Tellus (2006), 58A, 538–548
Printed in Singapore. All rights reserved

TELLUS

Comparison of error breeding, singular vectors, random perturbations and ensemble Kalman filter perturbation strategies on a simple model

By NEILL E. BOWLER*, *Met Office, Fitzroy Road, Exeter EX1 3PB, UK*

(Manuscript received 15 August 2005; in final form 2 June 2006)

ABSTRACT

An experiment has been performed, using a simple chaotic model, to compare different ensemble perturbation strategies. The model used is a 300 variable Lorenz 95 model which displays many of the characteristics of atmospheric numerical weather prediction models. Twenty member ensembles were generated using five perturbation strategies, error breeding, singular vectors, random perturbations (RPs), the Ensemble Kalman Filter (EnKF) and the Ensemble Transform Kalman Filter (ETKF). Based on normal verification methods, such as rank histograms and spread of the perturbations, the RPs method performs as well as any other method—this illustrates the limitations of using a simple model. Consideration of the quality of the background error information provided by the ensemble gives a better assessment of the ensemble skill. This measure indicates that the EnKF performs best, with the ETKF combined with RPs being the next most skillful. It was found that neither the ETKF, error breeding nor singular vectors provided useful background information on their own.

Central to the success of the EnKF is the localization of the background error covariance which removes spurious long-range correlations within the ensemble. Computationally efficient versions of the EnKF (such as the ETKF) cannot accommodate covariance localization and their performance is seen to suffer. Applying the ETKF to a series of local domains has been tested, which allows covariance localization whilst remaining computational efficient, and this has been found to be nearly as effective as the EnKF with covariance localization.

1. Introduction and motivation

The Met Office intends to develop a short-range (<3 d) ensemble prediction system based on a limited area model. This study aims to provide guidance on the development of initial conditions perturbations for that model using a simple model to test various strategies.

Ensembles provide a pragmatic approach to sampling the posterior pdf of a numerical weather prediction (NWP) model forecast. The uncertainties in the forecast are often partitioned into those arising from uncertainties in the initial conditions and uncertainties in the model formulation, and here the focus is on the former of these two.

Over the years a number of methods for the generation of perturbations to the initial conditions have been developed, such as error breeding (Toth and Kalnay, 1993, 1997), singular vectors (Buizza et al. 1993; Molteni et al., 1996), the ensemble Kalman filter (EnKF) (Evensen, 1994), perturbed observations (Houtekamer and Derome, 1995) and random perturbations (RPs) (Houtekamer et al., 1995; Du et al., 1997). Such a variety of methods has arisen because the dimension of the NWP systems is much greater than the number of ensemble members that can be afforded. In addition unbalanced perturbations (such as white noise) are unlikely to provide structures which grow into the type of structures observed as forecast error. From the existing perturbation methods it is necessary to choose one which will provide adequate performance for the proposed system. Recent studies (Wei and Toth, 2003; Buizza et al., 2005) have aimed to compare global EPS's from different centres, but it has proved difficult to disentangle the quality of the NWP model from the quality of the ensemble generation system. Nonetheless, these studies suggest that error breeding may be superior to singular vectors for short lead times. Studies on the ensemble transform Kalman filter (ETKF) have indicated that this is superior to error breeding (Wang and Bishop, 2003). Hence, it would seem that the ETKF is a suitable perturbation strategy to use for the new ensemble system. Hamill et al. (2000) compared error breeding, singular vectors and perturbed observations on a quasi-geostrophic channel model. Perhaps unsurprisingly the perturbed observation approach was close to ideal, although this

*Correspondence.
e-mail: Neill.Bowler@metoffice.gov.uk
DOI: 10.1111/j.1600-0870.2006.00197.x

approach is known to be computationally very costly. Anderson (1996) used the even simpler Lorenz 63 model to compare singular vectors, normal modes and error breeding as perturbation methods, and concluded that error breeding is best. The present study is similar to that of Hamill et al. (2000), but the focus will be on the EnKF and ETKF and whether these are superior to the more established methods of error breeding and singular vectors.

This test is based on the Lorenz 95 model (Lorenz, 1995; Lorenz and Emanuel, 1998). Care has been taken to implement a scheme where all the perturbation strategies are treated equally. The model used is considerably simpler than NWP models and therefore any results cannot be *directly* applied to NWP ensembles. However, results from these tests should provide further guidance on the perturbation strategy to be used at the Met Office.

The major difficulty in this study lies in the simplicity of the Lorenz 95 (L95) model, and whether results based on such a simple model can provide useful insight for ensembles based on NWP models. This paper is structured as follows. The L95 model is described in Section 2, the 3-D Var analysis scheme is detailed in Section 3, and the various perturbation strategies are described in Section 4. In Section 5, the perturbation strategies are compared using standard verification measures and the deficiencies of the simple model become apparent. In Section 6, the quality of the background error information provided by the ensembles is investigated, from which some useful results for NWP are derived and Section 7 contains the conclusions.

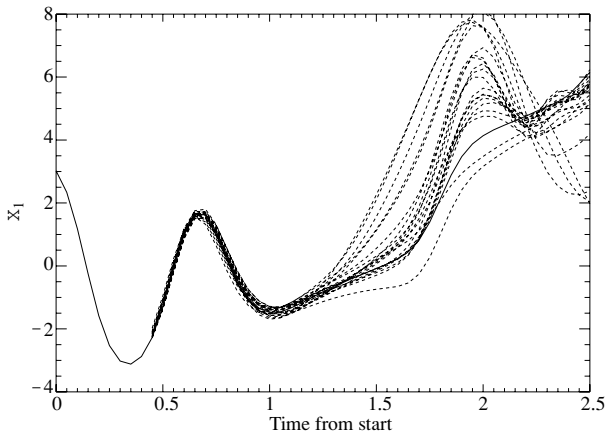

Fig. 1. Example trajectory of the first variable X_1 for the truth run (solid line) and the 20 ensemble members calculated using the EnKF (dashed lines). At all points the truth lies within the spread of the ensemble, and the ensemble spread increases with increasing lead time.

2. Lorenz 95 Model

The L95 model (Lorenz, 1995; Lorenz and Emanuel, 1998) is defined by the following system of differential equations

$$\frac{\mathrm{d}X_i}{\mathrm{d}t} = (X_{i+1} - X_{i-2})X_{i-1} - X_i + F, \tag{1}$$

where the variables X_i, $i = 1, 2, \ldots, N$, are defined on a cyclic chain such that $X_{-1} = X_{N-1}$, $X_0 = X_N$ and $X_1 = X_{N+1}$. Different behaviour is seen depending on the magnitude of the forcing, F. For small values of F the system converges to a fixed point ($X_i \to 0$ for all i when $F = 0$) whereas the system is chaotic for large values of F. In all the experiments here $F = 8$ is used, since this is well into the chaotic regime. The Runge–Kutta 4th order method was used to perform the time stepping, for intervals of $\delta t = 0.05$. One time unit is taken to be equivalent to 5 d, and a localized initial perturbation takes around 14 d (2.8 time units) to propagate around a 40 variable system (Lorenz and Emanuel, 1998).

For these experiments the system size is chosen to be $N = 300$, since this allows an ensemble of size $K = 20$, whilst still maintaining the ensemble size much less than the system size. The large size of the system means that a localized perturbation will only propagate around a small portion of the domain within the time of interest. Other configurations have been tested, including a 40-variable version with larger observation errors, and a configuration with a well observed land area. The effect of changing the configuration is discussed in Section 6. Figure 1 shows an example ensemble forecast using the ensemble Kalman filter. The solid line shows the true trajectory which the ensemble is attempting to capture, and this lies within the spread of the ensemble (dashed lines) for all lead times.

3. Analysis method

The Lorenz model was run with a time step of $\delta t = 0.05$ for 2000 time steps, the first 1000 being used to allow the system to 'spin-up'. Every time step observations were assimilated using a robust assimilation method which is independent of the ensemble perturbations. After each occasion of observations being assimilated an ensemble forecast with 20 members was run for 50 time steps. Hence a set of 1000 ensemble forecasts has been produced, with these forecasts being verified against a 'truth run'.

The observations were a perturbed version of the state of the 'truth run' with an error standard deviation of 0.1. A set of observations contains the values of all 300 state variables. The analysis update is performed using the 3-D Var method, but an exact minimization is performed using the serial updating approach (see EnKF Section 4.3). A control forecast is run, and this is updated by computing

$$\mathbf{X}^{ca} = \mathbf{X}^{cf} + \mathbf{P}^b\mathbf{H}^T(\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{Y} - \mathbf{H}\mathbf{X}^{cf}), \tag{2}$$

where $\mathbf{X}^{ca}$ is the control analysis, $\mathbf{X}^{cf}$ is the control forecast, $\mathbf{H}$ is the observation operator (the identity here), $\mathbf{P}^b$ is the background error covariance, $\mathbf{R}$ is the observation error covariance (diagonal matrix with elements equal to 0.01), and $\mathbf{Y}$ are the observations. $\mathbf{P}^b$ is assumed constant, as is normal in operational NWP

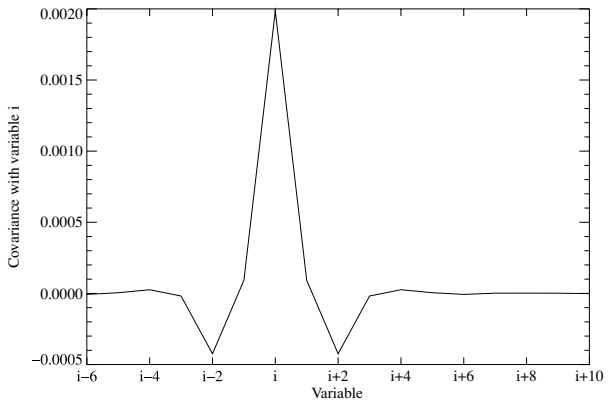

Fig. 2. Example row from the constant background error covariance matrix used in the 3-D Var analysis scheme. The matrix is symmetric and essentially zero for elements more than two variables away from the diagonal.

forecasts, and has been estimated from a long run of an EnKF ensemble. A row from this matrix is shown in Fig. 2 which is symmetric and essentially quin-diagonal.[1] $\mathbf{P}^b$ was also estimated from a long run of a singular vectors ensemble, but this resulted in an inferior analysis.

It is possible to improve this data assimilation by allowing a time-varying $\mathbf{P}^b$ (see Section 6.3 below). However, a constant $\mathbf{P}^b$ is used as this is what is used for operational NWP and it provides an analysis which is independent of the ensemble quality.

4. Perturbation strategies

4.1. Error breeding

To generate an ensemble using error breeding (Toth and Kalnay, 1993, 1997), random perturbations are allowed to grow through model integrations. At each assimilation time the perturbations are scaled down by an appropriate factor. This factor is chosen so that the spread of the perturbations to the analysis is a given constant, the size of this constant being tuned so that the forecast ensemble spread at the assimilation time is equal, on average, to the root mean square (RMS) error of the ensemble mean at that time. For each assimilation time, the first 10 perturbations are re-scaled by the appropriate factor. The second 10 perturbations are chosen to be the negative of the first 10, thus ensuring the perturbations are centred around the analysis.

The version of breeding used at NCEP also includes a regionally varying rescaling factor (Toth and Kalnay, 1997). This aims to ensure that the perturbation magnitude varies in accordance with average errors in the analysis. Since all the variables in the model are equivalent and the magnitude of the analysis error in this model does not vary with variable number only a simple global re-scaling is applied.

4.2. Singular vectors

The singular vector (SV) perturbations are found by calculating the fastest growing modes of a linearized version of the dynamics. The linearized version of the L95 model is

$$\frac{\mathrm{d}\,\delta X_i}{\mathrm{d}t} = X_{i-1}(\delta X_{i+1} - \delta X_{i-2}) + \delta X_{i-1}(X_{i+1} - X_{i-2}) - \delta X_i, \tag{3}$$

where terms of order δX^2 and above have been neglected. From this, the matrix transformation for the linearized dynamics may be obtained

$$\lim_{\delta t \to 0} \frac{\delta \mathbf{X}(t+\delta t)}{\delta t} = \mathbf{L}\,\delta \mathbf{X}(t). \tag{4}$$

$\mathbf{L}$ describes the instantaneous rate of change of a perturbation, and must be combined with $\mathbf{L}$ from different times to account for the Runge–Kutta 4th order time stepping that is used to update the model. This propagator matrix is found for each time step up to a lead time of 0.4 time units (believed equivalent to 2 d of an atmospheric model). These are multiplied together to give a final matrix $\mathbf{M}$ which describes the evolution of the perturbations over the next 0.4 time units. In accordance with the ECMWF system (Molteni et al., 1996) the right singular vectors of $\mathbf{M}$ are found. Unlike SVs for NWP the linearization of the model equations is exact, but in all other respects the SVs should be similar to those calculated operationally at ECMWF.

Ten of these SVs are chosen, using the overlap function of (Molteni et al., 1996) which aims to exclude SVs which are very similar. Since the SVs tend to be highly localized, the first 10 SVs are typically selected. In this study, the singular vectors are calculated using a Euclidean norm at initial and final times. This is a reasonable choice since all the variables of the model are equivalent and the background error covariance matrix is nearly diagonal (Fig. 2). Experiments using the analysis error covariance to provide the initial time norm (often referred to as Hessian singular vectors) produced very similar results to those presented here. The SVs are mixed using a rotation matrix, $\mathbf{S}$, so that the perturbations become

$$\mathbf{X}^a = \mathbf{VS}, \tag{5}$$

where the columns of $\mathbf{V}$ contain the SVs and the columns of $\mathbf{X}^a$ contain the perturbations to be added to the ensemble mean analysis. $\mathbf{S}$ is chosen to minimize the cost function

$$C_f = \sum_{j=1}^{K} \left[\sum_{i=1}^{N} \left(X^a_{j,i} \right)^8 \right]^{1/4}, \tag{6}$$

where $X^a_{j,i}$ is the resultant perturbation of variable i for member j. This minimization is performed to produce perturbations that are similar to the SVs, but have smaller local maxima and a more uniform spatial distribution. C_f is minimized by examining all

[1] A quin-diagonal matrix contains non-zero entries only in the five elements that are nearest the diagonal in each row.

the pairs of SVs and minimizing C_f for each pair in turn. The resultant perturbations are scaled so that the ensemble spread at the assimilation time is equal to the average RMS error of the ensemble mean forecast. The second 10 perturbations are set equal to the negative of the first 10 perturbations. This approach is chosen to closely mimic the approach taken by ECMWF (Molteni et al., 1996).

The approach used here differs from that used at ECMWF by two extra features of the ECMWF system. First, ECMWF combine the initial time singular vectors with evolved singular vectors (Barkmeijer et al., 1999). This aims to improve the ability of the ensemble to capture errors that are important in the early part of the forecast. Second, ECMWF calculate singular vectors over a number of regions, the northern and southern hemispheres and the tropics meaning that the number of singular vectors used is greater than the number of ensemble members. Both of these differences would likely improve the performance of the perturbation scheme, but are unlikely to have a large effect on the results seen later in this study (see Section 6.3).

4.3. Ensemble Kalman filter (EnKF)

The Kalman filter (Bozic, 1979; Kalman, 1960) is a data assimilation scheme which provides an optimal estimate of the true state of the system, given certain conditions. The filter works by maintaining an estimate of the true state, which is updated by a forecast of the state from the previous time and by observations. Errors in the observations, the state at the previous time and the model for propagating this state forward in time are accounted for (although here a perfect model framework is being used). The filter provides an optimal data assimilation system if the model and observation operator are linear, and all the errors are Gaussian. These conditions are not met by NWP models of the atmosphere. For NWP models the ensemble Kalman filter is a useful approximation of the Kalman filter, which allows the data assimilation to be computed on the basis of a limited size ensemble. It still assumes that all the errors are Gaussian (which will not be the case when non-linear models are used) but behaves well even when they are not.

As input the EnKF requires an ensemble of forecasts of the current state from which the background error covariance may be estimated

$$\hat{\mathbf{P}}^b = \frac{\mathbf{X}^f \mathbf{X}^{fT}}{K-1}, \tag{7}$$

where K is the size of the ensemble and the columns of $\mathbf{X}^f$ correspond to the difference between an ensemble member and the ensemble mean forecast. In common with others (Houtekamer and Mitchell, 1998) a localization function is applied to this covariance matrix

$$P^{b,l}_{i,j} = P^b_{i,j} C(|i-j|, D), \tag{8}$$

where C is the popular 5th order function of (Gaspari and Cohn, 1999). Throughout this work a covariance localization half-width of 24 variables is used, as this produces best performance for the analysis, using background error covariances from the EnKF (see Section 6.3). The localization excludes long range correlations in the background error, which are likely to be spurious. The ensemble mean analysis is calculated using

$$\bar{\mathbf{X}}^a = \bar{\mathbf{X}}^f + \mathbf{G}[\mathbf{Y} - \mathbf{H}(\bar{\mathbf{X}}^f)], \tag{9}$$

where the symbols take the meanings given in eq. (2), and $\mathbf{G}$ is known as the Kalman gain, given by

$$\mathbf{G} = \hat{\mathbf{P}}^{b,l}\mathbf{H}^T(\mathbf{H}\hat{\mathbf{P}}^{b,l}\mathbf{H}^T + \mathbf{R})^{-1}. \tag{10}$$

The inversion of the term on the right of eq. (10) can be very expensive for NWP models which use large numbers of observations. These equations are identical to eq. (2), the 3-D Var update equation, except that the background error covariance matrix $\hat{\mathbf{P}}^{b,l}$ is estimated from the ensemble rather than being static.

Equation (9) is correct for performing the update of the ensemble mean, or analysis, but to derive the correct distributions of analysis perturbations it would normally be necessary to perturb the observations, $\mathbf{Y}$. However, since $\mathbf{R}$ is diagonal it is entirely appropriate to process the observations serially (Whitaker and Hamill, 2002). In this case $\mathbf{H}\hat{\mathbf{P}}^{b,l}\mathbf{H}^T$ and $\mathbf{R}$ reduce to scalars. To update the perturbations without perturbing the observations $\mathbf{G}$ is multiplied by

$$\alpha = \left(1 + \sqrt{\frac{\mathbf{R}}{\mathbf{H}\hat{\mathbf{P}}^{b,l}\mathbf{H}^T + \mathbf{R}}}\right)^{-1}. \tag{11}$$

This algorithm performs better than the standard EnKF with perturbed observations (Evensen, 2004), and is much faster for the simple model used here.

A fixed inflation factor (Anderson, 2001) is applied to the analysis perturbations and along with covariance localization helps to ensure that the perturbations have the correct spread at the next time step. When using background error covariance information from the ensemble (see Section 6.3) the perturbations need only be inflated by 1% to ensure the correct spread. Since using a static background error covariance matrix degrades the skill, an inflation factor of 16.2% has to be used when a 3-D Var analysis scheme is used. This inflation factor is chosen to ensure that the ensemble perturbations have the right spread for a forecast with a lead time of 0.4 time units. The variable inflation factor of Wang and Bishop (2003) would have been used. However, the variable inflation factor did not behave stably, and made the results much worse. This is due to the algorithm introducing a feedback into the perturbation size, creating undesirable oscillations.

4.4. Ensemble transform Kalman filter (ETKF)

The ETKF (Bishop et al., 2001) is equivalent to the EnKF and allows for rapid calculation of the analysis perturbations. The

ETKF is approximately 10^5 times faster than the serial EnKF for current NWP models. The efficiency of the computation relies on the analysis perturbations being in the subspace of the forecast perturbations. Hence, covariance localization is not possible within the ETKF framework. The scheme centres around calculating a transform matrix, $\mathbf{T}$, which transforms the forecast perturbations according to

$$\mathbf{X}^a = \mathbf{X}^f \mathbf{T}. \tag{12}$$

The analysis perturbations for a given variable are therefore a linear combination of the forecast perturbations for that variable. Thus, the ETKF may be seen as being closely related to the error breeding method.

The transformation matrix is calculated as (Wang et al., 2004)

$$\mathbf{T} = \mathbf{C}(\Gamma + \mathbf{I})^{-1/2}\mathbf{C}^T, \tag{13}$$

where $\mathbf{C}$ and Γ are derived from the eigenvectors and eigenvalues of the matrix

$$\mathbf{E} = \frac{(\mathbf{R}^{-1/2}\mathbf{H}\mathbf{X}^f)^T(\mathbf{R}^{-1/2}\mathbf{H}\mathbf{X}^f)}{K-1}. \tag{14}$$

Since $\mathbf{X}^f$ contains the ensemble perturbations (and therefore must sum to zero) one of the eigenvalues of $\mathbf{E}$ will be zero. Γ is a $(K-1)\times(K-1)$ diagonal matrix containing all the eigenvalues of $\mathbf{E}$ except this zero eigenvalue, and $\mathbf{C}$ contains the corresponding eigenvectors, save the one relating to the zero eigenvalue. The perturbations thus generated will be orthogonal to each other in the space of the observations (Wang et al., 2004), and have the appropriate variance, as dictated by the Kalman filter formulation.

A fixed inflation factor is used to inflate the ETKF perturbations. Since the ETKF cannot support covariance localization a much larger factor is required than for the EnKF. An inflation factor of 96% was found to ensure that the perturbations were the correct size for a forecast with a lead time of 0.4 time units.

4.5. Random perturbations

The final perturbation scheme considered here is the simplest—adding white noise to the analysis. The perturbation at one point is independent of the perturbation at all other points, and conditioned according to

$$\sum_{j=1}^{K} X^a_{j,i} = 0 \qquad \text{for all } i \tag{15}$$

and

$$\sum_{j=1}^{K} \left(X^a_{j,i}\right)^2 = \sigma_n^2 \qquad \text{for all } i, \tag{16}$$

where σ_n^2 is chosen so that the spread of the ensemble is on average equal to the RMS error of the ensemble mean forecast at the assimilation time. The distribution of the noise is chosen to fit the distribution of the measured analysis error. This distribution is given in Fig. 3 and is the convolution of a Gaussian and a symmetric exponential distribution.

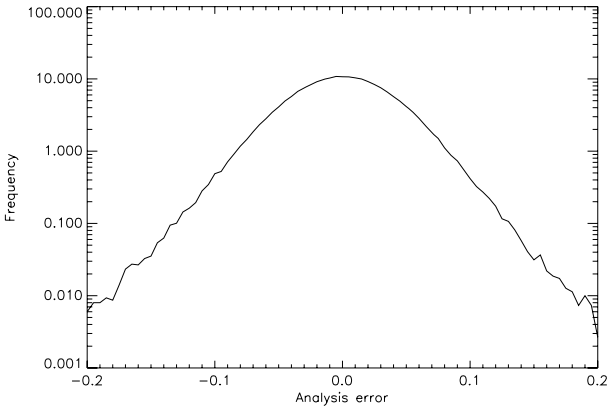

Fig. 3. Frequency distribution of the observed analysis errors for the 3-D Var system. This distribution is approximately the convolution of a Gaussian with a symmetric exponential distribution, indicating that large errors in the analysis are more common than in a Gaussian distribution.

To apply the above conditions on the perturbations, a random sample was first drawn from the appropriate distribution. In general the mean and variance of this sample do not meet the above conditions, so the mean and variance of the sample are found, and the perturbations are normalized before being applied.

5. Results

In this section, the perturbation schemes are compared using a number of different methods. The RMS spread of the ensemble, the skill of the ensemble mean and the rank histogram are used. The Brier skill score for forecasting the event $X_i > 2$ for all i has been considered, but is not presented since this yields no new information in addition to the other scores.

5.1. Skill of the ensemble mean

The greater the ability of an ensemble forecast to account for the likely errors in the forecast, the greater the skill of the mean of that ensemble forecast. Therefore, the skill of the ensemble mean can be taken as a (crude) measure of the quality of the ensemble. Figure 4 shows the RMS error of the ensemble mean forecast for the five perturbation strategies. The skill of the ensemble mean forecast for the EnKF and RP schemes are similar, with some suggestion that the EnKF is slightly superior. The ensemble mean of the ETKF is the next most skillful followed by the SV ensemble. The error breeding ensemble is substantially worse than the other methods. These rankings are likely to be largely due to the results for the spread of the ensemble versus the RMSE of the ensemble mean at longer lead times (see below).

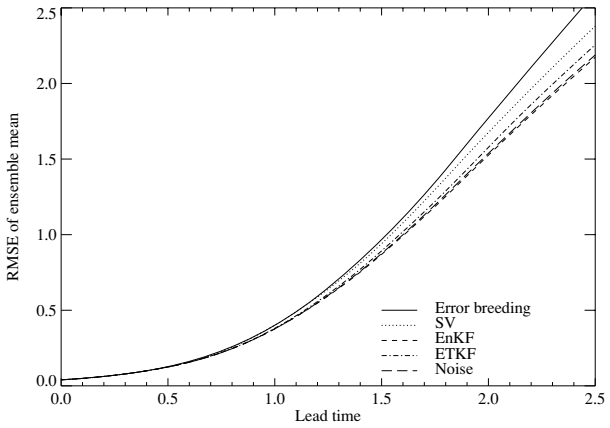

Fig. 4. RMS error of the ensemble mean. The EnKF and RPs have the lowest errors, performing slightly better than the ETKF. The SV ensemble is next best with the error breeding ensemble performing worst.

5.2. Spread–error of the ensemble

Ideally, the spread of the ensemble will be equal to the RMS error of the ensemble mean at all lead times. Figure 5 shows the RMS spread of the ensemble, divided by the RMS error of the ensemble mean. All of the ensembles have been tuned so that this ratio is equal to unity for the optimization time of the singular vectors (0.4 time units). Both the EnKF and ETKF show robust performance with the ensemble spread close to one for all lead times. There is little difference between these filters for up to lead time of 0.8 time units. After this time the performance of the ETKF ensemble deteriorates further. Initially the spread of the RPs ensemble decreases, but then levels off to a value very close to one.

Error breeding is clearly the worst performer amongst the schemes, showing growth that is much too slow. The poor performance is attributable to the perturbations being very similar, as is seen later (Section 5.4). The initial growth of spread of the SV ensemble is rapid. At around the optimization time (0.4 time units) the growth rate plateaus, after which the growth rate is too slow.

5.3. Rank histogram

If each ensemble member is an equally probable sample from the true pdf of the forecast, then the truth should have the same statistical properties as the ensemble members. For each variable one can order the ensemble members, and find where the truth lies within this order. If the ensemble forecast samples the true forecast pdf, then the truth is equally likely to lie in any of the $K + 1$ bins.

The error breeding and SVs schemes do not aim to provide a random sampling of the forecast pdf, rather they attempt to identify which errors are likely to grow rapidly. Computation of a pdf from the ensemble forecast is a non-trivial work. Nonetheless, the rank histogram can provide useful information on the quality of the ensemble schemes. In particular, if the number of times the truth lies outside the range of the ensemble is larger than would be expected, then this is a clear sign of problems with the ensemble.

The rank histogram for forecasts with a lead time of 0.4 time units (Fig. 6) indicates that the EnKF is superior to the other methods with its rank histogram being closer to flat than the

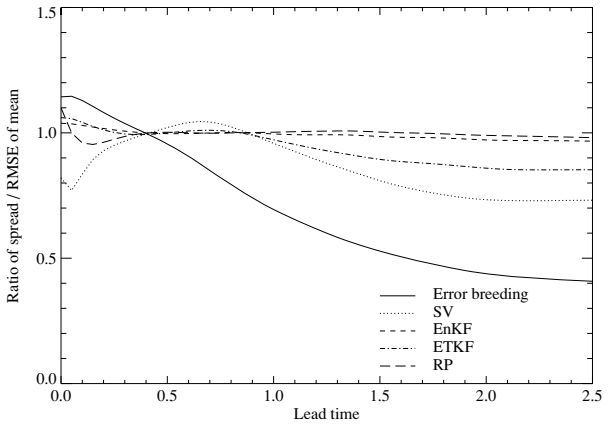

Fig. 5. Ratio of the RMS error of the ensemble mean to the spread of the ensemble. All of the schemes are tuned to have the correct spread at the optimization time of the singular vectors (0.4 time units). For the ensemble Kalman filters this ratio stays close to 1 for all lead times. However, the ensemble spread for the error breeding and singular vectors methods does not grow at the correct rate. The full EnKF appears to be slightly superior to the ETKF.

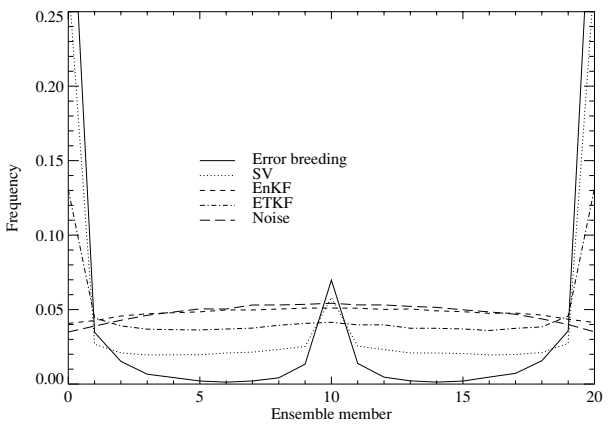

Fig. 6. Rank histogram for the five perturbation strategies, evaluated for the optimization time of the singular vectors (0.4 time units). All of the schemes have the correct spread at this time (see Fig. 5). Error breeding is clearly inferior to the other methods, and the full EnKF appears to have the best performance. The peak in the centre of the diagram for singular vectors and error breeding is due to the use of $+/-$ pairs in centring the perturbations.

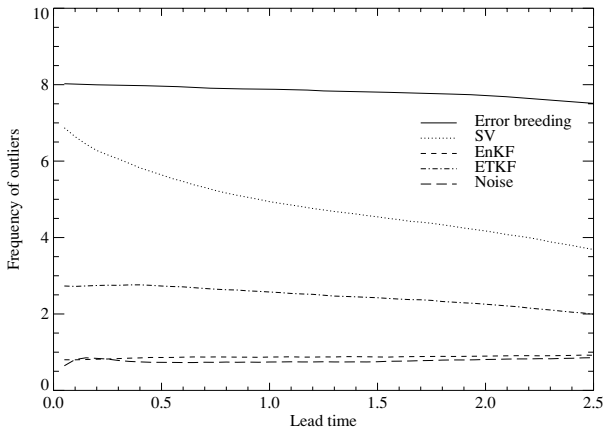

Fig. 7. The frequency of the excessive number of outliers for varying lead times. The EnKF and RP ensembles are close to perfect, and for the other methods the truth too often lies outside the range of the ensemble. As the spread of the SVs ensemble increases very rapidly initially, it appears to improve quickly with increasing lead time.

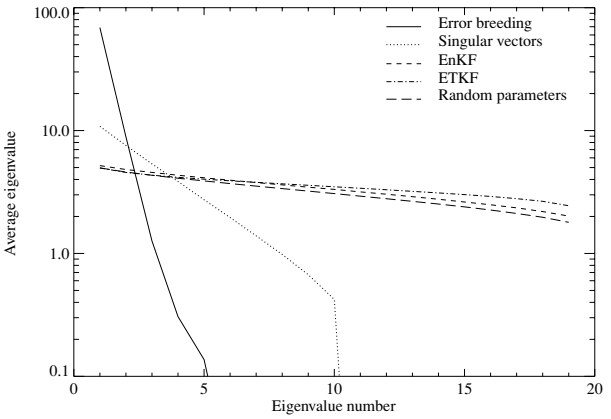

Fig. 8. Sum of the eigenvalues of the forecast perturbations in normalized observation space measured for a one time step forecast (the assimilation time). This indicates the extent to which the forecast perturbations are independent of each other. The SVs method uses $+/-$ pairs for centring the ensemble, and it can be seen that one half of these pairs is very similar to the other half. Both ensemble Kalman filters show good independence of the perturbations.

rank histograms for the other methods. The rank histogram for the random perturbations has too few outliers, indicating that the perturbations are less similar to each other than a random draw from the pdf of the forecast. This is because the ensemble mean of the perturbations is set to zero at every point (see eq. 15) which introduces a small anti-correlation between the perturbations. There is a peak in the centre of the rank histogram for the SV and error breeding perturbations which is a consequence of choosing $+/-$ pairs to centre the ensemble. The ETKF performance is superior to error breeding and SVs—even though all the ensembles have the same spread at this time.

A crucial aspect of rank histograms is the frequency with which the truth lies outside the spread of the ensemble. This frequency is a single-variable estimate of the skill of an ensemble, and the frequency of outliers, relative to the ideal value of $2/(K+1)$ is plotted in Fig. 7. Error breeding has a large number of outliers for all lead times. The SV ensemble has a quick reduction in the number of outliers, due to the initially rapid growth of the ensemble spread. The ETKF varies between having two to three times the ideal number of outliers, which is always less than for error breeding or SVs. The EnKF has slightly too few outliers, but is closer than all the other methods to the ideal frequency for all lead times. The RPs ensemble has even fewer outliers, for the reason discussed above. All of the ensemble forecasts have closer to the ideal number of outliers at longer lead times indicating that performance at short lead times is more informative than the performance at long lead times. This is a general feature of ensemble forecasts, as noted by Wei and Toth (2003).

5.4. Average eigenvalues

The average eigenvalues of the normalized forecast perturbations (Fig. 8) can provide useful information on the independence of the perturbations from one another (Wang and Bishop, 2003). These are the eigenvalues of the matrix given in eq. (14) which are routinely calculated as part of the ETKF computation. These are therefore an assessment of a single time step forecast. The sum of these eigenvalues is related to the spread of the ensemble at this time. At this point it is not the values that are of interest, but the slope the eigenvalue with ensemble member. For error breeding there is a great deal of similarity between the ensemble perturbations, indicated by the first eigenvalue being much larger than all the others. The singular vector perturbations show a much more even distribution of the eigenvalues, but the second 10 perturbations are very similar to the first 10, as might be expected. The EnKF, ETKF and RPs generate a broad spectrum of independent perturbations, with increased independence for the ETKF perturbations, presumably due to the orthogonalization that has been applied to this set.

6. Discussion

6.1. Performance of singular vectors, error breeding and random perturbations

In this study, the two most established methods of initial condition perturbations appear to be the worst performing methods. For error breeding this is due to the perturbations being very similar (as is seen in Fig. 8). It is less clear why the SVs ensemble performs poorly.

The SVs identify those structures which will grow most rapidly during the forecast. However, there is no guarantee that the analysis error projects significantly onto these vectors. The rank histogram suggests (Fig. 6) that the singular vectors are

often dissimilar to the analysis error. The SVs tend to be localized perturbations at initial time. Since a localized perturbation propagates over slightly less than 40 variables in the forecast time considered here, 10 independent localized perturbations will struggle to cover a 300-variable domain in the time. Hence, the rank histogram may be showing the fact that the SV ensemble has not perturbed all grid-points.

Changes to the configuration of the L95 model can lead to substantial improvements of the performance of the SV and ETKF ensembles. Tests have been performed with a 40-variable version of the model, using the standard observation errors (0.1), observation errors 10 times larger, and a configuration with a well and poorly observed region, simulating a land–sea contrast. For these tests the ratio of the ensemble mean error to ensemble spread was much closer to one for all perturbation methods. The SV and ETKF ensembles had far fewer outliers than for the 300-variable model. The error breeding ensemble performed much worse than the other systems. It was better for the case with small observation errors than for large, indicating that the scaling of the initial perturbations effects the quality of the error breeding ensemble. The performance of the RP ensemble did not change, still providing near-perfect forecasts. The EnKF was not tested for these configurations since the localization scale used in the 300-variable version is similar to the size of the 40-variable system.

Apart from the 'n' shaped rank histogram, Fig. 6, the RPs ensemble appears to perform extremely well. Since the average background error covariance has nearly all its weight at the origin, the errors in the analysis are typically uncorrelated, and hence white noise would be expected to closely reproduce the errors observed in the analysis. This, and the insensitivity of the L95 model to imbalance, mean that the random perturbations ensemble performs well.

For an NWP model one does not expect a RPs scheme to perform well. An NWP model has a vast state space ($\sim 10^8$ variables) which dwarfs the number of ensemble members which are normally run. Methods such as SVs and error breeding aim to sample only those initial condition perturbations that grow rapidly and therefore dominate the forecast error. For models as large as NWP models the efficient sampling of fast growing initial condition perturbations is important and RP methods are likely to lead to unreliable estimates of uncertainty, as was discussed by Lorenz (1965).

The RP ensemble performs extremely well for the all the configurations which have been tested here. Thus, one may be tempted to discard the Lorenz 95 model, concluding that it is too simple to have any relevance to NWP systems. However, by considering the background error covariance information derived from ensemble integrations (below) it is possible to construct tests for which random perturbations are considerably inferior to filtering methods, and it is on the basis of such tests that the conclusions of this work are based.

6.2. *Adding noise to the perturbations*

If it is chosen to run the EnKF without localization then one may take advantage of the considerable computational speed increase of the ETKF. However, from the good performance of the EnKF above it is evident that covariance localization brings considerable improvements to the ensemble, especially when the background error covariance estimate is used in an analysis scheme.

The effect of covariance localization is to ensure that the analysis update for distant points in the system are performed independently. Therefore, it is expected that the average background error covariance, $\mathbf{P}^b$, between distant points will be zero (as is the case, see Fig. 2). However, due to random sampling using a limited size ensemble the RMS of the background error covariance between distant points should be $1/\sqrt{K}$ times the average value of the background error variance. This is seen for the EnKF with localization (Fig. 9), which also shows a noticeable suppression of the variability in $\mathbf{P}^b$ within the localization area. It may be possible to improve the localization by fitting a correlation function to $\mathbf{P}^b$, rather than multiplying by another function. However, this would require knowing the appropriate function to which to fit.

The ETKF (or equivalently EnKF without localization) also provides a good estimate of the average background error covariance between distant points (indistinguishable from Fig. 2). However, the variability in $\mathbf{P}^b$, as measured by the RMS, is much reduced for distant points and much enhanced for nearby points, see Fig. 9. Therefore, the spurious long-range correlations introduced by the small sample size are squashed by the EnKF

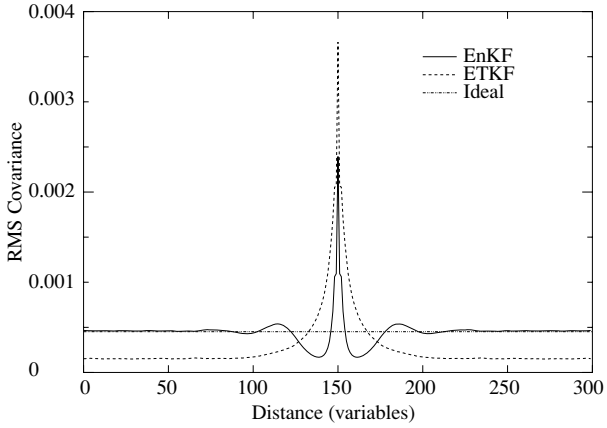

Fig. 9. The RMS of the background error covariance between a given point and variable number 150, $\sqrt{(\mathbf{P}^b)^2}$, estimated from the EnKF (solid line) and ETKF (dashed line) ensembles. The dotted line shows the value that would be expected for the EnKF at long distances, based on the updates of distant points being performed independently. The value of this quantity for distant points from the ETKF ensemble is suppressed due to spurious long-range correlations in the ensemble.

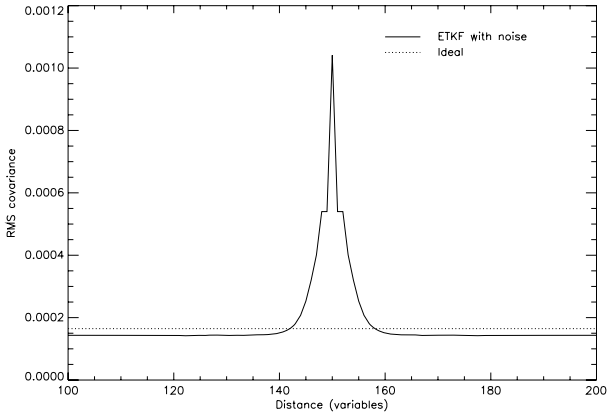

Fig. 10. The RMS background error covariance with variable 150 for the ETKF with added noise, parameters tuned as in Table 1. Also plotted is the ideal value this would take if the analysis update for distant points were independent. The value seen for the ETKF is still slightly suppressed.

processing, reducing the variability in $\mathbf{P}^b$ at long range. The effect of localization is to restore the variability in the background error covariance estimated from the ensemble. This squashing of the variability may be the effect responsible for the suppression of the variability for the EnKF with localization within the localization area.

To improve the quality of the ETKF perturbations without the cost of localization, Corazza et al. (2002) considered combining the ETKF with RPs and found considerable benefits. Since RPs are uncorrelated they may increase the variability of $\mathbf{P}^b$ at long range for the ETKF. Fig. 10 shows $\mathbf{P}^b$ for the ETKF combined with RPs and that the background error covariance derived from this ensemble can closely resemble the shape one would ideally expect.

6.3. Quality of the analysis

Given the near-perfect performance of the RP ensemble seen above it is difficult to assess the relative quality that the perturbation methods would have for an NWP model. Additionally it would be difficult to assess the benefits of combining the ETKF with RPs. However, the quality of the analysis that is produced when using the ensemble estimate of $\mathbf{P}^b$ in the 3-D Var scheme may used to measure the quality of the ensemble forecasts. The average RMS error of the analysis produced when calculated using the background error information from the ensemble is shown in Table 1. All these ensembles perform excellently when verified using the methods in section 5. However, it can be seen that the background error information provided by RPs is considerably inferior to that provided by the other methods.

Although the ETKF is incompatible with covariance localization, the $\mathbf{P}^b$ estimated by the ETKF may be localized before it is used as input to the analysis scheme. The results in Table 1 indicate that the EnKF with covariance localization is the best perturbation strategy, followed by the combined ETKF–RPs scheme. For this scheme, the RPs contribute around 7% of the variance of the total perturbations. The performance of error breeding is much improved by combining with RPs. This may explain why error breeding has not been seen to be clearly inferior to SVs for NWP models, since deterministic parameterizations with their numerous switches may provide some of the noise required to keep the bred modes distinct from each other.

It should be noted that using either the ETKF or error breeding alone did not provide adequate information and lead to the analysis irrevocably diverging from the true solution. However, if the ETKF or bred perturbations are 'refreshed' using RPs then considerable benefit is seen with an analysis having as little as 60% the error of the 3-D Var case. It is the interaction of the RPs with the dynamics that allows them to closely resemble structures seen in the analysis error. The EnKF is the only method that does not benefit from the addition of noise, with the analysis being superior to that obtained using other methods.

Combining SVs with RPs did not provide an improved analysis compared with the RPs alone. This indicates that the singular vector perturbations, even when combined in the complex way described in Section 4.2 do not provide an estimate of the analysis error. This is in agreement with the results found for NWP models (Fisher and Andersson, 2001).

Table 1. The best analysis which can be found for the various perturbation schemes when combined with the RPs. The EnKF with localization is clearly superior to the other methods. Combining singular vector perturbations with RPs produced an analysis which was inferior to using RPs alone. The analysis made when using the ETKF or error breeding alone diverged irrevocably from the truth

Method	Noise σ_n	Localization D	Mean spread	Analysis error
Static P^b	0	N/A	N/A	0.0400
EnKF	0	24	0.0173	0.0174
ETKF + RP	0.0064	7.5	0.0241	0.0249
Breeding + RP	0.0120	5	0.0277	0.0296
RP	0.0453	3.5	0.0400	0.0369

6.4. Localization and the ETKF

Another approach to improve the ETKF perturbations is to apply the ETKF over a series of local areas, similar to the local EnKF (LEKF) proposed by (Ott et al., 2004). They calculate the analysis for a given grid point as the solution of the EnKF over an area centred on that grid point. Only observations and model forecasts local to that point are considered in updating the ensemble at the point in question, so this is an effective form of covariance localization. Since the area over which the EnKF is solved only changes one grid-point at a time, the solution produced by this algorithm will vary smoothly over the model domain. However, this scheme is still potentially rather expensive—it requires that the EnKF is solved for every grid point and scales with the fourth power of the horizontal resolution.

With the ETKF, a rather helpful enhancement to the LEKF scheme presents itself. The solution of the ETKF gives a transform matrix, in which the analysis perturbations for some variable at a point are written as a linear combination of the forecast perturbations for that variable at the same point. By solving the ETKF over local areas one may produce the same solution as for the LEKF scheme. However, it is possible to solve the ETKF at a sparse set of points, and interpolate the transform matrix between these points. The difficulty is that the transform matrix calculated at one point needs to be consistent with the transform matrix calculated for all adjacent points. The determination of the transform matrix by the spherical simplex method (see above and Wang et al., 2004) may give the required consistency, since this necessarily requires that all the diagonal elements of the matrix are positive. Provided the areas over which the ETKF is solved are sufficiently large and the points are not too sparse, then the interpolation should not cause serious problems.

To demonstrate the effect of the interpolation of the transform matrix, the local ETKF scheme is run using an area size of 41 variables (very similar to a localization half-width D of 20 variables) for various numbers of points. Where the scheme is run for 300 points this corresponds to the original LEKF scheme, and where it is run for 150 points the transform matrix is calculated at every other point. The background error covariance matrix from this ensemble is calculated, localized in the standard manner, and used by the 3-D Var scheme. Table 2 gives the performance of the analysis when the local ETKF approach is used. There is little difference between the approaches for all the number of points used. If the local areas of adjacent points do not overlap, then the scheme is unstable, but it appears to work well in all other circumstances. Furthermore, since the quality of the analysis does not deteriorate with increased interpolation these results indicate that the transform matrix obtained in one region is very similar to that obtained in the adjacent regions. Given as areas for the LEKF with an NWP model are given at around 1000 km across for an ensemble of size 40 (Szunyogh et al., 2005), the local ETKF could safely interpolate between points separated by less than this distance, thus 200 local areas over the earth would probably suffice.

Table 2. Performance of the 3-D Var analysis scheme when the local ETKF scheme is used to specify the background error covariance. For all configurations local areas of 41 variables were used, and covariance localization applied to the $\mathbf{P}^b$ calculated from the ensemble. The interpolation does not appear to adversely affect the system, and the differences in performance are not believed to be significant.

Number of points	Localization D	Mean spread	Analysis error
300	20	0.0182	0.0179
100	20	0.0182	0.0178
60	20	0.0182	0.0178
20	20	0.0181	0.0177
10	20	0.0181	0.0179

7. Conclusion

Five ensemble perturbation schemes have been tested on a simple model in an environment without model error. Although these tests are on a simple model, it is hoped that they will provide guidance on which schemes are likely to provide the best ensemble perturbation strategy for short range weather forecasting. The standard verification scores indicated that RPs was at least as good as any other method. This result is not expected to be reproduced in NWP models so a different verification approach was developed, based on assessing the quality of an analysis which uses the ensembles to provide background error covariance information. These tests indicate that the ensemble Kalman filter is the best framework for generating initial condition perturbations for an ensemble. The localization of the background error covariance is central to the performance of the EnKF. Combining either the ETKF or error breeding with RPs brings considerable improvements on any of these perturbation schemes alone.

The potential problems introduced by covariance localization or adding noise does not appear to affect the results for this simple model. Although imbalance is an important issue in NWP models it has been shown (Mitchell et al., 2002) that covariance localization may be used without introducing serious imbalance. Model error must also be accounted for in NWP models, and one way of doing this is a 'stochastic physics' parameterization. Such schemes may improve the performance of the filtering perturbation methods (ETKF, error breeding) in the same way that combining these schemes with RPs has produced improved performance here. However, other authors (Mitchell et al., 2002) have found it difficult to devise a good model error parameterization.

This study has been limited to a simple model without model error and with a size of the model state much less than that used for NWP models. However, the system is sufficiently large that it is not a trivial matter to provide an accurate ensemble

forecast. This limitation on the system size may make the ETKF and EnKF appear more attractive than they would otherwise, since the size of the ensemble affects the determination of the perturbations. Model error (and its parameterization) is likely to make the perturbation schemes appear more equal (since the ensemble quality will depend sensitively on the parameterization quality).

Therefore, it is worth investigating the performance of the ETKF on atmospheric NWP models as this should prove to be a quick and effective method for generating ensemble perturbations. An EnKF which permits the localization of background error covariances is likely to be superior to the simple ETKF, and a cheap method for achieving this has been shown to exhibit good performance.

8. Acknowledgments

The author wishes to thank Andy Lawrence for providing the initial code for the ETKF from which this study has its origin. Thanks also to Craig Bishop, Xugang Wang, Ken Mylne, Jim Hansen, Martin Leutbecher, Istvan Szunyogh and Tom Hamill for their advice and comments on this paper.

References

Anderson, J. L. 1996. Selection of initial conditions for ensemble forecasts in a simple perfect model framework. *J. Atmos. Sci.* **53**, 22–36.

Anderson, J. L. 2001. An ensemble adjustment kalman filter for data assimilation. *Mon. Wea. Rev.* **129**, 2884–903.

Barkmeijer, J., Buizza, R. and Palmer, T. N. 1999. 3D-Var Hessian singular vectors and their potential use in the ECMWF ensemble prediction system. *Q. J. R. Meteorol. Soc.* **125**, 2333–2351.

Bishop, C. H., Etherton, B. J. and Majumdar, S. J. 2001 Adaptive sampling with the ensemble transform kalman filter. Part 1: theoretical aspects *Mon. Wea. Rev.* **129**, 420–436.

Bozic, S. M. 1979. Digital and Kalman Filtering. Edward Arnold (Publishers) Ltd., London.

Buizza, R., Tribbia, J., Molteni, F. and Palmer, T. 1993. Computation of optimal unstable structures for a numerical weather prediction model. *Tellus* **45A**, 388–407.

Buizza, R., Houtekamer, P. L., Toth, Z., Pellerin, G., Wei, M. and co-authors. 2005. Assessment of the status of global ensemble prediction. *Mon. Wea. Rev.* **133**, 1076–1097.

Corazza, M., Kalnay, E., Patil, D. J., Ott, E., Szunyogh, J. and co-authors 2002. Use of the breeding technique in the estimation of the background error covariance matrix for a quasi-geostrophic model. In: *Symposium on Observations, Data Assimilation, and Probabilistic Prediction*, *13–17 January 2002*, *Orlando*, *Florida* AMS, Boston, MA, 154–157.

Du, J., Mullen, S. L. and Sanders, F. 1997. Short-range ensemble forecasting of quantitative precipitation. *Mon. Wea. Rev.* **125**, 2427–2459.

Evensen, G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte-carlo methods to forecast error statistics. *J. Geophys. Res.–Oceans* **99**(C5), 10 143–10 162.

Evensen, G. 2004. Sampling strategies and square root analysis schemes for the EnKF. http://www.nersc.no/∼geir/EnKF/Publications/eve04a.pdf

Fisher, M. and Andersson, E. 2001. Developments in 4D-Var and Kalman filtering. *ECMWF technical memo no. 347* ECMWF.

Gaspari, G. and Cohn, S. E. 1999 Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.* **125**, 723–757.

Hamill, T. M., Snyder, C. and Morss, R. E. 2000 A comparison of probabilistic forecasts from bred, singular-vector, and perturbed observation ensembles. *Mon. Wea. Rev.* **128**, 1835–1851.

Houtekamer, P. L. and Derome, J. 1995. Methods for ensemble prediction. *Mon. Wea. Rev.* **123**, 2181–2196.

Houtekamer, P. L., Lefaivre, L. and Derome, J. 1995 The RPN ensemble prediction system. In: *Proceedings of the seminar on predictability* (Reading, Berkshire, UK, 1995), vol. II, ECMWF, pp. 121–146.

Houtekamer, P. L. and Mitchell, H. L. 1998. Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.* **126**, 796–811.

Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. *Trans. AMSE - J. Basic Eng.* **82**(D), 35–45.

Lorenz, E. N. 1965. A study of the predictability of a 28-variable atmospheric model. *Tellus* **17**, 321–333.

Lorenz, E. N. 1995. Predictability: a problem partly solved. In: *In Proceedings of the seminar on predictability*. Volume I, ECMWF, Reading, Berkshire, UK., pp. 1–18.

Lorenz, E. N. and Emanuel, K. A. 1998. Optimal sites for supplementary weather observations: simulation with a small model. *J. Atmos. Sci.* **55**, 399–414.

Mitchell, H. L., Houtekamer, P. L. and Pellerin, G. 2002. Ensemble size, balance, and model-error representation in an ensemble Kalman filter. *Mon. Wea. Rev.* **130**, 2791–2808.

Molteni, F., Buizza, R., Palmer, T. N. and Petroliagis, T. 1996. The ECMWF ensemble prediction system: methodology and validation. *Q. J. R. Meteorol. Soc.* **122**, 73–119.

Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E., and co-authors. 2004. A local ensemble kalman filter for atmospheric data assimilation. *Tellus* **56A**, 415–428.

Szunyogh, I., Kostelich, E., Gyarmati, G., Patil, D., Hunt, B., and co-authors. 2005. Assessing a local ensemble kalman filter: perfect model experiments with the National Centers for Environmental Prediction global model. *Tellus* **57A**, 528–545.

Toth, Z. and Kalnay, E. 1993. Ensemble forecasting at NMC: the generation of perturbations. *Bull. Am. Meteorol. Soc.* **74**, 2317–2330.

Toth, Z. and Kalnay, E. 1997. Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.* **125**, 3297–3319.

Wang, X. and Bishop, C. H. 2003. A comparison of breeding and ensemble transform kalman filter ensemble forecast schemes. *J. Atmos. Sci.* **60**, 1140–1158.

Wang, X., Bishop, C. H. and Julier, S. J. 2004. Which is better, an ensemble of positive–negative pairs or a centered spherical simplex ensemble? *Mon. Wea. Rev.* **132**, 1590–605.

Wei, M. and Toth, Z. 2003. A new measure of ensemble performance: perturbation versus error correlation analysis (PECA) *Mon. Wea. Rev.* **131**, 1549–1565.

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