Ensemble Transformation and Adaptive Observations

CRAIG H. BISHOP
Department of Meteorology, The Pennsylvania State University, University Park, Pennsylvania

ZOLTAN TOTH
GSC, NCEP, Camp Springs, Maryland

(Manuscript received 22 May 1998, in final form 14 October 1998)

ABSTRACT

Suppose that the geographical and temporal resolution of the observational network could be changed on a daily basis. Of all the possible deployments of observational resources, which particular deployment would minimize expected forecast error? The ensemble transform technique answers such questions by using nonlinear ensemble forecasts to rapidly construct ensemble-based approximations to the prediction error covariance matrices associated with a wide range of different possible deployments of observational resources. From these matrices, estimates of the expected forecast error associated with each distinct deployment of observational resources are obtained. The deployment that minimizes the chosen measure of forecast error is deemed optimal.

The technique may also be used to find the perturbation that evolves into the leading eigenvector or singular vector of an ensemble-based prediction error covariance matrix. This time-evolving perturbation “explains” more of the ensemble-based prediction error variance than any other perturbation. It may be interpreted as the fastest growing perturbation on the subspace of ensemble perturbations.

The ensemble-based approximations to the prediction error covariance matrices are constructed from transformation matrices derived from estimates of the analysis error covariance matrices associated with each possible deployment of observational resources. The authors prove that the ensemble transform technique would precisely recover the prediction error covariance matrices associated with each possible deployment of observational resources provided that (i) estimates of the analysis error covariance matrix were precise, (ii) the ensemble perturbations span the vector space of all possible perturbations, and (iii) the evolution of errors were linear and perfectly modeled. In the absence of such precise information, the ensemble transform technique links available information on analysis error covariances associated with different observational networks with error growth estimates contained in the ensemble forecast to estimate the optimal configuration of an adaptive observational network. Tests of the technique will be presented in subsequent publications. Here, the objective is to describe the theoretical basis of the technique and illustrate it with an example from the Fronts and Atlantic Storm Tracks Experiment (FASTEX).

1. Introduction

Humans are and always have been at the mercy of chaotic systems. The behavior of stock markets, politics, and the weather are all made difficult to forecast by the fact that apparently insignificant events sometimes have catastrophic consequences. People who attempt to forecast the evolution of chaotic systems need to be able to predict when and where small uncertainties will lead to forecast failure. If they have access to this information, then steps can be taken to ensure that the aspects of the system that require accurate observations receive the needed observations.

Since, in meteorology, the location of data-sensitive regions will vary from day to day, some sort of “adaptive” observational network would be required (Emanuel et al. 1995). The adaptive component might involve aircraft that drop instruments when over data-sensitive regions, pilotless aircraft or blimps that can be directed to data-sensitive regions, radiosonde stations that only release balloons when they lie in a data-sensitive region, and satellite-borne lidars that save power by only being switched on when they can be directed at data-sensitive regions. The cost-effective use of this adaptive component will require the ability to use a forecast or ensemble of forecasts initialized at \( t_a \) to predict which of the \( I \) possible deployments of observational resources at a future observation/analysis time \( t_f \) will minimize the expected prediction error of forecasts for the verification time \( t_y \), which are initialized with, inter alia, the supplemental data taken at \( t_f \). Note that in order to
leave sufficient time for aircraft, etc., to reach optimal observing sites, this prediction has to be made before \( t_{\text{fa}} \) at some decision time, \( t_d \) (cf. Fig. 1). The ensemble transform technique finds the deployment of observational resources that minimizes expected prediction error by first estimating the expected prediction error associated with all \( I \) possible deployments of adaptive observational resources and then searching this field of expected prediction error for the deployment of observational resources that minimizes it.

Accurate predictions of the forecast error associated with the \( i \)th possible deployment of observational resources could be made from an unthinkably large “ideal” ensemble forecast generated by a perfect model of the stochastic processes leading to analysis and prediction error [in Anderson’s (1997) terminology, this is an unconstrained ensemble]. In such an ensemble, the probability of a particular class of events occurring would be equal to the number of ensemble members that fell into this class divided by \( N \), \( N \) being the (very large) number of ensemble perturbations. The distribution of the values of ensemble variables would be consistent with the known statistics of observational and background errors associated with the \( i \)th possible deployment of observations at the future analysis time \( t_{\text{fa}} \). Although such an ensemble is unattainable before \( t_{\text{fa}} \) because, inter alia, one needs the observations at \( t_{\text{fa}} \) to find the best estimate of the state of the atmosphere about which to initialize the ensemble representing the uncertainty of forecasts from \( t_{\text{fa}} \), we discuss it since its statistical properties are what we seek to divine. This ideal ensemble would give accurate predictions of the distribution and range of possible prediction errors associated with the \( i \)th possible deployment of observations. The key features of this distribution could be synthesized in the ensemble’s perturbation covariance matrix

\[
P_i(t) = \frac{V_i(t)V_i(t)^T}{N},
\]

where \( V_i(t) \) is a matrix whose columns are the ensemble perturbations about the mean of the ensemble, \( N \) is the (unthinkably large) number of ensemble perturbations, and \( t \) is the time. The subscript \( i \) indicates to which of the \( I \) possible future observational networks the perturbation covariance matrix pertains. Note that \( V_i \) has \( N \) columns and \( M \) rows. The number of rows, \( M \), would typically be equal to the dimension of the state vector; for example, in a gridpoint model with six independent variables at each grid point, \( M \) would be equal to six times the number of grid points. (A list of important symbols is given in Table 1.)

If the mean of our ideal ensemble forecast were indistinguishable from the true state of the atmosphere and if the number \( N \) of ensemble perturbations tended to infinity, \( P_i(t) \) would be equivalent to the prediction error covariance matrix \( P_i(t) \) of forecasts made from the \( i \)th possible deployment of observational resources at the future analysis time \( t_{\text{fa}} \). Under these circumstances, the future analysis error covariance matrix \( A_i(t_{\text{fa}}) \) would be obtainable from (1) by simply evaluating the ensemble perturbations at their initialization time \( t_{\text{fa}} \); that is,

\[
P_i(t_{\text{fa}}) = A_i(t_{\text{fa}}).
\]

The significance of the aforementioned connection between the perturbation covariance matrix of an ideal ensemble and the prediction and analysis error covariance matrices mentioned above is twofold. First, it indicates how observations should affect ensemble spread. (Ensemble spread may be loosely defined as the average size of the differences between ensemble members.) The spread of an ensemble initialized at \( t_{\text{fa}} \) representing an observational network that included localized supplemental observations over an oceanic region should, in the supplemental observation region, be less than the corresponding spread of an ensemble that represented an alternative observational network that did not include supplemental observations. Subsequent to the observation time, the difference between the spreads of the two ensembles would propagate downstream with the evolving ensemble perturbations. If this difference in the spreads of the ensembles ended up extending over some localized verification region at the verification time \( t_v \), one could conclude that supplementary observations reduced expected prediction error. Second, the connection illustrates the validity of a fundamental assumption of the ensemble transform technique—that estimates of prediction and analysis error covariance matrices may, in principle, be made from appropriately configured ensemble perturbations.

Prediction of forecast error covariance matrices\(^\dagger\) represent key aspects of our uncertainty in weather forecasts. For example, one might be interested in using adaptive observations to minimize normalized predic-

\[^\dagger\] Daley (1991, chap. 13) describes how, in principle, accurate estimates of prediction and analysis error covariance matrices could be obtained via the computationally exorbitant Kalman–Bucy filter.
Table 1. Important symbols.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t_a )</td>
<td>Time at which control forecast and ensemble perturbations ( X(t) ) are initialized (( t_f ) precedes ( t_a )).</td>
</tr>
<tr>
<td>( t_f )</td>
<td>Future analysis time. The time at which supplemental observations are to be taken.</td>
</tr>
<tr>
<td>( t_c )</td>
<td>Verification time.</td>
</tr>
<tr>
<td>( t_d )</td>
<td>Decision time. Time by which the best possible deployment of observational resources must be identified.</td>
</tr>
<tr>
<td>( V_i(t) )</td>
<td>( M \times N ) matrix of near-infinite number ideal ensemble perturbations corresponding to the ( i )th possible deployment of observational resources. Each column represents an ensemble perturbation.</td>
</tr>
<tr>
<td>( N )</td>
<td>Near-infinite number of ideal ensemble perturbations.</td>
</tr>
<tr>
<td>( M )</td>
<td>Number of elements in the state vector.</td>
</tr>
<tr>
<td>( I )</td>
<td>Number of possible deployments of observational resources.</td>
</tr>
<tr>
<td>( P_e(t) )</td>
<td>Ensemble perturbation covariance matrix; also used as estimate of prediction error covariance matrix.</td>
</tr>
<tr>
<td>( A_i(t_{fa}) )</td>
<td>( M \times M ) analysis error covariance matrix corresponding to ( i )th possible deployment of observational resources.</td>
</tr>
<tr>
<td>( P_i(t) )</td>
<td>( M \times M ) prediction error covariance matrix corresponding to ( i )th possible deployment of observational resources.</td>
</tr>
<tr>
<td>( P^*_{fa}(t) )</td>
<td>Normalized prediction error covariance matrix localized to verification region.</td>
</tr>
<tr>
<td>( F^{1/2} )</td>
<td>( M \times M ) matrix that normalizes variables and sets them to zero in regions outside of the verification region.</td>
</tr>
<tr>
<td>( X_i(t) )</td>
<td>( M \times K ) matrix whose ( K ) columns are ensemble perturbations about a control or ensemble mean forecast.</td>
</tr>
<tr>
<td>( A_{fa}(t) )</td>
<td>Estimate of the ( i )th analysis error covariance matrix.</td>
</tr>
<tr>
<td>( K )</td>
<td>Number of ensemble perturbations.</td>
</tr>
<tr>
<td>( C )</td>
<td>( K \times K ) transformation matrix corresponding to the ( i )th possible deployment of observational resources.</td>
</tr>
<tr>
<td>( A_{fa}(t_{fa}) )</td>
<td>Ensemble-based estimate of ( i )th analysis error covariance matrix.</td>
</tr>
<tr>
<td>( Y_i(t) )</td>
<td>( M \times K ) matrix ( Y_i(t) = X_i(t) C ).</td>
</tr>
<tr>
<td>( H_i(t) )</td>
<td>Any ( M \times M ) matrix that can be written in the form ( G^{1/2} Y_i(t) Y_i(t)^T G^{1/2}/K ).</td>
</tr>
<tr>
<td>( t_f )</td>
<td>Time at which generalized error covariance matrix is defined.</td>
</tr>
<tr>
<td>( G^{1/2} )</td>
<td>An ( M \times M ) positive definite matrix that would typically be used to normalize and localize a state vector.</td>
</tr>
<tr>
<td>( b_k )</td>
<td>Orthonormal eigenvectors of the symmetric matrix ( Y_i(t) ).</td>
</tr>
<tr>
<td>( \beta_k )</td>
<td>( K \times K ) diagonal matrix whose ( K ) nonzero elements are ( b_k ).</td>
</tr>
<tr>
<td>( s^*<em>k(t</em>{fa}) )</td>
<td>The ( k )th orthonormal SV of the covariance matrix ( H_i(t) ).</td>
</tr>
<tr>
<td>( s_k(t) )</td>
<td>The dimensionalsized form of the perturbation that evolves into the ( k )th orthonormal SV of the covariance matrix ( H_i(t) ).</td>
</tr>
<tr>
<td>( S^*_k ), ( S_k )</td>
<td>( M \times K ) matrix whose columns are the orthonormal SVs of ( H_i(t) ).</td>
</tr>
</tbody>
</table>

- The ensemble transform technique utilizes linear combinations of \( K \) ensemble perturbations \( X_i(t) \) from an ensemble initialized at the analysis time \( t_a \) to estimate the \( I \) prediction error covariance matrices indicated by Eq. (2). In mathematical terms, \( \mathbf{X}_i(t) \) is an \( M \times K \) matrix whose \( K \) columns are linearly independent ensemble perturbations about a mean or control forecast. The basic strategy of the ensemble transform technique is to find a transformation matrix, \( \mathbf{C} \), such that an ensemble-based estimate of the analysis error covariance matrix at \( t_{fa} \) is given by

\[
A_{fa}(t_{fa}) = \frac{X_i(t_{fa}) \mathbf{C} \mathbf{C}^T X_i(t_{fa})^T}{K},
\]

while the corresponding prediction error covariance matrix at the verification time \( t_v \) is given by

\[
P_{fa}(t_v) = \frac{X_i(t_v) \mathbf{C} \mathbf{C}^T X_i(t_v)^T}{K}.
\]

In cases where the ensemble perturbations are small enough to obey linear dynamics, (4) gives the mapping of the analysis error covariance matrix \( \{3\} \) into the prediction error covariance matrix by the tangent linear propagator around the mean (or control) forecast from \( t_{fa} \) to \( t_v \).

- In finding transformation matrices corresponding to every feasible deployment of the adaptive component of observational resources, we obtain estimates of the prediction error covariance matrices corresponding to every feasible deployment of observational resources. The deployment that minimizes the chosen measure of prediction error is deemed optimal.

- As will be discussed in section 2, the \( i \)th transformation matrix \( \mathbf{C} \) is derived from the \( i \)th estimate or guess \( \mathbf{A}_{fa}(t_{fa}) \) of the analysis error covariance matrix that would result from the \( i \)th possible deployment of observational resources. However, if one substitutes \( \mathbf{C} \) into Eq. (3) one obtains an ensemble-based estimate of the analysis error covariance matrix \( \mathbf{A}_{fa}(t_{fa}) \). Thus, there are two approximations to the analysis error covariance matrix that need to be considered when considering the accurateness of the ensemble transform technique (cf. section 2c).

Since accurate methods of estimating the inverse analysis error covariance matrix are computationally exorbitant (Daley 1991, p. 384), the estimates \( \mathbf{A}_{fa}(t_{fa}) \) are necessarily crude. Nevertheless, qualitatively correct estimates of some aspects of the inverse analysis error covariance matrix either are available or could be made

\footnote{Note that since \( P^*_i(t) \) is symmetric, the eigenvectors of \( P^*_i(t) \) are the same as the singular vectors of \( P^*(t) \) (cf., e.g., Strang 1988).}
available with relatively little effort. For example, the
monthly average values of analysis error covariance can
be reasonably estimated by examining the differences
between independent data assimilation schemes that ac-
count for observational error, for example, Houtekamer
et al. (1996). Also, if we are mainly interested in large-
scale flows, analysis error covariance information can
also be inferred from balance considerations (Daley
1991, p. 167). Furthermore, in regions where no data
is available, the analysis error variance should generally
increase as the spread of an appropriately constructed
ensemble increases. Finally, we may be reasonably cer-
tain that the analysis error variance in some regions will
be reduced by taking supplemental observations in that
region.

Ideally, all of the above considerations should con-
tribute to the estimates \( A_{ei}(t_{fa}) \) used to derive the
transformation matrices \( C_i \). We must accept, however, that
there will be significant quantitative errors in these es-
timates. For the illustrative purposes of this paper, we
will ignore all flow-dependent aspects of the \( A_{ei}(t_{fa}) \) and
simply assume that they are diagonal. The diagonal
elements represent analysis error variances. Note, how-
ever, that the ensemble-based estimate \( A_{ei}(t_{fa}) \) will also tend to give high analysis error variance estimates in regions where the spread of the
basis ensemble perturbations \( X_i(t) \) is large. The accuracy of the ensemble-based estimate \( A_{ei}(t_{fa}) \) of analysis error covariance is critically important because it is from
\( A_{ei}(t_{fa}) \) that our prediction error covariance matrices
\( P_{ei} \) evolve.

In summary, use of the ensemble transform technique
to optimally deploy observational resources involves the
following steps:

1) creation of an ensemble forecast initialized at the
analysis time \( t_{fa} \) and calculation of the perturbations
of the ensemble members about the mean (or control)
forecast;
2) estimation at the future analysis time \( t_{fa} \) of the I
analysis error covariance matrices \( A_{ei}(t_{fa}) \) cor-
responding to the I distinct possible deployments of
observational resources;
3) calculation of the I corresponding transformation
matrices \( C_i \);
4) evaluation of the measure of prediction error for each
possible deployment of observational resources; and
5) implementation of the deployment of observational
resources, which minimizes the measure of prediction
error at \( t_{fa} \).

Some alternative strategies to the ensemble transform
technique for identifying optimal supplementary obser-
vation sites are described in Langland and Rohaly
(1996), Lorenz and Emanuel (1998), and Palmer et al.

Section 2a describes the mathematical basis of the
ensemble transform technique. Section 2b describes the
transformations that yield the singular vectors (SVs) of
ensemble-based estimates of the analysis and prediction
error covariance matrices. In section 2c, the relationship
between the ensemble-based estimate \( A_{ei}(t_{fa}) \) of analysis
error covariance is compared and contrasted with the
estimate \( A_{ei}(t_{fa}) \) from which it is derived.

We discuss in section 3 some calculations that were
used during the recent Fronts and Atlantic Strom Tracks
Experiment (FASTEX) (Joly et al. 1997) to help direct
aircraft to regions where supplemental observations
would improve forecasts over a region just west of the
British Isles. The ensemble used for these calculations
contains ensemble perturbations from both the U.S.
and Canadian national forecast centers. We illustrate the
time continuity of data-sensitive regions and we give
an example of the evolution of a perturbation that
evolves into the leading SV of the ensemble-based es-
imate of the prediction error covariance matrix. Con-
cluding remarks are given in section 4.

2. Theory

a. Defining the transformation matrices \( C_i \)

Let \( Y_{ei}(t) \) denote the \( i \)th set of transformed ensemble
perturbations; that is, let

\[
Y_{ei}(t_{fa}) = X_i(t_{fa}) C_i .
\]

(5)

The essential requirement of this transformation is that
in the ideal case of a perfect guess \( A_{ei}(t_{fa}) \) of the actual
analysis error covariance matrix \( A_{ei}(t_{fa}) \), the ensemble-

based estimate \( A_{ei}(t_{fa}) \) of the analysis error covariance
matrix should equal the actual analysis error covariance
matrix \( A_{ei}(t_{fa}) \) when the number of independent ensemble
perturbations \( K \) is equal to the dimension \( M \) of the state
vector. Mathematically, this requirement is expressed by

\[
A_{ei}(t_{fa}) = \frac{Y_{ei}(t_{fa}) Y_{ei}(t_{fa})^T}{K} = A_{ei}(t_{fa}) = A_{ei}(t_{fa}).
\]

(6)

Although the following theory can be developed with
little alteration in cases where \( A_{ei}(t_{fa}) \) is a rank-deficient
matrix, for the sake of simplicity, here we shall assume
that it is a full rank matrix. First, note that if (6) is
satisfied then the following equation must be satisfied:

\[
Y_{ei}(t_{fa}) [Y_{ei}(t_{fa}) A_{ei}^{-1}(t_{fa}) Y_{ei}(t_{fa})] Y_{ei}(t_{fa})^T = K Y_{ei}(t_{fa}) Y_{ei}(t_{fa})^T .
\]

(7)

Thus, provided \( C_i \) is chosen such that

\[
[C_i X_i(t_{fa}) A_{ei}^{-1}(t_{fa}) X_i(t_{fa}) C_i] = K I,
\]

(8)

where \( I \) is the identity matrix, then (7) is satisfied. Equation
(8) is solved by letting the columns of \( C \) be the \( K \)
orthogonal eigenvectors \( c_a \) of the \( K \times K \) symmetric
matrix \( X_i(t_{fa}) A_{ei}(t_{fa}) X_i(t_{fa}) \) with corresponding eigen-
values \( \gamma_a \) and letting the magnitudes of the \( c_a \) be such
that $c^T_i c_k = K/\gamma_k$ for $\gamma_k > 0$ and $c^T_i c_k = 0$ for $\gamma_k = 0$. Note that since $X_i(t_0) A_i(t_0)^{-1} X_k(t_0)$ is a positive definite matrix, none of the eigenvalues are negative.

If $C_i$ satisfies (8) and the inverse of $X_i(t_0) C_i$ exists, then

$$A_i(t_0)^{-1} = K[C_i X_i(t_0)^T]^{-1} [X_i(t_0) C_i]^{-1},$$

and, hence, (6) is satisfied as may be proven by taking the inverse of both sides of (8).

However, if the number of ensemble perturbations $K$ is less than the dimension $M$ of the state vector, then the inverse of $X_i(t_0) C_i$ will not exist and the equality (6) will not hold. Nevertheless, transforming ensembles such that (8) is satisfied when the inverse of $X_i(t_0) C_i$ does not exist provides approximate solutions that, as $K$ is increased, converge to accurate solutions to (6) in much the same way as finite difference solutions to differential equations converge to accurate solutions as grid spacing is reduced. Furthermore, as is argued in section 2c, in cases where $A_i(t_0)$ contains little or no flow-dependent information, it is possible that the ensemble-based estimate of the analysis error covariance matrix $A_i(t_0)$ might be closer to the actual analysis error covariance matrix $A_i(t_0)$ than the guess of the analysis error covariance matrix, $A_i(t_0)$.

Recall that the point of obtaining the transformation matrices $C_i$ is not to achieve the inherently redundant goal of obtaining an ensemble-based estimate $A_i(t_0)$ of the known estimate $A_i(t_0)$. It is to obtain ensemble-based estimates of the prediction error covariance matrix $P_i(t_0)$ associated with the $i$th possible deployment of adaptive observational resources at $t_0$ via Eq. (4).

If (i) the number of independent ensemble perturbations $K$ are equal to the dimension of the state vector $M$ and (ii) the estimate $A_i(t_0)$ was equal to the actual covariance matrix $A_i(t_0)$, then Eq. (6) would be satisfied. If, in addition, (iii) there were no model error and (iv) errors in the ensemble forecasts remained in the linear regime, then the following equation would hold:

$$P_i(t) = R A_i(t_0) R^T = \frac{R X_i(t_0) C_i C_i^T X_i(t_0) R^T}{K} = \frac{X_i(t_0) C_i C_i^T X_i(t_0)^T}{K} = P_i(t),$$

where $R$ was the linear tangent propagator that mapped the vector space of perturbations at the time $t_0$ to the vector space of perturbations at the later time $t$; that is, the ensemble transform prediction error covariance matrix $P_i(t)$ would be precisely equal to the actual prediction error covariance matrix $P_i(t)$.

### b. Finding the SVs of ensemble-based estimates of error covariance matrices

With good reason, Buizza and Palmer (1995) and Ehrendorfer and Tribbia (1997) and other works on predictability have emphasized that SVs maximize certain measures of growth. In contrast, works on data analysis such as those of Bretherton et al. (1992), Elsner and Tsonis (1996), and Cherry (1996), and textbooks on linear algebra such as Golub and Van Loan (1983) and Strang (1988) emphasize that SVs specifically refer to the orthonormal vectors required to perform a singular value decomposition of a matrix. In the case of a symmetric matrix such as the prediction or analysis error covariance matrix, the SVs are equal to the orthonormal eigenvectors of the matrix. Nevertheless, since one can always find cost functions that SVs maximize, as is discussed at the end of this section, defining SVs in terms of appropriate cost functions can be equivalent to defining them in terms of appropriately normalized matrices. We begin by showing how the SVs of the ensemble-based prediction and analysis error covariance matrices may be obtained via appropriate linear combinations of ensemble perturbations.

Ensemble-based estimates of both the prediction and analysis error covariance matrices may be written in the generalized form

$$G^{-1/2} Y_{e_i}(t_0) Y_{e_i}(t_0)^T G^{-1/2} = H_{e_i}(t_0).$$

To obtain the normalized and localized ensemble-based prediction error covariance matrix $P_{e_i}(t)$ from (11), one would take $G = F$ and $t_f = t_0$. To obtain the ensemble-based analysis error covariance matrix $A_{e_i}(t_0)$ from (11), one would take $G$ to be equal to the identity matrix $I$ and $t_f = t_0$.

Eigenvectors of the general form $H_{e_i}(t_0)$ may be found by determining transformation vectors $b_{e_i}$ such that

$$H_{e_i}(t_0)[G^{-1/2} Y_{e_i}(t_0) b_{e_i}] = \frac{G^{-1/2} Y_{e_i}(t_0) Y_{e_i}(t_0)^T G^{-1/2} Y_{e_i}(t_0) b_{e_i}}{K} = \beta_{e_i}[G^{-1/2} Y_{e_i}(t_0) b_{e_i}],$$

where $\beta_{e_i}$ is the eigenvalue and the bracketed term describes the eigenvector. This is readily achieved by letting $b_{e_i}$ be the orthonormal eigenvectors of the symmetric matrix $Y_{e_i}(t_0)^T G^{-1/2} Y_{e_i}(t_0)/K$ with corresponding eigenvalues $\beta_{e_i}$. To normalize the eigenvectors we note that their magnitude is given by

$$b_{e_i}^T Y_{e_i}(t_0)^T G^{-1} Y_{e_i}(t_0) b_{e_i} = K \beta_{e_i},$$

and, hence, the normalized eigenvectors or SVs of the covariance matrix $H_{e_i}(t_0)$ are given by

$$s_{e_i}(t_0) = (K \beta_{e_i})^{-1/2} G^{-1/2} Y_{e_i}(t_0) b_{e_i}.$$
vector, \( \mathbf{e} \), is given by \( \mathbf{e}^T \mathbf{H}_g(t_f) \mathbf{e} \), one can show from (15) that the direction (or structure) of the SV corresponding to the largest eigenvalue \( \beta_{\max} \) explains more error variance than any other unit vector. The sum of the \( K \) eigenvalues gives the total error variance.

To clarify the relationship between the SVs of the ensemble-based prediction error covariance matrix and the growth-maximizing SVs discussed in Houtekamer (1995), Ehrendorfer and Tribbia (1997), and Palmer et al. (1998), we make the following two observations. First, in the case where \( \mathbf{H}_g(t_f) \) represents the normalized and localized prediction error covariance matrix \( \mathbf{P}^*_g(t_{fa}) \), \( \mathbf{G} = \mathbf{F}, t_f = t_{fa} \), one can obtain the structure from which each SV evolves by evaluating (14) at \( t_{fa} \). The subsequent amplification characteristics of these structures can be inferred by using (5), (8), (13), (14), and the fact that the vectors \( \mathbf{b}_a \) are orthonormal to show that

\[
\frac{\mathbf{b}_a^T \mathbf{Y}_g(t_f) \mathbf{Y}_g(t_f) \mathbf{b}_a (K\beta_{\max})^{-1}}{\mathbf{b}_a^T \mathbf{Y}_g(t_{fa}) \mathbf{A}_g(t_{fa})^{-1} \mathbf{Y}_g(t_{fa}) \mathbf{b}_a (K\beta_{\max})^{-1}} = \frac{\mathbf{s}_a(t_f) \mathbf{F}^{-1} \mathbf{s}_a(t_f)}{\mathbf{s}_a(t_{fa}) \mathbf{A}_g(t_{fa})^{-1} \mathbf{s}_a(t_{fa})} = \beta_{\max},
\]

which gives the dimensional form of the SVs. Thus, the measure of growth, (16), of the structures that evolve into the SVs of \( \mathbf{P}^*_g(t_{fa}) \) is equal to the corresponding eigenvalues of the SVs. Since any vector in the subspace of ensemble perturbations may be written as a linear sum of the SVs, it can be shown that the perturbation in the subspace of ensemble perturbations that maximizes (16) is the SV with the largest eigenvalue \( \beta_{\max} \), that is, the leading SV of the ensemble-based prediction error covariance matrix.

Extending the seminal works of Farrell (1988, 1989), Houtekamer (1995), Ehrendorfer and Tribbia (1997), and Palmer et al. (1998) have noted that the perturbation that maximizes the right-hand side of (16) is the eigenvector \( \mathbf{s} \) corresponding to the largest eigenvalue \( \beta_{\max} \) of the generalized eigenvector problem

\[
(\mathbf{R}(t_{fa}; t_{fa})^{1/2})(\mathbf{F}^{-1} \mathbf{R}(t_{fa}; t_{fa})) \mathbf{s}(t_{fa}) = \beta_{\max} \mathbf{A}_g(t_{fa})^{-1} \mathbf{s}(t_{fa}),
\]

where \( \mathbf{R}(t_{fa}; t_{fa}) \) is the tangent linear propagator that maps the vector space of perturbations at \( t_{fa} \) to the vector space of perturbations at \( t_{fa} \) such that

\[
\mathbf{R}(t_{fa}; t_{fa}) \mathbf{s}(t_{fa}) = \mathbf{s}(t_{fa}).
\]

Generally, the eigenvectors of (18) would not be restricted to the subspace of ensemble perturbations and, hence, they would not be equivalent to the \( \mathbf{s}_a(t) \) described by (17) and (14). However, this alternative procedure would yield equivalent eigenvectors to the \( \mathbf{s}_a(t) \) described by (17) and (14) if, in Eq. (18), \( \mathbf{A}_g(t_{fa})^{-1} \) were replaced by the pseudoinverse of \( \mathbf{A}_g(t_{fa})^{-1} \).

The above discussion makes it clear that the ensemble transform technique could be interpreted as an SV-based technique for optimally deploying adaptive observational resources. However, while the SV-based technique discussed by Palmer et al. (1998) uses just one set of SVs that are somewhat consistent with the standard observational network, the ensemble transform technique calculates \( I \) different sets of SVs that are somewhat consistent with the \( I \) different possible deployments of adaptive observational resources. It is interesting to note that a set of the SVs of Palmer et al. (1998) could be used as the basis ensemble perturbations \( \mathbf{X}_i(t) \) for the ensemble transform technique. This would provide a rational framework for adding known information about the effect of different possible deployments of observational resources on analysis error to the growth information contained in Palmer et al.’s SVs. Hopefully, tests of this combined approach will be performed in the not too distant future.

c. The characteristics of \( \mathbf{A}_g(t_{fa}) \) and \( \mathbf{A}_u(t_{fa}) \)

Current estimates of the diagonal of the analysis error covariance matrix are likely to be superior to estimates of the off-diagonal components. The diagonal components give the variance of analysis error at individual grid points. In regions where many observations are taken, these variances have a relatively simple relationship to the accuracy and density of surrounding observations. In data-sparse regions, the accuracy of the first guess field also plays a significant role in determining the accuracy of the analysis error variance. In principle, the uncertainty of this first guess field could be predicted reasonably accurately by the spread of an appropriately configured ensemble forecast.

The off-diagonal components of the analysis error covariance matrix contain information about vertical and horizontal correlations between variables and also information about typical balances between wind and mass fields. These covariances are highly dependent on the background flow field. For example, near the highly curved flows of intense cyclones, the horizontal correlations and balance constraints are very different from what they are near the quasi-two-dimensional flows of cold fronts. Similarly, in regions of moist convective instability, vertical correlations and balance constraints are very different from what they are in convectively stable regions. Information about these covariances is contained in the differences between ensemble perturbations. This suggests that an effective strategy for estimating analysis error covariance would involve letting the diagonal of the analysis error covariance matrix be informed by both the distribution of observational resources and the ensemble spread and letting the off-diagonal components be informed by some outer product of the ensemble perturbations.

The method of estimating analysis error variance with an ensemble described in section 2a provides a means...
of achieving this goal. Information about how the geographical distribution of observational resources in the ith possible deployment of adaptive observational resources affects analysis error variance and information about the effect of ensemble spread on analysis error variance is easily incorporated into the diagonal of $A_{ei}(t_n)$. insisting that the transformed ensemble perturbations satisfy (8) for each observational network ensures that the transformed ensemble perturbations are informed by the distribution of observational resources. Using the outer product of the transformed ensemble perturbations to estimate analysis error covariance matrices [cf. Eqs. (3) and (5)] then ensures that the approximate covariance matrix $A_{gi}(t_n)$ is not only informed by the analysis error variance information contained in $A_{ei}(t_n)$, but also by the flow-dependent anisotropic error covariance information that is implicitly contained in the ensemble perturbations.

The weakest aspect of ensemble-based estimates of error covariance matrices is the fact that their rank is typically many orders of magnitude lower than the rank of the actual analysis error covariance matrix. Nevertheless, the successful preliminary work of Evensen (1994) and Houtekamer and Mitchell (1998) on ensemble Kalman filters suggests that low-rank flow-dependent error covariance matrices contain more useful information than higher-rank error covariance matrices that make no attempt to estimate the flow-dependent aspects of error covariance matrices. This idea is also supported by the work of Barkmeijer (1995) who demonstrated that a very limited set of ensemble perturbations can account for a large percentage of the structure of localized prediction error variance, and also by that of Lord (1996), who showed that the impact on forecasts for the United States of supplemental observations over the Pacific could be reasonably interpreted in terms of individual members of the 14-member National Centers for Environmental Prediction (NCEP) ensemble. This work leads us to speculate that, in some respects, the low-rank ensemble-based estimate of analysis error covariance $A_{gi}(t_n)$ may be superior to the higher-rank estimate $A_{gi}(t_n)$ from which it was derived. Further research will be required to properly address this issue.

Inevitably, the differences between the $I$ different ensemble-based estimates of analysis error covariance will be in the directions of the ensemble perturbations at $t_n$. Ideally, these differences should mimic the differences in the $I$ different analyses that would result from the $I$ different possible deployments of observational resources. The structure of these differences will depend on the type of data assimilation scheme used. The optimal but impractical data assimilation scheme is the Kalman–Bucy filter (cf. Daley 1991) or a 4D Var equivalent (Pires et al. 1996). Particularly in data-sparse regions, the large analysis increments produced by the Kalman–Bucy filter tend to occur in the directions of the leading eigenvectors of the prediction error covariance matrix.

Since well-constructed ensemble perturbations have a significant projection onto the leading eigenvectors of the prediction error covariance matrix, we anticipate that in data-sparse regions the ensemble transform technique should give a reasonable indication of how changing an observational network would change the accuracy of forecasts initialized with a sophisticated data assimilation scheme.

Another consequence of describing the analysis error covariance matrix in terms of ensemble perturbations is that regions of relatively large ensemble spread tend to be treated as regions where relatively large analysis error is likely. To the extent that the differences between the ensemble members at the initialization time $t_n$ accurately reflect the uncertainty in the initialization, this is a useful feature when the future analysis time $t_n$ corresponds to the analysis cycle immediately following $t_n$. Further research will be required to determine whether this is a helpful or detrimental feature when $t_n$ falls a number of analysis cycles after $t_n$.

3. Production of rms prediction error maps during FASTEX

Our primary purpose in this section is to illustrate how the theory described in the previous section can be used to direct dropsonde-equipped aircraft to regions where supplementary data is likely to improve the accuracy of a specific measure of forecast accuracy. Assessments of the skill of the ensemble transform technique and assessments of how its predictions compare with other methodologies will not be attempted here. Such studies are in progress and will be described in subsequent publications.

The ensemble used for the examples shown in this paper comprises eight independent members from NCEP and nine independent members from the Canadian Recherche en Prévision Numérique (RPN) ensemble. As is described below, the inverse analysis error covariance matrix $A_{gi}^{-1}(t_n)$ used for this demonstration is based on the spread of the RPN ensemble.

a. The ensemble

The NCEP ensemble perturbations are created using the breeding method and are called bred perturbations. The initial states of these perturbations are obtained by multiplying the difference between two perturbed non-linear 24-h ensemble forecasts by a smoothly varying function. This smoothly varying function gives the ensemble perturbations a geographical variation that mimics the geographical distribution of estimated analysis errors. In addition to reducing the magnitude of the ensemble perturbations, the smoothly varying function

\footnote{Note that these estimates are very crude.}
constrains the initial magnitudes of the bred vectors to be consistent with the typical variation in magnitude of analysis errors between continents and oceans. Details of the method are described in Toth and Kalnay (1993, 1997).

The RPN ensemble perturbations are created by running eight perturbed observation/assimilation/forecast cycles as well as one unperturbed control observation/assimilation/forecast cycle. To obtain an acceptable variance in the ensemble, the perturbed forecasts are obtained by adding noise to the observations that is up to 80% larger than estimated magnitudes of observational error. Only four of the ensemble members use surface observations of humidity in mountainous regions; otherwise, the same data assimilation scheme is applied to all the members of the ensemble. Different numerical models are realized by using a variety of parameterization schemes that are of a similar quality. Note that although there are some differences between the numerical models used to integrate the members of the RPN ensemble, the models probably bear more resemblance to each other than they do to the atmosphere. This is part of the reason that excessively large noise has to be added to the observations in order to obtain an adequate variance. In any case, the initial states of the RPN ensemble represent nine independent, reasonably accurate estimates of the state of the atmosphere. As such, the differences between the members of this ensemble at the initial time provide a reasonable estimate of the geographical distribution and spatial scale of analysis error on any particular day. The construction of the RPN ensemble is described in detail in Houtekamer et al. (1996).

In order to give equal weight to members of the RPN and NCEP ensemble when calculating the mean of the combined ensemble, the following normalization procedure was performed. First, the means and standard deviations of the streamfunction fields of the RPN and NCEP ensembles at the initialization time were calculated. Second, the amplitudes of the NCEP perturbations were adjusted at the initialization time so that the standard deviation of the NCEP streamfunction perturbations about the NCEP streamfunction mean was equal to the standard deviation of the RPN streamfunction perturbations about the RPN streamfunction mean. These scaled perturbations were then added to the NCEP mean to create a new normalized NCEP ensemble. The mean and perturbations of the ensemble comprising the RPN members and the normalized NCEP members were then calculated. Note that one may only obtain 16 independent perturbations from the 17-member ensemble.

b. The diagonal inverse analysis error covariance matrix \( \mathbf{A}^{-1}_{gi}(t_{fa}) \)

The normalized measure of analysis error implicit in (8) is \( x(t_{fa})^{T} \mathbf{A}^{-1}_{gi}(t_{fa}) x(t_{fa}) \), where \( x(t_{fa}) \) is an analysis error vector at the future analysis time. The transformation matrix \( \mathbf{C} \) ensures that all of the transformed perturbations have the same normalized size at the analysis time. The NCEP and RPN perturbations are produced by spectral numerical models that provide isotropic horizontal resolution. For display purposes, the model fields are stored on \( 2.5^\circ \times 2.5^\circ \) latitude–longitude grids. Such grids place a disproportionate number of points at the poles. Since the size of the contribution that a region makes to \( x(t_{fa})^{T} \mathbf{A}^{-1}_{gi}(t_{fa}) x(t_{fa}) \) is proportional to the density of grid points in the region, the use of such a grid would make reductions in analysis error over a certain region over the poles have a much larger effect on the measure of analysis error than a similar reduction in analysis error over the equator. To avoid this problem, we interpolated values from a \( 2.5^\circ \times 2.5^\circ \) latitude–longitude grid onto an equal area grid. The equal area grid was obtained by keeping zonal grid steps constant while letting meridional grid steps be proportional to the secant of latitude. The area of each grid box was approximately equal to \((250 \text{ km})^2\).

Another assumption we made concerned the size of the inverse analysis error covariance matrix \( \mathbf{A}^{-1}_{gi}(t_{fa}) \). In principle, this matrix can be used to constrain the amplitude of any analyzed variable, including “derived” variables such as radiance fields. For the sake of simplicity, here we chose to use a reduced inverse analysis error covariance matrix that only constrained the magnitudes of vorticity and streamfunction at 850, 500, and 250 mb over the Northern Hemisphere. We chose these two variables because the processes by which these fields evolve are well understood, because forecasters are familiar with these fields, and particularly because vorticity is a relatively easy field to track through time. Preliminary tests, to see if constraining a different set of quantities such as wind and temperature or just the two horizontal components of the wind would markedly change the location of the observation site that minimizes prediction error variance, have been performed. Generally, constraining vorticity and streamfunction gave optimal sites within \( 5^\circ \) of the sites obtained when temperature and wind were constrained. Detailed results of these tests will be presented in subsequent publications.

Since only Northern Hemisphere vorticity and streamfunction were to be constrained, we were able to work with reduced state vectors whose dimension \( L \) was equal to the sum of the number of Northern Hemisphere vorticity and streamfunction variables at 850, 500, and 250 mb [i.e., \( L \) equals the number of latitudinal grid levels in the Northern Hemisphere (23) multiplied by the number of longitudinal grid levels (144) multiplied by the number of vertical levels (3) multiplied by the number of variables (2); i.e., \( L = 19,872 \)].

The \( L \) diagonal elements of \( \mathbf{A}^{-1}_{gi}(t_{fa}) \) should represent the inverse of the analysis error variance of a particular variable at a particular grid point. The analysis error variance associated with an adaptive observational network may be viewed as being equal to the sum of a
The "background" analysis error variance associated with the standard observational network and a change in analysis error variance due to supplemental observations. This change in analysis error variance only affects variables in the vicinity of the site at which the supplemental observations are taken.

Although the variance of the RPN ensemble members about the mean of the RPN ensemble at the initialization time \( t_o \) provides an estimate of the analysis error variance associated with the background observational network, with only nine members it would generally give a far from perfect representation of the analysis error variance. Notably, if more perturbations were added to the RPN ensemble there would be less chance of extremely high or extremely low values of variance occurring. To ameliorate this problem, we first smoothed the RPN ensemble members with a diffusion-type filter. The filtered fields were rescaled to ensure that the globally averaged variance of the smoothed fields on constant pressure surfaces was the same as the corresponding variance of the unsmoothed fields. Once this had been done, the background analysis error variance of a gridpoint variable was taken to be a weighted sum of the gridpoint variance of the smoothed variable and the globally averaged isobaric variance. Details are given in appendix A.

Generally, analysis error variance is reduced in regions where supplemental observations are taken. The size and shape of the area affected by supplementary observation depends on the types of observing platforms available to take the supplementary observations. For example, two aircraft can take supplementary observations over a larger area than one. For simplicity, we assumed that supplementary observations would reduce analysis error in vorticity and streamfunction at 850, 500, and 250 mb over nine adjacent grid boxes that cover a rectangular (750 km)\(^2\) region.

The amount by which analysis error variance is reduced depends on factors such as the number, accuracy, and spacing of the observations taken, the data assimilation scheme, and also on the size of the background analysis error variance. If the background analysis error variance were small, as it would be over data-rich regions, then the addition of extra observations would not reduce the analysis error variance by very much. Conversely, if the background analysis error variance were large, as it typically would be over data-sparse regions, then supplemental observations would reduce the analysis error variance by a larger amount. To crudely represent these facts, for this illustrative example, we decided to reduce the variance of the variables in question to half their globally and isobarically averaged values over the (750 km)\(^2\) region of supplementary observations (cf. appendix A).

The \( J \) different observational networks were generated by placing the center of the supplemental observation region in \( J \) different locations. These locations were placed at each latitudinal grid point and every second longitudinal grid point of the equal area grid. These spacings corresponded to a 5° longitudinal spacing and a 3°–5° spacing in the latitudinal direction. The location of the center of the optimal supplemental observation region can only be determined to an accuracy equal to this spacing. Note that the resolution of the observational network is independent of the resolution of the ensemble perturbations.

The ensemble perturbations used for this illustration were run at T63/T62 truncation, which roughly corresponds to a 250 \( \times \) 250 km equal area grid.

Finally, it is worth noting that experience to date indicates that, with ensembles containing less than 25 independent members, the qualitative results from the ensemble transform technique are not particularly sensitive to the way in which \( \mathbf{A}_w(t_o) \) is defined. For example, if we choose to hold the elements corresponding to variances of streamfunction and vorticity at their global average outside the supplemental observation region and a smaller constant inside the observations region, we obtain similar results to those presented here. More information on these issues will be presented in subsequent publications.

c. Rms prediction error as a function of location of supplemental observation sites

In order to assess the effect of supplemental observations on prediction error, we must choose some sort of measure of prediction error. As discussed in the introduction, a wide variety of measures of prediction error may be derived from the normalized and localized prediction error covariance matrix \( \mathbf{P}_n^q(t) \). Arguably, the types of prediction error that one would most like to avoid are large and/or likely prediction errors. In principle, the leading SVs of \( \mathbf{P}_n^q(t) \) correspond to large and/or likely prediction errors. For this reason, we chose to measure prediction error by the square root of the prediction error variance explained by the first five SVs of \( \mathbf{P}_n^q(t) \); that is, the square root of the sum of the first five eigenvalues of \( \mathbf{P}_n^q(t) \) was taken to be our measure of prediction error. We shall subsequently refer to this measure of prediction error as the rms prediction error.

To produce the normalized and localized prediction error covariance matrix, \( \mathbf{F}^{1/2} \) was chosen so that the prediction error variance was proportional to the sum of the squares of normalized gridpoint vorticity and streamfunction at 850, 500, and 250 mb over a 1000-km radius disk centered in the northeastern Atlantic (see appendix B for details). Note that since the magnitude of streamfunction perturbations is roughly 10 orders of magnitude larger than vorticity perturbations, it is essential that these variables be normalized before including them in some measure of prediction error. Without such normalization, the measure of forecast error would be insensitive to forecast errors in the vorticity field.

By calculating the rms prediction error for each pos-
sible location of the supplemental observation area, we generated contour maps of rms prediction error as a function of the location of the center of the supplemental observation site (e.g., Fig. 2). Maps like these were used during FASTEX to construct flight plans for radiosonde-equipped aircraft.

d. Temporal continuity of rms prediction error minima

Analysis error can project onto a wide variety of wave types whose group velocities are very different. For example, Eady’s (1949) idealized model of midlatitude flow contains nonexponentially amplifying waves whose group velocity is equal to the mean geostrophic wind at the surface, exponentially amplifying waves whose group velocity is equal to the mean midlevel geostrophic wind, and still other nonexponentially amplifying waves whose group velocity is equal to the mean geostrophic wind at the tropopause. Idealized equivalent barotropic models of midlatitudes show that the energy associated with large-scale waves can move faster than the speed of the wind at any level, whereas the energy associated with small-scale waves is about the same as the mean zonal wind. Given such a wide range of group velocities, one would generally expect that 2-day forecasts would be sensitive to the quality of data over a larger region than 1-day forecasts. Furthermore, one would generally expect the region of sensitivity for the 2-day forecast to be further upstream from the verification region than the region of data sensitivity for the 1-day forecast.

In the 20 cases examined in preparation for this paper, it was found that the region of sensitivity approached the verification region as the future analysis time \( t_f \) approached the forecast verification time \( t_v \). Although the size of the data-sensitive region generally reduced as the target analysis time \( t_a \) approached the forecast verification time \( t_v \), this did not occur in the case to be discussed in the following. An explanation of why this sometimes occurs and sometimes does not occur is beyond the scope of this paper.

Figure 2 illustrates a case that shows how the region of sensitivity approached the verification region as \( t_f \) approached the forecast verification time \( t_v = 1200 \) UTC on 15 February 1997. Figure 2a indicates that if supplemental observations were to be taken 48 h after the initialization time \( t_i = 0000 \) UTC 11 February of the ensemble forecast (note that this is 60 h before the verification time), then the optimal site for supplementary observations would be centered over 275°E and 47.7°N.

Thus, in spite of the fact that the standard observational network samples this region relatively well, our relatively crude implementation of the ensemble transform technique indicates that it would still be worthwhile to take extra observations in this region. Our crude implementation of the technique makes this prediction because our estimate of the analysis error variance that would result from taking supplemental observations in this region is smaller than our estimate of the analysis error variance associated with the standard observational network. If more accurate estimates of the analysis error variances associated with the standard supplemented observational networks had been used, one might expect to find an rms prediction error minimum less deep than that obtained with the crude analysis error variance estimates used for the plot shown in Fig. 2a.

The 500-mb ensemble mean vorticity field for this chart shows that this region of sensitivity corresponds to an arm of anomalously high vorticity. A secondary minimum in prediction error variance appears to correspond to a supplementary observation site centered at 44.1°N, 315°E. This minimum corresponds to a weakly “bull’s-eye” of ensemble mean vorticity.

From the 500-mb ensemble mean vorticity and streamfunction fields on these charts, one may track the “arm” and bull’s-eye of 500-mb vorticity through to the verification time 108 h after the initialization time \( t_f \) of the ensemble forecast. Our subjective attempts to do this indicate that the arm of anomalous vorticity ends up on the eastern side of the verification region at the verification time while the bull’s-eye of vorticity appears to end up being absorbed in a large region of positive vorticity on the western side of the verification region (cf. Fig. 2e). Thus, error modes that originate near the arm of vorticity and that move at the same speed or slightly faster than the arm of vorticity will have an impact on the verification region. Similarly, error modes that originate near the bull’s-eye and that move with or slightly slower than the bull’s-eye will affect the verification region.

At those future analysis times more than 60 h after the initialization time \( t_f \), the impact of supplementary observations over the bull’s-eye of vorticity is very small. Presumably, this is because the modes emanating from the bull’s-eye grow at a slower rate after this time. Figures 2c–2e further illustrate that, as was anticipated, the optimal supplementary observation site approaches the verification region as the future analysis time \( t_f \) approaches the verification time \( t_v \).

The total scaled prediction error variance associated with the various possible locations of the supplementary observation site may be obtained by adding the prediction error variance explained by the leading five SVs to the prediction error variance explained by the trailing SVs. When the future analysis time precedes the verification time by 60 h (Fig. 2a) the scaled prediction error variance corresponding to observation sites away from the data-sensitive region is about 190, and about 90% of this variance is explained by the five leading SVs. As the target analysis time is brought closer to the verification time there is a gradual decrease in the magnitude of the scaled prediction error variance and also in the percentage of variance explained by the five leading SVs. When the target analysis time precedes the verification time by just 12 h, the scaled prediction error...
variance corresponding to observation sites away from the data-sensitive region is about 50, and about 80% of this variance is explained by the five leading SVs.

e. The structure of the leading SVs

As discussed in section 2b, there are two ways of constructing SVs. One method involves the construction of SVs using a linear tangent model, its adjoint, and a Lanczos algorithm together with some approximation to the inverse analysis error covariance matrix (Buizza and Palmer 1995). This approach can identify structures that explain more prediction error variance than any other structure in the tangent linear regime (Houtekamer 1995; Ehrendorfer and Tribbia 1997). When the energy-based norm is used in place of estimates of the analysis error covariance matrix, Buizza and Palmer (1995) find that important SVs are frequently localized in the hor-
Fig. 2. (Continued) (c) As in Fig. 2a but here \( t_f \) is 72 h into the forecast at 0000 UTC 14 Feb 1997. The optimal supplementary observation site corresponds to a minimum of 9.31 and is located at 44.8°N, 30.5°E.

(d) As in Fig. 2a but here \( t_f \) is 84 h into the forecast at 1200 UTC 14 Feb 1997. The optimal supplementary observation site corresponds to a minimum of 8.12 and is located at 31.5°E, 44°N.

Horizontal at the initial time with maximum amplitudes in the lower troposphere, whereas at the verification time \( t_v \), peak amplitudes are in the upper troposphere.

The ensemble transform technique provides another way of creating SVs. The leading SV produced by the ensemble transform technique represents the structure that accounts for more of the ensemble-based prediction error variance estimate than any other structure.

It is of interest to compare the structure of a leading SV of our ensemble-based prediction error covariance matrix with the SV structures found by Buizza and Palmer (1995). Figure 3a illustrates the initial streamfunction and vorticity structure of the SV, which explains the largest part of the prediction error variance associated with the observational network that includes a \((750 \text{ km})^2\) region of enhanced observational scrutiny centered at 44.1°N, 305°E at 0000 UTC on 14 February 1997. This location corresponds to the optimal supplementary observation site indicated by Fig. 2c. Figure 3b shows the streamfunction structure 36 h later at the optimization (verification) time. We note that this SV has a few similar characteristics to the structures found by Buizza and Palmer (1995). Initially (at \( t_f \)) it is relatively localized in the horizontal, with maximum amplitude in the mid-to lower troposphere, whereas at the verification time \( t_v \), peak amplitudes are in the upper troposphere. We also note that, in the near vicinity of the optimal observation site, the initial 500-mb vorticity field is about a quarter of a wavelength to the west (upshear) of the 850-mb vorticity field. This type of vertical tilting is somewhat similar to the tilts of the “optimal” perturbations discussed in Farrell (1989).
FIG. 2. (Continued) (e) As in Fig. 2a but here $t_a$ is 96 h into the forecast at 0000 UTC 15 Feb 1997. The optimal supplementary observation site corresponds to a minimum of 6.78 and is located at 48°N, 325°E. Unlike Figs. 2a–d, the bottom panel of this figure shows the vorticity and streamfunction fields of the ensemble mean at the verification time 108 h into the forecast at 1200 UTC 15 Feb 1997.

The apparent lessening of the low-level phase tilt with time is reminiscent of the nonmodal development also discussed by Farrell (1984).

Although the leading ensemble transform SV for the 72–108-h time interval exhibits many of the features identified by Buizza and Palmer (1995), note that the leading ensemble transform SV for the 48–108-h time interval shown in Fig. 4 does not. In particular, the leading ensemble transform SV for this time interval is not localized in the horizontal. This SV is so spread out that it would be very difficult to choose a supplementary observation site based on the amplitude of the SV. In contrast, the corresponding map of rms prediction error as a function of supplementary observation site shown in Fig. 2a clearly identifies a localized region near the North American Great Lakes as the optimal supplementary observation site. This example demonstrates that the ensemble transform technique can extract information from ensembles that is not apparent in the superficial structure of individual ensemble transform SVs. Such information may, however, be apparent in the structure of the SVs of the type described by Palmer et al. (1988).

4. Concluding remarks

The ensemble transform technique for identifying optimal observing networks is applicable to any system for which ensemble forecasts are available. Here, we have explored a meteorological application of the technique. It is based on the idea that the prediction error variance associated with any distribution of observational resources may be estimated via an appropriate transformation of an ensemble forecast. We have shown that the ensemble transform technique would precisely recover the prediction error covariance matrices associated with each possible deployment of observational resources provided that (i) estimates of the analysis error covariance matrix were precise, (ii) the ensemble perturbations span the vector space of all possible perturbations, and (iii) the evolution of errors were linear and perfectly modeled. When the above ideal conditions are not met, the ensemble transform approach provides a rational framework for combining informed guesses about the analysis error covariance matrix associated with different possible deployments of observational resources with useful error covariance and growth information contained within the evolving ensemble perturbations.

An interesting aspect of the ensemble transform technique is that it may be directly applied to ensembles of nonlinear forecasts and their corresponding nonlinear ensemble perturbations. On the one hand, as the ensemble transform approach to estimating prediction error variance relies on the assumption that linear combinations of perturbations evolve in accord with the governing equations, one would expect a degradation in the efficacy of the ensemble transform technique as perturbations grew to nonlinear amplitudes. On the other hand, nonlinear ensemble perturbations may simultaneously contain small-scale features that rapidly grow, saturate, and decay through nonlinear processes together with less rapidly growing larger-scale features whose dynamics remain linear over a longer period. These considerations suggest that the ensemble transform technique may continue to give reliable estimates of the prediction error variance associated with larger scales even when nonlinear processes are occurring at smaller scales.

Conveniently, the ensemble transform technique’s representation of prediction error variance in terms of nonlinear ensemble perturbations means that rapidly growing features that nonlinearly saturate at small amplitudes, such as localized thunderstorms, will not make
erroneously large contributions to estimates of prediction error variance. Purely linear models of prediction error growth cannot account for the slowing of prediction error growth by nonlinear processes.

Another interesting aspect of the SVs produced by ensemble transformation is that they are as “balanced” as the perturbations from which they are derived while their dynamics remains linear. To see that this is true, consider the following. Suppose all of the ensemble members satisfy some balance equation. If the ensemble perturbation amplitudes were small enough to be governed by linear dynamics, all of the ensemble perturbations would satisfy a linearization of the same balance equation about the ensemble mean. Furthermore, any sum of ensemble perturbations would satisfy this linearized balance equation. Since the SVs produced by ensemble transformation are just linear sums of ensemble perturbations, they too must satisfy the linearized balance equation.

In contrast, the balance of the energy norm–based SVs, which are the basis of the targeting technique suggested by Palmer et al. (1998), is relatively unconstrained. These SVs identify the structures that maximize perturbation energy growth. Thus, depending on the background flow and the resolution of the model, the SVs may be in some sort of partial geostrophic balance or they may be far from geostrophic balance. Szunyogh et al. (1997) found that the energy norm–based SVs produced by the linear tangent of a low-order GCM and its adjoint were abnormally far from geostrophic balance. Thorpe (1998) examined the balance of an SV of the higher-resolution model used in operations at the European Centre for Medium-Range Weather Forecasts (ECMWF). They found that the ratio of the mean magnitude of divergence over the mean magnitude of vorticity was 0.7 for the SV and 0.3 for the time-evolving basic state; in other words, the flow in the SV appeared to be significantly more out of balance than that in the regular forecast model.

There are a number of ways in which the ensemble transform technique can be tested. One method is to send dropwindsonde-equipped planes to the optimal supplementary observation sites identified by the ensemble transform technique. One then compares forecasts from analyses that do not use the supplementary data with forecasts that do use the supplementary data. If the difference field between the two forecasts reaches the verification region at the verification time, one may conclude that the ensemble transform technique correctly identified an upstream region that was dynamically linked to the verification region. On average one should expect that such well-placed observations would improve forecasts over the verification region.\(^4\) Tests like these were performed over the North Atlantic during FASTEX (cf. Szunyogh et al. 1998, manuscript submitted to Mon. Wea. Rev.) and over the North Pacific during the North Pacific Experiment (NORPEX 98) (T. Szunyogh 1998, personal communication). These experiments have confirmed that the ensemble transform technique is capable of identifying upstream regions that are dynamically connected to the verification region with 7 linearly independent members from the NCEP ensemble and even more reliably with 25 linearly independent members from the ECMWF ensemble (cf. Toth and Szunyogh 1997).

Note that the effectiveness of the ensemble transform technique depends on both the size and type of ensemble forecast used. Experience to date indicates that the results of the ensemble transform technique significantly improve as the number of ensemble perturbations is increased. In the first 48 h after initialization the ECMWF, RPN, and NCEP perturbations all have very different structures. Consequently, in this early period, the use of different ensembles in the ensemble transform technique is more likely to lead to different conclusions about the optimal deployment of observations than at later times. A detailed analysis of these dependencies remains a subject for future research.

A cheaper test of the ensemble transform technique similar to the type of test described above may be performed by deliberately degrading the analysis over the optimal supplementary observation site. One then checks to see whether the difference field between a forecast made from the degraded analysis and a forecast made from the standard analysis reaches the verification region at the verification time. Furthermore, one can use this approach to check whether the optimal supplementary observation site has a stronger dynamical connection with the verification region than sites surrounding it by degrading the analysis in regions just outside of the optimal supplementary observation site. Results of tests of this type are equivocal, however, because it is difficult to ensure that equivalent types of degradations to the analysis are made inside and outside the optimal supplementary observation site. Tests of this kind have already been made at NCEP (T. Szunyogh 1997, personal communication) and have indicated that the ensemble transform technique is indeed capable of identifying the upstream region with the strongest dynamical link to the verification region at the verification time.

Finally, the cheapest and perhaps most rigorous way of testing the ensemble transform technique is to write a Kalman–Bucy filter for a simplified model so that the actual evolution of the prediction and analysis error covariance matrices can be obtained. Within such simple models, one could test how many and what sort of ensemble perturbations would be required to accurately predict the prediction error covariance matrices associated with I possible deployments of observational resources. Tests of this sort are planned for the future.

With the 17-member ensemble used in this paper, the
ensemble transform technique allows the prediction error variance associated with different deployments of observational resources to be assessed very rapidly. On a Silicon Graphics (SGI) workstation, it took less than 10 min to calculate the prediction error variance associated with 294 distinct observational networks. (The required eigenvector/eigenvalue calculations were performed by the IMSL routine DEVCSF). Thus, the prediction error variance associated with each observational network was obtained in about 2.0 s. Importantly, the prediction error variance calculations associated with each observational network are completely independent of each other and are trivial to run in parallel. For example, 10 SGI workstations could have estimated the prediction error variance associated with 294 distinct observational networks in less than 1 min. For larger ensembles, the most time-consuming part of the operation will be the eigenvector/eigenvalue calculations. With nonvectorized code the time spent on these calculations is approximately proportional to the cube of the number of ensemble members. With a vectorized eigenvector/eigenvalue routine, the time spent is approximately proportional to the square of the number of ensemble members. Note that, if needed, further computational cost savings could be obtained by only calculating the most rapidly growing SVs and/or utilizing recent advances in parallel processing.

The number of elements in the observational network that could be varied on a day-to-day basis is enormous. One could increase (decrease) the number of daily radiosonde launches at radiosonde stations in data-sensitive (insensitive) regions. One could reduce the electric power requirements of space-based lidar by only taking measurements over data-sensitive regions. One could have a global fleet of pilotless planes sampling the world’s ocean basins in data-sensitive regions. To successfully manage such an adaptive network, one needs to be able to assess the relative merits of a very large number of possible sites.

Fig. 3. (a) The initial streamfunction and vorticity structure of the SV that explains the largest part of the estimate of prediction error variance of a 36-h forecast for 1200 UTC 15 Feb 1997. This particular estimate of prediction error variance assumes that supplementary observations are to be taken at the optimal site whose center is located at 305°E, 44°N (cf. Fig. 2c).
deployments of observational resources in a timely manner. We offer the ensemble transform technique as a possible means of meeting this need.

Acknowledgments. This work was initiated while Craig Bishop was employed by the Universities Space Research Association in the Mesoscale Processes Branch of the Goddard Space Flight Facility. The authors would like to thank Robert Adler and Michael Kalb for their support during this period. Since August of 1996, Craig Bishop’s research efforts have been supported by NSF Grant ATM 96-12502. The authors would also like to thank Eugenia Kalnay for her encouragement and enthusiasm and for making a 2-month work sojourn at NCEP possible for Craig Bishop. Thanks also to Pierre Dubreuil of the Division de Recherche en Prévision Numérique (RPN) for granting us permission to use the RPN ensemble and to Peter Hou-tekamer and Louis Lefaivre for making it happen. The development of the ideas in this paper was assisted and inspired by conversations with Tim DelSole, Martin Ehrendorfer, Kerry Emanuel, Peter Houtekamer, Eugenia Kalnay, Sharanya Majumdar, James Purser, Chris Snyder, Istvan Szunyogh, Milja Zupanski, and two anonymous referees.

APPENDIX A

Using the RPN Ensemble to Construct a Diagonal \( \mathbf{A}_{\mathbf{g}}(t_{\mathbf{m}}) \)

In order to avoid spuriously large variations of the background analysis error variance, we used a diffusion-type smoother on all fields in the RPN ensemble and then defined the background analysis error variance \( \overline{u}_{qr}^2 \) of some variable \( u \) at the \( r \)th grid point on the \( q \)th pressure level with the equation

\[
\overline{u}_{qr}^2 = a \left( \frac{1}{RK} \sum_{r=1}^{K} \sum_{i=1}^{R} (u_{iqr} - \overline{u}_q)^2 \right) + (2 - a) \left( \frac{1}{K} \sum_{i=1}^{K} (u_{iqr} - \overline{u}_q)^2 \right). \tag{A1}
\]

Here \( u_{iqr} \) refers to the smoothed value of \( u \) in the \( k \)th ensemble member at the \( r \)th grid point at the \( q \)th pressure level at the initial time of the ensemble forecast; \( R \) gives the total number of grid points on the \( q \)th pressure level; \( \overline{u}_q \) is the average value of \( u \) over all the grid points and all the ensemble members on the \( q \)th pressure level; and \( a \) is a constant whose value lies between 0 and 2. The first term enclosed in large brackets in (A1) gives the ensemble variance of \( u \) on the \( q \)th pressure level averaged over the whole globe. The second term in large brackets in (A1) gives the ensemble variance on the \( q \)th pressure level of the smoothed \( u \) at the \( r \)th grid point. Thus, Eq. (A1) defines the analysis error variance of a particular variable at a particular grid point to be a weighted sum of the gridpoint variance of the smoothed ensemble and the isobaric globally averaged variance. For FASTEX, we took \( a = 1.5 \). This choice weights the global average of variance more than the local variance.

Note that the global average of variance defined by (A1) is equal to twice the global average of the variance of the RPN ensemble. Although this means that the variance is assumed to be 160% larger than current estimates of average analysis error variance, the magnitude of the global average of the background analysis error variance has no impact on the ability of the technique to estimate the change in prediction error variance associated with a change in the deployment of observational resources. This ability is only affected by the assumed geographical variation of background analysis error variance. Results of tests on the sensitivity of prediction error estimates to changes in the way the inverse
1764 JOURNAL OF THE ATMOSPHERIC SCIENCES

Fig. 4. The initial vorticity field of the leading SV for a 60-h forecast for 1200 UTC 15 Feb 1997. This SV is for the observational network that includes supplementary observations centered around the optimal site at 48°N, 275°E (cf. Fig. 2a).

analysis error covariance matrix is chosen will be presented in future work. The tests performed so far have failed to indicate any significant sensitivity to the way in which the background part of the analysis error covariance is estimated.

In using (A1) to define the analysis error variance associated with the standard observational network at the future analysis time $t_{fa}$, we are effectively assuming that there will be no difference between the background analysis error variances at $t_a$ and $t_{fa}$. This assumption is reasonable provided there is no major regime change in the time between $t_a$ and $t_{fa}$.

In regions affected by supplemental observations, we let the analysis error variance be equal to half the isobaric globally averaged $\overline{u^2}$, that is,

$$\overline{u^2} = \frac{1}{RR} \sum_{k=1}^{K} \sum_{x=1}^{X} (u_{kx} - \overline{u}_{kx})^2$$  \hspace{1cm} (A2)

near supplemental observation sites.

APPENDIX B
Details of the Definition of $F^{-1/2}$

The magnitude of a sum such as that which defines the normalized prediction error variance $x(t_y)^T F^{-1/2} x(t_y)$ is roughly proportional to the number $L_p$ of terms in the sum. If $A_{gi}^{-1}(t_{fa})$ was equal to a constant $b$ multiplied by the identity matrix, then Eq. (8) implies that the transformed ensemble perturbations would satisfy

$$x(t_{fa}) A_{gi}^{-1}(t_{fa}) x(t_{fa}) = bL \tilde{x}^2 = K,$$  \hspace{1cm} (B1)

where $\tilde{x}$ is the globally averaged value of the elements in the state vector $x(t_{fa})$ and $L$ is the number of dimensions in the state vector. Equation (B1) indicates that $\tilde{x}$ is inversely proportional to $\sqrt{L}$. Since, in some sort of statistically averaged sense, the magnitude of perturbations in the verification region at the verification time $t_y$ is proportional to the average size of gridpoint values at $t_{fa}$, it follows that nonnormalized prediction error variance is roughly inversely proportional to the number $L$ of terms in the sum $x(t_{fa})^T A_{gi}^{-1} x(t_{fa})$. To prevent our normalized measure of prediction error variance from being strongly dependent on the size of the verification region and the number of variables $L$ in our reduced state vector, the normalized prediction error variance $x(t_y)^T F^{-1/2} x(t_y)$ was defined in the following manner. First, all of the variables are divided by the square root of the isobaric global average of their squared values at $t_{fa}$. Second, they are multiplied by the square root of $L/L_p$. Third, at the verification time we compute the sum of the squares of the resulting normalized gridpoint vorticity and divergence values at 850, 500, and 250 mb over a 1000-km radius disk centered in the northeastern Atlantic. Fourth, the square root of this sum is taken.

REFERENCES


Buizza, R., and T. N. Palmer, 1995: The singular-vector structure of


Thorpe, A. J., 1998: Predictability studies with FASTEX data using the ECMWF IFS. *Proc. ECMWF Workshop on Predictability*, Reading, United Kingdom, ECMWF.


——, and I. Szunyogh, 1997: Review of the use of ECMWF ensemble data for targeting upstream observations during the FASTEX field experiments. NOAA/NWS/NCEP Office Note 420, 131 pp. [Available from Environmental Modeling Center, 5200 Auth Rd., WWB, Rm. 207, Camp Springs, MD 20746.]