1.4 ENSEMBLE FORECASTS BASED ON ET WITH RESCALING

Mozheng Wei
SAIC at NOAA/NWS/NCEP, Environmental Modeling Center, 5200 Auth Road, Rm. 207, Camp Springs, MD 20746, USA (email: Mozheng.Wei@noaa.gov)
Zoltan Toth, Yuejian Zhu
NOAA/NWS/NCEP, Camp Springs, MD, USA
Richard Wobus
SAIC at NOAA/NWS/NCEP, Camp Springs, MD, USA

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 years, 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 of Wei *et al.* 2008, referred to as W08): 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.

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 data assimilation (DA) systems that generate the analysis fields. Comparisons of performance between the ECMWF and NCEP operational EFSs were described in Wei and Toth (2003). A recent comprehensive summary of these first generation methodologies and their performance at ECMWF, MSC and NCEP can be found in Buizza *et al.* (2005).

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, W08). The research and experiments on using ensemble transform (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.,* 2005). Since then, more experimental results with ensembles using ET and ETR have been presented and documented in Wei *et al.* (2006b) and W08. 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 ran 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 day forecasts. The NCEP operational configuration has been switched to that described in Fig. 1 of this paper since March 27, 2007 after the arrival of new NCEP supercomputers in early 2007. In this paper, The methodology and some results from the ET and ET with rescaling are briefly summarized.

Another technique in the second generation is based on the Ensemble Transform Kalman Filter (ETKF, Bishop *et al.* 2001). Wang and Bishop (2003) used ETKF to generate ensemble perturbations in an idealized observation framework. It was further extended to an operational environment with the NCEP operational model and real-time observations by Wei *et al.* (2006a) (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 if the Hessian-SV is computed with flow-dependent analysis PDF. It samples fastest growing directions for a specific lead time and norm. i.e. 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 centers 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-varaible Lorenz model, the author showed that EnKF performs best, the performance of ETKF with random perturbations is the next most skillful. It was also found that neither the ETKF, error breeding nor singular vectors provided useful background information on their own.

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 (2007) who found that: analysis error characteristics (e.g. statistics) are similar (to first approximation) to those of 6-hour 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 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 span maximum degrees of freedom within ensemble subspace. 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 centers with operational observation data. Before ensemble DA shows satisfying performance, ETR with repositioning (i.e. perturbed ensemble states are centered about the analysis field) offers a good solution for consistent DA/EFS generation, with 2-way exchange of information. Section 2 provides brief description of the ET formulation for initial perturbations. Section 3 presents some 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 were 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. 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.

The ETKF formulation 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). 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 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. The detail mathematical descriptions of ET and ET with recalling for ensemble forecasting and implementation at NCEP are given in Wei et al. 2006b and 2008. Here we just summarize the properties initial perturbations generated by ET with rescaling: (a) The initial perturbations will be centered 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 centered and distributed in different directions. The more ensemble members we have, the more orthogonal the perturbations will become. It is shown 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.

Both ETKF and ET methods resemble breeding in that they both dynamically cycle the fastest growing nonlinear 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 nonlinear model. 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). The ET method produces perturbations along the fastest growing directions that are constrained by the initial analysis error variances. 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.

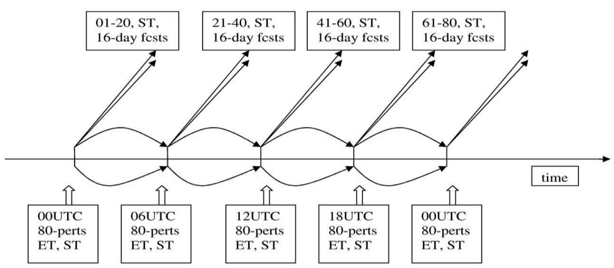

Fig.1. Schematic of the configuration of the 80-member ET-based ensemble experiment. At each cycle ET transformation is carried out in all 80 perturbations, followed by the ST transformation. ST is also imposed on the 20 perturbations that will be used for long-range forecasts.

3. RESULTS FROM ET WITH RESCALING, ET, BREEDING AND ETKF ENSEMBLES

3.1. Ensemble spread distribution

To understand 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. 2a we show the vertical distributions of energy spread for the analysis (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 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. 2b 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. 2a, 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 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.

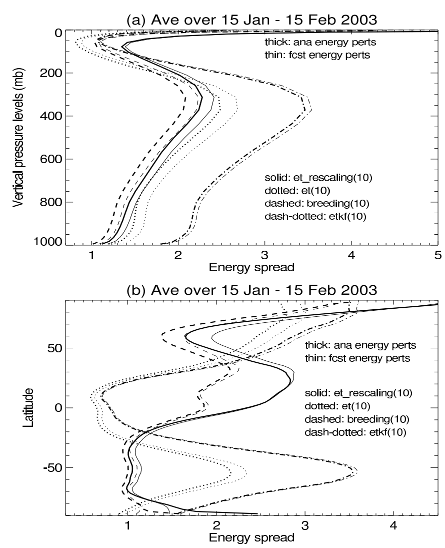

Fig.2. Energy spread distributions of ET with rescaling (solid), ET without rescaling (dotted), breeding (dashed) and ETKF ensemble perturbations (thick: analysis; thin: forecast). All the ensembles have 10 members and values are averaged over the period 15 Jan. – 15 Feb. 2003, with (a) vertical distribution as a function of pressure; (b) horizontal distribution by latitude.

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 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 500mb geopotential height for a 10-member ETR (solid), ET (dotted), breeding (dashed) and ETKF (dash-dotted) are shown in Figs. 3a, 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.

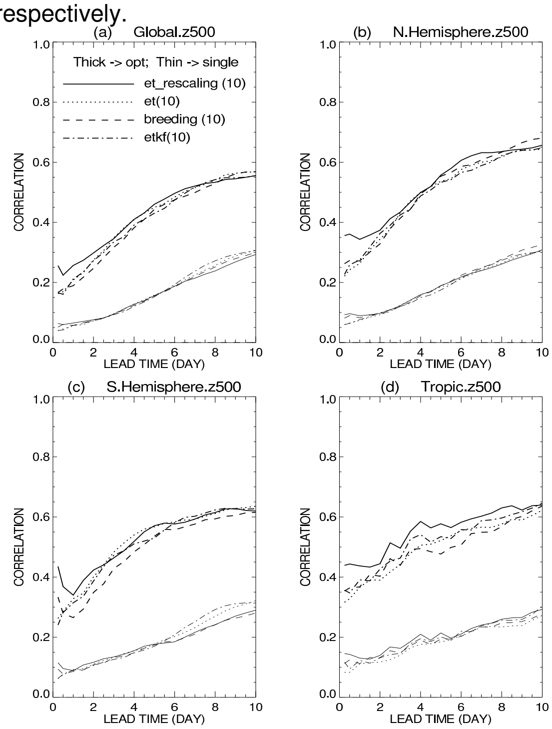

Fig. 3. PECA values for ET with rescaling (solid), ET without rescaling (dotted), breeding (dashed) and ETKF (dash-dotted) ensembles with 10 members for (a) the globe; (b) Northern Hemisphere; (c) Southern Hemisphere and (d) the tropics. Shown in thick and thin lines are PECA from the optimally combined perturbations and average PECA from the individual perturbations, respectively.

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. 2).

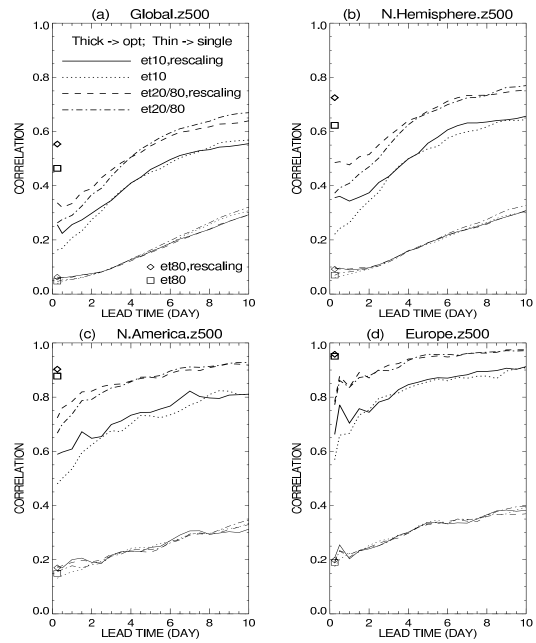

Fig. 4. PECA values for a 10-member ET with rescaling (solid), 10-member ET without rescaling (dotted), 20 of 80 member ET with rescaling (dashed) and 20 of 80 member ET without rescaling (dash-dotted) ensembles for (a) the globe; (b) Northern Hemisphere; (c) Southern Hemisphere and (d) the tropics. Shown in thick and thin lines are PECA from the optimally combined perturbations and average PECA from individual perturbations, respectively.

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. 4, 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 vs. dashed line; dotted vs. dash-dotted).

Also plotted in Fig. 4 are the PECA values from the optimally combined perturbations for 80-member ETR (diamond) and ET (square) 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 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% to 90% of forecast errors at 6-hour lead time if the analysis error is small and can be neglected. In all domains, the optimally combined 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. In practice a covariance localization would have to be applied to ensemble before it is used in 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.

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 (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.

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 can be decomposed into reliability, resolution and uncertainty components (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. 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 (Toth *et al.*, 2003).

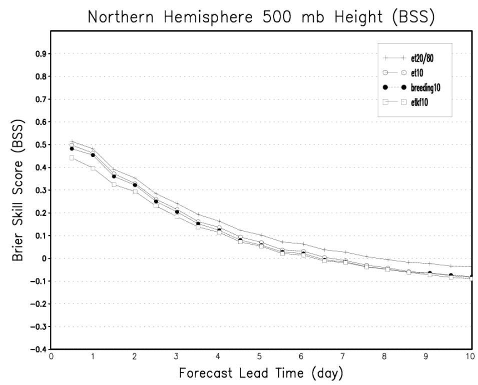

Fig. 5. Averaged Brier Skill Score of 500 mb geopotential height over the Northern Hemisphere for 20 of 80 member ET with rescaling (cross), 10-member ET with rescaling (open circle), 10-member breeding (full circle) and 10-member ETKF (open square) ensembles.

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 (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. 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. 5 shows the Brier Skill Score (BSS) for 500mb geopotential height over the Northern Hemisphere, which is calculated by using climatology as a reference forecast. 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.

Shown in Fig. 6 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. 7. 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.

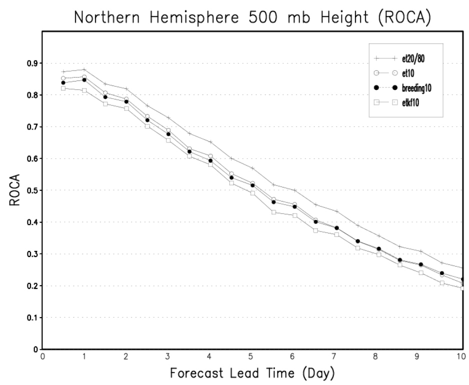

Fig.6. The same as Fig. 5, but for the relative operating characteristic area

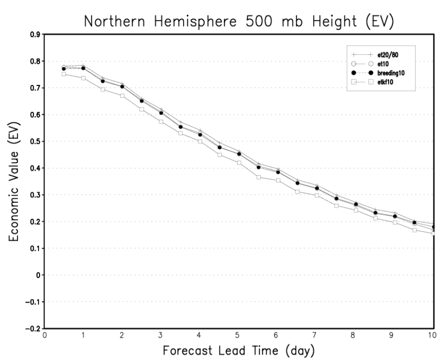

Fig. 7. The same as Fig. 5, but for the economic value.

4. DISCUSSION AND CONCLUSIONS

In this paper, we have described several experiments with four different initial perturbation generation techniques: breeding, ETKF, ET and ETR. All of these are based on the principle of the breeding method dynamically cycling nonlinear forecast perturbations. We have concentrated on the ensembles generated by ET and ETR, and compared 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.

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

- 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.
- 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 closer to being orthogonal the perturbations will be. In the limit of infinite number of ensemble members, the perturbations will be exactly orthogonal.
- 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.
- 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% 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 may be able to provide an efficient background covariance for the DA system.
- 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.

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.

Acknowledgements: We are grateful to many colleagues at NCEP/EMC for their help during this work.

REFERENCES

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. *Tellus,* **58A,** 538-548.

Buizza, R. and Palmer, T. N. 1995. The singular-vector structure of the atmospheric global circulation. *J. tmos. 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).

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

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.

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

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.

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.

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., Toth, Z., Wobus, R., Zhu, Y., Bishop, C. H. 2005. 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., 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. 2006b. 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,** 33pp**.**

Wei, M., Toth, Z., Wobus, R., Zhu, Y. 2008, Initial Perturbations Based on the Ensemble Transform (ET) Technique in the NCEP Global Operational Forecast system. *Tellus,* **60A,** 62-79.

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.