A Local Least Squares Framework for Ensemble Filtering

JEFFREY L. ANDERSON

NOAA/GFDL, Boulder, Colorado

ABSTRACT

Many methods using ensemble integrations of prediction models as integral parts of data assimilation have appeared in the atmospheric and oceanic literature. In general, these methods have been derived from the Kalman filter and have been known as ensemble Kalman filters. A more general class of methods including these ensemble Kalman filter methods is derived starting from the nonlinear filtering problem. When working in a joint state–observation space, many features of ensemble filtering algorithms are easier to derive and compare. The ensemble filter methods derived here make a (local) least squares assumption about the relation between prior distributions of an observation variable and model state variables. In this context, the update procedure applied when a new observation becomes available can be described in two parts. First, an update increment is computed for each prior ensemble estimate of the observation variable by applying a scalar ensemble filter. Second, a linear regression of the prior ensemble sample of each state variable on the observation variable is performed to compute update increments for each state variable ensemble member from corresponding observation variable increments. The regression can be applied globally or locally using Gaussian kernel methods.

Several previously documented ensemble Kalman filter methods, the perturbed observation ensemble Kalman filter and ensemble adjustment Kalman filter, are developed in this context. Some new ensemble filters that extend beyond the Kalman filter context are also discussed. The two-part method can provide a computationally efficient implementation of ensemble filters and allows more straightforward comparison of methods since they differ only in the solution of a scalar filtering problem.

1. Introduction

Interest in data assimilation methods using ensemble integrations of prediction models is growing rapidly in the atmospheric and oceanic communities. This is occurring because ensemble assimilation methods are maturing rapidly and because both prediction centers and research groups are becoming increasingly interested in characterizing more information about the probability distribution of the climate system than can be revealed by a single assimilated state estimate.

Ensemble assimilation methods were originally developed as computationally feasible approximate solutions of the nonlinear filtering problem patterned after the Kalman filter (Kalman and Bucy 1961; Courtier et al. 1993). This led to a sequence of related methods known as ensemble Kalman filters (Evensen 1994), which have been extended to increasingly general assimilation problems (Houtekamer and Mitchell 1998). More recently, other variants, still referred to as ensemble Kalman filters (Bishop et al. 2001; Anderson 2001; Pham 2001), have appeared in the literature demonstrating improved assimilation error characteristics and/

or decreased computational cost. Some of these were developed directly from the probabilistic statement of the nonlinear filtering problem, rather than starting from the Kalman filter. Developing filters in this context can lead to a more straightforward understanding of their capabilities for those not intimately related with the intricacies of the Kalman filter.

Here, a framework is developed in which many of the ensemble Kalman filter methodologies documented to date can be described while still supporting a more general class of ensemble filters. The derivation begins with the nonlinear filtering problem and applies a sequence of simplifying assumptions. The introduction of a joint state–observation space (Tarantola 1987) leads to an ability to deal with observations related to the model state variables by nonlinear functions. A least squares assumption (equivalent to assuming a local Gaussian relation among the prior joint state variables) has been made, sometimes indirectly, in many descriptions of ensemble Kalman filters. Here, that assumption is made explicitly and a significant simplification in the description of the algorithms results. Under the assumptions made here, the ensemble filter problem simplifies to an application of a nonlinear filter to a scalar, followed by a sequence of linear regressions. This simplification makes it easier to analyze the relative ca-

---

Corresponding author address: Dr. Jeffrey L. Anderson, NCAR/MMM, P.O. Box 3000, Boulder, CO 80307-3000.
E-mail: jla@ucar.edu

pabilities of a variety of ensemble filter implementations and can lead to reduced computational cost.

Section 2 derives this context for ensemble filtering and section 3 shows how several previously documented ensemble Kalman filters are related. Section 4 discusses details of methods for doing scalar assimilation problems while section 5 offers conclusions.

2. Ensemble filtering

To simplify notation, this section discusses only what happens at a single time at which observations become available. Discussion of how filter assimilations are advanced in time using ensemble methods and prediction models can be found in Anderson (2001, hereafter A01), Houtekamer and Mitchell (1998), and Jazwinski (1970). Basically, each ensemble member is integrated forward in time independently using a forecast model between times at which observations are available.

a. Joint state–observation space and Bayesian framework

The joint state–observation space (Tarantola 1987; A01) is defined by the joint space state vector:

$$\mathbf{z} = [\mathbf{x}, h(\mathbf{x})] = [\mathbf{x}, \mathbf{y}], \qquad (1)$$

where $\mathbf{x}$ is the model state vector; $\mathbf{y} = h(\mathbf{x})$, where h is the forward observation operator, defines the observations available at this time; and $\mathbf{z}$ is a vector of length $n + m$, where n is the number of state variables and m is the number of observations available at this time.

Using Bayesian statistics as in Jazwinski (1970) and A01, the distribution of the posterior (or updated) distribution $\mathbf{z}^u = [\mathbf{x}^u, \mathbf{y}^u]$ can be computed from the prior distribution, $\mathbf{z}^p = [\mathbf{x}^p, \mathbf{y}^p]$, as

$$\mathbf{p}(\mathbf{z}^u) = \mathbf{p}(\mathbf{y}^o \,|\, \mathbf{z}^p)\mathbf{p}(\mathbf{z}^p)/\text{normalization} \qquad (2)$$

where $\mathbf{y}^o$ is an $\mathbf{m}$-vector of the observed values available at this time. In the ensemble methods applied here, the normalization factor in the denominator is not used explicitly (Anderson and Anderson 1999).

One implication of (2) is that subsets of observations with independent (observational error) distributions can be assimilated sequentially. Let $\mathbf{y}^o$ be composed of s subsets of observations, $\mathbf{y}^o = \{ \mathbf{y}_1^o, \mathbf{y}_2^o, \ldots, \mathbf{y}_s^o \}$, where the distribution of the observational errors for observations in subset i is independent of the distribution for the observations in subset j, for $i \neq j$. Then

$$\mathbf{p}(\mathbf{y}^o \,|\, \mathbf{z}^p) = \prod_{i=1}^{s} \mathbf{p}(\mathbf{y}_i^o \,|\, \mathbf{z}^p). \qquad (3)$$

In particular, if the individual scalar observations in $\mathbf{y}^o$ have mutually independent error distributions, they can be assimilated sequentially in any order without changing the result in (2). This allows sequential assimilation with observational error distributions represented as Gaussians as long as the observational error covariance

matrix is diagonal. This was pointed out by Houtekamer and Mitchell (2001) in the ensemble Kalman filter context and used in A01. Equation (3) depends only on the observing system and makes no assumptions about the prior joint state distribution or how it is represented.

b. An ensemble method for the filtering problem

In ensemble methods for solving (2), information about the prior distribution of the state variables, $\mathbf{x}^p$, is available as a sample from N applications of a prediction model. An ensemble sample of the prior observation vector, $\mathbf{y}^p$, can be created by applying the forward observation operator, h, to each ensemble sample of $\mathbf{x}^p$.

Some Monte Carlo (ensemble) methods also have a weight, w, associated with each ensemble member. The possibility of weighted ensembles is not discussed in detail here, but the methods in this section are easily generalized to this case. Attempts to apply the most common types of weighting/resampling Monte Carlo algorithms in high-dimensional spaces have faced significant difficulties.

Observational error distributions of climate system observations are generally only poorly known and are often specified as Gaussian with zero mean (known instrument bias is usually corrected by removing the bias from the observation during a preprocessing step). Some observations have values restricted to certain ranges; for instance, precipitation must be positive. Redefining the observation variable as the log of the observation can lead to a Gaussian observational error distribution in this case (Tarantola 1987). Given Gaussian observational error distributions, observations can be decomposed into subsets where observational errors for observations in each subset are correlated but observational errors in different subsets are uncorrelated. In other words, $\mathbf{R}$, the observational error covariance matrix, is block diagonal with each block being the size of the number of observations in the corresponding observation subset. Error distributions for the different subsets are independent, so the subsets can be assimilated sequentially in (2) in an arbitrary order.

For many commonly assimilated observations, each scalar observation has an error distribution that is independent of all others, allowing each scalar observation to be assimilated sequentially (Houtekamer and Mitchell 2001). If the observational covariance matrix is not strictly diagonal, a singular value decomposition (SVD; equivalent to an eigenvalue decomposition for a symmetric positive-definite matrix like $\mathbf{R}$) can be performed on $\mathbf{R}$. The prior joint state ensembles can be projected onto the singular vectors and the assimilation can proceed using this new basis, in which $\mathbf{R}'$, the observational covariance matrix, is diagonal by definition. Upon completion of the assimilation computation, the updated state vectors can be projected back to the original state space. Given the application of this SVD, a mechanism for sequential assimilation of scalar observations im-

plies no loss of generality for observations with arbitrary Gaussian error distributions. In everything that follows, results are presented only for assimilation of a single scalar observation so that $m = 1$, with joint state space size $k = n + m = n + 1$. Allowing arbitrary observational error distributions represented as a sum of Gaussians is a straightforward extension to the methods described below.

c. Two-step data assimilation procedure

Following A01, define the joint state space forward observation operator for a single observation as the order $1 \times k$ linear operator $\mathbf{H} = [0, 0, \ldots, 0, 1]$. The expected value of the observation can be calculated by applying $\mathbf{H}$ to the joint state vector, $\mathbf{z}$, which is equivalent to applying the possibly nonlinear operator h to $\mathbf{x}$. The conversion of the possibly nonlinear h to the linear $\mathbf{H}$ is a primary motivation for applying ensemble filters in the joint state space.

The updated probability for the *marginal* distribution of the observation joint state variable, y, can be formed from Eq. (2) with the simple form

$$p_y(y^u) = \mathbf{p}(y^o \,|\, y^p)\mathbf{p}_y(y^p)/(\text{norm}), \qquad (4)$$

where the subscript on the probability densities indicates a marginal probability on the observation variable, y. The one-dimensional problem for this marginal distribution can be solved by a variety of methods, some of which are discussed in sections 3 and 4. Note that (4) does not depend on any of the model state variables.

This suggests a partitioning of the assimilation of an observation into two parts. The first determines updated ensemble members for the observation variable y given the observation, y^o. To update the ensemble sample of y^p, an increment, Δy_i, is computed for each ensemble member, $y_i^u = y_i^p + \Delta y_i$, $i = 1, \ldots, N$, where N is the ensemble size.

Given increments for the observation variable, the second step computes corresponding increments for each ensemble sample of each state variable, $\Delta x_{i,j}$ (i indexes the ensemble member and $j = 1, \ldots, k$ indexes which joint state variable throughout this report). This requires assumptions about the prior relationship between the joint state variables. Although reasonable alternatives exist (Tarantola 1987), the assumption used here is that the prior distribution is Gaussian (or a sum of Gaussians that allows generality). This is equivalent to assuming that a least squares fit (local least squares fit) to the prior ensemble members summarizes the relationship between the joint state variables.

Figure 1 depicts the simplest example in which there is only a single state variable, x. The observation variable, y, is related to x by the operator h, which is nonlinear in the figure. Increments for each ensemble sample of y have been computed. The corresponding increments for x are then computed by a global least squares fit (linear regression) so that
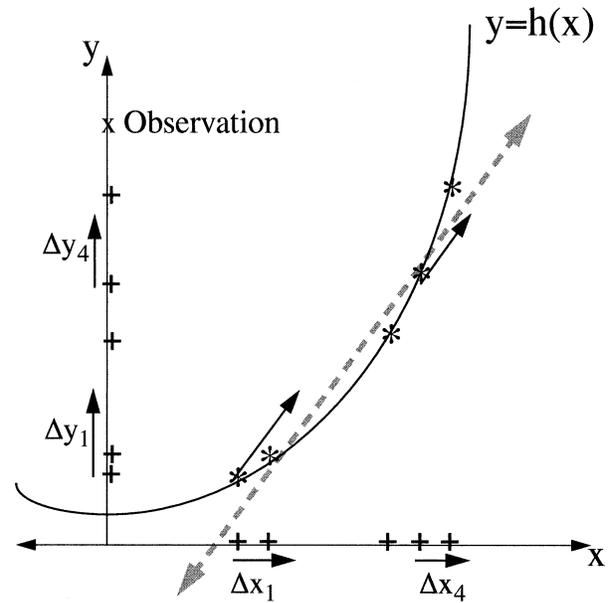
FIG. 1. An idealized representation showing the relation between update increments for a state variable, *x*, and an observation variable, *y*, for a five member ensemble represented by asterisks. The projection of the ensemble on the *x* and *y* axes is represented by a plus sign and the observation, y^o is represented by $\times$. In this case, *y* is functionally related to *x* by *h*. The gray dashed line shows a global least squares fit to the ensemble members. Update increments for ensemble members 1 and 4 for *y* are shown along with corresponding increments for the ensemble as a whole (thin vectors parallel to least squares fit) and for the *x* ensemble.

$$\Delta x_i = \frac{\sigma_{x,y}}{\sigma_{y,y}}\Delta y_i. \qquad (5)$$

The change in the ith ensemble sample of the state variable due to observation variable y is equal to the prior covariance of x with y, $\sigma_{x,y}$ divided by the prior variance of y, $\sigma_{y,y}$ times the change in the ith ensemble sample of the observation variable. This is just a statistical linearization and inversion of the observation operator h. This linearization can be done globally by computing the global sample covariance and using this for the regression for each ensemble member (Fig. 1).

The linearization can also be done locally (Fig. 2) by computing local estimates of covariance for each ensemble member. This can be done, for instance, by only using a set of nearest neighbors (in y, in x, or in some combined distance metric) to compute sample covariance. Figure 2 shows an idealized form of nearest neighbor linearization in which only a single closest ensemble member is used to compute the statistical linearization. Related methods for doing local Gaussian kernel approximations of this type can be found in Silverman (1986) and Bengtsson and Nychka (2001). When x is functionally related to y as in Fig. 2, local linearization methods like this can give significantly enhanced performance when h is strongly nonlinear over the prior ensemble range of y.
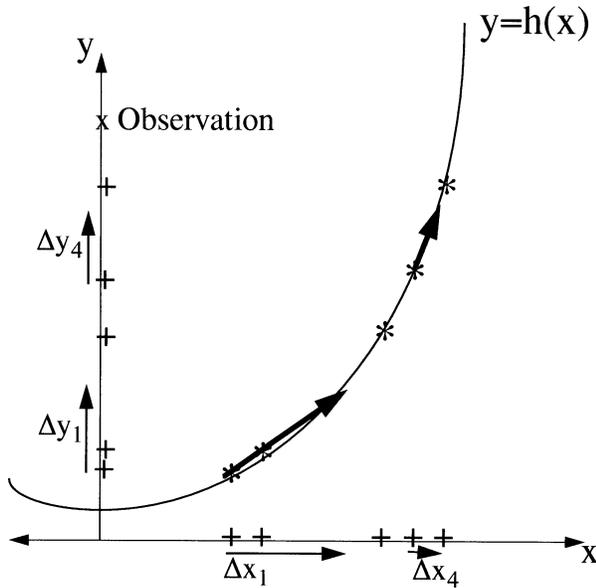
FIG. 2. As in Fig. 1 but showing the application of local least squares fits, in this case using only the nearest neighbor in *y*, to compute the updates for *x* given the updates for *y*. The local updates for the first and fourth ensemble members are shown by the black vectors.

FIG. 3. As in Fig. 1 but now $y = h(x_2)$, where x_2 is a second state variable that is moderately correlated with x_1. The thin dashed vector demonstrates the hazard of using local least squares fits when the observation variable *y* and the state variable x_1 are not functionally related.

If h is nonlinear as in the figures, the statistical linearization is only valid locally. To minimize errors due to the linearization, whether global or local linearizations are applied, it is desirable that the observation variable increments, Δy_i, should be as small as possible. This is discussed further in section 4c.

This two-step method can be extended trivially to problems with arbitrary numbers of state variables. When a global linearization is applied using a least squares fit, a single Gaussian is assumed to approximate the prior relation of the variables. The increments, $\Delta x_{i,j}$, for each ensemble sample of each state variable in terms of Δy_i can be computed independently by regression:

$$\Delta x_{i,j} = \frac{\sigma_{x_j,y}}{\sigma_{y,y}} \Delta y_i. \tag{6}$$

Again, local linearizations could be performed using Gaussian (or extended Gaussian) kernel methods (Tarantola 1987) in which only some subset of local information is used to compute the covariance from the ensemble sample. In (6), all relevant information about the prior covariance of the model state variables, **x**, needed to compute increments is contained in the correlation of the individual scalar state variables with the observation variable *y*.

When the state variable being updated and the observation variable are not functionally related, the use of local linearizations can be more problematic. Figure 3 shows an example where state variable x_1 is being updated by an observation, y^o. The expected value of the observation is $y = h(x_2)$, where x_2 is a second state

variable, here moderately correlated with x_1. In this case, the linear regression for x_1 performs a statistical linearization in the presence of noise. Using large (global) regressions is useful to filter out this noise. On the other hand, using local linearizations can help to resolve more of the structure of h. Applying local regressions that are based on too few ensemble members can lead to disastrous overfitting behavior as demonstrated by the application of an idealized single nearest-neighbor linearization in Fig. 3. Appropriate trade-offs in choosing local versus global linearizations are an important part of tuning ensemble filters for improved performance.

3. Relation to ensemble Kalman filters

A variety of ensemble Kalman filters have been described in the literature (see, e.g., Evensen and van Leeuwen 1996; Keppenne 2000; Mitchell and Houtekamer 2000; Pham 2001). Closely related methods for doing assimilation have been described by Lermursaix and Robinson (1999) and Miller et al. (1994). This section demonstrates that two of these, the perturbed observations ensemble Kalman filter (EnKF) and the ensemble adjustment Kalman filter (EAKF), can be recast in the two-step framework outlined in the previous section. At the heart of these ensemble Kalman filters is the fact that the product of the joint prior Gaussian with mean $\bar{\mathbf{z}}^p$, covariance Σ^p, and weight *w* and the Gaussian observation distribution with mean $\mathbf{y}^o$ and error variance **R** has covariance

$$\boldsymbol{\Sigma}^u = [(\boldsymbol{\Sigma}^p)^{-1} + \mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{H}]^{-1}, \tag{7}$$

mean

$$\bar{\mathbf{z}}^u = \boldsymbol{\Sigma}^u[(\boldsymbol{\Sigma}^p)^{-1}\bar{\mathbf{z}}^p + \mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{y}^o], \tag{8}$$

and an associated relative weight

$$D = \exp\{-[(\bar{\mathbf{z}}^p)^\mathrm{T}(\boldsymbol{\Sigma}^p)^{-1}\bar{\mathbf{z}}^p + (\mathbf{y}^o)^\mathrm{T}\mathbf{R}^{-1}\mathbf{y}^o$$
$$- (\bar{\mathbf{z}}^u)^\mathrm{T}(\boldsymbol{\Sigma}^u)^{-1}\bar{\mathbf{z}}^u]/2\}w, \tag{9}$$

as in A01.

The discussion that follows assumes sequential assimilation of scalar observations so that in (7)–(9), $\mathbf{y}^o$ is a vector of length 1 and $\mathbf{R}$ is a 1×1 matrix. Additional simplifications in computing the product of Gaussians can then be made easily. The order of the prior joint state covariance, $\boldsymbol{\Sigma}^p$, is $k \times k$, where $k = n + 1$. The updated covariance from (7) can be written

$$\boldsymbol{\Sigma}^u = [\mathbf{I} - (r + \sigma_{k,k})^{-1}\boldsymbol{\Sigma}^p_{0k}]\boldsymbol{\Sigma}^p, \tag{10}$$

where $\sigma_{k,k}$ is the prior observation variable error variance (the kth diagonal element of $\boldsymbol{\Sigma}^p$), r is the observation error variance (the only element of the 1×1 matrix $\mathbf{R}$), and $\boldsymbol{\Sigma}^p_{0k}$ is the matrix consisting of $\boldsymbol{\Sigma}^p$ with all elements except those in the last column set to 0. The last column is the prior covariance of each joint state variable with the observation variable. The change in the covariance due to the assimilation of a single observation is then

$$\Delta\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^u - \boldsymbol{\Sigma}^p = -(r + \sigma_{k,k})^{-1}\boldsymbol{\Sigma}^p_{0k}\boldsymbol{\Sigma}^p. \tag{11}$$

Substituting (10) into the expression for the updated mean from (8) gives

$$\bar{\mathbf{z}}^u = [\mathbf{I} - (r + \sigma_{k,k})^{-1}\boldsymbol{\Sigma}^p_{0k}]\boldsymbol{\Sigma}^p[(\boldsymbol{\Sigma}^p)^{-1}\bar{\mathbf{z}}^p + \mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{y}^o]. \tag{12}$$

Noting that $\boldsymbol{\Sigma}^p\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{y}^o = \boldsymbol{\Sigma}^p_{0k}\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{y}^o$ and $\boldsymbol{\Sigma}^p_{0k}\boldsymbol{\Sigma}^p_{0k} = \sigma_{k,k}\boldsymbol{\Sigma}^p_{0k}$, this becomes

$$\bar{\mathbf{z}}^u = [\mathbf{I} - (r + \sigma_{k,k})^{-1}\boldsymbol{\Sigma}^p_{0k}]\bar{\mathbf{z}}^p$$
$$+ r^{-1}[\mathbf{I} - (r + \sigma_{k,k})^{-1}\sigma_{k,k}]\boldsymbol{\Sigma}^p_{0k}\mathbf{H}^\mathrm{T}\mathbf{y}^o. \tag{13}$$

An equation for the change in the mean due to assimilating the observation is

$$\Delta\bar{\mathbf{z}} = \bar{\mathbf{z}}^u - \bar{\mathbf{z}}^p = (r + \sigma_{k,k})^{-1}\boldsymbol{\Sigma}^p_{0k}(\mathbf{H}^\mathrm{T}\mathbf{y}^o - \bar{\mathbf{z}}^p)$$
$$= (r + \sigma_{k,k})^{-1}(y^o - \bar{z}^p_k)[\sigma_{1,k}, \sigma_{2,k}, \ldots, \sigma_{k,k}]^\mathrm{T}, \tag{14}$$

where the single element of the vector $\mathbf{y}^o$ is y^o and $\bar{z}^p_k$ is the prior mean value of the observation variable.

Finally, in this case the weight D from (9) (a scalar) depends only on the observation and the observation variable and simplifies to

$$D = \exp[-(y^o - \bar{z}^p_k)^2/2(r + \sigma_{k,k})]w. \tag{15}$$

It is easily verified that computing the impact of the observation on each state variable independently in (11), (14), and (15) is equivalent to computing the impact on all state variables at once. This was pointed out, but not rigorously derived, in A01 where assimilations were performed by looking at the impact of an observation on each state variable in turn. More complete discussion of estimation theory as applied in the preceding two subsections can be found in Cohn (1997).

a. Perturbed observation ensemble Kalman filter

In its traditional implementation, the perturbed observation ensemble Kalman filter (Houtekamer and Mitchell 1998) uses a random number generator to sample the observational error distribution (specified as part of the observing system) and adds these samples to the observation, $\mathbf{y}^o$, to form an ensemble sample of the observation distribution, $\mathbf{y}^o_i$, $i = 1, \ldots, N$. In most implementations (Houtekamer and Mitchell 1998), the mean of the perturbations is adjusted to be 0 so that the perturbed observations have mean $\mathbf{y}^o$; other clever methods for perturbing the observations can preserve other aspects of the distribution (Pham 2001) (the discussion below applies whether adjustment to the means or other types of perturbation algorithms are used or not). Here, $\boldsymbol{\Sigma}^p$ is computed using sample statistics from the prior joint state ensemble and (7) is computed once to find the value of $\boldsymbol{\Sigma}^u$. Equation (8) is then applied N times with $\bar{\mathbf{z}}^p$ replaced with $\mathbf{z}^p_i$ and $\mathbf{y}^o$ replaced by $\mathbf{y}^o_i$ in the ith application to compute N ensemble members for $\mathbf{z}^u$. This method is described using more traditional Kalman filter terminology in Houtekamer and Mitchell (1998). 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 (2). Note that all ensemble members are assumed equally weighted in both the EnKF and EAKF so that (9) is not used; however, (9) may be relevant for other ensemble filtering methods (see section 4).

An equivalent two-step procedure for the EnKF begins by computing the update increments for the observation variable, y, a scalar problem independent of the other joint state variables (4). Perturbed observations are generated and (7) is used to compute an updated variance for the observation variable; all matrices here are order 1×1. Equation (8) is evaluated N times to compute y^u_i, with $\bar{\mathbf{z}}^p$ and $\mathbf{y}^o$ replaced by y^p_i and y^o_i, where the subscript refers to the value of the ith ensemble member.

Equation (14) applies in this case, since the updated covariance in the full dimension EnKF is computed by (7), and can be used to compute the increments for all other state variables given the value of $\Delta y_i = y^u_i - y^p_i$ (the kth component of the k-vector $\Delta\mathbf{z}_i$). All components of $\Delta\mathbf{z}_i$ can be computed from (14) as

$$\Delta\mathbf{z}_{i,j} = \frac{\sigma_{j,k}}{\sigma_{k,k}}\Delta y_i, \tag{16}$$

where the first subscript on $\Delta\mathbf{z}$ indexes the ensemble member and the second indexes the state variable. This

is the regression formula (6) presented in section 2c derived from the assumption of a Gaussian relation between the prior state variables. Implementing the EnKF in this two-step fashion gives results identical (to computational roundoff) to previous implementations of the EnKF.

b. Ensemble adjustment Kalman filter–square root filter

Bishop et al. (2001), Whitaker and Hamill (2002), and Pham (2001) have all described other ensemble filtering methods similar to the EAKF in A01. Tippett et al. (2003), providing an analysis of the work of these different authors, point out that the methods are roughly equivalent and suggest that the name deterministic square root filter (Andrews 1968) may be more appropriate.

The EAKF constructs an updated ensemble with a mean and sample variance that exactly satisfy (7) and (8). In A01 this is done by shifting the mean of the ensemble and then adjusting the spread of the ensemble around the updated mean using a linear operator $\mathbf{A}$:

$$\mathbf{z}_i^u = \mathbf{A}(\mathbf{z}_i^p - \overline{\mathbf{z}}^p) + \overline{\mathbf{z}}^u, \qquad (17)$$

where $\mathbf{A}$ satisfies $\mathbf{\Sigma}^u = \mathbf{A}\mathbf{\Sigma}^p\mathbf{A}^T$.

The EAKF can be recast in the two-step framework developed in section 2. Again, (4) implies that the observation variable y can be updated independently of the other joint state variables. Recalling that y is the kth element of the k-dimensional joint state vector, the updated variance for y can be written

$$\sigma_{k,k}^u = [(\sigma_{k,k}^p)^{-1} + r^{-1}]^{-1} \qquad (18)$$

using a scalar application of (7). Applying a scalar version of (8) to compute the updated mean, y^u, the updated value of y can be written

$$y_i^u = \alpha(y_i^p - \overline{y}^p) + \overline{y}^u, \qquad (19)$$

where

$$\alpha = [\sigma_{k,k}^u(\sigma_{k,k}^p)^{-1}]^{1/2} = [r(r + \sigma_{k,k}^p)^{-1}]^{1/2}.$$

For the change in the mean values, $\Delta \overline{z} = \overline{z}^u - \overline{z}^p$, Eq. (14) holds for the EAKF and this implies that (16) can be used to compute the changes in the state variables by regression from the change in the mean for $\mathbf{y}$.

It can also be shown that the regression formula can be applied for the adjustment of the ensemble members around the mean in the EAKF. Equation (10) can be rewritten by taking the square root of the operator as

$$\mathbf{\Sigma}^u = \mathbf{B}\mathbf{\Sigma}^p\mathbf{B}^T, \qquad (20)$$

where

$$\mathbf{B} = (\mathbf{I} + \beta\mathbf{\Sigma}_{0k}^p) \quad \text{and} \qquad (21)$$

$$\beta = (\sigma_{k,k}^p)^{-1}\left[\sqrt{r(r + \sigma_{k,k}^p)^{-1}} - 1\right]. \qquad (22)$$

If the deviation of the observation variable around

the mean is updated as in (19), applying the linear regression (16) for state variable j gives

$$\Delta z_{i,j} = \sigma_{j,k}(\sigma_{k,k})^{-1}\Delta y_i = \sigma_{j,k}(\sigma_{k,k})^{-1}(\alpha - 1)y_i^p. \qquad (23)$$

Then,

$$z_{i,j}^u = z_{i,j}^p + \sigma_{j,k}^p\Gamma y_i^p, \qquad (24)$$

where

$$\Gamma = (\sigma_{k,k}^p)^{-1}(\alpha - 1)$$
$$= (\sigma_{k,k}^p)^{-1}\left[\sqrt{r(r + \sigma_{k,k}^p)^{-1}} - 1\right]. \qquad (25)$$

Writing this in vector form for the updates of all joint state variables gives

$$\mathbf{z}_i^u = \mathbf{A}(\mathbf{z}_i^p - \overline{\mathbf{z}}^p), \qquad (26)$$

where

$$\mathbf{A} = \mathbf{I} + \Gamma\mathbf{\Sigma}_{0k}^p. \qquad (27)$$

The matrix $\mathbf{A}$ used in the regression update is identical to $\mathbf{B}$ in (21) demonstrating that using regression in the two-step framework is identical to the implementation of the EAKF described in A01.

In summary, the EnKF and EAKF assume a Gaussian relation between the variables in the joint state space prior distribution. However, these methods do not use (7) and (8) directly to compute an updated distribution. Both methods can be recast in terms of the two-step assimilation context developed in section 2c. First, update increments are computed for each ensemble sample of the observation variable using scalar versions of the traditional algorithms. Once increments for ensemble samples of the observation variable have been computed, (16) is used to solve for the increments, $\Delta z_{j,i}$, for each state variable in turn in terms of Δy_i by linear regression. The appendix discusses this method in the case where the prior covariance matrix is degenerate.

There are a number of implications about the computational complexity of ensemble (Kalman) filtering that can be drawn from (16). First, there is no need to compute the prior covariance among the model state variables (only the prior cross covariance of each state variable with the observation variable is needed, along with the variance of the observation variable) or the complete updated covariance, $\mathbf{\Sigma}^u$ (only the updated variance of the observation variable is needed). Second, once the observational variables are updated, the increments for the state variables depend only on ratios of prior (co)variances. Any multiplication of the prior covariance matrix by some type of covariance inflation factor, as is done in many existing ensemble Kalman filter implementations (Anderson and Anderson 1999; Whitaker and Hamill 2002), does not impact the solution to (16). The impacts of covariance inflation are still felt in the first step in which the increments of the observational variable are computed. Note that many types of covariance inflation directly increase the spread of the joint prior distribution.

There are important caveats to this discussion of enhanced computational efficiency. First, as noted in section 2b, correlations between observational errors require a rotation of the problem to apply the sequential methods discussed here. In the limit where most of the observational errors are correlated, the cost of doing this rotation could end up offsetting the savings from avoiding matrix multiplies. At present, most operational centers do not assume that there are correlated observational errors, but this could become important at a later date. If only certain sets of observations have correlated errors, a block diagonal covariance structure results and rotation can still be done more efficiently than computing the full matrix products. Second, the sequential method outlined here may present challenges for implementation on highly parallel computer architectures; relative performance on such hardware will require further analysis.

4. Additional methods for updating the observational variable ensembles

In this section, some additional methods for updating the observational variable are discussed. Once update increments for the observation variable are computed by one of the following methods, the rest of the joint state variables can be updated by linear regression using (16).

a. A kernel filter

If the prior distribution of the observation variable may have significant non-Gaussian structure, a kernel method similar to the one employed in Anderson and Anderson (1999) may be useful for computing the update increments. One simple example of kernel methods is the Fukunaga method (Silverman 1986). In this algorithm, the prior distribution is represented as a sum of Gaussians with identical variance but different means. The means are the individual prior ensemble samples and the variance is the prior sample variance multiplied by a scaling factor, η. The prior distribution is then

$$p(y^p) = \sum_{i=1}^{N} N(y_i^p, \, \eta \Sigma^P),$$

where $N(a, b)$ is a Gaussian with mean a and variance b.

The product of a prior expressed as a sum of Gaussians and a Gaussian observational distribution is equal to the sum of the products of the individual prior Gaussians and the observational Gaussian. The variance of all Gaussians summed in the product is identical in this case and can be computed by a single scalar application of (7). The means will all be different and can be computed by N scalar applications of (12). In the most naïve application of this method, an updated ensemble can be generated from this continuous representation by randomly sampling the sum of Gaussians as in Anderson and Anderson (1999).

This kernel method can be extended in a number of ways by allowing more general kernels. For instance, kernels with different means and different variances can be used following a variety of techniques like the class of nearest-neighbor methods (Silverman 1986; Bengtsson and Nychka 2001). In addition, kernels from the class of ''generalized'' Gaussians as described in Tarantola (1987) can lead to related kernel algorithms.

b. Quadrature product methods

Update methods that are based directly on ''quadrature'' solutions to (4) can also be used to find increments for observation variables. One implementation of such a method could begin by computing a continuous approximation to the prior distribution from the ensemble sample; again, kernel methods are an example. Quadrature methods can then be used to divide the real line into a set of intervals over which the product in the numerator of (4) is computed to approximate the updated distribution. An appropriate method can then be used to sample this updated distribution to generate new ensemble methods.

c. General requirements for an observation variable update

Several characteristics may be important for algorithms used to update the observation variables. First, low quality observations should have small impacts on the ensemble. For atmospheric and oceanic models, the prior distributions may be sampling model ''attractors'' that have a great deal of structure. Allowing low impact observations to change the ensembles has the potential to destroy valuable information. Pure resampling algorithms would be an example of an undesirable method. In this case, the prior ensemble would be converted to a continuous representation that would then be only subtly modified by a low information observation. This updated continuous distribution would then be resampled to generate an ensemble, leading to possibly large increments to ensemble members. The ensemble kernel filter as described above suffers from this deficiency and generally produces assimilations with larger ensemble mean error than do the EnKF and EAKF despite the fact that in many instances it produces more accurate samples of the updated observation variable distribution when the prior is significantly non-Gaussian. Modifications to the kernel filter that limit the impact of low information observations are required to make this method more generally useful.

For related reasons, it is desirable to limit the size of the increments for observation variables as noted in section 2. Since the regression used to update the state variables is a statistical linearization, it is likely to be an increasingly poor approximation as the increments increase. For instance, the updated mean and covariance of the *observation* variable for the EnKF would be the

same if the pairing between the updated, z_k^u, and prior, z_k^p, observational variable ensemble members was changed before the computation of update increments, Δz_k.

In some applications, the performance of the EnKF can be dramatically improved by pairing the updated observational variable ensemble members with the prior members so as to reduce the value of the update increments. The most obvious way to do this is to sort the prior and updated observational variable ensemble members and to associate the nth sorted updated ensemble member with the nth sorted prior ensemble member. Doing this reduces much of the difference between the EnKF and EAKF reported in A01. Doing ordered pairing also significantly improves the performance of kernel filter algorithms (section 4a) by reducing the size of the increments.

As an idealized example of this sorting procedure, suppose that for a three-member ensemble, the prior observation variable sample $z_k^p = \{1, 5, 7\}$ and the updated sample $z_k^u = \{4.9, 6.6, 1.2\}$. Unordered pairings like this can result when perturbed observations are used in the EnKF. Pairing these in this order gives increments $\Delta z_k = \{3.9, 1.6, -5.8\}$. If both sets are sorted, the statistics of each remains unchanged, but the increments are $\Delta z_k = \{0.2, -0.1, -0.4\}$. The errors resulting from the linear regression in the second step of the assimilation are clearly expected to be significantly less in the sorted case.

While many existing ensemble filters can be recast in the framework presented here, there are some that apparently cannot. For instance, the particle filter of Pham (2001) requires information about the joint distribution of observation and state variables when doing resampling. Methods like the singular second-order-exact EnKF of Pham (2001), however, can be expressed in the two-step framework.

5. Conclusions

A local least squares framework for ensemble filtering has been derived leading to a two-step ensemble filtering update procedure when a new observation becomes available. The first step is to compute update increments for each ensemble member of a prior estimate of the value of the observation. This can be done using a variety of algorithms including the perturbed observation ensemble Kalman filter and the ensemble adjustment Kalman filter. Other viable update methods, for instance a kernel filter, extend beyond the Kalman filter context and can be referred to more generally as ensemble filters.

The second step computes increments for each ensemble member of the prior estimate for each state variable by using the prior ensemble sample to do a linear regression of each state variable in turn on the observation variable. The increments for a given state variable are computed by multiplying the corresponding observation variable increment by the prior covariance of the state and observation variable and dividing by the prior variance of the observation variable.

Deriving a class of ensemble filters in this two-step context has a number of advantages. First, it is computationally more efficient than previous descriptions of ensemble Kalman filter algorithms in the literature when observations do not have correlated error distributions. The cost is expected to be dominated by the computation of the prior sample cross covariance of the observation and state variables and the variance of the observation variables. A second advantage is that much more elaborate and expensive ensemble update methods can be used because they need be applied only in a scalar fashion to the observation variables. A final advantage is that it is easier to understand differences between various filtering algorithms. Differences only need to be explored in a scalar context making the relative features of, for instance, the EnKF and EAKF much easier to understand.

By lowering the cost of existing filters and opening up a variety of new filter update algorithms, it is hoped that this local least squares framework can accelerate the development of ensemble filtering algorithms that are best suited for applications such as numerical weather prediction and ocean state estimation.

APPENDIX

Degenerate Prior Covariance

A potential complication occurs if Σ^p is degenerate so that its inverse is not defined. There are several reasons why Σ^p might be degenerate. First, it is possible that the joint state vector is of order greater than the size of the space that it spans. The most obvious instance occurs when an observation is a linear function of the prior model state variables (for instance if a state variable is observed directly). Second, details of the sample statistics of the prior ensemble could lead to degeneracy. For instance, if the ensemble is smaller than the size of the joint state space, Σ^p computed from the sample statistics must be degenerate. Third, details of the prediction model could lead to states that are confined to some submanifold of the model state space.

The two-step procedures for the EnKF and EAKF continue to be equivalent to the traditional implementations even if Σ^p is degenerate. In this case, (7)–(9) must be modified to replace the inverses with pseudoinverses. Let $\mathbf{U}^\mathrm{T}$ be an $\eta \times k$ orthogonal matrix whose rows are the set of left singular vectors of Σ^p corre-

sponding to nonzero singular values (η is the rank of $\boldsymbol{\Sigma}^p$). Applying $\mathbf{U}^\mathrm{T}$ to a vector in the original space gives the projection of that vector on the range of $\boldsymbol{\Sigma}^p$. The projection of the covariance on this space is $\mathbf{U}^\mathrm{T}\boldsymbol{\Sigma}^p\mathbf{U}$. The probability density of the prior distribution outside of this subspace is zero, so the updated distribution and hence $\Delta\mathbf{z}$ must lie in this subspace.

Next, define an $\eta \times \eta$ orthogonal matrix $\mathbf{B}^\mathrm{T}$ that performs a change of basis in the reduced SVD subspace. Let the $\eta \times k$ matrix $\mathbf{C}^\mathrm{T} = \mathbf{B}^\mathrm{T}\mathbf{U}^\mathrm{T}$ be defined so that the last row of $\mathbf{C}^\mathrm{T}$ is the projection of the observation variable, $\mathbf{z}_k = \mathbf{y}$, on the range of $\boldsymbol{\Sigma}^p$; the last column of $\mathbf{C}^\mathrm{T}$ is $[0, 0, \ldots, 0, \psi]^\mathrm{T}$. As long as the observation vector does not lie in the null space of $\boldsymbol{\Sigma}^p$, $\mathbf{C}^\mathrm{T}$ exists, but this must be the case for (2) to have a relevant solution. If the observation vector did lie entirely in the null space, then $\mathbf{p}(\mathbf{y}^o|\mathbf{z}^p)$ would be 0 with probability 1 and the result of (2) would be a delta function indicating a deterministic solution [and probably an improperly defined problem (Tarantola 1987)].

In the subspace spanned by the rows of $\mathbf{C}^\mathrm{T}$, the projection of the prior covariance has an inverse and (7)–(9) can be applied. The results can then be projected back to the original space:

$$\boldsymbol{\Sigma}^u = \mathbf{C}[(\mathbf{C}^\mathrm{T}\boldsymbol{\Sigma}^p C)^{-1} + \mathbf{C}^\mathrm{T}\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{H}\mathbf{C}]^{-1}C^\mathrm{T}. \quad (A1)$$

In this subspace, define that $\boldsymbol{\Sigma}'^u = \mathbf{C}^\mathrm{T}\boldsymbol{\Sigma}^u\mathbf{C}$, $\boldsymbol{\Sigma}'^p = \mathbf{C}^\mathrm{T}\boldsymbol{\Sigma}^p\mathbf{C}$, $\mathbf{z}'_i = \mathbf{C}^\mathrm{T}\mathbf{z}_i$ (an η vector), $y'_i = \psi y_i$ is the last element of z'_i, and $\sigma'^p_{\eta,\eta} = \psi^2\sigma^p_{k,k}$ is the prior variance of y'_i. Also note that $\mathbf{C}^\mathrm{T}\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{H}\mathbf{C}$ is a $\eta \times \eta$ matrix with all elements 0 except the last column of the last row, which is $r'^{-1} = \psi^2 r^{-1}$. Finally, define $\mathbf{R}'^{-1}$ as the 1×1 matrix with only element r'^{-1} and $\mathbf{H}'$ as the η vector $[0, 0, \ldots, 0, 1]$.

In the subspace, (10)–(16) hold for the primed quantities just defined. In particular, (16) in vector form gives

$$\Delta z'_i = (\sigma'^p_{\eta,\eta})^{-1}\boldsymbol{\Sigma}'^p[0, 0, \ldots, 0, \Delta y'_i]^\mathrm{T}. \quad (A2)$$

Converting this back to the original subspace gives

$$\Delta z_i = \mathbf{C}\Delta z'_i = (\sigma'^p_{\eta,\eta})^{-1}\mathbf{C}\boldsymbol{\Sigma}'^p [0, 0, \ldots, 0, \psi\Delta y_i]^\mathrm{T}. \quad (A3)$$

Using the fact that $\mathbf{C}^\mathrm{T}\mathbf{C} = \mathbf{I}$,

$$\Delta z_i = \psi^{-2}(\sigma^p_{k,k})^{-1}\mathbf{C}\boldsymbol{\Sigma}'^p\mathbf{C}^\mathrm{T}\mathbf{C}[0, 0, \ldots, 0, \psi\Delta y_i]^\mathrm{T}$$

$$= (\sigma^p_{k,k})^{-1}\boldsymbol{\Sigma}^p[0, 0, \ldots, 0, \Delta y_i]^\mathrm{T}, \quad (A4)$$

which is a vector form of (16).

This demonstrates that using the two-step procedure in the original space, even if $\boldsymbol{\Sigma}^p$ is degenerate, gives results corresponding to those given by the previously documented versions of the EnKF and EAKF.

REFERENCES

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.,* **129,** 2884–2903.

——, 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.

Andrews, A., 1968: A square root formulation of the Kalman covariance equations. *AIAA J.,* **6,** 1165–1168.

Bengtsson, T., and D. Nychka, 2001: Adaptive methods in numerical weather prediction. *Proc., First Spanish Workshop on Spatio–Temporal Modeling of Environmental Processes,* Benicassim, Castellon, Spain, Universitat Jaume I, 1–15.

Bishop, C. H., 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.

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.,* **126,** 1719–1724.

Cohn, S. E., 1997: An introduction to estimation theory. *J. Meteor. Soc. Japan,* **75,** 257–288.

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.

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.

Houtekamer, P. L., 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.

Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory.* Academic Press, 376 pp.

Kalman, R., and R. Bucy, 1961: New results in linear prediction and filtering theory. *Trans. AMSE J. Basic Eng.,* **83D,** 95–108.

Keppenne, C. L., 2000: Data assimilation into a primitive-equation model with a parallel ensemble Kalman filter. *Mon. Wea. Rev.,* **128,** 1971–1981.

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.

Miller, R. N., M. Ghil, and F. Gauthiez, 1994: Advanced data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sci.,* **51,** 1037–1056.

Mitchell, H. L., and P. L. Houtekamer, 2000: An adaptive ensemble Kalman filter. *Mon. Wea. Rev.,* **128,** 416–433.

Pham, D. T., 2001: Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Wea. Rev.,* **129,** 1194–1207.

Silverman, B. W., 1986: *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, 175 pp.

Tarantola, A., 1987: *Inverse Problem Theory.* Elsevier, 613 pp.

& Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: in press: Ensemble square-root filters. *Mon. Wea. Rev.,* in press.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.,* **130,** 1913–1924.