

# Quantifying predictability in a model with statistical features of the atmosphere

Richard Kleeman<sup>†</sup>, Andrew J. Majda, and Ilya Timofeyev

Courant Institute of Mathematical Sciences and Center for Atmosphere and Ocean Sciences, New York University, New York, NY 10012

Contributed by Andrew J. Majda, September 26, 2002

The Galerkin truncated inviscid Burgers equation has recently been shown by the authors to be a simple model with many degrees of freedom, with many statistical properties similar to those occurring in dynamical systems relevant to the atmosphere. These properties include long time-correlated, large-scale modes of low frequency variability and short time-correlated “weather modes” at smaller scales. The correlation scaling in the model extends over several decades and may be explained by a simple theory. Here a thorough analysis of the nature of predictability in the idealized system is developed by using a theoretical framework developed by R.K. This analysis is based on a relative entropy functional that has been shown elsewhere by one of the authors to measure the utility of statistical predictions precisely. The analysis is facilitated by the fact that most relevant probability distributions are approximately Gaussian if the initial conditions are assumed to be so. Rather surprisingly this holds for both the equilibrium (climatological) and nonequilibrium (prediction) distributions. We find that in most cases the absolute difference in the first moments of these two distributions (the “signal” component) is the main determinant of predictive utility variations. Contrary to conventional belief in the ensemble prediction area, the dispersion of prediction ensembles is generally of secondary importance in accounting for variations in utility associated with different initial conditions. This conclusion has potentially important implications for practical weather prediction, where traditionally most attention has focused on dispersion and its variability.

**P**redictability of dynamical systems relevant to the atmosphere and climate is a topic of enormous practical and theoretical interest. For several decades it has been recognized that a statistical framework is required for an adequate analysis of this subject (1–4).

Recently, ideas from information theory have made a natural appearance (5–7), because entropy measures offer a precise definition of the informational content of predictions based on probability distribution functions (pdfs).

Here we provide a concise summary of the relevant material contained in ref. 7. This reference contains considerably more detail and application to a range of dynamical systems relevant to climate and weather.

For any dynamical prediction there always exists uncertainty in the specification of initial conditions, and this may be described by a pdf. The time evolution of this pdf is at the heart of any analysis of the statistical prediction problem, which is characterized by two pdfs: the prediction distribution  $p$  and the climatological or equilibrium distribution  $q$ . The former is the time-evolved initial-condition pdf, and the latter is the asymptotic ( $t \rightarrow \infty$ ) distribution. In most practical situations there is considerable knowledge concerning  $q$  due to the long-term historical observation of the dynamical system (under the assumption of ergodicity). In terms of the informational content of a prediction, knowledge of  $q$  therefore may be considered prior information, because this is the information we have on a dynamical system *before a particular prediction is made*. One may quantify the additional information provided by  $p$  over and above the prior known  $q$ , and this will measure the utility of the

prediction. The relevant two-valued functional is known as the *relative entropy* and is given generically by

$$R(p, q) \equiv \int p \ln \left( \frac{p}{q} \right) dV. \quad [1]$$

The always nonnegative functional  $R$  has the attractive property of satisfying a generalized second law of thermodynamics for Markov processes in that it declines monotonically with time toward an asymptotic value of zero. Refer to ref. 8 for a short rigorous derivation of this interesting result. It is sometimes (9) deployed in applications of Boltzmann’s H theorem to nonequilibrium statistical mechanics problems. The property of  $R$  as a (nonsymmetric) distance function between pdfs also means that it plays an important role in the analysis of the approach to equilibrium of solutions of the Fokker–Planck equation (10).

The distributions  $p$  and  $q$  that one encounters in many practical contexts are apparently approximately Gaussian.<sup>‡</sup> In such a case an exact expression may be obtained for  $R$ . For the multivariate case, this is easily shown to be given by

$$R(p, q) = 0.5 * \left\{ \ln \left[ \frac{\det(\sigma_q^2)}{\det(\sigma_p^2)} \right] + \text{tr}[\sigma_p^2(\sigma_q^2)^{-1}] + (\tilde{\mu}_p - \tilde{\mu}_q^t)(\sigma_q^2)^{-1}(\tilde{\mu}_p - \tilde{\mu}_q) - n \right\}, \quad [2]$$

where  $\sigma_q^2$  is the covariance matrix of the equilibrium distribution,  $\sigma_p^2$  is the covariance matrix of the prediction distribution,  $\tilde{\mu}_q$  and  $\tilde{\mu}_p$  are the mean vectors of these two distributions, respectively, and  $n$  is the dimension of the state space under consideration. For pedagogical convenience we call the third term the *signal component*, and the sum of the remaining terms are termed the *dispersion component* of Gaussian relative entropy. In the univariate case, the dispersion component contributes to prediction utility when the prediction reduces the uncertainty from the uncertainty of the equilibrium or prior distribution. The signal component contributes when the mean of the prediction distribution differs significantly from that of the equilibrium distribution.

A fundamental and general question one can ask in predictability theory is: What determines *variations* in utility as a function of initial conditions? In particular, it is worth determining whether the Gaussian dispersion or signal component is a major determinant of such utility variation. This question is of some importance in a practical context, because a user of forecasts needs some guidance on whether a *particular* prediction is likely to be useful, and parameters such as the (Gaussian) signal or dispersion may be good indicators of this even when distributions are only approximately Gaussian. In general, the study of predictability in the atmospheric context has focused

Abbreviations: pdf, probability distribution function; TBM, truncated Burgers model.

<sup>†</sup>To whom correspondence should be addressed. E-mail: kleeman@cims.nyu.edu.

<sup>‡</sup>The limited size of available practical ensembles in practical situations makes it difficult to be completely precise on this point.

almost exclusively on variations with initial conditions of quantities that are functions of the second moments of  $p$  and  $q$  and thus are related to the second component above, i.e., the dispersion (however, see ref. 11 for another viewpoint). Ensemble spread<sup>§</sup> versus correlation skill diagrams are widely used (12, 13) despite the fact that this relationship often is not particularly strong. In fact, correlation skill may be shown theoretically (14) to depend on both the first and second moments of  $p$ , which suggests that the signal component might be a useful and neglected measure of practical prediction utility.

An interesting and simple counterexample to the conventional concentration on second moments is provided by linear constant coefficient stochastic differential equations with deterministic initial conditions. Such systems obviously have wide application and in particular have been proposed as simple representations of climate dynamical systems (15). For such dynamical systems (10), all distributions are Gaussian, and the prediction covariance matrix is *independent* of the initial conditions. On the other hand, the prediction mean vector for such a system is simply the dynamical propagator operator<sup>¶</sup> applied to the initial conditions, which obviously depends strongly on the particular initial conditions. Clearly in this class of dynamical systems it is only the signal term that is responsible for any variation with initial conditions of prediction utility.

Exploration in ref. 7 with a range of somewhat more complex climate models relevant to the El Niño phenomenon suggested that the signal component was generally more important than dispersion in determining utility variation. On the other hand, for the classical Lorenz (16) three-variable model of chaos dispersion dominated at least for short-range prediction. This suggested a possible qualitative difference in climate and weather prediction, because the Lorenz model is used often in the literature as a simple analog for the dynamical system underlying atmospheric systems (17). The severe spectral truncation of this model, however, has often led theoreticians in atmospheric and ocean dynamics (4) to consider higher order systems that exhibit a rather more stochastic as opposed to chaotic nature. In particular, the topographically forced barotropic potential vorticity equation has served often as a more realistic simple model of geophysical turbulence. This model can often be analyzed in terms of a statistical mechanical framework such as that described in ref. 18.

### A Simple Model of the Atmosphere

Recently an even simpler one-dimensional model of the type studied by Carnevale and Frederiksen (4) has been introduced and analyzed in some detail by A.J.M. and I.T. (19). It exhibits many of the desirable properties of the more complex atmospheric system but has the virtue of allowing a relatively complete analysis of statistical properties. The model is a spectrally truncated version of Burgers equation (referred to as truncated Burgers model or TBM),

$$u_t + \frac{1}{2} P_\Lambda (u_\Lambda^2)_x = 0, \quad [3]$$

where

<sup>§</sup>A Monte Carlo method known as ensemble prediction is commonly used in practical situations to attempt to approximate the prediction pdf. State-of-the-art numerical weather-prediction models have an order of  $10^7$  state variables, and thus this is a nontrivial exercise.

<sup>¶</sup>We are using the terminology of quantum mechanics here. The operator referred to is that for the corresponding dynamical system without stochastic forcing, which takes one from a state vector at one time to a new state vector at some later time.

$$u_\Lambda(t) = \sum_{|k| \leq \Lambda} u_k(t) e^{ikx} \quad u_{-k} = u_k^* \\ \text{and } P_\Lambda f \equiv f_\Lambda \equiv \sum_{|k| \leq \Lambda} f_k(t) e^{ikx}, \quad [4]$$

and typically values of at least  $\Lambda > 5$  are required for qualitative behavior to converge. In most of our previous work (and here)  $\Lambda = 50$ , and thus 100 real spectral components are retained. Majda and Timofeyev (19) showed that the equilibrium statistical mechanics of this model could be described by a canonical Gibbs probability measure,

$$G_\beta = C_\beta \exp \left( -\beta \sum_{k=-\Lambda}^{k=\Lambda} |u_k|^2 \right), \quad [5]$$

where  $\beta = \Lambda/E$ , with  $E$  the (kinetic) energy of the system, which can easily be shown to be conserved. (More precisely the energy is simply  $\int u^2 dx = \sum u_k^2$ .) The implication of Eq. 5 is that the equilibrium pdf for this model is Gaussian, with all spectral components having the same variance and zero mean and being statistically independent of each other.

The model has the interesting property that the decorrelation time scale of the spectral components is inversely proportional to their wave number. In other words, large-scale structures have low frequency variability and are much more persistent than the “weather modes” at smaller scales. Such a statistical property is a well known feature of the atmosphere and many other dynamical systems of physical interest (e.g., molecular biological systems). Furthermore this scaling behavior in the TBM is predicted by elementary theory and confirmed by numerical simulations.

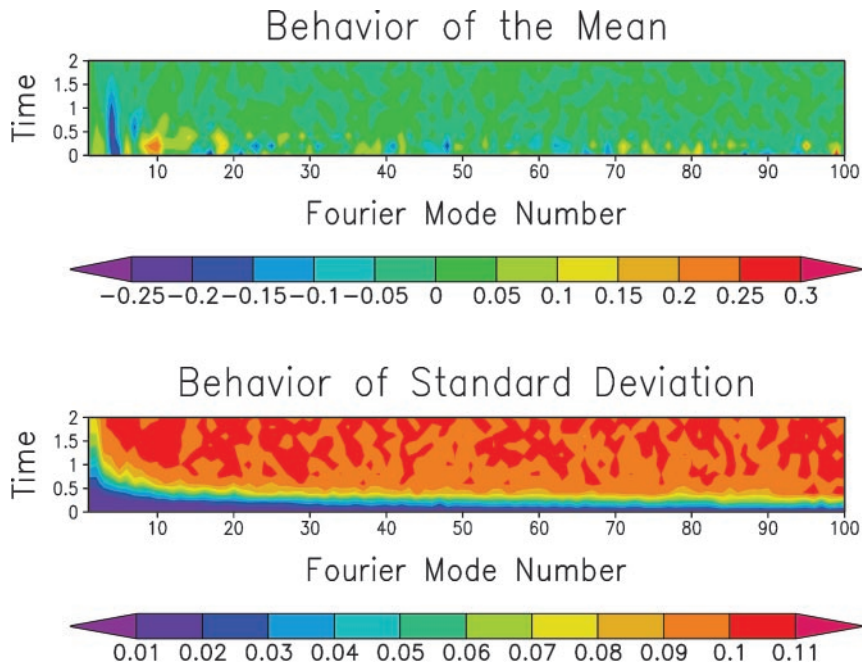
The two properties outlined above (a simple statistical equilibrium distribution and spatially scale-dependent decorrelation with many degrees of freedom) make the TBM a particularly attractive analog of more complex dynamical systems and an ideal vehicle to examine the developing ideas on predictability theory discussed above.

This article is organized as follows. The relaxation of the TBM toward its equilibrium distribution is analyzed in *Relaxation Behavior*. In *The Nature of Predictive Utility*, we use this analysis to examine the nature of predictive utility in the system and in particular study what determines variations in this quantity between different sets of initial conditions. As discussed, variation of predictability with initial conditions is a crucial theoretical and practical issue for dynamical systems. Finally, we provide a summary and discuss some research directions for the near future (*Summary and Discussion*).

### Relaxation Behavior

Statistical prediction may be viewed as the relaxation of a relatively tight probability distribution at the initial time toward an equilibrium distribution, which can be considered the climatological distribution. The initial time pdf can be considered as the uncertainty in the initial specification of the system’s state vector. In general, one would expect the mean of the initial-condition distribution to be drawn according to a pdf identical to that of climatology, because this is the historical distribution under the assumption of ergodicity. We adopt such an approach here.

Additionally, we assume that the initial-condition pdf is Gaussian, with a mean distributed as just discussed and a variance 4 orders of magnitude smaller than that of the equilibrium pdf (which also is Gaussian for the model currently under study, with variance of each mode equal to 0.1). This choice for the initial-condition variance is somewhat arbitrary and is in-

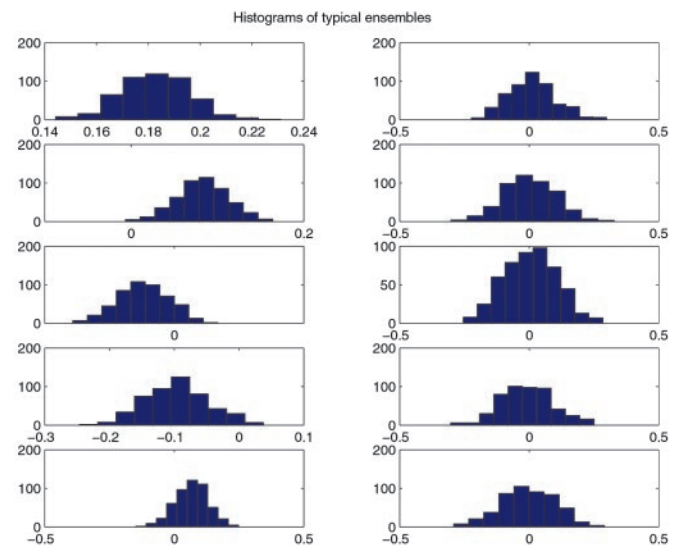


**Fig. 1.** Convergence of the first and second moments for a particular ensemble of predictions. The initial conditions are drawn from a Gaussian distribution with means drawn from the equilibrium distribution, which has a variance of 0.1 for each mode (see *Relaxation Behavior*).

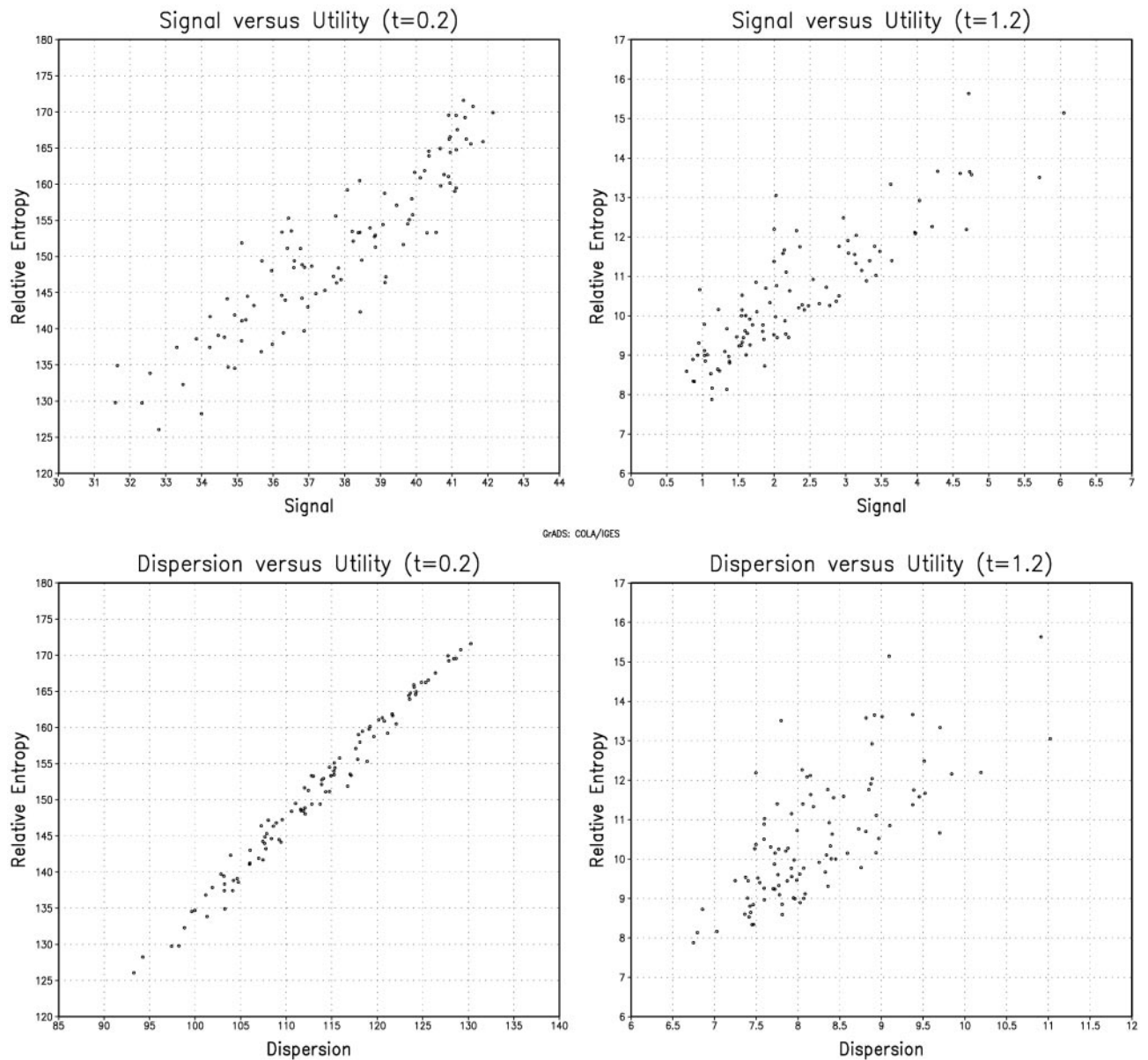
tended to represent the realistic scenario where uncertainty in the initial-conditions state vector is much less than the historical spread in the same vector (see below for further discussion). The relaxation behavior for a typical set of initial conditions is displayed in Fig. 1, which shows the evolution of the mean and standard deviation of the spectral components for a particular set of initial conditions. The quantities here are estimated by using ensemble methods. Thus, an ensemble of 500 members is used for this study, with initial conditions drawn according to the initial-condition pdf discussed above integrated forward in time until approximate equilibrium occurs. Each ensemble member represents a time integration of the TBM model. The technique of forward integration is a fourth-order Runge–Kutta scheme, and a pseudospectral technique is used to evaluate the nonlinear terms. As can be seen from Fig. 1 the smaller scale modes converge much more rapidly toward equilibrium for both first and second moments of the pdfs. Notice that the first moment can sometimes exhibit some oscillatory-like behavior as it converges to zero.

In general, the distributions governing the prediction pdfs appear to be approximately Gaussian at all lags. Fig. 2 shows the distributions for five of the modes at various prediction lags. The modes are chosen to be representative of the various spatial scales of the model. This degree of Gaussian behavior is rather surprising given the significant nonlinearity operating in the model (19). To check this result further we transformed the spectral modes (separately at all prediction times) to a basis in which all resulting modes were uncorrelated (the singular vector basis). Specifically this can be obtained by calculating the eigenvectors of the covariance matrix of the Fourier modes (for more detail see ref. 20). We then tested the Gaussianity of each transformed component using the Shapiro–Wilk  $W$  test (21). This latter reference explains in detail the derivation of the  $W$  statistic. Intuitively, if the data are plotted against a normal probability variate, then the  $W$  statistic represents the deviation from a straight line as measured by a correlation coefficient reduced from unity (it would be one in the case that the data were perfectly Gaussian). Results were computed (not shown) to determine when the test indicated non-Gaussianity at the 1%

confidence level. This was done for 100 different initial conditions at various prediction times. It was evident that only the final singular vector (which is dominated completely by small-scale features) showed any degree of non-Gaussian behavior and then only at small prediction times. Examination of the distribution for this singular vector shows that the deviation from Gaussian behavior takes the form of moderate bimodality (kurtosis). Interestingly, the first 10 (large-scale) singular vectors at such prediction time show a close correspondence with the first 10 spectral modes. For these large-scale “climate” modes the assumption of Gaussian behavior is universally an excellent approximation.



**Fig. 2.** Histograms of mode distributions of particular ensembles at particular times. (Left)  $t = 0.2$ . (Right)  $t = 0.8$ . Each row refers to a different Fourier mode from the model. From top to bottom, these are modes 5, 15, 25, 45, and 65.



**Fig. 3.** A scatter plot, for the 100 ensembles considered, of signal and dispersion versus relative entropy. The latter was calculated under the assumption that distributions are Gaussian and thus are the sum of signal and dispersion (see *The Nature of Predictive Utility*).

### The Nature of Predictive Utility

The approximately Gaussian nature of the prediction pdfs for the model under study considerably simplifies the (approximate) calculation of relative entropy. As discussed in the first section for the multivariate Gaussian case, the relative entropy is given by Eq. 2 and for pedagogical convenience may be decomposed exactly into two terms:

$$\text{dispersion} \equiv 0.5 * \left\{ \ln \left[ \frac{\det(\sigma_q^2)}{\det(\sigma_p^2)} \right] + \text{tr}[\sigma_p^2(\sigma_q^2)^{-1}] - n \right\}$$

$$\text{signal} \equiv 0.5 * (\tilde{\mu}_p - \tilde{\mu}_q)^t (\sigma_q^2)^{-1} (\tilde{\mu}_p - \tilde{\mu}_q)$$

As noted previously, the dispersion and signal measure rather different aspects of the prediction utility. In the case of “weakly” non-Gaussian distributions, it turns out to be still useful to determine which of these terms is important in determining relative entropy.

We conducted similar experiments with the TBM and calculated relative entropy according to Eq. 2 under the assumption that the prediction pdf is approximately Gaussian. If one were to drop this simplifying assumption, the *direct* calculation of relative entropy becomes generally prohibitively expensive (see ref. 7 for details about practicalities here), and new approximation methods are needed for systems with many degrees of freedom.

As was mentioned, we assumed that the initial pdf was Gaussian, with variance 4 orders of magnitude smaller than the equilibrium variance.<sup>||</sup> A consequence of this is that any initial conditions drawn from the equilibrium distribution will have essentially the same energy and consideration of Eq. 2, and the properties of the equilibrium pdf show immediately that this means that the time-0 signal term will also be equal for all initial

<sup>||</sup>The results described below are not qualitatively changed by varying the initial-condition pdf variance over 2 orders of magnitude.

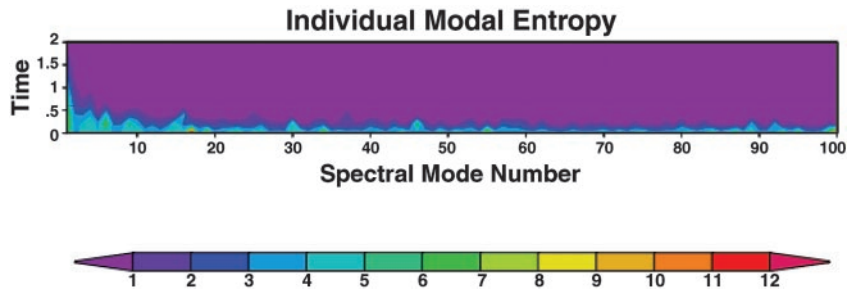


Fig. 4. The evolution of (univariate Gaussian) relative entropy with time for various Fourier modes and for a particular initial-condition ensemble.

conditions; thus, the dispersion component is automatically the most important measure of predictability variability at very short times. This property is somewhat artificial, because it is a consequence of the inviscid (conservative) nature of the TBM.

More realistic systems are obviously more dissipative and may not necessarily have this property.

Results for our set of 100 initial conditions are displayed in Fig. 3 for various prediction times. Recall that the prediction pdf

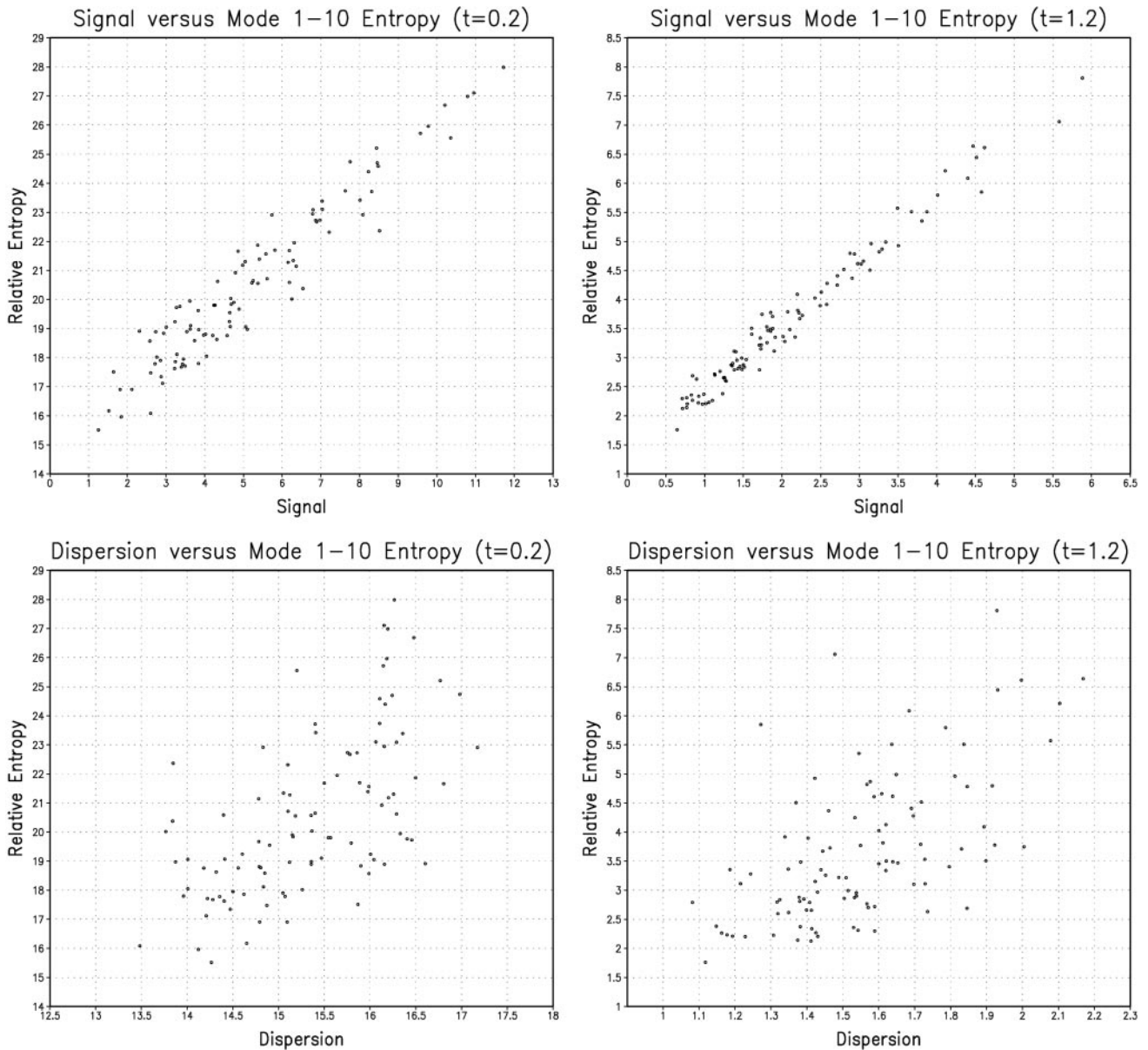


Fig. 5. The same as described for Fig. 3 but for multivariate Gaussian relative entropy of the first 10 Fourier modes (see text). These modes represent the large-scale, slowly evolving part of the flow.

statistics were obtained by using a 500-member ensemble. In ref. 19 it was found that a natural time scale in the TBM was that connected with shock formation from large-scale initial conditions (see figure 1 in ref. 19). This occurs at  $\approx t = 0.5$ , a time scale that is consistent with the relaxation process studied in the previous section. We shall refer to times shorter than this as short-range predictions and conversely for longer times.

In Fig. 3 we see that for short-range predictions, the signal and dispersion are of roughly equal importance in determining utility variation with initial conditions, but for longer ranges signal is somewhat more important in determining utility variability. Given the artificial nature of the dominance of dispersion at time 0, it is clear that signal is an important determinant of prediction-utility variability for the TBM. It is worth emphasizing that here we are interested in which parameter determines *variation* of utility with initial conditions and not the absolute value of the particular parameter. This latter quantity is somewhat arbitrary, because it depends on the assumption one makes about the tightness of the initial-condition pdf (a tighter value implies a higher absolute value of the dispersion and conversely).

In general, in prediction scenarios for climate one is interested in determining the large-scale component of the flow with low frequency variability. This separation of scales is the motivation for the climate-oriented stochastic modeling often used in studies of atmospheric dynamics (22–24). Stochastic modeling also is used extensively in climate systems that involve both the ocean and atmosphere (25–27). Here there is a much greater scale separation, with atmospheric transients providing the fast time-scale “stochastic” forcing for the slowly evolving ocean-controlled climate variables.

It is clear in the TBM that the large-scale spectral modes are much more predictable than the small-scale ones. This may be seen in Fig. 4 for one particular initial-condition set. Plotted is the evolution of utility as a function of spectral mode, and it is evident that the utility of the large-scale modes remains for a considerably longer period than the same quantity for the small-scale modes.

To examine the stochastic climate scenario we calculated the utility of the first 10 (large-scale) spectral modes. Fig. 5 shows the role of signal and dispersion in determining total large-scale utility. Rather strikingly, the signal component completely dominates utility variation at all prediction times. Similar results (not shown) also were found when even 20 and 40 modes were retained. This result indicates that in general signal is the main determinant of prediction utility in the TBM and that the equal signal/dispersion relation found for the total utility at short prediction times is really a consequence of the artificial constraint on initial conditions caused by the inviscid nature of the model, which automatically leads to the dominance of dispersion at very short times.

## Summary and Discussion

Relative entropy offers a very attractive means for quantifying the informational content of dynamical predictions. In the case that the probability distributions for both prediction and climatology (equilibrium) are Gaussian, a useful decomposition of this measure of utility into dispersion and signal is possible. In simple terms, the former measure the utility of uncertainty reduction through prediction, whereas the latter measures the degree to which the mean of a prediction differs from what one would expect in the absence of a dynamical prediction based on historical precedent.

Here we applied these ideas to a simple model with obvious similarities to the atmospheric dynamical system. The spectrally truncated Burgers equation has the property that large-scale structures are more persistent than those of small scale. In addition it has a particularly simple Gaussian equilibrium distribution, reflecting the fact that an equilibrium statistical mechanical formulation is possible. In addition we find that the prediction (nonequilibrium) distributions are also approximately Gaussian, which further facilitates the analysis of the system from the viewpoint of information theory.

We find that in general the signal component of relative entropy is significantly more important than dispersion. This result was particularly unexpected, because R.K. had found earlier (7) that dispersion was more important in the case of the Lorenz-63 (16) model. Given that the TBM system analyzed here is a many degree-of-freedom model with several important statistical features in common with the atmosphere that are absent in the Lorenz-63 model, this effect clearly deserves further investigation in more sophisticated models such as the barotropic potential vorticity equation. It is clear that if the current results hold in the more realistic context, then there are important implications for the rapidly developing field of statistical prediction. In particular, attention has focused to date in this field almost entirely on dispersion, and signal has been mainly ignored. Interestingly, this is not the case in climate prediction as noted in ref. 14.

Analysis in the present case was facilitated greatly by the approximate Gaussian nature of the prediction distributions. In the case of the Lorenz model such an assumption was not justified, because prediction distributions there are often highly bimodal (among other things). A priority of future work in applying information theory to dynamical prediction is the development of efficient methods for the calculation of entropy when many degrees of freedom are present in the system.

R.K. and A.J.M. thank Tapio Schneider for many useful conversations on predictability. R.K. was supported for this work partially by National Science Foundation Grant ATM-0071342 and National Aeronautics and Space Administration Grant NAG5-9871. A.J.M. is supported partially by National Science Foundation Grant DMS-9972865, Office of Naval Research Grant N00014-96-1-0043, and Army Research Office Grant DAAD19-01-10810. I.T. was funded as a postdoctoral fellow through the latter grants.

- Leith, C. E. (1974) *Mon. Weather Rev.* **102**, 409–418.
- Lorenz, E. N. (1965) *Tellus* **17**, 321–333.
- Epstein, E. S. (1969) *Tellus* **21**, 739–759.
- Carnevale, G. F. & Frederiksen, J. S. (1987) *J. Fluid Mech.* **175**, 157–181.
- Carnevale, G. F. & Holloway, G. (1982) *J. Fluid Mech.* **116**, 115–121.
- Schneider, T. & Griffies, S. M. (1999) *J. Clim.* **12**, 3133–3155.
- Kleeman, R. (2002) *J. Atmos. Sci.* **59**, 2057–2072.
- Cover, T. M. & Thomas, J. A. (1991) *Elements of Information Theory* (Wiley, New York).
- Ellis, R. S. (1999) *Physica D* **133**, 106–136.
- Gardiner, C. W. (1985) *Handbook of Stochastic Methods* (Springer, Berlin).
- Anderson, J. L. & Stern, W. F. (1996) *J. Clim.* **9**, 260–269.
- Palmer, T. N. & Tibaldi, S. (1988) *Mon. Weather Rev.* **116**, 2453–2480.
- Toth, Z. & Kalnay, E. (1997) *Mon. Weather Rev.* **125**, 3297–3319.
- Kleeman, R. & Moore, A. M. (1999) *Mon. Weather Rev.* **127**, 694–705.
- Kestin, T. S., Karoly, D. J., Yano, J.-I. & Rayner, N. A. (1998) *J. Clim.* **11**, 2258–2272.
- Lorenz, E. N. (1963) *J. Atmos. Sci.* **20**, 130–141.
- Palmer, T. N. (1993) *Bull. Am. Meteorol. Soc.* **74**, 49–65.
- Holloway, G. (1986) *Annu. Rev. Fluid Mech.* **18**, 91–147.
- Majda, A. J. & Timofeyev, I. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 12413–12417.
- Bretherton, C. S., Smith, C. & Wallace, J. M. (1992) *J. Clim.* **5**, 541–560.
- Royston, P. (1982) *Appl. Stat.* **31**, 115–124.
- DelSole, T. (1996) *J. Atmos. Sci.* **53**, 1617–1633.
- Majda, A. J., Timofeyev, I. & Vanden Eijnden, E. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 14687–14691.
- Achatz, U. & Branstator, G. (1999) *J. Atmos. Sci.* **56**, 3140–3160.
- Kleeman, R. & Power, S. B. (1994) *Tellus Ser. A* **46**, 529–540.
- Penland, C. & Sardeshmukh, P. D. (1995) *J. Clim.* **8**, 1999–2024.
- Kleeman, R. & Moore, A. M. (1997) *J. Atmos. Sci.* **54**, 753–767.