Tellus (2005), 57A, 1–11
Printed in UK. All rights reserved

TELLUS

A look at some details of the growth of initial uncertainties

By EDWARD N. LORENZ, *Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

(Manuscript received 21 October 2003; in final form 27 August 2004)

ABSTRACT

Because of the errors entailed in observing certain systems, the states that one might believe to be the true states form an ensemble, as do the states obtained from these states by forward extrapolation in time. We identify the uncertainty with the root-mean-square distance in state space of the ensemble members from their mean. We enumerate the properties of a special three-variable system that behaves chaotically, and we use the system to evaluate a logarithmic measure $\alpha\ (t_1, t_0)$ of the ratio of the uncertainty at a 'verifying time' t_1 to that at an 'observing time' t_0. With t_0 and t_1 as coordinates, we construct diagrams displaying contours of $\alpha\ (t_1, t_0)$. We find that the details of the diagrams tend to line up in the horizontal and vertical directions, rather than parallel to the diagonal where $t_1 = t_0$, as they would if $\alpha\ (t_1, t_0)$ depended mainly on the forecast range $t_1 - t_0$. The implication is that states at certain times t_1 are highly predictable, i.e. $\alpha\ (t_1, t_0) < \alpha\ (t, t_0)$ if t occurs somewhat before or after t_1, and that states at certain times t_0 are highly predictive, i.e. $\alpha\ (t_1, t_0) < \alpha\ (t_1, t)$ if t occurs somewhat before or after t_0. When observations at times preceding t_0 are combined with those at t_0, the greatest resulting reductions in uncertainty at t_1 occur when the states at the additional times are highly predictive. We speculate as to the applicability of these findings to larger systems.

1. Introduction

Speculations that the atmosphere and its surroundings form a chaotic dynamical system – one where future states depend sensitively on their antecedents – were abundant long before any reasonably convincing quantitative demonstrations could be offered. Nearly a century ago, in an essay concerning chance, Poincaré (1912) chose the atmosphere as one example of a system whose future states are in doubt. Much later, but still more than half a century ago, the implications of chaos for weather forecasting were aptly described by Eady (1951), who, after expressing the opinion that 'instability is a *normal* [his italics] feature of atmospheric motion,' stated the following:

> The practical significance of a demonstration that the motion is unstable is clear, for in practice, however good our network of observations may be, the initial state of motion is never given precisely and we never know what small perturbations may exist below a certain margin of error. Because the perturbation may grow at an exponential rate, the margin of error in the forecast (final) state will grow exponentially as the period of the forecast is increased, and this possible error is unavoidable whatever our method of forecasting. After a limited time interval, which, as we shall see, can be roughly estimated, the possible error will become so large as to make the forecast valueless. In other words, the set of all possible future developments consistent with our initial data is a divergent set and any direct computation will simply pick out, arbitrarily, one member of the set.

Eady was preparing his article just as the first computer-produced numerical integration to originate from an observed atmospheric state was being performed (Charney et al., 1950). It would be another five years before Phillips (1956) would construct the first GCM, and more than a decade before Leith, Mintz, and Smagorinsky would use GCMs to estimate the rate of divergence of the possible future developments (Charney et al., 1966). In the subsequent years, GCMs have evolved from general-circulation to global-circulation models, extending through the tropics, and from systems of a few hundred or a few thousand ordinary differential equations to some with several million. Here, bigger has meant better, if the quality of the forecasts produced is a valid measure. Systems that appear to be ever closer approximations to the real atmosphere continue to exhibit chaos, and they provide strong evidence favoring Eady's claim.

Meanwhile, with an eye toward a deeper understanding of certain properties of GCMs and more general chaotic systems, some investigators were making GCMs systematically smaller, most often by lowering the spatial resolution. Thus, with a 28-variable system we were able to examine changes in predictability that accompanied changes in the synoptic situation (Lorenz, 1965). More recently, we managed to reduce the system to a set of three ordinary differential equations (Lorenz, 1984, hereafter L84). While some might take exception to our description of the system as a GCM, there is no doubt as to its

chaotic behavior. The bulk of the present study is based on the behavior of this system, as revealed by numerical integrations.

Following standard practice in treating dynamical systems, we shall introduce, for any system, a state space or phase space – a multidimensional Euclidean space whose coordinates are the dependent variables of the system. Often we must first decide what quantities to use as dependent variables; in a global atmospheric model, for example, should we choose grid-point values or spherical-harmonic amplitudes? The choice having been made, an instantaneous state of the system becomes a point in state space, while a state varying through time becomes an orbit. A collection of states may become a sphere or some other object, which will be continually deformed as each point of the collection traverses its orbit. What might be called the climate of the system – the set of states that will be approached again and again by all or a large number of orbits, as opposed to those that will be permanently avoided – becomes the attractor, or the set of attractors.

Often we must distinguish between the existence of a state and our knowledge of the state. We can, when we wish, specify the state of a mathematically defined system precisely. A state of a real physical system such as the atmosphere, say the weather pattern 24 h ago, may also be a precisely defined entity, but at best we can know it to be one member of an infinite ensemble. As a measure of uncertainty in knowing the state, we shall choose the root-mean-square distance, in state space, of the points of the ensemble from their centroid, although other measures are possible and often preferable. Our concern will be with the growth or decay of uncertainty as each point in the ensemble advances along its orbit.

The scalar quantity most often associated with the growth of uncertainty is the leading Lyapunov exponent λ_1, which, for a chaotic system, is positive. The Lyapunov exponents $\lambda_1, \ldots, \lambda_K$, where K is the dimension of the state space, are long-term, actually infinite-term, properties of the system, and a number of different processes converge to them. The long-term average growth rate of the length of an infinitesimal line segment in phase space is λ_1, i.e. if $D(t)$ is the length acquired by the segment at time t, the limit as $t \to \infty$ of $(1/t) \log[D(t)/D(0)]$ is λ_1. If instead $D(t)$ is the area of an infinitesimal parallelogram, the above limit is $\lambda_1 + \lambda_2$. In general, two adjacent sides of the parallelogram will each grow as $\exp(\lambda_1 t)$, but the angle between them will decay as $\exp[-(\lambda_1 - \lambda_2)t]$. The sum $\lambda_1 + \cdots + \lambda_k$ is similarly related to the growth of the volume of a k-dimensional box. If $k = K$, the time-dependent vector to which all sides of the box ultimately become nearly parallel is the leading Lyapunov vector $\mathbf{L}_1$. The exponents may also be described in terms of long-term growth rates of the axes of an infinitesimal ellipsoid, but the former description leads to a more easily formulated computational algorithm, which we shall presently use to produce Fig. 1.

In applications, λ_1 is often little more than an order-of-magnitude estimate of the quantity in which we are interested. If, for example, we are dealing with a true state and an observed state that approximates it, either state may be viewed as the other plus a small perturbation, and, for a while at least, the growth rate of the uncertainty – the magnitude of the perturbation – will depend upon the form of the perturbation. It can also depend strongly upon the true state itself, and sometimes upon a regime of behavior in which the true state is occurring (see Palmer, 1988). We might try to remove this dependency by averaging

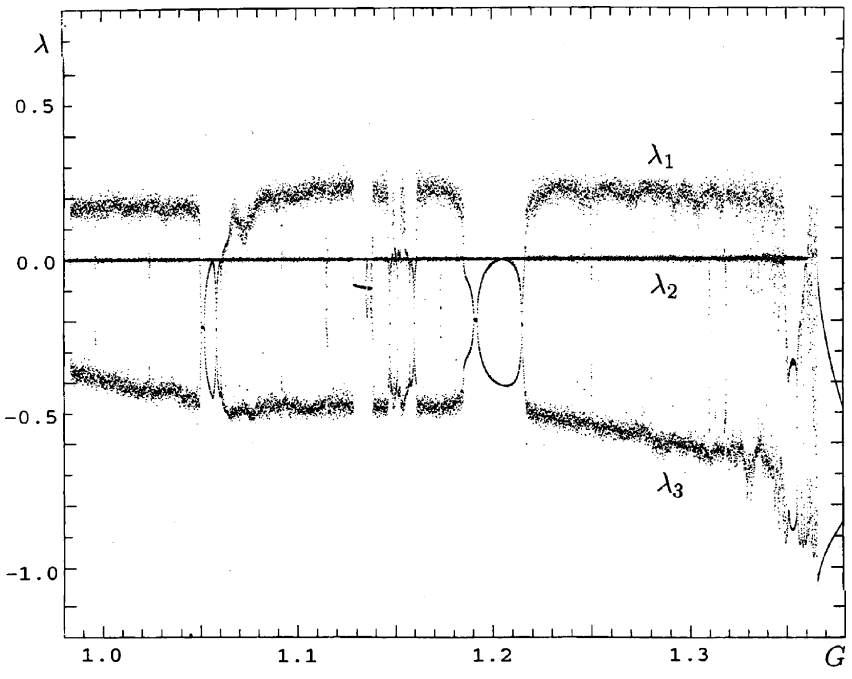

Fig 1. Variations of the Lyapunov exponents $\lambda_1, \lambda_2, \lambda_3$ of eqs. (1) (vertical scale) as G (horizontal scale) varies from 0.98 to 1.38. Points are plotted whenever G is a multiple of 0.00005. Each point is estimated from a 10-yr sample.

over an ensemble of perturbations, or an ensemble of true states, or both, but this can introduce new complications.

It has been pointed out on numerous occasions (e.g. Lacarra and Talagrand, 1988; Farrell, 1990; Trevisan, 1993) that the early growth rate of small uncertainties often exceeds a subsequent quasi-exponential rate; that this excess is systematic can become apparent if averaging is performed. Moreover, in most applications the initial uncertainties are not infinitesimal, and may not even qualify as small. Because the various ensemble members will ultimately be confined to the attractor, the uncertainty is limited by the diameter of the attractor, and the growth must slacken and eventually cease. 'Saturation' will then have occurred. We thus recognize three possible stages in the growth of uncertainties: an early stage, often with rapid growth; a mature stage, possibly with quasi-exponential growth; and a late stage, when saturation is imminent. With a realistically large initial uncertainty, it is quite possible that the slackening will commence just as the early rapid growth is ending, and the mature stage will be virtually absent.

The purpose of this paper is to examine in detail the manner in which step-by-step changes in the uncertainty at various stages of growth depend upon the succession of true states. We shall do this in the context of the three-variable model of L84, where the true state can visit most portions of the attractor during a relatively short time interval. We shall then speculate as to the applicability of the results to larger systems.

2. Documentation of the model

We feel that it is important, whenever speculations as to the behavior of a system are to be influenced by computations with a model, to have a thorough documentation of the model, which at the least should reveal the obvious ways in which the model resembles or fails to resemble the system. This would appear especially true when ostensibly the model differs considerably from the system, perhaps by being much simpler in structure. As already mentioned, the model in this case will be that introduced in L84 as the 'littlest GCM'.

Many of its interesting properties have by now been well documented (e.g. Lorenz, 1990; Trevisan, 1993; Aires and Rossow 2003), and here we shall simply note those that are relevant to the later sections of this work. The equations are

$$\mathrm{d}X/\mathrm{d}t = -Y^2 - Z^2 - aX + aF, \tag{1a}$$

$$\mathrm{d}Y/\mathrm{d}t = XY - bXZ - Y + G, \tag{1b}$$

$$\mathrm{d}Z/\mathrm{d}t = bXY + XZ - Z, \tag{1c}$$

where t is time. In their meteorological contexts, X represents the strength of a uniform westerly current (easterly when $X < 0$), identified geostrophically with a poleward temperature decrease, and driven by the forcing F, while Y and Z represent the components, in phase and in quadrature with the forcing G, of a large-scale superposed wave, assumed implicitly to tilt westward with increasing elevation. The linear terms represent thermal and mechanical damping, and the absence of coefficients for Y and Z indicates that the time unit is the damping time for the wave, assumed to equal 5 d. The quadratic terms containing b represent a simple translation of the waves by the zonal current, while the remaining quadratic terms describe exchanges of total energy, proportional to $X^2 + Y^2 + Z^2$, between the zonal flow and the waves. We shall frequently denote the point (X, Y, Z) in state space by the vector $\mathbf{X}$. In solving eqs. (1) numerically we shall invariably use the 'standard' fourth-order Runge–Kutta scheme with a time increment of 0.05 units, or 6 h.

An uncertainty at a time t_0 may be thought of as the magnitude of a perturbation $\mathbf{x}$ upon a basic state $\mathbf{X}$. If we let $\mathbf{W} = \mathbf{X} + \mathbf{x}$, we can determine $\mathbf{x}$ at a later time t_1 by numerically integrating eqs. (1) from the initial states $\mathbf{X}(t_0)$ and $\mathbf{W}(t_0)$, and then subtracting $\mathbf{X}(t_1)$ from $\mathbf{W}(t_1)$. Alternatively, if $\mathbf{x}$ is small – strictly speaking infinitesimal – we can derive from eqs. (1) the 'perturbation' equations

$$\mathrm{d}x/\mathrm{d}t = -ax - 2Yy - 2Zz, \tag{2a}$$

$$\mathrm{d}y/\mathrm{d}t = (Y - bZ)x + (X - 1)y - bXz, \tag{2b}$$

$$\mathrm{d}z/\mathrm{d}t = (bY + Z)x + bXy + (X - 1)z, \tag{2c}$$

for the components x, y, and z of $\mathbf{x}$. Because the coefficients in eqs. (2) depend upon $\mathbf{X}$, to integrate eqs. (2) we must integrate eqs. (1) and (2) as a system with six variables.

In either event, if $\mathbf{x}$ is small, the equations define $\mathbf{x}(t_1)$ as $\mathbf{A}(t_1, t_0)\mathbf{x}(t_0)$. Here, the vector $\mathbf{x}$ is treated as a matrix with one column, and $\mathbf{A}$ is a square matrix depending upon the values of $\mathbf{X}$ between t_0 and t_1 but independent of $\mathbf{x}$. The matrix $\mathbf{A}$ is easily determined, because each column is the vector $\mathbf{x}(t_1)$ obtained when $\mathbf{x}(t_0)$ is a unit vector directed along a coordinate axis.

We shall confine our attention to the well-studied values $a = 0.25$, $b = 4.0$, and $F = 8.0$, and all further references to eqs. (1) or (2) will imply that these values have been incorporated. With these values, we found in L84 that chaos occurred with $G = 1.0$, but not with the smaller values of G examined.

Figure 1 shows estimates of the three Lyapunov exponents, for values of G from 0.98 to 1.38, at intervals of 0.00005. For each value of G we have computed the exponents by making a basic run and three perturbed runs, which reveal the behavior of the three sides of an infinitesimal box, each run extending for 10 yr. We have then used the final state in the basic run as the initial state in the basic run for the succeeding value of G. Because the values of G are so closely spaced, the plotted values of each exponent appear to be organized into bands, with several breaks in the upper and lower bands. The finite vertical thickness of the bands reflects the sampling error; for a short range of values of G we produced segments of bands about half as wide by extending the runs to 40 yr instead of 10.

Positive values of λ_1, implying chaos, occupy much of the range of G, but a number of periodic windows – continua where chaos is absent – are evident, the most conspicuous extending from 1.186 to 1.217. A few additional computations with higher resolution have revealed still more windows, and the true number is presumably infinite. Except when all solutions approach a fixed point, which happens when G exceeds 1.367, one exponent must equal zero. This corresponds to a perturbation directed along the orbit, where the perturbed solution must continue to equal the basic one displaced by a fixed time, and long-term growth or decay cannot occur. The prominent zero line was not drawn as a reference line; it was formed by the computed values of λ_2 in the chaotic regions and λ_1 in the windows. The frequently occurring values of λ_1 between 0.22 and 0.23 imply a doubling time of about 15 d for small errors – very slow by real atmospheric standards. The slow growth rate may result from the very coarse horizontal resolution of the model, but may also occur simply because the model is not the real atmosphere.

On the basis of Fig. 1 we have chosen the value $G = 1.23$ for the chaotic system to be intensively studied; hereafter G will always equal 1.23, unless it is otherwise stated. Figure 2 shows the intersection of the attractor of the system with the plane $X = 1.0$, as approximated by 100 000 consecutive intersections of a single orbit with the plane. Its structure, with an infinite number of quasi-parallel curves separated by sometimes narrow and sometimes wide gaps, is typical of low-order chaos. Intersections with other planes show similar structure. Their finite diameters place a limit upon the ultimate size of uncertainties.

Figure 3 shows a time series for X for a two-year period, displayed as two 1-yr segments. To construct the figure we first designated a time as 'day −300', and let the 'true' state at this time be given by $X = 1.0$, $Y = 0.0$, and $Z = -0.75$, a point on or nearly on the attractor, as indicated by Fig. 2. We then integrated forward for 300 d to 'day 0'. The series shown begins at day 0.

Like the attractor, the series is typical of low-order chaos; there are no exact repetitions, but certain distinctive patterns occur frequently, while other conceivable ones are never found. Strong westerlies lasting a week or so seem to appear about once a month, while easterlies appear less often. The three months following day 156 nearly repeat those following day 34, while the four months after day 595 repeat equally closely those after day 433, but no extended portion of the second year superposes well on any part of the first, and, without additional output, one might wonder whether the two segments were produced by the same equations. The principal fluctuations of Y and Z, not shown, are due largely to translation of the waves, and tend to exhibit higher frequencies than those of X.

Figure 4 presents four estimates of the average behavior of initially small uncertainties. The computation procedure involves an ensemble of runs. For the first run only, not counted in the

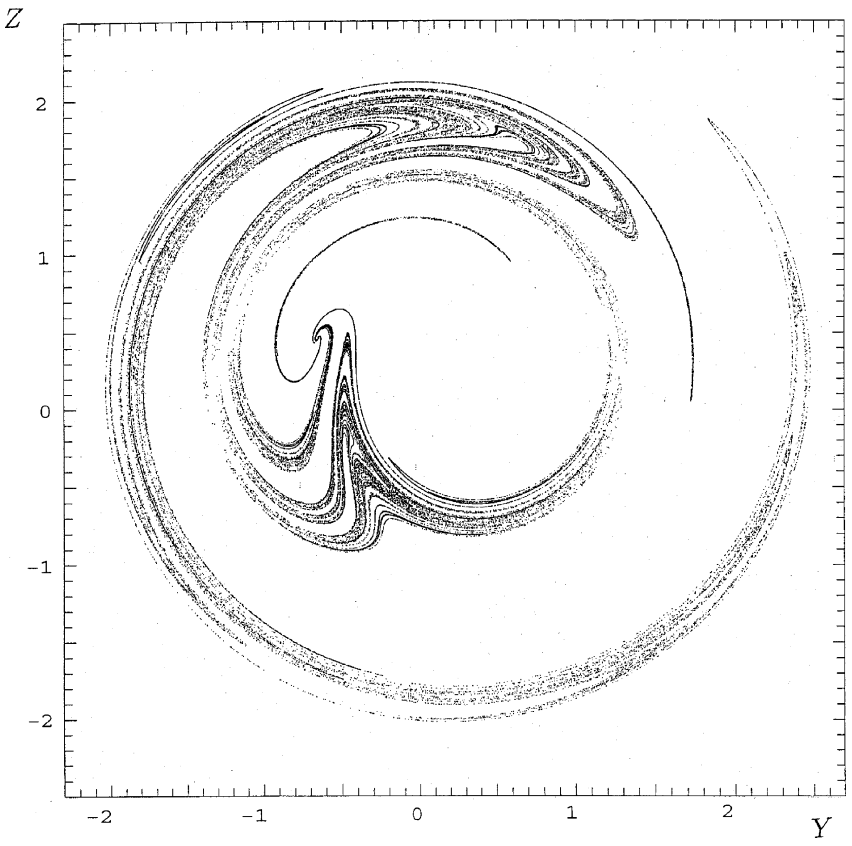

Fig 2. Intersection of the attractor of eqs. (1) when $G = 1.23$ with the plane $X = 1.0$. Coordinates are Y (horizontal) and Z (vertical).

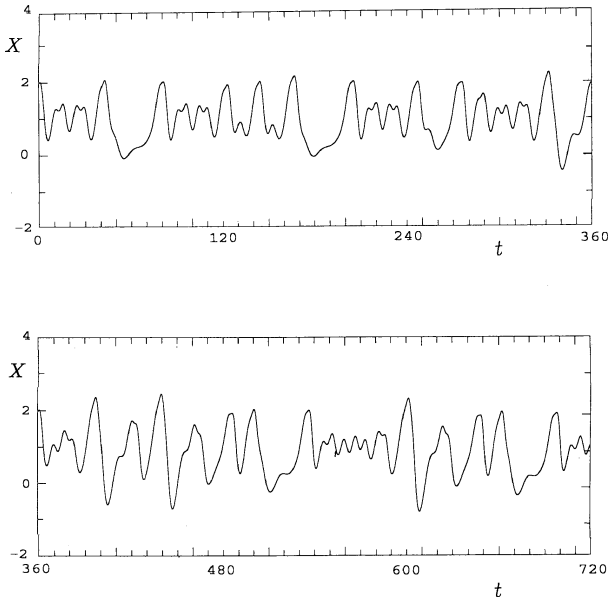

Fig 3. A 2-yr time series of X (vertical scales) satisfying eqs. (1) with $G = 1.23$, displayed as two 1-yr segments. Horizontal scales are times in days.

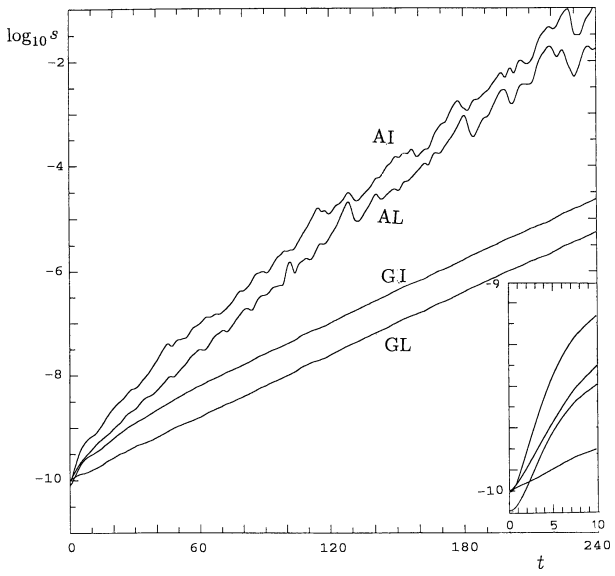

Fig 4. Estimates of average amplification of initially small uncertainties when two types of perturbation are superposed on each of 1000 basic states, in a system governed by solutions of eqs. (1) with $G = 1.23$. The horizontal scale is days following initial state, and the vertical scale is base-10 logarithm of uncertainty. 'G' denotes geometric average, 'A' denotes arithmetic average, 'L' indicates that initial ensemble members are parallel to the leading Lyapunov exponent, 'I' indicates that the initial ensemble is isotropic. The inset at the lower right is a four-times enlargement of the first 10 d.

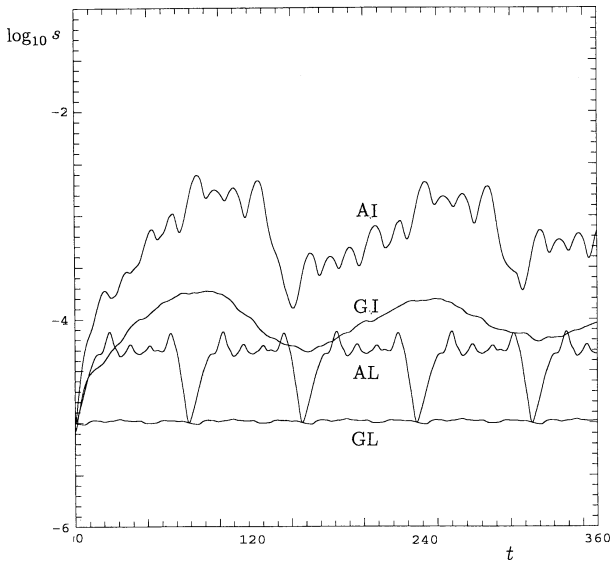

Fig 5. The same as Fig. 4, but with $G = 1.21$. Note the stretched vertical scale.

ensemble, we choose a basic initial state $\mathbf{X}(t_0)$ and two arbitrary unit perturbation vectors $\mathbf{x}_1(t_0)$ and $\mathbf{x}_2(t_0)$, which we call a 'Lyapunov perturbation' and an 'isotropic perturbation'. We let $\mathbf{X}_i(t_0) = \mathbf{X}(t_0) + \varepsilon\mathbf{x}_i(t_0)$ $(i = 1, 2)$, where ε is a chosen small quantity. We then integrate each state until time t_1, subtracting $\mathbf{X}(t)$ from $\mathbf{X}_i(t)$ at each time t to obtain $\varepsilon\mathbf{x}_i(t)$ and its magnitude $\varepsilon v_i(t)$. By time t_1, $\mathbf{x}_1$ is presumably nearly parallel to $\mathbf{L}_1$, because its component parallel to $\mathbf{L}_1$ grows while its other components do not. For initial values of $\mathbf{X}$ and $\mathbf{x}_1$ in each remaining run, we choose the final values of $\mathbf{X}$ and $\mathbf{x}_1/v_1$ in the preceding run, while choosing each component of the new $\mathbf{x}_2$ randomly from an isotropic Gaussian distribution, and then we repeat the procedure. Finally, the uncertainties εv_1 and εv_2 are averaged geometrically at each time over the ensemble of runs to produce the curves labeled GL (Lyapunov) and GI (isotropic), while the curves labeled AL and AI are root-mean-square values, where the averaging is arithmetic.

For Fig. 4, $\log_{10}\varepsilon = -10$, the ensemble size is 1000, and $t_1 - t_0 = 8$ months. The inset at the lower right magnifies the first 10 d. Only in curve GL is the growth rate constant, and in agreement with λ_1. In curve GI, the same growth rate becomes established, but only after a rapid early growth that has increased the uncertainty fourfold; there is no subsequent compensation. Arithmetic averages are necessarily larger than geometric averages when the quantities being averaged are not all equal, but it is apparent that they continue to grow more rapidly.

It is interesting to compare Fig. 4 with Fig. 5, which is constructed in the same manner, but with $G = 1.21$, when the system is not chaotic. The new curve GL is (nearly) horizontal, as expected, but the antics of the other curves might not have been anticipated. All of them show rapid early growth, which is thus clearly revealed as not necessarily indicating chaos. The curve AL has the 78-d periodicity of the system, and touches GL once every period, when the quantities being averaged are all the same.

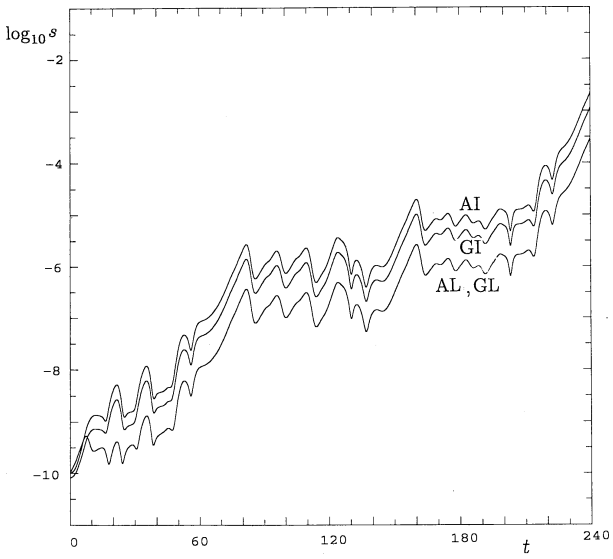

Fig 6. As in Fig. 4, with $G = 1.23$, but with 1000 perturbations of either type superposed on a single basic state.

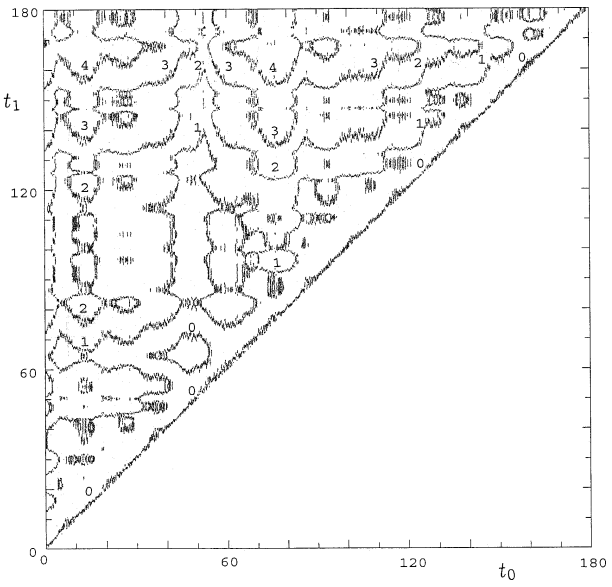

Fig 7. Base-10 logarithmic measure $\alpha\ (t_1, t_0)$ of ratio of uncertainty at time t_1 (vertical scale, in d) to uncertainty at time t_0 (horizontal scale, in d), when size-20 isotropic ensembles with uncertainty $\varepsilon = 0.00001$ are introduced at time t_0. Values are computed at 6-h intervals of t_0 and t_1. The number indicating the value of α on the contour has been placed on the high side of the contour.

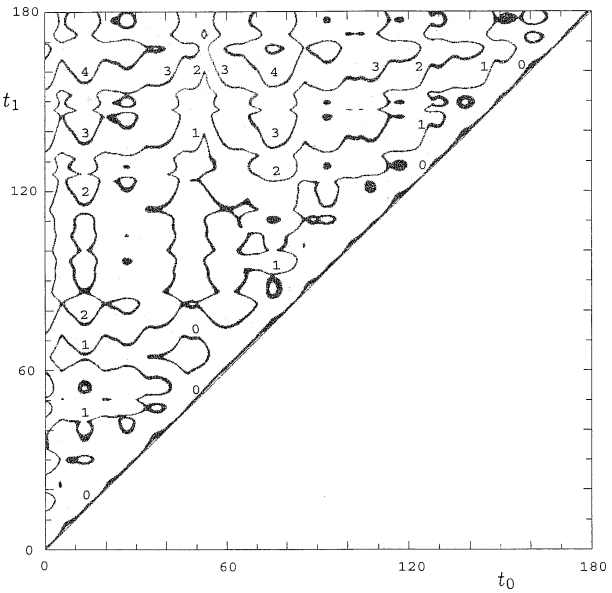

Fig 8. The same as Fig. 7, but with the computing procedure using eq. (6) (see text).

The curves GI and AI appear to have a doubled period, but, well beyond the range of the figure, GI flattens out, nearly an order of magnitude above GL, while AI acquires the 78-d periodicity.

The arithmetic averaging in Fig. 4 reduces the long-term doubling time to about 9 d – still rather slow – but two doublings occur in the first 5 d. Historically, this rapid early growth, not recognized then as being confined to a few days, was a factor in our decision to accept the model as a GCM.

Recall that in constructing Fig. 4 we superposed one Lyapunov and one isotropic perturbation on each of 1000 basic states. By contrast, in Fig. 6 we have superposed 1000 perturbations of either kind on the same basic state. After the first few weeks, when the mature stage seems to have set in, the curves closely parallel each other, suggesting that the spread in Fig. 4 resulted from the large number of basic states rather than perturbations. The curves AL and GL coincide, because the quantities being averaged, derived from the vector $\mathbf{L}_1$, are the same for the different ensemble members.

The total growth of uncertainty is comparable to that in Fig. 4, but, during the early weeks of the mature stage, the growth is more rapid, with doublings requiring about 4 d. After day 80, however, net growth ceases for two months, and then, after a brief resumption, ceases again. For the final month, the 4-d doubling reappears. The uncertainty also undergoes fluctuations with periods comparable to a week, often by a factor of 2 or 3 but sometimes by nearly an order of magnitude.

Other choices of the basic state produce different curves, some without the month-long interruptions of growth, but all with the shorter-period fluctuations. It is features that are characteristic of Fig. 6 but not present in Fig. 4 that will occupy the bulk of our subsequent attention.

3. Predictability and predictivity

Figures 7 and 8 show two triangular arrays that are much alike; the main difference is that Fig. 7 has a decidedly fuzzier appearance. In each figure the coordinates are times t_0 and t_1, measured in days, each extending from a time T_0 (day 0) to T_1 (day 180). The curves, or fuzzy approximations to them, are contours, at

unit intervals, of the base-10 logarithm $\alpha(t_1, t_0)$ of the ratio of the uncertainty at t_1 to that at t_0, when the uncertainty, presently to be more precisely defined, results from imperfect observations at t_0. In a meteorological context we are examining, for each time t_0, the quality of forecasts for each later time t_1, when the forecasts are based only on observations at t_0, and we should note that the uncertainty at t_1 results only from that at t_0, i.e. we are assuming 'perfect model' forecasts.

The number (0–4) adjacent to a contour has always been placed on the high side of the contour; thus, for example, the number '2' is located where α slightly exceeds 2.0, meaning that the uncertainty has amplified more than 100 times since t_0. The '2' contour was actually formed by placing a dot at every point (t_0, t_1) where α lies between 2.0 and 2.1, at 6-h intervals of t_0 and t_1; the other contours were produced similarly. Thus, slow variations of α may be recognized by wide contours, and very-small-scale structure by fuzziness.

To define α, we assume that a particular system governed by eqs. (1) possesses, at any 'observing time' t_0, a true state $\mathbf{U}(t_0)$, and that observing $\mathbf{U}$ entails an error of expected magnitude ε, as a consequence of which an observer will observe the state to be $\mathbf{V}(t_0)$. We assume that the error is systematic rather than random, so that different observers using the available observing procedure will make essentially the same error; otherwise the error could be considerably reduced by averaging the observations.

When we represent states by points in a three-dimensional state space, the different states $\mathbf{W}_1$, $\mathbf{W}_2$, ... that the observer, knowing that $\mathbf{V}$ involves an expected error, might believe to be the true state $\mathbf{U}$ will form an ensemble $M(t_0)$, and their expected distance from $\mathbf{V}$ will be ε. For definiteness we shall assume, not too realistically, that the members of M are distributed uniformly through a solid sphere $S(t_0)$, i.e. a sphere plus its interior, centered at $\mathbf{V}$, which must have radius $\varepsilon\sqrt{5/3}$ to make the root-mean-square distance from $\mathbf{V}$ equal to ε, while $\mathbf{V}$ is a single point in a solid sphere of similar radius centered at $\mathbf{U}$. The situation at time t_0 is illustrated schematically by the left portion of Fig. 9a, which shows $\mathbf{U}$, $\mathbf{V}$, and three sample points $\mathbf{W}$, all contained in S.

As time advances to an arbitrary 'verifying time' t_1, the points in S progress along their orbits. The situation is shown in the right portion of Fig. 9a. The shape of S (t_1) will be ellipsoidal if the uncertainty is small, but may become snake-like when it becomes large.

Although $\mathbf{V}(t_0)$ is a single point determined by $\mathbf{U}(t_0)$ and the observing procedure, we know in advance of the observation only the expected distance ε of $\mathbf{V}$ from $\mathbf{U}$. Hence, to our knowledge, $\mathbf{V}$ is simply one member of an ensemble, which is like M (t_0) except that it is centered at $\mathbf{U}$, while the ensemble M is one member of an ensemble $M^*(t_0)$ of ensembles, and S is one member of an ensemble of solid spheres, all contained in a solid sphere $S^*(t_0)$ having twice the radius of S, and centered at $\mathbf{U}$. As time advances, $M^*(t_0)$ and $S^*(t_0)$ are carried into $M^*(t_1)$ and $S^*(t_1)$. A complete schematic diagram would entail an infinite ensemble of diagrams

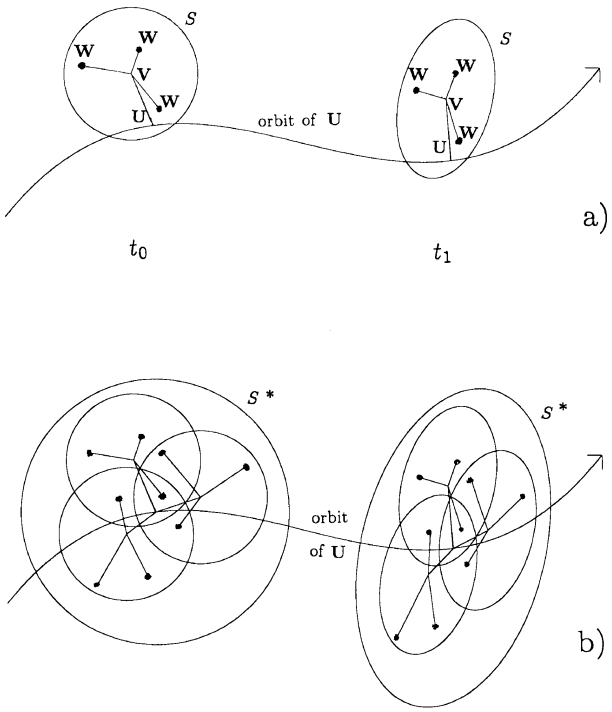

Fig 9. (a) Schematic diagram containing the true state $\mathbf{U}$, the observed state $\mathbf{V}$, shown connected to $\mathbf{U}$, the solid sphere or distorted solid sphere S centered at $\mathbf{V}$, and three sample states $\mathbf{W}$ that the observer might believe to be the true state $\mathbf{U}$, shown connected to $\mathbf{V}$, and contained in S, at time t_0 and at later time t_1. (b) Superposition of Fig. 9a (without labels on $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{W}$) and two similar diagrams, each with same state $\mathbf{U}$ but different conceivable observed states $\mathbf{V}$. The solid sphere S^* contains states $\mathbf{U}$, $\mathbf{V}$, $\mathbf{W}$ and solid spheres S.

like Fig. 9a; in Fig. 9b we offer an ensemble of size 3, and include $S^*(t_0)$ and $S^*(t_1)$.

We define the uncertainty $s(t_1)$, which depends also upon t_0 and ε, as the root-mean-square distance of all points $\mathbf{W}_1$, $\mathbf{W}_2$, ... in $M^*(t_1)$, i.e. in all ensembles $M(t_1)$, from their combined centroid $\{\mathbf{W}\}$. Here the braces denote an average over M^*. Thus, if $\mathbf{x}$ denotes $\mathbf{W} - \{\mathbf{W}\}$,

$$s^2(t_1) = \mathrm{tr}\left\{\mathbf{x}(t_1)\mathbf{x}^{\mathrm{T}}(t_1)\right\}. \tag{3}$$

If the uncertainty is small, $\{\mathbf{W}(t_1)\}$ is close to $\mathbf{U}(t_1)$ and, in any case, $\{\mathbf{W}(t_0)\} = \mathbf{U}(t_0)$, while s^2 $(t_0) = 2\varepsilon^2$. Finally we let

$$\alpha(t_1, t_0) = \log_B[s(t_1)/s(t_0)]. \tag{4}$$

For Figs. 7 and 8, the base B is 10.

To produce Figs. 7 and 8, we note that, given $\mathbf{U}$, a typical member $\mathbf{W}$ of M^* may be obtained by randomly selecting a point $\mathbf{V}$ from a distribution that is uniform in a solid sphere of radius $\varepsilon\sqrt{5/3}$, centered at $\mathbf{U}$, and then randomly selecting $\mathbf{W}$ from a similar distribution, centered at $\mathbf{V}$. We first use eqs. (1) to generate a sequence of true states $\mathbf{U}(t)$, at 6-h intervals, beginning with the state at day 0 used in constructing Fig. 3. Then, separately

for each choice of observing time t_0 from T_0 to T_1, we select K points $\mathbf{V}$ and their accompanying points $\mathbf{W}$, in the manner indicated above, to produce an ensemble of size K, intended to approximate $M^*(t_0)$. Again using eqs. (1) we advance each $\mathbf{W}$ forward to each verifying time t_1 from t_0 to T_1 to obtain $M^*(t_1)$, and finally we evaluate s and then α from eqs. (3) and (4).

For Fig. 7, $\varepsilon = 0.00001$, and the ensemble has only 20 members – a far cry from infinity. Figure 8 is identical to what we would have obtained with a very large value of K, although the procedure used to obtain it, to be described presently, is different. We see that even the small ensemble of Fig. 7 captures most of the visible details of Fig. 8.

Large ensembles can consume considerable computation time, and the time may be greatly reduced if we assume that ε is small, so that eqs. (2) apply and $\mathbf{x}_1 = \mathbf{A}_{10}\,\mathbf{x}_0$. Here $\mathbf{x}_0$, $\mathbf{x}_1$, and $\mathbf{A}_{10}$ denote $\mathbf{x}(t_0)$, $\mathbf{x}(t_1)$, and $\mathbf{A}(t_1, t_0)$, respectively. It follows from eq. (3) that

$$s^2(t_1) = \mathrm{tr}\left(\mathbf{A}_{10}\left\{\mathbf{x}_0\mathbf{x}_0^{\mathrm{T}}\right\}\mathbf{A}_{10}^{\mathrm{T}}\right). \tag{5}$$

Because $\mathbf{x}_0$ has been chosen randomly and isotropically, $\{\mathbf{x}_0\mathbf{x}_0^{\mathrm{T}}\} = \sigma^2\mathbf{I}$, where $\sigma^2 = 2\varepsilon^2/3$ is the variance of an individual component of $\mathbf{x}_0$. Hence, $s^2(t_1) = \sigma^2\mathrm{tr}(\mathbf{A}_{10}\mathbf{A}_{10}^{\mathrm{T}})$, and, because tr $(\mathbf{I}) = 3$,

$$\alpha(t_1, t_0) = \frac{1}{2}\log_B\left[\mathrm{tr}\left(\mathbf{A}_{10}\mathbf{A}_{10}^{\mathrm{T}}\right)/3\right]. \tag{6}$$

Equation (6) was used to produce Fig. 8, with its absence of fuzziness.

Although Figs. 7 and 8 readily reveal where α is small or large, our primary interest is in rather small-scale features, which are not too well resolved with the present contour spacing. Figure 10 is constructed like Fig. 8, except that the logarithmic base B is 2. We have not attempted to number the now closely spaced contours; approximate values are readily found by comparing Fig. 8.

If $\alpha(t_1, t_0)$ depended only upon the forecast range $t_1 - t_0$, the contours would be parallel to the main diagonal. Although there is some tendency for the contours nearest the diagonal to follow it, the most prominent property of Fig. 10 is the tendency for the smaller features, and especially the closed centers, to line up in two directions: the vertical and the horizontal. The tendency for $\alpha(t_1, t_0)$ to be abnormally high throughout a column, notably near days 13 and 76, but also near days 27 and 94, and low, especially near day 48 but also near days 6 and 86, illustrates the well-known dependence of the subsequent growth on the state when the uncertainty is introduced. Likewise, the equally evident tendency for α to be high in certain rows, notably near days 55, 83, and 168, but also near days 122, 145, and others, and low, especially near day 65 but also near days 50, 114, and others, illustrates the possibly less familiar dependence of the growth rate on the state when the uncertainty is occurring. In a meteorological context, we can describe the horizontal lining up by saying that some states, such as those near day 65, tend to be more predictable than others. Likewise, the vertical lining up implies that some states, like those near day 48, are more predictive than others; they are better predictors. The most predictable and also the least predictable states are in general neither the most nor the least predictive.

Figure 11 is constructed like Fig. 7, with $B = 10$, but with the initial uncertainty no longer small; $\varepsilon = 0.05$, about 5 percent of the somewhat indefinite saturation value. The ensemble size is 100. The '1' contour shares many details with the one in Fig. 7, but farther from the diagonal the final stage of growth has been entered and saturation has been reached, and '2', '3', and '4' contours no longer appear.

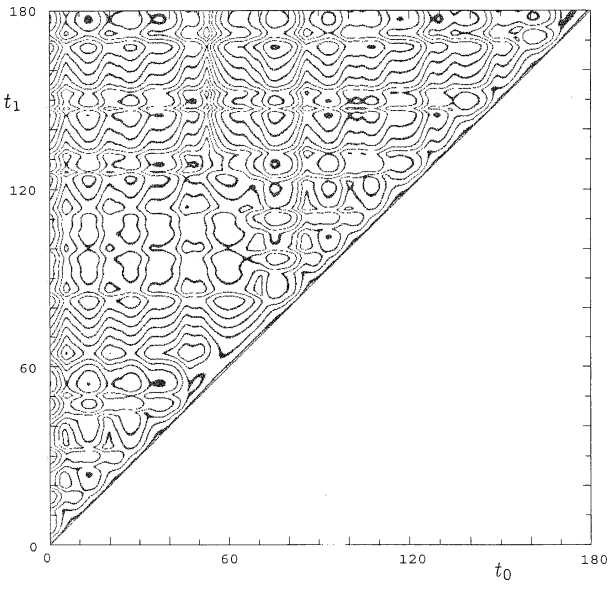

Fig 10. The same as Fig. 8, but for a base-2 logarithmic measure.

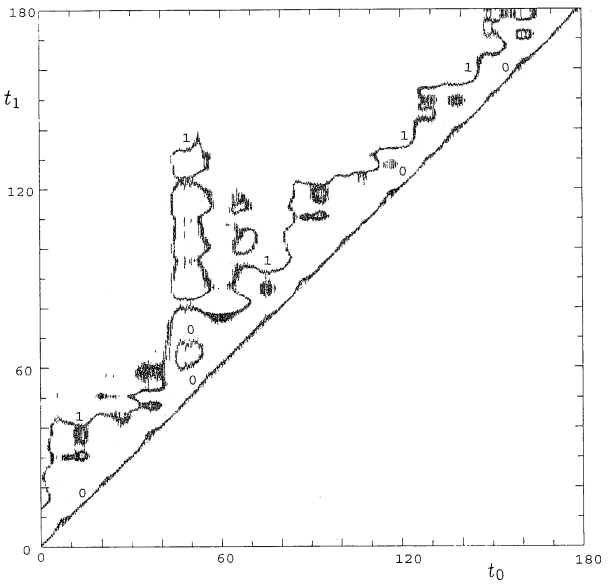

Fig 11. The same as Fig. 7, but with $\varepsilon = 0.05$, and with size-100 ensembles.

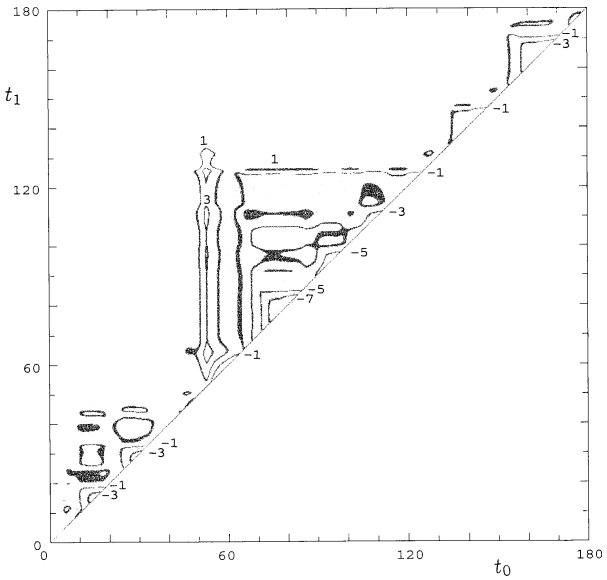

Fig 12. The same as Fig. 8, but for $\beta(t_1, t_0)$ (see text for definition), and for odd-integer values of base-2 logarithm.

A close look at the columns and rows in Fig. 10 suggests that, away from the diagonal, and so presumably in the mature stage of growth, $\alpha(t_1, t_0)$ may be closely approximated by the sum of a function of t_0 and a function of t_1; the two functions bear no obvious relation. If this is the case, subtracting one column, or one row, of $\alpha(t_1, t_0)$ from the field of $\alpha(t_1, t_0)$ should produce contours that are mainly vertically, or horizontally, oriented, while subtracting one column and one row should produce a field that is nearly uniform, away from the diagonal. We therefore let

$$\beta(t_1, t_0) = \alpha(t_1, t_0) - \alpha(t_1, T_0) - \alpha(T_1, t_0) + \alpha(T_1, T_0) \tag{7}$$

(we could use times other than T_0 and T_1), and in Fig. 12 we show odd-numbered contours of $\beta(t_1, t_0)$, with $B = 2$, for the same ranges of t_0 and t_1 as in Fig. 8. The use of odd contours serves to place the upper-left corner, where $\beta = 0$, midway between contours. As expected, there is a large contiguous area free of contours, where $-1 < \beta < 1$, but, perhaps less expected, the contours nearer the diagonal, where β is not nearly 0, have a decidedly step-like arrangement. On the diagonal, β appears to be always negative, although often small.

A plausible assumption is that the mature stage sets in when most perturbations in the ensemble have become nearly parallel to the leading Lyapunov vector $\mathbf{L}_1$. In a region of a diagram like Fig. 8, if this has happened, the increases or decreases in $\alpha(t_1, t_0)$ as t_1 moves up a column to a time t will then be independent of the column, i.e. independent of t_0, and

$$\alpha(t, t_0) - \alpha(t_1, t_0) = \Phi(t, t_1) \tag{8}$$

for some function Φ. If this region extends to the top of the diagram, where $t = T_1$, the first and third terms on the right of eq. (7) will sum to $-\Phi(T_1, t_1)$, while, if it also extends to the left edge, where $t_0 = T_0$, the second and fourth terms will sum to $\Phi(T_1, t_1)$, and $\beta(t_1, t_0)$ will vanish.

Our assumption therefore explains the large contiguous area free of contours in Fig. 12. An explanation for the step-like structure is more elusive.

4. Impact of past observations

Throughout our discussion we have been implying that if there is no basis for saying which members of a hypothetical set of nearly equal observed states are more like the true present state, there can be no basis for saying which of a set of perhaps widely differing predicted states will resemble a true future state. Often, however, there is a basis for saying which observed present states are more like the true state. This basis is the availability of past observations.

Consider again, but with new subscripts attached to some of the symbols, the scenario that led to the construction of Fig. 9. At time t_{01} an observer will observe the true state $\mathbf{U}(t_{01})$ to be $\mathbf{V}_1(t_{01})$, and will know that $\mathbf{U}$ is a point in a solid sphere $S_1(t_{01})$ centered at $\mathbf{V}_1$, but will not know which point, and hence at a later time t_{02} will know simply that $\mathbf{U}$ is a point in the distorted solid sphere $S_1(t_{02})$, shown as an ellipse in Fig. 9a. Assume now that t_{02} is an additional observing time rather than a verifying time. The observer will observe $\mathbf{U}(t_{02})$ to be $\mathbf{V}_2(t_{02})$, and will know that $\mathbf{U}$ is a point in a solid sphere $S_2(t_{02})$ centered at $\mathbf{V}_2$, again without knowing which point. Hence, $\mathbf{U}$ will be assumed to lie in the intersection of S_1 and S_2. This intersection is not empty, because it contains $\mathbf{U}(t_{02})$, but it may be much smaller than S_1 or S_2. If there are still more observing times $t_{03}, \ldots, t_{0N}$, the true state $\mathbf{U}(t_{0N})$ will be known to lie within the possibly very small intersection of $S_1(t_{0N}), \ldots, S_N(t_{0N})$, and the uncertainty may be very small. Figure 13, constructed for $N = 3$, and including a single verifying time t_1, illustrates the situation.

As in defining s and α, we treat the sequence of observations $\mathbf{V}_1, \ldots, \mathbf{V}_N$ as one member of an ensemble of conceivable sequences. The sequence of solid spheres $S_1, \ldots, S_N$ that are centered at $\mathbf{V}_1, \ldots, \mathbf{V}_N$ then becomes one member of an ensemble of sequences, and the intersection of $S_1(t_{0N}), \ldots, S_N(t_{0N})$ becomes one member of an ensemble $M^*(t_{0N})$ of conceivable intersections. Analogously to eq. (3), we define the uncertainty $s_N(t_1)$, for times t_1 following t_{0N}, as the root-mean-square distance, from their combined centroid, of all the points $\mathbf{W}$ in all the intersections comprising $M^*(t_1)$.

Just as we evaluated s by choosing a point $\mathbf{W}$ randomly from each member of an ensemble of solid spheres, so effectively we evaluate s_N by choosing a point $\mathbf{W}$ randomly from each member of an ensemble of intersections of distorted solid spheres. Our procedure for accomplishing this consists of randomly selecting points $\mathbf{V}_1(t_{01}), \ldots, \mathbf{V}_N(t_{0N})$, after having established a sequence of true states $\mathbf{U}(t)$. We then randomly select a point $\mathbf{W}(t_{01})$ from the solid sphere $S_1(t_{01})$ centered at $\mathbf{V}_1(t_{01})$ (see Fig. 13), and integrate eqs. (1) to obtain $\mathbf{W}(t_{02})$, which will lie in $S_1(t_{02})$. If

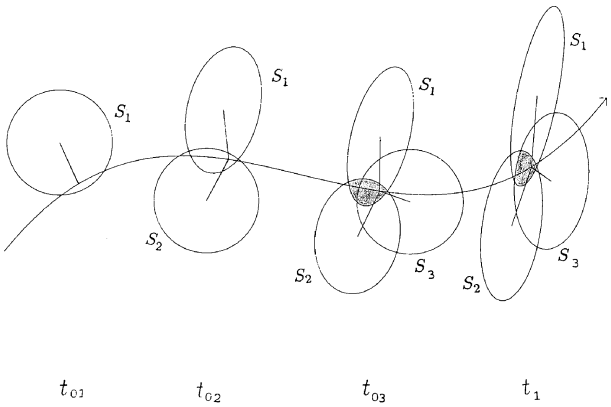

Fig 13. Schematic diagram like Fig. 9a, with true states **U** (not labeled) and observed states $\mathbf{V}_1$, $\mathbf{V}_2$, $\mathbf{V}_3$ determining solids spheres S_1, S_2, S_3 at observing times t_{01}, t_{02}, t_{03}, respectively, and states and distorted solid spheres into which these are carried at times t_{02}, t_{03}, and verifying time t_1. Shaded areas are regions common to S_1 and S_2 at t_{02} and common to S_1, S_2, S_3 at t_{03} and t_1.

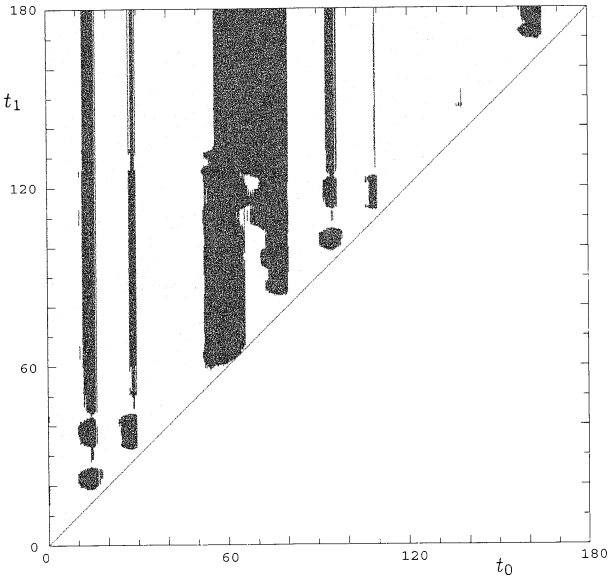

Fig 14. Ratio of $S(t_1)$, the uncertainty at verifying time t_1, when only observing time is t_0, to $S_4(t_1)$, the uncertainty at t_1 when observing times occur 15, 10, and 5 d before t_0 in addition to t_0. The shaded area indicates where $S(t_1)/S_4(t_1)$ exceeds 8.0. Computations were performed with $\varepsilon = 0.00001$ and ensembles of size 100.

$\mathbf{W}(t_{02})$ does not also lie in $S_2(t_{02})$ (and hence in the shaded area at t_{02} in Fig. 13), we discard **W** and select another point $\mathbf{W}(t_{01})$, repeating until we have made a selection that makes $\mathbf{W}(t_{02})$ lie in $S_2(t_{02})$. We then integrate eqs. (1) again to obtain $\mathbf{W}(t_{03})$, and again discard **W** if $\mathbf{W}(t_{03})$ does not lie in $S_3(t_{03})$ (and hence in the shaded area at t_{03}). When we have finally produced a point $\mathbf{W}(t_{0N})$ that lies in each of $S_1(t_{0N}), \ldots, S_N(t_{0N})$, it will be the first member of $M^*(t_{0N})$. We repeat, with different choices of $\mathbf{V}_1, \ldots, \mathbf{V}_N$, until we have amassed a suitably large ensemble of size K, from which we can compute s_N after advancing to t_1.

Our procedure is a variant of the 'shadowing' process, used to find exact solutions of a system that agree as closely as possible with a finite sequence of inexact values, such as might result from observations (Farmer and Sidorowich, 1991). Hansen and Smith (2001) have aptly illustrated the shadowing process by applying it to a sequence of 'observations' of the two-dimensional (Ikeda, 1979) system.

It is evident that in considering how past observations might affect present or future uncertainty, the number of combinations of past times at which observations might be introduced is enormous. To keep the paper reasonably concise, we shall treat a single case, where $N = 4$, and where the times t_{01}, t_{02}, and t_{03} precede the most recent observing time t_{04} by 15, 10, and 5 d, respectively. Instead of constructing a diagram of the ratio $s_N(t_1)/s_N(t_{04})$, analogous to Fig. 7, we shall evaluate the ratio $s(t_1)/s_N(t_1)$, i.e. the factor by which the uncertainty is reduced when past observations are added. We let $\varepsilon = 0.00001$ and $K = 100$, and, in Fig. 14, rather than showing a complete distribution, we simply place dots at those values of $t_0 (= t_{04})$ and t_1 where the reduction factor exceeds 8.0. Obtaining the 72 100 points **W** used in the evaluations required 32 740 917 tries.

We see first that the locations of greatest reduction tend to be organized into vertical bands. Comparison with Figs. 7, 8, or 10 shows that the large reduction factors occur when the state 5, 10, or 15 d before t_0 is more predictive than the state at t_0. The widest band includes the days that follow the highly predictive states near day 48 by 5–15 d, and also the days near day 76, which have very low predictivity.

Secondly, we see that, except for about 3 d following day 60, the shaded areas avoid the main diagonal where $t_1 = t_0$. Meteorologically, the incorporation of past observations has improved the forecasts more than the analyses, and the implication is that the greatest improvements in forecasting may come from producing analyses that are not simply more accurate per se, but are also more suitable for forward extrapolation.

Further calculations have indicated that it is difficult to make general statements about the number of past observations that yield significant reductions of uncertainty at various forecast ranges. It seems to be more important to add observations of those states, sometimes in the immediate and sometimes in the more distant past, that are especially predictive.

5. Discussion and conclusions

By superposing ensembles of perturbations on each of a succession of 'true' states of a simple system, we have obtained a succession of ensembles of perturbed states, in which any perturbed state might be mistaken for the true one. We have then examined the growth of uncertainty – the spread of the ensemble members about their mean – as each member advances in time. We have displayed our numerical findings in graphical form.

The most conspicuous feature of our diagrams is the strong dependence of the magnitude of the uncertainty at any particular time t_1, if the uncertainty is isotropic and of specified magnitude at an earlier time t_0, upon the characteristics of the states at times t_0 and t_1, rather than simply on the elapsed time $t_1 - t_0$. There are special times t_0 that are highly predictive – uncertainties introduced then do not grow so rapidly as those introduced somewhat earlier or later – and there are special times t_1 that are highly predictable – uncertainties introduced previously do not acquire the magnitude then that they reach somewhat earlier or later. In terms of weather prediction, it appears that knowing a state rather well at a highly predictive past time can contribute more than knowing the state equally well at one or even a succession of more recent but less predictive times.

Perhaps the most interesting question concerns whether these findings are true for many larger systems, and in particular for the global atmosphere or the large GCMs, or whether we have merely uncovered additional properties of our simple model. First of all, we do not expect the total atmosphere to show as much day-to-day variability in uncertainty as occurs in the model, simply because at any one time there are generally many synoptic systems, which at short range may be varying nearly independently. One system may be rapidly intensifying while another is decaying, and the effects may partially cancel. Thus, a single atmospheric state may act more like a medium-sized ensemble of states of the model. Of course the analogy is imperfect, because at longer range each synoptic system should influence all the others.

Viewed otherwise, a single realization of the model may act more like a restricted region of the atmosphere, or a restricted portion of a global model of the atmosphere – one about large enough to contain a single major synoptic system. We would expect the predictivity and the predictability within such a region to vary significantly with the synoptic situation. The uncertainty within a region with a rapidly intensifying system ought to act rather differently from that in one where not much is happening.

How might we test these ideas with real data, or at least with a fairly realistic GCM? The computational burden need not be a problem, because we could use the output of routine ensemble forecasts from an operational center. Further processing of the output could enable us to construct diagrams in the manner of Fig. 7, but with s and α evaluated over specified restricted regions.

Our findings seem relevant to the process of data assimilation, which, as currently carried out, is computationally very expensive. Might we do nearly as well by assimilating observations only at the highly predictive times, at least in certain regions, instead of at all times? It seems possible, but first we would require a means of identifying the predictive states when they occur, if, indeed there are states that are highly predictive in restricted regions. Recourse to routine ensemble forecasts might again prove advantageous.

In any event, we feel that anyone who might seek analogous results with state-of-the-art models would do well to work first with small but not-too-small models, perhaps with a few hundred variables and a dozen or more positive Lyapunov exponents. Possibly our findings would not hold up, but we have an intuitive feeling that they have wide applications.

6. Acknowledgments

I have been fortunate to have had the opportunity for frequent discussions of these ideas with my colleagues James A. Hansen and Kerry Emanuel. The 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

Aires, F. and Rossow, W. B. 2003. Inferring instantaneous, multivariate and nonlinear sensitivities for the analysis of feedback processes in a dynamical system: Lorenz model case-study. *Q. J. R. Meteorol. Soc.* **129**, 239–275.

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

Charney, J. G., Fleagle, R. G., Lally, V. E., Riehl, H. and Wark, D. Q. 1966. The feasibility of a global observation and analysis experiment. *Bull. Am. Meteorol. Soc.* **47**, 200–220.

Eady, E. T. 1951. The quantitative theory of cyclone development. In: *Compendium of Meteorology.* Am. Meteorol. Soc., Boston, MA, 464–469.

Farmer, J. D. and Sidorowich, J. J. 1991. Optimal shadowing and noise reduction. *Physica D* **47**, 373–392.

Farrell, B. F. 1990. Small error dynamics and the predictability of atmospheric flows. *J. Atmos. Sci.* **47**, 2409–2416.

Hansen, J. A. and Smith, L. A. 2001. Probabilistic noise reduction. *Tellus* **53A**, 585–598.

Ikeda, K. 1979. Multiple-valued stationary state and its instability of the transmitted light by a ring cavity system. *Opt. Commun.* **30**, 257.

Lacarra, J. and Talagrand, O. 1988. Short-range evolution of small perturbations in a barotropic model. *Tellus* **40A**, 81–95.

Lorenz, E. N. 1965. A study of the predictability of a 28-variable atmospheric model. *Tellus* **17**, 321–333.

Lorenz, E. N. 1984. Irregularity: a fundamental property of the atmosphere. *Tellus* **36A**, 98–110 (L84).

Lorenz, E. N. 1990. Can chaos and intransitivity lead to interannual variability? *Tellus* **42A**, 378–389.

Palmer, T. N. 1988. Medium and extended range predictability of the Pacific/North American mode. *Q. J. R. Meteorol. Soc.* **114**, 691–713.

Phillips, N. 1956. The general circulation of the atmosphere: a numerical experiment. *Q. J. R. Meteorol. Soc.* **82**, 123–164.

Poincaré, H. 1912. *Science et méthode*. Flammarion, Paris.

Trevisan, A. 1993. Impact of transient error growth on global average predictability measures. *J. Atmos. Sci.* **50**, 1016–1025.