

## Ensemble Transform with 3D Rescaling Initialization Method

JUHUI MA

*Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, Nanjing, and State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing, China*

YUEJIAN ZHU AND DINGCHEN HOU

*NOAA/NCEP/Environmental Modeling Center, College Park, Maryland*

XIAQIONG ZHOU AND MALAQUIAS PEÑA

*I.M. Systems Group, Inc., and NOAA/NCEP/Environmental Modeling Center, College Park, Maryland*

(Manuscript received 15 November 2013, in final form 2 July 2014)

### 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 variations of analysis error. However, because of a lack of suitable three-dimensional (3D) analysis error estimation, in the operational GEFS the mask is based on the estimated analysis error at 500 hPa 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 within the ensemble Kalman filter (EnKF) method (hereafter EnKF\_3DR).

Results from a set of experiments indicate that the 3D mask suppresses perturbations less in unstable regions. Relative to the ETR, the large amplitudes of the ET\_3DR initial perturbations at 500 hPa better reflect areas of baroclinic instability over the extratropics and deep convection over the tropics. Furthermore, the maxima of the vertical distribution for the ET\_3DR initial perturbations correspond to the heights of the subtropical westerly and tropical easterly jet regions. Such perturbations produce faster spread growths. Results with EnKF\_3DR also show benefits from an orthonormalization by the ensemble transform algorithm and amplitude constraint by the 3D mask rescaling. Thus, the EnKF\_3DR forecasts outperform 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 probabilistic forecast performance. The analysis errors can be decomposed into nongrowing and growing modes (Toth and Kalnay 1997). The nongrowing error modes, mainly stemming from observational errors, have a large dimensional subspace, and cannot be sampled well with a limited number of ensemble members;

thus, these error modes will typically lose their amplitude rapidly. The growing errors, dynamically generated from the short-range forecast in the analysis cycle by the use of initial conditions with errors and an imperfect model, amplify quickly and will eventually dominate the forecast error. Therefore, the success of an ensemble generation method lies in how well its perturbations sample the regions with the fastest-growing errors in the analysis.

The breeding vector (BV) method (Toth and Kalnay 1993, 1997) creates perturbations that grow rapidly by inserting 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).

---

*Corresponding author address:* Juhui Ma, Nanjing University of Information Science and Technology, 219 Ningliu Road, Nanjing, Jiangsu 210044, China.  
E-mail: juhui.ma@gmail.com

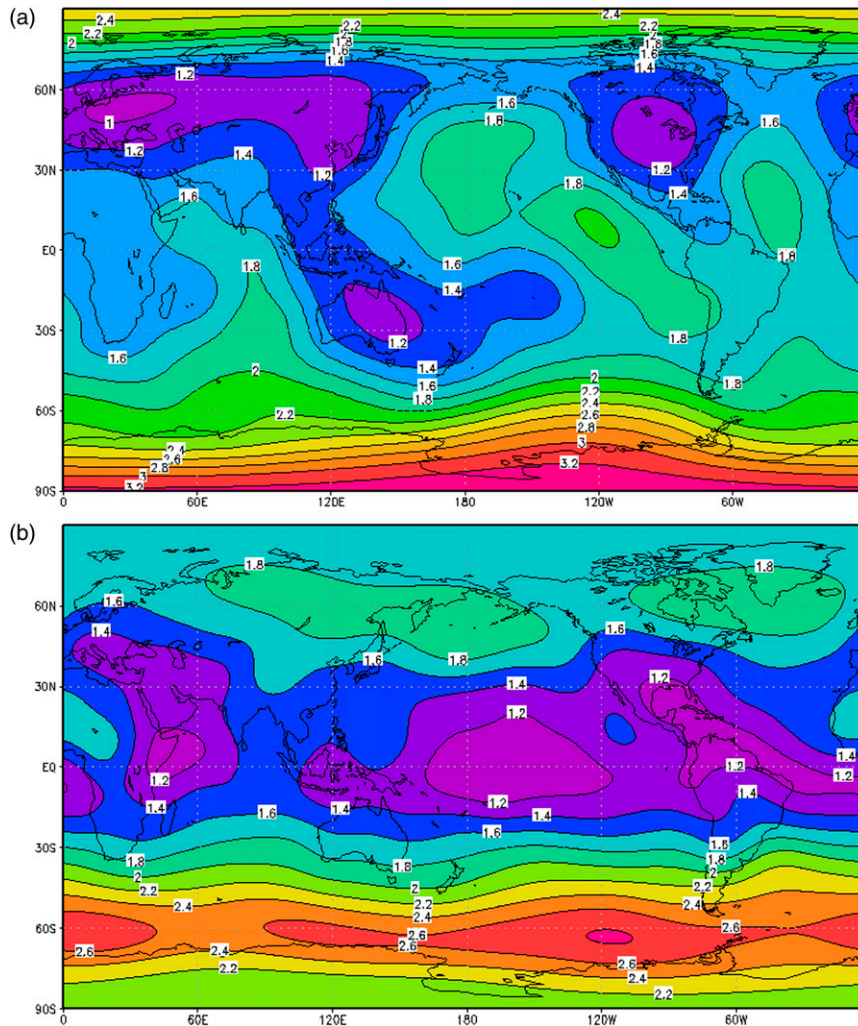


FIG. 1. The regional rescaling (a) 2D mask and (b) 3D mask at the 500-hPa level as the average for the period 1 Sep–30 Nov 2012. The contour interval is  $0.2 \text{ m s}^{-1}$ .

Therefore, an improved version of the BV, the ensemble transform (ET) method has been introduced to generate initial perturbations that are globally transformed from the forecast perturbations (Bishop and Toth 1999; Wei et al. 2008). As with the BV method, the ET also generates a flow-dependent error covariance spatial structure and can represent fast-growing components 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.

In fact, there are some shortcomings in the ET initialization. Although the ET perturbations are globally dependent on the analysis error variance during the

transformation, the distribution of the initial spread 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 by partitioning the global domain into latitude bands or latitude–longitude blocks, resulting in better agreement with the analysis error variance and improved ensemble performance.

In addition, the ET perturbations should project into fast-growing modes, but the distribution of the fast-growing modes is not the same as the analysis uncertainties, due to unevenly distributed observations and model errors. Therefore, the energy of the initial perturbations needs to be redistributed. At the National Centers for Environmental Prediction (NCEP), a regional rescaling process is imposed onto the ET initialization periodically to suppress high-amplitude perturbations in areas

where the analysis uncertainties are relatively low. This regional rescaling improves both the distributional spread of initial perturbations and most probabilistic scores with respect to the ET without rescaling (Wei et al. 2008). Magnusson et al. (2009) compared the ensemble transform with rescaling (ETR) and singular vector (SV) methods within the same model and data assimilation system. The results indicated that ETR had more advantages than SV. Wei et al. (2008) compared the ETR and ensemble transform Kalman filter (ETKF) methods and showed that ETR performed better than ETKF for most skill scores. Further, X. Zhou et al. (2014, unpublished manuscript) showed that the ensemble Kalman filter (EnKF) outperformed ETR in terms of the continuous ranked probability skill score over the Northern Hemisphere whereas the ensemble mean forecast is more skillful in ETR.

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 operational scheme, the mask is calculated using a long-term-averaged root-mean-square of analysis error variance in the kinetic energy norm at the 500-hPa level obtained from the variational data assimilation system (Szunyogh and Toth 2002; Wei et al. 2008). However, the current mask does not adequately represent analysis uncertainties for use in the context of the ensemble forecast system. First and foremost, the two-dimensional (2D) mask cannot represent the vertical structure of analysis uncertainties. To compensate for the resultant underestimate of analysis error, empirically derived additional inflation factors have to be applied to the mask for levels below 500 hPa. This is obviously suboptimal for application of regional rescaling. Second, the mask was computed from a decade's worth of past climatological data, during which the density and accuracy of observations, as well as the data assimilation technology itself, have greatly changed. Thus, there is a need to update the mask with new estimates of analysis errors based on the current real-time data assimilation system to make the initial perturbations more consistent with the observations and the data assimilation system. Wei et al. (2008) found that, due to the mask, the ETR failed to show high spread in the Southern Ocean storm-track area in comparison to the ET method, which indicated a more accurate time-dependent mask was necessary. Third, the total energy norm may be a more reasonable measure of the magnitude of initial perturbations than the kinetic energy norm. Palmer et al. (1998) found that total energy is more consistent with analysis error statistics than the streamfunction, enstrophy, or kinetic energy metrics. Some previous studies have designed different masks to

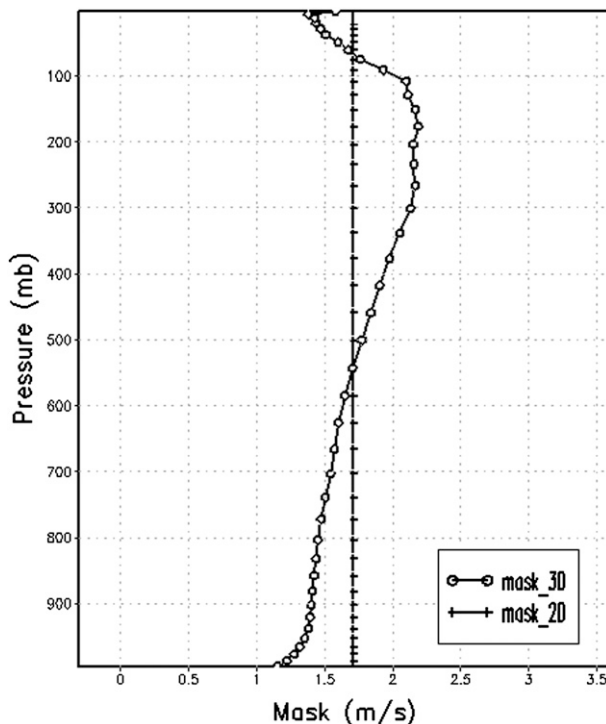


FIG. 2. The global average vertical distribution of the 2D mask and 3D mask for the period 1 Sep–30 Nov 2012.

address some of these issues. Wang and Bishop (2003) chose the square root of the seasonally and vertically averaged initial wind variance across the ETKF ensemble as the mask applied in the BV method. Magnusson et al. (2009) designed a vertically integrated estimation of analysis errors using the total energy norm from four-dimensional variational data assimilation (4D-Var) system as the mask. In this paper, a new mask will be defined by 3D analysis uncertainty measured by the 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 Global Ensemble Forecast System (GEFS) will be explored.

Relative to a variational data assimilation method with static background error, ensemble-based data assimilation provides the additional ability to provide flow-dependent estimates of the background error. Moreover, ensemble-based data assimilation generates an initial analysis for an ensemble of predictions in the next cycle and provides 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, including the Meteorological Service of Canada (MSC; Buehner et al. 2010a,b) and the European

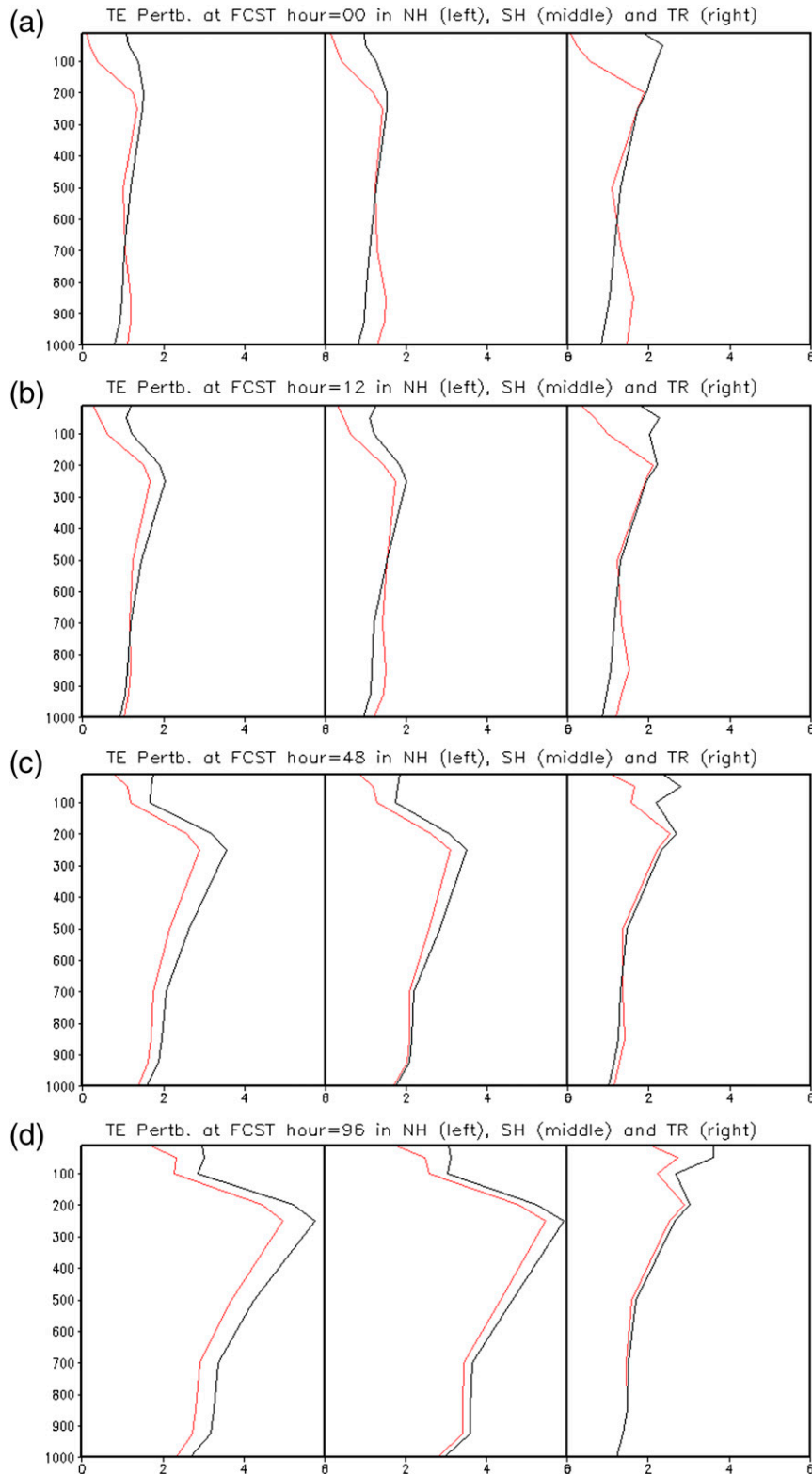


FIG. 3. The vertical profile of the square root of total energy ( $\text{m s}^{-1}$ ) of perturbations at (a) the initial time, (b) 12-h, (c) 48-h, and (d) 96-h forecast time for the ETR (red) and ET\_3DR (black) experiments as an average for the period 11 Sep–30 Nov 2012: (left) NH, (middle) SH, and (right) TR.

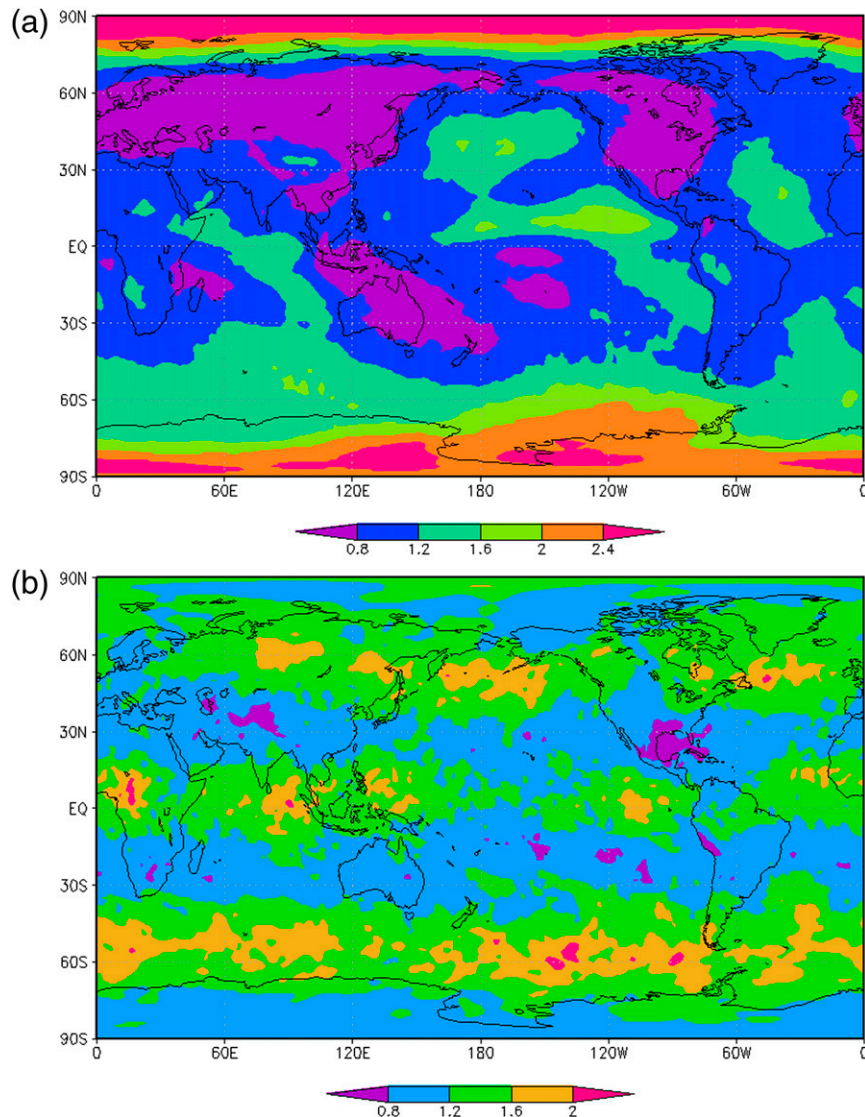


FIG. 4. The square root of total energy ( $\text{m s}^{-1}$ ) of initial perturbations for the (a) ETR and (b) ETR experiments at the 500-hPa level as an average for the period 11 Sep–30 Nov 2012.

Centre for Medium-Range Weather Forecasts (ECMWF; Buizza et al. 2008, 2010). A hybrid 3D-Var/EnKF data assimilation system became operational on 22 May 2012 at NCEP (Wang et al. 2013). In this system, the background error is created by a combination of static background error from the 3D-Var system and flow-dependent background error produced from the EnKF. Furthermore, EnKF perturbations are recentered within the hybrid analysis. The hybrid 3D-Var/EnKF has provided better analyses and forecasts than the previous operational 3D-Var system (see <http://www.emc.ncep.noaa.gov/GFS/impl.php>).

The availability of the EnKF in the NCEP Global Data Assimilation System (GDAS) provides an

alternative ensemble initial condition set for the operational GEFS. The performances of the EnKF and ETR perturbations in the NCEP operational environment are compared in X. Zhou et al. (2014, unpublished manuscript). It is found that the amplitude of initial perturbations is larger in EnKF than in ETR especially over the Southern Hemisphere, which leads to overdispersive ensemble spread and larger root-mean-square error (RMSE) of ensemble mean. Since the ETR method is able to maximize the effective degrees of perturbation freedom and constrain the amplitude of initial perturbations to vary in accordance with regional variations of analysis uncertainties without an undue burden on computer

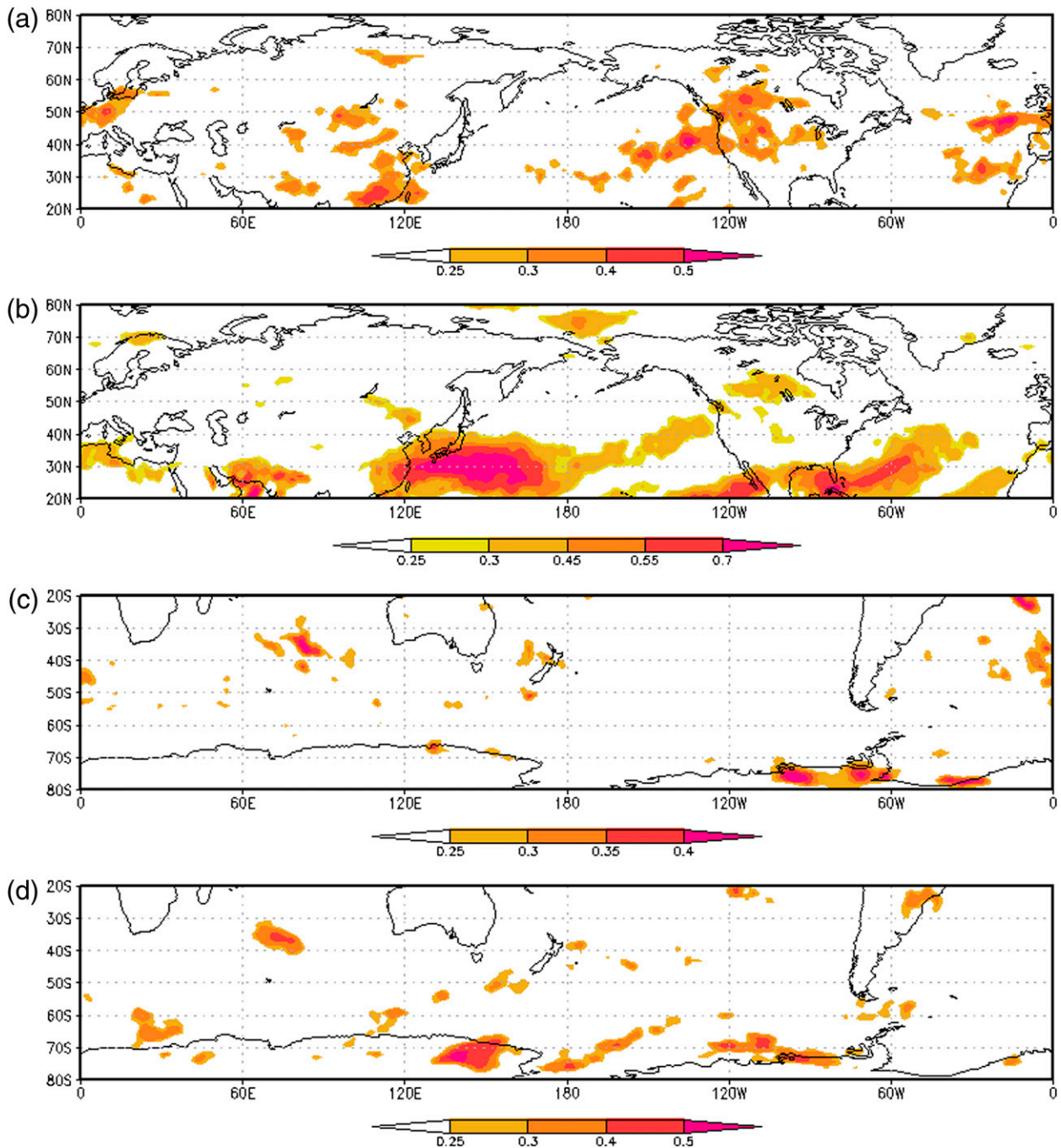


FIG. 5. The correlation coefficients between the Eady index and the square root of total energy of initial perturbations that are statistically significant at the 95% confidence interval for 500-hPa level over the period 11 Sep–30 Nov 2012 for the ETR experiments in the (a) NH and (c) SH; and the ET\_3DR experiments in the (b) NH and (d) SH.

resources, applying the ETR across other analysis ensembles (e.g., multicenter analyses, or analysis from ensemble-based data assimilation) may have a positive impact on the quality of initial conditions. In this study, the EnKF initial perturbations will be transformed and rescaled by the ensemble transform

with 3D rescaling (ET\_3DR) method and the impact will be explored.

In the next section, the methodology of ET\_3DR and the ensemble transform with 3D rescaling applied within the ensemble Kalman filter (EnKF\_3DR) are described. Section 3 investigates the horizontal and vertical

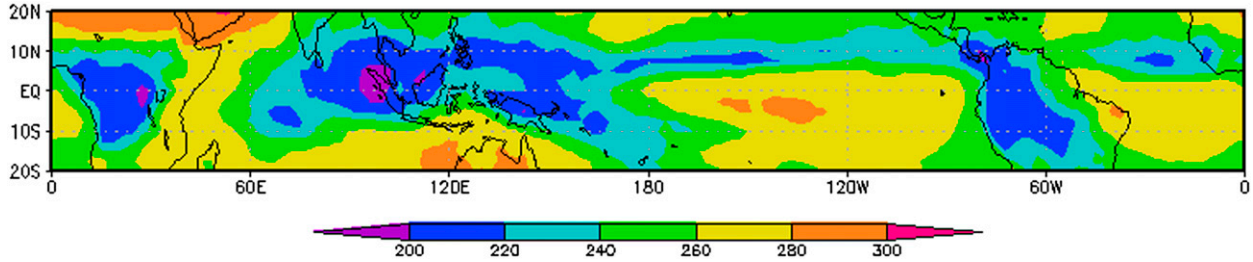


FIG. 6. The average OLR at the 500-hPa level over the period 11 Sep–30 Nov 2012. The contour interval is  $20 \text{ W m}^{-2}$ .

distributions of perturbations generated by the ETR and ET\_3DR, and compares their forecast performances. In section 4, the characteristics of perturbations generated by the EnKF and EnKF\_3DR are analyzed, and the methods' forecast skills are compared. The conclusions are summarized in section 5.

## 2. Initialization methodologies and experimental design

### a. Initialization methodologies

#### 1) THE ET\_3DR METHOD

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

$$\mathbf{X}^a = \mathbf{X}^f \mathbf{T}, \quad (1)$$

where  $n$  analysis perturbations  $\mathbf{x}_i^a$  ( $i = 1, 2, \dots, n$ ) are listed as columns in the matrix  $\mathbf{X}^a$ , and  $n$  forecast perturbations  $\mathbf{x}_i^f$  ( $i = 1, 2, \dots, n$ ) 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

$$\mathbf{T} = \mathbf{C}\mathbf{\Gamma}^{-1/2}, \quad (2)$$

where columns of the matrix  $\mathbf{C}$  contain the orthonormal eigenvectors ( $c_i$ ,  $i = 1, 2, \dots, n$ ) 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 ( $\lambda_i$ ,  $i = 1, 2, \dots, n$ ), in which the first  $n-1$  eigenvalues are nonzero and the last eigenvalue is zero. A diagonal matrix  $\mathbf{F}$  is defined by setting the zero eigenvalue in  $\mathbf{\Gamma}$  to a nonzero constant  $\alpha$ . The diagonal matrix  $\mathbf{P}^a$  contains the analysis error variances. Then, a simplex transformation (Wang et al. 2004) with  $\mathbf{C}^T$  is performed on all analysis perturbations to center them on the analysis as indicated by Eq. (13) of Wei et al. (2008), 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  $\gamma$  that is designed as

$$\gamma = \begin{cases} \frac{\text{mask}}{\text{pertb}}, & \text{if } \text{mask} < \text{pertb} \\ 1, & \text{if } \text{mask} \geq \text{pertb} \end{cases}. \quad (3)$$

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. In other words, the mask only suppresses the amplitude of perturbations in areas where the analysis uncertainties are relatively low, and not locally increases the amplitude.

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 500-hPa level obtained from a 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 (Magnusson et al. 2009) computed from 80-member EnKF analysis:

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

where  $u_i'$ ,  $v_i'$ , and  $T_i'$  ( $i = 1, 2, \dots, 80$ ) are the deviations corresponding to the  $i$ th EnKF member for the wind components and temperature. The quantity  $\kappa = c_p/T_r$  equals approximately  $4.0 \text{ J kg}^{-1} \text{ K}^{-2}$ , in which  $c_p$  is the special heat at constant pressure and  $T_r$  is the reference temperature. For the purpose of representing typical

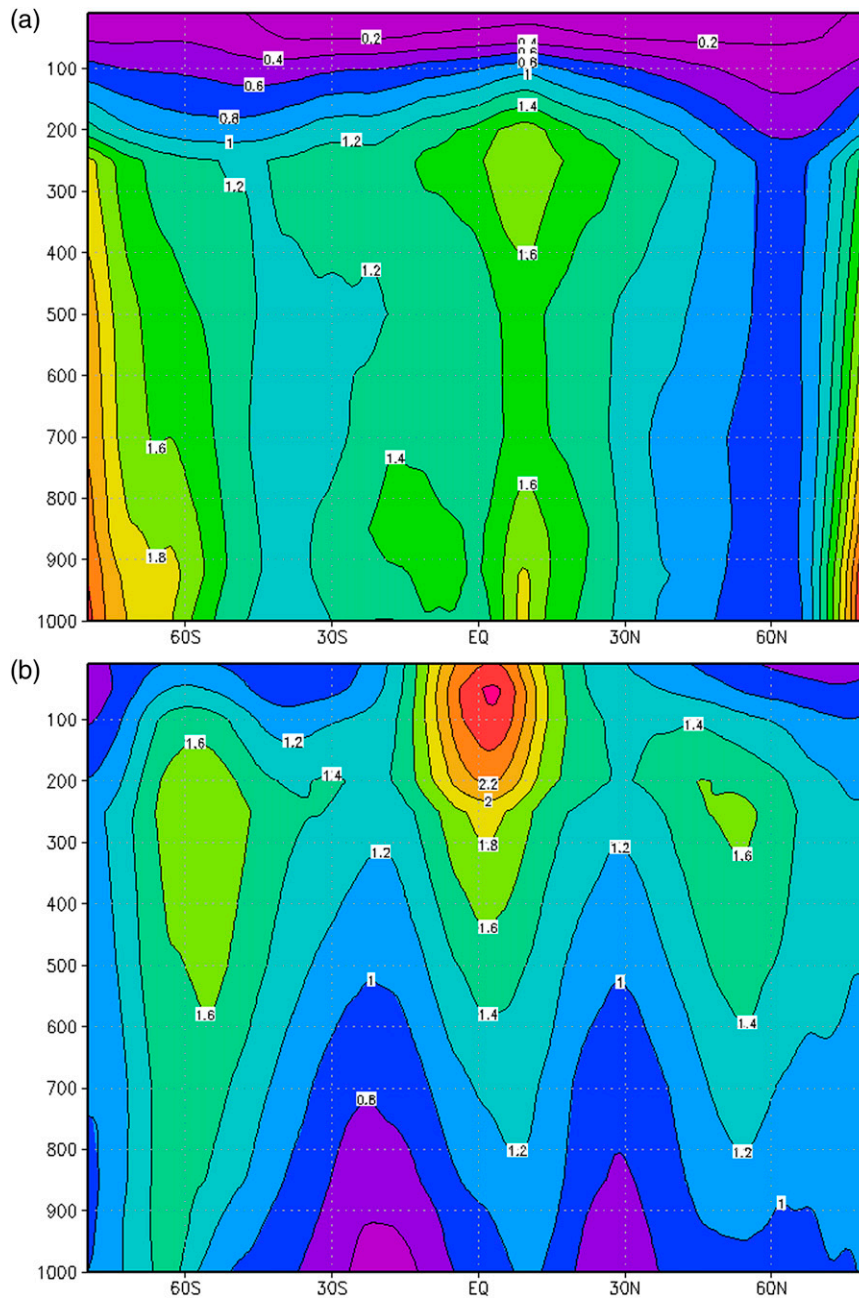


FIG. 7. The zonal average of the square root of total energy ( $\text{m s}^{-1}$ ) of initial perturbations for the (a) ETR and (b) ET\_3DR experiments as an average for the period 11 Sep–30 Nov 2012.

large-scale components in reference to most recent available set of analysis uncertainties, the decaying average method (Cui et al. 2012) is employed to accumulate the mask, given by

$$\text{TE}_{\text{ave}}(t) = (1 - w) \text{TE}_{\text{ave}}(t - 1) + w \text{TE}(t). \quad (5)$$

Here, the averaged mask  $\text{TE}_{\text{ave}}(t)$  is updated by the prior-period-averaged mask  $\text{TE}_{\text{ave}}(t - 1)$  and the most

recent  $\text{TE}(t)$  with the weight coefficient  $w$ . Sensitivity tests with different  $w$  (1%, 2%, and 5%, respectively) (not shown) indicate that the ensemble performance is not sensitive to the weight. The 2% weight is used in this study. To preserve most of the dynamical balance in the perturbations, the mask is smoothed horizontally with the Laguerre filter (King and Paraskevopoulos 1977), which is a smoothing filter based on Laguerre polynomials. The smoothing is controlled by a scaling parameter which



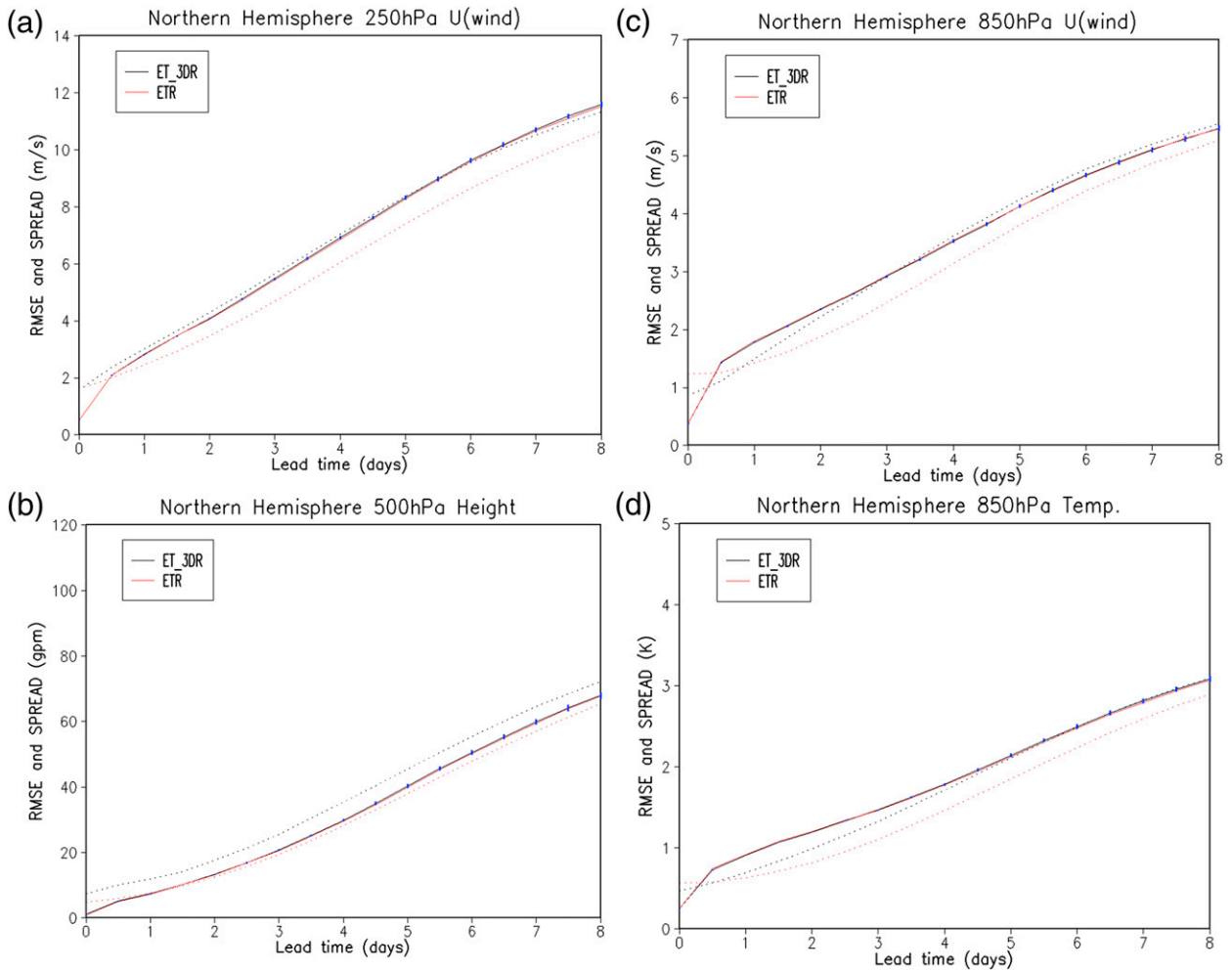


FIG. 8. The ensemble mean RMSE (solid) and ensemble spread (dashed) for (a) U250, (b) Z500, (c) U850, and (d) T850 over the NH for the period 11 Sep–30 Nov 2012. The vertical bars represent the [0.025, 0.975] confidence interval from a paired block bootstrap.

is proportional to the characteristic scale of the filter. In this study,  $\pi/40$  is chosen as the scaling parameter based on several experiments.

Figures 1a,b show the horizontal distribution of the 2D and 3D masks at 500 hPa 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. The analysis error should be not only associated with the observational 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 extratropics, the 3D mask has areas of maximum amplitude 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. This result may solve an existing problem in which the old method of rescaling cannot reduce

the amplitudes enough at 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 global average vertical distribution of the 2D and 3D masks over the same period as in Fig. 1. In the 3D mask, the amplitude increases with altitude, and then decreases after reaching a maximum between 300 and 100 hPa. This vertical structure cannot be represented with the 2D mask.

## 2) THE ENKF\_3DR METHOD

The following steps are performed to initialize the ensemble with the EnKF\_3DR method. First, the EnKF method (Whitaker and Hamill 2002) is used to generate an ensemble analysis. In this study, the 80-member 6-h EnKF forecasts from the previous cycle obtained from the NCEP hybrid 3D-Var/EnKF data assimilation system are used as the EnKF initial conditions, because operationally the EnKF analyses valid for this cycle will not have been

generated yet. Next, the 80-member ensemble set of deviations and their mean are generated. Finally, the ET\_3DR method as described in section 2a(1) is applied onto the 80 EnKF perturbations to generate 80 EnKF\_3DR perturbations.

### b. 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 and 42  $\sigma$ - $p$  hybrid vertical levels. The analysis is truncated from the T574L64 analysis provided by the NCEP GDAS. The ETR initial conditions are obtained from the operational GEFS. The EnKF initial conditions are 6-h forecasts from the previous cycle obtained from the operational hybrid 3D-Var/EnKF data assimilation system, again because in the operation the EnKF analyses valid for this cycle have not been generated yet. The methods used to generate the ET\_3DR and EnKF\_3DR ensemble initial perturbations are described in section 2a. The perturbations are updated every 6 h for the 80 members and only 20 members are chosen for medium-range forecasts due to limited computational resources. The simplex transformation is imposed on these 20 perturbations again to ensure they are centered around the analysis. 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 (0000 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 50- and 500-hPa geopotential height (Z50 and Z500); 850-hPa temperature (T850); and the 250-hPa, 850-hPa, and 10-m  $u$  components of wind (U250, U850, and U10m) over the extratropics of the Northern Hemisphere (NH; 20°–80°N), the extratropics of the Southern Hemisphere (SH; 20°–80°S), and the tropics (TR; 20°S–20°N).

## 3. ETR versus ET\_3DR

### a. Initial perturbation distribution

Figure 3 shows the vertical profile of the square root of the 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 a maximum also exists for ETR initial perturbations at 250 hPa, which is slightly smaller than in the ET\_3DR (left panel of Fig. 3a). The left panels of Figs. 3b–d show that the ET\_3DR grows

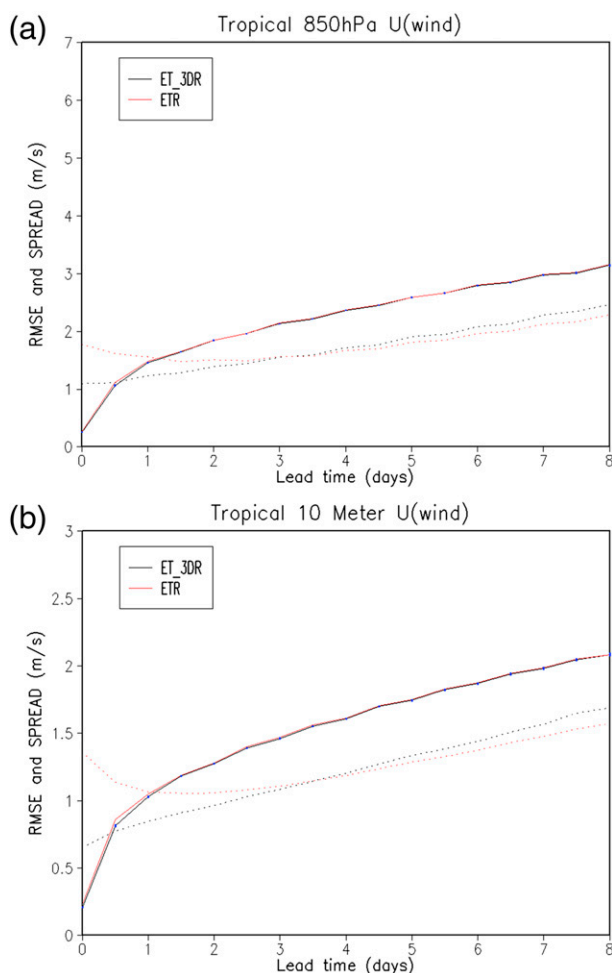


FIG. 9. The ensemble mean RMSE (solid) and ensemble spread (dashed) for (a) U850 and (b) U10m over the TR for the period 11 Sep–30 Nov 2012. The vertical bars represent the [0.025, 0.975] confidence interval from a paired block bootstrap.

faster than the ETR. After 12 h, the amplitude of ET\_3DR perturbations is close to that of the ETR below 700 hPa, and the difference becomes larger with time above 700 hPa. After 48 h, the perturbations of the ET\_3DR are larger than the ETR for all levels. Over the SH (middle panels of Figs. 3a–d), the results are similar to those in the NH. The growth rate over the TR (right panels of Figs. 3a–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 200 hPa (100 hPa) and their growth rates are comparable. At the lower levels, the amplitude of the initial perturbations is much larger than for the ET\_3DR, but their amplitudes become similar by 96 h. The fast growth of the ET\_3DR perturbations is probably because the 3D mask calculated from the EnKF analysis allows the perturbations to grow more in unstable regions than the 2D mask. To

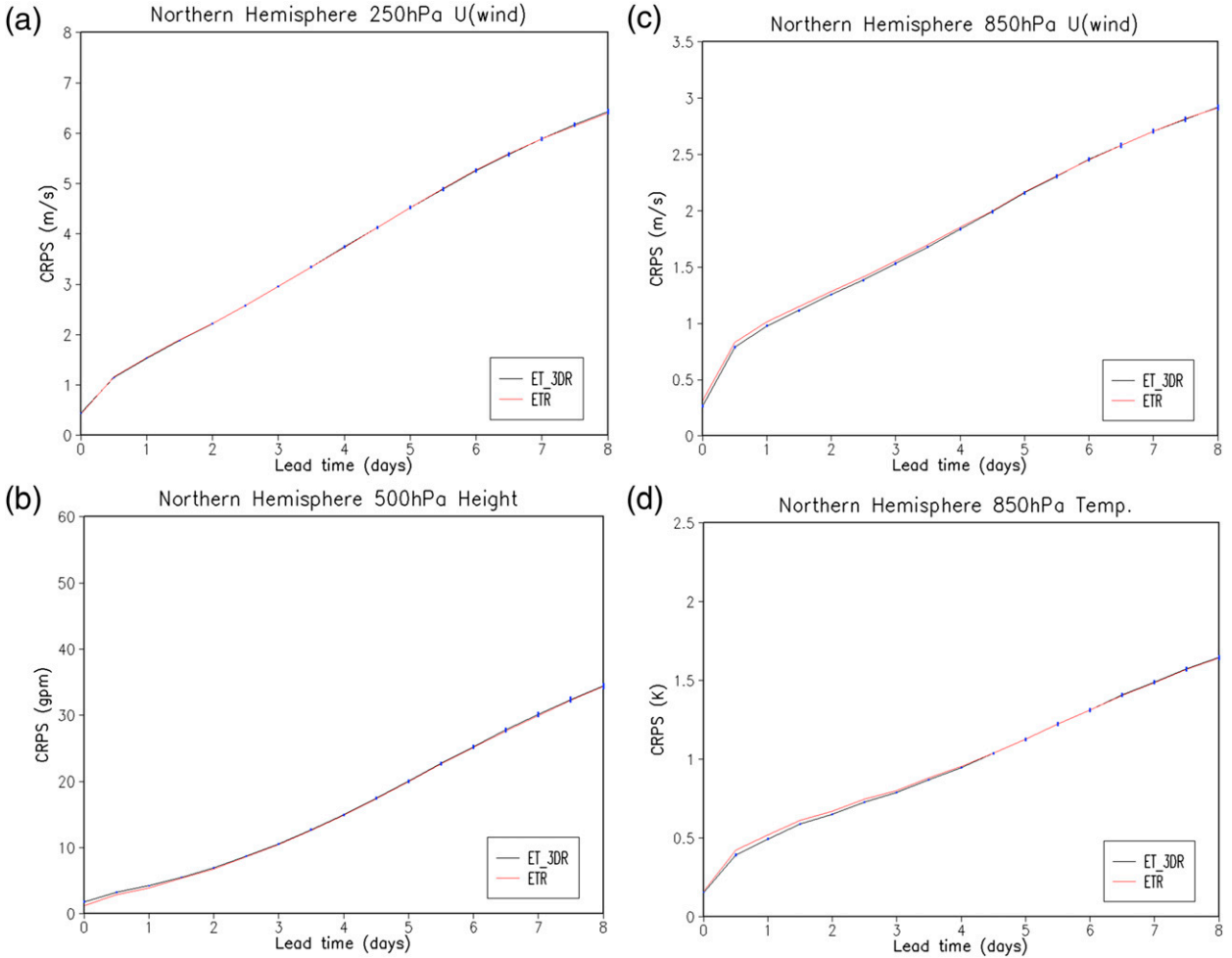


FIG. 10. The CRPS for (a) U250, (b) Z500, (c) U850, and (d) T850 over the NH for the period 11 Sep–30 Nov 2012. The vertical bars represent the [0.025, 0.975] confidence interval from a paired block bootstrap.

further illustrate the details of the initial perturbations, the horizontal and vertical distributions will be analyzed below.

Figure 4 shows the horizontal distribution of the square root of total energy of initial perturbations on the 500-hPa level for the two experiments. It is found that the distributions of initial perturbations for the ETR and ET\_3DR experiments are similar to their respective masks as shown in Fig. 1, which indicates that the regional rescaling masks have great impact on the resulting initial perturbations.

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. 5 (shading indicates correlations that are statistically significant at the 95% confidence interval). The Eady index, which is a simple measure of the most unstable baroclinic mode in a continuously stratified atmosphere, is defined as (Hoskins and Valdes 1990)

$$\sigma_E = 0.31 \frac{f}{N} \frac{du}{dz}, \tag{6}$$

where  $f$  is the Coriolis parameter,  $N$  is the static stability, and  $u$  is the magnitude of the vector wind. Here,  $N = (gd \ln \theta / dz)^{1/2}$  and  $du/dz$  are computed using the 300- and 1000-hPa potential temperature and wind from the NCEP Final Analyses. Over the NH (Figs. 5a,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. Notably, the correlation coefficients are higher than 0.6 and 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. 5c,d), the correlations in the two experiments are low compared to those over the NH. These results agree well with Magnusson et al. (2009).

Over the TR, deep convection has an important role in 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), a common surrogate for the intensity of tropical convection, is plotted in Fig. 6. Low values of OLR represent intense tropical convection. For the ET\_3DR (Figs. 4b and 6), the locations of the maxima of initial perturbations accurately coincide with zones of intense deep convection (low OLR) except for the maximum over the eastern Pacific Ocean. This connection cannot be detected at all in the ETR experiment (Figs. 4a and 6). These flow-dependent ET\_3DR initial perturbations will be beneficial for obtaining a sufficiently dispersed ensemble for use in medium-range forecasting.

Figure 7 shows the zonal average of the square root of total energy of initial perturbations. The distributions agree well with the respective masks for the ETR and ET\_3DR experiments. Since the 2D mask is vertically constant, the initial perturbations for the ETR are of a barotropic nature (Fig. 7a). Below the 200-hPa level, the minima of initial perturbations in both hemispheres are around 60°N and 40°S, respectively. Over the TR, there are two maxima at 10°N, located at 300 and 950 hPa. Above the 200-hPa level, the perturbations decrease with height around the globe. For the ET\_3DR (Fig. 7b), the large amplitudes of initial perturbations correspond to the locations of the atmospheric instability due to the 3D mask applied. The maxima are around 55°N and 55°S at 300 hPa, which correspond with the subtropical jet regions. Over the TR, the maximum occurs around 10°N at 100 hPa, near the tropical easterly jet region.

### b. Ensemble forecast skill

The verification methods used to evaluate the ensemble forecast skills for the ETR and ET\_3DR experiments include RMSE of the ensemble mean and the continuous

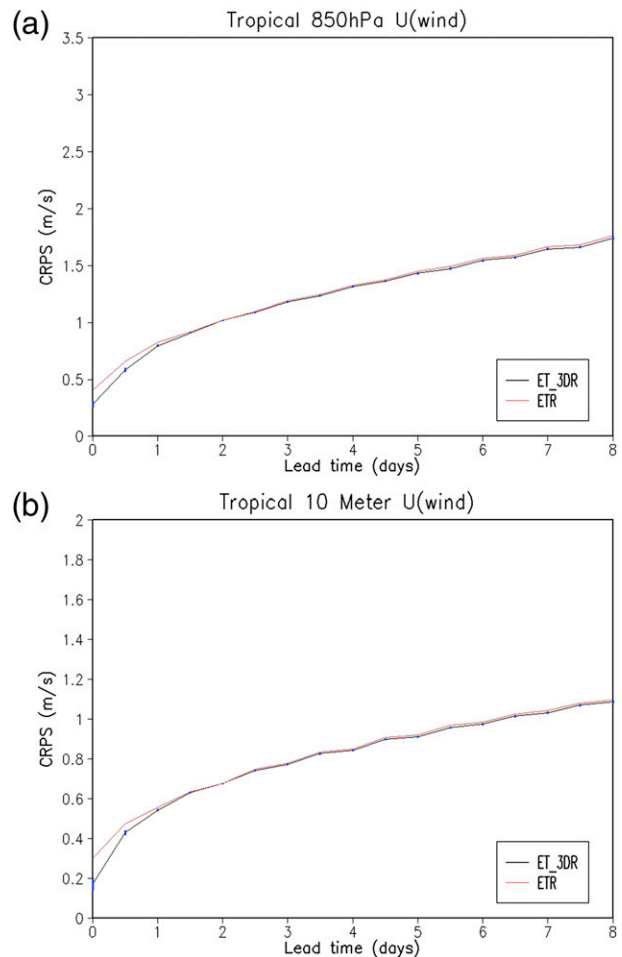


FIG. 11. The CRPS for (a) U850 and (b) U10m over the TR for the period 11 Sep–30 Nov 2012. The vertical bars represent the [0.025, 0.975] confidence interval from a paired block bootstrap.

ranked probability score (CRPS; Hersbach 2000). The paired block bootstrap algorithm is used to estimate the statistical significance of differences in scores. More details are available in Hamill (1999). In this study, the [0.025, 0.975]

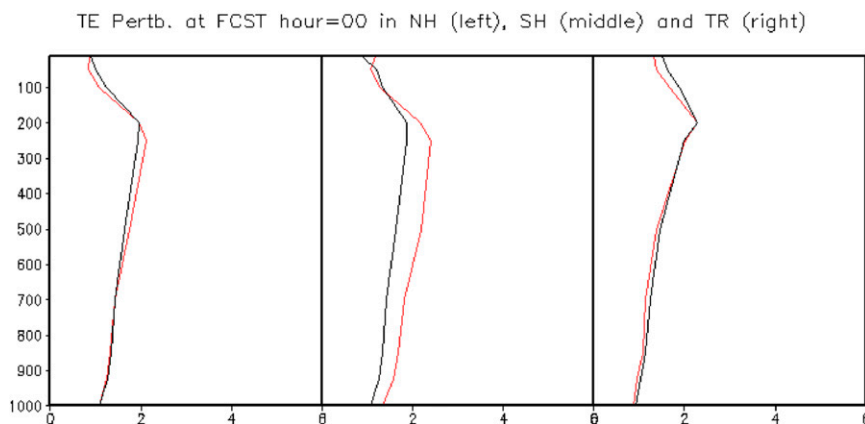


FIG. 12. The vertical profile of the square root of total energy ( $\text{m s}^{-1}$ ) of perturbations at the initial time for the EnKF (red) and EnKF\_3DR (black) experiments as an average for the period 11 Sep–30 Nov 2012: (left to right) NH, SH, and TR.

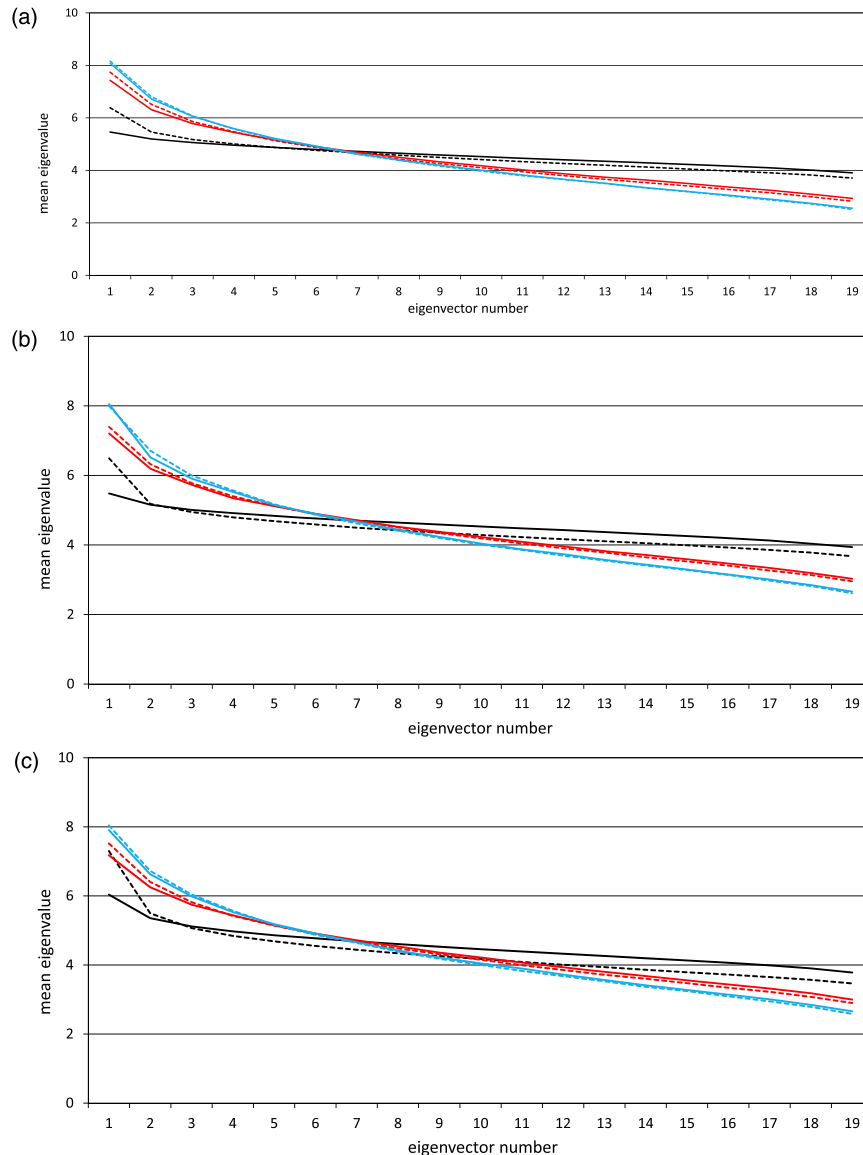


FIG. 13. The mean eigenvalue spectra of the covariance matrix for the square root of total energy ( $\text{m s}^{-1}$ ) of perturbations at the initial time (black), after 48 h (red), and 96 h (blue) for the ETR (dashed) and ET\_3DR (solid) experiments at the (a) 250-, (b) 500-, and (c) 850-hPa levels during the period 11 Sep–30 Nov 2012 over the globe.

confidence interval is computed from a bootstrap resampling from the 81 cases using 1000 random samples.

It is worth mentioning that the results of verification against the analysis can be affected by the analysis uncertainties. Especially, the analysis fields of winds in the tropics have large uncertainties due to a lack of reliable observations and because the forecast model is unable to capture convective scale variability. However, we could ignore this issue for long forecast lead time when the analysis uncertainty is relatively smaller compared to the forecast error itself.

### 1) RMSE AND ENSEMBLE SPREAD

Figures 8a–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 spread grows faster than that the ETR and maintains consistency with

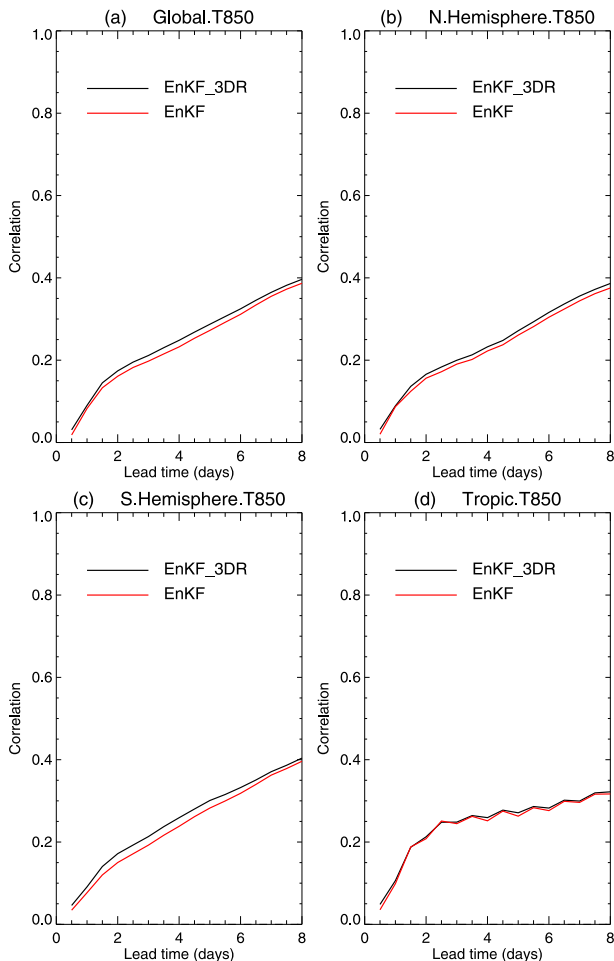


FIG. 14. PECA value for T850 over the (a) globe, (b) NH, (c) SH, and (d) TR as an average for the period 11 Sep–30 Nov 2012.

the RMSE for all lead times just as a perfect ensemble forecast system should (Fig. 8a). For the indirect model variable Z500, Fig. 8b shows that the ET\_3DR starts with 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 after 24 h this discrepancy is reduced and the spread approaches the RMSE gradually (Figs. 8c,d). Over the SH (not shown), the results are similar to those for the NH. Additionally, at Z50 (not shown) the two methods' ensemble mean RMSEs are close to each other and the ensemble spread for the ET\_3DR is larger than that for the ETR, but these measures grow in parallel. Because the amount of observations in the stratosphere is overwhelmingly less than in the troposphere, the analysis depends more on the forecast. For short lead times, the ET\_3DR produces much larger spread than the RMSE. With

increasing forecast length, the spread and RMSE become close. But the spread for the ETR is much smaller than its RMSE.

Figures 9a,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 days. The spread of ET\_3DR has a higher growth rate than the ETR especially for the first 2 days. As found in section 3a, this is attributed to the close connection between the ET\_3DR initial perturbations and the deep tropical convection. 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 physical processes, whereas those over the NH and SH are mainly influenced by 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 is the higher growth rate of the spread, which is especially obvious at lower levels and over the TR. The higher growth rate of the ET\_3DR spread does not lead to reduction of the RMSE, because the ensemble mean is not sensitive to ensemble spread.

## 2) CONTINUOUS RANKED PROBABILITY SCORE

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, the better is the quality of the probabilistic forecast. Over the NH, the CRPS for U250 is similar for the two experiments (Fig. 10a). The ETR has significantly smaller score than the ET\_3DR for the first 12 h for Z500 (Fig. 10b) probably due to the overdispersion as shown in Fig. 8b. 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 a statistically significantly better probabilistic forecast for the first 4 days than the ETR (Fig. 10c). For T850, the ET\_3DR has slightly, but statistically significantly, better performance for the first 2 days than the ETR (Fig. 10d). Over the SH (not shown), the results are generally similar to that over the NH, except that the ET\_3DR presents statistically significantly smaller scores than the ETR only for lead times up to 12 h for U850 and T850. Over the TR (Fig. 11), the ET\_3DR has statistically significantly better performance than the ETR for almost all lead times except for 1.5–2.5 days for

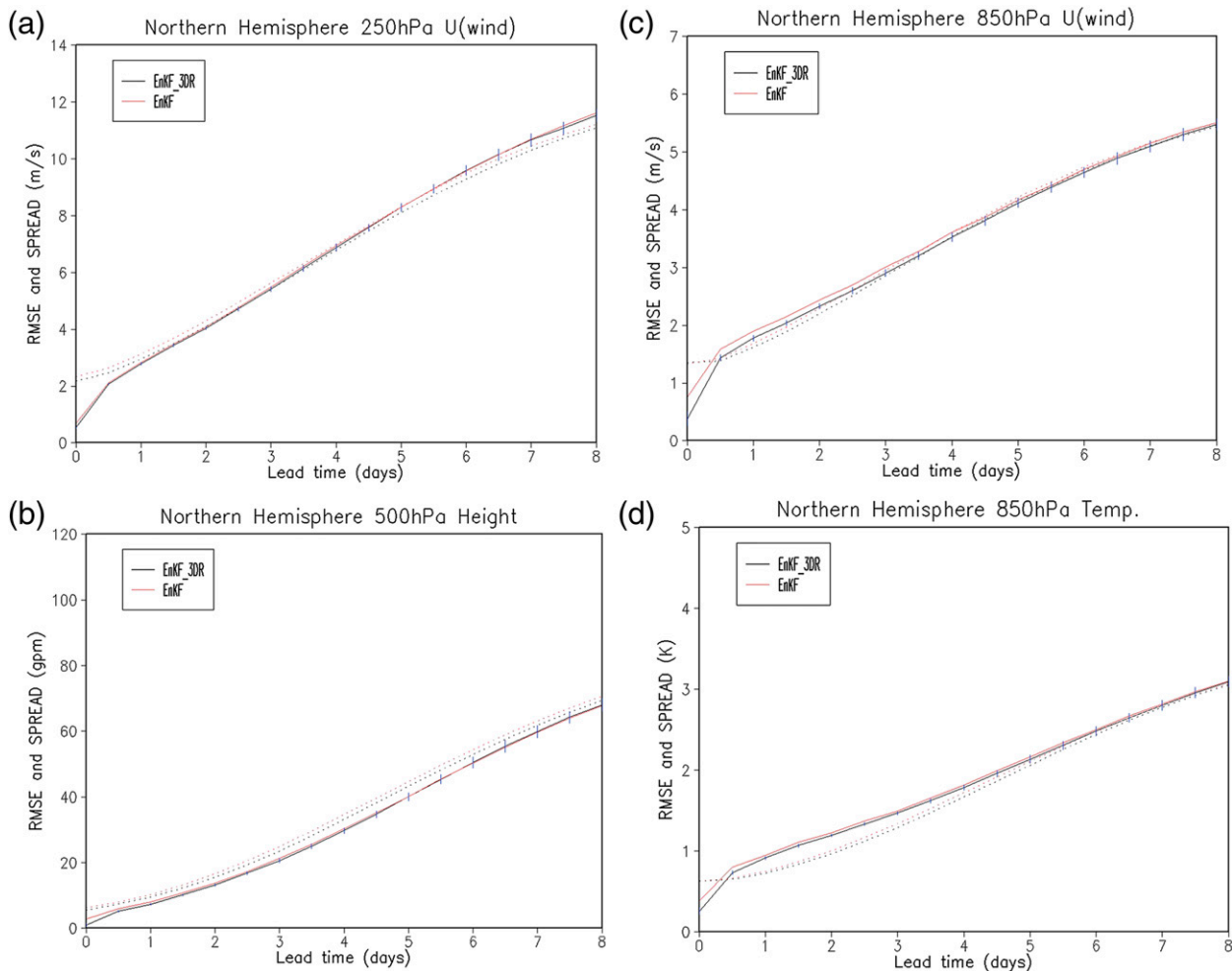


FIG. 15. The ensemble mean RMSE (solid) and ensemble spread (dashed) for (a) U250, (b) Z500, (c) U850, and (d) T850 over the NH for the period 11 Sep–30 Nov 2012. The vertical bars represent the [0.025, 0.975] confidence interval from a paired block bootstrap.

U850. For U10m, the ET\_3DR has a statistically significant advantage over the ETR for the first day. Large CRPS differences compared for short lead times are due to the spread differences in Fig. 9.

#### 4 EnKF versus EnKF\_3DR

##### a. The perturbations

Figure 12 shows the vertical profile of the square root of the total energy of initial perturbations for the EnKF and EnKF\_3DR experiments. The EnKF has larger initial amplitude than the EnKF\_3DR between 250 and 700 hPa over the NH (left panel of Fig. 12). The difference is much larger over the SH for levels below 200 hPa (middle panel of Fig. 12). The EnKF initial perturbations are slightly smaller than the EnKF\_3DR over the TR

(right panel of Fig. 12). In the data assimilation, the EnKF will not give enough weight to observations if the background error covariances are underestimated, so large perturbations of the EnKF are favorable. During the generation of the EnKF ensemble analyses, there is multiplicative inflation applied to account for unrepresented error sources (Whitaker and Hamill 2012). But large amplitude will not be beneficial for medium-range forecasting. Therefore, rescaling the EnKF with a mask may be helpful for the improvement of its performance.

Ensemble perturbations should span as many unstable dimensions of the atmospheric state as possible with a limited number of ensemble members. The eigenvalue spectra of the covariance matrix of the perturbations can be used to evaluate the distribution of the perturbation magnitudes across independent directions (Wang and

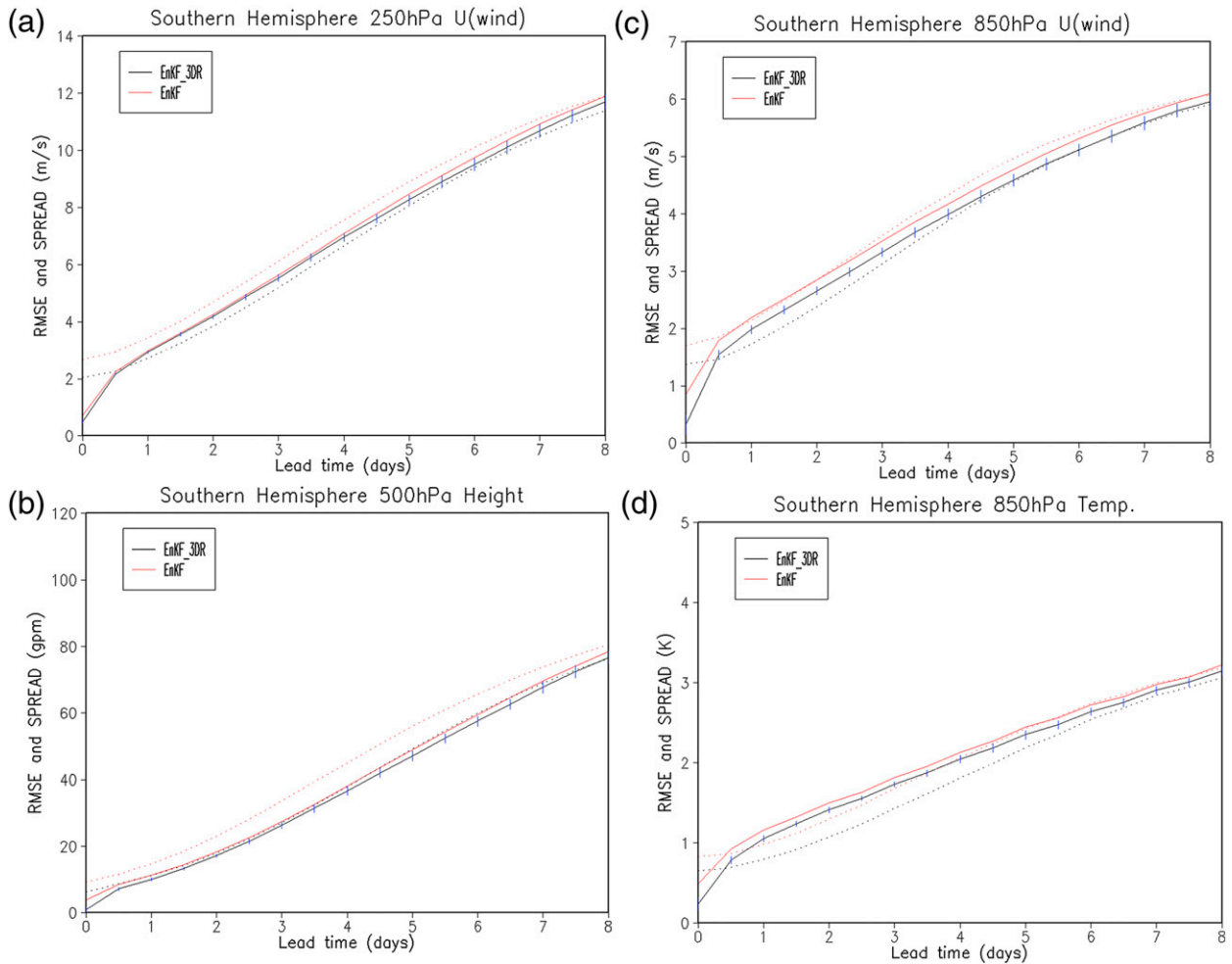


FIG. 16. As in Fig. 15, but over the SH.

Bishop 2003; Magnusson et al. 2008). Figure 13 shows the mean eigenvalue spectra for the square root of total energy of perturbations at different lead times and model levels during the period 11 September–30 November 2012 over the globe. It is found that the initial perturbations of the EnKF are overly accounted for by the direction of the first mode. The EnKF\_3DR has a flatter eigenvalue spectrum than the EnKF due to the orthonormalization by the ET. The result implies that the EnKF\_3DR members are more independent than the EnKF, which may have a potentially positive impact on the ensemble performance. With increasing forecast lead, the difference between the methods gradually decreases. It is also found that the differences at 250 (Fig. 13a) and 500 hPa (Fig. 13b) are slightly smaller than that at 850 hPa (Fig. 13c).

Perturbation versus error correlation analysis (PECA) values for T850 are shown in Fig. 14 and used

to evaluate the quality of ensemble perturbations by measuring how well ensemble perturbations can explain forecast error variance (Wei and Toth 2003). The results indicate that the EnKF\_3DR improves PECA values in all domains for all lead times, especially over the SH.

#### b. 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 3b.

##### 1) RMSE AND ENSEMBLE SPREAD

In Fig. 15, the RMSE and ensemble spreads 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. 15a,b), but the difference



is not statistically significant for U250 and only significant for the first day for Z500. Results for U850 and T850 (Figs. 15c,d) show that the EnKF\_3DR has significantly smaller RMSE than the EnKF for the first 3.5 days. 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 days for U250 and at all lead times for Z500 (Figs. 15a,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 increasing of the forecast lead time (Figs. 15c,d).

Over the SH, the RMSE in the EnKF\_3DR is slightly smaller than the EnKF for U250, but not significant (Fig. 16a). The EnKF\_3DR has significantly smaller RMSE than the EnKF out as far as 6.5-day lead time for Z500 (Fig. 16b). For U850 and T850, the EnKF\_3DR produces smaller RMSE than the EnKF, and the difference is statistically significant for all lead times (Figs. 16c,d). The spreads for U250 and Z500 in the EnKF\_3DR experiment are more consistent with the RMSEs, while in the EnKF experiment there is overdispersion (Figs. 16a,b). Similar to the result for the NH, the spread for U850 and T850 grows slower than the RMSE during the first 3–4 days, and then becomes almost equal to the RMSE with increasing of forecast lead.

Results for the TR (Fig. 17) show that U850 and U10m for both experiments appear to produce much lower spread compared to the RMSEs due to the undersampling of the model related errors. The spread for the EnKF\_3DR grows slightly slower than the EnKF, but the RMSE is still substantially smaller than for the EnKF.

## 2) 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. 18a and 19a). For Z500, the EnKF\_3DR produces a slightly better probabilistic forecast than the EnKF over the NH, and the difference is significant for up to 2 days (Fig. 18b). Over the SH, the improvement becomes more apparent by 6.5 days (Fig. 19b). For U850 and T850, the EnKF\_3DR has substantially better performance than the EnKF for both hemispheres. The difference is statistically significant until 4 days over the NH, and for all lead times over the SH (Figs. 18c and 19c). The results indicate that the large amplitudes degrade the performance of the EnKF and that applying ET\_3DR on the EnKF has a positive impact. Over the TR, for all lead

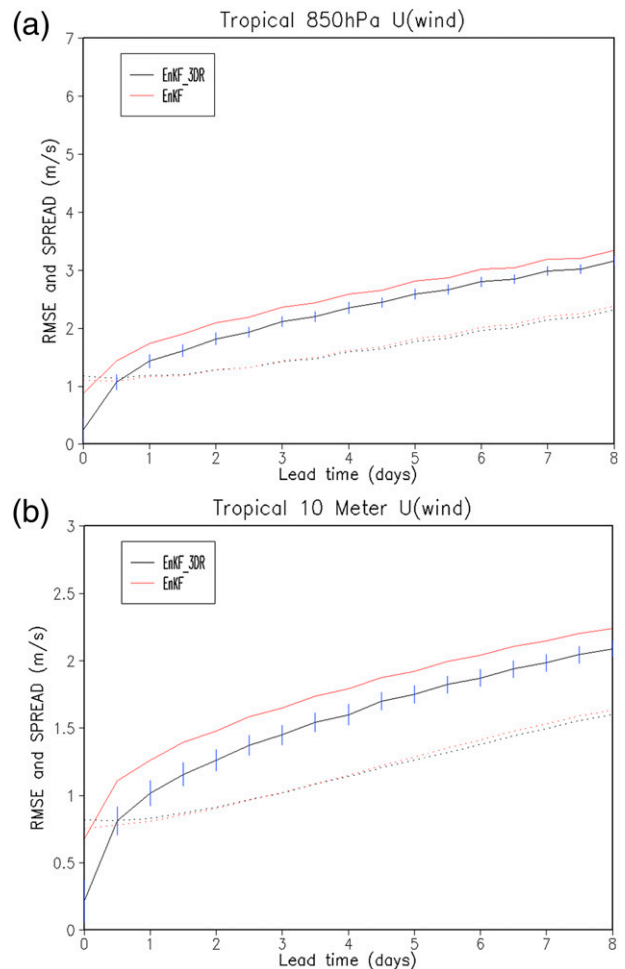


FIG. 17. The ensemble mean RMSE (solid) and ensemble spread (dashed) for (a) U850 and (b) U10 m over the TR for the period 11 Sep–30 Nov 2012. The vertical bars represent the [0.025, 0.975] confidence interval from a paired block bootstrap.

times, the EnKF\_3DR shows significantly better scores for both U850 and U10m (Fig. 20).

## 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 variations of analysis error. While the ETR used in the NCEP GEFS has improved the spread and probabilistic skill of the ensemble forecasts over both the BV and ET methods, its mask has several limitations, which we attempt to address in this study. There are three main modifications to the mask presented herein. First and foremost, for representing the vertical structure of analysis error, a 3D mask is employed instead of the original 2D mask. This is the most advantageous improvement of the ET\_3DR compared to the ETR. In the

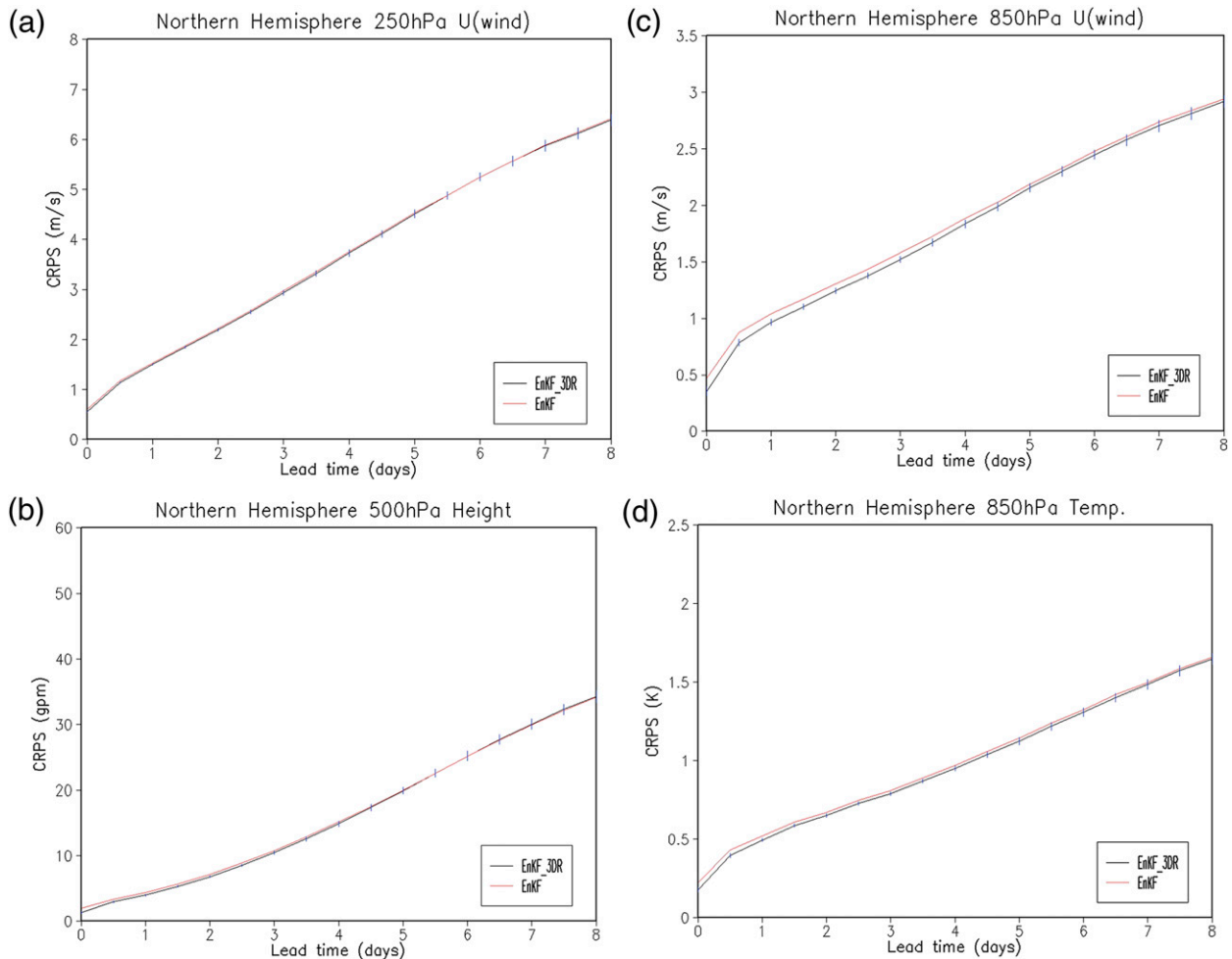


FIG. 18. The CRPS for (a) U250, (b) Z500, (c) U850, and (d) T850 over the NH for the period 11 Sep–30 Nov 2012. The vertical bars represent the [0.025, 0.975] confidence interval from a paired block bootstrap.

ETR method, due to the vertically constant mask used, additional inflation has been applied to the mask for levels from the model bottom to 500 hPa 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 observational density distribution. This new analysis error variance replaces the static analysis error variance. Third, a total energy norm is used instead of a kinetic 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 500 hPa 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 correlate better with areas of baroclinic instability over the NH and areas of deep convection over the TR, which is beneficial for obtaining a sufficiently dispersed ensemble in the medium range. The variations of the vertical distribution of the ETR perturbations are small due to the vertically constant mask, while the maxima of vertical distribution for the ET\_3DR perturbations correspond to the subtropical jet region and tropical easterly jet region. Since the amplitude of the initial perturbations for the ET\_3DR is more consistent with the typical locations of atmospheric instability, the spread grows much faster than for the ETR, especially at the lower levels and over

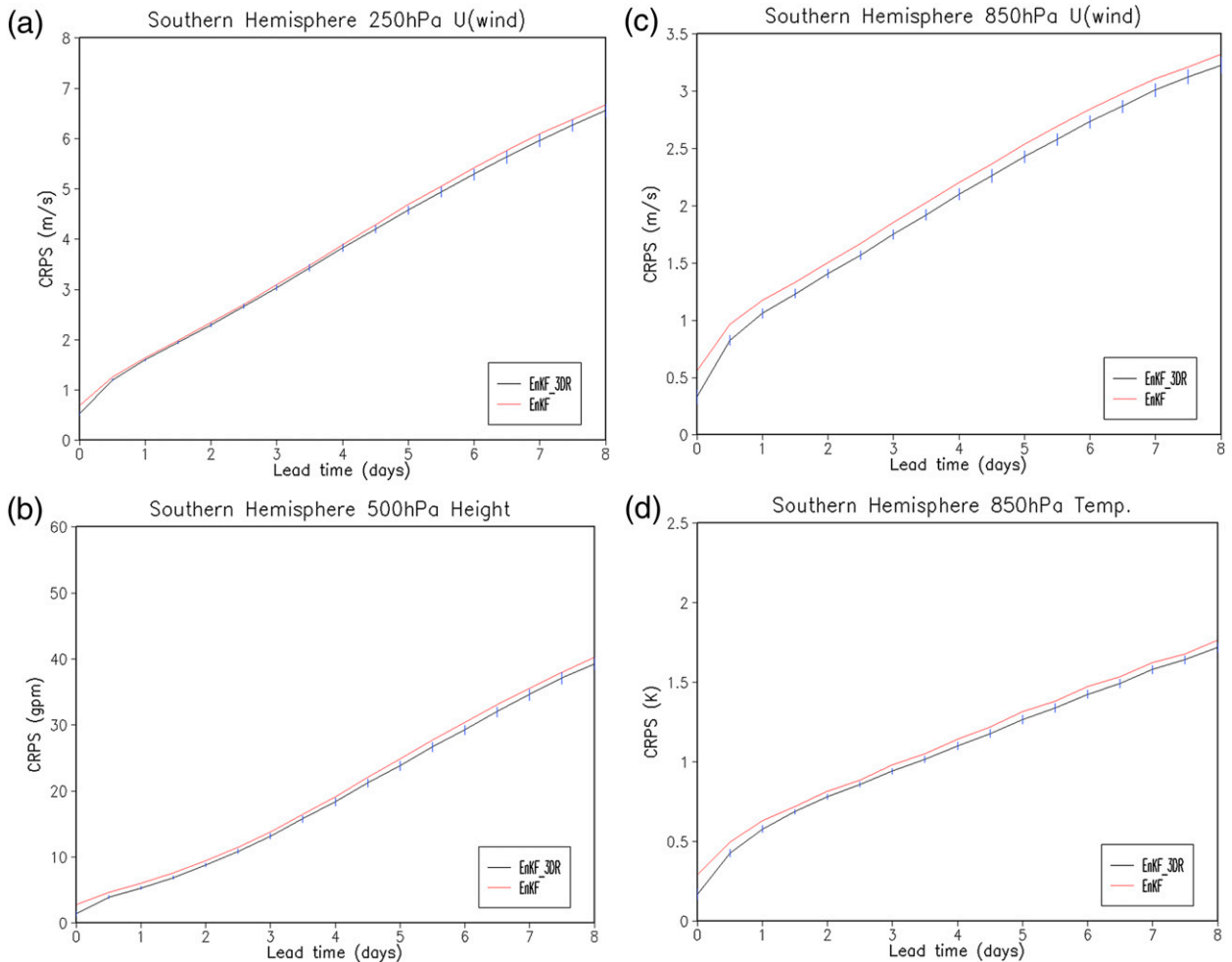


FIG. 19. As in Fig. 18, but over the SH.

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 and constrain the amplitude of initial perturbations without additional computing cost, with the availability of the EnKF analyses from the NCEP GDAS, the EnKF\_3DR method is designed in this study by applying ET\_3DR on EnKF perturbations. 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. Furthermore, the EnKF\_3DR perturbations can better explain forecast error variance measured by the PECA values. 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 to be the potential candidate for the NCEP operational GEFS initial perturbation method in the next implementation. However, from the results of this study, we 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 of using this strategy during other seasons. Furthermore, because of the merits of the ETR method, it may be also considered for application across multicenter sets of analyses within a consensus modeling ensemble framework.

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

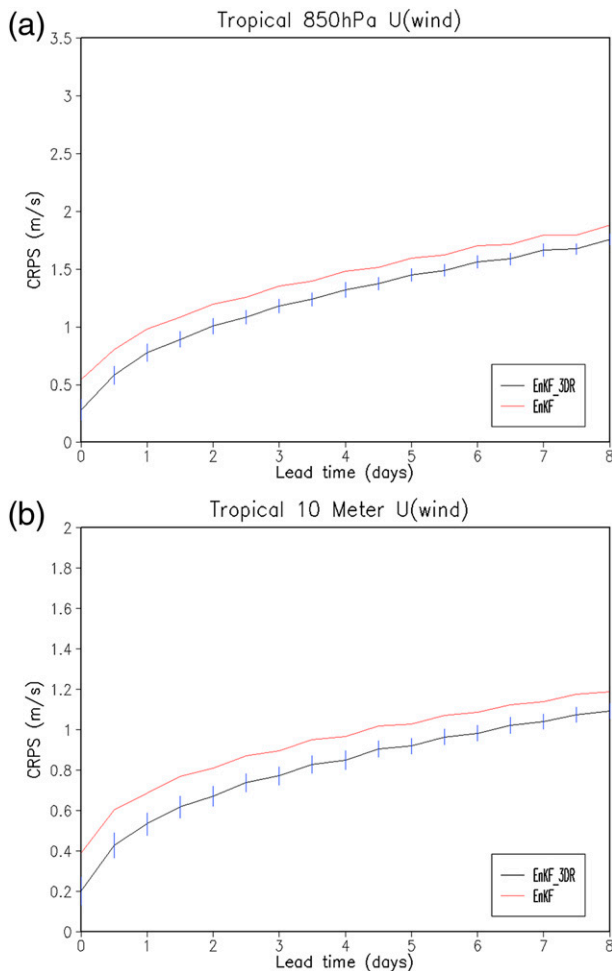


FIG. 20. The CRPS for (a) U850 and (b) U10m over the TR for the period 11 Sep–30 Nov 2012. The vertical bars represent the [0.025, 0.975] confidence interval from a paired block bootstrap.

**Acknowledgments.** The authors thank members of the Ensemble and Post Processing Team at EMC/NCEP for helpful suggestions during the course of this work. Thanks to Daryl Kleist for providing the data and code to read the EnKF ensemble analysis. The first author gratefully acknowledges the support of EMC and LASW State Key Laboratory Special Fund. We thank the three anonymous reviewers for their helpful comments and suggestions.

#### REFERENCES

- Annan, J. D., 2004: On the orthogonality of bred vectors. *Mon. Wea. Rev.*, **132**, 843–849, doi:10.1175/1520-0493(2004)132<0843:OTOOBV>2.0.CO;2.
- Bishop, C. H., and Z. Toth, 1999: Ensemble transformation and adaptive observations. *J. Atmos. Sci.*, **56**, 1748–1765, doi:10.1175/1520-0469(1999)056<1748:ETAAO>2.0.CO;2.
- 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, doi:10.1175/2009MWR3157.1.
- , —, —, —, 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.*, **138**, 1567–158, doi:10.1175/2009MWR3158.1.
- 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, doi:10.1175/MWR2905.1.
- , 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, doi:10.1002/qj.346.
- , —, —, and J. Haseler, 2010: Combined use of EDA- and SV-based perturbations in the EPS. *ECMWF Newsletter*, No. 123, ECMWF, Reading, United Kingdom, 22–28.
- Cui, B., Z. Toth, Y. Zhu, and D. Hou, 2012: Bias correction for global ensemble forecast. *Wea. Forecasting*, **27**, 396–410, doi:10.1175/WAF-D-11-00011.1.
- Hamill, T. M., 1999: Hypothesis tests for evaluation numerical precipitation forecasts. *Wea. Forecasting*, **14**, 155–167, doi:10.1175/1520-0434(1999)014<0155:HTFENP>2.0.CO;2.
- , 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, doi:10.1175//2559.1.
- Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting*, **15**, 559–570, doi:10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.
- Hoskins, B. J., and P. J. Valdes, 1990: On the existence of storm-tracks. *J. Atmos. Sci.*, **47**, 1854–1864, doi:10.1175/1520-0469(1990)047<1854:OTEOST>2.0.CO;2.
- Hou, D., Z. Toth, and Y. Zhu, 2006: A stochastic parameterization scheme within NCEP global ensemble forecast system. *18th Conf. on Probability and Statistics*, Atlanta, GA, Amer. Meteor. Soc., 4.5. [Available online at [https://ams.confex.com/ams/Annual2006/techprogram/paper\\_101401.htm](https://ams.confex.com/ams/Annual2006/techprogram/paper_101401.htm).]
- , —, —, and W. Yang, 2008: Impact of a stochastic perturbation scheme on global ensemble forecast. *19th Conf. on Probability and Statistics*, New Orleans, LA, Amer. Meteor. Soc., 1.1. [Available online at [https://ams.confex.com/ams/88Annual/techprogram/paper\\_134165.htm](https://ams.confex.com/ams/88Annual/techprogram/paper_134165.htm).]
- King, R. E., and P. N. Paraskevopoulos, 1977: Digital Laguerre filters. *Int. J. Circuit Theory Appl.*, **5**, 81–91, doi:10.1002/cta.4490050108.
- 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, doi:10.1007/s00376-012-1249-y.
- Magnusson, L., M. Leutbecher, and E. Källén, 2008: Comparison between singular vectors and breeding vectors as initial perturbations for the ECMWF Ensemble Prediction System. *Mon. Wea. Rev.*, **136**, 4092–4104, doi:10.1175/2008MWR2498.1.
- , J. Nycander, and E. Källén, 2009: Flow-dependent versus flow-independent initial perturbations for ensemble prediction. *Tellus*, **61A**, 194–209, doi:10.1111/j.1600-0870.2008.00385.x.
- McLay, J. G., C. H. Bishop, and C. A. Reynolds, 2008: Evaluation of the ensemble transform analysis perturbation scheme at NRL. *Mon. Wea. Rev.*, **136**, 1093–1108, doi:10.1175/2007MWR2010.1.

- , —, and —, 2010: A local formulation of the ensemble transform (ET) analysis perturbation scheme. *Wea. Forecasting*, **25**, 985–993, doi:[10.1175/2010WAF2222359.1](https://doi.org/10.1175/2010WAF2222359.1).
- Palmer, T. N., R. Gelaro, J. Barkmeijer, and R. Buizza, 1998: Singular vectors, metrics, and adaptive observations. *J. Atmos. Sci.*, **55**, 633–653, doi:[10.1175/1520-0469\(1998\)055<0633:SVMAAO>2.0.CO;2](https://doi.org/10.1175/1520-0469(1998)055<0633:SVMAAO>2.0.CO;2).
- 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, doi:[10.1175/1520-0493\(2002\)130<1125:TEOIHR>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<1125:TEOIHR>2.0.CO;2).
- Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330, doi:[10.1175/1520-0477\(1993\)074<2317:EFANTG>2.0.CO;2](https://doi.org/10.1175/1520-0477(1993)074<2317:EFANTG>2.0.CO;2).
- , and —, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319, doi:[10.1175/1520-0493\(1997\)125<3297:EFANAT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<3297:EFANAT>2.0.CO;2).
- 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, doi:[10.1175/1520-0469\(2003\)060<1140:ACOBAE>2.0.CO;2](https://doi.org/10.1175/1520-0469(2003)060<1140:ACOBAE>2.0.CO;2).
- , —, and S. J. Julier, 2004: Which is better, an ensemble of positive–negative pairs or a centered spherical simplex ensemble? *Mon. Wea. Rev.*, **132**, 1590–1605, doi:[10.1175/1520-0493\(2004\)132<1590:WIBAEO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1590:WIBAEO>2.0.CO;2).
- , 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.*, **141**, 4898–4117, doi:[10.1175/MWR-D-12-00141.1](https://doi.org/10.1175/MWR-D-12-00141.1).
- Wei, M., and Z. Toth, 2003: A new measure of ensemble performance: Perturbation versus error correlation analysis (PECA). *Mon. Wea. Rev.*, **131**, 1549–1565, doi:[10.1175/1520-0493\(2003\)131<1549:ANMOEP>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<1549:ANMOEP>2.0.CO;2).
- , —, 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, doi:[10.1111/j.1600-0870.2006.00159.x](https://doi.org/10.1111/j.1600-0870.2006.00159.x).
- , —, —, and —, 2008: Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system. *Tellus*, **60A**, 62–79, doi:[10.1111/j.1600-0870.2007.00273.x](https://doi.org/10.1111/j.1600-0870.2007.00273.x).
- Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924, doi:[10.1175/1520-0493\(2002\)130<1913:EDAWPO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2).
- , and —, 2012: Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Wea. Rev.*, **140**, 3078–3089, doi:[10.1175/MWR-D-11-00276.1](https://doi.org/10.1175/MWR-D-11-00276.1).