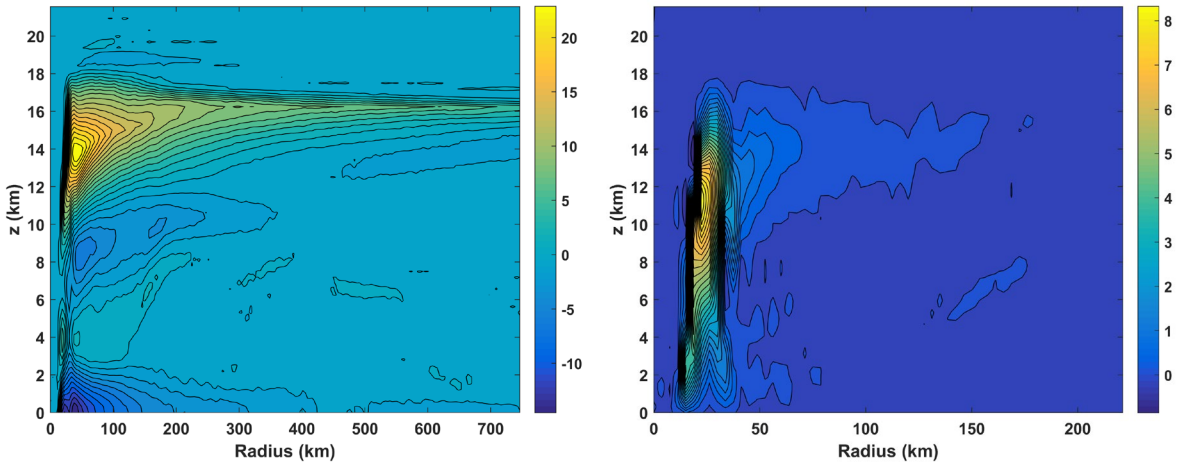


Figure 8.30: Radial-vertical section of radial (a) and vertical (b) velocities (ms^{-1}) averaged over one day during the quasi-steady phase of a tropical cyclone simulated using the axisymmetric, nonhydrostatic model of Rotunno and Emanuel (1987). Note that the radial scale in (b) is roughly one third that of (a).

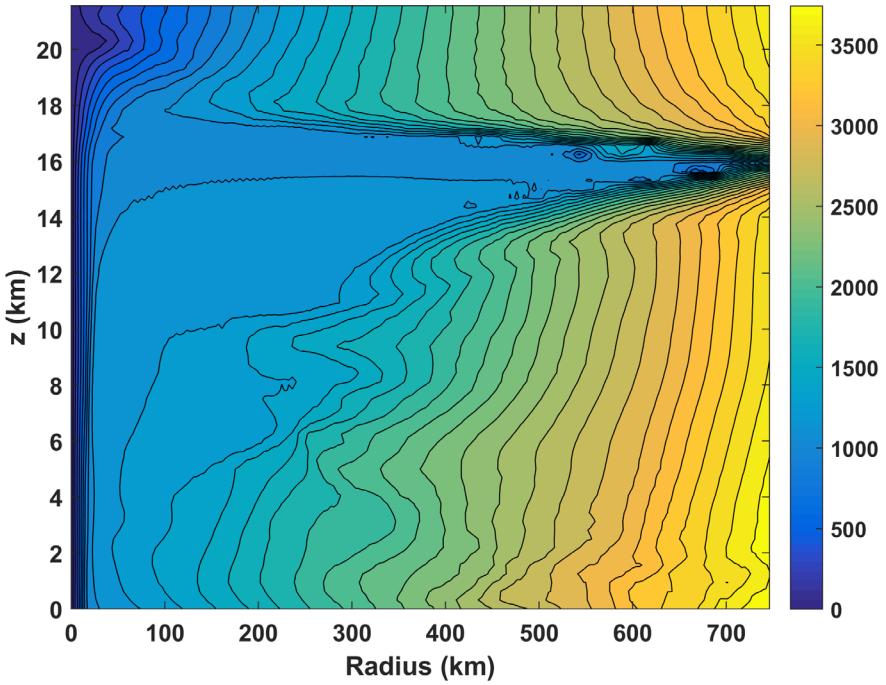


Figure 8.31: Radial-vertical section of the distribution of the square-root of the angular momentum given by equation (1).

The distribution of angular momentum in the numerically simulated storm used to construct Figure 8.30 is displayed in Figure 8.31. (Here the square root of M is shown to highlight detail in the inner core region.) Bearing in mind that the undisturbed angular momentum surfaces are vertical cylinders centered at the storm center, the inward deflection of the surfaces reflects the cyclonic winds of the storm. The most prominent feature of the angular momentum distribution is the outward protrusion of low values of M near the tropopause, corresponding to strongly anticyclonic flow near the storm top.

The strong radial inflow in the boundary layer clearly crosses M surfaces, and assuming that the flow is steady, this reflects the loss of angular momentum owing to the torque exerted by the surface upon the cyclonic flow. By the time the inflow reaches the eyewall, it has lost much of its original angular momentum, and this low M air then flows up the eyewall and out near the tropopause. Assuming that the azimuthal flow at the top of Figure 8.31 (in the lower stratosphere, around 22 km) nearly vanishes, and following the M surfaces downward to the outflow layer, one can see that they deflect inward into the outflow inside of about 400 km radius, corresponding to cyclonic flow, and outward outside 400 km, indicating anticyclonic flow. The thin but powerful anticyclonic lens near the tropopause is a highly prominent feature of developed tropical cyclones. In the real world, the anticyclone is usually displaced from the cyclone beneath it by environmental wind shear, and the outflow can be highly three-dimensional, often focused into curved jets. Figure 8.32 shows outflow streamlines in two numerically simulated tropical cyclones, one in a calm ambient environment and another in which there is both horizontal and vertical shear of the ambient horizontal wind (Rappin et al., 2011). In the first case, the outflow is quite axisymmetric, but when ambient shear flow is present, much of the outflow is focused into an anticyclonically curved jet north of the storm center. Flow at upper levels in association with a western North Pacific typhoon is shown in Figure 8.33. In this case, the outflow is focused into two anticyclonically curved jets, one north and one south of the storm center. The strong interaction of tropical cyclone outflow with other weather systems is evident in this figure, and is typical of real tropical cyclones, especially as they move into higher latitudes.

8.4.3 The eye and eyewall

The circulation in the eye of a tropical cyclone is especially unusual and interesting. Figure 8.34 shows a sketch of the secondary airflow in and near the eye of a mature storm, based on observations and numerical simulations. The primary circulation is cyclonic, with azimuthal velocity increasing outward to its maximum value in the eyewall. During rapid intensification, the azimuthal velocity can be near zero through much of the eye, increasing abruptly just inside the eyewall. In this case, the absolute value of the vertical component of vorticity has a sharp peak in the eyewall, and as we shall review in Chapter 11, this flow arrangement can be highly unstable to three-dimensional eddies, which are known to exist in real storms and often made visible by shallow clouds in the eye (Figure 8.35).

When the storm is nearly steady, the temperature in the eye is highly elevated, decreasing radially outward from a maximum on the central axis. Air gently subsides in the center of the eye, with compressional heating balanced by infrared radiative cooling to space. Subsidence velocities are two orders of magnitude smaller than ascent rates in the eyewall, averaging around $1\,cm\,s^{-1}$. Stronger subsidence is sometimes found just inside the eyewall, where

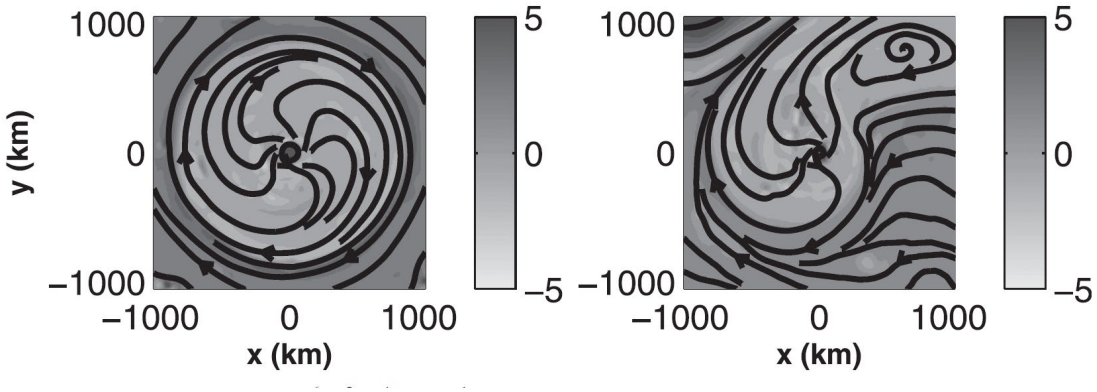


Figure 8.32: Streamlines (black) and potential vorticity (gray-shaded) at 13 km altitude in numerical simulations of tropical cyclones with constant Coriolis parameter in a calm environment (left) and in an environment with an ambient west-to-east jet stream north of the cyclone center (right). From Rappin et al. (2011)

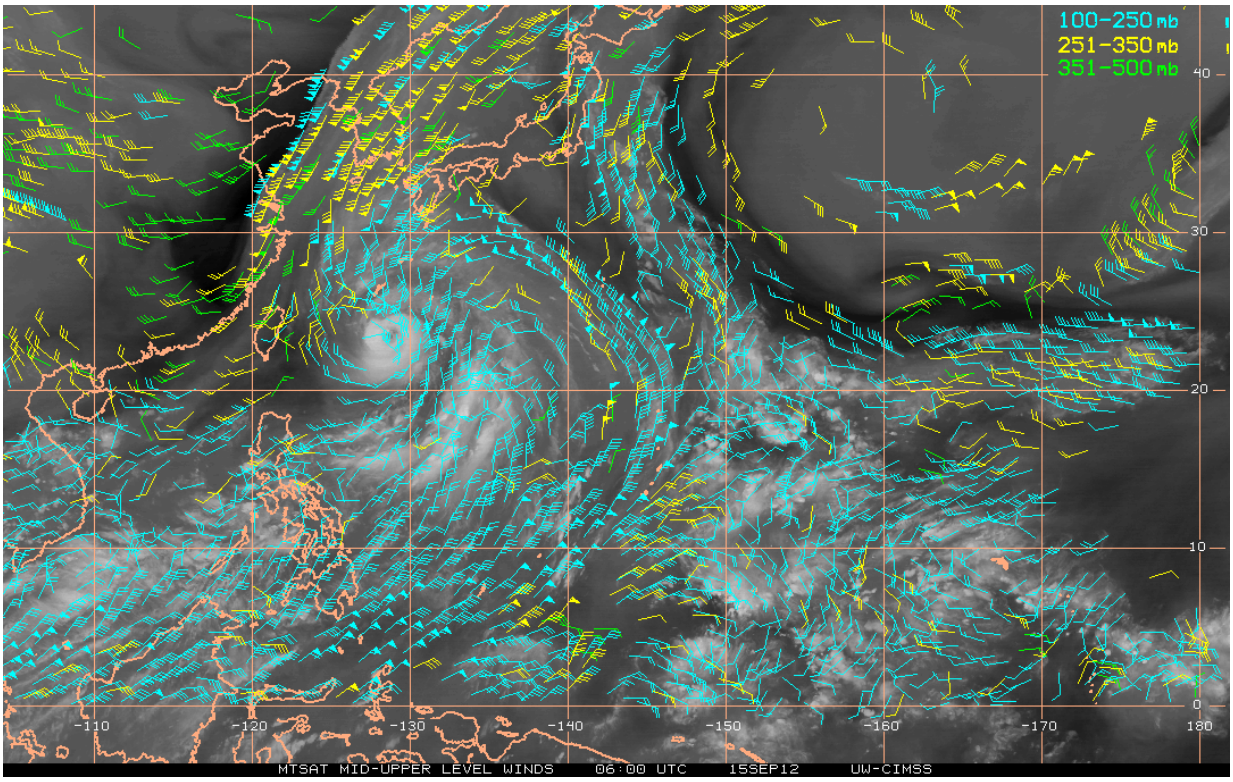


Figure 8.33: Airflow at three levels, indicated in the upper right, at 06 UTC on 15 September 2012 over the western North Pacific in association with Typhoon Sanba seen in the center left portion of the image. Note that the outflow is concentrated into two, anticyclonically curving jets.

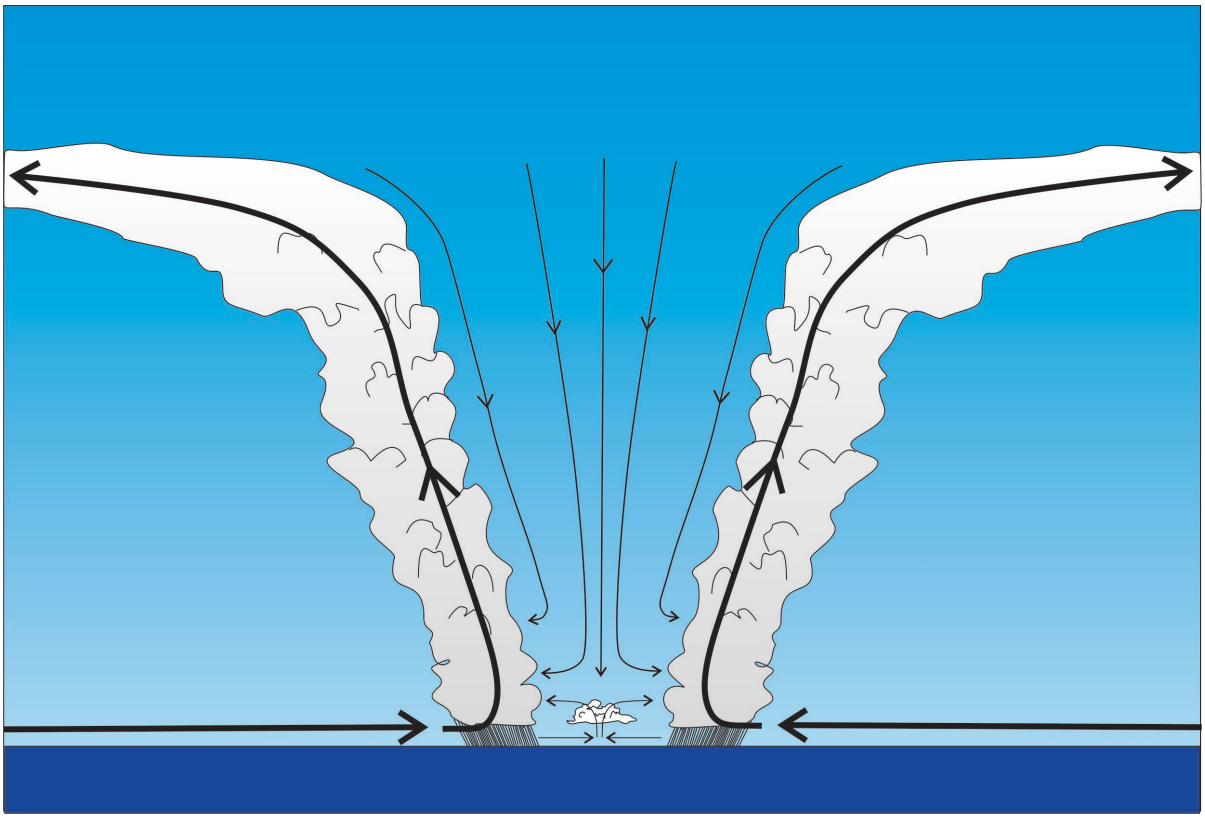


Figure 8.34: Schematic cross-section of airflow in the eye of a mature tropical cyclone. Bold arrows denote the main secondary circulation while the thin arrows denote airflow in the eye itself.

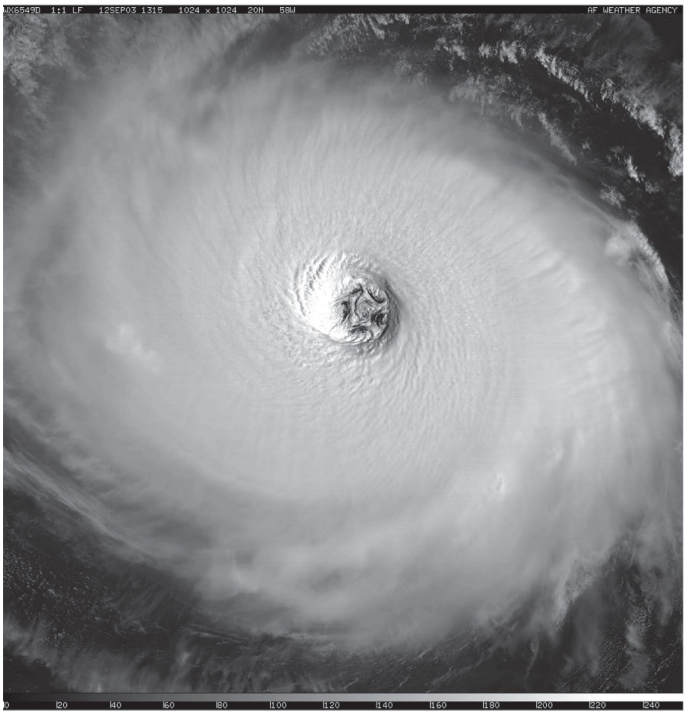


Figure 8.35: View of the eye of Atlantic Hurricane Isabel on 12 September, 2003, from a Defense Meteorological Program satellite

surfaces of constant potential temperature slope upward in association with negative radial temperature gradients, allowing air to descend along inward-sloping trajectories, more nearly adiabatically than is possible in the center of the eye.

The eyewall eddies also transport cooler air and hydrometeors into the eye, allowing the compressional heating to be balanced by turbulent enthalpy fluxes and evaporative and radiative cooling. Descent rates near the inner edge of the eyewall can reach many $cm\,s^{-1}$.

Because the boundary layer airflow is cyclonic, the torque exerted on it by surface friction causes an inward Ekman flow, just as in the eyewall and outer regions of the storm. But the very large inertial stability of the eye means that only a small radial velocity (a few ms^{-1}) is needed to balance the surface torque. Convergence of this radial Ekman flow leads to a weak, shallow updraft often made visible by a formation of shallow cumulus or stratocumulus clouds near the central axis; this is sometimes referred to as the *hub cloud*.

Where the shallow, relatively cool updraft meets the hot, downward flowing air, remarkably strong temperature inversions develop. The two air streams turn radially outward, toward the eyewall. The radial outflow crosses tightly packed surfaces of constant angular momentum; this is made possible by the inward transport of angular momentum by three-dimensional eddies.

The eye circulation above the boundary layer is thermally indirect, with relatively warm air sinking, and constitutes a giant reverse heat pump, driven mechanically by inward eddy transport of eyewall angular momentum, as we will discuss in detail in Chapter 9.

During episodes of rapid intensification, the eye itself is gradually spun up by three-dimensional eddies, accompanied by descent rates that may be substantially stronger than those associated with the steady state. The temperature inversion in the eye may at such times descend to quite low levels in response to the strong subsidence. Conversely, during periods of rapid weakening (for example, after the storm passes over land) the angular velocity of the eye may temporarily exceed that of the eyewall, causing an outward eddy flux of angular velocity and a reduced descent or even ascent in the eye. During such times, the boundary layer can deepen dramatically and the eye may fill with cloud.

We have referred many times to the eyewall without really defining it. It derives it name from the almost discontinuous circular wall of cloud seen from inside the eye (e.g. Figure 8.27), and it is the locus of strongest azimuthally averaged ascent in the storm (though local convective updrafts are usually strongest in spiral bands; see subsection 8.4.4). Broadly speaking it is an annulus of nearly saturated air, in which the important properties, like all three velocity components, temperature, and humidity can be nearly invariant with azimuth in strong storms, with individual convective cells hard to distinguish (see Figure 8.24). The region of nearly saturated air we refer to as the eyewall is generally thinner than the distance between its inner edge and the storm center, but as with many aspects of tropical cyclones this can vary greatly from one cyclone to the next.

As we will demonstrate in Chapter 10, the eyewall can also be defined as a region of intense frontogenesis, in which a combination of the Ekman radial flow and, during spin-up, the inviscid part of the radial flow, acts to strongly increase the magnitude of the radial gradient of moist entropy. Surface enthalpy fluxes are particularly important under the eyewall as they are not strongly countered by negative entropy fluxes in convective downdrafts. (This is why tropical cyclones can remain quite intense until the eyewall passes over land, and why it begins to spin

down quite rapidly when the eyewall makes landfall.) In important respects, the eyewall is the boiler room of the tropical cyclone.

Eyewalls are not always nearly circular. Wind shear in the large-scale environment can break the symmetry, and the very sharp maximum of vertical vorticity at the eyewall inner edge can be unstable, giving rise to eyewall mesovortices. These will be discussed in Chapter 11.

8.4.4: Spiral rainbands and concentric eyewalls

Perhaps the most visually spectacular features of tropical cyclone images from satellites and radar are the spiral bands of cloud and rain evident from just outside the eyewall to very large radii. These are readily apparent in satellite and radar imagery (Figures 8.23 and 8.24). A schematic showing a typical configuration of spiral bands, secondary eyewalls, and other precipitation features of tropical cyclones is presented in Figure 8.36. Detailed analysis of reconnaissance aircraft and dropsonde data show that the mesoscale structure of spiral bands in most respects resembles that of typical tropical and mid-latitude squall lines (Powell 1990). Ironically, some of the strongest convection in tropical cyclones, measured in terms of updraft velocity, occur in these spiral rainbands (Houze 2010). Rainfall can be very heavy but is usually of short duration. These bands can spawn waterspouts over the sea, and tornadoes over land (see subsection 8.4.5.4).

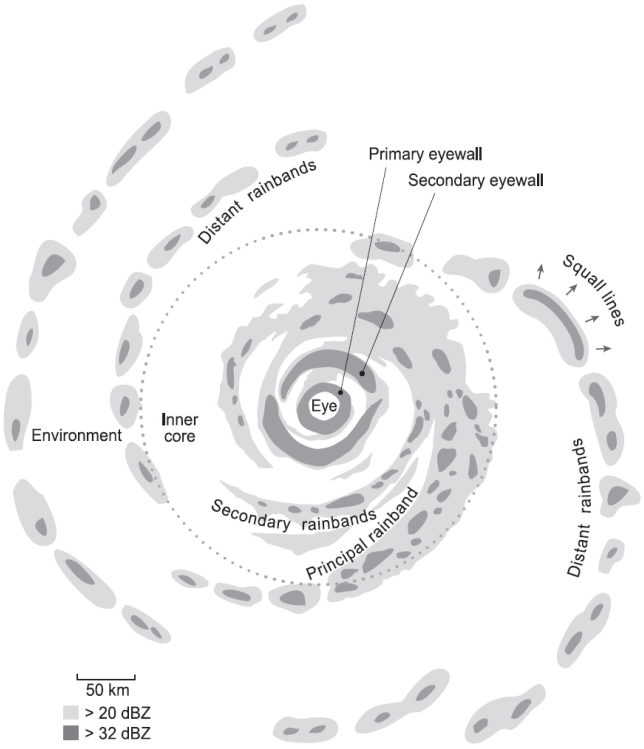


Figure 8.36: Typical distribution of radar reflectivity in a tropical cyclone, including spiral rainbands and primary and secondary eyewalls.

There are times and places where tropical cyclones have two eyewalls, one inside the other. In these cases, the inner eyewall is referred to as the primary, while the outer one is called the secondary eyewall. A nice example of a storm with primary and secondary eyewalls is shown in Figure 8.24. In most cases, the outer eyewall contracts over a period of roughly a day, "strangling" the inner eyewall and causing it to dissipate. The secondary eyewall becomes the primary eyewall. These *eyewall replacement cycles* (ERCs) can occur more than once in a single storm. The physics of secondary eyewall and ERCs will be discussed in Chapter 10.

8.4.5 Associated phenomena

8.4.5.1 Waves

The strong surface winds of tropical cyclones produce high amplitude ocean waves, and the rapid shift in wind direction near the storm center can create confused and dangerous sea states. For example, high seas associated with a western Pacific typhoon sank two destroyers and damaged much of the rest of Admiral Halsey's U.S. Third Fleet in December of 1944, near the end of World War II. Significant wave heights[7] in excess of 12 m have been reported, and individual wave heights can approach 35 m. Figure 8.37 displays the amplitude, direction and wavelength of ocean surface waves associated with Hurricane *Bonnie* of 1998 (Wright et al., 2001). The largest and steepest waves generally occur to the right of the storm track (left in the Southern Hemisphere). The wave field is strongly influenced by the intensity and geometry of the storm, its translation speed, and pre-existing ocean currents. Near shore, the wave field is made even more complex by refraction from coastlines.

[7] Generally, the average of the uppermost 1/3 of the probability distribution of trough-to-crest wave heights.

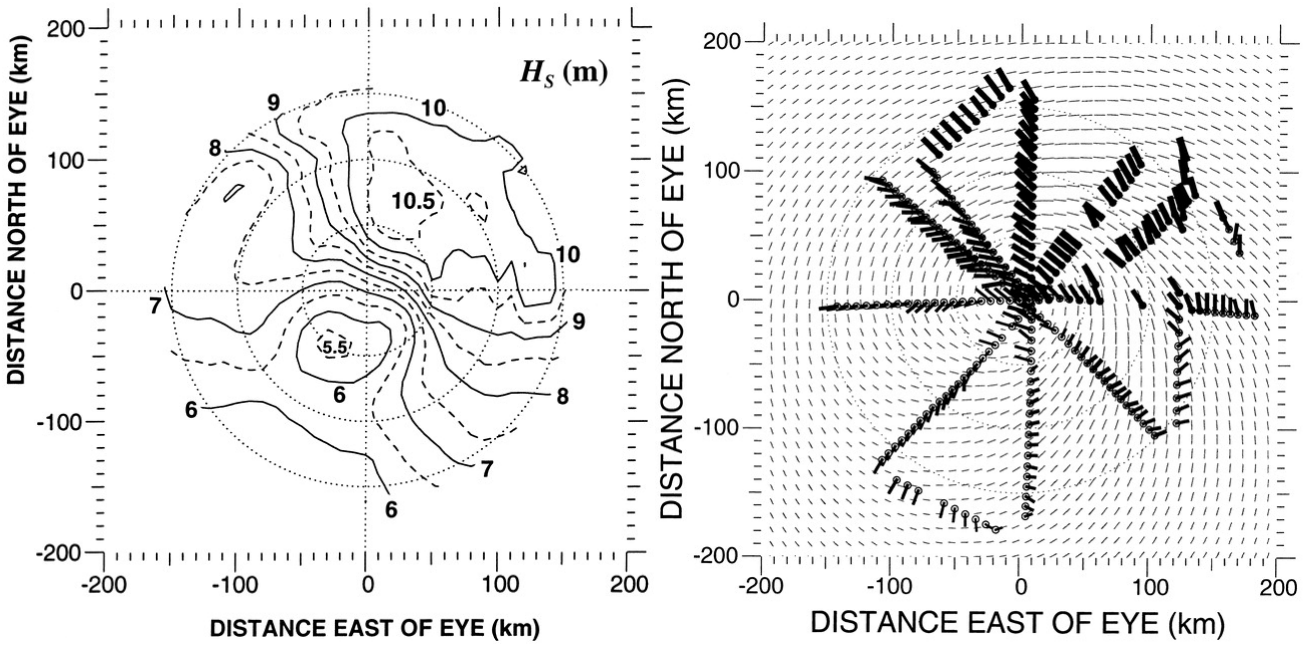


Figure 8.37: (a) Spatial variation in significant wave height measured by a scanning radar altimeter in Hurricane Bonnie on 24 Aug 1998. Contours for integer values of wave height (in meters) are solid and contours for integer values plus 0.5 m are dashed. Bonnie was moving toward the northwest at this time. (b) Hurricane Bonnie primary wave field. The circles indicate the data locations and the radials extend in the wave propagation direction a length proportional to the wavelength. The width of the radials is proportional to the significant wave height, so the aspect ratio is an indication of wave steepness. The short, narrow lines indicate the HRD surface wind analysis.

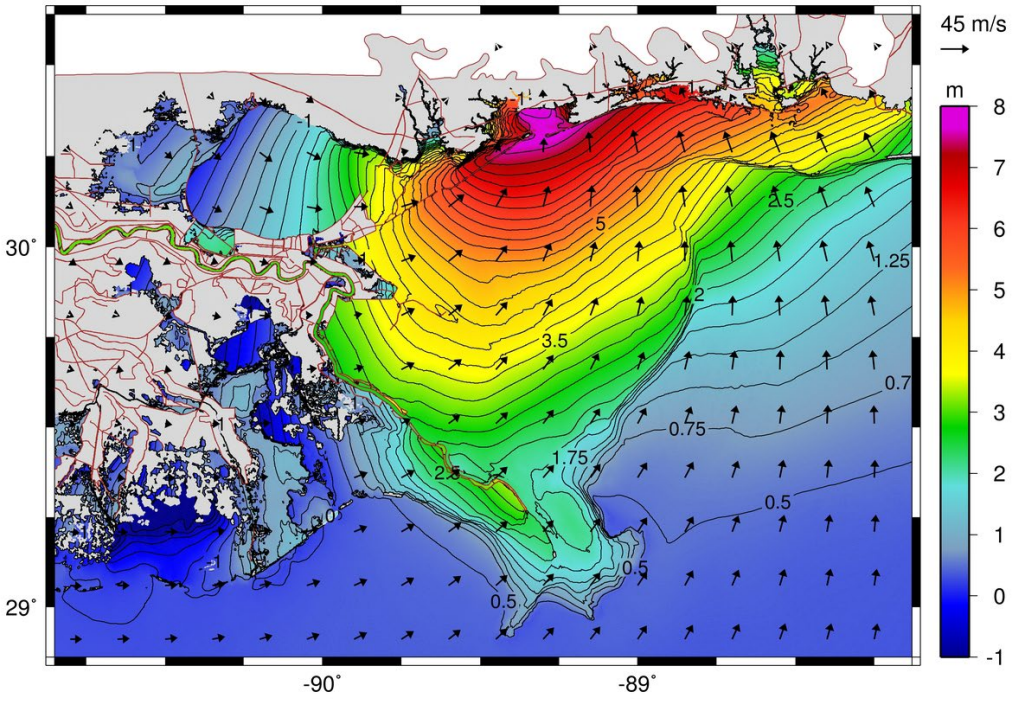


Figure 8.38: Surge heights in Hurricane Katrina as simulated by the Advanced Circulation Model (ADCIRC) of the Computational Hydraulics Laboratory at the University of Notre Dame. Black arrows indicate wind direction and speed.

8.4.5.2 Storm Surges

Storm surges are rapid rises in sea level that accompany tropical cyclones. They are responsible for most deaths and large damages from these storms. They are hydrodynamically nearly identical to tsunamis but they are driven by winds rather than moving sea floor. For this reason, storm surge damage resembles damage from tsunamis, with ships and debris swept far inland.

Technically, the *storm surge* is the storm-driven rise of the sea over and above the normal astronomical tide at the date and location of the event. The *storm tide* is the total storm-driven rise of water including both the surge and the astronomical tide. Flooding depends on the storm tide and thus, of course, on the state of the astronomical tide when the surge strikes. Where astronomical tides are associated with strong currents, the surge and the tide cannot be considered independently, as tidal currents affect the dynamics of the surge, so the storm tide is not always approximated by the arithmetic sum of the normal astronomical tide and the wind-driven surge. Figure 8.38 shows a numerically simulated storm surge accompanying Hurricane Katrina of 2005, the largest surge ever recorded along the U.S. coastline. Strong effects of the complex coastal geography are evident, and the peak surge is to the east of the storm track (not shown in this figure). The hydrodynamics of storm surges are well understood and simulated, as will be discussed in Chapter 13, and the main uncertainties in forecasting surges arise from errors in forecasting the storm track, intensity, size, and timing relative to astronomical tides. Flooding from storm surges can be exacerbated by surface waves.

8.4.5.3 Freshwater Flooding

Almost all the world's extreme rainfall records on time scales of 12 hours to 15 days were set in tropical cyclones affecting the volcanic island of Réunion in the tropical South Indian Ocean. Freshwater flooding from torrential rain is the leading cause of death in tropical cyclones globally, and in some of the drier coastal regions of the tropics, rainfall from tropical cyclones is a substantial fraction of total annual rainfall.

As a tropical cyclone moves inland, the strong winds of its inner core decay rapidly (see Chapter 10.4), but the winds further out decay somewhat more slowly, owing to the quadratic dependence of surface drag on wind speed. This means that the net Ekman mass flux toward the storm center decays relatively slowly with time, while the core of the storm may remain nearly saturated through a deep layer, unless ventilated as a result of environmental wind shear. As a result, storm total rainfall decays relatively slowly, compared to peak winds, as the cyclone moves inland. Moreover, in places like the eastern U.S., the cyclone may begin to interact with mountains after moving over the coastal plain, causing locally enhanced rainfall where the flow is upslope. Finally, tropical cyclones may interact with baroclinic systems as they move into higher latitudes, and that interaction can enhance upward motion (Chapter 12) and sustain or even increase rainfall rates. Very large storm total rainfall can occur when tropical storms stall, has happened in Houston Texas during Hurricane Harvey of 2017, which produced over 800 mm of rain in Harris County, leading to massive flooding.

Figure 8.39 displays the 100-year tropical cyclone-related storm total rainfall at the geographic centroid of each county in the eastern U.S., based on 37,000 tropical cyclone tracks produced by deterministic numerical simulations (Emanuel et al. 2008). In general, these extreme tropical cyclone storm total rain amounts increase toward the coast and southward. This reflects

increasing proximity to warm ocean waters. Note secondary maxima of rainfall along the Appalachian and Ozark Mountains.

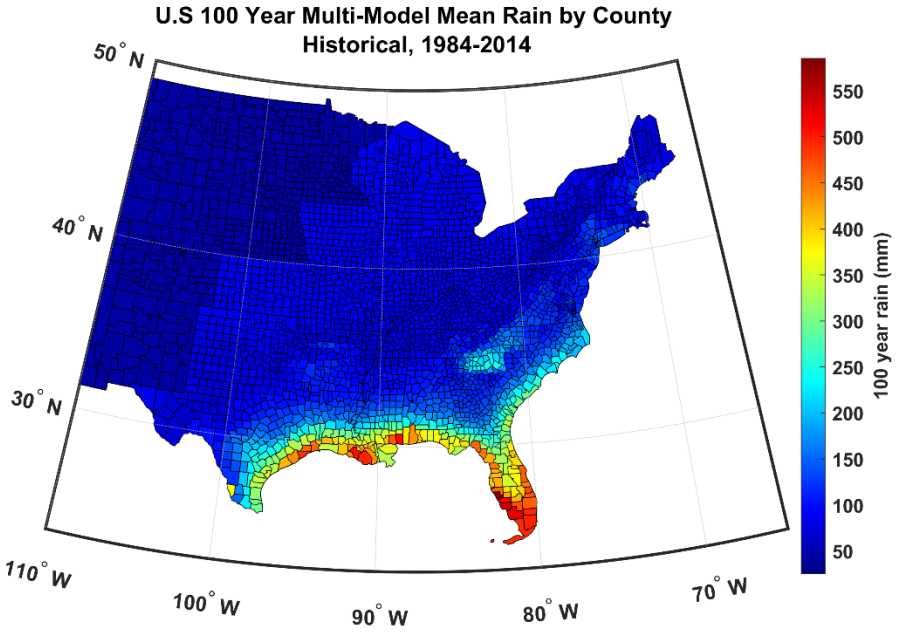


Figure 8.39: Storm total rainfall (mm) associated with tropical cyclones whose annual probability of exceeding the indicated storm total rain is 1%.

8.4.5.4 Tornadoes

As if high winds, floods, and storm surges were not enough, landfalling tropical cyclones often produce tornadoes, which can do considerable damage. Most of these occur in the right-front (left-front in the southern hemisphere) quadrant of the storm with respect to its direction of motion (Figure 8.40) and within 600 km inland of the point of landfall (Schultz and Cecil, 2009). A detailed analysis of the thermodynamic and kinematic environments of tropical-cyclone spawned tornadoes shows that there is far less buoyancy available than is typical of supercell-type thunderstorms of the type responsible for most destructive tornadoes, but this is compensated by strongly varying low-level wind shear, giving rise to storms with very high values of helicity (McCaul Jr., 1991). The strong low-level wind shear is produced by surface drag acting on the strong low-level gradient winds of tropical cyclones.

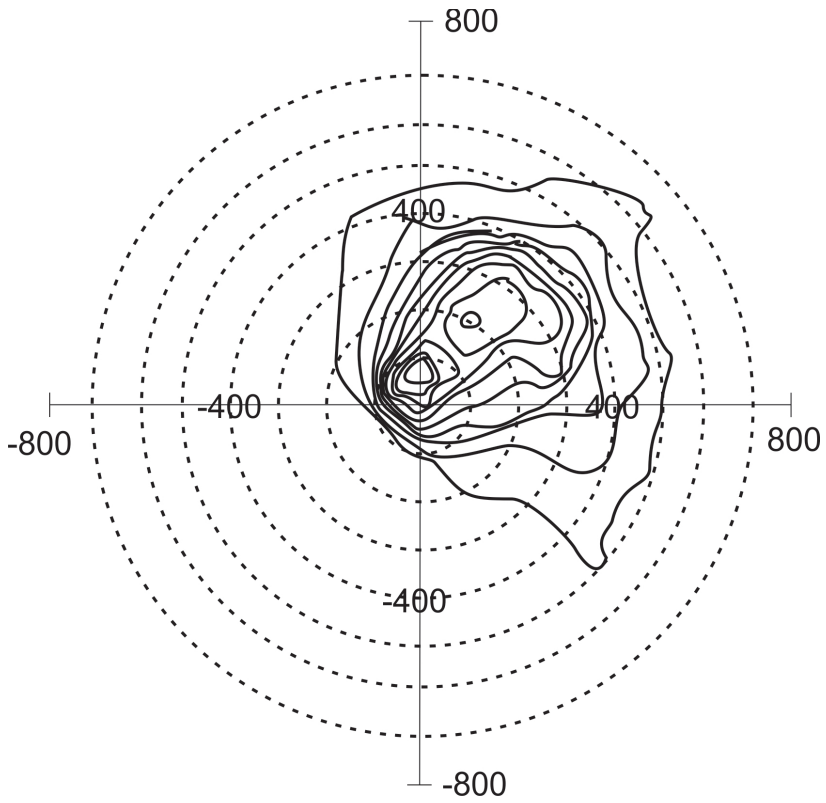


Figure 8.40: Two-dimensional histogram of tornado occurrence relative to storm motion vector (pointing toward top of page). The circles show distance from storm center in increments of 100 km and the contours are in intervals of 10 tornadoes per 100 km^2 bin spacing. After Schultz and Cecil (2009).

References

Cavicchia, L., H. Von Storch, and S. Gualdi, 2014: A long-term climatology of medicanes *Clim. Dyn.*, **48.**

Chavas, D. R., and K. A. Emanuel, 2010: A QuickSCAT climatology of tropical cyclone size. *Geophys. Res. Lett.*, **37**. 10.1029/2010GL044558.

Dean, L., K. Emanuel, and D. R. Chavas, 2009: On the size distribution of Atlantic tropical cyclones. *Geophy. Res. Lett.*, **36**. 10.1029/2009GL039051.

Emanuel, K., J. Callaghan, and P. Otto, 2008: A hypothesis for the re-development of warm-core cyclones over northern Australia. *Mon. Wea. Rev.*, **136,** 3863-3872.

Evans, C., R. S. Schumacher, and T. J. Galarneau Jr., 2011: Sensitivity in the overland reintensification of Tropical Cyclone Erin (2007) to near-surface soil moisture characteristics. *Mon. Wea. Rev.*, **139,** 3848-3870.

Hawkins, H. F., and S. M. Imbembo, 1976: The structure of a small, intense hurricane-Inez 1966. *Mon. Wea. Rev.*, **104,** 418-442.

Huffman, G. J., and co-authors, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeor.*, **8,** 38-55.

Jiang, H., and E. J. Zipser, 2010: Contribution of tropical cyclones to the global precipitation from eight seasons of TRMM data: Regional, seasonal, and interannual variations. *J. Climate*, **23,** 1526-1548.

Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, 2010: The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone best track data. Bull. Amer. Meteor. Soc., 91, 363–376.

Kossin, J. P., and Coauthors, 2007: Estimating hurricane wind structure in the absence of aircraft reconnaissance. *Wea. and Forecast.*, **22,** 89-101.

McCaul Jr., E. W., 1991: Bouyancy and shear characteristics of hurricane-tornado environments. *Mon. Wea. Rev.*, **119,** 1954-1978.

Rappin, E. D., M. C. Morgan, and G. J. Tripoli, 2011: The impact of outflow environment on tropical cyclone intensification and structure. *J. Atmos. Sci.*, **68,** 177-194.

Romero, R., and K. Emanuel, 2013: Medicane risk in a changing climate. *J. Geophys. Res.*, **118**. doi:10.1002/jgrd.50475.

Rotunno, R., and K. A. Emanuel, 1987: An air-sea interaction theory for tropical cyclones. Part II. *J. Atmos. Sci.*, **44,** 542-561.

Schultz, L. A., and D. J. Cecil, 2009: Tropical cyclone tornadoes, 1950-2007. *Mon. Wea. Rev.*, **137,** 3471-3484.

Tang, S., R. K. Smith, M. T. Montgomery, and M. Gu, 2016: Numerical study of the spin-up of a tropical low over land during the Australian monsoon. *Quart. J. Roy. Meteor. Soc.*, **142**. DOI: 10.1002/qj.2797.

Velden, C. S., and and co-authors, 2006: The Dvorak tropical cyclone intensity estimation technique: A satellite-based method that has endured for over 30 years. Bull. Amer. Meteor. Soc., 87, 1195–1210.

Wright, C. W., and Coauthors, 2001: Hurricane directional wave spectrum spatial variation in the open ocean. *J. Phys. Ocean.*, **31,** 2472-2488.

Yanase, W., and co-authors, 2016: Climatology of polar lows over the Sea of Japan using the JRA-55 reanalysis. *J. Climate*, **29,** 419-437.

Zhan, M., and H. von Storch, 2008: A long-term climatology of North Atlantic polar lows. *Geophy. Res. Lett.*, **35**. doi:10.1029/2008GL035769.

9. Axisymmetric, Steady-State Theory

Tropical cyclones are enormous heat engines that extract heat from the ocean at high temperature and expel it in the upper troposphere at low temperature. The fundamental driver is the thermodynamic disequilibrium that exists between the ocean and the tropical atmosphere above it, a disequilibrium that is a necessary consequence of RCE in a greenhouse atmosphere, as described in Chapter 3. Specifically, the requirement of no net energy flux through the ocean surface (which ignores any convergence of heat flow within the upper ocean itself) in thermal equilibrium results in

$$\rho_s C_k |\mathbf{V}| \left(h_0^* - h_b\right) = F_{SW\downarrow} - F_{LW\uparrow}, \tag{9.1}$$

where h_0^* is the saturation moist static energy of the sea surface, h_b is the actual moist static energy of the subcloud layer, ρ_s is the surface air density, C_k is the surface exchange coefficient for moist enthalpy, $|\mathbf{V}|$ is the average surface wind speed, $F_{SW\downarrow}$ is the net downward solar flux into the ocean, and $F_{LW\uparrow}$ is the net upward longwave radiative flux from the ocean surface. In writing (9.1) we have assumed that most of the turbulent flux from the sea surface is mechanically driven by wind flowing over the surface; in conditions of very light surface wind we would have to modify (9.1) to include the convective component of the turbulent fluxes near the surface.

The relation (9.1) can be turned around algebraically to relate the sea surface thermodynamic disequilibrium to the net surface radiative flux:

$$h_0^* - h_b = \frac{F_{SW\downarrow} - F_{LW\uparrow}}{\rho_s C_k |\mathbf{V}|}. \tag{9.2}$$

This shows that the thermodynamic disequilibrium between the ocean surface and the atmosphere just above it is a function of the net radiative flux into the sea and the average winds near the sea surface. The existence of this thermodynamic disequilibrium allows heat to be extracted from the ocean by windstorms which, through locally enhanced wind speeds, induce locally enhanced flows of heat to the atmosphere. This is important only over bodies of water; over land, the flux of heat is strongly limited by diffusion of heat through soils and rocks and, except under unusual conditions, is insufficiently strong to maintain a tropical cyclone.

The existence of a sufficient degree of thermodynamic disequilibrium enables *dissipative structures* to emerge in thermodynamic systems, as first shown by the Russian-Belgian physical chemist Ilya Prigogine[1] (e.g. Prigogine and Lefever 1968; Nicolis and Prigogine 1977). Prigogene showed that when systems are sufficiently far from thermodynamic disequilibrium, coherent, irreversible processes can drive the system to develop organized structures. Thus, Prigogene linked the existence of organized structures (of which we human beings are an example) to the existence of thermodynamic disequilibrium. Tropical cyclones are beautiful

[1] Prigogene won the 1977 Nobel Prize in chemistry for his work on dissipative structures.

examples of organized structures (that contrast with the chaotic nature or ordinary RCE states) that result from thermodynamic disequilibrium.

It follows that to have tropical cyclones, the thermodynamic disequilibrium given by (9.2) must be sufficiently large. Without a greenhouse effect, the net incoming radiation at the surface would be balanced by net outgoing infrared radiation, and there would be no thermodynamic disequilibrium. Thus, fundamentally, tropical cyclones are made possible by the greenhouse effect[2]. The relation (9.2) also demonstrates the important role of mean surface winds; the lighter the winds, the greater the disequilibrium. We expect the largest degree of disequilibrium in regions with high greenhouse gas content (including water vapor) and weak surface winds, both characteristics of tropical warm pools. Were we to modify (9.2) to account for ocean heat transport, we would see larger disequilibrium in regions of ocean heat flux convergence.

Note that the energy source for tropical cyclones is the flow of heat from the ocean to the atmosphere. It is not the latent heat released in convective clouds that drive such storms. Latent heat conversions occur close to local thermodynamic equilibrium and are thus best regarded as internal energy conversions, just as a counterweight is better regarded as part of an elevator's system, not as an external agent whose force is what causes elevators to move. Nor is phase change of water (or any other substance) necessary for tropical cyclones. Perfectly dry cyclones have been simulated in dry RCE states in which all of the surface turbulent heat flux is in the form of sensible heat (Mrowiec et al. 2011; Cronin and Chavas 2019). Tropical cyclones are driven by surface enthalpy fluxes, not latent heating. But as we shall see in Chapter 14, tropical cyclones in moist RCE states differ structurally from those in dry RCE states.

In this chapter, we will explore the physics of an idealized tropical cyclone that is axisymmetric and steady in time, and which exists over an ocean of fixed, constant sea surface temperature. By "steady", we mean that the macro-vortex is steady and we average in time over the small scale, high frequency fluctuations associated with small-scale turbulence and deep moist convection.

9.1 Tropical cyclones as Carnot engines

Consider the trajectory of an air parcel through a steady, axisymmetric tropical cyclone, as illustrated in Figure 9.1. The sample begins at point *A* in the figure and then flows around the circuit *A*-*B*-*C*-*D*-*A*. The shading in the figure shows the moist entropy per unit mass. Each segment of the circuit is described as follows:

1. *A*-*B*: Air spirals inward, losing angular momentum owing to surface friction and gaining enthalpy from the sea through turbulent fluxes. As the pressure drops radially inward, the air is expanding, but observations show that sensible heat flux from the sea is sufficient to prevent much adiabatic cooling, so we idealize this leg as being one of isothermal expansion. In addition to heat added through isothermal expansion and turbulent enthalpy fluxes from the sea, enthalpy is added by the dissipation of turbulence kinetic energy in the turbulent boundary layer.

[2] It follows that increasing greenhouse gas content will increase thermodynamic disequilibrium, barring an increase in mean surface wind speed, and this should increase tropical cyclone activity; this will be explored in Chapter 15.

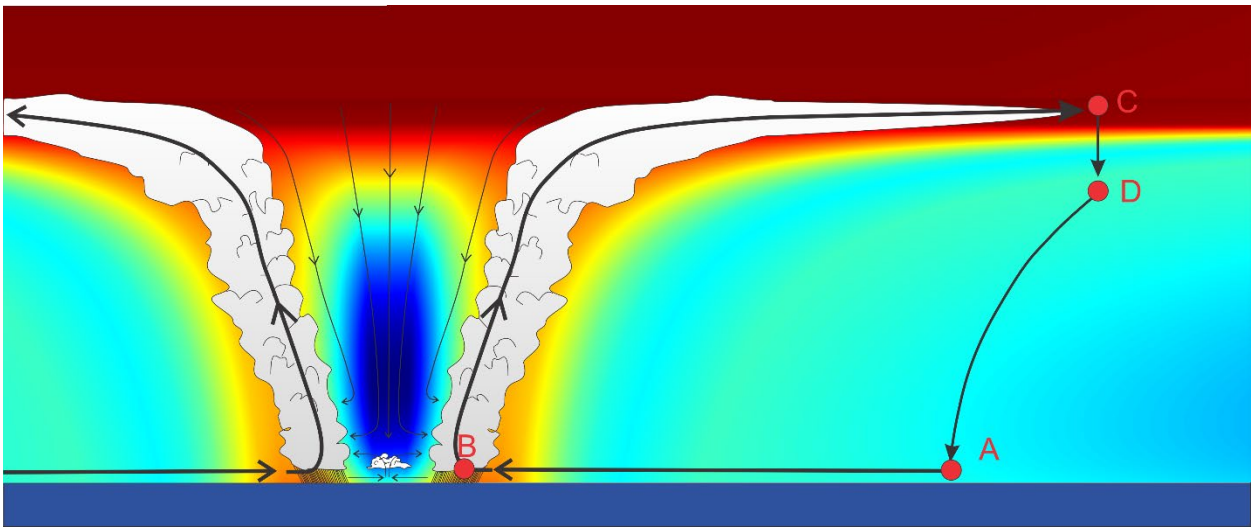


*Figure 9.1: Thermodynamic cycle of a steady, axisymmetric tropical cyclone. A sample of air begins at point **A**, spirals in isothermally toward the storm center in the boundary layer, acquiring enthalpy from the sea surface, and at **B** turns upward and flows moist adiabatically up and out to point **C**. It then descends isothermally from **C** to **D**, and follows an environmental moist adiabatic temperature profile while descending from **D** back to its starting point, **A**. The shading shows the moist entropy per unit mass, with warmer colors denoting higher values. The rapid upward increase of entropy in the stratosphere above the tropical cyclone is not shown here.*

2. *B*-*C*: The sample ascends reversibly and (moist) adiabatically, also conserving its angular momentum.

3. *C*-*D*: In reality, air flowing out of tropical cyclones tends to be exported, asymmetrically, to the far environment, and thus the energy cycle of real tropical cyclones is open. Moreover, real storms do not usually last long enough for the outflow to become truly steady. However, in axisymmetric models, with walls at large radii, the cycle is forced to be closed and yet the simulated cyclones exhibit realistic behavior. The sample loses the entropy it gained in *A*-*B* in this segment by infrared radiation, and does so nearly isothermally. At the same time, through eddy exchanges with the environment, this sample regains the angular momentum it lost during its inflow, *A*-*B*.

4. *D*-*A*: The sample descends back to its starting point, *A*. As it descends, it cools radiatively and loses entropy, but by mixing with shallow convective clouds and boundary layer turbulence, it regains the lost entropy by the time it reaches *A*. (The point *D* is defined so that its (moist) entropy is the same as that of *A*.) Since it loses entropy at a lower absolute temperature than it regains the lost entropy, this leg technically contributes to the production of irreversible entropy, and potentially work, along this leg. On the other hand, it turns out that nearly all (or perhaps all) of this extra irreversibly entropy production is lost the way it usually is in the tropical atmosphere, by mixing between dry and moist samples.

   This leg is far enough out that the tropospheric environment is essentially identical to that of the unperturbed environment of the storm. As emphasized throughout this book, the deep convecting portions of the tropical atmosphere are very nearly moist adiabatic, so that the saturation entropy, s^*, is nearly constant with height. Thus, leg *D*-*A* is also a leg along which the saturation entropy is constant. Thus, we could replace the radiative cooling and the mixing that happens along *D*-*A* by artificially spraying just enough liquid

water into the descending air to keep it saturated and at the same virtual temperature as along the original *D-A* path. Since the virtual temperature is unchanged, is follows that the dynamics are unchanged and thus we would get the same dissipation of kinetic energy as in the original cycle, but this time D-A is moist adiabatic. We lose both the additional potential work that comes from the loss and regain of entropy along the original D-A, but we also lose the additional irreversible entropy production by mixing; these two have to cancel exactly.

We next apply Carnot's theorem to this circuit. We begin by developing a rate equation for entropy by differentiating in time, following the sample, the entropy as defined by (2.69):

$$T\frac{ds}{dt} = \left(c_{pd} + c_l r_t\right)dT + \frac{d\left(L_v r\right)}{dt} - \alpha\left(1 + r_t\right)\frac{dp}{dt} - R_v T \ln\left(\mathcal{H}\right)\frac{dr_t}{dt}, \tag{9.3}$$

where c_{pd} is the heat capacity at constant pressure of dry air, c_l is the heat capacity of liquid water, r is the mixing ratio (mass of water vapor per unit mass of dry air), r_t is the total water mixing ratio (mass per unit mass of dry air), R_d is the gas constant for dry air, p is the pressure, L_v is the latent heat of vaporization, R_v is the gas constant for water vapor, and $\mathcal{H}$ is the relative humidity. We have made use of the identity $\alpha_d = \alpha\left(1 + r_t\right)$. The last term in (9.3) represents the irreversible source of entropy owing to phase change that occurs in conditions of thermodynamic disequilibrium, where $\mathcal{H}$ < 1. We are ignoring the ice phase in this development.

By taking the dot product if the vector momentum equation with the three-dimensional velocity vector, we obtain a relationship for the kinetic energy:

$$\frac{d}{dt}\left(\frac{1}{2}|\mathbf{V}|^2\right) = -\alpha\frac{dp}{dt} + \mathbf{V}\bullet\mathbf{F} - wg, \tag{9.4}$$

where $\mathbf{V}$ is the three-dimensional velocity, w is its vertical component, g is the acceleration of gravity, and $\mathbf{F}$ is the frictional source of momentum.

Eliminating pressure between (9.3) and (9.4) yields:

$$\begin{aligned} T\frac{ds}{dt} = \frac{d}{dt}\left[\left(c_{pd} + c_l r_t\right)T + L_v r + (1 + r_t)\left(\frac{1}{2}|\mathbf{V}|^2 + gz\right)\right] - \mathbf{V}\bullet\mathbf{F} \\ - \frac{dr_t}{dt}\left[R_v T \ln(\mathcal{H}) + \frac{1}{2}|\mathbf{V}|^2 + gz\right]. \end{aligned} \tag{9.5}$$

We next integrate (9.5) around the closed circuit *A-B-C-D-A* shown in Figure 9.1. In doing so, we recognize that the first term on the right side of (9.5) is a perfect derivative and thus integrates to zero around a closed circuit. The result of the integration is therefore

$$\oint T\frac{ds}{dt} = -\oint \mathbf{V}\bullet\mathbf{F} - \oint \frac{dr_t}{dt}\left[R_v T \ln(\mathcal{H}) + \frac{1}{2}|\mathbf{V}|^2 + gz\right]. \tag{9.6}$$

The first term on the right side of (9.6) represents frictional dissipation, while the last two terms are the work necessary to change the kinetic and potential energies of water substance. The term involving the relative humidity is the irreversible generation of entropy by phase changes of water under conditions of thermodynamic disequilibrium, in which $\mathcal{H} < 1$.

In a classical heat engine in which there is no phase change of water, the left side of (9.6) represents the net addition of heat over the cycle, and the right side, consisting only of the first term, is the work done on some external system. In the case of the tropical cyclone, the first term on the right represents the dissipation of wind energy. If we can estimate the net heating represented by the left side of (9.6) as well as the last integral involving water substance, we can thereby estimate the frictional dissipation of wind energy. This is a useful measure of the storm's power, including its potential for causing damage. Some of this dissipation can be regarded as work on done on an external system…in this case, the ocean. This energy is used to generate surface and internal waves, ocean currents, and turbulence in the ocean. The rest is dissipated by turbulence in the atmosphere and, to a lesser degree, exported as internal gravity and Rossby waves to the distance environment. An upper bound on internal frictional dissipation of wind energy can be estimated by assuming that it comprises all of the first term on the right side of (9.6).

We next introduce some mild approximations to (9.6). We first focus on the term involving relative humidity, $\mathcal{H}$. We shall assume that the inner leg *B*-*C* is water-saturated, so $\mathcal{H}$ =1 there and so there is no contribution to (9.6) on that leg. Leg *C*-*D* occurs very high in the system at correspondingly low temperature, so that the total water content r_t and its rate of change must be very small there, thus we also neglect the contribution of the non-equilibrium phase change term there. In the boundary layer, along *A*-*B*, we can compare the magnitude of the non-equilibrium phase change term to one of the important terms in the entropy tendency on the left side of (9.6), $\int_A^B L_v \frac{dr}{dt}$. If we assume that there is no condensed water close to the surface, it is immediately apparent that the ratio of these two terms is

$$\frac{-R_v T_b \ln(\mathcal{H})}{L_v},$$

where T_b is a representative absolute temperature in the boundary layer. Taking this temperature to be *300 K* and $\mathcal{H}$ =0.8, this ratio is about 0.01. Consequently, we can safely neglect its contribution to (9.6) along leg *A*-*B*.

Finally, as already stated, the non-equilibrium phase change term is the only substantial contribution to the entropy tendency itself along leg *D*-*A*. So we can neglect it there if we also take ds/dt to be zero on that leg.

Next we examine the middle term in brackets in (9.6). Even if $|\mathbf{V}| = 80\ ms^{-1}$, changes in the kinetic energy of water substance would be very small compared to changes in its potential energy, gz , for typical vertical displacements of water substance in tropical cyclones. For this reason, we neglect the work done to accelerate water substance in the atmosphere.

With these quite mild approximations, (9.6) can be very well approximated by

$$\oint T\frac{ds}{dt} \cong -\oint \mathbf{V}\bullet\mathbf{F} - \oint gz\frac{dr_t}{dt}. \tag{9.7}$$

Here it must be remembered that we are taking the left side of (9.7) to be zero along *D-A* as it cancels with the nonequilibrium phase change term along that leg.

This can be regarded as stating that the work against friction in a tropical cyclone is that available from the thermodynamic cycle minus the work used to lift water substance.

9.1.1. The differential Carnot cycle

While (9.7) is an interesting constraint on the total amount of dissipation that can occur in a tropical cyclone, it is just an integral constraint. Nevertheless, by considering two adjacent thermal cycles, we can derive local constraints on the magnitude of the winds. Consider two closed circuits that are very close to one another, as pictured in Figure 9.2

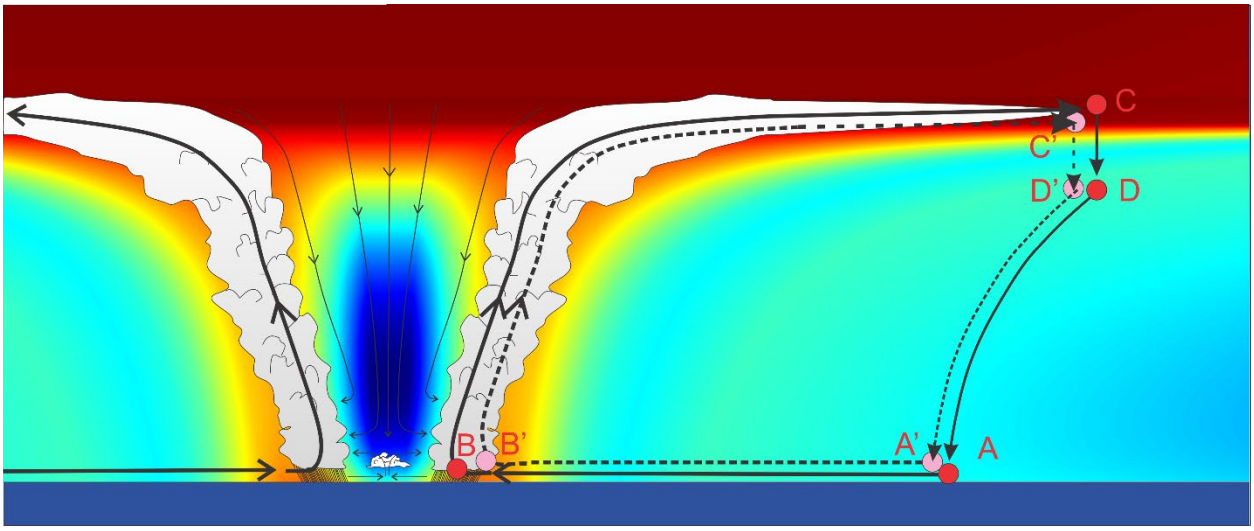


Figure 9.2: Same as Figure 9.1 but showing an additional thermodynamic cycle, A'-B'-C'-D'-A' directly adjacent to the first cycle.

We will perform the same closed integrals around this new, adjacent circuit as we did in the first instance. But since the return legs *D-A* and *D'-A'* occur in the quiescent environment, which has very small horizontal gradients, we will consider the thermodynamic properties at *A'* and *D'* to be identical to those at *A* and *D*, respectively. We then subtract the new circuit from the old one; symbolically we consider a differential circuit

$$\begin{aligned}
&[A \to B \to C \to D \to A] - [A' \to B' \to C' \to D' \to A'] \\
&\quad = [A \to B \to C \to D \to A] - [A \to B' \to C' \to D \to A] \\
&\quad = [B \to C \to C' \to B' \to B].
\end{aligned} \tag{9.8}$$

That is, the differential circuit consists in integrating around the closed loop *B-C-C'-B'-B* shown in Figure 9.2. While this differential circuit can no longer be associated with an air parcel trajectory, it nevertheless is a valid mathematical deduction from the two adjacent circuits.

So we will evaluate (9.7) around the loop *B-C-C'-B'-B*. In doing so, we note that there is no contribution to the left side of (9.7) from the two adiabatic legs *B-C* and *C'-B'*. We are also assuming that angular momentum in conserved on these two legs, and so there is no frictional dissipation along them and therefore no contribution of these two legs to the first term on the left of (9.7).

In evaluating the last term in (9.7) we take into account that, in reality, water is converted to precipitation and is removed from the sample by gravity along much of the vertical legs of the circuit. We will represent this, however, as water being conserved along the vertical branches up to some representative altitude $\overline{z}$, which we consider to be the mean altitude at which precipitation falls out of the upward moving air.

With these stipulations, (9.7) becomes

$$\int_{B'}^{B} T\frac{ds}{dt}+\int_{C}^{C'} T\frac{ds}{dt}=-\int_{B'}^{B}\mathbf{V}\bullet\mathbf{F}-\int_{C}^{C'}\mathbf{V}\bullet\mathbf{F}-\int_{C}^{C'} g\overline{z}\frac{dr_t}{dt} \tag{9.9}$$

Finally, we note that the state variables s and r_t must return to their starting values going around the closed circuit. This means that ds/dt evaluated between B' and B must be equal and opposite to its value evaluated between *C* and *C'*; likewise the value of dr_t/dt evaluated between *C* and *C'* must be equal and opposite to its value evaluated between *B* and *B'*.

Taking the limit as $B'\rightarrow B$ and $C'\rightarrow C,$ (9.9) becomes

$$\left(T_b-T_o\right)\left.\frac{ds}{dt}\right|_B=-\left.\mathbf{V}\bullet\mathbf{F}\right|_B-\left.\mathbf{V}\bullet\mathbf{F}\right|_C+g\overline{z}\left.\frac{dr_t}{dt}\right|_B, \tag{9.10}$$

where $T_o\equiv\left.T\right|_C$ is defined as the *outflow temperature*, and is an important factor in the thermodynamic generation of energy in a tropical cyclone. Note that the only terms that have to be evaluated at *C* are the outflow temperature (by definition) and the second frictional dissipation term on the right side of (9.10).

We are now in a position to make two more simplifications based on scale analysis. First, we compare the magnitude of the last term in (9.10) to the latent heating component of ds/dt on the left side of (9.10). That ratio is

$$\frac{g\overline{z}}{L_v}\frac{T_b}{T_b-T_o}$$

Using a generous estimate of $\overline{z}$ of 5 km and taking $T_b=300\,K$ and $T_o=220\,K$ (a fairly high value considering that typical outflow temperatures are closer to $200\,K$) gives a generous estimate of this ratio of 0.075. We conclude that the work done to lift water substance against gravity, while not entirely negligible, can be ignored to first order. But we note that including this term will not add appreciably to the complexity of our forthcoming estimate of surface frictional

dissipation, and that the work done to lift water substance subtracts from the energy available from the thermodynamic cycle to counter frictional dissipation.

Next, we compare the magnitudes of the outflow and boundary layer frictional dissipation terms (first and second terms on the right side of (9.10)). For the narrow purposes of this comparison, we suppose that the dissipation is predominantly related to the azimuthal component of the wind, V. For this component, in an axisymmetric flow, the frictional force is a sink or source of angular momentum:

$$\frac{dM}{dt} = rF_\theta, \tag{9.11}$$

where F_θ is the azimuthal component of the frictional force per unit mass, and M is the angular momentum per unit mass:

$$M = rV + \frac{1}{2} fr^2, \tag{9.12}$$

where f is the Coriolis parameter. Using (9.11) and considering only the azimuthal contributions to the two frictional dissipation terms, their ratio is

$$\frac{\mathbf{V}\bullet\mathbf{F}\big|_C}{\mathbf{V}\bullet\mathbf{F}\big|_B} \simeq \frac{r_B}{r_C} \frac{V_C \left.\dfrac{dM}{dt}\right|_C}{V_B \left.\dfrac{dM}{dt}\right|_B}, \tag{9.13}$$

where r_B and r_C are the radii and V_B and V_C the azimuthal velocities of points *B* and *C* in the thermodynamic cycle. But since, like entropy, angular momentum is a state variable that is conserved along the two adiabatic legs of the differential Carnot cycle, we must have

$$\left.\frac{dM}{dt}\right|_C = -\left.\frac{dM}{dt}\right|_B.$$

Thus (9.13) becomes

$$\frac{\mathbf{V}\bullet\mathbf{F}\big|_C}{\mathbf{V}\bullet\mathbf{F}\big|_B} \simeq \frac{-r_B V_C}{r_C V_B}. \tag{9.14}$$

before proceeding further, we note that this ratio vanishes in the event that $V_C = 0$. We also note that $r_C > r_B$ and that observations (Chapter 8) show that the largest azimuthal winds tend to occur quite low down in the storm near the eyewall, so that we expect that $|V_C| < |V_B|$. Thus we can be confident that if point *B* lies close to the radius of maximum winds, the magnitude of the ratio of frictional dissipation in the outflow to that in the boundary layer will be small. We can also show that this ratio becomes small in the limit of large r_C / r_B. Since the angular momentum at *C* is the same as that at *B*, we can write

$$V_C = \frac{M}{r_C} - \frac{1}{2} f r_C = \frac{r_B V_B + \frac{1}{2} f r_B^2}{r_C} - \frac{1}{2} f r_C. \tag{9.15}$$

Substituting this into the numerator of (9.14) gives

$$\frac{\mathbf{V}\bullet\mathbf{F}\big|_C}{\mathbf{V}\bullet\mathbf{F}\big|_B} \simeq \frac{1}{2}\frac{f r_B}{V_B} - \left(\frac{r_B}{r_C}\right)^2 \left(1 + \frac{1}{2}\frac{f r_B}{V_B}\right) \tag{9.16}$$

Therefore, in the limit of very large r_C / r_B, this ratio approaches $f r_B / 2V_B$. Even for a weak, broad vortex, taking $r_B = 100\, km$ and $V_B = 20\, ms^{-1}$ and using a typical tropical value of the Coriolis parameter of $5 \times 10^{-5} s^{-1}$, this ratio is 0.125. Therefore, as with the work used to lift water substance, outflow dissipation is not entirely negligible if we take r_C / r_B so be large, but can be ignored to first order. If we take point C to be closer to the radius at which the azimuthal velocity in the outflow vanishes, the outflow dissipation is even smaller compared to the boundary layer dissipation.

Therefore, to first order, we can approximate (9.10) by

$$\left(T_b - T_o\right)\left.\frac{ds}{dt}\right|_B \simeq -\mathbf{V}\bullet\mathbf{F}\big|_B. \tag{9.17}$$

Going back to the original Carnot cycle, we can conclude that (9.17) should be valid in a steady, axisymmetric tropical cyclone for any boundary layer point B that is broadly within the ascent region of the cyclone's inner core.

In the next section we use elementary boundary layer theory to relate the terms in (9.17) to surface fluxes.

9.1.2 Boundary layer sources of enthalpy and angular momentum

The physics of the tropical cyclone boundary layer controls the rates of flux of enthalpy from the sea to the air, which is the ultimate energy source for the storm, as well as most of the dissipation of kinetic energy in the storm. Because the air immediately above the sea surface is usually slightly cooler than the sea surface itself, and also because it is drier, the buoyancy flux is positive, but the winds are strong, yielding large values of $u*$ and therefore large values of the Obukhov length (Chapter 4; see equation (4.4)). For this reason, we can regard most of the boundary layer turbulence as have a mechanical origin in the cyclone's winds.

As we reviewed in Chapter 4, a common approximation made in boundary layer turbulence is that sources and sinks of momentum and other quantities are associated with the convergence of only the vertical component of the turbulent fluxes. That is, we approximate the boundary layer as being statistically homogeneous in the horizontal. We make this approximation here but do so with grave regrets, because owing to strong frontogenesis at the eyewall (see Chapter 10,

section 10.1.2), radial turbulent fluxes are an essential component of tropical cyclones as they prevent total collapse of the eyewall into a truly discontinuous front. The contribution of these horizontal turbulent fluxes to the total dissipation of kinetic energy in tropical cyclones remains a relatively unexplored problem.

With this approximation, the frictional dissipation term in (9,17) may be written

$$\mathbf{V}\bullet\mathbf{F} = -\mathbf{V}\bullet\frac{1}{\rho}\frac{\partial\boldsymbol{\tau}}{\partial z}, \tag{9.18}$$

where $\boldsymbol{\tau}$ is the vector horizontal turbulent stress. Integrating this over the depth of the boundary layer, $h,$ and using integration by parts:

$$\begin{aligned}\int_0^\delta \mathbf{V}\bullet\mathbf{F}dz &= -\int_0^\delta \mathbf{V}\bullet\frac{1}{\rho}\frac{\partial\boldsymbol{\tau}}{\partial z}dz = -\int_0^\delta \frac{\partial}{\partial z}\left[\frac{1}{\rho}\mathbf{V}\bullet\boldsymbol{\tau}\right]dz + \int_0^\delta \boldsymbol{\tau}\bullet\frac{\partial}{\partial z}\frac{\mathbf{V}}{\rho}dz \\ &= \mathbf{V}_s\bullet\frac{\boldsymbol{\tau}_s}{\rho_s} + \int_0^\delta \boldsymbol{\tau}\bullet\frac{\partial}{\partial z}\left(\frac{\mathbf{V}}{\rho}\right)dz,\end{aligned} \tag{9.19}$$

where we have defined the top of the boundary layer, $\delta,$ to be the level at which the turbulent stresses are assumed to vanish, and the subscript s in (9.19) denotes evaluation at the surface.

If the vertical wind shear is not too strong through the depth of the boundary layer, the last term in (9.19) may be somewhat smaller than the first, and so we may write, approximately,

$$\int_0^h \mathbf{V}\bullet\mathbf{F}dz \simeq \mathbf{V}_s\bullet\frac{\boldsymbol{\tau}_s}{\rho_s}. \tag{9.20}$$

One enormous challenge in understanding and simulating tropical cyclones is that, under very high wind conditions, the transition from ocean to atmosphere ceases to be an interface and becomes an emulsion, in which bubble-filled water gives way gradually to spray-filled air. At present, we do not know how to represent turbulent enthalpy and momentum fluxes through such an emulsion. We shall return to this interesting question in Chapter 13, but for now, we will bite the bullet and represent the surface stress using a classical aerodynamic flux formula, as reviewed in Chapter 4:

$$\boldsymbol{\tau}_s = -C_D\rho_s\,|\,\mathbf{V}_s\,|\,\mathbf{V}_s, \tag{9.21}$$

where C_D is a dimensionless drag coefficient. With this representation, (9.20) becomes

$$\int_0^\delta \mathbf{V}\bullet\mathbf{F}dz \simeq -C_D\,|\,\mathbf{V_s}\,|^3\,. \tag{9.22}$$

This dissipation of kinetic energy in the mechanical boundary layer varies as the cube of the mean wind speed. As detailed in Chapter 4, the value of C_D varies with the measurement altitude, as does the wind speed itself, so care must be taken in choosing a measurement altitude that is above the surface layer but not too deep within the boundary layer.

We next turn our attention to the entropy source term in (9.17). There are two main sources of entropy in the boundary layer: the flux of enthalpy from the sea surface (divided by the boundary layer temperature) and the heating that results when kinetic energy dissipates in the boundary layer.

Once again resorting to the aerodynamic representation of turbulent fluxes from a surface, the enthalpy flux may be written

$$F_h = C_k \rho_s \mid \mathbf{V_s} \mid \left(k_0^* - k_s\right), \tag{9.23}$$

where C_k is a dimensionless coefficient of enthalpy transfer, k_0^* is the saturation moist enthalpy at sea surface temperature and pressure, and k_s is the moist enthalpy of air near the surface.

The other important source of entropy in the boundary layer is internal dissipation of kinetic energy. Only that portion of the total dissipation that occurs locally in the atmospheric boundary layer contributes to the dissipative heating of the air; heat dissipated by turbulence in the ocean may be presumed to be mixed rapidly though the ocean mixed layer and thus not made immediately available to the atmosphere. What portion of the kinetic energy dissipated within the aforementioned emulsion winds up in the atmosphere is unknown. In addition, some of the wind stress creates surface waves and swell, which export energy out of the storm area, and also directly excites near-inertial oscillations in the ocean (Chapter 13). Because of the currently poor understanding of the magnitudes of these energy exports, we here simply assume that a fraction μ of the kinetic energy dissipated by friction at and near the surface, acts to heat the boundary layer[3].

With these assumptions, we can write the integral through the depth of the boundary layer of the left side of (9.17) as

$$\int_0^\delta \left(T_b - T_o\right) \frac{ds}{dt} = \frac{T_b - T_o}{T_b} \left[C_k \mid \mathbf{V_s} \mid \left(k_0^* - k_s\right) + \mu C_D \mid \mathbf{V_s} \mid^3 \right]. \tag{9.24}$$

9.1.3 Potential intensity

Integrating (9.17) through the depth of the boundary layer, using (9.22) and (9.24), and re-arranging the terms, yields an expression relating the surface wind speed to the local thermodynamic disequilibrium between the sea and the air:

$$\mid \mathbf{V_s} \mid^2 = \left[\frac{T_b - T_o}{T_b \left(1 - \mu\right) + \mu T_o} \right] \frac{C_k}{C_D} \left(k_0^* - k_s\right). \tag{9.25}$$

[3] The fraction μ has been derived by Edwards (2019) from a careful consideration of the effect of dissipative heating on the surface layer. His result is $\mu = 1 - \frac{1}{2}\frac{C_k}{C_d}$.

Note that this wind speed will be maximum when it is assumed that all the frictional dissipation is returned locally to the atmospheric boundary layer as heat $(\mu = 1)$ [4]. We define the resulting wind speed as the *potential intensity*, V_p :

$$V_p^2 = \frac{T_b - T_o}{T_o} \frac{C_k}{C_D} \left(k_0^* - k_s \right). \tag{9.26}$$

This may be regarded as an upper bound on the *surface* wind speed of a steady, axisymmetric tropical cyclone in a given thermodynamic environment. We emphasize that this is a surface wind speed estimate, which in this case is defined as the wind speed used in the aerodynamic flux formulae. It is not necessarily an upper bound on the wind speed anywhere within the tropical cyclone; indeed, as we saw in Chapter 8, the winds higher up within the boundary layer are often stronger, a point we shall return to in due course. Moreover, we have assumed that the entire macro-system is steady in time. There is nothing about (9.26) that precludes higher wind speeds in transient and/or non-axisymmetric wind features such as boundary layer eddies and vortices.

Also note that relaxing any of the assumptions we made along the way reduces this estimate of the wind speed. In particular, this estimate will be smaller if any or all of the following are accounted for:

- Some of the frictional loss of kinetic energy is used to heat the ocean, drive ocean currents, or create ocean wave energy
- Some of the frictional loss within the atmospheric boundary layer is not directly related to the surface stress (i.e. the last term in (9.19) is not negligible)
- Some of the atmospheric frictional loss occurs outside the boundary layer, particularly in the outflow region
- Some of the available energy is used to lift and accelerate water mass (last term in (9.10).

It is also important to recognize that (9.26) cannot be directly evaluated if all we know is the unperturbed thermodynamic environment of the tropical cyclone. That is because, without further development, we do not know *a priori* what the outflow temperature T_o is, nor do we know what the *local* atmospheric boundary layer moist enthalpy, k_s, is. Even if we know the sea surface temperature, the sea surface saturation enthalpy k_0^* also depends on the local surface pressure, which varies with the intensity of the storm, among other things. And finally, we still have no really good estimates of the surface exchange coefficients at high wind speeds, a point we will return to in Chapter 13.

We next turn to dynamical constraints on the maximum gradient wind speed and on the structure of steady, axisymmetric tropical cyclones.

[4] We note that uncertainty in the value of μ is wrapped up with uncertainty in the values of the surface exchange coefficients, insofar as potential intensity is concerned.

9.2. Tropical cyclones as quasi-balanced, slantwise-neutral vortices

Tropical cyclones evolve sufficiently slowly that, above the boundary layer, they may be considered to evolve through a succession of states in which gradient wind and hydrostatic balance apply. (Gradient balance may be a poor approximation in the outflow layer, for reasons that will be described later.) As they contain copious deep moist convection, and as their unperturbed environments tend to be moist-neutral, we may also consider the tropical cyclone outside its eye to be a moist-neutral vortex. In the next section, we explore the physics of slantwise convection and show that tropical cyclones above the boundary layer and outside the eye may be regarded as having zero saturation potential vorticity (SPV).

9.2.1. Slantwise convection and slantwise neutrality

Consider a thermally stratified, axisymmetric vortex in hydrostatic and gradient wind balance. Now imagine perturbing that vortex by displacing an axisymmetric "tube" of air of very small radial and vertical dimensions. This tube can be displaced in the radial and/or vertical directions. (This is analogous to displacing a "parcel" of air vertically to test its stability to convection.) The tube is so small in its vertical and radial dimensions that we can assume that its displacement will not perturb the pressure field.

We begin by writing the inviscid equations for conservation of radial and vertical momentum:

$$\frac{du}{dt} = -\alpha\frac{\partial p}{\partial r} + \frac{V^2}{r} + fV = -\alpha\frac{\partial p}{\partial r} + \frac{M^2}{r^3} - \frac{1}{4}f^2 r, \tag{9.27}$$

and

$$\frac{dw}{dt} = -\alpha\frac{\partial p}{\partial z} - g, \tag{9.28}$$

where α is the specific volume, p is pressure, V is the azimuthal velocity, f is the Coriolis parameter, and g is the acceleration of gravity. In writing the right side of (9.27), we have made use of the definition of angular momentum per unit mass, (9.12).

In the steady, balanced vortex, the pressure distribution, which is here assumed to remain invariant in time, is, from (9.27) and (9.28)

$$\bar{\alpha}\frac{\partial p}{\partial r} = \frac{\overline{M^2}}{r^3} - \frac{1}{4}f^2 r, \tag{9.29}$$

and

$$\bar{\alpha}\frac{\partial p}{\partial z} = -g, \tag{9.30}$$

where the overbar represents the balanced, background values. Then if we displace our hypothetical massless tube of air in the radial and vertical directions while conserving its angular momentum and moist or dry entropy (depending on whether the tube is water-saturated or not), we have, from (9.27) – (9.30)

$$\frac{du}{dt} = \frac{M^2 - \overline{M^2}}{r^3}, \tag{9.31}$$

$$\frac{dw}{dt} = g\frac{\alpha - \overline{\alpha}}{\overline{\alpha}} = \Gamma\left(s - \overline{s}\right), \tag{9.32}$$

where we have ignored the effect of perturbations in specific volume in writing (9.31) and in writing (9.32) we have made use of the relationship (2.49) or (2.75) between perturbations in specific volume and perturbations in entropy. The entropy in (9.32) should be considered moist entropy if the tube is saturated, or dry entropy if it is not; likewise, Γ should be taken as the moist adiabatic lapse rate in the first case, and the dry adiabatic lapse rate in the second.

Now consider a typical hurricane eyewall, which is saturated and in which the moist entropy increases inward and upward, while the angular momentum increases outward and downward. Suppose that the slope of the moist entropy surfaces is steeper than that of the angular momentum surfaces, as illustrated in Figure 9.3.

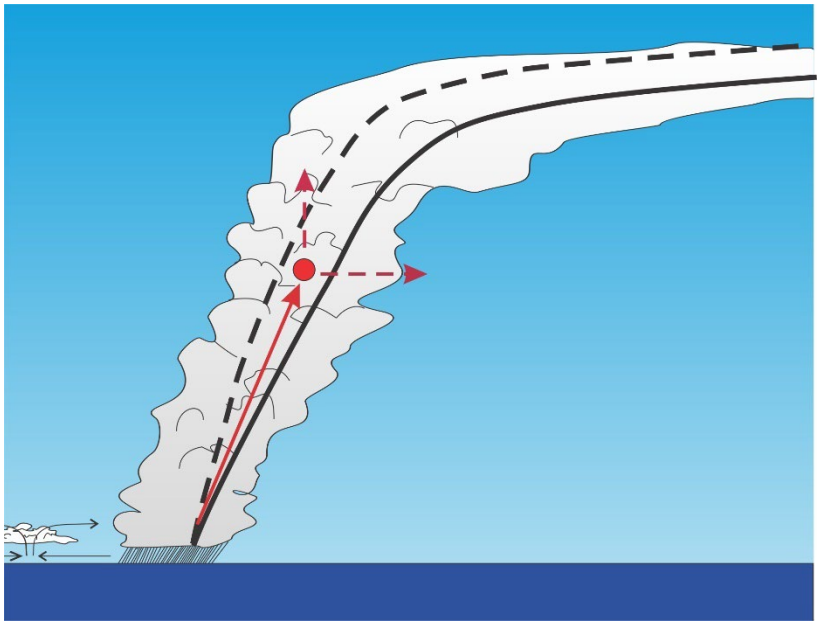


Figure 9.3: The eyewall of a steady, axisymmetric tropical cyclone showing a surface of constant angular momentum (solid black curve) and a surface of constant moist entropy (dashed black curve). A "tube" of air at cloud base is hypothetically lifted upward moist adiabatically between the two surfaces (red arrow), conserving its angular momentum and arriving at the filled red circle. The vertical and radial forces acting on the tube are indicated by the dashed red arrows.

First consider a tube of air that begins at cloud base and is hypothetically lifted moist adiabatically straight upward. Because the moist entropy increases upward in the eyewall, the sample will have a moist entropy that is less than that of its environment and, according to (9.32) will accelerate downward. The eyewall is stable to vertical displacements and thus to strictly vertical deep moist convection.

Next consider a strictly radial outward displacement. Since angular momentum increases outward, the tube will find itself in an environment of greater angular momentum and therefore,

according to (9.31), will accelerate radially inward, opposite to the sense of its initial displacement. This displacement is therefore stable, and the vortex is said to be *inertially stable*.

As we discussed extensively in Chapter 2, the criterion for convective stability is that the entropy should increase upward, and as we might deduce from the preceding paragraph, the criterion for inertial stability is that the angular momentum should increase outward.

But now suppose that the tube of air is displaced slantwise upward, at an angle between the slopes of the entropy and angular momentum surfaces, as illustrated by the red arrow in Figure 9.3. One can see that because it lies below its "home" entropy surface it will have a larger entropy than its environment and be accelerated upward, and because it lies radially inward of its home angular momentum surface, it will accelerate radially outward. The vector sum of these two accelerations is upward and outward, in the same direction as the initial displacement, and so the eyewall is unstable to slantwise (moist) convection, even though it is stable to both upright convection and inertial instability.

It is apparent that the condition for neutral stability is that the angular momentum surfaces and entropy surfaces should coincide. This is in fact a more general definition of convective neutrality than the conventional definition of zero buoyancy for a parcel lifted adiabatically in the vertical. The most general statement of convective neutrality is that parcels lifted adiabatically along vortex lines of the balanced flow should be neutrally buoyant. In an axisymmetric vortex, the vortex lines of the balanced flow are lines along angular momentum surfaces based on the gradient (rather than actual azimuthal) wind. Measurements made by research aircraft flying down balanced vortex lines in winter storms show moist adiabatic lapse rates (Emanuel 1988).

Just as we find nearly moist adiabatic lapse rates in deep convective atmospheres, even far away from actual deep convective clouds, we should expect to find nearly moist adiabatic lapse rates along angular momentum surfaces in tropical cyclones, even away from the core. Thus, neglecting the direct effects of water substance on density, we expect that above the boundary layer in the troposphere, saturation moist entropy, s^*, should be nearly constant on angular momentum surfaces. But as we shall see, in strong, well-developed tropical cyclones, strong subsidence outside the core can shut off deep convection entirely, and this is certainly the case in the eye; In such regions we might not expect to find moist convective neutrality.

Another consequence of slantwise moist neutrality is that the saturation potential vorticity, SPV, should vanish everywhere above the boundary layer and outside the eyes of tropical cyclones:

$$SPV \equiv \alpha\left(f\mathbf{k} + \nabla \times \mathbf{V}_g\right) \bullet \nabla s^* = 0, \tag{9.33}$$

where $\mathbf{k}$ is the unit vector in the vertical and $\mathbf{V}_g$ is the gradient wind. Note that SPV is not conserved except where the air is saturated, but because it is based on a state variable s^*, it is invertible, subject to gradient wind balance and appropriate boundary conditions, which are crucial in tropical cyclone dynamics. This means that, given the distribution of SPV and suitable boundary conditions, the system can be inverted to find the balanced flow and (virtual) temperature distributions.

9.2.2 Analogy to quasi-balanced Eady model

The classical Eady model of baroclinic instability (Eady 1949) posits a fluid of constant pseudo-potential vorticity (7.18) contained between two rigid lids. The fluid is adiabatic and inviscid, and in the quasi-geostrophic approximation, q_p is conserved and is therefore always constant. The entire system is determined by the space-time distribution of entropy (potential temperature) on the upper and lower boundaries, and therefore, conceptually and mathematically, all the system dynamics occurs on the two boundaries. By inverting q_p subject to the time-evolving boundary conditions on entropy, and by solving the quasi-geostrophic omega equation, one can diagnose the interior structure and ageostrophic velocities, including the vertical velocity, but it is not necessary to do so to evolve the system in time.

As with the Eady problem, the tropical cyclone has, outside its eye, constant potential vorticity; unlike the Eady problem, it is constant not because it is conserved (it isn't) but because it is constantly adjusted toward zero by (slantwise) convection. And the definition of the relevant PV, SPV as given by (9.33), is based on the saturation moist entropy and the gradient (not geostrophic) wind. Nevertheless, as with the Eady problem all of the time-evolving dynamics happen on the boundaries.

At the lower boundary, the space-time distribution of moist entropy is the relevant boundary condition. As discussed in section 9.1.2, the lower boundary layer entropy is determined not just by advection, as in the Eady model, but by surface fluxes and dissipative heating.

The upper boundary condition is much more difficult to formulate and at this writing there is no definite agreement on how to do so. We will return to a more detailed discussion of this issue presently; before we do so, we will describe the implications of the assumptions of gradient balance and zero SPV for the structure of tropical cyclones outside the eye and above the boundary layer.

9.2.3 Thermal wind balance and characteristic surfaces

We next demonstrate that thermal wind balance in a zero SPV vortex strongly constrains the structure of tropical cyclones outside their eyes and demonstrates that angular momentum and absolute temperature are the natural independent coordinates in such vortices.

We begin by writing an expression for gradient wind balance, (9.27), but this time do so in pressure coordinates:

$$\frac{\partial \phi}{\partial r} = \frac{M^2}{r^3} - \frac{1}{4} f^2 r, \tag{9.34}$$

where ϕ is the geopotential. In pressure coordinates, the hydrostatic equation is'

$$\frac{\partial \phi}{\partial p} = -\alpha. \tag{9.35}$$

Eliminating ϕ by cross-differentiation yields the thermal wind equation:

$$\frac{1}{r^3}\frac{\partial M^2}{\partial p} = -\frac{\partial \alpha}{\partial r}. \tag{9.36}$$

Expressing α as a function of the two state variables p and s*, we write (9.36) as

$$\frac{1}{r^3}\frac{\partial M^2}{\partial p} = -\left(\frac{\partial \alpha}{\partial r}\right)_p = -\left(\frac{\partial \alpha}{\partial s*}\right)_p \left(\frac{\partial s*}{\partial r}\right)_p = \left(\frac{\partial T}{\partial p}\right)_{s*} \left(\frac{\partial s*}{\partial r}\right)_p, \tag{9.37}$$

where we have used Maxwell's relation (2.74) with the approximation that $r_t << 1.$

Now if SPV given by (9.33) is zero, it follows that surfaces of constant $s*$ are also surfaces of constant absolute vorticity, so that the gradient of the former is perpendicular to the absolute vorticity. At the same time, it is easily shown that absolute vortex lines lie on angular momentum surfaces, because, from the definition of angular momentum (9.12),

$$\eta_z = \frac{1}{r}\frac{\partial M}{\partial r},$$
$$\eta_r = \frac{1}{r}\frac{\partial M}{\partial z},$$

where η_z and η_r are the vertical and radial components of the absolute vorticity. Therefore,

$$\frac{\eta_z}{\eta_r} = \frac{\partial M / \partial r}{\partial M / \partial z} = -\left(\frac{\partial z}{\partial r}\right)_M.$$

Vortex lines lie on surfaces of constant angular momentum. Therefore, zero SPV means that $s* = s*(M),$ the saturation entropy is a function of M alone in the steady state. (If zero SPV and thermal wind balance are maintained in the time-evolving flow, then $s*$ is a function of M and time.) So we can write (9.37) as

$$\frac{1}{r^3}\left(\frac{\partial M^2}{\partial p}\right)_r = \left(\frac{\partial T}{\partial p}\right)_{s*}\left(\frac{ds*}{dM^2}\right)\left(\frac{\partial M^2}{\partial r}\right)_p. \tag{9.38}$$

Dividing through by $\left(\frac{\partial M^2}{\partial r}\right)_p$ yields an expression for the slope of M surfaces:

$$\frac{1}{r^3}\left(\frac{\partial r}{\partial p}\right)_M = -\left(\frac{\partial T}{\partial p}\right)_{s*}\left(\frac{ds*}{dM^2}\right). \tag{9.39}$$

Note that since $s* = s*(M),$ the second factor in the right side of (9.39) is constant on surfaces of constant M. For this reason, (9.39) can be integrated upward along M surfaces to give an expression for the radius of each M surface as a function of absolute temperature and the gradient of $s*$ with respect to M :

$$\frac{1}{r^2} = \frac{1}{r_o^2} - 2\left(T - T_o\right)\frac{ds*}{dM^2}, \tag{9.40}$$

where the subscript o denotes evaluation at some particular level, which we here take to be in the outflow near the top of the cyclone. Since $ds*/dM^2$ is negative, this shows that the radii of angular momentum surfaces increase with decreasing absolute temperature (i.e. with altitude) and that the inverse squares of the radii increase linearly with temperature. This shows that the natural vertical coordinate in tropical cyclones is absolute temperature, and that angular momentum is a natural radial coordinate, with the radii of M surfaces being a dependent variable. The relation (9.40) gives the shape of characteristic surfaces in tropical cyclones, which are surfaces of constant M and s*, and are also vortex lines. Knowing r and $s*$ as functions of M gives a complete specification of the gradient wind and temperature above the boundary layer and outside the eye.

To solve (9.40) for $r(M)$, which would yield the full gradient wind distribution for the cyclone, we need to know $ds*/dM$ as a function of M and we also need to know r_o and T_o as functions of M. From a causal perspective, deep, moist slantwise convection originates in the boundary layer, and the principle of convective neutrality dictates that $s*$ in the free troposphere (along M surfaces) must equal the actual moist entropy, s, in the subcloud layer. Therefore, *the appropriate specification of the lower boundary condition amounts to specifying* ds/dM *in the boundary layer*.

An easy way to do this is to begin by writing the conservation equation for entropy in a coordinate system in which M serves as the radial coordinate. We can do this because M increases monotonically with radius, r, and so the coordinate transformation is well posed. In addition, we assume that moist entropy is well mixed in the vertical in the boundary layer, so that we can regard s as a function of time and M alone in the boundary layer:

$$s = s(t, M).$$

For this reason, the total time derivative of boundary layer entropy may be written

$$\frac{ds}{dt} = \frac{\partial s}{\partial \tau} + \frac{dM}{dt}\frac{\partial s}{\partial M}, \tag{9.41}$$

where we use τ for time in the first term on the right side to remind ourselves that the derivative in time is taken holding M (rather than r) constant. Following the derivation in section 9.1, in which surface enthalpy fluxes and dissipative heating are identified as the main sources of entropy in the boundary layer, we can write the integral through the depth of the boundary layer of (9.41) as

$$\int_0^{\delta}\left(\frac{\partial s}{\partial \tau} + \frac{dM}{dt}\frac{\partial s}{\partial M}\right)dz = \frac{C_k\,|\,\mathbf{V}_s\,|\left(k_0^* - k_s\right) + \mu C_D\,|\,\mathbf{V}_s\,|^3}{T_s}. \tag{9.42}$$

(Remember that μ is the fraction of dissipated kinetic energy that is assumed to occur locally in the atmospheric boundary layer.)

In deriving (9.42), note that there is no reference to radial velocity in the boundary layer, except, of course, that at the surface it contributes to the wind speed, $|\mathbf{V}_s|$.

The angular momentum sink in the boundary layer is given by

$$\int_0^h \frac{dM}{dt} dz = -C_D r_s |\mathbf{V}_s| V_s, \tag{9.43}$$

where V_s is the surface azimuthal wind speed and r_s reminds us that we are evaluating r at the surface here. Using this in the steady state version of (9.42) and remembering that ds/dM is not a function of height, we obtain

$$\frac{ds}{dM} = -\frac{C_k |\mathbf{V}_s| \left(k_0^* - k_s\right) + \mu C_D |\mathbf{V}_s|^3}{r_s C_D |\mathbf{V}_s| V_s T_s}, \tag{9.44}$$

Now if we multiply (9.40) through by M and use (9.44), equating ds/dM in the boundary layer with $ds*/dM$ just above the boundary layer, we obtain

$$\frac{M}{r^2} = \frac{M}{r_o^2} + \frac{T - T_o}{r_s V_s T_b}\left[\frac{C_k}{C_D}\left(k_0^* - k_s\right) + \mu |\mathbf{V}_s|^2\right]. \tag{9.45}$$

Remember that M is constant as we derived this by integrating along surfaces of constant M, so we can write

$$M = rV_g + \frac{1}{2} fr^2 = rV_{go} + \frac{1}{2} fr_o^2,$$

where V_{go} is the gradient wind at the outflow point. Using both these equalities in (9.45) gives

$$\frac{r_s}{r} V_g V_s = \frac{r_s}{r_o} V_{go} V_s + \frac{T - T_o}{T_b}\left[\frac{C_k}{C_D}\left(k_0^* - k_s\right) + \mu |\mathbf{V}_s|^2\right]. \tag{9.46}$$

A special case of (9.46) is its evaluation at the top of the boundary layer. Ignoring any difference between the radius of the M surface at the top and bottom of the boundary layer in the first two terms of (9.46), this special case reduces to

$$V_{gb} V_s = \frac{r_b}{r_o} V_{go} V_s + \frac{T_b - T_o}{T_b}\left[\frac{C_k}{C_D}\left(k_0^* - k_s\right) + \mu |\mathbf{V}_s|^2\right], \tag{9.47}$$

where V_{gb} is the gradient wind at the top of the boundary layer and r_b is the radius there.

Following the differential Carnot cycle development that led to (9.24), it can be shown that the first term on the right of (9.47) is related to the work done against friction in the outflow. Following the same scaling arguments as were presented in section 9.1.1, this work term can be neglected compared to the left side of (9.47) if we chose r_o to lie at or outside the radius where the azimuthal wind vanishes in the outflow. With this approximation, (9.47) becomes

$$V_{gb}V_s \cong \frac{T_b - T_o}{T_b}\left[\frac{C_k}{C_D}\left(k_0^* - k_s\right) + \mu\,|\,\mathbf{V}_s\,|^2\right]. \tag{9.48}$$

Comparing this to (9.24) leads to the interesting result that

$$V_{gb}V_s = |\,\mathbf{V}_s\,|^2\,. \tag{9.49}$$

The product of the gradient wind at the top of the boundary layer and the surface azimuthal wind speed equals the total surface wind speed squared. Since $V_s < \mathbf{V}_s$, it follows from (9.49) that

$$V_{gb} > |\,\mathbf{V}_s\,|\,. \tag{9.50}$$

The gradient wind at the top of the boundary layer must be larger than the total surface wind speed (which includes the contribution from the radial wind). Conversely, surface winds will generally be smaller than the gradient wind in steady-state conditions. It should be mentioned here that the inequality (9.50) does not preclude there being winds in the interior of the boundary layer that exceed both the gradient wind speed and the surface wind speed; in fact, there is much evidence that that is often the case, as will be explored in section 9.5. We also point out that the calculation of the gradient wind from (9.25) and (9.49) would require a somewhat detailed treatment of the boundary layer. Nonetheless, (9.25) is very useful by itself, as one often needs to know the surface wind speed.

Figure 9.4 compares the radial profile of surface azimuthal wind to the surface gradient wind *(left)* and the left side of (9.49) to the right side *(right)*, from an axisymmetric, nonhydrostatic, convection-permitting model with fixed, constant sea surface temperature run for long enough to achieve a quasi-steady state (Rotunno and Emanuel 1987). Figure 9.4 demonstrates that for this quasi-steady state of this simulation, while the surface azimuthal wind is less than the gradient wind outside the radius of maximum wind, the relation (9.49) is well satisfied.

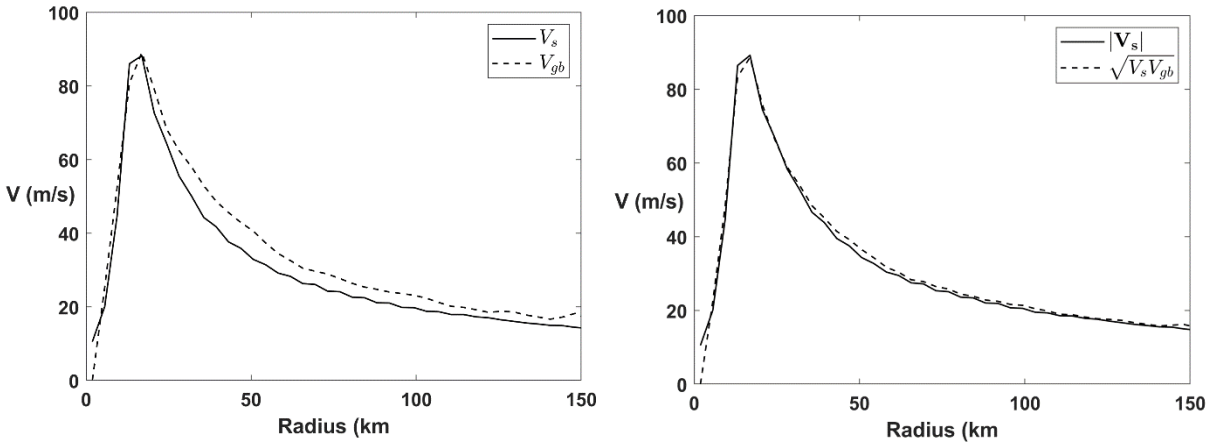


Figure 9.4: Radial profiles of the lowest grid point azimuthal velocity and boundary layer gradient wind (left), and lowest model level wind speed and the square root of the product of the lowest model level azimuthal wind and boundary layer gradient wind (right). These are taken from the quasi steady-states of a simulation using the model of Rotunno and Emanuel (1987).

Returning to (9.47), to close the system, one needs to know r_o, T_o, and k_s as a function of M, which, recall, serves as our independent radial variable. To fund these requires us to develop a theory of tropical cyclone outflow.

9.2.4 Theory of the outflow layer

The relations (9.25) and (9.48) for the boundary layer wind speeds should hold anywhere where the air above the boundary layer in saturated. (Note that (9.40) should hold everywhere outside the eye, but the boundary layer entropy equation must account for the sink of entropy by downdrafts outside the saturated core of the storm.) Observations clearly show that the surface wind speed reaches its maximum value well within the nearly saturated core, but equations (9.25) and (9.48) indicate that if the outflow temperature T_o is constant, wind speeds should increase with radius, owing to the large negative radial gradient of boundary layer enthalpy, k_s, there.

Recall that the first term on the right side of (9.47) vanishes if we assume that the outflow temperature is defined at the points that the angular momentum surfaces intersect a contour of vanishing gradient wind speed. Let's begin by looking at the variation of absolute temperature along the $V_g = 0$ contour in a numerical simulation using the axisymmetric, nonhydrostatic model of Rotunno and Emanuel (1987), run to a quasi-steady state as in Figure 9.4. In Figure 9.5, a family of surfaces of constant angular momentum, M, is plotted in a coordinate system in which azimuthal velocity is on the abscissa while the absolute temperature constitutes the ordinate.

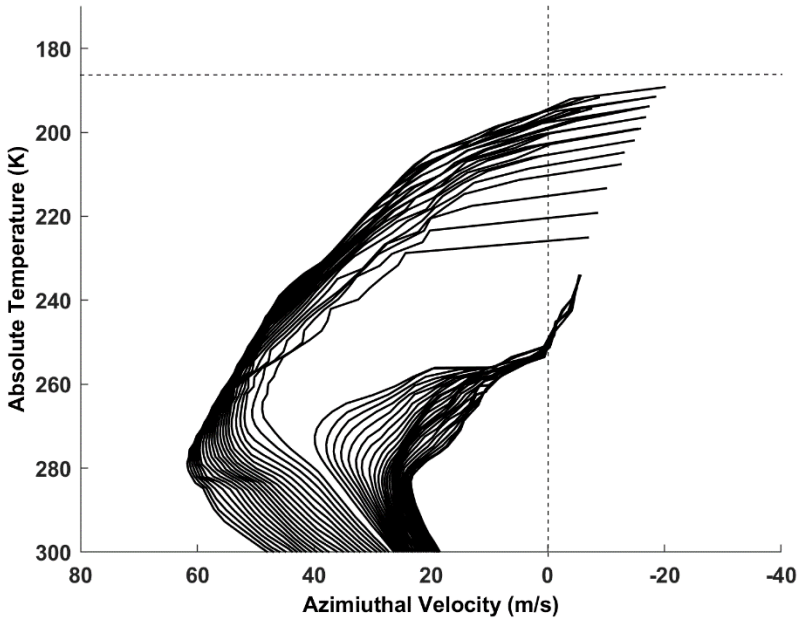


Figure 9.5: Contours of constant angular momentum plotted as a function of azimuthal velocity and absolute temperature from the quasi-steady state of a simulation using the nonhydrostatic, axisymmetric model of Rotunno and Emanuel (1987).

Clearly, the outflow temperature, defined at the intersection points of angular momentum surfaces with the zero azimuthal wind line, varies with M. This has also been found to be the case in simulations using three-dimensional, convection-permitting models (Tao et al. 2019). If we regard the distribution of T_o with M as the appropriate upper boundary condition, what determines this distribution?

As of this writing, this is regarded as an unsolved problem in tropical cyclone physics. In 2011, the author and his colleague (Emanuel and Rotunno 2011) proposed that shear in the outflow layer would lead to small-scale turbulence, the effect of which would be to hold the mean-flow Richardson Number at a critical value, given that turbulence is fast compared to the mean flow evolution. The Richardson Number is defined

$$Ri \equiv \frac{N^2}{\left|\frac{\partial \mathbf{V}}{\partial z}\right|^2}, \tag{9.51}$$

where N is the local buoyancy frequency and $\mathbf{V}$ is the horizontal wind vector. Stratified shear flows become unstable when Ri falls below a critical value that is order unity. The resulting instability is referred to as *Kelvin-Helmholtz instability* and leads to small-scale turbulence.

The nonhydrostatic, axisymmetric model of Rotunno and Emanuel (1987) uses a parameterization of small-scale turbulent diffusivity that depends, among other things, on the Richardson Number of the explicitly resolved flow, with a critical Ri of 1. There is no turbulent diffusivity when Ri is above this value.

Figure 9.6 shows the spatial distribution of vertical turbulent diffusivity, as a function of radius and height, averaged over the last 24 hours of a simulation of a mature tropical cyclone.

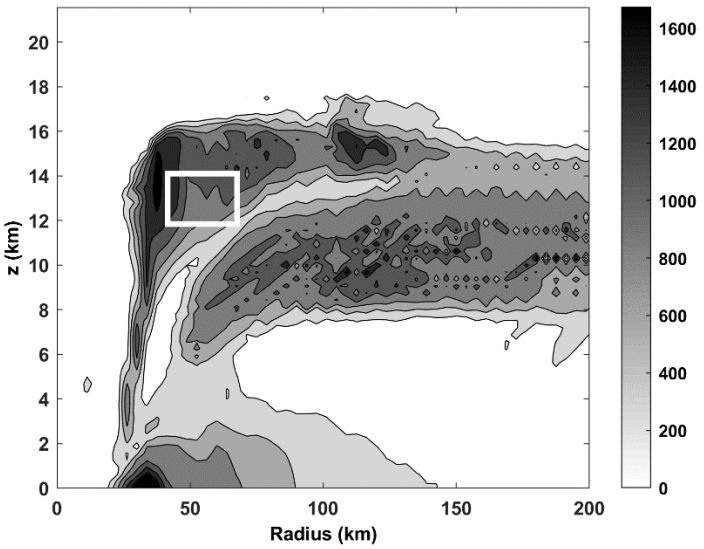


Figure 9.6: Vertical diffusivity ($m^2\ s^{-1}$) averaged over the last 24 hours of a numerical simulation of a tropical cyclone using the model of Rotunno and Emanuel (1987). The white box indicates an area in which sample of stratification and wind shear are collected.

It is clear that there is significant turbulence in most of the outflow layer; even more than in the boundary layer.

We tested the critical Richardson Number hypothesis by collecting co-located samples of wind shear and buoyancy frequency within the white rectangle shown in Figure 9.6. The result is shown in Figure 9.7.

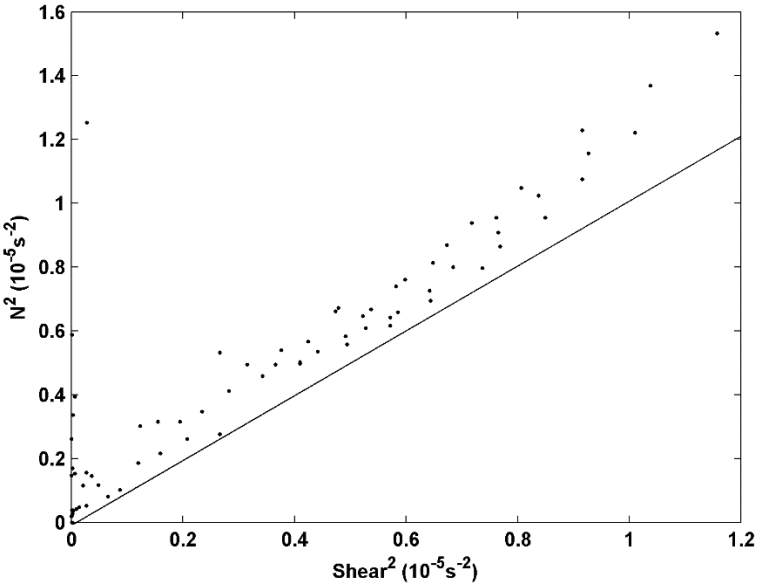


Figure 9.7: Square of the buoyancy frequency plotted against square of the vertical wind shear in samples collected from within the white box of Figure 9.6. Solid line shows Richardson Number =1.

The sample tend to line up in a region in which Ri is slightly larger than 1, suggesting that small-scale turbulence does tend to drive Ri back towards it critical value. (Note that several aspects of the computation performed here may distort the calculated value of Ri. First, we have neglected the contribution to Ri of vertical gradients in condensed water content, as it appears in eq. (28) of Rotunno and Emanuel (1987). Second, we have calculated Ri from the time-averaged wind and entropy fields rather than taking the time average of the instantaneous values of Ri.)

If we were to assume that the outflow has a Richardson Number close to its critical value, what does that imply about the upper boundary condition for the outflow temperature? We show here that, apart from a constant of integration, this assumption closes the mathematical system for the saturated inner core.

First, we neglect $\partial U / \partial z$ compared to $\partial V / \partial z$ in the denominator of (9.51), and, from the definition of angular momentum, substitute

$$\frac{\partial V}{\partial z} = \frac{1}{r}\frac{\partial M}{\partial z}. \tag{9.52}$$

Now the buoyancy frequency is a function of the entropy stratification and can be expressed as

$$N^2 = \Gamma_d \frac{\partial s_d}{\partial z} = \Gamma_m \frac{\partial s^*}{\partial z}, \tag{9.53}$$

where Γ_d and Γ_m are the dry and moist adiabatic lapse rates, respectively.

Since s^* is a function of M alone,

$$\frac{\partial s^*}{\partial z} = \frac{ds^*}{dM}\frac{\partial M}{\partial z}. \tag{9.54}$$

Using these in (9.51) gives

$$Ri \cong \frac{r^2 \Gamma_m \dfrac{ds^*}{dM}}{\left|\dfrac{\partial M}{\partial z}\right|}. \tag{9.55}$$

Thus, along any given streamline, Ri will be small at small radius and where $\left|\partial M / \partial z\right|$ is large.

If we next assume that in the outflow region, near radius $r = r_t$, Ri is equal to a critical value, Ri_c, then it follows from (9.55) that

$$\frac{\partial M}{\partial z} \cong \pm \frac{r_t^2 \Gamma_m \dfrac{ds^*}{dM}}{Ri_c}, \tag{9.56}$$

or, using

$$\frac{\partial s^*}{\partial z} = \frac{ds^*}{dM}\frac{\partial M}{\partial z},$$

this can be written

$$\frac{\partial s^*}{\partial z} \cong \pm \frac{r_t^2 \Gamma_m \left(\dfrac{ds^*}{dM}\right)^2}{Ri_c}. \tag{9.57}$$

Therefore, the assumption of critical Richardson number at some radius r_t constrains the vertical gradient of saturation entropy there. Since a negative vertical gradient of s^* would indicate absolute convective instability in saturated air, we choose the positive root in (9.57). Since s^* is a state variable, its vertical gradient dictates the vertical gradient of temperature, from which one can deduce the gradient of temperature with respect to s^* . Using the chain rule, we first write

$$\frac{\partial T}{\partial s*}=\left(\frac{\partial T}{\partial s*}\right)_p+\frac{\left(\frac{\partial T}{\partial p}\right)_{s*}}{\frac{\partial s*}{\partial p}}. \tag{9.58}$$

Using the definition of saturation entropy and the Clausius-Clapeyron equation, we can write the first term as

$$\left(\frac{\partial T}{\partial s*}\right)_p=\frac{\frac{T}{c_p}}{\left[1+\frac{L_v^2 q*}{R_v c_p T^2}\right]}. \tag{9.59}$$

Using the hydrostatic approximation for the second term in (9.58) and substituting (9.59), (9.58) can be written

$$\frac{\partial T}{\partial s*}=\frac{\frac{T}{c_p}}{\left[1+\frac{L_v^2 q*}{R_v c_p T^2}\right]}-\frac{\Gamma_m}{\frac{\partial s*}{\partial z}}. \tag{9.60}$$

Next, we substitute (9.57) for $\partial s*/\partial z$ in (9.60):

$$\frac{\partial T}{\partial s*}=\frac{\frac{T}{c_p}}{\left[1+\frac{L_v^2 q*}{R_v c_p T^2}\right]}-\frac{Ri_c}{r_t^2\left(\frac{ds*}{dM}\right)^2}. \tag{9.61}$$

It proves convenient to put this in the form of a gradient with respect to angular momentum rather than saturation entropy, which we can do by multiplying (9.61) through by $ds*/dM$, and to write the result as an expression for the $M-$ dependence of the outflow temperature:

$$\boxed{\frac{\partial T_o}{\partial M}=\frac{\frac{T_o}{c_p}\frac{ds*}{dM}}{\left[1+\frac{L_v^2 q*(T_o)}{R_v c_p T_o^2}\right]}-\frac{Ri_c}{r_t^2\left(\frac{ds*}{dM}\right)}.} \tag{9.62}$$

9.2.5 Putting it all together: A closed model of a slantwise neutral tropical cyclone

To formulate a complete model of a slantwise-neutral, quasi-balanced, steady-state tropical cyclones, we begin by re-writing the integrated thermal wind equation (9.40) as

$$\boxed{M_b = r_b^2\left(\frac{1}{2}f - \left(T_b - T_o\right)\frac{ds*}{dM}\right)}, \tag{9.63}$$

where we have made use of the fact that $V = 0$ at the outflow point. Next, we simplify (9.44) by making the approximation that $|\mathbf{V}_s| \simeq V_s$ and write that expression as

$$\frac{ds*}{dM} = -\frac{C_k}{C_D}\frac{s_0^* - s*}{r_s V_s} - \mu\frac{V_s}{r_s T_s}, \tag{9.64}$$

in which we have also used the approximation

$$\frac{k_0^* - k}{T_s} \cong s_0* - s*,$$

where s_0* is the saturation entropy of the sea surface and we have equated the boundary layer entropy with the saturation entropy of the free troposphere, assuming convective neutrality.

Next, we approximate V_s in (9.64) by $M/r_s - \frac{1}{2}fr_s$, giving

$$\left[\frac{T_s\left(1-\mu\right) + \mu T_o}{T_s}\right]\frac{ds*}{dM} = \frac{C_k}{C_d}\frac{s_0^* - s*}{r_s^2\left(T_s - T_o\right)\frac{ds*}{dM}}. \tag{9.65}$$

Finally, making the approximations $r_b \simeq r_s$ and $T_b \simeq T_s$, we eliminate r_b^2 between (9.63) and (9.65), resulting in a quadratic equation for $ds*/dM$:

$$\left(\frac{ds*}{dM}\right)^2 + 2\chi\frac{ds*}{dM} - \frac{\chi f}{T_b - T_o} = 0, \tag{9.66}$$

where

$$\boxed{\chi \equiv \frac{C_k}{C_D}\frac{s_0* - s*}{2M}\frac{T_b}{T_b\left(1-\mu\right) + \mu T_o}}. \tag{9.67}$$

The solution of (9.66) that has physical meaning is

$$\boxed{\frac{ds^*}{dM} = -\chi - \sqrt{\chi^2 + \frac{\chi f}{T_b - T_o}}.} \tag{9.68}$$

In all these derivations, we have ignored the difference between the surface temperature, T_s and the temperature at the top of the subcloud layer, T_b, and the difference between r_b and r_s.

To summarize, our system consists of the boxed equations (9.62) and (9.68) with the definition (9.67). These equations can be integrated, regarding the angular momentum, M, as the independent variable. Once s^* and T_o are found as functions of M, the physical radius can be found using (9.63), and then the azimuthal wind speed can be deduced from the definition of angular momentum, M.

To carry out the integration, we need to supply two boundary conditions. At an outer radius, r_{outer}, we specify that $V = 0$ and therefore the value of M there is $\frac{1}{2} f r_{outer}^2$. At this radius we demand that the saturation entropy, s^*, has its undisturbed environmental value s_e^*, and for V to vanish there, from (9.63) we must have $T_o = T_b$ at that outer radius. Using these boundary conditions, we can march (9.62) and (9.68) inward in angular momentum space, using (9.63) to diagnose r and the definition of M to diagnose V.

In carrying out this integration, we choose not to allow the outflow temperature T_o to become lower than the ambient, cold point tropopause temperature. We consider that the radius of maximum winds has been reached when T_o equals the ambient tropopause temperature T_t.

Before exploring numerical solutions of this system, we show that under some simplifying assumptions it admits analytic solutions. These assumptions are discussed and justified in Emanuel and Rotunno (2011), and here we merely summarize them:

1. We neglect dissipative heating, setting $\mu = 0$
2. We neglect the pressure dependence of the sea surface saturation entropy s_0^*, holding it constant
3. We neglect the last term under the square root in (9.68)
4. We neglect the first term on the right of (9.62)
5. We neglect the first term on the right of (9.63)
6. We use the approximation that $V_m >> \frac{1}{2} f r_m$, where V_m is the maximum azimuthal wind speed and r_m is the radius of maximum winds.

With these approximations, the analytic solution is easiest to write in terms of the angular momentum as a function of radius:

$$\left(\frac{M}{M_m}\right)^{2-\frac{C_k}{C_D}} = \frac{2\left(\frac{r}{r_m}\right)^2}{2-\frac{C_k}{C_D}+\frac{C_k}{C_D}\left(\frac{r}{r_m}\right)^2}, \tag{9.69}$$

where M_m is the approximate angular momentum at the radius of maximum winds: $M_m = V_m r_m$. The maximum wind speed in the solution is given by

$$V_m^2 \cong V_p^2 \left(\frac{1}{2}\frac{C_k}{C_D}\right)^{\frac{\frac{C_k}{C_D}}{2-\frac{C_k}{C_D}}}, \tag{9.70}$$

where here V_p is a nominal potential intensity under the assumptions 1 and 2 above:

$$V_p^2 \equiv \frac{C_k}{C_D}\left(T_b - T_t\right)\left(s_0 * - s_e *\right), \tag{9.71}$$

where $s_e *$ is the saturation entropy of the unperturbed free troposphere. Note that in contrast to the potential intensity expressions written earlier in this chapter, (9.71) is a closed expression involving purely environmental properties. The actual maximum wind speed given by (9.70) is a modification of this. In the special case that $C_k = C_D$, $V_m = V_p / \sqrt{2}$. The actual maximum wind speed is less than the nominal environmentally based potential intensity, given by (9.71), by a factor of $1/\sqrt{2}$. This in no way contradicts expressions (9.25) and (9.48), in which the boundary layer enthalpy or entropy is evaluated at the radius of maximum winds rather than in the far environment. In general, the boundary layer enthalpy and entropy will be larger at the radius of maximum winds than in the unperturbed environment, yielding a smaller potential intensity than given by (9.71) but a maximum wind speed that may indeed be consistent with (9.70).

The relations (9.70) and (9.71) show that the maximum wind speed has a complex-looking dependence on the ratio of the enthalpy exchange coefficient to the drag coefficient. Figure 9.8 graphs this dependence as the ratio of the maximum wind speed given by (9.70 and (9.71) to the quantity $\sqrt{\left(T_b - T_t\right)\left(s_0 * - s_e *\right)}$.

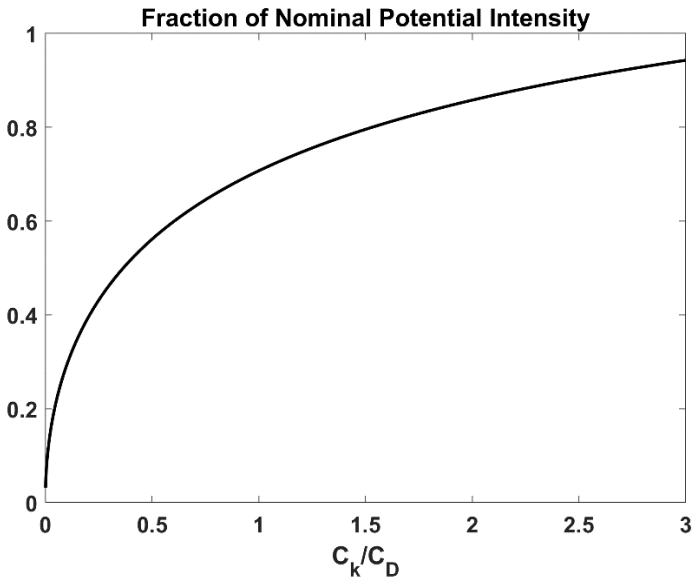


Figure 9.8: Ratio of the maximum wind speed given by (9.70) and (9.71) to the quantity $\sqrt{(T_b - T_t)(s_0^* - s_e^*)}$.

For very small values of C_k / C_D, the normalized intensity increases fairly rapidly (as the square root of C_k / C_D in the limit in which that ratio is very small), but for larger values of the ratio, the dependence is much weaker, asymptoting to $\sqrt{2}$ at very large values of the ratio.

This framework has three important length scales in it: The radius of maximum winds, r_m, the outer radius r_o at which the surface winds vanish, and the scale r_t that appears in (9.56). It turns out that only one of these three needs to be specified and the other two can be determined as part of the solution. Here we choose to specify r_o and determine both r_m and r_t. These solutions are

$$r_m \cong \left(\frac{1}{2}\right)^{\frac{3}{2}} \frac{f r_o^2}{\sqrt{(T_b - T_t)(s_0^* - s_e^*)}}, \tag{9.72}$$

and

$$r_t^2 = r_m^2 \frac{C_D}{C_k} Ri_c. \tag{9.73}$$

Curiously, the radius of maximum winds does not depend on the exchange coefficients but is sensitive to the outer radius.

These approximate analytic solutions may be compared to the full numerical solutions of (9.62), (9.63), and (9.68). To carry out these integrations, we first choose a large value of r_t and then integrate (9.62) and (9,68) inward starting from a chosen outer radius, r_o. We integrate all the way in to $r = 0$ and note the value of the outflow temperature T_o at the radius of maximum

winds. It will generally be larger than the desired boundary condition that $T_o = T_t$ at $r = r_m$. We then reduce the value of r_t and start over, repeating this cycle until $T_o = T_t$ at $r = r_m$. (The algorithm is inefficient but nevertheless fast.) Unlike the approximate analytic model, the numerical solution includes the full pressure dependence of the sea surface saturation entropy, s_0*, and can include any value μ of the coefficient governing dissipative heating. Unless specified as an option, the code does not make any of the assumptions (1-6) listed above. The reader interested in experimenting with this code and having access to MATLAB may download the model, *steady_state_model.m*, from https://zenodo.org/doi/10.5281/zenodo.10936382 .

To begin with, we compare the simplified analytic solution to a version of the numerical solution that applies approximations 1, 2, and 4 from the list above. This version does, however, include the Coriolis terms neglected in the simplified analytic model. Steady-state profiles of azimuthal velocity are shown for two different values of the outer radius in Figure 9.9 and compared to the analytic solutions. For the analytic model and the numerical solutions, the Coriolis parameter has been set to $5 \times 10^{-5} s^{-1}$, the surface exchange coefficients are equal to each other, the temperature of the boundary layer top is 300 K and that of the tropopause is 200 K. Reasonable values of the surface thermodynamic disequilibrium have been chosen.

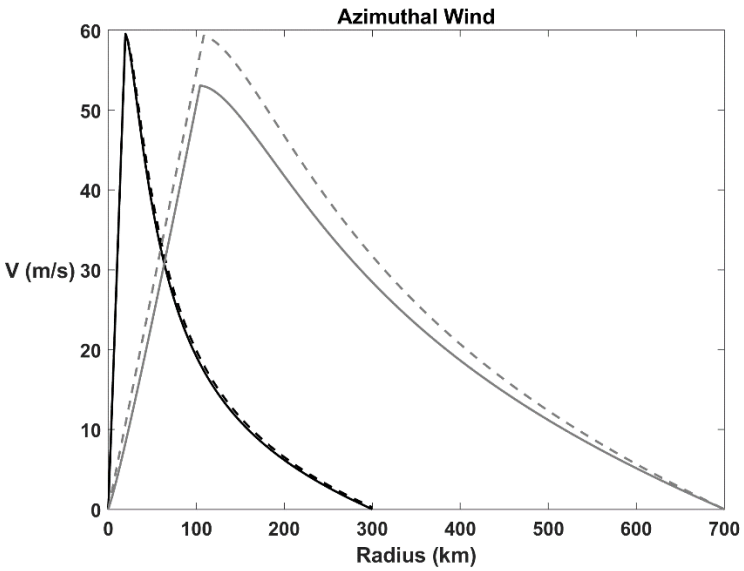


Figure 9.9: Steady-state radial profiles of azimuthal velocity from the numerical solution of (9.62), (9.63), and (9.68), neglecting dissipative heating and the pressure dependence of the surface saturation entropy (solid curves), for specified outer radii of 300 km (black) and 700 km (gray). These are compared to analytic solutions of the simplified model (dashed).

In the case of an outer radius of 300 km, the numerical and analytic solutions are virtually identical. But when the outer radius is 700 km, the numerical solution is slightly more concentrated but somewhat weaker than the analytic solution. This is because of the effect of the Coriolis terms missing from the analytic solution.

Physically, there are two effects that operate to weaken larger vortices in the same environment. The first is dissipation of wind energy in the outflow, which increases as the radius of maximum wind increases (see (9.16) and associated discussion). Quantitatively, though, this

explains at most about $1\,ms^{-1}$ of the $\sim 7\,ms^{-1}$ difference between the analytic and numerical profiles shown in Figure 9.9 in the case of the outer radius being 700 km. The larger of the two effects can understood by referring back to the expression for potential intensity, (9.26):

$$V_p^2 = \frac{T_b - T_o}{T_o}\frac{C_k}{C_D}\left(k_0^* - k_s\right).$$

Remember that this is not a closed expression. In the steady-state model developed here, the outflow temperature at the radius of maximum winds was assumed to be the tropopause temperature, T_t. But the enthalpy at the radius of maximum winds, k_s, is determined as part of the solution of the model. Recall that the assumption of moist convective neutrality links k_s to the temperature of the free troposphere above. The key is to realize that as one increases the diameter of the vortex, this temperature must increase, and therefore, so must k_s.To see why the temperature must increase, start with the gradient balance equation, applied near the surface of the vortex:

$$\alpha\frac{\partial p}{\partial r} = \frac{V^2}{r} + fV \tag{9.74}$$

Now define a nondimensional radius normalized by the radius of maximum winds:

$$r_n \equiv \frac{r}{r_m}.$$

With this substitution, integrate (9.74) inward from the outer radius to the radius of maximum winds (normalized radius of 1):

$$p_m = p_0 - \int_1^{r_o/r_m}\left(\frac{V^2}{r_n} + fVr_m\right)dr_n, \tag{9.75}$$

where p_m is the surface pressure at the radius of maximum winds and p_0 is the ambient surface pressure. Now *if* the wind profiles for different outer radii are self-similar (actually, they are not), then V would be, by definition, a universal function of r_n. But as one increases the diameter of the vortex, the second term on the right of (9.75) gets larger because r_m increases. Thus the surface pressure gets lower at the radius of maximum winds. Because the horizontal pressure gradients become weaker higher up in the storm, this implies, hydrostatically, that the temperature must increase and, owing to convective neutrality, k_s must increase as well, weakening the enthalpy flux and, with it, the intensity. This is the most important reason why larger tropical cyclones will be weaker, all other things being the same.

In the numerical solutions of this simple model, it is possible to include the effects of dissipative heating and of the surface pressure dependence of the sea surface saturation enthalpy, s_0*. Figure 9.10 compares the steady-state wind profile without either effect to profiles that include just dissipative heating, and both effects together.

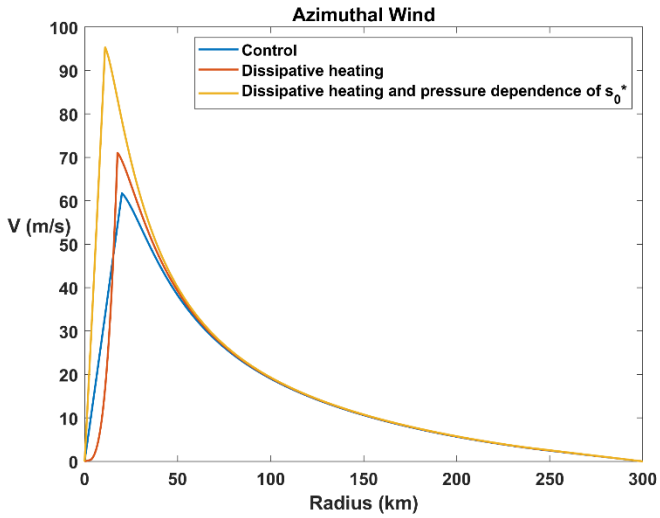


Figure 9.10: Steady-state radial profiles of surface azimuthal wind for the standard experiment with neither dissipative heating nor pressure dependence of s_0* *(blue), just dissipative heating (red), and both effects together (yellow).*

Dissipative heating increases the intensity by roughly 15%, moves the radius of maximum winds inward, and makes the profile more concave inside the radius of maximum winds. Allowing for the pressure dependence of the surface saturation entropy by itself (not shown here) raises the maximum intensity by slightly more (16%) than the dissipative heating alone. But including both effects together results in a 54% increase in the maximum wind speed, showing that the two effects feedback on one another and produce an increase somewhat larger than adding the two effects linearly.

In fact, the nonlinearity of the pressure dependence of the sea surface saturation entropy can cause the system to run away at sufficiently high sea surface temperature. In this case, as the system intensifies, the surface pressure drops at the radius of maximum winds, causing an increase in s_0*, further amplifying the intensity. At high temperature, this feedback increases more rapidly than the $|\mathbf{V}|^3$ dependence of the frictional dissipation, and the intensity continues to increase. The inference here is that, eventually, internal dissipation of wind kinetic energy becomes large enough to balance the energy input from the thermal cycle. The resulting storm has been called a *hypercane* and has been simulated with a full-physics, nonhydrostatic axisymmetric model (Emanuel et al. 1995). Hypercanes appear to have far more concentrated cores than regular tropical cyclones, and extend deep into the stratosphere where they would likely deposit large quantities of water, compromising the ozone layer. Under earth-like conditions, sea surface temperature would locally have to rise as high as 50°C to sustain a hypercane. This may have happened in the geological past as a result of a bolide impact or undersea flood basalt episodes.

Finally, to complete the description of the flow in this steady-state model, we estimate the radial and vertical components of the flow. In the steady state, the secondary circulation arises purely from the frictional inflow in the boundary layer and any sources and sinks of angular momentum aloft.

We can estimate the radial flow in the boundary layer from the steady-state balance of angular momentum, which can be written, for axisymmetric flow, as

$$\rho u \frac{\partial M}{\partial r} = -r \frac{\partial \tau_\theta}{\partial z}, \tag{9.76}$$

where τ_θ is the azimuthal component of the turbulent stress and we have neglected any vertical advection of angular momentum in the boundary layer. If we integrate (9,76) through the depth δ of the boundary layer, defined such that the azimuthal stress vanishes at its top, the result is

$$\overline{\delta u \rho \frac{\partial M}{\partial r}} = r\tau_{\theta s} = -C_D r \rho \,|\, \mathbf{V}_s \,|\, V_s, \tag{9.77}$$

where the overbar signifies an average over the depth of the boundary layer, $\tau_{\theta s}$ is the surface azimuthal stress, and we have substituted the aerodynamic drag formulation. We make several further simplifications by ignoring the vertical variation of density and $\partial M / \partial r$, replacing the surface wind speed by its azimuthal component, and ignoring the difference between V_s and the gradient wind speed, resulting in

$$\overline{u} \simeq \frac{-C_D r V^2}{\delta \, \partial M \big/ \partial r}. \tag{9.78}$$

With a specification of the boundary layer depth, the Coriolis parameter, and the drag coefficient, we can use (9.78) to estimate $\overline{u}$ from the modelled distribution of V.

Once we have $\overline{u}(r),$ we can calculate the vertical velocity at the top of the boundary layer from mass continuity:

$$w_b = -\frac{1}{r}\frac{\partial}{\partial r}\left(r\delta\overline{u}\right). \tag{9.79}$$

The radial and vertical velocities corresponding to the numerically integrated control case described earlier in this chapter are shown in Figure 9.11.

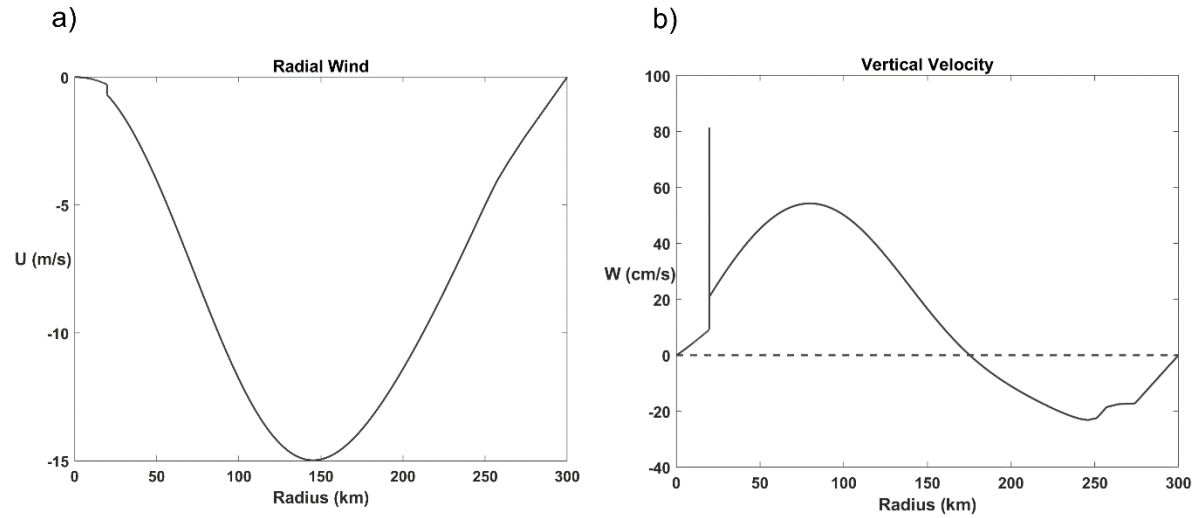


Figure 9.11: Radial (a) and vertical (b) velocities at the top of the boundary layer, calculated from the azimuthal solution shown by the solid blue curve in Figure 9.9 using (9.78) and (9.79). The vertical velocity is in cm s^{-1} and zero vertical velocity is denoted by the dashed line.

The radial velocity is inward everywhere and peaks in this case at about 140 km from the storm center, in contrast to the azimuthal velocity, which peaks at around 20 km. The vertical velocity has a prominent upward spike at the radius of maximum azimuthal wind. This is owing to the assumed discontinuity in the radial gradient of outflow temperature, which has been calculated from (9.61) from the radius of maximum winds outward, but has assumed to be constant inside the radius of maximum winds. The discontinuity in the gradient of outflow temperature at the radius of maximum winds introduces a discontinuity in the radial gradient of azimuthal wind, which, according to the denominator of (9.78), introduces a discontinuity in the radial gradient of the radial wind. This yields a delta function spike in w, according to (9.79). (The delta function is rendered finite by the numerical integration over finite intervals of M.)

Outside the spike of w at the radius of maximum winds, the vertical velocity has a broad positive peak at around 80 km radius, and changes sign to downward motion around 180 km.

Figure 9.11b reveals several unrealistic aspects of this steady-state model. First, the vertical motion inside the radius of maximum winds is upward, in contrast to observations which show gentle subsidence in the eye. There is no evidence, in either observations or full-physics models, of a spike of upward motion in the eyewall embedded in a broader region of ascent that maximizes well outside the radius of maximum winds; indeed, as we shall see, there is often strong descent not far outside the eyewall. Finally, the descent in the outer region in full-physics models is more on the order of 1 cm s^{-1} or less; indeed, larger descent rates would cause extreme subsidence warming.

The upper boundary condition on the model pertains specifically to the outflow. But for M surfaces originating at the top of the boundary layer in places where the vertical velocity is negative, there should be inflow and downflow and the physical reasoning behind our upper boundary conditions fails. Moreover, the flow of low entropy air into the boundary layer by both the mean flow and any convective downdrafts has been ignored. Consequently, the model presented here, if it is valid anywhere, should be valid in the nearly saturated inner core region of mature, axisymmetric tropical cyclones.

Another circumstance in which this model may have some validity is in the case of dry tropical cyclones, as described in the next sub-section.

9.2.6 Dry and moist tropical cyclones

Tropical cyclones are driven by surface enthalpy fluxes, and on the face of it, there is no obvious reason why they should not be possible in systems in which either there is no phase change of water or else the phase change is completely reversible; that is, the air is everywhere saturated, there is no precipitation, and condensed water evaporates in downward moving air just as it condenses in upward moving air.

Beginning with the work of Mrowiec et al. (2011), tropical cyclones have been simulated in atmospheres with no phase change of water and ambient dry adiabatic lapse rates in the troposphere, with large jumps between the surface and air temperatures. In this case, the thermodynamic disequilibrium between the surface and the atmosphere leads to purely sensible turbulent heat fluxes from the surface. Nevertheless, a vortex does spin up in such a dry atmosphere, but with some interesting differences in intensity and structure. Dry hurricanes can also develop in fully three-dimensional simulations, as shown by Cronin and Chavas (2019), who also provide a comprehensive analysis of the physics underlying the intensity and structure of the cyclones.

To illustrate some fundamental properties of dry tropical cyclones, we modify the nonhydrostatic, axisymmetric model of Rotunno and Emanuel (1987) to have no water vapor in the initial state and no evaporation from the ocean. The model is initialized using a sounding created with the MIT single-column model run into a steady state with no surface evaporation, creating an essentially dry adiabatic troposphere. The surface temperature is then modified to achieve a potential intensity comparable to that of the control (moist) simulation (about $80\ ms^{-1}$).

The evolution over time of the maximum surface azimuthal wind is shown in Figure 9.12, together with the surface wind from the moist control simulation. The dry vortex begins to intensify sooner than the moist control, but intensifies more slowly and achieves a maximum that is substantially less than the control, even though the potential intensity is very nearly the same.

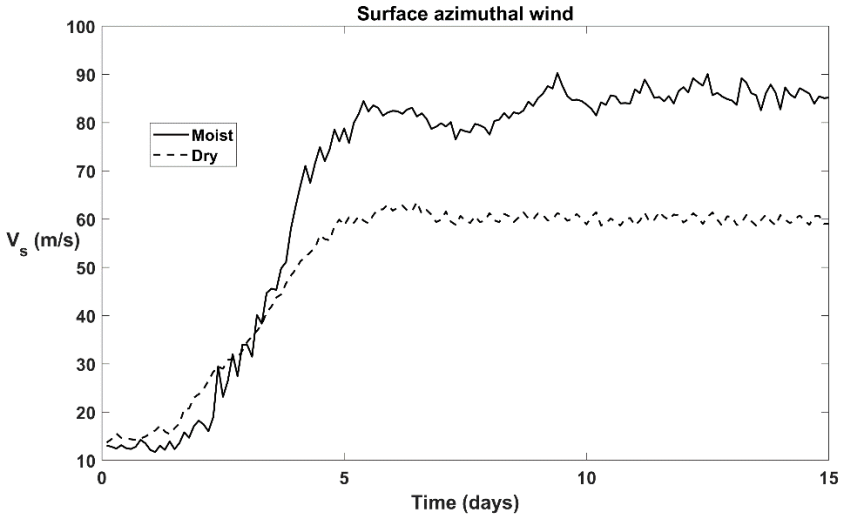


Figure 9.12: Evolution of the maximum surface wind in two simulations using the axisymmetric, nonhydrostatic model of Rotunno and Emanuel (1987). The control, moist simulation is shown by the solid curve, while the dry simulation is shown in dashed. The potential intensity in each simulation is around 80 ms^{-1}.

Figure 9.13 compares cross-sections through the mature stages of moist and dry simulations. The vortex in the dry simulation is shallower in physical space but spans roughly the same temperature difference across the tropopause as the moist simulation.

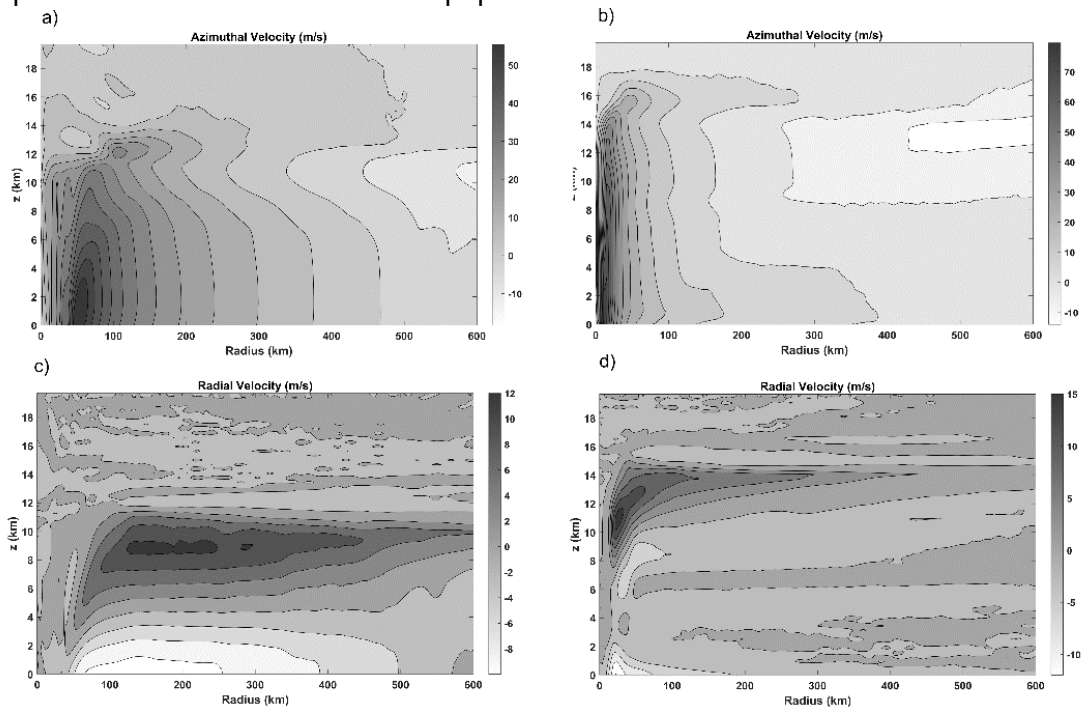


Figure 9.13: Contour plots of quantities averaged over a 72-hour period during the quasi-steady state of simulations using an axisymmetric, non-hydrostatic model. Top row: Azimuthal velocity (ms^{-1}) of a) dry and b) moist simulations. Bottom row: Radial velocity (ms^{-1}) in dry (c) and moist (d) simulations.

The moist vortex is much more concentrated, with a substantially smaller radius of maximum winds. Note the dramatically different structure of the radial velocity of the dry simulations. The radial flow varies roughly linearly with height through the depth of the circulation, whereas outflow in the moist simulation is strongly concentrated near the tropopause.

Plots of the turbulent diffusivity (not shown here) reveal, unsurprisingly, that most of the troposphere throughout the whole radial extent of the dry vortex is turbulent, in contrast with the moist simulation in which turbulence is confined mostly to the saturated region.

The presence of turbulence throughout much of the domain might imply that the critical Richardson Number hypothesis holds throughout the vortex. One way to test this idea is to compare the radial profile of azimuthal velocity produced by the numerical model with that of the simple steady state model described in section 9.2.5. In this case we use the numerical solution of the simple model, taking into account both dissipative heating and the pressure dependence of the sea surface enthalpy, and also use the same surface exchange coefficients, sea surface temperature, tropopause temperature, and environmental air temperature as the full Rotunno and Emanuel model. For the simple model we specify an outer radius of 500 km. The comparison of the simple and full models, shown in Figure 9.14. is excellent, suggesting that dry TC-like vortices are well represented by the steady-state model described in section 9.2.5. In the next section, we will see that this is not the case for outer regions of moist tropical cyclones.

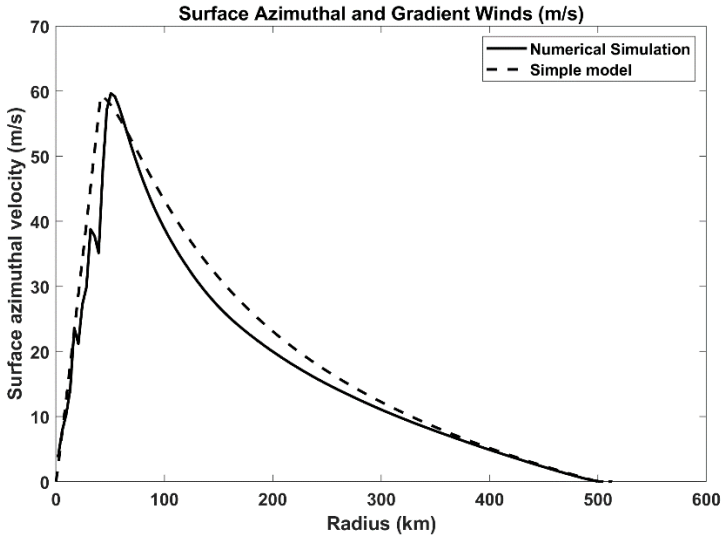


Figure 9.14: Radial profile of near-surface azimuthal velocity from the near steady-state of the dry simulation using the full model of Rotunno and Emanuel (1987; solid) and the simple steady-state model described in section 9.2.5 (dashed).

9.3 Axisymmetric steady structure theory

9.3.1. Outer wind field

The steady-state model developed in section 9.2 assumes that tropical cyclones have outflow density stratification that reflects Richardson numbers near their critical values. While this appears to be close to the case for numerically simulated dry vortices, it does not hold for the outer regions of moist tropical cyclones. Figure 9.15 compares the Rotunno and Emanuel (1987) model radial profile of surface azimuthal velocity to that the simple model, with the outer radius set to 290 km.

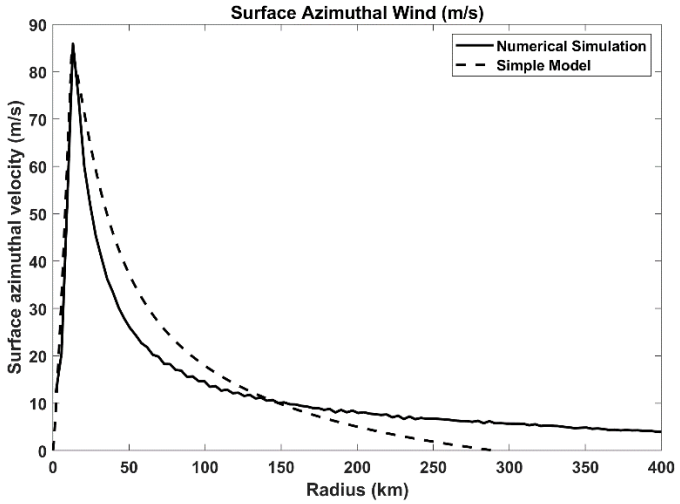


Figure 9.15: Radial profile of near-surface azimuthal velocity from the near steady-state of the fully moist simulation using the full model of Rotunno and Emanuel (1987; solid) and the simple steady-state model described in section 9.2.5 (dashed).

From just outside the radius of maximum winds outward, the simple model wind field does not match the numerically simulated wind profile very well. Outside the radius of maximum winds, the numerically simulated wind profile falls off much more steeply than the steady state model, but then far less steeply, forming a long tail out to well beyond 400 km.

Inspection of the fields generated by the full-physics axisymmetric model show that there is very little turbulence outside of about 50 km radius (whereas the radius of maximum winds is only about 12 km), and below a relatively thin outflow layer. This region, nearly devoid of deep convection, has a thermal stratification not very different from that of the unperturbed environment. Since horizontal temperature gradients are weak in this region, to a good approximation, there is a balance between radiative cooling and large-scale subsidence, as given by (3.21):

$$w_e = \frac{Q_{rad}}{\rho \, dh_d / dz}, \tag{9.80}$$

where Q_{rad} is the radiative cooling rate, ρ is the air density, and dh_d/dz is the vertical gradient of dry static energy. For a cooling rate of around $1^o C \, day^{-1}$ and a typical moist adiabatic lapse

rate of around $6^o C\ km^{-1}$, this amount to around $3\ mm\,s^{-1}$. Now compare this to the downward motion at the top of the boundary layer, in the outer region of the cyclone, given by the simple model described in section 9.2.5 and shown in Figure 9.11b. The vertical motion in the steady state is entirely owing to Ekman pumping/suction. This has a negative peak of about $200\ mm\,s^{-1}$, nearly two orders-of-magnitude larger than the downward motion that would be balanced by radiative cooling, as given by (9.80). This would result in very rapid warming of the air above the boundary layer and is incompatible with the assumption of a steady state.

The physics of the outer region of a moist tropical cyclone is clearly not governed by the hypothesis of Richardson Number criticality of the cyclone's overturning circulation, as postulated for the inner region. Instead it has been proposed (Emanuel 2004; Chavas and Emanuel 2014) that the low-level azimuthal flow profile in the outer region is such that the Ekman suction associated with it is exactly equal to the radiative subsidence rate given by (9.80). If this is true, the boundary layer momentum can be in equilibrium while at the same time the flow above the boundary layer can be in thermodynamic equilibrium.

Since neither the radiative cooling rate nor the thermal stratification are likely to vary greatly in the horizontal, we will assume that w_e as given by (9.80) is constant. The radial velocity in the boundary layer needed to balance the inflow of mass from above is given by integrating through the boundary layer the axisymmetric mass continuity equation:

$$w_e = -\frac{1}{r}\frac{\partial}{\partial r}\left(r\delta\overline{u}\right), \tag{9.81}$$

where $\overline{u}$ is the mass-weighted vertically averaged radial velocity in the boundary layer.

If we integrate this inward from a radius r_o at which the radial (and azimuthal) velocities are assumed to vanish, we arrive at

$$\overline{u}\delta = -\frac{1}{2}w_e\frac{\left(r_o^2 - r^2\right)}{r}. \tag{9.82}$$

By our hypothesis, the azimuthal velocity must be such that the Ekman radial velocity must be equal to that implied by the radiative subsidence, and given by (9.82). This Ekman velocity is given by (9.78):

$$\delta\overline{u} \simeq \frac{-C_D r V^2}{\partial M \big/ \partial r}. \tag{9.83}$$

Equating the radial velocities given by (9.82) and (9.83) yields an equation for the azimuthal velocity. Introducing $m \equiv rV$, this is

$$\frac{dm}{dr} = \frac{2C_D}{w_e}\frac{m^2}{r_o^2 - r^2} - fr, \tag{9.84}$$

where we have made use of the definition of angular momentum, $M = rV + \frac{1}{2} fr^2$.

Equation (9.84) is a first-order, nonlinear ordinary differential equation that has the canonical form of a Riccati equation, named after the 17th and 18th century Venetian mathematician Jacopo Francesco Riccati. Unfortunately, it does not have an analytic solution, but it is fairly straightforward to integrate it numerically[5]. Figure 9.16 shows such an integration, taking the outer radius to be $600\ km$, the radiative subsidence rate to be $2.2\ mm\,s^{-1}$, and the Coriolis parameter and drag coefficient to be the same as the values in the full-physics numerical simulation. Equation (9.84) is integrated in to the radius of maximum winds and compared to the steady-state azimuthal wind profile produced by the full-physics model, as in Figure 9.15.

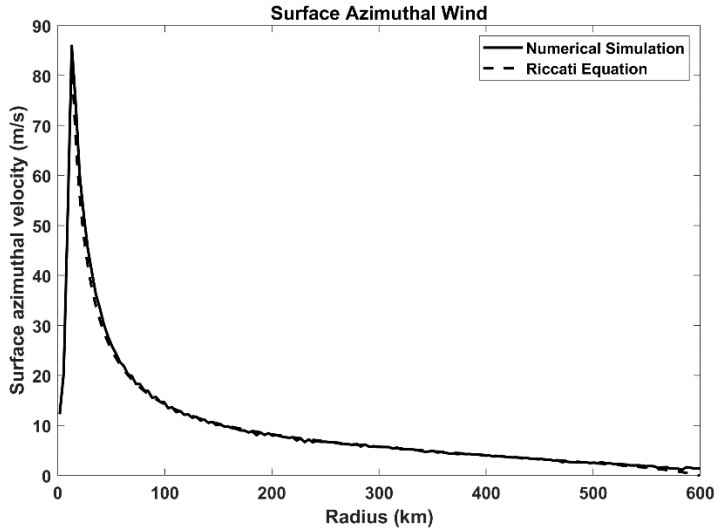


Figure 9.16: Radial profile of near-surface azimuthal velocity from the near steady-state of the fully moist simulation, as in Figure 9.15 (solid) compared to the solution of the Riccati equation (9.84) as described in the text (dashed).

It is hard to tell the difference between the solution of the Riccati equation and the results of the full-physics simulation. Chavas et al. (2015) compared solutions of (9.84) with radial profiles of surface azimuthal velocity observed in actual tropical cyclones and found quite good agreement.

The physics of the outer region is another illustration of the fact that the real importance of moisture to dynamics is not the condensation of water vapor but, rather, the irreversible fallout of condensed water, forcing the descending branch of the circulation to be dry and thereby to cause work to be done against the buoyancy force. In fact, Wang and Lin (2020) performed simulations with a full-physics axisymmetric model comparing moist and dry simulations, and a moist reversible simulation in which condensed water does not fall and the troposphere is largely saturated, with condensed water evaporating reversibly where the motion is downward. They showed that the dry and reversible moist simulations resemble each other, and their structure and behavior resemble that of the simple model described in section 9.2.5.

[5] Cronin (2023) has developed a fast algorithm for solving (9.84).

9.3.2 Merged inner and outer wind profiles

We expect that as one moves radially inward in a moist tropical cyclone, deep convection will eventually break out and the regime will transition to saturated upflow, where the inner core model developed in section 9.2.5 would prevail. Such a transition may be gradual, but here we model it as abrupt, and attempt to match the inner solution to the outer solution described above in section 9.3.1. Later in this chapter we will relax this assumption.

Because the Ekman flow solution given by (9.83) prevails everywhere and since $\delta\overline{u}$ must be continuous to avoid singularities in the vertical velocity given by mass continuity (9.81), we must insist that the right side of (9.83) be continuous across the radius at which we match the inner and outer solutions. This requires the ratio of V^2 to $\partial V / \partial r$ to be continuous. On the other hand, any upward jump in V itself, as one moves inward, would entail a jump in absolute angular momentum, M. Such an upward jump would be strongly inertially unstable, and any such instability would smooth out the discontinuity. For this reason, we take V to be continuous at the matching point, and thus, through the continuity of the right side of (9.83), $\partial V / \partial r$ must be continuous as well.

The inner solution, whether solved for numerically or through the analytic approximation given by (9.69) and (9.70), requires specification of the potential intensity, V_p the ratio of the surface exchange coefficients, and the radius of maximum wind, r_m. (The outer radius used in the inner core model is irrelevant to the merged solution, as the transition point will always be inside this outer radius.) The outer wind field, given as the solution of (9.84) requires specification of the surface drag coefficient, the radiative subsidence rate, w_e, the Coriolis parameter, f, and an outer radius.

Given the external parameters $V_p, C_k, C_D, w_e,$ and f, we could solve for the merged system in one of two ways. First, we could specify the radius of maximum winds, r_m, calculate the inner solution, and then solve for the outer solution (9.84), varying the outer radius, r_o, until the radial gradient of V matches the inner solution at the point that the two solutions intersect. Alternatively, we could specify the outer radius, r_o, and solve (9.84) for the outer wind field. Then we could apply the inner wind field model, varying r_m until we find a solution at which the radial slopes of V match at the radius where the two wind fields intersect. This means that we will always have to specify one or the other of r_o and r_m; we cannot determine the overall size of the cyclone by this technique.

Figure 9.17 shows an example of a merged solution. Chavas et al. (2015) compared such solutions to radial profiles of surface azimuthal wind in a set of observed tropical cyclones, finding very good agreement with observations. In each case, the overall storm size was chosen to minimize the difference between the merged model and observations.

Although merged solutions like these satisfy dynamical and thermodynamical constraints in the inner core, thermodynamic balance in the quiescent, subsiding air in the rain-free outer region, and dynamical balance in the boundary layer, they do not necessarily satisfy the thermodynamic

budget of the outer region boundary layer, where surface fluxes, radial advection, and subsidence of low entropy air into the boundary layer from aloft are all in play. We will tackle this issue in the following subsection.

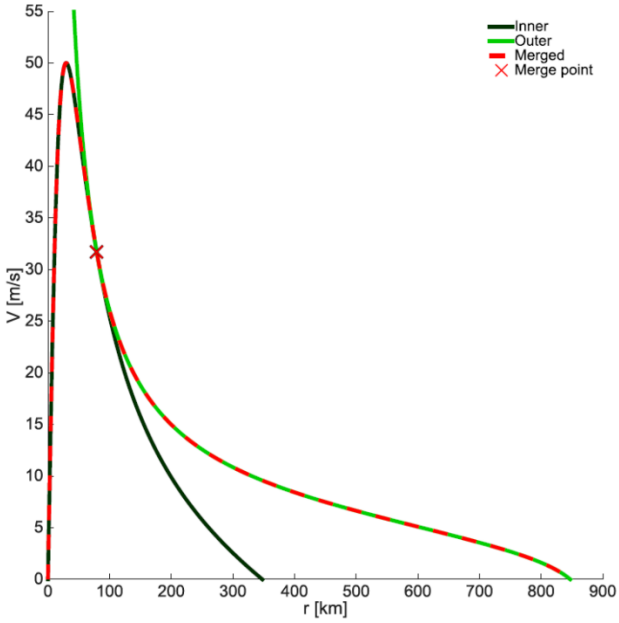


Figure 9.17: Example of an azimuthal wind profile representing the merger of an inner wind field (black) based on saturated moist conditions with a critical outflow Richardson number, and an outer wind field (green) based on matching Ekman suction to radiative subsidence. The dashed red line shows the merged solution, while the cross shows the matching point. The maximum wind speed has been specified as $50\ ms^{-1}$ *and the radius of maximum wind at 30 km, with* $C_k = C_D = 1\times10^{-3}$, $w_e = 2\,mm\,s^{-1}$, and $f = 5\times10^{-5}\,s^{-1}$ *; this yields an outer radius of 847 km.*

9.3.3 Thermodynamic balance in the outer boundary layer and equilibrium cyclone size

In formulating the outer boundary layer in section 9.3.1, we took for granted that the boundary layer air would gradually moisten as it moves radially inward, and that deep moist convection would erupt at the matching point. Yet, in the outer region, surface winds are weak, not providing enough enthalpy flux to counter the influx of low entropy (θ_e) air into the boundary layer, so the boundary layer air is likely to be stable to deep moist convection. Indeed, simulations with the moist, full-physics model of Rotunno and Emanuel (1987) show little or no deep convection outside roughly three times the radius of maximum winds.

The requirement that the boundary layer entropy increase inward at least as much as is needed to produce conditional instability may set a bound on the absolute storm size in the moist case. If the entropy increases too quickly, then moist convection will erupt in what had been assumed to be a deep-convection-free outer region. If it does not increase quickly enough, then convection will not be able to occur in the inner core, which we have assumed to be fully convecting.

Here we add to the merged inner and outer model described in the previous subsection by developing a simple model for the moist static energy of the boundary layer, and using that to ensure the thermodynamic consistency of the merged solution.

In constructing such a solution, it is helpful to identify four regions within tropical cyclones, as illustrated in Figure 9.18.

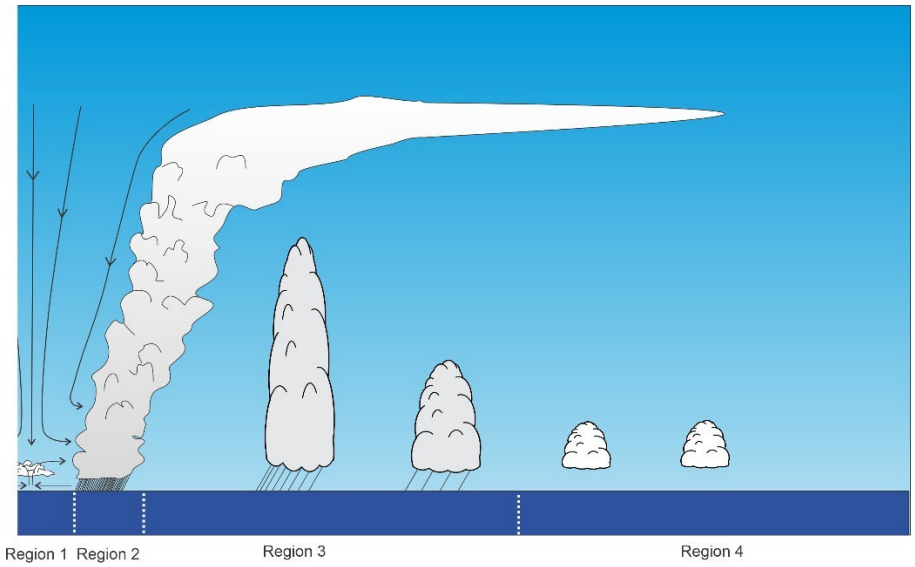


Figure 9.18: Four regions of a tropical cyclone, moving radially outward from the center, at left. Region (1) is the eye of the storm, discussed in section 9.5. Region 2 is the inner core, in which is air is saturated, or nearly so, on the large scale. Region 3 has some deep convection, but air is not saturated on the large scale. The far outer region, Region 4, is free of deep convection but may contain shallow convection.

Region 1 is the eye of the storm, discussed in section 9.5. It is free of deep convection though it may contain very shallow cumulus and or stratocumulus. Region 2 is the inner core, in which is air is saturated, or nearly so, on the large scale, and is ascending. Region 3 has some deep convection, but air is not saturated on the large scale, and on the storm scale air may be weakly ascending or descending. The far outer region, Region 4, is free of deep convection but may contain shallow convection; air is subsiding.

Of these regions, only Region 2 exists in all mature tropical cyclones, essentially be definition. Not all storms have eyes; even some well-developed tropical cyclones may have deep convection near their centers. Region 3 exists in some tropical cyclones, but in others, the shallow convection regime may extend almost all the way in to the Region 2, leaving no regime of deep convection into unsaturated air. These are known as “annular hurricanes” (Knaff et al. 2003). Finally, some tropical cyclones have at least some deep convection extending virtually all the way to the edge of the surface circulation. In this development, we will assume that Region 4 always exists. We will also postpone discussion of the eye, Region 1, to section 9.5.

First, let us consider some relationships that we expect to hold throughout Regions 2-4 (and some in region 1 as well). These are mass continuity (9.79):

$$w = -\frac{1}{r}\frac{d}{dr}\left(r\delta\overline{u}\right), \tag{9.85}$$

where w is the vertical motion at the top of the boundary layer, $\overline{u}$ is the depth-averaged radial velocity in the boundary layer, and δ is the depth of the boundary layer. (In the following development, for simplicity, we ignore height variations of horizontal velocities and moist static energy in the boundary layer.) Second is the momentum balance of the boundary layer, (9.78):

$$\delta\overline{u}\frac{dM}{dr} = r\tau_{\theta s} = -C_D r\,|\,\mathbf{V}\,|\,V, \tag{9.86}$$

where we have also ignored variations of density with depth in the boundary layer. Here, for convenience, we introduce a new variable y defined as

$$y \equiv \frac{r\,|\,\mathbf{V}\,|\,V}{dM\,/\,dr}, \tag{9.87}$$

so that

$$\delta\overline{u} = -C_D y, \tag{9.88}$$

and by mass continuity (9.85),

$$C_D\frac{dy}{dr} = -rw. \tag{9.89}$$

We also assume that an approximate form of the thermal wind relation (9.40) holds throughout. Making use of the relationship $Tds^* = dh^*$, where s^* is the saturation moist entropy and h^* is the saturation moist static energy, and defining outflow temperature, T_o, as the temperature along an angular momentum surface where the azimuthal wind vanishes, we have

$$\frac{T_b - T_o}{T_b}\frac{dh^*}{dM} = -\frac{V}{r}, \tag{9.90}$$

where we evaluate the expression just above the top of the boundary layer, where the temperature is T_b and we assume that the azimuthal velocity is the same as that in the boundary layer.

Region 4

We next turn our attention specifically to Region 4, whose dynamics we have already discussed in subsection 9.3.1. Namely, we set w at the top of the boundary layer to the subsidence velocity, w_e, as defined by (9.80), and the solve (9.89) for y, integrating inward from a specified outer radius. Given y, we then solve (9.87) for V, and (9.88) for δu. We also solve (9.90) for h^*, assuming that T_b and T_o are constant.

We solve for the boundary layer moist static energy, h_b, which should always be less than the saturation moist static energy aloft, by definition, so that deep convection cannot occur. But our solution will show that h_b increases inward, and may eventually become as large as the

saturation moist static energy of the free troposphere, h*, in which case we transition to Region 3 or possibly directly to Region 2.

The conservation relation for boundary layer moist static energy balances radial advection with surface fluxes and downward advection of lower moist static energy from above the boundary layer. Here we neglect one other possibly important source of moist static energy, which is the dissipation of wind energy, which should be relatively smaller in the outer region. The result is

$$\delta\overline{u}\,\frac{dh_b}{dr} = C_k\,|\,\mathbf{V}\,|\left(h_0^* - h_b\right) - w_e\left(h_b - h_m\right), \tag{9.91}$$

where C_k is the surface enthalpy exchange coefficient, h_0^* is the saturation moist static energy of the sea surface (which we approximate here as a constant), and h_m is some representative value of moist static energy above the boundary layer. If we use (9.88) for $\delta\overline{u}$ and make use of (9.87), the result is

$$\frac{dh_b}{dM} = \frac{1}{rV}\left[\frac{w_e}{C_D\,|\,\mathbf{V}\,|}\left(h_b - h_m\right) - \frac{C_k}{C_D}\left(h_0^* - h_b\right)\right]. \tag{9.92}$$

Integrating this inward allows us to determine the boundary layer moist static energy. Note that we have to integrate (9.92) inward simultaneously with (9.89), (9.87), and (9.90). Also note that we must, in Region 4, specify the lower tropospheric moist static energy, h_m as there is no deep convection to control it. Consequently, we regard it as a free, constant parameter, h_{m0}.

In performing the inward integration, we specify an outer radius, r_o, at which the azimuthal wind speed, V, vanishes. (We do not necessarily set $|\,\mathbf{V}\,|$ to zero there, allowing for background wind and/or a gust factor.) From (9,87) we see that y must vanish at the outer radius, while inward integration of (9.90) demands that we specify a value for the saturation moist static energy, h*, at the outer boundary.

Finally, the boundary layer moist static energy budget (9.92) demands a value of h_b at the outer boundary. But because V vanishes there, the quantity in brackets in (9.92) must also vanish. This shows that the value of the boundary layer moist static energy at the outer radius is a weighted average of the sea surface saturation moist static energy and the lower tropospheric moist static energy:

$$h_{b0} = \frac{w_e h_{m0} + C_k v_{gust} h_0^*}{w_e + C_k v_{gust}}, \tag{9.93}$$

where we have taken $|\,\mathbf{V}\,|$ at the outer radius to be equal to a surface wind gustiness measure, v_{gust}. If we omit any gustiness, then the only viable solution is $h_{b0} = h_{m0}$. (We might also regard v_{gust} as a proxy for background mean flow.) The outermost boundary layer is not observed to be anything like that dry, so omitting gustiness is not realistic. We also must choose the parameters in (9.93) to be such that the outer boundary layer moist static energy is less than the

tropospheric saturation moist static energy $\left(h_{b0} < h^*\right)$, else our supposition that deep most convection is absent will be wrong.

Also, inspection of (9.92) shows that if (9.93) is obeyed, the boundary layer moist static energy will always increase inward, just inside the outer radius, if $|\mathbf{V}|$ does so as well. Numerical integration of this system confirms this, but we postpone displaying this solution in favor of presenting the merged solutions for Regions 2, 3 and 4 together.

Region 3

If the solution just derived for Region 4 predicts that $h_b > h^*$ at any point before (outward of) the matching point to the solution of Region 2 is reached, deep convection is implied outside of Region 2. Note that there is a special class of solutions for which h_b becomes as large as h^* just as the matching point to Region 2 is reached; such solutions will have no Region 3. It is also important to note that if the outer radius is too small or the boundary layer moist static energy at the outer radius is too small, then the boundary layer will still be stable to deep convection at the matching point, on the inner side of which the troposphere is assume to be saturated and neutral to deep, slantwise convection. Since both h_b and h^* must be continuous across the matching point, such a solution is not viable. This implies that *there is an absolute lower limit on equilibrium storm size*.

We formulate the physics of region 3 in terms of the boundary layer quasi-equilibrium (BLQE) postulate described extensively in Chapter 3. Part of this formulation holds that in deep convecting regions, $h_b = h^*$, so we do not need to solve a boundary layer equation for h_b, we just solve the thermal wind equation (9.90). The boundary layer physics enters instead through the BLQE formulation of the convective updraft mass fluxes.

First recall that the net downward moist static energy flux into the boundary layer is $\left(M_c - w\right)\left(h_b - h_m\right)$, where M_c is the convective updraft mass flux and w is the large-scale vertical velocity near the top of the boundary layer. We equate this with the net source of moist static in the boundary layer:

$$\left(M_c - w\right)\left(h^* - h_m\right) = C_k\,|\mathbf{V}|\left(h_0^* - h^*\right) - \delta\overline{u}\,\frac{dh^*}{dr}, \tag{9.94}$$

where we have made use of the fact that $h_b = h^*$ in moist convecting regions and we have neglected the vertically integrated radiative cooling of the boundary layer. Now if we use the boundary layer momentum balance (9.86) for the radial velocity in the last term, and the thermal wind relation (9.90) for dh^*/dr, this can be written

$$\left(M_c - w\right)\left(h^* - h_m\right) = C_k\,|\mathbf{V}|\left(h_0^* - h^*\right) - \frac{T_b}{T_b - T_o}C_D\,|\mathbf{V}|V^2. \tag{9.95}$$

Also recall from Chapter 3 that the thermal equilibrium of the free troposphere requires that the subsidence in between clouds balances adiabatic and radiative cooling:

$$\epsilon_p M_c = w + \frac{Q_{rad}}{\rho \, dh_d / dz} = w + w_e, \tag{9.96}$$

where ϵ_p is the precipitation efficiency and we have made use of (9.80).

We can now eliminate M_c between (9.95) and (9.96) to derive a relation between h_m and the large-scale ascent rate:

$$h^* - h_m = \epsilon_p \, |\mathbf{V}| \frac{C_k \left(h_0^* - h^* \right) - \dfrac{T_b}{T_b - T_o} C_D V^2}{\left(1 - \epsilon_p \right) w + w_e}. \tag{9.97}$$

For the present purpose, we will be specifying $\epsilon_p, h_0^*, C_k, C_D, w_e,$ and $\left(T_b - T_o \right) / T_b$. To close the system we need to find the large-scale vertical velocity, w. To do so, we use the steady state version of the relation (3.64), governing the vertical mean moist static energy of the tropopause. The steady version of that relation, using the present notation, is

$$0 = -\delta \overline{u} \frac{dh^*}{dr} - Gw + C_k \, |\mathbf{V}| \left(h_0^* - h^* \right) - \mathbf{S} H w_e, \tag{9.98}$$

where G is the gross moist stability, H is a tropospheric depth scale, and $\mathbf{S}$ is the dry static stability, dh_d / dz. Once again using the boundary layer momentum balance (9.86) and the thermal wind relation (9.90), this can be written

$$w = \frac{1}{G} \left[|\mathbf{V}| \left(C_k \left(h_0^* - h^* \right) - C_D \frac{T_b}{T_b - T_o} V^2 \right) - \mathbf{S} H w_e \right]. \tag{9.99}$$

Therefore, in marching radially inward, we can use (9.99) to find w and (9.97) to find h_m.

Region 2

The solution for the inner core, assuming saturation and a critical Richardson Number in the outflow, was derived in section 9.2.5. For simplicity, rather than solving the full equation set, we here use the analytic approximation given by (9.69). Even though this solution is not valid in the eye, we carry it into the center for now, but return to the issue of eye dynamics in section 9.4.

Matched solutions

Given the solution in Regions 2-4, we first match the outer part of the solution to Region 2 to the solution in Region 3, if it exists, or else directly to Region 4. We match both the azimuthal wind speed, V and its radial slope.

Using this procedure, we construct solutions for the steady-state structure of tropical cyclones. In general, we find that as one varies the specified outer radius (with all other control parameters fixed), viable solutions exist only for a finite range of outer radii. Below some critical

outer radius, the boundary layer air flowing into Region 3 is stable to deep convection, which is inconsistent with the assumption that the inner core is fully convecting. At larger outer radii, Region 3 does not exist and the inner core (Region 2) transitions directly to the convection-free outer region (Region 4), with the boundary layer air becoming unstable just as it transitions from Region 4 to Region 2. For yet larger outer radii, all three Regions exist (not counting the eye). Beyond a second critical outer radius, matching of the inner core to the solution in Region 3 fails and we assume that no steady solution exists.

An example of a solution containing Regions 2, 3 and 4 is shown in Figure 9.19 for the following parameter values: $r_o = 800\,km$, $w_e = 3\ mm\,s^{-1}$, $V_p = 60\ ms^{-1}$, $\mathbf{S}H = 1\times10^4\ m^2 s^{-2}$, $G = 0.2\mathbf{S}H$, *gust factor* = $1.8\,ms^{-1}$, $f = 5\times10^{-5}\,s^{-1}$, $C_D = C_k = 1.5\times10^{-3}$, $\chi = 5$, and $\epsilon_p = 0.9$.

For this set of parameters, the profile of azimuthal velocity is not too terribly different from that shown in Figure 9.17, which matches the inner core directly to Region 4 without accounting for boundary layer moist static energy. The radius of maximum winds is only about *16 km* for these parameters.

Region 4 extends from about 110 km outward; there is, by definition, no deep convection and the large-scale descent is that of the specified radiative subsidence velocity. The free tropospheric moist static energy remains at the specified low value compared to its saturation value, which increases inward almost imperceptibly. The boundary layer entropy, which starts out quite low, owing to the small gust factor in the surface fluxes, recovers inward, reaching convective neutrality at the boundary with Region 3.

Region 3 extends from about 110 km inward to around 40 km. Here the middle troposphere is much moister and the boundary layer moist static energy remains neutral to deep moist convection. The deep convective mass flux is around $10\,cm\,s^{-1}$.

The inner core, by design here, is saturated through the whole troposphere and has the analytic approximation to the azimuthal wind profile given by (9.69). Upward motion in the core is about a factor of 6 stronger than in Region 3.

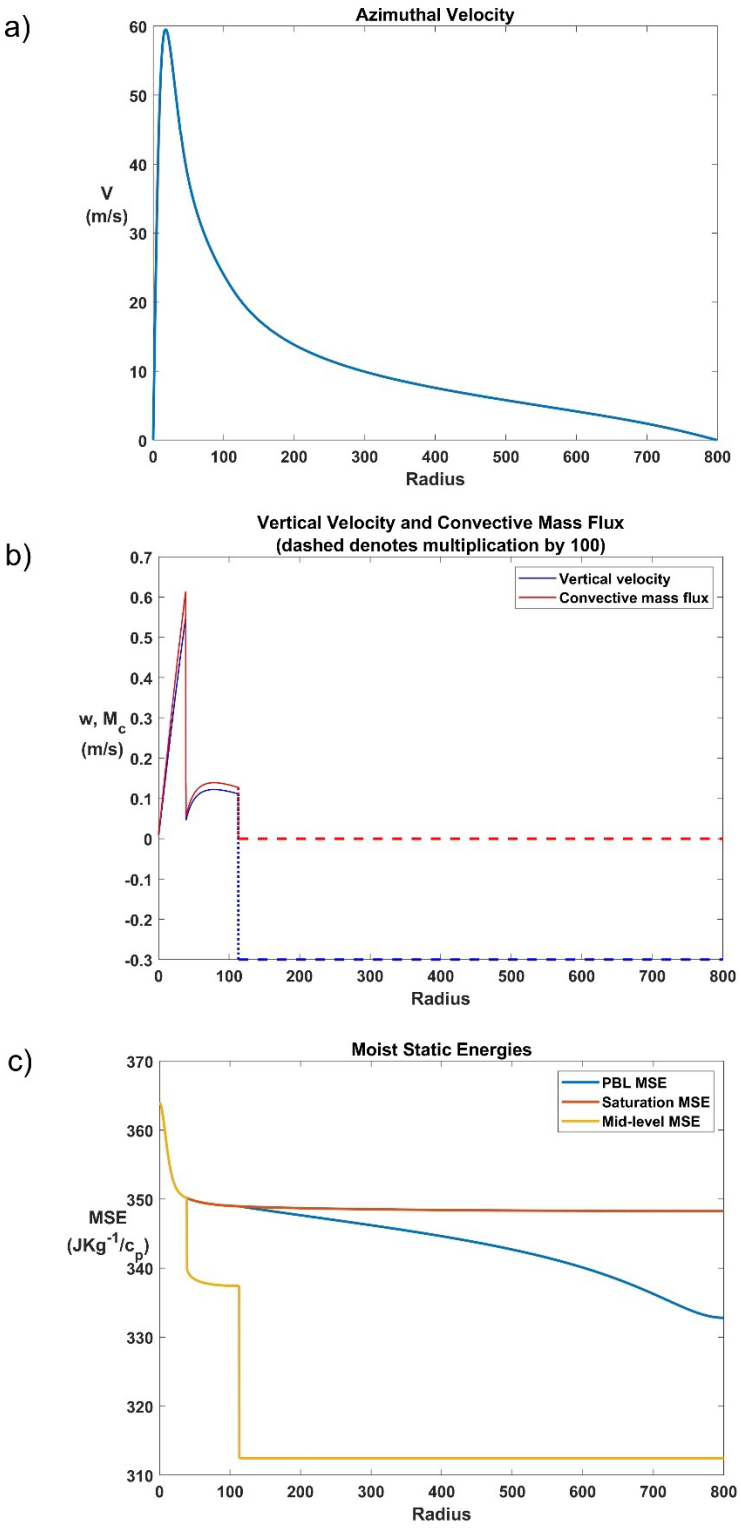


Figure 9.19: An example of a solution to the simple steady-state model, in which Regions 2-4 exist; see text for parameter values. Panel a) shows the azimuthal velocity field, panel b) displays both the large-scale vertical velocity and the convective updraft mass flux, and panel c) shows the moist static energies of the boundary layer and middle troposphere, as well as the saturation moist static energy of the free troposphere. Note that the Region 4 vertical velocities have been multiplied by 100.

Experiments varying the size of the outer radius, r_o, show that as it is reduced, the size of Region 3 shrinks and then vanishes at about $r_o = 650\,km$. Further reductions in outer radius result in the boundary layer moist static energy not recovering sufficiently, moving inward, so that it is still subcritical to deep moist convection moving into Region 2. Such solutions are therefore not viable. For these parameters, the equilibrium storm size cannot be less than about *650 km*.

As the outer radius is increased, Region 3 expands in size but the deep convective mass flux at its inner edge decreases. When r_o is greater than about *830 km*, the deep convective mass flux becomes negative at the inner edge of Region 3, and the solution is no longer viable.

For this set of parameters, then, the only viable steady solutions are for outer radii between 650 and 830 km. Accounting for the boundary layer moist static energy budget restricts, but does not entirely determine equilibrium storm size.

While the equilibrium size of tropical cyclones is naturally of interest, it may have little to do with the size of storms in nature. Experiments with full-physics numerical models suggests that it takes many tens of days for a tropical cyclone's size to equilibrate, which is somewhat longer than most storms last.

9.4 The tropical cyclone eye

Perhaps no other feature of tropical cyclones is a as prominent as its eye. Not all tropical cyclones have clearly defined eyes, and there are others whose eyes are invisible or nearly so in satellite imagery owing to high, cirrus overcast. Chapter 8.4.3 presents and reviews observations of tropical cyclone eyes. Here we briefly explore their dynamics.

The circulation is tropical cyclone eyes is thermally indirect, with relatively buoyant air sinking near the center and relatively less buoyant air ascending inside the eyewall. The eye circulation is mechanically driven from the eyewall, whose spin-up is strongly concentrated in a frontal zone some distance from the center of circulation (see section 10.1.2). Descent in the eye has a strong component proportional to the rate of spin-up of the eyewall and a much weaker component proportional to the ratio of radiative cooling to dry static stability (the radiative subsidence velocity); this latter component is the only one present in the steady state.

The essential dynamics of the eye can be understood with a simple thought experiment that could also be implemented as a laboratory experiment, as illustrated in Figure 9.20. A glass bucket or beaker mostly filled with water is placed on a steadily rotating turntable. After some time, the water in the vessel will be in solid body rotation with the same angular velocity as the turntable.

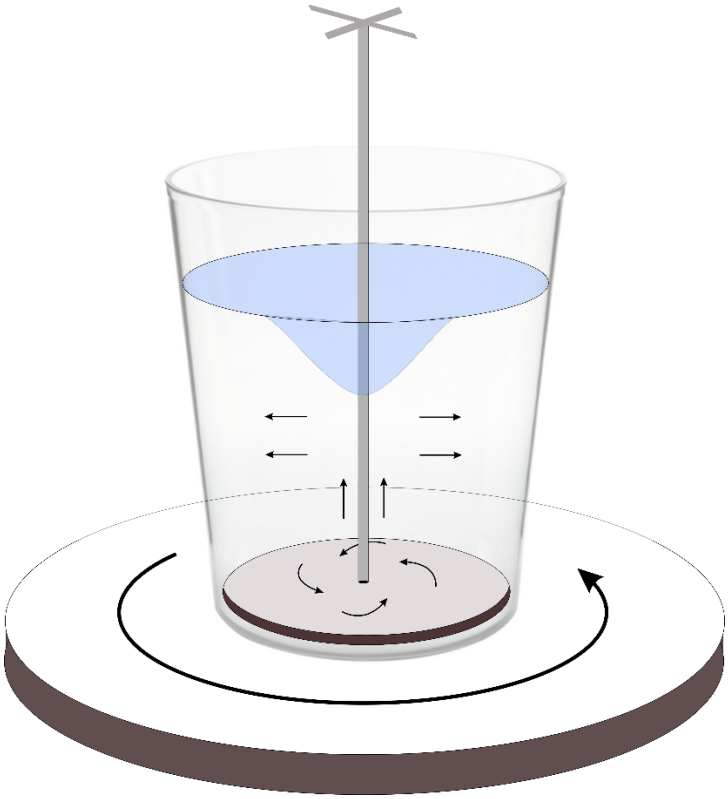


Figure 9.20: Thought experiment and possible laboratory experiment: A glass bucket or beaker of water is placed on a rotating turntable. A disk on a spindle is lowered to the bottom of the vessel and is kept stationary while the vessel rotates around it. The black arrows show the resulting motion of the water, which could be made visible by suspending small particles in the water.

Next, a solid disk on a spindle is lowered to the bottom but prevented from rotating. Owing to friction between the water and the stationary disk, an Ekman layer forms, with fluid spiraling inward and ascending in the inner portion of the vessel. By mass continuity, the water must flow outward above the Ekman layer, as shown by the black arrows in Figure 9.20.

The Ekman layer angular momentum balance is classical, and the same as we have assumed in our treatment of the tropical cyclone boundary layer earlier in this chapter: Loss of angular momentum to the stationary disk by surface friction is balanced by inward advection of higher angular momentum from the outer region.

Above the Ekman layer, the outflowing fluid initially advects low angular momentum outward, resulting in a spin-down of the fluid in the inner region. This results in a concave profile of azimuthal velocity, with angular velocity increasing outward. If the flow is laminar, the viscous diffusion will create and inward angular momentum flux that will eventually balance the outward advection of low angular momentum. If the flow is turbulent, as is the case in a tropical cyclone, the turbulent flux will also be down the gradient of angular velocity. Either way, the angular momentum balance in the interior can be expressed by

$$u\frac{\partial M}{\partial r} = \frac{1}{r}\frac{\partial}{\partial r}\left(r^3 \upsilon \frac{\partial}{\partial r}\left(\frac{V}{r} \right) \right), \tag{9.100}$$

in which u is the radial velocity, M is the angular momentum per unit mass, V is the azimuthal velocity, and υ is either the molecular kinematic viscosity, if the flow is laminar, or the turbulent diffusivity if it is turbulent.

Since we expect the radial velocity u to be positive above the boundary layer and $\partial M / \partial r$ will be positive (if for no other reason than negative gradients would be strongly unstable to inertial instability), then the right side of (9.100) must be positive, so that angular velocity will increase with radius: the profile of V will be concave. Since water is incompressible, the azimuthal velocity V in equilibrium cannot vary with height, as that would imply vertical pressure gradients, that cannot be balanced. We infer that neither u nor V can vary with height and we may replace the partial derivatives in (9.100) with total derivatives.

Moreover, the vertically integrated radial mass flux above the boundary layer must be equal and opposite to the radial mass flux in the boundary layer. The latter is given by (9.86) which we write here as

$$\delta u_b \frac{dM}{dr} = -C_D r V^2, \tag{9.101}$$

approximating $|\mathbf{V}|$ by V and denoting the boundary layer radial wind by u_b. Thus $\delta u_b = -Hu$, where H is the depth of the fluid above the boundary layer. Using this relation and combining (9.101) with (9.100) yields a single nonlinear ODE for the azimuthal velocity. Given that the cores of most tropical cyclones have angular velocities well in excess of the Coriolis parameter f, we also make the approximation $M \simeq rV$. The result may be written

$$\frac{d}{dr}\left(r^3 \frac{d}{dr}\left(\frac{V}{r} \right) \right) = \frac{C_D}{\upsilon H} r^2 V^2. \tag{9.102}$$

As this is a second-order ODE, two boundary conditions are required for its solution. To avoid any singularity, we must insists that $V = 0 \quad \text{at} \quad r = 0,$ and at the outer boundary, $V = V_0 \quad \text{at} \quad r = r_0,$ where r_0 is the radius of the vessel and the edge of the vessel has a velocity V_0. If we normalize radius by r_0 and velocity by V_0, then (9.102) simplifies to

$$\frac{d}{dr}\left(r^3 \frac{d}{dr}\left(\frac{V}{r} \right) \right) = \gamma r^2 V^2, \tag{9.103}$$

where

$$\gamma \equiv \frac{C_D r_0^2 V_0}{H\upsilon},$$

subject to $V = 0 \quad \text{at} \quad r = 0,$ and $V = 1 \quad \text{at} \quad r = 1.$ We numerically solve (9.103) by writing it as two, coupled first-order ODEs, marching inward from the outer boundary, and using a “shooting method” to satisfy the boundary condition at the central axis. Nondimensional solutions are shown in Figure 9.21.

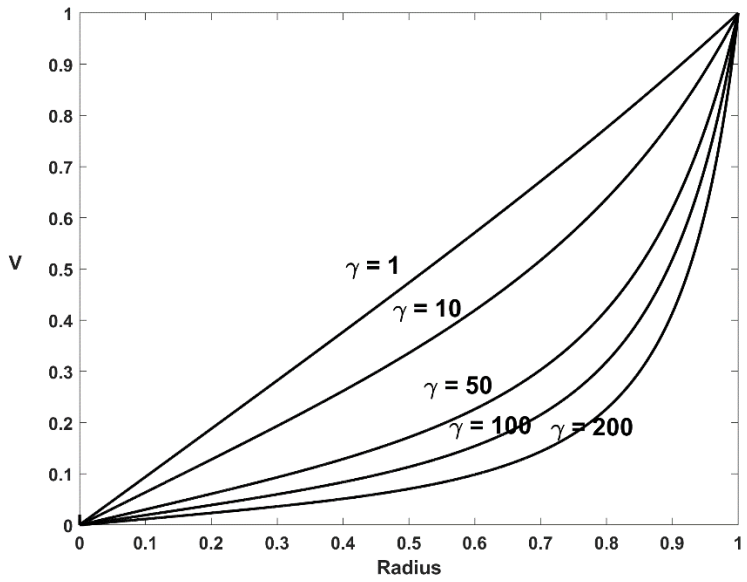


Figure 9.21: Nondimensional solutions of (9.103) for various values of the nondimensional parameter γ.

For rather viscous flow $(\gamma = 1)$, the fluid is nearly in solid body rotation (or constant angular velocity), with the flow speed increasing linearly outward. As the viscosity decreases, the flow profile becomes increasingly concave, with angular velocity and vorticity increasing outward, especially near the edge of the vessel. In the inviscid limit, the velocity and angular velocity are zero in the interior and have step functions at the edge of the vessel, while the vorticity has a delta function spike at the edge.

If the interior flow is turbulent, with vessel-filling eddies with a characteristic velocity V_0, then the turbulent diffusivity, υ, might be represented as the product of the vessel radius and the velocity V_0: $\upsilon \approx r_0 V_0$. In that case, $\gamma \simeq C_D r_0 / H$, independent of the vessel's rotation rate.

Figure 9.20 and the quantitative solutions shown in Figure 9.21 yield a fairly simple picture of the flow in our rotating vessel. This might be viewed as a key step to understanding the flow in real tropical cyclone eyes, but the latter differ from our rotating vessel in three important respects. First, the velocity and angular velocity of the eyewall decrease substantially with altitude, so that the airflow in the eye cannot be independent of altitude. Second, the tropical cyclone eye is density stratified, with dry entropy (or potential temperature) increasing upward. Third, there is radiative cooling in real TC eyes.

This density stratification allows the horizontal flow to vary with height without large vertical accelerations in response to perturbation vertical pressure gradients. To some degree of approximation, the stratification allows quasi-horizontal eddies to re-distribute angular velocity along isentropic surfaces so that at each level, radial advection of angular momentum by the azimuthal mean circulation can balance radial diffusion of angular velocity by turbulence. Vertical gradients of the azimuthal flow in the eye are nearly balanced by radial gradients of entropy. At the same time, radiative cooling must be balanced by a gentle downflow of air above the mechanical boundary layer.

To get some idea of the magnitude of the radial temperature gradients in tropical cyclone eyes, we will have a look again at the thermal wind equation (9.40), assuming that vertical lapse rates are approximately moist adiabatic above the boundary layer in the eye. Multiplying (9.40) through by absolute angular momentum per unit mass, M, we have

$$\left(T - T_o\right)\frac{ds*}{dM} = -\frac{V}{r}, \tag{9.104}$$

where, once again, T_o is the absolute temperature at which the azimuthal velocity vanishes along each M surface. This shows that as one travels upward along M surfaces, and temperature T decreases, the angular velocity of the flow must decrease. Assuming that the angular velocity V/r is approximately constant with radius at constant pressure, integrate (9.104) from the central axis to the radius of maximum wind, along an isobaric surface:

$$\left(\overline{T} - T_o\right)\left(s_c^* - s_e^*\right) = \int_0^{M_e} \frac{V}{r}\, dM, \tag{9.105}$$

where s_c^* and s_e^* are the saturation entropies at the storm center and in the eyewall, at the radius of maximum wind, respectively, $\overline{T}$ is the mean temperature on an isobaric surface and M_e is the angular momentum at the radius of maximum wind and at temperature T. As we are taking V/r to be approximately constant with radius, we can integrate the right side (9.105) directly. Approximating the angular momentum of the eyewall by rV, which is typically much larger than the Coriolis term, we have

$$s_c^* - s_e^* = \frac{V_m^2}{\overline{T} - T_o}, \tag{9.106}$$

where V_m is the maximum wind speed in the eyewall at temperature $\overline{T}$.

Finally, we can express (9.106) in terms of a radial temperature change using (2.49) and (2.75) for buoyancy, and neglecting the small total water content term, r_T:

$$B = \Gamma_d s_d{}' = \Gamma_m s*', \tag{9.107}$$

where the primes denote small fluctuations and, as a reminder, s_d is the dry entropy and Γ_d and Γ_m are the dry and moist adiabatic lapse rates, respectively. Using this in (9.106) yields

$$s_{dc} - s_{de} = \frac{\Gamma_m}{\Gamma_d}\frac{V_m^2}{T - T_o}. \tag{9.108}$$

The ratio of adiabatic lapse rates near the surface in the tropics is roughly 1/3, increasing with altitude an asymptotically approaching unity at low temperature. Near the tropical tropopause, this ratio is essentially unity. Note that as one travels up inside the eye, both V_m and $\overline{T}$ decrease.

If we evaluate (9.108) along an isobaric surface, and using the definition of dry entropy,

$$T_c = T_e e^{\frac{\Gamma_m}{\Gamma_d}\frac{V_m^2}{c_p(\overline{T}-T_o)}} \tag{9.109}$$

Suppose we evaluate this temperature difference in the middle troposphere, where $T_e \simeq 260K$ and $\Gamma_m / \Gamma_d \simeq 0.5$, and we assume an intense tropical cyclone with $V_m = 50\ ms^{-1}$ and $\overline{T} - T_o = 60\,K$ in the middle troposphere. This gives an upper eye temperature perturbation of around 5 K, relative to the eyewall.

How does the core of the eye of tropical cyclones manage to become this warm compared to the eyewall? As there is no release of latent heat there, and radiation only serves to cool the eye, there is only one possibility: adiabatic warming on descent. Therefore, as a tropical cyclone with a well-defined eye intensifies, there must be descent in the eye. According to (9.109), the warming can be thought of as occurring for two reasons: First, the eyewall itself warms up as the boundary layer moist entropy increases underneath it, so that T_e in (9.109) increases. Second, as the azimuthal velocity of the eyewall increases, to maintain thermal wind balance the eye must become progressively warmer than the eyewall, again according to (9.109). Assuming an approximately moist adiabatic lapse rate above the boundary layer, the amount of descent, Δz, associated with a given dry adiabatic temperature change ΔT, relative to an environmental (assumed) moist adiabatic lapse rate, is given by

$$\Delta z = \frac{-\Delta T}{\Gamma_d - \Gamma_m}. \tag{9.110}$$

For a moist adiabatic lapse rate that is half the dry adiabatic value and a temperature change of 5 K, this yields a 1 km downward displacement. Such large downward displacements also help explain why the air in tropical cyclone eyes above the boundary layer is so dry and free of cloud.

Another source of downward velocity in the eye, and the only one that is active in the steady state, is radiative cooling[6]. Typical cooling rates are of order $1K\,day^{-1}$. Therefore, a rapidly intensifying cyclone will be associated with much stronger eye descent than is caused by radiative cooling in the steady state. For example, the U.S. National Hurricane Center defines rapid intensification as maximum winds increasing by at least $17\,ms^{-1}$ in a 24-hour period. Assuming, for the moment, that the eye "keeps up" with the eyewall, maintaining solid-body rotation, equation (9.109) predicts that, in an intensification going from 30 to 47 ms^{-1}, the eye would warm by about 3 K (also accounting for the warming of the eyewall), with an implied descent rate of order $1\,cm\,s^{-1}$, about an order of magnitude greater than the radiative subsidence velocity.

Now refer back to Figure 8.34. Especially during rapid intensification, there must be descent in the eye above the boundary layer. But as long as the airflow at the surface is cyclonic, there must be radial inflow in the boundary layer, and upward motion near the center of the eye. The

[6] Close to the edge of the eye, cooling by evaporation of hydrometeors may also cause downward motion.

strong convergence (in the vertical) between the upward Ekman flow and the downward flow above the boundary layer typically creates a strong temperature inversion where the two streams meet. Below the inversion, the upward flowing air is usually at least partly cloudy. By mass continuity, there must be strong outward flow near the level of the inversion, and this would tend to make the radial profile of azimuthal velocity more concave there.

Because the Ekman convergence scales with the low-level azimuthal velocity, but the descent rate above the boundary layer scales with the rate of intensification (plus a smaller, more or less steady contribution from radiative cooling), the height of the eye inversion should reflect these different scalings, with low inversions during rapid intensification and higher ones in the steady and weakening phases.

Much of the above discussion should be tempered by observations that suggest that the time scale for the eye circulation to respond to changes in the eyewall is not that small. Especially after a period of rapid intensification, the azimuthal circulation in the eye can lag well behind that of the eyewall, with very large radial shear just inside the eyewall. The extreme concentration of the vertical and tangential components of vorticity are highly unstable and can lead to eyewall mesovortices, as discussed further in Chapter 11. The effect of these disturbances is to transfer angular velocity from the eyewall towards the center of the eye.

Figure 9.22 shows two distinct radial profiles of azimuthal velocity and vertical vorticity at 850 hPa in the same storm, Atlantic Hurricane Diana of 1984, separated by about 12 hours. (The first profile is based on 22 radial flight legs taken between 11 UTC September 11th and 00 UTC September 12th, while the second is based on 18 radial legs between 00 UTC and 12 UTC on September 12th.

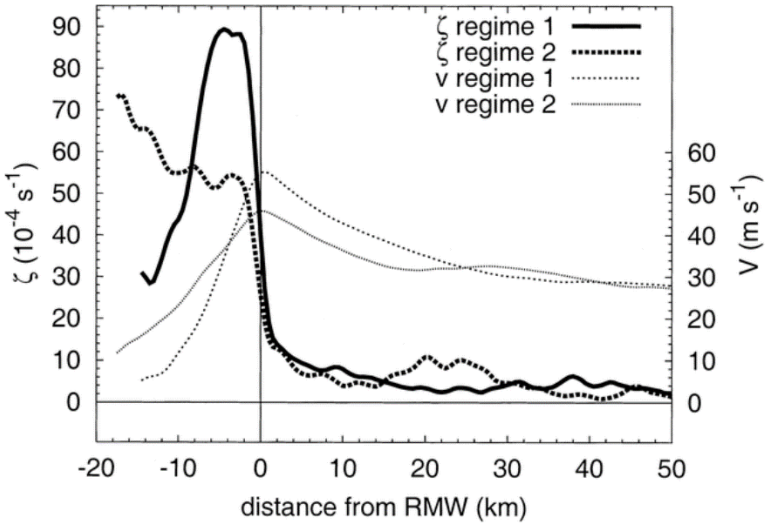


Figure 9.22: Vorticity (ζ) and tangential wind (V) profiles, averaged with respect to distance from the radius of maximum winds (RMW; km), within regime 1 and regime 2 at 850 mb in Atlantic Hurricane Diana (1984). Average profiles indicated as regime 1 are based on 22 radial legs flown during 1100 UTC 11 Sep to 0000 UTC 12 Sep. Average profiles indicated as regime 2 are based on 18 radial legs flown during 0000–1200 UTC 12 Sep.

Diana intensified rapidly between 12 GMT on September 10th and 00 GMT on September 12th, from a category 1 to a category 4 hurricane. Thereafter, it weakened. Regime 1 in Figure 9.22 was made from observations taken toward the end of the period of rapid intensification, while

Regime 2 was based on observations during the weakening phase. During rapid intensification, the velocity profile is concave and vorticity is highly concentrated just inside the eyewall. This is strongly unstable, as we will explore in Chapter 11. Only a few hours later, the high angular velocity near the eyewall has diffused into the center of the eye, and the vorticity is nearly constant across the eye. This shows that the diffusion of angular velocity is fast, but not fast enough to keep pace with rapid intensification.

In summary, the eye is mostly or entirely a passive, but nevertheless interesting response to the evolving eyewall circulation and to surface friction. The latter creates a turbulent Ekman layer with inflow and upward motion near the cyclone center, while the former leads to downflow in proportion to the rate of intensification of the azimuthal flow in the eyewall and the degree of radiative cooling in the eye. The two regimes are usually separated by a pronounced temperature inversion whose height reflects the relative magnitudes of the two processes.

9.5 The tropical cyclone boundary layer

The boundary layer is a critical component of tropical cyclones. Most of the important turbulent fluxes that drive and retard the storm are driven by shear and convection within boundary layer, and the ocean's response in the form of currents, waves, and storm surges are driven by turbulent stresses imposed from the boundary layer. Moreover, virtually all the wind damage done by these storms occurs within the boundary layer – it is where tropical cyclones interact with human beings and their infrastructure.

While the thermodynamic constraints developed earlier in this chapter place limits on the time-mean, axisymmetrically averaged surface wind, they do not constrain the magnitude of disturbances that are unsteady and/or asymmetric. Indeed, both observations and large-eddy simulations show important fluctuations in wind speed down to periods of just a few seconds as well as mesoscale organization in the boundary layer; these can and do have strong effects on the damage caused by hurricanes.

In the far outer region of tropical cyclones, the boundary layer physics are essentially the same as those operating normally in the tropical atmosphere, as described briefly in Chapter 4. But in the inner region, several factors combine to alter the physics in ways that have yet to be completely understood. First, the mean flow is strongly rotational, and this may alter the properties of the turbulence and affect the depth of the boundary layer. Second, the strong frontogenesis at the eyewall, to be discussed in Chapter 10, creates very strong radial gradients in the flow and thermodynamic variables, leading to strong radial turbulent fluxes. For this reason, the classical treatment of the boundary layer as statistically horizontally homogeneous breaks down. Finally, in strong storms, the air-sea interface itself breaks down into something more resembling an emulsion, where bubble-filled water gives way to spray-filled air, so that the classical formulations of interfacial fluxes may no longer apply. (We defer a detailed treatment of the air-sea interface to Chapter 13.) All of these factors make the eyewall region of a strong tropical cyclone a *terra incognita* (or, perhaps, *aqua incognita*) in great need of further research.

We begin with the simplest formulations of tropical cyclone boundary layers and progress to more comprehensive treatments.

9.5.1 Boundary layer equations

Thus far we have treated the tropical cyclone boundary layer as a slab whose properties do not vary with height or time and for which the azimuthal wind is specified (see equations (9.78)-(9.79) and Figure 9.11.) In reality, the azimuthal wind in the boundary layer can be substantially smaller or larger than the gradient wind, and in proper boundary layers, the gradient wind (i.e. the horizontal pressure gradient) is specified and the momentum equations are solved for the time-mean, azimuthally invariant horizontal wind components.

Thus, the external forcing for the boundary layer equations is the radial pressure gradient, and this is assumed steady and invariant with altitude within the turbulent boundary layer. Strong virtual temperature gradients and/or accelerations in the boundary layer can compromise this assumption.

First, it is convenient to divide the boundary layer variables into the time-mean flow and time-dependent fluctuations. For the primary variables $u, v, w,$ and $p,$ we can write

$$x = \overline{x}(r,z) + x'(r,z,t), \tag{9.111}$$

where x stands for any of the four dependent variables, the overbar represents the time-mean flow, and the prime represents fluctuations away from the time mean. The independent variables r, z and t represent radius, altitude, and time, respectively.

We begin with the Boussinesq form of the Navier-Stokes equation for the radial flow, u, phrased in cylindrical coordinates:

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial r} + w\frac{\partial u}{\partial z} = -\frac{1}{\rho_0}\frac{\partial p}{\partial r} + \left(f + \frac{v}{r}\right)v + \upsilon\nabla^2 u, \tag{9.112}$$

where p is pressure, ρ_0 is a mean, constant air density, f is the Coriolis parameter, which we assume constant here, and υ is the kinematic viscosity. We also write the Boussinesq form of the mass continuity equation as

$$\frac{\partial (ru)}{\partial r} + \frac{\partial (rw)}{\partial z} = 0. \tag{9.113}$$

Using (9.113) we may re-write (9.112) as

$$\frac{\partial u}{\partial t} + \frac{1}{r}\frac{\partial ru^2}{\partial r} + \frac{\partial uw}{\partial z} = -\frac{1}{\rho_0}\frac{\partial p}{\partial r} + \left(f + \frac{v}{r}\right)v + \upsilon\nabla^2 u. \tag{9.114}$$

Similarly, the momentum equation for the azimuthal component may be written

$$\frac{\partial v}{\partial t} + \frac{1}{r}\frac{\partial ruv}{\partial r} + \frac{\partial vw}{\partial z} = -\left(f + \frac{v}{r}\right)u + \upsilon\nabla^2 v. \tag{9.115}$$

To distinguish between the time -mean flow and the turbulent fluctuations within it, we apply the de-composition (9.111) to dependent variables in both equations, and then average the result over time, resulting in

$$\frac{1}{r}\frac{\partial r\overline{u}\,\overline{u}}{\partial r}+\frac{\partial \overline{u}\,\overline{w}}{\partial z}=-\frac{1}{\rho_0}\frac{\partial \overline{p}}{\partial r}+\left(f+\frac{\overline{v}}{r}\right)\overline{v}+\frac{\overline{v'^2}}{r}-\frac{1}{r}\frac{\partial r\overline{u'^2}}{\partial r}-\frac{\partial \overline{u'w'}}{\partial z}+\upsilon\nabla^2\overline{u}, \tag{9.116}$$

and

$$\frac{1}{r}\frac{\partial r\overline{u}\,\overline{v}}{\partial r}+\frac{\partial \overline{w}\overline{v}}{\partial z}=-\left(f+\frac{\overline{v}}{r}\right)\overline{u}-\frac{\overline{v'u'}}{r}-\frac{1}{r}\frac{\partial r\overline{u'v'}}{\partial r}-\frac{\partial \overline{v'w'}}{\partial z}+\upsilon\nabla^2\overline{v}. \tag{9.117}$$

The time derivatives vanish because we assume that the time-mean flow is steady, and all terms that are linear in the fluctuations vanish as well when averaged. When we average the mass continuity equation (9.113) in time we simply get

$$\frac{\partial\left(r\overline{u}\right)}{\partial r}+\frac{\partial\left(r\overline{w}\right)}{\partial z}=0. \tag{9.118}$$

The terms involving products of perturbations, which have been moved to the right sides of (9.116) and (9.117), represent the effects on the mean flow of the convergences of turbulent fluxes of momentum. Were it not for these terms, (9.116)-(9.118) would constitute a closed system for the three time-mean velocity components, provided the radial pressure gradient is specified.

In traditional boundary layer theory, the flow is assumed to be statistically horizontally homogeneous, and the turbulence not too far from being isotropic, so that the turbulent eddies have vertical and horizontal dimensions that are comparable. That being the case, the convergences of the vertical turbulent fluxes dominate the convergences of the horizontal fluxes. Also, the flow is assumed to be fully turbulent with very high values of the Reynolds Number, so that this viscous terms at ends of (9.116) and (9.117) are neglected. With these approximations, (9.116) and (9.117) become

$$\frac{1}{r}\frac{\partial r\overline{u}\,\overline{u}}{\partial r}+\frac{\partial \overline{u}\,\overline{w}}{\partial z}\simeq-\frac{1}{\rho_0}\frac{\partial \overline{p}}{\partial r}+\left(f+\frac{\overline{v}}{r}\right)\overline{v}-\frac{\partial \overline{u'w'}}{\partial z}, \tag{9.119}$$

$$\frac{1}{r}\frac{\partial r\overline{u}\,\overline{v}}{\partial r}+\frac{\partial \overline{w}\overline{v}}{\partial z}\simeq-\left(f+\frac{\overline{v}}{r}\right)\overline{u}-\frac{\partial \overline{v'w'}}{\partial z}. \tag{9.120}$$

Above the boundary layer, the turbulent fluxes are considered to be small, as is the mean radial velocity. Then, according to (9.119), we have gradient wind balance:

$$\frac{1}{\rho_0}\frac{\partial \overline{p}}{\partial r}=\left(f+\frac{\overline{v}}{r}\right)\overline{v}\equiv\left(f+\frac{v_g}{r}\right)v_g, \tag{9.121}$$

where v_g is defined to be the gradient wind and is a useful proxy for the horizontal pressure gradient. Can we regard this time-mean pressure gradient as being independent of altitude within the boundary layer?

To estimate the magnitude of the vertical variation of the radial pressure gradient within the boundary layer, we differentiate (9.121) with respect to altitude and use the hydrostatic equation to relate vertical pressure gradients to density, and the ideal gas law to relate density to virtual temperature, yielding

$$\left(f + \frac{2v_g}{r} \right) \frac{\partial v_g}{\partial z} \approx g \frac{\partial \ln T_v}{\partial r}. \tag{9.122}$$

If we take and extreme value of 5 Kelvins of virtual temperature increase over 100 km and a mean gradient wind of 50 ms^{-1}, this gives a change in gradient wind of roughly 2 ms^{-1} over 1 km of altitude. Given the other approximations we are making, assuming the gradient wind to be constant over the depth of the boundary layer is a minor approximation. Therefore, we shall take

$$\frac{1}{r} \frac{\partial r \overline{u}\, \overline{u}}{\partial r} + \frac{\partial \overline{u}\, \overline{w}}{\partial z} \simeq -\left(f + \frac{v_g}{r} \right) v_g + \left(f + \frac{\overline{v}}{r} \right) \overline{v} - \frac{\partial \overline{u'w'}}{\partial z}. \tag{9.123}$$

9.5.2 The slab boundary layer

An additional simplification is to average the momentum equations, (9.120) and (9.123), and the mass continuity equation, (9.118) over a prescribed or calculated depth of the boundary layer. Examples of slab boundary layer models include those of Smith (2003), Smith and Vogel (2008), and Smith and Montgomery (2008). Here we largely follow Kepert (2010), who also provides a nice review of tropical cyclone boundary layer physics and models.

What determines the depth of the boundary layer? There are several different definitions of boundary layer depth, including the altitude at which the turbulent fluxes vanish or become sufficiently small, the altitude at which the mean dry static stability jumps upward from near zero (or negative) inside the boundary layer, or perhaps the average altitude of cloud base. Sometimes these variously defined boundary layer depths coincide. For the time being, we will allow the boundary layer depth, h, to be a specified function of radius, and defined as the altitude at which the turbulent fluxes become quite small relative to their values in the interior of the boundary layer.

In performing the vertical integration of the momentum equations (9.123) and (9.120), there are quadratic terms in the velocities. Given that these velocities do, in reality, vary with height in the boundary layer, the vertical mean of their product is not equal to the product of means of the two variables. But given that there is no obvious way to represent these, we are forced to approximate these mean quadratic values as the product of the means of the two variables.

With this approximation, integrating (9.123), (9.120), and (9.118) through the depth of the boundary layer results in

$$\frac{1}{r} \frac{\partial r \overline{u}\, \overline{u}}{\partial r} + \frac{\overline{u}\, \overline{w}|_h}{h} - \frac{\overline{u'w'|_0}}{h} - \left(f + \frac{\overline{v}}{r} \right) \overline{v} = -\left(f + \frac{v_g}{r} \right) v_g, \tag{9.124}$$

where $\overline{u}\,\overline{w}|_h$ is the product of the time-mean radial and vertical velocities evaluated at the top of the boundary layer, and $\overline{u'w'}|_0$ is the surface turbulent vertical flux of radial velocity.

Consider first the term $\overline{u}\,\overline{w}|_h$. Where air is exiting the boundary layer $(w>0)$, we would not expect that to affect velocities inside the boundary layer, but where air is entering the boundary layer, we would expect it to import the radial velocity from just above the boundary layer. Here we approximate this latter quantity as zero. Therefore, we estimate $\overline{u}\,\overline{w}|_h$ as $\overline{u}\min\left(\overline{w}|_h,0\right)$.

We represent the surface flux of radial momentum, $\overline{u'w'}|_0$, by the conventional neutral aerodynamic flux formula (see section 4.3) as

$$\overline{u'w'}|_0 \simeq -C_D\sqrt{\overline{u}^2+\overline{v}^2}\,\overline{u}. \tag{9.125}$$

With these estimates, (9.124) becomes

$$\frac{1}{r}\frac{\partial r\overline{u}\,\overline{u}}{\partial r}+\frac{\overline{u}\min\left(\overline{w}|_h,0\right)}{h}+\frac{C_D\sqrt{\overline{u}^2+\overline{v}^2}\,\overline{u}}{h}-\left(f+\frac{\overline{v}}{r}\right)\overline{v}=-\left(f+\frac{v_g}{r}\right)v_g. \tag{9.126}$$

Using the same representations of surface fluxes and fluxes through the top of the boundary layer, (9.120) becomes

$$\frac{1}{r}\frac{\partial r\overline{u}\,\overline{v}}{\partial r}+\frac{\left(\overline{v}-v_g\right)\min\left(w\,|\,h,0\right)\right)}{h}+\frac{C_D\sqrt{\overline{u}^2+\overline{v}^2}\,\overline{v}}{h}+\left(f+\frac{\overline{v}}{r}\right)\overline{u}=0. \tag{9.127}$$

Finally, integrating (9.118) through the depth of the boundary layer yields

$$\overline{w}|_h=-\frac{h}{r}\frac{\partial}{\partial r}\left(r\overline{u}\right). \tag{9.128}$$

With specification of the radial profiles of gradient wind and boundary layer depth h, as well as the drag coefficient C_D and Coriolis parameter, f, (9.126) - (9.128) constitute a closed system. An easy way to solve it is to re-insert the time derivative terms into the momentum equations (9.126) and (9.127) and march forward in time until a steady state is achieved.

In many slab boundary layer models (e.g. see Kepert, 2010), the boundary layer depth is held constant. But basic theory (e.g. Rosenthal 1962; Eliassen and Lystad 1977), observations, and the solution of height-dependent models show that the depth of the boundary layer decreases inward in the inner region. The theory of turbulent Ekman layers suggests that the scales as the ratio of the friction velocity and a measure, I, of the inertial stability of the rotating flow. In the present notation,

$$h\approx\sqrt{\frac{C_D\left(\overline{u}^2+\overline{v}^2\right)}{I}}, \tag{9.129}$$

where

$$I \equiv \left(f + 2\frac{\overline{v}}{r} \right) \left(f + \frac{1}{r}\frac{\partial}{\partial r}\left(r\overline{v} \right) \right). \tag{9.130}$$

In the core of the vortex, where v is large and r is small, (9.129) predicts that the boundary layer depth should increase nearly linear with radius.

For simplicity, we hold the drag coefficient C_D constant in this model. In reality, C_D varies with wind speed and the surface wave state, as shall be discussed in Chapter 13.

Solutions of the system (9.126) – (9.130) are shown in Figure 9.23 for a particular choice of the input parameters. In this case, we have used the matched inner-outer wind profile discussed in section 9.3.2 as the gradient wind driver of the slab PBL model.

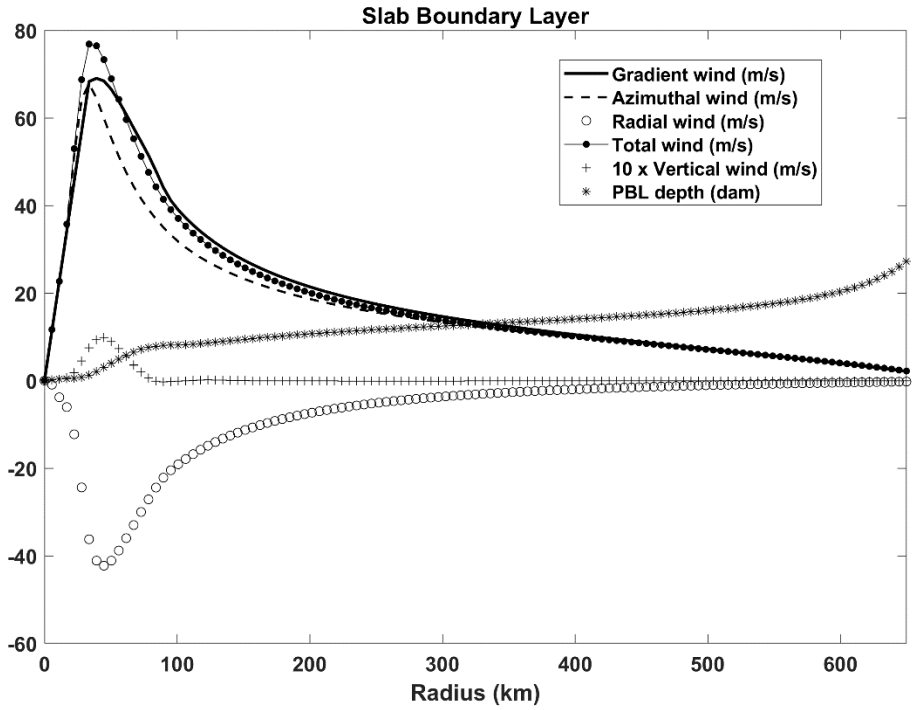


Figure 9.23: Solutions to the slab boundary layer model for $C_D = 1.5\times10^{-3}$, $f = 5\times10^{-5}\,s^{-1}$, $v_{gmax} = 70\,ms^{-1}$, $r_o = 700\,km$, $w_s = 10\,mm\,s^{-1}$, and $h = 1500\,m$, *where* v_{gmax} *is the peak gradient wind,* r_o *is the outer radius, and* w_s *is the radiative subsidence velocity.*

Note that the azimuthal wind component is substantially less than the gradient wind outside the radius of maximum winds, and has a peak value slightly less than the peak gradient wind. (For other values of the parameters, the peak azimuthal component can exceed the peak gradient wind.) The radial wind reaches its peak at a larger radius than does the gradient and azimuthal winds, and has a peak magnitude of more than half the peak gradient wind, so that the inflow angle at that radius is quite substantial, approaching 35°. The vertical velocity peaks at about the same radius as the radial velocity, somewhat outside the radius of maximum horizontal wind. The boundary layer depth is very shallow in the eye and increases rapidly across the

radius of maximum winds, reflecting the dramatic decrease of inertial stability from the eye to the outer core.

The peak total horizontal boundary layer velocity exceeds the peak gradient wind by about 12%, and this can be greater for different choices of the control parameters; this peak occurs inside the radius of maximum gradient wind. This is fundamentally owing to the inertia of the radial flow: air flowing rapidly inward "undershoots" the radius at which its angular momentum equals that of the gradient wind, carrying the higher angular momentum with it.

The reader is invited to explore the dependence of the slab boundary layer solutions on the shape of the gradient wind profile, the drag coefficient, the Coriolis parameter, and other parameters by running the script *PBLSlabModel.m,* available at https://zenodo.org/doi/10.5281/zenodo.10940645 .

Note that, according to (9.50), the surface wind speed must always be less that the peak gradient wind speed. Yet our slab model predicts a maximum wind speed that is greater than the maximum gradient wind speed. Since the slab wind speed represents a vertical average, this implies that the actual maximum wind speed in the boundary layer occurs away from the surface layer, within which the aerodynamic drag law is used to calculate drag. We shall show in the following section that this is indeed the case in vertically resolved boundary layers.

9.5.3. Vertically resolved tropical cyclone boundary layers

We return to our time-mean but vertically resolved equations, (9,123), (9.119) and (9.118). With a specified gradient wind, v_g, boundary layer depth, h, and boundary conditions at the surface and top of the boundary layer, this would be a closed system were it not for the turbulent flux terms (last terms on the right sides of 9.123 and 9.119). This presents the so-called "closure problem" – how to represent the turbulent fluxes in terms of the resolved variables, which in this case are the time-mean flow components.

In general, one expects fluxes to be down the gradient of the ensemble mean variable in question. This is exactly how molecular diffusion works. Thus we might represent the turbulent fluxes in (9.123) and (9.119) as

$$\overline{u'w'} = -K\frac{\partial \overline{u}}{\partial z}, \qquad \overline{v'w'} = -K\frac{\partial \overline{v}}{\partial z}, \tag{9.131}$$

where K is usually referred to as an eddy viscosity, as it has the units of molecular kinematic viscosity and would in fact be the molecular viscosity were the flow laminar. It has the dimensions of a length scale multiplied by a velocity scale, and can be loosely thought of as the product of typical eddy length and velocity scales. In nearly neutral boundary layers in which the turbulence is mostly mechanically driven, a logical candidate for an eddy velocity scale is the friction velocity, u_* (see Chapter 4), which is defined as the square root of the surface stress divided by air density and is independent of height. Close to the surface, the scale of eddies will be confined by their depth, whereas higher up in the boundary layer the eddies scale more nearly with the depth of the boundary layer, h. Kepert (2010) used the following formulation of the eddy missing length l:

$$l = \left[\left(kz \right)^{-1} + l_{\infty}^{-1} \right]^{-1}, \tag{9.132}$$

where k is the von Kármán constant (see Chapter 4) and l_{∞} is an asymptotic scale which Kepert (2010) took to be $60\,m$.

Therefore, the formulation used here for the eddy viscosity is

$$K = l C_D^{1/2} \mid \mathbf{V}_s \mid, \tag{9.133}$$

where $\mid \mathbf{V}_s \mid$ is the wind speed at the lowest model level. Note that Kepert used a wind-dependent drag coefficient.

The numerical solution of the equations (9.123), (9.119) and (9.118) requires a finite domain, and there is an issue of where the top of the domain should be located. In general, it should be above the region of maximum turbulence energy. If it is too low there will be significant turbulent fluxes at the boundary layer top and one may not claim that the model is simulating the whole of the boundary layer. If it is too high, then the model-top gradient wind, which does vary with height, may no longer represent that of the boundary layer. Kepert (2010) chose the domain top to be a little above 2 km altitude, with 15 computational layers from the surface to the top. He applied Neumann boundary conditions: all three velocity components are assumed to have zero vertical gradient at the model's upper boundary.

Figure 9.24 shows a solution of Kepert's height-resolved model for a gradient wind profile that peaks at $40\,ms^{-1}$. There are quite a few interesting aspects of these solutions.

The two horizontal wind components vary strongly with altitude in the boundary layer, as does the turbulent diffusivity, governed by (9.132). As in the slab model, the radial flow peaks outside the radius of maximum azimuthal winds, and in altitude it peaks remarkably close to the surface. The peak azimuthal wind is in the interior of the boundary layer, as predicted from the combination of (9.50) and the solutions of the slab model, and although not shown in the figure, this wind is approximately 6% stronger than the gradient wind.

As predicted by (9.129), the boundary layer depth increases with radius, with the largest slope in the inner region. The vertical velocity peaks at the top of the boundary layer, but decays only slowly into the free troposphere.

Unsurprisingly, the interior structure of the boundary layer is sensitive to the representation of turbulence, and there are many different representations in play as of this writing. Smith and Thomsen (2010) compare the results of running a full-physics, three-dimensional model using 5 different representations of boundary layer turbulence, documenting large differences in the internal structure of the boundary layer and in the overall intensity and intensification rates of modeled storms.

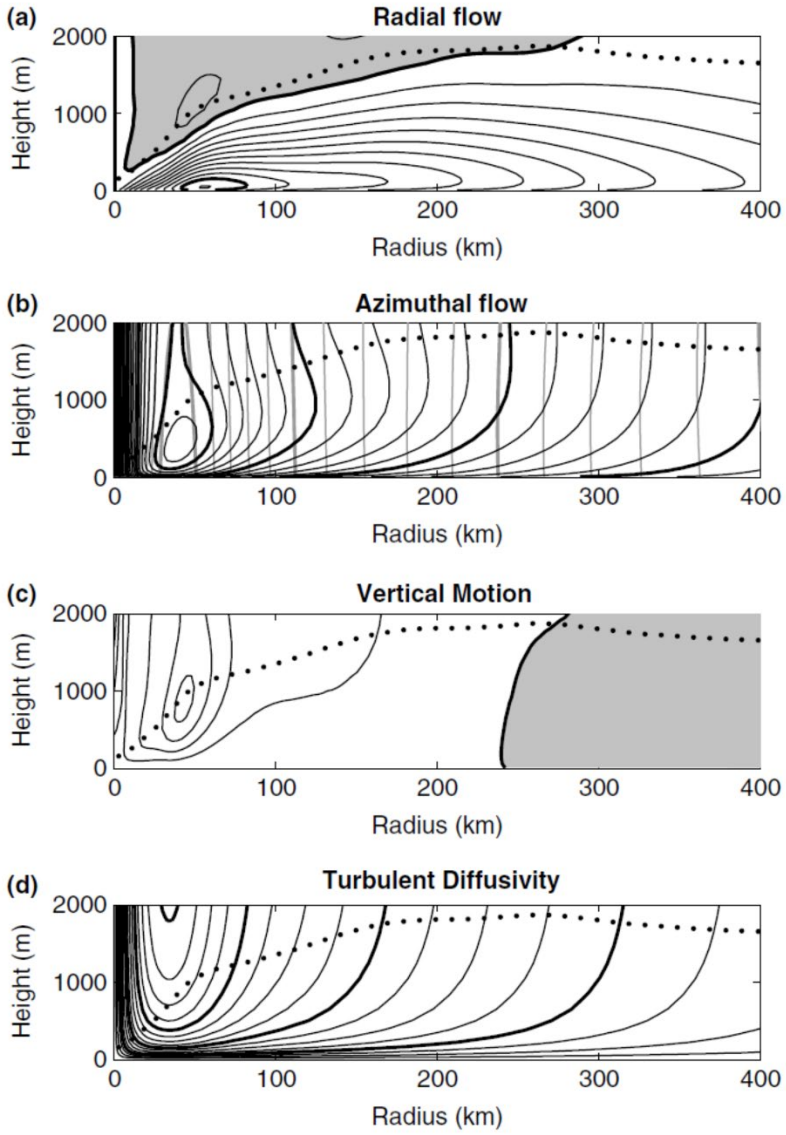


Figure 9.24: The boundary-layer flow in a stationary storm simulated by Kepert's height-resolving model. (a) Radius–height section of the radial wind, contour interval 1 ms^{-1}, multiples of 10ms^{-1} shown bold, positive values shaded. (b) Radius–height section of the azimuthal wind (black contours) and gradient wind (grey contours). Contour intervals 2ms−1, multiples of 10 ms−1 shown bold. (c) Radius–height section of the vertical velocity, contour interval 0.05 ms^{-1}, zero-line shown bold, subsidence shaded. (d) Radius–height section of the turbulent diffusivity, contour interval 5 ms^{-2}, multiples of 20 ms^{-2} shown bold. The dotted line in each panel indicates the level at which the stress magnitude reduces to 20% of its surface value.

Among other things, the turbulence scheme affects the relationship between the low-level wind speed used by the model to calculate surface fluxes, and the gradient wind, whose steady-state magnitude depends on the ratio of the drag and enthalpy flux coefficients. Thus, the representation of turbulence can affect the overall structure and intensity of the tropical cyclones.

We are presently at an exciting and productive juncture in understanding boundary layer turbulence and how best to represent it in models, thanks to the advent of large-eddy-resolving (LES) boundary layer models which, as their name implies, attempt to resolve three-dimensional turbulence down to the outer limits of the inertial subrange. (Below such limits, the turbulence is more nearly isotropic and homogeneous and is generally thought to be relatively well represented by simple turbulence schemes.) Such models resolve important structures in the boundary layer, including horizontal roll structures and eyewall mesovortices. We shall return to a discussion of such structures in Chapter 11, but here we briefly discuss the azimuthal-mean boundary layer characteristics in LES simulations.

A particular example of the behavior and structure of a tropical cyclone simulated by a full-physics, three-dimensional model is provided by Liu et. al (2021), who simulated western North Pacific Typhoon Matsa of 2005 using a model with a large set of two-way nested grids of progressively finer resolution, with horizontal grid spacings of the innermost grid of 333m, 111m, or 37m. The variable vertical grid spacing is roughly 70–100 m below 1 km and increases to about 400 m at 4 km altitude. (Note that one would prefer a vertical grid spacing more comparable to the horizontal spacing at these high resolutions.) In addition, the model was run at 333 m spacing but with a particular parameterization of turbulent fluxes in the boundary layer.

Figure 9.25 shows azimuthal average quantities at the three resolutions, together with the simulation using parameterized turbulence. In all three of the LES simulations, the peak azimuthal flow is somewhat stronger than in the case of parameterized turbulence, and the azimuthal mean flow does not change very much going from *111 m* to *33 m* resolution, suggesting that about 100 m grid spacing suffices for a robust simulation of the azimuthal mean flow. In all cases, but particularly the LES cases, the maximum radial inflow occurs remarkably close to the surface. Otherwise, the basic structure is quite similar to that based on a very simple parameterization of turbulence as shown in Figure 9.24, with the boundary layer depth increasing rapidly moving outward across the zone of peak vertical vorticity, the radial velocity peaking lower than and radially outward from the peak azimuthal wind, and the vertical velocity peaking well above the azimuthal velocity peak but at roughly the same radius as the peak surface winds.

The advent of LES simulations, made possible by rapidly increasing computer power, opens many avenues of research into boundary layer physics, including important transient wind events and structures, discussed further in Chapter 11. Such simulations also provide data against which to test representations of turbulence used in more coarsely resolved models.

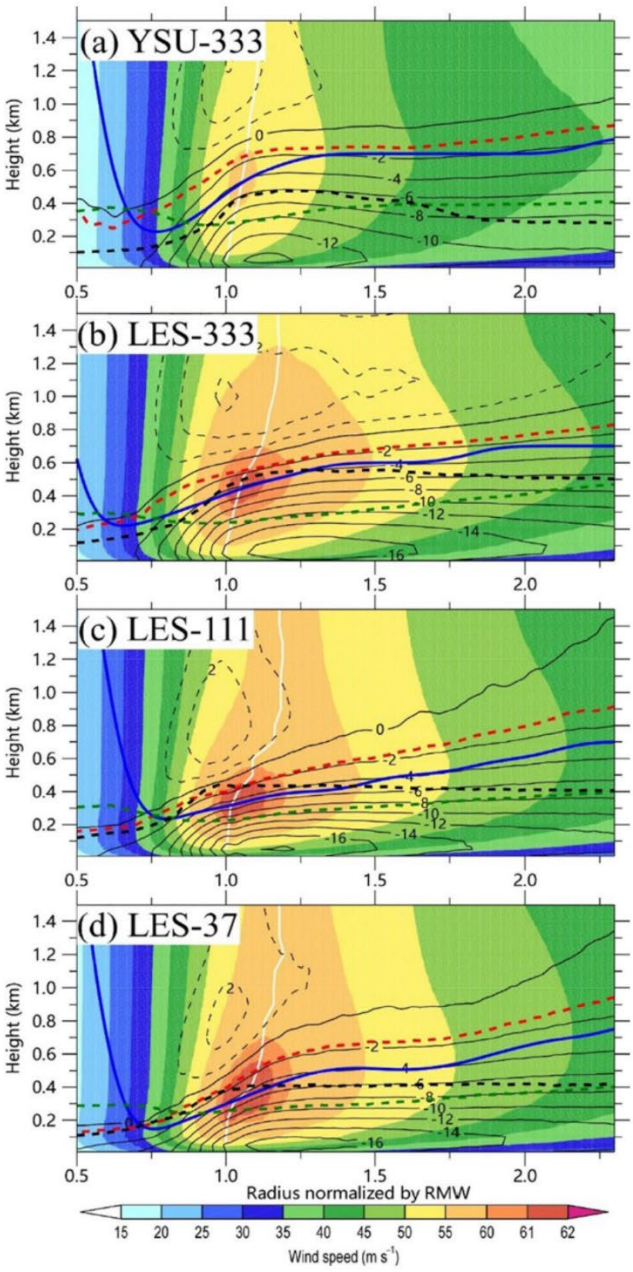


Figure 9.25: Radial-height cross-sections of azimuthal-mean tangential winds (shading, m s^{-1}) and radial winds (contour, m s^{-1}) below 1.5 km height averaged over 26–36 hr for the case with parameterized turbulence (YSU-333) (a), LES-333 (b), LES-111 (c), and LES-37 (d). The radius has been normalized by the radius of maximum azimuthal wind speed (RMW) at the surface. The contour interval is 2 m s^{-1}. The white solid line in all panels depicts the radius of maximum winds. The blue dashed line depicts the height of the maximum tangential wind speed, and the dashed red line depicts the height where the radial wind speed is 10% of the peak inflow. The dashed black line represents the height where the bulk Richardson number is equal to 0.25 and the dashed green line denotes the contour $\partial\theta_v / \partial z = 3\ K\ km^{-1}$. *The vertical and horizontal axes indicate the altitude (km) and the distance from the tropical cyclone center that is normalized by the radius of maximum winds at 10 m height, respectively.*

References

Chavas, D. R., and K. A. Emanuel, 2014: Equilibrium tropical cyclone size in an idealized state of axisymmetric radiative–convective equilibrium. *J. Atmos. Sci.*, **71**, 1663–1680.

Chavas, D. R., N. Lin, and K. Emanuel, 2015: A model for the complete radial structure of the tropical cyclone wind field. part i: comparison with observed structure. *Journal of the Atmospheric Sciences*, **72**, 3647–3662, https://doi.org/10.1175/jas-d-15-0014.1.

Cronin, T. W., 2023: An analytic model for tropical cyclone outer winds. *Geophysical Research Letters*, **50**, e2023GL103942, https://doi.org/10.1029/2023GL103942.

——, and D. R. Chavas, 2019: Dry and Semidry Tropical Cyclones. *Journal of the Atmospheric Sciences*, **76**, 2193–2212, https://doi.org/10.1175/jas-d-18-0357.1.

Eady, E. T., 1949: Long waves and cyclone waves. *Tellus*, **1**, 33–52.

Edwards, J. M., 2019: Sensible heat fluxes in the nearly neutral boundary layer: The impact of frictional heating within the surface layer. *Journal of the Atmospheric Sciences*, **76**, 1039–1053, https://doi.org/10.1175/JAS-D-18-0158.1.

Eliassen, A., and M. Lystad, 1977: The Ekman layer of a circular vortex. A numerical and theoretical study. *Geophysica Norvegica*, **31**, 1–16.

Emanuel, K., 2004: Tropical cyclone energetics and structure. *Atmospheric Turbulence and Mesoscale Meteorology*, R.R. and B.S. E. Federovich, Ed., Cambridge University Press, p. 240.

——, and R. Rotunno, 2011: Self-stratification of tropical cyclone outflow. Part I: Implications for storm structure. *J. Atmos. Sci.*, **68**, 2236–2249.

Emanuel, K. A., 1988: Observational evidence of slantwise convective adjustment. *Mon. Wea. Rev.*, **116**, 1805-1816.

——, K. Speer, R. Rotunno, R. Srivastava, and M. Molina, 1995: Hypercanes: A possible link in global extinction scenarios. *J. Geophys. Res.*, **100**, 13755–13765.

Kepert, J. D., 2010: Slab- and height-resolving models of the tropical cyclone boundary layer. Part I: Comparing the simulations. *Quart. J. Roy. Meteor. Soc.*, **134**, 1686–1699.

Knaff, J. A., J. P. Kossin, and M. DeMaria, 2003: Annular hurricanes. *Weather and Forecasting*, **18**, 204–223, https://doi.org/10.1175/1520-0434(2003)018<0204:AH>2.0.CO;2.

Liu, Q., L. Wu, N. Qin, and Y. Li, 2021: Storm-scale and fine-scale boundary layer structures of tropical cyclones simulated with the WRF-LES framework. *Journal of Geophysical Research: Atmospheres*, **126**, e2021JD035511, https://doi.org/10.1029/2021JD035511.

Mrowiec, A. A., S. T. Garner, and O. M. Pauluis, 2011: Axisymmetric hurricane in a dry atmosphere: Theoretical framework and numerical experiments. *J. Atmos. Sci.*, **68**, 1607–1619.

Nicolis, G., and I. Prigogine, 1977: *Self-organization in nonequilibrium systems*. John Wiley & Sons, 490 pp.

Prigogine, I., and R. Lefever, 1968: Symmetry Breaking Instabilities in Dissipative Systems. II. *The Journal of Chemical Physics*, **48**, 1695–1700, https://doi.org/10.1063/1.1668896.

Rosenthal, S. L., 1962: A theoretical analysis of the field of motion in the hurricane boundary layer. *Natioanl Hurricane Center Research Project*, Vol. 56 of, U. S. Department of Commerce.

Rotunno, R., and K. A. Emanuel, 1987: An air-sea interaction theory for tropical cyclones. Part II. *J. Atmos. Sci.*, **44**, 542–561.

Smith, R. K., and M. T. D.-:10. 1002/qj. 296 Montgomery, 2008: Balanced boundary layers used in hurricane models. *Quarterly Journal of the Royal Meteorological Society*, **134**, 1385–1395.

——, and S. D.-:10. 1002/qj. 216 Vogl, 2008: A simple model of the hurricane boundary layer revisited. *Quarterly Journal of the Royal Meteorological Society*, **134**, 337–351.

——, and G. L. D.-:10. 1002/qj. 687 Thomsen, 2010: Dependence of tropical-cyclone intensification on the boundary-layer representation in a numerical model. *Quarterly Journal of the Royal Meteorological Society*, **136**, 1671–1685.

Smith, R. K. D.-:10. 1256/qj. 01. 197, 2003: A simple model of the hurricane boundary layer. *Quarterly Journal of the Royal Meteorological Society*, **129**, 1007–1027.

Tao, D., K. Emanuel, F. Zhang, R. Rotunno, M. M. Bell, and R. G. Nystrom, 2019: Evaluation of the Assumptions in the Steady-State Tropical Cyclone Self-Stratified Outflow Using Three-Dimensional Convection-Allowing Simulations. *Journal of the Atmospheric Sciences*, **76**, 2995–3009, https://doi.org/10.1175/jas-d-19-0033.1.

Wang, D., and Y. Lin, 2020: Size and structure of dry and moist reversible tropical cyclones. *Journal of the Atmospheric Sciences*, **77**, 2091–2114, https://doi.org/10.1175/JAS-D-19-0229.1.

10. Axisymmetric Spin-up and Transients

Working backward from a description of the steady-state physics of axisymmetric tropical cyclones, we next inquire how incipient vortices spin up. Here we define "spin-up" as intensification that is driven primarily or exclusively by heat fluxes from the surface, and we assume that the inner core of the storm is water saturated, or nearly so, through the whole troposphere. This is, of course, an idealization that assumes that "genesis" has already occurred and that the system is unmolested by external influences such as cooling of the upper ocean and environmental wind shear. (Here we define genesis, somewhat tautologically, as development that occurs before the core is mostly water saturated and for which other processes, such as cloud-radiation interaction, are important, as will be discussed in Chapter 14.)

To make progress toward understanding spin-up, we make several simplifying assumptions and briefly discuss the conditions for their validity. The first, as implied by the title of this chapter, is that the process is essentially axisymmetric. By this we do not mean to imply that non-axisymmetric processes have no influence on development, merely that they are not essential. Real storms are, of course, never perfectly axisymmetric and so it is thus not possible, a priori, to assume axisymmetry. Indeed, several prominent theories of intensification are fundamentally asymmetric (e.g. Montgomery et al. 2006).

The second important assumption we make is that the azimuthal flow above the boundary layer is very nearly in gradient balance, in spite of the fact that it is evolving. Of course, to evolve at all, there must be radial flow, which means that there must be radial accelerations, which means that the flow cannot strictly be in gradient balance. Thus, we call the assumption of gradient wind balance a "quasi-balance" assumption, completely analogous to the classical quasi-geostrophic approximation often made in large-scale dynamics of the atmosphere and ocean.

At the heart of the quasi-geostrophic approximation lies the geostrophic momentum approximation, which is valid if parcels of fluid accelerate over time scales much larger than the inverse of the Coriolis parameter (Hoskins 1975). Similarly, gradient balance will apply if radial accelerations are small:

$$\left|\frac{du}{dt}\right| << \left(f + \frac{v}{r}\right)v \tag{10.1}$$

In the far outer region, this, as we might have guessed, reduces to the same time scale against which to measure the validity of the classical geostrophic momentum approximation. Conversely, in the rapidly rotating inner core, we can neglect the Coriolis term in (10.1) and approximate the magnitude of the radial acceleration by $u\partial u / \partial r$. If the variation of the radial velocity occurs over a length scale comparable to the radius itself, then this reduces to

$$|u| << v. \tag{10.2}$$

That is, if the inflow angle (the angle that the streamlines make to the azimuthal direction) is small, the flow will be in approximate gradient balance. Above the boundary layer, both observations and models show that (10.2) is well satisfied, although it can be violated in the

outflow layer, where at certain radii the radial flow can be as strong or stronger than the tangential flow. It is also violated within the boundary layer (see Chapter 9 section 9.5), but that does not affect the following development.

Finally, we will assume that the time scales of slantwise convection are shorter than those characterizing the spin-up of the vortex, so that approximate convective neutrality is maintained, and so saturation moist entropy surfaces and angular momentum surfaces coincide:

$$s^* = s^*\left(M, \tau\right), \tag{10.3}$$

where τ is the time over which the saturation entropy s^* varies on surfaces of constant angular momentum, M.

All these assumptions can be, and have been tested against full-physics, clouds-permitting axisymmetric simulations, and have been found to be reasonable, at least in the core of simulated tropical cyclones (Bryan and Rotunno 2009).

As was shown in Chapter 9, the assumption that angular momentum and saturation entropy surfaces are congruent is equivalent to assuming that the saturation potential vorticity (*SPV*, 9.33) is zero everywhere in the troposphere above the boundary layer. As we are assuming hydrostatic and gradient wind balance above the boundary layer, *SPV* is "invertible", given suitable boundary conditions (Hoskins et al. 1985). Since it is constant and invertible, any time dependence of the solutions must enter through the boundary conditions. In this respect, our model of spin-up resembles the well-known quasi-geostrophic version of the Eady model of baroclinic instability, in which the pseudo potential vorticity is a single constant and all the dynamics happens on the upper and lower boundaries (see section 9.2.2). In that case, the potential temperature on the boundaries is determined solely by advection. In our case, the distribution of moist entropy at the lower boundary is determined by frictionally driven flow across angular momentum surfaces, by surface fluxes, and by dissipation of wind energy into heat.

10.1 Elements of a simple model of inner core spin-up

The aforementioned forms the basis of a relatively simple model of tropical cyclone spin up. The "invertibility" principle takes the form of the thermal wind equation (9.40), developed with the approximations discussed above:

$$\frac{M}{r_b^2} = \frac{M}{r_o^2} - \left(T_b - T_o\right)\frac{\partial s^*}{\partial M}, \tag{10.4}$$

where r_b and r_o are the physical radii of angular momentum surfaces at the top of the boundary layer and at the tropopause, respectively, while T_b and T_o are the absolute temperature at the top of the boundary layer and at the tropopause, respectively. The partial derivative in (10.4), in contrast to (9.40), appears because now s^* is also a function of time.

Recall that in developing the expression for the outflow temperature, we defined r_o as the radius at which the azimuthal velocity vanishes along M surfaces. At that point, from the definition of M, $M = \frac{1}{2} f r_o^2$. Using this in the first term on the right side of (10.4) gives

$$\frac{M}{r_b^2} = \frac{1}{2} f - \left(T_b - T_o\right) \frac{\partial s^*}{\partial M},$$

or, equivalently,

$$r_b^2 = \frac{M}{\frac{1}{2} f - \left(T_b - T_o\right) \frac{\partial s^*}{\partial M}}. \tag{10.5}$$

This expression is supplemented by (9.62), which, when integrated, yields the outflow temperature T_o as a function of M.

The relations (10.3) and (10.5) suggest that angular momentum is a natural independent variable for this system, while the radius of M surfaces is a useful dependent variable. This suggests that we ought to learn to think about tropical cyclone spin-up in angular momentum coordinates. This is analogous to geostrophic coordinates that are used in semi-geostrophic theory (Hoskins 1975). We shall see that there are many advantages to using angular momentum coordinates in this problem.

Note that to solve (10.4) we will need to know, at each time, the distribution of s^* in M space, which will also allow us to solve (9.62) for the outflow temperature. As we are continuing to assume convective neutrality, the actual moist entropy in the boundary layer, s_b, must equal the saturation entropy s^* above the boundary layer. In Chapter 9, we already developed a budget equation for boundary layer entropy, (9.42), which we repeat here:

$$\delta \left(\frac{\partial s_b}{\partial \tau} + \frac{dM}{dt} \frac{\partial s_b}{\partial M} \right) = \frac{C_k \left| \mathbf{V}_s \right| \left(k_0^* - k_s \right) + \mu C_D \left| \mathbf{V}_s \right|^3}{T_b}, \tag{10.6}$$

where δ is the boundary layer depth and we have assumed that moist entropy is well mixed along M surfaces in the boundary layer. Recall that μ is a factor with values between zero and unity we can use to control the amount of dissipative heating.

Now if we assume that the sink of angular momentum is strictly owing the vertical convergence of azimuthal stress in the boundary layer, then

$$\frac{dM}{dt} = r_b \frac{1}{\rho} \frac{\partial \tau_\theta}{\partial z}, \tag{10.7}$$

where τ_θ is the azimuthal component of the turbulent stress. Integrating (10.7) in the vertical, assuming that turbulent stress vanishes at the top of the boundary layer, and using the aerodynamic drag formula, we have

$$\delta \frac{dM}{dt} = -C_D r_b \mid \mathbf{V}_s \mid V_s, \tag{10.8}$$

where we have also assumed that angular momentum does not vary much with height in the boundary layer.

Substituting this into (10.6) yields another form of the boundary layer moist entropy equation:

$$\delta \frac{\partial s_b}{\partial \tau} = C_D r_b \mid \mathbf{V}_s \mid V_s \frac{\partial s_b}{\partial M} + \frac{C_k \mid \mathbf{V}_s \mid \left(k_0^* - k_s\right) + \mu C_D \mid \mathbf{V}_s \mid^3}{T_b}, \tag{10.9}$$

Making the minor approximation that $\left(k_0^* - k_b\right) / T_s \simeq s_0^* - s_b,$ and that $s_b = s^*$ (moist convective neutrality), we write (10.9) as

$$\delta \frac{\partial s^*}{\partial \tau} = C_D r_b \mid \mathbf{V}_s \mid V_s \frac{\partial s^*}{\partial M} + C_k \mid \mathbf{V}_s \mid \left(s_0^* - s^*\right) + \mu C_D \frac{\mid \mathbf{V}_s \mid^3}{T_s}. \tag{10.10}$$

To avoid having to use a complex boundary layer model, we also approximate the total surface wind and the surface azimuthal wind by the gradient wind V_g :

$$\mid \mathbf{V}_s \mid \simeq V_s \simeq V_g, \tag{10.11}$$

This allows us to write the lower boundary condition as:

$$\delta \frac{\partial s^*}{\partial \tau} = C_D r_b V_g^2 \frac{\partial s^*}{\partial M} + C_k V_g \left(s_0^* - s^*\right) + \mu C_D \frac{V_g^3}{T_b}. \tag{10.12}$$

Given r_b as a function of the independent variable $M,$ we have, from the definition of M,

$$V_g = \frac{M}{r_b} - \frac{1}{2} f r_b. \tag{10.13}$$

In summary, our system consists of (10.5), (10.12), (10.13), and the outflow equation (9.62).

It is important to note that the time-dependent dynamics enter only through the boundary layer moist entropy equation (10.12), much like the Eady model of baroclinic instability, albeit with very different physics of boundary entropy evolution. This equation will only really be valid in the saturated core of the storm where there is convective neutrality and no downdrafts entering the boundary layer, either through convection or mean subsidence. It is therefore insufficient to model the whole time-dependent dynamics of tropical cyclones but, as we will see, it does capture the essence of TC spin-up. As with the Eady model, the interior flow is slaved to the evolution of entropy at the boundaries, really only the lower boundary in this case. The interior zero-SPV condition together with the upper boundary condition on T_o dictates how the boundary layer moist entropy inverts to provide the low-level gradient wind.

Before showing and discussing numerical solutions of this system, it is worth thinking carefully about the boundary moist entropy evolution as expressed by (10.12).

First note that each of the three terms in the right side of (10.12) have a surface exchange coefficient in front of them. Without surface fluxes and drag, nothing happens in this system, consistent with our understanding of the steady state physics of tropical cyclones.

The first term on the right of (10.13) is the advection of boundary layer moist entropy across M surfaces by the torque owing to surface drag. This is the radial "flow" in this system. It is negative definite because the gradient wind is positive while $\partial s^*/\partial M$ is always negative. Using (10.5) for $\partial s^*/\partial M$ and also making use of (10.13), the first term on the right of (10.12) becomes

$$-C_D \frac{V_g^3}{T_b - T_o}.$$

This can be combined with the last term in (10.12), and because $T_b - T_o < T_b$, this negative term will dominate. Thus, an alternative formulation of (10.12) (in the case that $\mu = 1$) is

$$\delta \frac{\partial s^*}{\partial \tau} = C_k V_g \left(s_0^* - s^* \right) - C_D V_g^3 \frac{T_o}{T_b} \frac{1}{T_b - T_o}. \tag{10.14}$$

A glance at (10.14) reveals that the time tendency of the saturation entropy will vanish when V_g equals the potential intensity at the radius of maximum winds.

The other term on the right side of (10.12) and (10.14) is the surface enthalpy flux, expressed approximately as a surface entropy flux. It drives the boundary layer entropy toward the saturation entropy of the sea surface, and is countered by the advection term.

It is important to note that aside from the Ekman flow across M surfaces, there are no terms in this system that in any way depend on the secondary circulation, consisting of the (non-Ekman) radial and vertical velocities. As in semi-geostrophic theory with constant interior potential vorticity, these components of the secondary circulation can be diagnosed from the solution of the balanced flow. We will return to this point presently.

10.1.1 Some approximate solutions for rates of intensification

Certain aspects of the simple model can be deduced simply by differentiating (10.14) with respect to the independent coordinate, angular momentum, M:

$$\begin{aligned} \delta \frac{\partial}{\partial \tau} \frac{\partial s^*}{\partial M} = & -C_k V_g \frac{\partial s^*}{\partial M} - C_D V_g^3 \frac{1}{\left(T_b - T_o \right)^2} \frac{\partial T_o}{\partial M} \\ & + \frac{\partial V_g}{\partial M} \left[C_k \left(s^*{}_0 - s^* \right) - 3 C_D V_g^2 \frac{T_o}{T_b} \frac{1}{\left(T_b - T_o \right)} \right]. \end{aligned} \tag{10.15}$$

Next, we substitute (10.4) into (10.15) to eliminate $\partial s^*/\partial M$:

$$\delta \frac{\partial}{\partial \tau} \frac{V_g}{r_b \left(T_b - T_o\right)} = -C_k V_g \frac{V_g}{r_b \left(T_b - T_o\right)} + C_D V_g^3 \frac{1}{\left(T_b - T_o\right)^2} \frac{\partial T_o}{\partial M} - \frac{\partial V_g}{\partial M} \left[C_k \left(s^*_0 - s^* \right) - 3 C_D V_g^2 \frac{T_o}{T_b} \frac{1}{\left(T_b - T_o\right)} \right]. \tag{10.16}$$

We now make two additional approximations to (10.16): We neglect the time dependence of the outflow temperature, T_o, on the left side and, in this same term and in the first term on the right of (10.16) , we approximate r_b as M / V_g, which is valid in the core, where $fr_b << V_g$. The result is

$$\delta \frac{\partial}{\partial \tau} V_g^2 = -C_k V_g^3 + C_D V_g^3 \frac{M}{\left(T_b - T_o\right)} \frac{\partial T_o}{\partial M} - M \frac{\partial V_g}{\partial M} \left[C_k \left(s^*_0 - s^* \right) \left(T_b - T_o\right) - 3 C_D V_g^2 \frac{T_o}{T_b} \right]. \tag{10.17}$$

For the vortex to amplify, the time tendency of the gradient wind must be positive at the radius of maximum winds, where $\partial V_g / \partial M = 0.$ A necessary condition for amplification is then

$$\frac{M}{\left(T_b - T_o\right)} \frac{\partial T_o}{\partial M} > \frac{C_k}{C_D}. \tag{10.18}$$

So there must be a positive gradient of outflow temperature across the radius of maximum winds, if the vortex is to amplify in this simple system.

This is a surprising result. Insofar as turbulence is necessary to create outflow of non-zero thickness, according to the critical Richardson Number hypothesis – and thereby sustain a positive gradient of outflow temperature with respect to angular momentum – then outflow turbulence is necessary for vortex amplification. Tropical cyclone intensification, in this view, is driven by boundary layer turbulent fluxes from the sea and outflow layer turbulence sustaining a gradient of absolute temperature across the outflow layer.

We can be a bit more quantitative about the rate of intensification. Let's always evaluate (10.17) at the radius of maximum winds, where we assume (as in the steady model) that $T_o = T_t,$ where T_t is the tropopause temperature:

$$2\delta \frac{\partial V_{gm}}{\partial \tau} = -C_k V_{gm}^2 + C_D V_{gm}^2 \frac{M_m}{\left(T_b - T_t\right)} \frac{\partial T_o}{\partial M}, \tag{10.19}$$

where V_{gm} is the maximum wind over radius at any given time, occurring when $M = M_m.$ Next, we substitute (9.62) for $\partial T_o / \partial M,$ neglecting the first term on the right side of (9.62) and also substituting from (10.4) for $\partial s^* / \partial M$:

$$2\delta \frac{\partial V_{gm}}{\partial \tau} = -C_k V_{gm}^2 + C_D V_{gm}^2 M_m \frac{Ri_c}{r_t^2} \frac{r_m}{V_{gm}}, \tag{10.20}$$

Recall that, in formulating the steady state model, we regarded Ri_c / r_t^2 as a constant that can be determined by specifying either the radius of maximum winds or the outer radius. What constant value should be used here? For consistency with the simple steady-state model developed in Chapter 9, the steady state value of the maximum winds should be given by V_{max}, equivalent to V_m given by (9.70) and (9.71). For this to be the steady state gradient wind attained according to (10.20), it must be true that

$$\frac{Ri_c}{r_t^2} = \frac{C_k}{C_D} \frac{1}{r_{max}^2}, \tag{10.21}$$

where r_{max} is the steady-state value of the radius of maximum winds. Note that (10.21) is identical to (9.73), obtained from the steady-state model. Using this together with $M_m \simeq r_m V_{gm}$ transforms (10.21) into

$$\frac{\partial V_{gm}}{\partial \tau} = \frac{C_k}{2\delta} V_{gm}^2 \left(\frac{r_m^2}{r_{max}^2} - 1 \right). \tag{10.22}$$

We make one final approximation, which is somewhat dubious – that the maximum wind speed always resides on the same M surface. If that is the case, and we again neglect Coriolis terms in the inner core, then $r_m = M_m / V_{gm}$ and $r_{max} = M_m / V_{max}$, and (10.22) reduces to

$$\frac{\partial V_{gm}}{\partial \tau} = \frac{C_k}{2\delta} \left(V_{max}^2 - V_{gm}^2 \right). \tag{10.23}$$

This has a simple analytic solution, which for the particular case that $V_{gm} = 0$ at time $\tau = 0$ is given by

$$V_{gm}(\tau) = V_{max} \tanh \left(\frac{C_k V_{max}}{2\delta} \tau \right). \tag{10.24}$$

At first blush, it might appear that the intensification rate does not depend on the drag coefficient, but remember that V_{max}, as given by (9.70) and (9.71) depends on both exchange coefficients. From (10.24) with (9.70) and (9.71), both the maximum intensity and the rate of intensification depend on the potential intensity.

This solution is plotted in Figure 10.1. The dimensional time scale over which the storm intensifies is $2\delta / C_k V_{max}$, which for typical values of the parameters is around 15 hours. If the maximum dimensional wind speed were 150 knots, the maximum 24-hour intensity change according to Figure 10.1 would be around 135 knots. This is somewhat more than the record rate of about 100 knots in 24 hours observed in eastern North Pacific Hurricane Patricia of 2015.

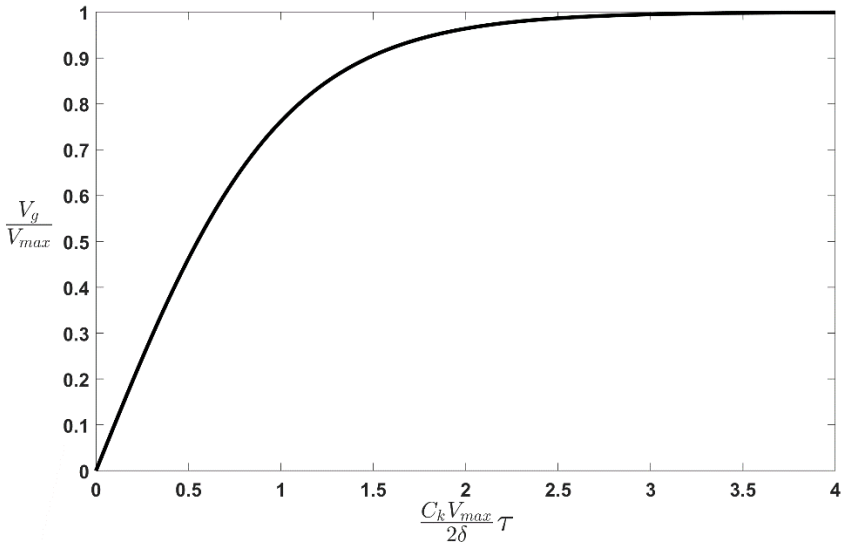


Figure 10.1: Solution of (10.24) showing the evolution of normalized gradient wind speed over normalized time.

10.1.2 Eyewall frontogenesis and eye formation

Here we will show that the assumption of a saturated, moist neutral core is inconsistent with the observation that the low-level flow is mostly or entirely cyclonic right into the center. First, note that, by symmetry, $V_g = 0$ at $r = 0,$ so according to (10.14) it is not possible for the temperature (saturation moist entropy) to increase with time there. Yet, from thermal wind balance, $\partial s^* / \partial M < 0$ everywhere that $V_g > 0$, so that s^* must reach its peak value on the axis. It cannot do this, according to (10.14), which must fail there.

This contradiction arises because we have ignored the special dynamics of the eye, as described in Chapter 9, section 9.4. Here, radial stresses with the eyewall force descent, the eye above the boundary layer warms and its s^* decouples from the boundary layer moist entropy. Therefore, our simple model, which assumes convective neutrality, is invalid in the eye.

Yet, the simple model actually predicts its own failure and shows us that, without radial diffusion, a discontinuity would form at the eye wall in a finite time. Such a discontinuity, or front, would be extremely unstable to non-axisymmetric eddies and quickly break down, with the resulting eddies diffusing angular velocity into the eye.

To see how our simple model predicts that this must occur, recall that it is phrased using angular momentum (rather than physical radius) as the independent radial coordinate. This coordinate transformation is valid as long as long as the Jacobian of the transform, $\partial M / \partial r$, is well behaved (no zeros or infinities). In the inner region, where we may approximate M by rV, we can write $r \simeq M / V$ and therefore, $\partial r / \partial M \simeq \left(V - M \partial V / \partial M\right) / V^2$. The transform fails if there are regions where $\partial V / \partial M \geq V / M$. The vorticity is given by

$$\zeta = \frac{1}{r}\frac{\partial M}{\partial r} = \frac{V^3}{M\left(V - M\dfrac{\partial V}{\partial M}\right)}, \tag{10.25}$$

and this will blow up at the point that the coordinate transform fails. This can happen inside the radius of maximum winds, where $\partial V / \partial M > 0$. We can keep integrating our angular momentum coordinate system in time beyond when this happens, but we will not be able to transform the solution back into physical space. The formation of the discontinuity is signified by vorticity approaching infinity.

What happens here is strongly analogous to the what happens in semi-geostrophic frontogenesis, and involves feedbacks between the balanced component of the flow and the non-balanced component, which is frontogenetic.

In the classical semi-geostrophic frontogenesis problem (Hoskins and Bretherton 1972), the background deformation that ultimately drives the frontogenesis is the large-scale, geostrophic deformation. In the case of tropical cyclones, however, the background deformation is the Ekman flow. In both cases, the direct contribution to frontogenesis by the background flow is augmented by an additional component owing to the non-balanced response. As the front becomes stronger, this second component becomes yet stronger, in a positive feedback loop that results in frontal collapse in a finite time.

To see how this works, let's take another look at the velocity evolution equation (10.17), which we slightly re-write as

$$\begin{aligned}\delta \frac{\partial}{\partial \tau} V_g^2 = C_D V_g^3 \frac{M}{\left(T_b - T_o\right)} \frac{\partial T_o}{\partial M} - C_k V_g^3 \\ + M \frac{\partial V_g}{\partial M} C_D \frac{T_o}{T_b}\left[3V_g^2 - V_{max}^2\right].\end{aligned} \tag{10.26}$$

We have seen that the first two terms cause the maximum wind speed to increase toward the potential maximum wind, while, of course, remaining zero at the storm center. So these two terms already increase $\partial V_g / \partial M$ inside the radius of maximum winds. But since V_g is also increasing with time, there is no guarantee that the vorticity, as given by (10.25), will blow up.

But now look at the last term on the right side of (10.26). If the peak gradient wind satisfies $3V_{gm}^2 > V_{max}^2$, then the term in square brackets in (10.26) will be positive just inside the radius of maximum winds, but become negative somewhere at smaller radii, where $3V_{gm}^2 < V_{max}^2$. Thus the last term in (10.26) acts to further increase $\partial V_g / \partial M$ inside the radius of maximum wind. It seems likely that the vorticity, given by (10.25), will eventually blow up, and this is confirmed by the full numerical solutions described in the next section. Indeed, the front forms close to the radius at which $V_{gm} = V_{max} / \sqrt{3}$.

One small departure of this system from the classical semi-geostrophic frontogenesis problem is that the background forcing remains constant over time in the latter case, whereas it increases with time as the tropical cyclone develops.

How does a frontal discontinuity (infinite vorticity) form in a finite time? To see that clearly, we should think about the problem in physical space. First, just as we see in M coordinates, the Ekman flow serves to increase the magnitude of the radial entropy gradient, $\partial s^* / \partial r$. But in physical space we have an additional radial velocity – beyond the Ekman flow – that can be diagnosed by first writing the azimuthal velocity equation:

$$\frac{\partial v}{\partial t} + u\left(f + \frac{V}{r} + \frac{\partial V}{\partial r} \right) = -F, \tag{10.27}$$

where F represents the frictional retardation of the flow and we have ignored vertical advection in the boundary layer. We can simply define the radial flow, u, as consisting of the sum of two components: an Ekman flow component, u_{Ek}, and a component, u_{in}, related to the intensification of the azimuthal wind:

$$u_{Ek} = \frac{-F}{\eta}, \qquad u_{in} = \frac{-\frac{\partial V}{\partial t}}{\eta}, \tag{10.28}$$

where η is the absolute value of the vertical component of vorticity:

$$\eta \equiv f + \frac{V}{r} + \frac{\partial V}{\partial r} = \frac{1}{r}\frac{\partial M}{\partial r}. \tag{10.29}$$

Of course, moist entropy is advected radially by both of these components of the radial flow. The first component is easily calculated from the surface stress, boundary layer depth, and near-surface wind field. The second component, u_{in}, can be diagnosed from the time-dependent solution of our system in angular momentum space, mapped back into physical space at each time step. It shows that where the azimuthal wind is increasing, there will be an inward radial flow. Because the absolute vorticity, the denominator in (10.28), decreases outward across the radius of maximum winds, the magnitude of the radial velocity increases outward. This serves to further hasten the compression of isolines of moist entropy, and is the feedback that leads to frontal collapse in a finite time.

If we do not do anything about this in our simple model, the coordinate transform will fail and we cannot map our solution back into physical coordinates. To prevent this failure, we take one or both of two approaches. In the first of these, after solving (10.5) for the radius at every time step, we regularize the radii by marching inward from the outer boundary and insisting that

$$r_i \le r_{i+1},$$

resetting $r_i = r_{i+1}$ where this condition is violated. Here i is just the index of the discrete set of radial points in a finite difference model in M space. Strictly speaking this is not an energy-conserving adjustment, but since we are dealing with a region of zero volume in physical coordinates, it will have a negligible effect on the volume-integrated kinetic energy.

In the second approach, we simply assume that inside the radius of maximum angular velocity, V_g / r_b, the gradient wind, determined by the dynamics described in section 9.4, obeys

$$V_g = r_b \left(\frac{V_g}{r_b} \right)\Big|_{max}, \tag{10.30}$$

where the factor in parentheses is the maximum angular velocity in the solution. But we continue to assume that thermal wind balance holds in the eye and that s^* continues to be invariant along M surfaces there. We also assume that the outflow temperature is constant and equal to the tropopause temperature T_t across the eye. Then, from the thermal wind relation (10.4), we have, in the eye,

$$\left(T_b - T_t\right) \frac{\partial s^*}{\partial M} = -\frac{V_g}{r_b} = -\left(\frac{V_g}{r_b} \right)\Big|_{max}, \tag{10.31}$$

where we have used (10.29). Thus, the eye, assumed to be in solid body rotation, has a constant gradient of saturation moist entropy with respect to angular momentum.

How do we know where the eyewall is? As we solve the simple model for the saturated inner core, described above, we constantly calculate (among other things), the quantity $-\partial s^* / \partial M$. Moving inward, where that quantity drops below the magnitude it would have from (10.31), we assume that we have crossed the inner edge of the eyewall into the eye, and apply (10.31) from that point inward. If we like, we can then solve the time-dependent equation (10.9) for the boundary layer moist entropy (which is, in the eye, decoupled from s^*) to ascertain that it remains less than s^*.

Note that the eyewall occurs at the point that V_g / r_b has its maximum value (and, by assumption, is constant inside of that radius). This necessarily occurs inside the point where V_g itself has a maximum value. Thus, in this simple model, the eyewall is inside the radius of maximum winds.

In solving our simple model numerically, we find the eyewall at every time step and apply (10.30) and (10.31) inside of it; we also solve the boundary layer moist entropy equation (10.9). By this means, we have the radial distributions of V_g, s^*, and (separately) s_b in the eye.

10.1.3 Full numerical solutions of the simple time-dependent model

Here we solve the system given by the time-dependent boundary condition (10.12) along with the diagnostic equations (10.5) and (10.13), and the one-dimensional ordinary differential equation for the outflow temperature, T_o, given by (9.62). We use the full form of the latter, except that given the observed low temperature of most of the outflow, we approximate the denominator of the first term on the right of (9.62) by unity. Also, to provide a smoother time integration, we relax the actual outflow temperature to the solution of (9.62) over time. We apply one or both of the aforementioned approaches to representing the eye. The model is run on a one-dimensional grid of points evenly spaced in angular momentum, and marched forward in time using a leap-frog scheme with an Asselin filter.

Figure 10.2 shows the evolution of the maximum wind speed for the case in which the physical radii in the eye are regularized, but without the additional constraint (10.31). In this simulation, dissipative heating has been switched off and the pressure effect on the surface saturation entropy is ignored. Choices of other parameters may be found in the header of the MATLAB script used here, *time_dependent_model.m,* available at https://zenodo.org/doi/10.5281/zenodo.10956451 .

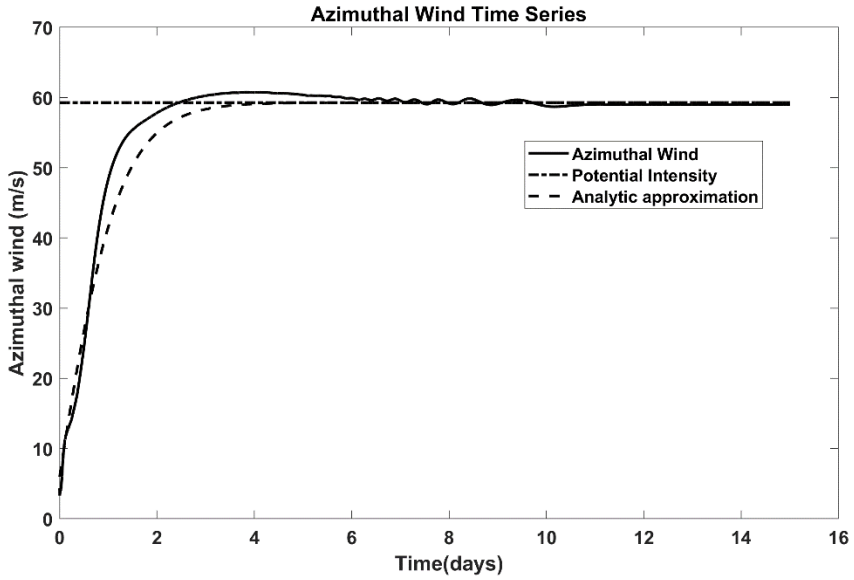


Figure 10.2: Evolution with time of the gradient wind in a simulation using the simple model described in this section. The solid black line shows the simulated gradient wind, and the dashed line shows the approximate solution given by (10.24) and using the same parameters as used in the full simulation. The flat, dash-dot line is the theoretical potential intensity.

The full model wind speed intensifies somewhat faster than the analytic approximation but asymptotically approaches the same steady-state value, which is also very close to the theoretical potential intensity given by (9.70) and (9.71).

The evolutions of the gradient wind, radial velocity, vertical velocity and saturation entropy are shown as time-radius (or Hovmöller) diagrams in Figure 10.3. The frontal discontinuity is evident in the gradient wind, vertical velocity and saturation entropy fields as a black curve inside the radius of maximum winds. The inner core gradient wind reaches a nearly steady state fairly quickly (after only about 4 days), whereas the vertical and radial velocities and the saturation entropy do not equilibrate until roughly 10 days.

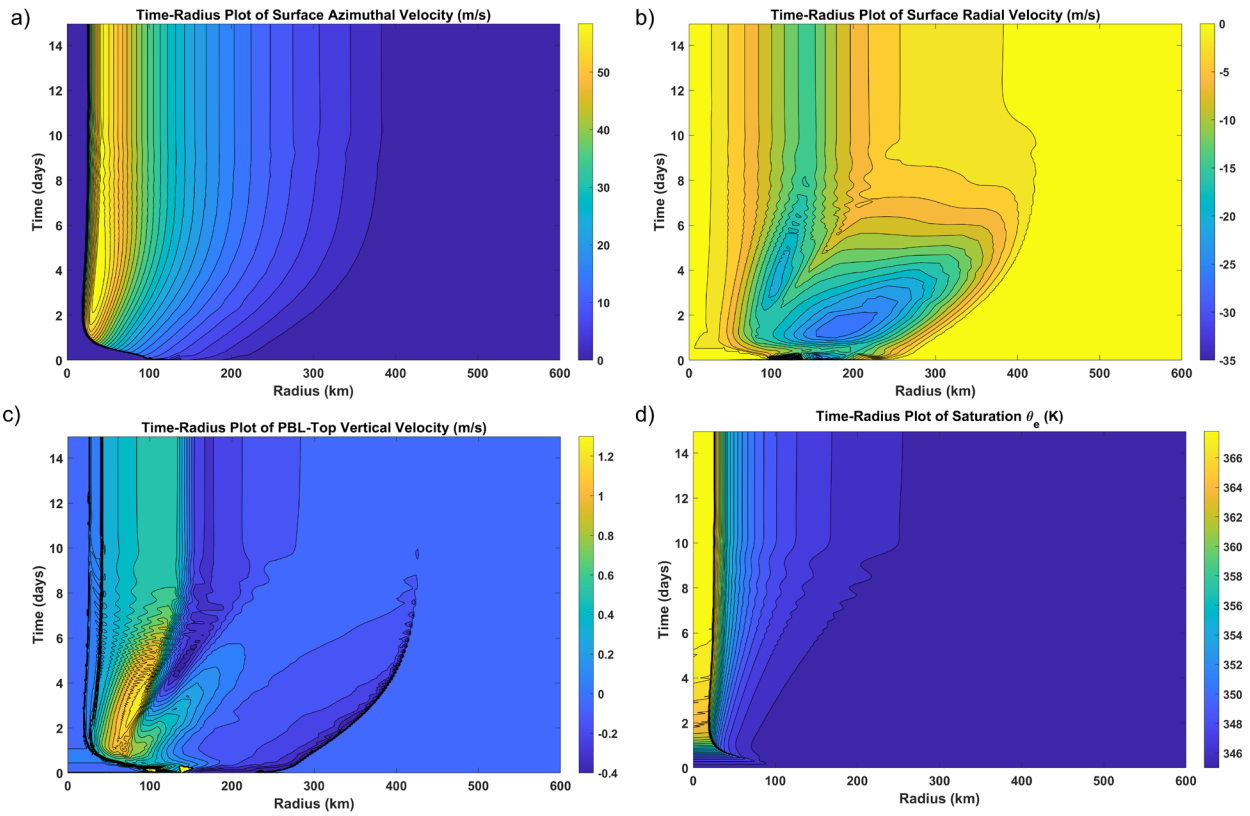


Figure 10.3: Time-radius plots of surface gradient wind (a), radial velocity (b), vertical velocity (c) and saturation entropy (d).

One can see that a discontinuity forms almost immediately and persists throughout the integration; this is evident by what appears to be a curved black line in Figure 10.3a inside of which the gradient wind vanishes. After roughly day 1, roughly the time when $V_{gm} = V_{max} / \sqrt{3}$, this discontinuity occurs inside the radius of maximum winds.

The radial velocity is largest during the most rapid intensification of the outer wind field, as one might expect from (10.28). Its peak value is somewhat larger than the peak value of the Ekman radial flow. It then settles down into a near steady state in which only the Ekman flow contributes to the radial velocity[1]. Up through about 7 days there are two peaks in the (negative) radial velocity; one is associated with the Ekman flow and the other with the spin-up component as defined by (10.28). Note that we would have obtained much larger Ekman radial velocity in the outer region, but we placed a minimum value of $\frac{f}{4}$ on the denominator of that term as defined in (10.28); in fact, the model wants to produce much smaller values of absolute vorticity

[1] Since the model is run in angular momentum coordinates, it is more convenient to diagnose the two components of radial velocity using it definition: $u = \frac{dr}{dt} = \frac{\partial r}{\partial \tau} + \frac{dM}{dt}\frac{\partial r}{\partial M}$. Comparing this to (10.28) we see that we can write $u_{in} = \frac{\partial r}{\partial \tau}$. (The other component, $\frac{dM}{dt}\frac{\partial r}{\partial M}$, is the same as the definition of u_{Ek} in (10.28).)

in the outer region. Note that this cap does not at all affect the solution we actually obtain in angular momentum coordinates; it just affects the diagnosed values of radial velocity and, through mass continuity, vertical velocities.

The vertical velocity field is surprisingly complex given the simplicity of the model. Recall that, because we have assumed that the atmosphere is saturated, the vertical velocity has no feedback on the flow. There are three curves in Figure 10.3c demarking discontinuities of the vertical velocity with radius. Although the radial velocity u must be continuous, its first derivative in radius may not be, and where it is not, there will be a jump in vertical velocity, as demanded by mass continuity.

The innermost curve of discontinuity in w coincides with the front evident in the gradient wind (Figure 10.3a). The discontinuity just outside this one occurs at the radius of maximum winds, where there is a large change in the radial gradient of radial velocity. Although it is not visible in the figure, a closer examination shows that the eyewall peak in vertical velocity is larger than the peak outside the eyewall after about day 4. Figure 10.4 shows the radial profile of vertical velocity at the end of the simulation. The two jumps in vertical velocity are clearly evident, as is a broad peak at a radius of around 120 km. Note also the rather steep decline in w at a radius of about 150 km. The third discontinuity evident in Figure 10.3c is at the outer radius of vanishing wind and is not present in the steady state.

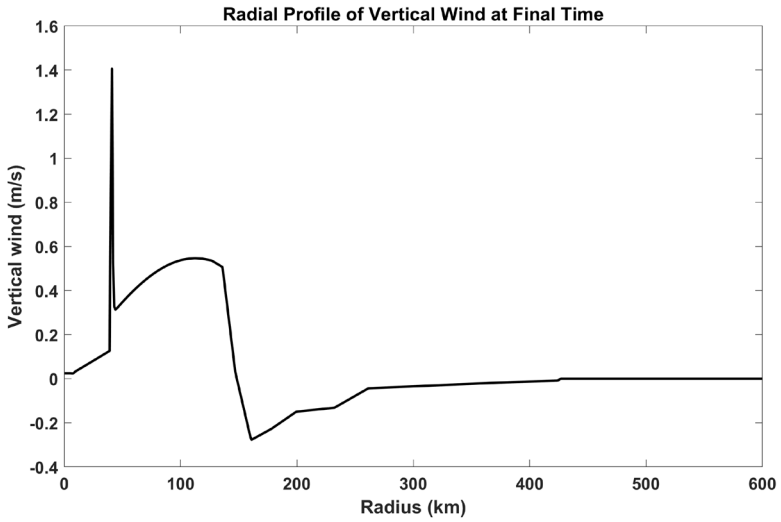


Figure 10.4: Radial profile of vertical velocity at the end of the simulation.

The saturation entropy distribution (Figure 10.3d) is rather unremarkable. While it is continuous across the front, its first derivative is not, coinciding with the discontinuity in gradient wind.

The frontogenesis inside the radius of maximum winds is exceptionally powerful in tropical cyclones. To see this, we can calculate a measure of frontogenesis, which is the time rate of change of the radial gradient of boundary layer entropy following parcels in the boundary layer. Differentiating the conservation equation for boundary layer entropy in physical coordinates and ignoring vertical advection of entropy in the boundary layer yields

$$\frac{d}{dt}\frac{\partial s_b}{\partial r} = -\frac{\partial u}{\partial r}\frac{\partial s_b}{\partial r} + \frac{1}{hT_s}\frac{\partial F_k}{\partial r}, \tag{10.32}$$

where F_k is the sum of the surface enthalpy flux and the dissipative heating and h is the boundary layer depth. The first term on the right side of (10.32) represents the deformation of the boundary layer entropy field and serves to increase the gradient of s_b where $\partial u / \partial r$ is negative. In Figure 10.5 we plot the radial distribution of the first term on the right side of (10.32) along the two components of the radial velocity ("Ekman" and "spin-up") and the full radial velocity at day 1 in the same simulation, when there is strong frontogenesis.

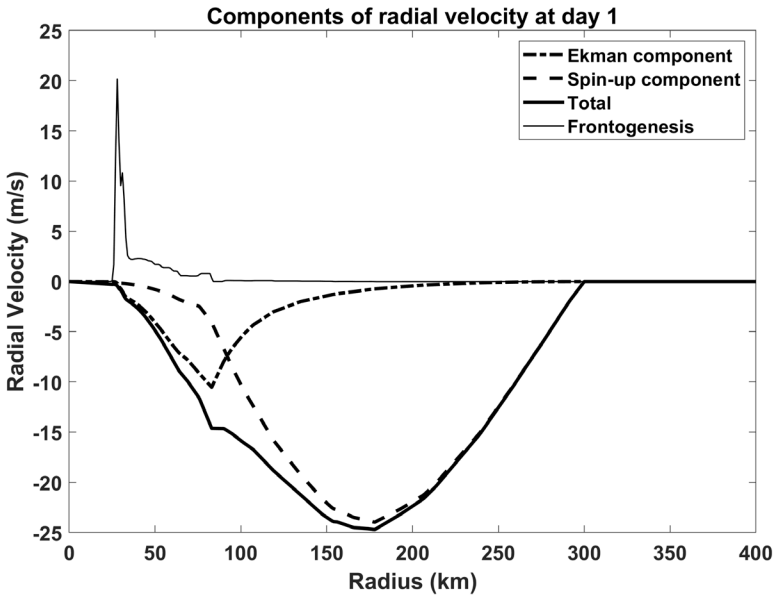


Figure 10.5: Two components of the radial velocity, the full radial velocity, and the frontogenesis function (first term on the right side of (10.32)) at day 1 of the simulation.

It is clear that, at this time, it is almost entirely the Ekman component of the radial velocity that drives the strong frontogenesis at the front itself. This turns out to be true during the entire simulation. As described previously, the Ekman flow increases with the intensity of the tropical cyclone, so there is positive feedback between the Ekman flow and the strong frontal gradients.

Simulations with diffusion of angular velocity into the eye, as represented by the implementation of (10.31), behave quite similarly to the simulation without it, described above, but the front is prevented from forming. The reader is encouraged to experiment with this simple model to obtain some appreciation of the dynamics of tropical cyclone intensification and its sensitivities to assumptions about the physics and to initial conditions.

10.1.4 A note on agradient winds

We have seen that, in angular momentum space, the radial velocity disappears from the governing equations and in its place is the total time derivative of the angular momentum. The non-Ekman part of the radial velocity is implicit and can be diagnosed from the solution. The

primary assumption in the balance model is that the azimuthal wind is nearly equal to its gradient value; that is, the vortex evolves through a sequence of balanced states, just as in quasi-geostrophic theory except that we assume gradient rather than geostrophic balance.

How good is this approximation? To answer this, we can diagnose the azimuthal component of the agradient wind from the radial momentum equation in angular momentum coordinates:

$$\frac{\partial u}{\partial \tau} + \frac{dM}{dt}\frac{\partial u}{\partial M} + fV_g + \frac{V_g^2}{r} \simeq fV + \frac{V^2}{r}, \tag{10.33}$$

where we have used the gradient wind as a proxy for the radial pressure gradient acceleration and we have ignored the vertical advection of radial velocity along angular momentum surfaces. Since we have calculated the time-evolving radial wind as a function of angular momentum, as well as the gradient wind itself, we can calculate the left side of (10.33) and from that diagnose the total azimuthal wind. Figure 10.6 displays a time-radius section of the agradient wind (difference between the total azimuthal wind and the gradient wind) corresponding to the simulation used to construct Figures 10.2-10.5.

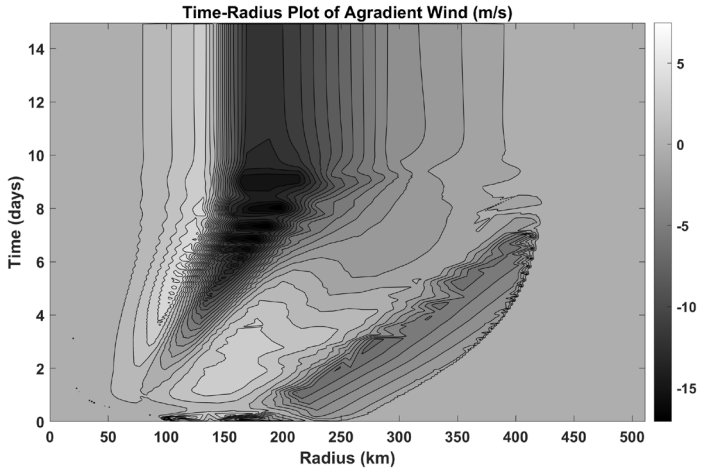


Figure 10.6: Time-radius section of agradient wind (solution of (10.33)) corresponding to the fields shown in Figure 10.3.

The field is complex, with negative values as large as $15\ ms^{-1}$, so the contribution is by no means trivial. But note, by comparing to Figure 10.3a, that the agradient wind is quite small near the radius of maximum gradient wind, so the peak wind speed is hardly affected. Toward the end of the simulation, the most negative agradient wind occurs outside the radius of minimum radial velocity, where parcels are accelerating inward; conversely, the largest positive values occur inside the radius of minimum radial velocity, where parcels are slowing down.

10.1.5 Fields above the boundary layer

In solving the simple model, we have effectively solved the time-dependent lower boundary condition for an inversion of the saturated potential vorticity field, which yields the flow and thermal fields throughout the vortex. Thus, the simple model actually provides the whole height-dependent, time-evolving fields of velocity and temperature.

To actually find these fields above the surface, we begin with a version of the thermal wind equation in angular momentum space, (9.40):

$$\frac{M}{r^2} = \frac{M}{r_b^2} + \left(T_b - T\right)\frac{ds*}{dM}, \tag{10.34}$$

where recall that r_b denotes the radius of angular momentum surfaces at the top of the boundary layer, where the temperature is T_b, and r is the radius of the same M surface where its temperature is T. Using $M = rV + \frac{1}{2}fr^2 = r_b V_b + \frac{1}{2}fr_b^2$, (10.34) may be written

$$\frac{V}{r} = \frac{V_b}{r_b} + \left(T_b - T\right)\frac{ds*}{dM}. \tag{10.35}$$

By definition, V vanishes when $T = T_o$, where T_o is the outflow temperature, which varies with M according to (9.62). Consequently,

$$\frac{V_b}{r_b} = -\left(T_b - T_o\right)\frac{ds*}{dM}. \tag{10.36}$$

Eliminating $ds*/dM$ between (10.35) and (10.36) gives an expression relating r to r_b along a surface of constant M, as a function of temperature, which serves as the effective vertical coordinate:

$$\frac{V}{r} = \frac{V_b}{r_b}\left[\frac{T - T_o}{T_b - T_o}\right]. \tag{10.37}$$

The angular velocity decreases linearly with temperature along an M surface. We have V_b and r_b at $T = T_b$ as a solution of our simple model, so we can find V/r all the way up the M surface. Using that and the definition of M, we can find V and r separately. This gives us the full, two-dimensional solution in $M - T$ space. (Mathematicians might be tempted to call this an "empty" solution!) If we can assume that the saturation entropy, $s*$, continues to be constant along M surfaces past the point where the temperature falls below T_o (a somewhat dubious assumption), we can carry the solution (10.37) all the way down to the ambient tropopause temperature.

To display the solution in more conventional coordinates, such as radius-altitude space, we first need to know the altitude corresponding to a temperature T along each M surface. To do this we first recall that such surfaces are also surfaces of constant $s*$. Because we know $s*$ and T, we can invert the definition of $s*$ to find the pressure, p. Once we know the pressure and temperature, we can calculate the saturation specific humidity, $q*$. Then we can integrate up a an inverse moist adiabat $\left(\partial z/\partial T\right)_{s*}$, to find the altitude. Alternatively, if we assume that

saturated moist static energy, h^*, is also nearly constant along a surface of constant s^*, then we can find the altitude directly from the definition of moist static energy, $gz = h^* - c_p T - L_v q^*$.

Once we know the radius, r, and the altitude, z, as a function of M and T, we can remap V and T from $M - T$ space into conventional $r - z$ space. Figure 10.7 shows these solutions for the final steady state corresponding to the solution displayed in Figure 10.3.

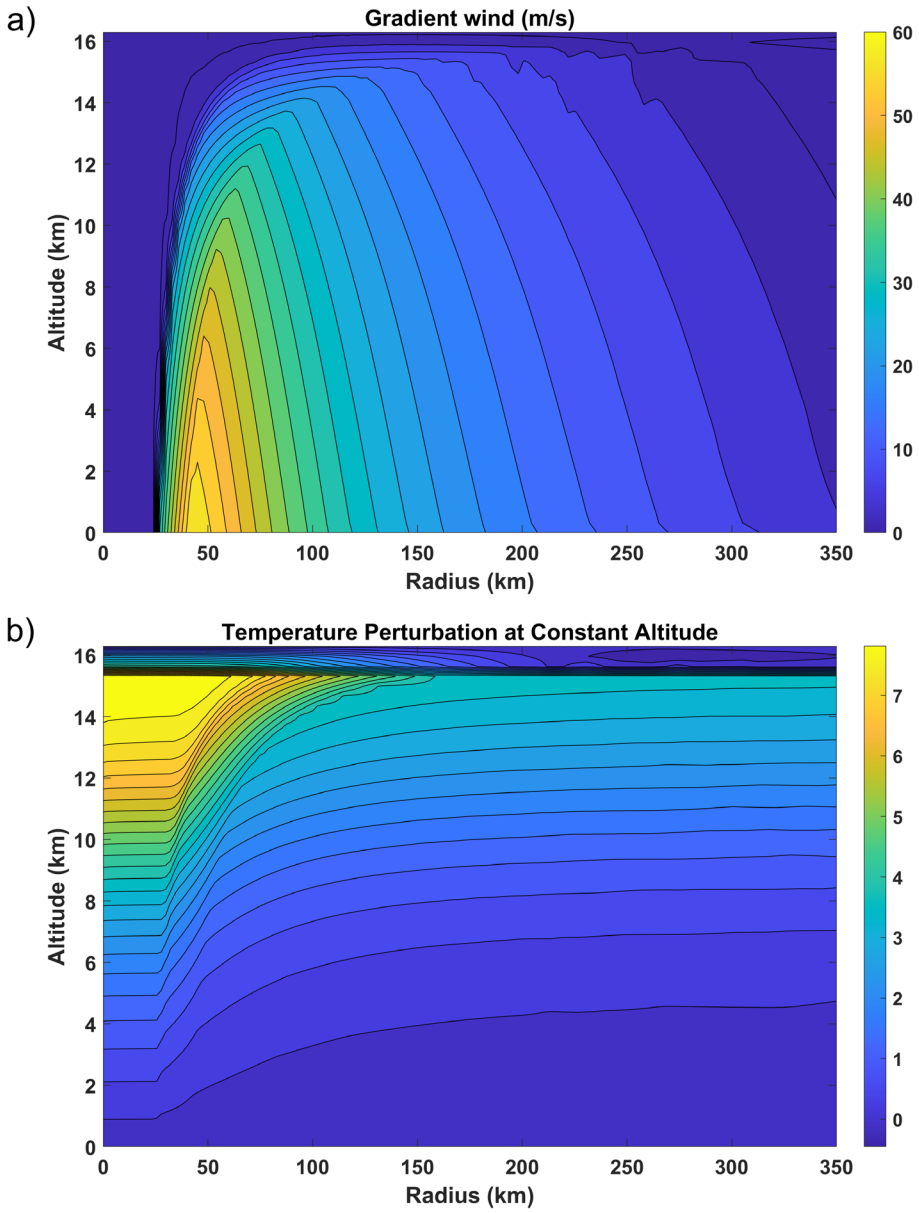


Figure 10.7: a) Azimuthal wind (m/s), and b) temperature perturbation (K) at constant altitude relative to the unperturbed environment.

The eyewall front is clearly evident in both temperature and wind, inside the radius of maximum wind. Technically, a discontinuity is present only at the surface and the gradient decays with altitude. Because there is no wind in the eye in this solution, there is also no temperature gradient.

What about the secondary solution? It is possible to derive this as well. We begin with the anelastic form of the mass continuity equation:

$$\frac{\partial(\rho r u)}{\partial r}+\frac{\partial(\rho r w)}{\partial z}=0, \tag{10.38}$$

where ρ is the air density. From this we can define a mass streamfunction, ψ, such that

$$\rho r u=-\frac{\partial \psi}{\partial z}, \quad \rho r w=\frac{\partial \psi}{\partial r}. \tag{10.39}$$

Substitution shows that (10.38) is obeyed. Then we can write

$$\frac{dr}{dt}=\frac{\partial r}{\partial \tau}+\frac{dM}{dt}\frac{\partial r}{\partial M}+w\frac{\partial r}{\partial Z}=u=-\frac{1}{\rho r}\frac{\partial \psi}{\partial z}. \tag{10.40}$$

where the derivative with respect to Z in the last term is a partial derivative holding time and angular momentum constant.

Using the chain rule, we may write

$$\frac{\partial \psi}{\partial Z}=\frac{\partial \psi}{\partial z}+\frac{\partial r}{\partial Z}\frac{\partial \psi}{\partial r}=\frac{\partial \psi}{\partial z}-\rho r w\frac{\partial r}{\partial Z}. \tag{10.41}$$

Using this for $\partial\psi/\partial z$ on the right side of (10.40) gives

$$\frac{\partial r}{\partial \tau}+\frac{dM}{dt}\frac{\partial r}{\partial M}+w\frac{\partial r}{\partial Z}=-\frac{1}{\rho r}\frac{\partial \psi}{\partial Z}+w\frac{\partial r}{\partial Z}. \tag{10.42}$$

The last terms on the right and left sides of (10.42) cancel. If we now use the hydrostatic equation for the first term on the right side of (10.42) we arrive at

$$\frac{\partial r}{\partial \tau}+\frac{dM}{dt}\frac{\partial r}{\partial M}=g\frac{\partial \psi}{\partial P}. \tag{10.43}$$

Once again, in the last term, the partial derivative holds time and angular momentum constant.

We use (10.43) as a diagnostic equation for the streamfunction, ψ. To do this, we first assume that the left side of (10.32) does not vary with height within the boundary layer, and that the second term on the left, which involves a source/sink of angular momentum, vanishes above the boundary layer. (Of course, in a steady flow, there must be a source of angular momentum somewhere, to balance the sink. We assume it is outside the domain we are going to use to construct radial-height cross-sections.) In the special case of steady flow, the first term in (10.43) also vanishes, and this means that ψ *is constant along angular momentum surfaces*

above the boundary layer in the steady state. This makes perfect sense, since M is conserved following the secondary flow above the boundary layer.

To integrate (10.43) upward to find the streamfunction above the boundary layer, we need to know what the time derivative of physical radius r is above the boundary layer. By substituting $V / r = M / r^2 - \frac{1}{2} f$ into (10.37) and differentiating in time, we have

$$\frac{\partial r}{\partial \tau} = \left[\frac{T - T_o}{T_b - T_o} \right] \frac{r^3}{r_b^3} \frac{\partial r_b}{\partial \tau}. \tag{10.44}$$

Using this in (10.43) and integrating upward yields the streamfunction. We can then re-map the streamfunction into physical space and use (10.39) to calculate the radial and vertical velocities. Figure 10.8 shows the streamfunction and associated radial and vertical velocities determined this way, for the final state of the simulation shown in figures 10.3-10.6 and corresponding to the gradient wind and temperature perturbation shown n Figure 10.7.

Above the boundary layer, the streamlines follow angular momentum surfaces (not shown here) because the flow is in a nearly steady state. Inflow is confined to the boundary layer, and there are two rather distinct outflows: one strongly concentrated in the outward-sloping eyewall, and a second, broader region above the boundary layer. This is related to the unrealistic Ekman suction into the boundary layer in the outer region and would likely not be present in a vortex in which the descent rate is limited by radiative cooling. A broad region of ascent gives way to intense updrafts in the eyewall near the radius of maximum winds, and along the inner edge of the eyewall front so prominent in Figure 10.7a.

it is instructive to compare these two-dimensional, steady fields to the fields produced by a full-physics, nonhydrostatic axisymmetric model for both dry and moist tropical cyclone simulations as shown in Figure 9.13. Because we have assumed moist (rather than dry) adiabatic lapse rates along M surfaces, the mapping from absolute temperature to altitude as a vertical coordinate is more informative in the moist case. For this reason, the azimuthal velocity 10.7a more nearly resembles the full-physics moist equivalent shown in Figure 9.13b than it does the dry solution shown in Figure 9.13a. On the other hand, the radial velocity of this idealized model does not strongly resemble those of either the moist or dry full-physics simulations (Figures 9.13 c and d).

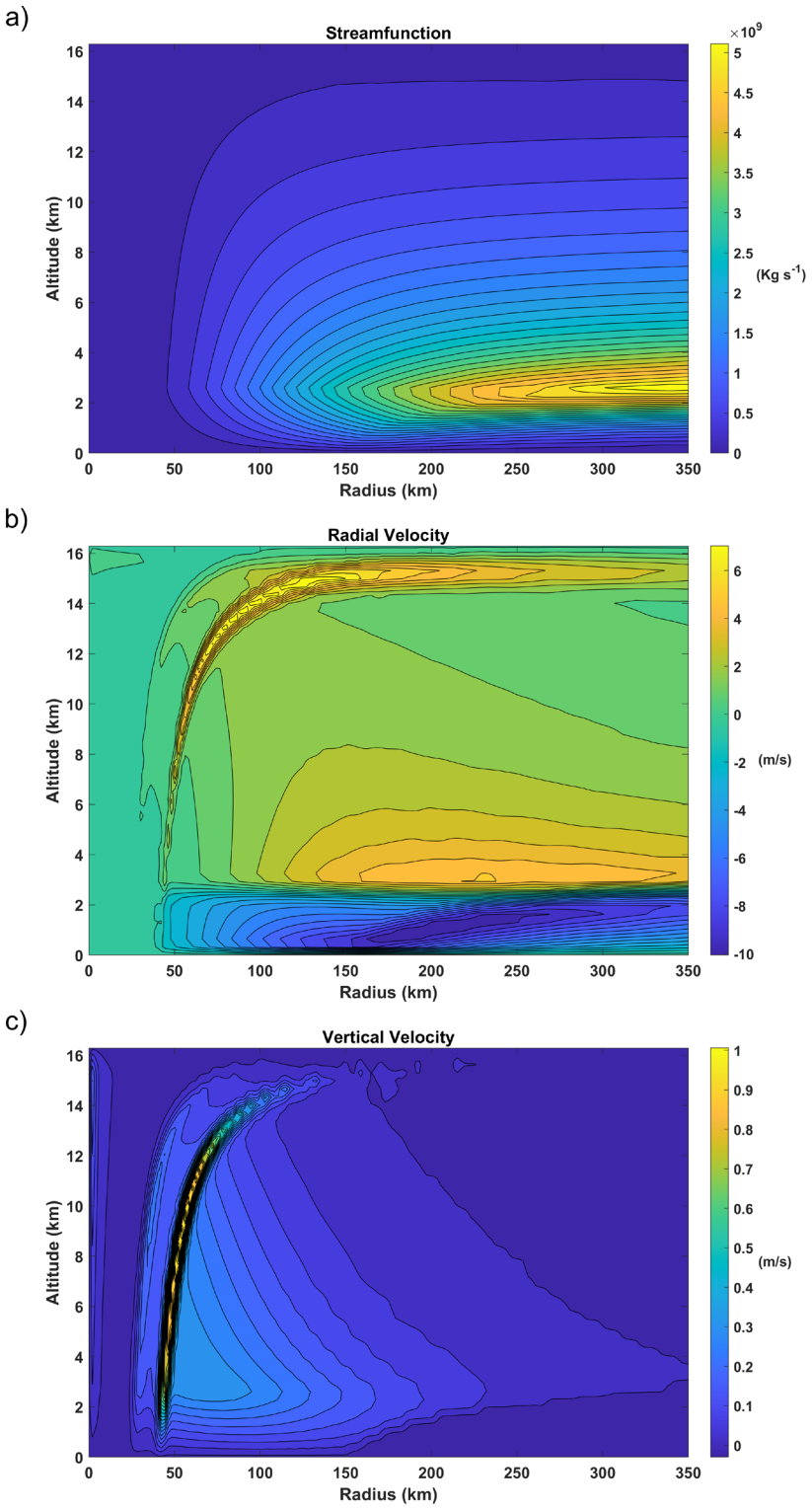


Figure 10.8: As in Figure 10.7 but showing the mass streamfunction (a), radial velocity (b), and vertical velocity (c) at the final time of the simulation.

10.2 Full-physics axisymmetric models

While the simple model described above is useful for understanding the basic physics of tropical cyclone intensification after the genesis stage, it relies on the validity of a set of approximations, including gradient wind and hydrostatic balance, an assumption of slantwise moist convective neutrality, and the assumption that all descent (as well as ascent) is moist adiabatic. This last assumption is particularly egregious as we know that much of the descent is dry and occurs under conditions of dry static stability. In addition, we have to make assumptions about boundary layer turbulence, including the depth of the boundary layer, and assume that turbulence in the outflow has the effect of holding constant a bulk Richardson Number.

Many of these assumptions can be discarded by integrating instead the full Navier-Stokes equations together with the first law of thermodynamics, mass continuity, and the equation of state, perhaps augmented by solving equations representing radiative transfer. These are integrated over a regular grid in radius-altitude space, retaining the assumption of axisymmetry. Turbulence, which we have seen is essential to tropical cyclone physics, must be parameterized. While it is today possible to run such simulations at high enough resolution to resolve the most important turbulent eddies, this is not really feasible in an axisymmetric framework, which forces turbulence to cascade energy from the smallest to the largest scales (Fjørtoft 1953). The formulation of turbulence must also handle the all-important surface fluxes of enthalpy and momentum, and most models use versions of the same aerodynamic flux formulae that we have employed in our theoretical and simple numerical models. One large problem, which we shall return to in Chapter 13, is that we do not yet understand how to formulate surface fluxes in very high wind regimes, where the surface is highly perturbed and much of the flux may be mediated by sea spray.

In most meteorological applications, it is assumed that turbulence acts mostly to redistribute conserved quantities in the vertical direction. Any horizontal fluxes are usually retained to ensure numerical stability of the integration. But, as we have seen, the eyewall of tropical cyclones is strongly frontogenetical and this would cause severe numerical difficulties were it not for horizontal diffusion. Thus, the representation of horizontal turbulent diffusion plays a nontrivial role in numerical simulation of tropical cyclones. The “triple point” where the eyewall front meets the ocean surface, is a particularly challenging place for numerical simulation, as complex turbulent processes in a strongly rotating fluid, causing strong vertical and radial diffusion, are in play with a highly agitated sea surface generating copious sea spray. This is a highly challenging aqua incognita in tropical cyclone physics.

Although moist convection is usually simulated explicitly in contemporary full-physics, axi-symmetric simulations, it is not resolved well and, perhaps more problematically, is being forced into an axisymmetric straight jacket, leading to possibly artificial upscale turbulent cascades. In nature, moist convection is usually fully three-dimensional.

Except in the special case of dry tropical cyclones, we also need to represent the cloud microphysical processes that govern the formation and possible re-evaporation of precipitation. These processes are important because they determine the water loading of clouds, which is an important contribution to their buoyancy, and also govern the strength of downdrafts forced by re-evaporation of condensate. The ice phase can also be important. For example, while vapor may condense into the supercooled liquid phase at temperature considerably below freezing, ice melts at temperatures at or above freezing, so the latent heat of fusion is released at

temperatures systematically lower than the temperatures at which it is absorbed as ice melts. This contributes negatively to the tropical cyclone heat engine.

The results of running experiments with the non-hydrostatic, axisymmetric model of Rotunno and Emanuel (1987) are scattered throughout this volume. This model, which is coded in FORTRAN, is available at https://zenodo.org/doi/10.5281/zenodo.10957525 together with scripts for analyzing the output. A more advanced model, which can be run either in axisymmetric or fully three-dimensional geometry, is also freely available from its developer, George Bryan[2].

10.3 Secondary eyewalls and eyewall replacement cycles

The existence of secondary eyewalls in tropical cyclones was recognized as early as the late 1950s (Fortner 1958) and by the early 1980s researchers had shown that the outer, or secondary eyewall, usually contracts over time and replaces the inner eyewall (Willoughby et al. 1982). Secondary eyewalls are clearly visible in radar images (e.g. Figure 8.23) and often in satellite imagery (Figure 10.9).

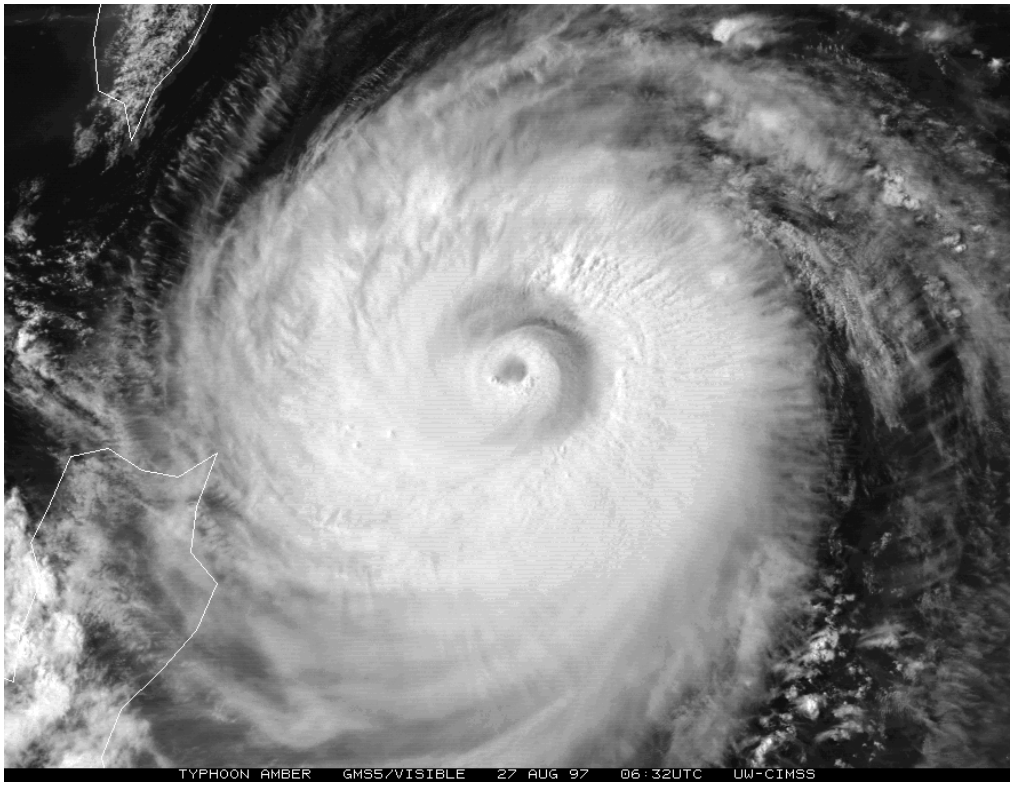


Figure 10.9: Geostationary Meteorological Satellite-5 (GMS-5) visible image of Typhoon Amber on 27 Aug 1997, showing a double eyewall structure.

[2] Go to https://www2.mmm.ucar.edu/people/bryan/cm1/

More recently, high-resolution, three-dimensional wind fields have been mapped in tropical cyclones using Doppler radar mounted on research aircraft (Marks and Houze 1984). Figure 10.10 shows a two-dimensional snapshot of winds and radar reflectivity in Hurricane Rita of 2005. Strong, slantwise updrafts are associated with each eyewall, though the outer eyewall is a bit shallower than the primary.

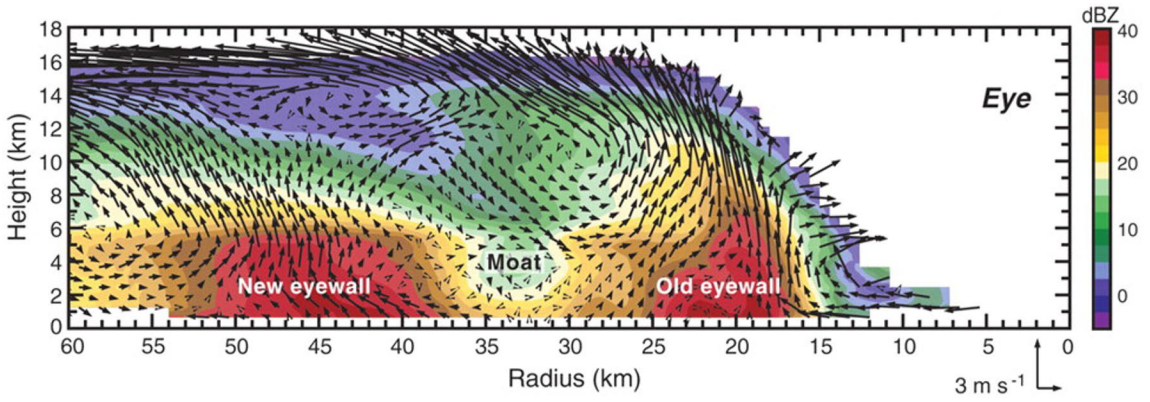


Figure 10.10: Vertical cross-section through Hurricane Rita of 2005, as surveyed by the NCAR Electra Doppler Radar (ELDORA) mounted on NRL's P-3 aircraft. The cross-section extends 60 km out from the storm center located at the right edge of the figure, and from the surface to 18-km altitude. The color shades show the radar reflectivity. The black arrows show the flow in the radius–height plane, with the scale given at the lower right. The outer reflectivity maximum and associated upward motion reveal a secondary eyewall that will move inward and eventually replace the primary eyewall. A reflectivity minimum with downward motion is evident between the two eyewalls.

Concentric eyewall cycles are associated with profound fluctuations in the maximum wind speed and central pressure of tropical cyclones, and are particularly common in intense storms. Hawkins and Helveston (2004) performed a global climatology of concentric eyewalls based on passive microwave imagery and found that of tropical cyclones whose maximum wind attained at least 120 kt, 40% of Atlantic and 80% of western North Pacific storms experienced at least one episode of concentric eyewalls. Kossin and Sitkowski (2009) showed that secondary eyewalls are more likely to form when the mid-tropospheric humidity is high, there is not much vertical shear of the environmental horizontal wind, and the ocean mixed layer is relatively deep. There is some evidence that secondary eyewalls are more likely to form in tropical cyclones with broad wind fields and to lead to a further expansion of the circulation (Huang et al. 2008; Rozoff et al. 2012).

At the time of this writing, there is no clear scientific consensus on the physical cause of secondary eyewalls and eyewall replacement, though a number of mechanisms have been proposed. There remains some question about whether the causes are external to the cyclone or result from some internal mechanism. For example, Kuo et al. (2004) proposed that the formation of secondary eyewalls could result from the barotropic interaction between the tropical cyclone and neighboring vorticity maxima, while Nong and Emanuel (2003) hypothesized that interactions of the high-level circulation with upper tropospheric disturbances could trigger secondary eyewall formation. On the side of internal processes, Abarca and Corbosiero (2011) showed, using a very high resolution, three-dimensional numerical model, that the development of secondary eyewall occurs near the stagnation radius for Rossby waves that exist on the strong radial gradients of potential vorticity in the inner cores of tropical cyclones.

There is also some disagreement about whether the development of secondary eyewalls might involve fundamentally asymmetric physics, or whether the basic physics can operate in axisymmetric frameworks. Some, but by no means all, integrations of axisymmetric numerical models exhibit spontaneous development of eyewall replacement cycles. An excellent example of such a spontaneous development within an axisymmetric simulation was presented by Shivamoggi (2022). Whether or not such phenomena in axisymmetric simulations have anything to do with secondary eyewalls in nature, it should be possible and would seem prudent to understand the physics operating in such models.

One interesting possibility is that the steady-state, axisymmetric solutions of the kind described in Chapter 9 are unstable. A hypothetically infinitesimal perturbation of boundary layer vorticity outside the primary eyewall, for example, may amplify owing to feedbacks among the Ekman flow, vorticity, and moist convection, as argued by Kepert (2013). Nonlinear effects involving, especially, the vertical advection of horizontal velocity have been argued to be fundamental to the formation of secondary eyewalls (Abarca et al. 2015). Especially in slab boundary layer models, shocks may develop owing to radial advection of radial velocity (Williams et al. 2013).

It may be that irreversible processes associated with moisture are essential to secondary eyewall formation. To begin with, secondary eyewalls have not been observed so far in simulations of dry tropical cyclones (Cronin and Chavas 2019), though there have been very few studies of these at the time of this writing. Some properties of moist cyclones with irreversible fallout of precipitation have no analogs in dry cyclones. For example, subsidence in the outer regions can cause a temperature inversion at low levels, which could allow CAPE to build up in the inflow. When, following inflowing air, the subsidence relaxes and is replaced by mean ascent, the inversion dissipates and the instability could be released, driving secondary eyewall formation (Abarca and Montgomery 2013).

Figure 10.11 shows an example of secondary eyewall formation and replacement in a simulation from the axisymmetric version of Cloud Model 1 (CM1; Bryan and Fritsch 2002).

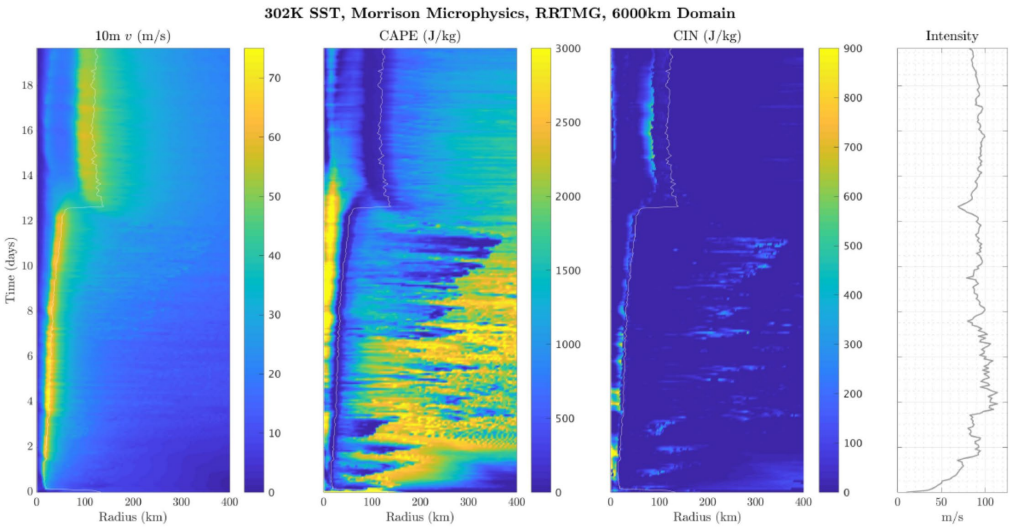


Figure 10.11: Hovmöller (time-radius) plots of near-surface azimuthal wind (left), Convective Available Potential Energy (CAPE; center), and Convective Inhibition (CIN; right) in an axisymmetric simulation using CM1. Time series of the maximum near-surface azimuthal wind is shown at far right.

The radius of maximum winds of the primary eyewall expands outward over time as the vortex intensifies. A secondary wind maximum develops at a radius of a few hundred km at about 12 days into the simulation, moves rapidly inward, and replaces the primary wind maximum. Note that the radius of maximum winds is appreciable larger after the replacement and indeed the whole wind field is broader.

The Hovmöller diagram of CAPE is interesting. Both the primary and secondary eyewalls are loci of zero CAPE, consistent with convective neutrality. Episodes of quite large values of CAPE are common outside of about 100 km radius during the first ten days of the simulation. These take the form of inward-propagating bands of high CAPE, sometimes terminating abruptly, giving way to small or zero values likely associated with cold pools originating in strong convective downdrafts. The onset of the secondary eyewall is associated with an abrupt cessation of strong convective activity outside the eyewall, as marked by the absence of high CAPE values and strong convective downdrafts. Episodes of convective inhibition also more or less cease after secondary eyewall formation. Whatever the cause of secondary eyewall development, it is associated in this simulation with a strong decline in deep convection in the outer regions of the cyclone.

Shivamoggi (2022) proposed that secondary eyewalls are the means by which tropical cyclones whose size is smaller than an equilibrium size transition to larger diameter. To help test this idea, she attempted to fit the matched inner-outer theoretical solution described in Chapter 9, section 9.3.2 to the radial profiles of azimuthal wind at various different times during CM1 simulations such as that used to construct figure 10.11. The result is shown in Figure 10.12. Before the eyewall replacement occurs, the match is not very good, with overly strong winds just outside the radius of maximum winds, and winds that are too weak in the far outer region. After secondary eyewall replacement occurs, around day 12, the agreement between the matched theoretical model and the actual azimuthal wind profile of the axisymmetric model is quite good, suggesting that the cyclone has transitioned to a state much closer to a plausible long-term equilibrium.

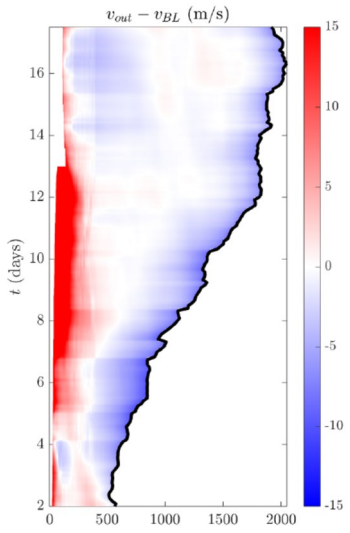


Figure 10.12: Difference (m/s) between the simulated azimuthal wind (left panel of Figure 10.111) and best fit theoretical wind profile as described in section 9.3.2. The tick black curve represents the outer radius as determined from the theoretical model.

Figure 10.12 shows that before eyewall replacement, the actual wind profile between the radius of maximum winds and the outer radius is steeper than the theoretical profile, which represents the azimuthal wind profile needed such that Ekman suction balances radiative cooling. The fact that the actual profile is steeper suggests that the Ekman suction is stronger than that needed to balance radiative cooling just above the boundary layer. Perhaps this leads to the formation of transient temperature inversions, as suggested by CIN build-up (right panel of Figure 10.11), and thereby to accumulation of CAPE. Whether by coincidence or not, the formation of the secondary eyewall is preceded by a strong convective downdraft evident in the CAPE field of Figure 10.11. Once the secondary eyewall takes over from the primary, the whole wind field is larger and the azimuthal wind profile closely matches theoretical expectations, suggesting that Ekman suction balances radiative subsidence in the outer region, preventing the buildup of CIN and CAPE.

Fang and Zhang (2012) compared two simulations of idealized tropical cyclones using a convection-permitting three-dimensional model. Both were performed in quiescent environments, but one was on an *f*-plane (constant Coriolis parameter) while the other was on a $\beta-$plane (constant meridional gradient of Coriolis parameter). As shown in the next chapter, Figure 11.23, the first develops only a primary eyewall while the second also develops a secondary eyewall through an asymmetric process we will examine in Chapter 11.

The physics of secondary eyewall formation remains an active field of research, the fruit of which may also aid in short-range prediction of secondary eyewall formation and replacement cycles.

10.4 Decay over land and cold water

When tropical cyclones move over land or oceanic regions of small potential intensity, they generally decay rather rapidly unless they interact constructively with potential vorticity anomalies in their environment or receive appreciable enthalpy flux from standing water and/or sufficiently hot soils with high heat conductivity (see Chapter 12, section 12.5 and Chapter 13, section 13.5). Here we consider the decay of tropical cyclones when neither of these influences is present. For clarity, we also assume that the storm is not subject to other environmental influences, such as wind shear, which might cause it to decay even over warm ocean water.

When the eyewall of the tropical cyclone moves over land, the enthalpy flux that sustains the storm is cut off, usually quite abruptly. The boundary layer moist entropy in the inner core begins a steep decline owing to the inward radial advection of environmental air in the boundary layer; this flushes out the central core of elevated moist entropy that had resulted from sea surface fluxes. The increased surface roughness leads to increased inflow, which may briefly enhance aera-integrated precipitation before the circulations decreases too much.

At first glance, it might seem that the rapid decrease in inner core boundary layer entropy would stabilize the atmosphere there and lead to a cessation of deep moist convection. This is not observed. The strong Ekman inflow maintains upward motion in the inner core, but because of the reduction of boundary layer moist entropy, the adiabatic cooling associated with the ascent cannot be completely compensated by latent heat, and so the inner core cools through the depth of the circulation. If we assume that slantwise moist neutrality is approximately

maintained, then, as before, the strength of the primary circulation is determined at each time mainly by the distribution of boundary layer moist entropy.

If we continue to assume slantwise moist neutrality and boundary layer control of the storm dynamics, we can apply the simple model developed in section 10.1 but starting with a mature cyclone and setting $C_k = 0$ to mimic landfall. Starting with (10.17) and following the subsequent development but with vanishing surface enthalpy flux, we arrive at

$$\delta \frac{\partial}{\partial \tau} V_g^2 = 3 C_D V_g^2 \frac{T_o}{T_b} M \frac{\partial V_g}{\partial M}. \tag{10.45}$$

We assume that this applies in the moist saturated region right up to the inner edge of the eyewall. Since (10.45) is first-order in the radial coordinate M, we need only one radial boundary condition and we will apply that at the outer limit of the vortex. We will not solve for the part of the solution inside the eyewall but assume that, broadly speaking, the eye spins down along with the eyewall. For simplicity, we hold T_o / T_b constant in radius and in time.

To solve (10.45), we first make the substitution

$$x \equiv 1 / V_g, \tag{10.46}$$

whereupon (10.45) becomes

$$\delta \frac{\partial x}{\partial \tau} = \frac{3}{2} C_D \frac{T_o}{T_b} \frac{\partial \ln(x)}{\partial \ln(M)}. \tag{10.47}$$

This has a separable solution of the form

$$x = \left(c_1 + c_2 \tau \right) Y(M), \tag{10.48}$$

where c_1 and c_2 are constants and $Y(M)$ satisfies

$$\frac{1}{Y^2} \frac{dY}{d \ln(M)} = \frac{2}{3} \frac{\delta c_2}{C_D} \frac{T_b}{T_o}, \tag{10.49}$$

whose solution is

$$\frac{1}{Y} = \frac{1}{Y_0} - \frac{2}{3} \frac{\delta c_2}{C_D} \frac{T_b}{T_o} \ln\left(\frac{M}{M_0} \right), \tag{10.50}$$

where Y_0 and M_0 are the values of Y and M at the inner edge of the eyewall.

We substitute this into (10.48) and result into (10.46), and apply the boundary condition that the gradient wind vanish at some outer radius where and the initial condition that $V_g = V_0$ at the inner edge of the eyewall. The result is

$$V = \frac{V_0}{1+\chi\left(\tau-\tau_0\right)}\left[1-\frac{\ln\left(M/M_0\right)}{\ln\left(M_{max}/M_0\right)}\right], \tag{10.51}$$

where τ_0 is the initial time and

$$\chi \equiv \frac{3}{2}\frac{T_o}{T_b}\frac{C_D}{\delta}\frac{V_0}{\ln\left(M_{max}/M_0\right)}. \tag{10.52}$$

(An expression similar to (10.51) was derived by Chen and Chavas (2021).) The wind speed decays algebraically (not exponentially) with a decay time scale on the order of $\delta / C_D V_0$. For a momentum mixing depth δ of 4 km, a drag coefficient of 1.5×10^{-3}, a starting wind speed of $50\ ms^{-1}$, an outer radius of 500 km, a radius of maximum winds of 25 km, $T_o/T_b = 2/3$, and $f = 5\times10^{-5}s^{-1}$, this algebraic decay time scale is about a day. (Doubling the drag coefficient to a value more typical of land, this would be reduced to 12 hours.)

The relation (10.51) gives the gradient wind outside the inner eyewall as a function of time and angular momentum. Once the solution is found, the definition of angular momentum can be inverted to find the physical radius r and the solution presented as a function of radius and time.

An example of this solution, using the parameter values mentioned above, is shown in Figure 10.13. Here we assume that the radial profile of gradient wind is linear between the storm center and the inner eyewall. As the wind decays, the radius of maximum winds expands outward. The wind profile has a shape different from the outer wind profile of the steady-state theoretical model.

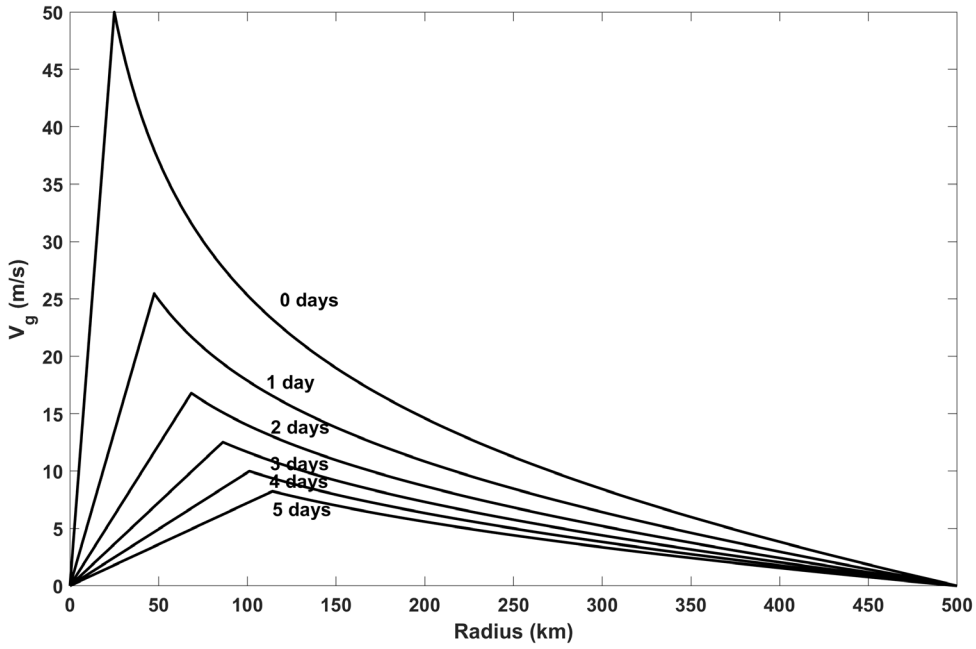


Figure 10.13: Radial profiles of gradient wind as solutions of (10.51) mapped into physical space at the initial time and for each day up to 5 days. It has been assumed that the wind profile inside the radius of peak winds is linear.

It is difficult to test simple model against historical observations. The best-track tropical cyclone database is mostly at 6-hour time resolution, which is not enough to distinguish the predicted algebraic decay from other plausible functions. Chen and Chavas (2021) compared the predictions of an expression similar to (10.51) to numerical simulations using the axisymmetric version of the CM1 model (Bryan and Fritsch 2002). Here we do a simple comparison to a single integration of the model of Rotunno and Emanuel (1987) (available at https://zenodo.org/doi/10.5281/zenodo.10957525), shown in Figure 10.14. The value of χ used here is about $1\,day^{-1}$. The algebraic decay (10.51) fits the simulated peak wind speed very well from landfall out to about 9 days. After that, the peak winds in the storm are found above the boundary layer.

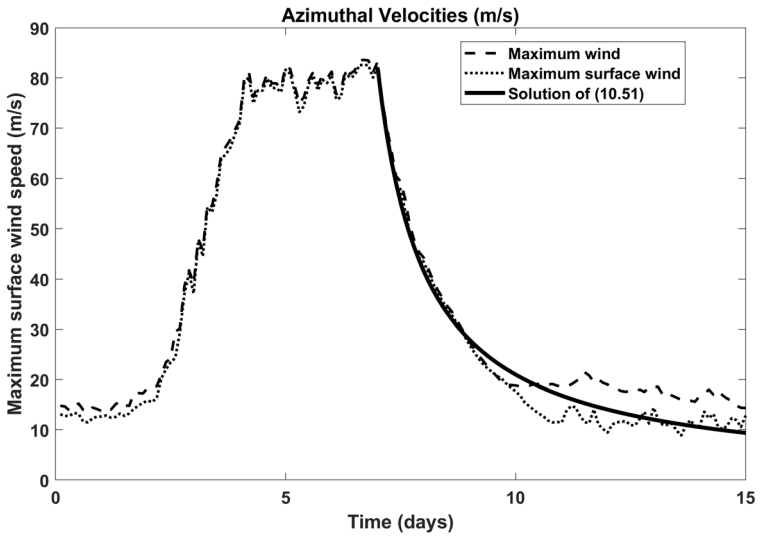


Figure 10.14: Maximum domain-wide wind speed (dashed) and maximum wind speed at lowest model level (dotted) in a simulation with the axisymmetric clouds-permitting model of Rotunno and Emanuel (1987). At 7 days into the simulation, the surface enthalpy exchange coefficient is set to zero, but the drag coefficient remains at its over-ocean value of 1.5×10^{-3}. *The solid line graphs (10.51) at the radius of maximum winds, using a value of* χ *of* $0.97\ day^{-1}$.

While surface winds decay quite rapidly after landfall (and more so when increases in surface roughness are accounted for), storm total rainfall usually diminishes less rapidly and may even increase if the cyclone interacts with steep topography and/or stalls, increasing the duration of heavy rains. Some of the most serious freshwater flooding disasters were caused by tropical cyclones whose winds had diminished to less than tropical storm strength. Thus, as tropical cyclones move inland, wind becomes progressively less of concern while flooding can become the more important risk.

When tropical cyclones move rapidly from warm to cold sea surfaces, the surface heat flux may change sign and the boundary layer may become stable. When this happens, the surface winds may diminish rapidly, exposing the winds at the top of the boundary layer to less surface drag then before the transition. In this situation, the surface winds may slow down far more rapidly than the winds above the stable boundary layer. As a rough guide, we might assume that shear-driven turbulence holds the bulk Richardson Number to a critical value. In this case, the shear across the depth h of the boundary layer would be given by $\Delta U = Nh/\sqrt{Ri_c}$, where N is the buoyancy frequency and Ri_c is the critical bulk Richardson Number. For example, if $N = 0.03$, $h = 500\,m$, and $Ri_c = 1$, $\Delta U = 15\,ms^{-1}$. Very large boundary layer shears were observed, for example, in the Great New England Hurricane of 1938, which passed from the warm waters of the Gulf Stream to the very cold water south of New England in just a few hours. Winds at the top of Blue Hill, about 200 m above sea level just south of Boston, exceeded $50\,ms^{-1}$ while remaining at or below about $15\,ms^{-1}$ in Boston. Likewise, winds of $35\,ms^{-1}$ were recorded atop the Empire State Building in New York City (about 400 m altitude), while remaining below $15\,ms^{-1}$ in Central Park.

10.5. Observed statistics of intensification and dissipation

Many factors influence the intensification and dissipation of tropical cyclones, including potential intensity, wind shear, dry air intrusions, and the character of the underlying surface. While the land surface characteristics, such as surface roughness, change slowly if at all over time, sea surface temperature and potential intensity vary mostly on season and interannual time scales and environmental wind and humidity can change appreciable on scales as short as a day or even a few hours. All of these factors conspire to yield a broad spectrum of intensification and dissipation rates. To quantify this spectrum, we examine historical tropical cyclone data (Knapp et al. 2010) for the North Atlantic region over the period 1979-2020, a period of relatively high quality and quantity of observations. For tropical cyclones that are not making landfall, we calculate the difference between the maximum surface wind speeds between each 6-hour data point, using a total of 9,882 6-hour point pairs. From these we calculate the cumulative probability of both intensification and decay. (For details, see Kowch and Emanuel (2015).) The results are displayed in Figure 10.15.

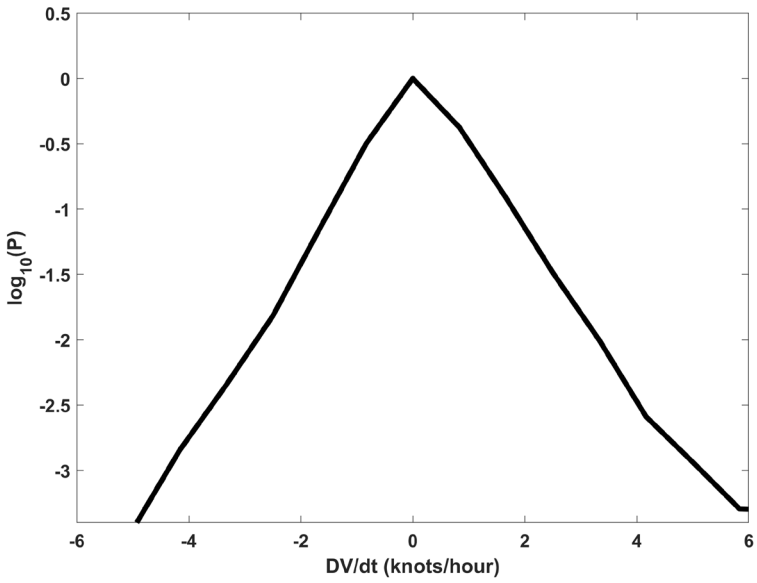


Figure 10.15: The base-10 logarithm of the cumulative probability of intensification/dissipation of North Atlantic tropical cyclones during the period 1979-2020. See text for details.

The logarithm of the probability density on either side of zero is practically a straight line, so that the probability density itself falls of exponentially with the magnitude of the intensification and dissipation rates. There is no obvious break anywhere in the distribution. The distribution is, however, noticeably skewed toward higher positive values (intensification). Had we included cyclones over land, the distribution would have been skewed negative.

Tropical cyclone forecasters are naturally concerned about rapidly intensifying storms and have defined “rapid intensification” as an increase of wind speed of more than 30 knots in 24 hours (at least 1.25 knots per hour sustained for day or more). There is a large and growing literature on the environmental or storm-scale factors that might lead to rapid intensification. But the lack of any clear break in the probability distribution of intensification rates (Figure 10.15) suggests that there is merely a continuum of intensification rates with no suggestion of any special

physics operating at high rates of intensification. Nevertheless, the exponential nature of tropical cyclone intensification probability presents a theoretical challenge; meeting that challenge would no doubt improve our overall understanding of tropical cyclone physics.

References

Abarca, S. F., and K. L. Corbosiero, 2011: Secondary eyewall formation in WRF simulations of Hurricanes Rita and Katrina (2005). *Geophy Res Lett*, **38**, https://doi.org/10.1029/2011GL047015.

——, and M. T. Montgomery, 2013: Essential dynamics of secondary eyewall formation. *J Atmos Sci*, **70**, 3216–3230.

Abarca, S. F., M. T. Montgomery, and J. C. McWilliams, 2015: The azimuthally averaged boundary layer structure of a numerically simulated major hurricane. *J. Adv. Model. Earth Syst.*, **7**, 1207–1219, https://doi.org/10.1002/2015MS000457.

Bryan, G. H., and J. M. Fritsch, 2002: A benchmark simulation for moist nonhydrostatic numerical models. *Mon Wea Rev*, **130**, 2917–2928.

——, and R. D.-:10. 1175/2009JAS3038. 1 Rotunno, 2009: Evaluation of an analytical model for the maximum intensity of tropical cyclones. *J Atmos Sci*, **66**, 3042–3060.

Chen, J., and D. R. Chavas, 2021: Can existing theory predict the response of tropical cyclone intensity to idealized landfall? *J. Atmospheric Sci.*, **78**, 3281–3296, https://doi.org/10.1175/JAS-D-21-0037.1.

Cronin, T. W., and D. R. Chavas, 2019: Dry and Semidry Tropical Cyclones. *J. Atmospheric Sci.*, **76**, 2193–2212, https://doi.org/10.1175/jas-d-18-0357.1.

Fang, J., and F. Zhang, 2012: Effect of beta shear on simulated tropical cyclones. *Mon. Weather Rev.*, **140**, 3327–3346, https://doi.org/10.1175/MWR-D-10-05021.1.

Fjørtoft, R., 1953: On the changes in the spectral distribution of kinetic energy for twodimensional, nondivergent flow. *Tellus*, **5**, 225–230, https://doi.org/10.1111/j.2153-3490.1953.tb01051.x.

Fortner, L. E., 1958: Typhoon Sarah, 1956. *Bull Amer Meteor Soc*, **39**, 633–639.

Hawkins, J. D., and M. Helveston, 2004: Tropical cyclone multiple eyewall characteristics. 26th Conference on Hurricanes an Tropical Meteorology, Miami, Amer. Meteor. Soc.

Hoskins, B. J., 1975: The geostrophic momentum approximation and the semi-geostrophic equations. *J. Atmospheric Sci.*, **32**, 233–242, https://doi.org/10.1175/1520-0469(1975)032<0233:TGMAAT>2.0.CO;2.

Hoskins, B. J., and F. P. Bretherton, 1972: Atmospheric frontogenesis models: Mathematical formulation and solution. *J Atmos Sci*, **29**, 11–37.

——, M. E. McIntyre, and A. W. Robertson, 1985: On the use and significance of isentropic potential vorticity maps. *Quart J Roy Meteor Soc*, **111**, 877–946.

Huang, Y. J., M. T. Montgomery, and C. C. Wu, 2008: Concentric Eyewall Formation in Typhoon Sinlaku (2008). Part II: Axisymmetric Dynamical Processes. *J Atmos Sci*, **69**, 662–674.

Kepert, J. D., 2013: How does the boundary layer contribute to eyewall replacement cycles in axisymmetric tropical cyclones? *J. Atmospheric Sci.*, **70**, 2808–2830, https://doi.org/10.1175/jas-d-13-046.1.

Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, 2010: The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone best track data. *Bull Amer Meteor Soc*, **91**, 363–376.

Kossin, J. P., and M. Sitkowski, 2009: An objective model for identifying secondary eyewall formation in hurricanes. *Mon Wea Rev*, **137**, 876–892.

Kowch, R., and K. Emanuel, 2015: Are special processes at work in the rapid intensification of tropical cyclones? *Mon Wea Rev*, **143**, 878–882, https://doi.org/10.1175/mwr-d-14-00360.1.

Kuo, H.-C., L.-Y. Lin, C.-P. Chang, and R. T. Williams, 2004: The formation of concentric vorticity structures in typhoons. *J Atmos Sci*, **61**, 2722–2734.

Marks, F. D., Jr., and R. A. Houze Jr., 1984: Airborne Doppler radar observations in Hurricane Debby. *Bull Amer Meteor Soc*, **65**, 569–582.

Montgomery, M. T., M. E. Nicholls, T. A. Cram, and A. B. Saunders, 2006: A vortical hot tower route to tropical cyclogenesis. *J. Atmospheric Sci.*, **63**, 355–386, https://doi.org/10.1175/jas3604.1.

Nong, S., and K. Emanuel, 2003: Concentric eyewalls in hurricanes. *Quart J Roy Meteor Soc*, **129**, 3323–3338.

Rotunno, R., and K. A. Emanuel, 1987: An air-sea interaction theory for tropical cyclones. Part II. *J Atmos Sci*, **44**, 542–561.

Rozoff, C. M., D. S. Nolan, J. P. Kossin, F. Zhang, and J. Fang, 2012: The Roles of an Expanding Wind Field and Inertial Stability in Tropical Cyclone Secondary Eyewall Formation. *J. Atmospheric Sci.*, **69**, 2621–2643, https://doi.org/10.1175/jas-d-11-0326.1.

Shivamoggi, 2022: Secondary eyewall formation as a response to evolving tropical cyclone wind structure. Massachusetts Institute of Technology, 125 pp. https://dspace.mit.edu/handle/1721.1/143163.

Williams, G. J., R. K. Taft, B. D. McNoldy, and W. H. D.-:10. 1002/jame. 20028 Schubert, 2013: Shock-like structures in the tropical cyclone boundary layer. *J. Adv. Model. Earth Syst.*, **5**, 338–353.

Willoughby, H. E., J. A. Clos, and M. G. Shoreibah, 1982: Concentric eyes, secondary wind maxima, and the evolution of the hurricane vortex. *J Atmos Sci*, **39**, 395-411.

11. Asymmetric Features: Theory and Numerical Experiments

Up until now we have idealized tropical cyclone as axisymmetric vortices, and that idealization allowed us to make some progress in describing intensification and the mature structure and intensity of tropical cyclones. Some real-world tropical cyclones look very axisymmetric in radar and satellite imagery, but even these storms usually have asymmetric features like spiral rain bands. Nascent cyclones are sometimes highly asymmetric, and interaction of tropical cyclones with non-axisymmetric features of the environment, such as shear of the background flow, shorelines and other topographic features, and other potential vorticity anomalies often introduces asymmetry into even mature storms. We will postpone discussion of asymmetries arising from these external reactions to Chapter 12. Here will focus on three classes of processes that can generate asymmetries: spontaneous symmetry breaking owing to instabilities of the circular vortex, asymmetries arising from storm translation relative to the sea surface, and the effects of the latitudinal variation of the Coriolis parameter.

11.1 Spontaneous Symmetry Breaking

In the absence of environmental asymmetries, axisymmetric vortices appear to be legitimate nonlinear solutions of the known equations. Both theory and axisymmetric models show that such vortices can amplify, maintain quasi-steady states, and decay in ways that are qualitatively and quantitatively consistent with observations. One may ask, however, whether such axisymmetric solutions may be unstable to asymmetric perturbations. (We have discussed, in Chapter 10, whether such solutions might be unstable to axisymmetric perturbations, giving rise to secondary eyewalls and replacement cycles.)

11.1.1 Eyewall mesovortices

Radar and satellite imagery usually exhibits a smooth, nearly circular eye in mature tropical cyclones, but on occasion the inner edge of the eyewall exhibits an array of vortices. There appear to be two classes of eyewall vortices: Mesoscale vortices on scales not much smaller than the eye itself, and somewhat smaller tornado-like vortices. Both of kinds of vortices are evident in Figure 11.1, showing radar reflectivity in Hurricane Harvey as it was making landfall on the coast of southeastern Texas. The first panel shows radar reflectivity on the scale of the storm, as recorded by the NOAA NEXRAD radar at Corpus Christi (station identifier KCRP) in the lower left side of the panel. This is an S-band radar transmitting pulses with a wavelength of about 10 cm. The other panels show images recorded by a “Doppler on Wheels” (DOW), a Doppler radar mounted on a truck and here positioned just to the west of the storm center at the time indicated. DOW transmits pulses at a wavelength of around 3 cm; this shorter wavelength therefore detects smaller particles than does NEXRAD, but at the cost of more attenuation of the signal. The DOW picks up four distinct mesovortices filling the eye and part of the eyewall, and the zoomed in portion of the western eyewall (panel d) shows distinct tornado-scale vortices in the Doppler velocity field.

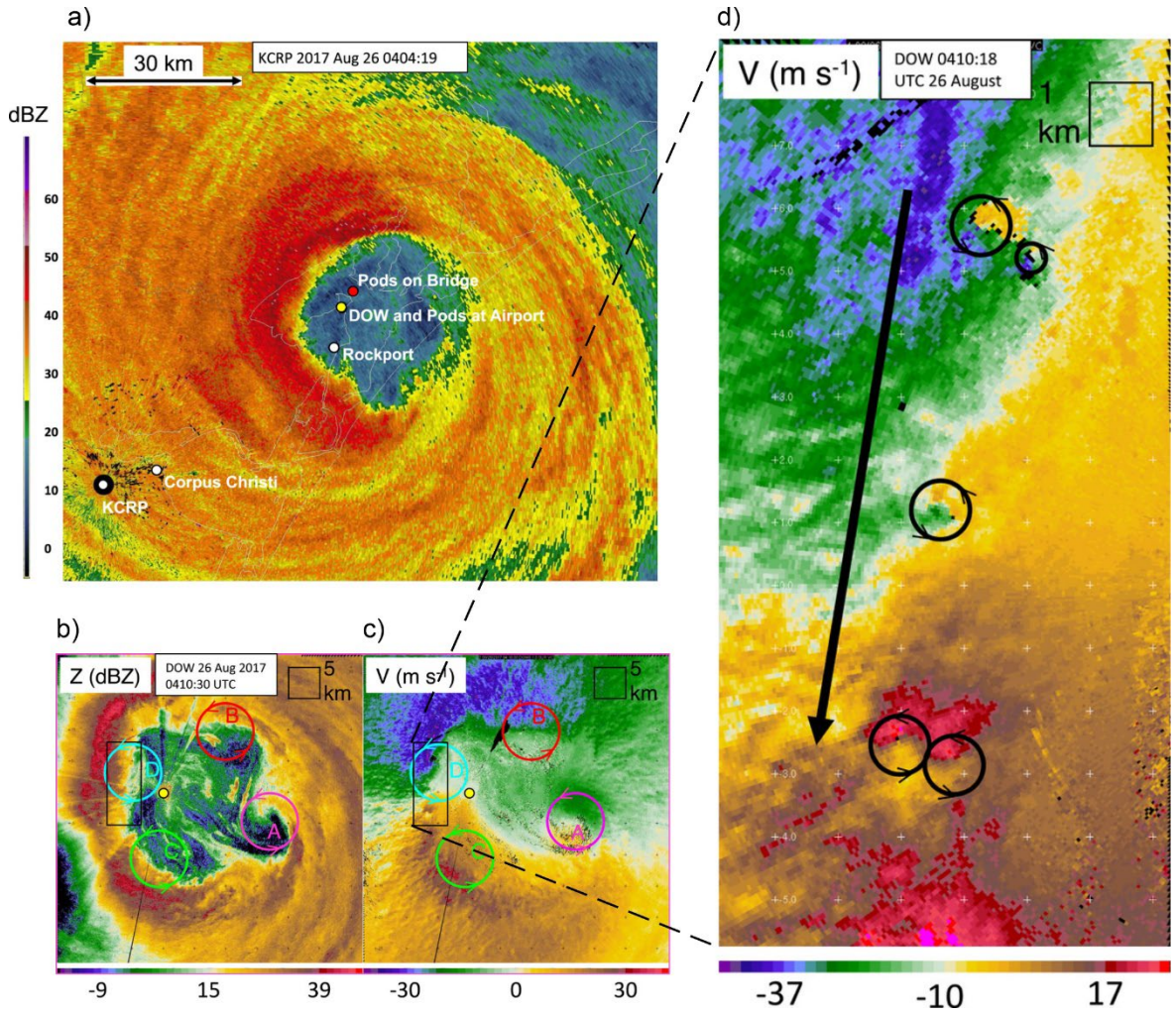


Figure 11.1: Radar returns associated with Hurricane Harvey on 26 August 2017 as it was making landfall near Rockport, Texas, USA. a) Radar reflectivity measured by the National Weather Service KCRP radar at 0404:19 UTC 26 Aug 2017 showing the eyewall of Harvey just after landfall. The radar is located near the lower left corner of the image, and Rockport is in the southwest part of the eye. The yellow dot shows the position of the Doppler on Wheels (DOW) whose data were used to make the radar images in (b)-(d). The DOW radar reflectivity (b) and Doppler velocity (c) at 0410:30 UTC 26 Aug 2017.The 4 colored circles show the locations of the eyewall mesovortices, and the reflectivity scale (dBZ) and Doppler velocity scale (m s^{-1}) are shown at the bottom of the panels. The black rectangle at left-center in b) and c) is magnified into panel d), in which tornado-scale vortices (TSVs) are indicated by black circles. These vortices, on the western eyewall, are translating rapidly southward.

The vortices were shown by Wurman and Kosiba (2018) to be associated with pockets of very high damage in the hurricane.

On 15 September, 1989, a NOAA WP-3D reconnaissance aircraft flew at an altitude of 450 m into the core of rapidly intensifying Hurricane Hugo. The aircraft encountered extreme turbulence in the eyewall and suffered a failure of one of its four engines, descending to an altitude of 270 m before enough fuel had been dumped to gain altitude. Subsequent analysis of the flight-level data by Marks et al. (2008) strongly suggested that the aircraft had flown through a tornado-scale vortex in the eyewall. The vortex had a horizontal scale of less than a kilometer and a peak vorticity of more than $0.1 s^{-1}$. The horizontal wind speed dropped from $82 ms^{-1}$ to

$25\,ms^{-1}$ in 6 seconds of flight time and the aircraft encountered updrafts of more than $20\,ms^{-1}$ and downdrafts as large as $9\,ms^{-1}$.

Mesovortices such as the larger ones evident in Figure 11.1, panels b) and c) have been simulated numerically using full-physics models designed to simulated real-world storms (e. g. Mashiko 2005) and in laboratory experiments (Montgomery et al. 2002), and in recent years, tornado-scale vortices have been simulated using very high resolution models with telescoping meshes (Wu et al. 2018). The structure and intensity of the simulated and laboratory vortices are consistent with observations by radar and aircraft.

In the rest of this section, we will review the extant theoretical understanding of mesovortices and tornado-scale vortices in tropical cyclone eyes. To set the stage, we present a short description of some fundamental waves and instabilities of vortical flow.

11.1.1.1 Rossby waves and associated instabilities

Shear flows with vorticity gradients support Rossby waves, a planetary scale form of which we dealt within our discussion of equatorial waves in Chapter 7. Here we present arguable the simplest flows that support Rossby waves and instabilities that arise from their interaction. This problem was first presented by Lord Rayleigh (1895).

Consider the barotropic flow whose variation in y is illustrated in Figure 11.2. The flow is piecewise continuous and all the meridional gradient of vorticity is concentrated in positive and negative delta functions at $y=+D$ and $y=-D$, respectively. We will formulate linear equations for small perturbations to this background state, solve them in each of the three regions, and match the solutions across the boundaries of the regions at $y=\pm D$.

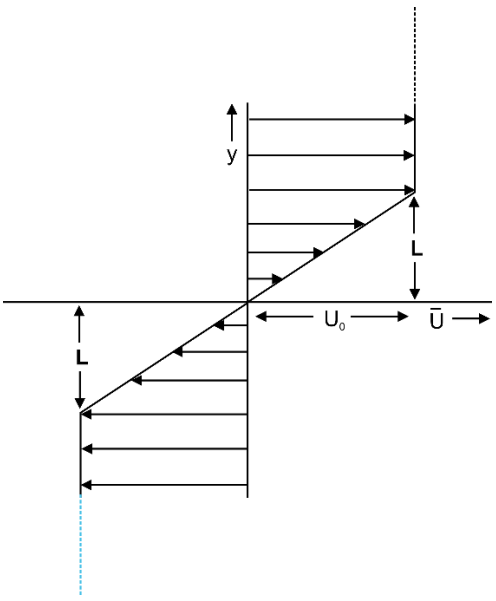


Figure 11.2: The Rayleigh shear flow problem. Flow in the x direction varies linearly in y between x=-L and x=+L. Beyond those limits the flow is constant in y and extends to $y=\pm\infty$.

The linearized momentum equations and the mass continuity equation linearized around this basic state are as follows. Overbars signify the background state shown in Figure 11.2; otherwise, the dependent variables are small amplitude perturbations:

$$\begin{aligned}
&\frac{\partial u}{\partial t}+\bar{U}\frac{\partial u}{\partial x}+v\frac{\partial \bar{U}}{\partial y}=-\alpha_0\frac{\partial p}{\partial x},\\
&\frac{\partial v}{\partial t}+\bar{U}\frac{\partial v}{\partial x}=-\alpha_0\frac{\partial p}{\partial y},\\
&\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}=0.
\end{aligned} \tag{11.1}$$

Here u and v are the westward and northward velocity components, α_0 is the (constant) specific volume, and p is the fluid pressure.

Consider solutions of the form

$$\begin{aligned}
u&=\tilde{u}(y)e^{ik(x-ct)},\\
v&=\tilde{v}(y)e^{ik(x-ct)},\\
p&=\tilde{p}(y)e^{ik(x-ct)},
\end{aligned} \tag{11.2}$$

where c may be complex. Substitution into (11.1) gives

$$\begin{aligned}
&ik(\bar{U}-c)\tilde{u}+\bar{U}_y\tilde{v}=-\alpha_0 ik\tilde{p},\\
&ik(\bar{U}-c)\tilde{v}=-\alpha_0\frac{d\tilde{p}}{dy},\\
&ik\tilde{u}+\frac{d\tilde{v}}{dy}=0.
\end{aligned} \tag{11.3}$$

By cross-differentiating, we can eliminate $\tilde{u}$ and $\tilde{p}$ from (11.3) to arrive at a single O.D.E. in $\tilde{v}$:

$$\frac{d^2\tilde{v}}{dy^2}-\tilde{v}\left(k^2+\frac{\bar{U}_{yy}}{\bar{U}-c}\right)=0. \tag{11.4}$$

Now note that in the interiors of each of the three regions in Figure 11.2, $\bar{U}_{yy}=0$, so (11.4) reduces to

$$\frac{d^2\tilde{v}}{dy^2}-k^2\tilde{v}=0 \tag{11.5}$$

within each region. Now we consider boundary conditions for solving (11.5). First, we impose the condition that the solutions remain bounded at $y=\pm\infty$:

$$\lim_{y\to\pm\infty} \tilde{v} = 0. \tag{11.6}$$

General solutions of (11.5) that satisfy (11.6) in each of the three regions are:

$$\begin{aligned} &\text{Top:} && \tilde{v} = Ae^{-ky} \\ &\text{Middle:} && \tilde{v} = Be^{-ky} + Ce^{ky} \\ &\text{Bottom:} && \tilde{v} = Fe^{ky} \end{aligned} \tag{11.7}$$

Next, we apply boundary conditions at the boundaries separating the regions. There are two fundamental requirements:

a. Fluid displacements must be continuous, and
b. Pressure must be continuous.

Because the basic state velocity, $\overline{U},$ is continuous across the regions, continuity of meridional displacement is equivalent to continuity of $\tilde{v}$ itself. Continuity of pressure implies, through the first and last relations of (11.3), that the quantity

$$(\overline{U} - c)\frac{d\tilde{v}}{dy} - \overline{U}_y \tilde{v} \tag{11.8}$$

is continuous across the region boundaries.

Before solving the full system, let's look at what happens at a single interface in isolation. We can do this by eliminating the bottom region an carrying the middle region down to $y \to -\infty$. In that case, $B = 0$ in (11.7) and applying the matching condition on $\tilde{v}$ and (11.8) yields the dispersion relation

$$c = U_0\left(1 - \frac{1}{kL}\right). \tag{11.9}$$

This is a Rossby wave that lives on the upper interface, at which the vorticity jumps from its negative value in the middle region to zero in the top region. It swims upstream, against the flow, but is Doppler shifted by the mean flow. When $k = L$, the wave is stationary in our chosen coordinate system. (Had we eliminated the top region instead, carrying the middle region to $y \to \infty,$ we would have found the dispersion relation for the lower interface, which is just (11.9) with the sign reversed. Again, this is a Rossby wave swimming upstream but Doppler shifted by the mean flow at the lower interface.))

Now let's solve the full problem by imposing the matching conditions at each of the two boundaries separating the regions yields a relation between the complex phase speed c and $k,$ yielding

$$\left(\frac{2Lkc}{U_0}\right)^2 = (2Lk-1)^2 - e^{-4kL}. \tag{11.10}$$

Note that the phase speed c is either purely real or purely imaginary, depending on the sign of the right side of (11.10): The waves are either neutral and propagating or amplifying (or decaying) and stationary. If c is imaginary, the exponential growth or decay rate is given by kc_i, where c_i is the imaginary part of the phase speed. The solutions to (11.10) are shown in Figure (11.3).

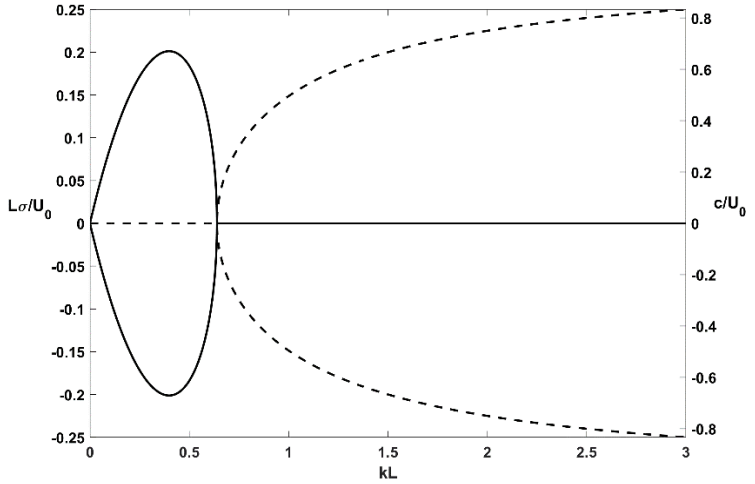


Figure 11.3: Solution of the dispersion relation (11.10). Solid curves show the nondimensional growth rate, read off the left axis, while the dashed curves show the nondimensional phase speed, read off the right axis. Both are functions of the nondimensional wavenumber k.

When the wavenumber is lower than a critical value (wavelength larger than a critical value), exponentially growing solutions exist. The peak nondimensional growth rate is 0.2012 at a nondimensional wavenumber of 0.398. This corresponds to a wavelength of about 8 times the width $(2L)$ of the shear zone. (In the limit that the width of the shear zone vanishes, so that there is a single interface across which the mean flow has a jump, the dimensional growth rate is $2U_0k,$ so that the most rapidly growing modes are very small in scale.)

The streamfunction associated with the most rapidly growing mode is displayed in Figure 11.4. Note that the modes tilt upshear – a universal characteristic of all shear-driven flow instabilities. The instability results from the mutual interaction of Rossby waves that live on the two planes on which there is a delta function of vorticity gradient. That interaction has two important aspects: First, the circulation around each vorticity anomaly on the interface changes the phase speed of the mode on the opposite interface in such a way that the two modes remain phase-locked. That is why their phase speed in this reference frame is zero. The second attribute is that the circulation associated with Rossby wave on each interface serves to amplify the anomaly on the opposite interface. The vorticity anomalies arise from the meridional displacement of the interface, and inspection of Figure 11.4 shows that the circulation associated with each interface vorticity anomaly increases the amplitude of the other.

For exponential instabilities of this kind, Rossby waves must be able to phase lock and amplify each other. Rayleigh and others developed necessary conditions for this to happen in general

barotropic shear flows. Here we review those conditions, which are a special case of the conditions for shear flow instabilities derived in Chapter 7, section 7.3.4.

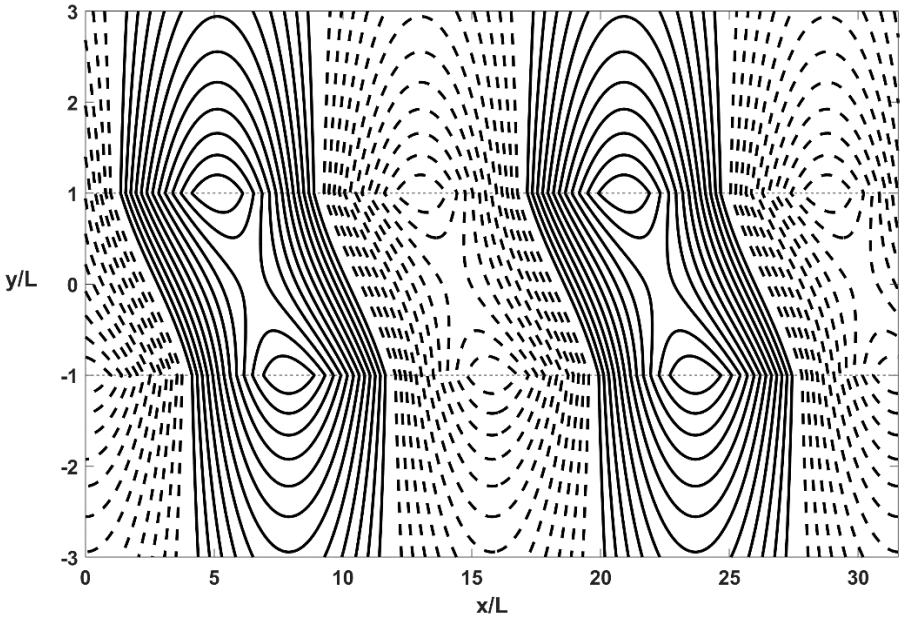


Figure 11.4: Streamfunction associated with the most rapidly growing mode of the Rayleigh shear problem, spanning two wavelengths

First, suppose we have a barotropic flow $\bar{U}(y)$ in which the vorticity gradient is confined to some finite region, so that

$$\lim_{y\to\pm\infty} \frac{d^2\bar{U}}{dy^2} = 0. \tag{11.11}$$

For such a flow, we expect perturbations to vanish at $y = \pm\infty$ since the refractive index for wave propagation, the vorticity gradient, vanishes there.

Now consider modal disturbances to such a flow. These are governed by (11.4), for the meridional structure of the meridional wind. Now multiply (11.4) through by the complex conjugate of $\tilde{v}$, $\tilde{v}^*$, and integrate the result through the whole domain:

$$\int_{-\infty}^{\infty} \left[\tilde{v}^* \frac{d^2\tilde{v}}{dy^2} - |\tilde{v}|^2 \left(k^2 + \frac{\bar{U}_{yy}}{\bar{U}-c} \right) \right] dy = 0 \tag{11.12}$$

Here we have made use of the fact that

$$\tilde{v}\tilde{v}^* = |\tilde{v}|^2,$$

where $|\tilde{v}|$ is the absolute value of $\tilde{v}$. Now the first term in the integrand can be integrated by parts:

$$\int_{-\infty}^{\infty} \tilde{v}^* \frac{d^2\tilde{v}}{dy^2} dy = \int_{-\infty}^{\infty} \frac{d}{dy}\left[\tilde{v}^* \frac{d\tilde{v}}{dy}\right] dy - \int_{-\infty}^{\infty} \left|\frac{d\tilde{v}}{dy}\right|^2 dy. \tag{11.13}$$

The first term on the right can be integrated exactly, but it vanishes because $\tilde{v} \to 0$ as $y \to \pm\infty$. Thus

$$\int_{-\infty}^{\infty} \tilde{v} \frac{d^2\tilde{v}}{dy^2} dy = -\int_{-\infty}^{\infty} \left|\frac{d\tilde{v}}{dy}\right|^2 dy. \tag{11.14}$$

Using (11.14), we may write (11.12) as

$$\int_{-\infty}^{\infty} \left[\left|\frac{d\tilde{v}}{dy}\right|^2 + |\tilde{v}|^2 \left(k^2 + \frac{\bar{U}_{yy}}{\bar{U} - c}\right)\right] dy = 0. \tag{11.15}$$

Remember that c is, in general, complex, so the real and imaginary parts of (11.15) must both be satisfied. The imaginary part of (11.15) is

$$c_i \int_{-\infty}^{\infty} \frac{\bar{U}_{yy}}{|\bar{U} - c|^2} |\tilde{v}|^2 \, dy = 0, \tag{11.16}$$

where c_i is the imaginary part of c, which is positive for growing disturbances.

The relation (11.16) shows that one of two things must be true: Either a) $c_i = 0$, or b) the integral in (11.16) vanishes. Thus, we may conclude the following:

1. A *necessary* condition for instability $(c_i > 0)$ is that $\bar{U}_{yy}$ change sign at least once within the domain. In other words, the mean state vorticity must have an extremum in the domain. But note that even if $\bar{U}_{yy}$ does change sign, this is no *guarantee* that the integral vanishes or that $c_i > 0$. This condition *is not* sufficient for instability.

2. If there is no extremum of vorticity within the domain, $c_i = 0$ and this is therefore a *sufficient* condition for stability.

Points 1 and 2 are saying the same thing in different ways.

Another theorem, due to Fjørtoft (1950), may be derived by looking at the real part of (11.15):

$$\int_{-\infty}^{\infty} \frac{\bar{U}_{yy}(\bar{U} - c_r)}{|\bar{U} - c|^2} |\tilde{v}|^2 \, dy = -\int_{-\infty}^{\infty} \left[\left|\frac{d\tilde{v}}{dy}\right|^2 + k^2 |\tilde{v}|^2\right] dy, \tag{11.17}$$

where c_r is the real part of c. Note that for *growing* disturbances, we are free to add any multiple of the integral in (11.16) to the left side of (11.17), since the former vanishes. We choose the multiplying factor to be c_r, giving

$$\int_{-\infty}^{\infty} \frac{\bar{U}\bar{U}_{yy}}{|\bar{U}-c|^2} |\tilde{v}|^2 \, dy = -\int_{-\infty}^{\infty} \left[\left| \frac{d\tilde{v}}{dy} \right|^2 + k^2 |\tilde{v}|^2 \right] dy. \tag{11.18}$$

Since the right-hand side of (11.18) is negative definite, so must be the left side. Therefore, *variations of $\bar{U}$ must be negatively correlated with variations of $\bar{U}_{yy}$ for growing disturbances*. (Again, this is a necessary but not sufficient condition for instability.) Figure 11.5 shows an example of a flow that the Fjørtoft theorem shows to be stable in spite of satisfying the Rayleigh necessary condition for instability, and a flow which satisfies both necessary conditions for instability.

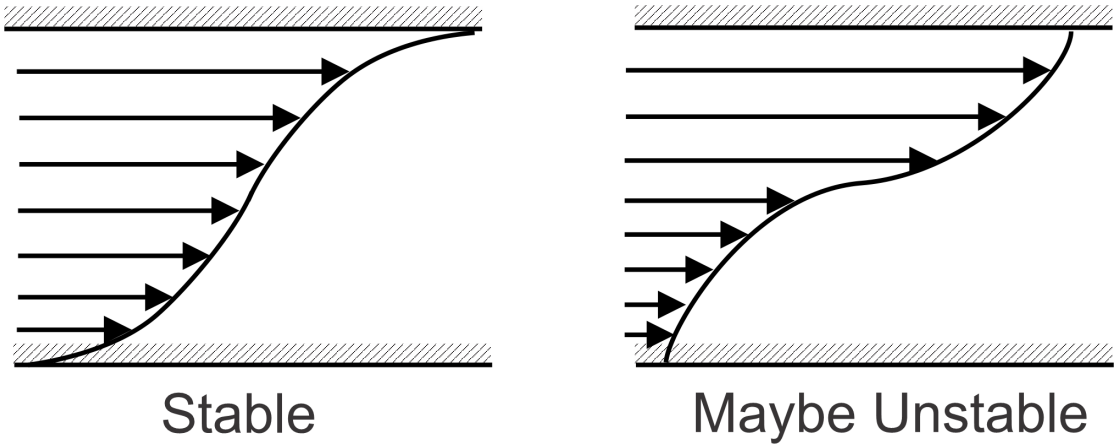


Figure 11.5: Two barotropic flow profiles, both of which satisfy the Rayleigh necessary condition for shear instability. The flow on the left is, however, stable according to the Fjørtoft theorem, while the flow on the right is not and therefore may be unstable.

The Rayleigh and Fjørtoft conditions can be interpreted in terms of mutually interacting Rossby waves. The Raleigh condition, that the vorticity gradient must change sign, ensures that there are at least two sets (or "flavors") of Rossby modes with opposite intrinsic (flow-relative) phase speeds that might interact with one another. The Fjørtoft conditions ensures that the variations of the background flow are such that the different "flavors" of Rossby can conceivably phase lock.

Armed with this understanding, let's now return to the problem of eyewall mesovortices. We saw in Chapter 10 that the eyewalls of tropical cyclones are highly frontogenetical. Were it not for radial diffusion of angular velocity, a discontinuity of azimuthal velocity, manifest as a delta function spike of vertical vorticity, would form in a finite time. Figure 11.6 shows the radial profile of vertical vorticity in the steady state achieved by the simple time-dependent model developed in Chapter 10, corresponding to the upper left panel of Figure 10.3. There is a strong maximum of vorticity corresponding to the inner edge of the eyewall. Outside the peak, the radial gradient of vorticity is negative, allowing for Rossby waves that propagate anticyclonically relative to the azimuthal flow. Inside the peak, Rossby waves travel cyclonically around the storm center

(relative to the background azimuthal flow). But the azimuthal wind itself peaks outside the vorticity peak, so the Doppler shifting it provides is in the right sense to allow the two flavors of Rossby wave to phase lock. Both the Rayleigh and Fjørtoft necessary conditions for instability are satisfied. While the peak of vorticity in this simple model is extreme, peaks in eyewall vorticity have been observed in aircraft and Doppler radar data in some tropical cyclones.

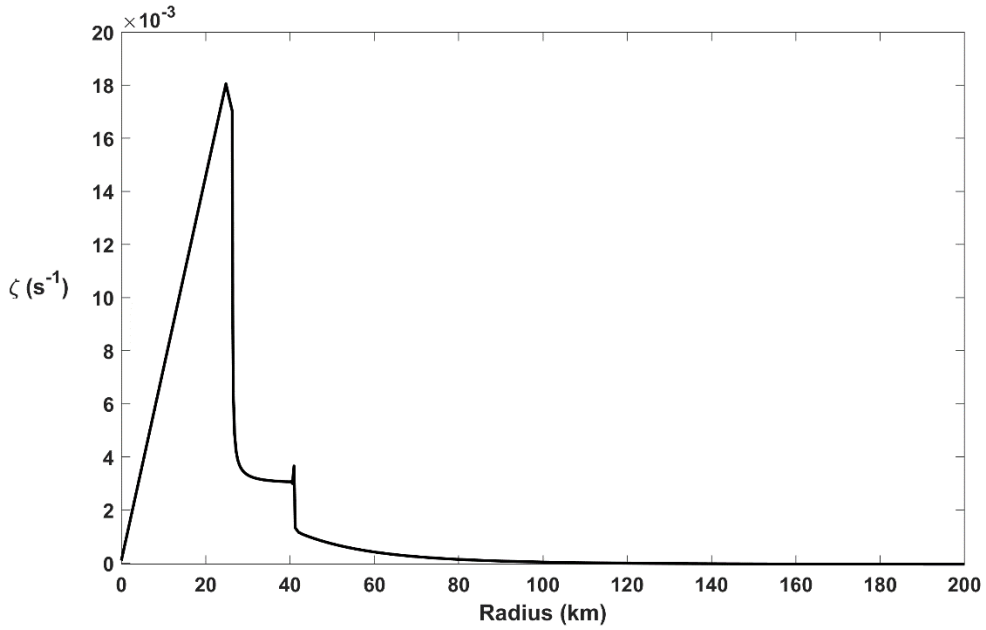


Figure 11.6: Radial profile of near-surface vertical vorticity in the steady-state achieved by the simple model developed in Chapter 10.

Yet it is not straightforward to carry deductions made from the simple Cartesian geometry of the Rayleigh problem to the circular geometry of tropical cyclones. The rotation of the flow imparts inertial stability to radial displacements, and the circular geometry limits the azimuthal wavelength to the circumference of the vorticity maximum. A comprehensive analysis of the stability of a cylindrical vortex sheet was undertaken by Rotunno (1978), who showed that, in the absence of any shear of the vertical velocity component across the vortex sheet, instability was only possible for disturbances with azimuthal wavenumber greater than 2. Shear of the vertical velocity destabilizes the flow and makes it possible to destabilize azimuthal wavenumbers of 1 and 2. For the typical situation in strong tropical cyclones, with updrafts outside the vorticity maximum and downdrafts inside, the most unstable modes would tend to tilt anticyclonically with height. However, close to the surface, where the vertical velocity vanishes, we could only have mesovortices with azimuthal wavenumbers of 3 and greater.

A number of numerical simulations have been carried out in idealized tropical cyclone vortices. For example, Schubert et al. (1999) confirmed the predictions of linear theory and carried out numerical simulations of the breakdown of initially axisymmetric rings of elevated vorticity. These typically broke down into 3, 4 or 5 mesovortices, giving the eyewall a polygonal appearance, as has been observed in satellite and radar imagery (e.g. Figure 11.1b). These mesovortices mix vorticity into the eye itself, smoothing out the initial vorticity peak and the radial profile of azimuthal wind, ultimately eliminating the instability and restoring a monopolar vortex.

Another complication in real world tropical cyclones is the existence of strong deformation as the radial inflow rapidly decelerates across the eyewall. As in the case of linear fronts, this deformation acts to suppress shear instabilities (e.g. Le Dizes et al. 1996), which may help explain why not all tropical cyclones exhibit eyewall mesovortices.

11.1.1.2: Inertial instabilities

Interacting Rossby waves are only one possible source of instability in rotating flows. We have seen that a change in sign of the vorticity gradient, and the possibility of Rossby waves phase locking are a necessary (but not sufficient) condition for this kind of instability. But other instability mechanisms have also been identified. Leibovich and Stewartson (1983) developed a sufficient condition for the instability of columnar vortices, which can be written

$$V\frac{d}{dr}\left(\frac{V}{r}\right)\left[\frac{d}{dr}\left(\frac{V}{r}\right)\frac{d}{dr}\left(rV\right)+\left(\frac{dw}{dr}\right)^2\right]<0, \tag{11.19}$$

where w is the vertical velocity. This was developed for vortices in an inertial coordinate system and would have to be modified for flows on an $f-$ plane. If vertical velocity is not present, (11.19) reduces to $\frac{d}{dr}\left(r^2V^2\right)<0$, which is the familiar criterion for pure inertial instability, which as described in Chapter 9, section 9.2.1. The author (Emanuel 1984) was able to show that the linear problem of inertial instability of an axisymmetric vortex with vertical motion is isomorphic to the classical problem with no vertical motion, in a local helical coordinate system rotating at angular velocity V/r in which all the shear (i.e. of both the relative V and w components) in these coordinates is contained in a plane that slopes in the vertical-azimuthal direction, as illustrated in Figure 11.7. This analysis also provided a growth rate for the most unstable modes.

Consequently, when radial shears of both the angular and vertical velocities are present, inertial instability can take place even when the angular momentum increases outward (so the flow is stable to classical inertial instability), and the most unstable modes will slope in the $\theta-z$ plane, taking the form of helical modes. The modes can be unstable outside the radius of maximum angular velocity, and these will be sloping backward with altitude, against the cyclonic mean flow. This is consistent with the structure of visible perturbations to the eyewall (see Figure 11.8), as argued by Bluestein and Marks (1987).

Whether the tornado-scale vortices of the kind described earlier and evident in Figure 11.1d are associated with the visible striations of cloud as seen in Figure 11.8 and/or with the helical form of inertial instability is unknown. The cause of these smaller scale vortices, which are consequential for conditions at the surface and for reconnaissance aircraft, clearly needs further research.

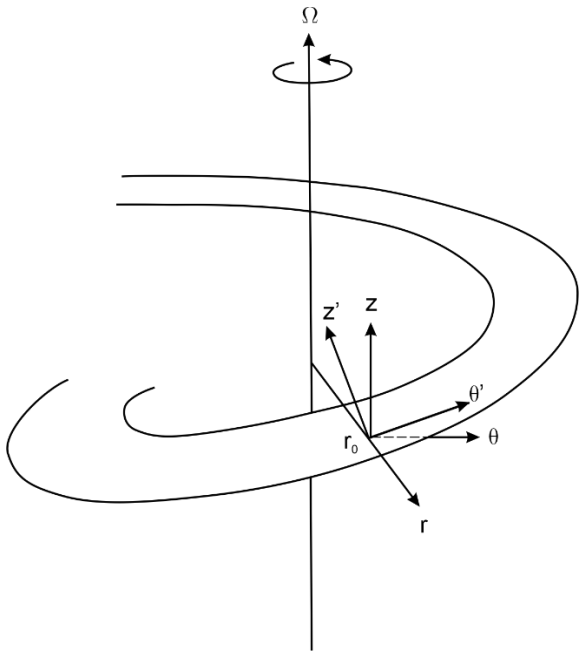


Figure 11.7: Showing the transformed coordinate system. Radius r, height z, and azimuth θ are the original coordinates, while r, z' and θ' are the transformed coordinates. The slope of the plane is that which maximizes the radial shear of the flow in the transformed plane, relative to rV_0'/r_0, where V_0' is the flow in the transformed plane at radius r_0.

Figure 11.8: The inner edge of the eyewall of Atlantic Hurricane Diana at 17:08 UTC on 11 September 1984. Note the cloud features sloping anticyclonically with height.

11.1.2 Asymmetric boundary layer structures

Detailed observations of air flow in tropical cyclone boundary layers reveal the existence of roll circulations, roughly aligned with the mean wind direction, that significantly modulate winds, including surface winds, in the boundary layer. An example of tropical cyclone boundary layer roll vortices, measured using a Doppler on Wheels (Wurman and Winslow 1998) is shown in Figure 11.9.

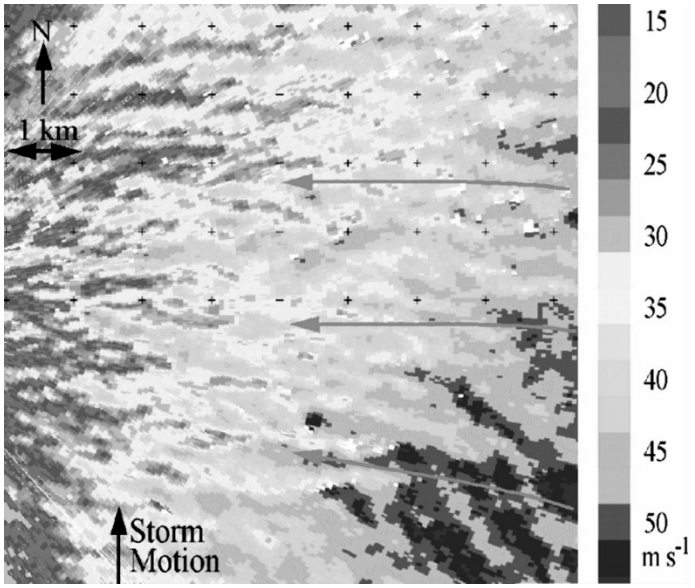


Figure 11.9: Doppler radar image in Atlantic Hurricane Fran of 1996, when it was just east of Wilmington, North Carolina (USA). The Doppler on Wheels radar was located just off the left edge of the image. The mean wind was from the east (long gray arrows) and the storm was moving northward. Wind streaks are aligned with the mean wind. Near the radar, at the left edge of the image, wind speeds across the rolls vary from $10\ ms^{-1}$ *to* $40\ ms^{-1}$, *with vertical vorticity values as large as* $0.1 s^{-1}$.

Roll vortices of horizontal scale less than 1 km, more or less aligned with the easterly flow are, prominent throughout the image. These span the depth of the boundary layer. Boundary layer rolls like these have also been measured directly with research aircraft flying in the boundary layer (Tang et al. 2021). Figure 11.10 shows an example of radial and vertical velocity measured by an aircraft flying across the direction of the wind at an altitude of 600 m, about 150 km from the center of western North Pacific Typhoon Nida of 2016. Consistent with Figure 11.9, the magnitudes of the horizontal and vertical velocities perturbations associated with the roll vortices are about the same as those of the mean flow, so that these roll vortices are quite strong. Power spectral analysis of aircraft-measured wind shows a prominent spectral peak at a wavelength very close to 1 km, so these rolls have spacings on the order of the boundary layer depth.

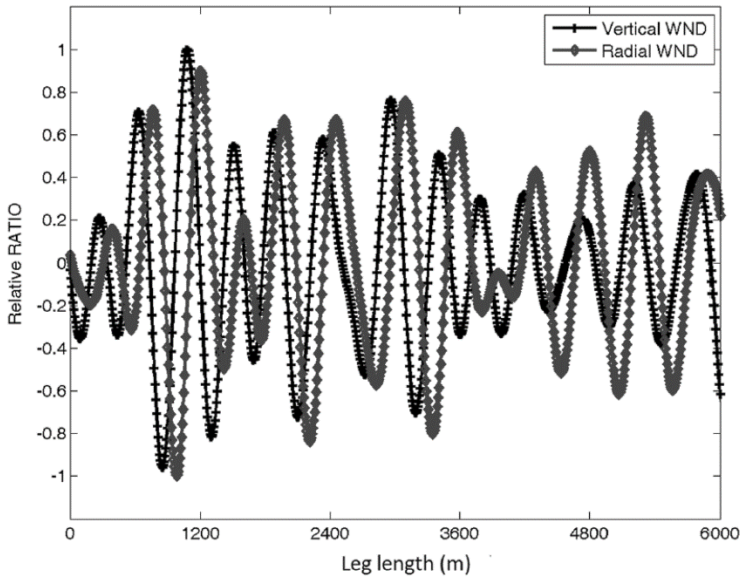


Figure 11.10: Vertical velocity (black), normalized by its peak value over the flight leg, and radial velocity (gray) normalized by the peak inflow velocity over the light leg, at an altitude of 600m in western North Pacific Typhoon Nida of 2016.

A schematic of the structure of roll vortices, based on observations, is presented in Figure 11.11. The rolls are nearly aligned with the vertically averaged boundary layer flow and transport momentum down its vertical gradient.

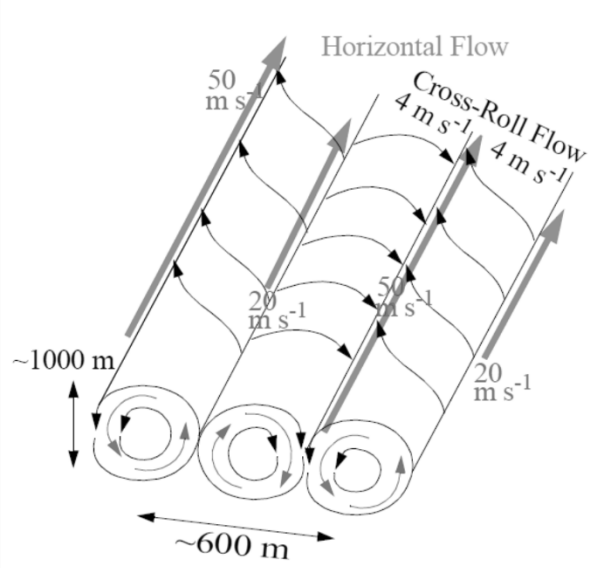


Figure 11.11: Schematic of boundary layer roll vortices in tropical cyclones, based on Doppler radar data. Air with high horizontal momentum from near the boundary layer top is transported downward, lower momentum air from near the surface is transported upward, so that these roll vortices transport momentum down its mean vertical gradient.

There is a rich literature on instabilities of rotating boundary layers, going back at least as far as the work of Gregory et al. (1955), Stern (1960), and Faller (1963). These and subsequent papers established that shear flows resulting from surface drag in rotating fluids can be unstable and result in horizontal vortices. Gregory et al. (1955) examined the stability of flows resulting from the insertion of a rotating disk into an otherwise stationary fluid and showed that the viscous boundary layers in such flows can be unstable to roll vortices. But a rotating disk in an otherwise stationary fluid is, in some sense, the reverse of the tropical cyclone, in which case a rotating fluid interacts with a stationary boundary. (On an *f*-plane, the setup considered by Gregory et al. would be analogous to a strong anticyclone interacting with the surface.) Stern found unstable modes that occupied the entire fluid layer, not just the boundary layer, even though they were driven exclusively by boundary layer shear. It is not clear whether such modes can exist in a fluid which is density stratified above the boundary layer.

The first experimental evidence for instabilities pertinent to the boundary layers of tropical cyclones was obtained by Faller (1963) using a rotating pan in which fluid was introduced at the outer edge and withdrawn from near the center. The resulting inflow spun up the fluid relative to the rotating pan; the retardation of this flow at the surface created an Ekman layer. By inserting dye into the fluid, Faller was able to detect and measure the properties of disturbances that formed in the boundary layer. Subsequent experiments by, e.g, Tatro and Mollo-Christensen (1967) and theoretical work by Lilly (1966) and Faller and Kaylor (1966, 1967) established a good theoretical and experimental understanding of Ekman layer disturbances, summarized nicely by Foster (2005).

Collectively, this work revealed at least three dynamical mechanisms at work in producing disturbances in rotating boundary layers. (We will not here consider another class of disturbances dominated by convective forcing.) The most fundamental of these is a shear flow instability of the kind discussed in the previous section, in which two flavors of Rossby wave are phase locked by shear of the mean flow and amplify one another. In this case, the Rossby waves are vertically, rather than horizontally, superposed. The most unstable modes are aligned about 15^{o} to the left of the gradient wind, forming inward spirals aligned more nearly with the vertically average boundary layer flow. The component of the mean boundary layer flow orthogonal to these bands have a (horizontal) vorticity maximum and satisfies both the Rayleigh and Fjørtoft necessary conditions for instability. The mean profile of velocity normal to bands spiraling cyclonically inward at an angle of 10^{o}, as measured by Faller (1963), is shown in Figure 11.12. The vorticity extremum occurs at a nondimensional altitude of about 2. The alignment of this kind of instability with the mean boundary layer wind as well as the depth and width of the most unstable linear mode are in good agreement with both laboratory experiments and with the observations of tropical cyclone boundary layer roll vortices reviewed earlier in this section.

A second type of instability was detected by Faller (1963) and in the linear stability analysis by Lilly (1966). This mode is more nearly parallel to the gradient wind and was identified by Lilly as a type of inertial instability. This mode is not clearly separated from the Rayleigh shear mode in diagrams of linear growth rate as a function of wavenumber and orientation, making it difficult to determine its role in boundary layers observed in tropical cyclones.

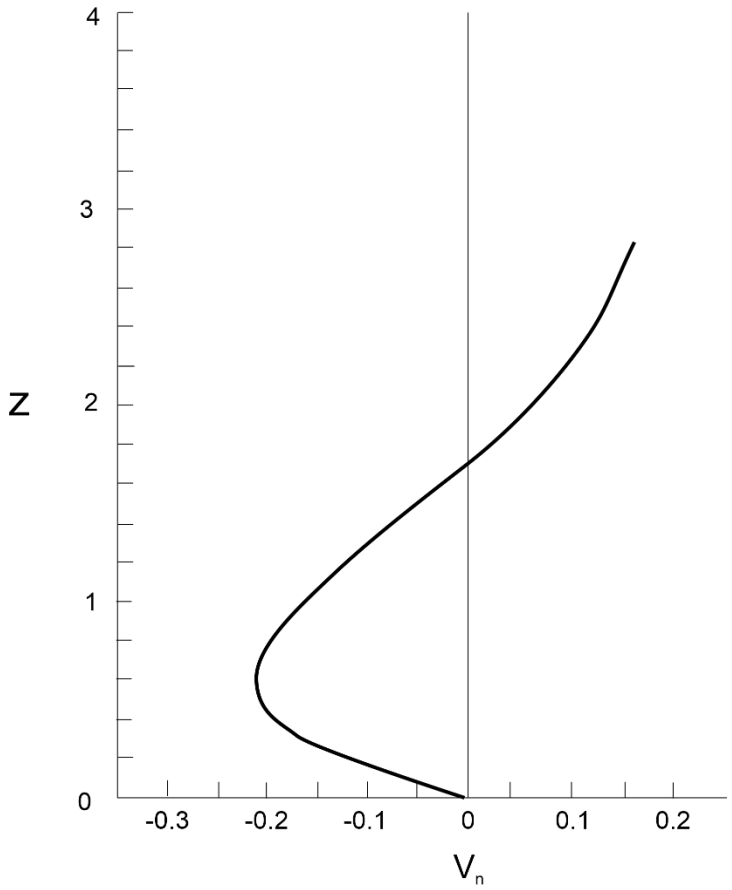


Figure 11.12: Profile of mean flow normal to bands oriented at a 10º angle to the gradient wind, as measured in the laboratory apparatus of Faller (1963). Both the velocity and the altitude are nondimensional; the altitude has been normalized by $\sqrt{\nu / \Omega}$, *where* ν *is the kinematic viscosity and* Ω *is the angular rotation rate of the apparatus.*

The third dynamical process may not be entirely separate from the first two. Faller and Kaylor (1967) undertook a stability analysis for both unstably and stably stratified flows with Ekman layers. In the latter case, there were new modes that appeared to represent a resonance between more typical boundary layer modes and inertia-gravity waves in the interior. These most unstable modes in the presence of stable stratification above the Ekman layer have somewhat longer wavelengths. Tatro and Mollo-Christensen (1967) also noticed that for certain parameters of their experimental setup, disturbances were found that extended through the whole depth of the fluid, rather than being confined to the boundary layer. In this case, because there was no density stratification, the resonance must have involved pure inertial oscillations in the interior, rather than inertia-gravity waves. Whether any of these domain-filling disturbances can be identified with the modes found by Stern (1960) remain an open question. It should also be remarked that some more contemporary stability analyses, such as that of Foster (2005), while largely supporting the validity of the earlier work, may have excluded domain-filling modes through the application of an upper boundary condition of zero disturbance amplitude. But observations of boundary layer rolls in tropical cyclones have so far failed to reveal much influence above the boundary layer.

The available evidence supports the idea that observed boundary layer rolls in tropical cyclones are the result of instability of boundary layer shear of the kind first identified by Faller (1963) with some possible influence of inertial instability of the kind first elucidated by Lilly (1966).

11.1.3 Spiral rainbands

One of the most distinctive features of tropical cyclones as seen in radar and satellite imagery are the spiral bands that wrap around the storm (see Figures 8.19 and 8.20). They were first documented in very early radar imagery of hurricanes (Wexler 1947)[1]. There have been quite a few observational studies of spiral rainbands (e.g. Senn and Hiser 1959; Barnes et al. 1983; Powell 1990; Gall et al. 1998; Corbosiero et al. 2006; Didlake Jr. and Houze R. A. 2013) culminating in a field project, the Hurricane Rainband And Intensity Change Experiment (RAINEX) of 2005 (Houze et al. 2006).

Sometimes there is only one obvious spiral band, or else one is clearly dominant, and in either case this is called the *principal band* (Willoughby 1988). In general, spiral bands appear to fall into roughly four categories: principal bands, secondary bands, distant bands, and inner bands (Houze 2010). Some of these are illustrated in Figure 11.13.

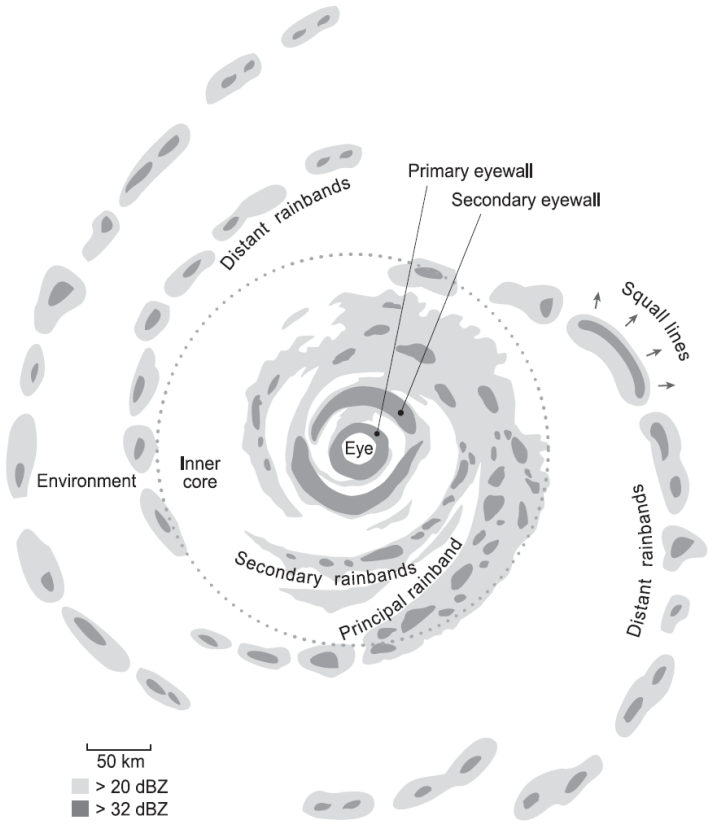


Figure 11.13: Schematic view of the radar reflectivity in a Northern Hemispheric tropical cyclone that exhibits a primary and secondary eyewall and various spiral rainbands.

[1] Drawings on ancient vases from the Caribbean region suggest that pre-Columbian cultures may have understood that tropical cyclones were vortices and possessed spiral structure (Ortiz 1947).

The outermost bands have structures that closely resemble tropical squall lines, and in general, the outer bands are more obviously convective in nature. As one proceeds radially inward, the bands take on a less cellular appearance and may be nearly continuous close to the core of the storm.

Aircraft measurements of bands away from the inner core show structures that closely resemble those of mid-latitude and tropical squall lines, with their leading edges facing radially inward; indeed, the bands are propagating anticyclonically with respect to the mean airflow. (In a storm-relative sense, they may be stationary or propagating slowly inward or outward (Powell 1990).) A cross-section, based on aircraft, airborne Doppler radar, and dropsondes, is displayed in Figure 11.14.

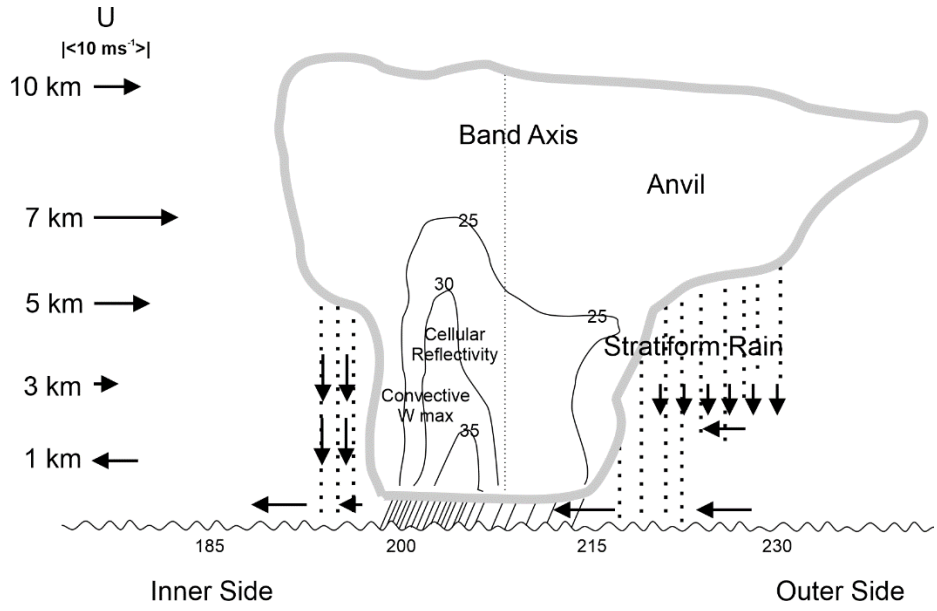


Figure 11.14: Schematic cross-section through a spiral rainband. Distances from the storm center (km) are indicated at bottom. The light gray curve represents the cloud boundary, and radar reflectivity (dbZ) is shown by the thin black curves. Horizontal airflow near the surface and updrafts and downdrafts are shown by arrows. The ambient cross-band airflow is shown at left.

The heaviest rain and strongest updrafts occur at the leading (inner) edge of the band. Evaporative cooling by the heavy rain creates a density current whose leading edge propagates inward. Evaporation and melting of hydrometeors in the trailing (outer) stratiform region also drives mesoscale descent and a downward-sloping inflow beginning at middle levels.

Spiral rainbands seem to be aligned nearly orthogonal to the vertical wind shear vector in the lower troposphere. Figure 11.15 shows a hodograph of winds averaged over a 40-by-40 km box straddling a rainband in Hurricane Earl of 1986, as determined by airborne Doppler radar. The mean shear between 0.8 km and 5.0 km is almost exactly perpendicular to the axis of the rainband. But while most tropical and extratropical squall lines propagate down the shear vector (Bluestein and Jain 1985; Robe and Emanuel 2001; Rotunno et al. 1988) relative to the mean flow, spiral rainbands strongly propagate upshear relative to the mean flow. This would seem to imply that different dynamics are at work in spiral bands, compared to most squall lines.

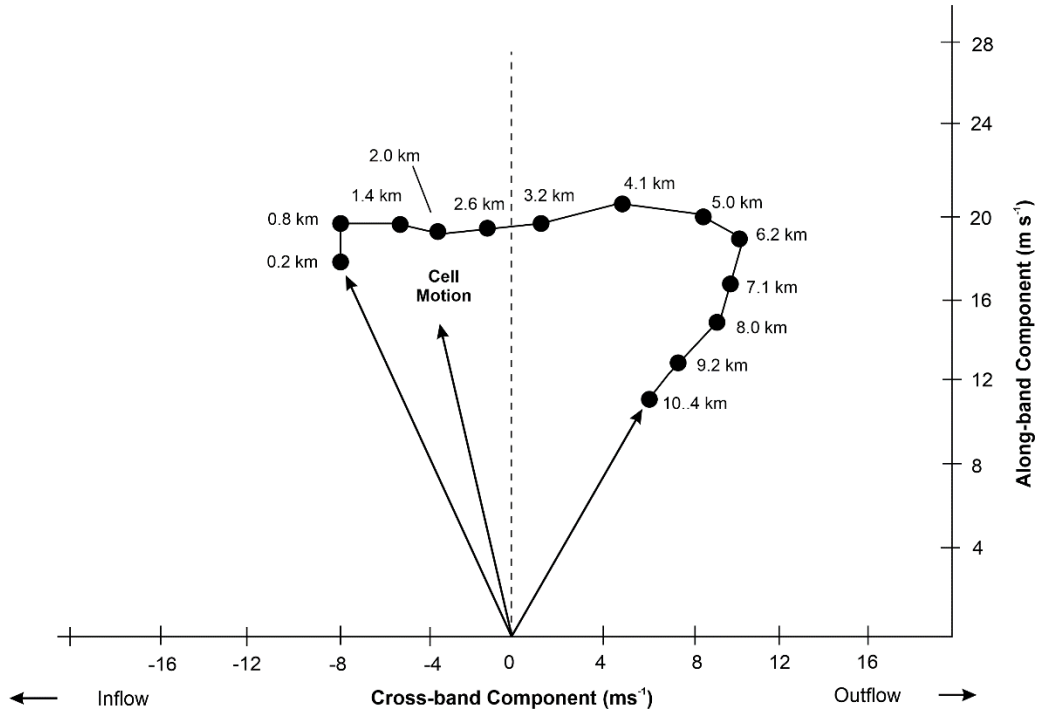


Figure 11.15: Hodograph in a coordinate system aligned with a segment of a spiral rainband in Hurricane Earl of 1986, averaged over a 40 x 40 km box straddling the band. The winds were measured by airborne Doppler radar.

The abundance of observational studies of spiral bands is matched only by the profusion of theories advanced to explain them. As of this writing, no single theory has become a generally accepted explanation for the existence of spiral bands. A comprehensive review of these theories is provided by Perez-Betancourt (2019).

Perhaps the earliest conjecture was made by Wexler (1947), who suggested that the mechanism of spiral banding was the same as that responsible for common “clouds streets” – rows of shallow cumulus clouds more or less aligned with low-level winds – which were in turn thought to result from dynamical and/or convective instabilities of the boundary layer. Non-precipitating convection in shear flow is relatively well understood today, and results in roll circulations whose axes are aligned with the shear. But spiral bands are more nearly orthogonal to tropical cyclone boundary layer shear, thus it is unlikely that they result from mere shallow convection in shear.

Ekman layer instabilities of the kind described in section 11.1.2 produce roll vortices aligned more nearly across the shear and thus could explain the orientation of spiral bands. But the most unstable horizontal wavelengths are of the order of the boundary layer depth, far too small to explain observed spiral bands, though as we saw in section 11.1.2, they are promising explanations for smaller scale roll vortices observed in the boundary layer.

Robe and Emanuel (2001) proposed that spiral bands are just squall lines in the particular low-level shear environment of tropical cyclones. But as we have seen, the flow-relative propagation direction of spiral bands is opposite to what is observed in ordinary squall lines, making that explanation suspect.

Senn and Hiser (1959) held that spiral bands are a manifestation of internal inertia-gravity waves generated in the storm core and propagating outward. Their work was followed by several further studies that pursued this hypothesis, and spiral modes that seemed to behave like inertia-gravity waves were diagnosed in numerical simulations of tropical cyclones (e.g. Anthes 1972; Kurihara and Tuleya 1974). These numerically simulated spiral bands propagated outward at unrealistically large speeds. A further problem with these early simulations, apart from coarse spatial resolution, was the presence of a rigid upper lid that would contain inertia-gravity waves; in reality, the upward component of the energy flux by such waves is substantial and much of the wave energy would be radiated into the stratosphere. Yet contemporary observations of tropical cyclones do reveal phenomena that strongly resemble inertia-gravity waves. Dunion (2014) observed diurnal pulses in satellite imagery of clouds tops in tropical cyclones and attributed them to inertia-gravity waves. This inference is strongly supported by numerical simulations undertaken by O'Neill et al. (2017). They showed that the waves in the outflow layer of their simulations were only apparent in the intrinsic frequency range between the inertial and buoyancy frequencies, in which inertia-gravity waves are possible. Nolan and Zhang (2017) detected inertia-gravity waves emanating from tropical cyclones, using satellite, aircraft, and surface data. These had radial wavelengths of 2-10 km and outward phase speeds of 20 – 25 ms^{-1}, generally too short and fast to explain spiral bands.

Inertia-gravity waves are part of a spectrum of waves and instabilities that are possible in cyclonic vortices. While observed spiral bands, almost by definition, have clouds and precipitation, it behooves us to step back and consider the spectrum of waves and possibly instabilities that are possible in vortices not encumbered by phases changes of water. What we learn from this exercise may not be so far off the example of "dry" vortices, given the behavior of the equatorial waveguide that we reviewed in Chapter 7. Recall that the observed spectrum of equatorially trapped waves, which are strongly coupled to clouds and precipitation, strongly resembles the neutral mode spectrum first described by Matsuno (1966), though with much reduced phase speeds. We showed that the main effect of condensation and evaporation in a convective quasi-equilibrium was to reduce the full baroclinic problem to the shallow water equations but with a much-reduced effective static stability. We might draw an analogy here by inquiring about the neutral and unstable modes of a shallow-water vortex on an *f-plane* (a limited region over which we consider the Coriolis parameter to be constant). Although this stipulation means that there is no gradient of planetary vorticity (β), there will be gradients of vorticity associated with the vortex itself, which will tail off to zero at large distance from the vortex center. Some of what we learn from this exercise might be carried over to the moist case for disturbances whose intrinsic time scales are long enough compared to moist convective time scales, in which case convective quasi-equilibrium should apply.

Another important qualitative difference from the equatorial β-plane is the strong differential advection of perturbations by the mean vortex flow. Among other aspects of the problem, the flow variation introduces the possibility of critical radii at which phase speed of particular disturbances matches the mean flow speed. The existence of such critical radii can greatly change the character of wave disturbances, in analogy with the existence of critical layers in the classical problem of vertically propagating gravity waves.

A suitable nondimensionalization of the shallow water equations in cylindrical coordinates results in an equation set with two nondimensional parameters, which are usually taken to be a

Froude Number and an inverse Rossby Number (e.g. see Ford 1994). These are defined, respectively,

$$F \equiv \frac{V_m^2}{gH}, \tag{11.20}$$

and

$$R \equiv \frac{fr_m}{V_m}, \tag{11.21}$$

where V_m and r_m are the maximum wind speed and radius of maximum winds in the unperturbed vortex, f is the Coriolis parameter, g is gravity, and H is the mean fluid depth. (Note that the Rossby number defined here is the inverse of the way it is usually defined.) The Froude number measures vortex flow speed compared to the shallow water gravity wave speed, while the Rossby number measures the Coriolis parameter relative to the vortex vorticity.

The structure and type of linear modes that can exist in any particular shallow-water vortex depends on the Froude and Rossby Numbers and, of course, on the radial profiles of velocity of the unperturbed vortex. An important conserved variable for the shallow water equations is the *potential vorticity,* q, defined

$$q = \frac{R + \mathbf{k} \bullet \nabla \times \mathbf{u}}{1 + Fh}, \tag{11.22}$$

where h is the departure (not necessarily small) of the fluid depth from its mean value H, normalized by H, $\mathbf{k}$ is the unit vector in the vertical, and $\mathbf{u}$ is the nondimensional horizontal velocity vector. In (11.22) and hereafter, all the velocities have been normalized by V_m, the length scales by r_m, the time scales by r_m / V_m, and the perturbation fluid depth by FH. It can be shown that $dq/dt = 0$.

The shallow water equations for the steady state axisymmetric vortex may be written

$$\frac{d\overline{h}}{dr} = R\overline{v} + \frac{\overline{v}^2}{r}, \tag{11.23}$$

and

$$\frac{d\overline{v}}{dr} + \frac{\overline{v}}{r} = \overline{q}\left[1 + F\overline{h}\right] - R. \tag{11.24}$$

The overbars here represent the steady state quantities. This is a second order system which can be solved subject to the specification of R, F and any one of $\overline{v}, \overline{h}$, or $\overline{q}$, together with the imposition of regular behavior at $r = 0$ and decay conditions at $r \to \infty$. Because of its fundamental importance in vortex dynamics, we shall generally specify the potential vorticity $\overline{q}$ and use (11.23) and (11.24) and the boundary conditions to calculate $\overline{v}$ and $\overline{h}$.

The linearized shallow water equations in cylindrical coordinates, governing perturbations to this mean vortex flow, are made nondimensional the same way (Ford, 1994). These can be solved by various numerical methods for any specified mean axisymmetric vortex, with the perturbations also subject to regularity at the axis and a decay condition as $r \to \infty$. Analytic solutions can be obtained in some special cases.

In general, similar to the equatorial β-plane, shallow water vortices admit to three general types of linear modes: inertia-gravity waves, Rossby waves (called “vortex Rossby waves” in the case of vortices (Montgomery and Kallenbach 1997)), and mixed Rossby vortex gravity-inertia waves. It can be shows that for rotationally dominated linear modes (in which magnitudes of vorticity perturbations greatly exceed those of divergence), horizontal gradients of potential vorticity play the same role as gradients of absolute vorticity in two-dimensional flow. Outside the eyewalls of tropical cyclones, the radial potential vorticity gradient is almost always negative (northern hemisphere), supporting Rossby waves that rotate anticyclonically around the storm center, relative to the vortex mean flow. The latter Doppler shifts these Rossby waves so that their ground-relative propagation may be cyclonic, anticyclonic, or stationary.

As noted above, the inertia-gravity wave solutions have intrinsic frequencies that are too large to explain spiral banding in tropical cyclones; they also flux energy upward and outward at fairly high rates.

MacDonald (1968) first suggested that spiral rainbands might be Rossby waves, based on very early satellite images of hurricanes. Solution of the linearized shallow water equations for mean state azimuthal profiles that resemble those of tropical cyclones by Zhong and Zhang (2014) show that the pure Rossby modes are generally confined to inner regions where the mean state potential vorticity gradient is large. Guinn and Schubert (1993) proposed that spiral bands are manifestations of nonlinear Rossby waves, seated in the inner region of large potential vorticity gradient, which have broken, shedding filaments of high potential vorticity into the outer region. Such tails of high potential vorticity are common in middle latitudes when Rossby waves participating in baroclinic instability break nonlinearly, shedding high potential vorticity filaments that extend equatorward and westward of the main baroclinic zones. In both cases, these nonlinear phenomena can extend well outside the critical radius (or latitude, in the case of middle latitude baroclinic waves) at which the linear wave speed matches the flow speed. Figure 11.16 shows an example of the distribution of potential vorticity in an initially elliptically shaped mass of potential vorticity. The resulting wavenumber 2 Rossby wave breaks, leading to long trailing tails of elevated potential vorticity that resemble spiral rainbands.

While the characterization of linear neutral normal modes may prove to be a useful first step toward understanding spiral rainbands, it raises the question of what processes might destabilize or otherwise excite such modes. Without such destabilization or excitation, the modes would rather quickly lose amplitude owing to frictional and/or thermal damping and radiation of energy upward and outward.

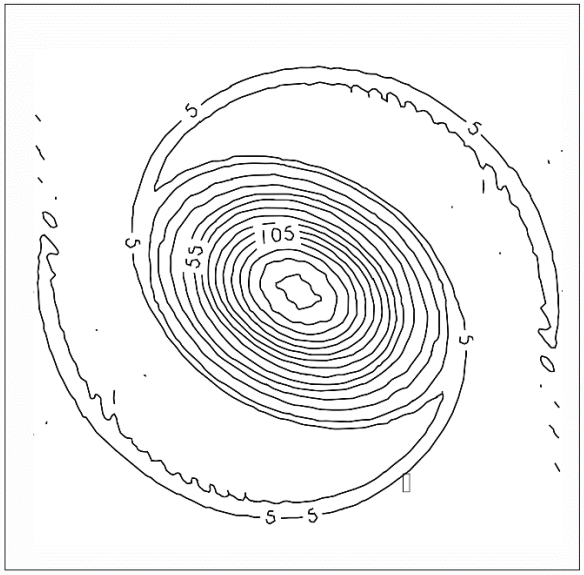


Figure 11.16: Distribution of potential vorticity 18 hours into a numerical simulation of the evolution of an initially elliptical mass of high potential vorticity in a shallow water system. The domain is 800 km x 800 km in size and the potential vorticity is labeled in units of $10^{-5} s^{-1}$, *with contours at intervals of* $10^{-4} s^{-1}$.

For the higher frequency inertia-gravity and mixed Rossby-inertia-gravity waves, relatively little is known about how they might be destabilized. Ford (1994) considered a shallow water vortex in which the potential vorticity was constant within a specified radius, and equal to the Coriolis parameter outside that radius. Rossby waves, in particular, are confined to the infinitesimal zone over which the potential vorticity jumps from f to its interior value. He found a spectrum of unstable normal modes, but for cyclonic vortices (interior potential vorticity greater than f) the growth rates are extremely small.

For the lower frequency Rossby modes for which vorticity is much larger than divergence, it can be shown that the Rayleigh and Fjørtoft necessary conditions apply, but with the radial gradient of potential vorticity taking the place of the radial gradient of absolute vorticity (see Chapter 7, section 7.3.4). That is, there must be two different “flavors” of Rossby wave (i.e., with opposite flow-relative phase speeds) embedded in a mean flow take it makes it possible for those two flavors to phase lock. This requires an extreme value of potential vorticity somewhere in the radial profile. We have already seen that this is a common feature of eyewalls and likely gives rise to the large mesoscale vortices observed in some tropical cyclone eyes. On the other hand, the radial gradient of potential vorticity outside the eyewall is usually negative everywhere, so the profile is necessarily stable to this kind of instability.

There is yet another intriguing possibility, arising from the baroclinity of tropical cyclones, which we have not yet discussed in this context. In an actual tropical cyclone, the flow varies with altitude and there are strong radial gradients of temperature, particularly within the eye. Relaxing the shallow water assumption and embracing a fully three-dimensional system, we again find a rich spectrum of low frequency Rossby waves for which the magnitude of the divergence is much less than that of the vorticity (Montgomery and Kallenbach, 1997). But we also introduce a kind of Rossby wave that is absent from the two-dimensional and shallow water systems.

To understand this, it is helpful first to define a potential vorticity that is conserved under adiabatic, frictionless processes in a fully three-dimensional, rotating, stratified fluid. That is Ertel's potential vorticity (Ertel 1942):

$$q_E = \alpha \nabla \theta \bullet \left[f\mathbf{k} + \nabla \times \mathbf{u} \right], \tag{11.25}$$

where α is the specific volume, θ is the potential temperature, and $\mathbf{u}$ is the full, three-dimensional velocity. Note that q_E is a function of both the thermal and velocity fields. If there is an additional, diagnostic relationship between the thermal and velocity fields, then (11.25) could be written as a function of a single variable.

In quasi-geostrophic theory (Charney 1948), the horizontal velocity is approximated to zero order by the geostrophic wind, and vertical advection is neglected except in the thermodynamic equation. The quasi-geostrophic form of the potential vorticity, which is conserved to leading order following the geostrophic flow, is

$$q_g = f_0 + \beta y + \nabla_H^2 \psi_g + \frac{1}{\rho_R} \frac{\partial}{\partial z} \left(\rho_R \frac{f_0^2}{N^2} \frac{\partial \psi_g}{\partial z} \right), \tag{11.26}$$

where $f_0 + \beta y$ is the Coriolis parameter linearized around some reference latitude, ρ_R is a reference density that varies only with altitude, N is an average buoyancy frequency corresponding to the reference state density; it too only varies with altitude, and ψ_g is the geostrophic streamfunction, defined such that the geostrophic winds are given by

$$v_g = \frac{\partial \psi_G}{\partial x}, \qquad u_g = -\frac{\partial \psi_G}{\partial y}. \tag{11.27}$$

Note that the Laplacian operator in (11.26) only involves horizontal derivatives. The conservation equation for q_g is

$$\left(\frac{\partial}{\partial t} + \mathbf{u}_g \bullet \nabla \right) q_g = 0, \tag{11.28}$$

provided the motion is frictionless and adiabatic.

Note that (11.26) is a linear elliptic relationship between the potential vorticity and the streamfunction. This is a strong analogy to the relationship between vorticity itself and streamfunction in two-dimensional flows, except that in the quasi-geostrophic case the elliptic. operator is in three dimensions rather than just two. To find the geostrophic flow and the temperature (which is proportional to the vertical derivative of the geostrophic streamfunction), one needs to invert the three-dimensional elliptic operator in (11.26), requiring boundary conditions at the top and bottom, not just the sides. This turns out to be a crucial distinction. Otherwise, the pair (11.26) and (11.28) are just like the equations relating streamfunction to vorticity, and conservation of vorticity, in two-dimensional flow. And, like the latter, one can form necessary conditions for instability based on integrals of the equations. But before we get to that, we need to talk about the upper and lower boundary conditions on ψ_G.

First, many atmospheric phenomena produce wind at the surface, so there is no reason to impose a condition that ψ_G must vanish. The first derivative of the geostrophic streamfunction is proportional, through the hydrostatic equation, to the potential temperature perturbation:

$$\frac{g}{\theta_0}\theta = f_0 \frac{\partial \psi_G}{\partial z}, \tag{11.29}$$

where θ_0 is a reference potential temperature and θ is the departure from the reference vertical profile of θ used to define N^2. If there exists a background gradient of potential temperature, then horizontal advection along the boundaries can produce temperature perturbations. The adiabatic quasi-geostrophic equation for potential temperature on a boundary along which the vertical velocity vanishes is

$$\left(\frac{\partial}{\partial t} + \mathbf{u}_g \bullet \nabla\right)\frac{\partial \psi_G}{\partial z} = 0. \tag{11.30}$$

Therefore, the boundary conditions that are imposed along horizontal boundaries at the top and bottom are dynamic rather than static.

The classical problem of baroclinic instability, posed to explain synoptic scale cyclone and anticyclones outside the tropics, considered the stability to small amplitude quasi-geostrophic perturbations, of zonally uniform baroclinic regions on a mid-latitude β-plane that extend indefinitely in the north-south and east-west directions. The baroclinic regions have westerly winds increasing with altitude, in thermal wind balance with am equatorward-directed temperature gradient. This basic state is a solution of the quasi-geostrophic equations, (11.26) (11.28), and (11.30); the question is whether such a state is stable to small amplitude perturbations.

Because the relationship between quasi-geostrophic potential vorticity and streamfunction, given by (11.28) is already linear, we only need to linearize the conservations equation for q_q (11.30) and the boundary condition (11.32). Denoting the basic state by overbars and perturbation quantities by primes, the problem to be solved is given by

$$\left(\frac{\partial}{\partial t} + \bar{U}\frac{\partial}{\partial x}\right)q_g{}' + \frac{\partial \psi'_g}{\partial x}\frac{d\bar{q}_g}{dy} = 0, \tag{11.31}$$

with

$$\frac{d\bar{q}_g}{dy} = \beta - \frac{1}{\rho_R}\frac{\partial}{\partial z}\left(\rho_R \frac{f_0^2}{N^2}\frac{\partial \bar{u}_g}{\partial z}\right) \tag{11.32}$$

and

$$q_g{}' = \nabla_H^2 \psi'_g + \frac{1}{\rho_R}\frac{\partial}{\partial z}\left(\rho_R \frac{f_0^2}{N^2}\frac{\partial \psi'_g}{\partial z}\right). \tag{11.33}$$

subject to the boundary conditions, at any rigid horizontal boundary,

$$\left(\frac{\partial}{\partial t}+\bar{U}\frac{\partial}{\partial x}\right)\frac{\partial \psi'_g}{\partial z}+\frac{\partial \psi'_g}{\partial x}\frac{g}{f\theta_0}\frac{d\bar{\theta}}{dy}=0 \quad \text{at boundaries.} \tag{11.34}$$

In writing (11.32) we have made use of (11.27) to relate the mean state streamfunction to the mean state zonal wind. Since none of the coefficients in (11.31) – (11.34) depend on x or y, the boundary conditions in those two coordinates are taken to be periodic.

Notice the very close mathematical similarity between the conservation equation for q_g' (11.31) and the boundary condition (11.34). Bretherton (1966) notices this and showed that one can solve the system (11.31) – (11.33) with the Dirichlet boundary condition, $\psi_g = 0$ if one replaces the definition of the mean gradient of potential vorticity (11.32) with

$$\frac{d\bar{q}_{ge}}{dy}=\beta-\frac{1}{\rho_R}\frac{\partial}{\partial z}\left(\rho_R\frac{f_0^2}{N^2}\frac{\partial \bar{u}_g}{\partial z}\right)+\left[\delta(0)-\delta(H)\right]\frac{f_0}{N^2}\frac{g}{\theta_0}\frac{d\bar{\theta}}{dy}, \tag{11.35}$$

where the $\delta's$ are Dirac delta functions at the surface $(z=0)$ and, if the tropopause is assumed to be a rigid boundary, at $z=H$. Here $d\bar{q}_{ge}/dy$ is the effective potential vorticity gradient of the mean state. This we regard our mathematical problem as the solution of (11.31), with $\bar{q}_{ge}$ substituting for $\bar{q}_g$, together with (11.33) and using (11.35), with $\psi_g = 0$ on horizontal boundaries and assuming periodic solutions in x and y.

We note again here the striking similarity of this problem to the Rayleigh problem of instability of two-dimensional shear flows described in section 11.1.1.1. Both involve solving a conservation equation for a vorticity variable and an elliptic diagnostic equation between that variable and the horizontal flow. The only difference in the quasi-geostrophic case is that the elliptic operator is in three dimensions rather than two…the Rossby waves can interact vertically as well as horizontally. And just as the mean flow vorticity gradient serves as a refractive index for Rossby waves in two-dimensional flows, the effective gradient of potential vorticity (11.35) serves as the refractive index for quasi-geostrophic Rossby waves. Rossby waves can live on interior gradients of potential vorticity and also on delta-function sheets of potential temperature gradients on boundaries. This second class of Rossby waves are trapped at boundaries and, for reasons we will reveal shortly, are called “Eady edge waves”.

Charney (1947) took as his basic state a baroclinic current in which the quasi-geostrophic potential vorticity has a constant positive gradient directed to the north while the potential temperature at the surface has a constant negative gradient. He took the flow to be unbounded in the vertical, and owing to the decrease of density with altitude, assumed that the perturbations vanish at $z=\infty$. His basic state density was given by $\rho_R = \rho_0 \exp(-z/H)$, where H in this case is a density scale height. With these assumptions, (11.35) becomes

$$\frac{d\bar{q}_{ge}}{dy}=\beta_1-\beta_2\delta(0), \tag{11.36}$$

where β_1 and β_2 are positive constants. The interior β_1 is associated with Rossby waves propagating westward relative to the mean flow, while the surface value β_2 is associated with eastward-propagating Eady edge waves.

Charney looked for modal solutions of the form $e^{ik(x-ct)}$, where k is the zonal wavenumber and c is a complex phase speed. If the imaginary part of c is positive, the disturbances are growing exponentially in time. Making this substitution, the solutions of (11.31) and (11.33) with (11.36) are mathematically complex and consist of, among other modes, a rich spectrum of exponentially amplifying disturbances with peak growth rates at distinct values of the zonal wavenumber. We note here that, in analogy with the Rayleigh shear problem, the instability results from the interaction of two flavors of Rossby wave – the westward propagating interior waves and the eastward-propagating Eady waves at the surface – and that the mean shear flow is such that these two waves can become phase locked.

Eady (1949) assumed instead that the basic state zonal flow has constant potential vorticity, so that $q_g' = 0$ everywhere for all time, but is bounded above by a rigid lid with the same mean state meridional temperature gradient. Thus, in his problem

$$\frac{d\overline{q}_{ge}}{dy} = \left[\delta(0) - \delta(H)\right]\frac{f_0}{N^2}\frac{g}{\theta_0}\frac{d\overline{\theta}}{dy}. \tag{11.37}$$

There are no interior Rossby waves, but there can be Eady edge waves on both the lower and upper boundaries. The lower one propagates eastward relative to the mean zonal wind, while the upper one propagates westward. The increase of mean zonal wind with altitude means that these two flavors of Rossby wave can potentially become phase locked. Indeed, the solution of the Eady problem, which is mathematically straightforward, shows a rich spectrum of unstable modes. Unlike the solutions of Charney's problem, there is shortwave cutoff: modes with wavelengths shorter than a critical value are neutral. The vertical penetration length of the Eady edge waves, which is proportional to their horizontal wavelength, is in this case too short to allow them to interact enough to destabilize one another. The maximum growth rate occurs for a wave whose length scales as the deformation radius, NH/f.

In section 11.1.1.1 we used integrals of the governing linear equations to develop the Rayleigh and Fjørtoft necessary conditions for instability. It is straightforward to do the same for the linearized quasi-geostrophic equations; this was accomplished by Charney and Stern (1962). The interested reader can consult that paper or Hoskins and James (2014) for this derivation. The result is the same: A necessary condition for instability is that the effective gradient of quasi-geostrophic potential vorticity (11.35) must change sign. This is true in the Charney problem because the θ gradient at the surface is negative while the interior potential vorticity gradient is positive. It is also true in the Eady problem because the θ gradient is negative at both boundaries, with the lower on contributing positively to the effective potential vorticity gradient and the upper one contributing negatively. The Fjørtoft condition is that the effective gradient of potential vorticity be positively correlated with the mean zonal wind. This is true in both the Charney and Eady problems.

Now (at last!) we return to the problem of spiral rainbands in tropical cyclones. Could they result from some form of interaction of low frequency Rossby waves?

First, it is helpful to extend the Charney-Stern and Fjørtoft theorems to the case of strong vortices, which cannot be regarded as quasi-geostrophic but are, nevertheless, quasi-balanced because, to leading order, the azimuthal winds are in gradient balance. This was accomplished by Montgomery and Shapiro (1995) and pertains to linear disturbances that are dominated by vorticity relative to divergence; that is, vortex Rossby waves.

First, they defined what we here refer to as a vortex potential vorticity, q_v. For the axisymmetric mean vortex (denoted by an overbar), this is

$$\overline{q}_v = N^2\overline{\eta} - \left(f + 2\frac{\overline{v}}{r}\right)\left(\frac{\partial \overline{v}}{\partial z}\right)^2, \tag{11.38}$$

where N is, as before, the buoyancy frequency of the mean state (a function of z only), $\overline{v}$ is the mean state azimuthal velocity, and $\overline{\eta}$ is the mean state absolute vorticity, defined

$$\overline{\eta} = f + \frac{1}{r}\frac{\partial}{\partial r}\left(r\overline{v}\right). \tag{11.39}$$

The generalized radial gradient of vortex potential vorticity that include boundary θ gradients as in (11.35) is then

$$\left(\frac{\partial \overline{q}_{ve}}{\partial r}\right)_\theta = \left(\frac{\partial \overline{q}_v}{\partial r}\right)_\theta + \frac{\overline{q}}{N^2}\frac{g}{\theta_0}\frac{\partial \overline{\theta}}{\partial r}\left[\delta(0) - \delta(H)\right]. \tag{11.40}$$

Note that the gradients of vortex potential vorticity are taken on surfaces of constant potential temperature, rather than on horizontal surfaces as in quasi-geostrophy.

Montgomery and Shapiro (1995) developed the vortex equivalents of the Charney-Stern and Fjørtoft necessary conditions for instability, namely, that $\left(\partial \overline{q}_{ve} / \partial r\right)_\theta$ as defined by (11.40) change sign in the domain, and that it be negatively correlated with $\overline{v}$.

Now, of course we have the problem that in real tropical cyclones, potential vorticity is hardly conserved, owing to strong diabatic heating and frictional dissipation. Outside the eyewall, whether there is deep convection in spiral and other bands but the air is not saturated on the mesoscale, there is no potential-vorticity-like quantity that is both conserved an invertible via balance equation. We can speculate, drawing from our experience with equatorially trapped waves in Chapter 7, that the most important effect of convection is to reduce, but not change the sign of, the effective static stability. In that view, for the slow Rossby modes, the Montgomery-Shapiro necessary conditions expressed through (11.40) are still valid, but with reduced N^2 that appears in it and in the definition of vortex potential vorticity, (11.38).

We have already seen, as discussed in section 11.1.1.1, that the eyewall often meets the Montgomery-Shapiro necessary conditions for instability, because the mean state $\overline{q}_v$ often peaks in or just inside the eyewall. But outside the eyewall, the gradient of $\overline{q}_v$ is usually

monotonic and negative. Thus there should be no contribution to instability from any reversal in internal $\overline{q}_v$ gradients. Along the surface, $\partial\overline{\theta}/\partial r$ is monotonically negative, the same sign as the internal $\overline{q}_v$ gradient, so there can be no unstable interaction between a vortex Eady edge wave and an internal vortex Rossby mode. There is, however, the possibility that the tropopause Eady edge waves could interact unstably with either the internal Rossby waves living on the radial potential vorticity gradient, or the surface edge wave.

Indeed, Perez-Betancourt (2019) suggested that unstable vortex Eady modes might exist owing to the fact that the radial potential temperature gradients are negative at both the surface and the tropopause, and the mean azimuthal wind shear is such that the two Eady waves could potentially become phase locked; that is, both of the Montgomery-Shapiro necessary conditions for instability are satisfied. She performed simulations with a fully axisymmetric, nonhydrostatic, cloud-resolving model phrase on an f -plane and with a full radiative transfer scheme. The initial classical radiative-convective equilibrium state rapidly breaks down into tropical cyclone-like vortices that possess, among other features, spiral rainbands.

Of particular interest was a set of simulations performed by Perez-Betancourt (2019) for the case of dry radiative-convective equilibrium, for which all of the surface thermodynamic disequilibrium is in a large jump from the imposed sea surface temperature to the atmosphere just above it. These, too, produced tropical cyclone-like vortices This system is free from complications stemming from phase change of water. On the other hand, the stratification of air below the outflow layer is essentially zero, so that quasi-geostrophy breaks down and one must consider nonhydrostatic baroclinic instability (Stone 1971). As the static stability becomes small, the wavelength of maximum instability in the Eady model begins to scale as $\Delta\overline{u}/f$, where $\Delta\overline{u}$ is the shear of the mean zonal wind. This is a mesoscale length scale which, for typical values in baroclinic currents in the atmosphere, is of order 100 km.

The vortices that develop in the simulations performed by Perez-Betancourt also develop prominent spiral bands. Figure 11.17 displays time-azimuth sections of radial velocity about 1.3 km above the surface, at three different radii. At 100 km radius, low azimuthal wavenumber disturbances propagating cyclonically around the center are evident between 2 and 6 days into the simulation. There propagation speeds range from about 2 ms^{-1} early on, to about 7 ms^{-1} by day 7, considerably slower than the mean azimuthal wind speed at this radius. They are also propagating radially outward (not shown) and can be found at larger radii after about 6 days. They are dominated by azimuthal wavenumber 2.

Further out in the storm, stationary, high wavenumber bands form early in the simulation but begin to merge into slowly propagating lower wavenumber modes later on. The lower wavenumber bands propagate cyclonically at around 6 ms^{-1} at 300 km and only a few ms^{-1} at 500 km radius; in both cases this is considerably smaller that the mean azimuthal flow at those radii.

The modes we are seeing here are likely to be vortex Rossby waves. This is consistent with their frequencies together with the fact that they are propagating anticyclonically with respect to the background mean azimuthal flow in the lower part of the circulation, as expected with an inward-directed potential vorticity gradient. But what is their source?

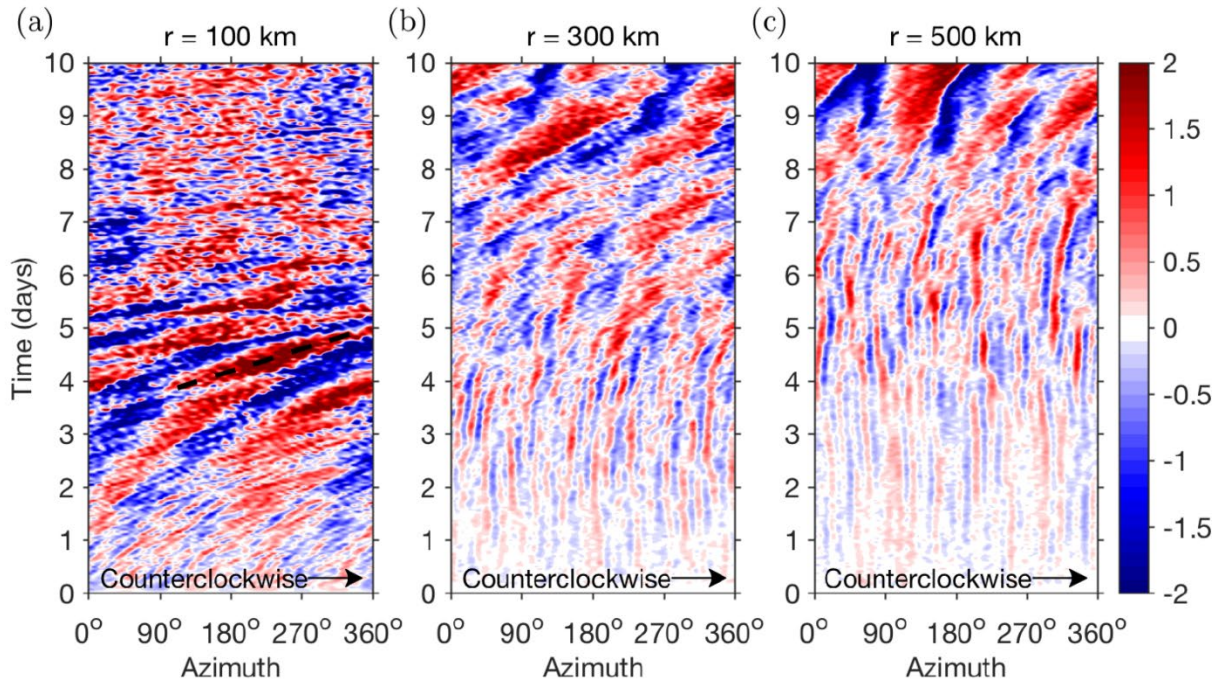


Figure 11.17: Time-azimuth sections of perturbations of radial velocity (ms^{-1}) from their azimuthal mean values, at 1.3 km altitude, from the dry, three-dimensional tropical cyclone simulations of Perez-Betancourt (2019). The velocities have been smoothed with a 20-km running mean. The cross sections are at radii of 100 km (a), 300 km (b) and 500 km (c).

As mentioned above, non-hydrostatic, non-geostrophic baroclinic instability could result for the interaction between the tropopause Eady edge waves, which propagate cyclonically relative to the background anticyclone at that level, and either or both the internal Rossby waves living on the internal radial potential vorticity gradient and the surface-based Eady edge waves.

Clues to the wave origin may be found in Fourier decomposition in azimuth of the disturbances at each level and radius. Figure 11.18 presents an example of this: The wavenumber 22 decomposition of various quantities in the altitude-azimuthal plane at 380 km radius. The high wavenumber spiral banding is pronounced at this time (3.5 days; Figure 11.18a). The temperature perturbations (panel b) show a strong maximum near the tropopause, suggesting the presence of an Eady edge wave there. There is also a local temperature perturbation maximum at the surface, suggesting that a surface edge wave may also be playing a role here. But the almost continuous-in-altitude along-band wind component (as associated vorticity) also suggests a prominent role for internal vortex Rossby waves. This may be a case in which banding is a manifestation of baroclinic instability resulting from the tropopause edge wave interacting with internal vortex Rossby waves and/or surface edge waves. The structure and dynamics of such waves would no doubt be affected by the strong radial flow near the surface and the tropopause, a component of background flow absent from the classical linear analyses of baroclinic instability. The azimuthal wave length of the wavenumber 22 disturbance is commensurate with the scaling of nonhydrostatic baroclinic instability, but the merging into much lower wavenumbers later in the simulation remains to be explained.

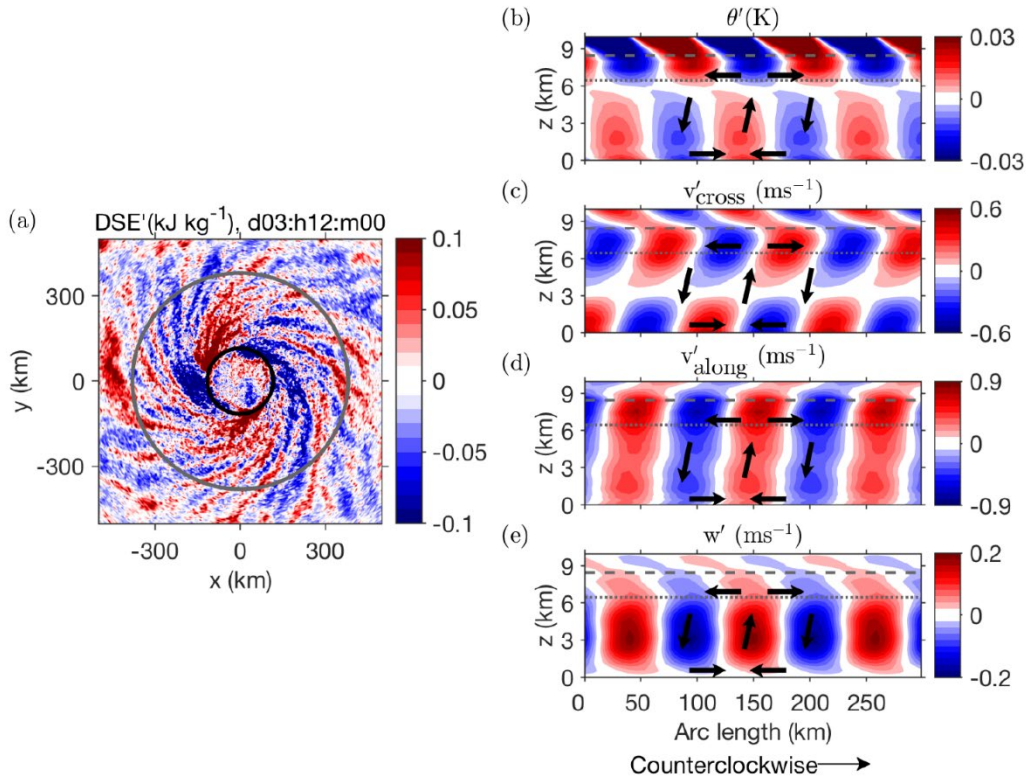


Figure 11.18: a) Departures of dry static energy (kJ Kg^{-1}) from its azimuthal mean value averaged over 1.3 to 5.5 km altitude at day 3.5 into the simulation of dry tropical cyclones by Perez-Betancourt (2019). The inner circle is at the radius of maximum winds and the outer circle is at a radius of 380 km. b) The wavenumber 22 component of perturbations to azimuthal mean components at a radius of 380 km. The horizontal wind is broken into cross-band and along-band components assuming a band orientation of 36º. Azimuthal distance is expressed in arc length (km). b) Potential temperature (K), c) cross-band wind component (ms^{-1}), d) along-band wind component (ms^{-1}), and e) vertical velocity (ms^{-1}). The tropopause, marked by a temperature inversion, is shown by the gray dashed line in the panels.

Whether the kind of banding operating in these dry simulations is dynamically similar to that observed in nature and in moist simulations also remains to be shown.

While there is an emerging consensus that tropical cyclone spiral rainbands see so clearly in radar and satellite imagery is a manifestation of vortex Rossby waves, much research remains to be done on the mechanism(s) in play, with baroclinic instability a prominent candidate.

11.2: Asymmetries owing to storm translation

Tropical cyclones are driven by surface enthalpy fluxes and retarded mostly by surface drag. Both of these fluxes are mostly wind-driven; surface buoyancy fluxes play a secondary role in tropical cyclones. Surface winds associated with the background flow in which tropical cyclones are embedded also contribute to the enthalpy and momentum flux. The net surface fluxes at any point in the circulation are rendered asymmetric by the background flow.

As an illustration, Figure 11.19 shows the perturbation absolute value of the surface wind speed from a constant background easterly flow of 5 m s^{-1}. This represents the addition of the constant background easterlies to an axisymmetric wind profile of the type described in Chapter 9, section 9.3.2, with a maximum wind speed of 60 m s^{-1}, an outer radius of 500 km, a Coriolis parameter of 5x10-5 s^{-1}, a radiative subsidence velocity of 2 mm s^{-1}, and a drag coefficient of $1.5x10^{-3}$.

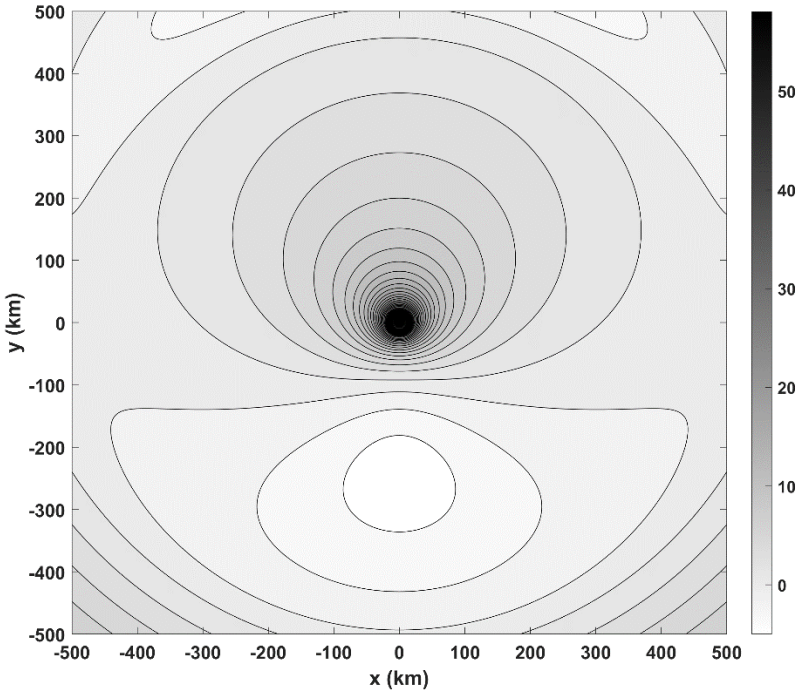


Figure 11.19: Perturbation of the absolute value of the surface wind speed from a constant background easterly wind of 5 m s^{-1}. The vortex has a peak axisymmetric wind speed of 60 m s-1 and extends to an outer radius of 500 km; see text for other vortex parameters. The Cartesian map is centered at the center of the axisymmetric vortex.

Close to the center of the vortex, where the axisymmetric wind component greatly exceeds the background wind speed, the effects of the latter are barely noticeable. But further away from the center, there is a pronounced asymmetry, with perturbation wind speeds becoming negative around 300 km south of the storm center.

It is important to recognize that the effect of background wind on the total local wind speed does not average out when taking the mean around circles centered at the vortex center. Representing the constant background zonal wind as u, the azimuthal component of the vortex

flow as V and the angle from true north as θ, the absolute value of the total azimuthal wind is given by

$$\begin{aligned} |\mathbf{V}| &= \sqrt{\left(V - u\cos(\theta)\right)^2 + u^2\sin^2(\theta)} \\ &= \sqrt{V^2 + u^2 - 2Vu\cos(\theta)}. \end{aligned} \tag{11.41}$$

We might suppose that, from an energetic standpoint, what matters is the *perturbation* enthalpy flux, whose azimuthal average is just the difference between the azimuthal mean of (11.41) and the background wind:

$$\overline{|\mathbf{V}|}' = \frac{1}{2\pi}\int_0^{2\pi}\sqrt{V^2 + u^2 - 2Vu\cos(\theta)}\,d\theta - u, \tag{11.42}$$

where the overbar signifies the azimuthal mean. Likewise, the relevant dissipation rate should be the perturbation of the azimuthal mean of wind speed cubed:

$$\overline{|\mathbf{V}|^3}' = \frac{1}{2\pi}\int_0^{2\pi}\left(V^2 + u^2 - 2Vu\cos(\theta)\right)^{3/2}d\theta - u^3. \tag{11.43}$$

Doing the integrals numerically, these two functions are shown in Figure 11.20.

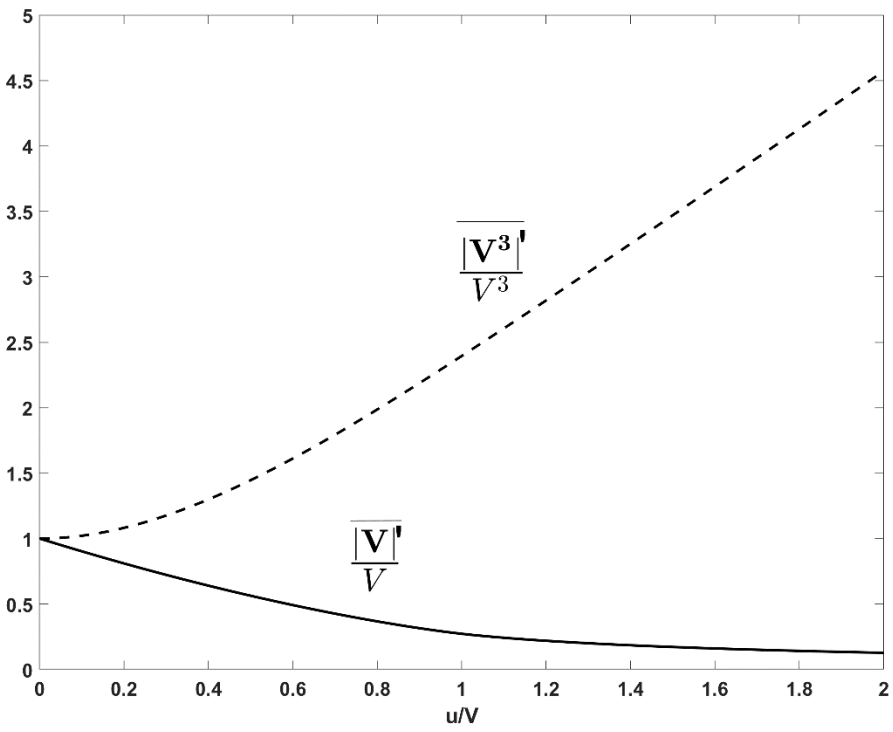


Figure 11.20: The perturbation azimuthally averaged surface wind speed divided by the circular vortex wind (solid), and the perturbation average wind speed cubed divided by the circular wind speed cubed (dashed), according to (11.42) and (11.43). Both are functions of the background zonal wind, u, divided by the circular vortex wind.

As the background wind increases relative to the vortex winds speed, the effective perturbation azimuthally averaged wind declines, while the perturbation average wind speed cubed increases. Because the former governs the enthalpy flux and the latter governs the kinetic energy dissipation, background wind is unfavorable to the energy cycle.

We can quantify this by equating kinetic energy dissipation to generation of energy through the thermal cycle, as in Chapter 9:

$$C_D\overline{|\mathbf{V}|'^3} = C_k\overline{|\mathbf{V}|'}\frac{T_s - T_o}{T_o}\left(k_0^* - k_s\right), \tag{11.44}$$

where the symbols (if not defined here) are as in Chapter 9 and we have assumed that dissipative heating is active. Using the definition of potential intensity (9.26) we can write (11.44) as

$$\overline{|\mathbf{V}|'^3} = V_p^2\overline{|\mathbf{V}|'}. \tag{11.45}$$

Given a background zonal wind u and a potential intensity V_p, (11.45), (11.42), and (11.43) can be solved for the energetic upper bound on V. We solve this system numerically and display the result in Figure 11.21, which displays the azimuthal wind normalized by the nominal potential intensity as well as the latter plus the background zonal wind, also normalized by the nominal potential intensity.

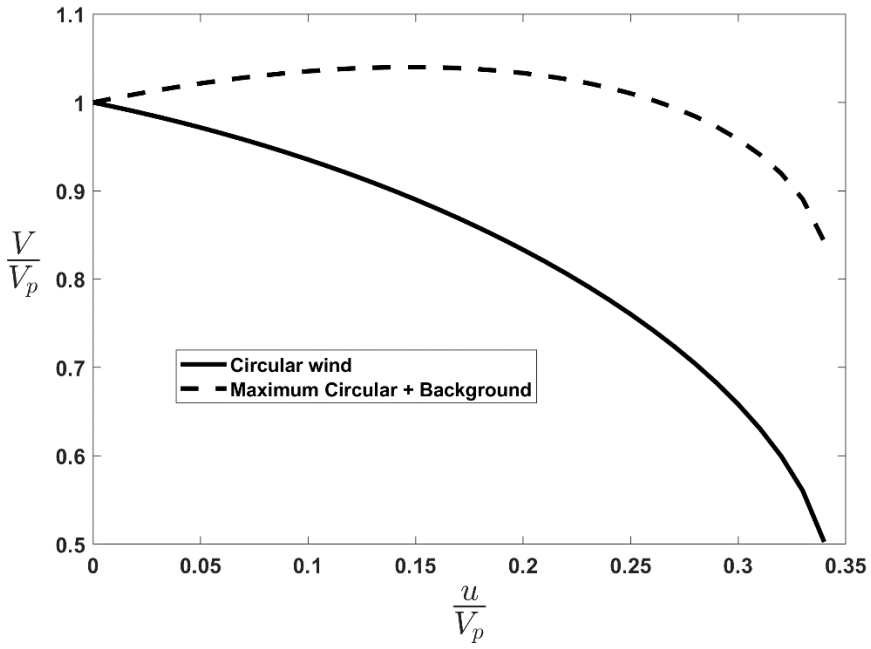


Figure 11.21: The energetic upper bound on the storm's maximum circular velocity component (solid) as a solution of (11.42), (11.43), and (11.45). The dashed line shows the maximum ground-relative wind speed as the sum of the circular wind and the background wind. Both quantities are normalized by the nominal potential intensity. No solution is possible for $u / V_p >\sim 0.34$.

No solution of this system is possible if the background wind is more than about a third of the nominal potential intensity. Below this value, the actual energetic upper bound on the circular component of wind falls off from the nominal potential intensity as the background wind increases, reaching a minimum of about ½ the nominal potential intensity just before the background wind exceeds it critical value of about 1/3. But note in Figure 11.21 that when one adds the background wind to the circular wind component to estimate the peak ground-relative wind (dashed line), there is hardly any change with background wind until the latter exceeds roughly 30% of the potential intensity. Thus the nominal potential intensity is a good bound on peak wind speed when the background wind is included in the net wind speed.

All of the above assumes that we can represent a tropical cyclone as a superposition of a circular vortex and a constant background flow. But this is not likely to be the case.

Return again to the map of surface wind speed displayed in Figure 11.19. Outside the eyewall, the surface winds speed has large asymmetry, which will lead to strongly asymmetric surface enthalpy fluxes and drag. This will, among other things, change the divergent part of the airflow in the boundary layer and introduce asymmetries into the frictionally produced vertical velocity at the top of the boundary layer. That, together with the asymmetric surface enthalpy flux, will force asymmetries in deep convection.

Shapiro (1983) presented a two-dimensional slab boundary layer model driven by a specified, axisymmetric pressure distribution and translating at a constant speed relative to the surface. The azimuthal variability in his model was represented by azimuthal wavenumbers 1 and 2, and the depth of the boundary layer was held constant.

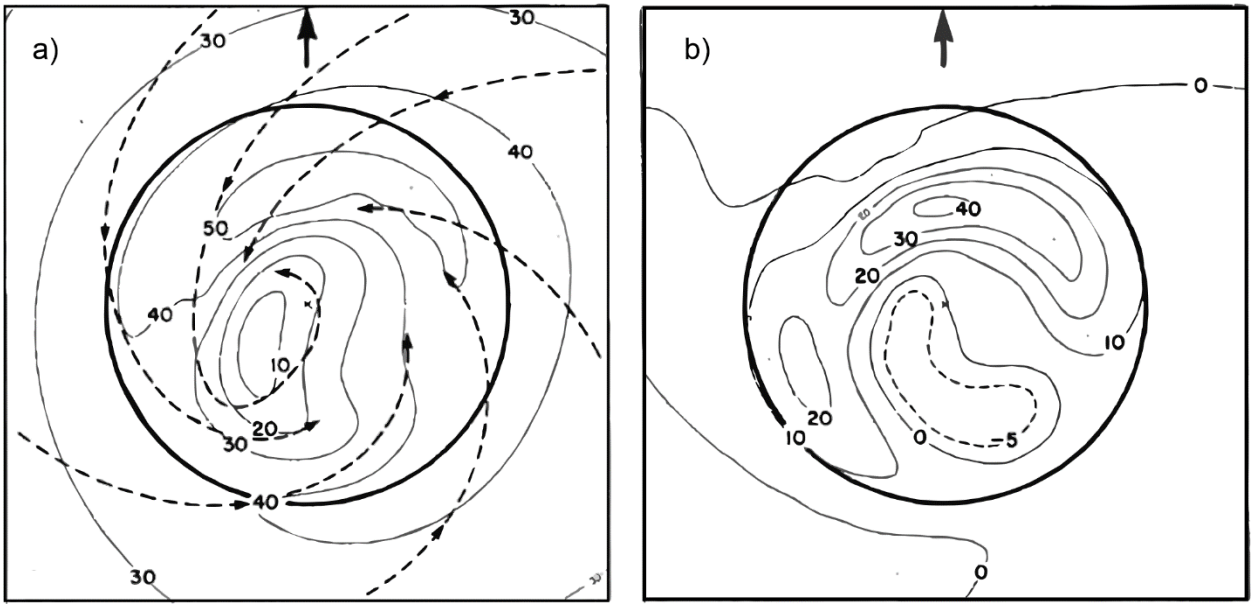


Figure 11.22: Boundary layer variables associated with a tropical cyclone-like axisymmetric gradient wind distribution moving in the direction of the arrows at 10 m s^{-1}. The gradient wind has a maximum value of 42 m s^{-1}, located 40 km from the storm center as shown by the circle. In a), thin black curves are isotachs of ground-relative wind (m s^{-1}) and heavy dashed arrows show streamlines. In b) the contours are velocity convergence in 10^{-4} s^{-1}.

Figure 11.22 shows some boundary layer variables associated with an axisymmetric pressure distribution with maximum gradient winds of 42 m s^{-1} at a radius of 40 km, translating northward at 10 m s^{-1}. This relatively large translation speed has a strong effect on boundary layer winds. The maximum winds occur now well inside the radius of maximum gradient wind (partly owing to the supergradient wind effects discussed in Chapter 9, section 9.5) and concentrated in the front sector of the cyclone, with respect to its translation. Strong winds are also weighted to the right side of this northern hemisphere vortex, reflecting the addition of the background flow to the vortex winds. Convergence is strongly asymmetric (Figure 11.22b), with the strongest convergence (and therefore ascent at the boundary layer top) located in the front sector of the vortex.

Shapiro's work, and subsequent research, demonstrates that background winds can have a strong effect on the structure of tropical cyclones. We now turn our attention to another environmental effect on tropical cyclone symmetry.

11.3 Asymmetries owing to beta

Mature tropical cyclones can be idealized as highly localized towers of high potential vorticity (or large negative values in the southern hemisphere). But near the tropopause, air rising through the eyewall flows out of the storm; this outflow may cover a large horizontal area though typically it is not very deep. This air as very low potential vorticity, having risen through the high wall with saturation potential vorticity near zero. At the very low temperature of the upper tropical troposphere, saturation potential vorticity and dry potential vorticity are very nearly equivalent.

These tropical cyclone-generated potential vorticity anomalies interact with any large-scale background gradients of potential vorticity that may be present. The simplest case is one in which we only have the gradient of planetary vorticity, β, to deal with, and we will approximate that as a constant.

In the lower and middle troposphere, on the east side of the cyclone, poleward flow will advect low values of planetary vorticity poleward, while on the west side, equatorward flow will advect higher values of planetary vorticity toward the equator. This advection will lead to a larger-scale cyclonic gyre west of the cyclone and an anticyclonic gyre to the east. As we will show in Chapter 12, the circulation around these gyres tends to advect the tropical cyclone poleward and westward. The gyres created by the interaction of the tropical cyclone with β are known as "β-gyres".

At the top of the storm, a similar interaction yields anticyclonic gyres to the equatorward and westward of the storm center, and cyclonic gyres poleward and eastward of the storm. This generates a equatorward and eastward directed vertical wind shear in the vicinity of the tropical cyclone.

Fang and Zhang (2012) showed, using numerical simulations, that this $\beta-$induced shear excites deep convection on the downshear side of the tropical cyclone, for reasons explained in Chapter 12. In this case, the downshear side is to the east and equatorward of the cyclone center.

Figure 11.23 compares the simulated radar reflectivity at 4 km altitude in two simulations performed by Fang and Zhang (2012): One with constant f and the other with constant β. These are shown at 96 and 114 hours into the two simulations.

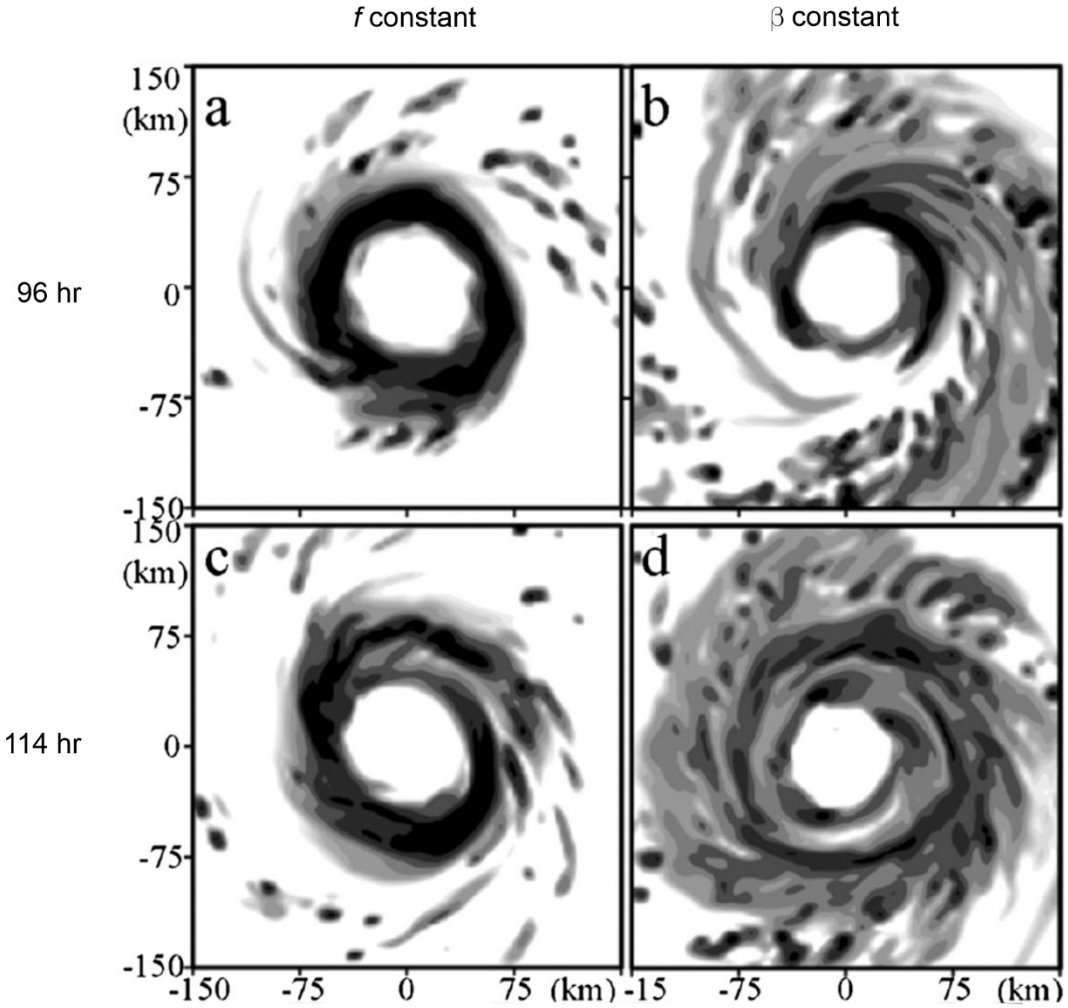


Figure 11.23: Radar reflectivity at 4 km altitude in two numerical simulations performed by Fang and Zhang (2012). Panels a) and c) are for a simulation on an f-plane, while b) and d) are for a simulation on a β-plane. The top and bottom row are at 96 hours and 114 hours into the simulations, respectively.

Compared to the f-plane simulation at 96 hours (upper-left panel), the $\beta-$plane simulation has much more convection outside the eyewall, particularly to the east and south of this simulated northern hemispheric tropical cyclone. But by 114 hours (bottom-right panel), the convection has largely axisymmeterized through a process explained by Montgomery and Kallenbach (1997), leaving a storm with a secondary eyewall and considerably more convection outside the primary eyewall.

A more detailed view of the time evolutions of these two simulations is presented in Figure 11.24.

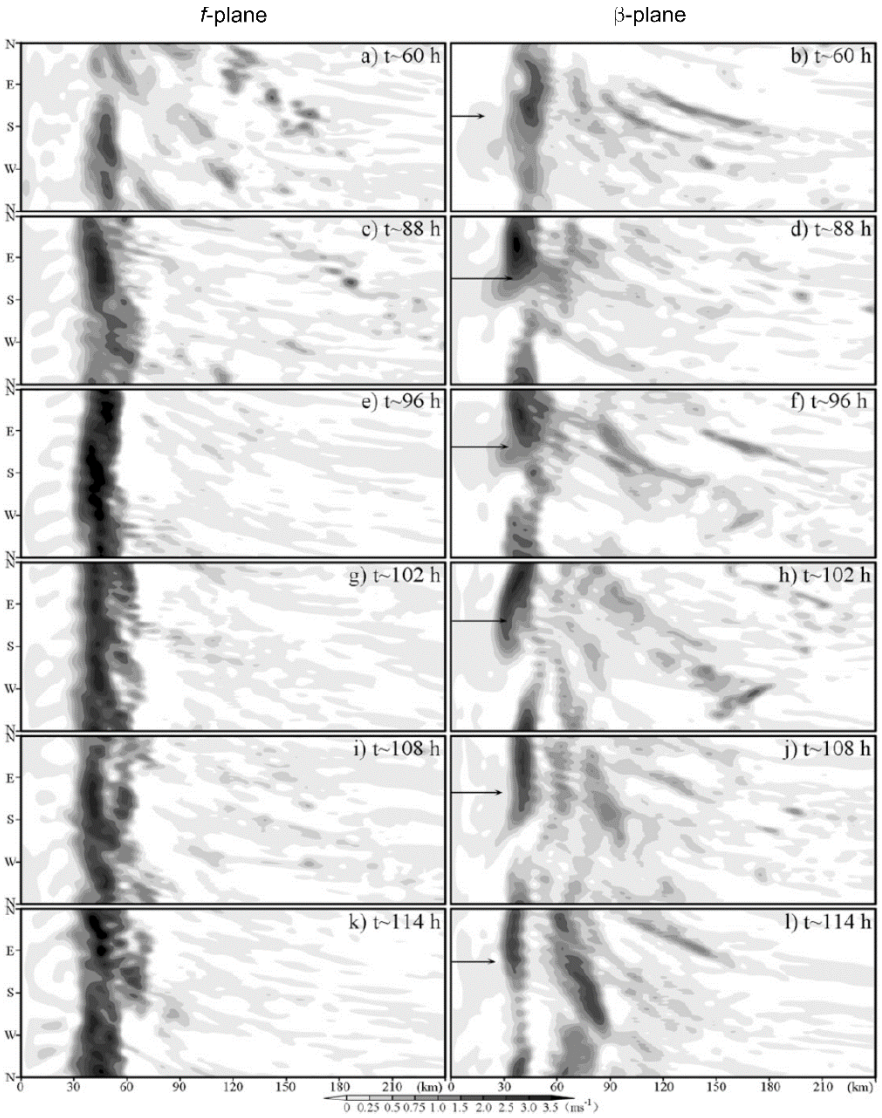


Figure 11.24: Evolution over time of the 1-hour average vertical velocity at 3 km altitude in the same two simulations by Fang and Zhang (2012) as shown in Figure 11.23. Time into the simulation is shown in the upper-right corner of each panel. The abscissa is distance from the storm center (km) while the ordinate is azimuth going counter-clockwise around the storm center.

While both simulations produce primary eyewalls, those in the $\beta-$plane simulation are less continuous in both time and azimuth. Deep convection breaks out south and east of the storm center in the $\beta-$plane simulation, and is arranged in cyclonic spirals. These propagate radially inward over time and eventually become axisymmeterized into a secondary eyewall by about 108 hours. This provides and alternative, non-axisymmetric explanation for the formation of secondary eyewalls, and in this case seems to have been produced by a two step process in which the $\beta-$induced shear excites convection equatorward and east of the storm center, and this deep convection is subsequently axisymmeterized into a secondary eyewall by the Montgomery-Kallenback process.

While the asymmetries described in this Chapter are, in a sense, induced by the tropical cyclone itself, asymmetry can also result from interactions with environmental shear and other features of the larger-scale environment, as described in the next chapter.

References

Anthes, R. A., 1972: Development of asymmetries in a three-dimensional numerical model of the tropical cyclone. *Mon. Weather Rev.*, **100**, 461–476, https://doi.org/10.1175/1520-0493(1972)100<0461:DOAIAT>2.3.CO;2.

Barnes, G. M., E. J. Zipser, D. Jorgensen, and F. Marks Jr., 1983: Mesoscale and Convective Structure of a Hurricane Rainband. *J. Atmospheric Sci.*, **40**, 2125–2137, https://doi.org/10.1175/1520-0469(1983)040<2125:macsoa>2.0.co;2.

Bluestein, H. B., and and M. H. Jain, 1985: Formation of mesoscale lines of precipitation: Severe squall lines in Oklahoma during the spring. *J Atmos Sci*, **42**, 1711–1732.

——, and F. D. Marks Jr., 1987: On the structure of the eyewall of Hurricane "Diana'' (1984): comparison of radar and visual characteristics. *Mon Wea Rev*, **115**, 2542–2552.

Bretherton, F. P., 1966: Critical layer instability in baroclinic flows. *Q. J. R. Meteorol. Soc.*, **92**, 325–334, https://doi.org/10.1002/qj.49709239302.

Brian Hoskins and Ian James, 2014: *Fluid dynamics of the mid-latitude atmosphere*. Wiley-Blackwell, 432 pp.

Charney, J. G., 1947: The dynamics of long waves in a westerly baroclinic current. *J Meteor*, **4**, 135–163.

Charney, J. G., 1948: On the scale of atmospheric motions. *Geof Publ*, **17**, 251–265.

Charney, J. G., and M. E. Stern, 1962: On the Stability of Internal Baroclinic Jets in a Rotating Atmosphere. *J. Atmospheric Sci.*, **19**, 159–172, https://doi.org/10.1175/1520-0469(1962)019<0159:otsoib>2.0.co;2.

Corbosiero, K. L., J. Molinari, A. R. Aiyyer, and M. L. Black, 2006: The Structure and Evolution of Hurricane Elena (1985). Part II: Convective Asymmetries and Evidence for Vortex Rossby Waves. *Mon. Weather Rev.*, **134**, 3073–3091, https://doi.org/10.1175/mwr3250.1.

Didlake Jr., A. C., and Jr. Houze R. A., 2013: Dynamics of the Stratiform Sector of a Tropical Cyclone Rainband. *J. Atmospheric Sci.*, **70**, 1891–1911, https://doi.org/10.1175/jas-d-12-0245.1.

Dunion, J. P., C. D. Thorncroft, and C. S. Velden, 2014: The Tropical Cyclone Diurnal Cycle of Mature Hurricanes. *Mon. Weather Rev.*, **142**, 3900–3919, https://doi.org/10.1175/mwr-d-13-00191.1.

Eady, E. T., 1949: Long waves and cyclone waves. *Tellus*, **1**, 33–52.

Emanuel, K. A., 1984: A note on the stability of columnar vortices. *J. Fluid Mech.*, **145**, 235–238, https://doi.org/10.1017/S0022112084002895.

Ertel, H., 1942: Ein neuer hydrodynamischer Erhaltungssatz. *Naturwissenschaften*, **30**, 543–544, https://doi.org/10.1007/BF01475602.

Faller, A., 1963: An experimental study of the instability of the laminar Ekman boundary layer. *J Fluid Mech*, **15**, 560–576.

Faller, A. J., and R. E. Kaylor, 1966: A numerical study of the instability of the laminar Ekman boundary layer. *J Atmos Sci*, **23**, 466–480.

Faller, A. J., and R. Kaylor, 1967: Instability of the Ekman spiral with applications to the planetary boundary layers. *Phys. Fluids*, **10**, S212–S219, https://doi.org/10.1063/1.1762452.

Fang, J., and F. Zhang, 2012: Effect of beta shear on simulated tropical cyclones. *Mon. Weather Rev.*, **140**, 3327–3346, https://doi.org/10.1175/MWR-D-10-05021.1.

Fjørtoft, R., 1950: Application of integral theorems in deriving criteria of stability for laminar flows and for the baroclinic circular vortex. *Geofys Publ*, **17**, 1–52.

Ford, R., 1994: The instability of an axisymmetric vortex with monotonic potential vorticity in rotating shallow water. *J Fluid Mech*, **280**, 303–334.

Foster, R. C., 2005: Why rolls are prevalent in the hurricane boundary layer. *J. Atmospheric Sci.*, **62**, 2647–2661, https://doi.org/10.1175/JAS3475.1.

Gall, R., J. Tuttle, and P. Hildebrand, 1998: Small-Scale Spiral Bands Observed in Hurricanes Andrew, Hugo, and Erin. *Mon. Weather Rev.*, **126**, 1749–1766, https://doi.org/10.1175/1520-0493(1998)126<1749:sssboi>2.0.co;2.

Gregory, N., J. T. Stuart, W. S. Walker, and E. C. Bullard, 1955: On the stability of three-dimensional boundary layers with application to the flow due to a rotating disk. *Philos. Trans. R. Soc. Lond. Ser. Math. Phys. Sci.*, **248**, 155–199, https://doi.org/10.1098/rsta.1955.0013.

Guinn, T. A., and W. H. Schubert, 1993: Hurricane spiral bands. *J. Atmospheric Sci.*, **50**, 3380–3403, https://doi.org/10.1175/1520-0469(1993)050<3380:hsb>2.0.co;2.

Houze, R. A., Jr., 2010: Clouds in tropical cyclones. *Mon. Weather Rev.*, **138**, 293–344, https://doi.org/10.1175/2009mwr2989.1.

——, and Coauthors, 2006: The Hurricane Rainband and Intensity Change Experiment: Observations and Modeling of Hurricanes Katrina, Ophelia, and Rita. *Bull. Am. Meteorol. Soc.*, **87**, 1503–1522, https://doi.org/10.1175/bams-87-11-1503.

Kurihara, Y., and R. E. Tuleya, 1974: Structure of a tropical cyclone developed in a three-dimensional numerical simulation model. *J. Atmospheric Sci.*, **31**, 893–919, https://doi.org/10.1175/1520-0469(1974)031<0893:SOATCD>2.0.CO;2.

Le Dizes, S., M. Rossi, and H. K. Moffat, 1996: On the three-dimensional instability of elliptical vortex subjected to stretching. *Phys Fluids*, **8**, 2084–2090.

Leibovich, S., and K. Stewartson, 1983: A sufficient condition for the instability of columnar vortices. *J Fluid Mech*, **126**, 335–356.

Lilly, D. K., 1966: On the instability of Ekman boundary flow. *J Atmos Sci*, **23**, 481–494.

MacDonald, N. J. D.-:10. 1111/j. 2153-3490. 1968. tb00358. x, 1968: The evidence for the existence of Rossby-like waves in the hurricane vortex. *Tellus*, **20**, 138–150.

Marks, F. D., P. G. Black, M. T. Montgomery, and R. W. Burpee, 2008: Structure of the eye and eyewall of Hurricane Hugo (1989). *Mon. Weather Rev.*, **136**, 1237–1259, https://doi.org/10.1175/2007MWR2073.1.

Mashiko, W., 2005: Polygonal eyewall and mesovortices structure in a numerically simulated Typhoon Rusa. *SOLA*, **1**, 29–32, https://doi.org/10.2151/sola.2005-009.

Matsuno, T., 1966: Quasi-geostrophic motions in the equatorial area. *J Meteor Soc Jpn.*, **44**, 25–42.

Montgomery, M. T., and L. J. Shapiro, 1995: Generalized Charney–Stern and Fjortoft theorems for rapidly rotating vortices. *J. Atmospheric Sci.*, **52**, 1829–1833, https://doi.org/10.1175/1520-0469(1995)052<1829:gcaftf>2.0.co;2.

Montgomery, M. T., and R. J. D.-:10. 1002/qj. 49712353810 Kallenbach, 1997: A theory for vortex Rossby-waves and its application to spiral bands and intensity changes in hurricanes. *Q. J. R. Meteorol. Soc.*, **123**, 435–465.

Montgomery, M. T., V. A. Vlaririmov, and P. V. Denissenko, 2002: An experimental study on hurricane mesovortices. *J. Fluid Mech.*, **471**, 1–32, https://doi.org/10.1017/S0022112002001647.

Nolan, D. S., and J. A. Zhang, 2017: Spiral gravity waves radiating from tropical cyclones. *Geophy Res Lett*, **44**, 3924–3931.

O'Neill, M. E., D. Perez-Betancourt, and A. A. Wing, 2017: Accessible environments for diurnal-period waves in simulated tropical cyclones. *J. Atmospheric Sci.*, **74**, 2489–2502, https://doi.org/10.1175/JAS-D-16-0294.1.

Ortiz, F., 1947: *El Huracán: Su Mitologia u Sus Simbolos*. Fondo de Cultera Economica,.

Perez-Betancourt, D., 2019: Formation and maintenance of tropical cyclone spiral bands in idealized numerical simulations. Massachusetts Institute of Technology, 114 pp. https://dspace.mit.edu/handle/1721.1/121760.

Powell, M. D., 1990: Boundary layer structure and dynamics in outer hurricane rainbands. Part I: Mesoscale rainfall and kinematic structure. *Mon Wea Rev*, **118**, 891-917.

Robe, F. R., and K. Emanuel, 2001: The effect of vertical wind shear on radiative-convective equilibrium states. *J Atmos Sci*, **58**, 1427–1445.

Rotunno, R., 1978: A note on the stability of a cylindrical vortex sheet. *J Fluid Mech*, **87**, 761–771.

——, J. B. Klemp, and M. L. Weisman, 1988: A theory for strong, long-lived squall lines. *J Atmos Sci*, **45**, 463–485.

Schubert, W. H., M. T. Montgomery, R. K. Taft, T. A. Guinn, S. R. Fulton, J. P. Kossin, and J. P. Edwards, 1999: Polygonal eyewalls, asymmetric eye contraction, and potential vorticity mixing in hurricanes. *J Atmos Sci*, **56**, 1197–1223.

Senn, H. V., and H. W. Hiser, 1959: On the origin of hurricane spiral bands. *J. Meteorol.*, **16**, 419–426, https://doi.org/10.1175/1520-0469(1959)016<0419:otoohs>2.0.co;2.

Shapiro, L. J., 1983: The asymmetric boundary layer flow under a translating hurricane. *J. Atmospheric Sci.*, **40**, 1984–1998, https://doi.org/10.1175/1520-0469(1983)040<1984:tablfu>2.0.co;2.

Stern, M. E., 1960: Instability of Ekman Flow at large Taylor Number. *Tellus*, **12**, 399–417, https://doi.org/10.1111/j.2153-3490.1960.tb01327.x.

Stone, P. H., 1971: Baroclinic stability under non-hydrostatic conditions. *J. Fluid Mech.*, **45**, 659–671, https://doi.org/10.1017/S0022112071000260.

Strutt, J. W., ed., 1895: On the stability, or instability, of certain fluid motions. *Scientific Papers: Volume 1: 1869–1881*, Vol. 1 of *Cambridge Library Collection - Mathematics*, Cambridge University Press, 474–487.

Tang, J., J. A. Zhang, P. Chan, K. Hon, X. Lei, and Y. Wang, 2021: A direct aircraft observation of helical rolls in the tropical cyclone boundary layer. *Sci. Rep.*, **11**, 18771, https://doi.org/10.1038/s41598-021-97766-7.

Tatro, P., and E. Mollo-Christensen, 1967: Experiments on Ekman layer instability. *J Fluid Mech*, **28**, 531–543.

Wexler, H., 1947: Structure of hurricanes as determined by radar. *Am N Acad Sci*, **48**, 821–844.

Willoughby, H. E., 1988: The dynamics of the tropical hurricane core. *Aust Meteor Mag*, **36**, 183–191.

Wu, L., Q. Liu, and Y. Li, 2018: Prevalence of tornado-scale vortices in the tropical cyclone eyewall. *Proc. Natl. Acad. Sci.*, **115**, 8307–8310, https://doi.org/10.1073/pnas.1807217115.

Wurman, J., and J. Winslow, 1998: Intense Sub-Kilometer-Scale Boundary Layer Rolls Observed in Hurricane Fran. *Science*, **280**, 555–557, https://doi.org/10.1126/science.280.5363.555.

——, and K. Kosiba, 2018: The role of small-scale vortices in enhancing surface winds and damage in Hurricane Harvey (2017). *Mon. Weather Rev.*, **146**, 713–722, https://doi.org/10.1175/mwr-d-17-0327.1.

Zhong, W., and D.-L. Zhang, 2014: An Eigenfrequency Analysis of Mixed Rossby–Gravity Waves on Barotropic Vortices. *J. Atmospheric Sci.*, **71**, 2186–2203, https://doi.org/10.1175/JAS-D-13-0282.1.

12. Interaction of Tropical Cyclones with their Atmospheric Environment

Up until now we have mostly explored tropical cyclone physics under the assumption that the environment is stationary. In reality, the tropical environment is dynamic, as described in Part 1 of this book, with winds and thermodynamic properties varying in space and time. Here we explore how tropical cyclones interact with dynamic environments, affecting their translation, intensity, and structure. Some of the material in this chapter is covered in a nice review of the subject by Rios-Berrios et al. (2023).

12.1 The movement of tropical cyclones

To a fair approximation, tropical cyclones move with the background flow in which they are embedded. In the case of background wind that is constant in altitude, latitude, and longitude, the Galilean invariance of the operative physics dictates that the storm will just go with the flow. (On the other hand, the surface fluxes are not invariant with background flow as they do vary with ground-relative wind speed. We saw in Chapter 11, section 11.2, that background flow does tangibly affect the storm physics through its effect on surface fluxes.)

There are two basic complications to this simple view of tropical cyclone motion. First, the flow does typically vary with altitude, latitude, and/or longitude, and in that case it is not clear how to average the flow to obtain a meaningful “steering flow” or whether that is even possible. Second, the circulation associated with the cyclone alters the background flow via several mechanisms. Even if there is no background flow, the cyclone will eventually produce one and begin to move, seemingly of its own accord. Reviews of these physics are provided by Wang et al. (1998), Chan (2005) and Ito et al. (2020).

We should mention that even in the simple case of a steady barotropic flow, and neglecting the alteration to that flow by the cyclone itself, the motion of the storm can still present a forecasting challenge. This is illustrated in Figure 12.1, depicting a stationary barotropic flow with a tropical cyclone approaching a saddle point in the flow. In this case, a small error in the initial location of the cyclone can lead to radically different future trajectories, even though the background flow itself is perfectly predictable. This is an example of Lagrangian chaos (e.g. Crisanti et al. 1991). To this we must also account for the time-evolving chaos of meteorological flows, which will place a definite limit on the theoretical predictability of tropical cyclone motion, among other properties (Lorenz 1969).

12.1.1 Tropical cyclone motion in flows with vertical shear

If, as is usually the case, the flow varies with height, the problem of tropical cyclone motion becomes somewhat more challenging. Empirically, tropical cyclones appear to move with a weighted vertical average of the flow in which they are embedded (Marks 1992), though the weighting function seems to be a function of storm intensity (Lin et al. 2023). Theoretically, the problem is difficult because there is no clearly conservative variable that is merely being advected by the background flow, as would be the case with a simple, passive, barotropic vortex.

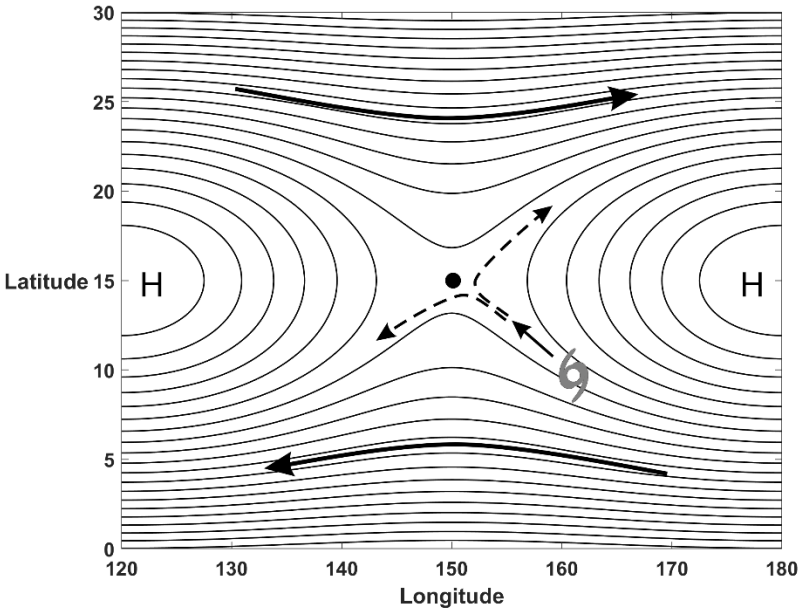


Figure 12.1: A tropical cyclone embedded in and moving with a stationary barotropic flow. Two high pressure systems, denoted by H, create a saddle point (black dot) in the streamlines. A small initial error in the tropical cyclone location can result in strong divergence of its future trajectory (dashed arrows); an illustration of Lagrangian chaos.

In Chapter 10 we idealized axisymmetric tropical cyclones as vortices with zero saturated potential vorticity. This condition is brought about my deep moist convection, which tends to drive density temperature lapse rates toward moist adiabatic along vortex lines, which in axisymmetric flows are equivalent to surfaces of constant angular momentum. To the extent that zero saturation potential is maintained, the properties of the cyclone are determined strictly by the distribution of entropy (or equivalent potential temperature) in the boundary layer. This is a cornerstone of our view of axisymmetric dynamics in an environment at rest with constant Coriolis parameter.

When the environmental flow has vertical shear, this framework runs into problems. First, as we shall see in section 12.2.1, the interaction of shear with the vortex produces asymmetric patterns of vertical motion, which can be strong enough to suppress deep convection where the motion is downward, producing anomalies of saturation potential vorticity. Secondly, as discussed in section 12.2.2, the interaction with shear can introduce unsaturated air into the storm's core, causing downdrafts that import relatively low entropy air unto the boundary layer. This produces asymmetries in boundary layer moist entropy and, through boundary layer quasi-equilibrium, asymmetries in deep convection. This may lead to partial or even complete decoupling of the surface vortex from the saturated air aloft. Without complete coupling, we are compelled to keep track of the separate trajectories of the (dry) potential vorticity and moisture anomalies. Unless they come back into alignment, the system is likely to dissipate.

Although deep convection and the secondary circulation of tropical cyclones act to make the vortex structure coherent through the depth of the troposphere, as a first step in understanding shear effects on tropical cyclone motion, it is helpful to take a step back and ignore the effects of the secondary circulation and deep convection, and consider the strictly dynamical coupling of the lower and upper troposphere. This coupling is manifest in quasi-balanced theory, in which

the potential vorticity is conserved and can be inverted in all three dimensions to recover the quasi-balanced part of the flow. A PV anomaly in the lower troposphere will "induce" a quasi-balanced flow in both the lower and in the upper troposphere, and vice versa. This dynamical coupling in the vertical is essential to understanding baroclinic dynamics, including baroclinic instability (Hoskins and James 2014). In fact, it is essential to the vertical coherence of nearly all vortices.

In the interests of keeping things simple while retaining the essential dynamics, we represent the tropical cyclone as a pair of potential vorticity anomalies stacked one on top of the other in a 2-layer, quasi-geostrophic fluid. In such a system, the quasi-geostrophic potential vorticity, q_i, in the *i*th layer, is related to the streamfunction ψ in both layers by

$$q_i = \nabla^2 \psi_i + \gamma\left(\psi_j - \psi_i\right) \quad \text{with } i = 1,2 \quad j = 3 - i, \tag{12.1}$$

where γ is the square root of the ratio of the radius of the vortex patches to the radius of deformation. There is a very rich literature, both in geophysics and in fluid dynamics, on the interaction of quasi-geostrophic vortices within a single layer and between two layers. The two-layer quasi-geostrophic formulation dates back to the work of Phillips (1954). Here we review applications of this framework to understanding the motion, and to some extent the structure of tropical cyclones in vertical shear flows.

A particularly simple example was developed and explored by Walsh and Pratt (1995) and Smith et al. (2000). They considered a flow with two cyclonic point potential vortices (in which the vorticity is a delta function in the horizontal plane), one in each layer. The upper of the two layers has a constant background zonal velocity U_2. A nice feature of point vortices is that, by symmetry, they cannot advect themselves; they can only advect each other and be advected by the background flow. The flow is also free of large-scale potential vorticity gradients in this example. The simple model is formulated as time rates-of-change of the positions of the two point vortices, in Cartesian coordinates represented as $\left(x_1, y_1\right)$ for the lower vortex and $\left(x_2, y_2\right)$ for the upper vortex. The lower vortex is advected by the upper vortex and vice versa, and the upper vortex is also advected by the constant background zonal flow. Walsh and Pratt (1995) and Smith et al. (2000) showed that the inversion, via (12.1), of the point potential vorticity associated with the two vortices leads to a coupled pair of ordinary differential equations for the differences between the meridional and zonal positions of the vortices in the two layers:

$$\frac{dX}{dt} = U_2 - \frac{\gamma\left(\Gamma_1 + \Gamma_2\right)}{X^2 + Y^2} Y, \tag{12.2}$$

and

$$\frac{dY}{dt} = \frac{\gamma\left(\Gamma_1 + \Gamma_2\right)}{X^2 + Y^2} X, \tag{12.3}$$

where $X \equiv x_2 - x_1, \quad Y \equiv y_2 - y_1$ and Γ_1 and Γ_2 are the circulation intensities of the two cyclonic point potential vortices. In addition, the centroid of the two vortices moves at a constant speed in the zonal direction:

$$\frac{dX_c}{dt} = \frac{\Gamma_2}{\Gamma_1 + \Gamma_2} U_2, \tag{12.4}$$

while the meridional position of the centroid of the upper- and lower-layer point vortices remains fixed at its initial value.

In essence, the whole system is moving eastward at a weighted vertical average of the background flow, where the weighting is by the circulation strength at each level. To the extent that this carries over to real tropical cyclones, whose circulation strength decays with altitude, we would expect the storms to move with a vertically averaged flow weighted toward the lower troposphere. In the case of intense tropical cyclones, whose circulation tends to penetrate deeper into the troposphere, we would expect the relevant steering flow to be weighted higher up than in the case of weak cyclones, as found empirically by Lin et al. (2023).

It is convenient to normalize the dependent and independent variables in (12.2) and (12.3) by

$$X, Y \to \frac{\gamma\left(\Gamma_1 + \Gamma_2\right)}{U_2}\left(X, Y\right), \quad t \to \frac{\gamma\left(\Gamma_1 + \Gamma_2\right)}{U_2^2} t. \tag{12.5}$$

With these normalizations, (12.2) and (12.3) become

$$\frac{dX}{dt} = 1 - \frac{Y}{X^2 + Y^2}, \tag{12.6}$$

and

$$\frac{dY}{dt} = \frac{X}{X^2 + Y^2}. \tag{12.7}$$

There are no coefficients in these equations, so that the entire system behavior collapses to the (nondimensional) relative initial positions of the upper and lower vortices. The nondimensional version of (12.4) for the centroid longitude is just

$$\frac{dX_c}{dt} = \frac{\Gamma_2}{\Gamma_1 + \Gamma_2}. \tag{12.8}$$

First note that $Y = 1, \quad X = 0$ is a particular, stationary solution of (12.6) and (12.7). This is an important special case that we will come back to. Relative to the lower-level vortex, the upper-level vortex is at the same longitude and displaced to the north (for eastward upper-level flow in the northern hemisphere) by one nondimensional unit, relative to the lower level vortex. In this special case, the flow associated with the upper-level vortex moves the lower-level vortex eastward, and the flow associated with the lower-level vortex retards the eastward motion of the upper-level vortex in just the right amount that the two vortices translate eastward at exactly the same speed, given by (12.8). If we linearize (12.6) and (12.7) about the stationary solution $Y = 1, \quad X = 0$, we see that this is an unstable equilibrium and so will not be observed.

We can solve (12.6) and (12.7) by integration in time. Alternatively, we can determine the solution by dividing (12.6) by (12.7) to obtain (after multiplying the result by X):

$$\frac{1}{2}\frac{d\left(X^2+Y^2\right)}{dY}=X^2+Y^2. \tag{12.9}$$

This can be solved analytically, yielding

$$X^2=ce^{2Y}-Y^2, \tag{12.10}$$

where c is any constant. A family of solutions of (12.10) for which X is real can be obtained by varying c. Note that the solutions are symmetric about $X=0$. Figure 12.2 shows the solutions (12.10) for a range of values of c.

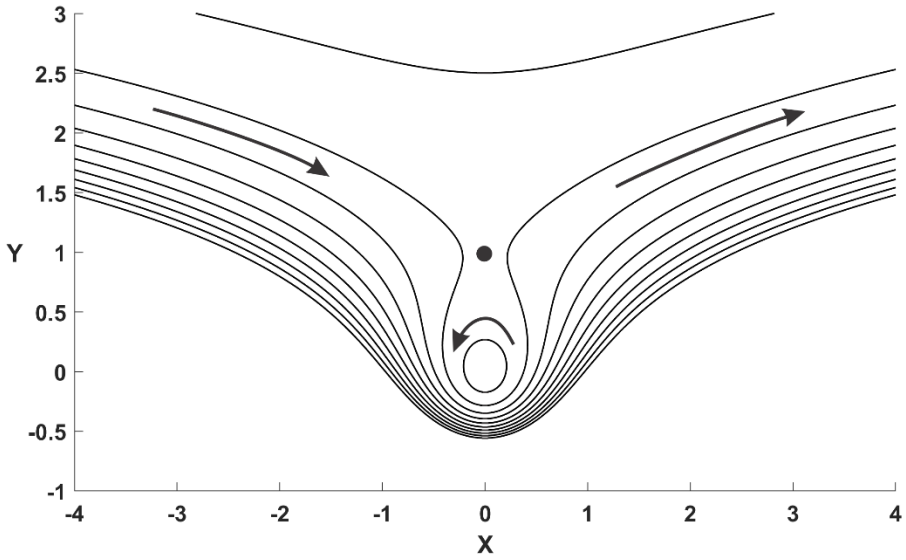


Figure 12.2: The solutions (12.10) for the (nondimensional) trajectory of the upper-level vortex relative to the lower-level vortex for a range of values of the constant c. The closed curves below the saddle point (black dot) at X=0 and Y=1 represent solutions for which the two vortices rotate about one another but do not diverge; in all the other solutions the upper-level vortex eventually diverges away from lower-level vortex.

The trajectories of the upper-level vortex with respect to the lower cyclone fall into two classes: Open trajectories in which the upper-level vortex is ultimately swept downshear away from its lower-level counterpart, and closed trajectories in which the two vortices rotate around each other. The streamline separating these two classes is the one that intersects $Y=1,\quad X=0$.

While Figure 12.2 depicts the motion of the upper cyclone with respect to the lower one, their ground-relative trajectories differ. Figure 12.3 shows the ground-relative motion for one of the oscillatory solutions and taking $\Gamma_1=\Gamma_2$ in (12.8).

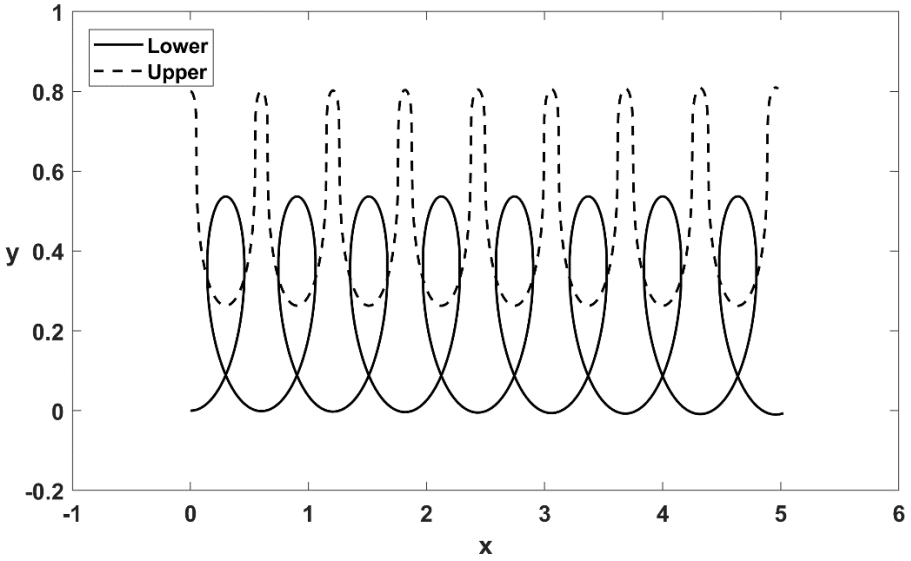


Figure 12.3: Trajectories in latitude and longitude of the lower (solid) and upper (dashed) vortices for $\Gamma_1 = \Gamma_2$ *, and* $X = 0, \quad Y = 0.8$ *at the initial time. The trajectories are run for 10 nondimensional time units.*

One limitation of using point vortices is that they are immutable. While they can be advected, their shape and magnitude are invariant. This proves to be a serious limitation in understanding the behavior of quasi-balanced vortices in shear.

In the absence of shear, like-signed point potential vortices can never spontaneously merge at the same level or become superimposed, one on top of the other, because to do so would violate conservation of energy. To see this, we first relate streamfunction and potential vorticity to the system energy by multiplying (12.1) by minus the streamfunction and integrating horizontally over an infinite plane, assuming that the streamfunction vanishes far from the point potential vortices. The result is

$$-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sum_{i=1,2} q_i\psi_i dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left\{\sum_{i=1,2}\left[\left(\frac{\partial\psi_i}{\partial x}\right)^2 + \left(\frac{\partial\psi_i}{\partial y}\right)^2\right] + \gamma\left(\psi_2 - \psi_1\right)^2\right\}dxdy = E, \qquad (12.11)$$

where E is the integrated energy of the system. The terms in the middle of the expression are the kinetic energy and potential energy, respectively. In the absence of external sources, the system energy E must be conserved. This equation applies to quasi-geostrophic systems in general, as long as they are adiabatic and frictionless.

Suppose we have two like-signed potential vortices separated by a large distance. We will suppose that the magnitudes of the vortices are equal and such that the integral on left side of (12.11) equals unity for each vortex separately. Since they are assumed to be far apart, the streamfunction of one vortex will be very tiny at the location of the other vortex, so can ignore that product of the PV of one vortex and the streamfunction of the other. Since we have two vortices, the total system energy is 2.

Now merge the two point potential vortices into a single vortex. Its potential vorticity is twice that of either of the original vortices, and likewise its streamfunction will have twice the magnitude,

so the integral in the left side of (12.11) will be 4 for the merged vortex[1]. This demonstrates that one would have to inject energy into the system to merge the point vortices. Absent that energy, there is no way for them to spontaneously merge.

When we have PV patches rather than points, the situation is quite different. It turns out that like-signed patches of potential vorticity anomalies at two levels will spontaneously align under certain conditions. Polvani (1991) examined such a system, introducing identical circular patches of potential vorticity at each of the two levels of a two-layer quasi-geostrophic model. Depending on the size of the patches, their initial separation, and the coupling parameter γ in (12.1), the patches may vertically align, as illustrated in Figure 12.4.

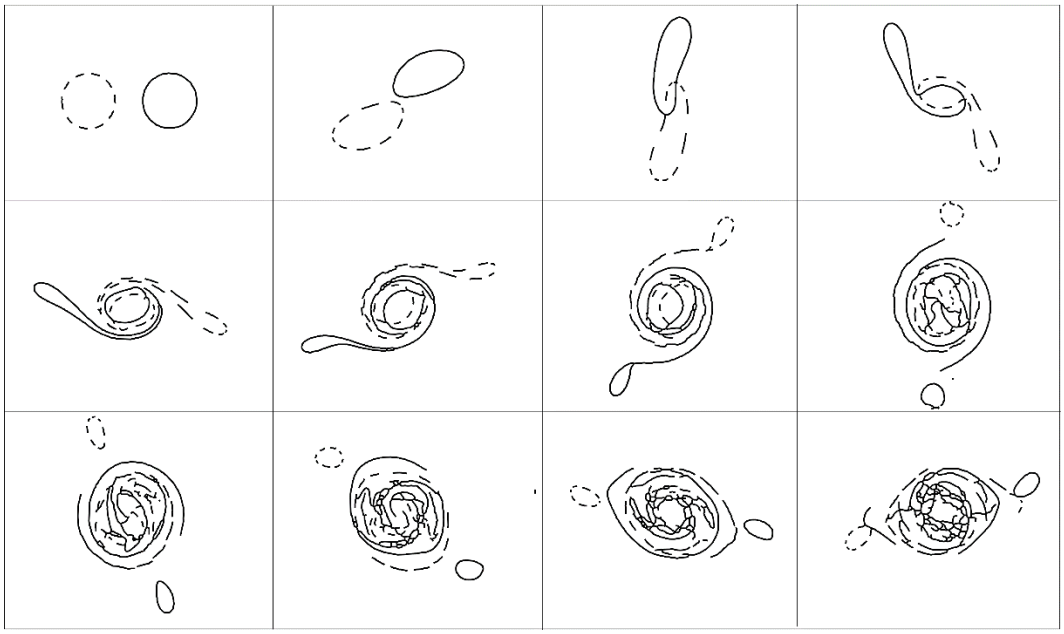


Figure 12.4: Evolution of two initially identical circular potential vorticity patches in the upper (solid) and lower (dashed) layers, separated horizontally by three times the deformation radius, for $\gamma = 2$. Time progresses rightward and downward in increments of 2 rotation periods. The single contour at each level separates high PV within the contour from low PV in the rest of the fluid.

In stark contrast to the case of point vortices, which rotate around one another in the absence of shear, these PV patches seem attracted to one another, and vertically align while shedding filaments of high PV. This shedding is essential to conserve energy during the superposition of the vortices.

Mutual rotation and possible merger can also take place horizontally, including in barotropic flows. The mutual cyclonic rotation of tropical cyclones was first described by Fujiwhara (1921) and is has been called the “Fujiwhara Effect” ever since. There is a rich literature on this subject. When two cyclonic, barotropic finite vortex patches of comparable magnitude interact, they rotate cyclonically around one another and will generally merge if the background flow brings them within a critical distance of one another, but if one is considerably stronger than the other the weaker vortex patch is instead deformed into filaments which do not merge with the

[1] In more general systems with finite, distributed potential vorticity, the more compact the distribution of like-signed potential vorticity, the larger the energy.

stronger vortex (Lander and Holland 1993; Ritchie and Holland 1993; Holland and Dietachmayer 1993).

The filamentation of the aligning PV patches is a fundamental process in vortex dynamics. The end-state quasi-circular blobs of PV have considerably less PV than they initially had, owing to the shedding of thin PV filaments. This is how the system manages to conserve energy while the vortices align. Owing to their geometry, the thin filaments have considerably less energy (left side of 12.11) than if the same volume of PV were in a circular patch. The inability of point vortices to shed PV filaments explains why they cannot spontaneously align.

If the vortices are too small, too far apart and/or if γ is too small, the vortex patches merely rotate around one another, as in the point vortex example in the absence of shear.

The two-layer quasi-geostrophic framework was extended to many layers by Reasor and Montgomery (2001), who affirmed the critical importance of the ability of the system to support vortex Rossby waves, which extract potential vorticity from the core of the vortex, allowing it to align vertically without violating energy conservation.

All of this brings us back to the motion of baroclinic vortices in a sheared environment. Of central importance is the question of vertical coherence: If the system is being sheared apart, it is difficult to define which piece of the system to track. Vertical coherence, even if it has an oscillatory component, alleviates this issue.

Research into the tendency for finite balanced vortices to vertically align was extended to the case with vertical wind shear by Reasor et al. (2004), using both a simple linearized model as well as a full primitive equations model. They showed that the vertical alignment process made possible by Rossby waves and filamentation can, under certain circumstances, allow fully three-dimensional tropical cyclone-like vortices to remain vertically aligned, if tilted, in the face of vertical wind shear. To a crude approximation, the aligned vortex behaves like the two-layer vortex at the stationary solution $Y = 1, \quad X = 0$ shown in Figure 12.2, except that this point is now stable thanks to the ability of the system to generate vortex Rossby waves and shed filaments.

Thus, to the extent that the two-layer point potential vortex model is any guide, tropical cyclones in vertically sheared flow should move with something like a circulation-weighted vertical average of the environmental flow, if conditions allow the system to remain coherent over time.

12.1.2 Effects of anticyclonic outflow on tropical cyclone motion

Up to this point in our discussion of tropical cyclone motion, we have assumed that the storms are cyclonic at all levels, whereas, in reality, the air circulation is anticyclonic near the storm top and far enough from the storm center. The perturbation vorticity in the outflow is of the opposite sign to that of the cyclonic flow (negative in the northern hemisphere). Although the outflow occupies a relatively thin layer, it can cover a large area and so can induce a non-trivial streamfunction response in the low troposphere.

To explore this effect, Wu and Emanuel (1993) developed a two-layer quasi-geostrophic model with a cyclonic point potential vortex in the lower layer and, horizontally co-located with this point vortex, a fixed source of zero potential vorticity air in the upper layer. This mimics the

outflow in tropical cyclones, which has approximately zero saturation potential vorticity. However, at the very low (< -60 C) temperatures of the upper troposphere, there is hardly any moisture, making the saturation and actual potential vorticity nearly equivalent.

The outflowing zero potential vorticity air in the upper troposphere is separated from the ambient, constant (nonzero) potential vorticity by a single contour. That contour is advected by the streamfunction determined by its shape (self-advection), by the upward influence of the point potential vortex in the lower troposphere, and by a constant, imposed background zonal wind to account for wind shear. (The background wind is zero in the lower layer.) The evolution of the upper level vortex is carried out using the technique of contour advection and surgery (Dritschel 1989) in which contour segments that fold back on themselves are reconnected…an irreversible process that can correspond to wave breaking. The lower layer point vortex is advected only by the flow induced by the upper-level potential vorticity distribution, as represented by the single contour.

In the case of no shear, the system remains axisymmetric and stationary, with a continuously expanding circular contour enclosing zero PV in the top layer. The application of shear from the west displaces the anticyclone downshear (eastward). The streamfunction induced in the lower layer by this anticyclone in the upper layer causes a poleward advection of the lower layer cyclone.

The location of the cyclone and the single PV contour in the upper layer is shown at 9 times in Figure 12.5 for a particular combination of nondimensional parameters.

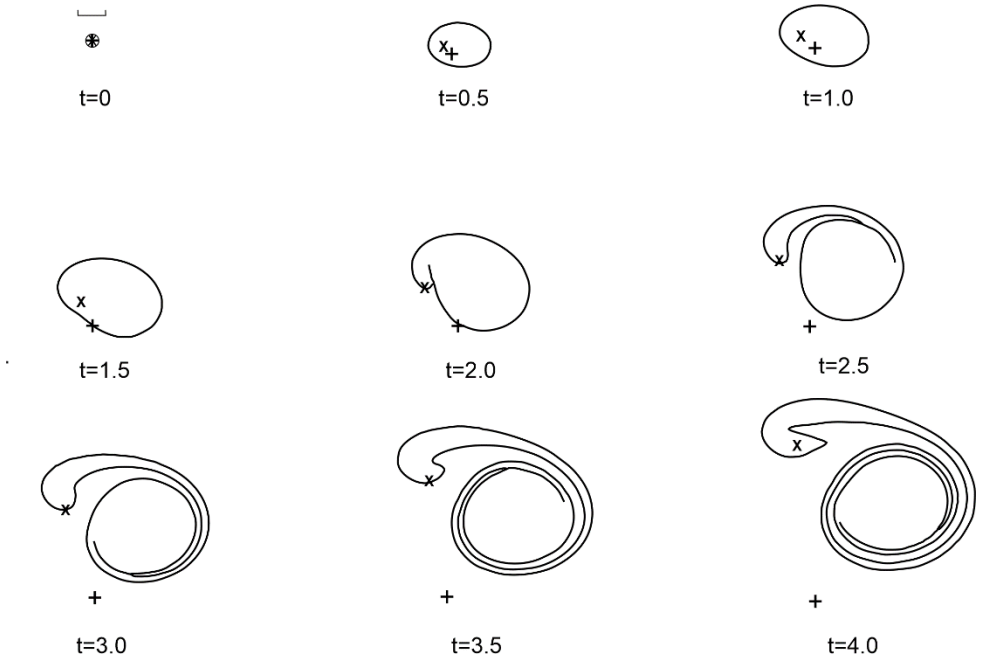


Figure 12.5: Position of the lower-layer cyclone, marked by x's, and the contour enclosing zero potential vorticity in the upper layer at 9 nondimensional times, for a particular choice of model parameters. The + signs show the initial position of the lower layer cyclone. The bracket in the upper left corner shows a scale of around 500 km for realistic values of the parameters.

As the upper anticyclone expands eastward, the circulation around the lower cyclone begins to advect it poleward and distort its shape. The cyclone is advected poleward and westward by the flow associated with the zero PV aloft. As time progresses, the influence of the cyclonic circulation on the upper anticyclone, as well as the self-interaction of the PV anomaly aloft, distorts the zero PV contour and causes it to roll up on itself.

With yet larger values of the imposed background shear, the upper anticyclone is swept so rapidly downshear that it has relatively little influence on the movement of the cyclone.

The trajectory of the surface cyclone is shown for three (nondimensional) values of the imposed westerly shear in Figure 12.6.

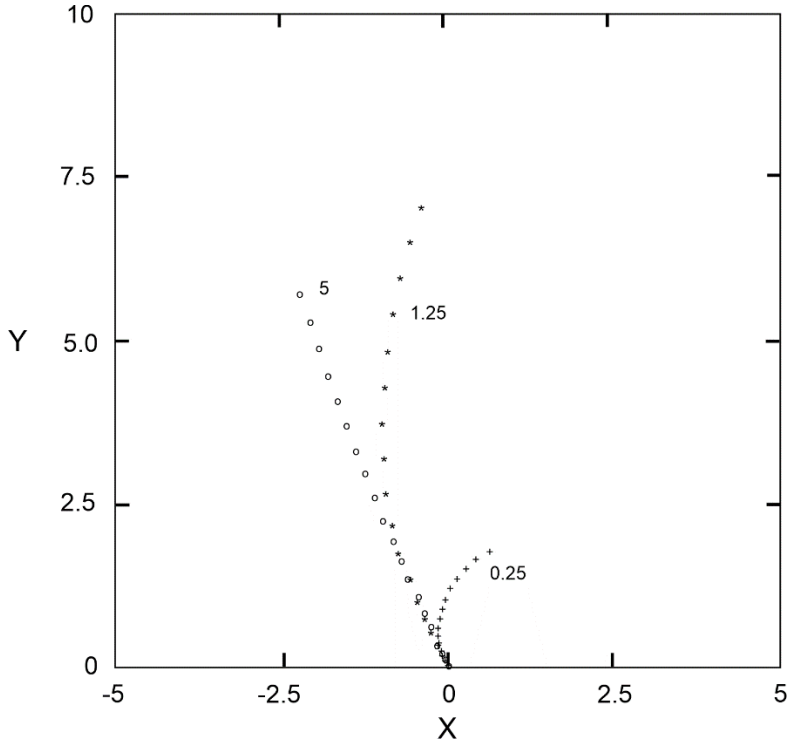


Figure 12.6: Trajectories of the lower layer cyclone for three different values of the imposed (nondimensional) west-to-east shear (numbers labeling curves). The points on the three curves are separated by equal increments of time. One nondimensional unit of X and Y is equivalent to roughly 500 km.

When the shear is quite weak, the cyclone drifts poleward and then eastward, quite slowly. Stronger shears elicit a stronger and more poleward drift, but only up to a point. The strongest shear among those driving the trajectories shown in Figure 12.6 produces weaker drift than in the case of intermediate strength shear. When the shear it strong, the anticyclonic PV is swept downshear too fast to have much effect on the lower-level cyclone.

When translated into dimensional values, the movement of the cyclone amounts to a few meters per second, or on the order of 100 km per day. We will see shortly that this is comparable to other mechanisms by which tropical cyclones alter their own steering flow.

12.1.3 Effects of environmental potential vorticity gradients

In Chapters 9 and 10 we developed axisymmetric models of tropical cyclones assuming they were embedded in a quiescent background and that the Coriolis parameter was constant. In nature, the atmosphere in which tropical cyclones are embedded contains gradients of vorticity that, among other things, support Rossby waves. The interaction of the strongly nonlinear vortical flows of tropical cyclones with background states that support Rossby waves can create dynamical responses that serve to advect the vortices relative to the unperturbed background flow and, perhaps, to drain some energy away from the cyclone.

In considering the dynamics of these interactions, it is helpful to remember that the tropical atmosphere in regions that support tropical cyclones is nearly moist adiabatic and is maintained in a state that, to a first approximation, is in radiative-convective equilibrium (RCE). In Chapter 7 we showed that small amplitude perturbations to RCE states near the equator tend to be confined to the first baroclinic mode, governed by shallow water equations, and a barotropic mode that satisfies the barotropic vorticity equation. We might imagine that the interaction of a tropical cyclone with its RCE environment might be describable, to a first approximation, in terms of these two modes interacting with a strong vortex. Historically, dating back to the work of Adem (1956), almost all studies of this interaction have focused on the barotropic component of this interaction. We will maintain that focus here but should bear in mind that the first baroclinic mode may also prove important (see, e.g., Korotaev 2022).

The simplest case treats a barotropic point vortex embedded in a constant background vorticity gradient, which may be owing to the meridional gradient of the Coriolis parameter and/or a cross-flow gradient of the relative vorticity associated with a background flow. The basic dynamics are illustrated in Figure 12.7.

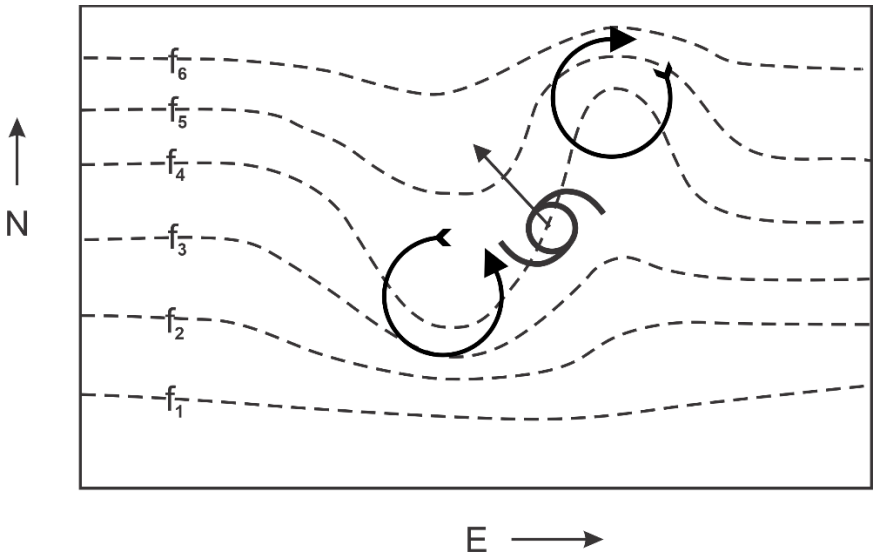


Figure 12.7: Barotropic effect of a northern hemisphere tropical cyclone on a planetary vorticity gradient. The flow around the cyclone advects the planetary vorticity in the manner shown, creating a negative anomaly to the northeast of the cyclone center, and a positive anomaly to its southwest. These anomalies are associated with anticyclonic and cyclonic flow anomalies, as shown, creating a northwestward advection of the cyclone's own vorticity. (In the southern hemisphere, the cyclone would move southwestward.)

Close to the cyclone center, the strong winds act on the background vorticity like a high-speed blender, axisymmeterizing the vorticity. But further from the vortex center, the slower deformation of the background absolute vorticity field results in an anticyclonic gyre poleward and eastward of the cyclone center, and a cyclonic gyre equatorward and westward of the center. These so-called “beta gyres” have poleward and westward flow across the cyclone center, advecting it in that direction, as shown in Figure 12.7.

After the work of Adem (1956), the problem of the motion of an isolated, barotropic vortex on a β-plane (constant gradient of planetary vorticity) attracted a great deal of attention from theorists (McWilliams and Flierl 1979; Holland 1983; Sutyrin 1988; Smith and Weber 1993; Sutyrin and Flierl 1994; Llewellyn Smith 1997; Schecter and Dubin 1999) and even some laboratory investigations (e.g. Firing and Beardsley 1976; Carnevale et al. 1991). The upshot of these investigation is that while the motion of isolated[2] cyclonic barotropic vortices on a β-plane is generally poleward and westward, it varies over time and depends on the structure and magnitude of the vortex relative to the background planetary vorticity gradient. An important nondimensional parameter influencing the motion is $\beta L^2 / V$, where β is the planetary vorticity gradient, and L and V are characteristic radial length and azimuthal velocity scales of the vortex.

Real tropical cyclones are baroclinic, and their upper regions rotate anticyclonically except very near the storm center. But anticyclonic vortices interact with the planetary vorticity gradient to produce westward and equatorward (rather than poleward) motion, so the immediate effect is to cause the upper anticyclone to move equatorward of the lower-level cyclone. This displacement of the upper anticyclone from the lower cyclone causes the two members to advect each other broadly eastward, countering, to some extent, the westward drift cause by the interaction of each with β. This effect is illustrated in Figure 12.8, which shows the trajectories of the streamfunction maximum in the lower layer of a 2-layer model with piecewise constant PV (Sutyrin and Morel 1997), as a function of the magnitude of the upper level PV patch relative to the resting state PV of the upper layer. The less cyclonic, or more anticyclonic, the upper vortex patch relative to the lower, the less westward the motion. Thus, compared to case of barotropic cyclones, the more realistic case of an anticyclone surmounting a cyclone tends to drift less westward and more poleward.

[2] In the fluid dynamical literature on vortex dynamics, an isolated vortex is one whose circulation vanishes at some finite radius.

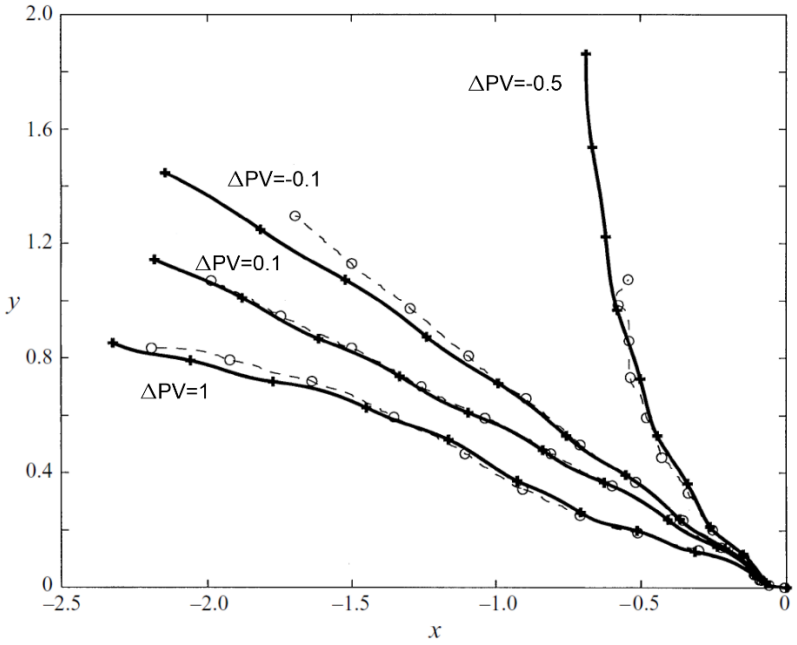


Figure 12.8: Trajectories in nondimensional Cartesian coordinates of the streamfunction maximum in the lower of two layers in a contour dynamics model. The curves are marked by the nondimensional difference between the PV inside the upper-level contour and the PV of the resting background state; the difference inside the lower layer contour is 1. The symbols on each curve are at equal increments of nondimensional time. The solid lines show approximate analytic solutions while the dashed curves show numerical solutions. The calculations are done over a time scale that is too short for appreciable radiation of Rossby waves to have occurred.

When the background flow itself has vorticity gradients, the net (planetary plus relative) vorticity (or potential vorticity) gradient is what matters for steering, but the flow itself introduces complications, including, for example, differentially advecting the storm-induced vorticity anomalies (Wang et al. 1998; Chan 2005; Ito et al. 2020). Moreover, vortices may radiate Rossby waves into the far environment, serving as a sink of vortex energy.

Combining the effects of vertical wind shear, upper level anticyclones, and background potential vorticity gradients (e.g. Dengler and Reeder 1997) results in a problem with so many interactions as to defy comprehension by most mortals. Yet understanding the individual processes and, at some level, their interaction is essential, not only to advancing the general understanding of tropical cyclone behavior, but to the development of improved forecast models.

12.1.4 Thermodynamic considerations

All the preceding discussion assumes that we are in a regime of dry dynamics in which one can assume that some form of potential vorticity is conserved and, under appropriate conditions, invertible to find the balanced flow. Steering, as such, was thought of as resulting from the background flow and the dynamical interaction of potential vorticity anomalies and gradients. Yet phase change of water is an important process in tropical cyclone dynamics and is a first-order source and sink of potential vorticity, rendering problematic this form of “PV thinking”.

In our investigation of axisymmetric tropical cyclone dynamics, we assumed that the saturation potential vorticity, given by (9.33), is always zero outside the eye, owing to the rapid adjustment to neutrality by slantwise moist convection. This has the gratifying effect of reducing the problem to the time-dependent spatial distribution of boundary layer moist entropy and to the distribution of vorticity near the tropopause.

Yet if this remained true in sheared, asymmetric tropical cyclones, then the cyclones would move with the boundary layer environmental winds. What we observe instead is that tropical cyclones move more nearly with a circulation-weighted vertical average of the environmental flow (Holland 1984), in agreement with elementary dynamics (subsection 12.1.1).

A key missing ingredient in the dynamics-only view of tropical cyclone motion is low- to mid-level moisture. As we shall see in section 12.2.2, the eyewall must remain nearly saturated through much of the troposphere for tropical cyclones to be maintained, otherwise evaporatively driven downdrafts quickly reduce boundary layer moist entropy. As shown in section 12.2.1, vertical shear interacting with a tropical cyclone produces a complex, asymmetric pattern of vertical motion, and it is often observed that deep moist convection disappears entirely from some sectors of the cyclone. In such regions, the assumption of slantwise convective neutrality fails and we cannot assume that the system moves with the environmental boundary layer winds.

We might infer from this that the disposition of lower tropospheric moisture, above the boundary layer, is of critical importance to tropical cyclone physics, including translation.

As a thought experiment, consider a tropical cyclone embedded in an environmental flow with a step function jump at the top of the boundary layer. Naively, ignoring complex nonlinear interactions, we might expect the moist column and vorticity above the boundary layer to be displaced downshear from the boundary layer vortex, leaving the latter exposed to a relatively dry column aloft and the moist column bereft of a low-level vortex. What might happen next?

For a tropical cyclone to re-form at the position of the low-level vortex, the whole column above it must first become nearly saturated (Chapter 14). This may take several days, whereas a new vortex can reform under the nearly saturated column aloft in less than a day (Emanuel 1989). Thus, we might expect the system as a whole to follow the low- to mid-level moisture (and vorticity) more than the surface vorticity in this case.

This is consistent with what seems to happen in at least some sheared tropical cyclones. For example, Molinari et al. (2006) showed that when Atlantic Tropical Storm Gabrielle of 2001 was exposed to strong wind shear, a new center reformed downshear of the original low-level vortex, following the region of strong convection (and presumably moisture) rather than the trajectory of the original low-level vortex.

This suggests that the motion of tropical cyclones in shear may depend at least as much on the disposition of column moisture as that of near-surface vorticity or boundary layer moist entropy. Shear also strongly affects storm intensity, as we shall discuss in section 12.2.

12.1.5 Forecasting tropical cyclone motion

In the time before numerical weather prediction and space-based observations, weather forecasters relied on climatology and sparse observations provided by surface stations, ships, and beginning in the 1930s, rawinsonde observations. Forecasts were often poor. For example,

while the Great New England Hurricane of 1938 was relatively well observed as it approached the southern U.S., forecasters predicted it would follow the normal climatology for similar storms and recurve out to sea before striking land. Instead, it roared into Long Island and New England with no warning, killing more than 600 people.

A large step forward was taken in the mid-1940s with the introduction of operational airborne hurricane reconnaissance in the North Atlantic and western North Pacific regions. Air crews rapidly became adept at finding and often penetrating the cores of tropical cyclones, but their estimates of storm intensity were at first based on visual observation of the state of the sea surface and were prone to large errors. The development of dropsondes – parachute-borne instrument packages deployed from the aircraft – led to much better estimates of surface pressure. Radar altimetry, introduced during World War II, together with direct pressure measurements, allowed aircraft to measure flight-level geopotential of pressure surfaces. The late 1950s saw the development and deployment of inertial navigation systems, which yielded more accurate storm center positions and allowed flight crews to estimate ground-relative speed. The vector difference between ground speed and airspeed gave the flight-level wind speed.

Progress in hurricane observations and track prediction was rapid during the 1970s and 80s. By the late 1970s, most tropical cyclones were being detected using satellite-borne visible and infrared imagery, and in 1982 Hurricane Debby became the first tropical cyclone to be surveyed by airborne Doppler radar, giving a three-dimensional view of the storm's wind field.

The first models used as guidance in hurricane forecasting were statistical, and it was not until the late 1980s that dynamical models began to be more skillful than statistical algorithms. The first statistical model used by NOAA's National Hurricane Center (NHC), deployed in 1969, was based on a combination of climatology and persistence. The first truly dynamical model to be used operationally as guidance to track forecasts was the Sanders Barotropic Tropical Cyclone Track Prediction Model (SANBAR; Burpee 2008), used by NHC from 1970 to 1989. This was based on a vertically averaged form of the vorticity equation, using initial conditions from operational analyses. By 1978, NOAA had developed high-resolution, storm-following regional primitive equation models driven by the output of global numerical weather prediction models.

These observational and modeling developments, guided in part by better understanding of the physics of tropical cyclone motion together with improving methods of assimilating observational data into numerical weather prediction models, led to much improved models and a steady increase in the skill of tropical cyclone track forecasts (Figure 12.9). Today, regional hurricane models are driven by ensembles of global forecasts, with the initial ensemble spread based on expected analysis errors. An example is provided in Figure 12.10 (Lin et al. 2020). Such forecasts provide measures of forecast uncertainty and represent a transition from storm-centric forecasts that focus on storm position to user-centric forecasts that emphasize threat probabilities.

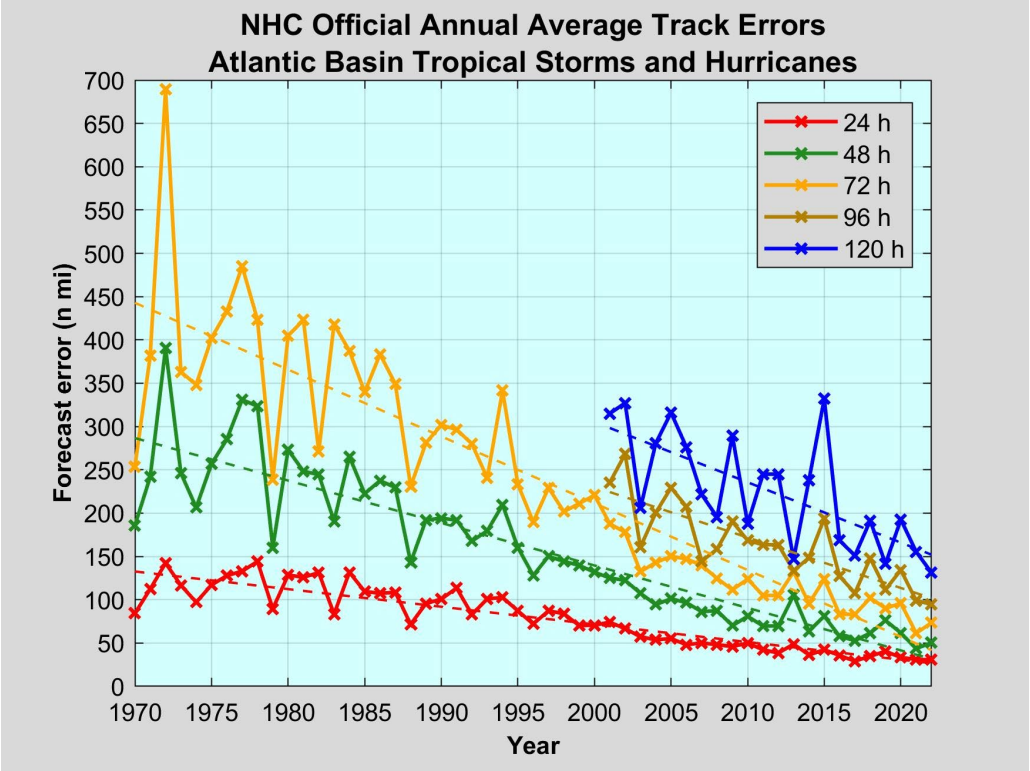


Figure 12.9: Annual average distance (nautical miles) between observed North Atlantic tropical cyclone center positions and those forecast by the NOAA National Hurricane Center for forecasts at 24 hrs (red), 48 hrs (Green), 72 hrs (orange), 96 hrs (brown), and 120 hrs (blue). Linear trends shown by dashed lines.

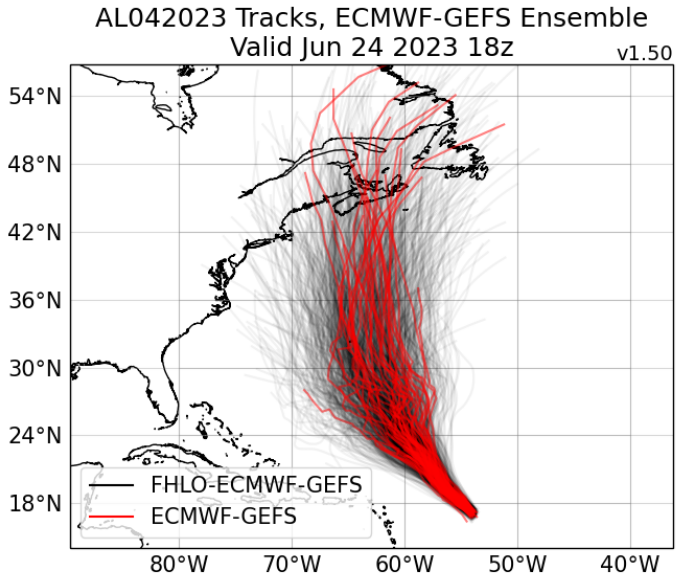


Figure 12.10: Real-time forecast tracks of the center of North Atlantic Tropical Storm Cindy of 2023. The red curves show a combination of 21 tracks of the storm center from NOAA's Global Ensemble Forecast System (GEFS) and 51 tracks from the European Center for Medium Range Forecasts (ECMWF) initialized at 18 GMT on June 24, 2023. The black curves are 1,000 forecast tracks statistically bootstrapped from the 72 global model forecast tracks. A separate specialized hurricane intensity model is applied to each of these 1,000 tracks. Based on Lin et al. (2020).

12.2 Effects of environmental wind shear on structure and intensity

The interaction of spatially varying environmental wind with tropical cyclones is complex and involves both dynamical and thermodynamic interactions. Observations leave no doubt that environmental wind shear has as strong effects on storm structure and intensity, as well as motion as reviewed in section 12.1. Asymmetries in the distribution of deep moist convection can be strong markers of shear effects, but the reader is cautioned against making causal inferences based on the distribution of convection. For example, weak convection ascending into anomalously warm air contributes to a positive correlation between heating and temperature and thus can lead to an increase in system kinetic energy, whereas strong deep convection ascending into anomalously cold air will reduce system kinetic energy.

We can, however, use deep convection as a diagnostic tool whose bulk properties are easily observed from satellite. We will continue to reply on the approximation of boundary layer quasi-equilibrium, and make use of (3.53), a diagnostic relationship for the deep convective updraft mass flux M_u :

$$M_u = \rho w_b + \frac{F_h}{h_b - h_m}, \tag{12.12}$$

where ρ is the air density, w_b the large-scale vertical velocity at the top of the boundary layer, F_h is the surface enthalpy flux, and h_b and h_m are the moist static energies of the boundary layer and lower to middle troposphere, respectively. This shows that deep convection is related to

1. Large-scale upward motion
2. Surface sensible and latent heat fluxes
3. Lower to middle level moisture, as represented by $h_b - h_m$.

External influences that alter vertical motion, surface enthalpy flux, and/or lower to middle tropospheric moisture should have a signature in the deep convective updraft mass flux. We begin by examining the dynamics controlling the interaction of intense cyclonic vortices with environmental wind shear.

12.2.1 Dynamics of shear-vortex interaction

In reality, tropical cyclones have strong secondary circulations and phase changes of water that destroy the conservation of ordinary (dry) potential vorticity, so useful in a first-order understanding of extratropical dynamics. Nevertheless, a basic understanding of the interaction of an isolated PV anomaly with background shear proves useful in understanding some aspects of tropical cyclone-shear interactions.

Consider first a quasi-geostrophic point potential vortex (a three-dimensional delta function of PV) within an infinite shear flow with otherwise constant PV, as illustrated in Figure 12.11.

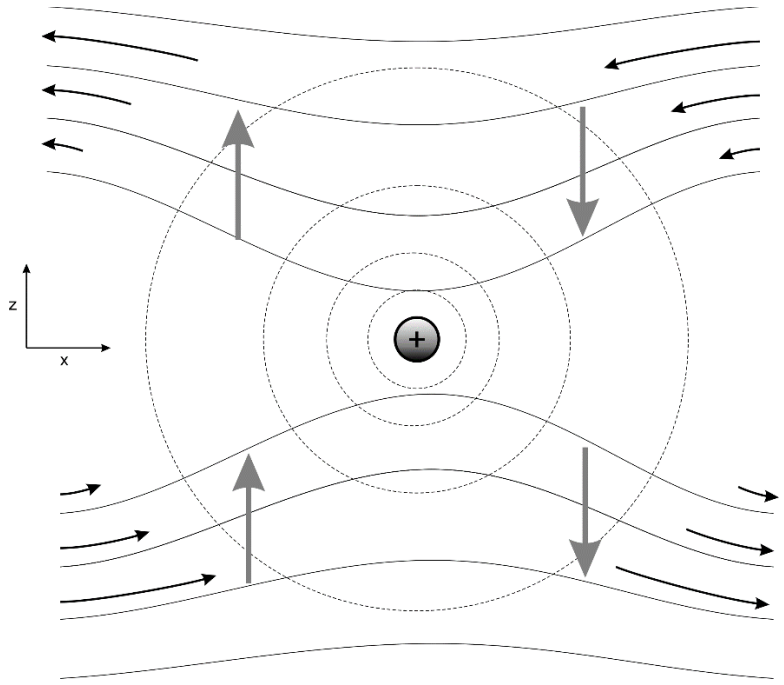


Figure 12.11: Cyclonic, quasi-geostrophic point potential vortex in the northern hemisphere, embedded in an easterly background shear flow. The horizontal coordinate has been scaled by the deformation radius. The delta function PV anomaly is at center; the solid curves represent surfaces of constant potential temperature, and the dashed circles are surfaces of constant geostrophic streamfunction. The black curved arrows represent the background shear flow deflected along potential temperature surfaces and the straight gray arrows represent the vertical velocity.

Since PV is conserved, the anomaly just moves with the mean flow at the same level, so we present the figure in a coordinate moving with that flow. In this coordinate the mean flow is zero at the level of the anomaly and all the fields are steady in time.

In quasi-geostrophic theory, and with horizontal coordinates scaled by the deformation radius, the geostrophic streamfunction ψ is related to the delta function PV anomaly at the origin by

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\psi=\delta(x)\delta(y)\delta(z). \tag{12.13}$$

The Green's function solution of this equation is

$$\psi=\frac{-1}{\sqrt{x^2+y^2+z^2}}, \tag{12.14}$$

and the meridional velocity and normalized potential temperature perturbations are then

$$V=\frac{\partial\psi}{\partial x}=\frac{x}{\left(x^2+y^2+z^2\right)^{3/2}}; \qquad \theta'\simeq\frac{\partial\psi}{\partial z}=\frac{z}{\left(x^2+y^2+z^2\right)^{3/2}}. \tag{12.15}$$

Inversion of the quasi-geostrophic cyclonic delta function of PV yields increased stratification near the PV anomaly, manifest as total potential temperature surfaces (including the background stratification, solid curves in Figure 12.11) squeezing together near the anomaly. It also yields cyclonic circulation that decays away from the anomaly.

Note that the approximations underlying quasi-geostrophic theory fail within roughly a deformation radius of the point vortex, because the vorticity owing to the curvature of the flow is comparable to or greater than the Coriolis parameter.

To conserve potential temperature, the easterly flow above the anomaly must descend to the east of the anomaly and ascend to its west; conversely, the westerly flow below the anomaly must ascend to the west and descend to the east. Additionally, the meridional flows associated with the PV anomaly advect the background gradient of potential temperature associated with the vertical shear, also leading to ascent downshear and descent upshear of the point potential vortex.

Likewise, the easterly flow above the anomaly experiences increasing relative vorticity as it approaches the anomaly from the east, and decreasing vorticity as it exits west of the PV anomaly. From the steady from of the quasi-geostrophic vorticity equation, this means that above the level of the anomaly, the stretching term in that equation, $f\frac{\partial w}{\partial z}$, must be positive to the east and negative to the west. Conversely, below the level of the PV anomaly, the stretching term is negative to the east and positive to the west. This implies that the vertical velocity, w, itself is negative everywhere east, or upshear, of the PV anomaly and positive to the west, or downshear, of the anomaly. This is consistent with our inference from the steady thermodynamic equation, as it must be in quasi-geostrophic theory. We can derive the vertical velocity from the steady states of either the quasi-geostrophic vorticity equation or the thermodynamic equation. We choose to use the latter here.

The dimensional form of the steady-state, adiabatic, quasi-geostrophic thermodynamic equation can be written for this constant shear flow

$$z\frac{\partial \overline{U}}{\partial z}\frac{\partial \theta'}{\partial x}+V\frac{\partial \overline{\theta}}{\partial y}+w\frac{\partial \overline{\theta}}{\partial z}=0, \tag{12.16}$$

where $\overline{\theta}$ refers to the background state. Using the thermal wind equation to replace $\partial \overline{U}/\partial z$, nondimensionalizing the result, and making use of the relations given in (12.15) gives, at $y=0$,

$$w=-\frac{\partial \overline{\theta}/\partial y}{\partial \overline{\theta}/\partial z}\frac{x\left(x^2+4z^2\right)}{\left(x^2+z^2\right)^{5/2}}. \tag{12.17}$$

Note that, unsurprisingly, the amplitude of the vertical motion is proportional to the slope of the background isentropic surfaces.

This function is plotted in Figure 12.12. Consistent with our qualitative inference from potential temperature conservation, there is descent upshear and ascent downshear of the point potential vortex, with no vertical motion at the longitude of the PV anomaly.

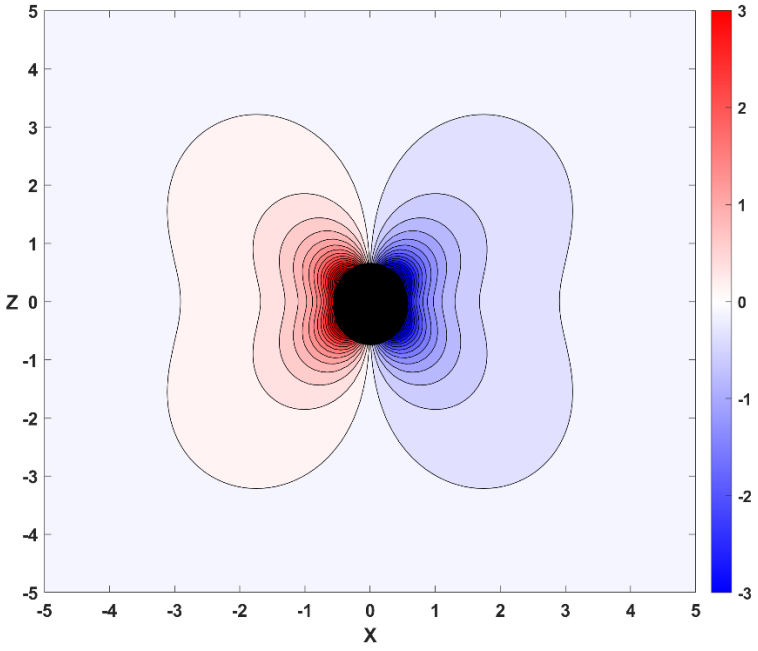


Figure 12.12: Pattern of vertical velocity around a three-dimensional point potential vortex at the origin embedded in a constant easterly shear flow, according to (12.17). This is graphed in the plane $y = 0$. *The horizontal coordinate is in deformation radii and the solution is blacked out inside 1 deformation radius.*

In general, we therefore expect upward motion, and from (12.12) enhanced deep convection, downshear of positive PV anomalies and suppressed deep convection upshear, assuming the other terms in (12.12) remain unchanged.

Note that the inversion operator to get from PV to the geostrophic streamfunction is linear, and so we can derive the total streamfunction by summing over a distribution of potential vorticity using its Green's function. We cannot do the same for the vertical motion, but must instead derive it from the quasi-geostrophic omega equation once the total distribution of streamfunction has been determined. Thus, even so simple a system as two-point potential vortices at different levels can have a vertical motion field that is too complex to visualize from first principles.

Figure 12.13 is a visible satellite image of Atlantic Tropical Storm Joaquin on September 29th, 2015. Almost all the deep convection is located downshear of the low-level circulation center, and deep convection in entirely suppressed upshear of the center, though bands of shallow convective clouds are visible. An example of a North Atlantic tropical cyclone exposed to northerly shear is shown in Figure 8.25.

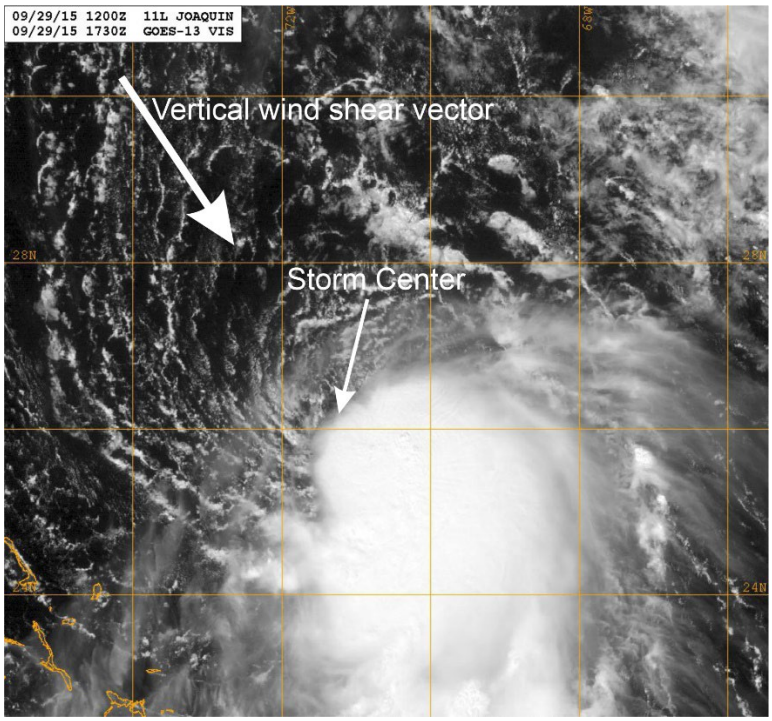


Figure 12.13: Visible satellite image of North Atlantic Tropical Storm Joaquin at 17:30 GMT on 29 September 2015. Arrows show the low-level circulation center and the direction of the tropospheric mean vertical wind shear vector at the time.

An important issue in tropical cyclone dynamics is the extent to which the cyclone can remain coherent over time in the presence of shear. Observations indicate that if the shear is too strong, or if the thermodynamic support is too weak, the locus of high column water vapor and deep convection becomes progressively separated from the low-level circulation. As mentioned in section 12.1.4, new low-level vortices sometime develop in the region of high column water vapor, but in other conditions the whole system decays over time.

We have already seen, in section 12.1.1, that patches of vorticity at two different levels can spontaneously align in the vertical in the absence of background shear. This work was expanded on and refined by Reasor and Montgomery (2001) and Schecter et al. (2002), who emphasized the importance of vortex Rossby waves and their damping in the vertical alignment process. To what extent can the dynamics of vertical alignment operate when vertical wind shear is present?

This problem was addressed by Reasor et al. (2004), who used a combination of theory and models to show that if the shear is not too strong, the alignment dynamics can serve to keep the system coherent despite the shear. The coherent vortex is tilted upward to the left of the shear vector, in the northern hemisphere, such that the shear induced by the vortex tilt locally cancels the imposed background shear, as illustrated in Figure 12.14.

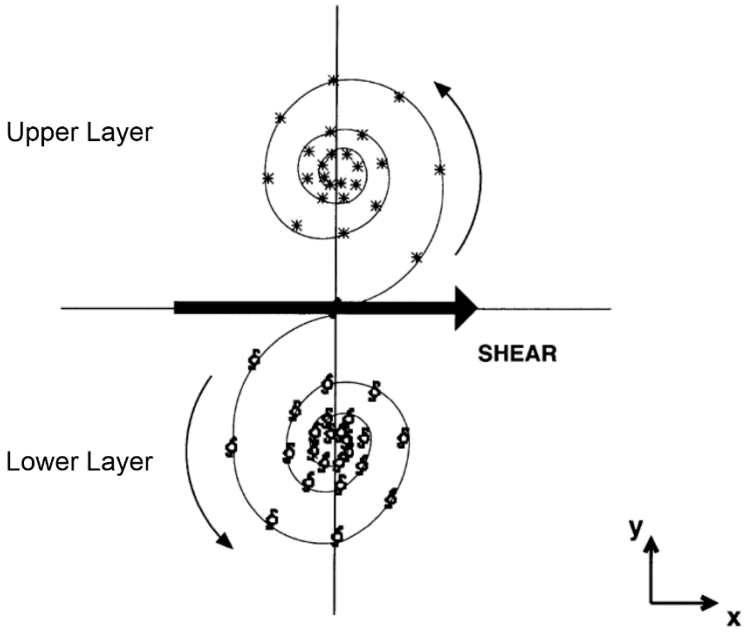


Figure 12.14: Alignment of a tilted cyclonic vortex in westerly shear flow in the northern hemisphere, in a simulation in which the vortex Rossby waves are damped. The lower layer vortex is at bottom while the upper layer vortex is at top. The spiral curves show the trajectories over time of the two vortex centers; the symbols are equally spaced in time. The circulation around the lower vortex, projected into the upper layer, moves the upper vortex westward against the background shear; conversely, the circulation around the upper vortex, projected into the lower layer, moves the lower vortex eastward so that the two vortices move coherently. A similar simulation but without damping of the vortex Rossby waves leads to corotation of the two vortices, as in Figure 12.3.

12.2.2 Thermodynamics of shear-vortex interaction

The axisymmetric theory of tropical cyclone intensification and maintenance assumes that the only strong influences on inner core boundary layer moist entropy are surface fluxes, dissipative heating, and radial advection. In reality, asymmetric features can act to import air with relatively low moist entropy into the core of the storm, reducing the all-important increase in boundary layer moist entropy in the inflow and thereby reducing the storm's intensity. This process was termed "ventilation" by Simpson and Riehl (1958), who analyzed mid-level airflow through two Atlantic hurricanes based on aircraft observations.

As we have seen in the previous subsection, the interaction of a vertically sheared flow with a strong vortex results in strong asymmetric patterns of both horizontal and vertical velocity. Although difficult to visualize, the tilted stable alignment of just two potential vorticity anomalies at different levels, as illustrated in Figure 12.14, results in complex three-dimensional trajectories that may (and often do) result in low moist entropy air from the middle tropospheric environment winding up in the boundary layer in the inner core, decreasing the magnitude of the radial moist entropy gradient and thus the wind speeds. In addition to the asymmetric vertical motion and the flow associated with the two vortex cores, these trajectories may be modified by vortex Rossby waves and by precipitating convection. In particular, evaporation of rain into relatively dry, low entropy air will produce strong downdrafts that carry the low entropy air into the boundary layer.

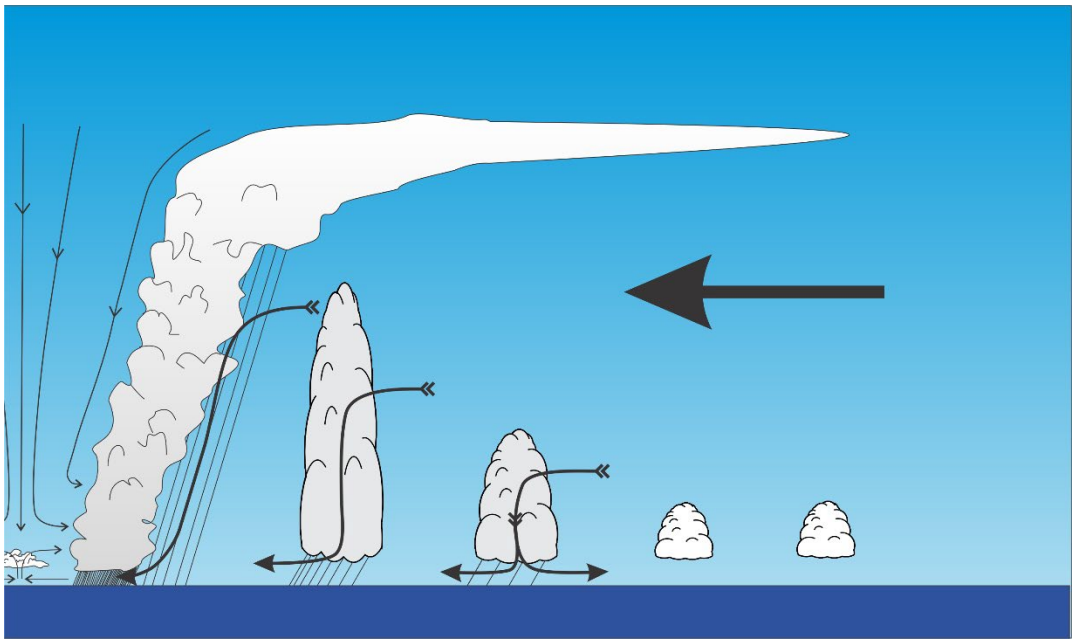


Figure 12.15: Interaction of a tropical cyclone with a sheared environment. While this is illustrated in a radial-vertical cross-section, the actual airflow is strongly three-dimensional. The thick horizontal arrow symbolizes an inward radial eddy flux of low moist entropy air from the middle tropospheric environment by asymmetric air flows of the type illustrated in Figure 12.14. The curved black arrows show possible routes taken by low entropy parcels, which may ultimately be cooled by evaporation of rain, sinking into the boundary layer either in rainbands outside the core or in the core itself. In both cases, downdrafts usually reduce the gradient in moist entropy from the storm core outward, weakening the circulation.

This process is illustrated schematically in Figure 12.15. A shear-induced inward eddy flux of low moist entropy air from the middle tropospheric environment results in some trajectories that sink into the boundary layer after being cooled by the evaporation of rain in deep convection. Moist entropy is nearly conserved along these trajectories, so that the low moist entropy air from the middle tropospheric environment ultimately winds up in the boundary layer, usually reducing the inward increase of moist entropy arising from surface fluxes and dissipative heating. (A case can be made that reducing the boundary layer entropy outside the core might actually increase the radial entropy gradient across the core.) Although Figure 12.15 is presented in a radial-vertical plane, it must be remembered that the trajectories of air parcels are strongly three-dimensional.

A quantitative theory for the effect of these asymmetric fluxes on the energy cycle was developed by Brian Tang (Tang 2010; Tang and Emanuel 2010, 2012). Here we present a somewhat simplified version of it.

Recall from Chapter 9 that the energy budget of an (unsheared) axisymmetric tropical cyclone can be expressed (combining (9.24) with (9.17) and (9.22))

$$C_D \mid \mathbf{V_s} \mid^3 = \frac{T_b - T_o}{T_b}\left[C_k \mid \mathbf{V_s} \mid \left(h_0^* - h_s\right) + C_D \mid \mathbf{V_s} \mid^3\right], \tag{12.18}$$

where we have set the coefficient of dissipative heating to unity. The left-hand side is the frictional dissipation in the boundary layer while the right-hand side is the thermodynamic efficiency multiplied by the net enthalpy source in the boundary layer, which is surface fluxes

and dissipative heating. (We have replaced the enthalpies in the surface fluxes with moist static energies, since the potential energy contributions cancel out.) Tang (2010) suggested that the effect of an inward and ultimately downward flux of low entropy (or moist static energy) air could be represented as proportional to the product of the vertical wind shear and the difference between the eyewall moist static energy and the moist static energy, h_m, of the mid-tropospheric environment:

$$F_{down} = -cV_{shear}\left(h_s - h_m\right), \tag{12.19}$$

where c is some (positive) nondimensional constant, V_{shear} is a measure of the magnitude of the environmental wind shear between low and middle levels of the troposphere, and we have assumed that the moist static energy of the eyewall equals the boundary layer moist static energy under the eyewall. The minus sign indicates that this is a negative enthalpy flux into the boundary layer. Inserting this on the right side of (12.18) as another contribution to the boundary layer moist static energy budget gives

$$C_D \mid \mathbf{V_s} \mid^3 = \frac{T_b - T_o}{T_b}\left[C_k \mid \mathbf{V_s} \mid \left(h_0^* - h_s\right) + C_D \mid \mathbf{V_s} \mid^3 - cV_{shear}\left(h_s - h_m\right)\right], \tag{12.20}$$

Collecting the $\mid \mathbf{V_s} \mid^3$ terms, dividing by C_D, and making use of the definition of potential intensity V_p given by (9.26) yields

$$\mid \mathbf{V_s} \mid^3 = V_p^2 \mid \mathbf{V_s} \mid - \frac{c}{C_D}\frac{T_b - T_o}{T_o}V_s\left(h_s - h_m\right). \tag{12.21}$$

If we now normalize $\mid \mathbf{V_s} \mid$ by V_p, $\mid \mathbf{V_s} \mid \rightarrow V_p V$, and once again make use of the definition of potential intensity (9.26), this may be written

$$V^3 - V - \Lambda = 0, \tag{12.22}$$

where the "ventilation parameter" Λ is defined

$$\Lambda \equiv \frac{V_s}{V_p}\frac{c}{C_k}\frac{h_s - h_m}{h_0^* - h_s}. \tag{12.23}$$

This will prove to be an important nondimensional measure of the thermodynamic effect of wind shear and was first identified by Tang (2010) and Rappin et al. (2010). If there is no shear, then $\Lambda = 0$ and V is either 0 or 1, as expected. Otherwise, depending on the value of Λ, there will be either no real roots of (12.22) or two real roots that coalesce to one root at a particular value of Λ. It is straightforward to show that there are no real roots of (12.22) if $\Lambda > 1/\sqrt{3} - 1/\sqrt{3}^3$, and at this critical value of Λ, $V = 1/\sqrt{3}$. The general solution of (12.22) is shown in Figure 12.16.

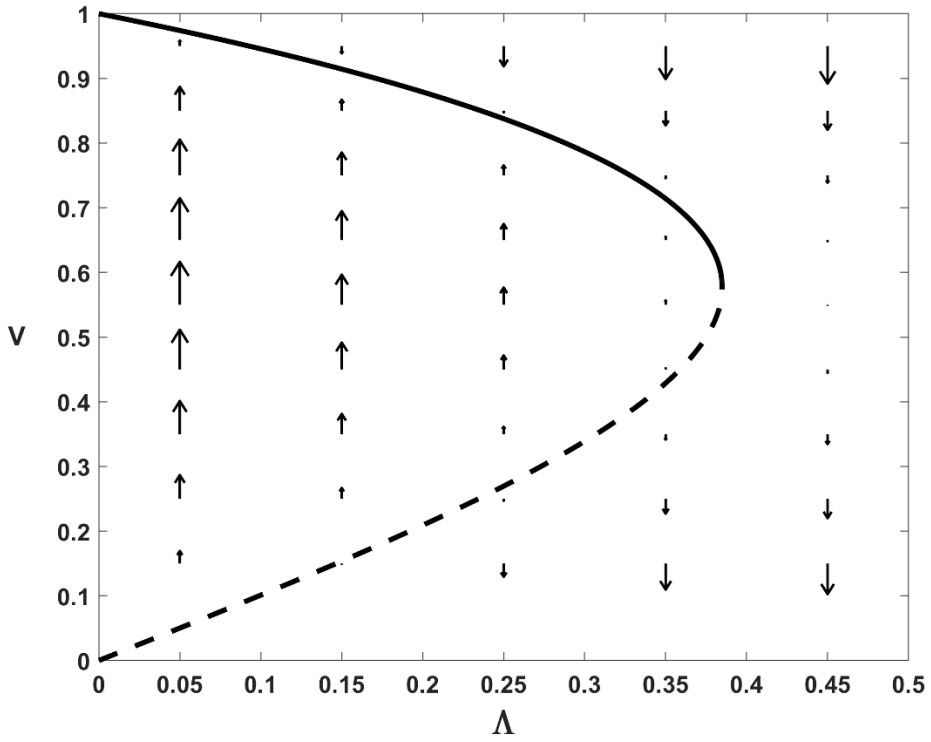


Figure 12.16: The solid and dashed curves show the two roots of (12.22) for subcritical values of the nondimensional ventilation parameter Λ. The arrows show minus the residual of (12.22) for a set of points in the $\Lambda - V$ plane.

The lower and upper roots are shown by dashed and solid curves, respectively, and the arrows show the sign and magnitude of minus the residual of (12.22) for a set of specified points in the $\Lambda - V$ plane. Upward-pointing arrows signify that energy generation by the thermal cycle exceeds dissipation, so that the vortex is amplifying. Clearly the lower root of (12.22) is unstable. The upper root shows that the steady-state wind intensity declines with increasing normalized shear and does not exist beyond the critical value of Λ.

This very simple treatment assumes, among other things, that the potential intensity itself is unaffected by ventilation. But Tang (2010) showed that reduction of eyewall enthalpy also leads to an increase in the outflow temperature, since parcels in the eyewall do not rise as high. Consequently, the thermodynamic efficiency and the potential intensity are reduced.

The solution of (12.22) shown in Figure 12.16 can be compared to observations of tropical cyclones using a standard historical TC data (e.g. IBTrACS, Knapp et al. 2010). Tang (2010) considered observations collected between 1981 and 2000, equatorward of 25^o latitude during July-October in the northern hemisphere and January-March in the southern hemisphere, noting their 24-hour wind intensity change, normalized by potential intensity. He used reanalysis data from the European Center for Medium Range Forecasts (ERA-Interim) to estimate the value of the ventilation parameter Λ. The result is displayed in Figure 12.17. There is a fair correspondence between the predicted response in Figure 12.16 and the observed response in Figure 12.17, suggesting that ventilation is indeed an important predictor of tropical cyclone intensification.

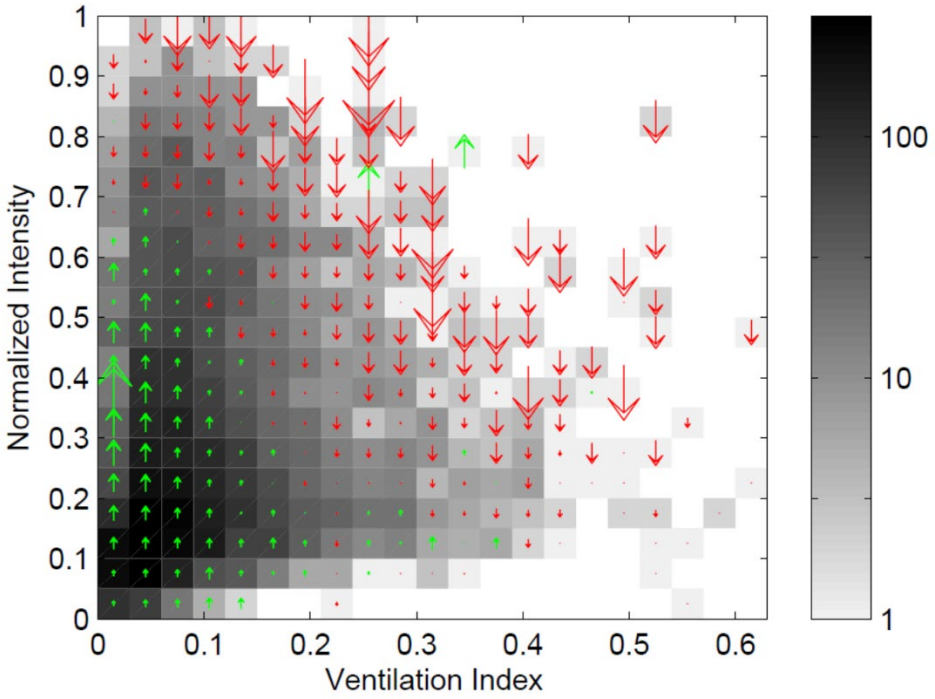


Figure 12.17: Number of daily tropical cyclone observations (shading) as a function of the normalized wind intensity and the ventilation parameter, together with 24-hour changes in normalized intensity (arrows). The maximum arrow length corresponds to a normalized intensity change of 0.4 over 24 hours, with intensification denoted by green arrows and dissipation by red arrows. The ventilation index has been calculated using ERA-interim data and the tropical cyclone data is from the best-track archive; see text for details.

While the addition of a single term to the tropical cyclone energy cycle (12.20) looks very simple, it belies a highly complex and fascinating account of what really happens in three dimensions and in time. Riemer et al. (2010, 2013), Alland et al. (2021a,b) and others conducted experiments with full-physics, three-dimensional numerical models at high spatial resolution to find out how low moist entropy air from the mid-tropospheric environment makes its way into the inner core of tropical cyclones. While there are several qualitatively different routes, Figure 12.18 summarizes the main pathways for low entropy air to ultimately lower the boundary layer moist entropy in the core. In association with the westerly shear, ascent at low levels is focused downshear and left (in the northern hemisphere) of the shear vector; in this case, to the northeast of the low-level vortex. Higher up (not shown), dry mid-tropospheric air enters the ascent region and rain evaporates into it, causing downdrafts just counterclockwise of the region of strongest ascent, bringing low moist entropy air into the boundary layer. As this low-entropy air circulates around the low-level vortex center, its moist entropy increases because of surface fluxes. Some of this air exits the system to the south, but most of the rest spirals into the eyewall as its moist entropy continues to increase. Although the moist entropy may ultimately reach values approaching those in unsheared storms in the same thermodynamic environment, it does so over a smaller area.

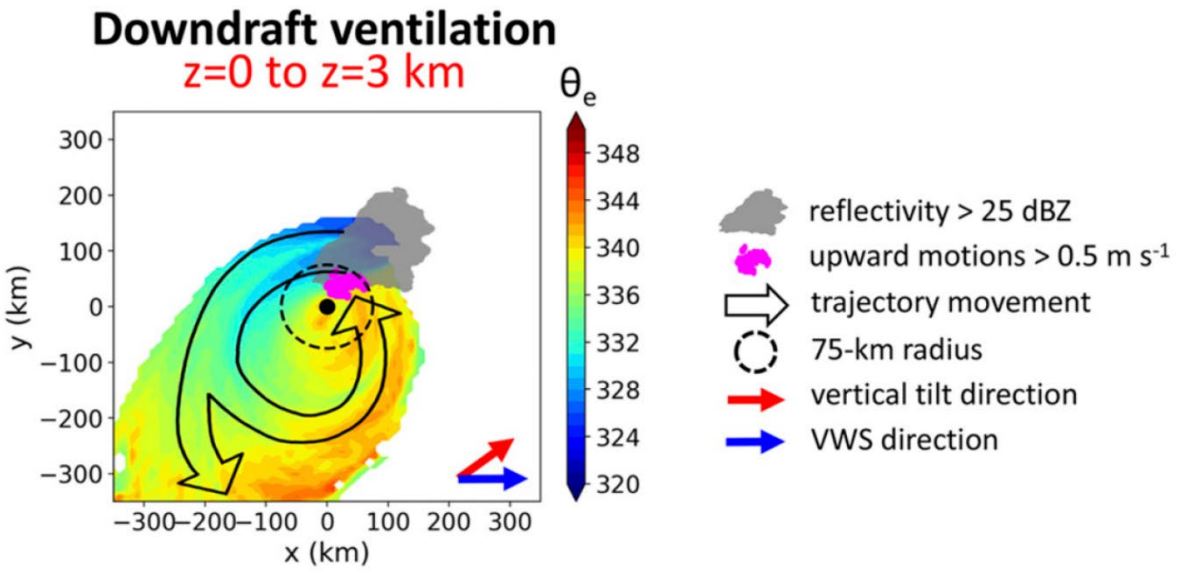


Figure 12.18: Summary sketch of downdraft ventilation of the boundary layers of numerically simulated tropical cyclones in a vertically sheared flow. The low-level circulation center is indicated by the black dot, and the colored shading shows boundary layer equivalent potential temperature. The shear vector is from the west and the vortex tilts upward to the northeast, as indicated by the red arrow at lower right. The gray shading shows high radar reflectivity and the magenta shading shows a region of strong upward motion at low levels. The open black arrows show airflow along trajectories.

Across all these idealized simulations of tropical cyclones in simple vertically sheared flows, the ultimate intensity, as measured by peak surface wind speeds, decreases with shear and with the dryness of the mid-tropospheric environment (i.e., with the ventilation parameter). Alland et al. (2021a) showed that the peak intensity is highly correlated with the inner core vertical mass flux and argued that the direct effect of the asymmetrical flushing of the boundary layer with low moist entropy air was to reduce or eliminate vertical mass flux in certain quadrants of the storm, thereby limiting storm intensity.

A practical consequence of shear interactions with tropical cyclones is to reduce the predictability of their intensity. Figure 12.19 shows the evolution of peak winds in idealized numerical experiments using a full-physics, convection-permitting, three-dimensional numerical model with four values of an imposed environmental wind shear (Zhang and Tao 2013). Twenty simulations were carried out for each value of environmental wind shear; these were identical to one another except for small differences in the moisture content of the inner core in the initial state. With even small imposed shear, the ensemble members show large spread during the period of most rapid intensification of the wind field, and the timing of intensification is exquisitely sensitive to the shear magnitude. Further experiments with the same model (Tao and Zhang 2015) show that 7.5 ms^{-1} of shear suffices to prevent intensification altogether, thus, ironically, restoring a high level of predictability. Given both practical and theoretical limits on the predictability of the larger scale environmental flow, sensitivity of tropical cyclone intensity to environmental shear places a strong limit on the predictability of tropical cyclone intensity (Emanuel and Zhang 2016, 2017).

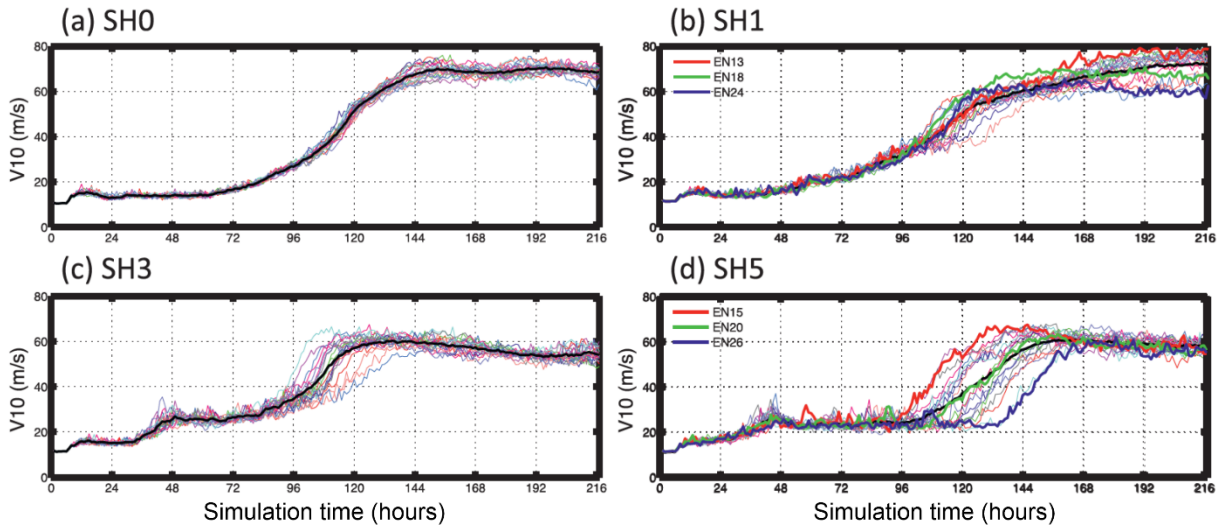


Figure 12.19: Peak surface (10 meter) wind speed as a function of time in simulations performed with a fully-physics, convection-permitting numerical model for four different values of an imposed environmental wind shear across the troposphere: none (a), 1 m s^{-1} (b), 3 m s^{-1} (c) and 5 m s^{-1} (d). Twenty ensemble members were run for each shear value, differing only by small perturbations of the initial inner core water vapor content.

12.3 Interaction with environmental potential vorticity

Here we treat two rather different ways that tropical cyclones interact dynamically with their environment: First, through their interaction with large-scale, quasi-linear distributions of potential vorticity, and second, through their interaction with more isolated potential vorticity anomalies in their environments.

12.3.1 Radiation of Rossby waves and other wave types from tropical cyclones

In section 12.1 we learned that the interaction of tropical cyclone circulation with large-scale gradients of potential vorticity in their environments leads to vortex motion, generally poleward and westward. The slow, flow-relative westward motion excites Rossby waves that carry energy away from the cyclone center, thereby weakening it. Beginning with the work of Flierl (1977), there is a rich literature on the decay of idealized vortices on $\beta-$planes. Much of this work was motivated by the desire to understand eddies shed from ocean currents like the Gulf Stream. Flierl (1977) showed that weak (linear) vortices on a $\beta-$plane move westward and decay quite rapidly by radiation of Rossby waves. But stronger disturbances can remain coherent for many rotation periods and their-spin down is dominated by Ekman pumping rather than radiation (McWilliams and Flierl 1979). Most of the work on Rossby wave radiation from vortices assumes that the vortices have non-zero circulation at infinite radius. (Point vortices are examples of such disturbances.) But actual hurricanes have wind fields that vanish at some finite radius (see Chapter 9, section 9.3) and as such have zero circulation there. Circulations of finite radial extent (sometimes referred to as “isolated vortices”) were shown by Carnevale et al. (1991) to behave differently from unbounded vortices, with a distinctly different beta-drift and much less Rossby wave radiation. The upshot of these studies is that Rossby wave radiation is probably not an important sink of energy in tropical cyclones, except perhaps in their formative stages when they are weak, or under synoptic-scale flow regimes that are especially conducive to

Rossby wave radiation. A nice summary of the theoretical and laboratory work on this subject is provided by Llewellyn Smith (1996). This is not a closed chapter in tropical cyclone research and more needs to be done to understand Rossby wave radiation from real tropical cyclones.

More recent research suggest that Rossby waves radiating from existing tropical cyclones might, under the right conditions, trigger the development of new tropical cyclones (Krouse et al. 2008; Krouse and Sobel 2010).

Rossby waves are not the only waves radiated by tropical cyclones. High frequency fluctuations in the storm cores and in association with individual convective events generate a rich spectrum of inertia-gravity waves that propagate laterally and vertically (e.g. Nolan and Zhang 2017). Curiously, some of these waves appear to be synchronized to the diurnal cycle (Dunion et al. 2014; O'Neill et al. 2017). Tropical cyclones also generate detectable, infrasonic atmospheric pressure waves (Hetzer et al. 2009), ocean acoustic signals (Wilson and Makris 2008), microseisms (Feng and Chen 2022), and acoustic-gravity waves in the ionosphere (Zakharov and Sigachev 2022). None of these higher frequency waves is regarded as a significant energy sink for tropical cyclones. But the creation and radiation of ocean surface waves may be a significant drain on wind energy, and will be described in more detail in Chapter 13.

12.3.2 Interaction of tropical cyclones with localized environmental potential vorticity anomalies

In section 12.1.1, we briefly described the interaction of two tropical cyclones, producing a cyclonic rotation of each storm around a common axis. Tropical cyclones also interact with weather systems in the upper troposphere. There are two reasons to be especially interested in such interactions. First, as we have argued all along in this book, in deep convective regimes the strong tendency toward moist convective neutrality ties balanced flows firmly to boundary layer moist entropy anomalies, and to temperature anomalies at the top of the adjusted layer (usually the tropopause). Potential vorticity anomalies can and do exist at and above the tropopause and can modulate quasi-balanced flows deep in the troposphere. Second, the characteristic surfaces of quasi-balanced flows are not vertical lines but rather vortex lines, which in axisymmetric systems lie along surfaces of absolute angular momentum. Whether or not tropical cyclones are strongly axisymmetric, their vortex lines can and do flare out to large radii (> 1000 km) at the storm top. External influences in the high troposphere can act along these characteristic surfaces to influence the surface circulation.

A hint of this can be detected in the thermal wind equation for slantwise neutral vortices (9.40), which we repeat here:

$$\frac{1}{r^2} = \frac{1}{r_o^2} - 2\left(T - T_o\right)\frac{ds*}{dM^2}, \tag{12.24}$$

where r_o denotes the radii of angular momentum surfaces in the upper troposphere. In evaluating (12.24) we usually take r_o to be very large so that the first term on the right side can be neglected. But by transporting angular momentum inward in the upper troposphere, external disturbances have the potential of reducing the effective value of r_o, increasing the intensity of the surface circulation.

How far away from the surface center could an upper tropospheric potential vorticity anomaly exert some influence on the storm core? A rough estimate can be made by using conservation

of angular momentum along a streamline originating at the top of the boundary layer at the radius of maximum winds. Given a radius of maximum winds r_m, maximum wind V_m, and an azimuthal wind V_o is the upper troposphere along the streamline in a steady state, conservation of angular momentum yields

$$r_m V_m \simeq r_o V_o + \frac{1}{2} f r_0^2, \tag{12.25}$$

where f is the Coriolis parameter, r_o is the radius of the streamline in the upper troposphere, and we have neglected the Coriolis term at the radius of maximum winds. We can solve this quadratic equation of get an estimate of r_o. Given a maximum wind speed of $60\, ms^{-1}$, a radius of maximum winds of $50\, km$, a Coriolis parameter of $5 \times 10^{-5}\, s^{-1}$, and an anticyclonic flow speed in the upper troposphere of $-10\, ms^{-1}$, the radius of the angular momentum surface would be about 600 km. We conclude that near-tropopause potential vorticity anomalies quite far away from the surface cyclone center could influence inner core surface conditions.

Some insight about what might happen can be gained by looking at the system energy from a potential vorticity perspective. The quasi-geostrophic equations have a superposition principle very similar to the two-layer shallow water version we derived as (12.11), but which is continuous in the vertical as well as horizontal dimensions. We do not derive that here but just state the result:

$$-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} q_g \psi_g \, dxdydz = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\left(\frac{\partial \psi_g}{\partial x}\right)^2 + \left(\frac{\partial \psi_g}{\partial y}\right)^2 + \frac{f^2}{N^2}\left(\frac{\partial \psi_g}{\partial z}\right)^2\right] dxdydz = E, \tag{12.26}$$

where ψ_g is the quasi-geostrophic streamfunction, q_g is the quasi-geostrophic potential vorticity, f is the Coriolis parameter, and N is the buoyancy frequency of the basic state (and can be, at most, a function of the vertical coordinate z). The last term in brackets on the right side of (12.26) is proportional to the square of the temperature perturbation away from the basic state and represents the potential energy of the perturbation. This version of the principle is for an infinite fluid; if there are horizontal boundaries and there are temperature perturbations on those boundaries, then there are additional terms in the superposition principle.

In spite of the added complexity, the interpretation is the same as for the two-layer shallow water version, (12.11): Bringing two like-signed potential vorticity anomalies into closer proximity increases the perturbation energy, E. This can happen (as in the two-layer system) when a cyclonic potential vorticity anomaly in the upper troposphere moves closer to the lower tropospheric potential vorticity anomaly associated with the cyclone itself. The maximum perturbation energy will occur when the upper tropospheric anomaly is directly overhead of the surface cyclone, assuming that potential vorticity is conserved for both anomalies.

There is a higher order approximation than quasi-geostrophy, from which we can learn something important for the interaction of upper tropospheric potential vorticity anomalies with tropical cyclones. This higher-order system is called semi-geostrophy (Hoskins and Bretherton 1972). Unlike quasi-geostrophy, semi-geostrophy retains the ageostrophic advection of both geostrophic momentum and potential temperature. This makes the equations quite complex, but

when they are phrased in a geostrophic flow-dependent coordinate system, called geostrophic coordinates, they reduce to a mathematical system almost (but not quite) identical to quasi-geostrophy[3]. As with quasi-geostrophy, the inversion operator is a three-dimensional Laplacian, but in the transformed coordinate. In this coordinate system "vertical" is actually upward along vortex lines of the geostrophic flow. We might imagine that in even more nonlinear systems like tropical cyclones, the characteristic surfaces are vortex lines of the balanced flow. We can prove that this is the case in strong axisymmetric vortices, for which the vortex lines lie along surfaces of absolute angular momentum.

As we have already seen, angular momentum surfaces can flare out to large radii in the upper troposphere, so superposition in more nonlinear but balanced flow is not in the vertical, but along the vortex lines of the balanced flow. Consequently, strong superposition can potentially happen when the near-tropopause potential vorticity anomaly is many hundreds of kilometers away from the surface cyclone.

If the atmosphere into which a near-tropopause disturbance is advected is not far from a state of radiative-convective equilibrium, then the effective value of N^2 in (12.26) is likely to be closer to $N^2\left(1-\epsilon_p\right)$, where ϵ_p is the precipitation efficiency, following the developments in Chapter 7. Because this is much smaller than the dry static stability, the vertical penetration depth should be somewhat larger when inverting a potential vorticity anomaly. In the tropical cyclone core, the precipitation efficiency is usually quite large, so that a near-tropopause potential vorticity anomaly that does not cause much a of surface signature in the normal RCE state might nevertheless have a significant surface signal in the TC core. Consequently, as a near-tropopause potential vorticity anomaly approaches an existing cyclone, the response may be quite abrupt once the anomaly starts to project down along vortex lines that reach the core of the tropical cyclone.

Early researchers, though they may not have been explicitly aware of the superposition principle, found empirically, from sparse observations, that tropical cyclones might be triggered, or an existing cyclone amplified, by interaction with suitable disturbances in the upper troposphere. For example, Riehl (1948) suggested that superposition of the low-latitude portions of extra-tropical upper-level troughs and easterly waves seemed to play a role in the genesis of some western North Atlantic tropical cyclones. Pfeffer (1955) suggested that inward transport of angular momentum by eddies could play a role in the intensification of tropical cyclones. Such an inward eddy flux is consistent with the superposition of near-tropopause potential vorticity with a surface cyclone.

Riehl (1950), while noting that tropical cyclones tend to form underneath upper tropospheric cyclones, was mainly concerned with how upper atmospheric disturbances could "channel" tropical cyclone outflow to provide an exit route for air ascending ion the cyclone's core. Many subsequent papers (e.g. Miller 1958; Colon and Nightingale 1963; Yanai 1968; Sadler 1976) pursued this line of thought concerning the role of upper level disturbances. With hindsight, the existence of outflow channels would not seem to be necessary for cyclone intensification, given that tropical cyclones have no trouble spinning up in axisymmetric models, where the constraint of inertial stability is particularly strong. Even so, most of these authors cited the presence of an

[3] The difference is that the semi-geostrophic version of potential vorticity is advected vertically as well as horizontally.

upper tropospheric trough westward and poleward of the surface cyclone, suggesting that superposition effects might have been active in the cases they examined.

Returning to the superposition idea, Hawkins and Rubsam (1968), presented the interesting case of the genesis of Atlantic Hurricane Hilda of 1964 from an African easterly wave. The genesis took place as a rather small-scale upper-level trough progressing southeastward from near Florida approached an easterly wave entering the northeast Caribbean. As the tropical cyclone intensified, its diabatic heating wiped out the near-tropopause potential vorticity anomaly that triggered it. Molinari and Vollaro (1989) showed that a period of significant intensification of Atlantic Hurricane Elena of 1985 coincided with a large inward cyclonic eddy angular momentum flux associated with the passage of a middle latitude trough north of the hurricane. The region of large eddy angular momentum flux tracked inward over a period of about a day, and when it reached the storm core, an extended period of rapid pressure falls followed. This is consistent with the superposition over time of a cyclonic near-tropopause potential vorticity anomaly with a tropical cyclone, taking into account the reduction of the effective static stability in the cyclone core.

A more systematic study of near-tropopause influences was undertaken by Reilly (1992), who examined cases of tropical cyclone intensification in the western North Pacific region. This region has the advantage of a high density of commercial aircraft activity, increasing the detection probability of small-scale features in the near-tropopause environment and, therefore, the likelihood that such features would be captured in reanalyses. He found that most pre-cyclonic disturbances were exposed to positive upper tropospheric potential vorticity advection just before being identified as tropical cyclones.

Hanley et al. (2001) undertook a study of upper-air interactions with all named Atlantic tropical cyclones between 1985 and 1996. In 78% of these cases, cyclonic superposition led to intensification of the surface winds and pressure field, provided the tropical cyclone was over relatively warm water. In the composite of these cases, intensification began soon after a small-scale upper-tropospheric PV maximum approached the cyclone center. The potential vorticity maxima subsequently weaken, most likely due to diabatic heating, and never cross over the cyclone centers. They found that if the approaching near-tropopause potential vorticity anomaly is too large and/or intense, the tropical cyclone is more likely to weaken than to amplify, probably because of the stronger vertical wind shear associated with the larger or stronger potential vorticity anomaly. This raises the question of whether there is an optimal size and strength of near-tropopause potential vorticity anomalies in terms of their potential for intensifying tropical cyclones.

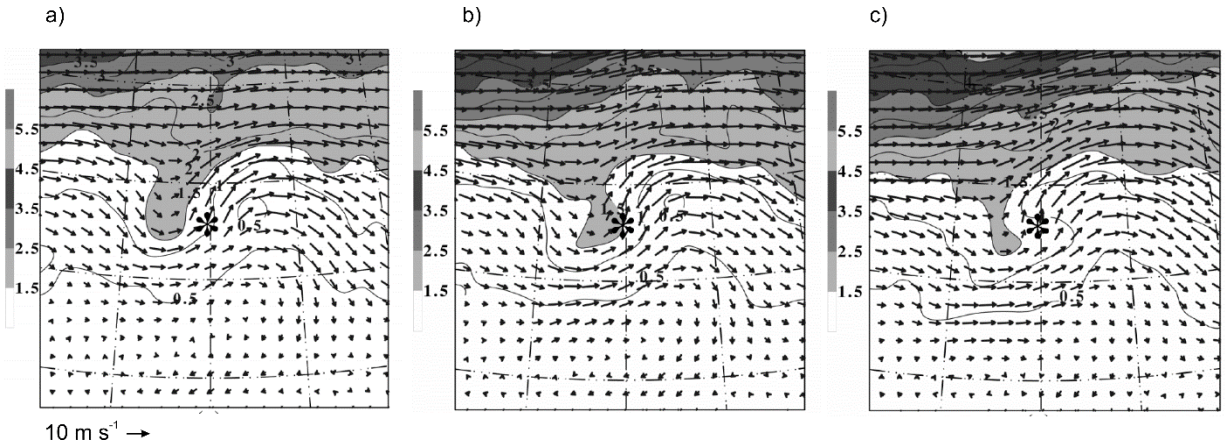


Figure 12.20: Composited 200 hPa charts showing Ertel's potential vorticity (increment is 0.5 PVU and values greater than 1.5 PVU are shaded as indicated) and horizontal wind vectors (scale shown at lower left). The composite is for 24 tropical cyclones that underwent total central surface pressure falls of at least 10 hPa after the central compositing time. The surface cyclone center position is shown by the asterisk. The panels are for 12 hours before the central time (a), the central time (b) and 12 hours after the central time (c). The increment in latitude and longitude is 10°.

A composite over 24 cases in which tropical cyclones deepened by more than 10 hPa after interacting with a near-tropopause potential vorticity anomaly is shown in Figure 12.20, extending over 24 hours centered on the time at which the central surface pressure began to fall. A very shortwave trough approaches the cyclone from the west, and close to the time to its nearest approach, the surface cyclone begins to intensify. The strong heating in the storm core destroys potential vorticity near the tropopause, giving rise to the illusion that the trough begins to move westward (panel *c* of Figure 12.10). These are transient interactions, with the cyclonic PV anomaly dissipating and/or progressing eastward away from the surface cyclone.

The complex interaction between tropical cyclones (or pre-cyclone disturbances) and near-tropopause potential vorticity anomalies remains a challenging and interesting research problem. Among many outstanding problems are the effects of (external) anticyclonic near-tropopause potential vorticity anomalies, and whether the current observational system is adequate for detecting small-scale but potentially influential near-tropopause potential vorticity anomalies.

12.4 Extratropical and tropical transitions

The interactions between tropical cyclones and near-tropopause potential vorticity anomalies described in the previous section are hit-and-run affairs in which the near-tropopause anomaly usually takes a beating, either dissipating or running off, somewhat diminished, to the east. The upper anomaly is small enough in scale that it is generally associated with very shallow baroclinicity in the high troposphere. But some poleward-moving tropical cyclones become embedded in the deep baroclinic flows of middle and higher latitudes, and begin to tap more substantially into the available potential energy of those flows, gradually taking on more and more of the characteristics of extratropical cyclones, developing fronts and highly asymmetric winds and precipitation. This process is known as *extratropical transition*. Famous examples include the Great New England Hurricane of 1938 (Brooks 1939; Avilés 2018) and Hurricane Hazel of 1954 (Gifford 2004). The Great New England Hurricane began as a "Cape Verde

Hurricane' making its way westward from the neighborhood of the Azores. As it approached the southeast U.S., it turned northward and interacted with a strong, deep baroclinic trough approaching from the west. The cyclone accelerated northward and reached a translation speed of close to $30\,ms^{-1}$ as it slammed into New England without any warning. Fatalities numbered over 600, and a wind gust of more than $83\,ms^{-1}$ was recorded by the Blue Hill Observatory, situated on a low hill just south of Boston, Massachusetts. This was considerably stronger than the storm's maximum wind while it was in the tropics, though Blue Hill has an elevation of 190 m. Hurricane Hazel of 1954 made landfall near the South Carolina – North Carolina border as a Category 4 storm with peak surface winds near $55\,ms^{-1}$. But rather than dying down at the usual rate as it progressed northwestward, Hurricane Hazel maintained very strong winds as it transitioned into an extratropical cyclone, retaining hurricane-force winds into Ontario, Canada, where it caused much damage.

During extratropical transition, the energetics of the system transitions from that of a heat engine driven by surface enthalpy fluxes to those of extratropical storms, driven by the conversion into kinetic energy of the available potential energy of the synoptic-scale baroclinic state. (Note that for a given wind shear magnitude, the temperature gradient and available potential energy increases poleward with the Coriolis parameter.) In intermediate stages of the transformation, both energy sources can be in play. The transition is accompanied by sometimes dramatic changes in the movement of the storm and in the structure, size, and magnitude of the surface wind and rain fields, and in extreme cases, the surface winds can amplify significantly, as in the case of the Great New England Hurricane of 1938.

In thinking through the dynamics and energetics of extratropical transition, quasi-balance potential vorticity-based thinking is helpful. In this regard, we make several key assumptions about the nature of the extratropical state that tropical cyclones interact with:

1. Quasi-geostrophic theory applies to the extratropical flow, at least qualitatively if not quantitatively
2. The quasi-geostrophic potential vorticity is a single constant through the whole troposphere with no lateral or vertical gradients
3. The troposphere is moist adiabatic along geostrophic vortex lines; i.e., it is neutral to slantwise moist convection (even though it is usually stable to vertical convection).
4. There is an equatorward-directed potential temperature gradient at the surface and/or the tropopause, with the latter considered to be a surface of constant potential vorticity.

Regarding Point 1, although even a modest tropical cyclone's core has a Rossby number much too high for quasi-geostrophy to hold, much of the interaction with baroclinic regions occurs at larger scales. As to Point 2, there is some evidence that potential vorticity is well-mixed in latitude, in middle latitudes, and that much of the Rossby wave activity has the character of Eady edge waves. Aircraft measurements (Emanuel 1988) and reanalysis data (Korty and Schneider 2007) suggest that away from arctic air masses, lapse rates of temperature are indeed close to moist adiabatic on geostrophic vortex lines. The last point reduces the middle-latitudes to something closely resembling the Eady model (Eady 1949), with constant (effective) potential vorticity in the interior and temperature gradients on the surface and tropopause, which act as though they were delta-function sheets of potential temperature gradient just inside the

two horizontal boundaries (Bretherton 1966): an effective negative gradient at the surface and an effective positive gradient at the tropopause.

For the sake of maximum simplicity, we will make one further idealization:

5. All the potential temperature gradient is confined to a delta function of latitude at both the surface and the tropopause. (The temperature itself is a step function at both boundaries.)

A tropical cyclone, represented conceptually as a point potential vortex in the lower troposphere surmounted by a chimney spewing negative potential vorticity anomaly into the upper troposphere, (as in Wu and Emanuel, 1993), approaches from the south (assuming a northern hemisphere configuration).

This conceptual idealization is illustrated in Figure 12.21, showing a point vortex with its anticyclonic outflow, approaching step functions of potential temperature at the surface and the tropopause. The baroclinicity is associated with a background shear flow from the west, represented in the diagram by a west wind at the tropopause.

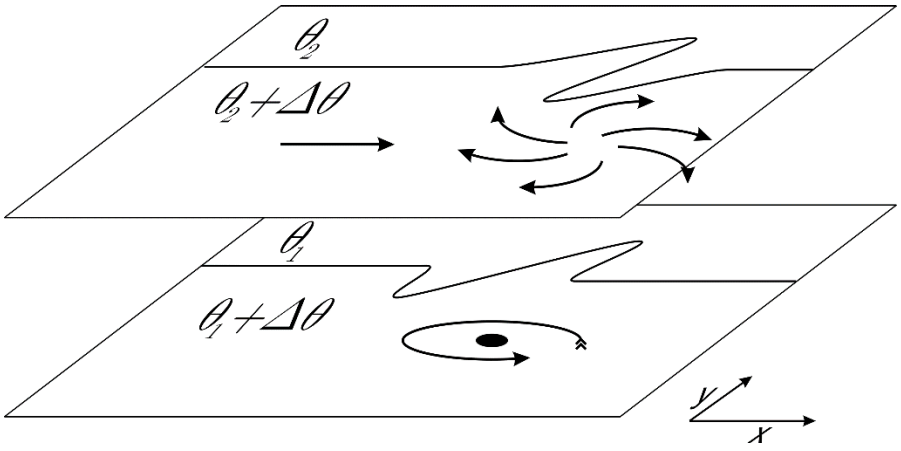


Figure 12.21: Conceptual idealization of extratropical transition, showing two planes: the surface and the tropopause in middle latitudes of the northern hemisphere. A tropical cyclone, represented here as a point potential vortex at the surface, approaches a baroclinic zone, represented by jumps of potential temperature θ at the surface and at the tropopause (wavy lines). The point potential vortex at the surface is accompanied by outflow of low potential vorticity air at the tropopause. The baroclinity is associated with shear from the west, represented here as a background flow from the west at the tropopause.

Even with this extraordinary degree of simplification and idealization, the number of dynamical interactions is large – too large to easily imagine what might happen. Among the various players are:

1. The point vortex at the surface
2. An expanding envelope of zero potential vorticity air at the tropopause, emanating from the tropical cyclone but affected by the surrounding flow
3. Eastward propagating Eady edge waves at the surface

4. Westward propagating Eady edge waves at the tropopause, Doppler shifted by the background westerly flow there

The number of interactions amongst these players is still much larger. For example, the motion of the point vortex at the surface will be influenced by

- Any large-scale background flow present at the surface
- Perturbations to the surface flow from Eady edge waves along the surface temperature jump
- Perturbations to the surface flow from Eady edge waves along the tropopause temperature jump
- Perturbations to the surface flow associated with the displaced envelope of zero potential vorticity air at the tropopause, emanating from the cyclone.

Note that even without the cyclone, the flow is unstable to baroclinic instability (the interaction of the upper and lower Eady edge waves) according to the theory of baroclinic instability (e.g. see Hoskins and James 2014).

Bell (1990) considered the interaction of a barotropic point vortex with a single barotropic vortex sheet, somewhat analogous to our own idealization as illustrated in Figure 12.21, except without the tropopause. Even with this great reduction in complexity, the solutions of this system are far from simple, with the cyclone's wind field potentially exciting Rossby waves (analogous to Eady edge waves) on the "front". Qualitatively adapting Bell's solutions to the circumstance of a cyclonic point potential vortex approaching a negative potential temperature jump from the south, two types of solution are possible. If the flow at the temperature jump is from the west, relative to the zonal movement of the vortex, then no stationary, forced Eady edge wave solution is possible since those waves travel exclusively eastward relative to the flow. Instead, the front and the flow associated with it bend northward at the zonal position of the vortex. This is illustrated in Figure 12.22a.

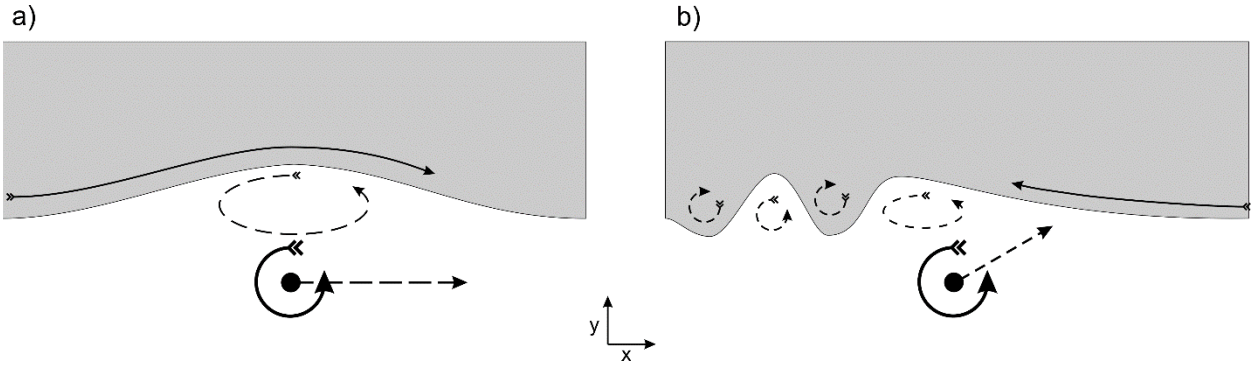


Figure 12.22: A point vortex interacting with a surface front in the northern hemisphere, with no potential vorticity anomalies above the surface. The gray shading denotes relatively cold air north of the front, with the point vortex denoted by the black dot south of the front. The dashed circular arrows denote the sense of circulation associated with the deformations of the front. In a) the mean flow is from the west relative to a coordinate system moving with the background flow at the latitude of the point vortex. In this case there can be no Eady edge waves as part of a steady solution. In b) the mean relative flow at the front is from the east and a train of stationary Eady edge waves extends west of the longitude of the point vortex. The induced motion of the point vortex is denoted by the straight dashed arrows in both cases.

The northward bend is associated with a warm temperature anomaly, which acts like a positive potential vorticity anomaly at the surface. The resulting cyclonice perturbation circulation advects the point vortex eastward with no meridional component of motion. Bell (1990) showed that one can then switch into a coordinate system moving with the point vortex (with its motion updated) and repeat the steady solution, iterating until a truly stationary solution is found.

If instead, the mean flow at the front is from the east relative to any zonal motion of the point vortex, then stationary Eady waves can exist and the solution has a train of such edge waves trailing off to the west of the longitude of the point vortex, as shown in Figure 12.22b. Now the circulation anomalies associated with the surface temperature perturbations are such as to induce a northeastward drift of the point vortex. These solutions will become invalid once the interaction becomes sufficiently nonlinear.

All the preceding assumes that the magnitude of the point potential vortex remains constant. This ignores all the complex thermodynamic and dynamic influences on the intensity of the surface vortex. Tropical cyclones undergoing extratropical transition are typically moving poleward and therefore encounter colder ocean waters if not landfall. Consequently, the potential intensity usually decreases, although as we will see later in this chapter, it can transiently increase again when large pools of cold air move over relatively warm water.

Given the complex behavior of even simple systems such as those described above, it is hardly surprising that most researchers studying extratropical transition rely on detailed analyses of observations and complex numerical simulations. (Still, it would be good if an enterprising student would code up the system illustrated in Figure 12.21, with a quasi-geostrophic point potential vortex and a closed, single expanding contour of zero potential vorticity above it, interacting with single contours of potential temperature at the lower and upper boundaries.)

In Chapter 8, section 8.3.8, we showed a particular example of an extratropical transitioning tropical cyclone: Atlantic Hurricane Floyd of 1999. Figure 8.15 shows a sequence of maps of Ertel's potential vorticity on a particular potential temperature surface as Floyd interacts with an upper tropospheric trough approaching from the northwest. In Figure 12.23 we present a sequence of cross-sections through the center of the Floyd's surface pressure center, showing potential vorticity. At 00 GMT on September 16, 1999, Hurricane Floyd is evident as a tall tower of high potential vorticity, extending upward to about 250 hPa in the southeast portion of the cross-section (Figure 12.23a). To its northwest, an upper tropospheric trough is manifest as a positive potential vorticity anomaly extending down to about 400 hPa, with a cold anomaly below it. Twelve hours later, about the time that Floyd makes landfall (Figure 12.23b), the two anomalies have drawn closer together, suggesting an increase in overall system energy according to the superposition principle. On September 17th at 00 GMT, the two anomalies are beginning to merge (Figure 12.23b) and twelve hours later there is a continuous PV tower and strong baroclinicity extending through the depth of the troposphere. Clearly, the portion of this PV tower that originated in the upper troposphere is playing an important role in the vertically integrated potential vorticity, and its merger with Floyd's PV is responsible for prolonging its intensity after landfall and contributing to the unusually heavy rains and flooding produced by the event.

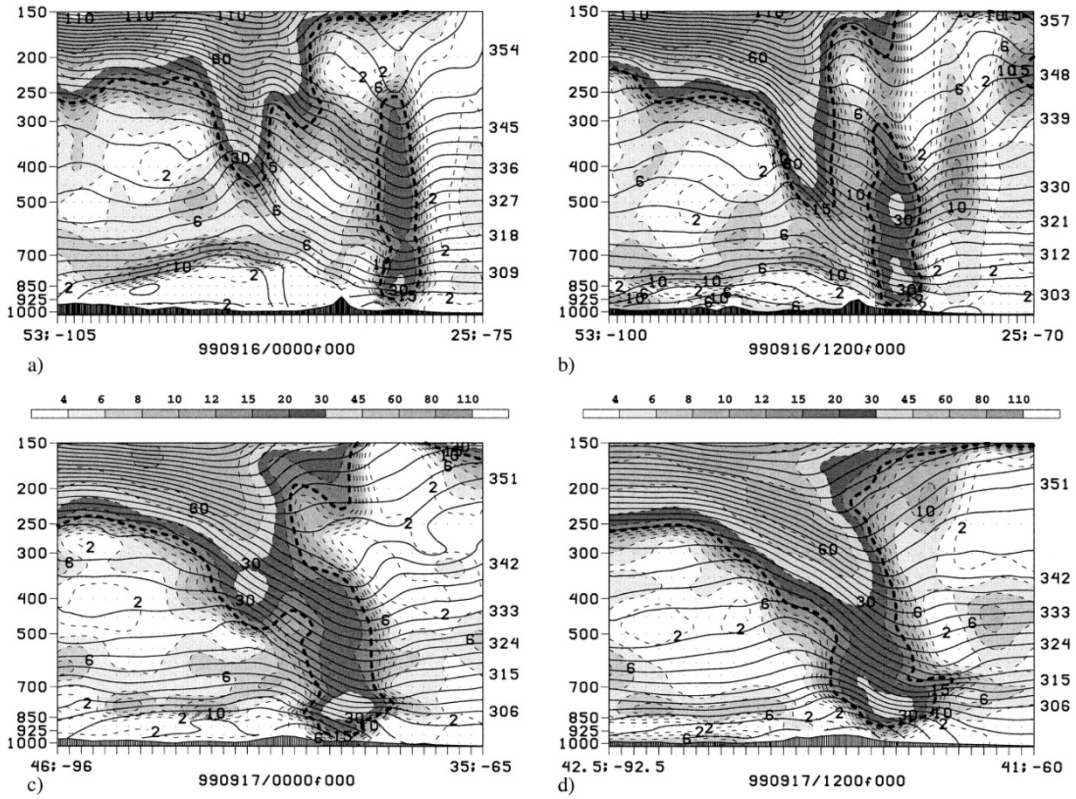


Figure 12.23: Sequence of cross-sections through Hurricane Floyd's surface center at a) 00 GMT 16 September, b) 12 GMT 16 September, c) 00 GMT 17 September, and d) 12 GMT 17 September, 1999. The first two cross-sections run from northwest (left) to southeast(right) while the last two run more nearly east-west. The cross-sections show potential vorticity (shaded) and potential temperature (solid black lines, contoured every 3 K).

There are many pathways to extratropical transition that can differ substantially in their details. Quite a few excellent overviews have been written on this topic (e.g. Hart and Evans 1991; Hanley et al. 2001; Jones et al. 2003; Evans and Hart 2003; Bieli et al. 2019; Sarro and Evans 2022).

Occasionally, disturbances that develop in middle and high latitudes, usually as extratropical cyclones but sometimes as mesoscale convective systems, move over relatively warm water and develop into tropical cyclones powered by surface enthalpy fluxes. Davis and Bosart (2004a) coined the term "tropical transition" to describe tropical cyclones that originate in extratropical baroclinic processes. Tropical transition usually involves some combination of classical baroclinic development that serves as a "trigger" for tropical cyclogenesis and a local increase in tropical cyclone potential intensity owing to deep cooling of the troposphere as part of the baroclinic process. As with extratropical transition, there are many pathways to tropical transition and several overviews have been published (Davis and Bosart 2003, 2004b; Bentley et al. 2016, 2017).

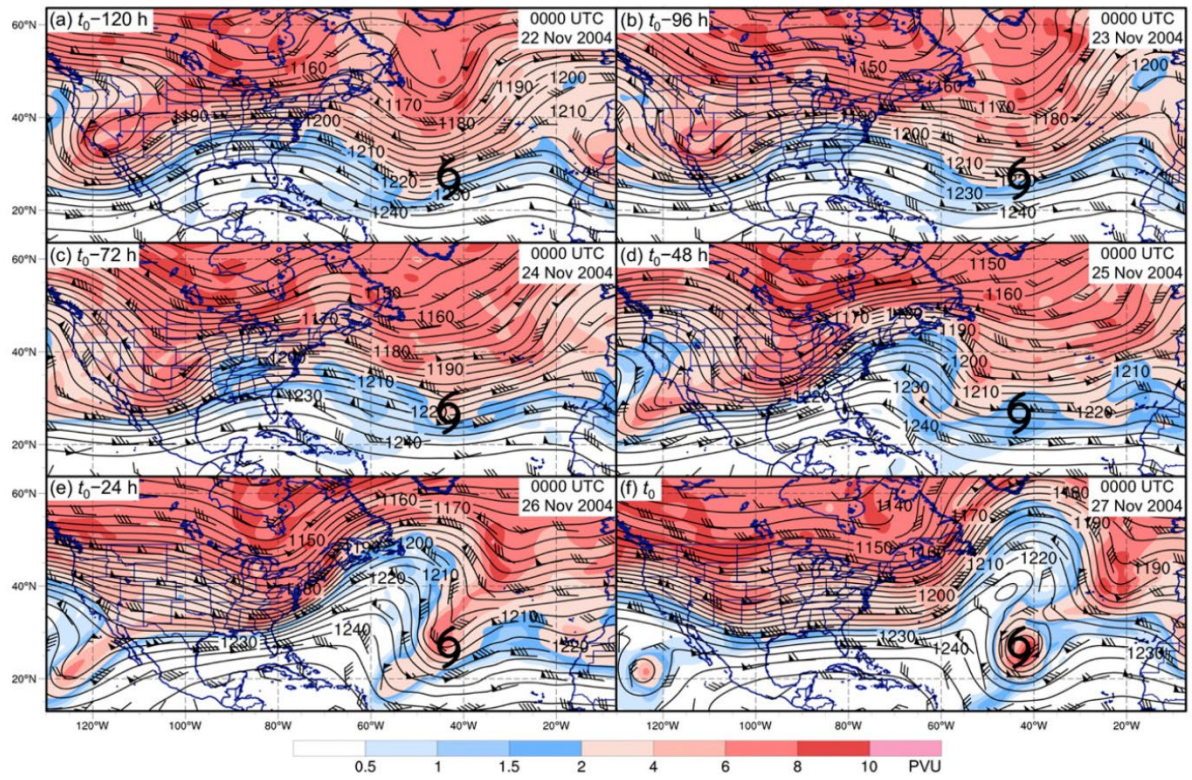


Figure 12.24: Development of subtropical cyclone Otto of 2004 over a 6-day period beginning at 00 GMT on 22 November 2004 (a-f). The shading shows the 150-250 hPa layer mean potential vorticity (PV, in PV units as shown by color bar at bottom), geopotential height at 200 hPa (black contours, every 5 decameters) and winds (flags and barbs, in knots). Subtropical cyclone Otto was considered to have formed at 00 GMT on 27 November. Its position at that time is shown by the hurricane symbol in each of the six panels.

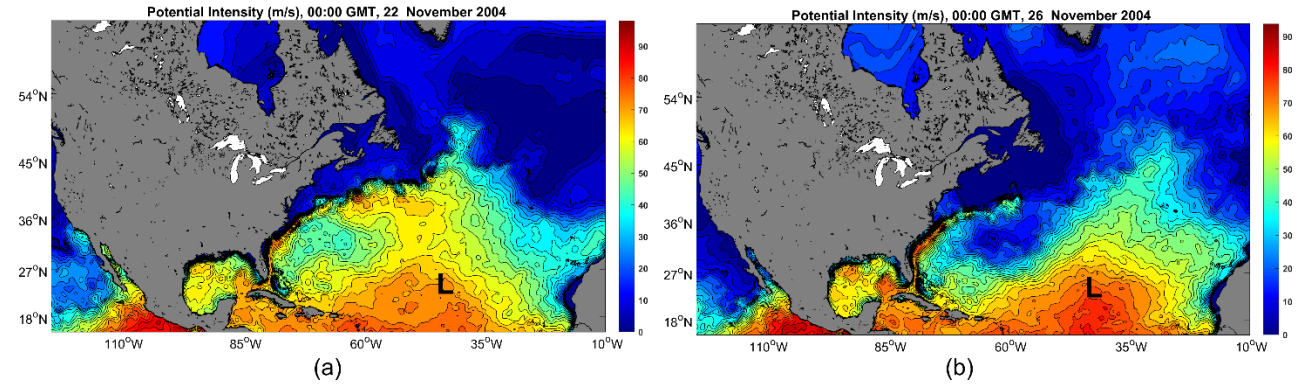


Figure 12.25: Tropical cyclone potential intensity (m/s) from ERA-5 reanalyses at 00 GMT on November 22 (a) and November 26 (b). The "L" marks the position of subtropical cyclone Otto of 2004.

A particular example of tropical transition is presented in Figures 12.24 and 12.25. Each of the six panels in Figure 12.24 shows upper tropospheric potential vorticity relative to the position of subtropical cyclone[4] Otto at the time of its formation, considered to be 00 GMT on 27 November 2004. The initial baroclinic development appears to be instigated by a shortwave appendage to

[4] "Subtropical cyclone" is a somewhat informal term that denotes a tropical cyclone that has at least some extratropical characteristics, such as a cold-core cyclone in the upper troposphere.

a longer synoptic-scale trough evident in the PV field over the central North Atlantic in panel *a* of Figure 12.24. While the shortwave trough in not evident in the PV field 24 hours later (panel *b*), it is present again after another day passes (panel *c*). Meanwhile, a pronounced ridge begins developing the far western North Atlantic, off the east coast of the U.S. As this ridge amplifies, it appears to lead to a Rossby wave breaking event downstream to the east, evident already at 00 GMT on November 25th (panel *d*) and becoming very pronounced on the 26th (panel *e*) as the ridge continues to build to the west. Finally, this wave breaks by the 27th, forming a cold-core cut-off cyclone directly above the subtropical cyclone development at the surface (panel *f*). Note that this type of interaction is much more sustained than the hit-and-run interactions described in section 12.3.2.

Figure 12.25 shows the potential intensity calculated from the algorithm described by Bister and Emanuel (2002) applied to ERA-5 reanalyses for 00 GMT on November 22nd and 26th (corresponding to panels *a* and *e* of Figure 12.24). The region where subtropical cyclone Otto forms already has high potential intensity $\left(\simeq 70\ ms^{-1}\right)$ on the 22nd, and this increases to roughly $75\ ms^{-1}$ by the 26th. It is very unlikely that this increase is caused by changes in the sea surface temperature over a mere 4 days; instead, we are seeing the effect of deep cooling of the upper troposphere associated with the development of the cold-core cut-off cyclone. (Note also the very large decline of potential intensity in the western North Atlantic associated with the strongly developing upper tropospheric ridge.) The increase of potential intensity assists the development of the subtropical cyclone but is not its primary cause in this case.

Tropical transitions of this kind are prominent at the margins of the tropical cyclone seasons in each basin, when there is normally too much shear, and perhaps not enough potential intensity, for tropical cyclogenesis to proceed without strong forcing. As shown in Figure 12.16, the stronger the ventilation parameter, the larger the initial amplitude of a disturbance must be to achieve intensification by the heat engine route. (If it is too strong, no disturbance can develop.) In the case of subtropical cyclone Otto, note the very strong shear that one can infer from the strong winds at 200 hPa during the formative stages of the cyclone. We can presume that this stage is almost purely baroclinic development. But by the time Otto is named on November 27th, there is almost no shear at the position of the center of the cut-off cyclone aloft.

Sometimes cyclones undergo multiple transitions between surface-driven and baroclinic energy sources. A good example of this is Atlantic Hurricane Sandy of 2012, which originated deep in the tropics as a classical tropical cyclone but ended its life as an extremely intense baroclinic cyclone that wreaked considerable damage in New York, New Jersey and southern New England (Sobel 2014). A sequence of maps showing Sandy's evolution is presented in Figure 12.26.

Sandy developed in the central Caribbean as a classical late-season hurricane and moved northward toward Florida. Beginning on October 26th, Sandy interacted with a shortwave trough in the subtropical jet, as apparent in the 340 K potential temperature contour plotted on the 2 PVU Ertel potential vorticity surface in Figures 12.26 *a* and *b*. (The two potential temperature contours plotted in Figure 12.26 are more or less representative of the locations of the polar and subtropical jet streams.)

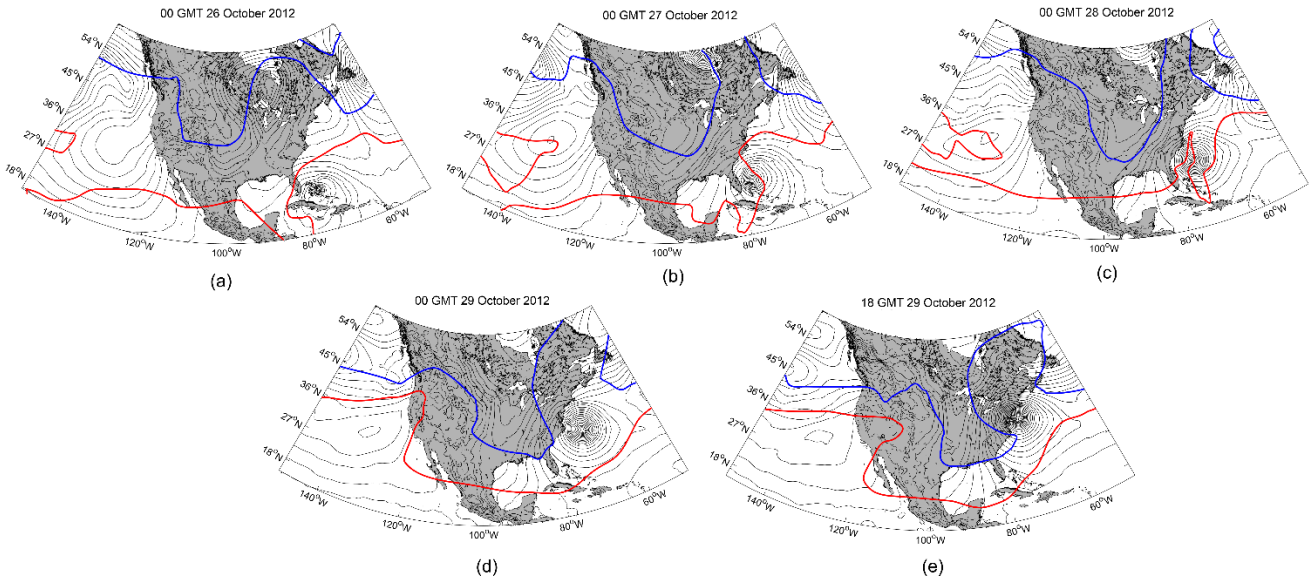


Figure 12.26: Surface pressure (black curves) and two potential temperature contours on the 2 PVU surface of constant Ertel potential vorticity: 310 K (blue) and 340 K (red). The maps are for 00 GMT on October 26th (a), October 27th (b), October 28th (c), October 29th (d) and 18 GMT on October 29th (e) in 2012. Hurricane Sandy is manifest as the tightly closed lower pressure just north of Cuba on the 26th, progressing northward to a position off the New Jersey shore at 18 GMT on the 29th.

This baroclinic interaction was strongest late on the 25th (not shown) and early on the 26th and Sandy reached its peak wind intensity near 00 GMT on the 26th. As the shortwave trough axis moved over the surface center on the October 27th, it was weakened by heating in Sandy's core and Hurricane Sandy itself weakened, but was still being sustained by strong surface enthalpy fluxes. On October 29th, Sandy interacted with a much deeper and stronger trough associated with the polar jet (blue contour in Figure 12.26) and deepened again, reaching a second peak intensity on the 29th. During this second deepening phase, the baroclinic interactions also resulted in a significant broadening of Sandy's wind and pressure field. (In terms of the radius of gale-force winds, Sandy was the largest hurricane on record in the North Atlantic.) It is important to note that once Sandy's surface center crossed over the north wall of the Gulf stream, then located at about 30° N at Sandy's longitude, it passed into a region of very small potential intensity (not shown) and its subsequent behavior was dictated by baroclinic dynamics. Hurricane Sandy was clearly affected by two distinct baroclinic interactions.

12.5 Polar lows and medicanes

We have thus far described transient baroclinic interactions that produce largely temporary effects on tropical cyclones, and extratropical and tropical transitions that are more sustained over time and result in more irreversible changes in storm structure and intensity. This kind of interaction involves the formation of deep cut-off cyclones, usually as a result of the breaking of tropopause-seated Rossby waves, which create a local thermodynamic environment that is conducive to tropical cyclones.

As a near-tropopause cut-off pool of large potential vorticity approaches a region, the air below it must cool to maintain thermal wind balance. To do so, it rises, which brings the air mass close

to saturation, with the coldest, most humid air right beneath the highest PV. The key point here is that the deep, cold column elevates tropical cyclone potential intensity, which may become large even if it was initially zero. The small wind shear near the PV core, high potential intensity, and near-saturated conditions are ideal for the formation of tropical cyclones. The development of such systems is favored where cold air masses, typically formed over continents during the cold season, flow out over relatively warm water. Examples include the Norwegian, Greenland and Barents Seas downwind from Greenland, the Sea of Japan downwind from China, the Gulf of Alaska, and the Labrador Sea downwind from Canada. A partial climatology of polar lows and an example of a polar low over the Sea of Japan was presented in Chapter 8, section 8.3.9.

A polar low over the Barents Sea, just about to make landfall at the northern tip of Norway, is shown in Figure 12.27.

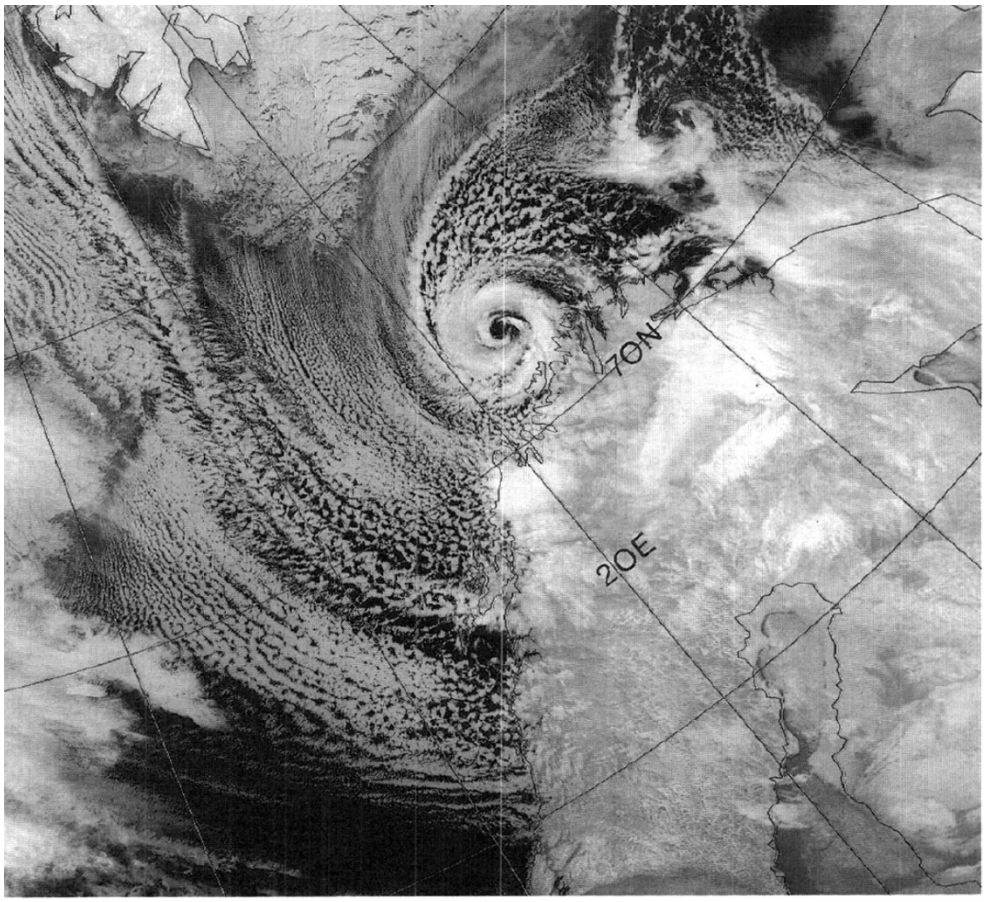


Figure 12.27: Infrared NOAA-9 satellite image of a polar low over the Barents Sea at 0224 GMT on 27 February 1987.

This small-scale cyclone brought strong winds ($\sim 20\ ms^{-1}$) to coastal points, but the storm dissipated rapidly after landfall. Like the polar low over the Sea of Japan presented in Figure 8.17, this cyclone had many of the visible characteristics of tropical cyclones, including a clear eye and spiral rainbands (or snowbands in this case). Surface air temperatures near such systems in the Barents Sea are usually below 0°C but are generally somewhat warmer close to the cyclones' core (Rasmussen 1985), reflecting the large surface fluxes of sensible as well as latent heat.

Surface pressure and 400 hPa geopotential at 02:00 GMT on 27 February, 1987, about the time of the image in Figure 12.27, are shown in Figure 12.28.

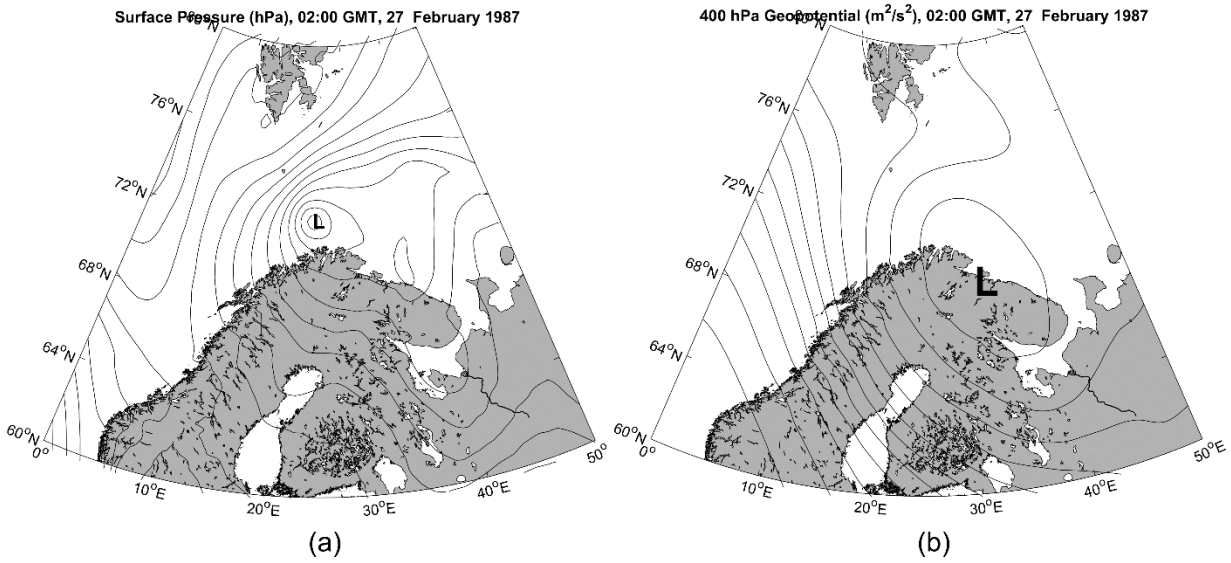


Figure 12.28: Surface pressure (a) and 400 hPa geopotential (b) at 02:00 GMT on 27 February 1987, from ERA-5 reanalyses. The surface pressure contours are in equal increments of 2.5 hPa, and the 400 geopotential contours are in equal increments of 500 m^2s^{-2}. The surface pressure and 400 hPa geopotential minima are denoted by "L".

The polar low is only a few hundred kilometers in diameter and is more or less under but displaced slightly northwest of the 400 hPa geopotential minimum.

The potential intensity at the same time is displayed n Figure 12.29. Normally, the potential intensity in this part of the world is zero at all times of the year. But when sufficiently cold air masses flow out over the relatively warm far North Atlantic, accompanied by cut-off cyclones aloft, the potential intensity can attain appreciable magnitudes, peaking at around 40 ms^{-1} in this case. The pool of high potential intensity is localized at the synoptic scale and transient in time.

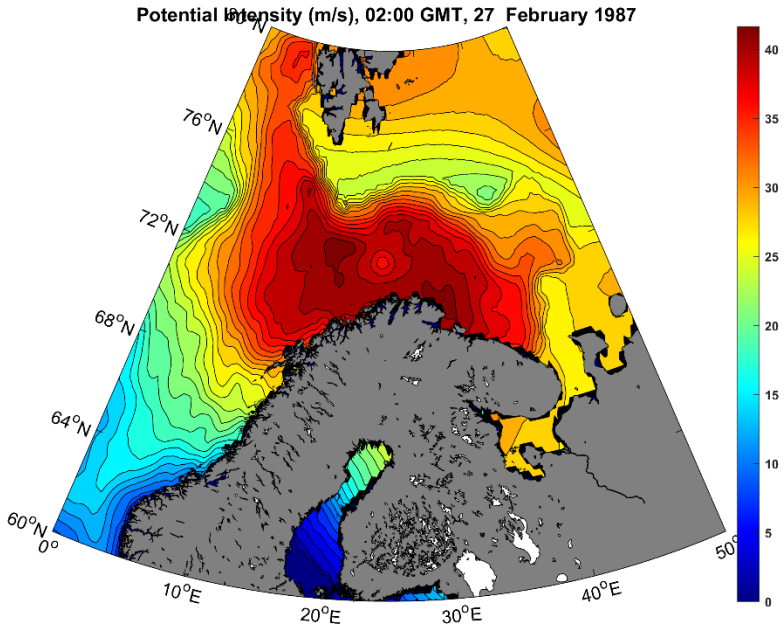


Figure 12.29: Potential intensity (m/s) at 02:00 GMT on 27 February, 1987, from ERA-5 reanalysis data.

For reviews of polar low research, see Rasmussen (2003) and Moreno-Ibáñez et al. (2021).

Polar lows also sometime occur over the Mediterranean in fall and winter, though they are hardly "polar" and the climatological potential intensity is not zero. They are usually classified separately from polar lows and are generally referred to as "Medicanes." But their synoptic-dynamic climatology is similar, forming under deep, cut-off synoptic-scale cyclones aloft, with cold air often flowing southward from Europe over the relatively warm waters of the Mediterranean[5].

A good example was the medicane of December 13-15, 2005 (Fita and Flaounas 2018). Satellite imagery at 00 GMT on December 15 (Figure 12.30) has the appearance of a classical tropical cyclone, with a clear eye, eyewall, and spiral banding. The surface cyclone developed directly underneath a cold-core cut-off geopotential minimum in the upper troposphere (Figure 12.31) and subsequently propagated eastward, dissipating a few days later in the eastern Mediterranean. The potential intensity (Figure 12.32) shows a synoptic-scale enhancement nearly collocated with the vertically stacked cyclone. This anomaly, reaching values over $50\ ms^{-1}$, is well above normal background values for the region and time of year. The resemblance to the high latitude polar low described earlier is striking.

[5] Not all tropical cyclone-like vortices over the Mediterranean develop like polar lows. For example, Medicane Ianos of September 17-18 2020 appears to have been an example of tropical transition. Although it was triggered by an upper tropospheric disturbance, potential intensity was high throughout the Mediterranean and did not appear to have been enhanced by the upper-level disturbance.

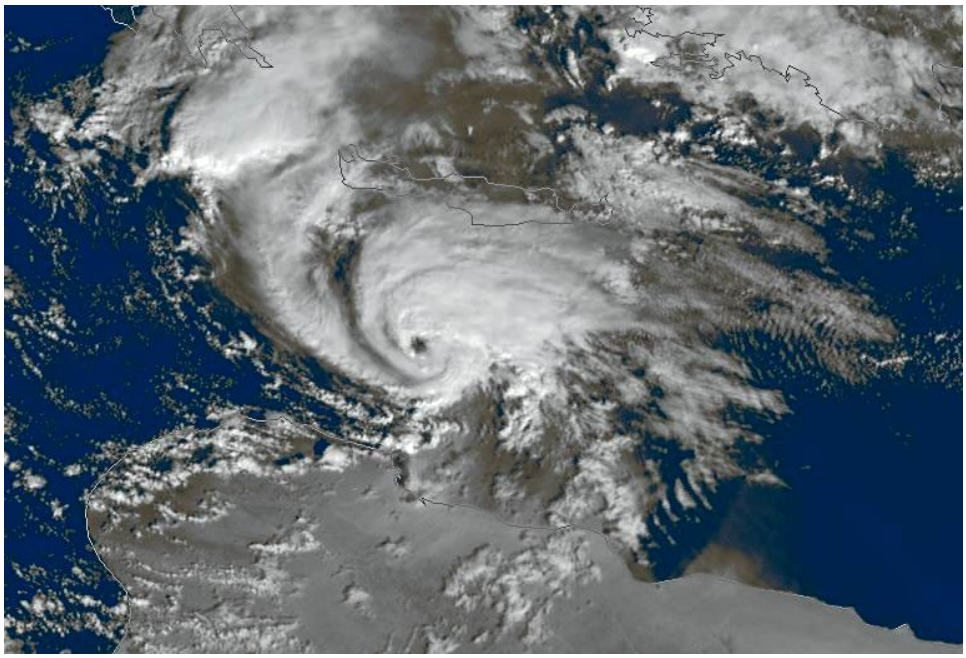


Figure 12.30: Visible satellite image of a medicane between Libya and Crete at 00 GMT on 15 December, 2005.

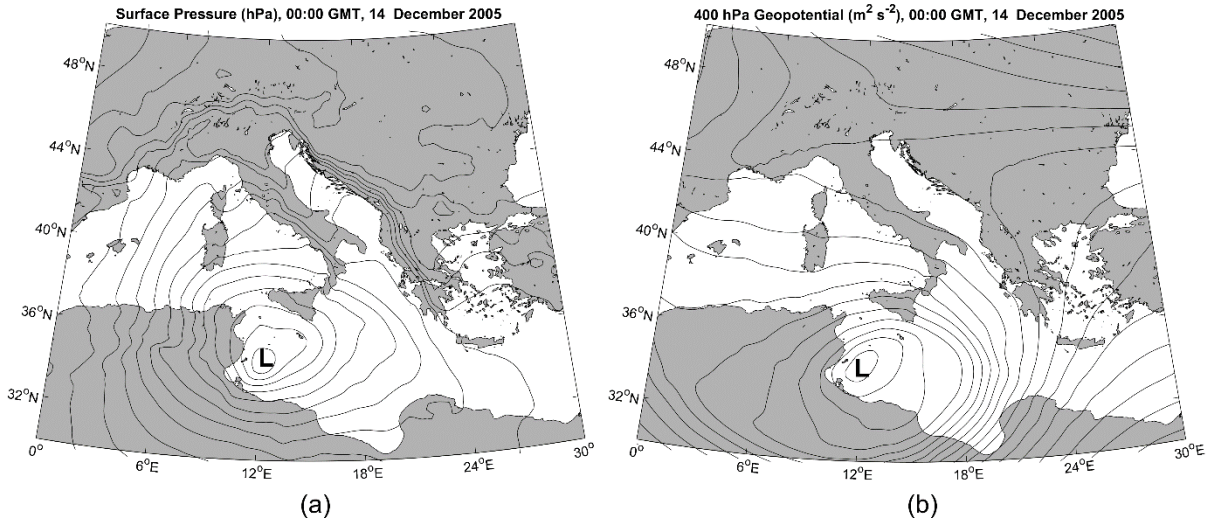


Figure 12.31: Surface pressure (a) and 400 hPa geopotential (b) at 00:00 GMT on 14 December 2005, from ERA-5 reanalyses. The surface pressure contours are in equal increments of 2.5 hPa, and the 400 geopotential contours are in equal increments of 250 m^2s^{-2}. The surface pressure and 400 hPa geopotential minima are denoted by "L".

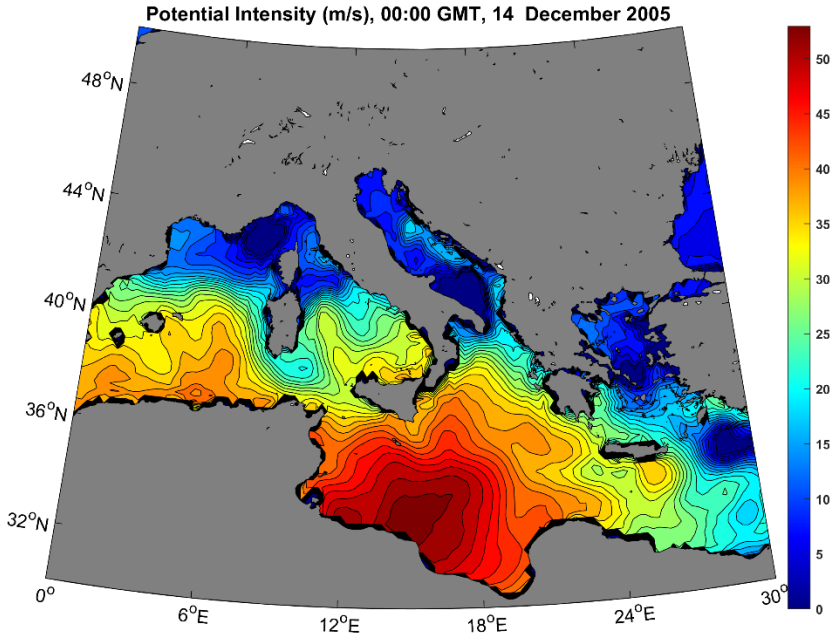


Figure 12.32: Potential intensity (m/s) at 00:00 GMT on 14 December 2005, from ERA-5 reanalysis data.

References

Adem, J., 1956: A series solution for the barotropic vorticity equation and its application to the study of atmospheric vortices. *Tellus*, **8**, 364–376.

Alland, J. J., B. H. Tang, K. L. Corbosiero, and G. H. Bryan, 2021a: Combined effects of midlevel dry air and vertical wind shear on tropical cyclone development. part I: Downdraft ventilation. *Journal of the Atmospheric Sciences*, **78**, 763–782, https://doi.org/10.1175/JAS-D-20-0054.1.

——, ——, ——, and ——, 2021b: Combined effects of midlevel dry air and vertical wind shear on tropical cyclone development. part II: Radial ventilation. *Journal of the Atmospheric Sciences*, **78**, 783–796, https://doi.org/10.1175/JAS-D-20-0055.1.

Avilés, L. B., 2018: *Taken by Storm, 1938: A Social and Meteorological History of the Great New England Hurricane*. American Meteorological Society, 292 pp.

Bell, G. I., 1990: Interaction between vortices and waves in a simple model of geophysical flow. *Physics of Fluids A: Fluid Dynamics*, **2**, 575–586, https://doi.org/10.1063/1.857757.

Bentley, A. M., D. Keyser, and L. F. Bosart, 2016: A Dynamically Based Climatology of Subtropical Cyclones that Undergo Tropical Transition in the North Atlantic Basin. *Monthly Weather Review*, **144**, 2049–2068, https://doi.org/10.1175/mwr-d-15-0251.1.

——, L. F. Bosart, and D. Keyser, 2017: Upper-tropospheric precursors to the formation of subtropical cyclones that undergo tropical transition in the North Atlantic basin. *Monthly Weather Review*, **145**, 503–520, https://doi.org/10.1175/MWR-D-16-0263.1.

Bieli, M., S. J. Camargo, A. H. Sobel, J. L. Evans, and T. Hall, 2019: A global climatology of extratropical transition. Part I: Characteristics across basins. *Journal of Climate*, **32**, 3557–3582, https://doi.org/10.1175/JCLI-D-17-0518.1.

Bister, M., and K. A. Emanuel, 2002: Low frequency variability of tropical cyclone potential intensity, 1: Interannual to interdecadel variability. *J. Geophys. Res.*, **107**, doi:10.1029/2001JD000776.

Bretherton, F. P., 1966: Critical layer instability in baroclinic flows. *Quarterly Journal of the Royal Meteorological Society*, **92**, 325–334, https://doi.org/10.1002/qj.49709239302.

Brooks, C. F., 1939: Hurricanes into New England: Meteorology of the storm of September 21, 1938. *Geographical Review*, **29**, 119–127, https://doi.org/10.2307/210069.

Burpee, R. W., 2008: The Sanders Barotropic Tropical Cyclone Track Prediction Model (SANBAR). *Meteorological Monographs*, **33**, 233–240, https://doi.org/10.1175/0065-9401-33.55.233.

Carnevale, G. F., R. C. Kloosterziel, and G. J. F. Van Heijst, 1991: Propagation of barotropic vortices over topography in a rotating tank. *Journal of Fluid Mechanics*, **233**, 119–139, https://doi.org/10.1017/S0022112091000411.

Chan, J. C. L., 2005: The physics of tropical cyclone motion. *Annual Review of Fluid Mechanics*, **37**, 99–128, https://doi.org/10.1146/annurev.fluid.37.061903.175702.

Colon, J. A., and W. R. Nightingale, 1963: Development of tropical cyclones in relation to circulation patterns at the 200-millibar level. *Monthly Weather Review*, **91**, 329–336, https://doi.org/10.1175/1520-0493(1963)091<0329:DOTCIR>2.3.CO;2.

Crisanti, A., M. Falcioni, A. Vulpiani, and G. Paladin, 1991: Lagrangian chaos: Transport, mixing and diffusion in fluids. *La Rivista del Nuovo Cimento (1978-1999)*, **14**, 1–80, https://doi.org/10.1007/BF02811193.

Davis, C. A., and L. F. Bosart, 2003: Baroclinically induced tropical cyclogenesis. *Monthly Weather Review*, **131**, 2730–2747, https://doi.org/10.1175/1520-0493(2003)131<2730:BITC>2.0.CO;2.

——, and L. Bosart, 2004a: The TT Problem. *Bull. Amer. Meteor. Soc.*, **85**, 1657–1666.

——, and L. F. Bosart, 2004b: The TT problem: Forecasting the tropical transition of cyclones. *Bulletin of the American Meteorological Society*, **85**, 1657–1662, https://doi.org/10.1175/BAMS-85-11-1657.

Dengler, K., and M. J. D.-:10. 1002/qj. 49712353909 Reeder, 1997: The effects of convection and baroclinicity on the motion of tropical-cyclone-like vortices. *Quarterly Journal of the Royal Meteorological Society*, **123**, 699–725.

Dritschel, D. G., 1989: Contour dynamics and contour surgery: Numerical algorithms for extended, high-resolution modelling of vortex dynamics in two-dimensional, inviscid, incompressible flows. *Computer Physics Reports*, **10**, 77–146, https://doi.org/10.1016/0167-7977(89)90004-X.

Dunion, J. P., C. D. Thorncroft, and C. S. Velden, 2014: The tropical cyclone diurnal cycle of mature hurricanes. *Monthly Weather Review*, **142**, 3900–3919, https://doi.org/10.1175/mwr-d-13-00191.1.

Eady, E. T., 1949: Long waves and cyclone waves. *Tellus*, **1**, 33–52.

Emanuel, K., and F. Zhang, 2016: On the predictability and error sources of tropical cyclone intensity forecasts. *J. Atmos. Sci.*, **73**, 3739–3747, https://doi.org/10.1175/JAS-D-16-0100.1.

——, and ——, 2017: The role of inner core moisture in tropical cyclone predictability and practical forecast skill. *J. Atmos. Sci.*, submitted.

Emanuel, K. A., 1988: Observational evidence of slantwise convective adjustment. *Mon. Wea. Rev.*, **116**, 1805-1816.

Emanuel, K. A., 1989: The Finite-Amplitude Nature of Tropical Cyclogenesis. *Journal of the Atmospheric Sciences*, **46**, 3431–3456, https://doi.org/10.1175/1520-0469(1989)046<3431:tfanot>2.0.co;2.

Evans, J. L., and R. E. Hart, 2003: Objective indicators of the life cycle evolution of extratropical transition for Atlantic tropical cyclones. *Monthly Weather Review*, **131**, 909–925, https://doi.org/10.1175/1520-0493(2003)131<0909:OIOTLC>2.0.CO;2.

Feng, X., and X. Chen, 2022: Rayleigh-wave dispersion curves from energetic hurricanes in the southeastern United States. *Bulletin of the Seismological Society of America*, **112**, 622–633, https://doi.org/10.1785/0120210192.

Firing, E., and R. C. Beardsley, 1976: The Behavior of a Barotropic Eddy on a β-Plane. *Journal of Physical Oceanography*, **6**, 57–65, https://doi.org/10.1175/1520-0485(1976)006<0057:tboabe>2.0.co;2.

Fita, L., and E. Flaounas, 2018: Medicanes as subtropical cyclones: the December 2005 case from the perspective of surface pressure tendency diagnostics and atmospheric water budget. *Quarterly Journal of the Royal Meteorological Society*, **144**, 1028–1044, https://doi.org/10.1002/qj.3273.

Flierl, G. R., 1977: The application of linear quasigeostrophic dynamics to Gulf Stream rings. *Journal of Physical Oceanography*, **7**, 365–379, https://doi.org/10.1175/1520-0485(1977)007<0365:TAOLQD>2.0.CO;2.

Fujiwhara, S. D.-:10. 1002/qj. 49704720010, 1921: The natural tendency towards symmetry of motion and its application as a principle in meteorology. *Quarterly Journal of the Royal Meteorological Society*, **47**, 287–292.

Gifford, J., 2004: *Hurricane Hazel: Canada's Storm of the Century*. Dundurn, 104 pp.

Hanley, D., J. Molinari, and D. Keyser, 2001: A composite study of the interactions between tropical cyclones and upper-tropospheric troughs. *Monthly Weather Review*, **129**, 2570–2584, https://doi.org/10.1175/1520-0493(2001)129<2570:ACSOTI>2.0.CO;2.

Hart, R. E., and J. L. Evans, 1991: A climatology of extratropical transition of Atlantic tropical cyclones. *J. Climate*, **14**.

Hawkins, H. F., and D. T. Rubsam, 1968: Hurricane Hilda, 1964. I: Genesis, as revealed by satellite photographs, conventional and aircraft data. *Mon. Wea. Rev.*, **96**, 428–452.

Hetzer, C. H., K. E. Gilbert, R. Waxler, and C. L. Talmadge, 2009: Generation of microbaroms by deep-ocean hurricanes. *Infrasound Monitoring for Atmospheric Studies*, A. Le Pichon, E. Blanc, and A. Hauchecorne, Eds., Springer Netherlands, 249–262.

Holland, G. J., 1983: Tropical cyclone motion: Environmental interaction plus a beta effect. *J. Atmos. Sci.*, **40**, 328–342.

Holland, G. J., 1984: Tropical cyclone motion. a comparison of theory and observation. *Journal of Atmospheric Sciences*, **41**, 68–75, https://doi.org/10.1175/1520-0469(1984)041<0068:TCMACO>2.0.CO;2.

——, and G. S. D.-:10. 1002/qj. 49711951408 Dietachmayer, 1993: On the interaction of tropical-cyclone-scale vortices. III: Continuous barotropic vortices. *Quarterly Journal of the Royal Meteorological Society*, **119**, 1381–1398.

Hoskins, B., and I. James, 2014: *Fluid dynamics of the mid-latitude atmosphere*. Wiley-Blackwell, 432 pp.

Hoskins, B. J., and F. P. Bretherton, 1972: Atmospheric frontogenesis models: Mathematical formulation and solution. *J. Atmos. Sci.*, **29**, 11–37.

Ito, K., C.-C. Wu, K. T. F. Chan, R. Toumi, and C. Davis, 2020: Recent progress in the fundamental understanding of tropical cyclone motion. *Journal of the Meteorological Society of Japan. Ser. II*, **98**, 5–17, https://doi.org/10.2151/jmsj.2020-001.

Jones, S. C., and Coauthors, 2003: The extratropical transition of tropical cyclones: forecast challenges, current understanding, and future directions. *Weather and Forecasting*, **18**, 1052–1092, https://doi.org/10.1175/1520-0434.

Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, 2010: The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone best track data. *Bull. Amer. Meteor. Soc.*, **91**, 363–376.

Korotaev, G. K., 2022: Emission of Rossby waves by synoptic vortices. *Water Resources*, **49**, 193–200, https://doi.org/10.1134/S0097807822020105.

Korty, R. L., and T. Schneider, 2007: A climatology of the tropospheric thermal stratification using saturation potential vorticity. *Journal of climate*, **20**, 5977–5991.

Krouse, K. D., and A. H. Sobel, 2010: An observational study of multiple tropical cyclone events in the western North Pacific. *Tellus A*, **62**, 256–265, https://doi.org/10.1111/j.1600-0870.2010.00435.x.

——, ——, and L. M. Polvani, 2008: On the wavelength of the Rossby waves radiated by tropical cyclones. *J Atmos Sci*, **65**, 644–654, https://doi.org/10.1175/2007jas2402.1.

Lander, M., and G. J. Holland, 1993: On the interaction of tropical-cyclone-scale vortices. I: Observations. *Quarterly Journal of the Royal Meteorological Society*, **119**, 1347–1361, https://doi.org/10.1002/qj.49711951406.

Lin, J., K. Emanuel, and J. L. Vigh, 2020: Forecasts of Hurricanes Using Large-Ensemble Outputs. *Weather and Forecasting*, **35**, 1713–1731, https://doi.org/10.1175/waf-d-19-0255.1.

——, R. Rousseau-Rizzi, C.-Y. Lee, and A. Sobel, 2023: An open-source, physics-based, tropical cyclone downscaling model with intensity-dependent steering.

Llewellyn Smith, S. G., 1996: Vortices and Rossby-wave radiation on the beta-plane. University of Cambridge, 167 pp. https://www.repository.cam.ac.uk/bitstreams/56cf5bca-dba2-4356-bf7e-c86866f5677d/download.

Llewellyn Smith, S. G., 1997: The motion of a non-isolated vortex on the beta-plane. *Journal of Fluid Mechanics*, **346**, 149–179, https://doi.org/10.1017/S0022112097006290.

Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289–307, https://doi.org/10.3402/tellusa.v21i3.10086.

Marks, D. G., 1992: *The beta and advection model for hurricane track forecasting.* Natl. Meteor. Center,.

McWilliams, J. C., and G. R. Flierl, 1979: On the evolution of isolated, nonlinear vortices. *Journal of Physical Oceanography*, **9**, 1155–1182, https://doi.org/10.1175/1520-0485(1979)009<1155:OTEOIN>2.0.CO;2.

Miller, B. I., 1958: On the maximum intensity of hurricanes. *J. Meteor.*, **15**, 184–195.

Molinari, J., and D. Vollaro, 1989: External influences on hurricane intensity, Pt. 1: Outflow layer eddy angular momentum fluxes. *J. Atmos. Sci.*, **46**, 1093–1105.

Molinari, J., P. Dodge, D. Vollaro, K. L. Corbosiero, and F. Marks Jr., 2006: Mesoscale Aspects of the Downshear Reformation of a Tropical Cyclone. *Journal of the Atmospheric Sciences*, **63**, 341–354, https://doi.org/10.1175/jas3591.1.

Moreno-Ibáñez, M., R. Laprise, and P. Gachon, 2021: Recent advances in polar low research: current knowledge, challenges and future perspectives. *Tellus A: Dynamic Meteorology and Oceanography*, **73**, 1–31, https://doi.org/10.1080/16000870.2021.1890412.

Nolan, D. S., and J. A. Zhang, 2017: Spiral gravity waves radiating from tropical cyclones. *Geophy. Res. Lett.*, **44**, 3924–3931.

O'Neill, M. E., D. Perez-Betancourt, and A. A. Wing, 2017: Accessible environments for diurnal-period waves in simulated tropical cyclones. *Journal of the Atmospheric Sciences*, **74**, 2489–2502, https://doi.org/10.1175/JAS-D-16-0294.1.

Pfeffer, R. L., 1955: A discussion of the balance of angular momentum in hurricanes. *Bull. Amer. Meteor. Soc.*, **37**, 234.

Phillips, N. A., 1954: Energy transformations and meridional circulations associated with simple baroclinic waves in a two-level, quasi-geostrophic model. *Tellus*, **6**, 274–286, https://doi.org/10.3402/tellusa.v6i3.8734.

Polvani, L. M., 1991: Two-layer geostrophic vortex dynamics. Part 2. Alignment and two-layer V-states. *Journal of Fluid Mechanics*, **225**, 241–270, https://doi.org/10.1017/S0022112091002045.

Rappin, E. D., D. S. Nolan, and K. A. D.-:10. 1002/qj. 706 Emanuel, 2010: Thermodynamic control of tropical cyclogenesis in environments of radiative-convective equilibrium with shear. *Quarterly Journal of the Royal Meteorological Society*, **136**, 1954–1971.

Rasmussen, E. A., 2003: Polar Lows. *A Half Century of Progress in Meteorology: A Tribute to Richard Reed*, R.H. Johnson and R.A. Houze, Eds., American Meteorological Society, 61–78.

Rasmussen, E. D.-:10. 1111/j. 1600-0870. 1985. tb00440. x, 1985: A case study of a polar low development over the Barents Sea. *Tellus A*, **37A**, 407–418.

Reasor, P. D., and M. T. Montgomery, 2001: Three-dimensional alignment and corotation of weak, TC-like vortices via linear vortex Rossby waves. *J. Atmos. Sci.*, **58**, 2306–2330.

Reasor, P. D., M. T. Montgomery, and L. D. Grasso, 2004: A new look at the problem of tropical cyclones in vertical shear flow: vortex resiliency. *Journal of the Atmospheric Sciences*, **61**, 3–22, https://doi.org/10.1175/1520-0469(2004)061<0003:ANLATP>2.0.CO;2.

Reilly, D. H., 1992: On the role of upper-tropospheric potential vorticity advection in tropical cyclone formation: Case studies from 1991. Massachusetts Institute of Technology, 124 pp. https://dspace.mit.edu/handle/1721.1/54988.

Riehl, H., 1948: *On the formation of west Atlantic hurricanes*.

Riehl, H., 1950: A model of hurricane formation. *J. Appl. Phys.*, **21**, 917–925.

Riemer, M., M. T. Montgomery, and M. E. Nicholls, 2010: A new paradigm for intensity modification of tropical cyclones: Thermodynamic impact of vertical wind shear on the inflow layer. *Atmos. Chem. Phys.*, **10**, 3163–3188.

——, ——, and ——, 2013: Further examination of the thermodynamic modification of the inflow layer of tropical cyclones by vertical wind shear. *Atmos. Chem. Phys.*, **13**, 327–346, https://doi.org/10.5194/acp-13-327-2013.

Rios-Berrios, R., P. M. Finocchio, J. J. Alland, X. Chen, M. S. Fischer, S. N. Stevensen, and D. Tao, 2023: A review of the interactions between tropical cyclones and environmental vertical wind shear.

Ritchie, E. A., and G. J. Holland, 1993: On the interaction of tropical-cyclone scale vortices. II: Interacting vortex patches. *Quart. J. Roy. Meteor. Soc.*, **119**, 1363–1397.

Sadler, J. C., 1976: A role of the tropical upper tropospheric trough in early season typhoon development. *Monthly Weather Review*, **104**, 1266–1278, https://doi.org/10.1175/1520-0493(1976)104<1266:AROTTU>2.0.CO;2.

Sarro, G., and C. Evans, 2022: An updated investigation of post-transformation intensity, structural, and duration extremes for extratropically transitioning North Atlantic tropical cyclones. *Monthly Weather Review*, **150**, 2911–2933, https://doi.org/10.1175/MWR-D-22-0088.1.

Schecter, D. A., and D. H. E. Dubin, 1999: Vortex motion driven by a background vorticity gradient. *Physical Review Letters*, **83**, 2191–2194.

——, M. T. Montgomery, and P. D. Reasor, 2002: A theory for the vertical alignment of a quasigeostrophic vortex. *Journal of the Atmospheric Sciences*, **59**, 150–168, https://doi.org/10.1175/1520-0469(2002)059<0150:ATFTVA>2.0.CO;2.

Simpson, R. H., and H. Riehl, 1958: Mid-tropospheric ventilation as a constraint on hurricane development and maintenance. Technical Conf. on Hurricanes, Amer. Meteor. Soc., D4-10.

Smith, R. K., and H. C. D.-:10. 1002/qj. 49711951314 Weber, 1993: An extended analytic theory of tropical-cyclone motion in a barotropic shear flow. *Quarterly Journal of the Royal Meteorological Society*, **119**, 1149–1166.

——, W. Ulrich, and G. Sneddon, 2000: On the dynamics of hurricane-like vortices in vertical-shear flows. *Quarterly Journal of the Royal Meteorological Society*, **126**, 2653–2670, https://doi.org/10.1002/qj.49712656903.

Sobel, A., 2014: *Storm Surge: Hurricane Sandy, Our Changing Climate, and Extreme Weather of the Past and Future*. First. Harper Wave, 336 pp.

Sutyrin, G., and Y. Morel, 1997: Intense vortex motion in a stratified fluid on the beta-plane: an analytical theory and its validation. *J. Fluid Mech.*, **336**, 203–220, https://doi.org/10.1017/S0022112096004685.

Sutyrin, G. G., 1988: Motion of an intense vortex on a rotating globe. *Fluid Dyn.*, **23**, 215.

Sutyrin, G. G., and G. R. Flierl, 1994: Intense Vortex Motion on the Beta Plane: Development of the Beta Gyres. *Journal of Atmospheric Sciences*, **51**, 773–790, https://doi.org/10.1175/1520-0469(1994)051<0773:IVMOTB>2.0.CO;2.

Tang, B., and K. Emanuel, 2010: Midlevel ventilation's constraint on tropical cyclone intensity. *Journal of the Atmospheric Sciences*, **67**, 1817–1830, https://doi.org/10.1175/2010jas3318.1.

——, and ——, 2012: A ventilation index for tropical cyclones. *Bulletin of the American Meteorological Society*, **93**, 1901–1912, https://doi.org/10.1175/BAMS-D-11-00165.1.

Tang, B. H.-An., 2010: Midlevel ventilation's constraint on tropical cyclone intensity. Massachusetts Institute of Technology, Department of Earth, Atmospheric, and Planetary Sciences, 195 pp. https://dspace.mit.edu/handle/1721.1/62321.

Tao, D., and F. Zhang, 2015: Effects of vertical wind shear on the predictability of tropical cyclones: Practical versus intrinsic limit. *Journal of Advances in Modeling Earth Systems*, **7**, 1534–1553, https://doi.org/10.1002/2015MS000474.

Walsh, D., and L. J. Pratt, 1995: The interaction of a pair of point potential vortices in uniform shear. *Dynamics of Atmospheres and Oceans*, **22**, 135–160, https://doi.org/10.1016/0377-0265(95)00402-V.

Wang, B., R. L. Elsberry, Y. Wang, and W. Luguang, 1998: Dynamics in tropical cyclone motion: A review. *Chin. J. of Atmos. Sci.*, **22**, 416–434.

Wilson, J. D., and N. C. Makris, 2008: Quantifying hurricane destructive power, wind speed, and air-sea material exchange with natural undersea sound. *Geophysical Research Letters*, **35**, https://doi.org/10.1029/2008GL033200.

Wu, C.-C., and K. A. Emanuel, 1993: Interaction of a baroclinic vortex with background shear: Application to hurricane movement. *J. Atmos. Sci.*, **50**, 62–76.

Yanai, M., 1968: Evolution of a tropical disturbance in the Caribbean Sea region. *Journal of the Meteorological Society of Japan. Ser. II*, **46**, 86–109, https://doi.org/10.2151/jmsj1965.46.2_86.

Zakharov, V. I., and P. K. Sigachev, 2022: Ionospheric disturbances from tropical cyclones. *Advances in Space Research*, **69**, 132–141, https://doi.org/10.1016/j.asr.2021.09.025.

Zhang, F., and D. Tao, 2013: Effects of vertical wind shear on the predictability of tropical cyclones. *J. Atmos. Sci.*, **70**, 975–983, https://doi.org/10.1175/jas-d-12-0133.1.

13. Interaction of tropical cyclones with the ocean and land surface

Tropical cyclones are driven by turbulent enthalpy fluxes from the surface, and retarded by turbulent momentum fluxes into the surface. Both these fluxes are strongly affected by the character of the underlying surface. In the widely used Monin-Obukhov theory reviewed in Chapter 4, section 4.2, the important parameters are the surface roughness length and the Monin-Obukhov length scale, which depends on the surface momentum and buoyancy fluxes. The resulting relationships between ensemble mean properties of the boundary layer and surface fluxes have been extensively tested through field experiments, most of which have been land-based. There have been fewer tests of these relations over the ocean, whose surface is strongly perturbed by the winds themselves. Very few observations have been made in the extreme conditions of intense tropical cyclones, whose physics are sensitive to surface fluxes in just such conditions.

Surface waves are one obvious way that winds perturb the ocean surface. In section 13.1, we explore how waves affect turbulent surface fluxes. Although storm surges have little effect on surface fluxes, they are enormously destructive and are therefore explored in section 13.2.

In section 13.3 we delve into the largely unsolved problem of how the generation of sea spray at hurricane-force winds can strongly alter those fluxes over water.

In our theoretical development, we have assumed that the sea surface temperature remains constant during passage of the tropical cyclone inner core. Observations make it clear that this is usually a poor approximation, and in section 13.4 we show how the generation of near-inertial oscillations by tropical cyclones leads to mixing to the surface of colder water from below the seasonal thermocline, reducing enthalpy fluxes to the atmosphere and thereby weakening TCs.

Unless they are strongly boosted by baroclinic processes, tropical cyclones nearly always spin down over land. But on rare occasions they are observed to re-intensify in the absence of significant baroclinicity. In at least some of these cases observations and theory suggest that fluxes of heat through initially hot, wet soils can sustain weak tropical cyclones. We review this evidence in section 13.5.

13.1: Ocean surface waves

The strong winds of tropical cyclones drive surface waves that are hazardous to ships and coastal residents and which affect the transfer of heat and momentum between the sea surface and the atmosphere. Understanding and modeling the generation and propagation of surface waves may therefore be necessary not only to quantify the wave hazard but also to predict the evolution of TCs.

The linear theory of surface waves on an incompressible fluid surmounted by an atmosphere of much smaller density yields a dispersion relation (Phillips 1966) relating the wave frequency, ω, to its horizonal wavelength, L:

$$\omega^2 = gk \tanh\left(kD\right), \tag{13.1}$$

where g is the acceleration of gravity, $k = 2\pi / L$ is the horizontal wavenumber, and D is the ocean depth. The phase speed of the waves is given by

$$c_p = \omega / k = \sqrt{\frac{g \tanh(kD)}{k}}, \tag{13.2}$$

and their group velocity is

$$c_g = \frac{\partial \omega}{\partial k} = \frac{1}{2}\sqrt{\frac{g \tanh(kD)}{k}}\left(1 + kD\frac{1 - \tanh^2(kD)}{\tanh(kD)}\right). \tag{13.3}$$

Note that for wavelengths that are very short compared to the ocean depth ($kD >> 1$), the $\tanh$ terms asymptotically approaches unity, while in the limit of small kD, neither the phase speed nor the group velocity depend on k and these "shallow water" waves are nondispersive. Hereafter we will confine ourselves to the deep-water limit:

$$\omega = \sqrt{gk}, \quad c_p = \sqrt{\frac{g}{k}}, \qquad c_g = \frac{1}{2}\sqrt{\frac{g}{k}}\ . \tag{13.4}$$

Note that in this deep-water limit, the group velocity is half the phase speed.

As wind blows across the water surface, a spectrum of surface waves develops. The spectral wave height variance density obeys (Hasselmann 1962)

$$\frac{\partial E}{\partial t} + c_{gx}\frac{\partial E}{\partial x} + c_{gy}\frac{\partial E}{\partial y} = S_{in} + S_{diss} + S_{nl}, \tag{13.5}$$

where the spectral wave height variance density $E(\omega, \theta, x, y, t)$ is a function of wave frequency ω, wave angle θ, the two horizontal coordinates x and y, and time. The spectral wave height variance propagates at the group velocity associated with the wave frequency. The source terms, in order, are wind input, S_{in}, dissipation, S_{diss}, and nonlinear wave-wave interactions, S_{nl}. Contemporary wave forecast models are based on the Hasselmann equation (13.5). Wind is the main net energy source for waves, and much of the energy input by the local wind is locally dissipated by wave breaking. Hasselmann (1962) proposed that nonlinear wave-wave interactions lead to an upscale energy cascade to long waves, and that prediction has been borne out by field experiments (Hasselmann et al. 1973).

Given unlimited space and time for a wave field to grow by wind forcing, the wind wave height variance spectrum would eventually reach and equilibrium state in which the local input of wind energy is balanced by local dissipation, mostly due to turbulence generated by wave breaking and spectral export to larger wavelengths. The time scale to achieve equilibrium is on the order of hours to days…as long as or longer than the time scales over which tropical cyclone cores affect the ocean locally. Consequently, sea states associated with tropical cyclone are usually not close to this equilibrium and must be calculated using (13.5) with appropriate specifications of the source terms.

A widely used metric of ocean waves is the significant wave height H_s, usually defined as the average over the highest 1/3 of waves in a sea state, which, at a given location and time, works out to

$$H_s = 4\sqrt{\iint E(\omega,\theta,x,y,t)d\theta d\omega}. \quad (13.6)$$

Figure 13.1 shows significant wave height as a function of both fetch (the distance over which the wind blows in a straight line in this case), and time from the onset of $20\ ms^{-1}$ winds, calculated from a widely used, public domain wave model (WAVEWATCH III, Abdolali et al. 2020).

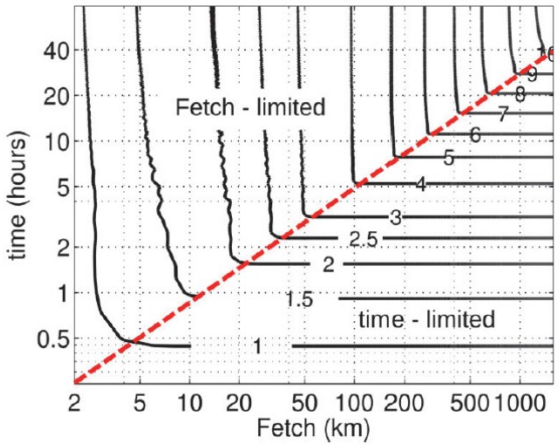


Figure 13.1: Significant wave height (m) as a function of fetch (km) and time (hrs) from the onset of $20\ ms^{-1}$ winds, based on a widely-used numerical model.

In this case, the saturation significant wave height is 10.8 m, but it would take several days and a thousand kilometers of fetch at this gale-force wind to get close to this value.

Using the same wave model, Moon et al. (2003) simulated waves associated with Atlantic Hurricane Bonnie of 1998. A snapshot of the simulated wave field is presented in Figure 13.2

The highest waves are well to the right of the storm center, with respect to Bonnie's translation direction. It may be tempting to attribute this rightward bias to the stronger winds in the right-hand sector of the circulation, and while that may be part of the explanation, the main source of this bias can be understood in reference to the deep-water wave dispersion relation and Figure 13.1. As evident in Figure 13.2, waves to the right of the storm track are moving in the same general direction as the cyclone. If we suppose that the newest, steepest waves are moving at roughly half the wind speed and that the energy propagation, which follows the group velocity, is half the phase speed, it follows that the energy propagation of these relatively young waves will be roughly ¼ of the local wind speed. Near the radius at which the wind speed is roughly 4 times the storm translation speed, the waves will be moving with the circulation and have time to grow to appreciable heights. The opposite is true in the left sector. (This is another reason for ships to avoid the dangerous right-hand semicircle of tropical cyclones.)

The longest members of the wave spectrum will have the largest group velocity and outrun the storm. In Figure 13.2 the wavelength of the dominant wave is quite large ahead of and to the

right of the storm's recent translation direction. (Before the days of aircraft, radar, and satellite data, observers on ships and near coastlines would time the period of swells…a decreasing period signaled that the storm was approaching.)

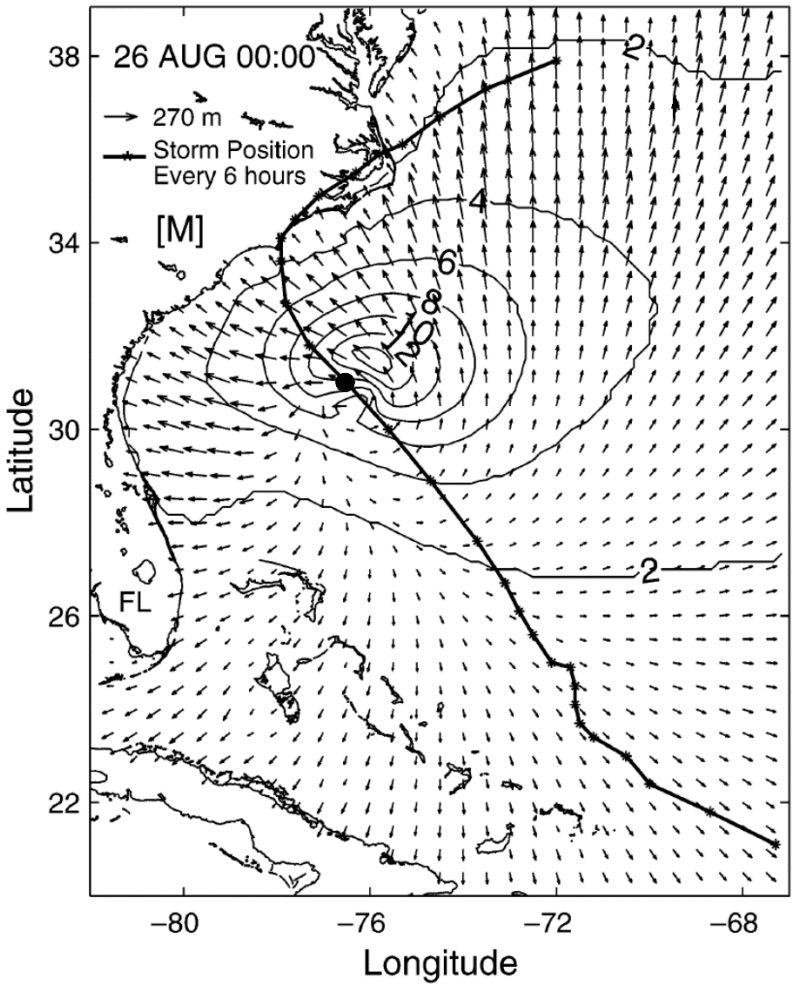


Figure 13.2: Simulated wave field associated with Hurricane Bonnie at 00 GMT 26 August 1998. The significant wave height (m) is contoured, and the arrows show the direction and length of the dominant wave at each location. (The arrow scale is shown in upper left.) Solid black curve shows the track of Hurricane Bonnie's center, with asterisks every 6 hours. The black dot shows Bonnie's current position.

While tropical cyclones can and do generate high waves, the highest recorded waves have been associated with strong extratropical cyclones. The significantly greater breadth of these storms allows for longer fetches over which surface wave packets can travel with the cyclone and extract energy from the wind.

We next consider the feedback of waves on tropical cyclones, through their effect on surface drag, which under neutral conditions is a function of wind speed and surface roughness. In the absence of sea spray, the variability of surface roughness length is mostly driven by relatively short (< 10 m wavelength) wind waves (Makin and Kudryavtsev 2002). These shorter waves may reach equilibrium fairly quickly. The earliest attempts to account for the wave contribution to surface roughness assumed local equilibrium, in which case dimensional considerations and

physical reasoning suggest that the surface roughness scales with the Charnock length (Charnock 1955):

$$z_0 = \alpha \frac{u_*^2}{g}, \tag{13.7}$$

where u_* is the friction velocity, g is the acceleration of gravity, and α is a nondimensional coefficient generally taken to be between 0.010 and 0.015. Calculation of such an equilibrium roughness length does not require a wave model. More recently, attempts have been made to refine this estimate, accounting for specific characteristics of the time-evolving sea state. Drennan et al. (2003) proposed a formulation that accounts for a measure of wave age:

$$z_0 = 3.35 H_s \left(\frac{u_*}{c_p} \right)^{3.4}, \tag{13.8}$$

where H_s is the significant wave height and c_p is the phase speed of the dominant wave (the wave corresponding to the spectral peak of the spectral wave height variance density). Calculation of the roughness length from (13.8) requires coupling to a wave model. Porchetta et al. (2019) refined this approach by also accounting for the angle θ between wave propagation and wind direction:

$$z_0 = 20 \cos\left(0.45\theta\right) H_s \left(\frac{u_*}{c_p} \right)^{3.8\cos\left(0.32\theta\right)}. \tag{13.9}$$

These formulations weight the actual drag to the right of the translation vector (in the northern hemisphere), where the significant wave heights and u_* are larger. This in turn suggests that storm that turn right, passing over the larger wave heights, may experience more dissipation than left turners.

While wave drag may indeed be important, it is possible that it may be overshadowed by spray drag, which we explore in section 13.3. We next turn our attention to a phenomenon which has no substantial feedback on tropical cyclones but is one of their deadliest side-kicks.

13.2 Storm surges

Historically, storm surges have taken more lives than wind. A single storm in East Pakistan (now Bangladesh) in 1970 is estimated to have drowned as many as half a million people. Hydrodynamically, storm surges are like tsunamis in that they have horizontal scales much larger than their depth and therefore obey the shallow water equations. But, of course, they are excited by wind rather than shaking seafloor or undersea landslides.

To a good approximation, storm surges are governed by the shallow water equations like those discussed in Chapter 6, section 6.2. But in this case, we must account for changes in the total fluid depth, variations in surface atmospheric pressure, and bottom friction. (The most sophisticated models account for baroclinic effects as well, but we ignore those here.) A commonly used equation set (e.g. Luettich et al. 1992) is

$$\frac{\partial \zeta}{\partial t} + \nabla \bullet \mathbf{V} h\zeta = 0, \tag{13.10}$$

$$\frac{\partial \mathbf{V}}{\partial t} + \left(\mathbf{V} \bullet \nabla\right)\mathbf{V} + f\mathbf{k} \times \mathbf{V} = -\alpha_0 \nabla p_s - g\nabla\zeta + \frac{\alpha_0}{h}\left(\boldsymbol{\tau}_s - \boldsymbol{\tau}_b\right), \tag{13.11}$$

where ζ is the departure of the surface height from the geoid, h is the total water column depth, $\mathbf{V}$ is the horizontal velocity vector, f is the Coriolis parameter, α_0 is the specific volume of seawater, p_s is the atmospheric surface pressure, $\boldsymbol{\tau}_s$ is the wind stress vector, and $\boldsymbol{\tau}_b$ is the bottom stress vector. (The momentum equation (13.11) usually contains internal diffusion terms as well.)

A simple understanding of why surges exist can be attained by considering the idealized case of a steady wind blowing onshore perpendicular to a north-south oriented coastline of infinite extent, with land to the west. We begin by first considering a "bathtub" ocean with constant undisturbed depth H rising abruptly to the shore. We neglect bottom stress and any variations in atmospheric surface pressure. The steady solution of (13.11) then just balances ocean onshore pressure gradients, represented by increasing sea elevation, against the applied surface wind stress:

$$g\frac{d\zeta}{dx} = \frac{\tau_x}{\rho\left(H + \zeta\right)}. \tag{13.12}$$

If we assume that $\zeta << H$, then the solution of (13.12), assuming that the wind stress vanishes a distance L of shore, is

$$\zeta \simeq \frac{1}{g\rho H}\int_L^x \tau_x dx, \tag{13.13}$$

where we have defined $x = 0$ to be the longitude of the coast. In this set-up, an onshore wind of strength U will be associated with a surface stress of $-C_D \rho_a U^2$, where C_D is the drag coefficient and ρ_a is the air density. For a hurricane-force wind of $50\, ms^{-1}$ extending *100 km* offshore, and taking $C_D = 2 \times 10^{-3}$, and a water depth of *10 m*, (13.13) would predict a water elevation of about *2.5 m* at the coastline.

A slightly more realistic case would be a uniformly sloping bottom, so that the undisturbed water depth approaches zero at the coast. Therefore, we take $H = \Gamma x$, where Γ is the bottom slope. Substituting into (13.12) gives

$$g\frac{d\zeta}{dx} = \frac{\tau_x}{\rho\left(\Gamma x + \zeta\right)}. \tag{13.14}$$

This time we cannot neglect ζ in comparison to Γx. We can reduce (13.14) to a one-parameter system by normalizing the dependent and independent variables in (13.14): $x \to Lx$ and $\zeta \to L\Gamma\zeta$. With the substitutions, (13.14) becomes

$$(x+\zeta)\frac{d\zeta}{dx} = -A, \tag{13.15}$$

where

$$A \equiv \frac{-\tau_x}{\rho g L \Gamma^2}. \tag{13.16}$$

For an onshore wind, A will be positive. As a boundary condition, we apply $\zeta = 0$ at $x = 1$ and assume as before that there is no wind stress beyond $x = 1$.

Equation (13.15) has closed-form solutions[1] in terms of Lambert functions, but these are fairly opaque, not lending much insight into the parameter dependence. It is easy to solve (13.15) numerically. Figure (13.3) compares the numerical solution of (13.15) for the nondimensional water depth at the coastline to an analytic approximation:

$$\zeta|_{x=0} \simeq A^{0.6}. \tag{13.17}$$

The comparison is quite good up to about $A = 10$. If we use the definition (13.16) and re-dimensionalize the approximate solution, then

$$\zeta|_{x=0} \simeq \left(\frac{-\tau_x}{\rho g}\right)^{0.6} L^{0.4} \Gamma^{-0.2}. \tag{13.18}$$

The surge magnitude increases a little more rapidly than the square root of the wind stress (or the first power of the wind speed) and a little more slowly than the square root of the fetch. Shallower bottom slopes produce larger surges, though the dependence is weak.

These solutions just show that onshore winds pile up water against coastlines; conversely, offshore winds lower the water level at and near the coast.

[1] Exact solutions of (13.15) satisfy the transcendental equation $A \log\left(\frac{x+\zeta-A}{1-A}\right) + \zeta = 0.$

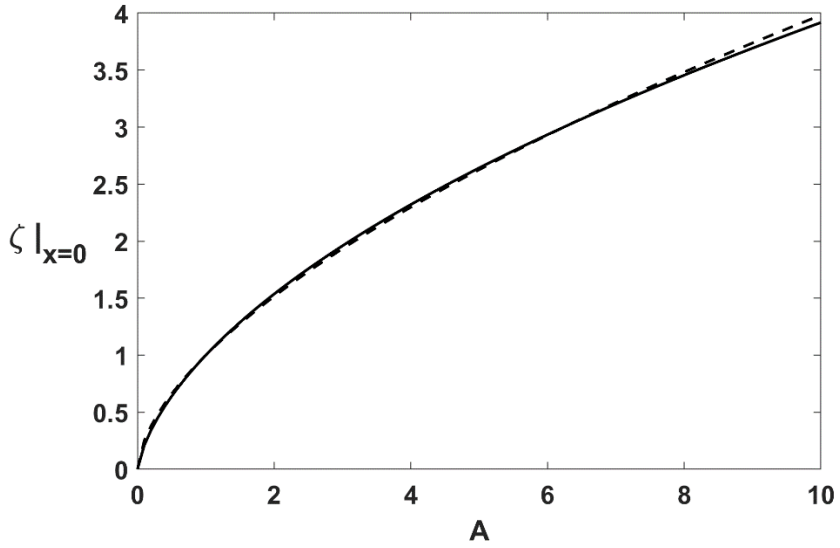


Figure 13.3: Numerical solution of (13.15) for the surge level at the coast (solid curve), and the analytic approximation (13.17) to it, (dashed).

The real world is considerably more complex. The actual wind speed strongly evolves in time, and is of finite lateral extent, setting up complex currents near the shore. The bottom is irregular and bottom friction can be important. The coastline itself is often highly complex, with bays, inlets, islands, and rivers. The surge problem is an example of a simple, though nonlinear, equation set but with boundary conditions that can be hugely complex. Moreover, surges interact with astronomical tides and other current systems, including river deltas. The atmospheric pressure term in (13.11) cannot be neglected in strong cyclonic storms; acting by itself, a *100 hPa* pressure fall would cause about a *1 m* elevation in sea level.

Coastal populations are often concentrated around bays, inlets, and river deltas. From a practical viewpoint, forecasts of surge levels in these regions are of critical importance, yet can require extraordinary horizontal resolution and thereby make unwieldy computational demands. One solution to this challenge is to solve the equations on unstructured grids that are conformally mapped to coastlines. This was the approach taken in developing the Advanced CIRCulation (ADCIRC) ocean model (Luettich et al. 1992). Figure 13.4 shows an example of an ADCIRC grid for the Gulf of Mexico. The grid areas are large in the central Gulf and become progressively finer near the shorelines. The land boundaries are reflective while the two open boundaries at the Yucatan Channel and at the entrance to the Straits of Florida are forced by tides but allow wave energy from the Gulf to pass through.

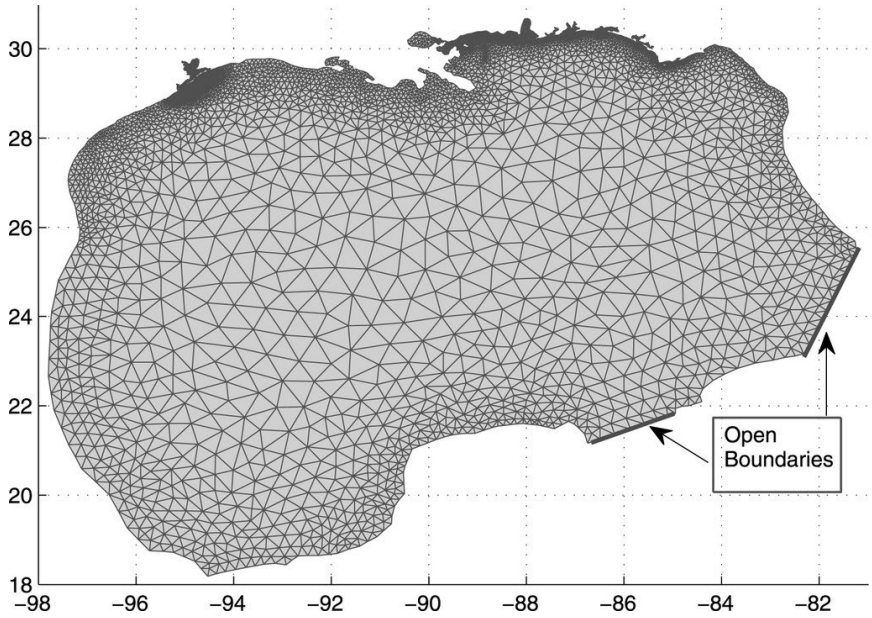


Figure 13.4: ADCIRC Gulf of Mexico grid containing 8,006 nodes and 14,269 elements. Open boundaries are denoted by bold boundary lines. The open boundaries are forced by the three most dominant diurnal and two most dominant semidiurnal tidal constituents in terms of amplitude. The land boundaries are reflective.

The peak surge levels driven by analyzed winds from Atlantic Hurricane Katrina of 2005, using the ADCIRC model, are shown in Figure 13.5 (Dietrich et al. 2012). The peak surge, which occurred in the vicinity of Bay St. Louis and Pass Christian, Mississippi, was about *8.5 m*. In Figure 13.5, note the discontinuity in water levels around the jetty extending southeast from near New Orleans. Though narrow, such jetties can have profound effects in the distribution and magnitude of storm surges.

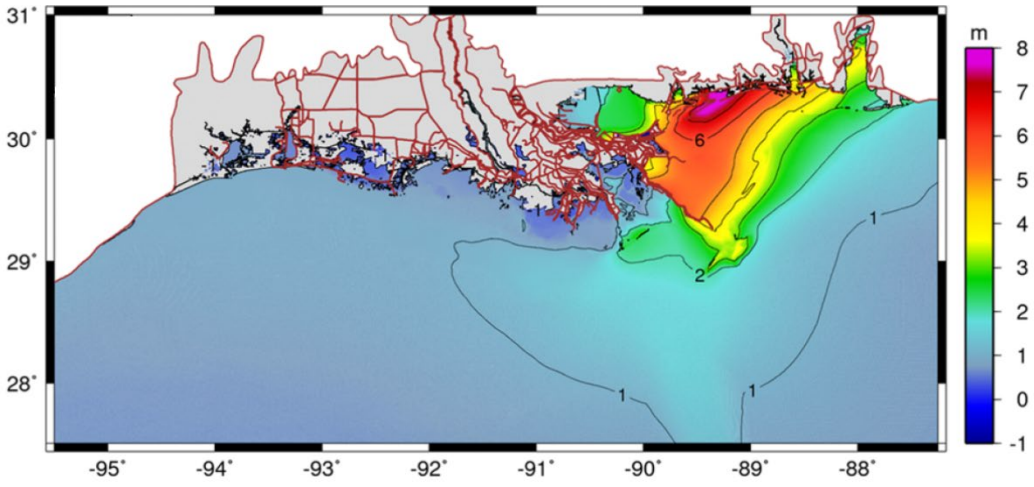


Figure 13.5: Peak surge level (m) in the northern Gulf of Mexico from the ADCIRC model driven by the observed, time-evolving winds of Hurricane Katrina of 2005.

Among the many computational challenges is modelling the over-land flow of water, including through and around urban infrastructure such as streets, buildings, and storm drains. The interaction of such flows with freshwater from torrential rains also remains a daunting problem.

Modeling storm surges is desirable not only for real-time forecasting but for long-term risk assessments, especially given that losses from storm surge flooding are among the costliest natural hazards in many nations.

13.4 Sea spray

Strong winds blowing over a water surface can generate and loft sea spray to altitudes commensurate with the Charnock length u_*^2 / g . The physics of how sea spray is generated and how it affects fluxes of momentum and enthalpy have become better understood in the last few decades, but there remains much uncertainty.

Photographs of the sea surface in tropical cyclones from reconnaissance aircraft (Black and Adams 1983) show that streamers of spray begin to develop when surface winds exceed about 20 ms^{-1}. When surface winds reach about 50 ms^{-1}, the boundary layer is filled with spray and the sea surface itself is hardly visible from the aircraft. Recently, saildrones (robotic sailboats) have been successfully deployed in fully developed hurricanes, and videos[2] taken from the saildrones confirm the presence of very large concentrations of spray at winds of about 40 ms^{-1}. Figure 13.6 shows a frame from one such video.

Figure 13.6: Frame from a video taken from a saildrone in the inner core of Atlantic Hurricane Fiona on 22 September, 2022. Surface winds were about 40 ms^{-1} at the time.

Still photographs and video footage show that spray mostly develops at the crest of waves that are breaking or on the verge of breaking. As wind speeds increase above 30-40 ms^{-1}, the shortest wind waves seem to be suppressed. These images also show that the seawater is filled with bubbles of air, and the eruption of these bubbles at the surface may also contribute to spray formation.

How does the presence of spray affect the all-important fluxes of momentum and enthalpy? To help understand the basic physics, we look at the limit of extremely strong wind speeds and

[2] See https://www.pmel.noaa.gov/saildrone-hurricane/

imagine that in such a limit, the air-ocean interface is transformed into an emulsion, as visualized in Figure 13.7. Moving upwards through the transition, bubble-filled water gradually gives way to spray-filled air, and the actual surface of the ocean becomes ill-defined. The figure shows the altitude at which the volumes of liquid water and air are equal.

Spray droplets are ejected upward and partially or completely evaporate; those that do not completely evaporate return to the ocean. The upward moving water mass (including the vapor) must be accelerated to some representative horizontal velocity of the air. Presumably, the vapor and smaller droplets reach higher altitudes on average and are accelerated to larger horizontal velocities. The acceleration of the water mass requires work and acts as drag on the airflow. The stress (force per unit area) acting on the airflow may be expressed

$$\tau_{spray} = -M_{spray} U, \tag{13.19}$$

where M_{spray} is the upward mass flux of water (including vapor) per unit area and U is a mean free stream horizontal velocity. (It is a curious fact that even when there is no spray, evaporation from a water surface into a moving airstream exerts a drag on the air.) Note that only droplets returning into the ocean will exert a stress on the water side; as there is a net mass exchange there is no requirement that the air- and water-side stresses be equal and opposite.

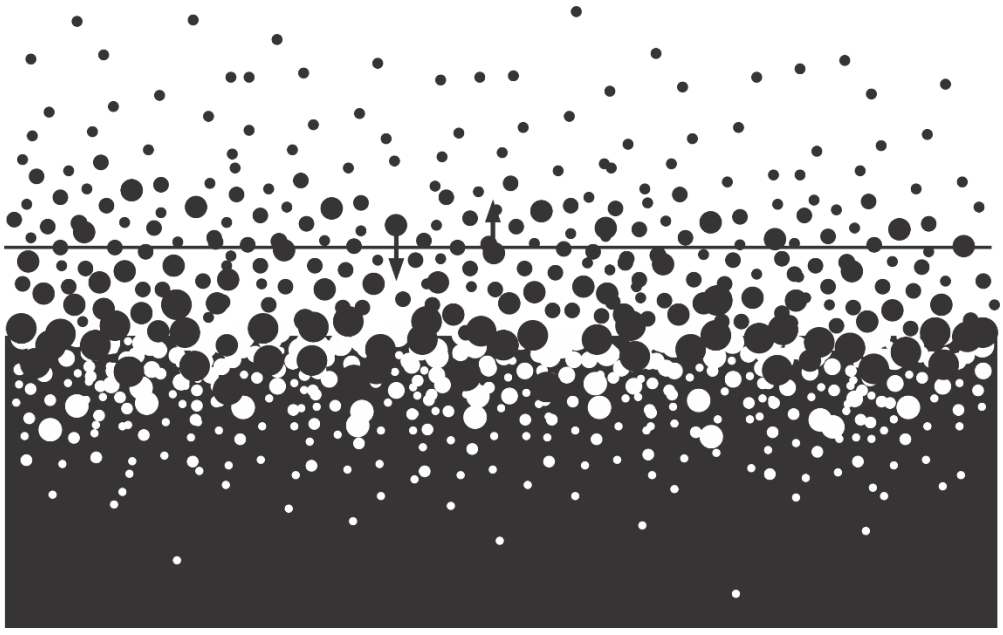


Figure 13.7: Visualization of the transition from bubble-filled water to spray-filled air at extreme wind speeds. The horizontal black line denotes the altitude at which, on average, the volumes of liquid water and air are equal.

It is helpful to consider the spectral distribution of this mass flux as a function of the size of water drops ejected. This spectral upward liquid water mass flux is referred to as the sea spray generation function (SSGF), and we will return presently to a more detailed discussion of this function.

Now consider the flux of enthalpy through the horizontal line in Figure 13.7 Here, by definition, the volumes of water and air are equal, but because the density of water is about 1,000 times

that of air at sea level, almost all the mass at the altitude is in the water drops. The flux of enthalpy will be dominated by the water drops, assuming their vertical motion is of the same (or greater) order of magnitude as that of air at this level. As a droplet travels upward, we will suppose that its temperature drops from the ocean temperature T_s to a representative air temperature T_a. The decrease of its enthalpy will therefore be $m_i c_l \left(T_s - T_a\right),$ where c_l is the specific heat capacity of liquid water and m_i is the mass of water in droplets of size category i. To conserve total system energy, that same quantity of enthalpy must be added to the air.

After that, a fraction of the droplet f_i evaporates. Here it is important to recognize that if $f_i = 1,$ i.e. the whole drop evaporates, there is no further enthalpy increase in the air because all the heat of vaporization must come from the air – we are just trading sensible for latent heat. But if not all of the droplet evaporates, the remaining mass, $m_i \left(1 - f_i\right),$ falls back into the ocean having cooled to a temperature $T_i.$ (These falling drops therefore cool the ocean.) The difference between the upward and downward enthalpy fluxes by the droplets from this second step is $m_i \left(1 - f_i\right)\left(T_a - T_i\right)$. Putting the two steps together, the flux of enthalpy from the ocean to the atmosphere by spray of size category i is

$$F_i = M_i c_l \left[T_s - T_a + \left(1 - f_i\right)\left(T_a - T_i\right)\right], \tag{13.20}$$

where M_i is the sea spray generation function. Andreas and Emanuel (2001) derived this relationship exactly from a rigorous analysis of thermodynamic energy conservation. Note that the integrated system enthalpy increases with time because of the dissipation of the kinetic energy of the water and air motion, though it is not so clear how this energy increase is partitioned between the atmosphere and ocean.

For typical conditions in the tropics, the second term on the right will be larger than the first term, simply because the difference between the droplet and air temperatures is usually larger than the difference between the air and ocean water temperatures. In the case of high latitude polar lows, however, the first term may be of equal or greater magnitude.

The net enthalpy flux is obtained by summing (13.20) over all drop size categories. Knowledge of the physics of drop evaporation is required to determine f_i and T_i, though we note that the wet bulb temperature, T_w is a rigorous lower bound on T_i.

Very small droplets will be lofted to greater height and will likely evaporate entirely, contributing little to the net enthalpy flux. At the other end of the size spectrum, very large droplets may not rise far enough to experience enthalpies appreciably different from that of saturation at sea surface temperature. Andreas (1992) estimated that most of the spray-associated enthalpy flux is accomplished by droplets of initial radii between $10\,\mu m$ and $300\,\mu m$. One great simplification for droplets in this size range is that one can assume, to a good approximation, that such droplets are re-entrant and cool right down to the wet-bulb temperature. Moreover, somewhat less than 1% of the mass of a droplet needs to evaporate to cool it down to the wet-bulb temperature (Andreas and Emanuel, 2001), so for these drops f_i in (13.20) can be very small, thereby contributing to a larger enthalpy flux.

Figure 13.5 shows the evolution of the temperature and radius of a droplet whose initial size is $100\,\mu m$ in a typical tropical boundary layer environment, using the model of Andreas (1989); (see also Andreas 1995). Note that the time axis is logarithmic. The droplet temperature drops to close to the wet-bulb temperature in less than 1 s. It remains near that temperature for almost 100 s, during which it would likely fall back into the sea, retaining most of its mass for several tens of seconds. If we assume that this drop re-enters the ocean somewhere between 1 and 50 second after it is forms, then to a good approximation $T_i = T_w$ and $f_i \cong 0$ in (13.20), so the latter becomes, to a good approximation

$$F_i \cong M_i c_l \left(T_s - T_w\right). \tag{13.21}$$

If we further assume that most of the spray mass ejected into the boundary layer consists of drops in the $10-300\,\mu m$ range, then summing (13.21) over all the drop sizes results in

$$F_{spray} \cong M_{spray} c_l \left(T_s - T_w\right). \tag{13.22}$$

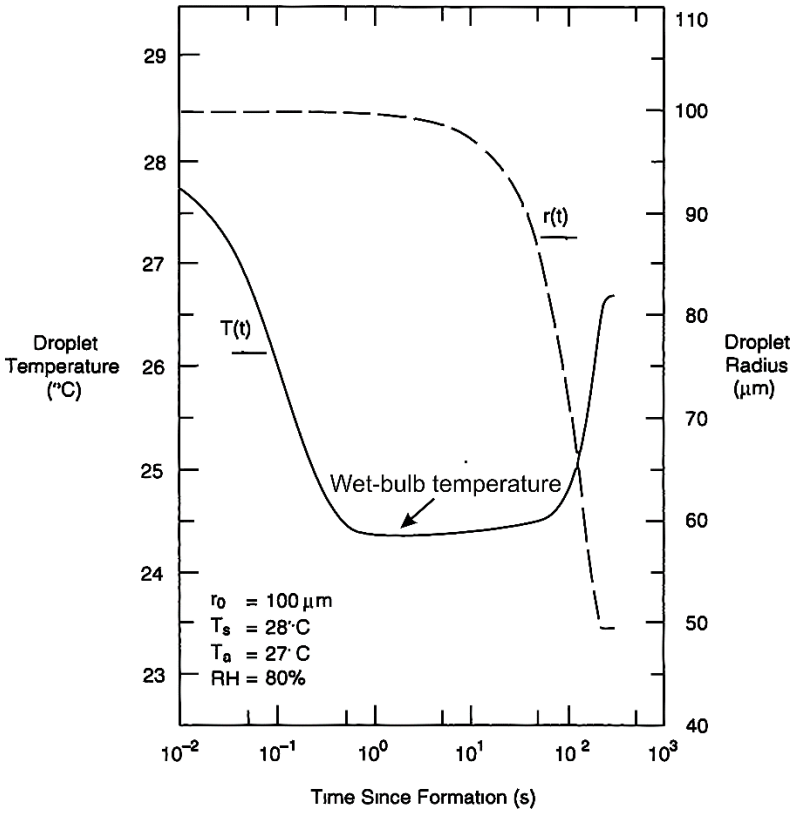


Figure 13.8: Evolution of temperature (°C; solid curve, left axis) and drop radius (μm; dashed curve, right axis) of a droplet of an initial radius of $100\ \mu m$ in a thermodynamic environment typical of a tropical boundary layer.

The problem of determining both the momentum and enthalpy fluxes collapses into a determination of the net spray mass flux, M_{spray}. Before we discuss approaches to estimating M_{spray}, we can make an interesting point about spray effects on hurricane potential intensity. Let

us conjecture that at extreme wind speeds, almost all the momentum and enthalpy flux is rate-limited by sea spray as opposed to conventional surface roughness. If we follow through the derivation of potential intensity in Chapter 9, but use (13.19) and (13.22) in place of conventional aerodynamics formulae for the surface momentum and enthalpy fluxes, we arrive at a different definition of potential intensity:

$$V_p^2 = \frac{T_s - T_o}{T_o} c_l \left(T_s - T_w \right), \tag{13.23}$$

where recall that T_o is the outflow temperature. The spray mass flux drops out, and no exchange coefficients appear in (13.23). For a sea surface temperature of $28^o C$, a near-surface relative humidity of 80%, and a modified thermodynamic efficiency of ½, (13.23) and the conventional formula for potential intensity both yield a potential intensity of around $77\, ms^{-1}$. But maintaining a relative humidity of 80% and the same modified thermodynamic efficiency, and increasing the sea surface temperature to $35^o C$ increases the conventional potential intensity to $95\, ms^{-1}$ but the potential intensity given by (13.23) only to about $81\, ms^{-1}$. Therefore, in a spray-dominated regime, potential intensity is less sensitive to temperature.

Of central importance to calculating spray momentum and enthalpy fluxes is the spray mass flux, M_{spray}. There have been many attempts to estimate this as a function of wind speed, mostly from field and laboratory experiments. Before we discuss these, a few theoretical inferences may prove helpful.

First, we consider a maximally simple system consisting of a semi-infinite depth of water with no density stratification, surmounted by a semi-infinite, unstratified atmosphere. We apply an infinitesimal horizontal pressure gradient to the atmosphere that, when integrated over the infinite depth of the fluid, yields a finite constant which, from overall atmospheric momentum balance must equal the surface wind stress when the system is in statistical equilibrium:

$$\int_0^\infty \frac{1}{\rho} \frac{\partial p}{\partial x} dz = u_*^2, \tag{13.24}$$

where ρ is the atmospheric density and u_* is the friction velocity. Specifying the pressure gradient is equivalent to specifying the surface stress in the statistical equilibrium state, and it is the properties of this state that interest us here. We also take the acceleration of gravity to be constant over this system.

Careful consideration of this system reveals seven dimensional control parameters: $u_*, g,$ the densities of water, $\rho_l,$ the density of air at the surface, $\rho_a,$ the kinematic viscosities of the two fluids, ν_l and ν_a, and the kinematic surface tension σ. Inspection of the governing equations shows that in this system, gravity and density only appear in a particular combination we refer to as the effective gravity:

$$g' \equiv g \frac{\rho_l - \rho_a}{\rho_l}. \tag{13.25}$$

Because the atmospheric density is three orders of magnitude smaller than the ocean density, there is hardly any difference between the actual and effective gravities. In reality, there are only 5 dimensional parameters governing this system and, according to the Buckingham Pi theorem, the system is governed by three nondimensional parameters. We first define three length scales:

$$L_c = \frac{u_*^2}{g'}, \qquad L_\nu = \frac{\nu_a}{u_*}, \qquad L_\sigma = \sqrt{\frac{\sigma}{g'}}. \tag{13.26}$$

The first is the Charnock length so important in surface waves, the second is a diffusive length scale, and the third can be interpreted as governing the radius of the largest water drop that is stable while falling through air.

Returning to the Pi theorem, we formulate three non-dimensional parameters:

$$R_u \equiv \frac{u_*^4}{\sigma g'}, \qquad R_\sigma \equiv \frac{\sigma}{\nu_l^{4/3} g'^{1/3}}, \qquad R_\nu \equiv \frac{\nu_l}{\nu_a}. \tag{13.27}$$

Note that only the first of these depends on u_*, and referring to (13.27), R_u can be interpreted as the ratio of the Charnock length to the radius of the largest water drop that remains stable falling through air. Consequently, for an atmospheric and ocean with fixed effective gravity, viscosities and surface tension, the physical character of the system should depend only on R_u. For example, the fractional area covered by whitecaps should be just a function of R_u and the spray mass flux per unit area should be given by

$$M_{spray} = \rho u_* \Phi\left(R_u\right), \tag{13.28}$$

where ρ is the density of water and Φ is a function that would have to be determined experimentally.

There have been many attempts to estimate spray mass fluxes both in the field, under high wind speed conditions, and in wind-wave laboratory flumes. Unfortunately, the results of these attempts differ over many orders of magnitude as (Sroka and Emanuel 2021; Ma et al. 2020; Xu et al. 2021). Figure 13.9 shows a summary of many of these estimates at the time of this writing. Note that these particular estimates are only at 10-meter wind speeds below hurricane-force, but even so, they differ by large amounts. Some of the differences may be owing to different sea states; the dimensional arguments presented above assume that ocean-atmosphere system is in a state of statistical equilibrium. (Curiously, the oldest estimate (dark blue line) and the most recent estimate (black dots) are reasonably close together in Figure 13.9. The dark blue line, from Andreas (1992), represents a cubic dependence of M_{spray} on U_{10}.)

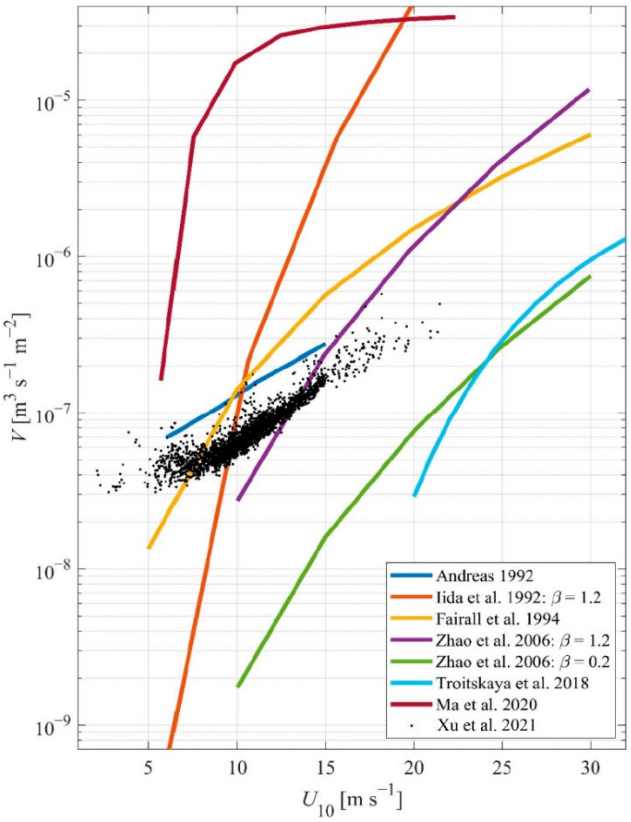
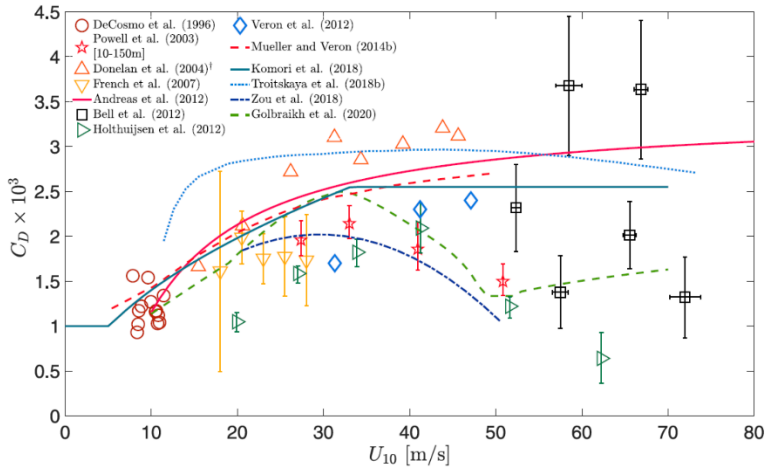


Figure 13.9: Various estimates of spray volume flux $(= M_{spray} / \rho_l)$ *as a function of 10-meter wind speed. The references in the legend are listed in the bibliography of this chapter.*

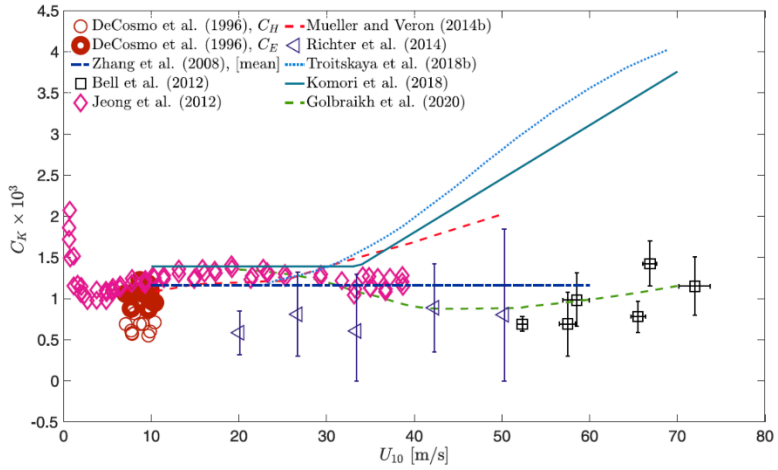
The current very large uncertainty in spray fluxes translates into uncertainty about their degree of importance in the structure and intensity of tropical cyclones and may have a bearing on the relative lack of progress in forecasting tropical cyclone intensity. Given limitations in laboratory and direct numerical simulations of sea spray, and the difficulty of directly measuring spray in the field, several attempts have been made to infer net turbulent surface fluxes from budget residuals. These estimates are for the total surface fluxes and are not able to distinguish spray-mediated fluxes from the net. Palmén and Riehl (1957) and Riehl and Malkus (1961) pioneered the estimation of surface wind stress from the residual in the angular momentum budget calculated from aircraft and dropsonde measurements in hurricanes, and this technique was further advanced by Hawkins and Rubsam (1968). More recently, Bell et al. (2012) used very dense radial deployments of GPS-based dropsondes in strong Atlantic hurricanes to estimate surface fluxes of angular momentum and energy. While broadly successful, the error bars on the estimated fluxes are large. Powell et al. (2003) used the vertical wind profiles measure by dropsondes in hurricanes to infer surface roughness. Their estimates of the surface drag coefficient show a peak around 30 ms^{-1}, declining with wind speed at higher speeds. Quite a few subsequent studies also concluded that the effective surface drag coefficient declines with wind speed for winds greater than marginal hurricane force. Theoretical studies attempt to explain this decline of the drag coefficient in terms of spray particle dynamics and the imposition of an effective stable stratification by sea spray (e.g. Makin 2005; Bye and Jenkins 2006; Kudryavtsev 2006; Kudryavtsev and Makin 2011; Kudryavtsev et al. 2012; Richter and Sullivan 2013). Good reviews of surface fluxes at high wind speeds are provided by Sterl (2017) and Sroka and

Emanuel (Sroka and Emanuel 2021). Various estimates of the surface drag and enthalpy exchange coefficients are compared in Figure 13.10. It has recently been proposed that the character of sea spray and surface fluxes may be affected by naturally occurring bio-surfactants (Vanderplow et al. 2020) and it has been conjectured that artificial surfactants might be used to suppress tropical cyclone intensity (Simpson and Malkus 1964; Simpson and Simpson 1966).

While detailed numerical simulations of tropical cyclones for research and forecasting have improved greatly over time, the aforementioned uncertainty in the physics and formulation of surface fluxes is regarded as a major remaining source of errors in tropical cyclone intensity forecasts. This weakness, however, presents an interesting opportunity, which is to use ensemble tropical cyclone data assimilation to back-off optimal values of the surface exchange coefficients and their dependence on surface winds. A good start in this direction was undertaken by Nystrom et al. (2021) who tested whether such a technique could recover the "correct" parameter dependencies from assimilating data from a control simulation into a parallel simulation in which the initial parameter values were changed. Among other issues, the efficacy of this technique depends to some extent on how good the rest of the model is; otherwise, the estimated surface exchange coefficients might partially compensate for other model defects. (This problem is not an issue in Nystrom et al.'s observing system simulation experiment.) Nevertheless, given the problems in estimating exchange coefficients at high wind speeds from theory, numerical simulations, and laboratory and field experiments, estimating these coefficients by optimizing their values and wind dependencies from high-quality forecast models might provide an interesting and practical addition to our collective understanding of surface fluxes at high winds speeds.



(a)



(b)

Figure 13.10: Various estimates of surface exchange coefficients. a) The sea surface drag coefficient C_D as a function of the 10-meter wind speed U_{10} from serveral studies. The plotted Donelan et al. (2004) data reflect the corrections published in Curcic and Haus (2020) for the momentum budget method. b) The sea surface enthalpy exchange coefficient C_k as a function of U_{10}. See Sroka and Emanuel (2021) for the references listed in the figure legends.

13.3: Near-inertial ocean current and temperature response

Surface waves are only one of a set of oscillations possible in the oceans. The others range from acoustic waves at the high frequency end of the spectrum down to Rossby waves at the low frequency end. Tropical cyclones can excite any of these modes.

Of special interest here, for reasons that shall become apparent, are near-inertial oscillations, i.e., oscillations whose frequency is close to the Coriolis parameter f.

Consider small perturbations to an inviscid, stably stratified Boussinesq fluid at rest on an f-plane. The linearized equation set for this case is

$$\begin{aligned}
\frac{\partial u}{\partial t} &= -\alpha_0 \frac{\partial p}{\partial x} + fv \\
\frac{\partial v}{\partial t} &= -\alpha_0 \frac{\partial p}{\partial y} - fu \\
\frac{\partial w}{\partial t} &= -\alpha_0 \frac{\partial p}{\partial z} + B \\
\frac{\partial B}{\partial t} &= -N^2 w \\
\frac{\partial u}{\partial x} &+ \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0,
\end{aligned} \tag{13.29}$$

where *u*, *v*, and *w* are the velocity perturbations, *p* is the perturbation pressure, *B* is the buoyancy perturbation, α_0 is the mean state (constant) specific volume, and N is the buoyancy frequency.

Eliminating all the variables in favor of the vertical velocity, w, results in

$$\frac{\partial^2}{\partial t^2} \nabla_3^2 w + N^2 \nabla_2^2 w + f^2 \frac{\partial^2 w}{\partial z^2} = 0, \tag{13.30}$$

where ∇_2^2 and ∇_3^2 are the two-and three-dimensional Laplacian operators, respectively. We look for regular eigenmode solutions to (13.30) of the form

$$w = w_0 e^{i(kx+ly+mz-\omega t)}, \tag{13.31}$$

where *k*, *l*, and *m* are the zonal, meridional, and vertical wavenumbers and ω is the frequency. Substituting (13.31) into (13.30) gives the dispersion relation for gravity-inertia waves:

$$\omega^2 = \frac{N^2\left(k^2 + l^2\right) + f^2 m^2}{k^2 + l^2 + m^2}. \tag{13.32}$$

In the limit of very long waves (compared to the deformation radius based on the scale $1/m$), for which $N^2\left(k^2 + l^2\right) << f^2 m^2$, the frequency approaches $\pm f$; these are the lowest frequency waves supported in this system and are referred to as inertial oscillations. The particle trajectories in such oscillations are clockwise circles in the northern hemisphere. Long, shallow waves in this limit are referred to as near-inertial waves. These are dispersive waves that propagate in all three spatial directions.

Tropical cyclones are very efficient producers of near-inertial oscillations in the ocean. To begin with, wind stress acting on the sea surface produces an acceleration in the ocean that is more or less uniformly distributed through the depth of the ocean mixed layer, typically around *50 m*.

But this stress acts over horizontal scales of hundreds of kilometers, so that the horizontal wavenumbers are small compared to the vertical wave numbers, i.e. $N^2\left(k^2+l^2\right) << f^2 m^2$.

Moreover, the turning of the wind direction as a tropical cyclone passes near a point on the ocean surface can resonantly excite inertial oscillations to the right of the track of the cyclone, in the northern hemisphere. This is illustrated in Figure 13.11. To the right (north, in this case) of the track of the cyclone, the wind stress rotates around the compass in a clockwise sense, in the same sense as particle trajectories in a near-inertial oscillation, so that wind energy is pumped into the oscillating current. The opposite is true to the left (south) of the track.

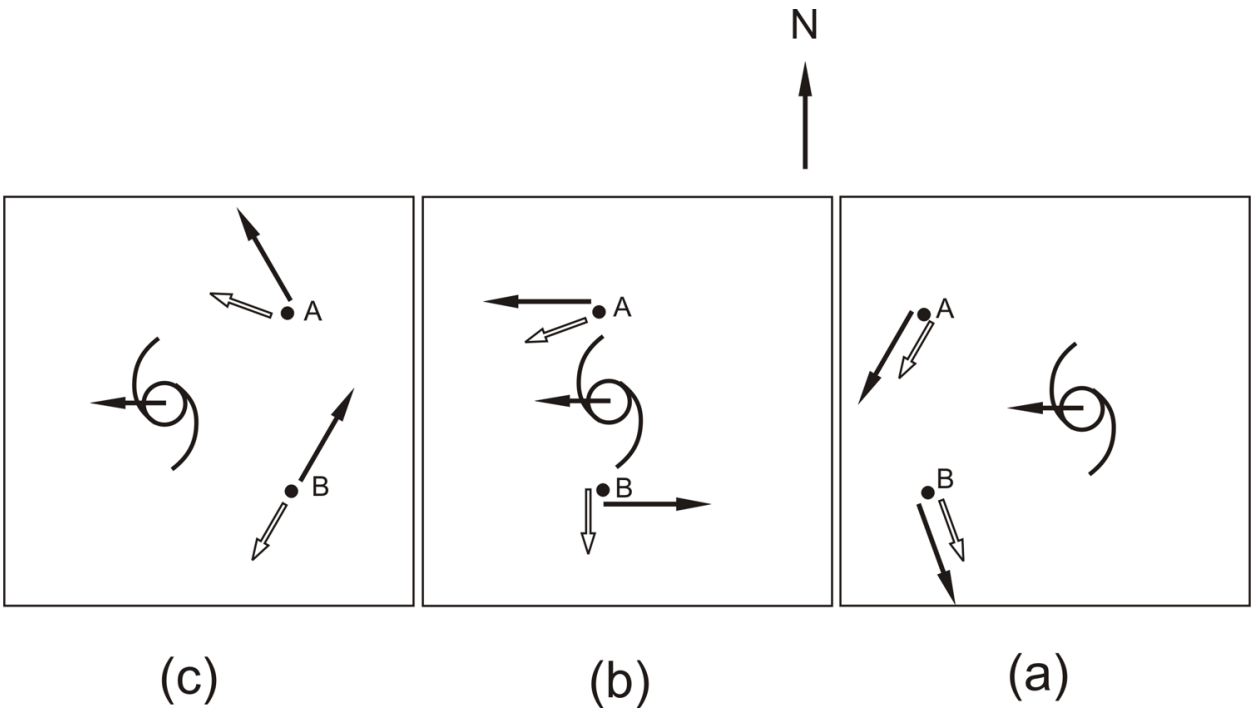


Figure 13.11: Illustration of the resonant excitation of ocean inertial oscillations by a westward-moving tropical cyclone in the northern hemisphere. The three maps represent a time sequence, in this case from right to left (a-c). The wind stress at fixed points like point A to the right of the track rotates clockwise around the compass, in the same sense as particle trajectories in a near-inertial oscillation, so that the wind stress continuously pumps energy into the oscillation. But at points to the left of the track, such as point B, the stress rotates around the compass in the opposite sense of particle trajectories in a near-inertial oscillation, so that after the time represented in panel (b), the wind stress damps the oscillation.

Observations and detailed numerical models of tropical cyclones indeed show a strong rightward bias of near-inertial current excitation in the ocean (leftward in the southern hemisphere). Figure 13.12 shows a snapshot of a numerical simulation using a full-physics ocean model driven by a steadily translating, westward moving tropical cyclone wind field in the northern hemisphere (Schade 1994). The tropical cyclone strength varies over time in response to variations in the ocean temperature. A strong, regular near-inertial oscillation is evident to the right of the storm track. In observations of real tropical cyclones, this inertial wake is evident over may cycles, but eventually the wave energy disperses downward into the deeper ocean and laterally as well. The currents can be very powerful, up to about 2 m s^{-1}.

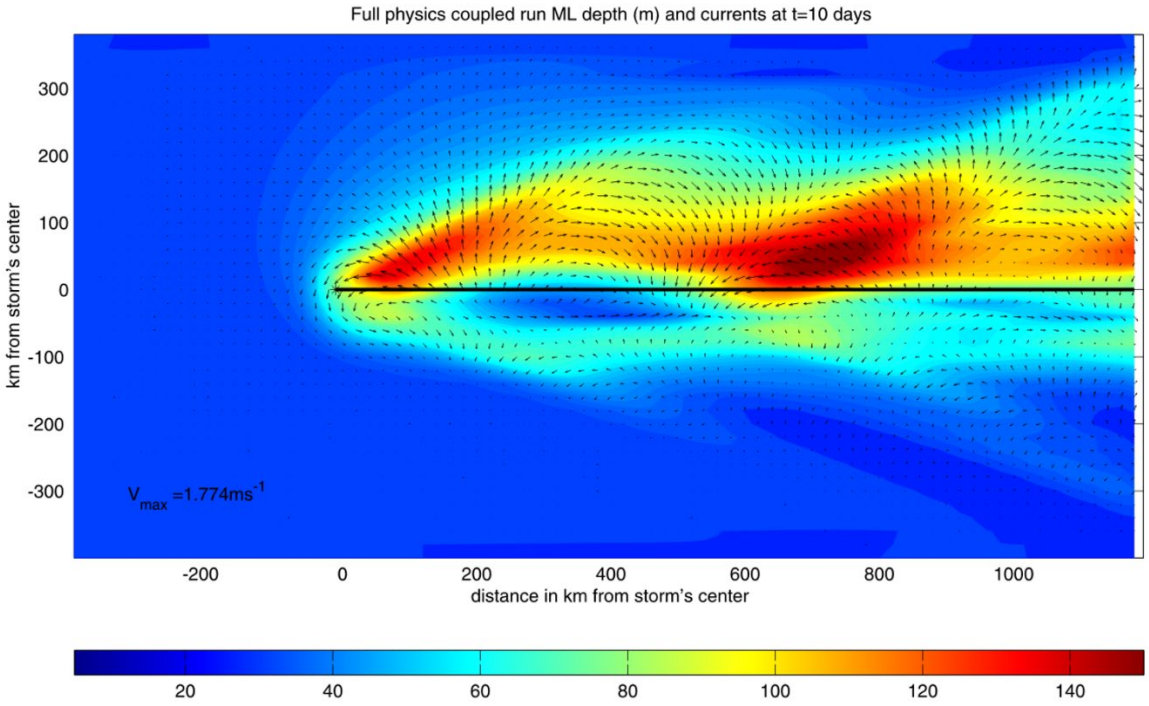


Figure 13.12: Ocean mixed layer currents (arrows) and mixed layer depth (m; shaded) in a numerical simulation of a steadily translating, westward moving tropical cyclone wind field in the northern hemisphere. The cyclone is translating at 7 m s^{-1}. The length of the arrows is proportional to the ocean mixed layer current speed; the maximum speed in the domain is 1.77 m s^{-1}. The horizontal black line represents the track of the cyclone; the asterisk at the left end of the line is the current position of the cyclone center. The undisturbed mixed layer depth is 40 m.

Notice also that the ocean mixed layer deepens rather profoundly to the right of the track. The physics underlying this ocean mixed layer deepening were described in detail by Price (1981). Turbulence in the mixed layer rapidly mixes bulk horizontal momentum throughout the layer, creating a jump in bulk velocity at the base of the mixed layer, where there is also a jump in potential density. This jump is unstable to Kelvin-Helmholtz instability, resulting in turbulence that entrains quiescent but usually colder waters from beneath the mixed layer. This deepens and cools the mixed layer. Figure 13.13 illustrates what a typical ocean temperature profile might look like before and after turbulent entrainment at the base of the mixed layer. Ignoring the dissipative heating associated with the turbulence, the mass-integrated enthalpy must be conserved, and assuming that temperature mixes linearly, the volumes of cooling and warming must be equal, as illustrated in the figure. Given a constant vertical temperature gradient Γ below the mixed layer, it is easy to show that

$$h\Delta T = h_0 \Delta T_0 + \frac{1}{2}\Gamma\left(h^2 - {h_0}^2\right), \tag{13.33}$$

where the quantities are defined in Figure 13.13.

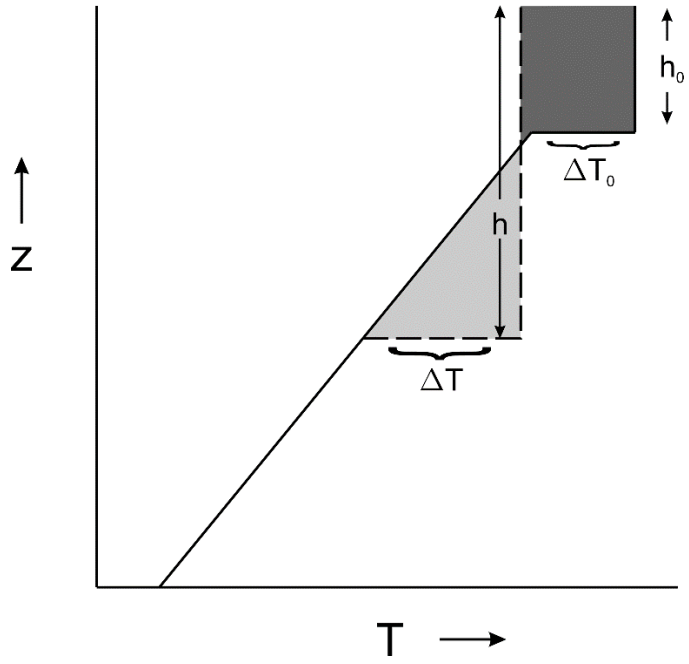


Figure 13.13: Sketch of a vertical temperature profile in the ocean before and after an entrainment event at the base of the mixed layer. The solid and dashed lines show the profiles before and after the mixing, respectively. The initial and final mixed layer depths are h0 and h, and the initial and final jumps in temperature at the base of the mixed layers are ΔT_0 and ΔT . The dark and light gray shadings show regions of cooling and warming, respectively; these areas must be equal for global enthalpy conservation.

As the turbulence is relatively fast, we might assume that it acts to maintain the relevant stability parameter in a near-neutral condition. On the basis of this reasoning as well as laboratory experiments with deepening mixed layers, Price (1979) showed that the relevant stability parameter is a bulk Richardson number defined

$$Ri \equiv \frac{gh\Delta\rho}{\rho U^2}, \tag{13.34}$$

where h is the mixed layer depth, $\Delta\rho$ is the (positive) jump in density across the base of the mixed layer, ρ is the mean water density, and U is the mixed layer flow speed. If we assume for the sake of simplicity that water density is controlled by temperature alone (in reality, it also depends on salinity and pressure), we can re-write (13.34) approximately as

$$Ri \simeq \frac{g\beta h\Delta T}{U^2}, \tag{13.35}$$

where β is the coefficient of thermal expansion of seawater. Eliminating ΔT between (13.33) and (13.35) gives

$$\frac{Ri}{g\beta}U^2 = h_0\Delta T_0 + \frac{1}{2}\Gamma\left(h^2 - {h_0}^2\right). \tag{13.36}$$

Assuming that the Richardson Number Ri remains close to a critical value, (13.36) shows that the velocity U and the mixed layer depth are proportional. (This ignores vertical advection, as would be induced by Ekman pumping, for example.) A casual inspection of Figure 13.12 confirms this relationship and explains why most of the mixed layer deepening is to the right of the track, where the inertial oscillations are resonantly forced.

This system can be closed by considering the momentum balance of the mixed layer. Assuming that, on the time scale of the passage of a tropical cyclone, all the momentum is confined to the mixed layer, conservation of momentum may be expressed

$$\left(\frac{\partial}{\partial t} + \mathbf{V}\bullet\nabla\right)(h\mathbf{V}) + f\hat{k}\times h\mathbf{V} = \boldsymbol{\tau}_s / \rho, \tag{13.37}$$

where $\mathbf{V}$ is the velocity vector, $\boldsymbol{\tau}_s$ is the wind stress, and we have ignored pressure gradients, which tend to be weak within inertial oscillations. In this notation, $U = |\mathbf{V}|$. Given wind stress and initial conditions, (13.36) and (13.37) can be solved for the evolution of mixed layer depth and velocity. Once those have been obtained, (13.33) can be solved for the temperature. Figure 13.14 shows the departure of the sea surface temperature from its initial value, corresponding to the fields shown in Figure 13.12. As expected, there is a reasonably good correspondence between the sea surface temperature depression and the degree of deepening of the mixed layer (Figure 13.12). Most of the cooling occurs behind and to the right of the storm track.

Almost all observed tropical cyclones produce cold wakes, and these are easily detected by comparing satellite-derived sea surface temperatures before and after the passage of the storms. Figure 13.15 shows the cold wake produced by Atlantic hurricane Gert of 1999. Temperature profiles taken before the passage of Gert and later in Gert's cold wake show the clear mixing signal of cooling of the surface waters and compensating warming below (Figure 13.15C).

In both the numerical simulation shown in Figure 13.14 and in the case of Hurricane Gert, the maximum sea surface anomaly is around 4°C. But so far as feedback is concerned, it is cooling under the storm's inner core that matters most, according to the arguments presented in Chapter 9. Inspection of Figure 13.14 shows that cooling near the current storm center position is only around 1°C. How much difference could that make to storm intensity?

We can address this question by asking how cold the sea surface would need to be under the eyewall of a hurricane for air saturated at that temperature to have the same enthalpy at the undisturbed boundary layer air. At such a sea surface temperature, it would be impossible for the entropy to increase as air flowed inward to the eyewall, and a tropical cyclone could not exist at all.

Equating these two enthalpies can be expressed:

$$c_p T_{seye} + L_v q^*\left(T_{seye}\right) = c_p T_a + \mathcal{H}_a L_v q^*\left(T_a\right), \tag{13.38}$$

where c_p is the heat capacity at constant pressure, L_v is the latent heat of vaporization, T_{seye} is the sea surface temperature under the eyewall, T_a is the ambient air temperature, and $\mathcal{H}_a$ is the ambient near-surface relative humidity.

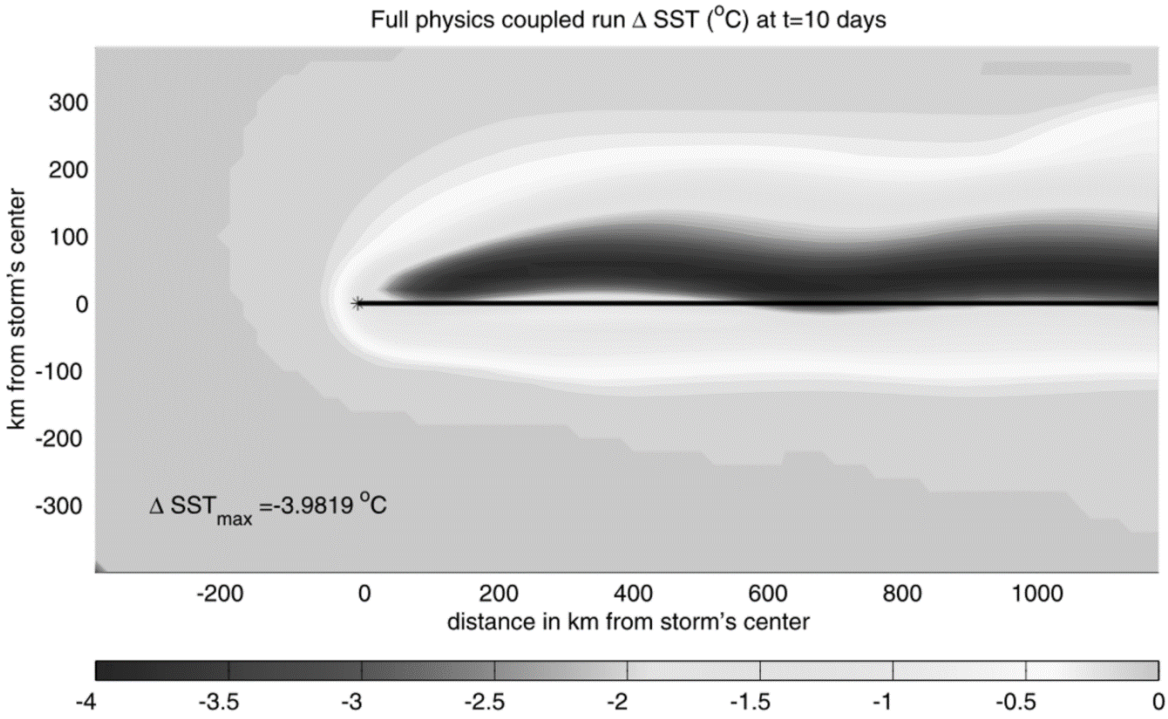


Figure 13.14: Sea surface temperature anomaly (°C) corresponding to the simulation shown in Figure 13.12. The maximum negative anomaly is 4 °C.

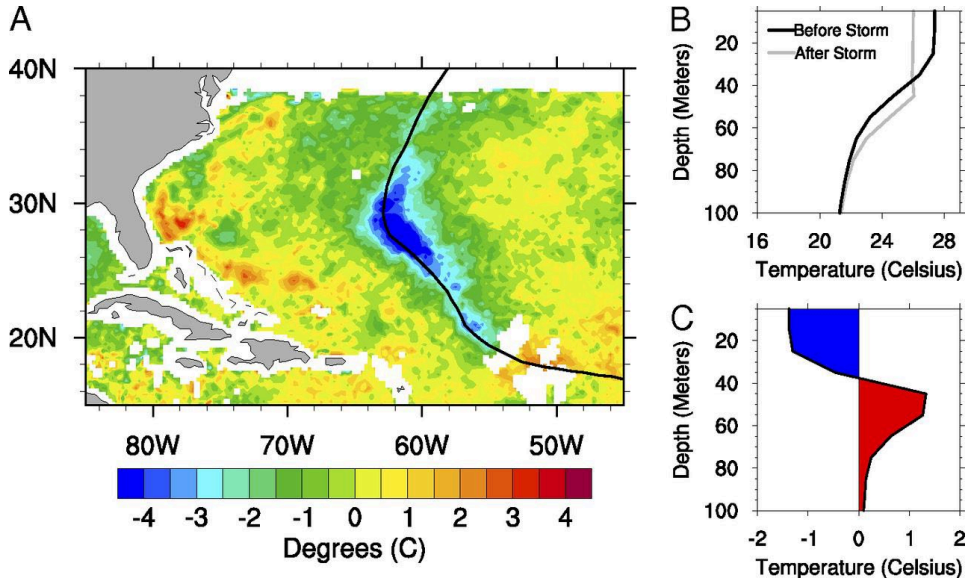


Figure 13.15: Upper-ocean response to Hurricane Gert in 1999: (A) observed surface temperature anomaly showing cooling along the storm track primarily as a result of vertical ocean mixing [the data are from the NASA Tropical Rainfall Measuring Mission (TRMM) Microwave Imager], (B) vertical ocean temperature profiles averaged over Hurricane Gert's track before (black curve) and after (gray curve) the storm [analyzed using the NASA ocean reanalysis from Estimating the Circulation and Climate of the Ocean, Phase II (ECCO2)], and (C) the difference between the two temperature profiles shown in (B), highlighting the cool anomaly (above 40 m) and warm anomaly (below 40 m) that resulted from vertical ocean mixing following storm passage. [Source: The figure and caption are from Sriver (2013).]

We can also write

$$T_{seye} = T_{sa} + \Delta T_s, \tag{13.39}$$

where ΔT_s is the change of sea surface temperature from its undisturbed value, T_{sa}, to under the eyewall. We can then write (13.38) as

$$c_p \Delta T_s + L_v q^*\left(T_{sa} + \Delta T_s\right) = c_p\left(T_a - T_{sa}\right) + \mathcal{H}_a L_v q^*\left(T_{sa} + T_a - T_{sa}\right). \tag{13.40}$$

Given that q^* is a function of temperature only (at a specified surface pressure), (13.40) can be solved numerically for ΔT_s given the ambient relative humidity and sea surface and air temperatures. But since ΔT_s and $T_{sa} - T_a$ are usually small quantities, we can linearize the Clausius-Clapeyron equation about the ambient sea surface temperature and find an approximate closed-form solution for ΔT_s:

$$\Delta T_s \simeq -\frac{L_v q^*\left(T_{sa}\right)\left(1-\mathcal{H}_a\right) + \left(T_{sa} - T_a\right)\left(c_p + \mathcal{H}_a \dfrac{L_v^2 q^*\left(T_{sa}\right)}{R_v {T_{sa}}^2}\right)}{c_p + \dfrac{L_v^2 q^*\left(T_{sa}\right)}{R_v {T_{sa}}^2}}. \tag{13.41}$$

In (13.41) R_v is the gas constant for water vapor. For an ambient sea surface temperature of 27°C, and ambient air temperature of 26°C and an ambient relative humidity of 80%, ΔT_s is about -3.4 °C. Consequently, a cooling of 3.4 °C would suffice to make a tropical cyclone thermodynamically impossible to sustain, and from this we see than even a 1-degree cooling of the eyewall should have an important negative effect on tropical cyclone intensity.

Note in Figure 13.12 that there is very little turning of the ocean velocity vector between the onset of storm winds and the arrival of the storm core. This is also true of the wind direction itself, if the storm is moving in a reasonably straight line. For the narrow purpose of estimating the actual sea surface temperature depression in the storm core, we might neglect both the nonlinear advection term and the Coriolis term, giving

$$\frac{\partial\left(hU\right)}{\partial t} = \tau_s / \rho, \tag{13.42}$$

where U is the scalar current speed and τ_s is the magnitude of the wind stress. Integrating in time and taking $U = U_0$ and $h = h_0$ at time $t = 0$ gives

$$hU = h_0 U_0 + \int_0^t \frac{\tau_s}{\rho} dt. \tag{13.43}$$

In (13.36), it is necessary that $\dfrac{Ri}{g\beta} {U_0}^2 = h_0 \Delta T_0$ so that the initial condition is satisfied. With that stipulation, our closed system consists of (13.36) and (13.43), and we can then use (13.33) to diagnose the temperature jump at the base of the mixed layer. Before doing this, it is helpful to normalize the variables, which we do as follows:

$$
\begin{aligned}
&h \to h_0 h, \\
&(\Delta T_0, \Delta T) \to \Gamma h_0 (\Delta T_0, \Delta T), \\
&U \to h_0 \sqrt{\frac{g\beta\Gamma}{Ri_c}} U,
\end{aligned}
\tag{13.44}
$$

where Ri_c is a critical Richardson Number. We also define a nondimensional time integral of the surface stress:

$$
\int \tau dt \equiv \frac{1}{\rho h_0} \sqrt{\frac{Ri_c}{g\beta\Gamma}} \int_0^t \tau_s dt. \tag{13.45}
$$

With these definitions, (13.36), (13.43), and (13.33) become, respectively,

$$
U^2 = \Delta T_0 + \frac{1}{2}\left(h^2 - 1\right), \tag{13.46}
$$

$$
hU = \sqrt{\Delta T_0} + \int \tau dt, \tag{13.47}
$$

and

$$
h\Delta T = \Delta T_0 + \frac{1}{2}\left(h^2 - 1\right). \tag{13.48}
$$

We note that the change over time of the nondimensional sea surface temperature is given by

$$
\Delta SST = 1 - h + \Delta T - \Delta T_0. \tag{13.49}
$$

There are just two nondimensional parameters in this system: The normalized initial temperature jump across the base of the mixed layer, ΔT_0, and the normalized time-integrated wind stress $\int \tau dt$. Eliminating U between (13.46) and (13.47) gives

$$
h^2 = \frac{1}{2} - \Delta T_0 + \sqrt{\left(\frac{1}{2} - \Delta T_0\right)^2 + 2\left(\sqrt{\Delta T_0} + \int \tau dt\right)^2}. \tag{13.50}
$$

Given this solution we use (13.48) to solve for ΔT and (13.49) to solve for ΔSST. Solutions for h and ΔSST are shown in Figure 13.16 as functions of the nondimensional ΔT_0 and the integral of the surface stress over time, $\int \tau dt$.

To a good approximation, the normalized mixed layer deepening rate is independent of the initial temperature jump across the base of the mixed layer (Figure 13.16a). In contrast, the sea surface temperature cools more rapidly over time (time-integrated surface stress) for larger initial mixed layer temperature jump (Figure 13.16b). For a given ΔT_0, the sea surface temperature changes nearly with the square root of $\int \tau dt$, especially at later times.

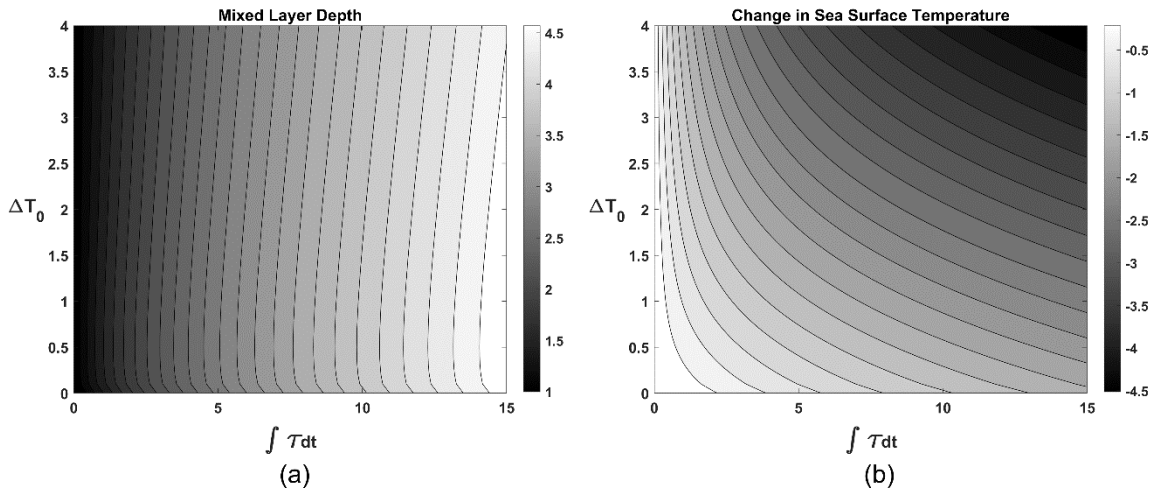


Figure 13.16: Nondimensional mixed layer depth (a) and nondimensional sea surface temperature anomaly (b) as functions of normalized surface stress integrated over time and the initial normalized temperature jump across the base of the mixed layer.

Schade (1994) developed a two-way interactive coupled ocean – tropical cyclone model in which the intensity of the tropical cyclone is affected by the sea surface temperature response of the underlying ocean, whose mixing is driven by the hurricane's wind stress field. The ocean was initially horizontally homogeneous, and the atmospheric environment was fixed with no wind shear. The translation speed and direction were fixed constants. Figures 13.12 and 13.14 are examples of output form the model. It was found that with constant translation speed, tropical cyclone intensity reached a steady state.

Schade and Emanuel (1999) identified five nondimensional parameters that governed the degree of negative feedback of ocean mixing on cyclone intensity. The negative ocean temperature feedback was measured by a nondimensional parameter F_{SST} defined by

$$F_{SST} \equiv \frac{\Delta p}{\Delta p\,|_{SST}} - 1, \tag{13.51}$$

where Δp is the steady-state central surface pressure drop and $\Delta p\,|_{SST}$ is it value when the sea surface temperature (SST) is held constant. The feedback factor is zero when there is no ocean feedback and -1 if the feedback suffices to kill the storm. The coupled model was run 2083 times with different nondimensional parameter combinations, and run out to 18 days, sufficiently long that the intensity had always reached a steady state, and F_{SST} was calculated for each of these simulations. An excellent predictor of F_{SST} was found as a function of the five nondimensional parameter combinations. Redimensionalizing the predictors, this function is

$$F_{SST} = -0.87e^{-z}, \tag{13.52}$$

where

$$z \equiv 0.55\,\eta^{-0.85}\left(\frac{h_0}{30\ m}\right)^{1.04}\left(\frac{u_T}{6\,m\,s^{-1}}\right)^{0.97}\left(\frac{\Delta p\,|_{SST}}{50\,hPa}\right)^{-0.78}\times$$
$$\left(\frac{f}{5\times10^{-5}\ s^{-1}}\right)^{0.59}\left(\frac{\Gamma}{8\times10^{-2}\ K\ m^{-1}}\right)^{-0.40}\left(\frac{1-\mathcal{H}_a}{0.2}\right)^{0.46}. \qquad (13.53)$$

Here η is a size scaling factor by which all the radial dimensions of the tropical cyclone are multiplied by before they are passed to the ocean model; this allows the effect of storm size on ocean feedback to be modeled. The other factors are h_0, the initial ocean mixed layer depth, u_T, the specified translation speed of the cyclone, f, the Coriolis parameter, Γ, the thermal stratification below the mixed layer, and $\mathcal{H}_a$, the relative humidity in the ambient boundary layer. The denominators of the factors are the control run values of the parameters.

Larger values of z in (13.53) translate to smaller negative feedback on cyclone intensity. Examining the exponents, there will be less negative feedback for smaller and more rapidly moving storms, deeper ocean mixed layers, smaller sub-mixed layer stratification, smaller potential intensity, and larger values of the Coriolis parameter.

Figure 13.17 shows how F_{SST} varies with translation speed and initial mixed layer depth, holding all the other parameters constant. Feedback is relatively weak for small storms moving rapidly over relatively deep mixed layers. The magnitude of the feedback suggests that ocean interaction can be a quantitatively important influence on storm intensity

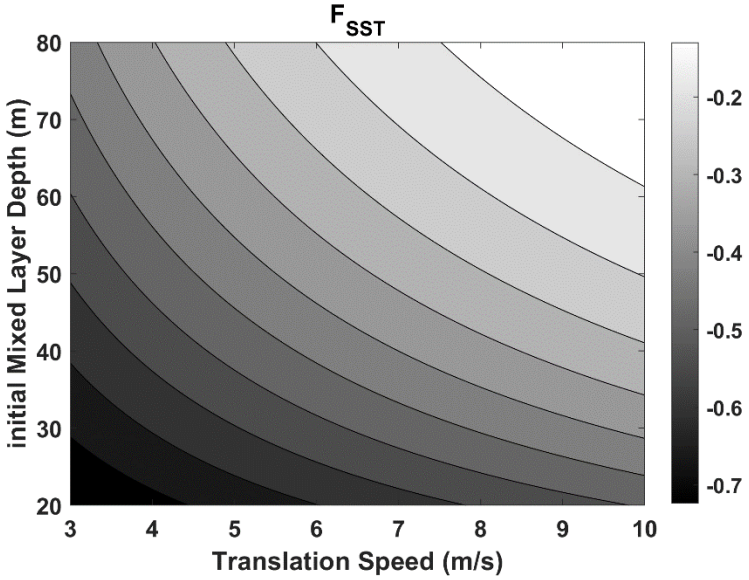


Figure 13.17: The ocean interaction feedback factor F_{SST} as a function of translation speed (ms^{-1}) and initial ocean mixed layer depth (m). All the other factors in (13.53) are set to unity.

Korty (2002) compared the results of running the fully coupled model of Schade (1994) with a version in which nonlinear advection, including vertical advection and horizontal pressure gradients are neglected, leaving mixing as the only operative process affecting sea surface temperature. In this version, each vertical column of the ocean behaves independently of the others, and are the conditions for the validity of our analytic solution, (13.46) – (13.49), except that the Coriolis terms are retained. Korty found that the results are nearly indistinguishable from the full physics simulations except when the cyclone is moving very slowly or doubles back over its own wake.

The coupled model and its simplification described here assume that oceanwater density is a function of temperature alone, but salinity can also be important in the density jump at the base of the mixed layer that defines the Richardson Number (13.35). Of particular interest is the circumstance in which mixed layer is relatively fresh compared to the waters just below. In this situation, the stable density jump associated with salinity reduces mixing and can allow the mixed layer to warm up seasonally more than it otherwise would, increase tropical cyclone potential intensity. Then, when the cyclone arrives, storm-induced mixing is also reduced, so the mixed layer does not cool as much or as rapidly (Ffield 2007). Some observational evidence suggests that the surface freshwater outflow from the Amazon and Orinoco Rivers damps ocean cooling for hurricanes like Irma of 2017 (Rudzin et al. 2020).

Forecasts of tropical cyclone intensity must account for ocean interaction, and the more intense the cyclone the more it matters (according to (13.53)). As an example, Figure 13.18 shows a "hindcast" of the intensity of Atlantic Hurricane Gert of 1999, using the coupled ocean-atmosphere model of Emanuel et al. (2004). In such a hindcast, the time-evolving storm environment and observed storm track and initial intensity are specified, and the model predicts the evolution of the cyclone's intensity. Figure 13.18 compares the observed maximum wind speed in Gert to that predicted by the coupled and uncoupled version of the models. (In the uncoupled version, the sea surface temperature field does not respond to the cyclone.)

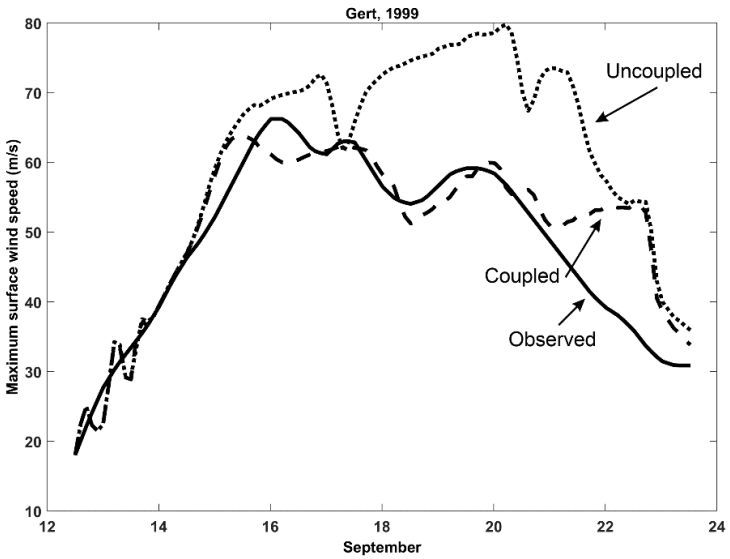


Figure 13.18: Evolution over time of the maximum surface wind (ms^{-1}) in Hurricane Gert of 1999 in observations (solid), coupled model (dashed) and uncoupled model (dotted), using the model described in Emanuel et al. (2004).

While the coupled model hindcast is by no means perfect, it is clearly far closer to what was observed than the hindcast using the uncoupled version.

One challenge in operational numerical forecasting of tropical cyclones is initializing the upper ocean. While surface temperature is well-measured today, real-time temperature and salinity profiles require in-situ observations, which may be sparse in some tropical cyclone-prone regions. Of particular concern are warm and cold mesoscale eddies on the ocean. In places like the Gulf of Mexico, such eddies can be nearly invisible in surface temperature fields, but strongly affect local ocean mixed layer depths. For this reason, surface cooling is suppressed in warm eddies and enhanced in cool eddies, even if there is no initial surface temperature manifestation (Jaimes and Shay 2009). There is a rich observational and numerical simulation-based literature on the effect of ocean eddies on tropical cyclone intensity (see the review by Zhang et al. 2021).

While initializing upper ocean conditions is challenging, greatly improved robotic in-situ sampling (Jayne et al. 2017) and operational analyses boosted by assimilating this and other in-situ and remote sensing observations are leading to much better initial conditions for coupled ocean-atmosphere forecast models.

13.5 Non-baroclinic re-intensification over land

In Chapter 10, section 10.4, we discussed the spin-down of tropical cyclone over land, assuming that the surface enthalpy flux becomes effectively zero. This is generally a good approximation, because while the uppermost few meters of a typical land surface can store as much or more heat than water, the conductivity of heat through rock and soils is generally far too small to provide significant enthalpy input to a cyclone. It is the low heat conductivity of land that prevents it from supporting surface-flux-driven cyclones.

Yet there are circumstances in which landfalling tropical cyclones are observed to re-intensify while well inland without any obvious baroclincity, even developing well-defined eyes in some cases. This kind of inland re-development seems to be most common over the hot deserts of northern Australia, and several case studies were presented by Emanuel et al. (2008), who christened such storms “Agukabams” from the aboriginal roots “agu”, meaning land, and “kabam”, meaning storm. For convenience, we will use that term here for cyclones that re-intensify over land without baroclinic influences. Figure 13.19 shows, from that paper, a microwave image of Tropical Cyclone Abigail a full four days after making landfall in northern Australia. It has a well-developed eye, highly unusual in cyclones so long after landfall.

We proposed that agukabams are fueled by unusually rapid transfer of heat from very hot, sandy soils that have been wetted by the outer rainbands of approaching cyclones. Soils like the sandy and massive earth soils of northern Australia can experience a three-fold increase thermal conductivity when wetted, enough to sustain a weak tropical cyclone. Soil temperatures can be as high as 33^{o}C down to depths of a half meter or more in the northern deserts of Australia in summer. Using a simple coupled atmosphere-soil model, we demonstrated that landfalling tropical cyclones could indeed undergo re-intensification when passing over sufficiently hot, sandy soils that had been saturated with water.

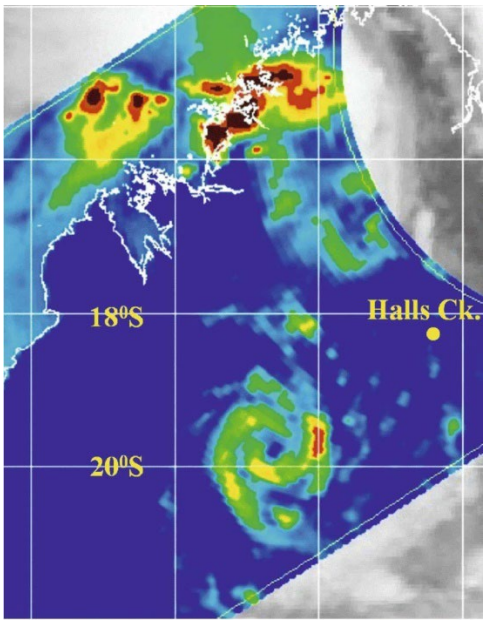


Figure 13.19: TRMM 85-GHz (horizontally polarized) microwave image 1903 UTC 2 March 2001 showing Tropical Cyclone Abigail in northwestern Australia, about four days after making landfall from the Gulf of Carpentaria.

Another interesting apparent case of an agukabam was Tropical Storm Erin, which made landfall in Texas and initially decayed over land (Figure 13.20). But quite suddenly, Erin re-intensified, with sustained winds reaching 25 ms^{-1} at 06 GMT on August 19th. It then decayed almost as rapidly. Figure 13.21 shows radar reflectivity at 10:08 GMT from the Oklahoma City NEXRAD radar. Erin has a well-developed eye at this time but had not shown a closed eye much before 04 GMT on the 19th. The National Hurricane Center best-track data shows peak winds of 25 knots (13 ms^{-1}) at 00 GMT and 50 knots (25 m s^{-1}) at 06 GMT.

By good fortune, Erin passed right through the Oklahoma Mesonet, a dense array of observing stations developed by University of Oklahoma and Oklahoma State University, and maintained by the Oklahoma Climatological Survey. The mesonet stations record standard meteorological variable and soil temperatures at various depths. Figure 13.22 shows the time series of soil temperature at Hinton, Oklahoma, at depths of 5, 10 and 30 cm. A strong diurnal cycle is evident in temperature at all three levels, until August 18th when the high clouds and outer rainbands of Erin arrive at Hinton. The amplitude of this cycle is much smaller at 30 cm, owing to the time it takes heat to diffuse downward through the soil. But as Erin's center approaches early on the 19th there is a precipitous drop in temperature at all three levels. In particular, the drop at 30 cm depth is much larger than temperature drops at that level associated with the normal diurnal cycle. We also notice from Figure 13.21 that rains extended well to the east and northeast, which is along Erin's forward path, so we may assume that soils were already quite wet by the time that Erin's core arrived.

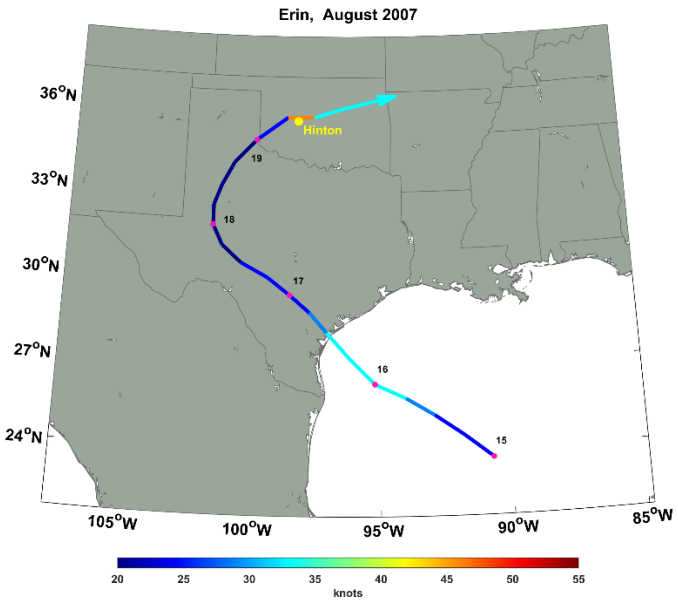


Figure 13.20: Track of Tropical Storm Erin of August, 2007. Red dots along track show the 00 GMT positions of the cyclone center labeled with dates of the month. The peak sustained wind speed is demarked by the track color according to the scale at bottom. The location of Hinton, Oklahoma is shown by the yellow dot.

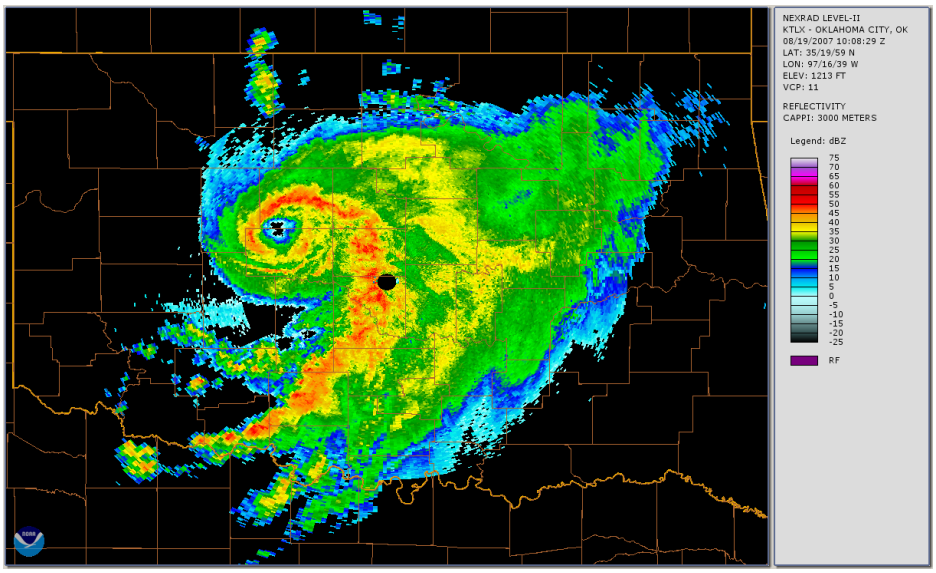


Figure 13.21: Radar reflectivity at 3 km altitude from the NOAA/NEXRAD radar at Oklahoma City (black dot) at 10:08 GMT on 19 August, 2007. The Oklahoma state boundary is visible near the periphery of the image. Reflectivity scale (dBZ) on right. Image produced by software available at https://www.ncdc.noaa.gov/wct/install.php .

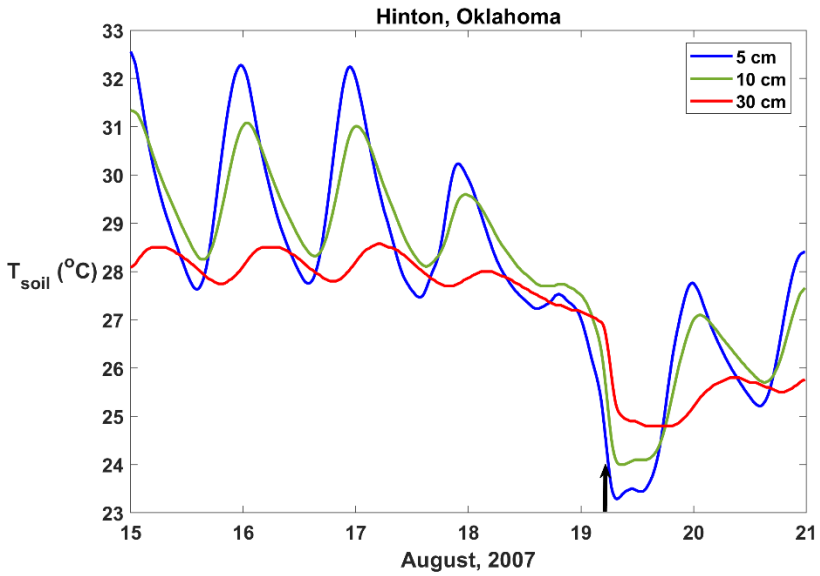


Figure 13.22: Time series of soil temperature at depths of 5, 10 and 30 cm at Hinton, Oklahoma. The black arrow at the bottom shows the approximate time of passage of the core of Tropical Storm Erin over Hinton.

Given these measurements, we can estimate the peak flow of heat from the soil surface to the atmosphere. The energy budget for the soil can be written

$$\int_0^H \rho_s c_s \frac{\partial T_s}{\partial t} dz = -F_s, \tag{13.54}$$

where ρ_s is the soil density, c_s is the soil heat capacity, F_s is the upward energy flux at the surface, and H is the depth over which a signal is evident. If, from Figure 13.22, we take an a temperature drop of 2.5 K over 40 cm of depth and, from tables of soil properties, use $\rho_s \simeq 1300\ Kg\,m^{-3}$, $c_s \simeq 1500\ J\,Kg^{-1}K^{-1}$, the surface energy flux amounts to about $180\ W\,m^{-2}$. From the Carnot cycle described in Chapter 9, equating surface frictional dissipation to the surface heat flux (including dissipative heating) multiplied by the Carnot efficiency, we have

$$\rho_a C_D V_s^3 = \frac{T_s - T_o}{T_o} F_s, \tag{13.55}$$

where ρ_a is the air density, C_D is the surface drag coefficient, V_s is the surface wind speed, T_s is the surface temperature, and T_o is the outflow temperature. If we take the inferred surface heat flux in Erin's core to be $180\ W\,m^{-2}$, the surface air density to be $1\ Kg\,m^{-3}$, the drag coefficient to be 2×10^{-3}, and the modified Carnot efficiency of the summertime mid-latitude atmosphere around Oklahoma to be about $1/3$, (13.55) yields a wind speed of about $30\ m\,s^{-1}$, not far from what was observed.

Yet the type of inland re-intensification exhibited by Erin, with no obvious baroclinic sources in the atmosphere, is rare. The key is to recognize that soils ahead of the cyclone must be both hot and wet. The heat conductivity of dry soil is not large enough to explain the drop in soil temperature seen in the measurements presented in Figure 13.22, so the soils must be quite

wet. The requirement that the soil must also be reasonably hot down to a depth of at least tens of centimeters suggests abundant sunshine for a period of many days in advance of the cyclone. The months leading up to Erin were exceptionally wet, and the period of March through July 2007 was, state-wide, the second wettest such period since records began in 1921 (Arndt et al. 2009). The days leading up to Erin were, by contrast, hot and sunny and Oklahoma Mesonet soil temperatures at a depth of 10 cm rose to as high as 40°C, allowing the uppermost soil to dry out (Arndt et al. 2009). This sequence of weather conditions may have been close to ideal for the re-intensification of Erin over Oklahoma, as the early rains of the storm re-moistened the upper soil, which was exceptionally hot. On the other hand, a modeling study by Evans et al. (2011) suggests that in this case, it was the antecedent rain over the preceding months that was important for Erin's re-intensification. Another key ingredient may be the absence of appreciable wind shear, because ventilation of the storm core with dry, low entropy air should have the same effect on inland cyclones as it does on tropical cyclones over the ocean.

Several studies support the idea that agukabams are powered by surface enthalpy fluxes under conditions in which the antecedent soils are hot and wet, and wind shear is light. In their study of monsoon depressions over India, Chang et al. (2009) demonstrated a positive correlation between cyclone intensity over land and antecedent soil temperature and moisture. They also showed that monsoon depressions maintained their intensity longer over land when there had been heavier rain during the preceding week. Andersen and Shepherd (2014) found that surface latent heat fluxes over land exceeded $70\,W\,m^{-2}$ in the 16 warm-core redeveloping systems they examined. Kishtawal et al. (2012) showed that the inland rate of decay tropical cyclones was slower for higher values of the thermal diffusivity of the underlying soils, suggesting that land surface heat flux may also have a positive effect on tropical cyclone intensity over land, even when it is not enough to cause re-intensification.

The effect of land surface enthalpy fluxes on post-landfall tropical cyclone behavior remains an interesting and active branch of research. Among several outstanding issues are whether large-scale land-use changes including agriculture and irrigation might affect the behavior of tropical cyclones over land, and whether climate change might also affect cyclone behavior by leading to changes in soil type, vegetations, temperature, and moisture.

References

Abdolali, A., A. Roland, A. van der Westhuysen, J. Meixner, A. Chawla, T. J. Hesser, J. M. Smith, and M. D. Sikiric, 2020: Large-scale hurricane modeling using domain decomposition parallelization and implicit scheme implemented in WAVEWATCH III wave model. *Coastal Engineering*, **157**, 103656, https://doi.org/10.1016/j.coastaleng.2020.103656.

Andersen, T. K., and J. M. Shepherd, 2014: A global spatiotemporal analysis of inland tropical cyclone maintenance or intensification. *International Journal of Climatology*, **34**, 391–402, https://doi.org/10.1002/joc.3693.

Andreas, E. L., 1989: *Thermal and size evolution of sea spray droplets*. US Army Cold Regions Research and Engineering Laboratory Hanover, NH, https://apps.dtic.mil/sti/pdfs/ADA210484.pdf.

——, 1992: Sea spray and the turbulent air-sea heat fluxes. *Journal of Geophysical Research: Oceans*, **97**, 11429–11441, https://doi.org/10.1029/92JC00876.

Andreas, E. L., 1995: The temperature of evaporating sea spray droplets. *Journal of Atmospheric Sciences*, **52**, 852–862, https://doi.org/10.1175/1520-0469(1995)052<0852:TTOESS>2.0.CO;2.

Andreas, E. L., and K. Emanuel, 2001: Effects of sea spray on tropical cyclone intensity. *J. Atmos. Sci.*, **58**, 3741–3751.

Arndt, D. S., J. B. Basara, R. A. McPherson, B. G. Illston, G. D. McManus, and D. B. Demko, 2009: Observations of the overland reintensification of Tropical Storm Erin (2007). *Bulletin of the American Meteorological Society*, **90**, 1079–1094, https://doi.org/10.1175/2009BAMS2644.1.

Bell, M. M., M. T. Montgomery, and K. Emanuel, 2012: Air-sea enthalpy and momentum exchange at major hurricane wind speeds observed during CBLAST. *J. Atmos. Sci.*, **69**, 3197–3222.

Black, P. G., and W. L. Adams, 1983: *Guidance for estimating surface winds based on sea state observations from aircraft and sea state catalog .* U.S. Dept. Commerce/NOAA,.

Bye, J. A. T., and A. D. Jenkins, 2006: Drag coefficient reduction at very high wind speeds. *Journal of Geophysical Research: Oceans*, **111**, https://doi.org/10.1029/2005JC003114.

Chang, H.-I., D. Niyogi, A. Kumar, C. M. Kishtawal, J. Dudhia, F. Chen, U. C. Mohanty, and M. Shepherd, 2009: Possible relation between land surface feedback and the post-landfall structure of monsoon depressions. *Geophysical Research Letters*, **36**, https://doi.org/10.1029/2009GL037781.

Charnock, H., 1955: Wind stress on a water surface. *Quarterly Journal of the Royal Meteorological Society*, **81**, 639–640, https://doi.org/10.1002/qj.49708135027.

Dietrich, J. C., and Coauthors, 2012: Performance of the unstructured-mesh, SWAN+ADCIRC model in computing hurricane waves and surge. *Journal of Scientific Computing*, **52**, 468–497, https://doi.org/10.1007/s10915-011-9555-6.

Drennan, W. M., H. C. Graber, D. Hauser, and C. Quentin, 2003: On the wave age dependence of wind stress over pure wind seas. *Journal of Geophysical Research: Oceans*, **108**, https://doi.org/10.1029/2000JC000715.

Emanuel, K., C. DesAutels, C. Holloway, and R. Korty, 2004: Environmental control of tropical cyclone intensity. *J. Atmos. Sci.*, **61**, 843–858.

Emanuel, K., J. Callaghan, and P. Otto, 2008: A Hypothesis for the Redevelopment of Warm-Core Cyclones over Northern Australia. *Monthly Weather Review*, **136**, 3863–3872, https://doi.org/10.1175/2008mwr2409.1.

Evans, C., R. S. Schumacher, and T. J. Galarneau, 2011: Sensitivity in the overland reintensification of Tropical Cyclone Erin (2007) to near-surface soil moisture characteristics. *Monthly Weather Review*, **139**, 3848–3870, https://doi.org/10.1175/2011MWR3593.1.

Fairall, C. W., J. D. Kepert, and G. J. Holland, 1994: The effect of sea spray on surface energy transports over the ocean. *Global Atmos. Ocean Sys.*, **2**, 121–142.

Ffield, A., 2007: Amazon and Orinoco River Plumes and NBC Rings: Bystanders or Participants in Hurricane Events? *Journal of Climate*, **20**, 316–333, https://doi.org/10.1175/JCLI3985.1.

Hasselmann, K., 1962: On the non-linear energy transfer in a gravity-wave spectrum Part 1. General theory. *Journal of Fluid Mechanics*, **12**, 481–500, https://doi.org/10.1017/S0022112062000373.

Hasselmann, K., and Coauthors, 1973: Measurements of wind-wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP). *Ergaenzungsheft zur Deutschen Hydrographischen Zeitschrift, Reihe A*,.

Hawkins, H. F., and D. T. Rubsam, 1968: Hurricane Hilda, 1964: II. Structure and budgets of the hurricane on October 1, 1964. *Mon. Wea. Rev.*, **96**, 617–636.

Iida, N., Y. Toba, and M. Chaen, 1992: A new expression for the production rate of sea water droplets on the sea surface. *Journal of Oceanography*, **48**, 439–460, https://doi.org/10.1007/BF02234020.

Jaimes, B., and L. K. Shay, 2009: Mixed layer cooling in mesoscale oceanic eddies during Hurricanes Katrina and Rita. *Monthly Weather Review*, **137**, 4188–4207, https://doi.org/10.1175/2009MWR2849.1.

Jayne, S. R., D. Roemmich, N. Zilberman, S. C. Riser, K. S. Johnson, G. C. Johnson, and S. R. Piotrowicz, 2017: The Argo Program. *Oceanography*, **30**, 18–28.

Kishtawal, C. M., D. Niyogi, A. Kumar, M. L. Bozeman, and O. Kellner, 2012: Sensitivity of inland decay of North Atlantic tropical cyclones to soil parameters. *Natural Hazards*, **63**, 1527–1542, https://doi.org/10.1007/s11069-011-0015-2.

Korty, R. L., 2002: Processes affecting the ocean's feedback on the intensity of a hurricane. 25th Conference on Hurricane and Tropical Meteorology, 14D.2, Amer. Meteor. Soc.

Kudryavtsev, V., V. Makin, and S. Zilitinkevich, 2012: *On the sea-surface drag and heat/mass transfer at strong winds*. KNMI,.

Kudryavtsev, V. N., 2006: On the effect of sea drops on the atmospheric boundary layer. *Journal of Geophysical Research: Oceans*, **111**, https://doi.org/10.1029/2005JC002970.

Kudryavtsev, V. N., and V. K. Makin, 2011: Impact of ocean spray on the dynamics of the marine atmospheric boundary layer. *Boundary-Layer Meteorology*, **140**, 383–410, https://doi.org/10.1007/s10546-011-9624-2.

Luettich, R. A., J. J. Westerink, N. W. Scheffner, and others, 1992: ADCIRC: an advanced three-dimensional circulation model for shelves, coasts, and estuaries. Report 1, Theory and methodology of ADCIRC-2DD1 and ADCIRC-3DL.

Ma, H., A. V. Babanin, and F. Qiao, 2020: Field observations of sea spray under Tropical Cyclone Olwyn. *Ocean Dynamics*, **70**, 1439–1448, https://doi.org/10.1007/s10236-020-01408-x.

Makin, V. K., 2005: A note on the drag of the sea surface at hurricane winds. *Boundary-Layer Meteorology*, **115**, 169–176, https://doi.org/10.1007/s10546-004-3647-x.

Makin, V. K., and V. N. Kudryavtsev, 2002: Impact of dominant waves on sea drag. *Boundary-Layer Meteorology*, **103**, 83–99, https://doi.org/10.1023/A:1014591222717.

Moon, I.-J., I. Ginis, T. Hara, H. L. Tolman, C. W. Wright, and E. J. Walsh, 2003: Numerical simulation of sea surface directional wave spectra under hurricane wind forcing. *Journal of Physical Oceanography*, **33**, 1680–1706, https://doi.org/10.1175/1520-0485(2003)033<1680:NSOSSD>2.0.CO;2.

Nystrom, R. G., S. J. Greybush, X. Chen, and F. Zhang, 2021: Potential for new constraints on tropical cyclone surface-exchange coefficients through simultaneous ensemble-based state and parameter estimation. *Monthly Weather Review*, **149**, 2213–2230, https://doi.org/10.1175/MWR-D-20-0259.1.

Palmén, E., and H. Riehl, 1957: Budget of angular momentum and energy in tropical cyclones. *Journal of Atmospheric Sciences*, **14**, 150–159, https://doi.org/10.1175/1520-0469(1957)014<0150:BOAMAE>2.0.CO;2.

Phillips, O. M., 1966: *The Dynamics of the Upper Ocean*. Cambridge University Press, 261 pp.

Porchetta, S., O. Temel, D. Muñoz-Esparza, J. Reuder, J. Monbaliu, J. van Beeck, and N. van Lipzig, 2019: A new roughness length parameterization accounting for wind–wave (mis)alignment. *Atmospheric Chemistry and Physics*, **19**, 6681–6700, https://doi.org/10.5194/acp-19-6681-2019.

Powell, M. D., P. J. Vickery, and T. A. Reinhold, 2003: Reduced drag coefficients for high wind speeds in tropical cyclones. *Nature*, **422**, 279–283.

Price, J. F., 1979: On the scaling of stress-driven entrainment experiments. *Journal of Fluid Mechanics*, **90**, 509–529, https://doi.org/10.1017/S0022112079002366.

——, 1981: Upper Ocean Response to a Hurricane. *Journal of Physical Oceanography*, **11**, 153–175, https://doi.org/10.1175/1520-0485(1981)011<0153:UORTAH>2.0.CO;2.

Richter, D. H., and P. P. Sullivan, 2013: Sea surface drag and the role of spray. *Geophysical Research Letters*, **40**, 656–660, https://doi.org/10.1002/grl.50163.

Riehl, H., and J. S. Malkus, 1961: Some aspects of Hurricane Daisy, 1958. *Tellus*, **13**, 181–213.

Rudzin, J. E., S. Chen, E. R. Sanabia, and S. R. Jayne, 2020: The air-sea response during Hurricane Irma's (2017) rapid intensification over the Amazon-Orinoco river plume as

measured by atmospheric and oceanic observations. *Journal of Geophysical Research: Atmospheres*, **125**, e2019JD032368, https://doi.org/10.1029/2019JD032368.

Schade, L. R., 1994: The ocean's effect on hurricane intensity. Mass. Inst. Tech., 127 pp. https://dspace.mit.edu/bitstream/handle/1721.1/11640/32142088-MIT.pdf?sequence=2&isAllowed=y.

——, and K. A. Emanuel, 1999: The ocean's effect on the intensity of tropical cyclones: Results from a simple coupled atmosphere- ocean model. *J. Atmos. Sci.*, **56**, 642–651.

Simpson, R. H., and J. S. Malkus, 1964: Experiments in hurricane modification. *Scientific American*, **211**, 27–37.

——, and J. Simpson, 1966: Why experiment on tropical hurricanes? Vol. 28 of, Trans. New York Acad. Sci., 1045–1062.

Sriver, R. L., 2013: Observational evidence supports the role of tropical cyclones in regulating climate. *Proceedings of the National Academy of Sciences*, **110**, 15173–15174, https://doi.org/10.1073/pnas.1314721110.

Sroka, S., and K. Emanuel, 2021: A review of parameterizations for enthalpy and momentum fluxes from sea spray in tropical cyclones. *Journal of Physical Oceanography*, **51**, 3053–3069, https://doi.org/10.1175/JPO-D-21-0023.1.

Sterl, A., 2017: Drag at high wind velocities-a review. *Roy. Netherland Met. Office. Tech. Rep*, **361**, 23.

Troitskaya, Yu., A. Kandaurov, O. Ermakova, D. Kozlov, D. Sergeev, and S. Zilitinkevich, 2018: The "bag breakup" spume droplet generation mechanism at high winds. Part I: Spray generation function. *Journal of Physical Oceanography*, **48**, 2167–2188, https://doi.org/10.1175/JPO-D-17-0104.1.

Vanderplow, B., A. V. Soloviev, C. W. Dean, B. K. Haus, R. Lukas, M. Sami, and I. Ginis, 2020: Potential effect of bio-surfactants on sea spray generation in tropical cyclone conditions. *Scientific Reports*, **10**, 19057, https://doi.org/10.1038/s41598-020-76226-8.

Xu, X., J. J. Voermans, H. Ma, C. Guan, and A. V. Babanin, 2021: A wind–wave-dependent sea spray volume flux model based on field experiments. *Journal of Marine Science and Engineering*, **9**, https://doi.org/10.3390/jmse9111168.

Zhang, H., H. He, W.-Z. Zhang, and D. Tian, 2021: Upper ocean response to tropical cyclones: A review. *Geoscience Letters*, **8**, 1, https://doi.org/10.1186/s40562-020-00170-8.

Zhao, D., Y. Toba, K. Sugioka, and S. Komori, 2006: New sea spray generation function for spume droplets. *Journal of Geophysical Research: Oceans*, **111**, https://doi.org/10.1029/2005JC002960.

14. The Genesis Problem

No aspect of tropical cyclones has proved more challenging than their genesis. After decades of research, there is little consensus on how tropical cyclone form, and our ability to forecast whether a particular cloud cluster or easterly wave will develop remains rudimentary. The one statistic we can forecast nearly perfectly is the global, annual number of storms which, as demonstrated in Figure 14.1, is remarkably stable at about 90. The interannual standard deviation is 9.5, very close to the square root of the mean as would be expected in pure Poisson variability. There is no significant long-term trend.

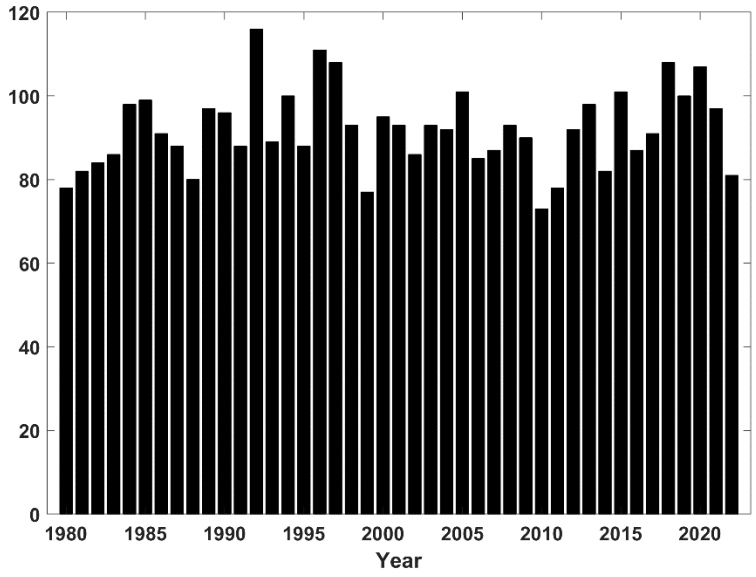


Figure 14.1: Annual, global number of tropical cyclones in the IBTrACS data set (Knapp et al. 2010), from 1980 through 2022. Only cyclones whose lifetime maximum surface winds reach at least 35 knots (18 ms^{-1}) are included. The long-term mean is 92 with a standard deviation of 9.5. Global tropical cyclone counts prior to about 1980 are compromised by poor or nonexistent satellite coverage.

At the time of this writing, there is no generally accepted theory for why there are O(100) tropical cyclones on our planet each year or why that number is so stable in spite of large inter-basin and seasonal variability.

Historically, there have been two threads of research on genesis. The first, which we here label the “local view”, focuses on how individual tropical cyclones come about, using observations and models. Early in the post-WWII renaissance of tropical cyclone research, it was recognized that most if not all tropical cyclones originate from independent disturbances such as African easterly waves (Dunn 1940; Riehl 1948a,b; Dunn 1951; Riehl 1951; Ramage 1959; Riehl 1975). Much research effort has been expended on whether and how such independent disturbances transform into tropical cyclones, partly to support efforts to forecast genesis in real time.

The second thread, begun by Palmén (1948) and expanded on by Gray (1975, 1979), focuses on susceptibility of the large-scale thermodynamic and kinematic environment to genesis. We label this thread the “global view”.

In this chapter, we will look at genesis through both of these lenses and attempt a synthesis that may help explain both how tropical cyclones develop and what controls their overall climatology.

14.1 The local view

We begin by looking at a strongly developing and a non-developing tropical Atlantic wave disturbance and see how they fit in with previous observational and field experimental inferences. We then proceed to a theoretical interpretation of these observations and conclude this section with a synthesis of the local view.

14.1.1 Case study: The development of Hurricane Ivan of 2004

Here we examine a particular case of tropical cyclogenesis that took place in the North Atlantic, leading to Hurricane Ivan of 2004. Figure 14.2 shows the track and maximum wind speeds in Hurricane Ivan.

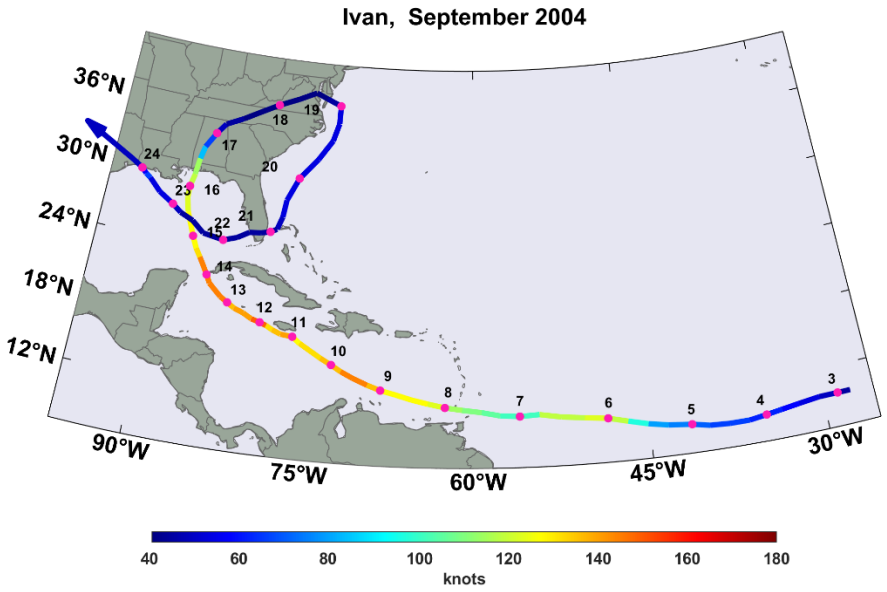


Figure 14.2: Track of the center of Hurricane Ivan of 2004. Numbers correspond to 00 GMT dates in September. Colors signify maximum wind speeds (kts).

Ivan followed a classical evolution into a "Cape Verde" hurricane from an easterly wave that moved off the west coast of Africa on August 31st. Its track after landfall in Alabama and the Florida panhandle was, however, unusual. We will focus on its development from September 1st to September 5th. Figure 14.3 shows pre-Ivan's evolution in 4 snapshots during this time period.

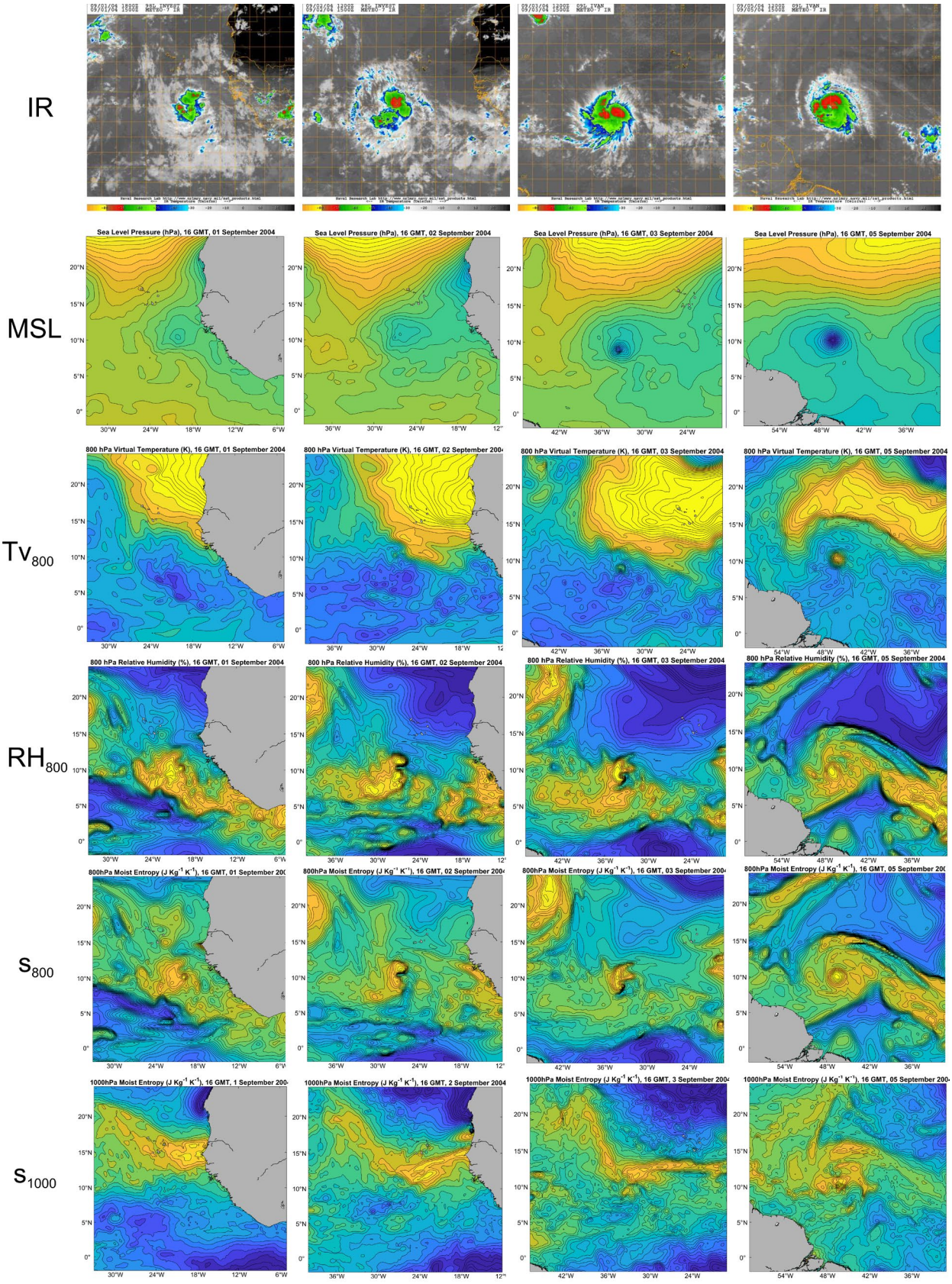

IR
MSL
Tv800
RH800
s800
s1000
Sea Level Pressure (hPa), 16 GMT, 01 September 2004
Sea Level Pressure (hPa), 16 GMT, 02 September 2004
Sea Level Pressure (hPa), 16 GMT, 03 September 2004
Sea Level Pressure (hPa), 16 GMT, 05 September 2004
800 hPa Virtual Temperature (K), 16 GMT, 01 September 2004
800 hPa Virtual Temperature (K), 16 GMT, 02 September 2004
800 hPa Virtual Temperature (K), 16 GMT, 03 September 2004
800 hPa Virtual Temperature (K), 16 GMT, 05 September 2004
800 hPa Relative Humidity (%), 16 GMT, 01 September 2004
800 hPa Relative Humidity (%), 16 GMT, 02 September 2004
800 hPa Relative Humidity (%), 16 GMT, 03 September 2004
800 hPa Relative Humidity (%), 16 GMT, 05 September 2004
800hPa Moist Entropy (J Kg-1 K-1), 16 GMT, 01 September 200
800hPa Moist Entropy (J Kg-1 K-1), 16 GMT, 02 September 2004
800hPa Moist Entropy (J Kg-1 K-1), 16 GMT, 03 September 2004
800hPa Moist Entropy (J Kg-1 K-1), 16 GMT, 05 September 200
1000hPa Moist Entropy (J Kg-1 K-1), 16 GMT, 1 September 200
1000hPa Moist Entropy (J Kg-1 K-1), 16 GMT, 2 September 2004
1000hPa Moist Entropy (J Kg-1 K-1), 16 GMT, 3 September 200
1000hPa Moist Entropy (J Kg-1 K-1), 16 GMT, 05 September 200
September 1
September 2
September 3
September 5

Figure 14.3: Evolution of Ivan from September 1st to September 5th, 2004 at 16 GMT, skipping September 4th for space reasons. The top row shows storm-centered infrared imagery from the Geostationary Operational Environmental Satellite (GOES), while all the other rows display fields from ERA5 hourly reanalyses. From row 2 to row 6: Mean sea-level pressure (from 1002 to 1018 hPa), 800 hPa virtual temperature (from 287 to 293 K), 800 hPa relative humidity (from 20 to 100 %), 800 hPa moist entropy (from 5780 to 5870 J Kg^{-1} K^{-1}) and 1000 hPa moist entropy (from 5840 to 5890 J Kg^{-1} K^{-1}).

The top row shows infrared imagery from a geostationary satellite while the other rows show fields from the European center for Medium Range Forecasts Re-Analysis 5 (ERA5; Hersbach et al. 2020). The reanalysis is based on running a global, high-resolution weather forecast model while assimilating most available observations, including from satellites, aircraft, rawinsondes, and surface-based platforms. The imagery and the ERA5 fields are aligned in space and time and follow the cyclone from a tropical wave on September 1st to a tropical storm on September 5th.

In interpreting these fields, one must bear in mind that re-analyses are not observations. There are very few in-situ observations in the region covered, and one must rely on the model itself and its ability to assimilate what observations there are.

A loosely organized, roughly circular cloud cluster moves away from the African coast on September 1st. A broad, weak area of low pressure accompanies the cluster, and a stronger cyclone has developed within it by September 3rd. At 800 hPa, the core of this disturbance is cold and humid (third and fourth rows of Figure 14.3) and while it has elevated moist entropy at 800 hPa (fifth row), there is a depression in the boundary layer (1000 hPa) moist entropy for the first two days of the evolution shown. Not until September 3rd does a high entropy core begin to develop in the boundary layer. The lower tropospheric part of the disturbance remains cold but by September 5th a very small, intense warm core has developed, visible in the 800 hPa virtual temperature field. (The disturbance has a warm core in the upper troposphere (not shown) through the whole period.)

Note that the high moist entropy core at 1000 hPa evident on September 5th interacts with a thin strip of high moist entropy moving off the central African coastline. This strip is muted at 800 hPa but by September 5th a broad region of elevated entropy surrounds the tight, high-entropy core of Tropical Storm Ivan.

Another feature of possible relevance is the large plume of hot but dry air emanating from the Sahara Desert, as visible in the 800 hPa virtual temperature and relative humidity. The moist entropy of this air is not very different from that of the surrounding air mass. By September 5th it has started to wrap around the western periphery of Ivan. Otherwise, the fields delineating Ivan's core and immediate surrounds are beginning to resemble those of a mature tropical cyclone by this time.

14.1.2: Case study: The non-development of Tropical Storm Gaston of 2010

The disturbance that eventually became Gaston, as weak tropical storm, followed an initial trajectory similar to that of Ivan of 2004, as shown in Figure 14.4.

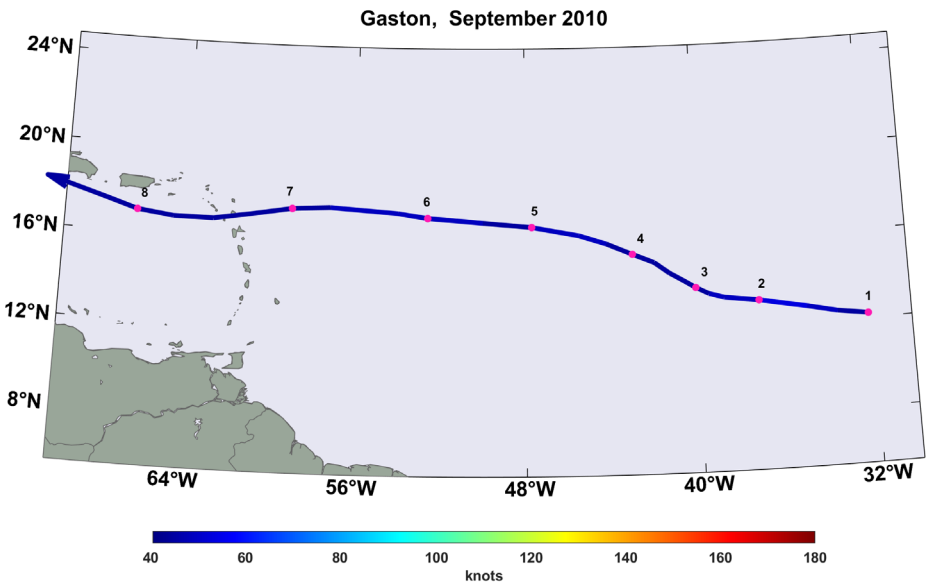


Figure 14.4: Track of Tropical Storm Gaston of 2010. As in Figure 14.2, the red dots denote the 00 GMT positions of the storm center at the September dates indicated by the numerals.

According to NOAA/National Hurricane Center records, Gaston's maximum winds peak at 35 knots (18 ms^{-1}) on September 2nd and 3rd and remained at or below 30 knots (15 ms^{-1}) thereafter.

Figure 14.5 shows the infrared presentation of Gaston, together with the virtual temperature and moist entropy at 800 hPa at three times. (Note the smaller scale of the maps in Figure 14.5 compared to Figure 14.3.) At 06 GMT on August 31st, an impressive cloud cluster is visible in the IR satellite image, but there is little indication of any disturbance at 800 hPa in either the ERA5 virtual temperature or moist entropy fields. (There is, however, a weak depression in sea level pressure, not shown here; co-located with a positive virtual temperature anomaly at 300 hPa.) The absence of a significant perturbation in the low-level thermodynamic fields suggests either that the cloud cluster does not yet have significant rotation or that the re-analysis has missed these features. Note the very hot Saharan air mass with a sharp southern boundary.

Several days later, at 00 GMT on September 3rd, the cloud cluster has almost disappeared from the infrared satellite image, yet by this time there are significant perturbations in the virtual temperature and moist entropy fields at 800 hPa. (It should be noted, though, that this is at about the local time of a diurnal minimum in moist convection over tropical oceans.) A negative virtual temperature anomaly has developed, apparently by virtue of Gaston's circulation penetrating the very hot Saharan airmass to its north, but a prominent positive moist entropy anomaly suggests that surface fluxes may be pumping high entropy into the lower troposphere. Also of interest is the band of high most entropy associated with the ITCZ to the south of Gaston. Sixty hours later, on September 5th, the cloud cluster appears to be re-intensifying (though note the opposite phase of the diurnal cycle from the previous image) and there are prominent positive anomalies in both virtual temperature and moist entropy at 800 hPa, more or less collocated with the cold cloud tops.

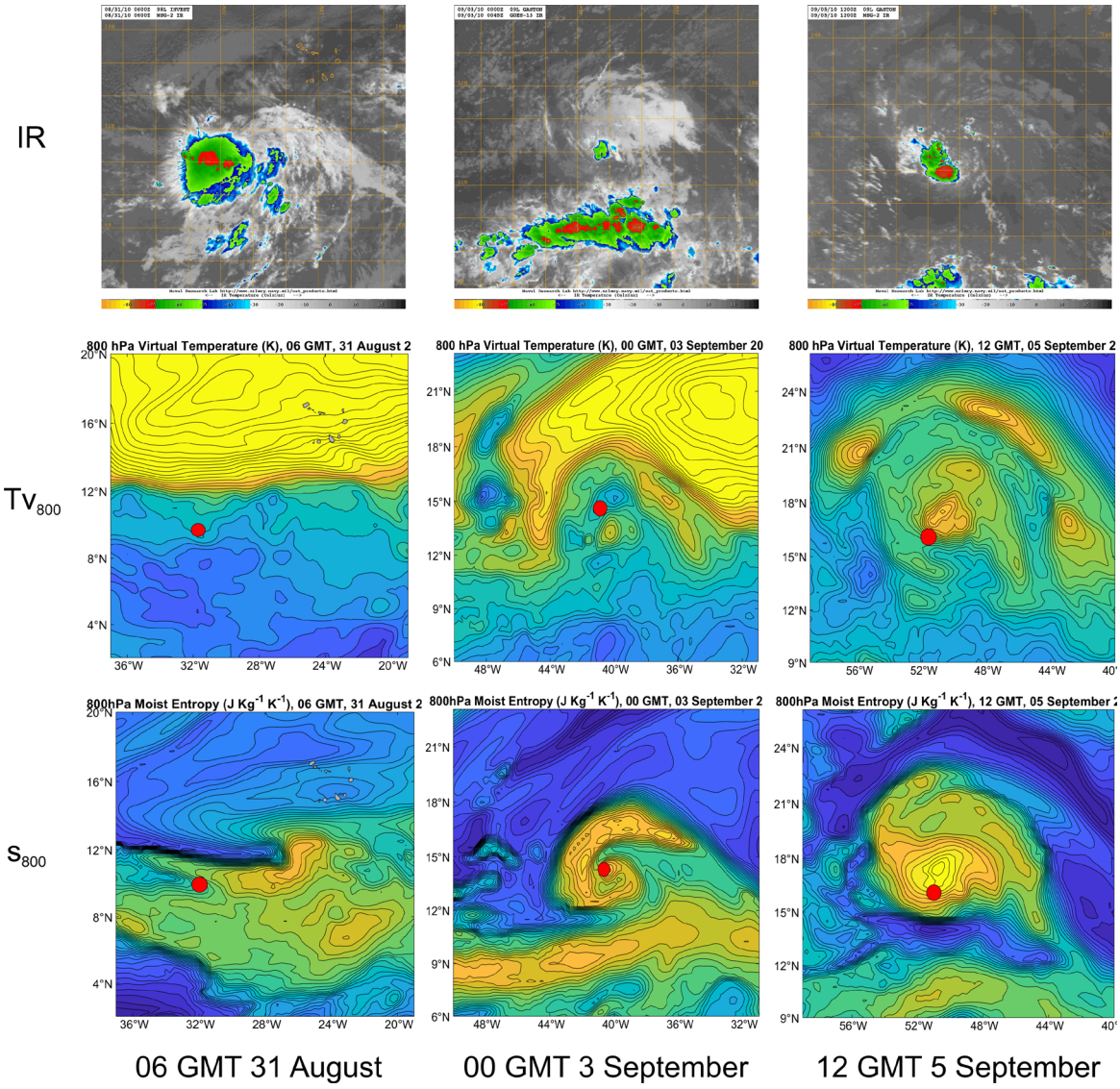


Figure 14.5: Evolution of Gaston from 06 GMT August 31st to 12 GMT September 5th, 2010. The top row shows storm-centered infrared imagery from the Geostationary Operational Environmental Satellite (GOES) while all the other rows display fields from ERA5 hourly reanalyses. Row 2: 800 hPa virtual temperature (from 287 to 293 K); Row 3: 800 hPa moist entropy (from 5790 to 5880 J Kg^{-1} K^{-1}). The red dot in the last two rows shows the approximate position of the IR temperature minimum (top row).

Comparing to Figure 14.3 for Ivan, and accounting for the different map scales (though the color scales span the same ranges), the low-level thermodynamic fields would seem to indicate a robustly intensifying system. But note that the warm core is not as concentrated as in the case of Ivan when it was of comparable age and strength. Moreover, according to the NOAA data, Gaston is weaker at this time than on September 3rd and never regained tropical storm intensity. What made the difference?

A clue to this enigma might be present in the relative humidity fields higher up, as shown in Figure 14.6, which compares the evolution of Hurricane Ivan of 2004 (top row) to that of Tropical Storm Gaston of 2010 (bottom row) in terms of the relative humidity at 600 hPa.

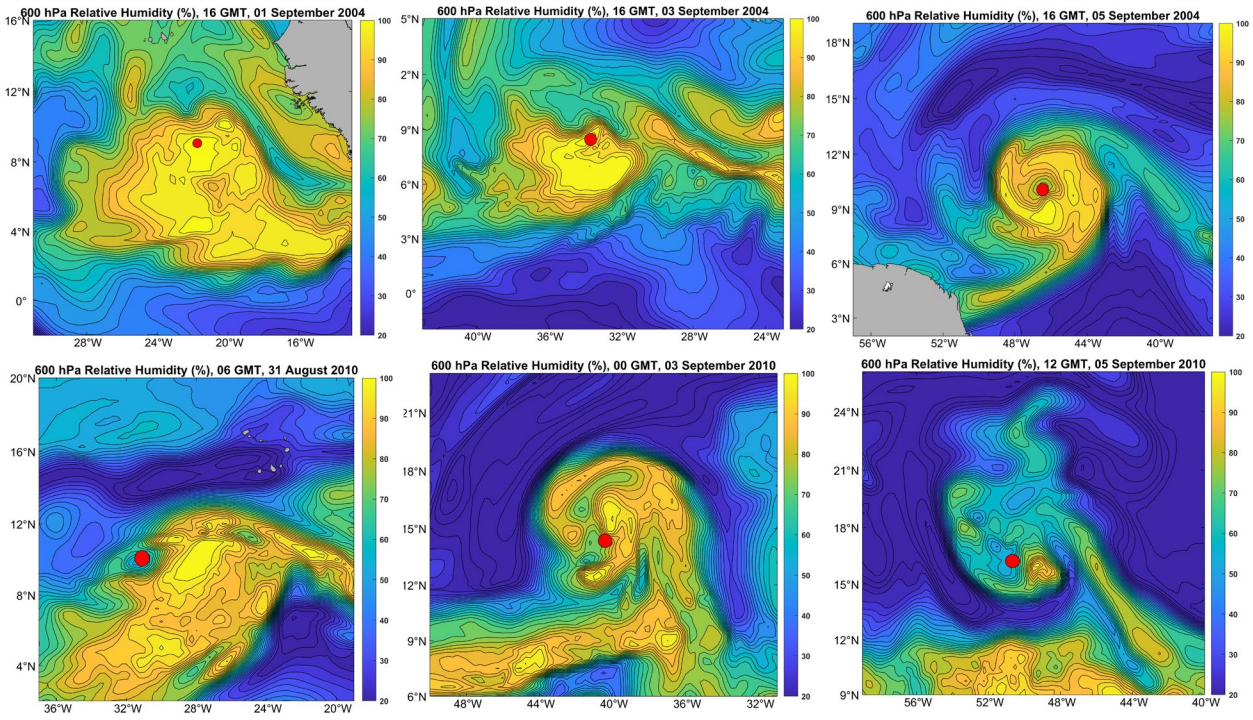


Figure 14.6: ERA5 relative humidity at 600 hPa during the development of Ivan of 2004 (top row) and Gaston of 2010 (bottom row). The Ivan sequence is for 16 GMT on September 1st, 3rd, and 5th, 2004 and the Gaston sequence is for 06 GMT on August 31st, 00 GMT on September 3rd and 12 GMT on September 5th, 2010. The red dot on each panel shows the approximate position of the coldest infrared cloud tops in each case. Note that an effort has been made to keep all of the maps at about the same scale, and the relative humidity color scale (20-100%) is identical across the maps.

Note that in Ivan's case, the 600 hPa relative humidity anomaly is closer to saturation and/or more closely aligned with the convection (as indicated by the cloud top temperature minimum), and in the last two columns, more coherent than was the case with Gaston. And while Gaston's 800 hPa moist entropy anomaly amplified between the second and third time periods (Figure 14.5), the 600 hPa relative humidity clearly dropped, along with the intensity.

A closer inspection of the bottom row of Figure 14.6 suggests that by September 3rd, dry air was beginning to wrap around the west side of Gaston, and by September 5th it had begun to wrap into the core of the system.

Now of course it is possible that the drying out of Gaston's core was an effect of its weakening rather than the cause, if all we have to go on is the observation of its drying out. Later in this chapter we will argue, on the grounds of theory and models, that the core drying is indeed key to understanding Gaston's demise and probably the non-development of most tropical disturbances.

14.1.3 Ivan and Gaston in the context of other observational studies and field programs

Hurricane Ivan of 2004 began life as a weak cyclonic disturbance with a distinct cold core in the lower troposphere, quite the opposite of what we expect and observe in mature tropical cyclones. This appears to be a common route to genesis. By the late 1940s, Riehl (1948a) had already noticed that most tropical cloud clusters have cold cores in the lower troposphere. We know that African easterly waves have maximum amplitude in the middle troposphere, around

600 hPa (see Chapter 7, section 7.3), and quasi-balance considerations dictate that the troposphere beneath this level must be anomalously cold. Analysis of aircraft, Doppler, and dropsonde data deployed into eastern North Pacific Hurricane Guillermo of 1991 showed a prominent lower tropospheric cold core with anticyclonic surface circulation, within which a smaller scale (relative) warm core cyclone developed and eventually became a hurricane (Bister and Emanuel 1997). The development of a small warm core within a broader cold core was quite similar to the development of Hurricane Ivan of 2004 between September 3rd and September 5th (see row 5 of Figure 14.3).

Raymond et al. (2011) analyzed data collected from seven aircraft missions into northwest Pacific tropical cyclone during the TPARC/TCS-08 project (THORPEX Pacific Asian Regional Campaign/Tropical Cyclone Structure experiment) in 2008. They concluded that “The formation of a strong midlevel circulation, with its associated cold core at low levels and warm core aloft, greatly aids the spin-up of a tropical cyclone by changing the vertical mass flux profile of deep convection from top heavy to bottom heavy”. We shall return to their point about the vertical profile of convective mass flux in the next subsection.

Another interesting route to tropical cyclogenesis is through mesoscale convective complexes that form over land and the drift out over warm water. Such land-based convective complexes commonly have mid-level meso-cyclones that are cold core at low levels (Bartels and Maddox 1991). A particularly interesting case of a strong mesoscale convective complex that first produced flash flooding in Pennsylvania and then formed a tropical cyclone once out over the Gulf Stream was documented by Bosart and Sanders (1981).

These observations, taken together with the route to the genesis of polar lows and medicanes described in Chapter 12, section 12.5, suggest the cold-core cyclones are suitable embryos within which to develop tropical cyclones. We will return to this subject in subsection 14.1.5.

Genesis is not always preceded by disturbances that have maximum amplitude well above the boundary layer. In the case of Gaston above, there is little evidence of a cold core in the lower troposphere, at least in the reanalysis data. (In Figure 14.5, there is evidence of a cold anomaly at 800 hPa on September 3rd, but in this case the anomaly seems to have resulted from Gaston’s circulating air mass invading the very hot Saharan air to its north.) Moreover, a careful analysis of a case of genesis in the western North Pacific by Montgomery and Smith (2012), using data from the Tropical Cyclone Structure (2008) (TCS08), showed little evidence of the an initial cold core in the lower troposphere. But, as described by Raymond et al. (2011), a mid-level vortex did develop as the system that became Typhoon Nuri reached the tropical depression stage.

Dunkerton et al.(2008) proposed that tropical cyclogenesis that is associated with tropical waves occurs within a ‘pouch’ surrounding a ‘sweet spot’ defined by the intersection of the wave axis with a critical latitude at which the phase speed of the wave equals the westward flow speed in the lower troposphere. This idea became known as the marsupial paradigm and was tested using observations made within Atlantic tropical disturbances during the Pre-Depression Investigation of Cloud Systems in the Tropics (PREDICT) experiment in 2010 (Montgomery et al. 2012; Wang et al. 2012). The basic idea is that this configuration of large-scale flow protects the inner core of nascent tropical cyclones from ventilation with low-entropy air in the lower to middle troposphere.

Early researchers, relying on sparse observations from ships and on island-based radiosondes, noted the steady progression of African easterly waves across the North Atlantic (Chapter 7, section 7.3). Dunn (1940) remarked that most of these waves crossed the North Atlantic with no development; Simpson et al. (1968) estimated that only about 10% of these waves developed during hurricane season. For some eighty years, we have observed that potential triggering disturbances are inefficient at generating tropical cyclones. At the same time, early researchers concluded that all tropical cyclones originate in pre-exiting disturbances of independent physical origin (e.g. Riehl 1948a,b; Dunn 1951; Riehl 1951; Ramage 1959; Riehl 1975) and this belief persists to this day (e.g. Vecchi et al. 2019; Sugi et al. 2020; Hsieh et al. 2020). Over the tropical Atlantic, it is clear that many tropical cyclones develop from easterly waves whose origin over Africa leaves little doubt that the formative mechanism is distinct from that of cloud clusters that form over the ocean. Historically, most observational studies have focused on the Atlantic owing to the comparatively rich observational data stream there. In the western Pacific region, in contrast, disturbances that lead to genesis originate mostly over the ocean (except in the relatively rare cases of storms that form over the eastern North Pacific or even the Atlantic and remain intact as they traverse the western Pacific). It is not entirely clear whether some of the western Pacific cyclones may form spontaneously, as happens in idealized models (section 14.2.2).

Because most of the Atlantic waves traverse the basin with little or no development, early researchers sought a third factor, beyond the presence of a tropical wave and a favorable thermodynamic and kinematic environment. Riehl (1948a,b) proposed that the third ingredient is a disturbance in the upper troposphere. Indeed, as we saw in Chapter 12, section 12.3, the quasi-balance superposition principle supports the idea that the superposition of a cyclonic disturbance in the upper troposphere with one at lower levels can increase the magnitude of surface winds, among other measures of intensity. The same physics as described in section 12.3 may operate during genesis, the only difference being the magnitude of the low-level cyclonic disturbance.

Riehl's ideas receive some support in observational studies. For example, Sadler (1967, 1976a,b) showed that transitions from tropical waves to tropical cyclones in the western North Pacific often occur when the waves approach tropical upper-tropospheric troughs (TUTTs). Hawkins and Rubsam (1968) showed that an Atlantic easterly wave developed fairly quickly into Hurricane Hilda of 1964 when it interacted with an upper tropospheric disturbance moving southeastward into the path of the wave. (Naturally, this may have been a coincidence.) The upper-level trough was remarkably small, having a half wavelength of a few hundred kilometers, and might have been missed altogether by the analysis had it not traveled through the dense radiosonde array then in place in Florida and the Caribbean. Reilly (1992) examined cases of western North Pacific cyclogenesis, using NOAA operational analyses with 2.5 degree resolution. He confined his analysis to regions rich in commercial aircraft observations and calculated the dot product of vertical shear with gradients of potential vorticity in the upper troposphere. He showed that some, but not all cases of tropical wave development were associated with anomalously high values of this quasi-balanced forcing parameter. Molinari et al. (1995) analyzed the interaction of a thin cyclonic potential vorticity filament in the upper troposphere interacting with Atlantic Hurricane Elena of 1985 and, beyond the superposition effect, emphasized the importance of the interaction of the upper tropospheric PV anomaly with the storm's anticyclonic outflow.

14.1.4: Inferences from numerical simulations

Beginning in the 1980s, a healthy interplay developed between field experiments and numerical simulations, with each approach informing the other. This subsection provides a brief synopsis of numerical experiments conducted to better understand tropical cyclogenesis.

Using a convection-permitting, nonhydrostatic, axisymmetric model, Rotunno and Emanuel (1987) showed that the RCE state of the tropical atmosphere was stable to the development of infinitesimal, tropical cyclone-like disturbances, absent interactive radiation. This was also shown to be true of a much simpler quasi-balanced axisymmetric model with parameterized convection (Emanuel 1989). But saturating the initial state of this same model, above the boundary layer, led to immediate amplification of even very weak initial vortices. It was proposed that downdrafts carrying low-entropy air into the boundary layer from the lower to middle troposphere overcame the tendency for surface enthalpy fluxes to elevate the moist entropy of the boundary layer in the nascent cyclone core. Artificially elevating the entropy of the lower and middle troposphere in the core of the initial cyclone stops this influx of low entropy to the boundary layer and allows the initial cyclone to amplify.

Partly on the basis of these findings, the Tropical EXperiment in MEXico (TEXMEX) was designed to test the hypothesis that near saturation of a mesoscale column is a prerequisite for tropical cyclogenesis. It was hypothesized that the condition of near saturation was brought about by sustained moist convection moistening a column to near saturation. However, the development of eastern North Pacific Hurricane Guillermo of 1991 proceeded from a column that had been both cooled and moistened by evaporation of rain from stratiform clouds. This column was therefore cold in the lower troposphere, and a smaller, relatively warm anomaly developed within it. The near-saturated condition appeared to have been brought about more by evaporation of rain falling from the stratiform cloud canopy than by detrainment from cumulus clouds.

Is it possible that stratiform rain from a cloud cluster could lead to genesis? To test this idea Bister (1996) inserted a “showerhead” at the center of an axisymmetric tropical cyclone model (Rotunno and Emanuel 1987) at an altitude of 4.375 km, for 36 hours, out to a radius of 116 km. (See the sketch of this configuration in Figure 14.7.) The relative humidity above the showerhead was maintained at 80% over the initial 36 hours, but no wind velocity perturbations were present in the initial condition. While the showerhead adds water mass to the system, it does not add either kinetic or moist static energy and in this respect is unique among numerical simulations of the genesis of single tropical cyclones. Water from the showerhead partially or completely evaporates as it falls, cooling and moistening the column. The sinking cool air produces inflow, maximizing in the lower to middle troposphere, and this spins up a mid-level meso-cyclone. The column under the showerhead approaches saturation by both cooling and moistening.

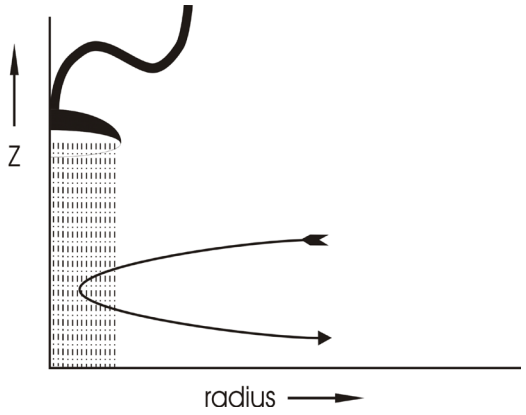


Figure 14.7: Configuration of a "showerhead" inserted into the center of the nonhydrostatic, axisymmetric model of Rotunno and Emanuel (1987). The showerhead is inserted at an altitude of 4.375 km, for 36 hours, out to a radius of 116 km, and the relative humidity is maintained at 80% in the column above the showerhead for the initial 36 hours. No kinetic or moist static energy perturbation is present in the initial state. Water from the showerhead partially or completely evaporates as it falls, cooling and moistening the column. The sinking cool air produces inflow, maximizing in the lower to middle troposphere, and this spins up a mid-level meso-cyclone. The column under the showerhead approaches saturation by both cooling and moistening.

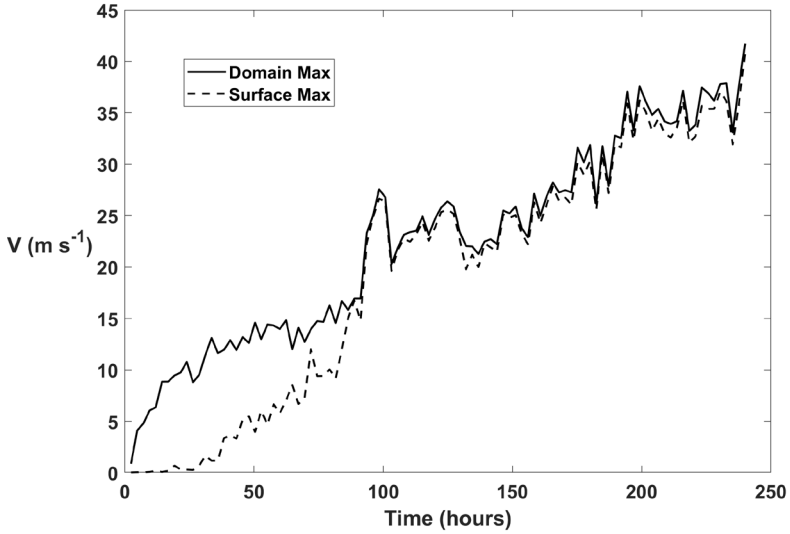


Figure 14.8: Time evolution of the domain-maximum azimuthal wind speed (m s^{-1}; solid) and maximum surface azimuthal wind speed (m s^{-1}; dashed) during a numerical simulation driven by the showerhead illustrated by in Figure 14.7.

The developments over time of the domain-maximum azimuthal wind and the maximum surface azimuthal wind are shown in Figure 14.8, for a simulation very similar to that reported by Bister (1996). The initial spin-up reflects the development of the mid-level meso-cyclone in response to the evaporative cooling. During the first 36 hours, while the meso-cyclone aloft is spinning up, there is essentially no surface wind. Over the next few days, the mesocyclone migrates downward as the condensation level lowers and the locus of maximum evaporation shifts downward. Sensible heat fluxes from the ocean cause shallow convection, which mixes angular momentum downward into the boundary layer.

At around 4 days (96 hours) into the simulation, and about 2.5 days after the showerhead has been turned off, deep convection erupts into the cold, humid core and rapid development of a deep, warm-core cyclone commences, with maximum winds at or near the surface. Figure 14.9 shows a cross-section of azimuthal velocity in a similar showerhead simulation at about the time the warm-core vortex forms inside the broader cold-core mesocyclone. This experiment suggests that given enough time, stratiform rain can lead to genesis.

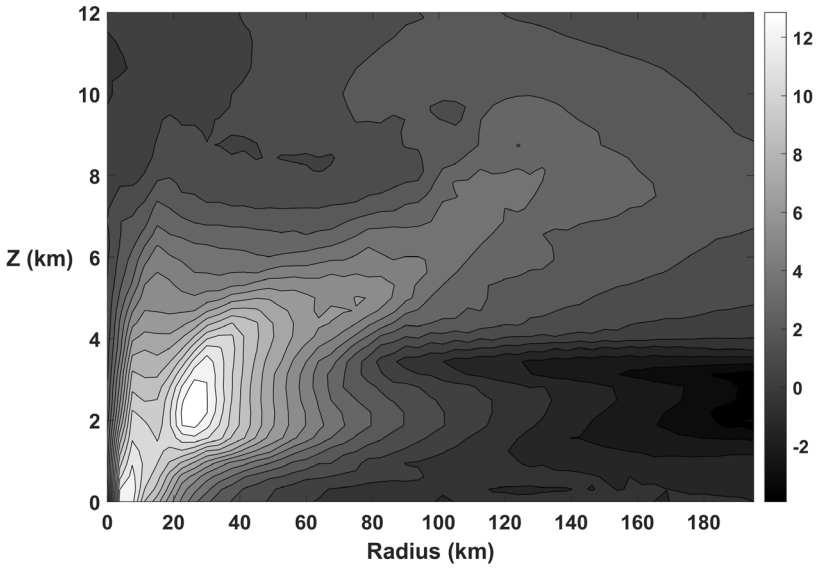


Figure 14.9: Cross-section of azimuthal wind speed (m s^{-1}) at about the time that a warm core vortex appears within the broader cold-core mesocyclone in a simulation using the non-hydrostatic, axisymmetric model of Rotunno and Emanuel (1987). This simulation was performed using a more recent version of the model, that is similar, but not identical, to that used to generate Figure 14.8.

Nolan (2007) performed fully three-dimensional simulations with the cloud-permitting Weather Research and Forecast Model (WRF) in an initial state that is ideal for tropical cyclone formation, with high sea surface temperature and a deep, convectively adjusted troposphere with no mean wind at any level. The model was initialized with a weak, warm-core vortex. Despite these ideal conditions, the initial surface vortex decayed and rapid development did not occur until 48-60 hours after initialization. During this gestation period, the troposphere gradually moistened from both detrainment and evaporation of precipitation from cumulonimbi. When the relative humidity in the core exceeded about 80%, systematic inflow ramped up at mid-levels and a mesoscale cyclone developed at those levels. As it intensified and the core humidity approached saturation, a smaller-scale, warm-core vortex developed within the larger-scale mesocyclone and intensified rapidly while evolving toward a classical mature tropical cyclone.

A somewhat similar study was conducted by Davis (2015) using a different model, but rather than initializing with a weak warm core vortex, he simulated radiative-convective equilibrium (on an *f*-plane, as in Nolan's study) starting from extremely small random perturbations. As in other studies of self-aggregating convection in rotating domains (see Chapter 3, section 3.4), convection slowly aggregated into non-rotating clumps after about 10 days, and these begin to exhibit rotation after around 20 days. These rotating clusters gradually merge into a single

system with a prominent mid-level vortex and a cold core in the lower troposphere. From this point on, the development is like that described by Nolan (2007), with relative humidity increasing at all levels, and a smaller scale, warm-core vortex developing very rapidly within the larger, cold-core mesocyclone. Once the mid-level mesocyclone forms, which takes more than 50 days from the initial state, the whole development proceeds at fast time scales of a few days consistent with those found by Nolan (2007) and Bister (1996). Davis (2015) emphasizes the importance of downdrafts that occur during the formation and development of the mesocyclone. These bring low entropy air from the lower and middle troposphere in contact with the warm sea surface, enhancing the surface enthalpy flux and thereby increasing the column-integrated moist static energy. They way also help to transport momentum from the mid-level cyclone down into the boundary layer, giving an impetus to the subsequent development of the warm-core vortex.

Using the WRF model, Nolan et al. (2007) explored the sensitivity of genesis to parameters and initial conditions, starting from radiative-convective equilibrium states on an *f*-plane with no mean wind. While spontaneous genesis occurred in as few as 15 days, the process was greatly accelerated by placing a weak vortex in the initial condition. While this vortex had peak amplitude at mid-levels, and was therefore cold core in the lower troposphere, it did extend down to the surface. The time to genesis, (defined as the first time the rate of central surface pressure fall exceeded 25% of its maximum value during the rapid intensification phase) decreased as the amplitude of the initial mesocyclone increased. They also showed that the time to genesis decreases with increasing potential intensity. This work, together with subsequent research, shows that while spontaneous genesis is theoretically possible in ideal environments, it takes tens of days, whereas even weak “triggering” disturbances can greatly accelerate the process, particularly if they have low-level cold cores that are nearly saturated.

Nolan and Rappin (2008) and Rappin et al. (2010) extended the work of Nolan et al. (2007) by including wind shear in the RCE state. As found in observational studies and other numerical work, wind shear impedes and may prevent genesis. Rappin et al. (2010) introduced a nondimensional “incubation parameter” (identical to Tang and Emanuel’s (2010) “ventilation parameter”) defined

$$\Lambda \equiv \frac{V_s}{V_p}\frac{h_b - h_m}{h_0^* - h_b}, \tag{14.1}$$

to quantify the effect of vertical wind shear on genesis. (See also equation (12.23) in Chapter 12.) Here, V_s is the magnitude of the imposed wind shear, V_p is the potential intensity, h_b is the environmental boundary layer moist static energy, h_m is a representative measure of the environmental moist static energy in the lower to middle troposphere, and h_0^* is the saturation moist static energy of the sea surface. They found that the time to genesis from a 10 m s^{-1} mid-level mesocyclone is highly correlated with this incubation parameter. (For a particular choice of the definition of time to genesis, the correlation coefficient is 0.94.). The greater the incubation parameter, the longer the time to genesis. Note, however, that Nolan et al. (2010) only considered a single magnitude of wind shear (5 m s^{-1} between 850 and 200 hPa), but Nolan and Rappin (2008) had explored the effects of a range of wind shear magnitudes. As shown in the theory developed by Tang and Emanuel (2010), there is a critical value of Λ beyond which intensification is impossible. Nolan et al. (2010) varied Λ by changing the imposed sea surface

temperature. At higher temperatures, both V_p and $h_0^* - h_b$ are larger, but this is overwhelmed by the increase in the subsaturation of the lower to middle troposphere as represented by $h_b - h_m$ in (14.1). Therefore, in RCE, increasing sea surface temperature holding wind shear constant should decrease the rate of cyclogenesis, if the statistics of triggering disturbances remain constant. We shall return to this issue in Chapter 15.

Dry air above the boundary layer has a strong influence on the effect of shear, as reflected in the incubation or ventilation parameter given by (14.1). But even in the absence of shear, the degree of sub-saturation and proximity of dry air to the nascent storm core has a strong influence on its development (Braun et al. 2012; Nolan and McGauley 2012). The storm core must become nearly saturated before strong development can occur, and this takes more time to accomplish if there is initially more dry air in or near the core.

While shear is almost always an obstacle to storm intensification, Nolan and McGauley (2012), using the WRF model, showed that moderate shear can hasten the formation of low-level vortices, at least given an initial mid-level vortex, but such vortices generally did not develop much after formation. This is also consistent with the observational study by Molinari et al. (2004). They analyzed the formation of Hurricane Danny of 1997 and showed that shear caused a sequence of cluster and mesovortex developments, the last of which finally developed into a tropical cyclone when the environmental shear had relaxed. They concluded that the shear assisted the earlier, sequential development.

These and other numerical simulations, combined with observationally based studies, lead to a conceptual view of how individual tropical cyclones develop, given a favorable environment. This is presented in the next subsection.

14.1.5: Synopsis of the local view of tropical cyclogenesis

Post-war researchers rapidly came to the view that tropical cyclogenesis does not take place spontaneously but needs to be triggered. As Riehl (1948a) put it, “In the cases of hurricane formation noted in the course of the study, deepening began, without exception, in pre-existing perturbations”. Riehl and Kleinschmidt understood that mature tropical cyclones are powered by enthalpy fluxes from the sea, but also suspected that other factors are at work in their genesis. They knew that only a small fraction of cloud clusters during hurricane season end up becoming tropical cyclones. Riehl, in particular, thought that the transformation of a wave disturbance into a tropical cyclone was cause by interactions with upper tropospheric disturbances (Riehl 1948b) and differential radiative cooling caused by the dense cirrus canopy of cloud clusters (Riehl 1963).

It is notable that European researchers had rejected the idea that genesis is powered by latent heat stored in the atmosphere, a hypothesis dating back to the work of Espy (1841). Already, von Hann (1901) had reasoned that ‘‘that since a thundercloud does not give any appreciable pressure-fall but generally even a pressure-rise, it would be unreasonable to assume that a magnifying of this process would cause the strongest pressure falls known’’ (quoted in Bergeron, 1954). Riehl, noting that “many of the heaviest tropical rainfalls occur without the existence of a cyclonic circulation, and even closed depressions with torrential precipitation often fail to deepen” calculated that releasing all the CAPE stored in the tropical atmosphere,

generously using surface-based lifting along a pseudo-adiabat, would lower the surface pressure only to about 1,000 hPa (Riehl 1954). He later remarked that “it has never been easy to see how such an increment [referring to observed pressure falls] can be obtained from internal storm characteristics only” (Riehl 1975). Nevertheless, many textbooks still claim that tropical cyclones are powered by latent heat release, and various forms of this idea continue to appear in the literature in connection with genesis (e.g. Ooyama 1964; Charney and Eliassen 1964; Montgomery and Enagonio 1998; Montgomery et al. 2006; Kilroy et al. 2017). Given the often magnificent appearance of cumulonimbi, the idea that deep moist convection drives genesis is understandable but undermined by the discovery that tropical cyclones can form in dry-adiabatic atmospheres with no moist convection (Mrowiec et al. 2011; Cronin and Chavas 2019; Velez-Pardo and Cronin 2023).

On the contrary, a strong case can be made that moisture in the atmosphere impedes genesis even though it permits ultimately stronger vortices given constant potential intensity (Cronin and Chavas 2019). In dry adiabatic atmospheres in RCE, there is no strong entropy minimum in the middle troposphere as we find in the moist case. Consequently, there is no reservoir of low entropy air to enter the boundary layer during genesis and prevent the boundary layer entropy from increasing. Experiments with a dry version of a nonhydrostatic axisymmetric model show that very small amplitude initial vortices amplify (Figure 14.10).

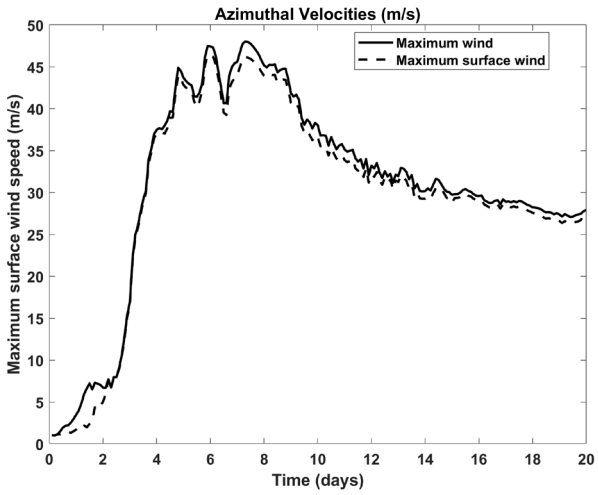


Figure 14.10: Domain-maximum (solid) and surface-maximum (dashed) azimuthal velocities in a dry version of the nonhydrostatic axisymmetric model of Rotunno and Emanuel (1987) with no damping of temperature perturbations. Initial warm-core vortex amplitude is about 1 m s^{-1}.

Note that in a dry RCE state, the ventilation (or incubation) parameter given by (14.1) vanishes because $h_b - h_m$ is nearly zero. Also, as noted before, as temperature increases in the moist RCE state, the ventilation parameter generally increases (depending to some extent on what forces the temperature increase), thereby increasing the main inhibition to genesis. Moisture makes genesis harder. (At the same time, other processes, like cloud-radiation feedbacks, that tend to enhance genesis may also increase with temperature.)

Through all the observational and numerical model studies reviewed in the previous section, with all the different particular routes to genesis, there is a common thread: Whatever happens, a mesoscale column of the atmosphere must become nearly water saturated before real

development can take place. Given that the density temperature lapse rates are seldom steeper than a reversible moist adiabat, this near saturated condition is equivalent to a state in which the moist static energy does not change much with altitude. (This is always true in dry RCE states since the static energy is constant with height in such states; there is no inhibition to overcome.)

The need for near saturation is consistent with the finding that in standard moist convection in an unsaturated environment, almost all of the irreversible entropy production is used up in mixing across water vapor gradients and by frictional dissipation of falling rain, with very little left over for kinetic energy dissipation (Pauluis and Held 2002). The environment must approach quite close to saturation before the system transitions to one in which dissipation of wind energy is the dominant irreversible process.

There are three ways to make the moist static energy constant with height starting from a typical tropical state (or a moist RCE state), as illustrated with the aid of Figure 14.11.

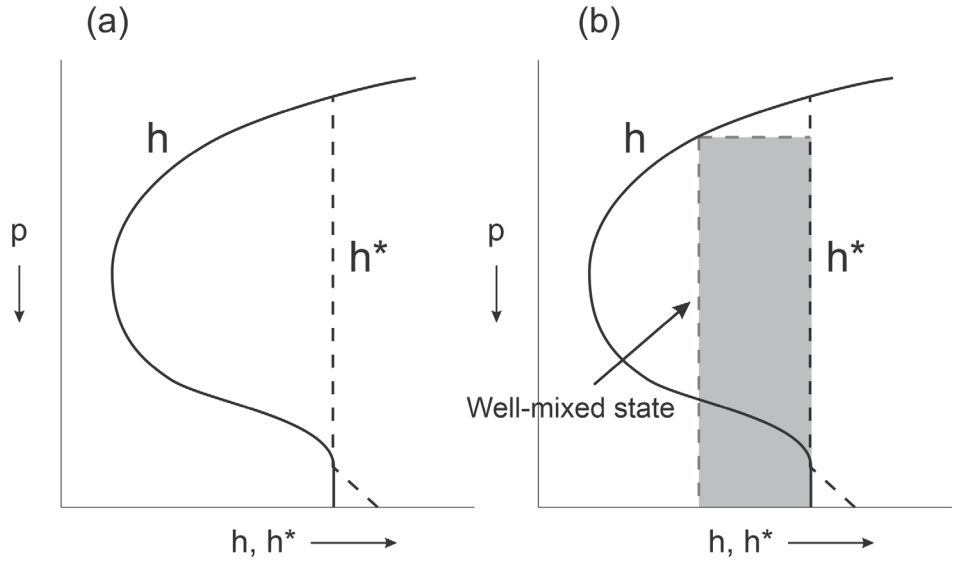


Figure 14.11: a) Typical profiles of moist static energy, h, and saturation moist static energy, h, in the tropical atmosphere. b) Same as a) but showing an additional profile (vertical dashed line in middle of plot) that represents both h and h* of a state that results from conservative vertical mixing of moist static energy. The difference between the initial and final profiles of h* is shaded in gray.*

There are three (not mutually exclusive) ways to create a saturated, vertically uniform distribution of moist static energy:

1. Add enough moist static energy to the column to bring it to saturation. This would entail filling in the whole area between the h and $h*$ curves in Figure 14.11a.

2. Mix the column vertically to homogenize the moist static energy without adding or subtracting column-integrated moist static energy. This state is indicated by the dashed vertical line in the center of Figure 14.11b. Note that the resulting column, having a smaller $h*$ than the initial state, will therefore be colder.

3. Lift the whole column over an extended time. Given a positive value of the gross moist stability, this will require work to be done on the column by some external disturbance. The final moist static energy profile may somewhat resemble the profile that results from mixing.

In the first case, large quantities of energy would have to be added to the column. For a randomly chose tropical sounding (the island of Koror in the western tropical North Pacific at 12 GMT on 1 April 2018), about $3.6 \times 10^8 \, J \, m^{-2}$ would have to be added to the column to bring it to saturation without changing its temperature. A surface enthalpy flux of $150 \, W \, m^{-2}$ would have to be sustained over about 28 days to achieve this, with complete blockage of outgoing longwave radiation by high clouds. This is about the time scale for spontaneous aggregation of moist convection through cloud-radiation interactions (Davis 2015; Wing et al. 2016). Very few tropical cloud clusters stay over warm water this long, so is unlikely that process 1) can lead to genesis by itself in the real world, though to be sure, cloud-radiation interactions contribute tangibly to genesis (Riehl 1963; Muller and Romps 2018).

Process 2) in its pure form does not change the column-integrated moist static energy (by design) and leads to a state that is water saturated but somewhat colder than the initial state. This state cannot exist on time scales longer than those associated with internal gravity waves unless the column is rotating and in thermal wind balance. It is important to note that cooling and moistening by evaporation of rain, accompanied by moist adiabatic adjustment, can also accomplish this. The operation of process 2) and/or 3) is consistent with the observation and modeling result that developing cloud clusters usually pass through a cold-core state with cyclonic vorticity at mid-levels. The process of bringing a column to saturation by process 2) or 3) can be substantially faster than by route (1). There is no need for a mid-level cyclonic disturbance to be present from the beginning; it should be regarded as something that occurs along the road to genesis.

The development of a tropical cyclone might be thought of as occurring in four stages, as illustrated in Figure 14.12.

In the first stage, a cloud cluster is triggered by some external disturbance such as a tropical wave or an extratropical front or upper tropospheric disturbance moving into the tropics. There are apparently many ways this can happen. The cloud cluster has a high, dense cirrus canopy with low effective infrared temperatures. Downdrafts import low entropy air from the lower and middle troposphere into the boundary layer. This stage can and often does take place over land.

During the second stage, much of the original deep convection diminishes and may even disappear, but the high cloud canopy remains largely intact. Evaporation of rain and mixing by shallow cumuli and congestus clouds cool and moisten the lower troposphere, and inflow at mid-levels gradually spins up a mesocyclone in the middle troposphere. This mesocyclone may slowly migrate downward, and shallow convection may transfer some angular momentum into the boundary layer. Any lowered boundary layer moist static energy begins to recover. The dense cirrus canopy blocks outgoing longwave cooling, allowing surface fluxes to gradually increase the column moist static energy. This gestation period can be disrupted or terminated by wind shear and/or passage over land or cold water; otherwise, it usually lasts a few days.

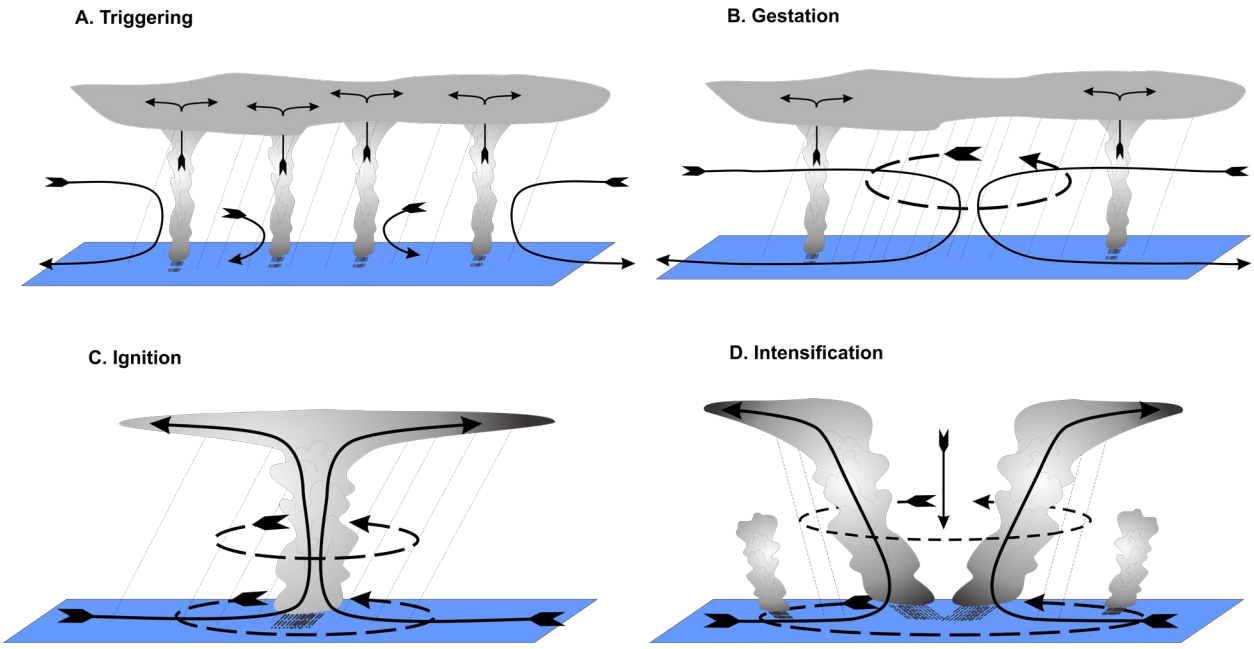


Figure 14.12: Four stages in the development of a tropical cyclone. A.) Triggering: A cloud cluster is initiated by one or more of a set of possible processes; b) Gestation: While the deep convection may diminish, the stratiform rain area persists and mid-level inflow continues, driven by evaporation of rain and melting of snow. The air cools and moistens, and the inflow gradually spins up a mid-level mesocyclone. This stage typically lasts a few days. C) Ignition: A new episode of deep convection penetrates the pre-moistened, cool air in the center of the mesocyclone. Downdrafts are weak and their moist entropy is not greatly depressed. Inflow and cyclonic spin up commence in the boundary layer. D) The WISHE feedback between intensifying surface winds and surface enthalpy fluxes leads to an intensifying tropical cyclone

When the air near the center of the mid-level mesocyclone becomes humid enough, and the boundary layer moist static energy has increased sufficiently, a new episode of deep convection erupts. Because the troposphere in the core of the system is nearly saturated, the deep convection can produce only weak downdrafts, though stronger downdrafts may occur outside the core. Inflow into the deep convection occurs in the boundary layer, and cyclonic vorticity develops rapidly there. In this ignition stage, the feedback between wind and surface enthalpy flux becomes active and the system begins to intensify as a tropical cyclone, leading to the final (and sometimes rapid) intensification phase.

This paradigm fits much of the field experimental observations and numerical modeling results presented in the previous subsections. On the basis of his analysis of satellite imagery, conventional observations, and airborne reconnaissance data collected in the early 1980s, Zehr (1992) developed a paradigm similar to the one described in Figure 14.12. He begins by stating that “Tropical cyclogenesis in the western North Pacific is typically characterized by an early convective maximum which precedes the initial appearance of a tropical depression” and suggests that this early maximum is a result of synoptic-scale forcing. He then states that “A weak mesoscale vortex is initiated which is located within the larger broad-scale cyclonic circulation of a pre-existing tropical disturbance…..This early convective maximum and associated mesoscale vortex formation precede first designation as a tropical storm by an average period of about 3 days.” Zehr includes our stage 2 as part of his first stage: “Stage 1 also includes the period following the convective maximum where the overall convection is considerably reduced. It is typically during this period that a distinct low-level circulation center can first be identified, and persistent cyclonically curved convective cloud lines are observed.”

Zehr's second stage is equivalent to our third: "The end of Stage 1 and the onset of Stage 2 is marked by the onset of increasing deep convective activity associated with a low-level circulation center. It is usually during Stage 2 that the disturbance is first designated as a tropical depression."

Based on their analysis of five tropical disturbances that formed in the northwest Pacific region in August and September of 2008, during the TPARC/TCS-08 project, Raymond et al. (2011) conclude that "The formation of a strong midlevel circulation, with its associated cold core at low levels and warm core aloft, greatly aids the spin-up of a tropical cyclone by changing the vertical mass flux profile of deep convection from top heavy to bottom heavy." They go on to remark that "This has two effects: (1) the enhancement at low levels of the convergence of mass and hence vorticity, thus aiding the spin-up of a warm-core vortex and (2) the suppression of the lateral export of moist entropy by deep convective inflows and outflows from the core of the developing system." The bottom-heavy upward mass flux by convection is also consistent with the absence or near absence of strong evaporatively driven downdrafts, enabling strong convergence and spin-up at low levels. This also has the effect of tamping down the inward flux of low entropy air at middle levels that characterizes the second stage of development in our paradigm.

The spin up at middle levels during stage 2 need not be initially concentrated in a single mesocyclone but may be distributed among a number of smaller mesocyclones which eventually merge into one (e.g. Davis 2015).

Raymond et al. (2011), along with Dunkerton et al. (2008), Wang et al. (2010), Montgomery and Smith (2012), and Montgomery et al. (2012) emphasize the vulnerability of stage 2 (in our nomenclature) to external disruption. They suggested that external flow configurations that are favorable to closed trajectories of air parcels in the lower to middle troposphere, relative to the translating cloud cluster, are necessary to avoid intrusions of low entropy air into the cores of developing disturbances. As the mesocyclone at mid-levels intensifies, its inertial stability also acts as a break on dry air intrusion.

Stage 2 can be skipped altogether with strong enough synoptic-scale forcing. For example, in the case of medicanes and polar lows, the large-scale trigger is itself a cold-core system with a humid interior, providing the favorable embryo in which a warm core tropical cyclone-like disturbance may amplify. The genesis of Atlantic Hurricane Diana of 1984 occurred as an intense cyclonic potential vorticity anomaly at the tropopause broke off from its parent Rossby wave. The formation of an intense, cold, saturated core through strong ascent in response to this development, together with the quasi-balanced forcing of cyclonic vorticity at the surface, both triggered and "ignited" the development of Diana (Bosart and Bartlo 1991).

This concludes our summary of some of the extant thinking on the local view of tropical cyclogenesis. We next turn to a global view of genesis.

14.2 The global view

The global view of tropical cyclogenesis inquires about the environmental factors that govern the statistics of genesis, including its global frequency, spatial and seasonal variations, and response to natural and anthropogenically forced variability. In a more advanced form, it could account for feedbacks of tropical cyclones on the climate system; a discussion of this idea will be deferred to Chapter 15.

14.2.1 Genesis indices

One approach to understanding the physical basis of the observed climatology of tropical cyclones is to see how it is correlated with known physical properties of the climate system. Palmén (1948) was probably the first to address the physical basis of the observed climatology of tropical cyclones. He believed that such storms are powered by CAPE stored in the tropical atmosphere and showed that they only occur over those portions of the oceans where CAPE is usually positive. He demonstrated that this condition is strongly correlated with region in which the sea surface temperature exceeds 26-27°C. Palmén also recognized that tropical cyclones do not develop where the magnitude of the Coriolis parameter is too small.

Gray (1975, 1979) developed the first genuine genesis index and tested it against observed spatial and seasonal variations of tropical cyclone activity. Gray's "seasonal genesis parameter" was defined (with some changes of notation) in terms of seasonal mean variables:

$$SGP = C \underbrace{f \frac{\zeta \pm 5\times10^{-6} s^{-1}}{|\Delta \mathbf{V}| + 3 m s^{-1}}}_{Dynamic} \underbrace{K_o \left(\theta_{eb} - \theta_{em} + 5K\right)\left(\mathcal{H} - 40\%\right)}_{Thermodynamic}, \tag{14.2}$$

where C is a (dimensional) constant, f is the Coriolis parameter, ζ is the low-level relative vorticity (in units of s^{-1}), $|\Delta \mathbf{V}|$ is the magnitude of the vector wind shear between 950 hPa and 200 hPa (in $m s^{-1}$), θ_{eb} and θ_{em} are the equivalent potential temperatures (in K) in the boundary layer and at 500 hPa, respectively, $\mathcal{H}$ is the relative humidity (in %) in the middle troposphere, and the "ocean thermal energy", K_o, is defined as

$$K_o \equiv \int_0^{60\,m} \rho_w c_w \left(T - 26\ ^oC\right) dz, \tag{14.3}$$

where ρ_w is the density of seawater, c_w is its heat capacity, and T is the seawater temperature in ^oC. The integrand in (14.3) is bounded below by zero, and the relative humidity in (14.2) is bounded below by zero and above by 100%. In the third factor in (14.2), the plus sign I used in the northern hemisphere, and the minus sign in the southern.

Gray emphasized that favorable conditions for genesis can be thought of as consisting of thermodynamic contributions and dynamic (or kinematic) contributions.

The index recognizes that tropical cyclones do not form very near the equator, where the Coriolis parameter vanishes, and may be more likely to form where the seasonal mean vorticity is already large, such as in a poleward-displaced ITCZ. His deep tropospheric shear factor appears in the denominator since shear is usually detrimental to genesis. It was also recognized

that a dry lower to middle troposphere is not conducive to genesis. But two of his factors have less physical justification. Gray, like Palmén and some researchers of that era, thought that the degree of conditional instability is important, and mistakenly equated that with the negative vertical gradient of moist entropy in the lower troposphere. Subsequent observational and model-based studies clearly show that genesis is inversely proportional to this gradient. But Gray's mid-tropospheric relative humidity factor works in the oppositive direction and it can be shown that the product of the two factors is nearly constant. This leaves only K_o as an important thermodynamic predictor.

But there are two problems with K_o. First, it has an arbitrary threshold sea surface temperature of 26°C, and while that worked reasonably well in the climate of that era, it is now understood that a threshold, even if it is justified, changes with climate. Second, although upwelling of cold water is an important influence on mature tropical cyclones, as we discussed in Chapter 13, there is little evidence that the temperature of water much below the surface has any effect on genesis.

Inspired by the success of Gray's seasonal genesis parameter, other genesis indices were developed more recently (Emanuel and Nolan 2004; Emanuel 2010; Tippett et al. 2011; Bruyère et al. 2012). Based on both theory and models, potential intensity was substituted for sea surface temperature and shear and vorticity factors were retained, although the latter appears to enter as more of a threshold effect (Tippett et al. 2011). Nondimensional numerical models (e.g. Emanuel 1989, 1995) point to the saturation deficit of the middle troposphere being the relevant humidity factor, rather than relative humidity (Emanuel 2010). Rappin et al. (2010) and Tang and Emanuel (2010) proposed that the humidity, shear, and potential intensity factors should enter as a particular nondimensional ratio of the shear and the potential intensity, all multiplied by a nondimensional measure of the saturation deficit of the middle troposphere. This nondimensional ratio was shown by Rappin et al. (2010) to be highly correlated with the length of time for a weak initial vortex to begin to amplify at a substantial rate. Genesis indices have been found to correlate with modulation of tropical cyclones by El Niño–Southern Oscillation (ENSO; Camargo et al. 2007a) and the Madden Julian oscillation (MJO; Camargo et al. 2009), and to account for some aspects of climate change simulated by climate models (Camargo et al. 2007b; Zhang et al. 2010; Korty et al. 2012a,b; Camargo 2013; Emanuel 2013, 2020).

The "genesis potential index" (GPI) developed by the author (Emanuel, 2010) was constrained to be dimensionally consistent, yielding a genesis rate in terms of number of cyclones per unit area per unit time. That index is

$$GPI \equiv C \frac{|\eta|^3 \, \chi^{-4/3} \left(V_{pot} - 35 ms^{-1} \right)^2}{(25 ms^{-1} + S)^4}, \tag{14.4}$$

where C is a dimensionless constant, η is the absolute vorticity, V_{pot} is the potential intensity (in ms^{-1}), S is the magnitude of the 250 hPa to 850 hPa wind shear (in ms^{-1}), and χ is a nondimensional measure of the subsaturation of the lower to middle troposphere:

$$\chi \equiv \frac{s^* - s_{600}}{s_0^* - s^*}, \tag{14.5}$$

where $s*$ is the saturation entropy of the troposphere (which is nearly constant with height in a moist adiabatic atmosphere but which is in practice evaluated at 600 hPa), s_{600} is the actual moist entropy at 600 hPa, and s_0^* is the moist entropy of air saturated at sea surface temperature and pressure. The factor $\left(V_{pot} - 35 ms^{-1}\right)^2$ is set to zero for all potential intensities less than $35 ms^{-1}$. The predictors in (14.4) were chosen to match the inputs to a simple, nondimensional numerical model (Emanuel 1995) and the various coefficients were chose to optimize the index's ability to predict spatial and temporal variability of observed genesis.

The exponents and numerical factors in (14.4) were optimized using monthly-mean values of the variables from reanalysis data. The GPI is designed to be used with monthly-mean inputs.

This and other *GPI*s are modestly successful at capturing the spatial distribution of genesis. Figure 14.13 compares the *GPI* given by (14.4) and applied to European Center CERA-20C monthly mean[1] reanalysis data (Laloyaux et al. 2018), summed over the period 1980-2010, with genesis points during the same period from standard IBTrACS historical data (Knapp et al. 2010).

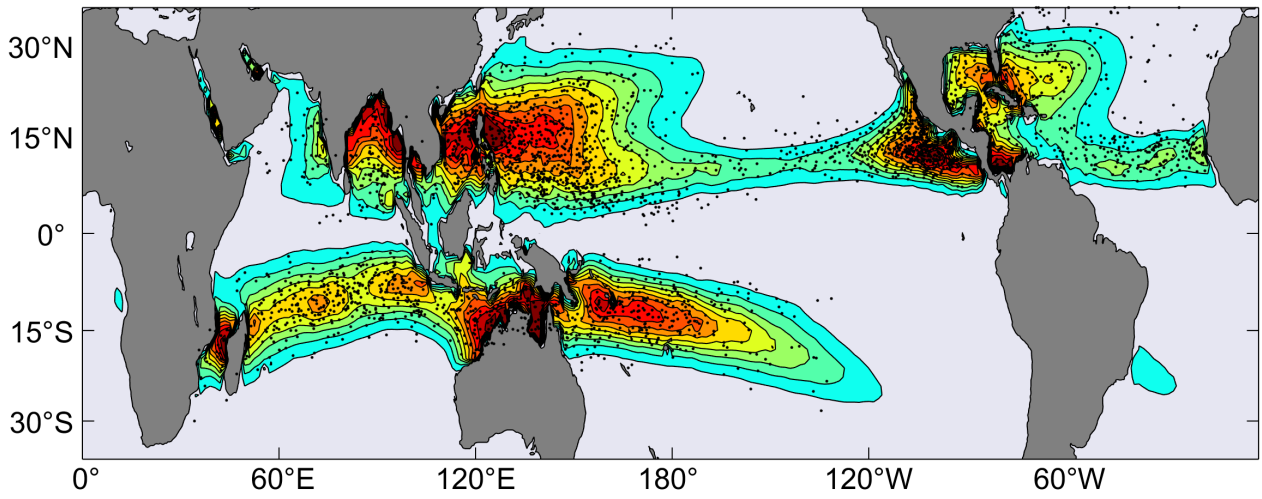


Figure 14.13: Genesis Potential Index given by (14.4) summed over the all the months from January, 1980 through December, 2010 (contours) from CERA-20C reanalyses, and first track points (black dots) in the IBTrACS historical data over the same period.

For the most part, there is a reasonable correspondence between the index and the density of genesis points. The index misses some subtropical developments, especially in the North Atlantic. This is likely owing to tropical transitions and to developments that take place under cold-core, cutoff cyclones aloft, in which the air is anomalously close to saturation and the potential intensity is elevated over its monthly mean values. Here the *GPI* based on monthly mean inputs may be somewhat smaller than the monthly mean of *GPI*s calculated with daily values. Also, in the central Pacific region as well as the North Indian Ocean, there are several

[1] For the shear in (14.4), shear magnitudes were calculated each day and then averaged over each month. For all other inputs to (14.4), monthly mean quantities were used.

instances of genesis equatorward of expectations based on the *GPI*. These are places where the daily values of relative vorticity may occasionally be somewhat larger than the monthly mean values, so here again the *GPI* based on monthly mean inputs may be smaller than that based on daily data and then averaged over a month.

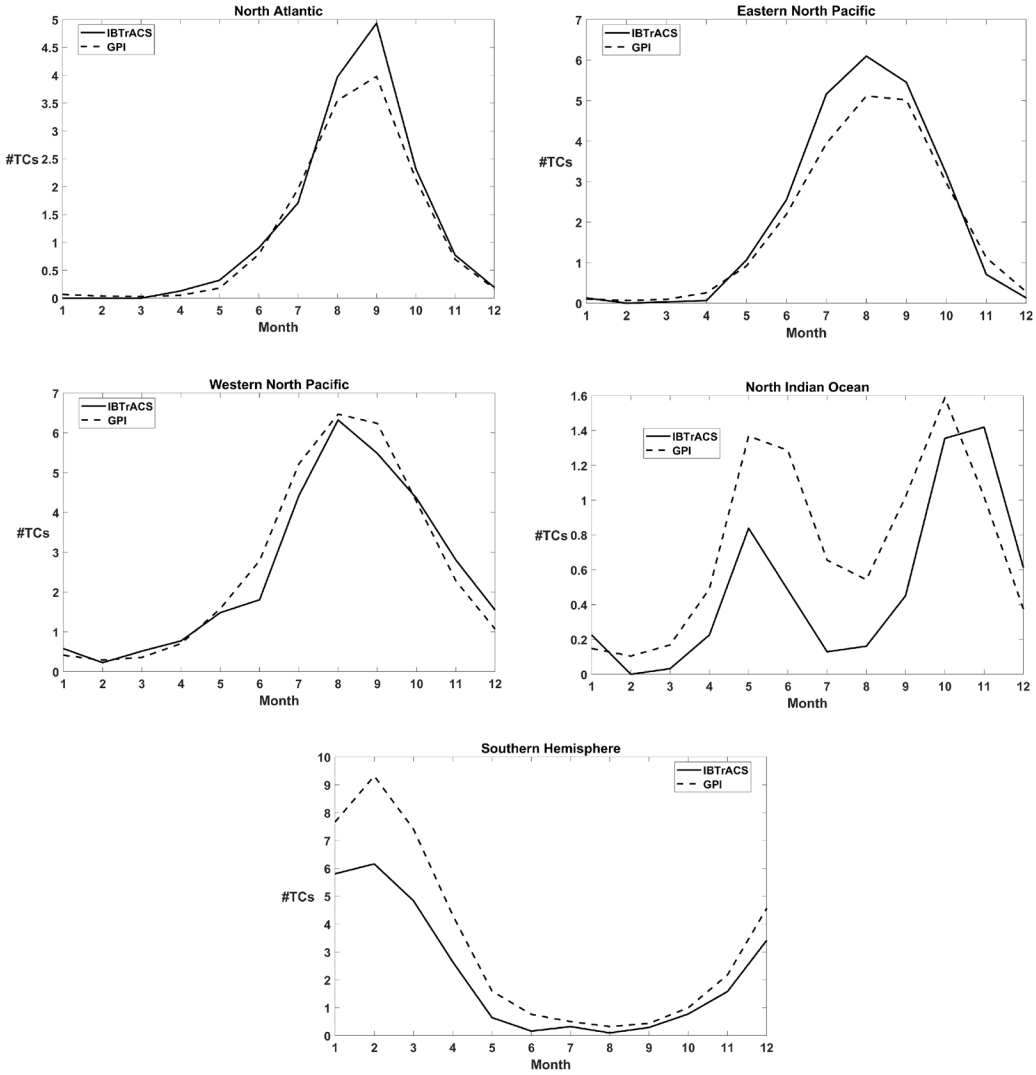


Figure 14.14: Monthly frequency of tropical cyclones in five ocean basins, from IBTrACS data (solid) and from the genesis potential index (GPI) given by (14.4) and applied to CERA-20C reanalyses (dashed). Both frequencies have been averaged over the period 1980-2010.

The seasonal cycle of the *GPI* given by (14.4) is compared to historical data in Figure 14.14. The index captures the seasonal cycle quite well, but over-predicts activity in the North Indian Ocean and in the southern hemisphere, and underpredicts overall activity in the North Atlantic and eastern North Pacific basins.

Comparison of the annually summed *GPI* given by (14.4), driven by CERA-20C reanalyses, with annual tropical cyclone frequencies in the North Atlantic region from 1980 to 2014 yields a correlation coefficient of about 0.75, so that more than half of the interannual variations in cyclone counts are “explained” by the *GPI*. Since roughly half the observed variance is consistent with a pure Poisson random process, the *GPI* accounts for most or all the rest. But

interannual variations of the *GPI* do not explain the observed interannual variability in other ocean basins, and by construction, *GPI* cannot explain the global mean count of about 90 storms per year, given the presence of the dimensionless constant C in (14.4), that must be tuned to give the correct annual global mean.

In summary, *GPI*s, which are based strictly on large-scale, time mean (usually monthly mean) variables, predict much of the spatial, and seasonal variability of tropical cyclones, and their response to large-scale phenomena such as ENSO and the MJO. They also predict much of the non-random interannual variability of tropical cyclones in the North Atlantic. The underprediction of genesis activity in the North Atlantic and eastern North Pacific may point to a role for triggering disturbances, like African easterly waves, in those places. Likewise, the overprediction of cyclones in the southern hemisphere may be owing to a dearth of triggering disturbances there. Finally, the over-prediction of genesis in the North Indian Ocean may simply be owing to there not being enough space for tropical cyclones to develop in the confines of the Bay of Bengal and the Arabian Sea.

Yet observations show that most if not all tropical cyclones originate in pre-existing disturbances. These disturbances are largely if not entirely filtered out in taking monthly means, so we are left with the seeming paradox that while cyclones originate in pre-existing disturbances, the climatology of those disturbances does not appear to play a large role in the climatology of tropical cyclones.

A strong clue to the resolution of this paradox is provided in the work of Patricola et al. (2018) and Danso et al. (2022). They examined tropical cyclone formation during the North Atlantic seasons of 2005 and 2020, respectively, using a regional, convection-permitting model (WRF; Skamarock et al. 2019) covering the tropical North Atlantic and driven by time-evolving sea surface temperatures and lateral boundary conditions supplied by the European Center for Medium-Range Weather Forecasts (ECMWF) ReAnalysis (ERA5; Hersbach et al. 2020). They ran a ten-member ensemble for the 2005 season and a three-member ensemble for the 2020 season, with ensemble members differing only in their initial conditions.

In each case, and for each ensemble member, they ran a control experiment with full, time-dependent boundary conditions and a second set in which a filter was applied at the eastern boundary of the domain, near the African coast, that filtered out all variability in the range of 2-10 days. This range includes African easterly waves (AEWs). They found that as many or more tropical cyclones formed when AEWs were suppressed, but under suppressed conditions storms tend to form further west and become more intense, perhaps because more of the tracks moved into the western Atlantic and Gulf of Mexico, where potential intensities are usually higher. Moreover, in the control experiment, the locations and timings of genesis were highly correlated among ensemble members and with measures of AEW location and time, while in the experiments with AEWs suppressed, the locations and timings of genesis were uncorrelated among ensemble members.

It would seem from these experiments that while potential initiating disturbances like AEWs indeed determine *when* and *where* tropical cyclones form, they may not determine *if* they form.

It should be recognized, however, that although 2-10 day variability was filtered from the eastern boundary and succeeded in suppressing AEWs, disturbances outside that frequency range could still pass through the eastern boundary, and variability at all sub-seasonal frequencies

could develop within the domain or enter through the northern and southern boundaries. But the lack of correlation of genesis location and timing among ensemble members with 2-10-day variability suppressed at the eastern boundary makes it unlikely that TCs were being triggered by any disturbances entering through the boundaries.

This view of genesis climatology suggests that genesis is more “fuel-limited” than “trigger-limited”.

It should also be recognized that there is appreciable interannual variability of AEW activity and those variations may be weakly correlated with Atlantic tropical cyclone metrics (Thorncroft and Hodges 2001). While such a correlation may point to a causal relationship between AEWs and genesis, it is also possible that larger-scale conditions affect both phenomena in the same direction.

Finally, at the opening of this chapter, we remarked that the observed global number of tropical cyclones is around 90, with a variance around that number consistent with pure Poisson variability. When we calculate that number from globally summed monthly GPI (14.4) applied to the CERA-20C reanalysis over the whole period 1901-201, we see a time series with no trend and with a variance about the mean of only about 38% of what would be expected from a pure Poisson process (Figure 14.15). This suggests that much of the observed global variability is related to randomness associated with triggering and/or sub-monthly variability of the large-scale environment.

We next turn to a description of tropical cyclone climatology in idealized climate states.

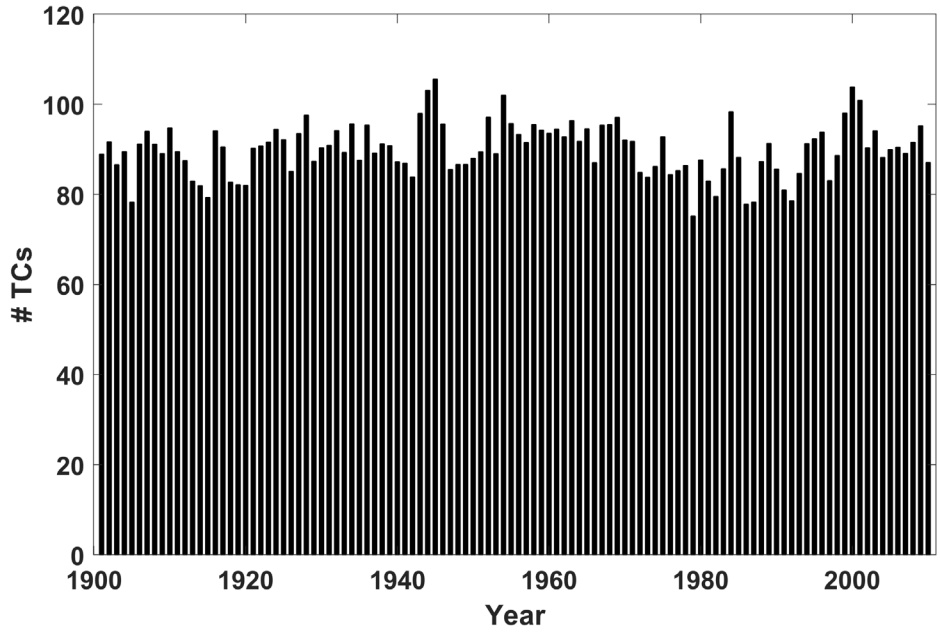


Figure 14.15: Globally and annually summed monthly genesis potential index (GPI) given by (14.4) and driven by CERA-20C reanalysis data, 1901-2010. The mean is 90 (by calibration to post-1980 observations) and the variance is 34, about 38% of what would be expected from a pure Poisson process.

14.2.2 "TC-World" Experiments

We have seen that environmental wind shear is detrimental to tropical cyclogenesis and development, and that most if not all real tropical cyclones are initiated by pre-existing disturbances that do not operate (principally or at all) through surface flux feedbacks. What happens when neither wind shear nor potential triggering disturbances are present?

One way to explore this issue is to find out what happens on a conceptually infinite $f-$plane with imposed, uniform sea surface temperature and calculated radiative transfer; in other words, RCE on an $f-$plane. This is been accomplished by extending the numerical framework used to explore self-aggregation of convection (Chapter 3, section 3.4) by including a constant Coriolis parameter. Much has been learned from this approach, which was pioneered by Bretherton et al. (2005). These have come to be known as "TC-world" experiments.

We have emphasized all along that tropical cyclones are surface flux-driven phenomenon, and that while phase change of water alters the intensity and structure of cyclones, it is not necessary to the phenomenon itself. Simulations of dry tropical cyclones were described in Chapter 9. This leads us to ask: Under what conditions do tropical cyclone-like vortices develop, as opposed to rotating convective structures typical of classical Rayleigh-Bénard convection (Emanuel 1994)?

The classical approach to understanding convection began with experiments by Henri Bénard (Bénard 1900) and was first treated theoretically, through linear stability analysis, by Lord Rayleigh (Rayleigh 1916). The approach considers the development of convection in a Boussinesq fluid between two rigid plates of infinite extent, the lower of which has a higher temperature than the upper. There is an extensive literature on this approach but little intersection between this community and the those who study atmospheric convection. Almost all the laboratory and theoretical studies of Rayleigh-Bénard convection consider aerodynamically smooth boundaries through which the fluxes of heat and momentum are always constrained by molecular diffusivities. Even with very strong thermal forcing, the solutions still depend on these diffusivities. By contrast, atmospheric boundary layer theory (e.g. Monin-Obukhov similarity (Monin and Obukhov 1954); see Chapter 4) ignores molecular diffusivities from the start, assuming that the lower boundary is aerodynamically rough. This difference between smooth and rough boundaries has divided the group of scientists studying convection into two essentially non-communicating groups for many decades.

Rotation was eventually added to the set-up (Chandrasekhar 1953, 1961). The upshot of these laboratory, theoretical, and eventually numerical studies, is that when the temperature difference between the two plates (or else the imposed heat flux through the boundary) exceeds a critical value, the system becomes linearly unstable and convection begins in the form of horizontally periodic cells. (The critical value depends on the imposed temperature difference or heat flux, the distance between the boundaries, the molecular diffusivities of heat and momentum, and the Coriolis parameter. These can be combined into three non-dimensional numbers that govern the behavior of the system.)

Under most circumstances, with fixed temperature boundaries, the aspect ratio of the convection is around unity – the horizontal distance between updrafts and downdrafts is on the order of the depth of the fluid. With background rotation, and with relatively weak thermal forcing, the updrafts are cyclonic in the lower half of the fluid and anticyclonic above that; and

vice-versa for the downdrafts. The spacing of these vortices remains on the order of the depth of the fluid. If the thermal forcing is sufficiently large to result in a highly turbulent flow, and if the rotation rate is high enough, domain-filling cyclones develop (e.g. Guervilly et al. 2014), but unlike tropical cyclones, these have height-invariant structure and are apparently driven by a local inverse cascade of energy from the convective-scale vortices, has had been envisioned by Montgomery et al. (2006) for tropical cyclones.

On the other hand, when the heat flux rather than the temperature itself is prescribed at the boundaries, the convective cells can become much broader than they are deep. And when the boundaries are permitted to be aerodynamically rough, so that mechanical turbulence limits the heat flux near the boundaries, the resulting convection can be of quite different character.

Recently, Velez-Pardo and Cronin (2023) explored the character of numerical simulations of Rayleigh-Bénard convection under different combinations of thermal boundary conditions, including fixed temperature, flixed heat flux, and aerodynamic fluxes. (They did not consider phase change of water.) Their findings are summarized in Figure 14.16.

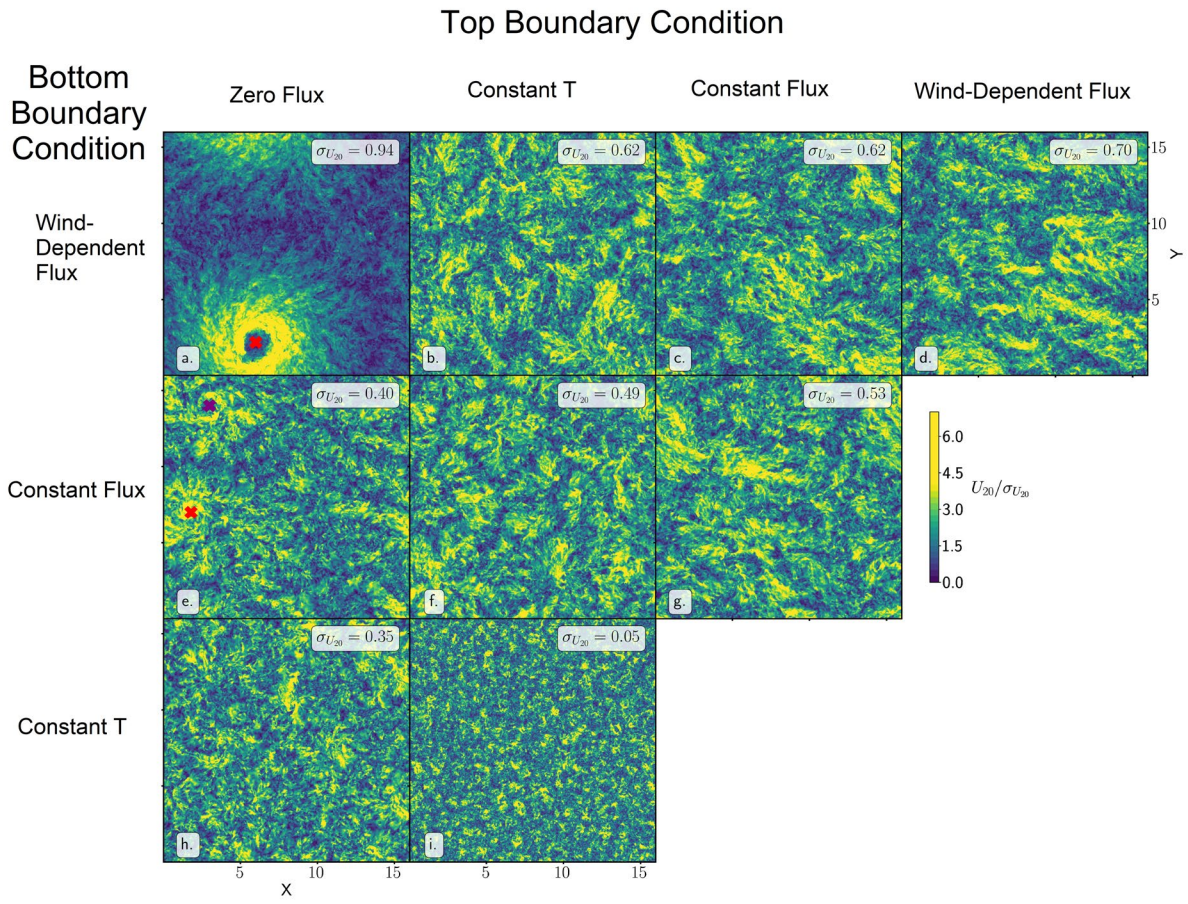


Figure 14.16: Horizontal wind at 20 percent of domain height for simulations with moderate Coriolis parameter. These snapshots were taken after 10 rotational periods into the simulations. Magnitudes are normalized by the standard deviation in each domain, and color shades saturate at 3.5 times the standard deviation. Rows represent different bottom thermal boundary conditions and columns are top boundary condition. Locations of near-surface pressure minima corresponding to persistent cyclonic vortices are indicated with "x" markers. Horizontal distances are in multiples of the fluid depth.

The figure shows snapshots of wind speed normalized by its domain-wide standard deviation at an elevation that is 20% of the domain depth, for a set of simulations in which the thermal forcing is sufficient to produce states with moderate to strong turbulence, and for a nondimensional measure of the Coriolis parameter to be moderate (see Velez-Pardo and Cronin 2023 for details). The snapshots are taken 10 rotational periods after initialization. Thermal boundary conditions are calculated by fixing the boundary temperature, fixing the boundary heat flux, or applying the aerodynamic flux formula (4.15), which depends on flow speed near the boundary, among other things. The type of boundary condition applied to the lower boundary is specified along the left side of Figure 14.16, while the top boundary condition is shown along the top. Note that for the combinations of zero flux at the top boundary and either constant flux or aerodynamic flux at the bottom boundary, the domain-mean temperature increases linearly in time. This is isomorphic to the case where we specify a constant, domain-mean heat sink in the interior of the fluid, somewhat like radiative cooling in RCE.

In all cases, no-slip boundary conditions are applied to the horizontal momentum equations. Only the upper left diagonal of the boundary condition matrix is shown because the model equations are invariant to changing the sign of the vertical coordinate and the temperature perturbation together.

For the classical fixed-temperature boundary conditions at both boundaries used by Rayleigh (1916), the separation between convective elements is about 1 fluid depth. When at least one of the two boundaries has either fixed thermal flux or uses the aerodynamic formula, the convection cells are somewhat broader than they are deep, by roughly a factor of 5.

No long-lived vortices are apparent in the fields except in the case of zero flux at the top boundary and either constant flux or aerodynamic flux at the bottom boundary. The latter case is equivalent to what most most TC-world simulations specify….no heat flux through the top of the domain and aerodynamic heat fluxes at the bottom, with (an implied) constant cooling applied through the interior. Only in this case do we see genuinely TC-like vortices emerge. The domain used here, in combination with other parameters, is only large enough to contain a single vortex in this case, but Cronin and Chavas (2019) showed that the separation distance between vortices (in their completely dry case) seems to scale as potential intensity divided by Coriolis parameter.

The use of aerodynamic flux conditions at the lower boundary is consistent with Monin-Obukhov boundary layer similarity theory (see Chapter 4) but are almost never used in theoretical or laboratory studies of Rayleigh-Bénard convection. This explains why the scientific communities that study Rayleigh-Bénard convection and tropical cyclones have remained largely independent.

This work shows that tropical cyclone-like vortices develop over aerodynamically rough boundaries in fluids for which surface thermodynamic disequilibrium is maintained by interior cooling. Phase change of water, though it has strong quantitative effects on vortex structure and intensity, is not necessary and is not the physics that truly distinguishes TC-like vortices from other forms of convection in rotating fluids.

With this background, we next turn to TC-world experiments with full radiation, phase change of water, and some representation of microphysical processes. Most of these have used convection-permitting models in which the convection is explicitly but crudely resolved, and have been run in doubly periodic or spherical domains with fixed or variable Coriolis parameter and specified, mostly constant, SSTs. The models are usually initialized by adding low-amplitude white noise to one or more model variables in a state that is otherwise at rest. Often the model is initially run in a domain that is too small to permit self-aggregation, so as to achieve something like classical RCE with random convection. Then the final state of such a simulation is unfolded into a larger domain (see Robe and Emanuel 1996 for details of this technique) and the integration is carried on in the large domain.

An example of a rotating self-aggregation experiments in a doubly periodic domain with constant Coriolis parameter and SST is the work of Wing et al. (2016). They ran the System for Atmospheric Modeling (SAM; Khairoutdinov and Randall 2003) in a *1536 km x 1536 km* domain with *3 km* horizontal grid spacing and *64* levels, with a fixed, constant SST of *305 K* and a Coriolis parameter of $5 \times 10^{-5}\ s^{-1}$. They initialized their simulations with average fields that resulted from averaging the last 20 days of a simulation over a much smaller *288 km x 288 km* domain, adding very small amplitude ($< 0.1\ K$) thermal noise to the lowest 5 levels of the domain. They ran 5 ensemble members, differing only in the particular realization of random noise in the initial conditions. Tropical cyclones spontaneously emerge in these simulations. The time evolution of the domain-maximum horizontal wind speed is shown in Figure 14.17 for each of the 5 ensemble members.

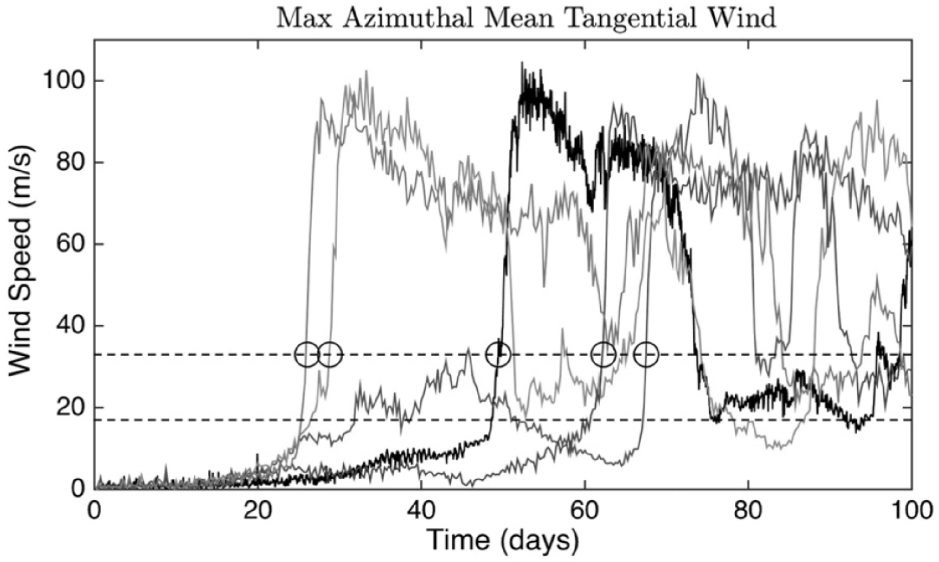


Figure 14.17: Evolution with time of the domain-maximum horizontal wind (ms^{-1}) for each of the 5 ensemble members in the rotating self-aggregation experiments performed by Wing et al. (2016).

After a gestation period ranging from about 25 to about 65 days, rapid cyclogenesis occurs. This time scale, which is longer than the observed lifetime of moist tropical cyclones, is about the time scale for the moist static energy to increase to saturation in a column in which outgoing longwave radiation has been mostly shut down by dense high clouds but surface enthalpy flux continues at its average rate. Other similar studies found similar gestation time scales and sensitivity to initial noise. This simulation resulted in only a single tropical cyclone in the doubly periodic domain, for the particular value of the Coriolis parameter used. Note in Figure 14.17 that the cyclones often collapse after 20 days or so, perhaps because they are being confined to a domain that is too small; this behavior is not observed in simulations with many tropical cyclones.

As in observational studies and many other simulations of genesis, a cold-core mid-level cyclone develops ahead of the surface cyclogenesis (Carstens and Wing 2020). Unlike in the simulations by Davis (2015), the mid-level vortex forms as a single entity in the Wing et al. (2016) simulations, again possibly because of the limited domain size.

How does spontaneous genesis occur in TC-World models? It turns out that the process is pretty much the same as occurs in the self-aggregation of non-rotating convection, described in detail in Chapter 3 section 3.4, but with a larger role for surface enthalpy flux feedback. Recall that we diagnosed the physics of self-aggregation using an equation for the variance of column-integrated moist static energy, (3.85), which we repeat here for convenience:

$$\frac{1}{2}\frac{\partial \overline{\hat{h}'^2}}{\partial t} = \overline{\hat{h}' F_{r0}'} - \overline{\hat{h}' F_{rt}'} + \overline{\hat{h}' F_{t0}'} + \overline{G\hat{\omega}'\hat{h}'}, \tag{14.6}$$

where $\hat{h}'$ is the fluctuation of column-integrated moist static energy, F_{r0} and F_{rt} are the upward radiative fluxes at the surface and tropopause, respectively, F_{t0} is the surface turbulent enthalpy flux (equivalent to the moist static energy flux), and G is the gross moist stability, which we assume to be positive. The overbars represent horizontal averages over distances that span many clouds. For the variance of column-integrated moist static energy to increase, the terms on the right side of (1.46) must sum to a positive result. Wing et al. (2016) calculated the terms in (14.6) in their simulations, dividing the radiative terms into shortwave and longwave components, with the results shown here in Figure 14.18.

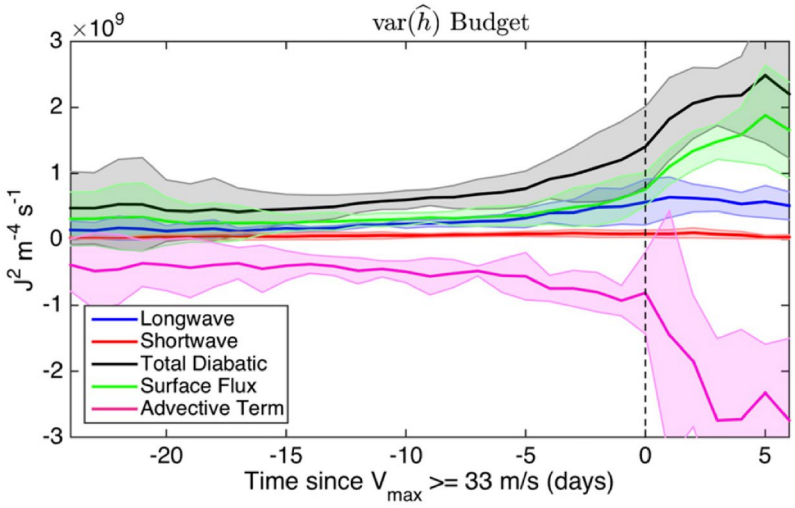


Figure 14.18: Evolution over time of the terms on the right side of (14.6). All 5 ensemble members are used, and composited with respect to the time that the wind speed first reaches hurricane intensity (33 m s^{-1}; see circles in Figure 14.17). The curves represent the ensemble mean and the shading indicates the standard deviation. The sum of the radiative and surface enthalpy flux terms is given in black; these are divided into longwave radiation (blue), shortwave radiation (red), and surface enthalpy flux (green). The advective term, a more complete version of the last term in (14.6), is shown in magenta.

It is clear from the variance budget that variance growth is owing to surface enthalpy fluxes and longwave radiation, which is likely owing mostly to high, dense clouds, as has been shown to be the case with non-rotating, self-aggregating convection. Late in the process, the surface flux feedback becomes dominant, as expected, but the longwave feedback remains positive and substantial. (In the non-rotating case, the surface flux feedback ultimately becomes negative.)

As shown by Muller and Romps (2018), when both radiative and surface flux feedbacks are suppressed, no spontaneous genesis occurs. Suppressing just the radiative feedback allows genesis to occur, but the timing is delayed and the cyclones do not become quite as strong. Cyclones can develop with radiative feedbacks only, but intensification is slower and the final intensity is considerably reduced.

Radiative feedback has been shown to accelerate the intensification phase of tropical cyclones in both TC-World simulations (Wing et al. 2016; Muller and Romps 2018) and in some real-world simulations using WRF (Ruppert et al. 2020).

Based on the non-convection-permitting, large-domain (*20,000 km x 20,000 km*) simulations by Held and Zhou (2008), which demonstrated that tropical cyclones become smaller and more densely packed as the Coriolis parameter is increased, Khairoutdinov and Emanuel (2013) performed convection-permitting simulations in a smaller domain but with an artificially inflated value of the Coriolis parameter. Specifically, they used a value of f of $5\times10^{-4}s^{-1}$ in a domain that was *2500 km* by *2500 km* with *3 km* horizontal grid spacing and *64* vertical levels. Their simulations were initialized from vertical soundings taken from a 100-day integration of a non-rotating version of the same model (SAM; Khairoutdinov and Randall 2003) in a much smaller (*96 km x 96 km*) domain, too small to allow for aggregation. This initialization procedure was applied for each SST used in the experiments. Small amplitude thermal noise was then added to the fields to initiate convection in the larger domain.

In these simulations, genesis occurred spontaneously after more than 60 days, a bit longer than the average in other studies of rotating self-aggregation (as shown in Figure 14.17). With the much larger Coriolis parameter, between 8 and 26 cyclones fill the doubly periodic domain, depending on other externally imposed conditions such as SST.

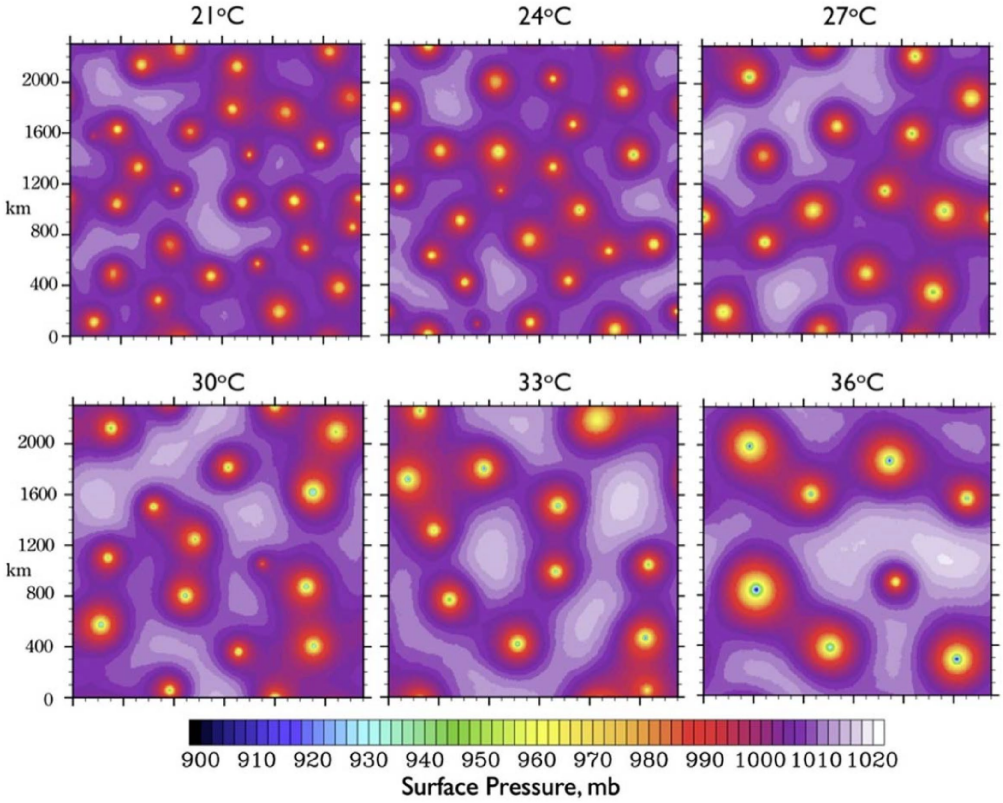


Figure 14.19: Snapshots of surface pressure taken after statistical equilibrium is reached in six experiments that differ only in the imposed sea surface temperature, as shown at the top of each panel.

Figure 14.19 shows snapshots of surface pressure for six different values of the imposed SST. These snapshots were made after statistical equilibrium was reached in each simulation. The intensity, as measured by minimum central pressure, and the diameter of the cyclones clearly increase with SST, and their number falls off correspondingly.

Enough simulations were performed to develop and evaluate scaling laws for the characteristics of tropical cyclones in the statistical equilibrium state. First, there is a clear relationship between cyclone diameter and the number of cyclones, N, since the cyclones seem to be densely packed:

$$N \simeq \frac{L^2}{D^2}, \tag{14.7}$$

where L is the length of the (square) domain and D is a characteristic distance between cyclone centers. Because it is easy to count cyclones, we will take (14.7) to be the working definition of D.

We next assume that in TC World, in contrast to the real world, the main source of irreversible entropy production is dissipation of kinetic energy in the boundary layer. If we define r_0 as a characteristic radius inside which most of the dissipation occurs, then the domain-average dissipation in the Carnot limit is

$$C_D V_p^3 \rho_a \frac{r_o^2}{D^2} \sim \varepsilon \dot{Q} = \varepsilon (F_{rt} - F_{r0}), \tag{14.8}$$

where C_D is the drag coefficient, ρ_a is the surface air density, V_p is the potential intensity (which we assume is the characteristic azimuthal velocity scale for TCs), F_{rt} is the net upward radiative flux at the top of the troposphere, F_{r0} is the net upward radiative flux at the surface, and ε is the modified thermodynamic efficiency defined

$$\varepsilon \equiv \frac{T_s - T_{cool}}{T_{cool}}, \tag{14.9}$$

where T_s is the surface air temperature and T_{cool} is now the inverse of the mean inverse temperature at which radiative cooling occurs in the system. The right-hand side is the Carnot production of mechanical energy, accounting also for dissipative heating, and the left-hand side is the average dissipation of kinetic energy.

We will test the hypothesis that the radius r_o inside of which most of the dissipation occurs scales linearly with the cyclone separation scale D:

$$r_o = \gamma D, \tag{14.10}$$

where γ is a constant which, on geometric grounds we assume obeys $0 < \gamma < 1$. If this turns out to be true, then (14.8) can be written

$$C_D \rho_a V_p^3 \gamma^2 \sim \varepsilon \dot{Q} = \varepsilon (F_{rt} - F_{r0}), \tag{14.11}$$

For the six simulations performed at different SSTs by Khairoutdinov and Emanuel (2013), Table 14.1 lists the equilibrium number of cyclones, the potential intensity based on domain mean atmospheric temperature and SST, the domain-mean kinetic energy of the lowest model level winds, the net radiative cooling, $F_{rt} - F_{r0}$, and the value of γ calculated from (14.11) assuming that $C_D = 1.5 \times 10^{-3}$, $\rho_a = 1\ Kg\ m^{-3}$, and $\varepsilon = 0.3$. (We will discuss the parameter α in due course.)

Table 14.1: TC World Statistics

SST (ºC)	21	24	27	30	33	36
# Cyclones	26	22	15	14	12	8
V_p (m s^{-1})	52.6	55.7	58.6	61.5	62.6	63.8
KE (J m^{-2})	0.34	0.38	0.43	0.45	0.46	0.49
$F_{rt} - F_{r0}\ \left(W\,m^{-2}\right)$	92.7	102.7	112.7	120.1	128.1	136.8
γ	0.36	0.34	0.33	0.32	0.32	0.32
$\alpha\ \left(\times 10^3\right)$	0.84	0.83	0.92	0.87	0.86	0.96

Note that γ is nearly constant, with a mean value of about 0.334. Using this mean value in (14.11) yields a regression coefficient of 0.99 between the predicted and actual values of the potential intensity. Also, as expected, the surface kinetic energy is highly correlated with the square of the potential intensity, also with a regression coefficient of 0.99.

Consequently, (14.11) tells us that the strength of the cyclones in TC World is proportional to the cube root of the rate of radiative cooling of the system, and is inversely proportional to the cube root of the drag coefficient. The horizontal scale of the cyclones seems to vary linearly with the separation distance between cyclones, but that does not tell us what this separation distance is. To estimate that, we bring in two further constraints. The first constraint is that the mean rate of latent heat release in the rain areas of the cyclones must equal to the rate of radiative cooling:

$$ML_v \Delta q^* = F_{rt} - F_{r0}, \tag{14.12}$$

where M is the updraft mass flux, L_v is the latent heat of vaporization, and Δq^* is the difference between the saturation specific humidity at cloud base and at cloud top. We next make two approximations to (14.12). First, we assume that q^* at cloud top is essentially zero, given the very cold temperature at the tops of the cyclones. We also assume that q^* at cloud base equals the actual value of specific humidity in the boundary layer, q_b, and in turn assume that the latter is the saturation specific humidity at SST and mean surface pressure p_0, multiplied by a near-surface relative humidity, which we take to be 0.8. Taken together, we are assuming that

$$\Delta q^* \simeq q_b \simeq 0.8 q^*(SST, p_0). \tag{14.13}$$

We can write the updraft mass flux as the product of air density ρ, updraft velocity w, and the fractional area σ occupied by the rainy cores of the cyclone:

$$M = \rho \sigma w. \tag{14.14}$$

From the Ekman radial velocity equation, (9.83), together with mass continuity, we can show that the vertical velocity in steady axisymmetric cyclone scales as

$$w \sim C_D V_p. \tag{14.15}$$

Combining (14.12) – (14.15) yields an expression for the fractional area:

$$\sigma \simeq \frac{F_{rt} - F_{r0}}{\rho C_D V_p L_v q_b}. \tag{14.16}$$

We next eliminate $F_{rt} - F_{r0}$ between (14.16) and (14.11):

$$\sigma = \frac{\gamma^2 V_p^2}{\varepsilon L_v q_b} \tag{14.17}$$

Note that although the potential intensity V_p increases with SST, it does not do so as rapidly as q^* and so (14.17) predicts that the fraction area occupied by the rainy cores of the cyclones decreases with increasing SST.

Finally, we represent the fractional area covered by the rainy areas as

$$\sigma = \frac{r_r^2}{D^2}, \tag{14.18}$$

where r_r is the radius of the rain area. We next make an ansatz which rests on rather scant evidence, which is that the inner core radius r_r equals $\alpha V_p / f$. where α is a numerically small nondimensional number. Combining this with (14.18) and (14.17) yields

$$D^2 = \frac{\alpha^2 \varepsilon L_v q_b}{f^2 \gamma^2}. \tag{14.19}$$

Using (14.7), this also gives us an estimate of the storm count:

$$N = \frac{L^2 f^2 \gamma^2}{\alpha^2 \varepsilon L_v q_b}. \tag{14.20}$$

Using the storm counts in Table 14.1 together with the specified domain length L, Coriolis parameter f, the mean value of γ form Table 14.1, and q_b from (14.13) and the SSTs in Table (14.1), we use (14.20) to calculate the value of α listed in Table 14.1. Although α does vary from 0.83×10^{-3} to 0.96×10^{-3}, it does not vary systematically with SST.

If we use the six-experiment mean values of α and γ from Table 14.1, we can summarize what we believe to be the correct scaling laws for $f-$ ***plane*** TC Worlds in statistical equilibrium:

$$V_p = \left(9 \frac{\varepsilon (F_{rt} - F_{r0})}{C_D \rho_a} \right)^{1/3}, \tag{14.21}$$

$$\sigma = 0.14 \frac{V_p^2}{\varepsilon L_v q_{SST}^*}, \tag{14.22}$$

$$D = 2.5 \frac{\sqrt{\varepsilon L_v q_{SST}^*}}{f}, \tag{14.23}$$

$$N = 0.18 \frac{L^2 f^2}{\varepsilon L_v q_{SST}^*}, \tag{14.24}$$

where q_{SST}^* is the saturation specific humidity at sea surface temperature and pressure.

We note that a somewhat better fit between predicted and simulated potential intensity can be obtained by modifying (14.21) to include a threshold:

$$V_p = \left(13.5 \frac{\varepsilon (F_{rt} - F_{r0})}{C_D \rho_a} - 1 \times 10^5 m^3 s^{-3} \right)^{1/3}, \tag{14.25}$$

with the provision that V_p cannot be negative. This may imply that a threshold level of thermodynamic disequilibrium between the surface and the atmosphere is required for organization into tropical cyclones. This is consistent with the work of Ilya Prigogine (Nicolis and Prigogine 1977), who argued that a finite degree of thermodynamic disequilibrium is necessary for self-organized structures to emerge, and is also consistent with some empirical genesis indices, such as that given by (14.4), which have a threshold potential intensity for tropical cyclone formation. Likewise, in their dry simulations, Velez-Pardo and Cronin (2023) found that large-scale persistent vortices in the case of a lower aerodynamic boundary condition only develop when a nondimensional measure of the thermal forcing (the convective Rossby Number) is large enough.

In their simulations using a global model with 25 km grid spacing but with constant SST and Coriolis parameter, Reed and Chavas (2015) found that the outer structure of their simulated vortices closely matches the prediction of the theoretical outer wind structure predicted by (9.84). A dimensional analysis of (9.84) reveals that if we again assume that the radii of maximum winds scale as $V_p / f,$ then the outer radius, which should define the storm spacing in equilibrium must obey

$$D = \frac{V_p}{f} F\left(\frac{C_D V_p}{w_e} \right), \tag{14.26}$$

where F is some nondimensional function. Moreover, from clear sky energy balance,

$$w_e = \frac{F_{rt} - F_{r0}}{\rho L_v q_b}. \tag{14.27}$$

It can easily be shown that if F is just the square root function, then (14.26) and (14.27) lead to (14.23). Indeed, it can be shown that the solution to (9.84) in the limit of large $C_D V_p / w_e$ gives an F that asymptotically approaches a square root function. Thus, provided $C_D V_p / w_e$ is large, the spacing of cyclones in TC-World is also consistent with the outer wind field given by (9.84) for a radius of maximum winds that scales at V_p / f.

The scaling law (14.25) tells us that cyclone intensity increases slowly with temperature and eventually saturates in RCE. It saturates because there is an upper bound on $F_{rt} - F_{r0}$ in RCE, for which F_{rt} vanishes (no net radiative flux at the top of the atmosphere) and, in very infrared-opaque atmospheres, F_{r0} is effectively bounded by the surface solar flux, since strange things happen when the net surface infrared flux becomes negative.

The fractional area covered by the rainy cores of the cyclone decreases with increasing temperature, according to (14.22), because the saturation specific humidity increases faster than the square of the potential intensity.

Finally, the number of cyclones per unit area drops rapidly with increasing temperature, according to (14.24), consistent with Figure 14.19. The storms become larger, however, and the net power dissipation increases slowly, along with the net radiative cooling of the atmosphere, as expressed by (14.11). In the real world, damage from tropical cyclones tends to increase with roughly the cube of the wind speed, and of course large storms will cause more damage than

smaller ones, all other things being equal. Consequently, if there were structures to damage in TC World, that damage would increase slowly with temperature.

In statistical equilibrium, the domain-average rain rate is constrained so that the net latent heat release balances the radiative cooling of the cloud layer. As shown in Table 14.1, this cooling rate, proportional to $F_{rt} - F_{r0}$, increases slowly with SST. On the other hand, the peak rainfall rate scales as wq_b, where w is a characteristic upward velocity in the storm core. Using the Ekman scaling (14.15), this means that the peak rain rates should scale as $C_D V_p q_b$. Peak rain rates rise even faster than Clausius-Clapeyron, because V_p also increases with SST.

The rapid increase of the maximum rain rate together with the slow increase of domain-average rainfall with SST is consistent with the decrease of the fractional area covered by the storm cores implied by (14.22) and observed in TC-World experiments.

Several variants on the original rotating RCE experiments have emerged in recent years. While most such experiments continue to be performed with fixed SSTs, a few have replaced those with shallow slab oceans (e.g. Zhou et al. 2017; Viale and Merlis 2017). In these simulations, the water underneath the eyewall cools in response to cloud shading and enhanced surface enthalpy flux because of strong winds. Eventually a new equilibrium is reached with less sunlight warming the surface under the cyclones' cores, and, owing to locally depressed SST, less enthalpy flux into the atmosphere. Yet the balance between Carnot cycle energy generation and surface kinetic energy dissipation, embodied in (14.11), still seems to hold, and because the net radiative cooling of the system is not very different from the fixed SST case, the power dissipation is also roughly the same.

Several simulations have been performed on constant SST aqua-planets (Shi and Bretherton 2014; Reed and Chavas 2015; Merlis et al. 2016; Chavas and Reed 2019; Walsh et al. 2020; Vu et al. 2021), and most of these have allowed the Coriolis parameter to vary with latitude (Shi and Bretherton 2014; Merlis et al. 2016; Chavas and Reed 2019; Walsh et al. 2020; Vu et al. 2021). The upshot is that β drift (see Chapter 12 section 12.1.3) causes cyclones to drift westward and toward higher latitudes. Shi and Bretherton (2014) showed that while these cyclones last for 60-120 days (unlike the doubly periodic f-plane where they last as long as the simulation runs), they eventually break up near the poles, perhaps owing to baroclinic instability of the vortices themselves. (Because they are slow to adjust to the poleward-decreasing Coriolis parameter, they arrive at higher latitude with sizes somewhat larger than the local deformation radius.) The transport of momentum by these poleward-drifting cyclones leads to a weak Hadley-like circulation even though there are no horizontal temperature gradients.

In their doubly-periodic $f-$ plane simulations, Carstens and Wing (2020, 2022) showed that tropical cyclones do not develop if the Coriolis parameter is too small, though that might be partially an artifact of using a domain size too small to accommodate the natural size of tropical cyclones when the Coriolis parameter is small. But on the sphere, it is quite possible that β limits tropical cyclone formation near the equator. Chavas and Reed (2019) present persuasive evidence that tropical cyclones have difficulty developing if their diameter exceeds the Rhines scale (Rhines 1975), which is $\sqrt{U/\beta}$, where U is a characteristic velocity of the outer region of a tropical cyclone. If we assume that equilibrium tropical cyclone size varies inversely with Coriolis parameter, as indicated by (14.23), then this scale will exceed the Rhines scale

sufficiently close to the equator. Chavas and Reed argued that the maximum genesis rate will occur around the latitude where the two scales are roughly equal.

In doubly periodic $f-$ plane experiments, and to some extent in constant SST simulations on aquaplanet spheres with variable Coriolis parameter, the number of tropical cyclones at any given time is a packing problem: How many cyclones of a given diameter (say, as given by (14.23)) will fit in the domain? But when one moves away from such idealizations toward more real-world conditions, for example, models with pole-to-equator SST gradients, the problem of cyclone frequency becomes more complex. For one thing, one has to deal with environmental wind shear, not just the shears associated with the cyclones themselves. The frequency of tropical cyclones ceases to become a function of SST alone and depends on the whole pattern of SST.

Merlis et al. (2013), by imposing a hemispherically asymmetric ocean heat flux in a climate model with 50-km horizontal grid spacing, showed that the frequency of tropical cyclogenesis increases as the ITCZ is displaced further poleward. This demonstrates that for the same overall climate forcing, regionality of climate, in this case imposed through ocean heat fluxes, can strongly affect the global frequency of tropical cyclones. While, as we have seen, uniform warming of a constant SST domain always results in fewer (but larger) cyclones, Merlis et al. show that warming accompanied by a shift in ocean heat transport can lead to increases of cyclone frequency. Consequently, the spatial variability of climate is important, not just the global mean state.

We can demonstrate the importance of such processes as ocean heat transport on just the potential intensity, putting aside wind shear and atmospheric humidity for the moment.

Consider first the heat budget of a small patch of ocean mixed layer, of depth h, as illustrated in Figure 14.20.

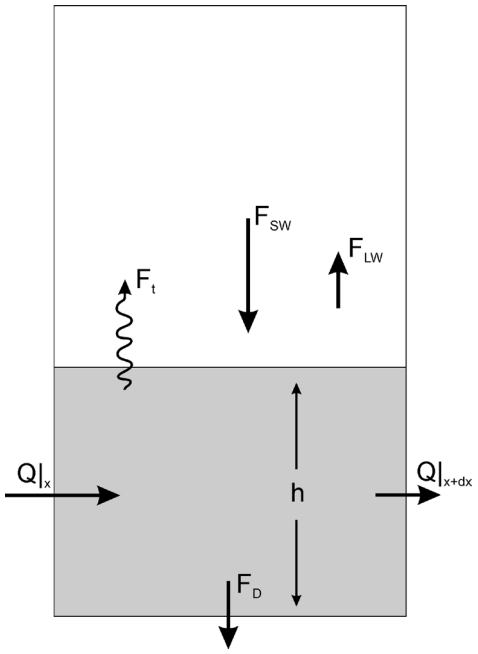


Figure 14.20: Heat budget of the ocean mixed layer (shading). The fluxes per unit area of shortwave and longwave radiation, turbulent enthalpy, and ocean mixed layer entrainment are denoted by F_{SW}, F_{LW}, F_t, and F_D, respectively. Vertically integrated lateral heat fluxes at the left and right sides of the box are $Q|_x$ and $Q|_{x+dx}$, respectively.

Assuming that the heat content of the patch of ocean mixed layer is not changing over time, the heat budget demands that

$$F_t = F_{SW} - F_{LW} - F_D - \nabla_2 \bullet \mathbf{Q} = \rho_a C_k |\mathbf{V}_a| \left(k_0^* - k_a\right), \tag{14.28}$$

where ∇_2 is the horizontal gradient operator, $\mathbf{Q}$ is the vertically integrated ocean horizontal heat flux vector, and we have used the neutral aerodynamic flux formula for F_t, with ρ_a the near-surface air density, C_k the enthalpy flux transfer coefficient, $|\mathbf{V}_a|$ the near-surface wind speed, k_0^* the saturation enthalpy of the air at SST and surface pressure, and k_a is the enthalpy of near-surface air. The fluxes per unit area of shortwave and longwave radiation, turbulent enthalpy, and ocean mixed layer entrainment are denoted by F_{SW}, F_{LW}, F_t, and F_D, respectively.

Now potential intensity is given by (9.26):

$$V_p^2 = \frac{T_b - T_o}{T_o} \frac{C_k}{C_D} \left(k_0^* - k_a\right). \tag{14.29}$$

Eliminating $C_k \left(k_0^* - k_a\right)$ between (14.29) and (14.28) yields:

$$V_p^2 = \frac{T_b - T_o}{T_o} \left(\frac{F_{SW} - F_{LW} - F_D - \nabla_2 \bullet Q}{\rho_a C_D |\mathbf{V}_a|} \right). \tag{14.30}$$

This shows that in equilibrium conditions, potential intensity is increased by increasing surface insolation, decreasing net longwave flux from the surface, decreasing detrainment of heat into the deeper ocean, increasing convergence of the vertically integrated ocean lateral heat flux, and decreasing near-surface wind speed. The assumption of a steady state is not valid over a seasonal cycle but is not too bad when averaged over a few years.

Changing turbulent heat flux, which in the tropics is carried mostly as a latent heat flux, will also influence the humidity of the lower and middle troposphere, which is an important contributor to genesis potential indices. Since, in the tropics, the convergence of the ocean heat flux is of the same magnitude as the net surface radiative flux (Trenberth and Caron 2001), it can have a strong influence on potential intensity, and through its modification of surface latent heat flux and atmospheric circulation, the humidity of the lower and middle troposphere, both of which affect tropical cyclone genesis frequency as estimated by indices such as (14.4).

Large-scale circulations in both the atmosphere and the ocean play a strong role in the climatology of tropical cyclones, circumscribing regions where such storms can form, even within the tropics, as illustrated by Figure 14.13. Perhaps the overall frequency of tropical cyclones is just a packing problem limited to these regions This hypothesis was tested by Hoogewind et al. (2020), but they found that the packing limit is about an order-of-magnitude higher than the observed frequency of cyclones, suggesting that other factors, such as the frequency of suitable triggering disturbances, may be playing a role.

In summary. the TC-World experiments have shown us that in idealized worlds with no SST gradients and therefore little or no external wind shear, tropical cyclones will develop and, unless they are too close to the equator, fill up the available space according to their diameters. But unless seeded with external triggering disturbances, the time to genesis is many tens of days – too long to develop in the real world, where disturbances would typically travel out of the area of enough genesis potential in less time than that. And, in the real world, the presence of external wind shear and the space time variability of it and other large-scale conditions (as measured by, e.g., genesis potential indices) and of triggering disturbances, strongly limits overall genesis rates.

14.2.3: Random seeding experiments

Random seeding was developed as part of an approach to estimating long-term tropical cyclone risk using physically based models. Accurate risk assessment depends, among other things, on having a sufficiently long record of events to make statistically robust estimates of hazards like wind and flooding, encompassing the all-important low-frequency tails of the probability distributions of those hazards. Analysis of damage from many different natural hazards reveals that long-term damage is dominated by events with return periods of a hundred or a few hundred years… rare enough that societies are not well adapted to them, but common enough to contribute to long-term damage. Historical records of the natural phenomena involved are rarely long enough to provide robust estimates of hazards with return periods of even 100 years. Climate change presents the additional complication of making the underlying statistics non-stationary.

On approach to making robust estimates of long-term risk is to numerically simulate many thousands of years of TCs using physically based models. At the same time, accurate numerical simulation of tropical cyclones requires grid spacings on the order of a few kilometers (Rotunno

et al. 2009). It is currently not possible to run even basin-sale models with such fine resolution for the equivalent of thousands of years.

To address this problem, a technique was developed for creating many thousands of tropical cyclone events driven by the relatively coarse output of global models and reanalyses (Emanuel et al. 2008). In the first step, the time-evolving output of such models or reanalyses is seeded randomly in space and time with weak, tropical depression-strength vortices. These are assumed to moved with a weighted vertical mean of the large-scale flow plus a $\beta-$ drift, as described in Chapter 12, section 12.1. Finally, a simple, deterministic coupled ocean-atmosphere TC model in run on each track to determine the evolution of the storm's intensity. The deterministic model, the Coupled Hurricane Intensity Forecasting System (CHIPS; Emanuel 1995), is axisymmetric but contains a parameterization of ventilation optimized to produce skillful real-time intensity forecasts (Emanuel and Rappaport 2000). It is coupled to a simple one-dimensional ocean model that incorporates the physics of storm-induced mixing of cold water to the surface. The axisymmetric model is based on quasi-balance equations written in angular momentum coordinates rather than in physical radius; this yields high spatial resolution in the core, where it is needed, with higher resolution at higher intensity.

The intensity model predicts that the vast majority of seeds perish immediately because they are in unfavorable environments. The survivors are regarded as constituting the climatology of tropical cyclones pertinent to the global model or reanalysis used to drive the procedure. The method is akin to biological natural selection, although it is more selection of those put down in favorable environments than survival of the fittest seeds.

The method has been shown to reproduce, with reasonable accuracy, the observed climatology of tropical cyclones, including their seasonal variation in each ocean basin, the probability distribution of their intensities, and their response to natural climate fluctuations like ENSO. We emphasize here that the method filters any triggering disturbances that might be present in the global model or reanalysis fields, so that the triggering used in the procedure is random, except for a dependence on latitude.

Figure 14.21 shows the results of applying this procedure to a global reanalysis (ERA5; Hersbach et al. 2020) over the years 1980-2020, in terms of the annual average numbers of tropical cyclones in each ocean basin. The rate of random seeding has been chosen to yield a reasonable global average annual number of storms. As with genesis indices, the global number must be assigned; it is not predicted by the technique (although its year-to-year variability is).

The technique performs quite well, with the important exceptions of the southwest Pacific Ocean and the eastern North Pacific region. In the latter case, the random seeding produces less than half the observed annual average number of cyclones. Danso et al. (2022) showed that in the absence of African easterly waves, North Atlantic tropical cyclones form further west than they do with AEWS are present, but at roughly the same rate. . Many eastern North Pacific tropical cyclones also develop from easterly waves (Molinari et al. 1997; Zehnder et al. 1999), but taking them away allows for precious little space for cyclones to develop from weaker noise, compared to the Atlantic case. (The distance from Dakar to Miami is about 6,500 km, whereas eastern Pacific tropical cyclones typically run out of positive genesis potential in about half that distance.) Consequently, it is possible that the frequency and characteristics of triggering disturbances plays a more important role in eastern North Pacific cyclogenesis than it does in other regions.

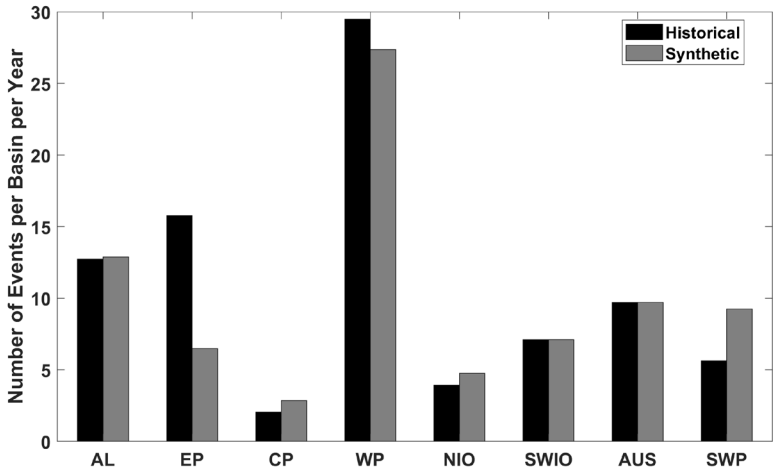


Figure 14.21: Average annual number of tropical cyclones over the period 1980-2020 in observations (IBTrACS; Knapp et al. 2010; black bars) and from random seeding (gray bars). The totals are presented for the North Atlantic (AL), eastern North Pacific (EP), central North Pacific (CP), western North Pacific (WP), North Indian Ocean (NIO), southwest Indian Ocean (SWIO), Australian region (AUS) and southwest Pacific (SWP).

14.3 Synthesis

The TC-World experiments tell us that under highly ideal conditions, tropical cyclones can develop spontaneously in RCE states owing to a combination of WISHE and radiative feedbacks. But the spontaneous emergence of cyclones takes tens of days – too long for cyclogenesis in the real world, where the background airflow would normally take growing disturbances out of favorable environments in less time. Nevertheless, these simulations and experiments conducted under more realistic conditions show that radiative feedbacks can accelerate cyclogenesis and that it is difficult if not impossible to develop tropical cyclones when both feedbacks are absent.

Case studies and numerical experiments under both idealized and real-world conditions show that tropical cyclones always (or almost always) develop out of pre-existing disturbances that originate at least somewhat independently. Field experimental studies and numerical experiments as well as theory, strongly suggest that a prerequisite for tropical cyclone development is the establishment of a nearly saturated mesoscale lower and middle troposphere above the subcloud layer. If such a region retains a nearly moist adiabatic lapse rate, it either must have a greatly elevated moist static energy or else it must be relatively cold compared to surrounding regions (or, more likely, a bit of both; see Figure 14.11). To prevent the cold anomaly from radiating away as gravity waves, it must be rotating and nearly in gradient and hydrostatic balance, which implies a mesocyclone aloft. Mesocyclones are observed along the route to genesis in many field experiments and most numerical simulations.

The experiments by Cronin and Chavas (2019) demonstrated that, for the same potential intensity, moist cyclones generally become stronger than dry cyclones. At the same time, the presence of moisture allows for a depression of mid-tropospheric moist entropy that cannot occur in a dry atmosphere with adiabatic lapse rates. Multiple lines of evidence show that such

a depression inhibits cyclogenesis and, in the presence of shear, weakens mature storms: moisture paradoxically makes storms potentially more intense and less likely. The physics of the inhibition seems to be tied to ventilation of the storm core by low moist entropy air in the lower to middle troposphere, made possible by the dynamic interaction between the TC and environmental wind shear. The nondimensional ventilation parameter, defined by (12.23), is an important parameter both for storm intensification and for mature intensity, and a similar parameter is part of many semi-empirical genesis indices like (14.4).

The early investigators of tropical cyclones, like Riehl (1951) and Dunn (1951), recognized that tropical cyclones occur when a suitable initiating disturbance occurs in a favorable environment. Between then and now we have come to understand more about what constitutes a favorable environment and physically why it is favorable, and also what makes for a suitable initiating disturbance. A remaining question is whether the observed rates of genesis are governed more by the degree of environmental salubriousness or by the frequency and character of initiating disturbances. While we have come to understand fairly well what controls the number of cyclones in idealized TC-Worlds, we also understand that observed genesis rates fall well short of expectations based on such experiments, which have little or no external wind shear.

In section 14.2.1 we reviewed the work of Patricola et al. (2018) and Danso et al. (2022) that suggested that while initiating disturbances help determine when and where genesis occurs, they do not seem to determine whether it occurs. This seems to imply that rates of genesis are more determined by the large-scale environment than by the frequency or character of initiating disturbances. This finding is also consistent with the success of genesis indices. At the same time, the long time scales needed for spontaneous genesis in highly favorable environments suggests that triggers are necessary in the real world. How do we reconcile this seeming paradox?

Here we develop a semi-quantitative conceptual framework for thinking about this problem. This framework is based on a pair of ordinary differential equations developed by the author (Emanuel 2017) for emulating the behavior of a deterministic tropical cyclone intensity model (Emanuel 1995) driven by real-world boundary conditions and used as an aid to forecasting actual storms in real time (Emanuel and Rappaport 2000). The CHIPS model is described briefly in section 14.2.3.

The emulator was designed to obey certain asymptotic limits, and in the absence of shear reduces to (10.23). It was refined and optimized against thousands of CHIPS intensity forecasts. The pair of ordinary differential equations are for two variables: the maximum gradient azimuthal wind speed, V_g, and a non-dimensional moisture variable m that varies between 0 and 1. They are as follows:

$$\frac{dV_g}{dt} = \frac{1}{2}\frac{C_D}{h}\left[\alpha\left(1-\epsilon-\kappa\right)V_p^2 m^3 - \left(1-\left(\epsilon+\alpha\kappa\right)m^3\right)V_g^2\right], \tag{14.31}$$

and

$$\frac{dm}{dt} = \frac{1}{2}\frac{C_D}{h}\left[\left(1-m\right)V_g - S\chi m\right], \tag{14.32}$$

where C_D is the drag coefficient, h is a boundary layer depth, α is an ocean-interaction parameter that varies with upper ocean properties, and various properties of the cyclone, ϵ is the raw thermodynamic efficiency:

$$\epsilon \equiv \frac{T_s - T_o}{T_s}. \tag{14.33}$$

The coefficient κ governs heating by isothermal expansion and is defined

$$\kappa \equiv \frac{\epsilon}{2} \frac{C_k}{C_D} \frac{L_v q_0^*}{R_d T_s}, \tag{14.34}$$

where T_s is the surface temperature, T_o is the tropical cyclone outflow temperature, C_k is the surface enthalpy exchange coefficient, L_v is the latent heat of vaporization, q_0^* is the surface saturation specific humidity at ambient environmental surface pressure, and R_d is the gas constant for dry air.

Finally, S is the magnitude of the 250 hPa – 850 hPa environmental wind shear, V_p is the potential intensity, and χ is a nondimensional measure of the saturation deficit of the lower to middle troposphere:

$$\chi = \frac{h_b - h_m}{h_0^* - h_b}, \tag{14.35}$$

where h_b and h_m are the moist static energies of the boundary layer and middle troposphere, and h_0^* is the saturation moist static energy of air at sea surface temperature and pressure. (Note that the χ factor was set to a constant 2.2 in Emanuel (2017)).

For our narrow purposes here, we neglect all ocean response, which entails setting $\alpha = 1$ in (14.31). While in practice, the parameters (like potential intensity and shear) vary along the tracks of tropical cyclones, we hold them constant here. We also nondimensionalize the equations by substituting nondimensional time and velocity scales:

$$\begin{aligned} t &= \frac{2h}{C_D V_p} \tau, \\ V_g &= V_p V. \end{aligned} \tag{14.36}$$

With these provisions, (14.31) – (14.32) become

$$\frac{dV}{d\tau} = \left(1 - a\right) m^3 - \left(1 - a m^3\right) V^2, \tag{14.37}$$

and

$$\frac{dm}{d\tau} = \left(1 - m\right) V - \Lambda m, \tag{14.38}$$

where

$$a \equiv \epsilon + \kappa, \tag{14.39}$$

and Λ is the ventilation parameter:

$$\Lambda \equiv \frac{S\chi}{V_p}. \tag{14.40}$$

We also hold a constant in this exercise.

We examine the behavior of the system comprised of (14.37) and (14.38). A detailed mathematical analysis of the system has been presented by Slyman et al. (2023).

First, depending on the value of the ventilation parameter Λ, (14.37) and (14.38) have either one real steady-state solution, or three. These are shown for the velocity V by the thick, black curves in Figure 14.22 for the case $a = 0.47$.

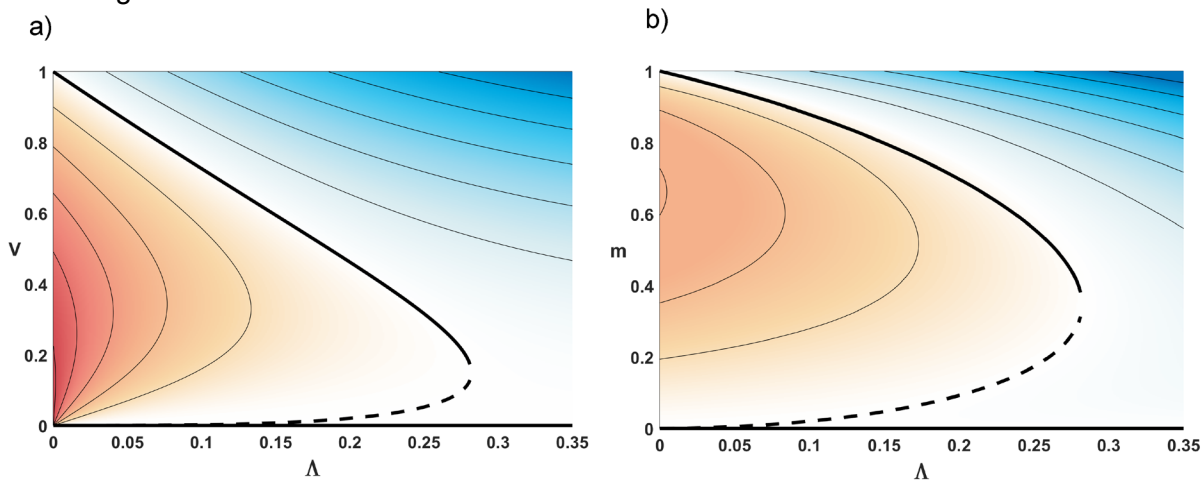


Figure 14.22: Regime diagrams for (14.37) and (14.38). The thick black curves show the steady state solutions for the nondimensional a) wind speed V and b) moisture variable m as functions of the ventilation parameter, Λ. Solid curves show stable states, and the dashed curves are unstable steady states. The colors and thin contours show a) the time rate of change of V for the special case that m is in a quasi-steady-state (nullcline of m), and b) the time rate of change of m for the special case that V is in a quasi-steady-state (nullcline of V) with warm colors denoting growth and cool colors denoting decay.

The state of no cyclone, $V = 0,\ m = 0,$ is always linearly stable, although through much of the range of ventilation parameter, it does not take much positive perturbation of V (but somewhat more of m) to tip the system into unstable growth. The shape of the steady-state solutions in 14.21a somewhat resemble the solutions of the system explored by Brian Tang (Tang 2010; Tang and Emanuel 2010) presented in Chapter 12, section 12.2 and graphed in Figures 12.16 and 12.17, but differ in details.

While the steady state solutions shown by the black curves in Figure 14.22 are universal, we show time rates-of-change of V and m for the special cases that the other variable is quasi-steady, which we obtain by setting the time derivative to zero in either (14.37) or (14.38). (In general, the evolutions of V and m depend on their initial values; the time rates-of-change shown in Figure 14.21 are not universal.)

In the special case of no shear ($\Lambda = 0$), the resting state is unstable and, in general, weak disturbances will grow slowly, depending on the initial values of V and m. In general, the growth of the cyclone is somewhat more sensitive to initial moisture than to initial velocity. This is illustrated by Figure 14.23, which shows the time evolution of wind speed for different values of initial wind speed for an initial moisture $m = 0$ (a), and for initial values of moisture given a weak initial vortex $V = 0.05$ (b).

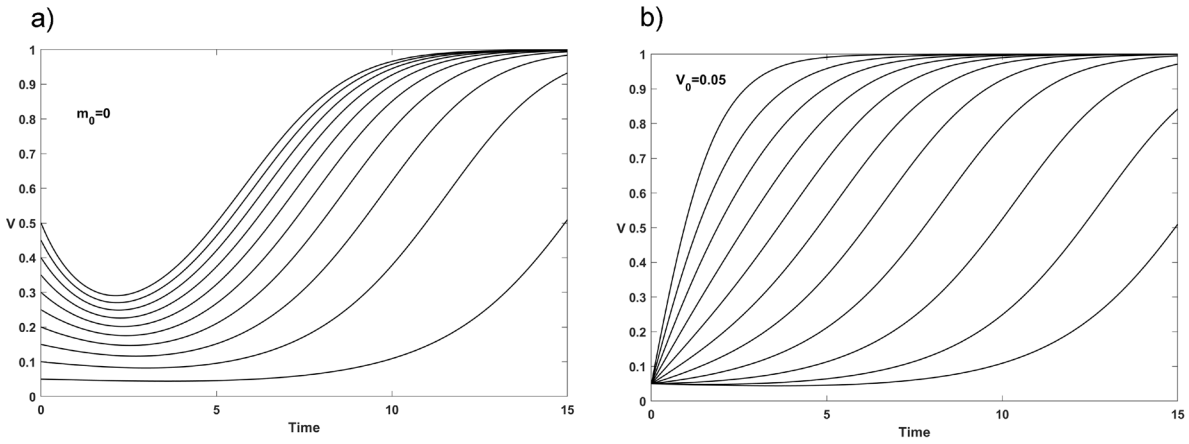


Figure 14.23: Evolution with nondimensional time of the nondimensional wind speed for a) zero initial moisture perturbation and initial winds speeds of from 0.05 to 1 in increments of 0.05, and b) initial wind speed of 0.05 for varying initial moisture perturbation from 0 to 1 in increments of 0.1. The velocity is normalized by potential intensity and one nondimensional time unit is approximately equal to 1 day. The ventilation parameter is zero in these integrations.

In the absence of any initial moisture perturbation, cyclones take a long time to develop. A weak disturbance ($V = 0.05$, Figure 4.23a) takes more than 10 time units (about 10 days) to develop, and strong initial vortices initially decay before finally amplifying, not reaching peak wind speed until about 10 time units. By contrast, even a weak disturbance will amplify quite rapidly with a large initial moisture perturbation (Figure 14.23b).

The effect of moderate ventilation on the evolution of disturbances is shown in Figure 14.24. Given a weak initial vortex, only perturbations with strong moisture anomalies amplify appreciably within 15 time units, and, consistent with Figure 14.22, disturbances amplify to a steady state well below the potential intensity. The initial vortex with the strongest moisture anomaly ($m = 1$) slightly overshoots the steady-state intensity before settling down.

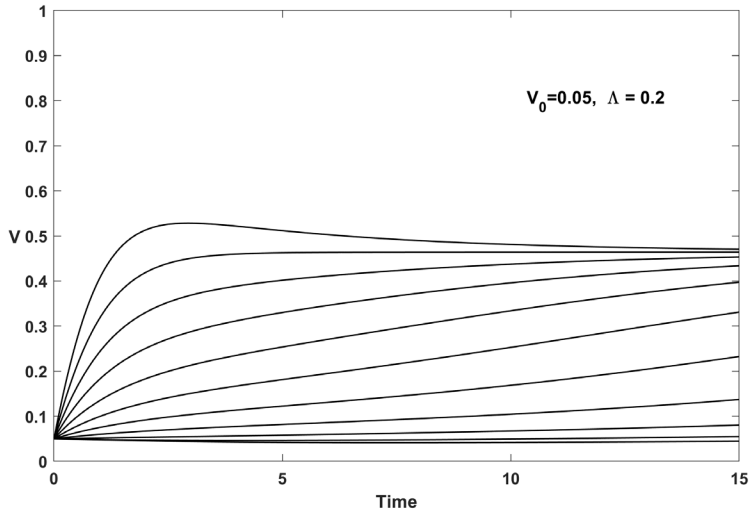


Figure 14.24: As in Figure 14.22 b but for a case with moderate shear ($\Lambda = 0.2$).

Using these results, we propose a conceptual framework for understanding tropical cyclogenesis based on the bifurcation diagram shown in Figure 14.22b. We choose to work in $m - \Lambda$ space because development is more sensitive to the initial values of m than to V. The conceptual framework is illustrated by Figure 14.25. Weather noise in the nondimensional moisture variable m moves the initial point in the diagram vertically, while noise in the nondimensional ventilation parameter moves the state horizontally. Together, they produce a joint probability distribution that is continuously explored by weather noise. The two probability distributions, and their joint probability, vary with location, season, and climate state. When the noise extends far enough into the upper left portion of the bifurcation diagram, and stays there long enough, a tropical cyclone will develop.

One might think of the moisture variable m as being associated with triggering disturbances and the ventilation parameter Λ as associated with the synoptic-scale environment. If the ventilation parameter frequently resides in the region of growth, then genesis frequency will be mostly sensitive to the frequency and amplitude of disturbances in the moisture variable m. On the other hand, if moisture disturbances frequently inhabit the region of strong growth, the overall frequency would be more determined by the probability distribution of Λ. The success of genesis indices and random seeding argues strongly in favor of the latter circumstance, except possibly in the eastern North Pacific region.

Finally, we have tacitly assumed that the probability distributions of triggering and ventilation are independent from each other and from the climatology of the tropical cyclones themselves. Yet there is growing evidence that this not the case, because tropical cyclones may dry out the tropical atmosphere, increasing lower to middle tropospheric saturation deficit (χ) and thereby increasing the ventilation parameter Λ, and may also alter ocean heat transport. These issues are tackled in Chapter 15.

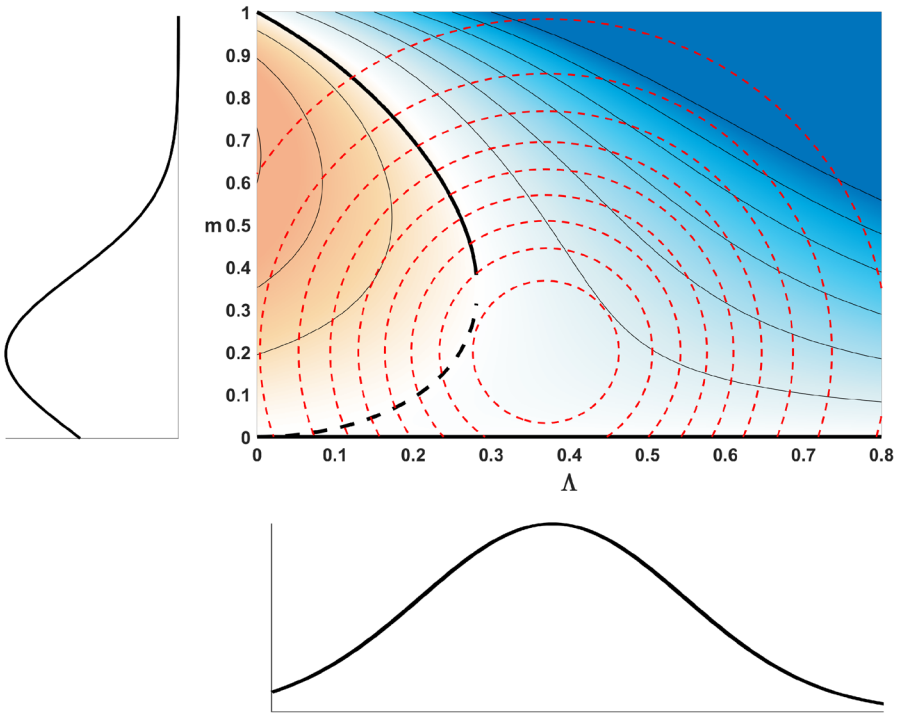


Figure 14.25: Illustrating the genesis conceptual framework based on the bifurcation diagram of Figure 14.21b. The panels at the left and bottom show probability distributions of weather noise in the nondimensional moisture variable m and ventilation parameter Λ, and the dashed red contours in the center panel show their joint probability.

References

Bartels, D. L., and and R. A. Maddox, 1991: Midlevel cyclonic vortices generated by mesoscale convective systems. *Mon. Wea. Rev*, 119.

Bénard, H., 1900: Étude expérimentale des courants de convection dans une nappe liquide. — Régime permanent : tourbillons cellulaires. *J. Phys. Theor. Appl.*, **9**, 513–524, https://doi.org/10.1051/jphystap:019000090051300.

Bergeron, T., 1954: The problem of tropical hurricanes. *Quart. J. Roy. Meteor. Soc.*, **80**, 131–164.

Bister, M., and K. A. Emanuel, 1997: The genesis of Hurricane Guillermo: TEXMEX analyses and a modeling study. *Mon. Wea. Rev.*, **125**, 2662–2682.

Bister, M. H., 1996: Development of tropical cyclones from mesoscale convective systems. Massachusetts Institute of Technology, 112 pp. https://dspace.mit.edu/handle/1721.1/57851.

Bosart, L. F., and F. Sanders, 1981: The Johnstown Flood of July 1977: A Long-Lived Convective System. *Journal of the Atmospheric Sciences*, **38**, 1616–1642, https://doi.org/10.1175/1520-0469(1981)038<1616:tjfoja>2.0.co;2.

Bosart, L. F., and J. A. Bartlo, 1991: Tropical storm formation in a baroclinic environment. *Mon. Wea. Rev.*, **119**, 1979–2013.

Braun, S. A., J. A. Sippel, and D. S. Nolan, 2012: The impact of dry midlevel air on hurricane intensity in idealized simulations with no mean flow. *Journal of the Atmospheric Sciences*, **69**, 236–257, https://doi.org/10.1175/jas-d-10-05007.1.

Bretherton, C. S., P. N. Blossey, and M. F. Khairoutdinov, 2005: An energy-balance analysis of deep convective self-aggregation above uniform SST. *J. Atmos. Sci.*, **62**, 4273–4292.

Bruyère, C. L., G. J. Holland, and E. Towler, 2012: Investigating the use of a genesis potential index for tropical cyclones in the North Atlantic basin. *Journal of Climate*, **25**, 8611–8626, https://doi.org/10.1175/jcli-d-11-00619.1.

Camargo, S., 2013: Global and regional aspects of tropical cyclone activity in the CMIP5 models. *J. Climate*, **26**, 9880–9902.

Camargo, S. J., K. A. Emanuel, and A. H. Sobel, 2007a: Use of a genesis potential index to diagnose ENSO effects on tropical cyclone genesis. *J Climate*, **20**, 4819–4834, https://doi.org/10.1175/Jcli4282.1.

——, A. H. Sobel, A. G. Barnston, and K. A. Emanuel, 2007b: Tropical cyclone genesis potential index in climate models. *Tellus A*, **59**, 428–443, https://doi.org/10.1111/j.1600-0870.2007.00238.x.

——, M. C. Wheeler, and A. H. Sobel, 2009: Diagnosis of the MJO Modulation of Tropical Cyclogenesis Using an Empirical Index. *J Atmos Sci*, **66**, 3061–3074, https://doi.org/10.1175/2009jas3101.1.

Carstens, J. D., and A. A. Wing, 2020: Tropical cyclogenesis from self-aggregated convection in numerical simulations of rotating radiative-convective equilibrium. *Journal of Advances in Modeling Earth Systems*, **12**, e2019MS002020, https://doi.org/10.1029/2019MS002020.

——, and ——, 2022: A spectrum of convective self-aggregation based on background rotation. *Journal of Advances in Modeling Earth Systems*, **14**, e2021MS002860, https://doi.org/10.1029/2021MS002860.

Chandrasekhar, S., 1953: Problems of stability in hydrodynamics and hydromagnetics: George Darwin lecture, delivered by Professor s. Chandrasekhar on 1953 Novemher 13. *Monthly Notices of the Royal Astronomical Society*, **113**, 667–678, https://doi.org/10.1093/mnras/113.6.667.

——, 1961: Hydrodynamic and Hydromagnetic Stability. Oxford University Press, New York. *654 pp.*,.

Charney, J. G., and A. Eliassen, 1964: On the growth of the hurricane depression. *J. Atmos. Sci.*, **21**, 68–75.

Chavas, D. R., and K. A. Reed, 2019: Dynamical aquaplanet experiments with uniform thermal forcing: System dynamics and implications for tropical cyclone genesis and size. *Journal of the Atmospheric Sciences*, **76**, 2257–2274, https://doi.org/10.1175/JAS-D-19-0001.1.

Cronin, T. W., and D. R. Chavas, 2019: Dry and Semidry Tropical Cyclones. *Journal of the Atmospheric Sciences*, **76**, 2193–2212, https://doi.org/10.1175/jas-d-18-0357.1.

Danso, D. K., C. M. Patricola, and E. Bercos-Hickey, 2022: Influence of African easterly wave suppression on Atlantic tropical cyclone activity in a convection-permitting model. *Geophysical Research Letters*, **49**, e2022GL100590, https://doi.org/10.1029/2022GL100590.

Davis, C. A., 2015: The formation of moist vortices and tropical cyclones in idealized simulations. *Journal of the Atmospheric Sciences*, **72**, 3499–3516, https://doi.org/10.1175/JAS-D-15-0027.1.

Dunkerton, T. J., M. T. Montgomery, and Z. Wang, 2008: Tropical cyclogenesis in a tropical wave critical layer: easterly waves. *Atmospheric Chemistry and Physics Discussions*, **8**, 11149–11292.

Dunn, G. E., 1940: Cyclogenesis in the tropical Atlantic. *Bulletin of the American Meteorological Society*, **21**, 215–229.

——, 1951: Tropical Cyclones. *Compendium of Meteorology*, T.F. Malone, Ed., American Meteorological Society, 887–901.

Emanuel, K., 2010: Tropical cyclone activity downscaled from NOAA-CIRES reanalysis, 1908-1958. *J. Adv. Model. Earth Sys.*, **2**, 1–12.

——, 2013: Downscaling CMIP5 climate models shows increased tropical cyclone activity over the 21st century. *Proc. Nat. Acad. Sci.*, **110**, 12219–12224, https://doi.org/10.1073/pnas.1301293110.

——, 2017: A fast intensity simulator for tropical cyclone risk analysis. *Nat. Hazards*, https://doi.org/10.1007/s11069-017-2890-7.

Emanuel, K., 2020: Response of Global Tropical Cyclone Activity to Increasing CO2 : Results from Downscaling CMIP6 Models. *Journal of Climate*, **34**, 57–70, https://doi.org/10.1175/jcli-d-20-0367.1.

Emanuel, K., and E. Rappaport, 2000: Forecast skill of a simplified hurricane intensity prediction model. Preprints of the 24th Conf. Hurricanes and Trop. Meteor., Ft. Lauderdale, FL, Amer. Meteor. Soc., Boston, 236–237.

——, R. Sundararajan, and J. Williams, 2008: Hurricanes and global warming: Results from downscaling IPCC AR4 simulations. *Bull. Amer. Meteor. Soc.*, **89**, 347–367.

Emanuel, K. A., 1989: The Finite-Amplitude Nature of Tropical Cyclogenesis. *Journal of the Atmospheric Sciences*, **46**, 3431–3456, https://doi.org/10.1175/1520-0469(1989)046<3431:tfanot>2.0.co;2.

Emanuel, K. A., 1994: *Atmospheric Convection*. Oxford Univ. Press, New York, 580 pp.

——, 1995: The behavior of a simple hurricane model using a convective scheme based on subcloud-layer entropy equilibrium. *J. Atmos. Sci.*, **52**, 3959–3968.

——, and D. S. Nolan, 2004: Tropical cyclone activity and the global climate system. *26th AMS Conference on Hurricanes and Tropical Meteorology*,.

Espy, J. P., 1841: *The Philosophy of Storms*. Little and Brown,.

Gray, W. M., 1975: *Tropical cyclone genesis*.

——, 1979: Hurricanes: Their formation, structure, and likely role in the tropical circulation. *Meteorology over the tropical oceans*, D.B. Shaw, Ed., Roy. Meteor. Soc., 155–218.

Guervilly, C., D. W. Hughes, and C. A. Jones, 2014: Large-scale vortices in rapidly rotating Rayleigh–Bénard convection. *Journal of Fluid Mechanics*, **758**, 407–435, https://doi.org/10.1017/jfm.2014.542.

von Hann, J., 1901: *Lehrbuch der Meteorologie (Textbook of Meteorology)*. C. H. Tauchnitz, 847 pp.

Hawkins, H. F., and D. T. Rubsam, 1968: Hurricane Hilda, 1964. I: Genesis, as revealed by satellite photographs, conventional and aircraft data. *Mon. Wea. Rev.*, **96**, 428–452.

Held, I., and M. Zhao, 2008: Horizontally homogeneous rotating radiative-convective equilibria at GCM resolution. *J. Atmos. Sci.*, **65**, 2003–2013.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hoogewind, K. A., D. R. Chavas, B. A. Schenkel, and M. E. O'Neill, 2020: Exploring controls on tropical cyclone count through the geography of environmental favorability. *Journal of Climate*, **33**, 1725–1745, https://doi.org/10.1175/JCLI-D-18-0862.1.

Hsieh, T.-L., G. A. Vecchi, W. Yang, I. M. Held, and S. T. Garner, 2020: Large-scale control on the frequency of tropical cyclones and seeds: a consistent relationship across a hierarchy of global atmospheric models. *Climate Dynamics*, **55**, 3177–3196, https://doi.org/10.1007/s00382-020-05446-5.

Khairoutdinov, M. F., and D. A. Randall, 2003: Cloud resolving modeling of the ARM summer 1997 IOP: Model formulation, results, uncertainties and sensitivities. *J. Atmos. Sci.*, **60**, 607–625.

——, and K. Emanuel, 2013: Rotating radiative-convective equilibrium simulated by a cloud-resolving model. *J. Adv. Model. Earth Sys.*, **5**, 816–825.

Kilroy, G., R. K. Smith, and M. T. D.-:10. 1002/qj. 2934 Montgomery, 2017: A unified view of tropical cyclogenesis and intensification. *Quarterly Journal of the Royal Meteorological Society*, **143**, 450–462.

Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, 2010: The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone best track data. *Bull. Amer. Meteor. Soc.*, **91**, 363–376.

Korty, R. L., S. J. Camargo, and J. Galewsky, 2012a: Tropical cyclone genesis factors in simulations of the last glacial maximum. *Journal of Climate*, **25**, 4348–4365, https://doi.org/10.1175/jcli-d-11-00517.1.

——, ——, and ——, 2012b: Variations in tropical cyclone genesis factors in simulations of the Holocene epoch. *Journal of Climate*, **25**, 8196–8211, https://doi.org/10.1175/jcli-d-12-00033.1.

Laloyaux, P., and Coauthors, 2018: CERA-20C: A Coupled Reanalysis of the Twentieth Century. *Journal of Advances in Modeling Earth Systems*, **10**, 1172–1195, https://doi.org/10.1029/2018MS001273.

Merlis, T. M., M. Zhao, and I. M. Held, 2013: The sensitivity of hurricane frequency to ITCZ changes and radiatively forced warming in aquaplanet simulations. *Geophysical Research Letters*, **40**, 4109–4114, https://doi.org/10.1002/grl.50680.

——, W. Zhou, I. M. Held, and M. Zhao, 2016: Surface temperature dependence of tropical cyclone-permitting simulations in a spherical model with uniform thermal forcing. *Geophysical Research Letters*, **43**, 2859–2865, https://doi.org/10.1002/2016GL067730.

Molinari, J., S. Skubis, and D. Vollaro, 1995: External influences on hurricane intensity. Part III: Potential vorticity structure. *Journal of Atmospheric Sciences*, **52**, 3593–3606, https://doi.org/10.1175/1520-0469(1995)052<3593:EIOHIP>2.0.CO;2.

——, D. Knight, M. Dickinson, D. Vollaro, and S. Skubis, 1997: Potential vorticity, easterly waves, and eastern Pacific tropical cyclogenesis. *Monthly Weather Review*, **125**, 2699–2708, https://doi.org/10.1175/1520-0493(1997)125<2699:PVEWAE>2.0.CO;2.

——, D. Vollaro, and K. L. Corbosiero, 2004: Tropical cyclone formation in a sheared environment: a case study. *Journal of the Atmospheric Sciences*, **61**, 2493–2509, https://doi.org/10.1175/jas3291.1.

Monin, A. S., and A. M. Obukhov, 1954: Basic laws of turbulent mixing in the surface layer of the atmosphere. *Tr. Akad. Nauk. SSSR Geophiz. Inst.*, **24**, 163–187.

Montgomery, M. T., and J. Enagonio, 1998: Tropical cyclogenesis via convectively forced vortex Rossby waves in a three-dimensional quasigeostrophic model. *J. Atmos. Sci.*, **55**, 3176–3207.

——, and R. K. Smith, 2012: The genesis of Typhoon Nuri as observed during the Tropical Cyclone Structure 2008 (TCS08) field experiment. Part 2: Observations of the convective environment. *Atmos. Chem. Phys.*, **12**, 4001–4009, https://doi.org/10.5194/acp-12-4001-2012.

——, M. E. Nicholls, T. A. Cram, and A. B. Saunders, 2006: A vortical hot tower route to tropical cyclogenesis. *Journal of the Atmospheric Sciences*, **63**, 355–386, https://doi.org/10.1175/jas3604.1.

Montgomery, M. T., and Coauthors, 2012: The pre-depression investigation of cloud-systems in the tropics (PREDICT) experiment: Scientific basis, new analysis tools, and some first results. *Bulletin of the American Meteorological Society*, **93**, 153–172, https://doi.org/10.1175/bams-d-11-00046.1.

Mrowiec, A. A., S. T. Garner, and O. M. Pauluis, 2011: Axisymmetric hurricane in a dry atmosphere: Theoretical framework and numerical experiments. *J. Atmos. Sci.*, **68**, 1607–1619.

Muller, C. J., and D. M. Romps, 2018: Acceleration of tropical cyclogenesis by self-aggregation feedbacks. *Proceedings of the National Academy of Sciences*, **115**, 2930–2935, https://doi.org/10.1073/pnas.1719967115.

Nicolis, G., and I. Prigogine, 1977: *Self-organization in nonequilibrium systems*. John Wiley & Sons, 490 pp.

Nolan, D. S., 2007: What is the trigger for tropical cyclogenesis? *Aust. Met. Mag.*, **56**, 241–266.

Nolan, D. S., and E. D. D.-:10. 1029/2008GL034147 Rappin, 2008: Increased sensitivity of tropical cyclogenesis to wind shear in higher SST environments. *Geophysical Research Letters*, **35**.

Nolan, D. S., and M. G. McGauley, 2012: Tropical cyclogenesis in wind shear: Climatological relationships and physical processes. *Cyclones: Formation, Triggers, and Control*, K. Oouchi and H. Fudeyasu, Eds., Nova Science Publishers.

Nolan, D. S., E. D. Rappin, and K. A. D.-:10. 1002/qj. 170 Emanuel, 2007: Tropical cyclogenesis sensitivity to environmental parameters in radiative–convective equilibrium. *Quarterly Journal of the Royal Meteorological Society*, **133**, 2085–2107.

Ooyama, K., 1964: A dynamical model for the study of tropical cyclone development. *Geofis. Int.*, **4**, 187–198.

Palmén, E., 1948: On the formation and structure of tropical hurricanes. *Geophysica*, **3**, 26–39.

Patricola, C. M., R. Saravanan, and P. D.-:10. 1002/2017GL076081 Chang, 2018: The response of Atlantic tropical cyclones to suppression of African easterly waves. *Geophysical Research Letters*, **45**, 471–479.

Pauluis, O., and I. M. Held, 2002: Entropy budget of an atmosphere in radiative-convective equilibrium. Part I: Maximum work and frictional dissipation. *J. Atmos. Sci.*, **59**, 125–139.

Ramage, C. S., 1959: Hurricane development. *Journal of Meteorology*, **16**, 227–237, https://doi.org/10.1175/1520-0469(1959)016<0227:hd>2.0.co;2.

Rappin, E. D., D. S. Nolan, and K. A. D.-:10. 1002/qj. 706 Emanuel, 2010: Thermodynamic control of tropical cyclogenesis in environments of radiative-convective equilibrium with shear. *Quarterly Journal of the Royal Meteorological Society*, **136**, 1954–1971.

Rayleigh, L. (John W. S., 1916: On convective currents in a horizontal layer of fluid when the higher temperature is on the underside. *Phil. Mag.;*, **32**, 529-546.

Raymond, D. J., S. L. Sessions, and C. L. D.-:10. 1029/2011JD015624 Carrillo, 2011: Thermodynamics of tropical cyclogenesis in the northwest Pacific. *Journal of Geophysical Research: Atmospheres*, **116**.

Reed, K. A., and D. R. Chavas, 2015: Uniformly rotating global radiative-convective equilibrium in the Community Atmosphere Model, version 5. *Journal of Advances in Modeling Earth Systems*, **7**, 1938–1955, https://doi.org/10.1002/2015MS000519.

Reilly, D. H., 1992: On the role of upper-tropospheric potential vorticity advection in tropical cyclone formation: Case studies from 1991. Massachusetts Institute of Technology, 124 pp. https://dspace.mit.edu/handle/1721.1/54988.

Rhines, P. B., 1975: Waves and turbulence on a beta-plane. *Journal of Fluid Mechanics*, **69**, 417–443, https://doi.org/10.1017/S0022112075001504.

Riehl, H., 1948a: *On the formation of west Atlantic hurricanes*.

Riehl, H., 1948b: On the formation of typhoons. *Journal of Meteorology*, **5**, 247–265, https://doi.org/10.1175/1520-0469(1948)005<0247:otfot>2.0.co;2.

——, 1951: Aerology of Tropical Storms. *Compendium of Meteorology*, T.F. Malone, Ed., American Meteorological Society, 902–913.

Riehl, H., 1954: *Tropical Meteorology*. McGraw-Hill, 392 pp.

——, 1963: Some relations between wind and thermal structure of steady state hurricanes. *J. Atmos. Sci.*, **20**, 276–287.

——, 1975: *Further studies on the origin of hurricanes*. Colorado State Univ.,.

Robe, F. R., and K. Emanuel, 1996: Dependence of tropical convection on radiative forcing. *J. Atmos. Sci.*, **53**, 3265–3275.

Rotunno, R., and K. A. Emanuel, 1987: An air-sea interaction theory for tropical cyclones. Part II. *J. Atmos. Sci.*, **44**, 542–561.

——, Y. Chen, W. Wang, C. Davis, J. Dudhia, and C. L. Holland, 2009: Large-eddy simulation of an idealized tropical cyclone. *Bull. Amer. Meteor. Soc.*, **90**, 1783–1788.

Ruppert, J. H., A. A. Wing, X. Tang, and E. L. Duran, 2020: The critical role of cloud–infrared radiation feedback in tropical cyclone development. *Proceedings of the National Academy of Sciences*, **117**, 27884–27892, https://doi.org/10.1073/pnas.2013584117.

Sadler, J. C., 1967: *The tropical upper tropospheric trough as a secondary source of typhoons and a primary source of tradewind disturbances*. Hawaii Institute of Geophysics,.

——, 1976a: *Tropical cyclone initiation by the tropical upper tropospheric trough*. Naval Environmental Prediction Research Facility,.

Sadler, J. C., 1976b: A role of the tropical upper tropospheric trough in early season typhoon development. *Monthly Weather Review*, **104**, 1266–1278, https://doi.org/10.1175/1520-0493(1976)104<1266:AROTTU>2.0.CO;2.

Shi, X., and C. S. Bretherton, 2014: Large-scale character of an atmosphere in rotating radiative-convective equilibrium. *Journal of Advances in Modeling Earth Systems*, **6**, 616–629, https://doi.org/10.1002/2014MS000342.

Simpson, R. H., N. FRANK, D. SHIDELER, and H. M. JOHNSON, 1968: Atlantic tropical disturbances, 1967. *Monthly Weather Review*, **96**, 251–259, https://doi.org/10.1175/1520-0493(1968)096<0251:ATD>2.0.CO;2.

Skamarock, W. C., and Coauthors, 2019: A description of the advanced research WRF version 4. *NCAR tech. note ncar/tn-556+ str*, **145**.

Slyman, K., J. A. Gemmer, N. K. Corak, C. Kiers, and C. K. R. T. Jones, 2023: Tipping in a low-dimensional model of a tropical cyclone.

Sugi, M., Y. Yamada, K. Yoshida, R. Mizuta, M. Nakano, C. Kodama, and M. Satoh, 2020: Future Changes in the Global Frequency of Tropical Cyclone Seeds. *SOLA*, **16**, 70–74, https://doi.org/10.2151/sola.2020-012.

Tang, B., and K. Emanuel, 2010: Midlevel ventilation's constraint on tropical cyclone intensity. *Journal of the Atmospheric Sciences*, **67**, 1817–1830, https://doi.org/10.1175/2010jas3318.1.

Tang, B. H.-An., 2010: Midlevel ventilation's constraint on tropical cyclone intensity. Massachusetts Institute of Technology, Department of Earth, Atmospheric, and Planetary Sciences, 195 pp. https://dspace.mit.edu/handle/1721.1/62321.

Thorncroft, C., and K. Hodges, 2001: African easterly wave variability and its relationship to atlantic tropical cyclone activity. *Journal of Climate*, **14**, 1166–1179, https://doi.org/10.1175/1520-0442(2001)014<1166:AEWVAI>2.0.CO;2.

Tippett, M. K., S. Camargo, and A. H. Sobel, 2011: A Poisson regression index for tropical cyclone genesis and the role of large-scale vorticity in genesis. *J. Climate*, **24**, 2335–2357.

Trenberth, K. E., and J. M. Caron, 2001: Estimates of Meridional Atmosphere and Ocean Heat Transports. *Journal of Climate*, **14**, 3433–3443.

Vecchi, G. A., and Coauthors, 2019: Tropical cyclone sensitivities to CO2 doubling: roles of atmospheric resolution, synoptic variability and background climate changes. *Climate Dynamics*, **53**, 5999–6033, https://doi.org/10.1007/s00382-019-04913-y.

Velez-Pardo, M., and T. W. Cronin, 2023: Large-scale circulations and dry tropical cyclones in direct numerical simulations of rotating Rayleigh-Bénard convection. *J. Atmos. Sci.*, In press.

Viale, F., and T. M. Merlis, 2017: Variations in tropical cyclone frequency response to solar and CO2 forcing in aquaplanet simulations. *Journal of Advances in Modeling Earth Systems*, **9**, 4–18, https://doi.org/10.1002/2016MS000785.

Vu, T.-A., C. Kieu, D. Chavas, and Q. Wang, 2021: A numerical study of the global formation of tropical cyclones. *Journal of Advances in Modeling Earth Systems*, **13**, e2020MS002207, https://doi.org/10.1029/2020MS002207.

Walsh, K. J. E., S. Sharmila, M. Thatcher, S. Wales, S. Utembe, and A. Vaughan, 2020: Real world and tropical cyclone world. part ii: sensitivity of tropical cyclone formation to uniform and meridionally varying sea surface temperatures under aquaplanet conditions. *Journal of Climate*, **33**, 1473–1486, https://doi.org/10.1175/JCLI-D-19-0079.1.

Wang, Z., M. T. Montgomery, and T. J. Dunkerton, 2010: Genesis of pre–Hurricane Felix (2007). Part I: The role of the easterly wave critical layer. *Journal of the Atmospheric Sciences*, **67**, 1711–1729, https://doi.org/10.1175/2009jas3420.1.

——, M. T. Montgomery, and C. Fritz, 2012: A First Look at the Structure of the Wave Pouch during the 2009 PREDICT–GRIP Dry Runs over the Atlantic. *Monthly Weather Review*, **140**, 1144–1163, https://doi.org/10.1175/mwr-d-10-05063.1.

Wing, A. A., S. J. Camargo, and A. H. Sobel, 2016: Role of radiative–convective feedbacks in spontaneous tropical cyclogenesis in idealized numerical simulations. *Journal of the Atmospheric Sciences*, **73**, 2633–2642, https://doi.org/10.1175/JAS-D-15-0380.1.

Zehnder, J. A., D. M. Powell, and D. L. Ropp, 1999: The interaction of easterly waves, orography, and the intertropical convergence zone in the genesis of eastern Pacific tropical cyclones. *Monthly Weather Review*, **127**, 1566–1585, https://doi.org/10.1175/1520-0493(1999)127<1566:TIOEWO>2.0.CO;2.

Zehr, R. M., 1992: *Tropical cyclogenesis in the western North Pacific*. NOAA,.

Zhang, Y., H. Wang, J. Sun, and H. Drange, 2010: Changes in the tropical cyclone genesis potential index over the western North Pacific in the SRES A2 scenario. *Advances in Atmospheric Sciences*, **27**, 1246–1258, https://doi.org/10.1007/s00376-010-9096-1.

Zhou, W., I. M. Held, and S. T. Garner, 2017: Tropical cyclones in rotating radiative–convective equilibrium with coupled sst. *Journal of the Atmospheric Sciences*, **74**, 879–892, https://doi.org/10.1175/JAS-D-16-0195.1.

15. Climate and Tropical Cyclones

Among the most interesting, vexing, and important issues concerning tropical cyclones is how and why they vary with climate. While the problem is usually posed as one of determining the strictly passive response of tropical cyclones to changing climate, it is possible that they have important feedbacks on climate, making them integral components of the climate system. If the latter turns out to be the case, then current global climate models, which woefully under-resolve TCs, are largely missing an important component of the climate system.

In this chapter we begin by reviewing what little is known, from an observational standpoint, about the relationship between TCs and climate, focusing on intraseasonal variability and ENSO and moving on to longer time scales revealed by geological records. We then turn to applications of theory and models to understanding the effect of changing climate on TCs, wrapping up with a discussion of possible feedbacks of TCs on climate.

15.1 Modulation of tropical cyclone activity by ENSO

Recognition of the modulation of North Atlantic tropical cyclone activity by ENSO goes back at least as far as the work of Gray (1984), and its effect on western North Pacific TC activity traces back to Chan (1985). The reader is referred to the author's review paper (Emanuel 2018) for a more comprehensive review of the history of research on ENSO and TCs, including modulation of TCs in other ocean basins.

In the case of the North Atlantic, El Niño suppresses TC activity; conversely, La Niña enhances it. During El Niño, the eastern equatorial Pacific warms dramatically, altering both the atmospheric thermal structure and circulation across a broad swath of the tropics (see Chapter 6). Specifically, the positive SST anomalies in the eastern equatorial Pacific induce a Gill-like atmospheric response (section 6.3.3), entailing a westerly wind shear anomaly broadly to the east of the SST anomaly. This shear anomaly is widely regarded as the main agent of suppression of North Atlantic TCs, though as we shall see later, this may not be the whole story.

The warming of the tropical troposphere proceeds on a time scale of months, which is quite similar to the thermodynamic response time of the ocean's mixed layer to perturbations in surface fluxes. Consequently, while El Niño heats the tropical troposphere, the tropical North Atlantic SSTs cannot keep up, and, following the Weak Temperature Gradient response, air subsides over the tropical Atlantic, drying out the middle troposphere and increasing the nondimensional saturation deficit as given by χ (Chapter 14, equation 14.5). This works in the same direction as increased wind shear, suppressing TC genesis. Tang and Neelin (2004) argued that the warming of the tropical atmosphere relative to North Atlantic SSTs would also suppress TCs by lowering potential intensity.

Even though Atlantic TC activity is suppressed in El Niño years, it can be enhanced along the U.S. coast, demonstrating that basin-wide metrics may not always be suitable proxies for coastal hazard risk (Kossin 2017).

Figure 15.1 shows correlations between Atlantic and eastern North Pacific power dissipation index (PDI), summed over all storms in the basin each year, and September sea surface temperatures at approximately 0.5-degree latitude resolution, from MERRA2 reanalyses data (NASA 2015), 1980 to 2022. The power dissipation index is a loose measure of the kinetic energy dissipated during the lifetime of a storm and is defined

$$PDI \equiv \int_0^\tau V_{\max}^3 dt. \tag{15.1}$$

where $V_{\max}$ is the maximum surface wind speed at a given time and τ is the lifetime of the storm. (The actual energy dissipation would include the air density and also integrate the wind speed cubed over the surface area of the storm, but there is seldom sufficient data to do this with any accuracy.)

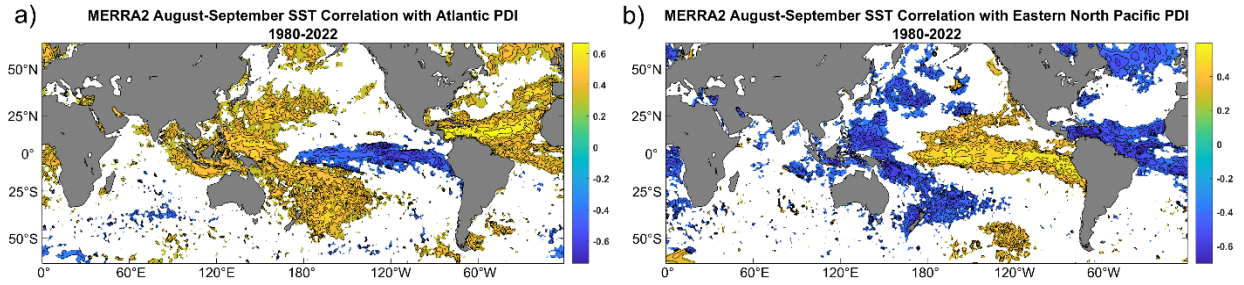


Figure 15.1: Correlation coefficient between annual times series of North Atlantic (a) and eastern North Pacific (b) power dissipation index (PDI) and September mean sea surface temperature at each point on the globe, from 1980 to 2022. Data are displayed only where the p-value of the linear regression is less than 0.15.

Atlantic hurricane power dissipation is large when a typical La Niña signal is present in the Pacific, with a cold eastern equatorial Pacific and warm water in the far western tropical Pacific. There is also a pronounced warm anomaly in the tropical North Atlantic genesis region. The signal is pretty much reversed for active seasons in the eastern North Pacific, with sea surface temperature anomalies reflecting typical El Niño conditions. (Not surprisingly, eastern North Pacific and Atlantic tropical cyclone activity metrics are anti-correlated.) Note that local sea surface temperature anomalies are not significantly correlated with PDI in the eastern North Pacific.

To get some insight into the physics underlying the modulation of tropical cyclones by ENSO, we first note that there is a reasonable correlation ($r \simeq 0.54$) between North Atlantic PDI and the genesis potential index (GPI) given by (14.4), averaged over the so-called Main Development Region (MDR, defined as 6°-18°N, 20°-60° W) and summed over all months of the year. Moreover, regressing SST against GPI rather than PDI yields a correlation structure and magnitude very similar to that shown in Figure 15.1a. Consequently, we may inquire which components of the GPI are most influential in ENSO modulation of North Atlantic PDI. We can do this by regressing PDI against the individual components that make up the GPI given by (14.4).

This regression shows that the potential intensity and shear are the two most important factors. The correlation coefficient for potential intensity alone is 0.58 and for shear alone is 0.47. Using

both potential intensity and shear together in a multiple linear regression yields a correlation coefficient of 0.6. The mid-tropospheric dryness parameter χ, given by (14.5), comes in a distant third with a correlation coefficient of only 0.12.

Yet these numbers do not tell the whole story. If we regress SST against the individual factors that make up the GPI we discover essentially no ENSO signal for either shear or potential intensity, even though there is a strong ENSO signal when we regress against the full GPI. Yet there IS a strong ENSO signal when we regress against χ itself, even though the direct correlation of χ with PDI is weak.

How are we to interpret these seemingly contradictory findings? Recall that ENSO only explains some of the variance of North Atlantic TC activity and GPI. When we correlate PDI with GPI and its components, we are including all the variability, not just the ENSO modulation. Local potential intensity and shear dominate the total variability of GPI, and both are strongly correlated with SSTs in the tropical and subtropical North Atlantic, but not with Pacific SSTs. Although mid-tropospheric dryness is a distant third in its contribution to GPI, it appears to be strongly correlated with ENSO signals, with correlation coefficients with equatorial Pacific SST as high as 0.6. Since the factors combine multiplicatively in the GPI, the ENSO modulation of χ modulates the whole GPI response.

Figure 15.2 is identical to Figure 15.1 except for PDI accumulated over western North Pacific tropical cyclones. Here there is a strong ENSO signal, but it peaks in the central equatorial Pacific rather than further east as with the eastern North Pacific and North Atlantic PDIs.

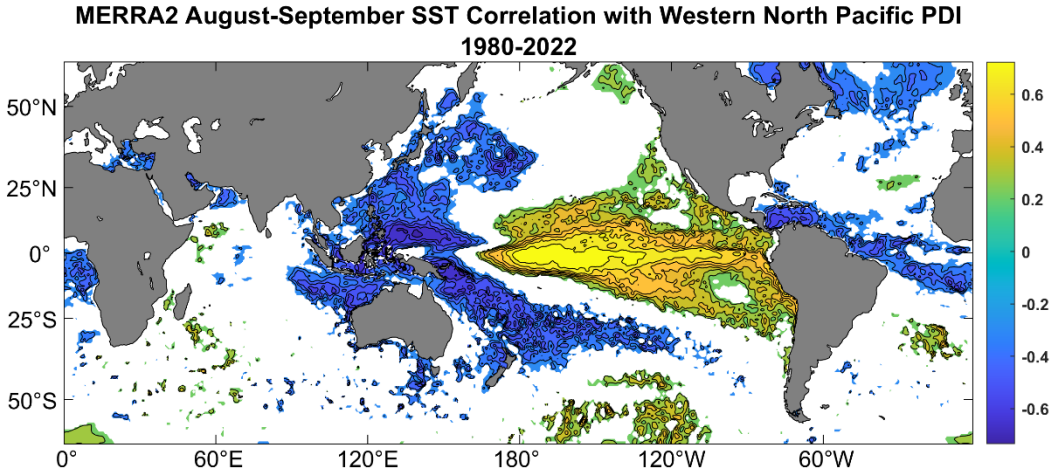


Figure 15.2: Correlation coefficient between annual times series of western North Pacific power dissipation index (PDI) and September mean sea surface temperature at each point on the globe, from 1980 to 2022. Data are displayed only where the p-value of the linear regression is less than 0.15.

Central Pacific El Niño events (often referred to as Niño Modoki) enhance western North Pacific TC activity, as measured by power dissipation. Unlike the case of the North Atlantic, there is no substantial correlation between western North Pacific PDI and genesis indices or their components. However, there is some correlation between the PDI of storms that form east of 160°W and potential intensity averaged over the months of July through November.

ENSO has some degree of predictability (Cane et al. 1986), which provides part of the basis for seasonal prediction of tropical cyclone activity.

15.2: Modulation of tropical cyclone activity by tropical intraseasonal variability

Many of the modes of tropical intraseasonal variability, described in detail in Chapter 7, have been found to modulate tropical cyclone activity in virtually all basins in which TCs are active. Recognition of the modulation of TCs by the MJO dates back at last as far as the work of Liebmann et al. (1994). (See the review by the author (Emanuel 2018) for a research history of this topic.)

Figure 15.3 presents evidence of modulation of TCs by the MJO as well as an indication of what physical factors may be responsible for this modulation. The black contours in the figure show observed TC genesis density anomaly for the years 1982-2007 as determined from historical observational data sets from the National Hurricane Center (NHC) and the Joint Typhoon Warning Center (JTWC), composited across combinations of 8 phases of the MJO determined from a method developed by Wheeler and Hendon (2004) applied to NCAR/NCEP reanalysis data over the same interval. Broadly, this MJO index is determined by the multivariate empirical orthogonal function (MV-EOF) analysis applied to seasonally detrended daily lower- and upper-level zonal winds and outgoing longwave radiation (OLR).

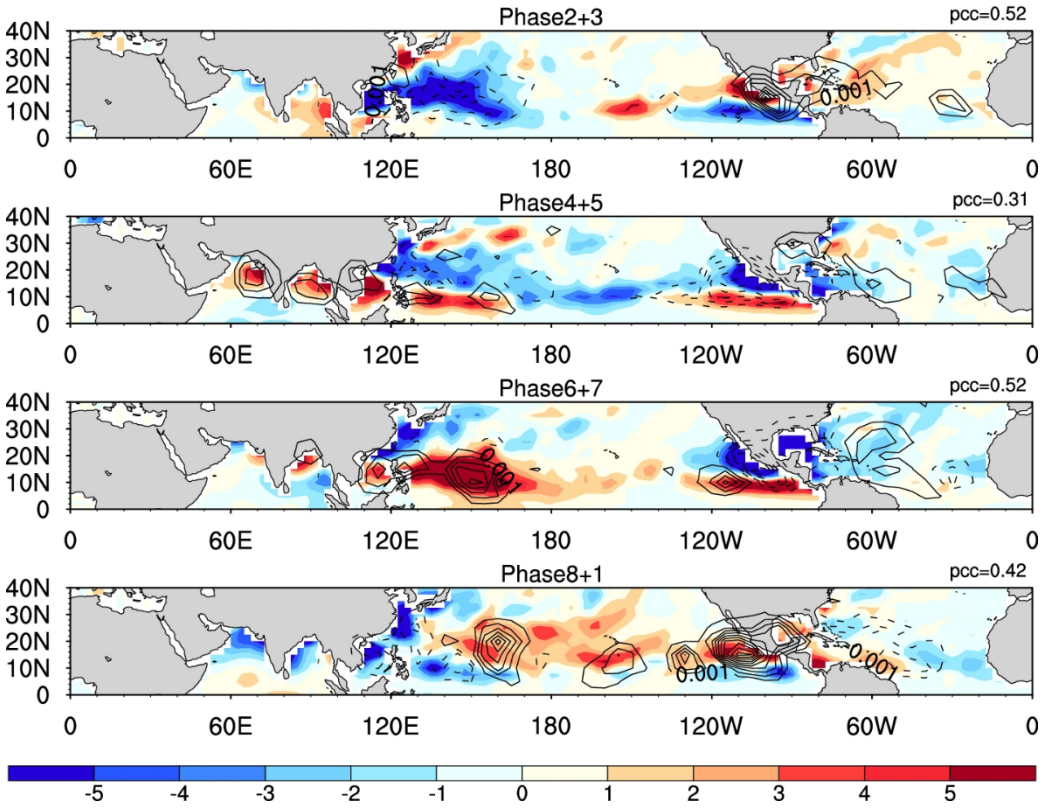


Figure 15.3: Genesis density anomaly (black contours) and Genesis Potential Index anomalies (shading) for combinations of 8 phases of the Madden-Julian Oscillation as determined by the method of Wheeler and Hendon (2004). The "pcc" numbers in the upper right corners of each panel show the pattern correlation of the two anomaly fields.

The shading in Figure 15.3 shows anomalies of the genesis potential index developed by Emanuel and Nolan (2004) applied to the NCAR/NCEP reanalysis data for the 8 MJO phases. This GPI is defined

$$GPI \equiv \left|10^5 \eta\right|^{3/2} \left(\frac{\mathcal{H}}{50}\right)^3 \left(\frac{V_{pot}}{70}\right)^3 \left(1 + 0.1 V_{shear}\right)^{-2}, \tag{15.2}$$

where η is the absolute vorticity in s^{-1}, $\mathcal{H}$ is the relative humidity at 700 hPa in percent, V_{pot} is the potential intensity in $m\,s^{-1}$, and V_{shear} is the magnitude of the vector shear from 850 to 200 hPa, in $m\,s^{-1}$.

Figure 15.3 shows broadly that the GPI anomalies oscillate in place, but are out of phase between the western and eastern North Pacific, possibly a result of eastward propagation. There is a reasonable correspondence between the GPI and genesis anomalies, as indicated by the modest values of the pattern correlations between the two fields, though there are quite a few discrepancies as well.

The authors of this figure, Zhao and Li (2019), determined the relative contributions of the various factors that make up the GPI given by (15.2) and showed that, in the western North Pacific and Gulf of Mexico regions, the MJO-modulated GPI is dominated by the relative humidity term, while shear dominates in the eastern tropical North Atlantic. In no case were fluctuations of the vorticity or potential intensity terms important.

Consequently, it would appear that the MJO modulates tropical cyclone activity principally through its effects on wind shear and mid-tropospheric humidity. The MJO does not modulate SST or atmospheric temperature enough to cause an appreciable signal in potential intensity, in contrast to the much slower ENSO, in which modulation of potential intensity is an important influence on TC activity, as shown in the previous section.

At shorter time scales, the effects of sea surface temperature variability become even less important. Numerous studies (e.g. Schreck et al. 2012; Wu and Takahashi 2018; Zhao et al. 2019; Lai et al. 2020; Wang et al. 2022) have demonstrated that genesis in the deep tropics can be strongly modulated by equatorial waves, including equatorial Rossby, mixed Rossby-gravity and, to a lesser extent, Kelvin waves. While it is not entirely clear what mechanisms are in play, modulation of low-level vorticity and low- to mid-level humidity appear to be the important factors. Many instances of genesis also appear to be modulated by synoptic scale waves (Zhao et al. 2019) and modulation by lower frequency disturbances may also act indirectly through their modulation of the synoptic scale disturbances.

While the modulation of genesis by equatorial and synoptic scale waves is of great interest for understanding genesis in general, and of practical interest for real-time forecasting of tropical cyclones, it is less clear that wave activity exercises much control over the longer-term climatology of TCs. The success of genesis indices based on monthly mean climatology in explaining longer-term spatial and temporal TC variability (Chapter 14) weighs against the idea of any strong rectification of genesis by higher frequency fluctuations. Even long-term variations

in ENSO states may have counter-intuitive effects on genesis activity; for example, Federov et al. (2010) found that imposition of a permanent El Niño state thought to represent the Pliocene period actually results in an increase in North Atlantic TC activity, opposite to what is observed in today's ENSO variations. For this reason, we next focus our attention on longer time-scale changes in climate.

15.3 Paleotempestology

In a seminal paper, Tor Bergeron (1954) wrote

> *"What kind of secular changes may have existed in the frequency and intensity of the hurricane vortices on our Earth? And what changes may be expected in future? … During certain geological epochs, hurricanes may have been just as frequent as the cyclones of our latitudes, or they may have occurred all over the oceans and within all coastal regions, and they may have been even more violent than nowadays. During other periods they may have been lacking altogether. In studying paleo-climatic and paleo-biological phenomena, especially along the coasts of previous geological epochs, it may be wise to consider such possibilities."*

Bergeron's prescient advice began to take root in the 1980s and the development and interpretation of proxies for tropical cyclones in the geological record has accelerated greatly in the last few decades. The proxies may be divided into two categories: mechanical proxies consisting of signatures of storm surges or strong currents in the sedimentary record, and geochemical signatures that so far rely on the observed fact the tropical cyclone rains are depleted in the heavier isotopes of oxygen and hydrogen. Nice reviews of the current field of paleotempestology are provided by Muller et al. (2017) and Oliva et al. (2017).

15.3.1 Tempestites

The study of ancient storms in the geologic record began with the discovery of tempestites, storm-influenced sedimentary deposits that formed on sea or lake beds owing to the action of currents and/or surface and internal waves (Myrow and Southard 1996). Their most prominent form is called hummocky cross-stratification, a type of cross-bedding in sedimentary deposit consisting of a series of hummock-like shapes that that cross-cut each other, as illustrated in Figure 15.4.

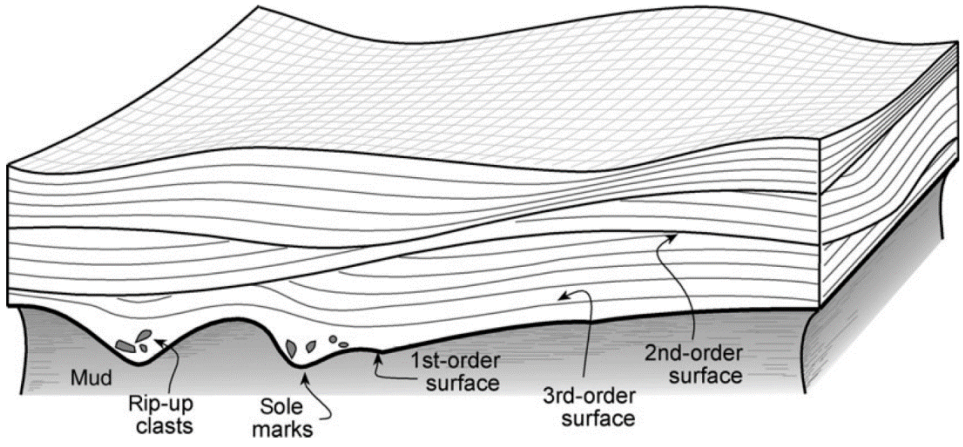


Figure 15.4: Sketch of the appearance of hummocky cross-stratification in sandy sediments. Hummocks are usually 10-50 cm high and horizontally spaced at from one to several meters. Sole marks are small-scale (usually of order 1 cm) grooves or marks at the bases of certain strata. Rip-up clasts are gravel-size chunks created when an erosive current flows over a bed of clay or mud and removes and transports pieces of sediment.

They can be seen in exposed sediment beds, including road-cuts. While there is near universal agreement that these arise from strong storms, the particular mechanisms are still debated. In most cases they appear to have been laid down in water depths between the depths at which typical surface waves have appreciable signal and depths below which tidal and storm-driven currents are largely absent.

At the time of this writing, there is no universally accepted explanation for hummocky cross-stratification, but most researchers believe that it is caused by some combination of uni-directional and oscillatory flows over sediments (Myrow and Southard 1996), which in turn are caused by wind-driven currents, long, high-amplitude surface waves, and/or internal waves traveling on a pycnocline (Morsilli and Pomar 2012; Pomar et al. 2012). Duke (1985) proposed that the majority (~70%) of these deposits were caused by tropical cyclones, with the remainder owing to strong winter storms. His findings were challenged (e.g. Swift and Nummedal 1987) on the basis of an argument that baroclinic storms should cause similar flows over shallow sediment beds as tropical cyclones; however, it should be noted that tropical cyclones are far more effective in generating near-inertial oscillations than extratropical cyclones (Price 1983). Observations of the upper ocean made during the passage of 2004 Atlantic Hurricane Frances (Sanford et al. 2011) show strong near-inertial wave energy propagating down to at least 200 m over the course of a few days, whereas the broader horizontal scale of inertial oscillations produced by baroclinic cyclones is associated with much slower vertical energy propagation (Thomas and Zhai 2022). Duke (1985) presented evidence that the preponderance of hummocky cross-stratification deposits are at low paleo-latitudes, suggesting that tropical cyclones might have been responsible for most of them, perhaps through their excitation of near-inertial oscillations. At the time of this writing, the extent to which tempestites may be used to infer storm activity in general and tropical cyclone activity in particular remains controversial (Barron 1989).

15.3.2 Overwash deposits and blue holes

Many coastal regions have barrier islands separated from the mainland by small lakes (including sinkholes) or marshes. These are often organically rich, and layers of mud build up over time. Storm surges can overtop the barrier islands, washing sand, shells, and other marine material into the lakes or marshes. Cores taken in such environments reveal layers of sand and other marine deposits interspersed with the organic material, which can be radiocarbon dated to determine its age and that of the interspersed sand layers. The latter can be identified by color, grain size, benthic foraminifera content, fraction of organic matter, and/or isotopic composition. This technique was pioneered by Davis et al. (1989) and Liu and Fearn (1993). A variant of this approach makes use of “blue holes”, submerged sink holes in carbonate shelves, which also trap coarse sediments washed into them during strong storms. An example of sand layers in a core extracted from a blue hole in the Bahamas is shown in Figure 15.5, and an exceptionally long time series from a core taken in a back barrier lagoon on a Caribbean island is displayed in Figure 15.6. The times series of storm proxies in the latter case is compared to proxies for El Niño events and tropical west African precipitation. Periods of heightened hurricane activity appear to coincide with reduced El Niño events and enhanced west African rainfall.

Overwash deposits are caused mostly by storm surges, some of which, at higher latitudes, may be caused by winter storms. Similar deposits can be result from tsunamis and care must be taken to exclude these from reconstructions of storm activity.

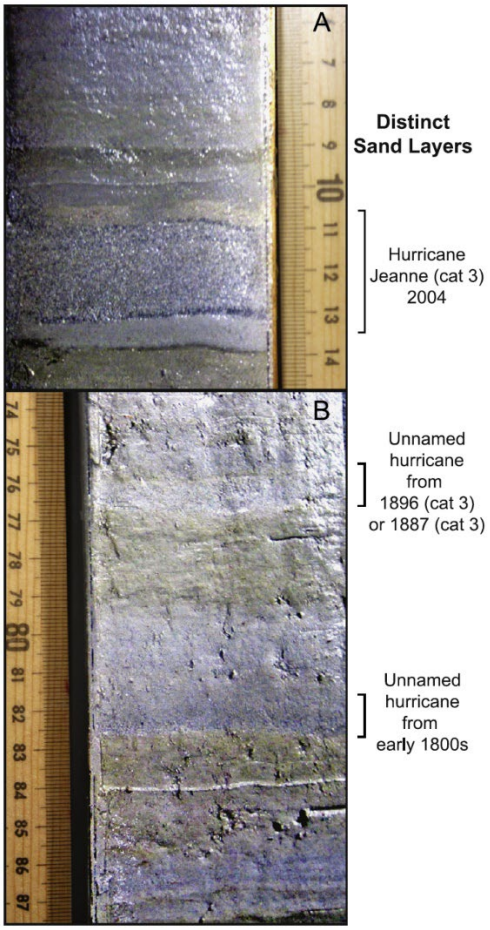


Figure 15.5: Cross-section of core from a blue hole at 69 m depth near Great Abaco Island in the Bahamas. (A) Sand layer at 11–12 cm produced by North Atlantic Hurricane Jeanne in 2004; (B) storm deposits at 76–76.5 cm and 82–82.5 cm.

A larger issue arises from the usually small sampling of events. Most cores sample only a handful of events, produced by strong storms that passed close to the site. Random clustering of events has sometimes been interpreted as long-term climate-driven variations in tropical cyclone activity (Wallace et al. 2021a). One possible solution to this problem is to infer climate-driven signals from networks of paleotempestology sites that are far enough apart to have sampled different individual events, but which span an area small enough to sample the same regional climatology (Wallace et al. 2021b).

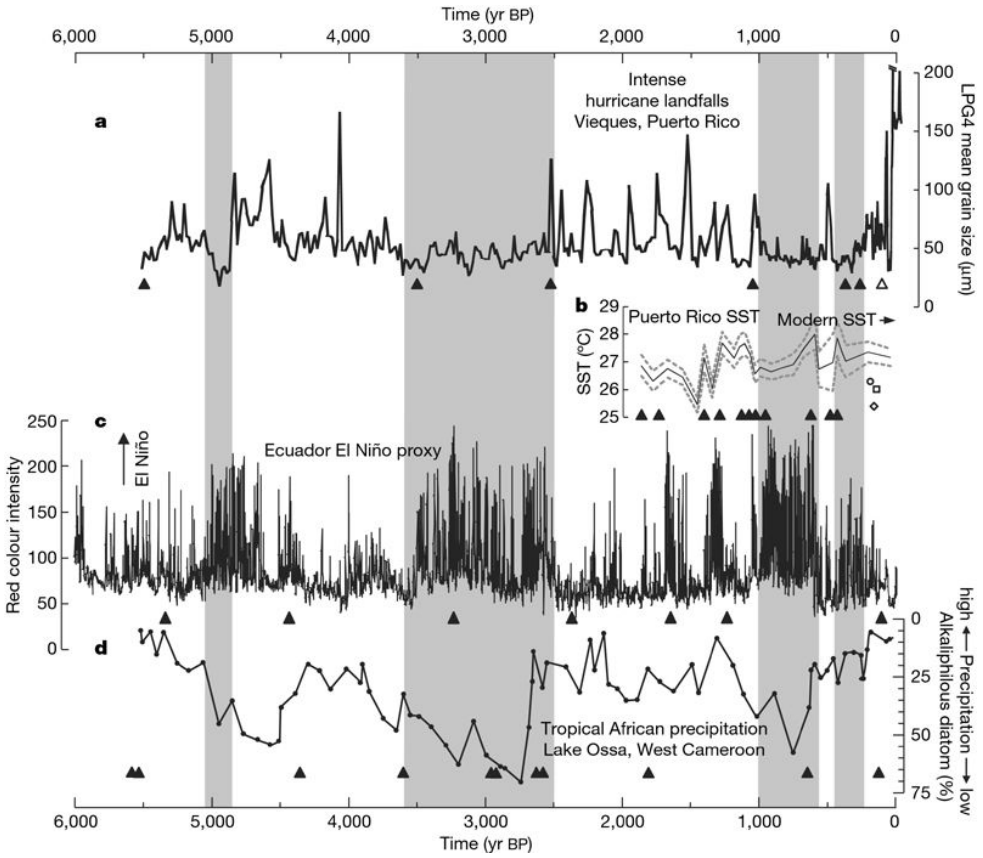


Figure 15.6: (a) Mean bulk grain-size record from a core sample taken from a back barrier lagoon on the Caribbean island of Vieques, Puerto Rico. Intervals of relatively few intense-hurricane-induced layers in all cores are noted with gray shading. (b) The thin line with the 2-std-dev uncertainty envelope (dashed lines) is a reconstruction of summer SSTs off Puerto Rico (core "PRP12"), and coral-based reconstruction of mean annual SSTs from La Parguera, Puerto Rico, are noted: 26.28C for 1700–1705 (open circle), 25.38C for 1780–85 (open diamond), and 26.08C for 1800–05 (open square). The modern mean annual SST is noted with an arrow. (c) El Niño proxy reconstruction from Laguna Pallcacocha, Ecuador. Peaks in red color intensity are documented as allochthonous material washed into the lake, primarily during strong El Niño events. (d) Changes in precipitation in West Cameroon inferred from alkaliphilous diatoms (thriving in alkaline conditions) from Lake Ossa. Radiocarbon-age control points are noted with filled black triangles below (a)–(d).

15.3.3: Beach deposits

In certain regions, notably Queensland in northeastern Australia, storm surges and tsunamis deposit ridges of sand and coral and shell fragments above the normal high tide levels. Most of these have long been attributed to storms (e.g. Chivas et al. 1986), and in Queensland and possibly other region in the deep tropics, this is likely to be the case. Yet some of the deposits are probably markers of tsunamis. The latter can sometimes be distinguished from the former by, e.g., the presence of boulders that have been lifted by tsunamis, when no such boulders are observed in known deposits by very intense historical tropical cyclones (Nott 2004).

It is thought that individual TCs leave a single ridge (Nott et al. 2009), but subsequent storms can leave additional ridges that may partially override pre-existing ridges (Nott and Hayne 2001). In locations with a long-term secular fall in sea level, the ridges may be separated by the normal landscape of the region (Nott et al. 2009). Material in the ridges can be radiocarbon dated to estimate when they were deposited.

Figure 15.7 displays a cross section through a sequence of storm deposits at two locations in Queensland.

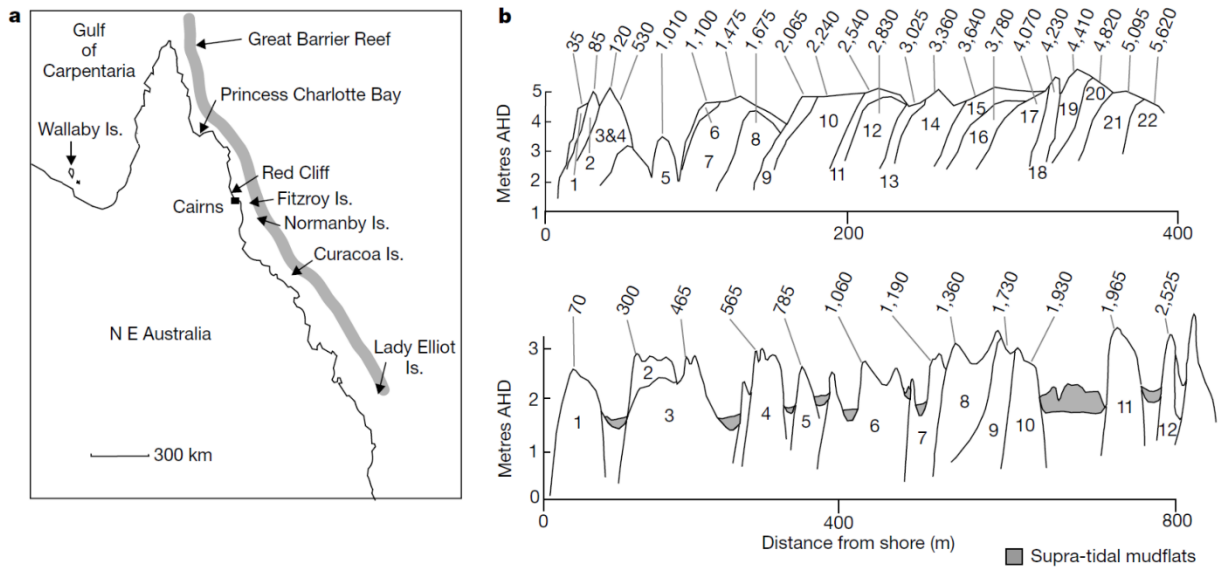


Figure 15.7: a) Map showing locations of beach ridges. b) Stratigraphy of beach ridges at Curacoa island (top) and Princess Charlotte Bay (bottom). Radiocarbon age (in yr BP) of each ridge is shown above the deposit.

Records of beach deposits can extend back as far as 6,000 years; much before that, such deposits would have been drowned by subsequent sea level rise at the end of the last glaciation. Note in figure 15.7 that the age of the ridges increases as one moves inland from the shore. Nott and Hayne (2001) argue that the ridges were deposited by intense storms, with central pressures less than 920 hPa. The records here suggest that in this region, intense landfalling tropical cyclones occur on average once in two or three centuries.

15.3.4 Geochemical proxies

We have already seen that radiocarbon techniques can be used to help date material in storm deposits, and cesium-137 and lead-210 have also been used to assist in dating. The isotopic composition of storm deposit material can also help determine whether the material is of marine origin.

Isotopes can also help detect TCs and other aggregated convective events. Rain in the cores of tropical cyclones has a distinct isotopic signature, being depleted (by up to 10 ‰) in the heavier stable isotopes of oxygen and hydrogen (Lawrence and Gedzelman 1996), probably because, out of thermodynamic necessity, there is much less evaporation of rain in the cores of tropical cyclones (and other types of aggregated convection) than in ordinary convection. Evaporation concentrates the heavier isotopes in the raindrops. The rainwater may be subsequently incorporated in cave deposits (speleothems) and/or trees.

An early attempt to look for such signatures was made by Malmquist (1997), who determined the isotopic composition of stalagmites in a Bermuda cave. Frappier et al. (2007) analyzed stalagmites from a cave in Belize, attaining monthly to weekly temporal resolution over a 24-

year period (1977-2000), during which 11 tropical cyclones passed over the region. The record of oxygen isotope ($\delta^{18}O$) excursions is compared to records of historical hurricane intensity and rainfall, to carbon isotope excursions ($\delta^{13}C$), and to a Southern Oscillation index (SOI) in Figure 15.8.

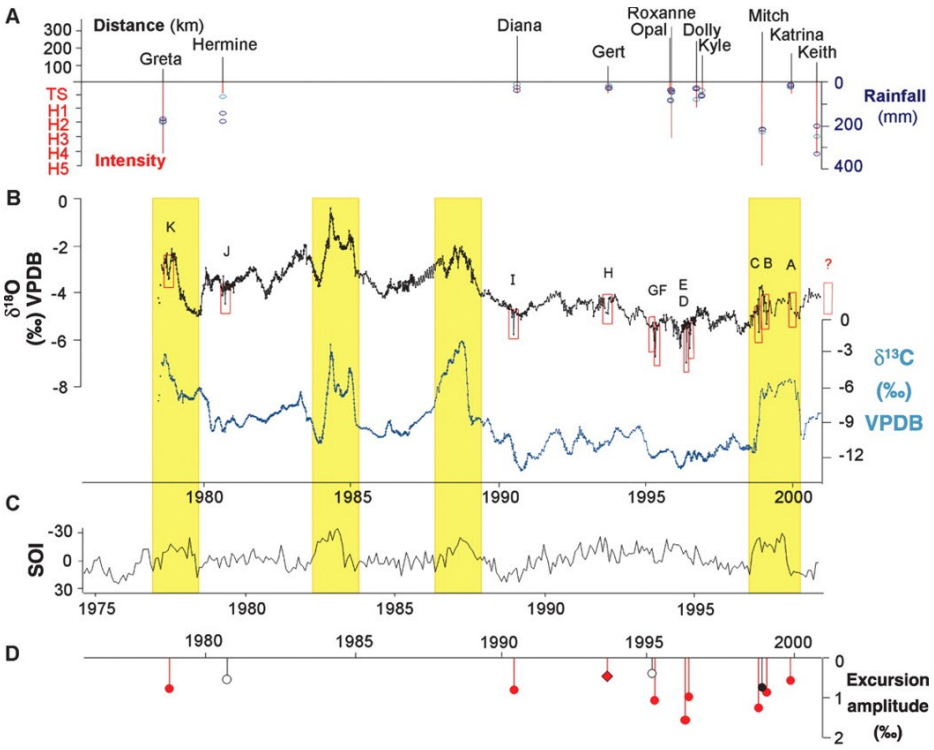


Figure 15.8: A) Record of historical hurricanes, with distance from cave to nearby storm tracks indicated by upward black bars, Saffir-Simpson intensity by downward red bars, and local rainfall at three nearby stations by blue circles; B) $\delta^{18}O$ in black and $\delta^{13}C$ in blue, with red boxes indicating downward excursions of $\delta^{18}O$ identified at tropical cyclone events; C) Southern Oscillation Index (SOI) on inverted scale, with yellow bars indicating major El Niño events which are accompanied by upward excursions of $\delta^{18}O$ and $\delta^{13}C$; D) Magnitudes of $\delta^{18}O$ excursions from background identified as tropical cyclone events, with red circles denoting correctly identified cyclones and black circles denoting excursions classified incorrectly as noncyclogenic (open) or cyclogenic (closed).

The very high frequency downward excursions of $\delta^{18}O$ correctly identified, for the most part, tropical cyclone events.

A similar analysis was performed by Haig (2014) on stalagmites from two caves in Australia, one on the northeast coast and the other near the west coast, with records extending back 1,500 and 700 years, respectively.

Denniston et al. (2023) compared isotopic analyses of stalagmites at two locations in western Australia; one recorded variations on monsoonal rains and the other mostly tropical cyclones. Based on these analyses and on tropical cyclones downscaled from global model simulations over a millennium, they concluded that TC activity in western Australia is highly correlated with the meridional position of the ITCZ, on multidecadal to centennial time scales.

Tropical cyclone rainwater is rapidly incorporated in near-surface soil water, which in turn can be taken up by trees, particularly those with shallow roots. The isotopic composition of the water is incorporated in tree trunks and can be analyzed in cores taken from the trees, as pioneered by Miller et al. (2006). Various refinements of this technique have been made (e.g. Li et al. 2011; Kagawa et al. 2015) and it has come to be regarded as a reliable proxy for tropical cyclone activity (e.g. Altman et al. 2021). Figure 15.9 shows a 354-year time series of $\delta^{18}O$ from trees in South Korea, with inferred tropical cyclone events. These are compared to historically documented events in the last half-century of the record.

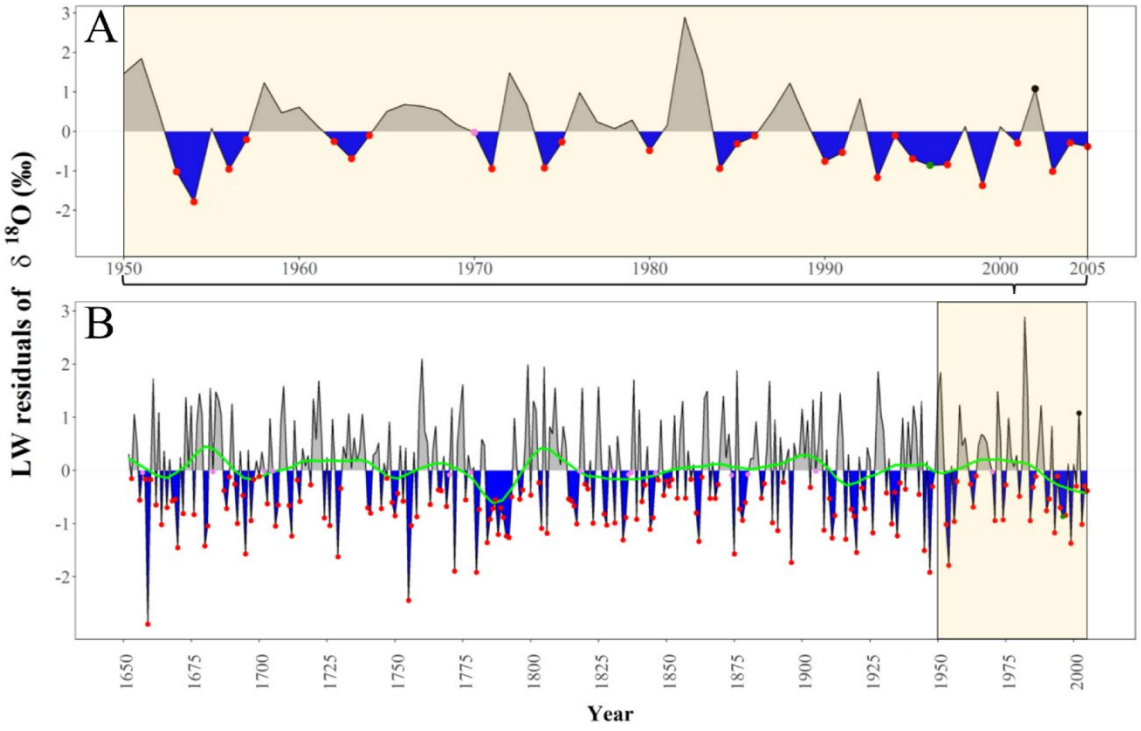


Figure 15.9: Residuals of $\delta^{18}O$ from the latewood (LW) portions of tree rings from a forest in South Korea. The residuals represent the difference between measured $\delta^{18}O$ and those from a 1-year autoregressive model, to minimize the effects of short-term auto-correlation in the time series, following the method of Miller et al. (2006). A): The modern instrumental period, 1950-2005, and B): the 354-year record (1652-2005) with a 25-year low-pass filter (green curve). The red dots in A correspond to known passages of tropical cyclones close to the analysis site, and the single green dot in 1996 indicates an observed TC that passes just outside the threshold for inclusion. The red dots in B mark years of TC events inferred from the Annals of the Joseon Dynasty (Lee and Hong 2012) during the period 1652–1903 and from other historical records thereafter. The pink dots in B represent probable TC events inferred from isotopic excursions that are not found in the historical records.

15.3.5: Historical data mining

In situ measurements if tropical cyclones are a comparatively modern development. Airborne reconnaissance began on the North Atlantic and western North Pacific region in the mid-1940s, and continues today only over the North Atlantic. Satellites were introduced in the 1960s and global coverage of the tropics was well established by about 1980. So called “best-track” data sets are based largely on instrumental measurements from ground stations and aircraft, and

inferences from satellite imagery. Today these data have been unified in the International Best Track Archive for Climate Stewardship (IBTrACS; Knapp et al. 2010) and are freely available. Until very recently, these data must be considered highly inhomogeneous, given large improvements in aircraft reconnaissance instrumentation and satellite technology over time, and great caution must be used to infer climate signals in such data, except perhaps for the highest frequency (e.g. ENSO) variability.

Paleotempestology seeks to lengthen the useful record of tropical cyclones, and so far we have discussed techniques based on proxies for TCs in the historical record. But there have also been attempts to extend the TC record, at least qualitatively, by examining historical accounts of extreme weather events, particularly in Asia, where such records may extend back for a thousand years or more. One early example is the work of Liu et al. (2001), who documented a 1,000-year history of typhoons striking Guangdong, Southern China, complied from semi-official local gazettes (see also Louie and Liu 2003). Trouet et al. (2016) examined historical records of Spanish shipwrecks in the Caribbean region and compared those to tree rings widths at Big Pine Key, Florida. Both records indicate reduced TC activity during the Maunder sunspot minimum, (1645-1715), when solar irradiance is thought to have been reduced.

Records of tropical cyclone activity have been deduced from historical archives of TCs affecting Japan (Grossman and Zaiki 2009), Madagascar (Nash et al. 2015), east China from 1450 to 1949 (Xu et al. 2012), Louisiana, USA (Mock 2008), the eastern and Central Caribbean since 1494 (Chenoweth and Howard 2023), the Mascarene Islands of the southwest Indian Ocean (Marriner et al. 2022), the Pacific coast of Mexico (Pazos and Mendoza 2013), eastern Australia (Callaghan and Power 2011), and the Yangtze River delta (Zhang et al. 2012) among other places.

15.3.6: Perspective on paleotempestology

The study of ancient tropical cyclones is a young and rapidly expanding venture that may prove our best hope for understanding the relationship between climate and TCs. Most of the geological proxies discussed here pertain to single locations, while historical records, other than maritime logs and shipwrecks, are inevitably weighted toward population centers. Each of the proxies has advantages and disadvantages, but most of them suffer from the fact that TCs, particularly strong TCs, are relatively rare phenomena. As demonstrated by Wallace et al. (2021a), Poisson noise in hurricane strikes at single locations can easily mislead one into deducing centennial-scale variability and trends, thus it seems likely that many of the published deductions of long-term variability and trends in paletempestological records are examples of being “fooled by randomness” (Taleb 2016). Understandably, the sites of storm records are chosen for a variety of mostly practical reasons, such as the existence of the right geological structures or places having long historical records, and maps of these sites have a helter-skelter quality (e.g. see Oliva et al. 2017). What is lacking at present is an optimal overall strategy for choosing new sites and for combining the various different proxies, using a network approach such as that advocated by Wallace et al. (2021b). Developing networks optimized to detect centennial and longer time scale variability and trends in TC activity holds great promise for improving our understanding of the relationship between climate and tropical cyclones.

15.4: Theoretical considerations

What can we say about the relationship between climate and tropical cyclone based on theory? As described in Chapter 14, we start with the strong handicap of not having an accepted theory for the global rate of genesis. While we do have a well-developed and tested theory for the number, size and intensity of tropical cyclones in "TC-World" numerical simulations, conducted in double periodic domains on $f-planes$ and with no external wind shear, it is clear that this scaling, at least for size and number, has little to do with the real world.

We do, however, have some understanding of the environmental factors that appear to control TC activity and characteristics. We next examine these factors individually.

15.4.1 Potential intensity

In Chapter 9 we derived a potential intensity from a differential Carnot cycle and also from thermal wind balance, assuming a vortex of zero saturation potential vorticity. In Chapter 14 we showed that potential intensity is also an important factor in semi-empirical genesis indices. Here we begin by showing that the lifetime maximum intensity (LMI) of observed storms, when normalized by the potential intensity at the place and time of LMI, forms a unique frequency distribution, once again pointing to the importance of potential intensity in real TCs. We then explore how climate controls potential intensity.

To show the universality of normalized LMI distributions, we first collect observed TCs over the period 1979-2022 from the IBTrACS database (Knapp et al. 2010). We find points over the ocean at which each TC reached its LMI. For this purpose, we only include storms whose LMI exceeded 15 ms^{-1}. We subtract 80% of the translation velocity at the time of LMI to estimate the circular wind component, which we take to be an estimate of the low-level gradient wind. We next find the potential intensity at these points by bi-linearly interpolating gridded monthly mean potential intensity derived from ERA5 reanalyses (Hersbach et al. 2020) using a potential intensity algorithm developed by Bister and Emanuel (2002). This algorithm accounts for heat input from isothermal expansion and dissipation of boundary layer turbulence. The potential intensity is also linearly interpolated to the day of LMI. Finally, we divide the modified LMI by the potential intensity at its date and location to create a normalized LMI.

A cumulative histogram of normalized LMI for North Atlantic TCs is displayed in Figure 15.10. The cumulative distribution seems to fall along two straight lines, determined by whether the normalized LMI is less than or greater than about 0.4. This characteristic shows up in sets of TCs spanning the same year in the eastern North Pacific, in the Southern Hemisphere in general, and is especially pronounced in the North Indian Ocean region. It is conspicuously absent in the western North Pacific region (Figure 15.11a), where the distribution is a single straight line. Further analysis demonstrates that, except in the western Pacific region, many tropical cyclones are limited by geography; they run out of space because they encounter either land or cold water. These make up a large proportion of the weaker storms in Figure 15.10 that have the steeper cumulative frequencies. Space is an especially strong limitation in the northern Indian Ocean (Figure 15.11b), so there are more storms in the part of the distribution that has a steeper slope.

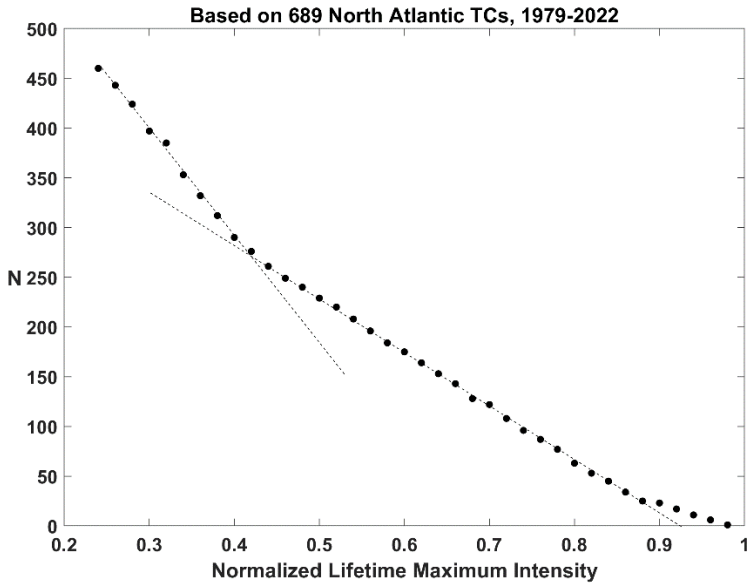


Figure 15.10: Cumulative histogram of normalized lifetime maximum intensity (LMI), defined as the ratio of a modified LMI to the potential intensity (from ERA5 reanalyses), for 689 North Atlantic TCs between 1979 and 2022. The recorded LMI has been modified by subtracting 80% of the storm's translation speed at the time of LMI, to estimate the circular gradient wind. Dashed lines show slopes along which the distribution is aligned.

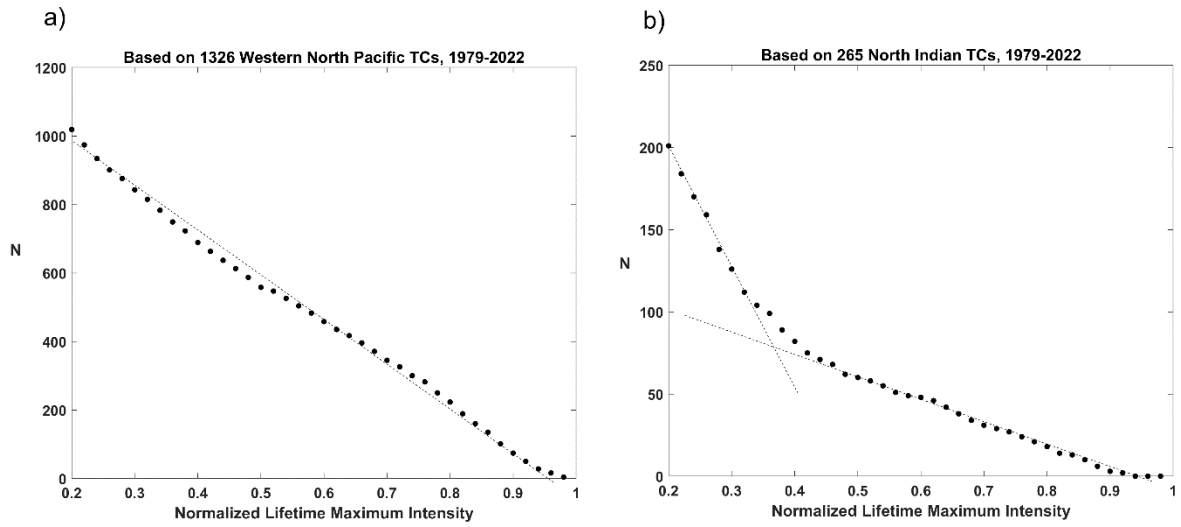


Figure 15.11: Same as Figure 15.10 but for the western North Pacific (a) and the North Indian Ocean (b).

A simple and intuitive interpretation of these results is that the LMI of a tropical cyclone is limited either by geography or by potential intensity; in either case, over the long run, the LMIs of individual storms are random up to a limit set by either the potential intensity or the time it takes to run out of space. We can test this idea by postulating that the limiting intensity V_{lim} of a tropical cyclone is given by

$$V_{lim} \equiv MIN\left(V_{pot}, \left(\frac{\partial V}{\partial t} \right)^{*} \Delta t_{max} \right), \tag{15.3}$$

where V_{pot} is the potential intensity at the time of peak intensity, $\left(\frac{\partial V}{\partial t} \right)^{*}$ is a characteristic maximum rate of intensification, and Δt_{max} is the time from genesis to maximum intensity. Choosing $\left(\frac{\partial V}{\partial t} \right)^{*}$ to be $4\ ms^{-1}$ in 6 hr and normalizing modified LMI by V_{lim} rather than by V_{pot} yields the cumulative distributions shown in Figure 15.12. To a large extent, the distributions collapse to single straight lines, suggesting that V_{lim} given by (15.3) is an appropriate bounding intensity. Note that relatively few storms are limited by landfall in the western North Pacific.

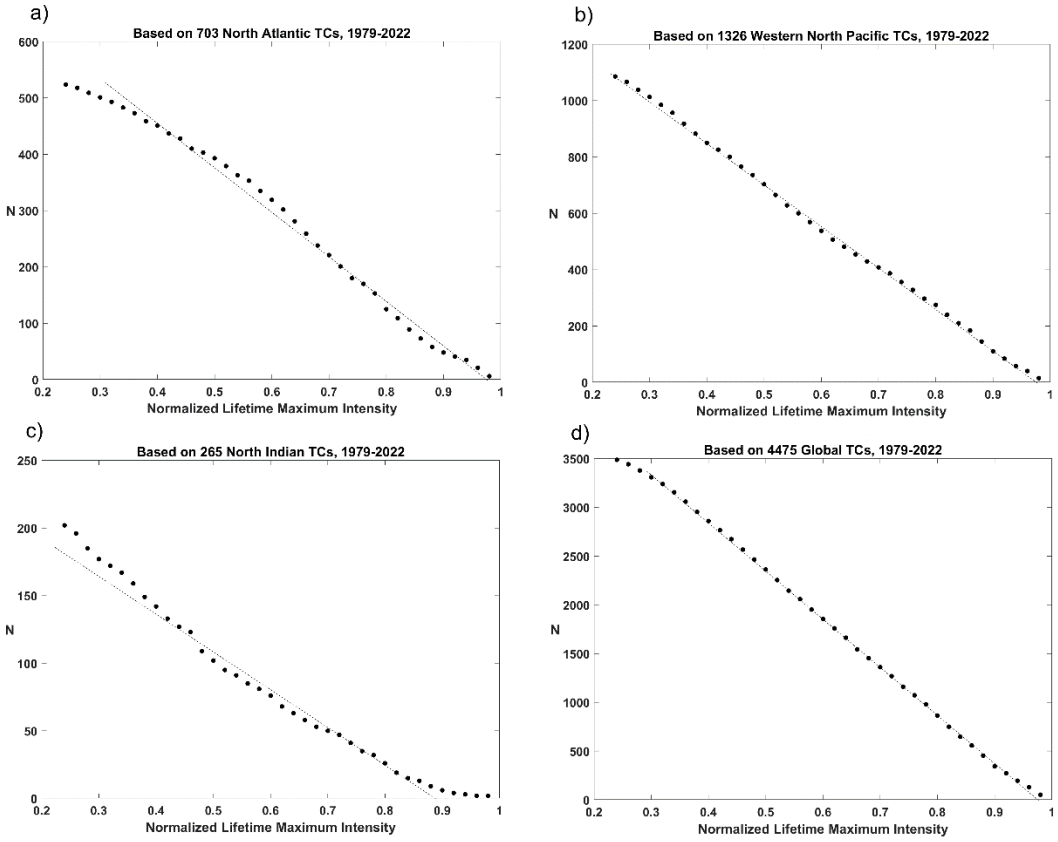


Figure 15.12: As in Figures 15.10 and 15.11 except that the LMIs are normalized by V_{lim}, as given by (15.3) rather than by the potential intensity (V_{pot}). The panels correspond to the North Atlantic (a), western North Pacific (b), North Indian Ocean, (c), and the whole globe (d).

Given the importance of potential intensity, we inquire about what aspects of the climate system control its magnitude and distribution. We approach this from the point of view of the vertically integrated heat budget of the ocean's mixed layer, which we described in Chapter 14, eliminating the surface thermodynamic disequilibrium term between the surface heat flux and the definition of potential intensity, resulting in equation (14.30), which we repeat here in a slightly different form:

$$d\rho_w C_l \frac{\partial T_s}{\partial t} = S + F_{IR\downarrow} - F_{IR\uparrow} - F_{ent} - F_h - d\nabla\bullet\mathbf{F}_{adv}, \tag{15.4}$$

where T_s is the surface temperature, T_t is the tropopause temperature, d is the ocean mixed layer depth, S is the net downward surface solar flux into the ocean, $F_{IR\downarrow}$ is the surface downward infrared flux, $F_{IR\uparrow}$ is the upward infrared flux from the sea surface, F_{ent} is the entrainment of enthalpy into the mixed layer owing to turbulence at its base (we define this to be a positive number so that in (15.4) it is a negative effect), $\mathbf{F}_{adv}$ is the horizontal enthalpy flux in the ocean mixed layer, C_D is the drag coefficient, ρ_a is the surface air density, and $|\mathbf{V_s}|$ is the surface wind speed (conventionally taken at 10 m altitude).

It follows that at the peak of the seasonal cycle, or averaging over a sufficiently long time scale, potential intensity is controlled mostly by surface solar and net infrared radiative flux, entrainment of cooler water through the base of the mixed layer, convergence of lateral ocean heat flux and, very importantly, surface wind speed. (Note that at very low surface wind speeds, much of the turbulence may be convectively driven and the neutral aerodynamic flux formula may be inaccurate.)

If we put aside, for the moment, the ocean-side fluxes F_{ent} and $\mathbf{F}_{adv}$, potential intensity would be controlled by surface radiative fluxes and wind speed. Increasing solar irradiance will increase potential intensity, both directly through the S term in (15.4) and indirectly, because increasing S will likely increase atmospheric temperature and, consequently, water vapor, increasing the downward infrared flux. Although the upward infrared flux also increases in radiative-convective equilibrium (RCE) states, it does not increase as fast as the downward flux. In the limit of high temperature, $F_{IR\downarrow}$ may gradually approach the upward flux $F_{IR\uparrow}$ from below, but is not likely to exceed it as the lower atmosphere is not likely to become warmer than the surface. (See Pierrehumbert 2010 for a discussion of this point.) It follows that at high temperature, in the absence of ocean-side enthalpy fluxes, potential intensity approaches the limiting quantity

$$V_p^2 \simeq \frac{T_s - T_t}{T_t} \frac{S}{C_D \rho_a |\mathbf{V_s}|}. \tag{15.5}$$

For a surface temperature of 300K, tropopause temperature of 200K, seasonal peak tropical surface solar irradiance of $250\,Wm^{-2}$, surface drag coefficient of 1.5×10^{-3} and an average surface wind speed of $5\,ms^{-1}$, (15.5) gives an upper bound potential intensity of about $125\,ms^{-1}$, somewhat larger than values calculated from observations, reflecting mostly the missing infrared terms.

Naturally, to calculate realistic solar and especially infrared fluxes, it is necessary to solve for at least radiative and convective heat transfer. To accomplish this, we turn to the RCE model presented in Chapter 3. Recall that this is a single-column model including a slab ocean mixed layer, which solves equations of radiative transfer and employs parameterizations of both moist and dry convection. We begin with a control experiment as described in Chapter 3, running the model into a state of RCE representative of a tropical ocean environment. We then calculate potential intensity for this and all subsequent simulations using the algorithm of Bister and Emanuel (2002). Next, we perform new RCE calculations in which we vary surface wind speed from its control value of $5\,ms^{-1}$ in increments of $1\,ms^{-1}$ down to $1\,ms^{-1}$; CO_2 concentration, from its control value of 360 ppm by successive doublings up to 11,520 ppm; and the solar constant from its control value of $1360\,Wm^{-2}$ in increments of $20\,Wm^{-2}$ up to $1480\,Wm^{-2}$. To simulate the effects of ocean heat flux, we simply vary the surface temperature. Details may be found in Emanuel and Sobel (2013). Rather than plotting the results as a function of these external parameters, we plot them, in Figure 15.13a, as a function of the RCE surface temperature.

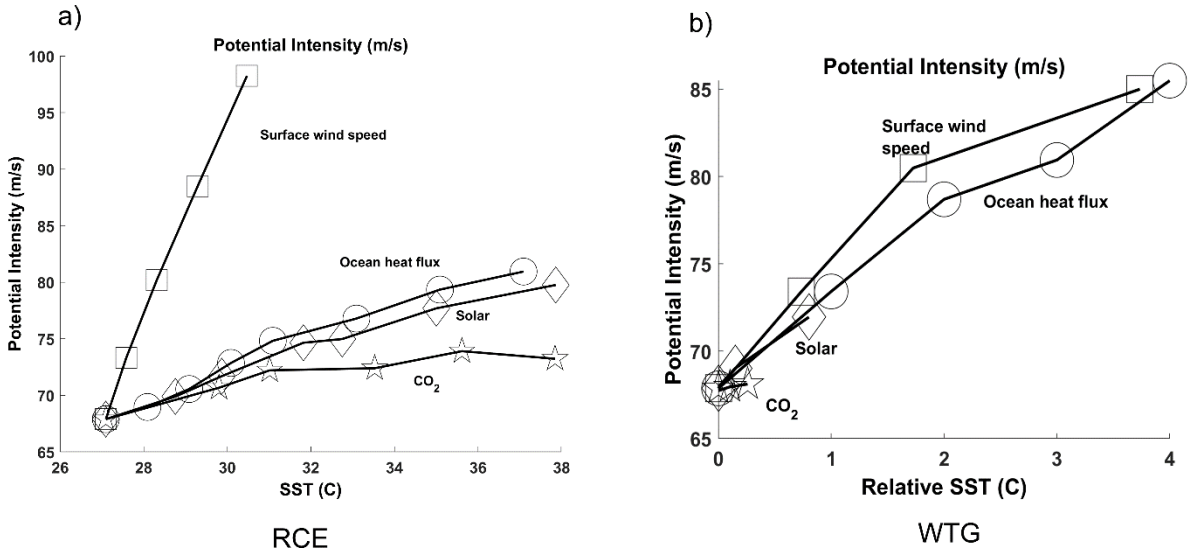


Figure 15.13: Potential intensity in radiative equilibrium states (a) and in weak-temperature-gradient (WTG) simulations (b), plotted as a function of absolute SST (a) and SST relative to the RCE state (b). Note that SSTs are also solutions of the single-column model. The experiments represent the results of varying the imposed surface wind speed (squares), ocean heat flux convergence (circles), solar constant (diamonds), and CO_2 concentration (stars).

The first thing to notice in Figure 15.13a is the very different slopes of the lines, depending on what external parameters have been varied; potential intensity is not a universal function of SST. Potential intensity varies most with SST when wind speed is varied and least when CO_2 is changed. Varying the solar constant and ocean heat flux convergence produces similar results; the slightly weaker slope of the solar curve is likely due to greater absorption of solar radiation by increased water vapor in warmer atmospheres, reducing the amount of sunlight reaching the surface. As predicted, increasing CO_2 also increases potential intensity, up to a limiting value. In fact, increasing CO_2 beyond this limit results in decreasing potential intensity, once again owing to increased absorption of sunlight by water vapor in warmer atmospheres.

In a second set of experiments, SSTs are allowed to respond only in a limited region, such that the free troposphere above 850 hPa is not allowed to vary from the control state, applying the weak-temperature gradient (WTG) approximation. The results are displayed in Figure 15.13b. Note that the axis scales are quite different from those in Figure 15.13a.

In these experiments, the saturation moist static energy above 850 hPa is fixed, because the temperature is fixed. If moist convective neutrality is maintained, the moist static energy of the boundary layer should be nearly equal to the (constant) saturation moist static energy of the free troposphere. Therefore, variations in the thermodynamic disequilibrium in the potential intensity must arise solely from variations in k_0^*, which depends only on SST. That is why all the curves in Figure 15.13b are nearly the same. What differences do exist may arise from the thermodynamic efficiency term in the potential intensity, given that there are differences in tropopause temperature.

In summary, changing surface wind speed causes large changes in potential intensity whether the changes are local or global. But local variations in surface solar flux and ocean heat flux convergence are somewhat more effective in changing potential intensity, per unit SST change,

than are global changes. Increasing long-lived greenhouse gases increases potential intensity, but only slowly. In contemplating the response of tropical cyclones to global warming, it is likely that the most important changes in potential intensity will arise indirectly from, e.g., changes in ocean circulations and cloudiness; the pattern of change may be more effective than the magnitude of the mean change in tropical ocean temperatures.

15.4.2: Normalized mid-tropospheric moisture

Besides potential intensity, mid tropospheric moisture is another strong determinant of tropical cyclone activity and constitutes and important part of genesis indices (see Chapter 14, section 14.2.1.) A non-dimensional measure of saturation deficit in the middle troposphere is χ, given by (14.5), which we repeat here:

$$\chi \equiv \frac{s^* - s_{600}}{s_0^* - s^*}, \tag{15.6}$$

where s^* is the saturation entropy of the troposphere (which is nearly constant with height in a moist adiabatic atmosphere but which is in practice evaluated at 600 hPa), s_{600} is the actual moist entropy at 600 hPa, and s_0^* is the moist entropy of air saturated at sea surface temperature and pressure. It appears as a factor in the ventilation index given by (14.1) and in the genesis index given by (14.4); it also falls out naturally in the non-dimensionalization of a numerical tropical cyclone model (Emanuel 1989, 1995). The larger its value, the longer it takes a given process to saturate the troposphere, an apparent necessary condition for genesis (Chapter 14). For developed storms, larger environmental values of χ make it possible to advect lower entropy into the storm's core when vertical shear is present.

We can use the same single-model framework we used in the previous section to explore how changing boundary conditions for RCE and WTG conditions affect χ. For this purpose, rather than using the 600 hPa moist entropy, s_{600}, in (15.6) we use the lowest value of moist entropy in the vertical profile; this will generally occur in the middle troposphere. Figure 15.14 follows the same format as Figure 15.13 but pertains to values of χ.

Increasing the temperature of the atmospheric column, whether by increasing CO_2, solar flux, or ocean heat flux convergence, warms the atmosphere and, assuming the relative humidity does not change much, increases the saturation deficit that constitutes the numerator of χ in (15.6). The air-sea thermodynamic equilibrium, comprising the denominator of (15.6), increases with increasing solar flux and ocean heat flux convergence but, for the reasons discussed in section 15.4.1, saturates as CO_2 is increased. Consequntly, χ increases more rapidly with SST when CO_2 is increased.

On the other hand, reducing surface wind speed cools the atmospheric column, reducing the saturation deficit and thereby decreasing χ.

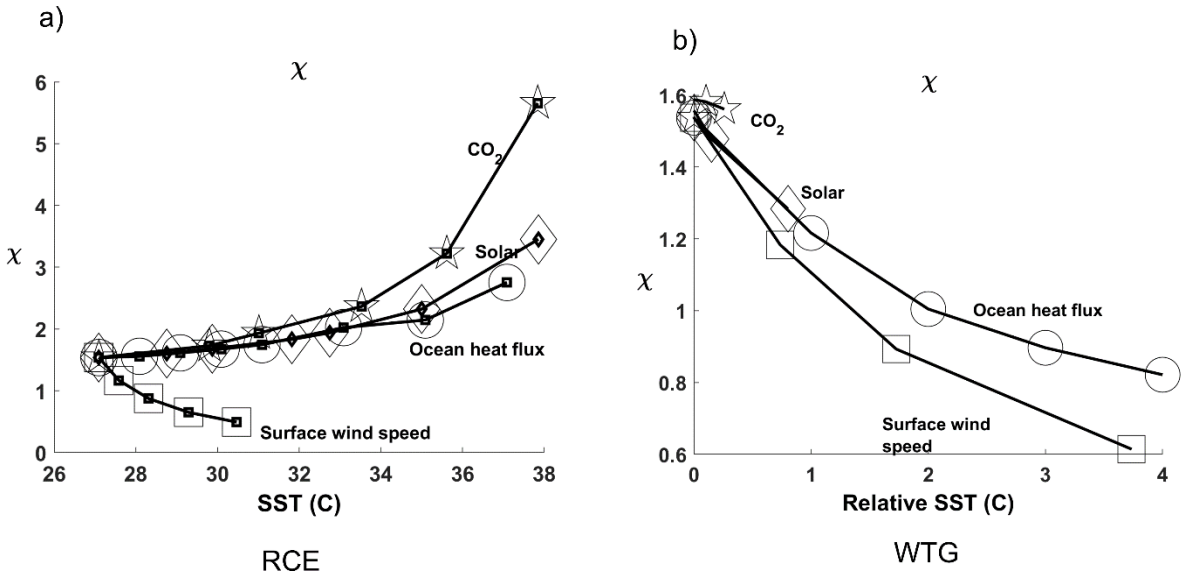


Figure 15.14: Nondimensional middle troposphere saturation deficit χ in radiative equilibrium states (a) and in weak-temperature-gradient (WTG) simulations (b), plotted as a function of absolute SST (a) and SST relative to the RCE state (b). Note that SSTs are also solutions of the single-column model. The experiments represent the results of varying the imposed surface wind speed (squares), ocean heat flux convergence (circles), solar constant (diamonds), and CO_2 concentration (stars).

In stark contrast, increasing any of the four driving factors decreases the saturation deficit χ in WTG experiments. As relative SST increases, upward motion ensues, moistening the column, and this outweighs increases in the air-sea-thermodynamic disequilibrium in the denominator of (15.6).

15.4.3: Thermodynamic component of a genesis potential index

We can combine the effects saturation deficit and potential intensity on genesis frequency by making use of just the thermodynamic part of the semi-empirical genesis index (GPI) given by (14.4). This thermodynamic contribution to GPI is then

$$GPI_{thermo} \equiv \chi^{-4/3}\left(V_{pot} - 35 ms^{-1}\right)^2 . \qquad (15.7)$$

The response of this component of the GPI to changing external influences on RCE and WTG states is displayed in Figure 15.15. Remarkably, in the case of RCE, the drying of the middle troposphere more than cancels increases in potential intensity for all but the changing surface wind speed experiments. In particular, increasing CO_2 lowers the thermodynamic component of GPI. In the case of decreasing surface wind speed, both components of GPI_{thermo} increase, yielding a large increase in their product.

Local forcing, as represented in the WTG experiments, has a very different effect, increasing GPI_{thermo} across all four forcing changes: decreasing wind speed and increasing solar, CO_2 and ocean heat flux convergence. In these cases, changes in χ and V_{pot} work in the same direction (see Figures 15.13b and 15.14b). Increasing ocean temperature while holding atmospheric

temperature fixed increases surface thermodynamic disequilibrium, increasing V_{pot}, while the resulting ascent moistens the column, decreasing χ.

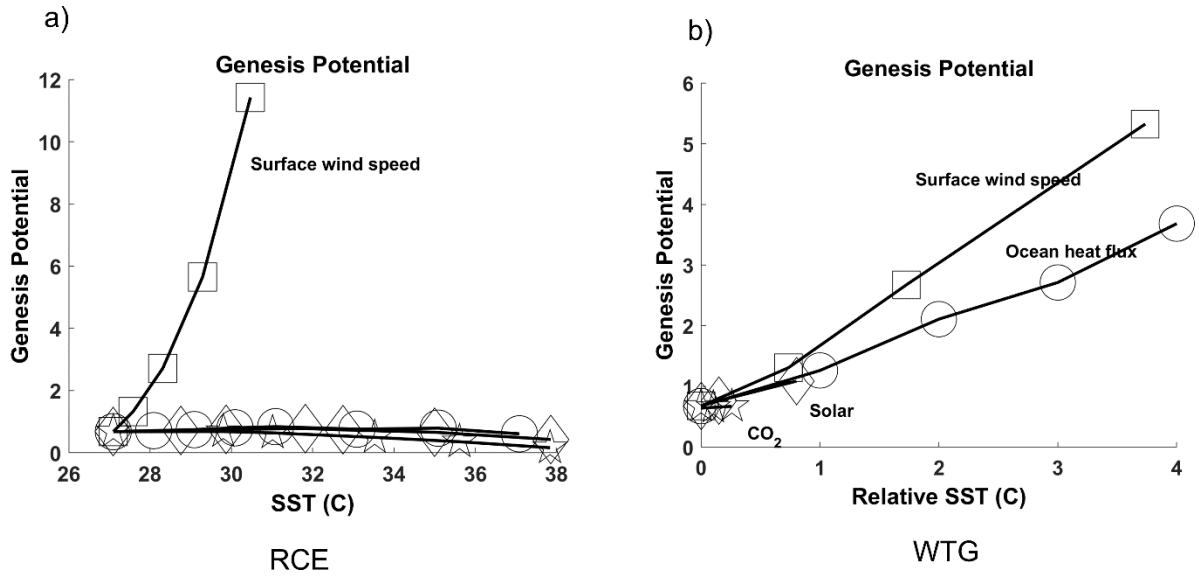


Figure 15.15: Thermodynamic component of the genesis potential index GPI_{thermo}, given by (15.11), in radiative equilibrium states (a) and in weak-temperature-gradient (WTG) simulations (b), plotted as a function of absolute SST (a) and SST relative to the RCE state (b). Note that SSTs are also solutions of the single-column model except when ocean heat flux is present. The experiments represent the results of varying the imposed surface wind speed (squares), ocean heat flux convergence (circles), solar constant (diamonds), and CO_2 concentration (stars).

To the extent that the thermodynamic component of the GPI represents the actual potential for tropical cyclogenesis, these results suggest that uniform warming has little effect on genesis; if anything, decreasing the frequency of TCs, unless some or all the warming is brought about by decreasing surface winds. On the other hand, climate change that enhances existing spatial patterns of SST, particularly if warm pools become warmer, may increase the incidence of tropical cyclones. One interesting way of increasing the spatial variance of tropical SSTs is to increase the magnitude of ocean heat transport, making warm pools warmer and cold pools colder. This argument assumes, though, that there are no large changes in the kinematic contributions to GPI, particularly wind shear.

15.4.4: Wind shear

Wind shear is generally detrimental to the formation and intensification of tropical cyclones, as discussed in Chapters 12 and 14. It has a major influence on the evolution of almost all TCs and is a strong concern of TC forecasters. Wind shear is associated with large-scale, slowly-evolving circulations such as monsoons and the Hadley and Walker cells, as well as smaller scale transient circulations such as the MJO and other equatorially disturbances and easterly waves. In certain regions such as the Caribbean, shear is strongly modulated by ENSO.

Yet at longer time scales, variations in GPI are strongly dominated by variations in its thermodynamic components. To show this, we form a time series of the natural logarithms of the various components (except for vorticity) of the GPI given by (14.4), consisting of monthly

averages of each of the terms averaged over the western North Pacific genesis region[1] and over a span of 41 years from 1979 to 2019. Recall that the GPI was optimized using monthly mean values, to best capture the geographic and seasonal variability of tropical cyclogenesis. Thus, the time series are 41x12 = 492 months long. First, we calculate the average the annual cycle over the 41-year period and subtract that from each 12-month sequence to de-seasonalize the time series. We next calculate the spectral power density of the time series, after detrending them. Finally, we perform a 4-point running mean of the power as a function of frequency. The result is displayed in Figure 15.16.

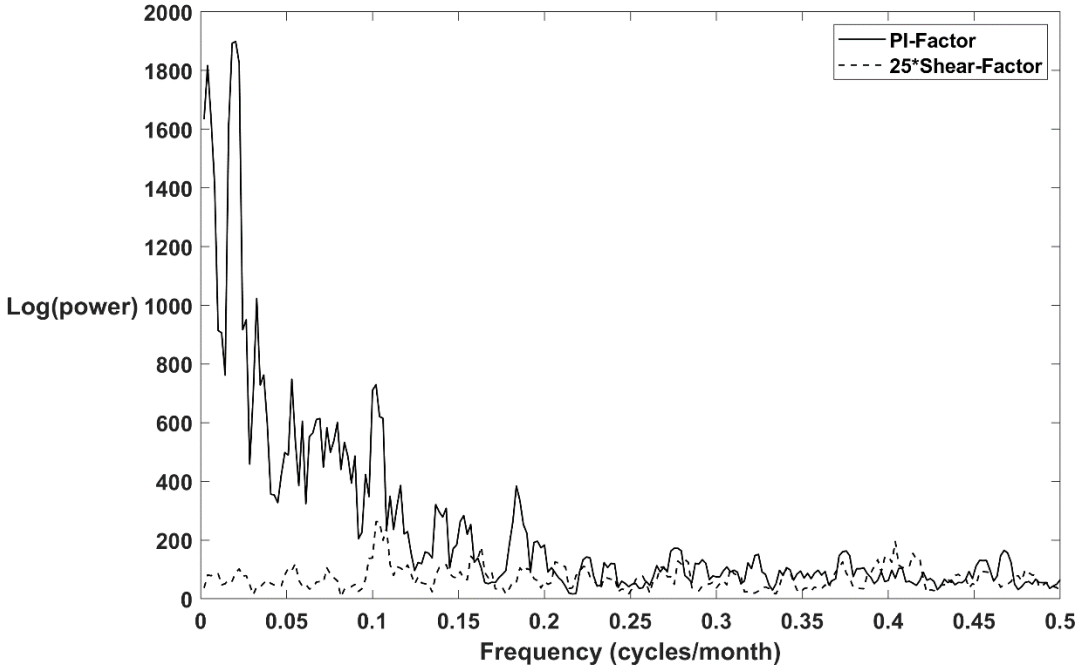


Figure 15.16: Power spectrum of variations in 2log(max(V_{pot}-35,1.001))(solid), and 4log(V_{shear}+25)(dashed). The latter has been multiplied by 25 for comparison.

Even after multiplication by a factor of 25, fluctuations in the log of the shear factor are smaller than those of the potential intensity and the latter has a much redder spectrum at time scales more than a year. The power in fluctuations of potential intensity has a strong spike at periods of 3-5 years, probably corresponding to ENSO. There is no prominent ENSO peak in the power spectrum of wind shear.

Figure 15.17 compares the same power spectrum of potential intensity to that of the χ factor in (14.4); note that the power has been multiplied by 10 Curiously, the power spectrum of χ is quite like that of potential intensity, except that there is far less power at periods of 1-2 years. The ENSO spike is clear in this power spectrum.

[1] Defined at the region between 5^o and 15^o N, and 130^o and 180^o E.

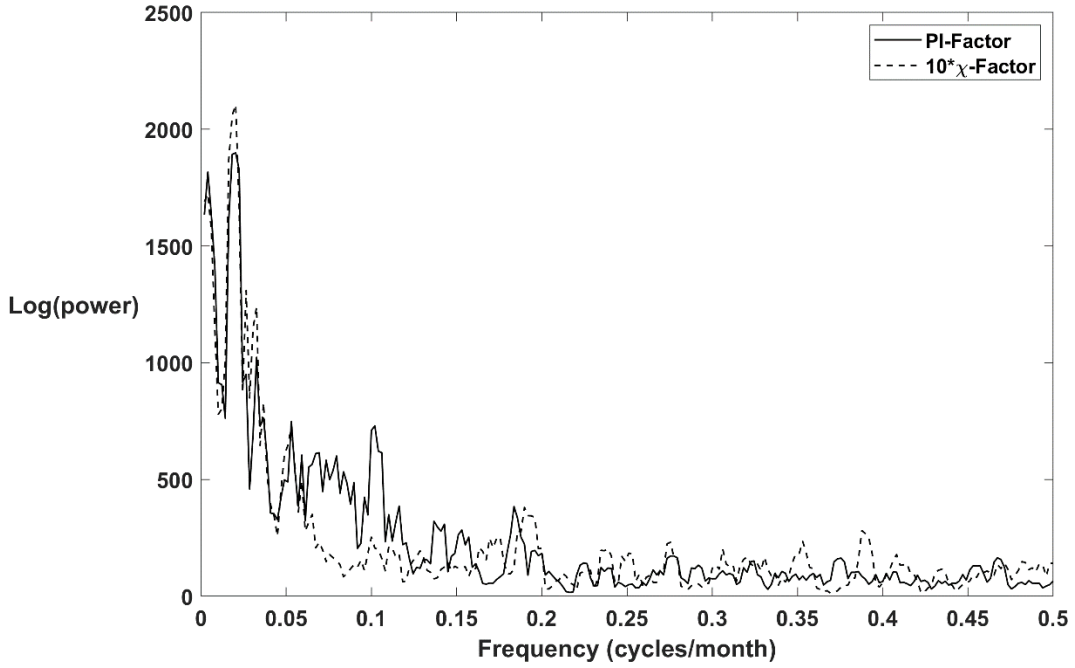


Figure 15.17: Power spectrum of variations in 2log(max(V_{pot}-35,1.001))(solid), and 1.33log(χ)(dashed). The latter has been multiplied by 10 for comparison.

We conclude from this that while changing shear exerts a crucial control on tropical cyclones at weather forecasting time scales, it does not seem very influential at times scales longer than a year. Since this analysis excludes the annual cycle, we cannot discount the possibility that modulations in the amplitudes of seasonal phenomena, like monsoons, could influence TCs through their effects on shear. Indeed, repeating the exercise for the North Indian Ocean (not shown here) reveals a strong signal in shear at annual time scales, expressing a modulation of the annual cycle.

15.4.5: The importance of the spatial distribution of tropical SST

At some level, the distribution of tropical temperatures must play a role in the location and magnitude of tropical cyclone activity. To take an extreme example: Suppose there were a strong equatorial peak in SST such that virtually all the convection in the deep tropics were concentrated at the equator. Owing to the lack of vorticity there, few TCs would form. If this super-ITCZ were displaced off the equator, one would expect TC activity to increase.

Figure 15.18 shows the spatial distribution of the logarithm of the potential intensity factor in the GPI given by (14.4). The potential intensity is derived from ERA-5 reanalyses and averaged over the period 1980-2020 and over the month each year at each grid point that has the maximum value of potential intensity over the 12 months of that year. The spatial distribution closely mirrors that of the underlying sea surface temperature, with broad, flat warm pools in the tropical Indian and western Pacific oceans, the Caribbean and Gulf of Mexico, and the eastern Pacific off Mexico. The steep shoulders of the distribution reflect the fast decline of the thermodynamic efficiency factor in the potential intensity as one travels from the tropics to

higher latitudes. The eastern Pacific equatorial cold tongue is readily apparent, and there is a weaker cold patch on the equator in the Atlantic.

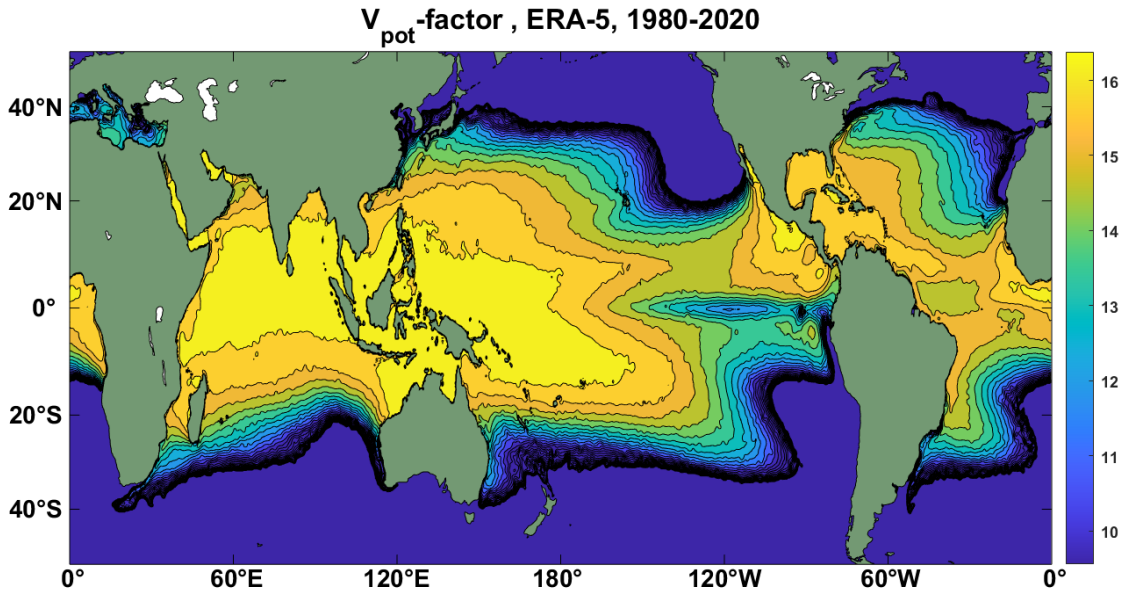


Figure 15.18: Spatial distribution of the annual maximum of the logarithm of the potential intensity factor in (14.4), from ERA-5 reanalyses, averaged over 1980-2020.

Compare this to the distribution of the logarithm of the χ factor in (14.4), shown in Figure 15.19. This has been calculated from the same reanalysis data, and over the month each year at each grid point that has the maximum value of potential intensity (not χ) over the 12 months of that year. Note that there is much more fine-scale structure in the χ field than in potential intensity, reflecting the fact that the variance of moisture in subsaturated atmospheres is not constrained by the those dynamics that enforce slowly varying virtual temperature. More importantly, the maximum χ factor in the central and eastern North Pacific is much more strongly concentrated than the maximum in potential intensity. The reasons for this were discussed in Chapter 5: Lateral advection of relatively low moist static energy in the boundary layer quenches convection at the outer boundaries of the region of high SST, and the convection is further concentrated by reduced gross moist stability and cloud-radiation feedbacks. Mid-level moisture tends to follow the deep moist convection. Since χ is an important factor in TC genesis and also plays a role in ventilation, its distribution is important for TC climatology.

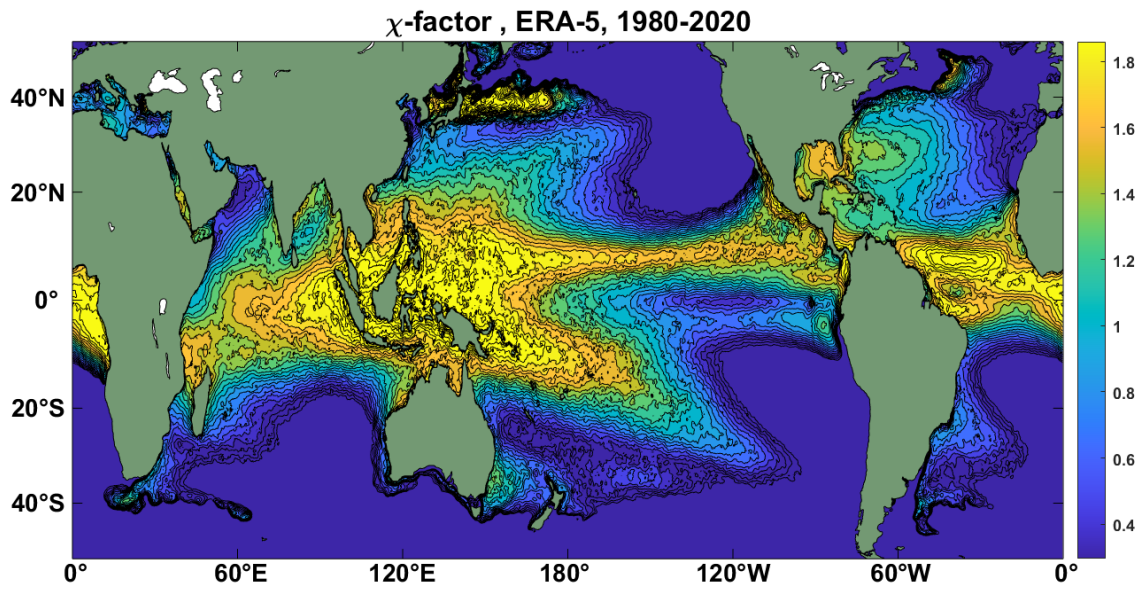


Figure 15.19: Spatial distribution of the logarithm of the χ factor in (14.4), from ERA-5 reanalyses, averaged over 1980-2020. In each year and at each grid point, χ is taken from the month of that year with the largest potential intensity.

Aquaplanet simulations show a strong correspondence between the latitude of the ITCZ and the frequency of tropical cyclones (Merlis et al. 2016; Viale and Merlis 2017), perhaps reflecting the shift of mid-level moisture maxima toward or away from the equator. As mentioned previously, analysis of oxygen isotopes in stalagmites in two Australian caves, one measuring the ITCZ position and the other the frequency of TCs, suggests that TCs are more frequent when the ITCZ is further south, over a 1500 year record (Denniston et al. 2023).

15.5: Historical trends in environmental parameters

The power spectra presented in Figures 15.15-15.17 were developed from detrended, de-seasonalized monthly time series. Here we examine whether there are discernible trends in the factors that make up the genesis index defined by (14.4). To detect such trends, we calculate the natural logarithm of the various factors in (14.4) from gridded reanalysis data, using the European Centre for Medium-Range Forecasts CERA-20C reanalysis set (Laloyaux et al. 2018), which extends from 1901 to 2010. This reanalysis differs from conventional reanalyses in that it assimilates no upper-air data but only surface pressure, surface air temperature, sea surface temperature, sea ice, and marine surface winds into a coupled atmosphere-ocean model. The idea is to avoid as much as possible the inhomogeneities introduced by rapidly evolving observations of the atmosphere above the surface, brought about by the introduction and evolution of rawinsondes, aircraft observations, and satellites. The gridded data are on a 1.5-degree global grid and, in keeping with the design of the genesis index, we use monthly mean data. In the first instance, the shear, potential intensity, vorticity, and χ are averaged over the Atlantic main development region, 6^{o}-18^{o} N, 20^{o}-60^{o} W and averaged from August through October.

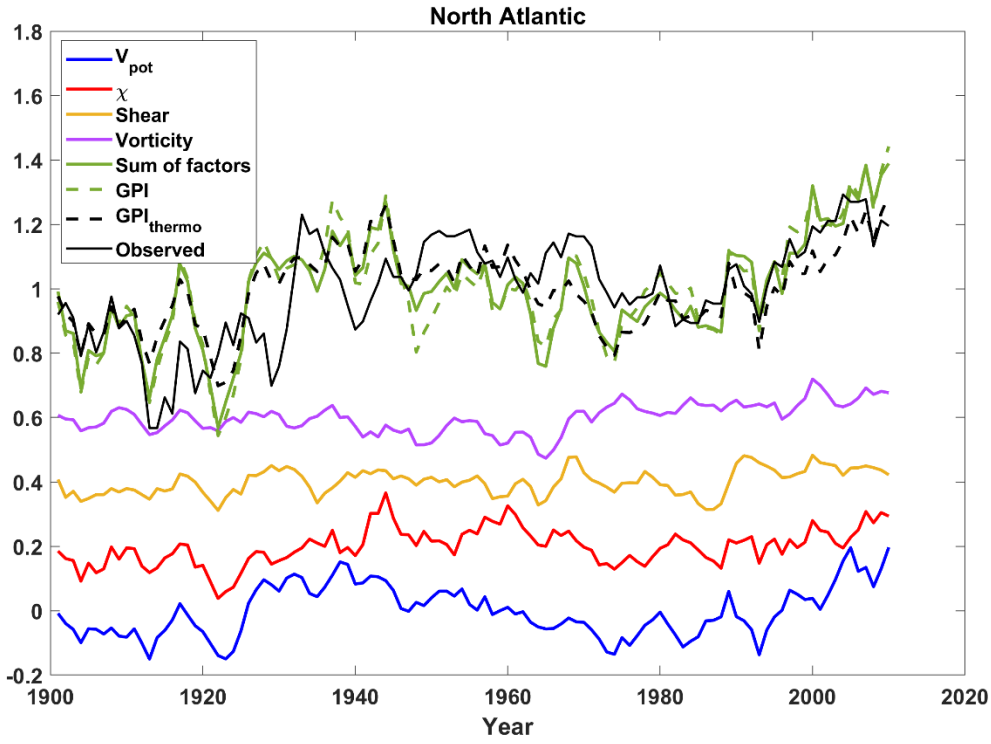


Figure 15.20: Logarithms of each of the terms in (14,.4) from monthly mean CERA-20C reanalysis data averaged over 6º-18º N, 20º-60º W and the months August-October. The long-term means of the natural logarithms are subtracted from each series, and the various series are offset from each other by increments of 0.2 for ease of comparison. The time series have been smoothed with a 4-year running average. Note that the signs of the exponents in (14.4) are preserved, so that positive values of each factor are positive contributions to GPI. The solid green line shows the sum of the 4 factors whereas the dashed green line displays the logarithm of the GPI calculated from (14.4) before averaging. The dashed black line shows the sum of just the thermodynamic contributions, V_{pot} and χ. The solid black line shows the natural logarithm of the observed North Atlantic tropical cyclone annual frequency from IBTrACS (Knapp et al. 2010), multiplied by 0.5, smoothed with a 4-year running average, and offset.

Once we calculate the natural logarithms of the factors in the GPI given by (14.4), we subtract from each its long-term mean value and plot the time series of each, offset by increments of 0.2 for ease of comparison. The results are displayed in Figure 15.20, which also shows the sum of just the thermodynamic contributions to the GPI (solid green line) and 0.5 times the natural logarithm of the observed North Atlantic annual tropical cyclone frequency from IBTrACS data (Knapp et al. 2010). All series have been smoothed with a 4-year running average.

First note that the sum of just the thermodynamic contributions to GPI (dashed black curve in Figure 15.20) is a reasonably good approximation to the full GPI. Of the two contributions, that from potential intensity is the more important. In this case, the small contribution from the kinematic terms arises mostly from variations in the absolute vorticity, not from shear. While the shear term does make a non-trivial contribution to the GPI on time scales up to a few years, it does not, in this case, contribute to the long-term upward trend.

Note also that the natural logarithm of the GPI, whether calculated directly or as the sum of the four contributions to it, is a reasonably good prediction of the natural logarithm of the observed North Atlantic tropical cyclone count. Here we have included only storms whose lifetime maximum intensity equals or exceeds $17\ ms^{-1}$. There is much controversy over the quality of the observed record of North Atlantic TCs much before 1970 (e.g. Vecchi and Knutson 2011), but widespread agreement that the records are unreliable before about 1980 in all other parts of the world oceans.

The same type of analysis applied to four other ocean basins is displayed in Figure 15.21.

Figure 15.21: As in Figure 15.20 except for a) the eastern North Pacific (5º – 18º N, 90º – 170º W) averaged over June - September, b) the western North Pacific (5º – 15º N, 130º – 180º E) averaged over July - November, c) the North Indian Ocean (5º – 20º N, 50º – 110º E) averaged over April - November, and d) all of the southern hemisphere (5º – 18º S, 60º – 180º E) averaged over January through May.

The net GPI outside the Atlantic shows no significant trend, except in the North Indian Ocean, where there is a weak upward trend. In the western North Pacific, there are significant trends in potential intensity and χ, but they oppose one another. Likewise, In the eastern North Pacific, there are significant trends in vorticity and shear, but they also tend to cancel. Of the five ocean basins considered here, the North Atlantic has the most significant multidecadal variability and trend.

15.6: Aerosol effects on tropical cyclones

Our atmosphere contains a highly diverse and variable population of aerosols. These arise naturally from volcanoes, wildfires, wind lofting of surface material, and evaporation of sea spray, leaving particles comprised of salt and organic compounds. Today, there are also strong anthropogenic aerosols sources from fires set for land clearing and, especially, from fossil fuel combustion. Secondary aerosols arise from chemical reactions involving gases emitted from the surface, including vegetation. Aerosols may be comprised of liquids, solids, or both.

Aerosols in the troposphere have characteristic residence times of a few weeks and are removed both by direct deposition on the surface and by precipitation. Aerosols injected into the stratosphere by volcanoes and pyrocumulonimbus (deep moist convective clouds driven by surface fires) may reside there for a few years.

Aerosols may affect tropical cyclone environments in several ways: by altering the optical properties of clouds, by changing the environmental moisture distribution, and by changing the thermodynamic disequilibrium at the sea surface through their effect on solar and longwave radiation.

Aerosols may more directly affect TCs through their effects on cloud microphysics, though this topic remains somewhat controversial. Tropospheric aerosols usually contribute to cloud condensation nuclei (CCN), and increasing CCN concentration generally results in a larger population of smaller cloud droplets. This usually decreases the precipitation efficiency of stochastic coalescence and leads to a decline in precipitation efficiency (Cotton et al. 2012). But it also may result in increased lofting of cloud droplets into the cold regions of clouds, enhancing latent heating by fusion and possibly making anvil clouds larger and more optically opaque to infrared radiation. Precipitation falling from higher in the cloud may be exposed to longer trajectories through unsaturated air, further reducing precipitation efficiency and causing stronger cold pools. But the increased evaporation of cloud and rain water, together with enhanced surface fluxes in stronger, colder outflow in the boundary layer, serve to moisten the column (Abbott and Cronin 2021), reducing the thermodynamic resistance to genesis as expressed by the nondimensional number χ given by (14.5). This, together with greater cloud-radiation feedback, could conceivably promote genesis.

Here we describe the somewhat better-known optical effect of aerosols acting on the large-scale environment. Some materials, like sulfates, are highly reflective and generally exert a cooling effect by back-scattering solar radiation to space. Aerosols containing compounds like black carbon can absorb sunlight and heat the atmosphere. Most real aerosol particles are composites made of two or more constituents. For example, particles may contain a mixture of inorganic ions, carbonaceous material, crustal elements, and water (Kocifaj et al. 2006).

15.6.1: Tropospheric aerosols and the late 20th century hurricane drought

The relatively short residence time of most aerosols in the troposphere means that their effects are concentrated within a distance of a few thousand kilometers from their source. Their effects on, e.g., surface temperature are therefore likely to be more regional than global. Following the physics described in section 15.4.1, the effect on important environmental controls of TC

activity, like potential intensity and χ, are likely to be stronger than the direct effects of globally uniform forcing such as that owing to long-lived greenhouse gases like CO_2.

A case in point involves anthropogenic sulfate aerosols that result, through various chemical pathways, from combustion of fossil fuels, particularly those with high sulfur content, leading to emissions of sulfur dioxide (SO_2). Aerosols with high sulfate content are highly reflective and result in a cooling of the surface and atmospheric column. Emissions of SO_2 increased rapidly after World War II, reaching a peak in the late 1970s, and then decreased just as rapidly owing mostly to legislation limiting sulfur emissions in many nations. It is estimated that the net effect of sulfate aerosols since 1950 has been to cool the surface by 0.4 K relative to the warming that would have occurred because of increasing concentrations of long-lived greenhouse gases (IPCC-6 2023).

The cooling effect of anthropogenic sulfate aerosols was by no means globally uniform. Industrial emissions of SO_2 are highly concentrated in the northern hemisphere, and in highly populated industrialized regions such as North America, Europe, and China. In boreal summer, low-level airflow around the Azores anticyclone carries pollution from European sources southward across the Mediterranean and over northern Africa, thence westward across the tropical North Atlantic. Figure 15.22 shows the history of SO_2 emissions from Europe (Smith et al. 2011).

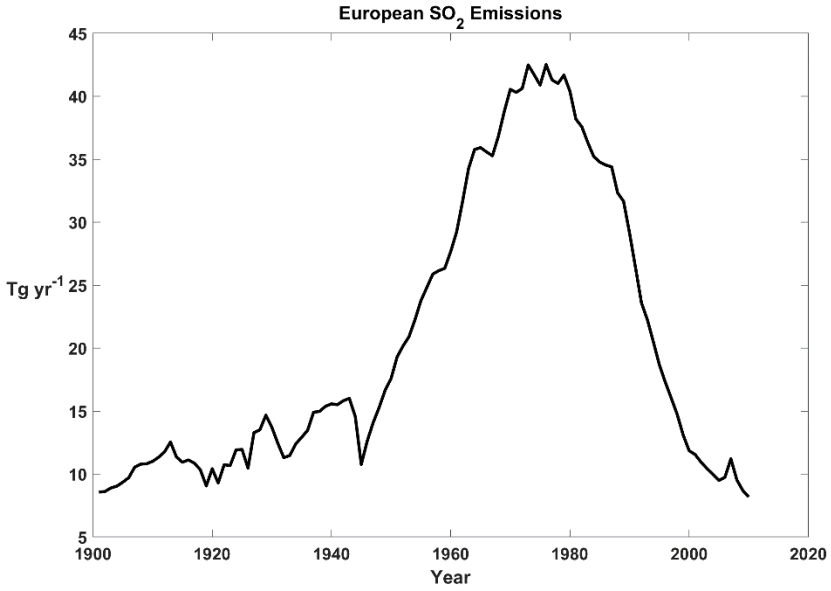


Figure 15.22: Sulfur dioxide emissions from Europe.

Given the short residence time of SO_2 in the troposphere, one may presume that concentrations of sulfate aerosol particles are strongly correlated with emissions on these time scales.

The high concentration of sulfate aerosols apparently had two strong effects on sea surface temperatures in the tropical North Atlantic (Rousseau-Rizzi and Emanuel 2022). First, the elevated concentration of sulfates over the tropical North Atlantic directly cooled that region. Second, the even higher concentrations over North Africa served to weaken the African summer

monsoon, greatly reducing rainfall in the Sahel region (Dong et al. 2014). The dry soils and elevated surface winds led to increased lofting of mineral dust (Prospero and Lamb 2003), which was then carried westward over the tropical North Atlantic by the prevailing easterly winds. The mineral dust also decreased solar radiation at the surface, more than doubling the direct effect of sulfate aerosols (Rousseau-Rizzi and Emanuel 2022).

Time series of the 10-year running mean of the logarithm of the potential intensity factor in (14.4) is compared to the negative of the logarithm of European SO_2 emissions, lagged by 2 years, in Figure 15.23. This shows a correspondence between the SO_2 emissions and depression of ocean temperatures in the main development region of the North Atlantic. As can be seen in the top of Figure 15.20, this was accompanied by a substantial depression of North Atlantic tropical cyclone activity.

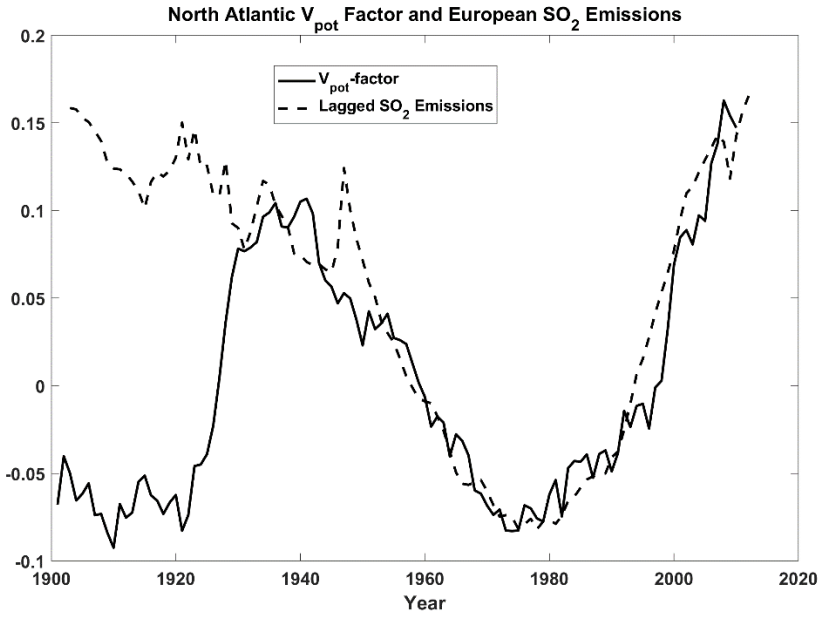


Figure 15.23: The natural logarithm of the potential intensity factor in the GPI given by (14.4), minus its mean over the record (solid), smoothed with a 10-year running average, compared to minus 0.15 times the log of the European SO_2 emissions (Tg yr^{-1}) plus 0.05. The SO_2 emissions have been lagged by two years.

15.6.2: Volcanic aerosols in the stratosphere

Volcanoes emit copious amounts of water vapor, carbon dioxide, sulfur dioxide, hydrogen sulfide, hydrogen fluoride, and hydrogen halides into the atmosphere, in addition to large quantities of ash and tephra. Average eruptions deposit all this material in the troposphere, where, as in the case of aerosols arising from other sources, it is removed by dry or wet deposition over a period of around two weeks. The eruptions may last from hours to years, but generally do not emit enough material to strongly affect atmospheric physics or dynamics.

Very strong eruptions produce plumes that can penetrate well into the stratosphere, injecting ash, water vapor and other gases. By various pathways, the SO_2 is converted to sulfate aerosols. As there is no precipitation in the stratosphere, the residence time for the aerosols is much longer… on the order of years. Sulfate is highly reflective and cools the troposphere and

surface. A recent example was the 1991 eruption of Mt. Pinatubo in the Philippines. The eruption sent a plume to an altitude of 40 km and ejected an estimated 17 million tons (about 15 Tg) of SO_2 into the stratosphere. The resulting sulfate increased the aerosol optical depth by between 10 and 100 times its value before the eruption, and produced a measurable surface cooling for almost 2 years, with a maximum magnitude of about 0.6 K.

While the surface cools, the lower stratosphere itself warms as a consequence of shortwave and longwave absorption by the aerosols (Robock 2000).

Stratospheric aerosols are rapidly homogenized in the east-west plane by mostly zonal winds, but take longer to spread meridionally. Unless the eruption is close to the equator, as was Pinatubo, the aerosols tend to stay in the hemisphere into which they were ejected. As with all radiative forcing, the pattern of the response may differ appreciably from that of the forcing. For example, there is some evidence from both observations and models that large tropical volcanic eruptions cause an El Niño-like response.

Looked at through the lens of a genesis potential index (GPI), large volcanic eruptions can conceivably affect genesis rates through alteration of potential intensity (both by surface cooling and near-tropopause warming), wind shear, and mid-tropospheric humidity. Yan et al. (2018) calculated GPI similar (but not identical) to that given by (14.4) to an ensemble of 1156-year global climate model simulations, comparing experiments with all known forcings to a set with volcanic forcings only. The latter were based on ice core estimates of aerosol loading (Gao et al. 2008). They found that large tropical volcanic eruptions in the tropics and the northern hemisphere lead to a significant suppression of tropical cyclone formation, particularly over the whole Pacific basin and the North Atlantic.

As an illustration, we calculated the GPI given by (14.4) from a single member of the ensemble used by Yan et al. (2018) from a simulation with volcanic forcing variability only. We summed the GPI over all of the tropics and each whole year and composited it over an interval of 11 years spanning three years before each eruption to seven years after. We included only tropical eruptions that are estimated to have released more than 15 Tg of SO_2, yielding a direct TOA radiative forcing of more than $2\,W\,m^{-2}$. Over the span of 1156 years, 29 eruptions met these criteria. Figure 15.24 shows the results. There is significant suppression of TC genesis for three years beginning with the year of the eruption. Yan et al. (2018) present a more detailed breakdown of which factors of the GPI contribute most to these changes, and they also show that southern hemisphere volcanoes tend to enhance, rather than suppress, TCs in the North Atlantic. Note that the changes in GPI are small enough that while they appear in composites, random noise or other climate forcing will often hide the volcanic signal accompanying any individual eruption.

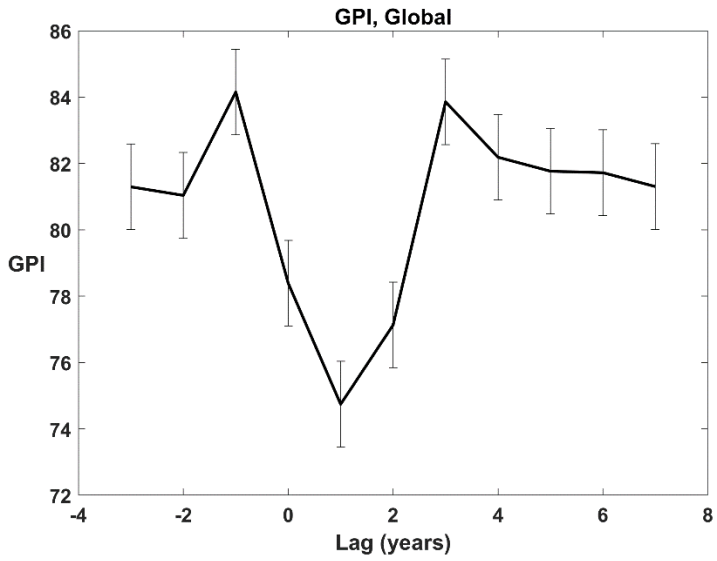


Figure 15.24: Genesis potential index (GPI) given by (14.4) and summed over the tropics (23.5°S to 23.5°N) and over all months of each year, normalized to an average of 80 TCs per year and composited around the year (year 0) of each of 29 volcanic eruptions whose sulfur emissions exceeded 15 Tg. Error bars show one standard deviation up and down based on the whole record of GPI. Climate model data from the National Center for Atmospheric Research CESM-LME model with volcanic-only forcing, as in Yan et al. (2018).

15.7: Greenhouse gas forcing: Global warming

Long-lived greenhouse gases, such as carbon dioxide (CO_2) and methane (CH_4), reside in the atmosphere on time scales that are much longer than atmospheric advective time scales and are therefore well homogenized, at least in the troposphere. However, the radiative forcing by such gases is by no means uniform. The top-of-the-atmosphere (TOA) forcing varies mostly with latitude because of the temperature-dependence of the forcing, while the direct surface radiative forcing by CO_2 and CH_4 is highly inhomogeneous because of the strong and spectrally overlapping absorption by water vapor and clouds (Huang et al. 2017). Referring back to section 15.4, it is apparent that the spatial pattern of the response to forcing, whether it is owing to the spatial variability of the forcing itself or to the spatial pattern of the multitudinous feedbacks, is much more important than the spatially averaged response. Consequently, the single-column response to greenhouse gas forcing, as illustrated in Figures 15.13-15.15, may be a poor guide to the actual global response.

Although it is tempting to focus on the usual large-scale predictors of TC activity, such as potential intensity, mid-tropospheric saturation deficit, and wind shear, one must also consider changes in storm tracks. It is possible that landfall of destructive storms may decrease even though global activity increases, or vice versa.

15.7.1: Behavior of TC-relevant environmental parameters

As part of the suite of experiments using global climate models participating in the Climate Model Intercomparison Project Sixth Assessment Report (CMIP6), which was conducted in

support of the Intergovernmental Panel on Climate Change Sixth Assessment Report (IPCC-6 2023), simulations were performed in which CO_2 was increased at the steady rate of 1% yr^{-1} for 150 years, up to a maximum of four times pre-industrial concentrations, with all other forcings remaining constant. Output of all CMIP6 simulations is publicly available.

To diagnose trends in important environmental quantities, we calculate the potential intensity, χ, and wind shear at each grid point using monthly mean air and sea surface temperatures, 600 hPa specific humidity, absolute vorticity (capped at 5×10^{-5} s^{-1}) and shear that is calculated as monthly averages of daily values. We do this for each month of each of the 150 years. We also combine these monthly means into a monthly mean genesis potential index (GPI), given by (14.4) but with the absolute vorticity capped, following Tippett et al. (2011). We then sum the GPI over all the months of the year, each year and at each grid point. For the potential intensity, χ, and wind shear, we ascribe to each year and each grid point the value of the quantity during the month of that year at which the potential intensity reaches a maximum value. Finally, we calculate the linear trend of each quantity at each grid point. We apply this procedure to each of ten climate models.

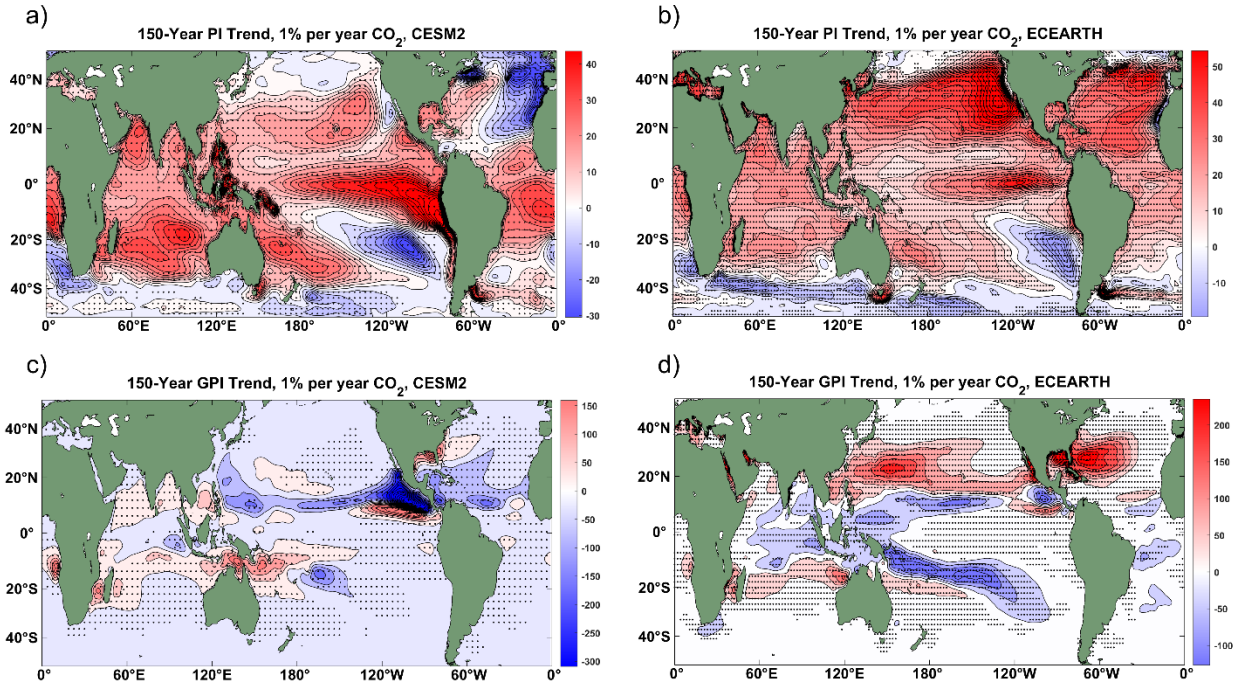


Figure 15.25: Linear trends of potential intensity (a and b, in m s^{-1} per century) and GPI given by (14.4) (c and d in percentage change per century relative to the first-year global mean of positive GPI values) for the NCAR CESM2 model (a and c) and the European Center EC-Earth3 model (b and d). Stippling denotes trends with p-values < 0.05.

Figure 15.25 shows maps of the linear trends of potential intensity (top row) and GPI (bottom row) from just two of the nine models; these two have the highest spatial resolution among the ten explored. Potential intensity increases in nearly all the TC genesis regions; decreasing only where it is already too small to support genesis, such as in the southeastern Pacific. In both models displayed in the figures, the eastern equatorial Pacific cold tongue warms up in the models, increasing potential intensity but in a region where it starts out small and, besides, the Coriolis parameter is too small to support genesis. This warming of the eastern equatorial

Pacific reflects a well-known tendency for global warming to increase the frequency of El Niño events in climate models, whereas observations show a declining frequency of such events (Sobel et al. 2023).

Apart from these commonalities, the patterns of potential intensity increase are quite different between these two models and indeed among the ten models explored here. Apart from the equatorial cold tongue, there is a general tendency for increases to be largest in the subtropics rather than the deep tropics. This may reflect the tendency for the values of potential intensity to saturate as CO_2 is increased, evident in Figure 15.13a. Across models, potential intensity generally increases more in the northern than in the southern hemisphere.

Trends in the GPI given by (14.4), shown in the bottom row of Figure 15.25, are highly inconsistent between the two models displayed, and indeed among all ten models. In CESM2, GPI decreases in the North Atlantic, except in the northern Gulf of Mexico and off the U.S. east coast, in the western North Pacific, and dramatically in the eastern North Pacific, while there are some increases across the southern hemisphere genesis regions. The pattern of GPI trends in the EC-Earth model is very different, with increases in the subtropical North Atlantic and western North Pacific and strong decreases in the South Pacific convergence zone.

To summarize the statistics of GPI and its factors over the ten models, we first calculate the base 10 logarithms of the GPI itself as well as the four factors that make it up, at each grid point and for each year, selecting the month of the year at each point and for each year as that with the largest potential intensity. From (14.4):

$$\begin{aligned}\log_{10}\left(GPI\right) = \log_{10}\left(C\right) + 3\log_{10}\left(\min\left(|\eta|, 5\times10^{-5}s^{-1}\right)\right) - \frac{4}{3}\log_{10}\left(\chi\right) \\ + 2\log_{10}\left(\max(V_{pot} - 35ms^{-1}, 0\right) - 4\log_{10}\left(1+0.04S\right).\end{aligned} \tag{15.8}$$

Now when we sum these over all the grid points, we exclude points at which any of the arguments of the logarithms is zero, which excludes points near the equator and all points at which the potential intensity is less than or equal to *35 ms^{-1}*. The summation yields time series of each factor as well as the total $\log_{10}\left(GPI\right)$, for each of the 10 models. For each model, we re-scale each factor in (15.8) by a common multiplicative factor such that the GPI itself is 80 at the initial time. Finally, we subtract the initial value of each factor and the total so that the series is relative to its initial value.

Figure 15.26 shows the evolution over the 150 years of each of the terms in (15.8), relative to their starting value. The bold curves show the multi-model means while the shading spans from the 10th to the 90th quantile among the ten models. The summed GPI increases in all ten models, though there is considerable scatter in the magnitude of the increase. The summed potential intensity also increases in all models, while the normalized mid-level saturation deficit, χ, increases (causing a negative contribution to GPI) in most but not all models, with very large uncertainty after the first 50 years or so.

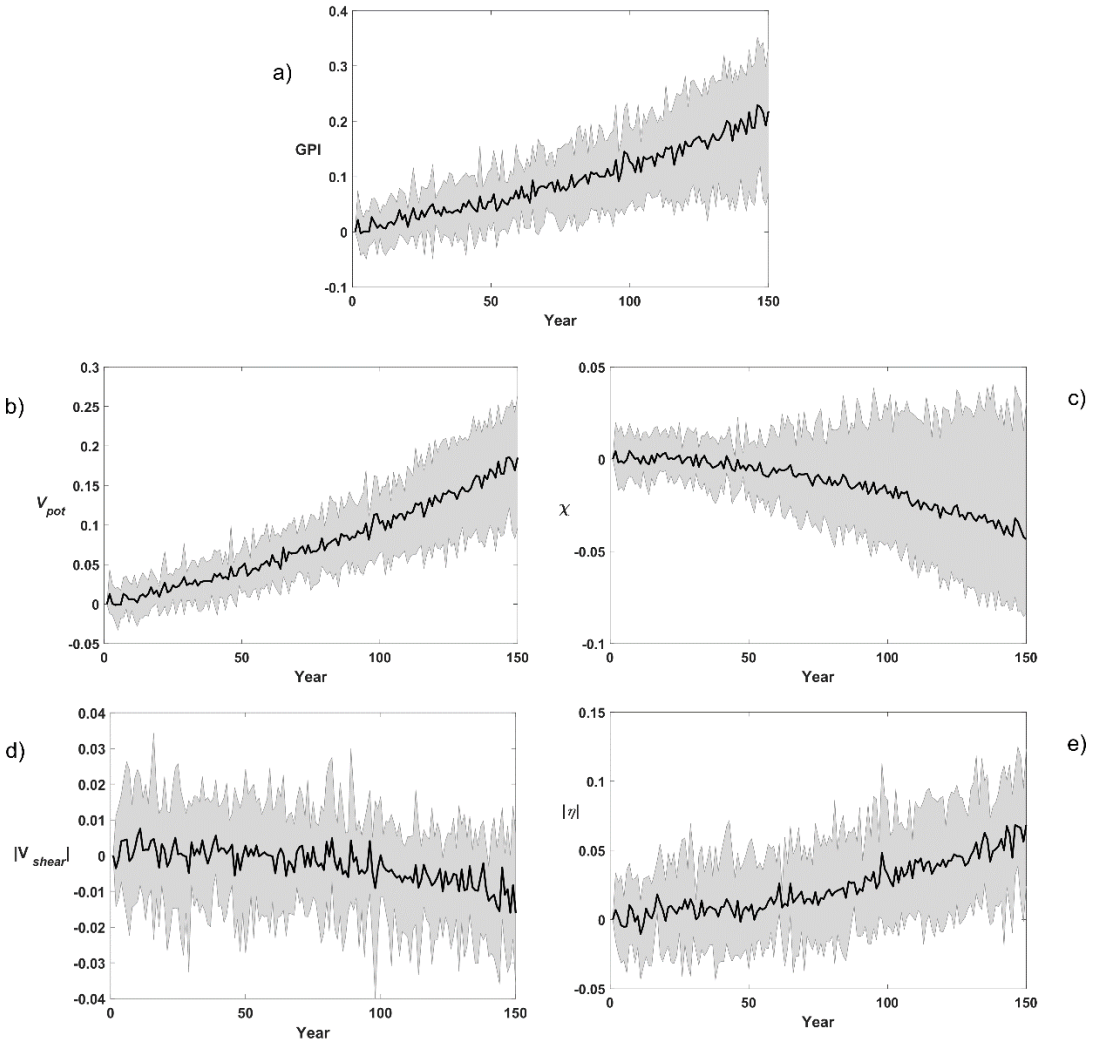


Figure 15.26: Evolution, over 150 years, of each of the terms in (15.8) relative to their initial values. a) GPI, b) potential intensity term, c) χ term, d) Shear term, and e) absolute vorticity term. The bold curves show the multi-model means and the shading spans from the 10th to the 90th quantile among the models.

The shear term makes a small though increasingly negative contribution (increasing shear), but there is much scatter across models. Curiously, the vorticity term makes increasingly positive contributions, perhaps reflecting the poleward expansion of the region in which potential intensity exceeds *35 ms^{-1}*.

The multi-model mean evolutions are compared with each other in Figure 15.27. The general increase in GPI is dominated by increasing potential intensity, while the increase in the vorticity term and the negative contribution from increasing saturation deficit nearly cancel each other, with very little contribution from shear. The increase in vorticity is more likely owing to the eastward expansion of the region of potential intensity in excess of *35 ms^{-1}* near the equatorial Pacific cold tongue than to local increases in cyclonic relative vorticity.

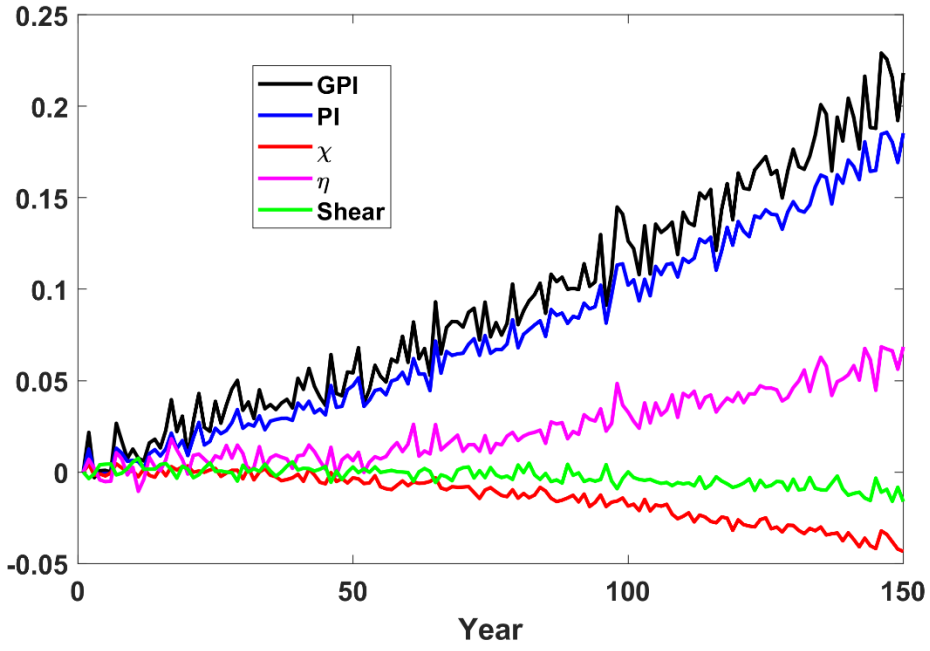


Figure 15.27: Comparison of the multi-model mean evolutions of the factors of (15.8). Note that the labels in the legend refer to the factors of (15.8) and not the quantities themselves.

The multi-model increase in GPI over 150 years is more than 60%, while the increase in the potential intensity factor is almost 50%, representing about a 25% increase in $V_{pot} - 35\ m\,s^{-1}$.

Recall that the single-column RCE model discussed earlier in this chapter predicts essentially no change in GPI, while the potential intensity increase that would result from a 1% increase in CO_2 over 150 years would be far less than that calculated from the climate models. Part of the difference is owing to our use of just the thermal part of GPI in the single-column approach (as wind shear and vorticity cannot be predicted with a single column), but much of it comes from the spatial pattern of warming in the climate models.

While these results are illuminating, they depend on the quality of the GPI given by (14.4) in predicting genesis. A different formulation that gives a stronger dependence on χ, without degrading the performance of the GPI, might suggest a qualitatively different response to increasing CO_2. An example is the genesis index developed by Tippett (2011), who used predictors similar to that of the index used here, but used Poisson regression applied to reanalysis data to find the optimal coefficients for the logarithms of the factors. When applied to a global warming simulation using a climate model with 50-km grid spacing (Camargo et al. 2014), the index yielded an increase in the global frequency of tropical cyclones. But when the index was re-trained on the output of the climate model, it predicted either an increase or a decrease in global activity depending on whether column relative humidity or saturation deficit was used as a predictor.

15.7.2: Direct numerical simulations

Tropical cyclone-like vortices were detected in some of the earliest simulations using global models (Manabe et al. 1970; Bengtsson et al. 1982; Broccoli and Manabe 1990). These models had low horizontal and vertical resolution. As with most global models, horizontal variations were represented by series of orthogonal functions, in this case representing wavelengths greater than about 1,000 km. Given the dimensions of real tropical cyclones, it is surprising that models with such low resolution can produce TC-like disturbances. A comprehensive review of TCs in climate models is provided by Camargo and Wing (2016).

Since the pioneering modeling studies, many research papers have been published on climate model TCs, as summarized by Knutson et al. (2020). Many of the models are spectral, but several use finite differences in the horizontal; in such cases it is conventional to describe their resolution in terms of equivalent grid point spacing. It should be remembered that it takes at least 4 grid points to resolve a single wave. All spectral models reported in Knutson et al. (2020) used triangular truncation, and it is conventional to describe their resolution in terms of the truncation of the zonal wavenumber. For example, "T42" denotes truncation at zonal wavenumber 42. If we denote this maximum wavenumber by k_{max}, then the minimum half-wavelength that could be partially resolved is given by

$$L_{min} = \frac{\pi a \cos(\theta)}{k_{max}}, \tag{15.9}$$

where a is the radius of the earth and θ is a representative latitude, which we here take to be 30°. We assume that a half-wavelength can be partially resolved by two grid points. Figure 15.28 shows a histogram of the minimum resolvable half-wavelength of global, and in some cases regional, models used in the Knutson et al. (2020) review.

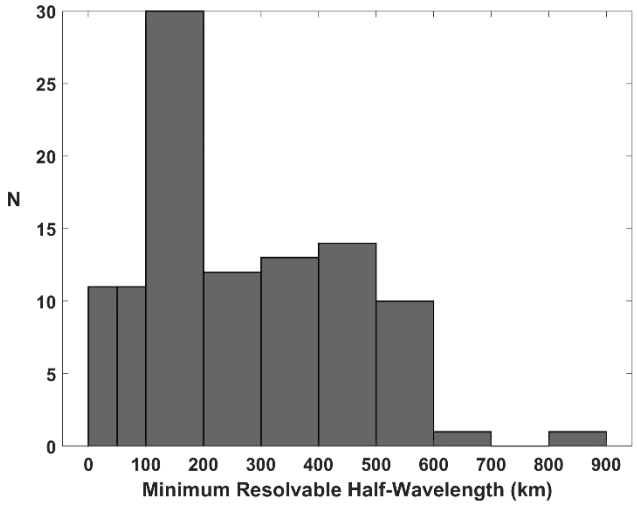


Figure 15.28: Histogram of minimum resolvable half-wavelength of the models used in the Knutson et al. (2020) review. Note that the first two bins are 50 km in width, while the rest are 100 km wide.

Most models cannot resolve features of dimensions less than 100 km. What does it take to properly resolve a tropical cyclone?

To help answer this question, we preform simulations with the convection-permitting, non-hydrostatic, axisymmetric tropical cyclone model of Rotunno and Emanuel (1987) (also see Chapter 10). Three simulations are performed that differ only in their radial grid spacing. The result is shown in Figure 15.29, with double the grid spacing (the minimum resolvable half-wavelength) displayed for ease of comparison with Figure 15.28.

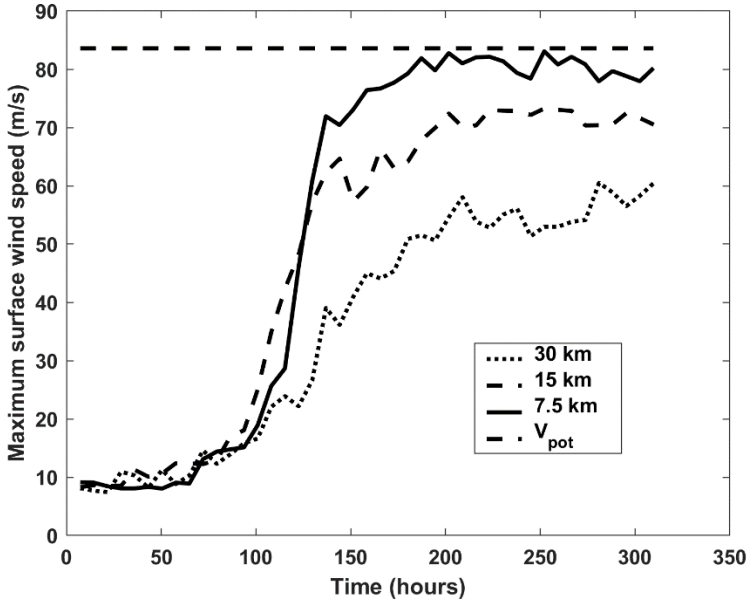


Figure 15.29: Evolution over time of the maximum surface wind speed in simulations using the axisymmetric model of Rotunno and Emanuel (1987). The legend is in terms of the minimum resolvable half-wavelength; the actual grid point spacing is half this value. The dashed horizontal line at top shows the theoretical potential intensity associated with the initial sounding, which is an RCE sounding from a single-column model.

These results suggest that the very highest resolution global models still greatly under-resolve tropical cyclones. A current drawback of high-resolution global models is that running them long enough to generate statistically robust climatologies of TCs may be too expensive. Recent studies confirm the sensitivity of global model TC simulation to resolution, up to the highest resolutions that have been used in reasonably long simulations (Roberts et al. 2020a).

Besides inadequate resolution, there are several other challenges in analyzing TCs in global climate models. While moist convection is clearly crucial in real world TCs, it cannot be resolved in any real sense by any current climate model. For this reason, most climate models include parameterizations of moist convection, which are among the most contentious features of models. Whatever else their shortcomings, they are not designed to represent slantwise convection, an example of which is the eyewall of a mature TC.

Whether or not a model has parameterized convection, if it has high enough resolution it may explicitly produce convection, which is very poorly resolved. This happens when grid point spacings or their spectral equivalent fall below roughly 50 km. Unless the convective

parameterization is absent or turned off, both parameterized and explicit convection may occur in the same region at the same time, complicating the interpretation of the physics. It is, in practice, difficult to determine whether changes in TC climatology brought about by changing resolution are owing changes in the relative importance of parameterized and explicit convection or better resolution of other features, such as the eyewall.

Although visual identification of real TCs in satellite imagery is relatively easy, objective identification of such storms in model output is far more challenging and quite a few different algorithms have been developed (see Bourdin et al. 2022 for a comparison of four of these algorithms). Virtually all such algorithms begin by detecting surface or near-surface signatures, such as a pressure minimum, or a maximum of vorticity or the Okubo-Weiss parameter (Tory et al. 2013). The latter focuses on the part of the vorticity that is free of deformation and is better at detecting TCs than the full vorticity itself. Other criteria are applied to, e.g., distinguish tropical from extratropical cyclones; these may include upper-level warm anomalies co-located with the surface cyclone, and humidity perturbations. Comparison among the different schemes reveals that TC counts and their dependence on climate both depend on the particular algorithm applied (Horn et al. 2014; Bourdin et al. 2022), especially in coarse resolution models (Camargo and Wing 2016).

The warming-induced changes in the frequency of all TCs and of Category 4-5 TCs in direct simulations by global and, in some cases, embedded regional models, as reported in Knutson et al. (2020), are shown in Figure 15.30. These are compiled without regard for the particular warming pathway used in each case. The vast majority of global models predict a decline in total numbers of TCs, in contradiction to trends of the genesis index explored in the previous subsection. The few models that predict an increase in frequency were from the CMIP5 generation of climate models, though there were other studies based on that generation that predicted declines using a different TC detection algorithm. A higher fraction of models predicted an increase in Category 4 and 5 TCs, consistent with the predicted increase in potential intensity across the major genesis regions.

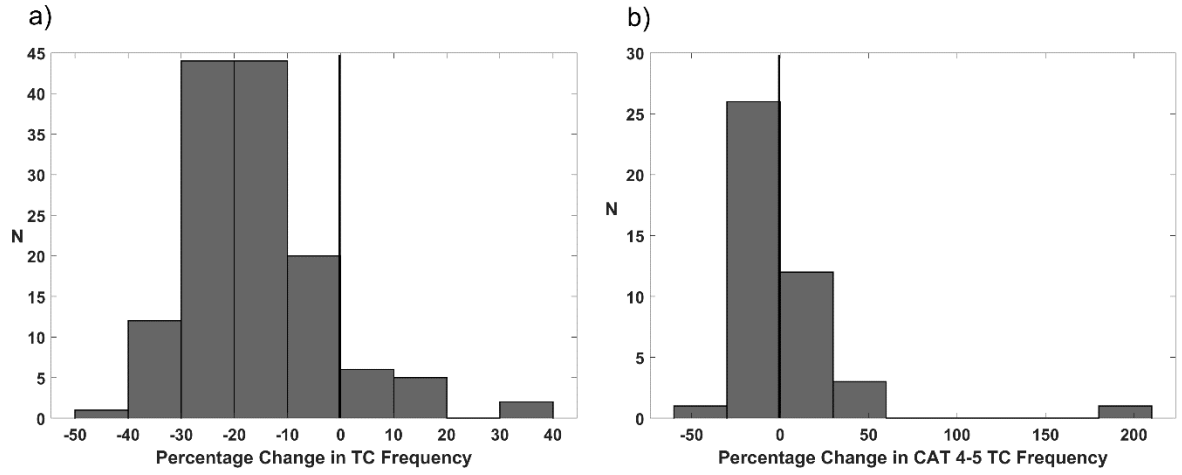


Figure 15.30: Percentage change in all TC frequency (a) and CAT 4-5 TC frequency (b) from present to warmer climate, in explicitly simulated TCs in global and regional models, from the study by Knutson et al. (2020). The warmer climates include a mixture of different emissions pathways and time periods.

Roberts et al. (2020b) analyzed TCs in the more recent CMIP6 generation of climate models, which include more high-resolution models. The best resolution models resolve a 50-km half-wavelength in our nomenclature. They broke the analyzed TC frequency down by basin and showed little change in the northern hemisphere across models, but a decided decline in southern hemisphere frequency.

The strong consensus of these modeling studies is that overall TC frequency should decline globally, with strong variations across models and individual ocean basins. But none of the models properly resolve TCs and so this conclusion should be regarded as provisional until much higher resolution models are applied to the problem. Some of the same models predict increases in the proportion of TCs that reach Category 4 and 5 status and – with important consequences – increasing rainfall.

15.7.3: Downscaled numerical simulations

The term “downscaling” has been used to describe a large variety of techniques for overcoming the limitations of low resolution described in the previous subsection. These range from embedding higher resolution models within global models down to the use of purely statistical relationships between global model variables and the variables representing small-scale phenomena.

The most straightforward technique is to embed within a global model a regional model of higher resolution. It is then possible to embed within the regional model a more localized model of even higher resolution, and so on. Embedded model downscaling can be two-way interactive when all scales of models are run simultaneously and feed information back and forth among themselves, or one-way noninteractive, in which the large-scale model does not receive information from the smaller domains. The higher resolution model can be run after the larger scale model if the appropriate initial and boundary conditions have been recorded. (We will henceforth refer to these just as two-way and one-way techniques.)

In the field of tropical cyclones, the first interactive two-way models were developed for numerical prediction of hurricanes; for example, Kurihara and Bender (1980) built a triple-mesh, two-way hurricane forecast model that was subsequently used to make the first broadly successful real-time numerical forecasts of tropical cyclones in the early 1990s (Kurihara et al. 1998; Bender et al. 2007). Beginning with the work of Camargo et al. (2007), regional one-way downscaling of global climate models has been used to help understand and predict the response of tropical cyclones to interannual climate variability. Knutson et al. (1998) were the first to use a (one-way) regional model to simulate the response of tropical cyclones, in the western North Pacific in this case, to increasing levels of CO_2.

Another approach is statistical downscaling, pioneered by Vickery et al. (2000), which models TC tracks using statistics culled from historical TC datasets such as IBTrACS (Knapp et al. 2010) and whose intensity is statistically modeled as function of environmental parameters like sea surface temperature. A recent incarnation of this technique by Bloemendaal et al. (2020) uses potential intensity, from historical or climate model data sets, to bound storm intensity. To apply this to climate change scenarios, Bloemendaal et al.(2022) used data from four high-resolution (25-50 km grid point spacing) climate models to modify their historically based TC track sets. They did this by applying the differences between historical and future simulations of

the four climate models to their previously created TC track sets based on IBTrACS data. This can be regarded as a way of bootstrapping the climate model's explicit track changes to far larger samples while avoiding the biases of the models' historical period TCs relative to IBTrACS. Yet changes in TC characteristics, including track, frequency, and intensity are still tied to those predicted by a limited set of global climate models while providing a statistically more robust and bias-corrected estimate of those changes.

Two other methods (Lee et al. 2018; Jing and Lin 2020) use statistical techniques to model genesis, tracks, and intensity, but use genesis and intensity models that are strongly constrained by known environmental parameters like potential intensity, wind shear, mid-level humidity, and genesis potential indices. These have become known as statistical-deterministic methods. The Lee et al. (2018) technique uses a beta-and-advection model for tracks, as in Emanuel (2006, 2008), while the Jing and Lin (2020) technique uses both local steering flow and historical track statistics to constrain TC tracks. The author's method (Emanuel et al. 2006, 2008) is the only one that uses an actual deterministic intensity model, but as that model is axisymmetric, shear effects have to be handled parametrically. Figure 15.31 compares the response of two of the three statistical-deterministic models to the response of high resolution (grid spacing of about 25 km) global model, all driven by an RCP 4.5 emissions scenario. The results are compared for the North Atlantic only.

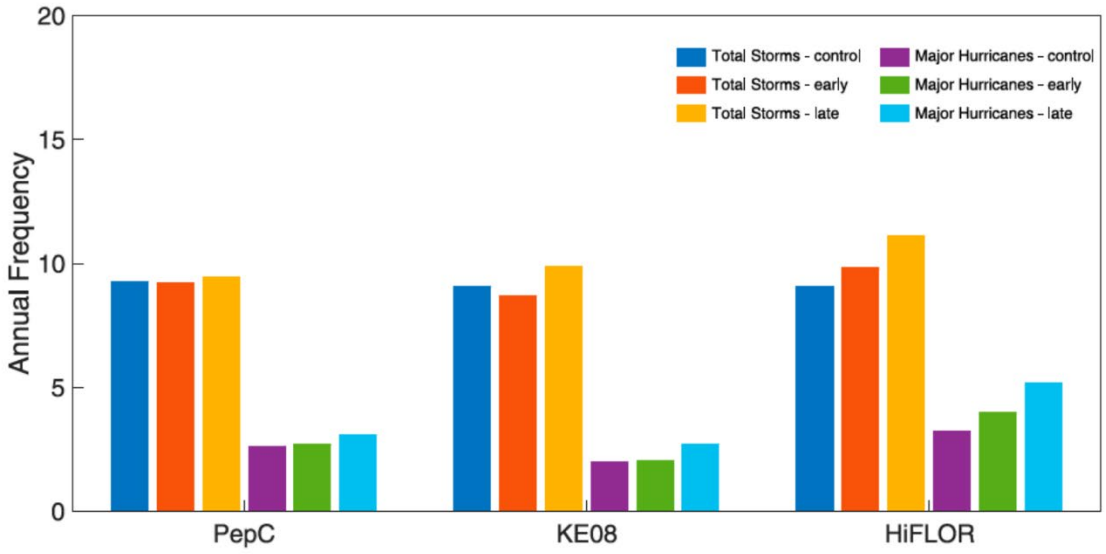


Figure 15.31: Comparison of total number of TCs (left-most 3 bars) and number of major hurricanes (right-most three bars) produced by the Jing and Lee (2020) downscaling (left), the Emanuel (2008) downscaling (center), and explicit TCs produced by the Geophysical Fluid Dynamics Laboratory (GFDL) High-Resolution Forecast-Oriented Low Ocean Resolution (HiFLOR) model (right). The downscalings have been applied to the same (HiFLOR) model. The dark blue and magenta bars are for the historical period 1896-2015, the orange and green bars are for 2016-2035 under RCP 4.5, and the yellow and light blue bars are for 2081-2100 under RCP 4.5. All results are for the North Atlantic basin.

In this case, North Atlantic TC frequency is not projected to change much by either downscaling method, but increases in the HiFLOR model itself. (HiFLOR also projects increasing frequency globally. Curiously, a lower resolution version of the same model projects globally declining frequency (Vecchi et al. 2019)). All three show an increase in major hurricanes, though the increase is larger in the explicit HiFLOR model.

Global downscaling experiments generally reveal an increase in the proportion of TCs that reach high intensity, across the various downscaling techniques, different GPIs used in some of the techniques, and across most GCMs that the techniques have been applied to. However, the change in downscaled TC frequency with increasing CO_2 depends on the technique itself, the form of the GPI used in some of the techniques, and on the climate model the downscaling is applied to. This dependence is large enough that not even the sign of the frequency change is robust.

TCs downscaled using statistical and statistical-dynamical methods have been used to help assess long-term TC risk. To do so, it is common to generate tens of thousands or hundreds of thousands of events. To be useful in assessing damage potential, the whole time-evolving wind field must be used; at the same time, storing wind over a very fine-scale grid with high time resolution for this number events may be expensive and/or infeasible. Models such as those of Bloemendaal et al.(2022), Lee et al. (2018) and Jing and Lin (2020) only predict storm position and intensity, and storing those variables at intervals of a few hours is entirely practical, even for 1 million events. But parametric models must then be used to determine the whole wind field given only the maximum wind speed. The model of Emanuel et al. (2008) deterministically predicts the axisymmetric gradient wind field as part of the solution of the deterministic intensity model, but only the predicted radii of maximum winds (and radius of any second wind maxima that may be present) are stored. Once an axisymmetric wind field is constructed, asymmetries arising from the presence of background flow and shear interactions can be added (Chang et al. 2020), and effects of topography, boundary layer stratification, and varying surface roughness can be accounted for by driving a boundary layer model with the time-evolving pressure field associated with the model's gradient winds.

While much of the focus of downscaling has been on quantities pertinent to wind damage, mortality and damage are more often caused by water in the form of storm surges and freshwater flooding from heavy rains. Downscaled TCs have been used extensively to drive hydrodynamic surge models (e.g. Lin et al. 2010, 2012, 2014; Aerts et al. 2014) as a means of estimating flooding risk from surges.

Estimating rainfall from TCs downscaled using statistical or statistical-dynamical techniques is somewhat more challenging, because the spatial and temporal distribution of rainfall is very much more complex than that of wind, especially over land. To deal with this, the author developed a rainfall algorithm (described in detail in Lu et al. 2018) that can be applied as a post-production step to TCs downscaled using statistical or statistical-dynamical techniques, as long as certain key environmental variables are stored together with the wind field parameters. In the first step, the vertical component of the flow at the surface is calculated as the dot product of the surface winds with the local topographic slope. In the second step, frictional convergence within the boundary layer is calculated from the local surface roughness and its local gradients; this is then added to the surface vertical velocity to obtain the vertical velocity at the top of the boundary layer. In the third step, the difference between the vertical velocity at the top of the boundary layer and that in the mid-troposphere is estimated from the equation for the vertical component of the vorticity:

$$\left(f+\zeta\right)\frac{\partial w}{\partial z}=\frac{\partial \zeta}{\partial t}+\mathbf{V}_H\bullet\nabla\left(f+\zeta\right), \tag{15.10}$$

where f is the Coriolis parameter, ζ is the vertical component of relative vorticity, w is vertical velocity, and $\mathbf{V}_H$ is the vector horizontal wind. In (15.10) we have neglected the tilting terms. Because we have the relative vorticity and its time derivative from the time-evolving wind field, we can calculate the absolute vorticity and the first term on the right side of (15.10). In practice, we do this in a coordinate system traveling with the TC. If there were no environmental wind, the last term on the right side of (15.10) would vanish for an axisymmetric storm. However, if there is wind shear, there will be a contribution from this last term, which we estimate from the vorticity field together with the unperturbed environmental wind shear.

Once these steps are taken, we can estimate the vertical motion in the middle troposphere by adding a representative mid-tropospheric altitude multiplied by $\partial w / \partial z$ from (15.10) and adding the result to w calculated at the top of the boundary layer. Finally, if we know the temperature of the middle troposphere, we can calculate the saturation specific humidity there, and assuming that the air ascending in a TC is nearly saturated, we have an estimate of the condensation rate. Multiplying this by an assumed precipitation efficiency yields an estimate of surface precipitation. This algorithm was tested extensively against radar and rain gauge data by Feldmann et al. (2019).

The power of downscaling is illustrated by Figure 15.32, constructed from downscaling 3,000 tropical cyclones, each passing within 150 km of Cairns, Australia, from each of 9 climate models and from two CMIP6 protocols: an historical period (1985-2014) and a future period (2071-2100) under socioeconomic pathway SSP3-7.0, a moderate global warming scenario. In addition, and for comparison, 4,500 TC events were downscaled from the ERA-5 reanalysis. The 58,500 events, in total, were downscaled using a statistical-dynamical technique (Emanuel et al., 2018).

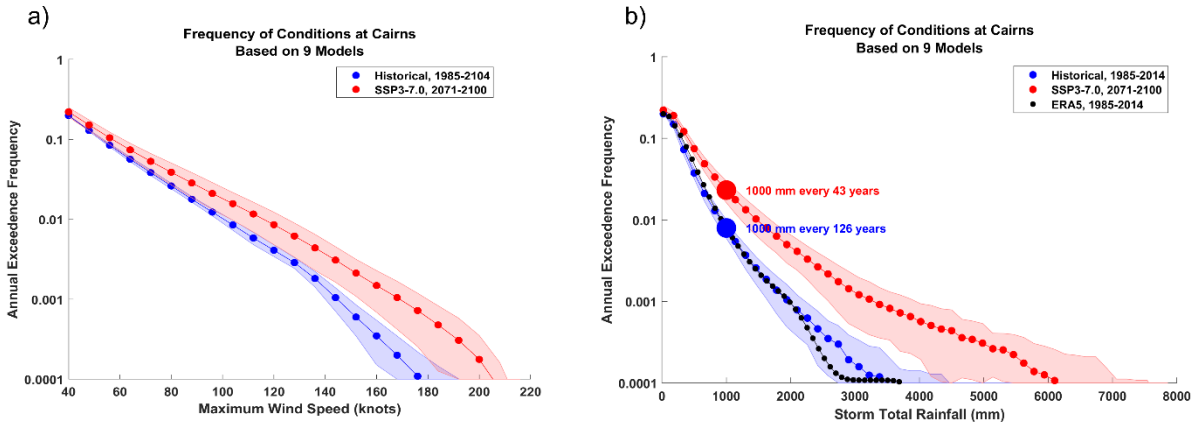


Figure 15.32: Annual exceedance frequencies of a) 10-m 1-minute wind speed, and b) storm total rainfall at Cairns, Australia, estimated by downscaling nine global climate models for two periods of time. The blue is for the historical period 1985-2014, while the red is for the future period 2071-2100 under CMIP6 protocol SSP3-7.0. The curves with the dots shown the multi-model means, while the shading shows one standard deviation up and down among the nine models. The black dots in b) are the results of downscaling the ERA-5 reanalysis over the period 1980-2022. The large dots in b) show the exceedance frequencies of events that produce 1,000 mm of rain at Cairns.

The frequencies of the downscaled events in all nine models and the ERA-5 reanalysis were calibrated to the same frequency of all events for the historical period.

The exceedance frequencies of 10-m 1-minute wind speeds at Cairns (Figure 15.32a) show that the overall frequency of winds speeds in excess of 40 kts increases slightly in the warmer climate, but the increase becomes substantially larger as the threshold wind speed increases, consistent with increasing potential intensity making storms stronger. There is considerable spread in wind frequencies among the nine models, but the two climates are reasonably well separated.

Changes in TC storm total rainfall are somewhat more dramatic (Figure 15.32b). Cyclone Jasper, so far the wettest tropical cyclone to affect Australia, caused extensive flooding in Queensland, dumping a meter of rain on the port city of Cairns and more than two meters in the hilly terrain to the west of that city. According to Figure 15.23b, a meter of rain in Cairns would have an annual probability of about one in 125 during the period 1985-2014, but the probability is expected to increase to about one in 40 by the end of this century, a three-fold increase. Linearly interpolating in time suggests that the probability of that amount of rain in Cairns increased by about 50% between 2000 and 2023, when Jasper occurred.

Consequently, while most discussions surround global changes in TC frequency, track, and intensity, the real issue is how the probability of truly destructive and/or life-threatening events changes. This is necessarily a local problem about which global or even basin-wide statistics may have little to tell us.

15.8: Possible feedbacks of tropical cyclones on climate

All the preceding discussion takes global and regional climate change as a given and attempts to estimate how TC activity responds to it. In reality, we must consider the possibility that TCs have significant feedbacks on climate itself. In principle, such feedbacks are intrinsic in coupled global climate models, but as discussed in subsection 15.7.2, these feedbacks may be poorly represented owing to inadequate resolution.

Before delving into this important if lightly explored subject, we observe that some popular reviews and undergraduate textbooks continue to claim that TCs are important agents in directly transporting heat and water from the tropics to higher latitudes. This idea is easily put to rest by scale analysis. Suppose that a single tropical cyclone whose inner core with elevated moist static energy h' over a diameter D and through a depth H is crossing a particular latitude circle with a poleward translation speed of U_t. Then the net instantaneous poleward thermodynamic energy transport by this single storm is $\rho HDU_t h'$, where ρ is a representative air density. Suppose further that 60 such events happen in the northern hemisphere every year. Then the average poleward enthalpy flux is $F = \alpha \rho HDU_t h'$, where α is 60 days per year or about 0.15. Using generous estimates of $\rho = 1 Kg\, m^{-3}$, $H = 10\, km$, $D = 100\, km$, $U_t = 10\, ms^{-1}$, and in calculating h' we take the core temperature perturbation to be $5\, K$ and moisture perturbation to be $0.01\, Kg\, Kg^{-1}$, we get $F \approx 5 \times 10^{13}\, W$. This can be compared to the observed peak poleward enthalpy flux by the atmosphere of about $4 \times 10^{15}\, W$. Thus, even with our generous scaling, the direct poleward energy flux by TCs is about 1% of the total flux. On the other hand, TCs moving into higher latitudes can excite baroclinic wave development (sections 12.3 and 12.4), which can lead to enhanced poleward heat transport. And, as we shall see in section 15.8.2, TCs,

though their effects on the upper ocean, might indirectly produce much larger horizontal enthalpy fluxes in the ocean and atmosphere.

Direct TC enthalpy fluxes aside, two types of potentially important feedbacks have been considered to date. The first is a feedback on radiation through modulation of clouds and water vapor, while the second involves a feedback on ocean heat transport.

15.8.1: Feedbacks on clouds and water vapor

Tropical cyclones, and indeed all forms of aggregated convection, tend to dry out the tropical troposphere, as described in Chapter 3 (section 3.4). There are two reasons for this: First, the characteristic length scale separating ascending moist regions from descending dry regions is large enough that lateral mixing of water vapor from the clouds to the clear region is too slow to allow much moistening of the descending air, at least judging from observations and the results of cloud-permitting simulations. Second, the precipitation efficiency is somewhat higher in aggregated convection than in normal RCE, because raindrops in the aggregated case usually fall through nearly saturated environments and thus do not evaporate much. (This lack of evaporation results in less concentration of heavier isotopes of hydrogen and oxygen in raindrops, a signature that is strongly evident in measurements of the isotopic composition of rain, as discussed in section 15.3.4.) At the same time, high clouds tend to cover a smaller fractional area in aggregated convection, providing less of a greenhouse effect (though they also increase the amount of sunlight reaching the surface). The drying of the atmosphere coupled with the reduction in high clouds increases outgoing longwave radiation, cooling the system. If the descent is strong enough, stratocumulus clouds can develop at the top of the boundary layer, reflecting sunlight and further cooling the system. Using an index of the degree of organization of deep tropical convection developed by Tompkins and Semie (2017), Bony et al. (2020) showed that net incoming top-of-the-atmosphere (TOA) radiation is strongly anti-correlated with both the degree of convective aggregation and tropospheric humidity. If it turns out that the degree of aggregation is correlated with temperature, this would constitute a negative feedback on changes in tropical temperature (Khairoutdinov and Emanuel 2010; Mauritsen and Stevens 2015). It is less certain what the relative roles of TCs and non-rotating aggregated convection are in this set of radiative feedbacks.

15.8.2: Feedbacks on ocean heat transport

On the face of it, it may be surprising that TCs, infrequent in time and affecting tiny fractions of the world oceans while they exist, might have a noticeable influence on ocean heat transport. Even more surprising might be the fact the while TCs are powered by extraction of heat from the ocean, when all is said and done, they end of inducing a heating of the upper ocean that far surpasses their heat extraction. These two concepts turn out to be strongly related.

First, let's explore why TCs end up heating the ocean. Back in Chapter 13 we discussed how TCs deepen the ocean mixed layer, cooling the surface and warming the region just below the unperturbed mixed layer (Figure 13.13 and Figure 15.33a). It is important to note that, aside from a very small net heating from dissipation of turbulence kinetic energy, the mixing does not change the column-integrated enthalpy of the ocean. But what happens after that?

The local cold anomaly at the surface will be associated with anomalously small turbulent and longwave radiative fluxes (Zhang et al. 2021), so that over a period of time, the ocean's mixed layer will recover towards its pre-cyclone state. Observations suggest that this recovery takes about one month and that considerable lateral mixing, owing to baroclinic instability, takes place during this recovery (Mei and Pasquero 2012). Very importantly, this anomalous surface heating does represent a net column heat input to the ocean. At the end of the recovery, one is left with a warm anomaly below the depth of the unperturbed mixed layer (Figure 15.33b).

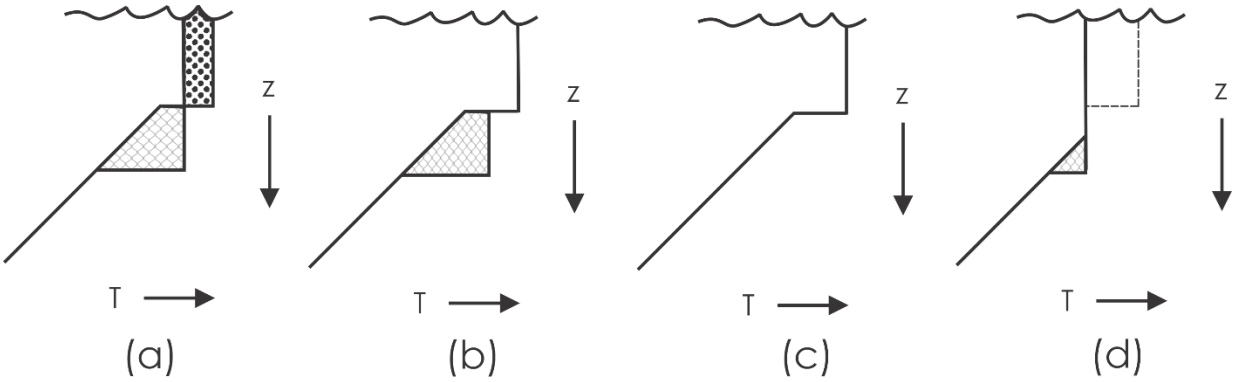


Figure 15.33: Three stages and one alternative stage in the thermodynamic perturbation of the upper ocean by a tropical cyclone. All panels show the unperturbed temperature of the mixed layer, below which there is a downward jump in temperature, giving way to a thermocline in which the temperature decreases gradually with depth. The rapid mixing induced by the tropical cyclone cools the original mixed layer (stippling in (a)) and warms the upper thermocline (netting in (a)), leaving the column-integrated enthalpy unchanged. After a month or so, the mixed layer temperature recovers (b), leaving a warm anomaly in the upper thermocline (netting in (b)). This stage involves a net, column-integrated heating. Finally, in (c), projection of the local warm anomaly onto various wave modes disperses the warm anomaly to the far field. Alternatively (and not mutually exclusively), seasonal deepening of the mixed layer in autumn returns the stored heat to the atmosphere (d).

In a climatological sense, some process must ultimately compensate for this warming of the upper thermocline. Because the warm anomaly is localized to the vicinity of the TC, it will project onto various ocean wave modes. Scott and Marotzke (2002) performed experiments with a simple ocean model in a square box representing the North Atlantic, and showed that steady vertical mixing across the tropical thermocline results in a westward surface current that turns poleward at the western boundary, with the opposite sign response at depth. These currents transport heat westward and then poleward. The mixing is particularly effective when it is located near the western boundary in the tropical thermocline. While the mixing in these idealized experiments was steady, Boos et al. (2004) showed that the transience of the mixing did not appreciably reduce the induced poleward heat transport by the oceans, especially if the mixing events were near the western boundary. Transient mixing did excite transient wave modes, particularly in the tropics.

The aforementioned scenario is most likely to run its course if the TC occurs early in the season. For later summer and autumn storms that are not at very low latitudes, autumn deepening of the mixed layer may mix out most or even all the stored warm anomaly, as illustrated in Figure 15.33d. Estimates from observations and models suggest that at least one third of the upper thermocline warm anomaly remains in the tropics for more than a year (Pasquero and Emanuel 2008), while Jansen (2010) estimated that only about a quarter of the heat anomaly created by TCs, on average, remains in the thermocline after subsequent autumn cooling.

The autumnal absorption of the upper thermocline warm anomaly has the effect of slowing the autumn cooling of the tropical atmosphere-ocean system. This in turn would increase the magnitude of the pole-to-equator temperature gradient and presumably increase the atmospheric poleward heat flux by baroclinic eddies in autumn. (Yet Hart (2011) finds a negative relationship between autumn TC activity and atmospheric poleward heat flux in the following winter.) On the other hand, the part of the warm anomaly that projects onto ocean wave modes can conceivably enhance poleward heat flux in the ocean, following the results of Soctt and Marotzke (2002) and Boos et al. (2004). Either way, it would appear that TCs may indirectly contribute, through their mixing of the upper ocean, to poleward heat transport by the atmosphere-ocean system.

An early estimate of the magnitude of lateral oceanic heat flux induced by TC mixing of the upper ocean was made by the author (Emanuel 2001). Using observed TC tracks and model-based estimates of TC mixing, he estimated that TCs induce $1.4 \pm 0.7\ PW$ of poleward heat transport, which is comparable to the peak oceanic poleward heat flux of around $2\ PW$ deduced from observations. (A petawatt (*PW*) is 10^{15} *W*.) However, he assumed no reabsorption of the upper thermocline warm anomaly during autumnal cooling, and that all the induced heat flux is poleward rather than zonal or equatorward. Subsequent work suggests that in reality, some of the induced oceanic heat flux from tropical cyclones is actually equatorward (Jansen and Ferrari 2009) and/or eastward, at least in the Pacific region (Federov et al. 2010). At higher latitudes, the enhanced ocean heat flux owing to TC-induced mixing, calculated by Jansen and Ferrari (2009), is on the order of a few tenths of a petawatt. Analysis of satellite-observed sea surface temperature recovery from TCs (Sriver et al. 2008) and sea surface altitude anomalies (Mei et al. 2013) also suggests that the actual TC-induced poleward heat flux is more in the range of $0.3 - 0.5\ PW$. Sriver et al. (2010) used satellite-derived estimates of upper ocean TC-induced mixing (Figure 15.34a) to drive an earth-system model of intermediate complexity. Comparison of the distributions of sea surface temperature in experiments with and without this enhanced mixing shows strong warming of the eastern Pacific upwelling region and cooling of the far western Pacific, with mild warming of the subtropical to polar North Atlantic (Figure 15.34b), consistent with the work of Jansen and Ferrari (2009) and Federov et al. (2010). A somewhat similar pattern of SST response was found by Korty et al. (2008) who also used a coupled model of intermediate complexity but with parameterized, interactive TC-induced mixing rather than specified mixing. The response has somewhat more warming at higher latitudes, implying a greater poleward heat transport.

While careful observational analysis has narrowed down the range of estimates of TC-induced warming of the upper ocean to around $0.2 - 0.6\ PW$, roughly 10%-30% of the observed peak poleward heat flux by the world oceans, modeling the response to this has thus far been compromised by lack of resolution of the atmosphere and/or ocean, and lack of seasonality and high frequency transience of TC-induced mixing. Some of the modeling has used specified atmospheric inputs, which ignores the return of TC-induced enthalpy anomalies from the ocean to the atmosphere in the colder seasons.

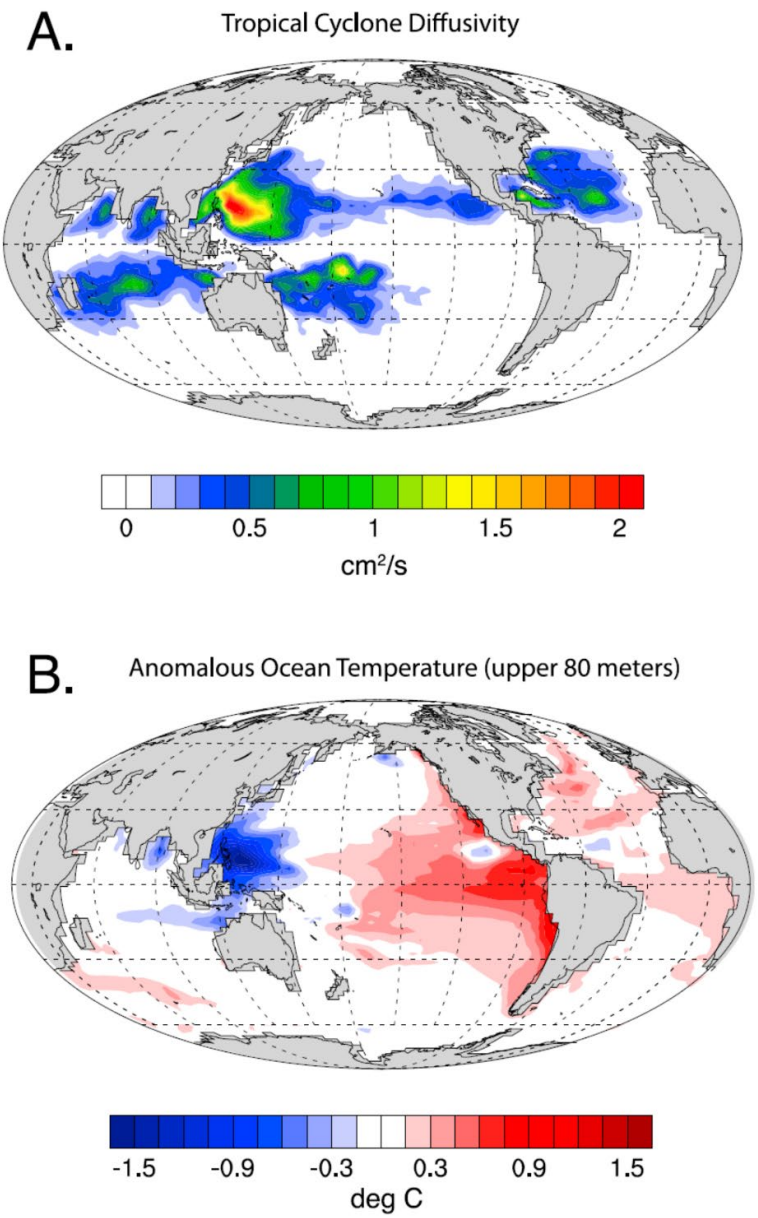


Figure 15.34: A): Tropical cyclone-induced upper ocean diffusivity from Sriver et al. (2008), and B) anomalous upper ocean temperature response in an earth-system model of intermediate complexity.

If TCs turn out to be an important driver of oceanic enthalpy fluxes, the response of TCs to climate change could constitute an important feedback on climate change (Emanuel 2001). For example, greater TC activity during the Pliocene period may have contributed to a state resembling a permanent El Niño in the eastern Pacific (Federov et al. 2010). The author (Emanuel 2002) hypothesized that interactive TC mixing of the upper ocean could lead to climate intransitivity, with two possible climate states – a state similar to today's and a far more equable climate with low pole-to-equator contrast – for the same external forcing. These and other aspects of the role of TCs and climate should remain important research problems for some time.

References

Abbott, T. H., and T. W. Cronin, 2021: Aerosol invigoration of atmospheric convection through increases in humidity. *Science*, **371**, 83–85, https://doi.org/10.1126/science.abc5181.

Aerts, J. C. J. H., W. J. W. Botzen, K. Emanuel, N. Lin, H. de Moel, and E. O. Michel-Kerjan, 2014: Evaluating flood resilience strategies for coastal megacities. *Science*, **344**, 473–475, https://doi.org/10.1126/science.1248222.

Altman, J., M. Saurer, J. Dolezal, N. Maredova, J.-S. Song, C.-H. Ho, and K. Treydte, 2021: Large volcanic eruptions reduce landfalling tropical cyclone activity: Evidence from tree rings. *Science of The Total Environment*, **775**, 145899, https://doi.org/10.1016/j.scitotenv.2021.145899.

Barron, E. J., 1989: Severe storms during Earth history. *Geological Society of America Bulletin*, **101**, 601–612.

Bender, M. A., I. Ginis, R. Tuleya, B. Thomas, and T. Marchok, 2007: The Operational GFDL Coupled Hurricane–Ocean Prediction System and a Summary of Its Performance. *Monthly Weather Review*, **135**, 3965–3989, https://doi.org/10.1175/2007MWR2032.1.

Bengtsson, L., H. Bottger, and M. Kanamitsu, 1982: Simulation of hurricane-type vortices in a general circulation model. *Tellus*, **34**, 440–457, https://doi.org/10.1111/j.2153-3490.1982.tb01833.x.

Bergeron, T., 1954: The problem of tropical hurricanes. *Quart. J. Roy. Meteor. Soc.*, **80**, 131–164.

Bister, M., and K. A. Emanuel, 2002: Low frequency variability of tropical cyclone potential intensity, 1: Interannual to interdecadel variability. *J. Geophys. Res.*, **107**, doi:10.1029/2001JD000776.

Bloemendaal, N., I. D. Haigh, H. de Moel, S. Muis, R. J. Haarsma, and J. C. J. H. Aerts, 2020: Generation of a global synthetic tropical cyclone hazard dataset using STORM. *Scientific Data*, **7**, 40, https://doi.org/10.1038/s41597-020-0381-2.

Bloemendaal, N., and Coauthors, 2022: A globally consistent local-scale assessment of future tropical cyclone risk. *Science Advances*, **8**, eabm8438, https://doi.org/10.1126/sciadv.abm8438.

Bony, S., A. Semie, R. J. Kramer, B. Soden, A. M. Tompkins, and K. A. Emanuel, 2020: Observed Modulation of the Tropical Radiation Budget by Deep Convective Organization and Lower-Tropospheric Stability. *AGU Advances*, **1**, e2019AV000155, https://doi.org/10.1029/2019av000155.

Boos, W. R., J. R. Scott, and K. A. Emanuel, 2004: Transient diapycnal mixing and the meridional overturning circulation. *Journal of Physical Oceanography*, **34**, 334–341, https://doi.org/10.1175/1520-0485(2004)034<0334:TDMATM>2.0.CO;2.

Bourdin, S., S. Fromang, W. Dulac, J. Cattiaux, and F. Chauvin, 2022: Intercomparison of four algorithms for detecting tropical cyclones using ERA5. *Geoscientific Model Development*, **15**, 6759–6786, https://doi.org/10.5194/gmd-15-6759-2022.

Broccoli, A. J., and S. Manabe, 1990: Can existing climate models be used to study anthropogenic changes in tropical cyclone climate? *Geophys. Res. Lett.*, **17**, 1917–1920.

Callaghan, J., and S. B. Power, 2011: Variability and decline in the number of severe tropical cyclones making land-fall over eastern Australia since the late nineteenth century. *Climate Dynamics*, **37**, 647–662, https://doi.org/10.1007/s00382-010-0883-2.

Camargo, S. J., and A. A. Wing, 2016: Tropical cyclones in climate models. *WIREs Clim. Change*, **7**, 211–237, https://doi.org/10.1002/wcc.373.

Camargo, S. J., H. Li, and L. Sun, 2007: Feasibility study for downscaling seasonal tropical cyclone activity using the NCEP regional spectral model. *International Journal of Climatology*, **27**, 311–325, https://doi.org/10.1002/joc.1400.

——, M. K. Tippett, A. H. Sobel, G. A. Vecchi, and M. Zhao, 2014: Testing the performance of tropical cyclone genesis indices in future climates using the hiram model. *Journal of Climate*, **27**, 9171–9196, https://doi.org/10.1175/jcli-d-13-00505.1.

Cane, M. A., S. E. Zebiak, and S. C. Dolan, 1986: Experimental forecasts of El Niño. *Nature*, **321**, 827–832, https://doi.org/10.1038/321827a0.

Chan, J. C. L., 1985: Tropical cyclone activity in the Northwest Pacific in relation to the El Niño/Southern Oscillation phenomenon. *Monthly Weather Review*, **113**, 599–606, https://doi.org/10.1175/1520-0493(1985)113<0599:TCAITN>2.0.CO;2.

Chang, D., S. Amin, and K. Emanuel, 2020: Modeling and parameter estimation of hurricane wind fields with asymmetry. *Journal of Applied Meteorology and Climatology*, **59**, 687–705, https://doi.org/10.1175/jamc-d-19-0126.1.

Chenoweth, M., and I. Howard, 2023: Hurricane impacts on land in the central and eastern Caribbean since 1494 CE from written records. *Earth and Space Science*, **10**, e2023EA002897, https://doi.org/10.1029/2023EA002897.

Chivas, A., J. Chappell, H. Polach, B. Pillans, and P. Flood, 1986: Radiocarbon evidence for the timing and rate of Island development, beach-rock formation and phosphatization at Lady Elliot Island, Queensland, Australia. *Marine Geology*, **69**, 273–287, https://doi.org/10.1016/0025-3227(86)90043-5.

Cotton, W. R., G. M. Krall, and G. G. Carrió, 2012: Potential indirect effects of aerosol on tropical cyclone intensity: convective fluxes and cold-pool activity. *Tropical Cyclone Research and Review*, **1**, 293–306, https://doi.org/10.6057/2012TCRR03.05.

Davis, R. A., S. C. Knowles, and M. J. Bland, 1989: Role of hurricanes in the Holocene stratigraphy of estuaries; examples from the Gulf Coast of Florida. *Journal of Sedimentary Research*, **59**, 1052–1061, https://doi.org/10.1306/212F90ED-2B24-11D7-8648000102C1865D.

Denniston, R. F., and Coauthors, 2023: Sensitivity of northwest Australian tropical cyclone activity to ITCZ migration since 500 CE. *Science Advances*, **9**, eadd9832, https://doi.org/10.1126/sciadv.add9832.

Dong, B., R. T. Sutton, E. Highwood, and L. Wilcox, 2014: The impacts of European and Asian anthropogenic sulfur dioxide emissions on Sahel Rainfall. *Journal of Climate*, **27**, 7000–7017, https://doi.org/10.1175/JCLI-D-13-00769.1.

Duke, W. L., 1985: Hummocky cross-stratification, tropical hurricanes, and intense winter storms. *Sedimentology*, **32**, 167–194, https://doi.org/10.1111/j.1365-3091.1985.tb00502.x.

Emanuel, K., 2002: A simple model of multiple climate regimes. *J. Geophys. Res.*, **107**, DOI 10.1029/2001JD001002.

Emanuel, K., 2018: 100 Years of Progress in Tropical Cyclone Research. *Meteorological Monographs*, **59**, 15.1-15.68, https://doi.org/10.1175/amsmonographs-d-18-0016.1.

Emanuel, K., and A. H. D.-:10. 1002/jame. 20032 Sobel, 2013: Response of tropical sea surface temperature, precipitation, and tropical cyclone-related variables to changes in global and local forcing. *J. Adv. Model. Earth Sys.*, **5**.

——, R. Sundararajan, and J. Williams, 2008: Hurricanes and global warming: Results from downscaling IPCC AR4 simulations. *Bull. Amer. Meteor. Soc.*, **89**, 347–367.

Emanuel, K. A., 1989: The Finite-Amplitude Nature of Tropical Cyclogenesis. *Journal of the Atmospheric Sciences*, **46**, 3431–3456, https://doi.org/10.1175/1520-0469(1989)046<3431:tfanot>2.0.co;2.

Emanuel, K. A., 1995: The behavior of a simple hurricane model using a convective scheme based on subcloud-layer entropy equilibrium. *J. Atmos. Sci.*, **52**, 3959–3968.

——, 2001: The contribution of tropical cyclones to the oceans' meridional heat transport. *J. Geophys. Res.*, **106**, 14,771-14,782.

——, and D. S. Nolan, 2004: Tropical cyclone activity and the global climate system. *26th AMS Conference on Hurricanes and Tropical Meteorology*,.

——, S. Ravela, E. Vivant, and C. Risi, 2006: A statistical-deterministic approach to hurricane risk assessment. *Bull. Amer. Meteor. Soc.*, **19**, 299–314.

Federov, A. V., C. M. Brierley, and K. Emanuel, 2010: Tropical cyclones and permanent El Niño in the early Pliocene epoch. *Nature*, **463**, 1066–1070.

Feldmann, M., K. Emanuel, L. Zhu, and U. Lohmann, 2019: Estimation of Atlantic Tropical Cyclone Rainfall Frequency in the United States. *Journal of Applied Meteorology and Climatology*, **58**, 1853–1866, https://doi.org/10.1175/jamc-d-19-0011.1.

Frappier, A. B., D. Sahagian, S. J. Carpenter, L. A. González, and B. R. Frappier, 2007: Stalagmite stable isotope record of recent tropical cyclone events. *Geology*, **35**, 111–114.

Gao, C., A. Robock, and C. Ammann, 2008: Volcanic forcing of climate over the past 1500 years: An improved ice core-based index for climate models. *Journal of Geophysical Research: Atmospheres*, **113**, https://doi.org/10.1029/2008JD010239.

Gray, W. M., 1984: Atlantic seasonal hurricane frequency. Part I: El Niño and 30 mb quasi-biennial oscillation influences. *Mon. Wea. Rev.*, **112**, 1649–1668.

Grossman, M., and M. Zaiki, 2009: Reconstructing typhoons in Japan in the 1880s from documentary records. *Weather*, **64**, 315–322.

Haig, J., J. Nott, and G. J. Reichart, 2014: Australian tropical cyclone activity lower than at any time over the past 550-1,500 years. *Nature*, **505**, 667-+, https://doi.org/10.1038/nature12882.

Hart, R. E. D.-:10. 1029/2010GL045612, 2011: An inverse relationship between aggregate northern hemisphere tropical cyclone activity and subsequent winter climate. *Geophysical Research Letters*, **38**.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Horn, M., and Coauthors, 2014: Tracking scheme dependence of simulated tropical cyclone response to idealized climate simulations. *Journal of Climate*, **27**, 9197–9213, https://doi.org/10.1175/JCLI-D-14-00200.1.

Huang, Y., Y. Xia, and X. Tan, 2017: On the pattern of CO2 radiative forcing and poleward energy transport. *Journal of Geophysical Research: Atmospheres*, **122**, 10,578-10,593, https://doi.org/10.1002/2017JD027221.

IPCC-6, 2023: *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press,.

Jansen, M., and R. D.-:10. 1029/2008GL036796 Ferrari, 2009: Impact of the latitudinal distribution of tropical cyclones on ocean heat transport. *Geophysical Research Letters*, **36**.

Jansen, M. F., R. Ferrari, and T. A. D.-:10. 1029/2009GL041808 Mooring, 2010: Seasonal versus permanent thermocline warming by tropical cyclones. *Geophysical Research Letters*, **37**.

Jing, R., and N. Lin, 2020: An environment-dependent probabilistic tropical cyclone model. *Journal of Advances in Modeling Earth Systems*, **12**, e2019MS001975, https://doi.org/10.1029/2019MS001975.

Kagawa, A., M. Sano, T. Nakatsuka, T. Ikeda, and S. Kubo, 2015: An optimized method for stable isotope analysis of tree rings by extracting cellulose directly from cross-sectional laths. *Chemical Geology*, **393–394**, 16–25, https://doi.org/10.1016/j.chemgeo.2014.11.019.

Khairoutdinov, M. F., and K. Emanuel, 2010: Aggregated convection and the regulation of tropical climate. 29th Conference on Hurricanes and Tropical Meteorology, Boston, MA, Amer. Meteor. Soc.

Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, 2010: The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone best track data. *Bull. Amer. Meteor. Soc.*, **91**, 363–376.

Knutson, T., and Coauthors, 2020: Tropical Cyclones and Climate Change Assessment: Part II: Projected Response to Anthropogenic Warming. *Bulletin of the American Meteorological Society*, **101**, E303–E322, https://doi.org/10.1175/bams-d-18-0194.1.

Knutson, T. R., R. E. Tuleya, and Y. Kurihara, 1998: Simulated increase of hurricane intensities in a CO2-warmed climate. *Science*, **279**, 1018–1020.

Kocifaj, M., M. Gangl, F. Kundracík, H. Horvath, and G. Videen, 2006: Simulation of the optical properties of single composite aerosols. *Journal of Aerosol Science*, **37**, 1683–1695, https://doi.org/10.1016/j.jaerosci.2006.08.002.

Korty, R., K. A. Emanuel, and J. R. Scott, 2008: Tropical cyclone-induced upper ocean mixing and climate: application to equable climates. *J. Climate*, **21**, 638–654.

Kossin, J. P., 2017: Hurricane intensification along United States coast suppressed during active hurricane periods. *Nature*, **541**, 390, https://doi.org/10.1038/nature20783.

Kurihara, Y., and M. A. Bender, 1980: Use of a movable nested-mesh model for tracking a small vortex. *Monthly Weather Review*, **108**, 1792–1809, https://doi.org/10.1175/1520-0493(1980)108<1792:UOAMNM>2.0.CO;2.

——, R. E. Tuleya, and M. A. Bender, 1998: The GFDL hurricane prediction system and its performance in the 1995 hurricane season. *Monthly Weather Review*, **126**, 1306–1322, https://doi.org/10.1175/1520-0493(1998)126<1306:TGHPSA>2.0.CO;2.

Lai, Q., J. Gao, W. Zhang, and X. Guan, 2020: Influences of the equatorial waves on multiple tropical cyclone genesis over the western North Pacific. *Terr. Atmos. Ocean Sci.*, **31**, 2027–2038, https://doi.org/0.3319/TAO.2020.03.20.01.

Laloyaux, P., and Coauthors, 2018: CERA-20C: A Coupled Reanalysis of the Twentieth Century. *Journal of Advances in Modeling Earth Systems*, **10**, 1172–1195, https://doi.org/10.1029/2018MS001273.

Lawrence, J. R., and S. D. Gedzelman, 1996: Low stable isotope ratios of tropical cyclone rains. *Geophys Res Lett*, **23**, 527–530.

Lee, C.-Y., M. K. Tippett, A. H. Sobel, and S. J. Camargo, 2018: An environmentally forced tropical cyclone hazard model. *Journal of Advances in Modeling Earth Systems*, **10**, 223–241, https://doi.org/10.1002/2017MS001186.

Lee, Uk, and Hong Yoon, 2012: Weather characteristics and efforts to reduce disasters during the reign of King Sejo in the Chosun Dynasty. *Atmosphere*, **22**, 199–208.

Li, Z.-H., N. Labbé, S. G. Driese, and H. D. Grissino-Mayer, 2011: Micro-scale analysis of tree-ring δ18O and δ13C on α-cellulose spline reveals high-resolution intra-annual climate variability and tropical cyclone activity. *Chemical Geology*, **284**, 138–147, https://doi.org/10.1016/j.chemgeo.2011.02.015.

Liebmann, B., H. H. Hendon, and J. D. Glick, 1994: The relationship between tropical cyclones of the western Pacific and Indian Oceans and the Madden-Julian Oscillation. *Journal of the Meteorological Society of Japan. Ser. II*, **72**, 401–412, https://doi.org/10.2151/jmsj1965.72.3_401.

Lin, N., K. A. Emanuel, J. A. Smith, and E. D.-:10. 1029/2009JD013630 Vanmarcke, 2010: Risk assessment of hurricane storm surge for New York City. *J. Geophys. Res.*, **115**.

——, K. Emanuel, M. Oppenheimer, and E. D.-:10. 1038/nclimate1389 Vanmarcke, 2012: Physically based assessment of hurricane surge threat under climate change. *Nature Clim. Change*,.

——, P. Lane, K. A. Emanuel, R. Sullivan, and J. P. Donnelly, 2014: Heightened hurricane surge risk in northwest Florida revealed from climatological-hydrodynamic modeling and paleorecord reconstruction. *J Geophys Res-Atmos*, **119**, 8606–8623, https://doi.org/10.1002/2014jd021584.

Liu, K., C. Shen, and K. Louie, 2001: A 1,000-Year History of Typhoon Landfalls in Guangdong, Southern China, Reconstructed from Chinese Historical Documentary Records. *Annals of the Association of American Geographers*, **91**, 453–464, https://doi.org/10.1111/0004-5608.00253.

Liu, K.-B., and M. L. Fearn, 1993: Lake-sediment record of late Holocene hurricane activities from coastal Alabama. *Geology*, **21**, 793–796.

Louie, K., and K. Liu, 2003: Earliest historical records of typhoons in China. *Journal of Historical Geography*, **29**, 299–316, https://doi.org/10.1006/jhge.2002.0453.

Lu, P., N. Lin, K. Emanuel, D. Chavas, and J. Smith, 2018: Assessing Hurricane Rainfall Mechanisms Using a Physics-Based Model: Hurricanes Isabel (2003) and Irene (2011). *Journal of the Atmospheric Sciences*, **75**, 2337–2358, https://doi.org/10.1175/jas-d-17-0264.1.

Malmquist, D. L., 1997: Oxygen isotopes in cave stalagmites as a proxy record of past tropical cyclone activity. Proceedings of the 22nd conference on hurricanes and tropical meteorology, Fort Collins, CO, Amer. Meteor. Soc., 393–394.

Manabe, S., J. L. Holloway, and H. M. Stone, 1970: Tropical Circulation in a Time-Integration of a Global Model of the Atmosphere. *Journal of Atmospheric Sciences*, **27**, 580–613, https://doi.org/10.1175/1520-0469(1970)027<0580:TCIATI>2.0.CO;2.

Marriner, N., D. Kaniewski, E. Garnier, M. Pourkerman, M. Giaime, M. Vacchi, and C. Morhange, 2022: Has the Anthropocene affected the frequency and intensity of tropical cyclones? Evidence from Mascarene Islands historical records (southwestern Indian Ocean). *Global and Planetary Change*, **217**, 103933, https://doi.org/10.1016/j.gloplacha.2022.103933.

Mauritsen, T., and B. Stevens, 2015: Missing iris effect as a possible cause of muted hydrological change and high climate sensitivity in models. *Nature Geosci.*, **8**, 346–351, https://doi.org/10.1038/NGEO2414.

Mei, W., and C. Pasquero, 2012: Restratification of the upper ocean after the passage of a tropical cyclone: a numerical study. *J Phys Oceanogr*, **42**, 1377–1401, https://doi.org/10.1175/Jpo-D-11-0209.1.

——, F. Primeau, J. C. McWilliams, and C. Pasquero, 2013: Sea surface height evidence for long-term warming effects of tropical cyclones on the ocean. *P Natl Acad Sci USA*, **110**, 15207–15210, https://doi.org/10.1073/pnas.1306753110.

Merlis, T. M., W. Zhou, I. M. Held, and M. Zhao, 2016: Surface temperature dependence of tropical cyclone-permitting simulations in a spherical model with uniform thermal forcing. *Geophysical Research Letters*, **43**, 2859–2865, https://doi.org/10.1002/2016GL067730.

Miller, D. L., C. I. Mora, H. D. Grissino-Mayer, C. J. Mock, M. E. Uhle, and Z. Sharp, 2006: Tree-ring isotope records of tropical cyclone activity. *P Natl Acad Sci USA*, **103**, 14294–14297, https://doi.org/10.1073/pnas.0606549103.

Mock, C. J., 2008: Tropical cyclone variations in Louisiana, U.S.A., since the late eighteenth century. *Geochemistry, Geophysics, Geosystems*, **9**, https://doi.org/10.1029/2007GC001846.

Morsilli, M., and L. Pomar, 2012: Internal waves vs. surface storm waves: a review on the origin of hummocky cross-stratification. *Terra Nova*, **24**, 273–282, https://doi.org/10.1111/j.1365-3121.2012.01070.x.

Muller, J., J. M. Collins, S. Gibson, and L. Paxton, 2017: Recent Advances in the Emerging Field of Paleotempestology. *Hurricanes and Climate Change: Volume 3*, J.M. Collins and K. Walsh, Eds., Springer International Publishing, 1–33.

Myrow, P. M., and J. B. Southard, 1996: Tempestite deposition. *Journal of Sedimentary Research*, **66**, 875–887, https://doi.org/10.1306/D426842D-2B26-11D7-8648000102C1865D.

NASA, 2015: Global Modeling and Assimilation Office (GMAO) (2015), inst3_3d_asm_Cp: MERRA-2 3D IAU State, Meteorology Instantaneous 3-hourly (p-coord, 0.625x0.5L42), version 5.12.4,. https://doi.org/10.5067/VJAFPLI1CSIV.

Nash, D. J., K. Pribyl, J. Klein, G. H. Endfield, D. R. Kniveton, and G. C. D. Adamson, 2015: Tropical cyclone activity over Madagascar during the late nineteenth century. *International Journal of Climatology*, **35**, 3249–3261, https://doi.org/10.1002/joc.4204.

Nott, J., 2004: The tsunami hypothesis—comparisons of the field evidence against the effects, on the Western Australian coast, of some of the most powerful storms on Earth. *Marine Geology*, **208**, 1–12, https://doi.org/10.1016/j.margeo.2004.04.023.

Nott, J., and M. Hayne, 2001: High frequency of “super-cyclones” along the Great Barrier Reef over the past 5,000 years. *Nature*, **413**, 508–512, https://doi.org/10.1038/35097055.

——, S. Smithers, K. Walsh, and E. Rhodes, 2009: Sand beach ridges record 6000 year history of extreme tropical cyclone activity in northeastern Australia. *Quaternary Sci Rev*, **28**, 1511–1520, https://doi.org/10.1016/j.quascirev.2009.02.014.

Oliva, F., M. Peros, and A. Viau, 2017: A review of the spatial distribution of and analytical techniques used in paleotempestological studies in the western North Atlantic Basin. *Progress in Physical Geography: Earth and Environment*, **41**, 171–190, https://doi.org/10.1177/0309133316683899.

Pasquero, C., and K. Emanuel, 2008: Tropical cyclones and transient upper-ocean warming. *J Climate*, **21**, 149–162, https://doi.org/10.1175/2007jcli1550.1.

Pazos, M., and B. Mendoza, 2013: Landfalling tropical cyclones along the eastern Pacific coast between the sixteenth and twentieth centuries. *Journal of Climate*, **26**, 4219–4230, https://doi.org/10.1175/JCLI-D-11-00411.1.

Pierrehumbert, R. T., 2010: *Principles of Planetary Climate*. Cambridge University Press, 652 pp.

Pomar, L., M. Morsilli, P. Hallock, and B. Bádenas, 2012: Internal waves, an under-explored source of turbulence events in the sedimentary record. *Earth-Science Reviews*, **111**, 56–81, https://doi.org/10.1016/j.earscirev.2011.12.005.

Price, J. F., 1983: Internal wake of a moving storm, Part I: Scales, energy budget, and observations. *J. Phys. Oceanogr.*, **13**, 949-965.

Prospero, J. M., and P. J. Lamb, 2003: African droughts and dust transport to the Caribbean: Climate change implications. *Science*, **302**, 1024–1027, https://doi.org/10.1126/science.1089915.

Roberts, M. J., and Coauthors, 2020a: Impact of model resolution on tropical cyclone simulation using the HighResMIP–PRIMAVERA multimodel ensemble. *Journal of Climate*, **33**, 2557–2583, https://doi.org/10.1175/JCLI-D-19-0639.1.

——, and Coauthors, 2020b: Projected Future Changes in Tropical Cyclones using the CMIP6 HighResMIP Multi-model Ensemble. *Geophysical Research Letters*, **n/a**, e2020GL088662, https://doi.org/10.1029/2020gl088662.

Robock, A., 2000: Volcanic eruptions and climate. *Reviews of Geophysics*, **38**, 191–219, https://doi.org/10.1029/1998RG000054.

Rotunno, R., and K. A. Emanuel, 1987: An air-sea interaction theory for tropical cyclones. Part II. *J. Atmos. Sci.*, **44**, 542–561.

Rousseau-Rizzi, R., and K. Emanuel, 2022: Natural and anthropogenic contributions to the hurricane drought of the 1970s–1980s. *Nature Communications*, **13**, 5074, https://doi.org/10.1038/s41467-022-32779-y.

Sanford, T. B., J. F. Price, and J. B. Girton, 2011: Upper-Ocean Response to Hurricane Frances (2004) Observed by Profiling EM-APEX Floats. *Journal of Physical Oceanography*, **41**, 1041–1056, https://doi.org/10.1175/2010jpo4313.1.

Schreck, C. J., J. Molinari, and A. Aiyyer, 2012: A global view of equatorial waves and tropical cyclogenesis. *Monthly Weather Review*, **140**, 774–788, https://doi.org/10.1175/MWR-D-11-00110.1.

Scott, J. R., and J. Marotzke, 2002: The location of diapycnal mixing and the meridional overturning circulation. *J. Phys. Ocean.*, **32**, 3578–3595.

Smith, S. J., J. van Aardenne, Z. Klimont, R. J. Andres, A. Volke, and S. Delgado Arias, 2011: Anthropogenic sulfur dioxide emissions: 1850–2005. *Atmospheric Chemistry and Physics*, **11**, 1101–1116, https://doi.org/10.5194/acp-11-1101-2011.

Sobel, A. H., and Coauthors, 2023: Near-term tropical cyclone risk and coupled Earth system model biases. *Proceedings of the National Academy of Sciences*, **120**, e2209631120, https://doi.org/10.1073/pnas.2209631120.

Sriver, R. L., M. Huber, and J. D.-:10. 1029/2007GC001842 Nusbaumer, 2008: Investigating tropical cyclone-climate feedbacks using the TRMM Microwave Imager and the Quick Scatterometer. *Geochemistry, Geophysics, Geosystems*, **9**.

——, M. Goes, M. E. Mann, and K. D.-:10. 1029/2010JC006106 Keller, 2010: Climate response to tropical cyclone-induced ocean mixing in an Earth system model of intermediate complexity. *Journal of Geophysical Research: Oceans*, **115**.

Swift, D. J., and D. Nummedal, 1987: Hummocky cross-stratification, tropical hurricanes and intense winter storms. *Sedimentology*, **34**, 338–344.

Taleb, N. N., 2016: *Fooled by randomness: The hidden role of chance in life and in the markets*. Editeurs divers USA,.

Tang, B. H., and D. J. D.-:10. 1029/2004GL021072 Neelin, 2004: ENSO Influence on Atlantic hurricanes via tropospheric warming. *Geophy. Res. Lett.*, **31**.

Thomas, L. N., and X. Zhai, 2022: Chapter 5 - The lifecycle of surface-generated near-inertial waves. *Ocean Mixing*, M. Meredith and A. Naveira Garabato, Eds., Elsevier, 95–115.

Tippett, M. K., S. Camargo, and A. H. Sobel, 2011: A Poisson regression index for tropical cyclone genesis and the role of large-scale vorticity in genesis. *J. Climate*, **24**, 2335–2357.

Tompkins, A. M., and A. G. Semie, 2017: Organization of tropical convection in low vertical wind shears: Role of updraft entrainment. *Journal of Advances in Modeling Earth Systems*, **9**, 1046–1068, https://doi.org/10.1002/2016MS000802.

Tory, K. J., R. A. Dare, N. E. Davidson, J. L. McBride, and S. S. Chand, 2013: The importance of low-deformation vorticity in tropical cyclone formation. *Atmospheric Chemistry and Physics*, **13**, 2115–2132, https://doi.org/10.5194/acp-13-2115-2013.

Trouet, V., G. L. Harley, and M. Domínguez-Delmás, 2016: Shipwreck rates reveal Caribbean tropical cyclone response to past radiative forcing. *Proceedings of the National Academy of Sciences*, **113**, 3169–3174, https://doi.org/10.1073/pnas.1519566113.

Vecchi, G. A., and T. R. D.-:10. 1175/2010JCLI3810. 1 Knutson, 2011: Estimating annual numbers of Atlantic hurricanes missing from the HURDAT database (1878-1965) using ship track density. *J. Climate*, **24**.

Vecchi, G. A., and Coauthors, 2019: Tropical cyclone sensitivities to CO2 doubling: roles of atmospheric resolution, synoptic variability and background climate changes. *Climate Dynamics*, **53**, 5999–6033, https://doi.org/10.1007/s00382-019-04913-y.

Viale, F., and T. M. Merlis, 2017: Variations in tropical cyclone frequency response to solar and CO2 forcing in aquaplanet simulations. *Journal of Advances in Modeling Earth Systems*, **9**, 4–18, https://doi.org/10.1002/2016MS000785.

Vickery, P. J., P. F. Skerjl, and L. A. Twisdale, 2000: Simulation of hurricane risk in the U. S. using empirical track model. *J. Struct. Eng.*, **126**, 1222–1237.

Wallace, E. J., S. Coats, K. Emanuel, and J. P. Donnelly, 2021a: Centennial-Scale Shifts in Storm Frequency Captured in Paleohurricane Records From The Bahamas Arise Predominantly From Random Variability. *Geophysical Research Letters*, **48**, e2020GL091145, https://doi.org/10.1029/2020GL091145.

——, S. G. Dee, and K. A. Emanuel, 2021b: Resolving long-term variations in North Atlantic tropical cyclone activity using a pseudo proxy paleotempestology network approach. *Geophysical Research Letters*, **48**, e2021GL094891, https://doi.org/10.1029/2021GL094891.

Wang, S., J. Fang, X. Tang, and Z.-M. Tan, 2022: A survey of statistical relationships between tropical cyclone genesis and convectively coupled equatorial Rossby waves. *Advances in Atmospheric Sciences*, **39**, 747–762, https://doi.org/10.1007/s00376-021-1089-8.

Wheeler, M. C., and H. H. Hendon, 2004: An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Monthly Weather Review*, **132**, 1917–1932, https://doi.org/10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2.

Wu, L., and M. Takahashi, 2018: Contributions of tropical waves to tropical cyclone genesis over the western North Pacific. *Climate Dynamics*, **50**, 4635–4649, https://doi.org/10.1007/s00382-017-3895-3.

Xu, M., Q. Yang, M. Ying, Z. Deng, and Z. Yang, 2012: Reconstruction of tropical cyclones affecting East China during 1450-1949 from historical documents. *Tropical Cyclone Research and Review*, **1**, 482–488, https://doi.org/10.6057/2012TCRR04.06.

Yan, Q., Z. Zhang, and H. Wang, 2018: Divergent responses of tropical cyclone genesis factors to strong volcanic eruptions at different latitudes. *Climate Dynamics*, **50**, 2121–2136, https://doi.org/10.1007/s00382-017-3739-1.

Zhang, J., Y. Lin, and Z. Ma, 2021: Footprint of tropical cyclone cold wakes on top-of-atmosphere radiation. *Geophysical Research Letters*, **48**, e2021GL094705, https://doi.org/10.1029/2021GL094705.

Zhang, X., Y. Ye, and X. Fang, 2012: Reconstruction of typhoons in the Yangtze River Delta during 1644-1949AD based on historical chorographies. *Journal of Geographical Sciences*, **22**, 810–824, https://doi.org/10.1007/s11442-012-0965-7.

Zhao, C., and T. Li, 2019: Basin dependence of the MJO modulating tropical cyclone genesis. *Climate Dynamics*, **52**, 6081–6096, https://doi.org/10.1007/s00382-018-4502-y.

Zhao, H., X. Jiang, L. Wu, and P. J. Klotzbach, 2019: Multi-scale interactions of equatorial waves associated with tropical cyclogenesis over the western North Pacific. *Climate Dynamics*, **52**, 3023–3038, https://doi.org/10.1007/s00382-018-4307-z.