A review of the deterministic ensemble Kalman filtering methods and related techniques

A scholarly paper submitted in partial fulfillment of the requirements of the degree of Master of Science in Meteorology

By

Takemasa Miyoshi

Department of Meteorology
University of Maryland

Advisor

Prof. Eugenia Kalnay

July 29, 2004

Abstract

The present paper aims to provide a brief review on several deterministic ensemble Kalman filtering (EnKF) methods and the related practical techniques to prevent filter divergence. Since Evensen (1994), several formulations of EnKF have been proposed, and Whitaker and Hamill (2002) suggested that a deterministic method, a.k.a. an ensemble square root filter (EnSRF, Andrews 1968), is expected to outperform the classical perturbed observation methods (e.g., Houtekamer and Mitchell 1998) especially in a limited ensemble size, which is usually the case in realistic atmospheric models. There are several EnSRFs proposed for atmospheric data assimilation, including an ensemble transform Kalman filter (ETKF) by Bishop et al. (2001), an ensemble adjustment Kalman filter (EAKF) by Anderson (2001), an EnSRF by Whitaker and Hamill (2002), all of which are effective when observational data are assimilated serially, and a local ensemble Kalman filter (LEKF) by Ott et al. (2002; 2004), where observations are assimilated simultaneously in a local patch. As Tippett et al. (2003) indicated, ensemble formulations in the deterministic method are not theoretically unique, and it is not clear if there is a unique preferable choice in the various possible implementations. In addition, there are several practical techniques to avoid filter divergence such as localization of the forecast error correlations, hybrid combinations with three-dimensional variational (3DVAR) method, a double ensemble method, covariance inflation, and random noise addition to analysis ensemble.

1 Introduction

Modern methods of data assimilation are a maximum likelihood estimation that combines background information (usually a forecast state) and measurements (observations) using their second order error statistics. It can be assumed that the measurement error statistics depends only on measurement systems that are independent of flow. However, the background error statistics depend on the flow, and the time evolution of second order error statistics requires N model time integrations, where N denotes the dimension of the phase space of the system. A realistic atmospheric model has millions of dimensions, which makes the direct computation impossible. Thus, it has been necessary for a long time to assume that the background error statistics are constant, i.e. flow-independent, in atmospheric data assimilation. This includes both the optimum interpolation (OI) and the three-dimensional variational (3DVAR) methods used in operational numerical weather prediction.

It is natural to anticipate that the inclusion of flow-dependence of error statistics should improve data assimilation substantially. The recent four-dimensional variational (4DVAR) method implemented at European Centre for Medium-Range Weather Forecasts (ECMWF), MeteoFrance, and Japan Meteorological Agency (JMA) provides an approach to include flow-dependence. Under the assumption of a perfect model, if the initial background error covariance for an assimilation interval is given by Kalman filter (KF), the final 4DVAR analysis is identical to the KF analysis. However, 4DVAR does not provide an updated analysis or background error covariance to start the next cycle. A method that does provide an estimate of the new error covariance is ensemble Kalman filtering (EnKF), first proposed by Evensen (1994), which explicitly includes the time evolution of error statistics, assuming that they are low-rank.

Since then, a number of approaches have been proposed for EnKF including various kinds of implementations for the same basic algorithm of Kalman filtering (Kalman 1960). Independent forecast-analysis cycles for each ensemble member are a typical implementation of EnKF (Evensen 1994; Houtekamer and Mitchell 1998). As Burgers et al. (1998) mentioned, observational data need to be perturbed in the algorithm so that analysis covariance is not underestimated. However, the inclusion of random perturbations in observational data introduces a new source of sampling errors. Thus, Whitaker and Hamill (2002) suggested that the perturbed observation method is expected to be worse than an alternative, deterministic method, especially in a limited ensemble size, which is usually the case in realistic atmospheric models. This

deterministic method solves data assimilation equations for the ensemble mean state and for the analysis error covariance. Analysis ensemble perturbations are computed by a linear combination of forecast ensemble perturbations that span the error covariance, which is why this approach is called square root filter (SRF). There have been several implementations proposed for the ensemble SRFs (EnSRFs) in atmospheric data assimilation. They include an ensemble transform Kalman filter (ETKF) by Bishop et al. (2001), an ensemble adjustment Kalman filter (EAKF) by Anderson (2001), an EnSRF (Andrews 1968) by Whitaker and Hamill (2002), all of which are effective when observational data are assimilated serially, and a local ensemble Kalman filter (LEKF) by Ott et al. (2002; 2004) where observations within a local patch are assimilated simultaneously. As Tippett et al. (2003) indicated, ensemble formulations in the deterministic method are not theoretically unique, and it is not clear if there is any preferable choice among the various possible implementations.

In addition, not only the implementations in SRFs in analysis ensemble update formula, but also practical techniques to avoid filter divergence (i.e., a temporary or permanent separation between the truth and the analysis) caused by sampling errors and covariance underestimation, introduce various different methods. The low-rank assumption that enables EnKF with relatively small ensemble size is a key source of sampling errors of covariance estimation, especially with the error correlation between distant points, which can be addressed by localization. Originally proposed by Houtekamer and Mitchell (1998), the serial treatment of observational data makes localization around observations preferable. The localization was done by defining a cut-off radius and by forcing zero correlation beyond the radius. To avoid the discontinuity in the localization, Houtekamer and Mitchell (2001) advanced the same idea to introduce a Schur product which is a smooth function similar to the Gaussian function but compactly supported (e.g., Gaspari and Cohn 1999; Hamill et al. 2001). Another approach is to introduce local analysis patches, which was introduced by Kalnay and Toth (1994) and Ott et al. (2002; 2004). Other than localization, hybrid use with three-dimensional variational (3DVAR) method provides another solution to avoid filter divergence. Defining error covariance as a weighted mean between flow-independent covariance in 3DVAR and flow-dependent covariance estimated by ensemble, the hybrid method performs significantly better than 3DVAR (Hamill and Snyder 2000; Corazza et al. 2002; Etherton and Bishop 2004) at a relatively low computational cost.

As for covariance underestimation, the main sources are model errors and nonlinearity.

Even under the perfect model assumption, model nonlinearity is an inevitable source because Kalman filtering is a linear estimation theory. Additionally, using the same ensemble to estimate the gain matrix for data assimilation and to obtain representation of analysis errors is a possible source of filter divergence, which was pointed out as an "inbreeding" problem by Houtekamer and Mitchell (1998). They designed a double ensemble method to avoid this problem. Whitaker and Hamill (2002) found that the double ensemble method outperforms single ensemble method in a given ensemble size, though the method used in EnSRFs requires significantly large additional computation. Covariance inflation, that simply multiplies covariance by a slightly larger number than 1, is a common solution for covariance underestimation. Since covariance inflation makes the filter suboptimal, Ott et al. (2004) created an enhanced variance inflation to minimize the suboptimality. Adding small random perturbations to each ensemble member is another solution, though it makes the filter suboptimal.

Thus, there are many possible combinations of SRFs and practical techniques, and it is not known if there is an optimal choice. The present paper aims to provide a brief review on the various methods and techniques.

The outline of the present paper is as follows. The Kalman filter algorithm is introduced, and the perturbed observation method and the EnSRF are described briefly in section 2. In section 3, several implementations of EnSRFs are described. Several practical techniques to prevent filter divergence are described in section 4. Summary and discussions are provided in section 5. The notation basically follows Ide et al. (1997) unless otherwise noted. Throughout the present proposal, N and m denote dimension of the system and ensemble size, respectively, and $\mathbf{I}$ denotes an identity matrix.

2 Overview of ensemble Kalman filtering

The Kalman filter algorithm is given by the following five equations (e.g., Jazwinski 1970; Gelb et al. 1974; Kalnay 2003):

$$\mathbf{x}_i^f = M\mathbf{x}_{i-1}^a \tag{2.1}$$

$$\mathbf{P}_i^f = \mathbf{M}_{\mathbf{x}_{i-1}^a} \mathbf{P}_{i-1}^a \mathbf{M}_{\mathbf{x}_{i-1}^a}{}^T + \mathbf{Q} \tag{2.2}$$

$$\mathbf{K}_i = \mathbf{P}_i^f \mathbf{H}^T [\mathbf{H}\mathbf{P}_i^f \mathbf{H}^T + \mathbf{R}]^{-1} \tag{2.3}$$

$$\mathbf{x}_i^a = \mathbf{x}_i^f + \mathbf{K}_i (\mathbf{y}_i^o - \mathbf{H}\mathbf{x}_i^f) \tag{2.4}$$

$$\mathbf{P}_i^a = [\mathbf{I} - \mathbf{K}_i \mathbf{H}]\mathbf{P}_i^f . \tag{2.5}$$

Here, $\mathbf{M}_{\mathbf{x}_{i-1}^a}$ denotes the tangent linear model corresponding to the nonlinear model M defined around $\mathbf{x}_{i-1}^a$. Define the square root $\mathbf{E}_{i-1}$ of the covariance matrix as $\mathbf{E}_{i-1}\mathbf{E}_{i-1}^T = \mathbf{P}_{i-1}$ and substitute into (2.2), which yields

$$\mathbf{E}_i^f \mathbf{E}_i^{fT} = \mathbf{M}_{\mathbf{x}_{i-1}^a} \mathbf{E}_{i-1}^a \mathbf{E}_{i-1}^{aT} \mathbf{M}_{\mathbf{x}_{i-1}^a}^T + \mathbf{Q} = \mathbf{M}_{\mathbf{x}_{i-1}^a} \mathbf{E}_{i-1}^a \left(\mathbf{M}_{\mathbf{x}_{i-1}^a} \mathbf{E}_{i-1}^a\right)^T + \mathbf{Q} . \tag{2.6}$$

Note that the square root is not unique because $\mathbf{E}\mathbf{U}$ is also a square root if $\mathbf{U}$ is any arbitrary unitary matrix ($\mathbf{U}\mathbf{U}^T = \mathbf{I}$). (2.6) gives ensemble-forecasting formula:

$$\begin{aligned} \mathbf{E}_i^f &= \mathbf{M}_{\mathbf{x}_{i-1}^a} \mathbf{E}_{i-1}^a = \mathbf{M}_{\mathbf{x}_{i-1}^a} \left(\delta\mathbf{x}_{i-1}^1 \quad \cdots \quad \delta\mathbf{x}_{i-1}^N\right) \\ &\cong \left(M(\mathbf{x}_{i-1}^a + \delta\mathbf{x}_{i-1}^1) - M(\mathbf{x}_{i-1}^a) \quad \cdots \quad M(\mathbf{x}_{i-1}^a + \delta\mathbf{x}_{i-1}^N) - M(\mathbf{x}_{i-1}^a)\right) \end{aligned} \tag{2.7}$$

where we ignore $\mathbf{Q}$, the error covariance introduced by model deficiencies, and $\delta\mathbf{x}_{i-1}^n$ denotes the nth column of the matrix $\mathbf{E}^a$, which can be regarded as an ensemble perturbation vector. Here, the tangent linear model is approximated by the original nonlinear model under the assumption that $\delta\mathbf{x}$ is small compared to $\mathbf{x}$. In realistic atmospheric models, N is a huge number of the order of millions, and (2.7) requires $N+1$ model integrations, which makes this method practically impossible. However, when the covariance matrix is degenerate, i.e. when most of the eigenvalues of the covariance matrix are very close to 0, which is usually the case in realistic atmospheric models, a covariance matrix could be approximated by relatively small ensemble size m much smaller than N. This fact enables the matrix $\mathbf{E}^a$ to be $N \times m$, thus, (2.2) can be realized by an ensemble forecasting with relatively small ensemble size.

The next step is to obtain the gain matrix (2.3). One approach is a sequential treatment of observations. Theoretically, uncorrelated observation can be treated serially (see for example,

Houtekamer and Mitchell 2001), which reduces the dimension of an observational space. Since the observational error correlation matrix $\mathbf{R}$ is usually almost diagonal, the reduced dimension of an observational space is an order of 1. Using the square root of covariance matrix, (2.3) can be written as

$$\mathbf{K}_i = \mathbf{E}_i^f (\mathbf{H}\mathbf{E}_i^f)^T [\mathbf{H}\mathbf{E}_i^f (\mathbf{H}\mathbf{E}_i^f)^T + \mathbf{R}]^{-1} \tag{2.8}$$

This equation requires transformation from ensemble to observational space and inverse of a matrix in observational space that is a scalar when observational errors are uncorrelated. Thus, using the ensemble, (2.3) can be directly computed. In the case that the dimension of observational space is larger than the ensemble size, (2.8) can be further simplified as

$$\mathbf{K}_i = \mathbf{E}_i^f [\mathbf{I} + (\mathbf{H}\mathbf{E}_i^f)^T \mathbf{R}^{-1} \mathbf{H}\mathbf{E}_i^f]^{-1} (\mathbf{H}\mathbf{E}_i^f)^T \mathbf{R}^{-1} \tag{2.9}$$

which requires inverse of an $m \times m$ matrix, where m denotes ensemble size as previously defined. The derivation of (2.9) is provided in Appendix A.

The final step is an ensemble update given by (2.5). In the classical approach to ensemble Kalman filtering, in which the forecast-analysis cycles are computed independently for each member, the analysis ensemble satisfies $\mathbf{E}_i^a = [\mathbf{I} - \mathbf{K}_i \mathbf{H}]\mathbf{E}_i^f$, which yields $\mathbf{P}_i^a = [\mathbf{I} - \mathbf{K}_i \mathbf{H}]\mathbf{P}_i^f [\mathbf{I} - \mathbf{K}_i \mathbf{H}]^T$ (e.g., Burgers et al. 1998; see also Appendix B). The factor $[\mathbf{I} - \mathbf{K}_i \mathbf{H}]$ represents reduction of covariance due to data assimilation, so the analysis covariance is reduced too much (cf. 2.5). Thus, Burgers et al. (1998) discussed the necessity of observational ensemble which is realized by perturbing the observations. Since each member is analyzed by (2.4), (2.5) is derived from (2.4) as follows in the presence of observational perturbations:

$$\mathbf{P}_i^a = [\mathbf{I} - \mathbf{K}_i \mathbf{H}]\mathbf{P}_i^f [\mathbf{I} - \mathbf{K}_i \mathbf{H}]^T + \mathbf{K}_i \mathbf{R} \mathbf{K}_i^T = [\mathbf{I} - \mathbf{K}_i \mathbf{H}]\mathbf{P}_i^f \tag{2.10}$$

(cf. (3) in Whitaker and Hamill 2002). The second term that comes from observational perturbation is important to prevent the excessive reduction of the estimation of the analysis error. This is the reason why perturbed observations are necessary in the classical EnKF (see also

Appendix B).

However, as Whitaker and Hamill (2002) suggested, perturbed observations introduce a new source of sampling errors, which is expected to be worse than deterministic algorithms, a.k.a. square root filters, especially in a limited ensemble size, which is usually the case in realistic atmospheric models. In general, deterministic algorithms solve (2.5) explicitly. Assume the analysis ensemble is given by linear combination of the forecast ensemble: $\mathbf{E}^a = \mathbf{E}^f \mathbf{T}$. Then, (2.5) is written as

$$\mathbf{E}^f \mathbf{T}\mathbf{T}^T \mathbf{E}^{f\,T} = [\mathbf{I} - \mathbf{K}\mathbf{H}]\mathbf{E}^f \mathbf{E}^{f\,T} \tag{2.11}$$

Solving for $\mathbf{T}$, one can get a solution for a deterministic ensemble update which satisfies (2.5). Note that the choice of $\mathbf{T}$, like the square root of covariance matrix, is not unique, i.e. $\mathbf{T}\mathbf{U}$ is also a solution. There are several implementations of EnSRFs that give solutions of (2.5), which are described in the next section (see also, Tippett et al. 2003).

3 EnSRFs

3.1 Serial method (Whitaker and Hamill 2002)

This method assumes ensemble update of the form

$$\mathbf{E}^a = [\mathbf{I} - \tilde{\mathbf{K}}\mathbf{H}]\mathbf{E}^f \tag{3.1.1}$$

In order for (3.1.1) to give a solution of (2.5),

$$[\mathbf{I} - \tilde{\mathbf{K}}\mathbf{H}]\mathbf{P}^f [\mathbf{I} - \tilde{\mathbf{K}}\mathbf{H}]^T = [\mathbf{I} - \mathbf{K}\mathbf{H}]\mathbf{P}^f \tag{3.1.2}$$

has to be satisfied (cf. (9) in Whitaker and Hamill 2002). In this algorithm, data assimilation is performed only once on the ensemble mean using (2.4), and (3.1.1) is used for ensemble update. The solution of $\tilde{\mathbf{K}}$ is given by

$$\tilde{\mathbf{K}} = \mathbf{P}^f \mathbf{H}^T [(\mathbf{H}\mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1/2}]^T [(\mathbf{H}\mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{1/2} + \mathbf{R}^{1/2}]^{-1} \tag{3.1.3}$$

(Andrews 1968) cf. (10) in Whitaker and Hamill (2002). This formula is used for correlated observations. If observations are uncorrelated, i.e. $\mathbf{R}$ is diagonal, each observation is treated serially, which makes the terms $\mathbf{H}\mathbf{P}^f \mathbf{H}^T$ and $\mathbf{R}$ scalar. In this case, (3.1.2) can be simplified, and assuming $\tilde{\mathbf{K}} = \alpha \mathbf{K}$ where α is a scalar value, one can get the formula for α:

$$\alpha = \left(1 + \sqrt{\frac{\mathbf{R}}{\mathbf{H}\mathbf{P}^f \mathbf{H}^T + \mathbf{R}}} \right)^{-1} \tag{3.1.4}$$

which was first derived by Potter in 1964, cf. (13) in Whitaker and Hamill (2002). Thus, the explicit computation of $\tilde{\mathbf{K}}$ can be avoided, and almost no additional computation is required.

3.2 ETKF (Bishop et al. 2001)

This algorithm assumes an ensemble update of the form

$$\mathbf{E}^a = \mathbf{E}^f \mathbf{T} \tag{3.2.1}$$

where $\mathbf{T}$ is an $m \times m$ matrix called transformation matrix. The solution of $\mathbf{T}$ is given by

$$\mathbf{T} = \mathbf{C}[\mathbf{\Gamma} + \mathbf{I}]^{-1/2} \tag{3.2.2}$$

cf. (11) in Bishop et al. (2001). Here, $\mathbf{C}$ is an $m \times m$ matrix composed of eigenvectors of $\mathbf{E}^{f\,T} \mathbf{H}^T \mathbf{R}^{-1/2\,T} \mathbf{R}^{-1/2} \mathbf{H} \mathbf{E}^f$, and $\mathbf{\Gamma}$ is a diagonal matrix composed of eigenvalues of the same matrix. Thus, the bracket of (3.2.2) is diagonal, so it is easy to compute the inverse of square root.

3.3 EAKF (Anderson 2001)

This algorithm aims to apply rotation and scaling (i.e. coordinate transformation) to the original space, in which the covariance can be treated as an identity matrix. The adjustment

process, denoted by $\mathbf{A}$, is applied to forecast ensemble to obtain analysis ensemble, that is, $\mathbf{E}^a = \mathbf{A}\mathbf{E}^f$. Let $\mathbf{F}\mathbf{D}\mathbf{F}^T$ be the singular value decomposition of $\mathbf{P}^f$. By virtue of non-negative definite symmetric matrix, singular value decomposition is equivalent to eigenvalue decomposition with $\mathbf{F}^T = \mathbf{F}^{-1}$. Because of the degeneracy of $\mathbf{P}^f$, there are only m non-negative eigenvalues (or singular values) at most, which are equal to the eigenvalues of the $m \times m$ matrix $\mathbf{E}^{f\,T}\mathbf{E}^f$. The adjustment $\mathbf{A}$ is given by

$$\mathbf{A} = \mathbf{F}\mathbf{D}^{1/2}\mathbf{U}[\mathbf{I} + \mathbf{D}]^{-1/2}\mathbf{D}^{-1/2}\mathbf{F}^T \tag{3.3.1}$$

as of given in the appendix A of Anderson (2001). Here, $\mathbf{U}$ is a unitary matrix ($\mathbf{U}\mathbf{U}^T = \mathbf{I}$). As Tippett et al. (2003) described, this is equivalent to the ensemble update of

$$\mathbf{E}^a = \mathbf{E}^f\mathbf{C}[\mathbf{I} + \mathbf{\Gamma}]^{-1/2}\mathbf{D}^{-1/2}\mathbf{F}^T\mathbf{E}^f \tag{3.3.2}$$

cf. (20) in Tippett et al. (2003). Here, $\mathbf{C}$ and $\mathbf{\Gamma}$ are the same as in (3.2.2) of the previous subsection. Because of the degeneracy of $\mathbf{P}^f$, $\mathbf{D}$ is an $m \times m$ matrix, and $\mathbf{F}$ is an $N \times m$ matrix in this equation. The difference from the ETKF is the factor $\mathbf{D}^{-1/2}\mathbf{F}^T\mathbf{E}^f$.

3.4 LEKF (Ott et al. 2002; 2004)

The important differences of LEKF are in the way of the localization, which is described in the next section. In this section, apart from the localization, we describe the data assimilation and ensemble update scheme in the LEKF which is the core concept of SRF.

A low-dimensional subspace (hereafter, “hat space”) spanned by the ensemble members is defined in the LEKF for the analysis step, and all the analysis computations are performed in this “hat space”. Generally, the forecast ensemble perturbations are not orthogonal to each other. Thus, an $N \times \hat{m}$ matrix $\mathbf{G}$, which is similar to the forecast ensemble $\mathbf{E}^f$, is formed from eigenvectors of $\mathbf{P}^f$ whose corresponding eigenvalues are not zero. Here, $\hat{m}$ denotes number of non-zero eigenvalues, which satisfies $\hat{m} \le m$. Thus, $\mathbf{G}\mathbf{G}^T$ gives a good approximation of $\mathbf{P}^f$. Now, $\mathbf{G}$ can be regarded as coordinate transformation from the N-dimensional real space to the $\hat{m}$-dimensional space that each eigenvector forms, i.e. “hat space”, because other

directions are not important (null space by virtue of the zero eigenvalues). Thus, an arbitrary vector in N-dimensional real space $\mathbf{v}$ is transformed to an $\hat{m}$-dimensional vector $\hat{\mathbf{v}}$ by $\hat{\mathbf{v}} = \mathbf{G}^T \mathbf{v}$. Similarly, an arbitrary $N \times N$ matrix $\mathbf{A}$ is transformed to an $\hat{m} \times \hat{m}$ matrix $\hat{\mathbf{A}}$ by $\hat{\mathbf{A}} = \mathbf{G}^T \mathbf{A} \mathbf{G}$. Since the other directions are not important as mentioned above, everything in the real space can be treated in the "hat space" throughout the whole data assimilation processes. Thus, Kalman filter equations (2.3), (2.4), and (2.5) are computed in the "hat space". Note that (2.1) and (2.2) cannot be done in the "hat space" because model integration requires real space. The coordinate transformation $\mathbf{G}$ is formed once at every analysis step, then the data assimilation and ensemble update is computed in the "hat space", and finally, updated ensembles are transformed to real space by the inverse transformation, which is just $\mathbf{G}$ itself.

To enable simultaneous treatment of observational data, the LEKF solves (2.2) and (2.3) by the following equations (cf. (22) and (23) of Ott et al. 2004):

$$\mathbf{x}_i^a = \mathbf{x}_i^f + \hat{\mathbf{P}}_i^a \hat{\mathbf{H}}^T \mathbf{R}^{-1} (\mathbf{y}_i^o - \mathbf{H}\mathbf{x}_i^f) \tag{3.4.1}$$

$$\hat{\mathbf{P}}_i^a = [\hat{\mathbf{P}}_i^{b\,-1} + \hat{\mathbf{H}}^T \mathbf{R}^{-1} \hat{\mathbf{H}}]^{-1} = \hat{\mathbf{P}}_i^b [\mathbf{I} + \hat{\mathbf{H}}^T \mathbf{R}^{-1} \hat{\mathbf{H}} \hat{\mathbf{P}}_i^b]^{-1} \tag{3.4.2}$$

Thus, the matrix inverse in the observational space can be avoided and (3.4.2) requires matrix inverse in the "hat space".

The ensemble update is formed as $\hat{\mathbf{E}}^a = \hat{\mathbf{E}}^f \mathbf{Y}$ which is similar to ETKF. To satisfy ensemble update formula (2.5) in the "hat space", $\hat{\mathbf{E}}^f \mathbf{Y}\mathbf{Y}^T \hat{\mathbf{E}}^{f\,T} = [\mathbf{I} - \hat{\mathbf{K}}\mathbf{H}]\hat{\mathbf{E}}^f \hat{\mathbf{E}}^{f\,T}$ has to be satisfied, which is the "hat space" version of (2.11). This $\mathbf{Y}$ is not unique as mentioned before, but the "optimal" choice is provided as (42) of Ott et al. (2004), that is,

$$\mathbf{Y} = [\mathbf{I} + \mathbf{E}^{f\,T} \hat{\mathbf{P}}^{f\,-1} (\hat{\mathbf{P}}^a - \hat{\mathbf{P}}^f) \hat{\mathbf{P}}^{f\,-1} \mathbf{E}^f]^{1/2} \tag{3.4.3}$$

4 Related techniques to prevent filter divergence

The previous section describes the theory of EnSRFs, but in practice, these are not sufficient for stable filtering. Usually, the assumptions of the Kalman filter are not strictly

satisfied because of various reasons such as model nonlinearity, systematic errors, non-Gaussian error statistics, and sampling errors due to limited ensemble size, which is why the filter tends to diverge. To prevent the filter divergence, several techniques have been proposed, which are described in this section.

4.1 Localization (Houtekamer and Mitchell 2001, Ott et al. 2004)

An important problem of EnKF is that the ensemble size m is merely of order 100 due to computational limitations, which is very small compared to the system size N that is usually an order of millions. Thus, the approximation $\mathbf{P} = \mathbf{E}\mathbf{E}^T$ using an $N \times m$ matrix $\mathbf{E}$ composed of ensemble contains large sampling errors especially in the error covariance between distant points (graphically, see for example, fig.6 of Houtekamer and Mitchell (1998); fig.10 and fig.11 of Hamill and Snyder (2000)). To prevent the sampling errors at large distances, covariance localization is a straightforward and simple solution. Houtekamer and Mitchell (1998) introduced a cut-off radius that defines the distance that an observation affects. This is also favorable since N becomes smaller in analysis computation, but the cut-off process is discontinuous, which causes noisiness in the analyzed field. Thus, Houtekamer and Mitchell (2001) advanced the idea to introduce a Schur product, which applies a smooth function known as a fifth-order piecewise rational function ((4.10) of Gaspari and Cohn (1999)) to the terms $\mathbf{P}^f\mathbf{H}^T$ and $\mathbf{H}\mathbf{P}^f\mathbf{H}^T$. This localization method is also used in Anderson (2001), and more precise explanation is found in Hamill et al. (2001).

A different method of localization was proposed by Ott et al. (2002; 2004), which utilizes local analysis patches based on the findings of Patil et al. (2001) and Kalnay and Toth (1994). A local patch around every grid point is constructed, and data assimilation is performed for every local patch independently. This procedure automatically forces zero covariance outside the patch. The overlapped analyses may be averaged to obtain a final product. This averaging process produces a smooth analysis. Each local patch has much smaller dimension than the whole domain (Patil et al. 2001), which makes the approximation $\mathbf{P} = \mathbf{E}\mathbf{E}^T$ much easier with a limited ensemble size.

4.2 EnKF/3DVAR hybrid method (Hamill and Snyder 2000)

With small ensemble size, the localization may not be sufficient to prevent filter

divergence; eventually, EnKF may not work at all, while 3DVAR is a rather robust method. The hybrid method aims to obtain better analysis than 3DVAR using flow-dependent information from ensemble. 3DVAR looks for the solution that minimizes the cost function

$$J(\mathbf{x}) = \tfrac{1}{2}[(\mathbf{x}-\mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^b) + (\mathbf{y}^o - \mathbf{H}\mathbf{x})^T \mathbf{R}^{-1}(\mathbf{y}^o - \mathbf{H}\mathbf{x})] \qquad (4.2.1)$$

where the background error covariance $\mathbf{B}$ is constant in time. Usually, dimension of $\mathbf{B}$ is so huge that many assumptions, such as isotropic and spatial uniformity, need to be made. With the assumptions, $\mathbf{B}$ is approximated as $\mathbf{B} = \mathbf{S}\mathbf{C}\mathbf{S}^T$, where variable transformation $\mathbf{S}$ makes the covariance $\mathbf{C}$ diagonal. The hybrid method mixes the constant covariance with flow-dependent covariance by

$$\mathbf{B} = (1-\alpha)\mathbf{E}^f \mathbf{E}^{f\,T} + \alpha\mathbf{S}\mathbf{C}\mathbf{S}^T \qquad (4.2.2)$$

where α is a numeric between 0 and 1 that determines the weight of flow-dependence. Thus, insufficient ensemble representation requires α close to 1, whereas better representation favors a smaller value of α.

4.3 Double ensemble

Houtekamer and Mitchell (1998) pointed out the usefulness of a double ensemble, which aims to prevent the “inbreeding” effect that is a possible source of filter divergence. In this method, two parallel ensemble data assimilation cycles are computed separately using the gain matrix estimated from the other batch of ensemble. Houtekamer and Mitchell (1998; 2001) described the usefulness of double ensemble by comparing to single ensemble with the same total ensemble size. Whitaker and Hamill (2002) also discussed the usefulness of double ensemble in their appendix. However, at the same time, they mentioned the increase of computational cost in an EnSRF because it requires two separate ensemble update processes.

4.4 Covariance inflation

Covariance inflation (e.g., Anderson and Anderson 1999; Whitaker and Hamill 2002) is a

simple and straightforward idea to prevent covariance underestimation. Covariance inflation multiplies a number slightly larger than 1 to covariance, i.e., $\widetilde{\mathbf{P}}^f = \mathbf{P}^f \times (1+\delta)$, where δ is a small positive number called an inflation factor. This process can be done after the assimilation process, i.e.,

$$\widetilde{\mathbf{P}}^a = \mathbf{P}^a \times (1+\delta) \tag{4.4.1}$$

which is equivalent to lengthening ensemble perturbation vectors by a factor $\sqrt{1+\delta}$. These processes make the filter suboptimal.

Ott et al. (2004) advanced the idea of "enhanced variance inflation" so that the suboptimality by the covariance inflation becomes minimal. The basic hypothesis is that the covariance underestimation is more important when a forecast departure ($\mathbf{y}^o - \mathbf{H}\mathbf{x}^f$) is large, which is based on the idea that the filter divergence means the larger and larger departure. A forecast departure is equivalent to summation of an observational error and a forecast error, where the former is constant in time. Consequently, a forecast departure represents a forecast error as well as an analysis error. Thus, the enhanced variance inflation defines an inflation factor δ proportional to analysis error, which is measured by the trace of $\mathbf{P}^a$, to yield

$$\widetilde{\mathbf{P}}^a = \mathbf{P}^a + \varepsilon \frac{tr(\mathbf{P}^a)}{m} \times \mathbf{I} \tag{4.4.2}$$

where $tr(\bullet)$ denotes trace, and ε is a constant. (4.4.2) inflates only diagonal components of $\mathbf{P}^a$, i.e. variance, which is different from (4.4.1) that inflates the whole covariance.

4.5 Adding random errors to the ensemble analyses (Corazza et al. 2002)

This method is adding small random errors to the ensemble analyses

$$\widetilde{\mathbf{E}}^a = \mathbf{E}^a + \boldsymbol{\mu}(0, \varepsilon) \tag{4.5.1}$$

where $\boldsymbol{\mu}(0, \varepsilon)$ denotes an $N \times m$ matrix whose components are random numbers with

zero-mean and standard deviation of ε where ε is a small number. (4.5.1) looks similar to the square root of covariance inflation (4.4.1), but (4.5.1) introduces another source of sampling errors. This method can be interpreted as a representation of model error covariance $\mathbf{Q}$ in (2.2) ($\mathbf{Q} = \boldsymbol{\mu}\boldsymbol{\mu}^T$), but random errors are added not to forecast ensemble but to analysis ensemble in (4.5.1). Since we do not have sufficient knowledge about model error covariance, the addition of perturbations can be interpreted as representing the model nonlinearity or other sources of model error covariance even under the perfect model assumption.

Alternatively, it has been known that random errors "seed" bred vectors to keep them "young" (Corazza et al. 2003), which is favorable to data assimilation. Corazza et al. (2002) found that adding random perturbations to bred vectors significantly improved data assimilation skill using an EnKF/3DVAR hybrid method in a quasi-geostrophic channel model. In addition, Miyoshi and Kalnay (2004) found that random error addition ("stochastic seeding") let bred vectors capture instabilities that are missed by the leading bred vector but captured by higher order orthogonal bred vectors using Lorenz 96 model (Lorenz 1996). In view of the similarity of breeding cycle to EnKF, adding random errors acts not only as a representation of the model error covariance but also as the "stochastic seeding" to keep analysis ensemble members "young" spanning additional possible unstable directions, though adding random errors might introduce another source of sampling errors.

5 Discussion

In the present paper, several formulations of EnSRF (a serial method, ETKF, EAKF, and LEKF) and several practical techniques to prevent filter divergence (localization, an EnKF/3DVAR hybrid method, a double ensemble method, covariance inflation, and random error addition to each ensemble member) are introduced and described. However, as Tippett et al. (2003) indicated, it is not clear if there is an optimal choice among the various possible formulations and techniques. Thus, it is important to compare the methods and techniques in a unified way, but there has been no such research so far; the comparison is still an open question.

Another important problem is on model errors, which is not mentioned in the present paper. It is known that advanced data assimilation schemes including EnKF are more sensitive to model errors compared to flow-independent schemes (e.g., Miyoshi 2004). A Kalman filter seeks a truly

optimal solution assuming a perfect linear model, thus, merely the nonlinearity of a model can introduce ill conditioning into a Kalman filter, which leads to filter divergence. It is natural to anticipate that an imperfect model that contains various kinds of biases can be a more significant source of ill conditioning. Thus, it has been a common understanding that the treatment of model errors is a very important issue in applying EnKF to realistic situations such as actual numerical weather prediction. Since EnKF has been mainly applied to ideal conditions under perfect model assumptions in related research so far, the problem had not been treated in a comprehensive manner.

Dee and da Silva (1998) discussed forecast bias estimation in the context of data assimilation. The basic idea is considering a forecast bias as a time evolving variable for which we also have observation: the departure $\mathbf{y}^o - \mathbf{H}\mathbf{x}^f$. Thus, if the time evolving law of the bias is given by $\mathbf{b}^f = \tilde{M}\mathbf{b}^a$ where $\mathbf{b}$ and $\tilde{M}$ denote a bias variable and a bias-evolving model, respectively, the most probable bias is estimated using Kalman filtering theory. Dee and da Silva (1998) used persistence model (identity) as the "forecast" model $\tilde{M}$ for the bias because of the lack of knowledge of the bias-evolving model. Although there have been some investigations on the dynamics of model errors (e.g., Zupanski 1997; Nicolis 2003), nothing seems to have provided an accurate bias-evolving model so far.

This basic concept naturally leads to the possibility of ensemble formulation of bias estimation. Thus, assuming ensemble of biases, the same technique as EnKF can be used. The source code for filtering may be shared, thus, no significant additional effort would be required. The flaw of the bias-evolving model could be treated by additional bias variables that define "bias of bias". In principle, this process could be repeated as many times as desired, though the additional variables require additional memory spaces and filtering computations.

Another issue is treating asynoptic observations such as satellite observations whose timing and place are not fixed. To treat observations of different timing at the same time, 4DVAR provides a common solution. However, 4DVAR requires tangent linear and adjoint versions of a forecast model, whose development and maintenance cost is large. Furthermore, 4DVAR does not provide an updated analysis error covariance to start the next assimilation cycle, so that alternative approaches such as combining 4DVAR with a reduced rank Kalman filter (RRKF, Fisher 1998; Fischer et al. 2001) have been tried without much success. An alternative way has been proposed by Hunt et al. (2004), so-called 4D-EnKF. The basic assumption of this method is

that ensemble members span all the analyzed directions in EnKF. In other words, linear combination of ensemble members can express all directions in analysis. Ensemble forecasting provides time evolution of ensemble members, thus, if observational data are expressed by a linear combination of ensemble members, the observations of different timing can be obtained by the same linear combination of the ensemble members. More precisely, if the timings of observation $\mathbf{y}_j^o$ and background $\mathbf{x}_i^f$ are different ($i \neq j$), the term $\mathbf{y}_j^o - \mathbf{H}\mathbf{x}_i^f$ can be treated as $\mathbf{y}_j^o - \mathbf{H}\mathbf{E}_j(\mathbf{E}_i^T\mathbf{E}_i)^{-1}\mathbf{E}_i^T\mathbf{x}_i^f$ (cf. (8) of Hunt et al. 2004) where each column of $\mathbf{E}$ is an ensemble member as defined before. Using this technique, observations of different timings can be treated in EnKF just by storing ensemble members at different timings.

EnKF has important advantages in the context of application in operational numerical weather prediction: (1) low cost of development and maintenance because it does not require tangent linear and adjoint versions of forecast models; (2) efficiency with ensemble prediction system that most operational numerical weather prediction centers already have. Because of the advantages, EnKF is a possible direction in the future operational ensemble forecast-analysis system, but at the moment, the new technique is not ready for operational implementation because of the lack of experience and the open questions discussed above. With further research and investigations, this new technique has great potential to be implemented operationally and to provide significantly improved products.

Acknowledgements

The author expresses deep gratitude to his advisor Prof. Eugenia Kalnay for her kind and valuable advising. The author thanks Dr. Joaquim Ballabrera for his valuable comments on the manuscript. The author's graduate study is supported by the Japanese Government Long-term Fellowship Program.

Appendices

A. Derivation of (2.9)

Using a formula of linear algebra:

$$[X_1 + X_{12}X_2^{-1}X_{21}]^{-1} = X_1^{-1} - X_1^{-1}X_{12}[X_2 + X_{21}X_1^{-1}X_{12}]^{-1}X_{21}X_1^{-1} \tag{A.1}$$

the RHS of (2.3) is transformed as follows:

$$\begin{aligned}
&\mathbf{P}\mathbf{H}^T[\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R}]^{-1} = \mathbf{E}(\mathbf{H}\mathbf{E})^T[\mathbf{H}\mathbf{E}(\mathbf{H}\mathbf{E})^T + \mathbf{R}]^{-1} \\
&= \mathbf{E}(\mathbf{H}\mathbf{E})^T[\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{H}\mathbf{E}[\mathbf{I} + (\mathbf{H}\mathbf{E})^T\mathbf{R}^{-1}\mathbf{H}\mathbf{E}]^{-1}(\mathbf{H}\mathbf{E})^T\mathbf{R}^{-1}] \\
&= \mathbf{E}(\mathbf{H}\mathbf{E})^T\mathbf{R}^{-1} - \mathbf{E}(\mathbf{H}\mathbf{E})^T\mathbf{R}^{-1}\mathbf{H}\mathbf{E}[\mathbf{I} + (\mathbf{H}\mathbf{E})^T\mathbf{R}^{-1}\mathbf{H}\mathbf{E}]^{-1}(\mathbf{H}\mathbf{E})^T\mathbf{R}^{-1} \\
&= \mathbf{E}[\mathbf{I} - (\mathbf{H}\mathbf{E})^T\mathbf{R}^{-1}\mathbf{H}\mathbf{E}[\mathbf{I} + (\mathbf{H}\mathbf{E})^T\mathbf{R}^{-1}\mathbf{H}\mathbf{E}]^{-1}](\mathbf{H}\mathbf{E})^T\mathbf{R}^{-1} \\
&= \mathbf{E}[[\mathbf{I} + (\mathbf{H}\mathbf{E})^T\mathbf{R}^{-1}\mathbf{H}\mathbf{E}] - (\mathbf{H}\mathbf{E})^T\mathbf{R}^{-1}\mathbf{H}\mathbf{E}][\mathbf{I} + (\mathbf{H}\mathbf{E})^T\mathbf{R}^{-1}\mathbf{H}\mathbf{E}]^{-1}(\mathbf{H}\mathbf{E})^T\mathbf{R}^{-1} \\
&= \mathbf{E}[\mathbf{I} + (\mathbf{H}\mathbf{E})^T\mathbf{R}^{-1}\mathbf{H}\mathbf{E}]^{-1}(\mathbf{H}\mathbf{E})^T\mathbf{R}^{-1}
\end{aligned} \tag{A.2}$$

B. Perturbed Observation

Let $\delta\mathbf{x}^f$, $\delta\mathbf{x}^a$, and $\delta\mathbf{y}^o$ be perturbations of $\mathbf{x}^f$, $\mathbf{x}^a$, and $\mathbf{y}^o$, respectively, that is, $\mathbf{x}^f = \overline{\mathbf{x}}^f + \delta\mathbf{x}^f$, and so on, where an overbar denotes a mean. Then, (2.4) is written as

$$\overline{\mathbf{x}}^a + \delta\mathbf{x}^a = \overline{\mathbf{x}}^f + \delta\mathbf{x}^f + \mathbf{K}[\overline{\mathbf{y}}^o + \delta\mathbf{y}^o - \mathbf{H}(\overline{\mathbf{x}}^f + \delta\mathbf{x}^f)] \tag{B.1}$$

Mean states satisfy (2.4), which yields

$$\delta\mathbf{x}^a = \delta\mathbf{x}^f + \mathbf{K}[\delta\mathbf{y}^o - \mathbf{H}\delta\mathbf{x}^f] \tag{B.2}$$

Without observational perturbation ($\delta\mathbf{y}^o = 0$), which is usual in independent data assimilation, (B.2) is simplified as

$$\delta \mathbf{x}^a = [\mathbf{I} - \mathbf{KH}]\delta \mathbf{x}^f \qquad \text{(B.3)}$$

This is equivalent to $\mathbf{E}^a = [\mathbf{I} - \mathbf{KH}]\mathbf{E}^f$, which gives excessive reduction of perturbation variance. In the presence of observational perturbation, however, (B.2) is the equation for ensemble update, whose perturbation variance is larger than that of (B.3) by the term $+\mathbf{K}\delta \mathbf{y}^o$. The covariance $\delta \mathbf{y}^o (\delta \mathbf{y}^o)^T = \mathbf{R}$ and cross terms such as $\mathbf{H}\delta \mathbf{x}^f (\delta \mathbf{y}^o)^T = 0$ yield the correct solution (2.10).

References

Anderson, J. L. 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.

Anderson, J. L., 2001: An Ensemble Adjustment Kalman 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.

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis Scheme in the Ensemble Kalman Filter. *Mon. Wea. Rev.*, **126**, 1719-1724.

Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive Sampling with Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. *Mon. Wea. Rev.*, **129**, 420-436.

Corazza, M, E. Kalnay, D. J. Patil, E. Ott, J. Yorke, I. Szunyogh, M. Cai, 2002: Use of the breeding technique in the estimation of the background error covariance matrix for a quasigeostrophic model. *AMS Symposium on Observations, Data Assimilation and Probabilistic Prediction*, Orland, Florida, 154-157.

Corazza, M., E. Kalnay, D. J. Patil, I. Szunyogh, S.-C. Yang, 2003: Keeping the bred vectors young: Impact on data assimilation. The presentation file available at http://www.atmos.umd.edu/~ekalnay/.

Dee, D. P. and A. M. da Silva, 1998: Data assimilation in the presence of forecast bias. *Quart. J. Roy. Meteor. Soc.*, **124**, 269-295.

Etherton, B. J. and C. H. Bishop, 2004: Resilience of Hybrid Ensemble/3DVAR Analysis Schemes to Model Error and Ensemble Covariance Error. *Mon. Wea. Rev.*, **132**, 1065-1080.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99** (C5), 10143-10162.

Fisher, M., 1998: Development of a simplified Kalman Filter. *ECMWF Tech. Memo.*, No. 260, 16pp.

Fisher, M., L. Isaksen, M. Ehrendorfer, A. Beck and E. Andersson, 2003: A critical evaluation of the reduced-rank Kalman filter (RRKF) approach to flowdependent cycling of background error covariances. *ECMWF Tech. Memo.*

Gaspari, G. and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723-757.

Gelb, A., J. F. Kasper, R. A. Nash, C. F. Price, and A. A. Sutherland, 1974: *Applied Optimal Estimation*. The M.I.T. Press, 374pp.

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

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.

Hamill, T. M., 2003: Ensemble-based Atmospheric Data Assimilation: A Tutorial. University of Colorado and NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado, 46pp.

Houtekamer, P. L. and H. L. Mitchell, 1998: Data Assimilation Using an Ensemble Kalman Filter Technique. *Mon. Wea. Rev.*, **126**, 796-811.

Houtekamer, P. L. and H. L. Mitchell, 1999: Reply. *Mon. Wea. Rev.*, **127**, 1378-1379.

Houtekamer, P. L. and H. L. Mitchell, 2001: A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. *Mon. Wea. Rev.*, **129**, 123-137.

Hunt, B. R., E. Kalnay, E. J. Kostelich, E. Ott, D. J. Patil, T. Sauer, I. Szunyogh, J. A. Yorke, and A. V. Zimin, 2004: Four-Dimensional Ensemble Kalman Filtering. *Tellus*, in press.

Ide, K., P. Courtier, M. Ghil, and A. C. Lorenc, 1997: Unified notation for data assimilation: Operational, sequential, and variational. *J. Meteor. Soc. Japan*, **75** (1B), 181-189.

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

Kalman, R. E., 1960: A New Approach to Linear Filtering and Prediction Problems. *J. Basic Eng., Trans. ASME*, 35-45.

Kalnay, E. and Z. Toth, 1994: Removing growing errors in the analysis cycle. *Preprints, Tenth Conference on Numerical Weather Prediction, Amer. Meteor. Soc.*, 212-216.

Kalnay, E., 2002: *Atmospheric modeling, data assimilation, and predictability*. Cambridge University Press, 341pp.

Miyoshi, T. and E. Kalnay, 2004: Ensemble forecasting using orthogonalized and stochastically seeded bred vectors in the Lorenz 96 model with orography. In preparation.

Miyoshi, T., 2004: Classical methods tour of advanced data assimilation using Lorenz 96 model. *A final report submitted to METO658E*, Department of Meteorology, University of Maryland, College Park.

Nicolis, C., 2003: Dynamics of Model Error: Some Generic Features. *J. Atmos. Sci.*, **60**, 2208-2218.

Ott, E., B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. J. Patil, and J. A. Yorke, 2002: Exploiting local low dimensionality of the atmospheric dynamics for efficient Kalman filtering. ArXiv:arc-ive/paper 0203058, http://arxiv.org/abs/physics/0203058.

Ott, E., B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. J. Patil, and J. A. Yorke, 2004: A Local Ensemble Kalman Filter for Atmospheric Data Assimilation. *Tellus*, in press.

Patil, D. J., B. R. Hunt, E. Kalnay, J. A. Yorke, and E. Ott, 2001: Local Low Dimensionality of Atmospheric Dynamics. *Phys. Rev. Lett.*, 5878-5881.

Parrish, D. F. and J. C. Derber, 1992: The National Meteorological Center's Spectral Statistical-Interpolation Analysis System. *Mon. Wea. Rev.*, **120**, 1747-1763.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble Square Root Filters. *Mon. Wea. Rev.*, **131**, 1485-1490.

van Leeuwen, P. J., 1999: Comment on "Data Assimilation Using an Ensemble Kalman Filter Technique". *Mon. Wea. Rev.*, **127**, 1374-1377.

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

Zupanski, D., 1997: A General Weak Constraint Applicable to Operational 4DVAR Data Assimilation Systems. *Mon. Wea. Rev.*, **125**, 2274-2292.