Ensemble Transform with 3D Rescaling Initialization Method

Juhui Ma[1], Yuejian Zhu[2], Dingchen Hou[2], Xiaqiong Zhou[3] and Malaquias Peña[3]

[1] Key Laboratory of Meteorological Disaster of Ministry of Education,

Nanjing University of Information Science and Technology, Nanjing 210044, China

[2] Environmental Modeling Center/NCEP/NOAA, College Park, MD 20740, USA

[3] I.M. Systems Group, Inc. (IMSG) at the Environmental Modeling Center/NCEP/NOAA,

College Park, MD 20740, USA

Manuscript submitted to

Monthly Weather Review

November 15, 2013

Corresponding author address:

Juhui Ma

1100 Perimeter Park Dr #104

Morrisville, NC 27560

Email: juhui.ma@gmail.com

Abstract

The Ensemble Transform with Rescaling (ETR) method has been used to produce fast growing components of analysis error in the NCEP Global Ensemble Forecast System (GEFS). The rescaling mask contained in the ETR method constrains the amplitude of perturbations to reflect regional changes of analysis error. However, due to a lack of suitable three-dimensional (3D) analysis error estimation, in the operational GEFS the mask is based on an estimated analysis error at 500hPa and is not flow-dependent but changes monthly. With the availability of an ensemble-based data assimilation system at NCEP, a 3D mask can be computed. This study generates initial perturbations by the Ensemble Transform with 3D Rescaling (ET_3DR) and compares the performance with the ETR. Meanwhile, the ET_3DR is also applied into the Ensemble Kalman Filter (EnKF) method (hereinafter referred to as EnKF_3DR).

Results from a set of experiments indicate that the 3D mask affects the amplitude of initial perturbations. Relative to the ETR, the large amplitudes of the ET_3DR initial perturbations at 500hPa connect better with baroclinic instability areas over the extra-tropics and deep convection areas over the tropics. Furthermore, the maxima of vertical distribution for the ET_3DR initial perturbations correspond to the subtropical jet region and tropical easterlies jet region. The better distribution of the perturbations is found to produce faster spread growths. Results with EnKF_3DR also show benefits. The variance along orthogonal basis vectors in the EnKF_3DR is maintained more than in the EnKF. Furthermore, it is found that the EnKF_3DR outperforms the EnKF.

1. Introduction

Ensemble generation methods seek to create a set of initial perturbations representative of analysis errors in a numerical weather prediction system with the goal to improve its probabilitic forecast performance. The analysis errors can be decomposed into nongrowing and growing (Toth and Kalnay, 1997). The nongrowing errors have large dimensional subspace, which cannot be sampled with a limited number of ensemble members and these errors will typically lose their amplitude rapidly. The growing errors amplify fast and dominate the short-range forecast error growth. Therefore, the success of an ensemble generation method lies on how well its perturbations sample the growing errors in the analysis.

The Breeding Vector (BV) method (Toth and Kalnay, 1993, 1997) creates perturbations that grow fast by inserting ("breeding") rescaled errors from previous cycles. After several cycles, the growing component amplifies, and the nongrowing component is eliminated. However, the BV method alone is insufficient to systematically capture all initial uncertainties (Annan, 2004; Buizza et al., 2005). Therefore, an improved version of the BV, the Ensemble Transform (ET) method is introduced to generate initial perturbations which are globally transformed from the forecast perturbations (Bishop and Toth, 1999; Wei et al., 2008). As the BV method, the ET also generates a flow-dependent spatial structure and is able to represent fast growing component of analysis errors with minimal computer expense. The advantages of the ET are that the perturbations have the maximum number of effective degrees of freedom and are more consistent with the data assimilation system due to their orthogonalization in the inverse analysis error variance norm (Wei et al., 2008); more importantly ET outperforms BV in standard probabilistic skill scores.

The analysis error variance decides the analysis perturbation globally during the transformation, but the initial spread distribution can be regionally inconsistent with the analysis error variance due to the limited ensemble size compared with the state dimension (McLay et al., 2008; Wei et al., 2008). McLay et al. (2010) performed the local ET with partitioning the global domain into latitude bands or latitude-longitude blocks, resulting in a better agreement with analysis error variance and improved ensemble performance. At NCEP, a simple remedy that regional rescaling process is imposed into the ET initialization periodically to make the amplitude of initial perturbations vary in accordance with regional changes of analysis uncertainties. This regional rescaling improved both the spread distribution of initial perturbations and most probabilistic scores with respect to the ET without rescaling (Wei et al. 2008). The regional rescaling factor is designed as the ratio of the mask and the square root of a special norm of analysis perturbations at each grid point. The choice of mask is the key to the regional rescaling. In the NCEP operation, the mask is calculated using a long term averaged root-mean-square of analysis error variance in the kinetic energy norm at the 500hPa level obtained from variational data assimilation system (Szunyogh and Toth, 2002; Wei et al., 2008). But the current mask is not adequate enough for using in the context of ensemble forecast system to represent the analysis uncertainties. First and foremost, the two-dimensional (2D) mask cannot represent the vertical structure of analysis uncertainties. To compensate for the underestimate of analysis error, extra inflations with empirical factor have to be applied to the mask for levels below 500hPa at NCEP. But it is obviously not optimal for regional rescaling. Second, the mask was computed from a past decade climatological data, during which the density and accuracy of observation, as well as the data assimilation technology all have greatly changed. Thus, there is a need to update the mask with the estimation of analysis errors from the current real time data assimilation system to make the initial perturbations more consistent with the observations and data assimilation system. As found in the study of Wei et al. (2008), compared with the ET method, the ETR failed to show high spread in the southern-ocean storm track area due to the mask, which indicated a more accurate time-dependent mask was necessary. Third, the total energy norm may be more reasonable to measure the magnitude of initial perturbations than the kinetic energy norm. Palmer et al. (1998) found that the total energy is more consistent with analysis error statistics than the streamfunction, enstrophy or kinetic energy metric. Some previous studies also considered parts of these problems above and designed different masks. Wang and Bishop (2003) chose the square root of the seasonally and vertically averaged initial ensemble wind variance from the Ensemble Transform Kalman Filter (ETKF) ensemble as the mask applied in the BV method. Magnusson et al. (2008) designed vertically integrated estimation of analysis errors using total energy norm from four-dimensional variational (4D-Var) assimilation system as the mask. In this paper, a new mask will be defined by 3D analysis uncertainty measured in total energy norm obtained from the 80-member ensemble analysis generated by the NCEP's hybrid 3D-Var/EnKF system (Wang et al., 2013). The sensitivity of ETR perturbations and forecast skill to the mask in the NCEP GEFS will be explored.

Relative to the variational data assimilation method with static background error, the ensemble-based data assimilation has the ability to provide flow-dependent estimates of the background error. Moreover, the ensemble-based data assimilation generates a set of initial analysis to initialize the ensemble of predictions in the next cycle and also to provide an estimate of the analysis error, which unifies the ensemble forecast and data assimilation steps. Consequently, many Numerical Weather Prediction (NWP) centers are adopting the use of ensemble technology to produce analysis in the data assimilation system and initial conditions in the ensemble prediction system simultaneously, such as the Meteorological Service of Canada (MSC) and the European Centre for Medium-Range Weather Forecasts (ECMWF). Researches demonstrate that the ensemble-based data assimilation is beneficial for both systems by producing flow-dependent estimates of analysis uncertainty and background error uncertainty (Buehner et al., 2010a, b; Buizza et al., 2008, 2010). A hybrid 3D-Var/EnKF data assimilation system became operational on 22 May 2012 at the NCEP (Wang et al., 2013). In this system, the background error is created by a combination of static background error from the 3D-Var and flow-dependent background error produced from the EnKF, and the EnKF perturbations are recentered on the hybrid analysis. The hybrid 3D-Var/EnKF provided better analyses and subsequent forecasts than the previous operational 3D-Var (at http://www.emc.ncep.noaa.gov/GFS/impl.php).

Since the ETR method is able to maximize the effective degrees of perturbation freedom without extra cost of computer resources, applying the ETR on other ensemble analysis (e.g. multi-center analysis, or analysis from ensemble-based data assimilation) may have a positive impact on the quality of initial conditions. The availability of the EnKF in the NCEP Global Data Assimilation System (GDAS) provides alternative ensemble initial conditions for the operational GEFS. The performance of the EnKF and ETR perturbations in the NCEP operational environment is compared and will be presented in another paper. In this study, the EnKF ensemble analysis will be transformed and rescaled by ET_3DR, and the impact will be explored.

In the next section, the methodology of ET_3DR and EnKF_3DR are described. Section 3 investigates the horizontal and vertical distributions of perturbations generated by the ETR and ET_3DR, and compares their forecast performances. In section 4, the characteristic of initial perturbations generated by the EnKF and EnKF_3DR is analyzed, and their forecast skills are compared. The conclusions are summarized in section 5.

2. Initialization methodologies and experimental design

2.1 Initialization methodologies

a. The ET_3DR method

In the ETR scheme (Wei et al., 2006; Wei et al., 2008), the analysis perturbations matrix $\mathbf{X}^a$ are generated from the forecast perturbations matrix $\mathbf{X}^f$ through an ensemble transformation matrix $\mathbf{T}$ as follows

$$\mathbf{X}^a = \mathbf{X}^f \mathbf{T}, \tag{1}$$

where n analysis perturbations $\mathbf{x}_i'^a$ (i=1, 2, …, n) are listed as columns in the matrix $\mathbf{X}^a$, and n forecast perturbations $\mathbf{x}_i'^f$ (i=1, 2, …, n) are listed as columns in the matrix $\mathbf{X}^f$. After the transformation, all perturbations are orthogonal. As shown in Wei et al. (2008) and Ma et al. (2012), the transformation matrix $\mathbf{T}$ is given by

$$\mathbf{T}=\mathbf{C}\mathbf{F}^{-1/2}, \tag{2}$$

where columns of the matrix $\mathbf{C}$ contain the orthonormal eigenvectors (c_i, i=1, 2, …, n) of the matrix $\frac{1}{n-1}(\mathbf{X}^f)^T(\mathbf{P}^a)^{-1}\mathbf{X}^f$, and the diagonal matrix $\mathbf{\Gamma}$ contains the corresponding eigenvalues (λ_i, i=1, 2, …, n), in which the first n-1 eigenvalues are non-zero and the last eigenvalue is zero. A diagonal matrix $\mathbf{F}$ is defined by setting the zero eigenvalue in $\mathbf{\Gamma}$ to a non-zero constant α. The diagonal matrix $\mathbf{P}^a$ contains the analysis error variances. Then, a simplex transformation is performed to mask analysis perturbations be centered on the analysis, but perturbations become quasi-orthogonal at this step.

To make the amplitude of initial perturbations vary in accordance with regional changes of analysis uncertainties, $\mathbf{X}^a$ is rescaled using a rescaling factor γ which is designed as

$$\gamma = \begin{cases} \dfrac{mask}{pertb}, & if\ mask < pertb \\ 1, & if\ mask \geq pertb \end{cases}. \tag{3}$$

Here, *mask* denotes a long term averaged root-mean-square of analysis error variance; *pertb* is the square root of a special norm from $\mathbf{X}^a$ at each grid point. If the ratio is larger than 1.0, the rescaling factor will be set to 1.0, which means the perturbations can grow freely; otherwise the amplitude will be rescaled to the size of the mask.

The mask used in the current NCEP operational GEFS is a 2D mask, which is computed from a long term averaged root-mean-square of analysis error variance in the kinetic energy norm at the 500hPa level obtained from variational data assimilation system (Szunyogh and Toth, 2002; Wei et al., 2008).

As discussed in section 1, an ensemble of analyses can be obtained from the NCEP operational hybrid 3D-Var/EnKF data assimilation system directly, which can provide a flow-dependent estimate of analysis error. In this study, the 3D mask is defined by the root-mean-square of the deviation total energy norm *TE* computed from 80-member EnKF analysis

$$TE = \frac{1}{80}\sum_{i=1}^{80}\sqrt{\frac{1}{2}(u_i'^2 + v_i'^2 + \kappa T_i'^2)}, \tag{4}$$

where u_i', v_i' and T_i' (i=1, 2, …, 80) are the deviation of the ith EnKF member from EnKF mean analysis for the wind components and temperature. $\kappa = \frac{c_p}{T_r}$ equals approximately $4.0JKg^{-1}K^{-2}$ in which c_p is the special heat at constant pressure and T_r is the reference temperature. For the purpose of representative of the typical large-scale components and considering the most recent behavior of analysis uncertainties meanwhile, the decaying average method (Cui et al., 2012) is employed to accumulate the mask, given by

$$TE_{ave}(t) = (1-w)TE_{ave}(t-1) + wTE(t)\,. \tag{5}$$

Here, the averaged mask $TE_{ave}(t)$ is updated by the prior period averaged mask $TE_{ave}(t-1)$ and the most recent $TE(t)$ with the weight coefficient w ($w = 2\%$ in this study). To preserve most of the dynamical balance in the perturbations, the mask is smoothed horizontally with a spectral filter.

Figure 1a, b show the horizontal distribution of the 2D and 3D mask at 500hPa over the period 1 September - 30 November 2012. The 2D mask obtained from static analysis error estimation, has large amplitude over the poorly observed oceans and small amplitude over the data-rich continents. Over the mid-latitudes, the 3D mask estimated from the EnKF data assimilation system has relatively small amplitude over the continents compared to the oceans, but this property is not quite obvious like the 2D mask. That is because the EnKF ensemble analysis used to produce the 3D mask is processed with a multiplicative inflation algorithm to account for unrepresented error sources during the generation (Whitaker and Hamill, 2012). The values of inflation are proportional to the amount observations reduce the ensemble spread, which are large in regions of dense observations. The analysis error should be not only associated with the observation network but also the distribution of the atmospheric instability (Hamill et al., 2003). The 3D mask is more flow-dependent relative to the 2D mask. For example, over the northern and southern extra-tropics, for the 3D mask we see that areas over the maximum amplitude are respective around 60˚N and 60˚S corresponding to main regions of baroclinic energy conversions, but the maximum areas are over the poles in the 2D mask. That may solve the problem that the old rescaling factor cannot reduce the amplitudes enough at the higher latitudes (Toth and

Kalnay, 1997). Another striking difference is located in the tropics, which will be discussed in the next section. Figure 2 shows the vertical profile of the 2D and 3D mask over the same period as Fig. 1. In the 3D mask, the amplitude increases with the altitude and decreases after it reaches the maxima between the 300hPa and 100hPa. The vertical structure cannot be represented with the 2D mask.

b. The EnKF_3DR method

As illustrated in the Fig. 3, the following steps are performed to initialize the ensemble with the EnKF_3DR method. Firstly, use the EnKF method (Whitaker and Hamill, 2002) to generate ensemble analysis. In this study, the 80 EnKF analyses are directly obtained from the NCEP hybrid 3D-Var/EnKF data assimilation system. Secondly, compute the ensemble mean analysis and the 80 ensemble deviations from the ensemble mean analysis. The root-mean-square of the deviation total energy norm is calculated by Eq. (4). Finally, apply ET_3DR method onto the 80 EnKF perturbations to generate 80 EnKF_3DR perturbations. The rescaling mask used in the ET_3DR method is the estimate of analysis error computed at the second step.

2.2 Experimental design

Four sets of ensemble generation experiments (ETR, ET_3DR, EnKF and EnKF_3DR) are performed using the NCEP Global Forecast System (GFS) model with a T254 horizontal resolution, 42 sigma-p hybrid vertical levels. The analysis is truncated from the T574L64 analysis provided by the NCEP GDAS. The initial perturbations for the ETR and EnKF experiments are obtained from the operational GEFS and hybrid 3D-Var/EnKF data assimilation system respectively. The methods used to generate the ET_3DR and EnKF_3DR ensemble initial perturbations are described in subsection 2.1. The perturbations are updated every 6-hour for 80-member and only 20-member is chosen for long forecasts due to limited computational resources.

The ET_3DR initial perturbation cycles are performed from 1 September to 30 November 2012 and the first 10 days are used for the system to spin-up. The 8-day long forecasts of the four sets experiments are produced once per day (00 UTC) between 11 September and 30 November 2012 (81 cases). To represent model error, all experiments use the Stochastic Total Tendency Perturbation (STTP) (Hou et al., 2006, 2008) as in the NCEP operational GEFS. Verification results are presented for 500hPa geopotential height (Z500); 850hPa temperature (T850); and 250hPa, 850hPa, 10m *u*-components of wind (U250, U850, and U10m) over the extra-tropics of the Northern Hemisphere (NH, 20˚-80˚N), the extra-tropics of the Southern Hemisphere (SH, 20˚-80˚S) and the Tropics (TR, 20˚S-20˚N).

3. ETR versus ET_3DR

3.1 Initial perturbation distribution

Figure 4 shows the vertical profile of the square root of total energy of perturbations at different lead times for the ETR and ET_3DR experiments. Over the NH, the ETR has larger initial amplitude compared to the ET_3DR at the lower levels and the other maximum of initial perturbations is at 250hPa, which is slightly smaller than the ET_3DR (the left panel of Fig. 4a). The left panel of Figs. 4b, c, d shows that the ET_3DR grows faster than the ETR. After 12-h, the amplitude of ET_3DR perturbations gets closer to the ETR below 700hPa, and the difference becomes larger than that at the initial time above 700hPa. After 48-h, the perturbations of the ET_3DR are larger than the ETR for all levels. Over the SH (the middle panel of Figs. 4a, b, c, d), the situation exhibits similarity with that over the NH. The growth rate over the TR (the right panel of Figs. 4a, b, c, d) is lower than that over the NH and SH. At the upper levels, the maximum of the ETR (ET_3DR) initial perturbations is at 200hPa (100hPa) and their growth rates are comparable. At the lower levels, to compensate for the slow growth of the ETR perturbations, the amplitude of its initial perturbations is much larger than that for the ET_3DR, but the amplitude of perturbations for the ETR is still caught up by the ET_3DR after 96-h. The fast growth of the ET_3DR perturbations compared to the ETR shown in Fig. 4 may be due to its structure of initial perturbations which could better sample the analysis error with applying the 3D regional rescaling mask. To further illustrate the details of the initial perturbations, the horizontal and vertical distributions will be analyzed below.

Figure 5 shows the horizontal distribution of the square root of total energy of initial perturbations on the 500hPa level for the two experiments. It is found that the regional rescaling masks applied for the ETR and ET_3DR experiments have great impact on their initial perturbations. For the ETR (Fig. 5a), the maxima are around the poles. In addition, the dominant feature of initial perturbations is the locations of the maxima and minima coinciding with the distributions of oceans and continents. The maxima are over the North Pacific, Atlantic and Indian Oceans, and the minima are located in the North America, Eurasia and Australia. For the ET_3DR (Fig. 5b), the large initial perturbations are over the meridional bands around 60°N, 60°S and the equator, which seem to be related to the baroclinic zones and tropical convection zones respectively. The flow-dependent ET_3DR initial perturbations will be beneficial for obtaining a sufficient dispersed ensemble in the medium range.

To investigate the connection between the initial perturbations and baroclinic instability, the correlation coefficients between the Eady index and the square root of total energy of initial perturbations over the NH and SH for both experiments are shown in Fig. 6 in which the shades indicate that the correlation is statistically significant at the 95% confidence interval. The Eady index which is a simple measure of the most unstable Eady mode, is defined as (Hoskins and Valds, 1990)

$$\sigma_E = 0.31 \frac{f}{N} \frac{du}{dz}, \qquad (6)$$

where f is the Coriolis parameter, N is the static stability, u is the magnitude of the vector wind. Here, $N = (g \frac{d \ln \theta}{dz})^{\frac{1}{2}}$ and $\frac{du}{dz}$ are computed using the 300hPa and 1000hPa potential temperature and wind from the NCEP Final Analyses. Over the NH (Figs. 6a, b), the areas of initial perturbations for the ET_3DR which are statistically significantly correlated with the Eady index are larger than the ones for the ETR. Especially, the correlation coefficients are higher than 0.6 even up to 0.8 over the western part of the Pacific Ocean and the Atlantic Ocean in the ET_3DR experiment. Over the SH (Figs. 6c, d), the correlations in the two experiments are low compared to those over the NH.

Over the TR, the deep convection has an important role on the development of perturbations through the release of latent heating. To illustrate the relationship between the initial perturbations and deep convection, the outgoing longwave radiation (OLR) which is a common measure of the intensity of the tropical convection is plotted in Fig. 7. The low value of OLR represents intense tropical convection. For the ET_3DR (Figs. 5b, 7), the locations of the maxima of initial perturbations accurately coincide with the intense deep convection zones (low OLR) except the maximum over the eastern Pacific Ocean. This connection cannot be detected at all in the ETR experiment (Figs. 5a, 7).

Figure 8 shows the zonal average of the square root of total energy of initial perturbations. For the ETR (Fig. 8a), below the 200hPa level, the minima of initial perturbations on both hemispheres are around 60°N and 40°S, respectively. Over the TR, there are two maxima at 10°N, 300hPa and 950hPa. Above the 200hPa level, the perturbations decrease with the height over the globe. For the ET_3DR (Fig. 8b), the amplitude is slightly larger over the SH than over the NH. The maxima are around 55°N and 55°S at 300hPa, which correspond with the subtropical jet regions. Over the TR, the maximum is around 10°N, 100hPa, near the tropical easterlies jet region.

3.2 Ensemble forecast skill

The verification methods used to evaluate the ensemble forecast skills with ETR and ET_3DR initial perturbations include Root Mean Square Error (RMSE; Toth et al., 2003) of ensemble mean and Continuous Ranked Probability Score (CRPS; Toth et al., 2003; Wilks, 2006). The paired block bootstrap algorithm (Hamill, 1999) is used to estimate the statistical significance of differences in scores. In this study, 95% confidence interval is computed from a bootstrap resampling using 1000 random samples of the 81 cases.

a. RMSE and ensemble spread

Figures 9a, b, c, d show the ensemble mean RMSE and ensemble spread for U250, Z500, U850, and T850 over the NH. Comparing the RMSE of the ETR and ET_3DR experiments, the results have no significant differences for all lead times. Regarding the ensemble spread, there are substantial differences between the two experiments. For U250, the ET_3DR and ETR have the same size of initial perturbations, but the ET_3DR grows faster than the ETR and keeps being consistent with the RMSE for all lead times as a perfect ensemble forecast system should do (Fig. 9a). For the indirect model variable Z500, figure 9b shows that the ET_3DR starts from a larger spread and overestimates the ensemble mean errors, but the amplitude of initial perturbations could be tuned further to give a similar spread to the errors at the initial time. For U850 and T850, the ET_3DR initial perturbations are much smaller than the ETR, but the spread catches up after 24-h and gets close to the RMSE gradually (Figs. 9c, d). Over the SH (Figs. 10a, b, c, d), the results are similar to the ones over the NH.

Figures 11a, b show the RMSE and spread for U850 and U10m since the wind field is of more interest than the mass field over the TR. The ETR starts from a much larger spread than the ET_3DR and decays during the first 2-d. The spread of ET_3DR has higher growth rate than the ETR especially for the first 2-d. As found in section 3.1, that is attributed to the close connection between the initial perturbations of the ET_3DR over the TR and the tropical deep convection which is an important factor for the development of perturbations. The ETR has significantly higher RMSE than the ET_3DR at 12-h lead time. Both experiments produce smaller spread than the RMSE. Because the growth in the ensemble spread over the TR is mostly determined by the physical processes, while that over the NH and SH is mainly influenced by the dynamic instability, sampling the model related errors plays a more important role on the ensemble spread over the TR.

Overall, the main advantage of the ET_3DR as the figures shown is the higher growth rate unlike the ETR method which artificially increases the amplitude of the initial perturbations to compensate for the low growth rate of spread with the cost of negative effects on the performance at short lead times. This advantage is especially obvious at lower levels and over the TR.

b. Continuous Ranked Probability Score (CRPS)

The CRPS is used to measure the reliability and resolution of ensemble-based probabilistic forecasts by calculating the distance between the predicted and the observed cumulative distribution functions of scalar variables. The smaller the score is, the better the quality of the probabilistic forecast is. Over the NH, the CRPS for U250 is similar for the two experiments (Fig. 12a). The ETR has significantly smaller score than the ET_3DR for the first 12-h for Z500 (Fig. 12b). There are more improvements on the probabilistic forecast score for lower levels compared to upper levels using the ET_3DR initial perturbations. For U850, the ET_3DR produces statistically significantly better probabilistic forecast for the first 4-d than the ETR (Fig. 12c). For T850, the ET_3DR has slightly but statistically significantly better performance for the first 2-d than the ETR (Fig. 12d). Over the SH (Figs. 13a, b, c, d), the results are generally similar to that over the NH, except that the ET_3DR presents statistically significantly smaller value than the ETR only for lead times up to 12-h for U850 and T850. Over the TR (Fig. 14), the ET_3DR has statistically significantly better performance than the ETR for almost all 8-d lead times except day 1.5-2.5 for U850. For U10m, the ET_3DR has statistically significant advantage over the ETR for the first 1-d.

4 EnKF versus EnKF_3DR

4.1 Initial perturbation

Initial perturbations should span as many unstable directions of the atmosphere as possible with limited ensemble members. The eigenvalue spectra of the covariance matrix of the initial perturbations can be used to evaluate the distribution of the amounts in independent directions. Figure 15 shows the mean eigenvalue spectra for Z500 during the period 11 September - 30 November 2012 over the globe. It is found that the initial perturbations of the EnKF are too much contained in the direction of the first mode. The EnKF_3DR has flatter spectra than the EnKF, implying that the ensemble members are more independent than the EnKF, which may have potentially positive impact on the ensemble performance.

4.2 Ensemble forecast skill

The results of the EnKF and EnKF_3DR experiments will be compared in this section using the same verification methods as in section 3.2.

a. RMSE and ensemble spread

In Fig. 16, the RMSE and ensemble spread for U250, Z500, U850 and T850 over the NH are shown. Comparing the RMSE, the EnKF_3DR is slightly better than the EnKF for U250 and Z500 (Figs. 16a, b), but the difference is not statistically significant for U250 and only significant for the first 1-d for Z500. Results for U850 and T850 (Figs. 16c, d) show that the EnKF_3DR has significantly smaller RMSE than the EnKF for the first 3.5-d. Regarding the ensemble spread, the growth rates are basically similar for the two experiments. For U250 and Z500, the initial spread for the EnKF_3DR is slightly smaller than the EnKF, and the spread for the EnKF_3DR is more consistent with the RMSE compared to the EnKF until 4-d for U250 and all lead times for Z500 (Figs. 16a, b). For U850 and T850, the spread grows somewhat slower than the RMSE for short lead times, but becomes almost equal to the RMSE with the increasing of forecast length (Figs. 16c, d).

Over the SH, the difference of the RMSE between the two experiments for U250 is also not significant for all lead times (Fig. 17a). The forecast length, during which the EnKF_3DR has significantly smaller RMSE than the EnKF, extends to 6.5-d for Z500 (Fig. 17b). For U850 and T850, the EnKF_3DR produces smaller RMSE than the EnKF, and the difference is statistically significant for all lead times (Figs. 17c, d). The spread for U250 and Z500 in the EnKF_3DR experiment is more consistent with the RMSE compared to the one in the EnKF experiment which is much larger than the RMSE (Figs. 17a, b). Similar to the one over the NH, the spread for U850 and T850 grows slower than the RMSE during the first 3 to 4-d, and then becomes almost equal to the RMSE with the increasing of forecast length.

Results for the TR (Fig. 18) show that U850 and U10m for both experiments appear to produce much less spread than the RMSE due to the under-sampling of model related errors. The spread for the EnKF_3DR grows slightly slower than the EnKF, but the RMSE is substantially smaller than the EnKF.

b. CRPS

The CRPS for U250 shows similar scores between the two experiments for both hemispheres and only differs significantly for the first 12-h over the SH (Figs. 19a and 20a). For Z500, the EnKF_3DR produces slightly better probabilistic forecast than the EnKF over the NH, and the difference is significant for up to 2-d (Fig. 19b). Over the SH, the improvement becomes more apparent that is significant for 6.5-d (Fig. 20b). For U850 and T850, the EnKF_3DR has substantially better performance than the EnKF for both hemispheres. The difference is statistically significant until 4-d over the NH, and all lead times over the SH (Figs. 19c and 20c). Over the TR, for all lead times, the EnKF_3DR shows significantly better score for both U850 and U10m (Fig. 21).

5 Conclusions and discussion

In the ETR method, the rescaling mask plays a critical role to constrain the amplitude of initial perturbations to reflect regional changes of analysis error. While the ETR used in the NCEP GEFS has proved to improve the spread and probabilistic skill of the ensemble forecasts over both BV and ET methods, its mask has several limitations, which in this study we attempt to address. There are three main modifications to the mask. First and foremost, for representing the vertical structure of analysis error, the 3D mask is employed instead of the 2D mask. This is the most advantage of the ET_3DR compared to the ETR. In the ETR method, due to the vertical-constant mask used, extra inflations have to be applied to the mask for levels from model bottom to 500hPa with empirical factors to compensate for the underestimate of analysis errors. Second, with the availability of an ensemble of analyses from the hybrid 3DVar-EnKF data assimilation system, on each data assimilation cycle a flow-dependent error variance is computed with real observations, which is associated with both the dynamics of the day and the observation density distribution. This new analysis error variance replaces the static analysis error variance. Third, the kinetic energy norm is changed into the total energy norm to measure the magnitude of initial perturbations. Results with the ETR and ET_3DR experiments performed from 11 September to 30 November 2012 using the NCEP GFS indicate that these updates have direct impact on the perturbations. The horizontal distribution of the ETR initial perturbations at 500hPa coincides with the distribution of oceans and continents, but is not consistent with the flow. Because of the flow-dependent mask applied in the ET_3DR, the large amplitudes of the initial perturbations connect better with the areas of baroclinic instability over the NH and the areas of deep convection over the TR, which is beneficial for obtaining a sufficient dispersed ensemble in the medium range. The difference of vertical distribution for the ETR perturbations is small due to the vertical-constant mask, while the maxima of vertical distribution for the ET_3DR perturbations correspond to the subtropical jet region and tropical easterlies jet region. Since the amplitude of the initial perturbations for the ET_3DR is better consistent with the distribution of the atmospheric instability regions, the spread grows much faster than the ETR, especially at the lower levels and over the TR. Consequently, the choice of mask is important to perturbation growth and ensemble performance for the NCEP GEFS.

Since the ETR method is able to maximize the effective degrees of perturbation freedom without extra cost of computer resources, with the availability of the EnKF analyses in the NCEP GDAS, the EnKF_3DR method is designed in this study by applying ET_3DR on EnKF ensemble analysis. The eigenvalue spectra of the covariance matrix of the initial perturbations show that the ensemble members of the EnKF_3DR are more independent than the EnKF. By evaluating the ensemble performance, it is found that the EnKF_3DR is substantially better than the EnKF, especially at the lower levels and over the TR.

The EnKF may be considered as the potential candidate of the NCEP operational GEFS initial perturbation method in the next implementation. However, from the results of this study, we can find that applying ET_3DR on EnKF is more beneficial for the improvement of the ensemble forecast performance than the EnKF. Further studies will explore the results using this strategy for other seasons. Furthermore, due to the merit of the ETR method, it may be also considered to apply on other ensemble analysis, such as multi-center analysis.

Although the results of this study indicate that the improvement of the mask benefits the ensemble performance, this regional rescaling is only a simplified remedy to the complex problem of making initial spread distribution agree with the analysis error variance regionally. Therefore, future research effort should be on practical accounting for all sources of analysis uncertainties.

Acknowledgments. The authors thank members of Ensemble and Post Processing Team at EMC/NCEP for helpful suggestions during the course of this work. Thanks to Daryl Kleist for providing data and code to read the EnKF ensemble analysis. First author gratefully acknowledges the support of EMC.

References

Annan, J. D., 2004: On the orthogonality of bred vectors. Mon. Wea. Rev., **132**, 843–849.

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

Buehner, M., P. L. Houtekamer, C. Charette, H. L. Mitchell, and B. He, 2010a: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP.

Part I: Description and single-observation experiments. *Mon. Wea. Rev.*, **138**, 1550–1566.

——, ——, ——, ——, and ——, 2010b: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: One-month experiments with real observations. *Mon. Wea. Rev.*, 1**38**, 1567–158.

Buizza, R., P. L. Houtekamer, Z. Toth, G. Pellerin, M. Wei, and Y. Zhu, 2005: A Comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems. *Mon. Wea. Rev.*, **133**, 1076–1097.

——, M. Leutbecher, and L. Isaksen, 2008: Potential use of an ensemble of analyses in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, **134**, 2051–2066.

——, ——, ——, and J. Haseler, 2010: Combined use of EDA- and SV-based perturbations in the EPS. *Newsletter* n. 123, ECMWF, Shinfield Park, Reading RG2-9AX, UK, 22–28.

Cui, B., Z. Toth, Y. Zhu, and D. Hou, 2012: Bias Correction for Global Ensemble Forecast. *Wea. Forecasting*, **27**, 396–410.

Hamill, T. M., 1999: Hypothesis tests for evaluation numerical precipitation forecasts. *Wea. Forecasting*, **14**, 155–167.

——, C. Snyder, and J. S. Whitaker, 2003: Ensemble forecasts and the properties of flow-dependent analysis-Error covariance singular vectors. *Mon. Wea. Rev.*, **131**, 1741–1758.

Hoskins, B. J., and P. J. Valdes, 1990: On the existence of storm-tracks. *J. Atmos. Sci.*, **47**, 1854–1864.

Hou, D., Z. Toth, and Y. Zhu, 2006: A stochastic parameterization scheme within NCEP global ensemble forecast system. In Proceedings of the *18th AMS Conference on Probability and Statistics*, 29 January –2 February 2006, Atlanta, Georgia.

———, ———, ———, and W. Yang, 2008: Impact of a stochastic perturbation scheme on NCEP global ensemble forecast system. In Proceedings of the *19th AMS Conference on Probability and Statistics*, 21–24 January 2008, New Orleans, Louisiana.

Houtekamer, P. L., and H. L. Mitchell, 2005: Ensemble Kalman filtering. *Quart. J. Roy. Meteor. Soc.*, **131**, 3269–3289.

Ma, J., Y. Zhu, R. Wobus, and P. Wang, 2012: An effective configuration of ensemble size and horizontal resolution for NCEP GEFS. *Adv. Atmos. Sci.*, **29**, 782–794.

Magnusson L., M. Leutbecher, E. Kallen, 2008: Comparison between singular vectors and breeding vectors as initial perturbations for the ECMWF Ensemble Prediction System. *Mon. Wea. Rev.*, **136**, 4092–4104.

McLay, J. G., C. H. Bishop, C. A. Reynolds, 2008: Evaluation of the Ensemble Transform Analysis Perturbation Scheme at NRL. *Mon. Wea. Rev.*, **136**, 1093–1108.

———, ———, and ———, 2010: A Local Formulation of the Ensemble Transform (ET) Analysis Perturbation Scheme. *Wea. Forecasting*, **25**, 985–993.

Palmer, T. N., R. Gelaro, J. Barkmeijer, and R. Buizza, 1998: Singular vectors, metrics, and adaptive observations. *J. Atmos. Sci*., **55**, 633–653.

Szunyogh, I., and Z. Toth, 2002: The effect of increased horizontal resolution on the NCEP global ensemble mean forecasts. *Mon. Wea. Rev.*, **130**, 1125–1143.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: the generation of perturbations. *Bull. Amer. Meteror. Soc.*, **174**, 2317–2330.

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

——, O. Talagrand, G. Candille, and Y. Zhu, 2003: Probability and Ensemble Forecasts. *Forecast Verification: A practitioner's Guide in Atmospheric Science*, Jolliffe and Stephenson, Wiley, 137–163.

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.

——, D. Parrish, D. Kleist, and J. Whitaker, 2013: GSI 3DVar-based Ensemble-Variational Hybrid Data Assimilation for NCEP Global Forecast System: Single Resolution Experiments. *Mon. Wea. Rev.*, doi:10.1175/MWR-D-12-00141.1, in press.

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

——, ——, ——, and ——, 2008: Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system. *Tellus*, **60A**, 62–79.

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

——, and ——, 2012: Evaluating methods to account for system errors in ensemble data assimilation. Mon. Wea. Rev., **140**, 3078–3089.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed., Academic Press, 627pp.

Figure Caption List

Figure 1: The regional rescaling (a) 2D mask and (b) 3D mask at the 500hPa level over the period 1 September - 30 November 2012. The contour interval is 0.2 m s^{-1}.

Figure 2: The vertical profile of the 2D mask and 3D mask over the period 1 September - 30 November 2012.

Figure 3: The flow chart of the EnKF_3DR method.

Figure 4: The vertical profile of the square root of total energy (m s^{-1}) of perturbations at (a) the initial time, (b) 12-h, (c) 48-h, and (d) 96-h forecast time for the ETR (dashed) and ET_3DR (solid) experiments at the 500hPa level as a mean for the period 11 September - 30 November 2012.

Figure 5: The square root of total energy (m s^{-1}) of initial perturbations for the (a) ETR and (b) ET_3DR experiments at the 500hPa level as a mean for the period 11 September - 30 November 2012.

Figure 6: The correlation coefficients between the Eady index and the square root of total energy of initial perturbations, which are statistically significant at the 95% confidence interval for 500hPa level over the period 11 September - 30 November 2012.

Figure 7: The average OLR at the 500hPa level over the period 11 September - 30 November 2012. The contour interval is 20W/m^2.

Figure 8: The zonal average of the square root of total energy (m s^{-1}) of initial perturbations for the (a) ETR and (b) ET_3DR experiments at the 500hPa level as a mean for the period 11 September - 30 November 2012.

Figure 9: The ensemble mean RMSE (solid) and ensemble spread (dashed) for (a) U250, (b) Z500, (c) U850 and (d) T850 over the NH. The vertical bars represent the 95% confidence interval from a paried block bootstrap.

Figure 10: Same as in Fig. 9, but over the SH.

Figure 11: The ensemble mean RMSE (solid) and ensemble spread (dashed) for (a) U850 and (b) U10m over the TR for the period 11 September - 30 November 2012. The vertical bars represent the 95% confidence interval from a paried block bootstrap.

Figure 12: The CRPS for (a) U250, (b) Z500, (c) U850 and (d) T850 over the NH for the period 11 September - 30 November 2012. The vertical bars represent the 95% confidence interval from a paried block bootstrap.

Figure 13: Same as in Fig. 12, but over the SH.

Figure 14: The CRPS for (a) U850 and (b) U10m over the TR for the period 11 September - 30 November 2012. The vertical bars represent the 95% confidence interval from a paried block bootstrap.

Figure 15: The mean eigenvalue spectra of the covariance matrix of the initial perturbations for Z500 during the period 11 September - 30 November 2012 over the globe.

Figure 16: The ensemble mean RMSE (solid) and ensemble spread (dashed) for (a) U250, (b) Z500, (c) U850 and (d) T850 over the NH. The vertical bars represent the 95% confidence interval from a paried block bootstrap.

Figure 17: Same as in Fig. 16, but over the SH.

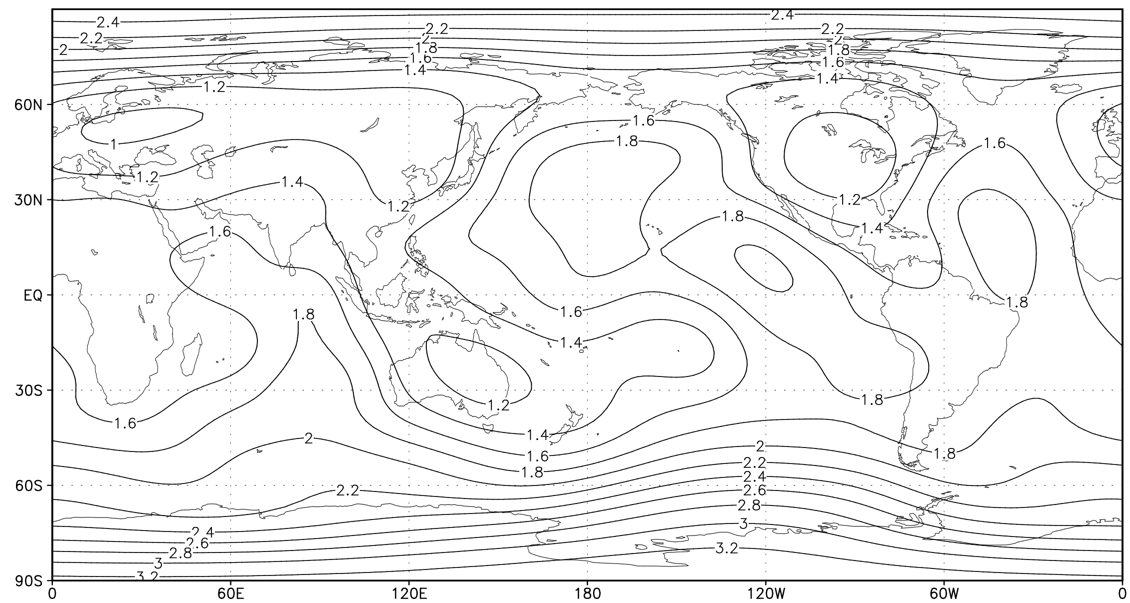
Figure 18: The ensemble mean RMSE (solid) and ensemble spread (dashed) for (a) U850 and (b) U10m over the TR for the period 11 September - 30 November 2012. The vertical bars represent the 95% confidence interval from a paried block bootstrap.

Figure 19: The CRPS for (a) U250, (b) Z500, (c) U850 and (d) T850 over the NH for the period 11 September - 30 November 2012. The vertical bars represent the 95% confidence interval from a paried block bootstrap.

Figure 20: Same as in Fig. 19, but over the SH.

Figure 21: The CRPS for (a) U850 and (b) U10m over the TR for the period 11 September - 30 November 2012. The vertical bars represent the 95% confidence interval from a paried block bootstrap.

(a)



(b)

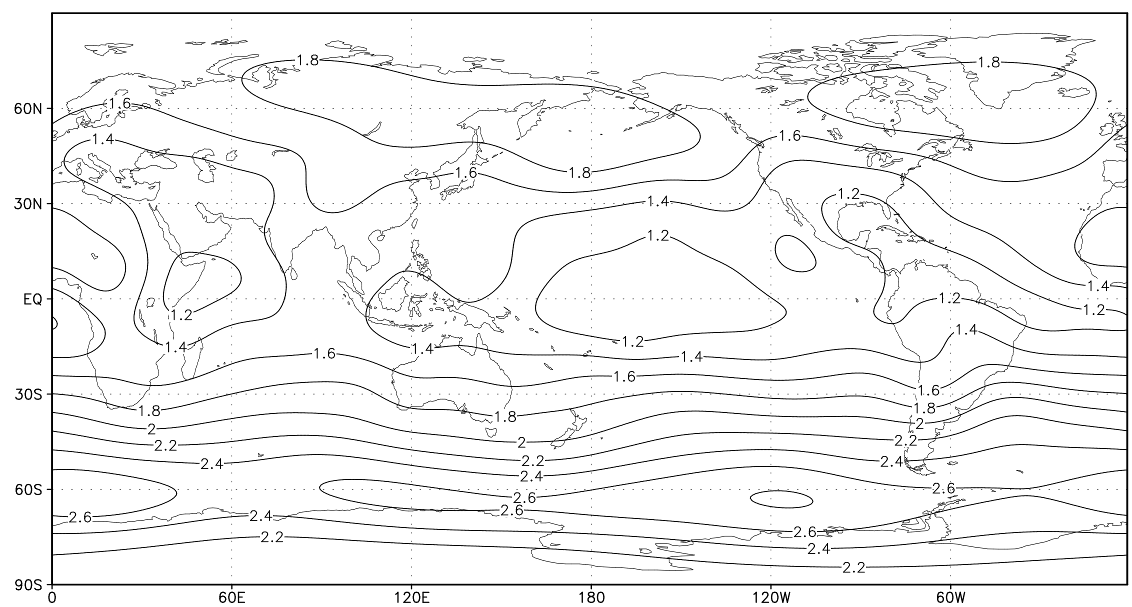


Fig. 1 The regional rescaling (a) 2D mask and (b) 3D mask at the 500hPa level over the period 1 September - 30 November 2012. The contour interval is 0.2 m s^{-1}.

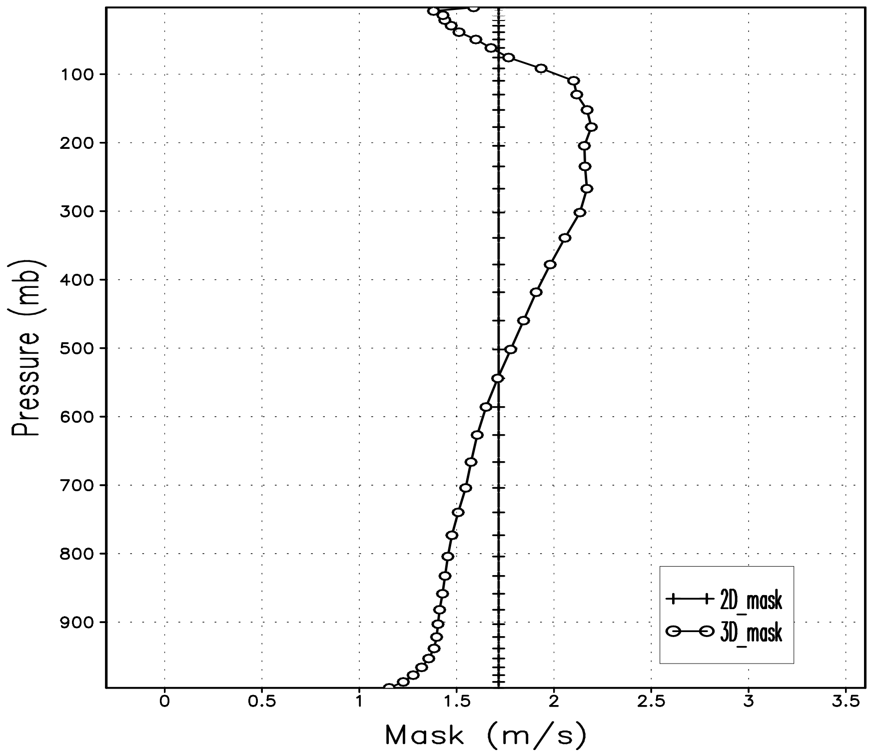


Fig. 2 The vertical profile of the 2D mask and 3D mask over the period 1 September - 30 November 2012.

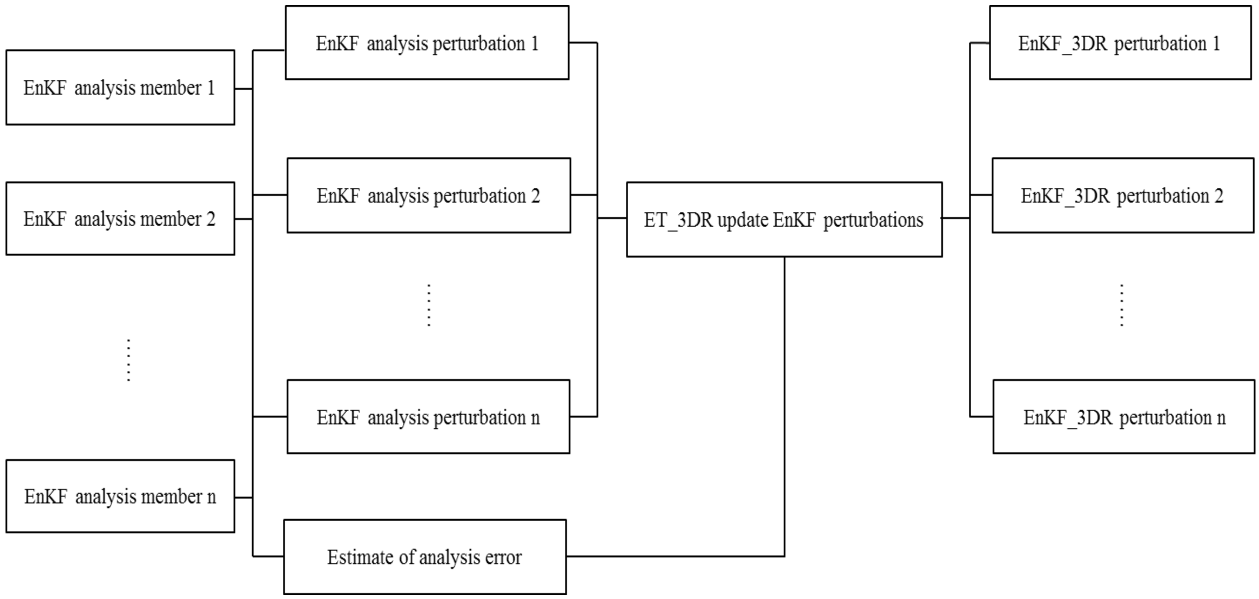


Fig. 3 The flow chart of the EnKF_3DR method.

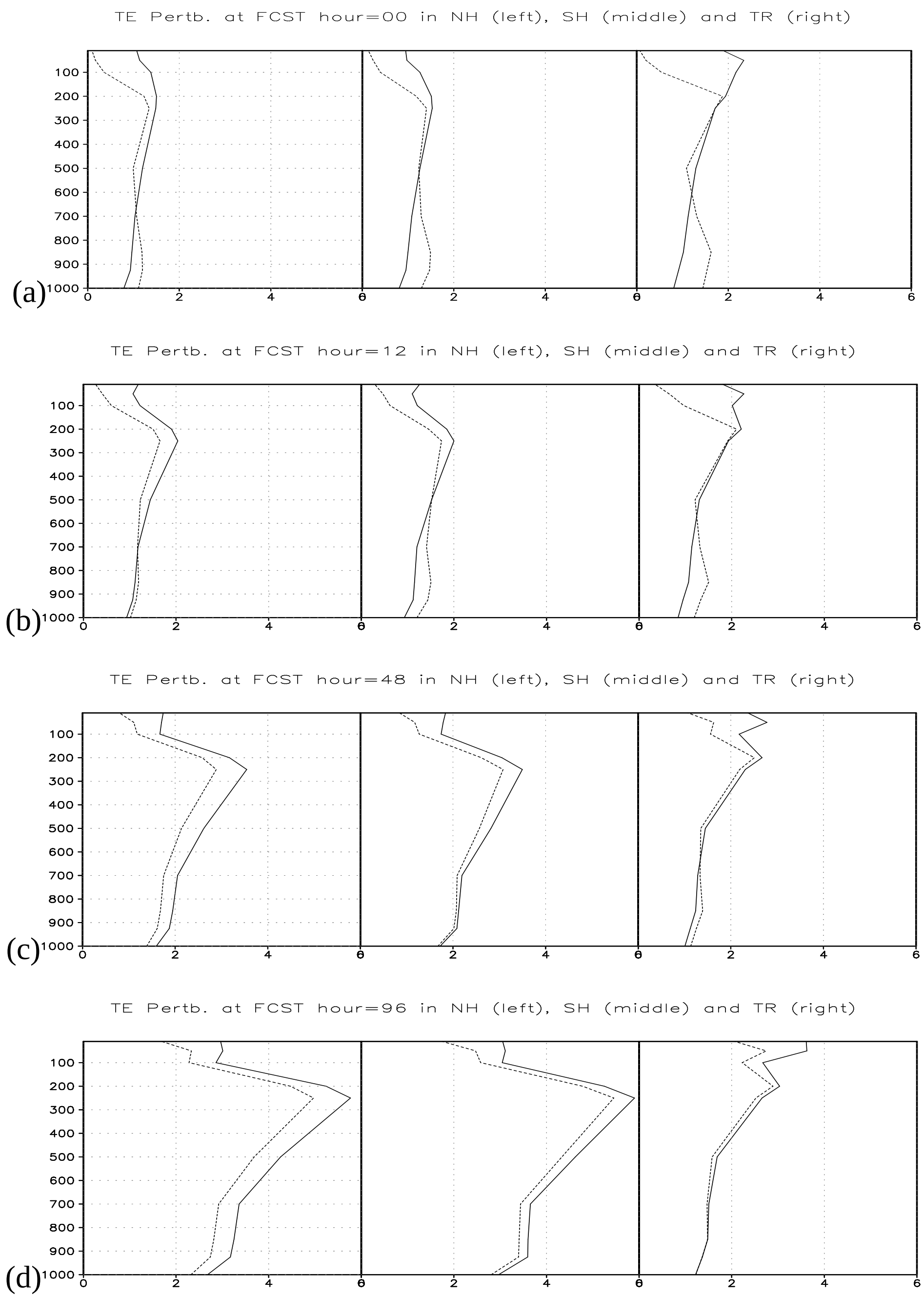


Fig. 4 The vertical profile of the square root of total energy (m s^{-1}) of perturbations (a) at the initial time, after (b) 12-h, (c) 48-h, and (d) 96-h for the ETR (dashed) and ET_3DR (solid) experiments at the 500hPa level as a mean for the period 11 September - 30 November 2012.

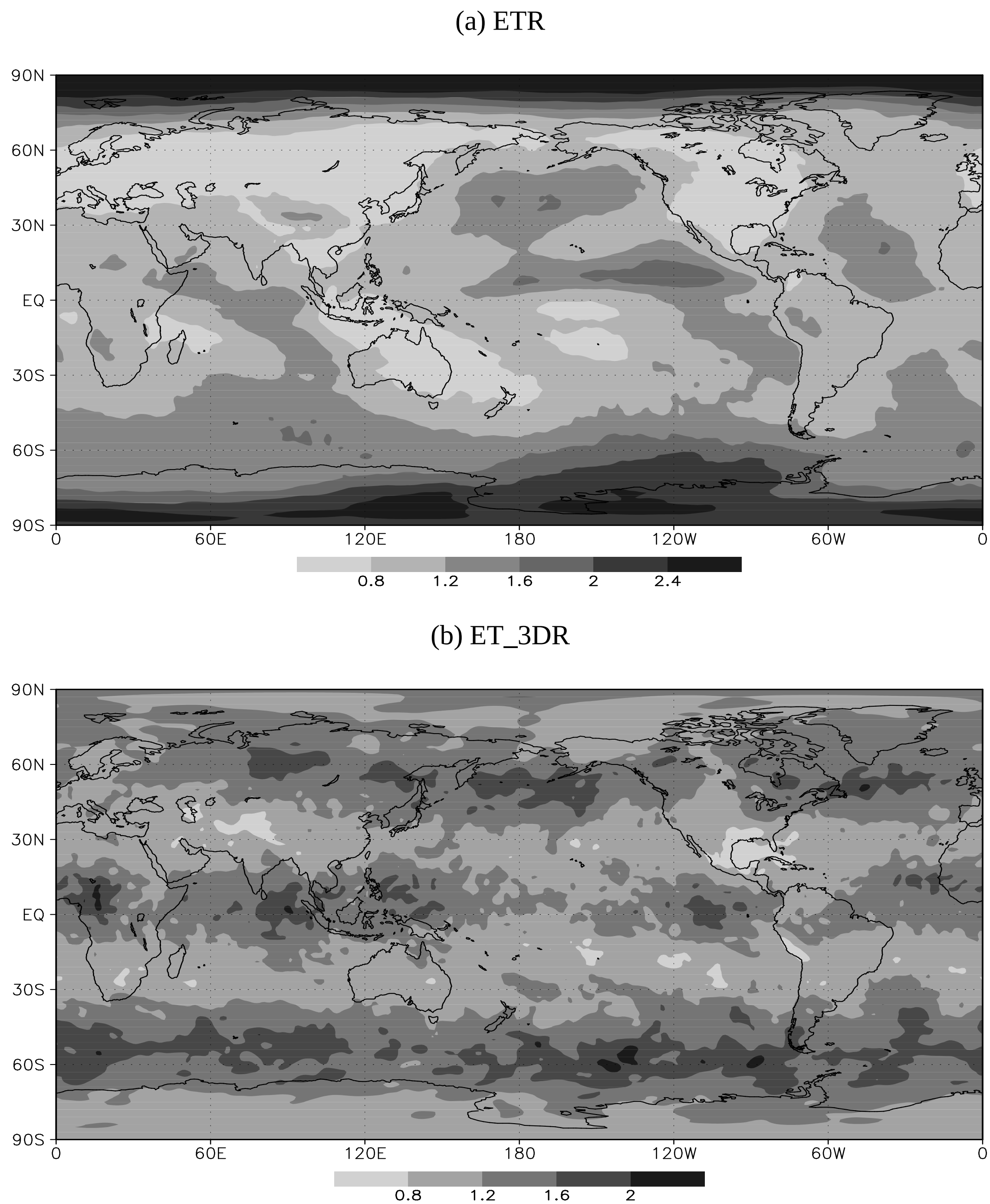
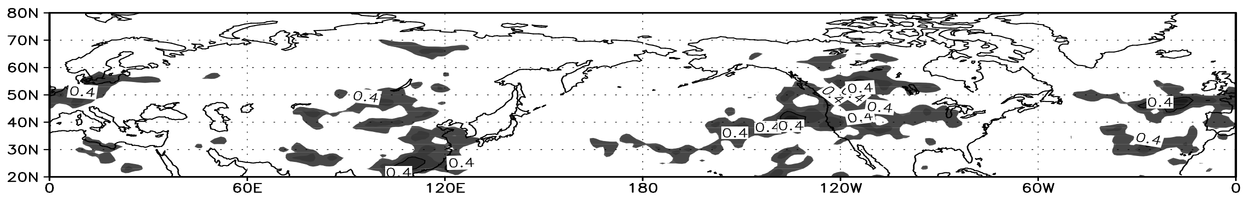
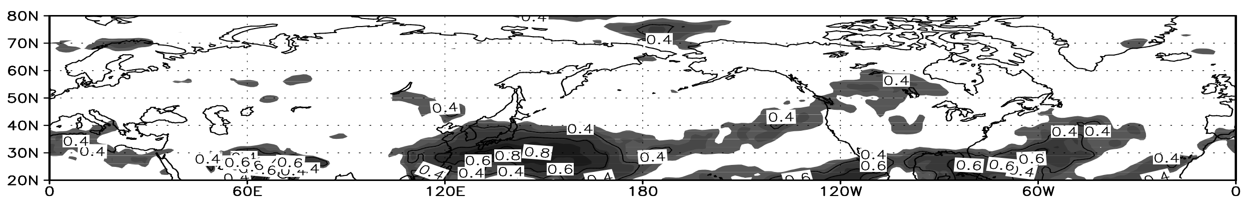


Fig. 5 The square root of total energy (m s^{-1}) of initial perturbations for the (a) ETR and (b) ET_3DR experiments at the 500hPa level as a mean for the period 11 September - 30 November 2012.

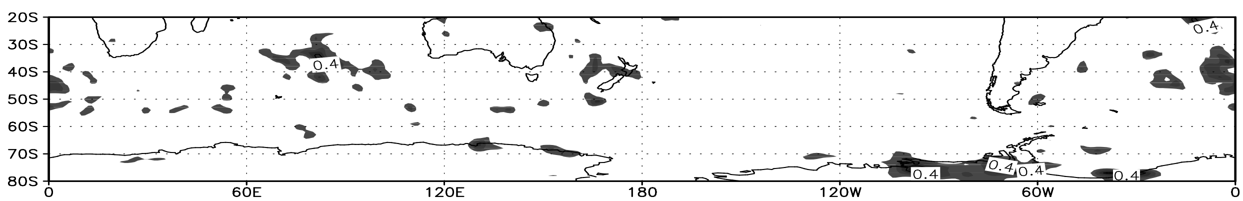
(a) ETR over NH



(b) ET_3DR over NH



(c) ETR over SH



(d) ET_3DR over SH

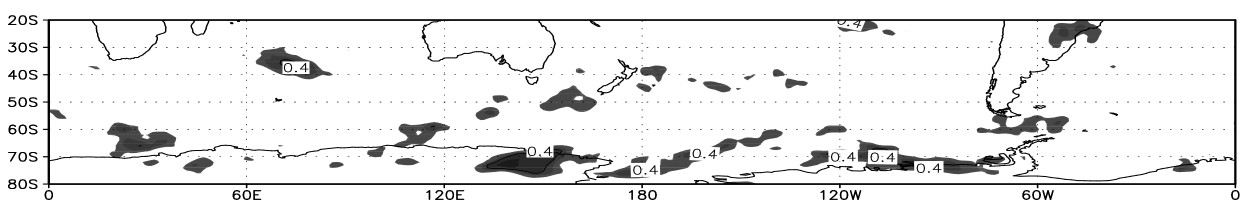


Fig. 6 The correlation coefficients between the Eady index and the square root of total energy of initial perturbations, which are statistically significant at the 95% confidence interval for 500hPa level over the period 11 September - 30 November 2012.

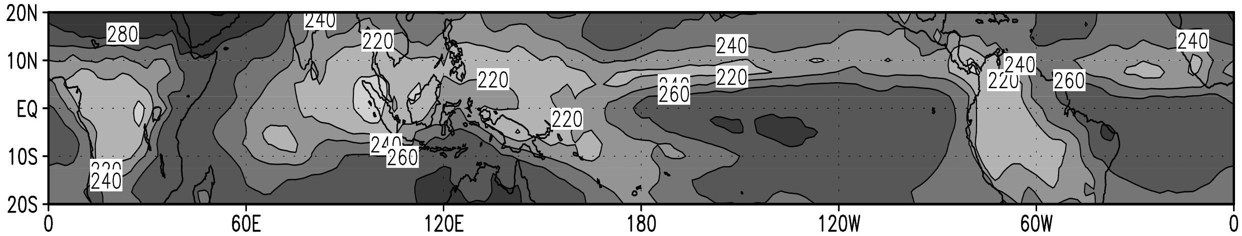


Fig. 7 The average OLR at the 500hPa level over the period 11 September - 30 November 2012. The contour interval is 20W/m^2.

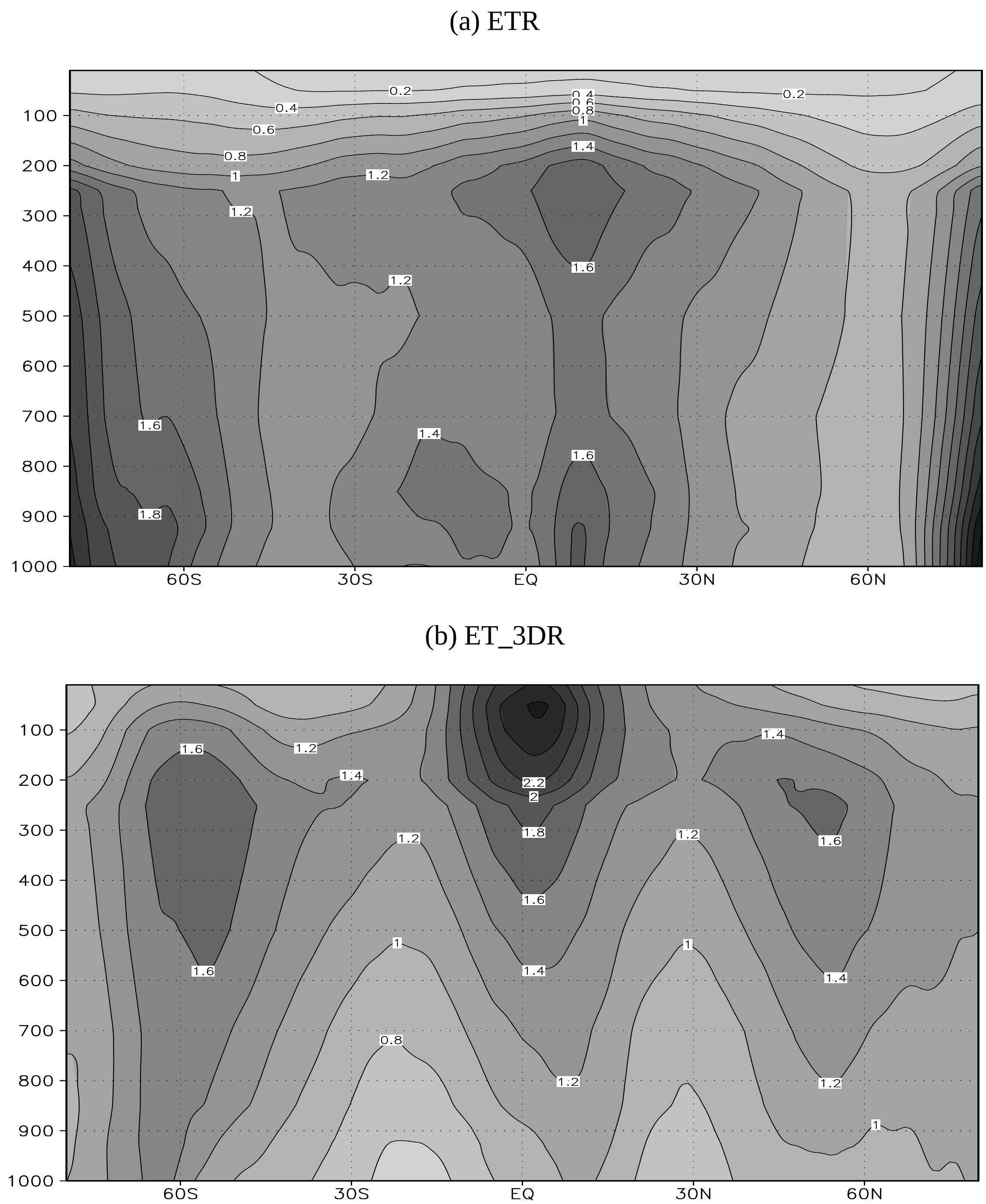
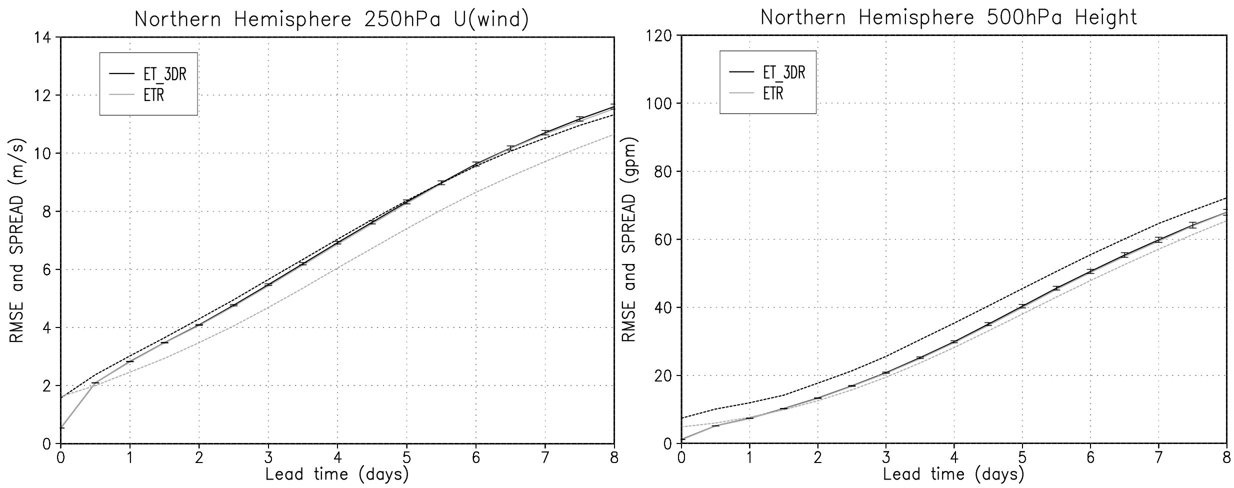


Fig. 8 The zonal average of the square root of total energy (m s^{-1}) of initial perturbations for the (a) ETR and (b) ET_3DR experiments at the 500hPa level as a mean for the period 11 September - 30 November 2012.

(a) (b)



(c) (d)

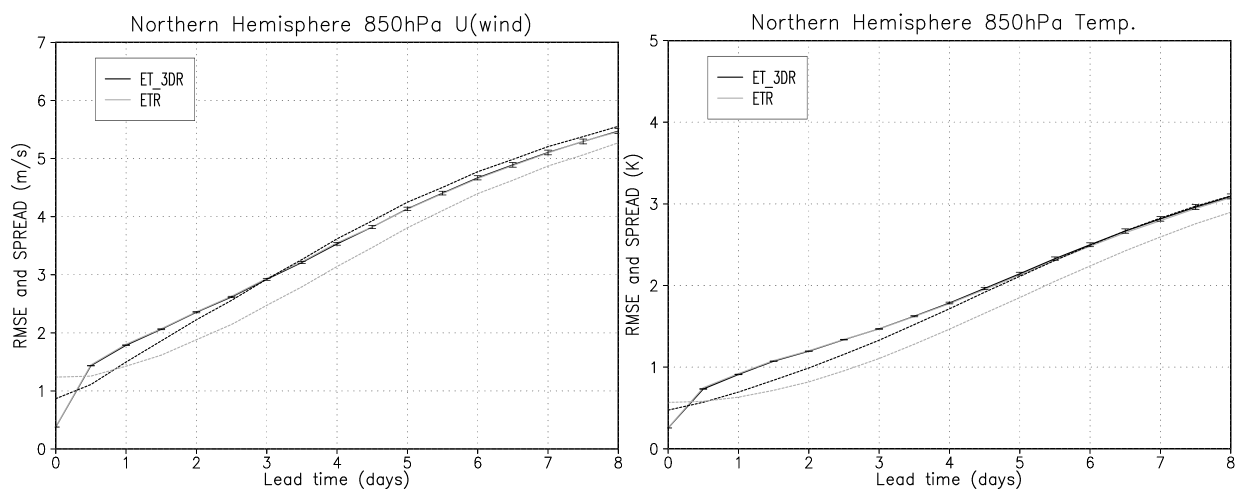


Fig. 9 The ensemble mean RMSE (solid) and ensemble spread (dashed) for (a) U250, (b) Z500, (c) U850 and (d) T850 over the NH. The vertical bars represent the 95% confidence interval from a paried block bootstrap.

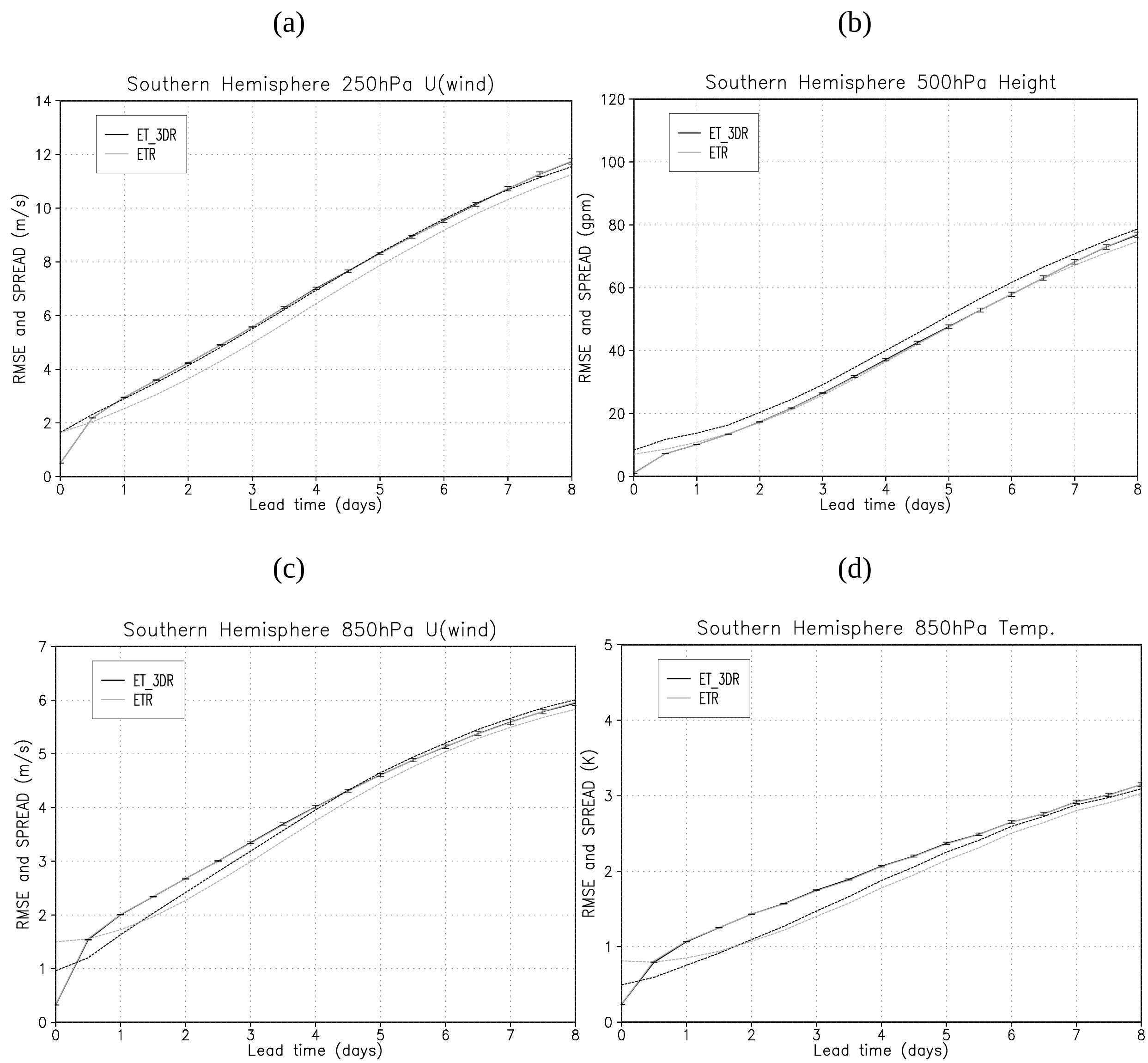


Fig. 10 Same as in Fig. 9, but over the SH.

(a) (b)

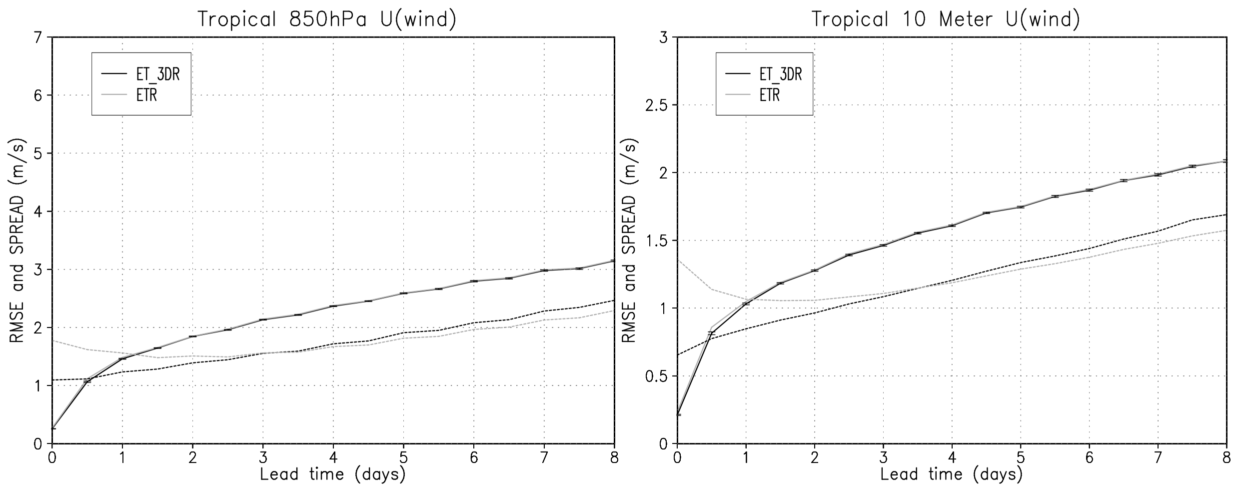


Fig. 11 The ensemble mean RMSE (solid) and ensemble spread (dashed) for (a) U850 and (b) U10m over the TR for the period 11 September - 30 November 2012. The vertical bars represent the 95% confidence interval from a paried block bootstrap.

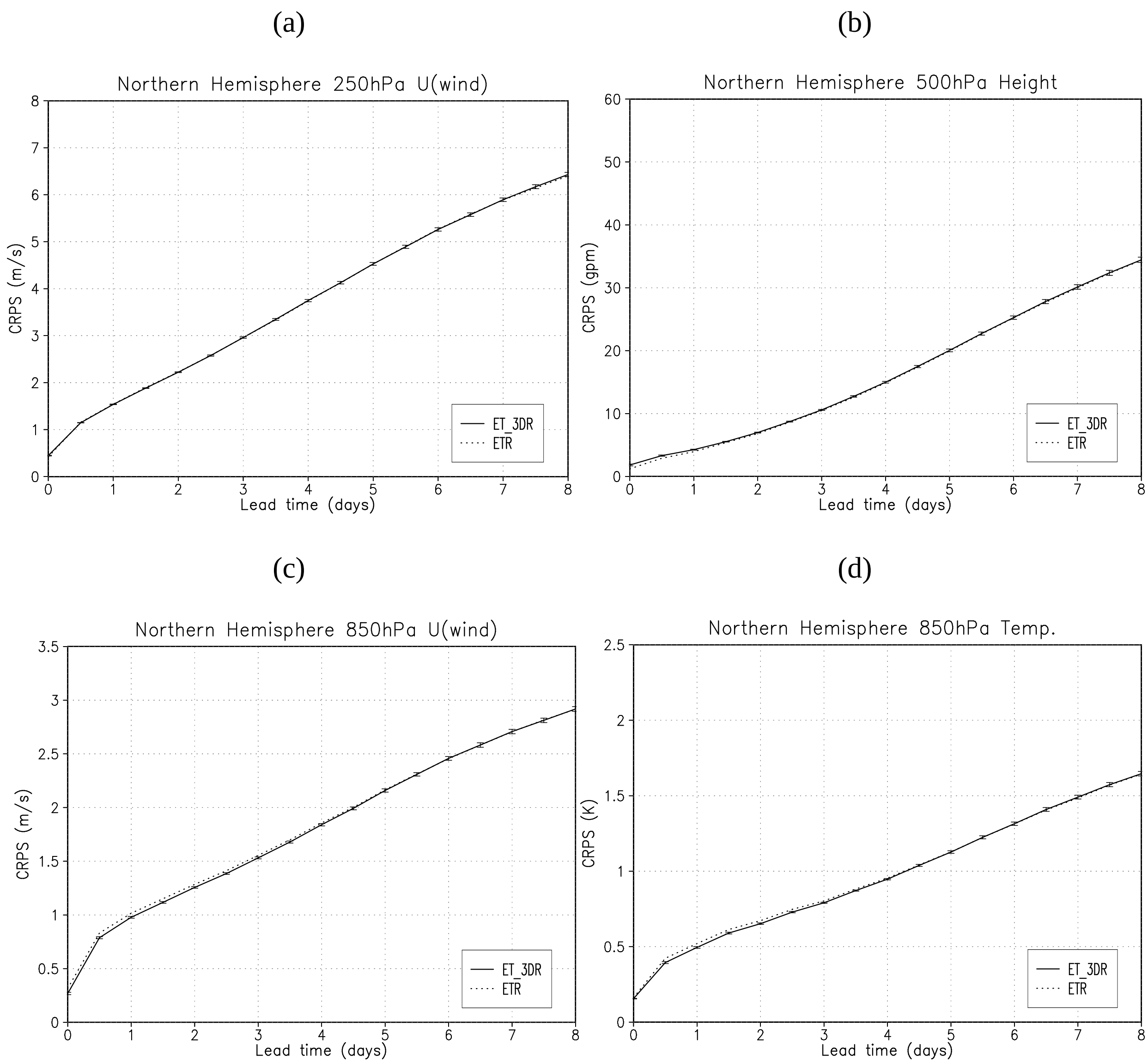


Fig. 12 The CRPS for (a) U250, (b) Z500, (c) U850 and (d) T850 over the NH for the period 11 September - 30 November 2012. The vertical bars represent the 95% confidence interval from a paried block bootstrap.

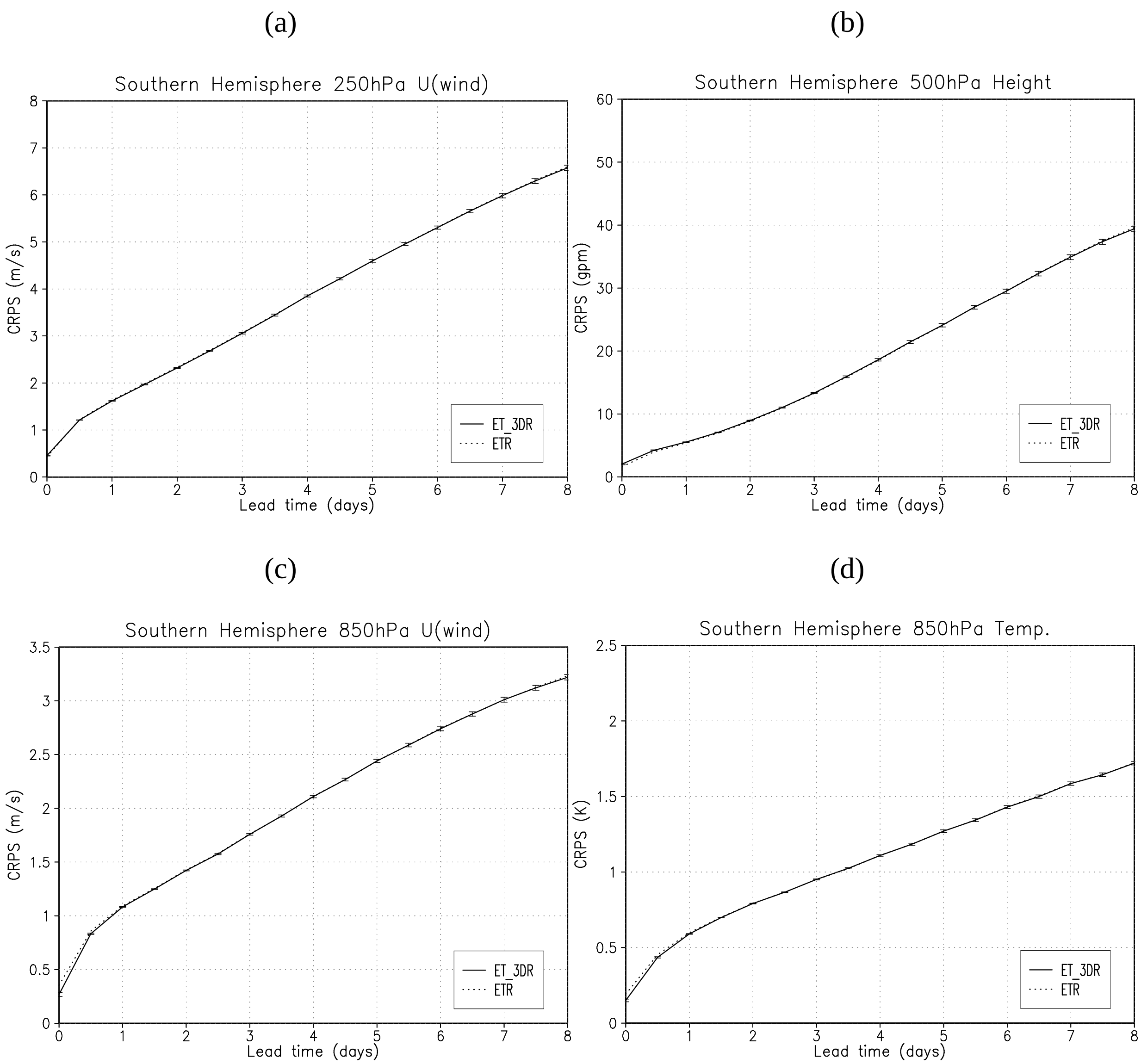


Fig. 13 Same as in Fig. 12, but over the SH.

(a) (b)

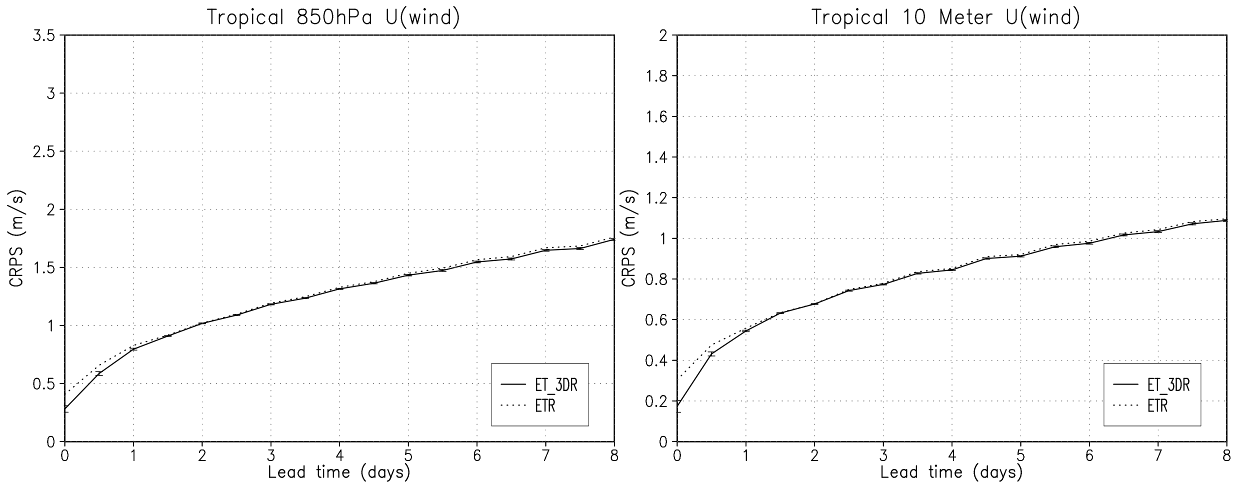


Fig.14 The CRPS for (a) U850 and (b) U10m over the TR for the period 11 September - 30 November 2012. The vertical bars represent the 95% confidence interval from a paried block bootstrap.

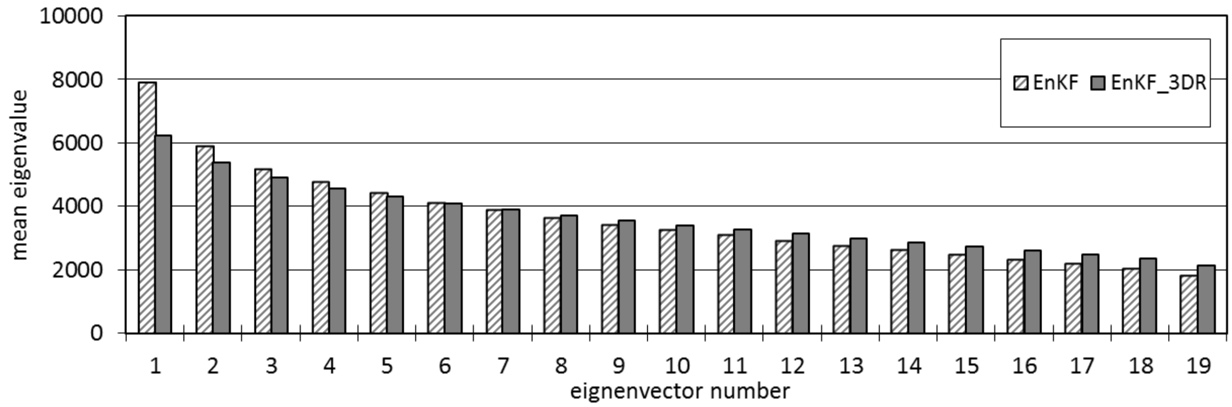


Fig. 15 The mean eigenvalue spectra of the covariance matrix of the initial perturbations for Z500 during the period 11 September - 30 November 2012 over the globe.

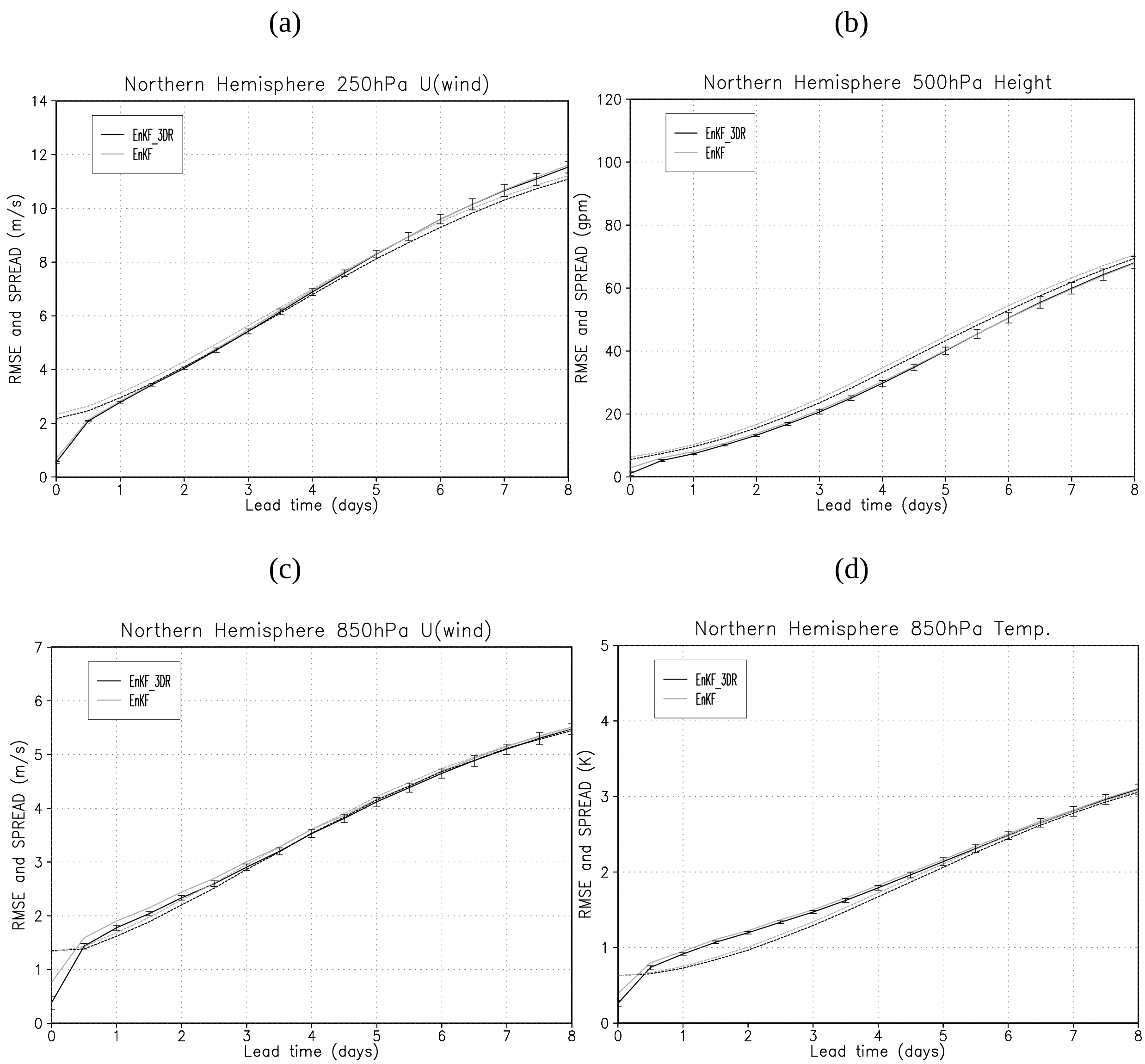


Fig. 16 The ensemble mean RMSE (solid) and ensemble spread (dashed) for (a) U250, (b) Z500, (c) U850, and (d) T850 over the NH. The vertical bars represent the 95% confidence interval from a paried block bootstrap.

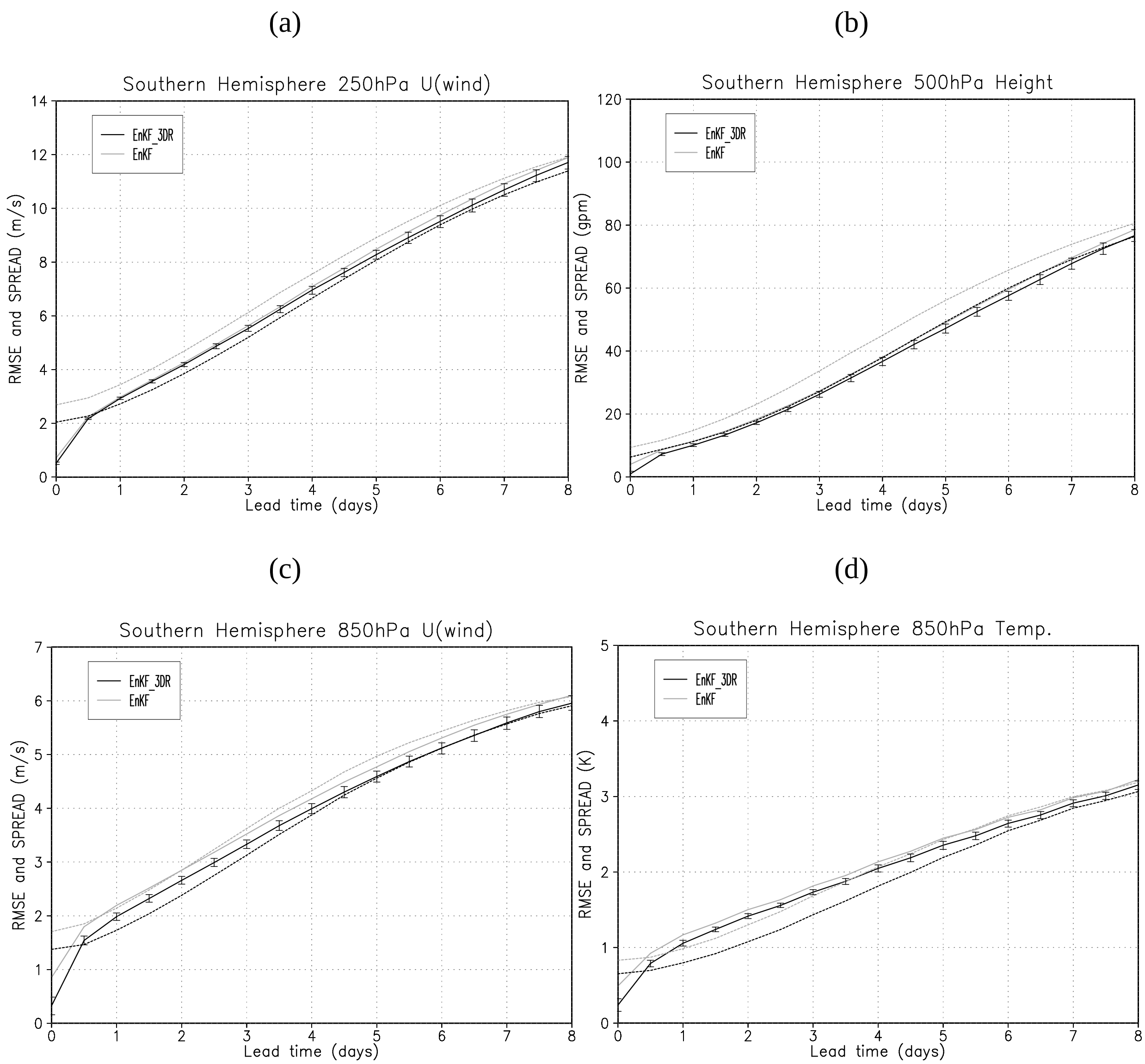


Fig. 17 Same as in Fig. 16, but over the SH.

(a) (b)

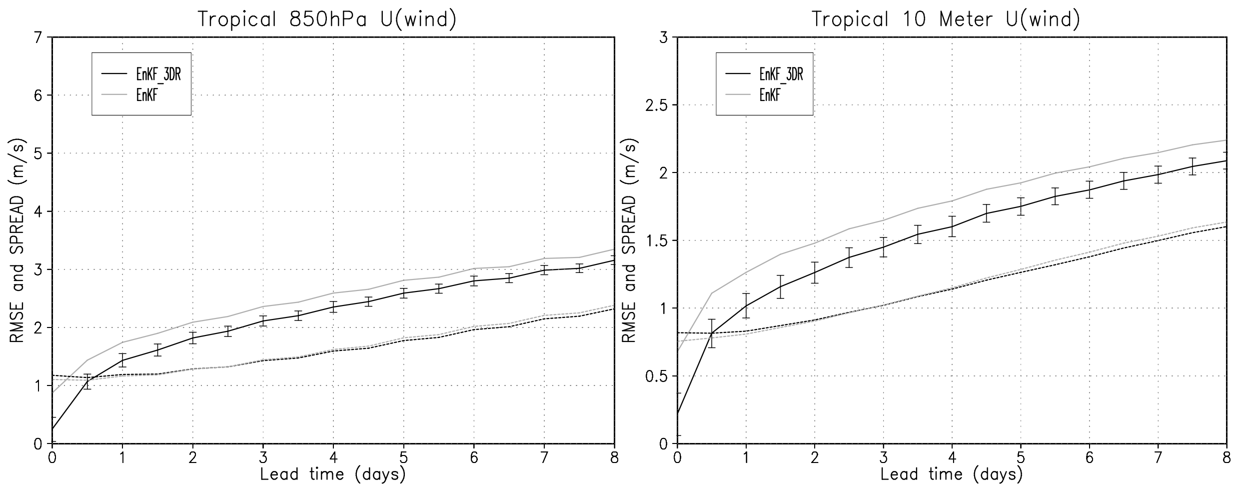
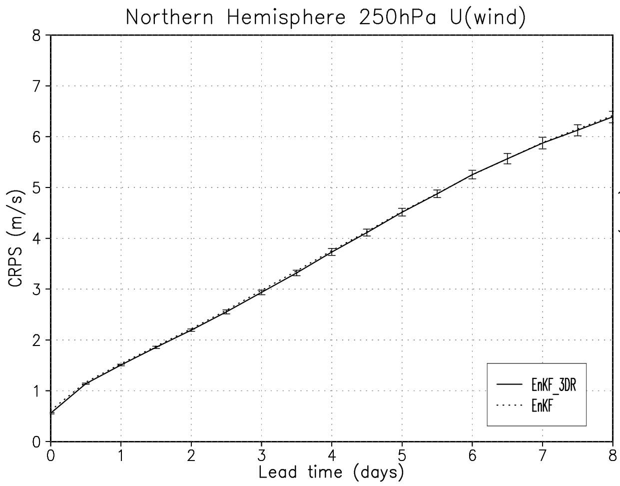
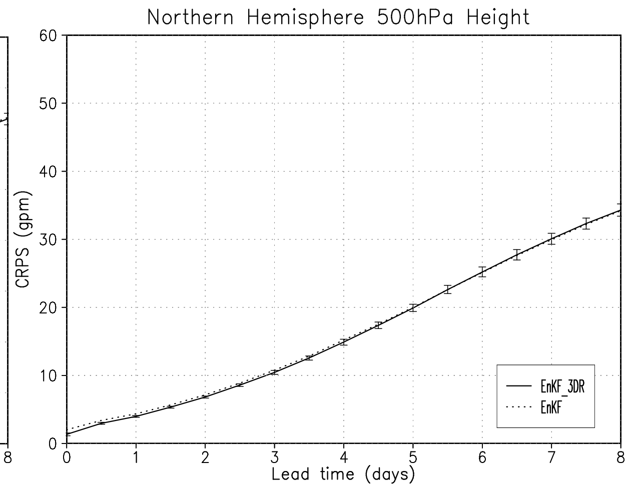


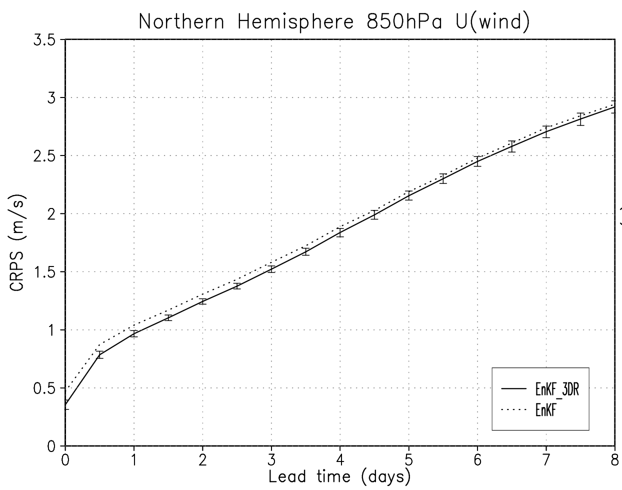
Fig. 18 The ensemble mean RMSE (solid) and ensemble spread (dashed) for (a) U850 and (b) U10m over the TR for the period 11 September - 30 November 2012. The vertical bars represent the 95% confidence interval from a paried block bootstrap.

(a) (b)





(c) (d)



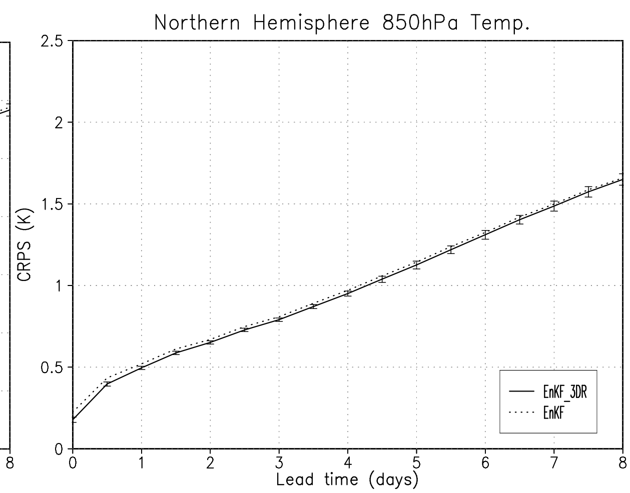


Fig. 19 The CRPS for (a) U250, (b) Z500, (c) U850 and (d) T850 over the NH for the period 11 September - 30 November 2012. The vertical bars represent the 95% confidence interval from a paried block bootstrap.

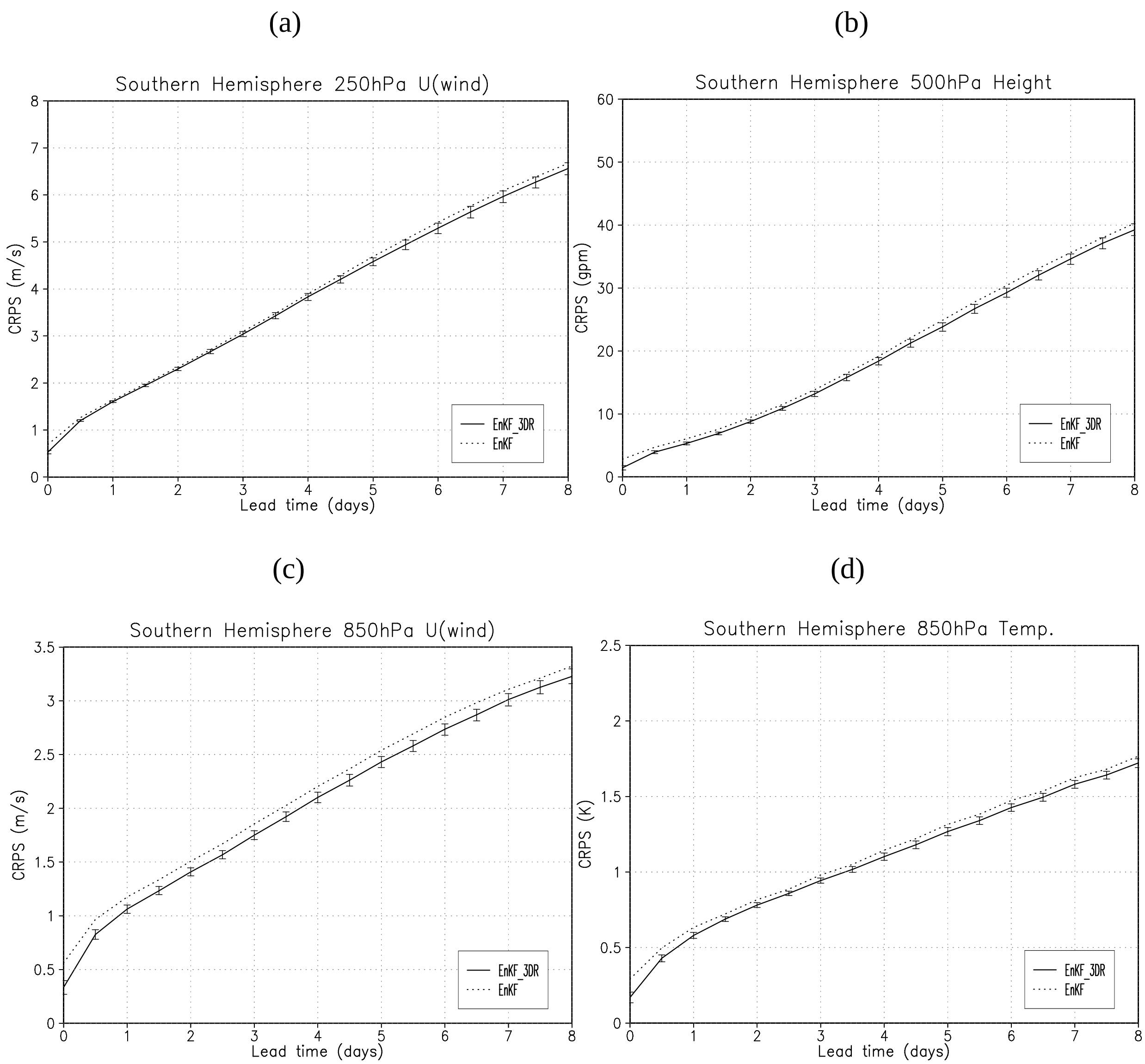


Fig. 20 Same as in Fig. 19, but over the SH.

(a) (b)

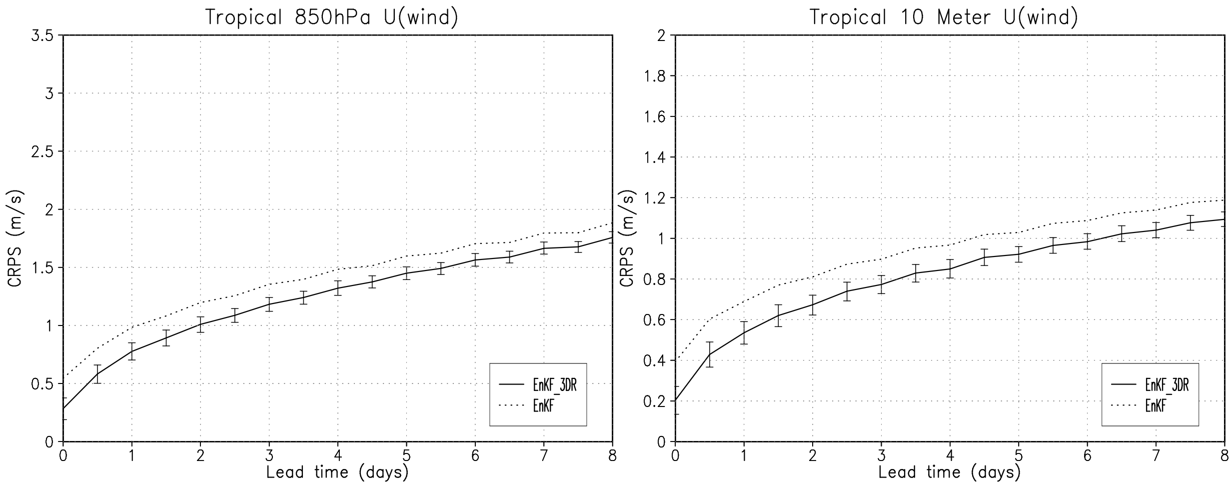


Fig. 21 The CRPS for (a) U850 and (b) U10m over the TR for the period 11 September - 30 November 2012. The vertical bars represent the 95% confidence interval from a paried block bootstrap.