

Information Theory and Predictability for Low-Frequency Variability

RAFAIL ABRAMOV, ANDREW MAJDA, AND RICHARD KLEEMAN

Courant Institute of Mathematical Sciences, Center for Atmosphere Ocean Science, New York University, New York, New York

(Manuscript received 29 March 2004, in final form 27 July 2004)

ABSTRACT

A predictability framework, based on relative entropy, is applied here to low-frequency variability in a standard T21 barotropic model on the sphere with realistic orography. Two types of realistic climatology, corresponding to different heights in the troposphere, are used. The two dynamical regimes with different mixing properties, induced by the two types of climate, allow the testing of the predictability framework in a wide range of situations. The leading patterns of empirical orthogonal functions, projected onto physical space, mimic the large-scale teleconnections of observed flow, in particular the Arctic Oscillation, Pacific–North American pattern, and North Atlantic Oscillation. In the ensemble forecast experiments, relative entropy is utilized to measure the lack of information in three different situations: the lack of information in the climate relative to the forecast ensemble, the lack of information by using only the mean state and variance of the forecast ensemble, and information flow—the time propagation of the lack of information in the direct product of marginal probability densities relative to joint probability density in a forecast ensemble. A recently developed signal–dispersion–cross-term decomposition is utilized for climate-relative entropy to determine different physical sources of forecast information. It is established that though dispersion controls both the mean state and variability of relative entropy, the sum of signal and cross-term governs physical correlations between a forecast ensemble and EOF patterns. Information flow is found to be responsible for correlated switches in the EOF patterns within a forecast ensemble.

1. Introduction

The successful prediction of low-frequency variability over the Northern Hemisphere on monthly to seasonal time scales depends on the ability to diagnose the temporal evolution of large-scale persistent teleconnection patterns, including the Arctic Oscillation (AO), North Atlantic Oscillation (NAO), and Pacific–North American pattern (PNA). It has been demonstrated that many of the main features of low-frequency variability can be captured in barotropic models despite their simplicity (Simmons et al. 1983; Selten 1995; Branstator and Haupt 1998). In this study, the low-frequency variability of the atmosphere is modeled through a standard T21 barotropic truncation on the sphere with realistic orography and two different types of climatological forcing. These two types of climatological forcing in the barotropic model are designed to emulate variability of the real atmosphere at two different heights, corresponding to the pressure of 500 and 300 hPa, respectively, thus providing dynamical regimes with different low-frequency chaotic dynamics and mixing. These two regimes have caricatures of the basic

teleconnection patterns such as the AO, NAO, and PNA among the leading empirical orthogonal functions (EOF). The predictability of the T21 barotropic model for two different dynamical regimes is studied through the use of forecast ensembles, augmented by a recently developed information theory framework for estimating forecast utility (Kleeman 2002; Kleeman et al. 2002; Majda et al. 2002, hereafter MKC; Abramov and Majda 2004, hereafter A–M).

As has been recognized by Lorenz (1963), due to the chaotic nature of underlying physical processes, weather forecasts are extremely sensitive to perturbations in initial conditions, and therefore a probabilistic framework is required to describe the predictive utility of a forecast. Ehrendorfer (1994) has suggested the Liouville equation as a potentially useful tool in predicting weather forecast skill. However, due to the numerical complexities, the only feasible way of solving the Liouville equation for a large number of degrees of freedom remains the Monte Carlo simulation through a finite sample ensemble of solutions. Studies of predictions with forecast ensembles were performed, among others, by Anderson and Stern (1996), Toth and Kalnay (1993), and Palmer (2000). Practice shows that, with current computational facilities, in order to perform real-time forecasts with large atmospheric GCMs the size of the prediction ensemble has to be small, about 10–50 members, depending on spatial resolution. Due

Corresponding author address: Dr. Rafail Abramov, Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012.
E-mail: abramov@cims.nyu.edu

to limited size, one has to gather as much information from each ensemble member as possible; therefore, ensemble generation technique becomes an important matter. Ehrendorfer and Tribbia (1997) show that, for correct error growth reconstruction, the fastest growing directions of the phase space have to be sampled. Two efficient methods of ensemble generation are usually used in practical ensemble forecasts: local Lyapunov vectors (Toth and Kalnay 1993, 1997; Kalnay 2003) and singular vectors (Palmer et al. 1993; Reynolds and Palmer 1998). The efficiency of prediction depending on ensemble size has been studied previously by Buizza and Palmer (1998). However, the present work develops several different uses of information theory (Cover and Thomas 1991) for medium-range ensemble forecasting in idealized models, so no attempt is made to emulate the difficulties with limited ensemble size for high-resolution multilayer models. Large, 500-member forecast ensembles with direct statistical significance and simple ensemble generation techniques in EOF and physical spaces are employed during the course of the numerical experiments in this paper.

The applicability of information theory for weather or climate prediction has been studied previously by Carnevale and Holloway (1982), Schneider and Griffies (1999), Roulston and Smith (2002). Kleeman (2002) has suggested the relative entropy as an estimate of predictive utility in an ensemble forecast relative to the climatological record, as well as a signal-dispersion decomposition. The Gaussian framework of relative entropy and signal-dispersion decomposition has been tested by Kleeman et al. (2002) for a simple 100-mode truncated Burgers–Hopf model with chaotic behavior and well-understood spectrum and autocorrelation time scaling [for complete model description and climatology see Majda and Timofeyev (2000, 2002), and Abramov et al. (2003)]. MKC developed a more sophisticated framework of predictability through relative entropy for non-Gaussian probability density functions, which includes an hierarchy of rigorous lower bounds on relative entropy through the statistical moments beyond the mean and covariance through maximum entropy optimization (Mead and Papanicolaou 1984). Abramov and Majda (2004) converted the non-Gaussian predictability framework into a practical tool through the hierarchy of lower bounds and a rapid numerical optimization algorithm. Recently, Cai et al. (2004) exhaustively tested several facets of the non-Gaussian information theoretical predictability framework in a simple chaotic mapping model with an explicit attractor ranging from Gaussian to fractal as parameters are varied. Finally, Kleeman and Majda (2004, manuscript submitted to *J. Atmos. Sci.*) have quantified the loss of information in coarse-grained ensemble estimators and applied these ideas to geophysical turbulence. For the non-Gaussian predictability framework, A–M provided the complete decomposition of the relative entropy into three parts: signal (originating from

the mean state), dispersion (originating from covariance and higher moments), and cross-term (includes both the mean and higher moments). Additionally, different applications of relative entropy as a predictability tool were developed in A–M; besides a straightforward measure of lack of information in the climate relative to the prediction ensemble, the relative entropy can be used in estimating lack of information in a forecast ensemble relative to the actual events (Roulston and Smith 2002), in evaluating additional information content in the skewness and higher-order forecast moments (non-Gaussianity), and the information flow between different subsets of phase space in an ensemble forecast (statistical correlation between different large-scale phenomena). In A–M all of these facets were demonstrated for the Lorenz '96 model (Lorenz and Emanuel 1998), including highly non-Gaussian behavior.

Here, for low frequency variability of a barotropic model climate, it is demonstrated how the ideas of measuring the lack of information in the climate, non-Gaussianity, generalized signal, and information flow can be used in practice. The structure of this paper is organized as follows. In section 2 the T21 barotropic model and the two types of barotropic climate are described. In section 3 the climatology (mean state statistics) is presented, including representations of large-scale EOFs in physical space and their statistics. The information theory framework used in this paper is summarized in section 4. The design of numerical experiments for predictability in a T21 barotropic model and the various uses of information theory are given in section 5. Finally, a summary discussion of results is presented in section 6.

2. The barotropic model

The model equations are

$$\frac{\partial \zeta}{\partial t} + \nabla^\perp \psi \cdot \nabla q = \left(-\frac{1}{\tau} + K\Delta^3 \right) \zeta + \zeta^*,$$

$$q = \Delta \psi + 2\Omega \sin\theta(1 + h). \quad (1)$$

The model is nondimensionalized using the radius of Earth as the length scale and the angular velocity of Earth rotation as the inverse of time scale, such that the nondimensional time unit within the model is equivalent to one day; $\psi = \psi(\psi, \theta, t)$ is the streamfunction, $\zeta = \zeta(\psi, \theta, t) = \Delta \psi$ is the relative vorticity, $q = q(\psi, \theta, t)$ is the potential vorticity, $h = h(\varphi, \theta)$ is the realistic Earth topography scaled by the height of the troposphere, $\tau = 15$ days is the Ekman friction time scale, and $\Omega = 1$ is the unit angular velocity of Earth's rotation. The scale-selective damping coefficient $K = 5.38 \times 10^{-10}$ (which has units length⁶/time) is chosen so that the damping time scale for the wavenumber 21 is 3 h.

The model equations in (1) are projected onto the triangular-shaped array of spherical harmonics with

maximum wavenumber 21 (standard T21 barotropic truncation). The T21 truncation is further restricted to the Northern Hemisphere with no flow across the equator, therefore the truncated flow is completely described by 231 spectral coefficients. The fourth-order explicit Adams–Bashforth multistep method is used to integrate the model in time, with discrete time step of 45 minutes.

The function $\zeta^* = \zeta^*(\varphi, \theta)$ is the time-independent forcing, which is chosen to roughly match the properties of model climatology with observations. In this work, two different types of steady forcing are used in the barotropic model to mimic a realistic climatology at 500-hPa and 300-hPa heights in the troposphere.

a. Barotropic model climate for 500-hPa height

This type of climate is proposed by Selten (1995). Here, ψ_{CL} denotes the 10-day climatological mean field computed from 10 winters from European Centre for Medium-Range Weather Forecasts (ECMWF) daily analysis of 300-hPa vorticity fields rescaled to 500 hPa via multiplication by the factor of 0.6. The first guess for the forcing is determined through the formula proposed by Roads (1987)

$$\zeta_{500hPa}^{*[1]} = \nabla^\perp \psi_{CL} \cdot \nabla q_{CL} + \left(\frac{1}{\tau} - K\Delta^3 \right) \zeta_{CL} + \overline{\nabla^\perp \psi'_{CL} \cdot \nabla \zeta'_{CL}}, \quad (2)$$

where ψ'_{CL} , ζ'_{CL} are the deviations of the 10-day running mean from the climatological mean state, and the overbar denotes the time average. Then, model runs for 2000 days are performed, and each next iteration of forcing is obtained by replacing the previous transient eddy forcing with the one from the current run:

$$\zeta_{500hPa}^{*[n]} = \zeta_{500hPa}^{*[n-1]} + \overline{(\nabla^\perp \psi'^{[n]} \cdot \nabla \zeta'^{[n]} - \nabla^\perp \psi'^{[n-1]} \cdot \nabla \zeta'^{[n-1]})}. \quad (3)$$

The third iteration of (3) is implemented in the model.

b. Barotropic model climate for 300-hPa height

This type of climate is proposed by Franzke et al. (2005). Here the same observational data is used as in 500-hPa climate, but no additional rescaling is performed. The first guess for the forcing is determined through the same formula as in (2). Model runs for 3600 days are performed so that each next iteration of forcing is obtained by replacing the model climatological mean state and transient eddy forcing from the previous run with the observed climatological mean state and transient eddy forcing:

$$\zeta_{300hPa}^{*[n]} = \zeta_{300hPa}^{*[n-1]} + (\zeta_{CL} - \zeta'^{[n-1]}) + \overline{(\nabla^\perp \psi'_{CL} \cdot \nabla \zeta_{CL} - \nabla^\perp \psi'^{[n-1]} \cdot \nabla \zeta'^{[n-1]})}. \quad (4)$$

The forcing is implemented after 30 iterations.

3. Climatology

Climatological properties of the T21 barotropic truncation are obtained through time averaging from a time series of 200 000 days for a single solution assuming ergodicity. The zero velocity field is taken as the initial condition for the integration, and an initial time interval of 10 000 days is skipped before beginning the time averaging. Hereafter, the statistical steady states resulting from the steady forcing in sections 2a and 2b, are called the 500-hPa and 300-hPa climates, respectively, in a slight abuse of terminology. The reader is urged to remember, however, that properties of statistical steady state in the barotropic model are governed by the artificially created constant forcing, rather than the actual atmospheric height.

a. Mean state and variability

The climatological mean state and its variability for both the 500-hPa and 300-hPa types of climate are shown in Fig. 1. Indeed, the 500-hPa mean state looks like a version of the 300-hPa mean state scaled by 0.6. The biggest difference between two mean states is the subequatorial band of high pressure in the 500-hPa climate over Africa, the Indian Ocean, and the western part of Pacific Ocean, which is missing in the 300-hPa climate. The variances of both types of climate also resemble each other, though to a smaller extent than the mean states. The average variability of the 500-hPa climate is nearly 3 times as small as that of the 300-hPa climate, as might be expected from the change in height. However, the relative variability over the North Pole for the 500-hPa climate is higher than that for the 300-hPa climate (the contour lines are more dense near the North Pole). In addition, the variability for the 300-hPa climate increases rather uniformly from the equator to the North Pole, whereas the 500-hPa climate has a pronounced local maximum over Europe.

b. Empirical orthogonal functions

Empirical orthogonal functions (EOFs) are computed in the energy metric (for details see Selten 1995). The projections on the EOFs are denoted as principal components (PC) in formulas utilized below, with such normalization that each PC has unitary climatological variance. Variances of the EOFs in the energy metric for both types of climate are shown in Fig. 2. In this study of low frequency variability, further attention is restricted to the first four energy metric EOFs. For the 500-hPa climate, the first four EOFs contain about 45% of total variance, of which roughly a half is concentrated in the first EOF. For the 300-hPa climate there is no large gap in variance between the first and second EOFs, but nevertheless the first four EOFs have about 41% of total variance. In physical space, EOFs 1–4 for 500-hPa and 300-hPa types of climate are shown in

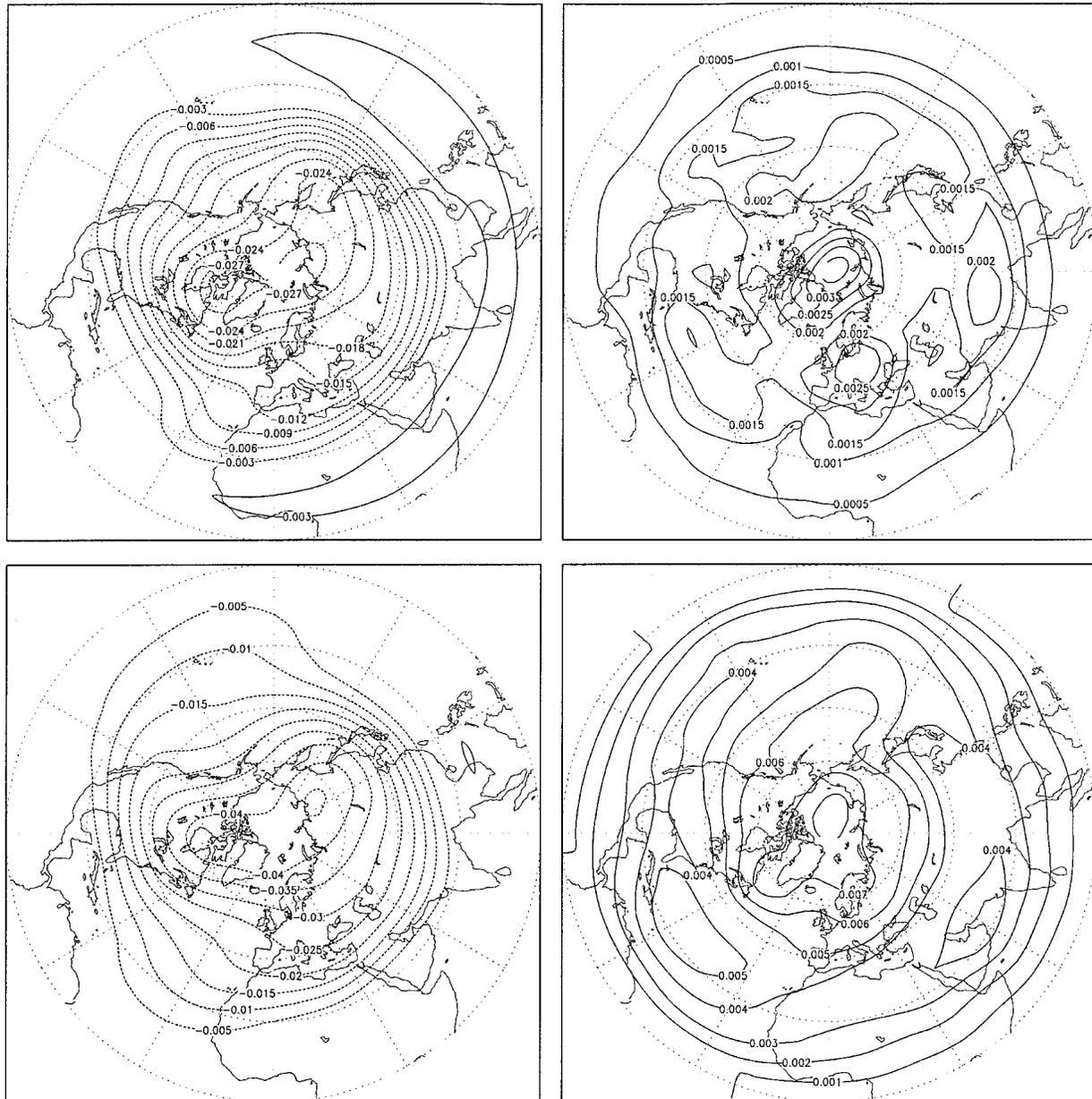


FIG. 1. The (left) mean state and (right) variability of the streamfunction for the (top) 500-hPa and (bottom) 300-hPa types of climate, displayed in model variables.

Figs. 3 and 4, respectively. The EOFs for the 500-hPa climate are characterized by smaller-scale structures than those for the 300-hPa climate. Especially it is noticeable for the second and fourth EOFs, where the wave train over southeastern Asia and the Pacific Ocean in the 500-hPa climate is replaced in the 300-hPa climate by one huge cyclone for the second EOF, and a cyclone–anticyclone pair for the fourth EOF. Because of these differences, the EOFs offer different zonal-blocking mechanisms depending on the type of barotropic model climate. Namely, while the first EOFs for

both types of climate strongly project on the Arctic Oscillation (AO), the second EOF for the 300-hPa climate projects on the Pacific–North American (PNA) pattern, whereas the third EOF for the 500-hPa climate resembles the North Atlantic Oscillation (NAO; Crommelin 2003). Of course, the realistic fidelity of the T21 barotropic model is far from that of high-resolution multilayer models; nevertheless these realistic features of the two barotropic model climates provide an interesting setting for low frequency variability for testing the predictability tools of information theory (A–M).

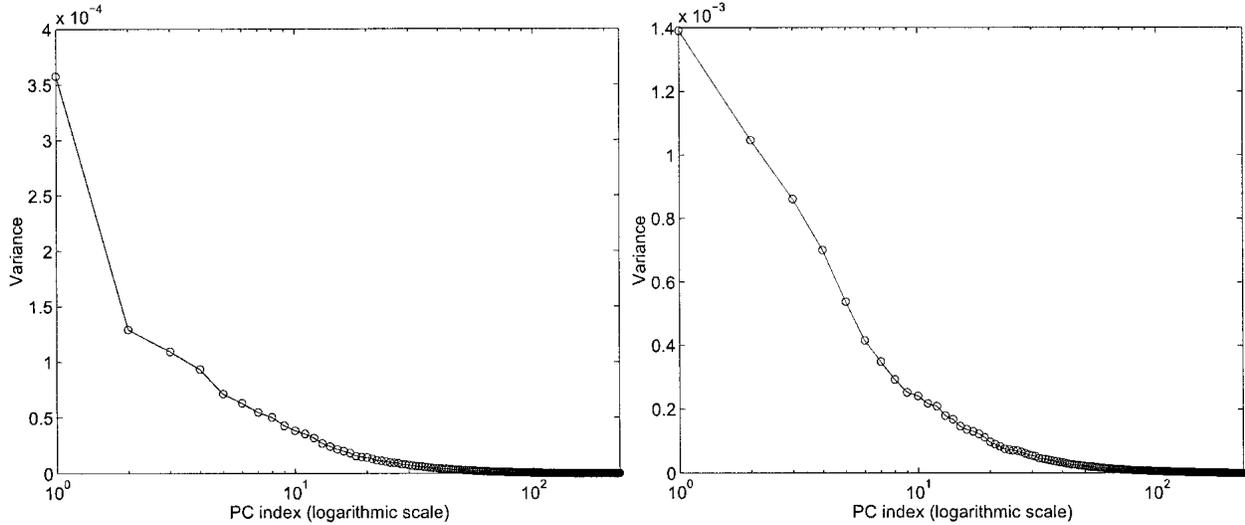


FIG. 2. Variances of energy metric EOFs for (left) 500- and (right) 300-hPa types of climate.

c. Time autocorrelations

The time autocorrelations of the PCs demonstrate the mixing rate and overall robustness of the dynamics and are given by the formula

$$C(\tau) = \overline{PC(t)PC(t + \tau)}, \quad (5)$$

where the overbar, as usual, denotes the time average. The time autocorrelations of the first four principal components for 500-hPa and 300-hPa types of climate are shown in Fig. 5. The rescaling from 300 to 500 hPa slows down the mixing, and is manifested by the significantly larger autocorrelation times for the PCs for the 500-hPa model climate. This is expected because the damping coefficients stay the same and the forcing is scaled down. However, this situation is natural—the higher velocity fields in the upper troposphere homogenize spontaneous structures much faster than in the lower troposphere, which corresponds to shorter mixing times. Here the 500-hPa climate has autocorrelation scale separation for PC1 relative to the rest (the correlation decay time is 50 days for PC1 and <25 days for the other three PCs for the 500-hPa climate). The more vigorous dynamics of the 300-hPa barotropic model climate do not exhibit autocorrelation scale separation (correlation decay times are all about 10 days for PC1–PC4).

d. Probability density functions

Probability density functions (PDFs) of the first four principal components for 500-hPa and 300-hPa types of climate are computed via standard bin counting from the time series of the 200 000-days-long climate trajectory and are shown in Fig. 6. The 300-hPa climate has smooth and nearly Gaussian PDFs for the first four PCs, whereas the PDFs for the 500-hPa climate are

skewed, and the PDF for PC1 exhibits signs of bimodality. The skewness and flatness for the first four PCs is shown in Table 1. The less robust dynamics of the 500-hPa climate results in more non-Gaussian shaping of the PDFs (the Gaussian skewness and flatness are 0 and 3, respectively). Table 1 shows that for the 300-hPa climate the skewness does not exceed 0.3 in magnitude, however, for the 500-hPa climate the skewness for PC1 (which is the least mixing) is nearly 0.4. Similar trends are found in the joint two-dimensional probability density functions of the first four principal components for the 500-hPa and 300-hPa types of climate, shown in Figs. 7 and 8, respectively. Contour plots for the 300-hPa climate are smooth and nearly round, which is evidence of high factorizability. However, the joint PDFs for the 500 hPa exhibit high peaks away from the origin. The impact of these different types of climate statistics on predictability is studied further in the paper.

4. The framework for relative entropy

Rigorous proofs of all mathematical statements given in this section, as well as more detailed discussion, can be found in MKC and A–M.

The lack of information in a prior probability density $\Pi(\mathbf{x})$ relative to a true probability density $p(\mathbf{x})$ is given,

TABLE 1. Skewness and flatness of PC1–PC4, 500- and 300-hPa types of climate.

| PC | 500-hPa climate | | PC | 300-hPa climate | |
|----|-----------------|----------|----|-------------------------|----------|
| | Skewness | Flatness | | Skewness | Flatness |
| 1 | 0.3905 | 2.086 | 1 | −0.1566 | 2.322 |
| 2 | −0.2831 | 2.467 | 2 | −0.1099 | 2.880 |
| 3 | 0.1758 | 3.144 | 3 | −0.2747 | 2.592 |
| 4 | 0.2314 | 2.760 | 4 | $−1.142 \times 10^{-2}$ | 2.736 |

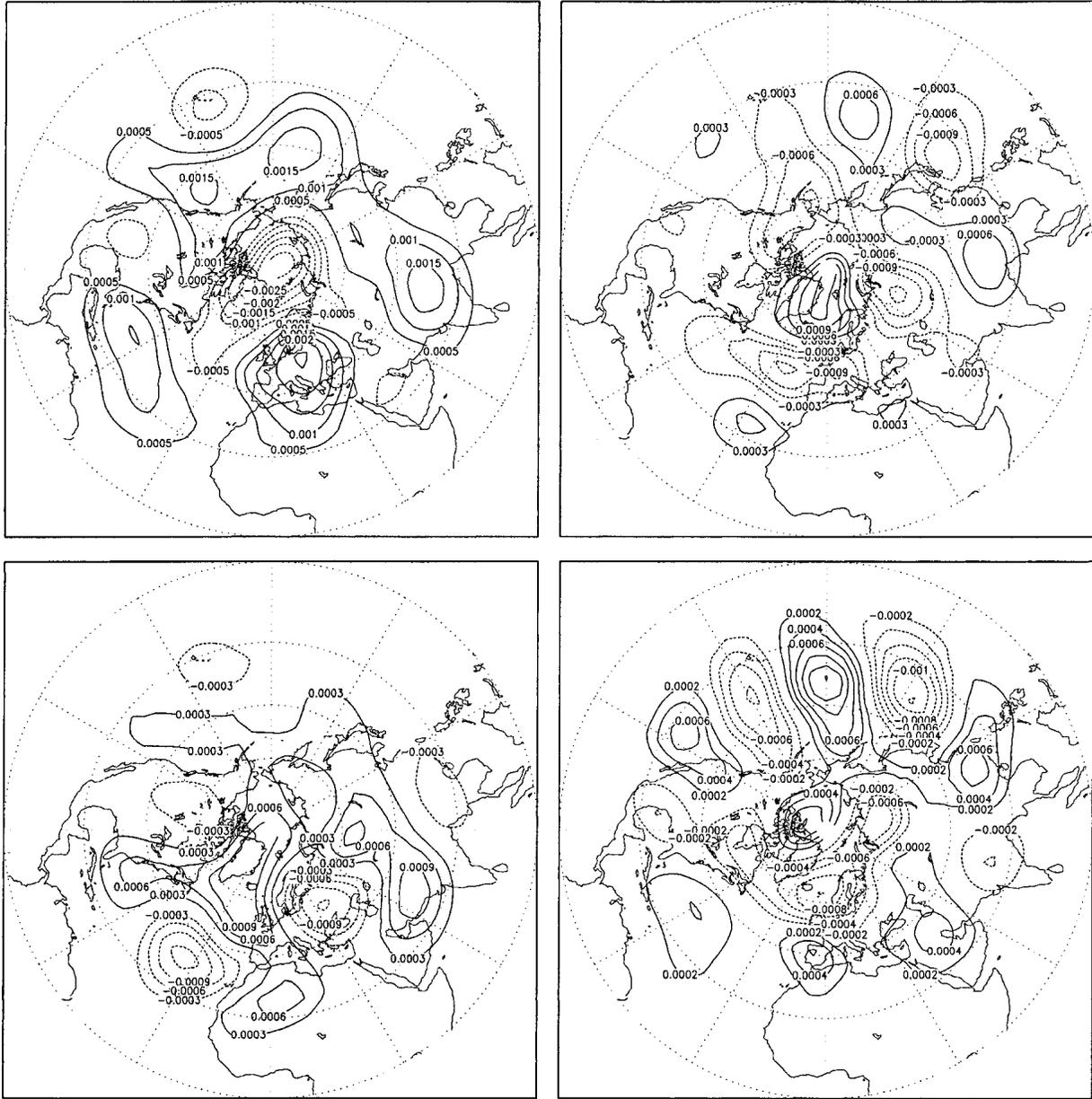


FIG. 3. The first four energy metric EOFs for 500-hPa climate. (upper left) EOF1, (upper right) EOF2, (lower left) EOF3, (lower right) EOF4.

in the context of the information theory (Cover and Thomas 1991), by the relative entropy

$$P(p, \Pi) = \int_{\mathbb{R}^N} p \ln(p/\Pi) dx. \quad (6)$$

The relative entropy in (6) possesses the following properties:

- $P(p, \Pi) \geq 0$ for any two probability densities p and Π ; $P(p, \Pi) = 0$ if and only if $p = \Pi$.
- $P(p, \Pi)$ is invariant under an arbitrary change of coordinates.

Throughout this paper, the relative entropy $P(p, \Pi)$ is used to estimate the lack of information in Π relative to p (or, alternatively, the information content in p relative to Π) in the following situations:

- lack of information in a climatological mean state relative to a forecast ensemble,
- lack of information in the Gaussian approximation to the forecast ensemble relative to forecast ensemble itself,
- the flow in time of the information content in a joint probability density for two phase space variables relative to the direct product of their one-dimen-

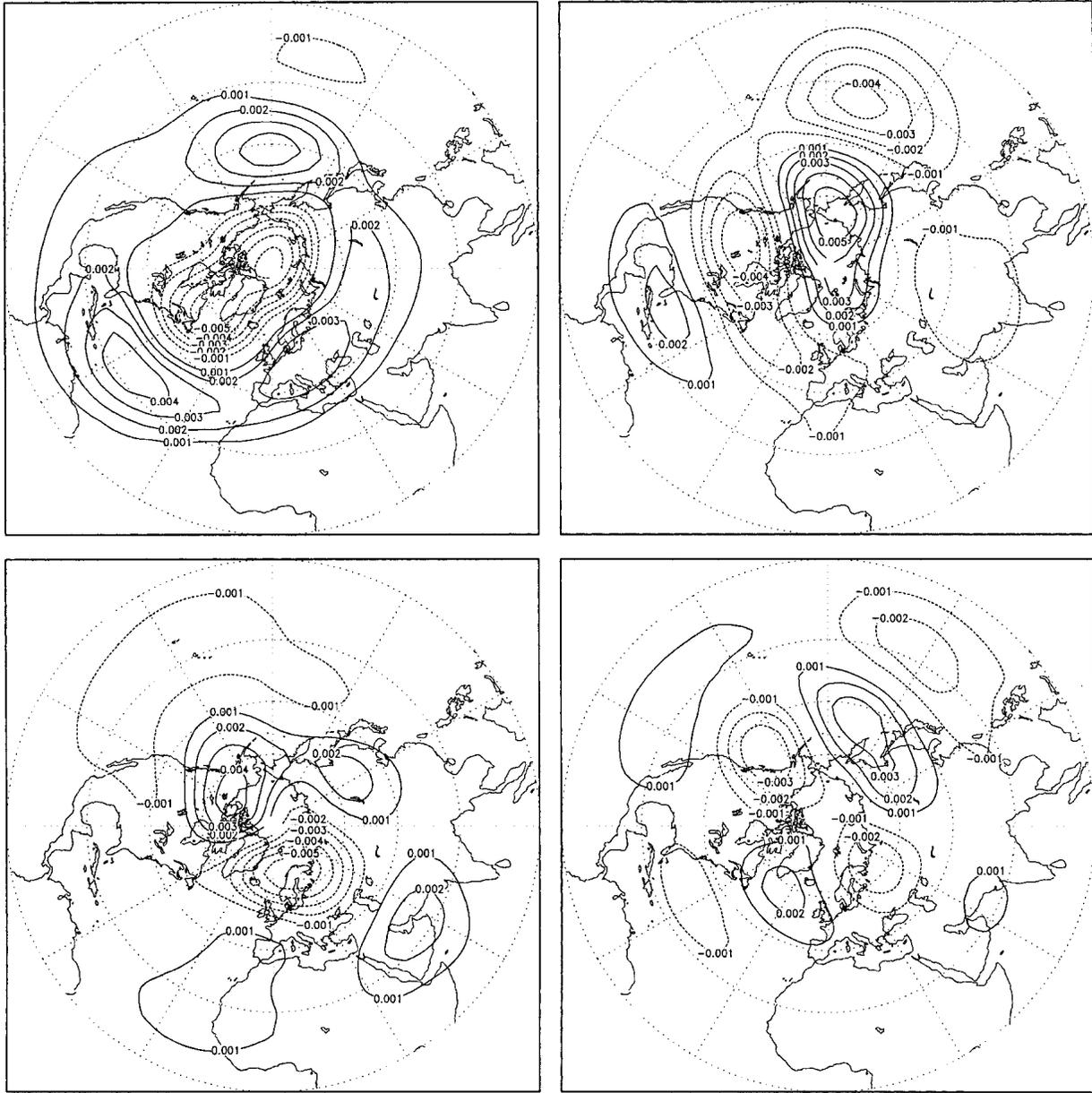


FIG. 4. As in Fig. 3 but for 300-hPa climate.

sional probability densities. This is the information flow in an ensemble prediction.

a. Two lower bounds on relative entropy

The need to compute relative entropy in (6) numerically for a large multidimensional phase space is simplified through the following assumption,

$$\Pi(\mathbf{x}) = \prod_i^N \times \Pi_i(x_i), \tag{7}$$

that is, throughout this paper the prior probability density $\Pi(\mathbf{x})$ is approximated by the product of its marginal

probability densities. As shown in MKC and A–M, this is an excellent assumption in many applications. With (7), the following identity holds:

$$P(p, \Pi) = P\left(p, \prod_i p_i\right) + \sum_i^N P(p_i, \Pi_i). \tag{8}$$

Since the relative entropy of any two probability densities is nonnegative, then relative entropy $P(p, \Pi)$ is bounded from below by

$$P(p, \Pi) \geq \sum_i^N P(p_i, \Pi_i). \tag{9}$$

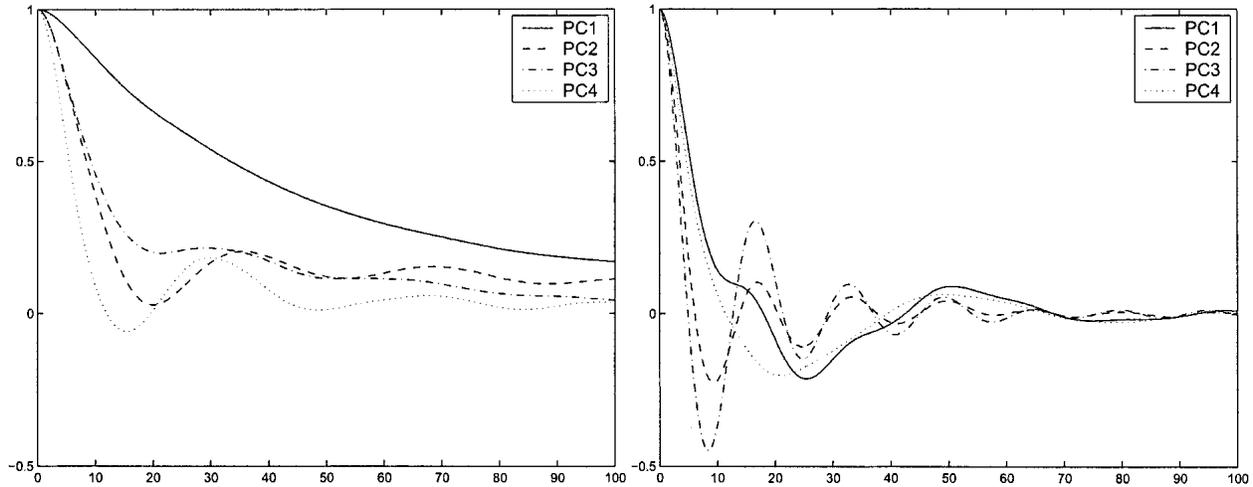


FIG. 5. The time autocorrelations of PC1–PC4 for (left) 500- and (right) 300-hPa climate.

The estimate in (9) is the first practical estimate for computations because, as shown later, the sum of N one-dimensional entropies can be computed quickly. Note that the first term on the right-hand side of (8) is zero only if phase space variables are statistically independent of each other; that is, they factor into a product distribution: therefore, it represents the mutual information (Cover and Thomas 1991) in the phase space coordinates for the probability density p . This mutual information in N variables is bounded from below by twice the averaged sum of the mutual information in each pair of variables, which produces the improved lower bound on relative entropy:

$$P(p, \Pi) \geq \frac{2}{N} \sum_{i=1}^{N-1} \sum_{j=i+1}^N P(p_{ij}, p_i \times p_j) + \sum_i P(p_i, \Pi_i). \quad (10)$$

The result in (10) is a symmetric version of the main theorem in section 4 of MKC, and this extension is presented in the appendix. Abramov and Majda develop optimization algorithms to estimate the two-dimensional mutual information in (10) rapidly. Thus, (10) allows for quick evaluation of lower bounds for information content in many variables. The change in time of the mutual information of two phase space variables during an ensemble prediction is the information flow.

b. Evaluation of relative entropy through a constrained optimization

The probability density $p(\mathbf{x})$ often represents the short-term ensemble forecast, and therefore it cannot be directly sampled over the phase space with good precision, unless the phase space is low-dimensional.

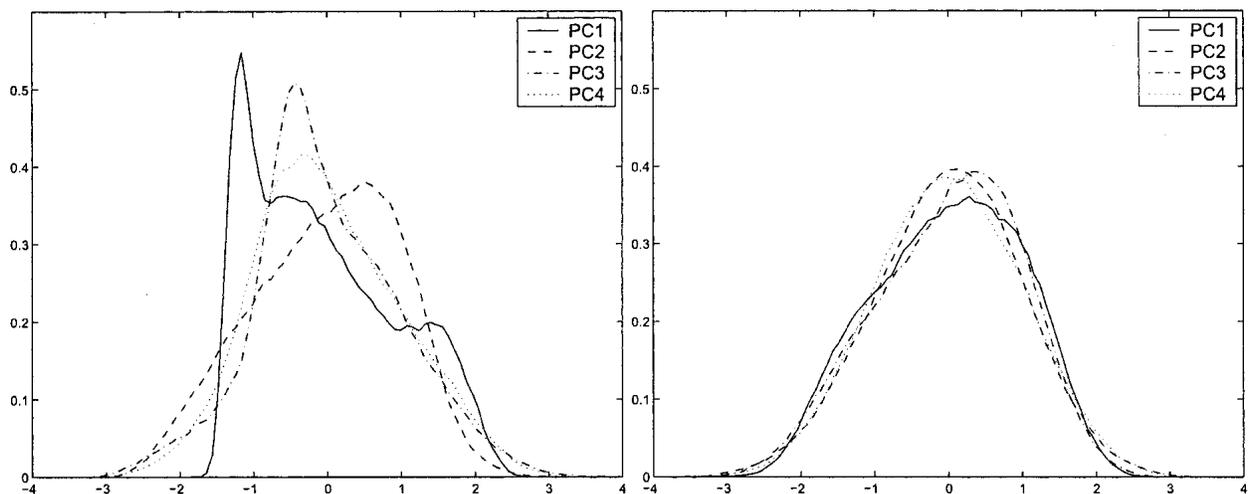


FIG. 6. The PDFs of PC1–PC4 for (left) 500- and (right) 300-hPa climate.

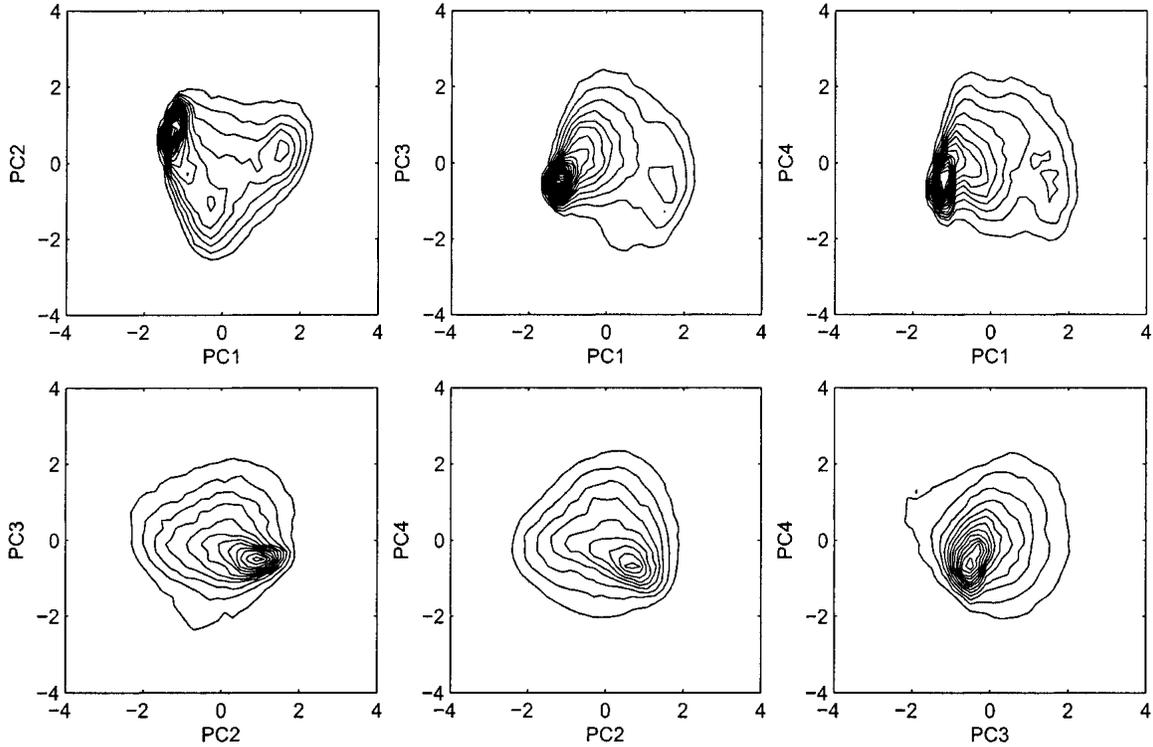


FIG. 7. The joint two-dimensional PDFs of the principal components PC1–PC4 displayed as contour plots, 500-hPa climate. Height increment per contour is 0.02.

Estimates for the loss of information in phase space sampling are given by Kleeman and Majda (2004, manuscript submitted to *J. Atmos. Sci.*). In some practical applications used in the course of this paper, the prior probability density $\Pi(\mathbf{x})$ is also related to the forecast ensemble. Thus, the direct numerical quadrature for relative entropy is inefficient.

Here, the moment optimization approach from MKC and A–M is adopted for evaluation of both $p(\mathbf{x})$ and $\Pi(\mathbf{x})$. Given the mean state and moments of p and Π

$$\begin{aligned} \bar{\lambda}_p &= \int \lambda p(\lambda), & f_i &= \int (\lambda - \bar{\lambda}_p)^i p(\lambda), & 0 \leq i \leq 2L, \\ \bar{\lambda}_\Pi &= \int \lambda \Pi(\lambda), & g_j &= \int (\lambda - \bar{\lambda}_\Pi)^j \Pi(\lambda), & 0 \leq j \leq 2M, \end{aligned} \tag{11}$$

the approximations p_{2L}^* and Π_{2M}^* are found through the maximization of the absolute (Shannon) entropy subject to the moment constraints in (11). A Lagrangian multiplier calculation yields the explicit form of p_{2L}^* and Π_{2M}^* as

$$\begin{aligned} p_{2L}^*(\lambda) &= \exp \left[\sum_{k=0}^{2L} \gamma_k (\lambda - \bar{\lambda}_p)^k \right], \\ \Pi_{2M}^*(\lambda) &= \exp \left[\sum_{k=0}^{2M} \alpha_k (\lambda - \bar{\lambda}_\Pi)^k \right], \end{aligned} \tag{12}$$

where γ_k and α_k are the Lagrangian multipliers for p and Π , respectively. The approximations in (12) yield an additional hierarchy of rigorous lower bounds for relative entropy, described in detail in MKC and A–M. It is important to note that use of only the first and second moments in (11) and (12) yields Gaussian distributions so that the explicit Gaussian formulas for relative entropy (Kleeman 2002) are the simplest lower bounds in the hierarchy (MKC; A–M). The high efficiency of the moment-constrained optimization is shown in Fig. 9, where the highly non-Gaussian climatological PDFs for the 500-hPa climate are compared to the Gaussian and four-moment optimized approximations, for both one- and two-dimensional estimates.

c. Signal–dispersion–cross-term decomposition

For the constrained lower bound approximations in (12), the relative entropy is evaluated through the explicit formula

$$P(p_{2L}^*, \Pi_{2M}^*) = S + D + \text{CT}, \tag{13}$$

where the three terms in the right-hand side of (13) are

$$\begin{aligned} \text{Signal:} & \quad S = - \sum_{k=1}^{2M} \alpha_k (\bar{\lambda}_p - \bar{\lambda}_\Pi)^k, \\ \text{Dispersion:} & \quad D = \gamma_0 - \alpha_0 + \sum_{k=2}^{2L} (\gamma_k - \alpha_k) f_k, \end{aligned}$$

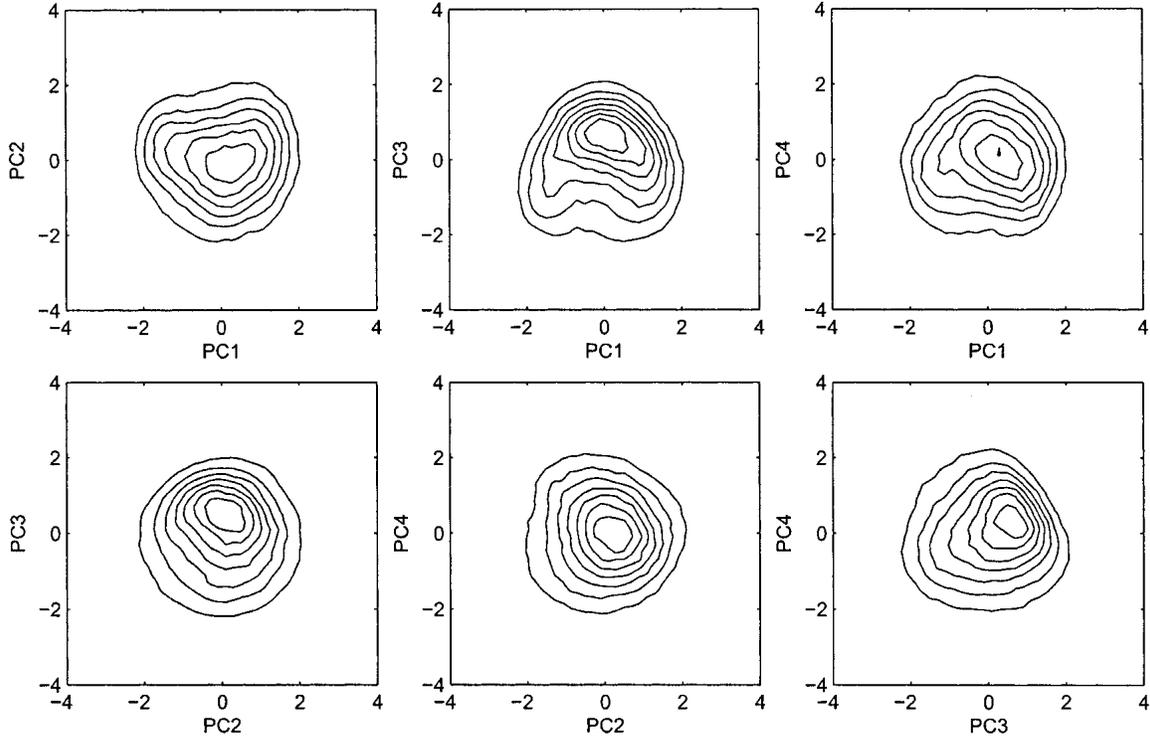


FIG. 8. As in Fig. 7 but for the 300-hPa climate.

Cross-term:

$$CT = - \sum_{k=2}^{2L} \sum_{m=k+1}^{2M} \frac{m!}{k!(m-k)!} \alpha_m f_k(\bar{\lambda}_p - \bar{\lambda}_\Pi)^{m-k}. \quad (14)$$

For a fixed Π , the signal S depends only on the expectation of p , the dispersion D depends only on the variance and higher moments of p , and the cross-term CT depends on both mean and moments of p . The signal, dispersion, and cross-term possess the following properties:

- if Π_{2M}^* is Gaussian (i.e., $\Pi_{2M}^* = \Pi_2^*$), then the cross-term CT is zero;
- if the means of p and Π are equal, then both signal and cross-term are zero;
- the dispersion D is always nonnegative; and
- the signal and cross-term do not depend on Lagrange multipliers for p and, therefore, do not require the optimization problem for p to be solved.

Following A–M, we propose the *generalized signal*

$$GS = S + CT \quad (15)$$

as a computationally inexpensive estimate for information content in p relative to Π in the case that Π represents the climatology.

5. Predictability

Here the information theoretical framework summarized in section 4 is applied to ensemble predictions for

the T21 barotropic model. This approach requires that various moments are measured for both the ensemble prediction and climatology. In the case of the climatological probability density, moments are computed, as usual, by time averaging of a long-term solution of T21 barotropic model. The forecast ensembles are computed through experiments with the following design.

- Each individual numerical realization of relative entropy for a forecast ensemble is computed from a 500-member statistical ensemble of solutions of T21 barotropic model in (1), where the ensemble itself is generated by perturbations with small variance around a single-time snapshot of single long-term climatological solution of (1).
- Statistical properties of relative entropy and its components are derived by averaging over a set of 100 forecast ensembles (ensemble of forecast ensembles). To generate these ensembles, 100 independent snapshots of a single long-term climatological solution were recorded, separated by an interval of 2000 days.

Generating the forecast ensemble around a snapshot of the climatological solution imitates the errors of observation, where the initial variance is derived from average precision of a measuring device. As a simple imitation of measurement error, it is common to generate the initial forecast ensemble via a Gaussian distribution, with the variance being several orders of magnitude smaller than climatological variance (Klee-

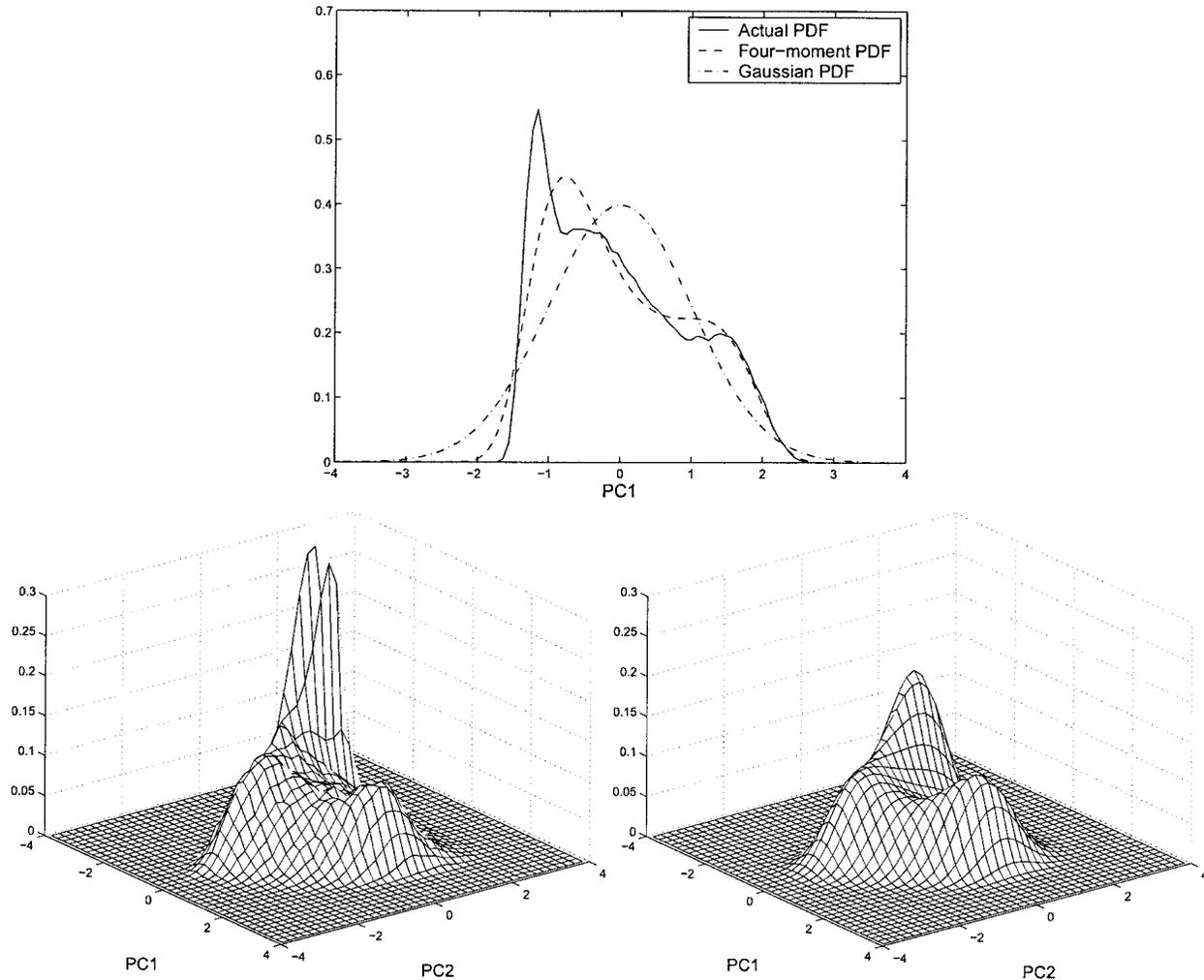


FIG. 9. The efficiency of the four-moment approximation to the actual PDFs. (top) Climatological PDF for 500-hPa PC1 (solid line), approximated by four-moment (dashed line), with $P(\Pi, \Pi_4^*) = 3.569 \times 10^{-2}$, and Gaussian (dot-dashed line), with $P(\Pi, \Pi_G) = 0.1565$ estimators. (bottom) Joint climatological PDF PC1-PC2 for (left) 500-hPa climate, approximated by the (right) two-dimensional four-moment estimator with $P(\Pi, \Pi_4^*) = 5.736 \times 10^{-2}$.

man 2002). In this study, two different techniques of generating initial ensembles are used.

- 1) The ensemble is generated by perturbing each principal component of the snapshot of the long-term climatological solution via a Gaussian distribution with variance 10^{-5} (equivalent to perturbing the snapshot along each EOF with variance $10^{-3}\%$ of the climatological variance in that direction).
- 2) The ensemble is generated by perturbing the streamfunction at each physical grid point independently via a Gaussian distribution with absolute variance 10^{-10} (500-hPa climate) and 10^{-9} (300-hPa climate) in the model variables. These values of initial variance in physical space are chosen to match approximately the initial variance of the corresponding PC ensemble to 10^{-5} .

The two methods of generating a forecast ensemble have different properties. Since the EOF basis organizes climatological variability into an ordered set of components, perturbations along these directions by a fixed fraction of total EOF variance produce correlated fluctuations in streamfunction with varying magnitudes in different regions of physical space. In contrast, the physical space method produces disordered small-scale fluctuations, which is what one expects to see in nature practically. Clearly the temporal response to such fluctuations could be different: while the small-scale fluctuations are quickly damped by diffusion, the large-scale fluctuations can persist for a long time. It turns out that the results of our study are usually insensitive to these two different ensemble generation techniques (with a few exceptions), and further comments on the ensemble generating method are made only when a difference in results is observed.

a. Relative lack of information in the climate

The time series of the lack of information in the climate PDFs relative to the ensemble forecast, averaged over 100 realizations for individual PCs 1 through 4, are shown in Fig. 10. All first four PCs have a universal decay pattern for any of the two types of climate and any of the two ensemble-generating methods. The rate of decay for relative information is 3 times faster in the 300-hPa climate (50 days against 150 days), regardless of ensemble generation method. The slower decay of the lack of information in the 500-hPa climate is apparently related to the more robust and chaotic dynamics of the 300-hPa climate. However, mixing does not control relative entropy. Like in the Lorenz '96 model (A-M), the lack of information in climate relative to forecast ensemble in a given PC is not characterized by the autocorrelation decay time in that PC, contrary to some expectations. The correlation functions for the first four

PCs are depicted in Fig. 5 and clearly have different correlation times, unlike relative entropy in Fig. 10. PC1 of the 500-hPa climate has small additional information compared to the other PCs (this is common for both types of ensemble generation), whereas all PCs of the 300-hPa climate share it nearly equally. This slight anomaly in behavior of 500-hPa PC1 might be the consequence of what was demonstrated in section 3—PC1 for the 500-hPa climate stands out by its high variability (Fig. 2) and autocorrelation scale separation (Fig. 5). However, the impact that such pronounced climate differences have on the ensemble-averaged relative entropy is rather small. The two ensemble-generating methods barely affect the results of Fig. 10. The only noticeable difference is that while for PC ensemble generation the relative entropy plots decay moderately at initial and later times, for physical space streamfunction ensemble generation the behavior is slightly different: the curves start out steeply with a high rate of decay,

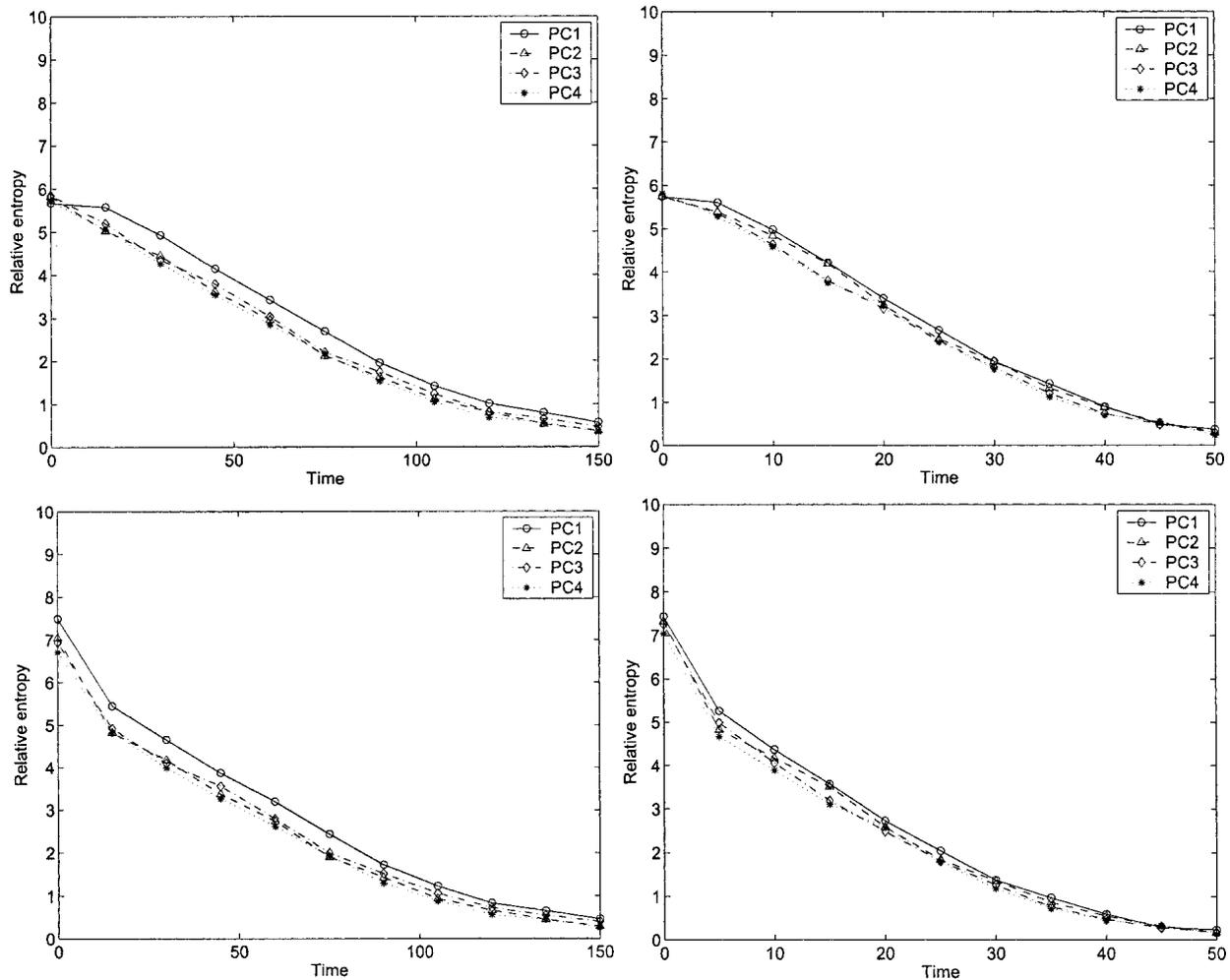


FIG. 10. Average (over 100 ensembles) lack of information in the climate relative to the 500-member ensemble forecast. The $P(p_4^*, \Pi_4^*)$ four-moment estimator is used. (left) The 500- and (right) 300-hPa climate. (top) PC ensemble breeding, (bottom) streamfunction ensemble breeding. Solid line is PC1, dashed line is PC2, dot-dashed line is PC3, dotted line is PC4.

then slow down to the same moderate rate as for PC ensemble generation. This difference is caused by the scale-selective diffusion: small-scale fluctuations, produced by streamfunction ensemble generation, are damped rapidly, which impacts relative entropy; however, the scale-ordered fluctuations of PC ensemble generation are more persistent and less vulnerable to small-scale diffusion. This behavior is common for all PCs, but manifests itself only locally for short initial times, and overall relative entropy decay stays unaffected.

The variability of relative entropy through the 100-member ensemble of ensemble predictions is shown in Fig. 11 via histograms for PC1. The histograms for both types of climate have similar spread at times that scale by the factor of 3 between climates. However, the shapes of histograms are different: while the relative entropy histograms for the 300-hPa climate have shapes close to a Gaussian (with the exception of the last snapshot at $t = 40$), for the 500-hPa climate only the first snapshot at time $t = 30$ looks Gaussian. The 500-hPa histograms for the second ($t = 60$) and third ($t = 90$) snapshots are visibly skewed.

b. Relative lack of information in the first two ensemble moments (non-Gaussianity)

In practical forecasts, usually only ensemble mean and spread (variability) are taken as indicators of predictive utility of that ensemble, which in the relative entropy framework corresponds to the Gaussian approximation of a forecast ensemble. Thus, a natural question is how much information the Gaussian approximation p_G lacks relative to that in the prediction ensemble p ? For example, this is important for detection of bimodality in a prediction ensemble. The relative entropy $P(p, p_G)$ is the measure of such a lack of information. Alternatively, one can view $P(p, p_G)$ as

the measure of non-Gaussianity: if p is Gaussian, then $P(p, p_G)$ is zero. In practice it is impossible to measure $P(p, p_G)$ with finite-sample ensemble forecasts. Instead, the lack of information in the Gaussian approximation p_G is measured relative to some higher-order approximation to p . In this paper, it is measured relative to the four-moment approximation p_4^* (section 4 and A–M). Since the mean states of p_4^* and p_G are equal, the signal and cross-term are zero, and the relative entropy consists entirely of dispersion. The plots averaged over 100 ensemble realizations $P(p_4^*, p_G)$ for PC1–PC4 are shown in Fig. 12. Temporal behavior of ensemble non-Gaussianity is universal for all PCs and both types of climate. First, there is a time interval where $P(p_4^*, p_G)$ is small (about three orders of magnitude smaller than the initial lack of information in the climate). Then, around a certain time ($t = 60$ for 500-hPa climate and $t = 20$ for 300-hPa climate), non-Gaussianity grows rapidly, and by the end of observations ($t = 150$ for 500-hPa climate and $t = 50$ for 300-hPa climate) it becomes only an order of magnitude smaller than the lack of information in the climate at that time. Note that the non-Gaussianity is always at least an order of magnitude smaller than the lack of information in the climate at the same time. Thus, the conventional wisdom, stating that it is enough to know only the mean and variance of the forecast ensemble to estimate its predictive utility relative to the climate, is correct for the T21 barotropic model. This behavior of ensemble non-Gaussianity is universal for all large-scale PCs; also, it occurs for both 500-hPa and 300-hPa type climates on time intervals, that scale by a factor of 3, like the decay of the lack of information in the climate. After that, non-Gaussianity in PC1 is generally greater than in the rest of PCs (20%–30% for 300 hPa, and more than 100% for 500 hPa).

Non-Gaussian growth is triggered by ensemble skew-

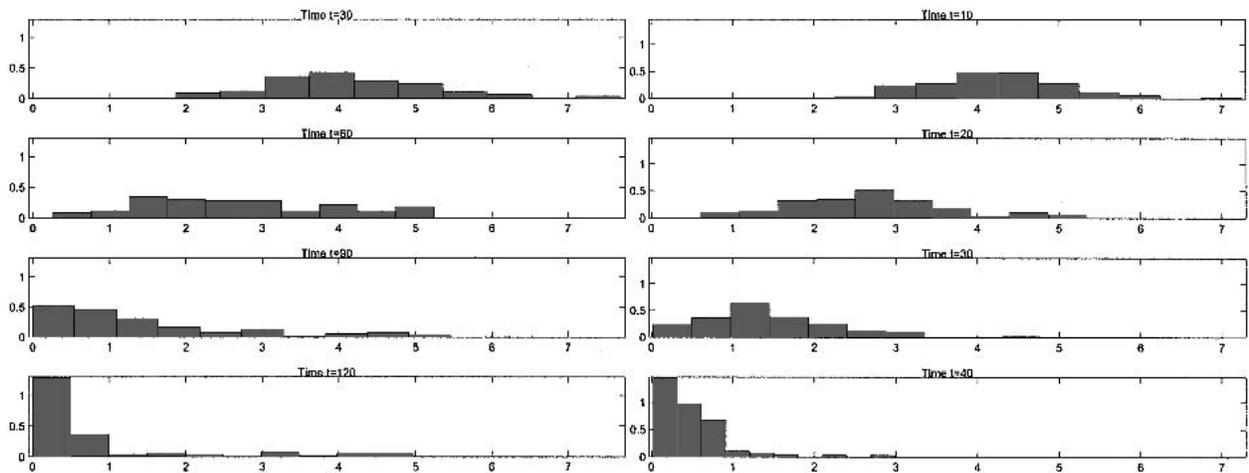


FIG. 11. Ensemble histograms of relative entropy $P(p_4^*, \Pi_4^*)$ for PC1. (left) Snapshots at $t = 30, 60, 90, 120$ for 500-hPa climate; (right) snapshots at $t = 10, 20, 30, 40$ for 300-hPa climate.

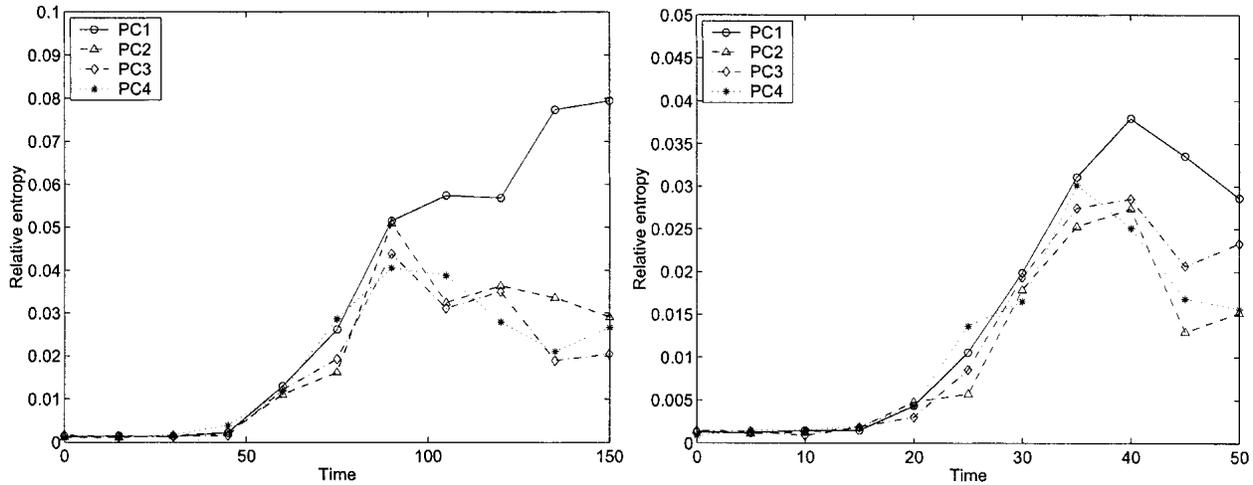


FIG. 12. Average (over 100 ensembles) lack of information $P(p_4^*, p_G)$ in the Gaussian forecast approximation relative to the four-moment forecast approximation. (left) The 500-hPa climate, (right) 300-hPa climate. Solid line is PC1, dashed line is PC2, dot-dashed line is PC3, dotted line is PC4. Results are shown for PC ensemble breeding.

ness and flatness when they deviate from their Gaussian values, which, in turn, could be induced by nonlinear effects in an ensemble with growing spread. Nonlinearity in a forecast ensemble should become effective when the average deviation of the ensemble members from their mean state exceeds some critical value. For practical purposes, this critical value is crudely estimated as 0.1, an order of magnitude smaller than the climatological variability of a principal component. The standard deviations of the forecast ensemble for the first four PCs, averaged over 100 ensembles, are shown in Fig. 13. By comparing Figs. 12 and 13, it is observed that average deviations from the ensemble mean become less than an order of magnitude smaller than 1 approximately at the same time as

non-Gaussianity starts developing. This observation is valid for both 500-hPa and 300-hPa types of climate for all four principal components shown, and serves as strong evidence that the non-Gaussian growth in the ensemble occurs due to nonlinearity. The precise reason why ensembles diverge with those particular rates, however, is not studied here.

c. Signal, dispersion, and cross-term contribution to variations in utility

Due to the different physical nature of information originating from signal, dispersion, and cross-term, it is important to know how these components contribute to the total relative entropy in measuring the lack of information in the climate. The ensemble of ensembles

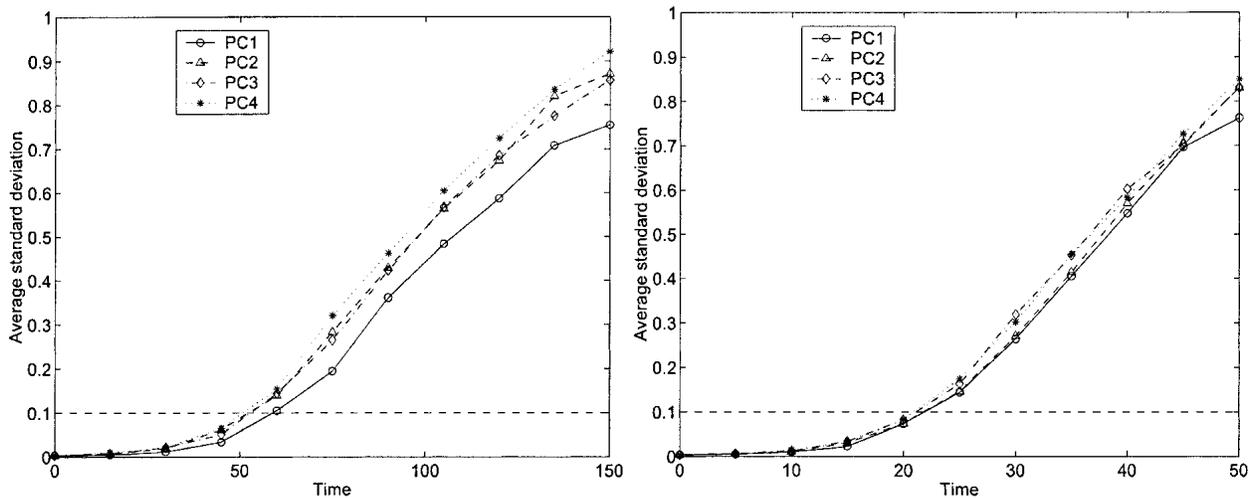


FIG. 13. Average (over 100 ensembles) standard deviations of 500-member ensemble. (left) The 500-hPa climate, and (right) 300-hPa climate; PC as in Fig. 12. Results are shown for PC ensemble breeding. Horizontal dashed line shows critical value 0.1.

mean contributions of signal, dispersion, and cross-term to the lack of information $P(p_4^*, \Pi_4^*)$ in the climate relative to the forecast ensemble for PC1 are shown in Fig. 14. For both 500-hPa and 300-hPa types of climate the situation is the same, namely, dispersion is the main contributor to the averaged relative entropy. Though only PC1 is shown, this situation is common for all large-scale PCs and does not depend on the type of ensemble generation used.

To understand the variability of the results in Fig. 14 over individual ensembles, the contribution of generalized signal and dispersion to variability of $P(p_4^*, \Pi_4^*)$ is shown in Fig. 15 for the 500-hPa climate via scatterplots of generalized signal and dispersion versus relative entropy for individual ensembles (Kleeman 2002). For both 500-hPa and 300-hPa (not shown) types of climate, the dispersion controls variability of relative entropy at all times. This can be seen from the fact that fluctuations of relative entropy and dispersion correlate (the scatterplots for dispersion are diagonally aligned), whereas the generalized signal does not change much. This situation for barotropic flow on the sphere contrasts strongly with that for both the truncated Burgers–Hopf model (Kleeman et al. 2002) and for standard models of baroclinic turbulence (Kleeman and Majda 2004, manuscript submitted to *J. Atmos. Sci.*), where the signal dominates variations in relative entropy for leading EOF modes. This situation also differs from the Lorenz '96 model (A–M), where the signal controls variability at later times, and from the Lorenz three-component model (Kleeman 2002) where non-Gaussian effects dominate at longer ranges. While not shown in detail here, in most situations dispersion is entirely controlled by its Gaussian part (terms of second and lower order), with the non-Gaussian part being negligible.

d. Realizations with high signal

The average contribution of signal to the total lack of information in the climate is negligible in comparison with the dispersion contribution. Thus, rare ensemble forecasts, when the signal is high compared to its average, are important. Due to the fact that signal depends only on ensemble mean, it naturally turns out that, whenever the signal in a given large-scale PC is by itself much higher than the average signal in the collection of large-scale PCs, the ensemble mean state (with subtracted climatological mean state) is strongly projected onto the corresponding EOF. In Fig. 16 two examples of such behavior are shown, one for the 500-hPa and another for the 300-hPa type of climate. The example for 500-hPa climate (on the left side of Fig. 16) illustrates a real situation that occurred when the signal in PC3 was high compared to the average signal in the first four PCs. The ensemble mean looks like the actual EOF3 in Fig. 3; in particular, the NAO-like pattern in EOF3 is repeated in the ensemble mean state. Similar behavior is observed for the 300-hPa climate (on the right side of Fig. 16). In this case, an ensemble prediction with high signal in PC2 was located, and indeed, like for the 500-hPa climate, the ensemble mean looks very similar to EOF2 (Fig. 4), including the PNA-like pattern.

In addition to the two examples above, systematic quantification of the correlation between high signal in a PC and projection onto the corresponding EOF is provided through the physical space correlation

$$\text{corr}(f, g) = \frac{(f, g)}{\|f\| \|g\|}, \quad (16)$$

where (f, g) is the energy metric inner product, and $\|f\| \equiv (f, f)^{1/2}$. Clearly, $\text{corr}(f, g) = \pm 1$ only if $f \equiv \pm \alpha g$ for

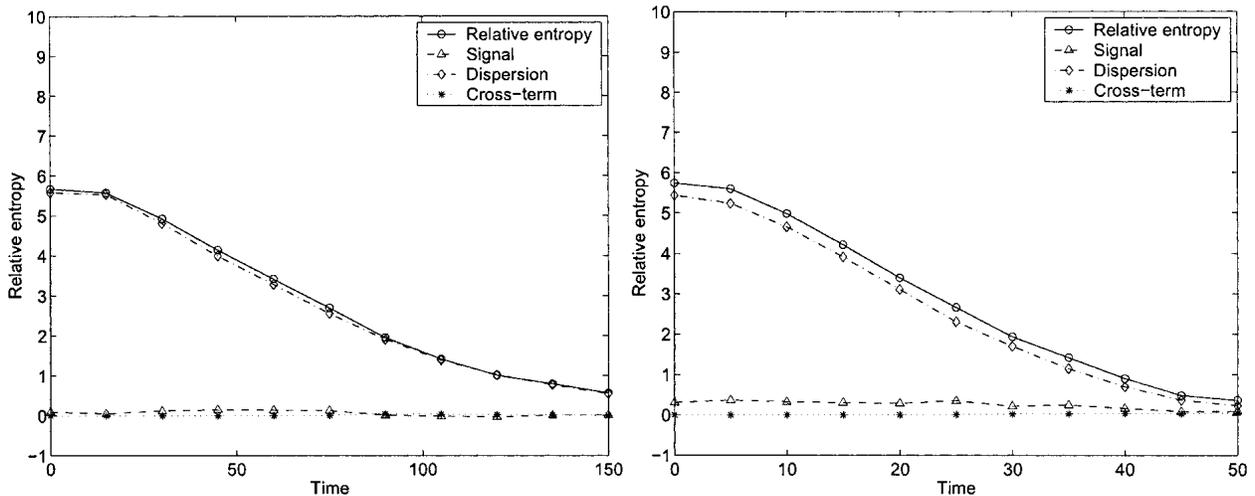


FIG. 14. Average (over 100 ensembles) signal–dispersion–cross-term contribution to the lack of information in the climate relative to forecast ensemble $P(p_4^*, \Pi_4^*)$, PC1. (left) The 500- and, (right) 300-hPa climate. Solid line is relative entropy, dashed line is signal, dot-dashed line is dispersion, dotted line is cross term.

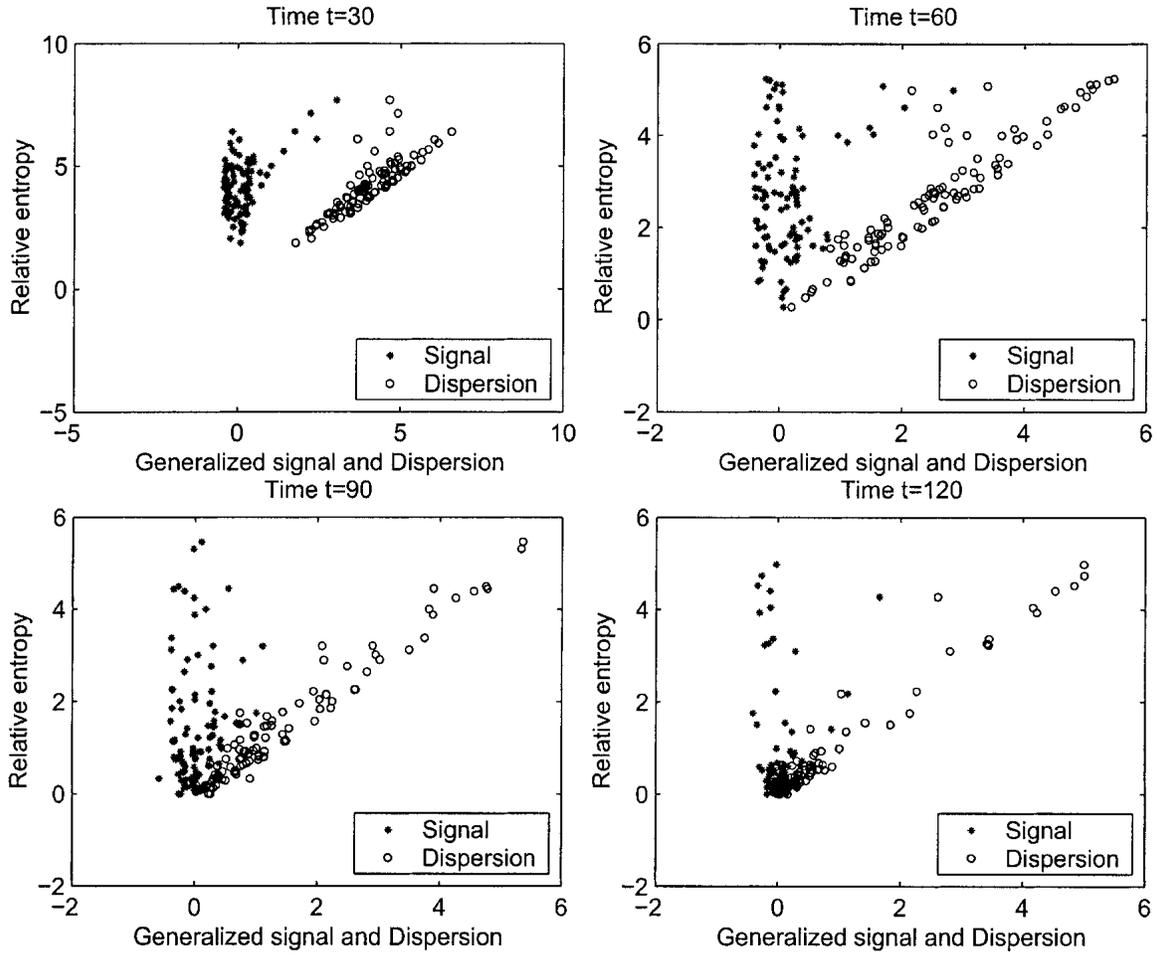


FIG. 15. Ensemble scatterplots of generalized signal and dispersion vs relative entropy $P(p_4^*, \Pi_4^*)$, time $t = 30, 60, 90, 120$ for PC1 and 500-hPa climate. Asterisks are the generalized signal, circles are dispersion.

some positive α (i.e., absolutely similar) and $\text{corr}(f, g) = 0$ if f and g are totally uncorrelated. In Fig. 17, for the ensemble of ensembles (100 realizations of a 500-member ensemble) and all discrete time snapshots, scatterplots of physical correlation between the ensemble mean and EOF1, are plotted versus a suitably normalized generalized signal in PC1

$$|\text{corr}(\text{ensemble mean, EOF1})| \text{ vs } \text{GS}(\text{PC1})/\text{GS}(\text{PC1} \dots \text{PC4}).$$

The scatterplots show strong correlation between high generalized signal and high similarity between ensemble mean state and EOF1: when the generalized signal is high, the physical correlation is also high; when the generalized signal is low, the correlation is close to zero. However, weaker correlation (a more blurred scatterplot) can be observed for the 500-hPa climate, whereas the scatterplot for the 300-hPa climate shows almost perfect correlation. This observation justifies

the use of the generalized signal as an inexpensive measure for some aspects of the lack of information in the climate: for high generalized signal in a PC, the physical pattern of the ensemble mean is highly predictable in terms of the projection onto the corresponding EOF. In fact, a high generalized signal in PC3 for the 500-hPa climate causes, with high probability, a NAO-like zonal-blocking structure to appear in the local flow; the same reasoning is valid for the 300-hPa climate, PC2, and the PNA (Fig. 16 is an example). The scatterplots are presented here only for PC1, but the same situation is observed as well for PC2, PC3, and PC4.

e. Information flow

As discussed below (10), the mutual information, $P(p_{ij}, p_i \times p_j)$ is an indicator of factorizability and mutual statistical dependence of involved phase space variables in an ensemble prediction; when $P(p_{ij}, p_i \times p_j)$ is small in magnitude, the p_i and p_j are nearly indepen-

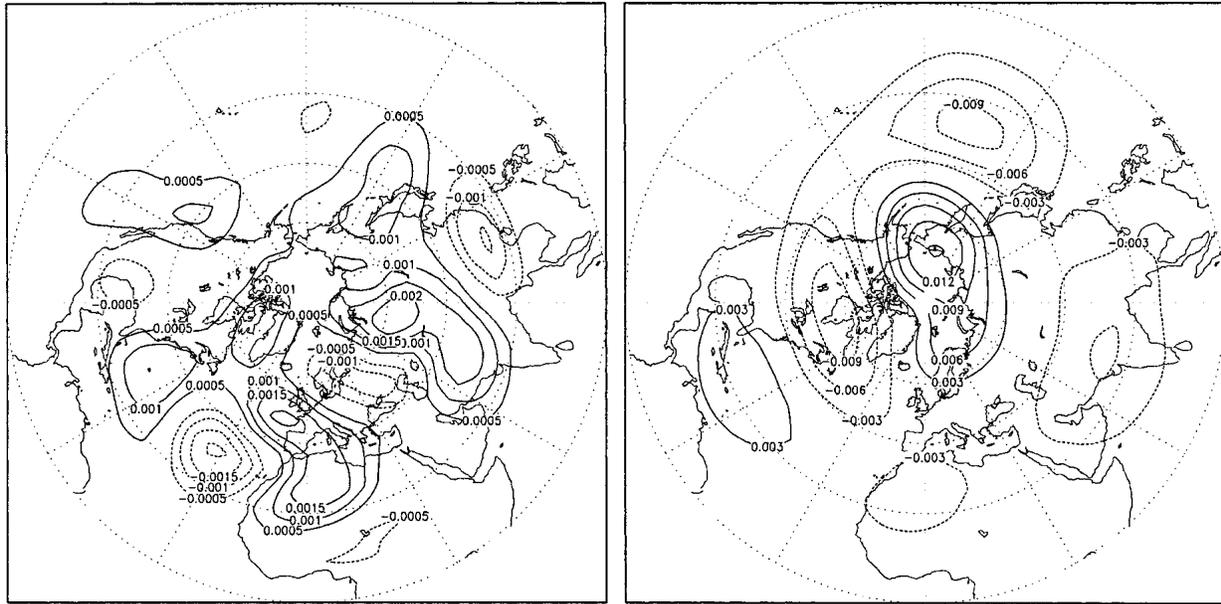


FIG. 16. Rare outcomes with high signal—physical similarity between ensemble mean and EOFs. (left) Ensemble mean for the 500-hPa climate strongly projects on EOF3 (Fig. 3) when generalized signal is high in PC3. (right) Ensemble mean for the 300-hPa climate strongly projects on EOF2 (Fig. 4) when generalized signal is high in PC2.

dent while large values yield strong dependence, and, according to (10), a large contribution to the information. The information flow among PCs is the temporal behavior of the mutual information as time evolves in an ensemble prediction. To evaluate the information flow in the forecast ensemble between two PCs, the relative entropy with the four-moment smooth approximation $P(p_{ij,A}^*, p_{i,A}^* \times p_{j,A}^*)$ is used. Averages over all prediction ensembles of the information flow for all pairs of first four PCs are shown in Fig. 18. Despite the

climate differences, here systematic trends in the information flow are similar for 500-hPa and 300-hPa types of climate. First, the average information flow has an ascent phase and a descent phase. The timing of the phase change roughly coincides with the beginning of non-Gaussian growth, which means that the average information flow reaches its maximum when nonlinear effects first appear in the forecast ensemble. It is shown later in the paper that the peak in mutual information indicates a significant increase in the ensemble cross-

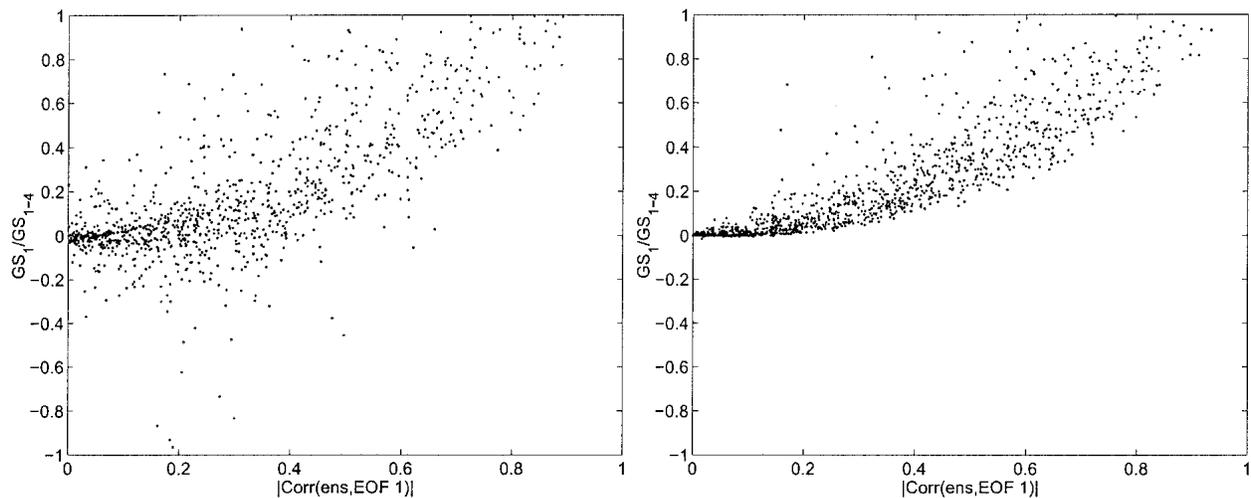


FIG. 17. The scatterplots $|\text{corr}(\text{mean}_{\text{ens}}, \text{EOF1})|$ vs $\text{GS}(\text{PC1})/\text{GS}(\text{PC1} \dots \text{PC4})$ for all time snapshots of 100 realizations of 500-member forecast ensemble. (left) The 500- and (right) 300-hPa climate.

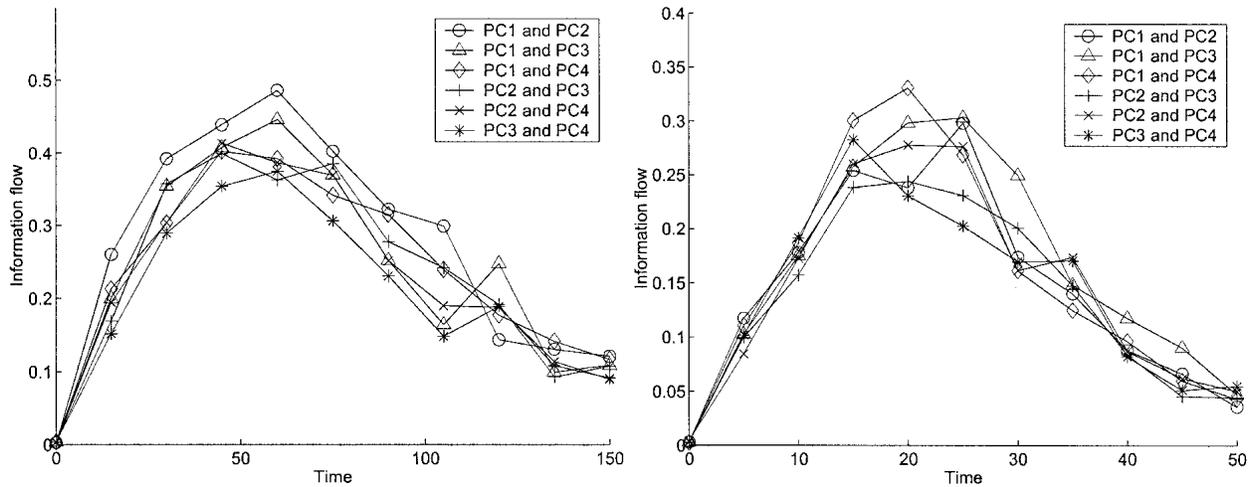


FIG. 18. Average (over 100 ensembles) information flow $P(p_{i,4}^* p_{j,4}^* \times p_{i,4}^*)$ for all pairs of principal components PC1 through PC4. (left) The 500- and (right) 300-hPa climate.

second moment relative to its variance. It means that the cross section of the ensemble pipe is being stretched anisotropically and rapidly along singular directions in phase space, which, in turn, triggers nonlinear behavior in the corresponding Liouville equation and growth of non-Gaussianity. Perhaps the rapid stretching of the ensemble pipe along chosen directions indicates aligning of ensemble members with eigenvectors of the linearized system, as pointed out by Toth and Kalnay (1997). Second, for 500-hPa climate, the average information flow PC1 \leftrightarrow PC2 is the largest of all (probably due to high climate non-Gaussianity of PC1), whereas for the 300-hPa climate such a trend is not observed. The overall information flow for 500-hPa climate is

greater than that for 300-hPa climate by a factor of 1.6, while the physical space ensemble generation does not result in any significant differences in information flow.

The average contribution of the information flow to the total lack of climate information is, however, small. In Fig. 19 the average (over 100 ensembles) of the relative lack of information in the climate for the principal components PC1 through PC4 is shown. The solid line in Fig. 19 denotes the improved two-dimensional estimator from (10), and the dashed line denotes the direct sum of one-dimensional estimators from (9). The average gain in information from the improved estimator does not exceed 7%–8% for either type of the climate. All of the above facts suggest that, while the gains in

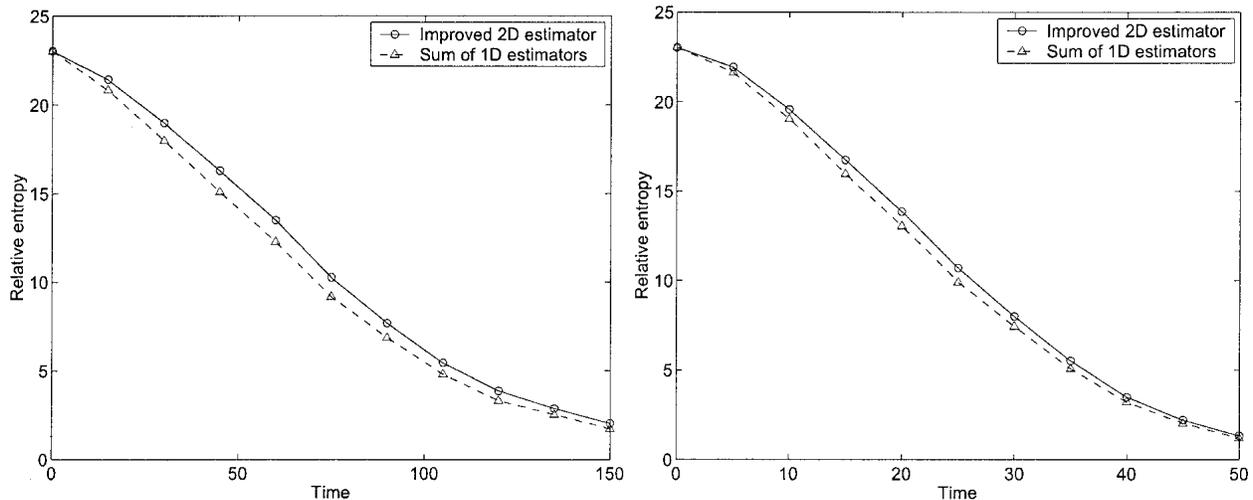


FIG. 19. Average (over 100 ensembles) relative lack of information in the climate for the principal components PC1 through PC4. Solid line is improved two-dimensional estimator from (10), dashed line is direct sum of one-dimensional estimators from (9) for (left) 500- and (right) 300-hPa climate.

estimating information with (10) rather than (9) are small (the one-dimensional estimators are good enough), the information flow can be used with high efficiency in establishing statistical connections between various large-scale flow structures in an ensemble prediction. These results for large-scale behavior in the barotropic model contrast with those for the Lorenz '96 model (A–M), where there are numerous simulations with large contribution of the flow of mutual information in (10) to the relative entropy.

f. Realizations with high information flow

It turns out that the magnitude of information flow $P(p_{ij,4}^*, p_{i,4}^* \times p_{j,4}^*)$ between two PCs depends largely on their normalized cross-second moment

$$M_{ij}^2 = \frac{\langle PC_i PC_j \rangle}{\langle PC_i^2 \rangle^{1/2} \langle PC_j^2 \rangle^{1/2}}, \quad (17)$$

The experimental dependence of information flow $P(p_{ij,4}^*, p_{i,4}^* \times p_{j,4}^*)$ on M_{ij}^2 is shown in Fig. 20 for all times and ensemble realizations. The magnitude of information flow highly correlates with magnitude of the cross-second moment except for a small number of outcomes, induced by non-Gaussian effects. In rare cases information flow can be high, which means creation and destruction of different EOF patterns in the ensemble prediction is statistically dependent. Large values of information flow $P(p_{ij,4}^*, p_{i,4}^* \times p_{j,4}^*)$ can only be induced by a structure of $p_{ij,4}^*$, which is not reproduced well by the direct product $p_{i,4}^* \times p_{j,4}^*$. In Fig. 21 one ensemble realization is shown, corresponding to a case when the information flows $PC1 \leftrightarrow PC3$ (500 hPa) was high compared to the average information flow. The information flow in Fig. 21 is high at various times because the

ensemble in the phase portrait is thin and long, and oriented at a nonzero angle to either of the two PC axes. The cross-second moment plays a big role in such an ensemble, and therefore it cannot be well approximated by a direct product of two one-dimensional PDFs. This phenomenon is illustrated in Fig. 22, where the result of Fig. 21 for the 500-hPa climate at time $t = 75$ is approximated by $p_{13,4}^*$ and $p_{1,4}^* \times p_{3,4}^*$. Clearly, the two-dimensional estimator is vastly superior to the direct product of one-dimensional estimators, when the ensemble second cross-moment is high.

High information flow is important from the physical point of view: with a long and thin ensemble aligned at nonzero angle with two PCs as in Fig. 21, changes in one PC will statistically cause changes in another PC. Good examples of such behavior are the situations in Fig. 21 with the 500-hPa climate at times $t = 45$ and $t = 75$, when the information flow $PC1 \leftrightarrow PC3$ is high. Depending on ensemble orientation (positive or negative slope), there are two mutually transversal scenarios of switching EOF patterns for high information flow in a pair of PCs. Differences between the EOF1–EOF3 patterns for the 500-hPa climate are shown in Fig. 23 for both switching scenarios (scenario with negative slope occurs at $t = 45$, and with positive at $t = 75$). The two scenarios induce different physical changes in the flow. The scenario at $t = 45$ stresses correlations in changes between AO-like and NAO-like zonal-blocking patterns, while the switching scenario at $t = 75$ (transversal to the one at $t = 45$) emphasizes other flow structures.

6. Summary

The relative entropy framework for predictability is tested here on the T21 barotropic model with two types of steady forcing. The resulting two types of climate

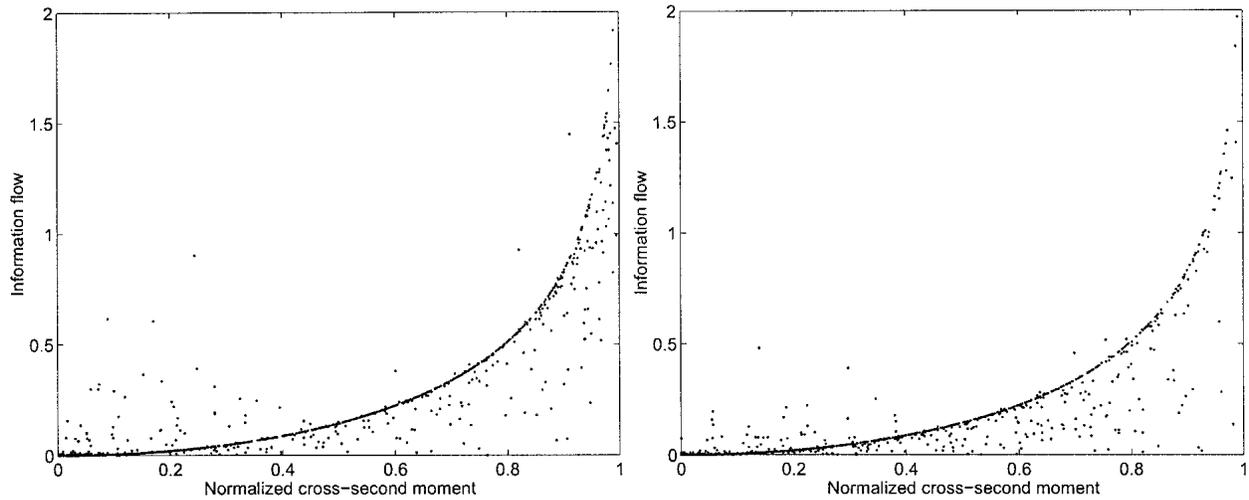


FIG. 20. Experimental dependence of information flow on absolute value of cross-second moment in (17). (left) The 500-hPa climate, $PC1 \leftrightarrow PC3$, (right) 300-hPa climate, $PC1 \leftrightarrow PC2$.

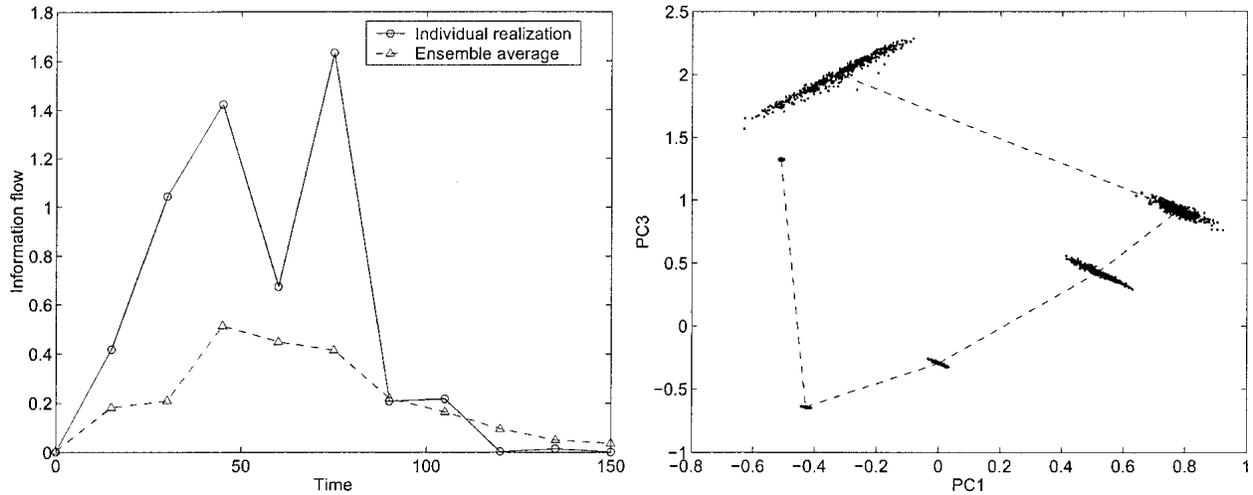


FIG. 21. (left) Individual realization with high information flow $PC1 \leftrightarrow PC3$ for 500-hPa climate. (right) Phase portrait of corresponding ensemble evolution of 500-hPa climate at time $t = 0, 15, 30, 45, 60, 75$.

correspond roughly to realistic flows at 500- and 300-hPa surface heights. Due to different height, these two flows, both being relevant to the actual physics of atmospheric low-frequency variability, have different dynamical and statistical properties. The 300-hPa climate is characterized by robust dynamics, uniform decay of autocorrelation time scaling, Gaussian-looking PDFs, and strong projection of the first two energy metric EOFs onto the Arctic Oscillation and Pacific–North American patterns. The 500-hPa climate has longer autocorrelation decay times with stronger scale separation, non-Gaussian PDFs, and its first and third energy metric EOFs project strongly on the Arctic Oscillation and North Atlantic Oscillation (Crommelin 2003). The different dynamics of the two types of barotropic model

climates impacts their predictability. This allows us to test the recent information theoretical framework for ensemble prediction summarized in section 4 in different regimes of low frequency variability in a simple barotropic model. Large (500 members) ensembles were utilized to avoid dealing with loss of information through small sample size effects, although this is an important practical topic discussed elsewhere (Kleeman and Majda 2004, manuscript submitted to *J. Atmos. Sci.*; Haven et al. 2004, manuscript submitted to *J. Comput. Phys.*). We summarize the results of the ensemble prediction study in the following remarks.

- The lack of information in the climate as estimated by $P(p_{1,4}^*, \Pi_4^*)$ has a different decay time scale for the

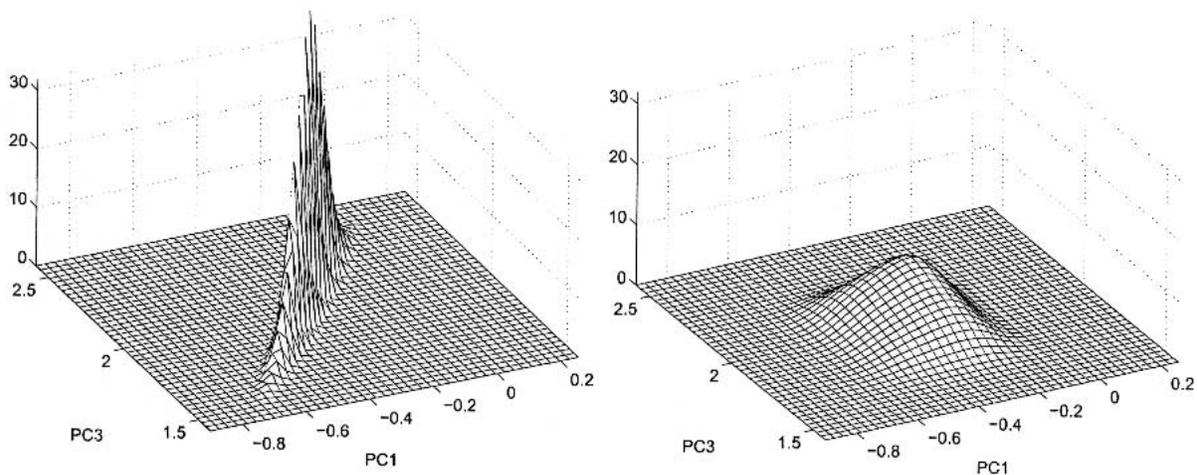


FIG. 22. Ensemble with high information flow $PC1 \leftrightarrow PC3$ for 500-hPa climate at time $t = 75$ (see Fig. 21), approximated by (left) $p_{13,4}$ and (right) $p_{1,4} \times p_{3,4}$.

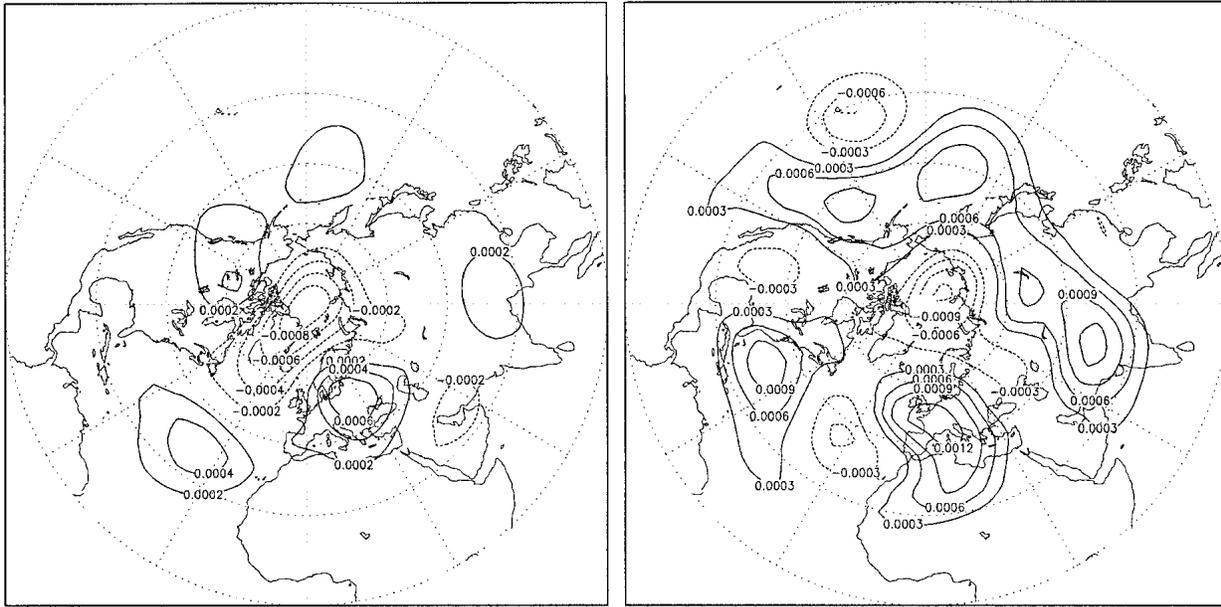


FIG. 23. Difference between realizations taken from opposite ends of 500-hPa ensemble (shown in Fig. 21), projected onto EOF1–EOF3 plane, at (left) time $t = 45$ and (right) $t = 75$.

different types of climate and slightly different qualitative properties for the different ensemble generation methods. For the 500-hPa climate the typical decay time of $P(p_4^*, \Pi_4^*)$ is 150 days, versus 50 days in case of the 300-hPa climate. Small-scale uncorrelated perturbations, originating from the physical space generation method, are rapidly damped by scale-selective diffusion, which causes the relative entropy for the physical space ensemble to decay faster at early times. The ensemble generation method does not, however, affect the global decay time of $P(p_4^*, \Pi_4^*)$. For any given combination of climate and ensemble generation, the behavior of principal components for large-scale EOFs is virtually identical. It is also worth mentioning that the decay rate of relative entropy in a PC is not controlled by its decorrelation time scale. Similar trend has also been encountered in the Lorenz '96 model (A–M). This is certainly a novel result, which should receive more attention in the authors' future works.

- The non-Gaussianity of the prediction ensemble as measured by $P(p_4^*, p_G)$ has different time scales for the different types of forcing, but exhibits no qualitative differences between different ensemble generation strategies. For both types of climate, there is a certain time (60 days for 500-hPa and 20 days for 300-hPa types of climate) at which all principal components for the large-scale EOFs simultaneously start growing. The ratio of time scales of non-Gaussian growth for different types of climate is proportional to the time scales for decay of climate lack of information. It is shown that for all large-scale EOFs and

both types of climate the non-Gaussian growth is initiated when the ensemble variability (standard deviation) reaches 10% of the climatological variability. This leads to the conclusion that non-Gaussian growth is initiated by nonlinear effects in the forecast ensemble. After the non-Gaussian growth is initiated, $P(p_4^*, p_G)$ for PC1 grows faster than for the rest of large-scale PCs.

- Dispersion controls both the mean over all ensembles and the variability of the lack of information in the climate relative to ensemble prediction for all large-scale EOFs and both types of climate. The signal, dispersion, and cross-term on average behave identically for all large-scale EOFs, but the decay time scale depends on the type of climate, while the ensemble generation method makes no difference in trends. Moreover, the dispersion itself is controlled entirely by its Gaussian part (terms of no more than second order).
- The generalized signal is important physically in individual ensemble predictions when it is large. For all regimes, there is a correlation between the generalized signal and the forecast ensemble mean state—when generalized signal in a given PC is high (compared to the other PCs), the ensemble mean state at that moment strongly projects on the corresponding EOF. In particular, AO, NAO, and PNA-like patterns in the EOFs are reproduced well by the ensemble mean in cases of high generalized signal. This result justifies the use of generalized signal as a stand-alone (and computationally cheap) measure of predictive utility as suggested earlier (A–M).

- The information flow between large-scale EOFs has common qualitative behavior for both types of climate. There are two stages in evolution of mutual information flow—the ascent stage and descent stage. The timing of the transition roughly coincides with the beginning of growth of non-Gaussianity in the ensemble, which means that information flow reaches its maximum when nonlinear effects first appear in the forecast ensemble. The combination of pairwise mutual information flows as an improvement to the direct sum of one-dimensional estimators for climate lack of information is not particularly efficient (average gain is 7%–8% for the first four EOFs) for the two barotropic climates studied here.
- High information flow between two PCs in an ensemble prediction is usually produced by a large cross-second moment in the forecast ensemble. Physically, projection of the forecast ensemble on these PCs becomes thin, long, and oriented at roughly equal angle to either of the PCs. Such a form of the forecast ensemble induces correlated changes in the predicted PCs along one of two transversal directions, depending on the sign of the ensemble slope. As discussed in the text, the two different scenarios of switching EOF patterns with high information flow stress changes of different low frequency physical structures in the flow.
- It is observed during the course of the paper, that the two ensemble breeding techniques are roughly equivalent for given initial variance of the forecast ensemble. However, if the ensemble was generated with the initial variance being 10% or more of climatological, its behavior could be sensitive to the breeding method, due to the rapid nonlinear anisotropic (because of high second cross-moment) stretching, which occurs at these values of ensemble variance.

In conclusion, the relative entropy framework turns out to be a versatile tool for providing measures and methods for predicting large-scale structures for low frequency variability in barotropic models. Following Anderson and Stern (1996), and Kleeman (2002), the use of the relative entropy as a measure of forecast utility compared with climatology has been utilized; also here the practical use of Generalized signal, ensemble non-Gaussianity, and mutual information flow is illustrated as stand-alone methods for predicting

physical features of the barotropic flow. Though it turns out that the predictability features for the low-frequency behavior of the T21 barotropic truncation are often controlled by Gaussian moments (dispersion by variance, information flow by cross-second moment), as pointed in the text, other recent studies (Abramov and Madja 2004; Cai et al. 2004) deal with nonlinear dynamics, where predictability strongly depends on non-Gaussian effects in a model such as Lorenz '96 with baroclinic effects. Also, in the barotropic models, the signal is only important for ensemble forecast utility when it is anomalously large, in contrast to the results from Kleeman et al. (2002) for the TBH model, and Kleeman and Majda (2004, manuscript submitted to *J. Atmos. Sci.*) for baroclinic turbulence, where the signal controlled variations in utility of large-scale patterns.

Acknowledgments. The authors thank Christian Franzke and Daan Crommelin for stimulating discussions, which were very helpful in the course of the work. Andrew Moore is thanked for useful discussions. Andrew Majda is partially supported by NSF Grant DMS-9972865, ONR Grant N00014-96-1-0043, and NSF CMG Grant DMS-0222133. Rafail Abramov is supported as a postdoctoral fellow through these research grants. Richard Kleeman is partially supported by NSF Grant ATM-0071342 and NSF CMG Grant DMS-0222133.

APPENDIX

The Inequality for Information Flow

For a probability density $p(\mathbf{x})$, a marginal probability density $p_{\mathbf{i}}(\mathbf{x}_{\mathbf{i}})$, $\mathbf{i} = i_1, \dots, i_{|\mathbf{i}|}$ is defined as

$$p_{\mathbf{i}}(\mathbf{x}_{\mathbf{i}}) = \int_{\mathbb{R}^{N-M}} p(\mathbf{x}) \prod_{k \neq i_j} dx_k, \quad (\text{A1})$$

where the integration is done over all dimensions except those with indices in \mathbf{i} . The following inequality holds for the marginal probability densities $p_{\mathbf{i}}$ of information flow (Majda et al. 2002)

$$P\left(p, \prod_{\mathbf{i}_j}^{N_{p_{\mathbf{i}}}} \times p_{\mathbf{i}_j}\right) \geq \sum_{j=0}^{(N_{p_{\mathbf{i}}}-3)/2} [P(p_{\mathbf{i}_{2j+1}, \mathbf{i}_{2j+3}}, p_{\mathbf{i}_{2j+1}} \times p_{\mathbf{i}_{2j+3}}) + P(p_{\mathbf{i}_{2j+2}, \mathbf{i}_{2j+3}}, p_{\mathbf{i}_{2j+2}} \times p_{\mathbf{i}_{2j+3}})], N_{p_{\mathbf{i}}} \text{ odd}, \quad (\text{A2})$$

$$P\left(p, \prod_{\mathbf{i}_j}^{N_{p_{\mathbf{i}}}} \times p_{\mathbf{i}_j}\right) \geq \sum_{j=0}^{(N_{p_{\mathbf{i}}}-4)/2} [P(p_{\mathbf{i}_{2j+1}, \mathbf{i}_{2j+3}}, p_{\mathbf{i}_{2j+1}} \times p_{\mathbf{i}_{2j+3}}) + P(p_{\mathbf{i}_{2j+2}, \mathbf{i}_{2j+3}}, p_{\mathbf{i}_{2j+2}} \times p_{\mathbf{i}_{2j+3}})] \\ + P(p_{\mathbf{i}_{N_{p_{\mathbf{i}}}-1}, N_{p_{\mathbf{i}}}}, p_{\mathbf{i}_{N_{p_{\mathbf{i}}}-1}} \times p_{\mathbf{i}_{N_{p_{\mathbf{i}}}}}), N_{p_{\mathbf{i}}} \text{ even}, \quad (\text{A3})$$

where N_{p_i} denotes the total number of marginal probability densities. The inequalities in (A2) and (A3) are not invariant under an arbitrary permutation of coordinate indices. In order to obtain the permutation-invariant version of (A2) and (A3), its right-hand side has to be averaged over all possible permutations of indices:

$$P\left(p, \prod_{\mathbf{j}}^{N_{p_i}} \times p_{\mathbf{i}}\right) \geq \frac{1}{N_{p_i}!} \sum_{\text{permutations}} \left[\text{right-hand side of (A2) or (A3)}\right]. \quad (\text{A4})$$

The summation over all permutations for an individual $P(p_{\mathbf{i}_j}, p_{\mathbf{i}} \times p_{\mathbf{j}})$ can be replaced by the nonrepeated summation, multiplied by the average number of equivalent outcomes per permutation (which is factorial of N_{p_i} divided by the binomial coefficient $C_2^{N_{p_i}}$). Since the number of terms in the right-hand side of either (A2) or (A3) equals $(N_{p_i} - 1)$, (A4) is rewritten as

$$P\left(p, \prod_{\mathbf{j}}^{N_{p_i}} \times p_{\mathbf{j}}\right) \geq \frac{N_{p_i}!(N_{p_i}-1)}{C_2^{N_{p_i}}N_{p_i}!} \sum_{j=1}^{N_{p_i}-1} \sum_{k=k+1}^{N_{p_i}} P(p_{\mathbf{i}_j\mathbf{i}_k}, p_{\mathbf{i}_j} \times p_{\mathbf{i}_k}). \quad (\text{A5})$$

Recalling that binomial coefficient is given by $C_k^n = n!/k!(n-k)!$, (A5) is written as

$$P\left(p, \prod_{\mathbf{i}}^{N_{p_i}} \times p_{\mathbf{i}}\right) \geq \frac{2}{N_{p_i}} \sum_{j=1}^{N_{p_i}-1} \sum_{k=k+1}^{N_{p_i}} P(p_{\mathbf{i}_j\mathbf{i}_k}, p_{\mathbf{i}_j} \times p_{\mathbf{i}_k}). \quad (\text{A6})$$

REFERENCES

Abramov, R., and A. Majda, 2004: Quantifying uncertainty for non-Gaussian ensembles in complex systems. *SIAM J. Sci. Stat. Comp.*, in press.

—, G. Kovačič, and A. Majda, 2003: Hamiltonian structure and statistically relevant conserved quantities for the truncated Burgers–Hopf equation. *Comm. Pure Appl. Math.*, **56**, 1–46.

Anderson, J., and W. Stern, 1996: Evaluating the potential predictive utility of ensemble forecasts. *J. Climate*, **9**, 260–269.

Branstator, G., and S. Haupt, 1998: An empirical model of barotropic atmospheric dynamics and its response to tropical forcing. *J. Climate*, **11**, 2645–2667.

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

Cai, D., K. Haven, and A. Majda, 2004: Quantifying predictability in a simple model with complex features, I: The perfect predictability scenario. *Stochast. Dyn.*, in press.

Carnevale, G., and G. Holloway, 1982: Information decay and the predictability of turbulent flows. *J. Fluid Mech.*, **116**, 115–121.

Cover, T., and J. Thomas, 1991: *Elements of Information Theory*. Wiley, 542 pp.

Crommelin, D., 2003: Regime transitions and heteroclinic connections in a barotropic atmosphere. *J. Atmos. Sci.*, **60**, 229–246.

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

—, and J. Tribbia, 1997: Optimal prediction of forecast error covariances through singular vectors. *J. Atmos. Sci.*, **54**, 286–313.

Franzke, C., A. Majda, and E. Vanden-Eijnden, 2005: Low-order stochastic mode reduction for a realistic barotropic model climate. *J. Atmos. Sci.*, in press.

Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 363 pp.

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

—, A. Majda, and I. Timofeyev, 2002: Quantifying predictability in a model with statistical features of the atmosphere. *Proc. Natl. Acad. Sci.*, **99**, 15 291–15 296.

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

—, and K. Emanuel, 1998: Optimal sites for supplementary weather observations. *J. Atmos. Sci.*, **55**, 399–414.

Majda, A., and I. Timofeyev, 2000: Remarkable statistical behavior for truncated Burgers–Hopf dynamics. *Proc. Natl. Acad. Sci.*, **97**, 12 413–12 416.

—, and —, 2002: Statistical mechanics for truncations of the Burgers–Hopf equation: A model for intrinsic stochastic behavior with scaling. *Milan J. Math.*, **70**, 39–96.

—, R. Kleeman, and D. Cai, 2002: A mathematical framework for quantifying predictability through relative entropy. *Meth. Appl. Anal.*, **9**, 425–444.

Mead, L., and N. Papanicolaou, 1984: Maximum entropy in the problem of moments. *J. Math. Phys.*, **25**, 2404–2417.

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

—, F. Molteni, R. Mureau, R. Buizza, P. Chapelet, and J. Tribbia, 1993: Ensemble prediction. *Proc. Vol. 1, of the ECMWF Seminar on the validation of Models over Europe*, Reading, United Kingdom, 21–66.

Reynolds, C. A., and T. N. Palmer, 1998: Decaying singular vectors and their impact on analysis and forecast correction. *J. Atmos. Sci.*, **55**, 3005–3032.

Roads, J., 1987: Predictability in the extended range. *J. Atmos. Sci.*, **44**, 3495–3527.

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

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

Selten, F., 1995: An efficient description of the dynamics of barotropic flow. *J. Atmos. Sci.*, **52**, 915–936.

Simmons, A., J. Wallace, and G. Branstator, 1983: Barotropic wave propagation and instability, and atmospheric teleconnection patterns. *J. Atmos. Sci.*, **40**, 1363–1392.

Toth, Z., 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.