

## Predictability in a Model of Geophysical Turbulence

RICHARD KLEEMAN AND ANDREW MAJDA

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

(Manuscript received 9 March 2004, in final form 30 August 2004)

### ABSTRACT

The nature of predictability is examined in a numerical model relevant to the midlatitude atmosphere and oceans. The approach followed is novel and uses new theoretical tools from information theory, namely entropy functionals, as measures of information content and their application to finite ensembles. Particular attention is paid here to the practical application of these methods to the problem of ensemble prediction in dynamical systems with state spaces of high dimensionality. In this case, typically only an estimate of the prediction probability distribution function is available at coarse resolution. A methodology for estimating the information loss implied by this limited knowledge is introduced and applied to the practical problem of measuring prediction information content in a model able to generate geophysical turbulence. The application studied here generates such turbulence through the mechanism of baroclinic instability via an imposed and constant mean vertical shear. In traditional studies in this area, considerable attention has been paid to variations in ensemble spread as the major determinant of how predictability may change as prediction initial conditions vary. The analysis here reveals that such a scenario neglects the important role of the so-called ensemble signal, which is related to the difference in the first moments of the prediction and climatological distributions. It is found, in fact, that this quantity is often a strong control over variations in predictability of the large-scale barotropic flow. An initial investigation of the role of non-Gaussian effects shows that for the univariate large-scale barotropic case, they are only of minor importance to variations in predictability.

### 1. Introduction

The problem of weather prediction has a long and interesting history both from a theoretical and practical perspective. Lorenz (1963) was among the first to recognize the extreme sensitivity of such predictions to small variations in the specification of initial conditions. In a series of papers in the succeeding decades, Lorenz essentially initiated some of the present considerable interest in chaotic dynamical systems (e.g., Ruelle and Takens 1971; Grassberger 1983). Later a theoretical framework for statistical prediction involving probability distribution functions (pdfs) was proposed by a number of authors including, for example, Epstein (1969) and Leith (1974). This approach has been particularly useful from a pedagogical viewpoint. From a practical perspective, the problem of how to implement the program of statistical prediction received much at-

tention in the past two decades (e.g., Murphy 1988; Toth and Kalnay 1993, 1997; Palmer et al. 1993; Molteni et al. 1996; Houtekamer et al. 1996; Ehrendorfer and Tribbia 1997; Buizza and Palmer 1998). The basic difficulty here is that generally only a relatively small ensemble estimate of the prediction pdf is practically available in what is a high dimensional dynamical system and in a situation where higher order moments of the pdf may contribute significantly (see Kleeman 2002).

Recently there has been a renewed theoretical interest in predictability, particularly from an information theoretical perspective [Carnevale and Holloway 1982; Schneider and Griffies 1999; Kleeman 2002, henceforth referred to as K02; Roulston and Smith 2002; see also Zhu et al. (1996) and Toth et al. (2003) for the connection to numerical weather prediction]. Such an approach is particularly attractive as it allows us to define transparent measures of the information content of predictions that have attractive theoretical properties. In particular, K02 has recently advocated the use of the relative entropy of the prediction and climatological pdfs as a measure of the utility of a particular statistical prediction. Conceptually before a prediction is made,

---

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

knowledge of a particular dynamical system is given by the climatological or equilibrium pdf. Once the prediction pdf is available, the informational inefficiency of ignoring this and using the prior knowledge is given by the relative entropy of the two distributions. Not surprisingly such a measure has also received considerable attention in the statistics literature [Bernardo and Smith (1994), hereafter referred to as BS94, gives a good overview] where it forms a part of the foundations of Bayesian theory. Here it is also referred to as a utility measure. In generic, nonrigorous terms we may formulate it as

$$D[p(x), q(x)] \equiv \int p(x) \ln \left[ \frac{p(x)}{q(x)} \right] dx, \quad (1)$$

where  $p(x)$  is the prediction pdf while  $q(x)$  is the equilibrium pdf that can be considered to be periodic in time.

In general, it is often a reasonable approximation to consider the time evolution of random variables in geophysical applications to be approximately Markovian. Such a property will hold if the entire nature of the pdf at time step  $t + 1$  can be derived from that at time  $t$ . Clearly the numerical formulation of most geophysical problems in time stepping form often ensures that their numerical approximation satisfies such a property.<sup>1</sup> If the Markovian property holds, then the relative entropy satisfies three particularly attractive properties:

- 1)  $D[p(t), q(t)] \geq 0 \quad \forall t$  Positivity

- 2)  $D[p(t_1), q(t_1)] \geq D[p(t_2), q(t_2)] \quad t_2 > t_1$

Temporal Monotonicity

- 3)  $D[p(\alpha), q(\alpha)] = D[P(F(\alpha)), Q\{P[F(\alpha)]\}$  Invariance,

where  $F: \Psi \rightarrow \Phi$  is a general nonlinear transformation of state-space variables with nonzero Jacobian. Rigorous demonstrations of the first two properties can be found in Cover and Thomas (1991), while the latter is shown in BS94 (p. 158) and Majda et al. (2002, henceforth referred to as MKC). It is worth emphasizing that other entropic measures do not satisfy any of these properties in general.

An interesting aspect of the use of this measure is the connection to nonequilibrium statistical mechanics. The second property above can be interpreted as a generalized second law of thermodynamics for Markov processes. In molecular statistical dynamical systems where this law was first proposed over a century ago by

Boltzmann (1995), the equilibrium pdf is actually uniform on an energy hypersphere and in this case if we assume energy conservation, then the relative entropy reduces to minus the absolute (standard) entropy [see Eq. (1) above] and so the usual formulation of the second law is recovered. Evidently in many problems of practical geophysical interest, the equilibrium pdf is far from uniform and in this case the relative entropy emerges as a particularly natural measure. In terms of the analogy with statistical mechanics, property 2 above shows that it is the degree of disequilibrium of the system at prediction time that measures the usefulness of the prediction. Further discussion of the equilibration process in stochastic systems and its relation to relative entropy may be found in Gardiner (2004).

It is important to emphasize that the monotonicity property above only holds rigorously when the entire state space of the dynamical system is considered. If a subspace is considered, then information may flow from the complement of the subspace into the subspace and this may cause the utility of the subspace to increase. Such an effect was noted in K02 in the case of ENSO, where information flowed from the subsurface to SST and thus increases in the utility of the latter quantity were sometimes observed. In general, if one chooses a subspace that explains much of the variation of state-space variables (as we shall do later in the text), this does not occur very often and, in our results, only occurred to a very minor extent. In this scenario, the neglected modes can be considered approximately to be a stochastic bath for the retained modes and then the reduced space behaves in a close to Markovian fashion.

At this point it is worth summarizing some of the advantages in using relative entropy to study predictability:

- 1) Relative entropy is a universal, intuitively transparent, and invariant measure of prediction utility.
- 2) Simpler diagnostics such as ensemble variance do not cover the range of effects covered by relative entropy. For example, that particular diagnostic says nothing about the signal to noise effect often important to practical prediction (see our conclusions on this matter later in the paper). It also says nothing about multimodality or kurtosis and other distribution shape issues. Of course, a range of diagnostics could be assembled to cover all these effects but a single functional covering many of them and on an equal footing has obvious advantages. The obvious interpretation of this functional in terms of information content and flow should also be of significant interest.

- 3) The perspective of prediction as an equilibration

<sup>1</sup> Of course if some aspect of the problem depends on values of the state-space variables from several time steps back in time, this will not hold. Such situations are, however, not very common.

process is, we believe, of fundamental physical importance. Entropy is the natural metric for the study of this relaxation process.

In K02, a range of stochastic models were examined that have direct relevance to the problem of climate prediction (the stochastic forcing represents the atmospheric time scales here) and it was found that the first moment of the prediction pdf was often the major control on variations in the utility of predictions with initial conditions. In fact, for stochastic differential equations with constant coefficients, it may be shown almost trivially that for deterministic initial conditions, the only control on utility variation is the first moment. This situation contrasts strongly with that normally assumed in atmospheric prediction, where it is often assumed that the pdf spread (the second moment) exercises such a control. A natural question then to ask is whether such a common assumption is justified within the general formulation of predictability that we have proposed. A first examination of this question was undertaken by Kleeman et al. (2002), using a particularly idealized model of the atmosphere and ocean known as the truncated Burgers model (see Majda and Timofeyev 2000). In this case, the pdfs are often close to Gaussian, which allows us to approximately calculate utility through exact formulae. It was found that the first moment again plays a very important role in controlling utility variations. Motivated by these results and the desire to introduce general techniques to deal with non-Gaussian pdfs and finite ensembles, we reexamine these ideas here in a model of geostrophic turbulence that has many of the physical features of the midlatitude atmosphere and ocean.

As alluded to above, a major problem in statistical prediction concerns calculation of the prediction pdf and its evolution. In practical situations, this function is estimated by the Monte Carlo technique of ensemble prediction. Here, one draws initial conditions according to some estimate of the initial condition pdf and calculates many trajectories. As the dimension of the state space increases, however, the estimation of the pdf becomes more and more difficult, a situation sometimes referred to as “the curse of dimensionality”.<sup>2</sup> The perspective we shall adopt here in response to this problem is derived from information theory. An ensemble estimate evidently implies a reduction in the amount of

information known about the prediction and, as we shall see in the next section, it is possible to quantify this loss of information rather precisely. Thus our philosophy is that ensemble prediction implies an information loss over the ideal of pdf prediction (which has been studied previously). The nature of this information loss is rather interesting and evidently of significant practical interest. We only begin our exploration of its consequences here.

In previous contributions, we have considered simple models relevant to both climate and atmospheric prediction. In this paper, we take the first step toward the consideration of realistic atmospheric prediction models. As our focus at this stage is still primarily methodological and didactic, we chose to investigate a model, which while still not approaching the complexity of numerical weather prediction or an ocean general circulation model, still has the dominant mechanism of midlatitude baroclinic turbulence. The quasigeostrophic two-layer model with uniform vertical shear meets this requirement and has also been extensively analyzed in the literature (see Salmon 1998, chapter 6, and references therein).

The remainder of this contribution is structured as follows. In section 2 we develop the tools for calculating the information content of an ensemble prediction. In section 3 the dynamical model to be studied is introduced, justified, and explored. In section 4, the methodology of section 2 is applied to the dynamical model and the question of what controls variations in utility is addressed. Section 5 contains a discussion and summary of the results.

## 2. Information loss due to ensemble estimation

As mentioned previously it is usually impossible to calculate the full evolution of the prediction pdf since the state spaces of dynamical systems of practical interest are often of very high dimensionality. In addition, the structure of the pdf can sometimes become highly non-Gaussian and again, as a consequence, difficult to estimate. As an example, K02 showed that the standard Lorenz 3 mode model has this property. Usually only a Monte Carlo estimation known as an ensemble is available, and often the size of this sample of the prediction pdf is considerably smaller than the state-space dimension. Various selective sampling techniques (singular and breeding vectors) are often deployed in an attempt to circumvent this problem (see, e.g., Palmer and Tibaldi 1988; Toth and Kalnay 1997). Here we adopt a different approach. An ensemble estimate implies that the full information of the prediction pdf is fundamentally unavailable to us. Indeed it may be shown rather

---

<sup>2</sup> It is worth pointing out that if one is only interested in one or two variables (as is often the case) rather than the full state space, as assumed in this discussion, then this issue may not arise since ensembles can generally be large enough to resolve adequately the corresponding pdf in the one or two dimensions.

easily (see below) that for realistic situations there is considerable loss of information implied for any ensemble that is within practical reach.<sup>3</sup> We present here a method for calculating such an information loss that relies on a coarse graining of a relevant subspace of state space. As we shall see, there are two sources of information reduction. The first is due to the coarse graining itself, which discards the finescale information, while the second is due to sampling error with respect to the chosen coarse graining.

To be more concrete, suppose one has a series of bins in state space for observing random variables of interest and determining their pdf and hence their information content. Clearly such an observing framework implies that we are throwing away information at scales that are smaller than our bins, hence, the first type of information loss. In addition when we use an ensemble we use the number of ensemble members falling into a particular bin as an estimate of the local bin probability. Of course, since we have an ensemble this (the bin probability) is only a sample quantity and will likely change (hopefully slightly!) if we rerun the ensemble. We have therefore an imprecise knowledge of our bin probabilities. This is the second type of information loss.

*Coarse-grained ensemble estimation*

In geophysical applications it is often possible to explain much of the variability of many dynamically relevant variables in a very high dimensional dynamical system with relatively few modes. These modes often tend to be large-scale spatially and of low frequency. This situation holds for the model we shall consider later. Providing that the number of these modes is not large, we can obtain useful estimates from ensembles of their information content albeit at fairly coarse resolution.

Let us suppose that this reduced space has dimension  $n$  and that we have a complete partitioning of this space into  $m$  bins or subsets  $X_i$  with  $i = 1, \dots, m$ . In general, one would expect that the number of bins  $m$  covering our space would greatly exceed the dimension of the space  $n$ . This is in order that there be adequate resolution of each dimension in our coarse graining. Given such a partitioning then an ensemble implies a frequency count  $f_i$  associated with every bin  $X_i$  (simply count the number of ensemble members passing through each bin). Providing that  $m \gg n$ , then many of

the  $f_i$  may be of significant size. As a concrete illustration, let us suppose we are interested in quartile information for each dimension of our space. Clearly in this case we require ensembles of size at least  $4^n$  for there to be many  $f_i$  of significant size (imagine a hypercube in  $n$  dimensions where each side has four divisions). For ensembles of size  $10^3$ – $10^4$  (the usual practical limit), this implies that approximately  $n < 7$  at least for quartile resolution. Higher resolution evidently requires larger ensembles.

In what follows, we make extensive use of standard techniques from Bayesian statistical analysis, which may be unfamiliar to some readers. We have found the book by Bernardo and Smith (1994) to be an excellent primer on these methodologies and thoroughly recommend chapters 1 and 3 particularly as background reading for this section.

Consider now the prediction pdf  $p$  on this reduced state space. If we integrate over each partition element  $X_i$  we obtain the coarse-grained discrete probability vector element  $p_i$ . Evidently we could estimate such a vector using the  $f_i$ . Consider now the conditional probability  $P(\mathbf{f}|\mathbf{p})$  that we observe  $f_i$  given that  $p_i$  holds. It follows from elementary probability theory that

$$P(\mathbf{f}|\mathbf{p}) = \frac{1}{Z(\mathbf{f})} \prod_{i=1}^m p_i^{f_i},$$

$$Z(\mathbf{f}) = \frac{\prod_{i=1}^m (f_i)!}{(N + m - 1)!},$$

$$N = \sum_{i=1}^m f_i.$$

Now Bayes theorem (see BS94) gives

$$P(\mathbf{p}|\mathbf{f}) \propto P(\mathbf{f}|\mathbf{p})P_{\text{pr}}(\mathbf{p}),$$

where  $P_{\text{pr}}(\mathbf{p})$  is the prior probability that a particular set of  $p_i$  occurs.<sup>4</sup> Without any evidence of what values  $p_i$  take, it is reasonable to take this prior probability as uniform; that is, in the absence of evidence there is no reason to expect that any one set of  $p_i$  is any more likely than any other. With this assumption we obtain

$$P(\mathbf{p}|\mathbf{f}) = \frac{C}{Z(\mathbf{f})} \prod_{i=1}^m p_i^{f_i} \equiv \Phi_{\mathbf{f}+1}(\mathbf{p}), \tag{2}$$

where  $\Phi$  is the so-called Dirichlet distribution. The first moment of this distribution then gives the expected  $p_i$ ,

<sup>3</sup> It is worth noting that there is also an information loss due to our uncertain knowledge of the initial condition (time zero) pdf. We do not consider this loss in this paper.

<sup>4</sup> In other words, it is prior to the ensemble observation of the frequencies  $f_i$ .



given the observed  $f_p$  and may be shown by direct analytical integration to be given by

$$\langle p_i \rangle = \frac{f_i + 1}{N + m}.$$

Now that we have calculated the distribution of the  $p_i$  we can calculate the expected information loss in assuming that  $p_i = \langle p_i \rangle$ . This is clearly

$$\text{EL}(\mathbf{f}) = \int \Phi_{f+1}(\mathbf{p}) D(\mathbf{p}, \langle \mathbf{p} \rangle) d\mathbf{p},$$

where  $D(\mathbf{p}, \langle \mathbf{p} \rangle)$  is the relative entropy of the coarse-grained pdfs  $\mathbf{p}$  and  $\langle \mathbf{p} \rangle$ . Using known analytical expression for moments of the Dirichlet distribution and the expected values of their logarithms it is possible to evaluate EL analytically with the result:

$$\text{EL}(\mathbf{f}) = \sum_{i=1}^m \langle p_i \rangle [\psi(f_i + 2) - \psi(n + m + 1) - \ln \langle p_i \rangle], \quad (3)$$

where  $\psi$  is the digamma function (Abramovitz and Stegun 1972, p. 258).

It may also be shown by the known form of the digamma function and elementary calculus that the expected information loss is minimized by choosing  $\langle p_i \rangle$  as our estimator of the coarse-grained pdf  $p_i$ . The relative entropy of the coarse-grained optimal prediction and climatological pdfs  $\langle p \rangle$  and  $\langle q \rangle$  may be easily evaluated:

$$D(\langle p \rangle, \langle q \rangle) = \sum_{i=1}^m \langle p_i \rangle \ln \left( \frac{\langle p_i \rangle}{\langle q_i \rangle} \right). \quad (4)$$

In general, in most practical contexts the loss of information due to sampling of the climatological distribution is considerably smaller than that due to the prediction ensemble sampling, since it is much larger. An intuitively appealing definition, then, of the utility of an ensemble prediction is therefore

$$\text{EU}(\langle p \rangle, \langle q \rangle) \equiv \max[D(\langle p \rangle, \langle q \rangle) - \text{EL}(\langle p \rangle), 0.0]; \quad (5)$$

that is, one reduces the coarse-grained information content by the information loss due to the prediction distribution sampling.

Another approach to this problem apart from Eq. (5) is to use the probability function defined in Eq. (2) to estimate the likely spread in the entropy. This has been pursued in Haven et al. (2004, manuscript submitted to *J. Comput. Phys.*).

### 3. Quasigeostrophic turbulence: Model and basic results

The midlatitude dynamical system that underlies both the atmosphere and ocean has been extensively studied in the past few decades [see Salmon (1998) for an excellent overview]. Central to these studies has been the so-called quasigeostrophic approximation of the primitive equations. This holds, crudely speaking, if the Rossby number is significantly less than unity and the Coriolis parameter does not vary greatly. Physically the approximation has the effect of filtering gravity waves and confining attention to low-frequency variability that is close to geostrophic balance. Many of the broad features of midlatitude variability that result from baroclinic and barotropic instability are well captured by models incorporating such an approximation.

Our aim in this contribution is to study the nature of predictability of as simple a system as possible that still retains the dominant physical instability mechanisms of the midlatitudes. The intention is to ensure that the main processes underlying turbulence generation in this region are retained in as simple a form as possible. This approach is motivated philosophically by the expectation that the basic predictability properties of the midlatitude dynamical system should follow from the nature of the turbulence there.

The simplest model meeting the above criteria is a two-level quasigeostrophic configuration that is externally forced by a mean vertical shear simulating the effect of differential meridional radiative forcing. We selected a two-layer version of the particular model of Smith et al. (2002) since the properties and nature of its turbulent cascade have received extensive discussion in the literature [see Salmon (1998), chapter 6 and references therein, as well as the comprehensive discussion and citations in Smith et al. (2002)].

The governing equations are prognostic in the potential vorticity  $q$  of the flow and for a two-level model (Salmon 1998, p. 111) with surface Ekman damping; orography and a constant mean vertical shear  $U \equiv U_1 - U_2$  may be written as

$$\frac{\partial q_i}{\partial t} + \mathcal{J}[\psi_i + \bar{\psi}_i, q_i + \bar{q}_i + R(i)h_b] = -\kappa R(i) \nabla^2 \psi_i + F_{\text{hyp}}, \quad (6)$$

where  $\psi_i$  is the streamfunction;  $\kappa$  the Ekman damping coefficient;  $h_b$  the orographic height function; and

$$R(1) = 0,$$

$$R(2) = 1.$$

The following relation holds between the potential vorticity and the streamfunction:

TABLE 1. Parameter values for the numerical experiments.

Experiment	$F = f_o^2 L^2 / 4\pi^2 g' H$	Beta = $\beta L^2 / 2\pi U_o$
Control	4.0	2.0
Small Rossby radius	40.0	2.0

$$q_1 + \bar{q}_1 = \nabla^2 \psi_1 + \beta y + S[(\psi_2 - \psi_1) + USy],$$

$$q_2 + \bar{q}_2 = \nabla^2 \psi_2 + \beta y + S[(\psi_1 - \psi_2) + USy],$$

with  $S = 4f_o^2/H^2N_m^2$  ( $H$  is the troposphere height while  $N_m$  is the Brunt-Väisälä frequency at the model mid-point vertically). The mean streamfunction is given by

$$\bar{\psi}_i = -U_i y,$$

in order to assure that the background state is geostrophic. Finally the term  $F_{\text{hyp}}$  represents a hyperviscosity, which in spectral space (the method used to solve the model) acts as a damping predominantly on the largest wavenumbers of the model. This term is used to ensure numerical stability and simulates the sink of energy at the smallest scales of the model [see Smith et al. (2002) for further discussion on the precise formulation]. Nondimensional parameter choices used in the numerical experiments below are detailed in Table 1. The symbols have the following meanings:  $f_o$  is the Coriolis parameter about which the beta plane used is constructed;  $L$  is the domain size in both directions;  $g'$  is the mean reduced gravity for the two-layer configuration; and  $\beta$  is the meridional gradient of the Coriolis parameter while  $U_o$  is the horizontal velocity scale.

For our initial numerical experiments reported here we chose to use a doubly periodic domain. Obviously this configuration choice may affect our conclusions, and more general numerical results also involving orography and spherical geometry will be reported in a future contribution. Given the numerically demanding ensemble experiments reported in the next section we chose to use a reasonably coarse horizontal resolution and retained 15 wavenumbers in the zonal and meridional directions. For the control experiment, this is adequate since, as has been shown in the literature, the turbulent cascade exchanges energy between baroclinic and barotropic modes at around the spatial scale of the Rossby radius, which for  $F = 4$  is quite well resolved by 15 modes. For the smaller Rossby radius ( $F = 40$ ) resolution is marginal for this spectral truncation; however, we tested the sensitivity of equilibrium behavior and equilibration time scale to an order of magnitude increase of resolution in both directions and noted little qualitative change in model behavior.

When the model described by Eq. (6) with the pa-

rameters of Table 1 is integrated from arbitrary initial conditions, the equilibration process is controlled primarily by the turbulent cascade rather than by Ekman spindown. When scaling is chosen appropriate to atmospheric conditions, the time scale involved is on the order of weeks rather than days (the Ekman time scale). The process by which the equilibrium turbulent cascade is maintained was studied by Salmon (1978, 1980) and Held and Larichev (1996) and is displayed in Fig. 12 of Salmon (1998). Energy injected by the large-scale (constant) mean shear into the baroclinic component of the model cascades via the nonlinearities of the model to smaller scales until it reaches the baroclinic Rossby radius. At this scale, transfer to the barotropic component of the model is possible. Barotropic energy at the conversion horizontal-scale cascades primarily to the larger scales where it is removed by the Ekman dissipation. Energy also cascades in both the barotropic and baroclinic modes to scales smaller than the conversion scale where it is removed by the hyperviscosity term of the model. When equilibrium is achieved most energy occurs in the large-scale barotropic modes. This behavior is depicted in Fig. 1, which depicts the barotropic and baroclinic energy spectrum of the control equilibrium (a large time average is used) as well as a typical snapshot from both vertical levels.

An important consequence of the equilibrium state of the turbulence from the viewpoint of theoretical predictability studies is that relatively few large-scale barotropic modes are required to explain much variance within the model. This was confirmed by performing a linear regression at each point of the domain between local streamfunction and the first two nonconstant complex barotropic spectral modes for an extended time period during equilibrium. With respect to the two-dimensional Fourier decomposition, the complex modes used have wavenumber vectors (1, 0) and (0, 1). Given the complex nature of the modes obviously four degrees of freedom are involved. In the control case these large-scale barotropic modes accounted for around 60% of the surface and 45% of the upper level streamfunction variance at any point. In the case that  $F = 40$ , the explained variance due to these modes was higher at around 95% at both levels. This reflected the more strongly peaked barotropic spectrum for this latter parameter setting.

The approach we follow here of concentrating on the predictability of the large-scale barotropic modes of the flow will be extended in future studies to consider other important physical variables (e.g., temperature) that depend significantly on the baroclinic component of the flow. A related study by Abramov and Majda (2004) with a simpler model has considered such variables.

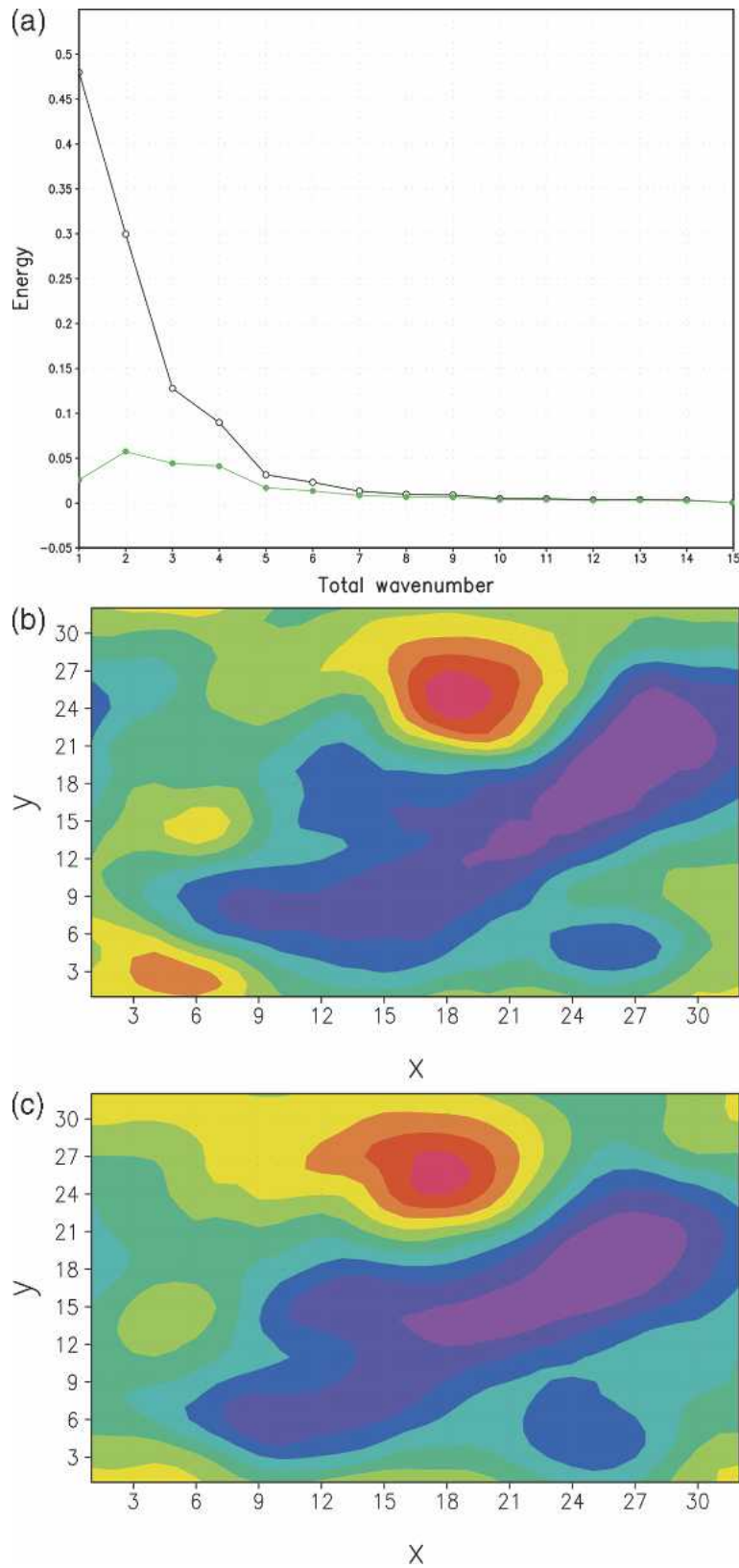


FIG. 1. Equilibrium behavior of the control run of the quasigeostrophic model. (a) Barotropic (black) and baroclinic (green) energy spectra. Snapshot of streamfunction at (b) upper and (c) lower levels.

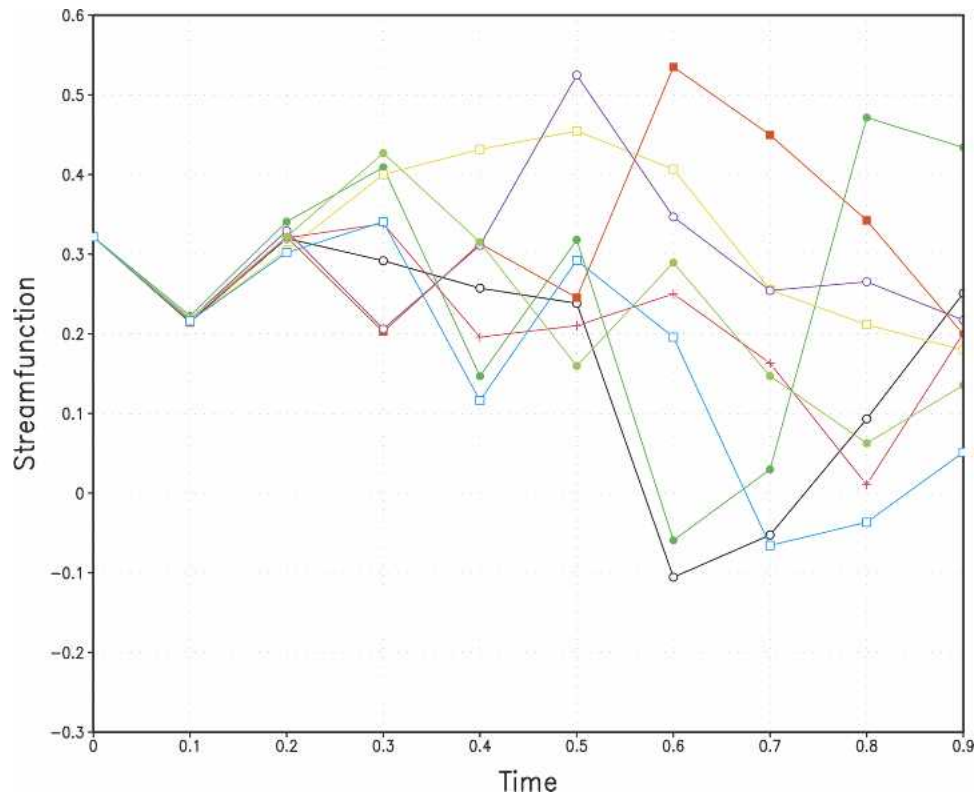


FIG. 2. Several ensemble members for a particular initial condition from the model control run ( $F = 4$ ; see text for further detail). Plotted is the real part of the  $(0, 1)$  spectral mode for various times.

It is important to reemphasize here the point made in section 2 that there is a fundamental limitation to the calculation of ensemble information content in a multivariate setting if one is restricted to ensemble sizes of order  $10^4$  or smaller. For such ensemble sizes there simply is not information available at fine scales for the multivariate case. Of course if only the univariate or bivariate cases are examined, considerably finer resolution may be used, as we shall see below.

#### 4. Predictability results

##### a. Experimental design

The motivation of the present study is to identify the nature of the variation in prediction utility with differing initial conditions. This variation is evidently of potentially great practical importance [see, e.g., Toth et al. (2003) for the numerical weather prediction context and earlier work]. To gain a representative view of such variations, we draw such initial conditions according to the climatological pdf. This is done by performing an extended integration of the model after it has achieved equilibrium and choosing the initial conditions at a sufficiently large equal time interval to ensure no correla-

tion of the state variables (which we take to be the streamfunction) from one set of initial conditions to the next.<sup>5</sup> At each initial condition set, an ensemble is generated by adding a small perturbation distributed according to a Gaussian with equal variance in all spectral components of the streamfunction. The standard deviation of this distribution was chosen to be 0.005 dimensionless units. For comparison, climatological standard deviation of the dominant spectral modes [the  $(1, 0)$  and  $(0, 1)$  components in spectral notation] is of order 1.5 units for the control experiment conducted here. Each ensemble member was integrated for 0.9 dimensionless time units that were roughly 60% of the way to equilibrium. A typical collection of ensemble member trajectories is depicted in Fig. 2, which plots the evolution of the real part of the  $(0, 1)$  spectral mode at the upper model level. To explore the predictability concepts introduced in section 2, a rather large 1000-member ensemble was produced for each of 50 initial condition sets. In practical applications, one is mostly currently restricted to smaller ensembles.

<sup>5</sup> This was identified by monitoring the total energy of the model and ensuring that it was quasi-steady in time.



As a first exploration of the predictability of this physically relevant system<sup>6</sup> we chose to explore the sensitivity of predictability to variations in the following parameter  $F$  of the flow that is related to the square of the inverse of the Rossby radius of deformation:

$$F = \frac{f_0^2 L^2}{4\pi^2 g' H}, \quad (7)$$

where  $H$  is the model vertical height while  $g'$  is the reduced gravity associated with the stratification that produces the mean vertical shear. We examined predictability at  $F = 4.0$  and  $40.0$ , which correspond very approximately to midlatitude atmospheric and oceanographic flow regimes, respectively. In the case of the  $F = 40.0$  setting, the climatological standard deviation of the dominant large-scale modes were a factor of 20 increased compared to the control case. We consequently increased the ensemble initial condition perturbations by the same factor.

### b. Coarse-grained entropy

As we have seen, a large amount of variance in this model may be explained by the first two nonconstant complex spectral barotropic modes. We chose therefore to partition a reduced four-dimensional subspace. With a 1000-member ensemble, we coarse grained each dimension into quartiles (with respect to the prediction ensemble), which implied that each partition box in the four-dimensional space had about 4 members for the prediction ensemble. For the climatological ensemble we took a large number ( $2 \times 10^4$ ) of basically uncorrelated snapshots from an extended equilibrium integration of the model. Given the much larger climatological ensemble size, partitions often had large number of climatological ensemble members within them (order 1000) but also often very few depending on how far from equilibrium the prediction ensemble was. The relaxation to equilibrium of all the initial conditions (with  $F = 4$ ) is depicted in Fig. 3, which shows the ensemble utility [see Eq. (5)] as a function of time. It is worth noting that for the present coarse graining the expected loss of information due to sampling [the term EL in Eq. (3)] is never larger than 0.1 and is relatively insensitive to initial conditions, prediction lead, and parameter settings. If we were to choose a larger number of bins, for example by considering quintiles rather than quartiles for each dimension, then the magnitude of this quantity

would at times approach the relative entropy estimate. This was one of the motivations for the particular choice of coarse graining adopted here.

Noteworthy is the significant fluctuation of utility from one initial condition to another. The variations in utility are strongly related to those derived under the assumption that all distributions are Gaussian (K02; Kleeman et al. 2002). As was discussed in this latter case, the utility  $R$  may be broken down into a term dependent on the first moments of the prediction pdf, which we call the signal and terms dependent on the second moments, which we call collectively the dispersion:

$$R = \frac{1}{2} \left[ \ln \left[ \frac{\det(\boldsymbol{\sigma}_q^2)}{\det(\boldsymbol{\sigma}_p^2)} \right] + \text{tr}[\boldsymbol{\sigma}_p^2(\boldsymbol{\sigma}_q^2)^{-1}] - n \right]$$

Dispersion

$$+ \frac{1}{2} (\boldsymbol{\mu}^p - \boldsymbol{\mu}^q)' (\boldsymbol{\sigma}_q^2)^{-1} (\boldsymbol{\mu}^p - \boldsymbol{\mu}^q) \quad \text{Signal},$$

where  $\boldsymbol{\sigma}^2$  and  $\boldsymbol{\mu}$  are the covariance matrices and mean vectors, respectively, and the sub/superscripts  $p$  and  $q$  refer to the prediction and climatological distributions, respectively. Variations in the coarse-grained utility here are generally strongly related to the signal term. This behavior is depicted in Fig. 4, which shows the relationships at dimensionless times of 0.2 and 0.7, which might be considered short and medium range predictions in weather nomenclature. It is worth noting also that in this case, the ensemble spread is generally not a good predictor of (coarse grained) utility. This kind of behavior has been noted in the past in the discussion of ensemble weather prediction skill (e.g., Van den Dool and Toth 1991; Toth et al. 2001) but has perhaps not received the prominence it deserves.

Interestingly there is a considerably uneven spread in the entropy at all leads with a few very high utility cases and many low cases, something that has also been reported in the weather prediction literature (e.g., Van den Dool and Toth 1991). A viewing of the high utility cases showed that they were consistently high for all prediction times.

The relaxation of pdfs to equilibrium is depicted in Fig. 5. In general, the climatological distribution is centered approximately on a sphere in the four-dimensional phase space with the distribution with respect to the radius of this sphere being approximately Gaussian. Points on such a sphere represent equal energy configurations in our reduced state space. Presumably this approximate energy conservation by the low wavenum-

<sup>6</sup> A future publication will explore the role of boundary condition, realistic orography, and spherical geometry.

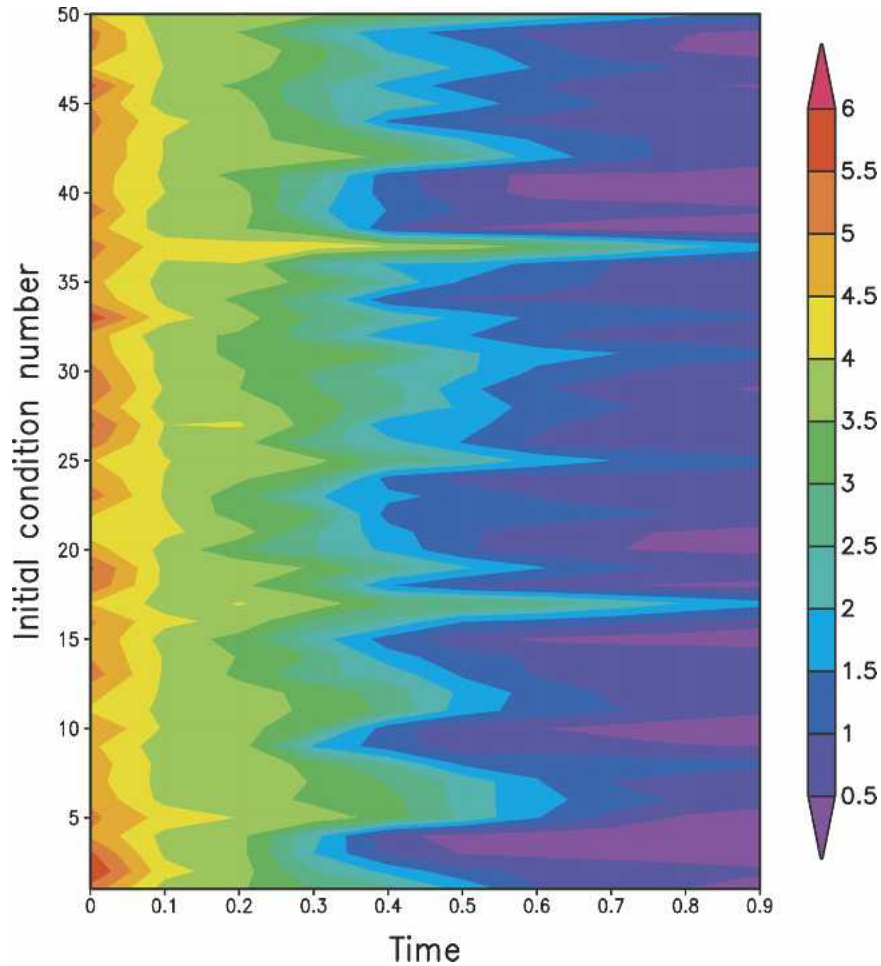


FIG. 3. Variation of the coarse-grained relative entropy as a function of time and initial condition. The entropy was calculated by the geometric partitioning of the reduced state space (see text). These results are for the control run with parameters as specified in Table 1, in particular  $F = 4$ .

ber barotropic modes represents a balance between energy injection from the higher wavenumber (barotropic) modes and dissipation at large scales by the Ekman friction. Prediction distributions at very short range are approximately Gaussian patches located at arbitrary points on or near this sphere. As time increases (Figs. 5c–e), this patch spreads, with some bias toward the meridional direction, around the sphere. Presumably this bias is caused by the beta effect. When viewed univariately (Figs. 5f), one notices some small non-Gaussianity apparently due to the spherical geometry that is guiding the relaxation.<sup>7</sup> It is interesting that

<sup>7</sup> In other words, this is if one were to consider the distribution of any one Fourier component (complex or real part).

somewhat similar behavior to this has been reported in the weather prediction context by Toth (1991). We examine how important non-Gaussian behavior is to utility/entropy below.

### c. Sensitivity to Rossby radius

The parameter  $F$  [see Eq. (7) above] controls the square of the ratio of the domain size to the model Rossby radius so larger values might be viewed (rather simplistically) as moving the model into a more oceanic regime. One might expect a priori predictability properties to be sensitive to this parameter in view of results reported elsewhere (K02). We therefore increased  $F$  from 4 to 40 and repeated our experiments from the previous subsection. The coarse-grained entropy of these modes was still dominated by the Gaussian signal

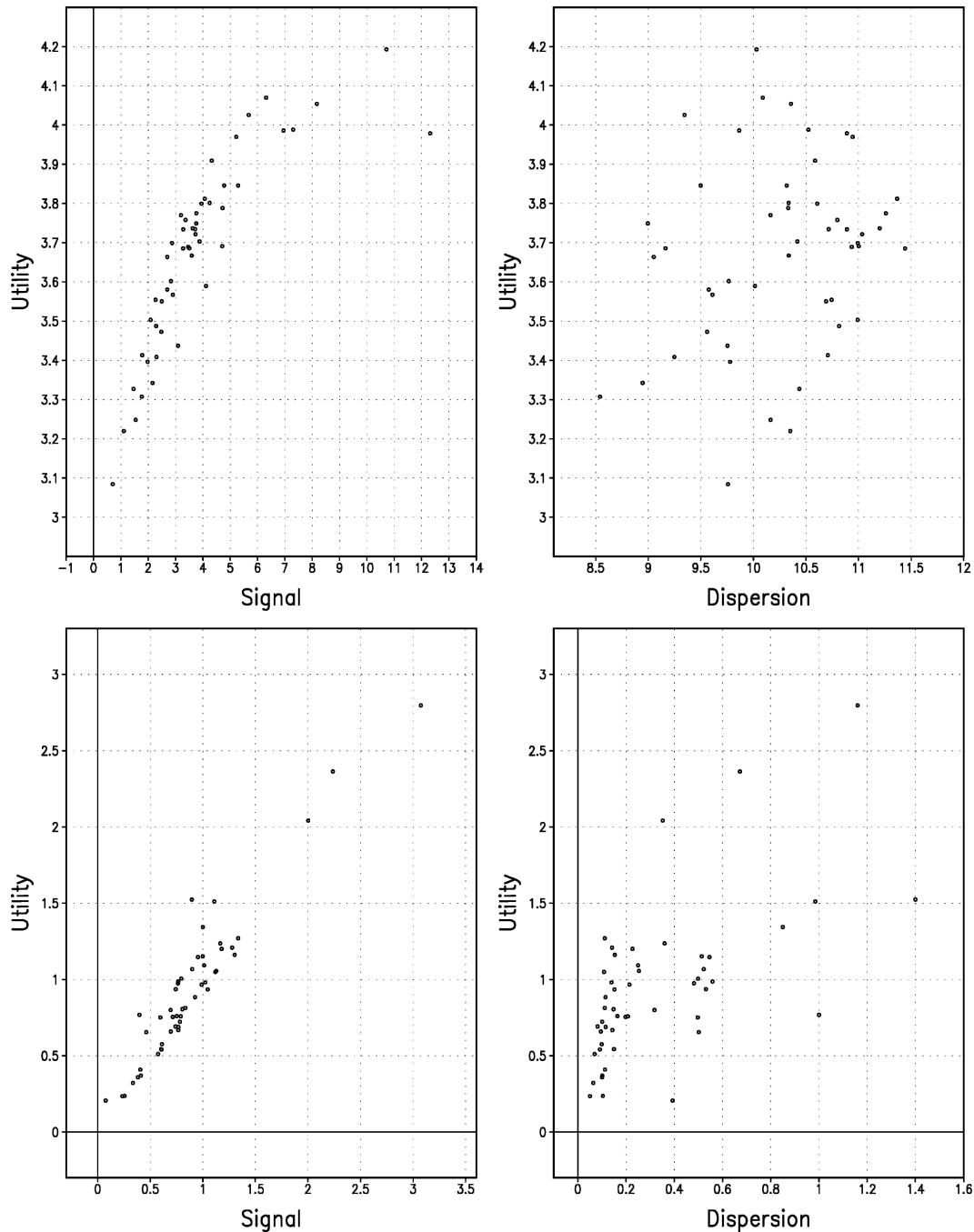


FIG. 4. Relationship of the (left) signal and (right) dispersion components of the Gaussian relative entropy with the coarse-grained relative entropy vs utility for prediction times of (top)  $t = 0.2$  and (bottom)  $t = 0.7$  for the control run ( $F = 4$ ) detailed in Table 1. Note that Gaussian relative entropy is also calculated in the four-dimensional reduced state space used for the coarse-grained entropy (and discussed in the text). The Gaussian functional is generally significantly higher than the coarse-grained one as the latter tends to miss considerable information due to the geometric partitioning, which is of course not assumed in the former case.

(see Fig. 6) at all time lags although perhaps not to quite the extent reported for  $F = 4$ .

Interestingly for longer prediction leads, the dispersion showed a triangular relationship to coarse-grained

utility with high dispersion being very often associated with high utility, whereas low dispersion was associated with both low and high utility situations. Such a relationship has been often reported in weather prediction

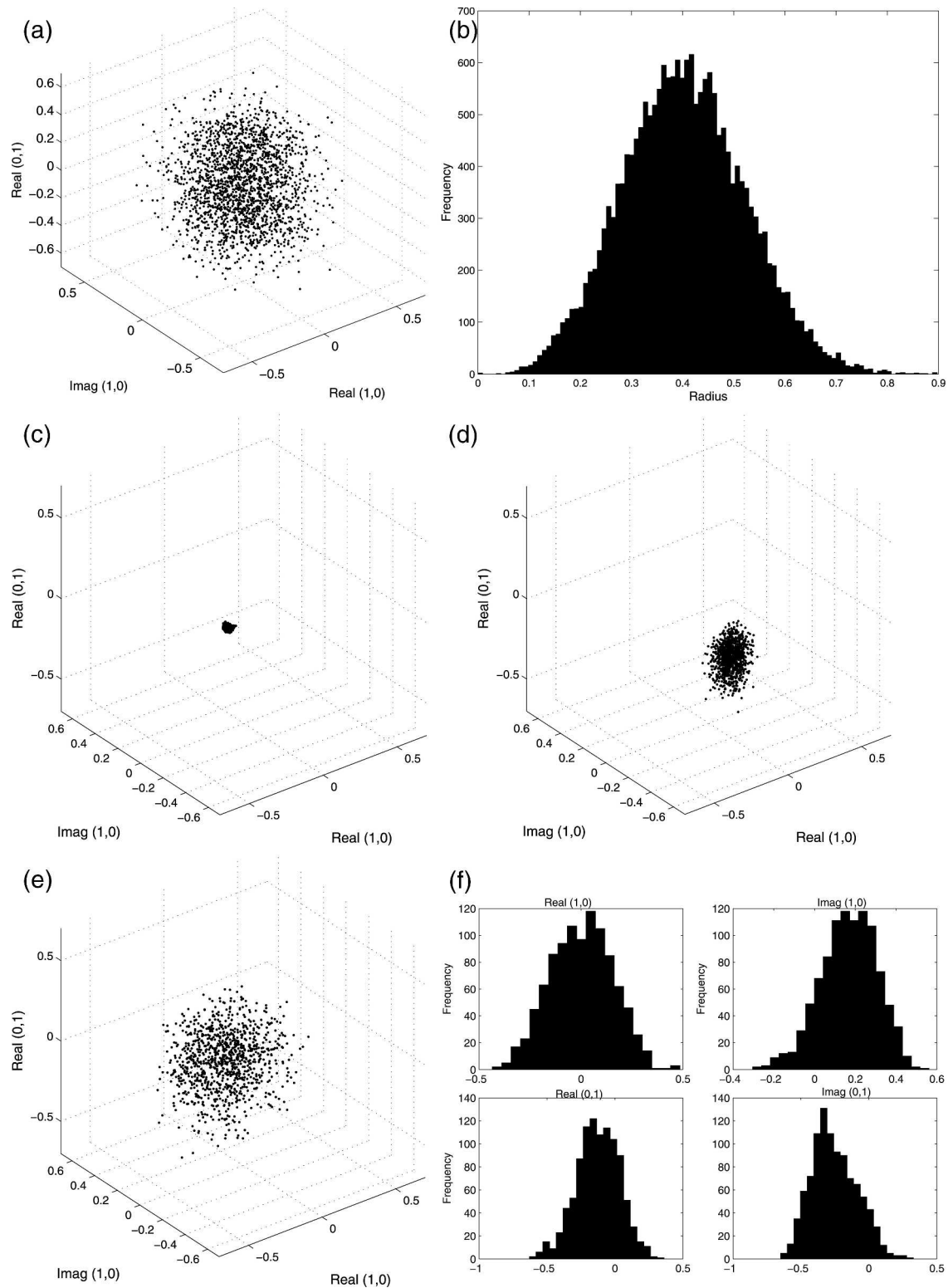


FIG. 5. Ensemble equilibration process for the control run ( $F = 4$ ; see Table 1). (a) 3D view of the 20 000-member equilibrium ensemble. The three dimensions used are a subset of the four-dimensional reduced state space. (b) Distribution for the radius vector in the reduced state space for the equilibrium ensemble. A particular prediction ensemble in the same 3D frame as (a) for times (c) 0.2, (d) 0.3, and (e) 0.5. Note the spreading about the sphere defined in (a). (f) Univariate distributions of the reduced state space are plotted for the case  $t = 0.7$ .

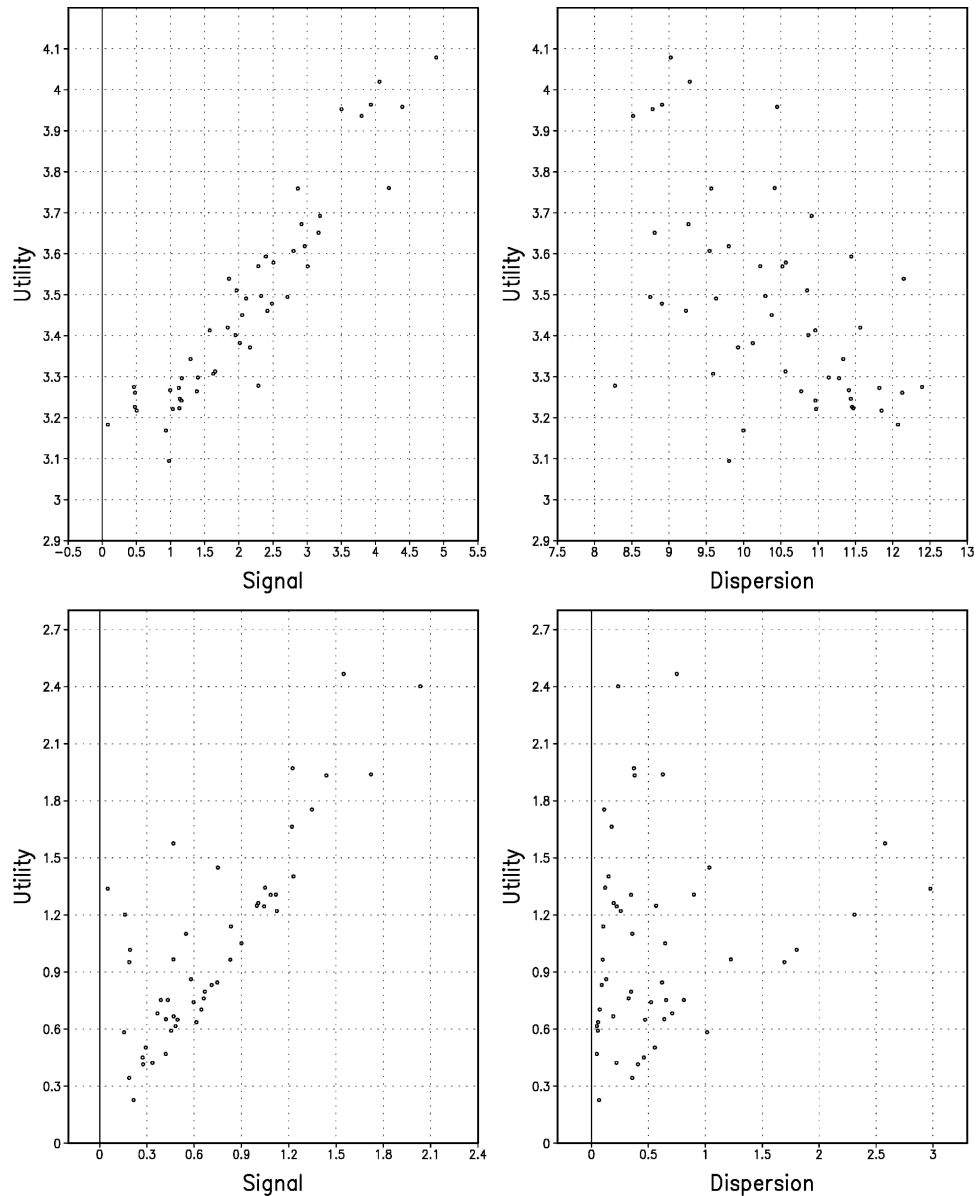


FIG. 6. Same as Fig. 4 but for the experiment with  $F = 40$ . See Table 1 for specific parameter settings.

studies of the relationship between skill and spread (the latter being closely related to the dispersion studied here).

#### d. Importance of non-Gaussianity to predictability

In section 4b, we noted some small non-Gaussianity in the prediction distribution due to the manner in which equilibration takes place. Given this, we decided to check the potential of such an effect to influence entropic measures. We chose to do this using univariate distribution since here there is generally enough parti-

tions for ensembles of size 1000 to be confident that non-Gaussian features of a distribution could be detected. Let us define the univariate relative entropy as the sum of all the relative entropies of the marginal distributions:

$$D_u(p, q) \equiv \sum_{i=1}^N D(p_i, q_i), \quad (8)$$

where  $p_i$  and  $q_i$  are the marginal distributions of the full pdfs  $p(x_1, x_2, x_3, \dots, x_N)$  and  $q(x_1, x_2, x_3, \dots, x_N)$  with respect to the state variable  $x_i$ .



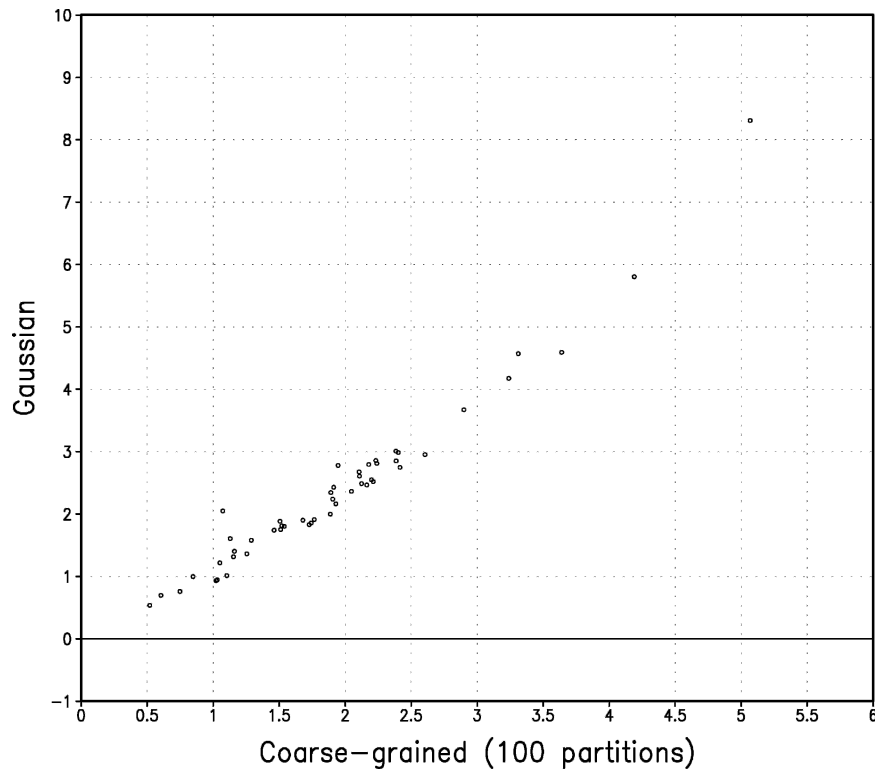


FIG. 7. Scatterplot of the univariate Gaussian relative entropy (see text) and that obtained by coarse graining each of the wavenumber-1 large-scale barotropic models at  $t = 0.7$ ; 100 partitions were used in the latter case, which implied a sample size of around 10 for each box and consequently very small sampling information loss.

Plotted in Fig. 7 is  $D_u$  for the first four Fourier modes discussed previously and for time 0.7 (results at other lags were similar). To perform this calculation, we divided the space for each mode into 100 partitions. Experience with synthetic ensembles drawn from Gaussian distributions shows that relative entropy is often close to converged with this number of partitions. What is shown in Fig. 7 is the relationship between the univariate entropy and the (univariate) entropy that would apply if the distributions were Gaussian. As can be seen the relationship between the two quantities is very strong suggesting that non-Gaussianity in this particular case is not very important to prediction utility variations. Naturally, one might expect this conclusion to be different if other variables such as temperature and precipitation were considered or if differing dynamical systems are examined (see, e.g., Abramov and Majda 2004).

## 5. Summary and discussion

A useful way of analyzing predictability in dynamical systems is through examination of the relaxation of pre-

diction (probability) distributions toward a quasi-stationary equilibrium distribution often called the climatological distribution. The degree of this disequilibrium may be measured rather precisely by the relative entropy of the two distributions. This functional corresponds with the informational inefficiency of assuming the climatological distribution when in fact the prediction distribution holds. From a Bayesian perspective, it thus represents the additional information brought to the table through the prediction process. Here one identifies the prior distribution with climatology and the posterior with the prediction distribution. Given this background the relative entropy measures rather transparently the utility of the prediction process.

In addition, the relative entropy satisfies a number of elegant mathematical properties including perhaps most importantly invariance under nondegenerate nonlinear transformations of state variables. This latter property would appear almost mandatory for a measure of predictability in geophysical systems where such state variable transformations are common. As a concrete illustration, the transformation from  $z$  vertical coordinates to sigma coordinates is a common (nonlinear)

transformation and one would not want results compromised by such a change.

In practical situations the full multivariate prediction and climatological distributions are unavailable since their time integration rapidly becomes infeasible as the dimension of state space increases. Instead one normally relies on Monte Carlo or ensemble methods to sample such distributions. This process must involve some reduction in information and hence the utility of the statistical predictions made. To analyze this loss one needs to adopt a particular coarse-graining frame of reference since probability density estimates at all points are obviously impossible. One obvious possibility for this is the geometric partitioning of state space.

Two forms of information loss are associated with such coarse-graining frames. First, the very act of coarse graining implies a discarding of information associated with the fine scales. Second, the remaining coarse-grained quantities are subject to sampling error, which again implies information loss. These two forms of loss are evidently connected since as one refines the coarse graining, one should expect the sampling error of the finer quantities to become larger. Such a trade-off in information loss could in fact be used to define an optimal coarse graining, a subject we will pursue further in a future publication. We analyzed in detail here the geometric coarse-graining strategy and derived expressions for the sampling information loss.

The mathematical machinery developed was then applied to one of the simplest models of midlatitude large-scale turbulence, namely a two-level quasigeostrophic model with constant vertical shear on a beta plane. Such a configuration simulates reasonably well the generation of turbulence through baroclinic instability and has a cascade from these baroclinic perturbations to dominant barotropic large-scale modes, which roughly approximates that thought to occur in the real atmosphere and ocean. The equilibrium configuration for this cascade shows typically that most energy is concentrated in the large-scale barotropic modes. This effect tends to be significantly greater in flows with a small rather than a large Rossby radius. Nevertheless, the first four Fourier modes typically explained a considerable fraction of the pointwise variation of the streamfunction at both vertical levels. Motivated by this, we simplified the predictability analysis by confining our attention in this study to this highly reduced state space. We plan to extend the dimension of the reduced space in further studies and also consider quantities such as the midlevel temperature field, which depends on the (neglected) baroclinic part of the flow.

A question of some practical importance in weather and ocean prediction concerns how predictability varies

from one forecast to another and what is the dominant control on such variations. Often one views skill spread diagrams under the assumption that variations in ensemble spread are the dominant control over predictability. The machinery we have developed here allows us to address these issues from a somewhat more fundamental viewpoint.

When a representative sample of initial conditions are chosen and the utility is calculated using the geometric partitioning strategy, there are often quite large variations at most prediction times (Fig. 3) and these variations are often not particularly well related to ensemble spread changes. On the other hand, they can be strongly related to a quantity referred to in previous publications as the signal. This is derived from the Gaussian expression for relative entropy and involves the difference in first moments of the prediction and climatology scaled by the climatological covariance matrix.

The importance of higher order moments—that is, non-Gaussianity—in causing variations in utility was also examined. This was done in a univariate context where we could have more confidence that such features could be adequately resolved by the 1000-member prediction ensemble. For large-scale barotropic modes we found no evidence that such moments were important to utility variation. In the future, we intend to revisit this issue in more detail and for other physical variables when more physically realistic atmospheric models are examined.

Another point worth discussing is the assumptions underlying this study. In particular, we are implicitly assuming both a perfect model as well as an accurate initial condition distribution of errors. Obviously such assumptions hold to varying degrees when practical prediction is attempted and model error becomes an important consideration. The perspective of this work is that the relations found here (and in future more complex models) are only relevant to the extent that such assumptions are close to being met. In other words, one would require that a good model [and initial condition (IC) distribution] be used. One would expect that in this situation the relationships reported here would still hold (and be useful) but be somewhat degraded by the presence of model error. In a practical sense it is often clear when a good model is used. Good prediction skill is but one indicator of this.

Finally it is worth emphasizing that the study reported here needs to be extended to models approaching the complexity of both modern numerical weather prediction and many level ocean general circulation models. We are presently in the process of completing

such a program using the tools that were explored in this paper.

*Acknowledgments.* The authors would like to acknowledge support from the National Science Foundation Grant CMG-0222133.

## REFERENCES

- Abramov, R., and A. J. Majda, 2004: Quantifying uncertainty for non-Gaussian ensembles in complex systems. *SIAM J. Sci. Stat. Comput.*, **26**, 411–447.
- Abramovitz, M., and I. A. Stegun, Eds., 1972: *Handbook of Mathematical Functions*. 9th ed. Dover, 1046 pp.
- Bernardo, J. M., and A. F. M. Smith, 1994: *Bayesian Theory*. John Wiley and Sons, 586 pp.
- Boltzmann, L., 1995: *Lectures on Gas Theory*. Dover, 490 pp.
- Buizza, R., and T. N. Palmer, 1998: Impact of ensemble size on ensemble prediction. *Mon. Wea. Rev.*, **126**, 2503–2518.
- Carnevale, G. F., and G. Holloway, 1982: Information decay and the predictability of turbulent flows. *J. Fluid Mech.*, **116**, 115–121.
- Cover, T. M., and J. A. Thomas, 1991: *Elements of Information Theory*. Wiley, 542 pp.
- Ehrendorfer, M., and J. J. Tribbia, 1997: Optimal prediction of forecast error covariances through singular vectors. *J. Atmos. Sci.*, **54**, 286–313.
- Epstein, E. S., 1969: The role of initial uncertainties in prediction. *J. Appl. Meteor.*, **8**, 190–198.
- Gardiner, C. W., 2004: *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*. 3d ed. Springer Series in Synergetics, Vol. 13, Springer, 415 pp.
- Grassberger, P., 1983: Generalized dimensions of strange attractors. *Phys. Lett.*, **97A**, 227–230.
- Held, I. M., and V. D. Larichev, 1996: A scaling theory for horizontally homogeneous, baroclinically unstable flow on a beta plane. *J. Atmos. Sci.*, **53**, 946–952.
- Houtekamer, P. L., L. Lefaivre, J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**, 1225–1242.
- Kleeman, R., 2002: Measuring dynamical prediction utility using relative entropy. *J. Atmos. Sci.*, **59**, 2057–2072.
- , A. J. Majda, and I. Timofeyev, 2002: Quantifying predictability in a model with statistical features of the atmosphere. *Proc. Natl. Acad. Sci. USA*, **99**, 15 291–15 296.
- Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, **102**, 409–418.
- Lorenz, E. N., 1963: Deterministic non-periodic flows. *J. Atmos. Sci.*, **20**, 130–141.
- Majda, A. J., and I. Timofeyev, 2000: Remarkable statistical behavior for truncated Burgers-Hopf dynamics. *Proc. Natl. Acad. Sci. USA*, **97**, 12 413–12 417.
- , R. Kleeman, and D. Cai, 2002: A framework of predictability through relative entropy. *Methods Appl. Anal.*, **9**, 425–444.
- Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliaigis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119.
- Murphy, J. M., 1988: The impact of ensemble prediction on predictability. *Quart. J. Roy. Meteor. Soc.*, **114**, 299–323.
- Palmer, T. N., and S. Tibaldi, 1988: On the prediction of forecast skill. *Mon. Wea. Rev.*, **116**, 2453–2480.
- , F. Molteni, R. Mureau, R. Buizza, P. Chapelet, and J. Tribbia, 1993: Ensemble prediction. *Proc. Validation Models Eur.*, **1**, 21–66.
- Roulston, M. S., and L. A. Smith, 2002: Evaluating probabilistic forecasts using information theory. *Mon. Wea. Rev.*, **130**, 1653–1660.
- Ruelle, D., and F. Takens, 1971: On the nature of turbulence. *Commun. Math. Phys.*, **20**, 167–192.
- Salmon, R., 1978: Two-layer quasi-geostrophic turbulence in a simple special case. *Geophys. Astrophys. Fluid Dyn.*, **10**, 25–52.
- , 1980: Baroclinic instability and geostrophic turbulence. *Geophys. Astrophys. Fluid Dyn.*, **15**, 167–211.
- , 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.
- Smith, S. K., G. Boccaletti, C. C. Henning, I. N. Marinov, C. Y. Tam, I. M. Held, and G. K. Vallis, 2002: Turbulent diffusion in the geostrophic inverse cascade. *J. Fluid Mech.*, **469**, 13–48.
- Toth, Z., 1991: Circulation patterns in phase space: A multinormal distribution? *Mon. Wea. Rev.*, **119**, 1501–1511.
- , and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330.
- , and —, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319.
- , Y. Zhu, and T. Marchok, 2001: The use of ensembles to identify forecasts with small and large uncertainty. *Wea. Forecasting*, **16**, 463–477.
- , O. Talagrand, G. Candille, and Y. Zhu, 2003: Probability and ensemble forecasts. *Environmental Forecast Verification: A Practitioner's Guide in Atmospheric Science*, I. T. Jolliffe and D. B. Stephenson, Eds., Wiley, 137–164.
- Van den Dool, H. M., and Z. Toth, 1991: Why do forecasts for “near normal” often fail? *Wea. Forecasting*, **6**, 76–85.
- Zhu, Y., G. Iyengar, Z. Toth, S. Tracton, and T. Marchok, 1996: Objective evaluation of the NCEP global ensemble forecasting system. Preprints, *15th Conf. on Weather Analysis and Forecasting*, Norfolk, VA, Amer. Meteor. Soc., J79–J82.