Information Theory and Predictability Lecture 10: Predictability of some simple models

1 Gaussian case

As was mentioned in Lecture 7 the case where probability distributions are Gaussian is both practically relevant since many complex system distributions are approximately Gaussian but also because analytical expressions are possible for entropic functionals. This enables us to make more concrete the meaning of prediction utility which we have already identified with relative entropy. Recall that it has the form:

$$\begin{aligned} D(p\|q) &= \tfrac{1}{2}\left[\log\left(\det\left(\boldsymbol{\sigma}_y\right)/\det\left(\boldsymbol{\sigma}_x\right)\right) + tr\left(\boldsymbol{\sigma}_x\boldsymbol{\sigma}_{\boldsymbol{y}}^{\boldsymbol{-1}}\right) - n\right] \quad \textit{Dispersion} \\ &+ \quad \tfrac{1}{2}(\overline{\mathbf{x}}-\overline{\mathbf{y}})^t\boldsymbol{\sigma}_y^{-1}(\overline{\mathbf{x}}-\overline{\mathbf{y}}) \quad \textit{Signal} \end{aligned} \tag{1}$$

where for our application the subscripts x and y refer to the prediction and equilibrium distributions p and q respectively while $\mathbf{x}$ and $\mathbf{y}$ are the respective values of the random vectors.

Notice we have decomposed this into two pieces (dispersion and signal) one of which contains the prediction covariance matrix (and not the prediction mean) while the other involves the prediction mean but not the covariance.

When the spread of the prediction distribution is small compared with that of the equilibrium distribution, as it often is for short range predictions, then the dispersion component of the relative entropy is dominated by the first term. As was noted in Lecture 7 this term is the difference in the entropies of the two distributions. Thus the dispersion term is in this case measuring the reduction in uncertainty of the random variables resulting from the prediction process.

The second expression "Signal" measures something quite different. It is large when the means of the two distributions are large relative to the equilibrium spread for a sufficient number of principal components of the equilibrium distribution. In an intuitive sense then this term is significant if the prediction means of enough principal components have "large anomalies".

A concrete example helps explain the signal utility. Suppose that the mean climatological temperature for New York in April is $50°F$ with a standard deviation of $5°F$ and suppose the prediction for tomorrow is $75°F$ with an expected uncertainty of $5°F$. Clearly such a prediction has utility and this derives not from any reduction in uncertainty (there is none here) but from the shift in the mean. In terms of the relative entropy the signal contribution to utility will be $(15/5)^2/2 = 4.5\,nats$ assuming of course that the relevant distributions are Gaussian (which they actually are to a close approximation).

2 Two simple examples

2.1 Lorenz Model

As mentioned in Lecture 8, in the 1950s it was realized that weather forecasts of sufficient duration show extreme sensitivity to initial condition errors. This realization ushered in an intense interest in the subject of chaos and associated concepts such as strange attractors and their fractal nature. The theoretical aspects of this relevant for predictability were discussed in Lecture 9. In the terminology we introduced earlier, the strange attractor of such dynamical systems is the equilibrium distribution. The fractal nature of the attractor shows how complex such distributions can be in even simple non-linear systems. It is interesting to see how the theoretical framework involving relative entropy introduced earlier works out in a classical example of such a low order dynamical system. The archetypal exampole of such chaotic systems was introduced by the MIT meteorologist Edward Lorenz in the early 1960s (see [5]). This system is 3 dimensional and satisfies the equations:

$$\begin{aligned}\frac{\partial x}{\partial t} &= -\sigma(x-y)\\ \frac{\partial y}{\partial t} &= \rho x - y - xz\\ \frac{\partial z}{\partial t} &= xy - \beta z\end{aligned}$$

where $\sigma = 10$, $\rho = 8/3$ and $\beta = 28$ are constants for the standard configuration. It is not possible to solve analytically for the time evolution of the prediction distribution within such a system because of it's non-linearity. Instead one must resort to numerical ensemble methods in that an ensemble must be generated from an assumed initial condition distribution and then probability densities at all places in state space estimated from the ensemble. More specific details are given below. Given the low dimensionality of the system it is possible to use very large ensembles (order 10^6 or greater) to perform such calculations. A typical such calculation is shown in Figure 1 which shows the evolution of an initially very tightly distributed (Gaussian) initial condition ensemble toward the equilibrium strange attractor distribution.

Observe that the initially Gaussian distribution becomes rapidly non-Gaussian (multi-modal in fact) and as it equilibrates the ensemble appears to "mix into" the strange attractor in a highly non-Gaussian fashion.

We can follow this equilibration process for a variety of different initial condition distributions using the relative entropy. In order to do this we need as already mentioned, to use an ensemble to estimate the relevant prediction probabilities. One approach to this mentioned in Lecture 9 is to divide a volume containing the attractor into a large number of bins or partitions and obtain using our huge ensembles an estimate of the distribution. This can then be used to obtain a coarse-grained estimate of the relative entropy. Another similar approach, which we follow here, is to calculate the number of prediction and equilibrium ensemble points lying within a small sphere around each prediction ensemble member and use these to obtain another coarse-grained estimate of relative entropy by simply averaging over the prediction ensemble. There are

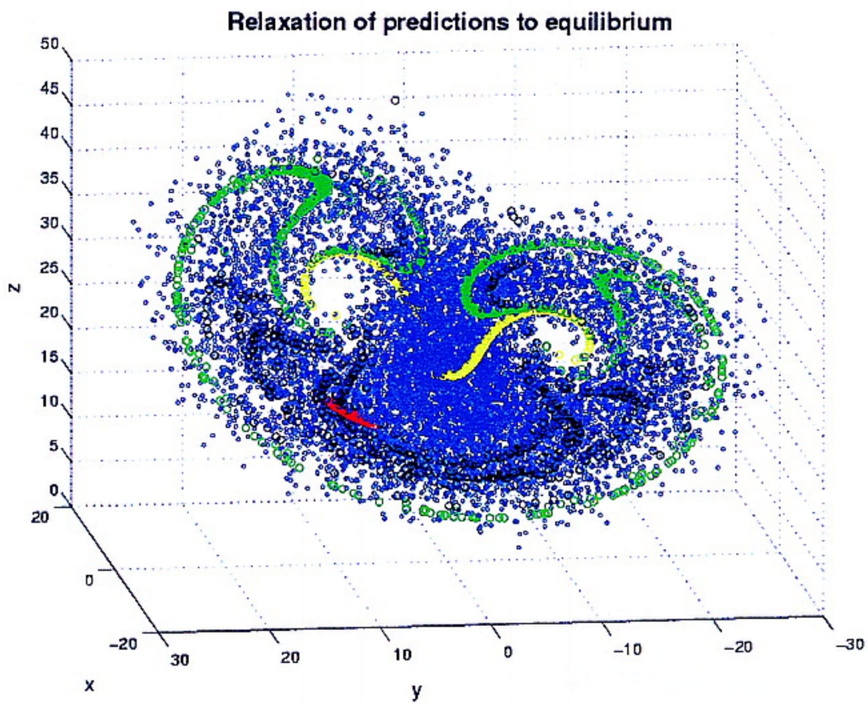

Figure 1: Ensemble relaxation to equilibrium in the Lorenz model. Different colors indicate different times with blue being the asymptotic strange attractor. The initial condition ensemble was located near the red cloud and the time ordering is red, yellow, green, black and blue.

evidently subtle technical issues involved in considering the continuous limit of bins of vanishing volume or similarly spheres of decreasing size. This follows from the fact that the equilibrium distribution is fractal in structure and hence highly discontinuous as well as having a non-integral dimension. Recall that the relative entropy can only be defined when the equlibrium distribution does not vanish. It follows that we must be rather careful about how we define the continuum limit. The interested reader is referred to [1] for further details.

It is interesting to see how relative entropy or utility varies with the choice of initial condition. This has practical implications since it is evidently useful to a forecaster to know whether their prediction is likely to have high or low utility and why. In this model we looked at this by picking a series of random initial conditions drawn from the strange attractor during a long integration and used them to set the mean values of an initial condition distribution. For the purposes of this exercise we assumed that the error vector for the initial conditions was constant but aligned in a tangential plane to the attractor. This is obviously rather a simplistic assumption but serves our present pedagogical puposes. Figure 2 shows a plot of relative entropy as a function of time and initial conditions.

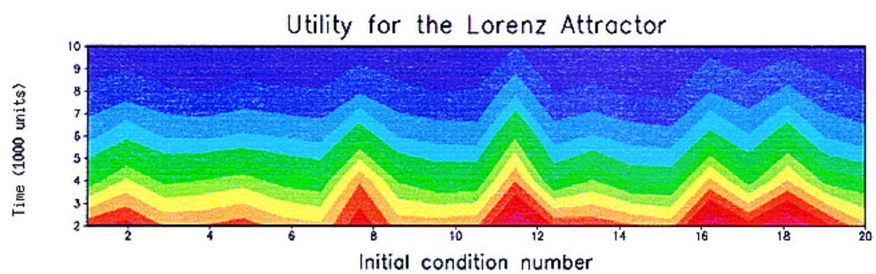

Figure 2: Relative entropy as a function of time (vertical axis) and initial condition (horizontal axis). The contour interval in 0.5 nats and the final contour line at the top of the plot is 0.5. For reference the color clouds in Figure 1 are at times 2,4,6 and 8.

Notice that there is some variation in utility with perhaps four out of the twenty predictions having high relative entropy for all times. Notice also the universal almost linear monotonic reduction with time. Interestingly the low and high utility predictions always come from the same part of the attractor as is shown in Figure 3.

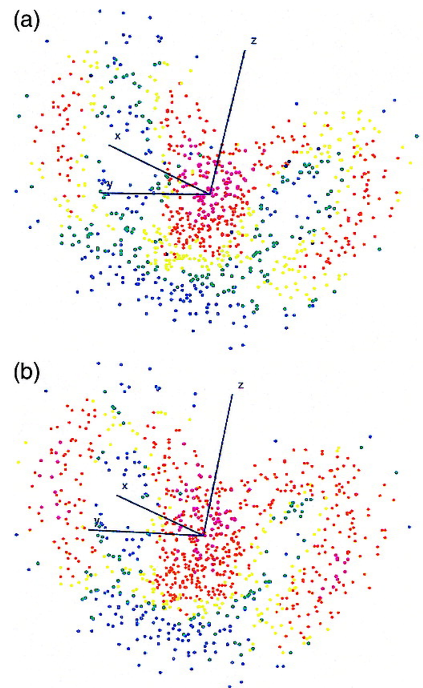

Figure 3: Initial condition origin of predictions of varying levels of utility at two times which are a substantial fraction of the equilibration time scale. The colors show the level of utility ranging from red for low utility predictions to blue for high utility.

Notice how the low utility predictions cluster around the "body of the butterfly". This region also happens to be the place where linear instability analysis reveals the greatest growth rates for small perturbations. Such a result suggests (naively) that it may be variations in ensemble spread due to linear instability variations which may be mainly responsible for utility variations. If the distributions involved were Gaussian then this would correspond to the dispersion component of equation (1) being very strongly correlated with utility. Now obviously Figure 1 above shows us that for longer range predictions, distributions are far from Gaussian and it turns out that indeed in that case the utility is not strongly related to (Gaussian) dispersion component or indeed to a Gaussian approximation for relative entropy (see Figure 4c). For short range predictions however which are approximately Gaussian because of our choice for the initial condition distribution, we might expect a strong relationship between actual relative entropy and the Gaussian approximation which indeed we do see (Figure 4a). Moreover, consistent with our argument above concerning linear instability, we find that this relationship is due to the dispersion component of the Gaussian expression for relative entropy (Figure 4b). It is rather interesting that this dispersion/instability/ensemble spread effect somehow transforms itself into some sort of higher order moment effect for longer range predictions.

Given the highly non-Gaussian distributions at longer prediction times these effects can be seen rather more cleanly by consideration of the relationship between relative entropy and the general (as opposed to Gaussian) entropy difference between the prediction and equilibrium ensembles. This is displayed in Figure 5a for predictions roughly half way to equilibrium. As may be noted, a very strong relationship exists between $D(p||q)$ and $H(q) - H(p)$ where p is the prediction distribution while q is the equilibrium distribution. For short range predictions this effect is consistent with the Gaussian results already seen but it is interesting to see that they persist into the non-Gaussian regime.

Also shown in Figure 5b is the relationship between relative entropy at different prediction times. It is notable that high utility predictions tend to remain high utility predictions (and conversely) as time proceeds. This effect we term durability. These results are consistent with the results of Figure 3.

Finally it is important to emphasize that for no prediction time is the Gaussian signal component ever strongly related to utility. We shall see below however that this latter result is not the case for other important dynamical systems.

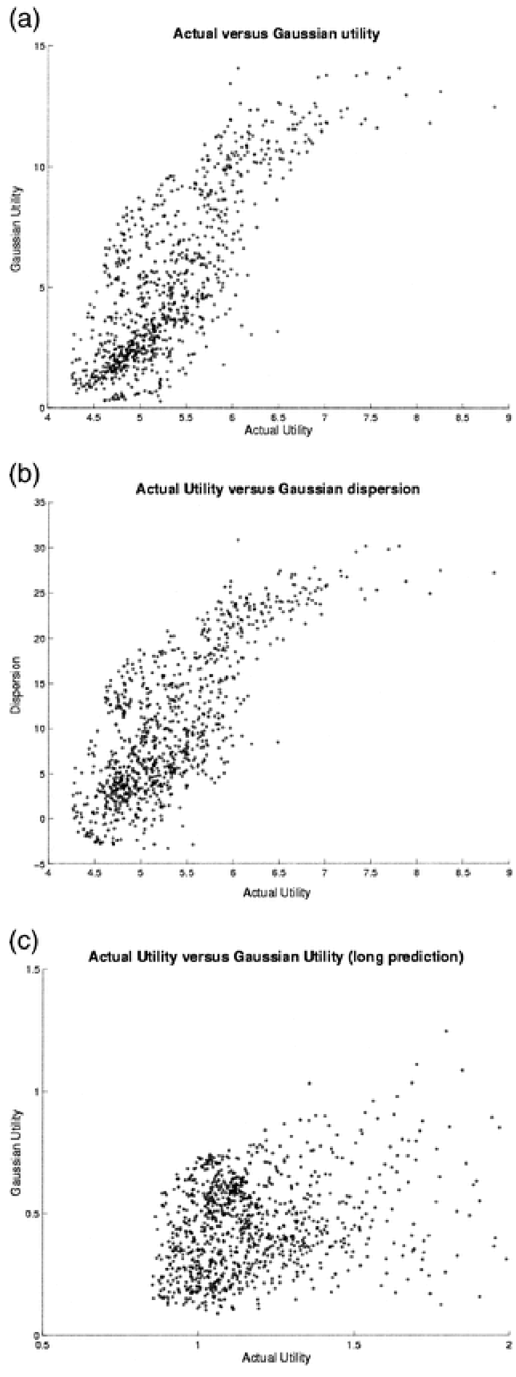

Figure 4: The relation between actual relative entropy and the relative entropy applicable if the distribution were assumed to be Gaussian. The first and last panels show the relationship for short and long times (with respect to equilibration) respectively. The middle panel shows that the reasonable relationship for short range predictions is due to the dispersion component of the Gaussian relative entropy.

2.2 Stochastically forced oscillator

The model considered in the previous section is a very simple example of a chaotic system. Another class of simple systems thought to be relevant for more realistic dynamical systems are the stochastic differential equations which we discussed in Lecture 5. Such models suppose that the dynamical system under consideration can be separated into fast and slow components with the former represented by some kind of stochastic forcing. An example where such modeling has been proposed is the El Nino/Southern Oscillation (ENSO) phenomenon. This coupled ocean atmosphere system is irregular with a spectral peak around a period of four years. In addition it shows strong cyclical behavior in time involving a rather small set of large scale ocean modes. Given this it has been proposed (see, for example, [4] and [6]) that a stochastically forced damped oscillator may be a good simple model of this phenomenon. Here the fast component of the system comes from atmospheric weather while the slow component derives from interior ocean adjustment. The simplest representation of such a system is given by the equations

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

where the constants β and γ are related to the oscillation period T and decay time τ by the relations

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

Physically in the ENSO case u_1 and u_2 represent a basin-wide heat content buildup and an Eastern Pacific SST increase respectively. The forcing F is white noise.

Comparison of equation (2) with the Fokker-Planck equation (2.1) Lecture 5 and it's reference "Langevin" equation (2.2) Lecture 5 shows that the Fokker-Planck coefficients A_i are linear functions of u_i while B_i are constants. The Fokker-Planck equation is analytically solvable for this case (exercise) and the solutions are bivariate Gaussian probability densities (see first equation of Lecture 7) in the case that the initial conditions are deterministic or Gaussian. The equlibrium distribution can also be shown to be of this form. Note that this distribution is continuous, differentiable and non-vanishing everywhere unlike the chaotic case considered above. This allows for the easy analytical calculation of continuous entropic functionals.

Interestingly it may be shown quite easily that the (Gaussian) solution of a general linear stochastic differential equation with additive noise has an associated covariance matrix which *do not depend on the particular initial conditions chosen.* Instead this matrix only depends on the linear coefficients and time. This fact has important implications for the predictability properties of such dynamical systems.

Again it is interesting to look at the behavior of relative entropy as a function of initial conditions and time. We chose deterministic initial conditions

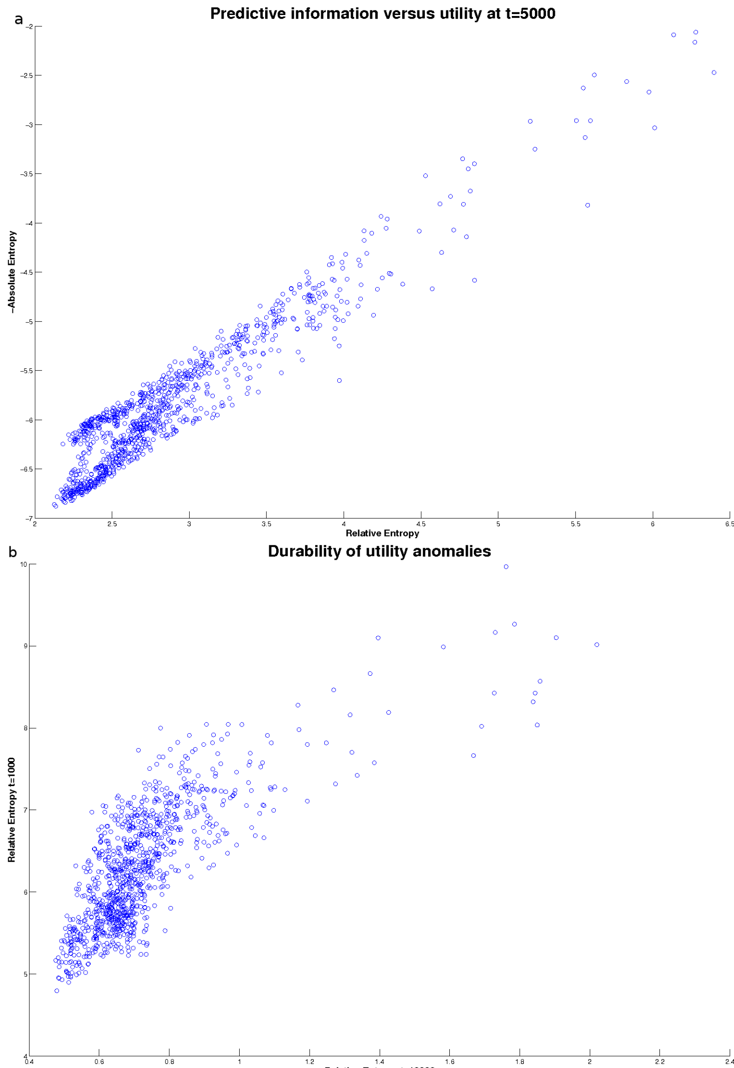

Figure 5: Top panel shows the relationship between entropy difference and relative entropy at a time roughly midway to equilibrium.predict. Bottom panel shows the relationship between relative entropy at early times with later times.

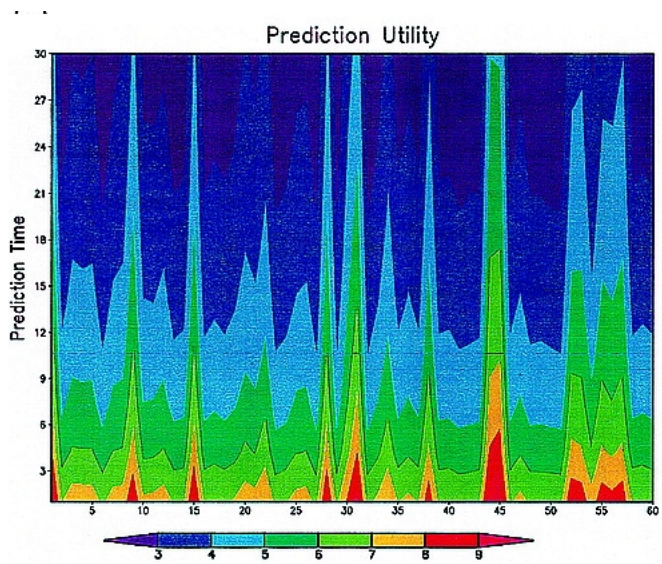

Figure 6: Relative entropy as a function of time and initial condition for the stochastically forced damped oscillator.

here (sampled from the equilibrium distribution) assuming that error growth in the system derives from small errors in the fast component (the atmosphere). Results are not very different if a similar choice of initial condition distribution to the previous model are assumed.

Figure 6 shows that there is considerable variation in utility from one initial condition to another and that the relative entropy tends to fall off in an approximately exponential fashion with time. The degree of variability with initial condition is noticeably higher than in the Lorenz case. Like that case however high (and low) utility tends to persist for the duration of a prediction.

Since the distributions for this model are all Gaussian we can analyze the causes of relative entropy variation exactly using equation (1). Given that the covariance matrix does not depend on the initial conditions used it is obvious that the large variations seen are due entirely to variations in the signal component. This situation is in sharp contrast to the Lorenz model.

As we saw above, the signal component is large when the prediction means of principal components of the equilibrium distribution are large relative to the equilibrium standard deviation of these components. For the particular model here one can show (exercise) that asymptotically u_1 and u_2 are uncorrelated implying that the relevant equilibrium principal components are actually just u_1 and u_2 themselves. In addition one can verify that the asymptotic values for the means of these components are zero. Consequently the signal and hence relative entropy will be large when the L_2 norm $\left(\frac{u_1}{\sigma_1^e}\right)^2 + \left(\frac{u_2}{\sigma_2^e}\right)^2$ is large (where σ_i^e are the equilibrium standard deviations). In other words when the scaled amplitude of the oscillation of the system is large.

In general this amplitude shows considerable variation for any particular realization of the stochastic process. The persistence of the signal throughout the high utility predictions evident in Figure 6 is due to the fact that it takes some considerable time in a prediction for the stochastic forcing to erode a highly anomalous set of initial conditions.

This picture of climate predictability variation due to signal variation has been confirmed in more realistic prediction models of ENSO. It suggests that there will be large variations in the utility of such predictions due simply to whether anomalies of the principal components of ENSO are large or not in the initial conditions. This accords with much practical experience of ENSO prediction.

2.3 Higher order realistic models.

The three dimensional model we examined in subsection 2.1 is rather too simple to be considered indicative of what is likely to occur in more realistic systems relevant to weather prediction. Indeed as the dimension of the system increases and the behaviour becomes better considered turbulent rather than chaotic, significant qualtitative differences regarding predictability emerge. Rather curiously the distributions observed in such more complex systems become increasingly more Gaussian as the dimension increases. If one makes the admittedly crude Gaussian approximation in such systems then it is possible to make progress on the predictability variation issues discussed above. This is a rather technical area which we do not enter into in any detail however the interested reader can find more information in the review paper of this area authored by the lecturer (see [3]). Using the Gaussian approximation a dynamical model relevant to the mid-latitude atmosphere has been analysed from the viewpoint of predictability variation. One might wonder given the very different results obtained in 2.1 and 2.2 which situation pertains in a realistic weather prediction system? Are variations primarily related to dispersion or signal variations? The results may be found in [2] and appear to indicate that signal is more important rather like in the stochastic case of section 2.2. These results are the first derived on this question and should be regarded at this stage given all the simplifications assumed, as preliminary. They are however tantalizing since understanding the causes of predicability variations is of immense practical importance.

References

[1] R. Kleeman. Measuring dynamical prediction utility using relative entropy. *J. Atmos. Sci.*, 59:2057–2072, 2002.

[2] R. Kleeman. Limits, variability and general behaviour of statistical predictability of the mid-latitude atmosphere. *J. Atmos Sci*, 65:263–275, 2008.

[3] R. Kleeman. Information theory and dynamical system predictability. *Entropy*, 13:612–649, 2011.

[4] R. Kleeman and S.B. Power. Limits to predictability in a coupled ocean-atmosphere model due to atmospheric noise. *Tellus A*, 46:529–540, 1994.

[5] E. N. Lorenz. Deterministic non-periodic flows. *J. Atmos. Sci.*, 20:130–141, 1963.

[6] C. Penland and P.D. Sardeshmukh. The optimal growth of tropical sea surface temperature anomalies. *J. Clim.*, 8(8):1999–2024, 1995.