



# Impact of variational assimilation using multivariate background error covariances on the simulation of monsoon depressions over India

M. Dhanya and A. Chandrasekar

Department of Earth and Space Sciences, Indian Institute of Space Science and Technology,  
Thiruvananthapuram, Kerala, India

*Correspondence to:* A. Chandrasekar (chandra@iist.ac.in)

Received: 14 October 2015 – Revised: 19 January 2016 – Accepted: 20 January 2016 – Published: 9 February 2016

**Abstract.** The background error covariance structure influences a variational data assimilation system immensely. The simulation of a weather phenomenon like monsoon depression can hence be influenced by the background correlation information used in the analysis formulation. The Weather Research and Forecasting Model Data assimilation (WRFDA) system includes an option for formulating multivariate background correlations for its three-dimensional variational (3DVar) system (cv6 option). The impact of using such a formulation in the simulation of three monsoon depressions over India is investigated in this study. Analysis and forecast fields generated using this option are compared with those obtained using the default formulation for regional background error correlations (cv5) in WRFDA and with a base run without any assimilation. The model rainfall forecasts are compared with rainfall observations from the Tropical Rainfall Measurement Mission (TRMM) and the other model forecast fields are compared with a high-resolution analysis as well as with European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim reanalysis. The results of the study indicate that inclusion of additional correlation information in background error statistics has a moderate impact on the vertical profiles of relative humidity, moisture convergence, horizontal divergence and the temperature structure at the depression centre at the analysis time of the cv5/cv6 sensitivity experiments. Moderate improvements are seen in two of the three depressions investigated in this study. An improved thermodynamic and moisture structure at the initial time is expected to provide for improved rainfall simulation. The results of the study indicate that the skill scores of accumulated rainfall are somewhat better for the cv6 option as compared to the cv5 option for at least two of the three de-

pression cases studied, especially at the higher threshold levels. Considering the importance of utilising improved flow-dependent correlation structures for efficient data assimilation, the need for more studies on the impact of background error covariances is obvious.

**Keywords.** Meteorology and atmospheric dynamics (mesoscale meteorology; tropical meteorology)

## 1 Introduction

Numerical simulation of a weather phenomenon requires an efficient numerical model as well as an accurate initial state of the system. Data assimilation is the technique used to estimate the optimal state of the atmospheric system, utilising available meteorological observations as well as background information. The background and the observation information are weighted utilising their respective error covariances. Accurate specification of both observation and background errors is important for the success of a data assimilation system since these contribute to the improved analysis. The observation error covariance matrix  $\mathbf{R}$  is determined by instrument errors and representation errors, the latter arising from the effect of unresolved scales in a model as well as errors that arise from the specification of the observation operator, which maps model space variables to observation space. However, exact determination of a background error covariance (BEC) matrix is not practically possible, as the “truth” is not known, and hence is estimated in data assimilation systems using different methods. Furthermore, for a typical numerical weather model with a state space of dimen-

sion  $\sim 10^7$ , the BEC matrix has a dimension of  $\sim 10^7 \times 10^7$ , which cannot be explicitly represented in a computing system. Better estimates of the BEC matrix can, in principle, lead to more accurate analyses. Hence the study of different methods of determination of background errors has become an important research area, with substantial literature available (Bannister, 2008a, b; Berre, 2000; Buehner, 2005, 2010; Derber and Bouttier, 1999; Fisher, 2003; Katz et al., 2011; Rakesh and Goswami, 2011)

There are three methods for diagnosing background error (BE) covariance, namely (i) using innovation statistics (Hollingsworth and Lönnberg, 1986), (ii) the National Meteorological Center (NMC) method (Parrish and Derber, 1992) and (iii) the analysis-ensemble method (Fisher, 2003). Usage of innovation (observation minus background) statistics has the limitation that this method requires a good-quality, homogeneous observation network. Furthermore, this method provides estimates of background error for observed quantities only. The NMC method approximates the background error statistics using the difference between two model forecasts of different lead times that are valid at the same time. The analysis system is run several times using perturbed inputs in the analysis-ensemble method. The difference between background fields in the various runs provides a surrogate of background errors in this method.

Despite having its own shortcomings, the most commonly used method for generating the BEC is the NMC method. It is known that the NMC method underestimates the variance of background error over data-sparse regions. Also, the duration of forecasts used to generate BEC is typically 12 to 48 h. Since the background fields used in the assimilation system would typically be of a shorter duration, the covariances of estimated background error could become broader than the actual background error (Fisher, 2003). The popularity of the method, however, arises from the fact that this method yields global statistics for BEC for model variables at all levels. Moreover, it is less expensive to implement the NMC method in an operational weather prediction environment since the forecasts required to generate the differences would be already available.

The NMC method has been implemented to estimate the BEC in the Weather Research and Forecasting (WRF) model's variational data assimilation system (WRFDA) (Barker et al., 2004). This method employs a control variable transform (CVT) in which the model variables are converted into the control variables of the data assimilation system. This CVT renders the BEC matrix diagonal and also imposes balance relations within the variables. The manner in which the CVT is specified can impact the data assimilation results. In WRFDA3.5.1, there are three choices for CVTs, referred to as the cv3, cv5 and cv6 options. With cv3, the control variables are described in physical space, while the control variables are defined in the eigenvector space in the cv5 and cv6 options. While the cv3 option uses a vertical recursive filter to model the vertical covariance, both the cv5 and cv6 option

use an empirical orthogonal function (EOF) approach to represent the vertical covariance. Previous studies (e.g. Routray et al., 2014) have shown that the simulation of mesoscale weather phenomena like monsoon depressions over India can be improved using domain-specific BEC statistics calculated using the cv5 option rather than BEC utilising the cv3 option in WRF. However, comparisons of the impact of assimilating observations by three-dimensional variation (3DVar) analysis using the cv5 and cv6 options for simulating weather phenomena over the Indian region are not available in the literature. The present study is an effort in that direction which examines how the difference in the formulation of CVTs in BEC formulation influences the impact of assimilating meteorological observations in the WRF model.

The BEC statistics generated here using the NMC method are indeed homogeneous, isotropic and stationary. Flow-dependent BEC statistics generated using ensemble methods and ensemble-based data assimilation systems provide improved analysis (Buehner, 2005). However, generating and maintaining a large-sized ensemble that can adequately sample the model's uncertainty space is challenging and computationally expensive. This is especially true in a limited-area model, where uncertainties in the model's lateral boundary conditions are also significant.

The influence of the CVT's formulation in BEC modelling is demonstrated here in the case of three monsoon depressions (MDs) that formed over the Indian region. The amount of summer monsoon rainfall over central India depends on the formation and passage of these systems. MDs are one of the most significant synoptic-scale disturbances that occur during the Indian summer monsoon season of June–September. Typically 6–7 MDs form over the Indian region during this period each year. These systems, which have a horizontal radial extent of 1000 km, contribute to heavy to very heavy rainfall over India during their passage. Several previous studies (e.g. Routray et al., 2014; Govindankutty and Chandrasekar, 2010; Sinha and Chandrasekar, 2010; Chandrasekar and Kutty, 2013) have reported that 3DVar assimilation of observations have resulted in improved simulation of MDs using the WRF model. However, the inclusion of additional multivariate correlation information in the background error formulation in the cv6 option has not been investigated by any of the above-mentioned studies.

Section 2 provides further details on the cv5 and cv6 options available with WRFDA and Sect. 3 describes the depression cases investigated here. Experimental design details are provided in Sect. 4, with the results discussed in Sect. 5. Section 6 briefly summarises the conclusions drawn from the current study.

## 2 Background error covariance modelling

Let  $\mathbf{x} \in \mathbb{R}^n$  denote the state vector of the system. If the true state of the system is represented as  $\mathbf{x}^t$  and a background state is represented as  $\mathbf{x}^b$ , then the error in the background state can be represented as  $\epsilon^b = \mathbf{x}^t - \mathbf{x}^b$ . It is assumed that the background error is unbiased, i.e.  $\langle \epsilon^b \rangle = 0$  where the angular brackets denote the expectation value. The 3DVar analysis is obtained by minimising a cost function defined as

$$\begin{aligned} J(\mathbf{x}) &= J_b + J_o \\ &= \frac{1}{2}[(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \\ &= (\mathbf{y} - H(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y} - H(\mathbf{x}))]. \end{aligned} \quad (1)$$

The analysis  $\mathbf{x} = \mathbf{x}^a$  represents a minimum variance estimate of  $\mathbf{x}^t$  given the observations  $\mathbf{y} \in \mathbb{R}^m$  as well as the error covariances of background and observations denoted by  $\mathbb{B}$  and  $\mathbb{R}$  respectively. The observation operator  $H$  provides a mapping from the model's grid space to the observation space.

The calculation of the background term  $J_b$  requires  $\sim O(n^2)$  calculations for a system with  $n$  degrees of freedom. For a numerical weather model with typically  $10^7$  degrees of freedom, the direct calculation of this term is not possible. To reduce the computational cost,  $J_b$  is calculated in terms of control variables  $\mathbf{v}$  defined via the relation  $\mathbf{x}' = \mathbf{U}\mathbf{v}$ , where  $\mathbf{x}'$  denotes the analysis increment,  $\mathbf{x}' = \mathbf{x} - \mathbf{x}^b$ . Using the incremental formulation (Courtier et al., 1994; Barker et al., 2004) and the control variables, Eq. (1) can be rewritten as

$$J(\mathbf{v}) = J_b + J_o = \frac{1}{2}\mathbf{v}^T \mathbf{v} + \frac{1}{2}(\mathbf{y} - H\mathbf{U}\mathbf{v})^T \mathbf{R}^{-1}(\mathbf{y} - H\mathbf{U}\mathbf{v}). \quad (2)$$

The transformation matrix  $\mathbf{U}$  is defined in such a way that the background error matrix can be represented as  $\mathbf{U}\mathbf{U}^T$ . In WRF 3DVar system, the CVT is implemented in three steps – a horizontal transform,  $\mathbf{U}_h$ ; a vertical transform,  $\mathbf{U}_v$ ; and a parameter transform,  $\mathbf{U}_p$  (Barker et al., 2004).

$$\text{i.e. } \mathbf{x}' = \mathbf{U}_h \mathbf{U}_v \mathbf{U}_p \mathbf{v} \quad (3)$$

The CVT aims to convert the BEC matrix into block-diagonal form. The horizontal transform  $\mathbf{U}_h$  is represented using recursive filters (Hayden and Purser, 1995; Purser et al., 2003) in WRF 3DVar. There are two free parameters associated with each variable for the recursive filter – the number of applications of the filter and the correlation length scale of the filter. The correlation length scale is estimated for each variable and vertical mode using the NMC method's accumulated forecast difference data processed as a function of grid-point separation (Barker et al., 2004). A tuning factor is applied to the length scale in order to reflect the actual correlation length scales in a domain.

In the vertical transform,  $\mathbf{U}_v$ , an empirical orthogonal function decomposition is performed on the vertical component of the background error,  $\mathbf{B}_v$ . The analysis increments

are projected onto the eigenvector space and the eigenvalues specify the relative weights of increments in the calculation of cost function. The parameter transform,  $\mathbf{U}_p$ , is applied so that the errors in the control variables are not correlated with each other and the BEC matrix is rendered block-diagonal. The increments in model variables  $u$  (zonal wind),  $v$  (meridional wind),  $T$  (temperature),  $p$  (pressure) and humidity ( $q$ ) are converted into a new set of variables such as stream function ( $\psi$ ), velocity potential ( $\chi$ ), temperature ( $T$ ), surface pressure (ps) and relative humidity (rh). The WRF-Var system provides the balance relations between the new set of variables using regression relations. After the “balanced part” of analysis variables is estimated, the “unbalanced part” is determined by subtracting the former from the full fields. Hence, while some fields are analysed in full, for some other variables the unbalanced parts are included in the analysis system. The control variable options cv5 and cv6 differ from each other in the specification of the balance relations between these control variables.

In the cv5 option, the analysis variables consist of the full fields corresponding to stream function and relative humidity and the unbalanced parts corresponding to the other variables included in the analysis. However, in the cv6 option, only stream function is analysed in full, while relative humidity and other variables comprise both balanced and unbalanced parts.

The control variables specified in the cv5 option are related as given below,

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

$$T_u(i, j, k) = T(i, j, k) - \sum_{l=1}^{N_k} \alpha_{\psi T}(i, j, k, l)\psi(i, j, l), \quad (4b)$$

$$ps_u(i, j) = ps(i, j) - \sum_{l=1}^{N_k} \alpha_{\psi ps}(i, j, l)\psi(i, j, l). \quad (4c)$$

Here,  $i$  and  $j$  denote the horizontal dimension index,  $k$  and  $l$  indicate the vertical sigma levels, and  $\alpha$  represents the various regression coefficients between the variables represented using the respective subscripts. The unbalanced parts of the fields are denoted using the subscript  $u$ .

The relations specified by Eq. (4) indicate the manner in which the various analysis control variables are related in the WRF-Var system. Here, the balanced part of velocity potential is related to the stream function alone and hence the possible relations that the divergent component of wind could have with other variables are not properly considered. Similarly, in cv5, temperature and surface pressure are neither related to each other nor to the moisture variable. Relative humidity field is not influenced by any of the model variables like temperature or wind, as per the relations in Eq. (4).

The balance relations defined in the cv6 option, on the other hand, include additional correlations between model

variables. The cv6 option has the following balance relations:

$$\chi_u(i, j, k) = \chi(i, j, k) - \alpha_{\psi\chi}(i, j, k)\psi(i, j, k) \quad (5a)$$

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

$$- \sum_{l=1}^{N_k} \alpha_{\chi_u T}(i, j, k, l)\chi_u(i, j, l) \quad (5b)$$

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

$$- \sum_{l=1}^{N_k} \alpha_{\chi_u ps}(i, j, l)\chi_u(i, j, l) \quad (5c)$$

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

$$- \sum_{l=1}^{N_k} \alpha_{\chi_u rh}(i, j, k, l)\chi_u(i, j, l)$$

$$- \sum_{l=1}^{N_k} \alpha_{T_u rh}(i, j, k, l)T_u(i, j, l)$$

$$- \sum_{l=1}^{N_k} \alpha_{ps_u rh}(i, j, k, l)ps_u(i, j, l). \quad (5d)$$

Here, additional correlation coefficients connect model variables in more ways than are available in the cv5 option. For example, the balanced parts of temperature and surface pressure are now also correlated with the unbalanced velocity potential. Hence, temperature and surface pressure are influenced by the divergent component of wind in the cv6 option unlike in the cv5 option. Similarly, additional correlations defined in the moisture variable make the moisture analysis multivariate in nature in the cv6 option (Chen et al., 2013).

In both the cv5 and cv6 option, the estimation of BECs proceed in the following five stages:

1. Calculation of standard perturbations from forecast differences. In the NMC method, this is calculated as  $\mathbf{x}' = \mathbf{x}_{T2} - \mathbf{x}_{T1}$ , where  $\mathbf{x}_{T2}$  and  $\mathbf{x}_{T1}$  are forecast difference times (e.g. 48 and 24 h for global, 24 and 12 h for regional).
2. Time/bin mean for each variable/level is removed so that zero-mean fields are obtained.
3. Regression analysis is performed and the various correlations are determined between the control variable fields. The unbalanced components of the fields are calculated.
4. The vertical component of the CVT is applied.
5. Recursive filter is used to provide horizontal correlations.

The cv5 and cv6 options in WRF-Var differ from each other in the step (iii) mentioned above.

### 3 Case description

#### 3.1 Case 1 – depression (29–31 May 2013)

Under the influence of a cyclonic circulation over the north-western Bay of Bengal off the Odisha–West Bengal coasts, a low-pressure area formed over the north of the Bay of Bengal and neighbouring areas. The system intensified into a depression and was centred near 21° N, 89.5° E, at 03:00 UTC on 29 May 2013. Moving north-northwestwards, the depression laid centred near 21.7° N, 88.8° E at 12:00 UTC on 29 May and crossed the West Bengal coast near 21.8° N, 88.7° E between 13:30 and 14:30 UTC on the same day. The above-mentioned depression caused heavy to very heavy rainfall during its passage over the Indian region.

#### 3.2 Case 2 – land depression (20–24 July 2014)

A land depression formed and laid centred over northeastern parts of Odisha and adjacent areas of Gangetic West Bengal on the morning of 21 July 2014. The depression moved west-northwestwards and weakened into a well-marked low-pressure area over western Madhya Pradesh and neighbouring areas on 23 July 2014.

#### 3.3 Case 3 – depression (29 July–1 August 2013)

A low-pressure area formed over northeast Bay of Bengal and adjacent Bangladesh and coastal areas of West Bengal on 29 July 2013. The system intensified into a Depression and laid centred near 21° N, 88° E. The depression moved north-westwards and crossed the coast of Odisha at 12:00 UTC on 30 July. The depression further moved westwards and week-end into a low-pressure system by 1 August 2013.

### 4 Experimental

The Advanced Research WRF (ARW) model version 3.5.1 is configured with two-way nested domains as shown in Fig. 1. The outer domain has 350 × 350 grid cells in the east–west and north–south directions with a horizontal resolution of 27 km, while the inner domain is configured with 451 × 451 grid points with a horizontal resolution of 9 km. There are 36 vertical levels in both domains. Physics parametrisation schemes utilised in this study are the same as in Routray et al. (2014) for simulating monsoon depressions. Further details on WRF model formulation and implementation are available in Skamarock et al. (2008) and Barker et al. (2004). All results discussed in this study are from the inner, higher-resolution domain.

The initial and boundary conditions for the model are obtained from the National Center for Environmental Predic-



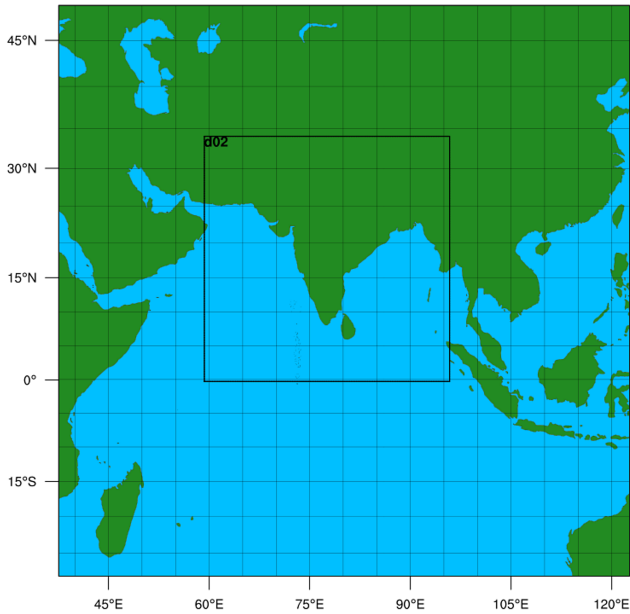


Figure 1. Model domains used in the study.

tion (NCEP) Global Forecast System (GFS) forecast fields of horizontal resolution  $0.5^\circ \times 0.5^\circ$ . Here, data assimilation is performed in the outer domain only. The inner domain has been initialised from the outer, coarse-resolution domain. For determining BEC using the NMC method, 12 and 24 h forecasts are generated for a period of 1 month during the 2013 summer monsoon period (1 June to 1 July 2013) for the outer domain. In the WRF 3DVar system, the forecast differences of 24 and 12 h forecast times are typically used for calculating background error using NMC method in the regional domain (Skamarock et al., 2008). The forecast perturbations generated using the differences of these forecasts are used for calculating BECs using both the cv5 and cv6 option. The conventional surface and upper air observations from the NCAR Computational and Information Systems Laboratory (CISL) Research Data Archive (<http://rda.ucar.edu/datasets/ds337.0>) are utilised here for assimilation. Also, Advanced Microwave Sounding Unit-A (AMSU-A) radiances (Baker et al., 2005; Singh et al., 2012) are assimilated in this study. In all three depression cases investigated here, the WRF model is integrated for 12 h without any assimilation for model spin-up. Using this model forecast as a background, 3DVar analysis is performed using both the cv5 and cv6 option once, with a specified observation window of  $\pm 3$  h around the analysis time, followed by free forecast (no assimilation of observations) for the next 48 h. Figure 2 shows the all observations typically available in the domain at the assimilation time. In order to bring out clearly the effect of data assimilation, a control (CTRL) run is performed with no assimilation of observations.

Assimilation of a single observation is a convenient and useful way to diagnose the properties of the BEC matrix. Let

us suppose that we have the observation of the  $k$ th element of the state  $\mathbf{x}$ . That is, here the observation is a scalar. Let the observation error be  $\sigma^2$  and the single observation be denoted as  $y$ . The observation operator  $H$  and its Jacobian  $\mathbf{H}$  become row vectors of zeroes apart from their  $k$ th element, which becomes 1. Calculating the gradient of  $J$  in Eq. (1) and rearranging the terms, we get the analysis increment at an element  $l$  as

$$\mathbf{x}_l^a - \mathbf{x}_l^b = \mathbf{B}_{lk} \frac{y - \mathbf{x}_k^b}{\mathbf{B}_{kk} + \sigma^2}. \quad (6)$$

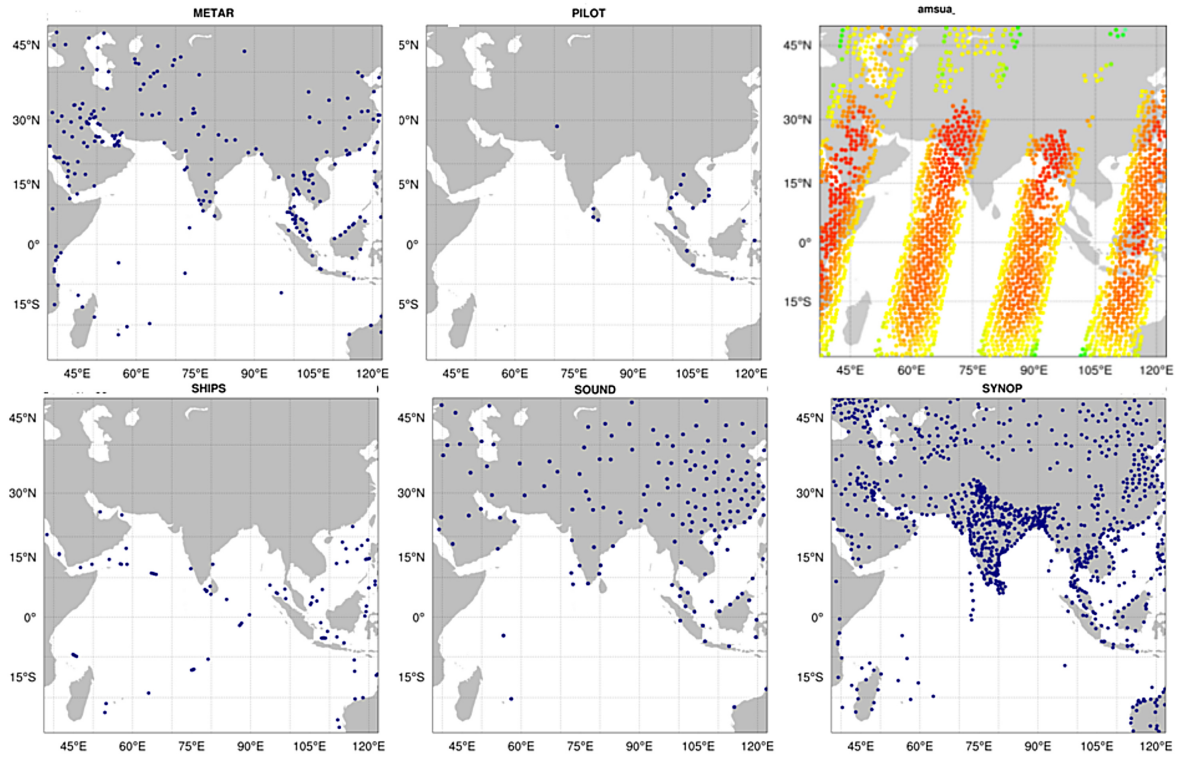
That is, the analysis increment is proportional to the element  $\mathbf{B}_{lk}$  of the BEC matrix, which denotes the covariance between elements at  $k$  and  $l$ . Hence, assimilating a single observation provides information regarding the structure of the BEC matrix. To investigate the impact of the formulation of control variables of BEC matrix, several single-observation assimilations have been performed.

The European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim reanalysis data (Dee et al., 2011) available at a horizontal resolution of  $0.125^\circ \times 0.125^\circ$  are used to evaluate the model forecasts here. Furthermore, a high-resolution analysis is also utilised here. The high-resolution analysis has been determined as follows. The GFS global analysis and all the observations (conventional and AMSU radiances) over the model domain are subjected to the 3DVar methodology using the cv5 option, throughout the model forecast period, with assimilation performed every 6 h. This uses the same horizontal and vertical resolution of the model grids as indicated in Sect. 4.

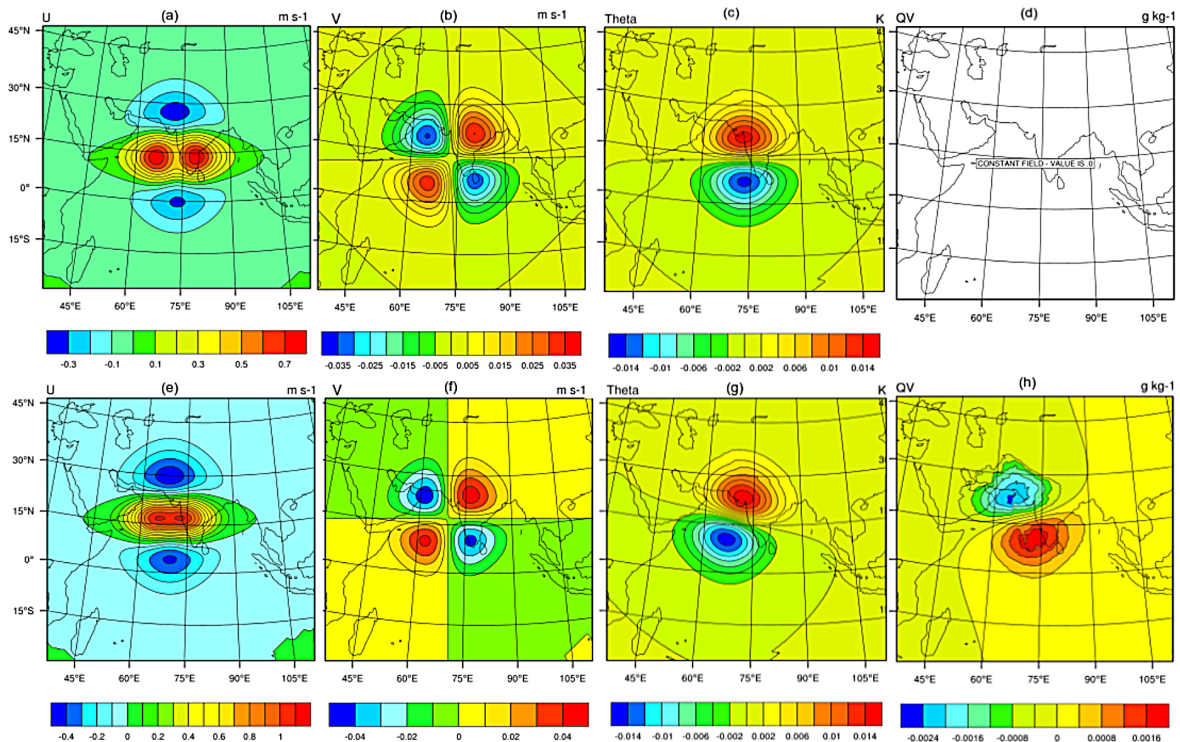
## 5 Results and discussion

### 5.1 Single-observation experiment

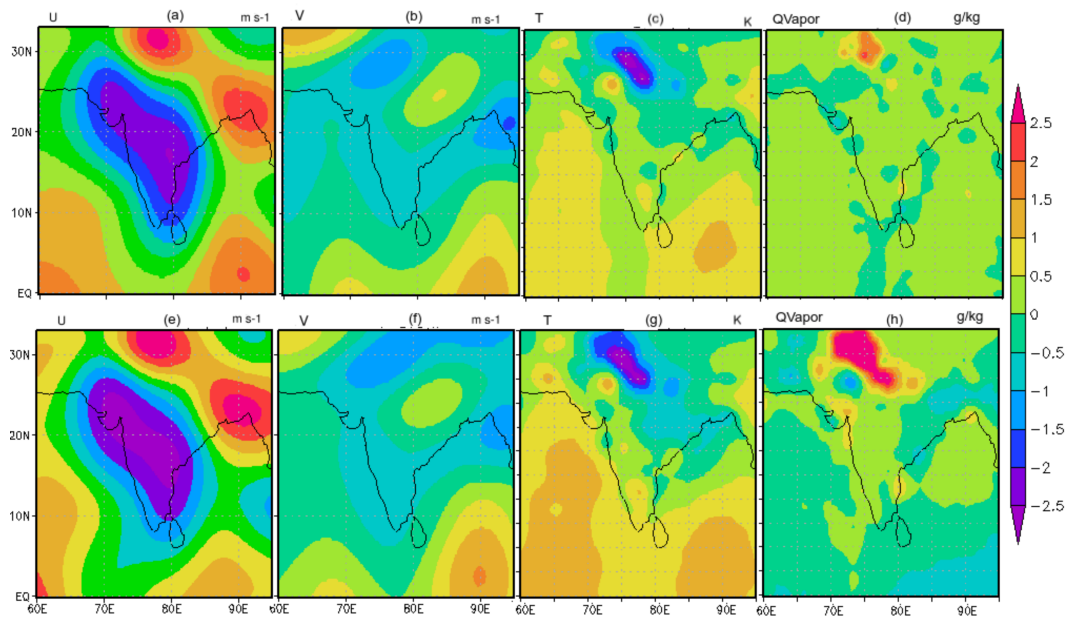
Figure 3 shows the analysis increments in  $u$ ,  $v$ ,  $\theta$  and  $q$  when a single  $u$  wind observation is assimilated at the middle of the domain at model sigma level 19 (same model level as in Routray et al., 2014). The  $u$ -wind observation which differs from the background with a magnitude of  $1 \text{ m s}^{-1}$  is assimilated using both the cv5 and cv6 option of the BEC matrix. Figure 3a–d provide the analysis increment in  $u$ ,  $v$ ,  $\theta$  and  $q$  respectively for the cv5 option, while Fig. 3e–h show the same for the cv6 options. The analysis increment patterns for the cv5 option is similar to that obtained by Routray et al. (2014) for the regional BE option, with a double maxima over the ocean region in the  $C(u, u)$  pattern (Fig. 3a and e). The inclusion of additional correlation functions in cv6 has resulted in a non-vanishing increment in moisture field on assimilation of wind information. Furthermore, the additional correlation coefficient  $\alpha_{\chi_u T}$  in the expression for  $T_u$  has resulted in a changed temperature field (Fig. 3g) for the cv6 option. Also, the location of the maximum impact of the observation is shifted in the cv6 run as compared with that



**Figure 2.** Observations available typically over the domain.



**Figure 3.** Analysis increment for zonal wind (a, e), meridional wind (b, f), potential temperature (c, g) and water vapour mixing ratio (d, h) when a single  $u$ -wind observation is assimilated at the middle of the domain at model level 19. Panels (a–d) are for the cv5 run and (e–h) for the cv6 run.



**Figure 4.** Analysis increment at model level 1 in zonal wind (a, e) meridional wind (b, f), temperature (c, g) and water vapour mixing ratio (d, h) for the cv5 option (a–d) and the cv6 option (e–h) for the first depression case due to assimilation of all available observations.

in the cv5 run. Since the covariance matrix is symmetric, assimilation of a single temperature observation will result in a relative rotation of increment pattern in the  $u$  field (Chen et al., 2013). Hence, the single-observation experiments reveal that the difference in formulation of control variables has resulted in different analysis increment structures in the cv6 option as compared to that in the cv5 option.

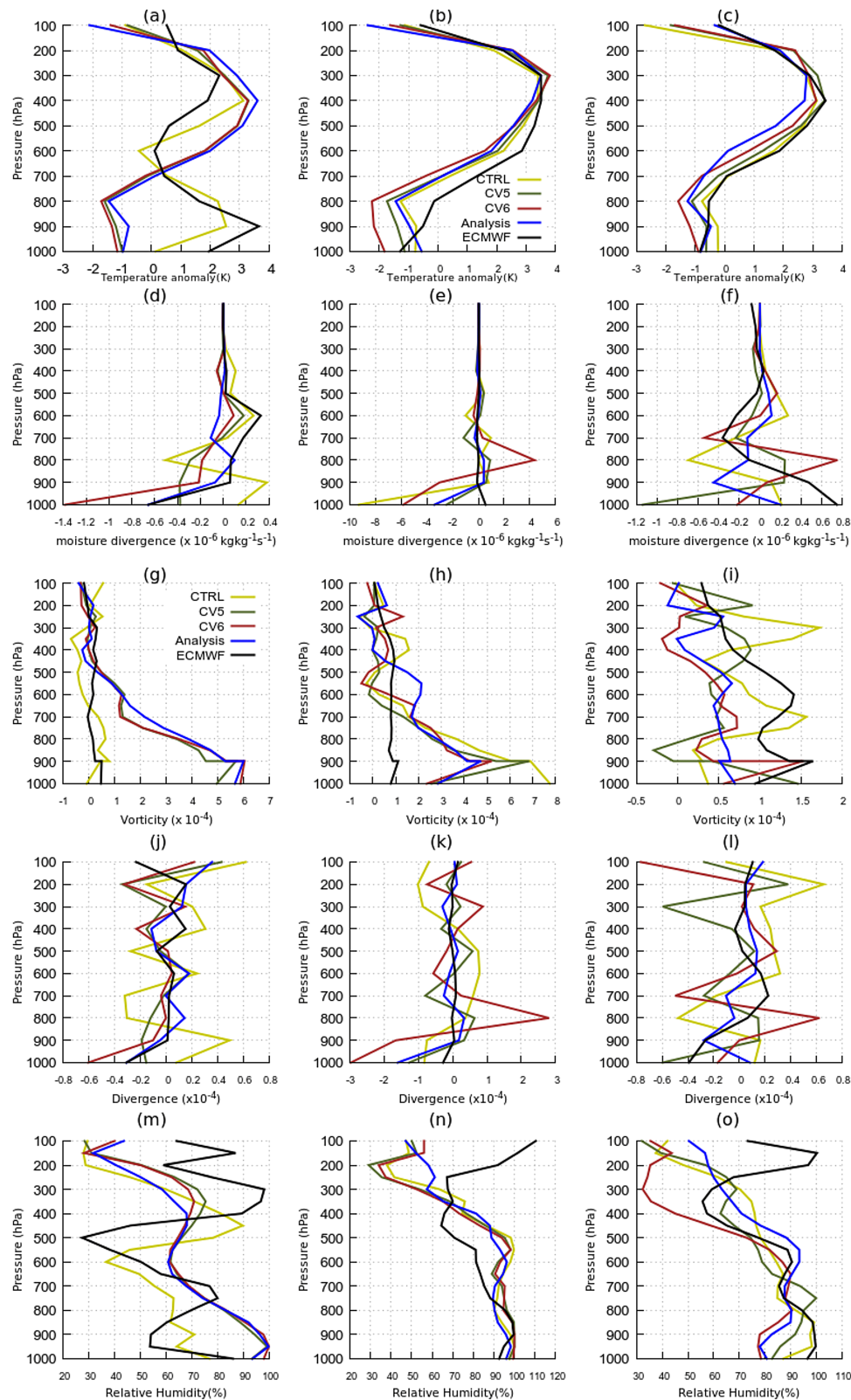
## 5.2 Assimilation results

Figure 4 shows the analysis increment for the cv5 and cv6 options at the model level 1 for zonal wind (a and e), meridional wind (b and f), temperature (c and g) and humidity (d and h) with all observations assimilated for the case 1 depression. Assimilation of all available observations (conventional and AMSU-A radiances) has yielded different impacts in the cv5 and cv6 cases as seen from the differing analysis increment patterns. The increments in  $u$ -wind shows a larger positive region over the oceans in the cv5 run as compared with the cv6 run. This indicates that the strength of the zonal wind has increased due to assimilation in the cv5 run as compared with the cv6 run. While the increment patterns in meridional winds and temperature at the surface level are fairly similar, assimilation of all observations has impacted the water vapour mixing ratio differently in the cv5 and cv6 cases. Additional correlation information in the BEC matrix formulation has resulted in larger increments in the cv6 option as compared with the cv5 option. The above results indicate that differing formulation of BEC matrix indeed impacts the assimilation of observations in the WRF 3DVar system.

The impact of assimilation on the structure and movement of the depressions is investigated by estimating the depression track errors for a 24 h forecast period with respect to India Meteorological Department (IMD) track reports. The track errors for the CTRL, cv5 and cv6 runs are shown in Table 1. In general, both the cv5 and cv6 run perform better than the CTRL run in terms of lower track errors. Here, the cv5 run has lower track errors as compared to the cv6 run. From Fig. 4 we can infer that the analysis increments between the cv5 and cv6 options are significantly different for the water vapour mixing ratio only. Therefore, it may be unrealistic to expect that the simulation of mean sea level pressure and hence the track errors will be improved by utilising the cv6 option instead of cv5.

Monsoon depressions are known to have cold core lows at the lower levels and warm core highs at the upper levels. It is known that systems with a cold core at lower levels typically intensify with increasing height, indicating that the monsoon depression has higher intensity at the lower to middle troposphere (Godbole, 1977; Sikka, 1977; Stano et al., 2002; Sørland and Sorteberg, 2015). To investigate how the various experiments have simulated this observed thermal structure of depressions, the profile of the temperature anomaly has been calculated over the depression centre at the initial time. For this a  $3^\circ \times 3^\circ$  box has been considered around the depression centre. The difference in average temperature in this box to that outside it is calculated. Figure 5a, b, and c show the vertical profile of the temperature anomaly at the analysis time at the depression centre for the case 1, case 2 and case 3 depressions respectively. All assimilation runs have





**Figure 5.** Vertical profiles of temperature anomaly (a, b, c), moisture divergence (d, e, f), relative vorticity (g, h, i), horizontal divergence (j, k, l) and relative humidity (m, n, o) at the analysis time of the cv5/cv6 sensitivity experiments over the depression centre for case 1 (left column), case 2 (middle column) and case 3 (right column) depressions.

**Table 1.** Track error in kilometres for the three depression cases with respect to IMD data.

Forecast hour	Case 1			Case 2			Case 3		
	CTRL	cv5	cv6	CTRL	cv5	cv6	CTRL	cv5	cv6
00:00	414.9	73.14	82.2	137.1	110.0	110.0	230.9	41.3	139.6
06:00	320.4	113.5	122.5	74.9	19.8	22.9	422.1	494.4	512.5
12:00	165.3	123.9	132.9	257.4	239.3	293.6	360.3	360.3	378.3
18:00	120.8	159.3	168.4	368.6	434.5	440.8	449.9	391.5	499.9
24:00	118.04	253.2	271.3	515.9	390.7	506.9	675.9	594.5	543.0

simulated the thermal structure of the depression well with respect to the high-resolution analysis as compared with the CTRL run, in all three cases. However, as compared with the ECMWF reanalysis, the simulated temperature anomalies have stronger cold cores at lower levels. For the case 1 depression, the ECMWF reanalysis reveals a system whose cold core, which is indicative of its intensity, weakens with height at the lower levels. This is different from the typically observed thermodynamic structure of the monsoon depressions. In all three depression cases, the cv6 runs have simulated a relatively stronger cold core at the lower levels, as compared to the CTRL and cv5 runs as well as the ECMWF and high-resolution analyses. This indicates that the thermodynamic structure of the monsoon depressions is modified by the cv6 analysis.

Figure 5d–f show the vertical profile of moisture divergence at the centre of the depression cases at the initial time of the forecast (+0 h forecast). Figure 5d shows that the cv6 run has contributed to higher moisture convergence values that are closer to the high-resolution analysis in the lower levels at the depression centre as compared with both the CTRL and cv5 run for the first depression case. In the second depression case shown in Fig. 5e, the cv6 run shows higher moisture convergence value close to the surface and relatively higher divergence values in the higher levels as compared with the cv5 run. Figure 5f, which shows the third depression case, indicates that surface moisture convergence values of cv6 are closer to the high-resolution analysis. The cv6 run simulated moisture convergence is closer to the ECMWF reanalysis as compared to the cv5 run at low levels for the first depression case.

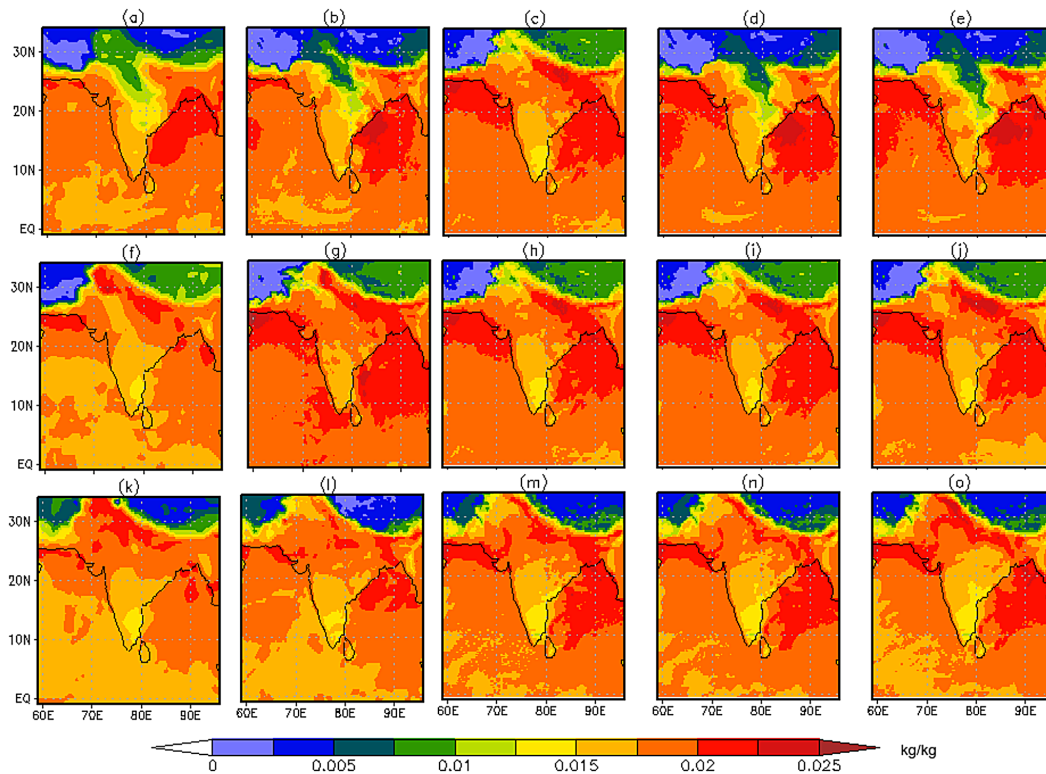
Figure 5g–i show the vertical profile of relative vorticity at the centre of the depression for all the depression cases at analysis time of the cv5/cv6 sensitivity experiments (+0 h forecast). For the second and third depression cases, at lower levels, the relative vorticity values for the cv6 runs are closer to the high-resolution analysis as compared to the cv5 and CTRL runs. The difference in relative vorticity profiles between the cv5 and cv6 options is not marked for all depression cases. However, the CTRL run's results for the first depression are very far from the analysis values at all levels. ECMWF reanalysis of the relative vorticity profile for the

first depression case also shows a weaker vortex at all the levels.

Vertical profiles of horizontal divergence over the depression centre at analysis time of the cv5/cv6 sensitivity experiments (+0 h forecast) for all three depression cases are shown in Fig. 5j–l respectively. For the first and third depression cases considered here, the cv6 option has simulated stronger convergence at the lower levels than the cv5 option. In the first case the simulated profile is closer to the ECMWF reanalysis.

Figure 5m–o show the vertical profiles of relative humidity at analysis time of the cv5/cv6 sensitivity experiments over the depression centre for all three depression cases. For the first depression case, the CTRL run simulates relatively drier atmosphere until the mid-troposphere, while the assimilation runs are closer to the high-resolution analysis and are moister. ECMWF reanalysis also indicates a drier atmosphere. This is consistent with the warmer core at around 900 hPa present in the ECMWF reanalysis. The ECMWF reanalysis also shows large relative humidity values for all depressions at the upper levels. For the first and second depression cases, the cv6 option simulates higher values of relative humidity at lower levels as compared to the high-resolution analysis as well as the CTRL and cv5 runs. In the case of the third depression, despite having lower relative humidity values at the initial time at lower levels, relative humidity simulated by the cv6 run is closer to the high-resolution analysis and the ECMWF reanalysis as compared to the cv5 and CTRL runs, especially in the middle levels.

The differences between the vertical profiles between the cv5 and cv6 runs with respect to the CTRL run at each model level at the beginning of the free forecast period are calculated. To estimate whether the assimilation has significantly altered the model fields, a Student *t* test (Wilks, 2011) has been performed by considering these differences. It is found that the differences in the vertical profiles of moisture convergence and relative vorticity between the reference (CTRL) run and the cv5/cv6 analyses are significant at the 95 % confidence level. The differences in relative humidity are found to be 95 % significant over the first 15 model levels only. However, the divergence profiles are found to be statistically significant at the 65 % level only.



**Figure 6.** Water vapour mixing ratio at level 1 compared with the ECMWF reanalysis at the analysis time of the cv5/cv6 sensitivity experiments at level 1 for case 1 (a–e), case 2 (f–j) and case 3 (k–o) depressions for ECMWF reanalysis (a, f, k), high-resolution analysis (b, g, l), CTRL run (c, h, m), cv5 run (d, i, n) and cv6 run (e, j, o).

Comparison of vertical profiles at the depression centre at analysis time of the cv5/cv6 sensitivity experiments shows that, in comparison with the cv5 option, the cv6 option has simulated the following:

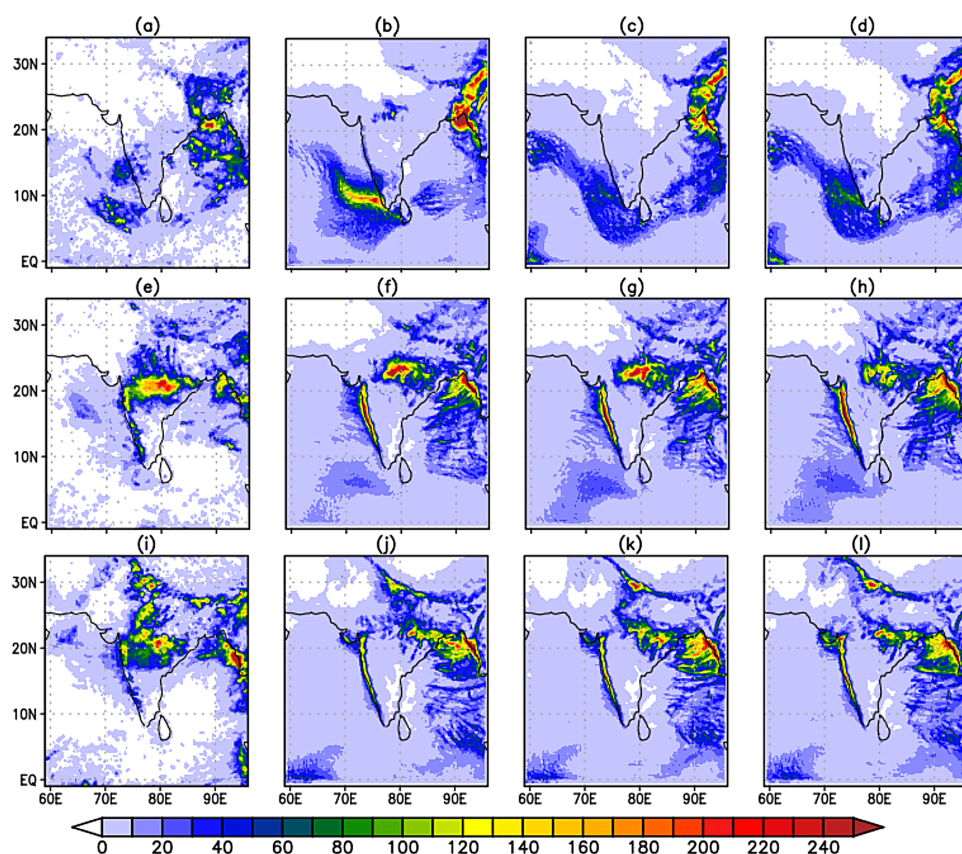
1. Stronger cold core low at lower levels (up to 800 hPa) for all three depression cases.
2. Stronger low-level moisture convergence for two of the three cases at the initial time.
3. Stronger low-level horizontal wind convergence for two of the three depression cases.
4. Higher values of relative humidity at lower levels in two out of three depression cases.

Considering the above, it is clear that utilising the cv6 option has modified the vertical structure at the depression centre at the initial time in two out of three cases considered here, resulting in higher relative humidity values, higher low-level moisture convergence and higher low-level horizontal convergence values. However, to infer definite and broad conclusions, more depression cases have to be investigated.

Figure 6 shows the surface level water vapour mixing ratio of the CTRL (c, h, m), cv5 (d, i, n) and cv6 run (e, j, o)

compared with the ECMWF reanalysis (a, f, k) and the high-resolution analysis (b, g, l) for all three depressions. In general, all the model simulations indicate more moisture than the ECMWF reanalysis. The cv6 run has simulated moderately higher mixing ratio values as compared with the CTRL and cv5 runs over the land region in two out of the three cases considered. With more moisture being simulated over land, both in spatial and vertical profiles along with larger horizontal moisture convergence, the cv6 run is expected to simulate rainfall better than the cv5 and CTRL runs.

The model-simulated 48 h accumulated rainfall is compared with Tropical Rainfall Measurement Mission (TRMM) rainfall observations to analyse the impact of BEC formulation on rainfall simulation in Fig. 7. The accumulated precipitation from TRMM and the CTRL, cv5 and cv6 runs for the three depressions considered here is shown in Fig. 7a–l. The top row (Fig. 7a–d) shows the first depression case, the middle row (Fig. 7e–h) shows the second depression case and the bottom row (Fig. 7i–l) shows the third depression case considered in this study. For all three cases, accumulated rainfall values indicate heavy rainfall over the Indian region due to the monsoon depressions. In the first depression case the model simulates excess rainfall over the head of the Bay of Bengal and northeastern regions of India. The TRMM observations show that the location of maximum precipitation is



**Figure 7.** 48 h accumulated precipitation for case 1 (a–d), case 2 (e–h) and case 3 (i–l) depressions from TRMM (a, e, i), CTRL run (b, f, j), cv5 run (c, g, k) and cv6 run (d, h, l).

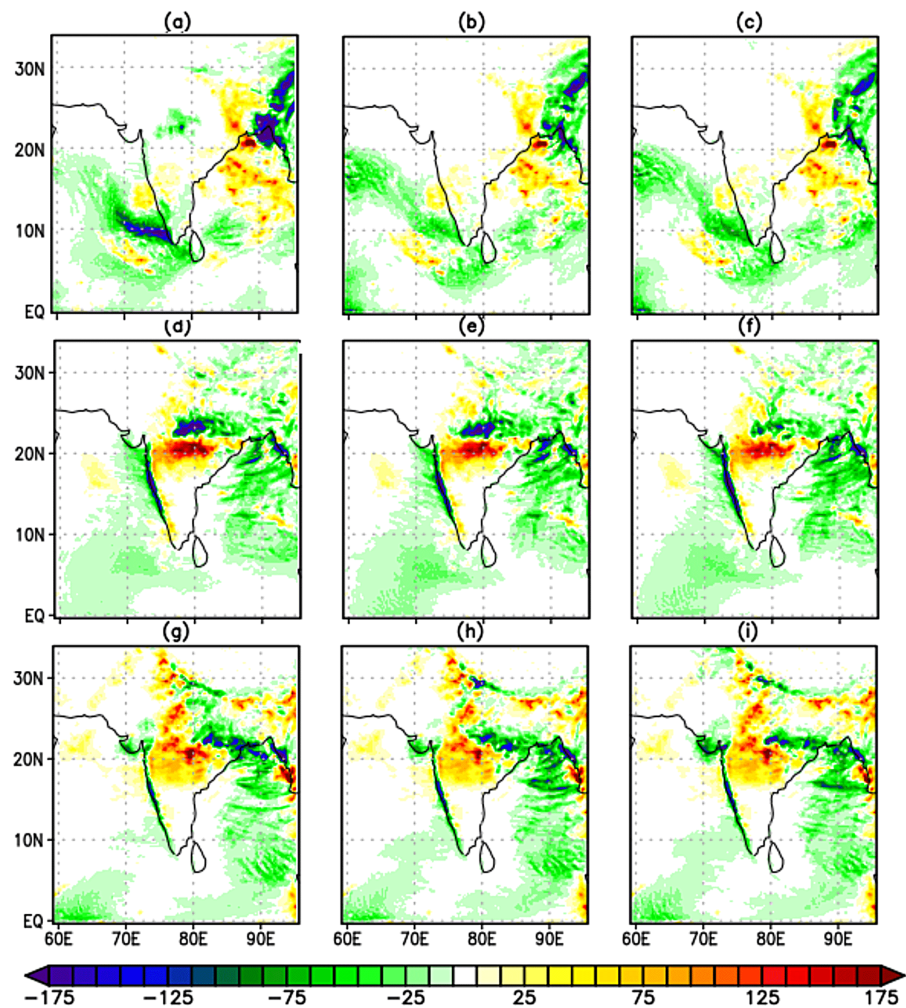
over the head of the bay. Furthermore, the CTRL run simulates erroneous rainfall maximum off the west coast of India, which is not seen in either the TRMM observations or assimilated runs. All the model runs show similar spatial pattern of precipitation for the second depression case. However, the intensity of rainfall is different in the cv6 run as compared with the CTRL and the cv5 runs. For the third depression case, the observed rainfall maximum over oceans is reproduced by all the model runs in terms of location and intensity. However, both the cv5 and cv6 run simulate rainfall closer to TRMM observations as compared to the CTRL run. The differences in 48 h accumulated rainfall simulated by the three model runs with respect to the TRMM observation are shown in Fig. 8. For all the depression cases, the maximum differences between TRMM observation and model simulations in the CTRL, cv5 and cv6 runs occur over the regions where TRMM observations show maximum accumulated rainfall.

Table 2 gives the location and magnitude of maximum 48 h accumulated rainfall from TRMM observations as well as the location error (in km) and magnitude error (in cm) in the model simulations of the same for the CTRL, cv5 and cv6 runs. The results show that in all the depression cases considered here, the model-simulated location and intensity

of maximum accumulated rainfall has errors when compared with TRMM observations. It is also found that the cv6 run has lower error in the location as well as magnitude of maximum rainfall when compared with both the CTRL and cv5 run in two out of three depression cases.

For quantitative verification of rainfall forecast skill, various skill scores like equitable threat score (ETS), bias score, false alarm ratio (FAR) and probability of detection (POD) have been calculated for 48 h accumulated rainfall with respect to TRMM observations for all three depression cases and are shown in Fig. 9. Towards this, model-simulated 48 h accumulated rainfall was regridded to the resolution of TRMM rainfall observations. The skill scores are calculated using the contingency table considering whether a forecast occurs or not (Wilks, 2011). The ETS estimates how well the observed event is forecast, taking into account the correct forecasts that can occur by chance. Bias score estimates the ratio of frequency of forecast events to the frequency of observed events, indicating whether there is over- or under-prediction by the model. The FAR gives the fraction of false alarms (model-simulated rainfall that is not observed) and POD gives the fraction of correctly forecast events. In general, assimilated runs perform better than the CTRL run in



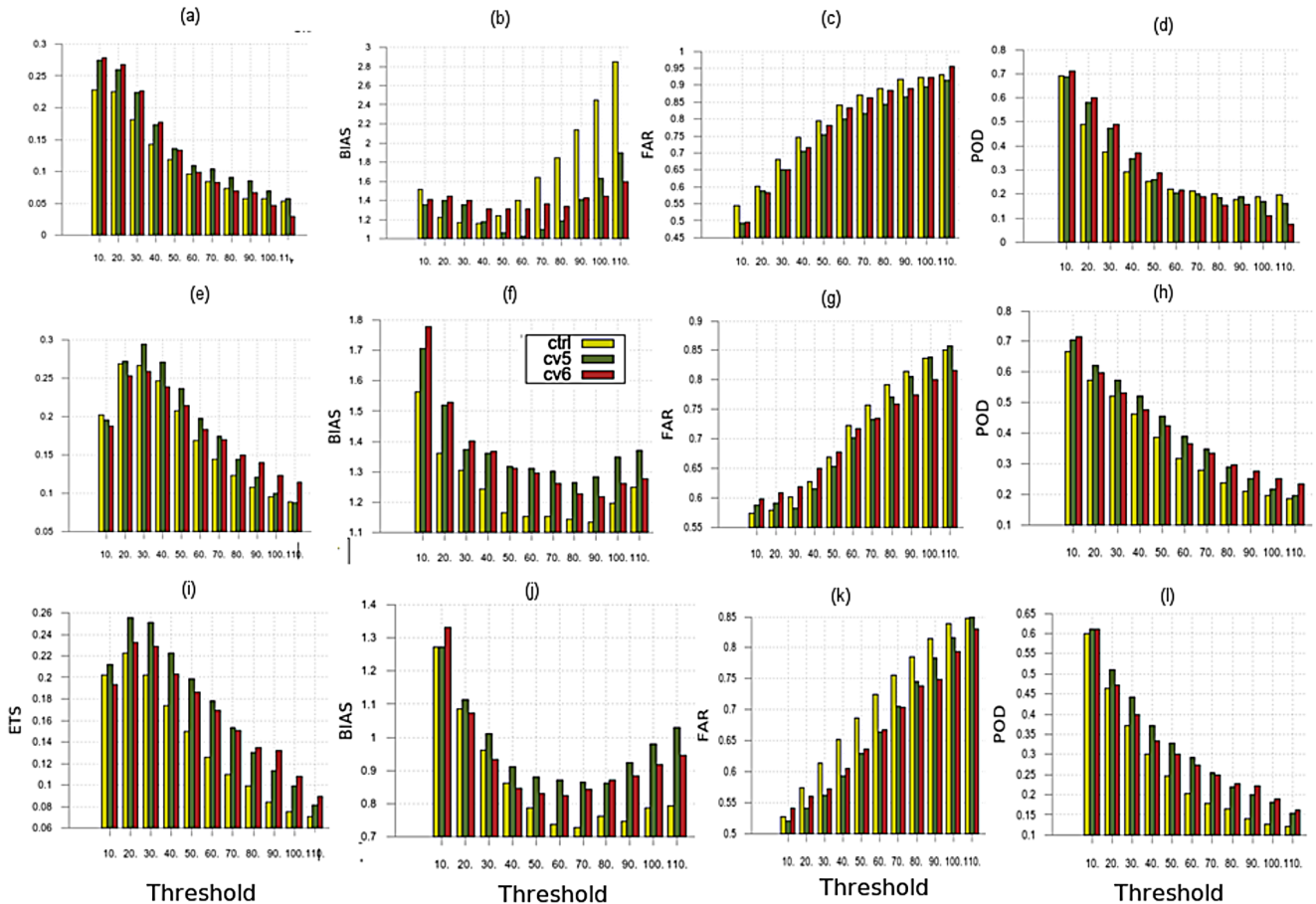


**Figure 8.** Differences in 48 h accumulated precipitation (in mm) from TRMM rainfall observation for case 1 (a–c), case 2 (d–f) and case 3 (g–i) for the CTRL (a, d, g), cv5 (b, e, h) and cv6 (c, f, i) runs.

**Table 2.** Location and intensity of maximum precipitation (cm) as obtained from TRMM for all three depressions together with location and magnitude error of maximum precipitation for the CTRL, cv5 and cv6 runs.

	TRMM		CTRL		cv5		cv6	
	location of max. rainfall	magnitude (cm) of max. rainfall	location error (km) of max. precipitation	magnitude error (cm) of max. precipitation	location error (km) of max. precipitation	magnitude error (cm) of max. precipitation	location error (km) of max. precipitation	magnitude error (cm) of max. precipitation
Depression 1	89.13° E, 20.88° N	35.7	1173	−11.8	257	−4.9	146	−4.4
Depression 2	80° E, 21.25° N	24.1	218	−21.6	211	−11.4	183	−2.1
Depression 3	79.65° E, 20.88° N	29.9	1325	−13.5	937	−15.2	943	−13.3





**Figure 9.** Skill scores of 48 h accumulated precipitation (mm) with regard to TRMM for case 1 (a–d), case 2 (e–h) and case 3 (i–l) depressions. Equitable threat score is shown in (a, e, i). Bias scores (b, f, j), false alarm ratio (c, g, k) and probability of detection (d, h, l).

all rainfall thresholds with higher ETS values, lower FAR and higher POD in all three cases. In the second and third depression cases, the cv6 runs have higher ETS score, lower bias, lower FAR and larger POD as compared with the cv5 and CTRL runs in the higher threshold regions. The higher ETS score, higher POD and lower FAR scores for all cases in the high rainfall thresholds also support the earlier inference (refer Table 2) that the cv6 runs perform better than the CTRL and cv5 runs for higher-intensity rainfall.

## 6 Conclusions

The aim of the current study was to investigate whether the formulation of control variables in the BEC matrix affect the impact of observations in a data assimilation system considerably. It is seen here that this does have a moderate influence on assimilation impact. Previous studies have pointed out the importance of using a regional BEC matrix for assimilation. Here, two formulations of the regional BEC matrix are compared with each other. The results of the present study indicate that the multivariate BEC formulation given by the cv6

option in WRFDA has a better positive impact on assimilation and forecast fields. As expected, the moisture field is seen to be impacted the most by using the cv6 option of the WRF 3DVar system. The cv6 option has simulated a stronger cold core at lower levels as compared with the cv5 run for the depressions considered here. Vertical profiles of moisture convergence indicate stronger moisture convergence values for two of the depressions at low levels in the cv6 run as compared with the cv5 run. Higher relative vorticity and horizontal divergence values are also seen at the lower levels in the cv6 option for two cases. More moisture is simulated in the cv6 run in the lower levels for two out of three cases considered here. Furthermore, in two of the three depression cases considered here, skill scores estimated for 48 h rainfall forecasts are better in the cv6 runs for the higher threshold levels as compared with the cv5 and CTRL run. The location and magnitude of maximum 48 h accumulated rainfall is also better simulated in the cv6 run when compared with the cv5 and CTRL run. However, broad conclusions on the impacts of using the cv6 option can only be arrived at by investigating more cases.

It is well known that BEC matrix values contribute to the performance of a data assimilation system. The results of the current investigation have shown moderate improvement in the simulation while using the cv6 option as compared to the cv5 option available in the WRF 3DVar system. Hence the results of the study indicate that the formulation of control variables in the 3DVar system does have an impact on the analysis as well as model forecast. The cv5 and cv6 options differ mainly in the formulation of moisture control variable. Moisture is a highly variable component in the atmosphere in both space and time. Especially in a tropical region like India where convection is a major driving factor for weather systems, better formulation of moisture fields in the analysis system can improve the analysis and consequently the model predictions. The BEC matrix is one factor which can impact the analysis considerably. Better formulations of the BEC matrix by incorporating flow-dependent information can further improve the analysis system.

**Acknowledgements.** The authors acknowledge the model and data support from the National Aeronautics and Space Administration (NASA) for TRMM data, NCEP/NCAR for WRF model and Indian Institute of Space Science and Technology (IIST).

The topical editor, M. Salzmann, thanks the three anonymous referees for help in evaluating this paper.

## References

- Baker, N. L., Hogan, T., Campbell, W., Pauley, R., and Swadley, S.: The impact of AMSU-A radiance assimilation in the US Navy's Operational Global Atmospheric Prediction System (NOGAPS), Tech. rep., DTIC Document, 2005.
- Bannister, R. N.: A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances, *Q. J. Roy. Meteor. Soc.*, 134, 1951–1970, 2008a.
- Bannister, R. N.: A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics, *Q. J. Roy. Meteor. Soc.*, 134, 1971–1996, 2008b.
- Barker, D. M., Huang, W., Guo, Y.-R., Bourgeois, A., and Xiao, Q.: A three-dimensional variational data assimilation system for MM5: Implementation and initial results, *Mon. Weather Rev.*, 132, 897–914, 2004.
- Berre, L.: Estimation of synoptic and mesoscale forecast error covariances in a limited-area model, *Mon. Weather Rev.*, 128, 644–667, 2000.
- Buehner, M.: Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting, *Q. J. Roy. Meteor. Soc.*, 131, 1013–1043, 2005.
- Buehner, M.: Error statistics in data assimilation: estimation and modelling, in: *Data Assimilation*, 93–112, Springer, Berlin Heidelberg, 2010.
- Chandrasekar, A. and Kutty, M. G.: Studies on the Impacts of 3D-VAR Assimilation of Satellite Observations on the Simulation of Monsoon Depressions over India, in: *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications* (Vol. II), edited by: Park, S. K. and Xu, L., 643–705, Springer, Berlin Heidelberg, 2013.
- Chen, Y., Rizvi, S. R., Huang, X.-Y., Min, J., and Zhang, X.: Balance characteristics of multivariate background error covariances and their impact on analyses and forecasts in tropical and Arctic regions, *Meteorol. Atmos. Phys.*, 121, 79–98, 2013.
- Courtier, P., Thépaut, J.-N., and Hollingsworth, A.: A strategy for operational implementation of 4D-Var, using an incremental approach, *Q. J. Roy. Meteor. Soc.*, 120, 1367–1387, 1994.
- Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., et al.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q. J. Roy. Meteor. Soc.*, 137, 553–597, 2011.
- Derber, J. and Bouttier, F.: A reformulation of the background error covariance in the ECMWF global data assimilation system, *Tellus A*, 51, 195–221, 1999.
- Fisher, M.: Background error covariance modelling, in: *Seminar on Recent Development in Data Assimilation for Atmosphere and Ocean*, ECMWF, Reading, UK, 45–63, 2003.
- Godbole, R. V.: The composite structure of the monsoon depression, *Tellus*, 29, 25–40, 1977.
- Govindankutty, M. and Chandrasekar, A.: Effect of 3DVAR assimilation of MODIS temperature and humidity profiles on the dynamic and thermodynamic features of three monsoon depressions over the Bay of Bengal, *Meteorol. Atmos. Phys.*, 107, 65–79, 2010.
- Hayden, C. M. and Purser, R. J.: Recursive filter objective analysis of meteorological fields: Applications to NESDIS operational processing, *J. Appl. Meteorol.*, 34, 3–15, 1995.
- Hollingsworth, A. and Lönnberg, P.: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field, *Tellus A*, 38, 111–136, 1986.
- Katz, D., Lawless, A. S., Nichols, N. K., Cullen, M., and Bannister, R. N.: Correlations of control variables in variational data assimilation, *Q. J. Roy. Meteor. Soc.*, 137, 620–630, 2011.
- Parrish, D. F. and Derber, J. C.: The National Meteorological Center's spectral statistical-interpolation analysis system, *Mon. Weather Rev.*, 120, 1747–1763, 1992.
- Purser, R. J., Wu, W.-S., Parrish, D. F., and Roberts, N. M.: Numerical aspects of the application of recursive filters to variational statistical analysis. Part I: Spatially homogeneous and isotropic Gaussian covariances, *Mon. Weather Rev.*, 131, 1524–1535, 2003.
- Rakesh, V. and Goswami, P.: Impact of background error statistics on 3D-Var assimilation: case study over the Indian region, *Meteorol. Atmos. Phys.*, 112, 63–79, 2011.
- Routray, A., Kar, S., Mali, P., and Sowjanya, K.: Simulation of Monsoon Depressions Using WRF-VAR: Impact of Different Background Error Statistics and Lateral Boundary Conditions, *Mon. Weather Rev.*, 142, 3586–3613, 2014.
- Sikka, D.: Some aspects of the life history, structure and movement of monsoon depressions, *Pure Appl. Geophys.*, 115, 1501–1529, 1977.
- Singh, R., Kishtawal, C., Pal, P., and Joshi, P.: Improved tropical cyclone forecasts over north Indian Ocean with direct assimilation of AMSU-A radiances, *Meteorol. Atmos. Phys.*, 115, 15–34, 2012.

- Sinha, P. and Chandrasekar, A.: Improvement of mesoscale forecasts of monsoon depressions through assimilation of QuikSCAT wind data: two case studies over India, *Open Atmospheric Science Journal*, 4, 160–177, 2010.
- Skamarock, W., Klemp, J., Dudhia, J., Gill, D., Barker, D., Duda, M., Huang, X.-y., and Wang, W.: A Description of the Advanced Research WRF Version 3. NCAR Technical Note NCAR/TN-475+STR, Tech. rep., NCAR, 2008.
- Sørland, S. L. and Sorteberg, A.: The dynamic and thermodynamic structure of monsoon low-pressure systems during extreme rainfall events, *Tellus A*, 67, doi:10.3402/tellusa.v67.27039, 2015.
- Stano, G., Krishnamurti, T., Kumar, T. V., and Chakraborty, A.: Hydrometeor structure of a composite monsoon depression using the TRMM radar, *Tellus A*, 54, 370–381, 2002.
- Wilks, D. S.: *Statistical methods in the atmospheric sciences*, vol. 100, Academic press, 260–269, 2011.