Selection of Initial Conditions for Ensemble Forecasts in a Simple Perfect Model Framework

JEFFREY L. ANDERSON

Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey

(Manuscript received 6 September 1994, in final form 11 January 1995)

ABSTRACT

An extremely simple chaotic model, the three-variable Lorenz convective model, is used in a perfect model setting to study the selection of initial conditions for ensemble forecasts. Observations with a known distribution of error are sampled from the "climate" of the simple model. Initial condition distributions that use only information about the observation and the observational error distribution (i.e., traditional Monte Carlo methods) are shown to differ from the correct initial condition distributions, which make use of additional information about the local structure of the model's attractor. Three relatively inexpensive algorithms for finding the local attractor structure in a simple model are examined; these make use of singular vectors, normal modes, and perturbed integrations. All of these are related to heuristic algorithms that have been applied to select ensemble members in operational forecast models. The method of perturbed integrations, which is somewhat similar to the "breeding" method used at the National Meteorological Center, is shown to be the most effective in this context. Validating the extension of such methods to realistic models is expected to be extremely difficult; however, it seems reasonable that utilizing all available information about the attractor structure of real forecast models when selecting ensemble initial conditions could improve the success of operational ensemble forecasts.

1. Introduction

For many years, the vast majority of operational forecasts of the atmosphere have used a single integration of a numerical weather prediction model started from a discrete observed state. However, because only sparse, inaccurate observations of the state of the atmosphere are available, the initial state of a prediction model is more appropriately represented as a probability distribution. A numerical forecast of the state of the atmosphere then entails estimating the probability distribution of the model variables at some ensuing time.

One possible approach to this problem of stochastic dynamic prediction was discussed by Epstein (1969). In this case the forecast model is based on stochastic differential equations, and the probability distribution of the initial and ensuing states are explicitly represented in the model. In general, this method involves such great expense that it is not practical for even the simplest of forecast models.

A second approach to stochastic-dynamic prediction, ensemble prediction, is now beginning to be used at operational prediction centers around the world. In ensemble prediction a number of discrete initial conditions are sampled from the observed probability distribution for the forecast model variables. Each of these initial conditions is then integrated as a separate forecast in a traditional "discrete" model. The distribution of the model variables in this ensemble of forecasts can then be viewed as a sample of the continuous probability distribution at some later time. Ensemble prediction was pioneered by a number of researchers including Gleeson (1970) and Leith (1974).

A number of problems remain in using ensembles of discrete forecasts to approximate the evolution of an initial probability distribution in a model. One fundamental problem is the interpretation of the ensemble forecast, especially when the ensemble size is very small compared to the number of dynamical degrees of freedom in the forecast model (Tracton and Kalnay 1993; Murphy 1990). A second problem is the selection of the initial ensemble members. In other words, how should the initial observed probability distribution be sampled in order to provide the most meaningful information from the ensemble forecast (Schubert et al. 1992).

Numerous approaches to selecting the initial ensemble members have been proposed. The oldest is Monte Carlo forecasting (Epstein 1969; Leith 1974). In Monte Carlo forecasts, the initial ensemble members are randomly selected from the best available approximation of the initial observed probability distribution. It can be demonstrated that this approach is generally a good way to sample a probability distribution (Lewis 1975) and that one must know a great deal about the

Corresponding author address: Dr. Jeffrey L. Anderson, Geophysical Fluid Dynamics Laboratory, Princeton University, P.O. Box 308, Princeton, NJ 08542.

probability distribution being sampled in order to sample more efficiently (Traub and Wozniakowski 1994). Despite this, because of the high costs involved in doing ensemble integrations, atmospheric scientists have attempted to devise more efficient ways of sampling than Monte Carlo.

One alternative to Monte Carlo sampling is lagged average forecasting (LAF) (Hoffman and Kalnay 1983). In LAF the initial ensemble sample for a time n is composed of a single discrete initial condition using observations at time n, plus some set of forecasts valid at time n that were produced using earlier observations. For instance, a three-member LAF ensemble might consist of a discrete "observed" initial condition, the 24-h forecast from discrete initial conditions 24 h earlier, and the 48-h forecast from discrete conditions 48 h earlier. The LAF sampling strategy is driven purely by pragmatism and does not explicitly attempt to make a good sample of the initial probability distribution. Despite this, there is some suggestion that LAF does a surprisingly good job at gleaning information about the initial probability distribution (Toth and Kalnay 1993). More sophisticated variants of LAF (Ebisuzaki and Kalnay 1991) have also been developed.

More recently, researchers have attempted to select a sample of the initial probability distribution using information about the numerical model (Molteni and Palmer 1993). In most of these methods the goal is to sample elements from the initial probability distribution that lead to the largest trajectory divergence in phase space. The objective of this type of ensemble forecasting is not necessarily to produce a good sample of the forecast probability distribution but, instead, to heavily sample the wings of the forecast probability distribution. During the initial linear phase of the evolution of the probability distribution, the wings of the distribution are in fact the least probable portion of the distribution. However, as noted by Mureau et al. (1993), it may be useful to have ensemble members that have as large a divergence as possible when attempting to use ensemble spread as a predictor of forecast skill. With this in mind, a number of techniques using various optimally growing perturbations derived from the forecast model have been used to generate ensemble initial conditions (Buizza et al. 1993; Buizza and Palmer 1994).

Despite enormous improvements over their predecessors, the current generation of forecast models are far from being perfect models of the real atmosphere. It seems highly unlikely that the model climate (the exact definition of climate in a model context will be discussed below) is consistent with the observed climate. An extended integration of these models probably never gets particularly "close" to any real observed condition in model phase space. Inconsistencies between the observed atmospheric state and the model's climate can be one cause of a period of transient evolution (spinup) in forecast model integrations. Although considering the implications of this spinup is vital in applying ensemble techniques to real forecast models, it introduces a level of complexity that is too great for this study. Therefore, the work presented here will be in a perfect model context (Leith 1974; Seidman 1981). The numerical model used is assumed to exactly represent the real world to be forecast. Observations of this world can be generated by taking snapshots during some long integration of the model and, if desired, adding some noise to these perfect observations. Such observations are guaranteed to be consistent with the long-term climate of the model, and the only source for spinup is the added observational noise (and possibly the numerical details of the model's time differencing).

Modern operational forecast models are extremely complex and require enormous amounts of computer time to integrate. The huge number of variables in these models can make forecast experiments difficult to analyze. This study will restrict attention to a maximally simplified model, which was selected to have the fewest possible number of variables while still producing "chaotic" behavior. Simple model studies like this have a long history in the atmospheric literature (Lorenz 1982). Experiments with simple models can give insight into the fundamental behavior of dynamical systems that is unattainable from more complicated models. However, great care must be exercised if techniques developed for simple models are ever to be applied to their vastly more complicated operational cousins.

This paper will examine the selection of initial conditions for ensemble forecasts in a simple, perfect model framework. Section 2 presents the numerical model and a quick discussion of the model's climate. Section 3 examines the selection of ensemble initial conditions for this model and, in particular, the impact of using information about the model's climate. Section 4 examines methods for finding the local structure of the model's climate and is followed by additional ensemble forecasting experiments in section 5. Section 6 provides a brief discussion of the implications of this work for more complicated models.

2. Model equations and attractor

The simple model used in this study is the three-variable convective model, whose derivation and physical interpretation can be found in Lorenz (1963). The differential equations for this model are

$$\dot{x} = -\sigma x + \sigma y \tag{1}$$

$$\dot{y} = -xz + rx - y \tag{2}$$

$$\dot{z} = xy - bz, \tag{3}$$

where the dot represents a derivative with respect to time. Throughout this study the parameters are set to

$\sigma = 10$, $r = 28$, and $b = 8/3$. This parameter range is the same as that used in Lorenz and is chosen because it demonstrates chaotic behavior. The two-step time differencing method described by Lorenz (1963) is used to integrate the system (1–3) with a nondimensional time step of 0.01. This time-stepping scheme is self-starting thus avoiding any of the *numerical* spinup problems mentioned in the introduction.

The structure of the attractor of the model has been documented in a host of publications (Lorenz 1987; Guckenheimer and Holmes 1983; Sparrow 1982). The attractor represents the equilibrated behavior of the model after it has been integrated a very long time from arbitrary initial conditions and is also referred to as the model climate. A brief description of some aspects of the attractor that are relevant to later sections is presented here.

Exploring the structure of a model attractor can be astronomically expensive. As noted by Lorenz (1987), little is actually known about the structure of the attractors of large dynamical systems. However, extremely long integrations of the Lorenz model are affordable and can reveal the nature of its attractor.

It is vital that these extended integrations be performed with high numerical precision for this model. It was found that long integrations became periodic with a period of several tens of millions of steps when evaluated with 32-bit floating point arithmetic. Obviously, any finite precision representation of the equations will eventually repeat; however, it is desirable that this repeat time be much longer than the time needed to sample the details of the attractor structure. Sixty-four bit floating point was used throughout this study, and no evidence of periodic model behavior was discovered. This is one example of how the attractor for the continuous equations and the numerical model can be fundamentally different. The remainder of this paper will be concerned with the attractor of the numerical model.

The attractor is composed of two distinct regions in phase space; the model tends to persist in one region of phase space for extended periods of time with occasional bifurcations causing a shift to the other region. The total nondimensional range of x, y, and z in the attractor is approximately 50 nondimensional units.

For this study, the most important aspect of the attractor is that it is not dense in the model's phase space [see Gleeson (1970) for a discussion of phase space]. Instead, the attractor is composed of a set of quasi-two-dimensional sheets that are imbedded in the three-dimensional phase space. Figure 1 shows three cross sections (Lorenz 1987) of the attractor generated by finding the first 10 000 points from an extended integration that lie within distance 0.01 (or about $1/5000$ of the total attractor size) of a given plane in phase space. The attractor is approximately two-dimensional when viewed on these scales.

Another method for studying the local attractor structure is to use analogs (Lorenz 1969; Toth 1991; Trevisan 1993). For example, a single point on the attractor is selected as a base point; the model is then integrated until 10 000 points with distance less than 1.0 from the base point are found. A study of analogs reveals that the attractor is composed of nearly two-dimensional sheets throughout the phase space. Locally (i.e., over a region with a radius about $1/50$ of the total attractor size), these sheets are very thin, with almost all points being within distance 0.0001 of a two-dimensional surface.

The two-dimensional structure of the attractor will be exploited in the ensemble forecasting experiments of the following sections. It is not clear how specialized the behavior demonstrated by this simple model may be since the structure of the attractors of higher dimensional models are not well documented. The attractors of some other low-order models, for instance the Lorenz three-variable dynamical model (Lorenz 1987), are also somewhat more complicated. Nevertheless, it seems plausible that higher-dimensional models might exhibit attractors that are of considerably lower dimension than the imbedding phase space of the model.

3. Ensemble forecasts in a perfect model

In this section, it will be demonstrated that failing to use information about the attractor structure when selecting ensemble initial conditions can lead to erroneous ensemble forecasts. The model described in the last section can be used as a perfect model to examine methods for selecting initial conditions for ensemble forecasts. Initially, the model is integrated for a very long time, at least 100 000 steps here, to ensure that it is on the attractor and no longer being influenced by the effects of arbitrary initial conditions. This equilibrated model is then taken as the "real world" physical system to be forecast.

Exact observations from the model consist of the values of x, y, and z. It is assumed that exact observations are impossible due to imperfections in the hypothetical observing system. Observations are generated by adding some random noise, selected from a given observational error distribution, to the states from the extended model integration. In this study, the distribution of observational error is assumed to be multinormal with standard deviation 1.0 for each of x, y, and z; the errors in the different variables are uncorrelated. The results are qualitatively insensitive to the size of the noise as long as the majority of the errors are small compared to the total size of the attractor.

Given a noisy observation and knowledge of the observational error distribution, a probability distribution for the "exact" observed state can be generated. The distribution for the exact observed state based *only* on a knowledge of the observational error distribution and the noisy observation is also multinormal with standard deviation 1.0 with the noisy observation as the mean.

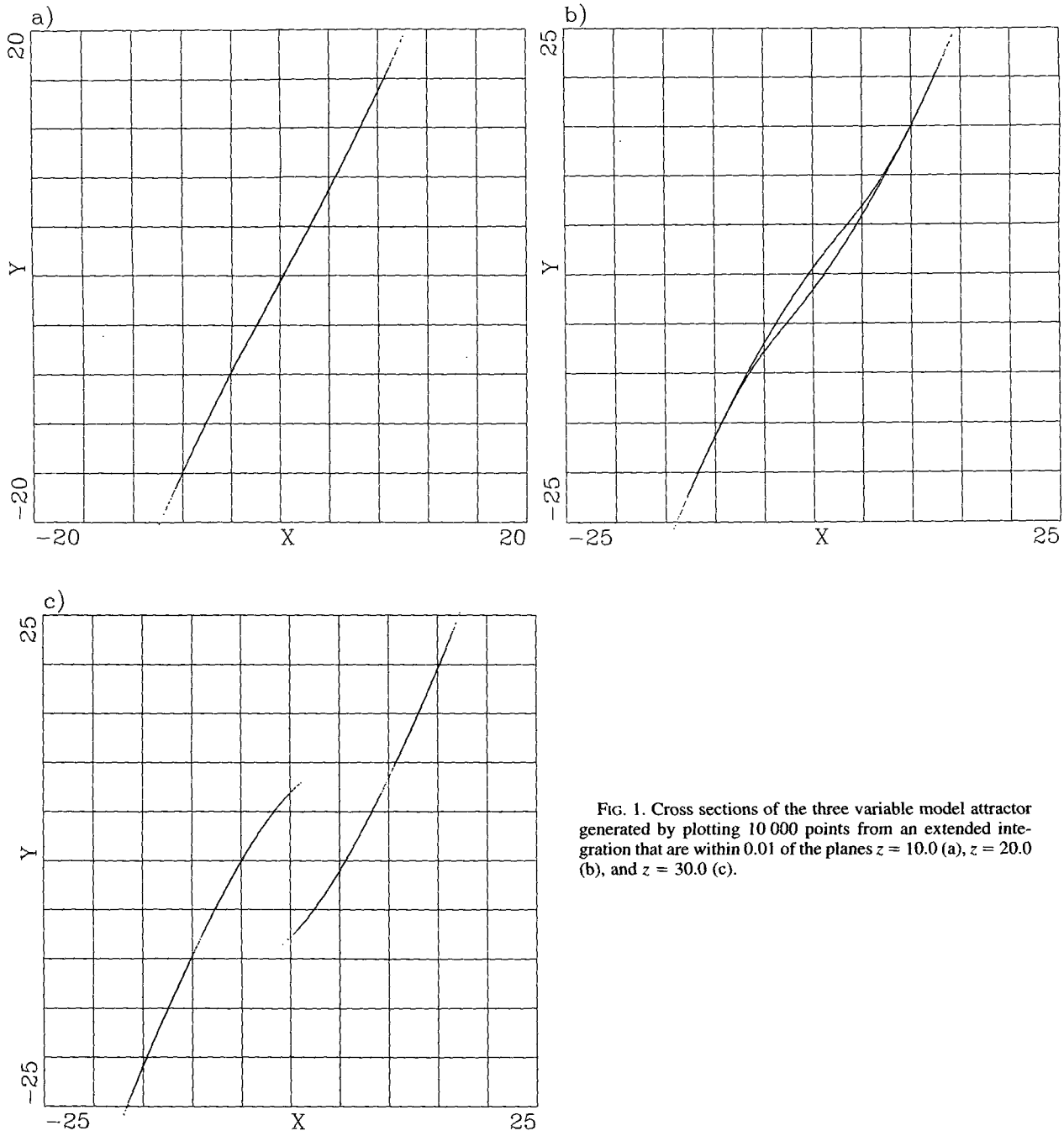

FIG. 1. Cross sections of the three variable model attractor generated by plotting 10 000 points from an extended integration that are within 0.01 of the planes $z = 10.0$ (a), $z = 20.0$ (b), and $z = 30.0$ (c).

In this perfect model context the observational error distribution and the noisy observation are not the only information available. The previous section has demonstrated that the attractor occupies only a small portion of the total model phase space. The exact observed state must lie on one of the quasi-two-dimensional sheets of the attractor. The rest of this section compares two types of ensemble forecasts. In the first set, the unconstrained ensemble, the only information used in generating the probability distribution of the initial conditions is the observational error distribution and the noisy observation. As noted above, the initial condition distribution in this case is a three-dimensional multinormal distribution.

The second set of forecasts, referred to as the *correct* ensemble forecasts, makes use of additional knowledge about the model attractor. For this ensemble the initial condition probability distribution is the conditional distribution of the three-dimensional unconstrained case given that the exact observation must lie on one of the

two-dimensional attractor sheets. Using the method outlined below, this is the best initial distribution that can be generated.

In this section, an additional constraint on the observational error will be introduced; this is that the noisy observation lies on the model attractor. This is obviously a highly artificial assumption, but it makes the initial discussion more straightforward. It should be noted that results based on this assumption will give a lower bound on the differences between the unconstrained and correct forecasts. Since the purpose of this section is to demonstrate that there are significant differences between forecasts from the unconstrained and correct distributions, this assumption is not a problem. The unconstrained forecasts with the noisy observation on the attractor are referred to as the unconstrained attractor (UA) ensemble forecasts. For multinormal observational error distributions, the correct ensemble initial distribution is not changed by the fact that the noisy observation is on the attractor. Figure 2 displays a schematic of the initial condition distributions for the correct and UA ensembles.

It is straightforward to generate a large sample of the three-dimensional UA initial condition distribution given a noisy observed point. Generating the quasi-two-dimensional distribution for the correct initial conditions is somewhat more difficult. First, a very large set of analogs to the noisy observed point is generated by integrating the model, sampling every 100th time step, and retaining all points that lie within a fixed distance of the noisy observed point. Then, this analog sample is itself selectively sampled to produce a set with the appropriate two-dimensional multinormal distribution. This can be done by rejecting a percentage of the analog points as a function of their distance from the observed point.

The initial distributions and the subsequent forecast distributions can be compared for the UA and correct cases. The differences between the two distributions can be regarded as errors in the UA case. The distributions will be compared using plots of the ensemble probability density projected onto planes in phase space. Both initial condition distributions are sampled with 100 000 ensemble members for the results shown here. This allows relatively smooth probability density plots to be made, even for forecast lead times that are well outside the "linear" range. Results for a single case will be presented; however, this case is representative of behavior found for noisy observations located anywhere on the model's attractor.

Figure 3 shows the projection on the local attractor "plane" of the initial ensemble distributions for the noisy observed point $x = 14.81$, $y = 12.65$, $z = 37.56$ (which lies on the model attractor) for the correct distribution (Fig. 3a) and for the difference between the UA and correct distributions (Fig. 3b). The UA case has probability densities that are concentrated too heavily near the center of the projected distributions. This

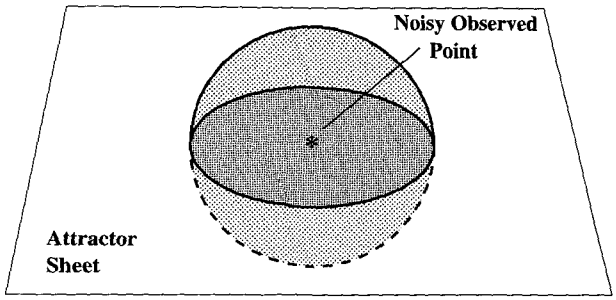

FIG. 2. Idealized representation of the initial condition distribution for the correct and UA ensembles. The UA distribution is represented by the sphere, while the correct distribution lies entirely on the quasi-two-dimensional attractor sheet. The asterisk represents the noisy observation, which lies on the attractor sheet in this case.

could have been deduced a priori; the density projection of a sphere onto a plane gives a higher concentration in the center, which falls off sharply toward the edges of the circular projection. Since the correct distribution is quasi-two-dimensional, it does not have the increased density in the center of the distribution that is found in the UA.

The evolution of the distributions as the ensembles are integrated can now be examined. During the early phases of the integration, the evolution of the shape of the ensemble probability density can be predicted well by linear theory. During this linear phase, both the two-dimensional correct and three-dimensional UA probability distribution ellipsoids are stretched or compacted along two and three axes, respectively. As the integration time increases, the three-dimensional UA distribution begins to collapse toward the attractor plane containing the two-dimensional correct distribution. This behavior continues on into the nonlinear evolution regime.

Figure 4 shows the projection on the x–y plane of the distribution for the correct ensemble after 50 time steps (it takes on the order of 100 time steps for the model to circulate around one of the two attractor regions) in a regime that is still mostly linear. The projections on the other coordinate planes look qualitatively similar. Figure 5 shows the difference between the x–y plane projections of the UA and correct ensemble distributions after 50 steps. Not surprisingly, the UA distribution continues to weight the central portion of the distribution too heavily, while the outer portions are not given sufficient probability. This difference is highlighted in Fig. 6, which shows the ratio of the UA to the correct distribution for the projection on the y–z plane. For the outer regions of the distribution, the UA ensemble has less than half of the correct probability while it is several times too large in the center.

As the integration time is further increased, the probability distributions begin to lose their ellipsoidal shape and become stretched and folded as is typical for chaotic systems (Guckenheimer and Holmes 1983). Points

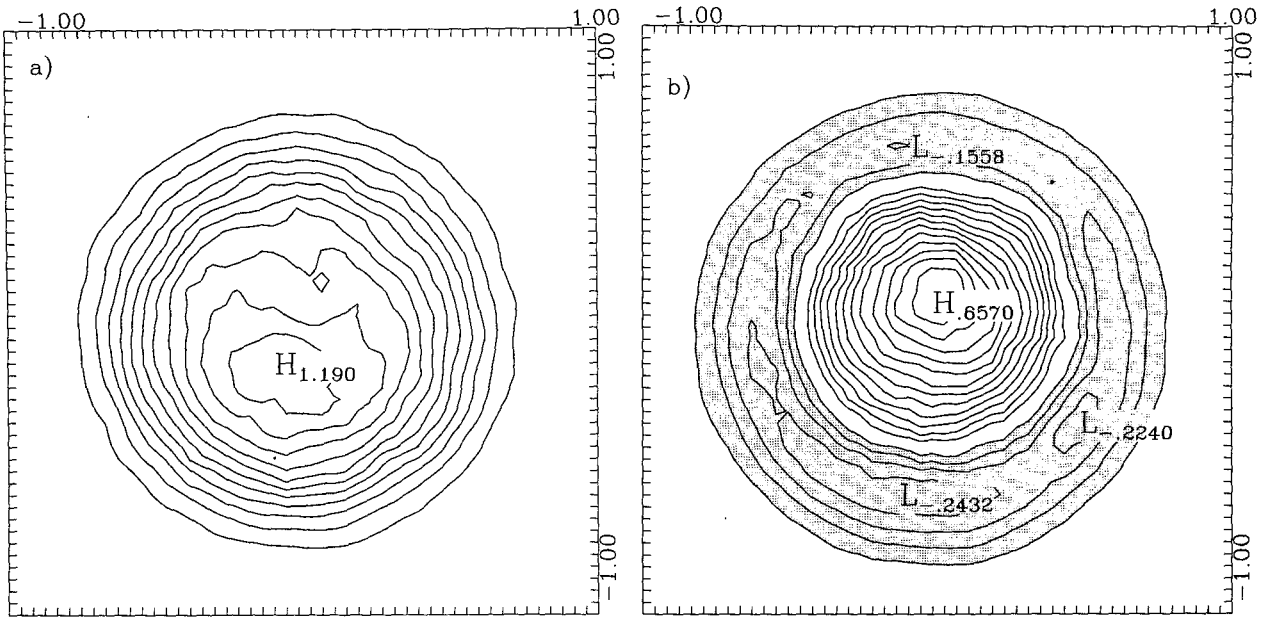

FIG. 3. Probability density projections on the local attractor plane for the correct (a) and for the UA minus correct (b) initial condition distributions for the noisy observed point (x = 14.81, y = 12.65, z = 37.56), which lies on the model attractor. Sampled values are placed in a 50 × 50 element binning grid, and a simple 9-point smoother is applied before contouring. Contour interval in (a) is 0.1 with the 0 contour omitted. Contour interval in (b) is 0.05 with regions less than −0.05 stippled.

that were initially on the fringe of the probability distributions are more likely to be stretched and twisted into long filaments at earlier times. However, it is these fringe regions that are insufficiently weighted by the UA distribution. Figure 7 shows a portion of the y–z plane projections for the correct and UA ensembles af-

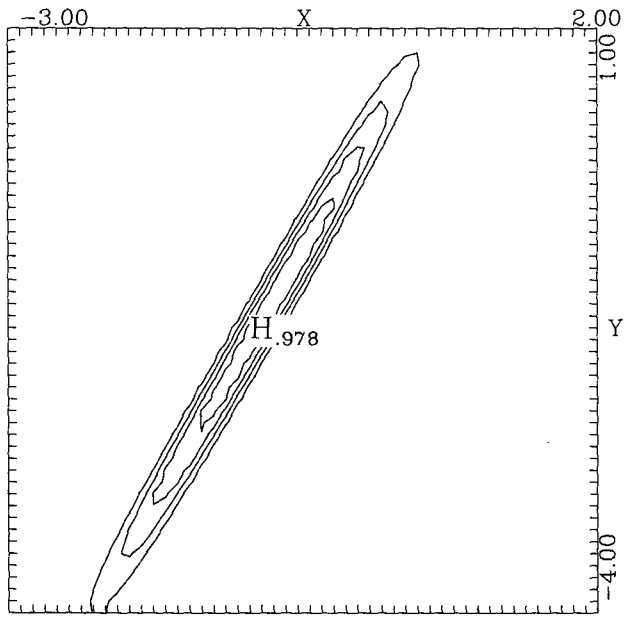

FIG. 4. Probability density projection on the x–y plane for the correct initial condition distribution of Fig. 3a after 50 steps of time integration. Contour interval is 0.2 with the zero contour omitted.

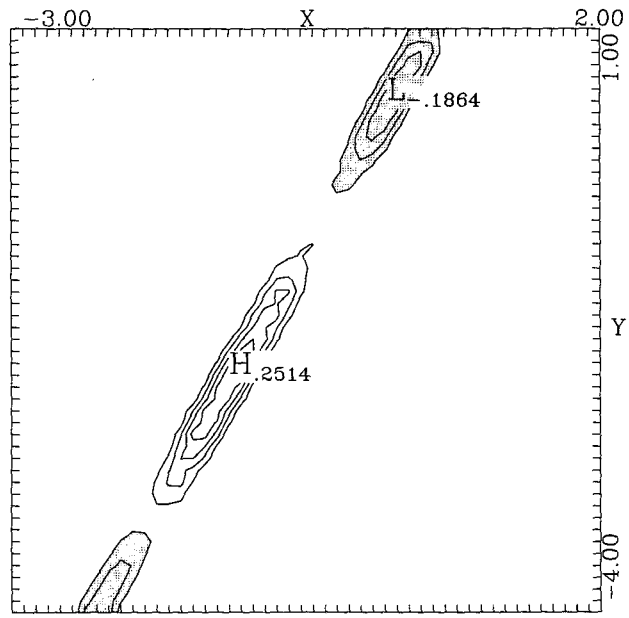

FIG. 5. Unconstrained attractor minus correct distribution probability density projection on the x–y plane after 50 steps of integration. Contour interval is 0.05 with the zero contour omitted; regions less than −0.05 are stippled.

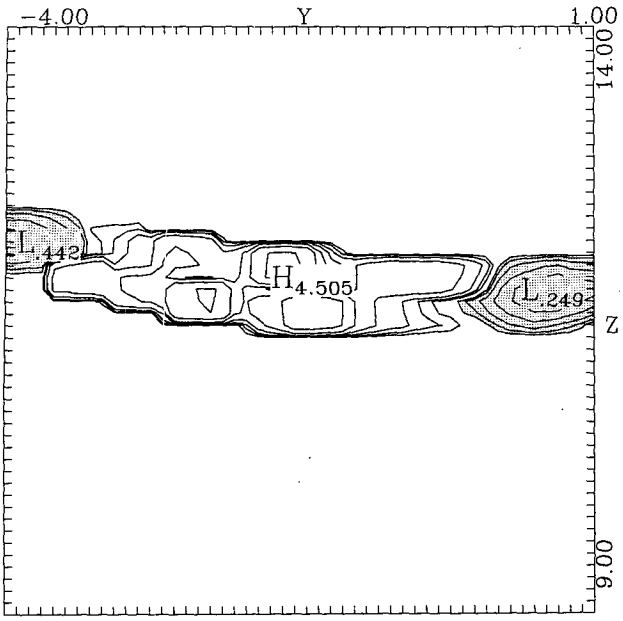

FIG. 6. Ratio between UA and correct probability density projections on the $y-z$ plane after 50 steps of integration. Contours are at 1.05, 1.10, 1.20, 1.40, 1.80, 2.60, 4.20, and the reciprocal of these values; nonzero values less than 1.0 are stippled. Bins where only one of the projections is 0 are given artificial values by increasing the number of points in the bin from 0 to 1..

ter 100 steps. As the correct distribution (Fig. 7a) shows, the distribution has been tremendously stretched by this time with one low-probability filament crossing the more probable "center" region of the distribution in this projection. The UA distribution (Fig. 7b) gives somewhat greater weight to the center region but gives far too little weight to the extended filament. A significant portion of the correct probability density lies in long stretched filament regions, which are all drastically underweighted (generally by a factor of 5 or more) by the UA distribution.

These results demonstrate that using knowledge about the model attractor can significantly influence the probability distribution of both the initial conditions and the forecasts. Errors resulting from using the UA initial conditions that do not make use of information about the attractor persist throughout the ensemble integration.

4. Finding the local attractor structure

In order to produce the correct ensemble initial conditions, it is necessary to compute the local structure of the attractor. This was done in the previous section using a long integration of the model to find analogs to the noisy observed point. This requires significant amounts of computation, even for an extremely simple model like the one used here. For large models, finding close analogs is practically impossible (Gutzler and Shukla 1984; Van den Dool 1994). In order to produce correct ensemble initial conditions in more realistic models, a less costly method for approximating the local attractor structure is essential.

Several possible alternative methods for finding the local attractor sheet in the simple model will be ex-

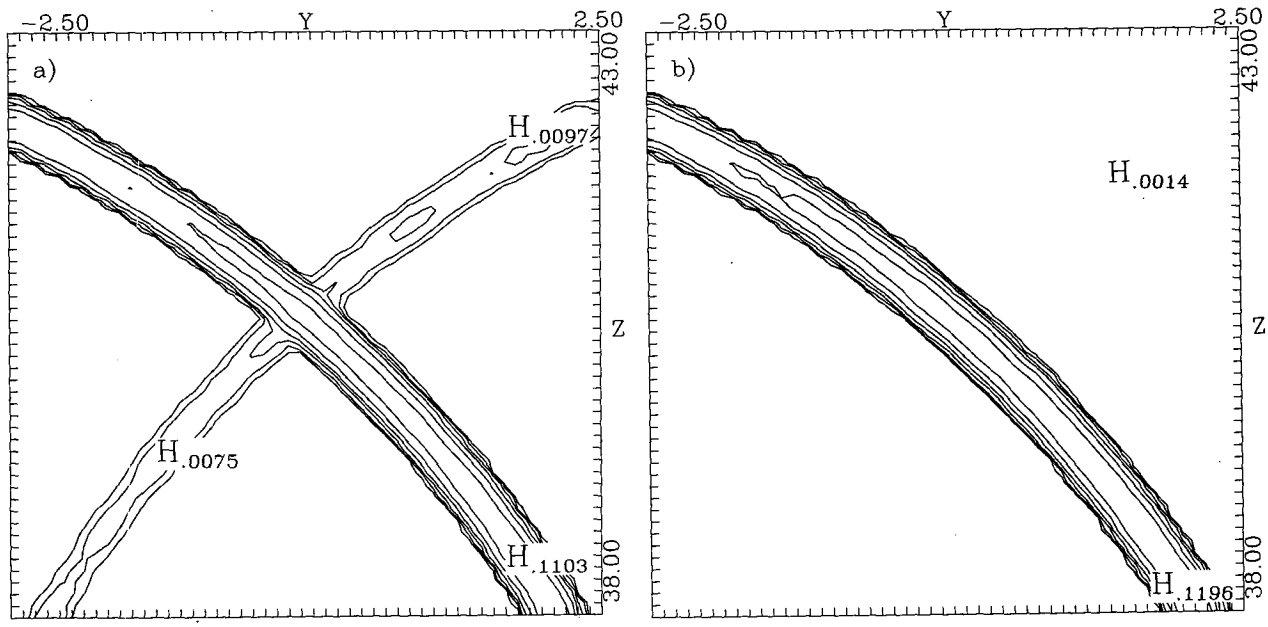

FIG. 7. Probability density projection on the $y-z$ plane after 100 steps of integration for the correct (a) and UA (b) ensembles. Contours are 0.002, 0.004, 0.008, 0.016, 0.032, and 0.064.

TABLE 1. Mean angle in degrees between singular vectors and vector perpendicular to local attractor plane as computed by analogs. Results are shown for optimization periods preceding and following the control points on the attractor by 1, 5, 10, and 20 steps. Angles are shown for all three of both the left and right eigenvectors.

Optimization period	Most stable left	Middle left	Most unstable left	Most stable right	Middle right	Most unstable right
1-step precursor	59.04	64.88	44.62	58.23	65.53	44.58
5-step precursor	50.43	66.18	54.76	56.44	65.07	47.21
10-step precursor	41.22	71.64	59.50	36.73	70.57	66.28
20-step precursor	41.07	69.55	60.59	37.81	69.77	66.53
1-step postcursor	59.16	64.95	44.40	57.47	65.38	45.74
5-step postcursor	48.19	68.44	54.11	55.16	65.15	48.72
10-step postcursor	40.47	73.35	57.95	36.19	66.75	69.53
20-step postcursor	39.27	63.91	69.09	41.82	65.66	64.85

amined in this section. Not purely by coincidence, the methods examined here are all related to methods that are used to select ensemble members in real forecast models at operational centers.

A set of 100 independent "test" points on the attractor is used to evaluate algorithms for finding the local attractor sheet in the simple model. For each of these 100 points the local attractor sheet is first computed by locating close "analog" points from an additional long integration of the model. As has already been noted, analog points within a radius of 1.0 (which defines "local" here) are very nearly coplanar. A perpendicular vector to the local analog attractor plane is computed by taking the average of all the cross-product pairs from a set of 10 close analogs.

The alternative algorithms compute a vector that is approximately perpendicular to the local attractor plane. The angle between these approximate vectors and the true (analog computed) vectors is computed for each of the 100 test points and used to evaluate the quality of the approximate algorithms.

a. Singular vectors

The first algorithm examined computes singular vectors (Farrell 1990) of the linear tangent propagator L of the simple model in the vicinity of the 100 test points (see the appendix for details of the definition and computation of singular vectors). A priori, it seems reasonable to assume that the most rapidly decaying structure, the singular vector with the smallest eigenvalue, might be closely related to the perpendicular to the local attractor sheet. This is a direction that generally decays very rapidly, as points off the attractor collapse onto the attractor. Growing singular vectors should be related to vectors on the attractor sheet. Since the attractor sheet is so nearly two-dimensional, one would not expect to have too much rapid growth projecting out of the plane of the attractor.

TABLE 2. Mean angle in degrees between purely real phase of normal mode eigenvectors and the vector perpendicular to local attractor plane as computed by analogs.

Mode	Angle
Most stable	22.65
Second most unstable	85.63
Most unstable	86.44

For completeness, a large number of singular vector predictors were compared to the true vectors. Singular vectors were computed for 1, 5, 10, and 20 time step optimization periods ending at t ($t = t_0 + n\Delta t$, $n = 1$, 5, 10, 20) and for the optimization period both preceding ($t_0 = t_p - n\Delta t$, where t_p is the time corresponding to the test point) and following ($t_0 = t_p$) the point on the attractor. In each case the singular vectors at both the initial and final times (the right and left eigenvectors, see appendix) were examined as potential predictors. Table 1 summarizes some of the results for the singular vector method. Even in the best cases the singular vectors were not particularly successful at predicting the orientation of the local attractor sheet. It was the singular vector with intermediate growth rate that seemed to be consistently the best predictor of the orientation of the attractor sheet.

b. Normal modes

As several long-running debates in the literature have demonstrated, if one can make an argument for the importance of singular vectors in a dynamical system, a similar argument can often be made for the exponentially growing normal modes of the linear tangent propagator L (see appendix; this is a more general definition of normal modes than has often been used in the atmospheric science literature). One might expect the most rapidly decaying normal mode to be approximately perpendicular to the local attractor sheet. Since L is not self-adjoint in the Lorenz model, the normal modes can be complex, and the eigenvectors are not generally orthogonal. Still, a single phase of an eigenvector can be used as a predictor of the local attractor sheet. Eigenvectors of L were computed for all the

cases outlined in the singular vector discussion above. However, the normal modes for a single time step, $n = 1$, with $t_0 = t_p$ were clearly superior to all others. The normal modes of this single-step propagator are simply the traditional normal modes of the model linearized around an observed state that have been discussed in many previous works. Table 2 summarizes the results for the 100 test points. The most rapidly decaying mode has a mean angle of approximately 22° to the analog vector, while the two other modes have angles of approximately 86°, indicating that they lie very nearly on the local attractor plane. The normal modes are clearly superior to singular vectors for the purpose of approximating the local attractor plane orientation in this model. Nevertheless, for some of the test points the normal mode method fails dismally. For instance, there are two cases for which the normal mode angle is between 50° and 60° for the most unstable mode. These failed cases seem to occur in regions where two separate attractor sheets are merging together.

A possible explanation for the superiority of the normal modes to the singular vectors as predictors of the local attractor structure follows. Suppose a point P_0 lies near the fringe of the local quasi-two-dimensional attractor sheet. A perpendicular vector, $\mathbf{X}_0$, can be constructed from the nearest point, Q_0, on the attractor plane (i.e., some local best fit plane to all points on the attractor) to P_0. If $\mathbf{X}_0$ projects heavily on a growing normal mode, then one would expect the point P_0 to be pushed farther away from the plane after an additional time step. After one time step, P_0 and Q_0 evolve into P_1 and Q_1, with the vector $\mathbf{X}_1$ from Q_1 to P_1 being parallel to and longer than $\mathbf{X}_0$. Since the normal mode vectors and the attractor plane orientation are relatively invariant over a few time steps, $\mathbf{X}_1$ would be nearly perpendicular to the local attractor plane and would also tend to have a large projection on the growing mode; P_1 would be pushed even farther from the plane after the next time step. Such a situation is not consistent with the fact that the attractor stays very nearly two-dimensional locally.

The same argument cannot necessarily be made for singular vectors because of their nonmodal nature. Suppose that P_0, Q_0, and $\mathbf{X}_0$ are defined as above. Suppose also that $\mathbf{X}_0$ projects heavily on a growing singular vector (i.e., the most rapidly growing right eigenvector of L^*L) with an optimization period of n time steps. After n additional time steps, P_0 and Q_0 evolve into P_1 and Q_1, and the distance between P_1 and Q_1 is greater than the distance between P_0 and Q_0. However, because the singular vectors are nonmodal, the vector from Q_1 to P_1 may not be close to perpendicular to the local attractor plane; it could even lie nearly on the attractor plane. The perpendicular vector, $\mathbf{X}_P$, from P_1 to the attractor plane could be shorter than $\mathbf{X}_0$, indicating that the point is being pulled back toward the local attractor sheet. Behavior like this would not be inconsistent with the nearly two-dimensional nature of the attractor. A similar argument can be made for most rapidly growing left eigenvectors of L^*L. Therefore, one expects that unstable normal modes must generally lie very nearly on the attractor plane, while growing singular vectors are not necessarily so constrained.

c. *Perturbed integrations*

In the perturbed integration method, the points that preceded each of the 100 test points in the nonlinear model integration by a certain number of steps, n, are located. Two additional randomly perturbed points close to each precursor point are constructed and integrated forward n steps to reach the time of the original test point. The vectors between the two perturbed points and the corresponding point in the original integration are orthogonalized and normalized to the original perturbation magnitude after each step (Fig. 8 demonstrates the procedure for $n = 3$). In the limit of small perturbation size, these vectors should converge to the two largest local Lyapunov vectors (Buizza and Palmer 1994). In general, given a large enough n, one might expect these vectors to lie on the local attractor sheet. However, it is also conceivable that such points might not converge if they were frequently affected by bifurcations of the attractor sheet.

The perturbed integration method was tried for precursor integrations of length $n = 1$, 5, 10, 20, 50, and 99 time steps. A number of normalization magnitudes were investigated, but the results were insensitive to this for magnitudes less than 1.0 and greater than 0.0001. The upper limit is apparently due to the fact that the attractor sheets begin to demonstrate significant curvature on these scales. The lower limit is determined by the local thickness of the attractor sheets. Results for the case with normalization amplitude 0.01 are

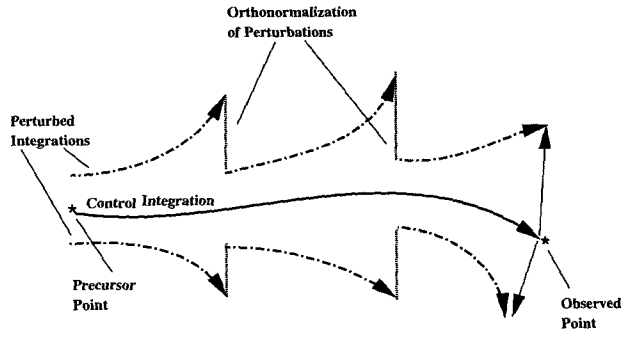

FIG. 8. Schematic diagram showing the perturbed integration method for generating an approximation to the local attractor sheet. The control integration is represented by the solid curve, and a pair of perturbed integrations by the dash-dotted curves. The dashed vertical lines represent the orthonormalization of the perturbations, which takes place after each time step. The two vectors projecting from the observed point at the end of the control integration should project heavily on the local attractor sheet.

TABLE 3. Mean angle between the cross product of two perturbed integration vectors and the vector perpendicular to local attractor plane as computed by analogs. Results are shown for perturbed integrations started at 1, 5, 10, 20, 50, and 99 steps previous to the control points. For these results the perturbation vectors were normalized to a magnitude of 0.01 after each step.

Steps	Angle
1	54.91
5	40.34
10	28.36
20	14.50
50	0.53
99	0.30

shown in Table 3. The approximation becomes increasingly accurate for longer precursor times, with mean angles between the true analog plane vector and the cross product of the two perturbation integration vectors being less than 1° for 50 and 99 steps. There are no significant outliers in these distributions, so it appears that this algorithm is robust throughout the attractor.

The results of this section have demonstrated that, given a point on the simple model's attractor, relatively inexpensive algorithms for finding the local structure of the attractor can be constructed. The normal mode approximation was fairly good and superior to the singular vector method, although it did break down in some regions of the attractor. The perturbed integration method is considerably more expensive than the normal mode method for low-order models like this. However, it produces better and more consistent results.

5. Ensemble forecasts without analogs

The previous section has demonstrated that, in this simple model, finding the orientation of the local attractor sheet is relatively simple given a point on the attractor. However, it may not be easy to find a point on the attractor given a point near the attractor. Recall that in section 3, the noisy observed points were constrained to lie on the attractor. This constraint is artificial even by the standards of the simple perfect model experiments being studied here.

At this point, the requirement that the noisy observation lie on the model attractor will be eliminated. Figure 9 displays a schematic of two additional ensemble distributions that will be studied. The unconstrained ensemble has the same three-dimensional multinormal distribution as the UA ensemble of section 3, except that its mean is not constrained to be on the attractor. The second new initial condition distribution is the constrained ensemble (Fig. 9). The relation between the unconstrained and the constrained ensembles is similar to that between the UA and correct ensembles in section 3 (Fig. 2). The constrained ensemble is the conditional distribution of the unconstrained distribution to a local "pseudoattractor" plane. This pseudoattractor plane is defined by applying the methods of the previous section to the noisy observed point that is not on the attractor plane. These algorithms will now find a plane in phase space passing through the noisy observation, but this plane is no longer a close approximation of the local attractor sheet. Nevertheless, as the discussion below will demonstrate, this pseudoattractor plane is likely to be approximately parallel to any nearby attractor sheet.

The correct distribution corresponding to the two new approximate distributions is identical to that studied in section 3 except that its mean is the point on the attractor that is nearest to the noisy observation (this statement is not necessarily true for observational error distributions that are not normal or in the vicinity of branching attractor sheets).

A large number of ensemble integrations were made for both constrained and unconstrained initial conditions for noisy observed points throughout the model attractor. When unconstrained forecasts are made for noisy observed points that are close (within distance 1.0) to the model attractor, the initial three-dimensional distribution collapses within approximately 50 time steps to a quasi-two-dimensional distribution. At the same time, the entire distribution gradually moves toward the nearby attractor sheet. If the noisy observed

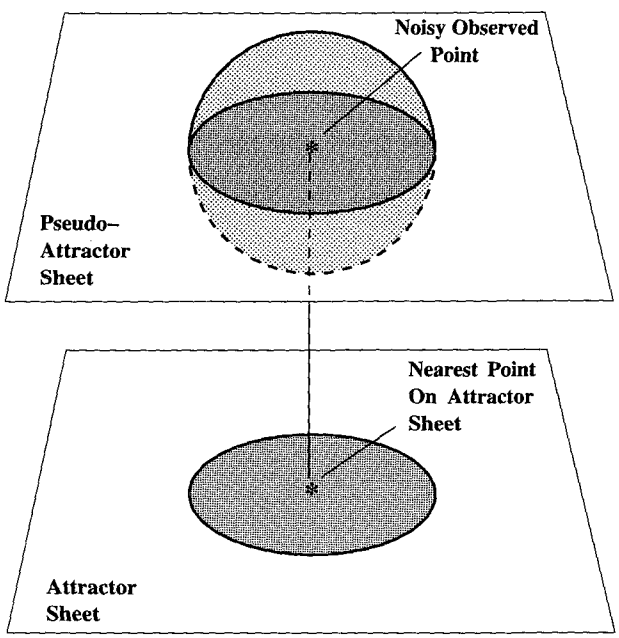

FIG. 9. Idealized representation of the initial condition distribution for the unconstrained, constrained, and correct initial condition distributions. The unconstrained distribution is represented by the sphere centered on the noisy observation, while the constrained is the circle confined to the pseudoattractor plane. The correct distribution is on the attractor sheet, which is approximately parallel to the pseudoattractor plane, and is centered about the point on that plane closest to the noisy observation.

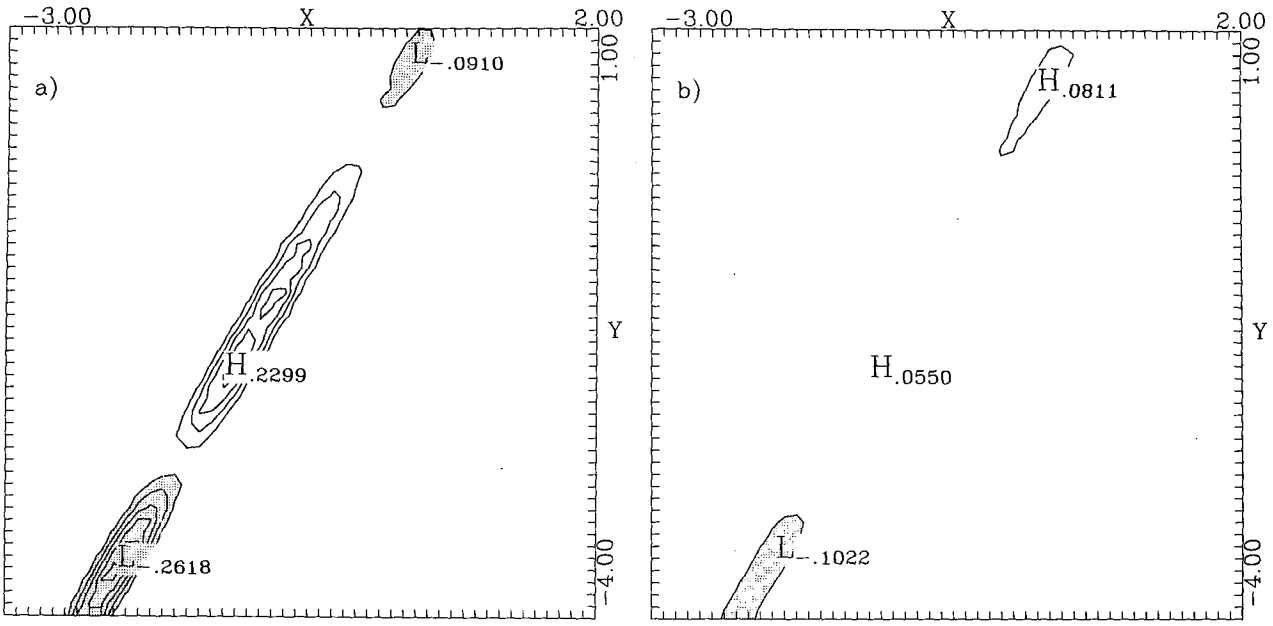

FIG. 10. Difference in probability density projection on the $x-y$ plane between the correct distribution and the unconstrained (a) and constrained (b) distributions after 50 steps of integration. Contours are as in Fig. 5.

point is far from the nearest attractor sheet, the three-dimensional cloud of points still collapses to a quasi-two-dimensional distribution in about 50 steps, but it may take much longer for this distribution to move onto an attractor sheet. The behavior of the constrained distributions is analogous; for noisy observations close to the attractor, the distribution remains quasi-two-dimensional and slowly moves toward the nearest attractor sheet. The constrained distribution might be expected to produce a good simulation of the evolution of the correct distribution since the dynamical behavior of perturbations in directions perpendicular to the local attractor planes is dominated by a decay toward the plane. This hypothesis was supported by tests on a large number of noisy observed points.

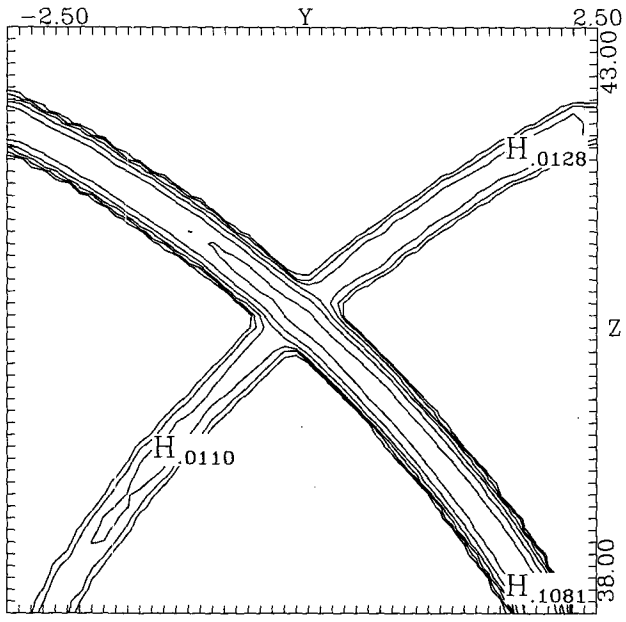

FIG. 11. Probability density projection on the $y-z$ plane after 100 steps of integration for the constrained ensemble. Contours are as in Fig. 7.

In order to demonstrate the behavior of the unconstrained and constrained ensembles, the noisy observed point of the examples in section 3 is shifted by a phase space distance of 0.5 in a direction perpendicular to the attractor plane. The noisy observed point in this revised example is $x = 14.38$, $y = 12.76$, $z = 37.74$. The correct distribution corresponding to this noisy observation is identical to that discussed in section 3. Both the unconstrained and constrained ensembles for this noisy observed point are sampled using 100 000 ensemble members.

Forecasts produced from the unconstrained distribution are somewhat less similar to the correct forecast distributions than were those produced by the UA ensemble in section 3. Figure 10a shows the difference in the $x-y$ plane projection of the probability density for 50 step forecasts between the unconstrained and correct ensembles and can be compared to Fig. 5.

Figure 10b shows the same difference but for the constrained ensemble and the correct ensemble. The differences are less than those found for the UA ensemble. The superior performance of the constrained ensemble continues into the nonlinear regime. Figure 11

shows the constrained ensemble distribution after 100 time steps corresponding to the UA and correct distributions in Fig. 7. The constrained distribution is very close to the correct and better approximates the probability of less likely events than either the UA or unconstrained ensembles. The unconstrained distribution (not shown) is more similar to the UA distribution and underweights the less likely events. This demonstrates that, at least in this particular simple model, using some information about the nearby attractor structure can produce better ensemble forecasts, even if the exact location of the nearby attractor remains unknown.

A brief discussion of how the constrained ensemble members were generated is appropriate. The Lorenz model was integrated backward 30 steps to find a precursor of the noisy observed point. For precursors of less than about 50 steps, the model can be integrated backward using the two-step method of the forward model applied to the negative of the local time tendency; even older precursors can be found successfully by explicitly inverting the two-step time-stepping method used in the forward model. Once the 30-step precursor point was found, the perturbed integration method of section 4 was used to find a local plane. Ensemble members were then selected from a binormal distribution on this plane. This adds an additional inaccuracy to the constrained distribution. Analog points on the quasi-two-dimensional model attractor are not exactly uniformly distributed locally. The correct distribution constructed in section 3 sampled this nonuniform structure by selecting the local distribution from a large set of analogs. The nonuniform behavior is reflected in the fact that the correct distribution of Fig. 3a is not a perfect ellipse; instead, the area of highest density is offset from the center of the distribution. The constrained initial condition distribution assumes local uniformity, and so the initial condition distribution contours (not shown) are circles. The similarity between the forecast distributions demonstrates that the departure from uniformity of the analogs is a minor effect in this case.

The results of the last three sections can now be combined in a method for ensemble forecasting (in this simple perfect model) that does not utilize analogs. Assume that one is given a noisy observation of the perfect model and a probability distribution for the observational error. Given a noisy observed point, the methods of section 4 can be used to compute the local plane that is approximately parallel to nearby attractor sheets. Next, given the distribution of the observational error, find the conditional distribution when restricted to the local plane. This distribution can then be sampled with an ensemble of whatever size is affordable. The ensemble members are then integrated to produce a sample of the forecast probability distribution.

This method could be combined with more sophisticated data assimilation techniques, for instance adjoint methods for four-dimensional assimilation (Derber 1989). This would allow one to find a better approximation to the exact observation than is given by a single noisy observation. However, it would also change the probability distribution of the error in the initial condition.

6. Implications for more complicated models

The differences between the constrained and unconstrained ensemble integrations demonstrated in the previous sections are somewhat subtle, but such differences might be much larger in higher-order models. In the Lorenz model the difference in dimension between the full three-dimensional phase space and the local attractor is only 1 (this statement is not entirely rigorous since the local attractor is technically only quasi-two-dimensional). However, in larger models the difference in dimension could be much larger. The degree to which the unconstrained initial conditions overweight the center of the distribution when projected onto the attractor is approximately proportional to σ^n, where σ is a constant greater than 1 and n is the difference in dimension. (This relation comes from projecting the mass of an $m + n$ dimensional hypersphere of uniform mass onto an m dimensional manifold and comparing the difference in weights between the projection on the center and the edge of the manifold). This relation suggests that the benefits of using the constrained ensemble initial conditions could be much greater in higher-order models. In addition, the dimension of the distribution that must be sampled by the ensemble is reduced when using the constrained initial conditions. In real models, where each integration can be costly, any method for reducing integrations that do not accurately sample the initial condition distribution could be of value.

There remain a number of technical difficulties with extending the methods of this report to more realistic models. One problem is that the local attractor structure of more complicated models is unknown (Lorenz 1987). Furthermore, it is too expensive to determine the attractor structure by the analog method presented here. The alternative algorithms presented for the Lorenz model in section 4 are implicitly dependent on the fact that the attractor is locally nearly two-dimensional. Without knowing some details about the local attractor structure, for instance the local dimension of the attractor, it becomes difficult to apply these algorithms to find ensemble initial conditions that lie on the attractor or on a surface parallel to the attractor.

An additional problem involves the expense of the algorithms of section 4 when they are applied to models with many degrees of freedom. Even if information about the local attractor structure is known, it may be too expensive to apply algorithms to approximate the local attractor. For instance, algorithms for computing normal modes or singular vectors cannot be applied to models as large as current operational NWP models.

Operational centers that generate ensembles using these modes generally use heuristic simplifications or compute the normal modes for models that are much simpler than the full forecast models (Mureau et al. 1993). Although application of the perturbed integration method could be considerably less expensive for large models, it is not entirely clear at this point how this algorithm can be extended to more complicated models. The perturbed integration method as developed here requires a precursor point of the observations; it is not clear that practical algorithms exist for finding precursors in large models.

Despite these difficulties, research into finding good initial ensemble members may still be productive. Because operational centers are, at least for the present, compelled to run forecast models at close to the highest resolution technologically possible, it seems unlikely that any center will run very large ensembles in the foreseeable future. The selection of good initial ensemble members becomes even more essential when a very small sample is being used to approximate the evolution of an observational distribution. It is quite possible that information from even a limited heuristic approximation to the types of algorithms presented in section 5 could be valuable in this context. In fact, the operational centers have already implicitly accepted this assumption by using heuristic algorithms to select initial ensemble members. Methods making use of the most rapidly growing normal modes and singular vectors to select ensemble initial conditions have been investigated at the European Center for Medium-Range Weather Forecasts (Mureau et al. 1993). The perturbed integration method for approximating the local attractor sheet is somewhat similar to the breeding cycle used operationally for ensemble selection at the National Meteorological Center (NMC) (Toth and Kalnay 1993; Kalnay and Toth 1994).

Although it is not immediately clear how to extend algorithms for finding the model attractor, at least a few suggestions can be made. First, if one is given a point on the attractor of a large model, it may be possible to discover something about the local structure of the attractor. If the attractor is not dense in the phase space, but instead has a reduced dimension of some sort (Houtekamer 1991), an extension of the perturbed integration method may be able to produce information about the attractor structure. One could integrate a large set of perturbations around a precursor state and orthonormalize these at each step. As the perturbed integration time increases, these perturbations should become increasingly rich in their projection on the local attractor. It should also be possible to find an approximation to the local dimension of the attractor by determining how many independent vectors are needed to form a basis for a significant portion of the subspace spanned by these perturbation vectors. Similar algorithms can be formulated for linear problems (Anderson 1991), and it is not immediately clear that this method would not be successful. If one could find the local attractor structure given a point on the attractor, then one also might be able to approximate the nearby attractor structure given a point near the attractor (a noisy observation) using the perturbed integration method. The existence of bifurcations and related complexities are still an unresolved problem for these approaches. Small additional complexities resulting from bifurcations were also ignored in sections 3–5 to simplify the presentation.

In order to investigate some of these possibilities, additional research with models of intermediate complexity is needed. Given current computational resources, it might be possible to repeat the experiments presented here in some slightly less simplified model such as the 28-variable model of Lorenz (1965). Such a model allows for much more variety in the local structure of the attractor and might be small enough that some information from analogs could be used. If the methods proved successful in this context, they could then be extended to more realistic models. Evaluating the methods in a more realistic context would be difficult since information about the local attractor structure from analogs would no longer be accessible.

Finally, all discussions to date have focused on perfect model experiments. Unfortunately, one is limited to a single realization of a perfect model experiment with the real atmosphere because no close analogs are ever likely occur (Van den Dool 1994). Algorithms for ensemble selection must be able to deal with both observational error [Kalnay and Toth (1994) addressed this directly with the breeding method] and a model attractor that is not the same as the attractor of the system being modeled (the real world). The easiest way to extend methods like those discussed is to view the combined observational/assimilation/forecast model system as a single stochastic dynamical system. In this case, an algorithm similar to the perturbed integration method might make use of "parallel" assimilation cycles in order to estimate the structure of the attractor in the real forecast model. Such a method would be somewhat similar to the NMC's "breeding" method, which makes use of a single assimilation.

7. Conclusions

Methods for selecting initial conditions for ensemble forecasts in a simple perfect model context have been presented. It has been demonstrated that using only information about the distribution of "observational" error in the perfect model when selecting ensemble members can result in an incorrect representation of both the initial condition probability distribution and the probability distribution of the forecasts. The correct initial and forecast distributions can be obtained only if information about the attractor structure (climate) of the perfect model is utilized. In the simple model used here the correct initial distribution is a quasi-two-di-

mensional structure, while the distribution based only on the observational error is a fully three-dimensional structure. Ensemble forecasts based solely on the observational error distribution overestimate the probability of points in phase space that are close to the observed point while more remote points are insufficiently weighted. This problem might be exacerbated in more complex models, although this has not been demonstrated.

Initially, the local attractor structure was found by using extremely long integrations of the forecast model and selecting close analog points to a noisy observation. This method is far too costly to apply to all but the simplest models. Therefore, a number of heuristic algorithms for finding the local attractor structure were presented. Three basic classes of algorithms, normal modes, singular vectors, and perturbed integrations, were investigated. These three methods are closely related to heuristic methods that have been used for ensemble initial condition selection in realistic forecast models. While the singular vector method gave some information about the local attractor structure, it was inferior to the normal mode method, which was in turn inferior to the perturbed integration method.

These three algorithms were first developed in a setting in which the observed point was known to be on the model attractor. This highly artificial condition was later removed without significantly affecting the results of the ensemble integrations.

Extending the results of this study to more realistic models will not be simple. It is not feasible to explicitly compute the local attractor structure in any significantly more complicated model, so it will become much more difficult to validate the algorithms presented. In addition, some a priori knowledge about the local attractor structure was exploited in the methods presented here; such information is not readily available in more realistic models. Despite these difficulties, there seems to be some hope that an extension of the perturbed integration method will be able to provide information about the local attractor structure in much more complicated models. This information in turn could be used to produce ensemble initial conditions that are good samples of the correct initial condition probability distribution and which may provide better operational ensemble forecasts.

Acknowledgments. The author is indebted to K. Miyakoda, A. Navarra, and three anonymous referees, whose suggestions significantly improved this study.

APPENDIX A

Singular Vectors and Normal Modes

Singular vectors (Buizza and Palmer 1994) are defined as the eigenvectors of $A = L^*L$, where $L = L(t, t_0)$ is the linear tangent propagator or resolvent (Talagrand and Courtier 1987) for the simple model and the asterisk represents the adjoint operator. The linear operator L gives the evolution of a small perturbation to the model state at time t_0, after it has been integrated to time t. The left eigenvector of A with the largest eigenvalue represents the final condition that demonstrates the largest growth from time t to t_0; the right eigenvector is the initial perturbation that results in the left eigenvector. The L2 norm in phase space has been used implicitly as a distance metric in this definition. The normal modes discussed in this paper are the eigenvectors of the resolvent L itself. For a single time step optimization period, these are equivalent to the eigenvectors of the linear tangent operator of the model.

REFERENCES

Anderson, J. L., 1991: The robustness of barotropic unstable modes in a zonally varying atmosphere. *J. Atmos. Sci.*, **48,** 2393–2410.

Buizza, R., and T. N. Palmer, 1994: The singular-vector structure of the atmospheric general circulation. *J. Atmos. Sci.*, **52,** 1434–1456.

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

Derber, J. C., 1989: A variational continuous assimilation technique. *Mon. Wea. Rev.*, **117,** 2347–2446.

Ebisuzaki, W., and E. Kalnay, 1991: Ensemble experiments with a new lagged analysis forecasting scheme. Research Activities in Atmospheric and Oceanic Modeling Rep. 15, World Meteorological Organization, 6.31–6.32.

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

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.

Guckenheimer, J., and P. Holmes, 1983: *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields.* Springer-Verlag, 453 pp.

Gutzler, D. S., and J. Shukla, 1984: Analogs in the wintertime 500 mb height field. *J. Atmos. Sci.*, **41,** 177–189.

Hoffman, R. N., and E. Kalnay, 1983: Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus,* **35A,** 100–118.

Houtekamer, P. L., 1991: Variation of the predictability in a low-order spectral model of the atmospheric circulation. *Tellus,* **43A,** 177–190.

Kalnay, E., and Z. Toth, 1994: Removing growing errors in the analysis cycle. Preprints, *10th Conf. on Numerical Weather Prediction,* Portland, OR, Amer. Meteor. Soc., 212–215.

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

Lewis, T. G., 1975: *Distribution Sampling for Computer Simulations.* Lexington Books, 333 pp.

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

——, 1965: A study of the predictability of a 28-variable atmospheric model. *Tellus,* **17,** 321–333.

——, 1969: Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.*, **26,** 636–646.

——, 1982: Low-order models of atmospheric circulations. *J. Meteor. Soc. Japan,* **60,** 255–267.

——, 1987: Deterministic and stochastic aspects of atmospheric dynamics. *Irreversible Phenomena and Dynamical Systems Analysis in Geosciences,* C. Nicholis and G. Nicholis, Eds., D. Reidel, 159–179.

Molteni, F., and T. N. Palmer, 1993: Predictability and finite-time instability of the northern winter circulation. *Quart. J. Roy. Meteor. Soc.,* **119,** 269–298.

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

Murphy, J. M., 1990: Assessment of the practical utility of extended range ensemble forecasts. *Quart. J. Roy. Meteor. Soc.,* **116,** 89–125.

Schubert, S. D., M. J. Suarez, and J.-K. Schemm, 1992: Persistence and predictability in a perfect model. *J. Atmos. Sci.,* **49,** 256–269.

Seidman, A. N., 1981: Averaging techniques in long-range weather forecasting. *Mon. Wea. Rev.,* **109,** 1367–1379.

Sparrow, C., 1982: *The Lorenz Equations: Bifurcations, Chaos and Strange Attractors.* Springer-Verlag, 269 pp.

Talagrand, O., and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Quart. J. Roy. Meteor. Soc.,* **113,** 1311–1328.

Toth, Z., 1991: Estimation of atmospheric predictability by circulation analogs. *Mon. Wea. Rev.,* **119,** 65–72.

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

Tracton, M., S. Kalnay, and E. Kalnay, 1993: Operational ensemble prediction at the National Meteorological Center: *Practical Aspects. Wea. Forecasting,* **8,** 379–398.

Traub, J. F., and H. Wozniakowski, 1994: Breaking intractability. *Sci. Amer.,* **270**(1), 102–107.

Trevisan, A., 1993: Impact of transient error growth on global average predictability measures. *J. Atmos. Sci.,* **50,** 1016–1028.

Van den Dool, H. M., 1994: Searching for analogues, how long must we wait? *Tellus,* **46A,** 314–324.