Ensemble Transform Kalman Filter-based ensemble perturbations in an operational global prediction system at NCEP

By MOZHENG WEI[1]*, ZOLTAN TOTH[2], RICHARD WOBUS[1], YUEJIAN ZHU[2], CRAIG H. BISHOP[3] and XUGUANG WANG[4], [1]*SAIC at NOAA/NWS/NCEP, Camp Springs, MD, USA;* [2]*NOAA/NWS/NCEP, Camp Springs, MD, USA;* [3]*Naval Research Laboratory, Monterey, CA, USA;* [4]*NOAA-CIRES/CDC, Boulder, CO, USA*

ABSTRACT

The initial perturbations used for the operational global ensemble prediction system of the National Centers for Environmental Prediction are generated through the breeding method with a regional rescaling mechanism. Limitations of the system include the use of a climatologically fixed estimate of the analysis error variance and the lack of an orthogonalization in the breeding procedure. The Ensemble Transform Kalman Filter (ETKF) method is a natural extension of the concept of breeding and, as shown by Wang and Bishop, can be used to generate ensemble perturbations that can potentially ameliorate these shortcomings. In the present paper, a spherical simplex 10-member ETKF ensemble, using the actual distribution and error characteristics of real-time observations and an innovation-based inflation, is tested and compared with a 5-pair breeding ensemble in an operational environment.

The experimental results indicate only minor differences between the performances of the operational breeding and the experimental ETKF ensemble and only minor differences to Wang and Bishop's earlier comparison studies. As for the ETKF method, the initial perturbation variance is found to respond to temporal changes in the observational network in the North Pacific. In other regions, however, 10 ETKF perturbations do not appear to be enough to distinguish spatial variations in observational network density. As expected, the whitening effect of the ETKF together with the use of the simplex algorithm that centres a set of quasi-orthogonal perturbations around the best analysis field leads to a significantly higher number of degrees of freedom as compared to the use of paired initial perturbations in operations. As a new result, the perturbations generated through the simplex method are also shown to exhibit a very high degree of consistency between initial analysis and short-range forecast perturbations, a feature that can be important in practical applications. Potential additional benefits of the ETKF and Ensemble Transform methods when using more ensemble members and a more appropriate inflation scheme will be explored in follow-up studies.

1. Introduction

It is well known that the atmosphere is chaotic, and its predictability is severely limited by both initial and model-related errors. A feasible way to improve a single, deterministic forecast is to use ensemble forecasting. Ensemble forecasts start from a set of different states that are approximated using a finite sample of initial perturbations. However, the nature of the best method to generate these initial perturbations for an ensemble forecasting system is still under investigation.

At the European Center for Medium-Range Weather Forecasts (ECMWF), singular vectors (SVs) are used to identify the directions of fastest forecast error growth for a finite time period (Buizza and Palmer, 1995; Molteni et al., 1996). Instead of using SVs, the National Centers for Environmental Prediction (NCEP) uses bred vectors (BVs) to sample amplifying analysis errors through breeding cycles that are similar to data assimilation cycles (Toth and Kalnay, 1993; 1997). However, both SVs and BVs cannot accurately represent the true uncertainties in analysis as we expect from a good ensemble forecast system. A comparison of the performances of the ECMWF and NCEP ensemble forecast systems was described by Zhu et al. in 1996 (personal communication), and a more recent comparison can be found in Wei and Toth (2003).

Another method is the perturbed observation (PO) approach developed at the Meteorological Service of Canada (MSC) (Houtekamer et al., 1996; Houtekamer and Mitchell, 1998). The

---

*Corresponding author.
e-mail: mozheng.wei@noaa.gov

PO approach generates initial conditions by assimilating randomly POs using different models in a number of independent cycles. The initial perturbations generated by the PO method are more representative of analysis uncertainties in comparison with SVs and BVs. A comprehensive summary of the current methodologies and performance of the three ensemble forecast systems from ECMWF, MSC and NCEP can be found in Buizza et al. (2005).

In this paper, we explore a method proposed by Wang and Bishop (2003) (referred to as WB) to generate the initial perturbations for global ensemble forecasts. The method is based on an Ensemble Transform Kalman Filter (ETKF) put forward by Bishop et al. (2001). The ETKF was initially applied to the adaptive sampling problem; for example, Majumdar et al. (2001; 2002). Later Wang and Bishop (2003) showed how it could be used to generate ensemble perturbations without having to perform data assimilation, while Etherton and Bishop (2004) showed how ETKF ensemble perturbations enabled a highly efficient hybrid data assimilation scheme. Although the ETKF formulation is derived from ensemble Kalman filter theory which is used for data assimilation, as in Wang and Bishop (2003), in this study, the ETKF is only used for ensemble generation alone. In this context, the ETKF transforms forecast perturbations into analysis perturbations in a manner consistent with the Kalman filter error covariance update equation. The ETKF transformation procedure requires as input the locations and error covariances of observations. It is similar to breeding cycles in that both schemes create analysis perturbations from forecast perturbations. The observational values are used only in computing inflation factors for adjusting the magnitudes of analysis perturbations. Ensemble Transform Kalman Filter analysis perturbations are then added to the analysis field produced by the NCEP operational data assimilation system (Parrish and Derber, 1992) instead of the analysis that could be produced by ETKF-based data assimilation. The reason for using the NCEP operational analysis field rather than an analysis based on some sort of ensemble Kalman filter is because the ETKF and other related ensemble-based data assimilation schemes (described below) have not yet been proven superior to the existing NCEP system. The question of whether such ensemble-based data assimilation schemes, including ETKF, can generate a good analysis with real observations is being pursued by a few major organizations (see discussion section).

WB compared the performance of the ETKF and breeding-based ensemble forecast systems. They showed that the ETKF ensemble produces better results than the breeding method in their experimental setup. However, their experiments were conducted in a simplified environment with an idealized observation system. It would be very interesting to understand how an ETKF-based ensemble forecast system works in an operational environment with real observations. Here are some major differences between WB and our experiments. First, the two models are different; our NCEP GFS model has a higher resolution

(T126L28) than the WB NCAR CCM3 model (T42L18), and we use fewer ensemble members (10) than WB (16). Second, whereas WB approximated the full observational network with a *fixed number* of rawinsonde-like stations, this study uses the full observational network with a *highly changeable number* of rawinsonde, aircraft satellite wind and other measurements that comprise the conventional observational network for NCEP's operational data assimilation. Thus, the observational operator in WB is simplified. In fact, an accurate computation of the observational operator is one of the major challenges in an operational data assimilation system. Third, our observations can be at any level and irregularly distributed, while WB's were assumed to be at only three prespecified levels. Fourth, our observational values are real for calculating inflation factors, while WB used re-analysis data as the observations. Fifth, our observation errors vary spatially and temporally, while WB computed the RMS with re-analysis data as the observational errors. As a matter of fact, WB used only two fixed values for temperature and wind observation error variances, respectively. Sixth, in WB's comparison, the magnitude of globally averaged breeding and ETKF initial ensemble variance is similar at all initialization times, whereas in the current comparison, average initial ensemble variance is larger for the ETKF than breeding. Furthermore, an interaction between the method used to compute the inflation factor and the varying number of observations from cycle to cycle may cause the initial ETKF ensemble variance to oscillate from one initialization time to the next.

Since under the limits of a very small ensemble (two members), the ETKF becomes equivalent to the breeding technique without paired perturbations and masking, the question of ensemble size is critical in any comparison between breeding and ETKF ensemble generation techniques. Wang and Bishop's (2003) experiments showed that an 8-member ETKF ensemble was not large enough to reliably resolve even large-scale geographical fluctuations in observational density. If limited computational resources limited one's ensemble size to eight members, then one would have had to apply some sort of masking (Toth and Kalnay, 1997) technique to Wang and Bishop's ETKF perturbations to reasonably represent the effect of observational density fluctuations on forecast error variance. Wang and Bishop (2003) did not apply masking to their perturbations because they found that increasing the ensemble size to 16 members was sufficient to crudely resolve the major fluctuations in observational density present in their simulated observational network. One of the objectives of this paper is to investigate whether a relatively small 10-member ETKF ensemble with no masking can outperform a similarly small breeding ensemble with masking. The choice of 10 members is motivated by the simple fact that, currently, NCEP is running a 10-member operational breeding ensemble.

The results from our experiments offer the first test of how a small ETKF ensemble works in an environment that is close to operations with real observations. The comparative evaluations of the ETKF and breeding methods will include the impact of

observations in different spaces, such as local, observational, 2-D and 3-D grid point spaces. The perturbation growth and effective number of degrees of freedom (EDF) of the subspaces spanned by the ETKF and breeding perturbations are compared.

Although the ETKF is not used for data assimilation in this study, the method of generating analysis perturbations (not analysis fields) from forecast perturbations is based on data assimilation principles. In fact, ETKF is one variant of ensemble-based Kalman square root filters (Tippett et al., 2003). Other closely related variants of ensemble-based Kalman filters are the Ensemble Adjustment Kalman Filter (EAKF) and Ensemble Square Root Filter (EnSRF) proposed by Anderson (2001) and Whitaker and Hamill (2002), respectively. A local Ensemble Kalman Filter (LEKF) was proposed by Ott et al. (2004) (also, see Szunyogh et al., 2005). All these methods (ETKF, EAKF, EnSR and LEKF) are deterministic solutions of ensemble Kalman filters, while the PO method is a stochastic solution (Houtekamer and Mitchell, 1998; Burgers et al., 1998). Lorenc (2003) has reviewed and compared different ensemble Kalman filters (such as ETKF, EAKF, EnSR and PO method) and 4-D Var for data assimilation.

The paper is organized as follows. Section 2 provides a brief basic description of the ETKF formulation. Also in this section, the experimental setup is described together with the real-time observation data. Section 3 presents the major results of our comparison. Discussion and conclusions are given in Section 4.

2. Methodology

2.1. Basic formulation

The initial perturbations of the NCEP global ensemble forecast system are generated by a breeding method. This method is well established, widely used and well documented. A description of the operational implementation at NCEP can be found in Toth and Kalnay (1993; 1997). More results and documents are available on the NECP ensemble forecast web site at http://wwwt.emc.ncep.noaa.gov/gmb/ens/index.html.

The ETKF formulation (Bishop et al., 2001) is based on the application of a Kalman filter, with the forecast and analysis covariance matrices being represented by k-forecast and k-analysis perturbations. Let

$$\mathbf{Z}^f = \frac{1}{\sqrt{k-1}} \big[\mathbf{z}_1^f, \mathbf{z}_2^f, \ldots, \mathbf{z}_k^f \big],$$

$$\mathbf{Z}^a = \frac{1}{\sqrt{k-1}} \big[\mathbf{z}_1^a, \mathbf{z}_2^a, \ldots, \mathbf{z}_k^a \big], \qquad (1)$$

where the n-dimensional state vectors $\mathbf{z}_i^f = \mathbf{x}_i^f - \mathbf{x}^f$ and $\mathbf{z}_i^a = \mathbf{x}_i^a - \mathbf{x}^a$ ($i = 1, 2, \ldots, k$) are k-ensemble forecast and analysis perturbations, respectively. n is the number of dimensions of the state vector in model space. In our experiments, $\mathbf{x}^f$ is the mean of k-ensemble forecasts and $\mathbf{x}^a$ is the analysis from the independent NCEP operational data assimilation system. Unless stated

otherwise, the lower and upper case bold letters will indicate vectors and matrices, respectively. The $n \times n$ forecast and analysis covariance matrices are formed, respectively, as

$$\mathbf{P}^f = \mathbf{Z}^f \mathbf{Z}^{f^T} \qquad \text{and} \qquad \mathbf{P}^a = \mathbf{Z}^a \mathbf{Z}^{a^T}, \qquad (2)$$

where T indicates the matrix transpose. For a given set of forecast perturbations $\mathbf{Z}^f$, the analysis perturbations $\mathbf{Z}^a$ can be determined by solving the Kalman filter error covariance update equation

$$\mathbf{P}^a = \mathbf{P}^f - \mathbf{P}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1} \mathbf{H} \mathbf{P}^f, \qquad (3)$$

where $\mathbf{R}$ is the $p \times p$ observational error covariance matrix for p observational values used in the NCEP operational data assimilation system and $\mathbf{H}$ is the linearized observational operator mapping the forecast grid point values onto the observational points. The ETKF transformation from forecast to analysis perturbations can be expressed as $\mathbf{Z}^a = \mathbf{Z}^f \mathbf{T}$. Inserting $\mathbf{P}^f = \mathbf{Z}^f \mathbf{Z}^{f^T}$ and $\mathbf{P}^a = \mathbf{Z}^f \mathbf{T} \mathbf{T}^T \mathbf{Z}^{f^T}$ in (3), one obtains an equation for $\mathbf{T}$. Bishop et al. (2001) showed that a solution to this equation is $\mathbf{T} = \mathbf{C}(\mathbf{\Gamma} + \mathbf{I})^{-1/2}$, where $\mathbf{C}$ contains the column orthonormal right SVs ($\mathbf{c}_i$) and $\mathbf{\Gamma}$ is a diagonal matrix containing squared singular values (λ_i) of

$$\mathbf{A}^f = \mathbf{R}^{-1/2} \mathbf{H} \mathbf{Z}^f = \mathbf{U} \mathbf{\Gamma}^{1/2} \mathbf{C}^T, \qquad (4)$$

that is, $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_k]$ and $\mathbf{\Gamma} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_k)$.

Although the forecast perturbations are by definition centred about the ensemble mean, i.e. $\sum_{i=1}^k \mathbf{z}_i^f = 0.0$, the analysis perturbations produced by the ETKF defined above are not necessarily centred around the analysis ($\sum_{i=1}^k \mathbf{z}_i^a \neq 0.0$). A simple transformation that will preserve $\mathbf{P}^a$ and centre the analysis perturbations about the analysis is the simplex transformation first proposed by Purser (1996) (see, also Julier and Uhlmann, 2002; Wang et al., 2004). As derived by Wang et al. (2004), $\mathbf{C}^T$ is one of the solutions of this transformation. Hence, in this paper, the spherical simplex form of the ETKF transformation $\mathbf{Z}^a = \mathbf{Z}^f \mathbf{T} \mathbf{C}^T$ will be used to create the initial ETKF perturbations.

Since the number of ensemble members is too small compared with the nominal degrees of freedom of model state space and since the model error is neglected, the analysis error covariance is greatly underestimated by the covariance of the transformed ensemble. Therefore, it is necessary to inflate the analysis perturbations. The inflation method proposed by Wang and Bishop (2003) assumes that the *global* sum of squares of the difference between a forecast and observation at the same time does not depend on the initialization of the forecast. It also assumes that the number, quality and location of observations are similar at all analysis times. While none of these assumptions are met in an operational system, one of the aims of this paper is to see whether the ETKF can outperform breeding even when the method of defining an inflation factor is ill posed. Further details of this inflation procedure can be found in Wang et al. (2004).

2.2. Experimental setup

Our experiments run from 31 December 2002 to 17 February 2003, however, our study will focus on the 32-d period from 15 January 2003 to 15 February 2003. There are 10 ensemble members in both the ETKF and breeding-based systems. The observations used are from the conventional data set in the NCEP global data assimilation system. This conventional data set contains mostly rawinsonde and various aircraft data, and wind data from satellites. Almost all the observational operators in the conventional data set are linear (Wan-shu Wu, personal communication). Both the ETKF and breeding ensembles are cycled every 6 hr in accordance with the NCEP data assimilation system, in which new observations are assimilated in consecutive 6-hr time windows centred at 00, 06, 12 and 18 UTC. The operational breeding system at NCEP was cycled every 24 hr at the time of the experiments, and was later upgraded to a 6-hr cycle in March 2004. This is the only difference between our experimental breeding system and the NCEP operational system.

The number of observations depends on the observation and telecommunication procedures and generally changes from one cycle to the next. Detailed observations can be found at NCEP web sites, such as http://www.emc.ncep.noaa.gov/gmb/ssaha. In general, the number of conventional observations per unit surface area is larger over North America, Western Europe and South-East Asia than other regions. The variation of total number of observations over the globe at different cycles for this time period is shown in Fig. 1. As usual, the number of observations over the Northern Hemisphere is much larger than that over the Southern Hemisphere (not shown). In the following two sections, we will present the results as described in the Introduction.

3. Results from a comparison between ETKF and breeding ensembles

3.1. Impact of observations on the ensemble spread

One of the main attractions of using an ETKF ensemble generation is that it allows ensemble variance to reflect the impact of variations in observational density on analysis and forecast error variance, provided the ensemble is large enough. To measure the impact of observations on ensemble variance, we will use a total energy measure of ensemble variance. This measure is considered the most appropriate for weather forecast and data assimilation (Palmer et al., 1998). For one perturbation, the total energy is computed from winds and temperature using

$$E(i, j, k) = \frac{1}{2}\left[u^2(i, j, k) + v^2(i, j, k) + \frac{C_p}{T_r}T^2(i, j, k)\right], \quad (5)$$

where i, j, k are indices for the horizontal and vertical directions in grid point space and u, v, T are the wind components (East–West, North–South) and temperature perturbations, respectively. $C_p = 1004.0 \text{ J kg}^{-1} \text{ K}^{-1}$ is the specific heat at constant pressure for dry air and T_r is the reference temperature, following the definition used in Rabier et al. (1996), Wang and Bishop (2003) and Wei and Toth (2003).

Figure 2 shows global distributions of the energy spread of analysis perturbations and the ratios of analysis and forecast spread averaged over all levels for both ETKF (left panel) and breeding (right panel) ensembles. For the ETKF ensemble (Fig. 2a), the energy spread of analysis perturbations in the Northern Hemisphere is generally lower than that in the Southern Hemisphere, particularly in the North American and Eurasian regions, due to the larger number of observations in these regions. The lowest energy spread is shown in the tropics where the error growth is small over 6-hr intervals.

A clearer picture of the impact from observations is given by the ratio of the analysis and forecast spread. This is shown in Fig. 2c. This ratio represents the rescaling factor from the forecast to analysis spread. In North America, Asia and Europe, where there are more data, the rescaling factors are low. In the Southern Hemisphere, the values of rescaling factors in the areas which are covered by satellite data are lower than in areas which are missed by the satellites. The energy spread distributions of analysis perturbations from breeding ensembles, shown in Fig. 2b, do not show the observation impact because observations are not used. The rescaling factors in breeding are designed empirically from climatology data, with lower scaling factors in the North American and Eurasian regions where traditionally
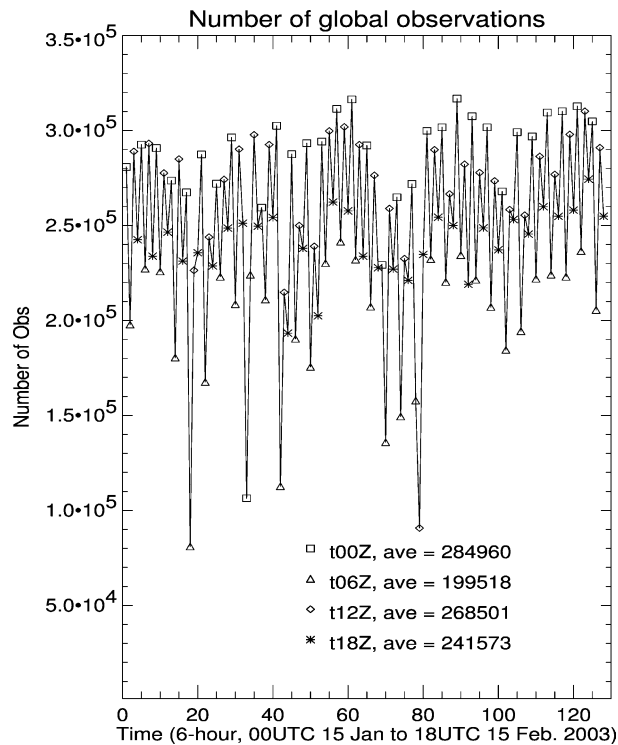

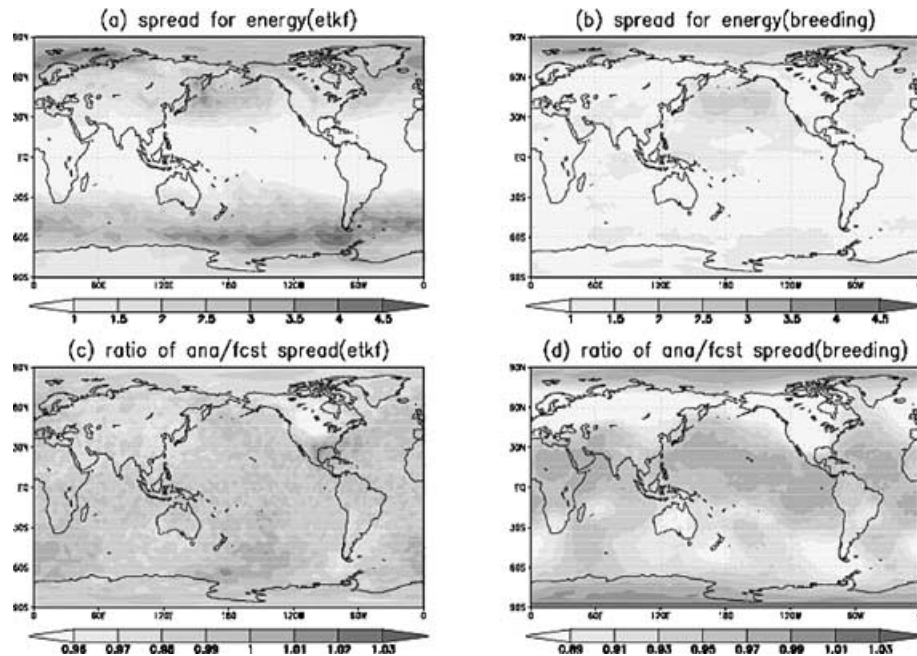*Fig. 1.* The number of observations at different cycles during the experimental period over the globe.

Fig. 2. Vertically averaged global distribution of energy spread of analysis perturbations and the ratios of the analysis and forecast spread, for both ETKF and breeding ensembles with (a) energy spread of ETKF; (b) energy spread of breeding ensemble; (c) ratio of analysis spread/forecast spread for ETKF and (d) ratio of analysis spread/forecast spread for breeding ensemble.

there are more observations. More details can be found in Toth and Kalnay (1993; 1997). It is obvious that the ETKF ensemble reflects time-dependent observations better than the breeding ensemble. Breeding initial spread is controlled by the mask which was designed to reflect the long-time average of analysis error variances. The rescaling factors in the breeding ensemble are particularly low in North America and Europe. One noticeable difference is that the ETKF rescaling factor distribution is noisier than that in the breeding ensemble. This noise is reminiscent of a similar plot shown in Wang and Bishop (2003) for their 8-member ETKF ensemble, but not of the plot corresponding to Wang and Bishop's (2003) 16-member ETKF ensemble. Thus, the noisiness of this plot suggests that with only 10 members the ETKF ensemble might benefit from some sort of masking (Toth and Kalnay, 1997). The relationship between forecast error variances and ensemble variances for both systems will be studied in detail in Subsection 3.6.

To see the vertical distributions of energy spread, we average the energy spread at all grid points at each level. In Fig. 3a, we show the vertical distributions of energy spread for the analysis (solid) and forecast (dotted) perturbations, and the rescaling factors (dashed) from both ETKF (thick lines) and breeding (thin lines) ensembles. In both ensemble systems, the analysis and forecast perturbations have relatively larger energy spreads between 600 mb and 200 mb. However, the averaged rescaling factors remain very uniform at all levels. The average values of both analysis and forecast perturbation spreads, over all levels, are larger in the ETKF ensemble than in the breeding ensemble.

They are 2.172 and 2.222 for the ETKF analysis and forecast perturbations, respectively, while for the breeding ensemble these values are 1.602 and 1.694. The generally larger spread for the ETKF is because the innovation-based inflation factor method is applied for the ETKF initial perturbations whereas no inflation is applied for the breeding whose initial perturbation magnitude is constrained by the mask only. The generally larger spread for the ETKF in this experiment setup may contribute to the fact that the individual and averaged bred perturbations grow faster than ETKF perturbations in most cases. This will be discussed in detail in the next sections.

Figure 3b shows the energy spread distribution of analysis and forecast perturbations by latitude for both ensemble systems. Unlike the distribution in the vertical direction in Fig. 3a, the latitudinal distributions of energy spread from the two ensemble systems are quite different. The result is consistent with the horizontal distributions in Figs. 2a and 2b. Generally, the ETKF ensemble has a lower energy spread in the tropics where baroclinic instability is relatively low, and a high spread near the North Pole. In the Southern Hemisphere, ETKF ensemble energy spread has a peak value around 50° South, close to the Southern Ocean track region. In contrast, the breeding ensemble has a lower energy spread mainly in the Southern Hemisphere; in particular, it has a minimum in the Southern Ocean storm track area. The failure to show higher spread in this region by the breeding ensemble is related to the mask imposed on the system (Toth and Kalnay, 1997). The results indicate that the mask used by the breeding ensemble system needs to be improved. A more
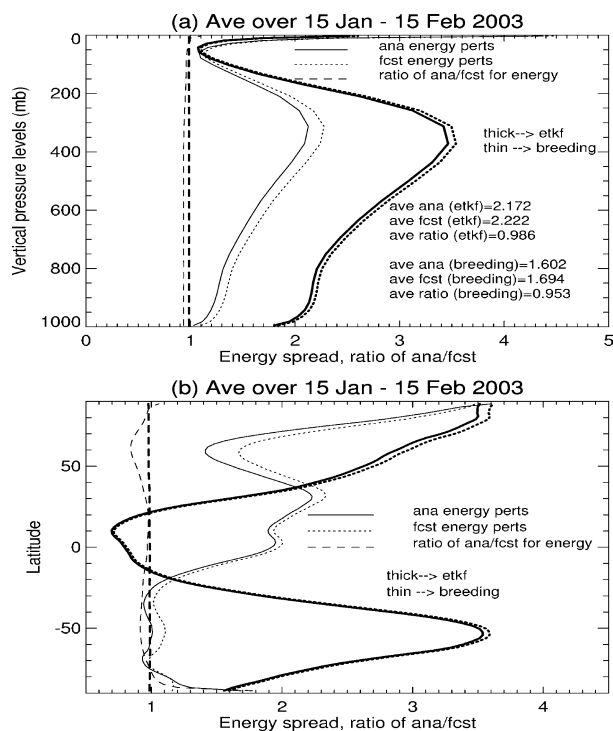
Fig. 3. Energy spread distributions of ETKF (thick) and breeding (thin) ensemble perturbations (solid: analysis; dotted: forecast). The ratio of analysis/forecast perturbations is indicated by the dashed line. All the values are averaged over the period 15 January–15 February 2003, with (a) vertical distribution as a function of pressure; (b) distribution by latitude.

accurate time-dependent mask can be built from the analysis error variances generated by a mature operational data assimilation system like NCEP 3-D Var.

3.2. Impact of WSR data

Having studied the impact from a large number of observations in the above subsection, we will look for signals from a small number of observations. Several days of Winter Storm Reconnaissance (WSR) data will be used to see if there is any influence from WSR data.

To test the impact of observations, we reran the ETKF experiments with slightly different observation data at particular times. In the new experiments, we removed the WSR data at 00 UTC on 19, 26, and 31 January and 01, 03, 08, and 09 February 2003. Details about the 2003 WSR data can be found at http://wwwt.emc.ncep.noaa.gov/gmb/targobs/target/wsr2003.html. Each experiment started from the same initial conditions as the original experiments for the previous cycle (i.e. 6-hr earlier). The new analysis perturbations on these seven days, at 00 UTC without WSR data, will be compared with those using the WSR data. On each day at 00 UTC, there are about 20 observations. Thus, in each of the seven cases, the difference between the experiments without and with WSR data will reflect

the impact of those 20 observations only. The average results of these seven cases are shown in Fig. 4.

Figure 4 shows the differences between the two experiments without and with WSR data for the vertically averaged analysis spread for temperature (Fig. 4a) and wind (Fig. 4b). The differences in the ratios between analysis and forecast spread from the two experiments are shown in Figs. 4c and 4d for temperature and wind, respectively. The black crosses indicate the locations of WSR data. It is clear that when WSR data are removed, analysis perturbations are larger over the region where the WSR was taken. Indeed, WSR data reduced the ensemble analysis variance by 1–2% for these seven cases with just a 10-member ensemble. These results demonstrate how increasing the observational density decreases ETKF ensemble variance. Note that in some areas outside the WSR data region, primarily near the equator, there is some noise. Convection near the tropics is more active than in other regions, and any differences, including slightly different initial conditions, which might come from the global model integration scheme will amplify quickly. An ensemble with large number of members may limit this kind of behaviour.

3.3. Variance distribution and the effective number of degrees of freedom of perturbations

Wang and Bishop's (2003) results indicated that the ETKF maintains significant variance in a substantially larger number of directions than breeding. Here, we investigate this hypothesis for the case of a small 10-member ensemble and real-time observations.

The forecast and analysis covariance matrices in normalized observational space are $\mathbf{A}^f \mathbf{A}^{f^T}$ and $\mathbf{A}^a \mathbf{A}^{a^T}$, respectively, where $\mathbf{A}^a = \mathbf{R}^{-1/2} \mathbf{H} \mathbf{Z}^a$ and $\mathbf{A}^f$ are defined in Section 2. The variances in different eigendirections are represented by the corresponding eigenvalues of the covariance matrices. Figures 5a and 5b show the averaged eigenvalues of $\mathbf{A}^f \mathbf{A}^{f^T}$ (a 6-hr forecast covariance matrix in normalized observational space) over the 32-d test period for the ETKF ensemble and breeding ensemble, respectively. In both schemes, there are only nine independent directions out of 10-ensemble members since the initial perturbations are centred around the analysis.

Figures 5a and 5b show that, as in Wang and Bishop (2004), the eigenvalue spectrum of the ETKF ensemble is significantly flatter than that of the breeding ensemble, if all nine nonzero eigenvalues are considered. However, the last four eigenvalues of the breeding ensemble are close to zero only because, by construction, the breeding ensemble is initialized with five pairs of identical but oppositely signed initial perturbations. As such, it is appropriate to note that the 1st and 5th ETKF eigenvalues are, respectively, 3.26×10^4 and 2.2×10^4, while the 1st and 5th eigenvalues of breeding are, respectively, 4.4×10^4 and 0.8×10^4. Hence, even when only the first five eigenvalues are considered, the eigenvalue spectrum of the ETKF is considerably
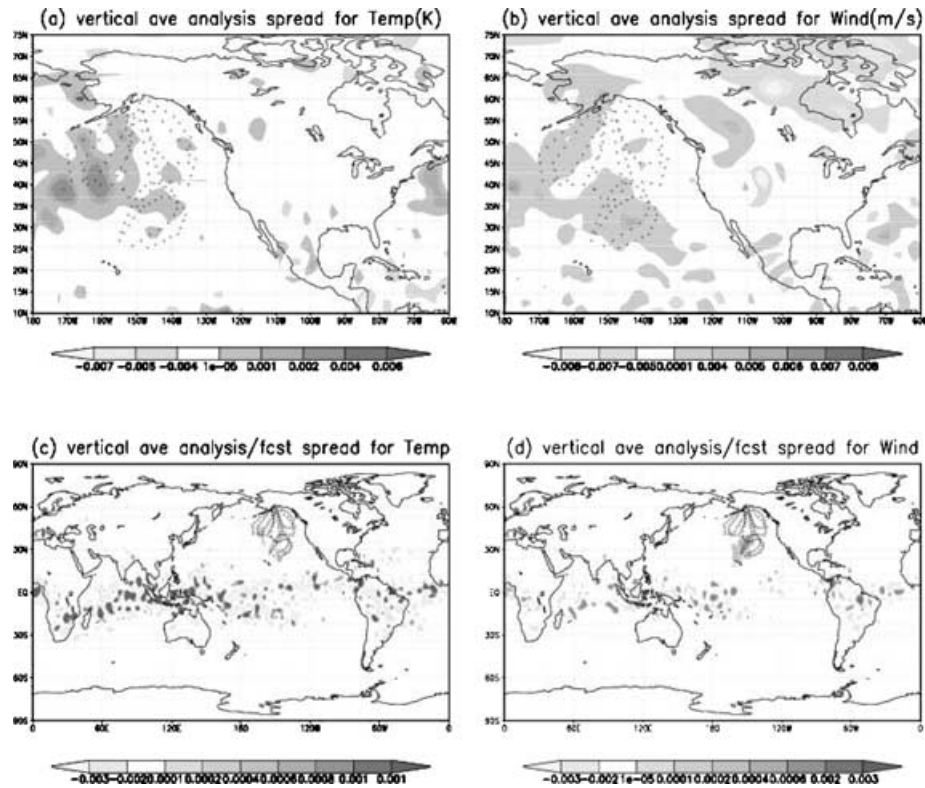
Fig. 4. The difference of vertically averaged analysis spread for (a) temperature and (b) wind, between two experiments with and without WSR data for ETKF system. The difference in analysis/forecast spread for (c) temperature and (d) wind, between the same two experiments.

flatter than that of the eigenvalue spectrum from the operational breeding scheme. Presumably, the reason for this difference is that Kalman filter error covariance update equation used by the ETKF accounts for the fact that the factor by which good data assimilation schemes reduce errors in any given direction is an increasing function of the error in the direction. Consequently, the ETKF procedure of transforming forecast perturbations into analysis perturbations explicitly flattens the eigenvalue spectrum.

A quantitative measure of the flatness of the spectrum is the number of degrees of freedom of the subspace spanned by the ensemble perturbations. Here, we use the dimension described in Patil et al. (2001). It was called the bred dimension by Patil et al. (2001), because the authors studied the subspace spanned by the BVs in their paper. A similar definition was used by Bretherton et al. (1999), where it was called the effective number of spatial degrees of freedom. It was called the Ensemble Dimension (E dimension) by Oczkowski et al. (2005) since it was used to measure the subspace of ensemble perturbations. Unlike the matrix rank that counts the number of nonzero singular values, this measure takes account of the relative values of variance in different directions, and removes the ambiguity of small nonzero variances due to, say, computing errors. We believe this definition is useful in measuring the dimensions of subspaces spanned by any vectors, not just ensemble perturbations. In this paper,

we call it the EDF of subspace spanned by the ensemble perturbations.

Figure 5a shows that in normalized observation space the EDF of the subspace spanned by the 10 ETKF ensemble forecast perturbations is 8.90, due to the variation of variances in different directions. It should be noted that the rank of the forecast covariance matrix is 9 when the relative variance values in different directions are not considered. The same time mean variance along different directions in the same normalized observational space for BVs is also computed. This is shown in Fig. 5b. As expected, the variances are overwhelmingly in the first five BVs, and one half of the BVs have variances close to zero. Hence, the ETKF spectrum is much more evenly distributed. The EDF in bred vector space is 5.89, which is much lower than that in the ETKF implementation. The main reason for this low dimensionality of bred vector space is that, as mentioned earlier, in the NCEP operational ensemble forecast system the initial BVs are in pairs, i.e. a plus/minus strategy was implemented (Toth and Kalnay, 1993; 1997). The same strategy was also employed in the ECMWF ensemble forecast system, where 25 SVs were added to and subtracted from the analysis in pairs to make 50 perturbed members (Molteni et al., 1996). It is expected that the EDF of the subspace spanned by initial SVs is also reduced by half. Since 28 September 2004, ECMWF has used multivariate Gaussian sampling from SVs to construct initial perturbations.
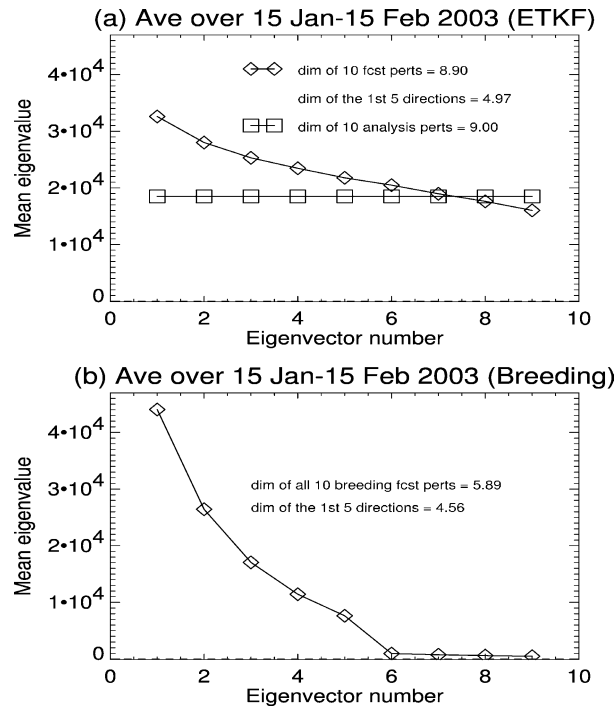
(a) Ave over 15 Jan-15 Feb 2003 (ETKF)

dim of 10 fcst perts = 8.90
dim of the 1st 5 directions = 4.97
dim of 10 analysis perts = 9.00

(b) Ave over 15 Jan-15 Feb 2003 (Breeding)

dim of all 10 breeding fcst perts = 5.89
dim of the 1st 5 directions = 4.56

Fig. 5. The averaged variance distributions along different eigendirections of forecast (diamond) and analysis (square) covariance matrices in the normalized observational space for the (a) ETKF ensemble and (b) breeding ensemble. Also, shown diagrams are the EDF values of 5 and 10 perturbations.

They are going to compare the results from both paired and non-paired strategies (Martin Leutbecher and Roberto Buizza, 2005, personal communication).

It is true that by using paired initial perturbations in an ensemble forecast system, like the NCEP and earlier ECMWF systems, EDF values of the ensembles are reduced by half. If we want to use an ensemble forecast system to produce a background covariance matrix for a data assimilation system, then the ensemble with a larger EDF value is surely better. However, the EDF is just one of many measures that have been used to verify ensemble perturbations. The compromise of EDF does not necessarily reduce the ensemble performance in other aspects. For instance, earlier experiments in the 1990s carried out at both ECMWF and NCEP showed that for the same number of members, the anomaly correlation score, which is the most frequently looked at by forecasters, were generally higher for paired ensembles than unpaired (Zoltan Toth and Roberto Buizza, 2005, personal communication). This was probably the main reason why both NCEP and ECMWF chose to use paired perturbations. In another comparison experiment, we carried out in Wei and Toth (2003), it was clearly shown that the PECA (perturbation versus error correlation analysis) value increases with the number of ensemble members in both NCEP and ECMWF operational paired systems. At NCEP, we also had results showing that other scores can be increased with a larger number of members in either the

paired or unpaired systems. These scores include ranked probability skill score, Brier skill score, ROC area and economic values. In the following, when the EDF is computed we may show the results for both 10 and 5 members from each system.

If we consider only the normalized observation subspace spanned by the first five directions with the largest variances, the EDFs are 4.97 and 4.56 for the ETKF and breeding perturbations, respectively. Thus, under this measure, the difference between the EDFs of the two systems are not as large as when we consider all 10 members.

Because the Kalman filter error covariance update equation can be explicitly shown to whiten (or flatten) the eigenvalue spectrum in normalized observation space, the EDF of the ETKF analysis perturbations has the maximum value in observational space. In grid point space, we expect the EDF values of perturbations to be smaller. Therefore, the impact of observations can be seen from a comparison of EDF values in observational and grid point spaces.

Figure 6a shows the temporal evolution of the EDF's for 10 three-dimensional temperature perturbations in grid point space over 32 d. The EDFs of the subspaces spanned by the 6-hr forecast perturbations are 8.4 and 5.8 for ETKF and breeding, respectively. Figure 6a also shows that the temporal variation of EDF over the 32-d period is greater for breeding than for the ETKF. In particular, around day 25 the breeding EDF decreases by about 20% compared with earlier values, while the ETKF EDF is largely unchanged.

Since observations are so irregularly distributed along the vertical levels and there are very different numbers of observations at different levels, the impact from different numbers of observations at different levels on the perturbation structures in the ETKF system can be studied by looking at the vertical distribution of EDF. Figure 6b shows the vertical distribution of time-averaged EDFs computed for 10 two-dimensional (horizontal) perturbations. The figure shows that while the two-dimensional EDFs of the breeding perturbations increase significantly from 900 mb to 400 mb, the corresponding ETKF EDFs show much less variation between 900 mb and 400 mb. The figure also shows that while the EDF of the breeding perturbations increases significantly between the analysis and the 6-hr forecast time, the ETKF EDFs shows a very slight decrease between analysis and forecast time.

As discussed earlier, it is impossible for the paired breeding scheme to have an EDF greater than 5 at the initial time. As such, it is of interest to consider EDF results when only five perturbations are considered. Figures 6c and 6d show that the EDFs of the two schemes are much closer in this case. They also show that even when only five perturbations are considered, the temporal and vertical variation of the EDF is greater for breeding than the ETKF.

In the following, we will look at the local EDF distributions at grid points for different pressure levels. Using the method described by Patil et al. (2001), we calculate the EDF of
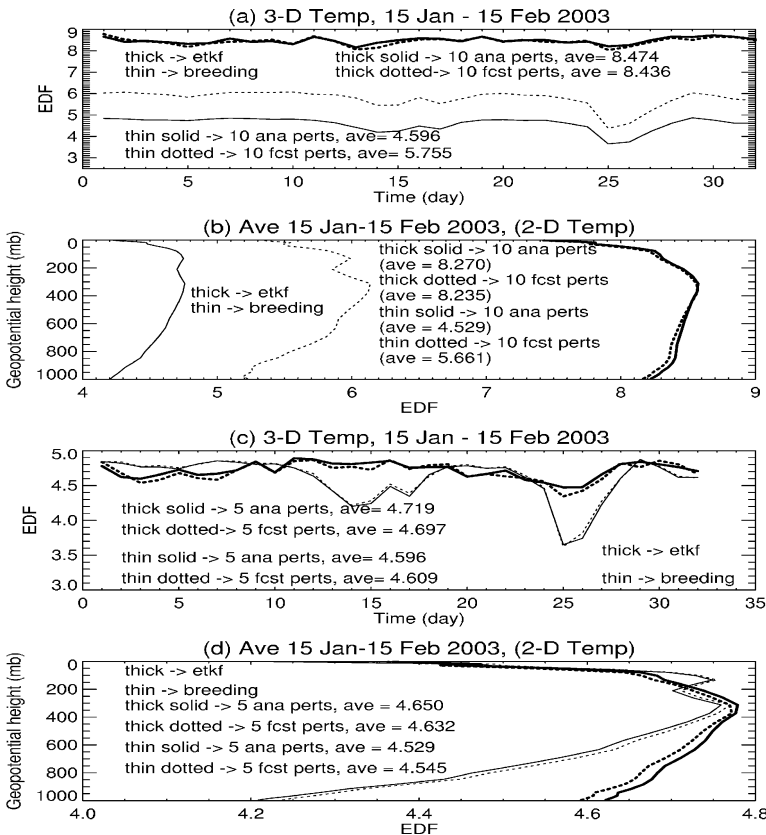
Fig. 6. The EDF of subspace spanned by temperature perturbations (solid: analysis; dotted: 6-hr forecast) from ETKF (thick) and breeding (thin) ensembles for (a) EDF of subspace spanned by 10 perturbations in three-dimensional grid point space, at different cycles during the period of experiments; (b) EDF of subspace spanned by 10 perturbations in two-dimensional grid point space at each pressure level; (c) as in (a), but for five perturbations; (d) as in (b), but for five perturbations.

horizontal subspaces spanned by the five analysis perturbations from each ensemble that covers only $(2L + 1)(2L + 1)$ horizontal grid points, where L is the number of grid points near the central points in each direction. The EDF value from this local subspace is defined as the EDF of the central grid point. The EDF distribution at each level can be calculated by moving the central grid point. The local EDF distribution measures the extent to which the ensemble perturbations are independent in the selected region. Perturbations constructed from different regions with different domain sizes (i.e., different values of L) will have different EDF values. Hence, the local EDF distributions give us
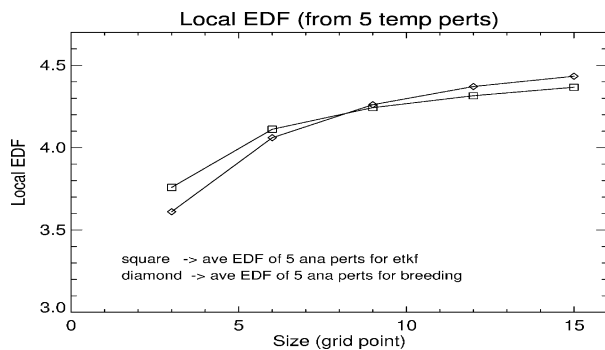

Fig. 7. Local EDF for different numbers of grid points for average local EDF as a function of L, which is described in the text.

information about how evenly the variances are distributed along different directions, and to what extent these ensemble perturbations are truly independent in that region. EDF was extensively used by Oczkowski et al. (2005) and Szunyogh et al. (2005), who related the local EDF distribution to local error growth and predictability in their data assimilation and predictability studies of the atmospheric system.

We note that the local EDF depends not only on the number of grid points we choose, but also the number of perturbations we use. In our experiments, the numbers of perturbations are the same for the two systems. To see the dependency of local EDF on the number of grid points, we carry out experiments for $L = 3, 6, 9, 12, 15$. Figure 7 shows the average EDF values over all three-dimensional grid point spaces for the five experimental cases for both ETKF (square) and breeding (diamond) ensembles. For perturbations in smaller areas, ETKF ensemble perturbations have higher degrees of freedom than the breeding perturbations, however, the bred perturbations have the advantage from $L = 9$ to $L = 15$. Since Figs. 6c and 6d show that the ETKF has higher EDF when all horizontal grid points are considered, it is clear that at some point between $L = 15$ and the L value that covers the globe, the ETKF's EDF must again exceed breeding's EDF. The results show that perturbations generated by both breeding and ETKF ensembles have different local EDF values for different local perturbations.

3.4. Amplification of perturbations

WB's results indicated that the growth rate of the most rapidly growing linear combination of ETKF perturbations significantly exceeded that of the corresponding optimal combination of breeding perturbations. Later experiments showed that the growth rates of perturbations in their global model were highly sensitive to the initial amplitude of the perturbations. In particular, it was found that perturbation growth rate increased as the size of the initial perturbations was diminished. While in WB the breeding technique was constructed so as to ensure that the breeding perturbations had about the same global amplitude as the ETKF perturbations, in the experiments reported here, the ETKF perturbations have a significantly larger amplitude than the breeding perturbations (see Fig. 3). Despite this discrepancy, it is of interest to compare growth rates between the two sets of ensemble perturbations. In addition, WB never compared the growth rates of individual perturbations from the two systems, but it is of considerable interest to measure this growth. The maximum amplification factor (AFs) from a linear combination of perturbations is calculated using a method similar to WB and Bishop and Toth (1999).

Figure 8 shows the AFs for different forecast lead times averaged from 00 UTC January 15 to 00 UTC 15 February 2003. The
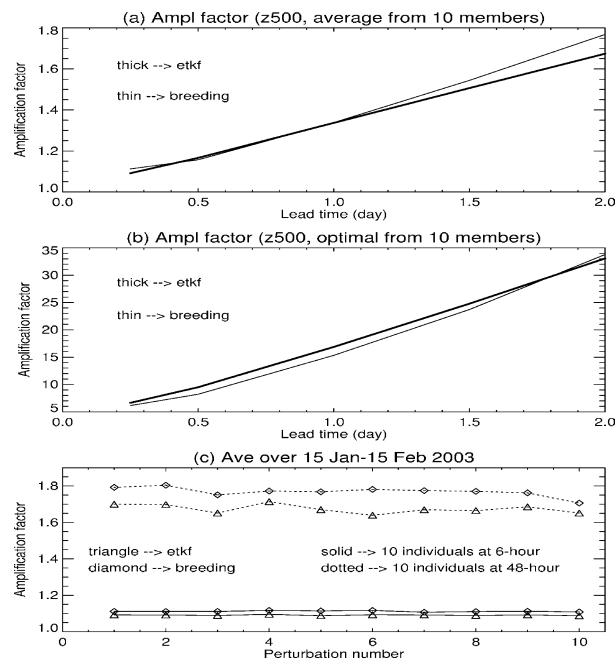

Fig. 8. Amplification factors of ensemble perturbations for (a) the average AF from 10 perturbations as a function of lead time (thick: ETKF; thin: breeding); (b) the maximum AF of optimally combined orthogonal perturbations from 10 original perturbations (thick: ETKF; thin: breeding); (c) the AF of 10 individual perturbations for two forecast lead times (solid: 6-hr; dotted: 48-hr; triangle: ETKF; diamond: breeding).

Table 1. Amplification factors of 500-mb geopotential height at 6-hr forecast lead time

Average AF for all individual perturbations				
	GL	TR	NH	SH
Breeding	1.112	1.282	1.105	1.103
ETKF	1.091	1.621	1.096	1.082

AFs are computed for both the individual perturbations and optimally combined orthogonal perturbations from both the ETKF and breeding-based systems. Figure 8a shows the average AFs for 10 perturbations at 500-mb geopotential height (thick: ETKF; thin: breeding) lines. It is clear that the average AF from individual perturbations in the breeding ensemble is larger than that from the ETKF ensemble for both shorter and longer lead times. As discussed below, we suspect the lack of ETKF growth is probably linked to the fact that the initial ETKF are significantly larger than initial breeding perturbations. We only show results out to 2 d, since the calculation of AF for optimally combined orthogonal perturbations assumes the perturbations are linear.

Shown in Fig. 8b are maximum AFs of the optimally combined orthogonal perturbations from the 10 original perturbations of both systems as a function of lead time. The AF of the ETKF ensemble is larger than that of the breeding ensemble for forecast lead time up to 1.8 d. While the individual ETKF perturbations grow slightly slower than the breeding, the maximum AF from the optimally combined orthogonal perturbations from 10 ETKF perturbations is larger than that from the breeding ensemble. This is related to the fact that the EDFs of the subspace spanned by the 10 ETKF perturbations are much larger than that from breeding ensemble (Figs. 6a and 6b) due to simplex and pairing schemes used in the two systems and also the whitening effect of the error covariance update equation used by the ETKF. For instance, if five members are chosen from each system, then the EDFs of subspace spanned by the five perturbations are similar for both the ETKF and breeding systems as shown in Figs. 6c and 6d, and the maximum AF of optimally combined perturbations is larger for breeding than ETKF (not shown). To see the growth rate of each individual perturbation from the two systems, we show the AF for each perturbation at 6-hr (solid) and 48-hr (dotted) lead times in Fig. 8c. At these two lead times, each breeding perturbation has a larger AF than the corresponding ETKF perturbation.

A likely reason for the individual bred perturbations having a larger AF than the ETKF perturbations for 500-mb geopotential height is that as mentioned previously, the AF is related to the initial perturbation size. The ETKF perturbations have a much larger spread below 150 mb (see Fig. 3a) than bred perturbations. This is also one of the reasons that ETKF perturbations have lower AF values, as shown in Fig. 8a. To demonstrate this, we compute the AFs of perturbations from both systems for different
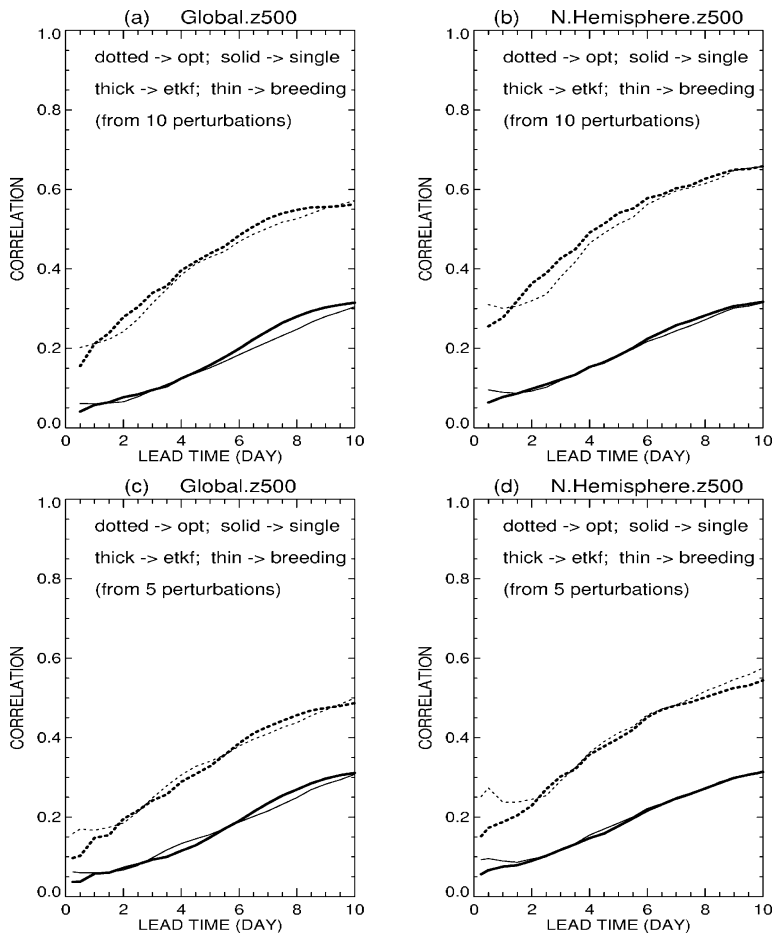
Fig. 9. The PECA values for ETKF (thick) and breeding (thin) ensembles from 10 (a and b) and 5 perturbations (c and d). Shown in dotted and solid lines are PECA from the optimally combined perturbations and the average PECA from individual perturbations.

regions. Figure 3b shows that the initial spread of ETKF perturbations is much larger than bred perturbations globally and in the Northern and Southern Hemisphere regions, but much smaller in the tropics. We then compare the AF values of perturbations from the two systems for 6-hr lead times in these different regions. Table 1 lists the average AF values of all individual perturbations for both ensemble systems in all these regions. In the global, Northern and Southern Hemisphere regions where the ETKF ensemble has a larger spread, the AFs of bred perturbations are larger. However, in the tropics where the ETKF has a smaller initial spread, the AFs of ETKF perturbations are larger.

3.5. *Representing forecast error covariance and reliability*

One measure of the performance of initial perturbations in ensemble forecasting is a direct comparison of the ensemble perturbations with the forecast errors. In this subsection, we use PECA to study the correlation between ensemble perturbations and forecast errors, as described in Wei and Toth (2003).

The PECA values from 10 perturbations for the two ensemble systems (thick: ETKF; thin: breeding) for the global and

Northern Hemisphere regions are displayed in Figs. 9a and 9b, respectively. Shown in dotted and solid lines are the PECA for the optimally combined perturbations and the averaged PECA from individual perturbations. At short forecast lead times breeding perturbations still have the advantage, but ETKF perturbations have higher PECA values than breeding perturbations after day 1. Perturbation versus error correlation analysis values for ETKF are increased more than for breeding compared with the 5-member results (Figs. 9c and 9d). This is particularly clear for PECA from optimally combined perturbations (dotted lines). This is also related to the fact that EDF in the 10-member ETKF ensemble is much higher than that in the 10-member breeding ensemble. The same results for five chosen perturbations for the two systems are shown in Figs 9c and 9d. For short forecast lead times (6 to 24 hr), bred perturbations have higher PECA values than the corresponding ETKF perturbations. If we consider data from only every 5th day as independent (not shown), the PECA values for the breeding method in the global and Northern Hemisphere domains for the 6- and 12-hr lead times are higher than for the ETKF method at the 90% (or higher) statistical significance level. The breeding and ETKF systems show similar PECA values beyond a 24-hr forecast lead time.
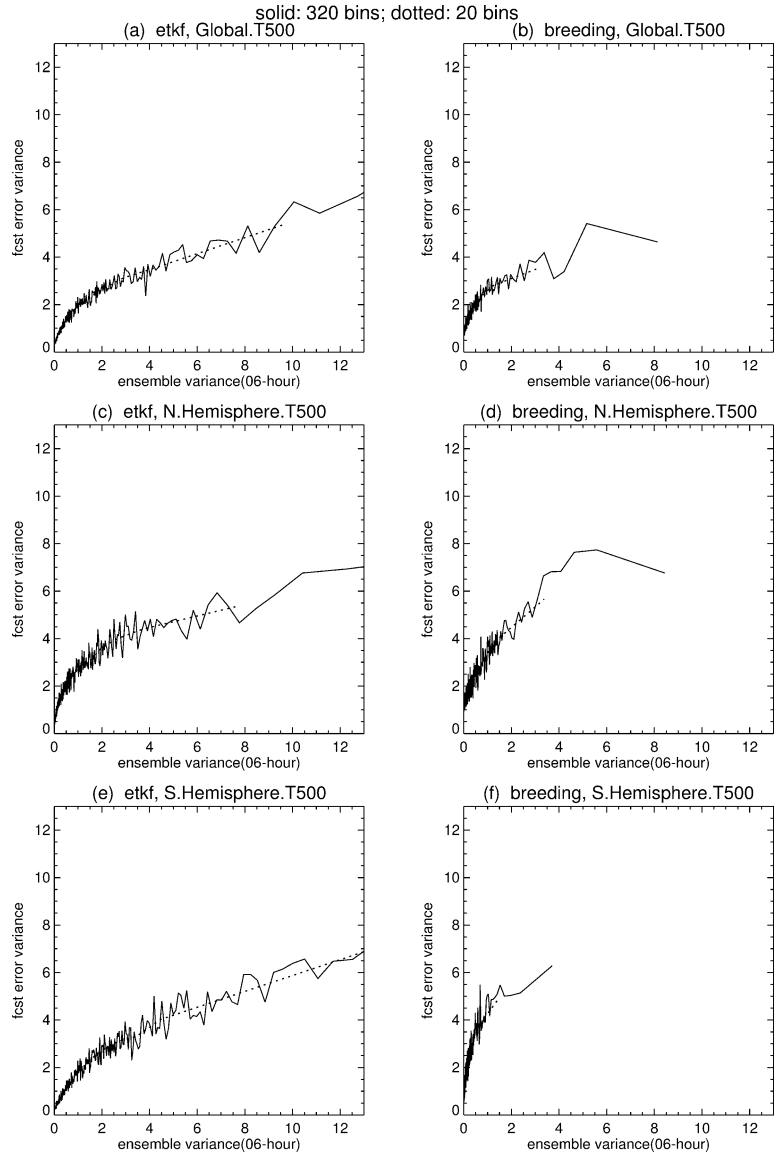
Fig. 10. Derived ensemble variance and forecast error variances at all grid points for 500-mb temperature, for ETKF (left panel) and breeding (right panel); for global (top), Northern (middle) and Southern (bottom) Hemisphere regions. The average value from each of 320 bins is indicated by solid lines. Dotted lines show the results from 20 bins only.

While PECA values indicate the correlations between ensemble perturbations and forecast errors, it is also interesting and important to compare the ensemble variance with the forecast variance for the two 10-member systems. To analyse how well the ensemble variance can explain the forecast error variance, we follow the method used in Majumdar et al. (2001; 2002) and Wang and Bishop (2003). First, we compute the ensemble variance and squared error of temperature at each grid point at the 500-mb pressure level for a 6-hr forecast lead time. A scatter plot (which is not shown) can then be drawn by using ensemble (abscissa) and squared forecast errors for all grid points. We next divide the points into 320 equally populated bins in order of increasing ensemble variance. The ensemble and forecast variances are then averaged within each bin. It is the averaged values from each bin that are plotted (solid) in Fig. 10. If the

number of bins is reduced, it is expected that the curve will be smoother. The result from 20 bins is shown by a dotted line. The variance relationship between ensemble and forecast is studied globally (top panel), and for the Northern (middle panel) and Southern (bottom panel) Hemispheres. The results for ETKF and breeding ensembles are shown in the left and right panels, respectively.

The results from the 20-bin case (dotted line) show that the range of forecast error variance (maximum minus minimum values) explained by the ensemble variance is larger for ETKF (5.03) than breeding (2.77) in the global region (Figs. 10a and 10b). This shows that ETKF is better than breeding at being able to distinguish times and locations where forecast errors are likely to be large from the times and locations where forecast errors are likely to be small. For the other two regions, the ranges of

forecast variances from ETKF are also slightly larger compared to the breeding ensemble.

The standard anomaly correlation scores for 500-mb height from the two 10-member systems show that the breeding ensemble has slightly higher scores than the ETKF in both Northern and Southern Hemispheres (not shown). It is known that these scores can be influenced by the magnitude of the initial spread (Buizza et al., 2005). As Fig. 3 shows, the initial spread of ETKF is generally larger in both Northern and Southern Hemispheres. This may reduce the anomaly correlation scores of the ETKF ensemble. A conclusion will be drawn from the future comparison with both systems having similar initial spread.

The analysis rank histograms (Toth et al., 2003) from the two systems (each with 10 members) for different forecast lead times are also studied (not shown). The results from the rank histograms for different lead times can be more concisely summarized in average percentage of excessive outliers (APEO) (Buizza et al., 2005). APEO is a measure of statistical reliability. It is the percentage of cases where the verifying analysis at any grid point lies outside the cloud of the ensemble in excess to what is expected by chance. A reliable ensemble will have a score of zero, while larger positive values indicate more outlier verifying analysis cases than expected from chance. The APEO values of 500-mb height in the Northern Hemisphere from the two ensembles show that the two systems have similar values for short lead times up to day 1. After day 1, the ETKF ensemble has lower APEO than the breeding system for up to nine days. The differences between the two systems are about 3–7%. After 9 to 10 d, the difference is reduced slightly. The results indicate that the ETKF ensemble has flatter rank histograms than the breeding ensemble and is more reliable. In general, APEO value could be reduced by larger spread. The fact that the ETKF ensemble has larger initial spread than breeding could have played a role the comparison.

3.6. Consistency between forecast and analysis perturbations

In this subsection, we evaluate the consistency between the forecast and corresponding transformed analysis perturbations. This consistency will be measured by computing the correlation between each of short-range ensemble forecast and its corresponding analysis perturbation. High correlation values indicate that the generation of new initial perturbations introduces minimal changes to the forecasts from which the analysis perturbations are derived; this means that there is a strong similarity between each short-range forecast and its corresponding analysis perturbation, and consequently, the individual perturbations exhibit a strong temporal consistency from one forecast cycle to the next.

Temporal consistency of ensemble forecasts from one cycle to the next in a general sense was discussed by Toth et al. (1997), and a quantitative measure of such consistency was offered in Toth et al. (2003). The temporal consistency of individual members, to

our knowledge, has not been explicitly discussed in the literature. However, such consistency may be a desirable characteristic of an ensemble forecast system for a number of applications. First, one can argue that the smaller the changes that the perturbation generation step introduces into the ensemble forecasts, the less the chance that the dynamically relevant information (e.g., the estimate of the fastest-growing perturbation directions) will be contaminated by any noise (i.e., dynamically not relevant information) in the process. A noise reduction in ensemble-based data assimilation has been shown to have a positive effect on the quality of the ensemble by, e.g., Whitaker and Hamill (2002). In their ensemble-based data assimilation work, Ott et al. (2004), in fact introduced a constraint aimed at limiting the changes applied to the forecast perturbations when deriving their analysis perturbation fields.

Second, the temporal consistency of ensemble perturbations as defined above can be useful in various applications of global ensemble forecasting, such as ocean wave, land surface, and hydrologic ensemble forecasting. Ocean waves, for example, are sensitive to wind forcing over a period of several days, and their numerical analysis is strongly dependent on past analysis of wind fields (Hendrik Tolman, 2005, personal communication). Ideally, one would desire to have a number of a series of perturbed analysis fields from the recent past that each constitutes a realistic perturbed scenario in time. Wind perturbations that are uncorrelated with perturbations at earlier times may cancel the ocean wave perturbations generated earlier and overall may spuriously reduce the magnitude of the ocean wave ensemble variance.

Third, statistical postprocessing and subjective forecast applications can potentially add extra value by utilizing the temporal consistency in the ensemble perturbations. Depending on the strength and time scale of the temporal correlations, the perturbed member with the best performance at a short lead time may produce one of the best members at subsequent initial times as well (Peter Manousos, 2005, personal communication).

As for the three main ensemble generation methods (Buizza et al., 2005), the SV-based methods (Buizza and Palmer, 1995), by definition, exhibit no temporal correlation as defined above. Perturbation methods using ensemble-based data assimilation techniques (Houtekamer et al., 1996) that have no built-in constraints can also be expected to yield low correlation values as well.

In the breeding ensemble, analysis perturbations are scaled from the 6-hr forecast perturbations. That is, $\mathbf{z}_m^a(i, j) = \alpha_m(i, j)\,\mathbf{z}_m^f(i, j)$, where $\alpha_m(i, j)$ is the rescaling factor derived from a mask field for ensemble member m and grid point i, j in horizontal space. In the case of a single global rescaling factor $\alpha_m(i, j) = constant$ at every cycle, the correlation between analysis and forecast perturbations will be 1.0. In this case, the spatial variations in analysis errors are not accounted for. In the procedure called regional rescaling, a mask, that has been constructed to describe spatial variations in analysis uncertainty, is used as a target for the amplitude of the analysis

perturbations [whose magnitude is measured using a strong spatial smoothing, see Toth and Kalnay (1997)]. At every grid point, the rescaling factor applied in the regional rescaling version of the breeding method is determined as a ratio of the target perturbation value given in the mask field with the spatially smoothed value of the forecast perturbation amplitude. It is expected that the correlation values between $\mathbf{z}_m^a$ and $\mathbf{z}_m^f$ in a bred ensemble that uses the regional rescaling procedure described above will be below one (due to the use of spatially dependent rescaling factors) but relatively high, due to the use of the strong smoothing factor used in the norm of the rescaling procedure (ensuring that the rescaling factors change in a smooth fashion in space).

In ETKF theory, the 6-hr forecast perturbations are transformed into analysis perturbations based on Kalman filter theory, taking the observation information into account, such as

$$\mathbf{Z}^a = \mathbf{Z}^f \mathbf{T} \mathbf{C}^T = \mathbf{Z}^f \mathbf{C} (\mathbf{\Gamma} + \mathbf{I})^{-1/2} \mathbf{C}^T. \tag{6}$$

The transformation from forecast to analysis perturbations can be described in three steps. First, the forecast perturbations $\mathbf{Z}^f$ are rotated by $\mathbf{C}$, then they are scaled by $(\mathbf{\Gamma} + \mathbf{I})^{-1/2}$. Finally, they are rotated again by $\mathbf{C}^T$ which is a simplex transformation. The main purpose of the simplex transformation is to centre the transformed perturbations around the analysis field while preserving analysis covariance. In the first step, the forecast perturbations are rotated into different directions, while the second step only rescales the rotated perturbations. It can be expected that without the last step, simplex transformation $\mathbf{C}^T$, the rotated and scaled perturbations would have a low correlation with the original forecast perturbations, depending on how much the perturbations are rotated. However, with the simplex transformation the rotated and scaled perturbations are rotated towards the directions that are opposite to the first-step rotation by $\mathbf{C}$. If the eigenvalue distribution $\mathbf{\Gamma}$ is completely flat, the correlation between $\mathbf{Z}^a$ and $\mathbf{Z}^f$ will be 1.0.

Shown in Fig. 11a are the averaged correlation values, over 10 members, between the forecast and analysis perturbations for ETKF (thick) and breeding (thin) ensembles at different times. The correlation between the forecast and analysis perturbations at each level is computed for both ensemble systems. The mean correlation over all levels is shown for each system. The results show that the mean correlation in the ETKF ensemble over different model levels is consistently higher than that in the breeding ensemble, although the mean correlation varies with time in both ensemble systems. At different pressure levels, the correlation between the corresponding forecast and analysis perturbations changes little for ETKF ensemble. However, for the breeding ensemble the correlation at different levels varies more, particularly at the top model level (2 mb, not shown). This variation with pressure level can be seen more clearly from Fig. 11b, which shows the vertical distribution of average correlation over time period from 15 January to 15 February 2003.
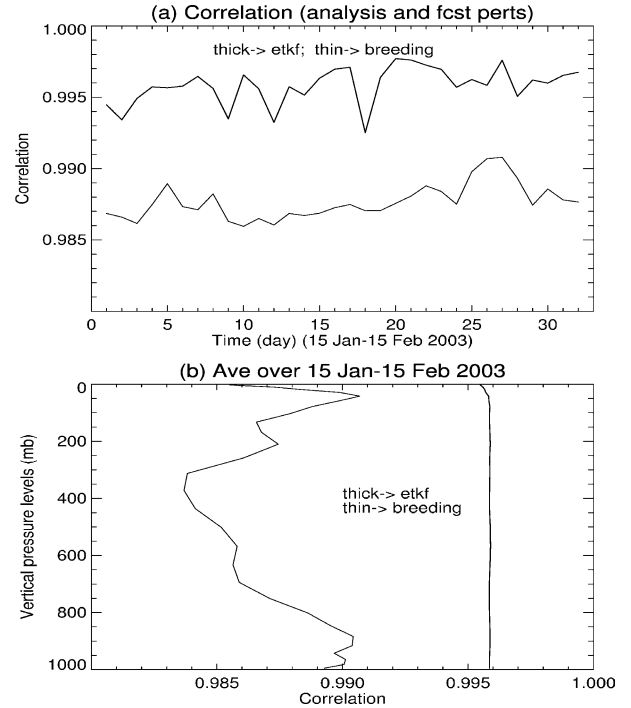

Fig. 11. Averaged correlation over 10 members between forecast and analysis perturbations for ETKF (thick) and breeding (thin) ensembles; (a) mean correlation as a function of time over all levels and (b) vertical correlation distribution averaged over time.

The average correlation over the experimental period is almost constant at different pressure levels for the ETKF ensemble, while the breeding ensemble shows larger variations at different levels.

The main reason for this extremely high correlation between analysis and forecast perturbations in ETKF is the simplex transformation. Equation (6) shows that the correlation in the ETKF ensemble is also influenced by the eigenvalue distribution $\mathbf{\Gamma}$. The eigenvalue distribution is determined by the number and locations of observations and the number of ensemble members. In our experiment, the eigenvalue distribution of the forecast covariance matrix in normalized observational space is shown (by diamonds) in Fig. 5a. The EDF is 8.9 out of nine independent forecast perturbations. The variances are quite evenly distributed in different directions. The correlation between analysis and forecast perturbations in the ETKF ensemble is changed little by this distribution. We notice that WB also showed a similar variance distribution in their model with ideal observations. We can reasonably expect that the analysis perturbations in most ETKF ensembles with simplex transformations have high correlations with the forecast perturbations, although the exact influence of the observations and the number of ensembles is hard to know. This feature makes the ETKF-based ensemble system particularly appealing.

4. Discussion and conclusions

In this paper, we have carried out experiments with two ensemble forecast systems based on two different techniques for generating initial perturbations: ETKF and breeding. Results are presented for a 32-d experimental period using the NCEP operational analysis/forecast system, and focusing on the characteristics of analysis and short-range forecast perturbations. One purpose of this comparison between the ETKF and breeding ensembles is to see if the ETKF-generated initial perturbations are more responsive to observation distributions and are representative of the analysis uncertainties, and whether the performance can be improved.

The properties of ETKF-generated perturbations are thoroughly studied from various aspects, such as the EDF of subspaces spanned by perturbations in local, observational, global 2-D and 3-D grid point spaces, and optimally combined orthogonal perturbations with the largest AFs. The relative strengths and weaknesses of the two systems are discussed and identified. The results presented in this paper for the first time offer a valuable, comprehensive description of the performance of an ETKF-based ensemble forecast system under a real-time observation environment.

The findings from our experiments are summarized as follows.

- The ETKF method is shown to produce initial perturbations whose variance, as desired, is influenced by variations in data coverage.

This is in contrast to some current operational techniques such as the breeding technique at NCEP and the SV technique, although other techniques such as the PO method used at MSC and Hessian SVs at ECMWF (Barkmeijer et al., 1999) are expected to produce a similar result.

- Due to the small number of ensemble members used in the ETKF experiment, the ETKF cannot represent the variations in analysis error variance on the global scale as well as breeding with geographical rescaling.
- The slope of the eigenvalue spectrum of the breeding ensemble covariance matrix is clearly steeper than that of the corresponding ETKF eigenvalue spectrum. The EDF of the 10-member ETKF ensemble is much larger than that of the paired breeding ensemble.

This is related to the ensemble centring strategies used in the two systems and also to the whitening effect of the error covariance update equation used by the cycling ETKF. A simplex centring method was used in the ETKF, while a paired centring scheme was used in the operational breeding system. For instance, if five members (one from each pair) are chosen from each system, the EDF values for the two systems are very similar. To test this issue more cleanly, we would have needed to follow Toth and Kalnay (1993; 1997) and Wang et al. (2004) and generate a nonpaired breeding ensemble to compare against the spherical simplex ETKF ensemble. Wang and Bishop (2003)

found that the ETKF-maintained variance in orthogonal directions much more effectively than breeding.

- Although the individual 10 bred perturbations grow faster than the ETKF perturbations, the optimal perturbation growth that can be found by linearly combining 10 forecast perturbations is larger for the ETKF than for breeding for optimization times less than 2 d. When only five perturbations are included in the optimization, optimally combined bred perturbations have higher growth rates for all lead times.

A good ensemble forecast system requires that the initial perturbations grow fast enough to match the growth rates of forecast errors. Calculations, not reported here, show that perturbation growth increases as perturbation amplitude is decreased. Our findings are consistent with this expectation: in the extra-tropics, where ETKF amplitude exceeds breeding amplitude, individual bred perturbations grow faster than individual ETKF perturbations whereas in the tropics, where ETKF amplitude is less than breeding amplitude, individual ETKF perturbations grow faster than individual breeding perturbations. In considering these results, it is worthwhile noting that the ETKF fits within the general breeding framework and can be viewed as a form of breeding in which the Kalman filter error covariance update equation is used to constrain the transformation of forecast perturbations into analysis perturbations.

- Perturbation versus error correlation analysis calculations indicate that at short lead time, bred perturbations can explain a larger portion of forecast error variance than ETKF perturbations. Beyond 1-d lead time, however, 10 ETKF perturbations are more efficient in explaining forecast error variance than the bred perturbations.

Note that PECA values quantitatively measure how well linear combinations of ensemble perturbations match the forecast errors. For longer forecast lead times any perturbation, including ETKF ensemble perturbations, will turn towards the leading Lyapunov vectors that are linked to the BVs (Wei, 2000; Wei and Frederiksen, 2004).

- Ensemble Transform Kalman Filter forecast error variance predictions were better than corresponding breeding predictions at distinguishing at times and locations where forecast errors were larger from times and locations where they were small.
- Both systems produce temporally consistent perturbation fields.

It is found that 10-member ETKF analysis perturbations have a very high correlation with forecast perturbations before the ETKF transformation. The average correlation values for the two systems are above 0.985 with the ETKF having slightly higher value than that in a breeding system with regional rescaling. This good feature of ETKF perturbations is due to the simplex transformation imposed.

- The forecast scores from the two 10-member systems are similar with only slight differences.

The results show that the breeding ensemble has slightly higher anomaly correlation than the ETKF ensemble in both Northern and Southern Hemispheres. This result may be influenced by the magnitude and geographical distribution of initial perturbation variances in the two systems compared, as well as the use of symmetric centring in the paired breeding scheme and spherical simplex centring in the ETKF scheme.

- The APEO from the ETKF ensemble has lower value than the breeding system in both Northern and Southern Hemispheres. The results are based on 10 members for both systems. This is consistent with the fact that the ETKF ensemble variance in this experiment setup is generally larger than the breeding ensemble variance.

We note that the above findings are from the experiments we have carried out so far. There are still some clear limitations in our study, such as

1. One should note that both theory and Wang and Bishop's (2003) results indicate that potential for the ETKF ensemble to outperform the masked breeding ensemble increases as the ensemble size is increased. Also, it is important to inflate the analysis perturbations properly. Here, we used the same inflation strategy as WB. It worked fine in their environment with idealized observation system. However in our operational environment with real observations, the inflation strategy needs to be improved. At present, how to correctly inflate the analysis variances remains a challenging research issue for the ensemble Kalman filter research community. To avoid the effect of possible ill-posed inflation factor on the ETKF in the current comparison with the breeding and also to ameliorate the effect of perturbation magnitude on the comparison of perturbation growth, one simple way would be to inflate the ETKF initial perturbations so that on a globally averaged basis the initial ensemble variance of the ETKF was equivalent to that of the operational breeding. Alternatively, to handle the problem of varying number of observations, instead of using just the current observations, we can also try to use previous two weeks' observations and follow the similar steps in WB to get the inflation factor. We plan to explore these possibilities in future work.

2. WB's experiments showed that an 8-member ETKF ensemble was not large enough to resolve geographical fluctuations in observation density, while their 16-member ensemble was large enough to resolve very large-scale fluctuations in observational density. The ETKF ensemble outperformed breeding system in WB's experiment. The key differences that make the results in this paper differ from those of WB can be attributed to (a) our observation system is very different from WB's as we summarized in the introduction; (b) sixteen members were used in their lower resolution system. We had only 10 members for our system with higher resolution; (c) the inflation scheme may not work well due to the large variations of observations in both space and time.

3. Only the so-called conventional data from the NCEP operational data assimilation system have been used. To include satellite data, more work is needed.

4. Ensemble Transform Kalman Filter analysis error estimates assume that the background error covariance matrix used in data assimilation is identical to the ensemble covariance matrix. This is not strictly correct. Since we did not use ETKF to carry out data assimilation, the analysis perturbations generated by ETKF are not centred around the analysis field generated by ETKF, but the analysis operationally produced by the NCEP 3-D-Var system. The ETKF analysis estimate is not fully consistent with the NCEP operational 3-D-Var analysis. NCEP's 3-D-Var operational data assimilation system (Parrish and Derber, 1992) assumes quasi-isotropic covariances of a different nature to those generated by a 10-member ETKF ensemble. It is expected that the background covariance matrix produced by the NMC method is more isotropic than that generated by the ensembles. This is particularly true for our small number of ensemble members. There are two ways of avoiding this inconsistency between the error covariance model assumed by the ETKF and that assumed by the data assimilation scheme. First, one could use ensemble-based data assimilation. As described in the introduction, ETKF, EAKF, EnSR and LEKF all are ensemble-based Kalman filters. A major intercomparison project has been initiated recently at NCEP in cooperation with the people who derived and formulated these filters at the NOAA Climate Diagnostics Center (NOAA CDC), University of Maryland and National Center for Atmospheric Research (NCAR). This project is supported by THORPEX (see http://box.mmm.ucar.edu/uswrp). The goal of this project is to compare the performance of each of these ensemble-based data assimilation schemes in an environment with real operational (and more sophisticated) models and data. The results will be compared with the benchmark NCEP operational data assimilation system. A second possibility would be to get the analysis uncertainty information from 3-D/4-D Var and feed it into the ensemble forecast system. We plan to explore this with respect to breeding techniques in the future. Orthogonalization and simplex transformations can be used to restrain initial perturbation variance. The results will be compared with 40-member ETKF.

5. Acknowledgments

References

Anderson, J. L. 2001. An ensemble adjustment Kalman filter for data assimilation. *Mon. Weather Rev.* **129**, 2884–2903.

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

Bishop, C. H. and Toth, Z. 1999. Ensemble transformation and adaptive observations. *J. Atmos. Sci.* **56**, 1748–1765.

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

Bretherton, C. S., Widmann, M., Dymnikov, V. P., Wallace, J. M. and Blade, I. 1999. The effective number of spatial degrees of freedom of a time-varying field. *J. Clim.* **12**, 1990–1999.

Buizza, R. and Palmer, T. N. 1995. The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.* **52**, 1434–1456.

Buizza, R., Houtekamer, P. L., Toth, Z., Pellerin, P., Wei, M. and Zhu, Y. 2005. A comparison of the ECMWF, MSC and NCEP global ensemble prediction systems. *Mon. Weather Rev.* **133**, 1076–1097.

Burgers, G., van Leeuwen, P. J. and Evensen, G. 1998. On the analysis scheme in the ensemble Kalman filter. *Mon. Weather Rev.* **123**, 1719–1724.

Etherton, B. J. and Bishop, C. H. 2004. Resilience of hybrid ensemble/3D-var analysis schemes to model error and ensemble covariance error. *Mon Weather Rev.* **130**, 1065–1080.

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

Houtekamer, P. L., Lefaivrem, L., Derome, J., Ritchie, H. and Mitchell, H. L. 1996. A system simulation approach to ensemble prediction. *Mon. Weather Rev.* **124**, 1225–1242.

Julier, S. J. and Uhlmann, J. K. 2002. Reduced sigma point filters for propagation of means and covariances through nonlinear transformations. *Proc. IEEE American Control Conf.*, Anchorage, AK, IEEE, 887–892.

Lorenc, A. C. 2003. The potential of the ensemble Kalman filter for NWP – a comparison with 4D-var. *J. R. Meteorol. Soc.* **129**, 3183–3203.

Majumdar, S. J., Bishop, C. H., Szunyogh, I. and Toth, Z. 2001. Can an Ensemble Transform Kalman Filter predict the reduction in forecast error variance produced by targeted observations?. *Q. J. R. Meteorol. Soc.* **127**, 2803–2820.

Majumdar, S. J., Bishop, C. H. and Etherton, B. J. 2002. Adaptive sampling with Ensemble Transform Kalman Filter. Part II: field program implementation. *Mon. Weather Rev.* **130**, 1356–1369.

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

Oczkowski, M., Szunyogh, I. and Patil, D. J. 2005. Mechanism for the development of locally low-dimensional atmospheric dynamics. *J. Atmos. Sci.* **62**, 1135–1156.

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

Palmer, T. N., Gelaro, R., Barkmeijer, J. and Buizza, R. 1998. Singular vectors, metrics, and adaptive observations. *J. Atmos. Sci.* **55**, 633–653.

Parrish, D. F. and Derber, J. 1992. The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Weather Rev.* **120**, 1747–1763.

Patil, D. J., Hunt, B. R., Kalnay, E., Yorke, J. A. and Ott, E. 2001. Local low dimensionality of atmospheric dynamics. *Phys. Rev. Lett.* **86**, 5878–5881.

Purser, R. J. 1996. Arrangement of ensemble in a simplex to produce given first and second-moments, NCEP Internal Report (available from the author at Jim.Purser@noaa.gov).

Rabier, F., Klinker, E., Courtier, P. and Hollingsworth, A. 1996. Sensitivity of forecast errors to initial conditions. *Q. J. R. Meteorol. Soc.* **122**, 121–150.

Szunyogh, I., Kostelich, E. J., Gyarmati, G., Patil, D. J., Hunt, B. R., and co-authors. 2005. Assessing a local ensemble Kalman filter: perfect model experiments with the NCEP global model. *Tellus* **57A**, 528–545.

Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. and Whitaker, J. S. 2003. Ensemble square root filters. *Mon. Weather Rev.* **131**, 1485–1490.

Toth, Z. and Kalnay, E. 1993. Ensemble forecasting at NMC: the generation of perturbations. *Bull. Am. Meteorol. Soc.* **174**, 2317–2330.

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

Toth, Z., Kalnay, E., Tracton, S. M., Wobus, R. and Irwin, J. 1997. A synoptic evaluation of the NCEP ensemble. *Weather Forecast.* **12**, 140–153.

Toth, Z., Talagrand, O., Candille, G. and Zhu, Y. 2003. Probability and ensemble forecasts. In: *Forecast Verification: A Practitioner's Guide in Atmospheric Science* (eds Ian T. Jolliffe and David B. Stephenson). John Wiley & Sons Ltd., England, 137–163.

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. Weather Rev.* **132**, 1590–1605.

Wei, M. 2000. Quantifying local instability and predictability of chaotic dynamical systems by means of local, metric entropy. *Int. J. Bifurcation Chaos* **10**, 135–154.

Wei, M. and Frederiksen, J. S. 2004. Error growth and dynamical vectors during southern hemisphere blocking. *Nonl. Proc. Geophys.* **11**, 99–118.

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

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