Ocean Dynamics (2004) 54: 561–569
DOI 10.1007/s10236-004-0104-9

Jeffrey D. Kepert

On ensemble representation of the observation-error covariance in the Ensemble Kalman Filter

Received: 24 February 2004 / Accepted: 10 September 2004

Abstract Evensen (2003) presents a modification of the Ensemble Kalman Filter (EnKF), in which the observation-error and background-error covariance matrices are both represented by ensembles, in contrast to the usual practice, where only the background error is so represented. It is shown that this modification can cause the ensemble to collapse to a single member, in the common situation where the number of observations is more than twice the number of ensemble members, and to be rank-deficient when the number of observations is greater than or equal to the ensemble size. It is also shown that some further modifications to the scheme, presented by Evensen as offering numerical efficiencies, can prevent this collapse. However, these latter modifications are shown in some simple numerical examples to require tuning to produce acceptable results, which are nevertheless inferior to those of the standard EnKF.

Keywords Data assimilation · Ensemble Kalman Filter · Observation-error covariance

1 Introduction

Since its introduction by Evensen (1994), the Ensemble Kalman Filter (EnKF) has been developed as an algorithm for data assimilation in meteorology and oceanography, which offers potentially important advantages over methods currently used operationally. In particular, the use of a Monte Carlo method to propagate the analysis-error covariance matrix into the future allows the development of flow-dependent structures in the background-error covariance matrix. This offers significant benefits over the use of fixed covariance models, such as are widely used in operational statistical interpolation and 3D-VAR schemes at present. Moreover, these benefits are achieved without the substantial development cost of coding and validating adjoint and tangent-linear versions of the forecast model, as required by the 4D-VAR algorithm.

These potential benefits have stimulated a number of implementations with several variations upon the basic algorithm. Here, the standard EnKF (SEnKF) is defined to be that of Evensen (1994), with the important modification that the observations are perturbed as described by Burgers et al. (1998) to avoid underestimation of the analysis covariance. A number of variants to the SEnKF have been proposed, some of which appear to hold considerable promise. However, they will not be studied here, as the purpose is rather to explore the properties of a recent modification due to Evensen (2003, henceforth E2003). In particular, the ensemble background covariances will not be localized as done by Houtekamer and Mitchell (1998), nor is their double EnKF adopted. Similarly, alternatives to perturbed observations as expounded by Anderson (2001), Bishop et al. (2001) and Whitaker and Hamill (2002) are not considered.

Evensen (2003) provides a broad, if brief, review of the various approaches, before going on to discuss the standard algorithm in more detail. In the development a further variant is introduced, in which the observation-error covariance in the expression for the gain matrix is estimated from the ensemble of perturbed observations, rather than being used directly. For convenience, this variant will be referred to here as the observation covariance ensemble Kalman filter (OCEnKF). Evensen shows that the OCEnKF can offer some numerical economies over the SEnKF, particularly in the case where the size of the ensemble is much less than the number of observations. These economies arise essentially because of the reduced rank of the innovation-error covariance matrix, that is, the covariance of the observations minus the (interpolated) background field.

Notation and analysis schemes are defined in Section 2. Section 3 contains an analytical proof that

Responsible Editor: Jörg-Olaf Wollff

J. D. Kepert
Bureau of Meteorology Research Centre,
GPO Box 1289K, Melbourne Vic 3001,
Australia
e-mail: J.Kepert@bom.gov.au

the OCEnKF leads in certain circumstances to the collapse of the filter, that is, to the situation where all the analysis members become identical. Numerical examples are also presented in which the ensemble-mean analysis contains features of erroneously large amplitude, which are interpreted in terms of the spectrum of the Kalman gain matrix. Evensen (2003) presented an alternative implementation of the OCEnKF which offers improved numerical efficiency. It is shown in Section 4 that this implementation contains an implicit perturbation to the innovation covariance matrix, which has the side effect of preventing collapse. Some experiments with this implementation strategy in a simple system are used to demonstrate, however, that the problems are only partially ameliorated, rather than completely solved, and that the results are inferior to those from the SEnKF.

2 Notation and definitions

Here, notation and analysis schemes are briefly defined. This is a slight extension of E2003's notation, and generally consistent with standard data assimilation notational conventions, except where these conflict with E2003.

Let P^b denote the background-error covariance matrix, R the observation-error covariance matrix, H the (linearized) observation operator, y a vector of m observations and x^b a vector containing the n-element background state. The analysis equation for Kalman Filter (KF) may be written

$$x^a = x^b + K(y - Hx^b), \tag{1}$$

where the Kalman gain matrix is

$$K = P^b H^T (HP^b H^T + R)^{-1} \tag{2}$$

and the analysis-error covariance is

$$P^a = (I - KH)P^b. \tag{3}$$

In the SEnKF, a matrix $A \in \mathbb{R}^{n\times N}$ containing the background state from the N ensemble members replaces x^b, and a matrix $D \in \mathbb{R}^{m\times N}$ containing the perturbed observations replaces y. An overbar denotes the ensemble mean and a prime the deviations from it, except for the observation deviations which (to be consistent with E2003) are denoted Υ. The analysis equation becomes

$$A^a = A + K_e(D - HA), \tag{4}$$

where the Kalman gain matrix is

$$K_e = P_e^b H^T (HP_e^b H^T + R)^{-1}, \tag{5}$$

the ensemble estimate of the background-error covariance is

$$P_e^b = \frac{A'A'^T}{N-1}, \tag{6}$$

and the analysis-error covariance is estimated with

$$P_e^a = (I - K_e H)P_e^b. \tag{7}$$

In the OCEnKF, R in Eq. (5) is replaced by an approximation calculated from the observation perturbations:

$$R_e = \frac{\Upsilon\Upsilon^T}{N-1} \tag{8}$$

and the subscript *ee* will be used on the gain and analysis covariance matrices to denote that both P^b and R are being represented by ensembles:

$$\begin{aligned} K_{ee} &= P_e^b H^T (HP_e^b H^T + R_e)^+ \\ &= A'A'^T H^T (HA'A'^T H^T + \Upsilon\Upsilon^T)^+, \end{aligned} \tag{9}$$

and

$$P_{ee}^a = (I - K_{ee}H)P_e^b. \tag{10}$$

The superscript + denotes the matrix pseudo-inverse, a generalization of the usual matrix inverse (Golub and van Loan, (1996), pp. 257–258), and used here because $HP_e^b H^T + R_e$ may be singular.

3 Proof that the OCEnKF collapses

3.1. Background

The analysis equation in the OCEnKF (Eq. 54 of E2003) may be written as the sum of its ensemble mean and perturbation parts:

$$\begin{aligned} A^a = \overline{A} &+ A'A'^T H^T (HA'A'^T H^T + \Upsilon\Upsilon^T)^+ (\overline{D} - H\overline{A}) \\ &+ A' + A'A'^T H^T (HA'A'^T H^T \\ &+ \Upsilon\Upsilon^T)^+ (\Upsilon - HA'). \end{aligned} \tag{11}$$

Now, combine the background ensemble under the observation operator HA and the perturbed observations D into a single matrix

$$Z = [HA,\, D] \in \mathbb{R}^{m\times 2N}, \tag{12}$$

where m is the number of observations and N the ensemble size. Define

$$X_N = I_N - 1_N \in \mathbb{R}^{N\times N}, \tag{13}$$

where I_N is the $N \times N$ identity matrix and 1_N is the $N \times N$ matrix with all entries $1/N$. Recall 1_N is the averaging operator, $A1_N = \overline{A}$. Note that $X_N^2 = X_N$ and $X_N^T = X_N$, and so X_N is both an orthogonal projection and its own pseudo-inverse. As $HAX_N = HA'$ and $DX_N = \Upsilon$, X_N is the projection onto the subspace of the space spanned by the ensemble where the ensemble mean is 0, and can be regarded as the perturbation operator. Also, let

$$X = \begin{bmatrix} X_N & 0 \\ 0 & X_N \end{bmatrix} \in \mathbb{R}^{2N\times 2N}, \tag{14}$$

which is similarly an orthogonal projection and its own pseudo-inverse, and observe that rank $(X) = 2N - 2$.

3.2 The first proof

Let $ZX = U\Sigma V^T$ be the singular-value decomposition of $ZX = [HA', \Upsilon]$ and $p = \text{rank}\ (ZX)$. It would normally be the case that the background and observation perturbations are linearly independent, so Z is of full rank and $p = \min(2N-2, m)$, but this is not necessary in what follows. In any event, at least the last two diagonal elements of Σ will be zero, and the corresponding columns of V are an orthonormal basis for the null space of X. They may thus be taken without loss of generality to be $(1, ..., 1, 0, ..., 0)^T/\sqrt{N}$ and $(0, ..., 0, 1, ..., 1)^T/\sqrt{N}$.

The perturbation part of Eq. (11) is now written as

$$
\begin{aligned}
&A' + K_{ee}(\Upsilon - HA') \\
&= A'\left[I_N - A'^T H^T (HA'A'^T H^T + \Upsilon\Upsilon^T)^+ (HA' - \Upsilon)\right] \\
&= A'\left\{I_N - [I_N\ 0](ZX)^T [ZX(ZX)^T]^+ (ZX)\begin{bmatrix} I_N \\ -I_N \end{bmatrix}\right\} \\
&= A'\left\{I_N - [I_N\ 0]V\Sigma^T\Sigma^{T+}\Sigma^+\Sigma V^T\begin{bmatrix} I_N \\ -I_N \end{bmatrix}\right\} \\
&= A'[I_N\ 0]V\left\{I_{2N} - \begin{bmatrix} I_p & 0_{p\times(2N-p)} \\ 0_{(2N-p)\times p} & 0_{(2N-p)\times(2N-p)} \end{bmatrix}\right\} V^T\begin{bmatrix} I_N \\ -I_N \end{bmatrix} \\
&= A'[I_N\ 0]V\begin{bmatrix} 0_{p\times p} & 0_{p\times(2N-p)} \\ 0_{(2N-p)\times p} & I_{2N-p} \end{bmatrix} V^T\begin{bmatrix} I_N \\ -I_N \end{bmatrix}. \qquad (15)
\end{aligned}
$$

Consider first the case where Z is of full rank and $p = 2N-2 \le m$, or $N \le m/2+1$. With the last two columns of V as described above, Eq. (15) becomes

$$
\begin{aligned}
&A'[I_N\ 0]V\begin{bmatrix} 0_{(2N-2)\times(2N-2)} & 0_{(2N-2)\times 2} \\ 0_{2\times(2N-2)} & I_2 \end{bmatrix} V^T\begin{bmatrix} I_N \\ -I_N \end{bmatrix} \\
&\quad = A'1_N = 0. \qquad (16)
\end{aligned}
$$

and the analysis ensemble has collapsed to a single member.

For typical global atmospheric assimilation systems, $m \sim 10^5 - 10^6$, so the condition $N \le m/2+1$ is likely to be met for any reasonably conceivable ensemble analysis system. The assumption that Z is of full rank is equivalent to assuming that the background ensemble (under the observation operator H) and the perturbed observations are linearly independent, and this will almost always be true. In fact, as the analysis ensemble is a subset of the subspace spanned by the background ensemble, it is clearly highly desirable that the latter be of full rank, so as to span as large a subspace of the model space as is possible. However, it can also be seen from the above argument that further loss of rank will occur in the analysis ensemble, in the case where Z is not of full rank.

Next, the case $N \le p = m < 2N-2$. It is clear from Eq. (15) that the analysis perturbations have rank at most $2N-m$, and calculation of the terms $[I_N, 0]V$ and $([I_N, -I_N]V)^T$ with the above choice for the last two columns of V shows that an additional dimension is lost, giving a rank of at most $2N-m-1$. Thus, the analysis ensemble will be rank-deficient whenever $N \le m$.

3.3 An alternative proof

A shorter proof of the rank deficiency of the analysis ensemble is now presented, which yields further insight into the reason for the collapse. However, unlike the previous proof, it does not lead to the results to be presented below on the spectra of the gain and analysis-error covariance matrices, nor to the reasons that E2003's numerical scheme avoids collapse. Thus, both approaches are useful.

The matrix R_e is symmetric and positive semidefinite, so there exists an orthonormal $S \in \mathbb{R}^{m\times m}$ such that $\widetilde{R}_e = SR_eS^T$ is diagonal. Now, use S to transform the perturbed observations $\widetilde{D} = SD$. The observation operator for $\widetilde{D}$ is SH, and the ensemble estimate of the transformed-observation-error covariance is $\widetilde{R}_e$. Precisely $m-N+1$ of the diagonal entries of $\widetilde{R}_e$ will be zero, because R_e is of rank $N-1$. That is, so far as the OCEnKF is concerned, $m-N+1$ of the transformed observations are perfect. Lorenc (2003, Appendix A) shows that assimilating a perfect observation into the EnKF removes a degree of freedom from the ensemble, and so the analysis ensemble in the OCEnKF will be of rank $N-(m-N+1) = 2N-m-1$. Examination of Lorenc's (2003) argument shows that his result is valid only if the ensemble has rank > 1; once sufficient of the perfect observations have been assimilated to reduce the rank of the ensemble to 1, no further reduction occurs.

Hence, complete collapse occurs whenever $2N-m-1 \le 1$, or equivalently $N \le m/2+1$, as shown above. Rank deficiency of the analysis ensemble will occur when $2N-m-1 < N$, or $N \le m$, as obtained above.

3.4 Eigenvalue analysis

When $N \le m/2+1$, HK_{ee} has precisely $N-1$ eigenvalues equal to 1, with the rest being 0. This result is now demonstrated, and its implications discussed.

If $[I_N, -I_N]^T$ in Eq. (15) is replaced with either $[I_N, 0]^T$ or $[0, I_N]^T$, two related results follow:

$$
\begin{aligned}
K_{ee}HA' &= A' \\
K_{ee}\Upsilon &= 0.
\end{aligned} \qquad (17)
$$

Using the second of these and the properties of the pseudo-inverse, the ensemble estimate of the analysis-error covariance matrix in the OCEnKF is

$$
\begin{aligned}
P^a_{ee} &= \frac{1}{N-1}[A' + K_{ee}(\Upsilon - HA')][A' + K_{ee}(\Upsilon - HA')]^T \\
&= P^b_e - P^b_e H^T K^T_{ee} - K_{ee}HP^b_e + K_{ee}(HP^b_e H^T + R_e) \\
&\quad \times K^T_{ee} = (I_n - K_{ee}H)P^b_e, \qquad (18)
\end{aligned}
$$

which must be zero when the analysis ensemble collapses. If e is an eigenvector of P_e^b with $\lambda \neq 0$ the corresponding eigenvalue, then

$$0 = H(I - K_{ee}H)P_e^b e = \lambda(I - HK_{ee})He \tag{19}$$

and so He is an eigenvector of HK_{ee} with eigenvalue 1. Furthermore, it follows from Eq. (9) that

$$\text{rank}(HK_{ee}) \leq \text{rank}(P_e^b) = N - 1, \tag{20}$$

and so HK_{ee} has precisely $N - 1$ non-zero eigenvalues, all of which are 1.

In contrast, consider the SEnKF with $R = I_m$ for simplicity. If

$$HP_e^b H^T = E\Lambda E^T \tag{21}$$

is the eigenvalue decomposition, then the gain matrix K_e satisfies

$$\begin{aligned} HK_e &= HP_e^b H^T (HP_e^b H^T + I)^{-1} \\ &= E(\Lambda(\Lambda + I)^{-1})E^T, \end{aligned} \tag{22}$$

and so the non-zero eigenvalues of HK_e, $\lambda_i/(1 + \lambda_i)$, are all less than 1.

With this decomposition, the analysis equation (with subsequent interpolation to observation space) for the SEnKF can be interpreted as a transformation of the innovations into the eigenspace of HK_e [left multiplication by E^T] followed by adjustment of their amplitude (left multiplication by $\Lambda(\Lambda + I)^{-1}$), and transformation back into the observation space (left multiplication by E) to give the analyzed increments. Here, modes present in the innovations which correspond to non-zero eigenvalues of $HP_e^b H^T$ are present in the analyzed increments, but with reduced amplitude, while modes corresponding to zero eigenvalues are eliminated. In contrast, all the eigenvalues in HK_{ee} in the OCEnKF are either 0 or 1, so the observed modes are either eliminated or unchanged in the analysis.

Interpretation of this is complicated by the fact that HK_{ee} is not in general symmetric, and the eigenvectors of $HP_e^b H^T$ and $HP_e^b H^T + R_e$ are different. Some examples will be presented below in which the result of HK_{ee}'s eigenvalues being either 0 or 1 is that the gravest and shortest modes are, respectively, preserved and eliminated, and therefore treated approximately correctly. However, there is an intermediate range which is moderately damped in the SEnKF, but appears in the OCEnKF analysis with spuriously large amplitude.

4 The impact of some approximations

Inspection of Eq. (15) shows that the collapse in the OCEnKF arises because ZX appears in both the gain matrix and the observation innovations. One might, therefore, suspect that a suitable approximation to, for instance, $(HA'A'^T H^T + \Upsilon\Upsilon^T)^+ = [(ZX)(ZX)^T]^+$ might eliminate the problem. Naturally, the primary consideration in making such an approximation should be that the analysis ensemble properly represents the analysis covariance matrix. Evensen (2003) presents an alternative solution strategy which offers numerical efficiencies, which will be shown to also have the effect of approximating the pseudo-inverse and preventing collapse. A different approximation strategy, which retains only the dominant modes in calculating the pseudo-inverse, will also be considered. However, it will be seen that both modifications to the OCEnKF produce results which are inferior to those of the SEnKF. A further strategy would be to filter the background covariance matrix, as suggested by Houtekamer and Mitchell (1998). This is not examined here as its effect would be to restore the innovation-error covariance matrix to full rank, thus eliminating any potential numerical benefit derived from its being rank-deficient. Given the lack of any efficiency gain through using R_e in a localized OCEnKF, one might as well revert to the SEnKF and reduce the overall sampling error. Finally, the possibility of a sequential version of the OCEnKF will be discussed.

Evensen (2003) suggests that the pseudo-inverse $(HA'A'^T H^T + \Upsilon\Upsilon^T)^+$ be calculated from the singular-value decomposition of the $m \times N$ matrix[1]

$$HA' + \Upsilon = U\Sigma V^T, \tag{23}$$

whence the pseudo-inverse becomes

$$(HA'A'^T H^T + \Upsilon\Upsilon^T)^+ \approx U\Sigma^+\Sigma^{T+}U^T. \tag{24}$$

This becomes an exact equality if E2003's Eq. (57),

$$HA'\Upsilon^T = 0, \tag{25}$$

is true. E2003 states that is equivalent to assuming that the ensemble and observation perturbations are uncorrelated, which is usually the case. This is correct, if one interprets Eq. (25) as being stochastically true, but any particular realization of uncorrelated perturbations will not satisfy Eq. (25) exactly, because of finite sample size.

In fact, Eq. (25) will be true for a particular ensemble realization only if the columns of Υ^T are chosen from the right null space of HA'. However, $HA' = HAX_N$, and so provided that HA is of full column rank, the null space consists of vectors with all components equal. That is, Eq. (25) is exactly satisfied only under the requirement that all the observation perturbation vectors are equal, which would rather defeat the purpose of perturbing the observations. If they are chosen otherwise, the violation of Eq. (25) amounts to an approximation of the pseudo-inverse.

The gain matrix and analysis covariance matrix in the OCEnKF under the approximation (24) will be denoted $\widehat{K}_{ee}$ and $\widehat{P}_{ee}^a$, respectively.

4.1 Sampling bias in covariance matrices

Before proceeding, it is necessary to first consider the spectra of P_e^b and R_e. For convenience, consider a

[1] Note that U, Σ and V here are different to those in Section 2.

periodic one-dimensional analysis domain $[0, m)$ in which the error covariances are Gaussian with unit variance and length-scale L; that is, the covariance between points x and y is given by

$$C(x,y) = \exp\left[-0.5(\min(|x-y|, 2\pi - |x-y|)/L)^2\right]. \quad (26)$$

The Gaussian function has the convenient property that its Fourier transform is also Gaussian, with length scale $1/(2\pi L)$; a wide function in physical space is narrow in Fourier space and vice versa. Moving from continuous to discrete space and dividing the domain into m equally spaced gridpoints $\{0, \ldots m-1\}$ with m even, the covariance matrix $Q \in \mathbb{R}^{m \times m}$ is defined by

$$(Q)_{jk} = C(x_j, x_k); \quad (27)$$

later this will be taken to be either the background-error or observation-error covariance matrix by appropriate choice of L.

Define also the discrete Fourier transform matrix $F \in \mathbb{C}^{m \times m}$ by

$$(F)_{jk} = \exp(-2\pi ijk/m), \quad (28)$$

then

$$F^* Q F = \sqrt{2\pi} L \mathrm{diag}\left\{ \exp\left[-\frac{1}{2}\left(\frac{2\pi L \min(j, m-j)}{m} \right)^2 \right], \right.$$
$$\left. j = 1, \ldots, m \right\} \quad (29)$$

where F^* denotes the complex conjugate transpose of F. The m elements of the diagonal of F^*QF are the Fourier coefficients and also clearly the eigenvalues; note that they are real and that all except the first and $m/2+1$'th are repeated. The corresponding columns of F are complex eigenvectors, but the vectors for each repeated eigenvalue are a complex conjugate pair and so by taking an appropriate linear combination, real eigenvectors can be found. The vectors corresponding to the non-repeated eigenvalues are real.

Now consider the Monte Carlo estimation of Q. That is, choose $B \in R^{m \times N}$ whose columns are N random samples over the analysis domain, chosen from a Gaussian distribution with covariance matrix Q, and form the ensemble covariance matrix

$$Q_e = \frac{1}{N-1}(B - \overline{B})(B - \overline{B})^T. \quad (30)$$

Note that

$$\langle \mathrm{trace}(Q_e) \rangle = \mathrm{trace}(Q) = m, \quad (31)$$

where the angle brackets $\langle . \rangle$ denote the expected value, while

$$\mathrm{rank}(B - \overline{B}) \leq N - 1, \quad (32)$$

so that Q_e has at most $N-1$ non-zero eigenvalues. These eigenvalues satisfy

$$\langle \mathrm{trace}(Q_e) \rangle = \sum_{i=1}^{n} \langle \lambda_i(Q_e) \rangle = m, \quad (33)$$

where $\lambda_i(Q_e)$ denotes the i'th-largest eigenvalue of Q_e. The mean of the non-zero eigenvalues thus has the expected value $m/(N-1)$.

Now consider two cases, the limit $L \to 0$, and L large. The first corresponds to the observation-error covariance matrix R, which is typically close to diagonal, while the second corresponds to the background-error covariance matrix P^b, which is relatively broad. Clearly, all of R's eigenvalues are unity. In contrast, R_e will have only $N-1$ non-zero eigenvalues, whose mean has the expected value $m/(N-1)$, which is larger than 1. This constitutes a sampling bias which guarantees that the spectrum of R_e will be very different to that of R.

For the large L case, P^b has a relatively broad covariance function, or in eigenspace by Eq. (29), a relatively narrow one in which many of the eigenvalues are very small. Thus the constraints Eq. (33) and that at most $N-1$ of the eigenvalues are non-zero, lead to a much less severe sampling bias in P_e^b's spectrum (and hence $HP_e^bH^T$'s), than in R_e's.

Finally, recall that for any symmetric positive semi-definite real matrices S and T, that

$$\lambda_i(S+T) \geq \max[\lambda_i(S), \lambda_i(T)] \geq 0 \quad (34)$$

(Golub and van Loan 1996, p. 396). Thus, the overestimate of the leading eigenvalues of R_e due to the sampling bias places a lower bound under the leading eigenvalues of $HP_e^bH^T + R_e$.

4.2 A simple example

These ideas are illustrated in Fig. 1, which was calculated for the above situation with $m = 128$ observations, one at each of the $n = 128$ gridpoints (so $H = I_n$), and $N = 64$ ensemble members. The background-error covariance P^b is defined as above with length scale $L = 8$, and the observation-error covariance is $R = I$. It can be seen in Fig. 1a that sampling error slightly inflates the first two eigenvalues of P_e^b, but the spectrum is otherwise well represented. The spectrum of R is flat, but the leading eigenvalues of R_e are substantially greater than those of R, consistent with the above analysis. In Fig. 1b the spectra of $P^b + R$ and $P_e^b + R$ are close to that of P_e^b and P^b at first, while their smaller eigenvalues converge to those of R. In contrast, the sampling bias in the leading part of the spectrum of R_e causes the spectrum of $P_e^b + R_e$ to diverge from that of P_e^b relatively early, and to be spuriously large in this range.

Figure 1c and d shows the spectrum of the Kalman gain and analysis covariance matrices for the various cases. Clearly, the assumption (25) improves the spectrum of $\widehat{K}_{ee}$ over that of K_{ee}, but at the cost that the leading eigenvalues now exceed 1. Consistent with this, the analysis covariance spectrum $\widehat{P}_{ee}^a$ now has too much energy in the leading modes.

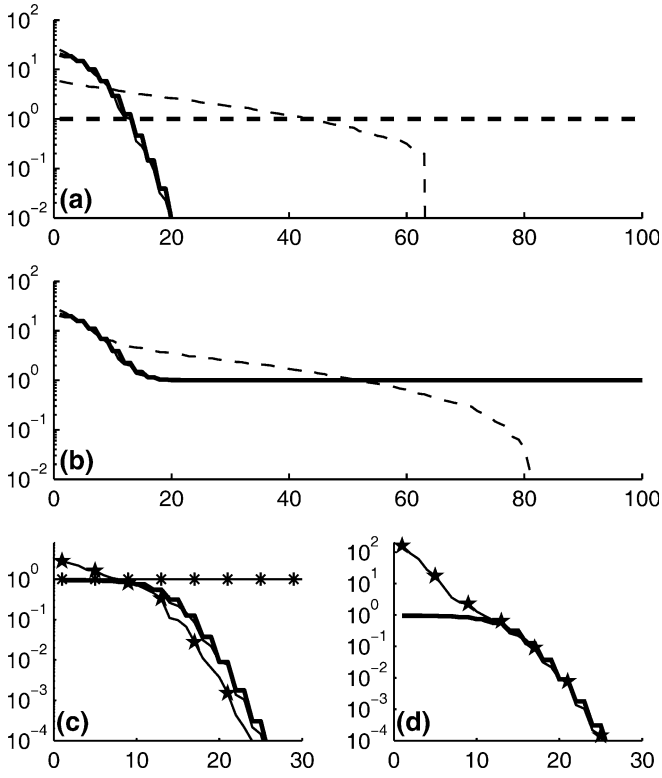

Fig. 1a–d Eigenvalue spectra for the case described in the text with $m = n = 128$, $L = 8$ and $N = 64$. **a** Background-error and observation-error covariance matrices: P^b (*heavy line*); P_e^b (*light line*, nearly obscured by the former); R (*heavy dashed*); R_e (*light dashed*). **b** Innovation-error covariance matrices: $P^b + R$ (*heavy line*); $P_e^b + R$ (*light line*, nearly obscured by the former); $P_e^b + R_e$ (*light dashed*). **c** Kalman gain matrices: K (*heavy line*) K_e (*light line*, nearly obscured by the former); K_{ee} (*line with asterisks*); $\widehat{K}_{ee}$ (*line with stars*). **d** Analysis covariance matrices: P_a (*heavy line*); P_e^a (*light line*, nearly obscured by the former); $\widehat{P}_{ee}^a$ (*line with stars*). P_{ee}^a is everywhere 0 (because of the ensemble collapse) and so does not appear on these axes

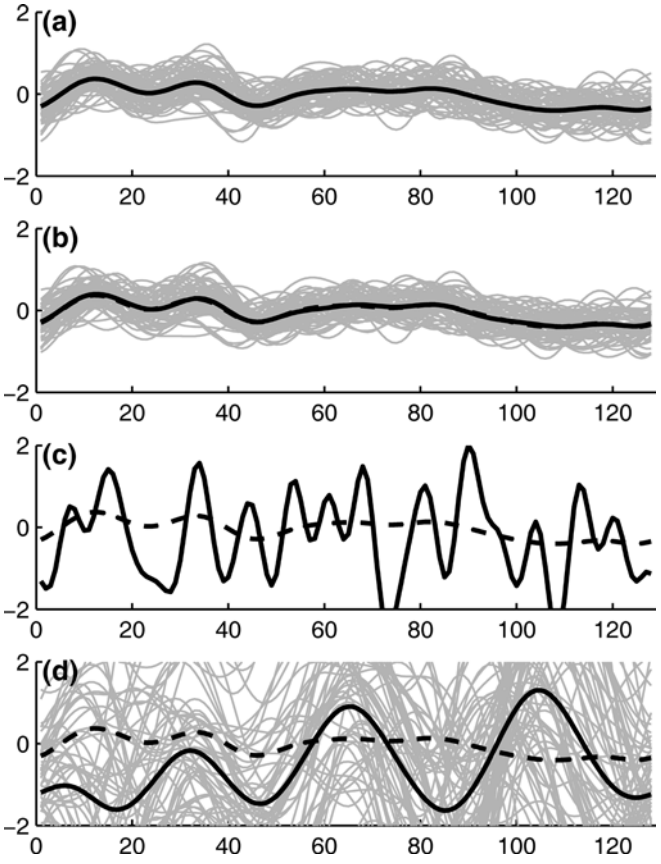

Fig. 2a–d Mean analysis (*heavy black line*) and ensemble members (*light grey lines*) for the case in Fig. 1 under various analysis schemes. The KF mean analysis from panel **a** is repeated in the lower panels as a *heavy dashed line*. **a** Analysis ensemble from applying KF to each member of the background ensemble (i.e. using P^b and R). **b** SEnKF (i.e. using P_e^b and R). **c** OCEnKF (i.e. using P_e^b and R_e). **d** OCEnKF with pseudo-inverse approximated by assuming $HA'\Upsilon^T = 0$

Figure 2 shows a realization of the analysis ensemble under the various schemes for this case. Here, the background ensemble was generated by random sampling of a multivariate Gaussian distribution with covariance matrix P^b, while the observations and their perturbations were independent Gaussian with mean 0 and variance 1. Figure 2a shows the mean and ensemble analyses using the "true" covariance matrices P^b and R, while Fig. 2b shows the situation for the SEnKF, with P_e^b and R. The third panel has the mean analysis for the OCEnKF; note that the analysis ensemble has here collapsed, and that the mean analysis contains an excess of energy at small scales, consistent with the spectrum of K_{ee} shown in Fig. 1. The final panel shows the effect of applying Eq. (25), and it is apparent that the spread is too large, and that the mean analysis contains too much energy at long scales. These are both consistent with the spectra of $\widehat{K}_{ee}$ and $\widehat{P}_{ee}^a$ in Fig. 1c,d.

Similar calculations were performed for a two-dimensional domain. Here, the spectrum of P^b was broader than in the one-dimensional case, length scales being equal. Although this gave greater sensitivity to the sampling bias discussed above than in the one-dimensional case, this sensitivity was nevertheless much less than for R. The effects of representing P^b and R by ensembles were very similar to those found in the one-dimensional case and presented above, as was the impact of the pseudo-inverse approximation (Eq. 24) discussed above, and further approximations to be examined below.

4.3. Truncating the SVD of $HA' + \Upsilon$

In the SVD (Eq. 23), Σ is a diagonal matrix which, assuming HA' and Υ have rank $N - 1$ and ignoring round-off error, has precisely one diagonal entry of 0. The matrix Σ^+ is calculated by replacing all non-zero elements of Σ^T by their reciprocal. However, round-off error would normally require that any elements smaller than some tolerance be regarded as zero, rather than being inverted.

The numerical noise in the singular-value decomposition is of order

$$
\begin{aligned}
&m\epsilon\|(HA' + \Upsilon)(HA' + \Upsilon)^T\|_2 \\
&\quad \le m\epsilon\big(\|HA'A'^T H^T\|_2 + \|\Upsilon\Upsilon^T\|_2 \\
&\quad + \|HA'\Upsilon^T + \Upsilon A'^T H^T\|_2\big) \\
&\quad \sim m\epsilon\Big(\sqrt{2\pi}LN\Big), \qquad (35)
\end{aligned}
$$

where ϵ is the numerical precision, $\|.\|_2$ denotes the matrix 2-norm, and the last line follows because $HA'A'^T H^T$ has the largest 2-norm under the assumptions in the previous subsection about the relative shapes of the spectra of P^b and R.

If the decomposition is done with, say, IEEE 64-bit numerical precision with $\epsilon \sim 2 \times 10^{-16}$, this will usually be much less than the tolerance applied by E2003, who chose this tolerance to be such that either 99% (in the text) or 99.9% (in the code in the Appendix) of the total variance in Σ is retained. The code sample uses the Eispack routine dgesvd to perform the SVD in which the initial d implies double precision. The tolerance applied by E2003 is therefore much more severe than round-off considerations would normally require, and therefore constitutes a second level of approximation in calculating the pseudo-inverse. Moreover, it is presented without any guidance as to how to choose this tolerance.

Figures 3 and 4 show the impact of making this truncation, for the same case as above. It is clear that the most severe truncation shown nearly eliminates the spurious eigenvalues of $\widehat{K}_{ee}$ which exceed 1, and greatly reduces the similar eigenvalues of $\widehat{P}^a_{ee}$. (Some of the apparent over estimation of the leading part of the spectrum of P^a is due to sampling error; for comparison, an ensemble estimate of this for the "true" case is also shown.) Note though that a substantial systematic overestimation remains in all cases. In Fig. 4 the two least severe truncation cases have too much ensemble spread, while all three of the mean analyses are significantly different to that calculated using the true P^b and R (Fig. 4, dashed).

4.4 Truncating the pseudo-inverse of $HA'A'^T H^T + \Upsilon\Upsilon^T$

Here another approximation strategy to the pseudo-inverse term, not considered by E2003, is explored. Given the excess energy at short scales in the analysis with the full pseudo-inverse in Figs. 1 and 2, and its attribution to the spectrum of K_{ee}, it is reasonable to consider whether removing some of the shorter scales, without making the assumption (25), would yield benefits. The effects of this are shown in Figs. 5 and 6. If fewer modes are retained, K_{ee} loses its flat spectrum. In this case, a threshhold of 99.9% gave the closest approximation of the spectrum of K_{ee} to K, while 90% gave the best approximation of the spectrum of P^a_{ee} to P^a, of those shown. Examination of the corresponding analysis ensembles and comparison to the "truth" in Fig. 1 confirm that this case produces reasonable behaviour. However, it is still clearly inferior to that obtained using the SEnKF, shown in Fig. 2b.

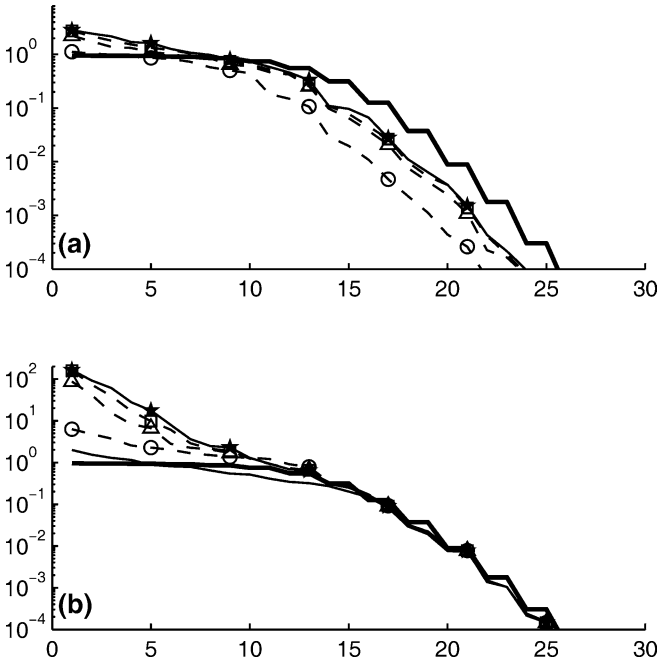

Fig. 3a,b Eigenvalue spectra of **a** the gain matrix and **b** the analysis-error covariance matrix, for the cases in Fig 1. KF (*heavy black line*), OCEnKF assuming $HA'\Upsilon^T = 0$ and no truncation (*thin black line with stars*) and with truncation of the SVD at the 90% (dashed with circles), 99% (*dashed with triangles*) and 99.9% (*dashed with squares*) points. The *lower panel* shows in addition P^a as estimated from the analysis ensemble for KF (*thin continuous line*)

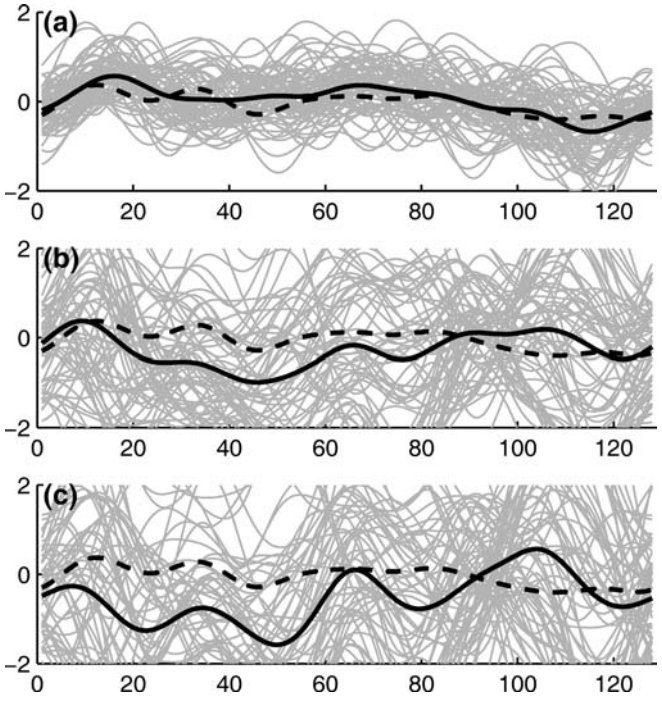

Fig. 4a–c As for Fig. 2, but for the OCEnKF analyses under the approximation $HA'\Upsilon^T = 0$ with truncation of the SVD at the **a** 90%, **b** 99% and **c** 99.9% points

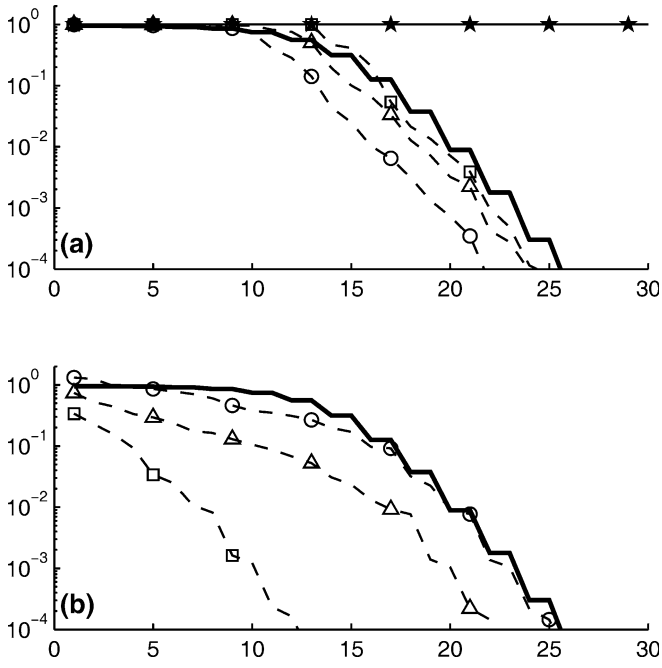

Fig. 5 Similar to Fig. 3, but without the approximation $HA'\Upsilon^T = 0$. KF (*heavy black line*), OCEnKF with no truncation (*thin black with stars*) and with truncation of the SVD at the 90% (*dashed with circles*), 99% (*dashed with triangles*) and 99.9% (*dashed with squares*) points. The curve for the untruncated analysis covariance P^a_{ee} is identically zero due to the ensemble collapse and does not appear

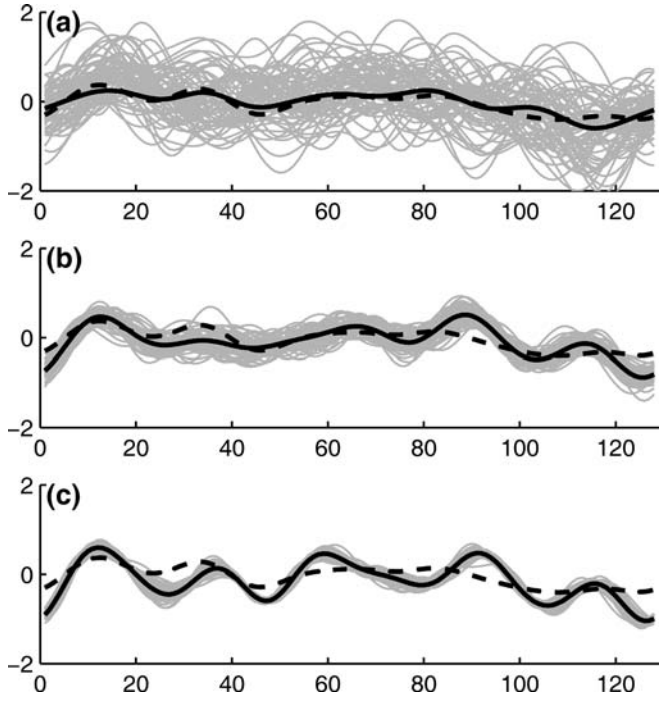

Fig. 6a–c Similar to Fig. 4, but without the assumption $HA'\Upsilon^T = 0$. The SVD is truncated at the **a** 90% (top), **b** 99% and **c** 99.9% points

Experimentation in this simple system with different ensemble sizes, number of observations, background length scales and relative magnitudes of observation and background error has shown that a universal choice of optimal truncation does not exist. Moreover, this approach is now numerically less efficient than the SEnKF, as directly calculating the inverse of $P^b_e + R$ is less expensive than proceeding via the SVD of $P^b_e + R_e$.

4.5 A sequential OCEnKF?

Several of the variants of the EnKF so far proposed have raised the possibility of a sequential approach; that is, of analyzing the observations in batches of a relatively small number at a time or even singly, with an update of the background covariance (using the ensemble) between batches. Provided the observation errors are uncorrelated between batches, this is equivalent to analyzing all the observations simultaneously.

For example, Houtekamer and Mitchell (2001) give a thorough discussion and analysis of the implementation of a sequential version of their double EnKF. A major reason for their use of a sequential algorithm was to be able to control the size of the linear system that needs to be solved at each analysis. As the cost of this part of their algorithm is proportional to the batch size cubed, this allows a significant control of the overall numerical cost. They presented results for several configurations of their system, with maximum batch sizes of 150, 300 and 1200 observations, and showed that the smaller batch size had the least numerical cost. In their experiments, the numerical cost of the largest batch size tested is less than double that of the smallest, so the sensitivity of cost to batch size is only moderate, and there is not a hard upper bound on the batch size.

The problem with rank deficiency in the OCEnKF arises when $N \leq m$, so it is clearly possible to avoid the loss of rank by analyzing the observations in batches of less than N at a time. Note that R_e for the full set of observations will not, in general, have off-diagonal entries of 0, even if R does, due to sampling error. Sequential analysis in the OCEnKF is thus equivalent to setting blocks, corresponding to observation batches, of off-diagonal elements of the full R_e to 0. This will clearly tend to increase the rank of $HP^b_eH^T + R_e$. If batch sizes of less than N are chosen, then this sequential version of the OCEnKF will produce a full-rank analysis.

However, as the batch size decreases, the cost of a direct solution of the SEnKF analysis for that batch also diminishes, and the relative numerical advantage of the OCEnKF becomes less. In fact, once the batch size is less than N, as is necessary to avoid loss of rank, $HP^b_eH^T + R_e$ for each batch will be of full rank and the numerical advantages of rank deficiency will be lost.

A sequential algorithm is inherently more expensive to parallelize than the direct, because the information from all ensemble members must be brought together after analyzing each observation batch, and interprocess communication is relatively expensive. Determining the optimum batch size depends on a trade-off between the cost of the analysis for each batch, the cost of

interprocess communication, and other factors as discussed by Houtekamer and Mitchell (2001). The OCEnKF, with its hard upper bound on batch size, will offer much less flexibility in this regard than the SEnKF.

Finally, ''buddy-check'' quality control algorithms similar to those of Lorenc (1981) are less satisfactory in a sequential setting, as an observation once used cannot be rejected, even if a significant number of observations in subsequent batches are found to disagree with it. The negative impact of this can probably be controlled in practise by not making the batch size too small. Again, a sequential version of the SEnKF offers superior flexibility to a sequential OCEnKF in this regard.

Thus, while the sequential approach to the OCEnKF may avoid the problem of collapse, there are some significant problems to overcome before it could be useful in real oceanographic or atmospheric assimilation systems, and its numerical advantages relative to a similarly sequential SEnKF would be less marked.

5 Conclusions

It has been shown analytically that the variant of the EnKF proposed by E2003 will collapse to a single member, provided that the ensemble size is not more than half the number of observations plus one. This result has been related to the spectrum of the associated gain matrix, which consists solely of 1's and 0's.

It was further demonstrated that some approximations, presented by E2003 as conducive to numerical efficiency, will prevent this collapse but at the cost of incorrectly representing both the analysis ensemble spread and the mean analysis. On the other hand, these problems do not arise in the standard version of the EnKF, in which the observation-error covariance matrix is used directly, rather than via a Monte Carlo representation.

In a crude sense, the result that better performance is attained in the SEnKF than in the OCEnKF is perhaps not surprising — after all, E2003's new scheme clearly increases the level of sampling error with which one must contend. However, the situation is more subtle than that. The (nearly) diagonal observation-error covariance matrix is a very poor candidate for Monte Carlo representation, which leads to a substantial systematic bias in its eigenvalues. This latter problem was shown to be considerably less severe for the relatively broadly structured background-error covariance matrix.

This last point may help explain an apparent peculiarity of atmospheric and oceanographic EnKF schemes. EnKF schemes for the atmosphere and ocean have produced satisfactory results with ensembles that are remarkably small, in the sense that the number of members is a small fraction of both the number of observations and the dimension of the state space. This may be because the important step is to properly represent the appropriate covariance matrices. In the atmosphere and ocean, the background-error covariance, with a relatively compact spectrum, can be represented with far fewer ensemble members than is required for the much flatter spectrum of the observation-error covariance. The arguments above suggest that this can lead to satisfactory performance of the EnKF, provided one applies the ensemble representation only to P^b, and not to R as well, as done in E2003.

Evensen (2003) claims some numerical advantages arise in his implementation of the EnKF when R is approximated by R_e, essentially because R_e is of reduced rank. While this claim is not disputed, it was shown here that it comes at a significant cost in terms of numerical accuracy. In particular, there is a marked degradation in the accuracy of the depiction of the analysis-error covariance. Since the main virtue of the EnKF is its ability to determine system state-dependent error statistics, it is suggested that the representation of R by R_e is unlikely to be worthwhile in practice.

Acknowledgements The author acknowledges useful discussions with Peter Steinle, and other participants at the EnKF workshop held in BMRC in November, 2003.

References

Anderson J (2001) An Ensemble Adjustment Kalman Filter for data assimilation. Mon Weath Rev 129: 2884–2903

Bishop C, Etherton B, Mujamdar S (2001) Adaptive sampling with the Ensemble Transform Kalman Filter, part I. Theoretical aspects. Mon Weath Rev 129: 420–436

Burgers G, van Leeuwen P, Evensen G (1998) Analysis scheme in the Ensemble Kalman Filter. Mon Weath Rev 126: 1719–1724

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

Evensen G 2003, The Ensemble Kalman Filter: theoretical formulation and practical implementation. Ocean Dynamics 53: 343–367. DOI 10.1007/210236-003-0036-9

Golub G, van Loan C (1996) Matrix computations. The Johns Hopkins University Press, Baltimore

Houtekamer P Mitchell H (1998) Data assimilation using an Ensemble Kalman Filter technique. Mon Weath Rev 126: 796–811

Houtekamer P, Mitchell H (2001) A sequential Ensemble Kalman Filter for atmospheric data assimilation. Mon Weath Rev 129: 123–137

Lorenc A (1981) A global three-dimensional multivariate statistical interpolation scheme. Mon Weath Rev 109: 701–721

Lorenc A (2003) The potential of the ensemble Kalman filter for NWP—a comparison with 4D-Var. Quart J Roy Meteorol Soc 129: 3183–3203

Whitaker J, Hamill T (2002) Ensemble data assimilation without perturbed observations. Mon Weath Rev 130: 1913–1924