ORIGINAL PAPER

Balance characteristics of multivariate background error covariances and their impact on analyses and forecasts in tropical and Arctic regions

Yaodeng Chen · Syed R. H. Rizvi · Xiang-Yu Huang · Jinzhong Min · Xin Zhang

Abstract For variational data assimilation, the background error covariance matrix plays a crucial role because it is strongly linked with the local meteorological features, and is especially dominated by error correlations between different analysis variables. Multivariate background error (MBE) statistics have been generated for two regions, namely the Tropics (covering Indonesia and its neighborhood) and the Arctic (covering high latitudes). Detailed investigation has been carried out for these MBE statistics to understand the physical processes leading to the balance (defined by the forecasts error correlations) characteristics between mass and wind fields for the low and high latitudes represented by these two regions. It is found that in tropical regions, the unbalanced (full balanced) part of the velocity potential (divergent part of wind) contributes more to the balanced part of the temperature, relative humidity, and surface pressure fields as compared with the stream function (rotational part of wind). However, the exact opposite happens in the Arctic. For both regions, the unbalanced part of the temperature field is the main contributor to the balanced part of the relative humidity field. Results of single observation tests and six-hourly data assimilation cycling experiments are consistent with the respective balance part contributions of different fields in the two regions. This study provides an understanding of the contrasting dynamical balance relationship that exists between the mass and wind fields in high- and low-latitude regions. The study also examines the impact of MBE on Weather Research and Forecasting model forecasts for the two regions.

1 Introduction

Numerical Weather Prediction (NWP) accuracy relies heavily on the quality of the initial atmospheric state (Caron and Fillion 2010). Data Assimilation is used to provide the initial atmospheric state for numerical weather prediction at many NWP centers (Derber and Bouttier 1999; Lorenc 2003 ; Barker et al. 2004; Sadiki and Fischer 2005; Rawlins et al. 2007). Use of a background state is very important, because in the absence of observations it provides a realistic reference atmospheric state (Bannister 2008a). Generally, the background state is a short-range (typically 6-h) forecast from a NWP model. Since the spatial and multivariate structure of the analysis increments is filtered by the background error (BE) statistics, the BE matrix plays a crucial role in meteorological data analysis (Derber and Bouttier 1999).

Derivation of a reasonable and accurate representation of background error covariances is a major challenge for variational data assimilation schemes (VAR) (Brousseau et al. 2011). In the absence of knowing the true atmospheric state, some basic assumptions are made to estimate the background error statistics. Some NWP centers rely on the NMC method (Parrish and Derber 1992). In this method, the forecast error is approximated with the difference between two NWP forecasts (typically, 12 and 24 h) valid for the same time, and these forecast error

Y. Chen (✉) · J. Min
Key Laboratory of Meteorological Disaster of Ministry of Education, University of Information Science and Technology, Nanjing 210044, China
e-mail: keyu@nuist.edu.cn

S. R. H. Rizvi · X.-Y. Huang · X. Zhang
National Center for Atmospheric Research, Boulder, CO 80302, USA

statistics are averaged over a certain time period (typically, 1 month). An alternative approach for estimating BE statistics is the ensemble method (Fisher 2003; Pereira and Berre 2006). In this method the forecast error is estimated with "ensemble minus ensemble mean". Technical details for using these two approaches for computing the BE statistics are described by Berre et al. (2006). Most often, the size of the BE covariance matrix is very large (typically, $10^6 \times 10^6$) and thus it cannot be stored in any computer memory. It is due to this reason that in VAR the analysis control variables are carefully designed, and the background error covariances are modeled using a suitable sequence of analysis control variable transforms (Derber and Bouttier 1999). This approach of applying background error in analysis control variable space has several advantages, such as reducing the size of background error matrix by making it diagonal, enhancing the physical balance constraints, and improving the pre-conditioning of the minimization (Bannister 2008b; Courtier and Talagrand 1990).

Most often in NWP, the balance between different atmospheric state variables is generally defined by "geostrophic" or "hydrostatic" types of diagnostic relationships. However, in VAR, the "balance relationship" across different analysis variables is generally defined using some regression coefficients between these variables. The desired regression coefficients are estimated with forecast errors using a standard regression technique. Thus the dynamical balance imposed by the NWP model on the forecast gets reflected in the corresponding BE statistics. Using regression coefficients, once the "balanced" part of any analysis variable is known, the corresponding "unbalanced" part is obtained by subtracting the balanced part from the respective full field. Thus in VAR, some variables are analyzed in full, while for other variables only the corresponding unbalanced parts are analyzed. Further details about how the balanced and unbalanced parts are computed will be discussed in a later section while describing the analysis control variables.

For any NWP model, the BE statistics depend on many factors and may vary from region to region. Some of the regional dependencies of BE statistics are due to observation density, the type of observations available, local meteorological features of the region, and the balance between different model state variables, etc. For example, in mid- and high-latitudes, the basic balance between mass and wind fields is dominated by the geostrophic relationship, whereas in the tropics, due to the weaker Coriolis force, the geostrophic effect is very small. Many authors have investigated different aspects of background errors in different regions, such as Sadiki et al. (2000), Montmerle et al. (2006), Michel and Auligne (2010). For regional analysis, it is very important to use the regional BE as it

reflects the local meteorological characteristics (Storto and Randriamampianina 2010).

In VAR, the balance relationship spreads information between different analysis variables. The spreading in space is handled suitably by the application of horizontal and vertical correlations. From studies such as, Berre (2000), Žagar et al. (2005), Berre et al. (2006), Caron and Fillion (2010), one can find the role of balance constraints across different analysis variables. However, in different latitude regions, little is known about the contribution of different analysis variables towards the balanced part of other variables. The impact of multivariate background error covariances on data assimilation and NWP model forecasts is also lesser known in different latitude domains. These two issues are the main focus of the present study.

This paper is organized as follows: Sect. 2 gives a brief description of the variational data assimilation system used in this study. Formulation of analysis control variables is described in Sect. 3. Details of the various experiments undertaken and some of the characteristics of the corresponding MBE (such as the balance relationship, eigenvalues and length-scale) used in these experiments are discussed in Sect. 4. The response of assimilating a single observation is discussed in Sect. 5. Results with month-long, six-hourly data assimilation cycling runs are discussed in Sect. 6. Conclusions are drawn in Sect. 7.

2 Variational data assimilation

In general, variational data assimilation schemes are designed to provide an analysis that minimizes an objective cost function (J), defined as

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) \\ + \frac{1}{2}(\mathbf{y}^0 - H(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y}^0 - H(\mathbf{x})) \qquad (1)$$

Here, $\mathbf{x}$ is the vector of the NWP model state variable (e.g. wind components, temperature, humidity, and surface pressure), $\mathbf{x}_b$ is the background vector, $\mathbf{y}^0$ is the observation vector, H is the nonlinear observation operator mapping model space to the observation space, and $\mathbf{B}$ and $\mathbf{R}$ are the background and observation error covariance matrices, respectively.

Most operational NWP centers use an incremental approach (Courtier et al. 1994) to solve the variational problem of minimizing the cost function (J). In this approach, the observations, the error covariances of observations and background, and the physical laws governing the NWP model state are all combined to produce

the analysis increment $\delta\mathbf{x} = (\mathbf{x} - \mathbf{x}_b)$. Since $\mathbf{B}$ is symmetric positive definite, it may be partitioned in terms of a lower triangular matrix $\mathbf{U}$, as $\mathbf{B} = \mathbf{U}\mathbf{U}^T$. Here, $\mathbf{U}^T$ is the transpose of $\mathbf{U}$. Following Derber and Bouttier (1999), let us define a set of analysis control variables ($\mathbf{v}$), as $\mathbf{U}\mathbf{v} = \delta\mathbf{x}$. Thus, in terms of analysis control variables ($\mathbf{v}$), the objective cost function may be written as

$$J(\mathbf{v}) = \frac{1}{2}\mathbf{v}^T\mathbf{v} + \frac{1}{2}(\mathbf{d} - \mathbf{HUv})^T\mathbf{R}^{-1}(\mathbf{d} - \mathbf{HUv}) \tag{2}$$

Here, $\mathbf{d} = \mathbf{y}^0 - H(\mathbf{x}_b)$ is the innovation vector, representing the departure between observation and the background, and $\mathbf{H}$ is the linearized version of the non-linear observation operator, H. Typically, the analysis control variable transform ($\mathbf{U}$) consists of a sequence of three transforms, the horizontal ($\mathbf{U}_h$), vertical ($\mathbf{U}_v$) and physical ($\mathbf{U}_p$), defined as

$$\mathbf{U} = \mathbf{U_p U_v U_h} \tag{3}$$

Since, $\mathbf{B}$ is represented as $\mathbf{U}\mathbf{U}^T$, the background error covariances may be specified in analysis control variable space (e.g. stream function, unbalanced part of velocity potential, unbalanced part of temperature, unbalanced part of surface pressure and relative humidity) via a sequence of control variable transforms defined in terms of $\mathbf{U}$ and $\mathbf{U}^T$, as $\mathbf{B} = \mathbf{U_p U_v U_h U_h^T U_v^T U_p^T}$.

The WRF data assimilation (WRFDA) system used in this study is a variational data assimilation system formulated in grid-point space (Barker et al. 2012). In this system, $\mathbf{U}_h$ is a recursive filter transform to impose the horizontal correlations, $\mathbf{U}_v$ is the application of vertical correlations through empirical orthogonal functions (EOF) of analysis control variables, and $\mathbf{U}_p$ changes the analysis control variables to model state variables using the statistical balance relationship. The choice of analysis control variables for the WRFDA system and the basic input for these three transforms will be discussed in the next section. In the WRFDA system, the input background field ($\mathbf{x}_b$) is the short-range (typically, 1–6 h) forecast from the WRF model (Skamarock et al. 2008). Further technical details about the WRFDA system may be found in Barker et al. (2004) and Huang et al. (2009).

The WRFDA system can ingest a wide variety of observation types, including conventional observations (surface, rawinsonde, dropsonde, aircraft, wind profiler and atmospheric motion vectors) and non-conventional data (radar reflectivity and radial velocity, GPS occultation and radiance data observed in different channels from a variety of satellite platforms). However, in the present study, only conventional observations are used. All the input observations are pre-processed using the WRFDA OBSPROC utility (Barker et al. 2003).

3 Formulation of analysis control variables

For any data assimilation system, its choice of analysis control variables makes it unique. The choice of control variables mainly depends on the type of analysis variables used, the definition of balance relationships across other analysis variables, and the application of background error covariances. The choice of balance relationships is important because they decide whether a particular variable will be analyzed as univariate or multivariate. The choice of analysis control variable also depends on how $\mathbf{B}$ is represented or applied in the respective variational data assimilation. Thus, before discussing the choice of analysis control variables in the current WRFDA, it is important to understand how background error statistics are computed for their application in the WRFDA system.

For computing BE for WRFDA, irrespective of the method used, the forecast errors are accumulated typically for a period of 1 month at least. Further, the computation of BE statistics proceeds sequentially in the following five steps:

a. Regression coefficients between different analysis variables are computed. These regression coefficients form the basis for the $\mathbf{U}_p$ transform.
b. Using regression coefficients in step a, compute the balanced part for all the desired analysis variables (not aimed to be analyzed as full). Remove the balanced part from the corresponding full variable to get the unbalanced part.
c. Compute the vertical error covariance matrix for all the 3D-variables (full field or the unbalanced part after step-b). Eigen-decomposition is done for the vertical error covariance matrix to get the eigenvector and eigenvalues. These eigenvalues and eigenvectors (EOF) form the basis for the $\mathbf{U}_v$ transform.
d. Each 3D analysis variable (after step b) is projected in EOF-space with the corresponding EOFs (computed in step c).
e. The desired length scale for the $\mathbf{U}_h$ transform is estimated for each of the analysis control variables (for 2D-variable this is the output of step b; otherwise it is the output of step d) using a Gaussian fit method as described in Barker et al. (2003). Here, it may be noted that for all the analysis control variables, the length scale does not vary horizontally, and for 3D-variables it is eigen-mode dependent.

The WRFDA system analyzes the stream function, velocity potential, temperature, surface pressure, and relative humidity. The current WRFDA system defines balance relationships between the stream function with velocity potential, temperature, and surface pressure. Thus, the analysis control variables for stream function (ψ) and

relative humidity (*rh*) are for the respective full field. For velocity potential, temperature, and surface pressure, the analysis control variables are the corresponding unbalanced part, defined as follows:

$$\chi_u(i,j,k) = \chi(i,j,k) - \alpha_{\psi\chi}(i,j,k)\psi(i,j,k) \tag{4}$$

$$T_u(i,j,k) = T(i,j,k) - \sum_{l=1}^{N_k} \alpha_{\psi T}(i,j,k,l)\psi(i,j,l) \tag{5}$$

$$P_u(i,j) = ps(i,j) - \sum_{l=1}^{N_k} \alpha_{\psi ps}(i,j,l)\psi(i,j,l) \tag{6}$$

Here, the indices *i* and *j* run over the horizontal dimensions of the geographical domain, *k* and *l* run over the N_k vertical sigma levels, and α represents the regression coefficients between the variables specified with the respective subscripts. In Eqs. (4) through (6), the second term on the right-hand side defines the balanced part of the velocity potential, temperature, and surface pressure, respectively.

In the WRFDA system, each of the 3D analysis control variables is represented in EOF space with their corresponding EOFs. Thus in the current WRFDA system, the analysis control variables are the stream function (ψ), the unbalanced part of the velocity potential (χ_u), the unbalanced part of temperature (T_u), the unbalanced part of surface pressure (ps_u), and the relative humidity (*rh*). An important point to note here is that the WRFDA analysis procedure also follows the five steps (a through e) mentioned above for the computation of BE statistics, but in reverse order. It is so because the analysis procedure starts with full/unbalanced variables in EOF space and delivers the analysis for the model state variables, whereas the BE computation procedure starts with the model state variable and produces the full/unbalanced variables in EOF space. The different orders in which the three control variable transforms ($\mathbf{U}_p$, $\mathbf{U}_v$, and $\mathbf{U}_h$) are carried out each have their own advantages and disadvantages. With the balance transform ($\mathbf{U}_p$), it is possible to apply a different degree of implicit geostrophic balance depending on, for example, the latitude. However, in this study the background errors used are latitude-independent. The vertical transform ($\mathbf{U}_v$) filters out the vertical correlation that is outside the space generated by the background vertical error covariances. At the same time, it has the advantage of saving memory/computation because most of the variance (99 %) may be explained by only the first few leading eigenvectors (EOFs) and so it is not necessary to include all the EOFs in the analysis procedure. The horizontal transform ($\mathbf{U}_h$) is applied using recursive filter in EOF's mode with uniform (not varying with latitude) length-scale and so it inhibits the possibilities of unbalanced part being scale-dependent, which is important for meso-scale. The WRFDA analysis procedure makes use of the BE statistics that are already computed offline for the regression coefficients, eigenvalues, eigenvectors, and length-scales.

It may be seen that in the current WRFDA system, temperature and surface pressure observations will influence the velocity potential via its balanced part contributed by the stream function. However, since there is no correlation used between the velocity potential and temperature or the velocity potential and surface pressure, neither temperature nor surface pressure observations can directly influence the divergent part of the wind. Due to similar reasons, the moisture observations will not have any impact on other variables like wind, temperature, and surface pressure. To overcome these limitations, six additional regression coefficients, namely $\alpha_{\chi_u T}$, $\alpha_{\chi_u ps}$, $\alpha_{\psi rh}$, $\alpha_{\chi_u rh}$, $\alpha_{T_u rh}$ and $\alpha_{ps_u rh}$ are introduced in defining the balance relationship across different analysis variables. Thus, the new set of equations defining the balance relationships across other analysis variables, parallel to Eqs. (4) through (6), is as follows:

$$\boldsymbol{\chi}_u(i,j,k) = \boldsymbol{\chi}(i,j,k) - \alpha_{\psi\chi}(i,j,k)\boldsymbol{\psi}(i,j,k) \tag{7}$$

$$T_u(i,j,k) = T(i,j,k) - \sum_{l=1}^{N_k} \alpha_{\psi T}(i,j,k,l)\psi(i,j,l)$$
$$- \sum_{l=1}^{N_k} \alpha_{\chi_u T}(i,j,k,l)\chi_u(i,j,l) \tag{8}$$

$$ps_u(i,j) = ps(i,j) - \sum_{l=1}^{N_k} \alpha_{\psi ps}(i,j,l)\psi(i,j,l)$$
$$- \sum_{l=1}^{N_k} \alpha_{\chi_u ps}(i,j,l)\chi_u(i,j,l) \tag{9}$$

$$rh_u(i,j,k) = rh(i,j,k) - \sum_{l=1}^{N_k} \alpha_{\psi rh}(i,j,k,l)\psi(i,j,l)$$
$$- \sum_{l=1}^{N_k} \alpha_{\chi_u rh}(i,j,k,l)\chi_u(i,j,l)$$
$$- \sum_{l=1}^{N_k} \alpha_{T_u rh}(i,j,k,l)T_u(i,j,l) - \alpha_{ps_u rh}(i,j,k)ps_u(i,j) \tag{10}$$

In Eqs. (7) through (10), the balanced parts for different variables are represented collectively with the 2nd term on the right-hand side. Thus, with the new setup of the analysis control variables, the definition of the balanced part of temperature and surface pressure has changed, and the relative humidity also has a balanced and unbalanced part. With the inclusion of correlation between temperature and surface pressure with unbalanced velocity potential, one can see that in the new analysis procedure, the temperature and surface pressure observations will also influence the divergent part of the wind. Additionally, the inclusion of

moisture correlations with all other analysis variables will also lead to multivariate moisture analysis. It is via these additional correlations that the moisture will have impact on other analysis variables like wind, temperature, and surface pressure, and vice versa.

The new set of analysis control variables, defined by (7)–(10), has already been successfully implemented within the framework of WRFDA system (Krysta et al. 2009). This gives an opportunity to study the impact of MBE on WRFDA analyses and the subsequent WRF model forecasts.

4 Formulation of the MBE statistics

4.1 Domains and experiments

Two domains, a tropical region (covering Indonesia and its neighborhood, representing a part of tropics) and an Arctic region (representing high latitudes), are configured at a horizontal resolution of 30 km with 51 vertical sigma levels (Table 1) and with the model top at 10 hPa. Exact geographical locations for these two domains are shown in Fig. 1. For both these domains, 12- and 24-h forecasts are generated for a period of 1 month (00 UTC of 15 July to 18 UTC of 15 August 2009) using the WRF model with all its default options. The initial and boundary conditions for WRF are prepared using the NCEP GFS analysis at 1° (~100 km) horizontal resolution. Forecast differences (perturbations) between 12- and 24-h forecasts valid at the same time are used as input to the NMC method for generating the MBE statistics. While creating the perturbations, forecasts both from 00 to 12 UTC initial conditions are used to avoid any systematic errors due to diurnal variation in the WRF forecasts. Thus, for each region, the NMC method is used with 62 forecast error samples to estimate the MBE statistics for each of the seven experiments. Here it may be noted that for any statistical estimate, large sample size is always good. In keeping with the practices followed at various operational NWP centers for estimating the background error statistics using NMC method, the forecast error sample size of 62 used in this study is sufficient. As a reference, NCEP used 30 samples for its SSI scheme (Parrish and Derber 1992) and 48 samples for its GSI scheme (Wu and Purser 2002), EC-MWF used 45 samples (Derber and Bouttier 1999). The MBE statistics corresponding to each of the seven experiments have been described below.
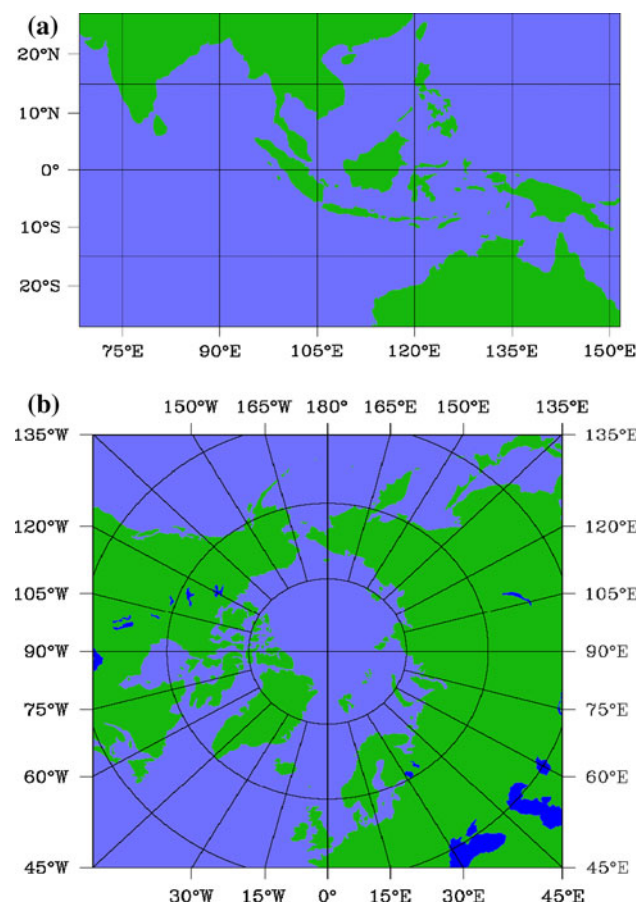
Seven experiments have been designed to illustrate the impact of MBE with the inclusion of different regression coefficients. The first experiment, Exp-1 (control run), is performed with the original formulation, which takes into account only three regression coefficients ($\alpha_{\psi\chi}$, $\alpha_{\psi T}$, and $\alpha_{\psi ps}$). In the subsequent experiments 2 through 7, the six new regression coefficients are included gradually, and thus defining accordingly the balanced part of the different analysis variables. Details, showing which of the nine regression coefficients are active in each of the seven experiments, are listed in Table 2. Thus, each experiment has its own set of analysis control variables and runs WRFDA with the corresponding MBE. Some important features of MBE statistics, such as the balanced contributions, eigenvalues, and horizontal length scales for all seven experiments, are discussed in the rest of this section.

4.2 Balanced part contributions

Figure 2 displays the contribution of different control variables in the balanced part of different variables for the two regions. It can be seen that in the tropical region, the contribution of unbalanced velocity potential in the balanced part of temperature, surface pressure, and relative

Table 1 Pressure estimate (based on 1000 hPa surface pressure) for 51-vertical sigma levels, used for tropical and Arctic regions

Sigma level	Pressure (hPa)	Sigma level	Pressure (hPa)	Sigma level	Pressure (hPa)	Sigma level	Pressure (hPa)	Sigma level	Pressure (hPa)
1	1000	12	830.71	23	450.55	34	185.23	45	45.64
2	994.06	13	803.98	24	418.87	35	168.4	46	37.72
3	986.14	14	774.28	25	389.17	36	152.56	47	29.8
4	978.22	15	742.6	26	361.45	37	136.72	48	21.88
5	968.32	16	707.95	27	334.72	38	122.86	49	15.544
6	957.43	17	671.32	28	309.97	39	109.99	50	12.772
7	931.69	18	632.71	29	286.21	40	97.12	51	10
8	915.85	19	592.12	30	263.44	41	85.24		
9	893.03	20	554.5	31	241.66	42	74.35		
10	877.24	21	517.87	32	221.86	43	64.45		
11	855.46	22	483.22	33	203.05	44	54.55		

Fig. 1 Geographical display of the regions under study **a** tropical, and **b** Arctic

Table 2 List of experiments, showing the regression coefficients that are active in each experiment

Experiment name	Active regression coefficients
Exp-1	$\alpha_{\chi\psi}$, $\alpha_{T\psi}$, $\alpha_{ps\psi}$
Exp-2	$\alpha_{\chi\psi}$, $\alpha_{Tt\psi}$, $\alpha_{ps\psi}$, $\alpha_{\chi_u T}$
Exp-3	$\alpha_{\chi\psi}$, $\alpha_{T\psi}$, $\alpha_{ps\psi}$, $\alpha_{\chi_u T}$, $\alpha_{\chi_u ps}$
Exp-4	$\alpha_{\chi\psi}$, $\alpha_{T\psi}$, $\alpha_{ps\psi}$, $\alpha_{\chi_u T}$, $\alpha_{\chi_u ps}$, $\alpha_{\psi rh}$
Exp-5	$\alpha_{\chi\psi}$, $\alpha_{T\psi}$, $\alpha_{ps\psi}$, $\alpha_{\chi_u T}$, $\alpha_{\chi_u ps}$, $\alpha_{\psi rh}$, $\alpha_{\chi_u rh}$
Exp-6	$\alpha_{\chi\psi}$, $\alpha_{T\psi}$, $\alpha_{ps\psi}$, $\alpha_{\chi_u T}$, $\alpha_{\chi_u ps}$, $\alpha_{\psi rh}$, $\alpha_{\chi_u rh}$, $\alpha_{T_u rh}$
Exp-7	$\alpha_{\chi\psi}$, $\alpha_{T\psi}$, $\alpha_{ps\psi}$, $\alpha_{\chi_u T}$, $\alpha_{\chi_u ps}$, $\alpha_{\psi rh}$, $\alpha_{\chi_u rh}$, $\alpha_{T_u rh}$, $\alpha_{ps_u rh}$

It may be noted that in each case, all the coefficients activated in the preceding experiments remains active in the experiments which follow

humidity is large as compared with the stream function, whereas for the Arctic region the contribution of stream function is greater than the unbalanced velocity potential, in the balanced part of the other analysis variables. As a result, the inclusion of additional correlations, between unbalanced velocity potential and temperature, and between unbalanced velocity potential and surface pressure, is expected to enhance the divergent part of the wind more in the tropical than in the Arctic region.

In addition, it is seen that the total contributions of the balanced part of surface pressure (from stream function and unbalanced velocity potential) is about 34 % in the tropical region and about 90 % in the Arctic region. For the Arctic region, most of the contribution (75 %) in the balanced part of surface pressure is from stream function. However, for the tropical region, it is the velocity potential via its unbalanced part which contributes 22 % to the balanced part for the surface pressure. This means that with the formulation of new analysis control variables in the tropical region, the velocity potential field will have greater impact on the surface pressure. It is also seen that, in both the tropical and Arctic regions, the balanced part of relative

humidity is mainly due to the unbalanced part of temperature. Thus in both the regions the temperature, via its unbalanced part, may affect the moisture analysis.

Here, it may be noted that Exp-3 collectively deals with the impact of the correlation of unbalanced velocity potential with both the temperature and surface pressure. The contributions of the unbalanced part of velocity potential and surface pressure on relative humidity is very small, and so no significant changes are expected between the output of experiments Exp-4, Exp-5, Exp-6 and Exp-7. The contribution of stream function to *rh* is very small, and so the output of Exp-3 and Exp-4 may not differ much. It is mainly due to these reasons that only the assimilation results corresponding to Exp-1, Exp-3 and Exp7 have been discussed in Sect. 6.
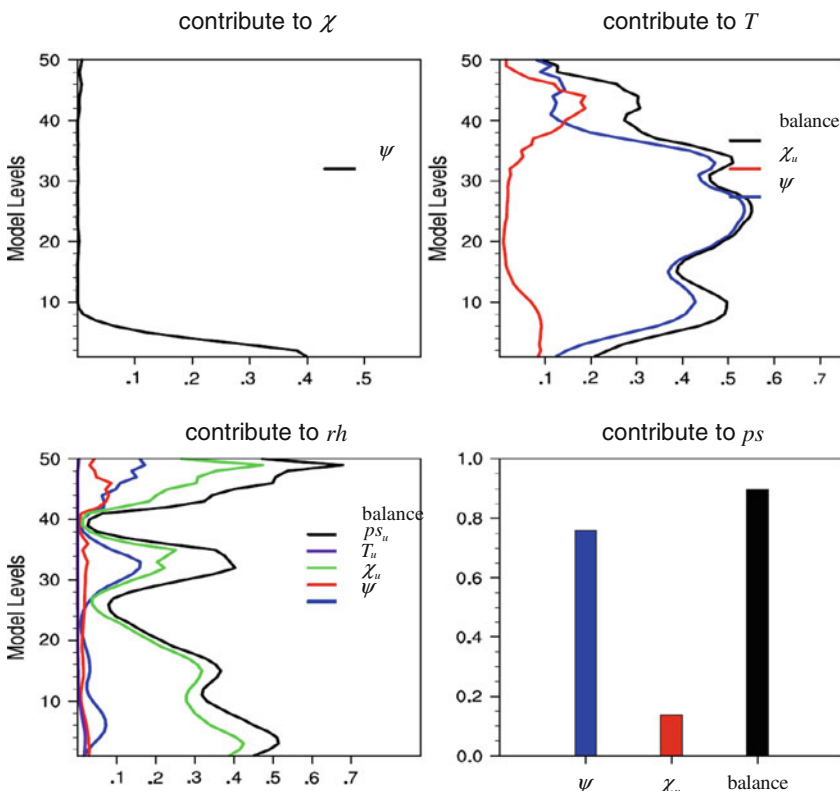
4.3 Eigenvalues and horizontal length-scales

For the tropical region, Fig. 3 displays eigenvalues and horizontal length scales of unbalanced temperature (T_u) and relative humidity (*rh*) analysis control variables corresponding to each of the seven experiments. In this figure, the *x*-axis represents the vertical mode number for the corresponding EOF and the *y*-axis displays the eigenvalues (Fig. 3a) and the horizontal length scale (Fig. 3b) for the corresponding EOF mode. For each of the seven experiments and for each variable the corresponding values of the horizontal length scales and eigenvalues are used in the respective $\mathbf{U}_h$ and $\mathbf{U}_v$ transforms.

Since in all seven experiments the first two analysis control variables, namely the stream function (ψ) and unbalanced velocity potential (χ_u), are the same, the corresponding eigenvalues and length-scales will not differ in all seven experiments (not shown in Fig. 3). However, for the unbalanced temperature and relative humidity control variables, some changes are seen both in the eigenvalues and the horizontal length-scales. With the inclusion of
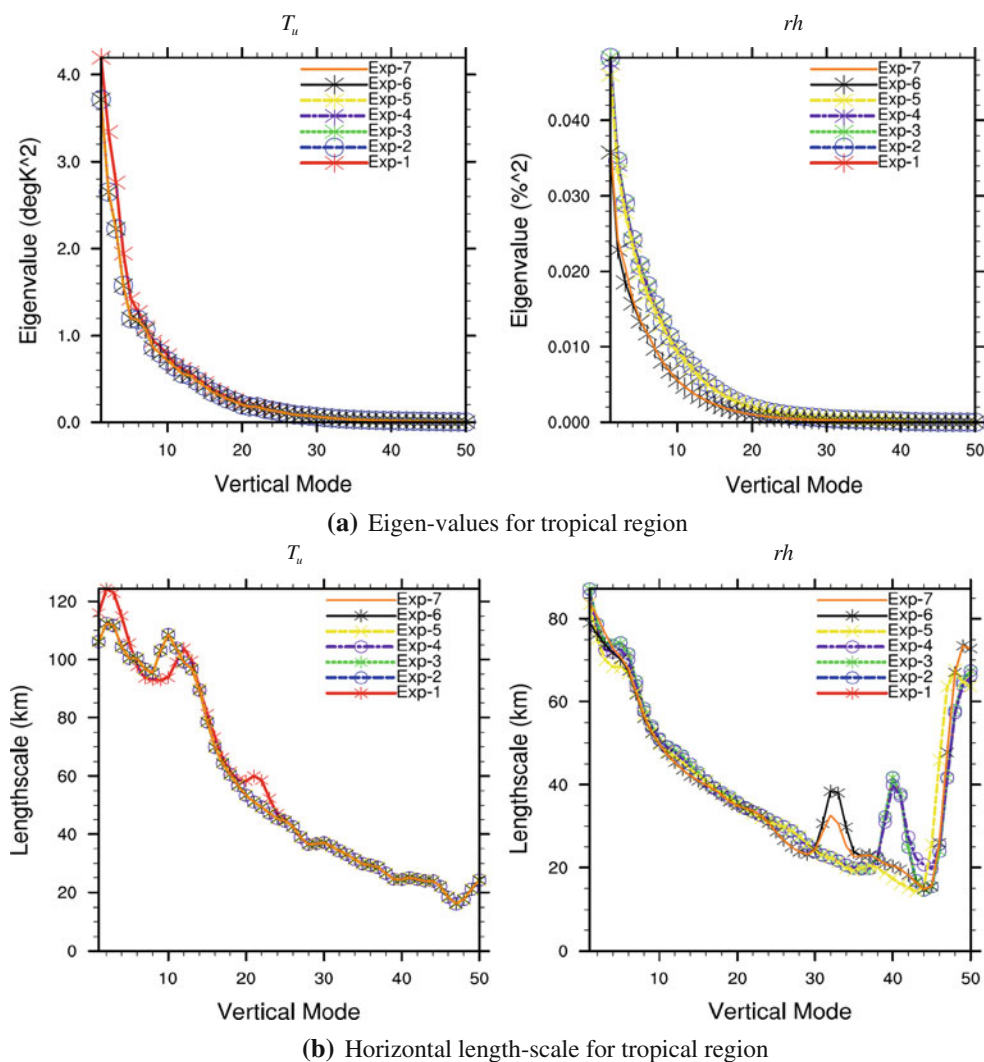
Fig. 2 Contributions to the balanced part of χ, *T*, *rh,* and *ps* from other variables, as shown in the respective legend or on the *x*-axis (in case of *ps*). The total balanced part is indicated as "balance" (*black*) for **a** tropical region, and **b** Arctic region
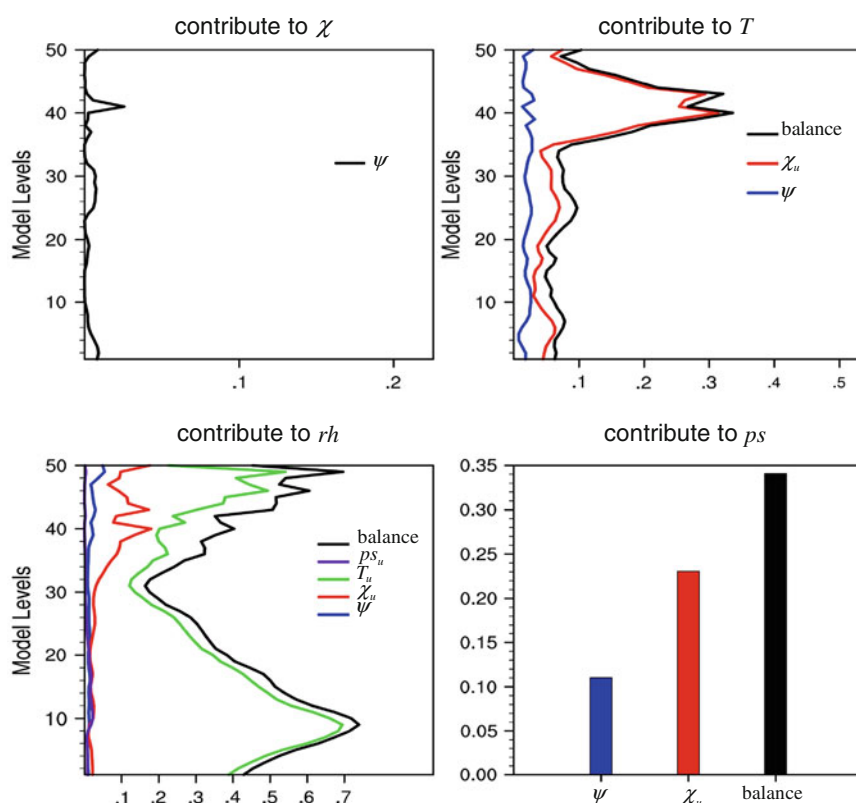
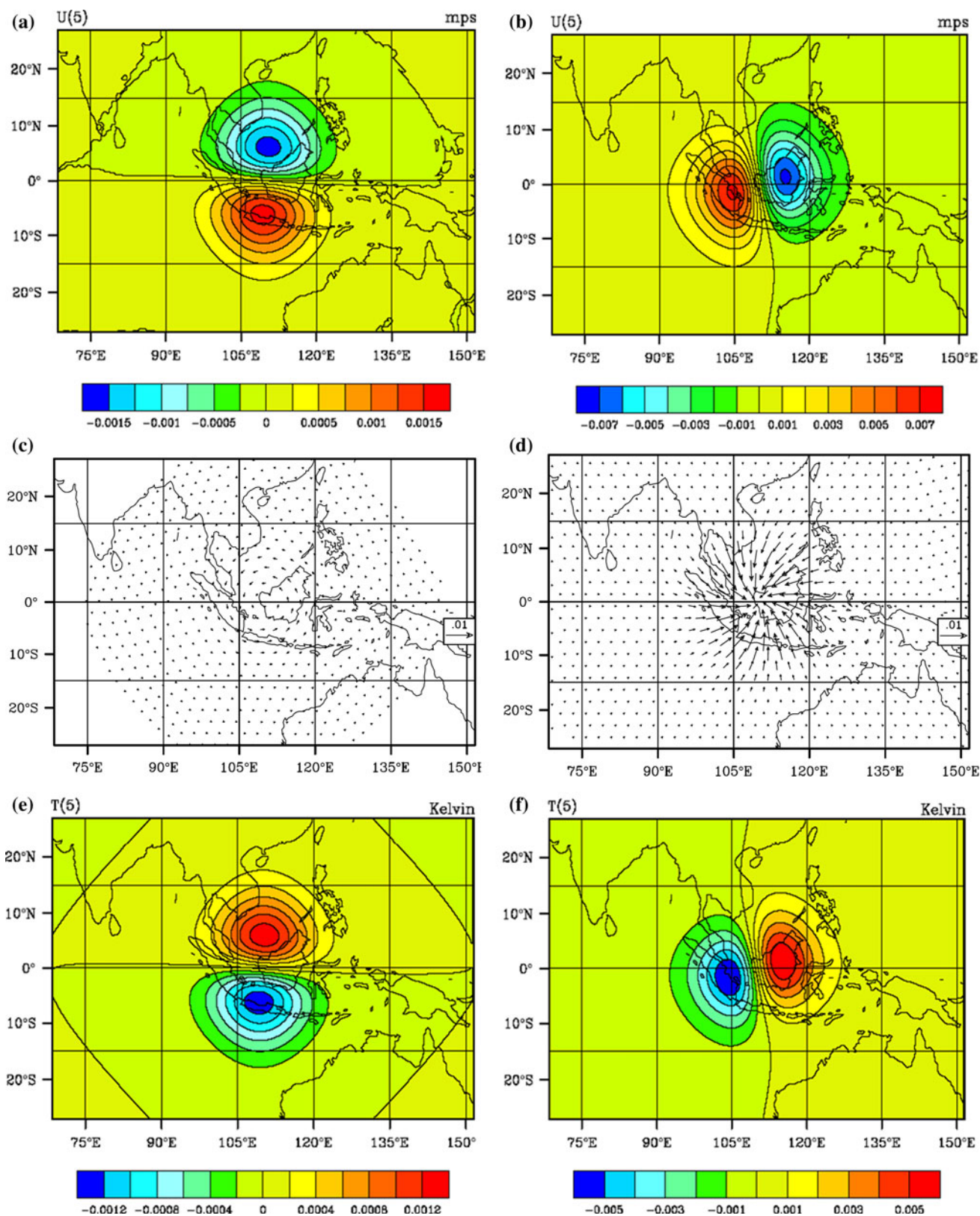

(a) Tropical region

(b) Arctic region

(a) Eigen-values for tropical region

(b) Horizontal length-scale for tropical region

Fig. 3 For the tropical region, display of **a** eigenvalues, and **b** horizontal length-scale for unbalanced temperature (T_u) and relative humidity (rh) analysis control variables, corresponding to all seven experiments

moisture correlations (Exp-4 to 7) the drop of eigenvalues for relative humidity with increasing mode number (lower right panel in Fig. 3a) is less as compared with Exp-1, Exp-2 or Exp-3, implying that the moisture analysis corresponding to Exp-4 to 7 may draw more information from the moisture observations. No significant difference is seen in the unbalanced temperature in different experiments (lower left panel in Fig. 3a). As shown in Fig 3b (lower left panel), for the first couple of modes (which are weighted most), the horizontal length-scale of unbalanced temperature is slightly smaller (~ 10 km) in all the experiments (Exp-2 to Exp-7) as compared with the control experiment (Exp-1). Thus with additional regression coefficients, the changes in the temperature analysis increments will have slightly less horizontal influence, compared with the temperature analysis increments from the $\alpha_{\psi T}$ regression coefficient in Exp-1. In all seven

experiments, the first 25 modes for horizontal length scales of relative humidity do not differ much. Some experiments show sharp fluctuations in the relative humidity length scales for the higher modes (>30). These fluctuations are due to the very small quantity of moisture that is represented by the higher modes of relative humidity. As a result, for higher modes there are insufficient numbers of moisture "bins" available to fit the Gaussian curve for estimating the horizontal length scales. In reality, there is very little moisture above 200 hPa, and the structure of relative humidity EOFs is almost flat above 30th sigma level (not shown). Thus, even if these higher modes for relative humidity are used in WRFDA, it may not have much effect on the moisture analysis.

For the Arctic region, the characteristics of eigenvalues and horizontal length scales are similar to the tropical region but the corresponding magnitudes are different (not

Fig. 4 For the tropical region, horizontal cross-section of wind vector and *u*-component of wind increment at 5th sigma level as a result of assimilating a single temperature observation at the same sigma level at the center of the region. **a** Exp-1 (w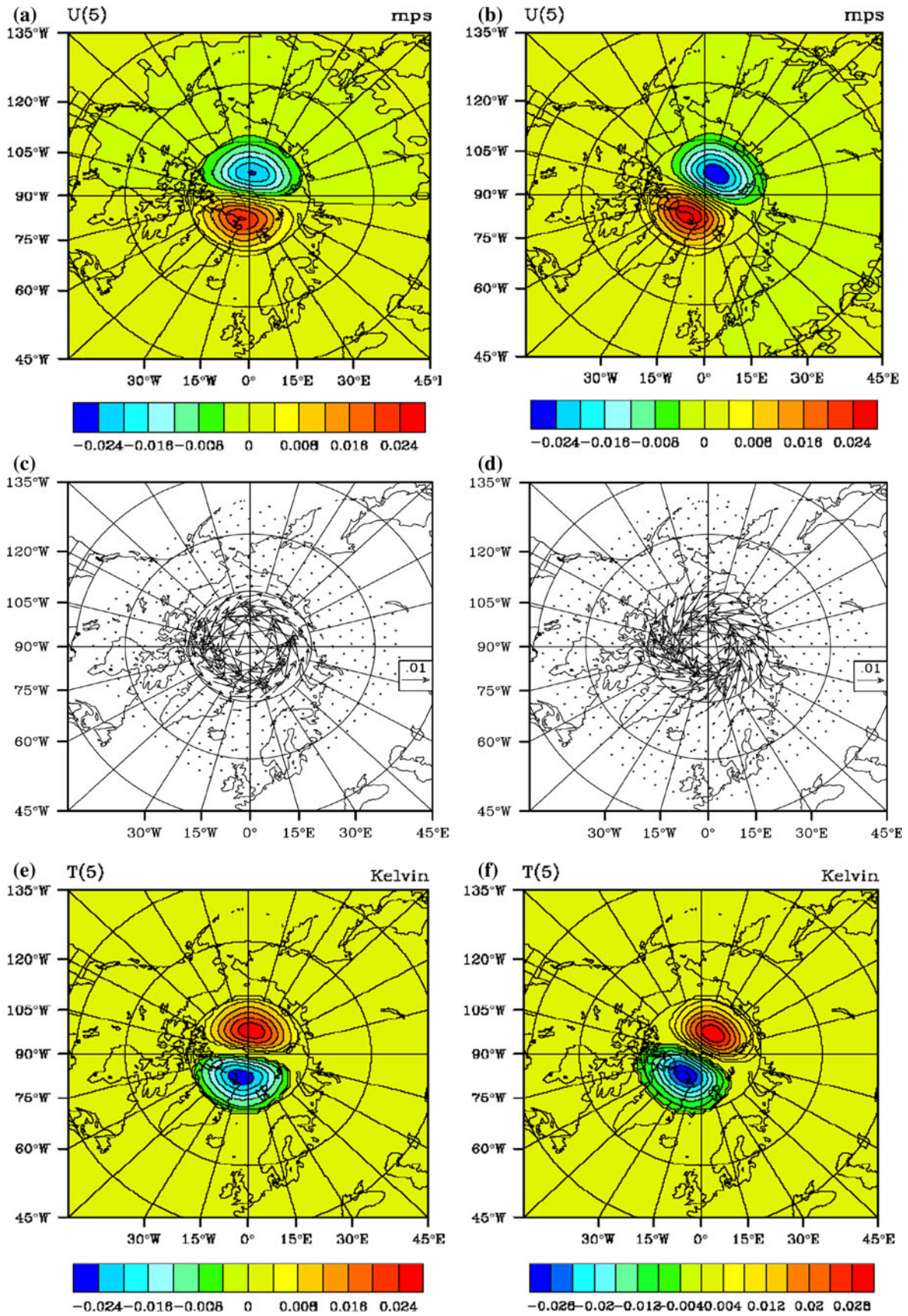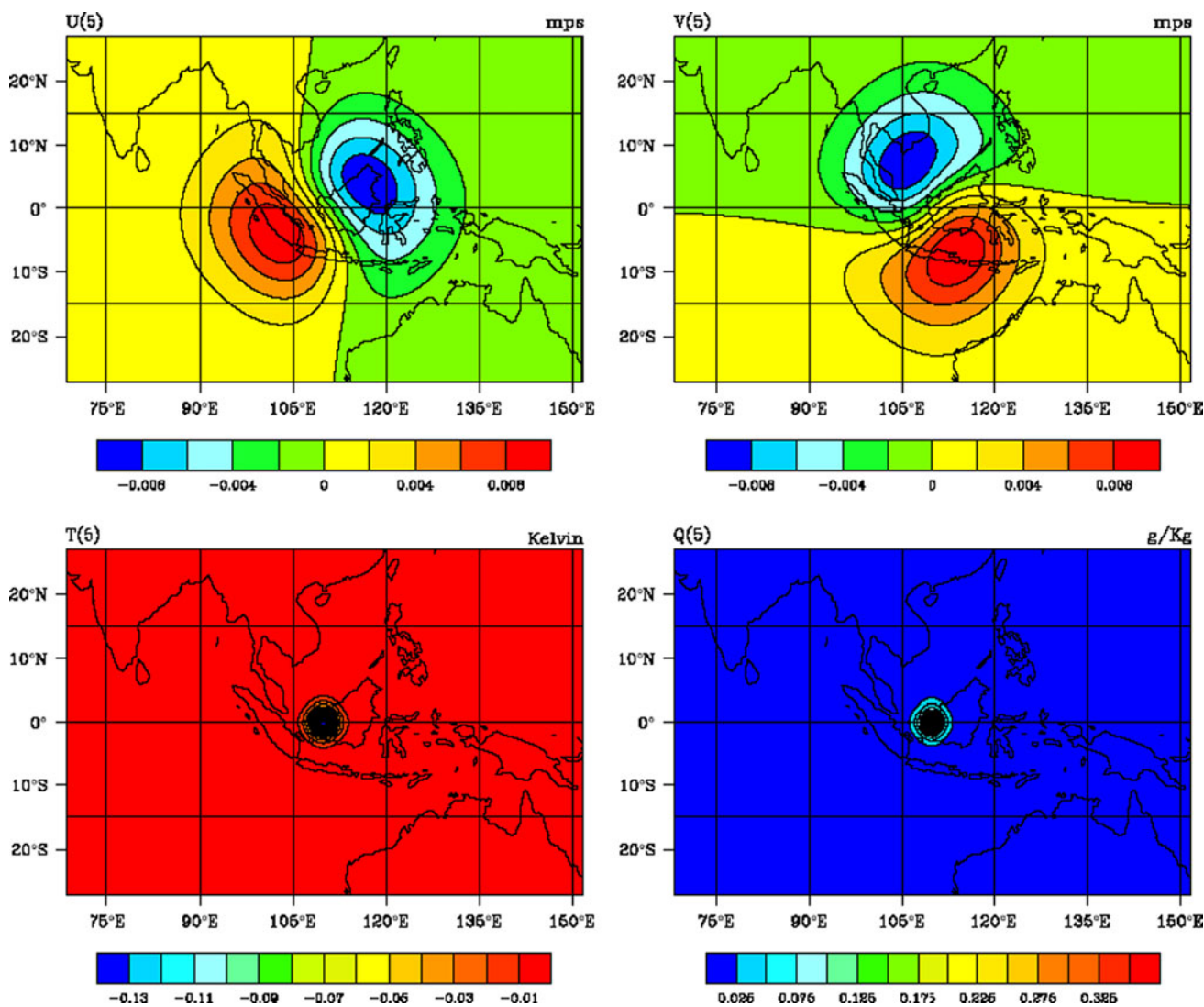ithout $\alpha_{\chi_u T}$), and **b** Exp-2 (with $\alpha_{\chi_u T}$). Horizontal cross-section of temperature increment at the 5th sigma level as a result of assimilating a single wind (*u*) observation at the same sigma level at the center of the region. **c** Exp-1 (without $\alpha_{\chi_u T}$), and **d** Exp-2 (with $\alpha_{\chi_u T}$)

Fig. 5 Same as Fig. 4, but for the Arctic region

Fig. 6 For the tropical region, horizontal cross-section of analysis increments for wind (*u* and *v* components), temperature, and specific humidity at the 5th sigma level as a result of assimilating a single moisture observation at the same sigma level at the center of the region for Exp-7

shown). As an example, for the Arctic region the horizontal length scale for the leading eigenvectors (first ten) of unbalanced velocity potential is about 300 km, whereas for the tropical region it is greater than 500 km. The larger length scale for unbalanced velocity potential may influence the divergent component of wind at larger scales in the tropical region than in Arctic region.

5 Single observation test results

To understand the responses of different regression coefficients and the overall structure of MBE, a series of single observation assimilation tests are undertaken for both the tropical and Arctic regions.

5.1 The effect of $\alpha_{\chi_u T}$

Results for the assimilation of a single temperature (*T*) or wind (*u*) observation in the tropical region suggest that, with the inclusion of the $\alpha_{\chi_u T}$ correlation (Exp-2), a slight increase in the magnitude of wind increment is seen (Fig. 4b) as compared with Exp-1 (Fig. 4a). The magnitude of convergence/divergence also increased due to additional contributions to the divergent component of wind with the inclusion of the $\alpha_{\chi_u T}$ term (Fig. 4c, d). Comparison of Fig. 4a and b also suggests that, due to enhanced convergence/divergence, there is a rotation in the location of maxima/minima of wind speed (*u*) increments with Exp-2 as compared with Exp-1. Due to the symmetric property of correlations, with the assimilation of a single wind (*u*)

(a) Bias profiles

(b) RMSE profiles

EXP-7

EXP-3
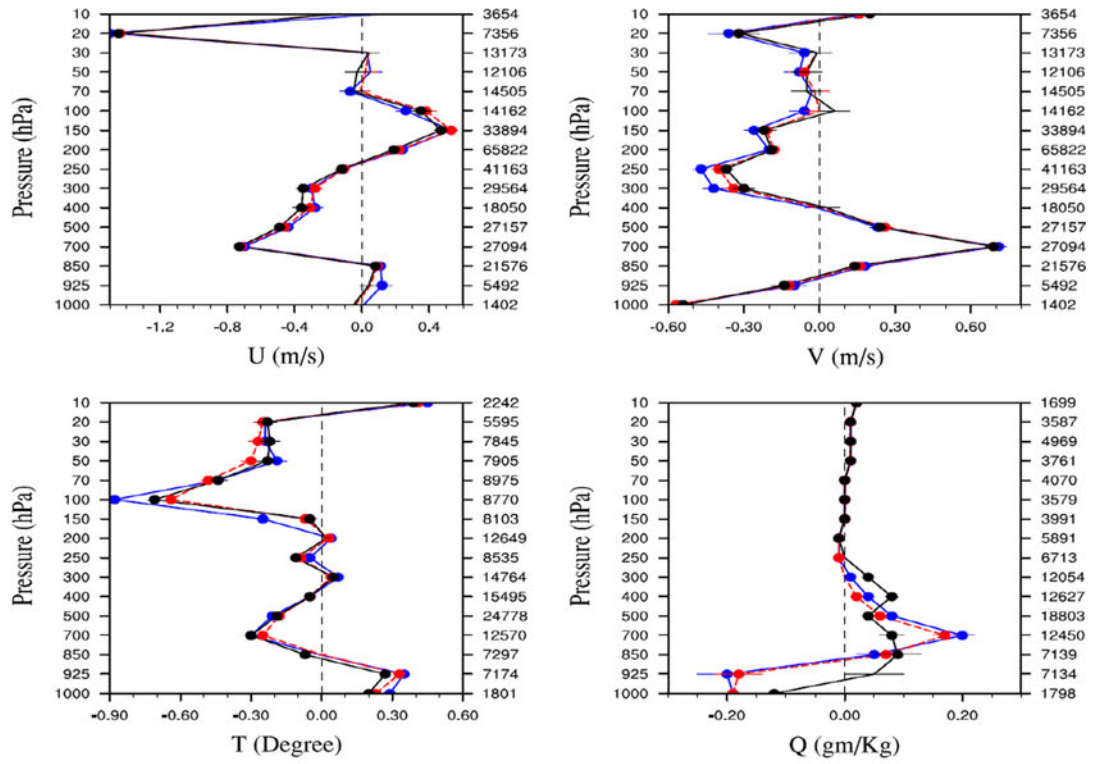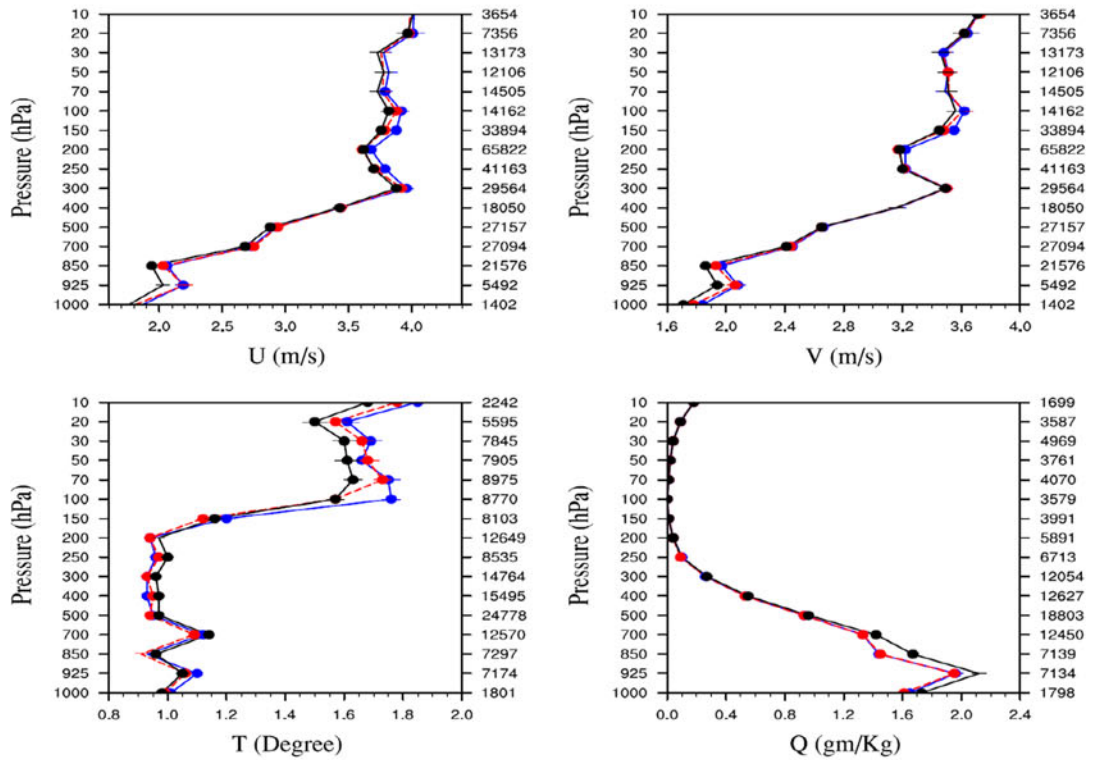
EXP-1

◀ **Fig. 7** Analysis verification scores for the tropical region (from 15 July to 15 August 2009), BIAS (**a**) and RMSE (**b**), for Exp-1, 3, and 7. Level wise, the total number of observations used in verification is displayed on the *right-hand side* of the *vertical axis*

observation, a similar response (rotation of maxima/minima) is also seen in the temperature increment (Fig. 4e, f).

Similar results are also seen for the Arctic region, with parallel runs of assimilating the single temperature (T) and wind (u) observations (Fig. 5). Since the contribution of the $\alpha_{\chi_u T}$ term is less in the Arctic as compared with the tropical region, an accordingly smaller increase in the magnitude of wind and temperature increment is seen in Exp-2 (Fig. 5, right panel) as compared with Exp-1 (Fig. 5, left panel). Due to the same reason, since the magnitude of convergence/divergence is also less, the rotation in the maxima/minima of the temperature and wind (u) increments is also relatively less in the Arctic (Fig. 5), as compared with the tropical region (Fig. 4).

5.2 The effect of $\alpha_{\chi_u ps}$

In tropical region, with the inclusion of the $\alpha_{\chi_u ps}$ term (Exp-3), a slight increase in the magnitude of surface pressure, temperature, and wind (u) analysis increment is observed with the assimilation of a single surface pressure observation (not shown). Since the contribution of the $\alpha_{\chi_u ps}$ correlation in the Arctic is less than that in the tropical region, an accordingly smaller increase in the magnitude of surface pressure, temperature, and wind (u) analysis increments is seen in the Artic region (not shown).

5.3 The effect of, $\alpha_{\psi rh}$, $\alpha_{\chi_u rh}$, $\alpha_{T_u rh}$ and $\alpha_{ps_u rh}$

As expected, with active $\alpha_{\psi rh}$, $\alpha_{\chi_u rh}$, $\alpha_{T_u rh}$ and $\alpha_{ps_u rh}$ terms (Exp-7), assimilation of a single moisture observation yields multivariate analysis increments, both in the tropical and Arctic regions. For the tropical region, the response of assimilating a single moisture observation on wind (u and v), temperature, and moisture analysis increments with MBE corresponding to Exp-7 is shown in Fig. 6. Similar multivariate response in analysis increments with the assimilation of single moisture observation is not possible with Exp-1, Exp-2 and Exp-3. Similar multivariate response in analysis increments is also seen in the Arctic region but with slightly less magnitude in the analysis increments, as compared with the tropical region (not shown).

## 6 Data assimilation results

For each of the seven experiments (Table 2), a parallel six-hourly cycling data assimilation experiment is run for a 1-month period, running from 00 UTC of 15 July to 18 UTC of 15 August 2009. Each experiment starts the first data assimilation cycle by running WRFDA at 2009071500 with the corresponding MBE, using the GFS analysis as the background. The background input for each following assimilation cycle is produced using the 6-h forecast initialized with the WRFDA analysis from the previous cycle. Parallel, 72-h forecasts are made with 00 and 12 initial conditions (ICs) produced in the respective data assimilation cycling experiments. In all the experiments, the same boundary conditions derived from the GFS analysis are used. All observations which are identified as "good" by the WRFDA "quality control procedure" in the control assimilation cycle run (Exp-1) are used in verifying the analyses and forecasts produced for each of the seven experiments. Verification scores of root mean square error (RMSE) and bias are computed for the zonal (u) and meridional (v) components of wind, temperature (T), and specific humidity (q). Figure 7 displays analysis verification scores for the tropical region corresponding to Exp-1, Exp-3, and Exp-7. Parallel results for 6-h forecast verification scores are shown in Fig. 8. Assimilation results with Exp-1 and Exp-3 are compared to understand the impact of unbalanced velocity potential (χ_u). For the tropical region, it is seen that both BIAS and RMSE analysis scores for Exp-3 are marginally better when compared with Exp-1 (Fig. 7). In addition, most of the improvements in the analysis with Exp-3 are retained in the 6-h WRF forecast (Fig. 8).
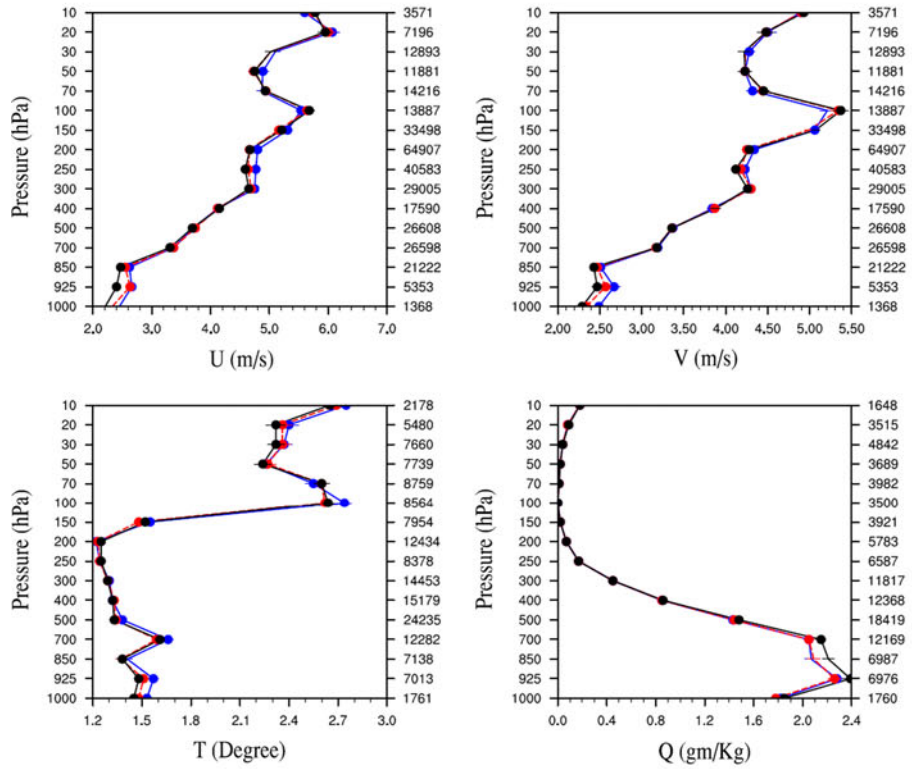
Parallel results for the Arctic region are shown in Figs. 9 and 10. For this region the improvement is relatively less than that seen in the tropical region, and it is mainly seen at the higher sigma levels. This is consistent with the contribution of unbalanced velocity potential for the two regions, as shown in Fig. 2. Thus in tropical region, inclusion of $\alpha_{\chi_u ps}$ and $\alpha_{\chi_u T}$ terms in Exp-3 helped improve wind analyses and 6-h forecasts because of better representation of the divergent part of wind in this region. These results are also consistent with the results from the assimilation of single observations, discussed earlier.

Exp-4 through Exp-7 differs from Exp-2 and Exp-3 in the formulation of the moisture analysis control variables. In Exp-4 through Exp-7, relative humidity correlations with stream function, unbalanced velocity potential, temperature and surface pressure are added gradually. These moisture correlations lead to the partitioning of relative humidity in balanced and unbalanced parts. It is seen that there is not much difference in the verification scores by activating $\alpha_{\psi r_h}$, $\alpha_{\chi_u rh}$ and $\alpha_{ps_u rh}$ (not shown), but there is some effect seen in activating $\alpha_{T_u rh}$. This is mainly because in both the regions the balanced part of relative humidity is largely contributed by the unbalanced part of temperature field, as

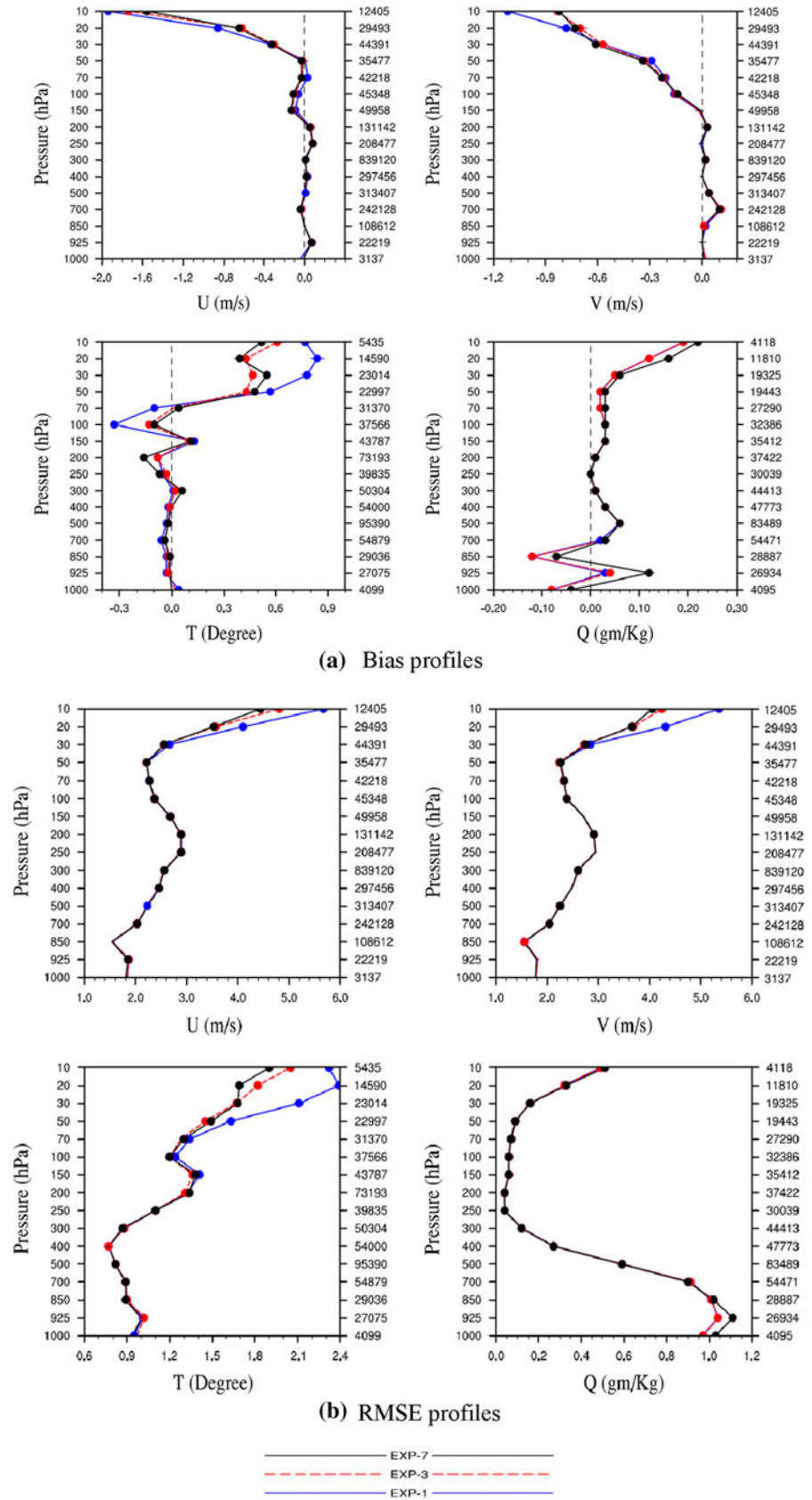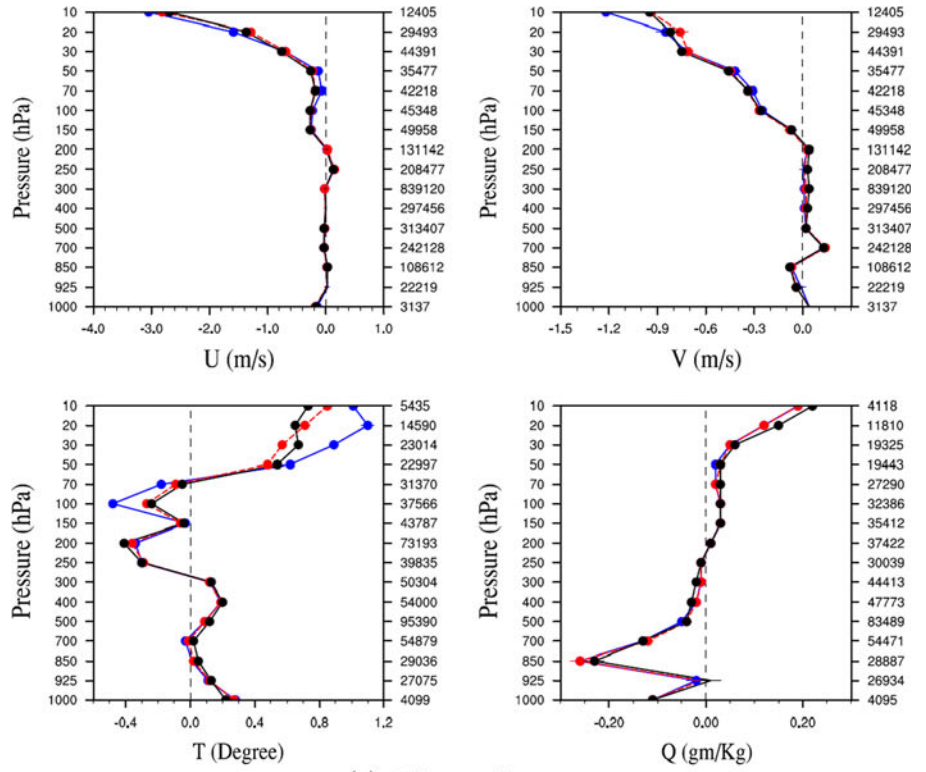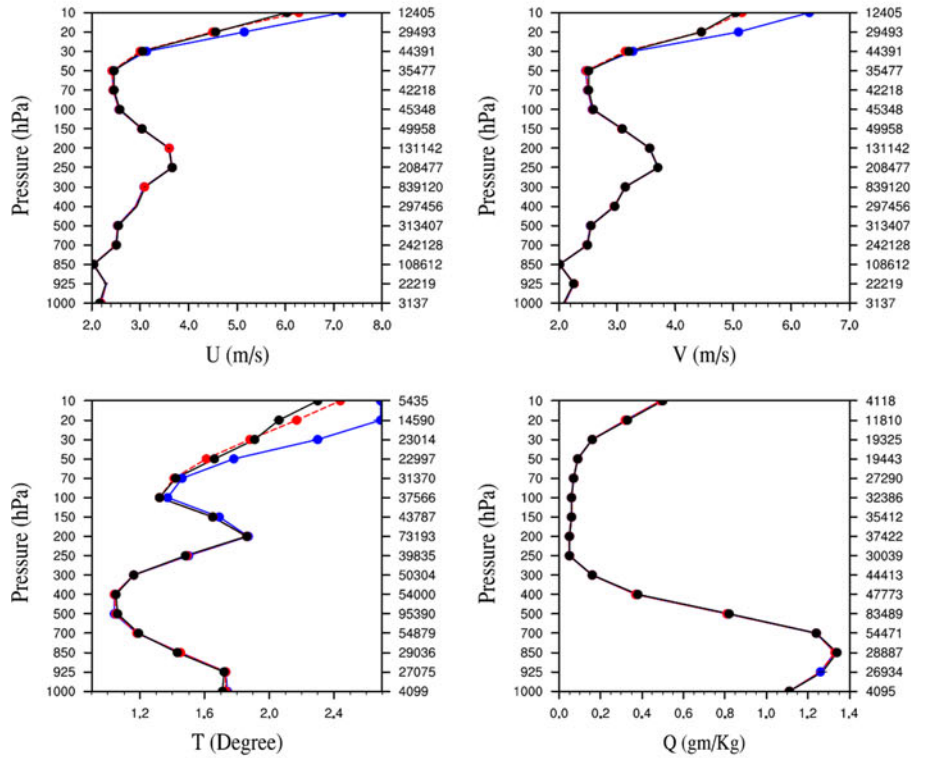**Fig. 8** Same as Fig. 7, but for the 6-h forecast
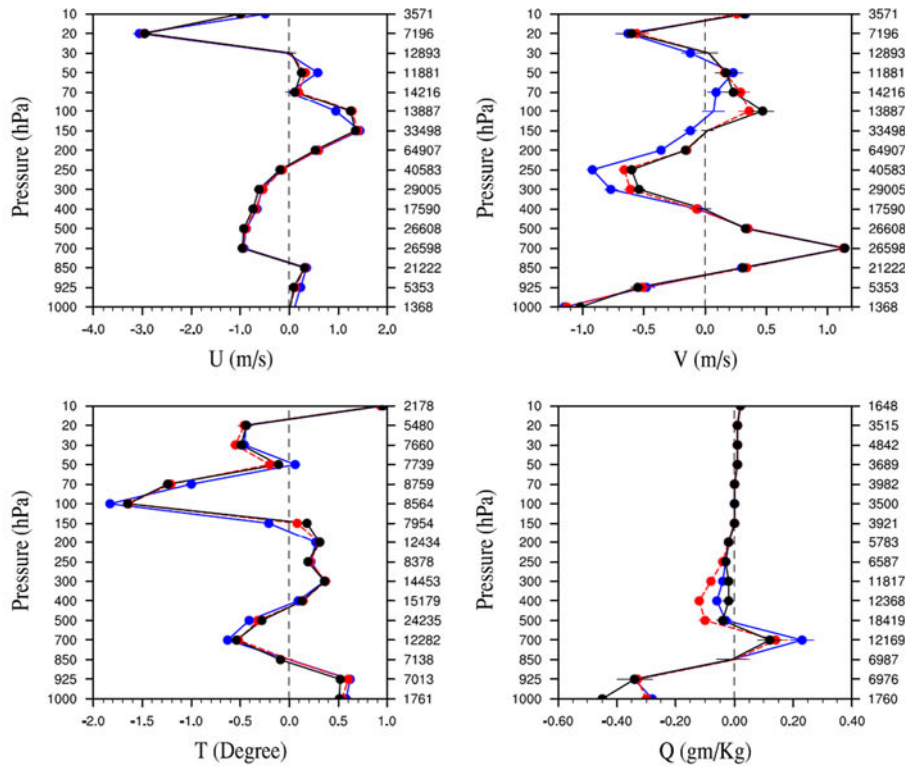
Fig. 9 Analysis verification scores for the Arctic region (from 15 July to 15 August 2009), **a** BIAS and **b** RMSE, for Exp-1, 3. and 7. Level wise, the total number of observations used in verification is displayed on the *right-hand side* of the *vertical axis*
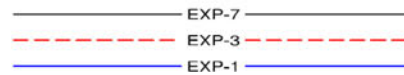

(a) Bias profiles

(b) RMSE profiles

EXP-7
EXP-3
EXP-1

Fig. 10 Same as Fig. 9, but for
the 6-h forecast

(a) Bias profiles

(b) RMSE profiles

Fig. 11 For the tropical region, 24-h forecast verification scores (from 15 July to 6 August 2009) for Exp-1 and Exp-7 **a** BIAS, and **b** RMSE
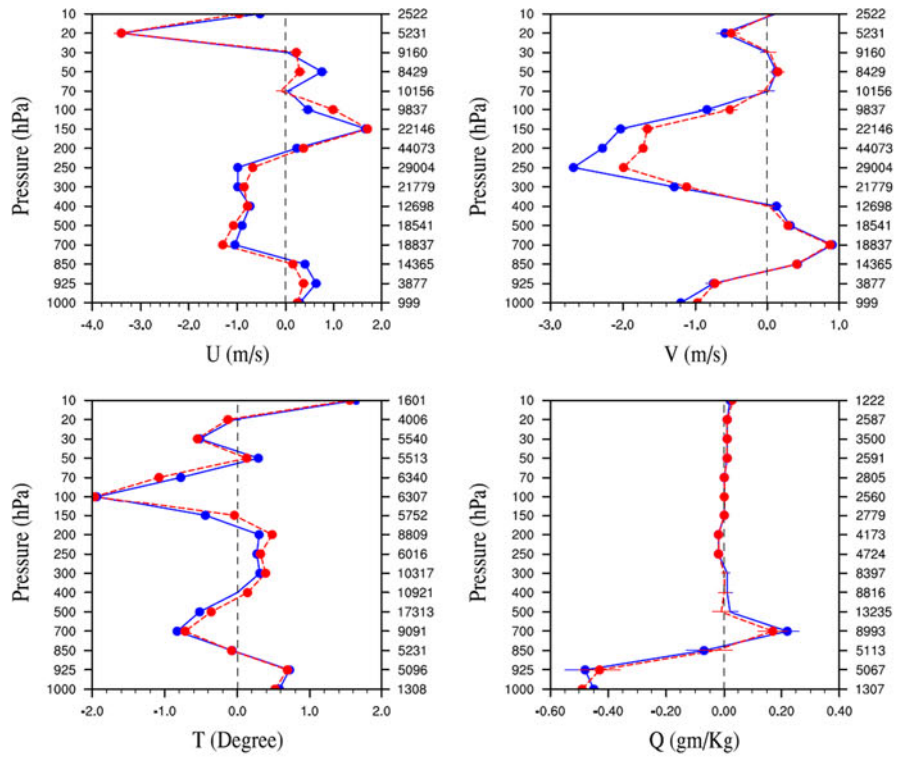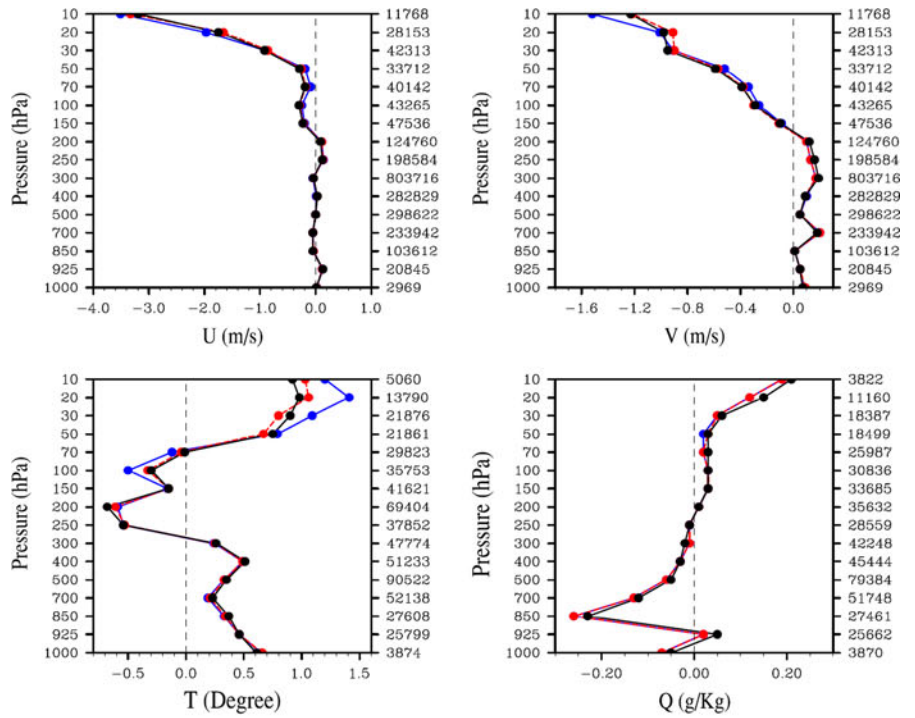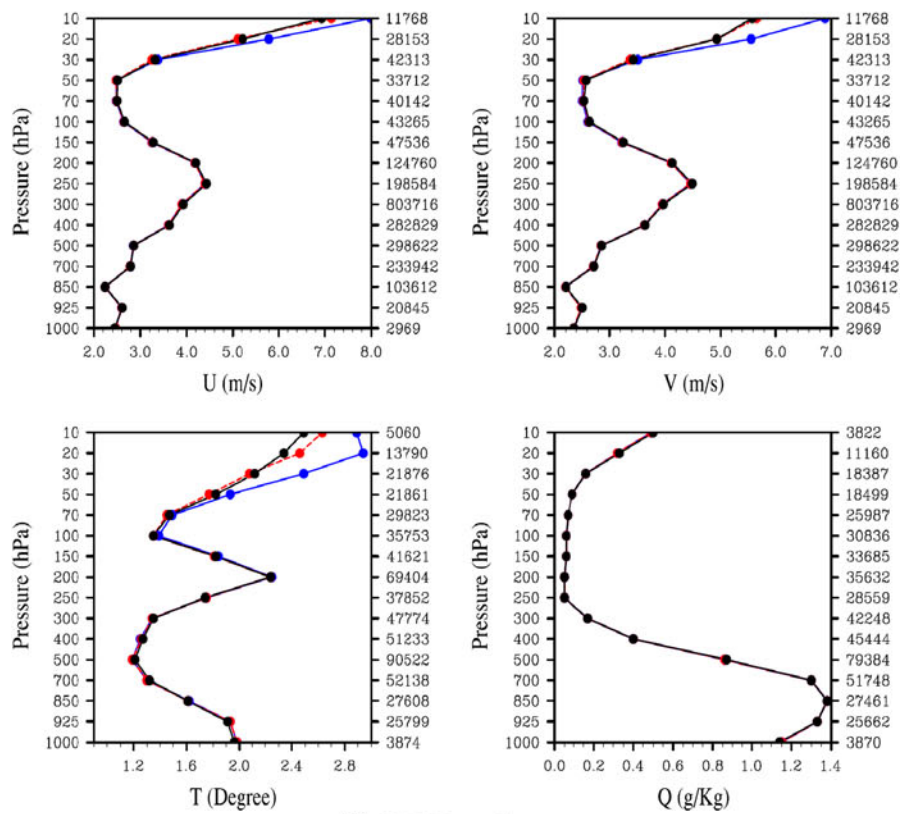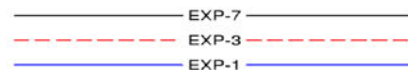

(a) Bias profiles
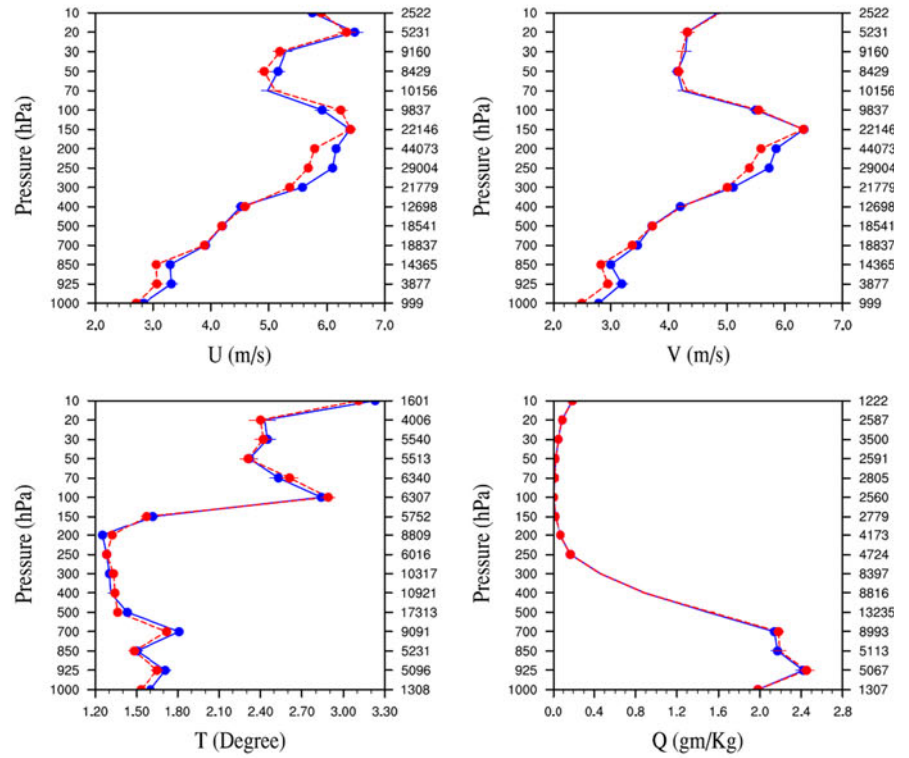
(b) RMSE profiles

Fig. 12 Same as Fig. 11, but for the Arctic region

(a) Bias profiles

(b) RMSE profiles

EXP-7 ———
EXP-3 – – –
EXP-1 ———

shown in Fig. 2. In the tropical region, comparison of analysis and 6-h forecast verification scores for Exp-3 and Exp-7 suggests that with the addition of moisture correlations, both analysis and 6-h forecast verification scores are slightly better for wind and temperature fields. However, in the 6-h forecast moisture fields, slight deterioration is seen in the lower levels. BIAS scores for analyses in the lower levels indicate that the analysis is "over-fitting" the moisture observations, suggesting moisture observation errors need to be tuned. For the Arctic region, not much difference is seen between Exp-3 and Exp-7 for wind and temperature analysis scores (Fig. 9 and 10).

For the tropical region, 24-, 48-, and 72-h forecasts from 00 to 12 UTC initial conditions are verified for a period of 20 days (00 UTC of 15 July to 18 UTC of 15 August 2009). Results for the verification of 24-h WRF model forecasts for Exp-1 and Exp-7, as shown in Fig. 11, indicate that the verification scores corresponding to Exp-7 (with all correlations included) are marginally better than the control run (Exp-1). However, for long-range forecast (48 and 72 h), no significant difference is seen between Exp-1 and Exp-7 (not shown).

For the Arctic region, a positive effect with MBE is seen, especially above the jet level both in the analyses (Fig. 9), 6-h forecasts (Fig. 10) and 24-h forecasts (Fig. 12). Like in the tropical region, no significant difference is seen in the long-range forecast for the Arctic region (not shown).

7 Summary and conclusions

For two regions, tropical (representing the Tropics) and the Arctic (representing higher latitudes), a variety of background error statistics with the inclusion of linear regression across different analysis control variables has been computed using a new formulation of multivariate background errors (MBE). It is seen that the characteristics of the background error covariance matrix for the tropical region differ significantly from those in the Arctic region. In the case of the tropical region, the contribution of velocity potential to the balanced part of other variables is much larger than that of the stream function. However, for the Arctic region the role of the stream function is more dominant compared with the velocity potential field. The total contribution of the balanced part of surface pressure is higher (about 90 %) in the Arctic region than in the tropical region (about 34 %). Both in the tropical and Arctic regions, contributions to the balanced part of relative humidity are mainly due to the unbalanced part of temperature. The unbalanced part of the surface pressure contributes very little to the balanced part of relative humidity. One month-long 6-h cycling data assimilation experiments for both regions suggest marginal improvement with the

inclusion of the correlations between the unbalanced velocity potential, temperature, and surface pressure. Since the divergent part of wind contributes more in the tropical than in the Arctic region, improvements in the tropical region are more apparent. Inclusion of additional moisture correlations did not show much difference in the analysis or the short-range NWP forecast, especially in the Arctic region.

Since the distribution of moisture highly depends on the synoptic situation, use of average (1 month) moisture correlations might not be very effective. Nevertheless, this study has built up necessary updates for MBE in the WRFDA system irrespective of how it gets generated (using either the NMC or the ensemble method). It is quite likely that moisture correlations with MBE input derived using case-based ensembles, either in pure 3D-VAR or in hybrid mode, might give better results. Since the quality of a 6-h model forecast is important for six-hourly data assimilation, this study has ensured that MBE has added some value to the 6-h forecast, especially in the tropical region. After gaining confidence, we are in the process of evaluating the impact of MBE on forecasts for some typical synoptic events. In addition, some changes are expected in MBE, especially due to different moisture distribution and forecast quality for winter and summer seasons.

References

Bannister RN (2008a) A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances. Q J R Meteor Soc 134:1951–1970

Bannister RN (2008b) A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics. Q J R Meteor Soc 134:1971–1996

Barker DM, Huang W, Guo Y-R, Bourges A (2003) A three-dimensional variational (3DVAR) data assimilation system for use with MM5. Section 3, NCAR Technical Note No. NCAR/TN-453 + STR. Available from MMM, National Center For Atmospheric Research, Boulder

Barker DM, Huang W, Guo Y-R, Xiao QN (2004) A Three-Dimensional (3DVAR) data assimilation system for use with MM5: implementation and initial results. Mon Weather Rev 132:897–914

Barker DM, Huang X-Y, Liu Z, Auligne T, Zhang X, Rugg S, Ajjaji R, Bourgeous A, Bray J, Che Y, Demirtas M, Guo Y-R, Henderson T, Huang W, Lin H-C, Michalakes J, Rizvi S, Zhang X (2012) The weather research and forecasting model's community variational/ensemble data assimilation system WRFDA. Bull Am Soc 93:831–843

Berre L (2000) Estimation of synoptic and mesoscale forecast error covariances in a limited-area model. Mon Weather Rev 128: 644–667

Berre L, Ştefănescu SE, Belo-Pereira M (2006) The representation of the analysis effect in three error simulation techniques. Tellus A 58:196–209

Brousseau P, Berre L, Bouttier F, Desroziers G (2011) Background-error covariances for a convective-scale data-assimilation system: AROME–France 3D-Var. Q J R Meteor Soc 137:409–422

Caron JF, Fillion L (2010) An examination of background error correlations between mass and rotational wind over precipitation regions. Mon Weather Rev 138:563–578

Courtier P, Talagrand O (1990) Variational assimilation of meteorological observations with the direct and adjoint shallowwater equations. Tellus 42:531–549

Courtier P, Thepaut JN, Hollingsworth A (1994) A strategy for operational implementation of 4D-Var using an incremental approach. Quart JR Meteor Sco 120:1367–1388

Derber JC, Bouttier F (1999) A reformulation of the background error covariance in the ECMWF global data assimilation system. Tellus A 51:195–221

Fisher M (2003) Background-error covariance modeling. ECMWF seminar on recent developments in data assimilation for atmosphere and ocean. ECMWF, Reading, pp 45–63

Huang XY, Xiao Q, Barker DM, Zhang X, Michalakes J, Huang W, Henderson T et al (2009) Four-dimensional variational data assimilation for WRF: formulation and preliminary results. Mon Weather Rev 137:299–314

Krysta M, Rizvi S, Huang XY (2009) A new formulation of WRFDA analysis control variables. 10th WRF Users' Workshop, 23–26 June 2009

Lorenc AC (2003) Modeling of error covariances by 4d-Var assimilation. Q J R Meteor Soc 129:3167–3182

Michel Y, Auligne T (2010) Inhomogeneous background error modeling and estimation over Antarctica. Mon Weather Rev 138:2229–2252

Montmerle T, Lafore JP, Berre L, Fischer C (2006) Limited-area model error statistics over Western Africa: comparisons with midlatitude results. Q J R Meteor Soc 132:213–230

Parrish DF, Derber JC (1992) The National Meteorological Center's Spectral Statistical Interpolation analysis system. Mon Weather Rev 120:1747–1763

Pereira M, Berre L (2006) The use of an ensemble approach to study the background error covariances in a global NWP. Mon Weather Rev 134:2466–2489

Rawlins F, Ballard SP, Bovis KJ, Clayton AM, Li D, Inverarity GW, Lorenc AC, Payne TJ (2007) The Met Office global four-dimensional variational data assimilation scheme. Q J R Meteor Soc 133:347–362

Sadiki W, Fischer C (2005) A posteriori validation applied to the 3D-VAR Arpège and Aladin data assimilation systems. Tellus A 57:21–34

Sadiki W, Fischer C, Geleyn JF (2000) Mesoscale background error covariances: recent results obtained with the limited-area model ALADIN over Morocco. Mon Weather Rev 128:3927–3935

Skamarock WC, Klemp J, Dudhia J, Gill DO, Barker DM, Duda MG, Huang XY, Wang W, and Powers JG (2008) A description of the advanced research WRF Version 3, available at: http://www.mmm.ucar.edu/wrf/users/docs/arw_v3.pdf

Storto A, Randriamampianina R (2010) Ensemble variational assimilation for the representation of background error covariances in a high-latitude regional model. J Geophys Res 115:D17204. doi:10.1029/2009JD013111

Wu W-S, Purser JR (2002) Three-dimensional variational analysis with spatially inhomogeneous covariances. Mon Weather Rev 130:2905–2916

Žagar N, Andersson E, Fisher M (2005) Balanced tropical data assimilation based on a study of equatorial waves in ECMWF short-range forecast errors. Q J R Meteor Soc 131:987–1011