

The diffuse ensemble filter

X. Yang and T. DelSole

Center for Ocean-Land-Atmosphere Studies, 4041 Powder Mill Rd., Suite 302, Calverton, MD, 20705, USA
George Mason University, Fairfax, VA, USA

Received: 7 May 2008 – Revised: 18 June 2009 – Accepted: 18 June 2009 – Published: 16 July 2009

Abstract. A new class of ensemble filters, called the Diffuse Ensemble Filter (DEnF), is proposed in this paper. The DEnF assumes that the forecast errors orthogonal to the first guess ensemble are uncorrelated with the latter ensemble and have infinite variance. The assumption of infinite variance corresponds to the limit of “complete lack of knowledge” and differs dramatically from the implicit assumption made in most other ensemble filters, which is that the forecast errors orthogonal to the first guess ensemble have vanishing errors. The DEnF is independent of the detailed covariances assumed in the space orthogonal to the ensemble space, and reduces to conventional ensemble square root filters when the number of ensembles exceeds the model dimension. The DEnF is well defined only in data rich regimes and involves the inversion of relatively large matrices, although this barrier might be circumvented by variational methods. Two algorithms for solving the DEnF, namely the Diffuse Ensemble Kalman Filter (DEnKF) and the Diffuse Ensemble Transform Kalman Filter (DETKF), are proposed and found to give comparable results. These filters generally converge to the traditional EnKF and ETKF, respectively, when the ensemble size exceeds the model dimension. Numerical experiments demonstrate that the DEnF eliminates filter collapse, which occurs in ensemble Kalman filters for small ensemble sizes. Also, the use of the DEnF to initialize a conventional square root filter dramatically accelerates the spin-up time for convergence. However, in a perfect model scenario, the DEnF produces larger errors than ensemble square root filters that have covariance localization and inflation. For imperfect forecast models, the DEnF produces smaller errors than the ensemble square root filter with inflation. These experiments suggest that the DEnF has some advantages relative to the ensemble square root filters in the regime of small ensemble size, imperfect model, and copious observations.

1 Introduction

It is well established that forecast ensembles in ensemble-based Kalman filters tend to collapse – that is, the forecast spread tends to shrink with time until the filter effectively rejects the observations.¹ The collapse of the ensemble implies that the forecast errors are underestimated and that the filter weights the first guess too heavily. Eventually, the forecast becomes so “overconfident” that the filter ignores the observations altogether. Two methods for avoiding filter collapse are covariance inflation (Anderson and Anderson, 1999) and localization (Hamill et al., 2001; Houtekamer and Mitchell, 2001). Covariance inflation attempts to avoid filter collapse by inflating the covariance of the ensemble by an empirical factor. However, covariance inflation alone cannot prevent filter collapse if the ensemble size is sufficiently small, as we will show. This result may be understood as follows. The full state space can be split into two subspaces: the space spanned by the ensemble, which we call the ensemble space, and the complement to the ensemble space, which we call the null space. Generally for atmospheric applications, the ensemble size is much less than the model dimension, so that the ensemble does not span the full model space, and hence the null space is very large. In essence, the ensemble filters, e.g., the ensemble Kalman filter (EnKF) (Evensen, 1994) and ensemble square root filters (Tippett et al., 2003), updates only those variables in the ensemble space. It follows that variables in the null space are not updated, which is equivalent to assuming that the forecast covariance of the null space vectors vanishes. Thus, no matter how much inflation is applied,

¹ Some papers refer to this phenomenon as filter divergence. For instance, Maybeck (1979) defined the filter divergence as “the filter underestimates its own errors, and it will not “look hard enough” at the measurements”. Unfortunately, the term divergence also is used in the Kalman filter literature to refer to different things. For instance, the filter divergence refers to modeling errors by Anderson and Moore (1979), while it refers to computational errors by Haykin (2001). To avoid confusion, we use the term ensemble collapse.



Correspondence to: X. Yang
(xyang@cola.iges.org)

this inflation only influences the ensemble space, leaving the variances in the null space zero and hence underestimated.

The above reasoning highlights a very unrealistic property of ensemble filters: they effectively assume that forecast errors in the null space vanish. Consequently, observations have no impact on the null space, regardless of how much the ensemble is inflated. This deficiency of ensemble filters deserves emphasis: if the ensemble size is small but the observations are abundant, the observations nevertheless are not used to modify the ensemble outside the space spanned by the first guess, no matter how many observations are available that would justify such modifications. This deficiency follows directly from the assumption that the forecast is “perfect” in the null space, an assumption that is grossly incorrect for atmospheric and oceanic data assimilation, in which the underlying forecast model is imperfect. The question arises as to whether a Kalman filter can be formulated in such a way as to avoid the assumption of vanishing forecast errors in the null space. In an abstract sense, a similar situation occurs in the initialization of a Kalman filter – the forecast covariance matrix generally is not available at the first time step. To deal with incompletely specified initial conditions, Ansley and Kohn (1985) proposed a method that is equivalent to assuming a diffuse prior distribution for the unspecified part of the initial state. A distribution is said to be diffuse if its covariance matrix is arbitrarily large (de Jong, 1991). The diffuse assumption often corresponds to the limit of complete lack of knowledge in Bayesian analysis, from which the Kalman filter can be derived (Maybeck, 1979). Ansley and Kohn (1985) and de Jong (1991) discuss the extension of the Kalman filter to partially diffuse covariance matrices.

The purpose of this paper is to develop an extension of ensemble filters to allow for arbitrarily large forecast errors. Our fundamental assumption is that the forecast errors orthogonal to the ensemble are uncorrelated with the errors in the ensemble, and are infinitely large. We call the resulting filters *Diffuse Ensemble Filters* (DEnFs). We propose two specific algorithms called the *Diffuse Ensemble Kalman Filter* (DEnKF) and the *Diffuse Ensemble Transform Kalman Filter* (DETKF). Our derivation of the DEnFs is essentially independent of Ansley and Kohn (1985) and de Jong (1991), as it is tailored to the special needs of an ensemble Kalman filter. It should be recognized, however, that the derivation of a diffuse filter is subtle. For instance, the filtering and limiting operations are not interchangeable, as noted by Ansley and Kohn (1985). Also, early derivations of diffuse filters were numerically inefficient. In the derivation presented here, the proof is general, direct, and yields a closed form set of equations.

Another approach to avoiding filter collapse is covariance localization. Covariance localization attempts to reduce the spurious correlations that inevitably arise from sample based estimates by taking the Schur product between the sample based estimate and a distance-dependent function that varies from unity at the observation location to zero at some pre-

defined radial distance. In order to maintain the positive definiteness of covariance matrices, the distance-dependent function used in the Schur product must itself be positive definite. This procedure can be interpreted as imposing structure on the error covariance, in which case the ensemble effectively gives information about many more degrees of freedom than just the ensemble space. Accordingly, covariance localization changes the rank of the forecast covariance; in particular, it usually eliminates the null space (as we will show). Thus, there can be no diffuse ensemble filters with localization, because under localization there is no null space for applying the diffuse assumption. However, localization alone still allows underestimation of covariances and hence most applications of covariance localization also apply covariance inflation.

The paper is organized as follows. The algorithm of DEnFs is presented in Sect. 2, and the experimental setup is described in Sect 3. Data assimilation experiments with the Lorenz 96 model are used to compare the diffuse ensemble filters and the ensemble filters in Sect. 4. Initialization using DETKF is presented in Sect.5. The paper ends with the conclusions and discussions in Sect. 6.

2 Derivation of the Diffuse Ensemble Filters

In this section we review traditional ensemble filters, use a simple example to illustrate some differences between diffuse and traditional filters, and then derive the Diffuse Ensemble Kalman Filter (DEnKF) and the Diffuse Ensemble Transform Kalman Filter (DETKF). We end this section by discussing additional generalizations of the diffuse filter.

2.1 The Ensemble Transform Kalman Filter (ETKF)

The Ensemble Transform Kalman Filter (ETKF) was proposed by Bishop et al. (2001) and clarified by Tippett et al. (2003). We briefly review this filter to establish notation and provide a reference for comparison. The standard Kalman Filter equations for the mean update and the analysis covariance matrix are (Maybeck, 1979, p117)

$$\bar{\Psi}^a = \bar{\Psi}^f + \mathbf{P}\mathbf{H}^T (\mathbf{R} + \mathbf{H}\mathbf{P}\mathbf{H}^T)^{-1} (\mathbf{o} - \mathbf{H}\bar{\Psi}^f) \quad (1)$$

$$\mathbf{P}^a = \mathbf{P} - \mathbf{P}\mathbf{H}^T (\mathbf{R} + \mathbf{H}\mathbf{P}\mathbf{H}^T)^{-1} \mathbf{H}\mathbf{P}, \quad (2)$$

where $\bar{\Psi}$ is the mean state vector, \mathbf{R} is the observation error covariance matrix, \mathbf{H} is the observation operator, \mathbf{P} is the forecast covariance matrix, and \mathbf{o} is the observation vector. Let the difference between the j -th ensemble member and the ensemble mean be denoted by the M -dimensional vector \mathbf{a}_j . For ensemble size N , let

$$\mathbf{A} = \frac{1}{\sqrt{N-1}} [\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_N]. \quad (3)$$

Then an unbiased estimate of the forecast covariance matrix is

$$\mathbf{P}_E = \mathbf{A}\mathbf{A}^T. \tag{4}$$

The ensemble Kalman Filter is obtained by substituting the sample covariance matrix \mathbf{P}_E for \mathbf{P} in (1) and (2). By invoking the Sherman-Morrison-Woodbury formula, it is straightforward to show that the resulting analysis covariance matrix can be written as

$$\mathbf{P}^a = \mathbf{A} \left(\mathbf{I} + \mathbf{A}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{A} \right)^{-1} \mathbf{A}^T. \tag{5}$$

An analysis ensemble matrix \mathbf{A}^a such that $\mathbf{P}^a = \mathbf{A}^a (\mathbf{A}^a)^T$ is derived by setting

$$\mathbf{A}^a = \mathbf{A} \left(\mathbf{I} + \mathbf{A}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{A} \right)^{-1/2}, \tag{6}$$

where the matrix in parentheses is a square root matrix. The square root matrix can be derived by computing the eigenvector decomposition

$$\left(\mathbf{I} + \mathbf{A}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{A} \right) = \mathbf{Y} \mathbf{D} \mathbf{Y}^T, \tag{7}$$

where \mathbf{Y} is unitary and \mathbf{D} is a diagonal element with positive diagonal elements, and then setting

$$\left(\mathbf{I} + \mathbf{A}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{A} \right)^{-1/2} = \mathbf{Y} \mathbf{D}^{-1/2} \mathbf{Y}^T. \tag{8}$$

As noted by Sakov and Oke (2008), the symmetric form of the square root defined in (8) preserves the ensemble mean.

We draw attention to the following fact. It is evident that the mean update is pre-multiplied by \mathbf{A} , and that the covariance update is pre- and post-multiplied by \mathbf{A} and \mathbf{A}^T , respectively. It follows that the mean and covariance updates occur only in the subspace spanned by the first guess ensemble. Therefore, the ensemble Kalman Filter does not modify any variable in the space orthogonal to the ensemble. This result is tantamount to assuming that the forecast covariance matrix vanishes in the null space, which of course is highly unrealistic, and the filter is overconfident in the null space. As we will see, this characteristic of the ensemble square root filter (ESRF) distinguishes it from the diffuse filter.

2.2 A simple example

In this section, we present a simple 2-dimensional example to illustrate some key properties of various filters. Without loss of generality we use a basis set in which the forecast covariance matrix is diagonal:

$$\mathbf{P} = \begin{pmatrix} p_E & 0 \\ 0 & p_N \end{pmatrix}. \tag{9}$$

Shortly, we will interpret p_E as the variance in ensemble space and p_N as the variance in the null space. Consider

the situation in which only two observations are available. Although general observation networks can be considered, this extra generality does not lead to substantial insights in this 2-D problem. Accordingly, we make the simplifying assumptions that \mathbf{H} and \mathbf{R} are diagonal:

$$\mathbf{H} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \mathbf{R} = \begin{pmatrix} r_E & 0 \\ 0 & r_N \end{pmatrix}. \tag{10}$$

The mean analysis under these assumptions is

$$\bar{\Psi}^a = \bar{\Psi}^f + \mathbf{P} \mathbf{H}^T \left(\mathbf{R} + \mathbf{H} \mathbf{P} \mathbf{H}^T \right)^{-1} \left(\mathbf{o} - \mathbf{H} \bar{\Psi}^f \right) \tag{11}$$

$$= \begin{pmatrix} \bar{\Psi}_E^f \\ \bar{\Psi}_N^f \end{pmatrix} + \begin{pmatrix} \frac{p_E}{p_E+r_E} & 0 \\ 0 & \frac{p_N}{p_N+r_N} \end{pmatrix} \begin{pmatrix} o_E - \bar{\Psi}_E^f \\ o_N - \bar{\Psi}_N^f \end{pmatrix} \tag{12}$$

$$= \begin{pmatrix} \bar{\Psi}_E^f + \frac{p_E}{p_E+r_E} \left(o_E - \bar{\Psi}_E^f \right) \\ \bar{\Psi}_N^f + \frac{p_N}{p_N+r_N} \left(o_N - \bar{\Psi}_N^f \right) \end{pmatrix}, \tag{13}$$

where the mean forecast is denoted $\bar{\Psi}^f = (\bar{\Psi}_E^f \ \bar{\Psi}_N^f)^T$. Similarly, the covariance matrix update is

$$\mathbf{P}^a = \mathbf{P} - \mathbf{P} \mathbf{H}^T \left(\mathbf{H}^T \mathbf{P} \mathbf{H} + \mathbf{R} \right)^{-1} \mathbf{H} \mathbf{P} \tag{14}$$

$$= \begin{pmatrix} p_E - \frac{p_E^2}{r_E+p_E} & 0 \\ 0 & p_N - \frac{p_N^2}{r_N+p_N} \end{pmatrix}. \tag{15}$$

Let us first consider the Kalman Filter solution for an ensemble size of two. In this case, the forecast covariance matrix is rank-1. If p_E is identified as the variance of the ensemble, then $p_N=0$. The mean update in this case is

$$\bar{\Psi}^a = \begin{pmatrix} \bar{\Psi}_E^f + \frac{p_E}{p_E+r_E} \left(o_E - \bar{\Psi}_E^f \right) \\ \bar{\Psi}_N^f \end{pmatrix}, \tag{16}$$

while the covariance update is

$$\mathbf{P}^a = \begin{pmatrix} p_E - \frac{p_E^2}{r_E+p_E} & 0 \\ 0 & 0 \end{pmatrix} \tag{17}$$

This solution reveals two key characteristics of the ensemble based Kalman Filter: the analysis increment (i.e., $\bar{\Psi}^a - \bar{\Psi}^f$) is confined to the ensemble space, and the covariance matrix update (i.e., $\mathbf{P}^a - \mathbf{P}$) is confined to the ensemble space. This means that the forecast in the null space is not modified; that is, $\bar{\Psi}_N^a = \bar{\Psi}_N^f$. The limit $p_N \rightarrow 0$ implies that the forecast in the null space has zero uncertainty, or equivalently that the forecast is “perfect.” This assumption is obviously unrealistic in genuine data assimilation problems in which nature is unknown.

Let us now consider the diffuse limit, which corresponds to the limit $p_N \rightarrow \infty$. This limit is easily evaluated as

$$\bar{\Psi}^a = \begin{pmatrix} \bar{\Psi}_E^f + \frac{p_E}{p_E+r_E} \left(o_E - \bar{\Psi}_E^f \right) \\ o_N \end{pmatrix}. \tag{18}$$

and

$$\mathbf{P}^a = \begin{pmatrix} \frac{1}{\frac{1}{r_E} + \frac{1}{p_E}} & 0 \\ 0 & r_N \end{pmatrix}. \quad (19)$$

The solution shows that the update in ensemble space is exactly the standard KF solution, while the update in the null space is replaced by the appropriate observation. This result is sensible, since the diffuse limit implies that the forecast is completely uncertain and so the analysis should reduce to the observation. In contrast to the ensemble based Kalman Filter, the update occurs in both the ensemble space and the null space.

2.3 The Diffuse Ensemble Filter

The basic assumption in the DEnFs is that the forecast errors orthogonal to the first guess ensemble are uncorrelated with the ensemble and have infinite covariance matrix. With this assumption, we will derive the algorithm to update the ensemble using the Kalman Filter. Let the SVD of the $M \times N$ matrix \mathbf{A} be

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T, \quad (20)$$

where \mathbf{S} is an $M \times N$ diagonal matrix, whose diagonal elements specify the non-negative singular values, ordered from the largest to smallest, and \mathbf{U} and \mathbf{V} are unitary (but having respective dimensions $M \times M$ and $N \times N$). At most, $N - 1$ diagonal elements of $\mathbf{S}\mathbf{S}^T$ are nonzero, since the ensemble mean has been subtracted from each member. Assume that exactly $N - 1$ singular values are nonzero. Furthermore, let the singular vectors be ordered such that the first $N - 1$ vectors are those with non-zero singular values. This ordering allows us to partition the singular vector matrix \mathbf{U} as

$$\mathbf{U} = [\mathbf{U}_E \ \mathbf{U}_N], \quad (21)$$

where \mathbf{U}_E denotes the $M \times (N - 1)$ matrix whose $N - 1$ column vectors are the singular vectors associated with non-zero singular values, and \mathbf{U}_N denotes the matrix containing the remaining singular vectors that span the null space. The forecast ensemble covariance matrix can then be written as

$$\mathbf{P}_E = \mathbf{U}_E \mathbf{S} \mathbf{S}^T \mathbf{U}_E^T = \mathbf{U}_E \mathbf{S}_E^2 \mathbf{U}_E^T \quad (22)$$

where \mathbf{S}_E is an $N - 1$ dimensional, square, diagonal matrix whose diagonal elements equal the non-zero singular values of \mathbf{A} .

To derive the diffuse ensemble filter, we start with the ‘‘inverse’’ form of the Kalman filter equations (Maybeck, 1979, Sect. 5.7), also known as the information filter, which are

$$\bar{\Psi}^a = \bar{\Psi}^f + \left(\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} + \mathbf{P}^{-1} \right)^{-1} \mathbf{H}^T \mathbf{R}^{-1} \left(\mathbf{o} - \mathbf{H} \bar{\Psi}^f \right) \quad (23)$$

$$\mathbf{P}^a = \left(\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} + \mathbf{P}^{-1} \right)^{-1}. \quad (24)$$

Since \mathbf{P}_E is not invertible, we cannot simply substitute $\mathbf{P} = \mathbf{P}_E$ in these equations as we did for the standard form of the Kalman filter equations. Accordingly, we invoke a fictitious ensemble whose covariance matrix is \mathbf{P}_N such that total forecast covariance

$$\mathbf{P} = \mathbf{P}_E + \mathbf{P}_N \quad (25)$$

is nonsingular. The first assumption of the diffuse filter is that \mathbf{P}_E and \mathbf{P}_N are orthogonal; i.e., $\mathbf{P}_E \mathbf{P}_N = \mathbf{P}_N \mathbf{P}_E = \mathbf{0}$. This implies that \mathbf{P}_N is of the form

$$\mathbf{P}_N = \mathbf{U}_N \Sigma \mathbf{U}_N^T \quad (26)$$

where Σ is a nonsingular matrix specifying the covariance matrix in the null space. Under this assumption the inverse forecast covariance matrix becomes

$$\mathbf{P}^{-1} = \mathbf{U} \begin{pmatrix} \mathbf{S}_E^{-2} & 0 \\ 0 & \Sigma^{-1} \end{pmatrix} \mathbf{U}^T. \quad (27)$$

The second assumption of the DEnFs is that $\Sigma^{-1} \rightarrow 0$. One way to interpret this limit is to define $\mathbf{P}_N = \alpha \mathbf{U}_N \Sigma' \mathbf{U}_N^T$, where Σ' is a constant, nonsingular matrix, and then take the limit $\alpha \rightarrow \infty$. In this case, $\Sigma^{-1} \rightarrow 0$ regardless of the detailed structure of Σ' ; that is, the limit is independent of the details of the forecast covariance in the null space. The diffuse limit is therefore

$$\mathbf{P}_{\text{dif}}^{-1} = \mathbf{U} \begin{pmatrix} \mathbf{S}_E^{-2} & 0 \\ 0 & 0 \end{pmatrix} \mathbf{U}^T = \mathbf{U}_E \mathbf{S}_E^{-2} \mathbf{U}_E^T. \quad (28)$$

The substitution $\mathbf{P}^{-1} \rightarrow \mathbf{P}_{\text{dif}}^{-1}$ in (23) and (24) may present problems because the matrix $\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} + \mathbf{P}^{-1}$ may be singular and therefore has no inverse. We show in the appendix that a necessary and sufficient condition for \mathbf{P}^a to be nonsingular is that the auxiliary matrix

$$\mathbf{W} = \mathbf{U}_N^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{U}_N \quad (29)$$

should be nonsingular. The restriction that \mathbf{W} be invertible can be interpreted as requiring that the observations project onto every degree of freedom in the null space. Loosely speaking, if \mathbf{W} is singular, then there exists a vector in the null subspace that is unobserved. This restriction is sensible in light of the fact that the null space has no model information under the diffuse assumption, so the only other information available for updating the null space must come from observations. Since \mathbf{P}^a is nonsingular in this case, it is full rank, indicating that the mean and covariance updates are not confined to the ensemble subspace. This represents a fundamental difference with other ensemble Kalman Filters.

To summarize, the mean update equation for the DEnF is

$$\bar{\Psi}_{\text{dif}}^a = \bar{\Psi}^f + \left(\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} + \mathbf{P}_{\text{dif}}^{-1} \right)^{-1} \mathbf{H}^T \mathbf{R}^{-1} \left(\mathbf{o} - \mathbf{H} \bar{\Psi}^f \right), \quad (30)$$

and the covariance update, derived by substituting (28) into (24), is

$$\mathbf{P}^a = \left(\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} + \mathbf{U}_E \mathbf{S}_E^{-2} \mathbf{U}_E^T \right)^{-1}. \quad (31)$$

The fact that \mathbf{P}^a is full rank when \mathbf{W} is full rank raises the question as to how to define an analysis ensemble. This question does not arise in traditional EnKFs because the analysis and forecast span exactly the same space and hence can be represented by the same number of basis vectors. In contrast, the DEnF may start with a small ensemble but leads to a full rank analysis covariance matrix that cannot be represented by an ensemble size smaller than or equal to the model dimension. Of the many approaches to deriving an ensemble filter that can be conceived, we present two: one based on perturbed observations, and one based on projecting the analysis into the ensemble space. At the end of this section we discuss alternative solution methods, including a method that relaxes the requirement that \mathbf{W} be nonsingular.

2.3.1 The Diffuse Ensemble Kalman Filter (DEnKF)

Houtekamer and Mitchell (1998) and Burgers et al. (1998) proposed what is now called the Ensemble Kalman Filter (EnKF), which is characterized by randomly perturbed observations. By analogy, we propose the Diffuse Ensemble Kalman Filter (DEnKF), in which the ensemble update for the i -th ensemble member is defined as

$$\Psi_i^a = \Psi_i^f + \left(\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} + \mathbf{U}_E \mathbf{S}_E^{-2} \mathbf{U}_E^T \right)^{-1} \mathbf{H}^T \mathbf{R}^{-1} \left(\mathbf{o}_i - \mathbf{H} \Psi_i^f \right), \quad (32)$$

where $i = 1, \dots, N$, $\mathbf{o}_i = \mathbf{o} + \mathbf{r}_i$, $\mathbf{r}_i \sim \mathbf{N}(\mathbf{0}, \mathbf{R})$, and $\mathbf{N}(\mu, \sigma^2)$ denotes a Gaussian distribution of mean μ and variance σ^2 . If the forecast covariance matrix based on the ensemble is full rank, \mathbf{U}_N equals $\mathbf{0}$, and the DEnKF reduces to the EnKF. Note that the analysis increment $\Psi_i^f - \Psi_a^f$ of the DEnKF is not restricted to the ensemble space, in contrast to the EnKF.

2.3.2 The Diffuse Ensemble Transform Kalman Filter (DETKF)

A deterministic diffuse filter can be derived by analogy with the ETKF (see Sect. 2.1). In this case, the mean update is given by the same equation as in the ETKF, namely (30). However, instead of using the full analysis covariance (31), we project \mathbf{P}^a onto the ensemble space. This projection implies that the ensemble is updated only in the space spanned by the first guess ensemble, just as in the ETKF. We show in the appendix that the final analysis update equation for the DETKF is

$$\mathbf{P}_{\text{dif}}^a = \mathbf{A} \left[\mathbf{I} + \mathbf{A}^T \mathbf{H}^T \left(\mathbf{R}^{-1} - \mathbf{R}^{-1} \mathbf{H} \mathbf{U}_N \mathbf{W}^{-1} \mathbf{U}_N^T \mathbf{H}^T \mathbf{R}^{-1} \right) \mathbf{H} \mathbf{A} \right]^{-1} \mathbf{A}^T. \quad (33)$$

Comparison of this equation with (5) reveals that the DETKF differs from the ETKF by an extra term in the matrix whose

inverse is taken. Furthermore, this extra term has the effect of *inflating* the analysis ensemble (i.e., $\mathbf{P}_{\text{dif}}^a - \mathbf{P}^a$ is positive semi-definite). This inflation reflects the fact that the DETKF accounts for uncertainty in the null space, whereas the ETKF effectively assumes the forecast in the null space is perfect. The DETKF and ETKF become identical if

$$\mathbf{U}_E^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{U}_N = 0, \quad (34)$$

because in this case the “extra” term in (33) vanishes. It is sensible that the DETKF and ETKF have the same ensemble spread when (34) is satisfied, because the observations in the ensemble space and null space are uncorrelated, in which case observations in the null space provide no information for updating the ensemble space.

The square root form of the DETKF is obtained by solving the eigenvalue decomposition

$$\mathbf{I} + \mathbf{A}^T \mathbf{H}^T \left(\mathbf{R}^{-1} - \mathbf{R}^{-1} \mathbf{H} \mathbf{U}_N \mathbf{W}^{-1} \mathbf{U}_N^T \mathbf{H}^T \mathbf{R}^{-1} \right) \mathbf{H} \mathbf{A} = \mathbf{Y} \mathbf{D} \mathbf{Y}^T, \quad (35)$$

where \mathbf{Y} is unitary and \mathbf{D} is a diagonal matrix with positive diagonal elements, and then defining

$$\mathbf{A}_{\text{dif}}^a = \mathbf{A} \mathbf{Y} \mathbf{D}^{-\frac{1}{2}} \mathbf{Y}^T, \quad (36)$$

which gives $\mathbf{P}_{\text{dif}}^a = \mathbf{A}_{\text{dif}}^a \left(\mathbf{A}_{\text{dif}}^a \right)^T$. If ensemble covariance is full rank, \mathbf{U}_N equals $\mathbf{0}$, and the DETKF reduces to the Ensemble Transform Kalman Filter (ETKF). Thus, the DETKF does not converge to the ensemble square root filter (ESRF) of Whitaker and Hamill (2002) as ensemble covariance goes to full rank, since the latter filter differs from the ETKF.

2.4 Alternative diffuse filters

We emphasize that the DEnKF and DETKF require inverting matrices of the order of the model dimension. For atmospheric and oceanic models, this dimension can easily exceed 100 000, which is clearly impractical. However, the DEnKF might be solvable using an equivalent variational method, just as large scale data assimilation problems are solved using variational methods at operational centers (Klinker et al., 2000). As is well known (Maybeck, 1979, p. 234), the mean update of the Kalman Filter equations minimizes the cost function

$$L = \left(\mathbf{o} - \mathbf{H} \bar{\Psi} \right)^T \mathbf{R}^{-1} \left(\mathbf{o} - \mathbf{H} \bar{\Psi} \right) + \left(\bar{\Psi} - \bar{\Psi}^f \right)^T \mathbf{P}^{-1} \left(\bar{\Psi} - \bar{\Psi}^f \right). \quad (37)$$

The first term can be interpreted as a “goodness of fit”, since it measures how close the state is to the observations, while the second term is a penalty function, since it increases with the distance between the state and first guess. Under the diffuse assumption, this cost function becomes

$$L = \left(\mathbf{o} - \mathbf{H} \bar{\Psi} \right)^T \mathbf{R}^{-1} \left(\mathbf{o} - \mathbf{H} \bar{\Psi} \right) + \left(\mathbf{U}_E^T \left(\bar{\Psi} - \bar{\Psi}^f \right) \right)^T \mathbf{S}_E^{-2} \left(\mathbf{U}_E^T \left(\bar{\Psi} - \bar{\Psi}^f \right) \right). \quad (38)$$

The latter cost function differs from the former in that the penalty function is evaluated only in the ensemble space. The

advantage of minimizing this cost function is that it can be solved with standard conjugate gradient methods without explicitly inverting the matrix \mathbf{W} . Unfortunately, the resulting solution gives only the mean update; how one can use (37) and (38) to generate an ensemble filter is unclear.

Another question is whether the restriction that \mathbf{W} be non-singular can be relaxed. One theoretical barrier to defining a diffuse limit when \mathbf{W} is singular is that it leads to a contradictory situation. Specifically, singular \mathbf{W} implies that neither the forecast ensemble nor the observations constrain a certain space. Indeed, it is possible to show that L is independent of the null vectors of \mathbf{W} , indicating that L does not constrain these vectors. Now, if neither the forecast nor the observations constrains part of the null subspace, then on what basis can one update this space? The solution to this problem is to apply the diffuse assumption only to the part of the null space that is constrained by observations. This can be accomplished by splitting the null space itself into two parts, one constrained by observations (identified by the range of \mathbf{W}), and one unconstrained by observations (identified by the null space of \mathbf{W}). Then, the diffuse assumption can be applied to the subspace that is constrained by observations, while the ‘‘perfect model’’ assumption can be applied to the subspace that is unconstrained by observations. This alternative diffuse filter will not be discussed further in this paper.

3 Experimental setup

The model used here is the Lorenz-96 model (Lorenz and Emanuel, 1998), which is governed by the equation

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + f_0, \quad (39)$$

where $i = 1 \dots J$ with cyclic indices. Here, J is 40 and f_0 is 8.0. The consecutive model states are obtained by integrating the model forward with the time interval 0.05, and a fourth-order Runge-Kutta numerical method is applied at each model time step. The truth is one single integration of the model. The observational data set was constructed by adding Gaussian white noise with zero mean and unit variance to the truth at each of the 40 grid points, thereby producing 40 observations at each time step.

In realistic data assimilation, the model is imperfect due to model errors, e.g., uncertain model parameters. In this study, we will conduct some data assimilation experiments with an imperfect model, defined as

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - \frac{x_i}{1.0 + d_i} + f_0 + f_i, \quad (40)$$

where the dissipation parameters d_i and forcing parameters f_i are randomly specified according to

$$\mathbf{f}_i \approx N(0, 4), i = 1 \dots J, \quad (41)$$

$$\mathbf{d}_i \approx N(0.5, 0.5), i = 1 \dots J. \quad (42)$$

The ensemble filters used here are the EnKF of Evensen (1994) and the ESRF of Whitaker and Hamill (2002). The initial ensemble members for the first data assimilation experiment are generated by adding independent, zero mean, normally distributed random numbers of variance 1.0 to the climatology of the long run with 30 000 time steps. The covariance inflation for all experiments in this study, when applied, is the adaptive covariance inflation algorithm proposed by Anderson (2007) or constant inflation (Anderson and Anderson, 1999). The localization applied here is the fifth order polynomial function of Gaspari and Cohn (1999) with half-width c . Localization half width c is 10 relative to the model domain size 40. If the distance between the observation and the state variable is greater than $2c$, then the localization function is zero, which implies that the observation has no impact on the state variable; otherwise, it approximates a Gaussian. The root mean square error (RMSE) is computed as the root mean square of the difference between the analysis and the truth over the 40 grid points and from model time steps 3000 to 6000.

To test the consistency between observations and filter output, we use the fact, as noted by (Maybeck, 1979, p229), that the Kalman filter predicts that the innovation vector

$$\mathbf{z} = \mathbf{o} - \mathbf{H}\bar{\Psi}^f \quad (43)$$

is a white Gaussian sequence with zero mean and covariance matrix

$$\mathbf{C} = \mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R}. \quad (44)$$

This fact allows us to construct an *innovation consistency function* (ICF). Specifically, if this assumption is correct, then the quadratic form

$$\text{ICF} = \mathbf{z}^T \mathbf{C}^{-1} \mathbf{z} \quad (45)$$

should have a chi-square distribution with degrees of freedom equal to the rank of \mathbf{C}^{-1} (Johnson and Wichern, 2002, Result 4.7). The above quadratic form is essentially the log-likelihood function, aside from irrelevant constant and multiplicative terms (Maybeck, 1979, p234). The 2.5% and 97.5% thresholds for a chi-squared distribution with 40 degrees of freedom are 24.4 and 59.3, respectively. Accordingly, the innovation vector is deemed inconsistent with the filter if ICF falls outside the interval (24.4, 59.3) more than 5% of the time. In the case of the DESRF, the evaluation of ICF is not straightforward since \mathbf{C} becomes unbounded. The evaluation of ICF for the DESRF is discussed in the appendix and shown to have a chi-squared distribution with 9 degrees of freedom (i.e., $40 - (30 + 1) = 9$) for ensemble size 10. The 2.5% and 97.5% thresholds for a chi-squared distribution with 9 degrees of freedom are 2.7 and 19, respectively. If the innovation vector falls outside this interval more the 5% of the time, then we conclude that the innovations are inconsistent with the filter.

4 Numerical results

Figure 1a–d shows a typical result for the truth, observation, forecast, and analysis by the ensemble square root filter at one grid point in the Lorenz-96 model. Note that the blue and green curves are superposed and undistinguishable. The innovation consistency function (ICF) is shown in Fig. 1e–h (for a longer time period). Note that the two ICF thresholds in panels e, f, and g are undistinguishable since ICFs are much larger than the two thresholds. Inspection of Fig. 1e–h shows that the innovations are consistent with the filter only if both covariance inflation and localization are applied (i.e., the ICF lies between the two dashed lines only in Fig. 1h). In other cases, the innovations are inconsistent with the filter. More importantly, the ensemble collapses in the cases illustrated in Fig. 1a–c – the analysis is weighted too heavily toward the model forecast, allowing the analysis to diverge from the observations. Interestingly, the ensemble square root filter with just localization still diverges (Fig. 1c and g) even though there is no null space. This may be due to the model non-linearity and underestimation of covariances by the sample ensemble.

The results for the DETKF are shown in Fig. 2a and c. The figures show that the amplitudes of the innovation vectors produced by the DETKF are too large relative to that assumed internally by the filter. However, in this case, there is no ensemble collapse. Instead, the analysis is weighted too heavily to the observations. Consequently, the analysis reveals much more high frequency noise than the truth, owing to the white noise in the observations. Just as with the ensemble filters, the DETKF might be improved with covariance inflation. Accordingly, we apply covariance inflation to the forecast ensemble (we do not inflate the null space covariances, since they are already inflated by the diffuse limit assumption). The ICF when covariance inflation is applied to the DETKF is shown in Fig. 2d, which reveals that inflation does indeed improve the consistency. It turns out that inflation also improves the RMSE of the analysis (not shown).

In order to avoid ensemble collapse due to the finite ensemble size and model non-linearities, two common methods, covariance inflation (Anderson and Anderson, 1999) and localization (Hamill et al., 2001; Houtekamer and Mitchell, 2001), are usually applied. The diffuse limit can be interpreted as an extreme example of inflation for the null space. Yet, even with infinite covariances in the null space, the diffuse filter still diverged. Similarly, in the ESRF with localization, there is no null space, yet the filter still diverges. Thus, an interesting conclusion from the above results is that the filter converges only when the covariance of both the ensemble space and the null space are inflated – inflating just one subspace is not enough to avoid filter collapse.

Covariance localization can not be implemented in the diffuse ensemble filters because it usually eliminates the null space by rendering the forecast covariance matrix full rank. Figure 3 shows the minimum spectrum of eigenvalues of the

forecast covariance matrix for 10 ensemble members with and without covariance localization. Without localization, the covariance matrix has 9 nonzero eigenvalues and 31 zero eigenvalues, which corresponds to the size of the ensemble space and null space respectively. All eigenvalues are nonzero when the covariance localization is applied, which implies that the localized covariance matrix is full rank and hence the null space is zero. The eigenvalue spectrum slope is deeper when the localization half width is larger. Note that covariance localization also intends to reduce sampling errors.

To investigate the sensitivity of the results to ensemble size, we show in Fig. 4a the performance of the ESRF and the DETKF, with inflation, as a function of ensemble size. For the ensemble size 41, there is no null space, so the DETKF is identical to the ETKF, and the values of RMSE for the two filters are almost the same (the small difference arises from the fact that the ESRF of Whitaker and Hamill (2002) differs from the ETKF). We see that the RMSE for the ESRF decreases dramatically and eventually the filter converges after 15 ensemble members. This implies that inflation alone can allow the filter to converge if the ensemble size is sufficiently large. Equivalently, if the ensemble size is too small, then inflation alone is not enough to prevent filter collapse. Thus, for small ensemble sizes relative to the model dimension, the DETKF may be an attractive alternative to the ETKF.

One can argue that the above test is not completely fair because the dynamical model is perfect in the sense that it is identical to the model that generates the truth. Consequently, the first guess of the dynamical model is very good, and therefore a filter that reduces to the first guess in the null space may perform preferentially better than a filter that does not. Accordingly, we consider a new test by using the imperfect model (40) to generate forecasts, but use the same set of observations generated by the original model (39). Note that the adaptive covariance inflation tends to be larger in the imperfect model case to account for model errors (Anderson, 2007). The resulting average RMSE as a function of ensemble size is shown in Fig. 4b. Compared to the perfect model scenario, the performance of the ESRF is dramatically degraded, especially for small ensemble sizes, while the performance of DETKF does not change much. This implies that DETKF outperforms the ESRF without localization for the imperfect model scenario.

Figure 5a shows the RMSE of the DEnKF and the EnKF with inflation as a function of ensemble size. For the ensemble size 41, there is no null space, so the DEnKF is identical to the EnKF. The RMSE for the EnKF decreases dramatically and eventually the filter converges after 20 ensemble members. When the ensemble size is smaller than 16, DEnKF performs better than EnKF. This implies that the diffuse EnKF outperforms EnKF in the regime of small ensemble sizes. The RMSE of EnKF is larger than that of ESRF (Figs. 4a and 5a), and the RMSE of DEnKF is also larger than that of DETKF (Fig. 5b). This indicates that sampling errors from perturbed

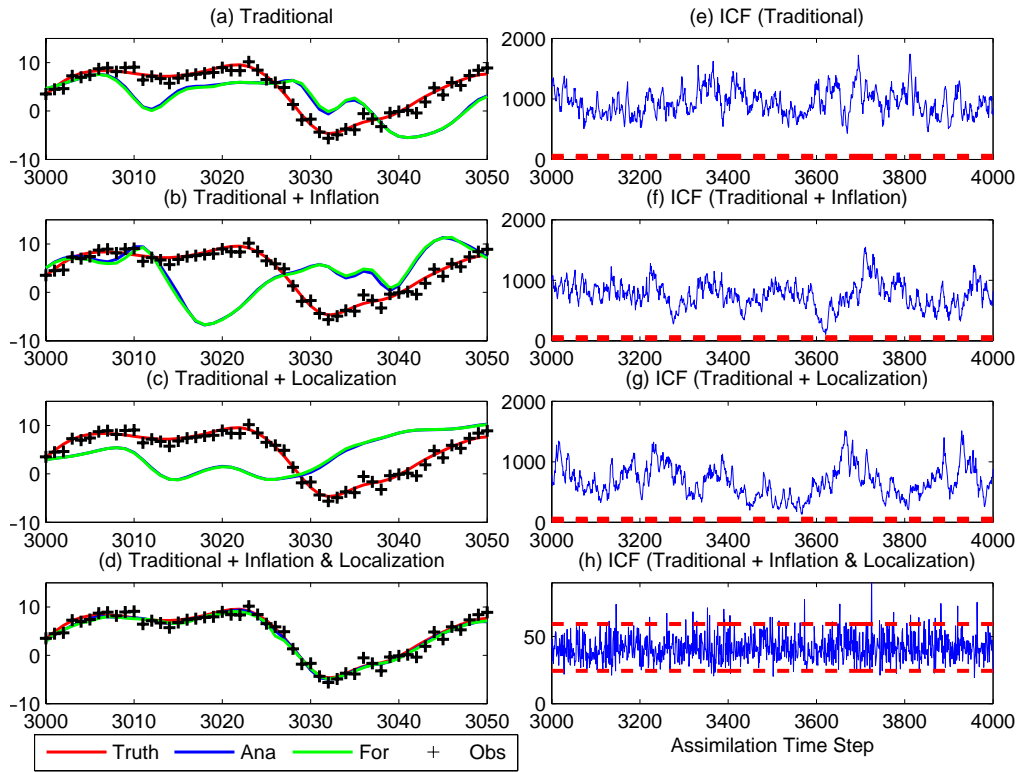


Fig. 1. Time series based on the Lorenz 96 model of the truth (red), the model forecast (green), the analysis (blue) and the observation (plus) at one grid point for (a) ESRF without inflation and localization, b) ESRF with inflation only, c) ESRF with localization only, and d) ESRF with localization and inflation. Time series of the innovation consistency function (ICF) for e) ESRF without inflation and localization, f) ESRF with inflation only, g) ESRF with localization only, h) ESRF with localization and inflation. Ensemble size is 10 for all experiments. Localization half width c is 10 relative to the model domain size 40. Red dashed line indicating the threshold value of ICF.

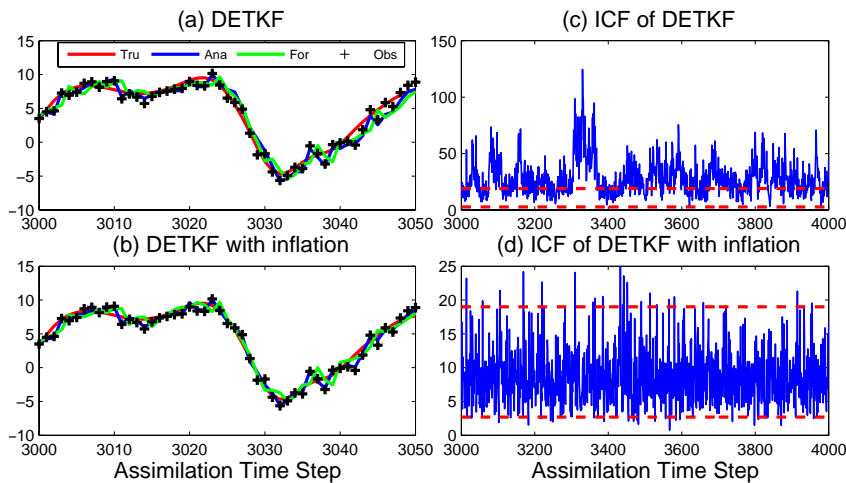


Fig. 2. Time series based on Lorenz 96 model of the truth (red), the model forecast (green), the analysis (blue) and the observation (plus) at one grid point for (a) DETKF without inflation, (b) DETKF with inflation. Time series of the innovation consistency function (ICF) for (c) DETKF without inflation, (d) DETKF with inflation. Ensemble size is 10 for all experiments. Red dashed line indicating the threshold value of ICF.

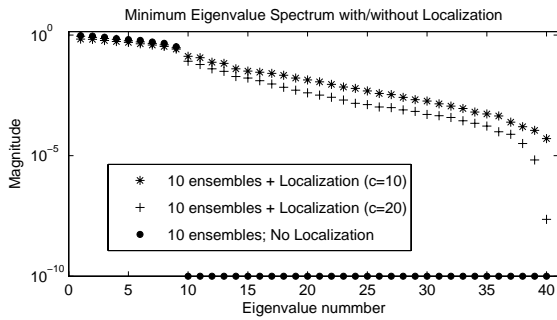


Fig. 3. Minimum of the ordered eigenvalues of the forecast covariance matrix for 10 ensemble members with and without covariance localization. The minimum is obtained from assimilation time steps 3000 to 6000, and localization was applied for $c=10$ and $c=20$, as indicated in the figure. Note that all 31 zero eigenvalues for 10 ensemble members without localization are set to 10^{-10} for plotting purpose.

observations in both EnKF and DEnKF degrade the performance of filters. This is the reason that in this study we focus on the performance of DETKF, rather than DEnKF.

5 Initialization using DESRF

Originally, the diffuse Kalman filter was designed to initialize the Kalman filter (de Jong, 1991; Koopman, 1997). Analogously, DETKF can be applied to initialize the ESRF. Here, we first run the DETKF for one time step to get the analyzed ensemble mean and perturbations, and then these optimal ensemble members are used to initialize the ESRF. Note that in this section the root mean square error (RMSE) is defined as the root mean square of the difference between the analysis and the truth over the 40 grid points. Figure 6a shows the RMSE as a function of assimilation time for the ESRF with and without using DETKF initialization with 20 ensemble members. The ESRF with standard initial ensembles of random Gaussian noise perturbations converges slowly to the optimal level of RMSE at around 500 assimilation time steps, while the ESRF, initialized with DETKF, converges rather quickly to the optimal level of RMSE at round 50 assimilation time step. After 500 assimilation time step, the RMSEs of these two different ensemble initializations are indistinguishable. The same experiment with 10 ensemble members plus localization reveals the similar results (Fig. 6b). This implies that initialization using DETKF accelerates the initial spin-up time for the ESRF.

6 Summary and discussion

This paper proposed a new type of filter called the Diffuse Ensemble Filter (DEnF). The DEnF assumes that the forecast errors in the space orthogonal to the first guess ensemble are

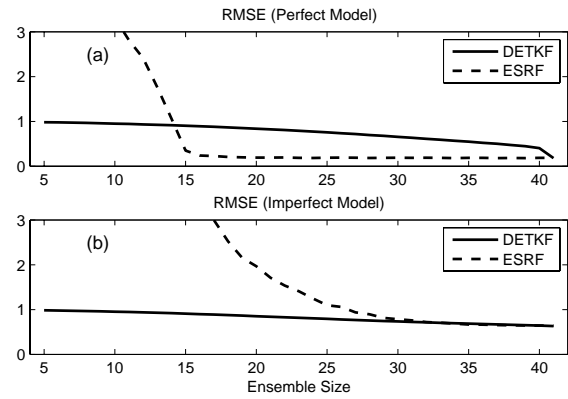


Fig. 4. The root mean square error (RMSE) as a function of ensemble size for the ESRF with inflation (dashed) and the DETKF with inflation (solid) using the (a) perfect and (b) imperfect models. Results are averaged over the 3000 to 6000 assimilation time step.

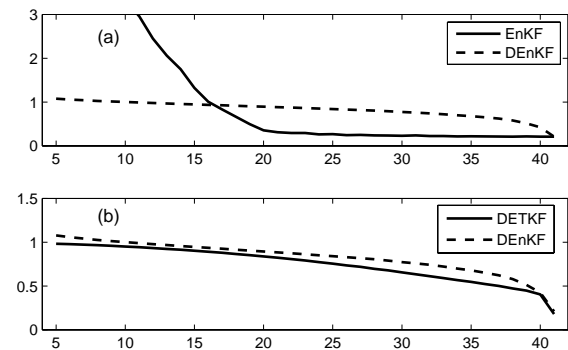


Fig. 5. The root mean square error (RMSE) as a function of ensemble size for (a) the EnKF with inflation (solid) and the DEnKF with inflation (dashed), (b) the EnKF with inflation (dashed) and the DETKF with inflation (solid) using the perfect model. Results are averaged over the 3000 to 6000 assimilation time step.

uncorrelated with the latter ensemble, and are infinite, corresponding to complete lack of information. Thus, in terms of the forecast covariance matrix in the null space \mathbf{P}_N , ensemble filters assume $\mathbf{P}_N \rightarrow 0$, while diffuse filters assume $\mathbf{P}_N \rightarrow \infty$. The limiting form of the DEnF can be derived in close form and does not depend on the detailed covariance in the null space. Importantly, the ensemble update in the DEnF is not confined to the space spanned by the first guess ensemble, in contrast to ETKF or the EnKF (Evensen, 1994; Burgers et al., 1998; Bishop et al., 2001; Tippett et al., 2003). Two diffuse filters are derived in this paper: one based on perturbed observations called the DEnKF, and one based on a deterministic square root filter called the DETKF. The DEnKF and the DETKF generally reduce to the EnKF and the ETKF respectively, when the ensemble size exceeds the dimension of the model, because in this case there is no null space in which

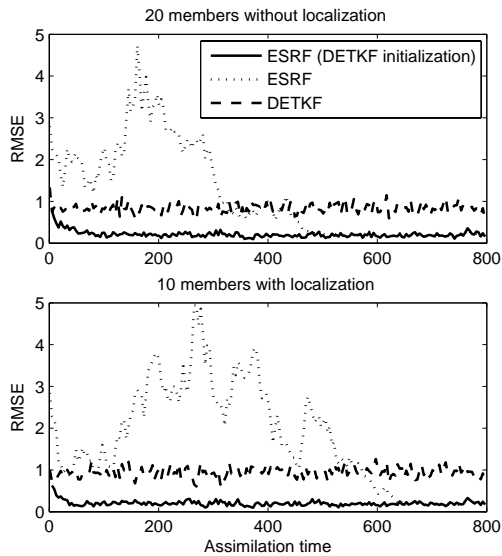


Fig. 6. The root mean square error (RMSE) between analysis and truth as a function of assimilation time for the ESRF with DETKF initialization (solid) and with random initial conditions (dotted) using (a) 20 ensemble members plus constant inflation and (b) 10 ensemble members plus constant inflation and localization. RMSE of DETKF (dashed) is plotted for reference. The inflation factor is 1.08 for (a), and 1.05 for (b).

to apply the diffuse assumption. The diffuse limit is well defined only in observation rich regimes (more precisely, the matrix \mathbf{W} defined in (29) is invertible). In the null space, the analysis produced by the DESRF is strongly coupled to the observations, consistent with assuming infinite forecast covariance in this space, whereas the analysis produced by traditional filters is strongly coupled to the first guess.

Numerical experiments presented in this paper demonstrate that the DETKF and DEnKF successfully prevent filter collapse for small ensemble sizes. Unfortunately, the amplitude of the innovation vectors produced by these filters are too large relative to that assumed internally in the filters. In addition, the analyses produced by the diffuse filters have significantly larger error than those produced by the ESRF with inflation and localization. Inflating the ensemble forecast covariance in the DETKF reduces the analysis errors, but does not reduce them as much as the ESRF with inflation and localization. To investigate the impact of using an imperfect forecast model, we conducted assimilation experiments using a forecast model in which the forcing and dissipation parameters were perturbed relative to the model that generated the truth. We found that the performance of the ESRF was significantly degraded by the presence of model errors, whereas the DETKF was not since it is less dependent on the first guess. These results suggest that the DETKF can outperform ESRF without localization in the more realistic case of small ensemble size and imperfect model, provided enough observations are available to render a well defined diffuse limit.

The DETKF also was found to dramatically accelerate the spin-up time of the ESRF. This result is consistent with the study of Zupanski et al. (2006), who found that the commonly used initial ensemble of uncorrelated random perturbations for the ESRF converged slowly, while initial perturbations that had horizontally correlated errors converged faster. Kalnay and Yang (2009) also found that the spin-up time of EnKF is longer than the corresponding spin-up time in variational methods, and they proposed a scheme to accelerate the spin-up of EnKF applying a no-cost Ensemble Kalman Smoother, and using the observations more than once in each assimilation window in order to maximize the initial extraction of information. We note that the DETKF still requires a guess for the initial condition and error covariances, unlike the diffuse Kalman filter (de Jong, 1991; Koopman, 1997).

A fundamental limitation of the DEnFs, as formulated here, is that it requires a relatively large number of observations. The precise condition is that the matrix \mathbf{W} defined in (29) needs to be invertible. For this operator to be invertible, the observations must be sufficiently numerous as to constraint the analysis in the null space. This constraint is a natural consequence of the diffuse assumption – since the forecast is completely uncertain in the null space, the only other information available for specifying the assimilation is the observations. That is, if neither the forecast nor observations are available in the null space, then there is no basis for estimating the corresponding state. With the emergence of copious data from satellites, this constraint might be satisfied for realistic atmospheric data assimilation. It is possible to generalize the DEnFs to situations in which \mathbf{W} is singular, but this approach was only outlined in this paper.

The limitation that \mathbf{W} be invertible is not only a theoretical limitation of diffuse filters, but also a practical limitation, because the dimension of this matrix is approximately equal to the model dimension minus the ensemble size. For atmospheric or oceanic models, this dimension can easily exceed 100 000, which is clearly impractical at the present time. We briefly described a variational solution for the DETKF that avoids inversion of \mathbf{W} .

A question relevant to all ensemble filters is whether the errors are treated appropriately across update steps. For instance, a vector may project in the ensemble space at one time and project in the null space at the next time. It seems unrealistic to treat the vector as completely unknown at the second step even though it formerly had finite variance at the first step. An equally compelling question arises with respect to ensemble filters – the vector that projects in the ensemble space first and then in the null space second is assumed to have finite uncertainty at the first step and vanishing uncertainty at the second step. In either case, filter performance might be enhanced by accounting for time correlation in the forecast errors, perhaps through an appropriate prior distribution.

The fact that diffuse filters do not perform as well as the ESRF with inflation and localization is instructive. In the DETKF, the covariances in the null space are inflated while the covariances in the ensemble space are not. Conversely, in the ESRF with inflation only, the covariances in the ensemble space are inflated while the covariances in the null space are not. Neither case produces as good an analysis as the ESRF with both inflation and localization. Presumably, the benefits of localization derive from the fact that the forecast errors of the system actually do have spatially local correlations. In other words, the first guess ensemble really does contain information about the null space, even though it is orthogonal to it. It would be interesting and more consistent to develop a filtering scheme that imposes this structure in the prior distribution of the forecast errors, rather than impose it empirically after the fact through the Schur product. Perhaps a better diffuse assumption is that the covariances approach a finite “climatological” value in the null space, with the details of the spatial correlations being estimated through bootstrapping, sub-sampling, or cross validation techniques.

Appendix A

Covariance Update of the DETKF

In this appendix we derive the analysis covariance matrix for the DETKF. First, we substitute the diffuse inverse covariance (28) into the “inverse” form of the analysis covariance (24):

$$\mathbf{P}^a = \left(\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} + \mathbf{U}_E \mathbf{S}_E^{-2} \mathbf{U}_E^T \right)^{-1} \quad (\text{A1})$$

$$= \mathbf{U} \left(\mathbf{U}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{U}^T + \begin{pmatrix} \mathbf{S}_E^{-2} & 0 \\ 0 & 0 \end{pmatrix} \right)^{-1} \mathbf{U}^T. \quad (\text{A2})$$

To examine when this inverse exists, let us define $\mathbf{Z}_E = \mathbf{R}^{-1/2} \mathbf{H} \mathbf{U}_E$ and $\mathbf{Z}_N = \mathbf{R}^{-1/2} \mathbf{H} \mathbf{U}_N$. Then

$$\mathbf{P}^a = \mathbf{U} \begin{pmatrix} \mathbf{Z}_E^T \mathbf{Z}_E + \mathbf{S}_E^{-2} & \mathbf{Z}_E^T \mathbf{Z}_N \\ \mathbf{Z}_N^T \mathbf{Z}_E & \mathbf{Z}_N^T \mathbf{Z}_N \end{pmatrix}^{-1} \mathbf{U}^T. \quad (\text{A3})$$

From standard theorems regarding the inverse of partitioned matrices (Horn and Johnson, 1985, p. 18), the above inverse exists if the following two matrices are invertible:

$$\mathbf{W} = \mathbf{Z}_N^T \mathbf{Z}_N \quad (\text{A4})$$

$$\mathbf{F} = \mathbf{S}_E^{-2} + \mathbf{Z}_E^T \left(\mathbf{I} - \mathbf{Z}_N \left(\mathbf{Z}_N^T \mathbf{Z}_N \right)^{-1} \mathbf{Z}_N^T \right) \mathbf{Z}_E.$$

However, \mathbf{F} is always invertible if \mathbf{W} is invertible. This can be seen by noting that $\mathbf{Z}_N \left(\mathbf{Z}_N^T \mathbf{Z}_N \right)^{-1} \mathbf{Z}_N^T$ is positive semi-definite, in which case \mathbf{F} can be seen to be the sum of a positive definite and positive semi-definite matrices, and hence must itself be positive definite, and thus invertible. This argument establishes that invertibility of \mathbf{W} is a sufficient condition for \mathbf{P}_a to exist.

It turns out that \mathbf{W} also is a necessary condition for \mathbf{P}_a to exist; that is, \mathbf{P}_a is nonsingular only if \mathbf{W} is nonsingular. To show this latter fact, we invoke standard theorems about the determinants (especially of partitioned matrices Johnson and Wichern, 2002, p. 204) to obtain

$$|\mathbf{P}_a| = |\mathbf{Z}_E^T \mathbf{Z}_E + \mathbf{S}_E^{-2}|^{-1} |\mathbf{Z}_N^T \mathbf{Z}_N - \mathbf{Z}_N^T \mathbf{Z}_E \left(\mathbf{Z}_E^T \mathbf{Z}_E + \mathbf{S}_E^{-2} \right) \mathbf{Z}_E^T \mathbf{Z}_N|^{-1} \quad (\text{A5})$$

$$= |\mathbf{Z}_E^T \mathbf{Z}_E + \mathbf{S}_E^{-2}|^{-1} |\mathbf{Z}_N^T \left(\mathbf{I} - \mathbf{Z}_E \left(\mathbf{Z}_E^T \mathbf{Z}_E + \mathbf{S}_E^{-2} \right) \mathbf{Z}_E^T \right) \mathbf{Z}_N|^{-1} \quad (\text{A6})$$

$$= |\mathbf{Z}_E^T \mathbf{Z}_E + \mathbf{S}_E^{-2}|^{-1} |\mathbf{Z}_N^T \left(\mathbf{I} + \mathbf{Z}_E \mathbf{S}_E^2 \mathbf{Z}_E^T \right)^{-1} \mathbf{Z}_N|^{-1}. \quad (\text{A7})$$

Since $\mathbf{Z}_E^T \mathbf{Z}_E + \mathbf{S}_E^{-2}$ is positive definite, it is invertible and the first determinant on the right side exists. Turning now to the second determinant, the matrix $\mathbf{I} + \mathbf{Z}_E \mathbf{S}_E^2 \mathbf{Z}_E^T$ is positive definite and so its inverse, call it \mathbf{B} , exists and also is positive definite. It remains, then, to show that $\mathbf{Z}_N^T \mathbf{B} \mathbf{Z}_N$ is nonsingular to establish that \mathbf{P}_a exists. The quadratic form $\mathbf{x}^T \mathbf{Z}_N^T \mathbf{B} \mathbf{Z}_N \mathbf{x} > 0$ if and only if $\mathbf{Z}_N \mathbf{x} \neq 0$, because \mathbf{B} is positive definite. But if $\mathbf{Z}_N \mathbf{x} \neq 0$, then $\mathbf{x}^T \mathbf{Z}_N^T \mathbf{Z}_N \mathbf{x} \neq 0$. We see then that if $\mathbf{Z}_N^T \mathbf{Z}_N$ is positive definite, then so is $\mathbf{Z}_N^T \mathbf{B} \mathbf{Z}_N$; conversely, if $\mathbf{Z}_N^T \mathbf{Z}_N$ is positive semi-definite, then so is $\mathbf{Z}_N^T \mathbf{B} \mathbf{Z}_N$. This result establishes that the second determinant on the right side exists if and only if \mathbf{W} is nonsingular. We conclude, then, that \mathbf{P}_a exists if and only if \mathbf{W} is invertible.

To derive the square root form of the filter, we project the covariance (A3) onto the ensemble space. This is done by pre- and post-multiplying \mathbf{P}^a by the projection matrix $\mathbf{U}_E \mathbf{U}_E^T$ giving

$$\tilde{\mathbf{P}}^a = \mathbf{U}_E \mathbf{U}_E^T \mathbf{U} \begin{pmatrix} \mathbf{Z}_E^T \mathbf{Z}_E + \mathbf{S}_E^{-2} & \mathbf{Z}_E^T \mathbf{Z}_N \\ \mathbf{Z}_N^T \mathbf{Z}_E & \mathbf{Z}_N^T \mathbf{Z}_N \end{pmatrix}^{-1} \mathbf{U}^T \mathbf{U}_E \mathbf{U}_E^T. \quad (\text{A8})$$

Since $\mathbf{U}_E^T \mathbf{U} = [\mathbf{I} \ 0]$, we need only the $(N - 1) \times (N - 1)$ upper block diagonal of the above inverse matrix. This block is readily computed from standard linear algebra formulas (Horn and Johnson, 1985, p. 18) as

$$\begin{aligned} \tilde{\mathbf{P}}^a &= \mathbf{U}_E \left(\mathbf{S}_E^{-2} + \mathbf{Z}_E^T \mathbf{Z}_E - \mathbf{Z}_E^T \mathbf{Z}_N \left(\mathbf{Z}_N^T \mathbf{Z}_N \right)^{-1} \mathbf{Z}_N^T \mathbf{Z}_E \right)^{-1} \mathbf{U}_E^T \\ &= \mathbf{U}_E \mathbf{S}_E \left(\mathbf{I} + \mathbf{S}_E \mathbf{Z}_E^T \left(\mathbf{I} - \mathbf{Z}_N \left(\mathbf{Z}_N^T \mathbf{Z}_N \right)^{-1} \mathbf{Z}_N^T \right) \mathbf{Z}_E \mathbf{S}_E \right)^{-1} \mathbf{S}_E \mathbf{U}_E^T. \end{aligned} \quad (\text{A9})$$

Inserting the identity matrix $\mathbf{I} = \mathbf{V}^T \mathbf{V}$ just before and after the term in parentheses and invoking the definitions of \mathbf{Z}_E , \mathbf{Z}_N , and (20) gives

$$\tilde{\mathbf{P}}^a = \mathbf{A} \left(\mathbf{I} + \mathbf{A}^T \mathbf{H}^T \left(\mathbf{R}^{-1} - \mathbf{R}^{-1} \mathbf{H} \mathbf{U}_N \left(\mathbf{U}_N^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{U}_N \right)^{-1} \mathbf{U}_N^T \mathbf{H}^T \mathbf{R}^{-1} \right) \mathbf{H} \mathbf{A} \right)^{-1} \mathbf{A}^T. \quad (\text{A10})$$

This equation is the covariance matrix for the DETKF given in (31).

Appendix B

The innovation consistency function for diffusive covariances

The innovation consistency function for the innovation vector is

$$\text{ICF}(N) = \mathbf{z}^T (\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})^{-1} \mathbf{z}. \quad (\text{B1})$$

Substituting (25) and (26) and (22) gives

$$\text{ICF}(N) = \mathbf{z}^T (\mathbf{H}\mathbf{U}_E \mathbf{S}_E^2 \mathbf{U}_E \mathbf{H}^T + \mathbf{H}\mathbf{U}_N \Sigma \mathbf{U}_N \mathbf{H}^T + \mathbf{R})^{-1} \mathbf{z}. \quad (\text{B2})$$

Applying the Sherman-Morrison-Woodbury formula gives

$$\text{ICF}(N) = \mathbf{z}^T (\mathbf{C}^{-1} - \mathbf{C}^{-1} \mathbf{H}\mathbf{U}_N (\Sigma^{-1} + \mathbf{U}_N^T \mathbf{H}^T \mathbf{C}^{-1} \mathbf{H}\mathbf{U}_N)^{-1} \mathbf{U}_N^T \mathbf{H}^T \mathbf{C}^{-1}) \mathbf{z}. \quad (\text{B3})$$

Taking the diffusive limit $\Sigma^{-1} \rightarrow 0$ gives

$$\text{ICF}(N) = \mathbf{z}^T (\mathbf{C}^{-1} - \mathbf{C}^{-1} \mathbf{H}\mathbf{U}_N (\mathbf{U}_N^T \mathbf{H}^T \mathbf{C}^{-1} \mathbf{H}\mathbf{U}_N)^{-1} \mathbf{U}_N^T \mathbf{H}^T \mathbf{C}^{-1}) \mathbf{z}. \quad (\text{B4})$$

Factoring this equation into square root form gives

$$\text{ICF}(N) = \mathbf{z}^T \mathbf{C}^{-1/2} (\mathbf{I} - \mathbf{C}^{-1/2} \mathbf{H}\mathbf{U}_N (\mathbf{U}_N^T \mathbf{H}^T \mathbf{C}^{-1} \mathbf{H}\mathbf{U}_N)^{-1} \mathbf{U}_N^T \mathbf{H}^T \mathbf{C}^{-1/2}) \mathbf{z} \quad (\text{B5})$$

$$= \mathbf{z}^T \mathbf{C}^{-1/2} (\mathbf{I} - \mathbf{G} (\mathbf{G}^T \mathbf{G}) \mathbf{G}^T) \mathbf{C}^{-1/2} \mathbf{z}, \quad (\text{B6})$$

where $\mathbf{G} = \mathbf{C}^{-1/2} \mathbf{H}\mathbf{U}_N$. The term in parentheses is idempotent, and therefore its rank is given by its trace, which is $M - N - 1$ (recall \mathbf{G} is an $M \times (M - (N + 1))$ matrix). Since \mathbf{C} is full rank, the rank of the total matrix in the ICF is $M - N - 1$. Therefore, the function $\text{ICF}(N)$ has a chi-squared distribution with $M - N - 1$ degrees of freedom.

Acknowledgements. This research is supported by NOAA grant NA06OAR4310001. We thank Chris Snyder, acting as reviewer, for numerous stimulating comments that led to substantial improvements in the manuscript. We also thank two anonymous reviewers for their constructive comments.

Edited by: O. Talagrand

Reviewed by: C. Snyder, T. Miyoshi, and another anonymous referee

References

- Anderson, B. D. O. and Moore, J. B.: Optimal Filtering, Dover Publications, 1979.
- Anderson, J. L.: An adaptive covariance inflation error correction algorithm for ensemble filters, *Tellus A*, 59, 210–224, 2007.
- Anderson, J. L. and Anderson, S. L.: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts, *Mon. Weather Rev.*, 127, 2741–2758, 1999.
- Ansley, C. F. and Kohn, R.: Estimation, filtering and smoothing in state space models with incompletely specified initial conditions, *Ann. Stat.*, 13, 1286–1316, 1985.

- Bishop, C. H., Etherton, B., and Majumdar, S. J.: Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects, *Mon. Weather Rev.*, 129, 420–436, 2001.
- Burgers, G., van Leeuwen, P. J., and Evensen, G.: On the Analysis Scheme in the Ensemble Kalman Filter, *Mon. Weather Rev.*, 126, 1719–1724, 1998.
- de Jong, P.: The diffuse Kalman Filter, *Ann. Stat.*, 19, 1073–1083, 1991.
- Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, 99, 1043–1062, 1994.
- Gaspari, G. and Cohn, S. E.: Construction of Correlation Functions in Two and Three Dimensions, *Q. J. Roy. Meteor. Soc.*, 125, 723–757, 1999.
- Hamill, T. M., Whitaker, J. S., and Snyder, C.: Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter, *Mon. Weather Rev.*, 129, 2776–2790, 2001.
- Haykin, S.: Kalman Filtering and Neural Networks, in: Kalman filters, edited by: Haykin, S., chap. 1, p. 284, John Wiley & Sons, 2001.
- Horn, R. A. and Johnson, C. R.: Matrix Analysis, Cambridge University Press, New York, 561 pp., 1985.
- Houtekamer, P. L. and Mitchell, H. L.: Data Assimilation Using an Ensemble Kalman Filter Technique, *Mon. Weather Rev.*, 126, 796–811, 1998.
- Houtekamer, P. L. and Mitchell, H. L.: A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation, *Mon. Weather Rev.*, 129, 123–137, 2001.
- Johnson, R. A. and Wichern, D. W.: Applied Multivariate Statistical Analysis, Pearson Education Asia, 2002.
- Kalnay, E. and Yang, S.-C.: Accelerating the spin-up of ensemble Kalman filtering, *Q. J. Roy. Meteorol. Soc.*, submitted, 2009.
- Klinker, E., Rabier, F., Kelly, G., and Mahfouf, J.-F.: The ECMWF operational implementation of four-dimensional variational assimilation. III: Experimental results and diagnostics with operational configuration, *Q. J. Roy. Meteorol. Soc.*, 126, 1191–1215, 2000.
- Koopman, S. A.: Exact Initial Kalman Filtering and Smoothing for Nonstationary Time Series Models, *J. Am. Stat. Assoc.*, 92, 1630–1638, 1997.
- Lorenz, E. N. and Emanuel, K. A.: Optimal sites for supplementary weather observations: simulation with a small model, *J. Atmos. Sci.*, 55, 399–414, 1998.
- Maybeck, P. S.: Stochastic models, estimation, and control, Academic Press, 423 pp., 1979.
- Sakov, P. and Oke, P. R.: Implications of the form of the ensemble transformations in the ensemble square root filters, *Mon. Weather Rev.*, 136, 1042–1053, 2008.
- Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M., and Whitaker, J. S.: Ensemble square-root filters, *Mon. Weather Rev.*, 131, 1485–1490, 2003.
- Whitaker, J. and Hamill, T. M.: Ensemble Data Assimilation Without Perturbed Observations, *Mon. Weather Rev.*, 130, 1913–1924, 2002.
- Zupanski, M., Fletcher, S. J., Navon, I. M., Uzunoglu, B., Heikes, R. P., Randall, D. A., Ringler, T. D., and Daescu, D.: Initiation of ensemble data assimilation, *Tellus A*, 58, 159–170, 2006.