Chin. Ann. Math.
DOI: 10.1007/s11401-009-0108-3

Chinese Annals of Mathematics, Series B

Quantifying Dynamical Predictability: the Pseudo-Ensemble Approach****

Jianbo GAO* Wenwen TUNG** Jing HU***

(*Dedicated to Professor Andrew Majda on the Occasion of his* 60*th Birthday*)

Abstract The ensemble technique has been widely used in numerical weather prediction and extended-range forecasting. Current approaches to evaluate the predictability using the ensemble technique can be divided into two major groups. One is dynamical, including generating Lyapunov vectors, bred vectors, and singular vectors, sampling the fastest error-growing directions of the phase space, and examining the dependence of prediction efficiency on ensemble size. The other is statistical, including distributional analysis and quantifying prediction utility by the Shannon entropy and the relative entropy. Currently, with simple models, one could choose as many ensembles as possible, with each ensemble containing a large number of members. When the forecast models become increasingly complicated, however, one would only be able to afford a small number of ensembles, each with limited number of members, thus sacrificing estimation accuracy of the forecast errors.

To uncover connections between different information theoretic approaches and between dynamical and statistical approaches, we propose an (ϵ, τ)-entropy and scale-dependent Lyapunov exponent — based general theoretical framework to quantify information loss in ensemble forecasting. More importantly, to tremendously expedite computations, reduce data storage, and improve forecasting accuracy, we propose a technique for constructing a large number of "pseudo" ensembles from one single solution or scalar dataset. This pseudo-ensemble technique appears to be applicable under rather general conditions, one important situation being that observational data are available but the exact dynamical model is unknown.

Keywords Dynamical predictability, Ensemble forecasting, Relative entropy, Kolmogorov entropy, Scale-dependent Lyapunov exponent
2000 MR Subject Classification 37, 60, 86

1 Introduction

One may consider atmospheric phenomena over a wide range of spatial and temporal scales as dynamical systems in which their future states evolve from their present states according to physical laws or mathematical rules. The dynamical predictability of the atmosphere is defined as the extents to which these systems can be predicted if an optimum procedure is used. Since the pioneering works of Thompson [46] and Lorenz [26–28], it has been widely recognized that in order to quantify the sensitive dependence of a prediction on initial errors, model imperfections,

Manuscript received April 1, 2009. Published online August 10, 2009.
*PMB Intelligence LLC, PO Box 2077, West Lafayette, IN 47996, USA. E-mail: jbgao@pmbintelligence.com
**Corresponding author. Purdue Climate Change Research Center, Department of Earth and Atmospheric Sciences, Purdue University, 550 Stadium Mall Dr., West Lafayette, IN 47907-2051, USA. E-mail: wwtung@purdue.edu
***PMB Intelligence LLC, PO Box 2077, West Lafayette, IN 47996, USA; Affymetrix, Inc., 3380 Central Expressway, Santa Clara, CA 95051, USA. E-mail: jhu@pmbintelligence.com
****Project supported by the National Science Foundation (Nos. CMMI-0825311, CMMI-0826119).

and external forcing, a probabilistic framework is required to describe the dynamical predictive utility of a forecast. Along with this recognition, ensemble forecasting has been developed and widely used (see [24] for a review of the history).

Fundamental research works on ensemble forecasting so far have followed two distinct paths. One is dynamical, focusing on constructing suitable forecast ensembles based on dynamical systems theory to provide as much information about the distribution of forecasts as possible. The other is statistical, focusing on quantifying the predictive utility using information theory. Along line (i), two efficient methods to generate ensembles have been proposed. One is local Lyapunov vectors (see [19, 47]) and bred vectors (see [19]), the other is singular vectors (see [38, 39]). Furthermore, Ehrendorfer and Tribbia [7] have shown that to correctly reconstruct the growth of forecast error, the fastest error-growing directions of the phase space have to be sampled, while Buizza and Palmer [3] have studied the dependence of prediction efficiency on ensemble size. Along line (ii), Shannon entropy (see [5, 25, 42]) and recently, relative entropy (see [1, 4, 5, 17, 20–22, 25, 32, 40, 42]) as means to quantify predictive utilities have been proposed.

Currently, with simple models, one would choose as many ensembles as possible, with each ensemble containing a large number of members. For example, with the Lorenz [26] model, Kleeman [20] constructed 10^3 ensembles, each containing 10^5 members. When the forecast models become increasingly complicated, however, one would only be able to afford a small number of ensembles, each with limited number of members, thus sacrificing estimation accuracy of the forecast errors.

In this paper, we consider three important problems. (1) Are there any fundamental connections between different information theoretic approaches to ensemble forecasting? (2) Can the dynamical and the statistical approaches to ensemble forecasting be connected and even unified? (3) Can dynamical predictability be quantified by constructing a large number of "pseudo" ensembles from one single solution or an observational dataset of an atmospheric phenomenon, such that the pseudo-ensemble technique requires significantly less computational time and data storage than the standard ensemble technique? To answer these questions, we propose a general theoretical framework based on the (ϵ, τ) entropy and the scale-dependent Lyapunov exponent to quantify information loss in ensemble forecasting.

The remainder of the paper is organized as follows. In Section 2, we consider general information flow in a dynamical system by introducing to the readers the concepts of phase-space, dynamical systems, and an entourage of entropy definitions. In Section 3, we discuss how in principle the quantification of information loss in the context of ensemble forecasting can be achieved with the Shannon entropy and relative entropy. Building on top of these two sections, in Section 4, we discuss in practice how the quantification of dynamical predictability can be achieved using the concept of scale-dependent Lyapunov exponent. In Section 5, we propose a pseudo-ensemble technique for quantifying dynamical predictability. In Section 6, we make concluding discussions to sort out when and where the pseudo-ensemble technique may be applicable.

2 Kolmogorov-Sinai Entropy and Information Flow in Dynamical Systems

2.1 Dynamical systems in the phase space

Entropy characterizes the rate of creation of new information (or loss of prior knowledge)

in a dynamical system. To set up a proper stage to consider information flow in a dynamical system, it is important to first briefly review the concept of phase space. For simplicity, let us consider a system with two state variables, X_1 and X_2. When monitoring the motion of the system in time, one can plot out the waveforms for $X_1(t)$ and $X_2(t)$. Alternatively, one can monitor the trajectory defined by $(X_1(t), X_2(t))$, where the time, t, appears as an implicit parameter. The space spanned by X_1 and X_2 is called the phase space (or state space). The dimension of the phase space is the number of degrees of freedom of the system. When the system is modeled by partial differential equations (PDEs) or stochastic partial (or ordinary) differential equations, the dimension of the phase space is infinite. However, the dimension of the attractor of the dynamics could still be quite low, even if the phase space is infinite-dimensional. For example, a stable fixed point solution has a dimension of zero, while a limit cycle solution has a dimension of one. In the following, when we refer to a high-dimensional dynamical system, we shall not only mean that the dimension of the phase space is high, but also that the dimension of the attractor is high. Therefore, a system modeled by PDEs is not considered a high-dimensional system if its dynamics is simple.

2.2 Entropy of a dynamical system

To calculate the entropy of the motion of a dynamical system, we partition the phase space into small boxes of size ϵ, compute the probability p_i that box i is visited by the trajectory, and finally calculate the Shannon entropy $I = -\sum p_i \ln p_i$. For many systems, when $\epsilon \to 0$, information linearly increases with time (see [2]):

$$I(\epsilon, t) = I_0 + Kt, \tag{2.1}$$

where I_0 is the initial entropy and K is the Kolmogorov-Sinai (K-S) entropy (to be more precisely defined shortly). Suppose that the system is initially in a particular region of the phase space, and that all initial probabilities are zero except the corresponding probability for that region, which is 1. Therefore, $I_0 = 0$.

We now consider three cases of dynamical systems: (i) deterministic, nonchaotic, (ii) deterministic, chaotic, and (iii) random. For case (i), during the time evolution of the system, phase trajectories remain close together. After a time T, nearby phase points are still close to each other, and can be grouped into some other small region of the phase space. Therefore, there is no change in information. For case (ii), due to exponential divergence, the number of phase space region available to the system after a time T is $N \propto \mathrm{e}^{(\sum \lambda^+)T}$, where λ^+ are positive Lyapunov exponents. Assuming that all of these regions are equally likely, we have $p_i(T) \sim \frac{1}{N}$, and the information function becomes

$$I(T) = -\sum_{i=1}^{N} p_i(T) \ln p_i(T) = \Big(\sum \lambda^+\Big)T. \tag{2.2}$$

Therefore, $K = \sum \lambda^+$. More generally, if these phase space regions are not visited with equal probability, then

$$K \le \sum \lambda^+. \tag{2.3}$$

Grassberger and Procaccia [15], however, suggested that the equality usually holds. Finally, for case (iii), we can easily envision that after a short time, the entire phase space may be visited. Therefore, $I \sim \ln N$. When $N \to \infty$, we have $K = \infty$.

We now define the K-S entropy precisely. Consider a dynamical system with F degrees of freedom. Suppose that the F-dimensional phase space is partitioned into boxes of size ϵ^F. Suppose that there is an attractor in phase space and consider a transient-free trajectory $\vec{x}(t)$ (i.e., one that has evolved onto the attractor). The state of the system is now measured at intervals of time τ. Let $p(i_1, i_2, \cdots, i_d)$ be the joint probability that $\vec{x}(t = \tau)$ is in box i_1, $\vec{x}(t = 2\tau)$ is in box i_2, $\cdots$, and $\vec{x}(t = d\tau)$ is in box i_d. The K-S entropy is then (see [15])

$$K = -\lim_{\tau\to 0}\lim_{\epsilon\to 0}\lim_{d\to\infty}\frac{1}{d\tau}\sum_{i_1,\cdots,i_d} p(i_1,\cdots,i_d)\ln p(i_1,\cdots,i_d). \tag{2.4}$$

Alternatively, we may first introduce the block entropy,

$$H_d(\epsilon,\tau) = -\sum_{i_1,\cdots,i_d} p(i_1,\cdots,i_d)\ln p(i_1,\cdots,i_d). \tag{2.5}$$

It is on the order of $d\tau K$; then we take difference between $H_{d+1}(\epsilon,\tau)$ and $H_d(\epsilon,\tau)$ and divide it by τ:

$$h_d(\epsilon,\tau) = \frac{1}{\tau}[H_{d+1}(\epsilon,\tau) - H_d(\epsilon,\tau)]. \tag{2.6}$$

Let

$$h(\epsilon,\tau) = \lim_{d\to\infty} h_d(\epsilon,\tau). \tag{2.7}$$

It is called the (ϵ,τ)-entropy (see [14]). The K-S entropy can also be obtained by taking proper limits in equation (2.7):

$$K = \lim_{\tau\to 0}\lim_{\epsilon\to 0} h(\epsilon,\tau) = \lim_{\tau\to 0}\lim_{\epsilon\to 0}\lim_{d\to\infty}\frac{1}{\tau}[H_{d+1}(\epsilon,\tau) - H_d(\epsilon,\tau)]. \tag{2.8}$$

In the following, we shall only deal with the situation where τ is the sampling time and is constant. For convenience, $h(\epsilon,\tau)$ is simply written as $h(\epsilon)$. Note that, in ergodic systems, the operation of $d\to\infty$ in equation (2.7) may be replaced by ensemble average.

3 Quantifying Predictability in Ensemble Forecasting by the Shannon and Relative Entropy

3.1 Increase of Shannon entropy in ensemble forecasting

To quantify information loss in ensemble forecasting, the key is to recognize that the size of the initial perturbations used to generate the ensembles is not infinitesimal, but finite. Therefore, their growth rate typically may not be given by the K-S entropy in equation (2.4), but depends continuously on the instantaneous size ϵ of the ensembles during the passage of time. Fortunately, the definition of the (ϵ,τ)-entropy in equation (2.7) makes it clear that in general, the rate of creation of new information I can be written as

$$\frac{\mathrm{d}I(\epsilon_t)}{\mathrm{d}t} = h(\epsilon_t), \tag{3.1}$$

where ϵ is written as ϵ_t to emphasize that, in the case of finite scales, the size of the ensembles is a function of time. Alternatively, we have

$$I(\epsilon_t) = I_0 + \int_0^t h(\epsilon_t)\mathrm{d}t. \tag{3.2}$$

Equation (3.2) may be considered as a generalization of equation (2.1).

3.2 Decay of relative entropy in ensemble forecasting

Next, we consider the decay of the relative entropy in ensemble forecasting. To facilitate coarse-grained partitioning of a phase space critical for defining dynamical entropies of Section 2, we consider discrete probabilities. Let $\{p_i : i = 1, 2, \cdots\}$ be the probabilities associated with the forecast ensembles in the i-th box and $\{q_i\}$ be the invariant measure or the equilibrium or climatological distribution of the ergodic system under consideration (see [20]). The discrepancy between them can be quantified by the relative entropy (see [6])

$$R = \sum_i p_i \ln\Big(\frac{p_i}{q_i}\Big), \tag{3.3}$$

which is always nonnegative and is zero if and only if $p = q$.

We now focus on the evolution of one particular ensemble occupying a space of size ϵ_0 that is around one of the chosen initial conditions, $\vec{x}(i)$. Denote it by Ensemble(i). We can partition the phase space with boxes of size ϵ_0 to include Ensemble(i) in one single box i. With this setting, $p_k = 1$ only when $k = i$ and zero elsewhere. Therefore, initially, for Ensemble(i), the relative entropy is

$$R^{(i)}(t = 0) = -\ln q_i, \tag{3.4}$$

where for clarity, the superscript (i) in the relative entropy is used to indicate that this is for Ensemble(i). At time τ, Ensemble(i) will have evolved into a different region in phase space, with some directions expanded and some others shrunk. If the dynamics is complicated, it will then occupy a number of boxes that are used to partition the phase space. For ease of presentation, we consider two cases of ϵ_0: infinitesimal and finite. Let us first consider the case of infinitesimal ϵ_0. For chaotic systems, assume that the dynamics of the system around $\vec{x}(i)$ have J positive local Lyapunov exponents $\lambda_1^{(i)} \geq \lambda_2^{(i)} \geq \cdots \geq \lambda_J^{(i)} > 0$, where the superscript i is used to indicate that these Lyapunov exponents are local Lyapunov exponents near $\vec{x}(i)$. Edges along the J Lyapunov directions should have expanded. The size of the ensemble now becomes $\epsilon_\tau = \epsilon_0 \mathrm{e}^{\lambda_1^{(i)}\tau}$, where $\lambda_1^{(i)}$ is the largest positive local Lyapunov exponent, and the number of phase space boxes N available to Ensemble(i) becomes $N \propto \mathrm{e}^{\left(\sum_j \lambda_j^{(i)}\right)\tau}$ or equivalently, $\mathrm{e}^{K^{(i)}\tau}$, where $K^{(i)}$ is the summation of the J local positive Lyapunov exponents and may be called local K-S entropy. In the case that the invariant measure is smooth on its support, then so long as ϵ_τ is small, we may assume that the phase space regions available to Ensemble(i) at time τ are equally likely. Indexing the boxes covering these regions by $l = 1, 2, \cdots$, we see that the relative entropy becomes

$$R^{(i)}(t = \tau) = -\sum_l \ln q_{il}^{(\tau)} - K^{(i)}\tau, \tag{3.5}$$

where we have used $q_{il}^{(\tau)}$ to denote the invariant measure in phase space regions available to Ensemble(i) at time τ.

Next, we consider the case of finite ϵ_0. Following the argument leading to equation (3.1), we readily see that equation (3.5) should be replaced by

$$R^{(i)}(t = \tau) = -\sum_l \ln q_{il}^{(\tau)} - h(\epsilon_\tau)^{(i)}\tau, \tag{3.6}$$

where $h(\epsilon_\tau)^{(i)} = h(\epsilon_\tau, \tau)^{(i)}$ is identified as the local (ϵ, τ)-entropy pertinent to Ensemble(i) at time τ.

We now consider a collection of ensembles. Loss of information in each ensemble may be described by equations (3.4)–(3.6). Averaging over all the ensembles, and assuming that initial conditions represent the invariant measure so that both $\langle -\ln q_i \rangle$ and $\left\langle -\sum_l \ln q_{il}^{(\tau)} \right\rangle$ are equal to the maximal prior information $I_{\max}$, available to the system with a partition using boxes of size ϵ_0, we arrive at

$$R(t) = I_{\max} - \int_0^t h(\epsilon_t)\mathrm{d}t. \tag{3.7}$$

Since $\int_0^t h(\epsilon_t)\mathrm{d}t$ is the major term that appears in both equations (3.2) and (3.7), we see that the Shannon entropy and the relative entropy approaches are closely related.

3.3 Scaling laws for $h(\epsilon_t)$

To evaluate equations (3.2) and (3.7), it is crucial to know functional forms for $h(\epsilon_t)$ as well as how ϵ grows with time t. The former was systematically studied by Gaspard and Wang [14]. The major results are summarized here.

(a) For deterministic chaos, when ϵ is small, $h(\epsilon_t)$ is the K-S entropy

$$h(\epsilon_t) \sim K. \tag{3.8}$$

Note that in this case, the term $\int_0^t h(\epsilon_t)\mathrm{d}t$ simply becomes Kt. In particular, when there is only one positive Lyapunov exponent λ_1, then K equals λ_1.

(b) For independent and identically distributed (iid) noise and chaotic signals on scales large enough such that the signals are quasi-independent,

$$h(\epsilon) \sim -D_I \ln \epsilon, \tag{3.9}$$

where D_I is the dimension for noise and the "information dimension" for chaotic signals (see, e.g., [13]).

(c) For random $\frac{1}{f^{2H+1}}$ processes, where $0 < H < 1$ is called the Hurst parameter which characterizes the correlation structure of the process: depending on whether H is smaller than, equal to, or larger than $\frac{1}{2}$, the process is said to have anti-persistent, short-range, or persistent long-range correlations (see [12, 13])

$$h(\epsilon) \sim \epsilon^{-\frac{1}{H}}. \tag{3.10}$$

Note that the famous Kolmogorov's $-\frac{5}{3}$ turbulence energy spectrum (see [23]) corresponds to $H = \frac{1}{3}$.

From a finite dataset, it is difficult to compute $h(\epsilon)$ accurately (see [14]). Fortunately, scaling laws similar to equations (3.8)–(3.10) recently have been found for the scale-dependent Lyapunov exponent (SDLE) (see [11, 13]), which can be readily computed from short datasets. The functional forms for $h(\epsilon)$ can therefore be inferred from the functional forms for the SDLE. Furthermore, the SDLE can be used to determine how ϵ grows with time. Before we discuss the SDLE in detail in Section 4, we illustrate the main results discussed so far by three examples.

3.4 Case studies

(1) Logistic map

We first study the logistic map

$$x_{n+1} = 4x_n(1 - x_n). \tag{3.11}$$

By a simple transformation (see [36])

$$x = \frac{1}{2}[1 - \cos(\pi y)],$$

one can readily show that this map is equivalent to the tent map

$$y_{n+1} = 1 - 2\left|y_n - \frac{1}{2}\right|. \tag{3.12}$$

Therefore, the Lyapunov exponent is $\ln 2 \approx 0.693$. The invariant measure for the logistic map is given by (see [36])

$$q^*(x) = \frac{1}{\pi\sqrt{x(1-x)}}. \tag{3.13}$$

It is non-Gaussian and has singularities at $x = 0$ and 1. See Figure 1 the solid black curve.

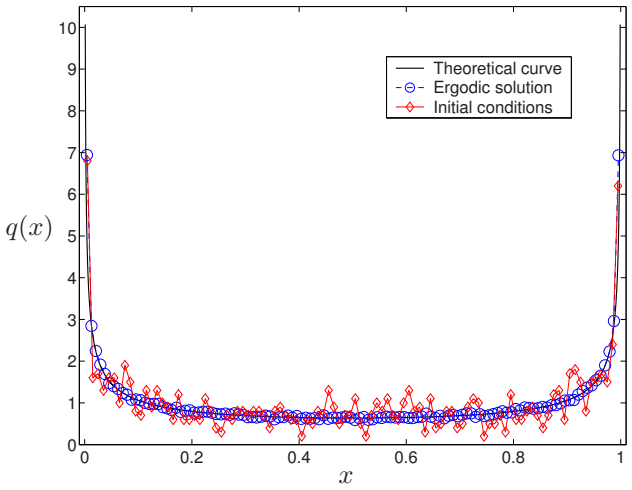

Figure 1 Density functions for the logistic map. The solid black curve is generated by equation (3.13). The blue curve designated by circle is estimated from an ergodic solution consisting of 10^5 points. The red curve designated by diamond is estimated from 1000 initial conditions chosen for constructing ensembles.

We have generated a transient-free ergodic solution consisting of 10^5 points for the logistic map. The probability density of these points is plotted in Figure 1 as the blue curve designated by circles. We then have chosen 1000 initial conditions as centers of ensembles. The density estimated from these 1000 points is shown in Figure 1 as the red curve designated by diamonds. Although with strong resemblance, its deviation from the true invariant measure (i.e., the black curve in Figure 1) is still discernible. Such minor deviation has interesting consequences, as will be discussed shortly. Around each chosen initial condition, we have constructed an ensemble containing a large number of members, according to a normal distribution with its standard deviation denoted by r_1. Let r_2 denote the size of intervals used to partition the phase space. Figure 2 shows the decay of the relative entropy with time for a number of combinations of r_1 and r_2, where for convenience of comparison, a dashed straight line with slope equal to the Lyapunov exponent is also plotted. From Figure 2(a) we observe that the decay of the relative entropy is indeed governed by the Lyapunov exponent, or equivalently, the K-S entropy. It is

interesting to note two types of "abnormal behaviors", as shown in Figure 2(b). (i) When r_1 and r_2 are both as small as 10^{-4}, and the number of members in each ensemble is only 10^4, the relative entropy does not decay to 0, as shown by the curve designated as circles in Figure 2(b). This is because each interval, on average, only contains 1 point, and hence, the estimation of density has considerable uncertainty. This feature can be eliminated if the number of members for each ensemble is increased to, for example, 10^5, as shown by the curve designated by circles in Figure 2(a). (ii) When $r_1 \ll r_2$, the relative entropy may first increase instead of decrease with time, as shown by the curve designated by pentagons in Figure 2(b). This is because when the evolution time t is small, each ensemble, even though it is spreading, will not only stay within an interval used for partitioning the phase space, but also accumulate around the two singular points, $x = 0$ and 1. The latter amounts to readjusting according to the invariant measure. The net result is then a fairly constant or even slightly increasing relative entropy for small evolution time.

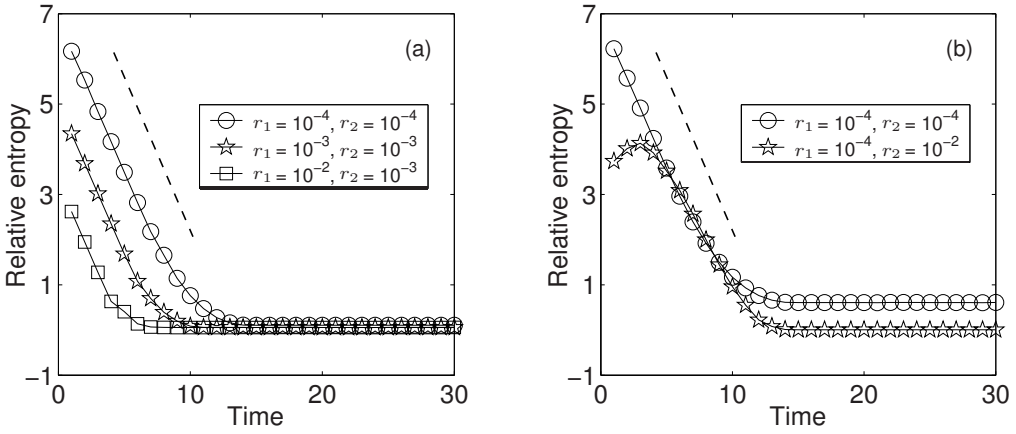

Figure 2 Decay of the relative entropy with time for the logistic map. The parameter r_1 denotes the standard deviation of the normally distributed ensemble centered around the chosen initial conditions, and r_2 denotes size of boxes used for partitioning the phase space. The slope for the dashed line equals the Lyapunov exponent. The case designated by circles in (a) contains 10^5 members for each ensemble, while all other cases contain 10^4 members for each ensemble.

(2) Lorenz '63 system

Next, we examine the Lorenz system with stochastic forcing:

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= -10(x-y) + D\eta_1(t), \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= -xz + 28x - y + D\eta_2(t), \\ \frac{\mathrm{d}z}{\mathrm{d}t} &= xy - \frac{8}{3}z + D\eta_3(t), \end{aligned} \tag{3.14}$$

where $D\eta_i(t)$, $i = 1, 2, 3$ are independent Gaussian noise forcing terms with mean 0 and variance D^2. The system is solved using the scheme of exact propagator (see [33]), where the exact solution of the Lorenz system is solved using a 4-th order Runge-Kutta method with a time-step of $h = 0.002$, and then a term $D\sqrt{h}\,W$, where W is a Gaussian noise of mean 0 and variance 1, is added to the corresponding equations to take into account the noise. In this

subsection, however, we only study the clean system (i.e., $D = 0$). In this case, it has three Lyapunov exponents, one positive, one zero, and the third negative with a large magnitude. Using standard methods for calculating the largest positive Lyapunov exponent λ_1, one readily finds

$$\lambda_1 = 0.905. \tag{3.15}$$

To evaluate the predictive utility, we have generated a transient-free solution consisting of 10^8 points sampled at $\tau = 0.06$. We have constructed two types of ensembles. In the first type, we have partitioned each coordinate of the phase space containing the attractor into 32 bins, thus obtained $32 \times 32 \times 32$ boxes. We then choose about 1500 boxes containing at least 500 points of the attractor as initiators of the ensembles. Within each box, we randomly choose 10^6 points as initial conditions of the members of each ensemble. Finally, we let all the members of the ensembles evolve with time and compute the relative entropy. The variation of the relative entropy with time is shown in Figure 3(a). We observe three interesting features. (a) It takes a considerable period of time for the members of the ensembles to adjust to the stage that $R(t)$ truly decays. (b) When $R(t)$ finally decays, the rate is 0.78, which is slightly smaller than the Lyapunov exponent given by equation (3.15). (c) The curve $R(t)$ is slightly oscillatory. Feature (c) has been observed by many other researchers (see, e.g., [44]). It is caused by the circular motions along either of the scrolls of the Lorenz attractor, and is related to the so-called hidden frequency phenomenon (see [13, 18, 35]). Although it has interesting implications to dynamical predictability, we shall not be further concerned about it in this paper. Instead, we shall focus on understanding features (a) and (b).

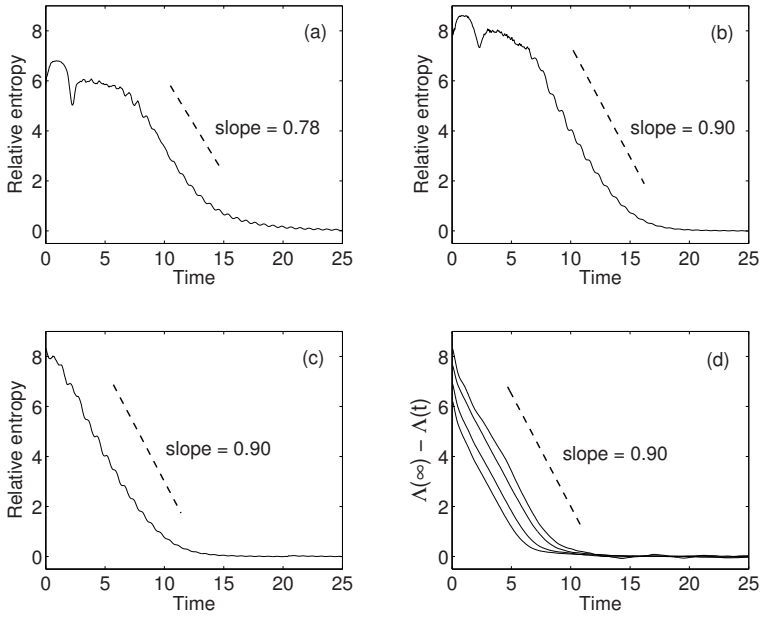

Figure 3 (a)–(c) Decay of the relative entropy with time for the clean Lorenz system. See the text for details about the computation. (d) The variation of $\Lambda(\infty) - \Lambda(t)$ with time. In the computation, a time series of 10^5 points was used, with embedding parameters $m = 4$ and $L = 2$. The four curves from bottom to top correspond to shells $(2^{-i-1}, 2^{-i})$, $i = 10, 11, 12$ and 13, respectively.

Let us try to understand feature (b) first. We have found that underestimate of the decay rate of $R(t)$ is due to the fairly large boxes used to partition the phase space. It can be remedied

by using smaller boxes. For example, if one partitions each coordinate of the phase space into 512 bins, then the decay rate of the relative entropy is given by λ_1, as shown in Figure 3(b). Note that in this case there are about 4.0×10^5 non-empty boxes, which is consistent with the capacity dimension D_0 defined by Gao et al. [13],

$$N \sim \epsilon^{-D_0} \tag{3.16}$$

with $D_0 = 2.06$ and $\epsilon = \frac{1}{512}$.

Let us now discuss the origin for the long initial adjustment time to understand feature (a). We have found two major reasons. One is due to the smallness of each ensemble. Having 10^6 members of an ensemble all within one box, the size of the ensemble is effectively much smaller than the box size, giving rise to the effect we have discussed with respect to the curve designated by pentagons in Figure 2(b). The other reason is related to the negative and zero Lyapunov exponents of the Lorenz system. Recall that the members of each ensemble were chosen uniformly within each box. Having a very large magnitude, the negative Lyapunov exponent of the Lorenz system causes the 3-D ensemble to rapidly collapse onto a plane-like structure, because the dimension of the Lorenz system is about 2.06. One direction of the 3-D ensemble as well as one direction of the plane-like structure is along the tangential direction of the flow, contributing to a Lyapunov exponent of zero magnitude (see [10, 16]). Both the negative and the zero Lyapunov exponents surely contribute significantly to the long initial adjustment time.

To shorten the initial adjustment time, a number of techniques may be used. For example, one could construct Lyapunov vectors or singular vectors (see, e.g., [19]). Here, we propose a simpler solution by constructing the following type of ensemble. Let $\vec{x}_i$ denote the center of Ensemble(i). Using part of the transient-free solution, we find the nearest neighbor of $\vec{x}_i$. Denote it by $\vec{x}_j$. To ensure that $\vec{x}_j$ is not along the tangential direction of $\vec{x}_i$, we require (see [9])

$$|i - j| > W, \tag{3.17}$$

where W is certain threshold value (determination of which from time series data will be discussed in Section 5.1). We approximate the most unstable direction along the trajectory near $\vec{x}_i$ by $\vec{x}_i - \vec{x}_j$, and randomly choose 10^6 points along this direction according to a Gaussian distribution, with the condition that the mean of the Gaussian distribution is at $\vec{x}_i$ and that the standard deviation of the Gaussian distribution is comparable to the size of the boxes used to partition the phase space. A case with box size of 2^{-9} is shown in Figure 3(c). Now we observe that the long initial adjustment period has gone and $R(t)$ decays with a rate almost ideally equal to λ_1.

To summarize, in quantifying predictability of the Lorenz '63 model, we have observed three distinct time scales. The first one characterizes how fast the constructed ensembles align onto the most unstable direction of the phase space. It is evident in Figures 3(a) and (b). The second time scale corresponds to the linear decay of the relative entropy, with the decay rate governed by the K-S entropy (or λ_1, in the case there is only one positive Lyapunov exponent). It is evident in Figures 3(b) and (c). The third time scale corresponds to the slow decay of the relative entropy to zero. It is evident in Figures 3(a)–(c).

(3) Lorenz '96 system

The Lorenz '96 model (see [29–31]) is supposed to represent some atmospheric quantity

equally spaced around a latitudinal circle. It is described by the following equations:

$$\frac{\mathrm{d}X_n}{\mathrm{d}t} = (X_{n+1} - X_{n-2})X_{n-1} - X_n + F, \quad n = 1, 2, \cdots, N, \tag{3.18}$$

with periodic boundary condition, i.e., $X_{-1} = X_{N-1}$, $X_0 = X_N$ and $X_{N+1} = X_1$. F is a positive constant representing external forcing. The term $-X_n$ represents dissipation, and the term $(X_{n+1} - X_{n-2})X_{n-1}$ represents advection. When $N = 40$, $F = 8$, the system has 13 positive Lyapunov exponents and a Kaplan-Yorke dimension of 27.1 (see [31]).

Now suppose that we partition the phase space containing the chaotic attractor into $r \times r \times \cdots \times r$ boxes. If a trajectory is long enough, then using equation (3.16) with $\epsilon = \frac{1}{r}$ and estimating D_0 by the Kaplan-Yorke dimension, we can expect about $r^{27.1}$ non-empty boxes. If those non-empty boxes are to be visited by the trajectory with equal probability, then the probability of visiting each non-empty box would be around $r^{-27.1}$ and the maximal information content is about $27.1 \times \log_2 r$ bits. Choosing a moderate r such as 10 already results in a formidably large number of $10^{27.1}$ non-empty boxes. Therefore, we have to conclude that the relative entropy approach is difficult to resolve small-scale behavior of the system. Consequentially, only the largest of the three time scales we identified for the Lorenz '63 model may be partially resolved by the relative entropy approach. This is the time scale that is most relevant to distributional analysis. Pertinent to this time scale, an interesting framework to effectively calculate the relative entropy from small samples has been proposed by Haven et al. [17]. To uncover the small scale behavior of the system, however, we need other concepts and methods, which are to be developed in the next two sections.

4 Quantifying Predictability by the Scale-Dependent Lyapunov Exponent

4.1 The scale-dependent Lyapunov exponent (SDLE)

The SDLE is defined in a phase space through consideration of an ensemble of trajectories. In the case of a scalar time series $x(1), x(2), \cdots, x(n)$, a suitable phase space may be obtained by using time delay embedding (see [37, 41, 45]) to construct vectors of the form

$$V_i = [x(i), x(i + L), \cdots, x(i + (m - 1)L)], \tag{4.1}$$

where m and L are called the embedding dimension and the delay time, respectively. For chaotic systems, m and L have to be chosen according to certain optimization criterion (see [13]). For a stochastic process, which is infinite-dimensional, the embedding procedure transforms a self-affine stochastic process to a self-similar process in a phase space. Therefore, specific values of m and L are not important, so long as $m > 1$. For example, one could simply choose $m = 2$, $L = 1$ (see [11, 13]).

To define the SDLE, let us denote the initial separation between two nearby trajectories by ϵ_0, and their average separation at time t and $t + \Delta t$ by ϵ_t and $\epsilon_{t+\Delta t}$, respectively. The SDLE $\lambda(\epsilon_t)$ is defined through equations (see [11, 13])

$$\epsilon_{t+\Delta t} = \epsilon_t \mathrm{e}^{\lambda(\epsilon_t)\Delta t} \quad \text{or} \quad \lambda(\epsilon_t) = \frac{\ln \epsilon_{t+\Delta t} - \ln \epsilon_t}{\Delta t}. \tag{4.2}$$

Equivalently, we have a differential equation for ϵ_t:

$$\frac{\mathrm{d}\epsilon_t}{\mathrm{d}t} = \lambda(\epsilon_t)\epsilon_t \quad \text{or} \quad \frac{\mathrm{d}\ln \epsilon_t}{\mathrm{d}t} = \lambda(\epsilon_t). \tag{4.3}$$

Given a discrete time series data, the smallest Δt possible is the sampling time τ.

In our recent work (see [11, 13]), we have found a number of distinctive scalings for the SDLE. Those parallel to equations (3.8)–(3.10) are summarized below.

(a) For deterministic chaos, for small ϵ, the SDLE equals the largest positive Lyapunov exponent λ_1:

$$\lambda(\epsilon_t) = \lambda_1. \tag{4.4}$$

(b) For noise, noisy chaos, and clean chaos on large scales where memory has lost,

$$\lambda(\epsilon_t) \sim -\gamma \ln \epsilon, \tag{4.5}$$

where γ is a parameter.

(c) For $\frac{1}{f^{2H+1}}$ processes, we have a diffusive scaling law:

$$\lambda(\epsilon_t) \sim \epsilon^{-\frac{1}{H}}. \tag{4.6}$$

In our previous studies (see [11, 13]), we illustrated the scaling laws described by equations (4.4) and (4.5) by the stochastic Lorenz '63 system (with parameters different from those used in equation (3.14)) and other chaotic systems, and the scaling law described by equation (4.6) by the prototypical models for $\frac{1}{f^{2H+1}}$ processes, the fractional Brownian motion model and the power-law distributed ON/OFF intermittency. For ease of discussions presented in Section 5, we wish to re-illustrate scaling laws described by equations (4.4) and (4.5), again using the stochastic Lorenz '63 system of equation (3.14). Figure 4 shows five curves, for the cases of $D = 0, 1, 2, 3, 4$. The computations are done by using $m = 4$, $L = 2$, and 10000 points of the x component of the Lorenz system. For the clean system, we observe two scaling laws. One is equation (4.4), $\lambda(\epsilon) \approx 0.9$, for small ϵ; the other is equation (4.5), for large ϵ where memory has been lost. For the noisy system, the scale region where the scaling law of equation (4.4) shrinks when the stochastic forcing is increased. Interestingly, although the part of the curve with $\lambda(\epsilon) \sim -\gamma \ln \epsilon$ shifts to the right when noise is increased, the parameter γ (which is around 1.09 here) appears to not depend on the noise strength.

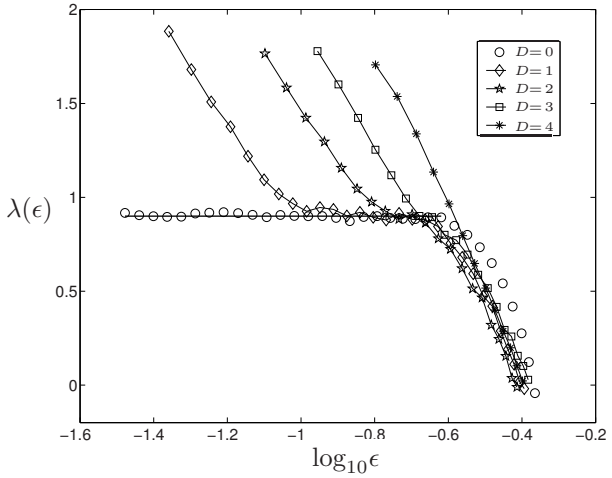

Figure 4 $\lambda(\epsilon)$ curves for the clean and the noisy Lorenz system.

Note that $\lambda(\epsilon)$ may assume different scaling laws on different scale ranges (see [11, 13]), such as chaotic scaling law on small scales and diffusive scaling law on large scales. In situations where directly calculating $h(\epsilon)$ from a given dataset is difficult, it would be reasonable to assume the same functional forms for $h(\epsilon)$ on the corresponding scale ranges. In particular, if a system only has one positive Lyapunov exponent, then $\lambda(\epsilon) = h(\epsilon)$.

4.2 SDLE in ensemble forecasting

We now examine the physical meaning of $\lambda(\epsilon)$ in the context of ensemble forecasting. First, we note that $\frac{1}{\lambda(\epsilon)}$ is closely related to the error doubling time (see [29]). More precisely, using equation (4.3), we have

$$\ln \epsilon_t = \ln \epsilon_0 + \int_0^t \lambda(\epsilon_t) \mathrm{d}t. \tag{4.7}$$

Letting $\epsilon_{T_{db}} = 2\epsilon_0$, we find the error doubling time T_{db} given by

$$\ln 2 = \int_0^{T_{db}} \lambda(\epsilon_t) \mathrm{d}t. \tag{4.8}$$

As the first approximation, we may consider $\frac{1}{\lambda(\epsilon)}$ to be proportional to the doubling time. Mintz [34] for the first time estimated the doubling time to be 5 days in the Mintz-Arakawa two-layer model. With the reduction in the initial uncertainties, it was hoped that the doubling time would increase when model complexity increases and more detailed physical processes are accounted for. However, the opposite has been observed. For example, a recent estimate of the doubling time with the ECMWF model is less than 2 days (see [43]). Lorenz [29] suggests that the major factor for the decrease in the doubling time has been the increase in resolution, introducing into the system finer-scale errors that amplify more rapidly than those at the coarser scales. Although spatial resolution is not equivalent to the scale ϵ discussed here, they are nevertheless closely related. Therefore, the estimates of the doubling times reported so far strongly suggest that scalings described by equations (4.5) and (4.6) are more relevant to reality.

Next, we note that $\mathrm{e}^{\int_0^t \lambda(\epsilon_t)\mathrm{d}t}$ is the very error growth curve most commonly used in ensemble forecasting. This is evident from equation (4.7).

5 The Pseudo-Ensemble Technique to Quantify Predictability

5.1 Flow chart

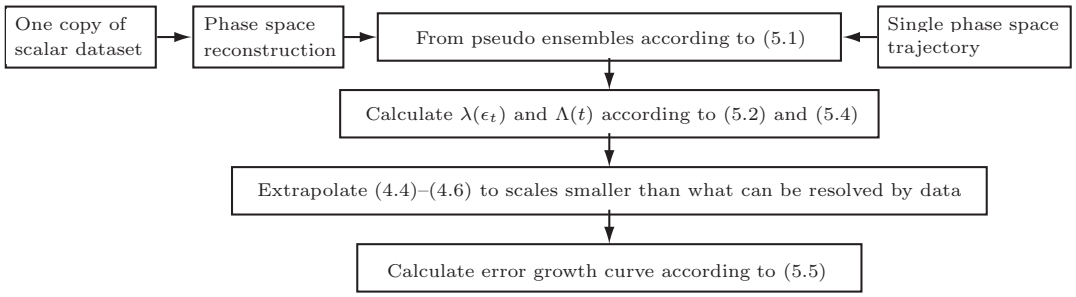

Figure 5 Flow chart for the pseudo-ensemble approach.

We can now explain the flow chart depicted in Figure 5. Assume that a discretized phase-space trajectory, denoted by $\{V_i : i = 1, 2, \cdots\}$, is known. The trajectory may be directly obtained by integrating a model, or obtained by embedding a scalar time series (either from a model output or from an observation) into a suitable phase space using equation (4.1). To compute the SDLE, we first find all the pairs of vectors in the trajectory with their distance being approximately ϵ, and then calculate their average distance after a time Δt. The first half of this description amounts to introducing a "shell" (indexed as k):

$$\epsilon_k \leq \|V_i - V_j\| \leq \epsilon_k + \Delta\epsilon_k, \quad |i - j| > W. \tag{5.1}$$

The parameters ϵ_k (the radius of the shell) and $\Delta\epsilon_k$ (the width of the shell) are arbitrarily chosen small distances. A shell may be considered as a differential element that would facilitate computation of conditional probability. The condition $|i - j| > W$, where W, an integer parameter (which is on the order of the embedding window size $(m-1)L$ for a time series), is used to ensure that the initial separation, $\|V_i - V_j\|$, has aligned with the most unstable direction of the motion (see [11, 13]). To expedite computation, it is advantageous to introduce a sequence of shells, $k = 1, 2, 3, \cdots$. With all these shells, we can monitor the evolution of all the pairs of vectors (V_i, V_j) within a shell and take average. When each shell is very thin, we have

$$\lambda(\epsilon_t) = \frac{\langle \ln \|V_{i+t+\Delta t} - V_{j+t+\Delta t}\| - \ln \|V_{i+t} - V_{j+t}\| \rangle}{\Delta t}, \tag{5.2}$$

where t and Δt are integers in unit of the sampling time, and the angle brackets denote average within a shell. It is clear that

$$\int_0^t \lambda(\epsilon_t) \mathrm{d}t = \langle \ln \|V_{i+t} - V_{j+t}\| - \ln \|V_i - V_j\| \rangle. \tag{5.3}$$

The time-dependent exponent curves corresponding to the sequence of shells are defined as (see [8–10])

$$\Lambda(t) = \langle \ln \|V_{i+t} - V_{j+t}\| - \ln \|V_i - V_j\| \rangle. \tag{5.4}$$

Therefore, equation (5.3) can be written as

$$\int_0^t \lambda(\epsilon_t) \mathrm{d}t = \Lambda(t). \tag{5.5}$$

Letting $t \to \infty$, we recognize that $\Lambda(\infty)$ plays the role of $I_{\max}$ in equation (3.7).

We now explain why the method can be termed pseudo-ensemble approach. For this purpose, let us denote the index i of vector V_i satisfying condition (5.1) by $i_1, i_2, \cdots, i_q$. For each i_l, $l = 1, 2, \cdots, q$, we collect all the indices j of vectors V_j satisfying the same condition. We consider each V_{i_l} as the center of an ensemble, and all the V_j's where $\|V_{i_l} - V_j\|$ satisfies condition (5.1) as the members of the ensemble. Therefore, averaging within a shell amounts to averaging over all the ensembles and the members within them. The method is "pseudo" because all the ensembles are constructed based on one single trajectory.

Since a given finite trajectory or dataset determines a finite minimal scale $\epsilon_{\min} = \min_{i,j} \|V_i - V_j\|$, a critical issue is to determine the behavior of the SDLE on scales smaller than $\epsilon_{\min}$ for

medium- and high-dimensional systems from that finite dataset. Here extrapolation to scales smaller than $\epsilon_{\min}$ using a priori knowledge of the process in the "unresolved" scale is of crucial importance. How this can be achieved will be made clearer after we examine two model systems.

5.2 Case studies

(1) Prediction in the Lorenz '63 model

In Subsection 3.4(2), we considered the predictability of the model using relative entropy. Now we consider it again using the pseudo-ensemble approach. Figure 3(d) shows four $\Lambda(\infty) - \Lambda(t)$ vs. t curves, where the four curves, from bottom to top, correspond to shells $(2^{-i-1}, 2^{-i})$, $i = 10, 11, 12$ and 13, respectively. In our computations, a scalar time series of 10^5 points was used, with embedding parameters $m = 4$ and $L = 2$. While all the curves show a neat straight-line portion with slope 0.9, the curve corresponding to the shell $(2^{-i-1}, 2^{-i})$, $i = 13$, is very similar to the curve shown in Figure 3(c).

To appreciate the amount of computations and data storage the pseudo-ensemble approach has saved, we note that to produce Figure 3(d), only a scalar time series of 10^5 points was used, with CPU time of a PC on the order of 1 minute. In contrast, to obtain each of the Figures 3(a)–(c), about 100 hours of CPU time were used to generate a reference trajectory of 10^8 points, and then 1500 ensembles, each with 10^6 members.

(2) Prediction in the Lorenz '96 model

In Subsection 3.4(3), we considered the predictability of this system and found that the small-scale behavior of the system cannot be readily characterized by the relative entropy approach. We now explain how the difficulty may be overcome by the extrapolation procedure employed in the pseudo-ensemble approach. Below, we shall focus on calculating the error growth curve, the most popular measure in ensemble forecasting.

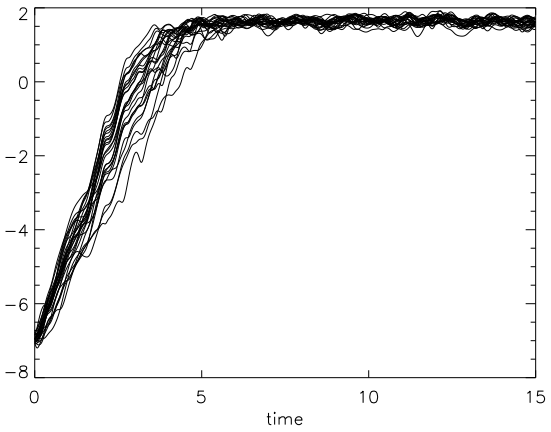

Figure 6 25 error growth curves for the ensemble forecasting of the Lorenz '96 model.

Following a simple setting used by Lorenz [29], we generated an ensemble with 250 members, each being 15 non-dimensional time units long. For each member, we computed an rms error growth curve. To obtain the average error growth for each component, we divided the rms error growth curve by N. To compare with the $\Lambda(t)$ curves of equation (5.4), such curves may

be plotted in semi-logarithmic scale. Figure 6 shows 25 of such curves. As expected, they are quite noisy. The mean of these 250 curves is plotted in Figure 7(a) as the blue curve. We observe that for small time t, the error linearly grows, with the slope being the largest positive Lyapunov exponent. To use the pseudo-ensemble approach, we sample a trajectory of 480 non-dimensional time units with a sampling time of 0.001, and embed the X_1 component to a phase space with $m = 45$ and $L = 2$. The $\Lambda(t)$ curve is shown as the red curve in Figure 7(a), and the corresponding SDLE is shown in Figure 7(b). We observe that the small-scale behavior of the system has been partially resolved. When the time series is longer, of course, the small-scale behavior of the system can be resolved better. A more effective approach, however, is to use extrapolation — using the knowledge that the system is clean, we can simply extend the plateau shown in Figure 7(b) to scales not resolved by the given dataset, and then integrate to get the error growth curve. Since this is a deterministic chaotic system, the extrapolation may be directly performed on the red error growth curve: simply extending the linear part of the red curve in Figure 7(a) would be sufficient. This is shown as the dash red curve in Figure 7(a). Note that it is smoother than the blue line in Figure 7(a), which is the mean of the 250 error growth curves.

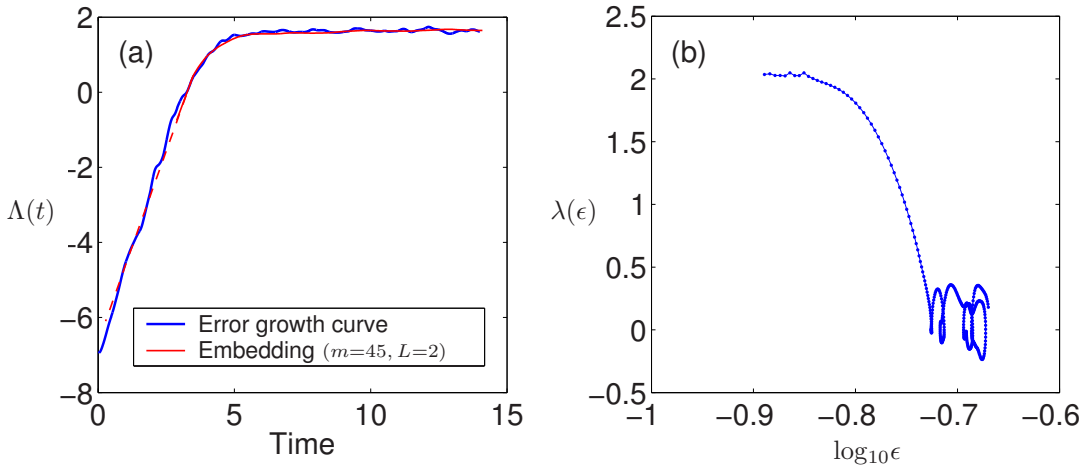

Figure 7 (a) Error growth curves for the Lorenz '96 system: the blue and red curves were obtained using standard and pseudo-ensemble approaches, respectively. See the text for details. (b) $\lambda(\epsilon)$ curve based on X_1 time series of the Lorenz '96 system.

6 Concluding Discussions

To uncover connections between different information theoretic approaches and between dynamical and statistical approaches in ensemble forecasting, and more importantly, to reduce computational complexity and data storage, we have proposed a general theoretical framework to quantify information flow in ensemble forecasting of dynamical systems using the concepts of (ϵ, τ) entropy and the SDLE, and developed a pseudo-ensemble approach. Studies of simple model systems using the relative entropy approach suggest the existence of three distinct time scales. The first one is dynamical and characterizes how fast the constructed ensembles align onto the most unstable direction of the phase space. When the ensembles are carefully constructed, this initial adjustment period of time can be reduced to almost zero. The second time

scale corresponds to the linear decay of the relative entropy, with the decay rate governed by the K-S entropy. The K-S entropy clearly connects the statistical and dynamical approaches in the study of ensemble forecasting. The third time scale corresponds to the slow decay of the relative entropy to zero. It is particularly relevant to distributional analysis. Study of the Lorenz '63 model using the proposed pseudo-ensemble approach indicates that the approach may expedite computations and reduce data storage by more than 4 orders of magnitude. The method is also shown to be effective in quantifying predictability in the Lorenz '96 model, which has 13 positive Lyapunov exponents and a fairly large Kaplan-Yorke dimension of 27.1.

While our analyses of the model systems show that the pseudo-ensemble approach is very promising, it is important to examine under what conditions the approach may be effective. Based on our discussion of the Lorenz '63 model, we have good reason to believe that the pseudo-ensemble approach should work well in low-dimensional chaotic systems. Therefore, we only need to focus on medium- and high-dimensional systems.

First, we consider infinite-dimensional noisy dynamics and $\frac{1}{f}$ processes. The latter are particularly relevant to the many convective and turbulent systems in the atmosphere (see, e.g., [49]). Since the SDLE for those systems is given by equations (4.4) and (4.6) (see [11, 13]) and can be readily estimated using only around 10^4 points, we therefore have good reasons to expect that the approach works well for those systems. In particular, if $\epsilon_{\min}$ is the smallest scale resolvable by data, to know the behavior of the SDLE on scales smaller than $\epsilon_{\min}$, we may be able to do simple extrapolation to those scales. Integrating $\lambda(\epsilon)$ then will allow us to obtain the error growth curve.

Next, we consider medium-dimensional dynamical systems. Here arises a major challenge: given a finite trajectory or a finite dataset, can the scaling of equation (4.4) be resolved? When the size of the dataset is large, we expect that equation (4.4) can be at least partially resolved. In this case, we can again extrapolate to scales smaller than $\epsilon_{\min}$ and get small scale behavior of the SDLE. When the size of the dataset is small, however, the scaling of equation (4.4) may not be resolved. In such a situation, if the model is unknown, then one has no way to truly find out the exact behavior on scales not resolved by the given dataset. However, if the model is known, then one can integrate the model two times to obtain a reference trajectory and a perturbed trajectory to roughly determine whether on small scales the dynamics are deterministic described by equation (4.4) or noisy described by equation (4.5). Integration twice may be called minimal ensemble forecasting. While it may remind one of the bred vector technique, it is actually simpler, because it does not require periodic or quasi-periodic normalization of the perturbed trajectory to ensure that the distance between the reference and the perturbed trajectory does not exceed certain threshold value. Compared with standard ensemble forecasting which involves a large number of ensembles, each with a lot of members, the combined pseudo-ensemble approach and the minimal ensemble forecasting can still tremendously reduce computational complexity and data storage. Our analysis of the Lorenz '96 model suggests that the proposed pseudo-ensemble approach should often be effective in dealing with medium-dimensional dynamical systems.

Finally, we consider complex spatial-temporal systems, which include realistic weather and climate models as special cases. To apply the proposed approach, one way is first to decompose the model output using Empirical Orthogonal Functions (EOFs, see, e.g., [48]), and then to

compute the SDLE for the principal component (PC) time series of EOFs. The key issue here is to find the scaling laws for the SDLE of the PC time series of EOFs. We envision that the SDLE may reveal different prediction time scales from the PC time series corresponding to different EOFs, if the system is truly multiscaled and complex.

To make our discussions complete, we now discuss how extrapolation to small scales can be done. Among the three scaling laws described by equations (4.4)–(4.6), which one should be used? When there is no a priori knowledge, this is purely a matter of preference. For example, some researchers may model certain large scale complex phenomenon by deterministic chaos, while others may model it by stochastic dynamics. Both are valid strategies so long as the phenomenon is well modeled. As for the question: which scaling law of equations (4.4)–(4.6) should be chosen, our opinion is that the best approach is to take into account as much information about the dynamics on a wide-range of scales as possible. One can hope that quite often information from various types of observational data would allow one to determine which scaling law is the best to be used for extrapolation.

The above discussions make it clear that the pseudo-ensemble technique may be applicable under rather general conditions. One particularly important situation is that observational data are available but the exact dynamical model is unknown. We emphasize that in such situations, using a priori knowledge to extrapolate scaling laws of equations (4.4)–(4.6) to scales smaller than $\epsilon_{\min}$ should be considered an integral part of the pseudo-ensemble approach.

Acknowledgements The authors thank Dr. Yinhe Cao for computing Figures 3(a) and (b), Mr. Jing Ai for computing Figure 3(c), and Professor Andrew Majda for making an enlightening comment, which is instrumental for us to improve the manuscript.

References

[1] Abramov, R., Majda, A. and Kleeman, R., Information theory and predictability for low frequency variability, *J. Atmos. Sci.*, **62**(1), 2005, 65–87.

[2] Atmanspacher, H. and Scheingraber, H., A fundamental link between system theory and statistical mechanics, *Found. Phys.*, **17**(9), 1987, 939–963.

[3] Buizza, R. and Palmer, T., Impact of ensemble size on ensemble prediction, *Mon. Weather Rev.*, **126**(9), 1998, 2503–2518.

[4] Cai, D., Haven, K. and Majda, A., Quantifying predictability in a simple model with complex features, *Stoch. Dyn.*, **4**(4), 2004, 547–569.

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

[6] Cover, T. and Thomas, J., Elements of Information Theory, Wiley-Interscience, New York, 1991.

[7] Ehrendorfer, M. and Tribbia, J., Optimal prediction of forecast error covariances through singular vectors, *J. Atmos. Sci.*, **54**(2), 1997, 286–313.

[8] Gao, J. B. and Zheng, Z. M., Local exponential divergence plot and optimal embedding of a chaotic time series, *Phys. Lett. A*, **181**(2), 1993, 153–158.

[9] Gao, J. B. and Zheng, Z. M., Direct dynamical test for deterministic chaos and optimal embedding of a chaotic time series, *Phys. Rev. E*, **49**(5), 1994, 3807–3814.

[10] Gao, J. B., and Zheng, Z. M., Direct dynamical test for deterministic chaos, *Europhys. Lett.*, **25**(7), 1994, 485–490.

[11] Gao, J. B., Hu, J., Tung, W.-W., et al, Distinguishing chaos from noise by scale-dependent Lyapunov exponent, *Phys. Rev. E*, **74**(6), 2006, 066204.

[12] Gao, J. B., Hu, J., Tung, W.-W., et al, Assessment of long range correlation in time series: How to avoid pitfalls, *Phys. Rev. E*, **73**(1), 2006, 016117.

[13] Gao, J. B., Cao, Y. H., Tung, W.-W., et al, Multiscale Analysis of Complex Time Series — Integration of Chaos and Random Fractal Theory, and Beyond, Wiley-Interscience, New York, 2007.

[14] Gaspard, P. and Wang, X. J., Noise, chaos, and (ϵ, τ)-entropy per unit time, *Phys. Rep.*, **235**(6), 1993, 291–343.

[15] Grassberger, P. and Procaccia, I., Estimation of the Kolmogorov entropy from a chaotic signal, *Phys. Rev. A*, **28**(4), 1983, 2591–2593.

[16] Haken, H., At least one Lyapunov exponent vanishes if the trajectory of an attractor does not contain a fixed point, *Phys. Lett. A*, **94**, 1983, 71–72.

[17] Haven, K., Majda, A. and Abramov, R., Quantifying predictability through information theory: Small sample estimation in a non-Gaussian framework, *J. Comput. Phy.*, **206**(1), 2005, 334–362.

[18] Hu, J., Tung, W.-W., Gao, J. B., et al, Uncovering structures in complex time series through scale separation, *Phys. Lett. A*, 2007, to appear.

[19] Kalnay, E., Atmospheric Modeling: Data Assimilation and Predictability, Cambridge University Press, Cambridge, 2003.

[20] Kleeman, R., Measuring dynamical prediction utility using relative entropy, *J. Atmos. Sci.*, **59**(13), 2002, 2057–2072.

[21] Kleeman, R. and Majda, A., Predictability in a model of geophysical turbulence, *J. Atmos. Sci.*, **62**(8), 2005, 2864–2879.

[22] Kleeman, R., Majda, A. and Timofeyev, I., Quantifying predictability in a model with statistical features of the atmosphere, *Proc. Natl. Acad. Sci.*, **99**(24), 2002, 15291–15296.

[23] Kolmogorov, A. N., The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers, *Dokl. Akad. Nauk SSSR*, **30**, 1941, 299–303.

[24] Lewis, J. M., Roots of ensemble forecasting, *Mon. Weather Rev.*, **133**(7), 2005, 1865–1885.

[25] Leung, L. Y. and North, G., Information theory and climate prediction, *J. Climate*, **3**(1), 1990, 5–14.

[26] Lorenz, E. N., Deterministic nonperiodic flow, *J. Atmos. Sci.*, **20**(2), 1963, 130–141.

[27] Lorenz, E. N., Predictability of a flow which possesses many scales of motion, *Tellus*, **21**(3), 1969, 289–307.

[28] Lorenz, E. N., Climate predictability, The Physical Bases of Climate and Climate Modeling, GARP Publication Series, **16**, World Meteorological Organization, 1975, 132–136.

[29] Lorenz, E. N., Predictability — a problem partly solved, Proceedings of the Seminar on Predictability, **1**, ECMWF, Reading, Berkshire, 1996, 1–18.

[30] Lorenz, E. N., Designing chaotic models, *J. Atmos. Sci.*, **62**(5), 2005, 1574–1587.

[31] Lorenz, E. N. and Emanuel, K. A., Optimal Sites for Supplementary Weather Observations: Simulation with a Small Model, *J. Atmos. Sci.*, **55**(3), 1998, 399–414.

[32] Majda, A., Kleeman, R. and Cai, D., A mathematical framework for quantifying predictability through relative entropy, *Methods Appl. Anal.*, **9**(3), 2002, 425–444.

[33] Mannella, R., Integration of stochastic differential equations on a computer, *Internat. J. Modern Phys. C*, **13**(9), 2002, 1177–1194.

[34] Mintz, Y., A very long-term global integration of the primitive equations of atmospheric motion: An experiment in climate simulation, WMO-IUGG Symposium on Research and Development Aspects of Long-range Forecasting, WMO Technical Note, **66**, 1965, 141–167.

[35] Ortega, G. J., A new method to detect hidden frequencies in chaotic time series, *Phys. Lett. A*, **209**(5–6), 1995, 351–355.

[36] Ott, E., Chaos in Dynamical Systems, Second Edition, Cambridge University Press, Cambridge, 2002.

[37] Packard, N. H., Crutchfield, J. P., Farmer, J. D., et al, Geometry from a time series, *Phys. Rev. Lett.*, **45**(9), 1980, 712–716.

[38] Palmer, T., Molteni, F., Mureau, R., et al, Ensemble prediction, Proceedings of the Validation of Models Over Europe, **1**, 1993, 21–66.

[39] Reynolds, C. and Palmer, T., Decaying singular vectors and their impact on analysis and forecast correction, *J. Atmos. Sci.*, **55**(19), 1998, 3005–3023.

[40] Roulston, M. and Smith, L., Evaluating probabilistic forecasts using information theory, *Mon. Weather Rev.*, **130**(6), 2002, 1653–1660.

[41] Sauer, T., Yorke, J. A. and Casdagli, M., Embedology, *J. Stat. Phys.*, **65**(3–4), 1991, 579–616.

[42] Schneider, T. and Griffies, S., A conceptual framework for predictability studies, *J. Climate*, **12**(10), 1999, 3133–3155.

[43] Simmons, A. and Hollingsworth, A., Some aspects of the improvement in skill of numerical weather prediction, *Quart. J. Roy. Meteor. Soc.*, **128**(580), 2002, 647–677.

[44] Smith, L. A., Ziehmann, C. and Fraedrich, K., Uncertainty dynamics and predictability in chaotic systems, *Quart. J. Roy. Meteor. Soc.*, **125**(560), 1999, 2855–2886.

[45] Takens, F., Detecting strange attractors in turbulence, Dynamical Systems and Turbulence, Lecture Notes in Mathematics, **898**, D. A. Rand and L. S. Young (eds.), Springer-Verlag, Berlin, 1981.

[46] Thompson P. D., Uncertainty of the initial state as a factor in the predictability of large scale atmospheric flow patterns, *Tellus*, **9**, 1957, 275–295.

[47] Toth, A. and Kalnay, E., Ensemble forecasting at NMC: The generation of perturbations, *Bull. Amer. Meteor. Soc.*, **74**(12), 1993, 2317–2330.

[48] Tung, W. W., Lin, C. C., Chen, B. D., et al, Basic modes of cumulus heating and drying observed during TOGA-COARE IOP, *Geophys. Res. Lett.*, **26**(20), 1999, 3117–3120.

[49] Tung, W. W., Moncrieff, M. W. and Gao, J. B., A systematic view of the multiscale tropical deep convective variability over the tropical western-Pacific warm pool, *J. Climate*, **17**(14), 2004, 2736–2751.