A Comparison between the 3/4DVAR and Hybrid Ensemble-VAR Techniques for Radar Data Assimilation

Hongli Wang[*1], Xiang-Yu Huang[1], Juanzhen Sun[1], Dongmei Xu[1], Shuiyong Fan[2], Jiqin Zhong[2] and Man Zhang[3]

[1]*National Center for Atmospheric Research, Boulder, CO, USA*
[2]*Beijing Meteorological Bureau, Beijing, China*
[3]*University of Colorado, Boulder, CO*

ABSTRACT

In variational data assimilation systems using climatological background error covariances, flow-dependent background error covariances can be introduced by hybrid ensemble-variational or 4-dimensional variational data assimilation techniques. In this paper, the features of climatological background error modeling via the NMC method are first investigated for the Weather Research and Forecasting Model's variational data assimilation (WRF-Var) system, then flow-dependent background error features introduced by hybrid ensemble-3D-Var or 4D-Var data assimilation techniques are investigated. The background error statistics are extracted from the short-term 3km resolution forecasts in June, July and August 2012. It is found that (1) background error variances vary month to month and also have a feature of diurnal variations in low-level atmosphere; (2) u- and v-wind variances are underestimated and their auto-correlations are overestimated when the default control variable option in the WRF-Var is used. Two additional control variable transforms are proposed and described to form background error covariance matrix via the NMC method. Single observation data assimilation experiments demonstrated capability of both hybrid and 4D-Var methods are effective to introduce flow-dependent background error covariances. A case study using WRF 4D-Var assimilating radar radial velocity observation shows that precipitation location forecast is slightly better than that 3D-Var or hybrid background error formulation.

Corresponding author:

Dr. Hongli Wang

hlwang@ucar.edu

MMM/NCAR

1. Introduction

The background error (BE) covariance matrix plays a key role in a variational data assimilation system by weighting the importance of a priori state, by smoothing and spreading information from observation points, and by imposing balance between the model control variables (Daley 1991; Bannister 2008a,b). However, the estimation of the BE statistics is not straightforward, since the truth is not known. Two methods are mainly used in current data assimilation systems. The so-called NMC (named for the National Meteorological Center, now called the National Centers for Environmental Prediction) method (Parrish and Derber 1992) is one approach that is widely employed to estimate the BE covariances. This method uses the differences between forecasts of different lengths, but valid at the same time, to evaluate the short-range forecast errors. An alternative method is to use an ensemble of short-term forecasts at a specific time to evaluate the BE covariances (Houtekamer et al. 1996; Fisher 2003).

Various control variable transforms (CVTs) have been used in variational data assimilation systems to model multivariate and univariate aspects of the BE covariances approximately in a compact and efficient way (Bannister 2008a,b). Three kinds of BE modeling for wind analysis are widely used in variational data assimilation. Vorticity and (unbalanced) convergence/divergence are used in data assimilation systems at ECMWF (Courtier et al. 1996) and Meteo France (Fischer et al. 2005). Streamfunction and unbalanced velocity potential are widely used as control variables in global data assimilation systems, and some regional data assimilation systems (e.g. Ingleby 2001; Barker et al. 2004, 2012; Zupanski 2005; Rawlins et al. 2007; Huang et al. 2009; Wang et al. 2013). Whereas, velocities are employed as control variables in data assimilation systems for mesoscale and convective scales (e.g. Zou et al. 1995; Sun and Crook 1997; Gao et al. 1999; Zupanski et al. 2005; Kawabata et al. 2011). The velocity control variables may be more suitable for mesoscale and convective scale data assimilation since past theoretical analysis found that velocity control variables could combine the background and observations for all scales (Xie et al. 2002; Xie and MacDonals 2011).

Assumptions are made to model BE covariances in an efficient and affordable way since the BE matrix is of high dimensions. The present numerical weather prediction model uses a large dimensional space, typically 10^7 dimensions or more, and so the BE matrix has 10^{14} elements, which cannot be explicitly modeled. In practice, it is usually assumed that the BE covariances are nearly homogeneous and isotropic. For example, the present Weather Research and Forecasting (WRF) model's variational data assimilation (WRF-Var) system assumes that the BE covariances are homogeneous and isotropic. It is noted that both choice of CVT and assumptions made to model BE covariances as mentioned above could have impact on extracting BE information from forecast examples using either NMC or the ensemble method.

The WRF-Var data assimilation system has been extensively used in research community and operational centers (Barker et al. 2012; Huang et al. 2013). For examples, WRF-Var was adopted in the Rapid Update Cycling Data assimilation and Forecasting System at Beijing Meteorological Bureau (BJ-RUC; Chen et al. 2009), which has been run in operation since June 2008. The WRF-Var system with radar data assimilation has shown consistently positive impact on precipitation prediction (Wang et al. 2013a). The operational application of the WRF-Var system at Taiwan's Central Weather Bureau (CWB) has significantly reduced typhoon track forecast errors (Hisao et al. 2012).

The NMC method has been employed in the specification of BE statistics for WRF-Var. WRF-Var is the basic component of the WRF model's community variational/ensemble data assimilation system (WRFDA; Barker et al. 2012). Barker et al. (2004) suggested to apply empirical tuning factors to the length scales calculated via the NMC method (ranging between 0.5 and 1). Past studies (e.g. Xiao and Sun 2007; Sugimoto 2009; Li et al. 2012; Sun et al. 2012; Wang et al. 2013a) showed that radar radial velocity data assimilation using WRF-Var system with reduced lengthscales improved analyses and forecasts. These provide the motivation to investigate the features of BE modeling via the NMC method to further improve the performance of the WRF-Var system with the aim at mesoscale and convective data assimilation using high-resolution observations such as radar data. Moreover, investigations on BE modeling for WRF-Var will also benefit WRFDA. Flow-dependent background error features introduced by hybrid ensemble-3D-Var or 4D-Var data assimilation techniques are investigated as well.

This paper is organized as follows. Section 2 provides a description of the NMC method and features of BE statistics over Beijing region. Two new CVTs are introduced to use climatological background errors with the NMC method in section 3. Single observation and real radar data assimilation experiments are presented in section 4. A summary and discussion is given in the final section.

2. Background error modeling

2.1 The NMC method

A common method to model BE covariance matrix is to take the difference between pairs of forecasts of different lead times but each valid at the same time (Parrish and Derber 1992). Forecast differences are usually calculated over a reasonably long period of time (e.g. a month). This makes the NMC method suitable for climatological forecast error statistics. In WRF-Var, the background error covariance matrix may be considered the following expression

$$\mathbf{B} \approx \overline{(\mathbf{x}^{24}-\mathbf{x}^{12})(\mathbf{x}^{24}-\mathbf{x}^{12})^{\mathrm{T}}} \quad (1)$$

where $\mathbf{x}^{\mathbf{24}}$ and $\mathbf{x}^{\mathbf{12}}$ are 24 h and 12 h forecasts respectively valid at the same time. The overbar denotes an average over time and/or space. The two forecasts can be written in terms of "truth" and their errors

$$\mathbf{x}^{24} = \mathbf{x}^{\mathrm{truth}} + \boldsymbol{\varepsilon}^{24} + \mathbf{b}^{24} \quad (2a)$$

$$\mathbf{x}^{12} = \mathbf{x}^{\mathrm{truth}} + \boldsymbol{\varepsilon}^{12} + \mathbf{b}^{12} \quad (2b)$$

Here, $\mathbf{x}^{\mathrm{truth}}$ is the true atmospheric state at the valid time. $\boldsymbol{\varepsilon}^{24}$ and $\boldsymbol{\varepsilon}^{12}$ are the random errors, and $\mathbf{b}^{24}$ and $\mathbf{b}^{12}$ are the biases in each forecast. Assuming there is no bias or the bias is constant in time, $\mathbf{b}^{24}=\mathbf{b}^{12}$, the forecast difference is

$$\mathbf{x}^{\mathrm{diff}} = \boldsymbol{\varepsilon}^{24} - \boldsymbol{\varepsilon}^{12} \quad (3)$$

The BE covariance matrix is written as

$$\begin{aligned}\mathbf{B} &\approx \overline{(\mathbf{x}^{\text{diff}})(\mathbf{x}^{\text{diff}})^{\mathrm{T}}} \\ &= \overline{(\boldsymbol{\varepsilon}^{24}-\boldsymbol{\varepsilon}^{12})(\boldsymbol{\varepsilon}^{24}-\boldsymbol{\varepsilon}^{12})^{\mathrm{T}}} \\ &= \overline{(\boldsymbol{\varepsilon}^{24})(\boldsymbol{\varepsilon}^{24})^{\mathrm{T}}} + \overline{(\boldsymbol{\varepsilon}^{12})(\boldsymbol{\varepsilon}^{12})^{\mathrm{T}}} - \overline{(\boldsymbol{\varepsilon}^{24})(\boldsymbol{\varepsilon}^{12})^{\mathrm{T}}} - \overline{(\boldsymbol{\varepsilon}^{12})(\boldsymbol{\varepsilon}^{24})^{\mathrm{T}}} \quad (4)\end{aligned}$$

It is seen that BE modeling using Eq.1 including three parts: 24 h BE, 12 h BE and their correlations.

2.2 $\mathbf{B}$ *modeling in WRF-Var*

In the WRF-Var system, a control variable transform $\delta\mathbf{x}=\mathbf{U}\mathbf{v}$ is used to model background errors. $\mathbf{v}$ represents control variable vector and $\delta\mathbf{x}$ stands for analysis increment vector. The $\mathbf{U}$ transform maps control variables from control space to analysis space. The CVT $\delta\mathbf{x}=\mathbf{U}\mathbf{v}$ is implemented through a series of operations $\delta\mathbf{x}=\mathbf{U}_p\mathbf{U}_v\mathbf{U}_h\mathbf{v}$ (Barker et al. 2004). The default control variables (CV option 5; CV5) in WRF-Var includes the streamfunction ψ, the unbalanced part of velocity potential χ_u, the unbalanced part of temperature T_u, the unbalanced part of surface pressure Ps_u, and pseudo–relative humidity RH. The term "unbalance" refers to the residual from the balance with the streamfunction.

The operators $\mathbf{U}_p$, $\mathbf{U}_v$ and $\mathbf{U}_h$ are briefly described here. Readers are referred to Barker et al. (2004) for details. The horizontal transform $\mathbf{U}_h$ is to model auto-correlation of a control variable using recursive filters. The horizontal correlations are assumed to be homogeneous (i.e. not dependent on geographic position) and isotropic for each control variable. The vertical transform $\mathbf{U}_v$ is performed via an empirical orthogonal function (EOF) decomposition of the vertical component of BE on model levels. The variances and vertical correlations of each control variable are modeled in this stage. In default CV5, the time- and domain-averaged vertical component of BE is used indicating that the variances and vertical correlations are constant on each model level and do not depend on geographic positions. The physical variable transform $\mathbf{U}_p$ involves balance transform and conversion of control variables to analysis variable increments. The statistical balance transform is applied in this stage, defined by

$$\begin{pmatrix}\psi\\ \chi\\ T\\ Ps\\ RH\end{pmatrix}=\begin{pmatrix}\mathbf{I}&0&0&0&0\\ \mathbf{C}_{\chi,\psi}&\mathbf{I}&0&0&0\\ \mathbf{C}_{T,\psi}&0&\mathbf{I}&0&0\\ \mathbf{C}_{Ps,\psi}&0&0&\mathbf{I}&0\\ 0&0&0&0&\mathbf{I}\end{pmatrix}\begin{pmatrix}\psi\\ \chi_u\\ T_u\\ Ps_u\\ RH\end{pmatrix} \quad (5)$$

where $\mathbf{I}$ is the identity matrix, and $\mathbf{C}_{\chi,\psi}$, $\mathbf{C}_{T,\psi}$ and $\mathbf{C}_{Ps,\psi}$ stand for statistical regression matrixes between χ, T, Ps and ψ. The analysis variables of u-wind (u), v-wind (v), and specific humidity (q) can be obtained by a transform as follows.

$$\begin{pmatrix} u \\ v \\ T \\ Ps \\ q \end{pmatrix} = \begin{pmatrix} \mathbf{C}_{u,\psi} & \mathbf{C}_{u,\chi} & 0 & 0 & 0 \\ \mathbf{C}_{v,\psi} & \mathbf{C}_{v,\chi} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{I} & 0 & 0 \\ 0 & 0 & 0 & \mathbf{I} & 0 \\ 0 & 0 & 0 & 0 & \mathbf{C}_{q,rh} \end{pmatrix} \begin{pmatrix} \psi \\ \chi \\ T \\ Ps \\ RH \end{pmatrix} \tag{6}$$

$\mathbf{C}_{u,\psi}$, $\mathbf{C}_{u,\chi}$, $\mathbf{C}_{v,\psi}$, $\mathbf{C}_{v,\chi}$ and $\mathbf{C}_{q,rh}$ map variables ψ, χ, T, Ps and RH to analysis variables u, v, T, Ps and q.

2.3 Features of $\mathbf{B}$ *variance modeling*

It is seen that the CVT and several assumptions are taken to model the BE matrix approximately in a compact and efficient way. A natural question to ask is: what information is lost/filtered out by CVT and the above assumptions?

The operational WRF 24 h and 12 h forecasts from BJ_RUC in June, July and Aug 2012 are used to calculate short-term BE statistics in this paper. The statistics are over the inner domain (D2, Fig.1) with 3 km grid spacing. Readers are referred to Chen et al. (2009) for a detailed description on BJ_RUC. The forecast differences between 24 h and 12 h forecasts at the same time are employed to model the BE statistics. The standard deviation in the three-month forecasts directly estimated by Eq. (1) ***without*** CVT is named as STD_NMC. The standard deviation derived from $\mathbf{B}$ that is generated by WRF-Var gen_be utility is named as STD_CV5. This is achieved through sampling the $\mathbf{B}$ matrix in control variable space and then computing statistics in analysis space after CVT (Andersson et al. 2000). 200 samples are taken in this paper. Single observation experiment is another alternative method to show BE variance in observation space, which will be used in section 3.

The monthly variations of background errors are found. Figure 2 depicts the BE variances for u, v, T, and RH. It is seen that the BE variances for wind and temperature in June and July are relatively larger than that in August. The diurnal variations of forecast error near surface are clearly shown in Fig. 3. The error variances for wind and temperature in the low atmosphere at evening12Z (local time 20Z) are larger than that at morning 00Z (local time 08Z). The above results indicate that even with the climatological BE statistics the time dependent variances can be achieved. Surface observations are important data sources for a regional rapid update cycle data assimilation system. The results indicate that BE covariances accounting for diurnal variation may benefit the surface data assimilation.

Practically CVT (CV5) is used to model background errors for WRF-Var. We also compare STD_NMC to that modeled by WRF-Var (STD_CV5) to answer the question: what signals are filtered by WRF-Var CVT? The vertical profiles of STD_NMC and STD_CV5 are plotted in Fig. 4. It is seen that STDs of all the variables are underestimated and especially for u, v. In the BJ_RUC system for radar data assimilation, the lengthscales are tuned to be half of the original ones. With tuned $\mathbf{B}$

(CV5), the u, v variances (STD_CV5_L05 in Fig. 4) are comparable to STD_NMC. This indicates CVT in WRF-Var may contribute to the significant underestimates of the background error STD of winds.

3. New CVTs for $\mathbf{B}$ modeling

In this paper, two CVTs are proposed to account background errors statistics in the NMC forecast differences. A natural choice is to use u and v as control variables for wind analyses since that u and v has been used as control variables in data assimilation systems for mesoscale and convective-scale forecasts (e.g. Zou et al. 1995; Sun and Crook 1997; Gao et al. 1999; Zupanski et al. 2005; Kawabata et al. 2011).

The new formulation, which uses u, v, T, Ps and RH_s (pseudo relative humidity) as control variables, is developed in WRF-Var for mesoscale and convective-scale data assimilation. The new control variable transform can be written as

$$\delta\mathbf{x} = \mathbf{U}_{p2}\mathbf{U}_{v}\mathbf{U}_{h}\mathbf{v}_{2} \tag{7}$$

We followed WRF-Var procedure to use recursive filter and EOFs to model horizontal and vertical correlations respectively which are implemented through $\mathbf{U}_v\mathbf{U}_h$. The homogeneous and isotropic filters, which are used for CV5, are applied to each control variable. In addition, the time- and domain-averaged vertical component of BE is used indicating that the BE statistics do not depend on geographic position. It is noted that $\mathbf{U}_{p2}$ in Eq.7 only involves conversion analysis increments relative humidity to specific humidity as shown in Eq.6. No statistical balance transform (Eq.5) is applied in this transform.

Another possible approach that directly uses forecast differences in the NMC method to form BE covariance matrix is described here. The analysis increment is expressed in a subspace expanded by the NMC forecast differences

$$\mathbf{x'} = \sum_{k=1}^{K}\left(\mathbf{a}_k \circ \mathbf{x}_k^d\right) \tag{8}$$

where K is the total number of forecast differences, and the vectors $\mathbf{x}_k^d$ ($k = 1, K$) is the kth unbiased difference of forecast pairs normalized by $K^{1/2}$.

$$\mathbf{x}_k^d = \left(\mathbf{x}_k^{diff} - \overline{\mathbf{x}}\right)/\sqrt{K} \tag{9}$$

In practice, the time-averaged bias $\overline{\mathbf{x}}$ is removed from the forecast differences. The vector $\mathbf{a}_k$ stands for the augmented control variables for the kth forecast difference. $\mathbf{a}_k$ will be called "alpha" control variable hereafter in this paper. The symbol $\circ$ denotes the Schur product of the vectors $\mathbf{a}_k$ and $\mathbf{x}_k^d$. Let $\mathbf{X'} = (\mathbf{x}_1^d, \mathbf{x}_2^d, ..., \mathbf{x}_k^d, ..., \mathbf{x}_K^d)$, it is obvious that $\left(\mathbf{X'}\mathbf{X'}^{\mathrm{T}}\right) \circ \mathbf{S}$ is the $\mathbf{B}$ covariance matrix defined in Eq.1 but with covariance localization defined by $\mathbf{S} = \left\langle \mathbf{a}_k(\mathbf{a}_k)^{\mathrm{T}} \right\rangle$.

The transform (Eq.8) has been developed in the WRF hybrid ensemble-3DVar data assimilation system. The readers are referred to Wang et al. (2008) for details. In their scheme, an ensemble of forecast perturbations was used to incorporate flow dependent error covariance of the day. We adopt this idea but use the NMC forecast differences instead of ensemble perturbations to form the BE covariance matrix.

In the WRF hybrid ensemble-3DVar system, both the horizontal and vertical localization can be applied. Specifically, the horizontal and vertical correlation localizations are implemented through recursive filters (Wang et al. 2008) and vertical correlation matrix respectively. A general formulation to form the vertical covariance matrix can be written as

$$Cov(l_1,l_2) = \exp\left(-\frac{d(l_1,l_2)^2}{D(l_1)^2}\right) \qquad (10)$$

where $Cov(l_1,l_2)$ represents the correlation between model levels l_1 and l_2. d is the distance in a specified coordinate between model level l_1 and l_2 and D stands for the level-dependent vertical localization radius.

The default vertical correlation matrix in WRF-Var is defined in model level space, specifically, $d(l_1,l_2) = l_2 - l_1$, $D(l_1) = 10\frac{l_1}{N}$, N is the total number of model levels.

$$Cov(l_1,l_2) = \exp\left(-\frac{(l_2-l_1)^2}{(10\frac{l_1}{N})^2}\right) \qquad (11)$$

It is seen that the level-dependent localization radius $D(l_1)$ only depends on the number index of model level indicating that observation at model level with large number index will be widely spread in vertical direction. This may reduce the impact of observations that are located in low model levels.

In addition to the above formulation (Eq.11), a specific application of Eq.10 in *height* coordinate is also tested

$$Cov(l_1,l_2) = \exp\left(-\frac{(Z(l_2)-Z(l_1))^2}{(\Delta Z(l_1))^2}\right) \qquad (12)$$

Z is domain-averaged height at a model level. For simplicity, a constant vertical localization radius $\Delta Z(l_1)$ is used as done by Li et al. (2012). The above two vertical localization specifications (Fig. 5) are examined and Eq. (12) is used in real radar data assimilation experiments.

In summary, we proposed two CVTs to incorporate climatological **B** with the NMC method in WRF-Var. The first one uses u, v, T, Ps and RH_s as control variables, which is named as CV7 in WRF-Var. The STD derived from the BE matrix using CV7 is shown in Fig. 4 (STD_CV7). It is shown that the use of u and v, which are WRF model prognostic variables, as control variables gives a good STD modeling. The other "alpha" control variable approach can provide geographic location dependent BE covariance modeling, which will be clearly shown in the single observation data assimilation experiments in next section.

4. Single observation data assimilation experiments

The **B** matrix weights the background state and spreads out observation information in horizontal and vertical directions in space. Increments from single observation experiments can be used to estimate BE variance and demonstrate how the BE covariance spreads the observation information spatially, which give a graphic representation BE structure function (Huang et al. 2009; Gustafsson et al. 2012).

4.1 Experimental design

To better understand the differences in BE representations in the three formulations, single observation data assimilation experiments are conducted to show BE covariance structures ($\mathbf{B}$). First the experiments (Table. 1) assimilating a single u observation are presented. Innovation is 1.0 m s^{-1}, and observation error is 1.0 m s^{-1}. Variance scaling factor for each control variable is 1.0 which is default value in WRF-Var. The single observation is set at location (271,212,25) in model grid.

The experiments CV5 and CV7 and NMC are designed to see impact of climatological background error formulations on analysis increments. The operational WRF 24 h and 12 h forecasts from BJ_RUC in June, July and Aug 2012 are used to calculate BE statistics for CV5 and CV7. The NMC forecast differences are used in experiment NMC. Compared to the above three experiments, three experiments ENS, 4DVAR-CV5 and 4DVAR-CV7 are carried out to show the capability of introducing flow-dependent error covariances by hybrid or 4D-Var methods. There are no vertical localization in experiments NMC or ENS. In experiment ENS, an ensemble of 20-member 12 h WRF forecasts at 0000 UTC 21 July 2012 is initiated from NCEP global ensemble forecast system. The assimilation time window is 30 minutes for the two 4DVAR experiments.

4.2 Analysis increments

In this subsection, first the horizontal and vertical structures of u analysis increments will be examined. Then multivariate features of analysis increments will be described. Vertical south-north section of u increments across the single observatoin location are shown in Fig. 6. The maximum values of the u analysis increments in CV5 and CV7 are 0.66 (Fig. 6a) and 0.89 m s^{-1} (Fig. 6b) respectively. The corresponding BE standard deviations are 1.39 and 2.84 m s^{-1}, respectively, indicating that CV5 underestimate wind variance than CV7. This result is consistent with the BE standard deviation estimations presented in section 2.2 (Fig. 4a). It is noted that the four experiments using alpha control variable produces almost the same value of the maximum analysis increments (Fig. 6c) to that of CV7. The two new formulations provide a consistent BE variance modeling.

In CV7, the lengthscale is directly calculated in u space so that it can be used as time- and domain averaged lengthscale reference for other experiments. Comparing the horizontal spread of observation information in Fig. 6a to Fig. 6b, it is found that CV5 overestimates horizontal lengthscale of auto-correlations. By reducing lengthscale by a half, the maximum value of u increment is increased to 0.88 m s^{-1} which is almost same to those in CV7 (Fig. 6b). The results confirm that reducing lengthscale increases the wind variance in Fig. 4. This may partially explain why radar data assimilation using

WRF-Var system with reduced lengthscales improved analyses and forecasts (Xiao and Sun 2007; Sugimoto 2009; Sun et al. 2012; Wang et al. 2013b).

For experiments NMC, ENS, 4DVAR-CV5 and 4DVAR-CV7, it is seen that the u analysis increment is unsymmetrical compared to the experiments CV5 and CV7. Small negative increments under about 10th model are derived by NMC differences, ensemble information or 4D-Var technique compared to the experiments CV5 and CV7. It is noted that for use of alpha control variables, results are sensitive to horizontal localization scale and vertical correlation matrix formulation.

The features of the multivariate analysis increments are analyzed. Figure 7 shows T increments by assimilating a single u observation. It is obvious that the amplitudes of increments are smaller in NMC than ENS. Vertical localization scales is required when alpha control variables are used.

4.3 Real radar data assimilation experiment

The impact of background formulations on the heaviest rainfall in 6 decades occurred in Beijing on 21 July 2012 are demonstrated in this section. The model domains are shown in Fig. 1. One-way nesting configuration is used in this study. The inner model domain has 550x424x38 grids with 3 km resolution. The WRF model is initiated at 1800 UTC 20 July 2012, and its 6 hour forecast at 0000 UTC 21 July 2012 is serviced as the background of data assimilation experiments. For model physics options please refer Chen et al's paper (2009). All the data assimilation experiments are over inner domain. Forecasts from three 3D-Var and hybrid data assimilation, and two 4D-Var data assimilation experiments are reported. For 4D-Var experiments, the time window is 18 minutes, and radar data are used every 6 minutes. Observed and forecasted 24h accumulate precipitation are shown in Fig. 8. It is seen that the 4D-Var experiment with the new background error formulation (CV7) shows that precipitation location forecast is slightly better than that with the standard CV5 or Hybrid background error formulation.

5. Summary and discussion

In this paper, the features of background error modeling via the NMC method are investigated in details for the WRF-Var system. The aim of this work is to further improve the performance of the WRF-Var system on mesoscale convective-scale data assimilation. The short-term regional 3km resolution forecasts in June, July and August 2012 from BJ-RUC are used to extract background error statistics. The two new CVTs are proposed and described to incorporate climatological **B** via the NMC method in WRF-Var. Up to the authors' knowledge, this is the first work using the two proposed CVTs to study climatological BE modeling in context of the WRF-Var system. The features of several CVTs are investigated in detail using three-month high-resolution operational forecasts.

The main work is summarized as follows.

- The BE variances of various variables vary from month to month, and the diurnal variation of BE in low-level atmosphere is also found.

- WRF-Var CV5 BE modeling underestimates wind error variance but overestimates wind error length scale.

- Two CVTs that are proposed to incorporate climatological **B** with the NMC method in WRF-Var are investigated. The first one uses u, v, T, Ps and RH_s as control variables. Another approach is to employ "alpha" control variables to incorporate location dependent error covariance. The detailed technical descriptions are provided for the two new formulations.

- A case study using WRF4D-Var with the new background error formulation (CV7) shows that precipitation location forecast is slightly better than that with the standard CV5 or Hybrid background error formulation.

The two proposed methods give good variance modeling in the NMC forecast differences. The use of the NMC forecast differences through alpha control variable has the benefits of incorporating the geographical dependent covariance information and producing multi-variate analysis. However, analysis increments are sensitive to horizontal and vertical localization radiuses using alpha control variable. Though the new proposed CV7 only produces the univariate analyses, the multi-variate analyses is achieved by use of hybrid or 4D-Var methods. The developments will benefit other components such as 4DVar, hybrid Var-ensemble data assimilation in WRF community data assimilation system (Barker et al. 2012; Huang et al. 2013).

In WRF-Var, the climatological statistical correlations between relative humidity and other control variables can be taken into account with CV option 6 (CV6) (Chen et al. 2013). These climatological statistical correlations between relative humidity and other variables can be achieved by the use of the alpha control variable as well. Moreover, CV7 can also be used to model BE covariance of day using an ensemble of short-term forecasts.

The monthly and diurnal variations of variances can be considered in climatological BE modeling in WRF-Var. In reality, the BE covariances may be substantially flow dependent. The current BE statistics using the NMC method may not be optimal to provide the BE covariance of the day for mesoscale and convective scale data assimilation. A super ensemble including the NMC forecast differences and short-term ensemble forecasts could be used to blend climatological error covariances and flow-dependent error covariances of the day in a hybrid system. Preliminary results from a case study using WRF4D-Var with the new background error formulation (CV7) showed that precipitation location forecast is slightly better than that with the standard CV5 or Hybrid background error formulation. The detailed investigations on reasons leading to good forecasts are underway.

References

Andersson, E., M. Fisher, R. Munro and A. McNally, 2000: Diagnosis of background errors for radiances and other observable quantities in a variational data assimilation scheme, and the explanation of a case of poor convergence. *Quart. J. Roy. Meteor. Soc.* **126**, 1455–1472.

Bannister, R. N., 2008a: A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances. *Quart. J. Roy. Meteor. Soc.*, **134,** 1951–1970.

Bannister, R. N., 2008b: A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics. *Quart. J. Roy. Meteor. Soc.*, **134,** 1971–1996.

Barker, D. M., W. Huang, Y.-R. Guo, A. J. Bourgeois, and Q. N. Xiao, 2004: A three-dimensional variational data assimilation system for MM5: Implementation and initial results. *Mon. Wea. Rev.*, **132**, 897–914.

Barker, D. M., 2005: Southern High-Latitude Ensemble Data Assimilation in the Antarctic Mesoscale Prediction System. *Mon. Wea. Rev.*, **133**, 3431–3449.

Barker, D. M., and Coauthors, 2012: The Weather Research and Forecasting (WRF) Model's Community Variational/Ensemble Data Assimilation System: WRFDA. *Bull. Amer. Meteor. Soc.*, **93**, 831–843.

Boilley, Alexandre, and Jean-François, Mahfouf, 2012: Assimilation of low-level wind in a high-resolution mesoscale model using the back and forth nudging algorithm. *Tellus A* 2012, **64**, 18697, http://dx.doi.org/10.3402/tellusa.v64i0.18697

Chen, M., S. Fan, J. Zhong, X.-Y. Huang, Y.-R. Guo, W. Wang, Y. Wang, and B. Kuo, 2009: A WRF-based rapid updating cycling forecast system of BMB and its performance during the summer and Olympic Games 2008. 10th WRF Users' Workshop, June 23-26, 2009, Boulder, CO. [Available online at http://www.mmm.ucar.edu/wrf/users/workshops/WS2009/abstracts/P3B-37.pdf.]

Chen, Y., S.R.H. Rizvi, X.-Y. Huang, J. Min, and X. Zhang, 2013: Balance characteristics of multivariate background error covariance and their impact on analyses and forecasts in Tropics and Arctic. *Meteor. Atmos. Phys.*, 121, 79-98.

Courtier P., E. Andersson, W. Heckley, J. Pailleux, D. Vasiljevic, A. Hollingsworth, M. Fisher, and F. Rabier, 1998: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. *Q. J. R. Meteorol. Soc.* **124**: 1783–1807.

Daley, R. 1991: *Atmospheric data analysis*. Cambridge University Press, pp. 460.

Fisher M., 2003: Background error covariance modelling. P. 45 – 64 in ECMWF Seminar on Recent developments in data assimilation for atmosphere and ocean, 8 – 12 September 2003. ECMWF: Reading UK.

Fischer C., T. Montmerle, L. Berre, L. Auger, S.E. Ştefãnescu, 2005: An overview of the variational assimilation in the ALADIN/France numerical weather prediction system. *Q. J. R. Meteorol. Soc.* **131**: 3477–3492.

Gao, J., M. Xue, A. Shapiro, and K. K. Droegemeier, 1999: A variational analysis for the retrieval of three-dimensional mesoscale wind fields from two Doppler radars, *Mon. Wea. Rev.,* **127**, 2128-2142.

Houtekamer PL, L. Lefaivre, J. Derome, H. Ritchie, and H.L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Weather Rev.* **124**: 1225–1242.

Gustafsson, N., X.-Y. Huang, X. Yang, K.S. Mongensen, M. Lindskog, O. Vignes, T. Wilhelmsson, and S. Thorsteinsson, 2012: Four-dimensional variational data assimilation for a limited area model. *Tellus A*, **64**, 14985, DOI: 10.3402/tellusa.v64i0.14985.

Hsiao, L-F, D-S Chen, Y-H Kuo, Y-R Guo, T-C Yeh, J-S Hong, C-T Fong, and C-S Lee, 2012: Application of WRF 3DVAR to Operational Typhoon Prediction in Taiwan: Impact of Outer Loop and Partial Cycling Approaches. *Wea. Forecasting*, **27**, 1249–1263.

Huang, X.-Y., and coauthors, 2009: Four-Dimensional Variational Data Assimilation for WRF: Formulation and Preliminary Results. *Mon. Wea. Rev.*, **137**, 299-314.

Huang, X.-Y., and coauthors, 2013: The 2013 WRFDA overview. 14th WRF Users' Workshop, Boulder, Colorado, 24-28 June 2013. Extended Abstract 1.2. [http://www.mmm.ucar.edu/wrf/users/workshops/WS2013/WorkshopPapers.php]

Ingleby, N., 2001: The statistical structure of forecast errors and its representation in The Met. Office global 3-D variational data assimilation scheme. *Quart. J. Roy. Meteor. Soc.*, **127**, 209–231.

Kawabata, T., T. Kuroda, H. Seko, and K. Saito, 2011: A cloud-resolving 4DVAR assimilation experiment for a local heavy rainfall event in the Tokyo metropolitan area. *Mon. Wea. Rev.*, **139**, 1911–1931.

Li, Y., X. Wang, and M. Xue, 2012: Assimilation of Radar Radial Velocity Data with the WRF Hybrid Ensemble–3DVAR System for the Prediction of Hurricane Ike (2008). *Mon. Wea. Rev.*, **140**, 3507–3524.

Parrish, D., and J. Derber, 1992: The National Meteorological Center's spectral statistical interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763.

Rawlins, F., and coauthors, 2007: The Met Office global four-dimensional variational data assimilation scheme. *Quart. J. Roy. Meteor. Soc.*, **133**, 347–362.

Skamarock, W. C., and Coauthors, 2008: A description of the advanced research WRF version 3. NCAR Tech. Note TN-475+STR, 113 pp.

Sugimoto, S., N. A. Crook, J. Sun, Q. Xiao, and D. M. Barker, 2009: An Examination of WRF 3DVAR Radar Data Assimilation on Its Capability in Retrieving Unobserved Variables and Forecasting Precipitation through Observing System Simulation Experiments. *Mon. Wea. Rev.*, **137**, 4011–4029.

Sun, J., and N. A. Crook, 1997: Dynamical and Microphysical Retrieval from Doppler Radar Observations Using a Cloud Model and Its Adjoint. Part I: Model Development and Simulated Data Experiments. *J. Atmos. Sci.*, **54**, 1642-1661.

Sun, J., S. B. Trier, Q. Xiao, M. L. Weisman, H. Wang, Z. Ying, M. Xu, and Y. Zhang, 2012: Sensitivity of 0–12-h Warm-Season Precipitation Forecasts over the Central United States to Model Initialization. *Wea. Forecasting*, **27**, 832–855.

Xiao, Q., and J. Sun, 2007: Multiple-Radar Data Assimilation and Short-Range Quantitative Precipitation Forecasting of a Squall Line Observed during IHOP_2002. *Mon. Wea. Rev.*, **135**, 3381–3404.

Wang, H., J. Sun, S. Fan, and X.-Y. Huang, 2013a: Indirect Assimilation of Radar Reflectivity with WRF 3D-Var and Its Impact on Prediction of Four Summertime Convective Events. *J. Appl. Meteor. Climatol.*, **52**, 889–902.

Wang, H., J. Sun, X. Zhang, X.-Y. Huang, and T. Auligné, 2013b: Radar Data Assimilation with WRF 4D-Var: Part I. System Development and Preliminary Testing. *Mon. Wea. Rev*. **141**, 2224–2244.

Wang, X., D. M. Barker, C. Snyder, and T. M. Hamill, 2008: A Hybrid ETKF–3DVAR Data Assimilation Scheme for the WRF Model. Part I: Observing System Simulation Experiment. *Mon. Wea. Rev.*, **136**, 5116–5131.

Xie, Y., C. Lu, and G. L. Browning, 2002: Impact of formulation of cost function and constraints on three-dimensional variational data assimilation. *Mon. Wea. Rev.*, **130,** 2433–2447.

Xie Y. and A. E. MacDonals, 2011: Selection of momentum variables for a three-dimensional variational analysis. *Pure and Applied Geophysics*, 169, 335–351.

Zou, X., Y.-H. Kuo, and Y.-R. Guo, 1995: Assimilation of atmospheric radio refractivity using a nonhydrostatic adjoint model. *Mon. Wea. Rev.*, **123**, 2229-2249.

Zupanski M, D. Zupanski, T. Vukicevic, K. Eis, and T. Vonder Haar. 2005: CIRA/CSU four dimensional variational data assimilation system. *Mon. Weather Rev.* **133**: 829–843.

Table. 1 List of single observation data assimilation experiments

Experiment name	CV	Descriptions
CV5	5	3DVAR
CV7	7	3DVAR
4DVAR-CV5	5	4DVAR
4DVAR-CV7	7	4DVAR
NMC	alpha	3DVAR; NMC differences; Length scale 60km
ENS	alpha	3DVAR; Ensemble differences; Length scale 60km

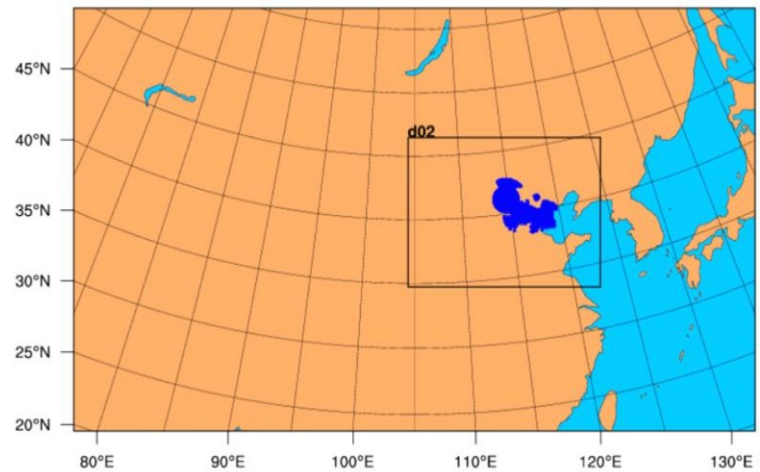

Fig. 1. BJ_RUC model domains and an example of radar observation distribution at 00Z July 21 2012.

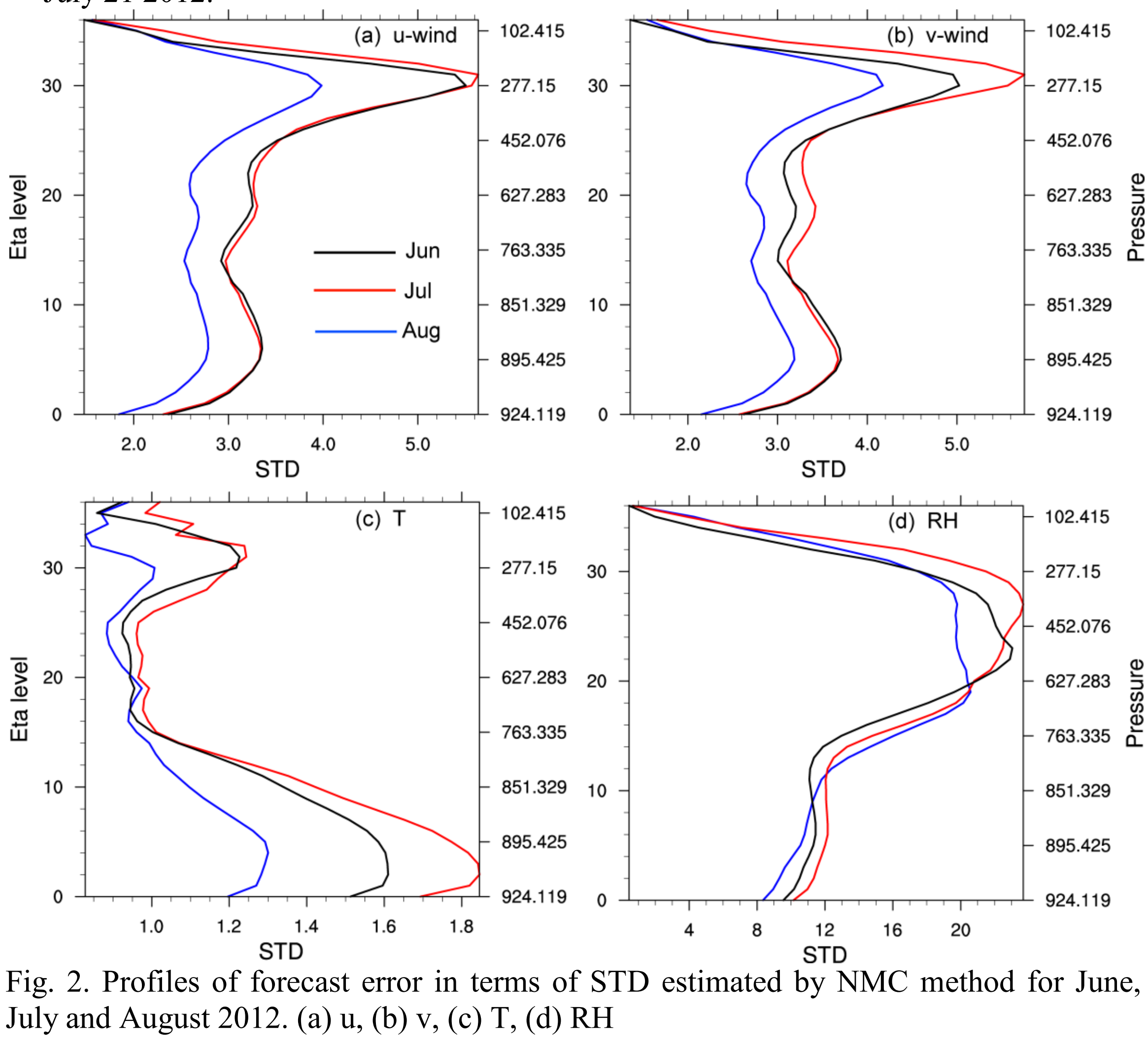

Fig. 2. Profiles of forecast error in terms of STD estimated by NMC method for June, July and August 2012. (a) u, (b) v, (c) T, (d) RH

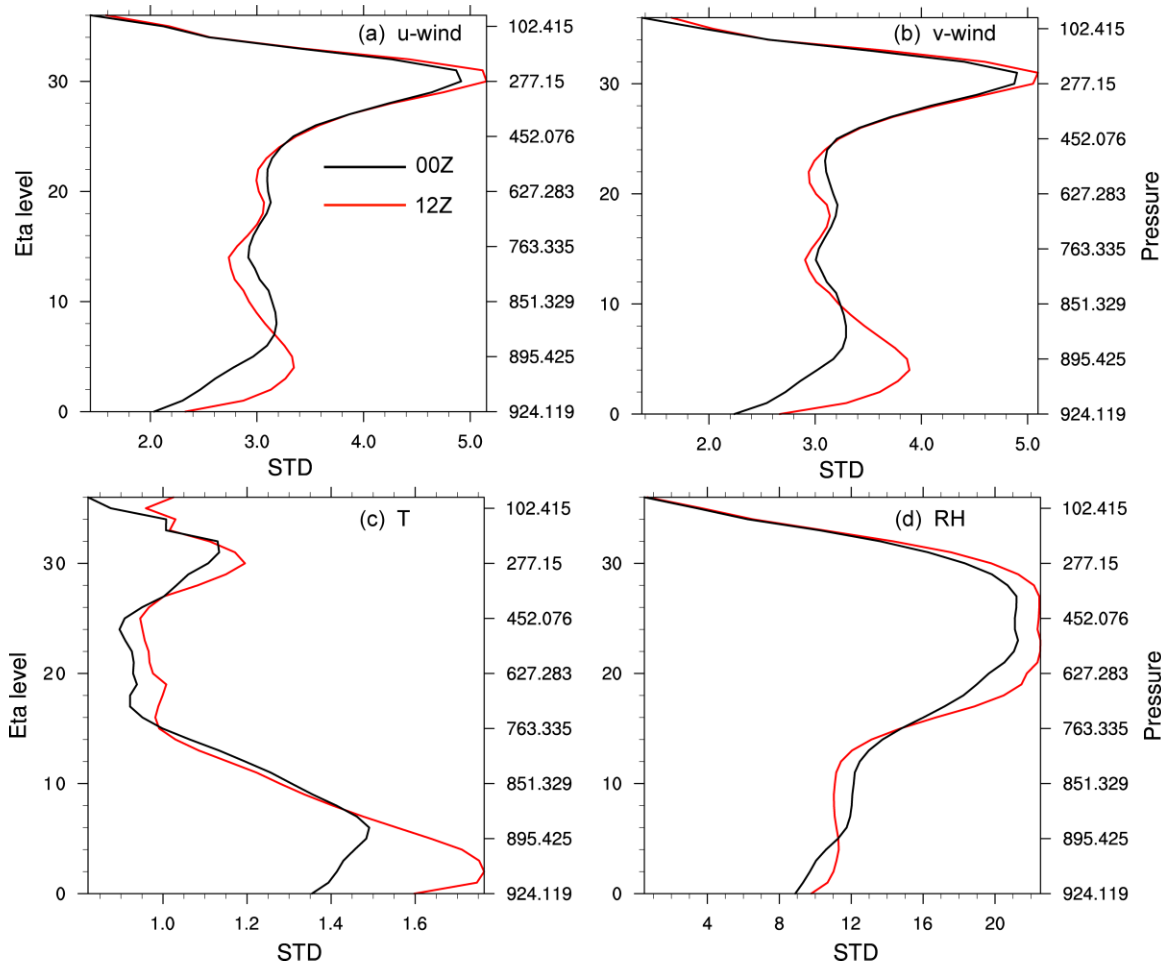

Fig. 3. Profiles of forecast error in terms of STD estimated by NMC method at 00 UTC (black curve) and 12 UTC (red curve) for (a) u, (b) v, (c) T, (d) RH

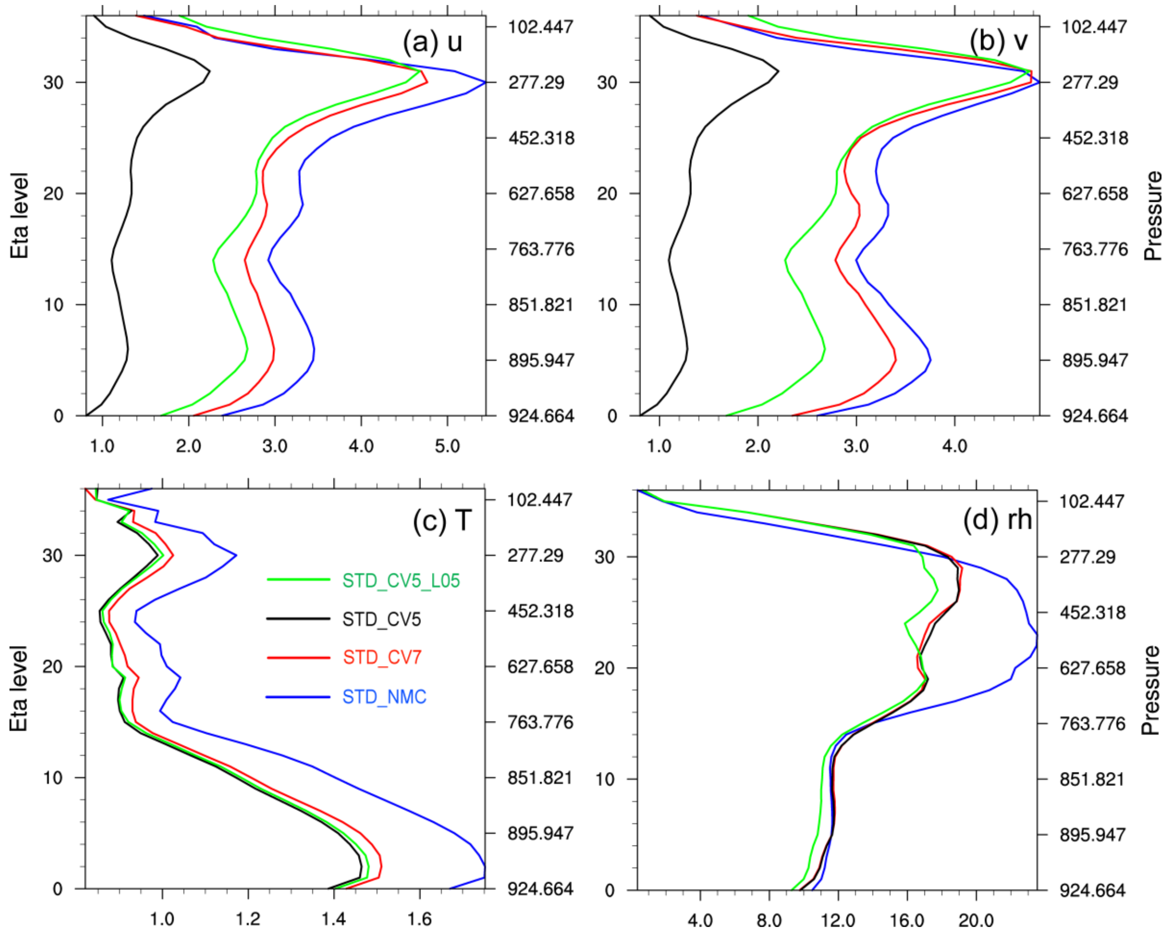

Fig. 4. The vertical profiles of STD_NMC, STD_CV5, STD_CV5_L05 and STD_CV7.

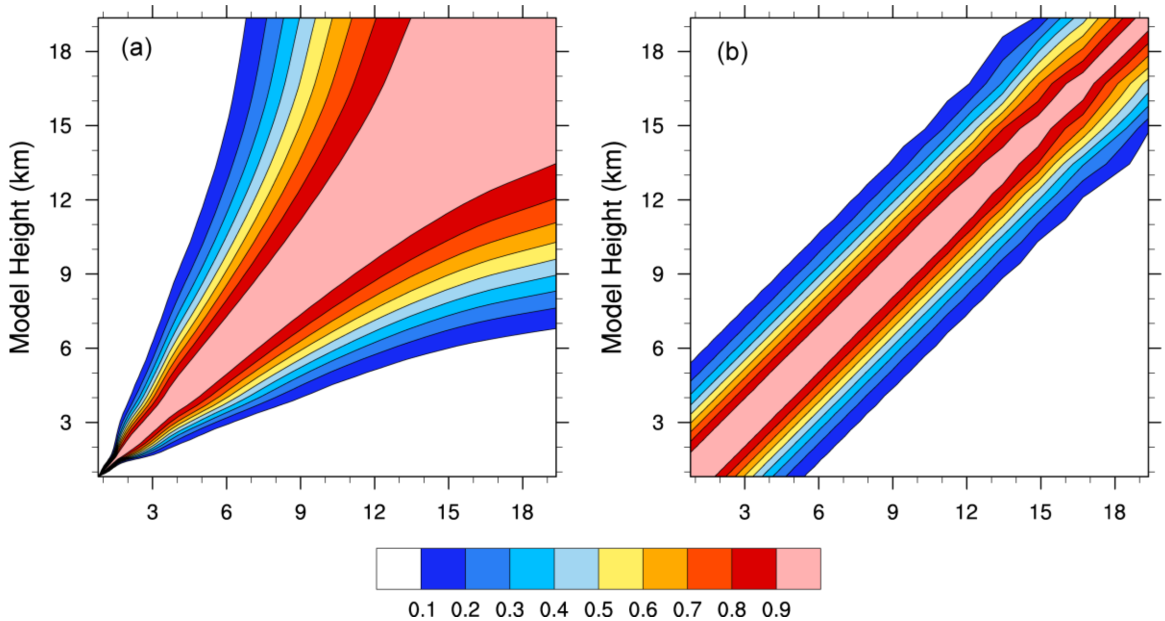

Fig. 5 The correlation matrix for alpha control variable localization for (a) Eq.11, and (b) Eq.12.

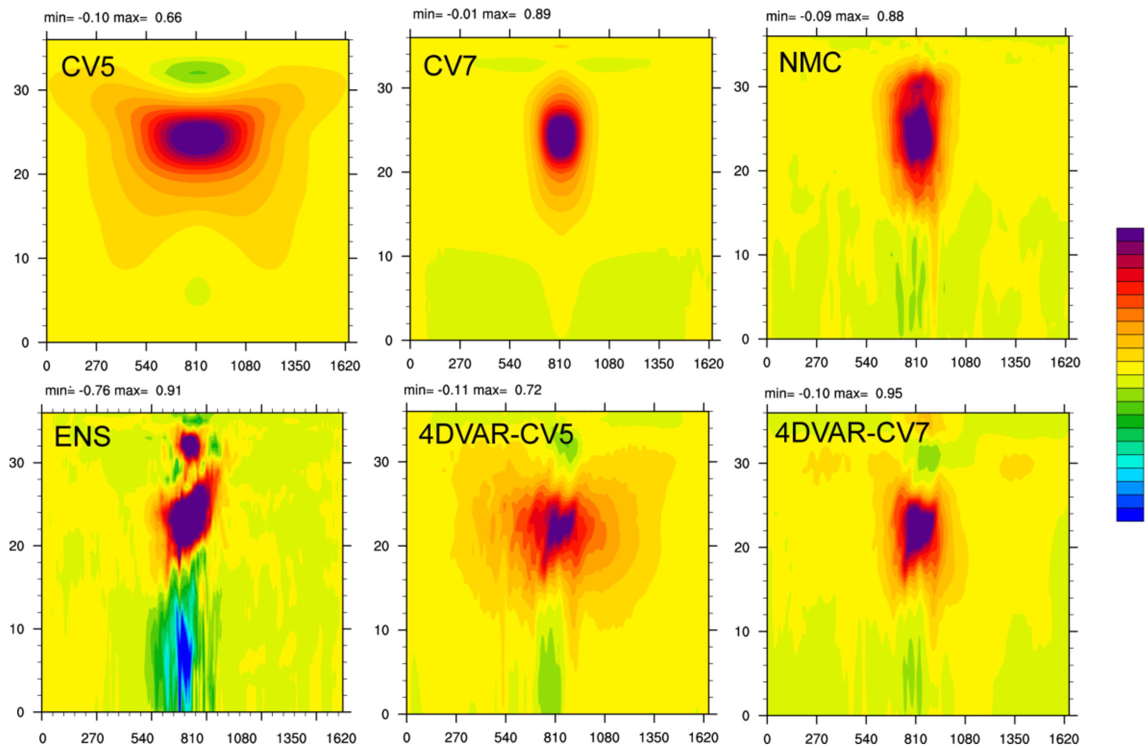

Fig. 6. Vertical west-east section of u increments for single u observation experiments. (a) CV5, (b) CV7, (c) NMC, (d) ENS, (e) 4DVAR-CV5, and (f) 4DVAR-CV7.

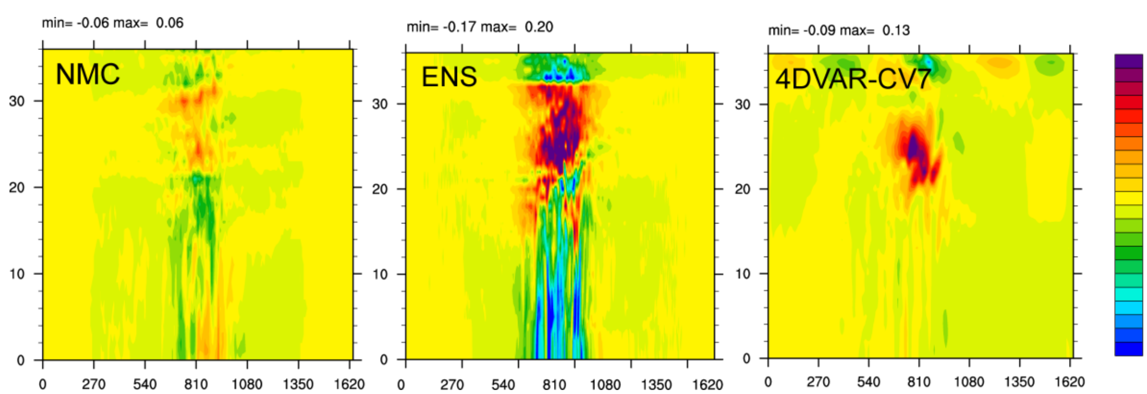

Fig. 7. Vertical west-east section of T increments by assimilating single u observation (a) NMC, (b) ENS, (c) 4DVAR-CV7.

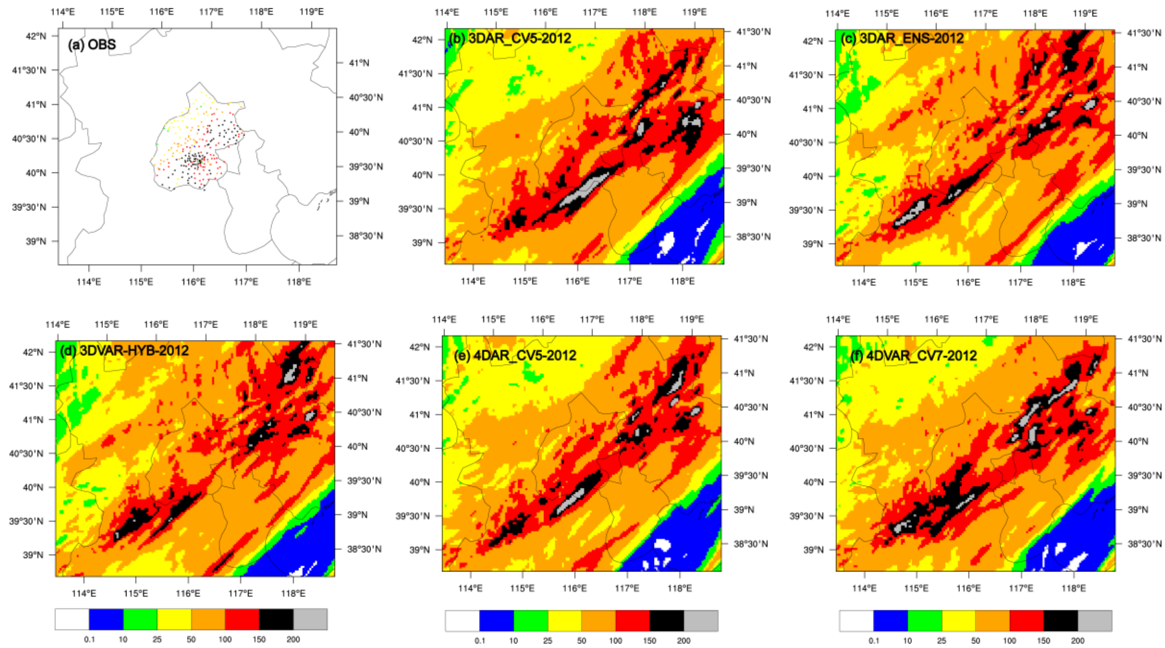

Fig. 8. 24h accumulated rainfall from 00Z 21 July 2012 to 00Z 22 July.