

Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system

By MOZHENG WEI^{1*}, ZOLTAN TOT², RICHARD WOBUS³ and YUEJIAN ZHU², ¹SAIC at NOAA/NWS/NCEP, Environmental Modeling Center, 5200 Auth Road, Rm. 207, Camp Springs, MD 20746, USA; ²NOAA/NWS/NCEP, Camp Springs, MD, USA; ³SAIC at NOAA/NWS/NCEP, Camp Springs, MD, USA

(Manuscript received 22 November 2006; in final form 9 August 2007)

ABSTRACT

Since modern data assimilation (DA) involves the repetitive use of dynamical forecasts, errors in analyses share characteristics of those in short-range forecasts. Initial conditions for an ensemble prediction/forecast system (EPS or EFS) are expected to sample uncertainty in the analysis field. Ensemble forecasts with such initial conditions can therefore (a) be fed back to DA to reduce analysis uncertainty, as well as (b) sample forecast uncertainty related to initial conditions. Optimum performance of both DA and EFS requires a careful choice of initial ensemble perturbations.

DA can be improved with an EFS that represents the dynamically conditioned part of forecast error covariance as accurately as possible, while an EFS can be improved by initial perturbations reflecting analysis error variance. Initial perturbation generation schemes that dynamically cycle ensemble perturbations reminiscent to how forecast errors are cycled in DA schemes may offer consistency between DA and EFS, and good performance for both. In this paper, we introduce an EFS based on the initial perturbations that are generated by the Ensemble Transform (ET) and ET with rescaling (ETR) methods to achieve this goal. Both ET and ETR are generalizations of the breeding method (BM).

The results from ensemble systems based on BM, ET, ETR and the Ensemble Transform Kalman Filter (ETKF) method are experimentally compared in the context of ensemble forecast performance. Initial perturbations are centred around a 3D-VAR analysis, with a variance equal to that of estimated analysis errors. Of the four methods, the ETR method performed best in most probabilistic scores and in terms of the forecast error explained by the perturbations. All methods display very high time consistency between the analysis and forecast perturbations. It is expected that DA performance can be improved by the use of forecast error covariance from a dynamically cycled ensemble either with a variational DA approach (coupled with an ETR generation scheme), or with an ETKF-type DA scheme.

1. Introduction

The goal of ensemble forecasting is to generate a sample of numerical forecasts that represent our knowledge about the possible evolution of a dynamical system. A set of ensemble forecasts must preferably reflect forecast uncertainty related to both initial value (analysis) and numerical model related errors. During the past 15 yr, various perturbation methods have been developed to achieve these goals.

As for initial ensemble perturbations, at a general level it is accepted that they must constitute a sample taken from a probability density function (PDF) that represents our best knowledge

about the state of the dynamical system (i.e. ‘analysis PDF’). Various initial perturbation methods differ in how they *estimate* the analysis PDF, and how they *sample* it.

The operational implementation of the *first generation* initial perturbation generation methods (Table 1): the Perturbed Observation (PO) method (Houtekamer et al., 1996), the Total Energy norm based Singular Vector (TE-SV) method (Buizza and Palmer, 1995; Molteni et al., 1996) and the breeding method (BM, Toth and Kalnay, 1993, 1997) were all limited in that for various reasons the sample they produced was not consistent with the analysis PDF.

(1) The PO method, used at the Meteorological Service of Canada (MSC), relied on an Optimal Interpolation (OI) analysis scheme that produced suboptimal analyses compared to a newer variational scheme developed at the same time at MSC.

*Corresponding author.

e-mail: mozheng.wei@noaa.gov

DOI: 10.1111/j.1600-0870.2007.00273.x

Table 1. First generation initial perturbation generation techniques

	Perturbed Observations (MSC, Canada)	Breeding with Regional Rescaling (NCEP, USA)	Singular Vectors with total energy norm (ECMWF)
Estimation of analysis uncertainty	Realistic through sample, case dependent patterns and amplitudes	Fastest growing subspace, case dependent patterns	No explicit estimate used, variance not flow dependent
Sampling of analysis uncertainty	Random for all errors, including non-growing, potentially hurts short-range performance	Non-linear Lyapunov vectors, subspace of fastest growing errors, some dependence among perturbations	Dynamically fastest growing in future, orthogonal
Consistency between EFS and DA system	Good, quality of DA lagging behind 3D-Var	Not consistent, time-constant variance due to use of fixed mask	Not consistent, potentially hurting short-range performance

This was due to the fact that OI scheme was much cheaper to run. In addition, POs can introduce undesirable noise into the forecasts (Whitaker and Hamill, 2002). As most observations are perturbed by a simulated observational error, followed by multiple numerical analyses, perturbations with the PO method represent a random sample from the suboptimal analysis PDF, containing both fast growing and neutral/decaying perturbation patterns with realistic amplitude. Since the PO method samples analysis error, therefore it is closer to a second generation method than other first generation methods.

(2) The TE-SV method assumes that each possible pattern with unit total energy in the three-dimensional and multivariate space of a numerical model is equally likely to be the error pattern in a numerical analysis. Sampling out of the assumed uniform distribution is done through the SV method that results in perturbations that produce the maximum linear growth over a pre-specified ‘optimization’ period. Growth and perturbation patterns beyond a transitional period are determined by the state dependent local Lyapunov characteristics of the system (supporting “sustainable” growth), while those during the transitional period are strongly influenced by the choice of the norm used in the definition of SVs. As Jon Alquist (personal communication, 1999) showed in a study, by the proper choice of the norm, *any* perturbation pattern can be made a leading SV. Subsequent research (Barkmeijer et al., 1999) demonstrated that when the assumption about uniform error distribution is replaced with a constraint based on information from the analysis error covariance matrix used in variational data assimilation (DA) techniques (Hessian-SV), the initial perturbation patterns become more similar to bred vectors (and distinctly different from the TE-SV vectors). Fisher and Andersson (2001) found that both ET-SV and Hessian-SV are poor approximation of the actual analysis error covariance.

(3) The BM method is based on the assumptions that, (i) analysis errors are dominated by short-range forecast errors (since the analysis fields are constructed through the use of dynamically cycled forecasts), Errico et al. (2007) confirmed that analysis errors to first approximation look like 6-h forecast errors. (ii) The proper sampling of the large number of possible

non-growing error patterns is not possible (and therefore, to avoid aliasing these patterns should not be sampled with realistic amplitude in initial perturbations). To simulate the error breeding process in an analysis cycle, a perturbation field in a breeding cycle is dynamically cycled through the use of two non-linear forecast integrations, where the difference between the two forecasts are periodically rescaled and then repositioned onto consecutive analysis fields (Toth and Kalnay, 1993). Note that the resulting perturbations are similar to those generated by the PO method except they represent only the growing, and ignore the neutral and decaying components of analysis error, providing a random sample of growing analysis errors.

In general the initial perturbations in first generation ensemble prediction/forecast systems (EPS or EFS) do not fully represent the uncertainties in analysis, as one expects from an ideal EFS. They are not consistent with the DA systems that generate the analysis fields. Comparisons of performance between the ECMWF and NCEP operational EFSs were described in Zhu et al. (1996, personal communication) and in Wei and Toth (2003). By using a quasigeostrophic channel model, Hamill et al. (2000) compared the performance of these three methods based on BM, SVs and PO. A more recent comprehensive summary of these first generation methodologies and their performance at ECMWF, MSC and NCEP can be found in Buizza et al. (2005). The results showed that the ECMWF ensemble has largest spread in the medium range which helped its performance for a number of measures in comparison with NCEP and MSC ensembles. The basic properties, including both advantages and disadvantages, are summarized in Table 1.

Research since the emergence of the first generation ensemble forecasting techniques in the 1990s has revealed that the addition of random perturbations slightly degrades the performance characteristics of ensemble-based DA systems (Anderson, 2001; Whitaker and Hamill, 2002). With an increased emphasis on the use of the analysis PDF for initial ensemble perturbation generation, recently a new, *second generation* of techniques have emerged (Table 2).

In this paper, a new ensemble forecasting method using the Ensemble Transform (ET) technique is introduced and tested

Table 2. Second generation initial perturbation generation techniques

	ETKF, perturbations influenced by forecasts and observations	ET/rescaling with analysis error variance estimate from DA	Hessian Singular Vectors
Estimation analysis uncertainty	Fast growing subspace, case dependent patterns and amplitudes	Fast growing subspace, case dependent patterns and amplitudes	Case-dependent variance info from analysis, amplitudes of SVs have to be specified
Sampling analysis uncertainty	Orthogonal in the normalized observational space	High EDF in ensemble subspace	Dynamically fastest growing in future
Consistency between EFS and DA system	Very good, however, quality of DA has not been proven better than 4D-Var in operational environment so far	Very good, DA provides good analysis for EFS which provides accurate forecast error covariance for DA	Possibly consistent (not used operationally by any known NWP centres)

for the generation of initial ensemble perturbations for the representation of analysis uncertainty. The ET technique was first proposed by Bishop and Toth (1999) in target observation studies. The research and experiments on using ET and ET with rescaling (ETR) for ensemble forecasts first started at NCEP before 2004, and the initial results were presented in the THORPEX Symposium in 2004 (Wei et al., 2005a). Since then, more experimental results with ensembles using ET and ETR have been presented and documented in Wei et al. (2005b, 2006b,c).

Another technique in the second generation is based on the Ensemble Transform Kalman Filter (ETKF). The ETKF method was proposed by Bishop et al. (2001) also for adaptive observation studies. Wang and Bishop (2003) used ETKF to generate ensemble perturbations in an idealized observation framework. ETKF was further extended to an operational environment with the NCEP operational model and real-time observations by Wei et al. (2006a) (hereafter referred to as W06). Please note that the Hessian singular vector based technique (Barkmeijer et al., 1999) can also be classified as a second generation method, as listed in Table 2 if the Hessian-SV is computed with flow-dependent analysis PDF. It samples fastest growing directions for a specific lead time and norm, that is, the Hessian-SV depends on a time interval you choose. Like TE-SV, the amplitude of initial Hessian SV has to be specified. The feedback from ensemble to the DA system was not explored and the Hessian-SV method has not been implemented operationally by any known numerical weather forecast centres so far. By using a much lower-order Lorenz 95 model, Bowler (2006) compared different initial perturbation generation techniques including ETKF, error breeding, singular vector, random perturbation and ensemble Kalman filter methods. Using a 300-variable Lorenz model, the author showed that EnKF performs best, the performance of ETKF with random perturbations is the next most skilful. It was also found that neither the ETKF, error breeding nor singular vectors provided useful background information on their own.

The ensemble based DA methods like ETKF share the principle of cycling the dynamically most relevant (i.e. without addi-

tion of any noise) fastest amplifying perturbations for describing the uncertainty in short range forecasts (Wang and Bishop, 2003; W06). Dynamically cycled ensemble perturbations may provide an ideal way of estimating the background (short range forecast) error covariance. At the same time, an advanced DA system may provide analysis error variance information to the ensemble generation schemes.

We will compare the results based on the four methods: BM, ETKF, ET and ETR. All four schemes belong to the same class of methods based on concept of breeding, involving the dynamical cycling of ensemble perturbations. This is based on the observation that since a modern NWP analysis method strongly rely on short range forecasts (Toth and Kalnay, 1993). This is supported by Errico et al. (2007) who found that: analysis error characteristics (e.g. statistics) are similar (to first approximation) to those of 6-h forecast error.

In the ET and ETR methods, the initial perturbations are restrained by the best available analysis variance from the operational DA system and centred around the analysis field generated by the same DA system. In this way, the ensemble system will be consistent with the DA. The perturbations are also flow dependent and orthogonal with respect to the inverse of analysis error variance. This will overcome some drawbacks in the current operational system resulting from paired perturbations (W06). Another advantage is that the ET/ETR technique is considerably cheaper than ETKF if the analysis variance information is available.

A common feature of the second generation techniques is that the initial perturbations are more consistent with the DA system. At NCEP we intend to develop an EFS that is consistent with the DA system that generates the analysis fields for the ensemble. This will benefit both EFS and DA systems. A good DA system will provide accurate estimates of the initial analysis error variance for the EFS, while a good, reliable EFS will produce accurate flow dependent background covariance for the DA system.

So far, no ensemble DA experiments have produced the analysis that is better than the product from the mature operational 3D/4D-Var systems at major weather forecasting centres with

operational observation data. Before ensemble DA shows satisfying performance, ETR with repositioning (i.e. perturbed ensemble states are centred about the analysis field) offers a good solution for consistent DA/EFS generation, with two-way exchange of information. It has a potential for smooth transition to operational ensemble DA when the performance of ensemble DA becomes satisfactory in an operational environment. The question of whether these ensemble based DA schemes, including ETKF, can generate a better analyses with real observations is currently being pursued by the developers of these schemes in collaboration with the authors at NCEP (see discussion section of W06).

Section 2 provides detailed descriptions of the ET formulation for initial perturbations. Section 3 introduces ETR and presents the major results from comparisons of ET/ETR with the NCEP operational bred perturbation-based ensemble system. Discussion and conclusions are given in Section 4.

2. Methodology

2.1. Basic formulation

Initial perturbations in the NCEP global EFS are generated by the BM with regional rescaling. This method is well established, documented and widely used. It dynamically recycles perturbations and is a non-linear generalization of the standard method which has been widely used for computing the dominant Lyapunov vectors (Wei, 2000; Wei and Frederiksen, 2004). A scientific description of the BM can be found in Toth and Kalnay (1993, 1997). Some limitations are that the variance is constrained statistically by a climatologically fixed analysis error mask and there is no orthogonalization between the perturbations due to the positive/negative paired strategy. More technical descriptions, documents and results are available on the NCEP ensemble forecast web site at <http://www.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. The application of ETKF to ensemble forecasting can be found in Wang and Bishop (2003) and Wang et al. (2004). More results about the characteristics of ETKF perturbations with NCEP real-time observations are described in W06. In the ETKF framework, the perturbations are dynamically cycled with orthogonalization in the normalized observational space. The ensemble variance is constrained by the distribution and error variance of observations. However, there are still some challenging issues in the ETKF based ensemble with real observations, such as perturbation inflation. Flow dependent inflation factors are hard to construct due to the fact that the number and positions of observations change rapidly from one cycle to the next. Since the ensemble mean from ETKF has yet to be improved to the level of the analysis from a mature variational DA

like the NCEP SSI (Parrish and Derber, 1992), the perturbations generated by ETKF have to be centred around the analysis field from SSI. In addition, the ETKF is much more expensive than breeding in an operational environment with real-time observations. More details can be found in W06.

The ET method was formulated in Bishop and Toth (1999) for targeting observation studies. In this paper, we adopt this technique for generating ensemble perturbations. Let

$$\begin{aligned}\mathbf{Z}^f &= \frac{1}{\sqrt{k-1}}(\mathbf{z}_1^f, \mathbf{z}_2^f, \dots, \mathbf{z}_k^f), \\ \mathbf{Z}^a &= \frac{1}{\sqrt{k-1}}(\mathbf{z}_1^a, \mathbf{z}_2^a, \dots, \mathbf{z}_k^a),\end{aligned}\quad (1)$$

where the n dimensional state vectors $\mathbf{z}_i^f = \mathbf{x}_i^f - \bar{\mathbf{x}}^f$ and $\mathbf{z}_i^a = \mathbf{x}_i^a - \bar{\mathbf{x}}^a$ ($i = 1, 2, \dots, k$) are k ensemble forecast and analysis perturbations for all model variables, respectively. In our experiments, $\bar{\mathbf{x}}^f$ is the mean of k ensemble forecasts and $\bar{\mathbf{x}}^a$ is the analysis from the independent NCEP operational DA system. Unless stated otherwise, the lower and upper case bold letters will indicate vectors and matrices, respectively. In the ensemble representation, the $n \times n$ forecast and analysis covariance matrices are approximated, respectively, as

$$\mathbf{P}^f = \mathbf{Z}^f \mathbf{Z}^{fT} \quad \text{and} \quad \mathbf{P}^a = \mathbf{Z}^a \mathbf{Z}^{aT}, \quad (2)$$

where superscript T indicates the matrix transpose. For a given set of forecast perturbations \mathbf{Z}^f at time t , the analysis perturbations \mathbf{Z}^a are obtained through an ensemble transformation \mathbf{T} such that

$$\mathbf{Z}^a = \mathbf{Z}^f \mathbf{T}. \quad (3)$$

In the ET method, we want to use analysis error variances from the best possible DA system to restrain the initial perturbations for our EFS. At NCEP, the best analysis error variances can be derived from the NCEP operational DA system which uses many kinds of real-time observations. In contrast, ETKF method uses observations and short time forecasts from previous cycle to specify the analysis error covariance matrix, which is then used to restrain the initial analysis perturbations for next cycle ensemble forecasts. More details can be found in Wang and Bishop (2003) and W06.

2.2. ET perturbations

Suppose \mathbf{P}_{op}^a is the diagonal matrix with the diagonal values being the analysis error variances obtained from the operational DA system, the ET transformation matrix \mathbf{T} can be constructed as follows. For an ensemble forecast system, the forecast perturbations \mathbf{Z}^f can be generated by eq. (1). One can solve the following eigenvalue problem.

$$\mathbf{Z}^{fT} \mathbf{P}_{op}^{a-1} \mathbf{Z}^f = \mathbf{C} \mathbf{\Gamma} \mathbf{C}^{-1}, \quad (4)$$

where \mathbf{C} contains column orthonormal eigenvectors (\mathbf{c}_i) of $\mathbf{Z}^{fT} \mathbf{P}_{op}^{a-1} \mathbf{Z}^f$ (also the singular vectors of $\mathbf{P}_{op}^{a-1/2} \mathbf{Z}^f$), and $\mathbf{\Gamma}$

is a diagonal matrix containing the associated eigenvalues (λ_i) with magnitude in decreasing order, that is, $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_k]$, $\mathbf{C}^T \mathbf{C} = \mathbf{I}$ and $\mathbf{\Gamma} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k)$. Since the forecast perturbations are centred around the mean, that is, $\sum_{i=1}^k \mathbf{z}_i^f = 0$, the sum of the columns of matrix $\mathbf{P}_{op}^{a-1/2} \mathbf{Z}^f$ is zero (\mathbf{Z}^f is defined in eq. (1)). Thus,

$$\mathbf{Z}^{fT} \mathbf{P}_{op}^{a-1} \mathbf{Z}^f \mathbf{1} = \mathbf{C}^T \mathbf{C}^T \mathbf{1} = 0, \quad (5)$$

where $\mathbf{1} = (1, 1, \dots, 1)^T$. Equation (5) indicates that the last eigenvalue is zero and its associated eigenvector is a constant, that is, $\lambda_k = 0$ and since each eigenvector is normalized, one has $\mathbf{c}_k = (\frac{1}{\sqrt{k}}, \frac{1}{\sqrt{k}}, \dots, \frac{1}{\sqrt{k}})^T$. From eq. (5), we have $\mathbf{\Gamma} \mathbf{C}^T \mathbf{1} = 0$, which means that

$$\lambda_j \sum_{i=1}^k \mathbf{C}_{ij} = 0, \quad j = 1, 2, \dots, k. \quad (6)$$

Since only the first $k - 1$ eigenvalues are non-zero, we have

$$\sum_{i=1}^k \mathbf{C}_{ij} = 0, \quad j = 1, 2, \dots, k - 1. \quad (7)$$

Hence the sum of the components of each of the first $k - 1$ eigenvectors is zero.

Now suppose $\mathbf{F} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{k-1})$ and $\mathbf{G} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{k-1}, \alpha)$ where α is a non-zero constant, that is, $\mathbf{G} = \text{diag}(g_1, g_2, \dots, g_k) = \begin{pmatrix} \mathbf{F} & 0 \\ 0 & \alpha \end{pmatrix}$ and $\mathbf{\Gamma} = \begin{pmatrix} \mathbf{F} & 0 \\ 0 & 0 \end{pmatrix}$

The ET analysis perturbations can be constructed through transformation $\mathbf{T}_p = \mathbf{C} \mathbf{G}^{-1/2}$ and

$$\mathbf{Z}_p^a = \mathbf{Z}^f \mathbf{T}_p = \mathbf{Z}^f \mathbf{C} \mathbf{G}^{-1/2}. \quad (8)$$

The analysis perturbations after ET are $\mathbf{Z}_p^a = \mathbf{Z}^f \mathbf{C} \mathbf{G}^{-1/2}$, since

$$\begin{aligned} \mathbf{Z}_p^{aT} \mathbf{P}_{op}^{a-1} \mathbf{Z}_p^a &= \mathbf{G}^{-1/2} \mathbf{C}^T \mathbf{Z}^{fT} \mathbf{P}_{op}^{a-1} \mathbf{Z}^f \mathbf{C} \mathbf{G}^{-1/2} \\ &= \mathbf{G}^{-1/2} \mathbf{C}^T \mathbf{C} \mathbf{\Gamma} \mathbf{C}^{-1} \mathbf{C} \mathbf{G}^{-1/2} \\ &= \begin{pmatrix} \mathbf{F}^{-1/2} & 0 \\ 0 & \alpha^{-1/2} \end{pmatrix} \begin{pmatrix} \mathbf{F} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{F}^{-1/2} & 0 \\ 0 & \alpha^{-1/2} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{I}_{(k-1) \times (k-1)} & 0 \\ 0 & 0 \end{pmatrix}. \end{aligned} \quad (9)$$

Equation (9) shows that the first $k - 1$ analysis perturbations are orthogonal with respect to an inverse analysis error variance norm. The analysis error covariance matrix can be approximated through analysis perturbations such as eq. (2) if the number of ensemble members is large, that is, when $k \rightarrow n$.

Let's look at the individual components of each perturbation. Equation (8) leads to:

$$(\mathbf{Z}_p^a)_{mq} = \sum_{i=1}^k \mathbf{Z}_{mi}^f \sum_{l=1}^k \mathbf{C}_{il} \mathbf{G}_{lq}^{-1/2} = \sum_{i=1}^k \mathbf{Z}_{mi}^f \mathbf{C}_{iq} g_q^{-1/2}. \quad (10)$$

If $q = k$, we get the last analysis perturbation

$$(\mathbf{Z}_p^a)_{mk} = \sum_{i=1}^k \mathbf{Z}_{mi}^f \mathbf{C}_{ik} g_k^{-1/2} = \frac{1}{\sqrt{\alpha k}} \sum_{i=1}^k \mathbf{Z}_{mi}^f = 0. \quad (11)$$

The sum of the analysis perturbations is

$$\begin{aligned} \sum_{q=1}^k (\mathbf{Z}_p^a)_{mq} &= \sum_{q=1}^k \sum_{i=1}^k \mathbf{Z}_{mi}^f \mathbf{C}_{iq} g_q^{-1/2} \\ &= \sum_{i=1}^k \sum_{q=1}^{k-1} \mathbf{Z}_{mi}^f \mathbf{C}_{iq} g_q^{-1/2} \neq 0. \end{aligned} \quad (12)$$

Equations (11) and (12) show that the last perturbation is a zero vector and the sum of all transformed perturbations (\mathbf{Z}_p^a) defined by eq. (8) is not zero, although the forecast perturbations (\mathbf{Z}^f) before transformation are centred. These two properties do not depend on the value of α . It is desirable that all initial perturbations are centred around the best possible analysis field in order to get better ensemble mean performance (Toth and Kalnay, 1997; Buizza et al., 2005; Wang et al., 2004; W06).

2.3. Centring perturbations

As shown in eq. (8), \mathbf{Z}_p^a are the perturbations, transformed by \mathbf{T}_p . These perturbations are not centred, the first $k - 1$ analysis perturbations are orthogonal in the norm described above and the last perturbation is zero. A transformation that will transform the $k - 1$ perturbations into k centred perturbations and preserve ensemble analysis covariance \mathbf{P}^a (see eq. 2) is the simplex transformation (ST) (Purser, 1996; Julier and Uhlmann, 2002; Wang et al., 2004; W06).

Let $\mathbf{E} = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_{k-1}]$, then $\mathbf{C} = [\mathbf{E}, \mathbf{c}_k]$. We can show that \mathbf{E}^T satisfies the conditions required for a simplex transformation as described in Wang et al. 2004. Equation (7) indicates that $\mathbf{E}^T \mathbf{1} = 0$, which is needed to centre the perturbations. Also $\mathbf{C}^T \mathbf{C} = \mathbf{I}$ means $\mathbf{E}^T \mathbf{E} = \mathbf{I}_{(k-1) \times (k-1)}$. This condition will keep the covariance matrix unchanged. The fact that the diagonal elements of $\mathbf{E} \mathbf{E}^T$ are equal will make the centred perturbations equally likely. This can be seen in the following subsection.

In the practical implementation, we use the transformation \mathbf{C}^T , which is a by-product from the eigenvalue solution in eq. (4), to act on all k perturbations \mathbf{Z}_p^a to produce k centred perturbations with simplex structure. This is equivalent to using \mathbf{E}^T acting on the first $k - 1$ orthogonal perturbations. Therefore, the final ET solution with ST is

$$\mathbf{Z}^a = \mathbf{Z}_p^a \mathbf{C}^T = \mathbf{Z}^f \mathbf{C} \mathbf{G}^{-1/2} \mathbf{C}^T. \quad (13)$$

Similar to eq. (10), let's look at the components of the new perturbations

$$\begin{aligned} \mathbf{Z}_{mq}^a &= \sum_{i=1}^k \mathbf{Z}_{mi}^f \sum_{l=1}^k \mathbf{C}_{il} g_l^{-1/2} \mathbf{C}_{lq}^T \\ &= \sum_{i=1}^k \mathbf{Z}_{mi}^f \sum_{l=1}^{k-1} \mathbf{C}_{il} \mathbf{C}_{lq} \lambda_l^{-1/2} + \frac{1}{k\sqrt{\alpha}} \sum_{i=1}^k \mathbf{Z}_{mi}^f. \end{aligned} \quad (14)$$

Since the second term on the right-hand side in above equation is zero, the final transformed k perturbations are not dependent on the value of α . In the experiments and the following NCEP

operational implementation (see the discussion in the end of the paper), we chose $\alpha = 1.0$ for simplicity. Following eq. (14), the sum of the final perturbations is:

$$\begin{aligned} \sum_{q=1}^k \mathbf{Z}_{mq}^a &= \sum_{i=1}^k \mathbf{Z}_{mi}^f \sum_{q=1}^k \sum_{l=1}^{k-1} \mathbf{C}_{il} \mathbf{C}_{ql} \lambda_l^{-1/2} \\ &= \sum_{i=1}^k \mathbf{Z}_{mi}^f \sum_{l=1}^{k-1} \mathbf{C}_{il} \lambda_l^{-1/2} \sum_{q=1}^k \mathbf{C}_{ql} = 0. \end{aligned} \quad (15)$$

Equation (7) is used in the last step of eq. (15). This shows that all perturbations after ET and ST transformations are centred.

2.4. Orthogonality of centred perturbations

Since all perturbations are centred, they are not strictly orthogonal as they were before ST. The ideal initial perturbations in an ensemble system must be centred and span a subspace that has maximum number of degrees of freedom. This will be further exploited numerically in our experiment. Let's now look at the orthogonality of the perturbations defined in eq. (13) in the following.

$$\begin{aligned} \mathbf{J} &= (\mathbf{P}_{op}^{a-1/2} \mathbf{Z}^a)^T (\mathbf{P}_{op}^{a-1/2} \mathbf{Z}^a) = \mathbf{Z}^{aT} \mathbf{P}_{op}^{a-1} \mathbf{Z}^a \\ &= \mathbf{C} \mathbf{G}^{-1/2} \mathbf{\Gamma} \mathbf{G}^{-1/2} \mathbf{C}^T \\ &= (\mathbf{E}, \mathbf{c}_k) \begin{pmatrix} \mathbf{I}_{(k-1)(k-1)} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{E}^T \\ \mathbf{c}_k^T \end{pmatrix} = \mathbf{E} \mathbf{E}^T. \end{aligned} \quad (16)$$

From $\mathbf{C} \mathbf{C}^T = \mathbf{I}$, one has $(\mathbf{E}, \mathbf{c}_k) \begin{pmatrix} \mathbf{E}^T \\ \mathbf{c}_k^T \end{pmatrix} = \mathbf{I}$. Consequently, eq. (16) results in

$$\mathbf{J} = \mathbf{E} \mathbf{E}^T = \mathbf{I} - \mathbf{c}_k \mathbf{c}_k^T \quad (17)$$

This equation shows that $\mathbf{J}_{ii} = 1 - 1/k$ and when $i \neq j$, we have

$$\mathbf{J}_{ij} = -\frac{1}{k}, \quad \lim_{k \rightarrow \infty} (\mathbf{J}_{ij}) = 0. \quad (18)$$

Equation (18) shows that for a finite number of ensemble members, the analysis perturbations after ET and ST transformations are not orthogonal. The properties of the initial perturbations generated from eq. (13) can be summarized as follows. (a) The initial perturbations will be centred around the analysis field to avoid degrading the score of the ensemble mean. (b) They have simplex, not paired, structure. The ST, which preserves the analysis covariance, ensures that the initial perturbations will have the maximum number of effective degrees of freedom (e.g. W06). The variance will be maintained in as many directions as possible within the ensemble subspace. (c) The perturbations are uniformly centred and distributed in different directions. The more ensemble members we have, the more orthogonal the perturbations will become. Equation (18) shows that if the number of ensemble members approaches infinity, then the transformed perturbations will be orthogonal under this norm. (d) Like the other perturbation generation methods used in

this study, the initial perturbations from ET have flow dependent spatial structure. (e) The covariance constructed from the initial perturbations is approximately consistent with the analysis covariance from the DA if the number of ensemble members is large.

The above properties of ET perturbations show that the ET method resembles breeding in that they both dynamically cycle the fastest growing non-linear perturbations. Unlike the SV method where perturbations are defined in a linear sense using the tangent linear model, the bred vectors evolve according to the dynamics of non-linear model. Linearity is assumed only in a limited sense when the non-linear perturbations are rescaled and repositioned to centre around the new analysis field. The bred vectors are generalizations of the dominant Lyapunov vectors. Dominant Lyapunov vectors together with the associated Lyapunov exponents are the fundamentals of non-linear dynamical systems; they characterize the intrinsic predictability of a dynamical system (Toth and Kalnay, 1993, 1997; Wei, 2000; Wei and Frederiksen, 2004). The ET method produces perturbations along the fastest growing directions that are constrained by the initial analysis error variances (eq. 3). The ET method can be considered as an extension of the well-established BM. In the special case where there are only two ensemble members, ET and breeding will produce the same perturbations.

2.5. 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 for ETKF are from the conventional data set in the NCEP global DA system. This conventional data set contains mostly rawinsonde and various aircraft data, and wind data from satellites. Details about the comparison between ETKF and breeding can be found in W06. The ETKF results displayed in most figures are mainly for comparison with various ET experiments. We also ran 10-member ET experiments with and without rescaling to compare with our previous experiments with breeding and ETKF.

In addition, we test ET experiments with more members. In particular, we run an 80-member ET at every cycle. However, due to the computing resource limit only 20 members will be integrated for long forecasts. The other 60 members are used only for cycling (integrated to 6 h). At every cycle, both ET and ST are imposed on all 80 members, followed by ST on the 20 members used for the long forecasts. At different cycles, a different 20 members will be used for long forecasts. A schematic of this configuration is depicted in Fig. 1. All the ensembles are cycled every 6 h in accordance with the NCEP DA system, in which new observations are assimilated in consecutive 6-h time windows centred at 00, 06, 12 and 18 UTC.

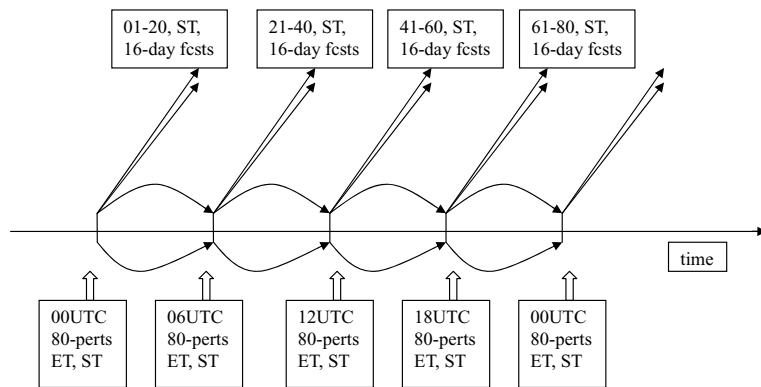


Fig. 1. Schematic of the configuration of the 80-member ET-based ensemble experiment. At each cycle ET transformation is carried out in all 80 perturbations, followed by the ST transformation. ST is also imposed on the 20 perturbations that will be used for long-range forecasts.

3. Results from ET with rescaling, ET, breeding and ETKF ensembles

3.1. Ensemble spread distribution

It is shown in W06 that an ETKF ensemble generation using real observations is able to produce initial perturbations that reflect the impact of observations even with only 10 members. 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 that are covered by satellite data are lower than in areas that are missed by the satellites. The impact of observations in ETKF was displayed in figs. 2 and 4 of W06. In the ET based ensemble system as described in the above section where the initial analysis variance is used to restrain the initial perturbations, it is natural to see if the initial spread distribution generated by the ET is influenced by the initial analysis variance. To make comparisons with our previous studies for ETKF and breeding, which were shown in fig. 2 of W06, we have computed the energy spread distribution for a 10-member ET with initial analysis variance drawn from the NCEP operational breeding mask (Toth and Kalnay, 1993, 1997). The mask in the NCEP operational global EFS represents the kinetic energy variance and is computed from a long time average of climatological data. It has lower scaling factors in the North American and Eurasian regions where traditionally there are more observations. Breeding initial spread is controlled by the mask, which was designed to reflect the long term average of analysis error variances.

Figure 2b shows the ratio of analysis and forecast spread averaged over all levels for ET. This ratio represents the rescaling factor from the forecast to analysis spread. It is clear that the rescaling factor distribution from ET based on eq. (13) is different from the rescaling factor in the 10-member breeding system (see fig. 2d of W06). Therefore, purely ET and ST transformations using the analysis error variances based on eq. (13) cannot restrain the initial spread distribution, although the analysis error variance decides the covariance structure of the ensemble perturbations. To have an initial spread distribution that is similar to the analysis error variance, we introduce ETR. In ETR, we impose a

regional rescaling process like operational breeding based EFS, that is, each initial perturbation after ET and ST from eq. (13) will be rescaled by the analysis error variance using

$$\mathbf{y}_m^a(i, j, l) = \alpha(i, j, l) \mathbf{z}_m^a(i, j, l), \quad (19)$$

where i, j, l are indices for the horizontal and vertical directions in grid point space; and $m = 1, 2, \dots, k$ is the index for the ensemble member. α is the rescaling factor derived from analysis error variance (\mathbf{P}_{op}^a) and the grid point values of analysis perturbations. The rescaling factor is defined as the ratio of the square root of kinetic energy from \mathbf{P}_{op}^a and the square root of the kinetic energy of analysis perturbations at each grid point. If the ratio is larger than one, the rescaling factor will be set to 1.0. For simplicity, we compute the rescaling factor for the model level at about 500 mb height. The same rescaling factor will be applied at all other levels (Toth and Kalnay, 1997). However, the rescaling factors can be computed and applied at all levels.

If the initial perturbations from eq. (13) are rescaled using eq. (19), we expect to get a distribution comparable to breeding. This result is shown in Fig. 2a. To test whether ET's failure to generate an initial spread similar to the analysis error variance is due to the small number of ensemble members, we calculate the ratio of analysis to forecast spread for temperature at 500 mb (T500) for an ensemble with 80 members as described in Fig. 1. The results with and without regional rescaling are shown in Fig. 2c and d, respectively. It appears that even with 80 members the ratio of analysis to forecast spread of T500 does not resemble that of the analysis error variance (Fig. 2d). After the regional rescaling is imposed as eq. (19), this ratio shows a strong similarity to that of the breeding ensemble (cf. fig. 2c and d of W06). In conclusion, the spread distribution of the initial perturbations generated by pure ET and ST transformations, based on eq. (13), does not reflect the initial analysis error variance without regional rescaling in spite of the good features summarized in the previous section.

To compute the vertical distributions of energy spread for the ensembles using different generation schemes, we average the energy spread of all grid points at each level. In Fig. 3a we show the vertical distributions of energy spread for the analysis

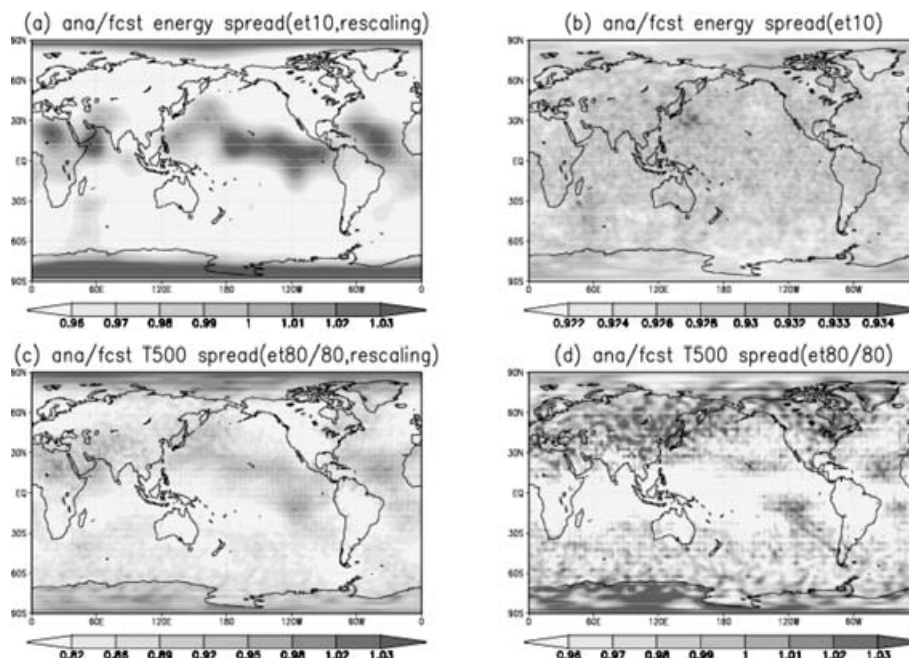


Fig. 2. Global distribution of the ratios of the analysis to forecast spread for ET based ensembles for (a) vertically averaged ratio of energy spread for a 10-member ET with rescaling; (b) vertically averaged ratio of energy spread for a 10-member ET without rescaling; (c) ratio of temperature spread at 500 mb for a 20 of 80 member ET with rescaling; and (d) ratio of temperature spread at 500 mb for a 20 of 80 member ET without rescaling.

(thick) and forecast (thin) perturbations from ETR (solid), ET (dotted), breeding (dashed) and ETKF (dash-dotted) ensembles. There are 10 members in all the ensembles. The results show that the analysis and forecast perturbations have the largest spread in terms of energy between 600 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 other three ensembles. The relatively larger spread in the ETKF is because the innovation-based inflation factor method did not work as ideally with real observations as with simulated observations (W06).

Figure 3b shows the energy spread distributions of analysis and forecast perturbations by latitude for 10-member ensemble systems using ETR, ET, breeding and ETKF. Unlike the vertical distribution in Fig. 3a, the latitudinal distributions of energy spread from ET and ETKF are similar with lower energy spread values near the tropics where baroclinic instability is relatively low, and a high spread near the North Pole. In the Southern Hemisphere, the ET and ETKF ensembles' energy spread have peak values at around 50° south, close to the southern ocean track region. However, different distributions are found in the ETR and the breeding ensembles. The spread distributions in these two systems are similar except for some differences in the tropics. Both ETR and breeding have lower energy spread values mainly in the Southern Hemisphere; in particular, both attain a minimum in the southern-ocean storm track area. The failure by

the ETR and the breeding ensembles to show higher spread in this region is related to the mask imposed on the system (Toth and Kalnay, 1997). Both ETR and breeding ensembles use the same rescaling method from the same mask. These results indicate that the mask used in our ensemble system needs to be improved. A more accurate time-dependent analysis error variance can be generated by a mature operational DA system like the NCEP 3-D VAR.

The temporal consistency of ensemble forecasts from one cycle to the next is also studied by computing the correlation between the forecast and analysis perturbations as shown in section 3.6 of W06. In that study, the temporal consistency was studied for the ETKF and breeding ensembles. In our current experiments, we also calculated the correlation between the forecast and analysis perturbations for ET and ETR. The results (not shown) indicate that ET without rescaling produces analysis perturbations with the highest correlations to the corresponding forecast perturbations. When rescaling is imposed, this correlation is decreased to a level similar to that of the ETKF perturbations. The breeding perturbations have a high time consistency with a correlation about 0.988, however, it is lower than the ET, ETR and ETKF.

3.2. Forecast error covariance

One good measure of ensemble forecast performance is a direct comparison of the ensemble perturbations to the forecast errors.

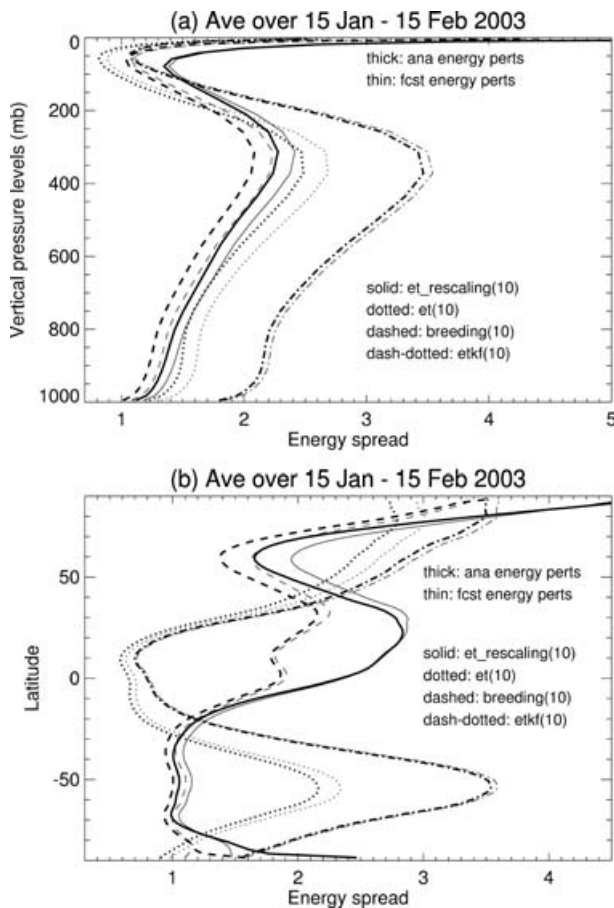


Fig. 3. Energy spread distributions of ET with rescaling (solid), ET without rescaling (dotted), breeding (dashed) and ETKF ensemble perturbations (thick: analysis; thin: forecast). All the ensembles have 10 members and values are averaged over the period 15 January–15 February 2003, with (a) vertical distribution as a function of pressure and (b) horizontal distribution by latitude.

We have computed the values of a measure called perturbation versus error correlation analysis (PECA). PECA measures how well ensemble perturbations can explain forecast error variance. It evaluates the performance of ensemble perturbations and perturbation generation technique. Apart from the PECA values averaged from individual perturbations, we also compute the PECA for the optimally combined perturbations. To do this, we linearly combine all the forecast perturbations so that the final combined perturbation is closest to the forecast error. A minimization problem will need to be solved for this. More details are available in Wei and Toth (2003).

The PECA values for 500 mb geopotential height for a 10-member ETR (solid), ET (dotted), breeding (dashed) and ETKF (dash-dotted) are shown in Fig. 4a–d for the globe, Northern and Southern Hemispheres, and the tropics. In each panel, the PECA for the optimally combined perturbations and the PECA

averaged from individual perturbations are displayed in thick and thin lines, respectively.

In each of these regions, ETR (solid) has the highest average PECA values (thin lines) for short lead times, with breeding (dashed) next. The gap between ETR and breeding is even larger for the optimally combined perturbations (thick). This is due to the structural difference between the two methods. The perturbations in ETR are simplex structures, while in breeding the positive/negative paired strategy is used. In a paired strategy, the effective number of degrees of freedom (EDF) of ensemble subspace is reduced by half by construction, while a simplex structure has a maximum EDF. It is interesting to see that the PECA values for both optimally combined and individual averages are similar for ET and ETKF. This is related to the fact that ET and ETKF have similar latitudinal distributions of energy spread (Fig. 3).

It is noteworthy that the rescaling imposed on the ET perturbations improves PECA values in almost all the domains we have chosen, particularly for the lead times up to a few days. In order to see the improvement in PECA from the increase of members, we compare a 10-member ET and a 20-of-80-member ET (see Fig. 1 for the configuration). In Fig. 5, we show PECA values for the 10-member ETR (solid) and ET (dotted), the 20-of-80-member ETR (dashed) and ET (dash-dotted) for Northern Hemisphere, North America, Europe and the globe. Again, the average PECA from the individual members and that from the optimally combined perturbations are indicated by thin and thick lines, respectively. It is clear that rescaling can increase the PECA value for a 20-member ensemble as well (see thick dashed and dash-dotted lines) as for a 10-member ET. Another message from this figure is that increasing the number of ensemble members will significantly increase the PECA value for optimally combined perturbations in all domains (thick solid versus dashed line; dotted versus dash-dotted).

Also plotted in Fig. 5 are the PECA values from the optimally combined perturbations for 80-member ETR (diamond) and ET (square) at a 6-h lead time. Since we have integrated only 20 members for the long forecasts due to computing resource limits, the remaining 60 members are integrated for only 6 h, for cycling. Again, rescaling increases the PECA values for the ET ensembles, especially for regions like North America, Northern Hemisphere and the globe. The difference between ET ensembles with and without rescaling is smaller over Europe. The PECA value for ETR is about 0.9 and 0.95 for North America and Europe, respectively. This means that the 80-member ET perturbations with rescaling can explain about 80–90% of forecast errors at 6-h lead time if the analysis error is small and can be neglected. In all domains, the optimally combined PECA values at a 6-h lead time from the 80-member ET are much larger than those from 20 members. This implies that the forecast error covariance at 6-h lead times constructed from the 80-member ET forecast perturbations will be a very good approximation to the real background covariance matrix, which can be used to

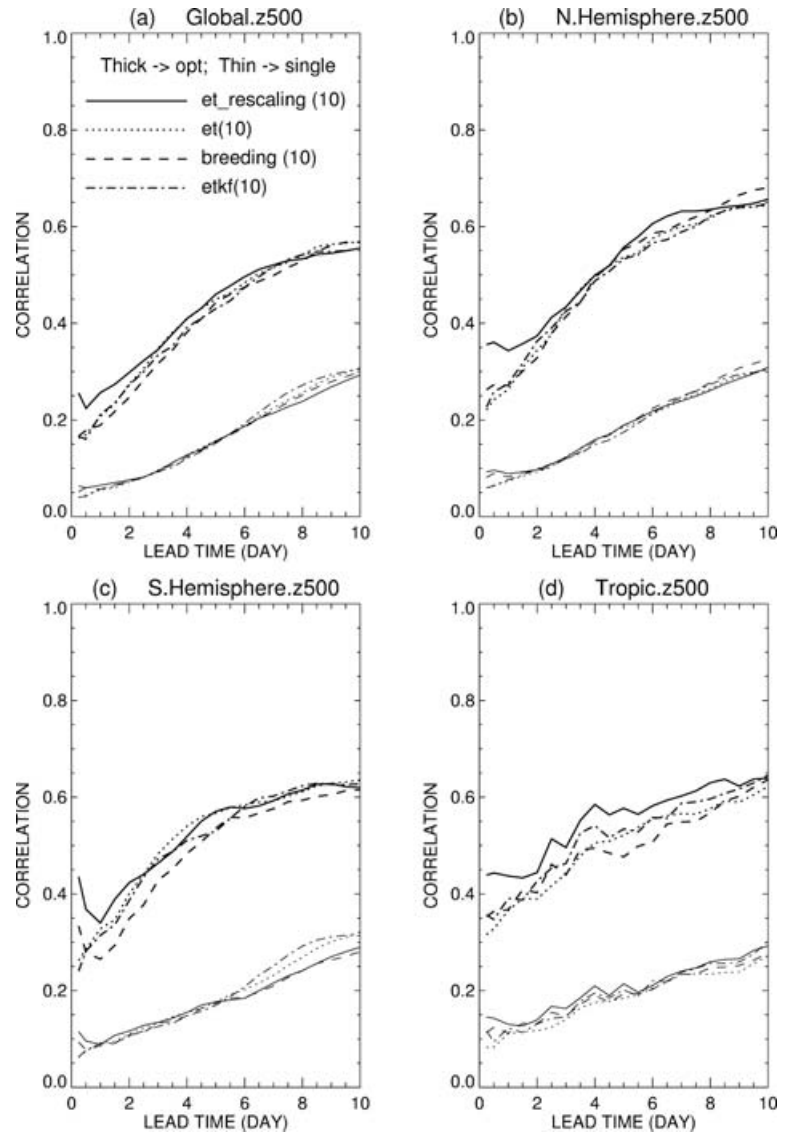


Fig. 4. PECA values for ET with rescaling (solid), ET without rescaling (dotted), breeding (dashed) and ETKF (dash-dotted) ensembles with 10 members for (a) the globe; (b) Northern Hemisphere; (c) Southern Hemisphere and (d) the tropics. Shown in thick and thin lines are PECA from the optimally combined perturbations and average PECA from the individual perturbations, respectively.

improve DA quality. In practice a covariance localization would have to be applied to ensemble before it is used in DA (Lorenc, 2003). Wei and Toth (2003) compared ensemble perturbations (from both NCEP and ECMWF) with the NMC method vectors that are commonly used to estimate background error covariance (Parrish and Derber, 1992). It was found that both NCEP and ECMWF perturbations are better able to explain the forecast errors than their respective NMC method vectors. Our long term goal at NCEP is to build an EFS that is consistent with the operational DA system. The DA system provides the best estimate of the analysis error variance needed to restrain initial perturbations for the EFS, while the ensemble system generates a better estimate of the background (6-h forecast) error covariance for the DA system.

Next we compare the ensemble variance and forecast error variance directly for the 10-member systems. To do this, we

compute ensemble variance and the squared forecast error of temperature for each grid point at the 500 mb pressure level for a 6-h lead time. A scatter plot can be drawn by using the ensemble variance (abscissa) and squared forecast errors for all grid points (not shown). Then we divide the points into 320 equally populated bins in order of increasing ensemble variance. The ensemble and forecast variances are averaged within each bin. It is the averaged values from all bins that are plotted. The ranges of forecast error variances that are associated the ranges of ensemble variances are the explained error variances statistically (see Majumdar et al., 2001, 2002; Wang and Bishop, 2003 and W06 for more details). Figure 6a shows the variance distributions in the Northern Hemisphere for a 10-member ETR and ET. The same is shown for a 10-member breeding ensemble and ETKF in Fig. 6b. In ET experiments, the performance is similar between ETR and ET, although ETR explains slightly less of the forecast

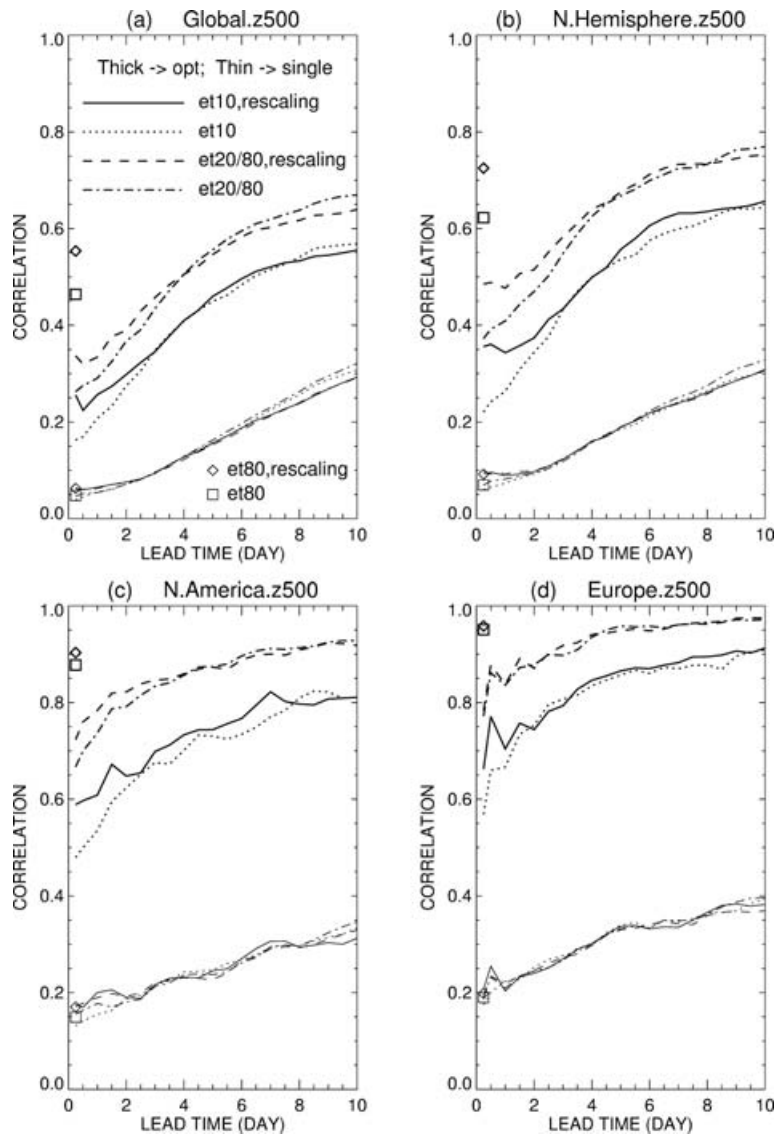


Fig. 5. PECA values for a 10-member ET with rescaling (solid), 10-member ET without rescaling (dotted), 20 of 80 member ET with rescaling (dashed) and 20 of 80 member ET without rescaling (dash-dotted) ensembles for (a) the globe; (b) Northern Hemisphere; (c) Southern Hemisphere and (d) the tropics. Shown in thick and thin lines are PECA from the optimally combined perturbations and average PECA from individual perturbations, respectively.

error variance in the Northern Hemisphere than pure ET without rescaling. The difference between them is larger over the global domain (not shown). We have tested the ET ensembles with more ensemble members, and found that increasing the membership will not increase the ensemble's ability to explain the forecast variance. Figure 6b shows the comparison between 10-member breeding and ETKF. For small amounts of variance, the breeding ensemble explains more of the forecast variance, while for larger forecast variance the ETKF ensemble is better.

3.3. Probabilistic forecasts

In this section, we will look at the probabilistic scores of the ensemble experiments we have done. Probabilistic scores have been frequently used for describing the performance of different ensemble systems (Buizza et al., 2005). Some scores, particularly

different skill scores described in the following will depend on the reference forecast. These scores will be different if a different reference forecast is used. The most commonly used reference forecast is the climatology (Wilks, 1995; Richardson, 2000; Zhu et al., 2002; Toth et al., 2003; Buizza et al., 2005). In this paper, climatology is also used as reference forecast in computing the probabilistic scores for all ensemble schemes studied. Please also note that our results will only focus on the 32-d period from 15 January 2003 to 15 February 2003. This is mainly due to the limitation of computing resources although longer period of experiments would be better. As a result, the scores reported here are less than ideal.

Since different probabilistic measures emphasize different aspects of ensemble forecasts, we will use several commonly used measures such as Brier Skill Score (BSS), Ranked Probability Skill Scores (RPSS), Economic Values (EV) and the area under the

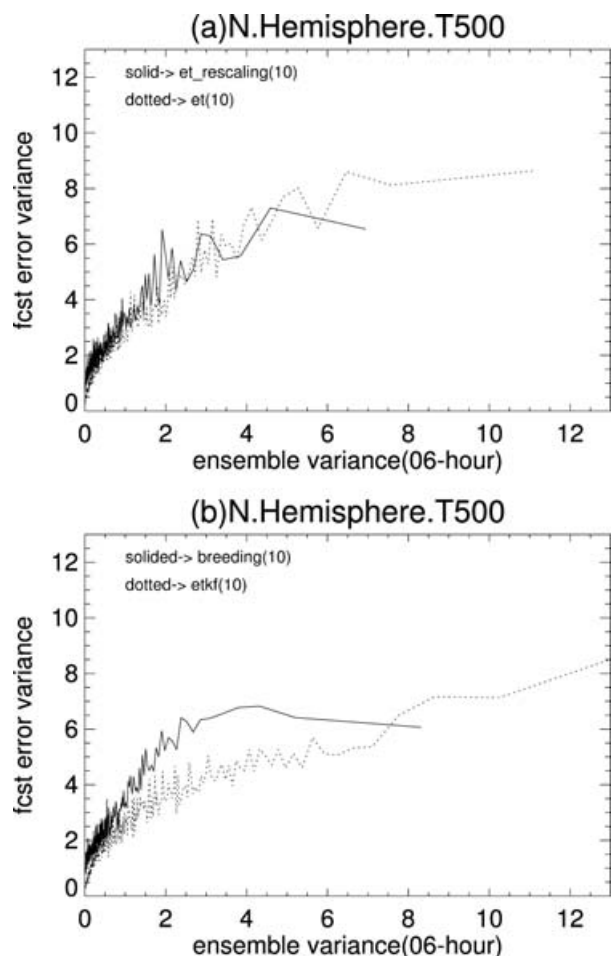


Fig. 6. Derived 10-member ensemble variance and forecast error variances at all grid points for 500 mb temperature over the Northern Hemisphere for (a) ET with rescaling (solid) and ET without rescaling; (b) breeding (solid) and ETKF (dotted).

Relative Operating Characteristic (ROC). One commonly used measure in probabilistic forecasts is the Brier score (BS). BS is actually the mean-squared error of the probability forecasts. It can be decomposed into reliability, resolution and uncertainty components (Wilks, 1995; Toth et al., 2003). However, it is the BSS that we normally prefer to use in measuring ensemble forecasts. BSS is a skill score based on BS, using climatology as a reference forecast. The maximum value of BSS is one, associated with a perfect forecast; while zero value indicates that the forecast is no better than climatology, the reference forecast. A common extension of BS to multi-event situations is the Ranked Probability Score (RPS). Unlike in the BS, the squared errors are computed with respect to the cumulative probabilities of the forecast and observation vectors. As with BSS, the Ranked Probability Skill Score (RPSS) based on RPS can also be defined by using climatology as the reference forecast (Wilks, 1995; Toth et al., 2003).

Economic value (EV) is based on a contingency table of losses and costs accrued by using ensemble forecasts, depending on the forecast and observed events (Richardson, 2000; Zhu et al., 2002). It also uses climatology as a reference forecast. ROC is based on 2×2 contingency tables containing the relative fractions of hits, misses, false alarms and correct rejections (Mason, 2003). The curve fitting method of Wilson (2000) is used in computing ROC. The ROC Area (ROCA) is the area under the ROC curve; the value of ROCA ranges from 1 for a perfect forecast to 0. A forecast with ROC area of 0.5 or less is not considered to be useful.

Figure 7 shows the Brier Skill Score (BSS) for 500 mb geopotential height over the Northern Hemisphere, which is calculated by using climatology as a reference forecast. The climatology forecast is divided into 10 equally likely (10%) events for computing the probability (see Richardson, 2000; Zhu et al., 2002; Toth et al., 2003 for more details). For shorter forecast lead times at least up to day 7, and for ensembles with 10 members ETR is best, while ETKF is the worst and breeding is in the middle. If we use 20 members out of the 80-member ETR as described in Fig. 1, its BSS value is higher than all the other experiments at all forecast lead times. The impact on Brier score by changing ensemble size was studied recently by Ferro (2007).

Shown in Fig. 8 is the ROCA for the same experiments over the Northern Hemisphere. ROCA is a measure of discrimination. The results show that a 10-member ETR is better than 10-member breeding, while a 10-member ETKF has the lowest value of ROCA. Again, when the ensemble membership is increased to 20 members out of 80-member ETR, the ROCA is significantly higher than for all the other three experiments with only 10 members. We have also computed the EV for all these ensemble systems, which is shown in Fig. 9. In terms of EV, the 10-member ETR is similar to the 10-member breeding, and both are better than the 10-member ETKF. Again, the 20 out of 80 member ETR is better than all the other ensembles.

4. A comparison of one-sided, paired and ET based ensembles

In addition to the experiments described above, we carried out two other experiments. One is a 20-member breeding ensemble. Instead of using the paired positive/negative perturbation strategy, we use a different centring method. In this method, we still use the operational mask to rescale the 20 forecast perturbations, followed by subtracting the mean of the perturbations from each perturbation. This will result in a new set of perturbations that are centred. We call it one-sided breeding, as compared with paired breeding. This experiment is referred to as ENS_c in Fig. 10. Another experiment is similar to the 20 out 80 member ETR. Instead of imposing ST on 20 members for each cycle, we impose ST on the first 10 members only. Then the negative parts of these 10 members are used to form 20 members for the long forecasts at each cycle. This experiment is referred to as ENS_p.

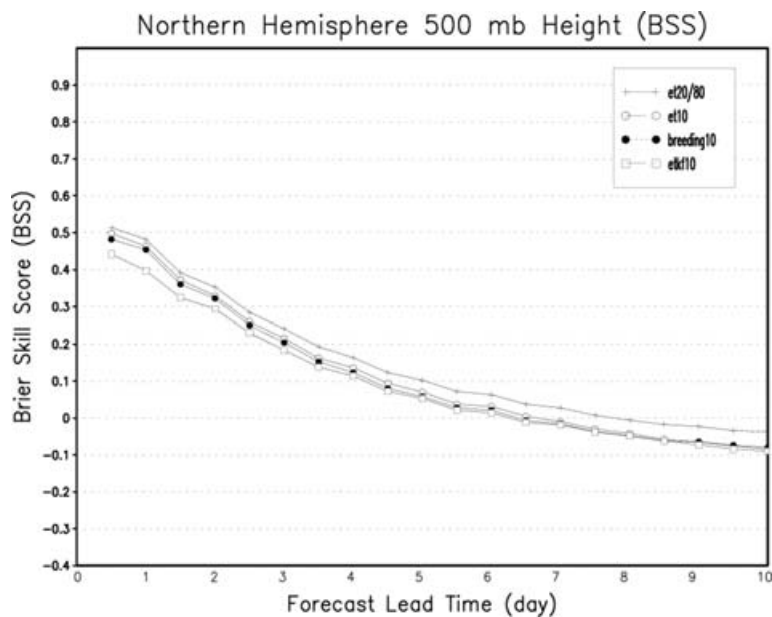


Fig. 7. Averaged Brier Skill Score of 500 mb geopotential height over the Northern Hemisphere for 20 of 80 member ET with rescaling (cross), 10-member ET with rescaling (open circle), 10-member breeding (full circle) and 10-member ETKF (open square) ensembles.

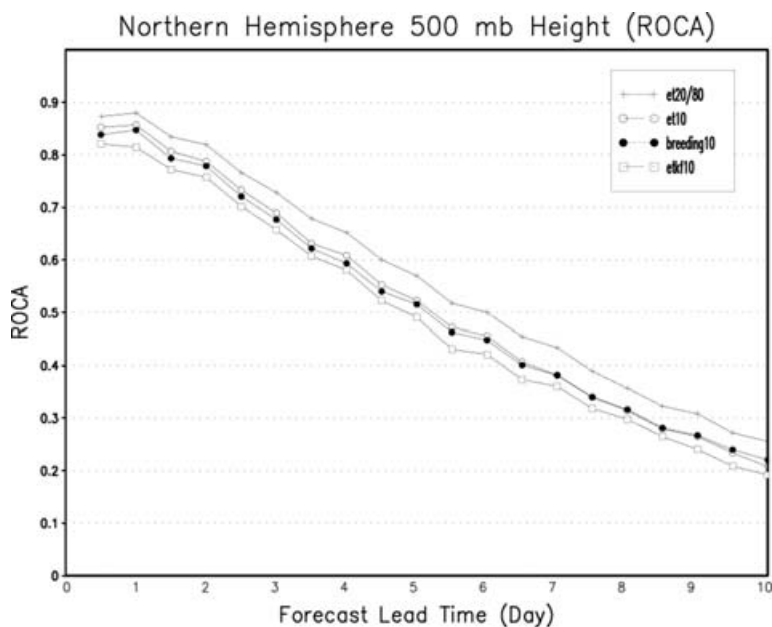


Fig. 8. The same as Fig. 7, but for the relative operating characteristic area.

in Fig. 10. The 20 out 80 member ETR experiment described back in Section 3 is indicated as ENS_s in Fig. 10.

Figure 10 shows the Ranked Probability Skill Score (RPSS) of 500 mb geopotential height over the Northern Hemisphere (NH). It appears that ENS_c for short lead times is similar to ENS_s, but is slightly better over the NH for larger lead times. The RPSS value of ENS_p is the lowest. These results show that either one sided breeding or ETR works better than the paired ensemble system. The paired ensemble also shows the worst scores in the other measures (not shown).

The results from other scores that are computed but not included here can be summarized as follows. In terms of the

range of forecast variance explained by the ensemble variance, as shown in Fig. 6, ENS_c is slightly better than ENS_s over the NH and globe, while they are similar in the Southern Hemisphere (SH). However, ENS_s has shown slightly better PECA values than ENS_c over the globe, NH, SH and the tropics (TR). RPSS values for ENS_s and ENS_c are similar over the SH, while ENS_s is better over the TR. In terms of BSS, ENS_s is similar to ENS_c over the NH and SH, but is better over the TR. ENS_s has higher values of ROC Area than ENS_c over the SH and TR. In the NH both have similar ROCA values. Both ensembles have shown similar EV values over the NH, SH and TR.

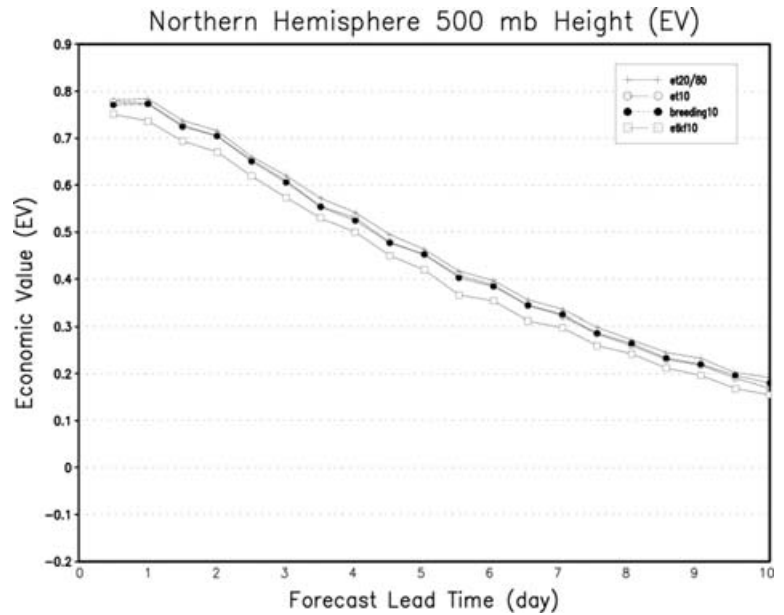


Fig. 9. The same as Fig. 7, but for the economic value.

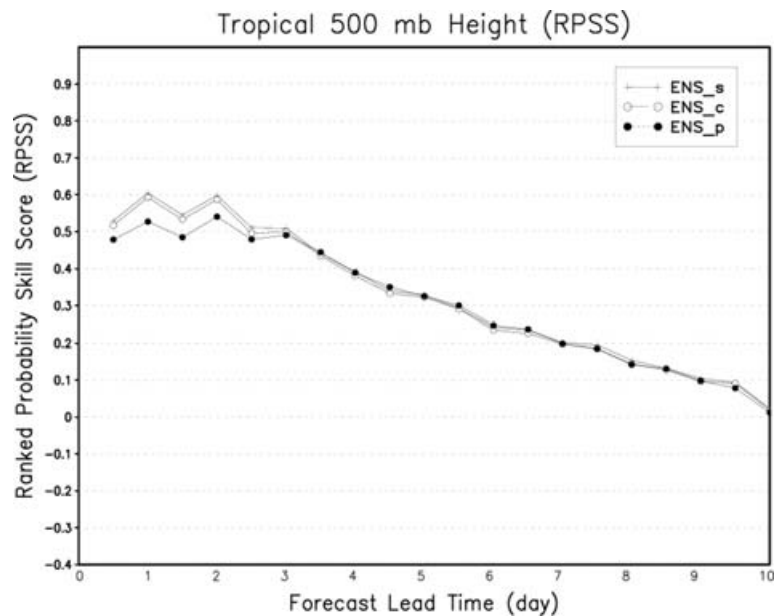


Fig. 10. Averaged Ranked Probability Skill Score of 500 mb geopotential height over the Northern Hemisphere for 20 of 80 member ET with rescaling (ENS_s, cross), 20-member one-sided breeding (ENS_c, open circle), and 20-member paired (full circle) ensembles.

The ENS_c one-sided breeding experiment has demonstrated good scores for some of the standard measures. In fact, its performance is better than was initially expected. Since each bred perturbation has to subtract the mean of all perturbations, the direction of each perturbation is going to be changed to a certain extent. To see how much change is going to be involved during the centring process, let us look at the temporal consistency of the perturbations. Again, we compute the correlations between the analysis and corresponding forecast temperature perturbations at the 500 mb level. The mean for 20 members is displayed in Fig. 11a as a function of time over the experimental period. As expected, the average value over the 32-d period is 0.906,

which is lower than for the breeding, ETKF and ET ensembles (see section 3.6 of W06 for more details). But the correlation value is still reasonably high.

In order to understand how independent the 20 perturbations in the ENS_c experiment are, we follow W06 and compute the effective number of degrees of freedom (EDF) of the subspace spanned by the 20 temperature analysis perturbations at the 500 mb geopotential height level. These results are shown in Fig. 11b. The average EDF value over the experimental period is 16.187, which is much higher than the paired ensembles, as expected. This is a little lower than those in ETKF and in ET with a simplex transformation. This result indicates that removing the

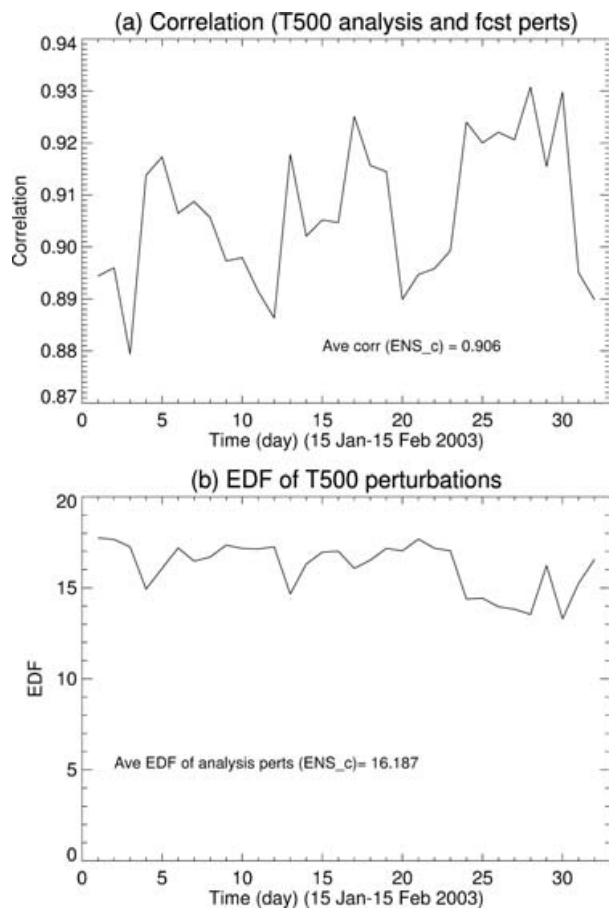


Fig. 11. (a) Average correlation over 20 members between the temperature forecasts and analysis perturbations at 500 mb geopotential height for a one-sided breeding ensemble. (b) The effective number of degrees of freedom of subspace spanned by 20 temperature analysis perturbations from a one-sided breeding ensemble.

mean of all perturbations from each perturbation does not alter the directions of the perturbations too much. Using one-sided breeding with this kind of simple centring strategy can significantly increase the EDF of the subspace spanned by the bred perturbations, compared to ordinary breeding with the paired perturbations. The better performance shown by this experiment with one-sided breeding is probably related to the increase of the EDF of the ensemble subspace.

5. Discussion and conclusions

In this paper, we have carried out several experiments with four different initial perturbation generation techniques: breeding, ETKF, ET and ETR. All of these are based on the principle of the BM dynamically cycling non-linear forecast perturbations. As in W06, results are presented for a 32-d experimental period using the NCEP operational analysis/forecast and observation systems. Beyond a detailed description of the theoretic

cal formulations of ET and ETR in generating initial perturbations, this paper also provides a comprehensive description of the performance of these techniques in terms of various commonly used measures—including probabilistic scores in an operational environment.

Various aspects of the properties of ETKF-generated perturbations using NCEP real observations have been studied in W06. In this paper, we concentrate on the ensembles generated by ET and ETR, and compare them to NCEP operational breeding. For scientific interest, in some figures we also include results with the ETKF from our previous study. Both ET and ETR are second generation techniques attempting to better link DA and EFS.

Apart from the four techniques that we have focused on in this paper, we also tested a one-sided BM. In this method, instead of pairing perturbations, the bred vectors are centred by simply removing the mean of all perturbations at each cycle. The results are compared with the ETR and two-sided positive/negative paired perturbation ensembles. For years, there has been some confusion within the ensemble forecasting research community as whether the one-sided perturbations or paired (positive plus negative) strategy should be used in operational forecasts. That issue is addressed in this paper as well.

Based on our experiments with different methods, our findings can be summarized as follows:

- (1) The ET/ETR method is an extension of breeding and is similar to breeding in that they both dynamically cycle the perturbations. In an ensemble with only two members, both methods should produce the same results.

- (2) Initial perturbations from ET and ST have simplex, not paired, structure. The ST, which preserves the analysis covariance, ensures that the initial perturbations will have the maximum number of effective degrees of freedom. The variance is maintained in as many directions as possible within the ensemble subspace. The perturbations are uniformly centred and distributed in different directions. The more ensemble members we have, the closer to being orthogonal the perturbations will be. In the limit of infinite number of ensemble members, the perturbations will be exactly orthogonal.

- (3) A purely ET method with ST cannot produce initial perturbations with a variance distribution that is similar to the initial analysis variance provided by the DA system, as desired. ETR can generate initial perturbations that have a variance distribution similar to the analysis variance, and maintains the large EDF of the ensemble subspace generated by ET and ST.

- (4) An important finding from this study is the difference in geographical distribution of spread in energy as a function of latitude. The energy spread distribution for ET without rescaling is surprisingly similar to the ETKF, with lower values in the tropics and higher spread in the extra-tropics of both hemispheres. On the other hand, the energy spread for ETR and breeding have higher values in the tropics and lower values in the

extra-tropics. The vertical distributions of energy spread for ETR and ET, breeding, and ETKF are similar.

(5) PECA results show that ET perturbations can explain an amount of forecast error similar to the breeding and ETKF perturbations, while the ETR has higher PECA values than the other three perturbations types over all regions at shorter forecast lead times. For larger lead times, the gap gets smaller. When the number of ensemble members is increased, the PECA value for the optimally combined perturbation is increased significantly. ETR performs better than ET without rescaling independent of ensemble size. When 80 perturbations are used, optimally combined perturbations from ETR can explain about 80–90% of the forecast error at a 6-h lead time over smaller regions like North America and Europe. This implies that the 80-member ensemble may be able to provide an efficient background covariance for the DA system.

PECA values quantitatively measure how well linear combinations of ensemble perturbations match the forecast errors (Wei and Toth, 2003). At longer forecast lead times any perturbation, including ET and ETKF ensemble perturbations, will turn towards the leading Lyapunov vectors that are linked to the bred vectors (Toth and Kalnay, 1997; Wei, 2000; Wei and Frederiksen, 2004).

(6) ET forecast error variance predictions were better than the corresponding breeding predictions by distinguishing times and locations with larger forecast errors from times and locations with smaller forecast error. When rescaling is imposed, this ability of variance prediction is downgraded slightly.

(7) All ensemble systems based on the four techniques produce temporally highly consistent perturbation fields.

All perturbations have a very high correlation with forecast perturbations before the transformations, with ET the highest and breeding the lowest. The correlations for ETR and ETKF are in the middle. The advantages of high temporal consistency in EFS were discussed in W06.

(8) In terms of probabilistic forecast capability, ETR has higher scores than breeding and ETKF in BSS, ROCA, EV and RPSS for the same number of ensemble members. Increasing the number of ensemble members generally increases all of these scores.

(9) Pattern anomaly correlations for the ensemble mean forecasts for all the EFS had only slight differences.

Results show that the breeding ensemble mean has a slightly higher pattern anomaly correlation than the ETKF ensemble or ETR and ET in the Northern Hemisphere for extended forecast lead times. However, it seems that this difference is not statistically significant. The anomaly correlation may be influenced by the magnitude and geographical distribution of the initial perturbation variance, as well as by the use of symmetric centring in

the paired breeding scheme or spherical simplex centring in the other schemes.

(10) Experiments using one-sided breeding have performed significantly better the paired perturbations with performance almost as high as that with ETR. This one-sided breeding system also has relatively good time consistency between the analysis and forecast perturbations. The EDF of ensemble subspace is also high.

The good performance by the one-sided breeding is related to the ensemble centring strategy. By simply removing the mean from all perturbations, the independence of these perturbations is preserved. The paired centring scheme reduces the EDF of the ensemble subspace by half, which may result in worse probabilistic scores. The simplex centring strategy used in the ET and ETR maximizes the EDF of ensemble subspace. This may contribute to the fact that ETR generally produces slightly higher probabilistic scores than one-sided breeding.

Our goal at NCEP is to build an EFS that is consistent with the DA system. The DA system provides an accurate analysis error variance for EFS in an operational environment using real observations, while the EFS can feed back the background covariance information into the DA system. This study is a step towards this goal.

Another choice is to use ensemble DA like ETKF which is in fact one of the ensemble-based Kalman square-root filters (Anderson, 2001; Whitaker and Hamill, 2002; Tippett et al., 2003; Ott et al., 2004; Zupanski, 2005). The quality of a DA system also depends on how the model errors and bias are taken into account. A new method of dealing with the model errors and bias has been proposed recently by Toth and Pena (2006). The new mapping paradigm is shown to be capable of reducing model-related errors greatly.

During the last 2 yr good progress has been made in ensemble DA (Szunyogh et al., 2005; Whitaker et al., 2006, personal communication). For example, Whitaker et al. (2006, personal communication) showed good results with conventional data, and satellite data will be tested. If the future experiments using all satellite data can produce better results and are more efficient than NCEP's operational 3D-Var DA system, the ensemble system could be implemented in a few years' time. Right now, NCEP's 3D-Var operational DA system (Parrish and Derber, 1992) provides the best estimate of the analysis error variance in an operational environment. To make use of the best available analysis error variance from the NCEP operational DA system, ETR is the best choice for generating initial perturbations for NCEP global EFS.

By the time that we completed the final version of this manuscript, the ETR method had been adopted and implemented successfully at NCEP on May 30, 2006 for operational forecasts. Due to the limitations on computing resources at the time of the implementation, the NCEP global EFS runs only 56 ET-generated members for the four daily cycles at 00Z, 06Z, 12Z

and 18Z. For each cycle, only 14 members are integrated for the 16 d forecasts. The NCEP operational configuration has been switched to that described in Fig. 1 of this paper since March 27, 2007. This was due to the arrival of new NCEP supercomputers in early 2007.

6. Acknowledgments

We are grateful to many colleagues at NCEP/EMC for their help during this work. Particularly, we thank Craig Bishop and Jim Purser for many helpful discussions. We are very thankful to David Parrish and Jun Du for their useful suggestions to the manuscript, and Mary Hart for improving the presentation. We thank two anonymous reviewers for their comments and suggestions that have helped improving the manuscript.

References

- Anderson, J. L. 2001. An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. 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. Roy. Meteor. 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. Wea. Rev.*, **129**, 420–436.
- Bowler, N. E. 2006. Comparison of error breeding, singular vectors, random perturbations and ensemble Kalman filter perturbation strategies on a simple model. *Tellus*, **58A**, 538–548.
- 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. Wea. Rev.*, **133**, 1076–1097.
- Errico, R., Yang, R., Masutani, M. and Woollen, J. 2007. The use of an OSSE to estimate characteristics of analysis error. *Meteorologische Zeitschrift*, in press.
- Ferro, C. A. T. 2007. Comparing probabilistic forecasting systems with the Brier Score. *Wea. Forecast.*, in press.
- Fisher, M. and Andersson, E. 2001. Development in 4D-Var and Kalman Filtering. ECMWF Technical Memorandum, No. 347. 36 pp.
- 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., Lefaiavrem, L., Derome, J., Ritchie, H. and Mitchell, H. L. 1996. A sytem simulation approach to ensemble prediction. *Mon. Wea. 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. In: *Proc. IEEE American Control Conf.*, Anchorage, AK, IEEE, 887–892.
- Lorenc, A. C. 2003. The potential of ensemble Kalman filter for NWP-a comparison with 4D-Var. *Quart. J. Roy. Met. 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? *Quart. J. Roy. Met. 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. Wea. Rev.*, **130**, 1356–1369.
- Mason, I. B. 2003. Binary Events. In: *Forecast Verification: A Practitioner's Guid in Atmospheric Science* (eds Ian T. Jolliffe and David B. Stephenson). John Wiley & Sons Ltd., England, 37–76.
- Molteni, F., Buizza, R., Palmer, T. and Petroliagis, T. 1996. The ECMWF ensemble prediction system: methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119.
- Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J. and co-authors. 2004. A Local ensemble Kalman filter for atmospheric data assimilation. *Tellus*, **56A**, 415–428.
- Parrish, D. F. and Derber, J. 1992. The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763.
- 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).
- Richardson, D. S. 2000. Skill and relative economic value of the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **126**, 649–667.
- 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. Wea. Rev.*, **131**, 1485–1490.
- Toth, Z. and Kalnay, E. 1993. Ensemble forecasting at NMC: the generation of perturbations. *Bull. Amer. Meteor. Soc.*, **174**, 2317–2330.
- Toth, Z. and Kalnay, E. 1997. Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319.
- Toth, Z., Talagrand, O., Candille, G. and Zhu, Y. 2003. Probability and ensemble forecasts. In: *Forecast Verification: A Practitioner's Guid in Atmospheric Science* (eds Ian T. Jolliffe and David B. Stephenson). John Wiley & Sons Ltd., England, 137–163.
- Toth, Z. and Pena, M. 2006. Data assimilation and numerical forecasting with imperfect models: the Mapping Paradigm. *Physica D*, **13** p., doi:10.1016/j.physd.2006.08.016.
- 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**, 1590–1605.
- 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 Toth, Z. 2003. A new measure of ensemble performance: perturbations versus Error Correlation Analysis (PECA). *Mon. Wea. Rev.*, **131**, 1549–1565.
- Wei, M. and Frederiksen, J. S. 2004. Error growth and dynamical vectors during southern hemisphere blocking. *Nonl. Proc. Geoph.*, **11**, 99–118.
- Wei, M., Toth, Z., Wobus, R., Zhu, Y. and Bishop, C. H. 2005a. Initial perturbations for NCEP Ensemble Forecast System. In: *Thorpe Symposium Proceedings for the First THORPEX Internal Science*

- Symposium 6-10 December 2004*, Montreal, Canada. The Symposium Proceedings in a WMO Publication 2005, **WMO TD No.1237, WWRP THORPEX No. 6**, 2005. p227–230.
- Wei, M., Toth, Z., Wobus, R., Zhu, Y., Hou, D. and co-authors. 2005b. NCEP Global Ensemble: recent developments and plans. In: *2nd SRNWP Workshop on Short Range Ensemble*, Bologna, Italy, 7–8 April, 2005. Available at http://smwp.cscs.ch/Lead_Centres/2005Bologna/Agenda.htm
- Wei, M., Toth, Z., Wobus, R., Zhu, Y., Bishop, C. H. and Wang, X. 2006a. Ensemble Transform Kalman Filter-based ensemble perturbations in an operational global prediction system at NCEP. *Tellus*, **58A**, 28–44.
- Wei, M., Toth, Z., Wobus, R., Zhu, Y. and Bishop, C. H. 2006b. The Use of Ensemble Transform Technique for Generating Initial Ensemble Perturbations. *NOAA Thorpex PI Workshop* at NCEP, Camp Springs, Maryland. Jan. 17–19, 2006. Available at <http://www.emc.ncep.noaa.gov/gmb/ens/THORPEX/PI-shop-2006.html>
- Wei, M., Toth, Z., Wobus, R. and Zhu, Y. 2006c. Initial Perturbations based on the Ensemble Transform (ET) Technique in the NCEP global Ensemble Forecast System. *US Department of Commerce, NOAA/NCEP Office Note No. 453*, 33 pp.
- Whitaker, J. S. and Hamill, T. M. 2002. Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924.
- Wilks, D. S. 1995. *Statistical Methods in the Atmospheric Sciences*. Academic Press, London, 464 p.
- Wilson, L. J. 2000. Comments on “Probabilistic Predictions of Precipitation Using the ECMWF ensemble prediction system”. *Wea. Forecast.*, **15**, 361–369.
- Zhu, Y., Toth, T., Wobus, R., Richardson, D. and Mylne, K. 2002. The economic value of ensemble-based weather forecasts. *Bull. Amer. Meteor. Soc.* **83**, 73–83.
- Zupanski, M. 2005. Maximum likelihood ensemble filter: theoretical aspects. *Mon. Wea. Rev.*, **133**, 1710–1726.