

The operational global four-dimensional variational data assimilation system at the China Meteorological Administration

Lin ZHANG,* Yongzhu LIU, Yan LIU, Jiandong GONG, Huijuan LU, Zhiyan JIN,
Weihong TIAN, Guiqing LIU, Bin ZHOU, and Bin ZHAO

Numerical Weather Prediction Center of China Meteorological Administration

*Correspondence to: Lin Zhang, China Meteorological Administration,
Zhongguancun South Avenue, Beijing, China.

E-mail: lzhang@cma.gov.cn

Running head: Global 4D-Var Data Assimilation at the CMA

Abstract

Since 1 July 2018, the GRAPES (Global/Regional Assimilation and PrEdiction System) global 4-dimensional variational (4D-Var) data assimilation system, has been in operation at the China Meteorological Administration (CMA). In this study, the GRAPES global 4D-Var data assimilation system is comprehensively introduced. This system applies the non-hydrostatic global tangent-linear model (TLM) and the adjoint model (ADM) for the first time. The use of a digital filter as a weak constraint is achieved. A series of linear physical processes is developed, including vertical diffusion, sub-grid scale orographic parameterization, large-scale condensation, and cumulus convection parameterization. The vertical diffusion and sub-grid scale orographic schemes are used in the operational suite and the linear convection parameterization and large-scale condensation scheme remain under assessment. The Lanczos and conjugate gradient (Lanczos-CG) algorithm and the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm are also developed. In terms of computational optimization, the total computational time of the GRAPES global TLM and ADM is approximately 3-fold that of the GRAPES global non-linear model (NLM).

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/qj.3533

Before it became operational, a one-year retrospective trial was performed on the GRAPES global 4D-Var data assimilation system. The entire system was stable, and the analysis and forecasting performances were significantly better than those of the 3D-Var data assimilation system, especially in the Southern Hemisphere.

Key Words: Global 4D-Var; GRAPES; CMA

1. Introduction

Four-dimensional variational assimilation (4D-Var) is an extension of three-dimensional variational assimilation (3D-Var) in the time dimension. 3D-Var ignores the temporal distribution of observations and assumes that all the observations within a time window are valid in the analysis time, whereas 4D-Var considers the observation time more accurately. 4D-Var can implicitly propagate the initial background error covariance during the assimilation window (Lorenc and Rawlins, 2005).

Prior to 2006, the global 4D-Var data assimilation system was in operation in the leading global numerical weather forecasting centres (Rabier et al., 2000; Janisková et al., 1999; Rawlins et al., 2007; Gauthier et al., 2007; Kadowaki, 2005). In recent years, a hybrid assimilation technique combining the ensemble method and the variational method has been rapidly developed. The European Centre for Medium-Range Weather Forecasts (ECMWF) implemented the ensemble of global 4D-Var data assimilations (EDA) to obtain the background error covariance, which changes with weather conditions. The original modelled background error covariance of global 4D-Var was improved, and a better analysis performance was achieved (Isaksen et al., 2010; Bonavita et al., 2012). By applying an extended control variable technique, the Met Office implemented the ensemble-based background error covariance into the global 4D-Var assimilation and developed the Hybrid-4DVar data assimilation system (Clayton et al., 2013). By linearly combining an ensemble of forecast samples, the Canadian Meteorological Centre (CMC) first obtained the temporal evolution of the analysis increment and then replaced the linear model in the 4D-Var with the temporal

evolution of the analysis increment and developed the 4D ensemble-variational (4DEnVar) data assimilation system (Buehner et al., 2015). The National Centers for Environmental Prediction (NCEP) also uses a 4DEnVar assimilation system (Kleist and Ide 2015). Overall, 4D-Var is the mainstream global data assimilation method, and the tendency in the future will also be to develop the hybrid ensemble-variational assimilation technique based on the current variational assimilation systems.

The Numerical Weather Prediction Center of the China Meteorological Administration (CMA) began developing a GRAPES global 4D-Var data assimilation system in 2008. On 1 June 2016, the GRAPES global 4D-Var system began operation using a 3D-Var analysis scheme (Wang et al. 2017) . At the end of 2017, the GRAPES global 4D-Var system began two sets of 4D-Var retrospective trials of 6-month duration. The operational stability and overall performance were confirmed. In April of 2018, the GRAPES global 4D-Var system was successfully implemented in the high-performance computer PI-SUGON, which had been recently purchased by the CMA. On 8 May 2018, a 4D-Var parallel test was started in a real-time environment. On 1 July 2018, the system began formal operation.

The incremental 4D-Var with a single outer loop is implemented in the GRAPES operational system. This system is designed to allow the multiple outer-loop technique to be applied. Since the computational efficiency of the nonlinear model integration and the observation data input is relatively low, the outer-loop iteration is not worthwhile for our system currently.

In this work, we systematically introduce, for the first time, the GRAPES global 4D-Var data assimilation system and focus on the work that updates the 3D-Var to 4D-Var due to length limitations. See Xue et al. (2008) for more details of GRAPES global 3D-Var system. In Section 2, the framework of the GRAPES global 4D-Var data assimilation system is introduced. The basic information on the GRAPES global tangent-linear model (TLM) and adjoint model (ADM) is given in Section 3. The assessment of the computational efficiency of the GRAPES global 4D-Var is provided in Section 4, followed by the retrospective trial results of the 4D-Var. The conclusion and discussion are given in the last section.

2. 4D-Var framework

2.1 Incremental 4D-Var

The GRAPES global 4D-Var data assimilation system applies the incremental analysis scheme proposed by Courtier et al. (1994). In this scheme, the definition of the cost function is :

$$J(\delta\mathbf{x}) = \frac{1}{2} \left(\delta\mathbf{x} - (\mathbf{x}^b - \mathbf{x}^g) \right)^T \mathbf{B}^{-1} \left(\delta\mathbf{x} - (\mathbf{x}^b - \mathbf{x}^g) \right) + \frac{1}{2} \sum_{i=0}^n (\mathbf{H}_i \mathbf{M}_{0 \rightarrow i} \delta\mathbf{x} + \mathbf{d}_i)^T \mathbf{R}_i^{-1} (\mathbf{H}_i \mathbf{M}_{0 \rightarrow i} \delta\mathbf{x} + \mathbf{d}_i) + J_c \quad (1)$$

Where \mathbf{x}^b is the background state; \mathbf{x}^g is the first guess; $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}^g$ is the analysis increment of the model variables; $\mathbf{d}_i = \mathbf{H}_i \mathbf{M}_{0 \rightarrow i} (\mathbf{x}^g) - \mathbf{y}_i$ is the observation increment at time i ; \mathbf{H}_i is the observation operator at time i ; $\mathbf{M}_{0 \rightarrow i}$ is the model integration from the analysis time to time i ; \mathbf{H}_i is the linear operator corresponding to \mathbf{H}_i ; $\mathbf{M}_{0 \rightarrow i}$ is the linear operator corresponding to $\mathbf{M}_{0 \rightarrow i}$; \mathbf{y}_i is the observation at time i ; \mathbf{B} is the error covariance matrix of \mathbf{x}^b ; \mathbf{R}_i is the observational error covariance matrix at time i ; and J_c is the weak constraint term based on the digital filter. To reduce the computational cost of the minimization, $\delta\mathbf{x}$ and \mathbf{B} are defined on the lower resolution grid than \mathbf{x}^b and \mathbf{x}^g . The interpolation operator is introduced to act on \mathbf{x}^b and \mathbf{x}^g .

The digital filter as a weak constraint is applied only to the increment at the time in the middle of the assimilation window. Defining $\delta\mathbf{x}_i = \mathbf{M}_{0 \rightarrow i} \delta\mathbf{x} - \mathbf{M}_{0 \rightarrow i} (\mathbf{x}^b - \mathbf{x}^g)$ represents the increment at time i , $\delta\mathbf{x}_{\frac{n}{2}}$ is the increment at the time in the middle of the assimilation window. The weak constrain term is :

$$J_c = \frac{1}{2} \left(\delta\mathbf{x}_{\frac{n}{2}} - \sum_{k=-\frac{n}{2}}^{\frac{n}{2}} \alpha_k \delta\mathbf{x}_{k+\frac{n}{2}} \right)^T \mathbf{Q} \left(\delta\mathbf{x}_{\frac{n}{2}} - \sum_{k=-\frac{n}{2}}^{\frac{n}{2}} \alpha_k \delta\mathbf{x}_{k+\frac{n}{2}} \right) \quad (2)$$

Where α_k is the coefficient of the digital filter, and its value can be found in Lynch and Huang (1992). \mathbf{Q} is a diagonal matrix composed of the inverse background error variance as Wee and Kuo (2004). It should be pointed out that the definition of \mathbf{Q} is empirical currently. We plan to use the energy norm to improve this weighting in the J_c term. And it is perhaps better to give $\delta\mathbf{x}_i = \mathbf{M}_{0 \rightarrow i} \delta\mathbf{x} - \mathbf{M}_{0 \rightarrow i} (\mathbf{x}^b) + \mathbf{M}_{0 \rightarrow i} (\mathbf{x}^g)$ which ensures that the non-linear model trajectory was balanced. At this time, the

Hessian matrix of the cost function (hereafter, the J_c term is ignored) is :

$$J''_{\delta \mathbf{x}} = \mathbf{B}^{-1} + \sum_{i=0}^n \mathbf{M}_{0 \rightarrow i}^T \mathbf{H}_i^T \mathbf{R}_i^{-1} \mathbf{H}_i \mathbf{M}_{0 \rightarrow i} \quad (3)$$

The rate of convergence of the 4D-Var minimization depends on the condition number (the ratio of the largest and smallest eigenvalues) of $J''_{\delta \mathbf{x}}$ which is very large. Therefore, operational variational assimilation systems generally use the so-called square root matrix of the background error covariance matrix \mathbf{U} ($\mathbf{U}\mathbf{U}^T = \mathbf{B}$) to perform the preconditioning transformation of the control variables (Courtier, et al., 1998; Gauthier, et al., 1999; Honda, et al., 2005; Lorenc, et al., 2000; Parrish and Derber, 1992). Using \mathbf{w} to represent the control variables after the preconditioning transformation, we have:

$$\delta \mathbf{x} = \mathbf{U} \mathbf{w} \quad (4)$$

After the preconditioning transformation of the control variables, the cost function is :

$$J(\mathbf{w}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{1}{2} \sum_{i=0}^n (\mathbf{H}_i \mathbf{M}_{0 \rightarrow i} \mathbf{U} \mathbf{w} + \mathbf{d}_i)^T \mathbf{R}_i^{-1} (\mathbf{H}_i \mathbf{M}_{0 \rightarrow i} \mathbf{U} \mathbf{w} + \mathbf{d}_i) \quad (5)$$

Furthermore, if approximate information of $J''_{\mathbf{w}}$ is available, the preconditioning transformation of the control variables can be further modified. For instance, if $\mathbf{E} \mathbf{E}^T \cong (J''_{\mathbf{w}})^{-1}$, then by using :

$$\delta \mathbf{x} = \mathbf{U} \mathbf{E} \mathbf{w}' \quad (6)$$

we obtain :

$$J''_{\mathbf{w}'} = \mathbf{E}^T J''_{\mathbf{w}} \mathbf{E} \quad (7)$$

The condition number of $J''_{\mathbf{w}'}$ is smaller than that of $J''_{\mathbf{w}}$, and the minimizing convergence is faster. In the GRAPES global 4D-Var assimilation system, \mathbf{E} is the preconditioning operator of the Lanczos and conjugate gradient (Lanczos-CG) algorithm (Fisher, 1995). In fact, in our scheme the preconditioning of the Lanczos-CG algorithm is conducted during the preconditioning transformation of the control variables. The minimization results of our scheme are consistent with those obtained by preconditioning the Lanczos-CG algorithm. However, our scheme is independent of the minimization algorithm, so that other minimization algorithms can also benefit from the preconditioning operator of the Lanczos-CG algorithm. The

preconditioning operator of Lanczos-CG algorithm can be used in the subsequent minimization after the first minimization if the multiple outer-loops are performed. Since the multiple outer-loops are not implemented in our system currently as mentioned above, we calculate the approximated eigenpairs of Hessian matrix from one DA cycle and conduct the preconditioning operator in the 4D-Var minimization at the subsequent times. It should be point out that this preconditioned algorithm is not used in the operational system. We are performing more experiments to design the operational scheme.

2.2 Control variable transformation

The state variables of the GRAPES global non-linear model (NLM) include the non-dimensional pressure π , the potential temperature θ , the east-west component u and the north-south component v of the horizontal wind, the vertical component \hat{w} of wind, and the specific humidity q , etc. (Chen et al., 2008).

Because the error covariance of different variables coexists with the error covariance of the same variable but at different spatial positions, the background error covariance matrix \mathbf{B} in terms of model variables is extremely complicated. Thus, the stream function ψ , unbalanced velocity potential χ^u , unbalanced non-dimensional pressure π^u and specific humidity q are used as the GRAPES global 4D-Var assimilation system variables which are assumed to be independent of each other. The corresponding background error covariance matrix for the uncorrelated variables is a block diagonal matrix. The GRAPES global 4D-Var assimilation system does not analyse the vertical component of wind. The relationship between the analysis increments of model variables and the uncorrelated variables is :

$$\begin{bmatrix} \delta u \\ \delta v \\ \delta \pi \\ \delta q \end{bmatrix} = \mathbf{P} \begin{bmatrix} \delta \psi \\ \delta \chi^u \\ \delta \pi^u \\ \delta q \end{bmatrix} \quad (8)$$

where the physical transformation operator \mathbf{P} is expressed as :

$$\mathbf{P} = \begin{bmatrix} -\frac{\partial}{\partial y} & \frac{\partial}{\partial x} & 0 & 0 \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{I} & 0 & 0 & 0 \\ \mathbf{M} & \mathbf{I} & 0 & 0 \\ \mathbf{N} & \mathbf{\Lambda} & \mathbf{I} & 0 \\ 0 & 0 & 0 & \mathbf{I} \end{bmatrix}$$

(9)

where \mathbf{M} represents the transformation from stream function to velocity potential, \mathbf{N} represents the transformation from stream function to non-dimension pressure, and $\mathbf{\Lambda}$ represents the transformation from unbalanced velocity potential to non-dimensional pressure. The balance operator \mathbf{N} involves solving the linear balance equation at each model layer and a vertical linear regression (Lorenc et al, 2000). And the \mathbf{M} and $\mathbf{\Lambda}$ matrices are calculated by linear regression (Wu et al, 2002).

In addition, the analysis increment of potential temperature θ is diagnosed from the analysis increment of dimensionless pressure π using the hydrostatic relationship.

Given the background error covariance matrix \mathbf{B}^u for the uncorrelated variable, the background error covariance matrix \mathbf{B} in terms of the model variable is:

$$\mathbf{B} = \mathbf{P} \mathbf{B}^u \mathbf{P}^T \quad (10)$$

\mathbf{U}^u is then constructed and used to perform the preconditioning transformation of the control variables. Here $\mathbf{U}^u \mathbf{U}^{uT} = \mathbf{B}^u$. After the transformation, the background error covariance matrix for the control variable can be simplified to a unit matrix.

$$\mathbf{U}^u = \boldsymbol{\Sigma}^u \mathbf{U}_h \mathbf{U}_v \quad (11)$$

$\boldsymbol{\Sigma}^u$ is a diagonal matrix composed of the corresponding background root mean square error of the uncorrelated variable; and \mathbf{U}_h and \mathbf{U}_v are the horizontal and vertical correlation transformation matrices, respectively.

The background error variance of the uncorrelated variable used in the GRAPES global 4D-Var assimilation system is obtained using the National Meteorological Centre (NMC) method and 1-year of GRAPES global model forecasting samples. We then zonally averaged the obtained 3-dimensional background error variance, obtaining the background error variance, which varies only along latitude and altitude.

The horizontal correlation model is obtained by applying a second order autoregressive correlation function. A spectral filter is then applied to calculate \mathbf{U}_h . The horizontal correlation scale is referred to the parameters provided by Ingleby (2001). The vertical correlation model is obtained from an ensemble of samples using statistical methods. Then, an empirical orthogonal function (EOF) is applied to obtain \mathbf{U}_v .

2.3 Minimization algorithm

In the GRAPES global variational assimilation system, we always used the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm (Liu and Nocede, 1989) to perform the minimization. The L-BFGS algorithm needs only to store the values of m pairs of cost function and gradients, thus its memory usage is very small. Moreover, m is an adjustable parameter. According to our experience, the better choice for m is 12.

In the last two years, we learned from the successful experience of the ECMWF (Fisher, 1998), and developed a Lanczos-CG algorithm in the GRAPES global 4D-Var data assimilation system. Compared to the L-BFGS algorithm, the Lanczos-CG algorithm has a higher convergence speed and a more stable convergence process. In a 1-month global 4D-Var experiment with the same experimental settings, on average the L-BGFS algorithm used 77 iteration steps, whereas the Lanczos-CG algorithm used 45 iteration steps (Liu and Zhang, 2018). Therefore, the Lanczos-CG algorithm is applied for operation of the GRAPES global 4D-Var data assimilation system.

In the GRAPES global 4D-Var data assimilation system, the criterion for minimizing convergence is :

$$\frac{\|\nabla_{w_k} J\|}{\|\nabla_{w_0} J\|} \leq \text{eps0} \quad (12)$$

where $\|\nabla_{w_0} J\|$ is the norm of the initial cost function gradient, $\|\nabla_{w_k} J\|$ is the norm of the cost function gradient after the k^{th} minimization iteration, and the default value of eps0 is 0.03.

3. GRAPES global TLM and ADM

The GRAPES global model uses a set of non-hydrostatic and fully compressible equations on the sphere, and the semi-Lagrangian semi-implicit temporal integration scheme (Chen et al., 2008). Based on the GRAPES global NLM code, we directly developed the non-hydrostatic dynamical frame code of GRAPES global TLM and ADM. We then developed a series of linear physical processes for the application of 4D-Var. To be consistent with the update of the dynamical frame of the GRAPES

global NLM, we update the dynamical frame codes of the GRAPES global TLM and ADM every year.

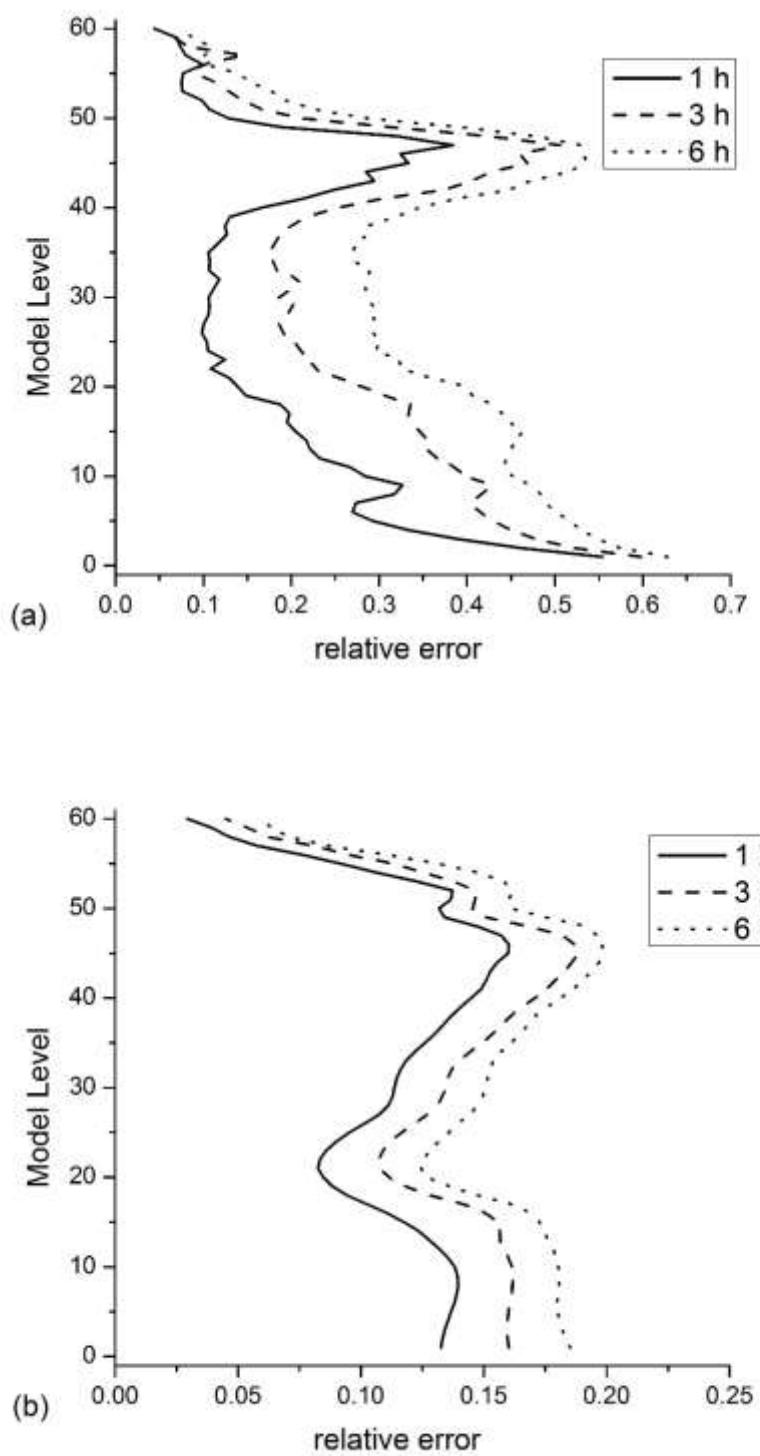
In the development of the global TLM and ADM, the standard correctness check is performed repeatedly using the equation (13) for arbitrary analysis increment $\delta\mathbf{x}$.

$$\langle \mathbf{M}\delta\mathbf{x}, \mathbf{M}\delta\mathbf{x} \rangle = \langle \delta\mathbf{x}, \mathbf{M}^T \mathbf{M}\delta\mathbf{x} \rangle \quad (13)$$

We have also evaluated the performance of the global TLM using the analysis increment as the initial field of the perturbed model variables and following the method as Radnóti et al. (2005) from ECMWF. The relative error in the tangent linear model is calculated as

$$r = \frac{\|\mathbf{M}(\mathbf{x}+\delta\mathbf{x}) - \mathbf{M}(\mathbf{x}) - \mathbf{M}\delta\mathbf{x}\|}{\|\mathbf{M}(\mathbf{x}+\delta\mathbf{x}) - \mathbf{M}(\mathbf{x})\|} \quad (14)$$

In these tests, the nonlinear and linear models are performed at the same resolution. So there is no interpolation operator act on $\delta\mathbf{x}$. From our results, the linear approximation for the non-dimensional pressure is the best and for the specific humidity is the worst (Figure 1). The error near the surface is similar to ECMWF's. And the error in the high level is also significant in our system. We applied the TAPENADE tool (Laurent, 2007) to generate a few tangent-linear and adjoint programs with relatively simple logical structures. We then tuned the programs manually to improve them. We manually programmed all the programs with complicated logical structures.



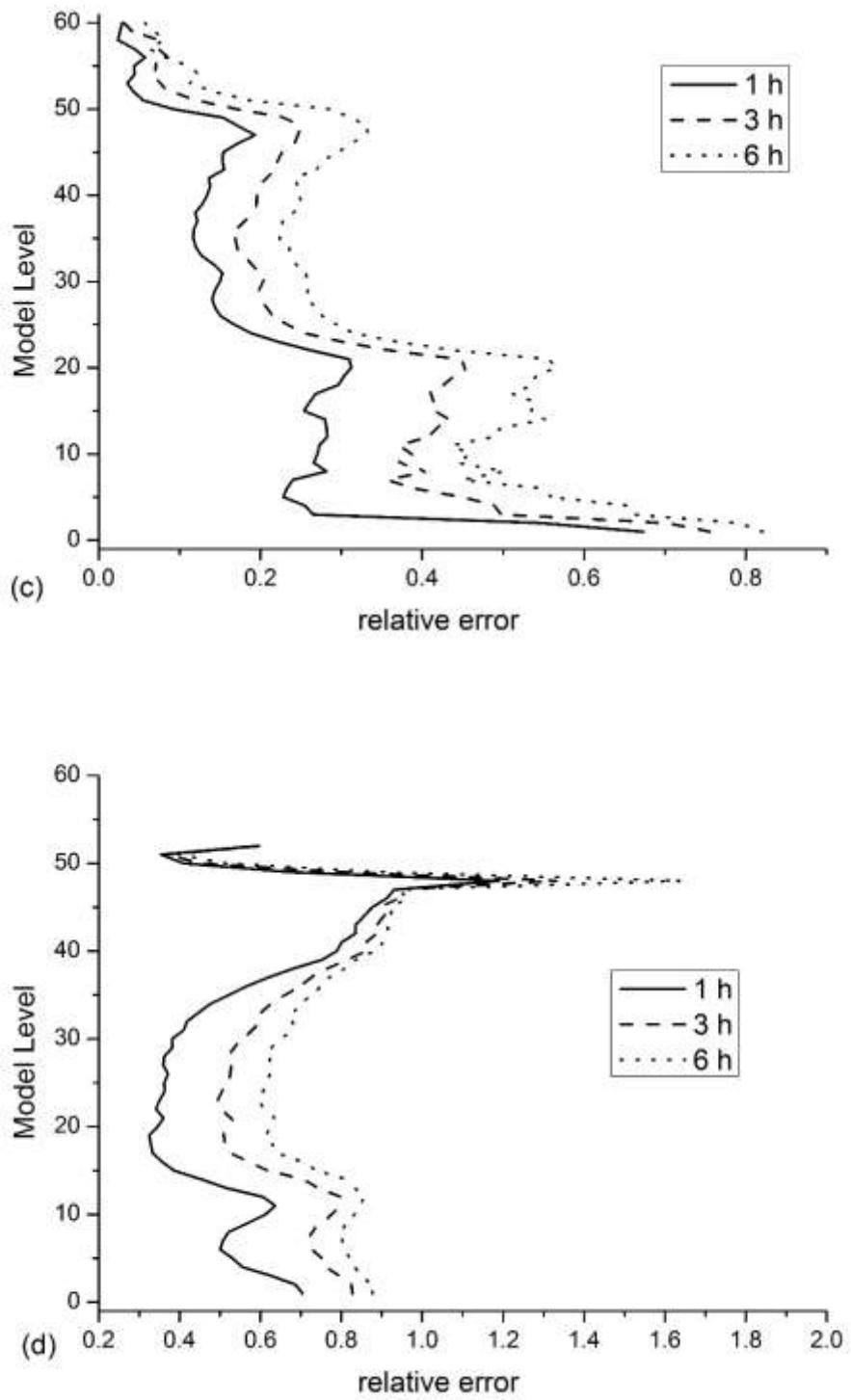


Figure 1: The relative error in the tangent linear model with simple physics with respect to the nonlinear forecast model with full physics at the resolution of 1.0° , the solid line gives the 1-h forecast results, the dashed line represents the 3-h forecast results, and the dotted line shows the 6-h forecast results, (a) u wind; (b) non-dimensional pressure; (c) potential temperature; (d) specific humidity.

The computation of the global TLM and ADM needs the model trajectories at different times. These model trajectories can be classified into two types: the first type is the model trajectory at the beginning of each integration step; the second type is the model trajectory, which is input by different computation modules within each integration step. To reduce the difference between the linear model and nonlinear model, the model trajectories of the global TLM and ADM at the beginning of each integration step are provided by the integration of the nonlinear forecast model. There is an interpolation step from the higher resolution to the lower resolution. It also includes the vertical interpolation in view of the different heights of the model surfaces at the different resolution. For the global TLM, the model trajectories input by different computation modules within each integration step are calculated by calling the codes of the forecast model. For the global ADM, these model trajectories are obtained in two ways. In the first way, the model trajectories are recalculated as in the TLM. This calculation is easy to conduct and the codes are straightforward, but it consumes more computational time. In the second way, the model trajectories are stored in the memory stacks during the integration of the global TLM, and then the model trajectories are fetched by the global ADM from the memory stacks. This approach can reduce the computational time but has a large memory requirement with complicated data flow, which complicates the maintenance of the codes. To reduce the computational time of the global ADM, the GRAPES global ADM primarily uses the memory transfer method to process the model trajectories, which is supplemented by the recalculation method.

Table 1 shows the computational times of the GRAPES global NLM, GRAPES global TLM, and GRAPES global ADM on the PI-SUGON high-performance computer (excluding the time of the initial value data I/O). Since the computational time of the models is influenced by the state of the high-performance computer, the results shown in Table 1 may be biased. However, it is still apparent that with an increasing number of computing cores, the accelerations of the global TLM and global ADM are better than that of the GRAPES global NLM. This is because when the number of computing cores is relatively small, the memory resources cannot fully meet the

demands of the global TLM and global ADM. Therefore, the computing efficiency is influenced. When using 1024 computing cores, the total time of the global TLM and global ADM is approximately 3-fold that of the GRAPES global NLM.

Currently, the linear physical processes in the GRAPES global TLM and ADM include vertical diffusion, sub-grid scale orographic parameterization, cumulus deep convection, and large-scale condensation. These linear physical processes are developed based on the medium-range forecast (MRF) non-regional boundary layer scheme by Hong (1996), the sub-grid scale topographic gravity wave parameterization scheme by Alpert (1988), the new simplified Arakawa-Shubert (NSAS) cumulus scheme by Han and Pan (2006), and the simplified large-scale cloud and precipitation scheme by Tompkins and Janiskova (2004), respectively.

Table 1. Computational time (unit: seconds) of the dynamical frame of the GRAPES global NLM, the GRAPES global TLM, and the GRAPES global ADM. The models have a horizontal resolution of 1° and 60 vertical layers. The forecasting period is 6 hours and the time for the integration step is 900 seconds.

	Number of processor			
	128	256	512	1024
NLM	13.6	8.5	4.7	5.2
TLM	25.5	16.7	9.1	6.7
ADM	35.0	21.0	11.2	7.9

We have confirmed that linear vertical diffusion and sub-grid scale orographic parameterization can improve the forecast of the global TLM and ADM. After the linear vertical diffusion and sub-grid scale orographic parameterization are switched on, the minimization process in the global 4D-Var is faster and more stable. Because the analysis and forecasting are improved, both schemes are used in the operation of the system. The linear convection parameterization and large-scale condensation scheme remain under assessment, and will be put into operation soon. The linear

physical processes will be described in another paper in more detail.

4. Computational efficiency

To improve the computational efficiency of the global 4D-Var data assimilation system, apart from the optimization of the global ADM described above, we applied an executable program to calculate the incremental 4D-Var analysis with one outer loop. This program allows data transfer through the memory, which can significantly reduce the read time of the inner loop. However, a problem in which the main program calls the global NLM, TLM, and ADM with different settings must be solved. The discrepancy of this scheme is that the higher resolution nonlinear model integration and the lower resolution linear model integration can compute only in parallel with the same number of computing cores, which is not flexible.

By applying the operation settings (Table 2), we used 256, 512, and 1024 cores to perform 50 minimization iterations of the global 4D-Var data assimilation on the PI-SUGON high-performance computer of the CMA. The computational times were about 50, 33, and 25 minutes (Table 3), respectively. Overall, increasing parallel computing can significantly improve the computational efficiency of the 4D-Var minimization in the inner loop. However, due to more data I/O, the parallel acceleration of the 4D-Var observation increment calculations is relatively low. When using 1024 cores, the computational time of the 4D-Var observation increment calculations can take 28% of the total time. Therefore, we will focus next on the optimization of the observation I/O.

Table 2. The operational settings for the GRAPES global 4D-Var data assimilation system

Horizontal resolution	0.25°/1.0° (outer loop/inner loop)
Vertical layers	60
Model integration step	300 s/900 s
Assimilation window	6 hours
Observation time slot	30 minutes

Maximum number of minimization iterations	50
---	----

Table 3. The computational time of the GRAPES global 4D-Var assimilation (unit: seconds) when using the operational setting scheme. The 4D-Var observation increment calculations include the 6-h nonlinear model forecast at high resolution. The 4D-Var minimization includes the tangent-linear model and adjoint model integrations during the inner loop iteration. The cost of these two parts is approximately equal to the total cost of 4D-Var analysis.

	Number of processor		
	256	512	1024
4D-Var observation increment calculations	640.7	437.8	424.1
4D-Var minimization	2354.9	1555.6	1096.1

5. Retrospective trials

We conducted two sets of 6-month global 4D-Var retrospective trials (June 2016–November 2016 and December 2016–May 2017, respectively). Four 6-hour global 4D-Var analyses were performed every day. The starting time of each assimilation window was 0300 UTC, 0900 UTC, 1500 UTC, and 2100 UTC. After the global 4D-Var analysis, the analysis field at each starting time of the assimilation window was generated. The 6-hour global model forecasting was then conducted, which provided the background field for the next global 4D-Var analysis. At the same time, the model forecasting results in the third hour after the starting time of each window were produced as the analysis fields at four standard times, namely, 0600 UTC, 1200 UTC, 1800 UTC, and 0000 UTC. They were used as the initial values for the 10-day global forecast.

The settings for the global 4D-Var retrospective trials are identical to those of the

operation. The maximum minimization iteration number in the inner loop is set to 50. The observations used in the retrospective trials are shown in Table 4. (In the operation, the FY3C MWHS2 observation is also used). For comparison, we also conducted a 3D-Var retrospective trial. There are two main differences between the 3D-Var and 4D-Var experiments. First, the 4D-Var applies a digital filter as a weak constraint to control the noise of gravity waves, whereas for the 3D-Var, the digital filter is applied independently after the 3D-Var analysis. Second, the horizontal resolution of the global 3D-Var is identical to that of the original operational system, with the inner-loop resolution of 0.25° . However, according to our previous experiments, increasing only the global 3D-Var inner-loop resolution has a slight influence on the analysis quality.

Table 4. The observational data used in the GRAPES global 4D-Var retrospective trials

Observation type	Observation element
TEMP	temperature, wind, and relative humidity
SYNOP	air pressure
SHIP	air pressure
AIREP	wind
SATOB	wind
SCATWIND	wind
GNSSRO	refractivity
NOAA15 AMSUA	radiance
NOAA18 AMSUA	radiance
NOAA19 AMSUA	radiance
METOP-A AMSUA	radiance
METOP-B AMSUA	radiance
ATMS AMSUA	radiance
AIRS	radiance

In the 4D-Var experiment of 1-year duration, the Lanczos-CG minimization of the global 4D-Var analysis was interrupted once at 9:00 on 9 March 2017. This interruption occurred because the leading minor of the tridiagonal matrix (see Eq. 20 of Fisher, 1998 for details) calculated from the Lanczos-CG iteration was not positive definite. There was no such interruption when using the L-BFGS algorithm. We believe that there was certain noise in the integration results of the GRAPES global ADM, which resulted in computational errors in the Lanczos iteration coefficient. When the errors exceeded an acceptable range, the Lanczos-CG algorithm was interrupted. In the application of the L-BFGS algorithm, if a proper convergence step length cannot be found, it remains possible to adjust the convergence direction. Therefore, the tolerance of the L-BFGS algorithm for computational errors may be higher.

Apart from the interruption mentioned above, all the global 4D-Var analysis-forecast cycles and the 10-day forecasts were normal. Therefore, the GRAPES global 4D-Var data assimilation system is sufficiently stable for operation.

On average, there are 43 iterations in the global 4D-Var assimilation; the convergence rate of the cost function is 29%; and the convergence rate of the gradient norm is 96%. For the global 3D-Var, the corresponding results are 41, 38%, and 97%, respectively. Overall, the convergence speed of the global 3D-Var assimilation is higher. As described above, the minimization convergence of variational assimilation is determined by the condition number of the Hessian matrix of the cost function. From the maximum eigenvalue of the Hessian matrix estimated using the Lanczos-CG algorithm, one can see that a considerable proportion of the 4D-Var results are exceptions. Under normal circumstances, the estimates of the maximum eigenvalue of the 4D-Var Hessian matrix range between 4000 and 5000, but approximately 10% of the estimates of the maximum eigenvalue exceeded 10^4 and approximately 4% exceeded 10^5 . In extreme cases, the estimates of the maximum eigenvalue exceeded 10^{10} . In these cases, the norm of the initial gradient also increased significantly, indicating that the integration results of the ADM had changed overall. In the global

3D-Var experiment, the estimates of the maximum eigenvalue of the Hessian matrix did not change dramatically. The estimates of the maximum eigenvalue exceeded 10^4 only 3 times with a highest value of 23451.12. This indicates that the problem with the 4D-Var assimilation is related to the model integration.

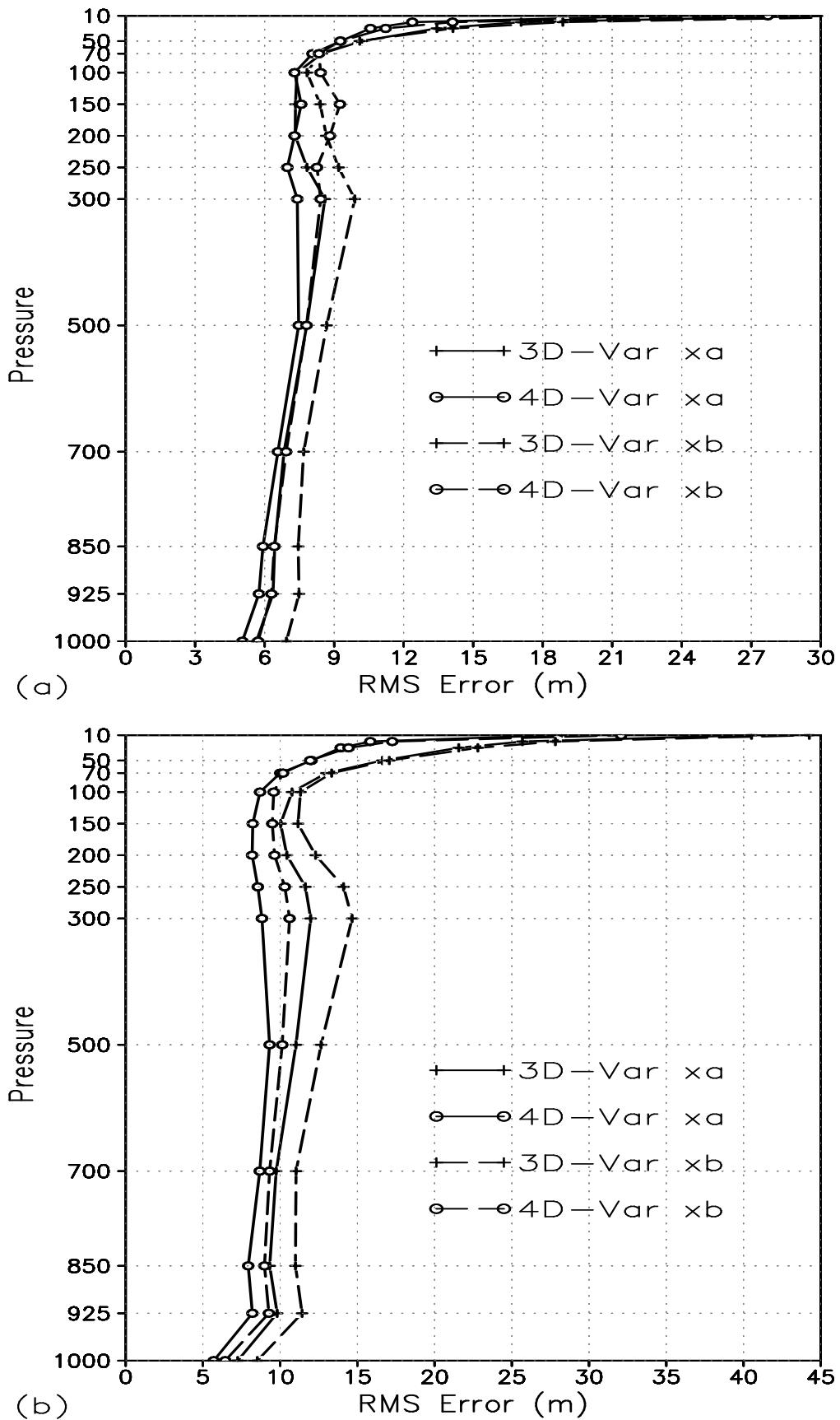
The 3D-Var assimilation uses 1.4 million pieces of data every day, whereas the 4D-Var assimilation uses 2.1 million pieces of data every day, which is 50% more than that of 3D-Var. The amount of observation data with higher temporal resolution is increased in the 4D-Var, especially for SHIP, SYNOP, SATOB, and AIREP, whose amounts increased by 344%, 309%, 153%, and 128%, respectively.

We divided the globe into 4 regions and statistically analysed the root mean square error (RMSE) and mean bias of the GRAPES global background field and analysis field. The 4 regions are the Northern Hemisphere (20°N - 90°N , 0°E - 180°W), the Southern Hemisphere (90°S - 20°S , 0°E - 180°W), the tropics (20°S - 20°N , 0°E - 180°W), and East Asia (15°N - 65°N , 70°E - 145°E).

Referenced to the ERA INTERIM reanalysis field, the RMSE of the analysis field of the GRAPES global geopotential height in the troposphere of the Northern Hemisphere ranges between 4 and 8 geopotential metres. The RMSE in the troposphere of the Southern Hemisphere ranges between 5 and 10 geopotential metres (Figure 2). Compared to the 3D-Var results, the RMSE of the 4D-Var is reduced by approximately 2-5 geopotential metres. However, the RMSE above 100 hPa increases significantly, especially for 10 hPa, where the RMSE exceeds 30 geopotential metres. The analysis quality in high layers is largely influenced by the model forecasting error and the bias correction of the satellite high-layer channel. In the next step, we will focus on increasing the top of the model, reducing the model forecasting error, and improving the effect of the satellite observation assimilation.

Figure 2 shows that the RMSE of the geopotential height background field (6-hour forecast) of the GRAPES global 4D-Var is smaller than that of the 3D-Var analysis, especially in the Southern Hemisphere. The 4D-Var analysis field, 4D-Var background field, 3D-Var analysis field, and 3D-Var background field have a RMSE in an ascending order. This is a very clear feature, indicating that the analysis quality

of the GRAPES 4D-Var is obviously better than that of the 3D-Var.



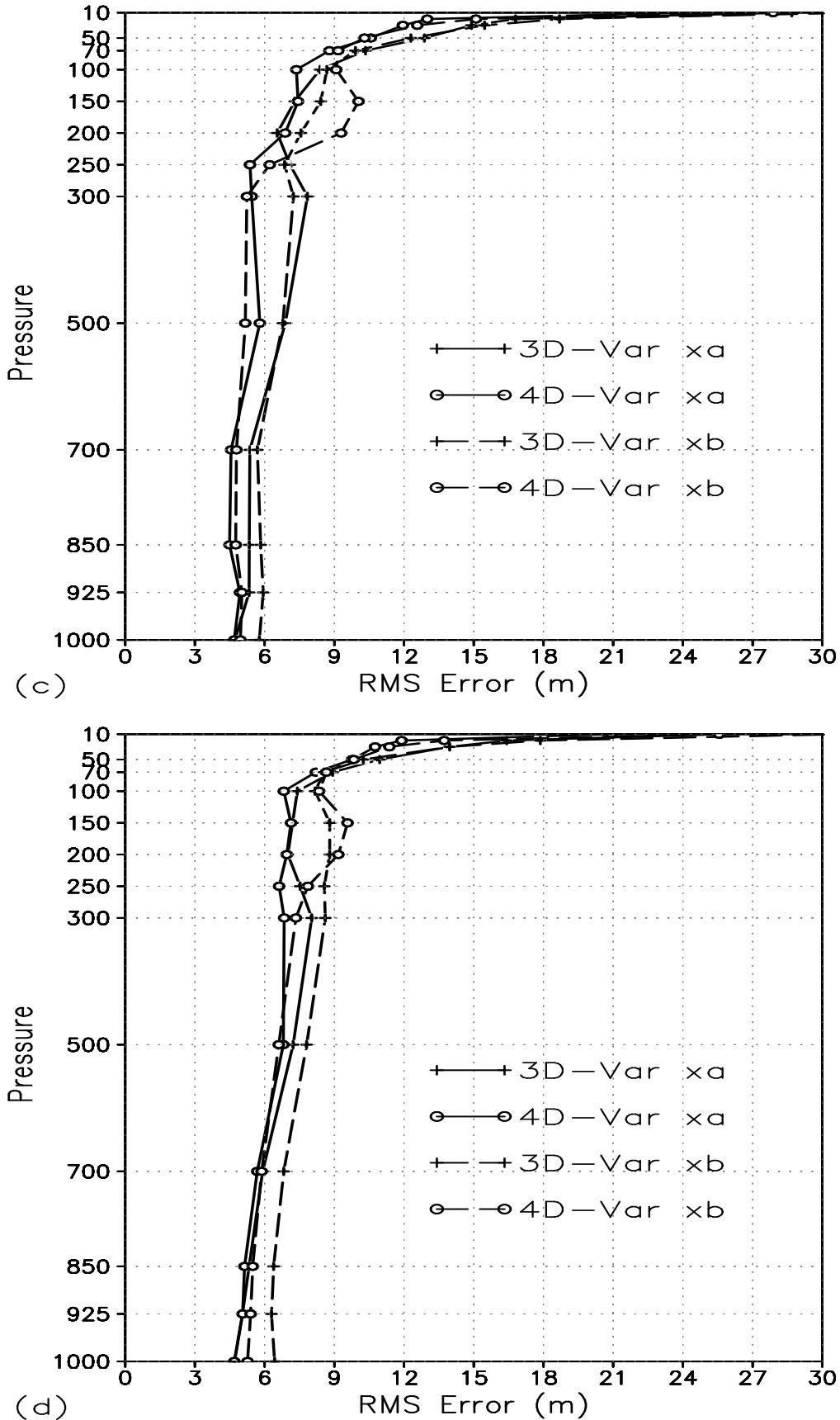
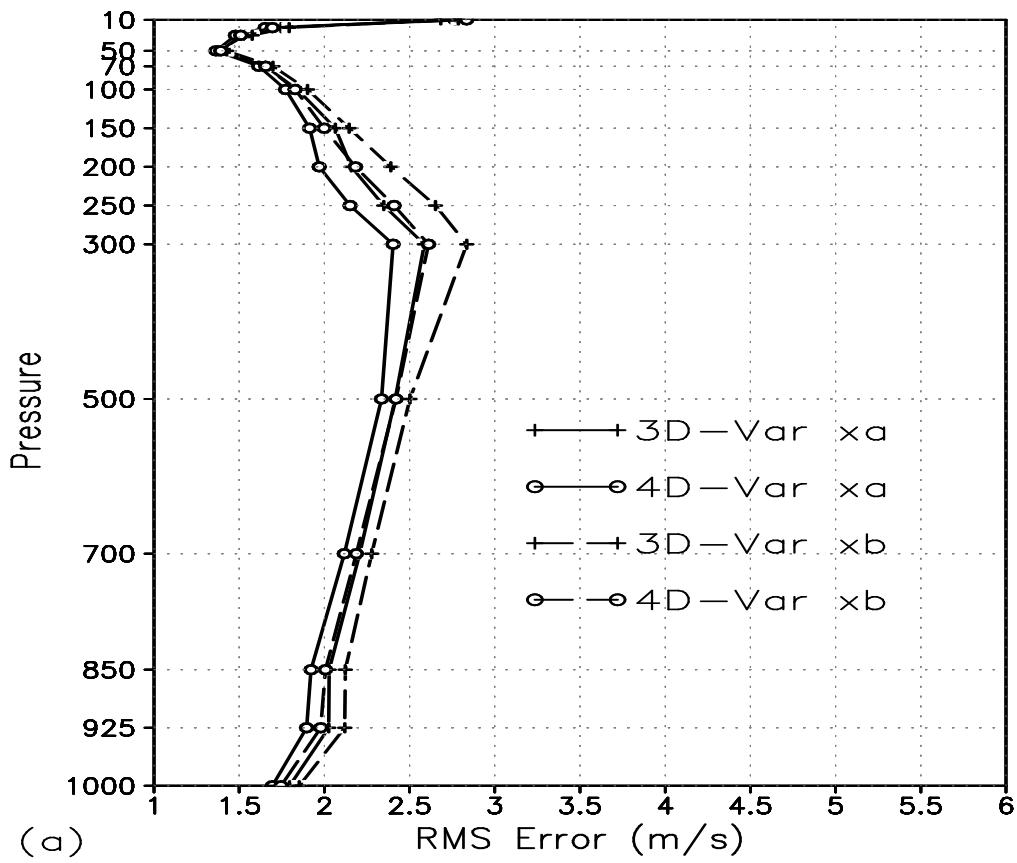
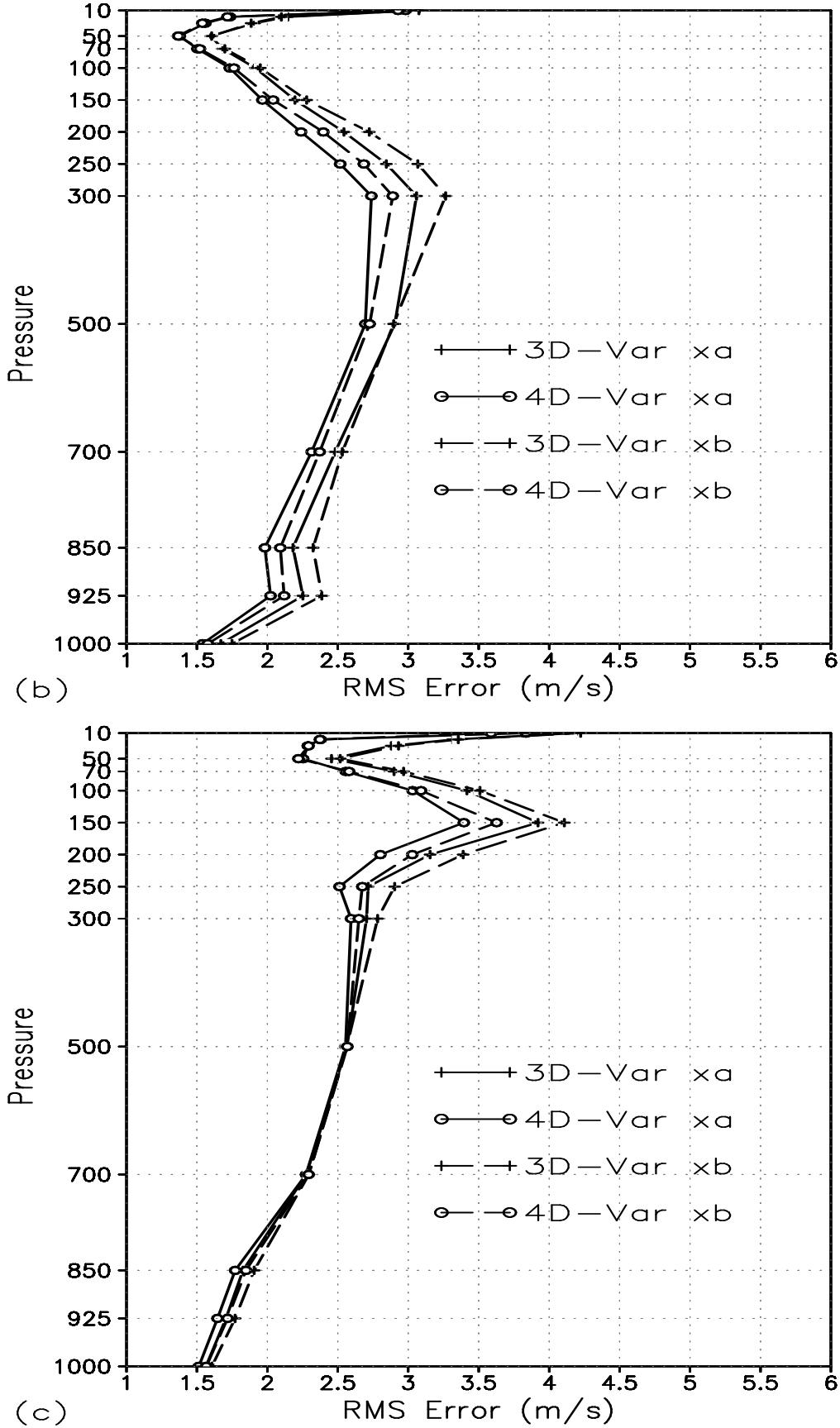


Figure 2. The RMSE (measured against of the ERA INTERIM reanalysis) of the background and analysis fields of the GRAPES global geopotential height in the (a) Northern Hemisphere, (b)

Southern Hemisphere, (c) tropics, and (d) East Asia. The curves with plus sign and open circle show the 3D-Var and 4D-Var results, respectively. The dashed and solid curves show the background field and analysis field, respectively.

In terms of wind field analysis, the major improvement provided by the 4D-Var is in the troposphere in the Northern Hemisphere, in the Southern Hemisphere and in the tropical higher layers (Figure 3). The biggest improvement is in the tropical higher layers, where the RMSE of U wind is reduced by 1 m/s. In terms of humidity analysis, the improvement provided by the 4D-Var is marginal (not shown).





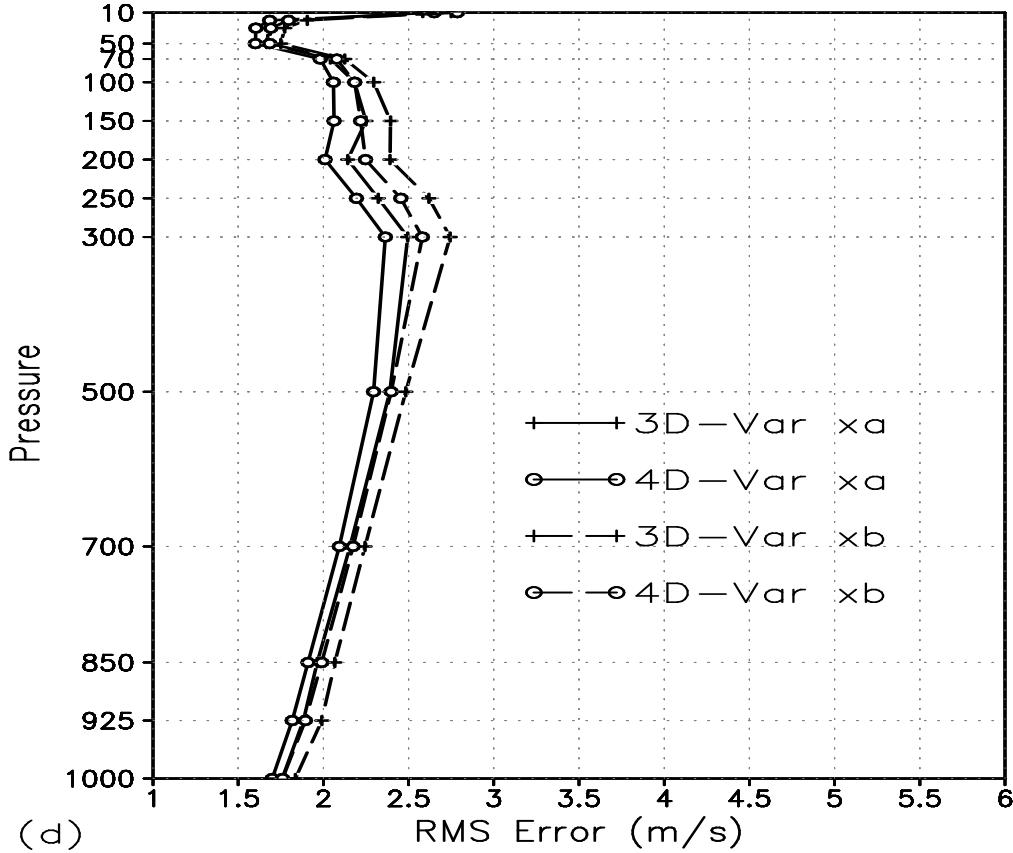


Figure 3. The RMSE (measured against of the ERA INTERIM reanalysis) of the U wind background and analysis fields in the (a) Northern Hemisphere, (b) Southern Hemisphere, (c) tropics, and (d) East Asia. The curves with plus sign and open circle show the 3D-Var and 4D-Var results, respectively. The dashed and solid curves show the background field and the analysis field, respectively.

On average, the performance of the GRAPES global 4D-Var is very stable. The 4D-Var showed advantages over the 3D-Var throughout the retrospective trials. As shown in Figure 4, the RMSE of the 500 hPa background field in the 4D-Var is approximately 7 geopotential metres, whereas that of the 3D-Var is approximately 9 geopotential metres. Note that there were two periods without the satellite radiance data between 15:00, 27 July and 21:00, 28 July and between 21:00, 17 October and 21:00, 18 October. In both periods, the 3D-Var and 4D-Var analyses showed abnormal results. The 3D-Var and 4D-Var rapidly returned to the normal standard after these two periods.

The average error of the background field relative to the observation field is a general

indicator for assessing the quality of an analysis system, since it represents the short-term (6 hours) forecasting level of the analysis field. When the radiosonde observation is used as a criterion, the RMSEs of the temperature and wind background fields of the GRAPES global 4D-Var are smaller than those of the 3D-Var by 1-6% (Figure 5). When the GNSSRO refractivity is used as a criterion, the RMSE of the background field of the GRAPES global 4D-Var is also smaller than that of the 3D-Var, by 1-4% (Figure 6), which is more evident in the Southern Hemisphere.

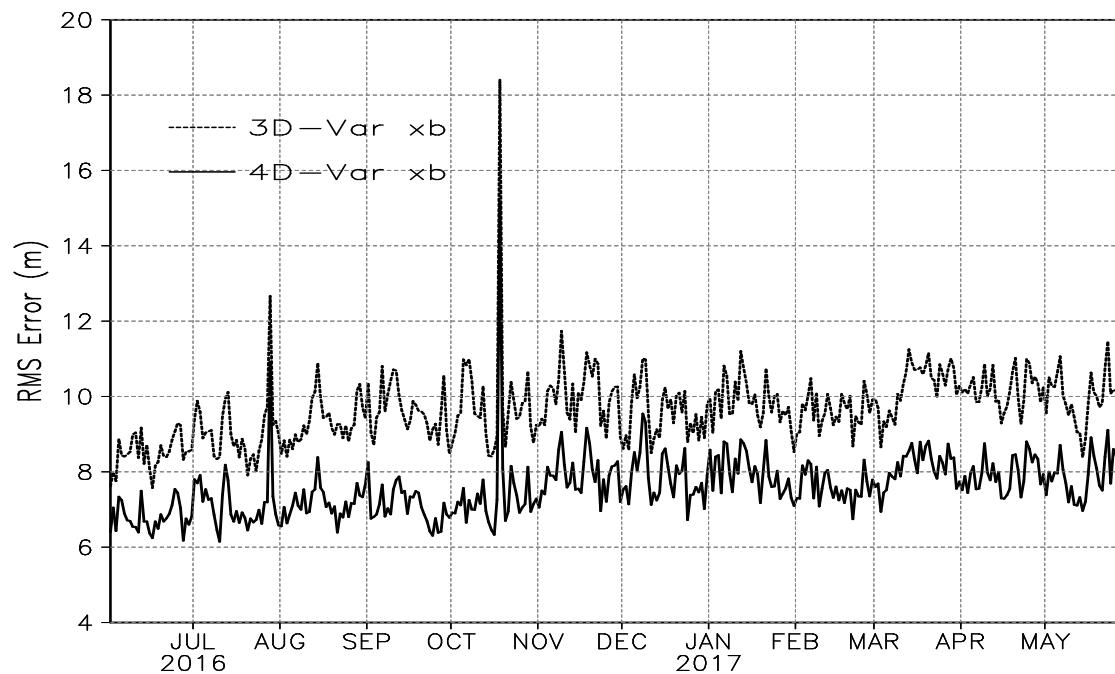


Figure 4. The temporal evolution of the global mean RMSE (measured against of the ERA INTERIM reanalysis) of the 500 hPa geopotential height. The dashed and solid curves show the 3D-Var and 4D-Var results, respectively.

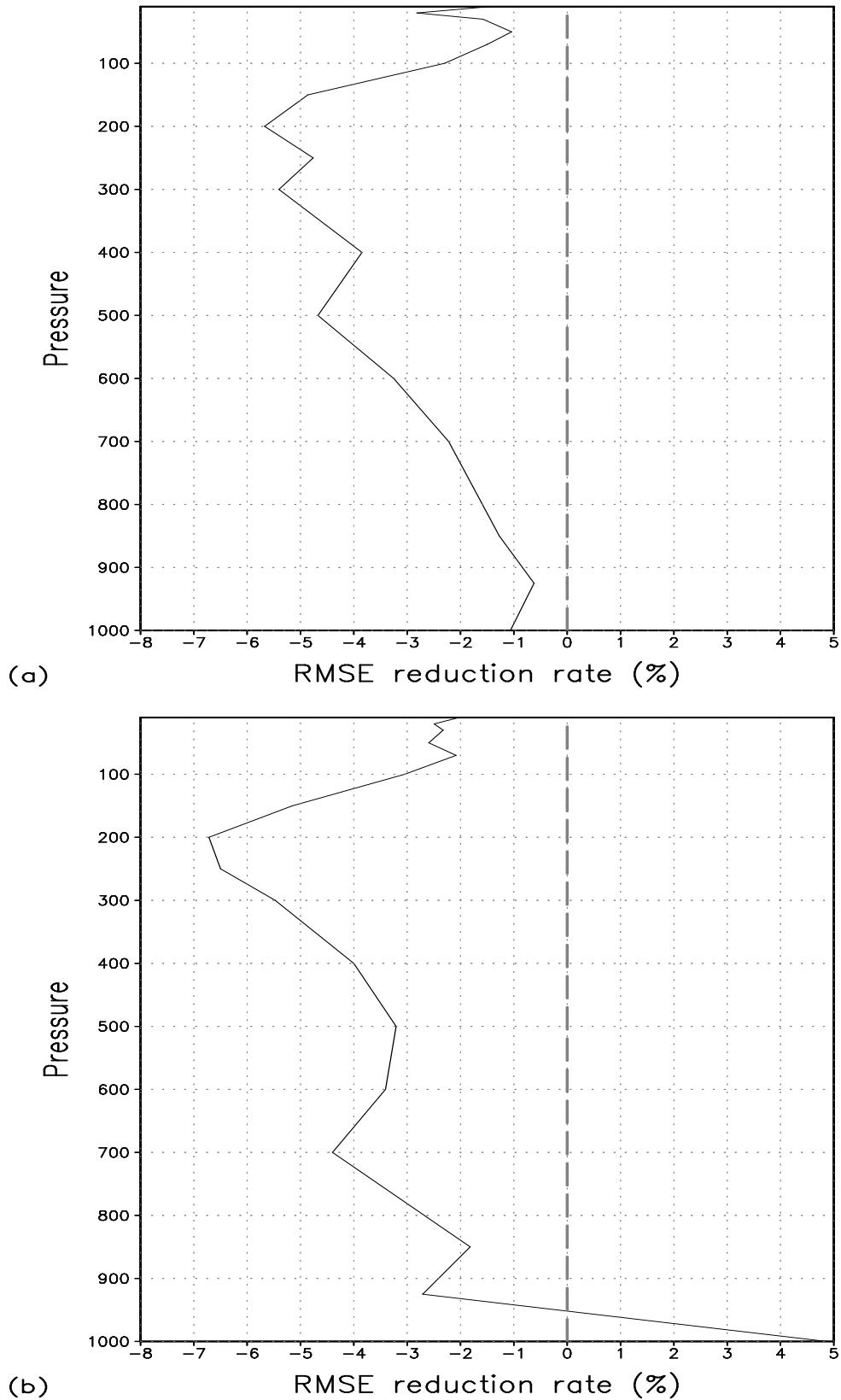


Figure 5. The RMSE reduction rate of the 4D-Var relative to the 3D-Var (global mean) for (a) temperature and (b) U wind when the radiosonde observation is used as a criterion.

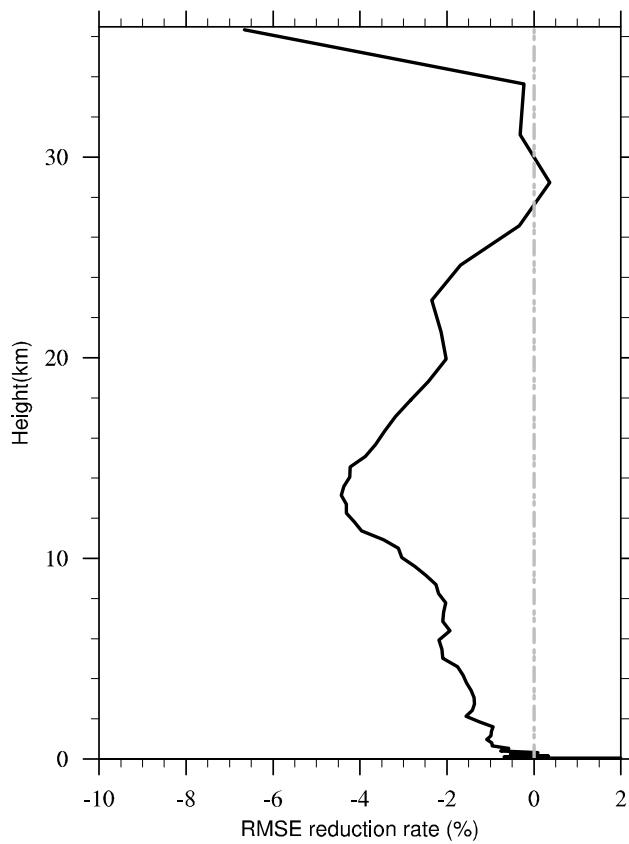


Figure 6. The RMSE reduction rate of the 4D-Var relative to the 3D-Var (global mean) when the occultation refractivity is used as a criterion.

Figure 7 shows the anomaly correlations of the 500 hPa geopotential height forecast field calculated using different analysis fields. The forecast skill of the GRAPES global 4D-Var is better than that of the 3D-Var, especially in the Southern Hemisphere. We also found that the advantage of the 4D-Var is more obvious when the anomaly correlations are calculated using the NCEP reanalysis field. This might be similar to the problem encountered by Clayton et al. (2013). In an assessment of the Met Office global hybrid-4D-Var assimilation system, Clayton et al. (2013) noted that using the own analyses to evaluate the data assimilation system might be problematic. For instance, if the analysis system modifies the background field only slightly, even though these modifications are not appropriate, they will result in better verification results.

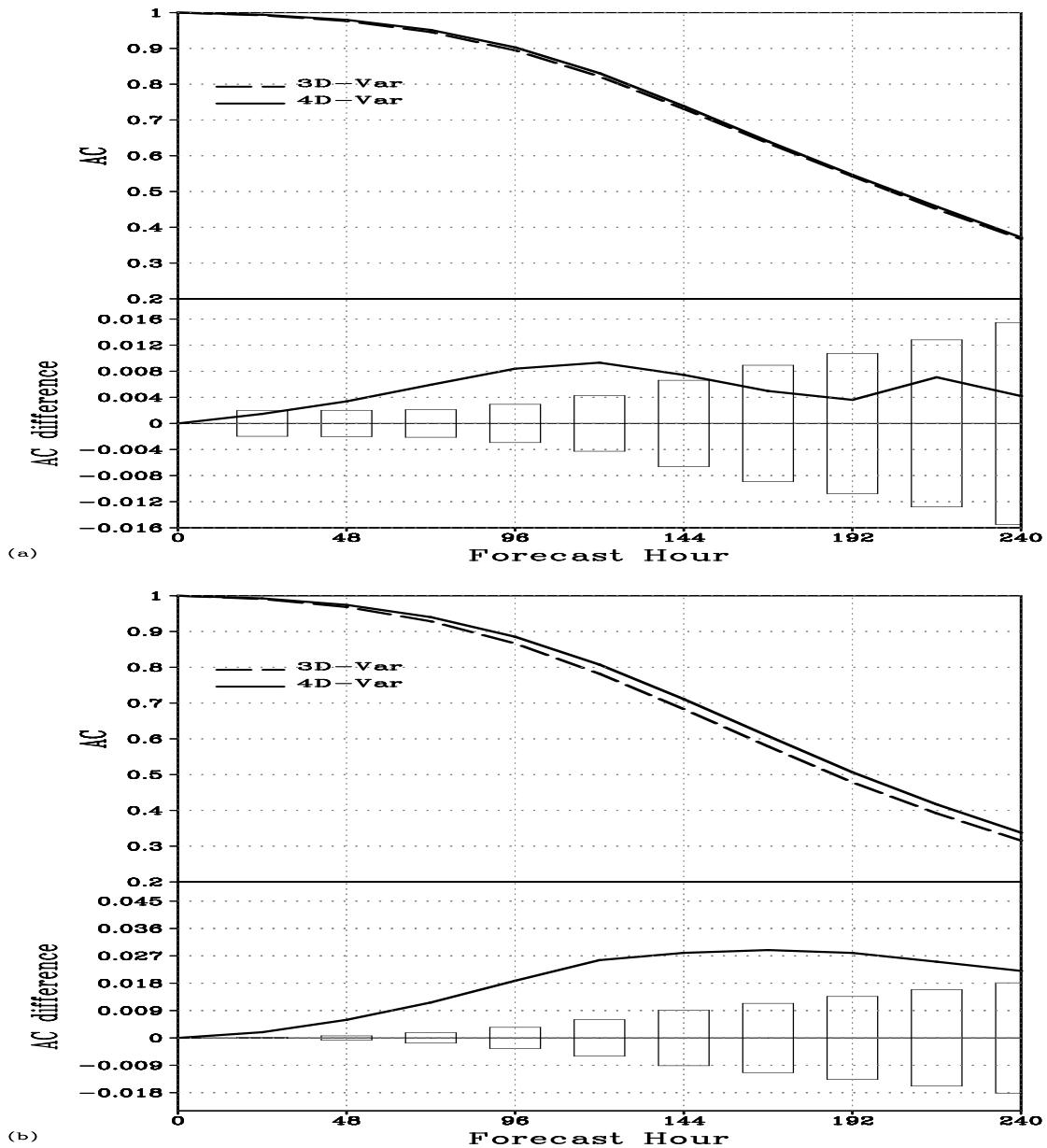


Figure 7. The anomaly correlations of the 500 hPa geopotential height for the (a) Northern Hemisphere, and (b) Southern Hemisphere. The dashed and solid curves show the 3D-Var and 4D-Var results, respectively. The AC difference for 4D-Var – 3D-Var and the 95% confidence threshold are also plotted in the bottom panels. The horizontal axis is the forecasting time in hours.

The 72-hour forecast skill of the geopotential height of the GRAPES global 4D-Var is clearly better than that of the 3D-Var in the troposphere (below 200 hPa) of the Northern Hemisphere. Throughout the Southern Hemisphere, the 4D-Var shows an improvement over 3D-Var in the 72-hour forecast skill of the geopotential height

(Figure 8).

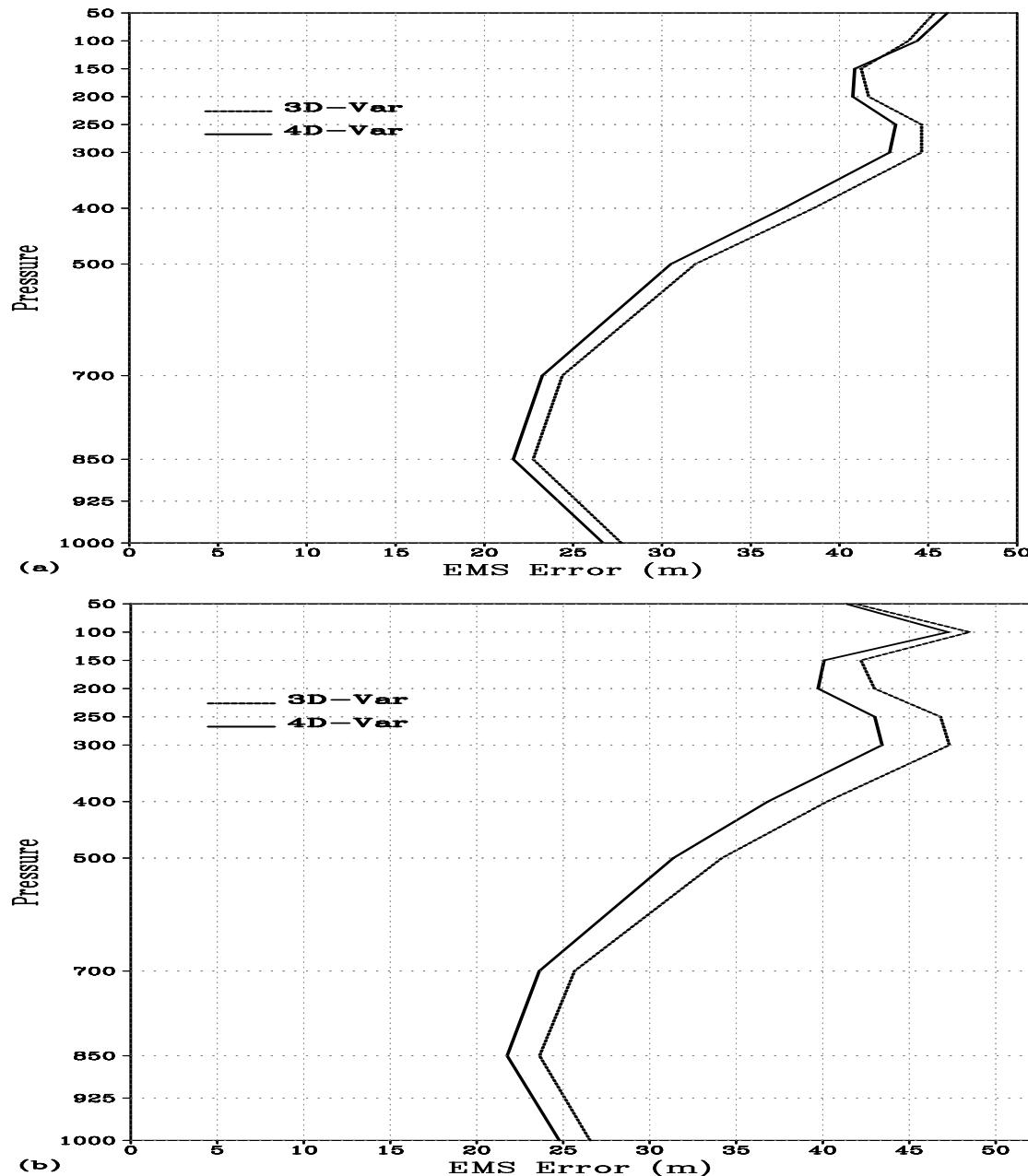


Figure 8. The vertical profile of the RMSE of the 72-hour geopotential height forecasts in the (a) Northern Hemisphere, and (b) Southern Hemisphere when the radiosonde observation is used as an inspection criterion. The dashed and solid curves show the 3D-Var and 4D-Var results, respectively.

By comparing the vertical distribution of the geopotential height forecast error on different time scales (Figure 9), one can see that the 4D-Var shows improvement below 250 hPa in the Northern Hemisphere relative to the 3D-Var. The improvement is particularly obvious for the forecast at 300 hPa over 5 days. However, positive errors occur in the upper layer of the model, primarily for the forecast at 100 hPa in the first 7 days, which is more obvious in the early stage. In the Southern Hemisphere, the 4D-Var shows an overall improvement in the forecast skill, which is most obvious above 500 hPa.

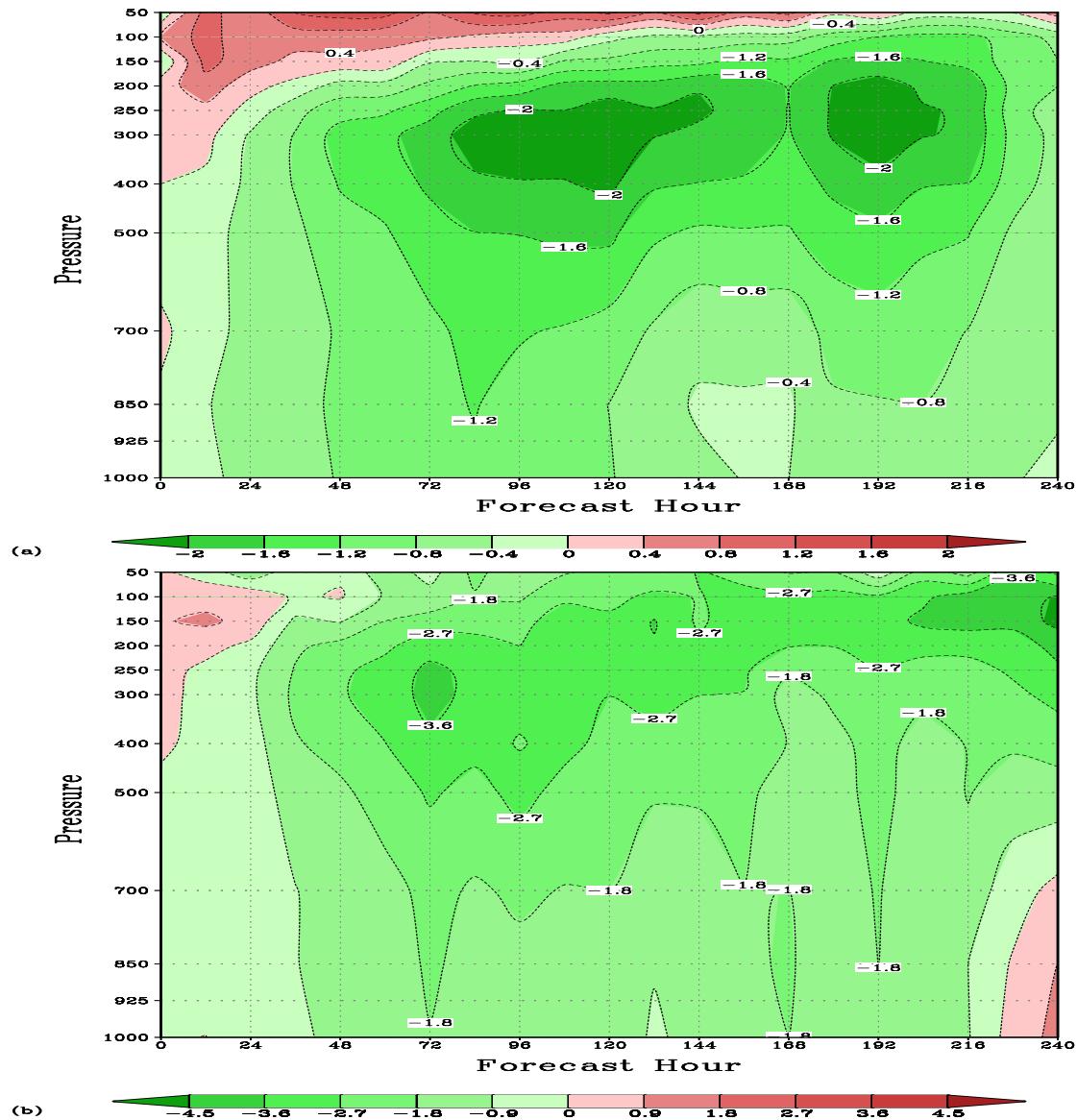


Figure 9. The reduction of the 4D-Var geopotential height forecast error relative to that of the 3D-Var (unit: geopotential metre) in the (a) Northern Hemisphere and (b) Southern Hemisphere against the radiosonde observation. The green shading indicates a reduction in the forecasting error and the red shading indicates an increase in the forecasting error. The horizontal coordinate is the forecasting time (unit: hour) and the vertical coordinate is the pressure (unit: hPa).

6. Discussion and conclusion

The GRAPES global 4D-Var data assimilation system of the CMA began operation on 1 July 2018. This system uses non-hydrostatic global TLM and ADM for the first time. In terms of computational optimization, the total computational time of the GRAPES global TLM and ADM is approximately 3-fold that of the GRAPES global NLM. Moreover, the GRAPES global 4D-Var data assimilation system uses an executable program to perform the incremental 4D-Var with a single outer loop. This reduces the data I/O through hard drives. On the high-performance computer PI-SUGON of the CMA, one operational global 4D-Var analysis takes 25 minutes when 1024 cores are used. This meets the requirement of timeliness in operation. In terms of analysis and forecast, the GRAPES global 4D-Var data assimilation system shows an overall improvement compared to the 3D-Var system, particularly in the Southern Hemisphere. The amount of observational data used in the GRAPES global 4D-Var assimilation system is 50% higher than that used in the 3D-Var. Our next focus is to take full advantage of the 4D-Var data assimilation system and to increase the observations with higher temporal resolution, especially the geostationary satellite observations. Currently, we are conducting assimilation studies on the observations from the Chinese Fengyun-4 (FY4) satellite. We already have preliminary results and will implement them very soon.

Another difference between the GRAPES global 4D-Var data assimilation system and other operational systems is that it uses the preconditioning operator of the Lanczos-CG algorithm to change the preconditioning of the control variables. We found that one construction of the Lanczos-CG preconditioning operator can accelerate the 4D-Var minimization convergence for a long time. We have performed

a one-month cycling experiments. In these experiments, we calculated the approximated eigenpairs of Hessian matrix during the 4D-Var minimization at the first DA cycle and conduct the preconditioning operator in the 4D-Var minimization at the subsequent times in one month. The averaged minimization iteration number from one-month experiments is reduced from 37 to 27 after the use of the preconditioning operator. We will use this scheme to perform the preconditioned Lanczos-CG algorithm operationally after the update of the operational 4D-Var system.

Relatively complete linear physical processes are also developed. But the linear convection parameterization and large-scale condensations scheme are not ready for the operational use. The improvement of the linear physics is the field we will focus on in the near future.

The development direction of the GRAPES global 4D-Var data assimilation system is an ensemble-variational assimilation. In fact, the Numerical Weather Prediction Centre of the CMA has begun to develop the GRAPES global Hybrid-4DVar assimilation system. The new system applies an extended control variable technique similar to that of the Met Office. This technique allows for implementation of the background error covariance, which changes with weather conditions, in the global 4D-Var. The background error covariance is statistically obtained from the ensemble-based data assimilation samples. The preliminary results show that the influence of the technique is neutral.

Acknowledgements

The development of the GRAPES global 4D-Var data assimilation system is a systematic project. Apart from the authors, many other colleagues have participated in the project. We sincerely thank the entire team for their cooperation! We also thank Dr. Lars Isaksen, Dr. Mike Fisher, and Dr. Elias Holm from the ECMWF and Dr. Dingming Li from the Met Office for their thorough discussions with the development team of the GRAPES global 4D-Var system.

References

- Alpert, J.C., 1988: Sub-grid scale mountain blocking at NCEP. 20 th Conf. WAF/16 Conf. NWP P2.4.
- Bonavita M, Isaksen L, H'olm E., 2012: On the use of EDA background error variances in the ECMWF 4D-Var. Q. J. R. Meteorol. Soc. 138: 1540–1559. DOI:10.1002/qj.1899.
- Buehner, M., R. McTaggart-Cowan, A. Beaulne, et al., 2015: Implementation of deterministic weather forecasting systems based on ensemble-variational data assimilation at Environment Canada. Part I: The global system. Mon. Wea. Rev., 143, 2532-2559, DOI: 10.1175/MWR-D-14-00354.1.
- Chen D H, Xue J S, Yang X S, et al., 2008: New generation of multi-scale NWP system (GRAPES): general scientific design. Chin Sci Bull, 53(22): 3433-3445
- Clayton, A. M., A. C. Lorenc, and D. M. Barker, 2013: Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. Q. J. R. Meteorol. Soc., 139, 1445–1461, doi:10.1002/qj.2054.
- Courtier P, Thepaut JN, Hollingsworth A., 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. Q. J. R. Meteorol. Soc., 120: 1367-1388.
- Courtier P, Andersson E, Heckly W, et al., 1998: The ECMWF implementation of three dimensional variational assimilation (3D-Var). Part I: Formulation. Q. J. R. Meteorol. Soc., 124:1783-1808.
- Fisher M., 1995: Estimating the covariance matrices of analysis and forecast error in variational data assimilation. ECMWF Tech. Memo. 220, ECMWF, Reading, UK.
- Fisher M. ,1998: Minimization algorithms for variational data assimilation. In Seminar on Recent developments in numerical methods for atmospheric modelling. ECMWF.364-385.
- Gauthier P, Charette C, Fillion L, et al., 1999: Implementation of a 3D variational data assimilation system at the Canadian Meteorological Centre. Part I: The global analysis. Atmosphere-Ocean, 37:103-156.
- Gauthier, P., M. Tanguay, S. Laroche, S. Pellerin, and J. Morneau, 2007: Extension of 3DVAR to 4DVAR: Implementation of 4DVAR at the Meteorological Service of

- Canada. Mon. Wea. Rev., 135, 2339–2354, doi:10.1175/MWR3394.1.
- Han J, Pan H L. 2006: Sensitivity of Hurricane Intensity Forecast to Convective Momentum Transport Parameterization. Mon. Wea. Rev., 1134(2): 664-674.
- Honda Y, Nishijima M, Koizumi K, et al., 2005: A pre-operational variational data assimilation system for a non-hydrostatic model at the Japan Meteorological Agency: Formulation and preliminary results. Q. J. R. Meteorol. Soc., 131:3465-3475.
- Hong, Song-You, Hua-Lu Pan, 1996: Nonlocal boundary layer vertical diffusion in a medium-range forecast model. Mon. Wea. Rev., 124: 2322-2339
- Ingleby N. Bruce, 2001: The statistical structure of forecast errors and its representation in the Met. Office Global 3-D Variational Data Assimilation Scheme, Q. J. R. Meteorol. Soc., 127:209-231.
- Isaksen L., M. Bonavita, R. Buizza, M. Fisher, J. Haseler, M. Leutbecher, Laure Raynaud, 2010: Ensemble of data assimilations at ECMWF. ECMWF Tech. Memo. 636, ECMWF, Reading, UK.
- Janisková M, Thépaut J-N, Geleyn, J-F. 1999: Simplified and regular physical parameterisations for incremental four-dimensional variational assimilation. Mon. Weather Rev., 127: 26–45.
- Kadowaki, T., 2005: ‘A 4-dimensional variational assimilation system for the JMA Global Spectrum Model’. Research Activities in Atmospheric and Oceanic Modelling. Working Group for Numerical Experimentation (WGNE), Blue Book Series (J.Coté, Ed).
- Kleist, D. T., and K. Ide, 2015: An OSSE-based evaluation of hybrid variational–ensemble data assimilation for the NCEP GFS. Part II: 4DEnVar and Hybrid Variants. Mon. Wea. Rev., 143, 452–470, doi:10.1175/MWR-D-13-00350.1.
- Laurent, H., 2007: TAPENADE, Automatic differentiation by program transformation. <http://www-sop.inria.fr/tropics>.
- Liu DC, Nocedal J., 1989: On the limited memory BFGS method for large scale optimization. Mathematical Programming, 45:503-528.
- Liu, Y., L. Zhang, Z. Lian, 2018: Conjugate Gradient Algorithm in the Four-Dimensional

- Variational Data Assimilation System of GRAPES (in Chinese). *Journal of Meteorological Research*, doi: 10.1007/s13351-018-8004-y
- Lorenc AC, Ballard SP, Bell RS, et al., 2000: The Met Office global three-dimensional variational data assimilation scheme. *Q. J. R. Meteorol. Soc.*, 126:2991-3012.
- Lorenc, A.C. and F. Rawlins, 2005: Why does 4D-Var beat 3D-Var? *Q. J. R. Meteorol. Soc.*, 131:3247-3257.
- Lynch P. and Xiang-Yu Huang, 1992: Initialization of the HIRLAM model using a digital filter. *Mon. Wea. Rev.*, 120: 1019-1033.
- Parrish DF, Derber JC., 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, 120:1747-1763.
- Rabier, F., H. Järvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Q. J. R. Meteorol. Soc.*, 126:1143-1170, doi:10.1002/qj.49712656415.
- Radnóti G, Trémolet Y, Andersson E, Isaksen L, Hólm E, Janisková M. 2005. Diagnostics of linear and incremental approximations in 4D-Var revisited for higher resolution analysis. *ECMWF Tech. Memo.* 467, ECMWF, Reading, UK.
- Rawlins, F., S. Ballard, K. Bovis, A. Clayton, D. Li, G. Inverarity, A. Lorenc, and T. Payne, 2007: The Met Office four-dimensional variational data assimilation scheme. *Q. J. R. Meteorol. Soc.*, 133:347–362, doi:10.1002/qj.32.
- Tompkins A M, Janisková M., 2004: A cloud scheme for data assimilation: Description and initial tests. *Q. J. R. Meteorol. Soc.*, 130(602):2495-2517.
- Wang, J., Lu H., Han W., et al., 2017: Improvements and performances of the operational GRAPES_GFS 3DVar system (in Chinese). *Journal of Applied Meteorological Science*, 28(1): 11-24. DOI: 10.11898/1001-7313.20170102
- Wee, TK and YH Kuo, 2004: Impact of a digital filter as a weak constraint in MM5 4DVAR: An observing system simulation experiment. *Mon. Wea. Rev.*, 132: 543-559.
- Wu W S, Purser R J, Parrish D F. 2002: Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.*, 130(12): 2905-2916.

Xue J S, Zhuang S Y, ZHU G F, et al., 2008: Scientific design and preliminary results of three-dimensional variational data assimilation system of GRAPES, Chin Sci Bull, 53(22): 3446-3457.

Figure legend section

Figure 1. The relative error in the tangent linear model with simple physics with respect to the nonlinear forecast model with full physics at the resolution of 1.0° , the solid line gives the 1-h forecast results, the dashed line represents the 3-h forecast results, and the dotted line shows the 6-h forecast results, (a) u wind; (b) non-dimensional pressure; (c) potential temperature; (d) specific humidity. **Figure 2.** The RMSE of the background and analysis fields of the GRAPES global geopotential height in the (a) Northern Hemisphere, (b) Southern Hemisphere, (c) tropics, and (d) East Asia. The curves with plus sign and open circle show the 3D-Var and 4D-Var results, respectively. The dashed and solid curves show the background field and analysis field, respectively.

Figure 3. The RMSE of the U wind background and analysis fields in the (a) Northern Hemisphere, (b) Southern Hemisphere, (c) tropics, and (d) East Asia. The curves with plus sign and open circle show the 3D-Var and 4D-Var results, respectively. The dashed and solid curves show the background field and the analysis field, respectively.

Figure 4. The temporal evolution of the global mean RMSE of the 500 hPa geopotential height. The dashed and solid curves show the 3D-Var and 4D-Var results, respectively.

Figure 5. The RMSE reduction rate of the 4D-Var relative to the 3D-Var (global mean) for (a) temperature and (b) U wind when the radiosonde observation is used as a criterion.

Figure 6. The RMSE reduction rate of the 4D-Var relative to the 3D-Var (global mean) when the occultation refractive rate is used as a criterion.

Figure 7. The anomaly correlations of the 500 hPa geopotential height for the (a) Northern Hemisphere, and (b) Southern Hemisphere. The dashed and solid curves show the 3D-Var and 4D-Var results, respectively. The AC difference for 4D-Var – 3D-Var and the 95% confidence threshold are also plotted in the bottom panels. The horizontal axis is the forecasting time in hours.

Figure 8. The vertical profile of the RMSE of the 72-hour geopotential height forecasts in the (a) Northern Hemisphere, and (b) Southern Hemisphere when the radiosonde observation is used as an inspection criterion. The dashed and solid curves show the 3D-Var and 4D-Var results,

respectively.

Figure 9. The reduction of the 4D-Var geopotential height forecast error relative to that of the 3D-Var (unit: geopotential metre) in the (a) Northern Hemisphere and (b) Southern Hemisphere against the radiosonde observation. The green shading indicates a reduction in the forecasting error and the red shading indicates an increase in the forecasting error. The horizontal coordinate is the forecasting time (unit: hour) and the vertical coordinate is the pressure (unit: hPa).