A Comparison of Breeding and Ensemble Transform Vectors for Global Ensemble Generation

DENG Guo[1*](邓 国), TIAN Hua[1](田 华), LI Xiaoli[1](李晓莉), CHEN Jing[1](陈 静), GONG Jiandong[1](龚建东), and JIAO Meiyan[2](矫梅燕)

1 *National Meteorological Center, China Meteorological Administration, Beijing* 100081

2 *China Meteorological Administration, Beijing* 100081

(Received November 25, 2010; in final form December 7, 2011)

ABSTRACT

To compare the initial perturbation techniques using breeding vectors and ensemble transform vectors, three ensemble prediction systems using both initial perturbation methods but with different ensemble member sizes based on the spectral model T213/L31 are constructed at the National Meteorological Center, China Meteorological Administration (NMC/CMA). A series of ensemble verification scores such as forecast skill of the ensemble mean, ensemble resolution, and ensemble reliability are introduced to identify the most important attributes of ensemble forecast systems. The results indicate that the ensemble transform technique is superior to the breeding vector method in light of the evaluation of anomaly correlation coefficient (ACC), which is a deterministic character of the ensemble mean, the root-mean-square error (RMSE) and spread, which are of probabilistic attributes, and the continuous ranked probability score (CRPS) and its decomposition. The advantage of the ensemble transform approach is attributed to its orthogonality among ensemble perturbations as well as its consistence with the data assimilation system. Therefore, this study may serve as a reference for configuration of the best ensemble prediction system to be used in operation.

Key words: breeding, ensemble transform, ensemble prediction system

Citation: Deng Guo, Tian Hua, Li Xiaoli, et al., 2012: A comparison of breeding and ensemble transform vectors for global ensemble generation. *Acta Meteor. Sinica*, **26**(1), 52–61, doi: 10.1007/s13351-012-0105-4.

1. Introduction

After the fundamental work of Lorenz (1963), the chaotic attribute of weather system is well-recognized by meteorologists, along with the realization of further limitation in model errors linked to the approximate simulation of atmospheric processes. These two sources of uncertainty limit the skill of single, deterministic forecasts in an unpredictable way, and thus ensemble prediction is put forward as a feasible way to complement a single, deterministic forecast with an estimate of the probability density function of forecast states. Ensemble of numerical forecasts from slightly perturbed initial conditions can have a beneficial impact on the skill of the forecast (Leith, 1974). To this end, ensemble forecasts start from a set of different states sampled from a probability density function. However, how to better generate these initial perturbations is still a research issue. The initial perturbation techniques applied at different forecasting centers are: singular vectors (SVs) at the European Centre for Medium-Range Weather Forecasts (ECMWF) (Buizza and Palmer, 1995), breeding vectors (BVs) at the National Meteorological Center, China Meteorological Administration (NMC/CMA), ensemble transform vectors at the US National Centers for Environmental Prediction (NCEP) (Wei et al., 2008), and the perturbed observation (PO) approach at the Meteorological Service of Canada (MSC) (Buizza et al., 2005). The ensemble transform initial perturbation technique was developed from the traditional breeding method, and it represents a nonlinear extension of the Lyapunov vectors. The ensemble transform technique is more consistent with the data assimilation systems (Wei et al., 2008). Hence, it is regarded as a promising method in operational forecast. In this study, based

Supported by the China Meteorological Administration Special Public Welfare Research Fund (GYHY200706001 and GYHY200906007) and National Natural Science Foundation of China (41075079).

*Corresponding author: deng719@cma.gov.cn.

on the spectral model T213 at the NMC/CMA, comparisons between two different initial perturbation methods (i.e., the breeding vector method and the ensemble transform method) are conducted to investigate the skill of three ensemble prediction systems. The result may serve as a reference for global ensemble prediction system (GEPS) upgrading at the national level forecast centers.

This paper is organized as follows. In Section 2, we briefly review the breeding vector and ensemble transform techniques. Section 3 describes verification results and analysis. The summary and discussion are given in the last section.

2. Initial perturbation methodologies and experiment design

The model used is the NMC/CMA T213/L31, a global model that adopts the semi-Lagrangian time-stepping scheme. The model calculates meteorological fields on three-dimensional grids with 31 vertical levels (thereby L31). The output for the horizontal fields is compressed in spherical harmonics with triangular truncation and a maximum resolution of 213 wavelengths on a great circle (thereby T213).

The initial perturbation methods using breeding vectors and ensemble transform vectors are used to sample and sort out the fastest-growing components of initial errors. The breeding approach attempts to simulate the development of growing errors in the analysis cycle (Toth and Kalnay, 1997), and the ensemble transform technique (Bishop and Toth, 1999; Wei et al., 2008) is an extension of breeding and thus has some similarities in breeding the perturbations in the dynamical cycles.

2.1 The breeding vector approach

The breeding vector (BV) approach simulates the development of growing errors in the analysis cycle. This procedure consists of the following steps: (a) add a small perturbation to the atmospheric analysis (initial state) at any given time t_0 (the small perturbation is obtained by using the difference between the forecasts valid at time t_0 but initiated at a short random time period prior to t_0); (b) integrate the model from both the perturbed and unperturbed initial conditions for a short period (from t_0 to t_1); (c) subtract one forecast from another; (d) scale down the difference field so that it has the same norm (i.e., amplitude of root mean square of rotational kinetic energy) as the initial perturbation; and (e) add this perturbation to the analysis at time t_1. The steps (b)–(e) are repeated forward in time. Note that once the initial perturbation is introduced in step (a), the development of the perturbation field is dynamically determined by the evolving atmospheric flow (Toth and Kalnay, 1993, 1997)

$$\boldsymbol{z}^a = \boldsymbol{z}^f \times r, \tag{1}$$

where $\boldsymbol{z}^f$ and $\boldsymbol{z}^a$ are ensemble forecast and analysis perturbations, respectively.

$$r = \mathrm{mask}_{(\lambda,\phi,t)}/k_{(\lambda,\phi,t)}, \tag{2}$$

where r is the regional rescaling factor at latitude λ, longitude ϕ, and time t; mask denotes the monthly uncertainty and is estimated by comparing the T213 operational analysis and the NCEP operational analysis; and k is derived from the every day forecast difference between a pair of ensemble members for all variables at all levels.

The growing component of the regionally varying uncertainty in the analysis is measured as the difference between parallel analysis cycles. The average difference field is then used as a mask in the regular rescaling process of the bred vectors to ensure that the initial ensemble perturbations have a spatial distribution and amplitudes similar to those of the analysis errors. Each bred perturbation is either added to or subtracted from the control analysis.

2.2 The ensemble transform vector approach

One of the initial perturbation methods, the ensemble transform (ET) vectors, was formulated in Bishop and Toth (1999) primarily for target observation studies. In this paper, we adopt this technique for ensemble forecasting. We follow Wei et al. (2008) in the perturbation matrix formulation. Let

$$\begin{aligned} \boldsymbol{Z}^f &= \frac{1}{\sqrt{k-1}}[\boldsymbol{z}_1^f, \boldsymbol{z}_2^f, \cdots, \boldsymbol{z}_k^f], \\ \boldsymbol{Z}^a &= \frac{1}{\sqrt{k-1}}[\boldsymbol{z}_1^a, \boldsymbol{z}_2^a, \cdots, \boldsymbol{z}_k^a], \end{aligned} \tag{3}$$

where the n-dimensional state vectors $\boldsymbol{z}_i^f = \boldsymbol{x}_i^f - \boldsymbol{x}^f$ and $\boldsymbol{z}_i^a = \boldsymbol{x}_i^a - \boldsymbol{x}^a (i = 1, 2, \cdots, k)$ are k ensemble forecasts and analysis perturbations, respectively. In our experiments, $\boldsymbol{x}^f$ is the mean of k ensemble forecasts, and $\boldsymbol{x}^a$ is the analysis from the independent T213 operational data assimilation system. Unless stated otherwise, the lower and upper case bold letters indicate vectors and matrices, respectively. The $n \times n$ forecast and analysis covariance matrices are formed, respectively, as

$$\boldsymbol{P}^f = \boldsymbol{Z}^f(\boldsymbol{Z}^f)^{\mathrm{T}}, \text{ and } \boldsymbol{P}^a = \boldsymbol{Z}^a(\boldsymbol{Z}^a)^{\mathrm{T}}, \tag{4}$$

where T indicates the matrix transpose. For a given set of forecast perturbations $\boldsymbol{Z}^f$ at time t, the analysis perturbations $\boldsymbol{Z}^a = (\boldsymbol{Z}^f)\boldsymbol{T}_{\mathrm{r}}$. Suppose we have obtained the analysis covariance matrix from the operational data assimilation system, then $\boldsymbol{P}^a = (\boldsymbol{Z}^f\boldsymbol{T}_{\mathrm{r}})(\boldsymbol{Z}^f\boldsymbol{T}_{\mathrm{r}})^{\mathrm{T}}$. The ET solution is $\boldsymbol{Z}^a = \boldsymbol{Z}^f\boldsymbol{T}_{\mathrm{r}}$, where $\boldsymbol{T}_{\mathrm{r}} = \boldsymbol{C}\boldsymbol{\Gamma}^{-1/2}$, $\boldsymbol{C}$ contains column orthonormal eigenvectors ($\boldsymbol{c}_i$) of $(\boldsymbol{Z}^f)^{\mathrm{T}}(\boldsymbol{P}^a)^{-1}\boldsymbol{Z}^f$, and $\boldsymbol{\Gamma}$ is a diagonal matrix containing the associated eigenvalues (λ_i), that is, $\boldsymbol{C} = [\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_k]$ and $\boldsymbol{\Gamma} = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_k)$. Although the forecast perturbations are, by definition, centered about the ensemble mean, i.e., $\sum_{i=1}^{k} \mathbf{z}_i^f = 0.0$, the analysis perturbations produced by the ET defined above are not centered around the analysis ($\sum_{i=1}^{k} \mathbf{z}_i^a \neq 0.0$). A simple transformation that will preserve $\boldsymbol{P}^a$ and center the analysis perturbations about the analysis is the simplex transformation. Similar to the ensemble transform Kalman filter (ETKF) experiments, $\boldsymbol{C}^{\mathrm{T}}$ is one of the solutions of this transformation. Hence, $\boldsymbol{Z}^a = \boldsymbol{Z}^f\boldsymbol{T}_{\mathrm{r}}\boldsymbol{C}^{\mathrm{T}}$ will be used as our initial analysis perturbations for the next cycle forecasts.

2.3 Comparison between the BV and ET methods

Figure 1 shows the basic characters of the two initial perturbation techniques. For the BV method, $\boldsymbol{P}1$ and $\boldsymbol{N}1$ are the pairs of positive and negative independent vectors, the amplitudes of which are controlled by a simple scaling coefficient while the directions of which are not changed in this process. As to the ET method, $\boldsymbol{P}1$, $\boldsymbol{P}2$, $\boldsymbol{P}3$, and $\boldsymbol{P}4$ are orthogonal vectors. To centralize all perturbed vectors (sum of all vectors should be equal to zero), like the BV method, perturbation amplitudes are scaled down by applying the mask derived in Eq. (2), and the directions of the vectors will be tuned by the ensemble transform process.

2.4 Numerical experiment design

2.4.1 Control run

The control run is the natural integration of the T213 model from 10 to 23 November 2007 with a lead-time of 10 days. The NMC operational analysis is used as observations for verification. The analysis field is obtained from the T213 analysis and assimilation system with 3DVAR and a multivariate spectral statistical interpolation (SSI) analysis scheme in which observations including the ATOVS (Advanced TIROS Operational Vertical Sounder; TIROS stands for Television and Infrared Observation Satellite) data are assimilated. The analysis is used as "true state" of atmosphere for verification in this paper.

2.4.2 14-member GEPS with initial perturbations generated by the BV method (BV14)

The model and model integration period as well as the analysis data are the same as the control run.

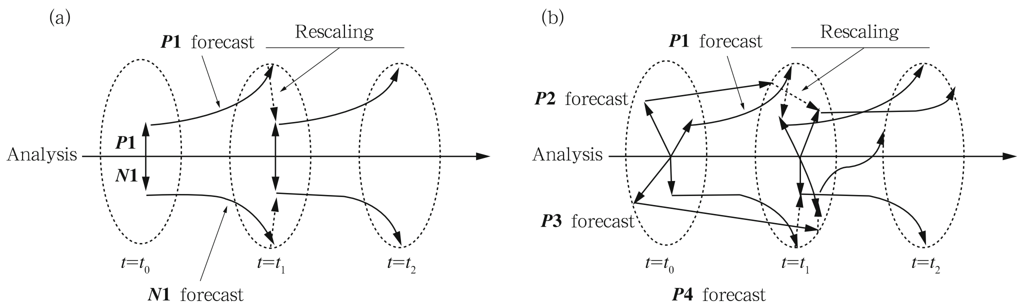

Fig. 1. Schematic diagrams of the breeding vector (a) and ensemble transform (b) methods.

We make use of the BV method to set up the GEPS, in which 14 members are used in the breeding scheme. To create a stable and fully developed ensemble system, the initial time for model integration is set at 0000 UTC 5 November 2007 so that it has a 5-day "spin-up" period. The initial perturbation is obtained by using the difference between the forecasts integrating from different initial time to a unified ending time. The T213 model runs at a cycle of 4 times a day, and each time it integrates at a 6-h lead time. After a 5-day ensemble cycling, the perturbations would have fully developed and thereafter, we make parallel comparisons between different GEPSs for the period 10–23 November 2007 as aforementioned.

2.4.3 14-member GEPS with initial perturbations generated by the ET technique (ET14)

All of the work flow is similar to that described in Section 2.4.2, except that we make use of the ET method to set up the GEPS.

2.4.4 24-member GEPS with initial perturbations generated by the ET technique (ET24)

Similar to the above, we make use of the ET method to set up the GEPS, but 24 members are used in the scheme.

To test the performance of different initial perturbation techniques under the same conditions, we run parallel tests with the BV based GEPS and the ET based GEPS, respectively. There are 14 ensemble members for both of them. Furthermore, to test the effect of ensemble size, we increase the member size of the ET based GEPS to 24. All the GEPSs run with a period of "spin up" to achieve adequate perturbation growth with different methods.

3. Results

To figure out the similarities and differences between the two methods, a rough comparison of perturbation evolution of 500-hPa geopotential height is conducted. Figures 2 and 3 show the ensemble mean and root mean square error (RMSE) of the 500-hPa geopotential height at the initial time of verification (0000 UTC 10 November 2007) and 10 days later (0000 UTC 20 November 2007). To compare equally populated ensembles, only 14 members for each method (BV14 and ET14) are used. Each ensemble is verified against the same analysis at the corresponding time.

At the initial verification time, the ensemble mean and RMSE of 500-hPa geopotential height by use of the BV and ET methods are almost the same, although each ensemble system has passed the 5-day spin-up with its own cycles (Fig. 2). Figure 3 shows the 10-day forecast valid at 0000 UTC 20 November 2007. It is shown that even after 10-day integrations, the forecast errors for the two methods are still very similar in both magnitude and error distributions,

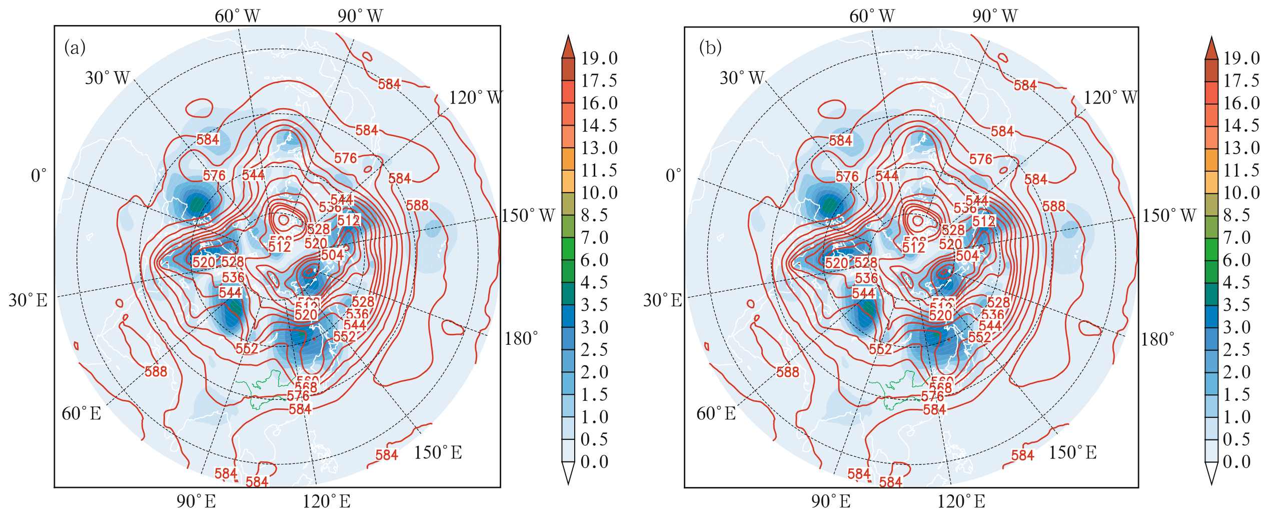

Fig. 2. Ensemble mean (contours; dagpm) and RMSE (shadings) of 500-hPa geopotential height at 0000 UTC 10 November 2007 by use of the BV (a) and the ET (b) methods. Contour interval is 4 dagpm and shading level interval is 0.5 dagpm.

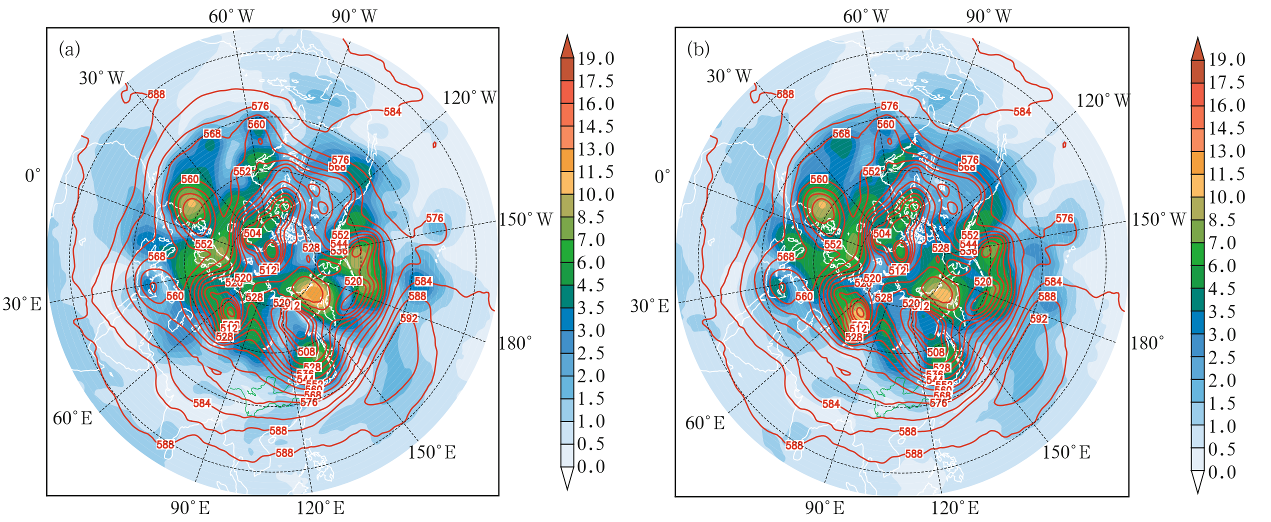

Fig. 3. As in Fig. 2, but at 0000 UTC 20 November 2007.

which proves that the perturbation structure for the ET and BV methods is very similar. This similarity in perturbation structure may partly result from the similarities of the two methods in creating the ensemble initial conditions, and it may also be due to the "inadequacy" of cycles at the verification time.

Different measures emphasizing different aspects of forecast performance can be used to assess the statistical reliability, resolution, and discrimination of a forecast system. In this study, the performance of the three EPSs will be compared using a comprehensive set of standard ensemble and probabilistic forecast verification methods, including calculations of the anomaly correlation coefficient (ACC), RMSE, the Brier skill score, the outlier statistics (a measure of reliability), and the area under the relative operating characteristics (ROCs; a measure of discrimination). Refer to Candille and Talgrand (2005), Atger (1999), Toth et al. (2003, 2005) for a detailed description of these scores.

3.1 Forecast skill for ensemble mean

The RMSE and ACC are influenced by both systematic errors (e.g., a low bias in ensemble spread, degrading reliability) and random error variance (reducing a forecast system's ability to distinguish among different events, leading to reduced resolution). Therefore, these two scores offer good measures for the overall forecast performance.

ACC is used to quantify the spatial correlation between forecast and observed deviations from climatology. It often serves as a basic score to evaluate the performance of a numerical weather prediction system. Normally, ACC = 0.6 (subjective) is defined as a basic synoptic-scale forecast skill.

For ACC, the ensemble result is also compared with that of the control forecast. It is worth noting that all GEPSs are more skillful than the control run in terms of ACC (see Fig. 4). The gain in predictability from running an ensemble (instead of a single control forecast) is about 0.5 day in view of the synoptic-scale forecast skill (ACC = 0.6), and the benefit of ensemble

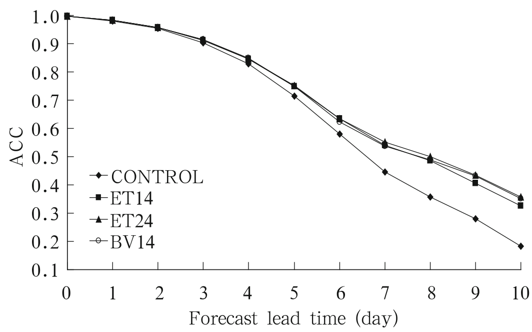

Fig. 4. Forecast skill (pattern ACC) for the control run (diamond) and the ensemble mean of the GEPS (rectangle for ET14, triangle for ET24, and open circle for BV14). Values are for 500-hPa geopotential height over the Northern Hemisphere, and if not specifically mentioned, all of the calculations in this article refer to this variable.

versus control run becomes more significant as the lead time increases. These gains are due to the nonlinear filtering effect that the ensemble averaging offers in terms of error growth reduction (Toth and Kalnay, 1997). Comparison among results of the three ensemble prediction tests show that although all the ensemble predictions gain some skills in general, the skills of the two ET based ensembles achieve a slight higher score than the BV method for less than 7-day forecast lead time. However, for longer lead time, ET14 gets worse compared with BV14, and ET24 gains a highest score throughout all the 10-day forecasts.

3.2 Ensemble mean error and spread

For a perfect ensemble prediction system, the spread of the system is equal to the RMSE of the ensemble mean. Figure 5 compares the RMSE of the ensemble mean forecast and the mean spread of corresponding test ensemble systems as a function of lead time. At the beginning of the test, the ensemble spread is slightly larger than the ensemble mean error, indicating a larger than desired initial spread. However, the error growth becomes larger than the ensemble spread, which indicates that all test ensemble systems underestimate ensemble spread. This is likely because not all possible sources of model related uncertainty are accounted for. Regarding the relationship between the control run (model integration with the best initial conditions or control analysis) and ensemble forecasts, it is obvious that all ensemble systems obtain smaller RMSE of the ensemble mean than that of the control run, suggesting the advantage of the ensemble system. It is seen that the BV based ensemble has the least spread at all forecast lead time (BV14-SP), and with the largest error growth among all ensembles except for the 9- and 10-day forecast lead time (BV14-RM). In comparison, the 14-member ET based ensemble shows a larger spread growth rate and smaller error growth than the BV method. Unlike the BV perturbation technique, the perturbation in each member of the ET based ensemble is almost orthogonal with others. The more ensemble members we have, the more orthogonality we could achieve within an ensemble system, which may explain why the 24-member ET based ensemble system (ET24-SP) possesses the best spread among all test systems. Furthermore, for all three ensemble systems, the smaller error growth is almost related to larger ensemble spread, indicating the consistent behavior of the ensembles.

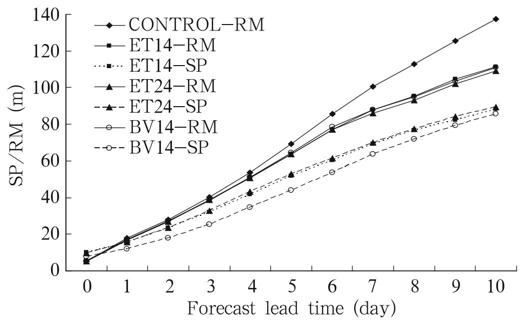

Fig. 5. Evolution of RMSE of the control run and the GEPS mean error (“RM”) and spread (“SP”) as a function of lead time for different initialization methods.

3.3 The continuous ranked probability score

To evaluate the global skill of an EPS, we check the CRPS score of the EPS (Candille et al., 2007). The CRPS measures the distance between the predicted and observed cumulative density functions (CDFs) of scalar variables.

$$\mathrm{CRPS} = \int_{-\infty}^{\infty} [P_{\mathrm{fcst}}(x) - P_{\mathrm{obs}}(x)]^2 \mathrm{d}x, \tag{5}$$

where $P_{\mathrm{obs}}(x)$ and $P_{\mathrm{fcst}}(x)$ are the observed and predicted CDFs, respectively. The CRPS is the generalization of the Brier score over all possible thresholds of the variable under consideration. The CRPS is negatively oriented, reaching its minimum value of zero for a perfect deterministic system. A higher value of the CRPS indicates a lower skill of the EPS. It has the advantage of being sensitive to the whole range of values of the parameter of interest, and does not depend on predefined classes at the same time.

Figure 6a shows the forecast skill of all the three ensembles at 10-day lead time for 500-hPa geopotential height. It is seen that the difference gets larger and larger as forecast time increases. Overall, the 24-member ET based ensemble system has the least CRPS score compared with the other two, and the score of the BV method is the highest within the forecast range, so it achieves the least forecast skill.

Figure 6b shows the time evolution of the CRPS

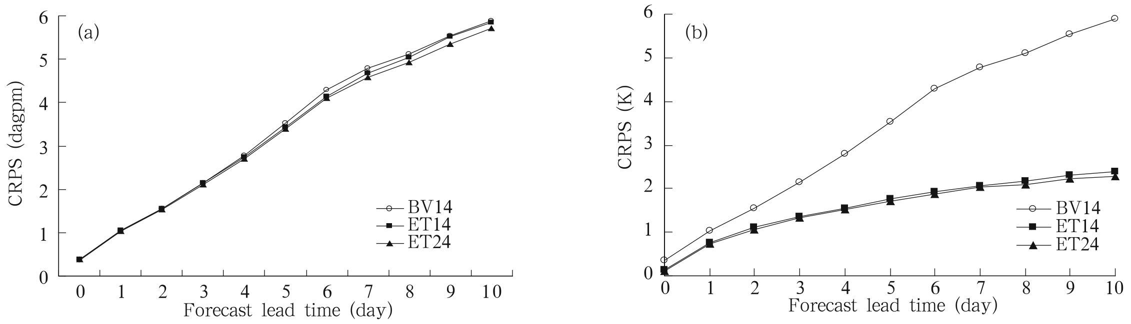

Fig. 6. CRPS score for different GEPS experiments for 500-hPa geopotential height (a) and 850-hPa temperature (b).

for temperature at 850 hPa (T850). Like the 500-hPa geopotential height (H500), the forecast skill from the best to the worst is listed as ET24>ET14>BV14. However, it is obvious that the difference between the two ET based ensembles is small, indicating that increasing of the ensemble size does not guarantee a significant gain in forecast skill by means of CRPS score. Nevertheless, there is a marked difference between the ET and the BV based ensemble forecast systems, implying that selection of the initial perturbation techniques is important for the prediction results of temperature and geopotential height.

The CRPS provides an all-inclusive measure of the skill of an EPS, and can be decomposed into reliability and resolution in order to evaluate the two main characteristics of a probabilistic prediction system (Toth et al., 2003). Reliability and resolution are two general attributes of forecast systems, and they determine the usefulness of a probabilistic forecast system. Hersbach (2000) and Candille et al. (2007) have proposed a reliability/resolution decomposition of the CRPS in the discretized case:

$$\mathrm{CRPS} = \mathrm{Reli} + \mathrm{Reso}, \tag{6}$$

$$\mathrm{Reli} = \sum_{i=0}^{N} g_i (o_i - p_i)^2, \tag{7}$$

$$\mathrm{Reso} = \sum_{i=0}^{N} g_i o_i (1 - o_i), \tag{8}$$

where N is the ensemble size, g_i is the average width of the bin i (Euclidean distance between consecutive ensemble members x_i and x_{i+1} for $0 < i < N$, and Euclidean distance between the observation and the outliers for $i = 0$ or $i = N$), o_i can be seen as the average frequency when the observation is less than the middle of the bin i, and p_i is the fraction i/N. Reli measures the reliability of the ensemble system and measures the difference between the resolution of the EPS and the uncertainty associated with the variable considered. It is noted that the uncertainty does not depend on the prediction system. Reliability indicates the property of statistical consistency between predicted probabilities and observed frequencies of occurrence of the event under consideration; while resolution shows the ability of a forecast system to discern sub-sample forecast periods with different relative frequencies of the event. Like the CRPS itself, the two components of the decomposition are negatively oriented (Hersbach, 2000), that is, the smaller those scores are, the better an EPS is. Reli is equal to 0 if the system is perfectly reliable and a significant positive value of Reli quantifies the lack of reliability of the system.

The CRPS shows that the overall forecast skill of the ET based ensemble is higher than the BV ensemble. However, as shown in Fig. 7a, the resolutions of all the three ensembles are almost the same. In other words, the skill propriety of the ET method comes from the attribute of reliability. This indicates that the ET method is superior to the BV in issuing the observed climatological distribution or reliability. In addition, the ensemble size contributes to reliability since the score of the 24-member ET ensemble is much smaller than that of the 14-member ET ensemble (Fig. 7b).

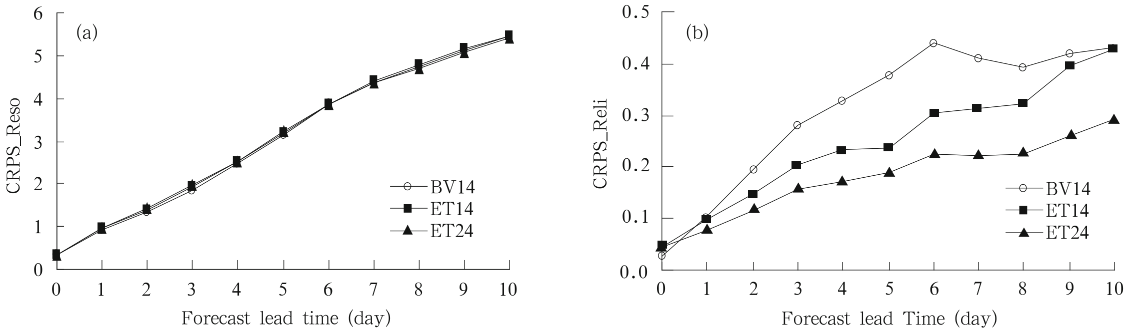

Fig. 7. Decomposition of the CRPS for different GEPS experiments for geopotential height at 500 hPa. (a) Resolution and (b) reliability.

3.4 The reduced centered random variable

As an important attribute of an EPS, the reliability property of different ensembles will be further explored. Let us consider a scalar variable x and an associated EPS that produces an ensemble of values from which the predicted PDF can be obtained.

The EPS is such a reliable mean that any verifying observation will be indistinguishable from the values predicted by the ensemble system. Considering observation x_0, the mean m, the observation error σ_0, and the standard deviation σ of corresponding ensemble of predictions, a new method to verify attributes of EPS is the reduced centered random variable (RCRV) (Candille et al., 2007), which is expressed as:

$$y = \frac{x_0 - m}{\sqrt{\sigma_0^2 + \sigma^2}}. \tag{9}$$

The average of y is defined as

$$b = [y]. \tag{10}$$

The variable b is computed over all grids of the ensemble system and represents the weighted bias between the ensemble and the observation. The standard deviation of y,

$$d = \sqrt{\frac{M}{M-1}E[(y-b)^2]}, \tag{11}$$

where M is the sample size on which the statistics are computed, and d is a measure of systematic over- or under-dispersion of the ensemble.

The first two moments of b and d provide a simple diagnostic of the indistinguishability between the values predicted by the ensemble and the observation. A perfectly reliable system has no bias (i.e., $b = 0$) and a dispersion equal to 1 ($d = 1$). A significant negative (positive) value of b indicates a negative (positive) bias. A value of d significantly greater (smaller) than 1 characterizes the under-dispersion (over-dispersion) of the system.

To further demonstrate the improvement of the ET ensemble prediction over the BV ensemble prediction, we show in Fig. 8 the bias and the dispersion derived from the RCRV. The significant reliability gain is mainly due to the dispersion improvement, which is significant for all forecast ranges (up to 0.2–0.3 K); while the bias difference is comparatively less, especially within the forecast range of 4-day lead time, and the normal difference between different ensembles is less than 0.05 K. Note that all ensembles are under dispersive ($d \sim$ [1.2, 1.9]) with positive bias ($b \sim$ [0.2, 0.6]). The dispersion of the 24-member ET ensemble is better than the 14-member ET ensemble, and the latter in turn is better than the BV ensemble. For the bias, the 24-member ET ensemble is the best for all forecast ranges, while the 14-member ET ensemble is better within the forecast range of days 2–7, but worse for days 8–10, compared with the breeding based ensemble prediction.

4. Conclusions and discussion

In this paper, we have configured three global ensemble prediction systems (GEPSs). The GEPSs have different numbers of ensemble members and apply the breeding vector (BV) method or the ensemble transform (ET) method to generate initial perturbations.

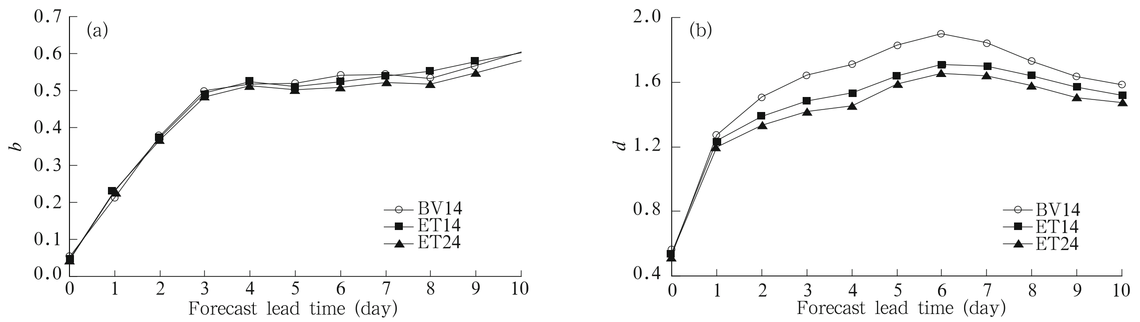

Fig. 8. Bias (a) and dispersion (b) between different GEPSs.

The forecast model used is the spectral model of T213/L31, which contains the 3DVAR assimilation module and could direct assimilate large amount of observations including the ATOVS data. To compare the performance of different ensemble systems, a series of ensemble verification methods are introduced. We mainly focus on the most important attributes of ensemble forecast systems such as forecast skill for ensemble mean, resolution, and reliability. The major conclusions are obtained as follows:

1) The forecast skill measured by ACC of the ensemble mean of 500-hPa geopotential height indicates that the predictability of the BV and ET based ensembles is quite similar, and all of them have obvious advantages over the single deterministic forecast.

2) The RMSE of the BV method is slight higher than the ET method, while the spread of the latter is a little larger than the former. Further, the increase rate of the ET is faster than the BV, indicating that the ET ensemble prediction exhibits a better skill on description of uncertainty.

3) Detailed evaluation of resolution and reliability (CRPS and RCRV) indicates that the superior skill of the ET results from reliability. Evaluation of both geopotential height and temperature fields shows similar results.

4) The size of the ensemble (number of ensemble members) in the ET ensemble prediction system not only extends predictability of ensemble mean compared with the deterministic forecast, but also increases the uncertainty description skill from different sides.

In short, comparisons between two different initial perturbation methods, i.e., the BV method and the ET method, show that the ET technique has a somewhat higher skill over the BV method. This is likely because that the former adopts the advantage of the breeding method in representing the nonlinear extension of the Lyapunov vectors; meanwhile, there is a comparatively better orthogonality among ensemble members and the orthogonality will increase as the number of ensemble members increases. This explains why the 24-member ensemble is superior to the 14-member ensemble for all evaluation results. Furthermore, the ensemble transform technique is more consistent with the data assimilation system; hence, it is a promising choice in ensemble prediction system construction for operation at the NMC of China.

Acknowledgments. We are grateful to Drs. Wei Mozheng and Zhu Yuejian for their help and suggestions on construction of the ensemble transform based ensemble prediction system at the NMC/CMA.

REFERENCES

Atger, F., 1999: The skill of ensemble prediction systems. *Mon. Wea. Rev.*, **127**(9), 1941–1953.

Bishop, C. H., and Z. Toth, 1999: Ensemble transformation and adaptive observations. *J. Atmos. Sci.*, **56**(11), 1748–1765.

Buizza, R., and T. N. Palmer, 1995: The singular vector structure of the atmosphere general circulation. *J. Atmos. Sci.*, **52**(9), 1434–1456.

——, P. L. Houtekamer, Z. Toth, et al., 2005: A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Mon. Wea. Rev.*, **133**(5), 1076–1097.

Candille, G., and O. Talgrand, 2005: Evaluation of probabilistic prediction systems for a scalar variable. *Quart. J. Roy. Meteor. Soc.*, **131**(609), 2131–2150.

—–, C. Cote, P. L. Houtekamer, et al., 2007: Verification of an ensemble prediction system against observations. *Mon. Wea. Rev.*, **135**(7), 2688–2699.

Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting*, **15**(5), 559–570.

Leith, C. S., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, **102**(6), 409–418.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**(2), 130–141.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC, the generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**(12), 2317–2330.

—–, and —–, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**(12), 3297–3319.

Toth, Z., O. Talagrand, G. Candille, and Y. Zhu, 2003: Probability and Ensemble Forecasts. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. Jolliffe, I. T., and D. B. Stephenson, Eds., Wiley, 137–163.

—–, —–, and Y. Zhu, 2005: The attributes of forecast systems: A framework for the evaluation and calibration of weather forecasts. *Predictability of Weather and Climate*. Palmer, T. N., and R. Hagedorn, Eds., Cambridge University Press, 584–595.

Wei, M., Z. Toth, R. Wobus, et al., 2008: Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system. *Tellus*, **60**A, 62–79.