

An Ensemble Adjustment Kalman Filter for Data Assimilation

JEFFREY L. ANDERSON

Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey

(Manuscript received 29 September 2000, in final form 11 June 2001)

ABSTRACT

A theory for estimating the probability distribution of the state of a model given a set of observations exists. This nonlinear filtering theory unifies the data assimilation and ensemble generation problem that have been key foci of prediction and predictability research for numerical weather and ocean prediction applications. A new algorithm, referred to as an ensemble adjustment Kalman filter, and the more traditional implementation of the ensemble Kalman filter in which “perturbed observations” are used, are derived as Monte Carlo approximations to the nonlinear filter. Both ensemble Kalman filter methods produce assimilations with small ensemble mean errors while providing reasonable measures of uncertainty in the assimilated variables. The ensemble methods can assimilate observations with a nonlinear relation to model state variables and can also use observations to estimate the value of imprecisely known model parameters. These ensemble filter methods are shown to have significant advantages over four-dimensional variational assimilation in low-order models and scale easily to much larger applications. Heuristic modifications to the filtering algorithms allow them to be applied efficiently to very large models by sequentially processing observations and computing the impact of each observation on each state variable in an independent calculation. The ensemble adjustment Kalman filter is applied to a nondivergent barotropic model on the sphere to demonstrate the capabilities of the filters in models with state spaces that are much larger than the ensemble size.

When observations are assimilated in the traditional ensemble Kalman filter, the resulting updated ensemble has a mean that is consistent with the value given by filtering theory, but only the expected value of the covariance of the updated ensemble is consistent with the theory. The ensemble adjustment Kalman filter computes a linear operator that is applied to the prior ensemble estimate of the state, resulting in an updated ensemble whose mean and also covariance are consistent with the theory. In the cases compared here, the ensemble adjustment Kalman filter performs significantly better than the traditional ensemble Kalman filter, apparently because noise introduced into the assimilated ensemble through perturbed observations in the traditional filter limits its relative performance. This superior performance may not occur for all problems and is expected to be most notable for small ensembles. Still, the results suggest that careful study of the capabilities of different varieties of ensemble Kalman filters is appropriate when exploring new applications.

1. Introduction

Methods used to produce operational forecasts of the atmosphere have been undergoing a gradual evolution over the past decades. Prior to the 1990s, operational prediction centers attempted to produce a single “deterministic” prediction of the atmosphere; initial conditions for the prediction were derived using an assimilation and initialization process that used, at best, information from a single earlier prediction. Since that time, the operational use of multiple forecasts, ensembles, has been developed in an attempt to produce information about the probability distribution (van Leeuwen and Evensen 1996) of the atmospheric forecast (Molteni et al. 1996; Tracton and Kalnay 1993; Toth and Kalnay 1993, 1997; Houtekamer et al. 1995).

Anderson and Anderson (1999, hereafter AA) devel-

oped a Monte Carlo implementation of the nonlinear filtering problem (Jazwinski 1970, chapter 6) for use in atmospheric data assimilation. The framework developed in AA allowed a synthesis of the data assimilation and ensemble generation problem. The method worked well in low-order systems, but it was not immediately clear how it could be applied to the vastly larger models that are commonplace for atmospheric and oceanic prediction and simulation.

The fundamental problem facing the AA method and a variety of other ensemble assimilation techniques, in particular the traditional ensemble Kalman filter (Evensen 1994; Houtekamer and Mitchell 1998; Keppenne 2000), that have been proposed for atmospheric and ocean models is that the sample sizes of practical ensembles are far too small to give meaningful statistics about the complete distribution of the model state conditional on the available observations (Burgers et al. 1998; van Leeuwen 1999). This has led to a variety of clever heuristic methods that try to overcome this prob-

Corresponding author address: Dr. Jeffrey L. Anderson, NOAA/GFDL, Princeton University, P. O. Box 308, Princeton, NJ 08542.
E-mail: jla@gfdl.gov

lem, for instance using ensembles to generate statistics for small subsets of the model variables (Evensen and van Leeuwen 1996; Houtekamer and Mitchell 1998).

The AA method has a number of undesirable features when applied sequentially to small subsets of model state variables that are assumed to be independent from all other subsets for computational efficiency. The most pathological is that prior covariances between model state variables in different subsets are destroyed whenever observations are assimilated. A new method of updating the ensemble in a Kalman filter context, called ensemble adjustment, is described here. This method retains many desirable features of the AA filter while allowing application to subsets of state variables. In addition, modifications to the filter design allow assimilation of observations that are related to the state variables by arbitrary nonlinear operators as can be done with traditional ensemble Kalman filters. The result is an ensemble assimilation method that can be applied efficiently to arbitrarily large models given certain caveats. Low-order model results to be presented here suggest that the quality of these assimilations is significantly better than those obtained by current state-of-the-art methods like four-dimensional variational assimilation (Le Dimet and Talagrand 1986; Lorenc 1997; Rabier et al. 1998) or traditional ensemble Kalman filters. Although the discussion that follows is presented specifically in the context of atmospheric models, it is also applicable to other geophysical models like ocean or complete coupled climate system models.

2. An ensemble adjustment Kalman filter

a. Joint state–observation space nonlinear filter

The state of the atmosphere, χ_t , at a time, t , has the conditional probability density function

$$p(\chi_t | \mathbf{Y}_t), \quad (1)$$

where \mathbf{Y}_t is the set of all observations of the atmosphere that are taken at or before time t . Following Jazwinski (1970) and AA, let \mathbf{x}_t be a discrete approximation of the atmospheric state that can be advanced in time using the atmospheric model equations:

$$d\mathbf{x}_t/dt = M(\mathbf{x}_t, t) + \mathbf{G}(\mathbf{x}_t, t)\mathbf{w}_t. \quad (2)$$

Here, \mathbf{x}_t is an n -dimensional vector that represents the state of the model system at time t , M is a deterministic forecast model, and \mathbf{w}_t is a white Gaussian process of dimension r with mean 0 and covariance matrix $\mathbf{S}(t)$ while \mathbf{G} is an $n \times r$ matrix. The second term on the right represents a stochastic component of the complete forecast model (2). In fact, all of the results that follow apply as long as the time update (2) is a Markov process. As in AA, the stochastic term is neglected initially. For most of this paper, the filter is applied in a perfect model context where

$$d\mathbf{x}_t/dt = M(\mathbf{x}_t, t) \quad (3)$$

exactly represents the evolution of the system of interest.

Assume that a set of m , scalar observations, \mathbf{y}_t^o , is taken at time t (the superscript o stands for observations). The observations are functions of the model state variables and include some observational error (noise) that is assumed to be Gaussian (although the method can be extended to non-Gaussian observational error distributions):

$$\mathbf{y}_t^o = \mathbf{h}_t(\mathbf{x}_t, t) + \boldsymbol{\varepsilon}_t(\mathbf{x}_t, t). \quad (4)$$

Here, \mathbf{h}_t is an m -vector function of the model state and time that gives the expected value of the observations given the model state and $\boldsymbol{\varepsilon}_t$ is an m -vector observational error selected from an observational error distribution with mean 0 and covariance \mathbf{R}_t ; m , is the size of the observations vector that can itself vary with time. It is assumed that the $\boldsymbol{\varepsilon}_t$ for different times are uncorrelated. This may be a reasonable assumption for many traditional ground-based observations although other observations, for instance satellite radiances, may have significant temporal correlations in observational error.

The set of observations, \mathbf{y}_t^o , available at time t can be partitioned into the largest number of subsets, $\mathbf{y}_{t,k}^o$, for which the observational error covariance between subsets is negligible; this is the equivalent of the observation batches used in Houtekamer and Mitchell (2001). Then,

$$\mathbf{y}_{t,k}^o = \mathbf{h}_{t,k}(\mathbf{x}_t, t) + \boldsymbol{\varepsilon}_{t,k}(\mathbf{x}_t, t), \quad k = 1, \dots, r, \quad (5)$$

where $\mathbf{y}_{t,k}^o$ is the k th subset at time t , $\mathbf{h}_{t,k}$ is an m -vector function (m can vary with both time and subset), $\boldsymbol{\varepsilon}_{t,k}$ is an m -vector observational error selected from an observational error distribution with mean 0 and $m \times m$ covariance matrix $\mathbf{R}_{t,k}$, and r is the number of subsets at time t . Many types of atmospheric observations have observational error distributions with no significant correlation to the error distributions of other contemporaneous observations leading to subsets of size one ($\mathbf{y}_{t,k}^o$ is scalar). Note that no restrictions have been placed on $\mathbf{h}_{t,k}$ (and \mathbf{h}_t); in particular the observed variables are not required to be linear functions of the state variables.

A cumulative observation set, \mathbf{Y}_τ , can be defined as the superset of all observations, \mathbf{y}_t^o , for times $t \leq \tau$. The conditional probability density of the model state at time t ,

$$p(\mathbf{x}_t | \mathbf{Y}_\tau), \quad (6)$$

is the complete solution to the filtering problem when adopting a Bayesian point of view (Jazwinski 1970). Following AA, the posterior probability distribution (6) is referred to as the *analysis probability distribution* or *initial condition probability distribution*. The forecast model (3) allows the computation of the conditional probability density at any time after the most recent observation time:

$$p(\mathbf{x}_t | \mathbf{Y}_\tau) \quad t > \tau. \quad (7)$$

This predicted conditional probability density is a forecast of the state of the model, and also provides the prior distribution at the time of the next available observations for the assimilation problem. The temporal evolution of this probability distribution is described by the Liouville equation as discussed in Ehrendorfer (1994). The probability distribution (7) will be referred to as the *first guess probability distribution* or *prior probability distribution* when used to assimilate additional data, or the *forecast probability distribution* when a forecast is being made.

Here, $\mathbf{Y}_{\tau, \kappa}$ is defined as the superset of all observation subsets $\mathbf{y}_{t, k}^o$ with $t \leq \tau$ and $k \leq \kappa$ (note that $\mathbf{Y}_{t, 0} = \mathbf{Y}_{t_p}$, where t_p is the previous time at which observations were available). Assume that the conditional probability distribution $\mathbf{p}(\mathbf{x}_t | \mathbf{Y}_{t, k-1})$ is given. The conditional distribution after making use of the next subset of observations is

$$\mathbf{p}(\mathbf{x}_t | \mathbf{Y}_{t, k}) = \mathbf{p}(\mathbf{x}_t | \mathbf{y}_{t, k}^o, \mathbf{Y}_{t, k-1}). \quad (8)$$

For $k = 1$, the forecast model (3) must be used to compute $\mathbf{p}(\mathbf{x}_t | \mathbf{Y}_{t_p})$ from $\mathbf{p}(\mathbf{x}_{t_p} | \mathbf{Y}_{t_p})$.

In preparation for applying the numerical methods outlined later in this section, define the joint state–observation vector (referred to as joint state vector) for a given t and k as $\mathbf{z}_{t, k} = [\mathbf{x}_t, \mathbf{h}_{t, k}(\mathbf{x}_t, t)]$, a vector of length $n + m$ where m is the size of the observational subset $\mathbf{y}_{t, k}^o$. The idea of working in a joint state–observation space can be used in a very general description of the filtering problem (Tarantola 1987, chapters 1 and 2). Working in the joint space allows arbitrary observational operators, \mathbf{h} , to be used in conjunction with the ensemble methods developed below. Following the same steps that led to (8) gives

$$\mathbf{p}(\mathbf{z}_{t, k} | \mathbf{Y}_{t, k}) = \mathbf{p}(\mathbf{z}_{t, k} | \mathbf{y}_{t, k}^o, \mathbf{Y}_{t, k-1}). \quad (9)$$

Returning to the approach of Jazwinski, Bayes's rule gives

$$\begin{aligned} \mathbf{p}(\mathbf{z}_{t, k} | \mathbf{Y}_{t, k}) \\ = \mathbf{p}(\mathbf{y}_{t, k}^o | \mathbf{z}_{t, k}, \mathbf{Y}_{t, k-1}) \mathbf{p}(\mathbf{z}_{t, k} | \mathbf{Y}_{t, k-1}) / \mathbf{p}(\mathbf{y}_{t, k}^o | \mathbf{Y}_{t, k-1}). \end{aligned} \quad (10)$$

Since the observational noise $\varepsilon_{t, k}$ is assumed uncorrelated for different observation times and subsets,

$$\mathbf{p}(\mathbf{y}_{t, k}^o | \mathbf{z}_{t, k}, \mathbf{Y}_{t, k-1}) = \mathbf{p}(\mathbf{y}_{t, k}^o | \mathbf{z}_{t, k}). \quad (11)$$

Incorporating (11) into (10) gives

$$\begin{aligned} \mathbf{p}(\mathbf{z}_{t, k} | \mathbf{Y}_{t, k}) \\ = \mathbf{p}(\mathbf{y}_{t, k}^o | \mathbf{z}_{t, k}) \mathbf{p}(\mathbf{z}_{t, k} | \mathbf{Y}_{t, k-1}) / \mathbf{p}(\mathbf{y}_{t, k}^o | \mathbf{Y}_{t, k-1}), \end{aligned} \quad (12)$$

which expresses how new sets of observations modify the prior joint state conditional probability distribution available from predictions based on previous observation sets. The denominator is a normalization that guarantees that the total probability of all possible states is 1. The numerator is a product of two terms, the first representing new information from observation subset

k at time t and the second representing the prior constraints. The prior term gives the probability that a given model joint state, $\mathbf{z}_{t, k}$, occurs at time t given information from all observations at previous times and the first $k - 1$ observation subsets at time t . The first term in the numerator of (12) evaluates how likely it is that the observation subset $\mathbf{y}_{t, k}^o$ would be taken given that the state was $\mathbf{z}_{t, k}$. This algorithm can be repeated recursively until the last subset from the time of the latest observation, at which point (3) can be used to produce the forecast probability distribution at any time in the future.

b. Computing the filter product

Applying (12) to large atmospheric models leads to a number of practical constraints. The only known computationally feasible way to advance the prior state distribution, \mathbf{x}_t , in time is to use Monte Carlo techniques (ensembles). Each element of a set of states sampled from (6) is advanced in time independently using the model (3). The observational error distributions of most climate system observations are poorly known and are generally given as Gaussians with zero mean (i.e., a standard deviation or covariance).

Assuming that (12) must be computed given an ensemble sample of $\mathbf{p}(\mathbf{x}_t | \mathbf{Y}_{t, k-1})$, an ensemble of the joint state prior distribution, $\mathbf{p}(\mathbf{z}_{t, k} | \mathbf{Y}_{t, k-1})$, can be computed by applying $\mathbf{h}_{t, k}$ to each ensemble sample of \mathbf{x}_t . The result of (12) must be an ensemble sample of $\mathbf{p}(\mathbf{z}_{t, k} | \mathbf{Y}_{t, k})$. As noted in AA, there is generally no need to compute the denominator (the normalization term) of (12) in ensemble applications. Four methods for approximating the product in the numerator of (12) are presented, all using the fact that the product of two Gaussian distributions is itself Gaussian and can be computed in a straightforward fashion; in this sense, all can be viewed as members of a general class of ensemble Kalman filters.

1) GAUSSIAN ENSEMBLE FILTER

This is an extension of the first filtering method described in AA to the joint state space. Let $\bar{\mathbf{z}}^p$ and Σ^p be the sample mean and covariance of the prior joint state, $\mathbf{p}(\mathbf{z}_{t, k} | \mathbf{Y}_{t, k-1})$, ensemble. The observation subset $\mathbf{y}^o = \mathbf{y}_{t, k}^o$ has error covariance $\mathbf{R} = \mathbf{R}_{t, k}$ (\mathbf{R} and \mathbf{y}^o are functions of the observational system). The expected value of the observation subset given the state variables is $\mathbf{h}_{t, k}(\mathbf{x}_t, t)$, as in (5), but in the joint state space this is equivalent to the simple $m \times (n + m)$ linear operator \mathbf{H} , where $H_{k, k+n} = 1.0$ for $k = 1, \dots, m$ and all other elements of \mathbf{H} are 0, so that the estimated observation values calculated from the joint state vector are $\mathbf{y}_{t, k} = \mathbf{H}\mathbf{z}_{t, k}$.

Assuming that the prior distribution can be represented by a Gaussian with the sample mean and variance results in the numerator of (12) having covariance

$$\Sigma^u = [(\Sigma^p)^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}]^{-1}, \quad (13)$$

mean

$$\bar{z}^u = \Sigma^u [(\Sigma^p)^{-1} \bar{z}^p + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{y}^o], \quad (14)$$

and a relative weight

$$\mathbf{D} = \exp\{-[(\bar{z}^p)^T (\Sigma^p)^{-1} \bar{z}^p + (\mathbf{y}^o)^T \mathbf{R}^{-1} \mathbf{y}^o - (\bar{z}^u)^T (\Sigma^u)^{-1} \bar{z}^u]/2\}. \quad (15)$$

These are an extension of Eqs. (A1)–(A4) in AA to the joint state space (S. Anderson 1999, personal communication). In the Gaussian ensemble filter, the updated ensemble is computed using a random number generator to produce a random sample from a Gaussian with the covariance and mean from (13) and (14). The *expected values* of the mean and covariance of the resulting ensemble are \bar{z}^u and Σ^u while the expected values of all higher-order moments should be 0. The weight, \mathbf{D} , is only relevant in the kernel filter method described in the next subsection, since only a single Gaussian is used in computing the product in the three other filtering methods described here.

2) KERNEL FILTER

The kernel filter mechanism developed in AA can also be extended to the joint state space. In this case, the prior distribution is approximated as the sum of N Gaussians with means z_i^p and identical covariances Σ^p , where z_i^p is the i th ensemble sample of the prior and N is the ensemble size. The product of each Gaussian with the observational distribution is computed by applying (13) once and (14) and (15) N times, with \bar{z}^p replaced by z_i^p in (14) and (15) and \bar{z}_u being replaced by z_i^u in (15) where z_i^u is the result of the i th evaluation of (14). The result is N new distributions with the same covariance but different means and associated weights, \mathbf{D} [Eq. (15)], whose sum represents the product. An updated ensemble is generated by randomly sampling from this set of distributions as in AA. In almost all cases, the *values* and *expected values* of the mean and covariance and higher-order moments of the resulting ensemble are functions of higher-order moments of the prior distribution. This makes the kernel filter potentially more general than the other three methods; however, computational efficiency issues outlined later appear to make it impractical for application in large models.

3) ENSEMBLE KALMAN FILTER

The traditional ensemble Kalman filter (EnKF hereafter) forms a random sample of the observational distribution, $\mathbf{p}(\mathbf{y}_{t,k}^o | \mathbf{z}_{t,k})$ in (12), sometimes referred to as perturbed observations (Houtekamer and Mitchell 1998). The EnKF uses a random number generator to sample the observational error distribution and adds these samples to the observation, \mathbf{y}^o , to form an ensemble sample of the observation distribution, \mathbf{y}_i , $i = 1, \dots, N$. The mean of the perturbations is adjusted to be

so that the perturbed observations, \mathbf{y}_i , have mean equal to \mathbf{y}^o . Equation (13) is computed once to find the value of Σ^u . Equation (14) is evaluated N times to compute z_i^u , with \bar{z}^p and \mathbf{y}^o replaced by the z_i^p and y_i , where the subscript refers to the value of the i th ensemble member. This method is described using more traditional Kalman filter terminology in Houtekamer and Mitchell (1998), but their method is identical to that described above. As shown in Burgers et al. (1998), computing a random sample of the product as the product of random samples is a valid Monte Carlo approximation to the nonlinear filtering equation (12). Essentially, the EnKF can be regarded as an ensemble of Kalman filters, each using a different sample estimate of the prior mean and observations. The updated ensemble has mean \bar{z}^u and sample covariance with an *expected value* of Σ^u , while the expected values of higher-order moments are functions of higher-order moments of the prior distribution.

Deriving the EnKF directly from the nonlinear filtering equation (12) may be more transparent than some derivations found in the EnKF literature where the derivation begins from the statistically linearized Kalman filter equations. This traditional derivation masks the statistically nonlinear capabilities of the EnKF, for instance, the fact that both prior and updated ensembles can have an arbitrary (non-Gaussian) structure. Additional enhancements to the EnKF, for instance the use of two independent ensemble sets (Houtekamer and Mitchell 1998), can also be developed in this context.

4) ENSEMBLE ADJUSTMENT KALMAN FILTER

In the new method that is the central focus of this paper, equations (13) and (14) are used to compute the mean and covariance of the updated ensemble. A new ensemble that has exactly these sample characteristics while maintaining as much as possible the higher moment structure of the prior distribution is generated directly. The method, referred to as *ensemble adjustment*, for generating the new ensemble applies a linear operator, \mathbf{A} , to the prior ensemble in order to get the updated ensemble

$$z_i^u = \mathbf{A}^T (z_i^p - \bar{z}^p) + \bar{z}^u, \quad i = 1, \dots, N, \quad (16)$$

where z_i^p and z_i^u are individual members of the prior and updated ensemble. The $(n+m) \times (n+m)$ matrix \mathbf{A} is selected so that the sample covariance of the updated ensemble is identical to that computed by (13). Appendix A demonstrates that \mathbf{A} exists (many \mathbf{A} 's exist since corresponding indices of prior and updated ensemble members can be scrambled) and discusses a method for computing the appropriate \mathbf{A} . As noted by M. K. Tippett (2000, personal communication), this method is actually a variant of a square root filter methodology. An implementation of a related square root filter is described in Bishop et al. (2001).

c. Applying ensemble filters in large systems

The size of atmospheric models and of computationally affordable ensembles necessitate additional simplifications when computing updated means and covariances in ensemble filters. The sample prior covariance computed from an N -member ensemble is nondegenerate in only $N - 1$ dimensions of the joint state space. If the global covariance structure of the assimilated joint state cannot be represented accurately in a subspace of size $N - 1$, filter methods are unlikely to work without making use of other information about the covariance structure (Lermusiaux and Robinson 1999; Miller et al. 1994a). When the perfect model assumption is relaxed, this can become an even more difficult problem since model systematic error is not necessarily likely to project on the subspace spanned by small ensembles.

One approach to dealing with this degeneracy is to project the model state onto some vastly reduced subspace before computing products, leading to methods like a variety of reduced space (ensemble) Kalman filters (Kaplan et al. 1997; Gordeau et al. 2000; Brasseur et al. 1999). A second approach, used here, is to update small sets of “physically close” state variables independently.

Let C be a set containing the indices of all state variables in a particular independent subset of state variables, referred to as a *compute domain*, along with the indices of all possibly related observations in the current joint state vector. Let D be a set containing the indices of all additional related state variables, referred to as the *data domain*. Then Σ_{ij}^u and \bar{z}_i^u , where $i, j \in C$, are computed using an approximation to Σ_{ij}^p in which all terms for which $i \notin C \cup D$ or $j \notin C \cup D$ are set to zero. In other words, the state variables in each compute domain are updated making direct use only of prior covariances between themselves, related observations, and also variables in the corresponding data domain. These subsets can be computed statically (as will be done in all applications here) or dynamically using information available in the prior covariance and possibly additional a priori information. The data domain state variables in D may themselves be related strongly to other state variables outside of $C \cup D$ and so are more appropriately updated in conjunction with some other compute set.

Additional computational savings can accrue by performing a singular value decomposition on the prior covariance matrix (already done as part of the numerical method for updating the ensembles as outlined in appendix A) and working in a subspace spanned by singular vectors with nonnegligible singular values. This singular vector filtering is a prerequisite if the size of the set $C \cup D$ exceeds $N - 1$, leading to a degenerate sample prior covariance matrix (Houtekamer and Mitchell 1998; Evensen and van Leeuwen 1996).

All results in the following use particularly simple and computationally efficient versions of the filtering algorithms. First, all observation subsets contain a single

observation; in this perfect model case this is consistent with the observations that have zero error covariance with other observations (Houtekamer and Mitchell 2001). Second, the compute domain set, C , also contains only a single element and the data domain D is the null set in all cases. The result is that each component of the mean and each prior covariance diagonal element is updated independently (this does *not* imply that the prior or updated covariances are diagonal). The joint state prior covariance matrix used in each update is 2×2 containing the covariance of a single state and the single observation in the current observational subset. In computing the products to get the new state estimate, the ensemble adjustment Kalman filter (EAKF) algorithm used here only makes use of singular value decompositions and inverses of 2×2 matrices; similarly, the EnKF only requires 2×2 matrix computations. Allowing larger compute and data domains would generally be expected to improve slightly the results discussed in later sections while leading to significantly increased constant factors multiplying computational cost.

d. Motivation for EAKF

This section discusses advantages of the EAKF and EnKF over the Gaussian and kernel filters, both referred to as *resampling* Monte Carlo (or just resampling) methods since a random sample of the updated distribution must be formed at each update step. Applying resampling filters locally to subsets of the model state variables as discussed in the previous subsection, one might expect the structure of the assimilated probability distributions to be simpler and more readily approximated by Gaussians. Subsets of state variables of size smaller than N can be used so that the problem of degenerate sample covariance matrices is avoided altogether. This can solve problems of filter divergence that result from global applications of resampling filters (AA). The state variables can be partitioned into compute and data subsets as described above, motivated by the concept that most state variables are closely related only to a subset of other state variables, usually those that are physically nearby. Ignoring prior covariances with more remote variables is expected to have a limited impact on the computation of the product. Similar approaches have been used routinely in EnKFs (Houtekamer and Mitchell 1998).

Unfortunately, resampling ensemble filters are not well suited for local application to subsets of state variables. Whenever an observation is incorporated, the updated mean(s) and covariance(s) are computed using Eqs. (13) and (14) and a new ensemble is formed by randomly sampling the result. Even when observations with a very low relative information content (very large error covariance compared to the prior covariance) are assimilated, this resampling is done. However, resampling destroys all information about prior covariances

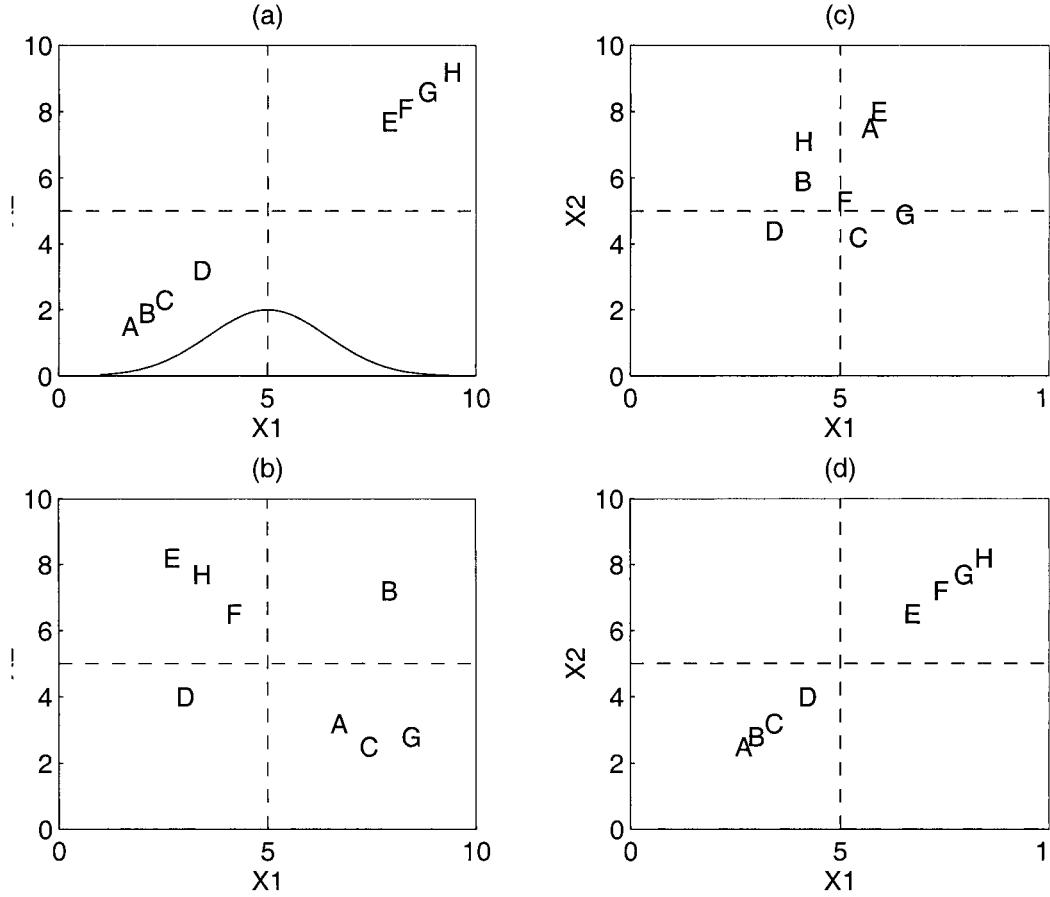


FIG. 1. Schematic showing results of applying different filters to two variables X_1 and X_2 in different compute subsets. (a) The prior distribution of an eight-member ensemble in the X_1 - X_2 plane and the solid curve is an idealized distribution for an observation of X_1 . The results of applying (b) a kernel resampling filter, (c) a single Gaussian resampling filter, and (d) an ensemble adjustment Kalman filter are depicted in the same plane. The distribution for an ensemble Kalman filter would look similar to (d) with some amount of additional noise added to the ensemble positions.

between state variables in different compute subsets. The assumption that the prior covariances between different subsets are small is far from rigorous in applications of interest, so it is inconvenient to lose all of this information every time observations become available.

Figure 1a shows an idealized representation of a system with state variables X_1 and X_2 that are in different compute domains. An idealized observation of X_1 with Gaussian error distribution is indicated schematically by the density plot along the X_1 axis in Fig. 1a. Figure 1d shows the result of applying an EAKF in this case. The adjustment pulls the value of X_1 for all ensemble members toward the observed value. The covariance structure between X_1 and X_2 is mostly preserved as the values of X_2 are similarly pulled inward. The result is qualitatively the same as applying a filter to X_1 and X_2 simultaneously (no subsets). Figure 1c shows the results of applying a single Gaussian resampling filter and Fig. 1b the result of a multiple kernel resampling filter as in

AA. The resampling filters destroy all prior information about the covariance of X_1 and X_2 .

There are other related problems with resampling ensemble filters. First, it is impossible to meaningfully trace individual assimilated ensemble trajectories in time. While the EAKF maintains the relative positions of the prior samples, the letters in Figs. 1b and 1c are scrambled throughout the resulting distributions. This can complicate diagnostic understanding of the assimilation. Trajectory tracing is easier in the EnKF than in the resampling filters, but, especially with small ensembles, less straightforward than in the EAKF due to the noise added in the perturbed observations.

Second, if only a single Gaussian kernel is being used to compute the product, all information about higher-order moments of the prior distribution is destroyed each time data are assimilated (Fig. 1c). Anderson and Anderson (1999) introduced the sum of Gaussian kernels approximation to avoid this problem. In Fig. 1b, the projection of higher-order structure on the individual

state variable axes is similar to that in Fig. 1d, but the distribution itself winds up being qualitatively a quadrupole because of the loss of covariance information between X_1 and X_2 .

These deficiencies of the resampling ensemble filters occur because a random sampling of the updated probability distribution is used to generate the updated ensemble. In contrast, the EAKF and EnKF retain some information about prior covariances between state variables in separate compute subsets as shown schematically in Fig. 1d for the EAKF (a figure for the EnKF would be similar with some amount of additional noise added to the ensemble locations). For instance, observations that have a relatively small information content make small changes to the prior distributions. Most of the covariance information between variables in different subsets survives the product step in this case. This is particularly relevant since the frequency of atmospheric and oceanic observations for problems of interest may lead to individual (subsets of) observations making relatively small adjustments to the prior distributions.

The EAKF and EnKF also preserve information about higher-order moments of prior probability distributions as shown in Fig. 1d. Again, this information is particularly valuable when observations make relatively small adjustments to the prior distributions. For instance, if the dynamics of a model are generating distributions with interesting higher moment structure, for instance a bimodality, this information can survive the update step using the EAKF or EnKF but is destroyed by resampling with a single Gaussian kernel.

Individual ensemble trajectories can be meaningfully traced through time with the EAKF and the EnKF although the EnKF is noisier for small ensembles (see also Figs. 3 and 9). If observations make small adjustments to the prior, individual ensemble members look similar to free runs of the model with periodic small jumps where data are incorporated. Note that the EAKF is deterministic after initialization, requiring no generation of random numbers once an initial ensemble is created.

The EAKF and EnKF are able to eliminate many of the shortcomings of the resampling filters. Unlike the resampling filters, they can be applied effectively when subsets of state variables are used for computing updates. The EAKF and EnKF retain information about higher-order moments of prior distributions and individual ensemble trajectories are more physically relevant leading to easier diagnostic evaluation of assimilations. All of these advantages are particularly pronounced in instances where observations at any particular time have a relatively small impact on the prior distribution, a situation that seems to be the case for most climate system model/data problems of interest.

e. Avoiding filter divergence

Since there are a number of approximations permeating the EAKF and EnKF, there are naturally inaccuracies in the prior sample covariance and mean. As for other filter implementations, like the Kalman filter, sampling error or other approximations can cause the computed prior covariances to be too small at some times.

The result is that less weight is given to new observations when they become available resulting in increased error and further reduced covariance in the next prior estimate. Eventually, the prior may no longer be impacted significantly by the observations, and the assimilation will depart from the observations. A number of sophisticated methods for dealing with this problem can be developed. Here, only a simple remedy is used. The prior covariance matrix is multiplied by a constant factor, usually slightly larger than one. If there are some local (in phase space) linear balances between the state variables on the model's attractor, then the application of small covariance inflation might be expected to maintain these balances while still increasing uncertainty in the state estimate. Clearly, even if there are locally linear balanced aspects to the dynamics on the attractor, the application of sufficiently large covariance inflations would lead to significantly unbalanced ensemble members.

The covariance inflation factor is selected empirically here in order to give a filtering solution that does not diverge from the observations while keeping the prior covariances small. For all results shown, a search of covariance inflation values is made until a minimum value of ensemble mean rms error is found and results are only reported for these tuned cases. The impacts of covariance inflation in the EnKF are explored in Hamill et al. (2001). More sophisticated approaches to this problem are appropriate when dealing with models that have significant systematic errors (i.e., when assimilating real observations) and are currently being developed.

f. "Distant" observations and maintaining prior covariance

As pointed out in section 2d, one of the advantages of the EAKF and EnKF is that they can maintain much of the prior covariance structure even when applied independently to small subsets of state variables. This is particularly important in the applications reported here where each state variable is updated independently from all others. If, however, two state variables that are closely related in the prior distribution are impacted by very different subsets of observations, they may end up being too weakly related in the updated distribution.

One possible (expensive) solution, would be to let every state variable be impacted by all observations. This can, however, lead to another problem that has been noted for the EnKF. Given a large number of observations that are expected to be physically unrelated to a particular state variable, say because they are observations of physically remote quantities, some of these observations will be highly correlated with the state

variable by chance and will have an erroneous impact on the updated ensemble. The impact of spuriously correlated remote observations can end up overwhelming more relevant observations (Hamill et al. 2001).

Following Houtekamer and Mitchell (2001), all low-order model results here multiply the covariances between prior state variables and observation variables in the joint state space by a correlation function with local support. The correlation function used is the same fifth-order piecewise rational function used by Gaspari and Cohn [(1999), their equation (4.10)] and used in Houtekamer and Mitchell. This correlation function is characterized by a single parameter, c , that is the half-width of the correlation function. The Schur product method used in Houtekamer and Mitchell can be easily computed in the single state variable cases presented here by simply multiplying the sample covariance between the single observation and single state variable by the distance dependent factor from the fifth-order rational function.

3. Results from a low-order system

The EAKF and EnKF are applied to the 40-variable model of Lorenz [(1996), referred to hereafter as L96; see appendix B], which was used for simple tests of targeted observation methodologies in Lorenz and Emanuel (1998). The number of state variables is greater than the smallest ensemble sizes (approximately 10) required for usable sample statistics and the model has a number of physical characteristics similar to those of the real atmosphere. All cases use synthetic observations generated, as indicated in (4), over the course of a 1200 time step segment of a very long control integration of the 40-variable model. Unless otherwise noted, results are presented from the 1000 step assimilation period from step 200 to 1200 of this segment. Twenty-member ensembles are used unless otherwise noted.

For all L96 results reported, for both the EAKF and the EnKF, a search is made through values of the covariance inflation parameter and the correlation function half-width, c . The covariance inflation parameter is independently tuned for c 's of 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, and 0.35 in order to minimize the rms error over the assimilation period. For the smallest c , state variables are only impacted by observations at a distance of less than 0.10 (total of 20% of the total domain width of 1.0) while in the 0.35 case, the correlation function actually wraps around in the cyclic domain allowing even the most distant observations to have a nonnegligible impact on a state variable.

a. Identity observation operators

In the first case examined, the observational operator, \mathbf{h} , is the identity (each state variable is observed directly), the observational error covariance is diagonal with all elements 4.0 (observations have independent

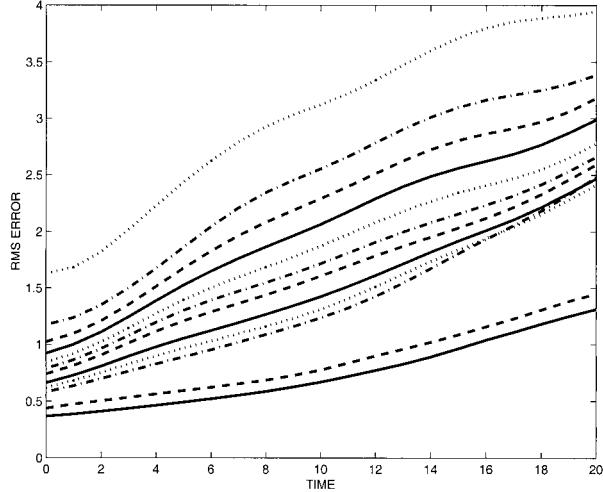


FIG. 2. Rms error as a function of forecast lead time (lead time 0 is the error of the assimilation) for ensemble adjustment Kalman filters with a 10-member ensemble (lowest dashed curve) and a 20-member ensemble (lowest solid curve) and for four-dimensional variational assimilations that use the model as a strong constraint to fit observations over a number of observing times. In generally descending order, the number of observation times used by the variational method is two (dotted), 3 (dash-dotted), 4 (dashed), 5 (solid), 6 (dotted), 7 (dash-dotted), 8 (dashed), 10 (solid), 12 (dotted), and 15 (dash-dotted).

error variance of 4), and observations are available every time step. As discussed in detail in AA, the goal of filtering is to produce an ensemble with small ensemble mean error and with the true state being statistically indistinguishable from a randomly selected member of the ensemble. For the EAKF, the smallest time mean rms error of the ensemble mean for this assimilation is 0.390 for a c of 0.3 and covariance inflation of 1.01. Figure 2 shows the rms error of the ensemble mean for this assimilation and for forecasts started from the assimilation out to leads of 20 assimilation times for steps 100–200 of the assimilation (101 forecasts; this period is selected for comparison with four-dimensional variational methods in section 5). Results for the EAKF would be virtually indistinguishable if displayed for steps 200–1200.

Figure 3a shows a time series of the “truth” from the control run and the corresponding ensemble members (the first 10 of the total of 20 are displayed to reduce clutter) and ensemble mean from the EAKF for variable X_1 . There is no evidence in this figure that the assimilation is inconsistent with the truth. The truth lies close to the ensemble mean (compared to the range of the variation in time) and generally is inside the 10 ensemble members plotted. The ensemble spread varies significantly in time; for instance, the ensemble is more confident about the state (less spread) when the wave trough is approaching at assimilation time 885 than just after the peak passes at time 875. The ability to trace individual ensemble member trajectories in time is also clearly demonstrated; as noted in section 2 this could

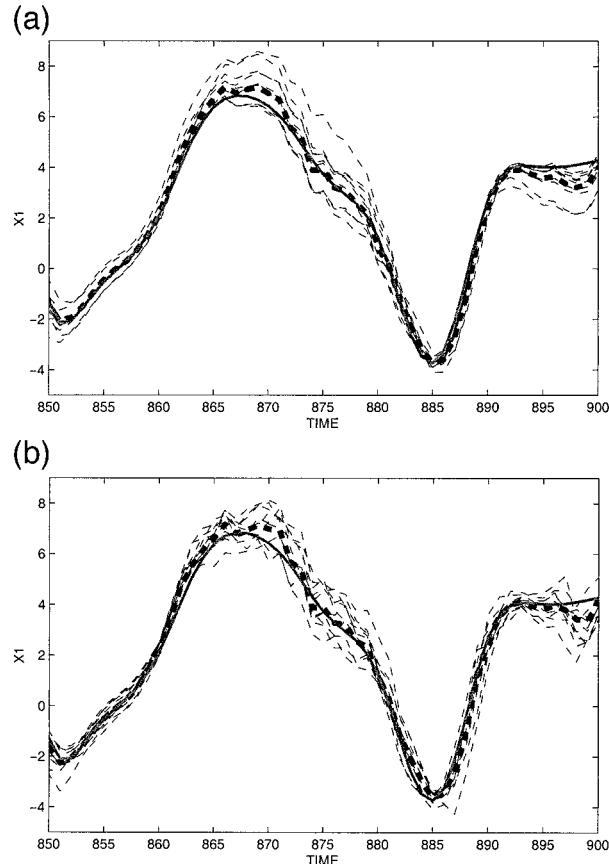


FIG. 3. Time series of “truth” from long control run (solid gray), ensemble mean (thick dashed), and the first 10 of the 20 individual ensemble members (thin dashed) for variable X_1 of the L96 model from assimilation times 850–900 using (a) an ensemble adjustment Kalman filter and (b) an ensemble Kalman filter.

not be done in resampling methods. As an example, notice the trajectory that maintains a consistently high estimate from steps 870 through 880.

Figure 4 displays the rms error of the ensemble mean and the ensemble spread (the mean rms difference between ensemble members and the ensemble mean) for the X_1 variable for assimilation times 850–900. There is evidence of the expected relation between spread and skill; in particular, there are no instances when spread is small but error is large. For steps 200–1200, the rms error of the ensemble mean and the ensemble spread have a correlation of 0.351. The expected relation between spread and skill (Murphy 1988; Barker 1991) will be analyzed in detail in a follow-on study.

Figure 5 shows the result of forming a rank histogram [a Talagrand diagram; Anderson (1996)] for the X_1 variable. At each analysis time, this technique uses the order statistics of the analysis ensemble of a scalar quantity to partition the real line into $n + 1$ intervals (bins); the truth at the corresponding time falls into one of these $n + 1$ bins. A necessary condition for the analysis ensemble to be a random sample of (6) is that the distri-

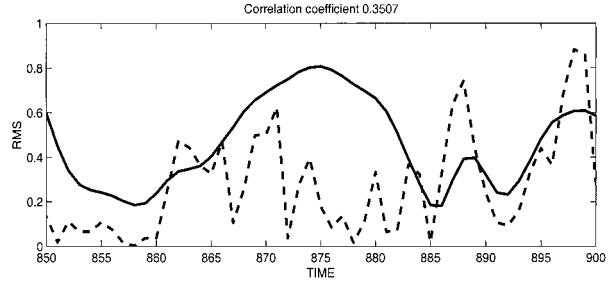


FIG. 4. Time series of rms error of ensemble mean from ensemble adjustment Kalman filter assimilation (dashed) and mean rms difference between ensemble members and the ensemble mean (spread, solid) for variable X_1 of the L96 model from assimilation times 850–900 of the same assimilation as in Fig. 3.

bution of the truth into the $n + 1$ bins be uniform (Anderson 1996). This is evaluated with a standard chi-square test applied to the distribution of the truth in the $n + 1$ bins. The null hypothesis is that the truth and analysis ensemble are drawn from the same distribution. Figure 5 does not show much evidence of the pathological behavior demonstrated by inconsistent ensembles, for instance clumping in a few central bins or on one or both wings. Obviously, if one uses large enough samples the truth will always be significantly different from the ensemble at arbitrary levels of confidence. Assuming that the bin values at each time are independent, the chi-square test applied to Fig. 5 gives 38.15, indicating a 99% chance that the ensemble was selected from a different distribution than the truth for this sample of 1000 assimilation times. However, the bins occupied by the truth on successive time steps are not independent (see for instance Fig. 3a) as is assumed by the chi-square test. This implies that the chi-square result assumes too many degrees of freedom indicating that the distribution is less uniform than it is in reality (T. Hamill 2001, personal communication).

Another simple method for evaluating whether the truth is similar to a randomly selected ensemble member is to compute the ratio of the time-averaged rms error of the ensemble mean to the time-averaged mean rms error of the ensemble members (this can be done for the complete state vector or individual state components). As shown by Murphy (1988, 1990), the expected

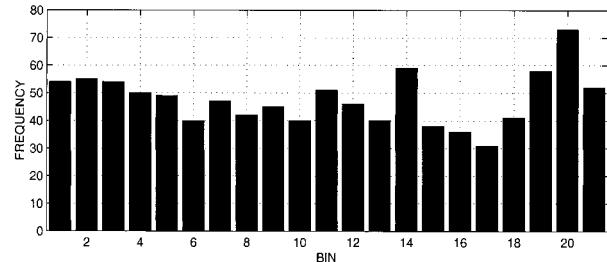


FIG. 5. Rank probability histogram (Talagrand diagram) of the true solution for X_1 within the 20-member ensemble of the ensemble adjustment Kalman filter assimilation for assimilation times 200–1200.

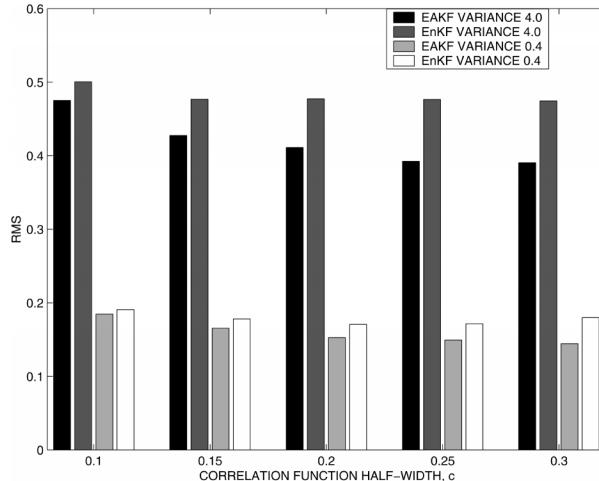


FIG. 6. Time mean of rms error of ensemble mean for steps 200–1200 for identity observation assimilations with observational error variances of 4.0 and 0.4 for 20-member EAKF and EnKFs as a function of correlation function half-width, c , for c ranging from 0.10 to 0.30.

value of this ratio (referred to as the rms ratio hereafter) should be

$$E(\text{Ra}) = \sqrt{(N + 1)/2N}$$

if the truth is statistically indistinguishable from a member of the analysis ensemble for an N -member ensemble. In the following, the ratio of Ra for a given experiment to the expected value,

$$r = (\text{Ra})/E(\text{Ra}),$$

referred to as the normalized rms ratio, is used to evaluate ensemble performance. For this assimilation, r for the complete state vector is 1.003, close to unity but indicating that the ensemble has slightly too little uncertainty (too little spread).

The same experiment has been run using only a 10-member ensemble. Results are generally slightly worse, as shown by the rms error curves as a function of lead time in Fig. 2. Using ensembles much smaller than 10 leads to sample covariance estimates that are too poor for the filter to converge. Using ensembles larger than 20 leads to small improvements in the rms errors.

It is important to examine the rate at which ensemble mean error and spread grow if the assimilation is turned off to verify that the EAKF is performing in a useful fashion. In this case, the forecast error growth plot in Fig. 2 shows that error doubles in about 12 assimilation times.

For comparison, the EnKF is applied to the same observations from the L96 model and produces its best assimilation with a time mean rms error of 0.476 for c of 0.25 and covariance inflation of 1.12 (see Fig. 6). Time series of the individual assimilation members for the EnKF (Fig. 3b) are somewhat noisier than those for the EAKF and in some places (like the one shown in

Fig. 3b) it can become challenging to trace individual ensemble members meaningfully in time.

The EAKF and EnKF were also applied in an experiment with the observational error covariance decreased by a factor of 10 to 0.4. In this case, the best EAKF result produced a time mean rms error of 0.144 for correlation function half-width c of 0.30 and covariance inflation of 1.015 while the best EnKF had rms error of 0.171 for c of 0.20 and covariance inflation of 1.06 (Fig. 6). The ratio of the best EAKF to EnKF rms is 0.842 for the reduced variance case, a slight increase from the 0.819 ratio in the larger variance case. The ratio of this reduced observational error variance to the “climatological” variance of the L96 state variables could be argued to be more consistent with the ratio for the atmospheric prediction problem.

An EnKF with two independent ensemble sets (Houtekamer and Mitchell 1998) was also applied to this example. Results for a pair of 10-member EnKF ensembles were worse than for the single 20-member ensemble. This paired EnKF method was evaluated for all other EnKF experiments discussed here and always produced larger rms than a single EnKF. Given this degraded performance, there is no further discussion of results from paired ensemble EnKFs.

b. Nonlinear observations

A second test in the L96 model appraises the EAKF's ability to deal with nonlinear forward observation operators. Forty observations placed randomly in the model domain are taken at each assimilation time. The observational operator, \mathbf{h} , involves a linear interpolation from the model grid to the location of the observation, followed by a squaring of the interpolated value. The observational errors are independent with variance 64.0. In this case, the EAKF with covariance inflation of 1.02 and correlation function half-width c of 0.30 produces a time mean rms error of 0.338 (Fig. 7) while the normalized rms ratio r is 1.002. The results of the EAKF in this case are qualitatively similar to those discussed in a related assimilation experiment, which is discussed in more detail in section 4. The EnKF was also applied in this nonlinear observations case giving a best time mean rms of 0.421 for c of 0.25 and covariance inflation of 1.12 (Fig. 7). A number of additional experiments in this nonlinear observation case are examined in the next subsection.

c. Comparison of EAKF and EnKF

The results presented to date suggest that the inclusion of noise in the EnKF through the use of perturbed observations may lead to degraded performance relative to the EAKF. One potential problem with ensemble Kalman filters in general is the impact of spurious correlations with physically unrelated variables (Hamill et al. 2001). This is the motivation for limiting the impact of

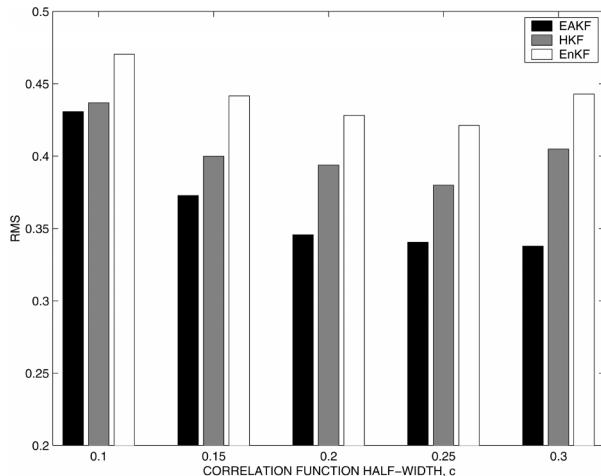


FIG. 7. Time mean of rms error of ensemble mean for steps 200–1200 in nonlinear observation assimilations for 20-member EAKF, EnKF, and a hybrid filter described in the text HKF as a function of correlation function half-width, c , for c ranging from 0.10 to 0.30.

spatially remote observations through a correlation function like the one used in the results above. Figure 7 shows the impact of varying c , the half-width of the correlation function, on EAKF and EnKF performance for the nonlinear observations case described in section 3b. For the EAKF, the rms reduces monotonically over this range as c is increased (Fig. 7). For all values of c , the EnKF produces greater rms than the EAKF; however, it does not show a monotonic decrease in rms with c . Instead, the EnKF rms has a minimum for $c = 0.25$ as shown in Fig. 7. If a correlation function is not used at all (same as limit of c becomes large), the rms error of the EAKF is 0.49, considerably greater than for c of 0.3, but the EnKF diverges from the truth for all values of covariance inflation. It is not surprising that rms decreases as c is increased, allowing more observations to impact each state variable. The increase in rms for very large c is consistent with Hamill et al. (2001); as more observations with weak physical relation are allowed to impact each state variable, the assimilation will eventually begin to degrade due to spurious correlations between observations and state variables. This behavior is exacerbated in the EnKF since the noise introduced can itself lead to spurious correlations.

Figure 6 shows this same behavior as a function of c in the identity observation cases discussed in section 3a. The relative differences between the EAKF and EnKF rms are slightly smaller for the case with reduced observational error variance of 0.4. Again, this is expected as the noise introduced through perturbed observations is smaller in this case and would be expected to produce fewer large spurious correlations.

In the EAKF, the limited number of physically remote variables in the 40-variables model is barely sufficient to see a slight decrease in rms when all variables are allowed to impact each state variable. In models like

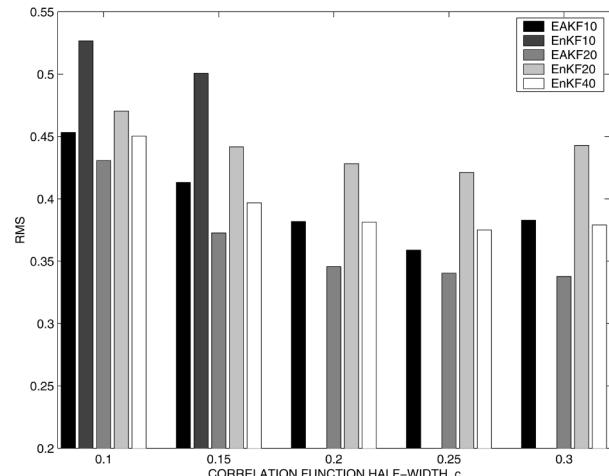


FIG. 8. Time mean of rms error of ensemble mean for steps 200–1200 in nonlinear observation assimilations for 10- and 20-member ensemble adjustment filters (EAKF10 and EAKF20) and for 10-, 20-, and 40-member ensemble Kalman filters (EnKF10, EnKF20, and EnKF40) as a function of correlation function half-width, c , for c ranging from 0.10 to 0.30.

three-dimensional numerical weather prediction models with many more physically remote state variables, the EAKF should be subject to more serious problems with spurious remote correlations.

Figure 8 shows the impact of ensemble size on the EAKF and EnKF for the nonlinear observations case from section 3b. As ensemble size decreases, the problem of spurious correlations should increase (Hamill et al. 2001). For the EAKF, a 10-member ensemble produces rms results that are larger than those for 20 members for all values of c , and the relative degradation in performance becomes larger as c increases. For the 10-member ensemble, the EAKF has a minimum rms for c of 0.25 indicating that the impact of spurious correlations is increased. Results for a 40-member EAKF ensemble are very similar to those for the 20-member ensemble and are not plotted in Fig. 8. Apparently, sampling error is no longer the leading source of error in the EAKF for ensembles larger than 20 in this problem.

The EnKF also shows more spurious correlation problems for 10-member ensembles; for values of c greater than 0.15 the EnKF diverged for all values of covariance inflation. For c equal to 0.15 and 0.10, the EnKF did not diverge but did produce rms errors substantially larger than the 10-member EAKF or the 20-member EnKF. These results further confirm the EnKF's enhanced sensitivity to spurious correlations.

For all the cases examined in low-order models, the EnKF requires a much larger covariance inflation value than does the EAKF. Optimal values of covariance inflation for the EnKF range from 1.08 to 1.16 for 20-member ensembles. For the EAKF, the optimal values range from 1.005 to 1.02. For 40-member ensembles, optimal values of covariance inflation were somewhat smaller, especially for the EnKF, but the EnKF values

were still much larger than those for the EAKF. The larger values of covariance inflation are required because the EnKF has an extra source of potential filter divergence since only the expected value of the updated sample covariance is equal to that given by (13). By chance, there will be cases when the updated covariance is smaller than the expected value. In general, this is expected to lead to a prior estimate with reduced covariance and increased error at the next assimilation time, which in turn is expected to lead to an even more reduced estimate after the next assimilation. Turning up covariance inflation to avoid filter divergence at such times leads to the observational data being given too much weight at other times when the updated covariance estimates are too *large* by chance. The net result is an expected degradation of EnKF performance.

To further elucidate the differences between the EAKF and EnKF, a hybrid filter (referred to as HKF hereafter) was applied to the nonlinear observation case. The hybrid filter begins by applying the EnKF to a state variable–observation pair. The resulting updated ensemble of the state variable has variance whose expected value is given by (13), but whose actual sample variance differs from this value due to the use of perturbed observations. As a second step, the hybrid filter scales the ensemble around its mean so that the resulting ensemble has both the same mean and sample variance as the EAKF. However, the noise introduced by the perturbed observations can still impact higher-order moments of the state variable distribution and its covariance with other state variables. Figure 7 shows results for the EAKF, HKF, and EnKF for a range of correlation function c 's. In all cases, the rms of the HKF is between the EAKF and EnKF values, but the HKF rms is much closer to the EAKF for small values of c . As anticipated, the values of covariance inflation required for the best rms for the HKF are smaller than for the EnKF, with values ranging from 1.01 for c of 0.10 to 1.04 for c of 0.20, 0.25, and 0.30. The HKF experiment can be viewed as isolating the impacts of enhanced spurious correlations from the impacts of the larger covariance inflation required to avoid filter divergence in the EnKF. For small c , almost all the difference between the EnKF and EAKF is due to the enhanced covariance inflation while for larger c , most of the degraded performance is due to enhanced impact of spurious correlations.

The EnKF's addition of random noise through “perturbed observations” at each assimilation step appears to be sufficient to degrade the quality of the assimilation through these two mechanisms. The L96 system is quite tolerant of added noise with off-attractor perturbations decaying relatively quickly and nearly uniformly toward the attractor; the noise added in the EnKF could be of additional concern in less tolerant systems.

4. Estimation of model parameters

Most atmospheric models have many parameters (in dynamics and subgrid-scale physical parameterizations)

for which appropriate values are not known precisely. One can recast these parameters as independent model variables (Derber 1989), and use assimilation to estimate values for the unknown parameters. Ensemble filters can produce a sample of the probability distribution of such parameters given available observations.

To demonstrate this capability, the forcing parameter, F , in the L96 model is treated as a model variable (the result is a 41-variable model) and the EAKF is applied to the extended model using the same set of observations as in the nonlinear observation case described in section 3b. For assimilation steps 200–1200, the EAKF with covariance inflation of 1.02 and correlation half-width c of 0.30 produces a time mean rms error of 0.338 while the normalized rms ratio r is 0.996 indicating that the ensemble has slightly too much spread. There is no good benchmark available to which these values can be compared, but they suggest that the EAKF is working appropriately in this application. It is interesting to note that the rms error is nearly identical to that obtained in the experiment in section 3b in which F was fixed at the correct value.

The time mean rms error for F is 0.0232 over steps 200–1200. Figure 9a shows a time series of F from this assimilation. The “true” value is always 8, but the filter has no a priori information about the value or that the value is constant in time. Also, there are no observations of F , so information is available only indirectly through the nonlinear observations of the state variables; all observations are allowed to impact F . The assimilation is more confident about the value of F at certain times like time 925 than at others like time 980. The chi-square for F over the assimilation from steps 200 to 1200 is very large indicating that the truth was selected from a different distribution. However, as shown in Fig. 9a there is a very large temporal correlation in which bin is occupied by the truth, suggesting that the number of degrees of freedom in the chi-square test would need to be modified to produce valid confidence estimates.

Estimating state variables in this way may offer a mechanism for tuning parameters in large models (Houtekamer and Lefavre 1997; Mitchell and Houtekamer 2000), or even allow them to be time varying with a distribution. It remains an open question whether there is sufficient information in available observations to allow this approach in current-generation operational models. Given the extreme difficulty of tuning sets of model parameters, an investigation of the possibility that this mechanism could be used seems to be of great importance.

One could further extend this approach by allowing a weighted combination of different subgrid-scale parameterizations in each ensemble member and assimilating the weights in an attempt to determine the most appropriate parameterizations. This would be similar in spirit to the approaches described by Houtekamer et al. (1996) and might be competitive with methods of generating “superensembles” from independent models

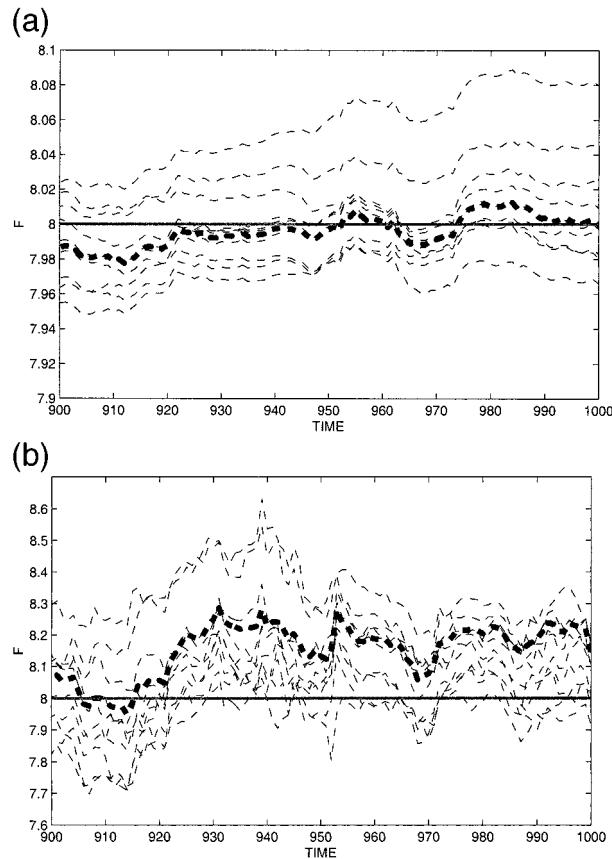


FIG. 9. Time series of truth from long control run (solid gray), ensemble mean (thick dashed) and the first 10 of the 20 individual ensemble members (thin dashed) for the model forcing variable, F , of the L96 model from assimilation times 900–1000 for an assimilation with nonlinear observations operator described in text; results from (a) an ensemble adjustment Kalman filter and (b) an ensemble Kalman filter.

(Goerss 2000; Harrison et al. 1999; Evans et al. 2000; Krishnamurthi et al. 1999; Ziehmann 2000; Richardson 2000).

The best EnKF result for this problem had an rms error of 0.417 for c of 0.20 and covariance inflation 1.08. However, the rms error in F is 0.108, about four times as large as for the EAKF. Figure 9b shows a time series of the EnKF estimate of the forcing variable, F , for comparison with Fig. 9a. The spread and rms error are much larger and the individual EnKF trajectories display a much greater high-frequency time variation than did those for the EAKF.

The introduction of noise in the EnKF is particularly problematic for the assimilation of F because, in general, all available observations are expected to be weakly, but equally, correlated with F . There is no natural way to use a correlation function to allow only some subset of observations to impact F as there was for state variables. The result is that the EnKF's tendency to be adversely impacted by spurious correlations with weakly related observations has a much greater impact than

for regular state variables. This result suggests that the EnKF will have particular difficulty in other cases where a large number of weakly correlated observations are available for a given state variable, for instance certain kinds of wide field of view satellite observations.

5. Comparison to four-dimensional variational assimilation

Four-dimensional variational assimilation methods (4DVAR) are generally regarded as the present state of the art for the atmosphere and ocean (Tziperman and Sirkes 1997). A 4DVAR has been applied to the L96 model and results compared to those for the EAKF. The 4DVAR uses the L96 model as a strong constraint (Zupanski 1997), perhaps not much of an issue in a perfect model assimilation. The 4DVAR optimization is performed with an explicit finite-difference computation of the derivative, with 128-bit floating point arithmetic, and uses as many iterations of a preconditioned, limited-memory quasi-Newton conjugate gradient algorithm (NAG subroutine E04DGF) as are required to converge to machine precision (in practice the number of iterations is generally less than 200). The observations available to the 4DVAR are identical to those used by the EAKF, and the number of observation times being fit by the 4DVAR is varied from 2 to 15 (cases above 15 began to present problems for the optimization even with 128 bits).

Figure 2 compares the rms error of the 4DVAR assimilations and forecasts to those for the EAKF assimilations out to leads of 20 assimilation times for the first case presented in section 3. All results are the mean for 101 separate assimilations and subsequent forecasts, between assimilation steps 100–200. As the number of observation times used in the 4DVAR is increased, error is reduced but always remains much greater than the EAKF error. The 4DVAR cases also show accelerated error growth as a function of forecast lead compared to the EAKF when the number of observation times for the 4DVAR gets large, a symptom of increasing overfitting of the observations (Swanson et al. 1998). An EAKF with only 10 ensemble members is still able to outperform all of the 4DVAR assimilations (Fig. 2).

The EAKF outperforms 4DVAR by using more complete information about the distribution of the prior. In addition to providing better estimates of the state, the EAKF also provides information about the uncertainty in this estimate through the ensemble as discussed in section 3. Note that recent work by Hansen and Smith (2001) suggests that combining the capabilities of 4DVAR and ensemble filters may lead to a hybrid that is superior to either. Other enhancements to the 4DVAR algorithm could also greatly enhance its performance. Still, these results suggest that the EAKF should be seriously considered as an alternative to 4DVAR algorithms in a variety of applications.

6. Ease of implementation and performance

Implementing the EAKF (or the EnKF) requires little in addition to a forecast model and a description of the observing system. The implementation of the filtering code described here makes use of only a few hundred lines of Fortran-90 in addition to library subroutines to compute standard matrix and statistical operations. There is no need to produce a linear tangent or adjoint model [a complicated task for large models; Courtier et al. (1993)] nor are any of the problems involved with the definition of linear tangents in the presence of discontinuous physics an issue (Vukicevic and Errico 1993; Miller et al. 1994b) as they are for 4DVAR methods.

The computational cost of the filters has two parts: production of an ensemble of model integrations, and computation of the filter products. Integrating the ensemble multiplies the cost of the single model integration used in some simple data assimilation schemes by a factor of N . In many operational atmospheric modeling settings, ensembles are already being integrated with more conventional assimilation methods so there may be no incremental cost for model integrations.

As implemented here, the cost of computing the filter products at one observation time is $O(mnN)$, where m is the number of observations, n is the size of the model, and N is the ensemble size. The impact of each observation on each model variable is evaluated separately here. The computation for a given observation and state variable requires computing the 2×2 sample covariance matrix of the state variable and the prior ensemble observation, an $O(N)$ operation repeated $O(mn)$ times. In addition, several matrix inverses and singular value decompositions for 2×2 matrices are required (cost is not a function of m , n , or N). The computation of the prior ensembles of observed variables for the joint state-observation vector is also required, at a cost of $O(m)$. It is difficult to envision an *ensemble* scheme that has a more favorable computational scaling than the $O(mnN)$ for the methods applied here. The cost of the ensemble Kalman filter scales in an identical fashion as noted by Houtekamer and Mitchell (2001).

7. Filter assimilation in barotropic model

The limitations of the resampling filter in AA made it impossible to apply to large systems with reasonable ensemble sizes. In this section, an initial application of the EAKF to a larger model is described. The model is a barotropic vorticity equation on the sphere, represented as spherical harmonics with a T42 truncation (appendix C). The assimilation uses the streamfunction in physical space on a 64-point latitude by 128-point longitude grid (total of 8192 state variables).

The first case examined is a perfect model assimilation in which a long control run of the T42 model is used as the truth. To maintain variability, the model is forced as noted in the appendix. Observations of stream-

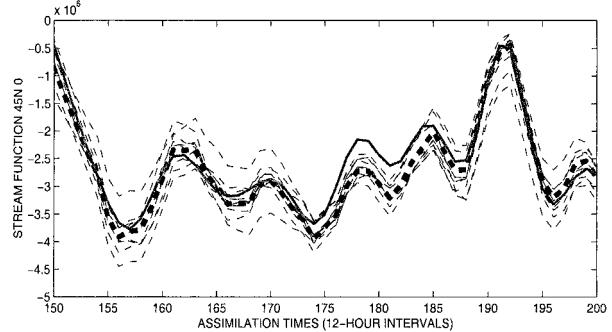


FIG. 10. Time series of “truth” from long control run (solid gray), ensemble mean from ensemble adjustment Kalman filter assimilation in global barotropic model (thick dashed), and the first 10 of the 20 individual ensemble members (thin dashed) for streamfunction at 45°N , 0° . Observations are available every 12 h and consist of 250 points placed randomly on the surface of the sphere excluding the longitude belt from 60° to 160°E where there were no observations; the observational error standard deviation was $1 \times 10^6 \text{ m}^2 \text{ s}^{-1}$.

function are available every 12 h at 250 randomly chosen locations on the surface of the sphere excluding the longitude belt between 60°E and 160°E where there are no observations. An observational error with standard deviation $1 \times 10^6 \text{ m}^2 \text{ s}^{-1}$ is added independently to each observation. A covariance inflation factor of 1.01 is used with a 20-member ensemble. In addition, only observations within 10° of latitude and $\cos^{-1}(\text{lat}) \times 10^{\circ}$ of longitude are allowed to impact any particular state variable. This limitation is qualitatively identical to the cutoff radius employed by Houtekamer and Mitchell (1998). In later work, Houtekamer and Mitchell (2001) report that their use of a cutoff radius when using an EnKF leads to discontinuities in the analysis. Here, this behavior was not observed, presumably because the EAKF does not introduce the noise that can impact correlations in the EnKF and because state variables that are adjacent on the grid are impacted by sets of observations that have a relatively large overlap. One could implement smoother algorithms for limiting the spatial range of impacts of an observation as was done with the correlation function in the L96 results in earlier sections.

Figure 10 shows time series of the truth, the ensemble mean, and the first 10 ensemble members for a grid point near 45°N , 0° . Figure 11 shows the corresponding rms error of the ensemble mean and the ensemble spread for the same variable. The rms streamfunction error is consistently much less than the observational error standard deviation, even though only 250 observations are available. The truth generally stays inside the first 10 ensemble members in Fig. 10. The chi-square statistic for the bins over the 100 observation time interval from times 100 to 200 is 30.19, corresponding to a 93% chance that the truth was not picked from the same distribution as the ensemble. In general, for this assimilation, a sample of 100 observation times is enough to distinguish the truth from the ensemble at about the 90%

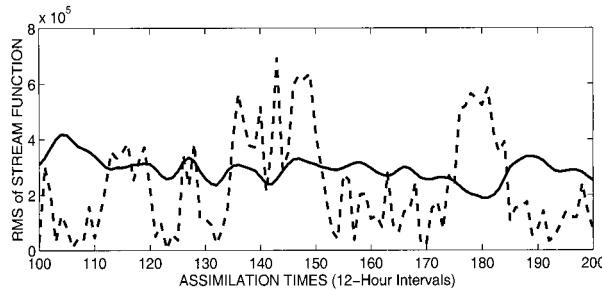


FIG. 11. Time series of rms error of ensemble mean from ensemble adjustment Kalman filter assimilation (dashed) and mean rms difference between ensemble members and the ensemble mean (spread, solid) for streamfunction at 45°N , 0° for the same assimilation as in Fig. 10.

confidence level. The normalized rms ratio r is 1.026 indicating that in general this ensemble assimilation is somewhat too confident (too little spread).

Figure 12 plots the error of the ensemble mean streamfunction field at assimilation time 200. All shaded areas have error magnitude less than the observational standard deviation. The largest errors are in the region between 60°E and 160°E where there are no observations. The areas of smallest error are concentrated in areas distant from and generally in regions upstream from the data void.

As noted in section 3, it is important to know something about the error growth of the model when the data assimilation is turned off in order to be able to judge the value of the assimilation method. For this barotropic model, the ensemble mean rms error doubles in about 10 days.

The second case examined in this section uses the same T42 model (with the weak climatological forcing removed) to assimilate data from the National Centers for Environmental Prediction (NCEP) operational analyses

for the winter of 1991/92. The “observations” are available once a day as T21 truncated spherical harmonics and are interpolated to the Gaussian grid points of the T42 model being used. This interpolation is regarded as the truth and observations are taken at each grid point by adding observational noise with a standard deviation of $1 \times 10^6 \text{ m}^2 \text{ s}^{-1}$. This is a particularly challenging problem for the EAKF because the T42 model has enormous systematic errors at a lead time of 24 h. The result is that the impact of the observations is large while the EAKF is expected to work best when the impact of observations is relatively small (see section 2c).

In addition, the EAKF as described to date assumes that the model being used has no systematic errors. That is obviously not the case here and a direct application of the filter method as described above does not work well. A simple modification of the filter to deal with model systematic error is to include an additional parameter that multiplies the prior covariance, Σ_p , only when it is used in (14) to compute the updated mean. Setting this factor to a value greater than 1 indicates that the prior estimate of the position of the mean should not be regarded as being as confident as the prior ensemble spread would indicate. In the assimilation shown here, this factor is set to 1.02. A covariance inflation factor must also continue to be used. Because error growth in the T42 barotropic model is much slower than that in the atmosphere, this factor is much larger here than in the perfect model cases and serves to correct the systematic slowness of uncertainty growth in the assimilating model. Covariance inflation is set to 1.45 here.

Figure 13 shows a time series of the truth, ensemble mean, and the first 10 ensemble members from the T42 assimilation of NCEP data for streamfunction near 45°N , 0°E , the same point shown in the perfect model results earlier in this section. The ensemble assimilation

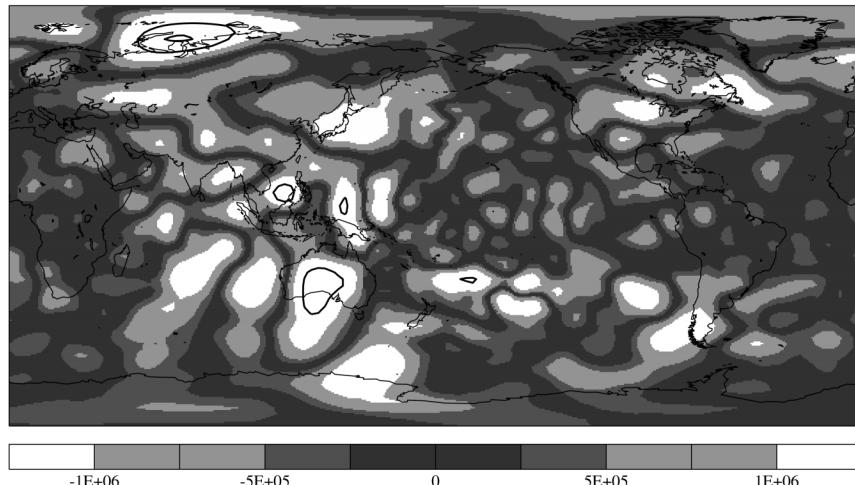


FIG. 12. Error of ensemble mean of assimilation at assimilation step 200 for the same assimilation as in Fig. 10. In addition to shading, contours are plotted with an interval of 1×10^6 for absolute values greater than 1×10^6 .

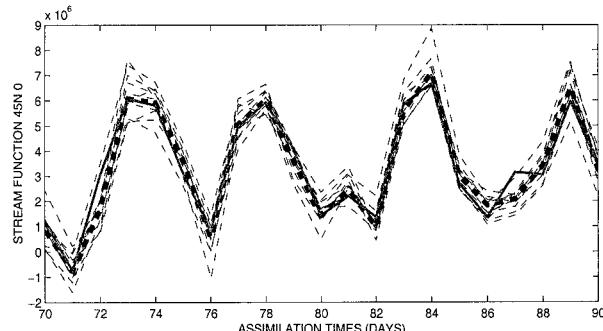


FIG. 13. Time series of “truth” from NCEP analyses (solid gray), ensemble mean from ensemble adjustment Kalman filter assimilation (thick dashed), and the first 10 of the 20 individual ensemble members (thin dashed) for streamfunction at 45°N , 0° from a T42 barotropic model. Observations are available at each model grid point once per day with observational error standard deviation $1 \times 10^6 \text{ m}^2 \text{ s}^{-1}$.

clearly tracks the observed data, which have much higher amplitude and frequency temporal variability than is seen in the perfect model in Fig. 10. Although the truth frequently falls within the 10 ensemble members, this variable has a chi-square statistic of 46.00, which gives 99% confidence that the truth is not drawn from the same distribution as the ensemble given 100 days of assimilation starting on 11 November 1991. Given the low quality of the model, these results still seem to be reasonably good. Figure 14 plots the error of the ensemble mean on 19 February 1992, a typical day. All shaded areas have ensemble mean error less than the observational error standard deviation with dark shaded regions having less than 25% of this error. These results give some encouragement that practical assimilation schemes for operational applications could be obtained if the EAKF were applied with a more realistic forecast model and more frequent observations.

8. Conclusions and future development

The EAKF can do viable data assimilation and prediction in models where the state space dimension is large compared to the ensemble size. It has an ability to assimilate observations with complex nonlinear relations to the state variables and has extremely favorable computational scaling for large models. At least in low-order models, the EAKF compares quite favorably to the four-dimensional variational method, producing assimilations with smaller error and also providing information about the distribution of the assimilation. Unlike variational methods, the EAKF does not require the use of linear tangent and adjoint model codes and so is straightforward to implement, at least mechanistically, in any prediction model. The EAKF is similar in many ways to the EnKF, but uses a different algorithm for updating the ensemble when observations become available. The EnKF introduces noise by forming a random sample of the observational error distribution and this noise has an adverse impact on the quality of assimilations produced by the EnKF.

It is possible that additional heuristic modifications to the EnKF could make it more competitive with the EAKF. Comparing the EAKF to other methods in large models is impossible at present. Both of these points underscore the need to develop some sort of data assimilation testbed facility that allows experts to do fair comparisons of the many assimilation techniques that are under development.

The EAKF can be extended to a number of other interesting problems. The version of the filter used here is currently being used in a study of adaptive observing systems (Berliner et al. 1999; Palmer et al. 1998). Just as the ensemble can provide estimates of the joint distribution of model state variables and observed variables, it can also provide estimates of joint distributions

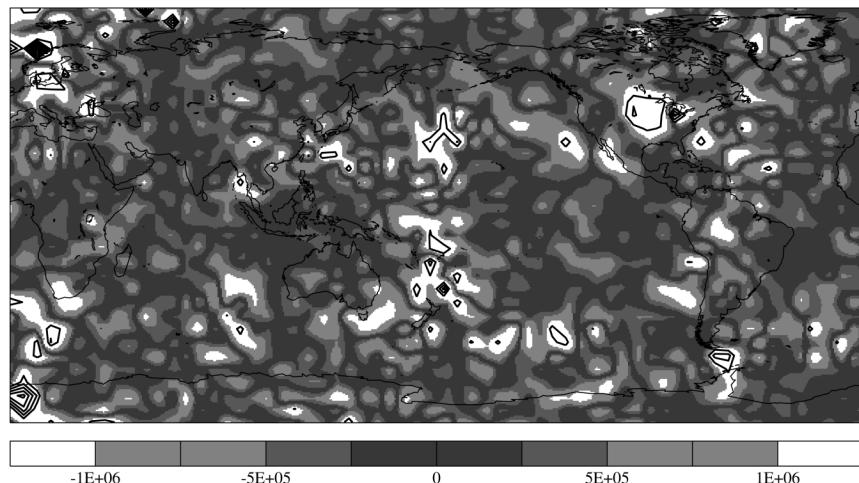


FIG. 14. Error of ensemble mean of assimilation at day 200 for the same assimilation as in Fig. 13. In addition to shading, contours are plotted with an interval of 1×10^6 for absolute values greater than 1×10^6 .

of the model state at earlier times with the state at the present time. Likewise, joint distributions of the state variables at different forecast times can be produced. These joint distributions can be used to examine the impact of observations at previous times, or during a forecast, on the state distribution at later times, allowing one to address questions about the potential value of additional observations (Bishop and Toth 1999). In a similar spirit, the ensemble filter provides a potentially powerful context for doing observing system simulation experiments (for instance Kuo et al. 1998).

Another product of the filter assimilation is estimates of the covariances between state variables or state variables and observations (Ehrendorfer and Tribbia 1997). These estimates are similar to those that are required for simpler data assimilation schemes like optimal interpolation but also may be useful for theoretical understanding of the dynamics of the atmosphere (Bouttier 1993). Time and spatial mean estimates of prior joint state–observation covariances could be generated through an application of the EAKF over a limited time and then used as input to a less computationally taxing three-dimensional variational technique. Initial tests of this method in a barotropic model have been promising.

Despite the encouraging results presented here, there are a number of issues that must still be addressed before the EAKF could be extended to application in operational atmospheric or oceanic assimilation. The most serious problem appears to be dealing with model uncertainty in a systematic way. In the work presented here, the covariance inflation factor has been used to prevent model prior estimates from becoming unrealistically confident. The current implementation works well in perfect model assimilations with homogeneous observations (observations of the same type distributed roughly uniformly in space), but begins to display some undesirable behavior with heterogeneous observations. In the barotropic model with a data void this was reflected as an inability to produce good rms ratios in both the observed and data-void areas. Reducing the covariance inflation factor when the spread for a state variable becomes large compared to the climatological standard deviation (not done in the results displayed here) solves this problem. Another example of this problem occurs when observations of both temperature and wind speed are available in primitive equation model assimilations. Clearly, a more theoretically grounded method for dealing with model uncertainty is needed. Nevertheless, the covariance inflation approach does have a number of desirable features that need to be incorporated in a more sophisticated approach. Operational atmospheric models tend to have a number of balances that constrain the relation between different state variables. If the problem of model uncertainty is dealt with in a naive fashion by just introducing some unstructured noise to the model, these balance requirements are ignored. As an example, in primitive equation applications, this results in excessive gravity wave noise

in the assimilation (Anderson 1997). The covariance inflation approach maintains existing linear relations between state variables and, so, produces far less gravity wave noise in primitive equation tests to date. The EnKF introduces noise when computing the impact of observations on the prior state and this noise may also lead to increased gravity wave noise in assimilations.

Dealing with the more serious model errors that occur in assimilation of observed atmospheric data requires even more careful thought. Introducing an additional parameter that controls the confidence placed in prior estimates of the mean is able to deal with a number of model biases, but a more theoretically grounded approach would be desirable.

Ongoing work with the EAKF is addressing these issues and gradually expanding the size and complexity of the assimilating models. Initial results with coarse-resolution dry primitive equation models are to be extended to higher resolutions with moist physics. The filter is also scheduled to be implemented in the Geophysical Fluid Dynamics Laboratory Modular Ocean Model for possible use in producing initial conditions for seasonal forecast integrations of coupled models.

Acknowledgments. The author would like to thank Tony Rosati, Matt Harrison, Steve Griffies, and Shaoqing Zhang for their comments on earlier versions of this manuscript. The two anonymous reviewers were unusually helpful in finding errors and improving the originally submitted draft. Conversations with Michael Tippett, Peter Houtekamer, Ron Errico, Tom Hamill, Jeff Whitaker, Lenny Smith, and Chris Snyder led to modifications of many of the ideas underlying the filter. Jim Hansen's extremely comprehensive review found several key errors in the appendix. Finally, this work would never have proceeded without the insight and encouragement of Stephen Anderson.

APPENDIX A

Ensemble Adjustment

This appendix describes a general implementation of the EAKF; refer to the last paragraph of section 2 for details on how this method is applied in a computationally affordable fashion. Let $\{\mathbf{z}_i^p\}$, $(i = 1, \dots, N)$ be a sample of the prior distribution at a time when new observations become available with the subscript referring to each member of the sample (an N -member ensemble of state vectors). The prior sample mean and covariance are defined as $\bar{\mathbf{z}}^p$ and $\Sigma = \Sigma^p$. Assume that $\mathbf{H}^T \mathbf{R}^{-1} \mathbf{y}^o$ and $\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}$ are available at this time with \mathbf{y}^o the observations vector, \mathbf{R} the observational error covariance, and \mathbf{H} the linear operator that produces the observations given a joint state vector.

Since Σ is symmetric, a singular value decomposition gives $\mathbf{D}^p = \mathbf{F}^T \Sigma \mathbf{F}$, where \mathbf{D}^p is a diagonal matrix with the singular values, μ^p of Σ on the diagonal and \mathbf{F} is a

unitary matrix ($\mathbf{F}^T \mathbf{F} = \mathbf{I}$, $\mathbf{F}^{-1} = \mathbf{F}^T$, $(\mathbf{F}^T)^{-1} = \mathbf{F}$). Applying \mathbf{F}^T and \mathbf{F} in this fashion is a rotation of Σ to a reference frame in which the prior sample covariance is diagonal.

Next, one can apply a scaling in this rotated frame in order to make the prior sample covariance the identity. The matrix $(\mathbf{G}^T)^{-1} \mathbf{F}^T \Sigma \mathbf{F} \mathbf{G}^{-1}$, where \mathbf{G} is a diagonal matrix with the square root of the singular values, μ^p , on the diagonal, is the identity matrix, \mathbf{I} .

Next, a singular value decomposition can be performed on the matrix $\mathbf{G}^T \mathbf{F}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{F} \mathbf{G}$; this is a rotation to a reference frame in which the scaled inverse obser-

vational “covariance” matrix, $\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}$, is a diagonal matrix, $\mathbf{D} = \mathbf{U}^T \mathbf{G}^T \mathbf{F}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{F} \mathbf{G} \mathbf{U}$, with the diagonal elements the singular values, μ . The prior covariance can also be moved to this reference frame, and it is still the identity since \mathbf{U} is unitary, $\mathbf{I} = \mathbf{U}^T (\mathbf{G}^T)^{-1} \mathbf{F}^T \Sigma \mathbf{F} \mathbf{G}^{-1} \mathbf{U}$.

The updated covariance can be computed easily in this reference frame since the prior covariance inverse is just \mathbf{I} and the observed covariance inverse is diagonal. The updated covariance can then be moved back to the original reference frame by unrotating, unscaling, and unrotating. (Note that \mathbf{G} is symmetric.)

More formally, the updated covariance can be evaluated as

$$\begin{aligned}\Sigma^u &= (\Sigma^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} = (\mathbf{F}^T)^{-1} \mathbf{G}^T (\mathbf{U}^T)^{-1} [\mathbf{U}^T (\mathbf{G}^T)^{-1} \mathbf{F}^T (\Sigma^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{F} \mathbf{G}^{-1} \mathbf{U}] \mathbf{U}^{-1} \mathbf{G} \mathbf{F}^{-1} \\ &= (\mathbf{F}^T)^{-1} \mathbf{G}^T (\mathbf{U}^T)^{-1} \{ [\mathbf{U}^{-1} \mathbf{G} \mathbf{F}^{-1} (\Sigma^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}) (\mathbf{F}^T)^{-1} \mathbf{G}^T (\mathbf{U}^T)^{-1}]^{-1} \} \mathbf{U}^{-1} \mathbf{G} \mathbf{F}^{-1} \\ &= (\mathbf{F}^T)^{-1} \mathbf{G}^T (\mathbf{U}^T)^{-1} \{ [\mathbf{U}^{-1} \mathbf{G} \mathbf{F}^{-1} \Sigma^{-1} (\mathbf{F}^T)^{-1} \mathbf{G}^T (\mathbf{U}^T)^{-1} + \mathbf{U}^{-1} \mathbf{G} \mathbf{F}^{-1} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} (\mathbf{F}^T)^{-1} \mathbf{G}^T (\mathbf{U}^T)^{-1}]^{-1} \} \mathbf{U}^{-1} \mathbf{G} \mathbf{F}^{-1} \\ &= (\mathbf{F}^T)^{-1} \mathbf{G}^T (\mathbf{U}^T)^{-1} \{ [(\mathbf{U}^T (\mathbf{G}^T)^{-1} \mathbf{F}^T \Sigma \mathbf{F} \mathbf{G}^{-1} \mathbf{U})^{-1} + \mathbf{U}^T \mathbf{G}^T \mathbf{F}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{F} \mathbf{G} \mathbf{U}]^{-1} \} \mathbf{U}^{-1} \mathbf{G} \mathbf{F}^{-1}.\end{aligned}$$

The first term inside the square brackets is just \mathbf{I} , and the second is $\text{diag}[\mu_1, \mu_2, \dots]$, so the term inside the curly brackets is $\text{diag}[1/(1 + \mu_1), 1/(1 + \mu_2), \dots]$. This can be rewritten as $\mathbf{B}^T (\mathbf{G}^T)^{-1} \mathbf{F}^T \Sigma \mathbf{F} \mathbf{G}^{-1} \mathbf{B}$, where

$$\mathbf{B} = \text{diag}[(1 + \mu_1)^{-1/2}, (1 + \mu_2)^{-1/2}, \dots].$$

Then, $\Sigma^u = \mathbf{A} \Sigma \mathbf{A}^T$, where $\mathbf{A} = (\mathbf{F}^T)^{-1} \mathbf{G}^T (\mathbf{U}^T)^{-1} \mathbf{B}^T (\mathbf{G}^T)^{-1} \mathbf{F}^T$.

The mean of the updated distribution needs to be calculated to compute the \mathbf{z}_i^u . Once the updated sample covariance has been computed as outlined above, the mean is calculated easily as $\bar{\mathbf{z}}^u = \Sigma^u (\Sigma^{-1} \bar{\mathbf{z}}^p + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{y}^o)$. For computational efficiency, Σ^{-1} can be computed by transforming back from the rotated sample singular value decomposition (SVD) space in which it is diagonal.

As noted above, being able to write $\Sigma^u = \mathbf{A} \Sigma \mathbf{A}^T$ enables an update of the prior sample, $\{\mathbf{z}_i^p\}$, to get an updated sample, $\{\mathbf{z}_i^u\}$ as

$$\mathbf{z}_i^u = \mathbf{A}^T (\mathbf{z}_i^p - \bar{\mathbf{z}}^p) + \bar{\mathbf{z}}^u.$$

An understanding of this update process follows from the discussion above. After applying the rotation, scaling, and rotation operators \mathbf{U}^T , $(\mathbf{G}^T)^{-1}$, and \mathbf{F}^T to the prior sample, it is in a space where the prior sample covariance is \mathbf{I} and the observational covariance is diagonal. One can then just “shrink” the prior ensemble by the factor $1/(1 + \mu_i)$ independently in each direction to get a new sample with the updated covariance in this frame. The rotations and scaling can then be inverted to get the final updated ensemble.

If the sample prior covariance matrix is degenerate (for instance if the ensemble size, N , is smaller than the size of the state vectors), then there are directions in the state space in which the ensemble has no variance.

Applying the SVD to such sample covariance matrices actually results in a set of $m < N$ nonzero singular values and $N - m$ zeros on the diagonal of \mathbf{D}^p . All the computations can then be performed in the m -dimensional subspace spanned by the singular vectors corresponding to the m nonzero singular values. In addition, there may be some set of singular values that are very small but nonzero. If care is used, these directions can also be neglected in the computation for further savings.

APPENDIX B

The Lorenz 1996 Model

The L96 model is a variable size low-order dynamical system used by Lorenz (1996) and more recently by others including Lorenz and Emanuel (1998). The model has N state variables, X_1, X_2, \dots, X_N , and is governed by the equation

$$dX_i/dt = (X_{i+1} - X_{i-2})X_{i-1} - X_i + F,$$

where $i = 1, \dots, N$ with cyclic indices. The results shown are for parameters with a sensitive dependence on initial conditions: $N = 40$, $F = 8.0$, and a fourth-order Runge–Kutta time step with $dt = 0.05$ is applied as in Lorenz and Emanuel.

APPENDIX C

Nondivergent Barotropic Model

A spherical harmonic model of the nondivergent barotropic vorticity equation on the sphere is used with a transform method for nonlinear terms performed on a nonaliasing physical space grid with 128 longitude

points and 64 Gaussian latitude points for a total of 8192 grid points. A time step of 1800 s is used with a third-order Adams–Bashforth time step, which is initialized with a single forward step followed by a single leapfrog step. A ∇^8 diffusion on the streamfunction is applied with a constant factor so that the smallest resolved wave is damped with an *e*-folding time of 2 days. When run in a perfect model setting, a forcing must be added to the model to induce interesting long-term variability. In this case, the zonal flow spherical harmonic components are relaxed toward the observed time mean zonal flow for the period November through March 1991–92, with an *e*-folding time of approximately 20 days.

REFERENCES

- Anderson, J. L., 1996: A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Climate*, **9**, 1518–1530.
- , 1997: The impact of dynamical constraints on the selection of initial conditions for ensemble predictions: Low-order perfect model results. *Mon. Wea. Rev.*, **125**, 2969–2983.
- , and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741–2758.
- Barker, T. W., 1991: The relationship between spread and forecast error in extended-range forecasts. *J. Climate*, **4**, 733–742.
- Berliner, L. M., Z.-Q. Lu, and C. Snyder, 1999: Statistical design for adaptive weather observations. *J. Atmos. Sci.*, **56**, 2536–2552.
- Bishop, C. H., and Z. Toth, 1999: Ensemble transformation and adaptive observations. *J. Atmos. Sci.*, **56**, 1748–1765.
- , B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436.
- Bouttier, F., 1993: The dynamics of error covariances in a barotropic model. *Tellus*, **45A**, 408–423.
- Brasseur, P., J. Ballabriga-Poy, and J. Verron, 1999: Assimilation of altimetric data in the mid-latitude oceans using the singular evolutive extended Kalman filter with an eddy-resolving, primitive equation model. *J. Mar. Syst.*, **22**, 269–294.
- Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724.
- Courtier, P., J. Derber, R. Errico, J.-F. Louis, and T. Vukicevic, 1993: Important literature on the use of adjoint, variational methods and the Kalman filter in meteorology. *Tellus*, **45A**, 342–357.
- Derber, J. C., 1989: A variational continuous assimilation technique. *Mon. Wea. Rev.*, **117**, 2437–2446.
- Ehrendorfer, M., 1994: The Liouville equation and its potential usefulness for the prediction of forecast skill. Part I: Theory. *Mon. Wea. Rev.*, **122**, 703–713.
- , and J. J. Tribbia, 1997: Optimal prediction of forecast error covariances through singular vectors. *J. Atmos. Sci.*, **54**, 286–313.
- Evans, R. E., M. S. J. Harrison, R. J. Graham, and K. R. Mylne, 2000: Joint medium-range ensembles from The Met. Office and ECMWF systems. *Mon. Wea. Rev.*, **128**, 3104–3127.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10 143–10 162.
- , and P. J. van Leeuwen, 1996: Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasigeostrophic model. *Mon. Wea. Rev.*, **124**, 85–96.
- Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757.
- Goerss, J. S., 2000: Tropical cyclone track forecasts using an ensemble of dynamical models. *Mon. Wea. Rev.*, **128**, 1187–1193.
- Gordeau, L., J. Verron, T. Delcroix, A. J. Busalacchi, and R. Murtugudde, 2000: Assimilation of TOPEX/POSEIDON altimetric data in a primitive equation model of the tropical Pacific Ocean during the 1992–1996 El Niño–Southern Oscillation period. *J. Geophys. Res.*, **105**, 8473–8488.
- Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, 2776–2790.
- Hansen, J. A., and L. A. Smith, 2001: Probabilistic noise reduction. *Tellus*, in press.
- Harrison, M. S. J., T. N. Palmer, D. S. Richardson, and R. Buizza, 1999: Analysis and model dependencies in medium-range ensembles: Two transplant case-studies. *Quart. J. Roy. Meteor. Soc.*, **125A**, 2487–2515.
- Houtekamer, P. L., and L. Lefavre, 1997: Using ensemble forecasts for model validation. *Mon. Wea. Rev.*, **125**, 2416–2426.
- , and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.
- , and —, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123–137.
- , L. Lefavre, and J. Derome, 1995: The RPN Ensemble Prediction System. *Proc. ECMWF Seminar on Predictability*, Vol. II, Reading, United Kingdom, ECMWF, 121–146.
- , —, —, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**, 1225–1242.
- Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory*. Academic Press, 376 pp.
- Kaplan, A., Y. Kushnir, M. A. Cane, and M. B. Blumenthal, 1997: Reduced space optimal analysis for historical data sets: 136 years of Atlantic sea surface temperatures. *J. Geophys. Res.*, **102**, 27 835–27 860.
- Kepenne, C. L., 2000: Data assimilation into a primitive-equation model with a parallel ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 1971–1981.
- Krishnamurti, T. N., C. M. Kishtawal, T. E. Larow, D. R. Bachiuchi, Z. Zhang, C. E. Williford, S. Gadgil, and S. Surendran, 1999: Improved weather and seasonal climate forecasts from multi-model superensemble. *Science*, **285**, 1548–1550.
- Kuo, Y.-H., X. Zou, and W. Huang, 1998: The impact of Global Positioning System data on the prediction of an extratropical cyclone: An observing system simulation experiment. *Dyn. Atmos. Oceans*, **27**, 439–470.
- Le Dimet, F.-X., and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus*, **38A**, 97–110.
- Lermusiaux, P. F., and A. R. Robinson, 1999: Data assimilation via error subspaces statistical estimation. Part I: Theory and schemes. *Mon. Wea. Rev.*, **127**, 1385–1407.
- Lorenc, A. C., 1997: Development of an operational variational assimilation scheme. *J. Meteor. Soc. Japan*, **75**, 229–236.
- Lorenz, E. N., 1996: Predictability: A problem partly solved. *Proc. ECMWF Seminar on Predictability*, Vol. I, Reading, United Kingdom, ECMWF, 1–18.
- , and K. A. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.*, **55**, 399–414.
- Miller, R. N., M. Ghil, and F. Gauthiez, 1994a: Advanced data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sci.*, **51**, 1037–1056.
- , E. D. Zaron, and A. F. Bennet, 1994b: Data assimilation in models with convective adjustment. *Mon. Wea. Rev.*, **122**, 2607–2613.
- Mitchell, H. L., and P. L. Houtekamer, 2000: An adaptive ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 416–433.
- Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–120.

- Murphy, J. M., 1988: The impact of ensemble forecasts on predictability. *Quart. J. Roy. Meteor. Soc.*, **114**, 463–493.
- , 1990: Assessment of the practical utility of extended range ensemble forecasts. *Quart. J. Roy. Meteor. Soc.*, **116**, 89–125.
- Palmer, T. N., R. Gelaro, J. Barkmeijer, and R. Buizza, 1998: Singular vectors, metrics, and adaptive observations. *J. Atmos. Sci.*, **55**, 633–653.
- Rabier, F., J.-N. Thepaut, and P. Courtier, 1998: Extended assimilation and forecast experiments with a four-dimensional variational assimilation system. *Quart. J. Roy. Meteor. Soc.*, **124**, 1–39.
- Richardson, D. S., 2000: Skill and relative economic value of the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **126**, 649–667.
- Swanson, K., R. Vautard, and C. Pires, 1998: Four-dimensional variational assimilation and predictability in a quasi-geostrophic model. *Tellus*, **50**, 369–390.
- Tarantola, A., 1987: *Inverse Problem Theory*. Elsevier Science, 613 pp.
- Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330.
- , and —, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319.
- Tracton, M. S., and E. Kalnay, 1993: Operational ensemble prediction at the National Meteorological Center: Practical aspects. *Wea. Forecasting*, **8**, 379–398.
- Tziperman, E., and Z. Sirkes, 1997: Using the adjoint method with the ocean component of coupled ocean–atmosphere models. *J. Meteor. Soc. Japan*, **75**, 353–360.
- van Leeuwen, P. J., 1999: Comments on “Data assimilation using an ensemble Kalman filter technique.” *Mon. Wea. Rev.*, **127**, 1374–1377.
- , and G. Evensen, 1996: Data assimilation and inverse methods in terms of a probabilistic formulation. *Mon. Wea. Rev.*, **124**, 2898–2912.
- Vukicevic, T., and R. M. Errico, 1993: Linearization and adjoint of parameterized moist diabatic processes. *Tellus*, **45A**, 493–510.
- Ziehmann, C., 2000: Comparison of a single-model EPS with a multi-model ensemble consisting of a few operational models. *Tellus*, **52A**, 280–299.
- Zupanski, D., 1997: A general weak constraint applicable to operational 4DVAR data assimilation systems. *Mon. Wea. Rev.*, **125**, 2274–2292.