The Impact of Dynamical Constraints on the Selection of Initial Conditions for Ensemble Predictions: Low-Order Perfect Model Results

Jeffrey L. Anderson

Geophysical Fluid Dynamics Laboratory, Princeton University, Princeton, New Jersey

ABSTRACT

A number of operational atmospheric prediction centers now produce ensemble forecasts of the atmosphere. Because of the high-dimensional phase spaces associated with operational forecast models, many centers use constraints derived from the dynamics of the forecast model to define a greatly reduced subspace from which ensemble initial conditions are chosen. For instance, the European Centre for Medium-Range Weather Forecasts uses singular vectors of the forecast model and the National Centers for Environmental Prediction use the ''breeding cycle'' to determine a limited set of directions in phase space that are sampled by the ensemble forecast.

The use of dynamical constraints on the selection of initial conditions for ensemble forecasts is examined in a perfect model study using a pair of three-variable dynamical systems and a prescribed observational error distribution. For these systems, one can establish that the direct use of dynamical constraints has no impact on the error of the ensemble mean forecast and a negative impact on forecasts of higher-moment quantities such as forecast spread. Simple examples are presented to show that this is not a result of the use of low-order dynamical systems but is instead related to the fundamental nature of the dynamics of these particular low-order systems themselves. Unless operational prediction models have fundamentally different dynamics, this study suggests that the use of dynamically constrained ensembles may not be justified. Further studies with more realistic prediction models are needed to evaluate this possibility.

1. Introduction

Because it is impossible to measure precisely the current state of the atmosphere and because numerical prediction models are only approximations of the true atmospheric dynamics, it is natural to view the task of predicting the future state of the atmosphere in a stochastic framework. In recent years, the world's operational atmospheric prediction centers have increasingly tended to view the forecast problem in this light. This has led to the development of ensemble prediction systems at most of the operational centers (Tracton and Kalnay 1993; Palmer et al. 1993; Molteni et al. 1996).

The enormous number of variables (degrees of freedom) in modern global atmospheric prediction models, typically on the order of 1 000 000, has naturally led to grave concerns about whether the small number of ensemble forecasts that can be produced with available computer resources can reasonably sample the probability distributions to be predicted. The general conclusion has been that it is impossible to sample adequately the full complexity of the prediction model's phase space (Gleeson 1970). Instead, a priori attempts are made to identify a vastly reduced set of directions in phase space that are the most important to sample when producing an ensemble forecast (Houtekamer 1995; Palmer et al. 1993); these directions are frequently determined using information from the dynamics of the forecast model (Mureau et al. 1993; Toth and Kalnay 1993).

One primary difference between the ensemble prediction schemes of the various operational centers is the way in which this subset of important directions is chosen. For instance, the National Centers for Environmental Prediction (NCEP) use directions determined from a set of leading ''breeding'' vectors (Toth and Kalnay 1996). The European Centre for Medium-Range Weather Forecasts (ECMWF) makes use of optimal perturbations that are obtained from the leading singular vectors of a linearized version of their forecast model (Buizza et al. 1993). Most other major centers are producing ensembles with variants of these two methods. Similar methods have also been used in nonoperational studies of the sensitivity of the evolution of the atmosphere in both global (Farrell 1990) and regional models (Vukicevic and Raeder 1995; Errico and Vukicevic 1992).

Much of the utility of ensemble (Monte Carlo) techniques is based on their interpretation as a random sam-

Corresponding author address: Dr. Jeffrey L. Anderson, GFDL, Princeton University, P.O. Box 308, Princeton, NJ 08540.
E-mail: jla@gfdl.gov

ple of a probability distribution (Epstein 1969; Leith 1974). When the search for ensemble members is limited by dynamical constraints, the ensemble is no longer a random sample, and great care must be taken in interpreting quantities derived from the ensemble. In the following, the value of using dynamical constraints in the selection of initial conditions for ensemble forecasts is examined in a perfect model study where the only source of error is assumed to be a prescribed observational error distribution (analysis error distribution in a real model). The study is mostly performed in two very low-order (three-variable) dynamical systems. The results are extrapolated to the vastly larger and more complex operational prediction models; however, it is fair to point out in advance that this extrapolation must be regarded with caution. Further studies with models of intermediate size and complexity are still needed to better understand the impact of dynamical constraints.

The utility of dynamically constraining ensembles has been addressed several times before (Houtekamer 1995; Brooks et al. 1995; Buizza and Palmer 1995). In the study most closely related to the present results, Houtekamer and Derome (1995) examined this problem in a global three-level quasigeostrophic model with moderate horizontal resolution. Their conclusions were that the use of dynamically constrained initial conditions had relatively little impact on ensemble predictions. The present study differs from theirs in several ways. First, as noted in section 2, this study is concerned with evaluating more than the ensemble mean forecast; the results for the ensemble mean are consistent with Houtekamer and Derome's conclusion that dynamical constraints have relatively little impact. Second, this study is able to examine a very large sample of different ensemble forecasts due to the simplicity of the dynamical systems used. This allows one to have considerably greater confidence in the results presented.

Section 2 discusses the selection of measures for evaluating ensemble predictions. Section 3 defines the types of dynamical constraints studied here and presents results for linear models. Section 4 describes corresponding nonlinear experiments whose results are presented in section 5. Section 6 presents a discussion of the results and their possible implications for large prediction models, while section 7 provides conclusions and some suggestions for further research.

2. Evaluating ensemble forecasts

Since all observations have an associated measurement error, the observed (analyzed) state of the atmosphere can be represented as a probability distribution. The forecast problem is then to predict how this probability distribution is evolved in time by a model. In the following, an ensemble of initial states is used to sample this observed probability distribution and is integrated to produce an estimate of the probability distribution at some forecast time (Leith 1974). This sec-

tion describes a set of measures that are used to evaluate the quality of ensemble forecasts of the probability distribution. The evaluation of the forecast of the complete probability distribution is one thing that distinguishes this study from some earlier evaluations of ensemble forecasts (Houtekamer and Derome 1995).

Careful definitions of terminology are essential to the following discussion. The term "member" refers to an individual forecast that is part of an ensemble for a given initial condition probability distribution. "Set" refers to a collection of k-member ensembles; a set is composed of "samples," each of which is a k-member ensemble.

a. Ensemble mean forecast

The root-mean-square distance (rms) in phase space between the mean of an ensemble of forecasts and the verifying truth is the principal tool for measuring forecast error here. The anomaly correlation (AC) is also frequently used to evaluate the error of atmospheric forecasts. Results are not shown for AC; however, the results of section 3 are unchanged if one substitutes AC for rms, and the results of sections 5 and 6 were found to be qualitatively unchanged. The conclusions of this study would remain unchanged if AC had been substituted for rms everywhere.

b. Consistency

If the verifying truth is indistinguishable from a randomly selected member of an ensemble forecast over a large set of forecast cases, the ensemble forecasts are said to be consistent with the truth. Consistency is evaluated here using a binning method (Anderson 1996b; Harrison et al. 1995; Molteni et al. 1996) that is also referred to as Talagrand diagrams. The binning method evaluates the consistency of ensemble predictions of a scalar, x, by using the individual ensemble forecasts of x to partition the real line. For a consistent ensemble the truth should fall into each bin with equal probability. Given a large set of ensemble forecast cases and the verifying truth for each, the distribution of the truth in the bins should be uniform. The standard chi-square test is applied to evaluate if the distribution of the truth in the bins is uniform; the significance of the chi-square test is presented here to allow comparison between different sets of ensemble forecasts. Small values of the chi-square significance mean that the distribution of the truth is unlikely to be uniform, which implies that the ensemble forecasts are not consistent with the truth. For a variable that is consistent, one expects the values of the significance to be approximately uniformly distributed over the interval 0–1 (i.e., a random sample from a uniform distribution will produce a chi-square significance of 0.1 or less 10% of the time and a value greater than 0.9 10% of the time). Consistency is related to other uses of ensemble forecasts, such as predicting forecast skill or producing "clusters" (Brankovic et al.

1990; Molteni and Tibaldi 1990; Cheng and Wallace 1993) of forecasts. Inconsistent forecasts will generally be inferior to consistent forecasts for such applications. It is important to remember that a consistent forecast is not necessarily a skillful forecast; for example, ensembles selected randomly from the model climate distribution will be consistent. Therefore, consistency should only be used as a measure of forecast quality in concert with other measures such as those discussed in the previous subsection.

c. Spread versus skill

The second moment (variance) and even higher moments of a forecast probability distribution can be predicted using ensemble forecasts. To date, the relation of a measure of the second moment (spread) to the ensemble mean forecast error has received the most attention (Kalnay and Dalcher 1987; Molteni and Palmer 1991; Barkmeijer 1993; Houtekamer 1993). Traditionally, the quality of the ensemble forecast's variance prediction has been judged by examining the correlation of the ensemble spread and the forecast error of the ensemble mean over a large set of forecasts (Barker 1991; Wobus and Kalnay 1995). Although this correlation is a poor measure of the quality of the second-moment forecast from an ensemble (Houtekamer 1993), examples of the correlation of the spread and the ensemble mean rms are discussed to facilitate comparison with the results of previous studies.

Another evaluation of the spread versus skill relation is performed by evaluating the consistency of the ensemble predictions of forecast error and the verifying forecast error between the ensemble mean and the truth. One cannot get a random sample of the distribution of ensemble mean rms error directly from an ensemble of forecasts. With an n-member ensemble, one might consider examining the distribution of rms differences between a given ensemble member and the mean of the remaining $n - 1$ members; this would produce n values. However, these rms differences are not independent; this is easily seen by considering the limiting case of a two-member ensemble in which this would produce two identical values.

In order to produce a random sample of the ensemble mean rms, one needs a sample of the mean of the distribution that is independent of the randomly chosen members. One can then compute the rms difference between each of the n ensemble members and this independent mean. These n predictions of the rms error can be used to partition the real line into $n + 1$ bins. For an ensemble that is consistent with the truth, the verifying rms error of the ensemble mean should fall into each of these bins with equal probability. The same chi-square method described above can be used to evaluate this. This also measures how well one can predict the expected value of the ensemble mean rms error from the "spread" of the ensemble.

All that is needed now is an independent sample of the mean of the probability distribution. As is discussed in sections 3 and 4, an independent mean is available in the ensemble forecasts made here, at least for short forecast lead times.

d. Inclusiveness

Another measure of the quality of an ensemble forecast examines whether extreme outliers of the probability distribution are appropriately sampled. The rms error of the member with the largest rms in an ensemble is used as a crude measure of how well the edges of the probability distribution are sampled by the ensemble.

One might also be interested in evaluating the worst forecast "bust" over a large set of ensemble forecasts. Here, the worst forecast bust is defined as that ensemble forecast for which the ensemble member with the minimum rms (the best forecast in the ensemble) is farthest from the verifying truth. Because this is an extremely unstable statistic, the notion of worst forecast bust is measured by averaging the rms error of the best ensemble member for each of the forecast cases that fall in the top 1% of worst forecast busts found in a large set. This quantity is related to how often the ensemble forecast completely fails to represent the true state.

It appears that there is no generally accepted measure of inclusiveness in this sense. In fact, there seem to be almost as many measures as there are operational prediction organizations. Therefore, this measure should be regarded as one possible measure that is appropriate to some subset of users of ensemble forecasts.

3. Constrained and unconstrained ensembles

A number of the world's operational prediction centers are producing constrained initial conditions for their ensemble predictions. Here, a "constrained" ensemble initial condition is defined as an ensemble in which only some subspace of the complete model phase space is sampled. Constrained ensemble forecasts include the use of the singular value decomposition (singular vectors) at ECMWF (Molteni et al. 1996) and the breeding cycle at NCEP (Toth and Kalnay 1996). In this section, constrained and corresponding unconstrained ensemble initial condition distributions are compared in a general perfect model framework.

Suppose one is given a perfect prediction model that operates on points in an n-dimensional real (one can easily generalize to complex) phase space (Gleeson 1970) along with a long model trajectory $T(t)$, where t is time, that corresponds to the "truth" of the system being predicted. Also assume that observed points can be generated for each true point by adding a random selection from a prescribed observational error distribution E.

Given an observed point $O(t)$ and the observational

error distribution, an ensemble of initial conditions that is consistent with the corresponding truth $T(t)$ can be generated by subtracting random samples of E from $O(t)$ (Leith 1974). In addition, all ensemble initial conditions used here are made symmetric around $O(t)$ by doubling the size of the ensemble. If U_i is a member of the ensemble, then the point $V_i = 2O(t) - U_i$ is also added as a member. This procedure guarantees that the ensemble mean of the initial conditions is $O(t)$. Similar procedures are used at the operational centers to minimize ensemble mean forecast error at early forecast leads (Molteni et al. 1996; Toth and Kalnay 1993). Toth and Kalnay (1996) found that having paired perturbations reduces error at longer lead times, too.

This set of ensemble initial conditions is referred to as an unconstrained ensemble initial condition distribution in what follows, and forecasts generated from these initial conditions are referred to as unconstrained ensemble forecasts. The complete unconstrained ensemble is used to compute the ensemble mean and for the measures of inclusiveness described in the previous section. Since only that half of the unconstrained ensemble that is randomly selected (the U_i) is consistent with the truth, only these ensemble members are used for forming bins to evaluate measures of consistency. If one uses both members of the pairs for forming bins, the truth is not consistent, even at the initial time.

Suppose that an m-dimensional subspace of the n-dimensional phase space is specified by m vectors representing directions around $O(t)$ that are believed to be of unusual interest. For instance, these vectors could be the m leading singular vectors (ECMWF) or breeding vectors (NCEP). The constrained ensemble initial condition samples the projection of the observational error distribution on this subspace. It can be generated by randomly selecting ensemble members in the same fashion as for the unconstrained ensemble initial conditions but then projecting onto the m-dimensional subspace. Again, the constrained ensemble is made symmetric around $O(t)$ by doubling the ensemble size. Figure 1 illustrates the unconstrained and constrained initial condition probability distributions that are sampled by the ensemble initial conditions. The constrained ensemble uses all the information about the observational error distribution that can be expressed in the constrained subspace. This is the only fair way to compare the constrained and unconstrained ensembles, since any other sampling technique in the constrained subspace would unfairly penalize the constrained ensembles.

By definition, the half of the constrained ensemble initial conditions that is randomly selected [i.e., not the part of the ensemble that is added to make it symmetric around $O(t)$] is consistent with the truth in the m-dimensional subspace, but it is inconsistent with the truth in the remaining $(n - m)$-dimensional subspace of the phase space (given that the observational error's projection onto these remaining directions is not a delta function, which seems a reasonable assumption). In oth-
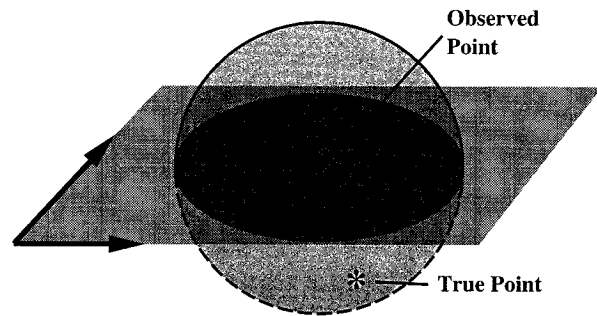

FIG. 1. An idealized depiction of the constrained and unconstrained probability distributions. The unconstrained distribution is represented by the spherical distribution centered around the observed point and enveloping the true point. The constrained distribution is the circular distribution centered on the observed point and on a plane defined by the two vectors that are dynamically constrained "important" directions in phase space.

er words, the comparison of the constrained and unconstrained methods is as fair as possible since the constrained methods know everything about the projection of the probability distribution onto the directions they sample.

Now, suppose that the forecast model is a linear operator, L. If L is applied to the constrained and unconstrained ensembles, the ensemble means are identical to the result of applying L to the observed point $O(t)$ (Leith 1974; Vukicevic 1991). The unconstrained ensemble forecast is a random sample of the truth (consistent) in this perfect model study. The constrained ensemble continues to be consistent in the m-dimensional subspace that results from operating on the original constrained subspace with L, but it also continues to be inconsistent in the remaining $n - m$ directions. The fact that it is inconsistent in some directions implies that the spread and skill relation is also inconsistent.

Spread–skill consistency can also be evaluated in this linear model context. The observed point $O(t)$ is by construction the mean of both the probability distribution being sampled and the full ensemble, so it can be used for the computation of the expected rms error of the ensemble mean (see section 2). In the linear model, $L[O(t)]$ remains the mean of the distribution, so it can be used to compute a forecast distribution of ensemble mean rms error at all lead times. A consistent sample of the expected ensemble mean error can be produced by computing the error between $L[O(t)]$ and the elements of the randomly chosen half of the unconstrained ensemble. Once again, the failure of the constrained distribution to sample some directions in phase space means that the constrained distribution does not demonstrate this spread–skill consistency.

This section demonstrates that in a linear system there may not be a reason to use a constrained ensemble, given the measures being used here. The ensemble mean errors are identical and the unconstrained ensemble is consistent (both for the distributions and for spread versus

skill), while the constrained ensemble is only consistent over some subspace. The evaluation of inclusiveness is difficult enough that it is addressed only empirically through the nonlinear experiments of the next section. Even if the constrained subspace were known a priori to contain most of the interesting dynamics, there might still be no reason to use a constrained ensemble. These same results also apply to the early times of a forecast when the behavior of a nonlinear forecast model is quasi-linear (Errico et al. 1993; Ehrendorfer and Tribbia 1997).

4. Design of nonlinear experiments

This section describes simple, nonlinear perfect model experiments that evaluate some of the impacts of nonlinear models on the results of the previous section. Emphasis is given to determining if nonlinear effects could make the use of constrained ensemble initial conditions superior to the use of unconstrained ensembles. The experiments are described in terms of a general nonlinear dynamical system and then applied to several systems (see the appendix) in the next section.

First, a large number of true points on the attractor of the dynamical system are generated by doing an initial very long spinup integration, followed by a continued integration in which the model state is periodically sampled. In the real world, true states like this can never be known because of inevitable observational errors.

Next, an observed state corresponding to each true point is generated by adding a random sample from some observational error distribution to the true state vector. For the results shown here, the observational error distribution is a random independent selection from a normal distribution with a given standard deviation for each of the components of the true state vector [independent identically distributed (iid) normal]. A number of other types of observational error distributions were also examined. These included independent selections from uniform distributions and the use of normal distributions for error distance in which the individual error components are not independent. Use of these different observational error distributions had no qualitative impact on the results.

As in the discussion of the previous section, an unconstrained ensemble distribution is generated for each observed state by subtracting a random sample of the observational error distribution from the observed state for each ensemble member. Again, a second ensemble member that is symmetric around the observed point is included for each randomly selected member. The randomly selected half of the unconstrained ensemble is then consistent with the observational error distribution by definition.

A constrained ensemble distribution is also generated for each observed state as in the previous section. The constrained ensembles studied here are restricted to a two-dimensional subspace of the model phase space.

Given two orthonormal vectors that span this subspace, the observational error distribution can be projected on the subspace, resulting in a two-dimensional probability density distribution. The constrained ensemble is composed of random samples of this reduced dimension probability density distribution. This projection of the observational error distribution onto the constrained subspace allows the constrained ensemble to use all information about the observational error distribution that can possibly be expressed in the subspace. As noted in section 3, this is the only fair way to compare constrained versus unconstrained ensembles, since any other method of sampling in the constrained space would unfairly penalize the constrained methods.

A variety of algorithms for selecting the constrained subspace have been examined. The two primary examples, perturbed integrations and singular vectors were discussed in detail for the Lorenz-63 system (see appendix) in Anderson (1996a).

The perturbed integration method (PI hereafter), is similar to the breeding cycle used at NCEP. To find the m most important directions, m randomly selected orthogonal vectors of very small amplitude are added to the true point at the start of the long spinup integration. Each of these perturbed states is integrated along with the true point. After each time step, the vectors representing the difference between the perturbed states and the true state are orthonormalized (to the same small amplitude), and the resulting orthogonal vectors are added to the true state to form a new set of perturbations. The perturbation vectors should converge to the local Lyapunov vectors of the dynamical system after a long integration (Buizza and Palmer 1995). The PI method differs from the operational breeding cycle in several ways. First, the breeding cycle normalizes to finite amplitude perturbations while the PI method normalizes to a very small amplitude. Second, the breeding cycle perturbs around the analysis state (which introduces discontinuities in the trajectory followed during the computation of the bred vectors), rather than the truth. Perturbed integrations around the observed points instead of the truth were also investigated; the results were qualitatively indistinguishable from those for the PI method and are not discussed further. It is possible that the application of the breeding cycle within the context of an analysis system could significantly change the behavior of the bred vectors; this possibility needs to be investigated further in more realistic models. Operationally produced breeding vectors have also been used for other applications, such as reducing analysis errors or selecting optimal locations for targeted observations. The present study only attempts to assess the use of the PI method for producing ensemble forecasts in a somewhat unrealistic environment where the initial errors are unrelated to the basic flow.

The use of the singular vector decomposition (SVD) is the other primary method for selecting constrained directions. The SVD method used here is essentially

identical to those used operationally (Molteni et al. 1996) and in the study of Houtekamer and Derome (1995). A description of the use of this method in one of the low-order models discussed hereafter can be found in Anderson (1996a). A number of different optimization times have been examined for the SVDs; the impact of the optimization time on the results is discussed in the following sections.

The evaluation of spread–skill consistency in the fully nonlinear model is somewhat problematic. At the initial time, the observed point is the mean of the probability distribution and the full ensemble, and can be used to compute samples of the ensemble mean rms error (section 2) as in the linear problem of section 3. At later times, the ensemble mean of the full ensemble can be used as a proxy for the mean of the forecast distribution. A random sample of the expected rms error of the ensemble mean can be generated from the rms distance between each member of the randomly selected half of the ensemble and the mean of the full ensemble. However, in the nonlinear model, there is no guarantee that the full ensemble mean remains consistent with the forecast distribution's mean (Leith 1974) as forecast lead extends past the linear range. This can lead to a loss of consistency between the ensemble prediction of ensemble mean rms error and the verifying rms error of the ensemble mean at long forecast leads. A consistent ensemble forecast will begin with a consistent relation, but this consistency may gradually disappear as the forecast extends into the nonlinear regime.

5. Nonlinear model results

Comparisons between constrained and unconstrained ensembles are presented for a pair of three variable dynamical systems in this section. Both models can be viewed as low-order analogs of the large-scale atmospheric dynamics (Lorenz and Krishnamurthy 1987; Lorenz 1984). The first model has an unusually straightforward attractor that can facilitate understanding the differences between the different ensembles. The attractor of the second system is considerably more complicated and may provide a better analog to the behavior found in real forecast models.

All results shown in this section are averages over 10 000 sample sets of ensemble forecasts. This large set is sufficient to produce extremely stable statistics for most of the quantities evaluated; splitting the set into two 5000 sample halves had little impact on the results.

Each experiment was performed for 2, 9, and 99 member pair ensembles; results are generally only presented for two member pair ensembles (i.e., two pairs of integrations). This choice seems most compatible with the very small ensemble sizes (especially relative to the phase space size) that can be produced operationally. However, it is important to recall that the ensemble forecasts here are simply being used to sample the probability distributions of various scalar quantities.

The size of the phase space and its complexity have no direct impact on the ability of an ensemble to sample such quantities (Cramer 1966; see his chapter 14). Where results for ensembles with more than two member pairs are presented (see Fig. 9 and related discussion), they should *not* be used to compare constrained to unconstrained ensemble performance. For a fair comparison of larger ensemble sizes, one would have to increase the dimension of the constrained subspace, which was not done for the 9 and 99 member pair ensembles here.

a. Lorenz-63

The first model examined is the popular Lorenz-63 system (see appendix; Lorenz 1963), which has been used in a number of ensemble forecast investigations (Palmer 1993; Houtekamer and Derome 1994). All results shown are for an observational error distribution with standard deviation of 1.0, which is approximately 2% of the total attractor range of this model. Forecast integrations extend to 1.0 nondimensional time units, which is somewhat longer than the time for an average attractor trajectory to orbit one of the two lobes of the attractor. As shown below, it is also a long enough integration that the error of individual ensemble members is beginning to saturate; later stages of the forecasts are clearly in the nonlinear regime. The SVD results shown are for an optimization time of 1.0 time units. Optimization times ranging from a single time step (0.01 time units) through 2.0 time units were also tested to assess the impact of the optimization time on the SVD results.

Figure 2 displays the rms error of the ensemble mean as a function of forecast lead time, along with the mean error of the individual ensemble members and the error of the ensemble members with the maximum and minimum rms as a function of lead time for the unconstrained ensembles and the constrained SVD ensemble. The curves for the rms of the ensemble mean are extremely similar demonstrating that the constrained ensemble does not offer an advantage for this quantity. The minimum rms curves are also very similar, but the curve for the mean error of the individual members and the maximum error curve show the SVD ensemble having slightly lower rms. The brief initial decrease in the rms error quantities is a reflection of the collapse of off-attractor quantities onto the attractor.

All four rms error curves are nearly identical after about 0.05 time units for the unconstrained and PI constrained ensembles (not shown). The rms errors of the 100 worst forecast busts as defined in section 2 are also indistinguishable for both the PI and SVD constrained ensembles and the unconstrained ensemble.

Although the rms of the ensemble mean is the same for the constrained and unconstrained ensembles, there are significant differences in the consistency results. Figure 3 shows the consistency for the x and z variables
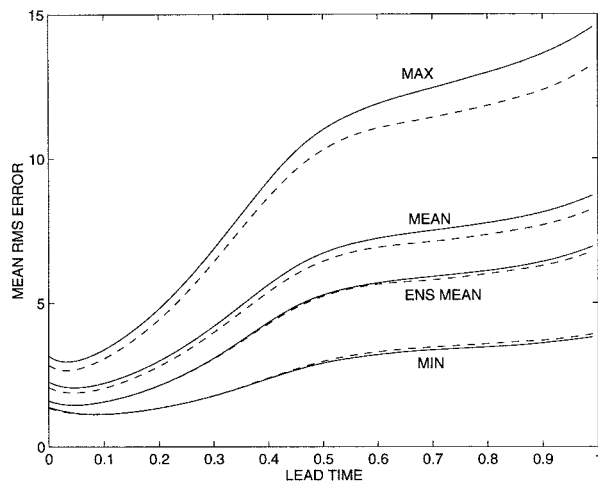
FIG. 2. Average rms error as a function of lead time for unconstrained (solid) and SVD constrained (dashed) two-pair ensembles for the L63 model; observational error standard deviation is 1.0. The upper and lower curves are for the maximum and minimum rms, respectively; the second set of curves from the bottom is the rms of the ensemble mean; and the final set is the mean rms for the ensemble. All results are averages over 10 000 samples. The minimum and ensemble mean curves are nearly identical for the unconstrained and constrained.
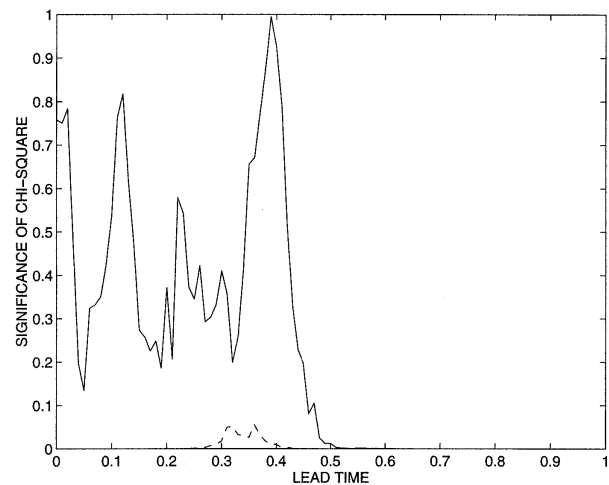
FIG. 4. Consistency of ensemble "spread" and rms error of ensemble mean measured by the significance of the chi-square test as a function of lead time for the same L63 experiments as in Fig. 2. Results are shown for the unconstrained (solid), PI constrained (dashed, only briefly nonzero), and SVD constrained (indistinguishable from zero) ensembles.

of the Lorenz-63 system as a function of forecast lead time (results for the y variable are qualitatively similar to those for x). For all three variables, the SVD constrained ensemble is highly inconsistent at all lead times with significance values generally less than 1×10^{-10}. The unconstrained ensemble is consistent by definition at the initial time of the forecast. For the x and y variables, the unconstrained ensemble also appears to be consistent with the truth at all leads while the constrained PI ensemble is inconsistent until about time 0.1

and consistent thereafter. For the z variable, both the PI and unconstrained ensembles are consistent initially but rapidly become inconsistent as forecast lead increases past time 0.1; at later times both ensembles regain consistency. A more detailed analysis and some alternative measures need to be developed to better understand this behavior.

Figure 4 shows the consistency of the ensemble spread and skill as a function of lead. Again, the SVD ensemble is highly inconsistent at all lead times. For this quantity, however, the PI ensemble is also highly inconsistent, while the unconstrained ensemble is con-
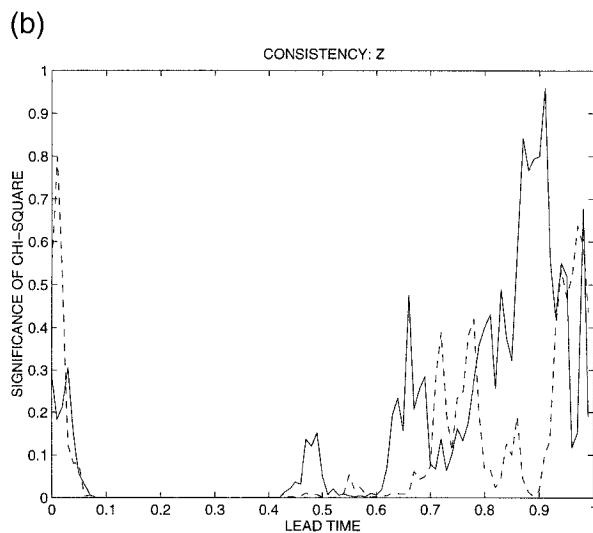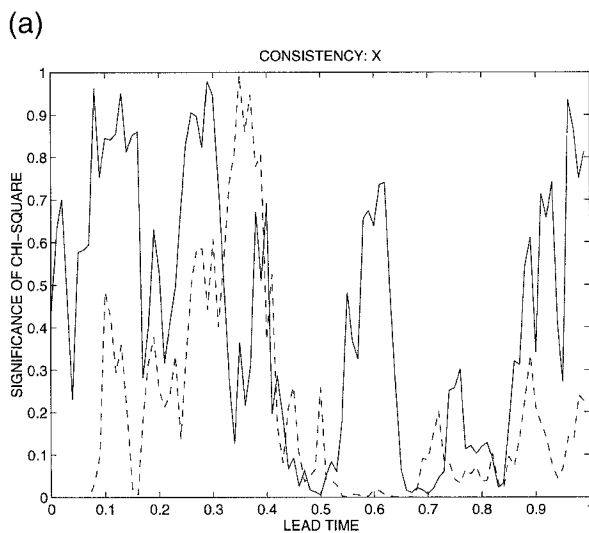
(a)

(b)

FIG. 3. Consistency of ensemble predictions and truth measured by the significance of the chi-square test as a function of lead time for the x and z variables of the L63 model for the same experiment as in Fig. 2. Results are for the unconstrained ensemble (solid), the PI constrained ensemble (dashed), and the SVD ensemble, which is indistinguishable from zero in both panels.

sistent until lead time 0.5. The conclusion is that the unconstrained ensemble can be used to make reliable predictions of forecast skill out to time 0.5 but that the two constrained ensembles can not.

Much of the behavior of these three ensembles in the Lorenz-63 model can be explained since both the equilibrium and nonequilibrium dynamics of this system are relatively straightforward. The large-scale structure of the attractor consists of nearly flat sheets that are attached at one end and trajectories rotate around these sheets, sometimes switching from one to another. The nonequilibrium dynamics in the vicinity of the attractor is also simple; points close to the attractor are rapidly pulled in toward the nearest attractor sheet. If one can capture the details of that part of the observational error distribution that is parallel to the local attractor sheet, one can predict most of the evolution of the probability distribution. This is reflected by the small differences between the unconstrained and PI constrained distributions in this model. As noted in Anderson (1996a), the two leading vectors computed by the PI method are very nearly parallel to the local attractor sheet over the vast majority of the Lorenz-63 attractor. Within about 0.1 time units, both the two-dimensional PI ensemble and the three-dimensional unconstrained ensemble are pulled onto the nearest attractor sheet (in the vicinity of the region where the two sheets are joined somewhat more complicated behavior is possible; see section 6 and Fig. 11). For the x and y variables for which the unconstrained distribution remains consistent, the PI distribution also becomes consistent as the two distributions collapse to the same plane.

Although the primary off-attractor dynamics is a collapse to the nearby attractor sheet, there is also a shear in phase space in directions along the attractor. The consistency of spread and skill is affected by this shear as the PI and unconstrained distributions collapse to the attractor. The PI distribution may not sample points that contain the true point (because it only samples in directions parallel to the local attractor), while the unconstrained does sample the true point in all cases (see additional discussion in section 6 and Fig. 10). For this reason, the unconstrained ensemble is able to maintain spread skill consistency for a much longer period of time.

The SVD constrained ensemble performs more poorly than the PI in this case (in terms of consistency with the truth) because the two leading singular vectors are on average not nearly as close to parallel to the local attractor sheet (Anderson 1996a). The SVD ensemble therefore tends to fail to sample the important part of the probability distribution parallel to the local attractor and ends up being inconsistent with the truth. It is interesting to note that the singular vectors are not such poor predictors of the local attractor structure that the rms of the ensemble mean is degraded.
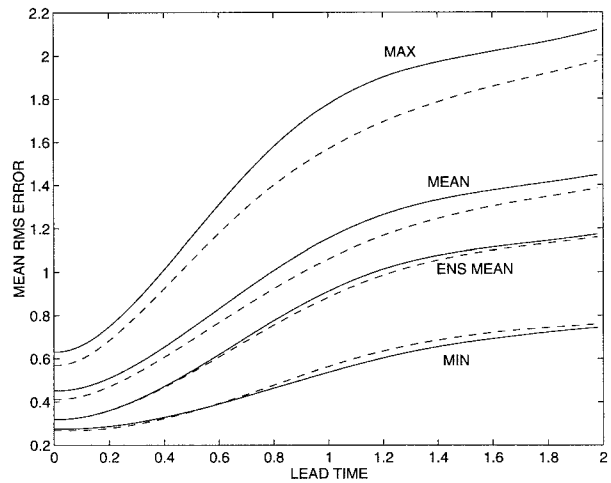

FIG. 5. Average rms error as a function of lead time for unconstrained (solid) and SVD constrained (dashed) two-pair ensembles for the L84 model; observational error standard deviation is 0.2. The upper and lower curves are for the maximum and minimum rms, respectively; the second set of curves from the bottom is the rms of the ensemble mean; and the final set is the mean rms for the ensemble. All results are averages over 10 000 samples. The minimum and ensemble mean curves are nearly identical for the unconstrained and constrained.

b. Lorenz-84

The Lorenz-63 model's extremely simple attractor is not a very rigorous test of ensemble initial condition distributions. In particular, because the PI method is so successful in sampling the relevant phase space directions, this model can cast undue aspersions on the SVD method. The Lorenz-84 model (Lorenz 1984; see appendix) is another three-variable model that has a more complicated attractor structure. For this model, the attractor dimension is greater than 2 and not nearly integral (Leonardo 1995) so that neither of the constrained direction methods can give an accurate depiction of the local attractor structure in all cases. Ehrendorfer and Tribbia (1997) examine the use of the SVD method for predicting the evolution of the second moment of forecast probability distributions in this model.

All results shown are for an observational error distribution with standard deviation of 0.2, which is approximately 2% of the total attractor range of this model. Forecast integrations extend to 2.0 nondimensional time units, which is somewhat longer than the time for an average attractor trajectory to go once around the central portion of the attractor. As shown below, it is also a long enough integration that the error of individual ensemble members is beginning to saturate. The SVD optimization time for the results shown is 1.0 time units.

Figure 5 shows the rms error curves for the unconstrained and SVD ensembles; in this case, the curves for the PI ensembles are indistinguishable from those for the SVD. In both cases, the rms error of the ensemble mean forecast is essentially identical for unconstrained and constrained ensembles at all forecast leads. Again,
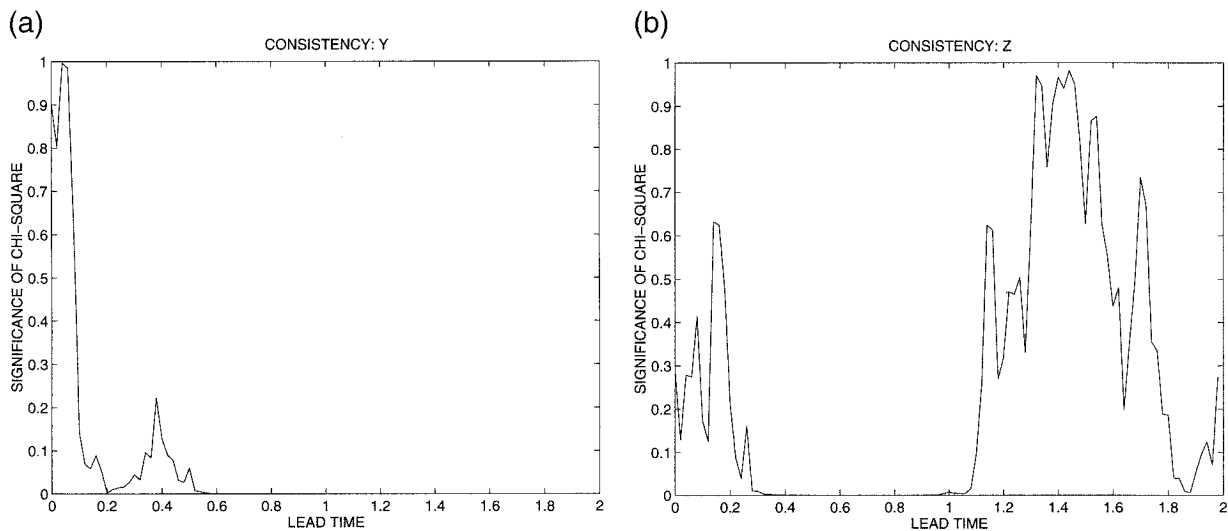
(a)

CONSISTENCY: Y

(b)

CONSISTENCY: Z

FIG. 6. Consistency of ensemble predictions and truth as measured by the significance of the chi-square test as a function of lead time for the y and z variables of the L84 model in the same experiment as in Fig. 5. Results are shown for the unconstrained ensemble and the PI and SVD constrained ensembles, which are both indistinguishable from zero throughout both panels.

the mean error of individual ensemble members is larger in the unconstrained ensemble, demonstrating that this ensemble is making a more efficient sampling of the phase space.

Figure 6 shows the consistency for the y and z variables. For all of x, y, and z, both constrained ensembles are highly inconsistent for all lead times with values of significance rarely exceeding 1×10^{-10}. For the y variable, the unconstrained ensemble is somewhat consistent out to about time 0.5; its consistency for x (not shown) is similar but disappears even earlier. For z (Fig. 6), the unconstrained ensemble becomes inconsistent for

lead times from 0.3 to about 1.0 and regains consistency thereafter.

Figure 7 shows the spread–skill consistency for the L84 model. Again, both constrained distributions are highly inconsistent at all leads while the unconstrained ensemble retains consistency to about time 0.5. One could also look at the more traditional spread–skill correlation as a function of lead time, which is shown in Fig. 8. All three ensembles have a low spread–skill correlation at early lead times because the initial condition probability distribution is fixed and therefore uncorrelated with the initial error by definition (Barker 1991; Wobus and Kalnay 1995). This demonstrates just one
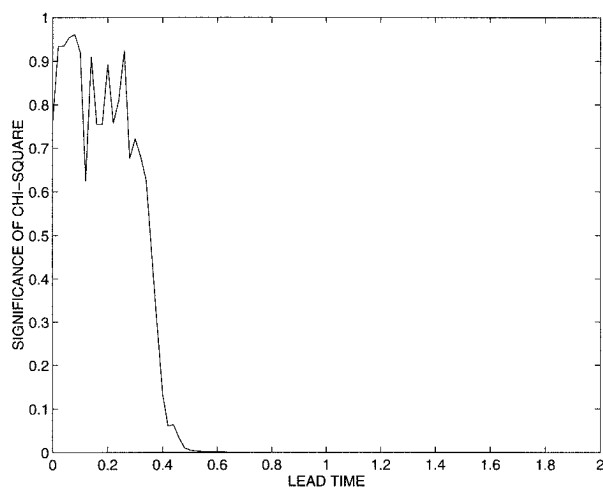
FIG. 7. Consistency of ensemble spread and ensemble mean rms error as measured by the significance of the chi-square test as a function of lead time for the same L84 experiments as in Fig. 5. Results are shown for the unconstrained and for the PI and SVD constrained ensembles, which are both indistinguishable from zero throughout.
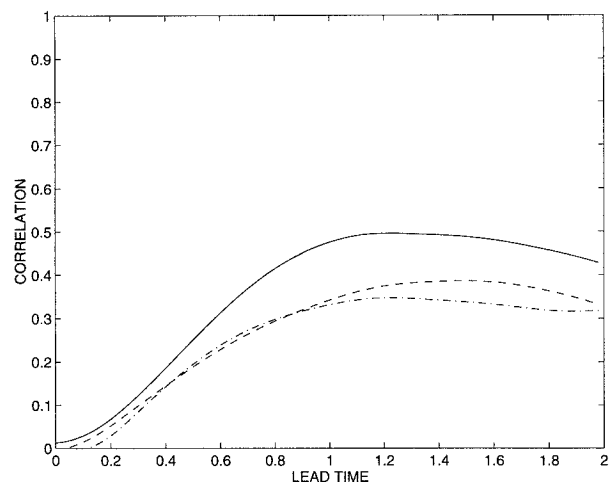
FIG. 8. Correlation between ensemble spread and ensemble mean rms error as a function of lead time for the same experiment as depicted in Fig. 5. Results are shown for the unconstrained ensemble (solid) and the PI (dashed) and SVD (dash-dotted) constrained ensembles.
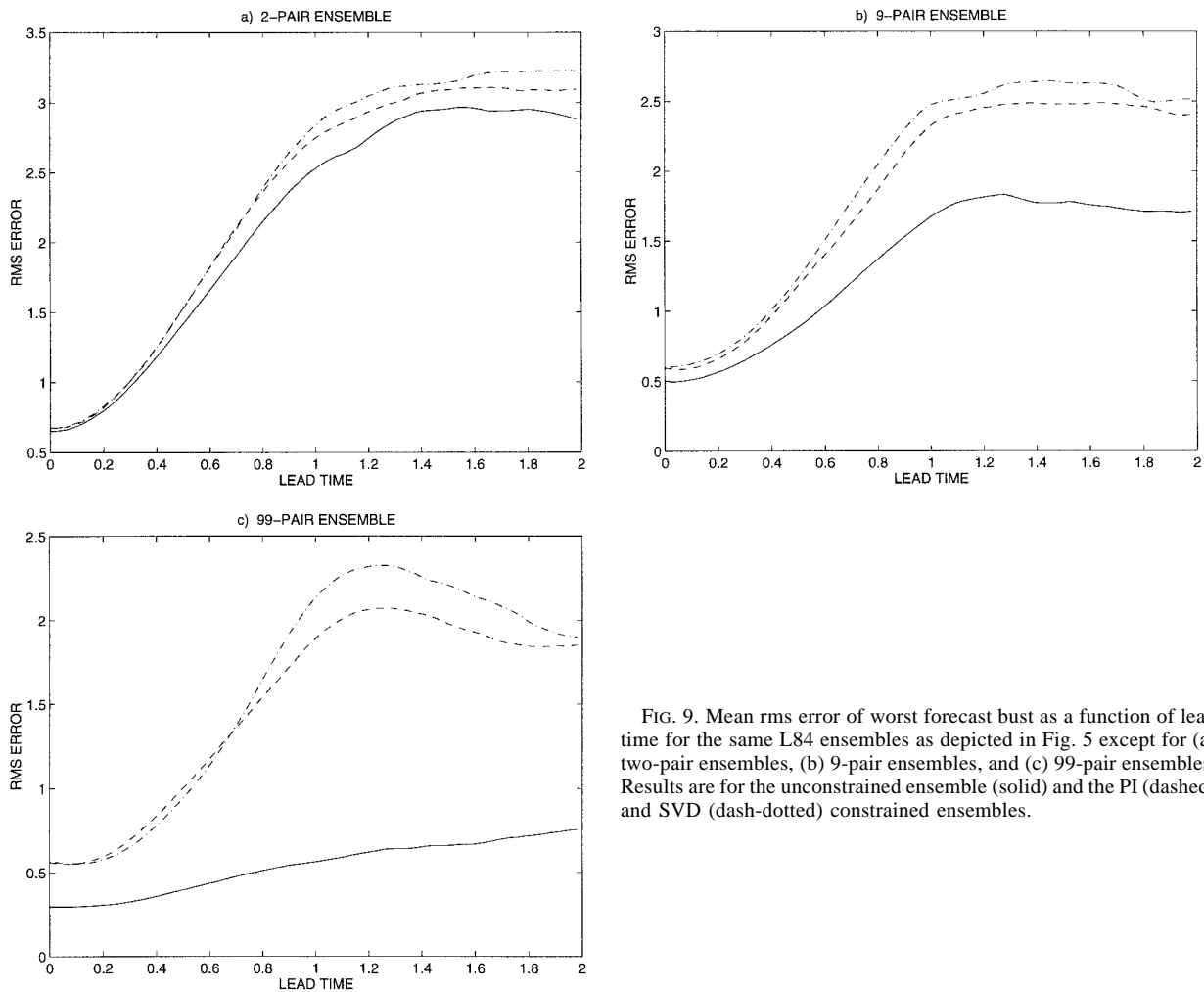
FIG. 9. Mean rms error of worst forecast bust as a function of lead time for the same L84 ensembles as depicted in Fig. 5 except for (a) two-pair ensembles, (b) 9-pair ensembles, and (c) 99-pair ensembles. Results are for the unconstrained ensemble (solid) and the PI (dashed) and SVD (dash-dotted) constrained ensembles.

of the many problems of using spread–skill correlation. However, at later times, the unconstrained distribution has a significantly higher correlation between spread and skill. This enhanced correlation extends well beyond time 0.5 at which the very sensitive consistency test shows the spread and skill to become inconsistent. This loss of consistency may be because the full ensemble mean is no longer consistent with the mean of the true distribution at these lead times (section 4). The conclusion is that the unconstrained distribution is more useful for making predictions of forecast skill.

Figure 9 shows the average rms of the worst forecast busts as defined in section 2 for a variety of ensemble sizes. For the two member pair ensembles shown in Fig. 9a, the unconstrained ensemble has rms slightly smaller than for the two constrained ensembles. Since this is likely to be an extremely unstable statistic, this result was verified with much larger sets (50 000 samples) and found to be stable. In this model, the constrained ensembles are more likely to have forecasts that entirely miss the truth than is the unconstrained ensemble.

To date, only two member pair ensembles have been discussed since this seems to be the most similar to the small ensembles used by operational centers. In the experiments with larger ensembles discussed next, the constrained ensembles are unfairly restricted to two-dimensional subspaces, so these results should not be used to compare the abilities of constrained and unconstrained ensembles. It is of interest, however, to use considerably larger ensembles to better understand the behavior of the complete probability distributions and to make sure that the small ensemble size is not contributing qualitatively to the results presented. For this reason, Fig. 9 also contains plots of the worst forecast busts for the Lorenz-84 models for 9 and 99 member pair ensembles. As the ensemble size increases, the ratio between the worst forecast busts for the constrained versus the unconstrained ensembles get progressively larger. For all other fields discussed previously, there is no qualitative difference between the larger ensemble size results and the two member pair ensemble results. This experiment is not meant to suggest that operational centers would

choose to search a number of directions in phase space that is less than the size of the ensemble.

The SVD method has an additional free parameter, the optimization time interval for the computation of the SVDs. As noted above, the results shown here are for an optimization time of 1.0 time units in both models, but SVD constrained ensembles were also evaluated for optimization times ranging from 0.01 through 2.0 time units. In operational forecasting the length of the optimization time has significant impacts on the structure of the SVDs and on the forecast results. In general, operational considerations dictate that the optimization time should be bounded above by the time interval up to which the evolution of the finite amplitude ensemble perturbations are well approximated by linear theory (Buizza 1995).

In the results here, the length of the SVD optimization had negligible impact on the results although the singular vectors themselves are sensitive to this parameter (Anderson 1996a). The results described in this section occur not because the SVD method gives a poor set of constrained directions, but because sampling in any constrained two-dimensional subspace leads to similar results.

6. Interpretation and discussion

Dynamically constrained ensemble initial conditions are a response to the staggering size of the phase spaces associated with modern atmospheric prediction methods. Since one cannot hope to sample each of these directions independently, one can argue that the most appropriate use of resources is to sample only in a subset of directions that are believed a priori to be of most importance. In one argument, those directions that are believed to support the most rapid error growth have been judged to be the most important (Palmer et al. 1993). In other cases, those directions onto which the analysis error is believed to project most heavily are selected as most important (Toth and Kalnay 1996).

The results discussed in the previous sections suggest that there are instances in which an a priori selection of important directions for constrained initial conditions may not be a useful approach. For the sake of argument, suppose that algorithms exist to determine precisely all the important (constrained) directions for an atmospheric prediction model using whichever definition of importance is deemed appropriate. If the remaining directions (referred to in this section as unconstrained directions) are in fact irrelevant to the evolution of the forecast model, then there is no apparent harm in sampling in these directions. The projection of the ensemble initial conditions onto the unconstrained directions will evolve with no significant impact on the important constrained directions by definition. This was demonstrated for linear dynamics in section 3.

An additional simple demonstration can be made by adding a large number of additional ''irrelevant'' de-

grees of freedom to the Lorenz-84 model and investigating the behavior of constrained and unconstrained ensembles. A 1000-variable model composed of the Lorenz-84 three-variable system and 997 additional variables whose dynamics consists of a uniform exponential decay to zero is examined (see appendix). The experimental design is the same as in section 5 with the observational error distribution being iid normal for each of the 1000 variables. Because all of the additional variables are decaying, both the PI and SVD methods of selecting a pair of constrained directions have negligible projections onto any of the auxiliary directions in the phase space. The two leading vectors for each method are essentially identical to the Lorenz-84 in the remaining three phase-space dimensions. Results were examined for a reduced set of 1000 samples and for ensemble sizes of 2 and 9 pairs. The results for this 1000-variable auxiliary model are qualitatively similar to those for the Lorenz-84 model for the x, y, and z variable consistency; the spread–skill consistency; and for the correlation of spread and skill. For the additional 997 variables, the unconstrained ensemble is consistent by definition at all leads while the constrained ensembles are both inconsistent. The rms error of the ensemble mean forecasts from the constrained and unconstrained distributions are indistinguishable at all forecast leads. There are some differences in the maximum and minimum ensemble member rms and in the mean rms of the ensemble members. Initially, the unconstrained ensemble has higher values for these quantities than does the constrained ensemble. As the auxiliary variables decay and the three standard variables evolve, these three rms error quantities from the unconstrained ensemble collapse toward the constrained values, and by time 1.0, the behavior is similar to that for the original L84 cases (even if the decay time for the auxiliary variables is infinite). This same behavior is reflected in the worst forecast bust for the two ensemble pair cases. For early lead times the unconstrained ensemble has slightly larger mean rms for the worst forecast bust, but by time 1.0, the unconstrained has significantly smaller errors than the constrained cases.

This test is a particularly rigorous one for the unconstrained ensembles. There are fewer than three important directions in this model by any of the standard definitions and the constrained ensembles are given high quality information about two of these directions. The unconstrained ensemble has to deal with nearly 1000 additional unimportant directions. For the measures of forecast quality and the dynamical systems investigated here, even a huge number of unimportant directions does not impact the quality of the unconstrained forecast.

To date, it has been assumed that the unconstrained directions are entirely irrelevant to the evolution of the constrained directions. If this is not the case, it becomes increasingly difficult to justify the use of constrained ensembles. Two simple idealized examples, which reflect behavior that occurs in both the L63 and L84 models,
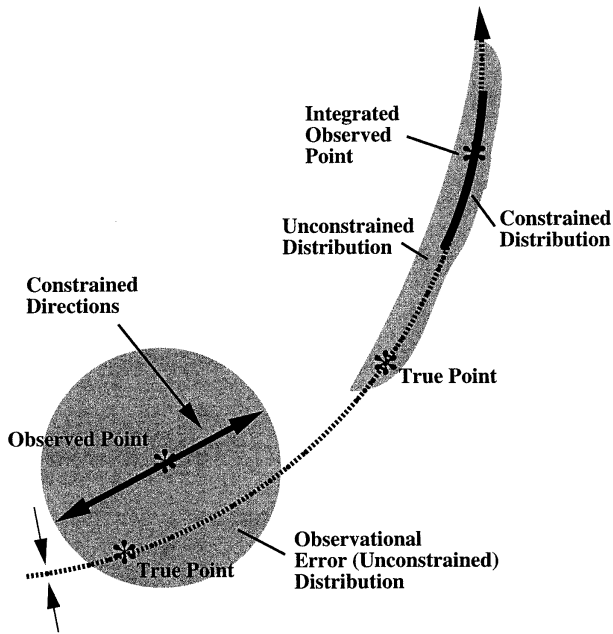
FIG. 10. Evolution of idealized constrained and unconstrained initial condition distributions in an attractor with shear. The attractor is represented by the dotted line. The initial true point and observed point are marked by the asterisks in the lower left portion of the figure. The observed point is surrounded by the circular unconstrained initial distribution; the constrained directions are depicted by the dark line with two arrows through the observed point. The true point and the corresponding constrained and unconstrained distributions at a forecast time are depicted in the upper right portion of the figure.
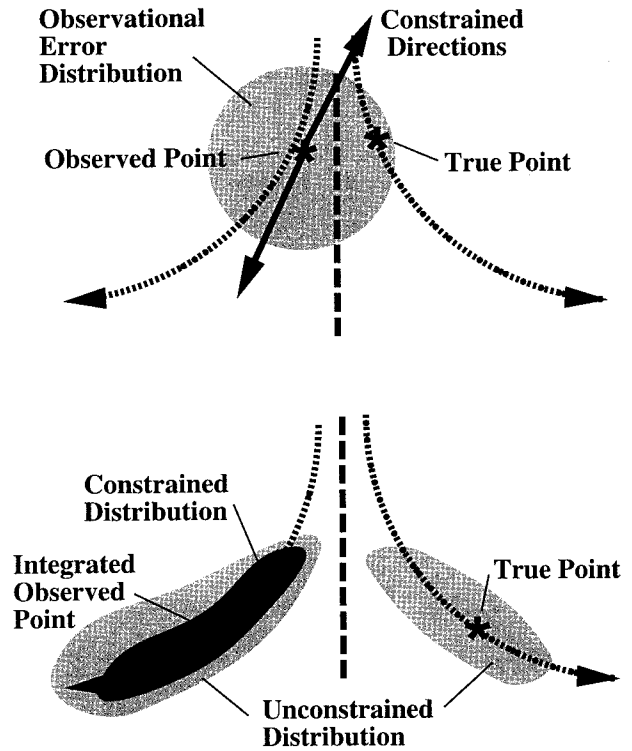

FIG. 11. Evolution of idealized constrained and unconstrained initial condition distributions in an attractor with multiple "branches." The upper panel shows the initial conditions as in Fig. 10; the heavy dashed line separates points that are attracted to the right branch from those attracted to the left. The lower panel shows the resulting distributions at a forecast time.

are used to demonstrate the potential pitfalls of the constrained ensemble approach. In both examples, the constrained directions are assumed to be perfectly computed to give information about local important directions (the attractor), while the unconstrained directions are still assumed to decay quickly onto the attractor. It is important to keep in mind that these idealized examples are individual cases, while all behavior discussed to date has been the mean over a large set of ensemble forecasts.

In the first example, depicted in Fig. 10, the dotted line represents an idealized attractor and the shaded circle an observational error distribution (unconstrained initial condition distribution) centered around an observed point. It is assumed that all points off the attractor quickly collapse onto the attractor. However, in this example it is also assumed that there is a shear perpendicular to the attractor so that points above the attractor move more quickly in a direction parallel to the attractor than points below the attractor. The constrained and unconstrained distributions that result after some time are shown in the upper part of Fig. 10. That portion of the probability mass that was above the attractor has been swept ahead of the true point during its collapse toward the attractor, while the smaller portion that was originally below the attractor has lagged behind the true point. Since the constrained initial condition distribution is entirely above the attractor, the constrained distri-

bution is completely ahead of the true point. A portion of the unconstrained distribution does remain around the true point. The result in this case is that the constrained distribution is clearly inconsistent with the truth, while the unconstrained distribution, although spread over a larger area, does contain the truth.

An even more drastic example in which failing to sample "unimportant" directions can be problematic is depicted in Fig. 11. The attractor is locally represented by the two dotted curves that could have resulted from a recent bifurcation (qualitatively similar behavior occurs in both L63 and L84 models) or could represent two pieces of the attractor that simply happen to approach each other for a time (similar behavior occurs in L84 model). Points to the right of the dotted line are attracted to the right portion of the attractor; points on the left of the line are attracted to the left. In the example, the true point is on one branch of the attractor, but the observed point is close to the other. The initial observational error distribution (unconstrained ensemble) and the constrained directions are shown at the top of the figure. Much of the unconstrained distribution and all of the constrained distribution move along the left branch of the attractor as time passes. However, as the lower part of Fig. 11 shows, the true point and some

portion of the unconstrained distribution lie on the right side of the attractor. In this case, the constrained distribution is highly inconsistent with the truth. Cases similar to this example were found to be the cause of many of the largest forecast bust cases in both the L63 and L84 model integrations. The behavior depicted in Fig. 11 might happen frequently in real forecast models if statistically significant clusters (Brankovic et al. 1990) are occurring in operational forecasts; it remains unclear at this time if such behavior is relevant for atmospheric prediction.

The examples above demonstrate that unconstrained ensembles can have a number of advantages over constrained ensembles. If consistency is the goal, there appears to be no justification for the use of constrained ensembles. For the simple models examined here, there is also no difference between the rms error of the ensemble mean forecasts. However, it is possible to construct dynamical systems for which the rms error of the ensemble mean is better or worse for the constrained than for the unconstrained ensembles. One can also construct different measures for the ensemble mean skill that give different expected values for the constrained and unconstrained ensembles. The important point is that it is the details of the dynamics and the error measure that determine the relative ensemble mean errors of the constrained and unconstrained ensembles, and that the size of the phase space and the attractor is not of direct relevance. For the simple systems discussed here, the dynamics do not support the use of constrained ensembles to improve the ensemble mean forecast. If this is not the case for realistic forecast models, it should be rigorously demonstrated as an argument to support the use of constrained ensembles.

Understanding why the errors of the ensemble mean forecasts are so nearly identical in the models used here is also of interest. It seems likely that this is related to the fact that the models examined have only quadratic nonlinearities. For linear forecast models, one can demonstrate a number of results concerning the errors of constrained and unconstrained ensembles (Ehrendorfer and Tribbia 1997). They show that an optimal sample of the forecast error covariance can be obtained by sampling the leading singular vectors in the linear regime. However, the results here extend well into the nonlinear regime; this can be verified by noting that the errors of the discrete forecast produced from the single observed point are considerably larger than the errors of the ensemble means by the end of the forecasts. Understanding why the constrained and unconstrained errors are so similar might allow an a priori statement about the expected results in larger models.

The results of Ehrendorfer and Tribbia (1997) also point out the importance of the measure used for evaluating ensemble forecasts. They demonstrate that a set of perturbations along the leading singular vectors is optimal to sample the forecast error covariance. This same set of perturbations is not optimal under the measures suggested here for evaluating ensemble forecasts as demonstrated in previous sections.

In order to evaluate unconstrained forecasts in operational models, one must have some means for providing a random sample of the analysis error. One could explicitly use information about the error covariances (and higher moments) to attempt to construct explicit descriptions of the error distribution. While information about error covariances is part of analysis systems, this information is considered to be highly uncertain and it would be difficult and expensive to use.

Instead, it might be possible to use the power of Monte Carlo methods to produce random samples of the analysis error distribution. To do this, one could perform a number of different assimilations of data following the example of Houtekamer and Derome (1995) (also Houtekamer et al. 1996), each starting from a randomly selected initial condition and each using independent samples of observations at each step during the assimilation.

To do this properly would in fact require completely independent sets of observations (i.e., a number of completely redundant global observing systems). Since such redundancy is clearly out of the question, one must attempt to produce random samples of the observational error distribution associated with the global observing system at each observing time. One method for attempting this would involve adding on random samples from the observational error distribution of each individual observation (or related group of observations) (Evensen 1994). Houtekamer and Derome did this for upper-air observations, noting that there is a covariance in the errors for measurements from a single radiosonde (Lonnberg and Hollingsworth 1986). The advantage to this type of method is that the error distributions of individual observations appear to be somewhat better known than the entire global analysis error distribution.

So far this section has argued that constrained ensemble forecasts may not be useful. However, there are cases in which constraining ensemble forecasts is clearly useful. The first case is when one knows the details of the local attractor structure and its position. As demonstrated in Anderson (1996a), in such cases it is valuable to use only ensemble initial conditions that lie on the attractor. While such an approach would be nice, it seems unlikely that algorithms for finding the position of the local attractor will ever be practical in models as large and complicated as operational forecast models.

A related case where constrained ensembles are potentially useful is when an initialization is applied to assimilated data before using them as initial conditions for a forecast. If one selects an unconstrained ensemble of initial conditions and then applies initialization to all members, the result is typically an ensemble with far too little variance (Hollingsworth 1980). Instead, it may be more useful to apply an initialization and then only search for ensemble members in directions that satisfy the balance represented by the initialization. Again, it

may be very difficult to design algorithms for finding such constrained directions in large forecast models.

There may be instances in which it is computationally more economical to compute a constrained ensemble than an unconstrained one. For instance, the breeding cycle at NCEP is extremely inexpensive as a means for generating constrained initial conditions (Toth and Kalnay 1993). If algorithms for attempting to find an unconstrained distribution prove to be expensive, this would argue against their use.

Finally, as already noted, there may be dynamical systems or error measures for which constrained ensembles are superior to unconstrained. More research is needed to determine what aspects of a particular norm or dynamical system are important for this determination.

7. Conclusions and future work

For the low-order models discussed here, the previous sections have demonstrated the relative merits of constrained and unconstrained ensembles. In general, for the measures proposed in section 2, unconstrained ensembles perform at least as well as, and in many cases much better than, constrained ensembles. These results suggest that operational ensemble prediction systems that make use of constrained initial conditions may be suboptimal.

There is a long history of using such simple model results as proxies for results that are too expensive to obtain in large atmospheric prediction models (Lorenz 1963). Nevertheless, there is no guarantee that low-order results can be generalized to much larger and more complicated systems. To further strengthen the results, tests of constrained versus unconstrained ensembles need to be made in higher-order models. Houtekamer and Derome (1995) have already taken a significant step in this direction, although they only evaluated first-moment quantities. Examining higher moments of the ensemble predictions in their quasigeostrophic forecast system would be of interest.

The relation between constrained and unconstrained ensembles also needs to be examined in operational prediction models. Because of the great expense involved in running these models, only a small set of relatively small ensembles would be possible. The results here suggest that the use of small (two pair) ensembles does not present a problem in comparing the two types of ensembles. Using such tiny ensembles would in turn allow a greater number of cases to be evaluated to increase the significance of the results. It would also be of interest to examine particular cases, especially those in which the two types of ensembles produced very different results, in order to see if this is due to some of the types of behavior described in section 6.

Problems remain with this method. In particular, without having sets of independent observations, it requires special procedures to guarantee that the Monte Carlo assimilated ensemble initial conditions are statistically consistent with the truth. Further work on Monte Carlo assimilations in low-order forecast system simulations like that of Houtekamer are needed to resolve these problems.

Problems also remain in assimilation cycles that require explicit initialization, for instance, to remove gravity wave noise. As pointed out above, initialization tends to reduce the variance of an ensemble. Again, special efforts are required to ensure that ensemble initial conditions retain sufficient variance in the presence of an initialization. Fortunately, it appears that the most recently developed operational assimilation systems no longer require an explicit initialization (Parrish and Derber 1992).

The results presented here also have implications for nonoperational problems such as evaluating the sensitivity of a particular set of initial conditions on the resulting evolution of the atmosphere. For instance, using SVDs to determine the directions in which the evolution would be most sensitive is only relevant in a linear regime and fails to give information about the likelihood of such extreme perturbations. Just as in the prediction problem, one must carefully assess what question one desires to answer with an ensemble and then decide whether constrained or unconstrained initial condition distributions are the most appropriate for addressing the question at hand.

APPENDIX

Dynamical Systems

a. Lorenz-63

The Lorenz-63 model (Lorenz 1963) is defined by the three differential equations

$$\dot{x} = -\sigma x + \sigma y, \tag{A1}$$

$$\dot{y} = -xz + rx - y, \tag{A2}$$

$$\dot{z} = xy - bz, \tag{A3}$$

where the dot represents a derivative with respect to time. The parameters are set to $\sigma = 10$, $r = 28$, and $b = 8/3$, a regime with known chaotic behavior. The two step self-starting time differencing scheme of Lorenz (1963) is used with a nondimensional time step of 0.01.

b. Lorenz-84

The Lorenz-84 model (Lorenz 1984) is defined by the three differential equations

$$\dot{x} = -y^2 - z^2 - ax + aF, \qquad \text{(A4)}$$

$$\dot{y} = xy - bxz - y + G, \qquad \text{(A5)}$$

$$\dot{z} = bxy + -xz - z. \qquad \text{(A6)}$$

The parameters are set to $a = 0.25$, $b = 4$, $F = 8$, and $G = 1.25$. The time differencing scheme is the same as for the Lorenz-63 model.

c. Lorenz-84 with auxiliary variables

The m-variable auxiliary model has three variables governed by the Lorenz-84 equations (A4)–(A6) and $m-3$ additional variables that are governed by an exponential decay to zero with a prescribed e-folding time R:

$$\dot{w}_i = -(1/R)w_i, \qquad i = 4, \ldots, m. \qquad \text{(A7)}$$

REFERENCES

Anderson, J. L., 1996a: Selection of initial conditions for ensemble forecasts in a simple perfect model framework. *J. Atmos. Sci.,* **53,** 22–36.

——, 1996b: A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Climate,* **9,** 1518–1530.

Barker, T. W., 1991: The relationship between spread and forecast error in extended range forecasts. *J. Climate,* **4,** 733–742.

Barkmeijer, J., 1993: Local skill prediction for the ECMWF model using adjoint techniques. *Mon. Wea. Rev.,* **121,** 1262–1268.

Brankovic, C., T. N. Palmer, F. Molteni, S. Tibaldi, and U. Cubasch, 1990: Extended-range predictions with ECMWF models: Time-lagged ensemble forecasting. *Quart. J. Roy. Meteor. Soc.,* **116,** 867–912.

Brooks, H. E., M. S. Tracton, D. J. Stensrud, G. DiMego, and Z. Toth, 1995: Short-range ensemble forecasting: Report from a workshop, 25–27 July 1994. *Bull. Amer. Meteor. Soc.,* **76,** 1617–1624.

Buizza, R., 1995: Optimal perturbation time evolution and senstivity of ensemble prediction to perturbation amplitude. *Quart. J. Roy. Meteor. Soc.,* **121,** 1705–1738.

——, and T. N. Palmer, 1995: The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.,* **52,** 1434–1456.

——, J. Tribbia, F. Molteni, and T. Palmer, 1993: Computation of optimal unstable structures for a numerical weather prediction model. *Tellus,* **45A,** 388–407.

Cheng, X., and J. M. Wallace, 1993: Cluster analysis of the Northern Hemisphere wintertime 500-hPa height field: Spatial patterns. *J. Atmos. Sci.,* **50,** 2674–2696.

Cramer, H., 1966: *Mathematical Methods of Statistics.* Princeton University Press, 575 pp.

Ehrendorfer, M., and J. J. Tribbia, 1997: Optimal prediction of forecast error covariances through singularvectors. *J. Atmos. Sci.,* **54,** 286–313.

Epstein, E. S., 1969: Stochastic dynamic prediction. *Tellus,* **21,** 739–759.

Errico, R. M., and T. Vukicevic, 1992: Sensitivity analysis using an adjoint of the PSU–NCAR mesoscale model. *Mon. Wea. Rev.,* **120,** 1644–1660.

——, ——, and K. Raeder, 1993: Examination of the accuracy of a tangent linear model. *Tellus,* **45A,** 462–477.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.,* **99,** 10 143–10 162.

Farrell, B., 1990: Small error dynamics and the predictability of atmospheric flows. *J. Atmos. Sci.,* **47,** 2191–2199.

Gleeson, T. A., 1970: Statistical-dynamical prediction. *J. Appl. Meteor.,* **9,** 333–344.

Harrison, M. S. J., D. S. Richardson, K. Robertson, and A. Woodcock, 1995: Medium-range ensembles using both the ECMSF T63 and unified models—An initial report. UKMO Tech. Rep. 153, 25 pp. [Available from U.K. Meteorological Office, London Road, Bracknell, Berkshire RGIZ 2SZ, U.K.]

Hollingsworth, A., 1980: An experiment in Monte Carlo forecasting procedure. *Proc. ECMWF Workshop on Stochastic Dynamic Forecasting,* Reading, United Kingdom, ECMWF, 65–85.

Houtekamer, P. L., 1993: Global and local skill forecasts. *Mon. Wea. Rev.,* **121,** 1834–1846.

——, 1995: The construction of optimal perturbations. *Mon. Wea. Rev.,* **123,** 2888–2898.

——, and J. Derome, 1994: Prediction experiments with two-member ensembles. *Mon. Wea. Rev.,* **122,** 2179–2191.

——, and ——, 1995: Methods for ensemble prediction. *Mon. Wea. Rev.,* **123,** 2181–2196.

——, L. Lefaivre, J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.,* **124,** 1225–1242.

Kalnay, E., and A. Dalcher, 1987: Forecasting forecast skill. *Mon. Wea. Rev.,* **115,** 349–356.

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

Leonardo, A., 1995: Numerical studies on the Lorenz-84 atmospheric model. Department of Mathematics Tech. Rep., 49 pp. [Available from Dept. of Mathematics, Utrecht University, Box 80010, Utrecht 3508 TA, the Netherlands.]

Lonnberg, P., and A. Hollingsworth, 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part II: The covariance of height and wind errors. *Tellus,* **38A,** 137–161.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.,* **20,** 130–141.

——, 1984: Irregularity: A fundamental property of the atmosphere. *Tellus,* **36A,** 98–110.

——, and V. Krishnamurthy, 1987: On the nonexistence of a slow manifold. *J. Atmos. Sci.,* **44,** 2940–2950.

Molteni, F., and S. Tibaldi, 1990: Regimes in the wintertime circulation over the northern extratropics. II: Consequences for dynamical predictability. *Quart. J. Roy. Meteor. Soc.,* **116,** 1263–1288.

——, and T. N. Palmer, 1991: A real-time scheme for the prediction of forecast skill. *Mon. Wea. Rev.,* **119,** 1088–1097.

——, R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy Meteor. Soc.,* **122,** 73–120.

Mureau, R., F. Molteni, and T. N. Palmer, 1993: Ensemble prediction using dynamically-conditioned perturbations. *Quart. J. Roy. Meteor. Soc.,* **119,** 299–323.

Palmer, T. N., 1993: Extended-range atmospheric prediction and the Lorenz model. *Bull. Amer. Meteor. Soc.,* **74,** 49–66.

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

Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.,* **120,** 1747–1763.

Tracton, S., and E. Kalnay, 1993: Operational ensemble forecasting at NMC: Practical aspects. *Wea. Forecasting,* **8,** 379–398.

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

——, and ——, 1996: Ensemble forecasting at NMC and the breeding method. NMC Office Note 407, 58 pp. [Available from NOAA/NWS/NCEP/EMC, 5200 Auth Road, Camp Springs, MD 20746.]

Vukicevic, T., 1991: Nonlinear and linear evolution of initial forecast errors. *Mon. Wea. Rev.,* **119,** 1602–1611.

——, and K. Raeder, 1995: Use of an adjoint model for finding triggers for alpine lee cyclogenesis. *Mon. Wea. Rev.,* **123,** 800–816.

Wobus, R. L., and E. Kalnay, 1995: Three years of operational prediction of forecast skill at NMC. *Mon. Wea. Rev.,* **123,** 2132–2148.