U.S. Department of Commerce
National Oceanic and Atmospheric Administration
National Weather Service
National Centers for Environmental Prediction
5200 Auth Road
Camp Springs, MD 20746

Office Note 453

Initial Perturbations Based on the Ensemble Transform (ET) Technique in the NCEP Global Ensemble Forecast System

Mozheng Wei* , Zoltan Toth, Richard Wobus and Yuejian Zhu
NOAA/NWS/NCEP, Environmental Modeling Center, Camp Springs, Maryland

September 19, 2006

* Corresponding author address: Mozheng Wei, NCEP Environmental Modeling Center, 5200 Auth Road, Rm. 207, Camp Springs, MD 20746.
E-mail: Mozheng.Wei@noaa.gov

ABSTRACT

Limitations on the initial perturbations used in the global operational ensemble forecast system at the National Centers for Environmental Prediction (NCEP) include the use of a climatologically fixed estimate of the analysis error variance and nonorthogonal paired bred vectors. In order to address these shortcomings, we introduced initial perturbations generated by the Ensemble Transform (ET) and ET with rescaling methods and compared them with the breeding ensemble in the NCEP operational environment.

Both ET and ET with rescaling are second generation methods and generate initial perturbations that are consistent with the operational data assimilation (DA) system. The best possible initial analysis error variance from DA is used to restrain the initial perturbations that are orthogonal with respect to an inverse analysis error variance norm. In addition, a simplex transformation (ST) is imposed to ensure that the initial perturbations are centered and span a subspace with the maximum number of degrees of freedom. The variance is maintained in as many directions as possible within the ensemble subspace. It is shown that the perturbations are uniformly centered and distributed in different directions. The more ensemble members we have, the more orthogonal the perturbations will become. In the limit of infinite number of ensemble members, the perturbations will be orthogonal.

Results show that the overall difference is not large although ET with rescaling performs best in almost all probabilistic scores and in terms of the forecast error explained by the perturbations. The forecast error variance can be explained best by pure ET with ST, which also has the highest time consistency between the analysis and forecast perturbations. The anomaly correlation of the ensemble mean from the breeding ensemble is slightly higher than that of the others for longer forecast lead times.

A new one-sided breeding, which is centered by removing the mean of all perturbations, is tested for the same experimental period. It shows higher probabilistic scores than the paired ensemble. It is also found that the one-sided bred perturbations span a high number of degrees of freedom and show a strong high time consistency.

1. Introduction

Ensemble forecasts start from a set of different initial states that are constructed from a group of initial perturbations. These initial perturbations are supposed to be sampled from a probability density function that represents the initial-value related uncertainty. However, how to best generate these initial perturbations for an ensemble forecasting system (EFS) is still a research issue.

The first operational EFSs were implemented in 1992 at the US National Centers for Environmental Prediction (NCEP) and the European Center for Medium-Range Weather Forecasts (ECMWF) (Buizza *et al.* 2005). Singular vectors (SVs) are used at ECMWF to identify the directions of fastest forecast error growth for a finite time period in the total energy norm (Buizza and Palmer, 1995; Molteni *et al.* 1996). Bred vectors (BVs) are used at NCEP to sample amplifying analysis errors through breeding cycles. Generation of BVs are similar to data assimilation (DA) cycles (Toth and Kalnay 1993; 1997). Frederiksen *et al*. (2004) successfully used a breeding based ensemble for predicting blocking events in the Southern Hemisphere. However, both SVs and BVs do not represent the true uncertainties in analysis, as one expects from an ideal EFS. Initial perturbations based on both

SVs and BVs are not consistent with the DA systems that generate the analysis fields. Comparisons of performance between the ECMWF and NCEP EFSs were described in Zhu *et al.* (1996, personal communication) and in Wei and Toth (2003). Both BV- and SV- based techniques at NCEP and ECMWF can be classified as *first generation* initial perturbation techniques as listed in Table 1.

Another major method among first generation techniques is the perturbed observation (PO) approach developed at the Meteorological Service of Canada (MSC) (Houtekamer et al. 1996). The PO based EFS became operational at MSC in 1998. In this approach, the initial conditions are generated by assimilating randomly perturbed observations 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. By using a quasigeostrophic channel model, Hamill *et al.* (2000) compared the performance of these three methods based on BVs, 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 basic properties, including both advantages and disadvantages, are summarized in Table 1.

A new initial perturbation technique that is based on the Ensemble Transform (ET) method was first presented by the first author in the First THORPEX International Science Symposium, Montreal, Canada, 6-10 December 2004 and published by WMO (Wei *et al.* 2005a), Some preliminary experimental results have been shown in Wei *et al.* (2005b, 2006a). The main properties of the ET based initial perturbations are briefly summarized by Wei (2006). In this note, we give a detail description of the ET formulation and comprehensive results. Results from the EFS based on ET will be compared with those from Ensemble Transform Kalman Filter (ETKF) and NCEP operational BV-based ensemble systems. Both ET and ETKF based techniques may be classified as *second generation* initial perturbation techniques for reasons that will be explained below. 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. The ETKF and ET methods were proposed initially by Bishop *et al.* (2001) and Bishop and Toth (1999), respectively, for adaptive observation studies. Wang and Bishop (2003) later 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.* (2006b) (referred to as W06). By using a much lower-order Lorenz 95 model, Bowler (2005) compared different initial perturbation generation techniques including ETKF, error breeding, singular vector, random perturbation and ensemble Kalman filter methods.

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 the ensemble and DA systems. A good DA system will provide accurate estimates of the initial analysis error variance for the ensemble system, while a good, reliable ensemble system will produce accurate flow dependent background covariance for the DA system.

ETKF is in fact one of the ensemble-based Kalman square-root filters (Tippett *et al.* 2003). However, at NCEP the quality of DA from ETKF has yet to be demonstrated to be superior to a mature 3D-VAR system with real observations, such as the NCEP operational 3D-VAR system. In this case, the perturbations have to be centered about the operational analysis field generated by the operational 3D-Var. Other closely related ensemble-based Kalman filters are the Ensemble Adjustment Kalman Filter (EAKF, Anderson, 2001), Ensemble Square-root Filter (EnSRF, Whitaker and Hamill, 2002) and Maximum Likelihood Ensemble Filter (MLEF, Zupanski, 2005). A Local Ensemble Kalman Filter (LEKF) was proposed by Ott *et al.* (2004) (also see Szunyogh *et al.*, 2005). ETKF, EAKF, EnSR, MLEF and LEKF are all deterministic solutions of ensemble Kalman filters. 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). A new mapping paradigm introduced in this method is shown to be capable of reducing model-related errors greatly.

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). To get better performance before DA quality from the ETKF outperforms the NCEP 3D-VAR system, the ensemble cloud has to be adjusted to center around the analysis field generated by the reliable 3D-VAR system as shown in W06). In the ET method, the initial perturbations are restrained by the best available analysis variance from the operational DA system and centered 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 of ET is that the ET technique is considerably cheaper than ETKF if the analysis variance information is available from a mature DA system.

Section 2 provides a brief basic description of the ET formulation for initial perturbations. Section 3 presents the major results from comparisons of ET 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 breeding method with regional rescaling. This method is well established, documented and widely used. It dynamically recycles perturbations and is a nonlinear 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 breeding method 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 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. 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 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 quality of DA from ETKF has yet to be improved to the level of a mature variational DA like the NCEP SSI (Parrish and Derber, 1992), the perturbations generated by ETKF have to be centered 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

$$\mathbf{Z}^f = \frac{1}{\sqrt{k-1}}[\mathbf{z}_1{}^f, \mathbf{z}_2{}^f, ..., \mathbf{z}_k{}^f], \qquad \mathbf{Z}^a = \frac{1}{\sqrt{k-1}}[\mathbf{z}_1{}^a, \mathbf{z}_2{}^a, ..., \mathbf{z}_k{}^a] \tag{1}$$

where the *n* dimensional state vectors $\mathbf{z}_i{}^f = \mathbf{x}_i{}^f - \mathbf{x}^f$ and $\mathbf{z}_i{}^a = \mathbf{x}_i{}^a - \mathbf{x}^a$ $(i = 1, 2, ...k)$ are k ensemble forecast and analysis perturbations, respectively. 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 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

$$\mathbf{P}^f = \mathbf{Z}^f \mathbf{Z}^{f^T} \quad \text{and} \quad \mathbf{P}^a = \mathbf{Z}^a \mathbf{Z}^{a^T}, \tag{2}$$

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

$$\mathbf{Z}^a = \mathbf{Z}^f \mathbf{T} \tag{3}$$

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 all kinds of real-time observations.

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 equation (1). One can solve the following eigenvalue problem.

$$\mathbf{Z}^{f^T} \mathbf{P}_{op}{}^{a^{-1}} \mathbf{Z}^f = \mathbf{C} \boldsymbol{\Gamma} \mathbf{C}^{-1} \tag{4}$$

Where $\mathbf{C}$ contains column orthonormal eigenvectors ($\mathbf{c}_i$) of $\mathbf{Z}^{f^T} \mathbf{P}_{op}{}^{a^{-1}} \mathbf{Z}^f$ (also the singular vectors of $\mathbf{P}_{op}{}^{a^{-1/2}} \mathbf{Z}^f$), and $\boldsymbol{\Gamma}$ is a diagonal matrix containing the associated eigenvalues (λ_i) with magnitude in decreasing order, that is, $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_k]$, $\mathbf{C}^T \mathbf{C} = \mathbf{I}$ and $\boldsymbol{\Gamma} = diag(\lambda_1, \lambda_2, ..., \lambda_k)$. Since the forecast perturbations are centered around the mean, i.e. $\sum_{i=1}^{k} \mathbf{z}_i{}^f = 0.0$, the sum of the columns of $\mathbf{P}_{op}{}^{a^{-1/2}} \mathbf{Z}^f$ is zero. Thus,

$$\mathbf{Z}^{f^T} \mathbf{P}_{op}{}^{a^{-1}} \mathbf{Z}^f \mathbf{1} = \mathbf{C} \boldsymbol{\Gamma} \mathbf{C}^T \mathbf{1} = 0 \tag{5}$$

where $\mathbf{1} = (1,1,.....1)^T$. Equation (5) indicates that that the last eigenvalue is zero and its associated eigenvector is a constant, i.e. $\lambda_k = 0.0$ and since the each eigenvector is normalized, one has $\mathbf{c}_k = (\frac{1}{\sqrt{k}}, \frac{1}{\sqrt{k}}, ... \frac{1}{\sqrt{k}})^T$. From equation (5), we have $\mathbf{\Gamma C}^T \mathbf{1} = 0$, which means that

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

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

$$\sum_{i=1}^{k} \mathbf{C}_{ij} = 0, \quad \mathrm{j} = 1, 2, ... \mathrm{k\text{-}1} \tag{7}$$

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

Now suppose $\mathbf{F} = diag(\lambda_1, \lambda_2, ..., \lambda_{k-1})$ and $\mathbf{G} = diag(\lambda_1, \lambda_2, ..., \lambda_{k-1}, \alpha)$ where α is a nonzero constant, i.e. $\mathbf{G} = diag(g_1, g_2, ..., g_k) = \begin{pmatrix} \mathbf{F} & 0 \\ 0 & \alpha \end{pmatrix}$ 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

$$\mathbf{Z}_p{}^a = \mathbf{Z}^f \mathbf{T}_p = \mathbf{Z}^f \mathbf{C}\mathbf{G}^{-1/2} \tag{8}$$

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

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

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, i.e. when $k \to n$.

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

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

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

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

The sum of the analysis perturbations is

$$\sum_{q=1}^{k} (\mathbf{Z}_p{}^a)_{mq} = \sum_{q=1}^{k} \sum_{i=1}^{k} \mathbf{Z}_{mi}{}^f \mathbf{C}_{iq} g_q{}^{-1/2} = \sum_{i=1}^{k} \sum_{q=1}^{k-1} \mathbf{Z}_{mi}{}^f \mathbf{C}_{iq} \lambda_q{}^{-1/2} \neq 0 \tag{12}$$

Equations (11) and (12) show that the last perturbation is a zero vector and the sum of all perturbations defined by equation (8) is not zero, although the forecast perturbations are centered. These two properties do not depend on the value of α. It is desirable that all initial perturbations are centered 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. Centering Perturbations.

As shown in equation (8), $\mathbf{Z}_p{}^a$ are the transformed perturbations by $\mathbf{T}_p$. These perturbations are not centered, 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 centered 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 and W06).

Let $\mathbf{E} = [\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_{k-1}]$, 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 center the perturbations. Also $\mathbf{C}^T \mathbf{C} = \mathbf{I}$ means $\mathbf{E}^T \mathbf{E} = \mathbf{I}_{(k-1)(k-1)}$. 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 centered 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 equation (4), to act on all k perturbations $\mathbf{Z}_p{}^a$ to produce k centered 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

$$\mathbf{Z}^a = \mathbf{Z}_p{}^a \mathbf{C}^T = \mathbf{Z}^f \mathbf{C} \mathbf{G}^{-1/2} \mathbf{C}^T \tag{13}$$

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

$$\mathbf{Z}_{mq}{}^a = \sum_{i=1}^{k} \mathbf{Z}_{mi}{}^f \sum_{l=1}^{k} \mathbf{C}_{il} g_l{}^{-1/2} \mathbf{C}^T{}_{lq} = \sum_{i=1}^{k} \mathbf{Z}_{mi}{}^f \sum_{l=1}^{k-1} \mathbf{C}_{il} \mathbf{C}_{ql} \lambda_l{}^{-1/2} + \frac{1}{k\sqrt{\alpha}} \sum_{i=1}^{k} \mathbf{Z}_{mi}{}^f \tag{14}$$

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 equation (14), the sum of the final perturbations is:

$$\sum_{q=1}^{k} \mathbf{Z}_{mq}{}^{a} = \sum_{i=1}^{k} \mathbf{Z}_{mi}{}^{f} \sum_{q=1}^{k} \sum_{l=1}^{k-1} \mathbf{C}_{il} \mathbf{C}_{ql} \lambda_l^{-1/2} = \sum_{i=1}^{k} \mathbf{Z}_{mi}{}^{f} \sum_{l=1}^{k-1} \mathbf{C}_{il} \lambda_l^{-1/2} \sum_{q=1}^{k} \mathbf{C}_{ql} = 0 \tag{15}$$

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

2.4. Orthogonality of Centered Perturbations.

Since all perturbations are centered, they are not strictly orthogonal as they were before ST. The ideal initial perturbations in an ensemble system must be centered 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 orthognality of the perturbations defined in equation (13) in the following.

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

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 equation (16) results in

$$\mathbf{J} = \mathbf{E}\mathbf{E}^{T} = \mathbf{I} - \mathbf{c}_k \mathbf{c}_k{}^{T} \tag{17}$$

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

$$\mathbf{J}_{ij} = -\frac{1}{k}, \qquad \lim_{k \to \infty} (\mathbf{J}_{ij}) = 0 \tag{18}$$

Equation (18) shows that for a finite number of ensemble members, the analysis perturbations after ET and ST transformations are not orthogonal. The perturbations are uniformly centered and distributed in different directions. The larger the number of ensemble members, the more orthogonal the perturbations will become. If the number of ensemble members approaches infinity, then the transformed perturbations will be orthogonal under this norm.

The properties of the initial perturbations generated from equation (13) can be summarized as follows. (a) The initial perturbations will be centered around the analysis field to improve the score of 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

centered and distributed in different directions. The more ensemble members we have, the more orthogonal the perturbations will become. (d) The initial perturbations have flow dependent spatial structure if the analysis error variance is derived from operational DA system at every cycle. (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 perturbations. The breeding technique generates perturbations along the fastest growing directions without assuming the linearity of the initial errors. The bred vectors are generalizations of the dominant Lyapunov vectors. Dominant Lyapunov vectors together with the associated Lyapunov exponents are the fundamentals of nonlinear 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 breeding method. 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 Dec 2002 to 17 Feb 2003, however, our study will focus on the 32-day period from 15 Jan 2003 to 15 Feb 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 hours). 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 hours in accordance with the NCEP DA system, in which new observations are assimilated in consecutive 6-hour time windows centered at 00, 06, 12 and 18 UTC.

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.

Fig. 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. (4) 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 Eqs. (3) and (4) 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 impose a regional rescaling process like operational breeding based EFS, i.e., each initial perturbation after ET and ST from eq. (4) will be rescaled by the analysis error variance using

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

where α is the rescaling factor derived from analysis error variance; $i,\ j,\ l$ are indices for the horizontal and vertical directions in grid point space; and $m = 1, 2, \ldots k$ is the index for the ensemble member.

If the initial perturbations from Eq.(4) are rescaled using Eq. (5), 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 500mb (T500) for an ensemble with 80 members as described in Fig. 1. The results with and without regional rescaling are shown in Figs. 2c and 2d, respectively. It appears that even with 80 members the ratio of analysis to forecast spread of T500 doesn't resemble that of the analysis error variance (Fig. 2d). After the regional rescaling is imposed as Eq. (21), this ratio shows a strong similarity to that of the breeding ensemble (Fig. 2c). In conclusion, the spread distribution of the initial perturbations generated by pure ET and ST transformations, based on Eqs. (3) and (4), doesn't 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 (thick) and forecast (thin) perturbations from ET with rescaling (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 relatively larger energy spreads between 600mb and 200mb. 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).

Fig. 3b shows the energy spread distributions of analysis and forecast perturbations by latitude for 10-member ensemble systems using ET with rescaling,

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 degrees south, close to the southern ocean track region. However, different distributions are found in the ET with rescaling and the breeding ensembles. The spread distributions in these two systems are similar except for some differences in the tropics. Both ET with rescaling 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 ET with rescaling and the breeding ensembles to show higher spread in this region is related to the mask imposed on the system (Toth and Kalnay 1997). These results indicate that the mask used by the breeding 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 ET with rescaling. 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 the lowest when compared with the ET, ET with rescaling, 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. We have computed PECA values as described in Wei and Toth (2003) for all the ensemble systems mentioned in the previous paragraph. The PECA values for 500mb geopotential height for a 10-member ET with rescaling (solid), ET (dotted), breeding (dashed) and ETKF (dash-dotted) are shown in Figs. 4a, b, c and 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, ET with rescaling (solid) has the highest average PECA values (thin lines) for short lead times, with breeding (dashed) next. The gap between ET with rescaling 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 ET with rescaling 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 ET with (solid) and without (dotted) rescaling, the 20-of-80-member ET with (dashed)

and without (dash-dotted) rescaling 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 vs. dashed line; dotted vs. dash-dotted).

Also plotted in Fig. 5 are the PECA values from the optimally combined perturbations for 80-member ET ensembles with (diamond) and without (square) rescaling at a 6-hour 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 hours, for cycling. Again, rescaling increases the PECA values for the ET ensembles, especially for large domains like the Northern Hemisphere and the globe. The difference between ET ensembles with and without rescaling is smaller for smaller domains, such as North America and Europe. The PECA value for ET with rescaling 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% to 90% of forecast errors at 6-hour lead time. In all domains, the PECA values at a 6-hour lead time from the 80-member ET are much larger than those from 20 members. This implies that the forecast error covariance at 6-hour 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 improve DA quality. 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-hour 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 500mb pressure level for a 6-hour 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 (see Majumdar *et al*. 2001 and 2002, Wang and Bishop 2003 and W06 for more details). Fig. 6a shows the variance distributions in the Northern Hemisphere for a 10-member ET with and without rescaling. The same is shown for a 10-member breeding ensemble and ETKF in Fig. 6b. In ET experiments, the performance is similar between ET with and without rescaling, although ET with rescaling explains slightly less of the forecast 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. Fig. 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. 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 Sores (RPSS), Economic Values (EV) and the area under the 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 a perfect forecast with 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 2x2 contingency tables containing the relative fractions of hits, misses, false alarms and correct rejections (Mason 2003). 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.

Fig. 7 shows the Brier Skill Score (BSS) for 500mb geopotential height over the Northern Hemisphere, which is calculated by averaging the BSS for 10 climatologically equally likely events using climatology as a reference forecast. For shorter forecast lead times at least up to day 7, and for ensembles with 10 members ET with rescaling is best, while ETKF is the worst and breeding is in the middle. If we use 20 members out of the 80-member ET with rescaling as described in Fig. 1, its BSS value is higher than all the other experiments at all forecast lead times.

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 ET with rescaling 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 ET with rescaling, 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 ET with rescaling is similar to the 10-member breeding, and both are better than the 10-member ETKF. Again, the 20 out of 80 member ET with rescaling 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 centering 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 centered. 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 ET with rescaling. 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 in Fig. 10. The 20 out 80 member ET with rescaling experiment described back in Section 3 is indicated as ENS_s in Fig. 10.

Fig. 10 shows the Ranked Probability Skill Score (RPSS) of 500mb 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 ET with rescaling 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 and 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 score, 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.

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 centering 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 500mb 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 day 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 500mb 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 mean of all perturbations from each perturbation doesn't alter the directions of the perturbations too much. Using one-sided breeding with this kind of simple centering 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 EFS based on four different techniques for generating initial perturbations: ET, ET with rescaling, breeding and ETKF. As in W06, results are presented for a 32-day experimental period using the NCEP operational analysis/forecast and observation systems. We have studied the characteristics of the analysis and short range forecast

perturbations, and analyzed the forecast performance using various commonly used probabilistic measures.

As stated in the Introduction, our purpose is to find an initial perturbation generation technique that will use the initial analysis error variance from the best possible DA system to control the initial perturbations for the ensemble system. In our next step, the forecast covariance from our ensemble will be used as the background covariance for the NCEP DA system. This study is the first step towards our long term goal of building a global EFS at NCEP that will be consistent with our operational DA system.

Various aspects of the properties of ETKF-generated perturbations using NCEP real observations have been thoroughly studied in W06. The relative strengths and weaknesses of ETKF and breeding were discussed and identified. In this paper, we concentrate on the ensembles generated by ET and ET with rescaling, and compare them to NCEP operational breeding. For scientific interest, in some figures we also compare our results with the ETKF results from our previous study. Both ET and ET with rescaling are second generation techniques. This paper not only gives a detailed description of the theoretical formulations of ET and ET with rescaling in generating initial perturbations, but 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.

Apart from the four techniques that we have focused on in this paper, we also tested a one-sided breeding method. In this method the bred vectors are centered by simply removing the mean of the all perturbations at each cycle. The results are compared with the ET with rescaling 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:

- The ET 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.
- 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 centered and distributed in different directions. The more ensemble members we have, the more orthogonal the perturbations will be. In the limit of infinite number of ensemble members, the perturbations will be orthogonal.
- 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. ET with rescaling 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.
- An important finding from this study is the difference in the energy spread distribution 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 spreads for ET with rescaling and breeding have higher values in the tropics and lower values in the extra-tropics. The vertical distributions of energy spread for ET with and without rescaling, breeding, and ETKF are similar although ETKF has the largest initial spread.

- PECA results show that ET perturbations can explain an amount of forecast error similar to the breeding and ETKF perturbations, while the ET with rescaling has much higher PECA values than the other three perturbations 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. ET with rescaling always has an advantage over ET without rescaling independent of ensemble size. When 80 perturbations are used, optimally combined perturbations from ET with rescaling can explain about 80% to 90% of the forecast error at a 6-hour lead time over smaller regions like North America and Europe. This implies that the 80-member ensemble can 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 perturbations, including ET and ETKF ensemble perturbations, will turn toward the leading Lyapunov vectors that are linked to the bred vectors (Toth and Kalnay 1997, Wei 2000, and Wei and Frederiksen 2004).

- ET forecast error variance predictions were better than the corresponding breeding predictions at distinguishing times and locations with larger forecast errors from times and locations where they were small. When rescaling is imposed, this ability of variance prediction is downgraded slightly.
- All ensemble systems based on the four techniques produce temporally 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 ET with rescaling and ETKF are in the middle. The advantages of high temporal consistency in EFS were discussed in W06.

- In terms of probabilistic forecast capability, ET with rescaling 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.
- The forecast scores from anomaly correlations for all the ensemble systems had only slight differences.

Results show that the breeding ensemble has a slightly higher anomaly correlation than the ETKF ensembles or ET with and without rescaling in the Northern Hemisphere for large 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 variances, as well as by the use of symmetric centering in the paired breeding scheme or spherical simplex centering in the other schemes.

- Experiments using one-sided breeding have shown better performance scores than the paired ensembles. 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 centering strategy. By simply removing the mean from all perturbations, the independence of these perturbations is preserved. The paired centering scheme reduces the EDF of the ensemble subspace by half, which may result in worse probabilistic scores. The simplex centering strategy used in the ET and ET with rescaling maximizes the EDF of ensemble subspace. This may contribute to the fact that ET with rescaling generally produces higher probabilistic scores.

The above findings are from the different experiments we have carried out so far using different techniques. Our goal at NCEP is to build an ensemble system that is consistent with the DA system and which provides an accurate analysis error variance in an operational environment using real observations to restrain

our initial perturbations. The EFS will supply the operational DA system with a good background covariance estimate. The different ensemble Kalman filter-based DA schemes (ETKF, EAKF, EnSR and LEKF) discussed in the introduction are still being tested and are being pursued by various organizations in cooperation with people at NCEP (see the discussion in W06). Good progress has been made during the last two years, and the results are compared with the benchmark NCEP operational DA system. At the time of writing, some groups have shown good results with conventional data and are testing satellite data (Whitaker *et al*. 2006). 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, ET with rescaling is the best choice for generating initial perturbations for NCEP global EFS.

By the time when we complete the final version of this manuscript, the ET with rescaling method was 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. At each cycle, only 14 members are integrated for the 16 day forecasts. It is planned that the operational configuration will be switched to that described in Fig. 1 of this paper after the NCEP supercomputers are upgraded in early 2007. Those results will be reported in the near future.

Acknowledgements: 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.

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. 2005. Comparison of error breeding, singular vectors, random perturbations and ensemble Kalman filter perturbation strategies on a simple model. (submitted to *Tellus* **A**).

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.

Frederiksen, J. S., Collier, M. A. and Watkins, A. B. 2004. Ensemble prediction of blocking regime transitions. *Tellus,* **56A**, 485-500.

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., Lefaivrem, 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. *Proc. IEEE American Control Conf.*, Anchorage, AK, IEEE, 887-892.

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 Verifiction: 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., Corazza,M., Kalnay, E., Patil, D.J. and Yorke, J.A. 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., Kalnay, E., Ott, E., and York, J.A. 2005. Assessing a local ensemble Kalman filter: perfect model experiments with the NCEP global model. *Tellus*, **57A**, in press

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. Meteror. 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 Verifiction: 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,* in press.

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. of Bifurcation and 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. in Geoph.*, ***11***, 99-118.

Wei, M., Toth, Z., Wobus, R., Zhu, Y., Bishop, C. H. 2005a. Initial Perturbations for NCEP Ensemble Forecast System. Thorpex Symposium Proceedings for t*he 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. Cui,B. 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. 2006. A Very Brief Summary of Ensemble Transform (ET) Based Initial Perturbations in NCEP Global Ensemble Forecast System. *NOAA/NCEP/EMC internal document,* June 26, 2006 (available from the author or EMC director).

Wei, M., Toth, Z., Wobus, R., Zhu, Y., Bishop, C. H. 2006a. 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., Zhu, Y., Bishop, C. H. and Wang, X. 2006b. Ensemble Transform Kalman Filter-based ensemble perturbations in an operational global prediction system at NCEP. *Tellus,* **58A**, 28-44.

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

Whitaker, J. S., Hamill, T.M., Wei, X., Song, Y. and Toth, Z. 2006. Ensemble data assimilation with the NCEP global forecast system. (To be submitted to *Mon. Wea. Rev.*)

Wilks, D. S. 1995. Statistical Methods in the Atmospheric Sciences. Cambridge Press, 547pp.

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.

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	Realistic through sample, case dependent patterns and amplitudes	Fastest growing subspace, case dependent patterns	No explicit estimate, not flow dependent
Sampling	Random for all errors, including non-growing, potentially hurts short-range performance	Nonlinear Lyapunov vectors, subspace of fastest growing errors, some dependence among perturbations	Dynamically fastest growing in future, quite 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

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	Fast growing subspace, case dependent patterns and amplitudes	Fastest growing subspace, case dependent patterns	Case dependent Variance
Sampling	Orthogonal in the normalized observational space	Orthogonal in analysis covariance norm	Dynamically fastest growing in future
Consistency between EFS and DA system	Very good, however, quality of DA lagging behind 4D-Var	Very good, DA provides good analysis for EPS which provides accurate forecast error covariance for DA	Consistent

Figure Captions

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.

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 500mb for a 20 of 80 member ET with rescaling; and (d) ratio of temperature spread at 500mb for a 20 of 80 member ET without rescaling.

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 Jan. – 15 Feb. 2003, with (a) vertical distribution as a function of pressure; (b) horizontal distribution by latitude.

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.

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.

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

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.

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.

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.

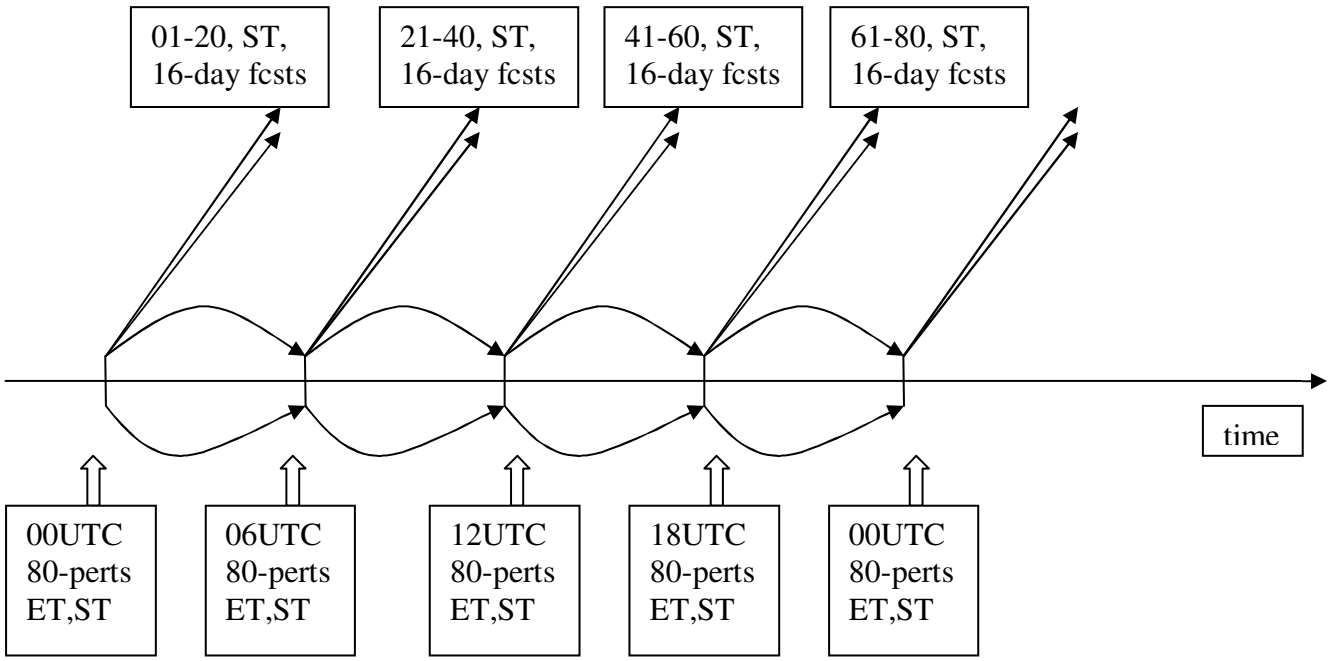

Fig.1. Schematic of the configuration of the 80-member ET-based ensemble experiment. At each cycle ET transformation is carried out in all 80 perturbations, followed by the ST transformation. ST is also imposed on the 20 perturbations that will be used for long-range forecasts.

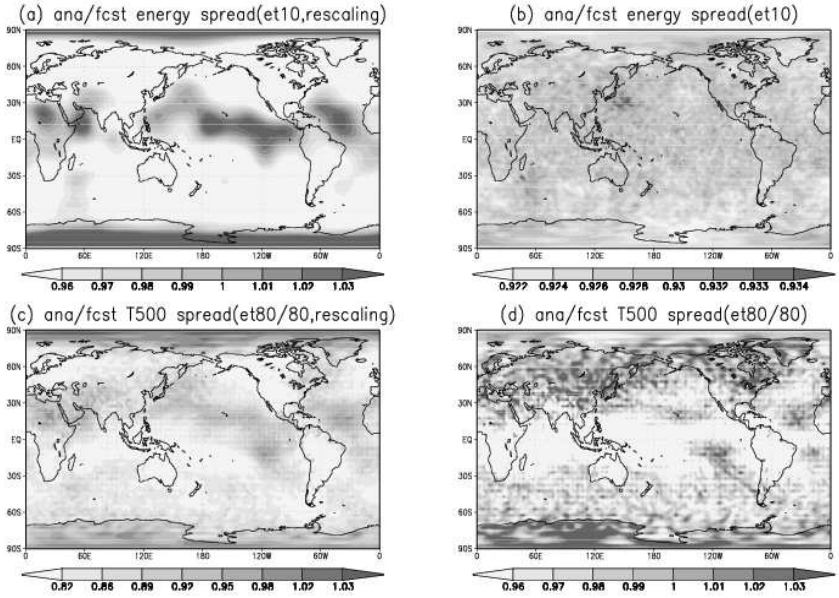

Fig. 2. Global distribution of the ratios of the analysis to forecast spread for ET based ensembles for (a) vertically averaged ratio of energy spread for a 10-member ET with rescaling; (b) vertically averaged ratio of energy spread for a 10-member ET without rescaling; (c) ratio of temperature spread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

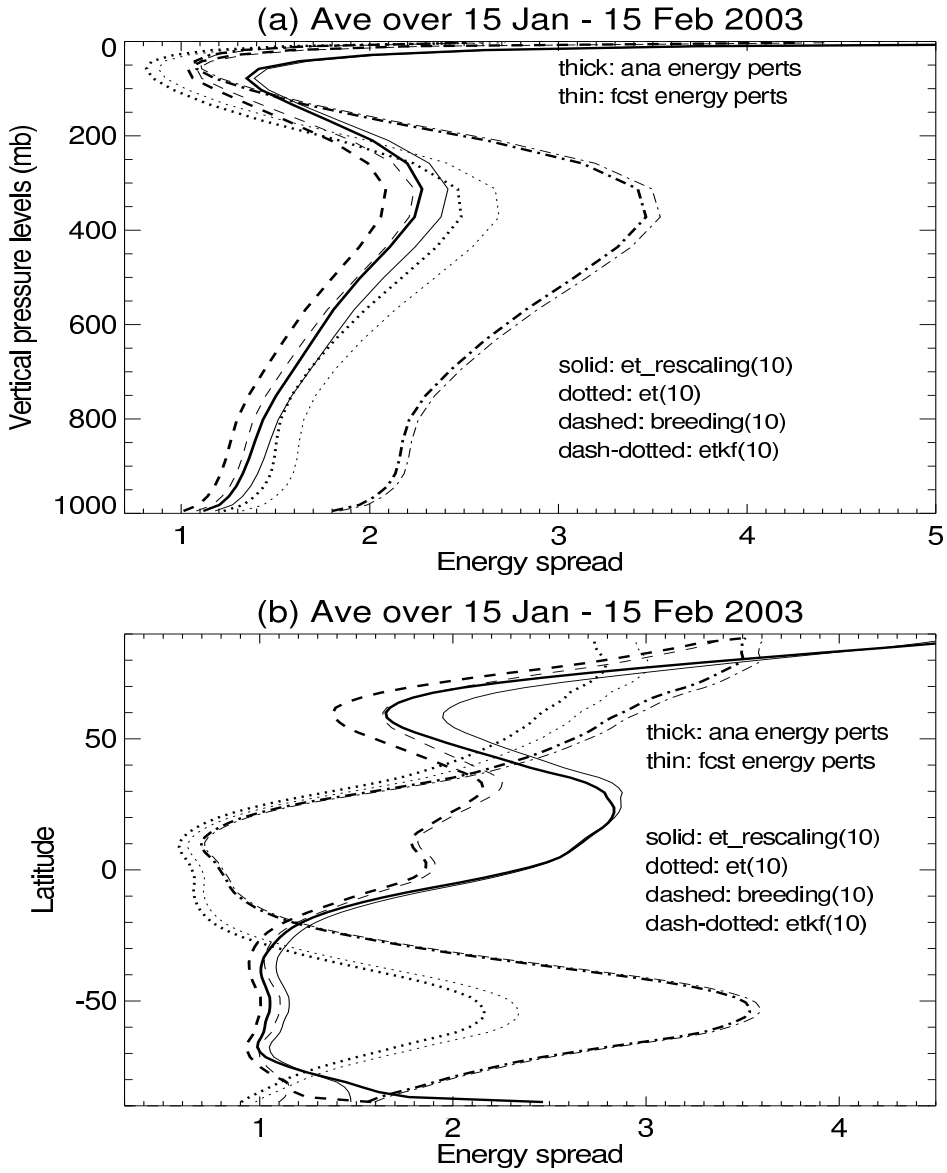

Fig.3. Energy spread distributions of ET with rescaling (solid), ET without rescaling (dotted), breeding (dashed) and ETKF ensemble perturbations (thick: analysis; thin: forecast). All the ensembles have 10 members and values are averaged over the period 15 Jan. – 15 Feb. 2003, with (a) vertical distribution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

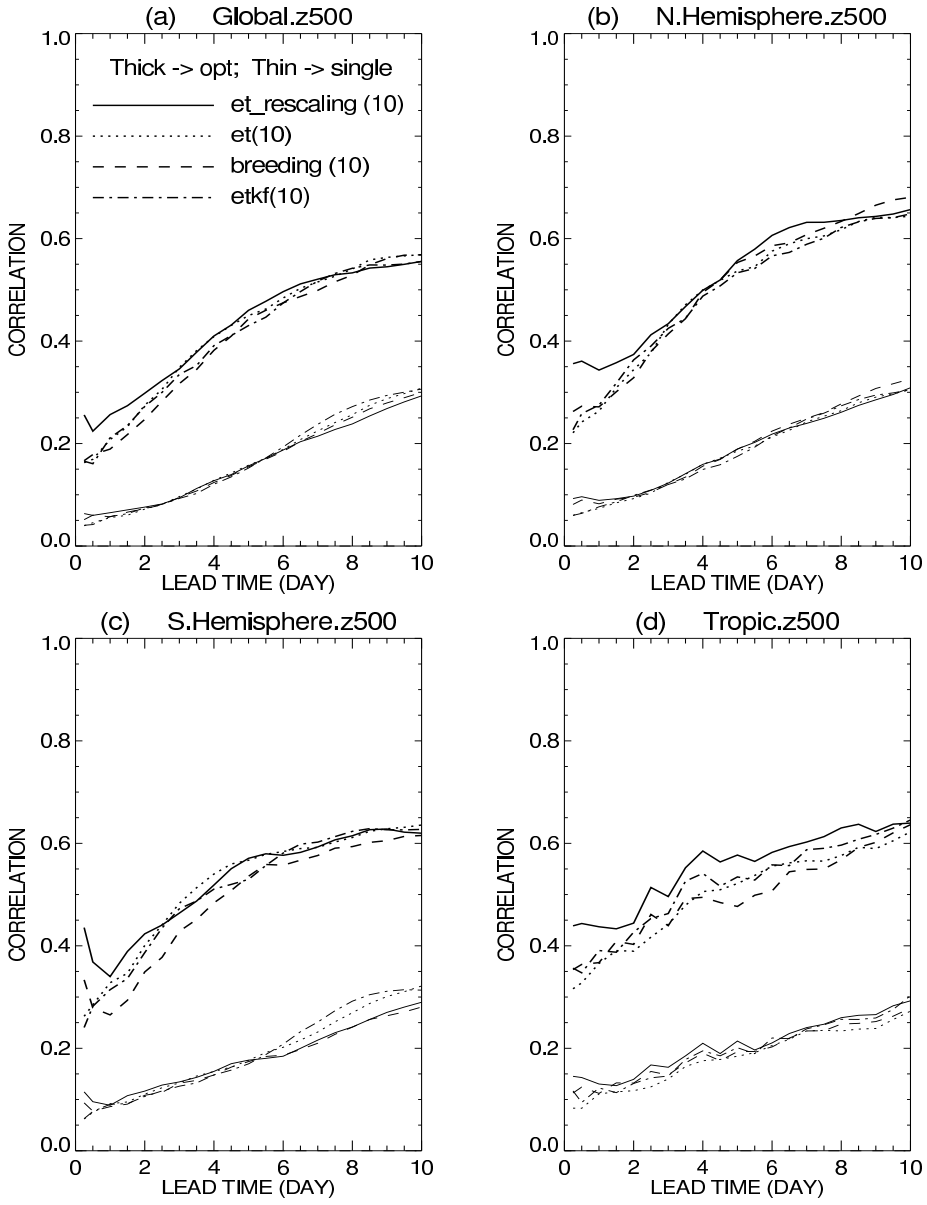

Fig. 4. PECA values for ET with rescaling (solid), ET without rescaling (dotted), breeding (dashed) and ETKF (dash-dotted) ensembles with 10 members for (a) the globe; (b) Northern Hemisphere; (c) Southern Hemisphere and (d) the tropics. Shown in thick and thin lines are PECA from the optimally combined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

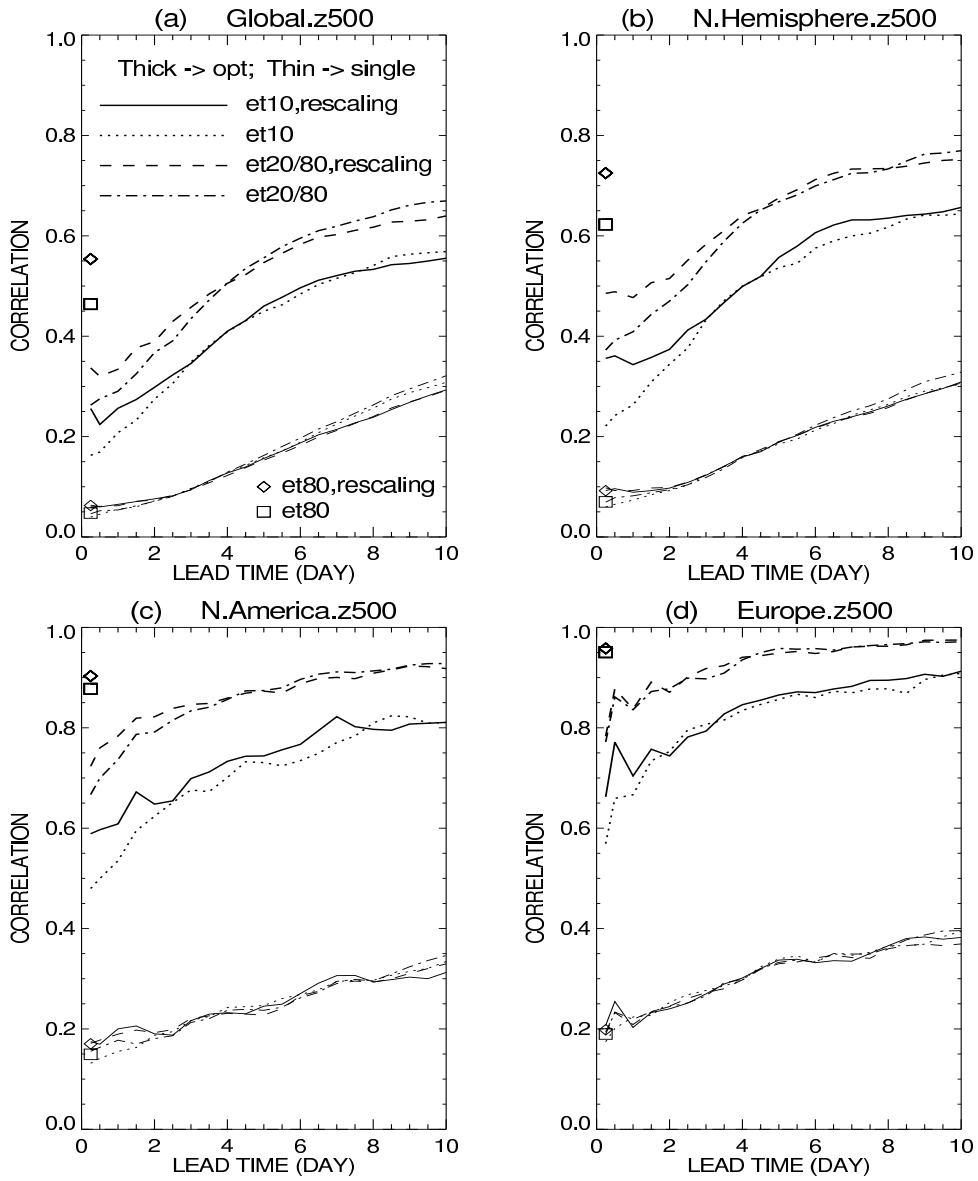

Fig. 5. PECA values for a 10-member ET with rescaling (solid), 10-member ET without rescaling (dotted), 20 of 80 member ET with rescaling (dashed) and 20 of 80 member ET without rescaling (dash-dotted) ensembles for (a) the globe; (b) Northern Hemisphere; (c) Southern Hemisphere and (d) the tropics.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

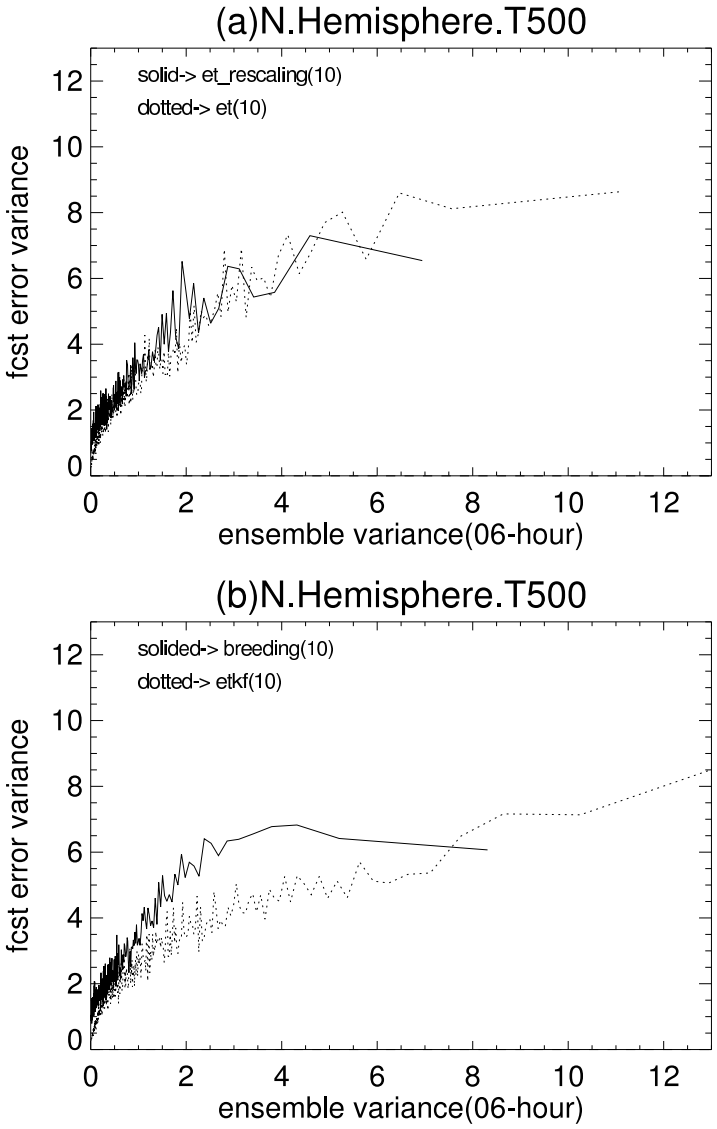

Fig. 6. Derived 10-member ensemble variance and forecast error variances at all grid points for 500mb temperature over the Northern Hemisphere for (a) ET with rescaling (solid) and ET without rescaling; (b) breeding (solid) and ETKF (dotted).

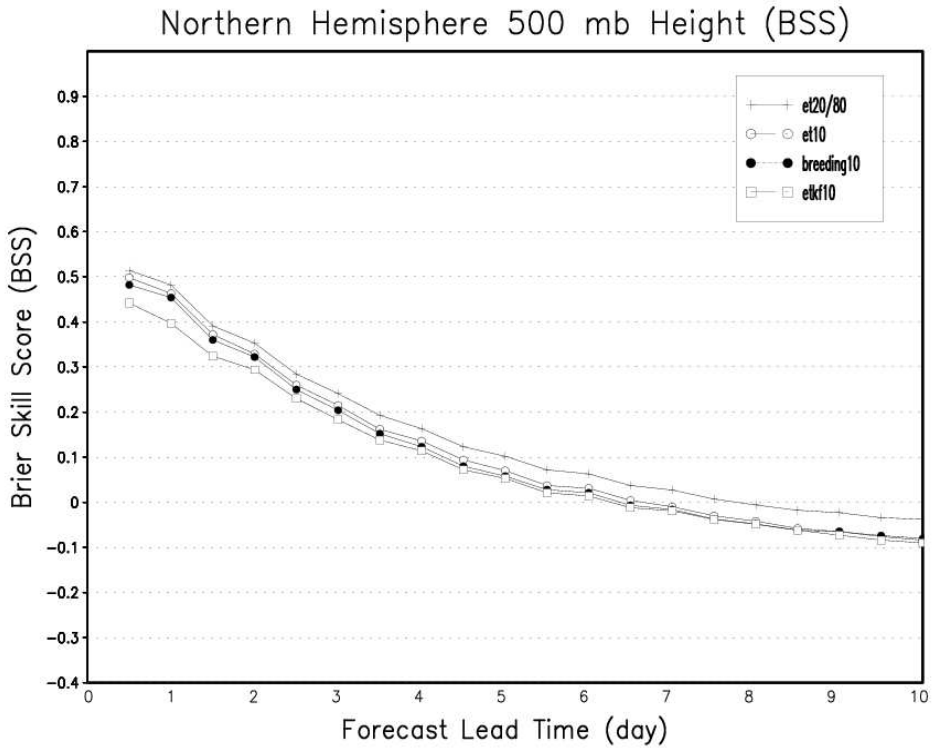

Fig. 7. Averaged Brier Skill Score of 500 mb geopotential height over the Northern Hemisphere for 20 of 80 member ET with rescaling (cross), 10-member ET with rescaling (open circle), 10-member breeding (full circle) and 10-member ETKF (open square) ensembles.

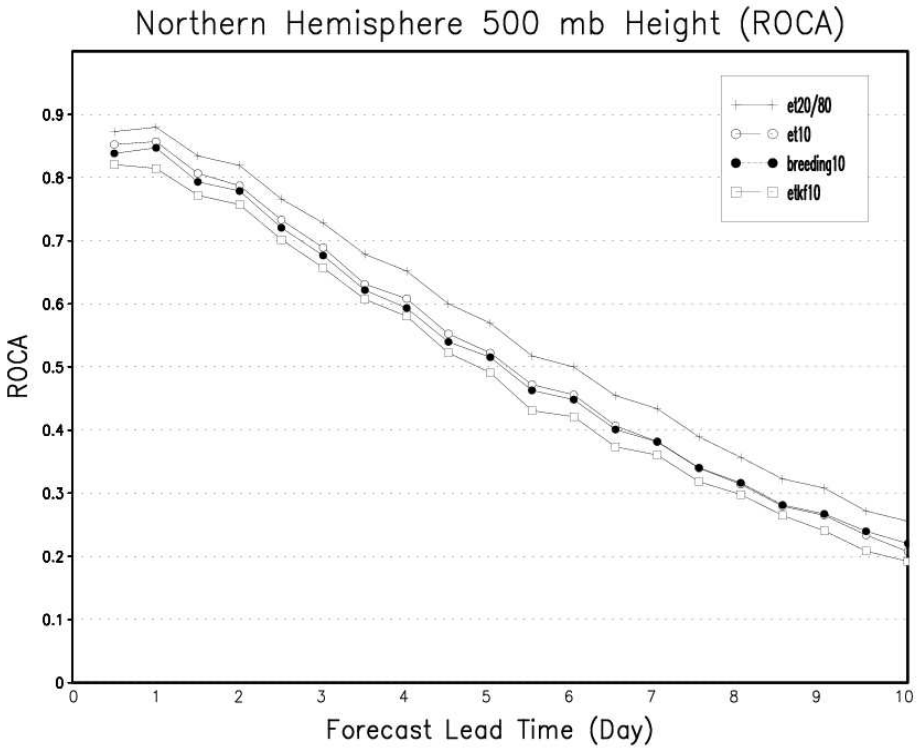

Fig.8. The same as Fig. 7, but for the relative operating characteristic area.

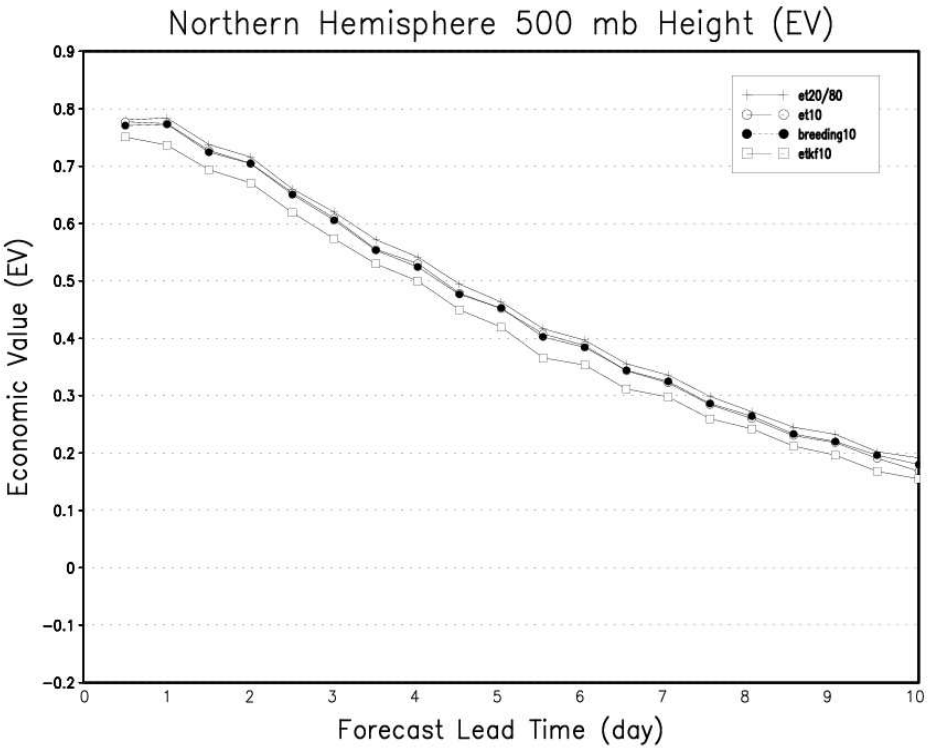

Fig. 9. The same as Fig. 7, but for the economic value.

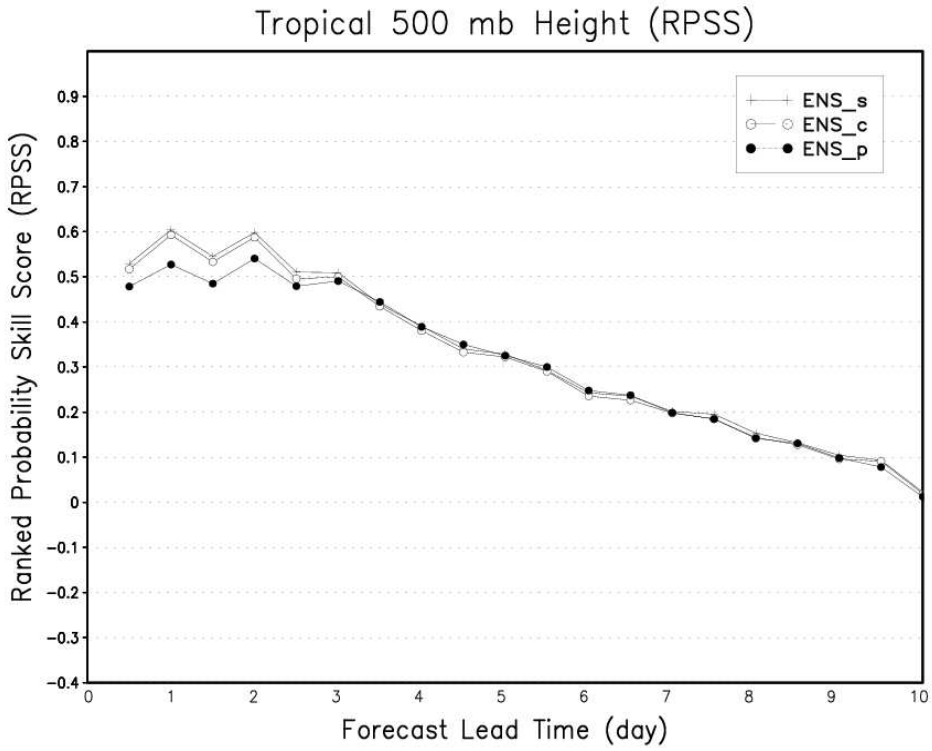

Fig. 10. Averaged Ranked Probability Skill Score of 500 mb geopotential height over the Northern Hemisphere for 20 of 80 member ET with rescaling (ENS_s, cross), 20-member one-sided breeding (ENS_c, open circle), and 20-member paired (full circle) ensembles.

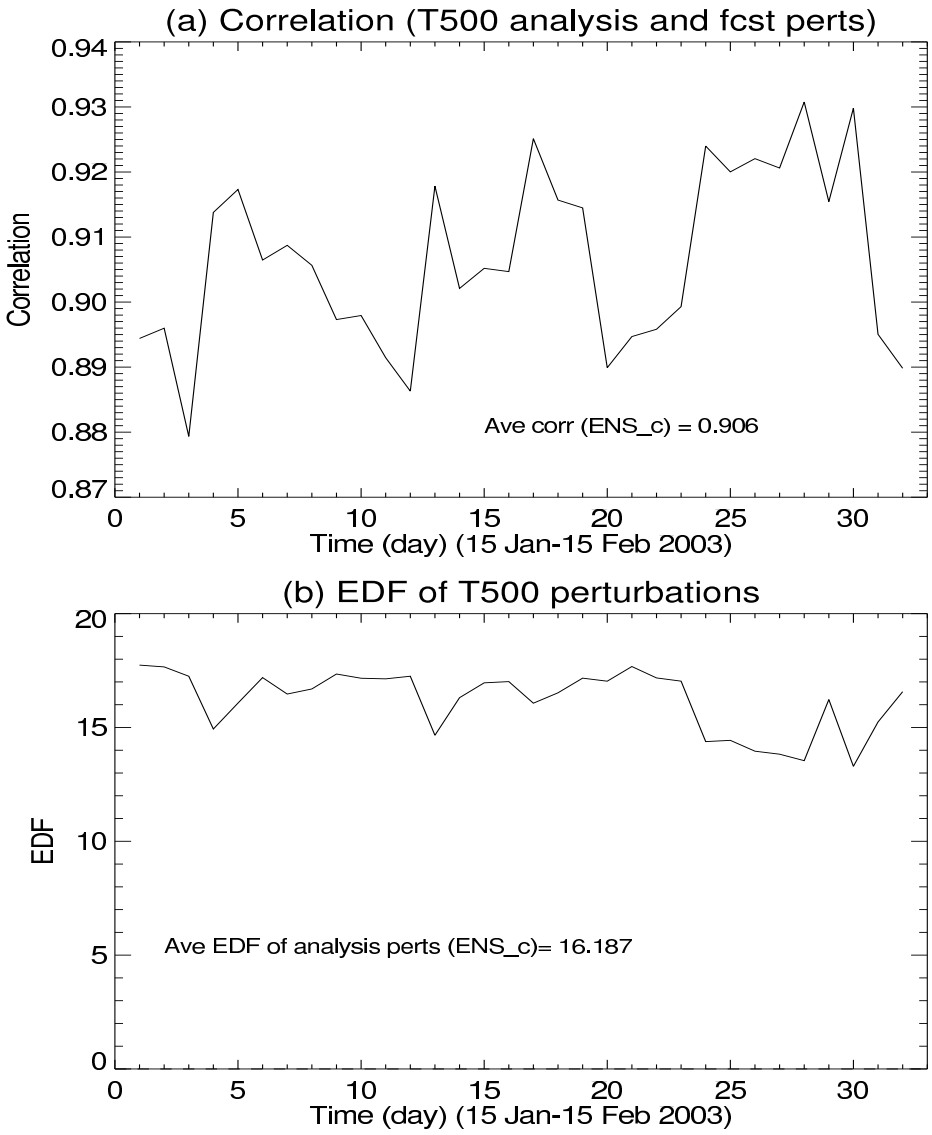

Fig. 11. (a) Average correlation over 20 members between the temperature forecasts and analysis perturbations at 500 mb geopotential height for a one-sided breeding ensemble. (b) The effective number of degrees of freedom of subspace spanned by 20 temperature analysis perturbations from a one-sided breeding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