ATMOSPHERIC MODELS AS DYNAMICAL SYSTEMS

Edward N. Lorenz
Department of Earth, Atmospheric, and Planetary Sciences
Massachusetts Institute of Technology
Cambridge, MA 02139, U.S.A.

ABSTRACT

We describe various types of approximation which have been introduced into the atmospheric equations to convert them into models. These models may be treated as dynamical systems. We examine one model in detail, and we enumerate some atmospheric problems where a nonlinear-dynamical approach might yield beneficial results.

1. Introduction--models

The laws which govern the atmosphere may be expressed as a system of nonlinear equations. Deducing the typical behavior of the atmosphere from these equations constitutes a challenging problem in nonlinear dynamics. In attacking this problem one might expect to be guided by some of the recent studies in dynamical-systems theory, and one's first reaction to our title might be, "Why models of the atmosphere? Why not the real thing?" To understand our preference for models one needs to know what constitutes a dynamical system. One must also take a close look at the real atmosphere, and at the nature of the systems which comprise most atmospheric models.

We sometimes define a dynamical system as a finite system of coupled deterministically formulated ordinary differential equations in as many dependent variables [1]. Sometimes we relax the requirements to allow a countably infinite number of equations. Sometimes our systems consist of difference equations rather than differential equations. Whatever modifications we may permit, our interest is mainly in the long-term properties of typical solutions of the equations, rather than in methods of finding the solutions. We expect to encounter some special solutions, perhaps steady or periodic, whose properties differ considerably from those of most other solutions, but we expect that in some meaningful sense the special solutions will form a set of measure zero, so that their properties will not contribute to the overall average behavior.

What about the system of equations representing the laws which govern the atmosphere? Among these laws are the fundamental laws of hydrodynamics and thermodynamics, and we ordinarily take the attitude that they are known. A few details still elude us; for example, we do not know what determines just when a cloud, consisting of suspended water droplets or ice crystals, will release its water in the form of larger rain drops or snowflakes. Nevertheless, we are reasonably confident that a system obeying the atmospheric equations, as we have formulated them, will closely resemble the real atmosphere in its gross features and in many of its details.

What are typical solutions of these equations like? The equations are highly nonlinear, the most prominent nonlinear terms representing the quadratically nonlinear process of advection--the transport of momentum, heat, or moisture by the atmospheric motion. Any time-dependent solutions which we may be skillful or fortunate enough to discover by analytic procedures are likely to represent highly specialized behavior. In principle we can obtain typical solutions to any desired degree of approximation by numerical integration, although the actual task may be impractical. However, if our assumption regarding the exactness of the equations is correct, we can determine the nature of the typical solutions by observing the behavior of the atmosphere itself.

An outstanding characteristic of the atmosphere is the simultaneous presence of features of many spatial and temporal scales, and, in particular, many horizontal scales. There are globe-encircling westerly-wind currents, culminating in the jet streams. There are migratory vortices of subcontinental size, whose progression is responsible for many of the day-to-day weather changes in middle latitudes. There are tropical hurricanes, otherwise known as typhoons or tropical cyclones, which are less extensive but equally vigorous. There are intricately structured thunderstorms, comparable in size to large mountains. There are fair-weather cumulus clouds, often no larger than small hills. There are individual wind gusts, sometimes only broad enough to sway a single tree at one moment. Our list is but a sampling.

The above are not simply features which may be present in a correct solution of the equations; they are features which must be present in almost all time-dependent solutions. Any solution which describes only the meanderings of a westerly current, or only the progression of a chain of cyclonic and anticyclonic vortices, is a special solution, belonging to the set of measure zero whose existence we have noted.

It is evident that we lack the means for representing, even at a single instant, global fields of wind, temperature, and moisture which contain several thousand thunderstorms and hundreds of millions of gusts. In short, we are limited by the speed and capacity of today's most powerful computers, or of our brains, from determining typical solutions of the most realistic atmospheric equations which we can formulate. As a dynamical system the real atmosphere does not lend itself to convenient investigation.

In view of these limitations, how is it possible for dynamic meteorology, which was actually a well-established discipline long before the advent of computers, to accomplish anything? Several lines of pursuit are available.

We may use the equations, without actually solving them, to study various atmospheric phenomena and processes. For example, we may derive from the exact equations an expression for the time derivative of the total energy of the vortices, and we may identify the various terms in the expression with particular physical processes. If adequate observational data are available, we may then evaluate the long-term averages of the various terms, and learn which physical processes play leading roles.

Alternatively, we may introduce various approximations. A common procedure consists of linearizing the equations. The great advantage of linear systems, aside from relative ease of solution, is superposability of solutions. Thus, we may find solutions in which all features have 3000-kilometer wave lengths, and others in which they all have 3-kilometer wave lengths, and we should then be able to study large-scale vortices and cumulus-cloud circulations independently of one another. Of course, any direct influence of one feature on the other will be suppressed.

With the advent of computers, numerical methods of solution have become popular, although many dynamicists still find the earlier procedures more appealing. As we have seen, essentially exact numerical solutions of the exact equations are unattainable, and again we must introduce various approximations, but there is no need to remove the nonlinearity.

Along with the adoption of numerical techniques has come a change of perspective. Whereas we were previously content to find approximate solutions of the equations governing the atmosphere, we now take the attitude that we are finding exact solutions of models of the atmosphere. In short, our atmospheric models are simply systems of equations, derived by introducing various approximations into the exact equations, and arranged so as to be amenable to analytic or numerical integration. When the model is to be handled numerically, we may regard the finite-differencing scheme, and even the round-off procedure, as a part of the model. This contrasts with the situation in some fields, where models are often constructed simply by postulating interrelations among various features.

Many types of approximation are in common use [2]. First, we may simplify the physical nature of the atmosphere or its surroundings. For example, we often ignore the presence of gaseous, liquid, and solid water, and treat the atmosphere as an ideal gas. This step appears to handle the largest-scale features reasonably well, although it would be fatal in dealing with phenomena like tropical hurricanes, which depend upon water for their origin and maintenance. Likewise, we often replace the spherical surface of the earth by an infinite or bounded plane. We represent the effect of the earth's rotation by a force--the Coriolis force--which deflects the wind to the right in the northern hemisphere and to the left in the southern. We assume that features which would develop under such conditions are qualitatively like those which do develop above a rotating spherical earth. The globe-encircling westerly current, for example, would become rectilinear, but its vertical and cross-latitude structure might remain virtually unchanged. In conjunction with the latter simplification, we often omit the earth's topographic features.

We may instead modify or eliminate certain physical processes. If we replace the equation of vertical motion by the hydrostatic approximation, which balances gravity against the vertical pressure force, and equates the pressure at a point to the weight of a column of air extending upward from that point, we obtain a considerably simpler system which is incapable of propagating sound waves, but is scarcely distinguishable from the exact equations in its treatment of systems larger than thunderstorms. If we also replace one equation of horizontal motion by the geostrophic approximation, which balances the Coriolis force against the horizontal pressure force, and effectively equates the pressure to a stream function for the wind, so that a low pressure center and a cyclonic vortex become equivalent, we obtain a still simpler system which is incapable of propagating inertial-gravity waves, but handles the largest-scale atmospheric features fairly well outside of the tropics. A combination of the hydrostatic and geostrophic approximations equates the vertical shear of the wind to the horizontal temperature gradient. Sometimes we simply discard annoying terms from an equation with little regard for their physical meaning.

Further approximations are necessary if numerical methods of solution are to be used. The exact equations, and the equations obtained from them by introducing various physical simplifications, are generally formulated as a set of partial differential equations. If radiative heating and cooling enter explicitly, the equations will also contain integrals. Before the usual numerical procedures can be carried out, the field of each dependent variable must be represented by its values at a finite grid of points, and finite differences and sums must replace derivatives and integrals. Alternatively each variable may be expressed as a series of spherical harmonics or other orthogonal functions; multiple Fourier series may be used if plane geometry has been introduced. The equations are then transformed into a countably infinite system of ordinary differential equations, with a countably infinite number of terms in each equation. Again, all but a finite number of equations, and all but a finite number of terms in each equation, must be discarded before numerical integration can begin.

If we intend to use our model to study the smaller scales, we can resolve these scales by restricting the model to a limited area. We may include the influence of larger-scale features through prescribed boundary conditions. If instead our purpose is to study the larger scales, we must discard the small scales. However, the small scales influence the large scales; the circulation within each cumulus cloud, for example, can carry significant amounts of heat and moisture to higher levels. We are therefore well advised to include additional terms in our model, representing the probable influence of an extensive ensemble of small-scale features.

How well do models with both physical and mathematical simplifications perform? Many of them have been constructed for the purpose of weather forecasting. These typically contain thousands of equations; the chief limitation to their size is the speed and capacity of the computers which are compatible with the budgets of the various weather services. The forecasts produced by the largest models, with several hundred thousand variables, compare favorably with forecasts produced by other means, although they are far from perfect.

Models used primarily for research are sometimes equally large, but, when only qualitatively correct results are desired, they are often made much smaller. The most highly simplified models are the "low-order models", which are often designed to study specific phenomena, and where, ideally, one retains the minimum amount of physics and the minimum resolution needed to describe the phenomenon of interest [3]. Low-order models typically have fewer than a hundred variables, and sometimes fewer than ten. Not surprisingly, some of the lowest-order atmospheric models have become some of the most intensively studied dynamical systems.

In the following paragraphs we shall describe how a dynamical-systems approach may be applied to a specific model. We shall then enumerate several problems where this approach may yield beneficial results.

2. A simple model

For a model which is readily treated as a dynamical system, we choose what is perhaps the simplest set of equations which can make some claim to being a model of the global atmospheric circulation [4]. This low-order model contains only three ordinary differential equations:

$$dX/dt = -(Y^2 + Z^2) - a(X - F) \ , \tag{1}$$

$$dY/dt = XY - bXZ - (Y - G) \ , \tag{2}$$

$$dZ/dt = bXY + XZ - Z \ . \tag{3}$$

In Eqs. (1)-(3), X represents the intensity of the middle-latitude westerly wind current in the northern or southern hemisphere; the two hemispheres may be treated as mirror images of each other. Simultaneously, X represents the cross-latitude temperature gradient in either hemisphere. The wind and temperature fields are assumed to be in permanent geostrophic balance, so that a single variable can represent both. We shall refer to these combined fields as the zonal flow, using the term "flow", as we often do in meteorology, to refer not only to the motion field but also to the pressure and temperature fields which must accompany it. The horizontal and vertical structures of the zonal flow are prespecified, and only the intensity is allow to vary.

The variables Y and Z represent the cosine and sine phases of a chain of vortices superposed on the zonal flow. The horizontal and vertical structures of the vortices are prespecified, and only their longitude and intensity are allowed to vary. Relative to the zonal flow, the vortices are scaled so that $X^2 + Y^2 + Z^2$ is proportional to the total (kinetic plus potential plus internal) energy.

The vortices are linearly damped by viscous and thermal processes, and the damping time for the vortices is chosen as the time unit. The constant a is the reciprocal of the damping time for the zonal flow, and we let $a < 1$. In interpreting our results we let one time unit equal five days.

The vortices are constrained to tilt westward with increasing elevation, whence, under the assumed geostrophic balance, the poleward-moving air is warmer than the equatorward-moving air at the same latitude, and the net effect of the vortices is to transport heat poleward, thus reducing the temperature gradient. This effect accounts for the terms $-(Y^2 + Z^2)$ in Eq. (1). At the same time the transport does not alter the total energy of the atmosphere, so that the energy extracted from the zonal flow must be absorbed by the vortices; hence the terms XY in (2) and XZ in (3). The variables are scaled to make the coefficients of these terms equal to unity.

In addition to strengthening the vortices, the zonal flow transports them eastward (or westward, if $X < 0$). The constant b measures the ratio of the transport rate to the amplification rate, and we assume that $b > 1$.

The principal external driving force--the contrast between the low-latitude and high-latitude solar heating--acts directly on the zonal flow, and is represented by aF. A secondary driving force, which varies with longitude, and may be assumed to depend upon the contrasting thermal properties of the oceans and continents, acts on the vortices, and is represented by G.

In view of the simplicity of the included physical processes, it is evident that we could have constructed essentially the same model by simply postulating relationships among the variables, as is commonly done in some other disciplines. Actually, however, the model may be derived through systematic simplifications of the exact equations, including omission of moisture, introduction of the hydrostatic and geostrophic approximations, and extreme truncation.

As a dynamical system, Eqs. (1)-(3) possess a rich bifurcation structure, and the solutions exhibit many forms of behavior as the four parameters are altered. In this description we shall confine our attention to the fixed values $a = 1/4$, $b = 4$, and $F = 8$, and examine the changes in behavior as G increases from zero. Changing the sign of G has no effect other than changing the signs of Y and Z.

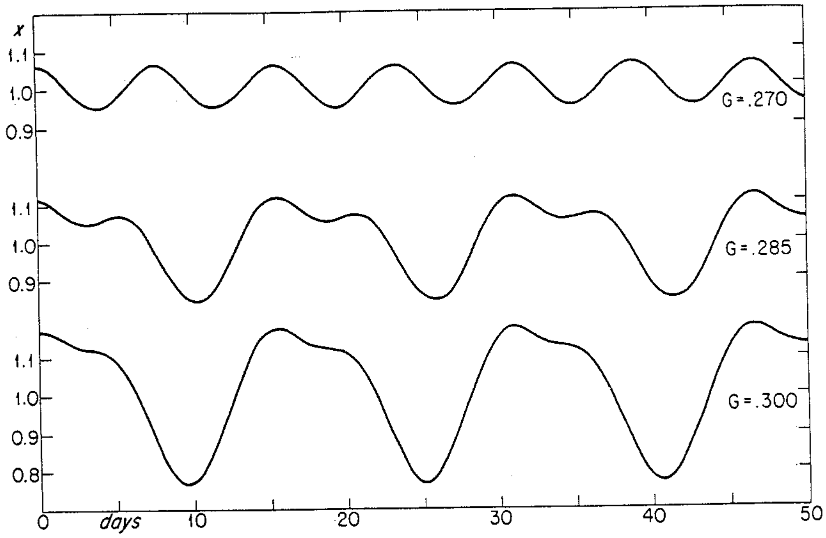

Fig. 1. Variations of X in the stable periodic solution of Eqs. (1)-(3), for G = 0.270, 0.285, and 0.300. In each case a = 0.25, b = 4.0, and F = 8.0.

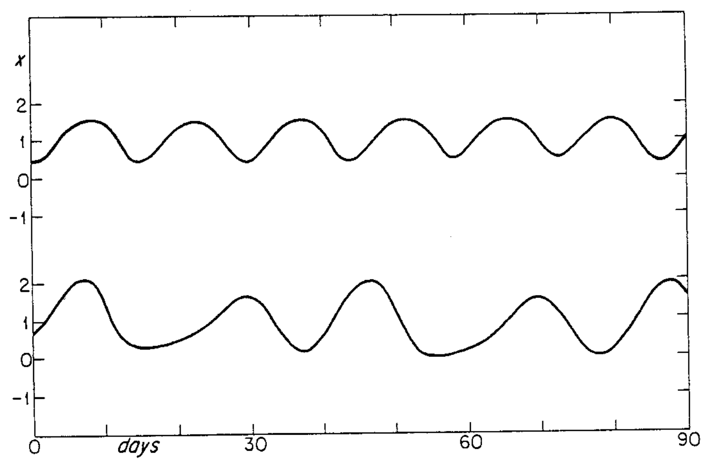

Fig. 2. Variations of X in the two distinct stable periodic solutions of Eqs. (1)-(3), for G = 0.8. As in Fig. 1, a = 0.25, b = 4.0, and F = 8.0.

When $G = 0$, the equations possess the single steady solution $X = F$, $Y = 0$, $Z = 0$, representing undisturbed eastward flow. This is seen to be unstable if $F > 1$, in which case there is a periodic solution $X = 1$, $Y = R \cos bt$, $Z = R \sin bt$, with $R^2 = a(F - 1)$, representing steadily progressing vortices. By transforming Y and Z to R and θ, where $\tan \theta = Z/Y$, we readily see that this solution is stable.

When G acquires a small positive value, we may expect a modified periodic solution in which the vortices tend to intensify when $Y > 0$ and weaken when $Y < 0$. A resultant effect of these variations of R will be oscillations of a similar period in the zonal flow X; these in turn will produce additional variations in the behavior of the vortices.

Explicit solutions when $G > 0$ may be sought numerically. We find that the anticipated behavior does occur until G reaches 0.277, when the solution becomes unstable. The bifurcation at this value at first resembles a classical period-doubling bifurcation [5], with oscillations occurring at the original frequency, but with weaker oscillations alternating with stronger ones. However, when G reaches 0.294 the weaker maximum disappears, at least in the variations of X, and the frequency has indeed been halved. Fig. 1 compares the variations of X for $G = 0.27$, 0.285, and 0.30.

No further doublings in this solution are evident, but at $G = 0.75$ a new periodic solution is born, and there are two disjoint attractors. Fig. 2 shows the variations of X for the two periodic solutions, for $G = 0.8$.

The new solution soon enters a period-doubling sequence, and becomes chaotic when G reaches 0.85, but the older solution remains stable, although its basin of attraction becomes increasingly smaller, up to $G = 0.99$, when it is swallowed up by the chaotic solution. For most values of G from 0.99 to 1.367, where a new stable steady solution appears, there is a single strange attractor, but, within this range of G, there are some intervals, notably near $G = 1.19$, where the solution is periodic. Such periodic windows are common occurrences in systems containing very few variables, and are probably rarer in more detailed atmospheric models.

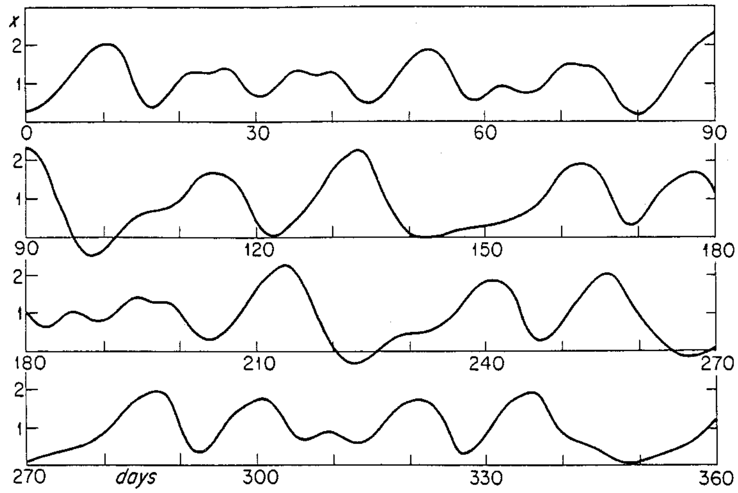

Fig. 3. Variations of X in an aperiodic solution of Eqs. (1)-(3), for G = 1.1, during a one-year interval, shown as four consecutive 90-day segments. As in Fig. 1, a = 0.25, b = 4.0, and F = 8.0.

Fig. 3 shows the variations of X during a one-year interval, when G = 1.1, displayed as four consecutive 90-day segments. The lack of periodicity is apparent, and there is some tendency to switch back and forth between weaker more rapid oscillations, like those in the upper curve in Fig. 2, and stronger less rapid oscillations, characteristic of the lower curve.

Fig. 4 shows the intersection of the attractor with the plane Z = 0, when G = 1.1. An intricate structure is evident. Qualitatively the appearance of the attractor is about the same on either side of a periodic window, and it contrasts with the small collection of points which would replace Fig. 4 in the window. Fig. 5 shows one half of the intersection of the attractor with the plane Y = 0, with higher resolution, and many details, including the interior blank areas, are more easily seen.

What does this analysis tell us about the real global atmospheric circulation? It certainly does not reveal what processes maintain the zonal flow and the vortices; at most it indicates that the processes which we have included in the model may be capable of doing so. It

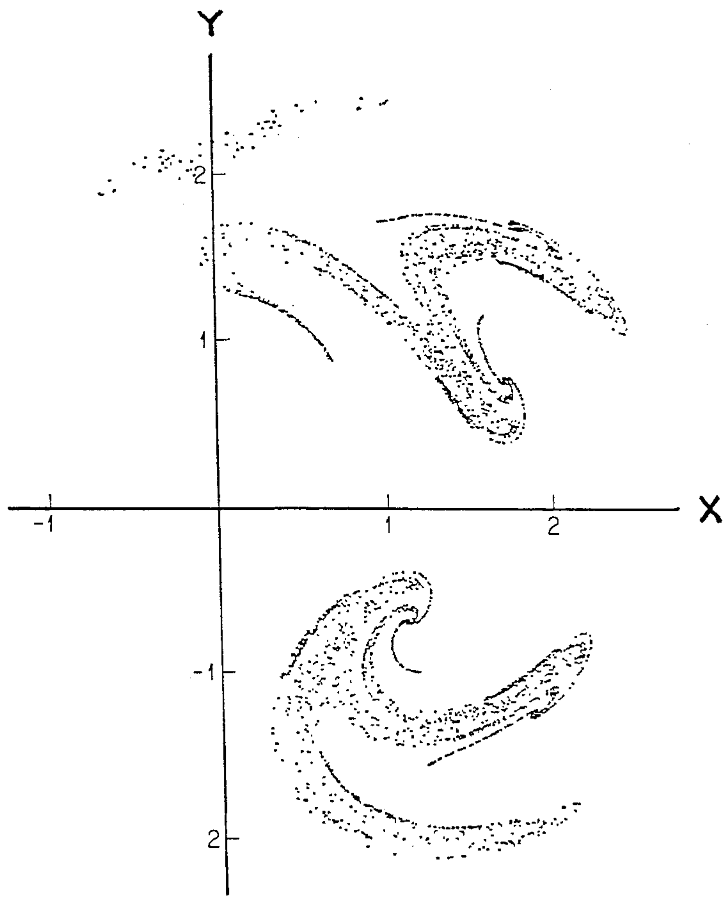

Fig. 4. The intersection of the attractor of Eqs. (1)-(3), for the conditions of Fig. 3, with the plane Z = 0, as represented by 3000 successive crossings of a single orbit.

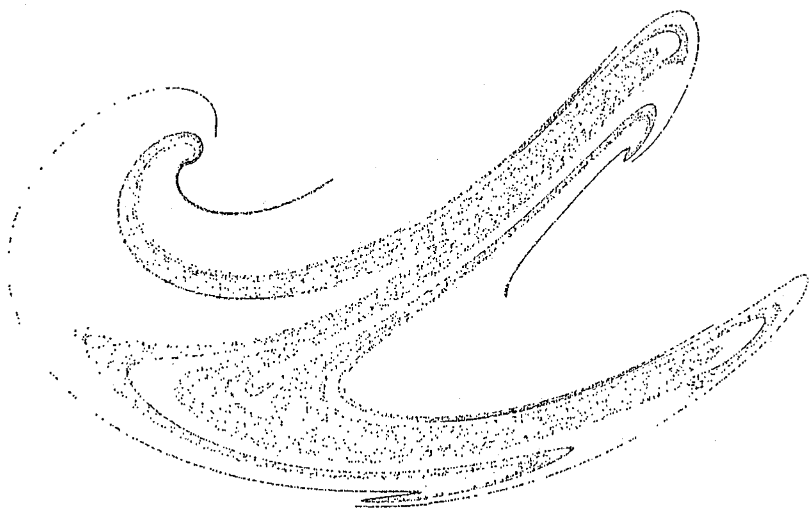

Fig. 5. A portion of the intersection of the attractor of Eqs. (1)-(3), for the conditions of Fig. 3, with the plane Y = 0, as represented by 12000 crossings of a single orbit.

does not, for example, imply that a transfer of energy from the zonal flow, rather than a transfer from smaller-scale features, is the process which maintains the vortices, since the model is incapable of saying anything about the smaller-scale features.

Perhaps the property of the model's behavior which most closely resembles its real atmospheric counterpart is the erratic variations of the zonal flow, as displayed in Fig. 3. Various conflicting explanations for such variations in the real atmosphere have been offered. The model tells us that there is no certainty that variable external activity is involved; in the model everything external is steady. Moreover, complicated internal mechanisms need not be involved; everything in the model is simple.

In the real atmosphere vortices seem to be always present; the flow never becomes purely zonal. One sometimes assumes, however, that the flow would become zonal, at least temporarily, if one waited long enough, and, in attempting to explain certain occurrences, one sometimes feels compelled to explain how they could have evolved from an essentially zonal state, or perhaps even from a state of rest. In our model the attractor clearly excludes a cylinder surrounding the X-axis, where the energy of the vortices is small. This suggests that, in the real atmosphere, states which are nearly zonal may never be approached, much less attained. The frequently used initial conditions in numerical integrations of various model equations, consisting of a zonal flow plus a small perturbation, may therefore be illogical, even if convenient.

3. Conclusion—atmospheric problems

In looking at Eqs. (1)-(3) as a dynamical system, we have gained some insight into one atmospheric problem—the coupling of the zonal flow and the large vortices—but the simplicity of the model has prevented us from treating all aspects of the problem. We could have learned more, for example, by using a model which does not assume that the vortices transport heat poleward, and instead determines its own heat transport, or a model which does not presuppose geostrophic balance, and instead produces its own balance. Meanwhile, there are numerous other atmospheric problems which presumably can be profitably investigated by constructing appropriate models and treating them as dynamical systems.

Probably the classical example of irregular or chaotic behavior is turbulent behavior. Atmospheric turbulence is especially complicated because of its inhomogeneity and intermittency. In the lowest few meters of the atmosphere, where most of us spend most of our lives, the vertical extent of a turbulent eddy is limited by the proximity of the ground or the sea, but no such limitation exists farther aloft. Turbulence covers a wide range of scales, and even the largest vortices possess some of the properties of anisotropic turbulent eddies. The volume of research in atmospheric-turbulence theory has been enormous, but many basic questions remain unanswered, and it has recently been predicted that future research will involve the concepts of strange attractors and coherent structures [6].

Coherent structures, which pose another problem, are in a sense the antithesis of turbulence, but they are equally nonlinear. They consist of features which retain their form over extended time intervals, even though they may be superposed on an essentially turbulent background. The classical example is the soliton. Coherent atmospheric structures appear to be more easily found in models than in the atmosphere itself [7], but there is some evidence that tropical hurricanes are structures of this sort.

A more specialized problem is initialization. This is one step in the process of numerical integration of operational weather-forecasting models which do not preassume geostrophic balance, and it is needed because small errors in wind and temperature observations usually produce observed initial states where the geostrophic unbalance is much greater than in the true initial state. Large-amplitude inertial-gravity waves then ensue, and contaminate the forecast. Initialization procedures attempt to replace the observed initial state by a slightly different state which is in approximate geostrophic balance, and will remain in approximate geostrophic balance as the forecast evolves. A number of initialization procedures have been developed [8], but even the more successful ones are often awkward to apply. As a problem in dynamical-systems theory, initialization may be equated with seeking a special invariant manifold, sometimes called the "slow manifold" [9]. The problem has been approached via invariant-manifold theory [10].

We close with a somewhat more detailed discussion of another problem--predictability. This concerns the extent to which it is possible to predict various aspects of the weather at various ranges. The limiting factor is the sensitive dependence of atmospheric models, and presumably of the real atmosphere, on initial conditions; our observations will not distinguish among a number of nearly identical states, and, since these states will develop differently, there will be no basis for choosing among a number of considerably different future states.

The key quantity is the rate at which small differences between states will amplify, traditionally expressed in terms of a doubling time. For a simple system like Eqs. (1)-(3), this is proportional to the reciprocal of the largest Lyapunov exponent. One might suppose that a similar relationship would hold in more general models, but

[redacted]

actually this is not the case, if the model has sufficiently high resolution.

The reason is again the abundance of scales found in the atmosphere. Errors in observing smaller-scale features, especially the more energetic ones, will grow rapidly; the error in observing the details of a thunderstorm, if such observations are indeed performed at all, should amplify at least as rapidly as the thunderstorm itself, doubling in half an hour or less. In short, the largest Lyapunov exponenets of high-resolution atmospheric models, and of the atmosphere itself, are associated with small scales.

These same errors, however, soon acquire limiting amplitudes, at a time when the errors in the larger scales are just beginning to reveal their growth. The latter errors, aside from growing more slowly and therefore being associated with smaller Lyapunov exponents, continue to grow much longer, generally doubling in two days or more [11], and they ultimately acquire much larger amplitudes than the small-scale errors; they therefore provide the major contribution to the total error in the forecast.

We need to know, then, not only how rapidly small-amplitude errors of all scales will depart from their initial magnitudes, but also how slowly large-amplitude errors will approach their limiting magnitudes. Such information is especially pertinent to extended-range prediction and the prediction of climate. We also need to know how errors in separate weakly coupled scales will influence each other. Some relevant individual studies have been performed [12], but the development of a coherent theory presents an especially challenging problem in nonlinear dynamics.

Acknowledgment. This work has been sponsored by the Climate Dynamics Program of the Atmospheric Science Section, National Science Foundation, under Grant 82-14582 ATM.

REFERENCES

1. Guckenheimer, J., and Holmes, P., 1983: Nonlinear oscillations, dynamical systems, and bifurcations of vector fields. Springer-Verlag, New York, 453 pp.

2. Holton, J. R., 1979: An introduction to dynamic meteorology, 2nd ed. Academic Press, New York, 391 pp.

3. Lorenz, E. N., 1982: Low-order models of atmospheric circulations. J. Meteorol. Soc. Japan, 60, 255-267.

4. Lorenz, E. N., 1984: Irregularity: a fundamental property of the atmosphere. Tellus, 36A, 98-110.

5. Feigenbaum, M. J., 1978: Quantitative universality for a class of nonlinear transformations. J. Stat. Phys., 19, 25-52.

6. Tennekes, H., 1985: A comparative pathology of atmospheric turbulence in two and three dimensions. Turbulence and predictability in geophysical fluid dynamics and climate dynamics, M. Ghil, R. Benzi, and G. Parisi, eds., Italian Phys. Soc., North Holland Pub. Co., Amsterdam, 45-70.

7. Malanotte Rizzoli, P., 1984: Coherent structures in planetary flows as systems endowed with enhanced predictability. Predictability of fluid motions, G. Holloway and B. J. West, eds., Amer. Inst. Physics, New York, 223-245.

8. Daley, R., 1981: Normal mode initialization. Rev. Geophys. and Space Sci., 19, 450-468.

9. Leith, C. E., 1980: Nonlinear normal mode initialization and quasi-geostrophic theory. J. Atmos. Sci., 37, 958-968.

10. Kopell, N., 1985: Invariant manifolds and quasi-geostrophic theory. Physica, 14D, 203-215.

11. Lorenz, E. N., 1982: Atmospheric predictability experiments with a large numerical model. Tellus, 34, 505-513.

12. Lorenz, E. N., 1984: Estimates of atmospheric predictability at medium range. Predictability of fluid motions, G. Holloway and B. J. West, eds., Amer. Inst. Physics, New York, 133-139.