

Study on ETKF-Based Initial Perturbation Scheme for GRAPES Global Ensemble Prediction*

MA Xulin^{1,2†}(马旭林), XUE Jishan²(薛纪善), and LU Weisong¹(陆维松)

¹ Key Laboratory of Meteorological Disasters of Ministry of Education, NUIST, Nanjing 210044

² Chinese Academy of Meteorological Sciences, Beijing 100081

(Received September 11, 2008)

ABSTRACT

Initial perturbation scheme is one of the important problems for ensemble prediction. In this paper, ensemble initial perturbation scheme for Global/Regional Assimilation and PrEdiction System (GRAPES) global ensemble prediction is developed in terms of the ensemble transform Kalman filter (ETKF) method. A new GRAPES global ensemble prediction system (GEPS) is also constructed. The spherical simplex 14-member ensemble prediction experiments, using the simulated observation network and error characteristics of simulated observations and innovation-based inflation, are carried out for about two months. The structure characters and perturbation amplitudes of the ETKF initial perturbations and the perturbation growth characters are analyzed, and their qualities and abilities for the ensemble initial perturbations are given.

The preliminary experimental results indicate that the ETKF-based GRAPES ensemble initial perturbations could identify main normal structures of analysis error variance and reflect the perturbation amplitudes. The initial perturbations and the spread are reasonable. The initial perturbation variance, which is approximately equal to the forecast error variance, is found to respond to changes in the observational spatial variations with simulated observational network density. The perturbations generated through the simplex method are also shown to exhibit a very high degree of consistency between initial analysis and short-range forecast perturbations. The appropriate growth and spread of ensemble perturbations can be maintained up to 96-h lead time. The statistical results for 52-day ensemble forecasts show that the forecast scores of ensemble average for the Northern Hemisphere are higher than that of the control forecast. Provided that using more ensemble members, a real-time observational network and a more appropriate inflation factor, better effects of the ETKF-based initial scheme should be shown.

Key words: GRAPES, ensemble transform Kalman filter (ETKF), initial perturbation, ensemble prediction

Citation: Ma Xulin, Xue Jishan, and Lu Weisong, 2009: Study on ETKF-based initial perturbation scheme for GRAPES global ensemble prediction. *Acta Meteor. Sinica*, **23**(5), 562–574.

1. Introduction

Attributed to initial errors, model errors, and weather pattern interaction, the forecast skill of numerical weather forecast has been limited to the uniform certainty forecast for the nonlinear atmospheric motion. Leith (1974) pointed out that ensemble prediction could improve forecast skill, and since then the research of ensemble initial perturbation has been greatly developed. Many schemes such as Monte Carlo stochastic perturbations (Hollingsworth, 1980) and time lagged average perturbations (Hoffman and Kalnay, 1983), aimed at estimating analy-

sis error probability distribution function, have been successively developed. Based on the analysis of numerical prediction error, the Breeding of Growing Modes (BGMs) and improved BGM schemes (Toth and Kalnay, 1993, 1997) used at the U.S. National Centers for Environmental Prediction (NCEP) have been developed. These techniques obtained a group of pairing initial perturbations through adjusting analysis errors. The Singular Vector (SV) method (Buizza and Palmer, 1995; Molteni et al., 1996) was successfully applied at the European Centre for Medium-range Weather Forecasts (ECMWF) by identifying the most rapidly increasing modes in the operational

*Supported by the National Natural Science Foundation of China under Grant Nos. 40675064, 40518001, and 40675062, and CMA NWP Innovational Research Project—"Key Technology of Global Operational Data Assimilation System".

[†]Corresponding author: xulinma@nuist.edu.cn.

ensemble prediction. The ensemble initial perturbations constructed by observational perturbation scheme (Houtekamer et al., 1996; Houtekamer and Mitchell, 1998) also well presented the uncertainties of analysis (Buizza et al., 2005). Over the past decade, Chinese scientists have embarked on the development of ensemble forecast and achieved some exciting results (Fan, 1999; Li and Chen, 2002; Yang et al., 2002; Chen et al., 2003). Zhong and Wang (2004) studied the feasibility of physical ensemble for weather and climate feature of the abnormal flood during summer in the eastern China by use of the physical ensemble construction method of MM5 limited regional model. Chen and Yan (2005) designed a new initial perturbation method for ensemble mesoscale heavy rain prediction, known as the Different Physical Mode Method (DPMM). To improve the ensemble forecast, the approach adopted by the DPMM attempted to generate initial perturbations that had reasonable mesoscale circulation structures, and could reflect the uncertainty in convective instability. The BGM and SV method, however, are still confined to precisely describing the analysis uncertainties (Wei and Toth, 2003), while the DPMM only applied well to the mesoscale weather system research. Therefore, more investigations on ensemble forecast initial perturbation schemes are needed.

Ensemble initial perturbation schemes aim to generate unrelated initial perturbations so as to make ensemble initial perturbation covariance interpret analysis error covariance of initial time more accurately. Therefore, error structure and perturbation amplitudes illustrated by initial perturbations become two key problems in the research of ensemble initial perturbations. The error structure reflects the normal modes of analysis error variance distribution in the phase space, while amplitude of the initial perturbations should represent equivalent relations between analysis error and forecast error, as well as reflect directly on the spread of the forecast ensemble. Using a constant factor, BGM scheme's initial perturbations cannot reflect the influence of the spatial variations of observation density and accuracy on the analysis fields, as well as the adjustments of different forecast

error variations. Hamill et al. (2000) pointed out that regional rescaling factor adopted by improved BGM scheme can also bring about greater illusive noise of ensemble spread. The observational perturbation scheme could lead to smaller ensemble spread with leading integration time, though it can absorb observation information. Derived from the ensemble transform Kalman filter (ETKF) theory (Bishop et al., 2001), the ETKF initial perturbation scheme (Wang and Bishop, 2003) offset the disadvantages of BGM scheme by using a transform matrix restricted in ensemble subspace. It transforms forecast perturbations into analysis perturbations in accordance with Kalman filter error covariance equation. In addition, ETKF scheme is less costly. Bishop et al. (2003) pointed out that ETKF initial perturbation scheme can estimate the prediction and analysis error covariance, more accurately thus making it better than BGM ensemble scheme. Wang and Bishop (2003) also showed similar results by comparing BGM and ETKF ensemble scheme.

In China, the research of ensemble initial perturbation is still immature. On the other hand, when GRAPES global numerical forecast system has been achieved primarily, and the related scientific problems of ensemble forecast are necessary to develop. This paper attempts to establish the GRAPES global ensemble prediction system (GEPS) after preliminary research of ETKF initial perturbation had been carried out to investigate the structure of initial perturbations, the characters of the increasing perturbations, and the forecast efficiency. Section 2 mainly deals with the ETKF initial perturbation scheme, while ensemble forecast experiments are carried out in Section 3. Section 4 discusses the feature and analysis of the GRAPES ensemble forecast initial perturbations. Section 5 focuses on the verification of the ensemble forecast. Finally, summary and conclusions are given in Section 6.

2. ETKF initial perturbation scheme

The ensemble transform Kalman filter (ETKF) is an adaptive observational method first derived from

Kalman filter theory. It can approximately represent the analysis error covariance matrix and the forecast error covariance matrix in terms of ensemble perturbations to construct ensemble initial perturbations. ETKF initial perturbation scheme extends the BGM scheme while getting rid of unchanged analysis error variance and obtaining orthogonal perturbations. In terms of uncertainty of model error, observational error, and subsequent analysis error covariance, ETKF therefore is a suboptimal Kalman filter (Daley, 1991). Compared to other ensemble Kalman filter, forecast error covariance in ETKF is only to approximately estimate the analysis error covariance at the initial time, but does not update the ensemble mean state.

For the ensemble forecast, \mathbf{X}^f and \mathbf{X}^a are defined as forecast and analysis perturbations in contrast to control forecast, respectively. Mathematically, $\mathbf{X}^f = (x_1^f - x_0^f, x_2^f - x_0^f, \dots, x_K^f - x_0^f)$ and $\mathbf{X}^a = (x_1^a - x_0^a, x_2^a - x_0^a, \dots, x_K^a - x_0^a)$, where x_i^f and x_i^a ($i = 1, 2, \dots, K$) denote the K ensemble forecast and analysis perturbations. The superscript "a" is the analysis from data assimilation, while the subscript "0" denotes control forecast. The ensemble-based forecast and analysis covariance matrices can then be expressed as

$$\mathbf{P}^f = \mathbf{Z}^f \cdot (\mathbf{Z}^f)^T, \quad (1)$$

$$\mathbf{P}^a = \mathbf{Z}^a \cdot (\mathbf{Z}^a)^T, \quad (2)$$

where \mathbf{P}^f and \mathbf{P}^a refer to ensemble forecast and analysis error covariance, respectively. The superscript "T" indicates the matrix transpose. Let $\mathbf{Z}^f = \mathbf{X}^f / \sqrt{K-1}$, $\mathbf{Z}^a = \mathbf{X}^a / \sqrt{K-1}$, and K is the number of the ensemble members. For a given set of ensemble forecast perturbations \mathbf{Z}^f , the analysis perturbations \mathbf{Z}^a can be solved from the Kalman filter error covariance update equation

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

where \mathbf{R} is observational error covariance matrix, and refers to linear observational operator that converts forecast grid value to observational station. ETKF transforms forecast perturbations into analysis perturbations through Eq. (4).

$$\mathbf{Z}^a = \mathbf{Z}^f \cdot \mathbf{T}, \quad (4)$$

where \mathbf{T} is transformation matrix from forecast perturbations into analysis perturbations in the ensemble

subspace. Substituting Eqs. (1) and (2) into Eq. (3), we can obtain transformation matrix \mathbf{T} ,

$$\mathbf{T} = \mathbf{C}(\mathbf{\Gamma} + \mathbf{I})^{-1/2}. \quad (5)$$

Notes that columns of the matrix \mathbf{C} contain the eigenvectors of $(\mathbf{Z}^f)^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{Z}^f$, and the corresponding eigenvalues are nonzero elements of the diagonal matrix $\mathbf{\Gamma}$. Since $(\mathbf{Z}^f)^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{Z}^f$ is a real symmetric matrix, the corresponding eigenvalues are real, and the eigenvectors are orthogonal. That the analysis perturbations produced by ETKF initial perturbation scheme are orthogonal in standard observational space is an important feature. In Eq. (3), the error covariance varies with the differences of observation density and the precision in the spatial distribution due to the effect of observation operator \mathbf{H} and observation error \mathbf{R} . Consequently, for a given set of ensemble forecast in an optimal assimilation scheme, forecast perturbations covariance matrix from ensemble is equal to the real forecast error covariance matrix, therefore the analysis perturbation covariance matrix transformed by ETKF must be strictly equal to real analysis error covariance matrix.

The forecast perturbations are centered about the ensemble mean (or control forecast), i.e., $\sum_{i=1}^K z_i^f = 0.0$. However, this does not ensure that the corresponding analysis perturbations produced by ETKF to necessarily have the same feature ($\sum_{i=1}^K z_i^a \neq 0.0$). Thus, to ensure that the analysis error covariance is invariable, as well as to center the analysis perturbations about the control analysis, we employ the spherical simplex centering scheme (Wang et al., 2004) to create the initial ETKF perturbations. That is, to change Eq. (4) into

$$\mathbf{Z}^a = \mathbf{Z}^f \mathbf{T} \cdot \mathbf{C}^T, \quad (6)$$

where \mathbf{C}^T is transpose matrix of the matrix \mathbf{C} in Eq. (5). This centering scheme is capable of maintaining the original attributes of initial ensemble perturbations when the numbers of the ensemble members are not too large.

Since the degrees of freedom of forecast model state space are larger than the number of ensemble

members and that model error is not considered, in practice, the analysis error covariance obtained from the covariance of the transformed ensemble in Eq. (2) is much smaller than the error covariance. To avoid this problem, the inflation factor method (Wang and Bishop, 2003) is adopted to efficiently ensure that the global ensemble error of 12-h forecast in observational space is identical to the global control forecast error. Mathematically,

$$\mathbf{Z}_i^a = \mathbf{Z}_i^f \mathbf{T}_i \cdot \mathbf{C}_i^T \cdot \prod_i, \quad (7)$$

where \prod_i is the scalar inflation factor at time t_i ($i = 1, 2, \dots$, refers to different cycling time), and it is defined as

$$\prod_i = \prod_{i=1} \sqrt{\alpha_i} = \sqrt{\alpha_1 \alpha_2 \cdots \alpha_i}. \quad (8)$$

Parameter α_i is expressed as $\alpha_i = (\tilde{\mathbf{d}}_i^T \tilde{\mathbf{d}}_i - N) / \sum_{k=1}^{K-1} \lambda_i$, and $\tilde{\mathbf{d}}_i$ is the difference between observation value and forecast value in observational space at time t_i , also referred as innovation. K is the ensemble member, and λ_i is the diagonal element of Γ in Eq. (5). N denotes the valid observation number. α_i in Eq. (8) is solved statistically for global observational system with the assumption that the forecast errors and the analysis error are uncorrelated. When the 12-h ensemble perturbation variance is not equal to the control forecast error variance, α_i is forced to adjust to make ensemble perturbation variance in accordance with control forecast error variance in the global observational space.

Using information from observation data, ETKF initial perturbation method transforms the mode of ensemble perturbations in error subspace and adjusts the amplitudes of perturbations according to error covariance update equation. Accordingly, it establishes the relationship between analysis perturbations and the spatial distribution and quality of observations, and properly reduces larger variance in larger ensemble and fewer adjustments to smaller ensemble. This method thus efficiently gets rid of rigid estimates of analysis error variance due to the lack of the information of observation data in BGM method while still holding the orthogonal characters in observational space.

3. Ensemble forecast experiments

3.1 Simulated observation data

To simplify the problem, we assume that the spatial distribution of observation system is invariant in the cycling process of ensemble forecast system, and this simplified simulated observational system is used to evaluate the effect of ETKF initial perturbation scheme and its response to the decrease of forecast error. We assumed that the observational system only consists of three elements: wind (u, v) and temperature (T) at 850, 500, and 200 hPa. The analysis fields of T213 with a horizontal resolution of 0.5625° at 0000 and 1200 UTC are interpolated horizontally to the

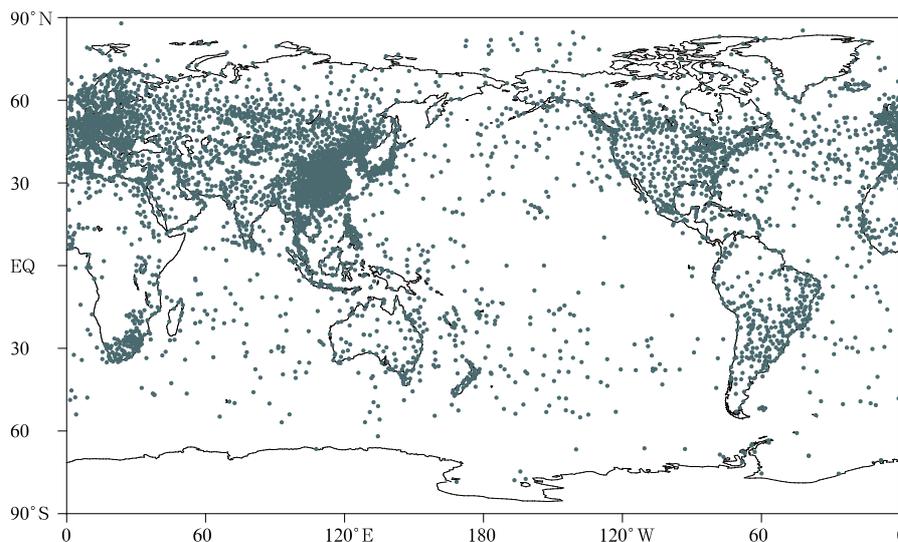


Fig. 1. Network of simulated observational stations. Black dots denote locations of observational stations.

5186 global observational stations (Fig. 1) to create simulated observation data. Even though these settings of the experiment are not coherent to real observations, they do not affect the numerical experimental results of ETKF initial perturbation scheme. As the analysis of operational T213 has already included the dynamical restrictions of initial guess and some information about real observation data, the observation errors of ETKF scheme are slightly smaller than that of the operational three-dimensional variational (3DVAR) data assimilation system. Suppose observation errors are uncorrelated, the observation error covariance matrix \mathbf{R} is diagonal matrix.

3.2 Construction of initial perturbations

The experiments employ GRAPES global model and 3DVAR data assimilation system. The horizontal resolution of the forecast model and data assimilation system is $1^\circ \times 1^\circ$, with 31 model levels under terrain-following coordinate and 17 pressure levels, respectively. The background fields for the data assimilation system come from 12-h forecasts of the operational T213, and the analysis assimilated with TEMP, SYNOP, SHIP, and AIR data is used as initial values for the global forecast model, to make 96-h forecast as control forecast of ensemble forecast. After the addition of stochastic perturbations into the control variables of 3DVAR data assimilation system, a set of raw initial perturbations with dynamical and physical constraints are obtained at the start time of ensemble forecast. These initial perturbations are used in Eq. (5) to calculate transform matrix, and the inflation factor is solved by Eq. (8). Finally, the update ensemble analysis perturbations for the next time are obtained from Eq. (7). The sum of analysis perturbations and control forecasts become the model initial values for each ensemble forecast member to make a set of 96-h ensemble forecast. Subsequently, the new ETKF ensemble analysis perturbations, which are generated from 12-h forecast of all ensemble members, are combined with the control forecast to construct the initial perturbations. Then, the cycle of ETKF ensemble forecast is carried out, with a time interval of 12 h and an ensemble forecast time 96 h. In our ensemble forecast experiments, the ensemble members consist of control

forecast and 13 ensemble perturbations, i.e., 14 ensemble members. The time periods are from 2 July to 30 October 2006, with startup time at 0000 and 1200 UTC each day. The value of parameters is usually unstable in the initial periods of forecast cycle, due to the incomplete harmony of the raw initial perturbations generated by 3DVAR data assimilation system and initial perturbations constructed by ETKF. After 3 to 4 days of auto adjustment, α will oscillate around the value of 1 and the ensemble initial perturbations tend to be reasonable and stable. To remove the influence of perturbation adjustment during the initial periods, the following sections are devoted to study the 52-day ensemble forecast results from 10 July 2006.

4. Analysis of ensemble initial perturbation

The quality of initial perturbations directly affects the skill of ensemble forecast. Whether initial perturbations are capable of precisely reflecting the main characteristic modes and proper amplitude of analysis error variance is one of the key criterions for the effectiveness of initial perturbation scheme. A good initial perturbation method can capture the increasing possible analysis errors, meanwhile, each initial perturbation can maintain appropriate spread in forecast valid time, and then these ensemble forecasts can more accurately represent the actual atmospheric state.

4.1 The influence of observation to ensemble network density

One of the main features of ETKF initial perturbation scheme is that ensemble variance can exactly reflect the impact of variations of observational density on analysis error variance and forecast error variance when the ensemble member is large enough. Thus, to measure this influence of observations to ensemble variance, we will investigate the total energy of ensemble variance pointed out by Palmer et al. (1998). This method is considered most suitable for weather forecast and data assimilation. For one perturbation, one defines the total energy from wind and temperature using $\frac{1}{2}[u'^2(i, j, k) + v'^2(i, j, k)] + \frac{c_p}{T_r}T'^2(i, j, k)$, where u', v', T' are perturbations of the wind

components and temperature, respectively; c_p is the specific heat at constant pressure for dry air; T_r is the reference temperature; and i, j, k are indices for the horizontal and vertical directions in grid space (Wei and Toth, 2003; Wang and Bishop, 2003). First, the total energy of each perturbation member in each vertical level is computed, and then the total energy is averaged for all vertical levels and all ensemble members. Figure 2 shows the global distribution of energy spread of ETKF ensemble analysis perturbations averaged vertically. From Fig. 2, it can be seen that the perturbation amplitude of the total energy spread of analysis perturbations in the Northern Hemisphere is relatively large, whereas the Southern Hemisphere represents approximately zonal distribution. The total energy spread between 30°N and 30°S is the lowest in the tropics, which implies that the forecast error growth is small in these areas over 12-h intervals. The well consistency between the total energy spread and the observational distribution (Fig. 1) indicates that ETKF initial ensemble perturbation scheme can adjust ensemble variance properly through factors based on observational density so that initial perturbation variance can precisely reflect the features of analysis error variance. As mentioned in the introduction, the rescaling factor of BGM scheme is an empirical statistical value obtained from climatic data, which means

that it can only implicate the average state of analysis error variance over a long time without the influence of observations. The corresponding ensemble analysis perturbations, therefore, can hardly precisely suggest the attributes of analysis error variance. On the other hand, ETKF ensemble can absorb almost all the information contained in the observations spatial distributions, then adjust the effect observations on ensemble analysis error variance, and further improve the quality of ensemble initial perturbations.

Figure 3 shows the variation of total energy spread and ratio of the analysis (solid line) and 12-h forecast (dashed line) perturbations with the change in latitude (Fig. 3a) and height (Fig. 3b). Figure 3a indicates that the total energy spread is smaller in the tropical area where baroclinic instability is relative low, while the spread is larger in the middle and high latitude in the Northern and Southern Hemispheres. The difference between forecast and analysis perturbations spread at about 60°N or 60°S is apparently larger than that in the lower latitude. This is consistent with the horizontal distributions in Fig. 2. In order to analyze the vertical distribution of total energy spread, the averaged total energy at all grids in each vertical level is calculated. Figure 3b gives the total energy spread of the analysis (solid line) and forecast (dashed line) ensemble perturbations, as well as the

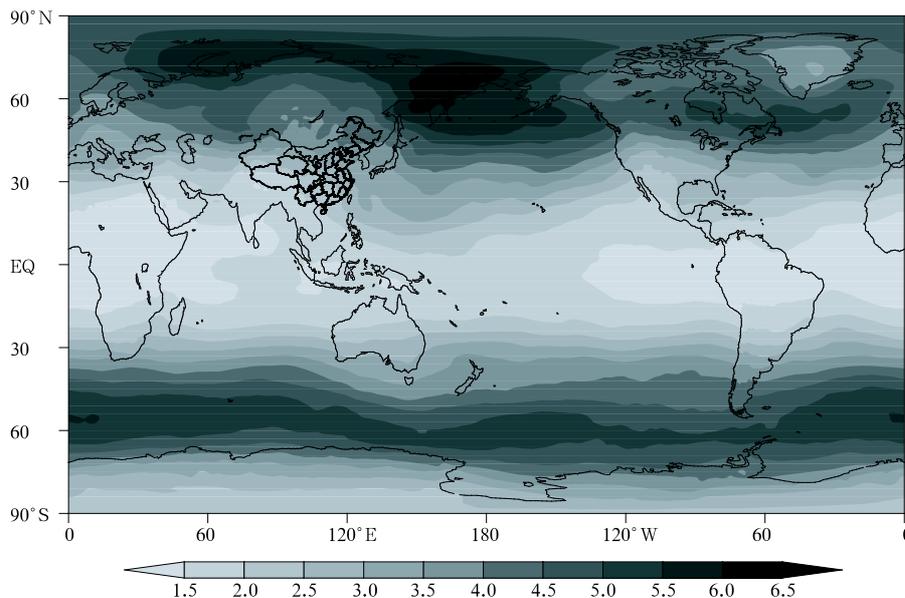


Fig. 2. Vertically averaged global distribution of energy spread (unit: J kg^{-1}) of analysis perturbations.

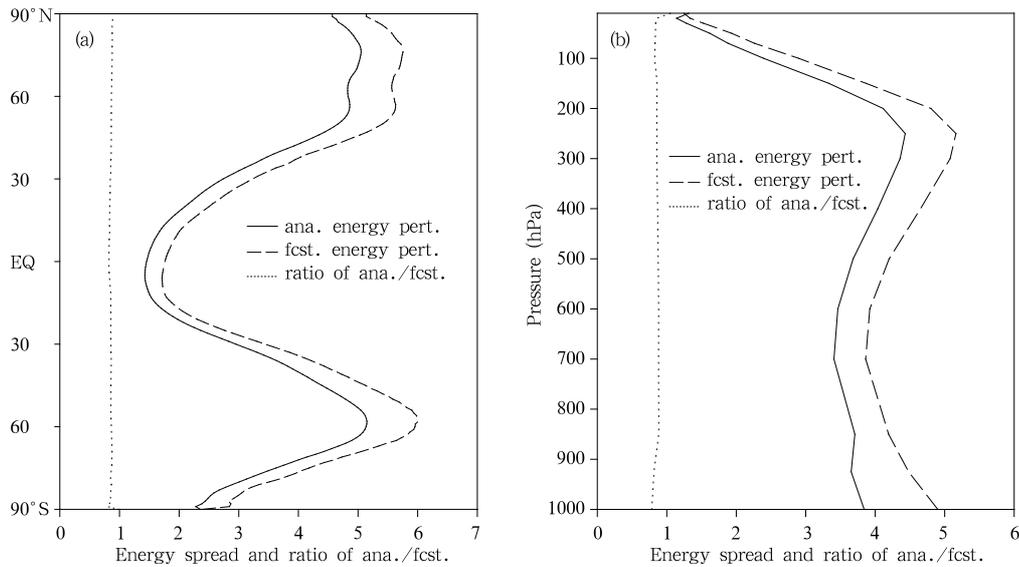


Fig. 3. Energy spread (unit: J kg^{-1}) distributions of ensemble perturbations (solid: analysis; dashed: forecast) and the ratio of analysis/forecast perturbations (dotted line). All the values are averaged over 52 days with (a) vertical distribution as function of pressure and (b) distribution by latitude.

vertical distributions of the ratio (dotted line) of the analysis and 12-h forecast perturbations, namely rescaling factor. Obviously, the total energy spread of analysis and forecast perturbations between 500 and 200 hPa are larger, similar to the results of Wei et al. (2006), whereas it decreases from lower level to higher level until 850 hPa and the differences between them are much larger. The averaged rescaling factor (dotted line) is almost the same except slightly smaller value below 850 hPa in Fig. 3b. This maybe due to the configuration of surface layer in GRAPES model, which needs further investigation. The ETKF initial perturbations scheme uses an inflation factor that is interrelated to innovation in data assimilation to adjust the analysis perturbations at each time so that appropriate spread of the initial ensemble perturbations could be kept all the time.

4.2 The variance distribution of ensemble perturbations

Figure 4 shows the time-averaged eigenvalue distribution of analysis (Fig. 4a) and 12-h forecast (Fig. 4b) perturbations of 14 ensemble members along different eigen-directions of the covariance matrices in the normalized observational space. For forecast and

analysis error covariance matrices, each eigenvalue represents the time-averaged ensemble variance. Of the 14 initial ensemble members, though centralized about control forecast, only 13 are orthogonal and linearly uncorrelated to each other. As shown in Fig. 4a, the 13 eigenvalues of analysis perturbations are almost the same, while the value of first rank for the 12-h forecast perturbations reaches 30936, and the minimum value is 14292. Obviously, the eigenvalue spectrum of analysis perturbations is much more evenly distributed than that of 12-h forecast perturbations. From Eqs. (5) and (6), the forecast perturbations are first rotated by matrix \mathbf{C} , then adjusted by diagonal matrix $(\mathbf{\Gamma} + \mathbf{I})^{-1/2}$, and finally rotated by \mathbf{C}^T to get the transformed analysis perturbations. The transform matrix in Eq. (5) is solved through error update Eq. (3), derived from optimal data assimilation scheme that has filtering functions. Hence, along eigen-directions in normalized observational space the eigenvalue spectrum of analysis error covariance matrix is more evenly distributed than that of forecast error covariance, which indicates that ETKF ensemble can equally distribute error variance into all eigen-directions in the ensemble subspaces, so as to incorporate more ensemble forecast error variance along all

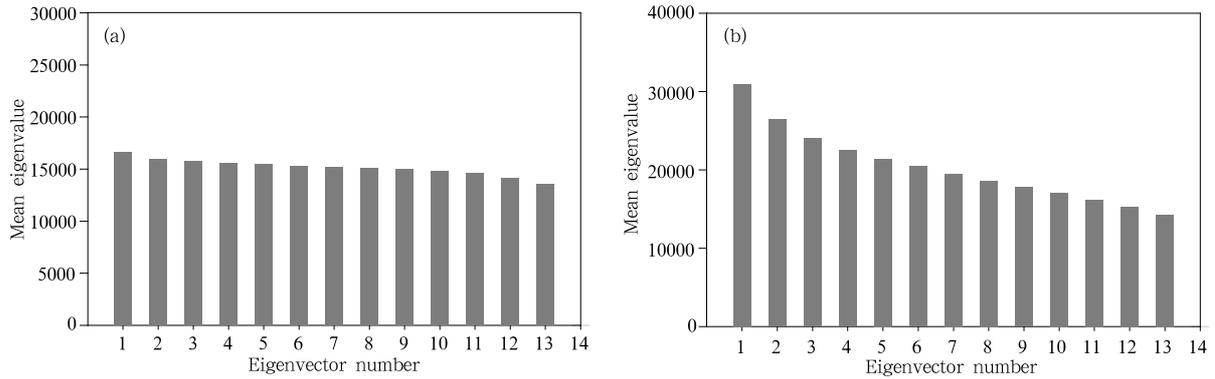


Fig. 4. The time-averaged eigenvalue distributions along different eigen-directions of covariance matrices in the normalized observational space. (a) Analysis and (b) 12-h forecast.

orthogonal eigen-directions in the subspaces. Consequently, this is an important feature that makes the ETKF scheme much better than other initial perturbation schemes.

4.3 The growth of ensemble perturbations

For linearly dynamic propagating forecast errors, the linear combination of the ensemble initial perturbations can represent the initial structures of the main eigenvector of the forecast error covariance. For actual nonlinear dynamic forecast system, the ensemble forecast can also identify large forecast errors brought by the rapid growth analysis errors. The appropriate growth of ensemble perturbations not only indicates that the initial perturbation scheme has the ability of capturing the main modes of analysis perturbation error variance, but also can affect the ensemble spread. Therefore, the growth of ensemble perturbations is an important aspect which can reflect the attributes of ensemble initial perturbations. Palmer et al. (1998) pointed out that the growth of SVs of total energy modes in the ensemble space can reflect approximately the growth characteristics of forecast error variance. Following the method employed by Bishop and Toth (1999) and Wang and Bishop (2003), we compute the maximal growth of total energy for 96-h forecast lead times in the GRAPES ensemble subspaces, as well as the energy growth averaged over all perturbations. The fastest energy growth is the mean value of the maximal energy growth over 52 days, while the mean energy growth is calculated by the growth of all growth

modes in ensemble subspace.

The fastest energy growth and mean energy growth as a function of forecast lead time in the whole experiment period (Fig. 5) show that the fastest energy growth of global average ensemble perturbations (solid line) has an evident growing tendency in the forecast lead time (Fig. 5a). The growing range is relatively smaller before 24 h, but growth increases clearly after 24 h, suggesting that the ETKF initial perturbation scheme for GRAPES global forecast model has a ability to capture the normal norms of the rapidly growing analysis error. Hence, this scheme is capable of reflecting the growth features of forecast error, and keeps the appropriate spread of ensemble forecast. As for different regions, the growth in the Northern Hemisphere (dashed line) is quite similar to that in the Southern Hemisphere (dashed-dot line) before 24 h while the largest energy growth after 24 h in the Northern Hemisphere tends to be below that of global, which is the opposite to the Southern Hemisphere. Literature review (Wang and Bishop, 2003) shows that for global ensemble prediction, the ensemble perturbation growth is more sensitive to the amplitude of initial perturbation, and is decreasing with the increase of initial perturbations. The lack of observations in the Southern Hemisphere (Fig. 1), which induces the small initial perturbations as a whole, is one possible reason to cause the difference in the fastest energy growth in the Southern and Northern Hemisphere. Figure 5b shows the average total energy norm growth of 14 perturbations. In general, it reflects the growth characteristics

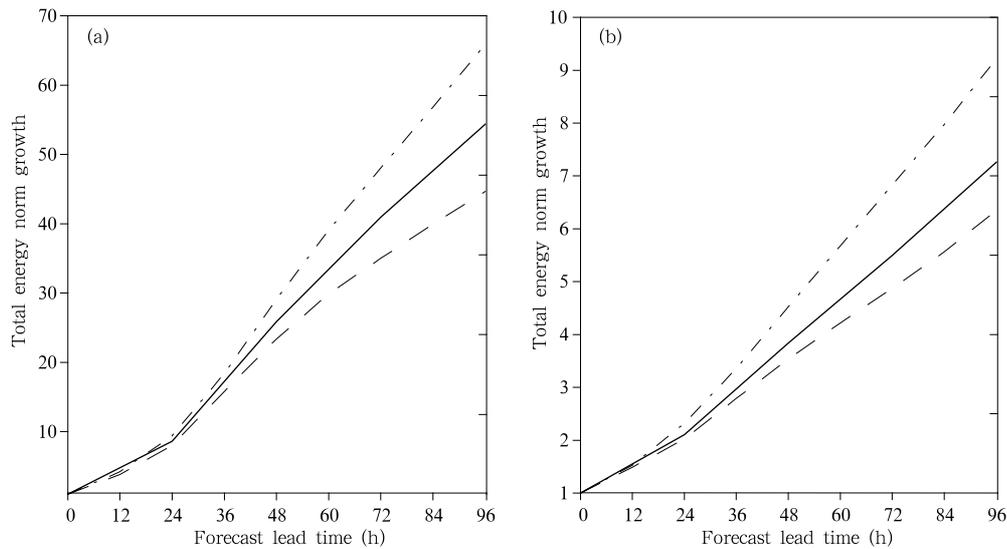


Fig. 5. The total energy norm growth in ensemble subspace as a function of lead time. (a) The fastest growth, (b) the average total energy norm growth from 14 perturbations. Solid, dashed, and dashed-dot lines denote the global, Northern and Southern Hemispheres, respectively.

of ensemble perturbations. The whole tendency of the average energy norm growth is almost identical to the fastest growth, which approximately maintained the ensemble spread in forecast lead time. The ETKF scheme can not only precisely capture the analysis error normal modes that grow rapidly and make initial perturbations increase in an appropriate way in the ensemble subspace, but also keep the ensemble spread in a suitable range. Note that the energy growth mentioned above is calculated under the assumption that forecast error is linearly increasing. Generally, the linear relation is maintained for about 2 days. Thus, the one after 48-h forecast time in Fig. 5 is only an approximation of the total energy growth.

4.4 Correlation analysis of ensemble perturbations and forecast error

Traditional statistical verification of the quality of ensemble forecast is confined to the ensemble initial scheme, the forecast model, and assimilation scheme. The perturbations and errors correlation analysis (PECA; Wei and Toth, 2003) seems to be more suitable for evaluating the performance of initial perturbations of ensemble forecast. It measures the quality of the initial perturbations through en-

semble perturbations interpreting forecast error variance and directly compares the ensemble perturbations with forecast errors reducing the effect of initial error caused by analysis quality. On average, the more correlative the ensemble perturbations and forecast error, the more realistic the ensemble perturbations will depict the state of the real atmosphere. The value of PECA is the ensemble average of the correlation coefficients of ensemble perturbations and forecast error. Here, the forecast error is the difference of the control forecast and the analysis.

For the ensemble perturbations and corresponding forecast error from 14 members, the average PECA values is computed at different times and over different areas (Fig. 6). The PECA value in the Northern Hemisphere (short dashed line) is larger than that of the global (solid line) and the Southern Hemisphere (long dashed line) in 48-h forecast time, while the PECA value in the Southern Hemisphere is much lower, especially evident before 24 h. This implies that GRAPES ensemble members generated by ETKF initial perturbation scheme have better quality in the Northern Hemisphere than in the Southern Hemisphere, and this is due to the scarcity of observations and the relative worse quality of the observations in

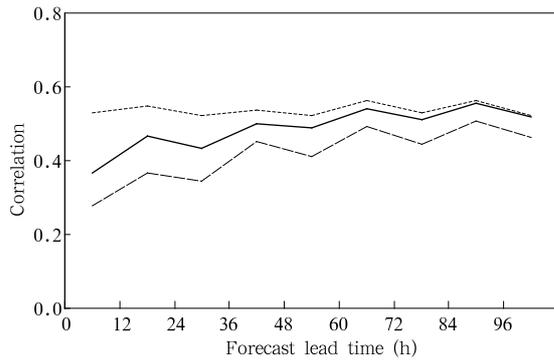


Fig. 6. The PECA values averaged over 52 days for ETKF ensembles from 14 perturbations as a function of lead time. The solid line denotes global area, and the short and long dashed lines represent the Northern and Southern Hemisphere, respectively.

the Southern Hemisphere influencing the performance of initial perturbations to some extent. In the 96-h forecast time, the values of PECA in the Northern Hemisphere have little variation, while it fluctuates greatly in the Southern Hemisphere but with a gradually increasing trend. The global PECA value starts lower due to the influence of the Southern Hemisphere. Subsequently, it increases gradually with the increase of forecast time. The above analysis shows that GRAPES global ensemble perturbations constructed by ETKF scheme can basically simulate the real state distributions of atmosphere in 96-h forecast time.

5. Quality verifications of ensemble forecast

5.1 A case study on subtropical high

In order to further verify the qualities of the ensemble forecast, we utilize simulated observations to make 96-h ensemble forecast for subtropical high at 0000 UTC 22 July 2006, with the same configurations as in the experiments. Figure 7 shows the 5880-gpm contour line with lead time of 72 h at 500 hPa for the 14 ensemble members, with time interval of 12 h. From Fig. 7, it can be seen that the ensemble mean at each forecast time, in general, is better than a single control forecast relative to real atmosphere, especially in the forecast of the westward and northward position of the subtropical high ridge. Although there is still a certain difference between the simple ensemble averages

and real atmosphere, it is some members more accurately predict the position and the moving tendency of the subtropical high ridge. The ensemble mean after 36-h forecast time tends to show greater advantages than the single control forecast. Thus, ETKF ensemble forecasts can effectively describe the probability distributions of the forecast to real atmospheric state. The initial forecast (0 h) has a reasonable ensemble spread and the ensemble mean is much closer to the real atmosphere. Furthermore, the ensemble mean of initial forecast is fairly uniform with contour line of control forecast, which indicates the efficiency of the spherical simplex centering scheme. With the extension of the leading time, the differences between ensemble members gradually tend to increase, especially in the unstable areas of weather system. However, the differences do not spread disorderly, and the growth of forecast error of ensemble members is also controlled within a reasonable scope. While the ensemble spread, which is too large by 96 h (not shown), is beyond the reasonable range of ensemble forecast, and makes the forecast meaningless. This may attribute to the quality of initial perturbations of ensemble members. The performance of forecast model and the initial values of the model are some other causes.

For the ETKF-based GRAPES global ensemble prediction system (GEPS), as shown in Fig. 7, the growth of forecast error is reasonable, which indicates that initial perturbations can capture the rapidly increasing analysis error modes, and the 14 perturbations can basically reflect the uncertainty of real atmosphere state in the forecast.

5.2 Statistical test of abnormal correlation for global height at 500 hPa

Figure 8 shows averaged abnormal correlation coefficient (ACC) of 500-hPa height over 52 days for the Northern Hemisphere. The solid line denotes averaged ensemble forecast and the dashed line denotes the control forecast. The skill of ensemble average is better than that of the single control forecast from the verifications of ACC at 500-hPa height in the 96-h forecast time, and shows increasing trend with the extension of the forecast time. Although the numerical improvement of this forecast skill is not quite apparent

(one of the main reasons for the forecast quality is that the ensemble number is comparatively less), it has preliminarily shown the advantages owned by the

ensemble forecast. Thus, this shows that the GRAPES global ensemble prediction has some practical prediction effect; the ETKF initial perturbations is able to

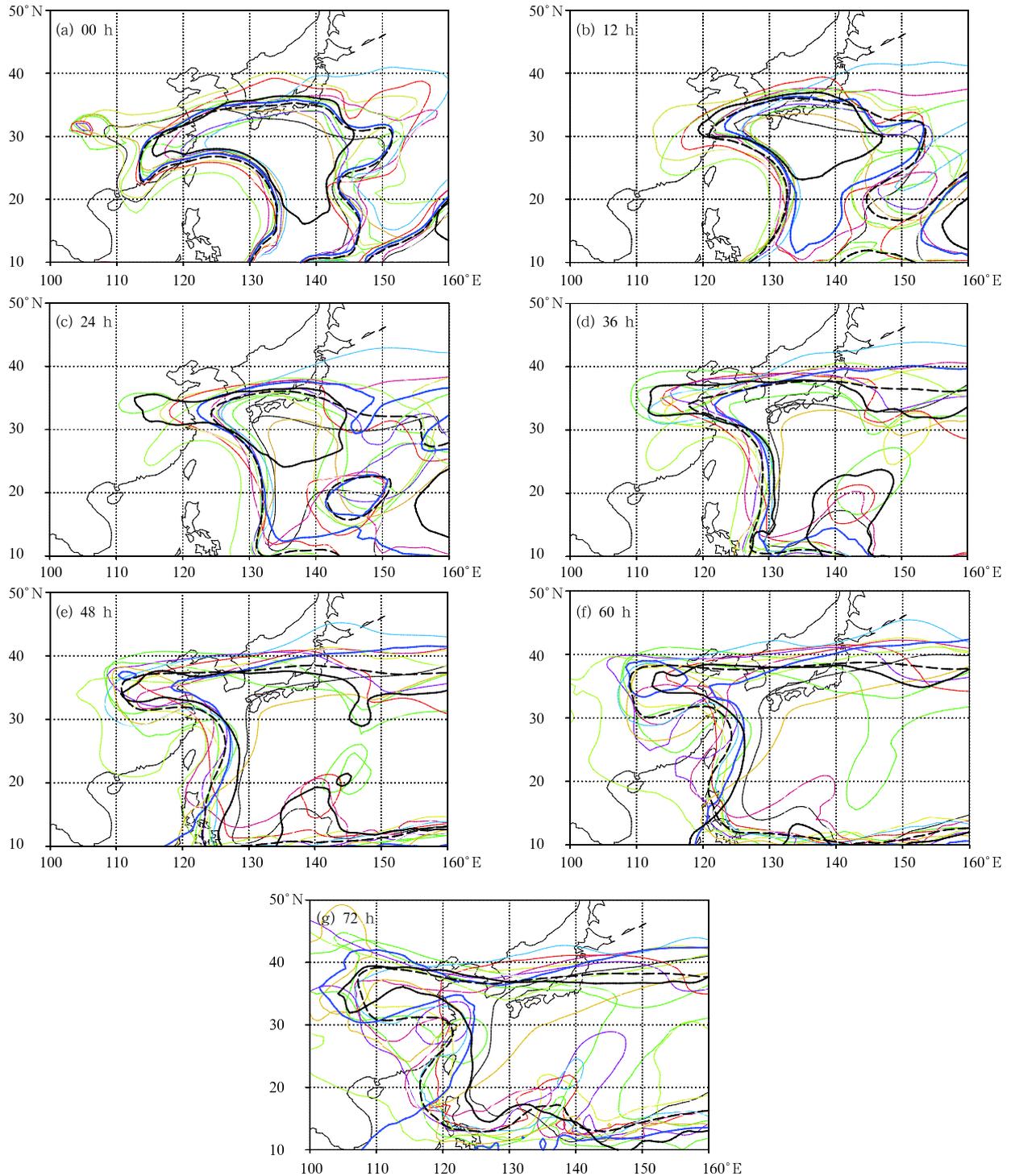


Fig. 7. 500-hPa 5880-gpm specified line for the ensemble forecast over 72-h lead time from 0000 UTC 22 July 2006. The thick-black solid lines denote the real fields, the thick-black dashed lines for the ensemble mean, and the thick-blue solid lines for the control forecast. The thin-color solid lines denote the ensemble members, respectively.

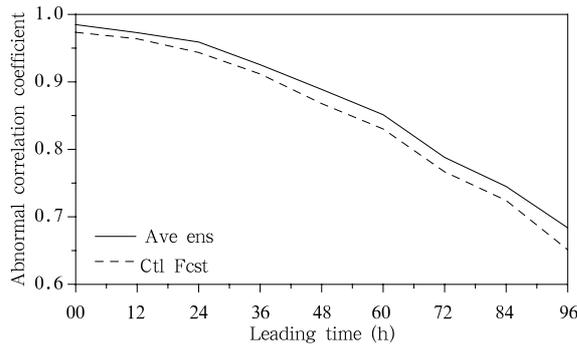


Fig. 8. Averaged abnormal correlation coefficient at 500 hPa over the 52 days for the Northern Hemisphere. The solid line denotes averaged ensemble forecast, and the dashed line denotes the control forecast.

describe the structure characteristics of analysis error at initial time and maintain a reasonable growth of forecast perturbations.

6. Summary and discussions

Based on ETKF initial perturbation method, this paper adopts the way of simulated observations, introduces the spherical simplex centering scheme that is relatively simple and effective, and the initial perturbations scheme is developed for GRAPES global ensemble prediction system. The experimental results show that the GRAPES ensemble initial perturbations are able to reflect the distributions of the main normal modes of analysis error variance in the Northern Hemisphere, and maintain the appropriate initial and forecast spread with the reasonable perturbation amplitude in 96-h forecast time. In the Southern Hemisphere, the perturbation effect is slight worse because of the scarcity of observations. As a whole, the growth of perturbation basically consistent with that of forecast error, and the ensemble variance is able to explain accurately more variance of forecast error. Thus, the ensemble forecast members can describe reasonably the developing state of the real atmosphere. The experiments carried on 14 ensemble members show that GRAPES global ensemble prediction system based on the ETKF initial perturbation scheme has development potential and application value.

In this paper, we carry out the experiments of GRAPES global ensemble prediction using relatively few ensemble perturbations. As known, the ensemble

member has direct influence on initial perturbations. Due to the limited number of ensemble member, the transformed matrix in the ensemble subspace cannot completely reflect the characteristic mode distributions of analysis error, thus restricting the performance of ETKF scheme. The use of more ensemble members should be able to fully reflect the advantages owned by ETKF initial perturbation scheme.

REFERENCES

- Bishop, C. H., and Z. Toth, 1999: Ensemble transformation and adaptive observations. *J. Atmos. Sci.*, **56**, 1748–1765.
- , B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436.
- , C. A. Reynolds, and M. K. Tippett, 2003: Optimization of the fixed global observing network in a simple model. *J. Atmos. Sci.*, **60**, 1471–1489.
- Buizza, R., and T. N. Palmer, 1995: The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.*, **52**, 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**, 1076–1097.
- Chen Jing, Xue Jishan, and Yan Hong, 2003: The uncertainty of mesoscale numerical prediction of South China heavy rain and the ensemble simulations. *Acta Meteor. Sinica*, **61**, 432–446. (in Chinese)
- and Yan Hong, 2005: A new initial perturbation method of ensemble mesoscale heavy rain prediction. *Chinese J. Atmos. Sci.*, **29**, 717–726. (in Chinese)
- Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press, 457 pp.
- Fan Xingang, 1999: A global study on ensemble prediction method. *Acta Meteor. Sinica*, **57**, 74–83. (in Chinese)
- Hamill, T. M., C. Snyder, and R. E. Morss, 2000: A comparison of probabilistic forecasts from bred, singular vector, and perturbed observation ensembles. *Mon. Wea. Rev.*, **128**, 1835–1851.
- Hoffman, R. N., and E. Kalnay, 1983: Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus*, **35A**, 100–118.

- Hollingsworth, A., 1980: An experiment in Monte Carlo forecasting procedure. ECMWF Workshop on stochastic dynamic forecasting, ECMWF.
- Houtekamer, P. L., L. Lefaivre, J. Derome, et al., 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**, 1225–1242.
- , and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.
- Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, **102**, 409–418.
- Li Zechun and Chen Dehui, 2002: The development and application of the operational ensemble prediction system at national meteorological center. *Journal of Applied Meteorological Science*, **13**(1), 1–15. (in Chinese)
- Molteni, F., R. Buizza, T. N. Palmer, et al., 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119.
- Palmer, T. N., R. Gelaro, J. Barkmeijer, et al., 1998: Singular vectors, metrics, and adaptive observations. *J. Atmos. Sci.*, **55**, 633–653.
- Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330.
- , and —, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319.
- Wang, X., and C. H. Bishop, 2003: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.*, **60**, 1140–1158.
- Wei, M., and Z. Toth, 2003: A new measure of ensemble performance: perturbations versus error correlation analysis (PECA). *Mon. Wea. Rev.*, **131**, 1549–1565.
- , —, and R. Wobus, 2006: Ensemble transform Kalman filter-based ensemble perturbations in an operational global prediction system at NCEP. *Tellus*, **58A**, 28–44.
- Yang Xuesheng, Chen Dehui, and Leng Tingbo, 2002: The comparison experiments of SV and LAF initial perturbation techniques used at the NMC ensemble prediction system. *Journal of Applied Meteorological Science*, **13**(1), 62–66. (in Chinese)
- Zhong Ke and Wang Hanjie, 2004: The physical ensemble technique of the regional climate simulation. *Acta Meteor. Sinica*, **62**, 776–781. (in Chinese)