Characteristics of Initial Perturbations in the Ensemble Prediction System of the Korea Meteorological Administration

JUN KYUNG KAY AND HYUN MEE KIM

Atmospheric Predictability and Data Assimilation Laboratory, Department of Atmospheric Sciences, Yonsei University, Seoul, South Korea

(Manuscript received 7 August 2013, in final form 20 January 2014)

ABSTRACT

In this study, the initial ensemble perturbation characteristics of the new Korea Meteorological Administration (KMA) ensemble prediction system (EPS), a version of the Met Office Global and Regional Ensemble Prediction System, were analyzed over two periods: from 1 June to 31 August 2011, and from 1 December 2011 to 29 February 2012. The KMA EPS generated the initial perturbations using the ensemble transform Kalman filter (ETKF). The observation effect was reflected in both the transform matrix and the inflation factor of the ETKF; it reduced (increased) uncertainties in the initial perturbations in regions with dense observations via the transform matrix (inflation factor). The reduction in uncertainties is generally governed by the transform matrix but locally modulated by the inflation factor. The sea ice significantly affects the initial perturbations near the lower boundary layer. The large perturbation energy in the lower stratosphere of the tropics was related to the dominant zonal wind, whereas the perturbation energy in the upper stratosphere of the winter hemispheres was related to the dominant polar night jet. In the early time-integration stage, the initial perturbations decayed in the lower troposphere but grew rapidly in the mid- to upper troposphere. In the meridional direction, the initial perturbations grew greatest in the northern polar region and smallest in the tropics. The initial perturbations maintained a hydrostatic balance, especially during the summer in both hemispheres and during both the summer and winter in the tropics, associated with the smallest growth rates of the initial perturbations. The initial perturbations in the KMA EPS appropriately describe the uncertainties associated with atmospheric features.

1. Introduction

Because the atmosphere is a chaotic system, initial condition errors and imperfect numerical forecast models limit predictions of future states of the atmosphere. If information regarding the distribution of forecast errors is provided, then the imperfections of deterministic predictions can be mitigated by obtaining more information on the probabilistic forecast (Palmer 1999). Ensemble prediction (EP) estimates forecast uncertainties by integrating multiple initial conditions sampled from the error probability of the initial state. To predict the uncertainties associated with the forecast under the perfect model assumption, initial conditions for EP should estimate the probability density function (PDF) associated with the error of the initial state (Ehrendorfer 1994).

Forecast uncertainties depend on the growth of the initial state errors that are differences between initial ensemble members and their mean field; therefore, initial ensemble perturbations should have realistic growth rates that reflect the dynamic property of error growth in the atmosphere (Palmer 1993; Magnusson et al. 2008).

The numerical weather prediction (NWP) centers have developed and implemented various initial perturbation generation methods for EP: the total energy norm singular vector (TESV) methods (Buizza and Palmer 1995; Molteni et al. 1996) of the European Centre for Medium-Range Weather Forecasts (ECMWF), the breeding method (Toth and Kalnay 1993) of the National Centers for Environmental Prediction (NCEP), the ensemble transform (ET) technique (McLay et al. 2008; Wei et al. 2008) that is a new method of the NCEP, and the ensemble transform Kalman filter (ETKF) method (Bowler et al. 2008, 2009) of the Met Office (UKMO). The TESV method generates the perturbations with the greatest linear growth over the optimization time interval (Buizza et al. 2005; Kim and Morgan 2002) and the breeding
method calculates the leading Lyapunov vectors of model solutions associated with the error of the greatest nonlinear growth (Toth and Kalnay 1993). Both the TESV and breeding method consider the dynamical growth of the initial perturbations, but may not draw the initial ensembles from the PDF of the analysis error.

The ETKF method is based on the ensemble square root filter (EnSRF) data assimilation scheme (Tippett et al. 2003). Using the information of forecast and observational uncertainty, the ETKF generates initial ensemble perturbations based on Kalman filter (KF) analysis error covariance statistics. Because the calculations associated with the ETKF can be performed rapidly, this method effectively generates ensemble initial conditions. To estimate the analysis and forecast error covariance appropriately using a finite number of ensembles, each ensemble member should be maintained independently (i.e., orthogonally) so that the finite number of ensembles could represent most of the analysis and forecast uncertainties. A concentrated variance to a specific subset of ensemble members among the entire set of ensemble members cannot represent forecast uncertainties appropriately. Therefore, the even distribution of the variance of the initial ensemble perturbations to the orthogonal ensemble directions is a desirable property when generating the initial ensemble perturbations. Wang and Bishop (2003) first implemented the ETKF without localization as an ensemble generation method and compared this method with a masked-breeding method using a global climate model [Community Climate Model (CCM3)] with a fixed artificial observation network. They found that the variance of the initial ensemble perturbations from the ETKF reflected the spatial variability of observational density and that this variance was more evenly distributed along the orthogonal ensemble direction for the ETKF method than for the masked-breeding method. Wei et al. (2006) applied the methodology of Wang and Bishop (2003) to the NCEP operational ensemble prediction system (EPS) using their operational observations. In contrast to Wang and Bishop (2003), Wei et al. (2006) showed that the effect of the real observations on the initial ensemble variance was not large on the global scale due to small ensemble sizes (i.e., 10 members).

Miyoshi and Yamane (2007) used a local ETKF approach (LETKF; an ETKF method with localization as a data assimilation scheme) in an AGCM model [the Atmospheric General Circulation Model for the Earth Simulator (AFES)] and performed observing system simulation experiments (OSSEs) using a real observation network; they showed that the initial ensemble spread from the LETKF was able to represent the analysis error. Bowler et al. (2008) described the overall characteristics and performance of the Met Office Global and Regional Ensemble Prediction System (MOGREPS) and demonstrated that the horizontal structure of the initial ensemble spread from the ETKF was related to the distributions of dynamically active areas and observational densities in the midtroposphere. Bowler et al. (2009) adopted localization of the ETKF in MOGREPS and verified that the localization made the latitudinal distributions of the initial ensemble spread appropriately reflect the ensemble-mean error distribution. Nevertheless, most ETKF studies have focused on the general features of the initial perturbations, but have not discussed the characteristics of the ETKF components composing the initial perturbations in detail.

Since 2011, the Korea Meteorological Administration (KMA) has operationally implemented the Unified Model (UM) and its related pre-/postprocessing systems, imported from the Met Office (Kay et al. 2013). The new EPS of the KMA is based on MOGREPS and uses ETKF with localization as the initial ensemble perturbation generation method. To verify that the ETKF method generates operational ensembles appropriately, it is necessary to understand the characteristics of the initial ensemble perturbations over the globe for statistically significant time periods. Therefore, the present study analyzes the global characteristics of the ETKF components and ETKF-based initial ensemble perturbations in the KMA EPS from 1 June to 31 August 2011 (hereafter JJA) and from 1 December 2011 to 29 February 2012 (hereafter DJF). In particular, the role of each ETKF component in the generation of initial ensemble perturbations is investigated, and the ability of the initial ensemble perturbations to represent the initial state uncertainties is examined from various dynamic viewpoints.

Section 2 presents the details of the system configuration, the theoretical description of the ETKF, and the experimental framework. Section 3 describes the characteristics of the transform matrix and the inflation factor of the ETKF. Section 4 presents the structure and property of the initial ensemble perturbations. Section 5 provides a summary and discussion.

2. Methodology

a. The KMA EPS

The KMA EPS used in the present study is fundamentally the same as the original MOGREPS. MOGREPS consists of a global and regional ensemble system (Bowler et al. 2008). The global ensemble system used in the present study has a horizontal resolution of approximately 40 km and 70 vertical levels (N320L70) and 24 members (23 perturbed forecasts and 1 unperturbed control forecast). The local ETKF generates initial
perturbations (see Bowler et al. 2009). The 23 initial perturbations were added to the analysis derived from the four-dimensional variational data assimilation (4DVAR) system of the KMA UM; the sum of the analysis and the perturbations are then integrated forward for 10 days using the KMA UM with the same resolution used in the KMA EPS. The global ensemble run is performed twice daily (0000 and 1200 UTC). MOGREPS considered model uncertainties using stochastic-physics schemes that consist of “random parameters” and “stochastic convective vorticity” schemes (Bowler et al. 2008). The random parameters scheme (Bright and Mullen 2002) introduces uncertainties into the empirical parameters of the physical parameterization schemes. The stochastic convective vorticity scheme (Gray and Shutts 2002) addresses a potential vorticity (PV) anomaly dipole similar to the one typically associated with a mesoscale convective system (MCS). The horizontal components of wind \(\nu'\) and \(\nu''\), potential temperature \(\theta'\), Exner pressure \(\pi'\), and specific humidity \(q'\) perturbations were variables used in MOGREPS.

b. ETKF formulation

ETKF is a type of ensemble square root filter that estimates forecast and analysis error covariance with ensemble perturbations of size \(N\) based on the optimal Kalman filter theory. If the size of the ensemble perturbations is \(N\), then the forecast and analysis perturbation vectors \(z'_f = x'_f - x_f\) and \(z'_a = x'_a - x_a\) \((i = 1, 2, \ldots, N)\) are used to compose the matrix:

\[
Z' = \frac{1}{\sqrt{N-1}} (z'_1, z'_2, \ldots, z'_N) \quad \text{and} \quad Z'' = \frac{1}{\sqrt{N-1}} (z'_1, z'_2, \ldots, z'_N),
\]

where \(x'_f\) and \(x'_a\) are the \(i\)th ensemble forecast and analysis fields, respectively; \(\bar{x}'\) and \(\bar{x}''\) are the ensemble-mean forecast and analysis from \(N\) members, respectively; and the overbar represents the expectation operator. The forecast and analysis covariance matrices are expressed as

\[
P'_f = (Z'_f)^T \quad \text{and} \quad P'' = (Z'')^T,
\]

where \(T\) indicates the matrix transpose. The analysis ensemble perturbation matrix \(Z''\) is calculated by solving the Kalman filter update equation:

\[
P'' = P'_f - P'_f H (HP'_f H^T + R)^{-1} H P'_f,
\]

where \(R\) is the observational error covariance matrix that includes the errors of representativeness and observation operator. In this system, \(R\) is defined as a diagonal matrix and \(H\) is a linearized form of the observation operator \(H\).

If the analysis ensemble perturbations are derived by transforming the forecast ensemble perturbations as

\[
Z'' = Z' T,
\]

then the analysis error covariance matrix in Eq. (2) can be rewritten as \(P'' = Z' T (Z' T)^T\). By substituting this analysis error covariance matrix into Eq. (3), the transform matrix \(T\) is derived, as described by Wang et al. (2004):

\[
T = C (\Gamma + I)^{-1/2} C^T,
\]

where \(\Gamma\) is an \((N - 1) \times (N - 1)\) diagonal matrix containing all of the eigenvalues of \(R^{-1/2}HZ'Z'H^{-1/2}\).

\[
C = \text{eigenvector matrix of Eq. (6)}, \quad \text{and} \quad I \text{ is the identity matrix.}
\]

In MOGREPS and the KMA EPS, \(H(x') - H(x'')\) is used instead of \(HZ'\) in Eq. (6). The observational information matrix \(R\) and the form \(H\) are used in Eq. (6) to incorporate the observation effects.

If the ensemble size is smaller than the degrees of freedom of the model state, then the analysis error covariance is underestimated because the forecast error covariance is not fully estimated by the ensemble members. The inflation factor \(\Pi\) is applied to the initial ensemble perturbations in Eq. (4) to solve

\[
\Pi_n = \Pi_{n-1} \sqrt{\frac{\text{trace}(d_n d_n^T) - \text{trace}(R)}{\text{trace}(HP_n H^T)}},
\]

where \(n\) represents the time steps and \(d_n = y_n - H(x'_n)\) is an innovation vector that represents the difference between observations \(y\) and the 12-h ensemble-mean forecast in observational space verified at the time of observation. The inflation factor ensures that the sum of the spreads of the ensemble 12-h forecast matches the sum of the variations in the error of the ensemble-mean 12-h forecast in the observational space over the verification region (Bowler et al. 2008, 2009; Kay et al. 2013).

The two perturbation matrices \(Z'\) and \(Z''\) are square roots of the forecast and analysis covariance matrices and have the same rank as their covariance matrices. The initial ensemble perturbations are orthogonal to each other in the observational space normalized by the square root of the observation error (Wang and Bishop...
If the ensemble mean becomes the best estimate of the true state (i.e., analysis), the sum of the ensemble perturbations should be zero (Wang et al. 2004; Wei et al. 2006; McLay et al. 2008). By applying the spherical simplex method (Wang et al. 2004) that postmultiplies the initial analysis perturbations by the transpose of \(C\) [see Eq. (5)], the initial ensemble perturbations become unbiased and the sum of the initial ensemble perturbations becomes zero.

The ensemble-based data assimilation system uses a covariance localization to decrease the long-range spurious correlations due to the limited ensemble size (Kay et al. 2013). The basic idea of covariance localization is to reduce the elements of the forecast error covariance to remove the correlation between distant grid points (Houtekamer and Mitchell 2001; Hamill et al. 2001). The aforementioned covariance localization cannot be implemented in ETKF because the forecast error covariance is not explicitly calculated. Rather, the horizontal localization of ETKF in MOGREPS is realized by dividing the globe into 92 centers of approximately equal distance; the ETKF is then calculated using the observations taken from a specific radius of influence of each center to improve the likelihood that the spread of EP is represented as a function of latitude (Bowler et al. 2009). The KMA EPS does not currently apply a vertical localization.

c. Experimental design

The characteristics of the initial ensemble perturbations of the KMA EPS were investigated for JJA and DJF. The observations used to calculate the initial ensemble perturbations in the ETKF included sonde, surface observation, aircraft, satellite wind (Satwind), Advanced Television Infrared Observation Satellite (TIROS) Operational Vertical Sounder (ATOVS), scatterometer wind (Scatwind), Atmospheric Infrared Sounder (AIRS), GPS radio occultation (GPS RO), and Infrared Atmospheric Sounding Interferometer (IASI), all of which are conventional observations collected for the KMA operational data assimilation system (Table 1). The observational information corresponding to the observations in Table 1 is used in Eq. (6) to calculate the transform matrix. The number of observations used is shown in Fig. 1. The inflation factors in Eq. (7) are calculated using sonde and ATOVS observations with 12-h ensemble forecasts because the error estimates of these observations are relatively easily obtained in the KMA data assimilation system (Kay et al. 2013). The inflation factors were calculated vertically for four layers: the lower troposphere (from the surface to 2 km), the mid- to upper troposphere (from approximately 2 to 15 km), the stratosphere (from approximately 16 to 56 km), and the mesosphere (from approximately 62 km to the top of the model (76 km)).

The detailed method of calculating the inflation factor is described in Kay et al. (2013).

3. ETKF component characteristics

a. The transform matrix and the inflation factor

When the initial ensemble perturbations are generated from Eqs. (4), (5), and (7), the 12-h ensemble forecast perturbations are rotated by the eigenvector matrix of the forecast covariance matrix \(C\) and rescaled by \((\Gamma + I)^{-1/2}\). Therefore, the initial ensemble perturbations normalized by the square root of \(R\) have the same direction as the eigenvector of the 12-h forecast covariance matrix in observational space (Wang and Bishop 2003). As mentioned in section 1, even distribution of the variance of the initial ensemble perturbations to the orthogonal ensemble directions (i.e., flatter eigenvalue spectrum) is one of the desirable properties when generating the initial ensemble perturbations. As the eigenvalue spectrum of the forecast covariance matrix is flatter, the forecast ensemble is equally rescaled in the eigenvector direction by \((\Gamma + I)^{-1/2}\). Thus, the eigenvalue spectrum of the forecast covariance matrix provides information regarding how well the initial ensemble perturbations include equally filtered information in the direction of the eigenvectors.

Figure 2a shows the global eigenvalue spectrum of the forecast covariance matrix; in other words, all observations on the globe are used to calculate the transform matrix. Figure 2b shows the average of the 92 eigenvalue spectra calculated from all localization centers. The eigenvalue spectra in Figs. 2a and 2b are normalized by their leading eigenvalues. The flatter slope of the eigenvalue spectra implies that a large number of eigenvectors is necessary to explain the overall forecast uncertainties. In contrast, the steeper slope of the spectra implies that a relatively smaller number of eigenvectors can explain much of the forecast uncertainties (Szunyogh et al. 2007; Kim and Jung 2009). The eigenvalue spectra of JJA are slightly flatter than those of DJF in the global domain (Fig. 2a), which implies that the forecast uncertainties are governed by a large number of atmospheric systems during JJA compared with DJF. The slopes of the eigenvalue spectra of JJA and DJF do not significantly differ for localized domains (Fig. 2b). The average of the 92 eigenvalue spectra has a steeper slope than the global eigenvalue spectra; thus, the average variance in the different eigenvector directions of the forecast covariance vector decreases via localization. The localization in the covariance matrix increases the rank of the forecast error covariance (Hamill et al. 2000; Oke et al. 2007), and averaged eigenvalue spectra over increasing ranks show
increasing flatness because the increased rank of the forecast error covariance helps the ensembles to get more effective degrees of freedom to fit the observations (Ehrendorfer 2007). However, the localization removes and consequently ignores the information associated with the trailing modes of the eigenvalue spectrum (Hacker et al. 2007). Although the localization of this system improves the relationship between the ensemble spread and the forecast error of the ensemble mean as a function of latitude (Bowler et al. 2009), localization produces eigenvalues with less variability than the global eigenvalues.

To investigate the spatial distribution of the variance of the uncertainty spanned by the initial ensemble perturbations following Patil et al. (2001), the ensemble dimension at each localization center was calculated as

$$
\Psi(\sigma_1, \sigma_2, \ldots, \sigma_{N-1}) = \frac{\left(\sum_{i=1}^{N-1} \sigma_i^2\right)^2}{\sum_{i=1}^{N-1} \sigma_i^4},
$$

where $\sigma_i$ is the $i$th singular value of the 12-h forecast covariance matrix in the observational space normalized by the square root of $\mathbf{R}$. The largest value of the ensemble dimension is $N - 1$, which is possible when the uncertainty variance is evenly distributed across all ensemble perturbations (Oczkowski et al. 2005).

Figure 3 shows the spatial distribution of the ensemble dimensions. For both JJA and DJF, these dimensions are large over East Asia and America in the tropical region (TR) and greater in the extratropical Northern Hemisphere (NH) compared with the extratropical Southern Hemisphere (SH). On average, continents have greater ensemble dimensions than oceans over the same latitudes. This distribution of the ensemble dimension is similar to that in Kuhl et al. (2007), which investigated the ensemble dimension using the NCEP global model and LETKF. Two factors determine the ensemble dimension distributions. First, because small ensemble dimensions are related to large forecast error growth in local regions (Oczkowski et al. 2005; Szunyogh et al. 2005; Kuhl et al. 2007), large ensemble dimensions in the TR are caused by small forecast error growth during JJA and DJF (see Reynolds et al. 1994; Straus and Paolino 2009). The large ensemble dimension in the TR

<table>
<thead>
<tr>
<th>Observation type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sonde</td>
<td>Radiosonde observations (on land, ship, and mobile platforms), dropsonde observations, and wind profiler observations</td>
</tr>
<tr>
<td>Surface</td>
<td>Surface-based observations at or near the earth’s surface</td>
</tr>
<tr>
<td>Aircraft</td>
<td>Aircraft-based observations reported by the Aircraft Meteorological Data Reporting (AMDR) system and aircraft reports (AIREPs)</td>
</tr>
<tr>
<td>Satwind</td>
<td>Atmospheric wind observations derived from successive geostationary satellite observations</td>
</tr>
<tr>
<td>ATOVS</td>
<td>Advanced Television Infrared Observation Satellite (TIROS) Operational Vertical Sounder system observations</td>
</tr>
<tr>
<td>Scatwind</td>
<td>Sea wind observations obtained from the Quick Scatterometer (QuikSCAT)</td>
</tr>
<tr>
<td>AIRS</td>
<td>Atmospheric Infrared Sounder observations</td>
</tr>
<tr>
<td>GPS RO</td>
<td>Global positioning system radio occultation observations</td>
</tr>
<tr>
<td>IASI</td>
<td>Infrared Atmospheric Sounding Interferometer observations</td>
</tr>
</tbody>
</table>

FIG. 1. Observation numbers (contour) used to calculate the transform matrix at each local center averaged for (a) JJA and (b) DJF. The observation locations are interpolated into model grids.
indicates that the ensemble perturbations are highly independent; as a result, the variability of the uncertainty along different directions is maintained. In contrast to the forecast error in the midlatitude, the forecast error in the TR is not explained by several major systems, but associated with many small systems of complicated dynamics. Because of this feature, the ensemble dimensions are relatively large in the TR. Therefore, increasing the ensemble size might result in more accurate estimations of the forecast error structure in the TR. Second, because the transform matrix in Eq. (5) is calculated in the regions where the observations are made, the observation density affects the ensemble dimension. A high density of observations over the NH continents (Fig. 1) leads to smaller forecast errors in the NH; therefore, the ensemble dimension in the NH is generally greater than that in the SH for both JJA and DJF. In contrast, the ensemble dimension over the ocean is relatively small, which is associated with large forecast errors due to sparse observations. The ensemble dimensions in the northern polar (NP) region differ from those in the southern polar (SP) region. In the NP region, the ensemble dimensions are large during JJA but small during DJF. The seasonal change of the ensemble dimension in the NP region is associated with a small error growth (i.e., a large ensemble dimension) in the summer and a large error growth (i.e., a small ensemble dimension) in the winter due to baroclinic instability. Therefore, the ensemble dimension in the NP region depends on the atmospheric dynamic features. However, these seasonal changes are not observed in the SP region because this region has fewer observations (Fig. 1); as a result, the observational density exerts more influence on the seasonal change of the ensemble dimension than the dynamic atmospheric features in the SP region.

The inflation factor in Eq. (7) increases the spread of the initial ensemble perturbations to match their mean error. Figure 4 shows the average inflation factors applied to the troposphere. The distribution of the sonde and ATOVS observations used to calculate the inflation factor is similar to the distribution of the inflation factor itself. The regions of dense observations have greater inflation factors because more observations are used to calculate the inflation factor. In contrast, the regions over the SH and the ocean have smaller inflation factors.

b. The combined effect of the transform matrix and the inflation factor on the initial ensemble perturbations

Observational information reduces the forecast uncertainty in the ETKF method by applying the transform matrix to the forecast ensemble perturbations in the Kalman filter update equation. To investigate the observation effect on the initial ensemble perturbations, the ratio of the spread of the initial ensemble perturbations with regard to that of the 12-h ensemble forecasts was calculated (Fig. 5). This ratio effectively identifies the spatial distributions of rescaling factors (i.e., the transform matrix) that reduce the forecast spread to the initial ensemble spread using the observational information and the inflation factor (Wang and Bishop 2003; Wei et al. 2006). According to Wang and Bishop (2003), the ETKF, the optimal data assimilation scheme, attenuates the forecast uncertainty represented by the forecast ensemble spread to generate the initial ensemble spread via observational information; therefore, the ratio is expected to be less than 1. Wang and Bishop (2003) also showed that this ratio is lower in regions with dense observations compared with those with sparse observations due to the filtering effect of the ETKF (hereafter called the transform effect). In contrast, the value of
the inflation factor increases in regions with dense observations. Therefore, the ratio of the spread of the initial ensemble perturbations with regard to the spread of the 12-h ensemble forecast in Fig. 5 shows the combined effect of both the transform matrix and the inflation factor.

The ratio is greater in the TR than at higher latitudes and is generally higher in the SH than in the NH because there are more observations in the NH (Fig. 5), which implies that the transform effect governs the general features of the ratio in the latitudinal direction. The ratios do not change significantly with model variables; however, they do change with seasons. In the NH midlatitudes, the ratio over the continents is relatively greater than that over the ocean (except in North America during DJF). This finding implies that the inflation factor effect predominates in these regions. The ratios exhibit less seasonal variation in the SH compared with the NH, and they are relatively greater over land regions with a large amount of satellite observation data (cf. Fig. 5 with Fig. 4). This finding implies that the inflation factor effect also predominates over the land in the SH. Therefore, the transform effect governs the general features of the ratio along the latitudinal direction, but the inflation factor effect governs the local features of the ratio along the longitudinal direction.
which implies that the ratio generally follows the transform effect and is locally modulated by the inflation factor. The distribution of the ratios in the present study is similar to that of the ensemble transform rescaled to consider the climatologically derived analysis error variance [cf. Fig. 5 with Fig. 2a in Wei et al. (2008)], which implies that the initial ensemble perturbations of the ETKF in the KMA EPS are consistent with the climatologically derived analysis error variance.

To investigate the vertical structure change of the ensemble perturbations that result from the application of the transform matrix and the inflation factor, two areas were chosen: one is the northern midlatitude area (20°–60°N) where the average ratio is less than 1, and the other is the area in which the ratio is greater than 1 (denoted A in Fig. 5a). Figure 6 shows the vertical profiles of the initial ensemble and the 12-h forecast ensemble spreads over the northern midlatitude area and area A. In the northern midlatitude, the spread of the initial ensemble perturbations was generally smaller than that of the forecast ensemble perturbations during JJA and DJF, and distinctly smaller approximately

Fig. 4. Number of sonde and ATOVS observations (contours) used to calculate the inflation factor at each local center and the averaged inflation factors in the troposphere (shaded) for (a) JJA and (b) DJF. The observation locations and inflation factors at 92 local centers are interpolated into model grids.
below 50 and above 10 hPa (Figs. 6a,c). This finding implies that the transform effect generally dominates the northern midlatitude area. In area A above 1 hPa (i.e., the midstratosphere), the initial ensemble spread is smaller than the forecast ensemble spread (Figs. 6b,d), which implies that the inflation factor suppresses the transform effect that usually decreases forecast uncertainty throughout most of the atmosphere below 1 hPa.

Overall, the transform effect and inflation factor oppositely affect the initial uncertainties. The transform effect reduces the size of the initial perturbations, but the inflation factor increases that of the initial perturbations. The transform effect affects the general features of the initial perturbations, but the inflation factor affects the local features of the initial perturbations.

4. Initial ensemble perturbation characteristics

a. The structure of the initial ensemble perturbations

The total energy (TE) norm of the initial ensemble perturbations was calculated as described by Bannon (1995), Ehrendorfer and Errico (1997), Rabier et al. (1996), and Wang and Bishop (2003):

\[
TE = \frac{1}{2}(u'^2 + v'^2) + \frac{R}{2\kappa T_{ref}}\theta'^2 + \frac{1}{2}RT_{ref}\pi'^2,
\]

where \( u', v', \theta', \) and \( \pi' \) are variables in the KMA EPS; \( R \) is the gas constant for dry air; \( \kappa (\approx 0.286) \) is the ratio of \( R \) to \( c_p \) (the specific heat of dry air at a constant pressure); \( \gamma (\approx 1.4) \) is the ratio of \( c_p \) to \( c_v \) (the specific heat of dry air at a constant volume); \( T_{ref} \) is a reference temperature of 300 K; and \( \pi \) is taken from the analysis of the KMA UM. The first term on the right-hand side of Eq. (9) represents the kinetic energy (KE), whereas the second and third terms together represent the potential energy (PE) (Eckart 1960).

Figure 7 shows the TE of the initial ensemble perturbations averaged during JJA and DJF at several vertical layers, as well as the vertical profiles of the TE for six regions: the SP region from 90°S to 70°S, the SH from 70°S to 20°S, the TR from 20°S to 0°N, the NH from 20°N to 70°N, the NP region from 70°N to 90°N, and the entire globe. At the lower boundary layer, the TE during JJA is large along the border between Antarctica and the ocean in the SH and large over the Arctic Ocean in the NH (Fig. 7a); conversely, the TE during DJF is large over the Arctic Ocean (Fig. 7f), which implies that sea ice constitutes an important source of uncertainty regarding the initial ensemble perturbations near the surface. In the troposphere, a large amount of TE is found in the midlatitude during JJA and DJF, isolated over Africa, the eastern Pacific region, and central Asia in the NH; the area of large TE in the SH is located over the ocean and Antarctica (Figs. 7b,g). In the lower- to midstratosphere, the TE is concentrated in the TR from approximately 20 to 30 km (from 50 to 10 hPa) and forms a zonal band during JJA and DJF (Figs. 7c,h).
the altitude increases above 10 hPa, the TEs during JJA and DJF are located broadly in the winter hemispheres (i.e., SH, Fig. 7d; NH, Fig. 7i). The TE is greater during DJF (Fig. 7i) compared with JJA (Fig. 7d) near the top of the model. As the altitude increases, the TE decreases to its minimum in the stratosphere and then increases at the top of the model (Figs. 7e,j).

Figure 8 shows the vertical cross sections of the zonally averaged TE of the initial ensemble perturbations superimposed on the zonal wind of the KMA UM analysis. The initial ensemble perturbations have the greatest TE above the upper stratosphere of the winter hemisphere where the polar night jets are strong. Over

the TR, the large TE of the initial ensemble perturbations was concentrated in the lower stratosphere where the dominant zonal winds have large spatial variations and directional changes with altitude. Kawatani and Hamilton (2013) demonstrated that the descending easterly wind was predominant during JJA 2011, and this finding was confirmed by the KMA UM analysis (Fig. 8a). During DJF, the easterly wind regime was predominant in regions above 40 hPa, and the westerly regime located below the easterly regime is weaker than the easterly regime (Fig. 8b). As Figs. 7c and 7h show, the zonally averaged TE of the initial ensemble perturbations over the TR in the stratosphere is greater along

---

**FIG. 6.** The vertical profiles of the zonally averaged initial ensemble spread (solid line) and 12-h forecast ensemble spread (dashed line) for zonal wind averaged over (top) the NH (20°–60°N) and (bottom) area A during (a),(b) JJA and (c),(d) DJF.
Fig. 7. The vertically averaged TE of the initial ensemble perturbations (shaded: J kg⁻¹) at each layer during (left) JJA and (right) DJF in the (a),(f) lower boundary; (b),(g) troposphere; (c),(h) lower to midstratosphere; and (d),(i) top of the model. (e),(f) The vertical profiles for horizontally averaged TE (contour: J kg⁻¹) over the NP (gray), NH (purple), TR (blue), SH (green), and SP (red) regions, as well as the entire globe (black) for (left) JJA and (right) DJF.
the dominant easterly wind during DJF compared with JJA (cf. Fig. 8b with Fig. 8a). From the surface to the lower troposphere, the initial ensemble perturbation TE is greater in the winter polar regions, which is due to the uncertainty associated with the sea ice (see Figs. 7a,f).

Figure 9 presents the vertically and zonally averaged ensemble-mean TE, KE, and PE of the initial ensemble perturbations and the standard deviation of the TE for each vertical layer identified in section 2c. The TE, KE, and PE are vertically integrated considering the density in each layer and normalized by the sum of the density of all vertical layers at each model grid. During JJA and DJF, the magnitude of KE was greater than that of PE, which is consistent with the results of Bowler et al. (2008), who used MOGREPS without localization. In the lower troposphere, the KE and PE were large and showed large variations in the ensemble members in the NP and SP regions. The TE was especially large in the winter polar regions and relatively weak in the TR (Figs. 9a,e); Bowler et al. (2009) reported similar findings, showing underestimated ensemble spreads in the TR. The PE was relatively large in the NP and SP regions; however, the magnitude of the PE was approximately one order of magnitude lower than that of the KE.
Vertically and zonally averaged zonal and meridional winds contribute equally to the KE in all regions except the TR where the zonal wind perturbation energy is approximately 2 times greater than the meridional wind perturbation energy. In the mid- to upper troposphere, the TE during JJA is large in the SP region (Fig. 9b), and the TE during JJA and DJF is large in the TR (Figs. 9b,f). In the stratosphere, the TE during JJA and DJF was large and had a high standard deviation in the TR (Figs. 9c,g). In the stratosphere, the dominant
component of the TE in the TR is the zonal wind, which is primarily affected by the various waves generated by convective activities and other forcings in the TR, as well as by ozone concentrations (Holton 2004). These factors might affect the direction and magnitude of background zonal wind and the initial ensemble perturbations. Although the magnitude of the TE in the mesosphere was largest in the vertical distribution (Figs. 7e,j), few vertical model layers in the mesosphere rendered the vertically integrated ensemble-mean TE (normalized by the sum of the density of all vertical layers) relatively small (Figs. 9d,h). The TE was large in the winter hemispheres and reached maximum values at the midlatitudes (Figs. 9d,h). Enomoto et al. (2010) showed that the large TE in the midlatitudes at the upper layer is associated with a temporal change in zonal wind and temperature.

b. The growth rate of the initial ensemble perturbations

The perturbation growth rates are important factors in estimating the quality of the initial ensemble perturbations (Magnusson et al. 2008). To reflect the instability in the atmosphere, the initial ensemble perturbations should grow appropriately to span the subspace of the forecast error throughout the model integration. This property is a critical constraint needed to identify the initial ensemble perturbations because randomly sampled perturbations from analysis uncertainty tend to decay during the early stages of integration and consequently degrade the quality of the EPS. Wang and Bishop (2003) showed that the growth rate of the initial ensemble perturbations from the ETKF was greater than those of singular vectors and breeding methods. In this study, to estimate the growth rate of the initial ensemble perturbations from the ETKF, the Lyapunov exponent, the exponential growth rate of the perturbations in a chaotic system (Palmer 1993; Magnusson et al. 2008), is calculated as

\[ \lambda = \frac{1}{\Delta t} \ln \left( \frac{\| \Delta x(t + \Delta t) \|}{\| \Delta x(t) \|} \right), \]

where \( \Delta t = 6 \) h and \( \| \Delta x(t) \| \) is the amplitude of the perturbations at time step \( t \). Because the 10-day ensemble forecasts in the KMA EPS are produced in the limited troposphere (i.e., from the surface to 200 hPa), the growth rate of the initial ensemble perturbations was calculated for regions below 200 hPa.

Figure 10 shows the exponential growth rate of the zonal wind perturbations at 850, 500, and 200 hPa for a 48-h forecast lead time. Because Eq. (10) assumes that the perturbation growth is linear, the verification time is set to 48 h. Overall, the growth rates during DJF are lower than those during JJA. During both periods, the growth rates at 850 hPa averaged over the entire globe had negative values for the short forecast lead time, which implies that the initial ensemble perturbations decay during this time (Figs. 10a,d). Because the magnitude of the 12-h forecast ensemble perturbations is smaller than that of the initial ensemble perturbations at 850 hPa (Figs. 6a,c), negative growth rates occur at 850 hPa. As the forecast lead time increases, the growth rate over the entire globe acquires similarly positive values. The greatest growth rate occurs in the SH, whereas the smallest rate occurs in the TR at 850 hPa (Figs. 10a,d). Compared to other regions, the error growth rates in the TR are small due to the weaker energy source for dynamical growth, as described in Whitaker and Hamill (2012). At 500 and 200 hPa, the initial ensemble perturbations exhibited the greatest positive growth rate at the first time step; subsequently, the growth rates decreased as the forecast lead time increased (Figs. 10b,c,e,f). The perturbations in the NP region had the highest growth rate for all forecast lead times, whereas those in the TR had the lowest growth rate (Figs. 10b,c,e,f).

c. The hydrostatic balance of the initial ensemble perturbations

The initial ensemble perturbations from the ETKF, which are constrained by the forecast error covariance \( \mathbf{P}_f \), generally satisfy geostrophic and hydrostatic balances. The hydrostatic balance of the initial ensemble perturbations is necessary because the imbalanced initial ensemble perturbations grow during the model integration and degrade the quality of the ensemble forecasts. To measure how this balance is maintained with the initial ensemble perturbations, hydrostatically balanced potential temperature perturbations were calculated using the initial ensemble perturbations from the KMA EPS and UM analyses (see Vetra-Carvalho et al. 2012):

\[ \theta_H' = \frac{(e^{-1} - 1)q'\theta}{[1 + (e^{-1} - 1)q]} + \frac{c_p}{g} \frac{d\pi'}{dz} \theta^2 [1 + (e^{-1} - 1)q], \]

where \( e \) is the ratio of the mass of liquid water to dry air, \( q \) is the specific humidity, and \( \theta \) and \( q \) are taken from the analysis of the KMA UM.

The correlations between the hydrostatically balanced potential temperature perturbations (\( \theta_H' \)) and the potential temperature perturbations (\( \theta' \)) from the ETKF were calculated for each grid and then averaged horizontally over the NP, NH, TR, SH, and SP regions (Fig. 11). The correlations were greater than 0.9 during JJA and
DJF, with generally similar patterns among whole vertical layers; this finding implies that the initial ensemble perturbations satisfy the hydrostatic balance. Near the surface and at altitudes above 6.6 km, the correlations during JJA and DJF were generally close to 1. The correlations were between 0.9 and 1 from 0.7 to 6.6 km. The correlations in the stratosphere are larger than those in the troposphere due to higher static stability in the stratosphere than in the troposphere. The correlations during JJA and DJF at the midtroposphere (approximately 6.6 km) were smaller in the NP region (Fig. 11a) and greater in the TR (Fig. 11c), which is consistent with the largest and smallest growth rates in the NP region and the TR, respectively (see Figs. 10b,c,e,f). The correlations in the NH were greater during JJA compared with DJF (Fig. 11b); in contrast, the correlations in the SH and the SP region during JJA were smaller than those during DJF (Figs. 11d,e). The greater correlations during DJF across the entire globe (Fig. 11f) were caused by the greater correlations during DJF in the TR, SH, and SP regions.

Overall, the initial ensemble perturbations maintained hydrostatic balance during the summer and winter in the TR and during the summer in each hemisphere, consistent with the smaller perturbation growth rates.

5. Summary and discussion

To accurately estimate the forecast uncertainty in the ensemble prediction system, initial ensemble perturbations...
should represent the uncertainty of the atmospheric state at the time of analysis. Due to the limited ensemble size, the initial perturbations should sample the probability distribution of the analysis, thereby maintaining the independence between the ensemble members with appropriate growth rates. In this paper, the characteristics of the initial ensemble perturbations based on the ETKF in the high-resolution operational EPS of the KMA were investigated over two periods: from 1 June to 31 August 2011 (JJA) and from 1 December 2011 to 29 February 2012 (DJF).

The characteristics of the transform matrix and the inflation factor of the initial ensemble perturbations were investigated. The eigenvalue spectrum and the ensemble dimension of the transform matrix, which reduces forecast uncertainty with observational information, were analyzed. The eigenvalue spectra during JJA and DJF were similar across global and localized domains. The global eigenvalue spectrum was flatter than the average of the localized eigenvalue spectra; thus, localization in the ETKF might bias the filtering of forecast uncertainty in certain eigenvector directions. Overall, the ensemble dimension (a measure of how the analysis variances are distributed over the eigenvector directions) was large over regions with small errors or small error growth and with dense observations (e.g., the TR and the summer hemispheres) during JJA and DJF.

In contrast to the transform matrix, which reduces the forecast uncertainty to produce the initial ensemble perturbations, the inflation factor increases the spread of the initial ensemble perturbations to match the ensemble-mean error. Therefore, the combined effects of these components are needed to understand the characteristics of the initial ensemble perturbations from the ETKF. In this study, the combined effects were investigated using the ratio of the initial ensemble spread with regard to the 12-h forecast ensemble spread. The vertically averaged ratio was large near the TR and generally greater over the continent compared with over the ocean at the given latitude during JJA and DJF. This finding partially contradicts those of previous studies (e.g., Wang and Bishop 2003; Wei et al. 2006) concerning the effect of the transform matrix because the inflation factor based on Eq. (7) had greater values in regions with dense observations. As Whitaker et al. (2008) discussed, the multiplicative inflation factor based on Eq. (7) might produce unrealistic forecasts or analysis error variance due to the heterogeneous observation distributions. The use of other types of inflation factors [e.g., the additive or the combined additive and multiplicative inflation factor discussed by...
Whitaker and Hamill (2012)] might address different aspects of the initial ensemble perturbations.

The ensemble dimension is greater and the ratio is generally smaller in the NH compared with the SH, which indicates that the dense observations in the NH reduce forecast error and that the initial ensembles approximate the observations. In contrast, in the SH, the sparse and localized distributions of the observations differentiate the initial ensembles from the observations and the reduction of the forecast error variance by the transform matrix is restricted with regard to magnitude and direction. To solve this issue, the bias correction of the forecast error variances estimated via ensemble or statistical adjustments of the ensemble spread to the observations (Wang et al. 2007) may be adopted in future studies.

The initial ensemble perturbation energy in the boundary layer is located along the sea ice in the NP and SP regions. In the troposphere, the perturbation energy is greatest near Antarctica and located at the midlatitude of the winter hemispheres. The perturbation energy in the lower and midstratosphere is concentrated zonally in the TR, and the depth of the large TE layer depends on the vertical variation of the dominant zonal winds in this layer. From the upper stratosphere to the model top, the perturbation energies are broadly spread across the winter hemispheres where the strong polar night jets are located.

The growth rates of the ensemble perturbations were calculated for the troposphere. In the lower troposphere, the growth rates averaged across the globe were negative for short forecast lead times during JJA and DJF; then the growth rates subsequently increased and converged to positive values. The perturbations showed the smallest and largest growth rates in the TR and the SH, respectively. In the mid- and upper troposphere, the growth rates were greatest at the first forecast time step, and then decreased as the forecast lead time increased. The perturbations showed the largest and smallest growth rates in the NP regions and the TR, respectively, during JJA and DJF. On average, the perturbations increased more rapidly during JJA compared with DJF.

The hydrostatic balance of the initial ensemble perturbations were investigated by comparing the hydrostatically balanced potential temperature perturbations with the potential temperature perturbations generated from the ETKF. Hydrostatic stability is an instability criterion associated with the growth of baroclinic perturbations (Fleagle 1955); therefore, the hydrostatic balance of the initial perturbations appears to be consistent with the growth rate of the initial perturbations. During JJA and DJF, the initial perturbations were in hydrostatic balance, and these perturbations experienced the smallest growth rates in the TR. In contrast, the initial perturbations were in relatively weak hydrostatic balance in the NP regions where the greatest increase in the growth rate occurred. In the lower troposphere from approximately 0.7 to 6.6 km, the initial perturbations departed from hydrostatic balance during JJA and DJF. The initial perturbations were closer to hydrostatic balance states in the SH during DJF compared with JJA; the same was true in the NH during JJA compared with DJF. These findings imply that hydrostatic balance is maintained in the summer in both hemispheres and during JJA and DJF in the TR, and is associated with the smallest growth rates of the initial ensemble perturbations.

The observational effect should be reflected appropriately in the ETKF to represent the initial uncertainties by the initial ensemble perturbations. Because the distribution and the density of the observations affect the transform matrix and the inflation factor in the ETKF, they also influence the characteristics and quality of the initial ensemble perturbations. The adoption of different algorithms for the inflation factor or a bias correction scheme of forecast error might improve the quality of the initial ensemble perturbations generated by the ETKF in the KMA EPS, by relaxing the effect of the observational distribution smoothly along the latitudinal and longitudinal directions. From a dynamic perspective, the initial ensemble perturbations of the ETKF in the KMA EPS are generally in hydrostatic balance and exhibit appropriate growth rates, which implies that the initial ensemble perturbations appropriately describe the uncertainties associated with atmospheric dynamic features. In the future, additional studies of the meteorological phenomena in the upper layer using the initial ensemble perturbations of the KMA EPS will be conducted.

Acknowledgments. The authors thank the two anonymous reviewers for their valuable comments. This study was supported by the Korea Meteorological Administration Research and Development Program under Grant CATER 2013–2030. The authors thank Dr. Young-You Park and Ms. Joohyung Son for their support in the early stages of this study and the Numerical Weather Prediction Division of the Korea Meteorological Administration and the Met Office for providing the resources necessary for this study.

REFERENCES


