Measuring Dynamical Prediction Utility Using Relative Entropy

RICHARD KLEEMAN

Courant Institute of Mathematical Sciences, New York, New York

ABSTRACT

A new parameter of dynamical system predictability is introduced that measures the potential utility of predictions. It is shown that this parameter satisfies a generalized second law of thermodynamics in that for Markov processes utility declines monotonically to zero at very long forecast times. Expressions for the new parameter in the case of Gaussian prediction ensembles are derived and a useful decomposition of utility into dispersion (roughly equivalent to ensemble spread) and signal components is introduced. Earlier measures of predictability have usually considered only the dispersion component of utility. A variety of simple dynamical systems with relevance to climate and weather prediction is introduced, and the behavior of their potential utility is analyzed in detail. For the climate systems examined here, the signal component is at least as important as the dispersion in determining the utility of a particular set of initial conditions. The simple ''weather'' system examined (the Lorenz system) exhibited different behavior with the dispersion being more important than the signal at short prediction lags. For longer lags there appeared no relation between utility and either signal or dispersion. On the other hand, there was a very strong relation at all lags between utility and the location of the initial conditions on the attractor.

1. Introduction

Fundamental limits to predictability have received considerable attention in recent years due to the pioneering work of Lorenz (1963), who showed that extreme sensitivity of weather predictions to the specification of initial conditions means that detailed forecasts are, in general, impossible beyond a certain time limit (later found by many to be around 2 weeks). This pioneering study led over the following decades to the extensive study of chaotic dynamics (e.g., Ruelle and Takens 1971; Grassberger 1983; Eckmann and Ruelle 1985).

Motivated by this fundamental atmospheric uncertainty, the concept of statistical prediction involving an ensemble of possible projections has become commonplace in weather and climate prediction (e.g., Leith 1974; Palmer et al. 1993; Toth and Kalnay 1993; Shukla 1998). From a theoretical perspective, interesting new generalized methods for defining predictability involving information theory and chaotic dynamical concepts have also been introduced (Carnevale and Holloway 1982; Smith 1996; Smith et al. 1999; Schneider and Griffies 1999). A notable feature of predictability is that ensemble spread may vary considerably indicating that certain predictions may be much more reliable than others. Some progress (Palmer et al. 1993; Toth and Kalnay 1993; Moore and Kleeman 1998; Palmer 2000) has occurred in utilizing this measure as an indicator of forecast skill. This notion has been formalized into a parameter of so-called ''potential predictability.'' This measures ensemble spread relative to the equilibrium or climatological spread. For the univariate case one may write

$$\mathrm{PP} = 1 - \frac{\sigma_E^2}{\sigma_c^2}, \tag{1}$$

where σ_E^2 and σ_c^2 are the ensemble and climatological variances, respectively. Such a measure implies that potential predictability declines through a forecast from a near-perfect value of $\mathrm{PP} \doteq 1.0$ for well-observed initial conditions through to the no-predictability case $\mathrm{PP} \doteq 0.0$ as $t \to \infty$.

Measures similar to PP have seen extensive application in analysis of various predictability scenarios [see, however, Smith (1996) for alternative atmospheric viewpoints]. As a measure of forecast utility however PP is of somewhat less value since although it takes into account uncertainty it does not reference it to anything except the equilibrium dispersion. Interestingly in climate prediction this is not the case. Here both in the case of atmospheric (Madden 1981; Shukla 1998) and ocean–atmosphere El Niño–Southern Oscillation (ENSO) prediction (Kleeman and Moore 1999) other forms of referencing are commonly employed. In the first case, variance due to ensemble spread is compared

Corresponding author address: Dr. Richard Kleeman, Courant Institute of Mathematical Sciences, 251 Mercer Street, New York, NY 10012.
E-mail: kleeman@cims.nyu.edu

to that due to (low frequency) boundary condition variation. In the second case, the variance is compared to the mean squared value of the prediction and this is found to be mathematically related to the widely used correlation skill measure. Both these ideas essentially boil down to determining a signal-to-noise ratio for predictions. A very simple example illustrates why this concept is important for prediction utility:

Suppose one is interested in forecasting a single variable with unit climatological variance. A particular ensemble prediction for this variable may have mean value $+2.0$ with ensemble spread of 1.0. The ensemble spread may be this large because of the increased instability of the initial conditions relative to the typical forecast. According to Eq. (1) we have PP = 0.0 and yet clearly the prediction has considerable value simply because it is forecasting a very large departure from normal conditions. If the ensemble spread were referenced to the mean-squared prediction on the other hand, we obtain for the "potential prediction utility" PPU[1]

$$ \text{PPU} = \frac{\mu^2}{\mu^2 + \sigma^2} = \frac{4}{5}, \qquad (2) $$

that is, showing considerable utility. This simple example actually occurred in the prediction of the huge 1997 El Niño (Moore and Kleeman 1999).

It is clear that in order to measure prediction utility we need to consider the behavior of the *total* forecast distribution relative to the equilibrium distribution not simply a comparison of the second moments.

In the next section we formalize a new measure of utility using information theoretic concepts and show that is has a number of very desirable properties. In section 3 we apply the new measure to a number of interesting (simple) dynamical examples. Section 4 contains a discussion, summary, and ideas for the practical use of the measure proposed here.

2. Formal definition of utility

Consider the following (classical) perfect model scenario: Due to uncertainty in the value of the initial conditions their values are given by a particular probability distribution p. This distribution evolves in time as a (statistical) prediction progresses. Given a reasonable dynamical system asymptotically this distribution approaches an equilibrium distribution q. If we assume ergodicity (i.e., that the long-time behavior of the system matches its equilibrium behavior), then this distribution can be thought of as the climatological distribution.

How should one measure the utility of a particular prediction? For the sake of clarity we will consider first the case where a perfect model is available. A very ap-

pealing way of measuring the usefulness of a prediction is to ask how much additional information is added to a particular situation by its availability. Obviously in a practical situation one already has information to hand on the past or climatological behavior of the system so a prediction should add to this. Information theory (e.g., Cover and Thomas 1991) provides a very natural measure of precisely this known as *relative entropy R*. This gives the information loss sustained by assuming climatology when the prediction distribution is available. If a discrete set of states are being predicted this is given by

$$ R = \sum_i p_i \ln\left(\frac{p_i}{q_i}\right), \qquad (3) $$

where q_i is the climatological distribution and p_i is that for the prediction. This parameter is also known as the Kullback Leibler distance as it measures the distance between the distribution $\vec{p}$ and $\vec{q}$ and only vanishes when they are identical. Another very attractive property of R is that if the dynamical process being modeled is Markov (an excellent approximation for the case considered here of perfect geophysical dynamical models) and $\vec{q}$ is the equilibrium (or asymptotic) distribution then R always decreases monotonically with time (Cover and Thomas 1991, section 2.9). This property is often referred to as a generalized second law of thermodynamics and interestingly only holds for relative entropy and not absolute entropy. In our context it means that due to chaos, prediction model utility always declines (monotonically) with the length of the forecast. At a sufficiently long lag utility approaches zero as the prediction distribution approaches the equilibrium distribution.

Given the above discussion, in the present contribution we shall use the terms relative entropy and prediction utility interchangeably. We shall also distinguish between prediction utility and the more general term predictability. This latter term has a variety of definitions in the literature and so we choose to coin a new term "utility" in order to distinguish our measure of predictability from others.

a. Practical considerations

The definition of predictive utility given above is idealistic in the sense that it takes no account of the physical accuracy of the prediction model. In other words it is a perfect model measure. In practical situations one would like to take into account errors in the model as well. In principle this could be achieved by also computing the relative entropy between the model prediction distribution and that appropriate to the real world. This quantity measures the amount of information lost by making the inaccurate model ensemble prediction. Actually determining the real world distribution is, however, a challenging task as in general only one realization actually occurs [see Smith (1996) for a careful and interesting discussion on this point]. This practical prob-

---

[1] The precise functional form of PPU is naturally unimportant. We use that deployed by Kleeman and Moore (1999).

(a)

One realization from the simplest stochastic oscillator

(b)

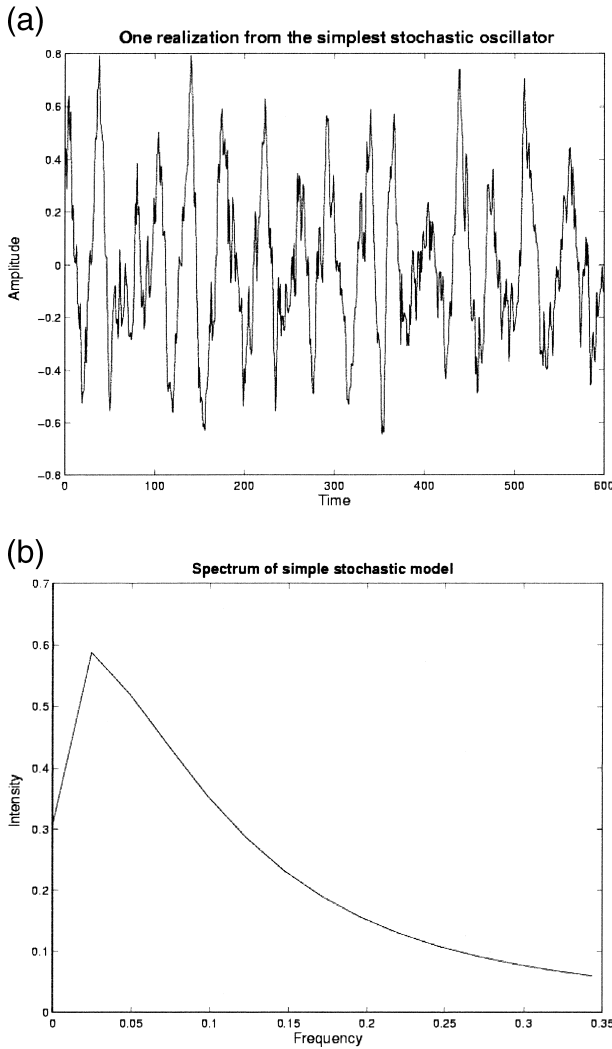**Spectrum of simple stochastic model**

FIG. 1. (a) A single realization from a simple stochastic oscillator. The integration extends over many cycles of the oscillator. (b) The spectrum of the oscillator calculated under the (true) assumption that it is an AR(2) process.

lem will of course occur no matter what kind of measure of predictability is deployed. Discussion of this rather subtle issue is deferred to a future publication, our aim here is try to understand how prediction utility may be affected by dynamical effects and so a perfect model scenario is considered appropriate.

3. Utility behavior for different dynamical systems

a. Gaussian distributions

In the case that the prediction and equilibrium distributions p and q are Gaussian of finite dimension n, a closed form analytical expression may be obtained for the relative entropy. Let us assume that the first and second moments of these distributions are denoted by μ_i^p, $(\sigma_p^2)_{ij}$ and μ_i^q, $(\sigma_q^2)_{ij}$, respectively. Further let us in-

troduce the continuous distribution form for relative entropy (Cover and Thomas 1991, chapter 9):

$$ R = \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 \cdots \int_{-\infty}^{\infty} dx_n p \, \ln\frac{p}{q}. \quad (4) $$

Given the standard form of Gaussian distributions (e.g., Gardiner 1985) it is straightforward to show that

$$ R = \frac{1}{2}\left\{ \ln\left[\frac{\det(\sigma_q^2)}{\det(\sigma_p^2)}\right] + \text{tr}[\sigma_p^2(\sigma_q^2)^{-1}] \right. $$
$$ \left. + (\vec{\mu}^p - \vec{\mu}^q)^t(\sigma_q^2)^{-1}(\vec{\mu}^p - \vec{\mu}^q) - n \right\}. \quad (5) $$

Below we shall refer to the first two terms minus n as the *dispersion* component and the third term as the *signal* component of the relative entropy. It is worth noting that the first term is the regular entropy measure proposed and extensively analyzed by Schneider and Griffies (1999). It is rather revealing to consider the univariate specialization of this equation:

$$ R = \frac{1}{2}\left[\ln\left(\frac{\sigma_q^2}{\sigma_p^2}\right) + \frac{\sigma_p^2}{\sigma_q^2} + \frac{(\mu^p)^2}{\sigma_q^2} - 1 \right], $$

where we are assuming without loss of generality that the equilibrium distribution has zero mean. It is clear now that for Gaussian distributions the effects of PP and PPU expressed through Eqs. (1) and (2) are both incorporated into the relative entropy measure of utility. Such a result is hardly surprising since both the relative dispersion of prediction and climatology as well as the mean value of the prediction are important in measuring how "different" the prediction and climatological distributions are from each other and hence determining the information content of the prediction. For the concrete example considered in section 1, it is easily seen that the dispersion part (as well as the first term) of the relative entropy vanishes and we have

$$ R = \frac{1}{2}\frac{(\mu^p)^2}{\sigma_q^2} = 2. $$

b. Stochastically forced damped linear oscillator

This is an interesting first example to consider since exact analytical solutions are possible and this simple model has been proposed by several authors as a "null hypothesis" for a model of the El Niño–Southern Oscillation (see, e.g., Kestin et al. 1998). Consider the following two-dimensional stochastic differential equation:

$$ \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}_t = \begin{pmatrix} 0 & 1 \\ \beta & \gamma \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \begin{pmatrix} 0 \\ F \end{pmatrix}, \quad (6) $$

where F is white with variance C and mean zero. Without the forcing, it is easily shown that damped oscillations occur with period T and damping time τ where we have

$$\gamma = -\frac{2}{\tau} \qquad \beta = -\left(\frac{4\pi^2}{T^2} + \frac{1}{\tau^2}\right). \qquad (7)$$

A realization of this stochastic differential equation with $\tau = T = 36$ months is displayed in Fig. 1 together with a spectral analysis. Clearly the oscillation period T is still noticeable but as the spectrum shows considerable broadening has occurred due the forcing. The statistical solution of these equations for the covariances and means of u_1 and u_2 has been discussed by Gardiner (1985). The covariance matrix at time t (given a deterministic set of initial conditions at time 0) is given by

$$\begin{bmatrix} \langle u_1, u_1 \rangle & \langle u_1, u_2 \rangle \\ \langle u_2, u_1 \rangle & \langle u_2, u_2 \rangle \end{bmatrix}$$

$$= \int_0^t dt' \, \exp[(t - t')A]\begin{pmatrix} 0 & 0 \\ 0 & C \end{pmatrix} \exp[(t - t')A^t],$$

where

$$A = \begin{pmatrix} 0 & 1 \\ \beta & \gamma \end{pmatrix}.$$

Further the mean vector at time t is given by

$$\begin{pmatrix} \overline{u}_1 \\ \overline{u}_2 \end{pmatrix} = \exp(tA)\begin{bmatrix} u_1(0) \\ u_2(0) \end{bmatrix}.$$

Since the equations are linear it follows that all probability distribution functions will be Gaussian providing that the stochastic forcing has this property that we assume. In order to evaluate R we therefore require the covariance and means of the transient and equilibrium (i.e., as $t \to \infty$) ensembles. Analytical solutions can be obtained in a straightforward way by an evaluation[2] of $\exp(sA)$

$$\exp(sA) = e^{-s/\tau}\begin{bmatrix} \cos\omega s + \dfrac{1}{\omega\tau}\sin\omega s & \dfrac{1}{\omega}\sin\omega s \\ -\dfrac{1}{\omega}\left(\dfrac{1}{\tau^2} + \omega^2\right)\sin\omega s & \cos\omega s - \dfrac{1}{\omega\tau}\sin\omega s \end{bmatrix} \qquad \omega \equiv \frac{2\pi}{T}.$$

The equilibrium distribution is obtained easily from this and the means of this distribution are zero as well as the covariance between u_1 and u_2. The equilibrium variances are given by

$$\sigma_1^2 = \frac{C\tau^3}{4(1 + \omega^2\tau^2)} \qquad \sigma_2^2 = \frac{C\tau}{4}.$$

Calculation now of the relative entropy or prediction utility R for all prediction ensembles for this dynamical system is straightforward.

A very important property of this dynamical system is the fact that the covariance of transient distributions is independent of the initial conditions for a particular prediction. This means that only the signal component of the prediction utility R shows any variation with initial conditions. This is a striking counterexample to the widespread perception that ensemble spread is the main determinant of potential forecast skill. Here the ensemble spread is identical for all predictions of a given time and yet the prediction utility R can actually vary quite markedly. Figure 2a shows how the utility can vary from initial condition to initial condition: Utility at various prediction lags is shown from a particular set of (randomly chosen) 60 initial conditions drawn from the realization of the stochastic system displayed in Fig. 1. The probability distribution of utility at 12 months is shown in Fig. 2b. This was constructed using 10 000

initial conditions drawn at random from the realization of Fig. 1. Thus a prediction from this dynamical system can be considerably more useful than normal simply because (by chance) it has a particular set of initial conditions that "contain a large signal."

Viewing of Fig. 2a shows, as predicted in the previous section, prediction utility drops monotonically with time and obviously approaches zero as the ensemble relaxes toward the equilibrium or climatological distribution. It is important to note however that this property holds only when the state vector for the *entire* dynamical system is used to calculate utility. If only part of this vector is used, then this no longer holds since information can flow from one part of the state space to another. For our particular example one can calculate the utility of just the variable u_2:

$$R_2 = \frac{1}{2}\left\{\ln\left[\frac{\sigma_c^2(u_2)}{\sigma^2(u_2)}\right] + \frac{\sigma^2(u_2)}{\sigma_c^2(u_2)} + \frac{(\overline{u}_2)^2}{\sigma_c^2(u_2)} - 1\right\}.$$

Plotted in Fig. 3 is R_2 for the case that the initial conditions have the form $u_2(0) = 0$; $u_1(0) = c$ and one notes the short term rise in utility.

In terms of the damped oscillation of this system the variables u_1 and u_2 are phase shifted by 90° and so it follows that information contained in the variable

---

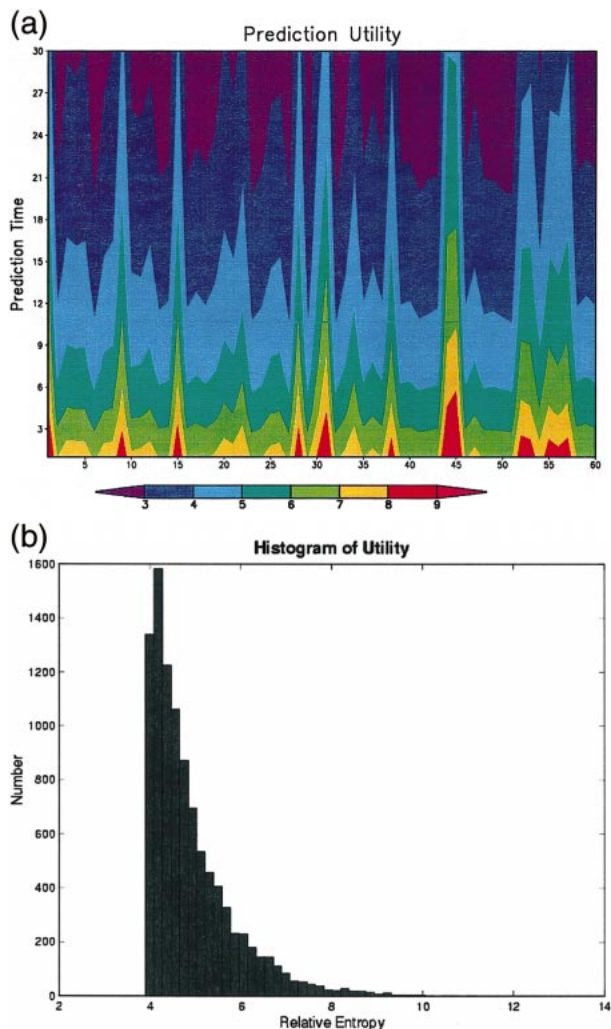[2] This is obtained by diagonalizing A.

FIG. 2. (a) The utility at various times of 60 randomly chosen predictions from a simple stochastic oscillator. (b) The distribution of utility at a given time for the simple stochastic oscillator.
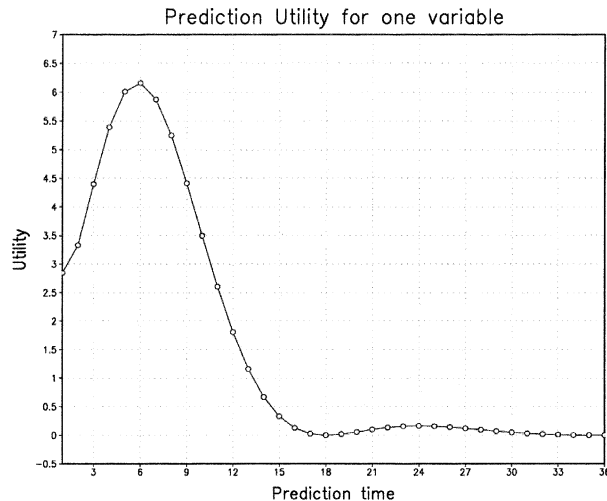

FIG. 3. The utility at various times of one particular variable from a simple stochastic oscillator. Note the increase in utility for the short-range prediction.

c. Linear oscillator with varying stability

Evidence from ENSO dynamical models (e.g., Moore and Kleeman 1997; Chen et al. 1997) suggests that the simple model of the previous subsection should be modified to take into account potentially large variations in the stability of the system caused by both the annual and ENSO cycles. We modified the model then by allowing the parameter τ to vary periodically with time. We chose periods P for this variation of $T/3$ and T and assumed a sinusoidal variation in $1/\tau$ of the form
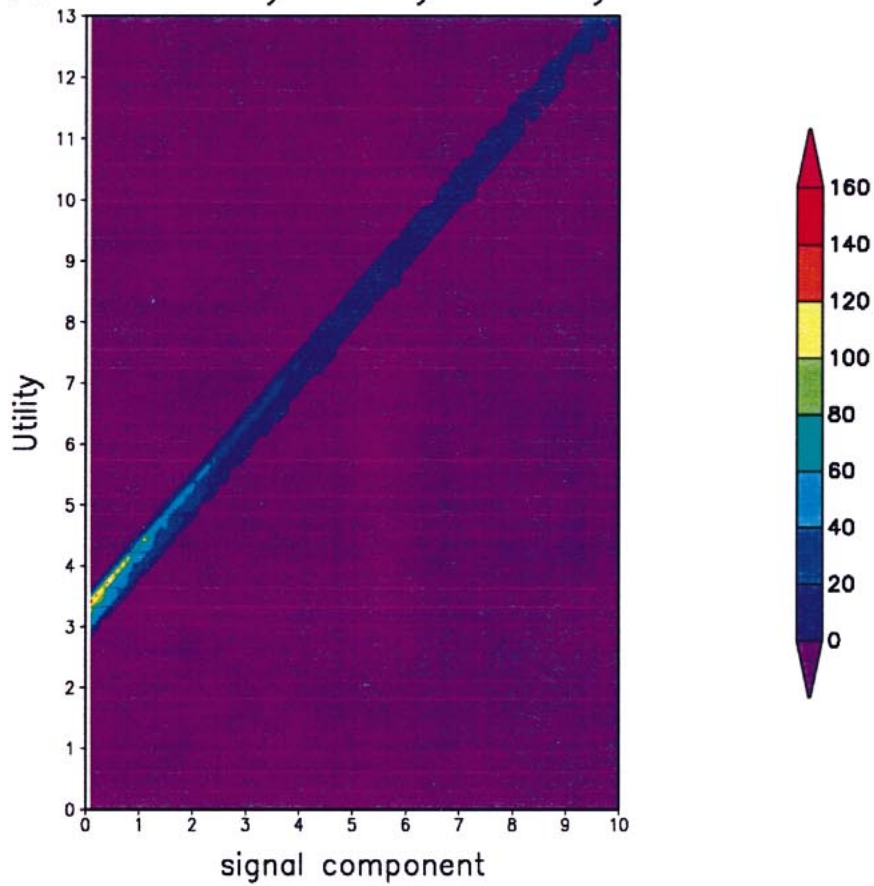
$$\frac{1}{\tau} = \left(1 + 2\sin\frac{2\pi t}{P}\right)\Big/ T. \qquad (8)$$

Clearly for certain times the oscillator is now highly unstable and so one might expect large variations in ensemble spread depending on the particular initial conditions chosen. Despite the varying stability of our new system all ensemble distributions are still Gaussian[3] and we may still therefore use the expressions of Eq. (5) to calculate utility. Completely analytical expressions are now not easily obtained and so we rely on numerical models of our equations to estimate the first and second moments of the prediction ensembles. All results reported here were checked for convergence with respect to ensemble size.
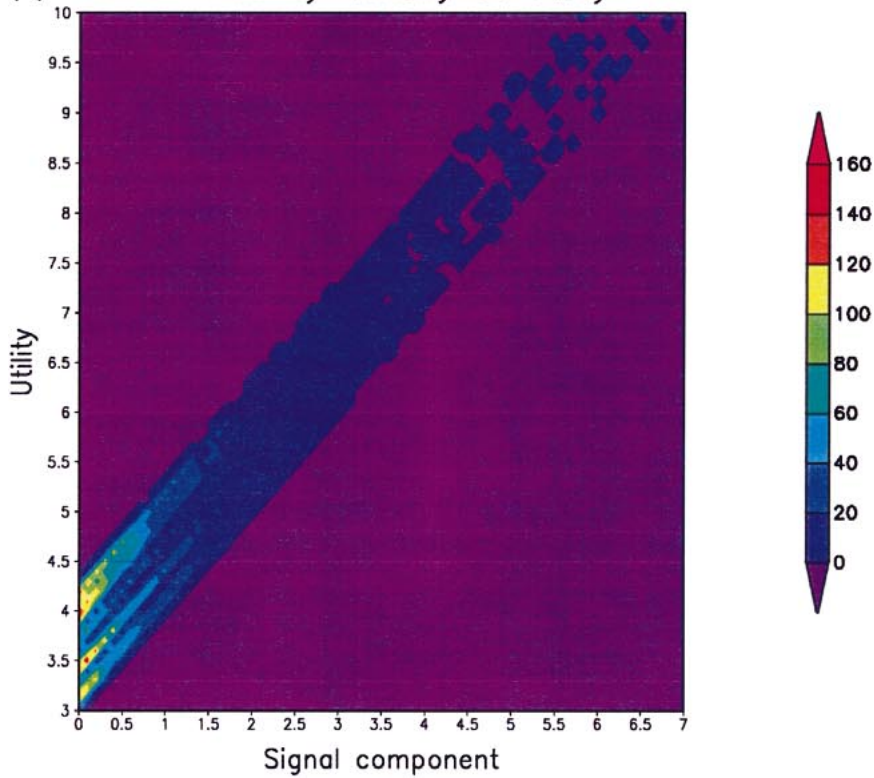
It is interesting now to compare the relative importance of the dispersion and signal term in the relative utility. This can be assessed by plotting the signal term versus the total utility for a large number (10 000) of randomly chosen initial conditions and this may be seen in Fig. 4. Here the probability density for each point on

---

[3] See Gardiner (1985, chapter 4). Note that prediction ensemble distributions become non-Gaussian only when the operator A becomes nonlinear.

u_1 can appear in the variable u_2 one-quarter of a period later. This situation has practical application because in the analogy to ENSO discussed above, the variable u_2 can be considered to measure eastern Pacific sea surface temperature (SST) anomaly and therefore be a measure of the global atmospheric effects of this phenomenon. The other variable u_1 is uncorrelated with these effects and can dynamically be considered to represent subsurface oceanic temperature perturbations that do not influence SST such as those occurring in the western Pacific. Thus information about the ocean subsurface that has no immediate utility in global climate prediction can be quite useful some nine months later when it strongly influences eastern Pacific SST (and hence global climatic phenomenon). This fact forms much of the physical basis of current ENSO prediction.

(a) Probability density of utility

(b) Probability density of utility

the plot is estimated using a ''circle of influence'' measure, that is, the number of sample points lying within a suitably small radius of parameter space[4] was calculated and used to estimate density. Results are shown for prediction times of one-third of the oscillators period (i.e., 12 months for the system displayed in Fig. 1). Figure 4a shows the results when the stability varies with period $T/3$ and while it is clear that dispersion has some effect on utility it is still the signal term that appears more important overall. In Fig. 4b the case where the stability varies with period T is depicted and now it is apparent that dispersion becomes more important to utility although it is clear that the signal term still remains very important.

The robustness of these results was tested by varying the stability parameters in Eq. (8) quite significantly. The only parameter causing a relative change in the importance of the dispersion and signal terms was the period of the stability cycle.

d. A stochastically forced coupled ocean–atmosphere model

The ENSO phenomenon has received considerable attention in recent years from mathematical modelers (see, e.g., Zebiak and Cane 1987; Ji et al. 1994; Kirtman and Shukla 1998) and considerable success has been obtained both in realistic dynamical simulation as well as prediction. The phenomenon is broadband with a spectral intensity peak of around 4 yr. Recently the causes of the broadband (as opposed to oscillatory) behavior of the phenomenon have received considerable attention. A leading candidate to explain this (see, e.g., Penland and Sardeshmukh 1995; Kleeman and Moore 1997) has been stochastic forcing of the low-frequency climate system by climatically unpredictable atmospheric transients such as those prevalent in the deep Tropics. Models of this form are able to accurately reproduce the observed irregularity of ENSO very robustly (see, e.g., Moore and Kleeman 1999; Thompson and Battisti 2000). In addition these models have considerable predictive skill (Kleeman et al. 1995), which adds credibility to the stochastic scenario. The models are also computationally inexpensive and so they are useful vehicles for examining the nature of prediction utility in the ENSO context (see, however, the discussion in section 2a concerning imperfect practical models). Here we use the stochastic model of Moore and Kleeman (1999), which consists of an intermediate coupled ocean–at-

mospheric forced by stochastic input which has the spatial structure of the first two stochastic optimals which represent the most efficient ways to induce variance growth within the stochastic dynamical system [see Kleeman and Moore (1997) for details on this terminology].

Examination of the ensemble behavior for a variety of dynamically interesting variables from the model shows that the short range (up to around 6–9 months) ensembles are Gaussian to a reasonable approximation. Beyond this and certainly for the equilibrium probability distribution, there is evidence of non-Gaussianicity in the form of a weak bimodality. Whether this is a feature of the real system or not is unclear as there is not really a sufficiently reliable dataset available to decide this property with confidence. In order then to calculate utility efficiently for this system, we confine ourselves to short range predictions and estimate the equilibrium distribution using a hypothesis of ergodicity for the system and a very long (10 000 year) integration. Restriction to short-range predictions is necessary to make this undertaking feasible since only the variance and mean of the ensembles need calculation rather than the entire distribution, which would converge more slowly with ensemble size. We took 100 sets of randomly chosen initial conditions from the very long integration mentioned above and constructed 100 member ensembles each of 6-month's duration. This ensemble size was sufficient to ensure convergence of the second moments of quantities examined. For the equilibrium distribution we used the adaptive mixtures algorithm (Priebe 1994) to estimate the distribution as a (positive) sum of Gaussian distributions with different first and second moments. The utility was calculated with respect to the variable known as Niño-3, which is the generally accepted global parameter of the ENSO state and measures the average sea surface temperature anomaly in the eastern equatorial Pacific (values greater than say 1.0 are commonly referred to as El Niño while values less than around −1.0 are called La Niña). Since we are not calculating the utility of the full state variable (this is typically of order 2000 in dimension for this model), we may expect that the utility of Niño-3 predictions will not universally decline as indeed was noted.

Displayed in Fig. 5 is a plot of utility against forecast lag for a sample of 20 of the initial conditions. Note that most of the time there is a monotonic decline but not always as mentioned. Note also the large variation in the value of the utility reflecting the apparently large fluctuation in the potential usefulness of different ENSO predictions.

Despite the equilibrium distribution not being Gauss-

---

[4] This was chosen for convenience to be 0.1. Note the axes scale in Fig. 4 for comparison.

FIG. 4. (a) The probability distribution of predictions from a stochastic oscillator as a function of signal and utility. The stochastic oscillator has stability that varies with a period of one-third of the period of the oscillation. (b) The same as (a) but with the stability cycle period extended to be equal to that of the oscillator.
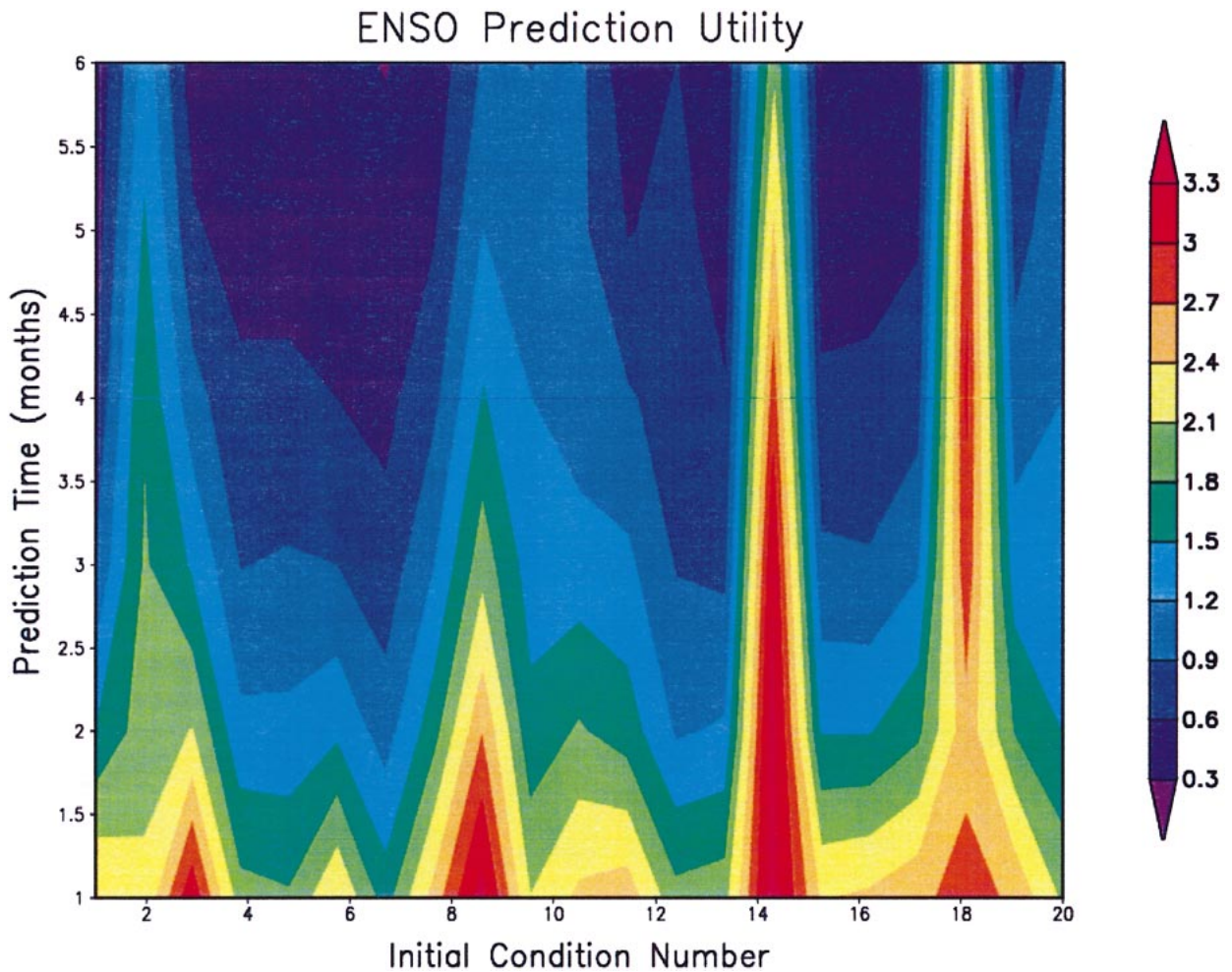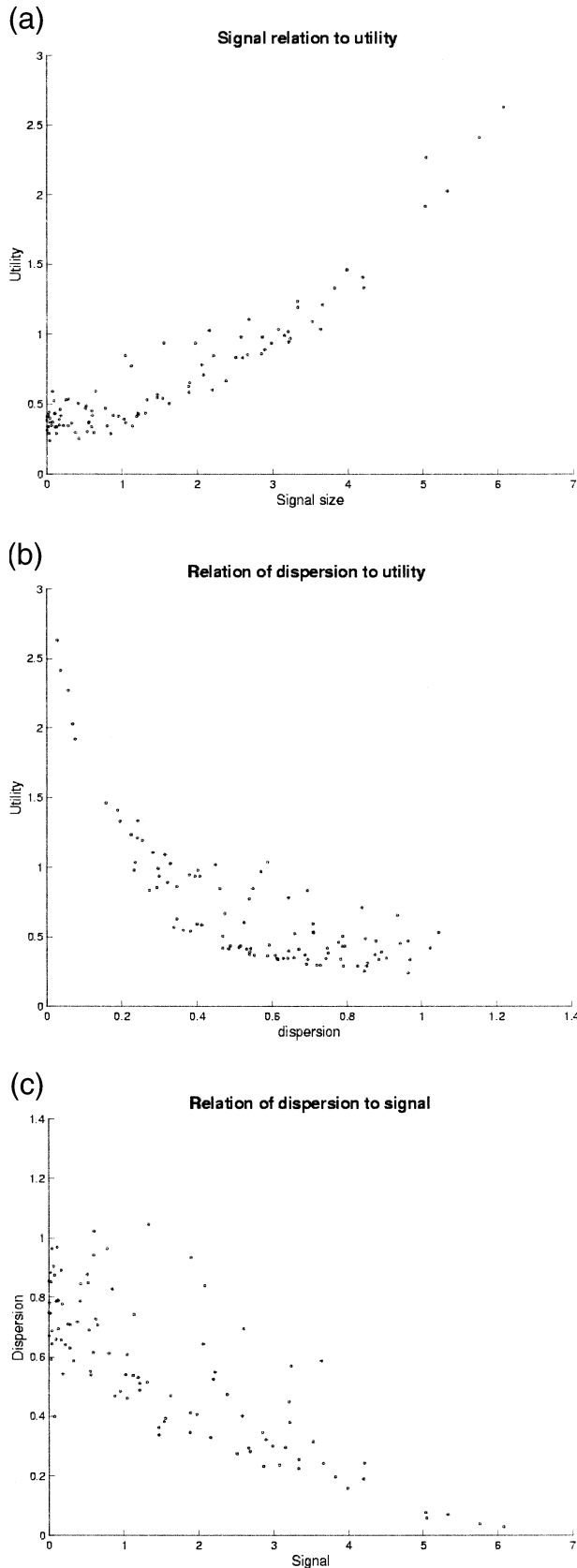
ENSO Prediction Utility

FIG. 5. The utility of Niño-3 predictions at varying lags from a stochastically forced coupled ocean–atmosphere model of ENSO (see text). There are 20 randomly chosen 6-month predictions displayed.

ian in this case, it is still interesting to see whether the signal and/or dispersion are indicators of utility. Displayed in Figs. 6a and 6b are plots of ensemble signal (the mean-squared value of Niño-3 for the ensemble) versus utility and dispersion versus utility. We see that both these parameters show some relationship with utility although the relationship with signal seems stronger. Interestingly in the case of this dynamical system (unlike those simple systems considered previously) dispersion and signal are not independent of each other which is an indication of nonlinearity. Figure 6c shows how these quantities are related to each other for the 100 ensemble predictions. The reason for this (nonlinear) relationship is as follows: The ensemble spread (or variance growth) depends on the local stability of the initial conditions used in predictions. Moore and Kleeman (1997) showed that this stability can be strongly influenced by the particular phase of ENSO that the initial condition comes from and in particular when the amplitude of the ENSO is large, the instability is reduced mathematically because of a nonlinearity in the ocean component of the model,

which restricts the magnitude of SST anomalies to being smaller than a fixed upper bound.

e. Lorenz attractor

The models considered thus far may be taken as analogs for the kind of behavior one might expect to encounter in climate prediction where there is a very clear separation between the slow scales (which are considered climate variables) and the fast scales, which are considered to be essentially stochastic. For the case of weather prediction this separation is less apparent (see, however, e.g., Egger and Schilling 1984) and so one may expect predictability to possibly have a different nature. There are many simple dynamical systems available that are analogs for weather dynamics and we plan to investigate a number of these in more detail in a future publication. Here we confine our attention to perhaps the best known of such systems, namely, the Lorenz system (Lorenz 1963), which exhibits chaotic behavior and has a noninteger dimensional attractor (Nayfeh and

(a)

Signal relation to utility

(b)

Relation of dispersion to utility

(c)

Relation of dispersion to signal

Balachandran 1995). It has three state variables satisfying

$$\dot{x} = -\sigma(x - y) \qquad \dot{y} = \rho x - y - xz$$
$$\dot{z} = xy - \beta z.$$

In this study we chose $\sigma = 10$, $\rho = 8/3$, and $\beta = 28$, which represent fairly typical values from the vast literature on this system. For the numerical results to be reported below a standard leapfrog method of integration was deployed to obtain solutions with a time step of 0.001.

One of the difficulties in applying the formalism of the previous section to the Lorenz system concerns defining an appropriate formula for relative entropy. Since the dimensionality of both the prediction ensembles and the complete attractor are less than 3, one must proceed with considerable caution. In principle, a rigorous formulation is possible for the usual (absolute) entropy involving Lebesgue measurable sets (Badii and Politi 1997, p285), which are required to define integrals on noninteger dimension manifolds. Using such methodology the author was able to calculate the absolute entropy of the attractor by using the technique of saturation curves (Nayfeh and Balachandran 1995, section 7.9), which have been typically used to estimate the so-called information dimension of the attractor. By choosing successively smaller spheres of radius r to estimate the probability density of the attractor $P(\vec{x})$, one calculates the attractor expectation of the logarithm of this density:

$$E = -\langle \ln[P(\vec{x})] \rangle$$

and plots it against $\ln(r)$. The slope of this curve becomes constant for small enough r and a sufficiently large sample from the attractor {required to adequately estimate $\ln[P(\vec{x})]$} and its value (~ 2.04) is the so-called information dimension. It can be demonstrated (see appendix) then that the intercept ($E \sim 3.38$) of the linear section of this curve serves as an adequate definition of the absolute entropy. Practical problems, however, occur in the calculation of the relative entropy because the dimensionality of the prediction ensemble appears for practical purposes to be less than that of the full attractor.[5] This behavior is illustrated graphically in Fig. 7 where initial conditions from an arbitrarily chosen point are perturbed along a plane lying approximately within the attractor. An ensemble sample of 1000 is

---

[5] The dimension of the prediction manifold will actually not change from that at the initial time but the effects of the cascade to smaller scales (effectively mixing) ensure that for practical purposes the dimension appears low (see Fig. 7 for intuitive insight on this point).

$\leftarrow$

FIG. 6. (a) The relationship between the utility of 100 6-month Niño-3 predictions and the Gaussian signal of the predictions. (b) Same as (a) but for the relationship with Gaussian dispersion. (c) The relationship between Gaussian signal and dispersion.
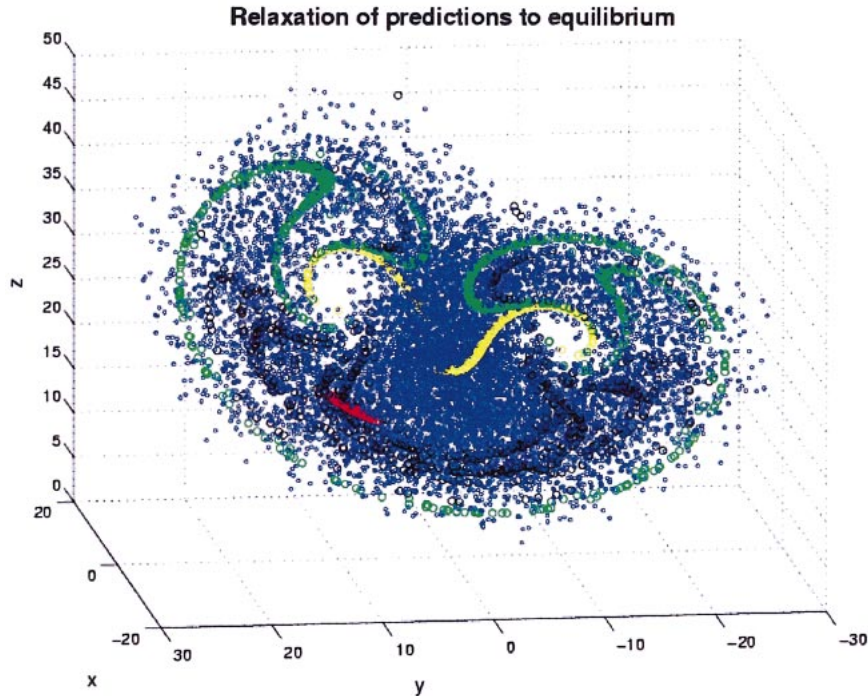
FIG. 7. The relaxation of an ensemble of predictions for the Lorenz model from a tight set of initial conditions. Different colors show the ensemble behavior at different times with red showing it at $t = 2000$; yellow at $t = 4000$; green at $t = 6000$; and black at $t = 8000$. The blue points show the equilibrium distribution. Transient (as opposed to equilibrium) distributions are shown as larger points for clarity.

chosen according to a Gaussian distribution lying in this plane with a uniform (very small) standard deviation. The figure shows the evolution of this sample back toward the equilibrium attractor distribution with different colors representing ensembles at different prediction times. As can be seen (and has been commented on often in the literature), the ensemble rapidly elongates in a preferred direction and this ''string'' convolutes slowly to fill the equilibrium attractor (represented by blue points). In practical terms, it is difficult to estimate the dimension (and then the intercept) of these prediction manifolds because very large sample sizes are required to carry out the saturation curve technique.

Given these problems we chose to evaluate relative entropy based on a fixed (small) value of r for the evaluation of the probability density. This corresponds also to the practical situation where knowledge of the probability density function (pdf) is subject to observational uncertainty. Thus we define

$$\mathrm{RE}(r) \equiv \left\langle \ln \frac{P_r(\vec{x})}{Q_r(\vec{x})} \right\rangle, \qquad (9)$$

where the subscript for the P (prediction) and Q (climatology) distributions means that they are evaluated with reference to a finite-resolution radius r. The expectation brackets are taken to mean with respect to the prediction ensemble. This measure of information con-

tent for the prediction ensemble may be interpreted as the information available at resolution r. The definition given in Eq. (4) measures the information content of the prediction at *all* resolutions.

A sample of 1000 randomly chosen initial conditions from the attractor were chosen and ensembles of 100 000 members were then constructed for each initial condition using the (very tight) Gaussian distribution discussed above.[6] Climatological probability distributions on the prediction ensembles [cf. Eq. (9)] were estimated using 10^6 points from the complete attractor, which were obtained by integrating the system for 5×10^9 time units and sampling every 5000 time units. We are assuming that the system is ergodic, which allows us to infer the equilibrium distribution from a long-time average. Values for RE(r) were estimated for $r = 0.1$, which is reasonably high resolution for this attractor from a practical viewpoint.[7] Values were calculated at time intervals of 2000 up to a limit of 20 000 by which

---

[6] The convergence of the relative entropy with respect to sample size was carefully checked and a prediction ensemble of 100 000 was found to be more than enough to ensure accuracy in the first decimal place of the estimates presented below.

[7] Since the attractor has ''size'' around 10 units for all dimensions, this value for r represents knowledge of the dynamical system two orders of magnitude smaller than the typical excursion, i.e., good accuracy.
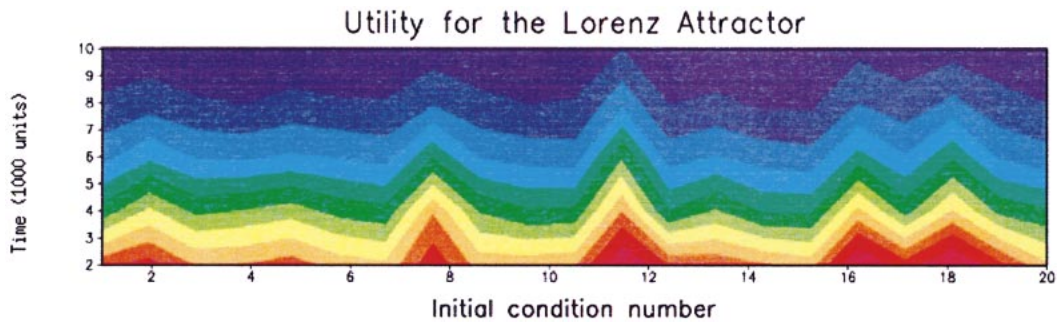
Utility for the Lorenz Attractor

FIG. 8. Variation in utility at different times for 20 randomly chosen predictions from the Lorenz system. The units on the vertical axis are in multiples of 1000.

stage there was typically little discernible difference between the prediction ensemble and the equilibrium attractor.

The typical behavior of utility with time is shown in Fig. 8, which displays results for 20 randomly chosen initial conditions. At most time lags there was a noticeable spread in the values of the utility for differing initial conditions. Shown in Fig. 9 is the distribution in values at $t = 4000$ and $t = 8000$. Also notable was the fact that this spread in utility tends to follow the topology of the attractor. In other words, initial conditions drawn from certain regions of the attractor tend to have higher prediction utility than those from others. Moreover this ''regionalization'' of utility was consistent throughout the prediction time interval so that predictions drawn from a particular part of the attractor tend to maintain their high or low utility right throughout the prediction. These effects are illustrated in Fig. 10 where the utility of predictions at $t = 4000$ and $t = 8000$ are displayed for the entire sample of 1000 initial conditions. The degree of utility is color coded according to a rainbow schema with high utility predictions having a violet color and low utility predictions having a red color. This dependence of utility on the location of initial conditions on the attractor has also been noted by Palmer (1993) for other measures of predictability.

It is interesting to consider how the utility here relates to more traditional measures of predictability. This was examined in two ways: first the prediction and equilibrium ensembles were assumed (for the sake of argument) to be Gaussian[8] and the dispersion and signal components calculated according to Eq. (5). Second the three-dimensional ensemble spread ($= \sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2}$), which is a measure of predictability often examined in atmospheric contexts, was compared with utility to determine if it has any skill in determining utility. For short-range predictions ($t = 2000$) it was found that the relative entropy calculated according to a Gaussian assumption showed a quite strong relation to the measure in Eq. (9). This (nonlinear) scatter relation is shown in Fig. 11a and a

decomposition of the Gaussian measure shows that most of the relation is due to the dispersion (rather than signal) component (Fig. 11b). For longer lags, this relationship is no longer a very good one as we see in Fig. 11c, which applies at $t = 8000$. The three-dimensional ensemble spread statistic was somewhat less skillful at predicting utility than dispersion. Figure 12 shows the scatterplot relation between spread and utility at $t = 2000$ and $t = 4000$. Clearly there is some relation at the short lag but it is probably not as clear as that for dispersion. For the longer lag the relation is poor (and worse than that for dispersion). The good relationship between dispersion and utility noted for short lags suggests that a relatively straightforward generalization of ensemble spread to a multidimensional environment (see Schneider and Griffies 1999) could be productive.

It is worth comparing the results found here to those found by Smith et al. (1999). These authors found that there was some return of predictability at longer prediction lags for the Lorenz model. Their definition of predictability was related to our Gaussian dispersion [see Eq. (5) and the discussion following it] so a direct comparison of results is not strictly possible, since our measure clearly contains first (and higher) moments of the pdf's (the signal in the Gaussian context) as well as second moments. It should be noted also that while the generalized second law of thermodynamics will hold for utility defined by Eq. (4) and for pdf's evolving according to a time-stepping algorithm for the Lorenz system [this has been rigorously demonstrated by Cover and Thomas (1991)] it need not *necessarily* hold for the coarse-grained version of relative entropy [Eq. (9)] since the dynamical system may not necessarily be Markovian at coarse scales even when it is for all scales. To check this possibility we carefully examined our large (1000 member) sample for monotonicity of relative entropy on the timescales examined by Smith et al. (1999). Very occasionally, relative entropy showed a small increase with time; however, the effect was probably not statistically significant. Overwhelmingly, relative entropy showed a decline for almost all 1000 initial conditions and at all prediction times. This leads one to the initial conclusion that it is the difference in the measures of

---

[8] As can be seen from Fig. 7 this is far from a reasonable assumption for both the equilibrium and prediction ensembles.
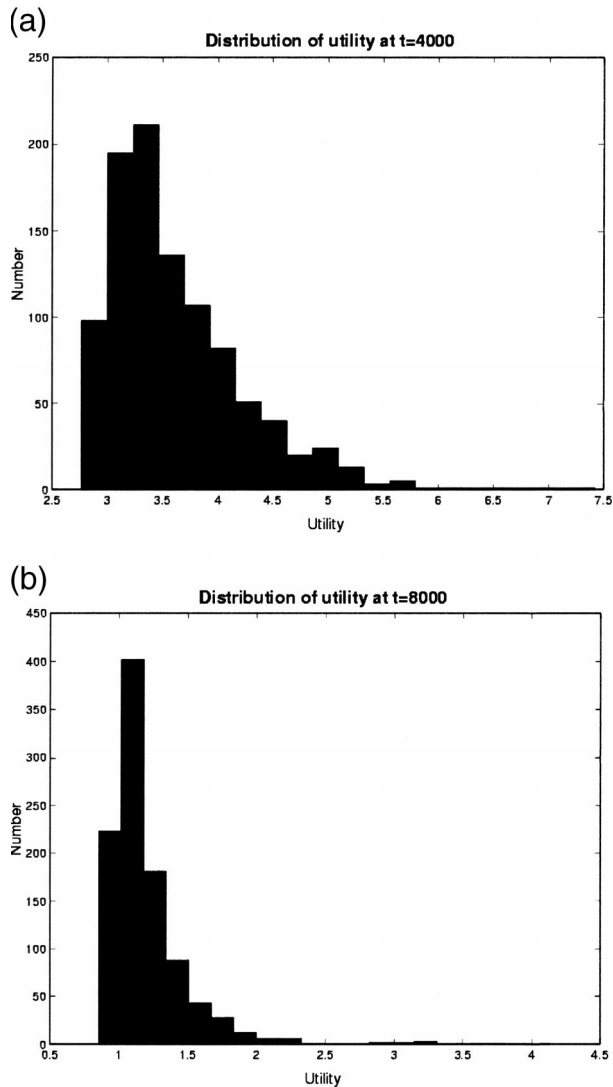
(a)

FIG. 9. (a) The distribution of utility for the Lorenz system at $t = 4000$. (b) The same as (a) but at $t = 8000$.
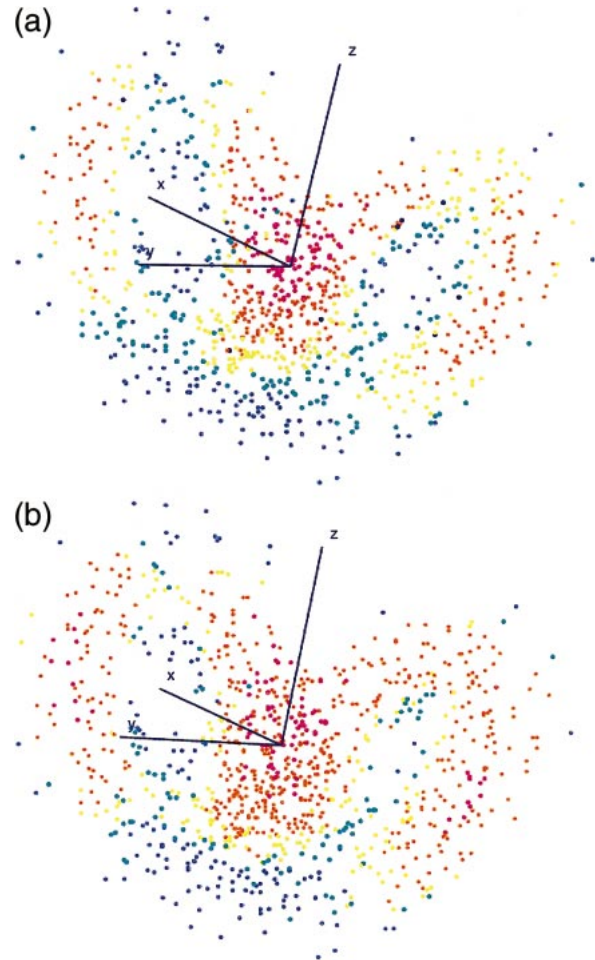

(a)

(b)

FIG. 10. (a) A three-dimensional view of utility as a function of initial condition location for the Lorenz system at $t = 4000$. The five colors (red, orange, yellow, green, and blue) show points with increasing values of utility. The color selection of utility range is chosen to give roughly equal numbers of points for each category. (b) The same as (a) but at $t = 8000$.

predictability used here and in Smith et al. (1999) that may account for the differing conclusions regarding the predictability of the Lorenz system. These subtle issues are currently being further investigated by the author and coworkers.

4. Summary and conclusions

A natural new measure of prediction utility for dynamical systems that is derived from information theory is introduced. It measures the additional information provided by a prediction over that already available (and usually well known) from the climatological or equilibrium distribution. This measure is well known in information theory and is referred to there as relative entropy. It has the intuitively very appealing property that for Markov processes it declines monotonically to zero

with increasingly long-range predictions. Thus as is intuitively obvious, utility of predictions declines with time until asymptotically they are of no use since they contain no information that is not already known from extensive historical observation. This property of entropy (known as the generalized second law of thermodynamics) is only applicable to relative entropy and in fact does not hold for absolute entropy (see Cover and Thomas 1991). Another way of viewing this measure of utility is as the distance between the prediction ensemble probability distribution and the climatological distribution. It is also worth emphasizing that this law holds only for state space as a whole. If a subset is considered (e.g., a single variable) there can be increases in utility since information can pass from one variable to another within the system.

It is useful to consider precisely what utility or relative entropy measures from an information theoretical
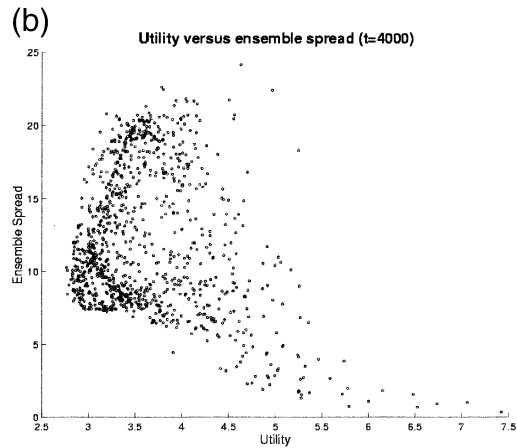
(a)

Actual versus Gaussian utility

(b)

Actual Utility versus Gaussian dispersion

(c)

Actual Utility versus Gaussian Utility (long prediction)

(a)

Utility versus ensemble spread (t=2000)
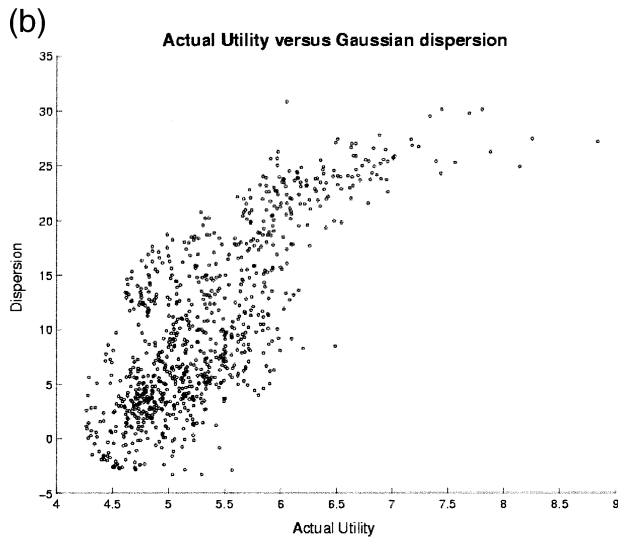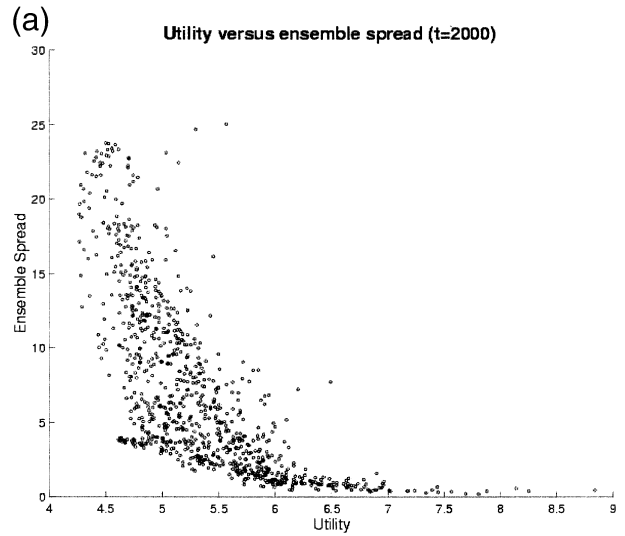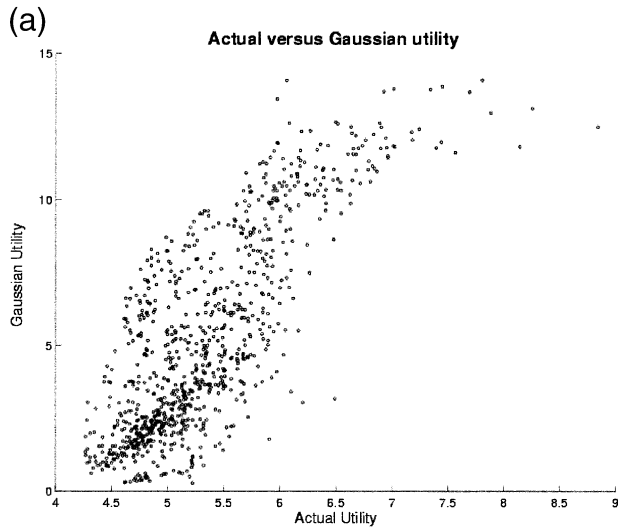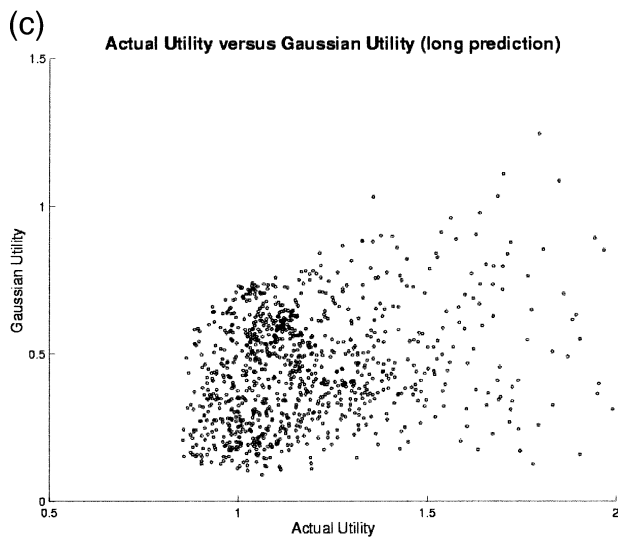
(b)

Utility versus ensemble spread (t=4000)

FIG. 12. (a) The calculated utility vs ensemble spread for the Lorenz system at $t = 2000$. (b) The same as (a) but at $t = 4000$. It is worth comparing this figure with Fig. 7.

perspective as this gives a concrete shape to this rather abstract measure. Thus knowledge of the state variables of a dynamical system before a prediction is made can come from many sources. Climatological (equilibrium) information is the prior knowledge we have chosen to emphasize in this contribution (this is described by the q distribution discussed previously) as this is typically what is available in most practical situations. It may be however that there are other situations where different prior information might be available (such as when only a limited amount of historical data is available) and then a different q would be appropriate reflecting this different prior knowledge. The utility measure gives the

←

FIG. 11. (a) The calculated utility vs Gaussian utility for the Lorenz system at $t = 2000$. (b) The same as (a) but for Gaussian dispersion. (c) The same as (a) but for $t = 8000$.

precise amount of additional information (measured in bits) provided by the prediction over that available to an observer before the prediction was made. If no prior information was available, then the relative entropy reduces to (minus) the usual absolute entropy which then effectively measures the uncertainty in the prediction since obviously the mean of the prediction distribution in this case has no intrinsic value since it cannot be compared with anything. In high-dimensional systems such as the atmosphere, specification of the climatological distribution may prove challenging; however, the present formalism allows for this situation as q simply represents what prior knowledge is available.

An explicit analytical expression for utility is possible in the case that both the prediction and climatological ensembles are Gaussian. This expression involves *both* the first (mean) and second (covariance) moments of the prediction ensemble. Such a result is hardly surprising given the distance interpretation of relative entropy and shows that this measure is different from the often considered potential predictability, which involves only the second moments of the prediction ensemble. Analytical expressions are also no doubt possible for other fixed non-Gaussian distributions but we defer such analysis to a future publication. For Gaussian distributions a very convenient separation of utility into signal and dispersion components is possible. The former is simply a function of the mean vector of the prediction ensemble whereas the latter is only a function of the prediction ensemble covariances. In previous approaches the signal contribution to the "predictability" of a system has tended to be overlooked. Here the relation between these two important contributors to utility is made transparent.

A concrete situation where the utility defined here is a useful measure (as opposed to previously proposed measures) can be found in ENSO prediction. Here ensemble dispersion often does not vary much from one prediction to another whereas the amplitude of the dominant ENSO oscillation can vary significantly in different initial conditions (compare the 1980s with the late 1970s or the early 1990s). In the Gaussian context this means that the signal term will significantly contribute to the usefulness of the prediction. The present formalism enables us to take into account this effect while retaining a measure of the usefulness that comes from a reduction in uncertainty. The generality of the approach as well as its clear formulation in terms of information thus makes relative entropy a very attractive measure of predictability.

In some rough sense, this separation of utility into signal and dispersion mirrors the two forecast statistics (anomaly correlation and rms error) often used to evaluate the practical skill of both weather and climate predictions. Kleeman and Moore (1999) showed that anomaly correlation for a perfect model requires the first moments of the prediction ensemble while evidently rms error is simply a function of the second moments.

The behavior of prediction utility is shown to strongly depend on the nature of the dynamical system under consideration. In stochastic models that serve as analogs for important climatic dynamical systems (e.g., ENSO) it is demonstrated that the signal component is often more important than the dispersion component, a result often not appreciated in analyzing climate predictability. These models are very linear in nature (the ENSO coupled model is weakly nonlinear in an amplitude limiting sense) and it will be interesting to see if the conclusions regarding signal hold for more nonlinear stochastic systems (see below).

In simple models that might be considered analogs for weather prediction such as the Lorenz system, the description of prediction utility appears complex and of a quite different character to the stochastic climate models. For short prediction lags it appears that there is a reasonable relation between Gaussian dispersion and utility whereas the Gaussian signal term is not very well related. There is also some relation with the conventional ensemble spread statistic although it is not as clear as the dispersion relation. For longer lags there appear no good relationships between utility and Gaussian terms or ensemble spread. On the other hand, utility is seen to be a strong function of the position of initial conditions on the attractor of this dynamical system and this relationship is consistent right throughout the predictions (useful predictions at a particular lag are useful at all other lags and conversely). In other words, useful and less useful predictions at all lags tend to come from the same regions of the attractor. Such a robust result suggests that considerably more analysis of predictability for such systems and their generalizations to higher-order systems should be a high priority. There are clear potential benefits to ensemble weather prediction in a better understanding of these kinds of behavior. Systems exhibiting a more stochastic and Gaussian (as opposed to chaotic) behavior have often been advocated as models for the weather dynamical system (see, e.g., Carnevale and Fredriksen 1987; Majda and Timofeyev 2000) and clearly such systems deserve further investigation using the present formalism. All this work is presently under way and will be reported on elsewhere; however, preliminary results suggest that the signal is more important than dispersion in such systems (see Kleeman et al. 2001, submitted to *Physica D*).

The measure introduced here can be compared with that recently advocated by Schneider and Griffies (1999). Their measure is the arithmetic difference in the absolute entropy of the prediction and climatological (or prior) distributions. It therefore measures the reduction in uncertainty of the prediction state vector over that of the climatological state vector. For Gaussian distributions their measure reduces to the first term in Eq. (5).

Finally it is worth remembering that the approach advocated here is based conceptually on a perfect model approach (see section 2a). The author is currently ex-

tending it to take into account model error using plausible assumptions about the nature of this quantity and this will be reported on elsewhere.

APPENDIX

Defining Entropy on Noninteger Dimensional Attractors

Here we derive a practical method for estimating entropy for dynamical systems whose equilibrium (climatological) probability distributions are strange attractors and hence have noninteger dimensionality. Let us assume that we have M data points available on the equilibrium manifold and let us define a finite resolution entropy as follows:

$$E_M(r) = -\frac{1}{M} \sum_{i=1}^{M} \ln\left[\frac{N_i(r)}{Mr^d}\right], \tag{A1}$$

where $N_i(r)$ is the number of data points within a Euclidean distance r of data point i on the manifold and d is the information dimension of the attractor (see Nayfeh and Balachandran 1995), which is defined as

$$d = \lim_{\substack{M \to \infty \\ r \to 0}} \frac{-\frac{1}{M} \sum_{i=1}^{M} \ln\left[\frac{N_i(r)}{M}\right]}{\ln(r)}.$$

We now define the entropy for all scales as

$$E = \lim_{\substack{M \to \infty \\ r \to 0}} E_M(r).$$

This definition may be compared with that used traditionally for integer dimension attractors, namely,

$$E \equiv -\int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 \cdots \int_{-\infty}^{\infty} dx_n \, p \, \ln p,$$

where p is the probability density function. The probability of points from the dynamical system being within a sphere of radius r at a point i on the manifold one may evidently estimate as $[N_i(r)]/M$ where the estimate becomes precise in the limit that $M \to \infty$. Further the probability *density* at point i is evidently

$$p_i = \lim_{\substack{M \to \infty \\ r \to 0}} \frac{N_i(r)}{M\alpha(n)r^n},$$

where $\alpha(n)$ is a constant depending only on the dimension $n(\pi, 4\pi, (4/3)\pi, \ldots)$. Given that spheres in (A1) are implicitly weighted according to their likelihood on the attractor, that is, by the probability function $p_i\alpha(n)r^n$, it follows in a straightforward manner that the two definitions are identical up to a constant that depends only on n. Absolute entropy is only definable up to a constant in any case (see Cover and Thomas 1991) so our definition agrees adequately with the usual one in the case of integer dimension. The usual method (Nayfeh and Balachandran 1995, section 7.9) of calculating the information dimension also serves as a method for calculating the entropy: one calculates $S(r, M) = -1/M \sum_{i=1}^{M} \ln[N_i(r)/M]$ and plots this against $\ln(r)$ for successively smaller values of r. For r sufficiently small this relation is linear and since the definition (A1) implies that

$$E_M(r) = S(r, M) + d \ln(r),$$

it follows that the intercept of this linear relation with the $\ln(r) = 0$ axis of the plot is a good estimate of E. Clearly the above method could be extended to a definition of relative entropy as well providing the dimensionality of both the prediction ensemble and the equilibrium ensemble are estimated. As was noted above, however, this can pose practical problems.

REFERENCES

Badii, R., and A. Politi, 1997: *Complexity: Hierarchical Structures and Scaling in Physics.* Cambridge Nonlinear Science Series, Vol. 6, Cambridge University Press, 318 pp.

Carnevale, G. F., and G. Holloway, 1982: Information decay and the predictability of turbulent flows. *J. Fluid Mech., 116,* 115–121.

——, and J. Frederiksen, 1987: Nonlinear stability and statistical mechanics of flow over topography. *J. Fluid Mech., 175,* 157–181.

Chen, Y.-Q., D. S. Battisti, T. N. Palmer, J. Barsugli, and E. S. Sarachik, 1997: A study of the predictability of tropical Pacific SST in a coupled atmosphere–ocean model using singular vector analysis: The role of the annual cycle and the ENSO cycle. *Mon. Wea. Rev., 125,* 831–845.

Cover, T. M., and J. A. Thomas, 1991: *Elements of Information Theory.* Wiley, 576 pp.

Eckmann, J.-P., and D. Ruelle, 1985: Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys., 57,* 617–656.

Egger, J., and H. D. Schilling, 1984: Predictability of atmospheric low-frequency motions. *Predictability of Fluid Motions,* G. Holloway and B. J. West, Eds., AIP-Conference Proceedings, Vol. 106, American Institute of Physics, 149–158.

Gardiner, C. W., 1985: *Handbook of Stochastic Methods, for Physics, Chemistry, and the Natural Sciences.* Springer-Verlag, 442 pp.

Grassberger, P., 1983: Generalized dimensions of strange attractors. *Phys. Lett. A, 97,* 227–230.

Ji, M., A. Kumar, and A. Leetmaa, 1994: A multiseason climate forecast system at the National Meteorological Center. *Bull. Amer. Meteor. Soc., 75,* 569–577.

Kestin, T. S., D. J. Karoly, J.-I. Yano, and N. A. Rayner, 1998: Time frequency variability of ENSO and stochastic simulations. *J. Climate, 11,* 2258–2272.

Kirtman, B. P., and J. Shukla, 1998: Current status of ENSO forecast skill: A report to the Climate Variability and Predictability (CLIVAR) Numerical Experimental Group. Lamont-Doherty Earth Observatory.

Kleeman, R., and A. M. Moore, 1997: A theory for the limitation of ENSO predictability due to stochastic atmospheric transients. *J. Atmos. Sci., 54,* 753–767.

——, and ——, 1999: A new method for determining the reliability of dynamical ENSO predictions. *Mon. Wea. Rev.,* **127,** 694–705.

——, ——, and N. R. Smith, 1995: Assimilation of subsurface thermal data into a simple ocean model for the initialization of an intermediate tropical coupled ocean–atmosphere forecast model. *Mon. Wea. Rev.,* **123,** 3103–3114.

Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.,* **102,** 409–418.

Lorenz, E. N., 1963: Deterministic non-periodic flow. *J. Atmos. Sci.,* **20,** 130–141.

Madden, R. A., 1981: A quantitative approach to long-range prediction. *J. Geophys. Res.,* **86,** 9817–9825.

Majda, A. J., and I. Timofeyev, 2000: Remarkable statistical behavior for truncated Burgers–Hopf dynamics. *Proc. Natl. Acad. Sci.,* **97,** 12 413–12 417.

Moore, A. M., and R. Kleeman, 1997: The singular vectors of a coupled ocean–atmosphere model of ENSO. Part I: Thermodynamics, energetics and error growth. *Quart. J. Roy. Meteor. Soc.,* **123,** 953–981.

——, and ——, 1998: Skill assessment for ENSO using ensemble prediction. *Quart. J. Roy. Meteor. Soc.,* **124,** 557–584.

——, and ——, 1999: Stochastic forcing of ENSO by the intraseasonal oscillation. *J. Climate,* **12,** 1199–1220.

Nayfeh, A. H., and B. Balachandran, 1995: *Applied Nonlinear Dynamics: Analytical, Computational, and Experimental Methods.* Wiley, 685 pp.

Palmer, T. N., 1993: Extended-range atmospheric prediction and the Lorenz model. *Bull. Amer. Meteor. Soc.,* **74,** 49–65.

——, 2000: Predicting uncertainty in forecasts of weather and climate. *Rep. Prog. Phys.,* **63,** 71–116.

——, F. Molteni, R. Mureau, R. Buizza, P. Chapelet, and J. Tribbia, 1993: Ensemble prediction. *Proc. Seminar on Validation of Models over Europe,* Vol. 1, Shinfield Park, Reading, United Kingdom, European Centre for Medium-Range Weather Forecasts, 21–66.

Penland, C., and P. D. Sardeshmukh, 1995: The optimal growth of tropical sea surface temperature anomalies. *J. Climate,* **8,** 1999–2024.

Priebe, C. E., 1994: Adaptive mixtures. *J. Amer. Stat. Assoc.,* **89,** 796–806.

Ruelle, D., and F. Takens, 1971: On the nature of turbulence. *Comm. Math. Phys.,* **20,** 167–192.

Schneider, T., and S. M. Griffies, 1999: A conceptual framework for predictability studies. *J. Climate,* **12,** 3133–3155.

Shukla, J., 1998: Predictability in the midst of chaos: A scientific basis for climate forecasting. *Science,* **282,** 728–731.

Smith, L. A., 1996: Accountability and error in non-linear forecasting, in 1995. *Proc. Seminar on Predictability,* Vol. 1, Shinfield Park, Reading, United Kingdom, European Centre for Medium-Range Weather Forecasts, 351–368.

——, C. Ziehmann, and K. Fraedrich, 1999: Uncertainty dynamics and predictability in chaotic systems. *Quart. J. Roy. Meteor. Soc.,* **125,** 2855–2886.

Thompson, C. J., and D. S. Battisti, 2000: A linear stochastic dynamical model of ENSO. Part I: Development. *J. Climate,* **13,** 2818–2832.

Toth, Z., and E. Kalnay, 1993: Operational ensemble prediction at the National Meteorological Center: Practical aspects. *Bull. Amer. Meteor. Soc.,* **74,** 2317–2330.

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