Nat Hazards (2014) 71:659–682
DOI 10.1007/s11069-013-0942-1

ORIGINAL PAPER

# A new ensemble-based data assimilation algorithm to improve track prediction of tropical cyclones

**Deepak Subramani · R. Chandrasekar · K. Srinivasa Ramanujam · C. Balaji**

Received: 23 September 2012 / Accepted: 27 October 2013 / Published online: 16 November 2013


**Abstract** This paper proposes a new ensemble-based algorithm that assimilates the vertical rain structure retrieved from microwave radiometer and radar measurements in a regional weather forecast model, by employing a Bayesian framework. The goal of the study is to evaluate the capability of the proposed technique to improve track prediction of tropical cyclones that originate in the North Indian Ocean. For this purpose, the tropical cyclone Jal has been analyzed by the community mesoscale weather model, weather research and forecasting (WRF). The ensembles of prognostic variables such as perturbation potential temperature ($\theta$, $k$), perturbation geopotential ($\phi$, $m^2/s^2$), meridional ($U$) and zonal velocities ($V$) and water vapor mixing ratio ($q_v$, kg/kg) are generated by the empirical orthogonal function technique. An over pass of the tropical rainfall-measuring mission (TRMM) satellite occurred on 06th NOV 0730 UTC over the system, and the observations from the radiometer and radar on board the satellite(1B11 data products) are inverted using a combined in-home radiometer-radar retrieval technique to estimate the vertical rain structure, namely the cloud liquid water, cloud ice, precipitation water and precipitation ice. Each ensemble is input as a possible set of initial conditions to the WRF model from 00 UTC which was marched in time till 06th NOV 0730 UTC. The above-mentioned hydrometeors from the cloud water and rain water mixing ratios are then estimated for all the ensembles. The Bayesian filter framework technique is then used to determine the conditional probabilities of all the candidates in the ensemble by comparing the retrieved hydrometeors through measured TRMM radiances with the model simulated hydrometeors. Based on the posterior probability density function, the initial conditions at 06 00 UTC are then corrected using a linear weighted average of initial ensembles for the all prognostic variables. With these weighted average initial conditions, the WRF model has been run up to 08th Nov 06 UTC and the predictions are then compared with observations and the control run. An ensemble independence study was conducted on the basis of which, an optimum of 25 ensembles is arrived at. With the optimum ensemble size, the sensitivity of prognostic variables was also analyzed. The model simulated track when

D. Subramani · R. Chandrasekar · K. S. Ramanujam · C. Balaji (✉)
IIT Madras, Chennai, India
e-mail: balaji@iitm.ac.in

compared with that obtained with the corrected set of initial conditions gives better results than the control run. The algorithm can improve track prediction up to 35 % for a 24 h forecast and up to 12 % for a 54 h forecast.

**Keywords** Ensemble data assimilation · Bayesian framework · EOF · Remote sensing · TRMM · WRF · Cyclone track forecast

## 1 Introduction

The prediction of the track and intensity of tropical cyclones continues to pose considerable challenge to meteorologists. Accurate prediction of track and intensity of cyclones holds the key for early warning and mitigation. With the availability of state-of-the-art computational facilities and high resolution regional mesoscale models, it is possible to predict tracks with increasing accuracy. Even so, typically landfall errors for a 72 h forecast of a tropical cyclone remain around 300–400 km (Pattanaik and Rama Rao 2009; Sandeep et al. 2007; Singh et al. 2008). Mohanty et al. (1999) analyzed the Orissa super cyclone (1999) using National Center for Atmospheric Research (NCAR) mesoscale model and concluded that (i) the error in track forecast is proportional to the position error in the initial vortex, (ii) the track forecast could be improved by reducing error in the initial conditions. Sandeep et al. (2007) indicated that the poor track prediction of a cyclonic system is due to the poor representation of the initial vortex position. In view of this, they investigated the impact of NCAR-AFWA(Air Force Weather Agency) synthetic vortex scheme for modifying the initial vortex position of tropical cyclones over the Arabian sea during Nov 2003 and then concluded that the simulations of cyclones with the synthetic vortex scheme give better results in terms of track, intensity and structure of cyclone, when compared with simulations without the synthetic vortex scheme. Severe tropical cyclones over the Bay of Bengal during 2007–2010 were analyzed by Raju et al. (2012) with WRF, and they concluded that while WRF model can simulate the cyclonic system with good accuracy in terms of track and intensity, it can be further improved by an accurate representation of initial vortex position and inclusion of available satellite observations over open ocean, as the system is evolving. From the above studies, it is clear that the forecast skill of any regional mesoscale weather model strongly depends on the model initial/boundary conditions. The prediction skill also depends on the model physics parameterization schemes/dynamics and representation of topographical data like albedo, snow cover, vegetation index etc.

$$\text{Forecast error} = f\left(\begin{bmatrix}\text{Error}\\ \text{in}\\ \text{initial/boundary}\\ \text{conditions}\end{bmatrix}, \begin{bmatrix}\text{Error in}\\ \text{model physics}\\ \text{parameterization}\\ \text{/dynamics}\end{bmatrix}, \begin{bmatrix}\text{Error}\\ \text{in}\\ \text{topography}\\ \text{data}\end{bmatrix}\right) \quad (1)$$

Equation 1 indicates the forecast error is a function of error in initial and boundary conditions, Limitations in model physics parametrization and dynamics and Error in surface boundary conditions. Physics parameterizations are the heart of all mesoscale weather models that represent the atmospheric processes as they create and mimic the atmosphere. A mesoscale weather model contains several physics parameterizations, and under each parameterization, different schemes are available(i.e., have been developed). Several

researchers have worked on these schemes, with a view to identify the best schemes for predicting certain kinds of weather systems. Chandrasekar and Balaji (2012) analyzed the sensitivity of the cyclone Jal to physical parameterizations by using WRF. They concluded that different combinations of different schemes give rise to different results and finally obtained the best combination of physics parameterization schemes for track prediction based on the tropical cyclone Jal. However, the best combination of physics schemes was seen to over predict the intensity of the same cyclone and also gave different results when the grid size and number of nesting were changed. The key result from the study was the difficulty to generalize the best physics schemes that will work for all systems, particularly tropical cyclones.

Weather models represent the dynamics of the atmosphere processes by solving nonlinear equations by using numerical techniques that have a certain order of accuracy due to truncation, discretization errors while approximating the Taylor's series and round off errors. Furthermore, accurate measurement of topography details of the region under consideration is difficult and suffers from a lack of regular updating. Hence, controlling the errors in model physics/dynamics and the topography database can be achieved only up to a certain extent. Most regional mesoscale models are initialized by the Global Forecast System (GFS) model outputs. Because of the low resolution of GFS data, the mesoscale features of the weather systems are invariably not represented accurately, because of which errors in the initial and boundary conditions are introduced. Data assimilation techniques are the only way to improve those initial conditions by assimilating observations, whether from space borne or land-based instruments. Srinivas et al. (2012) analyzed the impact of conventional and satellite observation data on the prediction of two cyclonic storms in the Bay of Bengal by the three-dimensional variational (3DVAR) data assimilation technique and concluded that assimilation with the conventional observational data gives poor prediction. This is mainly because of the fact that most of the conventional data are available over land because of which large errors are introduced in the initial stages of the cyclonic systems, where they are developing in the open ocean. Zhao et al. (2005) studied the impact of assimilating the Advanced microwave sounder unit-A(AMSU-A) retrieved wind and temperature in the case of typhoon Dan with four-dimensional variational assimilation(4DVAR) system. They concluded that the assimilation of satellite observations has a positive impact on the typhoon track forecast. From the above, it is clear that assimilating satellite observations particularly from a microwave, sounder or radiometer holds the key to improving forecast skill, as satellite observations can give us a better estimate of the initial vortex location and intensity of the cyclone. Singh et al. (2008) assimilated QuickSat satellite wind observations for the simulation of the track of the Orissa super cyclone by using 3DVAR and concluded that wind observations from QuickSat have great impact and improving the cyclone predictions. The focus of this study is the assessment of the impact of assimilating the hydrometeors profiles, viz., cloud liquid water (CW), cloud ice (CI), precipitationwater (PW) and precipitation ice (PI) retrieved from tropical rainfall measuring mission (TRMM) microwave imager (TMI) brightness temperatures (BT) using a WRF-Radar-Radiometer combined algorithm in ensemble-based data assimilation.

The introduction and governing equations of the WRF and model domain, physics and dynamic option details are given in Sect. 2. Section 3 details the data that have been used in this study. The ensemble sampling and the retrieval algorithm are discussed in Sects. 4 and 5, respectively. The details of equations pertaining to the observation operator are discussed in Sect. 6. Section 7 elucidates the data assimilation technique. Sections 8 and 9 present the results and conclusions of this study, respectively.

## 2 Weather research and forecasting

The WRF model is a popular community mesoscale weather model developed by the NCAR. The WRF, as like other mesoscale models, obtains its initial and boundary conditions from the GFS and downscales them into the required grid resolution.

### 2.1 Governing equations

In general, the WRF dynamic solver solves the following equations in the region of interest, with specified grid resolution and time step.

The momentum equations

$$\frac{\partial U}{\partial t} + m_x\left[\frac{\partial Uu}{\partial x} + \frac{\partial Vu}{\partial y}\right] + \frac{\partial \Omega u}{\partial \eta} + \left(\frac{m_x}{m_y}\right) + \left(\frac{\alpha}{\alpha_d}\right)\left[\mu_d\left(\frac{\partial \phi'}{\partial x} + \alpha_d\frac{\partial p'}{\partial x} + \alpha'_d\frac{\partial \overline{p}}{\partial x}\right) + \frac{\partial \phi}{\partial x}\left(\frac{\partial \phi'}{\partial \eta} - \mu'_d\right)\right] = F_U \tag{2}$$

$$\frac{\partial V}{\partial t} + m_y\left[\frac{\partial Uv}{\partial x} + \frac{\partial Vv}{\partial y}\right] + \left(\frac{m_y}{m_x}\right)\frac{\partial \Omega v}{\partial \eta} + \left(\frac{m_x}{m_y}\right) + \left(\frac{\alpha}{\alpha_d}\right)\left[\mu_d\left(\frac{\partial \phi'}{\partial x} + \alpha_d\frac{\partial p'}{\partial x} + \alpha'_d\frac{\partial \overline{p}}{\partial x}\right) + \frac{\partial \phi}{\partial x}\left(\frac{\partial \phi'}{\partial \eta} - \mu'_d\right)\right] = F_V \tag{3}$$

$$\frac{\partial W}{\partial t} + \left(\frac{m_x m_y}{m_y}\right)\left[\frac{\partial Uw}{\partial x} + \frac{\partial Vw}{\partial y}\right] + \frac{\partial \Omega v}{\partial \eta} + \left(\frac{g}{m_y}\right) + \left(\frac{\alpha}{\alpha_d}\right)\left[\frac{\partial p'}{\partial \eta} - \overline{\mu_d}(q_v + q_c + q_v)\right] + \frac{\mu' g}{m_y} = F_W \tag{4}$$

Energy equation

$$\frac{\partial \Theta}{\partial t} + m_x m_y\left[\frac{\partial U\theta}{\partial x} + \frac{\partial V\theta}{\partial y}\right] + m_y\frac{\partial \Omega\theta}{\partial \eta} = F_\Theta \tag{5}$$

Continuity equation

$$\frac{\partial \mu'_d}{\partial t} + m_x m_y\left[\frac{\partial U}{\partial x} + \frac{\partial V}{\partial y}\right] + m_y\frac{\partial \Omega}{\partial \eta} = 0 \tag{6}$$

Equation of geopotential

$$\frac{\partial \phi'}{\partial t} + \mu_d^{-1}\left[m_x m_y\left(U\frac{\partial \phi}{\partial x} + V\frac{\partial \phi}{\partial y}\right) + m_y\Omega\frac{\partial \phi}{\partial \eta} - m_g g W\right] = 0 \tag{7}$$

Continuity equation of moisture

$$\frac{\partial Q_m}{\partial t} + m_x m_y\left[\frac{\partial Uq_m}{\partial x} + \frac{\partial Vq_m}{\partial y}\right] + m_y\frac{\partial \Omega q_m}{\partial \eta} = F_{Q_m} \tag{8}$$

The above equations are also known as Euler equations for the basic prognostic equations which govern the dynamics of the atmosphere. The WRF also uses the following diagnostic relations.

Equation of state

$$p = p_0 \left( \frac{R_d \theta_m}{p_0 \alpha_d} \right)^{\gamma} \tag{9}$$

$$\theta_m = \theta \left( 1 + \left( \frac{R_v}{R_d} \right) q_v \right) \approx \theta (1 + 1.61 q_v) \tag{10}$$

Inverse density

$$\frac{\partial \phi'}{\partial \eta} = -\overline{\mu}_d \alpha'_d - \alpha_d \mu'_d \tag{11}$$

$$\alpha = \alpha_d (1 + q_v + q_c + q_r + q_i + ...) \tag{12}$$

Linear relations

$$p = \overline{p} + p' \tag{13}$$

$$\phi = \overline{\phi} + \phi' \tag{14}$$

$$\alpha = \overline{\alpha} + \alpha' \tag{15}$$

$$\mu_d = \overline{\mu}_d + \mu'_d \tag{16}$$

where

$p$ pressure (pa)
$p_0$ reference sea-level pressure ($10^5$pa)
$\overline{p}$ reference state pressure (pa)
$p'$ perturbation pressure (pa)

$\phi$ geopotential ($m^2/s^2$)
$\overline{\phi}$ geopotential for reference state ($m^2/s^2$)
$\phi'$ perturbation geopotential ($m^2/s^2$)

$\alpha$ inverse density of air ($m^3$/kg)
$\alpha_d$ inverse density of dry air ($m^3$/kg)
$\overline{\alpha}_d$ inverse density of dry air for the reference state ($m^3$/kg)
$\alpha'_d$ perturbation inverse density of dry air ($m^3$/kg)

$\mu$ hydrostatic pressure difference between surface and top of the model (pa)
$\mu_d$ dry hydrostatic pressure difference between surface and top of the model (pa)
$\overline{\mu}_d$ reference state dry hydrostatic pressure difference between surface and top of the model (pa)
$\mu'_d$ perturbation hydrostatic pressure difference between surface and top of the model (pa)

$\theta$ is potential temperature (K)
$\Theta$ coupled potential temperature ($\mu\ \theta$)
$(\Theta - T_0)$ perturbation potential temperature (K)
$T_0$ reference sea-level temperature (usually 270–300 K)
$\theta_m$ virtual temperature or moist potential temperature (K);

$u$ horizontal component of velocity in $x$-direction (m/s)
$v$ horizontal component of velocity in $x$-direction (m/s)
$w$ vertical component of velocity (m/s)
$U$ coupled horizontal component of velocity in $x$-direction ($\mu\ u$)
$V$ coupled horizontal component of velocity in $y$-direction ($\mu\ v$)
$W$ coupled vertical component of velocity ($\mu\ w$)
$\Omega$ coupled coordinate velocity

$q_m$ is the generic mixing ratios for moisture (kg/kg) and $q_m = q_v + q_c + q_r + q_i + \ldots$
$q_v$, $q_c$, $q_r$, $q_i$ are mixing ratios for water vapor, cloud, rain and ice, respectively, and all are in (kg/kg)
$Q_m$ generic coupled moisture variable ($\mu_d\ q_m$)

$m_x$, $m_y$ are map scale factors ($m$)
$R_d$, $R_v$ are gas constants of dry air 287 J/kg k and vapor 461.6 J/kg k, respectively
$\rho$, $\rho_d$ are density (kg/m$^3$) of air and dry air, respectively
$g$ acceleration due to gravity (9.81 m/$s^2$)
$\gamma$ is heat capacity ratio of air($\gamma = c_p/c_v$)
$F_U, F_V, F_W, F_\Theta$ and $F_{Q_m}$ are forcing terms arising from model physics

Equations 2–9 are integrated by the WRF dynamic solver using a third-order Runge-kutta (RK3) time integration scheme, detailed in Skamarock and Klemp (2008). The model physics and other details are clearly described in the NCAR technical report by Skamarock et al. (2005).

### 2.2 Model domain, physics and dynamic options

Figure 1 shows the model domain that covers the region spanning 77E to 93E and 6N to 22N with a single domain. The grid resolution is 6 km, and 291 grid points are used in both the East-West and North-South directions. A sensitivity of physics parameterizations for cyclone JAL has been conducted by Chandrasekar and Balaji (2012) for both track and wind prediction. The set of physics parameterization schemes for the best wind prediction from (Chandrasekar and Balaji 2012) has been used in this study. Incidentally, the same set of physics parametrization schemes has been used for developing the in-house algorithm of retrieval of geophysical parameters from satellite microwave radiances and radar reflectivities. Table 1 indicates the set of physics schemes, and Table 2 lists the dynamic options used in this study.

## 3 Data

The United States Geological Survey (USGS) 30“ resolution terrain topographical data have been used in the WRF preprocessing system (WPS). Data based on 0.5° resolution GFS real-time prediction from the National Center for Environmental prediction (NCEP) have been used as initial and boundary conditions. Simulations were initiated on 06th Nov 2010 00 UTC with lateral boundary conditions up to 08 Nov 2010 06 UTC. The Joint Typhoon Warning Center(JTWC) observed cyclone track data are used to compare the model simulated track, and TMI 1B11 (10.65, 19.35, 21, 37, and 85.5 GHz) BT data are used to retrieve hydrometeors profiles that are assimilated simultaneously.

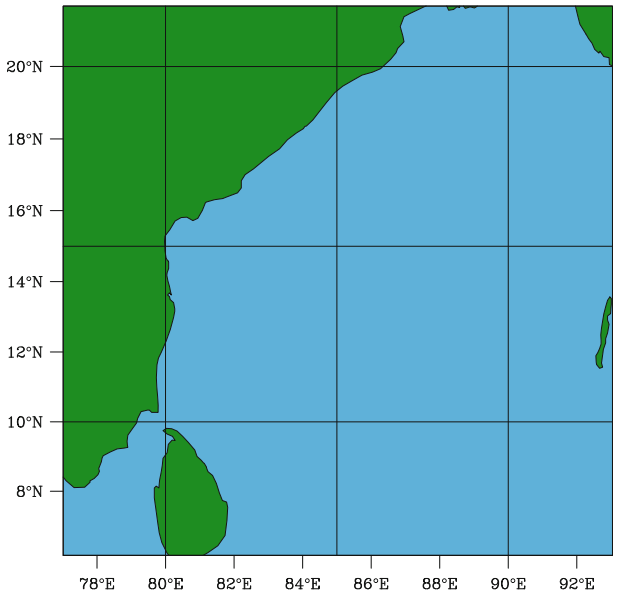


**Fig. 1** Model domain employed in this study

**Table 1** Best combination of physics schemes for wind prediction

| S. no. | Parametrization | Scheme |
|---|---|---|
| 1 | Cumulus parameterization | Grell-Devenyi ensemble scheme (GD) |
| 2 | PBL | Mellor-Yamada-Janjic (Eta) TKE scheme (MYJ) |
| 3 | Microphysics | WRF single moment 3-class simple ice scheme (WSM3) |
| 4 | Surface layer physics | Monin-Obukhov (Janjic Eta) scheme (JAN) |
| 5 | Land surface model | Pleim-Xu scheme (PLEIM) |
| 6 | Longwave radiation physics | Rapid radiative transfer model (RRTM) |
| 7 | Shortwave radiation physics | Rapid radiative transfer model for global (RRTMG) |

## 4 Ensemble sampling

The Empirical Orthogonal function (EOF) technique helps in synthetically generating a random atmospheric profile that is thermodynamically and physically consistent with a database of profiles. Zhang and Krishnamurti (1997, 1999) used the EOF technique to perturb the initial and conditions (only temperature and wind fields) and ran the Florida State University Global Spectral Model(FSUGSM) with different perturbed initial conditions. They concluded that the ensemble technique reduces the uncertainty in the initial conditions and that the average of ensemble forecast gives better results when compared with the control run in terms of track and intensity forecast. The EOF is a powerful technique to obtain synthetic profiles that double up as ensembles and uses principal component analysis to reduce the randomness in the generated profiles, thereby giving better control of ensembles generated in the family. The EOF requires a database to start the ensemble generation process. In this study, the database comes from the NCEP GFS

**Table 2** Model dynamics and domain details

| | |
|---|---|
| Dynamics | |
| Equation | Non-hydrostatic |
| Time integration scheme | Third-order Runge-Kutta scheme |
| Horizontal grid type | Arakawa-C grid |
| Domain | |
| Map projection | Lambert conformal mapping |
| Central point of the domain | 85 E, 14 N |
| No of domains | 1 |
| No of vertical layers | 27 sigma levels |
| Horizontal grid distance | 6 km |
| Time step | 18 sec |
| No of grid points | 291 in both (E-W) and (N-S) |

data(2010 Nov 06 00 UTC) which is used as initial conditions for the WRF. For a fuller discussion of generation of ensembles using the EOF technique, please refer to Tatarskaia et al. (1998).

The generation of synthetic profiles involves the following steps:

1. Extracting the shape information from a database through the covariance matrix.
2. Calculating principal components of this covariance matrix by performing an eigenvalue analysis.
3. Generating the synthetic profile by using random number vectors and eigenvectors and eigenvalues of principal components.

Let $X$ be the variable profile for which EOF is to be used to generate synthetic profile. $X$ is of the form NPROF $\times$ NGATE, where NGATE is the number of vertical levels of the profile and NPROF is the number of profiles in the database. Each profile is contained in each row of matrix $X$. First, the covariance matrix, $\mathbf{B}$ of $X$ is calculated as follows

$$B_{ik} = E(X_i' X_k') \tag{17}$$

$$\begin{aligned} X_i' &= X_i - E(X_i) \\ i, k &= 1 \cdots NGATE \end{aligned} \tag{18}$$

$B$ is a NGATE $\times$ NGATE matrix. Next, this covariance matrix is used to set up an eigenvalue problem as follows

$$\mathbf{B}\phi = \lambda\phi \tag{19}$$

$\phi$ is the matrix of eigenvectors, and $\lambda$ is the vector of eigenvalues. To reduce the dimensionality and the randomness in the synthetic profile, a principal component analysis is performed. Negative eigenvalues are discarded, and the positive eigenvalues are arranged in descending order. The ratio of eigenvalue to the sum of all eigenvalues gives the contribution of the corresponding eigenvector to the covariance matrix, $B$. The cumulative sum of the eigenvalues thus obtained gives the contribution of those many eigenvectors to $B$. A threshold of 0.99 is set, and the number of eigenvectors required for explaining 99 % variation in the data is obtained. After performing the principal component analysis(PCA), the number of principal components is determined as $N$. The new vector $X$ is calculated as follows

$$X_{\text{perturbed}} = X_{\text{unperturbed}} + \sum_{v=1}^{N} \zeta_v \sqrt{\lambda_v} \phi_v \tag{20}$$

$\zeta_v$ is a normal random number and $X_{\text{unperturbed}}$ is the vector of initial conditions obtained from NCEP. Depending on the requirement, any ensemble family size can be obtained by this technique.

### 4.1 Variables perturbed

From Eqs. 2–8, one can see that the WRF dynamic solver solves the set of equations with the following prognostic variables.

1. Perturbation geopotential ($\phi$,m$^2$/s$^2$)
2. Perturbation potential temperature ($\theta$, $K$)
3. $X$-wind($U$, m/s) and $Y$-wind velocity ($V$, m/s)
4. Water vapor mixing ratio ($q_v$, kg/kg)

The above variables are perturbed using the EOF technique explained above, and the vertical structure of the unperturbed and perturbed ensembles are shown in Figs. 2, 3, 4 and 5. Figure 6 shows the surface contour plots of water vapor mixing ratio, and Fig. 7 shows the maximum sustained wind speed at 10 m height of the unperturbed and the random perturbed ensembles.

## 5 Retrieval algorithm

The methodology followed here is to first invert the BT's measured by the TMI to obtain the state of the atmosphere at a particular instant of time and space, corresponding to the cyclone for which track improvement is desired, that can be assimilated into WRF. For this inversion a combined WRF-radar-radiometer physics-based retrieval algorithm is used. The algorithm looks at an a priori of a high-quality database prepared from WRF simulations matched up with TRMM Precipitation Radar (PR) and TMI. From the retrieved state of atmosphere, four hydrometeors are used for assimilation.

### 5.1 Overview

The procedure used to retrieve the state of the atmosphere is outlined in Figure 8. The strength of this Bayesian algorithm used here is the availability of a high-quality database of geophysical profiles and a forward model, viz., Micro-tropiques which can simulate BT corresponding to nine channels of TMI. Micro-tropiques is an in-house polarized radiative heat transfer equation detailed in Deiveegan et al. (2008). The algorithm, itself, consists of four steps.

1. The BT's for each pixel in the a priori database are simulated by Micro-tropiques ($BT_{mt,j}$, $1 < j < 400$).
2. The BT's of the pixel for which inversion is to be performed ($\text{BT}_{\text{tmi}}$ are obtained from 1B11 data products from Mirador).
3. The simulated BT of each and every pixel from the database is compared with TMI BT of the pixel to be retrieved to obtain the likelihood of the pixels ($X_j$) under consideration being the pixel to be retrieved ($\mathbf{X_{tmi}}$).

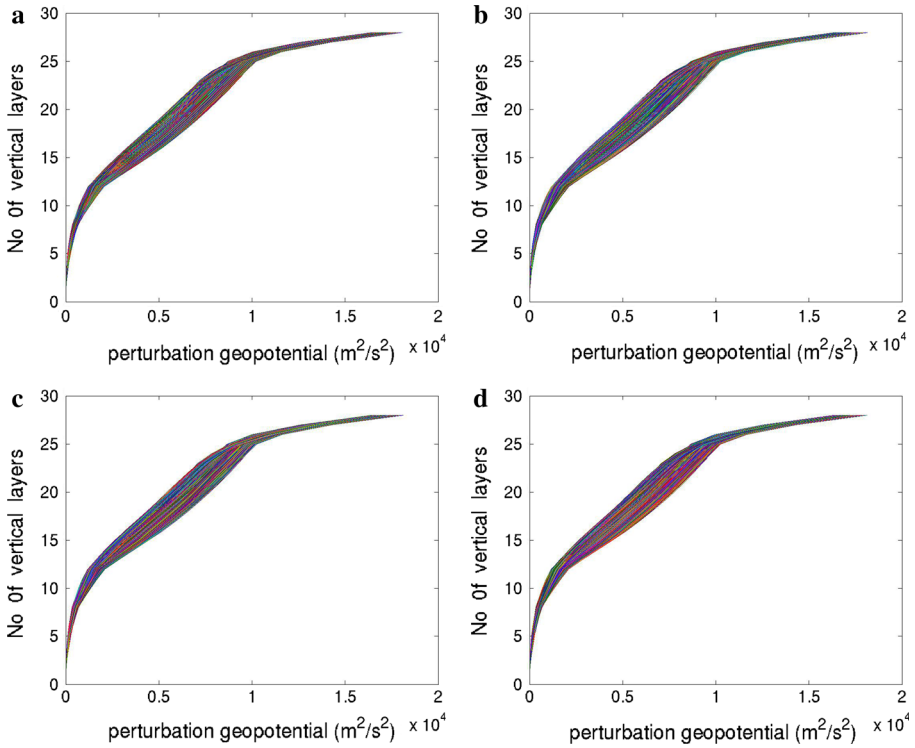


**Fig. 2** Vertical structure of perturbation geopotential (**a**) unperturbed GFS initial condition (**b–d**) are random samples of perturbed initial conditions. All are at 2010 Nov 06 00UTC

4. The likelihood is calculated as follows:

$$L(\mathbf{X_{tmi}}|\mathbf{X_j}) = \mathbf{exp}\left(-\frac{\sum_{i=1}^{14}(\mathbf{BT_{mt,j}} - \mathbf{BT_{tmi}})^2}{2\sigma^2}\right) \tag{21}$$

The posterior probability density function (PPDF) is constructed using the likelihoods calculated above. The fourth step is to take an expected value of the a-posteriori to obtain the retrieved profile.

## 5.2 Database

The algorithm, as explained in previous section, uses a database which is made using a judicious combination of WRF model predictions, radar reflectivity measured by PR and the BT measured by TMI. It is because of this combination that this algorithm is referred to as combined WRF-radar-radiometer algorithm. The formation of this database involves two match-up procedures using two different forward models. The initial database is formed from simulations of WRF which uses the best known physics and dynamics options. However, the thermodynamic and hydrometers profiles simulated by WRF still

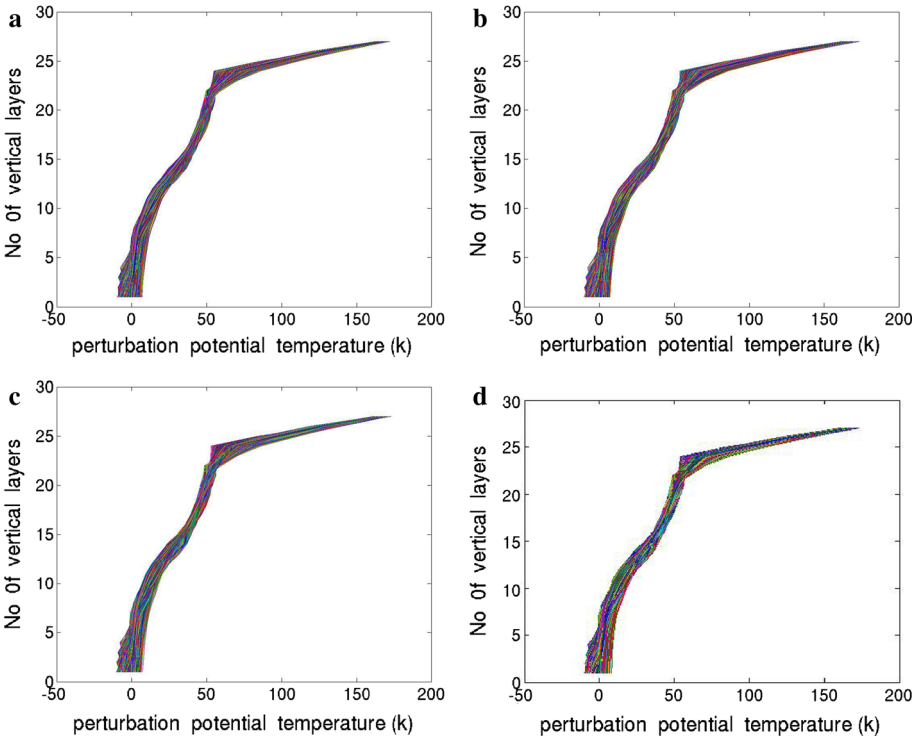


**Fig. 3** Vertical structure of perturbation potential temperature (**a**) unperturbed GFS initial condition (**b**–**d**) are random samples of perturbed initial conditions. All are at 2010 Nov 06 00UTC

have some model errors due to assumption of the same set of physics and dynamics options for all pixels irrespective of the inhomogeneity observed within the field of observation.

### *5.2.1 WRF profiles*

Vertical profiles of pressure (P), temperature (T), relative humidity (RH), CW, CI, PW and PI for 14 severe cyclonic storms that originated in north Indian Ocean region bounded between 0°–25°N and 60°–100°E during 2003–2010 are simulated by WRF to give these parameters corresponding to the time of TRMM overpass. In total, there were 31 such overpasses during these 14 cyclones covering the cyclone at various stages of development. The profiles are generated at 6 km pixel resolution in 72 pressure levels corresponding to TRMM PRs vertical resolution. WRF simulations are available at the entire domain under consideration. However, TMI and PR have limited footprint due to instrument capabilities. Therefore, a collocation procedure is required to identify common pixels. Here a minimum distance strategy is adopted to perform a two way collocation between WRF, TMI, and PR pixels. At the end of collocation, a total of 1,124 common pixels are identified. Now the database consists of WRF simulations, PR reflectivity, and TMI BT for some common 1,124 pixels. Using this information, rain and ice profiles can be matched up in two steps to obtain a high-quality database.

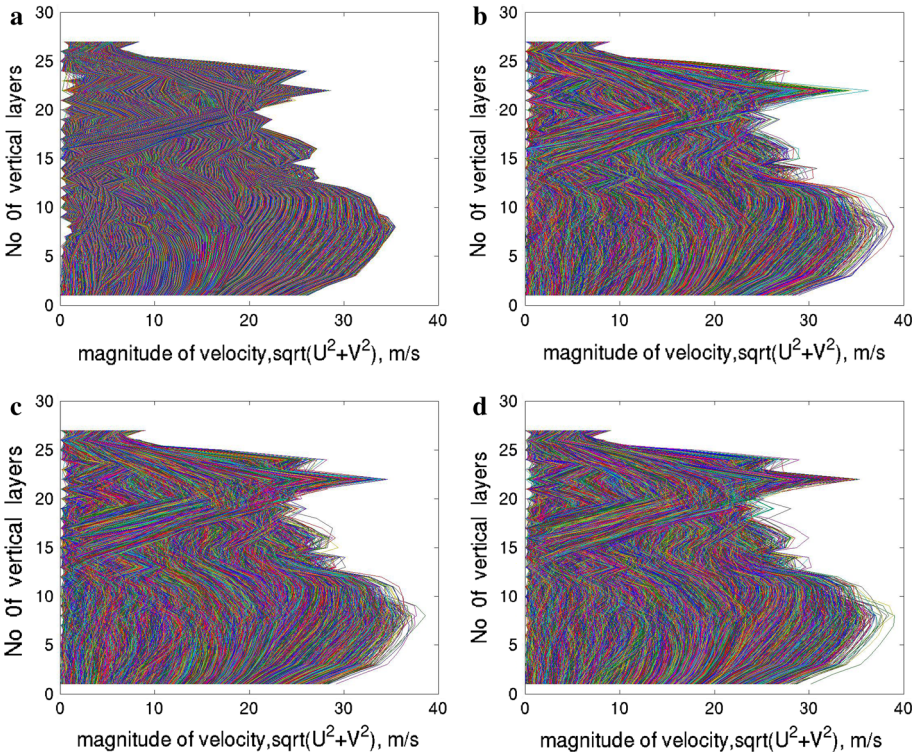


**Fig. 4** Vertical structure of the magnitude of velocity (**a**) unperturbed GFS initial condition (**b**–**d**) are random samples of perturbed initial conditions. All are at 2010 Nov 06 00UTC

### *5.2.2 Liquid profile match-up*

The 1,124 pixels are split into two sets of 600 and 524 each. Another 400 heavy rain WRF pixels are added to the set of 600 to form a diverse a priori of 1,000 WRF profiles. These profile are input to QuickBeam, a radar simulator package and the radar reflectivities at 72 PR levels (250–18,000 m) are simulated at 13.8 GHZ frequency corresponding to the PR. TRMM PR measured reflectivities from the set of 524 profiles are compared with the a priori reflectivities to estimate the likelihoods, and the PPDFs are obtained. The expected values from the PPDF form the matched-up cloud and rain profiles.

### *5.2.3 Ice profile match-up*

The TRMM PR is not sensitive to CI and PI contents, and hence, there is a need to use the TMI scattering channels to match up the ice profiles. The 524 PI matched-up profiles are converted to 14 TMI levels. The data is further split into two sets of 124 and 400 each. The first set of 124 matched-up WRF profiles in used for matching up the ice profiles, while the remaining are used for testing. The match-up is done by computing the BT corresponding to the scattering channels 37 and 85 GHz and using the Bayesian approach outlined Sect. 5.2.2. Figure 9 shows typical distributions of the four retrieved hydrometeors retrieved from TMI observations. A fuller discussion of the retrieval algorithm is available in (Ramanujam et al. 2012).

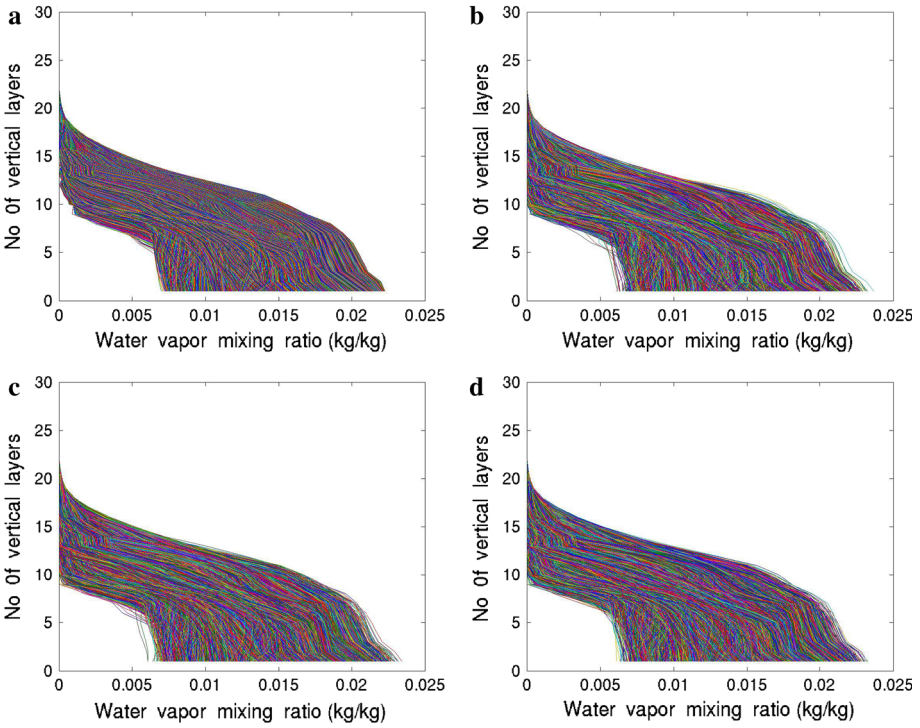


**Fig. 5** Vertical structure of water vapor mixing ratio (**a**) unperturbed GFS initial condition (**b–d**) are random samples of perturbed initial conditions. All are at 2010 Nov 06 00UTC

## 6 Observation operator

For data assimilation techniques, in general, the observation operator consists of a set of equations or a forward model which will convert the model atmospheric variables into available observation form. In this study, Eqs. 26–29 are used as the observation operator from which the four hydrometeors, viz., CW, CI, PW and PI are calculated from WRF output variables, viz P, T, RH, cloud mixing ratio ($Q_c$) and rain mixing ratio ($Q_r$).

$$P_{\text{sat}} = 61.078\left(\frac{7.5T - 2048.625}{T - 25.85}\right) \tag{22}$$

$$P_v = \text{RH} \times \frac{P_{\text{sat}}}{100} \tag{23}$$

$$P_{\text{air}} = P - P_v \tag{24}$$

$$\rho = 100\left(\frac{P_d}{R_d \times T} + \frac{P_v}{R_v \times T}\right) \tag{25}$$

where $P_{\text{sat}}$ is saturation pressure, $P_v$ is vapor pressure and $P_d$ is dry air pressure $R_v$ is gas constant of vapor(461.495)and ($R_d$) is gas constant of dry air (287.058).

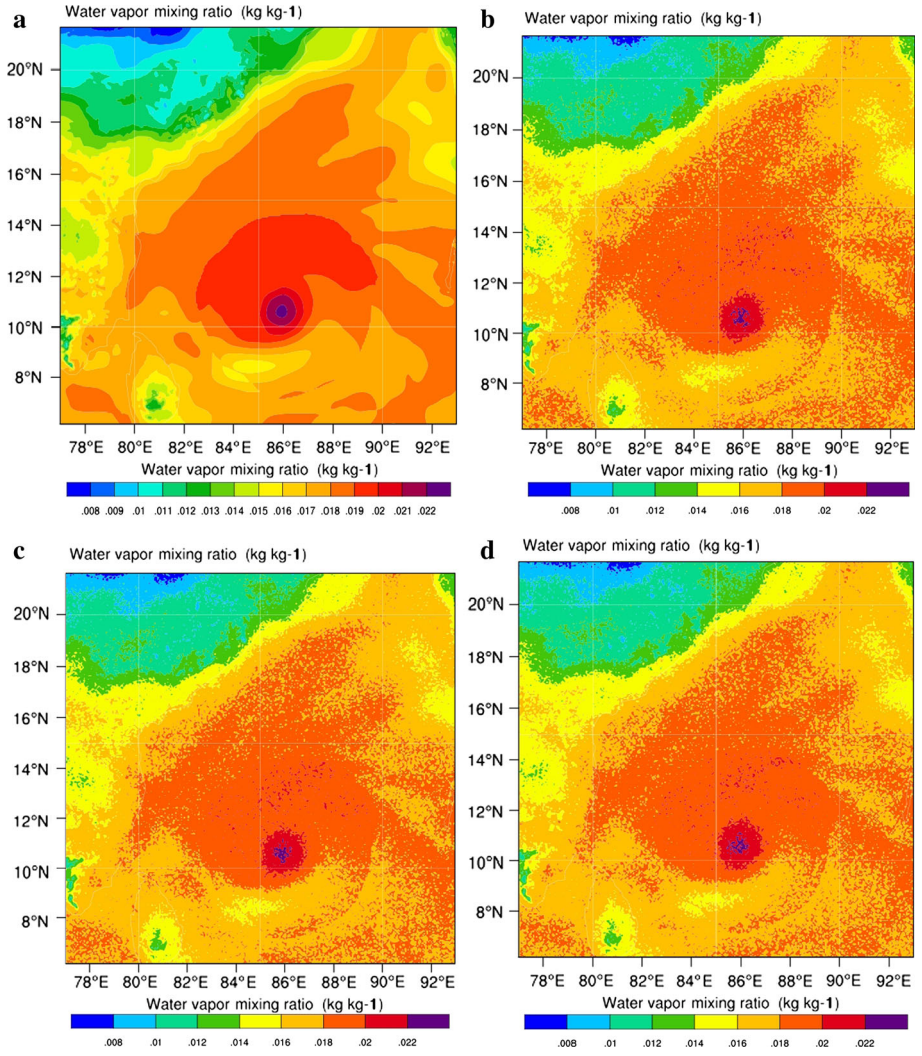


**Fig. 6** First level water vapor mixing ratio (**a**) unperturbed GFS initial condition (**b–d**) are random samples of perturbed initial conditions. All are at 2010 Nov 06 00UTC

$$\text{If height} \leq 5\,\text{km} \quad CW = 1000 \times Q_c \times \rho \quad \text{CI} = 0 \tag{26}$$

$$\text{If height} \geq 5\,\text{km} \quad CI = 1000 \times Q_c \times \rho \quad CW = 0 \tag{27}$$

$$\text{If height} \leq 5\,\text{km} \quad \text{PW} = 1000 \times Q_r \times \rho \quad \text{PI} = 0 \tag{28}$$

$$\text{If height} \geq 5\ \text{km} \quad \text{PI} = 1000 \times Q_r \times \rho \quad \text{PW} = 0 \tag{29}$$

At a height of 5 km, the temperature is around 273 K. Hence, below 5 km, the cloud and precipitation contents are in the form of water, and above 5 km, they are in the form of

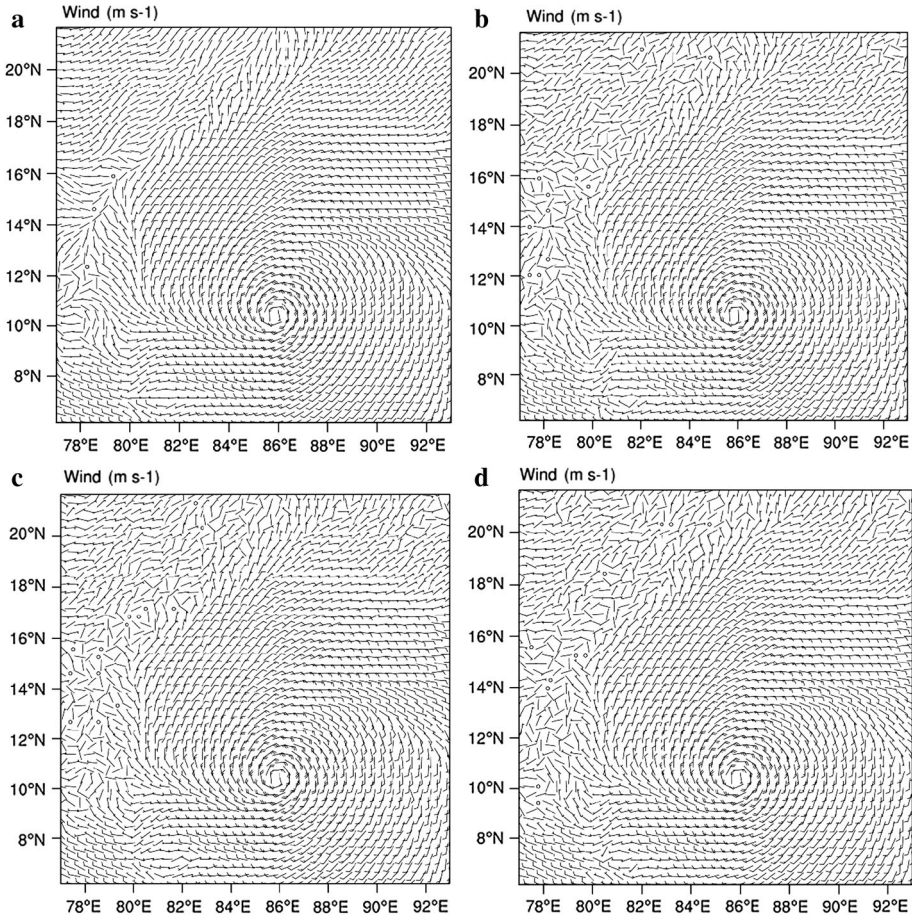


**Fig. 7** Maximum sustained wind at 10 m (**a**) unperturbed GFS initial condition (**b–d**) are random samples of perturbed initial conditions. All are at 2010 Nov 06 00UTC

ice. Figure 10 shows the basic atmospheric variables which are converted into observations variables. Figure 11 indicates the basic four hydrometeors that are calculated from the observation operator. The ordinate of the figures represents the vertical layer used in the TRMM retrieval algorithm and does not correspond to the actual height.

## 7 Data assimilation methodology

### 7.1 Motivation

The availability of the "observed" state of atmosphere data retrieved from TRMM observations using a "high-quality" database provides unprecedented application opportunities. The motivation of the present study is to use optimization techniques to perform data assimilation with the retrieved hydrometeors to reduce the track error. Of the various data assimilation techniques available natively in WRF, there is no scope to assimilate

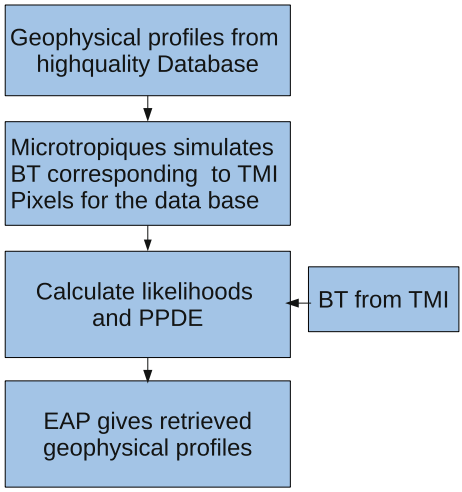


**Fig. 8** Flow chart of the algorithm developed for retrieval of hydrometeors

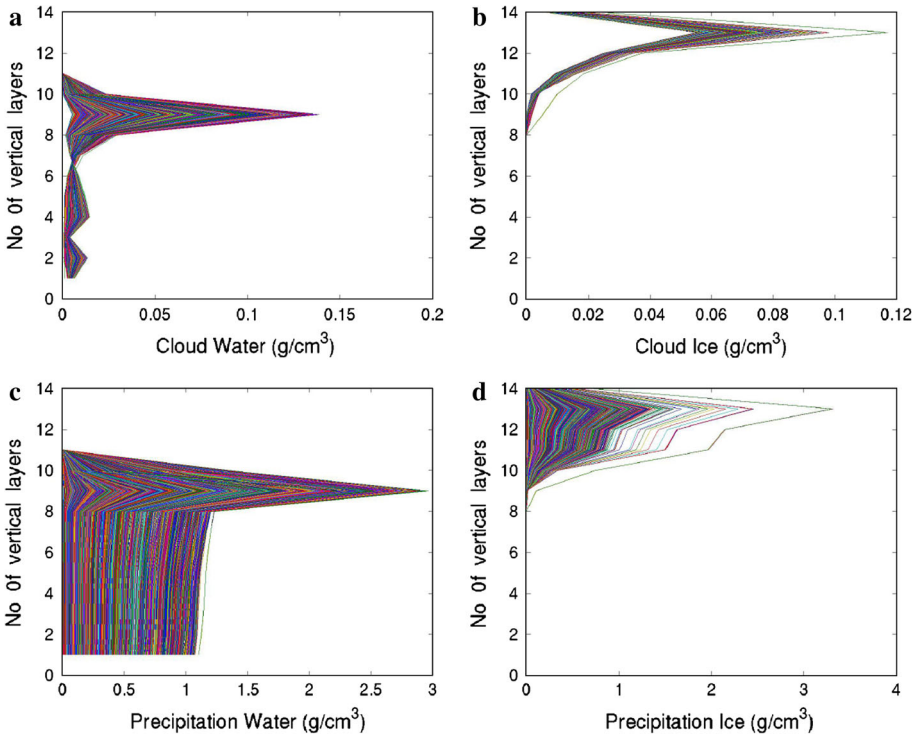


**Fig. 9** Retrieved hydrometeors from TMI. **a** Cloud water, **b** cloud ice, **c** precipitation water, **d** precipitation ice. All are at 2010 Nov 6 7.30UTC

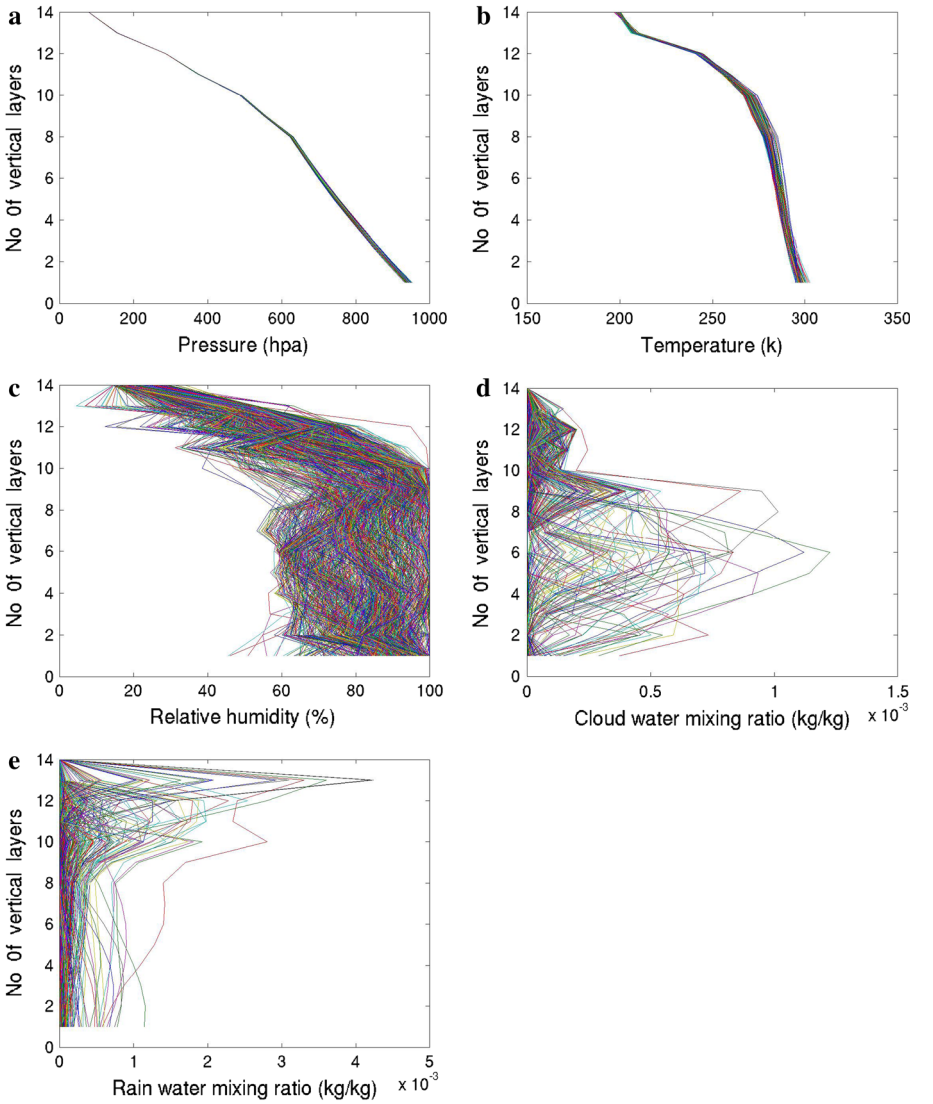


**Fig. 10** Basic atmospheric variables from WRF. **a** Pressure, **b** temperature, **c** relative humidity, **d** cloud water mixing ratio, **e** rain water mixing ratio. All are at 2010 Nov 6 7.30UTC

hydrometeors information, and so, a new algorithm is required for assimilating the hydrometeors. A stochastic Monte Carlo based ensemble algorithm using a Bayesian likelihood estimator is proposed and tested here.

## 7.2 Overview

The flowchart of the algorithm developed in this study is shown in Figure 12. The algorithm is essentially a Monte Carlo technique. First at 7.30 h before a TRMM overpass over

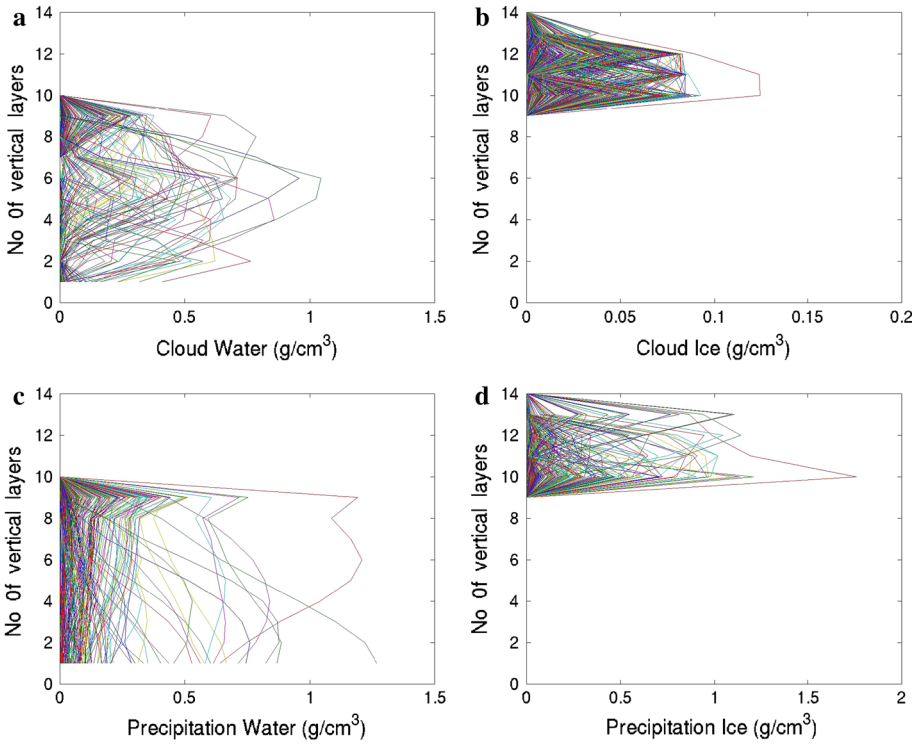


**Fig. 11** Calculated hydrometeors from the observation operator. **a** Cloud water, **b** cloud ice, **c** precipitation water, **d** precipitation ice. All are at 2010 Nov 6 7.30UTC

the cyclone under consideration, the time is set to $t = 0$. This is the initial condition with which WRF will start predicting the cyclone track. From here, the algorithm consists of six steps.

1. Preparation of a family of ensembles (with N ensembles) of initial conditions of variables considered. The ensemble sampling is performed using the EOF already discussed.
2. The WRF is initialized with the prepared N ensembles and simulations are carried out for a time period of 7.30 h corresponding to the TRMM overpass. The WRF simulates N sets of atmosphere in the entire domain under consideration at $t = 7.30$ h.
3. Retrieval the state of the atmosphere from TMI BT at TMI pixels for this instance of time using the WRF-radar-radiometer algorithm.
4. Identification of common pixels for TMI and WRF through a collocation procedure performed between TMI pixels and WRF pixels. A minimum distance strategy is used for this collocation.
5. A comparison of the hydrometers data at these common pixels obtained by two ways—WRF and retrieved atmosphere to obtain likelihoods of the initial condition ensembles. A Bayesian likelihood estimator is used in this step and a PPDF is constructed.
6. Completion of the data assimilation procedure by finding the best set of initial conditions that agrees well with observation retrieved from satellite. The expected

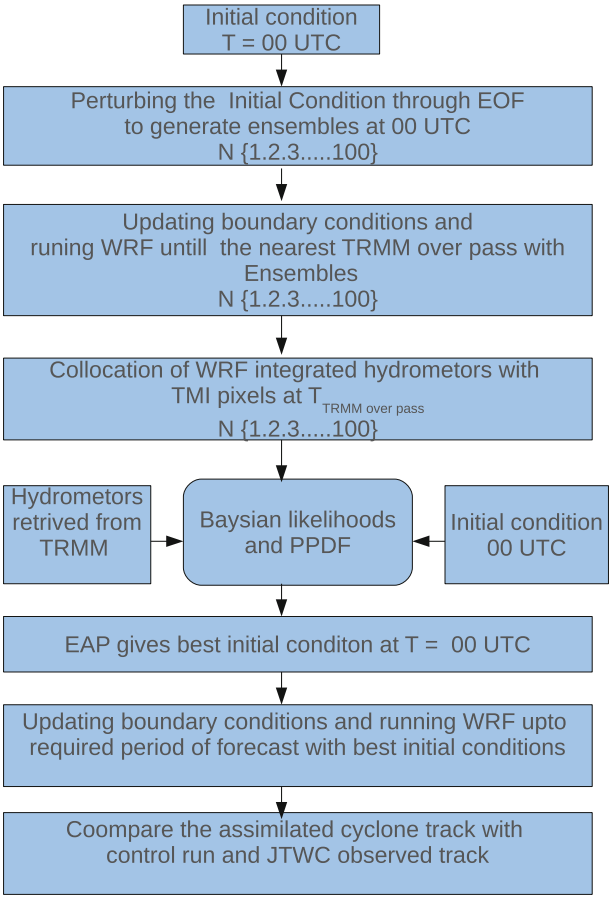


**Fig. 12** Flow chart of the new ensemble-based assimilation algorithm developed in this study

value of the a posteriori is the best estimate of initial conditions. At the end of these six steps, a "best" set of assimilated initial condition is obtained using which subsequent 54-h forecasts are performed using WRF.

### 7.3 Bayesian likelihood estimation

The Bayesian likelihood of the ensemble forecasts is calculated by comparing the four hydrometers data at collocated pixels. The likelihood is calculated as follows:

$$L_{(\text{ensemble})} = \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{14}\sum_{k=1}^{4}\left(\frac{HM(k)_{\text{ret:}i,j} - HM(k)_{\text{ens:}i,j}}{HM(k)_{\text{ret:}i,j} - HM(k)_{\text{ctr:}i,j}}\right)^2\right] \tag{30}$$

Here $HM(k)$,$k = 1{:}4$ represents the 4 hydrometeors. $n$ is the number of collocated pixels for which retrieved and WRF profiles are both available. $j$ is the index for vertical levels which are 14 in number. The denominator of the right side of Eq. 30 is the standard variance, $\sigma^2$ of the study. The value $\sigma^2$ is of the order of the squared error between the

unperturbed WRF forecast and the retrieved atmosphere. This stems from considering the unperturbed WRF forecast as the background information in this assimilation problem. The goal is to maximize L(ensemble). This is the same as minimizing $-\ln(\mathrm{L(ensemble)})$. This problem corresponds to a weighted least squares problem in the parlance of optimization. All the ensembles of a family are given uniform probability and the likelihoods are transferred to Bayesian probabilities and the PPDF is obtained using the following formula:

$$P_{(\mathrm{ensemble}_i)} = \frac{L(\mathrm{ensemble_i})}{\sum_{j=1}^{N} L(\mathrm{ensemble}_j)} \tag{31}$$

where $N$ is the number of ensemble to be considered. The expected a-posteriori gives the best set of initial conditions for the variable(s) under consideration. More details of the Bayesian likelihood estimation and PPDF are available in Balaji (2011). This best set of initial conditions is used to run WRF again from $t = 0$ to $t = 54$ h for the assimilated 54-h forecast. When the corrected set of initial conditions is used, the corresponding boundary conditions have to be updated using WRF Data assimilation (DA) update boundary conditions script. The key novelties of the present work are

1. Assimilation of hydrometeors from rainy pixels using RTE simulations and measured radiances with Bayesian retrievals.
2. Back correction of prognostic variables using the ensemble technique in a Bayesian framework.
3. Estimation of likelihoods from only the collocated pixels thereby considerably reducing interpolation error.

## 8 Results and discussion

Cyclone JAL (2010) is considered here for testing the data assimilation algorithm. To evaluate the efficacy of the assimilation, the simulated cyclone track is compared with the JTWC observed track which is considered as the gold standard in this study. To fine-tune the algorithm, detailed sensitivity studies that include a detailed ensemble study and a variable sensitivity study were performed. The results of these are reported in the ensuing sub sections.

### 8.1 Ensemble sensitivity study

A study was conducted to test the sensitivity of the number of ensembles on the reduction in track errors. The data assimilation algorithm is tested with 25, 50, 75 and 100 ensembles by perturbing the initial conditions of all variables simultaneously. Figure 13 shows the PPDF of the four above-mentioned four different set ensembles of 25, 50, 75 and 100 ensembles, respectively. The abscissa on these plots shows the index of the profile, and the ordinate shows the probability density from the PPDF plots; it is clear that use of the mean of the PPDF is a much better strategy than seeking only the maximum of the posterior. Figure 14 shows the propagation of the simulated cyclones with the JTWC observed and the control run tracks. Table 3 presents the error of the forecast cyclone with respect to the JTWC observed track. Also, shown in the table are 24- and 54-h average errors in the four simulations. From the results of the 25, 50, 75 and 100 ensembles reported in the table, it is

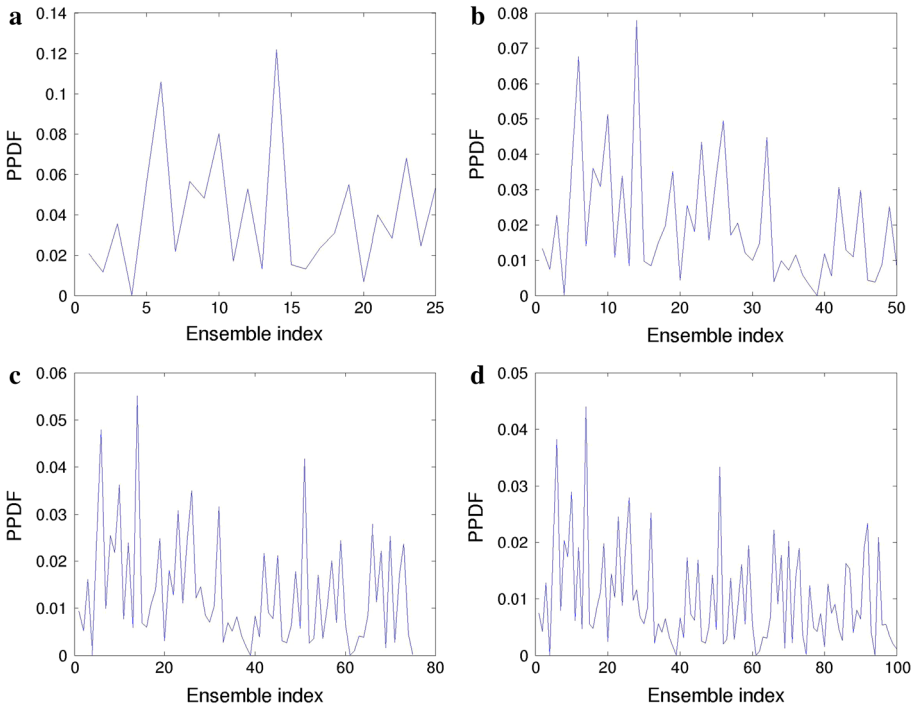


**Fig. 13** PPDF's of the samples in the ensemble sensitivity study for ensemble size **a** 25, **b** 50, **c** 75, **d** 100

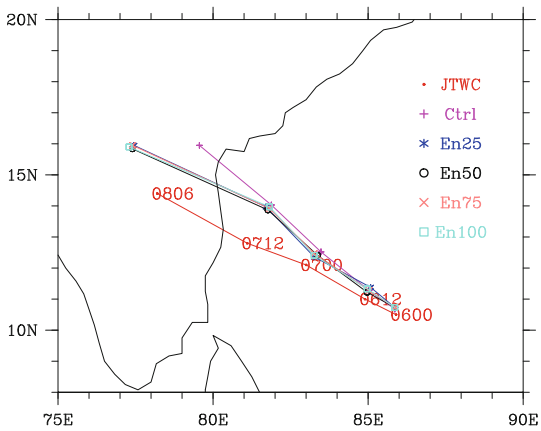


**Fig. 14** Track propagation of cyclone Jal in the ensemble sensitivity study

seen that all ensembles give better results than the control run. Furthermore, ensemble sizes of 25 and 50 give better results compared with ensembles 75 and 100. The average forecast error between an ensemble sizes of 25 and 50 is only 2.3 km for the case of a 24-h forecast and 2.5 km for the case of a 54-h forecast. In view of this, an ensemble family size of 25 ensembles is adequate for assimilating hydrometeors to improve track prediction.

**Table 3** Error in track (km) in the ensemble sensitivity study

| Time UTC | Forecast hour | Ctrl | En25 | En50 | En75 | En100 |
|---|---|---|---|---|---|---|
| 600 | 0 | 20.5 | 26.4 | 26.4 | 26.4 | 26.4 |
| 606 | 6 | 29.5 | 24.4 | 29.5 | 37.9 | 46.4 |
| 612 | 12 | 33.8 | 47.3 | 25.9 | 33.8 | 37.6 |
| 618 | 18 | 68.6 | 44.0 | 58.6 | 54.0 | 63.5 |
| 700 | 24 | 70.8 | 40.6 | 53.7 | 49.8 | 45.0 |
| 706 | 30 | 99.9 | 102.8 | 89.2 | 86.3 | 99.9 |
| 712 | 36 | 161.2 | 146.8 | 139.5 | 147.8 | 152.9 |
| 718 | 42 | 171.8 | 156.3 | 151.0 | 186.3 | 169.7 |
| 806 | 54 | 228.4 | 190.8 | 183.2 | 193.6 | 194.4 |
| | 24 h avg | 44.6 | 36.5 | 38.8 | 40.4 | 43.8 |
| | 54 h avg | 98.3 | 86.6 | 84.1 | 90.6 | 92.9 |

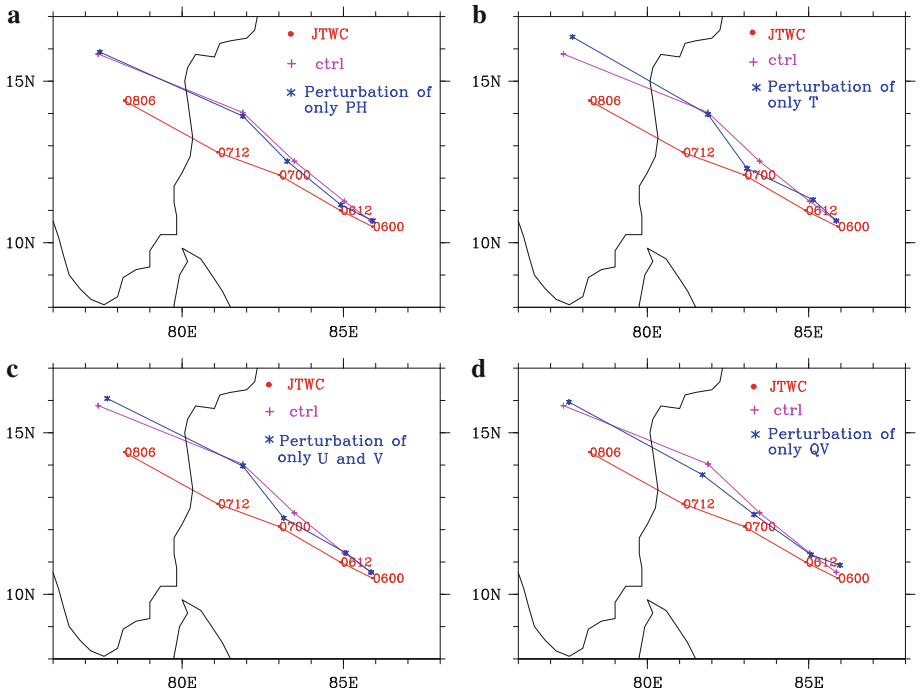


**Fig. 15** Track propagation of cyclone Jal in the variable's sensitivity study perturbation only in **a** PH, **b** T, **c** UV, **d** QV

## 8.2 Variable sensitivity study

The data assimilation algorithm is also tested by correcting the initial conditions of one variable at a time. Figures 15a–d show the track propagation of simulated track along with JTWC and control run track for perturbation of (a) $\phi$, (b) $\theta$, (c) $U$, $V$ and (d) $q_v$, respectively. From the figures, it is clear that large differences arise in the track skill, depending

**Table 4** Track propagation error in the variable sensitivity study

| Time UTC | Forecast hour | Ctrl | PH | T | UV | QV | ALL |
|---|---|---|---|---|---|---|---|
| 600 | 00 | 20.5 | 20.0 | 20.5 | 20.5 | 44.7 | 26.4 |
| 606 | 06 | 29.5 | 33.0 | 24.4 | 27.0 | 33.0 | 24.4 |
| 612 | 12 | 33.8 | 18.7 | 45.2 | 36.8 | 31.9 | 47.3 |
| 618 | 18 | 68.6 | 22.9 | 29.3 | 57.4 | 57.4 | 44.0 |
| 700 | 24 | 70.8 | 54.8 | 24.9 | 33.1 | 53.6 | 40.6 |
| 706 | 30 | 99.9 | 105.4 | 87.1 | 91.7 | 97.1 | 102.8 |
| 712 | 36 | 161.2 | 151.2 | 156.2 | 156.2 | 121.2 | 146.8 |
| 718 | 42 | 171.8 | 177.3 | 171.5 | 150.7 | 156.3 | 156.3 |
| 806 | 54 | 228.4 | 185.5 | 226.7 | 192.7 | 185.8 | 190.8 |
| | 24 h Avg | 44.6 | 29.9 | 28.9 | 35.0 | 44.1 | 36.5 |
| | 54 h Avg | 98.3 | 85.4 | 87.3 | 85.1 | 86.8 | 86.6 |

on the variable that is assimilated. Table 4 presents the track errors resulting from the assimilation of one of these variables one at a time. From the table, it is seen that there is no significant difference in the 54-h average forecast error between them including the one where all variables are assimilated. However, all the simulations produce better results compared with the control run. The 24-h average forecast error, it seem that the perturbation of only the $\phi$ gives a track error of 29.9 km error with while the perturbation of $\theta$ gives an error of 28.9 km, which are very close to each other. Large track errors are seen when only the $q_v$ is perturbed, where in the resulting error is equal to that of the control run. Table 4 also shows that assimilation of all the variables at a time gives a 24-h average forecast error the 36.5 km that lies between the error obtained by assimilating only $\theta$ (lower) and only $q_v$ (highest). So, for a short-term forecast, it is better to assimilate all the variables. However, for a long-term forecast, assimilation of any or all variables leads to similar errors, all of which are lower than the error associated with the control run.

## 9 Conclusions

This study attempted to improve the forecast skill of track prediction in numerical weather prediction models by a Monte Carlo, ensemble-based Bayesian assimilation algorithm. The proposed algorithm was then tested on cyclone JAL. The algorithm assimilates hydrometeors from a combined WRF-radar-radiometer algorithm that was developed in-home based on observations from TRMM. The results are encouraging and show that it is possible to reduce forecast errors significantly in the short term when assimilation is based on microwave observations. From the ensemble sensitivity study, it was seen that the assimilation can reduce track error up to 12 % in a 54-h forecast and up to 18.16 % in a 24-h forecast with an ensemble family size of 25. From a study of sensitivity of the variable to be assimilated it was concluded that the track forecast can be further reduced to 35.20 % in a 24-h forecast with the same ensemble family size of 25. This study has also brought to focus of the impact of GFS boundary conditions on the simulations. From the variable sensitivity study, it was seen that the 54-h average forecast error shows no significant sensitivity to the variables being assimilated because of the strong influence of GFS boundary conditions. Even so, this algorithm can improve the track forecast up to

35 % in a 24-hour forecast, and a potency of full assessment of it can only be ascertained with more studies. This study mainly demonstrated the efficacy of combining an algorithm with a Monte Carlo ensemble-based Bayesian assimilation algorithm for track reduction with a NWP model. Further studies are required to accurately assess the real impact of this assimilation algorithm in cyclone predictions that can be applied in a variety of situations.

## References

Balaji C (2011) Essentials of thermal system design and optimization. Ane Publications, New Delhi

Chandrasekar R, Balaji C (2012) Sensitivity of tropical cyclone Jal simulations to physics parameterizations. J Earth Syst Sci 121(4):923–946

Deiveegan M, Balaji C, Venkateshan SP (2008) A polarized microwave radiative transfer model for passive remote sensing. Atmos Res 88(3):277–293

Mohanty UC, Mandal M, Raman S (2004) Simulation of Orissa super cyclone (1999) using PSU/NCAR mesoscale model. Natural Hazards 31(2):373–390

Pattanaik DR, Rama Rao YV (2009) Track prediction of very severe cyclone Nargis using high resolution weather research forecasting (WRF) model. J Earth Syst Sci 118(4):309–329

Raju PVS, Potty J, Mohanty UC (2012) Prediction of severe tropical cyclones over the Bay of Bengal during 2007–2010 using high-resolution mesoscale model. Natural Hazards 63(3):1361–1374

Ramanujam S, Radhakrishnan C, Subramani D, Chakravarthy B (2012) On the effect of non-raining parameters in retrieval of surface rain rate using TRMM PR and TMI measurements. IEEE J Sel Topics Appl Earth Obs Remote Sens 5(3):735–743

Sandeep S, Chandrasekar A, Dash SK (2007) Impact of modification of initial cyclonic structure on the prediction of a cyclone over the arabian sea. Natural Hazards 41(3):487–499

Singh R, Pal PK, Kishtawal CM, Joshi PC (2008) The impact of variational assimilation of SSM/I and QuikSCAT satellite observations on the numerical simulation of Indian Ocean tropical cyclones. Weather Forecast 23(3):460–476

Skamarock WC, Klemp JB (2008) A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. J Comput Phys 227(7):3465–3485

Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker DM, Wang W, Powers JG (2005) A description of the advanced research WRF version 2. Technical report, DTIC Document

Srinivas CV, Yesubabu V, Hari Prasad KBRR, Venkatraman B, Ramakrishna SSVS (2012) Numerical simulation of cyclonic storms fanoos, nargis with assimilation of conventional and satellite observations using 3DVAR. Natural Hazards 63(2):867–889

Tatarskaia MS, Lataitis RJ, Stankov BB, Tatarskii VV (1998) A numerical method for synthesizing atmospheric temperature and humidity profiles. J Appl Meteorol 37(7):718–729

Zhang Z, Krishnamurti TN (1997) Ensemble forecasting of hurricane tracks. Bull Am Meteorol Soc 78(12):2785–2795

Zhang Z, Krishnamurti TN (1999) A perturbation method for hurricane ensemble predictions. Mon Weather Rev 127(4):447–469

Zhao Y, Wang B, Ji Z, Liang X, Deng G, Zhang X (2005) Improved track forecasting of a typhoon reaching landfall from four-dimensional variational data assimilation of AMSU-A retrieved data. J Geophys Res 110:D14101