3

Predictability – a problem partly solved

Edward N. Lorenz
Massachusetts Institute of Technology, Cambridge

Ed Lorenz, pioneer of chaos theory, presented this work at an earlier ECMWF
workshop on predictability. The paper, which has never been published

3.1 Introduction

As I look back over the many meetings that I have attended, I recall a fair number of times when I have had the pleasure of being the opening speaker. It's not that this is necessarily a special honour, but it does allow me to relax, if not to disappear

altogether, for the remainder of the meeting. On the present occasion, however, I find it is a true privilege to lead off. This is both because the subject of the seminar, predictability, is of special interest to me, and because much of the significant work

in this field has taken place here at the European Centre

In the ensuing discussion I shall frequently assume that our system is the atmosphere and its surroundings – the upper layers of the oceans and land masses – although I shall illustrate some of the points with rather crude models. By regularly calling our system the 'atmosphere' I do not mean to belittle the importance of the non-atmospheric portions. They are essential to the workings of the atmospheric portions, and, in fact, prediction of oceanic and land conditions can be of interest for its own sake, wholly apart from any coupling to the weather.

A procedure for predicting the evolution of a system may consist of an attempt to solve the equations known or believed to govern the system, starting from an observed state. Often, if the states are not completely observed, it may be possible to infer something about the unobserved portion of the present state from observations of past states; this is what is currently done, for example, in numerical weather prediction (see, for example, Toth and Kalnay, 1993). At the other extreme, a prediction procedure may be completely empirical. Nevertheless, whatever the advantages of various approaches may be, no procedure can do better than to duplicate what the system does. Any suitable method of prediction will therefore constitute, implicitly in more detail in the concluding section, any shortcoming in the extrapolation procedure will have a similar effect. Systems of this sort are now known collectively as *chaos*. In the case of the atmosphere, it should be emphasised that it may be difficult to establish the absence of an intrinsic basis for discriminating among several estimates of an initial state, and the consequent intrinsic unpredictability; some estimates that now seem reasonable to us might, according to rules that we do not yet appreciate, actually be climatologically impossible and hence rejectable, while others might, according to similar rules, be incompatible with observations of earlier states.

3.2 First estimates of predictability

Two basic characteristics of individual chaotic dynamical systems are especially relevant to predictability. One quantity is the leading Lyapunov number, or its logarithm, the leading Lyapunov exponent. Let us assume that there exists a suitable measure for the difference between any two states of a system – possibly the distance between the points that represent the states, in a multidimensional phase space whose coordinates are the variables of the system. If two states are infinitesimally close, and if both proceed to evolve according to the governing laws, the long-term average factor by which the distance between them will increase, per unit time, is the first Lyapunov number. More generally, if an infinite collection of possible initial states fills the surface of an infinitesimal sphere in phase space, the states that evolve from them will lie on an infinitesimal ellipsoid, and the long-term average factors by which the axes lengthen or shorten, per unit time, arranged in decreasing order, are the Lyapunov numbers. The corresponding Lyapunov exponents are often denoted by $\lambda_1, \lambda_2, \ldots$; a positive value of λ_1 implies chaos (see, for example, Lorenz, 1993). Unit vectors in phase space pointing along the axes of the ellipsoid are the Lyapunov vectors; each vector generally varies with time.

Our interest in pairs of states arises from the case when one member of a pair is the true state of a system, while the other is the state that is believed to exist. Their difference is then the *error* in observing or estimating the state, and, if the assumed state is allowed to evolve according to an assumed law, while the true state follows the true law, their difference becomes the *error* in prediction. In the meteorological community it has become common practice to speak of the *doubling time* for small errors; this is inversely proportional to λ_1 in the case where the assumed and true laws are the same.

The other quantity of interest is the size of the attractor; specifically, the average distance ρ between two randomly chosen points of the attractor. The attractor is simply the set of points representing states that will occur, or be approximated arbitrarily closely, if the system is allowed to evolve from an arbitrary state, and transient effects associated with this state are allowed to die out. Estimation of these quantities is fairly straightforward for mathematically defined systems – ordinarily [illegible]

may be difficult to deduce.

The third quantity that would seem to be needed for an estimate of the range of acceptable predictability is the typical magnitude of the error in estimating an initial state, ostensibly not a property of the system at all, but dependent upon our observing and inference techniques. For the atmosphere, we have a fair idea of how well we now observe a state, but little idea of what to expect in the years to come. Even though we may reject the notion of a future world where observing instruments are packed as closely as today's city dwellings, we do not really know what some undreamed-of remote-sensing technique may some day yield. However, assuming the size of an initial error, taking its subsequent growth rate to be given by λ_1, and recognising that the growth should cease when the predicted and actual states become as far apart as randomly chosen states – when the error reaches *saturation* – we can easily calculate the time needed for the prediction to become no better than guesswork.

How good are such naive estimates? We can demonstrate some simple systems where they describe the situation rather well, at least on the average. One system is one that I have been exploring in another context as a one-dimensional atmospheric model, even though its equations are not much like those of the atmosphere. It contains the K variables $X_1, \ldots, X_K$, and is governed by the K equations

$$dX_k/dt = -X_{k-2}X_{k-1} + X_{k-1}X_{k+1} - X_k + F, \tag{3.1}$$

where the constant F is independent of k. The definition of X_k is to be extended to all values of k by letting X_{k-K} and X_{k+K} equal X_k, and the variables may be thought of as values of some atmospheric quantity in K sectors of a latitude circle. The physics of the atmosphere is present only to the extent that there are external forcing and internal dissipation, simulated by the constant and linear terms, while the quadratic

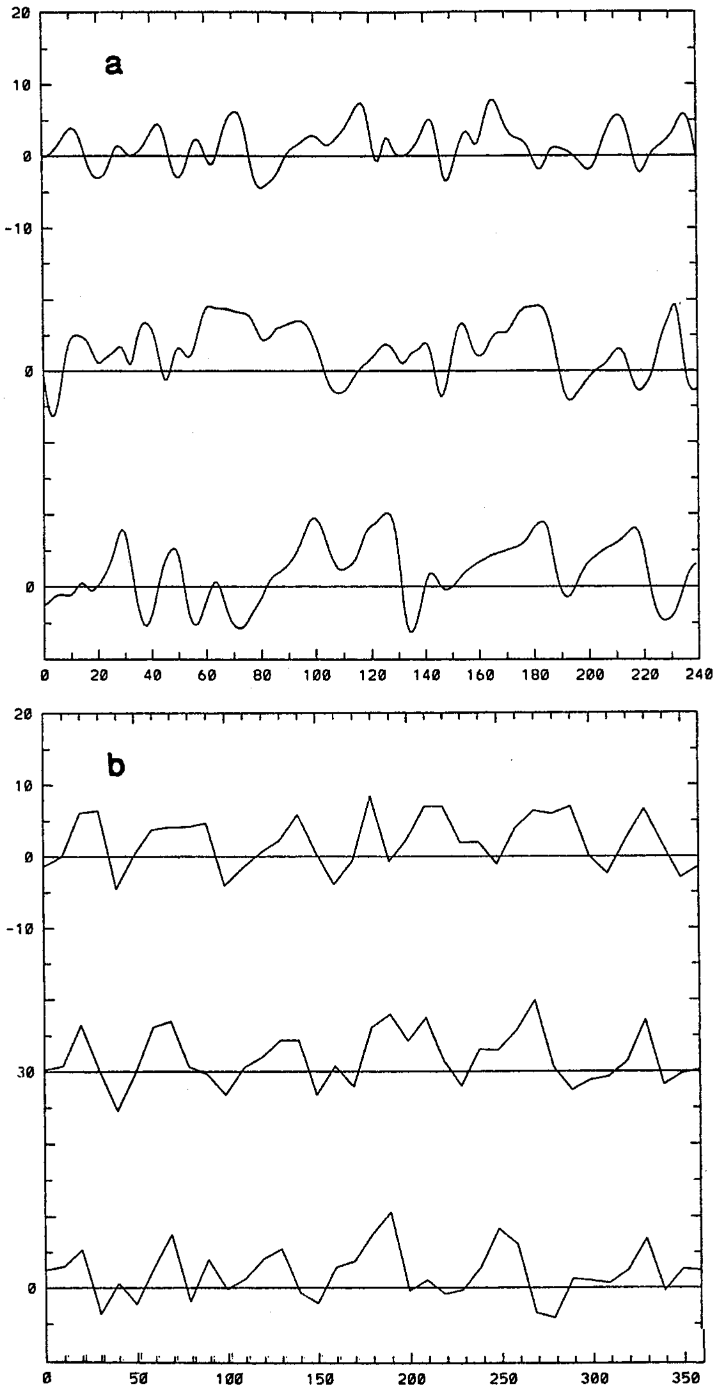

Figure 3.1 (a) Time variations of X_1 during a period of 180 days, shown as three consecutive 60-day segments, as determined by numerical integration of Eq. (3.1), with $K = 36$ and $F = 8.0$. Scale for time, in days, is at bottom. Scales for X_1 in separate segments are at left. (b) Longitudinal profiles of X_k at three times separated by 1-day intervals, determined as in (a). Scale for longitude, in degrees east, is at bottom. Scales for X_k in separate profiles are at left.

use the final values, which should be more or less free of transient effects, as new 'true' initial values, to be denoted by X_{k0}.

From Figure 3.1 we may gain some idea as to the resemblance or lack of resemblance between the *behaviour* of the model variables and some atmospheric variable such as temperature. Figure 3.1(a) shows the variations of X_1 during 720 time steps, or 180 days, beginning with the new initial conditions. The time series is displayed as three 60-day segments. There are some regularities – values lie mostly between -5 and $+10$ units, and about 12 maxima and minima occur every 60 days – but there is no sign of any true periodicity. Because of the symmetry of the model, all 36 variables should have statistically similar behaviour. Figure 3.1(b) shows the variations of X_k with k – a 'profile' of X_k about a 'latitude circle' – at the initial time, and one and two days later. The principal maxima and minima are generally identifiable from one day to the next, and they show some tendency to progress slowly westward, but their shapes are continually changing.

To produce the upper curve in Figure 3.2(a) we make an initial 'run' by choosing errors e_{k0} randomly from a distribution with mean 0 and standard deviation ε, here equal to 0.001, and letting $X'_{k0} = X_{k0} + e_{k0}$ be the 'observed' initial values of the K variables. We then use Eq. (3.1) to integrate forward from the true and also the observed initial state, for $N = 200$ steps, or 50 days, obtaining K sequences $X_{k0}, X_{k1}, \ldots, X_{kN}$ and K sequences $X'_{k0}, X'_{k1}, \ldots, X'_{kN}$, after which we let $e_{kn} = X'_{kn} - X_{kn}$ for all values of k and n.

We then proceed to make a total of $M = 250$ runs in the same manner, in each run letting the new values of X_{k0} be the old values of X_{kN} and choosing the values of e_{k0} randomly from the same distribution. Finally we let $e^2(\tau)$ be the average of the K values e^2_{kn}, where $\tau = n\Delta t$ is the prediction range, and let log $E^2(\tau)$ be the average of the M values of log $e^2(\tau)$, and plot $E(\tau)$ against the number of days (5τ), on a logarithmic scale. (The lower curve is the same except that the vertical scale is linear.)

For small n we see a nearly straight sloping line, representing uniform exponential growth, with a doubling time of 2.1 days, agreeing with λ_1, until saturation is approached. For large n we see a nearly straight horizontal line, representing saturation. It should not surprise us that the growth rate slackens before saturation is reached, rather than continuing unabated up to saturation and then ceasing abruptly.

The alternative procedure of simply letting $E^2(\tau)$ be the average value of $e^2(\tau)$, i.e. averaging the runs arithmetically instead of geometrically, would lead to a figure much like Figure 3.2(a), but with the sloping line in the upper curve indicating a

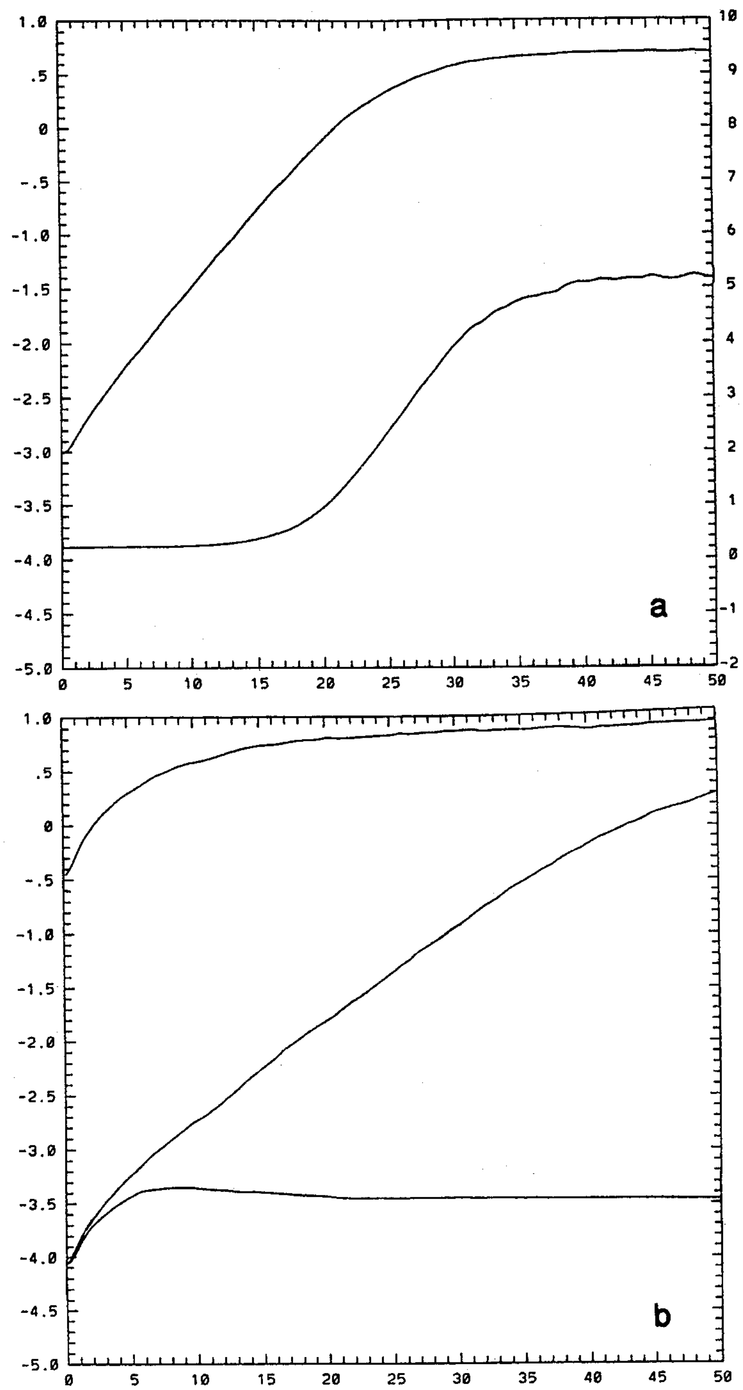

Figure 3.2 (a) Variations of average prediction error E (lower curve, scale at right) and $\log_{10} E$ (upper curve, scale at left) with prediction range τ (scale, in days, at bottom), for 50 days, as determined by 250 pairs of numerical integrations of Eq. (3.1), with $K = 36$ and $F = 8.0$ (as in Fig. 3.1). (b) The same as (a), but for variations of $\log_{10} E$ only, and as determined by 1000 pairs of integrations of Eq. (3.1), with $K = 4$, and with $F = 18.0$ (upper and middle curves, with different initial errors), and $F = 15.0$ (lower curve).

3.3 Atmospheric estimates

Some three decades ago a historic meeting, organised by the World Meteorological Organization, took place in Boulder, Colorado. The principal topic was long-range weather forecasting. At that time numerical modelling of the complete global circulation was just leaving its infancy; the three existing state-of-the-art models were those of Leith (1965), Mintz (1965), where A. Arakawa also played an essential role, and Smagorinsky (1965).

At such meetings the greatest accomplishments often occur between sessions. In this instance Jule Charney, who headed a committee to investigate the feasibility of a global observation and analysis experiment, persuaded the creators of the three models, all of whom were present, to use their models for predictability experiments, which would involve computations somewhat like those that produced Figure 3.2(a). On the basis of these experiments, Charney's committee subsequently concluded that a reasonable estimate for the atmosphere's doubling time was five days (Charney *et al.*, 1966). Taken at face value, this estimate offered considerable hope for useful two-week forecasts but very little for one-month forecasts.

The Mintz–Arakawa model that had yielded the five-day doubling time was a two-layer model. Mintz's graphs showed nearly uniform amplification before saturation

As with the Mintz–Arakawa model, the doubling times of the recent models appear consistent with the values of λ_1 for these models. Obviously not all of them can indicate the proper value of the exponent for the real atmosphere, and presumably none of them does.

Our reason for identifying the time unit in the model defined by Eq. (3.1) with five days of atmospheric time is now apparent. With $K = 36$ and $F = 8.0$ or 10.0, and indeed with any reasonably large value of K and these values of F, the doubling time for the model is made comparable to the times for the up-to-date global circulation models.

3.4 The early stages of error growth

Despite the agreement between the error growth in the simple model, and even in some global circulation models, with simple first estimates, reliance on the leading Lyapunov exponent, in most realistic situations, proves to be a considerable oversimplification. By and large this is so because λ_1 is defined as the long-term average growth rate of a very small error. Often we are not primarily concerned with averages, and, even when we are, we may be more interested in shorter-term behaviour. Also, in practical situations the initial error is often not small.

K reduced to 4 and F increased to 18.0, and with $\varepsilon = 0.0001$. Also, because so few variables are averaged together, we have increased M to 1000. Between about 6 and 30 days the curve has a reasonably uniform slope, which agrees with λ_1, and indicates a doubling time of 3.3 days, but during the first 3.3 days the average error doubles twice. Systems exhibiting anomalously rapid initial error growth are in fact not uncommon. Certainly there are practical situations where we are mainly interested in what happens during the first few days, and here λ_1 is not always too relevant.

This phenomenon, incidentally, is in this case not related to the chaotic behaviour of the model. The lower curve in Figure 3.2(b) is like the middle one, except that F has been reduced to 15.0, producing a system that is not chaotic at all. Again the error doubles twice during the first six days, but then it levels off at a value far

below saturation. If ε had been smaller, the entire curve would have been displaced downward by a constant amount.

When the initial error is not particularly small, as is often the case in operational weather forecasting, λ_1 may play a still smaller role. The situation is illustrated by the upper curve in Figure 3.2(b), which has been constructed exactly as the middle curve, except that $\varepsilon = 0.4$, or 5% of saturation, instead of 0.001. The rapid initial error growth is still present, but, when after four days it ceases, saturation is already being approached. Only a brief segment between 4 and 8 days is suggestive of 3.3-day doubling.

The relevance of the Lyapunov exponent is even less certain in systems, such as more realistic atmospheric models or the atmosphere itself, where different features possess different characteristic time scales. In fact, it is not at all obvious what the leading exponent for the atmosphere may be, or what the corresponding vector may look like. To gain some insight, imagine a relatively realistic model that resolves larger scales – planetary and synoptic scales – and smaller scales – mesoscale motions and convective clouds; forget about the fact that experiments with a global model with so many variables would be utterly impractical with today's computational facilities. Convective systems can easily double their intensity in less than an hour distinct time scales. The model has been constructed by coupling two systems, each of which, aside from the coupling, obeys a suitably scaled variant of Eq. (3.1). There are K variables X_k plus JK variables $Y_{j,k}$, defined for $k = 1, \ldots, K$ and $j = 1, \ldots, J$, and the governing equations are

$$dX_k/dt = -X_{k-1}(X_{k-2} - X_{k+1}) - X_k - (hc/b)\sum_{j=1}^{J} Y_{j,k}, \tag{3.2}$$

$$dY_{j,k}/dt = -cbY_{j+1,k}(Y_{j+2,k} - Y_{j-1,k}) - cY_{j,k} + (hc/b)X_k. \tag{3.3}$$

The definitions of the variables are extended to all values of k and j by letting X_{k-K} and X_{k+K} equal X_k, as in the simpler model, and letting $Y_{j,k-K}$ and $Y_{j,k+K}$ equal $Y_{j,k}$, while $Y_{j-J,k} = Y_{j,k-1}$ and $Y_{j+J,k} = Y_{j,k+1}$. Thus, as before, the variables X_k can represent the values of some quantity in K sectors of a latitude circle, while the variables $Y_{j,k}$, arranged in the order $Y_{1,1}, Y_{2,1}, \ldots, Y_{J,1}, Y_{1,2}, Y_{2,2}, \ldots, Y_{J,2}, Y_{3,1}, \ldots$, can represent the values of some other quantity in JK sectors. A large value of J implies that many of the latter sectors are contained in one of the former, and we may think of the variables $Y_{j,k}$ as representing a convective-scale quantity, while, in view of the form of the coupling terms, the variables X_k should represent something that favours convective activity, possibly the degree of static

instability

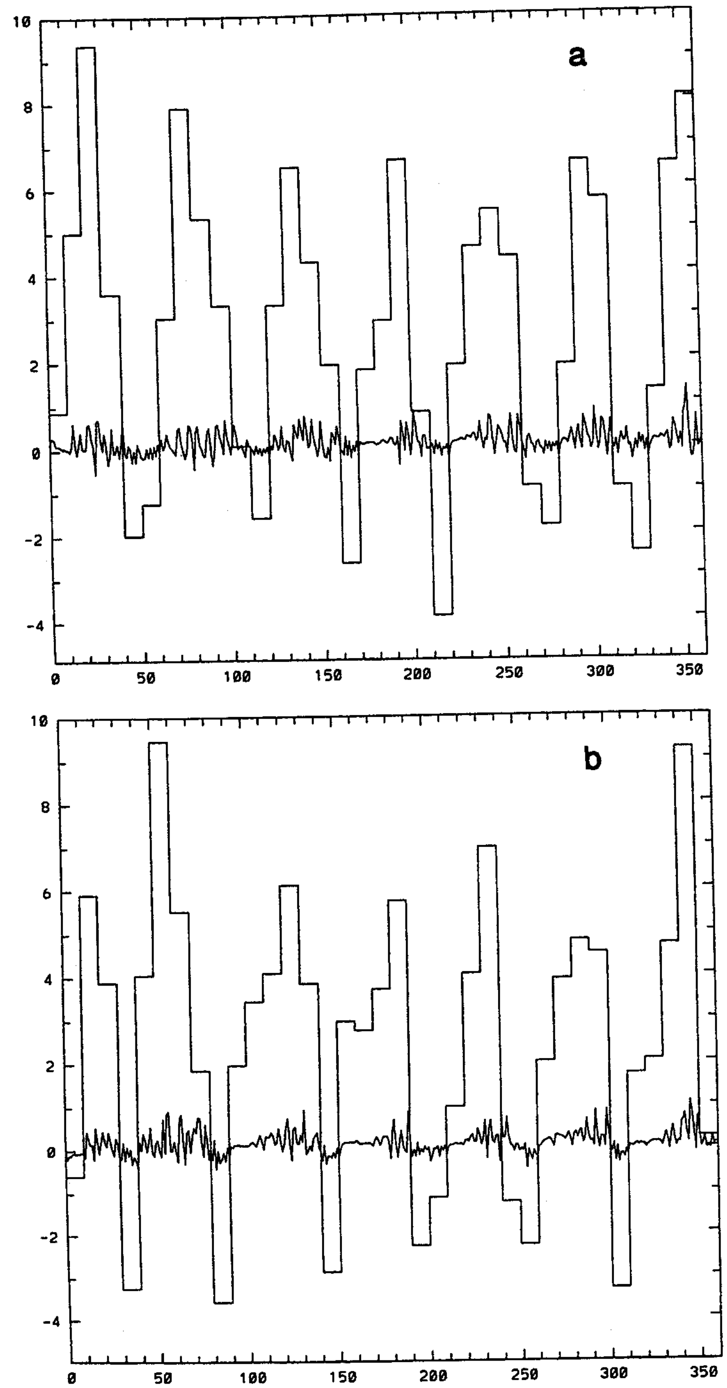

Figure 3.3 (a) Longitudinal profiles of X_k and $Y_{j,k}$ at one time, as determined by numerical integration of Eqs. (3.2) and (3.3), with $K = 36$, $J = 10$, $F = 10.0$, $c = 10.0$, $b = 10.0$, and $h = 1.0$. Scale for longitude, in degrees east, is at bottom. Common scale for X_k and $Y_{j,k}$ is at left. (b) The same as (a), but for a time two days later.

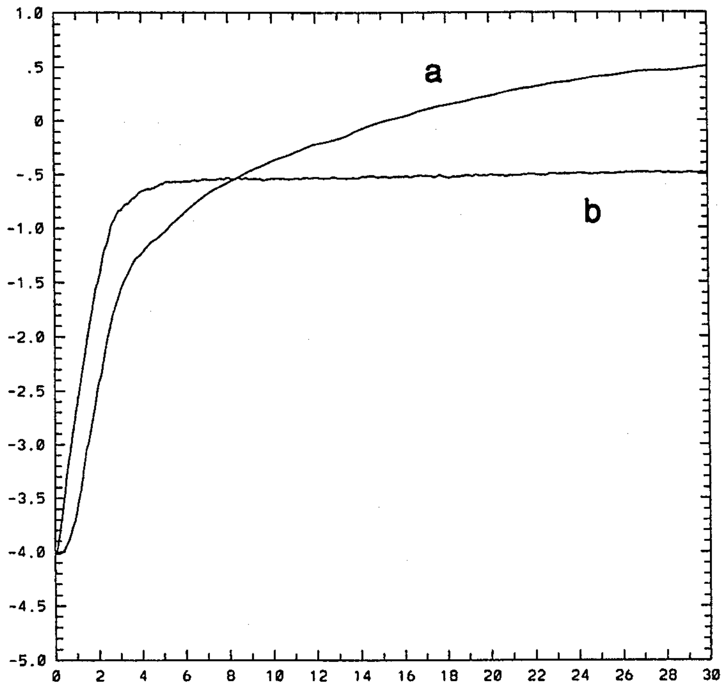

Figure 3.4 Variations of $\log_{10} E$ (scale at left) with prediction range τ (scale, in days, at bottom), shown separately for large scales (variables X_k, curve a) and small scales (variables $Y_{j,k}$, curve b), for 30 days, as determined by 25 pairs of integrations of Eqs. (3.2) and (3.3), with the parameter values of Figure 3.3.

continue to grow, at a slower quasi-exponential rate comparable to what appears in Figure 3.2(a), doubling in about 1.6 days. Finally they approach their own saturation level, an order of magnitude higher than that of the small-scale errors. Thus, after the first few days, the large-scale errors behave about as they would if the forcing were slightly weaker, and if the small scales were absent altogether.

In a more realistic model with many time scales or perhaps a continuum, we would expect to see the growth rate of the largest-scale errors subsiding continually, as, one after another, the smaller scales reached saturation. Thus we would not expect a large-scale-error curve constructed in the manner of Figure 3.4 to contain an approximate straight-line segment of any appreciable length.

We now see the probable atmospheric significance of the error doubling times of the various global circulation models. Each doubling time appears to represent the rate at which, in the *real* atmosphere, errors in predicting the features that are resolvable by the particular model will amplify, after the errors in unresolvable features have reached saturation. Of course, before accepting this interpretation, we must recognise the possibility that some of the small-scale features will not saturate rapidly; possibly they will act in the manner of coherent structures.

3.5 The late stages

As we have seen, prediction errors in chaotic systems tend to amplify less rapidly, on the average, as they become larger. Indeed, the slackening may become apparent long before the errors are close to saturation, and thus at a range when the predictions are still fairly good. For Eq. (3.1), and in fact for the average behaviour of some global atmospheric circulation models, we can construct a crude formula by assuming that the growth rate is proportional to the amount by which the error falls short of saturation. We obtain the equation

$$(1/E)dE/d\tau = \lambda_1(E^* - E)/E^*, \tag{3.4}$$

where E^* denotes the saturation value for E. Equation (3.4) possesses the solution

$$E = E^*(1 + \tanh(\lambda_1\tau))/2, \tag{3.5}$$

if the origin of τ is the range at which $E = E^*/2$. The well-known symmetry of the hyperbolic-tangent curve, when it is drawn with a linear vertical scale, then implies that the rate at which the error approaches saturation, as time advances, equals the that the onsets of coming phases may also possess some predictability. Again, the phenomenon should lead to an ebbing of the late-stage growth rate.

Perhaps less important but almost certainly more predictable than the ENSO-related features are the winds in the equatorial middle-level stratosphere, dominated by the quasi-biennial oscillation (QBO). Even though one cannot be certain just when the easterlies will change to westerlies, or vice versa, nor how the easterlies or westerlies will vary from day to day within a phase, one can make a forecast with a fairly low expected mean-square error, for a particular day, a year or even several years in advance, simply by subjectively extrapolating the cycle, and predicting the average conditions for the anticipated phase. Any measure of the total error that gives appreciable weighting to these winds is forced to approach saturation very slowly in the latest stages.

Looking at still longer ranges, we come to the question, 'Is climate predictable?' Whether or not it is possible to predict climate changes, aside from those that result from periodic or otherwise predictable external activity, may depend on what is considered to be a climate change.

Consider again, for example, the ENSO phenomenon. To some climatologists, the climate changes when El Niño sets in. It changes again, possibly to what it had previously been, when El Niño subsides. We have already suggested that climatic changes, so defined, possess some predictability.

To others, the climate is not something that changes whenever El Niño arrives or leaves. Instead, it is something that often remains unchanged for decades or longer, form. If instead of parametrising these features we omit them altogether, the models will still produce synoptic systems that behave rather reasonably, even though the actual forecasts will suffer from the omission. Evidently this is because the features that are small in scale are relatively small in amplitude, so that their influence acts much like small random forcing.

Moving to longer time scales, we find that some models yield rather good simulations of the behaviour of the ENSO phenomenon, even if not good forecasts of individual occurrences, without including the accompanying synoptic systems in any more than parametrised form. Here the synoptic systems do not qualify as being small in amplitude, but they appear to be rather weakly coupled to ENSO, so that again they may act like small random forcing.

Similarly, climatic fluctuations with periods of several decades or longer have more rapid oscillations superposed on them, ranging in timescale all the way from ENSO and the QBO to synoptic and small-scale features. Certainly these fluctuations

the climate itself, so that climate can constitute a dynamical system? If this is not the case, are these features nevertheless coupled so weakly to the climate that they act like small random forcing, so that climate still constitutes a dynamical system? Or do they act more like strong random forcing, so that climate does not qualify as a dynamical system, and prospects for its prediction are not promising? At present the reply to these questions seems to be that we do not know.

extrapolation. In the case of the atmosphere, the inevitable small-scale features will work like randomness.

I have confined my quantitative discussions to results deduced from pairs or ensembles of numerical solutions of mathematical models with various degrees of sophistication, but alternative approaches have also been exploited. Some studies have been based on equations whose variables are ensemble averages of error magnitudes. These equations have been derived from conventional atmospheric models, but, to close the equations, i.e., to limit the number of variables to the number of equations, it has been necessary to introduce auxiliary assumptions of questionable validity (see, for example, Thompson, 1957; Lorenz, 1969a). Results agree reasonably well with those yielded by more conventional approaches.

There have also been empirical studies. Mediocre analogues – pairs of somewhat similar states – have been identified in northern hemisphere weather data; their

(1982). Atmospheric predictability experiments with a large numerical model. *Tellus*, **34**, 505–13.

(1993). *The Essence of Chaos*. University of Washington Press.

Mintz, Y. (1965). Very long-term global integration of the primitive equations of atmospheric motion. In *WMO-IUGG Symposium on Research and Development Aspects of Long-range Forecasting*, WMO Technical Note 66, 141–67.

Palmer, T. N. (1988). Medium and extended range predictability and stability of the Pacific/North American mode. *Quart. J. Roy. Meteor. Soc.*, **114**, 691–713.

Simmons, A. J., R. Mureau and T. Petroliagis (1995). Error growth and estimates of predictability from the ECMWF forecasting system. *Quart. J. Roy. Meteor. Soc.*, **121**, 1739–1771.

Smagorinsky, J. (1965). Implications of dynamic modelling of the general circulation on long-range forecasting. In *WMO-IUGG Symposium on Research and Development Aspects of Long-range Forecasting*, WMO Technical Note 66, 131–7.

(1969). Problems and promises of deterministic extended range forecasting. *Bull. Am. Meteorol. Soc.*, **50**, 286–311.

Thompson, P. D. (1957). Uncertainty of initial state as a factor in the predictability of large-scale atmospheric flow patterns. *Tellus*, **9**, 275–95.

Toth, Z. and E. Kalnay (1993). Ensemble forecasting at NMC: the generation of perturbations. *Bull. Am. Meteorol. Soc.*, **74**, 2317–30.

Zebiak, S. E. and M. A. Cane (1987). A model El Niño-Southern Oscillation. *Mon. Weather Rev.*, **115**, 2262–78.