A Four-Dimensional Ensemble-Variational (4DEnVar) Data Assimilation System for Global NWPs: System Description and Primary Tests

S. J. Zhu[1], B. Wang[2,1,4,5], L. Zhang[3], J. J. Liu[2,4], Y. Z. Liu[3], J. D. Gong[3], S. M. Xu[1], Y. Wang[1], W. Y. Huang[1], L. Liu[1], Y. J. He[2] and X. J. Wu[3]

[1]Department of Earth System Science, Tsinghua University, Beijing 100084, China.

[2]State Key Laboratory of Numerical Modelling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, 100029, China.

[3]Research and Development Division, Numerical Weather Prediction Center of China Meteorological Administration, Beijing, 100081, China

[4]Innovation Group 311020008, Southern Marine Science and Engineering Guangdong Laboratory, Zhuhai, China

[5]School of Ocean, University of Chinese Academy of Sciences, Qingdao, 266400, China.

Corresponding author: Bin Wang (wab@lasg.iap.ac.cn)

Key Points:

- A DRP-4DVar-based 4DEnVar data assimilation system with the flow-dependent BEC was developed for global numerical weather predictions
- The deterministic forecast initialized from the 4DEnVar ensemble mean analysis has comparable performance to the 4DVar system
- Higher quality of analyses and ensemble forecasts can be produced by the 4DEnVar system relative to the 4DVar system

Abstract

A four-dimensional ensemble-variational (4DEnVar) data assimilation (DA) system was developed for global numerical weather predictions (NWPs). Instead of using the adjoint technique, this system utilizes a dimension-reduced projection (DRP) technique to minimize the cost function of the standard four-dimensional variational (4DVar) DA. It dynamically predicts ensemble background error covariance (BEC) initialized from its previous inflated analyses and realizes the flow-dependence of BEC in the variational configuration during the assimilation cycle. These inflated analyses, linear combinations of the ensemble analyses increment and balanced random perturbations, aim to prevent the predicted BEC from underestimation as well as to implicitly achieve the hybrid of the flow-dependent and static BEC matrices. A limited number of leading eigenvectors of the localization correlation function are selected to filter out the spurious correlations in the BEC matrix (B-matrix). In order to evaluate the new system, single-point observation experiments (SOEs) and observing system simulation experiments (OSSEs) were conducted with sounding and cloud-derived wind data. The flow-dependent characteristic was verified in the SOEs that utilized the localized ensemble covariance and compared with that of 4DVar. In the OSSEs, 4DEnVar reduced the analysis errors compared with 4DVar. The deterministic forecast initialized from the 4DEnVar ensemble mean analysis has better (worse) performance in the medium-range (long-range) forecasts in the Northern Extratropics and opposite performance in the Southern Extratropics, and exhibits slightly worse effects in the Tropics. Moreover, the ensemble mean forecast initialized from the 4DEnVar ensemble analyses has higher forecast skills than 4DVar.

Plain Language Summary

Medium-range numerical weather prediction (NWP) is of great significance to disaster mitigation and improvement of human living standards. It aims to predict weather states for future 1-10 days from the current state by solving the initial value problem of a set of partial differential equations. Data assimilation (DA) is one of the key techniques to improve forecast skills, which attempts to provide an optimal estimation of the current state by combining observations and forecasts. This study developed a four-dimensional ensemble-variational (4DEnVar) DA system for global NWPs using the dimension-reduced projection (DRP) four-dimensional variational

(4DVar) approach. Compared with the standard 4DVar, which is generally recognized as one of the most advanced DA methods, this new system has three unique features. First, it dynamically estimates background error covariance (BEC) in the DA cycle instead of adopting a pre-estimated static BEC as 4DVar does. Second, it uses an ensemble covariance without the Gaussian error assumption as in 4DVar. Third, it can avoid using adjoint models and handle nonlinear problems well. It shows more obvious flow-dependence of the BEC, smaller analysis errors, and better ensemble mean forecast skills than 4DVar, and comparable skills of deterministic forecast initialized from the ensemble mean analysis to 4DVar.

1 Introduction

Accurately predicting future weather and climate states is of great significance to disaster mitigation and to the improvement of human living standards. The accuracy of global numerical weather prediction (NWP) can be significantly improved through the use of new types of data such as from satellites (Simmons & Hollingsworth, 2002). Therefore, it is necessary to develop an effective data assimilation (DA) system to make good use of observations to provide more accurate initial conditions (ICs) for NWPs.

The four-dimensional variational (4DVar) DA is recognized as one of the most advanced DA methods. This method produces the analysis field constrained dynamically and physically (Rabier et al., 2000; Wang et al., 2010a), and implicitly implements the flow-dependent background error covariance (BEC) matrix (B-matrix), which propagates information within the assimilation window by the tangent linear model (TLM) and the adjoint model (ADM; Lorenc, 2003). The uses of the ADM (Lewis & Derber, 1985; Le Dimet & Talagrand, 1986) and the incremental 4DVar scheme (Courtier et al. 1994) make the operational application of 4DVar possible (Rabier et al., 2000; Gauthier & Thépaut, 2001; Koizumi et al., 2005; Rawlins et al., 2007; Gauthier et al., 2007; Zhang et al., 2019). However, this advanced DA method has not been applied in most NWP centers in the world, except for very few major advanced centers, e.g., the European Centre for Medium-range Weather Forecasts (ECMWF). In addition, the standard 4DVar approach fails to dynamically update the B-matrix during the assimilation cycle, given that it uses the modeled climatological covariance model that can only be implicitly developed within the assimilation window (Buehner et al., 2010a).

Ensemble Kalman Filter (EnKF) is another commonly used ensemble DA method. It is based on the Monte Carlo method that uses a number of ensembles to estimate the B-matrix with the explicit flow-dependent characteristic spanning the assimilation windows (Evensen, 1994) without modeling the B-matrix nor using the ADM. Moreover, EnKF has the advantage of saving time, implicitly through concurrently generating the ensembles on a parallel computer system due to the mutual independence of ensemble members. Due to these advantages, EnKF has been applied to many models, e.g., the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) Model (Whitaker et al., 2008, 2009), the Global Environmental Multiscale (GEM) Model (Buehner et al., 2010a, 2010b). There have been some studies comparing the

performance of the variational and EnKF systems. Whitaker et al. (2008) compared 3DVar and EnKF using low-resolution operational model and observations, except satellite radiation, and found that the ensemble system outperforms the 3DVar system, especially in data-sparse areas. Whitaker et al. (2009) further compared 3DVar, 4DVar and EnKF using sparse surface pressure observations, and discovered that 4DVar and EnKF have comparable performance. Buehner et al. (2010b) found slight degradations (improvements) in the short-range (medium-range) forecasts based on the EnKF ensemble mean analysis over the 4DVar-based forecasts in the Extratropics. There is not enough evidence to prove that the forecast provided by EnKF is better than that provided by 4DVar for the NWP models. Also, it is noted that the limited size of ensembles can result in sampling errors in the ensemble B-matrix.

However, the ensemble method can provide the explicit flow-dependent information for the variational method (Houtekamer et al., 2005; Whitaker et al., 2008, 2009; Buehner et al., 2010a, 2010b). Likewise, the variational method can supply the ensemble method with proven modules, e.g., quality control and minimization iteration modules (Courtier et al. 1994; Zhang et al., 2019). Therefore, several hybrid DA methods combining the variational and ensemble ideas have continuously been developed (Hamill & Snyder, 2000; Lorenc, 2003; Qiu et al., 2007; Liu et al., 2008, 2009; Tian et al., 2008, 2011; Wang et al., 2010a).

Different methods of incorporating the ensemble covariance make the classification of hybrid methods different. The hybrid ensemble-4DVar methods are mainly divided into En4DVar methods that include the ADM and 4DEnVar methods that avoid the ADM. En4DVar methods typically incorporate the ensemble covariance into the variational framework by a weighted sum of the static and ensemble covariances (Hamill & Snyder, 2000) or extending the original control variables by the control variables preconditioned by the square root of the ensemble covariance (Lorenc, 2003). Also, En4DVar methods can use the ensemble information to estimate the parameters of the covariance model for variational systems (Lei et al., 2020). In addition, the effects of the hybrid BEC on forecast skills have been investigated in simple models (Hamill & Snyder, 2000), regional models (Wang et al., 2008a, 2008b; Zhang & Zhang, 2012) and global models (Raynaud et al., 2011; Bonavita et al., 2012; Buehner et al., 2010a, 2010b, 2013, 2015; Clayton et al., 2013; Lorenc, 2015; Wang et al., 2013; Wang & Lei, 2014; Kleist & Ide, 2015a, 2015b).

4DEnVar method, which applies the variational framework and the idea of using ensembles valid at multiple time slots to avoid the ADM to obtain the optimal analysis, is an efficient DA method. Several ensemble-based methods, which can reduce the dimension from the model space to a subspace composed of a limited number of base vectors in optimization and avoid the use of the ADM, have been proposed in recent decade (Qiu et al., 2007; Tian et al., 2008; Wang et al., 2010a). The dimension-reduced projection 4DVar (DRP-4DVar) is one of the 4DEnVar methods that has been successfully applied in regional meso-scale weather forecasts (Wang et al., 2010a; Zhao & Wang, 2010; Liu & Wang, 2011; Zhao et al., 2012) and global decadal climate predictions (He et al., 2017, 2020a, 2020b; Li et al., 2021a, 2021b; Shi et al., 2021). In global medium-range NWPs, this approach has not been widely applied and systematically evaluated, although a DRP-4DVar system (Shen et al., 2015) was preliminarily established using an old version of the global forecast system of the Global/Regional Assimilation and Prediction System (GRAPES-GFS) based on the 3DVar system of this version (Chen et al., 2008; Xue et al., 2008). This method uses a limited number of base vectors composed of initial perturbations to project the incremental analysis in model space onto a low-dimensional subspace spanned by these base vectors, and directly obtains an optimal analysis solution to the minimization of the 4DVar cost function in the subspace. Furthermore, this method calculates the gradient of the cost function based on the statistical relationship between the model space and observation space, thereby avoiding the use of the ADM (Wang et al., 2010a).

The limited ensemble size may result in introducing sampling errors, which can lead to spurious correlations in the B-matrix (Evensen, 2003), and localization techniques (Liu et al., 2009; Hamill et al., 2001; Wang et al., 2010b, 2018) can effectively alleviate the aforementioned problem and ameliorate analyses and forecasts. Given that conducting localization in model space is quite inconvenient in implementation and computationally expensive for the non-sequential ensemble methods, adopting ensemble-sample-based subspace localization schemes is thought to be an economical choice (Wang et al, 2018). Localization is typically conducted as a Schür product between the ensemble-based B-matrix and the correlation matrix composed of the elements calculated by the correlation model related with their coordinates, so how to decompose the correlation matrix to avoid the expensive multiplication between high-dimensional matrices caused by the Schür product is the key to reduce computational costs. A limited number of leading eigenvectors expressed by orthogonal functions (e.g., empirical orthogonal function, sine function

and spherical harmonic function) were used to expand the correlation function so that the high-dimensional correlation matrix is decomposed into the sum of a set of products between an eigenvector and its transpose (Liu et al., 2009; Buehner et al., 2010a, 2010b; Bishop et al., 2011; Kuhl et al., 2013; Wang et al., 2010b, 2018). This approach not only alleviates the spurious correlations and rank deficiency of the B-matrix, but also efficiently produces the extended ensemble samples, which converts a very costly Schür product between two high-dimension matrices to much more economical Schür products between ensemble samples and eigenvectors.

Motivated by these studies, many research and operational centers have not only established their standalone variational systems, but also have been developing hybrid DA systems for their global NWPs. These centers realized the explicit flow-dependence of the B-matrix based on the original standard 4DVar framework, so that the forecast skills were further improved. The ECMWF (Bonavita et al., 2012) and Météo-France (Raynaud et al., 2011) have developed hybrid DA systems, which include ensemble information estimated by an ensemble of 4DVars. The Met Office incorporated the flow-dependent BEC estimated by EnKF into the 4DVar system to develop a hybrid system (Clayton et al., 2013). Unlike these systems relying on the ADM, some centers have developed 4DEnVar systems avoiding the use of the ADM. Environment Canada combined the static BEC with the 4D ensemble BEC obtained from EnKF to develop a 4DEnVar system, which is considered to be a potential alternative to 4DVar considering the simplicity, computational efficiency and forecast quality (Buehner et al., 2010a, 2010b, 2013, 2015). The Met Office developed a hybrid 4DEnVar system (Lorenc et al., 2015; Bowler et al., 2017a) and used an ensemble of 4DEnVars instead of the ETKF system to generate ensembles for the hybrid system (Bowler et al., 2017b). Wang et al. (2013) and Kleist & Ide (2015a) proved the benefits of including ensemble BECs into 3DVar. Then, perturbations valid at multiple time slots during the assimilation window were used to estimate the 4D ensemble BEC to develop a 4DEnVar system in NCEP (Wang et al., 2014; Kleist & Ide., 2015b).

This study focuses on developing a 4DEnVar system for global NWPs based on the DRP-4DVar approach, which not only can be an alternative to DA system to provide deterministic forecasts, but also to provide ensemble forecasts for hybrid systems. Compared with the theoretical DRP-4DVar proposed by Wang et al (2010a), the method used for the 4DEnVar system was improved in its inflation, localization and sampling. The successful applications of the 4DEnVar algorithm and economical localization technique provide a good foundation to develop the

4DEnVar system. As the first and necessary step to evaluate the impact of the 4DEnVar system on analyses and forecasts, single-point observation experiments (SOEs) and observing system simulation experiments (OSSEs) were conducted. SOEs are easy to study the flow-dependent characteristic of the BEC. OSSEs can help us evaluate the realistic analysis error because the "truth" state is known. The remainder of the paper is organized as follows. Section 2 introduces the formulation, localization and inflation techniques of the 4DEnVar system. Section 3 follows with the implementation of the 4DEnVar system and the experiment descriptions. Section 4 evaluates the performance of the 4DEnVar system on analyses and forecasts relative to the 4DVar system. The summary and the prospect for future work are presented in the last section.

2 Description of method

2.1 Incremental 4DVar algorithm

The variational system used in this paper (Zhang et al, 2019) adopts the incremental 4DVar scheme (Courtier et al., 1994), which usually obtains the optimal analysis of IC by minimizing a cost function on a low-resolution grid:

$$J[\delta x(t_0)] = \frac{1}{2}\{\delta x(t_0) - [x^b(t_0) - x^g(t_0)]\}^T B_0^{-1}\{\delta x(t_0) - [x^b(t_0) - x^g(t_0)]\} + \frac{1}{2}\sum_{i=0}^{n}[\mathbf{H}_i\delta x(t_i) - d_i]^T R_i^{-1}[\mathbf{H}_i\delta x(t_i) - d_i], \quad (1)$$

where x^b is the background state vector, x^g is the first guess vector, $\delta x = x - x^g$ is the perturbation of the IC, $\delta x(t_i) = \mathbf{M}_i\delta x(t_0)$ is the perturbation at time t_i, B is the static B-matrix, and R is the observation error covariance matrix. $d_i = y_i^o - H_i[x^g(t_i)]$ contains the observation innovations at time t_i, y_i^o contains the observations at time t_i, involving the observation operator H_i at time t_i and the nonlinear forecast model integration M_i from the analysis time to time t_i. $\mathbf{H}_i$ is the tangent linear observation operator corresponding to H_i, and $\mathbf{M}_i$ is the TLM of M_i.

The convergence rate of the gradient for the optimization problem is dependent on the condition number of the Hessian matrix (Zupanski, 1996). Operational DA systems generally reduce the condition number of the Hessian matrix of Eq. (1) through the preconditioning transformation, that is $\delta x = Uv$. Thus, the modeled climatological BEC can be estimated by

$$B_c = UU^T, \quad (2)$$

where v is the preconditioned state variable vector. U contains the physical transformation operator that transforms independent variables to model variables, the diagonal matrix composed of the background error variance square root of the independent variables, and the background error correlation transformation matrix (Zhang et al., 2019). After the aforementioned preconditioning transformation, Eq. (1) becomes

$$J[v(t_0)] = \frac{1}{2}[v(t_0)]^T[v(t_0)] + \frac{1}{2}\sum_{i=0}^{n}[\mathbf{H}_i\mathbf{M}_iUv(t_0) - d_i]^T R_i^{-1}[\mathbf{H}_i\mathbf{M}_iUv(t_0) - d_i]. \quad (3)$$

In order to minimize Eq. (3), the corresponding gradient should equal to zero, and the ADM is required for calculating the gradient of the cost function. Moreover, the calculation of the ADM requires the backward model trajectories, which are typically provided by the nonlinear forecast model and expensive in calculation and storage. Thus, the adjoint-free DA method may be an efficient alternative to 4DVar for operational applications.

2.2 4DEnVar

2.2.1 Algorithm

The 4DEnVar algorithm is based on the DRP-4DVar approach (Wang et al, 2010a) that projects the initial increment δx in model space onto the subspace expanded by a limited number of IC perturbation samples as its basis vectors, and obtains the optimal solution directly in the subspace.

For the convenience of implementing 4DEnVar in the standard 4DVar framework, the IC perturbation samples are obtained by the "randomcv" method, which uses the variational variable transform $\delta x = Uv$ to produce an ensemble of balanced and reasonable perturbations (Baker, 2005). $X = [\delta x_1, \delta x_2, \cdots, \delta x_K]$ contains the IC perturbation samples, where K is the ensemble size. The corresponding observational perturbation samples $Y = [\delta y_1, \delta y_2, \cdots, \delta y_K]$ are calculated using TLMs and tangent linear observation operators. Thus, an ensemble of IC perturbation samples and observational perturbation samples are chosen to define the following projection matrices:

$$\begin{cases} p_x = \frac{1}{\sqrt{K-1}}\left[\delta x_1 - \overline{\delta x}, \delta x_2 - \overline{\delta x}, \cdots, \delta x_K - \overline{\delta x}\right] \\ p_y = \frac{1}{\sqrt{K-1}}\left[\delta y_1 - \overline{\delta y}, \delta y_2 - \overline{\delta y}, \cdots, \delta y_K - \overline{\delta y}\right] \end{cases}, \tag{4}$$

where $\begin{cases} \overline{\delta x} = \frac{1}{K}[\delta x_1 + \delta x_2 + \cdots + \delta x_K] \\ \overline{\delta y} = \frac{1}{K}[\delta y_1 + \delta y_2 + \cdots + \delta y_K] \end{cases}$. Defining $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_K)^T$ as a K-dimensional vector composed of the weight coefficients of the basis vectors, δx and $\mathbf{H}_i \delta x(t_i)$ can be projected onto the subspace spanned by the ensemble samples via the following transformation:

$$\begin{cases} \delta x = p_x \alpha \\ \mathbf{H}_i \delta x(t_i) = p_y(t_i)\alpha \end{cases}, \tag{5}$$

where $p_y(t_i)$ is the observational projection matrix at time t_i. Thus, the ensemble BEC can be represented by

$$B_e = p_x p_x^T \tag{6}$$

and the new cost function with respect to α can be written as:

$$J[\alpha(t_0)] = \frac{1}{2}[\alpha(t_0)]^T[\alpha(t_0)] + \frac{1}{2}\sum_{i=0}^{n}\left[p_y(t_i)\alpha(t_0) - d_i\right]^T R_i^{-1}\left[p_y(t_i)\alpha(t_0) - d_i\right]. \tag{7}$$

To minimize Eq. (7), α must satisfy $\left[\frac{\partial J}{\partial \alpha(t_0)}\right]^T = 0$. Here, no ADM is needed. It is noted that a degraded analysis may result from the approximation in Eq. (5) due to the much smaller ensemble size than the dimension of the original IC perturbation, which can be alleviated by localization techniques.

2.2.2 Localization

The major drawback to the ensemble-based method is its spurious correlations and very small rank in the BEC due to the limited number of the IC perturbation samples, which excessively constrains the solving subspace of the optimal analysis increment. Localization is considered to be an effective technique to alleviate the aforementioned problems (Hamill et al., 2001).

The localized B-matrix can be typically expressed as the Schür product between the ensemble BEC B_e and the correlation matrix of the covariance localization C. Because the direct use of the localized B-matrix may lead to much computational cost according to Wang et al.

(2018), this matrix should be expressed in a form that can be used easily and economically. The correlation matrix can be approximately decomposed into a limited number of leading eigenvectors and extended IC perturbation samples can be obtained:

$$Ep_x = \left[\left(p_{x,1} \circ \boldsymbol{\rho}_{x,1}, \cdots, p_{x,1} \circ \boldsymbol{\rho}_{x,L}\right), \cdots, \left(p_{x,K} \circ \boldsymbol{\rho}_{x,1}, \cdots, p_{x,K} \circ \boldsymbol{\rho}_{x,L}\right)\right], \tag{8}$$

where $\boldsymbol{\rho}_{x,j}$ $(j = 1, 2, \cdots, L)$ is a leading eigenvector in model space, and L is the number of the selected leading eigenvectors according to the cumulative contribution of variance. In implementation, each leading eigenvector can be decomposed into zonal, meridional and vertical components: $\boldsymbol{\rho}_{x,j} = \boldsymbol{\rho}^z_{x,j_z} \circ \boldsymbol{\rho}^m_{x,j_m} \circ \boldsymbol{\rho}^v_{x,j_v}$. The EOF decomposition method is used to obtain the zonal and vertical components:

$$\begin{cases} \boldsymbol{\rho}^z_{x,j_z} = \boldsymbol{E}^z_{x,j_z}\left(\lambda^z_{x,j_z}\right)^{1/2} \\ \boldsymbol{\rho}^v_{x,j_v} = \boldsymbol{E}^v_{x,j_v}\left(\lambda^v_{x,j_v}\right)^{1/2} \end{cases}, \tag{9}$$

where $\boldsymbol{E}^z_{x,j_z}$ and $\boldsymbol{E}^v_{x,j_v}$ are eigenvectors for the zonal and vertical components, respectively, obtained using the empirical orthogonal decomposition. λ^z_{x,j_z} and λ^v_{x,j_v} are their corresponding eigenvalues. Then, the sine expansion method is utilized (Wang et al., 2018) to obtain the meridional component:

$$\boldsymbol{\rho}^m_{x,j_m} = \boldsymbol{E}^m_{x,j_m}\beta^{1/2}_{x,j_m}\,. \tag{10}$$

Here, $\boldsymbol{E}^m_{x,j_m}$ is the eigenvector for the meridional component, and β_{x,j_m} is its eigenvalue. When defining the correlation function model, we used the GC correlation function (Gaspari & Cohn, 1999) for the horizontal components:

$$C(r) = \begin{cases} -\frac{1}{4}r^5 + \frac{1}{2}r^4 + \frac{5}{8}r^3 - \frac{5}{3}r^2 + 1, 0 \le r \le 1 \\ \frac{1}{12}r^5 - \frac{1}{2}r^4 + \frac{5}{8}r^3 + \frac{5}{3}r^2 - 5r + 4 - \frac{2}{3}r^{-1}, 1 < r \le 2 \\ 0, 2 < r \end{cases} \tag{11}$$

Here, r is defined as the dimensionless latitude and longitude distance. The following correlation function is used for the vertical component:

$$C(r) = \frac{1}{1.0 + K_p r^2}. \tag{12}$$

where r is defined as the dimensionless logarithmic pressure distance.

According to Eq. (8) and ignoring the time-variation of the localization leading eigenvectors, the extended observational perturbation samples can be then represented as

$$Ep_y = \left[\left(p_{y,1} \circ \boldsymbol{\rho}_{y,1}, \cdots, p_{y,1} \circ \boldsymbol{\rho}_{y,L}\right), \cdots, \left(p_{y,K} \circ \boldsymbol{\rho}_{y,1}, \cdots, p_{y,K} \circ \boldsymbol{\rho}_{y,L}\right)\right]. \quad (13)$$

Redefining the control variables as an $K \times L$-dimensional vector β, the analysis increment and observational increment can be modified as

$$\begin{cases} \delta x = Ep_x \beta \\ \mathbf{H}_i \delta x(t_i) = Ep_y(t_i)\beta \end{cases}. \quad (14)$$

Finally, the localized cost function is formulated on the extended sample space. In the generation of the extended observational perturbation samples, the TLM is called for only K times. On one hand, the ensemble size can be greatly increased from the original samples to the extended samples without any additional computational cost for TLM calling, and the leading eigenvectors in both the model and observation space can be pre-calculated according to the coordinates of the model grid and observation locations. On the other hand, spurious correlations among the original samples can be significantly eliminated, and the calculation accuracy of the cost function and its gradient can be improved given that the extended samples have better independence from each other than the original samples.

2.2.3 Inflation

The 4DEnVar system uses an inflation technique similar to the Relaxation-to-prior-perturbations (RTPP; Zhang et al., 2004) to mitigate the filter divergence problem during the assimilation cycle. Different from the RTPP, this inflation adopts random perturbations with balance constraints from the static BEC of the standard 4DVar system.

60 random samples with balance constraints $(\delta x_k)_r$ $(k = 1, 2, \cdots, 60)$ are obtained by the "randomcv" method introduced in Section 2.2.1 (Baker, 2005). They are linearly combined with the 60 analysis increment samples $(\delta x_k)_a$ $(k = 1, 2, \cdots, 60)$ using the weights $\gamma_1 = 0.2$ for the former and $\gamma_2 = 0.9$ for the latter to achieve the inflation of the ensemble BEC for the next assimilation after a number of model integrations from the beginning to the end of the assimilation window initialized by the inflated analysis increments $(\delta x_k)_a^{inf}(k = 1, 2, \cdots, 60)$, where

$$\begin{cases} (\delta x_k)_a^{inf} = \gamma_1 (\delta x_k)_r + \gamma_2 (\delta x_k)_a \\ (\delta x_k)_f = M\left[x_b + (\delta x_k)_a^{inf}\right] \end{cases} \quad (k = 1, 2, \cdots, 60). \tag{15}$$

An obvious advantage of this inflation method is that it is convenient and easy to generate the random samples with balance constraints directly through the preconditioning process of the 4DVar system. Moreover, the inflation technique implicitly incorporates the climatological BEC into the ensemble BEC to construct the hybrid B-matrix, which can be represented by

$$(B_e)_{inf} = (p_x)_f (p_x)_f^T. \tag{16}$$

Here, $(p_x)_f = \left[(\delta x_1)_f, (\delta x_2)_f, \cdots, (\delta x_{60})_f\right]$ contains the updated inflated analysis increments. Thus, the inflation method not only alleviates the underestimation of the B-matrix, but also implicitly realizes the hybrid BEC for the 4DEnVar system.

3 Experimental design

3.1 Implementation of 4DEnVar system

In this study, the model used in the 4DEnVar system is the GRAPES-GFS model version 3.0 (Su et al., 2020) and contains 87 vertical levels. The horizontal resolution of the system is $0.5° \times 0.5°$ for the outer loop and $1.0° \times 1.0°$ for the inner loop. The 4DEnVar system combines the ensemble BEC estimated by 60 samples and the original variational framework to solve the assimilation problem, and is evaluated in comparison to the available 4DVar system (Zhang et al, 2019) with the same model and same resolutions. The schematic flowchart in Figure 1 describes the operational process of the 4DEnVar system. In order to mitigate the sampling errors and spurious correlations in the BEC due to the limited ensemble size (Hamill et al., 2001; Lorenc et al., 2003; Wang et al., 2010b, 2018), the localization scheme is designed according to the implementation introduced in Section 2.2.2, with 7° for the filtering radius in the horizontal direction and 3 for the filtering parameter K_p in the vertical direction. The minimization problem of the 4DEnVar system is solved in the subspace spanned by the extended samples derived from the Schür products between the ensemble members and the leading eigenvectors of the localization correlation function. The 4DEnVar system not only can realize the implicit flow-dependence of the BEC within the assimilation window like the standard 4DVar system, but also can achieve the explicit flow-dependent BEC from one assimilation window to the next. During the assimilation

cycle, the perturbed observations are continuously assimilated into the ensemble samples, and the flow-dependent ensemble samples are updated every 6 hours.

To alleviate the underestimation of the BEC in the assimilation cycle, an inflation technique based on the random perturbations with balance constraints is applied. Collaborated with the localization and perturbing techniques of observation and SST, the inflation may alleviate the filtering divergence problem during the assimilation cycle. Observational perturbations are obtained by superimposing normal distribution random perturbations with zero as their expectations (or mean values) and the observation errors as their standard deviations onto the observations. SST perturbations are produced similarly except that the standard deviations of random perturbations adopts the SST analysis errors.

3.2 Experiment design

In order to evaluate the performance of the 4DEnVar system efficiently, the OSSE is considered as one of the best choices. Here, two OSSEs are designed using the $0.25° \times 0.25°$ version of GRAPES-GFS for both the 4DEnVar and standard 4DVar systems. The OSSE for the latter is to provide a reference for comparisons.

A previous study has demonstrated that the 4DVar system using the GRAPES-GFS model significantly outperforms the 3DVar system using the same model on both analyses and medium-range forecasts, especially in the Southern Hemisphere (Zhang et al., 2019). OSSEs can be used to fairly evaluate the performance of the assimilation system (Wang et al., 2008a; Wang et al., 2010a; Kleist et al., 2015a, 2015b). In order to further study the influence of the 4DEnVar system, comparisons between it and the 4DVar system are necessary.

The time period of both experiments was about one month (0900 UTC 13 September 2016 - 0900 UTC 11 October 2016) after a 2-day assimilation cycle covering the period from 0900 UTC 11 September 2016 to 0900 UTC 13 September 2016 to alleviate the influence of the spin-up. Only the inflation coefficient tuning experiments were analyzed for about one weak (0900 UTC 13 September 2016 - 0900 UTC 18 September 2016). The analysis time was taken at the beginning of the assimilation window. In the OSSEs, the results from an uninterrupted free run with the higher-resolution ($0.25° \times 0.25°$) version of GRAPES-GFS were used as the “truth” state. To eliminate the impact of spin-up, the “truth” state was initiated from the time 24 hours prior to the

analysis time of the first assimilation window with the ERA-5 reanalysis field as the IC, which was verified to be consistent with the realistic atmospheric state in terms of geopotential height and precipitation in the first 8 days. For example, we investigated the rationality of the "truth" state in the Northern and Southern Extratropics based on a comparison of the 500hPa geopotential height between the ERA-Interim reanalysis and the "truth" state at 1200 UTC on 14, 16 and 18 September 2016 (Figure 2). Figure 2a shows the 500hPa geopotential height from the ERA-Interim reanalysis in the Northern Extratropics at 1200 UTC on 14 September 2016, with a low-pressure system near the Arctic and 4 troughs extended from the low-pressure system near 60°E, 180°, 120°W and 30°W. The low-pressure system extends along 180° and 30°W, and the locations and intensities of other main systems change slightly as the integration time increases (Figures 2e and 2i). The "truth" state captures these main features and their time-variations (Figures 2b, 2f and 2j). Similarly, Figure 2c shows the results from the ERA-Interim reanalysis in the Southern Extratropics. A low-pressure system exists near the Antarctic at 180° with 3 troughs near 0°, 90°W and 90°E, and some troughs at low and middle latitudes. As the integration time increases, the intensity of the low-pressure system near the Antarctic weakens and a high value center appears near 60°E, and the locations and intensities of the main systems at low and middle latitudes change slightly (Figures 2g and 2k). The "truth" state simulates these main systems well (Figures 2d, 2h and 2l). In general, the "truth" state reasonably captures the main features and the time-variations of the 500hPa geopotential height from the ERA-Interim reanalysis and gradually degrades following the increase of integration time.

The "observations" were produced by interpolating the "truth" state to the positions at which sounding and cloud-derived wind observations are located, and then superimposing normal distribution random perturbations with zero as their expectations and the observation errors as their standard deviations onto them. Figure 3 shows the spatial distribution of these observations. Sounding observations are typically sampled in the continental areas of the Northern Extratropics and are valid at 1200 UTC 13 September 2016. Cloud-derived wind observations are sampled every 30 minutes, mainly in the central and eastern North Pacific, the eastern South Pacific, the northern Indian Ocean, the Atlantic Ocean, as well as some continents such as the America and Africa. The observation errors were taken the same as the 4DVar system.

For the first assimilation window of the 4DVar system, the background was obtained from a 15-h forecast by the $0.5° \times 0.5°$ version of GRAPES-GFS initialized from the 6-h forecast of the ERA-Interim dataset, so that it is different from the "truth" state. Meanwhile, for the first assimilation window of the 4DEnVar system, 60 IC samples were generated by superimposing 60 random perturbation samples onto this background. These perturbation samples were generated according to the "randomcv" method introduced in Section 2.2.1 (Baker, 2005). The background for each assimilation window of the 4DEnVar system is the ensemble mean of the IC samples of this window, which are derived from 60 6-h forecasts by the $0.5° \times 0.5°$ version of GRAPES-GFS with 60 inflated analysis samples produced in the previous assimilation window as their ICs, respectively, except for the first assimilation window. The 4DEnVar system has the same background as the 4DVar system in the first assimilation window because the ensemble mean of 60 superimposed random perturbation samples is zero.

In addition, based on the OSSE for the 4DEnVar, two sets of SOEs were also conducted for both the 4DEnVar and 4DVar systems within a 6-h window covered the period from 0900 UTC 13 September 2016 to 1500 UTC 13 September 2016 after a 2-day assimilation cycle to verify the flow-dependent characteristic of the BECs. Both sets of SOEs adopted the same filtering radius that is $15°$ in the horizontal direction and the same background that is the ensemble mean of the IC samples produced by the 4DEnVar system. In each set of SOE, the DA system assimilated the single-point observation valid at the beginning, middle and end of the assimilation window, respectively. The first single-point observation valid at 1200 UTC 13 September (i.e., at the middle of the window) was selected from the "observations" in the OSSEs, which is the single-point temperature observation located upstream at the top of the short-wave ridge in the middle troposphere, and the other two at the beginning and end of the window took the same location and observation innovation as the first. The observation error was set to 0.95, and the observation innovation was -1.53 K.

4 Results

4.1 Single-point observation experiments

Figures 4a and 4d show the analysis increments from the 4DVar and 4DEnVar systems, respectively, which are produced by assimilating the same single-point temperature observation at

401 the beginning of the assimilation window in two SOEs. Both increments show the maximum
402 negative values near the observation as a response to the low temperature observation. The 4DVar
403 analysis increment of temperature appears a quasi-Gaussian distribution around the observation
404 location (Figure 4a). Given that the analysis time the resultant analysis increment is obtained is the
405 same as the time the single-point observation is located at, this distribution is reasonable. In
406 contrast, the 4DEnVar analysis increment of temperature, obtained using the ensemble BEC,
407 extends along the gradient of geopotential height, which is consistent with the northwestern
408 background flow. This visually demonstrates the explicit flow-dependence of the BEC of the
409 4DEnVar system. Furthermore, satisfactorily, no spurious correlations are sighted near the analysis
410 increment produced by the 4DEnVar system when the signal of the observation is preserved in the
411 analysis increment (Figure 4d). Both experiments also show cyclone wind responses around the
412 temperature increments, which suggests that the BECs satisfy some balance constraints. These
413 results are consistent with the SOEs introduced in Kleist et al. (2015b).

414 We further investigated the implicit flow-dependence of the BECs of both systems within
415 the assimilation window by visualizing the analysis increments obtained by assimilating single-
416 point observations valid at different time levels. Unlike those with the observation at the beginning
417 of the assimilation window, the maximum negative values of the increments produced by both
418 systems shift towards the northwest of the observation. Moreover, the further the observation is
419 located from the analysis time, the more the increments from both systems extend along the
420 gradient of geopotential height (Figures 4b, 4c, 4e and 4f). These suggest that 4DVar realized the
421 evolution of BEC within the assimilation window using the TLM and the ADM, while 4DEnVar
422 did through the statistical relationship between the model space and observation space.

423 4.2 Observing system simulation experiments

424 We divided the globe into four regions for the statistics and analysis of the following
425 indicators, including the Northern Extratropics (20°N~90°N, 180°W~180°E; NH-X), Southern
426 Extratropics (20°S~90°S, 180°W~180°E; SH-X), East Asia (15°N~65°N, 70°E~145°E; EA) and
427 Tropics (20°S~20°N, 180°W~180°E; TR). Considering that the root mean square error (RMSE)
428 mainly measures the random error that is not as correctable as the systematic bias, it is usually
429 applied to statistically analyze the random errors of the background and analysis fields. To exclude

the systematic error from the RMSE, we use a metrics called anomaly RMSE (ARMSE) instead of RMSE (He et al. 2020a):

$$ARMSE = \sqrt{\frac{\sum_{n=1}^{N} w_{(n)} \times \left(M_{(n)} - truth_{(n)} - bias\right)^2}{\sum_{n=1}^{N} w_{(n)}}}. \quad (17)$$

Here, $M_{(n)}$ and $truth_{(n)}$ represent the analysis (or background) and the "truth" state at the n-th grid point, respectively. $w_{(n)}$ denotes the weighted coefficients at the n-th grid point, and $bias = \frac{\sum_{n=1}^{N} w_{(n)} \times \left(M_{(n)} - truth_{(n)}\right)}{\sum_{n=1}^{N} w_{(n)}}$ represents the systematic bias.

4.2.1 Inflation impact

In order to study the impact of the inflation technique introduced in Section 2.2.3 on the 4DEnVar system, we compared the forecast skills of the experiments initialized by the 4DEnVar system adopting different sets of inflation coefficients.

The inflation coefficients in the 4DEnVar system were set to be larger for the ensemble analyses than for the random perturbations with balance constraints, so that a significant reduction of the flow-dependent characteristic of the ensemble covariance was avoided during the assimilation cycle. We tested several sets of inflation coefficients, including $(0.1, 1.0)$, $(0.2, 0.9)$, $(0.3, 0.8)$ and $(0.5, 0.6)$, where the first and second numbers in the parentheses are the coefficients for the random perturbations and ensemble analyses, respectively, and the results of the overall scores of the experiments were given in Figure 5. The scorecard shows that the 4DEnVar system using the inflation coefficient of $(0.2, 0.9)$ has the best performance in reducing ARMSE of the forecast initialized from the ensemble mean analysis than those of tests using other sets of inflation coefficients. A proper inflation improve the ensemble spread, but too strong inflation may lead to excessive spread of some variables in the ensemble. These suggest that the inflation technique can help the 4DEnVar system reduce the forecast errors.

4.2.2 Analysis error

To facilitate comparisons with reanalysis data that are located at the middle of the assimilation window, all background and analysis fields from both DA systems are transformed from the beginning to the middle of the window through 3-h forecasts using the $0.5° \times 0.5°$

version of GRAPES-GFS, which is similar to Zhang et al. (2019). Figure 6 shows the vertical profiles of the ARMSEs of the background and analysis fields from the 4DVar and 4DEnVar systems relative to the "truth" state. On one hand, comparing with the background fields, the analysis fields from both assimilation approaches basically improve most variables at most vertical levels. These analyses significantly reduce the ARMSE of zonal wind at almost all vertical levels in the Northern Extratropics, Southern Extratropics, East Asia and Tropics (Figures 6a-6d). No significant differences of temperature between the backgrounds and analyses can be observed (Figures 6e-6h) except that the 4DEnVar system improves the temperature in the middle and higher troposphere in the Southern Extratropics (Figure 6f) and the temperatures from the analyses in the lower troposphere in the East Asia by both the 4DEnVar and 4DVar systems are obviously degraded (Figure 6g). As for the specific humidity, no distinct changes from the backgrounds to analyses can be found except the degradations near the surface by both assimilation approaches (Figures 6i-6l). On the other hand, 4DEnVar fully outperforms 4DVar on the backgrounds and analyses of zonal wind, temperature and specific humidity in the aforementioned four regions. 4DEnVar makes the biggest improvement in zonal wind (temperature) in the stratosphere in the East Asian (Tropics) relative to 4DVar (Figures 6c and 6h). Significant improvements in specific humidity by 4DEnVar are mainly in the lower troposphere comparing with 4DVar.

The analysis error structures of the 4DEnVar and 4DVar experiments are very similar (Figure 7 left and middle), which are also consistent with the analysis error structures of the 3DVar experiment and the corresponding 3D hybrid assimilation experiment in Kleist et al. (2015a). As shown in Figures 7a-7b, the zonal wind error maxima are distributed in the middle and upper troposphere at middle latitudes in the Southern Extratropics, and large zonal wind errors even extend to the lower troposphere near 60 °S. Compared with 4DVar, 4DEnVar reduces the analysis errors of zonal wind mainly at the latitudes between 60°S and 60°N, although it increases the analysis errors in the middle and higher troposphere at high latitudes in the Northern Extratropics and the analysis errors in the stratosphere at middle and high latitudes in the Southern Extratropics (Figure 7c). Large temperature errors in the analyses of both assimilation approaches are concentrated in the lower troposphere, especially in the region from the Antarctica to 60°S, which extend to the middle and upper troposphere near 60°S (Figures 7d-7e). 4DEnVar has smaller ARMSEs of temperature than 4DVar over most latitudes except for lower troposphere at high latitudes and stratosphere at the latitudes around 60°S in the Southern Extratropics (Figure 7f).

Specific humidity shows analysis error structures quite different from the zonal wind and temperature, which have semicircular shapes located between 60°S and 60°N in the lower and middle troposphere (Figures 7g-7h). In the regions large humidity errors locate at, 4DEnVar improves the accuracies of almost all humidity analyses (Figure 7i). In a word, 4DEnVar reduces most analysis errors of zonal wind, temperature and specific humidity in comparison to 4DVar.

4.2.3 Forecast skill

From the above discussions, it can be found that the analysis accuracy of the 4DEnVar system is basically higher than that of the 4DVar system. Based on these encouraging results, our attention is now drawn to the impact of these more realistic analysis ICs on the forecasts. We want to know whether the improved analysis IC can lead to improved forecasts. For this reason, the analysis fields at 0900 UTC covered the period from 0900 UTC 13 September 2016 to 0900 UTC 11 October 2016 produced by the 4DEnVar and 4DVar systems were used as ICs to conduct a set of 10-day forecasts. Similar to the analyses that were extended to the middle of the assimilation window for evaluation, these forecasts with 3-hour extension for each lead forecast day were used for evaluation. Because the 4DEnVar is an ensemble-based assimilation approach that produced 60 analysis ICs in the OSSE, 60 sets of 10-day forecasts were obtained using these analysis ICs. For convenience of comparing with the single set of 10-day forecast initialized from the 4DVar analysis, the ensemble mean 10-day forecast initialized from the 60 sets of 4DEnVar analyses was used. Also, for a more comprehensive comparison with 4DVar, the results of the deterministic forecast initialized from the 4DEnVar ensemble mean analysis were given. The forecasts were evaluated using the "truth" state as the reference and adopting the anomaly correlation coefficient (ACC) and ARMSE as the metrics.

ACC is one of the important metrics to investigate the skill of a forecast, which is used to qualitatively measure the similarity between the anomalies of this forecast and the "truth" state. In terms of this metrics, the 4DEnVar-based 10-day deterministic and ensemble mean forecasts of 500hPa geopotential height have comparable skills to and higher skills than the 4DVar-based forecast on most lead forecast days, respectively (Figure 8). In the Northern Extratropics, the 4DEnVar-based deterministic and 4DVar-based forecasts have comparable skills on the lead days 1-5, and the former has slightly higher skills on the lead days 6-8 and slightly lower skills on the lead days 9-10 (Figure 8a). In contrast, in the Southern Extratropics (Figure 8b), where

observations are much sparser than in the Northern Extratropics (Figure 3), the former has skills comparable to or even slightly lower than the latter on the first 6 lead days and slightly higher skills than the latter on the lead days 7-10. Similar to the Northern Extratropics, the 4DEnVar-based deterministic forecast has comparable or even slightly higher skills on the lead days 1-7, but slightly lower skills on the lead days 8-10 in the East Asia (Figure 8c). In the Tropics, the 4DEnVar-based deterministic forecast has lower skills on the first 5 lead days, and slightly higher skills on the lead days 6-10 than the 4DVar-based forecast (Figure 8d). In comparison, the 4DEnVar-based ensemble mean forecast has significantly higher forecast skills on almost all lead forecast days than the 4DEnVar-based deterministic and 4DVar-based forecasts (Figures 8a-8d). In particular, the most significant improvements on the last few days in the Southern Extratropics can be easily sighted in the 4DEnVar-based ensemble mean forecast comparing with the 4DVar-based forecast. In summary, more accurate ICs from 4DEnVar generally achieve to higher forecast skills in the ensemble mean forecast on almost all lead days than those from 4DVar in the single forecast.

ARMSE is also an indispensable metrics to evaluate the skill of a forecast, which is used to quantitatively measure the difference between the anomalies of this forecast and the "truth" state. To facilitate the comparison between 4DEnVar and 4DVar on their contributions to forecast skill, the difference of ARMSE between 4DEnVar and 4DVar is used, on which the confidence test is conducted. The difference of ARMSE with a negative (positive) value indicates a further improvement (degradation) of the forecast by 4DEnVar comparing with that by 4DVar. Figure 9 shows the differences of ARMSE of the 500hPa geopotential height forecasts between 4DEnVar and 4DVar. The skill of the 4DEnVar-based forecast under this metrics basically matches that under the metrics of ACC, i.e., the 4DEnVar-based deterministic and ensemble mean forecasts have comparable performance to and better performance than the 4DVar-based forecast on most lead days, respectively. This deterministic forecast has a performance comparable to or even better than the 4DVar-based forecast on the first 9 lead days (the first 7 lead days), while the errors are larger on the last 1 lead day (the last 3 lead days) in the Northern Extratropics (East Asia) as shown in Figures 9a and 9c. In contrast, the former has larger errors on the first 6 lead days (the first 4 lead days) and smaller errors on the last few days than the latter in the Southern Extratropics (Tropics) from Figures 9b and 9d. In addition, the 4DEnVar-based ensemble mean forecast reduce errors significantly relative to the two abovementioned deterministic forecasts on almost all lead

days. In particular, the most significant improvements by the 4DEnVar-based ensemble mean forecast relative to the 4DVar-based forecast can be sighted on the last few lead days in the Southern Extratropics (Figure 9b), suggesting that the analysis ICs from the 4DEnVar system may have much better capability to reduce the forecast errors of 500hPa geopotential height in the regions with sparse observations. Due to the maximum improvements by 4DEnVar in the ensemble mean forecast of the 500hPa geopotential height on the lead day 10 in both the Northern and Southern Extratropics comparing with the 4DVar-based forecast (Figures 8a-8b and 9a-9b), the horizontal distributions of the forecasts on this lead day were also analyzed and compared. Figure 10 shows the "truth" state and the 240-h forecasts of the 500hPa geopotential height respectively initialized from the 4DVar and 4DEnVar analyses on 1200 UTC 13 September 2016 in the Northern Extratropics. In the "truth" state, a low-pressure system is distributed around the Arctic, with three troughs near 75°E, 160°W and 60°W, respectively. In addition, there is a high value center at middle and high latitudes near the longitude 0° (Figure 10a). The main circulation situations in the 4DVar-based (4DEnVar-based deterministic) forecast are basically similar to those in the "truth" state (Figures 10b-10c), although the low-pressure trough (high value center) near 160°W (the longitude 0°) is not correctly presented. Moreover, the main circulation situations in the 4DEnVar-based ensemble mean forecast is similar to the 4DEnVar-based deterministic forecast. As shown in Figure 10e, the 4DVar-based forecast has large errors with a "negative-positive-negative" distribution between 90°W and 60°E at middle and high latitudes. It also presents significant positive errors near 160°W at middle latitudes. In comparison, the 4DEnVar-based deterministic forecast mainly reduces the magnitude of the errors in the regions between 90°E to 30°W at middle and high latitudes, especially the regions near 160°W and 40°W at middle and high latitudes, but increases the errors in the regions between 30°W and 60°E at middle and high latitudes, especially the regions near 30°E at high latitude (Figure 10f). In contrast, compared with the 4DVar-based and 4DEnVar-based deterministic forecasts, the 4DEnVar-based ensemble forecast significantly reduces errors in almost all regions, except the region between 30°E to 60°E. Similar to Figure 10, Figure 11 shows the results in the Southern Extratropics. There is a low-pressure system near the Antarctica, which extends out three troughs near 90°W, 150°E and 60°E, respectively, in the "truth" state (Figure 11a). The 4DVar-based forecast and the 4DEnVar-based deterministic and ensemble mean forecasts basically represent the circulation situations in the "truth" state (Figures 11b-11d), but the first does not capture the troughs at 90°W very well. The

forecast errors of 4DVar in the Southern Extratropics are significantly larger than in the Northern Extratropics (Figure 11e), while the 4DEnVar-based deterministic forecast reduces almost all the significant forecast errors of 4DVar except the regions between 90°E and 0° (Figure 11f). In addition, the 4DEnVar-based ensemble mean forecast significantly reduces the errors compared with the 4DVar-based and 4DEnVar-based deterministic forecasts (Figure 11g). Overall, in terms of the 500hPa geopotential height forecast, the 4DEnVar-based deterministic forecast has comparable performance to the 4DVar-based forecast in the Northern Extratropics and slightly better performance than the 4DVar-based forecast in the Southern Extratropics. Moreover, the 4DEnVar-based ensemble mean forecast significantly reduce the errors, especially in the Southern Extratropics, which is consistent with the conclusions obtained by the ACC and ARMSE metrics.

The 4DEnVar-based forecast of geopotential height also has similar performances at most other vertical levels (Figure 12). The 4DVar-based forecast errors show the largest in the Southern Extratropics, followed by the Northern Extratropics and East Asia, and the smallest in the Tropics. In the Northern Extratropics, the 4DVar-based forecast has significant errors at the vertical levels between 400hPa and 200hPa, which keeps increasing following the lead time and reaches the maximum on the lead day 10. These errors gradually extend to lower and upper levels following the lead time (Figures 12a, 12d, 12g and 12j). As shown in Figure 12b, the 4DEnVar-based deterministic forecast has a similar error structure to the 4DVar-based forecast except for larger errors in the stratosphere, which extends to the troposphere on the lead days 1 and 10, and smaller errors in the lower stratosphere and middle troposphere on the lead days 5-8, which extends to the lower troposphere on the lead days 7-8. The maximum improvements (degradations) by the 4DEnVar-based deterministic forecast comparing with 4DVar are located between 400hPa and 200hPa (above 100hPa) on the lead days 6-7 (the lead days 1-9). In the Southern Extratropics, the 4DEnVar-based deterministic forecast has larger errors on the lead days 1-6 than the 4DVar-based one, especially in the stratosphere on almost all lead days, and smaller errors on the last 3 lead days (Figure 12e). In the East Asia, the 4DEnVar-based deterministic forecast has a performance comparable to and even better than the 4DVar-based one on the lead days 1-7 except in the upper stratosphere. However, it has larger errors on the last 3 lead days, especially between 600-100hPa (Figure 10h). In the Tropics, it has a degradation near the stratosphere (the lower and middle troposphere) on the lead days 1-9 (1-3) as shown in Figure 12k. In contrast, the 4DEnVar-based ensemble mean forecast reduces errors at almost all regions and vertical levels on the middle and

later lead days compared with the 4DVar-based forecast, especially on the later days in the Northern and Southern Extratropics, where the errors of the 4DVar-based forecast are the largest. The places where the 4DEnVar-based deterministic forecast becomes worse than the 4DVar-based forecast are largely improved in the ensemble mean forecast, especially for the lead days 8-10 in the East Asia (Figures 12c, 12f, 12i and 12l).

The 4DVar-based zonal wind forecast has an error structure similar to the geopotential height forecast at most vertical levels with the largest errors in the Southern Extratropics and smallest errors in the Tropics (Figures 13a, 13d, 13g and 13j). The improvement and degradation locations of the 4DEnVar-based deterministic forecast compared with the 4DVar-based forecast are generally consistent with the geopotential height forecast (Figures 13b, 13e, 13h and 13k). However, quite different from the geopotential height forecast, the 4DEnVar-based deterministic forecast of zonal wind reduces the errors relative to the 4DVar-based forecast at almost all vertical levels on the first lead day and the improvement extends to the lead day 9 (the lead day 7) in the middle and upper troposphere of the Northern Extratropics (East Asia) as shown in Figure 13b, 13e, 13h and 13k. In contrast, the 4DEnVar-based ensemble mean forecast performs better in almost all situations than the 4DVar-based forecast, especially on the middle and later lead days (Figures 13c, 13f, 13i and 13l). In addition, the locations where the 4DEnVar-based deterministic forecast deteriorates compared with 4DVar largely get improved in the 4DEnVar-based ensemble mean forecast, especially for the lead days 2-7 in the Southern Extratropics and the lead days 8-10 in the East Asia (Figures 13f and 13i). However, the 4DEnVar-based ensemble mean forecast is still worse than the 4DVar-based forecast on the first 5 lead days around the stratosphere in the Southern Extratropics and Tropics, which is similar to the geopotential height forecast (Figures 12f and 12l).

The error distribution of the 4DVar-based temperature forecast is not quite the same as those of the geopotential height and zonal wind forecasts, with the error size sorting as same as in geopotential height and zonal wind forecasts (Figures 14a, 14d, 14g and 14j). The locations of the improvements and degradations in the 4DEnVar-based deterministic temperature forecast relative to the 4DVar-based forecast is similar to the zonal wind forecast but with the following differences. The locations of the largest improvements in the 4DEnVar-based deterministic forecast in the Southern Extratropics and East Asia compared with the 4DVar-based forecast extend to the lower and middle troposphere. In addition, the ranges of the degradations of the 4DEnVar-based

deterministic forecast compared with the 4DVar-based forecast also increase (Figures 14b, 14e, 14h and 14k). In contrast, the 4DEnVar-based ensemble mean forecast is still significantly improved for almost all situations compared with 4DVar (Figures 14c, 14f, 14i and 14l).

The 4DVar-based specific humidity forecast has an error structure different from other variables, with large errors between 900hPa and 700hPa in each of four regions, increasing with the lead time and reaching a maximum since the lead day 9 (Figures 15a, 15d, 15g and 15j). Compared with the 4DVar-based forecast, the 4DEnVar-based deterministic forecast reduces errors on the lead days 1-7 in the Northern Extratropics and East Asia, while degradation occurs on the lead days 8-10 (Figures 15b and 15h). The 4DEnVar-based deterministic forecast reduces errors in the Southern Extratropics (Tropics) except the lead days 3-7 (4-10) compared with 4DVar (Figures 15e and 15k). In contrast, 4DEnVar-based ensemble mean forecast stably reduces errors for almost all regions and lead days, and the largest improvements can be found where and when large forecast errors of 4DVar are located (Figures 15c, 15f, 15i and 15l) .

4.2.4 Computational efficiency

The computational efficiency of the 4DEnVar system is also a key concern. Taking the 6-h assimilation window (0900 UTC 11 September 2016 - 1500 UTC 11 September 2016) as an example, the computational time was about 25 minutes for a 4DVar DA using 480 cores on the high-performance computer PI-SUGON of the China Meteorological Administration. In comparison, the 4DEnVar system took only 13 minutes since the ensemble members of the 4DEnVar system are independent and all members can be analyzed concurrently using a total of 60×480 cores. The aforementioned results may be slightly impacted by several factors, such as the high-performance computer state, but overall, the 4DEnVar system has the advantage of good parallel efficiency and scalability and is thereby timesaving when more computational resources are used.

5 Summary and discussion

In this study, a new 4DEnVar DA system was developed based on the DRP-4DVar approach. This system is novel in some aspects. It introduces the idea of ensemble into the variational framework to achieve the time-variant BEC and minimizes the cost function without using the ADM. It can easily take full advantage of the mature variational framework and

implicitly include the hybrid BEC into the pure ensemble covariance. An inflation method similar to RTPP (Zhang et al., 2004) based on balanced random perturbations produced using the static B-matrix of 4DVar is applied to alleviate the filter divergence during the assimilation cycle, which can be conveniently and efficiently implemented. Moreover, a limited number of leading eigenvectors of the localization correlation function are used to perform the localization of the B-matrix and rapidly increase the ensemble size without any extra model integrations.

Preliminary tests including SOEs and OSSEs were conducted to evaluate the performance of the 4DEnVar system, using the 4DVar system as a reference for comparison. The OSSE-based one-month DA cycles using both systems were conducted, in which the first 2-day cycles were used for spin-up. The SOEs were conducted at and the OSSE results were evaluated and compared since the ninth assimilation window. The SOEs show that both 4DEnVar and 4DVar assimilated the single-point observation effectively and satisfied certain balance constraints. Moreover, 4DEnVar using the ensemble BEC exhibits obvious flow-dependent features.

In the OSSEs, we first tuned the inflation coefficients. The weight of the ensemble analyses kept larger than that of the balanced random perturbations to avoid too much loss of flow-dependent information during the assimilation cycle. It is found that the forecast errors grow when the inflation is too weak or too strong. Finally, the weights 0.9 for the ensemble analysis and 0.2 for the balanced random perturbations were thought to be the proper inflation coefficients, which optimally reduced the ARMSEs of the forecast initialized by the 4DEnVar ensemble mean analysis.

The results of the OSSE-based one-month DA cycle show that the 4DEnVar and 4DVar analyses can significantly improve the dynamic variables such as the zonal wind, and has slight impacts on the thermodynamic variables such as the temperature and specific humidity compared with their backgrounds. In addition, 4DEnVar outperforms 4DVar in terms of ARMSE on the background and analysis fields of model variables. Compared with the 4DVar analysis, the 4DEnVar analysis can significantly reduce the errors in the model variables at low and middle latitudes. The greatest improvements in the zonal wind and temperature by the 4DEnVar analysis located mainly near the stratosphere, although there is significant degradation at high latitudes in the Southern Extratropics. The greatest improvements in the specific humidity are mainly in the lower troposphere. In conclusion, the analysis error of the 4DEnVar system is basically smaller than that of the 4DVar system.

The effect of the 4DEnVar system on the forecast is also evaluated. From the ACC and ARSME indicators of the 500hPa geopotential height forecasts, the 4DEnVar-based deterministic forecast has a performance comparable to or slightly better than the 4DVar-based forecast in the short and medium range in the Northern Extratropics and East Asia, and slightly worse performance in the long range. In the Southern Extratropics and Tropics, the results are roughly opposite to those in the Northern Extratropics and East Asia. Moreover, the 4DEnVar-based ensemble mean forecast is comparable to or even better than the 4DVar-based forecast except for the first few days in the Tropics. The geopotential height forecasts at other vertical levels are generally consistent with the results of the 500hPa geopotential height, but both the deterministic and ensemble mean forecasts from the 4DEnVar system show significant degradations near the stratosphere.

The 4DEnVar-based deterministic forecasts of the zonal wind and temperature are similar to and better than that of the geopotential height. In particular, these deterministic forecasts outperform the 4DVar-based forecasts in almost all regions and vertical levels on the first lead day, and this positive effect continues in the upper troposphere in the Northern Extratropics and East Asia to the middle and later lead days, but the temperature forecasts are slightly worse than the zonal wind. The improvements of the 4DEnVar-based deterministic specific humidity forecast over the 4DVar-based one are mainly located on the early lead days. In contrast, the 4DEnVar-based ensemble mean forecast of the zonal wind, temperature and specific humidity show significant improvements over both the 4DVar and 4DEnVar-based deterministic forecasts, with the greatest improvements on the later lead days.

Overall, the 4DEnVar system shows great promise in terms of reducing the analysis errors and producing high quality ensemble forecasts. The significantly improved ensemble forecasts from the 4DEnVar system suggest that they have the potential to provide high quality flow-dependent ensemble BECs for hybrid DA systems. Moreover, 4DEnVar is more timesaving than 4DVar given that the ensemble members of 4DEnVar can be predicted and analyzed concurrently. There is still much room for further improving the performance of the 4DEnVar system. For example, the current localization used here does not include the balance constraints and a balanced localization calls for further study. The localization length-scales currently adopted in the 4DEnVar system are constant and adaptive length-scales are expected in the future. The current 4DEnVar system borrows the low-resolution TLM of the 4DVar system and the NLM with the

same resolution may provide more accurate and more timesaving ensemble forecasts. In addition, assimilation experiments using real observations, especially satellite radiance observations, should be further carried out to evaluate the performance of the 4DEnVar system.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (2018YFC1506700). The assimilation and forecast experiments were performed on the high-performance computer PI-SUGON of the China Meteorological Administration. The sounding and cloud-derived wind observations were provided by the Global Telecommunications System (https://public.wmo.int/en/programmes/global-telecommunication-system). The ERA-5 reanalyses were downloaded from https://apps.ecmwf.int/data-catalogues/era5/?class=ea&stream=oper&expver=1&type=an. The 6-h forecasts of ERA-Interim dataset and the ERA-Interim reanalyses were downloaded from https://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/. We thank the editor and reviewers for giving helpful suggestions to improve the manuscript.

References

Barker, D. M. (2005). Southern high-latitude ensemble data assimilation in the Antarctic mesoscale prediction system. *Monthly Weather Review, 133*(12), 3431-3449.

Bishop, C. H., Hodyss, D., Steinle, P., Sims, H., Clayton, A. M., Lorenc, A. C., Barker, D. M., & Buehner, M. (2011). Efficient ensemble covariance localization in variational data assimilation. *Monthly Weather Review, 139*, 573-580.

Bonavita, M., Isaksen, L., & Hólm, E. (2012). On the use of EDA background error variances in the ECMWF 4D-Var. *Quarterly Journal of the Royal Meteorological Society, 138*(667), 1540-1559.

Bowler, N. E., Clayton, A. M., Jardak, M., Jermey, P. M., Lorenc, A. C., Wlasak, M. A., Barker, D. M., Inverarity, G. W., & Swinbank, R. (2017a). The effect of improved ensemble covariances on hybrid variational data assimilation. *Quarterly Journal of the Royal Meteorological Society, 143*: 785-797.

Bowler, N. E., Clayton, A. M., Jardak, M., Lee, E., Lorenc, A. C. and Piccolo, C., Pring, S. R., Wlasak, M. A., Barker, D. M., Inverarity, G. W., & Swinbank, R. (2017b). Inflation and localization tests in the development of an ensemble of 4D-ensemble variational assimilations. *Quarterly Journal of the Royal Meteorological Society*, *143*(704): 1280-1302.

Buehner, M., Houtekamer, P. L., Charette, C., Mitchell, H. L., & He, B. (2010a). Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: Description and single-observation experiments. *Monthly Weather Review, 138*(5), 1550-1566.

Buehner, M., Houtekamer, P. L., Charette, C., Mitchell, H. L., & He, B. (2010b). Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: One-month experiments with real observations. *Monthly Weather Review, 138*(5), 1567-1586.

Buehner, M., Morneau, J., & Charette, C. (2013). Four-dimensional ensemble-variational data assimilation for global deterministic weather prediction. *Nonlinear Processes in Geophysics, 20*(5), 669-682.

Buehner, M., Mctaggart-Cowan, R., Beaulne, A., Charette, C., Garand, L., Heilliette, S., Lapalme, E., Laroche, S., Macpherson, S. R., Morneau, J., & Zadra, A. (2015). Intercomparison of deterministic weather forecasting systems based on ensemble–variational data assimilation at Environment Canada. Part I: The global system. *Monthly Weather Review, 143*, 2532-2559.

Chen, D. H., Xue, J. S., Yang, X. S., Zhang, H. L., Shen, X. S., Hu, J. L., Wang, Y., Ji, L. R., & Chen, J. B. (2008). New generation of multi-scale NWP system (GRAPES): general scientific design. *Chinese Science Bulletin, 53*(22), 3433–3445.

Clayton, A. M., Lorenc, A. C., & Barker, D. M. (2013). Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Quarterly Journal of the Royal Meteorological Society, 139*(675), 1445-1461.

Courtier, P., Thépaut, J. N., & Hollingsworth, A. (1994). A strategy for operational implementation of 4D-Var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society, 120*(519), 1367-1387.

785 Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using
786 Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*
787 *Oceans, 99*(C5), 10143–10162.

788 Evensen, G. (2003). The ensemble Kalman filter: Theoretical formulation and practical
789 implementation. *Ocean Dynamics, 53*(4), 343-367.

790 Gaspari, G., & Cohn, S. E. (1999). Construction of correlation functions in two and three
791 dimensions. *Quarterly Journal of the Royal Meteorological Society, 125*(554), 723-757.

792 Gauthier, P., & Thépaut, J. N. (2001). Impact of the digital filter as a weak constraint in the
793 preoperational 4dvar assimilation system of Météo-France. *Monthly Weather Review, 129*(8),
794 2089-2102.

795 Gauthier, P., Tanguay, M., Laroche, S., & Pellerin, S. (2007). Extension of 3dvar to 4dvar:
796 Implementation of 4dvar at the meteorological service of Canada. *Monthly Weather*
797 *Review, 135*(6), 2339-2354.

798 Hamill, T. M., & Snyder, C. (2000). A hybrid ensemble Kalman filter–3d variational analysis
799 scheme. *Monthly Weather Review, 128*(8), 2905-2919.

800 Hamill, T. M., Whitaker, J. S., & Snyder, C. (2001). Distance-dependent filtering of background
801 error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review, 129*(11),
802 2776-2790.

803 He, Y., Wang, B., Liu, M., Liu, L., Yu, Y., Liu, J., Li, R., Zhang, C., Xu, S., Huang, W., Liu, Q.,
804 Wang, Y., & Li, F. (2017). Reduction of initial shock in decadal predictions using a new
805 initialization strategy. *Geophysical Research Letters, 44*(16), 8538-8547, DOI:
806 10.1002/2017GL074028.

807 He, Y., Wang, B., Liu, L., Huang, W., Xu, S., Liu, J., Wang, Y., Li, L., Huang, X., Peng, Y., Lin,
808 Y., & Yu, Y. (2020a). A DRP-4DVar-based coupled data assimilation system with a simplified
809 off-line localization technique for decadal predictions. *Journal of Advances in Modeling Earth*
810 *Systems, 12*, e2019MS001768. https://doi.org/10.1029/2019MS001768.

811 He, Y., Wang, B., Huang, W., Xu, S., Wang, Y., Liu, L., Li, L., Liu, J., Yu, Y., Lin, Y., Huang,
812 X., & Peng, Y. (2020b). A new DRP-4DVar-based coupled data assimilation system for

decadal predictions using a fast online localization technique. *Climate Dynamics, 54*(19), 3541-3559, DOI: 10.1007/s00382-020-05190-w.

Houtekamer, P. L., Mitchell, H. L., Pellerin, G., Buehner, M., Charron, M., Spacek, L., & Hansen, B. (2005). Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Monthly Weather Review, 133*(3), 604-620.

Kleist, D. T., & Ide, K. (2015a). An OSSE-based evaluation of hybrid variational-ensemble data assimilation for the NCEP GFS. Part I: System description and 3D-hybrid results. *Monthly Weather Review, 143*, 433-451.

Kleist, D. T., & Ide, K. (2015b). An OSSE-based evaluation of hybrid variational–ensemble data assimilation for the NCEP GFS. Part II: 4DEnVar and hybrid variants. *Monthly Weather Review, 143*(2), 452-470.

Koizumi, K., Ishikawa, Y., & Tsuyuki, T. (2005). Assimilation of precipitation data to the JMA mesoscale model with a four-dimensional variational method and its impact on precipitation forecasts. *Sola, 1*, 45-48.

Kuhl, D. D., Rosmond, T. E., Bishop, C. H., Mclay, J., & Baker, N. L. (2013). Comparison of hybrid ensemble/4dvar and 4dvar within the NAVDAS-AR data assimilation framework. *Monthly Weather Review, 141*(8), 2740-2758.

Le Dimet, F. X., & Talagrand, O. (1986). Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus A: Dynamic Meteorology and Oceanography, 38*(2), 97-110.

Lei, L., Whitaker, J. S., Anderson, J. L., & Tan, Z. (2020). Adaptive localization for satellite radiance observations in an ensemble Kalman filter. *Journal of Advances in Modeling Earth Systems, 12*, e2019MS001693. https://doi.org/ 10.1029/2019MS001693

Lewis, J. M., & Derber, J. C. (1985). The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus A: Dynamic Meteorology and Oceanography, 37*(4), 309-322.

Li, F., Wang, B., He Y., et al. (2021a). Important role of North Atlantic air–sea coupling in the interannual predictability of summer precipitation over the eastern Tibetan Plateau. *Climate Dynamics, 56*, 1433-1448.

Li, F., Wang, B., He Y., et al. (2021b). Improved decadal predictions of East Asian summer monsoon with a weakly coupled data assimilation scheme. *International Journal of Climatology*, 1-22.

Liu, C., Xiao, Q., & Wang, B. (2008). An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test. *Monthly Weather Review, 136*(9), 3363-3373.

Liu, C., Xiao, Q., & Wang, B. (2009). An ensemble-based four-dimensional variational data assimilation scheme. Part II: Observing system simulation experiments with advanced research WRF (ARW). *Monthly Weather Review, 137*(5), 1687-1704.

Liu, J., & Wang, B. (2011). Rainfall assimilation using a new four-dimensional variational method: A single-point observation experiment. *Advances in Atmospheric Sciences*, *28*(4), 735-742.

Lorenc, A. C. (2003). The potential of the ensemble Kalman filter for NWP—a comparison with 4D-Var. *Quarterly Journal of the Royal Meteorological Society, 129*(595), 3183-3203.

Lorenc, A. C., Bowler, N. E., Clayton, A. M., Pring, S. R., & Fairbairn, D. (2015). Comparison of hybrid-4DEnVar and hybrid-4DVar data assimilation methods for global NWP. *Monthly Weather Review, 143*(1), 212-229.

Qiu, C., Zhang, L., & Shao, A. (2007). An explicit four-dimensional variational data assimilation method. *Science in China Series D: Earth Sciences, 50*(8), 1232-1240.

Rabier, F., Jarvinen, H., Klinker, E. Mahfouf J. F., & Simmons, A. (2000). The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Quarterly Journal of the Royal Meteorological Society, 126*(564), 1143-1170.

Rawlins, F., Ballard, S. P., Bovis, K. J., Clayton, A. M., Li, D., Inverarity, G. W., Lorenc, A. C., & Payne, T. J. (2007). The Met Office global four-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, *133*(623), 347-362.

Raynaud, L., Berre, L., & Desroziers, G. (2011). An extended specification of flow-dependent background error variances in the Météo-France global 4D-Var system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 607-619.

Shen, S., Liu, J., & Wang, B. (2015). Evaluation of the historical sampling error for global models. *Atmospheric and Oceanic Science Letters, 8*(5), 250–256.

Shi, P., Lu, H., Leung, L., He, Y., Wang, B., Yang, K., et al. (2021). Significant land contributions to interannual predictability of East Asian summer monsoon rainfall. *Earth's Future, 9*, DOI: 10.1029/E2020EF001762.

Simmons, A. J., & Hollingsworth, A. (2002). Some aspects of the improvement in skill of numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society, 128*, 647-677.

Su, Y., Shen, X., zhang, H., & Liu, Y. (2020). A study on the three-dimensional reference atmosphere in GRAPES_GFS: Constructive reference state and real forecast experiment (in Chinese). *Acta Meteorologica Sinica, 78*(6):962-971. DOI: 10.11676/qxxb2020.075.

Tian, X., Xie, Z., & Dai, A. (2008). An ensemble-based explicit four-dimensional variational assimilation method. *Journal of Geophysical Research: Atmospheres, 113*, D21124. DOI: 10.1029/2008JD010358

Tian, X., Xie, Z., and Sun, Q. (2011). A POD-based ensemble four-dimensional variational assimilation method. *Tellus, 63A*(4), 805-816.

Wang, B., Liu, J. J., Wang, S., Cheng, W., Liu, J., Liu, C., Xiao, Q., & Kuo, Y. (2010a). An economical approach to four-dimensional variational data assimilation. *Advances in Atmospheric Sciences, 27*(4), 715-727.

Wang, B., Liu, J., Lu, B., & Tan, X. (2010b). An orthogonal expansion of filtering function in localization and its application to solution of CNOP. *Geophysical Research Abstracts, 12*, EGU2010-6786, EGU General Assembly 2010.

Wang, B., Liu, J., Liu, L, Xu, S., & Huang, W. (2018). An approach to localization for ensemble-based data assimilation. *Plos One, 13*(1), e0191088. https://doi.org/10.1371/journal.pone.0191088.

Wang, X., Barker, D. M., Snyder, C., & Hamill, T. M. (2008a). A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part I: Observing system simulation experiment. *Monthly Weather Review, 136*(12), 5116-5131.

Wang, X., Barker, D. M. , Snyder, C., & Hamill, T. M. (2008b). A hybrid ETKF–3DVAR data assimilation scheme for the WRF model. Part II: Real observation experiments. *Monthly Weather Review, 136*(12), 5132-5147.

Wang, X., Parrish, D., Kleist, D., & Whitaker, J. (2013). GSI 3DVAR-based ensemble–variational hybrid data assimilation for NCEP global forecast system: Single-resolution experiments. *Monthly Weather Review, 141*(11), 4098-4117.

Wang, X., & Lei, T. (2014). GSI-based four-dimensional ensemble–variational (4DEnsVar) data assimilation: Formulation and single-resolution experiments with real data for NCEP global forecast system. *Monthly Weather Review, 142*(9), 3303-3325.

Whitaker, J. S., Hamill, T. M., Wei, X., Song Y., & Toth Z. (2008). Ensemble data assimilation with the NCEP global forecast system. *Monthly Weather Review, 136*, 463-482.

Whitaker, J. S., Compo, G. P., & Thépaut, J. N. (2009). A comparison of variational and ensemble-based data assimilation systems for reanalysis of sparse observations. *Monthly Weather Review, 137*(6), 1991-1999.

Xue, J., Zhuang, S., Zhu, G., Zhang, H., Liu, Z., Liu, Y., & Zhuang, Z. (2008). Scientific design and preliminary results of three-dimensional variational data assimilation system of GRAPES. *Chinese Science Bulletin, 53*(22), 3446-3457.

Zhang, F. Q., Snyder, C., & Sun, J. (2004). Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble kalman filter. *Monthly Weather Review, 132*, 1238-1253.

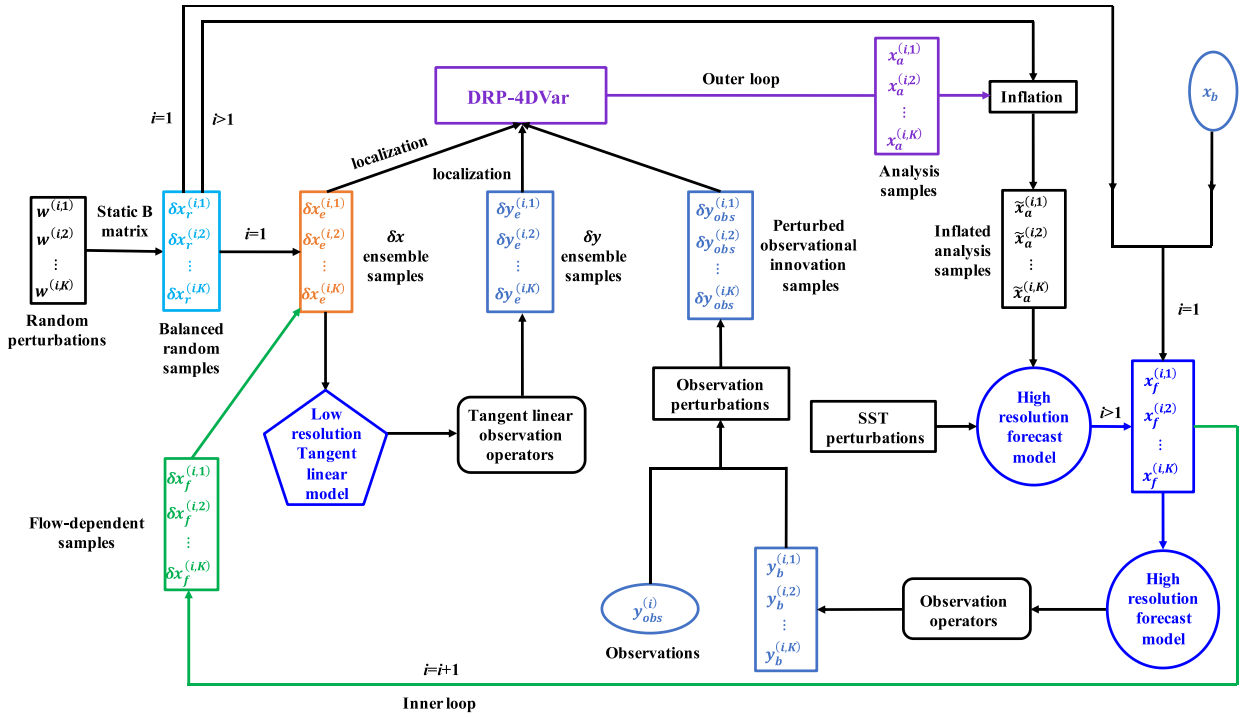
Zhang, L., Liu, Y. Z., Liu, Y., Gong, J., Lu, H., Jin, Z., Tian, W., Liu, G., Zhou, B., & Zhao, B. (2019). The operational global four-dimensional variational data assimilation system at the China Meteorological Administration. *Quarterly Journal of the Royal Meteorological Society, 145*, 1882–1896.

Zhang, M., & Zhang, F. (2012). E4dvar: Coupling an ensemble Kalman filter with four-dimensional variational data assimilation in a limited-area weather prediction model. *Monthly Weather Review, 140*(2), 587-600.

Zhao, J., & Wang, B. (2010). Sensitivity of the DRP-4DVar performance to perturbation samples obtained by two different methods. *Journal of Meteorological Research, 24*(5), 527-538.

927 Zhao, Y., Wang, B., & Liu, J. J. (2012). A DRP-4DVar data assimilation scheme for typhoon
928 initialization using sea level pressure data. *Monthly Weather Review, 140*(4), 1191-1203.

929 Zupanski, M. (1996). A preconditioning algorithm for four-dimensional variational data
930 assimilation. *Monthly Weather Review, 124*(11), 2562-2573.



931

932 **Figure 1**. Schematic of the 4DEnVar system used in this study.

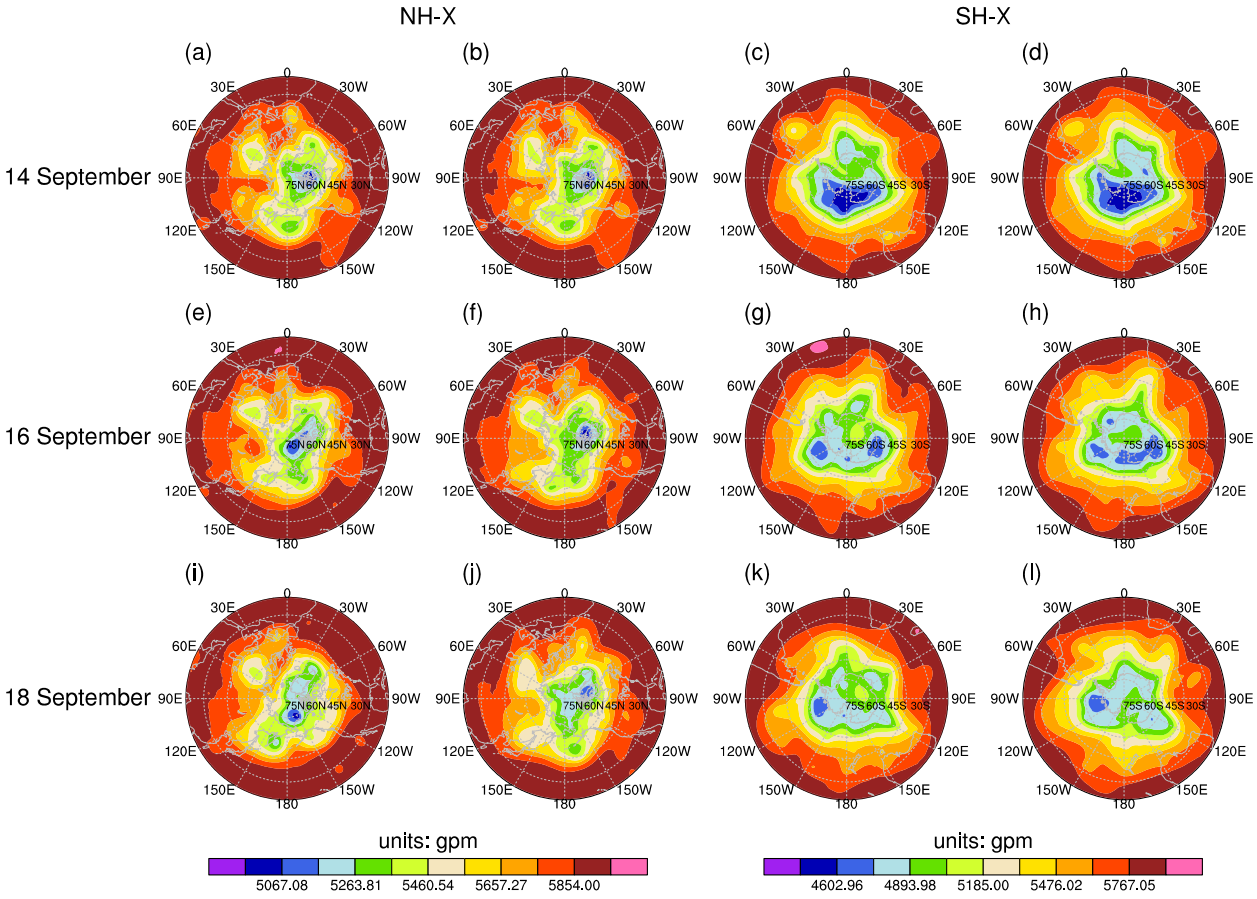


Figure 2. The 500hPa geopotential height at 1200 UTC on 14 September 2016 (top), 16 September 2016 (middle) and 18 September 2016 (bottom) from the ERA-Interim reanalysis (left) and the "truth" state (middle left) in the Northern Extratropics (20°N~90°N, 180°W~180°E), and the results in the Southern Extratropics (20°S~90°S, 180°W~180°E; the ERA-Interim reanalysis, middle right; the "truth" state, right) are plotted.

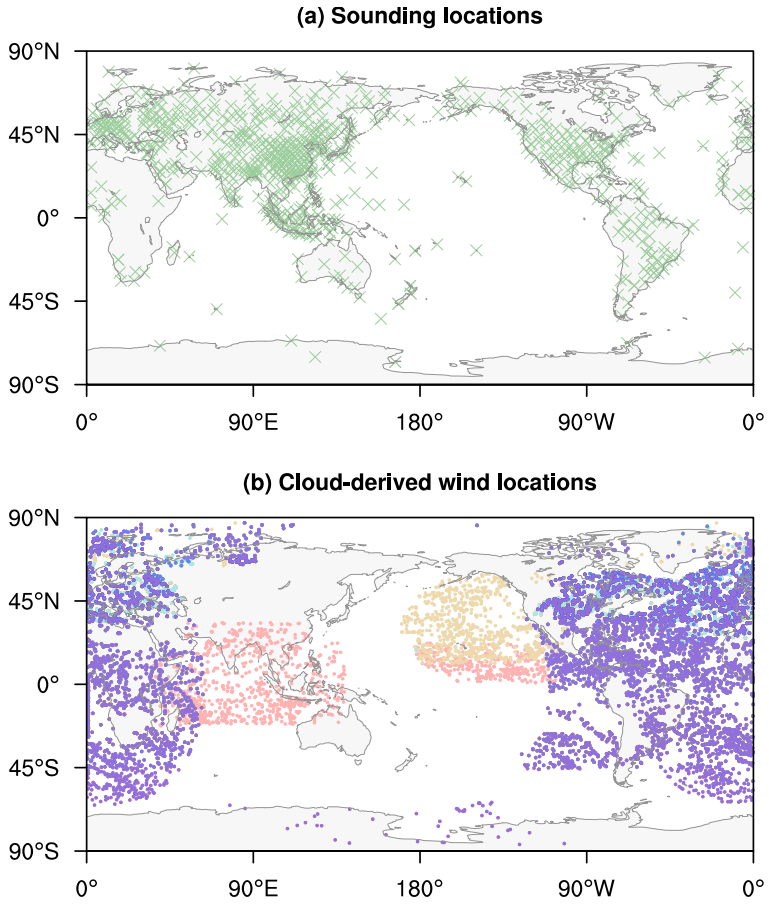


Figure 3. Locations of (a) sounding and (b) cloud-derived wind observations covered the period from 0900 UTC 13 September 2016 to 1500 UTC 13 September 2016. Different colored dots in (b) indicate different sampling times.

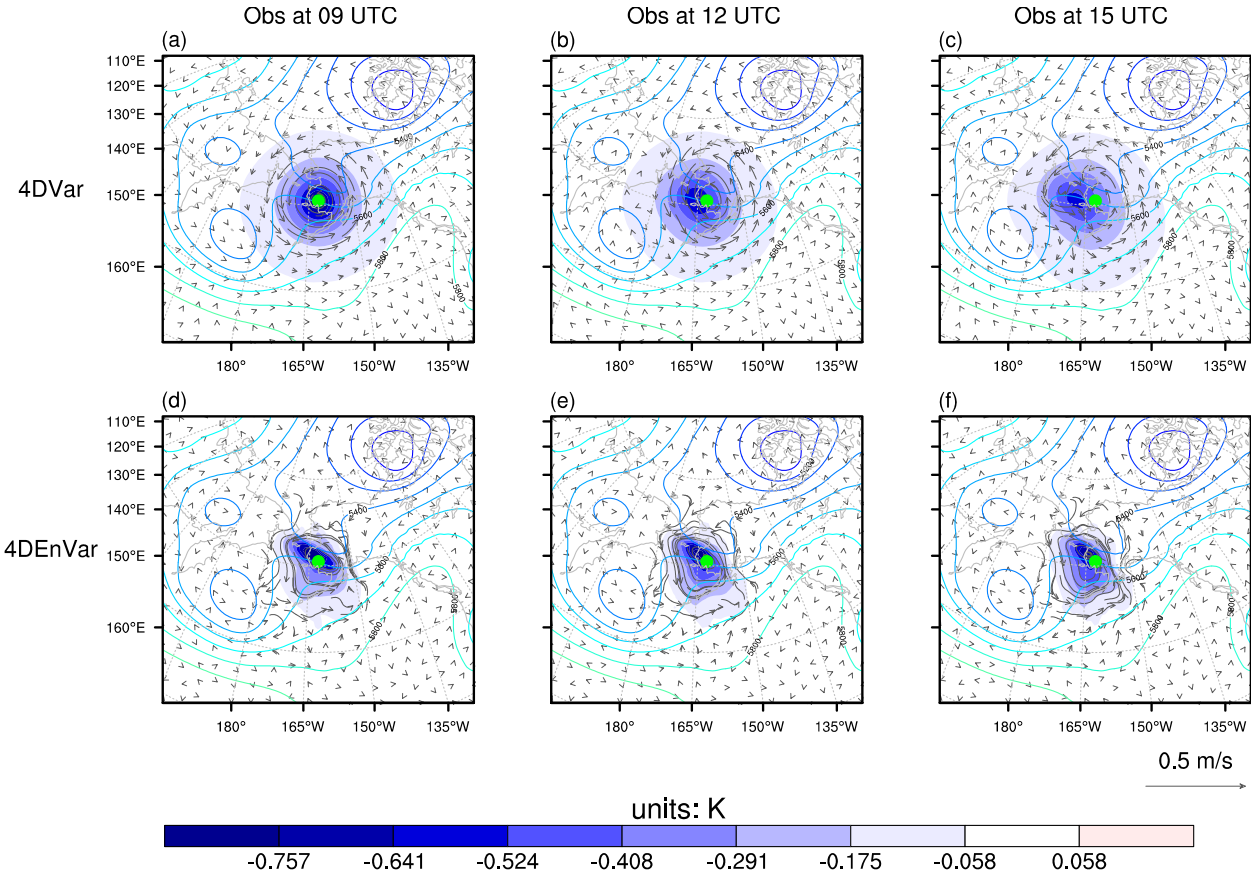


Figure 4. The temperature (shading; units: K) and vector wind (vector; units: m/s) analysis increments from assimilating the single-point temperature observations valid at the (left) beginning, (middle) middle and (right) end of the assimilation window for (top) 4DVar and (bottom) 4DEnVar on the model level closest to the single temperature observation assimilated, which locates at 500hPa (marked with a green dot). The solid contour is the 500hPa background field geopotential height (units: gpm) valid at the beginning of the assimilation window, when the analysis time is taken.

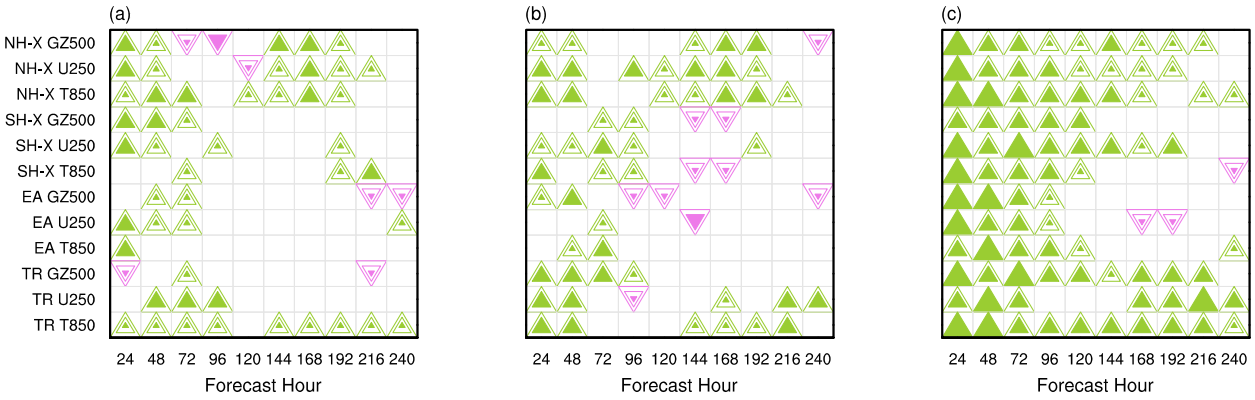
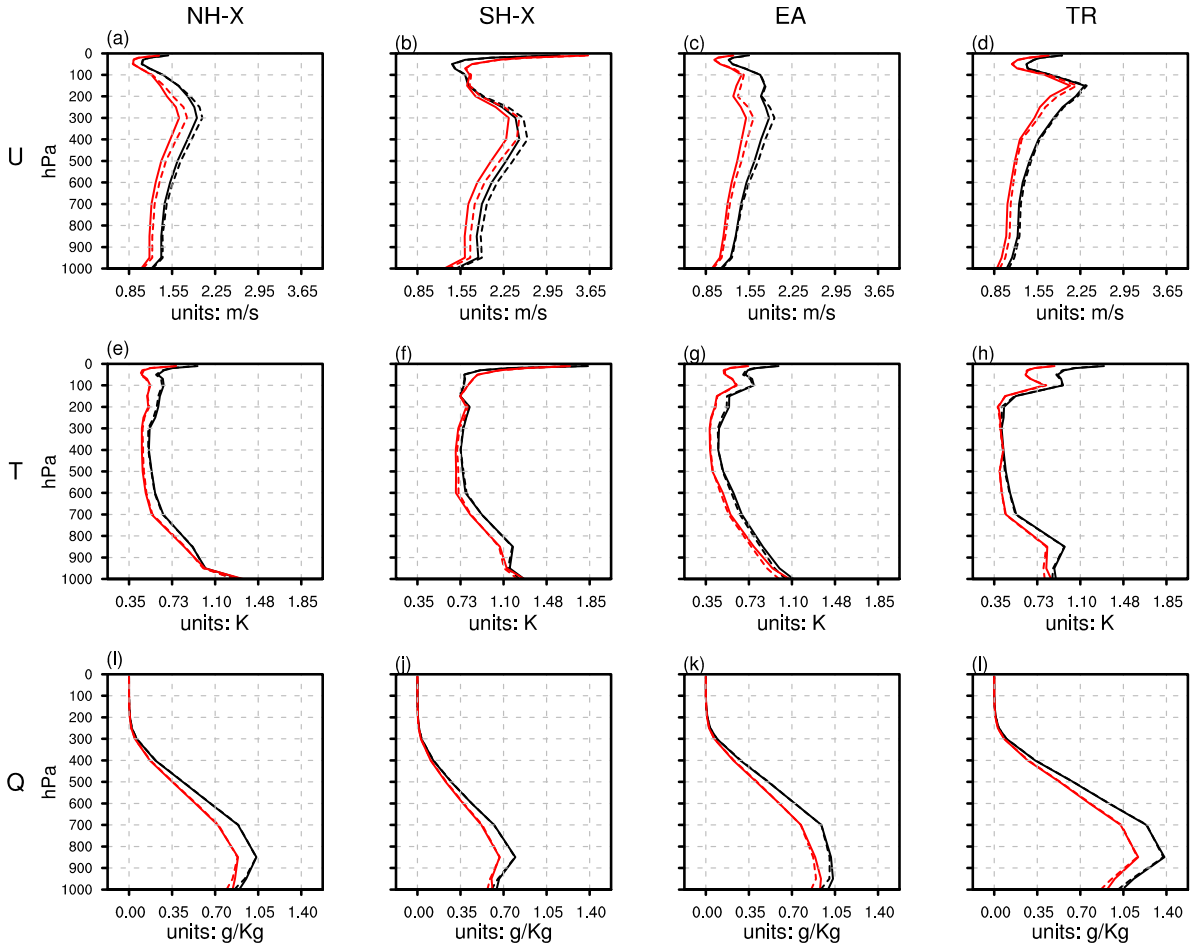


Figure 5. The scorecard of the ARMSE error mean calculated for the forecast initialized by the ensemble mean analysis from the 4DEnVar system adopting inflation coefficients of $(0.2, 0.9)$ against the forecasts from the 4DEnVar system adopting inflation coefficients of (a) $(0.1, 1.0)$, (b) $(0.3, 0.8)$ and (c) $(0.5, 0.6)$ with identical settings, respectively. If the former forecast has a significantly lower (higher) ARMSE error than the latter, then a green upward-pointing (red downward-pointing) triangle is plotted. The corresponding color outline is the three sizes that each symbol can be plotted, and the sizes from large to small corresponds to error mean differences greater than 3 times, between 1 times and 3 times, and between 0.5 times and 1 times the t value of the 95% confidence level, i.e., fairly significant, significant, and insignificant. The graph is not shown when the error mean differences are less than 0.5 times the t value of the 95% confidence level, indicating equivalent.



964

965 **Figure 6**. Vertical profiles of the ARMSE (verified relative to the "truth" state) of the background
966 (dashed line) and analysis (solid line) fields of the zonal wind (top; units: m/s), temperature
967 (middle; units: K) and specific humidity (bottom; units: g/Kg) in the Northern Extratropics
968 (20°N~90°N, 180°W~180°E; left), Southern Extratropics (20°S~90°S, 180°W~180°E; middle
969 left), East Asian (15°N~65°N, 70°E~145°E; middle right), and Tropics (20°S~20°N,
970 180°W~180°E; right). The black and red lines show the 4DVar and 4DEnVar results, respectively.

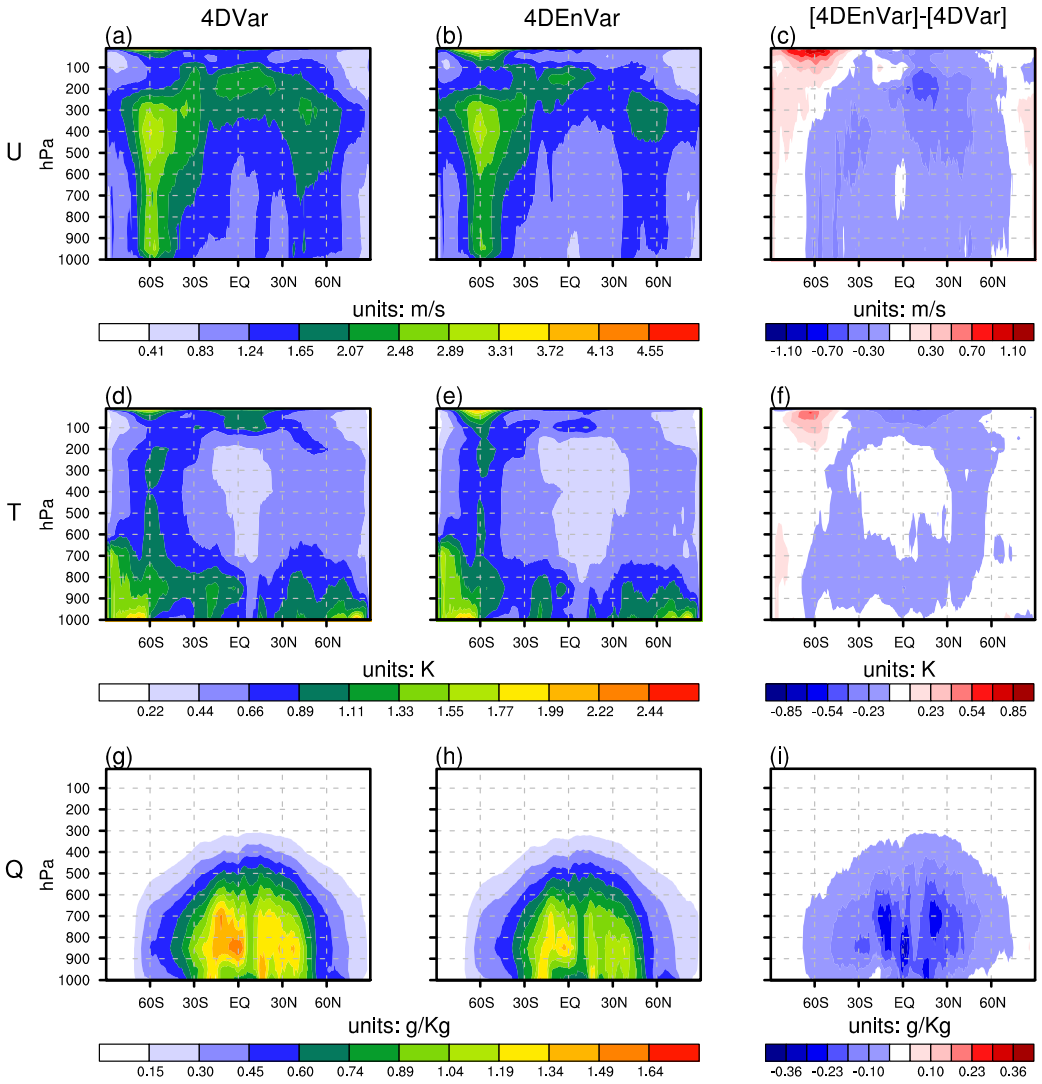


Figure 7. The pressure versus latitude plots of the ARMSEs (verified relative to the "truth" state) of the zonal wind (top; units: m/s), temperature (middle; units: K) and specific humidity (bottom; units: g/Kg) analyses of 4DVar (left), 4DEnVar (middle) and the ARMSE differences between 4DEnVar and 4DVar (right), respectively.

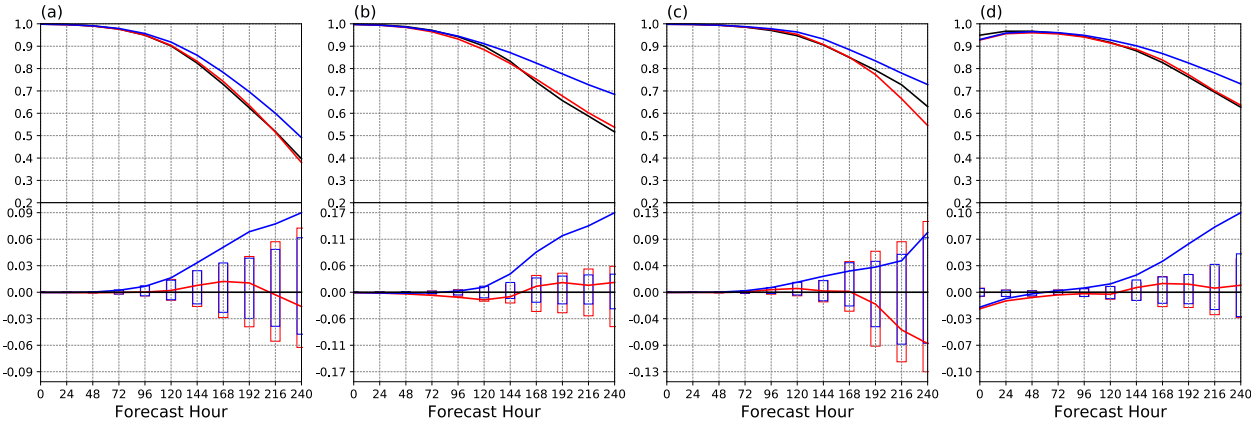


Figure 8. The anomaly correlation coefficients (ACCs) of the 4DVar-based (black line), 4DEnVar-based deterministic (red line) and the 4DEnVar-based ensemble mean (blue line) forecasts of the 500hPa geopotential height against the "truth" state in the (a) Northern Extratropics (20°N~90°N; 180°W~180°E), (b) Southern Extratropics (20°S~90°S; 180°W~180°E), (c) East Asia (15°N~65°N, 70°E~145°E), and (d) Tropics (20°S~20°N, 180°W~180°E). The corresponding ACC differences between the 4DEnVar-based deterministic forecast and the 4DVar-based forecast (red line), between the 4DEnVar-based ensemble mean forecast and the 4DVar-based forecast (blue line), and the 95% confidence thresholds are also plotted in the bottom.

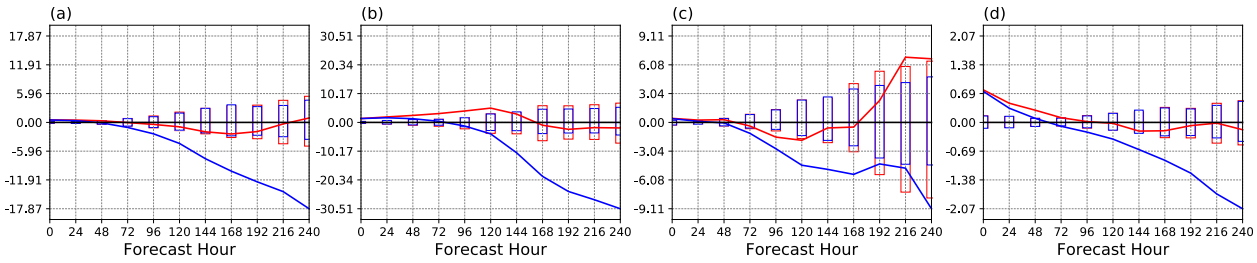
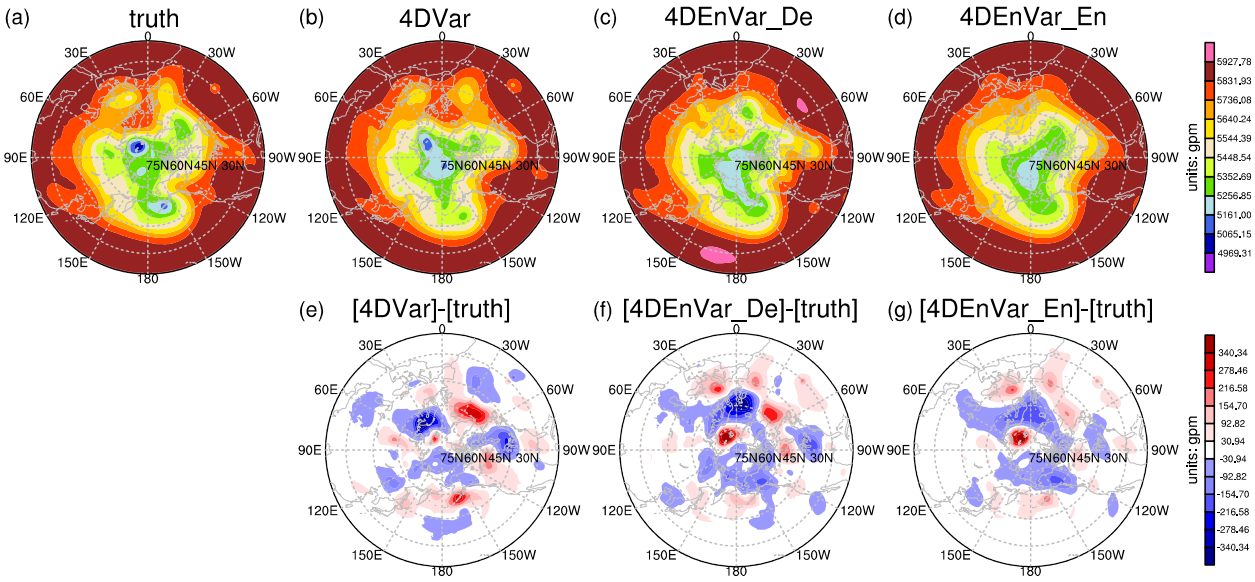


Figure 9. The ARMSE differences between the 4DEnVar-based deterministic forecast and the 4DVar-based forecast (red line) , between the 4DEnVar-based ensemble mean forecast and the 4DVar-based forecast (blue line) of the 500hPa geopotential height against the "truth" state in the (a) Northern Extratropics (20°N~90°N; 180°W~180°E), (b) Southern Extratropics (20°S~90°S; 180°W~180°E), (c) East Asia (15°N~65°N, 70°E~145°E), and (d) Tropics (20°S~20°N, 180°W~180°E). The bar charts show the 95% confidence thresholds.



992

993 **Figure 10**. The horizontal distributions of the 240-h forecast of the 500hPa geopotential height in
994 the Northern Extratropics (20°N~90°N; 180°W~180°E) for (a) the "truth" state, (b) the 4DVar-
995 based forecast, (c) the 4DEnVar-based deterministic forecast and (d) the 4DEnVar-based ensemble
996 mean forecast. The differences (e) between the 4DVar-based forecast and the "truth" state, (f)
997 between the 4DEnVar-based deterministic forecast and the "truth" state and (g) between the
998 4DEnVar-based ensemble mean forecast and the "truth" state are also plotted, respectively.

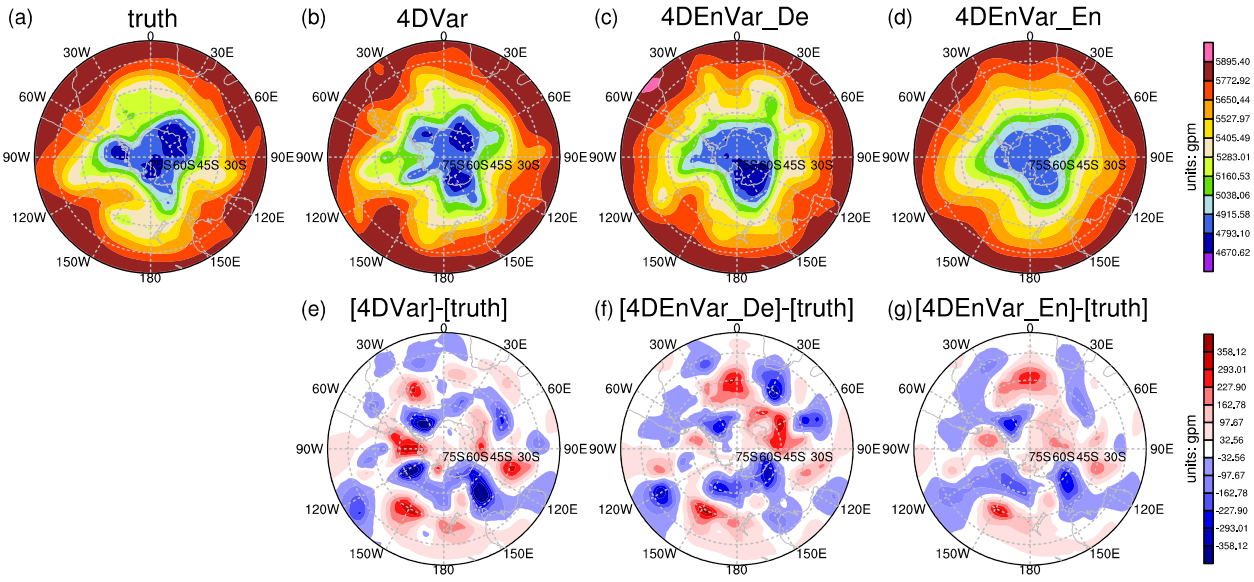


Figure 11. As in Figure 10, but showing the results in the Southern Extratropics (20°S~90°S; 180°W~180°E).

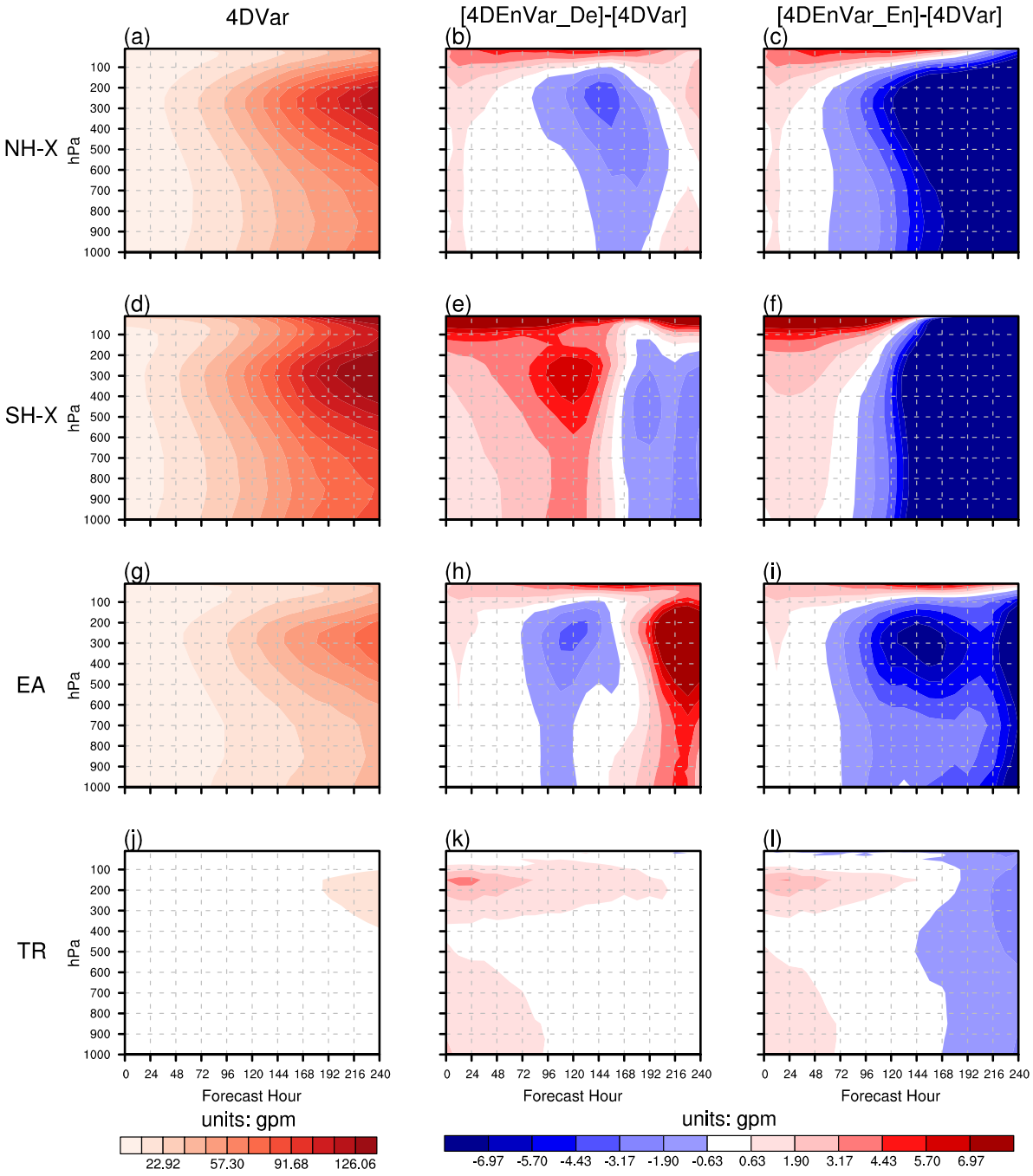


Figure 12. The ARMSEs of the geopotential height forecasts (units: gpm) initiated from the 1200 UTC analyses of the 4DVar experiment as a function of lead time (left) in the (a) Northern Extratropics (20°N~90°N; 180°W~180°E), (d) Southern Extratropics (20°S~90°S; 180°W~180°E), (g) East Asia (15°N~65°N, 70°E~145°E), and (j) Tropics (20°S~20°N, 180°W~180°E). The differences of ARMSE between the 4DEnVar-based deterministic forecast and the 4DVar-based forecast, and between the 4DEnVar-based ensemble mean forecast and the 4DVar-based forecast are plotted in (middle) and (right), respectively.

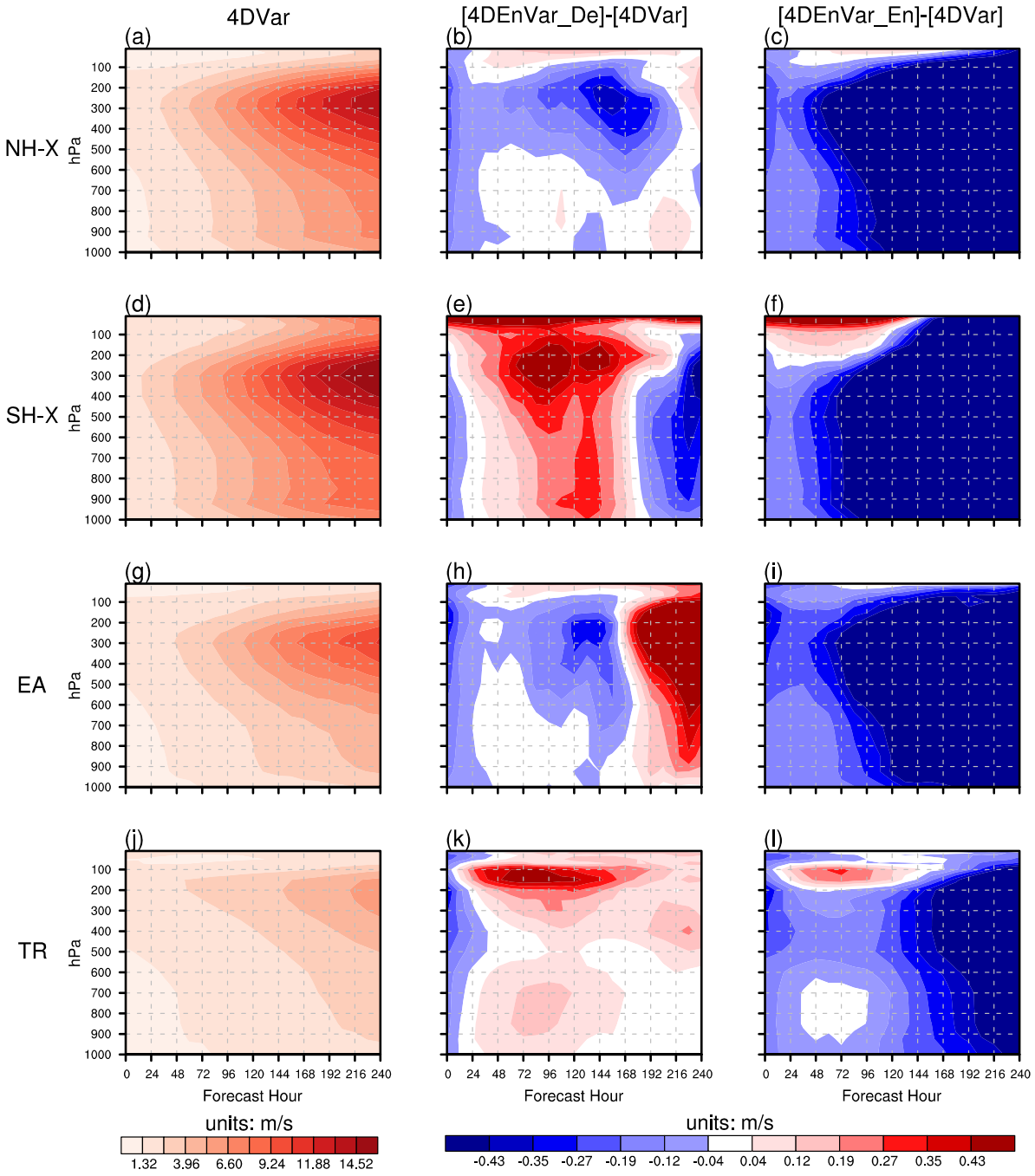


Figure 13. As in Figure 12, but showing the results of the zonal wind forecasts.

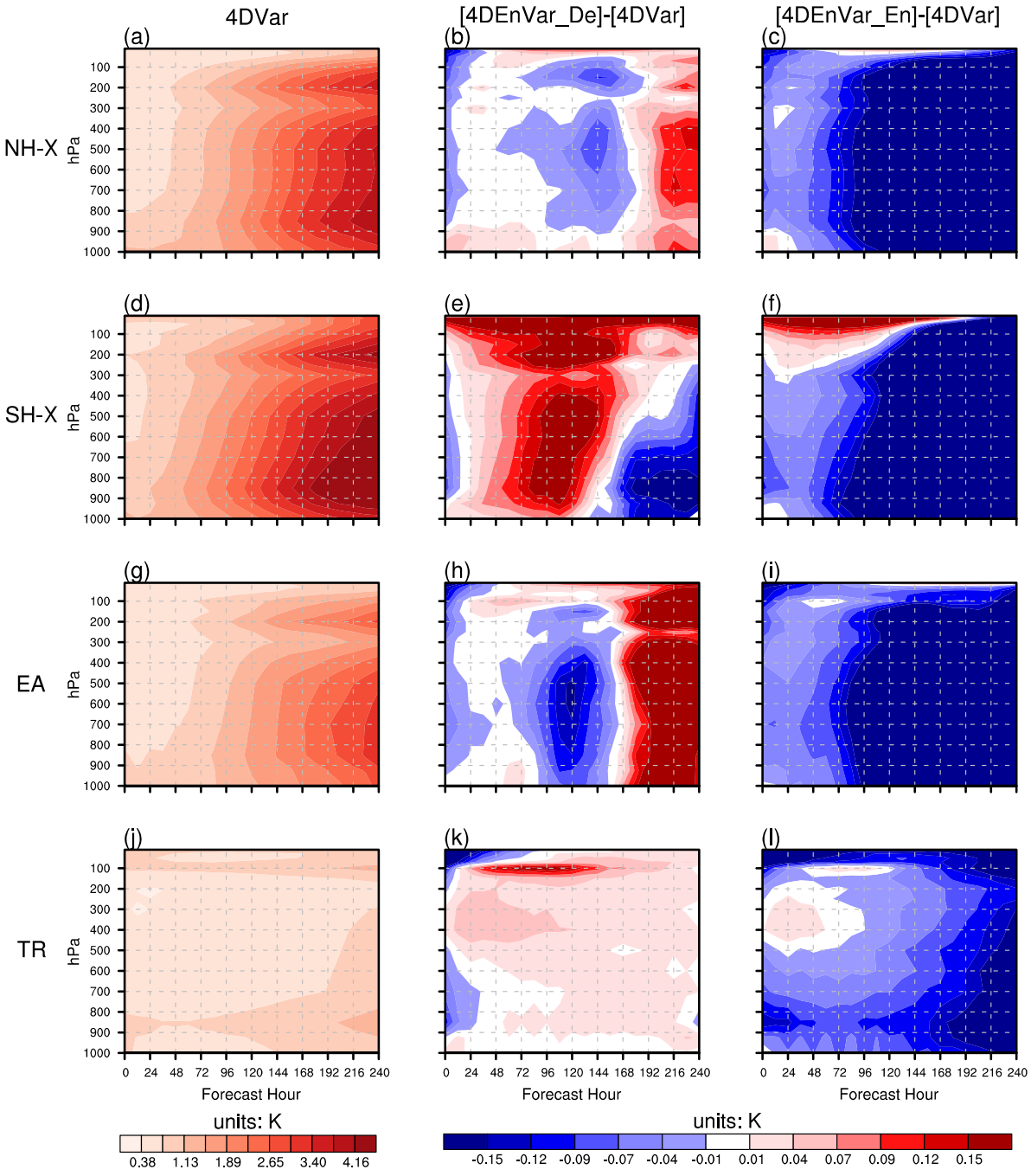


Figure 14. As in Figure 12, but showing the results of the temperature forecasts.

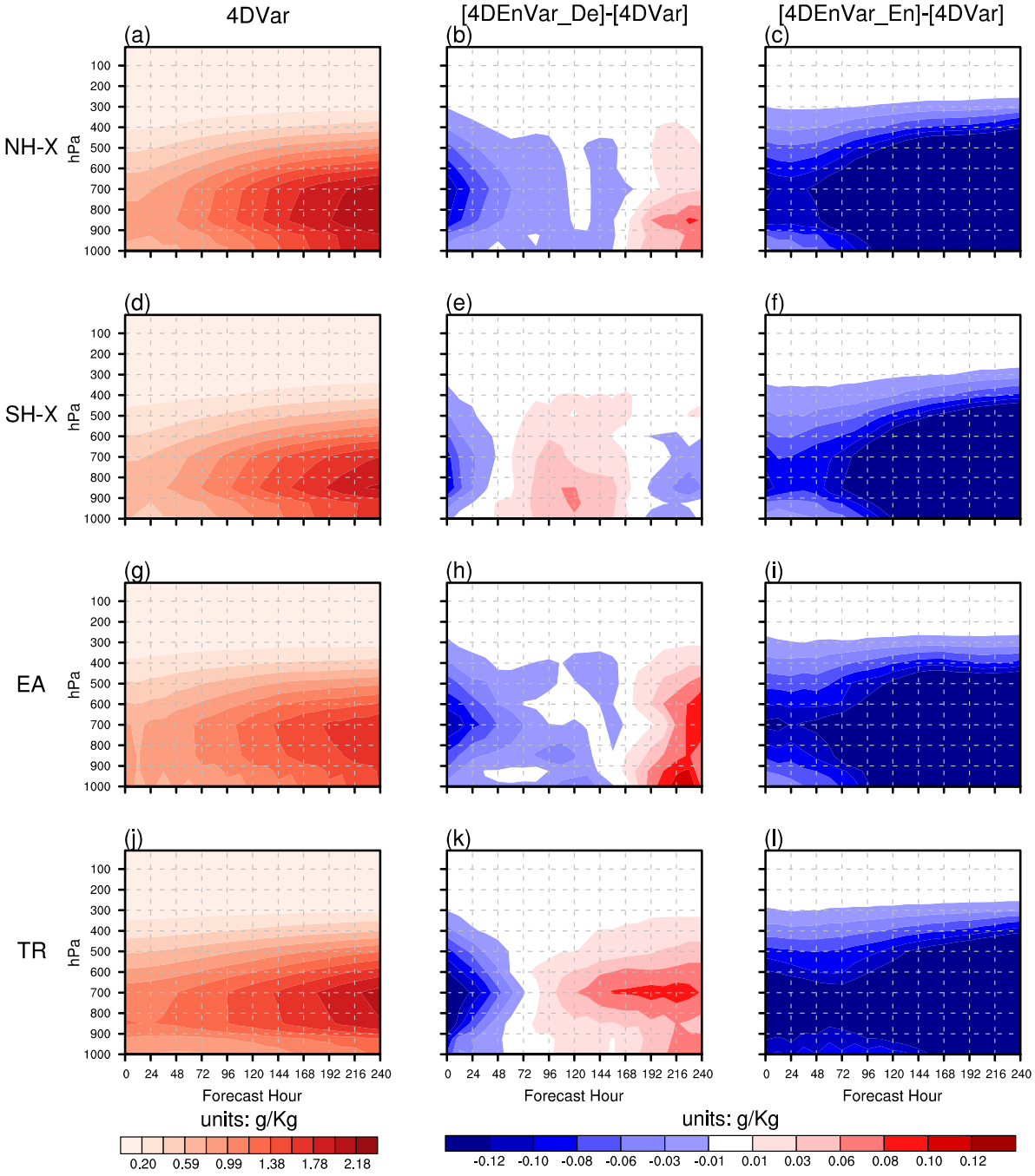


Figure 15. As in Figure 12, but showing the results of the specific humidity forecasts.