Regimes in Simple Systems

EDWARD N. LORENZ

Massachusetts Institute of Technology, Cambridge, Massachusetts

(Manuscript received 8 September 2005, in final form 7 November 2005)

ABSTRACT

Dynamical systems possessing regimes are identified with those where the state space possesses two or more regions such that transitions of the state from either region to the other are rare. Systems with regimes are compared to those where transitions are impossible.

A simple one-dimensional system where a variable is defined at N equally spaced points about a latitude circle, once thought not to possess regimes, is found to exhibit them when the external forcing F slightly exceeds its critical value F^* for the appearance of chaos. Regimes are detected by examining extended time series of quantities such as total energy. A chain of k^* fairly regular waves develops if $F < F^*$, and F^* is found to depend mainly upon the wavelength $L^* = N/k^*$, being greatest when L^* is closest to a preferred length L_0. A display of time series demonstrates how the existence and general properties of the regimes depend upon L^*.

The barotropic vorticity equation, when applied to an elongated rectangular region, exhibits regimes much like those occurring with the one-dimensional system. A first-order piecewise-linear difference equation produces time series closely resembling some produced by the differential equations, and it permits explicit calculation of the expected duration time in either regime. Speculations as to the prevalence of regimes in dynamical systems in general, and to the applicability of the findings to atmospheric problems, are offered.

1. Introduction

The numerous quantities whose numerical values depend upon the current state of the atmosphere–ocean–earth system undergo fluctuations on many time scales. The longer-period oscillations, where one type of behavior may persist for years or even decades before giving way to another, are sometimes regarded as changes in a regime of behavior or even changes of climate. I am not sure that any distinction between regime changes and climate changes is generally agreed upon, but my preference is to regard a change as a change of regime when we can expect, perhaps because of our knowledge of what has happened in the past, that in due time the system will reacquire its earlier type of behavior, and will subsequently continue to oscillate between the two types, or perhaps among several. When the climate changes, history assures us that it also will change again, but there is no reason to expect it to revert to its former course; something entirely new may develop, particularly if new external conditions prevail. If one accepts this distinction, the oscillations to be encountered in this work are best regarded as changes of regime.

Among the most familiar regime changes in the ocean and atmosphere are the well documented shifts between El Niño and La Niña, the extreme phases of the El Niño–Southern Oscillation (ENSO) phenomenon (for a review see Philander 1990). ENSO is an example of an oscillation whose underlying mechanism appears to be completely internal to the ocean and atmosphere (Zebiak and Cane 1987). With the distinction that I have favored, its continued approximate repetitions disqualify it as a climate-change phenomenon. If at some future date a full cycle should find itself regularly requiring several decades or only a few months rather than the current several years to complete, the climate could be said to have changed.

Claims for regimes in atmospheric behavior date back at least to the work of Sir Gilbert Walker (Walker 1923, 1924), who identified a North Atlantic Oscillation, a North Pacific Oscillation, and a Southern Oscillation. For a while much of the meteorological commu-

Corresponding author address: Edward N. Lorenz, Dept. of Earth, Atmospheric, and Planetary Sciences, Room 54-1622, Massachusetts Institute of Technology, Cambridge, MA 02139.
E-mail: jmsloman@mit.edu

nity tended to regard these oscillations as uninteresting or even suspect, partly because the Southern Oscillation became less evident in the decades following Walker's studies, but the discovery of a connection between the long-recognized and predominantly oceanic El Niño and the ostensibly atmospheric Southern Oscillation has reversed this tendency (for a historical account see Rasmusson and Carpenter 1982), and the North Atlantic Oscillation is also receiving much current attention (Wallace 2000). The latter oscillation involves shifts in the latitude of the strongest zonally averaged westerly winds, and similar shifts evidently take place in the Southern Hemisphere (Kidson 1988). Interest in these phenomena has led to the recent identification of an Arctic Oscillation (Thompson and Wallace 1998).

Since regimes evidently do occur in the atmosphere, one would expect that sufficiently realistic models would be able to reproduce them, and this is indeed so. Two-level models have been especially popular, and these have included both primitive equation (PE) models (Hendon and Hartmann 1985; Robinson 1991) and quasigeostrophic (QG) models (Kravtsov et al. 2003), but the models used have ranged from the National Center for Atmospheric Research (NCAR) Community Climate Model (Branstator 1992), a multilevel PE model with rather detailed representations of the physical processes, through a 20-level PE model (Akahori and Yoden 1997), a 7-level PE model (Yu and Hartmann 1993), and a 5-level PE model (James and James 1992), each without moisture and with idealized mountains or no mountains at all, to a 5-level QG model (Itoh and Kimoto 1999) and a 3-level QG model (Kondrashov et al. 2004), and ultimately to barotropic (one level) models (Legras and Ghil 1985; Crommelin 2003) and some highly truncated barotropic systems (Reinhold and Pierrehumbert 1982; De Swart and Grasman 1987). The coupled ocean–atmosphere model of Zebiak and Cane (1987) likewise produces regimes. Except in the simplest cases, these studies have sought to duplicate specific real-world occurrences. The latitude shifts of the strongest westerlies, readily captured with the multilevel models, appear to be largely a barotropic phenomenon occurring within a baroclinic system (Wallace 1983; Kravtsov et al. 2005).

Some investigators prefer to regard the occurrence of regimes as synonymous with low-frequency variability. Others ask that the distribution of some key quantity be bimodal. It appears easier to detect bimodality in the output of a simple model than in that of a more elaborate one, or in real atmospheric data. I shall return to bimodality in the concluding section.

The purpose of this study is to demonstrate that regimes occur in a variety of simple systems not restricted to those that seek to model real physical systems where regimes are known to occur. More generally, it is proposed that the presence of regimes, like the presence of chaos, is simply a property that some dynamical systems possess while others do not. At least in the present work, every system found to possess regimes proves to be chaotic, and, since there are plenty of chaotic systems where regimes do not occur, regimes are evidently in a certain sense less common than chaos.

Many—perhaps most—familiar dynamical systems contain one or more constants whose numerical values must be specified. Strictly speaking, changing the values of the constants produces a new dynamical system, and the set of systems produced by such changes constitutes a family of dynamical systems. A family is nevertheless often referred to simply as a system.

The subsequent sections will examine two families defined by equations that were originally formulated with meteorological applications but not regimes in mind, and a secondary purpose of this study is to add to what is known about the first of these systems. A simpler system where regimes have intentionally been built in will then be introduced. The presence or absence of regimes will be determined by performing extended numerical runs and then examining the resulting time series of some fundamental quantity, such as total energy. Series of this sort, suggesting regimes, appear in Fig. 10 of Legras and Ghil (1985), Fig. 5 of Crommelin (2003), and Fig. 5 of Kravtsov et al. (2005). It will not be concluded without reservation that regimes exist in any particular instance unless the types of behavior during separate long intervals are so obviously different that a statistical test would be superfluous.

2. Regimes and related phenomena

Consider a dynamical system whose state space contains two disjoint regions A and B, with successions of states in B being recognizably different from those in A. The types of behavior in A and B will be called mode A and mode B, respectively. If certain conditions, presently to be enumerated, are met, modes A and B also qualify as regimes A and B.

The system necessarily satisfies the condition, to be denoted by A0, that transitions of the state from A to B do not occur, or else the condition denoted by A1 that they do. Conditions B0 and B1 relating to transitions from B to A are analogously defined. Combining the condition A0 or A1 with B0 or B1, one can identify four categories of dynamical system. Of special interest in category A1B1 are the average duration D_A of mode A; that is, the average time between a transition to A and the next transition from A, and an analogously defined

average duration D_B. If it is assumed that regardless of the category the initial state may lie in either A or B, the properties of the categories are, briefly:

A0B0: Two attractors, one in A, one in B.
A0B1: One attractor, in A. Long transient in B possible.
A1B0: One attractor, in B. Long transient in A possible.
A1B1: One attractor, partly in A, partly in B. Regimes or events possible.

I have chosen not to regard transitions in category A1B1 as regime changes unless D_A and D_B are both long. If a series consisting mostly of states in A is punctuated with brief visits to B, or vice versa, I find the visits better described as events.

The relevance of the separate categories, when one is interested mainly in regimes, is that category A1B1, if regimes are present, can easily be mistaken for A0B1 or A1B0 with long transient behavior, or even for A0B0, if the length of the record or the numerical run is not long enough. The danger of a misidentification can be reduced by lengthening the record, but sometimes this is not practical.

There remains the question of how long is "long." In the real atmosphere, "long" might mean a year, but one can imagine a model atmosphere where all weather phenomena have lifetimes of a year or more, and here one would presumably not want to say that all phenomena constitute regimes. Let us insist that anything that we call a regime have a long duration compared to the time scale of some other significant oscillation that the system is undergoing. In addition, if we wish to avoid any claim that regimes have simply been built into a system, we should insist that the regime time scale be long compared to any time constant in the equations, such as a dissipation time, sometimes appearing as its reciprocal, a coefficient of friction.

3. Regimes in a one-dimensional system

The first family of systems to be studied contains N dependent variables $X_1, \ldots, X_N$. The definition of X_n is extended to all values of n by letting $X_{n+N} = X_n$, so that the variables form a cyclic chain. The governing equations are

$$dX_n/dt = -X_{n-2}X_{n-1} + X_{n-1}X_{n+1} - X_n + F, \qquad (1)$$

where t is time. The variations of X_n are intended to simulate the behavior of some atmospheric quantity at N equally spaced grid points about a latitude circle, but to my knowledge, the system cannot be obtained by truncating any realistic atmospheric model.

The quadratic terms are like the advective terms in many atmospheric models in that together they do not alter the value of some nonnegative quantity, in this case the sum of the squares of the variables, to be regarded as proportional to the total energy. The linear term represents thermal and mechanical damping, and its coefficient has been eliminated by choosing the damping time, assumed to be five days, as the time unit. The constant term, representing external forcing, does not vary with n or t.

The properties of the system have been documented by Lorenz (1996) and Lorenz and Emanuel (1998), and most recently and most thoroughly by Lorenz (2005, hereafter L05). Certain functions of all N variables are found to play a significant role in the system's behavior. These include the mean $\mu = [x_n]$, the mean square $s^2 = [x_n^2]$, the variance $\sigma^2 = [x_n^2 - \mu^2]$, and the j-lag autocorrelation $r_j = [x_n x_{n+j} - \mu^2]/\sigma^2$, the square brackets denoting an average over n. Since N is finite, the most general state may be written

$$X_n = \mu + \sum A_k \cos(kn\alpha - \varepsilon_k), \qquad (2a)$$

where $\alpha = 2\pi/N$, in which case

$$r_j = \sum A_k^2 \cos(jk\alpha) \Big/ \sum A_k^2, \qquad (2b)$$

the summations over k running from 1 to $N/2$ or $(N - 1)/2$.

Solutions of Eq. (1) are best found by numerical integration. Here the "standard" fourth-order Runge–Kutta scheme with a time increment of 0.05 units or 6 h has invariably been used. For $N > 3$ the solutions prove to be stably periodic if F is small enough and chaotic if F is large enough.

For any chosen values of N and F, the presence or absence of regimes will be determined mainly by visually examining time series of the energy s^2. In L05 I stated that Eq. (1) appears "not to produce significant variations with periods of several months or longer, at least when chaos is fully developed." I have more recently found that for many, but not all, values of N, regimes do appear when F slightly exceeds its critical value for the appearance of chaos.

A case that unequivocally exhibits regimes occurs when $N = 21$ and $F = 5.1$. Figure 1a shows a 50-yr time series of s^2 for this case, displayed as two 25-yr segments. There are two distinct types of behavior—a less active mode A where s^2 generally remains between 6.0 and 7.0, and a more active mode B where the range of s^2 is several times as large. Each mode generally persists for over a year, and sometimes more than five years.

Before accepting the modes as regimes, one needs to confirm the presence of fluctuations that are too rapid

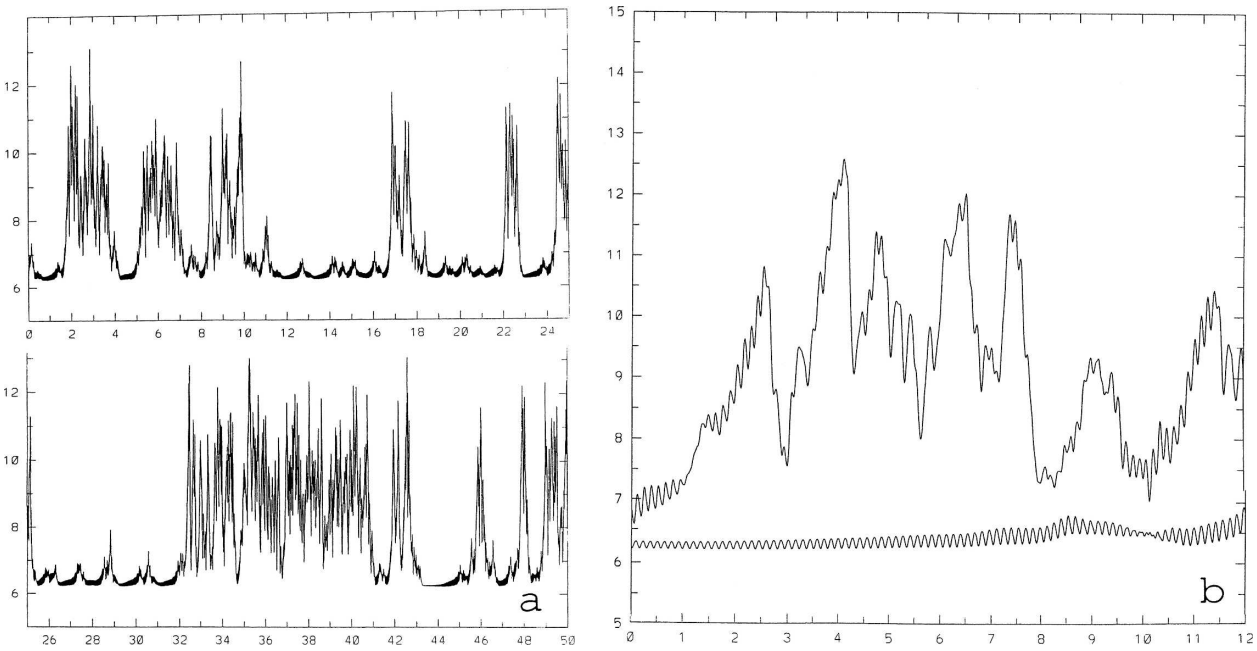

FIG. 1. (a) Fifty-year time series of s^2 produced by Eq. (1) with $N = 21$ and $F = 5.1$, displayed as two 25-yr segments. Horizontal scales are time in years. Vertical scales are units of s^2. (b) Two superposed 1-yr segments beginning 8 months into first and second years of Fig. 1a. Horizontal scale is time in months from start of each segment. Vertical scale is units of s^2.

for Fig. 1a to resolve. Figure 1b superposes two consecutive 1-yr segments, beginning 8 months into the first and second years of Fig. 1a. The less active year is dominated by weak oscillations with periods of about 4 days. The more active year also exhibits the 4-day oscillations, but week-to-week and month-to-month variations now dominate. Periods much shorter than those most conspicuous in Fig. 1a are indeed present, and since the equations also have no long time constants, the modes qualify as regimes A and B on all counts.

Figures 2a and 2b contrast the states that characterize the regimes. Each shows a sequence of cross-longitude profiles of X_n, which are complete states, at 12-h inter-

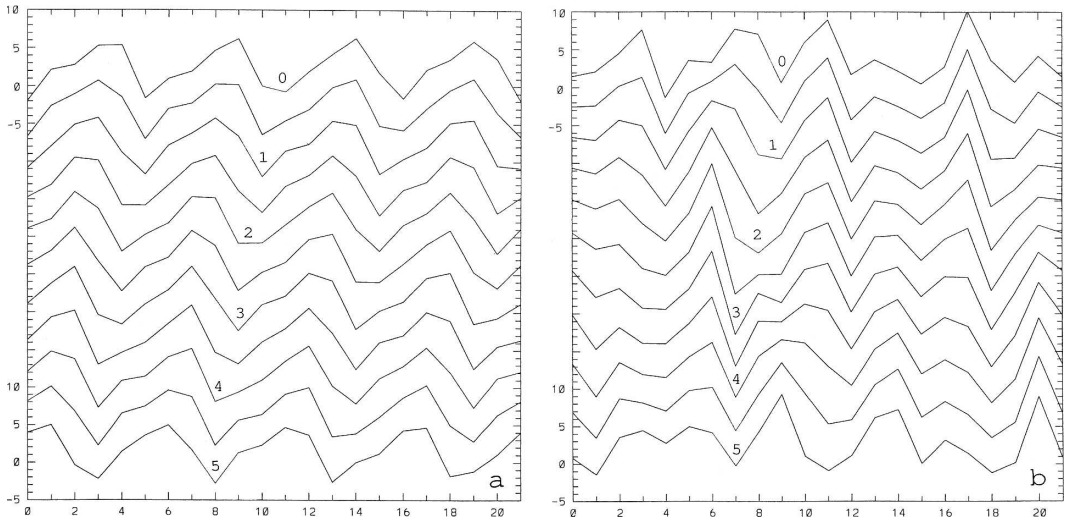

FIG. 2. (a) Profiles of X_n at 12-h intervals beginning 4 months into lower curve of Fig. 1b. Horizontal scale is gridpoint number. Vertical scales are units of X_n, for top and bottom profiles only. Numbers beside profiles are time in days from initial profile. (b) The same, but beginning 4 months into upper curve of Fig. 1b.

vals, beginning 4 months into one or the other series in Fig. 1b, and extending over 5 days. In Fig. 2a there is a clear 4-wave pattern. The waves modulate their shape, but not greatly. In Fig. 2b things are more irregular. One can most frequently count six distinct maxima and minima, but it is hard to identify by eye a dominating wavenumber.

Figure 3 presents 25-yr time series of s^2, all for $N = 21$, but for a sequence of values of F. In the first two series only mode A is present, and in the second the irregular spacing of the small peaks suggests incipient chaos. Mode B appears when $F = 5.05$ and becomes more dominant as F becomes larger, until in the final series the occurrences of mode A last only marginally long enough to qualify as regimes.

It appears, in fact, at least in this case, that the weak chaos typified by the behavior when $F = 5.0$ and the strong chaos prevailing when $F = 5.2$ or 5.25 are qualitatively different. Moreover, regimes seem to show up at the transition not from periodicity to weak chaos, but from weak to strong chaos.

Of course the series shown are only samples, but I have found no instances of regimes when $N = 21$ and $F = 4.95$ or 5.0. Some samples with $F = 5.05$ or 5.15 are hard to distinguish from the one shown for $F = 5.1$, but the increasing preference for mode B as F increases is obvious. The range of F for which regimes are clearly present is only a small fraction of the value of F.

For general values of N the convention to be adopted is that mode A is the one favored by smaller values of F, while mode B is favored by larger values. Figure 4a compares the first 25 yr of Fig. 1 when $N = 21$ and $F = 5.1$ with similar series for neighboring values of N and appropriate accompanying values of F. The series for $N = 20$ and $N = 21$ are a good deal alike, with some differences in the details, but when $N = 22$ mode A remains fairly active, while occurrences of mode B are often of short enough duration to qualify as events. The series with $N = 19$ looks unmistakably different. Qualitatively the series differ mainly in the nature of mode A, and the differences become apparent in Fig. 4b, where each series is a 1-yr mode-A segment of the corresponding series in Fig. 4a. Besides being distinguished by different appearances, the series pictured in Fig. 4 occur with distinctly different values of F. These distinctions are not unrelated.

In each series the selected value of F slightly exceeds the one needed to initiate chaos. When $N = 4$, chaos occurs only when $F > 11.96$, and one might have supposed that with successive increases in N successively smaller values of F would suffice to produce chaos, but clearly this is not the case. Figure 5a shows what actually happens.

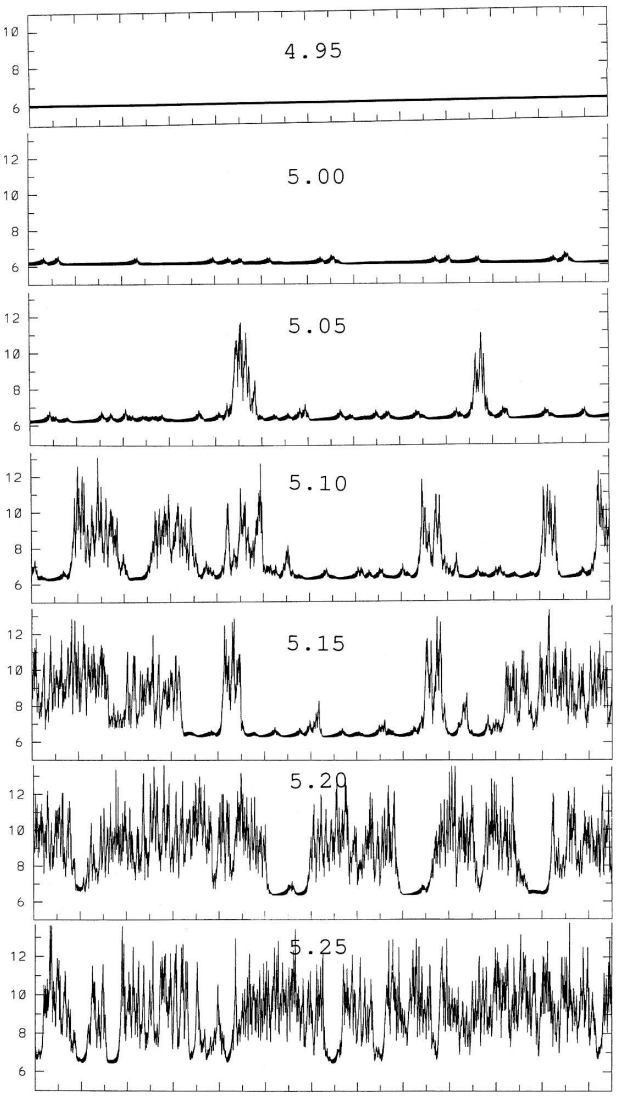

FIG. 3. Twenty-five-year time series of s^2 produced by Eq. (1) with $N = 21$, and with values of F indicated by numbers above series. Horizontal scale is time in years. Vertical scales are units of s^2.

The figure has been produced by first choosing, for each value of N, a value of F so small that chaos is patently not present, and then increasing F in increments of 0.01 and estimating, for each value of F, the leading Lyapunov exponent λ_1 from a 10-yr run. Before chaos sets in λ_1 should be zero, but, since the 10-yr runs are samples, a small positive or negative value may appear. What is shown in Fig. 5a is the value of F, to be called $F^*(N)$, for which λ_1 first exceeds 0.001.

The computations do not take into account the additional possibility of periodic windows when F exceeds F^*; that is, continua of F where chaos is absent. Such a window appears, for example, when $N = 30$ and $4.61 < F < 4.77$, even though $F^*(30) = 3.99$. They also ignore

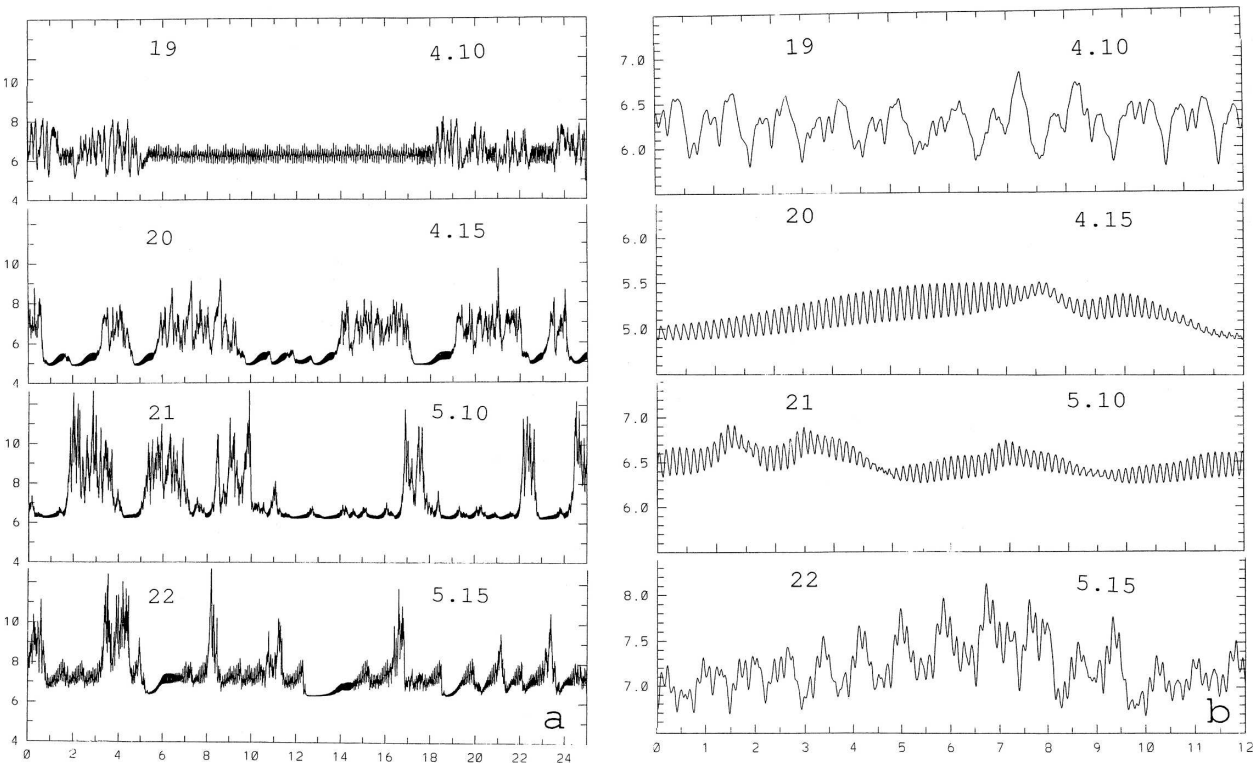

FIG. 4. (a) Twenty-five-year time series of s^2 for values of N and F indicated by numbers above series at left and right, respectively. Horizontal scale is time in years. Vertical scales are units of s^2. (b) One-year segments of curves in (a). Horizontal scale is time in months from start of each segment. Vertical scales are units of s^2.

the possibility that a chaotic solution with a separate basin of attraction from that of the periodic solution can exist even when $F < F^*$, and hence that the category is A0B0. This occurs, for example, when $N = 27$ and $5.51 < F < 5.64$, and when $N = 38$ and $5.30 < F < 5.99$. One might more properly call F^* the smallest value of F for which chaos must occur.

In Fig. 5a the vertical line at each value of N extends to $F^*(N)$. In the upper panel, except at the far left, a distinct peak appears in F^* at every fifth or sixth value of N up to 100, with decidedly smaller values between the peaks. The lower panel, extending N to 200, shows this behavior continuing nearly unabated.

Evidently the peak values of F^* are highest when N is closest to being a multiple of some fixed number L_0, which seems to be very close to 60/11, or 5.455; note the especially high peaks when $N = 60$, 120, or 180. A reasonable hypothesis is that just before chaos sets in, and sometimes even afterward during mode A, the cross-longitude profile of X_n consists of a whole number $k^*(N)$ of identifiable waves, like those in Fig. 2a, with a wavelength of $L^*(N) = N/k^*(N)$ grid points. Successive peaks in F^* correspond to unit increases in k^*. Evidently there is a preferred wavelength, namely L_0, such that waves of this length are the most stable, in the sense that the highest values of F are needed to destabilize them. When L^* cannot be close to L_0 because N is not an approximate multiple of L_0, a smaller value of F will suffice.

Figure 5b shows the values of F^* that appear in Fig. 5a, for $N > 36$, plotted against $L^*(N)$ instead of N. Most of them closely fit a piecewise-smooth curve with a sharp peak where $L^* = L_0$. A cluster of 13 points to the right of the curve fails to conform. When N is small, the fit (not shown) is not so close; evidently the process that favors wavelengths near L_0 requires that a reasonable number of waves be present.

To add further support to the hypothesis that F^* depends on L^* we note first that, as established in L05, the steady solution $X_n = F$ is unstable with respect to small-amplitude sinusoidal disturbances of wavelength L if $\cos(2\pi/L) - \cos(4\pi/L) > 1/F$. It follows that if $F > 1$, for example, the steady state is unstable with respect to waves where $4 < L < 6$. If N is fairly large there will be several wavenumbers k for which N/k lies in this range; note that L_0 lies in the range. Incipient waves will grow to finite size, and perhaps become distorted by overtones, but apparently when F becomes large

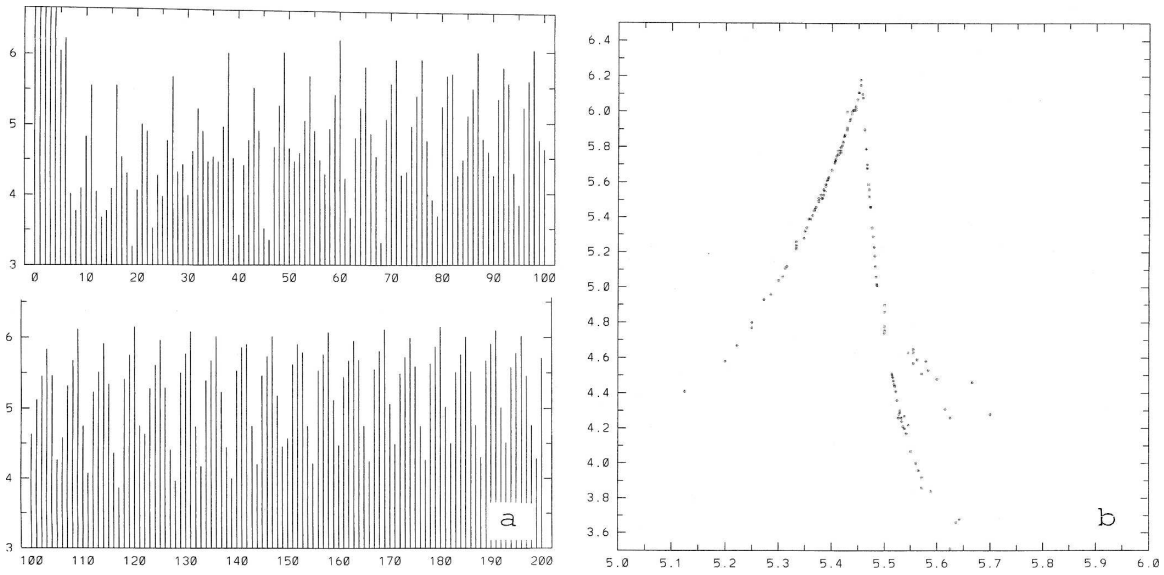

FIG. 5. (a) Minimum values of F for which chaos must occur, indicated by tops of vertical lines, for values of N indicated by horizontal scales. Vertical scales are units of F. (b) Minimum values of F for which chaos must occur, for values of wavelength L^* indicated by horizontal scale. Vertical scales are units of F.

enough they can become unstable with respect to additional disturbances, and a new pattern will develop. Sometimes this will consist of waves of a more favored length.

For an illustration of this process, other than a schematic one, it is desirable to have a measure K that can be evaluated for any state, and will be an integer when a distinct number of waves is apparent, as in Fig. 2a where K should simply equal 4. If the profile of X_n were a perfect sine curve with k waves, K should equal k, and k could be found without actually counting the waves by evaluating the autocorrelation r_1, since then, according to Eq. (2b), $r_1 = \cos(k\alpha)$. A typical profile is not a pure sine curve, but it appears that, when $F < F^*$, the profile is often closely approximated by a sine wave plus its first overtone. Prompted by this observation, I have chosen for K a function of r_1, r_2, and r_3 that will reduce to l, the smaller of two integers l and m, whenever the state is the sum of two sine curves with wavenumbers l and m, regardless of whether $m = 2l$.

To determine K, note that if $A_k = 0$ in Eq. (2a) except when $k = l$ or m, Eq. (2b) indicates that

$$[\cos(jl\alpha) - r_j]A_l^2 + [\cos(jm\alpha) - r_j]A_m^2 = 0, \quad (3a)$$

for any integer j. Since $\cos(2l\alpha) = 2c_l^2 - 1$ and $\cos(3l\alpha) = 4c_l^3 - 3c_l$, where $c_l = \cos(l\alpha)$, with analogous expressions for $\cos(2m\alpha)$ and $\cos(3m\alpha)$, it follows that

$$(c_l^j - \rho_j)A_l^2 + (c_m^j - \rho_j)A_m^2 = 0 \quad (3b)$$

for $j = 1$, 2 or 3, where $\rho_1 = r_1$, $\rho_2 = (r_2 + 1)/2$, and $\rho_3 = (r_3 + 3r_1)/4$. Eliminating A_l^2 and A_m^2 and then c_l or c_m, one finds that both c_l and c_m satisfy the quadratic equation

$$(\rho_2 - \rho_1^2)c^2 - (\rho_3 - \rho_2\rho_1)c - (\rho_2^2 - \rho_3\rho_1) = 0, \quad (4)$$

and, if c_l is the larger root, $\cos(K\alpha) = c_l$.

When the profile consists of more than two sine curves, K so determined need not be an integer, and its exact meaning is not obvious, but since, as indicated in Eq. (2b), the autocorrelations are completely determined by the spectral amplitudes A_k, K must also be a property of the spectrum. Extensive numerical tests with prespecified wavenumbers and amplitudes indicate that K is a weighted average of the prespecified numbers, with the lower numbers weighted more heavily; the other root of Eq. (4) would lead to a quantity where the higher numbers are weighted more heavily.

In addition, one may construct time series of K from runs where other quantities have already been examined. In Fig. 6 the upper curve is the top row of Fig. 1a, inverted for comparison with the lower curve, which a time series of K for the same 25 yr. There is nearly perfect correspondence between the resolved features, and, where regime A is least active, K correctly equals 4. Evidently K can serve as well as s^2 in detecting the presence of regimes. The wild fluctuations of K during regime B imply that the spectrum is likewise fluctuating wildly.

The illustrations whose desirability prompted the introduction of K may now be produced by choosing N

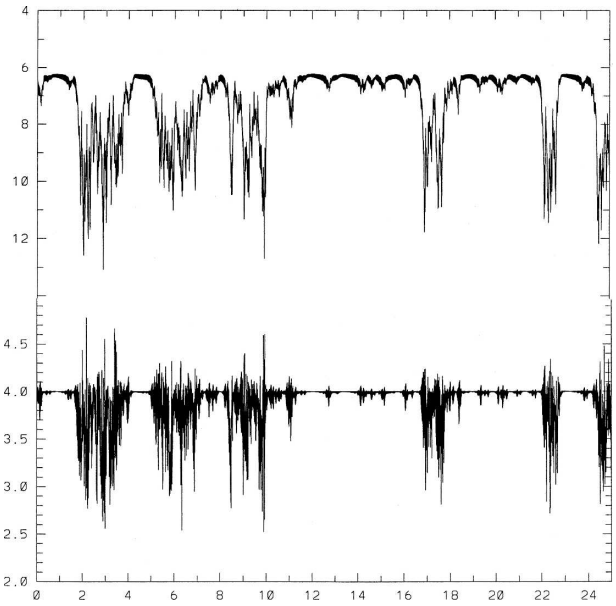

FIG. 6. Time series of (top) s^2 and (bottom) K for first 25 yr of Fig. 1a. Horizontal scale is time in years. Vertical scales are units of s^2 (top, increasing downward) and wavenumber (bottom).

and then, starting with a low value of F, increasing F at a fixed very slow rate during a run until instability is encountered, constructing time series of K during the process. In Fig. 7a, where $N = 96$, a run is initialized with a wavenumber-23 profile, making $L = 4.17$, and F increases from 1.5 to 6.5 at the rate of 0.01 units per year. At first $K = 23$, but the pattern becomes unstable before F reaches 2.0, and a stable 20-wave profile where L is closer to L_0 develops. The new pattern in turn becomes unstable and gives way to a stable 19-wave profile, which subsequently gives way to an 18-wave profile, with $L = 5.33$. At $F = 5.35$ this pattern also becomes unstable, and, there being no integer k that makes L closer to L_0, chaos sets in.

If rotated 90° counterclockwise, Fig. 7a may be interpreted as a plot of F against L, with L increasing to the right. The top point of the now-vertical line where $k = 18$, with $L = 5.33$ and $F = 5.35$, lies on or close to the curve in Fig. 5b, and the tops of the other now-vertical lines lie on what seems to be a leftward extension of this curve. As a test of this assumption, Fig. 7b has been constructed by superposing 45 runs, including the one in Fig. 7a, using each value of N from 96 to 100 and, for each N, initializing successively with a 16-wave to a 24-wave profile. The tops of the vertical segments all fit the same piecewise-smooth curve, at least to the left and somewhat to the right of the peak, and near the peak the curve coincides with the curve of Fig. 5b, to within the limits of the resolution. Evidently the defining property of the curve is instability rather than chaos per se, but, at least when $L < L_0$, this instability implies chaos when there is no wavelength closer to L_0 that can develop. When $L > L_0$, the smooth periodic behavior, on becoming unstable, often gives way to more complicated periodic behavior, and chaos sets in only after F becomes considerably larger. A consequence is the nonconforming cluster of points in Fig. 5b.

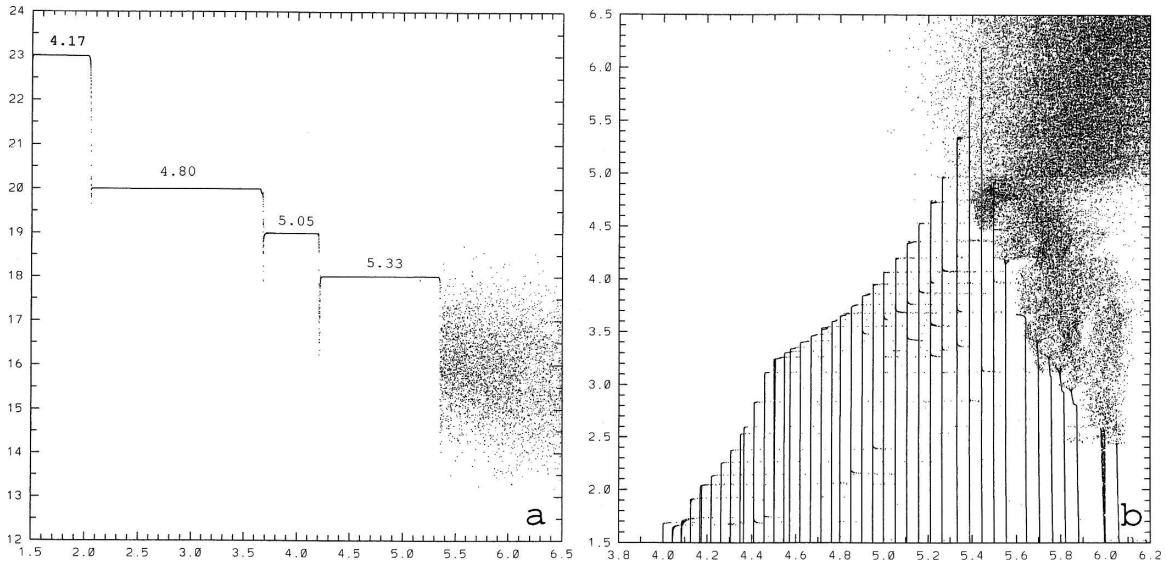

FIG. 7. (a) Variations of K with $N = 96$, with F increasing from 1.5 to 6.5, initialized with 23-wave profile of X_n. Horizontal scale is units of F. Vertical scale is wavenumber. Numbers above horizontal segments are wavelength L in grid intervals. (b) Superposition of 45 curves as in (a) with $N = 96, \ldots, 100$, each initialized with 16-, . . . , 24-wave profiles of X_n, and with pattern rotated 90° from (a). Horizontal scale is wavelength L in grid intervals. Vertical scale is units of F.

The sharpness of the peak at $L = L_0$, apparent in both Fig. 5b and Fig. 7b, strongly suggests that it is formed by the crossing of two separate smooth curves whose definitions extend to values of F above the peak. Such an occurrence would be expected if there are two distinct types of behavior with respect to which of the solutions where K is steady can become unstable. When $L > L_0$, these solutions first encounter instability with respect to one type, as F increases; when $L < L_0$, the other type develops first. This conclusion is further supported by the abrupt change in slope near $L = 4.5$. Presumably the left-hand smooth curve is only piecewise smooth, and consists of segments of separate smooth curves, and a third type of behavior develops first when $L < 4.5$.

Returning to regimes, one can conclude that differences in wavenumber, rather than differences in N per se, are responsible for the differences in the values of F needed to produce them, and for their differences in appearance. Figure 8 has been constructed with this conclusion in mind; it shows 20-yr time series of K, in each case with F at or slightly above $F^*(N)$, for values of N from 15 to 41, and it is intended to display the regimes for those values of N where they occur.

Values of N increase to the right in each row, and from row to row, as in reading a book. Each row contains values of N with the same value of k^*, ranging from 3 to 7, so that within each row L^* must increase to the right. I have found no objective rule that invariably locates the values of F producing regimes, and so, somewhat arbitrarily, in preference to subjective selection, I have used the value where the leading Lyapunov exponent λ_1, evaluated in the same manner as when determining F^*, reaches 0.2. Where F had previously been chosen subjectively, as in producing Fig. 4, there is rather good agreement.

Ironically, it is precisely those values of F in which we are most interested—those that produce regimes—for which the procedure is most suspect. The computed values of λ_1, and hence the value of F where λ_1 reaches 0.2, can depend very much on whether the 10-yr sample consists primarily of less active or more active behavior. One could increase the sample lengths, but, even without doing so, attempts to find regimes by altering F, for those values of N where regimes had not already appeared, generally failed. The lone exception was the leading series, with $N = 15$, where lowering F to 4.3 would greatly increase the typical duration of mode A, and produce a series much like the one shown for $N = 20$.

The general similarities in the patterns as one moves upward or downward in the figure, and the sometimes abrupt changes as one moves horizontally, are inescapable. A chemistry student will probably note an analogy with the periodic table of the elements, where, in the early portion, increasing an atomic number by 8 leads to an element with rather similar properties.

Especially noticeable are the pure wavenumbers that characterize much of regime A in the left-hand portion, where $L^* < L_0$. Also, the regime-B fluctuations of K strongly favor values lower than k^* there, as if emulating the final series of the previous rows. When $N = 20$ there is even a two-year interval when K falls back to 3.

When $N = 27$ or 38, L is close to L_0 and regimes do not appear; instead, as already noted, there are multiple attractors. Closer to the center, where L slightly exceeds L_0, K no longer stays so close to k^* during mode A. The mode-B fluctuations are more evenly centered about k^*, and, especially when $N = 17$, they tend to be of short duration, often resembling events. The presence of regimes is at best marginal.

Farther right, and especially in the upper right, regimes are in evidence again. There is now a tendency for K to exceed k^* during mode B, and, when $N = 25$, K advances to a pure wavenumber 5 during the third year.

Fascinating things can happen when N is very large, and multiple basins of attraction become more prevalent, as do appearances of three or more regimes. However, consideration of these ramifications is beyond the scope of this investigation.

4. The barotropic vorticity equation

The most famous prognostic equation in meteorological history may well be the barotropic vorticity equation, expressing the constancy of absolute vorticity $\eta = \zeta + f$ at a point moving with the flow. Here ζ is the vorticity relative to the rotating earth, while f, the Coriolis parameter, is the additional vorticity due to the earth's rotation. The equation was proposed by Rossby et al. (1939) as being relevant to the propagation of waves in the middle-latitude westerly wind belt, and was used by Charney et al. (1950) to produce the first moderately successful numerical weather forecast from real weather observations. The number of subsequent studies in which it has played a major role is almost uncountable.

To Rossby's formulation one must add damping and external forcing to prevent the new values of absolute vorticity from simply being rearrangements of the old ones. They will, of course, depend upon the chosen field of forcing and the nature of the damping.

The velocity and the vorticity are commonly expressed in terms of a streamfunction ψ. If following Rossby (1939) we replace the earth's surface by a plane with rectangular coordinates x and y increasing

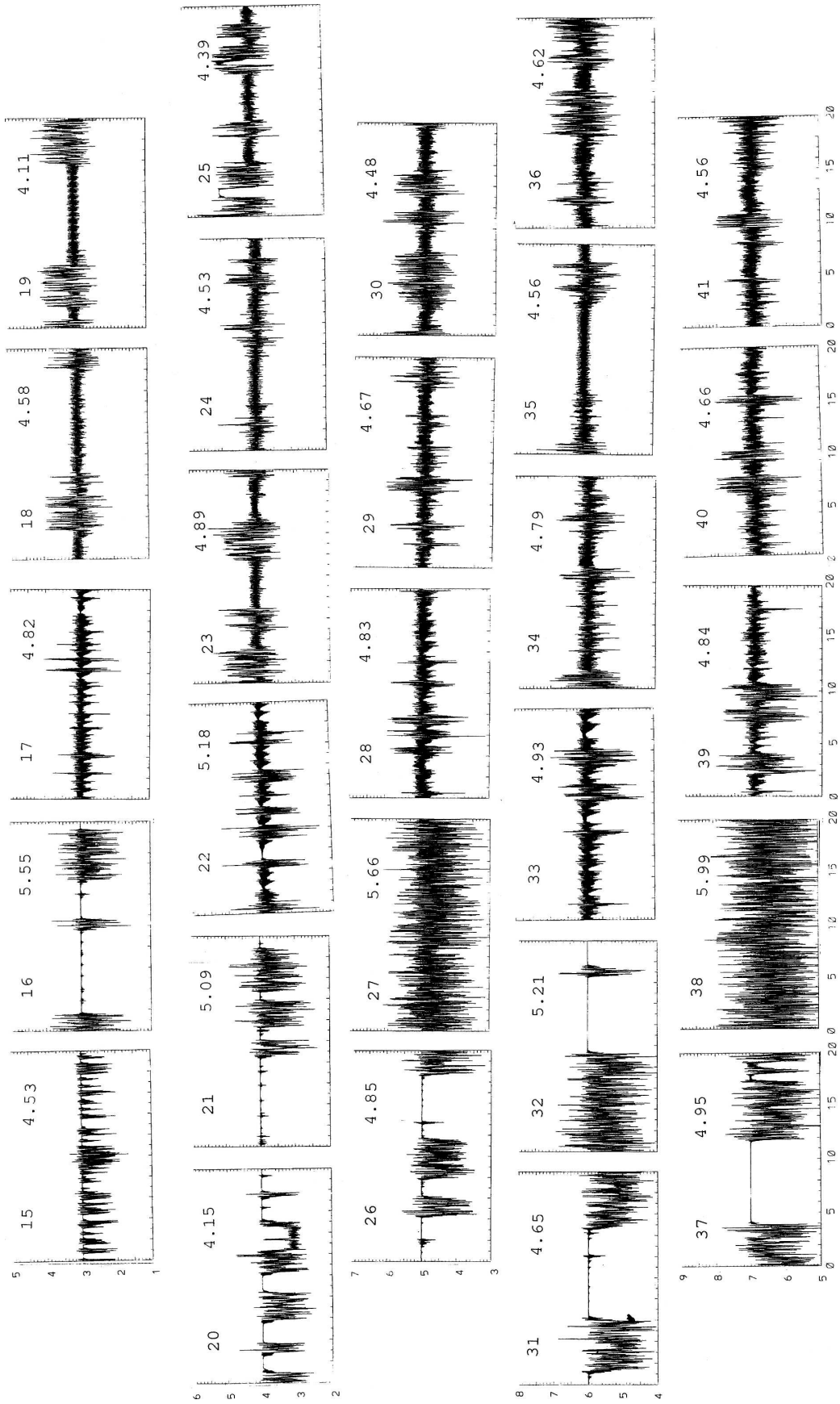

FIG. 8. Twenty-year time series of K, for values of N indicated at top left and F at top right of each panel. Each row of panels contains series with same dominant wavenumber, from (top row) 3 to (bottom row) 7. Horizontal scales appearing on lowest panels are time in years. Vertical scales appearing on lhs panels are wavenumber.

eastward and northward, and if subscripts t, x, and y denote partial differentiation, the vorticity equation may be written

$$\nabla^2\psi_t = -\psi_x\nabla^2\psi_y + \psi_y\nabla^2\psi_x - \beta\psi_x - \varepsilon(\nabla^2\psi - \nabla^2\phi), \quad (5)$$

where $\nabla^2\psi = \psi_{xx} + \psi_{yy} = \zeta$, ε is a damping coefficient, $\beta = df/dy$, and the forcing field ϕ is the field that ψ would approach if the adiabatic terms were absent. In the present formulation ψ will be defined over a rectangular region where x and y extend, respectively, from 0 to D_N and 0 to D_M, with $D_N > D_M$. Periodic boundary conditions; that is, $\psi(x + D_N, y) = \psi(x, y)$ and $\psi(x, y + D_M) = \psi(x, y)$, effectively extend the definition of ψ to the infinite plane.

There is no need to demonstrate that the barotropic vorticity equation can produce regimes. This has been aptly accomplished by various investigators who have used it to simulate real atmospheric regimes (e.g., Legras and Ghil 1985). Often these regimes consist of shifts in the latitude of the strongest westerlies. In the present formulation with periodic boundary conditions there is nothing to distinguish one latitude from another except, in some cases, the external forcing ϕ. When ϕ does not vary with latitude, anything producing a slight cross-latitude displacement need not be immediately compensated for, so that displacements can be cumulative, and the resulting slow cross-latitude drift may look locally like a regime change. Such a change need not involve any variations in total energy or prevailing wavenumber.

What is less obvious is whether the equation can produce regime changes resembling those produced by Eq. (1), where wavenumber variations play a major role. The purpose of this section is to show that some of the findings of the previous section can indeed be extended to a system [Eq. (5)] that bears a much closer resemblance to the atmosphere than Eq. (1).

For numerical computation, ψ and ζ will be represented by their values at the intersections of N longitudes and M latitudes, numbered from 0 to $N - 1$ and 0 to $M - 1$. With the periodic boundary conditions, $\psi_{nM} = \psi_{n0}$ and $\psi_{Nm} = \psi_{0m}$, the subscripts indicating longitude and then latitude. Adjacent longitudes or latitudes will be separated by the distance D, so that $D_N = ND$ and $D_M = MD$. The vorticity, at $n = 2$ and $m = 2$ for example, is taken to be

$$\zeta_{22} = (\psi_{12} + \psi_{32} + \psi_{21} + \psi_{23} - 4\psi_{22})/D^2. \quad (6)$$

Equation (6) is solved for ψ by Fourier-transforming the M values of ζ at each longitude, performing a simple matrix inversion for each Fourier component, and transforming back again, while the sum of the adiabatic terms in Eq. (5) is evaluated by the energy-and-enstrophy-conserving scheme of Arakawa (1966).

Analogously to the cyclic conditions in Eq. (1), the periodic boundary conditions allow one to express ψ as a sum of components, namely,

$$\psi_{nm} = \sum A_{kl}\cos(kn\alpha_N + lm\alpha_M - \varepsilon_{kl}), \quad (7)$$

where $\alpha_N = 2\pi/N$ and $\alpha_M = 2\pi/M$, and k runs from 0 to $N/2$ and l from $-M/2$ to $M/2$ (or from 1 to $M/2$ when $k = 0$) in the summation. For each component, k is the zonal wavenumber, while a measure of the scale is afforded by $H_{kl}^2 = 4 - 2\cos(k\alpha_N) - 2\cos(l\alpha_M)$, which approximates $(k\alpha_N)^2 + (l\alpha_M)^2$ when k and l are small. The largest-scale components are those where H_{kl} is smallest, and the contribution of a component to the mean-square wind speed is $H_{kl}^2A_{kl}^2/(2D^2)$.

In the forthcoming examples $D = 400$ km and $M = 20$, so $D_M = 8000$ km, while $\beta = 1.2 \times 10^{-11}$ m^{-1} s^{-1} and $\varepsilon = 1.16 \times 10^{-6}$ s^{-1} = (10 days)$^{-1}$. The forcing field ϕ will contain the single component $\phi_{nm} = Fb\cos(m\alpha_M)$, independent of longitude, where $b^2 = 2D^2/H_{01}^2$; this makes F the equilibrium root-mean-square wind speed. The values of N and F will vary from case to case.

Note that with D held fixed, changing N is not equivalent to changing the horizontal resolution, as might be done in updating an operational forecasting model. It corresponds instead to changing the circumference of the earth. It is therefore an exploratory procedure, and has no obvious counterpart in studies that are strictly concerned with the earth as it is. It may well have one in studies that compare several planets.

Since the adiabatic terms conserve both energy and enstrophy, and the damping is not scale-selective, increases in the energy of a small scale, if not externally forced, must be accompanied by increases in a large scale, at the expense of an intermediate one (Fjörtoft 1953). The scale of the chosen forcing, H_{01} or approximately α_M, is intermediate, since $N > M$, but the only components with larger scales; that is, with $H_{kl} < H_{01}$, are those where $l = 0$, whence H_{kl} approximates $k\alpha_N$, and where $k\alpha_N < \alpha_M$, or $N > kM$. To produce a chain of six waves ($k = 6$), for example, N must then exceed 120; from computations it appears that N must be near 200.

The total number of variables will therefore be two orders of magnitude greater than in the previous section. In addition, the amount of computation per variable per time step is an order of magnitude greater, and 3 h seems to be the greatest allowable time increment. It is thus beyond the scope of this study to treat the barotropic vorticity equation with the detail that was

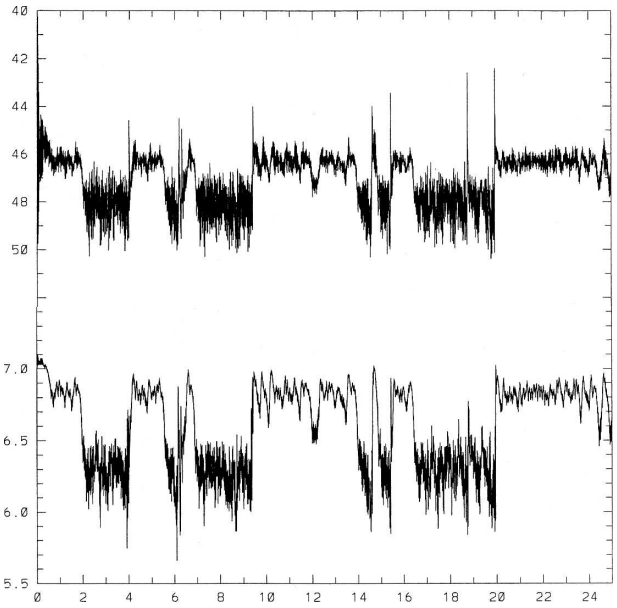

FIG. 9. Time series of (top) V and (bottom) K for same 25 yr, produced by Eq. (5) with $M = 220$ and $F = 150$ m s^{-1}. Horizontal scale is time in years. Vertical scales are m s^{-1} (top) and wavenumber (bottom).

given to Eq. (1). Instead, it will simply be demonstrated that Eq. (5) can produce regimes resembling those produced by Eq. (1), for some values of N, and will produce no regimes at all for others.

Figure 9a shows a 25-yr time series of V, the global root-mean-square wind speed, generated with $N = 220$ and $F = 150$ m s^{-1}, and drawn with V increasing downward. The extremely large equilibrium wind speed F is evidently not accompanied by comparable values of V, although V is rather large by real tropospheric standards. We see intervals sometimes longer than 4 yr when V oscillates within a narrow range about 46 m s^{-1}, interspersed with intervals of comparable length when V varies through a larger range about a slightly higher mean.

As previously, one may define a wavenumber index K in terms of the 1-, 2-, and 3-lag autocorrelations, again using Eq. (4), the lags being in the x direction. Figure 9b shows the corresponding 25-yr series of K. As in Fig. 6 there is nearly perfect agreement between the resolved features.

A decision that the two modes of behavior constitute regimes should again imply that significant shorter-period fluctuations of V or K are present. These are clearly revealed by Fig. 10, where the top panel shows the second and third years of Fig. 9b, which are typical of regimes A and B. The latter regime is dominated by an approximate 11-day periodicity, where the frequent double peaks imply superposed weaker fluctuations of even shorter period. In regime A there is also much very-short-period activity, although no single period dominates.

The remaining panels show "weather maps"—fields of ψ—for the day near month 8 of the top panel when K attains its peak value of 6.94, and the one near month

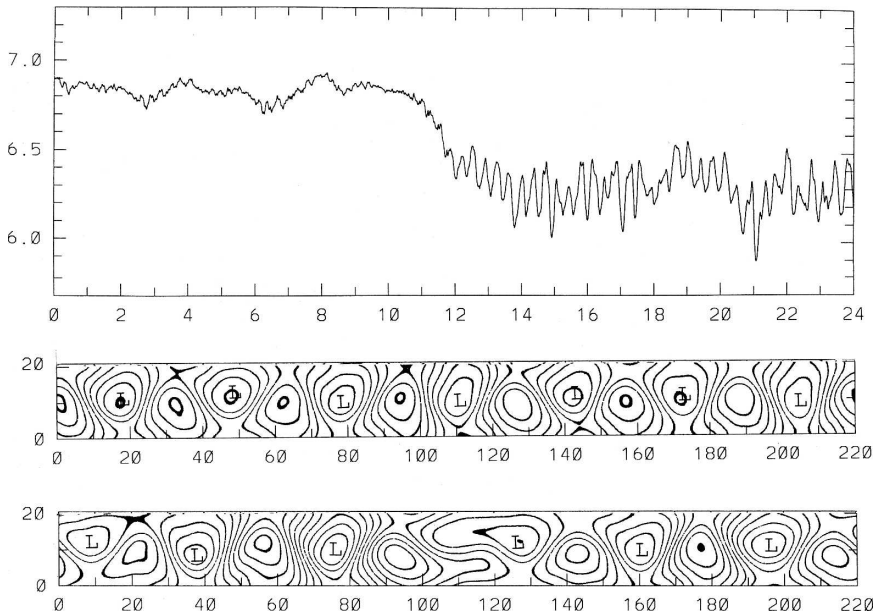

FIG. 10. (top) Second and third years of lower curve in Fig. 9. Horizontal scale is time in months. Vertical scale is wavenumber. (middle and bottom) Fields of ψ at month 8 and month 21 of (top). Low centers are labeled "L." Horizontal scales are longitude numbers. Vertical scales are latitude numbers.

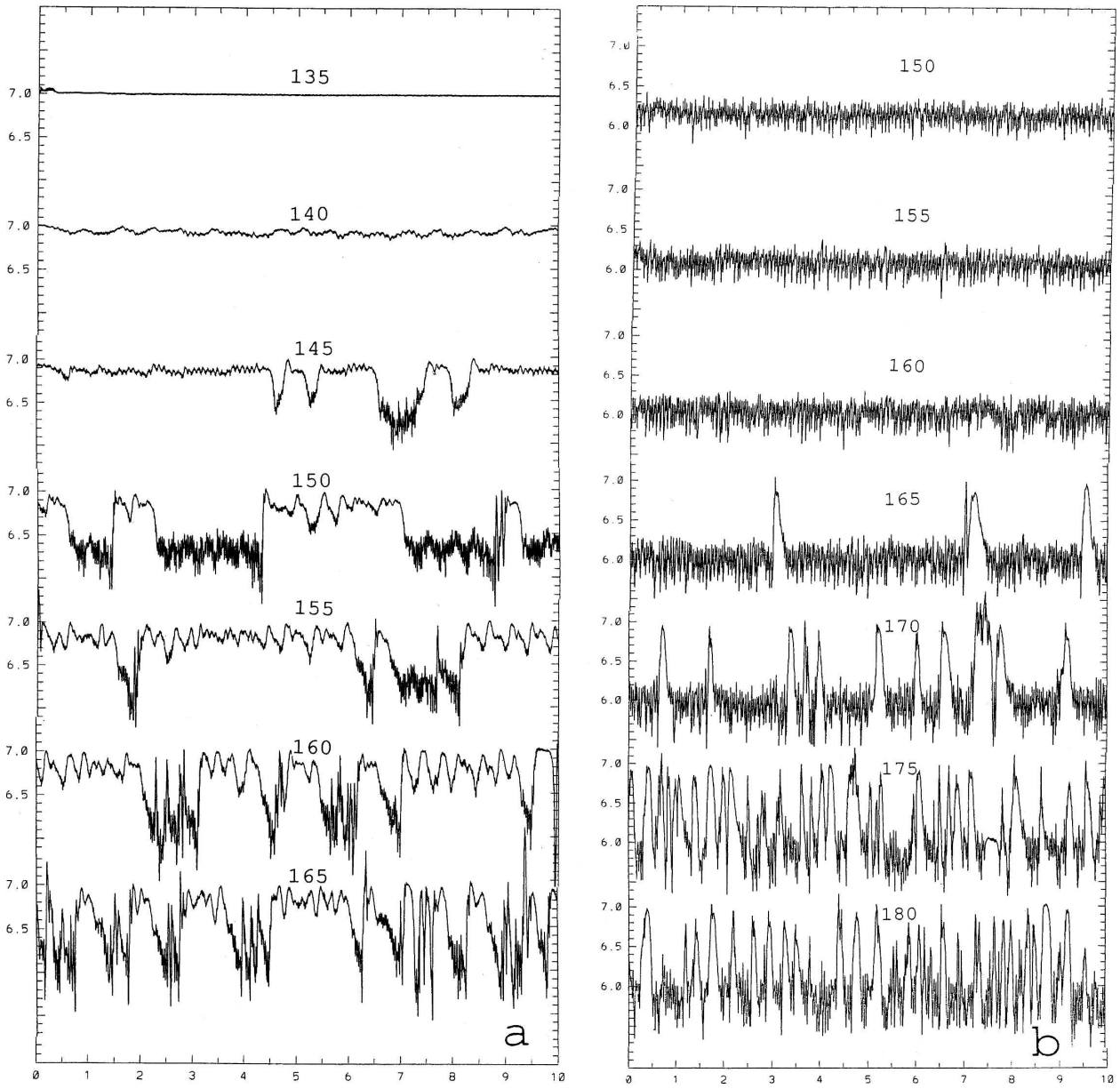

FIG. 11. (a) Ten-year time series of K produced by Eq. (5) with $N = 220$ and values of F (in m s^{-1}) above series. Horizontal scale is time in years. Vertical scales are wavenumber. (b) The same, but with $N = 210$.

21 when K reaches its minimum of 5.89. The former map reveals a chain of seven fairly regular highs and lows. In the latter there are only six, but, in the central longitudes, one high and one low have long extensions; here a closed high and a closed low were dissipating a few days before, and they are to regenerate a few days afterward. During the frequent days of regime B when K is fluctuating near 6.3, the maps generally display distorted 7-wave patterns.

Figure 11a shows that, at least when $N = 220$, regimes occur only through a fairly narrow range of F, shortly after chaos sets in. The curves are 10-yr series of K. When $F = 135$ or 140, only mode A appears; it is weakly chaotic at 140. Mode B appears at $F = 145$, and when $F = 150$ or 155, and probably 145 or 160, it qualifies as regime B. As F increases still more, the appearances of mode B look more like events. The qualitative similarity to Fig. 3, similarly produced for Eq. (1), hardly needs to be pointed out.

The preferred wavelength seems to be somewhat longer than 30 grid intervals. In Fig. 11b, N has been lowered a fraction of a wavelength to 210. Unlike what

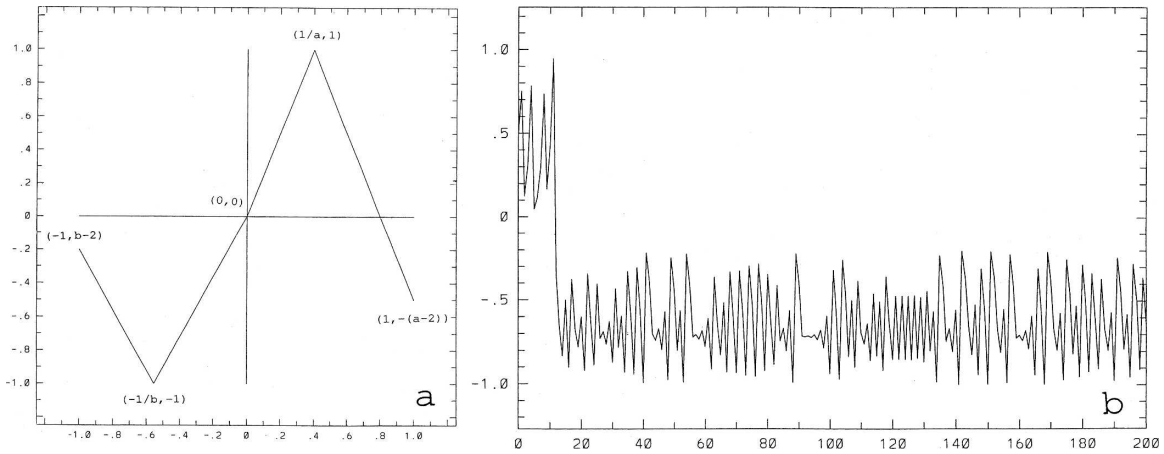

FIG. 12. (a) Graph of y_{m+1} against y_m given by Eqs. (8) with $a = 2.5$ and $b = 1.8$. Coordinates of ends of segments are shown. (b) 200-iteration time series of y for conditions of (a). Horizontal scale is iteration number. Vertical scale is units of y.

happens when $N = 220$, the mode with the lower value of K prevails at the lower values of F. When F reaches 165, a considerably higher value than the one needed to excite a second mode in Fig. 11a, the new mode appears, but its appearances qualify only as events. When F reaches 175, neither mode has a duration exceeding a few months. The effect of lowering N is somewhat like that of moving N from the beginning of a row in Fig. 8 to near the middle of the previous row.

Thus, as with Eq. (1), not all values of N lead to regimes. A comprehensive analysis of a wide range of values of N would be further complicated by the frequent appearance of more than two modes of behavior.

5. A model of the models

The regimes that are produced by the barotropic vorticity equation and the simpler Eq. (1) occur when specific conditions are met. These conditions, and also those that lead to an absence of regimes, can be modeled by a simple family of first-order difference equations, say $y_{m+1} = g(y_m)$, where y_m is the value of the single variable y that follows an initial value y_0 by m iterations. A cubic function for g will suffice (Lorenz 1965), but a piecewise linear function makes the system easier to analyze.

One such function is formed by joining a variant of the well-studied tent map (Weisstein 2003) to an inverted tent map, yielding the system

$$y_{m+1} = -ay_m + 2 \quad \text{if} \quad y_m \geq 1/a, \tag{8a}$$

$$y_{m+1} = ay_m \quad \text{if} \quad 1/a \geq y_m \geq 0, \tag{8b}$$

$$y_{m+1} = by_m \quad \text{if} \quad 0 \geq y_m \geq -1/b, \tag{8c}$$

$$y_{m+1} = -by_m - 2 \quad \text{if} \quad -1/b \geq y_m, \tag{8d}$$

where a and b are prechosen constants between 1 and 3. The positive and negative portions of the y axis will constitute the regions A and B. It is evident that $1 \geq y_{m+1} \geq -1$ if $1 \geq y_m \geq -1$.

Figure 12a is a graph of y_{m+1} against y_m, drawn with $a = 2.5$ and $b = 1.8$. One can see that with a and b so chosen, $y_{m+1} < 0$ if $y_m > 0.8$, whereupon a transition from A to B occurs. It is likewise apparent that $y_{m+1} < 0$ if $y_m < 0$, so that transitions from B to A are impossible. The category is therefore A1B0. More generally, in a–b space, the curves separating A1 from A0 and B1 from B0 are given simply by $a = 2$ and $b = 2$.

Figure 12b shows a 200-iteration series of y produced by Eq. (8), with a and b as in Fig. 12a. A comparison of the short-period fluctuations with those in Fig. 4b suggests that one iteration of Eq. (8) could be the equivalent of anything from two days to a week of the output of Eq. (1). Not surprisingly, simulating a century with Eqs. (8) is nearly three orders of magnitude faster than with Eq. (1) and about six orders faster than with Eq. (5), which in turn is many orders faster than an operational forecasting model.

With $a > 2$ and $b > 2$, permitting transitions in either direction, the appearance of the short-period fluctuations would not be greatly changed from Fig. 12b. The closer a and b are to 2, the less frequent the transitions, and hence the longer the durations D_A and D_B, defined in section 2.

Of special interest is the case where there exist positive integers I and J such that

$$a - 2 = b^{-J}, \tag{9a}$$

$$b - 2 = a^{-I}. \tag{9b}$$

Given I and J, Eqs. (9) are readily solved numerically for a and b, which will lie between 2 and 3. The range

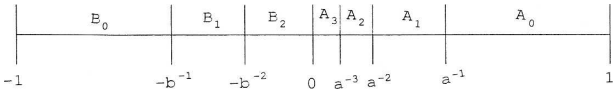

FIG. 13. The segments $B_0, \ldots, B_J, A_I, \ldots, A_0$ when $I = 3$ and $J = 2$. Values of y at end points of segments are shown.

of y from -1 to $+1$; that is, the intervals B and A, may then be divided into $I + J + 2$ segments $B_0, \ldots, B_J$, $A_I, \ldots, A_0$, bounded by the values $-1, -b^{-1}, \ldots, -b^{-J}, 0, a^{-1}, \ldots, a^{-1}, 1$ of y, as illustrated in Fig. 13 for the case $I = 3$ and $J = 2$, with the property that the map of each segment is another complete segment, or the sum of several complete segments. For example, A_1 maps onto A_0 (if $I > 1$), while A_0 maps onto the sum of $A_0, \ldots, A_I$ and B_J. Hence, if in an ensemble of solutions the probability density of y is uniform within each segment, it will remain so. It also appears that for most initial states the long-term probability density will approach uniformity in each segment as $n \to \infty$, although a proof, which would resemble a proof of an ergodic theorem, will not be offered here. This uniformity allows one to determine explicit expressions for D_A and D_B, and hence for the probabilities $P = D_A/(D_A + D_B)$ and $Q = D_B/(D_A + D_B)$ that for a randomly chosen value of m, y_m lies in A or B, and the probability $R = 1/(D_A + D_B)$ that y_m lies in A and y_{m+1} lies in B, which must equal the probability that y_m lies in B and y_{m+1} lies in A. Note that $P = D_A R$ and $Q = D_B R$.

Let the probabilities that y_m lies in $B_0, \ldots, B_J$, $A_I, \ldots, A_0$ be $Q_0, \ldots, Q_J, P_I, \ldots, P_0$, respectively. Because of the uniform probability density in each segment A_l or B_l, P_k is the sum over l of the product of each probability P_l or Q_l with the fraction of A_l or B_l that maps onto A_k. Equating these quantities yields

$$P_k = P_{k+1} + a^{-k-1}P_0 \quad \text{if} \quad k < I - 1, \tag{10a}$$

$$P_{I-1} = (1 - a^{-1})P_I + a^{-I}P_0, \tag{10b}$$

$$P_I = a^{-1}P_I + a^{-I}P_0/(a - 1) + (b - 2)Q_0/(b - 1). \tag{10c}$$

The final term in Eq. (10c) is simply the probability R of a transition from B to A. Analogous relations hold for $Q_0, \ldots, Q_L$.

Direct addition of Eqs. (10) simply confirms that $(a - 2)P_0/(a - 1) = R$, since $P_1, \ldots, P_K$ cancel out. Since $P = P_0 + \ldots + P_K$, a suitable linear combination of Eqs. (10a) and (10b) shows that $(a - 1)P = [I(a - 2) + a]\, P_0$ and, since $P = D_A R$, it follows that

$$D_A = a/(a - 2) + I. \tag{11a}$$

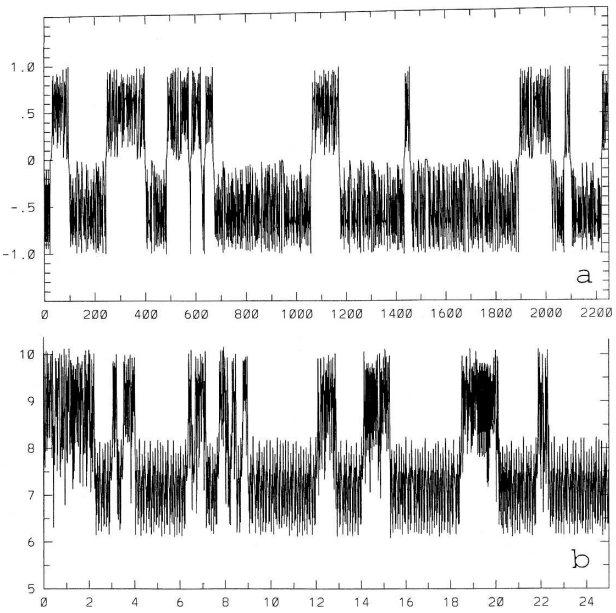

FIG. 14. (a) A 2250-iteration time series of y produced by Eqs. (8) with $I = 6$ and $J = 5$. Horizontal scale is iteration number. Vertical scale is units of y. (b) Twenty-five-year time series of s^2 produced by Eq. (1) with $N = 7$ and $F = 4.4$. Horizontal scale is time in years. Vertical scale is units of s^2.

Likewise

$$D_B = b/(b - 2) + J, \tag{11b}$$

and, as always, $R = 1/(D_A + D_B)$.

With the convention that one iteration equals 4 days, Fig. 14a presents a 25-yr time series produced by Eqs. (8) with $I = 6$ and $J = 5$, making $a = 2.0302$ and $b = 2.0143$. The regimes are as apparent as in Fig. 4a. An obvious dissimilarity is the lack of a noticeable difference between the amplitudes of the fluctuations in regimes A and B. Producing an amplitude difference with an equation like Eq. (8) could be accomplished by including an additional constant in the formulation.

Meanwhile, such an inclusion is not always wanted. Figure 14b has been constructed like the curves in Fig. 4a, with Eq. (1) but with $N = 7$ and $F = 4.4$. Despite some systematic differences the resemblance between the series in Fig. 14 is almost uncanny.

An extension of the run to 1 000 000 iterations has yielded 4663 sign changes in each direction, with an average duration of 72.4 iterations for regime A and 142.1 iterations for regime B. These sample results compare well with the expected values $R = 0.004\,559$, $D_A = 73.3$, and $D_B = 146.0$ evaluated from Eqs. (11).

If a and b are specified instead of I and J, with $a > 2$ and $b > 2$, Eqs. (9) may be solved for I and J. In general the resulting values will not be integers, and Eqs. (11)

will no longer be valid, but they appear to offer good approximations, especially when a and b are close to 2. One may then specify desired durations D_A and D_B and solve Eqs. (11) for a and b. Regimes can indeed be made to order.

6. Concluding remarks

The atmosphere–ocean–earth system is continually undergoing changes of regime. Indeed, the system is so intricate that different portions or features can simultaneously undergo their own regime changes—ENSO and the North Atlantic Oscillation, to name just two.

Many regimes are reproducible by suitable mathematical models. Some of these have been formulated and integrated for the express purpose of studying regimes. As one might have suspected, however, more general models, including some that were developed without anticipated application to the regime problem, often exhibit regimes. The present work has been concerned mainly with models of this sort.

The search for regimes has disclosed some previously unrecognized properties of the now often-used Eq. (1). Perhaps most important is the seesawing behavior, as N continually increases, of the minimum value of F for which chaos must occur.

How common are regimes in more general systems? They occur when the possible states fall into two or more sets, and transitions from one set to another are rare; this situation seems to be favored by the presence of chaos, as seen in the following scenario.

Most familiar systems contain numerical values somewhere in their formulation, and generally one can alter these values at will. Consider the case where wild chaotic fluctuations occur within either of two sets of states, but transitions are absent; each set has its own basin of attraction. An increase in a constant such as the intensity of external forcing may render the chaotic fluctuations still more intense, until one "rogue" fluctuation breaks through to what had been the other basin; the two basins have now merged. If this happens in both directions, regimes will result. Further increases in the constant may further intensify the fluctuations, until breakthroughs become so frequent that only events rather than regimes remain. One may therefore anticipate that regimes will be restricted to a rather narrow range of the constant.

Is chaos required for this scenario? Even some very simple systems possess stable periodic solutions that undergo long complicated behavior before repeating, but a sudden penetration into a new part of state space appears more likely to happen when chaotic fluctuations are occurring than when things are behaving more regularly. Since chaos typically occurs for a wide range of parameter values, the presence of regimes in a given dynamical system should be much less likely than the presence of chaos. Possibly regimes are only somewhat less likely than chaos to appear somewhere within an entire family of dynamical systems.

Although this study is largely concerned with regimes for their own sake, it is relevant to compare the regimes produced here with those encountered in the real atmosphere. Several nontrivial differences show up.

First, it is not clear that atmospheric regime changes, particularly those consisting of latitudinal shifts of the westerlies, involve significant changes in such global quantities as total energy—especially the abrupt changes often produced by the models. Next, atmospheric regimes do not seem to be associated with incipient chaos, or transitions from weak to strong chaos; atmospheric chaos tends to remain well developed. Finally, the regimes found in the models tend to occur within such narrow ranges of certain constants that it would be a rather unlikely coincidence for the corresponding atmospheric constants to lie in the necessary ranges. Here one should note, however, that many so-called atmospheric constants are not true constants; their values are controlled by the atmosphere's behavior. Perhaps, in some cases, there is a process within the atmosphere that forces a parameter to assume a value that favors regimes. The process might, for example, raise the value of the parameter whenever transitions from one mode to another become too rare, and lower it when they become too frequent. Such an idea seems difficult to verify with observations, so it must remain pure speculation until a model that produces the process is established.

Is it then proper to use Eq. (1), or perhaps a simple difference equation like Eqs. (8), as a tool in studying atmospheric regimes? If the purpose is to study the effect of regimes, as for example in seeing how the presence of regimes influences predictability, and if the equation is used with a single choice of parameter values, such use seems quite legitimate. A cubic difference equation was in fact used to illustrate the process and potential benefits of ensemble forecasting (Lorenz 1965), at a time when a demonstration with something even as simple as the barotropic vorticity equation could have been prohibitively expensive.

If instead the cause of the regimes is relevant, application of Eq. (1) to the atmosphere is somewhat questionable. It is noteworthy that when one replaces Eq. (1) by the more "atmospheric" barotropic vorticity equation, one is apparently forced to use parameter values well outside the atmospheric range to produce regimes resembling those of Eq. (1).

What about that often elusive property bimodality? In the upper panel of Fig. 4b and the upper right of Fig. 8, for example, where $N = 19$, there are clearly two types of extended-term behavior, but the distributions of the displayed quantities s^2 and K are not bimodal, nor are those of the dependent variables X_n from which s^2 and K are derived. What appears to be bimodal is something like the temporal standard deviation of s^2 within one-month intervals, but an evaluation of the probability density of this quantity from a 100-yr run indicates that it is not. The standard deviation within 6-month intervals passes the test. Indeed, it seems that in most systems one can discover something bimodal by searching enough. For instance, if a quantity is Gaussian with zero mean, its cube root is bimodal. Assuming that the time scales are appropriate, bimodality may be a sufficient condition for the existence of regimes, but my preference is not to regard it as necessary.

Meanwhile, the conditions for regimes that I have considered essential may be overly restrictive. They do not appear to permit all cases of low-frequency variability, and they may well exclude those where there are no abrupt transitions. It is even possible that I have disqualified my leading example—ENSO—which, at least in certain models, is characterized by long oceanic time constants.

Acknowledgments. I have benefited greatly from many discussions of this topic with my colleague James A. Hansen. I wish to thank Kirk Bryan for bringing to my attention some of the recent studies where regimes have been modeled. I thank an anonymous reviewer for pointing out a likeness between Eq. (1) and the equations of the celebrated Fermi–Pasta–Ulam (Fermi et al. 1955) experiment, consideration of which, although beyond the scope of this study, may lead to fruitful future research. This work has been supported by the Large-Scale Dynamic Meteorology Program, Lower Atmosphere Research Section, Division of Atmospheric Sciences, National Science Foundation under Grant ATM-0216866.

REFERENCES

Akahori, K., and S. Yoden, 1997: Zonal flow vacillation and bimodality of baroclinic eddy life cycles in a simple global circulation model. *J. Atmos. Sci.,* **54,** 2349–2361.

Arakawa, A., 1966: Computational design for long-term numerical integration of the equations of fluid motion: Two-dimensional incompressible flow. Part I. *J. Comput. Phys.,* **1,** 119–143.

Branstator, G. W., 1992: The maintenance of low-frequency atmospheric anomalies. *J. Atmos. Sci.,* **49,** 1924–1945.

Charney, J. G., R. Fjörtoft, and J. von Neumann, 1950: Numerical integration of the barotropic vorticity equation. *Tellus,* **2,** 237–254.

Crommelin, D. T., 2003: Regime transitions and heteroclinic connections in a barotropic atmosphere. *J. Atmos. Sci.,* **60,** 229–246.

De Swart, H. E., and J. Grasman, 1987: Effect of stochastic perturbations on a low-order spectral model of the atmospheric circulation. *Tellus,* **39A,** 10–24.

Fermi, E., J. Pasta, and S. Ulam, 1955: Studies in nonlinear problems. I. Los Alamos Rep. LA 1940. [Reprinted (1974) in *Nonlinear Wave Motion,* A. C. Newell, Ed., American Mathematical Society, 143–156.]

Fjörtoft, R., 1953: On the changes in the spectral distribution of kinetic energy for two-dimensional, nondivergent flow. *Tellus,* **5,** 225–230.

Hendon, H. H., and D. L. Hartmann, 1985: Variability in a nonlinear model of the atmosphere with zonally symmetric forcing. *J. Atmos. Sci.,* **42,** 2783–2797.

Itoh, H., and M. Kimoto, 1999: Weather regimes, low-frequency oscillations, and principal patterns of variability: A perspective of extratropical low-frequency variability. *J. Atmos. Sci.,* **56,** 2684–2705.

James, I. N., and P. M. James, 1992: Spatial structure of ultra-low frequency variability of the flow in a simple atmospheric circulation model. *Quart. J. Roy. Meteor. Soc.,* **118,** 1211–1233.

Kidson, J. W., 1988: Interannual variations in the southern hemispheric circulation. *J. Climate,* **1,** 1177–1198.

Kondrashov, D., K. Ide, and M. Ghil, 2004: Weather regimes and preferred transition paths in a three-level quasigeostrophic model. *J. Atmos. Sci.,* **61,** 568–587.

Kravtsov, S., A. W. Robertson, and M. Ghil, 2003: Low-frequency variability in a baroclinic β channel with land–sea contrast. *J. Atmos. Sci.,* **60,** 2267–2293.

——, ——, and ——, 2005: Bimodal behavior in the zonal mean flow of a baroclinic β-channel model. *J. Atmos. Sci.,* **62,** 1746–1769.

Legras, B., and M. Ghil, 1985: Persistent anomalies, blocking and variations in atmospheric predictability. *J. Atmos. Sci.,* **42,** 433–471.

Lorenz, E. N., 1965: On the possible reasons for long-period fluctuations of the general circulation. WMO–IUGG Symp. on Research and Development Aspects of Long-Range Forecasting, Tech. Note 66, World Meteorological Organization, 203–211.

——, 1996: Predictability: A problem partly solved. *Proc. Seminar on Predictability,* Vol. 1, Reading, United Kingdom, ECMWF, 1–18.

——, 2005: Designing chaotic models. *J. Atmos. Sci.,* **62,** 1574–1587.

——, and K. A. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.,* **55,** 399–414.

Philander, S. G. H., 1990: *El Niño, La Niña and the Southern Oscillation.* Academic Press, 289 pp.

Rasmusson, E. M., and T. H. Carpenter, 1982: Variations in tropical sea surface temperature and surface wind fields associated with the Southern Oscillation/El Niño. *Mon. Wea. Rev.,* **110,** 354–384.

Reinhold, B. B., and R. T. Pierrehumbert, 1982: Dynamics of weather regimes: Quasi-stationary waves and blocking. *Mon. Wea. Rev.,* **110,** 1105–1145.

Robinson, W., 1991: The dynamics of low-frequency variability in a simple model of the global circulation. *J. Atmos. Sci.,* **48,** 429–441.

Rossby, C.-G., 1939: Relation between variations in the intensity

of the zonal circulation of the atmosphere and the displacement of the semi-permanent centers of action. *J. Mar. Res.,* **2,** 38–55.

Thompson, D. W. J., and J. M. Wallace, 1998: The Arctic Oscillation signature in wintertime geopotential height and temperature fields. *Geophys. Res. Lett.,* **25,** 1297–1300.

Walker, G. T., 1923: Correlation in seasonal variations of weather. VIII: A preliminary study of world weather. *Mem. Indian Meteor. Dept.,* **24,** 75–131.

——, 1924: Correlation in seasonal variations of weather, IX: A further study of world weather. *Mem. Indian Meteor. Dept.,* **24,** 275–332.

Wallace, J. M., 1983: The climatological mean stationary waves: Observational evidence. *Large-Scale Dynamical Processes in the Atmosphere,* B. J. Hoskins and R. P. Pearce, Eds., Academic Press, 27–53.

——, 2000: North Atlantic Oscillation/annular mode: Two paradigms—One phenomenon. *Quart. J. Roy. Meteor. Soc.,* **126,** 791–805.

Weisstein, E. W., 2003: Tent map. *CRC Concise Encyclopedia of Mathematics,* 2d ed., Chapman and Hall/CRC, 2961 pp.

Yu, J.-Y., and D. L. Hartmann, 1993: Zonal flow vacillation and eddy forcing in a simple GCM of the atmosphere. *J. Atmos. Sci.,* **50,** 3244–3259.

Zebiak, S. E., and M. A. Cane, 1987: A model El Niño–Southern Oscillation. *Mon. Wea. Rev.,* **115,** 2262–2278.