Climate Dynamics
Lecture 10: Predictability in dynamical systems relevant to the atmosphere and ocean.

1 Introduction

Fundamental limits to predictability have received considerable attention in recent years due to the pioneering work of the atmospheric scientist Lorenz in the early 1960s. He showed that extreme sensitivity of weather predictions to the specification of initial conditions means that detailed forecasts are, in general, impossible beyond around 2 weeks. This pioneering study led over the following decades to the extensive study of chaotic dynamics by mathematicians and physicists such as Ruelle, Takens , Grassberger and Eckmann amongst many others. The original practical motivation of prediction has been somewhat lost in this onslaught of almost pure mathematics, elegant and inspiring as it may be. In this Lecture we present an approach aimed at rebuilding links between this mathematical stream and the more practical motivation of predicting quantities of utility to the general community.

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. This approach to predictability was pioneered by Leith in the early 1970s and implemented in a practical context by Palmer and Kalnay in the early 1990s. Mathematically, statistical prediction has as its basic focus of study, the time evolution of probability distribution functions (p.d.f.s). Ensemble prediction aims to produce representations of such evolving p.d.f.s by a Monte Carlo method of choosing samples of initial conditions drawn from the p.d.f. appropriate for the initial conditions (i.e. assuming we know the nature of the uncertainty at the beginning of our statistical prediction). Mathematically one might attempt to derive expressions for the evolution of p.d.f.s (for conservative systems these can be shown to satisfy the Liouville equation) however from a practical viewpoint this is an extremely difficult program because of the non-linearity of the equations and the very high dimensionality of the dynamical system (since the equations are p.d.e.'s in principle the dimensionality is infinite but a numerical discretization usually reduces this to a very large finite number of order 10^7 for state of the art weather models).

2 Quantifying uncertainty and prediction information content

A particularly attractive way of studying the evolution of statistical variables is provided by information theory which was invented by Claude Shannon in the late 1940s. This theory has since been applied by Jaynes and others to develop a deeper understanding of statistical mechanical systems as we will see shortly.

The informational content of a random vector variable may be defined precisely using the concept of entropy. Thus Shannon showed that the entropy associated with a particular p.d.f. may be written as

$$E = -\int p(\overrightarrow{x}) ln(p(\overrightarrow{x})) d\overrightarrow{x}$$

As the uncertainty associated with $\overrightarrow{x}$ increases (i.e. p becomes more uniform) then it's informational content E drops.

In a practical situation one typically has prior information regarding $\overrightarrow{x}$ namely it's historical behavior. This information can be identified readily with the asymptotic or equilibrium $p.d.f.$ of the dynamical system using an ergodic hypothesis. One can now ask how much additional information is obtained by providing the prediction $p.d.f.$ given the prior knowledge about the dynamical system. This quantity intuitively represents the usefulness of the prediction since it measures the additional information we have about the system by making a statistical prediction. Again it is possible to quantify this additional information using the concept of relative entropy:

$$R = \int p(\overrightarrow{x}) ln(\frac{p(\overrightarrow{x})}{q(\overrightarrow{x})}) d\overrightarrow{x}$$

where $q(\overrightarrow{x})$ is the equilibrium or prior known distribution (often referred to as the climatological distribution). It is instructive to calculate this quantity for Gaussian distributions q and p since many variables of interest in practical situations have such a distribution property approximately perhaps due to the central limit theorem of statistics. The result for the univariate case is

$$R = \frac{1}{2}\left[\ln\left(\frac{\sigma_q^2}{\sigma_p^2}\right) + \frac{\sigma_p^2}{\sigma_q^2} + \frac{(\mu_p - \mu_q)^2}{\sigma_q^2} - 1\right] \quad (1)$$

where σ^2 and μ are the variances and means of the prediction p.d.f. and equilibrium p.d.f.. The first two terms here represents the additional information provided by reducing the uncertainty of the statistical variable relative to that already known. The third term represents the information added by predicting a large deviation from the historical behavior of the system. We will refer to these two components respectively as the dispersion and signal contributors to Gaussian relative entropy. They are strongly related (although not directly) to traditional measures of the skill of statistical predictions (RMS error and anomaly correlation skill).

3 The second law of thermodynamics and the time dependent nature of predictability

If one views the prediction process as the relaxation of a prediction p.d.f. towards its equilibrium or asymptotic distribution one obtains a very natural connection with non-equilibrium statistical mechanics which is all about the same process for large scale molecular dynamical systems. In a certain sense the prediction problem represents a generalization of statistical mechanics which after all represents very particular (albeit complicated for liquids and solids) dynamical systems. This analogy and its relation to information theory was first commented upon by Jaynes in the late 1950s. Relative entropy has a very natural role within this general picture of prediction theory. For Markov dynamical processes it may be shown rigorously that the relative entropy declines monotonically (obviously towards zero as time becomes large). Such processes are defined as those in which a random variable at time step t depends only on random variables at time step $t-1$ and not on earlier times. For most systems of interest to us this condition is met as this is how numerical models of the atmosphere and ocean effectively work (i.e. via a time stepping algorithm). This principle of monotonic decline of relative entropy represents the generalization of the second law of thermodynamics from statistical mechanics. In that context the equilibrium distribution is typically uniform and so relative entropy reduces to minus the ordinary entropy (q.f. the definitions of these quantities above) and hence ordinary entropy increases monotonically.

Another practical issue in predictability studies concerns variation with initial conditions. Predictions commencing from certain times may be more useful than others due to factors such as the instability of the flow or the particular structure of the initial conditions themselves. If the p.d.f.s were close to Gaussian (as they often are) then equation (1) (or its multivariate generalization) suggests that factors controlling the prediction signal term or the prediction dispersion term may be responsible for such variations in predictability. The latter term is essentially controlled by the instability of the flow. The former term is not so easy to pin down but an example to be discussed below will bring out its significance more clearly. Which of these terms is more important overall to predictability is an interesting question and it turns out to depend on the particular dynamical system under consideration.

4 A simple relevant example

We consider firstly a stochastically forced oscillator as it illustrates some interesting effects. It is also an interesting 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 Nino/Southern Oscillation (ENSO). Consider the following two dimensional stochastic differential equation:

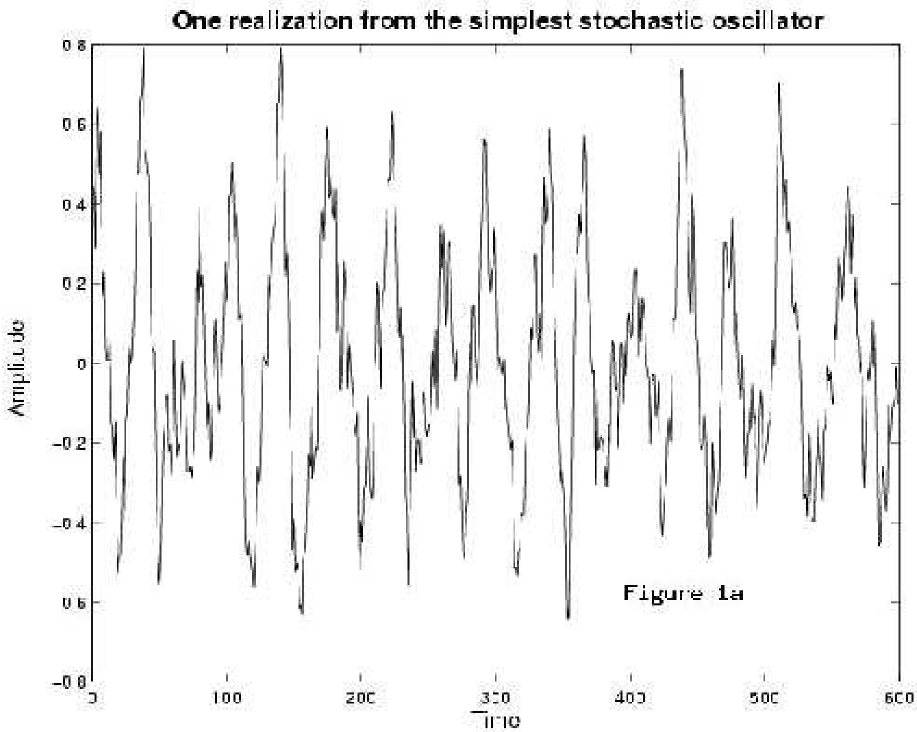
$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix}_t = \begin{pmatrix} 0 & 1 \\ \beta & \gamma \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \begin{pmatrix} 0 \\ F \end{pmatrix} \tag{2}$$

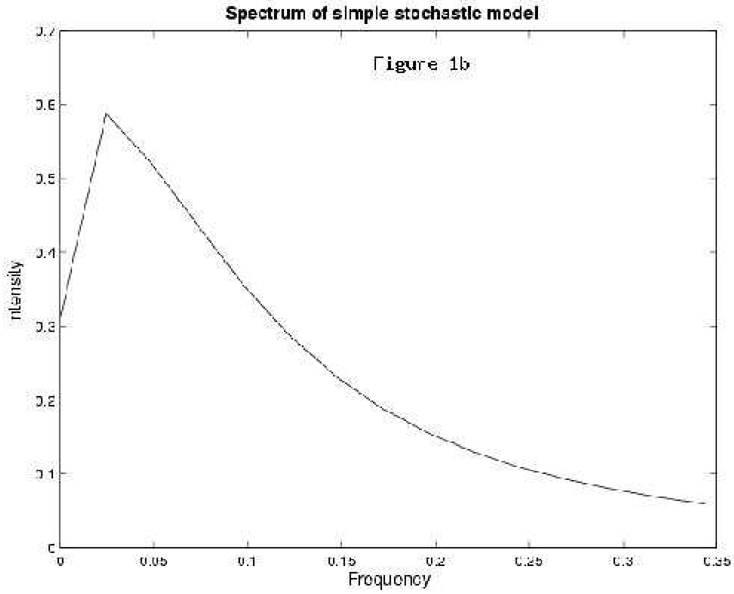
where F is white with variance C and mean zero. Without the forcing, it is easily shown that damped oscillations occur with period T and damping time

τ where we have

$$\begin{aligned} \gamma &= -\frac{2}{\tau} \\ \beta &= -\left(\frac{4\pi^2}{T^2} + \frac{1}{\tau^2}\right) \end{aligned}$$

A realization of this stochastic differential equation with $\tau = T = 36$ is displayed in Figure 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 in standard texts on stochastic differential equations (e.g., Gardiner, 1985). The covariance matrix at time t (given a deterministic set of initial conditions at time 0) is given by

$$\begin{pmatrix} \langle u_1, u_1\rangle & \langle u_1, u_2\rangle \\ \langle u_2, u_1\rangle & \langle u_2, u_2\rangle \end{pmatrix} = \int_0^t dt' \exp\left((t-t')A\right) \begin{pmatrix} 0 & 0 \\ 0 & C \end{pmatrix} \exp\left((t-t')A^t\right)$$

where $A = \begin{pmatrix} 0 & 1 \\ \beta & \gamma \end{pmatrix}$. Further the mean vector at time t is given by

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

Since the equations are linear it follows that all probability distribution functions will be Gaussian providing that the stochastic forcing has this property which 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 of $\exp(sA)$[1]. 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. Calculation now of the relative entropy R (or prediction utility as we shall call it) for all prediction ensembles for this dynamical system is straightforward.

[1] This is obtained by diagonalizing A

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 counter-example to a 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 Figure 1. The probability distribution of utility at 12 months is shown in Figure 2b.

Figure 2a

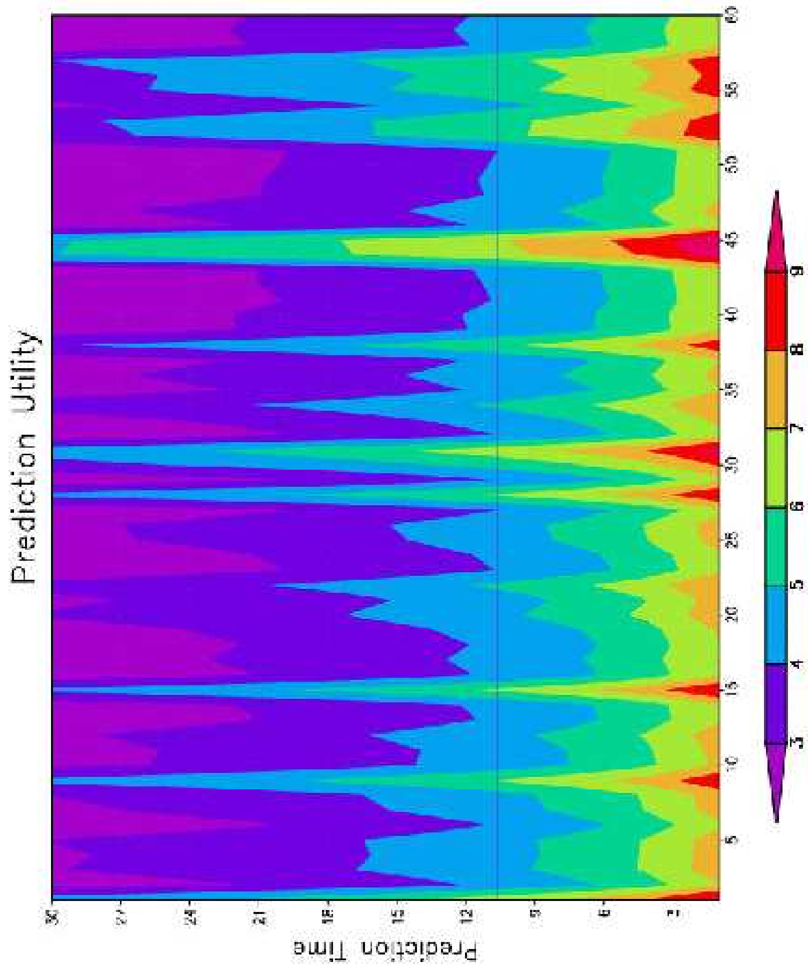
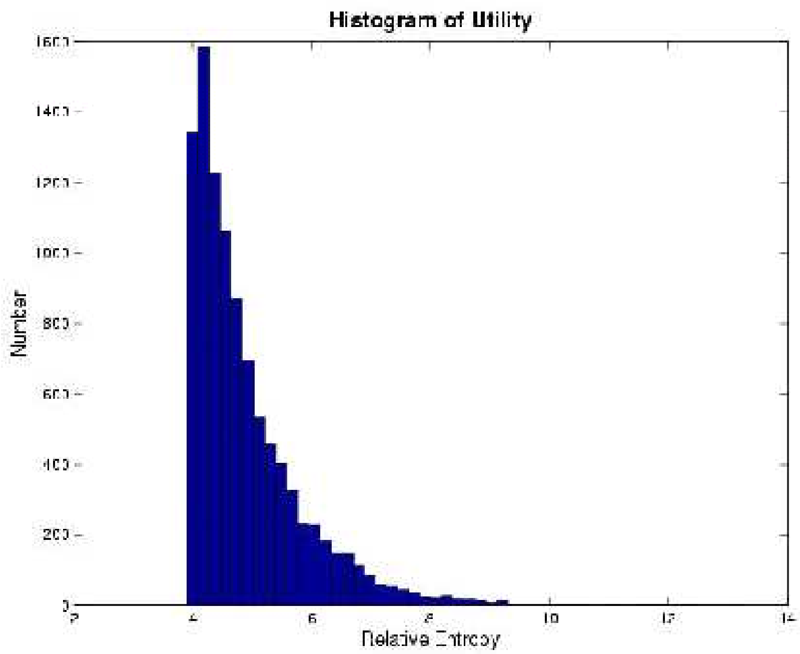


Figure 2b



This was constructed using 10000 initial conditions drawn at random from the realization of Figure 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 which "contain a large signal".

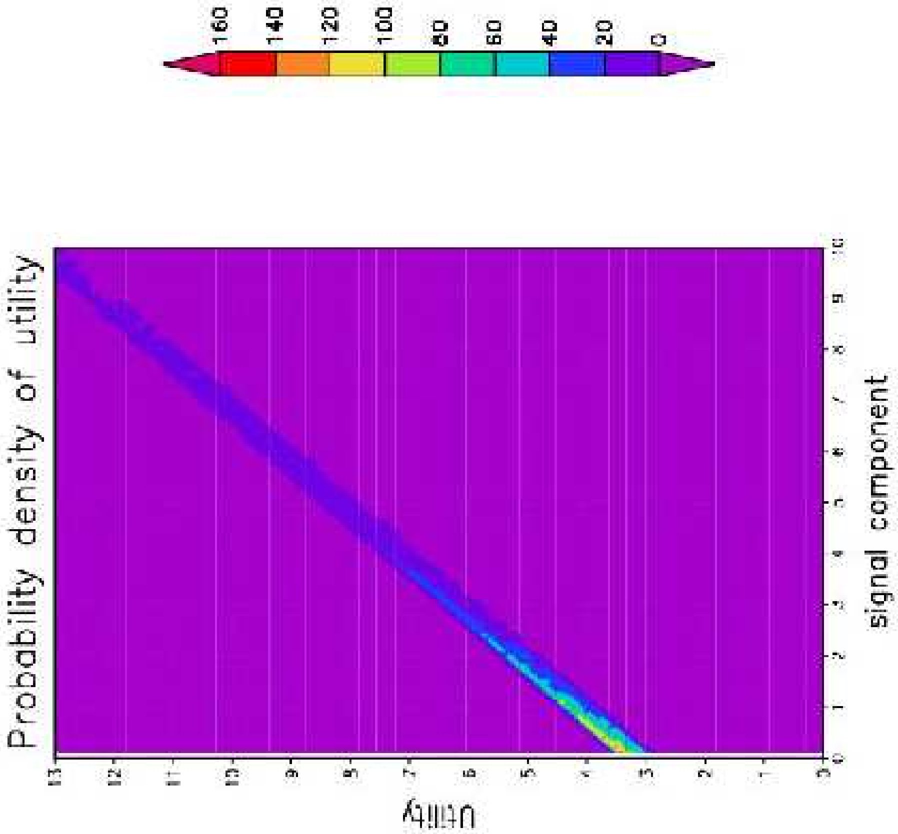
Evidence from ENSO dynamical models suggests that the simple model above should be modified to take into account potentially large variations in the stability of the system caused by both the annual and ENSO cycles. We modified the model then by allowing the parameter τ to vary periodically with time. We chose periods P for this variation of $T/3$ and T and assumed a sinusoidal variation in $1/\tau$ of the form

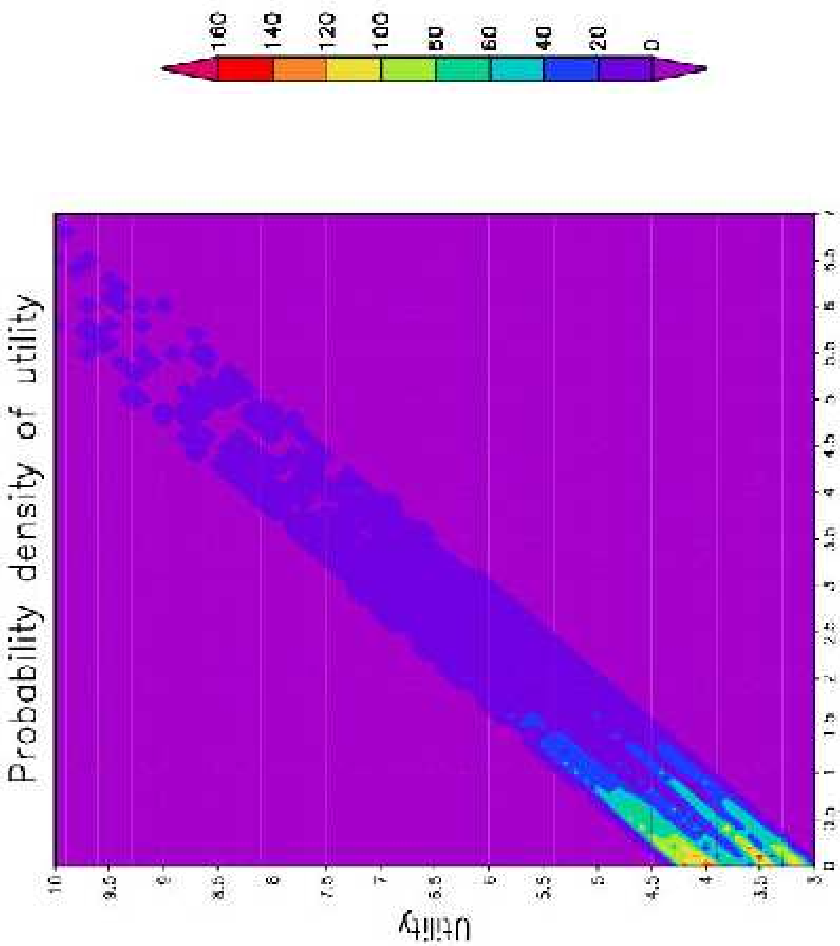
$$1/\tau = \left(1 + 2\sin\frac{2\pi t}{P}\right)/T \tag{3}$$

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[2] and we may still therefore use the expressions of equation (**??**) to calculate utility. Completely analytical expressions are now not easily obtained and so we must rely on numerical models of our equations to estimate the first and second moments of the prediction ensembles. Results were checked at all times 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 entropy. This can be assessed by plotting the signal term versus the total relative entropy for a large number (10,000) of randomly chosen initial conditions and this may be seen in Figure 3.

[2] See Gardiner, 1985 Chapter 4. Note that prediction ensemble distributions only become non-Gaussian when the operator A becomes non-linear.


Probability density of utility
signal component
utility
160
140
120
100
80
60
40
20
0



Here the probability density for each point on the plot is estimated using a "circle of influence" measure i.e. the number of sample points lying within a suitably small radius of parameter space[3] 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 Figure 1). Figure 3a 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 Figure 3b 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 equation (3) quite significantly. The only parameter causing a relative change

[3]This was chosen for convenience to be 0.1 Note the axes scale in Figure 4 for comparison with this.

in the importance of the dispersion and signal terms was the period of the stability cycle.

5 Further results and future research

The methodology and analysis discussed above can be applied to an arbitrary dynamical systems. In more general non-Gaussian cases the calculation of the entropy becomes non-trivial particularly when the dimension of the system is high. One example studied is the well known Lorenz model which is often cited (erroneously?) as a simple model of the atmosphere and it's chaotic behavior. Here it was found that for short range prediction dispersion is a much more important than signal to relative entropy which is counter to the stochastic model results discussed above. This result apparently does not carry over to higher dimensional systems (the Lorenz system is a 3 dimensional truncation of the large system discussed in the introduction to this Lecture). We examined a system with dimension 100 and found that signal was far more important to predictability. As the dimension of such systems increase they tend to become less "chaotic" and more "stochastic" in character consistent with our findings above. This research is ongoing and graduate students are welcome to become part of it as part of their project.