

# Limits, Variability, and General Behavior of Statistical Predictability of the Midlatitude Atmosphere

RICHARD KLEEMAN

*Courant Institute of Mathematical Sciences, New York University, New York, New York*

(Manuscript received 3 July 2006, in final form 19 March 2007)

## ABSTRACT

The nature of statistical predictability is analyzed in a T42 global atmospheric model that is able to adequately capture the main features of the midlatitude atmosphere. Key novel features of the present study include very large prediction ensembles and information theoretic techniques. It is found globally that predictability declines in a quasi-linear fashion with time for short-term predictions (3–25 days), while for long ranges (30–45 days) there is an exponential tail. In general, beyond 45 days the prediction and climatological ensembles have essentially converged, which means that beyond that point, atmospheric initial conditions are irrelevant to atmospheric statistical prediction.

Regional predictions show considerable variation in behavior. Both of the (northern) winter storm-track regions show a close-to-quasi-linear decline in predictability toward a cutoff at around 40 days. The (southern) summer storm track shows a much more exponential and considerably slower decline with a small amount of predictability still in evidence even at 90 days. Because the winter storm tracks dominate global variance the behavior of their predictability tends to dominate the global measure, except at longer lags. Variability in predictability with respect to initial conditions is also examined, and it is found that this is related more strongly to ensemble signal rather than ensemble spread. This result may serve to explain why the relation between weather forecast skill and ensemble spread is often observed to be significantly less than perfect. Results herein suggest that the ensemble signal as well as spread variations may be a major contributor to skill variations. Finally, it is found that the sensitivity of the calculated global predictability to changes in model horizontal resolution is not large; results from a T85 resolution model are not qualitatively all that different from the T42 case.

## 1. Introduction

Limitations to weather prediction have been a subject of intense investigation for many years. Lorenz (1969) put forward the idea that unresolved small scales in the initial conditions of a prediction would be transferred upscale and eventually swamp the synoptic scales. He verified this using a simple turbulence model with an upscale (inverse) energy cascade, which is typical of fully developed two-dimensional (2D) turbulence. This idea was further developed and verified using more sophisticated models of such turbulence by Leith (1971), Leith and Kraichnan (1972), and Kraichnan and Montgomery (1980), and recently in direct numerical simulations of an inverse energy cascade by

Boffetta and Musacchio (2001). Charney (1971), Kraichnan and Montgomery (1980), and Salmon (1998) have discussed in detail the close relation between classical 2D turbulence and the quasigeostrophic turbulence thought to be present in the midlatitude atmosphere and oceans.

It is often argued on theoretical (Lilly 1983) or observational (Nastrom and Gage 1985) grounds that at small scales in the atmosphere a forward, rather than inverse, energy cascade is more likely. The former cascades are more typical of 3D turbulence. If this is true, then there are evident implications for predictability limits, as discussed in detail in the review by Palmer (2000).

More recently, Tribbia and Baumhefner (2004) have revisited this problem using a suite of realistic general circulation models of varying resolutions. They found that error growth was particularly pronounced in a wavenumber band associated with baroclinic instability, and argued that disturbances/errors there drew sig-

---

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

nificant energy from the mean state. Disturbances in this wavenumber range were, in the view of these authors, “seeded” from the unresolved smaller scales in a manner similar to that proposed more broadly by Lorenz (1969).

A very natural general framework for considering all of these issues may be found in the idea (and practice) of statistical prediction (see, e.g., Leith 1974; Toth and Kalnay 1993; Ehrendorfer 1994a,b; Palmer 2000). Here, one implicitly assumes that initial condition variables are random, with a probability distribution defined by the observing procedure.<sup>1</sup> As prediction time increases this distribution evolves and eventually relaxes toward some kind of equilibrium distribution. Based on empirical evidence, one usually assumes ergodicity, which implies that the equilibrium distribution may be obtained from a very long time integration of the dynamical system. In the past five years there has been considerable theoretical study of this distributional relaxation process using the tools of information theory (see, e.g., Schneider and Griffies 1999; Kleeman 2002; Roulston and Smith 2002; Kleeman and Majda 2005; Del-Sole 2004; Kleeman 2007b).

In stochastic differential equation theory (e.g., Gardiner 2004), the natural tool for studying this relaxation of distributions is the relative entropy, which is a functional of the prediction and equilibrium distributions. Interestingly, it also happens to be a measure of the perfect model prediction utility (see Kleeman 2002). This is because the equilibrium distribution can be considered a reasonable *prior* distribution before observation of initial conditions and running of the prediction model while the prediction distribution constitutes an obvious *posterior* distribution. The quantitative measure of the shift between prior and posterior distributions given by the relative entropy is known in Bayesian statistical theory (see Bernardo and Smith 1994) as the *utility* of the process causing the shift, that is, in this case the prediction process.

Another way of viewing this relaxation process is as a measure of the importance of initial conditions to statistical prediction. The “closer” the prediction distribution is to the equilibrium<sup>2</sup> distribution, the less im-

portant initial conditions are to the statistical prediction. Such a viewpoint has an obvious application to the problem of atmospheric climate prediction, where the equilibrium distribution may be influenced by boundary conditions such as SST.

A great practical problem in implementing statistical prediction occurs because of the high dimensionality of the atmospheric dynamical system. In general, this means that one must resort to Monte Carlo sampling of the appropriate distributions, a process known as ensemble prediction [for particular practical approaches to this see, e.g., Toth and Kalnay (1993) and Palmer (2000)]. In general, for practical situations, the size of ensembles is considerably less than the dimension of the weather models being used. This implies a substantial loss of information relative to that which would be available if one had access to the full distribution. Recently, some theoretical progress has been made in rigorously analyzing this situation (see Kleeman 2007b). We summarize the relevant details from this analysis in the next section because it is particularly pertinent to our study here.

Given the uncertainties still apparent in our detailed understanding of predictability limits,<sup>3</sup> the theoretical developments detailed above were seen as an opportunity to revisit this problem from a rather different perspective. Another reason is that advances in computer capabilities now allow considerably larger ensembles than were previously possible, which enables us to carry through our theoretical program satisfactorily.

A final question of considerable practical interest concerns variations in predictability with initial conditions: Why are some predictions so much more skillful than others? A particular emphasis in the literature on this topic (see Palmer 2000) has been the so-called skill–spread relation where the skill of predictions is plotted against the ensemble spread. This reflects the intuitive idea that when the mean state is most unstable, predictions are least predictable (and vice versa). In general, while some such relation is often observed, it is frequently not as strong as one might expect (see Buizza and Palmer 1998). Recently in the climate prediction literature (see Kleeman and Moore 1999; Grötzner et al. 1999; Tang et al. 2005) it has been suggested that other factors in the initial conditions beyond instability may be primarily responsible for at least correlation skill variation. In particular, the amplitudes of especially “persistent” oscillatory patterns were found

---

<sup>1</sup> Naturally there are considerable challenges that have only been partially solved involved in deducing such distributions from the observational network.

<sup>2</sup> Note the rather important distinction here (unimportant to weather prediction) between equilibrium and climatological distributions; the former will be in equilibrium to the particular boundary conditions under consideration while the latter will generally be the equilibrium distribution given *mean* boundary conditions.

---

<sup>3</sup> An example of such is the precise relative roles of atmospheric initial and boundary conditions in atmospheric climate predictions of a particular duration.

to be significant. Our theoretical framework enables us to examine this question now for weather prediction.

This paper is structured as follows: Section 2 discusses the theoretical predictability tools to be used as well as the atmospheric model chosen for analysis. Section 3 examines the relaxation behavior of atmospheric ensembles as well as the factors behind variations in prediction utility with different initial conditions. Section 4 further analyzes the origins of the relaxation behavior, and looks at regional variations in predictability as well as the issue of the resolution dependency of results. Section 5 contains a discussion, a summary, and plans to further investigate the findings of this research.

## 2. Numerical model and theoretical tools

In this contribution we shall focus on the predictability of the midlatitudes and ignore the effects of convection. This choice is motivated by two considerations: First, the depiction of convection in present weather and climate models is still usually highly parameterized because it needs to describe some very important unresolved scales of motion in a statistical sense. In particular, the cloud scale of several kilometers is beyond the capabilities of the present generation models. Given this, it is not surprising that there are often significant problems in the simulation of important convective variability, such as the Madden–Julian oscillation (see, e.g., Lin et al. 2006). Second, physical parameterizations add significantly to computational expense, and here we wish to consider very large ensembles. We shall examine the limitations of our choice herein in the discussion section below.

Motivated by this we chose to use the first-generation Portable University Model of the Atmosphere (PUMA) primitive-equation model developed at the University of Hamburg (see Leslie and Fraedrich 1997). The configuration had five vertical ( $\sigma$ ) levels and, in our standard experiments, a spectral resolution of T42, which amounts to approximately  $3^\circ$ . Radiative cooling and convection are depicted by a simple Newtonian linear damping toward a specified vertical profile. The relaxation temperature profile is zonally uniform but varies meridionally in such a way as to simulate a northern winter season. Vertically, the temperature relaxation profile is close to a moist adiabat in the Tropics, but becomes more stable at higher latitudes. Details are provided in appendix B. Momentum dissipation is included with a simple linear Rayleigh friction term, which is enhanced in the lowest vertical level. The model also includes a representation of orography, which has the effect of making the mean state zonally variable, and in particular in the Northern

Hemisphere this locates storm tracks in the North Atlantic and North Pacific. The climatological performance of the model configuration is detailed in Kleeman (2007a) and is reasonably realistic. It locates the jet streams of both hemispheres in approximately the correct meridional location, and their strength is reasonable compared to northern winter observations. There is a noticeable and realistic zonal variation in the northern jet stream, with the jet being strongest over the North American/Atlantic sector. Pointwise variance of the streamfunction is concentrated in the North Atlantic and North Pacific midlatitudes in approximate agreement with observations. A qualitative comparison of synoptic disturbances in these regions with observations shows them to be reasonably similar in horizontal structure and propagation direction. The theoretical machinery to be utilized here can be found in Kleeman and Majda (2005) and Kleeman (2007b). We summarize the relevant material here; however, proofs and considerably more detail and discussion can be found in these sources.

The perspective we adopt is that ensembles represent a sampling of an underlying and unknown probability density function (pdf). This sampling philosophy also implies a discretization reference frame because one must define a partitioning of state space in order to use the ensemble as a sample estimate of the (now discrete) pdf. Such a partitioning naturally implies that the pdf is then only being estimated at a particular resolution. Now if one chooses a partitioning such that very few ensemble members are contained in any particular partition element (i.e., subregion of state space), then the sample estimate provided may have considerable sampling error associated with it. In Kleeman and Majda (2005) it was shown that this error could be quantified as information loss and then compared with the information content of the ensemble. It was argued that the choice of partitioning must ensure that this loss should be small relative to the information content being calculated. In practical terms this means that the number of ensemble members per partition is required to be greater than approximately five.

It also should be observed that for a sufficiently large sample, for which this information loss is not an issue, a refinement of the partition results in increased information (see Cover and Thomas 2006; Kleeman 2007b) because one is taking into account more of the fine structure of the pdf. There is, therefore, a trade-off between losing information because of sampling error and losing it because of coarse graining. In general, the optimal point occurs when there are around five ensemble members in partitions. If one ensures as part of a partitioning choice that this condition is met, then the

loss of information resulting from sampling is much smaller than the practical calculated entropies discussed below.

Consider now the importance of state-space dimensionality: It is clear that as the dimension of the state space increases the number of partitions *per dimension* must shrink rather rapidly so that sampling information loss is not significant. The implication, then, is that the degree of multivariate behavior possible for pdfs at adequate resolution is rather severely constrained by ensemble size. In fact, a simple calculation shows that for practical ensembles (i.e., with sizes of less than  $10^4$  or so) the degree of multivariate behavior possible with four partitions per dimension is at most around six.

One might ask whether this problem could be avoided by fitting certain reasonable distributions (such as Gaussians or their generalizations) to the ensemble. One finds, however, that the constraints described above then appear in a different form: First, the information content obtained from the fitted distributions are, for high-order multivariate distributions, considerably larger than those obtained using the optimal partitioning strategy discussed above. The apparent reason for this is that there must be considerable sampling error or information loss associated with the particular fitted distribution because it is rather uncertain. This uncertainty also implies that the numerical problem associated with the fitting can often become quite ill posed because there are many almost equally well-fitted distributions.

Given these issues, Kleeman (2007b) proposed the study of the information content of an ensemble associated with multivariate behavior of a particular degree. More precisely, suppose we have  $n$  random variables  $X_i$  with corresponding multivariate distribution  $p(X_1, X_2, \dots, X_n)$ . Consider now all possible marginal distributions of order  $m < n$   $p(X_{j_1}, X_{j_2}, \dots, X_{j_m})$ , with  $j_k \leq n$ . We define now the *marginal relative entropy* of degree  $m$  by

$$D^m(p\|q) \equiv \frac{1}{C_m^n} \sum_{j_1, j_2, \dots, j_m} D[p(X_{j_1}, X_{j_2}, \dots, X_{j_m}) \| q(X_{j_1}, X_{j_2}, \dots, X_{j_m})],$$

where  $q(X_{j_1}, X_{j_2}, \dots, X_{j_m})$  is the corresponding equilibrium distribution and the summation is over all possible distinct permutations  $j_1, j_2, \dots, j_m$ . For a given partitioning of state space the marginal entropies can also be shown to satisfy an inequality hierarchy, that is,

$$D^1(p\|q) \leq D^2(p\|q) \leq \dots \leq D^n(p\|q) = D(p\|q), \quad (1)$$

where it will be noted that the final member of the chain is the full relative entropy. This hierarchy of inequalities has the following natural interpretation: each marginal relative entropy is measuring the average information content involved in all possible multivariate fluctuations of given order  $m$  of the  $n$  random variables. Intuitively, as one accounts for higher-order fluctuations (bivariate, trivariate, and so on), one would expect to account for more and more of the information content of the full  $n$ -dimensional pdf. This is precisely the reason for the inequality hierarchy above. When the final point  $m = n$  is reached, then all multivariate variations are accounted for and so the full relative entropy is recovered.

Consider now the optimal partitioning strategy discussed previously. It is clear that the number of partitions per dimension will then strongly decrease as we consider the calculation of marginal relative entropies of increasing degree. Remembering that because refinement of partitioning implies an *increase* in relative entropy, it is clear that if we adopt an optimal partitioning strategy then the hierarchy of Eq. (1) may no longer apply. We will examine this point again in the next section when we deal with practical atmospheric ensembles.

Another approach to the above ensemble information content methodology is to assume in the limit of an infinite ensemble that a particular form of distribution holds. In the present context such a calculation is interesting and of practical significance because, as we shall see later, it is generally the case that the “converged” distributions likely appear to be close to Gaussian. We shall therefore also calculate the relative entropy under the assumption that the converged distributions are Gaussian. This is of course a sample estimate because we are estimating the covariances and means from our ensembles.<sup>4</sup> It also represents the information associated with only the first two cumulants of the converged distribution in which no assumptions regarding form are made (see Majda et al. 2002 for more detail). One could make less restrictive assumptions than Gaussianity; however, it is not clear exactly what form these should take, that is, which higher-order cumulants should be included and which should be neglected. This is a subject for future work. A motivation for the Gaussian assumption is the central limit theorem, which states that a sufficiently large sum of non-Gaussian random variables results in a nearly Gaussian

<sup>4</sup> For the large ensembles we shall deal with the sample error here is small. The actual error can be estimated by looking at the very long time prediction value of the relative entropy, as we shall see below.

variable. It appears that as the dimensionality of the dynamical system increases, as it does for realistic systems, then apparently many random variables of interest are effectively large sums of other random variables and so by the central limit theorem approximately Gaussian.

### 3. Initial conditions probability distribution

Errors arise in the initial conditions of weather predictions for two principal reasons—first because the observational networks that determine the current state of the atmosphere are less than perfect, and second because the models utilized to make forecasts have misrepresentations of physical processes in the atmosphere. This is primarily due to a lack of model resolution, which forces the parameterization of unresolved scales of motion.

In this study we are ignoring errors of the second kind by making the perfect model assumption. We thus are considering errors that arise solely from the imperfections in the observational network.

In practical weather prediction many methods have been used to process observational data for the purpose of initializing prediction models. The following three main types of techniques have been utilized:

- 1) optimal interpolation of earlier predictions and observations (see, e.g., Lorenc 1986),
- 2) variational assimilation of observations into a model over a particular time window (see, e.g., Rabier et al. 2000), and
- 3) Kalman filtering a combination of forward integrations of a model with observational data (see, e.g., Houtekamer and Mitchell 1998).

In all of these cases the prediction model itself is used in a fundamental way to “fill in the gaps” of the observational network. This interpolation is required to adequately initialize a particular prediction model on its numerical grid.

As a consequence of following such analysis methods, error estimates for the initial conditions reflect aspects of model dynamics as well as deficiencies in the actual observational network, which we mentioned above. Disentangling these combined effects is evidently a complex, interesting, and important subject of study. Because our primary focus in this contribution is upon the statistical evolution of errors and not on the initialization process, we defer a comprehensive and rigorous analysis of this situation and its effect on statistical predictability to a future contribution.

In place of the sophisticated methodologies of initialization mentioned above, we shall instead fix our ini-

tialization distribution by appealing to a rather idealized picture of deficiencies in the observation network; in particular, we shall assume that observations of prognostic variables are available only at a coarse horizontal resolution, which amounts to an integrated knowledge of variables in a particular “observation box.” The latter implies naturally errors in the pointwise-defined initial conditions. The coarse resolution of the box also implies that errors in model variables at grid points with close horizontal proximity will show significant correlation with each other. Using these basic precepts we build a multivariate Gaussian distribution for our initial conditions. The particular assumed horizontal structure is similar in several respects to that assumed in optimal interpolation for forecast errors (see Lorenc 1986).

In addition to the above covariance structure we also assume that the means for the initial condition distribution are obtained at random from an extended integration of the prediction model. This ensures that predictions are always close to the natural “attractor” of the prediction model. This prescription is also conceptually consistent with our perfect model assumption. The exact details and a justification of the distribution may be found in appendix A.

### 4. Limits and variability of predictability

The theoretical framework discussed above was now applied to the model detailed. Ensembles were constructed by sampling from a particular choice for the initial condition distribution discussed above. We shall assume, as discussed, that the primitive-equation prognostic variables are distributed according to a Gaussian (multivariate) distribution, which has a horizontal decorrelation scale of 1000 km and mean values drawn randomly from the equilibrium distribution with variances two orders of magnitude smaller than the equilibrium distribution. The variances chosen here are probably larger than one might expect in a well-observed region.

In the first numerical experiment, which explored predictability limits, we chose a very large ensemble of 9600 members. The large size was used in order to explore convergence with respect to the order of computed marginal relative entropy. In later experiments, aimed at examining variations in predictability with respect to the choice of initial conditions, and also in a higher-resolution run, we reduce this to 1000 members and compute only marginal entropies of order three or less.

To make the calculation of marginal relative entropy practical we require a reduced state space, because all possible permutations of state variables must be com-

puted [see Eq. (1)]. Accordingly, we calculated the first 60 EOFs from an equilibrium run using a multivariate calculation that used the full set of prognostic variables (divergence, vorticity, temperature, and the log of surface pressure) at all five vertical levels. The four dynamically different state variables were rescaled to have equal mean global variance before the EOF calculation was performed in order that all should be given equal weight in the reduced state space. The divergence was rescaled with the same factor as the vorticity because they have the same physical units. Without such a rescaling, results are dominated by the temperature variable. We also sampled the equilibrium run every 90 days to ensure temporal correlation. The set of calculated EOFs account for approximately 90% of the variance of the complete set of prognostic state variables. It was found that all results shown below were not qualitatively sensitive to cutoff dimension, a result checked by examining all results using 20 and 40 EOFs. The first EOF is shown in Fig. 1. Notice that overall it has the greatest amplitude in the northern jet stream and storm-track regions. It is also interesting to note that our results were insensitive to the state-space variable(s) selected. We also examined just streamfunction as a state-space variable and found little qualitative difference in the results.

As noted previously, an important practical consideration in the calculation of entropic functionals is the partitioning used. We chose to use the finest partitioning consistent with ensuring that the information loss resulting from sampling is reasonably small. Such a strategy is optimal in extracting information. We also chose to partition based on retaining an equal number of prediction ensemble members in each dimension partition. The number of partitions per dimension for the marginal entropy calculations in the first experiment is listed in Table 1.

#### *a. Basic behavior*

The results for a particular initial condition distribution are shown in Fig. 2. Plotted are the first four marginal entropies as well as the hypothetical Gaussian relative entropy discussed above.

The latter has been divided by 60 to facilitate comparison with the marginal entropies. It is much larger because it represents the information contained in a hypothetical converged distribution while the marginal entropies reflect the (severely) reduced information content of an ensemble. It is rather curious that despite this difference in calculation basis it appears to behave very similarly after rescaling. The correspondence is not perfect however, as we shall see later when we examine predictability variation. All entropic measures

qualitatively show the same behavior, namely, a quasi-linear drop for the first month or so of a prediction and some evidence of an exponential tail between 30 and 50 days. Beyond this there is approximately flat behavior for all measures. The level beyond 60 days is not zero, as one might expect from the convergence of prediction and climatological distributions, because of the issue of ensemble sampling. Thus, if one takes two different climatological ensembles of size 9600 and calculates the marginal and Gaussian relative entropies then one obtains a small positive residual equal to that seen in the long lead values of Fig. 2.

A number of other features are notable in the behavior of the entropies. First, for longer lags there is little difference in the behavior of the second- through fourth-order marginal entropies, suggesting that some kind of rough convergence is occurring. Second, there is a hint of “leveling off” of the higher marginal entropies for short lags. This is not apparent for the univariate and Gaussian measures. The difference is due to the relative coarseness of resolution used for the higher-order measures. For short lags the partitioning used ensures that the coarse cases do not resolve well the climatological distribution. Recall that partitioning is chosen with respect to the prediction distribution. For longer lags as the ensembles approach each other in spread the partitioning resolves better both distributions.

#### *b. Regional variations in predictability*

We next examined the variation of predictability behavior in different storm-track regions of the globe. In particular, we examined the Southern Ocean ( $70^{\circ}$ – $28^{\circ}$ S and all longitudes), the North Atlantic ( $25^{\circ}$ – $56^{\circ}$ N,  $76^{\circ}$ – $174^{\circ}$ W), and the North Pacific ( $25^{\circ}$ – $56^{\circ}$ N,  $143^{\circ}$ E– $115^{\circ}$ W). The first region is dynamically very different from the last two because in this study it lies in the summer hemisphere and thus has a mean state likely to be much less baroclinically unstable. The last two regions have very similar mean states.

In each regional area we recalculated the first 60 EOFs in order to obtain a dynamically relevant reduced state space in each case. The explained variance of these new sets of 60 EOFs was higher than the corresponding global set. The ensemble predictions were identical to the global case. We simply restricted attention to the specific region and projected onto the regional EOFs rather than the global ones.

Predictability results are displayed in Fig. 3 and show that the summer storm-track region indeed behaves very differently from the two northern winter regions, which appear almost identical in their behavior. In particular, the summer region shows a much slower decline

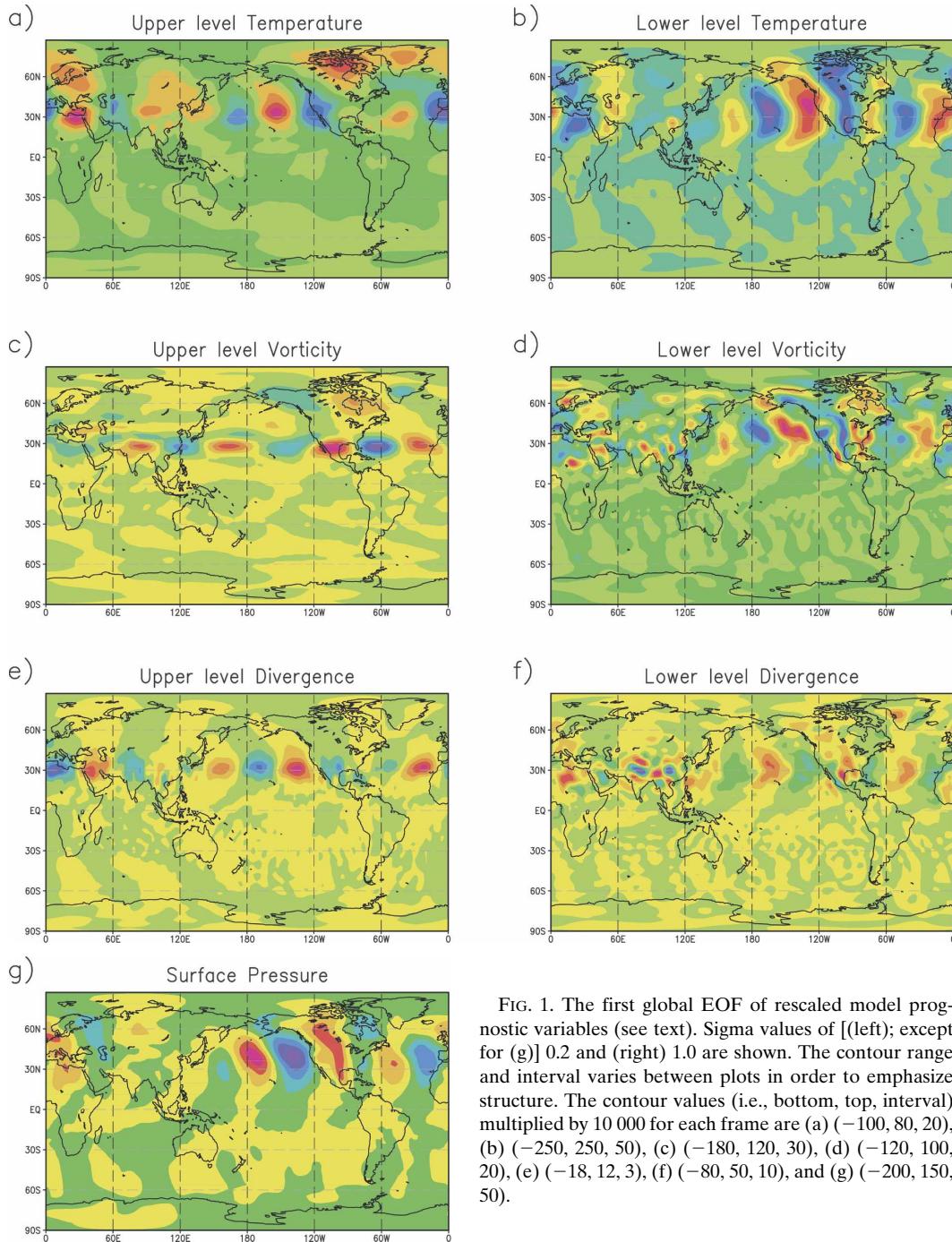


FIG. 1. The first global EOF of rescaled model prognostic variables (see text). Sigma values of [(left); except for (g)] 0.2 and (right) 1.0 are shown. The contour range and interval varies between plots in order to emphasize structure. The contour values (i.e., bottom, top, interval) multiplied by 10 000 for each frame are (a) (−100, 80, 20), (b) (−250, 250, 50), (c) (−180, 120, 30), (d) (−120, 100, 20), (e) (−18, 12, 3), (f) (−80, 50, 10), and (g) (−200, 150, 50).

in the relative entropy measures with statistical convergence of the prediction ensemble to the climatological ensemble still not quite complete after 90 days. The two northern winter regions by contrast showed somewhat faster predictability decline than the global case, with convergence almost complete by approximately 30 days. There is also a general impression that the predictability decline in the summer case is exponential in

character, while the winter case is almost quasi linear with a well-defined cutoff point in time beyond which there is no statistical predictability due to initial conditions.

*c. Variation in predictability with initial conditions*

This issue is of some obvious practical interest because a priori information about the likely skill of a

TABLE 1. Partitioning of state space for the various marginal relative entropies.

Marginal order	Partitions per dimension	Total partitions
1	1024	1024
2	32	1024
3	10	1000
4	6	1296

particular forecast has utilitarian consequences [a good review can be found in the work of Palmer (2000)]. Considerable attention in the literature in this respect has focused on the relation between prediction skill and ensemble spread. Some useful results have been obtained (see, e.g., Buizza and Palmer 1998); however, the relationship found has often been less than ideal.

Our present study allows us to reexamine this issue using rather clear theoretical measures of predictability with large ensembles. The focus here has been on using the relative entropy as a measure of the convergence of the prediction to the climatological ensemble, however, one may also interpret it clearly as the perfect model utility of the statistical prediction. This is the case because it measures the overall shift in expected prediction ensemble once initial conditions and dynamics are taken into account (see Kleeman 2002; Bernardo and Smith 1994).

If one assumes Gaussian ensembles, then this utility measure involves both the first and second cumulants of the ensembles, which makes it more similar to an anomaly correlation rather than an RMS skill measure (see Kleeman and Moore 1999) which only depends on the second cumulant.<sup>5</sup> In fact, one may write the Gaussian relative entropy as the sum of a piece depending only on the prediction's second cumulant (dispersion) and a piece depending only on the prediction's first cumulant (signal):

$$\text{RE} = \frac{1}{2} \left\{ \ln \left[ \frac{\det(\sigma_p^2)}{\det(\sigma_q^2)} \right] + \text{tr}[\sigma_p^2(\sigma_q^2)^{-1}] - n \right\} \text{Dispersion} + \frac{1}{2} (\boldsymbol{\mu}^p - \boldsymbol{\mu}^q)^t (\sigma_q^2)^{-1} (\boldsymbol{\mu}^p - \boldsymbol{\mu}^q) \text{Signal}, \quad (2)$$

where the subscripts  $p$  and  $q$  refer to prediction and climatological ensembles, respectively. The interpretation of the first piece is the prediction utility associated with reduction in uncertainty, while the second piece

<sup>5</sup> Recall that we are considering perfect model skill measures here. Naturally, in the case of model error the RMS error will also depend on the mean as well.

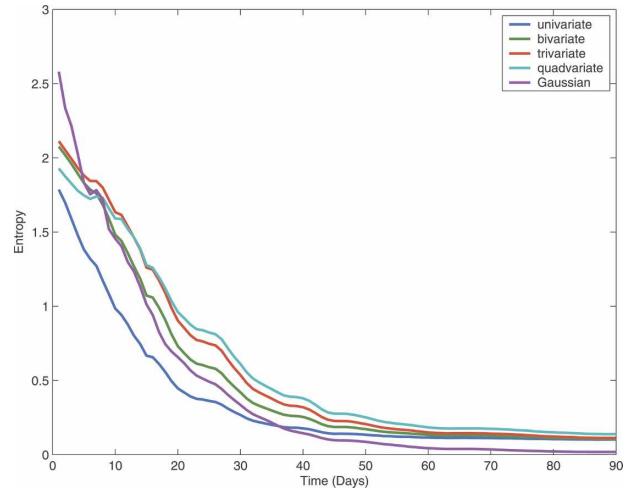


FIG. 2. Functional describing global predictability at various prediction times. Shown are various orders of marginal relative entropy (see text) as well as the hypothetical Gaussian relative entropy that applies if the converged distributions are Gaussian. The Gaussian functional has been rescaled by dividing by 60 for ease of viewing.

represents the utility associated with shifts in the means of the ensemble from the climatological case. Some reflection shows that both pieces have clear practical value to consumers of forecasts.

An interesting question is whether variations in prediction utility with respect to differing mean initial conditions are more closely associated with variations in either dispersion or signal. Clearly, an attempt to find a relationship between ensemble spread and anomaly correlation implicitly assumes that the former is the case. In fact, we find that the opposite is true.

To examine this we obtained 48 uncorrelated initial conditions from a climatological integration, which were used as the means for initial condition ensembles constructed using the method described previously. To make this exercise computationally practical a smaller ensemble of 1000 members was then constructed for each initial condition. The second and third marginal entropies were then calculated for each initial condition using 5 and 10 partitions per dimension. The Gaussian entropy was also computed and split according to Eq. (2).

In Fig. 4 we have plotted the correlations between the various entropies at different prediction times. The relationship between the marginal entropy and the Gaussian entropy is always quite strong (usually 0.8–0.9), which is evidence of the near-Gaussian nature of ensembles in atmospheric prediction. Note, however, that it is not as high as the correlation between different marginal entropies, suggesting some small but nonzero role for non-Gaussianity. This is only a tentative con-

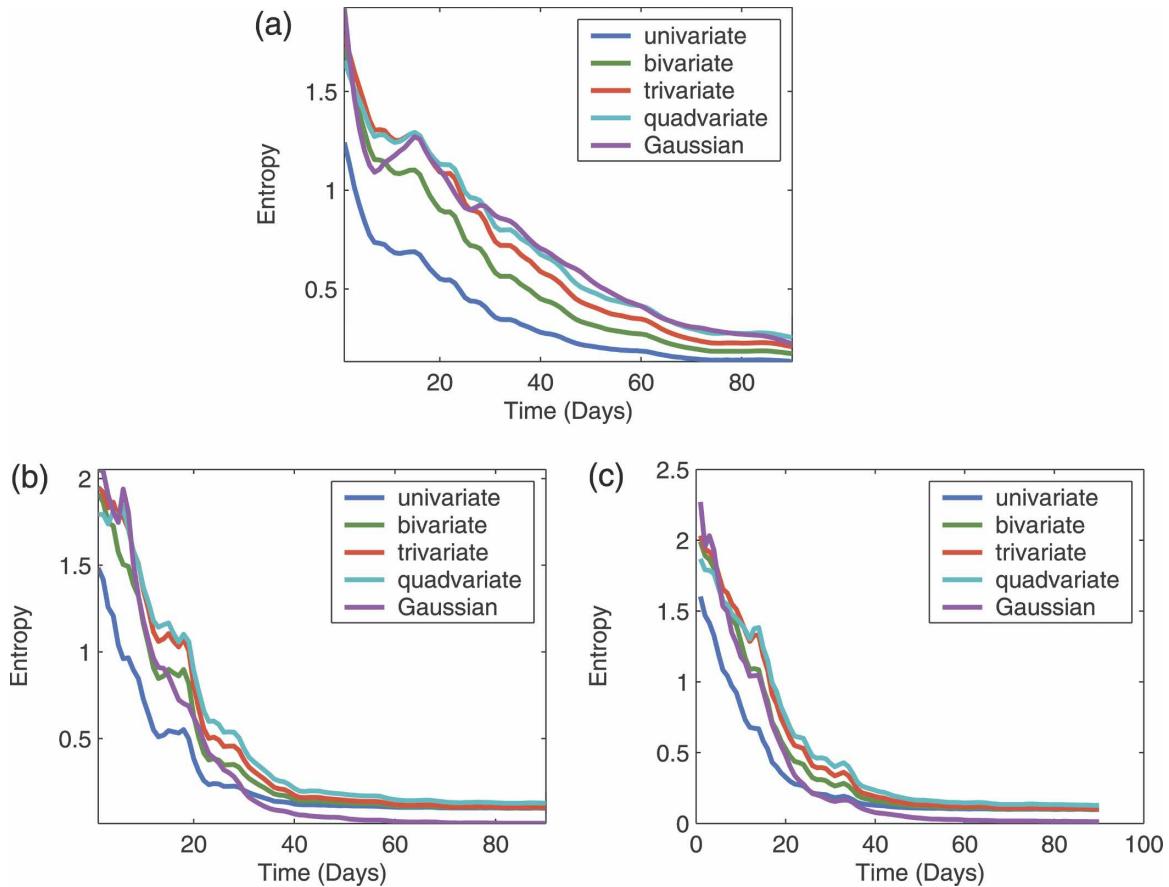


FIG. 3. Same as for Fig. 2, but broken into different geographical regions of particular dynamical interest: (a) Southern Ocean, (b) North Atlantic, (c) North Pacific.

clusion of course, and requires further detailed investigation that the author is currently undertaking and will document elsewhere.

The interesting aspect about Fig. 4 is that of the two Gaussian pieces, it is very clearly the signal that is more closely associated with marginal entropy variation than the dispersion. This holds at all prediction times, but most noticeably for shorter-range, more skillful predictions. In this respect the nature of primitive-equation midlatitude atmospheric predictability variation appears more similar to that seen previously in baroclinic quasigeostrophic turbulence (see Kleeman and Majda 2005), climate prediction (see Tang et al. 2005), and stochastically forced systems (see Kleeman 2002) rather than in chaotic systems like the Lorenz (1963) model (see Kleeman 2002). This idea is not new to weather prediction (see, e.g., van den Dool and Toth 1991), however the present work confirms it in a very general statistical predictability setting with very large ensembles.

Given this important role for the signal, it is interesting to view its variation with initial condition and pre-

diction time. This is plotted in Fig. 5, and two things are apparent. First, there can be considerable variation in this quantity from one set of initial conditions to another, and occasionally it is unusually high even for 2-week predictions. Second, it is reasonably coherent in time or, in other words, if the signal is large for short-range predictions it will tend to be also large for longer-range predictions (and vice versa).

An important caveat to the above results concerns the method used to produce the initial condition distribution. As noted in appendix A, the fixed covariance matrix assumed will ensure that at time zero there will be no variation in the dispersion component between differing initial conditions. With a more practical assimilation system one would not expect this to be the case because model dynamics then play a role in determining error covariances. It is worth observing, however, that the conclusion regarding dominance of the signal over dispersion is true at all prediction times, not just very short ones. Evidently, for lags of the order of 2–10 days, the prediction distribution is very heavily influenced by model dynamics in a way that is some-

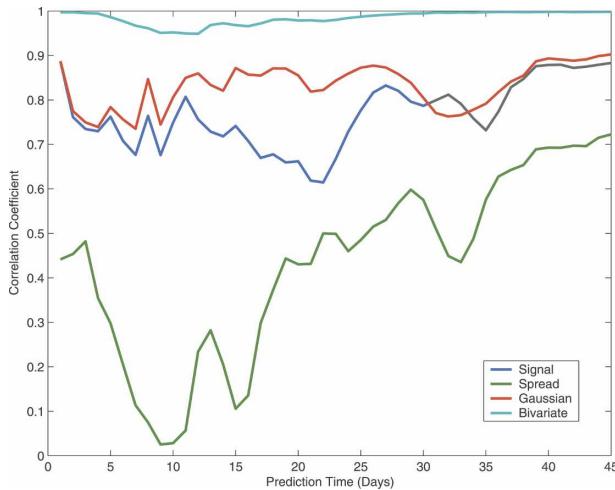


FIG. 4. Correlation of the third-order marginal relative entropy with various other functionals of interest. Correlations were computed by considering different initial conditions as data points and are plotted as a function of prediction time. Other functionals considered include the second-order (bivariate) marginal entropy, and the Gaussian relative entropy (see Fig. 1) and its constituent pieces, signal and dispersion (see text).

what analogous to practical data assimilation, and yet the signal term actually is increasingly dominant over dispersion during that period. The dispersion term however does grow from zero to a correlation of 0.45 in the first day, which does suggest that practical assimilation systems may have a significantly larger role for dispersion, at least for short-range predictions.

#### d. Sensitivity to resolution

In general in models of the type deployed here, scale-dependent dissipation is used to control small-scale noise. As resolution increases this typically acts primarily on smaller and smaller scales, which are well below the synoptic scales that dominate the leading part of the EOF spectrum that we have used to analyze predictability. Effectively, then, as we increase resolution we reduce dissipation of modes in this part of the spectrum, and at the same time increase the effective stochastic backscatter from previously unresolved smaller scales (see, e.g., Frederiksen and Davies 1997). We tested the sensitivity of our results to these effects by using a T85 rather than T42 version of the PUMA model. The effective horizontal resolution of this configuration was exactly double that of the standard model. We kept the vertical resolution at the standard five levels, and in all other respects, except for internal dissipation, the new configuration was identical to the standard one.

The internal dissipation used in the model was an order-8 hyperviscosity. The coefficient for this term was

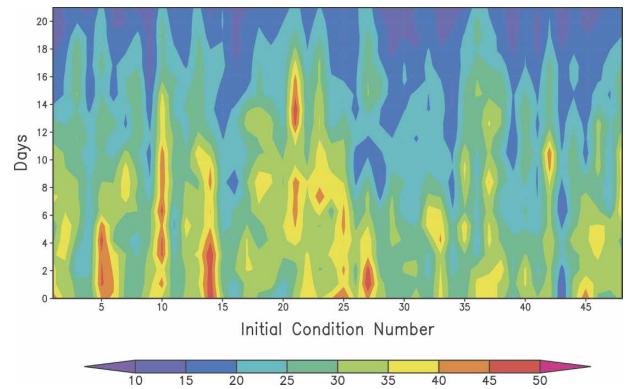


FIG. 5. The signal component of Gaussian relative entropy plotted as a function of prediction time and particular initial condition (there are 48 considered).

adjusted with resolution so that the dissipation time scale of the highest wavenumber remained constant at 4 days for both the T85 and T42 experiments.

Increasing resolution forces one to choose a significantly shorter time step for numerical stability, meaning that the computational cost of our high-resolution model is more than an order of magnitude higher. We thus restricted our attention to smaller ensembles, and in particular considered a 960-member ensemble.

At a higher resolution an EOF analysis similar to that reported above for T42 revealed, perhaps not surprisingly, a slower convergence with respect to explained variance. Thus, 100 modes explained around 55% of variance, rather than the 95% in the T42 case. Our experiments were conducted with 100-mode-reduced state space. As an initial condition distribution we used the identical one to that used in the T42 case to produce Fig. 1. The reduced ensemble meant that we were only able to calculate the first three marginal entropies, and these were performed using 121, 11, and 5 partitions per dimension, respectively. Again, the Gaussian relative entropy was also calculated and rescaled by the (new) reduced space dimension for ease of viewing. Results are displayed in Fig. 6 and show that, similar to the T42 case in Fig. 2, for all considered entropic functionals there is essential convergence of prediction and climatological ensembles by about 45–50 days. The general qualitative behavior is relatively unchanged, although there is some reduction in all relative entropies at short leads. This is possibly because the smaller total variance of the reduced state space for the high-resolution case may mean that more information is being omitted in the present case relative to the control case. The relative robustness of the results to rather large changes in horizontal resolution is a little surprising. It suggests that the stochastic backscatter from very small-scale

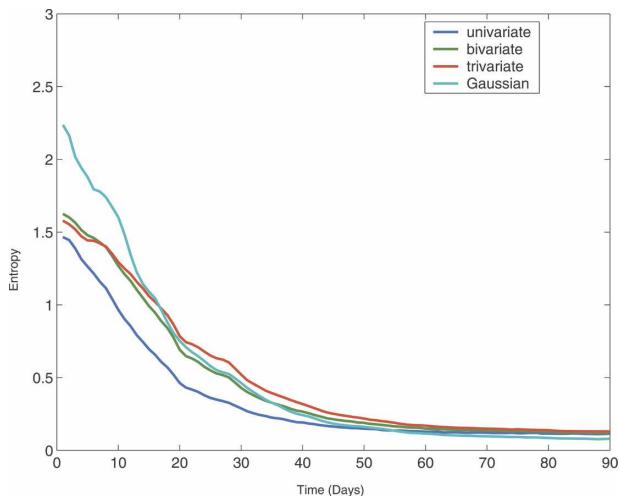


FIG. 6. Same as Fig. 1, but for a high-resolution (T85) version of the numerical model. The Gaussian marginal entropy has here been rescaled by 100 rather than 60 (see text). The fourth-order marginal relative entropy has been omitted because smaller ensembles were used.

wavenumbers is not very important to the basic statistical predictability properties of the atmosphere. Naturally, more detailed study of this issue is warranted.

**5. Summary and discussion**

In this contribution, for the first time we have investigated perfect atmospheric statistical predictability comprehensively using very large ensembles and tools from information theory. The model used is quite simplified compared to state-of-the-art numerical weather prediction models; however, it exhibits reasonably realistic behavior in the midlatitude storm tracks.

Our principal novel result is that in the winter storm tracks, model prediction and climatological ensembles are essentially indistinguishable beyond approximately 45 days. This has important consequences for atmospheric climate prediction because it suggests that beyond this time there is no value in incorporating initial condition data, and only boundary condition data are relevant to predictions. This qualitative result appears robust to significant increases in the horizontal resolution of the model used. It also seems robust to the dynamical variables used in the predictability analysis (see Kleeman 2007b).

A further novel result concerns regional variation in potential predictability. We find that in the southern summer hemisphere storm tracks, statistical predictability is at least twice as long as in the northern winter hemisphere. In addition, the decline of predictability as measured by the relative entropy of the prediction and

climatological ensembles was of a much more marked exponential character, whereas the winter case showed a more linear behavior. It seems plausible that these two different predictability regimes are consequences of the different nature of the geophysical turbulence operating in the various regions. The mean state vertical shear resulting from the local jet stream is perhaps a factor of 5 larger in the winter case.

A final major novel result concerns variations in predictability with differing initial conditions. This problem is of major practical interest to forecasters, and many attempts have been made to relate skill of predictions to ensemble spread variations. Our measure of predictability—relative entropy—is derived from information theory and may be interpreted as the utility of a statistical prediction given a perfect model. We find that variations in this measure are primarily not due to variations in prediction ensemble spread, but to variations in the anomalous means of these ensembles. We refer to the first effect as dispersion and the second as signal.

This result needs further detailed investigation because our method of constructing initial conditions ensures that differing initial conditions have the same initial condition spread. There are many options used in data assimilation to produce such initial conditions, and often they use the prediction model in a fundamental way to interpolate sparse observational data to the model grid. This ensures that the flow instability of the initial state will likely have some role in setting initial errors. Nevertheless, despite this caveat we find that after a prediction of 5–10 days, during which such instabilities may be expected to strongly influence ensemble spread, that the importance of signal relative to dispersion is actually somewhat increased.

A limitation of the present study is the neglect of convective processes. Apart from the computational overhead involved with the inclusion of these effects, there is often a serious problem of model verisimilitude. It is well known, for example, that models often have difficulty in simulating both the amplitude and period of low-frequency large-scale variations in convection, such as the Madden–Julian oscillation (MJO; see Lin et al. 2006). This may be due to the lack of horizontal resolution of the turbulence because cloud-resolving models often (and not surprisingly) appear to simulate aspects of convection better (see, e.g., Ziemaski et al. 2005). The precise manner in which the conclusions presented here are modified by the consideration of convection is likely to depend on details of the convective “turbulent cascade,” and so in the view of this author they will need to be studied in a number of models with very different convection parameteriza-

tions. One can easily imagine that if a model has too much variability at very high frequencies and high wavenumbers (convective “noise”), and too little associated with low-frequency and large-scale coherent structures, then estimates of predictability may be unduly pessimistic.

*Acknowledgments.* The author wishes to thank Tim DelSole, Adam Monahan, Brian Farrell, and Ilya Timofeyev for stimulating and useful discussions when this work was presented at several seminars and workshops. This work was supported by NSF Grants CMG 0417728 and ATM 0430889.

## APPENDIX A

### Initial Condition Distribution

Following the general philosophy outlined in the main body of the text, we assume that all prognostic variables are governed by a multivariate distribution that ensures a fixed horizontal decorrelation scale. Specifically, we assume that the set of prognostic variables  $x_i$  have the distribution

$$P(x_1, \dots, x_N) = D \exp \left[ - \sum_{i,j} C_{ij}^{-1} (x_i - \mu_i)(x_j - \mu_j) \right], \quad (\text{A1})$$

where the  $\mu_i$  are the means of the corresponding prognostic variables, which have a covariance matrix  $C_{ij}$ . We assume that the  $\mu_i$  are drawn randomly from a long integration of the atmospheric model. The covariances are assumed to have the form

$$C_{ij} = S_i S_j \exp[-(d^{ij}/a)^2] V_{ij}, \quad (\text{A2})$$

where  $d^{ij}$  is the horizontal distance between the various prognostic variables. Here,  $a$  is set at one-sixth the earth’s radius (i.e., around 1000 km);  $V_{ij} = 1$  if  $x_i$  and  $x_j$  are on the same vertical level and the same type of prognostic variable (i.e., temperature, vorticity, divergence, or surface pressure), and zero otherwise. This ensures no correlation between different vertical levels or different prognostic variables. This formulation is perhaps a little oversimplified because one would expect some correlation between different prognostic variables at a particular locality and also in the vertical.

The prognostic variables are assumed to have variance  $S_i^2$ , which is set at 1/100 of the climatological variance of the particular prognostic variable. These were determined by a long integration of the atmospheric model.

The idealized formulation given above can be justified qualitatively as follows: If one takes a prognostic

variable from the model and performs a local horizontal average to roughly represent the coarse-resolution value of this variable, one can use such an integrated variable to represent approximately an observation of the prognostic variable at coarse resolution. The discrepancy between this “observation” and the actual value of the variable at the center of the local averaging region then is an estimate of the observational error resulting from the lack of horizontal resolution. One can now obtain statistics for this estimated observation error using a long integration of the atmospheric model. The “errors” are approximately Gaussian with a horizontal decorrelation scale that varies directly with the degree of coarsening chosen to produce the observations.

In the initial condition experiments described in section 4c, it is worth noting that the error covariances will not vary from one mean initial condition to another because of our fixed choices outlined in Eq. (A2). A practical data assimilation methodology will not have this property because model dynamics will play a complex role in determining error covariances.

The initial conditions drawn from the distribution (A1) may be unbalanced quasi geostrophically, and thus possibly generate undesirable gravity waves. To circumvent this, a gravity wave time filter described in Lynch and Huang (1994) was applied to the prognostic variables for all ensemble members before a prediction integration.

## APPENDIX B

### Relaxation Temperature

The model temperature is relaxed using Newtonian cooling with a damping time of 15 days toward a zonally and vertically varying profile  $T_R(\phi, \sigma)$ , which is intended to represent the radiative convective profile for the global atmosphere in the northern winter. It is given as a function of latitude  $\phi$  and the vertical coordinate  $\sigma$  by

$$T_R(\phi, \sigma) = T_R(\sigma) + f(\sigma) \left[ \Delta T_{\text{NS}} \frac{\mu}{2} - \Delta T_{\text{EP}} \left( \mu^2 - \frac{1}{3} \right) \right], \quad (\text{B1})$$

where  $\mu = \sin\phi$ ;  $\Delta T_{\text{NS}} = -53$  K is the temperature difference between the north and south poles; and  $\Delta T_{\text{EP}} = 70$  K is the mean equator pole temperature difference. The first term  $T_R(\sigma)$  is the global mean vertical temperature profile, which is determined as follows: It is assumed that the relaxation temperature is a hyperbolic function of  $z$ , which is asymptotically isothermal as  $z \rightarrow \infty$  and asymptotically a constant gradi-

ent of  $6.5 \text{ K (km)}^{-1}$  as  $z \rightarrow -\infty$ . At  $z = 0$  it is assumed to have value  $288 \text{ K}$  and at the tropopause (assumed at  $12 \text{ km}$ ) it is assumed to be warmer than the constant gradient from the surface by an offset of  $2 \text{ K}$ . These four conditions are sufficient to uniquely define a regular conic section hyperbola. The profile was converted to sigma coordinates using the hydrostatic relation and by assuming that the surface profile temperature on nonzero orography is calculated using the  $z$  profile just discussed. For example, an orography of  $1 \text{ km}$  has an assumed surface profile value of a little more than  $281.5 \text{ K}$ .

The function  $f(\sigma)$  is a “surface intensification” factor that makes the atmosphere more stable as the surface cools from its equatorial value. It is given by

$$f(\sigma) = \sin \left[ \frac{\pi(\sigma - \sigma_T)}{2(1 - \sigma_T)} \right] \quad (\text{B2})$$

for  $\sigma > \sigma_T$  the sigma value of the tropopause and zero elsewhere.

REFERENCES

Bernardo, J. M., and A. F. M. Smith, 1994: *Bayesian Theory*. John Wiley and Sons, 586 pp.

Boffetta, G., and S. Musacchio, 2001: Predictability of the inverse energy cascade in 2D turbulence. *Phys. Fluids*, **13**, 1060–1062.

Buizza, R., and T. N. Palmer, 1998: Impact of ensemble size on ensemble prediction. *Mon. Wea. Rev.*, **126**, 2503–2518.

Charney, J. G., 1971: Geostrophic turbulence. *J. Atmos. Sci.*, **28**, 1087–1095.

Cover, T. M., and J. A. Thomas, 2006: *Elements of Information Theory*. 2nd ed. Wiley-Interscience, 748 pp.

DelSole, T., 2004: Predictability and information theory. Part I: Measures of predictability. *J. Atmos. Sci.*, **61**, 2425–2440.

Ehrendorfer, M., 1994a: The Liouville equation and its potential usefulness for the prediction of forecast skill. Part I: Theory. *Mon. Wea. Rev.*, **122**, 703–713.

—, 1994b: The Liouville equation and its potential usefulness for the prediction of forecast skill. Part II: Applications. *Mon. Wea. Rev.*, **122**, 714–728.

Frederiksen, J., and A. G. Davies, 1997: Eddy viscosity and stochastic backscatter parameterizations on the sphere for atmospheric circulation models. *J. Atmos. Sci.*, **54**, 2475–2492.

Gardiner, C. W., 2004: *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*. Springer Series in Synergetics, Vol. 13, Springer-Verlag, 415 pp.

Grötzner, A., M. Latif, A. Timmermann, and R. Voss, 1999: Interannual to decadal predictability in a coupled ocean–atmosphere general circulation model. *J. Climate*, **12**, 2607–2624.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.

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

—, 2007a: Information flow in ensemble weather predictions. *J. Atmos. Sci.*, **64**, 1005–1016.

—, 2007b: Statistical predictability in the atmosphere and other dynamical systems. *Physica D*, **230**, 65–71.

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

—, and A. J. Majda, 2005: Predictability in a model of geostrophic turbulence. *J. Atmos. Sci.*, **62**, 2864–2879.

Kraichnan, R., and D. Montgomery, 1980: Two-dimensional turbulence. *Rep. Prog. Phys.*, **43**, 548–619.

Leith, C. E., 1971: Atmospheric predictability and two-dimensional turbulence. *J. Atmos. Sci.*, **28**, 145–161.

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

—, and R. Kraichnan, 1972: Predictability of turbulent flows. *J. Atmos. Sci.*, **29**, 1041–1058.

Leslie, L. M., and K. Fraedrich, 1997: A new general circulation model: Formulation and preliminary results in a single and multiprocessor environment. *Climate Dyn.*, **13**, 35–43.

Lilly, D., 1983: Stratified turbulence and the mesoscale variability of the atmosphere. *J. Atmos. Sci.*, **40**, 749–761.

Lin, J.-L., and Coauthors, 2006: Tropical intraseasonal variability in 14 IPCC AR4 climate models. Part I: Convective signals. *J. Climate*, **19**, 2665–2690.

Lorenç, A. C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **112**, 1177–1194.

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

—, 1969: Predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289–387.

Lynch, P., and X.-Y. Huang, 1994: Diabatic initialization using recursive filters. *Tellus*, **46A**, 583–597.

Majda, A. J., R. Kleeman, and D. Cai, 2002: A framework of predictability through relative entropy. *Methods Appl. Anal.*, **9**, 425–444.

Nastrom, G., and K. Gage, 1985: A climatology of atmospheric wavenumber spectra observed by commercial aircraft. *J. Atmos. Sci.*, **42**, 950–960.

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

Rabier, F., H. Järvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.*, **126**, 1143.

Roulston, M. S., and L. A. Smith, 2002: Evaluating probabilistic forecasts using information theory. *Mon. Wea. Rev.*, **130**, 1653–1660.

Salmon, R., 1998: *Lectures on Geophysical Fluid Dynamics*. Oxford University Press, 378 pp.

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

Tang, Y., R. Kleeman, and A. Moore, 2005: On the reliability of ENSO dynamical predictions. *J. Atmos. Sci.*, **62**, 1770–1791.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330.

Tribbia, J., and D. Baumhefner, 2004: Scale interactions and atmospheric predictability: An updated perspective. *Mon. Wea. Rev.*, **132**, 703–713.

van den Dool, H., and Z. Toth, 1991: Why do forecasts for near normal often fail? *Wea. Forecasting*, **6**, 76–85.

Ziemiński, M. Z., W. W. Grabowski, and M. W. Moncrieff, 2005: Explicit convection over the western Pacific warm pool in the community atmospheric model. *J. Climate*, **18**, 1482–1502.