Ensemble Square-Root Filters[*]

MICHAEL K. TIPPETT[†]

International Research Institute for Climate Prediction, Palisades, New York

JEFFREY L. ANDERSON

GFDL, Princeton, New Jersy

CRAIG H. BISHOP

Naval Research Laboratory, Monterey, California

THOMAS M. HAMILL AND JEFFREY S. WHITAKER

NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado

November 18, 2002

[†]Corresponding author:IRI/LDEO 223 Monell, PO Box 1000 / 61 Route 9W, Palisades, NY 10964-8000. Phone: 845-680-4420. Fax: 845-680-4865. Email: tippett@iri.columbia.edu

Abstract

Ensemble data assimilation methods assimilate observations using state-space estimation methods and low-rank representations of forecast and analysis error covariances. A key element of such methods is the transformation of the forecast ensemble into an analysis ensemble with appropriate statistics. This transformation may be preformed stochastically by treating observations as random variables, or deterministically by requiring that the updated analysis perturbations satisfy the Kalman filter analysis error covariance equation. Deterministic analysis ensemble updates are implementations of Kalman square-root filters. The nonuniqueness of the deterministic transformation used in square-root Kalman filters provides a framework to compare three recently proposed ensemble data assimilation methods.

1. Introduction

Data assimilation addresses the problem of producing useful analyses and forecasts given imperfect dynamical models and observations. The Kalman filter is the optimal data assimilation method for linear dynamics with additive, state-independent Gaussian model and observation error (Cohn 1997). An attractive feature of the Kalman filter is its calculation of forecast and analysis error covariances, in addition to the forecast and analyses themselves. In this way, the Kalman filter produces estimates of forecast and analysis uncertainty, consistent with the dynamics and prescribed model and observation error statistics. However, the error covariance calculation components of the Kalman filter are difficult to implement in realistic systems because of (*i*) their computational cost, (*ii*) nonlinearity of the dynamics and (*iii*) poorly characterized error sources.

The ensemble Kalman filter (EnKF), proposed by Evensen (1994), addresses the first two of these problems by using ensemble representations for the forecast and analysis error covariances. Ensemble size limits the number of degrees of freedom used to represent forecast and analysis errors, and Kalman filter error covariance calculations are practical for modest-sized ensembles. The EnKF algorithm begins with an analysis ensemble whose mean is the current state-estimate or analysis and whose statistics reflect the analysis error. Applying the full nonlinear dynamics to each analysis ensemble member produces the forecast ensemble; tangent linear and adjoint models of the dynamics are not required. Statistics of the forecast ensemble represent forecast errors; in its simplest form, the EnKF only accounts for forecast error due to uncertain initial conditions, neglecting forecast error due to model deficiencies. The forecast ensemble mean and covariance are then used to assimilate observations and compute a new analysis ensemble with appropriate statistics, and the cycle is repeated. The new analysis ensemble can be formed either stochastically (Houtekamer and Mitchell 1998; Burgers et al. 1998) or deterministically (Bishop et al. 2001; Anderson 2001; Whitaker and Hamill 2002). Deterministic methods were developed to address

the adaptive observational network design problem and to avoid sampling issues associated with the use of "perturbed observations" in stochastic analysis ensemble update methods.

The EnKF and other ensemble data assimilation methods belong to the family of *square-root filters* (SRFs), and a purpose of this paper is to demonstrate that deterministic analysis ensemble updates are implementations of Kalman SRFs (Bierman 1977; Maybeck 1982; Heemink et al. 2001). An immediate benefit of this identification is a framework for understanding and comparing deterministic analysis ensemble update schemes (Bishop et al. 2001; Anderson 2001; Whitaker and Hamill 2002). SRFs, like ensemble representations of covariances, are not unique. We begin our discussion in Section 2 with a presentation of the Kalman SRF; issues related to implementation of ensemble SRFs are presented in Section 3; in Section 4 we summarize our results.

2. The Kalman SRF

Kalman SRF algorithms, originally developed for space-navigation systems with limited computational word length, demonstrate superior numerical precision and stability compared to the standard Kalman filter algorithm (Bierman 1977; Maybeck 1982). SRFs by construction avoid loss of positive definiteness of the error covariance matrices. SRFs have been used in Earth Science data assimilation methods where error covariances are approximated by truncated eigenvector expansions (Verlaan and Heemink 1997).

The usual Kalman filter covariance evolution equations are

$$\mathbf{P}_k^f = \mathbf{M}_k \mathbf{P}_{k-1}^a \mathbf{M}_k^T + \mathbf{Q}_k \,, \tag{1}$$

$$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f \,, \tag{2}$$

where $\mathbf{P}_k^f$ and $\mathbf{P}_k^a$ are respectively the $n \times n$ forecast and analysis error covariance matrices at time t_k, $\mathbf{M}_k$ is the tangent linear dynamics, $\mathbf{H}_k$ is the $p \times n$ observation operator, $\mathbf{R}_k$ is the $p \times p$ observation error covariance, $\mathbf{Q}_k$ is the $n \times n$ model error covariance matrix and $\mathbf{K}_k \equiv$

$\mathbf{P}_k^f \mathbf{H}_k^T \left(\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k \right)^{-1}$ is the Kalman gain; n is the dimension of the system state, and p is the number of observations. The error covariance evolution depends on the state estimates and observations through the tangent linear dynamics $\mathbf{M}_k$. The propagation of analysis errors by the dynamics with model error acting as a forcing is described by Eq. (1). Equation (2) shows how an optimal data assimilation scheme uses observations to produce an analysis whose error covariance is less than that of the forecast.

The forecast and analysis error covariance matrices are symmetric positive-definite matrices and can be represented $\mathbf{P}_k^f = \mathbf{Z}_k^f \mathbf{Z}_k^{fT}$ and $\mathbf{P}_k^a = \mathbf{Z}_k^a \mathbf{Z}_k^{aT}$ where the matrices $\mathbf{Z}_k^f$ and $\mathbf{Z}_k^a$ are *matrix square-roots* of $\mathbf{P}_k^f$ and $\mathbf{P}_k^a$ respectively; other matrix factorizations can be used in filters as well (Bierman 1977; Pham et al. 1998). A covariance matrix and its matrix square-root have the same rank or number of nonzero singular values. When a covariance matrix $\mathbf{P}$ has rank m, there is an $n \times m$ matrix square-root $\mathbf{Z}$ satisfying $\mathbf{P} = \mathbf{Z}\mathbf{Z}^T$; in low-rank covariance representations the rank m is much less than the state-space dimension n. This representation is not unique; $\mathbf{P}$ can also be represented as $\mathbf{P} = (\mathbf{Z}\mathbf{U})(\mathbf{Z}\mathbf{U})^T$ where the matrix $\mathbf{U}$ is any $m \times m$ orthogonal transformation $\mathbf{U}\mathbf{U}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I}$. The projection $\|\mathbf{x}^T\mathbf{Z}\|^2 = \mathbf{x}^T\mathbf{P}\mathbf{x}$ of an arbitrary n-vector $\mathbf{x}$ onto the matrix square-root $\mathbf{Z}$ is uniquely determined, as is the subspace spanned by the columns of $\mathbf{Z}$.

Covariance matrix square-roots are closely related to ensemble representations. The sample covariance $\mathbf{P}_k^a$ of an m-member analysis ensemble is given by $\mathbf{P}_k^a = \mathbf{S}\mathbf{S}^T/(m-1)$ where the columns of the $n \times m$ matrix $\mathbf{S}$ are mean-zero analysis perturbations about the analysis ensemble mean; the rank of $\mathbf{P}_k^a$ is at most $(m-1)$. A matrix square-root of the analysis error covariance matrix $\mathbf{P}_k^a$ is the matrix of scaled analysis perturbation ensemble members $\mathbf{Z}_k^a = (m-1)^{-1/2}\mathbf{S}$.

The Kalman SRF algorithm replaces error covariance evolution equations (1) and (2) with equations for the evolution of forecast and analysis error covariance square-roots $\mathbf{Z}_k^f$ and $\mathbf{Z}_k^a$ in a manner that avoids forming the full error covariance matrices. If the model error covariance $\mathbf{Q}_k$ is

neglected, (1) can be replaced by

$$\mathbf{Z}_k^f = \mathbf{M}_k \mathbf{Z}_{k-1}^a \, . \qquad\qquad (3)$$

In the ensemble context, (3) means to apply the tangent linear dynamics to each column of the $\mathbf{Z}_{k-1}^a$, that is, to each scaled analysis perturbation ensemble member. Practically, (3) can be implemented by applying the full nonlinear dynamics to each analysis ensemble member. For what follows, we only assume that the forecast error covariance matrix square-root $\mathbf{Z}_k^f$ is available and do not assume or restrict that it be calculated from (3). Section 3b discusses more sophisticated methods of generating $\mathbf{Z}_k^f$ that can include estimates of model error and give a forecast error covariance matrix square-root whose rank is greater than the number of perturbations evolved by the dynamical model.

Next, analysis error covariance equation (2) is replaced with an equation for the analysis error covariance square-root $\mathbf{Z}_k^a$. This equation determines how to form an analysis ensemble with appropriate statistics. Initial implementations of the EnKF formed the new analysis ensemble by updating each forecast ensemble member using the same analysis equations, equivalent to applying the linear operator $(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)$ to the forecast perturbation ensemble $\mathbf{Z}_k^f$. This procedure gives an analysis ensemble whose error covariance is $(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)^T$ and includes analysis error due to forecast error; the Kalman gain $\mathbf{K}_k$ depends on the relative size of forecast and observation error, and the factor $(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)$ shows how much forecast errors are reduced. However, in this procedure the analysis ensemble does not include uncertainty due to observation error and so underestimates analysis error. A *stochastic* solution to this problem proposed independently by Houtekamer and Mitchell (1998) and Burgers et al. (1998) is to compute analyses using each forecast ensemble member and, instead of using a single realization of the observations, to use an ensemble of simulated observations whose statistics reflect the observation error. This method is

equivalent to the analysis perturbation ensemble update

$$\mathbf{Z}_k^a = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{Z}_k^f + \mathbf{K}_k\mathbf{W}_k \,, \tag{4}$$

where $\mathbf{W}_k$ is a $p \times m$ matrix whose m columns are identically distributed, mean-zero, Gaussian random vectors of length p with covariance $\mathbf{R}_k/m$. The perturbed observation analysis equation (4) gives an analysis perturbation ensemble with correct expected statistics:

$$\begin{aligned}
\left\langle \mathbf{Z}_k^a \left(\mathbf{Z}_k^a\right)^T \right\rangle &= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^f(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T \\
&= \mathbf{P}_k^a \,.
\end{aligned} \tag{5}$$

However, the perturbed observation approach introduces an additional source of sampling error that reduces analysis error covariance accuracy and increases the probability of underestimating analysis error covariance (Whitaker and Hamill 2002). A Monte Carlo method avoiding perturbed observations is described in Pham (2001). The singular evolutive interpolate Kalman (SEIK) filter uses both deterministic factorization and stochastic approaches.

Kalman SRFs provide a *deterministic* algorithm for transforming the forecast ensemble into an analysis ensemble with consistent statistics. The "Potter method" for the Kalman SRF analysis update (Bierman 1977) is obtained by rewriting (2) as

$$\begin{aligned}
\mathbf{P}_k^a = \mathbf{Z}_k^a\mathbf{Z}_k^{aT} &= \left[\mathbf{I} - \mathbf{P}_k^f\mathbf{H}_k^T\left(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T + \mathbf{R}_k\right)^{-1}\mathbf{H}_k\right]\mathbf{P}_k^f \\
&= \mathbf{Z}_k^f\left[\mathbf{I} - \mathbf{Z}_k^{fT}\mathbf{H}_k^T\left(\mathbf{H}_k\mathbf{Z}_k^f\mathbf{Z}_k^{fT}\mathbf{H}_k^T + \mathbf{R}_k\right)^{-1}\mathbf{H}_k\mathbf{Z}_k^f\right]\mathbf{Z}_k^{fT} \\
&= \mathbf{Z}_k^f\left(\mathbf{I} - \mathbf{V}_k\mathbf{D}_k^{-1}\mathbf{V}_k^T\right)\mathbf{Z}_k^{fT} \,,
\end{aligned} \tag{6}$$

where we define the $m \times p$ matrix $\mathbf{V}_k \equiv \left(\mathbf{H}_k\mathbf{Z}_k^f\right)^T$ and the $p \times p$ innovation covariance matrix $\mathbf{D}_k \equiv \mathbf{V}_k^T\mathbf{V}_k + \mathbf{R}_k$. Then the analysis perturbation ensemble is calculated from

$$\mathbf{Z}_k^a = \mathbf{Z}_k^f\mathbf{X}_k\mathbf{U}_k \,, \tag{7}$$

where $\mathbf{X}_k\mathbf{X}_k^T = (\mathbf{I} - \mathbf{V}_k\mathbf{D}_k^{-1}\mathbf{V}_k^T)$ and $\mathbf{U}_k$ is an arbitrary $m \times m$ orthogonal matrix. As formulated, the updated ensemble $\mathbf{Z}_k^a$ is a linear combination of the columns of $\mathbf{Z}_k^f$ and is obtained by inverting the $p \times p$ matrix $\mathbf{D}_k$ and computing a matrix square-root $\mathbf{X}_k$ of the $m \times m$ matrix $(\mathbf{I} - \mathbf{V}_k\mathbf{D}_k^{-1}\mathbf{V}_k^T)$.

7

3. Ensemble SRFs

a. Analysis ensemble

In many typical Earth Science data assimilation applications the state-dimension n and the number of observations p are large, and the method for computing the matrix square-root of $(\mathbf{I} - \mathbf{V}_k \mathbf{D}_k^{-1} \mathbf{V}_k^T)$ and the updated analysis perturbation ensemble $\mathbf{Z}_k^a$ must be chosen accordingly. A *direct* approach is to solve first the linear system $\mathbf{D}_k \mathbf{Y}_k = \mathbf{H}_k \mathbf{Z}_k^f$ for the $p \times m$ matrix $\mathbf{Y}_k$, that is, to solve

$$(\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k) \mathbf{Y}_k = \mathbf{H}_k \mathbf{Z}_k^f , \tag{8}$$

as is done in the first step of the PSAS algorithm (Cohn et al. 1998). Then, the $m \times m$ matrix $\mathbf{I} - \mathbf{V}_k \mathbf{D}_k^{-1} \mathbf{V}_k^T = \mathbf{I} - (\mathbf{H}_k \mathbf{Z}_k^f)^T \mathbf{Y}_k$ is formed, its matrix square-root $\mathbf{X}_k$ computed and applied to $\mathbf{Z}_k^f$ as in (7). Solution of (8), even when p is large, is practical when the forecast error covariance has a low-rank representation and the inverse of the observation error covariance is available (see appendix). Iterative methods whose cost is on the order of the cost of applying the innovation covariance matrix are appropriate when the forecast error covariance is represented by a correlation model.

When observation errors are uncorrelated, observations can be assimilated one at a time or *serially* (Houtekamer and Mitchell 2001; Bishop et al. 2001). For a single observation, $p = 1$, $\mathbf{V}_k$ is a column-vector, and the innovation $\mathbf{D}_k$ is a scalar. In this case, a matrix square-root of $(\mathbf{I} - \mathbf{V}_k \mathbf{D}_k^{-1} \mathbf{V}_k^T)$ can be computed in closed form by taking the *ansatz*

$$\mathbf{I} - \mathbf{D}_k^{-1} \mathbf{V}_k \mathbf{V}_k^T = (\mathbf{I} - \beta_k \mathbf{V}_k \mathbf{V}_k^T)(\mathbf{I} - \beta_k \mathbf{V}_k \mathbf{V}_k^T)^T , \tag{9}$$

and solving for the scalar β_k, which gives $\beta_k = \left(\mathbf{D}_k \pm (\mathbf{R}_k \mathbf{D}_k)^{1/2} \right)^{-1}$. The analysis ensemble update for $p = 1$ is

$$\mathbf{Z}_k^a = \mathbf{Z}_k^f (\mathbf{I} - \beta_k \mathbf{V}_k \mathbf{V}_k^T) ; \tag{10}$$

see Andrews (1968) for a general solution involving matrix square-roots of $p \times p$ matrices. At observation locations, the analysis error ensemble is related to the forecast error ensemble by $\mathbf{H}_k \mathbf{Z}_k^a = (1 - \beta_k \mathbf{V}_k^T \mathbf{V}_k) \mathbf{H}_k \mathbf{Z}_k^f$. The scalar factor $(1 - \beta_k \mathbf{V}_k^T \mathbf{V}_k)$ has absolute value less than or equal to one and is positive when the plus sign is chosen in the definition of β_k.

In Whitaker and Hamill (2002) the analysis perturbation ensemble is found from

$$\mathbf{Z}_k^a = (\mathbf{I} - \tilde{\mathbf{K}}_k \mathbf{H}_k) \mathbf{Z}_k^f, \tag{11}$$

where the matrix $\tilde{\mathbf{K}}_k$ is a solution of the nonlinear equation

$$(\mathbf{I} - \tilde{\mathbf{K}}_k \mathbf{H}_k) \mathbf{P}_k^f (\mathbf{I} - \tilde{\mathbf{K}}_k \mathbf{H}_k)^T = \mathbf{P}_k^a. \tag{12}$$

In the case of a single observation, a solution of (12) is

$$\tilde{\mathbf{K}}_k = \left[1 + (\mathbf{R}_k/\mathbf{D}_k)^{1/2} \right]^{-1} \mathbf{K}_k = \beta_k \mathbf{Z}_k^f \mathbf{V}_k, \tag{13}$$

where the plus sign is chosen in the definition of β_k. The corresponding analysis perturbation ensemble update

$$\mathbf{Z}_k^a = (\mathbf{I} - \tilde{\mathbf{K}}_k \mathbf{H}_k) \mathbf{Z}_k^f = (\mathbf{I} - \beta_k \mathbf{Z}_k^f \mathbf{V}_k \mathbf{H}_k) \mathbf{Z}_k^f = \mathbf{Z}_k^f (\mathbf{I} - \beta_k \mathbf{V}_k \mathbf{V}_k^T), \tag{14}$$

is identical to (10). Observations with correlated errors, e.g., radiosonde height observations from the same sounding, can be handled by applying the whitening transformation $\mathbf{R}_k^{-1/2}$ to the observations to form a new observation set with uncorrelated errors.

Another method of computing the updated analysis ensemble is to use the Sherman-Morrison-Woodbury identity (Golub and Van Loan 1996) to show that

$$\mathbf{I} - \mathbf{V}_k \mathbf{D}_k^{-1} \mathbf{V}_k^T = \left(\mathbf{I} + \mathbf{Z}_k^{fT} \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k \mathbf{Z}_k^f \right)^{-1}. \tag{15}$$

The $m \times m$ matrix on the right hand side of (15) is practical to compute when the inverse observation error covariance matrix $\mathbf{R}_k^{-1}$ is available. This approach avoids inverting the $p \times p$ matrix $\mathbf{D}_k$

and is used in the *ensemble transform Kalman filter* (ET KF) where the analysis update is (Bishop et al. 2001)

$$\mathbf{Z}_k^a = \mathbf{Z}_k^f \mathbf{C}_k (\mathbf{\Gamma}_k + \mathbf{I})^{-1/2} \, ; \tag{16}$$

$\mathbf{C}_k \mathbf{\Gamma}_k \mathbf{C}_k^T$ is the eigenvalue decomposition of $\mathbf{Z}_k^{fT} \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k \mathbf{Z}_k^f$. Note that the matrix $\mathbf{C}_k$ of orthonormal eigenvectors is not uniquely determined.[1] Comparison with (15) shows that $\mathbf{C}_k (\mathbf{\Gamma}_k + \mathbf{I})^{-1} \mathbf{C}_k^T$ is the eigenvalue decomposition of $\mathbf{I} - \mathbf{V}_k \mathbf{D}_k^{-1} \mathbf{V}_k^T$ and thus that $\mathbf{C}_k (\mathbf{\Gamma}_k + \mathbf{I})^{-1/2}$ is a square-root of $(\mathbf{I} - \mathbf{V}_k \mathbf{D}_k^{-1} \mathbf{V}_k^T)$.

In the *ensemble adjustment Kalman filter* (EAKF) the form of the analysis ensemble update is (Anderson 2001)

$$\mathbf{Z}_k^a = \mathbf{A}_k \mathbf{Z}_k^f \, ; \tag{17}$$

the ensemble adjustment matrix $\mathbf{A}_k$ is defined by

$$\mathbf{A}_k \equiv \mathbf{F}_k \mathbf{G}_k \tilde{\mathbf{C}}_k (\mathbf{I} + \tilde{\mathbf{\Gamma}}_k)^{-1/2} \mathbf{G}_k^{-1} \mathbf{F}_k^T \, , \tag{18}$$

where $\mathbf{P}_k^f = \mathbf{F}_k \mathbf{G}_k^2 \mathbf{F}_k^T$ is the eigenvalue decomposition of $\mathbf{P}_k^f$ and the orthogonal matrix $\tilde{\mathbf{C}}_k$ is chosen so that $\tilde{\mathbf{C}}_k^T \mathbf{G}_k \mathbf{F}_k^T \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k \mathbf{F}_k \mathbf{G}_k \tilde{\mathbf{C}}_k = \tilde{\mathbf{\Gamma}}_k$ is diagonal.[2] Choosing the orthogonal matrix $\tilde{\mathbf{C}}_k$ to be $\tilde{\mathbf{C}}_k = \mathbf{G}_k^{-1} \mathbf{F}_k^T \mathbf{Z}_k^f \mathbf{C}_k$ gives that $\tilde{\mathbf{\Gamma}}_k = \mathbf{\Gamma}_k$ and that the ensemble adjustment matrix is

$$\mathbf{A}_k = \mathbf{Z}_k^f \mathbf{C}_k (\mathbf{I} + \mathbf{\Gamma}_k)^{-1/2} \mathbf{G}_k^{-1} \mathbf{F}_k^T \, . \tag{19}$$

The EAKF analysis update (17) becomes

$$\mathbf{Z}_k^a = \mathbf{Z}_k^f \mathbf{C}_k (\mathbf{I} + \mathbf{\Gamma}_k)^{-1/2} \mathbf{G}_k^{-1} \mathbf{F}_k^T \mathbf{Z}_k^f \, , \tag{20}$$

The EAKF analysis ensemble given by (20) is the same as applying the transformation $\mathbf{G}_k^{-1} \mathbf{F}_k^T \mathbf{Z}_k^f$ to the ET KF analysis ensemble. The matrix $\mathbf{G}_k^{-1} \mathbf{F}_k^T \mathbf{Z}_k^f$ is orthogonal and is, in fact, the matrix

---

[1]For instance, the columns of $\mathbf{C}_k$ that span the $(m-p)$-dimensional null-space of $\mathbf{Z}_k^{fT} \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k \mathbf{Z}_k^f$ are determined only up to orthogonal transformations if the number of observations p is less than the ensemble size m.

[2]The appearance of $\mathbf{G}_k^{-1}$ in the definition of the ensemble adjustment matrix $\mathbf{A}$ seems to require the forecast error covariance $\mathbf{P}_k^f$ to be invertible. However, the formulation is still correct when $\mathbf{G}_k$ is $m' \times m'$ and $\mathbf{F}_k$ is $n \times m'$ where m' is the number of nonzero eigenvalues of $\mathbf{P}_k^f$.

TABLE 1. Summary of analysis ensemble calculation computational cost as a function of forecast ensemble size m, number of observations p and state dimension n.

Analysis method	Cost
Direct	$O(m^2 p + m^3 + m^2 n)$
Serial	$O(mp + mnp)$
ET KF	$O(m^2 p + m^3 + m^2 n)$
EAKF	$O(m^2 p + m^3 + m^2 n)$

of right singular vectors of $\mathbf{Z}_k^f$. Therefore, $\mathbf{C}_k(\mathbf{I} + \mathbf{\Gamma}_k)^{-1/2}\mathbf{G}_k^{-1}\mathbf{F}_k^T\mathbf{Z}_k^f$ is a matrix square-root of $(\mathbf{I} - \mathbf{V}_k \mathbf{D}_k^{-1}\mathbf{V}_k^T)$.

Beginning with the same forecast error covariance, the direct, serial, ET KF and EAKF produce different analysis ensembles that span the same state-space subspace and have the same covariance. Higher order statistical moments of the different models will be different, a relevant issue for nonlinear dynamics. The computation costs of the direct, ET KF and EAKF is seen in Table 1 to scale comparably (see the appendix for details). There are differences in precise computational cost; for instance, the EAKF contains an additional SVD calculation of the forecast with cost $O(m^3 + m^2)$. The computational cost of the serial filter is less dependent on the rank of the forecast error covariance and more sensitive on the number of observations. This difference is important when techniques to account for model error and control filter divergence, as described in the next section, result in an effective forecast error covariance dimension m much larger than the dynamical forecast ensemble dimension.

b. Forecast error statistics

In the previous section we examined methods of forming the analysis ensemble given a matrix square-root of the forecast error covariance. There are two fundamental problems associated with directly using the ensemble generated by (3). First, ensemble size is limited by the computational cost of applying the forecast model to each ensemble member. Small ensembles have few degrees of freedom available to represent errors and suffer from sampling error that further degrades forecast error covariance representation. Sampling error leads to loss of accuracy and underestimation of error covariances that can cause filter divergence. Techniques to deal with this problem are distance-dependent covariance filtering and covariance inflation (Whitaker and Hamill 2002). Covariance localization in the serial method consists of adding a Schur product to the definition of $\tilde{\mathbf{K}}$ (Whitaker and Hamill 2002). Similarly, observations effecting analysis grid points can be restricted to be near-by in the EAKF (Anderson 2001).

The second and less easily resolved problem with using the ensemble generated by (3) is the neglect of model error and resulting underestimation of the forecast error covariance. Since there is little theoretical knowledge of model error statistics in complex systems, model error parameterizations combined with adaptive methods are likely necessary (Dee 1995). When the model error covariance $\mathbf{Q}_k$ is taken to have large-scale structure, a reasonable representation is an ensemble or square-root decomposition $\mathbf{Q}_k = \mathbf{Z}_k^d \mathbf{Z}_k^{dT}$ where $\mathbf{Z}_k^d$ is an $n \times q$ matrix. Then, a square-root of $\mathbf{P}_k^f$ is the $n \times m$ matrix

$$\mathbf{Z}_k^f = \begin{bmatrix} \mathbf{M}\mathbf{Z}_k^a & \mathbf{Z}_k^d \end{bmatrix} \, , \tag{21}$$

where $m = m_e + q$ and m_e is the number of dynamically evolved forecast perturbations. With this model error representation, ensemble size grows by q with each forecast/analysis cycle. Ensemble size can be limited by computing the singular value decomposition of the ensemble and discarding components with small variance (Heemink et al. 2001). A larger ensemble with evolved analysis

error and model error could be used in the analysis step, and a smaller ensemble used in the dynamical forecast stage. When the model error covariance $\mathbf{Q}_k$ is approximated as an operator, for instance using a correlation model, Lanczos methods can be used to compute the leading eigenmodes of $\mathbf{M}_k \mathbf{Z}_{k-1}^a (\mathbf{M}_k \mathbf{Z}_{k-1}^a)^T + \mathbf{Q}_k$ and form $\mathbf{Z}_k^f$ (Cohn and Todling 1996). Such a forecast error covariance model would resemble those used in "hybrid" methods (Hamill and Snyder 2000). In this case, the rank of $\mathbf{Z}_k^f$ can be substantially larger than the forecast ensemble size, making the serial method attractive. Monte Carlo solutions are another option as in Mitchell and Houtekamer (2000) where model error parameters were estimated from innovations and used to generate realizations of model error. Perturbing model physics, as done in *system simulation*, explicitly accounts for some aspects of model uncertainty (Houtekamer et al. 1996).

4. Summary and Discussion

Ensemble forecast/assimilation methods use low-rank ensemble representations of forecast and analysis error covariance matrices. These ensembles are scaled matrix square-roots of the error covariance matrices, and so ensemble data assimilation methods can be viewed as square-root filters (SRFs) (Bierman 1977). After assimilation of observations, the analysis ensemble can be constructed stochastically or deterministically. Deterministic construction of the analysis ensemble eliminates one source of sampling error and leads to deterministic SRFs being more accurate than stochastic SRFs in some examples (Whitaker and Hamill 2002; Anderson 2001). SRFs are not unique since different ensembles can have the same covariance. This lack of uniqueness is illustrated in three recently proposed ensemble data assimilation methods that use the Kalman SRF method to update the analysis ensemble (Bishop et al. 2001; Anderson 2001; Whitaker and Hamill 2002). Identifying the methods as SRFs allows a clearer discussion and comparison of their different analysis ensemble updates.

Accounting for small ensemble-size and model deficiencies remains a significant issue in en-

semble data assimilation systems. Schur products can be used to filter ensemble covariances and effectively increase covariance rank (Houtekamer and Mitchell 1998, 2001; Hamill et al. 2001; Whitaker and Hamill 2002). Covariance inflation is one simple way of accounting for model error and stabilizing the filter (Hamill et al. 2001; Anderson 2001; Whitaker and Hamill 2002). Hybrid methods represent forecast error covariances with a combination of ensemble and parameterized correlation models (Hamill and Snyder 2000). Here we have shown deterministic methods of including model error into a square-root or ensemble data assimilation system when the model error has large-scale representation and when the model error is represented by a correlation model. However, the primary difficulty remains obtaining estimates of model error.

Nonuniqueness of SRFs has been exploited in Estimation Theory to design filters with desirable computational and numerical properties. An open question is whether there are ensemble properties that would make a particular SRF implementation better than another, or if the only issue is computational cost. For instance, it may be possible to choose an analysis update scheme that preserves higher-order, non-Gaussian statistics of the forecast ensemble. This question can only be answered by detailed comparisons of different methods in a realistic setting where other details of the assimilation system such as modeling of systematic errors or data quality control may prove to be as important.

APPENDIX

Computational costs

Here we detail the computational cost scalings summarized in Table 1. All the methods require applying the observation operator to the ensemble members to form $\mathbf{H}_k \mathbf{Z}_k^f$, and we do not include its cost. This cost is important when comparing ensemble and non-ensemble methods, particularly for complex observation operators. The cost of computing $\mathbf{H}_k \mathbf{Z}_k^f$ is formally $O(mnp)$, but may be significantly less when $\mathbf{H}_k$ is sparse or can be applied efficiently. We also assume that the inverse observation error covariance $\mathbf{R}_k^{-1}$ is inexpensive to apply.

a. Direct method.

1. Solve $(\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k)\mathbf{Y}_k = \mathbf{H}_k \mathbf{Z}_k^f$ for $\mathbf{Y}_k$. If $\mathbf{R}^{-1}$ is available, the solution can be obtained using the Sherman-Morrison-Woodbury identity (Golub and Van Loan 1996)

$$(\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k)^{-1} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{H}_k \mathbf{Z}_k^f (\mathbf{I} + (\mathbf{H}_k \mathbf{Z}_k^f)^T \mathbf{R}^{-1}(\mathbf{H}_k \mathbf{Z}_k^f))^{-1}(\mathbf{H}_k \mathbf{Z}_k^f)^T \mathbf{R}^{-1},$$

   and only inverting $m \times m$ matrices. Cost: $O(m^3 + m^2 p)$.

2. Form $\mathbf{I} - (\mathbf{H}_k \mathbf{Z}_k^f)^T \mathbf{Y}_k$. Cost $O(pm^2)$.

3. Compute matrix square-root of the $m \times m$ matrix $\mathbf{I} - (\mathbf{H}_k \mathbf{Z}_k^f)^T \mathbf{Y}_k$. Cost: $O(m^3)$.

4. Apply matrix square-root to $\mathbf{Z}_f$. Cost: $O(m^2 n)$.

Total cost: $O(m^3 + m^2 p + m^2 n)$.

b. Serial method.

For each observation:

15

1. Form $\mathbf{D}$. Cost: $O(m)$.

2. Form $\mathbf{I} - \beta\mathbf{V}\mathbf{V}^T$ and apply to $\mathbf{Z}_k^f$. Cost: $O(nm)$.

Total cost: $O(mp + mnp)$.

c. ET KF.

1. Form $\mathbf{Z}^{fT}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{Z}^f$. Assume $\mathbf{R}^{-1}$ inexpensive to apply. Cost: $O(m^2 p)$.

2. Compute eigenvalue decomposition of $m \times m$ matrix. Cost: $O(m^3)$.

3. Apply to $\mathbf{Z}^f$. Cost: $O(m^2 n)$.

Total cost: $O(m^2 p + m^3 + m^2 n)$.

d. EAKF.

Cost in addition to ET KF is:

1. Eigenvalue decomposition of $\mathbf{P}^f$ (low-rank). Cost: $O(m^2 n + m^3)$.

2. Form $\mathbf{F}^T\mathbf{Z}^f$. Cost: $O(m^2 p)$.

Total cost: $O(m^2 p + m^3 + m^2 n)$.

REFERENCES

Anderson, J. L., 2001: An Ensemble Adjustment Filter for Data Assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.

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

Bierman, G. J., 1977: *Factorization methods for Discrete Sequential Estimation*. Mathematics in Science and Engineering, Vol. 128, Academic Press, 241 pp.

Bishop, C. H., B. 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: On the 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.

—, and R. Todling, 1996: Approximate data assimilation schemes for stable and unstable dynamics. *J. Meteor. Soc. Japan*, **74**, 63–75.

—, A. M. da Silva, J. Guo, M. Sienkiewicz, and D. Lamich, 1998: Assessing the Effect of Data Selection with the DAO Physical-space Statistical Analysis System. *Mon. Wea. Rev.*, **126**, 2913–2926.

Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.*, **123**, 1128–1145.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 1043–1062.

Golub, G. H., and C. F. Van Loan, 1996: *Matrix Computations*. 3d ed. The Johns Hopkins University Press, 694 pp.

Hamill, T. M., and C. Snyder, 2000: A Hybrid Ensemble Kalman Filter-3D Variational Analysis Scheme. *Mon. Wea. Rev.*, **128**, 2905–2919.

—, 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.

17

Heemink, A. W., M. Verlaan, and A. J. Segers, 2001: Variance Reduced Ensemble Kalman Filtering. *Mon. Wea. Rev.*, **129**, 1718–1728.

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.

—, L. Lefaivre, J. Derome, H. Ritchie, and H. L. Mitchel, 1996: A System Simulation Approach to Ensemble Prediction. *Mon. Wea. Rev.*, **124**, 1225–1242.

Maybeck, P. S., 1982: *Stochastic models, estimation, and control*, Vol. 1. Academic Press, 423 pp.

Mitchell, H. L., and P. L. Houtekamer, 2000: An Adaptive Ensemble Kalman Filter. *Mon. Wea. Rev.*, **128**, 416–433.

Pham, D., 2001: Stochastic Methods for Sequential Data Assimilation in Strongly Nonlinear Systems. *Mon. Wea. Rev.*, **129**, 1194–1207.

—, J. Verron, and M. Roubaud, 1998: A singular evolutive extended Kalman filter for data assimilation in oceanography. *J. Mar. Syst.*, **16**, 323–340.

Verlaan, M., and A. W. Heemink, 1997: Tidal flow forecasting using reduced rank square filters. *Stochastic Hydrology and Hydraulics*, **11**, 349–368.

Whitaker, J., and T. M. Hamill, 2002: Ensemble Data Assimilation Without Perturbed Observations. *Mon. Wea. Rev.*, **130**, 1913–1924.