# Impact of Ensemble Derived Flow-dependent Background Error Covariance in a Data Assimilation System for Regional-scale NWP model

*A thesis submitted*
*in partial fulfillment for the degree of*

**Doctor of Philosophy**

***by***

**Rekha Bharali Gogoi**


**Department of Earth and Space Sciences**

**Indian Institute of Space Science and Technology**

**Thiruvananthapuram - 695547, India**

**November**

सत्यमेव जयते

भारतीय अंतरिक्ष विज्ञान एवं प्रौद्योगिकी संस्थान
(वि.अ.आयोग अधिनियम 1956 की धारा-3 के अधीन मानित विश्वविद्यालय घोषित)
भारत सरकार, अंतरिक्ष विभाग, वलियमला पोस्ट, तिरुवनंतपुरम 695 547 भारत

INDIAN INSTITUTE OF SPACE SCIENCE AND TECHNOLOGY
(A Deemed to be University u/s 3 of the UGC Act, 1956)
Government of India, Department of Space
Valiamala P. O, Thiruvananthapuram 695 547 India

www.iist.ac.in

# CERTIFICATE

This is to certify that the thesis entitled "**Impact of Ensemble Derived Flow-dependent Background Error Covariance in a Data Assimilation System for Regional-scale NWP model**" by **Rekha Bharali Gogoi** to the Indian Institute of Space Science and Technology, Thiruvananthapuram, in partial fulfillment for the award of the degree of **Doctor of Philosophy** is a bona fide record of research work carried out by her under our supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institution or University for the award of any degree or diploma.

10/11/2021

**Dr. Arup Borgohain**
Co-supervisor
North Eastern Space Applications Centre
Umiam, Meghalaya
November 2021

16/11/2021.

**Dr. Govindan Kutty M**
Supervisor
Department of Earth and Space Sciences
Thiruvananthapuram, Kerala
November 2021

10.11.21

Counter signature of Director with seal

डॉ. एस पी अग्रवाल / Dr. S P Aggarwal
निदेशक / Director
उत्तर पूर्वी अंतरिक्ष उपयोग केंद्र/North Eastern Space Applications Centre
भारत सरकार/Govt. of India. अंतरिक्ष विभाग/Dept. of Space
उमियम/Umiam - 793103, मेघालय/Meghalaya

16/11/2021

Counter signature of HOD with seal

विभागाध्यक्ष / Head of the Department
पृथ्वी एवं अंतरिक्ष विज्ञान विभाग
Department of Earth and Space Sciences
भारतीय अंतरिक्ष विज्ञान एवं प्रौद्योगिकी संस्थान
Indian Institute of Space Science and Technology
अंतरिक्ष विभाग, भारत सरकार
Dept. of Space, Govt. of India
तिरुवनंतपुरम / Thiruvananthapuram - 695 547

# DECLARATION

I declare that this thesis entitled "**Impact of Ensemble Derived Flow-dependent Background Error Covariance in a Data Assimilation System for Regional-scale NWP model**" submitted in partial fulfillment of the degree of Doctor of Philosophy is a record of original work carried out by me under the supervision of **Dr. Govindan Kutty M** and **Dr. Arup Borgohain** and has not formed the basis for the award of any other degree or diploma, in this or any other Institution or University. In keeping with the ethical practice in reporting scientific information, due acknowledgments have been made wherever the findings of others have been cited.

Thiruvananthapuram-695547
November 2021

Rekha Bharali Gogoi
(SC15D027)

# ACKNOWLEDGEMENTS

Here, I want to start with a famous quote by Swami Vivekananda-"Take up one idea. Make that one idea your life; dream of it; think of it; live on that idea". I had also taken up an idea to do a Ph.D. from IIST, and a journey started from the north-eastern corner of India to the extreme southern part of India. During that journey, I had faced many ups and downs, but there were many hands to support me in living my dream.

First of all, I would like to extend my sincere gratitude to my supervisor Dr. Govindan Kutty M for his continuous support throughout my Ph.D. journey. I would like to thank him for showing patience and motivating me to achieve the targeted goal on time. Moreover, his immense knowledge of the research topic has made my job much easier. Due to the sudden spread of the COVID pandemic across the globe, I could not physically visit IIST. But, whenever I had any doubts, Sir was always there to answer my questions over the mail or call. It would not have been possible to bring the thesis to this shape without his guidance.

My co-supervisor, Dr. Arup Borgohain has also supported me in doing my best to successfully complete my Ph. D. work. I would like to acknowledge the support provided by him during my Ph.D. journey. I would also like to thank the respected members of my doctoral committee: Dr. Anantharaman Chandrasekar, Dr. Radhika Ramachandran, Dr. Samir Mandal, and Dr. Sarvesh Rajput, for their timely review of my research work and insightful comments.

My sincere thanks goes to Director, IIST, for giving me the opportunity to pursue Ph.D. at IIST and Director, NESAC, for allowing me to enroll for Ph.D. Here, I would also like to mention Dr. PG Diwakar, former Director of NESAC. His short stint of six months at NESAC paved the way

for my Ph.D. degree. Getting study leave on the job is never easy as we are assigned many responsibilities. But according to Diwakar Sir, higher studies are always crucial for employees for the organization's growth. Thank you, Sir, for believing in me. I also like to take the opportunity to thank Director, IITM, Pune, for allowing me to access their High Power Computing facility for the timely completion of my research work. Further, I want to mention Aswin Maheswar, Lab Assistant, Department of Earth and Space Sciences. Since, I could not go to IIST because of the pandemic; he has taken all care regarding printing and binding of my work report at the time of submission of thesis as well as pre-synopsis. I would like to extend my deepest thanks to Mr. Aswin for all his support.

It will be incomplete if I don't mention my friends, who are constantly very supportive. First, I would like to name my best friend, Achinta Kumar Gogoi, who has seen me in the agony I have gone through at times during this expedition. He is always by my side whenever I need his suggestions and support. I am very grateful to Achinta for everything he did for me. Second, I would like to thank my young friends Pragya, Rohith, Deepak, and Reshmi from IIST for their suggestions and support. I would also like to thank my childhood friends Asomi, Sangeeta, Lipsikha, Mitali, Meghali, and Chanda, who are always happy about my success. A special thanks to my colleagues, especially, Dr. Jonali Goswami, Dr. Dibyajyoti Chutia, and Dr. Shyam S. Kundu for their constant suggestions. Whenever I faced any hurdles Jonali Ma'am was always there to support me.

Last but not least, it's my family, the central pillar of my success. When I decided to pursue a Ph.D., I was the mother of three years old child. I was in a dilemma; while I wanted to do a Ph.D. from IIST, I didn't want to leave my child alone. The research work at IIST and my official responsibility at NESAC were complimenting each other. The Ph.D. would allow me to learn

more so that I could implement the same for my organizational growth. So, I didn't want to miss the chance. However, it was mandatory to spend six months at the IIST campus to complete the course work. But as a mother, my daughter is paramount. Nevertheless, my in-laws and parents turned out to be the savior. They support me throughout my journey without uttering a single word. While my parents have allowed me to get enough education to get a coveted job in ISRO, my in-laws have allowed me to go further. Without their support, the journey of my Ph.D. would not have been possible. “Thank you” is a very simple word to express my gratitude towards my parents, in-laws, husband, and an understanding daughter Pahi. I wholeheartedly thank all the other members of my family, especially my sisters Doli and Rashmi who are always proud of my small deeds.

The whole journey taught me that nothing is impossible; only we need to start working for fulfilling our dreams. Finally, a big thanks to God for everything and thanks to everyone else I couldn't name here.

Rekha Bharali Gogoi

# Abstract

The background error covariance (BEC) in Three-dimensional variational (3DVAR) data assimilation (DA) system is static and lacks information about the flow-conditions of the day while advanced DA algorithms like Ensemble Kalman Filter provides flow-dependent BEC information. A hybrid ensemble–variational (HYBRID) DA system combines the flow-dependent information from the ensemble DA system in the variational DA framework. The research work examines the impact of ensemble derived flow-dependent ensemble error covariance in 3DVAR DA system for short-range forecasts using Weather Research and Forecast model. The forecasted wind, temperature, and rainfall from the assimilation experiments are verified against corresponding observations. Evaluation of short-range forecasts during the Indian summer monsoon indicates that the flow-dependent ensemble BEC in 3DVAR has systematically improved the forecast when compared to traditional 3DVAR experiments. More specifically, the rainfall forecast skill is superior in HYBRID experiments as compared to 3DVAR. The rainfall forecasts in convection-permitting resolution are validated against 746 telemetric rain gauge observations over the Karnataka state shows higher quantitative precipitation forecast skill in HYBRID system than 3DVAR towards the later stages of DA cycling. Assessment of short-range forecast of landfalling tropical cyclones (TC) over the Bay of Bengal (BoB) indicates that the use of flow-evolving ensemble error covariance in the 3DVAR system has reduced the TC position and intensity errors in the analysis. However, adding more weights to the ensemble error covariance term in the 3DVAR cost function has not shown any significant improvements. The forecasts from HYBRID analysis outperform 3DVAR forecasts by reducing TC track forecast error. The impact of assimilated observations can be dependent on many factors in a DA system including data quality control, preprocessing, skill of the model, and the DA algorithm. Studies have been conducted to

understand the differences and similarities in the impact of observations assimilated by two different DA algorithms. Evaluations on the impact of INSAT-3D Atmospheric Motion Vectors (AMV) observations with and without flow-dependent BEC in 3DVAR DA system are performed for short-range forecasts of the Indian summer monsoon. The satellite AMV observations show a more relative impact in HYBRID, with 77% and 71% improvement for wind and tropical temperature as compared to 3DVAR. Incorporating AMV shows substantial improvement in the forecast of landfalling TC in HYBRID than in 3DVAR DA system. Furthermore, the assimilation of AMV observation in HYBRID shows improved skill scores for quantitative precipitation forecast. The final objective of the thesis is to quantify the impact of flow-evolving BEC for the convective scale DA system. In this study, retrieved rainwater and water vapor from radar reflectivity are incorporated in HYBRID DA system to understand thunderstorms over northeastern region of the Indian subcontinent. The results indicate that assimilation of radar observations enhances the quantitative precipitation skill scores in both HYBRID and 3DVAR DA system. No significant impact of HYBRID DA system in rainfall forecast is observed.

# Table of Contents

## LIST OF FIGURES WTH CAPTIONS

# CHAPTER 1

# Introduction

The Numerical Weather Prediction (NWP) model is being widely utilized for weather forecasts and the ability of forecast the weather accurately has significantly increased, recently. The NWP is an initial value problem, and its ability to represent the future state of atmosphere depends primarily on the initial conditions (IC). Data assimilation (DA) is the mathematical approach that optimally combines the information of observations with the model information to obtain a precise initial state (Daley 1997; Li and Xu 2011). Different DA techniques namely successive correction method (SCM), nudging, optimal interpolation (OI), Kalman filter, and variational methods have been developed since 1950 (Kalnay 2003). Advanced DA techniques employed in NWP models are found to be more effective in prescribing accurate initial state of the system (Brennan et al. 2009; Hsiao et al. 2012; Wang et al. 2013c; Counillon et al. 2014; Fierro et al. 2015; Rakesh et al. 2015; Zhang et al. 2015), and therefore the SCM and OI methods soon have been replaced by modern-day variational and ensemble Kalman filter (EnKF) methods for the operational forecast.

The variational DA techniques such as three-dimensional variational (3DVAR) and four-dimensional variational (4DVAR) systems have been in use operationally at most NWP centers for more than a decade (e.g., Parrish and Derber 1992; Lorenc et al. 2000; Kleist et al. 2009). The 3DVAR method is mainly based on the iterative minimization of a quadratic cost function to obtain the best estimate of the atmospheric state or the analysis. The cost function measures

the distance of the analysis from the model short-range forecast, commonly known as background and the observations. The fundamental difference between 3DVAR and 4DVAR is that 4DVAR includes integration of adjoint and tangent linear model during the cost function minimization procedure, and thereby incorporating the implicit flow-dependent error covariance in the DA process (e.g., Sun and Wang 2013; Chu et al. 2013; Wang et al. 2013a; Mazzarella et al. 2017). However, the computational cost involved in performing 4DVAR is significantly higher than the 3DVAR DA system (Yang et al. 2014). The three-dimensional variational (3DVAR) data assimilation (DA) system is cost-effective because it uses a background error covariance (BEC) matrix that is either completely static or only weakly coupled to the dynamics of the forecast model (e.g., Barker et al. 2004). However, the drawback of static BEC is that the observations assimilated will make local, isotropic increments without proper flow-dependent extrapolation. Several studies examined the role of static BEC in the performance of the 3DVAR DA system (Rakesh and Goswami 2011; Routray et al. 2016).

More advanced DA algorithms like EnKF (e.g., Houtekamer and Mitchell 2001; Evensen 2003) incorporate flow-dependent BEC information that would provide "errors of the day," which is estimated from the ensemble of nonlinear model forecasts. The ensemble covariance assigns appropriate weight to the observation information depending on the flow of the day. However, the major shortcoming of the EnKF DA system is that it introduces spurious long-distance correlation and underestimates the magnitude of error variance when the ensemble size is limited (Hamill et al. 2001; Barker et al. 2011). Lorenc (2003) has listed issues that are related to the sampling error in ensemble based DA schemes and proposed an extended control variable approach to incorporate ensemble covariance matrix in the variational framework. The

representation of flow-dependent ensemble BEC in a variational framework is initially studied by Barker (1999), which is found to be effective than EnKF DA approach even with limited computational resources. Several authors have also reported benefits of HYBRID compared to a standalone EnKF DA system (e.g., Campbell et al. 2010; Wang 2010).

Hamill and Snyder (2000) tested the effectiveness of flow-dependent covariance developed from the ensemble of short-range forecasts in the variational DA framework and found improvements in the analysis with respect to 3DVAR. The approach which incorporates ensemble-generated BECs in variational framework is commonly referred to as a hybrid ensemble-variational DA system (hereafter HYBRID). Wang et al. (2008) has introduced a HYBRID data assimilation system for Weather Research and Forecast (WRF) model based on WRF 3DVAR. The preliminary results from observation system simulation experiments reveal that the HYBRID analysis are better than 3DVAR and the improvements are more significant over data sparse region. The subsequent study with assimilation of real observations demonstrates better wind as well as temperature forecast using the flow dependent HYBRID DA system compared to 3DVAR (Wang et al. 2008b). The improvement is due to the fact that the HYBRID system is capable of adjusting the background over the data-sparse region using observation from data-rich region by following the background flow. Similarly, Prasad et al. (2016) examined the performance of HYBRID DA scheme in National Center for Medium Range Weather Forecasting (NCMRWF) Global Forecast System over the Indian region and found that HYBRID DA significantly improved the forecasted temperature and wind field, particularly over the upper troposphere. Another benefit of HYBRID is reported by Gao et al. (2013) for storm-scale simulations using radar DA. The results indicate that HYBRID can fit the radar reflectivity better than the 3DVAR and EnKF which in turn reduce the storm spin-up time.

The past literatures clearly advocate the benefit of using HYBRID DA compared to stand-alone EnKF and 3DVAR DA system for simulation of weather events at different scales ranging from global to convective scale. However, very few systematic study has been done over the Indian region using a limited area model. The current study investigates the impact of ensemble derived flow-dependent error covariance in a DA system over the Indian province by comparing the forecast performances of 3DVAR and HYBRID DA approaches.

It is known that the impact of observations may vary depending on many factors in the DA assimilation system, such as data quality control, preprocessing, and specification of BEC. To understand this further, the present study addresses the impact of Atmospheric Motion Vectors (AMV) in 3DVAR and HYBRID DA system. AMV is satellite-derived wind observations obtained by continuously tracking regions of clouds or water vapor using satellite images. AMV provides wind information with good areal coverage, particularly over the data-sparse oceanic area. Several studies have shown the benefit of assimilating AMVs on improving the weather forecasts over the tropics (e.g.,Velden et al. 1992; Leslie et al. 1998; Soden et al. 2001; Rani and Das Gupta 2014; Mounika et al. 2018). A study by Kaur et al. (2015) during the Indian summer monsoon (ISM) reported the positive impact of Kalpana-1 AMVs assimilation over the tropical region using 3DVAR. Deb et al. (2016) demonstrate the reduction of track forecast errors of the cyclonic storm NANAUK over the Arabian Sea when INSAT-3D AMV observations are assimilated. Another study by Kumar et al. (2017) reports improved model simulated wind speed, temperature, and moisture analyses and subsequent model forecasts over the Indian Ocean, Arabian Sea, Australia, and South Africa with INSAT-3D AMV DA in WRF-3DVAR. Zhang et al. (2018) demonstrate the impact of assimilation of enhanced AMVs using HYBRID ensemble–variational DA in improving the track and intensity forecasts of hurricanes. Wu et al.

(2014) demonstrate the favorable impact of assimilation of AMVs from Multifunctional Transport Satellite (MTSAT) using ensemble Kalman filter in the track forecast of Typhoon Sinlaku. Sawada et al. (2019) show improvement in tropical cyclone forecaston assimilating high-resolution AMVs derived from Himawari-8 in the HYBRID DA system. The study reveals that incorporating AMVs using background error covariance created from the HWRF ensemble forecast contributes to a significant reduction in negative intensity bias and error. There is a substantial benefit to the tropical cyclone (TC) size forecast. Also, Sawada et al. (2020), using an HYBRID DA system, investigate the assimilation impact of rapid-scan (RS) AMVs derived from Himawari-8 on three TC forecasts in the western North Pacific during 2016. The results show that the assimilation of RS-AMVs can improve the track forecast skill, while the weak bias or slow intensification bias increases at the shorter forecast lead time. Previous studies have shown that in the presence of flow evolving BEC in HYBRID DA system the observations are effectively assimilated compared to traditional 3DVAR DA systems. However, it has been observed that no such study has been conducted for the Indian domain using satellite data in the HYBRID DA system.

Apart from satellite DA, further studies have been conducted using high resolution radar observations for the forecast of thunder storms using 3DVAR and HYBRID DA systems. Radar data assimilation has been proved to be valuable for the convective scale weather forecast in many previous studies using different data assimilation techniques such as Newton relaxation approximation (Jones and Macpherson 1997), 3DVAR (Abhilash et al. 2007; Routray et al. 2010; Wang et al. 2013a), 4DVAR ( Sun and Zhang 2008; Sun and Wang 2013;Wang et al. 2013b; Thiruvengadam et al. 2020) and ensemble Kalman filter (EnKF) (Dowell et al. 2011; Gao and Min 2018; Tong et al. 2020) . The radar observations can be assimilated into the model either by

the direct method (Xiao et al. 2005) or indirect (Wang et al. 2013a). In the direct method, the radar reflectivity and radial wind directly ingested into the numerical model using the DA techniques while in the indirect method, the derived parameters such as rainwater, water vapor from radar reflectivity is assimilated (Li et al. 2018). A study by Wang et al. (2013a) has pointed out drawbacks of direct assimilation of radar observations for simulation of the convective system, especially when IC depict a dry environment unfavorable for convection initiation. The performance of direct assimilation of radar observations are found to be not very remarkable when the magnitude of rainwater in the reflectivity-rainwater relations is smaller. It is found that for dry environments, the warm rain process may not be turned on during the DA process, and the direct assimilation of radar observations may fail to capture the convection (Zhang et al. 2009). However, indirect assimilation of radar observations is expected to provide an environment conducive for convection even when the first guess is dry due to the assimilation of derived water vapor from radar reflectivity (Wang et al. 2013a).

In this study, radar data assimilation experiments are conducted to simulate thunderstorms developed over India's north-eastern region (NER). The NER region of India exhibits a complex topography with two-thirds of the area covered by hilly terrain. Also, this region experiences the highest thunderstorm activity compared to the rest of India. Such occurrence is supported by local influences such as orography of the region, insolation, and moisture advection in the presence of suitable wind conditions (Tyagi 2007). A study has been performed to understand the impact of indirect assimilation of radar data observations in the forecast of thunderstorms over NER region.

## 1.1 Objective of this thesis

The impact of flow-dependent error covariances compared to static BEC is well documented in several studies for different geographical domains globally. However, the impact of error covariance on a DA system may depend on the flow regimes under consideration. It is in this context a systematic study has been undertaken to estimate and quantify the impact of flow dependent ensemble error covariance in a DA system over the Indian subcontinent during various seasons using a hybrid ETKF-3DVAR DA approach. The specific objectives of this study are listed below:

- Quantifying the impact of flow-dependent ensemble error covariance on the short range forecasts of Indian summer monsoon Season in a limited-area model.
- Quantifying the impact of flow-dependent ensemble error covariance on the forecast of land-falling tropical cyclones formed over the Bay of Bengal
- Identifying the difference in the impact of INSAT-3D Atmospheric Motion Vectors when ensemble error covariance is used in 3DVAR framework on the forecasts of Indian summer monsoon rainfall
- Identifying the difference in the impact of INSAT-3D Atmospheric Motion Vectors when ensemble error covariance is used in 3DVAR framework on the forecasts of land-falling tropical cyclones over the Bay of Bengal
- Understanding the role of ensemble error covariance in convection-permitting resolution using Radar data assimilation for heavy rainfall forecasts

The thesis is organized as follows.

Chapter 2 summarizes the limited-area NWP model and DA techniques used in this study. Additionally, the methodologies used for validation of model analysis and forecast are also included in this chapter. Chapter 3 contains the results obtained from the impact study of flow-dependent ensemble error covariance in DA system for the short range forecast of Indian summer monsoon season short-range forecasts. Results from the impact of flow-dependent ensemble error in DA system covariance on the forecast of land-falling tropical cyclones formed over the Bay of Bengal are shown in Chapter 4. Chapter 5 shows the impact of INSAT-3D Atmospheric Motion Vectors when ensemble error covariance is used in 3DVAR framework on the forecasts of Indian summer monsoon rainfall. Chapter 6 explores the difference in the impact of INSAT-3D Atmospheric Motion Vectors in the presence and absence of ensemble error covariance in 3DVAR framework on the forecasts of land-falling tropical cyclones over the Bay of Bengal. The results from radar data assimilation using 3DVAR and HYBRID DA system for heavy rainfall forecasts over north-eastern region of Indian subcontinent are shown in Chapter 7 while the conclusions and future directions of this study are listed are listed in Chapter 8.

# Chapter 2

# Data and Methodology

Numerical Weather Prediction (NWP) models predict the weather in advance by solving differential equations representing the evolution of the atmosphere, with a given set of initial conditions. The performance of NWP models can be enhanced by improving the initial conditions using data assimilation (DA) techniques. This study uses a widely used NWP model named the Weather Research and Forecasting (WRF) model for model simulations and two DA systems available with the WRF model, namely 3DVAR and HYBRID are also used. A brief introduction to the WRF model and data assimilation methods used in this study is given in this chapter. This chapter also describes the data used for model initialization as well as data used for assimilation. Finally, the chapter concludes with details about the different verification procedure used for model validation.

## 2.1 Limited-area Model Overview

WRF model is a next-generation mesoscale non-hydrostatic NWP model developed jointly by the National Oceanic and Atmospheric Administration (NOAA), the Naval Research Laboratory (NRL), the University of Oklahoma, the Air Force Weather Agency (AFWA), and the Federal Aviation Administration (FAA). It is a mathematical model that represents the atmosphere based on the atmospheric governing equations. The model serves as a common platform for real-time numerical weather forecast, regional climate modeling, idealized simulations, data assimilation, and other atmospheric research, with the ability to resolve meteorological systems with spatial scales ranging from tens of meters to thousands of kilometers.

There are two central dynamical cores in the WRF model: the ARW (Advanced Research WRF) and the NMM (Non-hydrostatic Mesoscale Model). While the ARW is maintained by NCAR's Mesoscale and Microscale Meteorology (MMM) Laboratory, the NMM is maintained by the National Centers for Environmental Prediction (NCEP). The study uses the WRF-ARW Modeling System, which mainly consists of the WRF Preprocessing System (WPS), ARW solver, WRF-VAR, and post-processing graphics tools (Skamarock et al. 2019) as shown in Figure2.1.

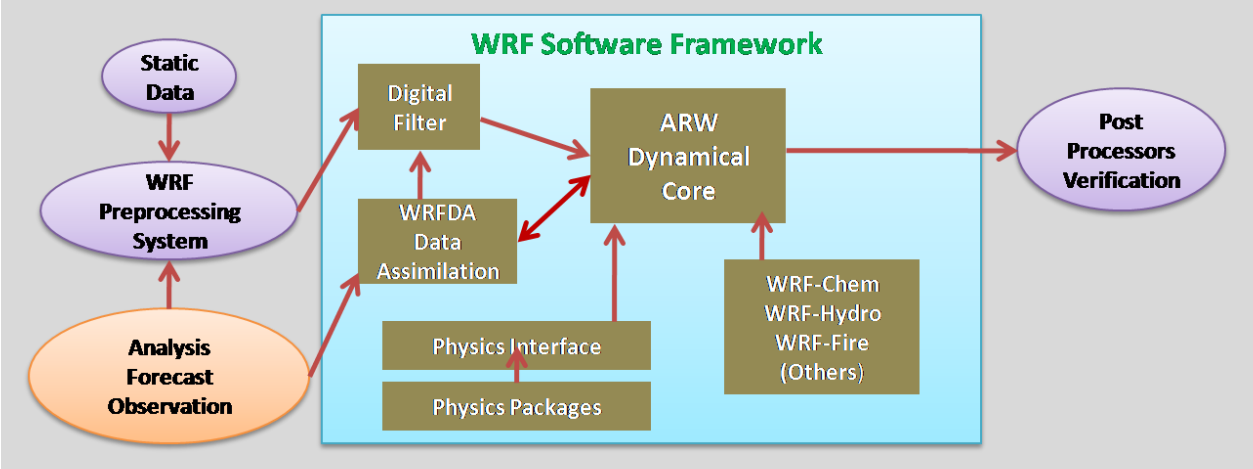


**Figure 2.1: WRF-ARW framework**

### 2.1.1 WRF Preprocessing System (WPS)

WPS mainly consists of three subsystems: Geogrid, Ungrib, and Metgrib. The different functionalities of each component of WPS are: (i) Geogrid creates the simulation domain with proper projection information (WRF-ARW currently supports Lambert conformal, polar stereographic, Mercator, and latitude-longitude projections) and interpolate the terrestrial data such as terrain, landuse, and soil types to the particular domain; (ii) Ungrib does decompress and

read meteorological data obtained from global models in GRIB format, and re-write in an intermediate form; (iii) Metgrid reads the intermediate data and interpolates the meteorological information to the simulation domain.

### 2.1.2 ARW solver

The ARW dynamic solver includes Eularian fully compressible non-hydrostatic equations in flux form with hydrostatic options and supports different nesting options: two-way nesting with multiple nests and nest levels, one-way nesting and moving nests.

The flux form Eularial equations of WRF-ARW are given below,

$$\partial_t U + (\mathbf{\nabla}.\mathbf{V}u) - \partial_x\left(p\phi_\eta\right) + \partial_\eta(p\phi_x) = F_U \tag{2.1}$$

$$\partial_t V + (\mathbf{\nabla}.\mathbf{V}u) - \partial_y\left(p\phi_\eta\right) + \partial_\eta\left(p\phi_y\right) = F_V \tag{2.2}$$

$$\partial_t W + (\mathbf{\nabla}.\mathbf{V}w) - g\left(\partial_\eta p - \mu\right) = F_W \tag{2.3}$$

$$\partial_t \Theta + (\mathbf{\nabla}.\mathbf{V}\theta) = F_\Theta \tag{2.4}$$

$$\partial_t \mu + (\mathbf{\nabla}.\mathbf{V}) = 0 \tag{2.5}$$

$$\partial_t \phi + \mu^{-1}[(\mathbf{V}.\mathbf{\nabla}\phi) - gW] = 0 \tag{2.6}$$

In the equations 2.1 to 2.3, η denotes the terrain-following hydrostatic-pressure vertical coordinate and defined as $\eta = (p_d - p_t)/\mu$ , where $\mu = p_s - p_t$. Here, $p_d$ is the hydrostatic component of the pressure of dry air and $p_s$ and $p_t$ represents values of $p_d$ along the surface and top boundaries of the model domain, respectively.

The flux form variables in the equations are defined as $\mathbf{V} = \mu \mathrm{v} = (U, V, W)$, $\Omega = \mu\dot{\eta}, \Theta = \mu\theta$, where $\mu(x, y)$ is the mass per unit area within the atmospheric column in the model domain (x,y). $\mathrm{v} = (u, v, w)$ refers to the covariant velocities in the two horizontal and vertical direction and $\omega = \dot{\eta}$ refers to thecontravariant vertical velocity. Among the other variables available in the equations, $\theta$ is the potential temperature, $\phi = gz$ is the geopotential, p is the pressure and $F_U, F_V, F_W, F_\Theta$ are the forcing variables occurring from model physics, turbulent mixing, spherical projections, and the Earth's rotation.

Further, the diagnostics relation for inverse density and equation of state are given as follows:

$$\partial_\eta \phi = -\alpha\mu \qquad (2.7)$$

$$p = p_0(R_d\theta/p_0\alpha)^\gamma \quad (2.8)$$

Here, $R_d$ is the gas constant for dry air, $p_0$ is the reference pressure and $\gamma$ is the ratio of the heat capacity for dry air. ARW uses Arakawa C-grid staggering option to discretize the equations horizontally and terrain-following vertical coordinates for vertical grid spacing. Finally, the numerical integrations are performed during Runge-Kutta 2$^{nd}$ and 3$^{rd}$order time-step options.

The atmosphere is a chaotic system comprised of different processes occurring at varied scales ranging from planetary scale to molecular diffusion scale. It is not possible to explicitly represent small-scale processes using mathematical models either due to the incomplete understanding of the physical processes or the lack of computational facility, leading to simplifying the equations. While the dynamic core solves the mathematical equations on the discrete grids, it neglects the small-scale processes within the grid, which are of utter importance for representing convective scale processes, land-atmosphere interaction processes, etc. The parameterization schemes in the

Physics package of the WRF model takes care of the sub-grid scale activities. The dynamic core calls the required Physics package through an interface as shown in Figure 2.1 for its use to represent the unresolved processes.

Some of the important parameterization schemes in the WRF model can be listed as microphysics, cumulus convection, planetary boundary layer (PBL), land-surface model, and radiation. The interactions among the various parameterization schemes are shown in Figure 2.2.

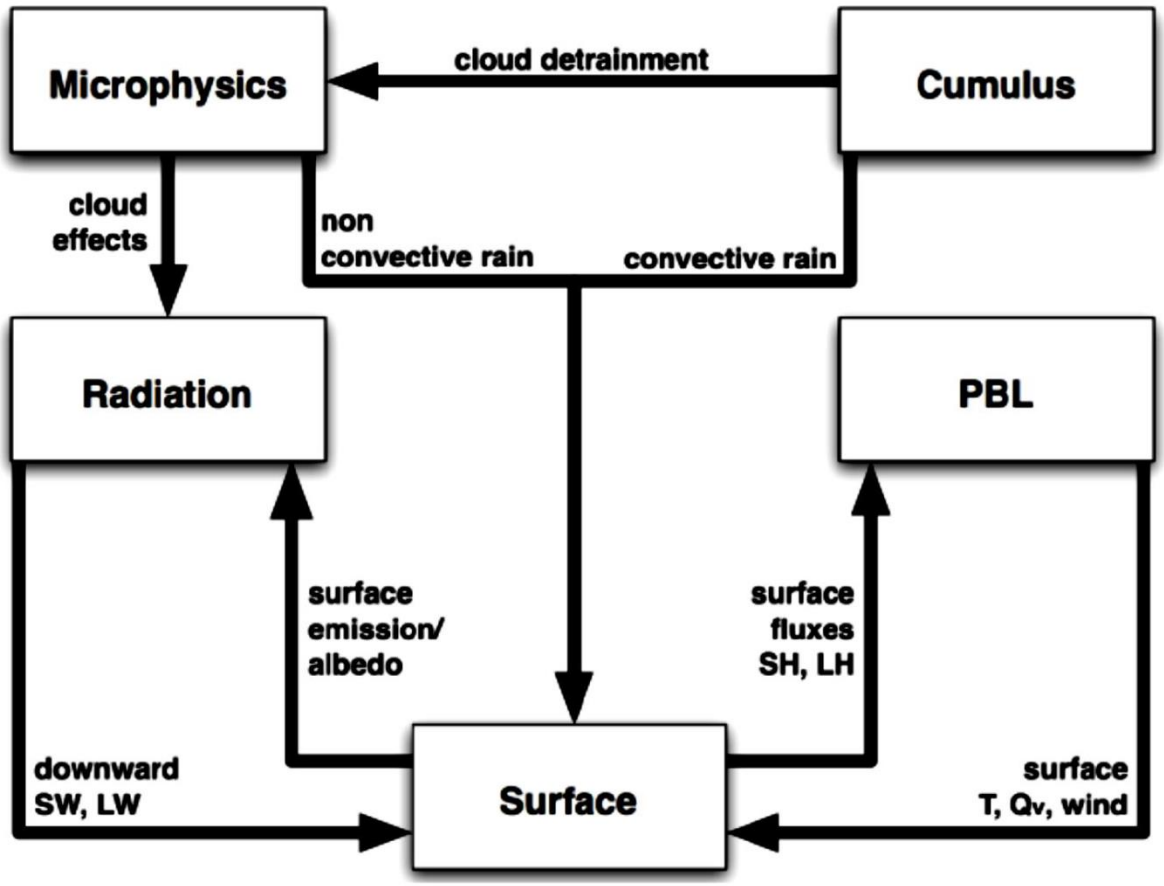


**Figure 2.2:** Schematic depiction of the communications between different parameterization schemes in WRF model

*Microphysics Schemes*

Microphysics parameterization (MP) schemes resolve the grid-scale or explicit water vapor, cloud, and precipitation processes in the model. The MP schemes available with the WRF model can be categorized into the single-moment and double-moment. While the single-moment MP schemes predict only the mixing ratios of the hydrometeors, the double-moment strategies predict both the mixing ratio and the number concentrations of the hydrometeors (Lim and Hong 2010). However, the double moment schemes are computationally expensive compared to single moment schemes. In this study, a mixed phase WRF single moment MP scheme is used which considers five classes of hydrometeors namely, cloud ice, cloud water, water vapor, snow and rain (Hong et al. 2004). Therefore, it is symbolized as WSM5. It allows super cooled liquid water to form and let the snow to melt gradually below the melting point.

*Cumulus Schemes*

The model unresolved sub-grid scale cloud and precipitation processes such as vertical fluxes due to unresolved updrafts and downdraft entrainment and detrainment processes, compensating motion outside the clouds, etc., are handled by the cumulus parameterization (CP) schemes. The current study uses a mass flux based scheme called Kain-Fritsch (KF) scheme (Kain 2004)for cumulous parameterization. The scheme is responsible for estimating the existing instability in the atmosphere along with the probable characteristics of the convective clouds.

*Surface Layer*

The atmospheric surface layer is the lowest part of the planetary boundary layer (PBL) in contact with the Earth's surface consists of almost 10% of the PBL (Stull 1988). Its accurate

representation is essential to determine the land–atmosphere interaction precisely. In the WRF-ARW model, the land-surface model and PBL schemes receive the friction velocities and exchange coefficient information from the surface layer scheme to calculate surface heat and moisture fluxes and surface stress, respectively.

*Land-Surface Model*

All the physical schemes are related to the land-surface models (LSMs) in some way or other. The LSM models provide heat and moisture fluxes over land points and sea-ice points by incorporating atmospheric information from the surface layer scheme, radiative forcing from the radiation scheme, and precipitation forcing from the microphysics and convective schemes, collectively with internal information on the land's state variables and land-surface properties. The flux information act as the lower boundary condition in the PBL schemes for vertical transport. The LSMs also handle the thermal and moisture fluxes in multilayer soil, together with 82 vegetation, root, and canopy effects and surface snow-cover prediction. It does not provide tendencies, but responsible for updating the land's state variables which include soil temperature profile, the ground (skin) temperature, soil moisture profile, snow cover, and possibly canopy properties. In WRF-ARW model, the LSMs are regarded as one-dimensional column models, as there is no horizontal interaction between neighboring points in the LSM. The WRF-ARW includes seven different LSM options and the most popular one is Noah LSM. The LSM is a 4-layer soil temperature and moisture model with canopy moisture and snow cover prediction. It accounts for root zone, evapotranspiration, soil drainage and runoff, vegetation categories, monthly vegetation fraction, and soil texture. It provides sensible and latent heat fluxes and also predicts soil ice, and fractional snow cover effects.

*Planetary boundary layer*

The PBL schemes account for the columnar vertical sub-grid scale fluxes due to eddy transport and also handle vertical diffusion. By incorporating surface flux information from the surface layer and land-surface schemes and horizontal diffusion information from appropriate horizontal diffusion option in the WRF-ARW model, PBL schemes estimates the flux profile within the well-mixed boundary layer and the stable layer. Consequently, PBL schemes provide temperature, moisture (including clouds), and horizontal momentum tendencies in the entire atmospheric column. There are twelve PBL schemes available in the WRF-ARW model and the commonly used schemes are Yonsei University (YSU) PBL and Mellor-Yamada-Janjic (MYJ) PBL. In this study the YSU PBL scheme is used which is the modified version of the medium-range forecast (MRF) PBL, with the difference that YSU PBL includes counter-gradient terms to represent fluxes due to non-local gradients. A study by Hong et al. (2006) demonstrates that the modification in YSU PBL results in a well-mixed boundary-layer profile; on the contrary, MRF PBL produces a pronounced over-stable structure in the upper part of the mixed layer.

*Atmospheric Radiation*

Solar radiation is the primary source of energy for the Earth and atmosphere and is responsible for the heat, mass, and momentum fluxes in the free atmosphere and at the surface. It is also the driving force for atmospheric circulations. The primary goal of the radiative parameterization is to estimate the total radiative flux at a particular location by encompassing the effects of radiative flux divergence and surface downward longwave and short-wave radiation for the ground heat budget. WRF-ARW includes seven longwave schemes and eight shortwave schemes for radiation parameterization and among them Rapid Radiative Transfer Model (RRTM) scheme is

commonly used for longwave parameterization and Dudhia scheme for short-wave parameterization. RRTM scheme uses look-up tables to accurately represent the processes due to water vapor, ozone, $CO_2$, trace gases, and cloud optical depth and providesatmospheric heating from radiative flux divergence (Mlawer et al. 1997). The Dudhia short-wave scheme has downward integration of solar flux that accounts for clear-air scattering, water vapor absorption, and cloud albedo and absorption (Dudhia 1989). It utilizes look-up tables for clouds but does not use sub-grid cloud fractions, only uniformly clear or cloudy within a model layer.

### 2.1.3 Data Assimilation systems

NWP model requires accurate estimate of the initial state of the atmosphere, which can be achieved through employing Data Assimilation (DA) systems. Data assimilation technique scales the distance between short-range model forecast (hereafter, background) and observation by the errors associated with background and observations, respectively, to produce accurate initial conditions (Kalnay 2003).

#### *2.1.3.1 Basic concept of DA system*

The fundamental concept of DA is based on the least-squares approach that aims at to find the final estimate by minimizing the error made in the analysis. Cressman analysis, optimal interpolation (OI), variational analysis, and the Kalman filter are different DA techniques that are rooted from the least square approach. One of the basic optimality properties of the linear least-square estimate is that it satisfies the properties of the Best Linear Unbiased Estimator (BLUE). According to Gauss-Markov theorem, the sample variance of the least-squares estimate is the minimum among all the linear unbiased estimates (Lewis et al. 2006).

**Proof of Gauss-Markov theorem:**

A least squares problem can be formulated as:

$$f(x) = \frac{1}{2}(z - \mathbf{H}x)^T \mathrm{H}^{-1}(z - \mathbf{H}x) \quad (2.9)$$

where, z is a linear estimation problem and it can be stated as:

$$z = \mathbf{H}x + v$$

Here, *x* is the unknown and $x \in R^n$; z is the known observation and $z \in R^m$; **H** is an known matrix and $\mathbf{H} \in R^{m \times n}$; v is the additive random noise. It is assumed that $(1)\, E(v) = 0$; $(2)\, E(v^T v) = \mathbf{R}$ **,** **R** is symmetric and positive definite; (3) *x* and *v* are uncorrelated.

The main objective of the problem is to minimize *f(x)* with respect to *x*. It can be obtained by computing the gradient of *f(x)* and equating to zero.

Eq. (2.9) gives,

$$\nabla f(x) = (\mathbf{H^T R^{-1} H})x - (\mathbf{H^T R^{-1}})z = 0 \quad (2.10)$$

or

$$\hat{x}_{LS} = (\mathbf{H^T R^{-1} H})^{-1}\mathbf{H^T R^{-1}}\, z \quad (2.11)$$

where, $\hat{x}_{LS}$ is the least square estimate.

Now, $Cov(\hat{x}_{LS}) = E[(\hat{x}_{LS} - x)(\hat{x}_{LS} - x)^T] = (\mathbf{H^T R^{-1} H})^{-1}$ (2.12)

Now, following Gauss-Markov theorem, we need to prove that,

$$Cov(\hat{x}) \geq Cov(\hat{x}_{LS})$$

where, $\hat{x} = \mathrm{G}z$ is an arbitrary linear estimate of $x$ and $G \in R^{m \times n}$.

Now,

$$E(\mathrm{Gz}) = \mathrm{G}\mathbf{H}E(x) + \mathrm{G}E(v)$$

$$= \mathrm{G}\mathbf{H}E(x) \qquad (2.13)$$

It can be seen from Eq. (2.13) that $\hat{x}$ will be unbiased when $\mathrm{GH} = \mathrm{I_n}$.

Again, taking the following difference, we get,

$$\hat{x}_{LS} - \mathrm{Gz} = [(\mathbf{H^T R^{-1} H})^{-1}\mathbf{H^T R^{-1}} - \mathrm{G}]\mathrm{z} = \mathrm{D}z \qquad (2.14)$$

Covariance of Gz can be defined as:

$$Cov(\mathrm{Gz}) = E[(\mathrm{Gv})(\mathrm{Gv})^{\mathrm{T}}] = \mathrm{G}\mathbf{R}\mathrm{G}^{\mathrm{T}}$$

Using Eq. (2.14), we get,

$$Cov(\mathrm{Gz}) = [(\mathbf{H^T R^{-1} H})^{-1}\mathbf{H^T R^{-1}} - \mathrm{D}]\mathbf{R}[(\mathbf{H^T R^{-1} H})^{-1}\mathbf{H^T R^{-1}} - \mathrm{D}]^{\mathrm{T}}$$

$$= \quad (\mathbf{H^T R^{-1} H})^{-1} + \mathrm{D}\mathbf{R}\mathrm{D}^{\mathrm{T}} - [(\mathbf{H^T R^{-1} H})^{-1}\mathbf{H^T R^{-1}}]\mathbf{R}\mathrm{D}^{\mathrm{T}}$$

$$-\mathrm{D}\mathbf{R}[((\mathbf{H^T R^{-1} H})^{-1}\mathbf{H^T R^{-1}} \qquad (2.15)$$

Using identity for unbiasedness, we get,

$$\mathrm{I} = \mathrm{G}\mathbf{H} = [(\mathbf{H^T R^{-1} H})^{-1}\mathbf{H^T R^{-1}} - \mathrm{D}]\mathbf{H} = \mathrm{I} - \mathrm{D}\mathbf{H}$$

It shows that DH $= 0$. Substituting the value of DH in Eq. (2.15), we get,

$$Cov(\mathrm{Gz}) = (\mathbf{H^T R^{-1} H})^{-1} + \mathrm{D}\mathbf{R}\mathrm{D}^{\mathrm{T}}$$

$$= Cov(\hat{x}_{LS}) + \mathrm{D}\mathbf{R}\mathrm{D}^{\mathrm{T}}$$

Since, R is a symmetric and positive definite matrix and $\mathrm{DRD}^{\mathrm{T}}$ is symmetric, for any y $\in R^n$,

$$y^{\mathrm{T}}\ \mathrm{D}\mathbf{R}\mathrm{D}^{\mathrm{T}}\mathrm{y} = (\mathrm{D}^{\mathrm{T}}\mathrm{y})^{\mathrm{T}}\mathbf{R}(\mathrm{D}^{\mathrm{T}}\mathrm{y})$$

For a non-zero vector y, $\mathrm{D}^{\mathrm{T}}$y can be zero. Therefore $\mathrm{DRD}^{\mathrm{T}}$ is positive and semi-definite; which shows that:

$$Cov(\hat{x}) \geq Cov(\hat{x}_{LS}).$$

It is proved that a least square estimate is a Best Linear Unbiased Estimator.

*2.1.3.2 WRF-VAR*

WRF-VAR is the variational data assimilation module of the WRF system. It is based on an incremental variational data assimilation technique, and offers three different data assimilation platforms: 1) three-dimensional variation data assimilation (3DVAR); 2) four-dimensional variation data assimilation (4DVAR); 3) hybrid Ensemble Transform Kalman Filter (ETKF)-3DVAR DA system.

In WRF-VAR, analysis is performed on an un-staggered Arakawa A-grid. Analysis increments are interpolated to staggered Arakawa C-grid and added to the background, also called first guess, to get the final analysis of the WRF-model grid. The observed data can be ingested into

the model in either ASCII format via the "OBSPROC" utility or directly in the "PREPBUFR" format. Apart from other conventional data, WRF-VAR can assimilate satellite radiances in BUFR format and radar reflectivity and radial velocity in the ASCII format, having multiple outer loops to address the nonlinearity.

The impact of observations may vary significantly among the different DA systems, and one such factor that affects the impact of observations is the specification of background error covariance. In the variational method, the horizontal component of the background error is represented via a recursive filter (for regional) or power spectrum (for global). The vertical component is applied through projections on climatologically generated averaged eigenvectors and its corresponding Eigenvalues. Horizontal and vertical background errors are non-separable. Each eigenvector has its horizontal climatologically determined length scale. Preconditioning of the background part of the cost function is done via the control variable transform *U* defined as $B = UU^T$. It includes the "gen_be" utility to generate the climatological background error covariance estimate via the NMC-method or ensemble perturbations.

The 3DVAR DA system in traditional variational framework and ensemble based HYBRID DA system is employed in this study.

*3DVAR*

3DVAR is a least-squares based three-dimensional variational data assimilation technique and it mainly aims on to iteratively minimize the quadratic cost function (Barker et al. 2004):

$$J(x) = \frac{1}{2}(x - x_b)^T \mathrm{B}^{-1}(x - x_b) + \frac{1}{2}\left(y_o - H(x)\right)^T \mathrm{R}^{-1}\left(y_o - H(x)\right) \qquad (2.16)$$

The cost function can be derived based on Bayesian approach, where it is assumed that given the background field, the true state is a realization of a random process defined by the following prior probability distribution function (pdf):

$$p(x_b|x) = \frac{1}{(2\pi)^{n/2}|\mathrm{B}|^{1/2}} e^{-\frac{1}{2}[(x_b-x)^T \mathrm{B}^{-1}(x_b-x)]} \quad (2.17)$$

The observation likelihood is defined as,

$$p(y_o|x) = \frac{1}{(2\pi)^{n/2}|\mathrm{R}|^{1/2}} e^{-\frac{1}{2}[(y_o-H(x))^T \mathrm{R}^{-1}(y_o-H(x))]} \quad (2.18)$$

Here, both the observations' likelihood and prior pdf are considered to follow Gaussian distribution. Finally, given the new observations $y_o$, the posteriori probability distribution of the true state is defined as,

$$p(x|y_o) \propto p(y_o|x)p_B(x_b|x) \propto e^{-\frac{1}{2}[(y_o-H(x))^T \mathrm{R}^{-1}(y_o-H(x))-\frac{1}{2}[(x_b-x)^T \mathrm{B}^{-1}(x_b-x)]} \quad (2.19)$$

The maximum of the posterior probability is obtained by the minimization of the cost function as shown in Eq. (2.16). The key objective of the minimization problem is to attain a posteriori maximum likelihood of the true state of the atmosphere with prior knowledge of background state ($x_b$) and observations ($y_o$) (Lorenc 1986). In other words, it is to find an optimal analysis ($x_a$) that minimizes the cost function $J(x)$, where the analysis can be attained by solving the following equation:

$$\nabla_x J(x_a) = 0 \quad (2.20)$$

To attain an exact solution, the second term in Eq. (2.16) needs to be expanded by linearizing *H* around the background field and it can be expressed as shown below:

$$y_o - H(x) = y_o - H[x_b + (x - x_b)]$$

$$= \{y_o - H(x_b)\} - \mathrm{H}(\mathrm{x} - x_b)$$

Subsequently, Eq. (2.16) can be rewritten as:

$$2J(x) = (x - x_b)^T \mathrm{B}^{-1}(x - x_b) +$$

$$[\{y_o - H(x_b)\} - \mathrm{H}(x - x_b)]^T \mathrm{R}^{-1}$$

$$[\{y_o - H(x_b)\} - \mathrm{H}(x - x_b)] \qquad (2.21)$$

Further, applying the rules of transpose matrix product, we can expand Eq. (2.21) as follows:

$$2J(x) = (x - x_b)^T \mathrm{B}^{-1}(x - x_b) + (x - x_b)^T \mathrm{H}^T \mathrm{R}^{-1} \mathrm{H}(\mathrm{x} - x_b)$$

$$-\{y_o - H(x_b)\}^T \mathrm{R}^{-1}\, \mathrm{H}(\mathrm{x} - x_b)$$

$$-(x - x_b)^T \mathrm{H}^T \mathrm{R}^{-1} \mathrm{H}\{y_o - H(x_b)\}$$

$$+\{y_o - H(x_b)\}^T \mathrm{R}^{-1}\{y_o - H(x_b)\} \qquad (2.22)$$

Since, $J(x)$ is a quadratic function, the gradient of $J(x)$ can be expressed as

$$\nabla J(x) = \mathrm{B}^{-1}(x - x_b) + \mathrm{H}^T \mathrm{R}^{-1} \mathrm{H}(x - x_b) - \mathrm{H}^T \mathrm{R}^{-1} \mathrm{H}\{y_o - H(x_b)\}$$

Here H is the Jacobian of $H$ and $H$ is the nonlinear observation operator that converts analysis to observation space; B is the Background error covariance (BEC) and R is the observation error matrix.

Eq. (2.20) gives

$$(B^{-1} + \mathrm{H}^T\mathrm{R}^{-1}\mathrm{H})(x_a - x_b) = \mathrm{H}^T\mathrm{R}^{-1}\{y_o - H(x_b)\}$$

or

$$x_a = x_b + (B^{-1} + \mathrm{H}^T\mathrm{R}^{-1}\mathrm{H})^{-1}\mathrm{H}^T\mathrm{R}^{-1}\{y_o - H(x_b)\}$$

or

$$x_a = x_b + \mathrm{W}\{y_o - H(x_b)\} \quad (2.23)$$

$$\mathrm{W} = (\mathrm{B}^{-1} + \mathrm{H}^T\mathrm{R}^{-1}\mathrm{H})^{-1}\mathrm{H}^T R^{-1} \quad (2.23a)$$

Eq. (2.23) is the formal solution of the 3DVAR problem. For practical uses, the solutions are obtained by the iterative minimization method like conjugate gradient method in the analysis control variable space.

Eq. (2.23a) can be written in an equivalent form as given below:

$$\mathrm{W} = (B^{-1} + \mathrm{H}^T R^{-1}\mathrm{H})^{-1}\mathrm{H}^T R^{-1} = (\mathrm{BH}^{\mathrm{T}})(\mathrm{R} + \mathrm{HBH}^{\mathrm{T}})^{-1} \quad (2.23b)$$

This equivalence follows from using the Sherman-Morrison-Woodbury matrix identity. There is significant importance of the equivalence in terms of cost effectiveness as inversion of $(\mathrm{R} + \mathrm{HBH}^{\mathrm{T}})$ is much more cost effective than inversion of $(B^{-1} + \mathrm{H}^T R^{-1}\mathrm{H})$, when the observation space is smaller than the dimension of state space.

In Eq. (2.23b), $(\mathrm{BH}^{\mathrm{T}})(\mathrm{R} + \mathrm{HBH}^{\mathrm{T}})^{-1}$ is same as the expression for gain matrix in OI and Kalman filtering, where OI and Kalman filtering are two sequential DA systems rooted from the best linear unbiased estimator (BLUE). The derivation of the background error covariance matrix B in Kalman filter is different from both OI and 3DVAR. Unlike OI and 3DVAR, where B is

assumed to be constant in time, B is updated from the previous analysis time to the new analysis time in Kalman filter method.

*Hybrid ETKF-3DVAR*

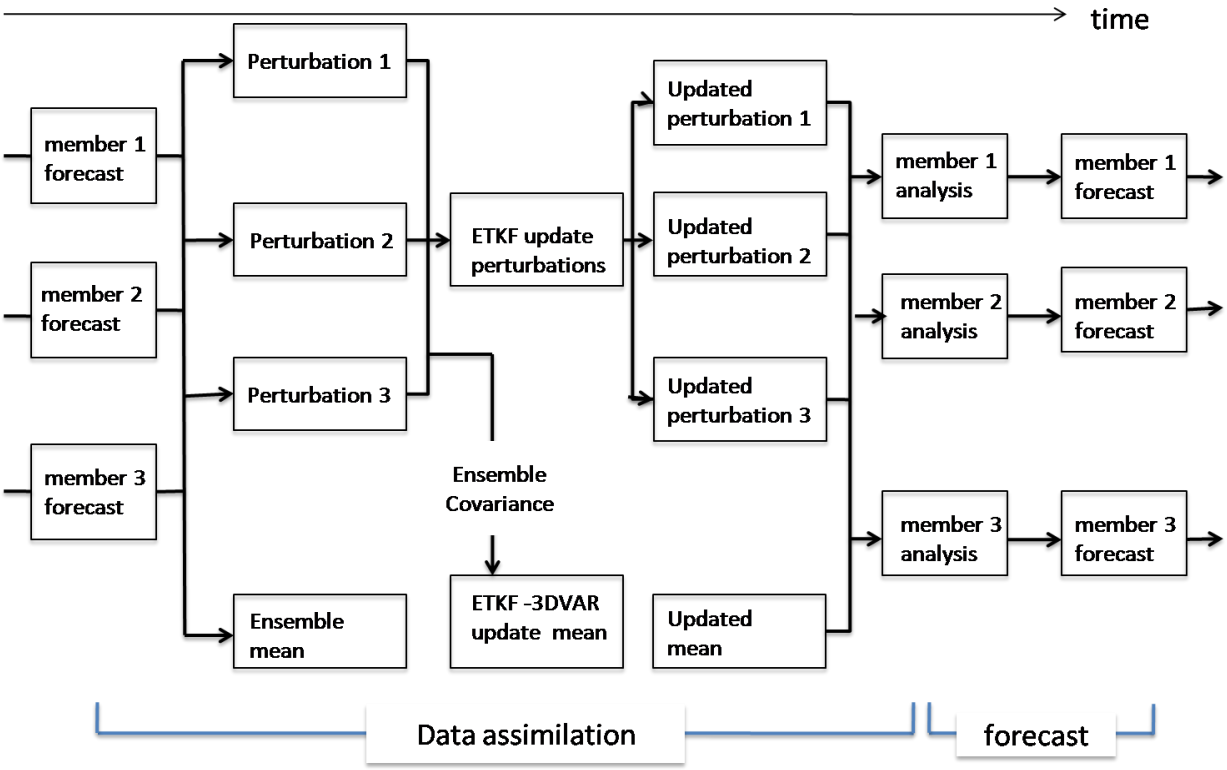


**Figure 2.3:** Illustration of the hybrid ETKF–3DVARanalysis and the ensemble generation cycle for a hypothetical three-member ensemble.

Figure 2.3 (Wang 2011) illustrates how the hybrid ETKF–3DVAR data assimilation cycle works. If we start with an ensemble of K background forecasts at time $t_0$, the following four steps are then repeated for each data assimilation cycle:

1) Update the ensemble mean by the hybrid ensemble-3DVAR method.

2) Update the forecast perturbations using the ETKF.
3) Add the updated ensemble perturbations to the updated ensemble mean to generate K initial ensemble members
4) Make K forecasts starting from the K initial ensemble members forward to the next analysis time.

The steps 1 and 2 will be described in sections a and b:

*a. Incorporating an ensemble in WRF 3DVAR using extended control variables*

We first introduce the terms generally used in the hybrid ensemble–3DVAR framework and then explain how it is applied within WRF 3DVAR.The analysis increment of the HYBRID, denoted as $x'$, is a sum of two terms, defined as

$$x' = x'_1 + \sum_{k=1}^{K}(a_k \circ x_k^e) \tag{2.24}$$

The first term, $x_1^{'}$ in Eq. (2.24) is the increment associated with the WRF 3DVAR static background covariance. The second term is the increment associated with the flow-dependent ensemble covariance. In the second term of Eq. (2.24), $x_k^e$ is the $k^{th}$ ensemble perturbation normalized by $\sqrt{K-1}$ where K is the ensemble size:

$$x_k^e = \frac{x_k - \bar{x}}{\sqrt{K-1}} \tag{2.25}$$

In Eq. (2.25), $x_k$ is the $k^{th}$ ensemble forecast and $\bar{x}$ is the mean of the K-member ensemble forecasts. The vectors $a_k$, k = 1. . . K, denote the extended control variables for each ensemblemember. The symbol $\circ$ denotes the Schur product (element by element product) of the vectors $a_k$ and $x_k^e$. In other words, the second term of Eq. (2.24) represents a local linear

combination of ensemble perturbations. The coefficient $a_k$ for each member varies in space as discussed later, which determines the ensemble covariance localization scale.

The analysis increment $x'$ is obtained by minimizing the following hybrid cost function:

$$J(x_1', a) = \beta_1 J_1 + \beta_2 J_e + J_o$$

$$= \beta_1 \frac{1}{2}(x_1')^T B^{-1}(x_1') + \beta_2 \frac{1}{2}(a)^T A^{-1}(a) + \frac{1}{2}(y^{o\prime} - \boldsymbol{H}x')^T R^{-1}(y^{o\prime} - \boldsymbol{H}x') \qquad (2.26)$$

As compared with a normal 3DVAR cost function, a weighted sum of $J_1$ and $J_e$ terms in Eq. (2.26) replaces the usual background term in hybrid. In Eq. (2.26), $J_1$ is the traditional WRF 3DVAR background term associated with the static covariance B, $J_e$ is a vector formed by concatenating K vectors $a_k$, k = 1. . . K. In other words $a^T = a_1^T, a_2^T, \ldots \ldots a_k^T$ and the variation of the extended control variables $a_k$ are controlled by the localization matrix **A**. Here, $J_o$ is the observation term. As the traditional 3DVAR, $y^{o\prime}$=$y^o - H(x^b)$ is the innovation vector. Here $y^o$denotes the observation, $x^b$ is thebackground forecast, and H is the nonlinear observation operator. In this study, the background forecast $x^b$ is the ETKF ensemble mean forecast. Here $\boldsymbol{H}$ is the linearized observation operator and *R* is the observation error covariance.

In Eq. (2.26), there are two factors $\beta_1$ and $\beta_2$ that define the weights placed on the static background error covariance and the ensemble covariance. To conserve the total background error variance, $\beta_1$ and $\beta_2$ are constrained by

$$\frac{1}{\beta_1} + \frac{1}{\beta_2} = 1 \qquad (2.27)$$

To further comprehend the HYBRID system defined by equations (2.24) – (2.27), Wang et al. (2007) explicitly proved that the solution from these equations is equivalent to the solution by

minimizing a cost function where the BEC was explicitly defined as a sum of the static covariance and the ensemble covariance with localization applied through the Schur product:

$$J(x') = \frac{1}{2}x'^{T}(\frac{1}{\beta_1}B + \frac{1}{\beta_2}P^{e}\circ S)^{-1}x' + \frac{1}{2}\left(y^{0'} - Hx'\right)^{T}R^{-1}\left(y^{0'} - Hx'\right) \quad (2.28)$$

Where $P^e$ is the ensemble covariance defined as:

$$P^e = \sum_{k=1}^{K} x_k^e\,(x_k^e)^T \quad (2.29)$$

Wang et al. (2007) states that given the covariance, $\langle a_k(a_k)^T\rangle$ = S, k=1,……,K the covariance of the second term in Eq. (2.24) satisfies

$$\langle\sum_{k=1}^{K}(a_k\circ x_k^e)[\sum_{k=1}^{K}(a_k\circ x_k^e)]^T\rangle = P^e\circ S \quad (2.30)$$

Based on equations (2.28) – (2.30), effectively the correlation matrix in the second term of Eq. (2.26) performs covariance localization on the ensemble covariance.

*b. The ETKF ensemble generation scheme*

The forecast ensemble perturbations are updated by ETKF using the following transformation matrix T:

$$\mathbf{T} = \mathbf{C}(\Gamma + \mathbf{I})^{-\frac{1}{2}}\mathbf{C}^{\mathrm{T}} \quad (2.31)$$

where Γ represents the eigenvalues and $\mathbf{C}$ represents the eigenvectors obtained by the singular value decomposition of $(\mathbf{HX^b})^{\mathrm{T}}\mathbf{R^{-1}HX^b}$. The under-sampling problem associated with ETKF due to small ensemble size is dealt with adding inflation factors Π and ρ that increases the analysis error variance (Wang et al. 2008):

$$\mathbf{X}^{a} = \mathbf{X}^{b}\mathbf{C}(\rho\Gamma + \mathbf{I})^{-\frac{1}{2}}\mathbf{C}^{T}\Pi \qquad (2.32)$$

Since covariance localization has not been applied to the ETKF formulation, large inflation factors are used to improve the systematic underestimation in the error variance. The initial perturbations for ensembles are obtained as random draws from the static BEC of 3DVAR.

*Stochastic Kinetic Energy Back-scatter Scheme (SKEBS)*

SKEB accounts for the model error due to inadequately resolving subgrid-scale processes in ensemble forecasts. In Stochastic parameterizations, each ensemble member is perturbed by a stochastic forcing term representing the statistical fluctuations in the subgrid-scale fluxes (stochastic diabatic tendencies) and altogether unrepresented interactions between the resolved and unresolved scale (stochastic kinetic energy backscatter). To calculate a stochastic kinetic energy source, random stream function perturbations $\psi'(x, y, t)$ and temperature perturbations $T'(x, y, t)$ are introduced, with a prescribed kinetic-energy spectrum. The effective stream function perturbations $\psi'(x, y, t)$ are given by

$$\psi'(x, y, t) = rD(x, y, t)\psi'(x, y, t) \qquad (2.33)$$

Where x is the zonal and y the meridional direction in physical space, and t denotes the time. Here, $(x, y, t)$ is the local, instantaneous dissipation rate, $\psi'(x, y, t)$ is a 2D stream function pattern with a prescribed kinetic energy spectrum, and r is the parameter "backscatter ratio."The spatial and temporal characteristics of the perturbation pattern are controlled by expanding the stream function pattern $\psi'(x, y, t)$ in spectral space and evolving each wavenumber as a first-order autoregressive process. If D is constant, then these characteristics will directly transfer to

the effective stream function perturbations $\psi'$. However, if D is a function of space and time, then the spatial and temporal characteristics will be the convolution between $(x, y, t)$ and $\psi'(x, y, t)$.

### 2.1.4 Post-processing Graphics Tools

The various post-processing programs supported under the WRF-ARW framework are NCAR Graphics Command Language (NCL), RIP4 (based on NCAR Graphics), GrADS, and Vis5D. This study uses NCL for data visualization and statistical analysis.

## 2.2 Data Used

### 2.2.1 Data for WRF model initialization

The terrestrial data needed by WPS is obtained from the WRF download page. The terrestrial data contains soil categories, land use category, terrain height, annual mean deep soil temperature, monthly vegetation fraction, monthly albedo, maximum snow albedo, and slope category. Because these data are time-invariant, they only need to be downloaded once. The data sets are available in resolutions of 30", 2', 5', and 10'; here, " denotes arc seconds and ' denotes arc minutes. The user need not download all available resolutions for a data set, although the interpolated fields will generally be more representative if a resolution of data near to that of the simulation domain is used.

The meteorological data for model initialization is obtained from the National Center for Environmental Prediction (NCEP) Global forecast system (GFS) data at $0.5^0 \times 0.5^0$ gridded resolution. GFS is a weather forecast model covering the entire globe at a horizontal resolution

of 28 kilometers and provides several variables such as temperatures, winds, precipitation, soil moisture, and atmospheric ozone concentration, etc. Operational forecasters use the dataset to predict weather out to 16 days in the future. In this study 6 hourly, GFS data is used to generate the model initial and boundary conditions.

### 2.2.2 Data for Assimilation

#### *2.2.2.1 Conventional Data*

The current study utilized conventional in-situ observations such as temperature, specific humidity, surface pressure, and wind from surface synoptic observations, radiosonde, ships, buoys, and wind reports from satellites and aircraft available from the Global Telecommunication System (GTS). The GTS plays a significant role in transmitting global meteorological data from both satellites and in-situ measurements to the weather research and forecaster community. NCEP collects the data from various organizations worldwide and formats it to BUFR (Binary Universal Form for the Representation of meteorological data) for their processing needs.

#### *2.2.2.2 INSAT-3D AMV*

INSAT-3D is an Indian geostationary meteorological satellite with the imager and the sounder onboard. The imager has one visible channel and five infrared (IR) channels, namely short-wave IR (SWIR), mid-wave IR (MIR), water vapor (WV), and two split thermal IR (TIR-1, TIR-2) channels. The channels have different ground resolutions. Both visible and SWIR channels are at 1 km resolution. The MIR and TIRs channels are at 4 km ground resolution, and the WV channel is comparatively at a coarser resolution of 8 km. INSAT-3D sounder has 18 IR and one

visible channel. Out of the 18 IR channels, six bands are in the SWIR; five are in the MIR; seven channels are in the longwave infrared (LWIR). The ground resolution of all sounder channels is 10 km. Three consecutive INSAT-3D images of 30-minute intervals are used to determine the AMVs, which consists of the following steps 1) Image registration, thresholding, filtering, 2) Features/tracer selection and tracking, 3) Quality control and 4) Height assignment (Deb et al. 2016; Sankhala et al. 2020). For this study, AMVs were retrieved from IR channel data from 100 hPa to 950hPa atmospheric levels. Low-level MIR and visible channels extended from 600 hPa to 950 hPa, and upper-level WV channel data ranges from 100 to 500 hPa are used. A recent study shows that INSAT-3D AMV is useful in understanding the intra seasonal monsoon variability of the Indian Summer Monsoon (Sankhala et al. 2019). AMV data for this study is obtained from https://www.mosdac.gov.in/.

*2.2.2.3 Radar Data*

The current work utilizes Cherrapunji Doppler Weather Radar (DWR) (Lat. 25.26° N and Long. 91.73° E) data in a numerical weather prediction model for the prediction of thunderstorm. It is an S-band radar that provides data through a volume scans. The radar completes one volume scan in 11 minutes, comprising of 360 degree azimuth scan for 10 elevation angles ranging from 0.5 to 21 degrees. The DWR covers 250 km (up to 500 km only for Z) with a spatial resolution of 300 m. The initial quality check is done by the software developed by BAL, India. To preprocess the data further, a python module has been developed. Since the DWR data resolution is very high compared to the model domain (3km), the quality checked radar data in azimuth-range format is transformed to the Cartesian grid with the map projection same as the model using the python module. Using the same python module data has been quality checked to discard

data beyond the limit of 2-30 m/s and reflectivity out of the range 10–55 dBz. Additionally, FORTRAN codes are developed to convert the processed data to WRFDA compatible format for data assimilation.

### 2.3 Validation

The verification of rainfall forecast is done using standard statistical skill scores such as Bias Score (BS) and Equitable Threat Score (ETS). BS and ETS are calculated based on a contingency table (Table 2.3), which shows the frequency of "YES" and "NO" forecasts and occurrences. The combinations of hits (a), misses (b), false alarms (c), and correct negatives (d) are shown in Table 1.The bias score in equation (2.34) is used to evaluate the model's tendency to underpredict (BS < 1) or overpredict an event (BS > 1). ETS defined by equation (2.35) represent the observed and forecast events that are correctly predicted, adjusted for the frequency of hits that would be expected to occur simply by random chance (Schaefer 1990), where equation (2.36) represents the random chance that both the forecast and observation coincide. The ETS ranges from –1/3 to 1, with a value of 1 signifying the ideal association between predicted and observed rain occurrence.

**Table 2.3: Contingency table**

| Observed | Forecast | |
|---|---|---|
| | YES | NO |
| YES | a | b |
| NO | c | d |

$$BS = \frac{a+b}{a+c} \tag{2.34}$$

$$ETS = \frac{a - a_{random}}{a+c+b-a_{random}} \tag{2.35}$$

$$a_{random} = \frac{(a+c)(a+b)}{a+b+c+d} \tag{2.36}$$

Three other statistical measures, namely, root mean square error (RMSE), mean error (ME), and improvement parameter (η) are used for spatial verification. RMSE is one of the standard statistical methods to evaluate the accuracy of a model. It is the squared root of the average of the square of forecast ($X_i$) departure from the observation ($O_i$) as shown in equation (2.37)

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(O_i - X_i)^2} \tag{2.37}$$

where $i = 1, \ldots \ldots, N$ is the number of forecasts and N, which will vary based on the total number of forecasts from each experiments. Smaller the values of RMSE, better is the model performance. The mean bias (ME), also known as Bias is the average forecast departure from observations as shown in equation (2.38):

$$Bias = \frac{1}{N}(O_i - X_i) \tag{2.38}$$

Improvement parameter (η) in percentage is defined as below:

$$\text{Improvement parameter, } \eta = [1-\frac{RMSE_{Hybrid}}{RMSE_{3DVAR}}] \times 100 \tag{2.39}$$

Positive values of η represent percentage improvement due to HYBRID experiments with respect to 3DVAR. Two-tailed Student's t test has been applied to identify the most significant zones at 90% and 95% confidence level

Furthermore, the impact of flow dependent error covariances in the variational framework is evaluated for the simulation of tropical cyclones (TCs) over the Bay of Bengal. The tropical cyclones considered in this study originated within the geographical range of 0ºN to 25ºN and 75ºE to 100ºE and were persistent within this region until landfall. Therefore, the position of the TCs are searched within this area by obtaining the minimum sea level pressure (MSLP) at the analysis time as well as at every 6 hourly forecast interval till the time of landfall. The location of the MSLP is treated as the simulated position of the TCs and position error is calculated for each TC by taking the great circle distance between the observed best track locations and the simulated position. The intensity error is calculated by taking the difference between the observed and simulated MSLP (hPa). The maximum wind speed (MWSP) is also calculated within the same region for both analysis and 6 hourly forecast intervals and the difference of observed MWSP (m/s) from the simulated MWSP (m/s) is considered as another measure of intensity error.

# *CHAPTER 3

# Impact of flow-dependent error covariance in 3DVAR DA system: Evaluation of short range forecasts during Indian summer monsoon

## 3.1 Introduction

The rainfall variability during the Indian summer monsoon (ISM) plays a crucial role in the agricultural productivity of India and, eventually, in the country's economic development. A large part of the Indian landmass receives 70-80% of its annual rainfall during the summer monsoon season (June to September), with maximum observed rainfall during July and August every year. Prior studies confirmed the effectiveness of 3DVAR DA system in improving forecasts over the Indian subcontinent during the ISM season (Routray et al. 2005; Rakesh et al. 2009; Kumar et al. 2011). However, a fewer studies are available that quantifies the impact of flow dependent error covariance in the DA system for the simulation of ISM using a limited area model. In the present study, the extent to which the initial conditions and model forecasts can be improved by incorporating flow-dependent dynamics in a new flow regime is explored using WRF model. Additionally, the improvements due to ensemble BEC in convection permitting resolution is validated using high-density observational data sets from telemetric rain gauge (TRG) observations over Karnataka state.

A study by Kutty and Wang (2015) examined the impact of flow-dependent ensemble covariance using HYBRID DA system in the National Center for Environmental Prediction (NCEP) Global Forecast System (GFS) model. The results indicate that the quality of forecasts is better in HYBRID initialized forecasts than that from the 3DVAR DA system. Similarly, Prasad et al. (2016) investigated the performance of HYBRID DA scheme in National Center for Medium



*_The content of Chapter 3 is taken from "**Gogoi, R. B.**, Kutty, G., Rakesh, V., & Borgohain, A. (2020). Comparison of the Performance of Hybrid ETKF-3DVAR and 3DVAR Data Assimilation Systems on Short-Range Forecasts during Indian Summer Monsoon Season in a Limited-Area Model. **Pure and Applied Geophysics, 177(10), 5007-5026."**_

Range Weather Forecasting (NCMRWF) Global Forecast System over the Indian region and found that HYBRID DA considerably improved the forecasted temperature and wind field, mainly over the upper troposphere. The current study compares the skill of model forecasts initialized from 3DVAR and HYBRID DA system using the WRF model for month-long, short-range rainfall forecasts during ISM at a convection-permitting resolution.

### 3.2 Model Description and Configurations

The WRF model is configured for three nested domains using the two-way nested method (Figure 3.1). The outer domain (D01) that covers the entire monsoon prevailing region has a horizontal resolution of 27 km with 350 × 350 grid points in the east-west and north-south directions. The inner domain (D02) and the innermost domain (D03) have horizontal resolutions of 9 km and 3 km. The innermost domain is centered over the state of Karnataka in India, where a dense set of rain gauge observations were available for validation. The number of vertical levels used is 36, with the model top at 50 hPa.

The parameterization schemes used are WRF single-moment five-class (WSM5) for microphysics, Kain-Fritsch for cumulus, Yonsei scheme for PBL, Unified Noah land surface model, Rapid Radiative Transfer Model (RRTM), and Dudhia for longwave and shortwave radiation schemes, respectively. The initial and lateral boundary conditions for the model initialization are generated using the National Center for Environmental Prediction (NCEP) Global forecast system (GFS) data at $0.5^0 \times 0.5^0$ gridded resolution. It is important to mention here that the assimilations are performed only in the outermost domain at 27 km resolution, and the interpolated fields from the outer domains are used to initialize the nested domains.

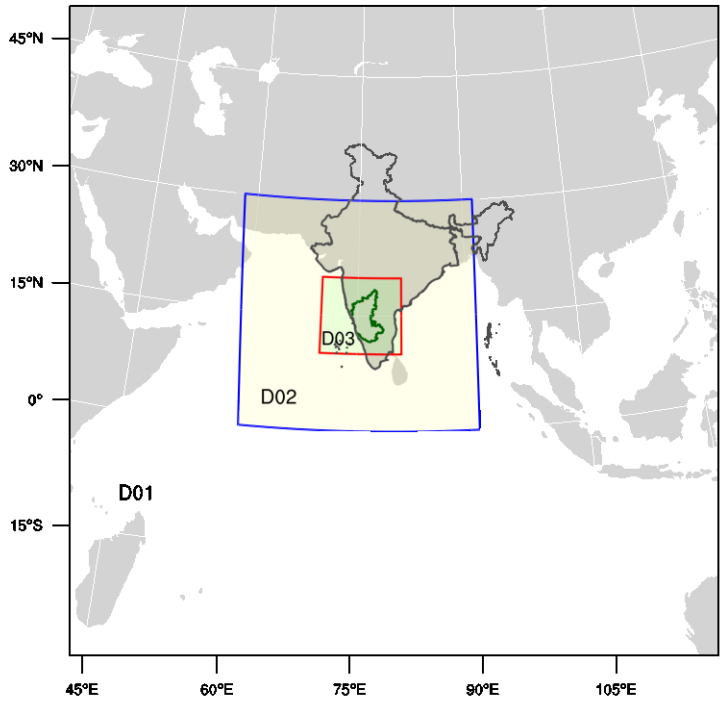


**Figure 3.1:** Model configuration deployed in this study; the innermost domain (D03) over the Karnataka State of India which is highlighted in the inset figure.

### 3.3 Experimental design and validation

In order to bring out the impact of ensemble-generated error covariance in the 3DVAR framework, four experiments are conducted. The experiment, which assimilated observations using the 3DVAR DA system, is denoted as 3DVAR. The experiments with HYBRIDDA system are configured in three variants by changing the tunable parameters $\beta_1$and $\beta_2$ that controls the contributions of the static and ensemble covariance in the HYBRID minimization cost function. The experiments are named HYBRID25, HYBRID50, and HYBRID75, where HYBRID25 signifies 25% (75%) weights assigned to ensemble (static) BECs, HYBRID50 represents an equal contribution of $\beta_1$ and $\beta_2$ and HYBRID75 represents 75% (25%) weights

assigned to ensemble (static) BEC. For data assimilation, the conventional *in situ* observations and satellite-derived wind observations available from the Global Telecommunication System (GTS) are used at every 12 h interval in ±3 h time window for ~4 weeks period during July 2013 with continuous cycling starting from 0000UTC 1 July 2013 to 0000 UTC 30 July 2013. Afterward, 48 h free forecast is launched from each 0000 UTC DA analysis during July 2013.To avoid the spin-up issues, the ensembles are initialized at 0000UTC 30 June 2013, 24 h prior to the first analysis time, by adding 50 random perturbations from WRF 3DVAR (Barker et al. 2004) and subsequently, advanced in time using ARW-WRF for a period of 24 h. The first guess for 3DVAR and HYBRID DA system is taken from the 24 h ensemble mean forecast at 0000UTC 1 July 2013, ensuring that HYBRID and 3DVAR experiments started from the same background forecast.

A systematic evaluation of the impact of flow-dependent error covariance in variational DA systems is carried out quantitatively in this study. First, the validation of rainfall forecast over the Indian landmass for D1 is performed by comparing model simulated rainfall with the India Meteorological Department (IMD; Rajeevan et al. 2008) gridded rainfall data available at 0.25˚×0.25˚ horizontal resolution. The gridded data is prepared using IMD archived daily rainfall data measured by rainfall gauge stations spread across India(Pai et al. 2015). Both observed and forecast data are brought to a common grid of 0.25˚×0.25˚ resolution prior to comparison. As a second set of data, a dense set of telemetric rain gauge (TRG) observations over Karnataka state are employed for validating the high-resolution rainfall forecasts for domain 3.A detailed description of the observation datasets used and their errors are available in Rakesh et al. (2015). The forecast value closest to the observation location is considered to verify the gridded forecast with the point observation location.

Furthermore, for spatial verification of the large-scale features of monsoon circulations simulated by the experiments, the European Center for Medium Range Weather Forecasting re-analysis (ERA) interim analyses at 12.5 km resolution is used. Before the statistical verification, the model forecasts are brought to the ERA interim grid resolution. The conversions of grid resolution in this study are performed using bilinear interpolation for all the cases.

The verification of rainfall forecast is done using standard statistical skill scores such as Bias Score (BS) and Equitable Threat Score (ETS). Three other statistical measures, such as root mean square error (RMSE), mean error (ME), and improvement parameter ($\eta$) are used for spatial verification. Additionally, two-tailed Student's t test has been applied to identify the most significant zones at 90% and 95% confidence levels.

### 3.4 Results and Discussion

This section evaluates the performance of traditional 3DVAR and HYBRID DA system in the short-range forecast during the Indian summer monsoon (ISM).

#### 3.4.1 Analysis and forecast verification

Figure 3.2 depicts vertical profiles of the analysis fit to observations for variously configured HYBRID and 3DVAR experiments where the analysis fields are averaged over the 60 data assimilation cycles. It is to be mentioned here that the Radiosonde observations used to calculate the analysis fit are assimilated to the analysis. Thus, the results cannot be treated as a measure of analysis error; rather, it represents the extent to which analysis from each DA scheme is closer to the observations.

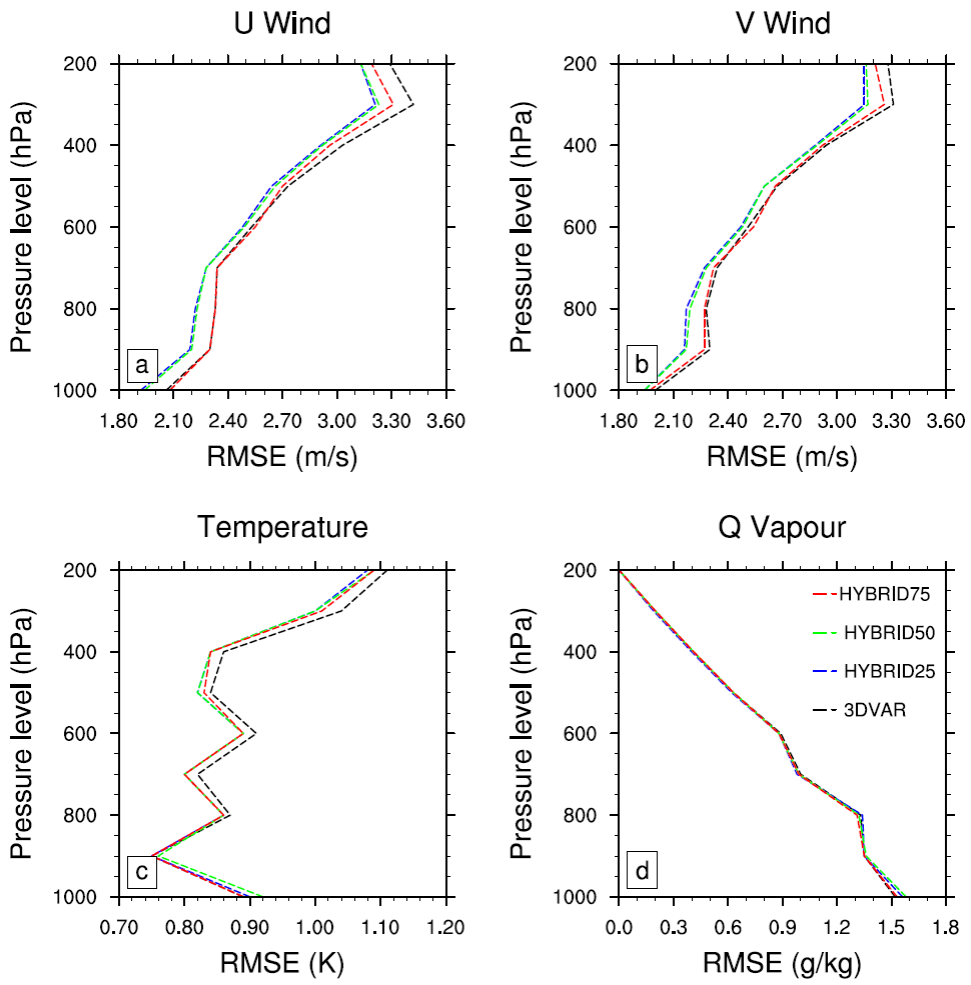


**Figure 3.2:** Vertical profiles of root mean square fit of analyses to Radiosonde observations for different prognostic variables; analyzed by 3DVAR (black line), HYBRID 25 (blue line), HYBRID 50 (green line) and HYBRID75 (red line) experiments.

For wind and temperature, the HYBRID analysis fits better to the observations than 3DVAR analysis. Among the HYBRID experiments, analysis from HYBRID25 and HYBRID50 are closer to observations than HYBRID75 analysis. For specific humidity, analysis from all the experiments describes a comparable fit to the Radiosonde observations, in general.

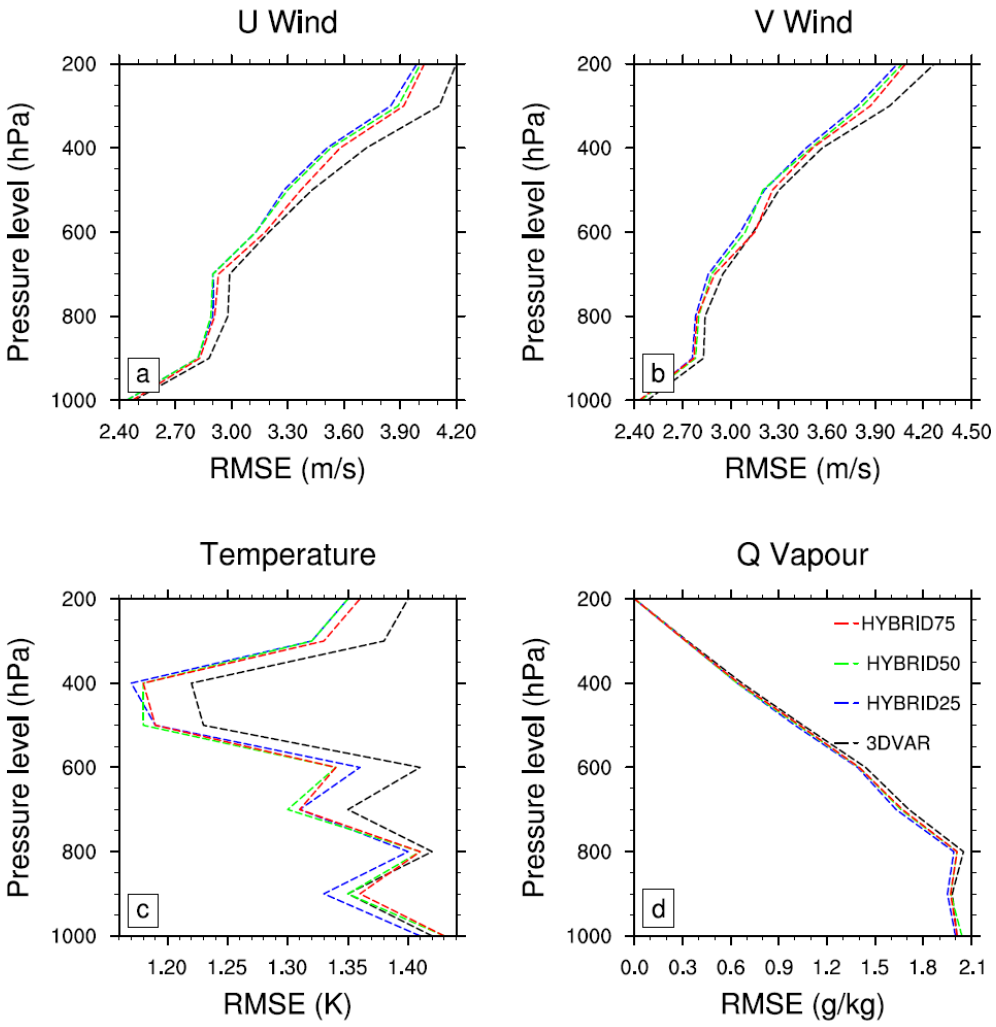


**Figure 3.3:** Vertical profiles of root mean square errors in 24 h forecasts from different experiments compared to radiosonde observations for different variables; 3DVAR (black line), HYBRID 25 (blue line), HYBRID 50 (green line) and HYBRID 75 (red line).

Figure 3.3 shows RMSE of 24 h wind, temperature, and specific humidity forecast validated against the Radiosonde observations. For the zonal and meridional wind, RMSE in HYBRID experiments is smaller than 3DVAR in almost all levels. The improvements in HYBRID experiments are more significant over the upper levels. The HYBRID50 and HYBRID75 experiments show improvements above 850 hPa, for temperature forecast as compared to

3DVAR experiment. Wang et al. (2008) have documented analogous results while comparing the performance of 3DVAR against HYBRID experiments. For specific humidity, HYBRID experiments show minimal improvements over 3DVAR (Figure 3.3d).

### 3.4.2 Validation of Rainfall over the Indian land mass

Figure 3.4a represents the observed monthly mean precipitation from IMD for July 2013, which depicts maximum rainfall along the west coast of the peninsular regions and over the central and northeastern parts of the Indian subcontinent. The model experiments simulated the rainfall spreading reasonably well, yet there is substantial underestimation in the magnitude of rainfall maxima (Figure not shown). Although all the experiments show wet bias and dry bias all over the Indian landmass, a notable strong dry bias is seen over the south-central region of the Indian subcontinent. However, this dry bias is significantly reduced in HYBRID experiments, mainly in HYBRID75 compared to the 3DVAR experiment. The improvement in the pattern of rainfall distribution in HYBRID compared to 3DVAR experiments may be attributed to the better depiction of dynamic and thermodynamic features associated with flow-dependent adjustments in the HYBRID analysis. The spatial distribution of the improvement parameter ($\eta$) computed using equation (2.28) for July 2013 is shown in Figure 3.4f-h. The positive (negative) value of $\eta$ depicts improvement (degradation) in the forecast.

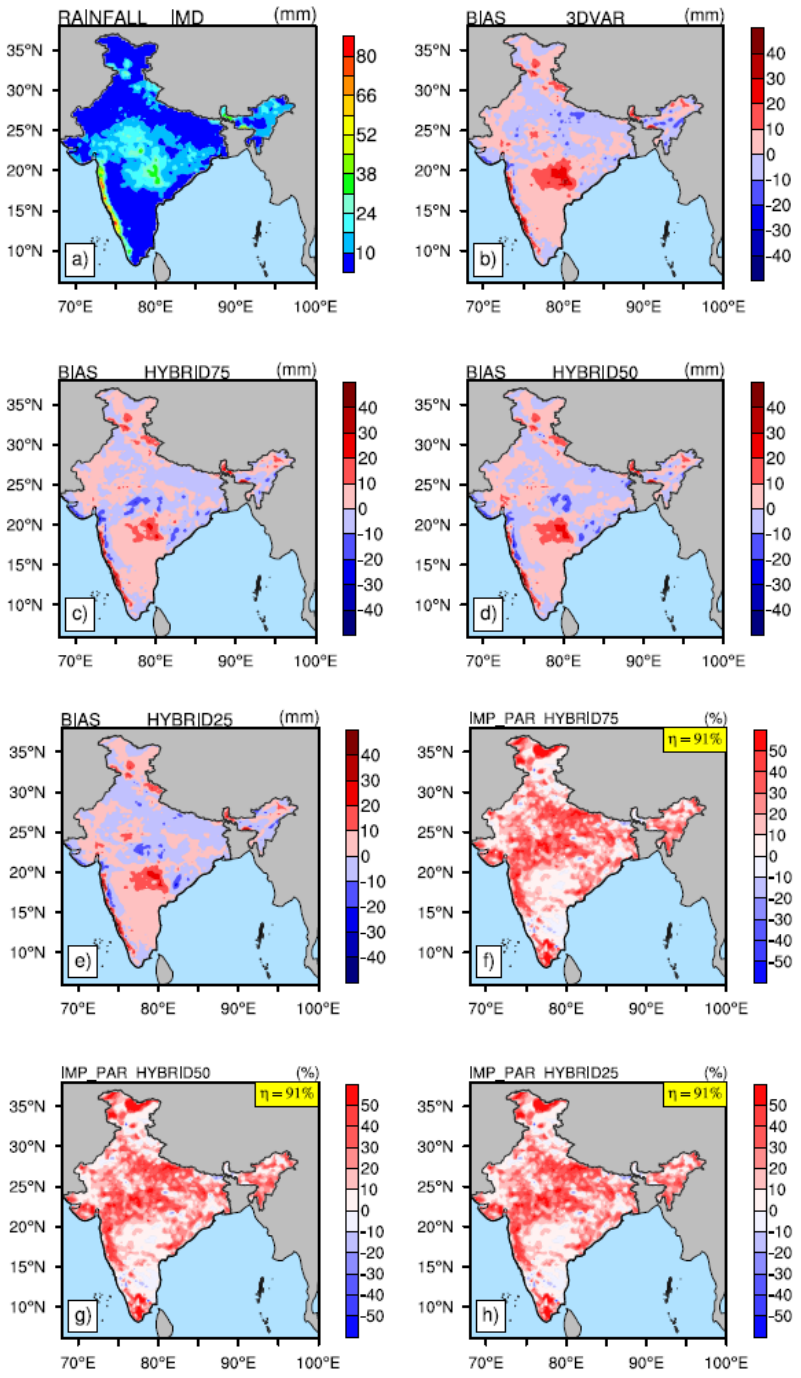


**Figure 3.4**: (a) Observed spatial distribution of monthly averaged (July) 24 h accumulated rainfall (mm/day) and BIAS in forecasted rainfall from(b) 3DVAR (c) HYBRID75 (d) HYBRID50 and (e) HYBRID25. The improvement parameter (%) where positive values represents improvements by HYBRID experiments are also shown (f,g and h).

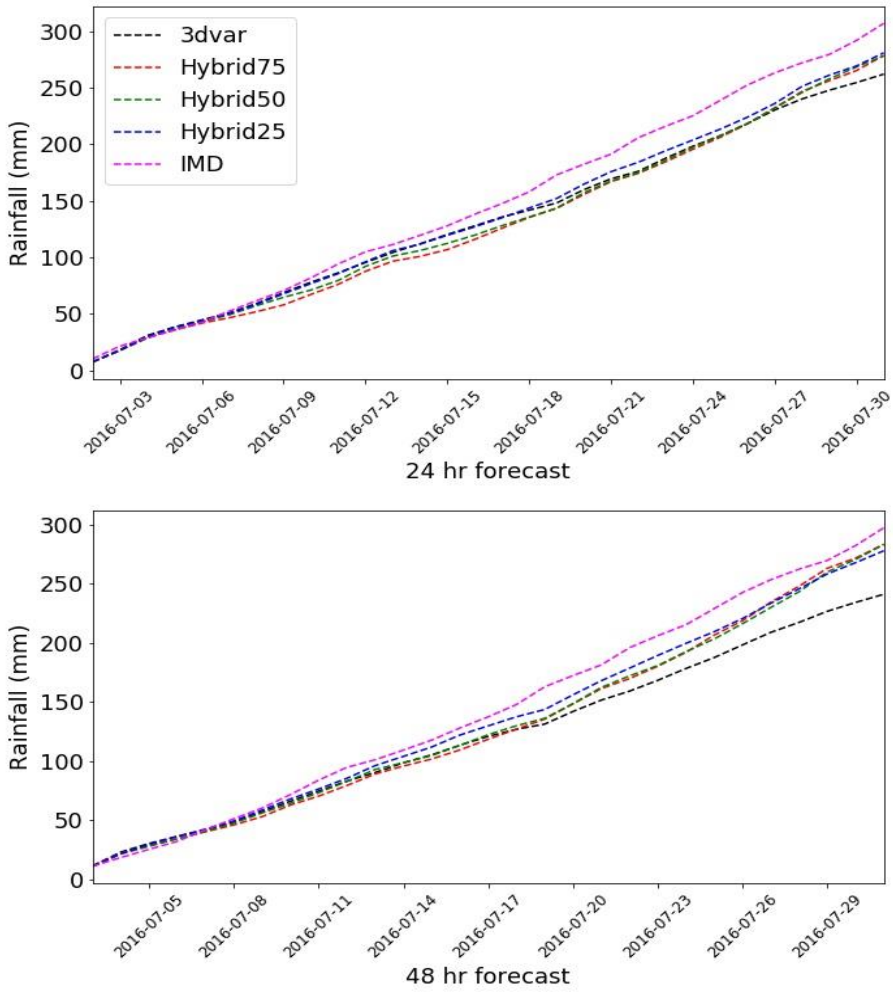


**Figure 3.5:** The cumulative rainfall over the Indian land mass from domain 1 aggregated over all the 24 hour forecasts (top) and 48 hour forecast (bottom) throughout the July, 2013

The rainfall simulation from HYBRID experiments is closer to IMD rainfall observations. Substantial improvement in HYBRID run is observed during later stages of the assimilation cycle for 24 h and 48 h forecasts compared to the 3DVAR experiment.

While comparing average forecasts in previous sections gave a generalized idea about the qualitative forecast skill of the model experiments, quantitative rainfall forecast skill has been addressed in this section. The rainfall forecast skill of the NWP models is usually validated quantitatively using Equitable Threat Score (ETS) and Bias Score (BS). Here, ETS and BS scores are used to evaluate the quantitative precipitation forecast (QPF) skill scores of 3DVAR and HYBRID experiments.

As shown in Figure 3.5, the cumulative effect of the DA system becomes more apparent as the number of data assimilation cycles increases. Therefore, the skill scores are computed in two phases; the first phase is for the assimilation cycle starting from 1 – 15 July 2013, the second phase is for the assimilation cycle from 16- 31 July 2013. Figure 3.6a-dshows the ETS of 24 h accumulated rainfall forecasts for the first and second phase evaluated against the IMD gridded rainfall. As expected, the second phase shows higher ETS skill scores compared to the first phase and BS close to 1.0 for all the DA experiments. The ETS values are gradually increasing with increasing rainfall thresholds for all the experiments indicating a higher QPF skill for rainfall at higher thresholds. Overall, the skill scores are better in HYBRID than in 3DVAR experiments for both phases. Among HYBRID experiments, ETS values are higher for HYBRID50 in the first phase and HYBRID75in the second phase for all the thresholds. The BS for 24 h rainfall forecasts (Figure 3.6b) indicates underestimating rainfall in all the model experiments. Among the experiments, the forecast initialized from HYBRID shows better BS when compared to 3DVAR across all the thresholds.

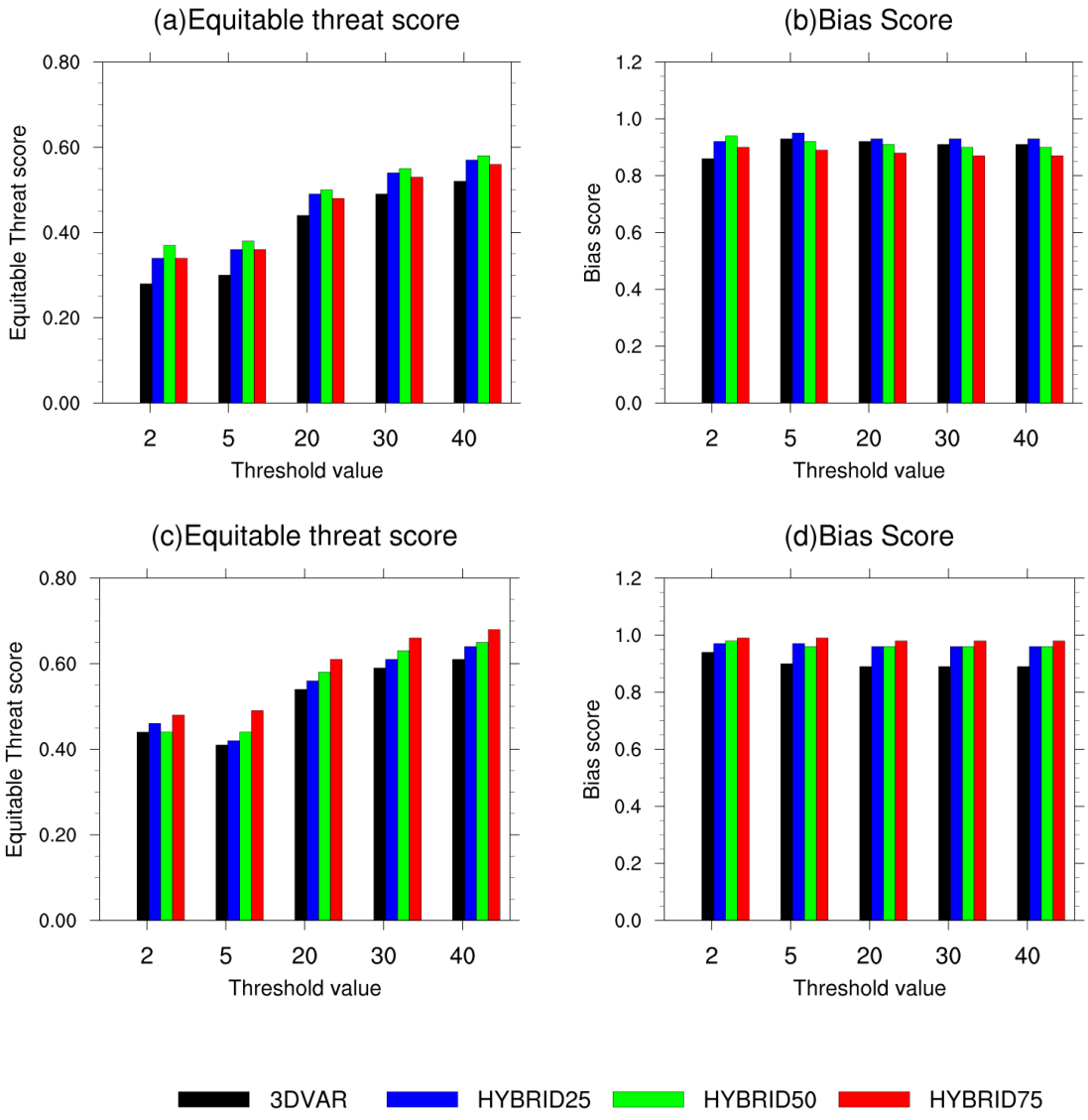


**Figure 3.6:** The (a) ETS and (b) Bias scores for different rainfall thresholds computed over the Indian land mass from domain 1 (d01) simulations averaged over the 24 hour forecasts valid from 2nd July 2013 to 16th July 2013; the (c) ETS and (d) Bias scores for different rainfall thresholds computed over the Indian land mass from domain 1 (d01) simulations averaged over the 24 hour forecasts valid from 17th July 2013 to 31th July 2013

The HYBRID experiments show BS values closer to one during the second phase of data assimilation cycles. Similar to the results obtained for 24 h, HYBRID experiments outperform 3DVAR experiments by enhancing the precipitation forecast skill for 48 h forecast (Figure not shown).

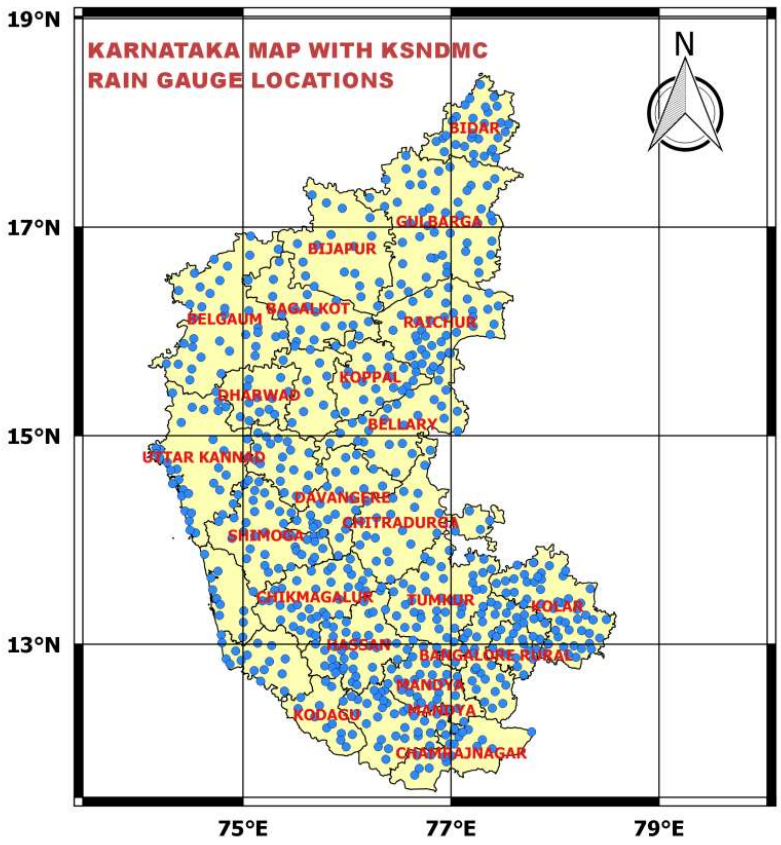


**Figure 3.7:** Spatial distribution of 746 KSNDMC rain gauge stations spread over Karnataka.

### 3.4.3 Validation of high resolution Rainfall forecast over Karnataka

The forecast skill of QPF depends strongly on the model resolution, and several past studies show that model simulations at grid spacing of less than 4 km generate a more realistic

simulation of convective systems (Weisman et al. 1997; Speer and Leslie 2002). However, the assessment of the rainfall forecast is mainly challenging due to the unavailability of high dense rainfall observations. This study uses the rainfall observations from a dense automated rain gauge network of 746 stations available over Karnataka (Figure 3.7) to evaluate rainfall forecast at convection-permitting resolution from the innermost domain (d03).

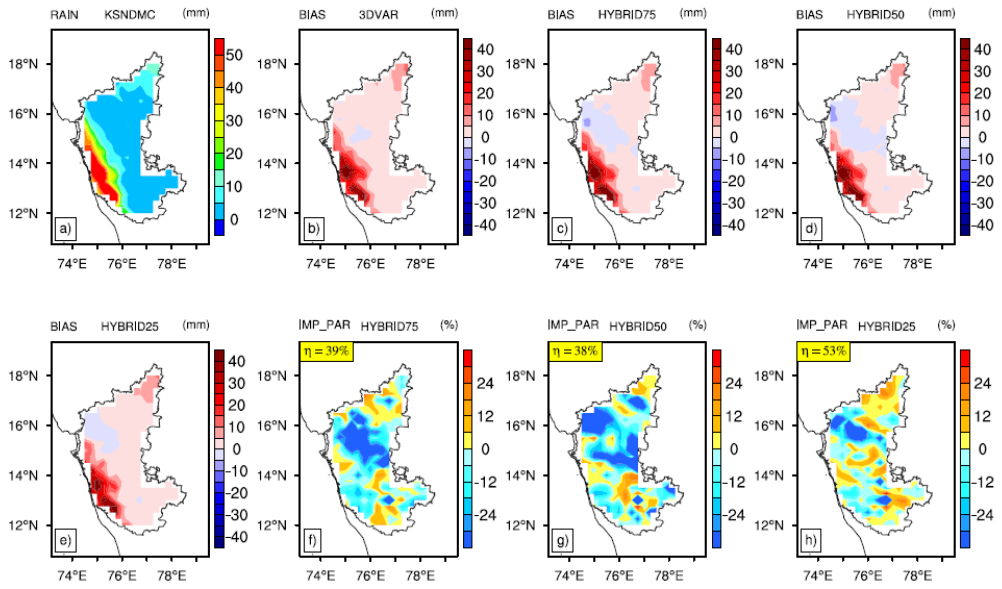


**Figure 3.8:** (a) Observed spatial distribution of monthly averaged (July) 24 h accumulated rainfall (mm/day) and BIAS in forecasted rainfall from (b) 3DVAR (c) HYBRID75 (d) HYBRID50 and (e) HYBRID25 domain 3 (d03) simulations. The improvement parameter (%) are also shown (f, g and h) and highlighted in yellow box.

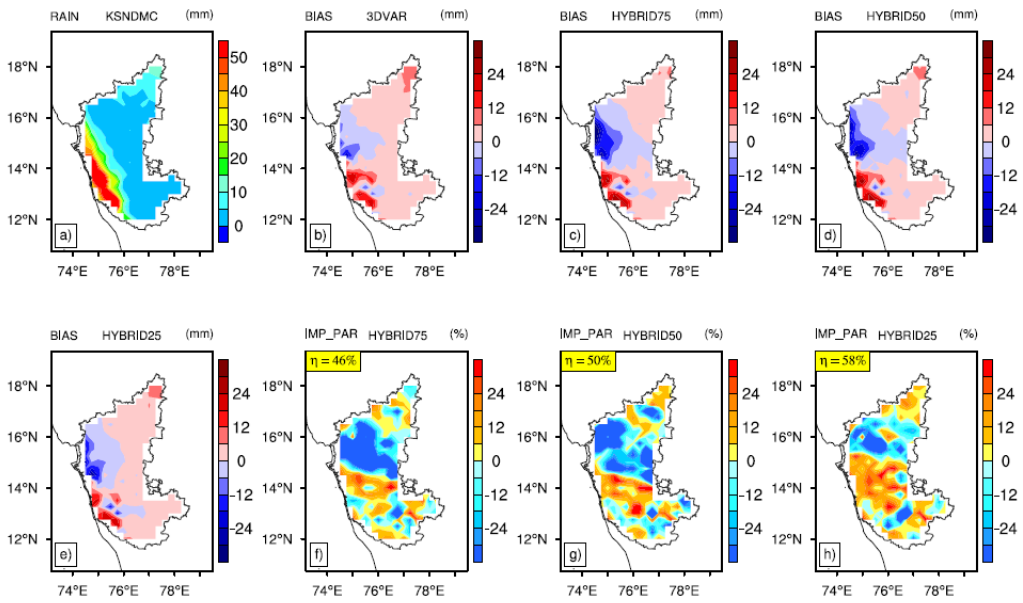


**Figure 3.9:** Same as Figure 3.8a, but from domain 3 (d03) simulations averaged over all the 48 hour forecasts throughout the July, 2013.

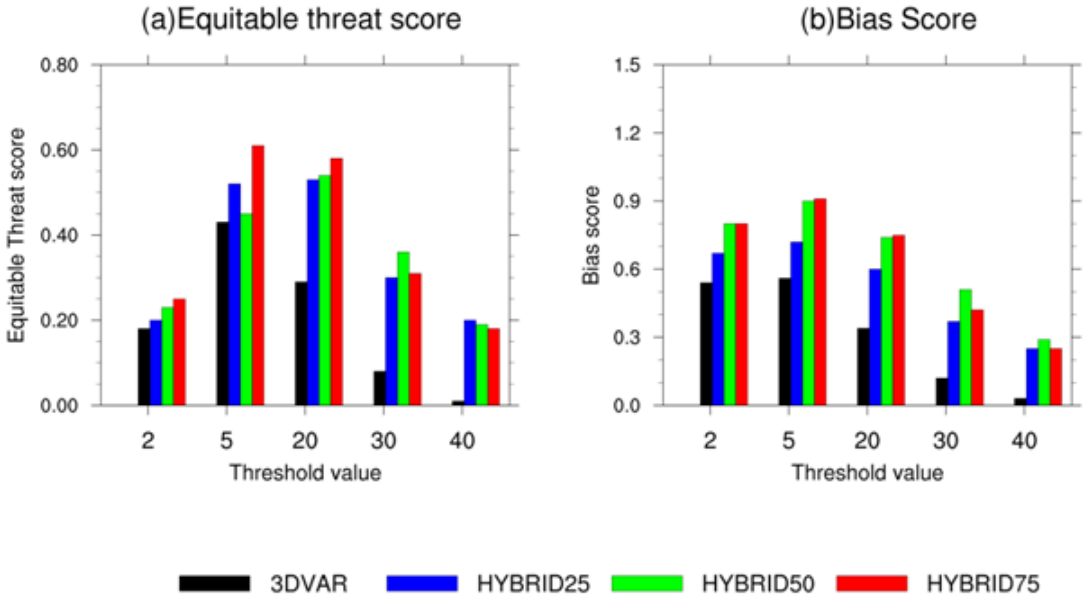


**Figure 3.10:** The (a) ETS and (b) Bias scores for different rainfall thresholds computed over the Karnataka state of India from domain 3 (d03) simulations averaged over the 24 hour forecasts valid from 17th July 2013 to 31$^{th}$ July 2013

Figure 3.8 shows the observed monthly-averaged 24 h accumulated precipitation from TRG of Karnataka state. In the 24 h model forecast (Figure 3.8b-e), a strong dry bias in all the experiments over the southwestern part of Karnataka can be seen. The 3DVAR and HYBRID experiments represent almost similar results, except for the slight wet bias that extends from west to east in the HYBRID run. In the 48 h forecast, a prominent wet bias and the dry bias over the western regions of Karnataka state can be seen (Figure 3.9b-e). The improvement parameter for 24 h forecast (Figure 3.8f-h) shows that the percentage of improvement is 53%,38%,and 39%,while for 48 h forecasts (Figure 3.9f – h), the percentage of improvement is 58%, 50% and 46% for HYBRID25, HYBRID50, and HYBRID75 experiments, respectively. Similar to that in the previous section, the model skill scores are computed for convection-permitting rainfall forecast using TRG observations over Karnataka in two phases. Figure 3.10 illustrates ETS and BS for the second phase, revealing that HYBRID experiments depicts better skill compared to 3DVAR experiment for most rainfall thresholds towards the later stage of forecast.

### 3.4.4 Validation of Monsoon circulation systems

It is evident from the previous section that the use of flow-dependent ensemble BEC in the variational framework improves the rainfall forecast. However, the improvement in rainfall distribution and intensity is strongly related to the precise representation of the observed large-scale flow patterns. Therefore, it is expected that the HYBRID DA system with flow-evolving estimates of BECs to provide optimal analysis with a realistic representation of flow patterns that may, in turn, lead to the improved QPF. Here, the large-scale features in 3DVAR and HYBRID experiments are compared against the ERA interim analysis to understand this

further. Since, no significant difference among the HYBRID experiments is observed in the analysis of model variables and rainfall forecast, the HYBRID75 run is considered representative of HYBRID experiments in this section for convenience

The ISM circulation is characterized by the onset of Low-Level Jet (LLJ), which is commonly known as Findlater Jet over the Arabian Sea (Findlater 1978). It is to mention here that, in this study, LLJ characteristics seen in 850 hPa level is only discussed. Figure 3.11 shows the general features of monthly averaged LLJ from the ERA-interim analysis and 24 h forecast initialized from 3DVAR and HYBRID analysis. Though the flow patterns in the experiments are comparable to the ERA-Interim analysis, in general, the magnitude of wind in both the experiments is underestimated, especially over the peninsular regions of the Indian subcontinent. Moreover, the westerly flow becomes more north-westerly in the 3DVAR forecast, near the southern tip of the Indian subcontinent than the ERA-Interim analysis and HYBRID experiment (Figures 3.11a-c). Also, the magnitude of westerly wind over the peninsular Indian landmass in 3DVAR is considerably lower than HYBRID and, it could be the potential reason for improved dry bias observed over the Indian landmass in 3DVAR experiments compared to HYBRID experiments. It is worth noting that RMSE in simulated wind fields (Figures 3.11d-e) are higher in 3DVAR as compared to HYBRID, in general. Furthermore, Figure 3.11f shows statistically significant differences between the 3DVAR and HYBRID experiment over the eastern and western equatorial Indian Ocean and the upper peninsular regions, where considerable rainfall bias is observed.

Another important circulation system during the Indian summer monsoon is the upper-level circulation at 200 hPa, characterized by strong easterly winds and an anticyclone centered over

the Tibetan Plateau as shown in Figures 3.12a-c. It is observed that the 3DVAR experiment show higher RMSE values over the equatorial Indian Ocean due to the anomalous northward turn of easterly winds. Further, the vector difference of the simulated wind in HYBRID and 3DVAR depicts enhanced southerly wind near the south of the Indian subcontinent in the HYBRID experiment (Figure 3.12f). Overall, the results imply that wind patterns in the lower and upper troposphere show marked improvement in HYBRID simulations. This may be attributed to the decrease in dry bias in the rainfall simulations over the Indian landmass in HYBRID experiments.

Wind shear is another parameter known to be instrumental in northward propagation of the monsoon(Jiang et al. 2004) and it is computed as the vector difference of wind at 850 hPa and 200 hPa. From Figure 13a, it can be seen that ERA analysis depicts strong wind shear over the core monsoon regions like the equatorial Indian Ocean and peninsular India. The RMSE in simulated wind shear is lower in HYBRID than in 3DVAR, as observed from Figures 3.13b and 13c.The HYBRID and 3DVAR runs show significant differences in the wind shear at a confidence interval of 95% (Figure 3.13d). The HYBRID experiment also demonstrates positive shear in the wind over the Indian landmass, which indicates faster progression of monsoon currents in HYBRID than in 3DVAR experiments.

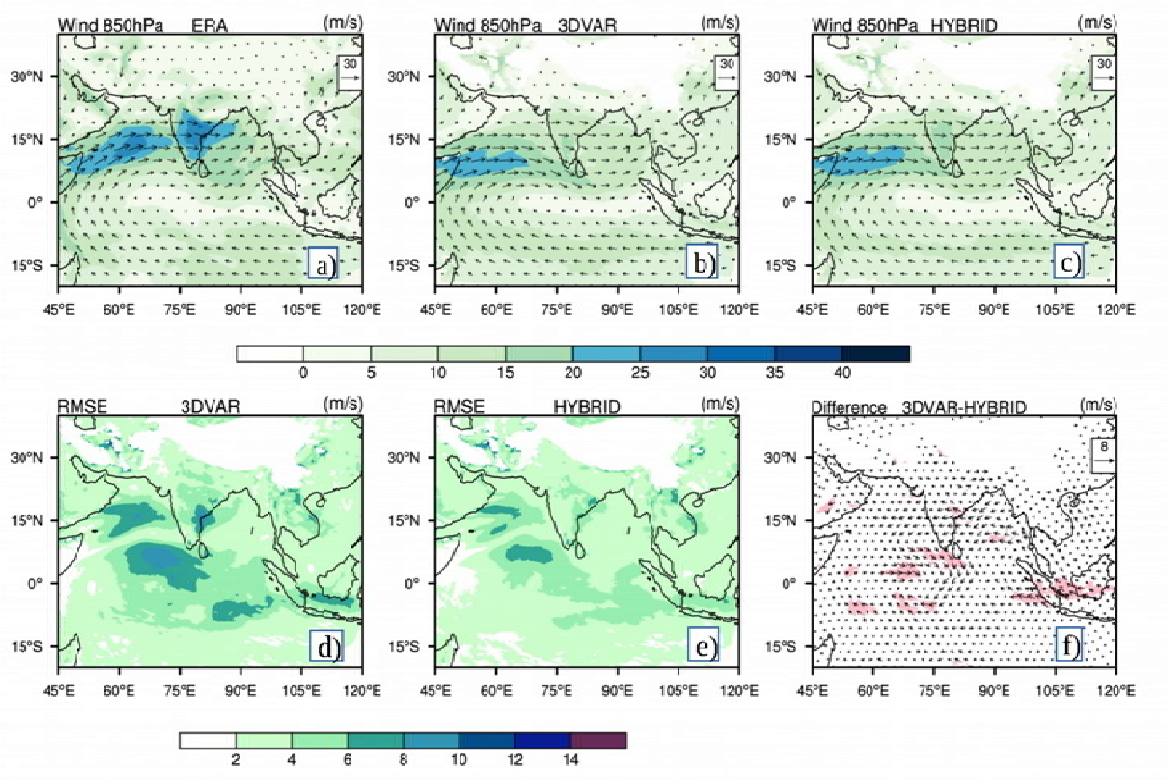


**Figure 3.11:** Monthly averaged 850 hPa wind from (a) ERA interim and 24 hour forecast from (b) 3DVAR and (c) HYBRID. RMSE in 24 h forecast from assimilation experiments with respect to ERA interim for (d) 3DVAR (e) HYBRID. The difference between 3DVAR and HYBRID with statistical significance at 90% is also shown (f). Spatial correlation of g) 3DVAR and h) HYBRID with ERA interim.

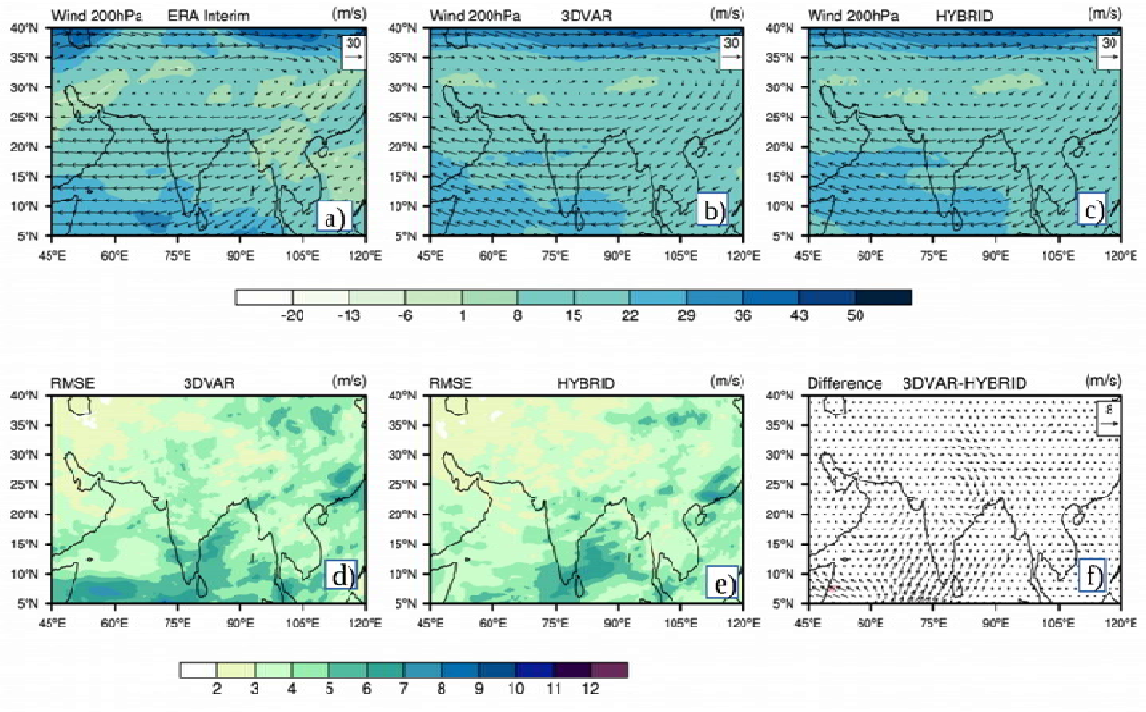


**Figure 3.12:** Same as Figure 3.11, but mean wind at 200 hPa level.

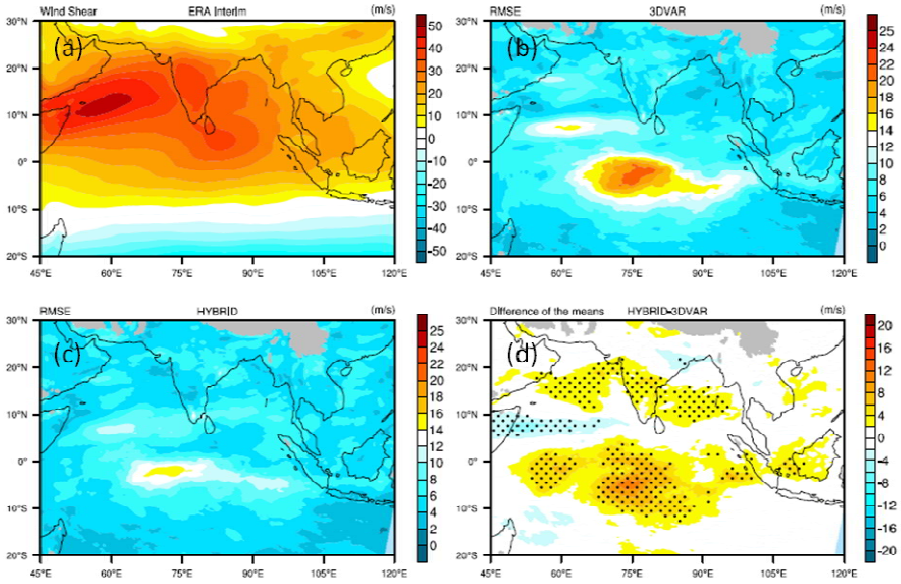


**Figure 3.13:** Monthly averaged vertical wind shear from (a) ERA interim and RMSE in wind shear computed for (b) 3DVAR and (c) HYBRID. The significant differences in the simulated mean wind field from HYBRID and 3DVAR at 95 % confidence level is also shown (d).

### 3.5 Summary

Since the Indian economy is profoundly dependent on the summer monsoon rainfall, its accurate prediction is of socio-economic importance. However, the monsoon rainfall prediction is challenging as the associated features span over a wide range of spatial and temporal scales extending from weather to climate scales. This study attempts to assess the impact of ensemble-based flow-dependent error covariance in the 3DVAR DA framework in improving the short-range forecast of Indian summer monsoon rainfall. The experiments have been conducted over the Indian summer monsoon region for July 2013. The continuous DA cycling is performed for ~4 weeks using 3DVAR and a HYBRID DA system that initialized 48 h model forecasts daily.

The results show that the use of flow-dependent ensemble BEC in 3DVAR systematically improves the forecast. The RMSE from the HYBRID experiments is systematically smaller than the 3DVAR experiments for the temperature and wind field in almost all vertical levels. In addition, HYBRID experiments have shown improved skill in the QPF during the Indian summer monsoon. The spatial comparison shows that the dry bias over the upper peninsular regions and wet bias over the central and northern parts of the Indian subcontinent is lower in HYBRID experiments compared to the 3DVAR experiment. The convective-scale experiments are conducted at higher resolution, and the results are validated against dense TRG network observations over the Karnataka state. Significant improvement in HYBRID runs during the later stages of data assimilation cycles is observed than the 3DVAR experiment. However, no significant difference among the HYBRID experiments has been observed for both analysis of the model prognostic variables and rainfall forecast.

The general circulation features of monsoons in 3DVAR and HYBRID experiments are comparable to those in the ERA-interim analysis. However, the magnitude and direction of lower and upper-level wind over the Indian monsoon region are better forecasted in HYBRID compared to the 3DVAR experiment. This is proposed as the reason for reducing dry bias observed over the Indian landmass in the HYBRID run.

This study is a preliminary effort to comprehend the effect of flow-dependent ensemble BEC in the 3DVAR DA system during the Indian summer monsoon season using a limited area model. Overall, the results are encouraging, and the study provides dynamically consistent, objective improvements in the initial conditions of the WRF model. To gain more insight into the relative advantages and disadvantages, the results of HYBRID DA systems should be compared with those of advanced DA systems such as 4DVAR. Moreover, the data assimilation was performed in the coarser-resolution domain for this study. The improvement in HYBRID DA may be more significant if assimilation is performed in finer resolution.

# *Chapter 4

## Impact of flow-dependent error covariance in 3DVAR DA System: Evaluation of short range forecast of tropical cyclones over Bay of Bengal

### 4.1 Introduction

Tropical cyclone (TC) is among the most devastating natural calamity associated with damaging winds and rainfall and often cause flooding in coastal and inland areas. Even though numerical weather prediction (NWP) models for the prediction of TCs were initiated quite early (e.g., Ooyama 1969), the forecast skill at necessary accuracy continues to be challenging. One of the main cause of such forecast errors is inaccurate initial conditions (Kalnay 2003). The TCs generally form over oceans, where direct in situ measurements are sporadic or absent, making it difficult to generate accurate initial conditions for NWP models.

It is found from previous studies that incorporating flow-dependent ensemble BEC in DA system could correct background TC forecast by spreading information from observations effectively (Torn and Hakim 2009; Hamill et al. 2011b, a). The HYBRID DA system has made immense developments in the TC forecast along with its environment over the Atlantic and Pacific oceans than 3DVAR counterpart (Wang 2011; Li et al. 2012; Shen and Min 2015; Xu et al. 2016; Lu et al. 2017). In another study by Kutty et al. (2018) has examined the model error impact in the HYBRID DA system for the forecast of TCs over the Bay of Bengal (BoB).

Even though the efficiency of static BEC has been reported in several studies (Rakesh and Goswami 2011; Dhanya and Chandrasekar 2016; Routray et al. 2016; Srinivas et al. 2017), studies regarding the effect of flow-dependent ensemble covariance in the variational DA



*The content of Chapter 4 is taken from Kutty, G., **Gogoi, R.,** Rakesh, V., & Pateria, M. (2020). Comparison of the performance of HYBRID ETKF-3DVAR and 3DVAR data assimilation scheme on the forecast of tropical cyclones formed over the Bay of Bengal. **Journal of Earth System Science, 129(1), 1-14.**

framework for forecast of TCs over the North Indian Ocean (NIO) are limited. This study investigates the competency of HYBRID flow-dependent ensemble DA system for a new flow regime in which similar studies are rare. Since the dynamics of error growth behaves differently in diverse flow regimes, the impact of flow-dependent BEC in a data assimilation system might exhibit regional variations. In this study, TC forecasts initialized from HYBRID and 3DVAR DA systems with flow-evolving ensemble-based BEC using WRF model over BoB, is evaluated statistically.

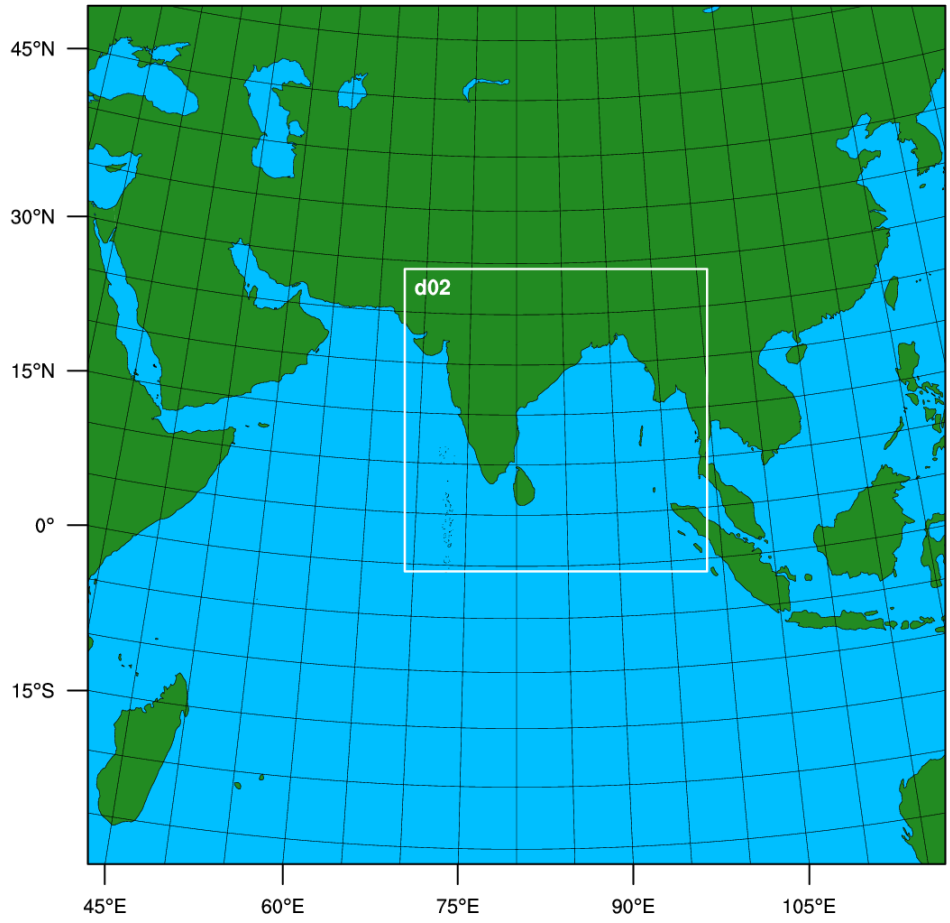


**Figure 4.1:** Model configuration deployed in this study.

### 4.2 Model Description and Configurations

The model is configured with two domains: the outer domain with 27 km and the inner domain with 9 km horizontal grid spacing (Figure 4.1). There are 36 vertical levels for both domains. The model simulations use the same set of physical parameterization schemes as mentioned in Chapter 3.

### 4.3 **Experimental design and validation**

All the assimilations are performed in the outer domain of 27 km resolution. The model initialization is started with data obtained from the 24 h NCEP GFS model forecasts, and initial ensembles are perturbed by taking random perturbations, as mentioned in Chapter 3. The initial 75 ensemble members used for conducting the HYBRID experiments are created 24 hours before the first analysis time to take care of the spin-up error.

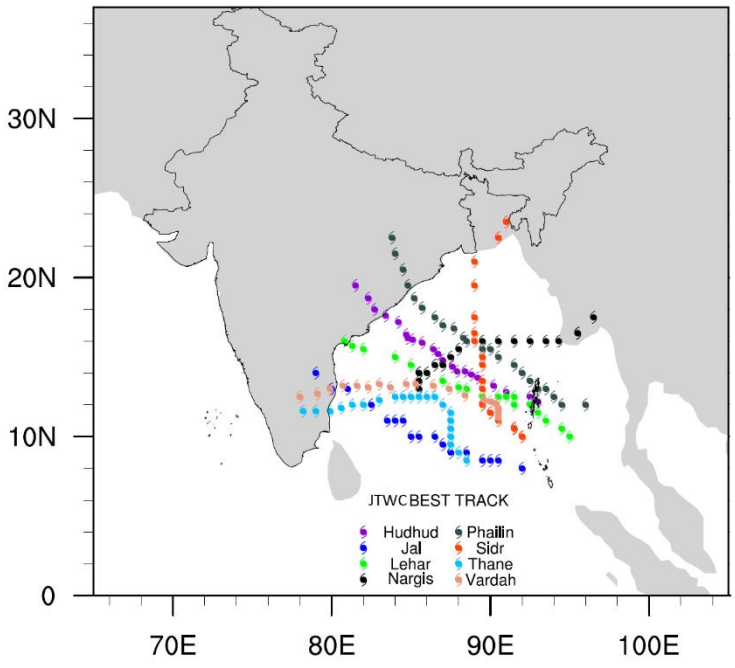


**Figure 4.2:** Best track of tropical cyclones obtained from JTWC.

In this study, eight land falling tropical cyclones that occurred over the BoB are used to assess the contribution of flow-evolving error covariance in the 3DVAR DA framework on the TC forecast. The selection of the TCs is made based on their varied intensity categories along with diverse tracks followed by the TCs before landfall. Figure 4.2 shows the observed track of the TCs obtained from the Joint Typhoon Warning Center (JTWC). It is clearly visible that excluding cyclone Nargis, the other TC cases have originated in BoB and moves in a usual northwestward direction. Three different experiments are performed for each of the TCs and the designs of the experiments are given in Table 1.

**Table1:** Design of experiments

| Data Assimilation System | Weight assigned for HYBRID (%) | | OSE name |
|---|---|---|---|
| | $\beta_1$ | $\beta_2$ | |
| HYBRID | 50 | 50 | HYBRID50 |
| HYBRID | 25 | 75 | HYBRID75 |
| 3DVAR | Not required | | 3DVAR |

DA experiments are conducted following (Wang 2011) by assimilating GTS based observations at every 12 h interval for a 48 h duration prior to the free forecast. The detailed information of the TCs along with DA duration is provided in Table 2. Finally, the free forecasts initialized from the last DA cycle of all the TCs are continued until landfall.

This study quantifies the effectiveness of flow-dependent error covariances by computing track error concerning simulated TC position and intensity error concerning simulated minimum sea level pressure (MSLP) and maximum wind speed (MWSP) with respect to the observed best track of JTWC.Further, Tropical Rainfall Measuring Mission (TRMM) data is used for validation of rainfall forecast. In addition, the bootstrap resampling method is used to make sure the statistical implication of the results by resampling the data distribution 10,000 times.

**Table2**: Details of eight TCs used in this study. Here, ESCS represents Extremely Severe Cyclonic Storm, VSCS represents Very Severe Cyclonic Storm, SSCS represents Severe Super Cyclonic storm

| **Serial No** | **Name** | **Duration of 12 hourly data assimilation cycles starting from the 00 UTC of the given date to the next 48 hour** | **Maximum sustained surface wind speed (kt)** | **TC category** |
|---|---|---|---|---|
| **1** | Sidr | 11$^{th}$ Nov 2007 | 115 | ESCS |
| **2** | Nargis | 29$^{th}$ Apr 2008 | 90 | ESCS |
| **3** | Jal | 4$^{th}$ Nov 2010 | 60 | SSCS |
| **4** | Thane | 26$^{th}$ Dec 2011 | 75 | VSCS |
| **5** | Phailin | 8$^{th}$ Oct 2013 | 115 | ESCS |
| **6** | Lehar | 24$^{th}$ Nov 2013 | 75 | VSCS |
| **7** | Hudhud | 8$^{th}$ Oct 2014 | 100 | ESCS |
| **8** | Vardah | 8$^{th}$ Dec 2016 | 70 | VSCS |

## 4.4 Results and Discussion

This section evaluates the performance of traditional 3DVAR and HYBRID DA system in the simulation of TC over the BoB.

### 4.4.1 Skill of the ensemble system

The relationship between spread and error is examined to understand the probabilistic skill of the ensemble members that estimate the flow-dependent background error covariance. For a perfect ensemble system, the total spread of ensembles is likely to have the same magnitude as forecast uncertainty. Vertical profiles of ensemble spread and root mean square error (RMSE), averaged over 8 cases are shown in Figure 4.3.

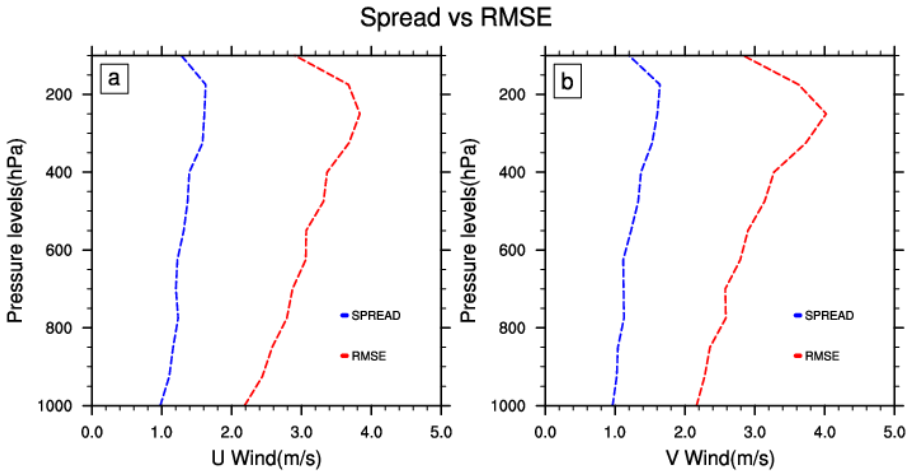


**Figure 4.3:** Ensemble spread (blue) and RMSE (red) at the first analysis cycle for variables (a) U-Wind (b) V-Wind

The ensemble spread is consistently lower than the RMSE for U-wind and V-wind at all the levels. This basically shows that the ensemble system considered in this study is under-dispersive, which points to the potential suboptimal behavior of the set of ensembles by underestimating the error variance. It is well-known that model uncertainties are usually not well represented in ensemble-based DA schemes (e.g., Berner et al. 2015). Thus, we speculate that the unrepresented model errors might have led to loss of ensemble variance and an under-dispersive ensemble system.

### 4.4.2 Analysis of TC position and intensity

The absolute errors of the analysis of track and intensity simulated by the three experiments for the eight TCs are shown in Figure 4.4. The results point o that the initial TCs positions are considerably improved in the HYBRID experiments' analysis than the 3DVAR experiments. The position errors are also relatively less in HYBRID analysis than in 3DVAR for 7 out of 8 TC cases. However, no apparent benefit is noticeable among the two HYBRID experiments in decreasing the initial position error. The experiments with the HYBRID DA system have performed better than 3DVAR in analyzing the wind speed in 4 out 8 TCs. For the rest of the TCs, the errors are more or less similar, excluding one TC. Similarly, the MSLP errors are reduced in the HYBRID experiment compared to the 3DVAR for most cases.

Figure 4.5 shows the analysis error that is averaged over all the TC cases indicated a 20% and 22% reduction in TC position error for HYBRID75 and HYBRID50 simulations, respectively, compared to 3DVAR (Figure 4.5). Furthermore, the relative improvements compared to 3DVAR for TC MWSP (MSLP) are 15% (9%) and 17% (17%), respectively,

for HYBRID50 and HYBRID75 experiments. Overall, the HYBRID method performed better than 3DVAR in analyzing both the initial position as well as intensity. However, the better performance of HYBRID compared to 3DVAR is more substantial for the initial position than intensity. Among the HYBRID experiments, while the HYBRID75 shows modest improvement compared to HYBRID50 in analyzing MWSP and MSLP, the analyzed TC initial position analysis is better simulated by HYBRID50 than HYBRID75. Furthermore, the analysed TC structures from all the experiments are examined for comprehending the factors influencing the differences in track position and intensity between the DA experiments.

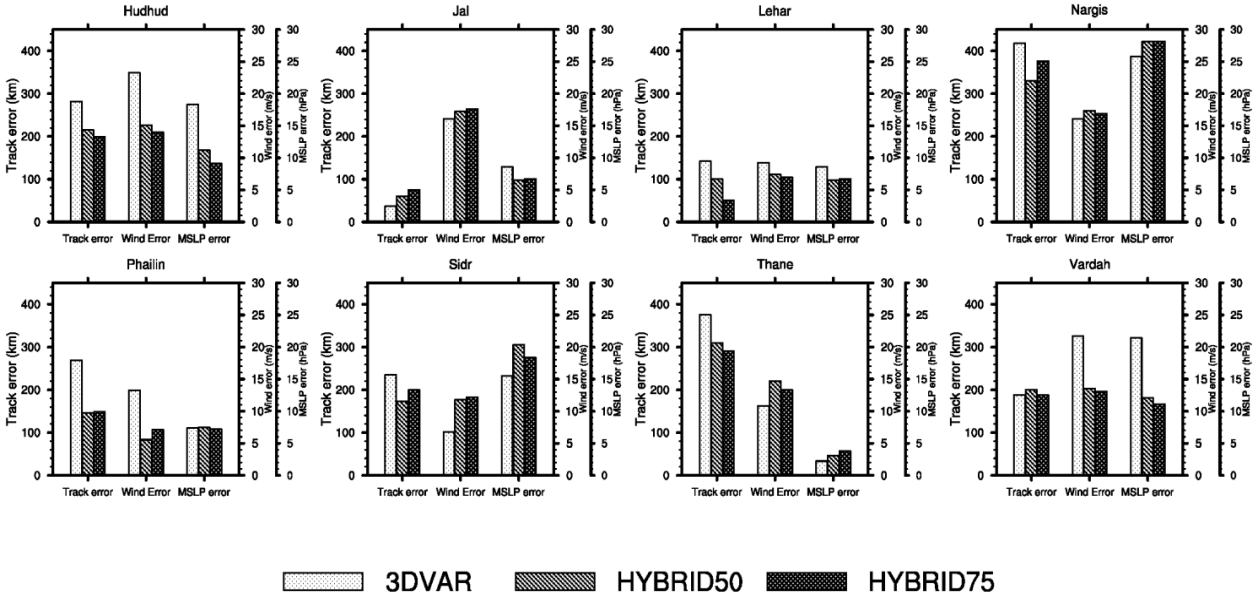


**Figure 4.4:** Analysis error averaged over all the DA cycles for track (km), maximum wind speed (m/s), and minimum sea-level pressure (MSLP) (hPa) from different experiments for each Tropical Cyclone cases.

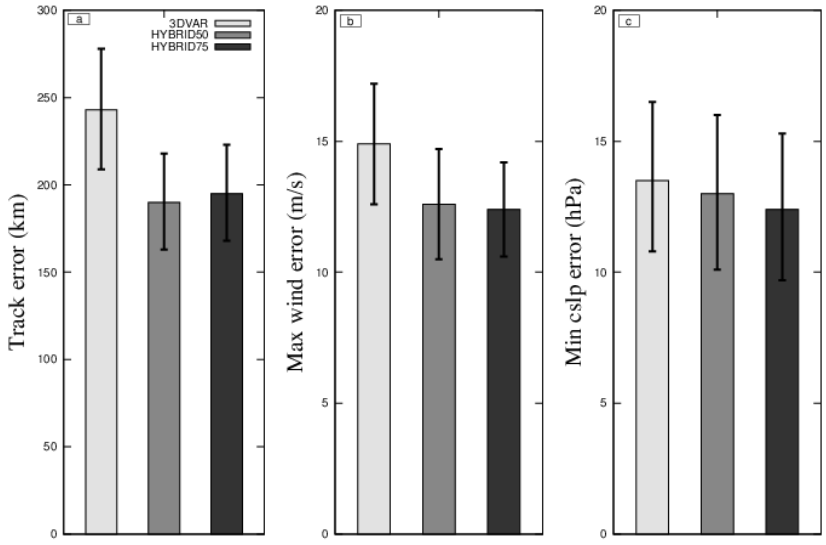


**Figure 4.5:** The analysis error averaged over all the cyclones and DA cycles for the (a) Initial position (b) Maximum wind speed (c) Minimum Sea level pressure. The error bars specify $5^{th}$ and $95^{th}$ percentiles obtained from bootstrap resampling

### 4.4.3 Analyzed TC structure

The average MSLP for the TC cases covering a $15^0 \times 15^0$ grid around the center of the TC is shown in Figure 4.6. The 3DVAR analysis results in a stronger vortex, which indicates higher intensity compared to HYBRID. It is also evident that 3DVAR analyzed a fall in pressure <980 hPa, whereas HYBRID experiments illustrated MSLP values >986 hPa. Furthermore, the composites of vertical cross-sections of horizontal wind speed overlaid with potential temperature simulated by the experiments are shown in Figure 4.7. A warm-core structure is evident from the figure, which depicts a descending bending of the potential temperature contours for all the experiments. However, the downward bending is more noteworthy in the 3DVAR experiment than HYBRID, particularly beyond 400 hPa.

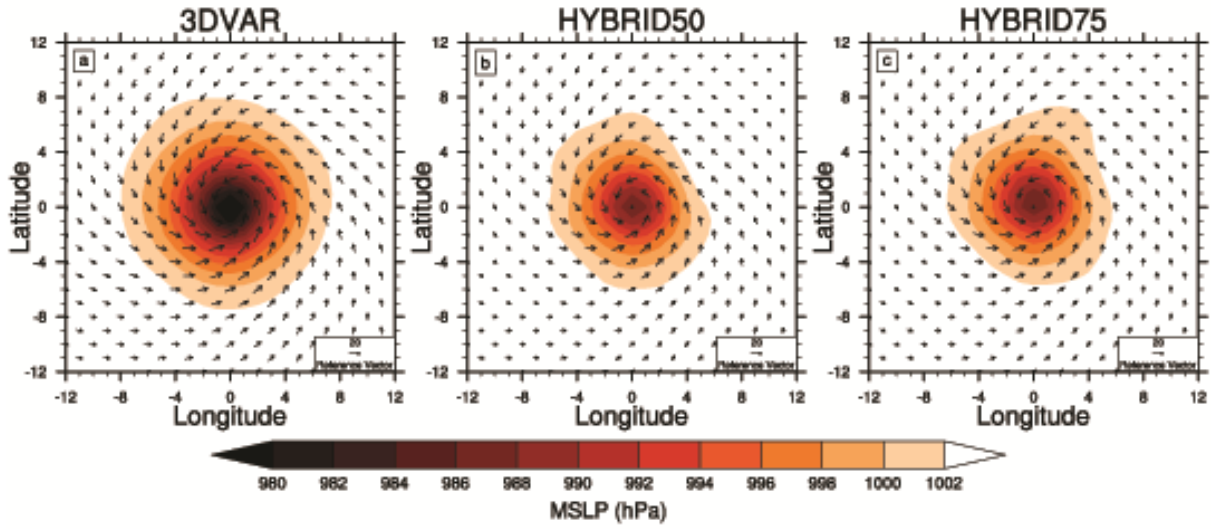


**Figure 4.6: C**omposite of analyzed MSLP (hPa) around the TC center averaged over all latitudes in a $15^{0} \times 15^{0}$ box for all the TC cases.

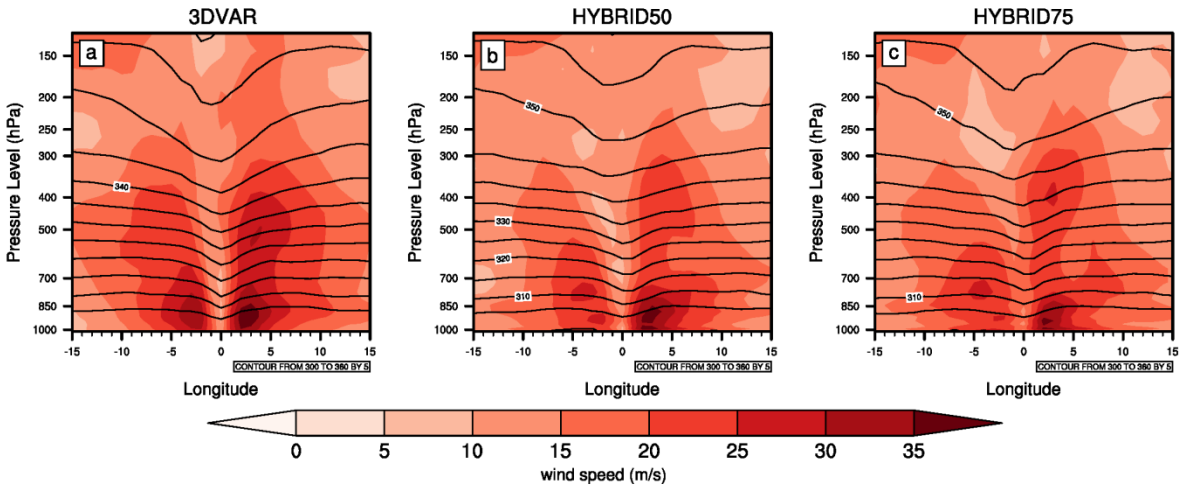


**Figure 4.7:** Composite of averaged vertical cross-section covering all latitudes in a $15^{0} \times 15^{0}$ box in terms of potential temperature (˚K; contour) and wind speed (shaded).The vertical cross-section is taken at the center of each tropical cyclone

Furthermore, the magnitude of wind speed on both the sides of the composite TC is larger in the 3DVAR experiment than in the HYBRID experiments.

Overall, the results reveal that 3DVAR analyzed a stronger and warmer TC core compared to HYBRID. Among the HYBRID experiments, HYBRID50 analysis depicted a stronger TC in terms of horizontal wind speed and potential temperature compared to HYBRID75 analysis. Though the differences amongst the experiments are observed all through the model vertical levels, more significant impacts are visible at lower levels. It could be because more observations were assimilated at surface levels than upper levels.

### 4.4.4 Track and intensity forecast

Figure 4.8 shows the average errors in track, MSLP, and MWSP forecasts for different experiments. It is observed that the HYBRID experiments outperformed the 3DVAR experiment in simulating TC tracks for all the forecast hours (Figure 4.8a). For 3DVAR, HYBRID50, and HYBRID75 experiments, the track errors are 325, 182, and 173 km, respectively. The smaller error bars from bootstrap resampling in HYBRID results compared to 3DVAR also indicated that improvements due to flow-dependent BEC are statistically significant.

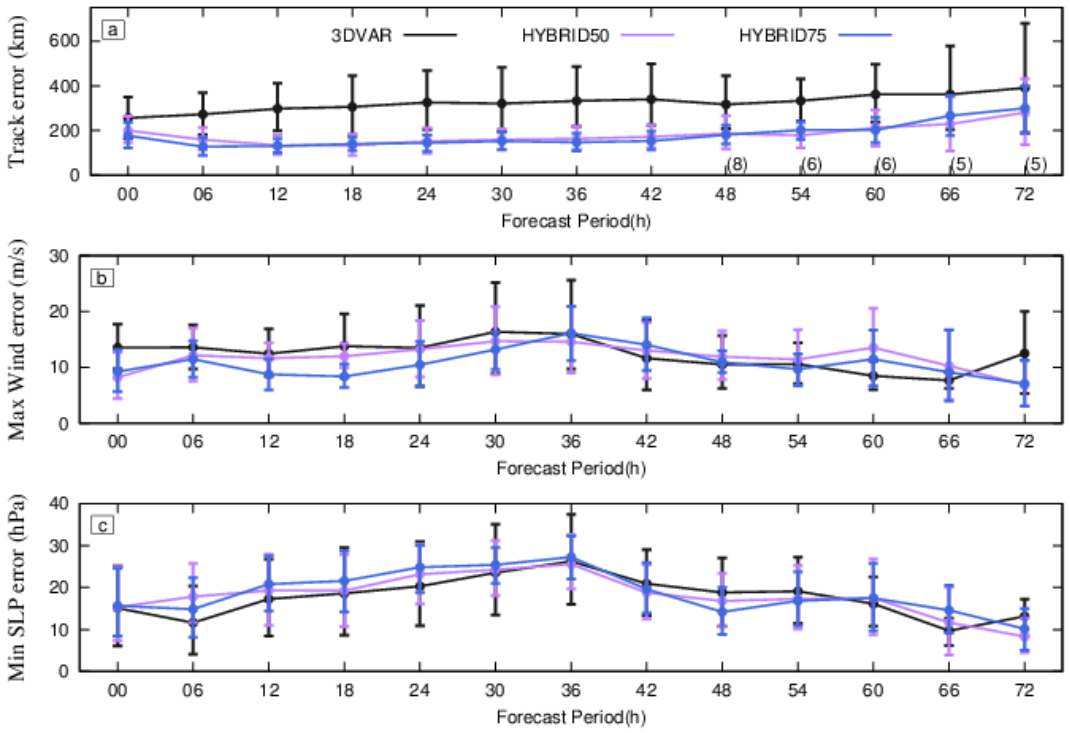


**Figure 4.8:** Mean errors in forecast of (a) Track position (b) Maximum wind speed (c) Minimum sea level pressure during the forecast intervals. The error bars represent 5$^{th}$ and 95$^{th}$ percentiles obtained by resampling using the bootstrap method. In the top panel (a), the number within a bracket shows the number of TCs available at the forecast time for validation.

On the contrary to the TC track forecast, HYBRID experiments did not show substantial improvement in MWSP prediction, except for some improvement, which has been observed during the initial forecast hours, mainly for the HYBRID75 experiment (Figure 4.8b). The relative reformation in the forecast of MWSP from HYBRID50 and HYBRID75 experiments than 3DVAR is 9% and 22%, respectively. However, the MSLP forecast errors from individual experiments did not show significant differences among themselves for the majority of the forecast intervals (Figure 4.8c).

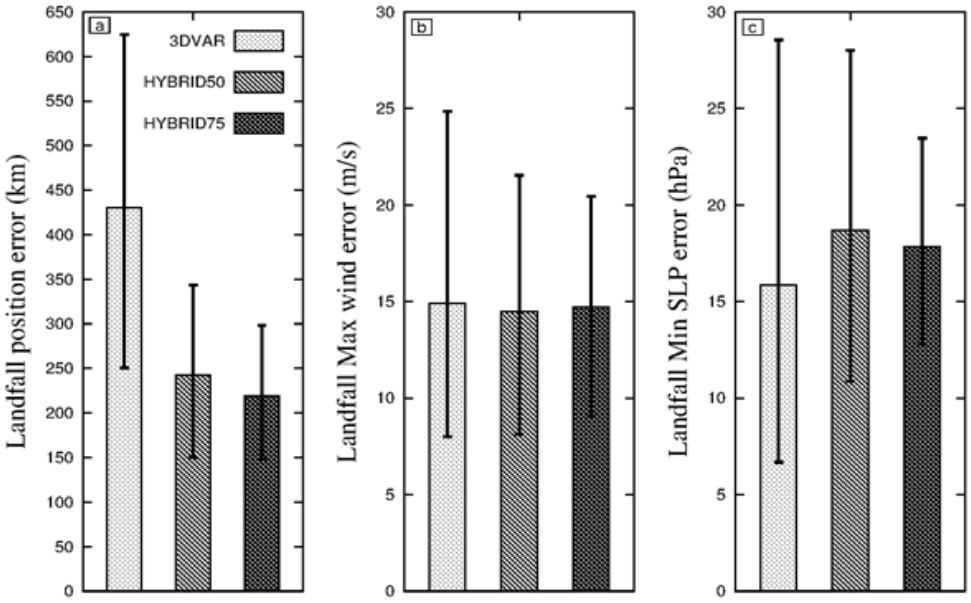


**Figure 4.9:** The averaged error of forecast in the (a) TC position, (b) Maximum wind speed (c) MSLP for each of the experiments during landfall. The averaging is done over the 8 TCs. The error bars stand for 5$^{th}$ and 95$^{th}$ percentiles calculated from bootstrap resampling

Figure 4.9 shows the forecast errors of position, MWSP, and MSLP at the time of TC landfall for 3DVAR, HYBRID50, and HYBRID75 are shown in. The average position error at the time of landfall for the 3DVAR experiment is 430 km. In comparison, the errors for HYBRID75 and HYBRID50 runs are 219 and 243 km, respectively. Thus, the relative improvements in TC position are 49% and 43% for HYBRID75 and HYBRID50 experiments compared to 3DVAR (Figure 4.9a).On the other hand, the HYBRID experiments did not show a notable advantage in intensity forecast over 3DVAR DA at the time of landfall(Figure 4.9b,c). However, the uncertainties in 3DVAR are higher than HYBRID, as depicted by the large error bars in 3DVAR as shown in Figure 4.9.

### 4.4.5 Rainfall validation

The model skill for rainfall forecast is evaluated using ETS and Bias scores. The ETS of 24 hour accumulated precipitation forecasts from different experiments during the TC landfall for individual TCs are shown in Figure 4.10.

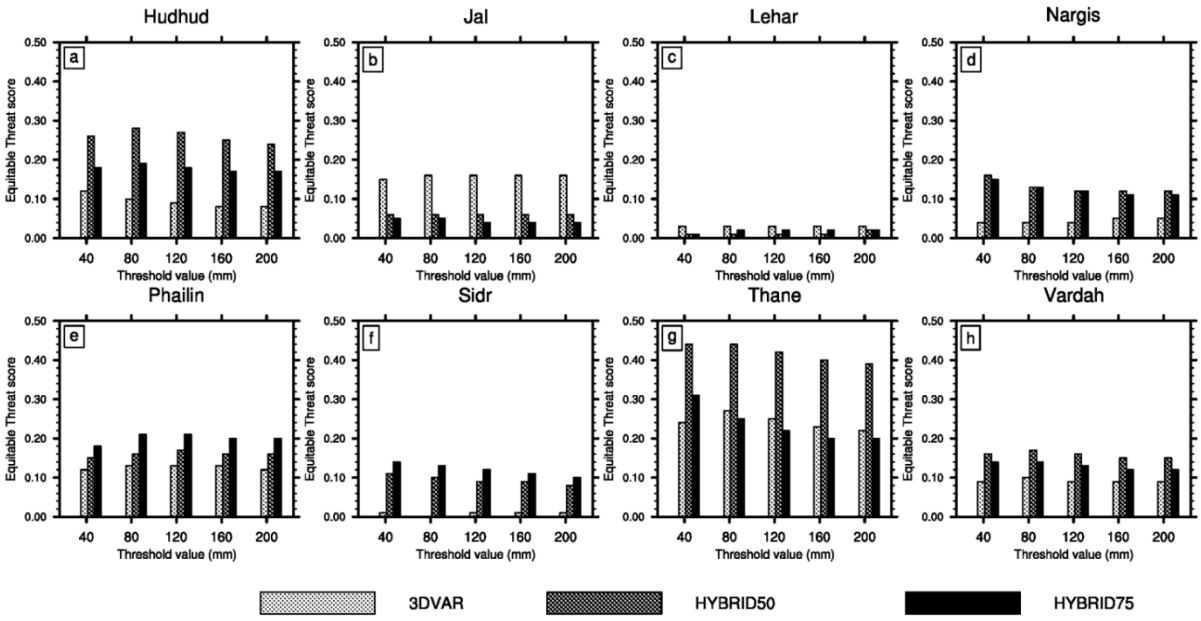


**Figure 4.10:** The ETS for forecasted 24 hours accumulated precipitation prior to the landfall of (a) Hudhud (b) Jal (c) Lehar (d) Nargis (e) Phailin (f) Sidr (g) Thane (h) Vardah.

The results signify that the rainfall forecast skill is comparatively higher for those cyclones, which pursue a typical northwest ward track compared to those that follow an altered track. The skill scores obtained for the HYBRID experiments are elevated compared to 3DVAR in most of the cyclone cases, all through the rainfall thresholds. The improvement in the precipitation forecasts from HYBRID experiments is partly because of the position accuracy in predicting landfall with a flow-dependent error covariance in the DA system.

It is also seen that the ETS values somewhat declined with a rise of rainfall threshold consistent with previous studies (Duc et al. 2013). Bias score value less (greater) than 1.0 indicates the tendency of the model to underestimate (overestimate) the precipitation amount. In general, the WRF model overestimated rainfall for almost all the TCs at the various rainfall thresholds (Figure not shown). It is found that consistent with the low ETS values, the overestimation of rainfall is very high for cyclone Thane. It may be noted here that the Thane cyclone followed a non-conventional south-westward track during the landfall (Figure 4.2). However, the use of flow-dependent error covariances substantially reduces the bias in rainfall forecast for cyclone Thane.

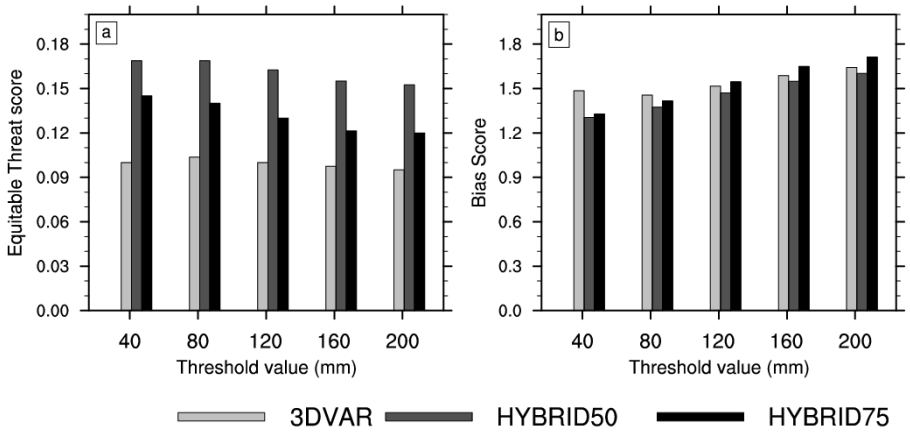


**Figure 4.11:** The averaged skill scores (a) ETS, (b) Bias score calculated for past 24 hours accumulated rainfall forecast from the landfall of TCs.

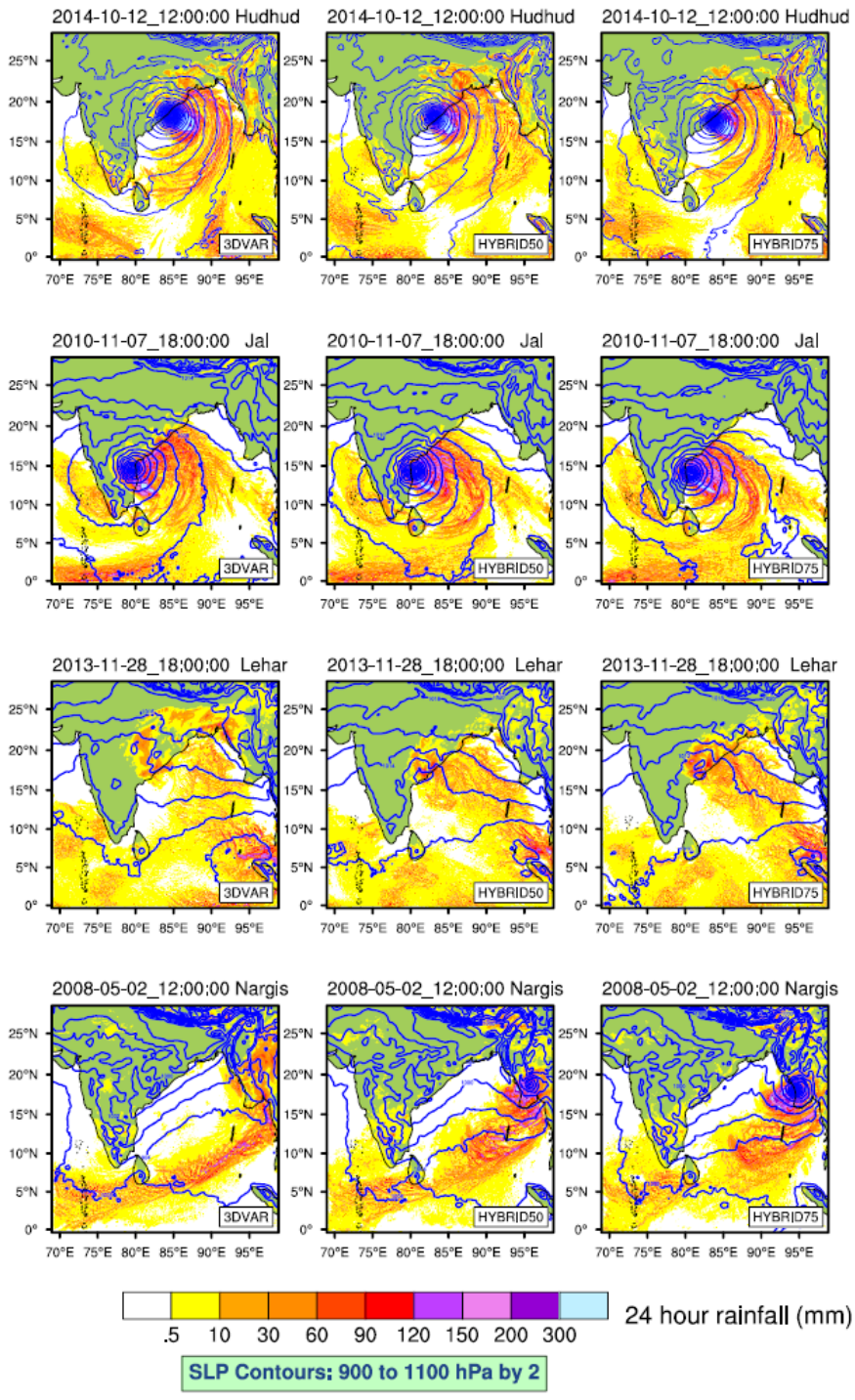


**Figure 4.12:** The spatial distribution of 24 h accumulated rainfall forecast overlaid with MSLP contours for TCs, namely for tropical cyclones Hudhud, Lehar, Jal, and Nargis

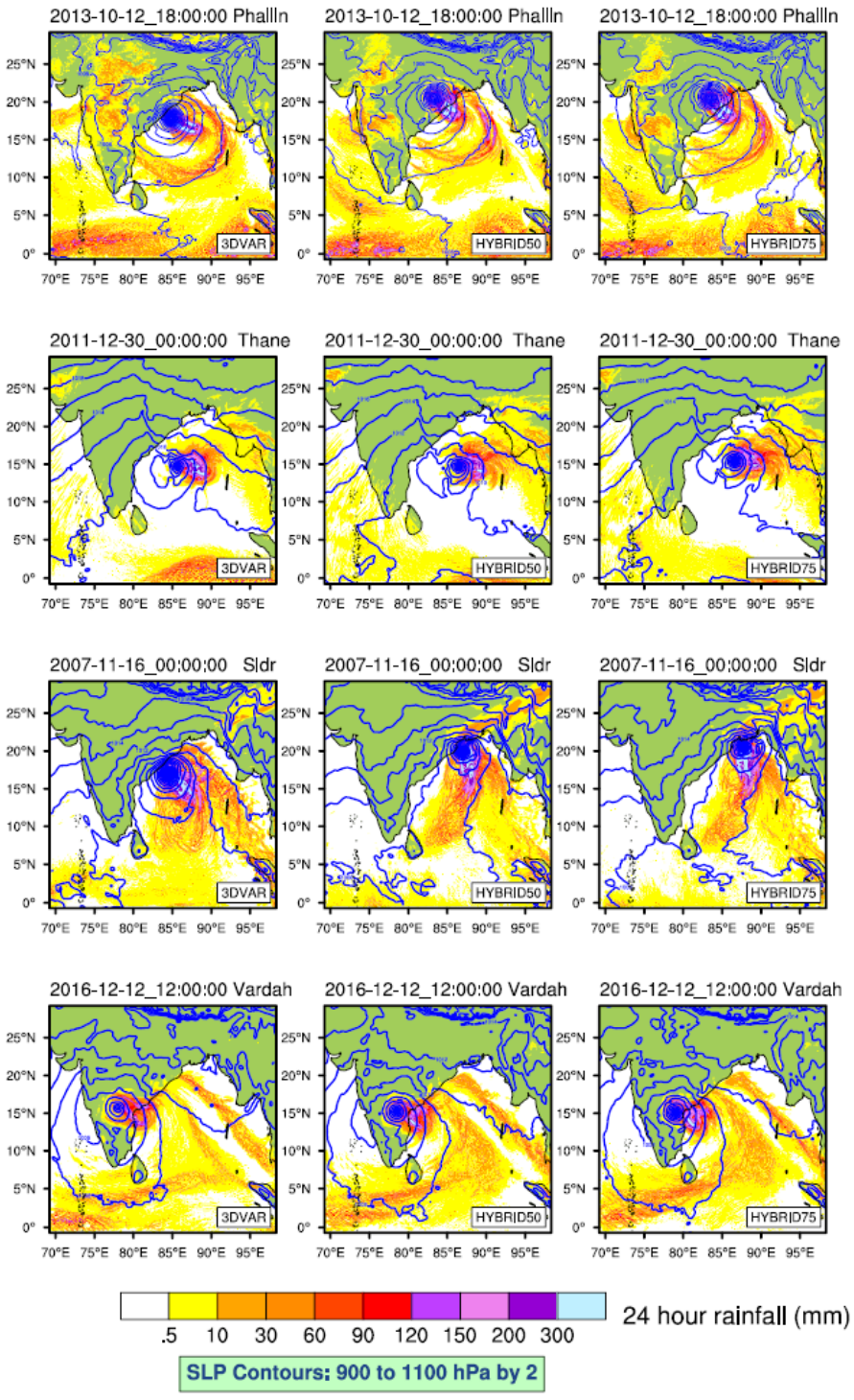


**Figure 4.13:** Same as Figure 4.12, but for TCs, namely Phailin, Thane, Sidr, and Vardah.

The averaged ETS values calculated over all the TCs showed improved rainfall forecast skill of the HYBRID experiments compared to 3DVAR at all the rainfall thresholds. Among the HYBRID experiments, the performance of HYBRID50 is better than HYBRID75 (Figure 4.11a). The average bias scores also reveal overprediction of rainfall over the verification domain in the three experiments, and bias increased with rainfall thresholds (Figure 4.11b). Furthermore, the spatial distributions of 24 h accumulated rainfall for the TC cases are shown in Figure 4.12 and Figure 4.13. To visualize the cyclone locations, the rainfall distribution is overlaid with mean sea level pressure. The HYBRID experiments show considerable improvements than 3DVAR in terms of the spatial distribution of precipitation.

**4.5 Summary**

In this study, the efficiency of flow-dependent BEC incorporated in a 3DVAR DA system is investigated for the forecast of tropical cyclones that occurred over the Bay of Bengal. The evaluations are carried out for eight land-falling tropical cyclones with varying ranges of intensities using the 3DVAR and ensemble-based HYBRID DA techniques available in the WRF DA system. The results from the experiments are inter compared and validated against the JTWC best track data and TRMM precipitation data. The TC position and intensity analysis from the HYBRID DA system are significantly better than the 3DVAR. The HYBIRD75 showed marginal advantages over HYBRID50 regarding a decrease in the analysis of TC position and intensity errors. The 3DVAR analysis depicted a prominent warm-core and a more vigorous vortex structure than HYBRID50 and HYBRID75 experiments. The TC structure was slightly weakened in the HYBRID75 experiment compared to HYBRID50. The forecasted track and landfall positions are notably better with

HYBRID compared to the 3DVAR DA system. However, in contradiction of the track prediction, improved performance in the intensity forecast is not noticeable with HYBRID DA compared with 3DVAR.The quantitative precipitation forecast during landfall time is improved with HYBRID compared 3DVAR DA system in the majority of the TC cases, all through the rainfall thresholds. Therefore, it can be inferred from the results that the superior rainfall forecasts by the HYBRID DA system compared to 3DVAR is due to the improved TC track prediction in the HYBRID simulations.The results of this study corroborate with Wang et al. (2013), in general. In the HYBRID experiments, the BEC weightings show nominal sensitivity in the analysis and forecast.

# *Chapter 5

## Assimilation of INSAT-3D Atmospheric Motion Vectors in DA system with and without flow-dependent error covariance: Impact evaluation in short range forecasts during Indian summer monsoon

### 5.1 Introduction

The rainfall over the Indian landmass during the monsoon season is largely due to India's surrounding marine regions. However, conventional observations over these regions are scant, making it difficult to simulate monsoon systems using the NWP model accurately. In such a scenario, the role of the satellite is significant that provides high-resolution global coverage of atmospheric data at a regular interval compared to the in-situ measurements. A recent study shows that INSAT-3D AMV is found to help understand the intraseasonal monsoon variability of ISM (Sankhala et al. 2019). In the current study, the impact of INSAT-3D AMVs in the HYBRID and 3DVAR DA system is evaluated for ISM simulations.

Several studies have revealed the superior performance of the flow-dependent BEC in the HYBRID system compared to the static BEC in 3DVAR in assimilating satellite wind data effectively (e.g., Zhang et al. 2018; Sawada et al. 2019). However, no such studies have been carried out using INSAT-3D AMVs over the Indian monsoon region.The current study attempts to enumerate the impact of INSAT-3D AMV assimilation in the HYBRID DA system using the WRF model for a month-long period of July 2016. The fundamental goal of this study is to understand how similar or different is the impact of INSAT-3D AMV observations in HYBRID and 3DVAR DA systems.



*The content of Chapter 5 is taken from **Gogoi, R.B.**, Kutty, G. & Boroghain, A. (2021) Intercomparison of the impact of INSAT-3D atmospheric motion vectors in 3DVAR and hybrid ensemble-3DVAR data assimilation systems during Indian summer monsoon. **Theoretical and Applied Climatology** (2021). https://doi.org/10.1007/s00704-021-03649-2*

## 5.2 Model Description and Configurations

Here, the WRF model of version 3.8.1 (Skamarock et al. 2008) is used for conducting the model simulations over a domain encompassing the monsoon region in and around India, as shown in Figure 5.1. The model is configured for a single domain at 27 km resolution with 36 vertical levels. The set of parameterization schemes used in this study are the same as mentioned in Chapter 3 and the WRF model is initialized by taking initial and boundary conditions from NCEP GFS data.

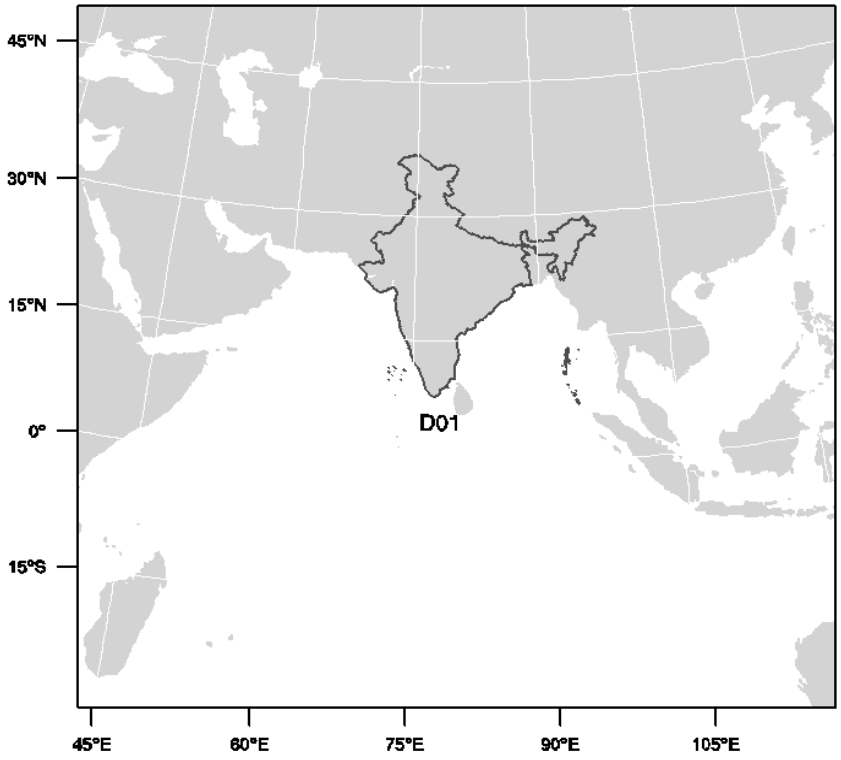


**Figure 5.1:** Model configuration deployed in this study.

### 5.3 Data used for assimilation

INSAT-3D is a meteorological satellite launched by the Indian Space research organization (ISRO) in July 2013. It is placed in the $82^0$ east longitude with the imager and the sounder onboard. The imager consists of one visible and five infrared (IR) channels, whereas the sounder consists of one visible and 18 IR channels. The details of the channels for both imager and sounder are given in Table 5.1.

Table 5.1: Details of INSAT-3D imager and sounder channels

| IMAGER CHANNELS | | |
|---|---|---|
| Channels | Resolution (km) | Wavelength (μm) |
| Visible | 1 | 0.55 - 0.75 |
| Short-wave IR (SWIR) | 1 | 1.55 - 1.70 |
| Mid-wave IR (MIR) | 4 | 3.80 - 4.00 |
| Water vapor (WV) | 8 | 6.50 - 7.10 |
| Thermal IR 1 (TIR-1) | 4 | 10.3 - 11.3 |
| Thermal IR 2 (TIR-2) | 4 | 11.5 - 12.5 μm |
| SOUNDER CHANNELS | | |
| Channels | Resolution (km) | Wavelength (μm) |
| Visible | 10 | 0.695 |
| Long-wave IR – 1 | 10 | 14.71 |
| Long-wave IR – 2 | 10 | 14.37 |
| Long-wave IR – 3 | 10 | 14.08 |
| Long-wave IR – 4 | 10 | 13.64 |
| Long-wave IR – 5 | 10 | 13.37 |
| Long-wave IR – 6 | 10 | 12.66 |
| Long-wave IR – 7 | 10 | 12.02 |
| Mid-wave IR – 1 | 10 | 11.03 |
| Mid-wave IR – 2 | 10 | 9.71 |
| Mid-wave IR – 3 | 10 | 7.43 |
| Mid-wave IR – 4 | 10 | 7.02 |
| Mid-wave IR – 5 | 10 | 6.51 |
| Short-wave IR – 1 | 10 | 4.57 |
| Short-wave IR – 2 | 10 | 4.52 |
| Short-wave IR – 3 | 10 | 4.45 |
| Short-wave IR – 4 | 10 | 4.13 |
| Short-wave IR – 5 | 10 | 3.98 |
| Short-wave IR – 6 | 10 | 3.74 |

The AMVs are determined using three successive INSAT-3D images at 30-minutes temporal gap by following the steps as mentioned below (Sankhala et al. 2020):

1) Image registration, thresholding, filtering,

2) Features/tracer selection and tracking,

3) Quality control,

4) Height assignment.

In this study, AMVs determined from low-level visible and MIR channels range from 600 hPa to 950 hPa, and upper-level WV channel data extended from 100 to 500 hPa are used. Conventional in situ observations and satellite-derived wind observations available from the GTS are also used for assimilation. AMV data is obtained from https://www.mosdac.gov.in/ for this study.

### 5.4 Experimental design and validation

The impact assessment of INSAT-3D AMVs in two DA systems, 3DVAR and HYBRID, is done by conducting four Observation System Experiments (OSEs), the details of which are given in Table 5.2. The DA experiments are performed at cyclic mode by assimilating the observations at every 12 h gap initiated from 0000 UTC 01 July 2016 to 0000 UTC 30 July 2016, and finally, a 48 h free forecast is instigated from each 0000 and 1200 UTC DA analysis for July 2016. It is to mention here that, in the HYBRID experiment equal weight has been assigned to the static and flow dependent error covariances.

**Table 5.2: Design of Experiments**

| Data Assimilation System | Observations | OSE name |
|---|---|---|
| 3DVAR | GTS | 3DVAR |
| 3DVAR | GTS+INSAT-3D | 3DVAR_AMV |
| HYBRID | GTS | HYBRID |
| HYBRID | GTS+INSAT-3D | HYBRID_AMV |

Here, 50 ensemble members are generated by adding random perturbations obtained from WRF 3DVAR. In order to avoid the spin-up issues that arise in the NWP model, the ensembles are initialized at 0000UTC 30 June 2016, 24 h prior to the first analysis time. The ensemble members are then integrated for the next 24 h and subsequently updated using ETKF. All the DA experiments using 3DVAR and HYBRID are initialized at 0000UTC 1 July 2016 using the ensemble mean. The observation error statistics for INSAT-3D AMVs have been adapted from Kumar et al. (2017), while the errors for GTS data are taken from NCEP statistics.

In this study, IMD gridded rainfall is used for rainfall validation, ERA-interim data are used for spatial verification of model-simulated wind and thermodynamic parameters at different pressure levels. For validation over oceanic region, wind data are obtained from the Advanced Scaterometer (ASCAT) onboard the European meteorological operational satellite MetOp-A at 25 km resolution. More details about ASCAT data can be found in http://apdrc.soest.hawaii.edu/data/data.php.

The statistical evaluations are performed using different verification metrics; namely, mean error (ME) or Bias, root mean square error (RMSE), and improvement parameter (η). In addition, Bias Score (BS) and Equitable Threat Score are used for rainfall forecast validation.

## 5.5 Results and Discussion

This section evaluates the impact of INSAT-3D AMVs in the 3DVAR and HYBRID DA system in the short-range forecast during the Indian summer monsoon (ISM).

### 5.5.1 Analysis and forecast profile verification

Figure 5.2 shows the root mean square fit of the monthly averaged vertical profile of the analysis to the observed zonal (U Wind) and meridional (V Wind) winds, temperature, and water vapor mixing (Q Vapour) ratio obtained from radiosonde observations available with the NCEP data. Averaging of the analysis field is done over the domain as shown in Figure 5.1 and the 60 DA cycles. The results show that the HYBRID fits more closely to the observations of U Wind and V Wind components than HYBRID_AMV and the 3DVAR experiments. Whereas, for temperature and Q Vapour analysis field, both the 3DVAR experiments show almost similar fit, but slightly better than HYBRID experiments.

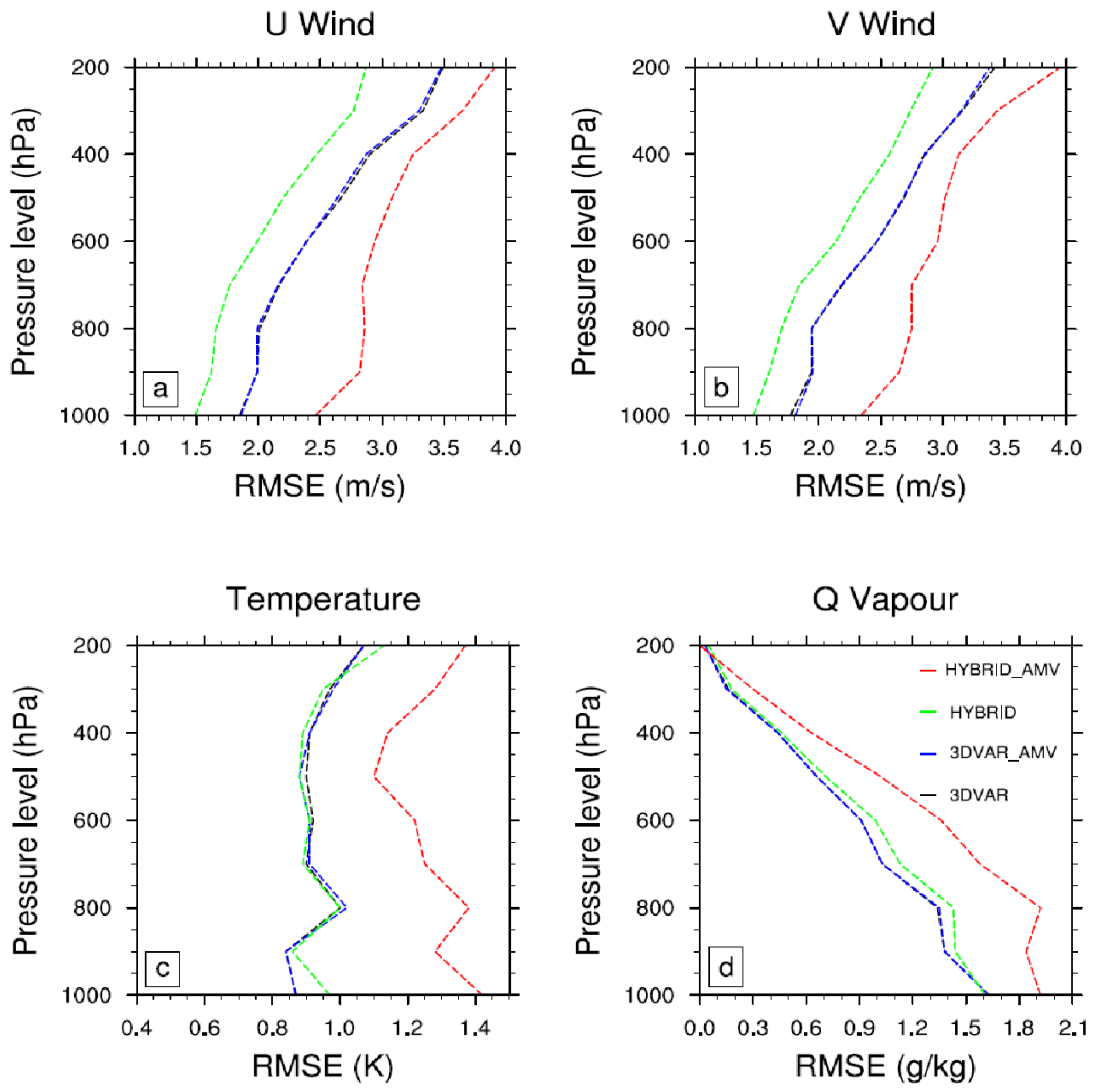


**Figure 5.2:** Vertical profiles of root mean square fit of analysis to radiosonde observations for different prognostic variables; analyzed by 3DVAR (black line),3DVAR_AMV (blue line), HYBRID (green line) and HYBRID_AMV(red line) experiments.

It is also important to mention here that the analysis fit to observation does not represent the analysis error. As the observed radiosonde data are already assimilated to the analysis, the root mean square fit only shows how close the analysis fields are to the observations. A study by

Wang (2008) reveals that the analysis fit to the observations depends on the various factor of the BEC setting in the DA system, such as background error variance, larger correlation length scale. Another study by Zhang et al. (2011) demonstrates that the better fit of analysis to observation does not necessarily result in a better forecast.

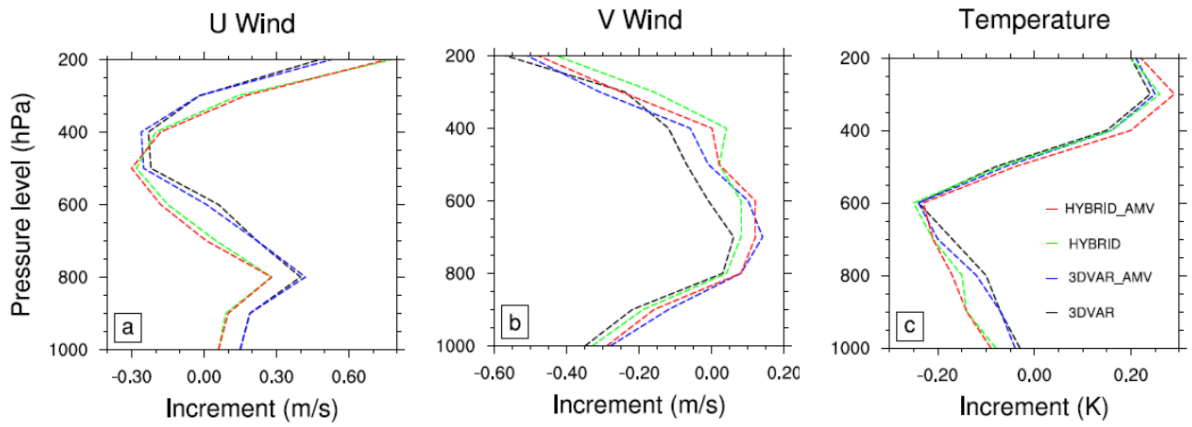


**Figure 5.3:** Vertical profiles of analysis increment (Analysis-Background) for three different prognostic variables; analyzed by 3DVAR (black line), 3DVAR_AMV (blue line), HYBRID (green line) and HYBRID_AMV (red line) experiments.

Figure 5.3 shows the vertical profile of analysis increment at theradiosonde locations of the NCEP data, where the impact of the two DA systems in the analysis increment is quite visible. Both the

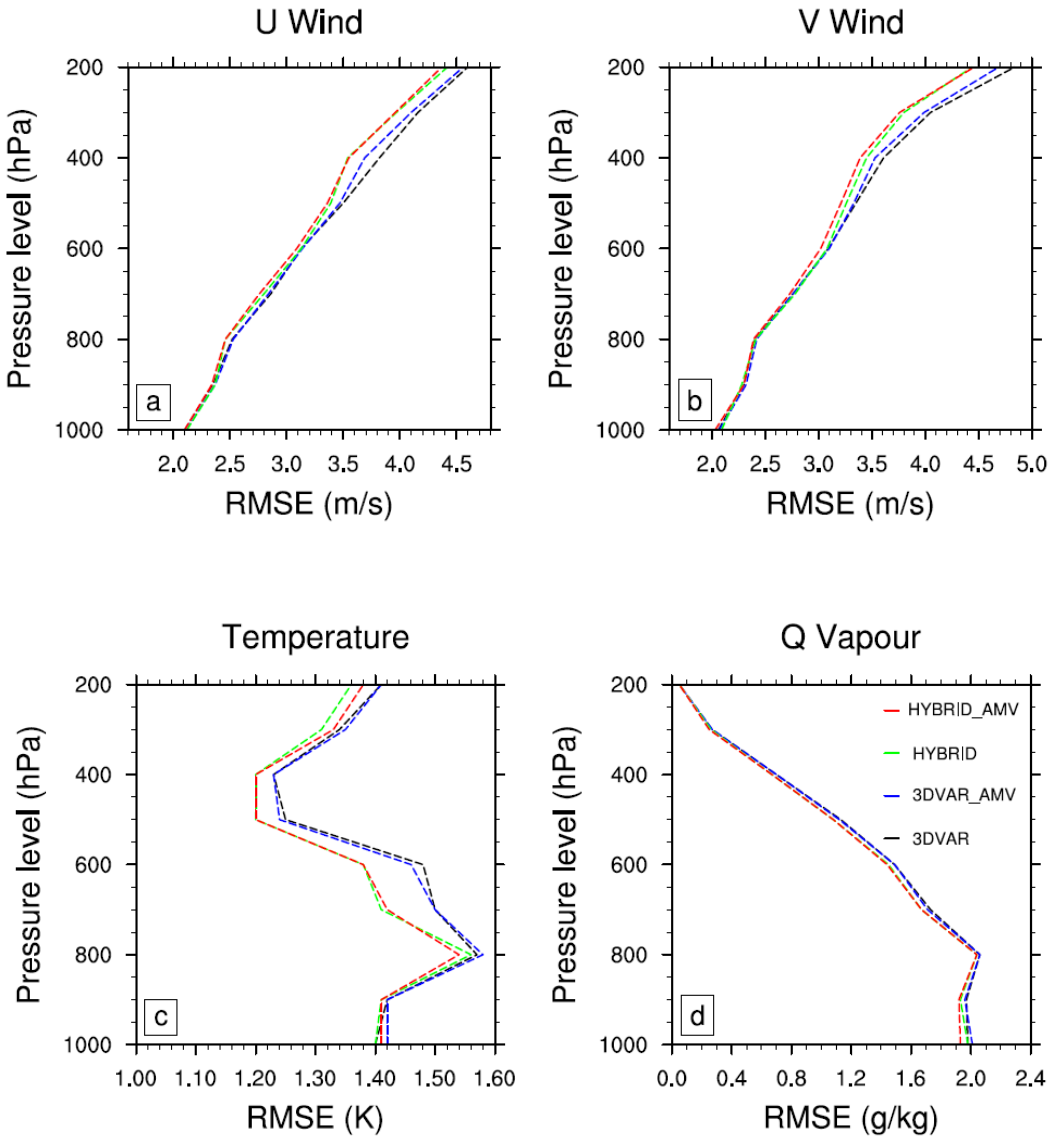


**Figure 5.4**: Vertical profiles of root mean square errors in 24 h forecasts from different experiments compared to radiosonde observations for different variables; 3DVAR (black line), 3DVAR_AMV (blue line), HYBRID (green line) and HYBRID_AMV (red line).

HYBRID experiments show significant upper-level zonal wind increment from 400 to 200 hPa. Although no remarkable impact of AMV assimilation is observed in any of the DA systems for zonal wind, the AMV influence is seen in the 3DVAR DA experiment for meridional wind

analysis increment from 800 to 400 hPa. The result for the temperature field shows a negative increment for all the experiments from 1000 hPa to 500 hPa, with HYBRID experiments having a higher magnitude of increment. Above 500 hPa, HYBRID_AMV results in a larger positive increment compared to the rest of the experiments.

Furthermore, the vertical profile of the monthly averaged 24 h forecast of the four variables is compared with radiosonde observations and shown in Figure 5.4, where the forecasts are initiated from the analysis at 0000 UTC of each day July 2016. The results specify that both the HYBRID experiments are more accurate than the 3DVAR experiments. Wind forecasting primarily reflects an improvement in the upper troposphere level, whereas the improvement in temperature forecasting in the HYBRID experiments is significantly more prominent at atmospheric levels ranging from 800 to 200 hPa than in the 3DVAR experiments. AMV shows no significant impact on HYBRID, while a moderate impact of AMV is evident in the 3DVAR DA system. However, it is to be noted that the forecast validations are performed regarding radiosonde observations that are not present over oceans. Therefore, it is not feasible to remark on the utility of the INSAT-3D AMVs from this result when verified over limited radiosonde observations.

### 5.5.2 Spatial forecast verification

The spatial representation of three atmospheric fields, including wind at 850 hPa level, tropospheric temperature (TT) averaged over 700 to 200 hPa level, and relative humidity (RH) from ERA-interim reanalysis data for July 2016 is shown in Figure 5.5. a,e,i. The spatial distribution of the improvement parameters ($\eta$) for these variables is shown in Figure 5.5b-d,f-h,j-l. The improvement parameters' positive (negative) values reflect percentage improvement

(degradation) in the model forecast compared to the 3DVAR experiment. HYBRID experiments show higher positive η values for wind and TT variables, which indicates the positive impact of flow-dependent error covariances in HYBRID. In addition, AMV DA in HYBRID depicts significant improvement compared to the rest of the experiments with 77% and 70% positive η values for wind and TT forecast, respectively. Although the improvement percentage for the moisture variable is not very remarkable for any of the experiments compared to 3DVAR, AMV DA demonstrates marginal improvement in both 3DVAR and HYBRID DA systems.

A dipole-like pattern with positive and negative impact is observed over the western Arabian Sea (AS) due to the presence of Low Level Jet (LLJ) in Figure 5.5 b-d. LLJ is also known as Findlater Jet (Findlater 1978) is an important lower level circulation feature of the Indian summer monsoon, which carries extensive moisture from the Indian Ocean producing rainfall over the Indian subcontinent. Therefore, the improvement in the forecast of LLJ is expected to improve the rainfall over Indian landmass.

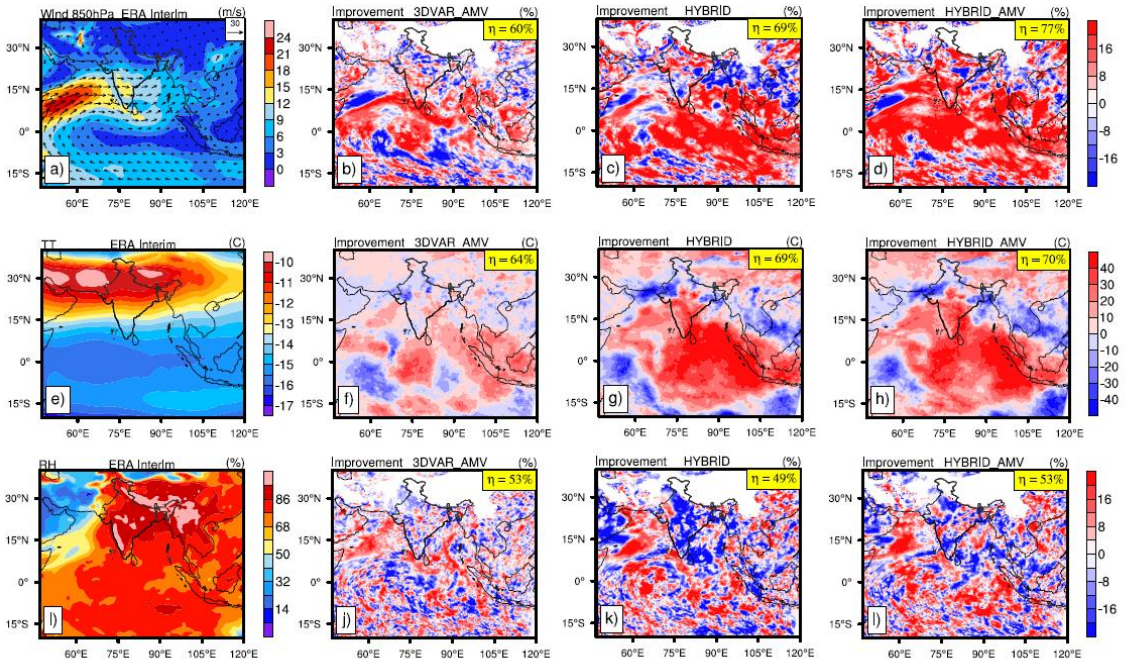


**Figure 5.5**: a) Monthly averaged wind (m/s) at 850 hPa level, e) tropospheric temperature (TT) ($C^o$) averaged over 200 hPa to 700 hPa level,i) relative humidity (RH) (%) at 850 hPa level from ERA interim and spatial distribution of improvement parameter (η ) in 24 h forecast of wind (b-d), TT (f-h) and RH (j-l) from 3DVAR_AMV, HYBRID and HYBRID_AMV system with respect to 3DVAR system.

Apart from wind, a substantial improvement of 70 % positive η values for the TT variable is observed in the HYBRID_AMV experiment (Figure 5.5h). The improvement percentage for RH is not very significant compared to wind and TT. However, marginal improvement due to AMV DA is observed both in 3DVAR and HYBRID.

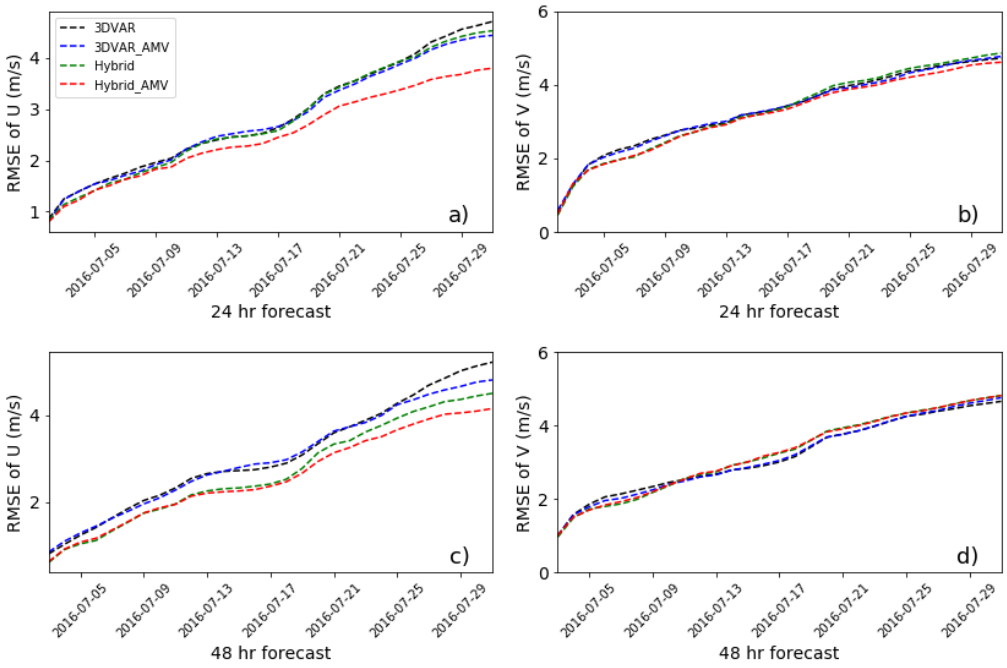


**Figure 5.6:** RMSE in 24 h (a,b) forecast and 48 h (c,d) forecast of U(m/s) and V(m/s) at 850 hPa calculated over Arabian Sea from assimilation experiments with respect to ERA interim data for 3DVAR, 3DVAR_AMV, HYBRID, HYBRID_AMV

Figure 5.6 shows the time series of area-averaged RMSE of zonal and meridional wind over the AS, respectively, for both 24 h and 48 h forecasts. The verification is confined to AS to understand the impact of experiments on LLJ. It is evident from Figure 6a and 6c that the assimilation of INSAT-3D AMV observations has significantly reduced the 24 h and 48 h forecasts errors for zonal winds in HYBRID_AMV experiment as compared to other experiments, and the impact of assimilation increases with increase in time that is indicative of the cumulative impact of assimilation. Though the HYBRID experiment does not significantly improve 24 h forecasts, the reduction in forecast errors is evident in the 48 h forecast. In the

3DVAR DA system, the positive impact of INSAT3D-AMV observations is noticeable during the later DA cycling hours.

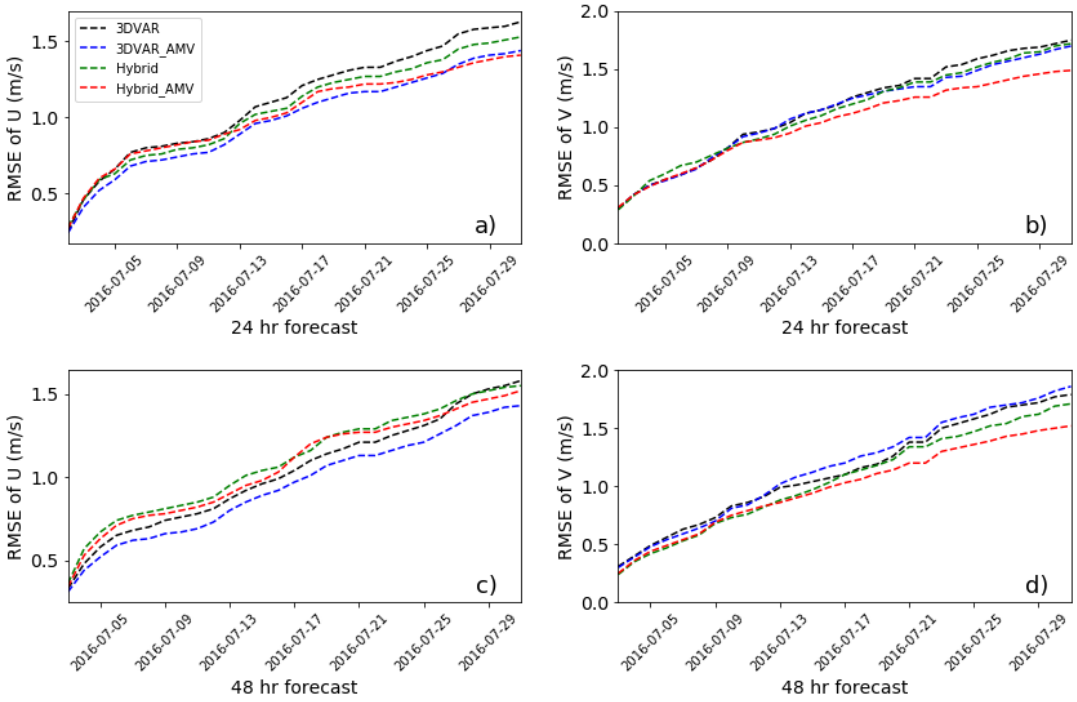


**Figure 5.7:** RMSE in 24 h forecast (a,b) and 48 h forecast (c,d) of 10m-U(m/s) and 10m-V(m/s) calculated over Arabian Sea from assimilation experiments with respect to ASCAT data for 3DVAR, 3DVAR_AMV, HYBRID, HYBRID_AMV

To explore further on the impact of observations closer to the surface, the 10-m surface wind forecast is evaluated with respect to ASCAT wind observations over the AS region. Figure 5.7shows the time series plot of monthly averaged RMSE of 24 h and 48 h near surface zonal and meridional wind forecast over the AS. Similar to the results obtained in ERA-interim validation, the assimilation of AMV observations has produced substantial improvements in the wind forecasts in both 3DVAR and HYBRID DA systems, in general. However, the 3DVAR_AMV experiment depicts a higher reduction in forecast errors than the other

experiments for zonal wind. The improvement in meridional wind component is more pronounced in HYBRID_AMV run as compared to other experiments for both 24 h and 48 h forecasts.

### 5.5.3 Rainfall forecast verification

Figure 5.8 shows the mean error (Bias) of the model simulated 24 h rainfall forecast with respect to IMD gridded rainfall. The HYBRID experiments show lower Bias in precipitation as compared to the 3DVAR run. It can be seen that the wet bias in Central India (CI) and dry bias to the south of CI have shown considerable reduction in HYBRID experiments. The AMV experiments do not depict any significant change in Bias when compared to its corresponding control DA experiments. Further, the difference of 3DVAR forecasted rainfall from 3DVAR_AMV experiments shows no significant difference zone. The difference of 3DVAR from HYBRID, HYBRID_AMV experiments shows a considerable difference zone over the same region where HYBID has improved wet bias and dry bias compared to 3DVAR. HYBRID_AMV also indicates a significant difference over India's southern CI and northeastern region for both 24 h and 48 h forecast (Figure not shown).

To quantitatively evaluate the precipitation forecasts, skill scores such as ETS and Bias scores are calculated for various experiments. It is to be noted that the skill scores are calculated in two phases of experiments: Phase-1 (2 – 16 July 2016) and Phase-2 (17 – 31 July 2016), which is represented in Figure 9. The skill of 24 h precipitation forecast for HYBRID experiments is found to be higher than 3DVAR experiments in Phase-2towards higher rainfall thresholds, which is evident from ETS values (Figure 5.9c). The HYBRID_AMV experiment shows improved skills for precipitation forecast for higher rainfall thresholds as compared to the

3DVAR_AMVexperiment in Phase-2. Bias score indicates that all the experiments in Phase-1 and Phase-2 show overestimation of rainfall, and the results are more pronounced in 3DVAR experiments in Phase-2.

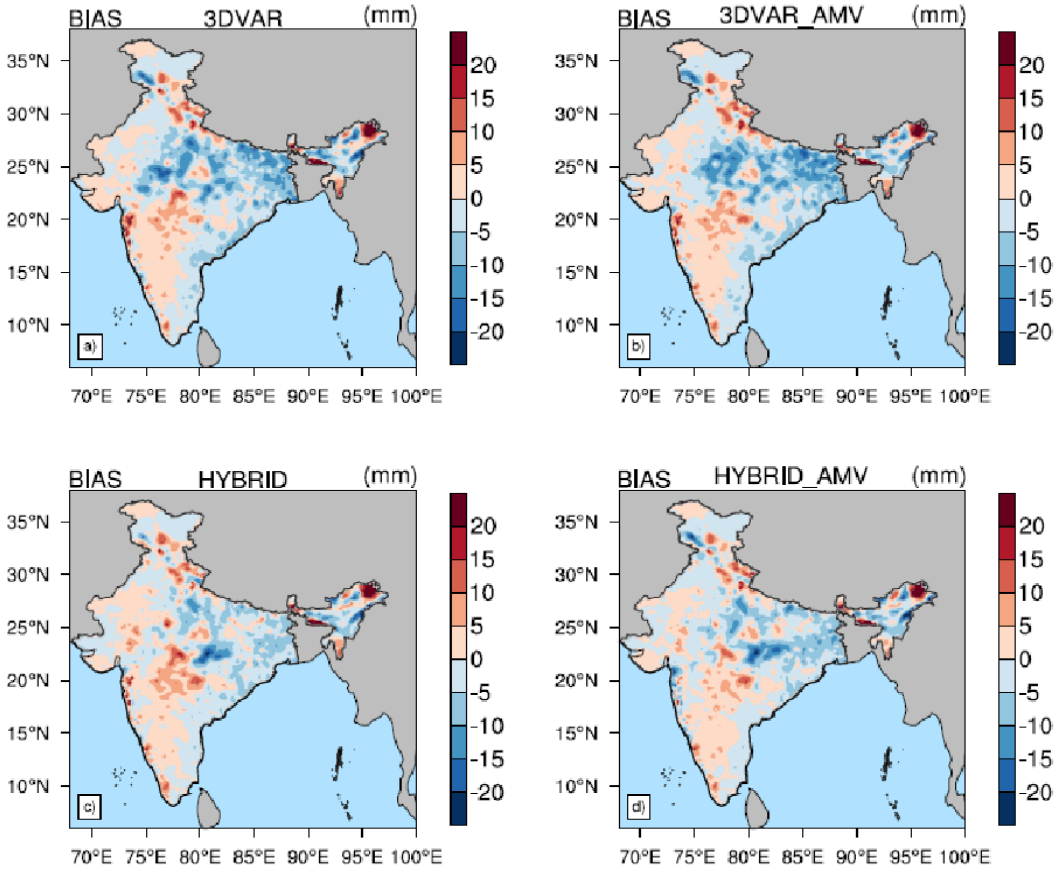


**Figure 5.8:** BIAS of monthly averaged (July) 24 h forecasted rainfall (mm/day) with respect to IMD gridded rainfall for (a) 3DVAR (b) 3DVAR_AMV (c) HYBRID and (d) HYBRID_AMV

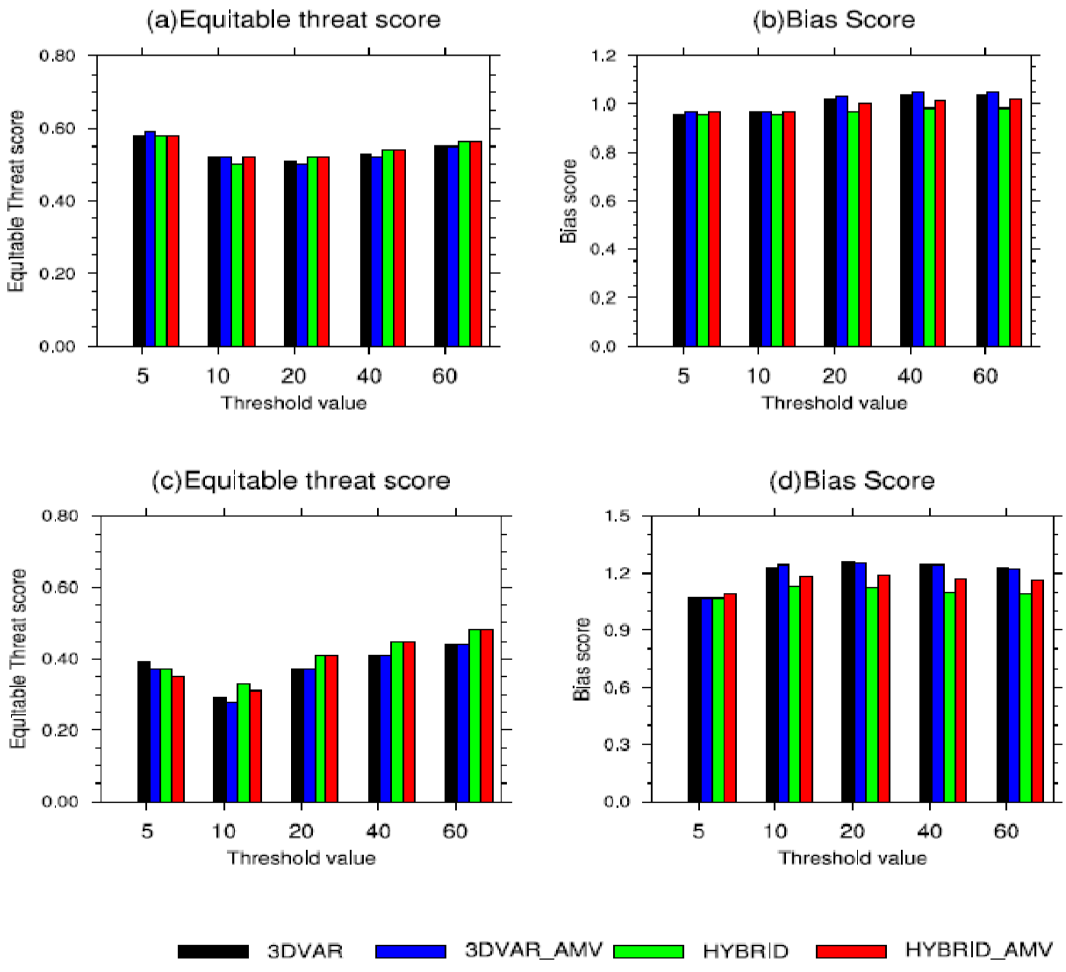


**Figure 5.9:** (a) ETS and (b) Bias scores valid from 2nd July 2016 to 16th July 2016 (first phase) and the (c) ETS and (d) Bias scores valid from 17th July 2016 to 31st July 2016 (second phase) for different rainfall thresholds computed over the Indian land mass averaged over the 24 hour forecasts

Similarly, 48 h rainfall forecast results do not show substantial improvement in both the HYBRID experiment in Phase-1 (Figure 5.10a). However, in Phase-2, the skill scores indicate modest improvements in rainfall forecast in moderate-high rainfall threshold for HYBRID experiment compared to that of 3DVAR (Figure 5.10c). Furthermore, the assimilation of AMV observations does not significantly improve the precipitation forecast in both 3DVAR and HYBRID experiments.

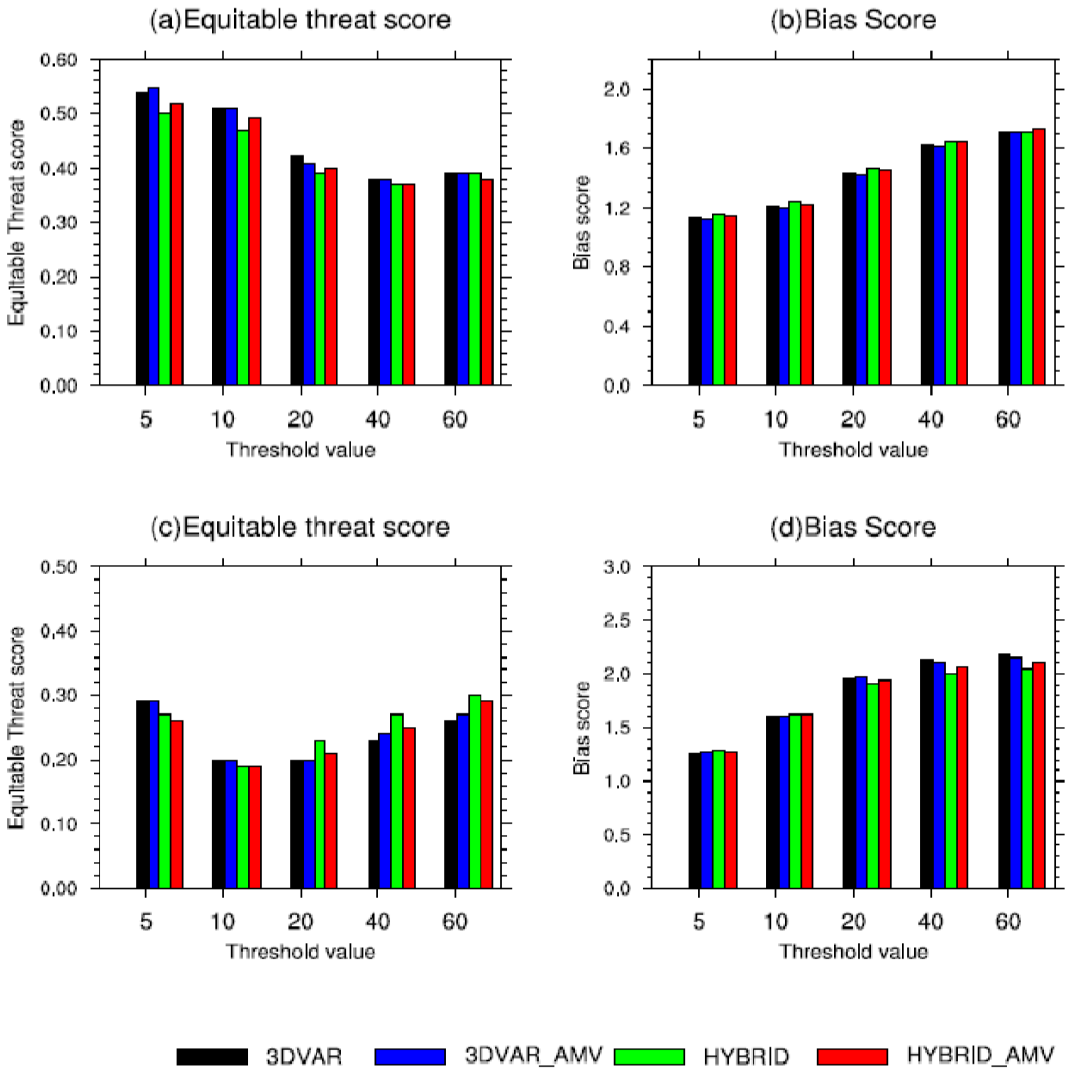


**Figure 5.10:** Same as Figure 5.9, but for the 48 hour forecasts

## 5.6 Summary

In this study, the impact of the assimilation of INSAT-3D AMV in the two DA systems for short-range forecast during the Indian summer monsoon season is evaluated. The DA systems used in this study include 3DVAR and HYBRID ETKF-3DVAR available in the WRF modeling system. The DA cycling experiments are performed for the ~4 week period of July 2016and a 48 h model forecast is generated from each analysis.

The results indicate that 3DVAR analysis fits more closely with the observations than HYBRID analysis. The domain-wide verification over the Indian landmass with respect to radiosonde observations discloses that forecasts in HYBRID experiments are more accurate than the 3DVAR experiments, in general. The wind forecasts show more improvements near the upper troposphere for the HYBRID run, with the slight impact of INSAT-3D AMV observations. In comparison with the forecasts from HYBRID analysis, the impact of INSAT-3D AMV observations is more pronounced in the 3DVAR DA system for wind forecasts over land. The spatial distribution depicts the positive impacts of INSAT-3D AMV observations across the whole domain for both HYBRID and 3DVAR experiments. The AMV observations show a superior relative impact in HYBRID than in 3DVAR, and the relative improvement in comparison to 3DVAR is 77% for wind and 71 % for tropospheric temperature. Time evolution of forecast errors with respect to ERA-Interim analysis in the zonal wind over the Arabian Sea indicates a larger growth rate in the 3DVAR experiment in comparison to the HYBRID experiment, while the assimilation of AMV observations considerably reduces forecast errors in both DA systems. The HYBRID_AMV experiments show improvement in the meridional component of near-surface winds when validated against ASCAT observations. The skill scores

for quantitative assessment of precipitation forecast specify a modest improvement in rainfall for the HYBRID run only. Incorporating the AMVs in the DA process does not considerably enhance the skill of 24 h and 48 h rainfall forecast.

The present study attempts to quantify the impact of INSAT-3D AMV observation in the 3DVAR and the hybrid ETKF-3DVAR DA system. The HYBRID DA system incorporates flow-dependent ensemble BEC that generates optimal analysis through increments consistent with the background flow and responds adaptively to the change in the observing system. Hence, it is expected that the impact of the observing system may vary depending on the DA system used. As a matter of fact, the results from the study indicate that the impact of INSAT-3D AMV observations varies in 3DVAR and HYBRID DA systems. Furthermore, the impact of the new observing system shows more value to the advanced DA systems such as HYBRID than the traditional 3DVAR approach. However, the ensemble system needs to be correctly configured for the DA system to perform optimally.

# *Chapter 6

## Assimilation of INSAT-3D Atmospheric Motion Vectors in data assimilation system with and without flow-dependent error covariance: Impact evaluation in short range forecasts of tropical cyclones over Bay of Bengal

### 6.1 Introduction

The performance of data assimilation methods is controlled by the good quality observations spread over the study area. For tropical cyclone, which originates and spends most of their duration over the ocean, the unavailability of conventional data poses difficulties in its accurate prediction. However, meteorological satellites with fairly good coverage over the ocean serve as a valuable source of atmospheric observations for assimilation. Among the different satellite products, the wind vector is shown to impact the forecast of TCs significantly. Several former studies have revealed the positive impact of the AMV observations in the TC forecast (Soden et al. 2001; Deb et al. 2011; Lim et al. 2019; Lewis et al. 2020). Studies have shown assimilation of AMV has significantly improved the initial position errors along with the track forecasts of cyclones formed over the BoB using the 3DVAR DA system(e.g., Deb et al. 2010; Greeshma et al. 2015). The positive influence of HYBRID DA in improving TC prediction has also been reported in several research articles (Hamill et al. 2011; Wang 2011; Shen et al. 2016; Lu et al. 2017; Kutty et al. 2018; Malakar et al. 2020a). However, studies with AMV DA in the HYBRID DA system are rare for TC simulation over the North Indian Ocean (NIO). The present work examines the impact of INSAT-3D AMVs in 3DVAR and the HYBRID DA systems with the major objective to understand how different or similar is the influence of INSAT-3D AMV when



*The content of Chapter 6 is taken from ***Gogoi, R.B.***, Kutty, G. & Boroghain, A. (2021) Intercomparison of the Impact of INSAT-3D Atmospheric Motion Vectors in 3DVAR and Hybrid Ensemble-3DVAR Data Assimilation Systems on the Forecast of Tropical Cyclones formed over the Bay of Bengal. **Modeling Earth Systems and Environment, 1-11.***

assimilated by 3DVAR compared HYBRID DA system for five TCs over the BoB using WRF model.

### 6.2 Model Description and Configurations

The WRF-ARW model version 3.9.1 (Skamarock et al. 2008) is used for simulation in this study. The model configuration (Figure 4.1) and parameterization schemes are same as mentioned in Chapter 4. The model is initialized with NCEP GFS analysis data. A recent article by Malakar et al. (2020b) has reported that GFS analysis is most suitable for the TC evolution study over the NIO.

### 6.3 Data used for assimilation

For the assimilation experiments, GTS-based conventional in-situ observations and satellite winds available from GTS and INSAT-3D are used in this study. A screenshot of the data spread over the study area is shown in Figure 6.1. Here, GTS-AMV represents atmospheric motion vectors obtained from NCEP GTS data. The sources of AMV data is given in Table 6.1. Detailed information about the AMV types can be found in https://rda.ucar.edu/datasets/ds735.0/.

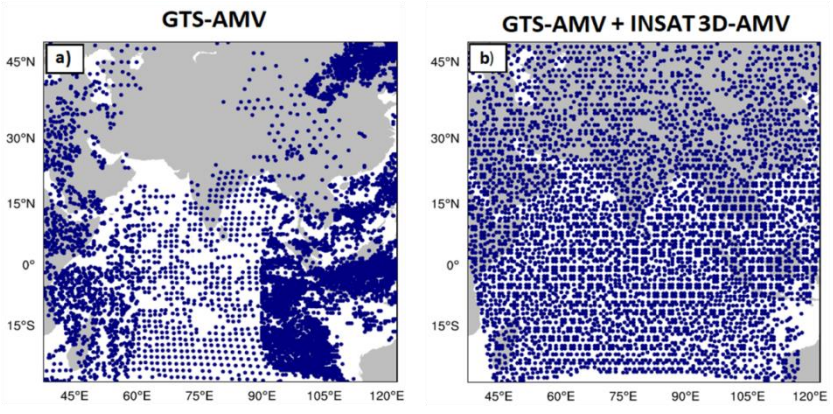


**Figure 6.1:** A screenshot that shows the spreading of atmospheric motion vector obtained from (a) GTS and (b) both GTS and INSAT-3D available at 00 UTC 10 October 2014

Table 6.1: Observation platform and AMV types of NCEP GTS data

| **SATELLITE** | **AMV TYPE** |
|---|---|
| GOES | SWIR, LWIR, WV, Visible |
| INSAT-3D | LWIR, WV, Visible |
| MTSAT/JMA | IR, WV, Visible |
| METEOSAT/EUMETSAT | IR, WV, Visible |
| AQUA/TERRA MODIS | IR, WV |
| NOAA-series/METOP-series | LWIR |
| SNPP/NOAA-20 | LWIR |

It is clearly visible that INSAT-3D AMV, together with the AMV observations available from the GTS, depicts a broader spatial coverage over the computational domain.

### 6.4 Experimental design and validation

Based on the INSAT-3D data availability, five TCs with varying intensity centered over the BoB over a period of 2014 to 2020 are considered. The details pertaining to the selected TCs are given in Table 6.1 and best track obtained from IMD is represented in Figure 6.2. The impact of INSAT-3D AMV in 3DVAR and HYBRID DA systems is evaluated by conducting four different observation system experiments, namely 3DVAR, 3DVAR-AMV, HYBRID, and HYBRID-AMV. While 3DVAR-AMVexperiment assimilates INSAT-3D AMV along with GTS observations, the 3DVAR run incorporates GTS observations alone using the 3DVAR DA assimilation approach. Similarly, HYBRID-AMV and HYBRID experiments are performed with and without INSAT-3D AMV observations in the HYBRID DA system, respectively. Assimilation is performed only in the parent domain with 27 km resolution, and the free forecasts

are achieved until the TC made landfall. It is to mention here that, in the HYBRID experiment equal weight has been assigned to the static and flow dependent error covariances.

**Table 6.1:** Details of tropical cyclones used in this study. SCS stands for Severe Cyclonic Storm. Other details are same as Table 4.2

| Serial No | Name | Duration of 12 hourly data assimilation cycles starting from the 00 UTC of the given date to the next 48 hour | Maximum sustained surface wind speed (kt) | TC category |
|---|---|---|---|---|
| **1** | Hudhud | 8th Oct 2014 | 100 | ESCS |
| **2** | Vardah | 8th Dec 2016 | 70 | VSCS |
| **3** | Titli | 8th Oct 2018 | 80 | VSCS |
| **4** | Bulbul | 6th Nov 2019 | 75 | VSCS |
| **5** | Amphan | 17th May 2020 | 135 | SCS |

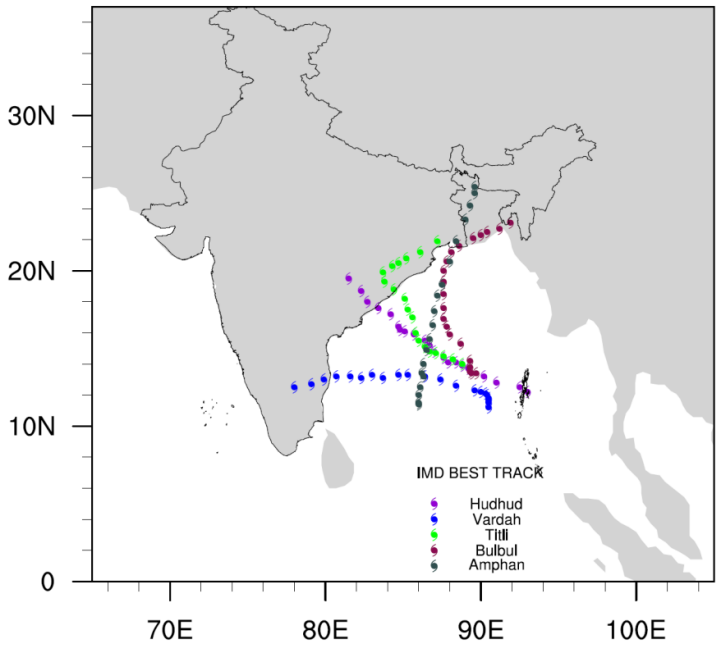


**Figure 6.2:** IMD observed based best track of the five tropical cyclones

## 6.5 Result and Discussion

The evaluations of the impact of the INSAT-3D AMV on HYBRID and 3DVAR DA system are achieved by comparing the track, intensity and time of landfall of TCs formed over BoB. Bootstrap resampling method is employed to evaluate the statistical significance of the result. Model simulated rainfall forecast during near landfall time is compared with Global Precipitation Measurement mission (GPM) data and equitable threat score is used for the validation of precipitation.

### 6.5.1 Analysis track and intensity

Figure 6.3 shows the absolute error in analyzed track position, maximum sustained wind error, and minimum sea level pressure error of the TC. The results show that the errors in the HYBRID experiment are lower than rest of the experiments both in terms of initial track and intensity. The intensity in respect of MSWS is improved significantly in HYBRID experiments for three cyclones, namely Hudhud, Vardah, and Amphan. The average errors from the analysis of 3DVAR-AMV, HYBRID, HYBRID-AMV shows relative improvement in the initial track by 4%, 18.85% and 19.19%, respectively, while that for intensity in terms of MSLP (MSWS) is 8.69% (5.77%), 30.35%(23.40%) and 31.60% (16.98%)respectively in comparison with 3DVAR experiment. The 3DVAR-AMV shows nominal improvements over 3DVAR, whereas the impact of AMV observations is not noteworthy in HYBRID DA experiments.

The comparative improvements found in HYBRID analysis than 3DVAR are further evaluated by assessing the vertical structure of the TCs as shown in Figure 6.4. Here, the composite of vertical cross-section represents the strength of tropical cyclones in respect of wind speed

(shaded) overlaid with potential temperature. All the experiments could simulate the core of the TCs, whereas 3DVAR experiments represent a vigorous cyclonic core compared to HYBRID with strong wind speed on both sides of the cyclonic center. AMV DA shows no significant impact on 3DVAR. On the contrary, the impact of AMV DA is noticeable in HYBRID, where the simulated strength of the cyclones is comparatively weak than in other experiments, mainly above 500 hPa. The overlying contours of potential temperature depict a prominent downward curving at the center of the TCs for all the experiments representing a warm core. However, HYBRID-AMV analysis shows a shallow bending compared to the rest of the experiments, mostly above 300 hPa. The overall results point toward that the analysis of cyclonic core simulated by 3DVAR is much stronger than the HYBRID simulations. At the same time, among the HYBRID experiments, HYBRID-AMV analysis depicts a weak cyclonic core.

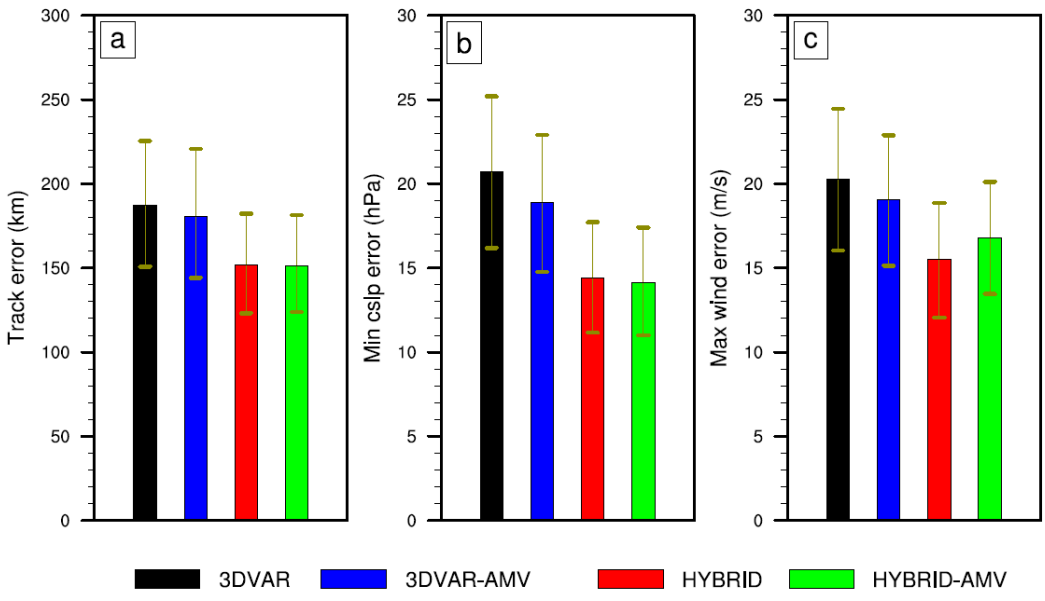


**Figure 6.3:** The error in analysis averaged over all the cyclones and DA cycles for the (a) Initial position (km) (b) Minimum Sea level pressure (hPa) (c) Maximum wind speed (m/s). The error bars specify 5$^{th}$ and 95$^{th}$ percentiles obtained from bootstrap resampling

### 6.5.2 Forecast track and intensity

Figure 6.5 represents the evolution of track forecast error with time for each TC considered in this study. The HYBRID experiment shows improvements in track forecast when compared to the 3DVAR experiment, and assimilation of INSAT-3D AMV observation depicts a positive impact on both the DA systems. However, the influence of INSAT-3D AMV observations is more substantial in 3DVAR than in HYBRID experiments. Previous studies by Kelly et al. (2007) have documented that the advanced DA systems with flow-dependent BEC are more efficient in propagating information from data rich to data void region. The apparent larger impact of INSAT-3D AMV in 3DVAR could be because of large data void regions in the absence of INSAT-3D AMV observations compared to the HYBRID DA system.

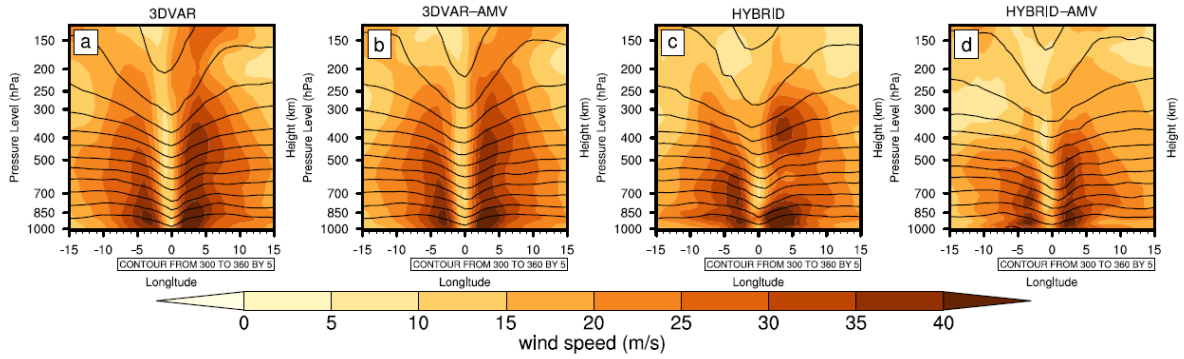


**Figure 6.4:** Composite of vertical cross-section covering all latitudes in a $15^0 \times 15^0$ boxin terms of potential temperature (˚K; contour) and wind speed (shaded). The vertical cross-section is taken at the center of each tropical cyclone

Furthermore, the time evolution of absolute error in model simulated MSWS at the 10-m level for individual cyclones is shown in Figure 6.6. Although the results do not reveal any significant

improvements in the intensity forecast by assimilating the INSAT-3D AMV observations, the HYBRID-AMV experiment indicates improvements in the intensity prediction near landfall.

### 6.5.3 Landfall

Figure 6.7 demonstrates landfall position and intensity forecast error averaged over all TC cases. The average landfall position error is 260, 180, 160, and 93 km for 3DVAR, 3DVAR-AMV, HYBRID, and HYBRID-AMV. The results clearly establish that incorporating INSAT3D AMV observations in the HYBRID DA system is more effective than 3DVAR. Landfall position is one of the critical predicted factor s from the disaster mitigation point of view, and the assimilation of INSAT-AMV observation is seen to be critical for DA systems. The average MSWS error during landfall measured for all cyclones shows better HYBRID-AMV performance than other experiments. Further, the smaller error bars in HYBRID-AMV testify to the statistical significance of the end result (Figure 6.7c).

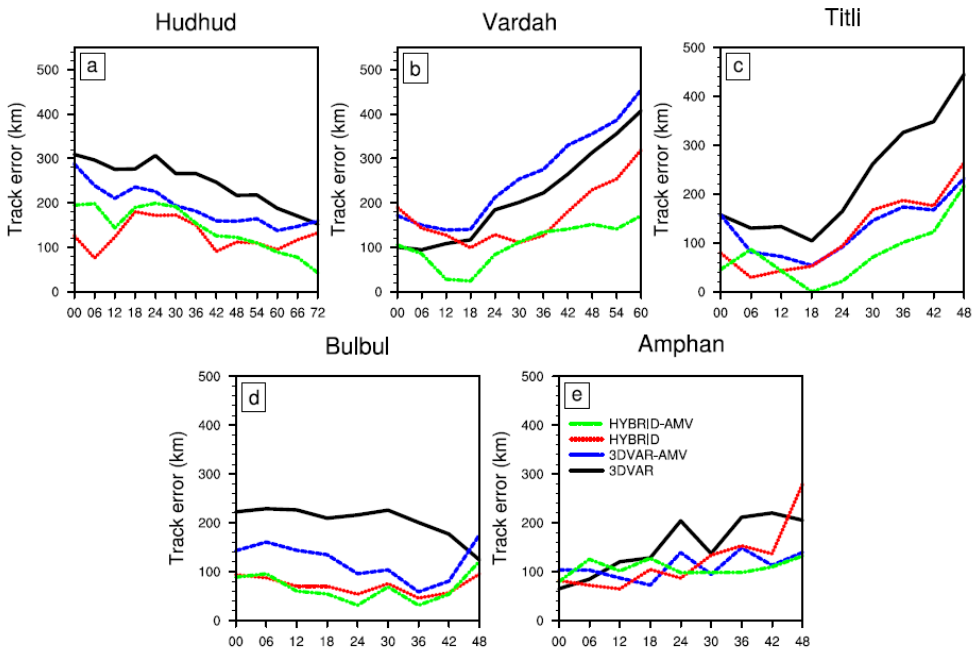


**Figure 6.5:** Track Forecast error for individual TCs from each experiments. The different colors represent different experiments: Green line (HYBRID-AMV), Red line (HYBRID), Blue line (3DVAR-AMV), Black line (3DVAR)

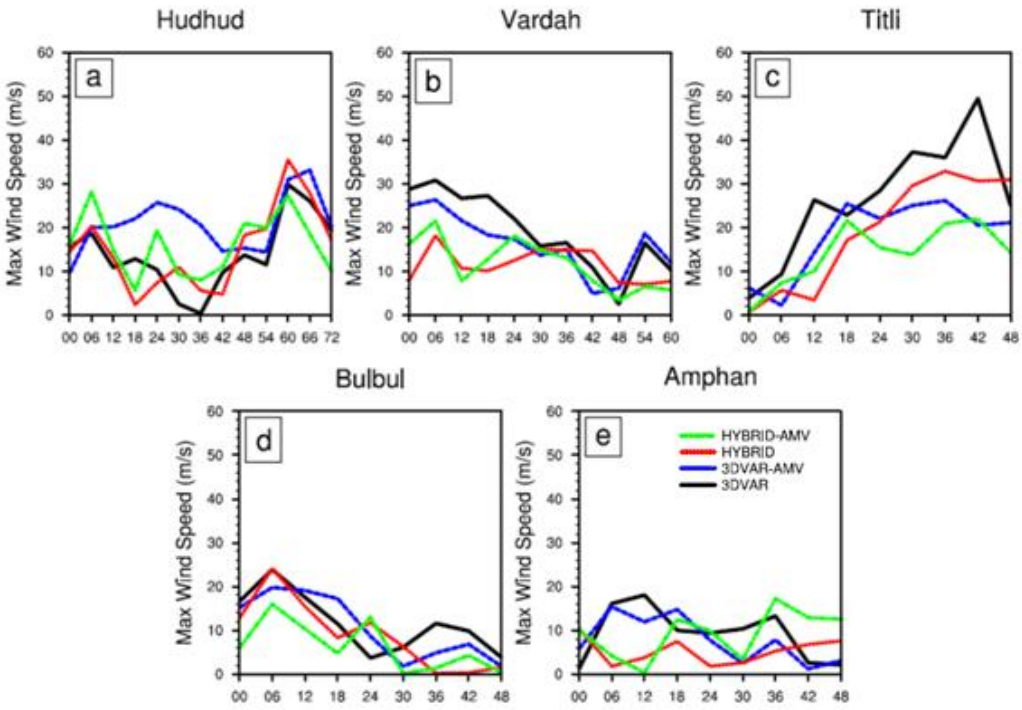


**Figure 6.6:** Forecast of intensity error in respect of maximum sustained wind speed (MSWS) for individual TCs from different experiments.

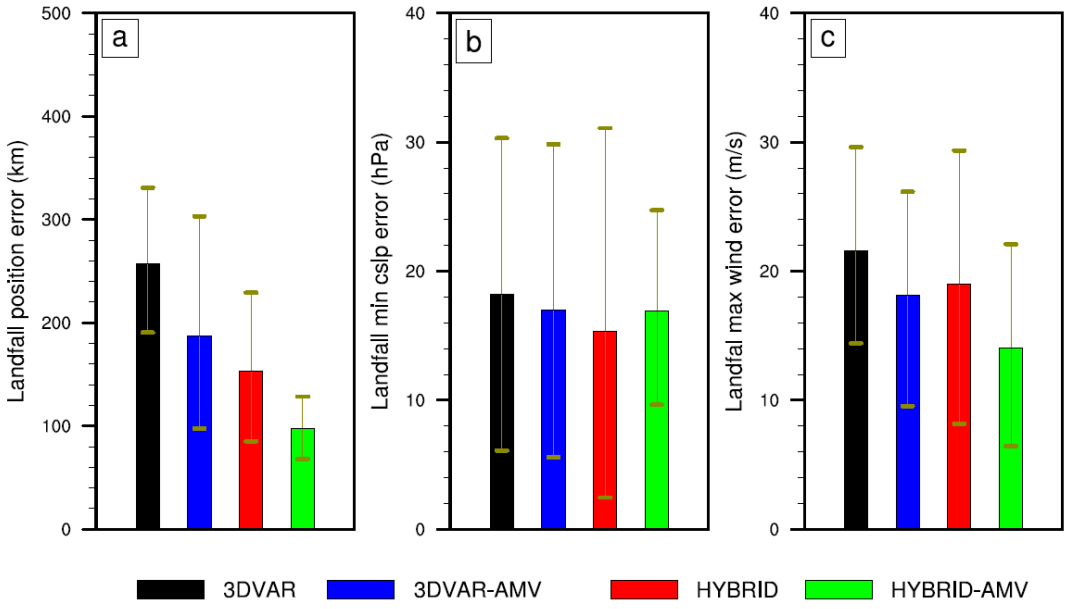


**Figure 6.7:** Same as Figure 6.4, but for averaged forecast error during landfall.

### 6.5.4 Rainfall Forecast

Figure 6.8 shows comparison of the 24 hour accumulated rainfall distribution overlaid by MSLP contours during the landfall of the TCs along with the GPM observations. All the simulations have represented the rainfall distribution around the center of TC fairly well. However, the inland rainfall distribution is vastly underestimated by the model simulations, which is more evident for TCs such as Titli and Amphan. For TC Amphan, intensity of rainfall around the TC core is underpredicted in all the experiments. The skill scores for the rainfall specify that HYBRID-AMV has a superior skill for rainfall forecast than other experiments, in general (Figure 6.9 ). The ETS values of 3DVAR-AMV and HYBRID-AMV simulations are substantially higher for Hudhud, and Vardah compared to its corresponding control experiments.

## 6.5 Summary

In this study, the performance of 3DVAR and HYBRID DA techniques with INSAT-3D Atmospheric Motion Vectors are compared for the forecasts of landfalling TCs originated over the BoB using the WRF model. The track and intensity of TCs are evaluated by comparing with IMD best track data and the quantitative precipitation forecasts at the time of landfall are validated with respect to GPM rainfall data. The results confirm that the initial position and intensity error of TCs are lower in HYBRID compared to 3DVAR analysis. However, assimilation of INSAT-3D AMV observations has revealed only slight improvements in the initial position and intensity using both the DA systems.On the contrary, the assimilation of AMV observations has significantly improved the HYBRID DA system's track forecast compared to 3DVAR, in general. The average position error at the moment of landfall is 260, 180, 160, and 93 km for 3DVAR, 3DVAR-AMV, HYBRID, and HYBRID-AMV runs,

respectively. The relative improvements in intensity compared to 3DVAR for TC are 18%, 13% and 36% respectively, for 3DVAR-AMV, HYBRID, and HYBRID-AMV.

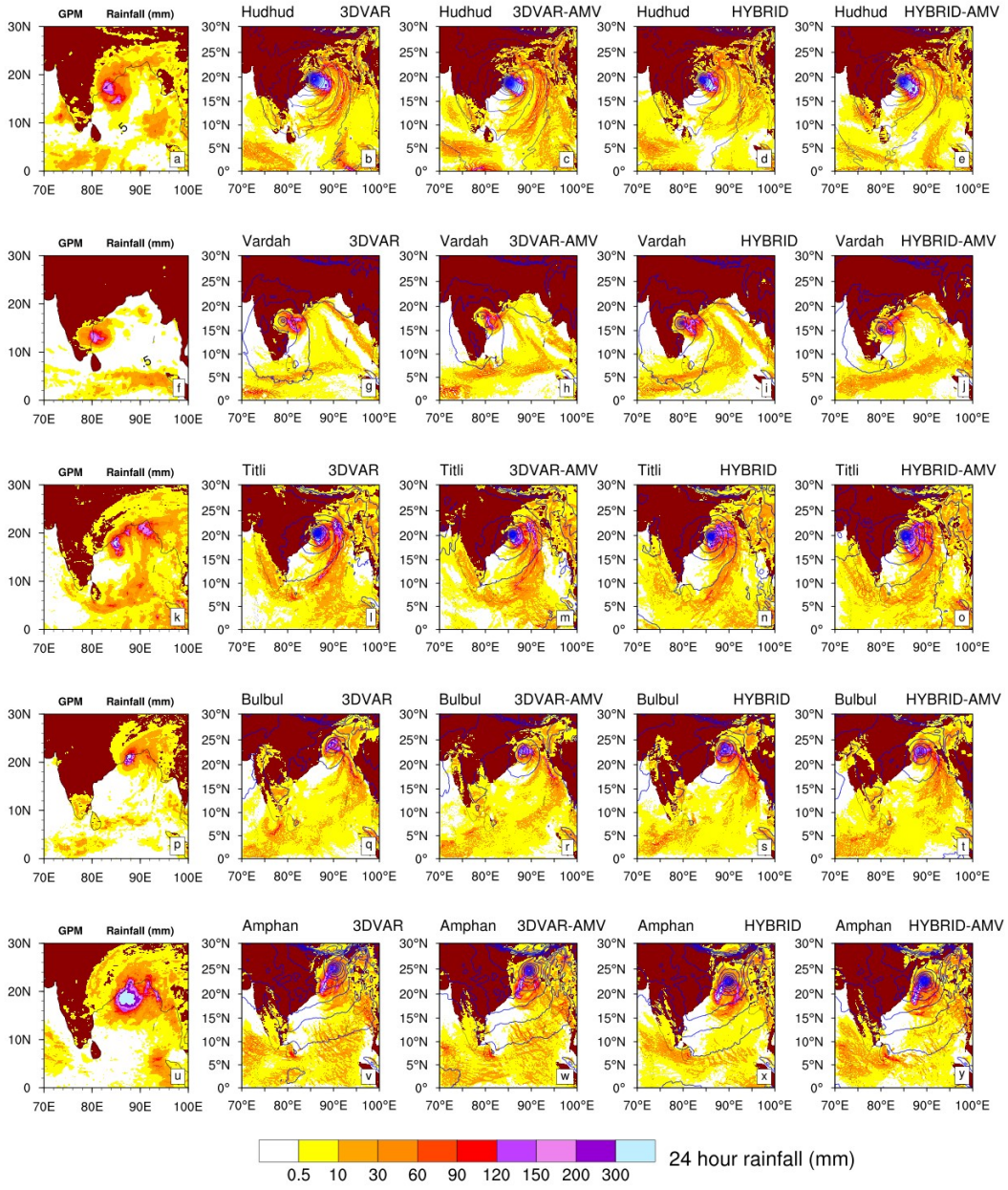


**Figure6.8:**Observed spatial distribution of 24 hour rainfall during landfall of TC (a) Hudhud, (f) Vardah, (k) Titli, (p) Bulbul, (u) Amphan and comparison with rainfall forecast from (b,g,l,q,v) 3DVAR, (c,h,m,r,w) 3DVAR-AMV, (d,i,n,s,x) HYBRID and (e,j,o,t,y) HYBRID-AMV experiments for the respective TCs. Blue contours represent model simulated sea level pressure (SLP)

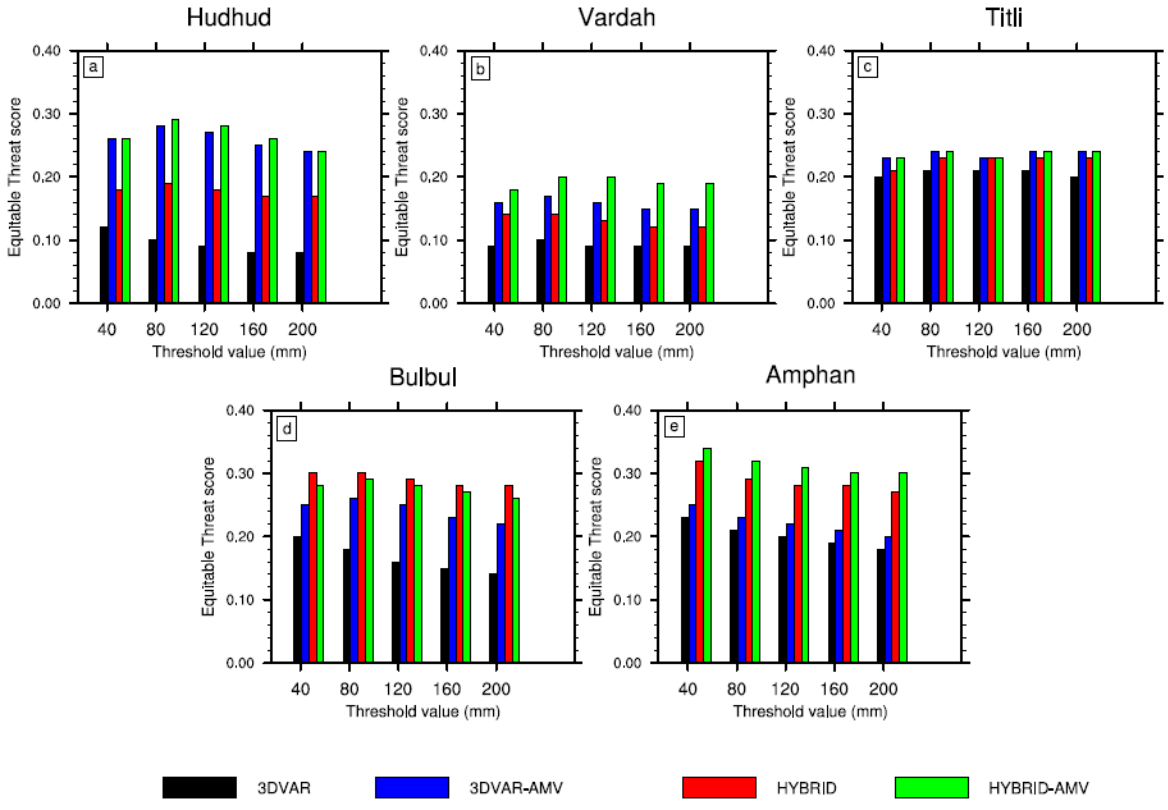


**Figure 6.9:** The equitable threat scores for 24 hours total rainfall forecasts at the time of landfall of TCs (a) Hudhud (b) Vardah (c) Titli (d) Bulbul (e) Amphan

Assimilation of INSAT-3D AMV observations has improved the forecast of TC landfall locations in all the experiments with more improvement using the HYBRID DA system. Rainfall forecast is significantly improved in both the data assimilation system by incorporating the INSAT-3D AMV data. In addition, the HYBRID-AMV experiment show improved skill scores for precipitation over all the other experiments, in general. Overall, assimilation of INSAT-3D AMV observations in the HYBRID DA system reduces the relative errors in landfall position significantly and with minor improvement in the intensity.

# Chapter 7

## Impact of flow-dependent error covariance in 3DVAR DA System in convection permitting resolution: Indirect assimilation of Doppler Weather Radar Reflectivity in the forecasts of Thunderstorms over North-eastern Region of India

### 7.1 Introduction

Numerical Weather Prediction models are used globally for early warning of catastrophic weather events such as Flood, Thunderstorm, Lightning, Tropical Cyclone etc. Though we have made remarkable improvements in the forecast of weather systems during the last decade, the prediction of severe convective systems such as thunderstorms, extreme localized rainfall, and lightning are yet very challenging. One potential reason for frequent failure of localized convection forecast is the lack high resolution of observation over a region that would represent the small scale features of a convective system. The observations from Doppler Weather Radar (DWR) play a significant role in the forecast of convective processes due to its high spatiotemporal resolution. In the last decade, several studies successfully assimilated high resolution DWR data in NWP models (He et al. 2020; Lai et al. 2020; Li et al. 2018; Li et al. 2013; Prasad et al. 2014; Routray et al. 2013; Srivastava et al. 2011). Reflectivity and radial velocity are two variables from DWR that are being ingested to NWP model. These parameters have substantial importance as reflectivity provides information on the hydrometeors whereas radial velocity holds information on convective motions of atmosphere.

Although DWR is an essential data source for NWP models, radar reflectivity gives only information about precipitation but no direct information about water vapor, temperature, and other fields. Therefore, in a comparatively dry environment where no prior instability

information is available, the direct assimilation of radar reflectance cannot represent convection initiation. Therefore, Wang et al. (2013a) introduced a new indirect method of assimilation of radar data that uses derived moisture products from reflectivity instead of direct reflectivity.

This study investigates the impact of assimilation of water vapor and rainwater derived from radar reflectivity in 3DVAR and HYBRID DA system for simulation of thunderstorms over the north-eastern region (NER) of India

### 7.2 Model Description and Configurations

The WRF model is configured with a single domain at 3 km horizontal grid spacing over the NER of India (Prasad et al. 2014) with 360 x 360 horizontal grids and 36 vertical levels. The domain which includes the name of the north-eastern states frequently affected by the thunderstorms, is shown in Figure 7.1. The initial and boundary conditions are interpolated from the NCEP GFS analysis and forecast at $0.25^0 \times 0.25^0$ gridded resolution. The various parameterization schemes used for the simulations are the same as those mentioned in Chapter 3.

### 7.3 Data used for Assimilation

The DWR located near Cherrapunji, Meghalaya is S-band radar that provides data through a volume scans. The radar completes one volume scan in 11 minutes, comprising of 360 degree azimuth scan for 10 elevation angles ranging from 0.5 to 21 degrees. The DWR covers a distance of 250 km (up to 500 km only for Z) with spatial resolution of 300 m. The initial quality check is done by the software developed at Indian Space Research Organization, India. To preprocess the data further, a python module has been developed. Since the resolution of the DWR data is very high as compared to the model domain, the quality checked radar data in azimuth-range

format is transformed to the Cartesian grid with the map projection same as the model using the python module. Using the same python module data has been quality checked to discard data beyond the limit 10–55dBz. Additionally, FORTRAN codes are developed to convert the processed data to WRFDA compatible format for data assimilation. Along with the observations from DWR, conventional observations are also used for assimilation in this study.

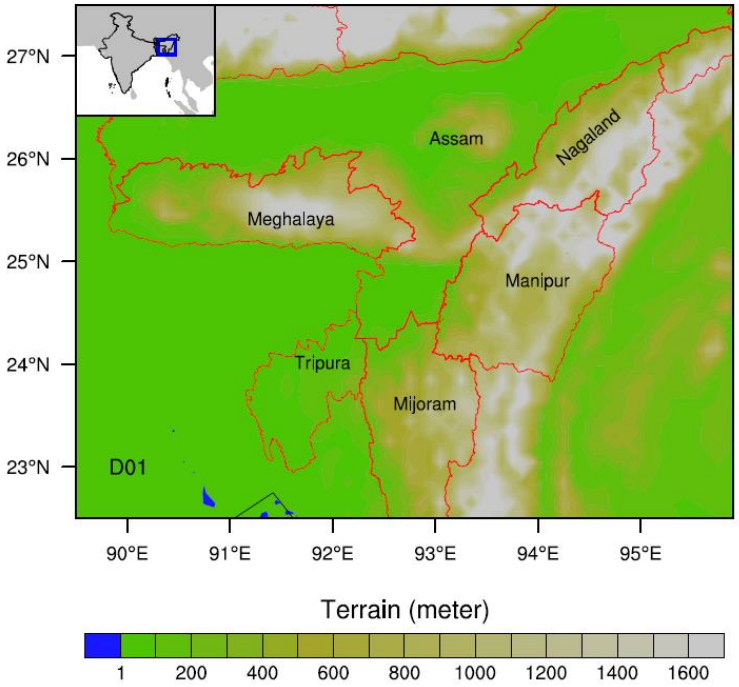


**Figure 7.1:** Model domain (D01) used in this study. Red lines represent the state boundaries and black fonts represent the state names of north-eastern region of India that are affected during the thunderstorm events. Background is the terrain height in meter.

### 7.3.1 Assimilation of retrieved rainwater and water vapor from radar reflectivity

To assimilate retrieved rainwater and water vapor, two additional terms are added into Eq. (2.16) following Wang et al. (2013):

$$J_{qr} = \frac{1}{2}\left(\mathrm{q_r} - \mathrm{q_r^b}\right)^{\mathrm{T}} \mathrm{B_{qr}^{-1}}\left(\mathrm{q_r} - \mathrm{q_r^b}\right) + \frac{1}{2}(\mathrm{q_r} - \mathrm{q_r^o})^{\mathrm{T}} \mathrm{R_{qr}^{-1}}(\mathrm{q_r} - \mathrm{q_r^o}) \tag{7.1}$$

and

$$J_{qv} = \frac{1}{2}(\mathrm{q_v} - \mathrm{q_v^o})^{\mathrm{T}} \mathrm{R_{qv}^{-1}}(\mathrm{q_v} - \mathrm{q_v^o}) \tag{7.2}$$

where $\mathrm{q_r}$ , $\mathrm{q_v}$ stands for rainwater and water vapor of the atmospheric state ; $\mathrm{q_r^o}$, $\mathrm{q_v^o}$ stands for rainwater and water vapor from radar reflectivity; $\mathrm{R_{qr}}$ and $\mathrm{R_{qv}}$ are the observation error variance of rainwater and water vapor; $\mathrm{q_r^b}$ is the background rainwater and $\mathrm{B_{qr}}$ is the background error matrix.

Given the radar reflectivity, rain water can be derived from the following equation (7.3):

$$Z = c_1 + c_2 log_{10}(\rho q_r) \tag{7.3}$$

where $Z$ is the reflectivity (dbZ), $\rho$ air density (kg $\mathrm{m^{-3}}$) and $q_r$ is the rain water mixing ratio (g $\mathrm{kg^{-1}}$); $c_1$ and $c_2$ are constants with values 43.1 and 17.5, respectively, following Sun and Crook (1997).

For water vapor retrieval, it is assumed that when the $Z$ value exceeds a specified threshold value above cloud base, the in-cloud relative humidity is 100%. Here, the threshold value is fixed as 30 dBz. It is resulted from the assumption that the estimated water vapor $q_v^o$ is equal to the saturation water vapor that is calculated using pressure and temperature of the background.

### 7.4 Thunderstorm Cases

Two pre-monsoon thunderstorm (TS) events within the vicinity of NER are considered for this study.

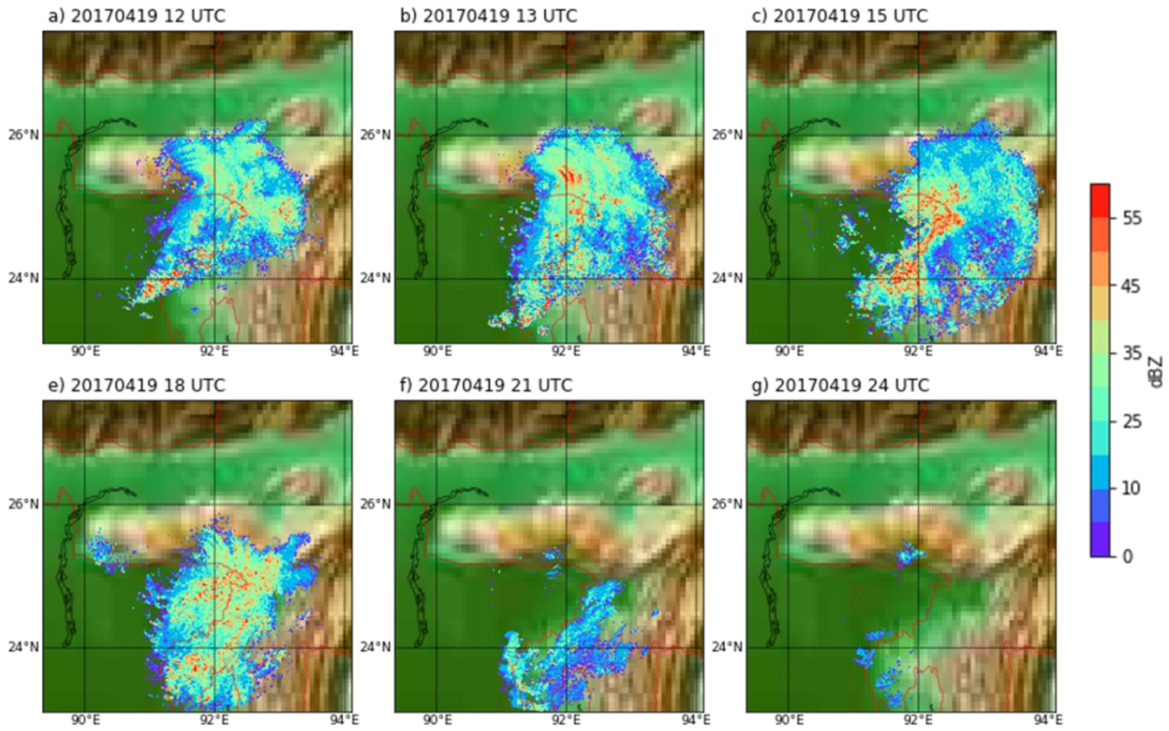


**Figure 7.2:** Maximum reflectivity (dBz) observed in Cherrapunji Radar data during a thunderstorm event on April 19, 2017 valid from 12 UTC to 24 UTC. This event is considered as Case-1 in this study

**Case-1**: The first TS event considered here is a synoptic scale system, which is initially developed over the Tibetan plateau that shifted towards NER of Indian subcontinent. The TS is initiated on April 19, 2017 at 12 UTC over Meghalaya and Assam. The system remained there till 13 UTC, and then moved towards southern Assam, Tripura and Mizoram. It has developed

into a more intense thunderstorm during 15 UTC over the same region producing heavy rainfall and the system started weakened by 21 UTC of April 19, 2017 as shown in Figure 7.2.

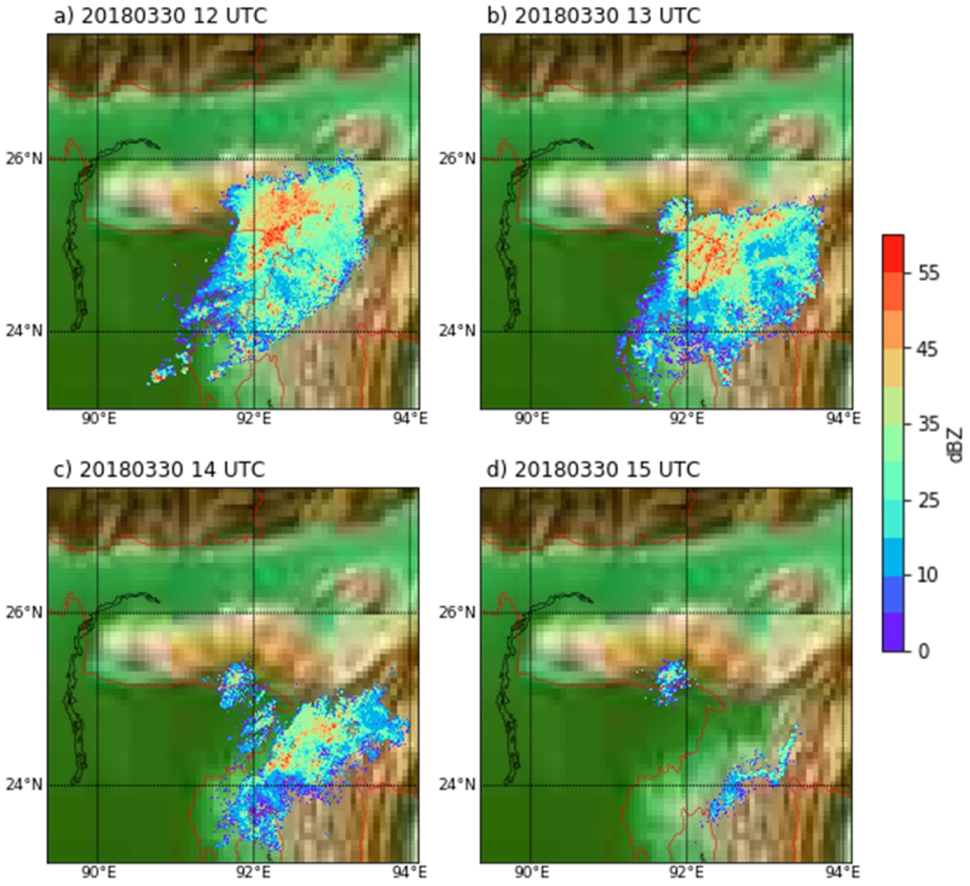


**Figure 7.3:** Maximum reflectivity (dBz) observed in Cherrapunji Radar data during a thunderstorm event on March 30, 2018 valid from 12 UTC to 15 UTC. This event is considered as Case-2 in this study

**Case-2**: The second TS event considered here is initiated on March 30, 2018 at 12 UTC and dissipated by 15 UTC of the same day. The TS developed locally over the Meghalaya region and gradually moved towards southern Assam, which finally dissipated over Mizoram at 15 UTC. The storm has affected Meghalaya, Southern Assam, Manipur, and some parts of Mizoram that is shown in the Figure 7.3.

### 7.5 Experimental design and validation

In order to analyze the impact of radar data assimilation in convective scale, four experiments are designed. Two of these experiments assimilated conventional observations and satellite AMV using 3DVAR and HYBRID DA systems while the other two experiments assimilated DWR observations along with the GTS observations using 3DVAR (3DVAR-RQ) and HYBRID (HYBRID-RQ) DA systems. In each case, model ensembles are integrated for 6 hour before the first data assimilation cycle. The second assimilation cycle is performed 12 h after the first assimilation and free forecast is performed thereafter. The ensemble members are generated using WRF-3DVAR CV3 BEC by adding 30 random perturbations (Barker et al. 2004). The current study employed NMC method to estimate the region-specific static BEC from WRF 12 and 24hr forecast differences averaged over the month of April using CV5 option.

For validation, model simulated rainfall forecast is compared with GPM rainfall measurement. Statistical evaluation of rainfall forecast is conducted by calculating ETS and BIAS score. Prior to comparison, all the data are brought to a common grid resolution using bilinear interpolation.

### 7.6 Result and Discussion

In this study, the impact of derived moisture field from radar reflectivity in the simulation of thunderstorms using 3DVAR and HYBRID is evaluated for both model analysis and forecast

variables. The development of a thunderstorm system depends on various convective parameters such as Convective Available Potential Energy (CAPE), Convective Inhibition Energy (CINE), mid-level humidity and wind shear. A study by Westermayer et al. (2017) has demonstrated the significant role of mid-level relative humidity in the formation of thunderstorm. This study assesses the influence of radar reflectivity in the model simulated moisture analysis fields namely water vapor mixing ratio (Qv) and relative humidity (Rh).

### 7.6.1 Analysis field

Figure 7.4 shows a longitudinal vertical cross-section of Qv analysis field for both the TS cases. Since there are no station data available to ensure the formations of thunderstorms over this region, the observed radar reflectivity is used as the indicator for the TS. A value of reflectivity more than 20 dBz is considered as the indication of the formation of TS. For Case-1, the model simulated Qv represents a wet environment mostly from the surface to 750 hPa level for all the experiments (Figure 7.4a-d). No significant impact of radar DA is observed for both 3DVAR and HYBRID DA systems. However, for Case-2, the impact of radar DA is evidently seen. It is observed from Figure 7.4e to 7.4g that in 3DVAR and HYBRID experiments that assimilated conventional observations, a dry atmospheric condition in vertical with low Qv values exists that are not favorable for the initiation of convection. In contrast, the assimilation of DWR observations enhances the moisture conditions from surface to 700 hPa level for 3DVAR and HYBRID experiments, as shown in Figure 7.4f to 7.4h. Additionally, the results show no significant variations among the two DA systems used in this study.

Similarly, Figure 7.5 shows the longitudinal vertical cross-section of the RH analysis field for both cases. For Case-1, the HYBRID experiment depicts enhanced mid-level moisture around

600 hPa compared to the 3DVAR experiment (Figure 7.5a-d). Therefore, it can be stated that the HYBRID experiment simulates a more conducive environment for TS than the 3DVAR experiment for the synoptic TS case. However, assimilation of DWR observations has not significantly impacted the analysis of 3DVAR or HYBRID.

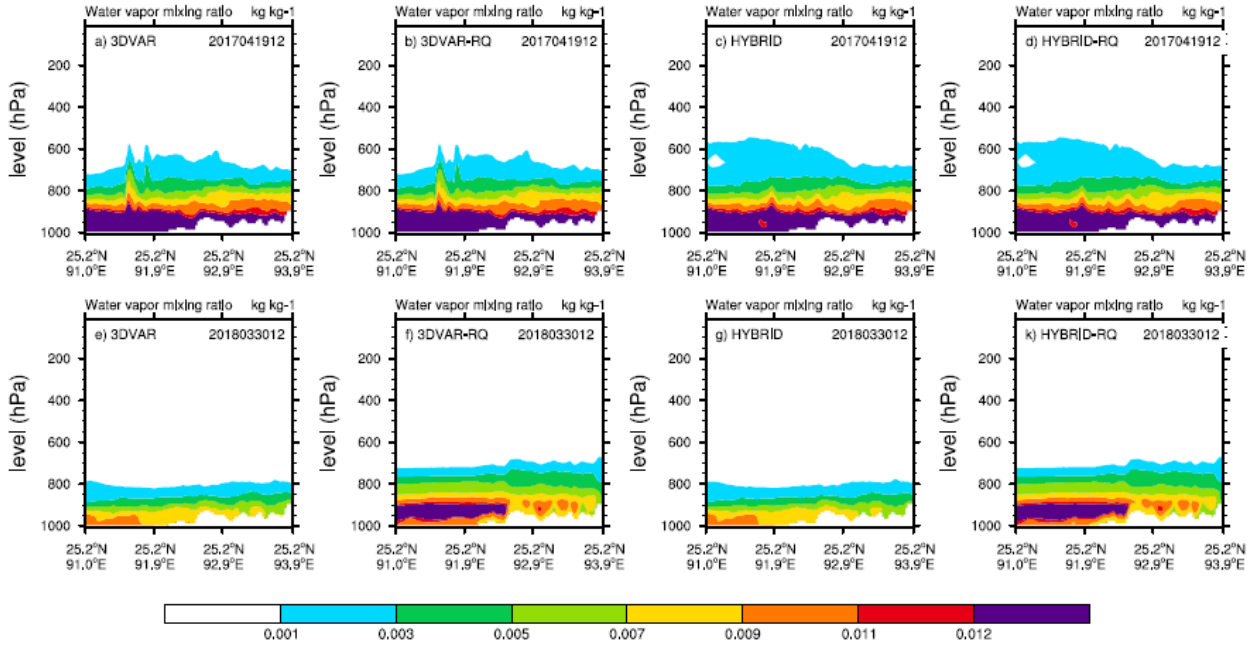


**Figure 7.4:** Longitudinal vertical cross-section of water vapor mixing ratio analysis field passes through the thunderstorm core for both Case-1 (12 UTC 19 April 2017) and Case-2 (12 UTC 30 March 2018) from a,e) 3DVAR, b,f) 3DVAR-RQ, c,g) HYBRID and d,h) HYBRID-RQ experiments

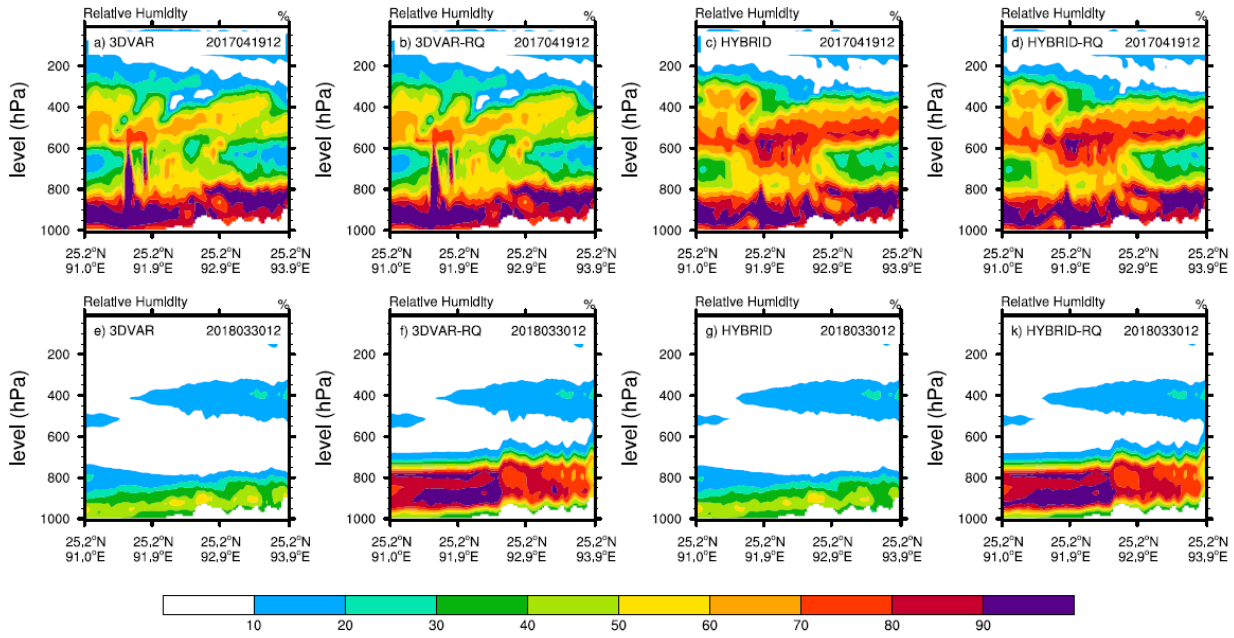


**Figure 7.5:** Same as Figure 7.4, but for relative humidity

For Case-2, assimilation of GTS observations alone in 3DVAR and HYBRID represents a relatively dry environment with less than 50 % of RH over the lower levels of atmosphere with dry conditions prevailing over the mid-level troposphere. But, the assimilation of DWR observations has enhanced the moisture over the lower levels of the atmosphere.

### 7.6.2 Rainfall forecast

Figure 7.6a and 7.6f show the accumulated rainfall from GPM data valid during the forecast period from 12 UTC April 19 to00 UTC April 20 of the year 2017 for Case-1 and 12 UTC March 30 to 15 UTC March 30 of the year 2018 for Case-2 respectively. For the synoptic TS Case-1, all the model simulations represent the spatial distribution of rainfall well (Figure 7.6b-e). However, it is observed from the GPM rainfall data that there exists a rainfall patch of intensity ranging from 20 mm to 100 mm towards the southern Assam bordering Tripura. The same is not

captured in any of the 3DVAR experiments. The missing rainfall patch in 3DVAR is noticeable in HYBRID experiments to some extent, HYBRID-RQ in particular. The improvement in rainfall forecast from HYBRID experiments for Case-1 may be attributed to the more conductive environment simulated by HYBRID for TS initiation, as observed in Figures 7.5c and 7.5d. For Case-2, a significant impact of radar DA is observed for both the DA system (Figure 7.6h,j). The assimilation of only GTS data into 3DVAR and HYBRID failed to capture the rains spread over southern Assam, Tripura, Manipur and Mizoram completely (Figure 7.6i,k). However, improvements in 3DVAR-RQ and HYBRID-RQ are seen in almost all areas except southern Tripura. This may be because the improved moisture analysis field with radar data assimilation improves rainfall forecast in Case-2.

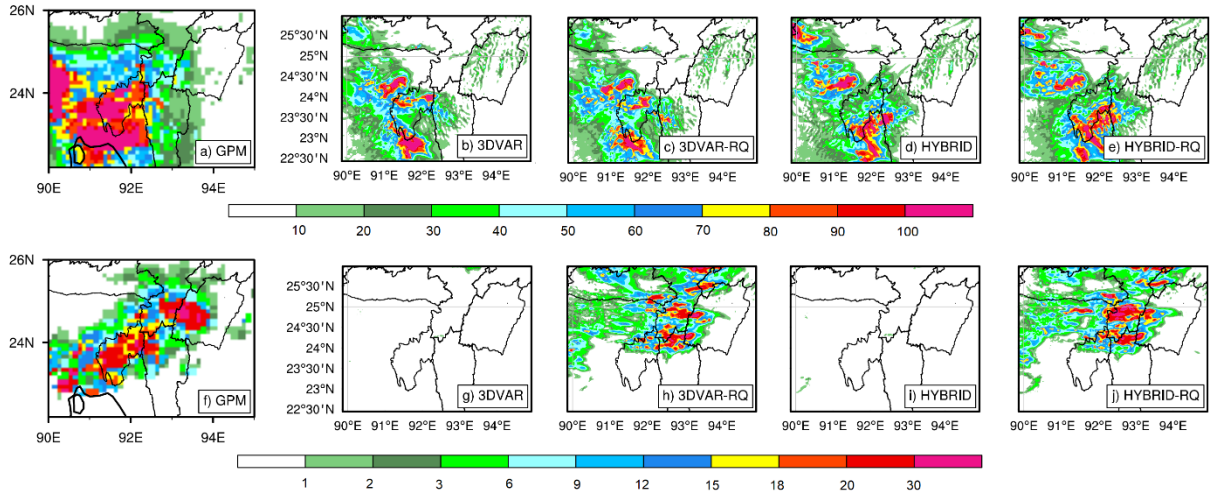


**Figure 7.6:** a) Accumulated rainfall calculated from GPM data; Model simulated rainfall forecast from b) 3DVAR, c)3DVAR-RQ, d) HYBRID and e) HYBRID-RQ valid from 20170419 12 UTC to 20170420 00 UTC. f) Accumulated rainfall calculated from GPM data; Model simulated rainfall forecast from b) 3DVAR, c)3DVAR-RQ, d) HYBRID and e) HYBRID-RQ valid from 20180330 12 UTC to 20180330 15 UTC

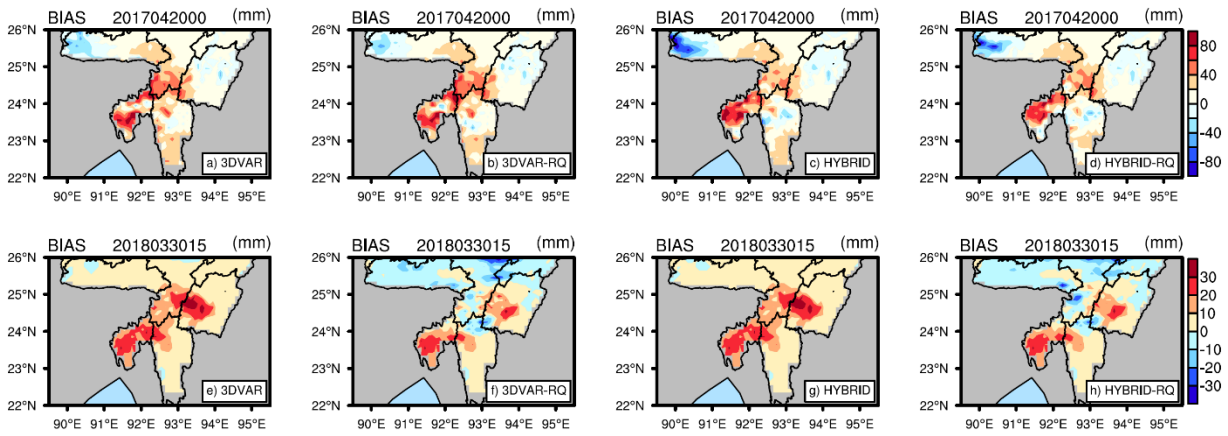


**Figure 7.7:** Bias of model simulated accumulated rainfall forecast valid from 20170419 12 UTC to 20170420 00 UTC with respect to GPM observed rainfall for a) 3DVAR, b) 3DVAR-RQ, c) HYBRID, d) HYBRID-RQ ; Bias of model simulated accumulated rainfall forecast valid from 20180330 12 UTC to 20180330 00 UTC with respect to GPM observed rainfall for e)3DVAR, f) 3DVAR-RQ, g)HYBRID, h) HYBRID-RQ

Figure 7.7 shows the bias of model-simulated accumulated rainfall forecast for both cases with respect to GPM observed rainfall for 3DVAR and HYBRID experiments. For Case-1, a dry bias is observed in 3DVAR experiments over the southern Assam and Mizoram, which can be seen, reduced in HYBRID experiments. For Case-2, a prominent dry bias is observed in 3DVAR and HYBRID experiments without DWR assimilation. The observed bias has reduced in 3DVAR-RQ and HYBRID-RQ over southern Assam and Meghalaya. The assimilation of DWR observations in 3DVAR and HYBRID DA systems has resulted in a better simulation of the thunderstorm for the localized TS event.

Further, the validation of model-based rainfall forecast is evaluated quantitatively by estimating model skill scores. The skill scores for Case-1 are calculated for five threshold values 20 mm, 30 mm, 40 mm, 70 mm, and 80 mm. The ETS values for experiments with DWR observations

are found to be nominally higher than the experiments without DWR observations. However, the performance of 3DVAR and HYBRID DA systems with and without DWR observations does not show significant differences in the rainfall skill score (Figure 7.8a). Bias score in HYBRID experiments shows improvement as compared to the 3DVAR experiment (Figure 7.8b). For Case-2, skill scores are calculated for the threshold values 2 mm, 5mm, 10 mm, 20 mm, and 30 mm. The skill scores are found to be close to zero for 3DVAR and HYBRID experiments without DWR assimilation, as the experiments fail to simulate rainfall in the absence of DWR observations. (Figure 7.9a). Though the ETS values are higher for the HYBRID-RQ experiment, the Bias score values show no significant change among both the experiments (Figure 7.9b).

### 7.7 Summary

This study evaluates the impact of the derived moisture field from radar reflectivity in 3DVAR and flow-dependent HYBRID DA system for the thunderstorm simulation over the NER of India. Two thunderstorm cases are considered: a locally developed short-duration TS and a synoptic-scale long-duration TS system.

The vertical structure of the analysis fields shows that the impact of DWR observations is prominent for a locally developed TS event when initial model state variables depict a dry environment compared to the TS system developed due to synoptic forcing for which the background moisture fields are already wet. The assimilation of DWR observations has not shown any significant impact on the synoptically forced TS. However, improvement due to the

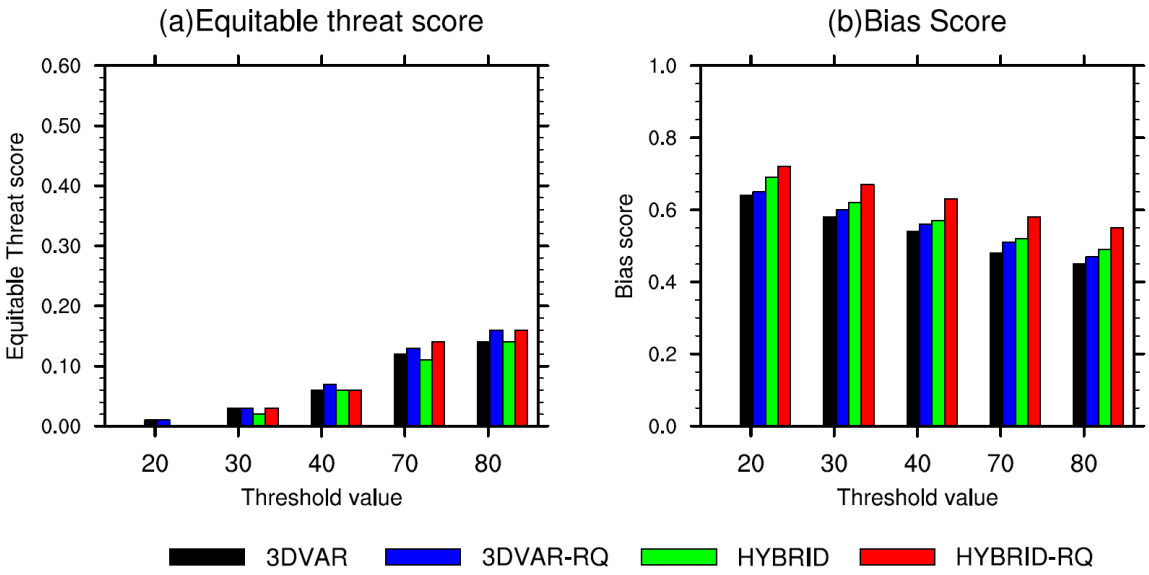


**Figure 7.8:** The (a) ETS and (b) Bias scores valid from 12 UTC 19 April 2017 to 00 UTC 20 April 2017 for different rainfall thresholds calculated for all the experiments.

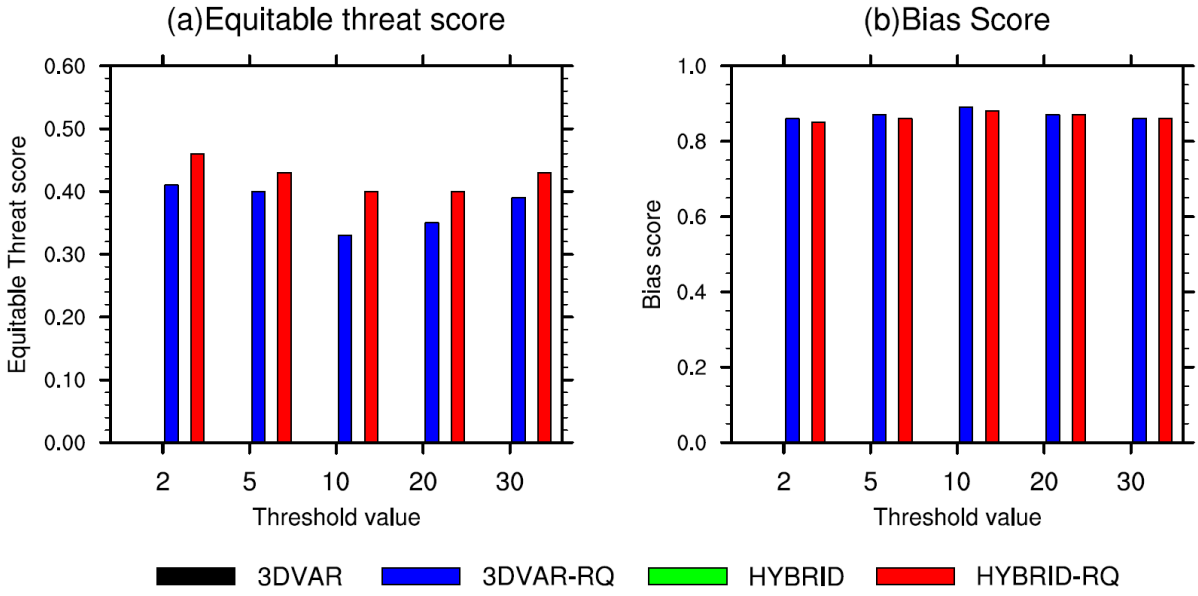


**Figure 7.9:** The (a) ETS and (b) Bias scores valid from 12 UTC 30 March 2018 to 15 UTC 30 March 2018 for different rainfall thresholds calculated for all the experiments. Here, skill scores of 3DVAR and HYBRID experiment is 0.

Incorporation of flow-dependent error covariance in the 3DVAR DA system is evident in simulating the moisture conditions for the development of TS in the analysis. Furthermore, the HYBRID experiments simulate a stronger TS system as compared to the 3DVAR experiment. The results clearly show a significant impact of radar data assimilation in the simulation of the localized convective system. However, no substantial impact of due to flow-evolving BEC is observed. This may be because this study has not utilized radar data to update the ensemble members using ETKF. Furthermore, the data assimilation is conducted every 12 hourly, which is too large for convective-scale simulations.

# Conclusions

The specification of background error covariance (BEC) has important role in the performance of data assimilation (DA) system. Studies have shown that assimilated observations are better and more effectively exploited when flow-dependent error covariance is used in the DA system. The three-dimensional variational (3DVAR) DA system is cost-effective because it uses a BEC matrix that is either completely static or only weakly coupled to the dynamics of the forecast model. However, the drawback of static BEC is that the observations assimilated will make local, isotropic increments without proper flow-dependent extrapolation. Several studies examined the role of static BEC in the performance of the 3DVAR DA system. The representation of flow-dependent ensemble BEC in a variational framework is found to be more effective than standalone Ensemble Kalman Filter (EnKF) DA approach especially when computational resources are limited. The approach which incorporates ensemble-generated BEC in variational framework is commonly referred to as a hybrid ensemble-variational DA system (HYBRID). The study compares the impact of flow-dependent ensemble BEC in 3DVAR DA system for the prediction of weather events in a flow regime that encompasses the Indian subcontinent.

In Chapter 3, the key research question that has been addressed is how does the flow-evolving BEC has affected the short-range rainfall forecasts during Indian summer monsoon at a convection-permitting resolution? The experiments have been carried out using the WRF model for month-long, short-range rainfall forecasts over the Indian summer monsoon region for July 2013. The results suggest that the use of flow-dependent ensemble BEC in 3DVAR has systematically improved the forecast. The root mean square error (RMSE) from HYBRID experiments is systematically smaller than 3DVAR experiments for temperature and wind field

in almost all the vertical levels. Further, HYBRID experiments have shown better skills in quantitative precipitation forecast (QPF) during Indian summer monsoon. On spatial scales, the 3DVAR experiment shows a dry bias over the upper peninsular regions and wet bias over the central and northern parts of the Indian subcontinent. The magnitude of the wet and dry bias is lower in HYBRID experiments as compared to the 3DVAR experiment. High-resolution experiments are conducted at the convection-permitting resolution, and the results are validated against dense TRG network observations over the Karnataka state. The result shows significant improvement in HYBRID runs during the later stages of data assimilation cycles compared to the 3DVAR experiment. The synoptic meteorological features of the monsoon in 3DVAR and HYBRID experiments are comparable to those in the ERA-interim analysis. However, the magnitude and direction of lower- and upper-level wind over the Indian monsoon region are better forecasted in the HYBRID than in the 3DVAR experiments. The information on wind is critical to understanding the moisture transport from the Indian Ocean, and hence accurate forecasting of wind may reduce the bias in the moisture over the Indian landmass. The improved forecast in wind variables is proposed as the reason for the reduction in dry bias observed over the Indian landmass in the HYBRID experiments. To summarize, the results of study provides dynamically consistent, objective improvements in the initial conditions of the WRF model when flow-dependent BEC is incorporated. However, nominal sensitivity of HYBRID experiment to the BEC weight in the analysis and forecast has been observed.

The key research question that is addressed in Chapter 4 is how effective the flow-dependent BEC in the 3DVAR DA framework in the forecast of tropical cyclones (TC) formed over the Bay of Bengal? To that end, eight landfalling TC of various intensities are considered. The results from the two state-of-the-art DA systems are compared and verified against the JTWC best track

data and TRMM precipitation observations. The HYBRID experiments analyzed the TC position and intensity significantly better compared to the 3DVAR DA system. The HYBIRD75 with 75% weight assigned to flow-dependent error covariance has marginal improvements over HYBRID50 with 50% weight assigned to flow-dependent error covariance in terms of reduction in analyzed TC position and intensity errors. The 3DVAR analysis depicts a stronger vortex and a prominent warm-core TC structure compared to HYBRID50 and HYBRID75 experiments. The TC structure tends to weaken in HYBRID75 as compared to HYBRID50.The errors in track forecast and landfall position of TC are smaller with HYBRID compared to the 3DVAR experiment. In contrast, the intensity forecast shows no substantial improvement with HYBRID compared to the 3DVAR DA system. The skill scores of quantitative precipitation forecast obtained for HYBRID50 and HYBRID75 experiments are higher than 3DVAR in most of the cyclone cases for all rainfall thresholds. The relative improvement in precipitation forecasts in HYBRID experiments compared to 3DVAR can be attributed to improved TC track prediction in the HYBRID DA system. Similar to the finding of Wang et al. (2013), BEC weightings found to have limited sensitivity in HYBRID analysis and forecast.

It is known that the impact of observations may vary depending on many factors in the DA assimilation system, such as data quality control, preprocessing, and specification of BEC. To understand the changes in the impact due to the specification of BEC, the Chapter 5 addresses the impact of Atmospheric Motion Vectors (AMV) in 3DVAR and HYBRID DA system. The impact of INSAT-3D AMV observation in the 3DVAR and the HYBRID DA system is evaluated for monsoon rainfall simulation. The HYBRID DA system incorporates flow-dependent ensemble BEC that generates optimal analysis through increments that are consistent with the background flow and respond adaptively to the change in the observing system. Therefore, it is

expected that the impact of the observing system may vary depending on the DA system used. The DA cycling experiments are performed for the ~4-week period of July 2016, and a 48 h model forecast is generated from each analysis. The results indicate that 3DVAR analysis fits more closely with the observations than HYBRID analysis. The domain-wide verification with respect to radiosonde observations reveals that forecasts in HYBRID experiments are more accurate than the 3DVAR experiments, in general. The wind forecasts show more improvements near the upper troposphere for the HYBRID run, with the neutral impact of INSAT-3D AMV observations. Compared with the forecasts from HYBRID analysis, the impact of INSAT-3D AMV observations is more pronounced in the 3DVAR DA system for wind forecasts over land. Geographical distribution depicts the positive impacts of INSAT-3D AMV observations across the whole domain in both 3DVAR and HYBRID DA systems. The AMV observations show a larger relative impact in HYBRID than in3DVAR, and the relative improvement compared to 3DVAR is 77% for wind and 71% for tropospheric temperature. Time evolution of forecast errors with respect to ERA-Interim analysis in the zonal wind over the Arabian Sea indicates a larger growth rate in the 3DVAR experiment in comparison to the HYBRID experiment, while the assimilation of AMV observations considerably reduces forecast errors in both DA systems. The HYBRID DA system that assimilated AMV observations show improvement in the meridional component of near-surface winds when validated against ASCAT observations. The HYBRID run reduces the bias in precipitation forecast, especially when AMV observations are incorporated. The skill scores for quantitative evaluation of precipitation forecast indicate a modest improvement in rainfall for the HYBRID run, and incorporating the AMV observation does not considerably enhance the skill of 24 h and 48 h rainfall forecast. As expected, the results from the study indicate that the impact of INSAT-3D AMV observations varies in 3DVAR and HYBRID DA systems. Furthermore, the impact of the new observing system shows more value

to the advanced DA systems such as HYBRID than the traditional 3DVAR approach. However, the ensemble system needs to be properly configured for the DA system to perform optimally.

The Chapter 6 address the impact of INSAT 3D for the simulation of tropical cyclones in 3DVAR and HYBRID DA systems over the BoB. Five tropical cyclones with intensity ranging from very severe cyclonic storm to super cyclone are considered. The track and intensity of TC are evaluated with respect to IMD best track data, and the quantitative precipitation forecasts during the landfall are validated with respect to GPM rainfall data. Results indicate that the error in initial position and intensity of TC is lower in HYBRID than in 3DVAR analysis. Assimilation of INSAT-3D AMV observations has shown nominal improvements in the initial track position and intensity of TC. In contrast, the forecast of the track shows positive impact of AMV in both the DA system. Assimilation of INSAT-3D AMV observations has improved the forecast of TC landfall locations in both the DA systems, and the impact is more substantial in the HYBRID-AMV experiment. The relative improvements in intensity compared to 3DVAR for TC are 18%, 13%, and 36%, respectively, for 3DVAR-AMV, HYBRID, and HYBRID-AMV. Rainfall forecast has improved significantly in both HYBRID and 3DVAR when INSAT-3D AMV data are incorporated. HYBRID experiments show improved quantitative precipitation skill scores for precipitation forecast compared to other experiments, in general.

In Chapter 7, the impact of the assimilation of the derived moisture field from radar reflectivity in 3DVAR and flow-dependent HYBRID DA system for the simulation of two thunderstorms (TS) over the northeastern region (NER) of India is evaluated. Among the two thunderstorm cases, one is driven by synoptic-scale weather system (Case-1) and the other is locally developed (Case-2). The initial conditions for Case-1 depicts a wet environment whereas for Case-2, the initial conditions depicts a dry background moisture field, which is not suitable for convection

initiation The vertical structure of the moisture analysis field shows significant impact of radar DA for Case-2 when compared to Case-1 in both 3DVAR and HYBRID DA system. Notable improvement of rainfall forecast is observed when radar observations are assimilated for the locally developed Case-2 thunderstorm, both in 3DVAR and HYBRID DA system. However, the impact of flow-dependent HYBRID DA is only marginal in both the thunderstorm cases.

**Future direction**

To gain further insight into the relative advantages and disadvantages, the results from the HYBRID DA system need to be compared with an advanced DA system such as 4DVAR. Further, the ETKF DA system does not use covariance localization for avoiding spurious long-distance correlations due to limited ensembles. To ameliorate such an effect in ETKF, an inflation factor, which is typically larger than other EnKF methods, is used in this study. Efforts are required to implement covariance localization in the ETKF DA system to reduce the sampling error. Further, the performance of flow-dependent ensemble covariance can be improved by accounting the model imperfections to the ensemble configuration. In addition, satellite radiance observations has not been considered for assimilation in this work, and it is expected that the direct assimilation of satellite radiance observations may improve the forecast manifold. A study by Emanuel and Zhang (2017) has revealed that the improvement in forecasting cyclone intensity is not just a matter of improved initialization of wind and the associated thermal field alone but about high-quality initialization of tropospheric water vapor. Compared to land, conventional observations over the ocean are either very minimal (e.g., buoy, aircraft) or not available. Since the primary source of observations over the oceans is dominated mainly by satellites, the direct assimilation of satellite radiance may further improve the intensity forecast of tropical cyclones. Additionally, the performance of the data assimilation systems needs to be tested for high-

resolution configurations to get more insights into the relatively lower impact of flow-dependent BEC in TC intensity prediction.

Further, for convective scale thunderstorm simulations, no significant impact of HYBRID is observed. This may be because this study has not utilized radar data assimilation to update the ensemble members using ETKF. In addition to that, the data assimilation cycling needs to be performed more frequently for convective-scale simulations. More frequent assimilation and updating the ensemble members with radar data can further improve results.

## REFERENCES


1. Abhilash S, Das S, Kalsi SR, et al (2007) Assimilation of Doppler weather radar observations in a mesoscale model for the prediction of rainfall associated with mesoscale convective systems. J earth Syst Sci 116:275–304

2. Barker DM (1999) Var scientific development paper 25: The use of synoptic-dependent error structure in 3DVAR. UK MET Met Off Tech Reports

3. Barker DM, Huang W, Guo Y-R, et al (2004) A three-dimensional (3DVAR) data assimilation system for use with MM5: Implementation and initial results. Mon Wea Rev 132:897–914

4. Barker DM, Lorenc A, Clayton A (2011) Hybrid variational/ensemble data assimilation

5. Berner J, Fossell KR, Ha S-Y, et al (2015) Increasing the skill of probabilistic forecasts: Understanding performance improvements from model-error representations. Mon Weather Rev 143:1295–1320

6. Brennan MJ, Hennon CC, Knabb RD (2009) The operational use of QuikSCAT ocean surface vector winds at the National Hurricane Center. Weather Forecast 24:621–645

7. Campbell WF, Bishop CH, Hodyss D (2010) Vertical covariance localization for satellite radiances in ensemble Kalman filters. Mon Weather Rev 138:282–290

8. Chu K, Xiao Q, Liu C (2013) Experiments of the WRF three-/four-dimensional variational (3/4DVAR) data assimilation in the forecasting of Antarctic cyclones. Meteorol Atmos Phys 120:145–156

9. Counillon F, Bethke I, Keenlyside N, et al (2014) Seasonal-to-decadal predictions with the ensemble Kalman filter and the Norwegian Earth System Model: a twin experiment. Tellus A Dyn Meteorol Oceanogr 66:21074

10. Daley R (1997) Atmospheric Data Assimilation (gtSpecial IssueltData Assimilation in Meteology and Oceanography: Theory and Practice). J Meteorol Soc Japan Ser II 75:319–329

11. Deb SK, Kishtawal CM, Kumar P, et al (2016) Atmospheric Motion Vectors from INSAT-3D: Initial quality assessment and its impact on track forecast of cyclonic storm NANAUK. Atmos Res 169:1–16

12. Deb SK, Kishtawal CM, Pal PK (2010) Impact of Kalpana-1-derived water vapor winds on Indian Ocean tropical cyclone forecasts. Mon Weather Rev 138:987–1003

13. Deb SK, Kumar P, Pal PK, Joshi PC (2011) Assimilation of INSAT data in the simulation of the recent tropical Cyclone Aila. Int J Remote Sens 32:5135–5155

14. Dhanya M, Chandrasekar A (2016) Impact of variational assimilation using multivariate background error covariances on the simulation of monsoon depressions over India. In: Annales Geophysicae. pp 187–201

15. Dowell DC, Wicker LJ, Snyder C (2011) Ensemble Kalman filter assimilation of radar observations of the 8 May 2003 Oklahoma City supercell: Influences of reflectivity observations on storm-scale analyses. Mon Weather Rev 139:272–294

16. Duc L, Saito K, Seko H (2013) Spatial-temporal fractions verification for high-resolution ensemble forecasts. Tellus A Dyn Meteorol Oceanogr 65:18171

17. Dudhia J (1989) Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. J Atmos Sci 46:3077–3107

18. Evensen G (2003) The ensemble Kalman filter: Theoretical formulation and practical implementation. Ocean Dyn 53:343–367

19. Fierro AO, Clark AJ, Mansell ER, et al (2015) Impact of storm-scale lightning data assimilation on WRF-ARW precipitation forecasts during the 2013 warm season over the contiguous United States. Mon Weather Rev 143:757–777

20. Findlater J (1978) Observational aspects of the low-level cross-equatorial jet stream of the western Indian Ocean. In: Monsoon Dynamics. Springer, pp 1251–1262

21. Gao J, Smith TM, Stensrud DJ, et al (2013) A real-time weather-adaptive 3DVAR analysis system for severe weather detections and warnings. Weather Forecast 28:727–745

22. Gao S, Min J (2018) Comparison of 3DVar and EnSRF Data Assimilation Using Radar Observations for the Analysis and Prediction of an MCS. Adv Meteorol 2018:

23. Greeshma MM, Srinivas C V, Yesubabu V, et al (2015) Impact of local data assimilation on tropical cyclone predictions over the Bay of Bengal using the ARW model. Ann Geophys 33:

24. Hamill TM, Snyder C (2000) A hybrid ensemble Kalman filter--3D variational analysis scheme. Mon Weather Rev 128:2905–2919

25. Hamill TM, Whitaker JS, Fiorino M, Benjamin SG (2011a) Global ensemble predictions of 2009�s tropical cyclones initialized with an ensemble Kalman filter. Mon Weather Rev 139:668–688

26. Hamill TM, Whitaker JS, Kleist DT, et al (2011b) Predictions of 2010�s tropical cyclones using the GFS and ensemble-based data assimilation methods. Mon Weather Rev 139:3243–3247

27. Hamill TM, Whitaker JS, Snyder C (2001) Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter. Mon Weather Rev 129:2776–2790. https://doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2

28. He Z, Wang D, Qiu X, et al Application of radar data assimilation on convective precipitation forecasts based on water vapor retrieval. Meteorol Atmos Phys 1–19

29. Hong S-Y, Dudhia J, Chen S-H (2004) A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. Mon Weather Rev 132:103–120

30. Hong S-Y, Noh Y, Dudhia J (2006) A new vertical diffusion package with an explicit treatment of entrainment processes. Mon Weather Rev 134:2318–2341

31. Houtekamer PL, Mitchell HL (2001) A sequential ensemble Kalman filter for atmospheric data assimilation. Mon Weather Rev 129:123–137

32. Hsiao L-F, Chen D-S, Kuo Y-H, et al (2012) Application of WRF 3DVAR to operational typhoon prediction in Taiwan: Impact of outer loop and partial cycling approaches.

Weather Forecast 27:1249–1263

33. Jiang X, Li T, Wang B (2004) Structures and mechanisms of the northward propagating boreal summer intraseasonal oscillation. J Clim 17:1022–1039

34. Jing-Yao L, Bao-De C, Hong L, et al (2013) Typhoon track forecast with a hybrid GSI-ETKF data assimilation system. Atmos Ocean Sci Lett 6:161–166

35. Jones CD, Macpherson B (1997) A latent heat nudging scheme for the assimilation of precipitation data into an operational mesoscale model. Meteorol Appl A J Forecast Pract Appl Train Tech Model 4:269–277

36. Kain JS (2004) The Kain--Fritsch convective parameterization: an update. J Appl Meteorol 43:170–181

37. Kalnay E (2003) Atmospheric modeling, data assimilation and predictability. Cambridge university press

38. Kaur I, Kumar P, Deb SK, et al (2015) Impact of Kalpana-1 retrieved atmospheric motion vectors on mesoscale model forecast during summer monsoon 2011. Theor Appl Climatol 120:587–599

39. Kelly G, Thépaut J-N, Buizza R, Cardinali C (2007) The value of observations. I: Data denial experiments for the Atlantic and the Pacific. Q J R Meteorol Soc 133:1803–1815. https://doi.org/https://doi.org/10.1002/qj.150

40. Kleist DT, Parrish DF, Derber JC, et al (2009) Introduction of the GSI into the NCEP global data assimilation system. Weather Forecast 24:1691–1705

41. Kumar P, Deb SK, Kishtawal CM, Pal PK (2017) Impact of assimilation of INSAT-3D retrieved atmospheric motion vectors on short-range forecast of summer monsoon 2014 over the south Asian region. Theor Appl Climatol 128:575–586

42. Kumar P, Singh R, Joshi PC, Pal PK (2011) Impact of additional surface observation network on short range weather forecast during summer monsoon 2008 over Indian subcontinent. J earth Syst Sci 120:53–64

43. Kutty G, Muraleedharan R, Kesarkar AP (2018) Impact of Representing Model Error in a Hybrid Ensemble-Variational Data Assimilation System for Track Forecast of Tropical Cyclones over the Bay of Bengal. Pure Appl Geophys 175:1155–1167

44. Kutty G, Wang X (2015) A comparison of the impacts of radiosonde and AMSU radiance observations in GSI based 3DEnsVar and 3DVar data assimilation systems for NCEP GFS. Adv Meteorol 2015:

45. Lai A, Min J, Gao J, et al (2020) Assimilation of Radar Data, Pseudo Water Vapor, and Potential Temperature in a 3DVAR Framework for Improving Precipitation Forecast of Severe Weather Events. Atmosphere (Basel) 11:182

46. Leslie LM, LeMarshall JF, Morison RP, et al (1998) Improved hurricane track forecasting from the continuous assimilation of high quality satellite wind data. Mon Weather Rev 126:1248–1258

47. Lewis JM, Lakshmivarahan S, Dhall S (2006) Dynamic data assimilation: a least squares approach. Cambridge University Press

48. Lewis WE, Velden CS, Stettner D (2020) Strategies for Assimilating High-Density Atmospheric Motion Vectors into a Regional Tropical Cyclone Forecast Model (HWRF). Atmosphere (Basel) 11:673

49. Li H, Xu JP (2011) Development of data assimilation and its application in ocean science. Mar Sci Bull 30:463–472

50. Li H, Xu X, Hu Y, et al (2018) Assimilation of Doppler radar data and its impact on prediction of a heavy Meiyu frontal rainfall event. Adv Meteorol 2018:

51. Li H, Xu X, Hu Y, et al (2013) Assimilation of Doppler radar data and its impact on prediction of a heavy Meiyu frontal rainfall event. Mon Weather Rev 141:2224–2244

52. Li Y, Wang X, Xue M (2012) Assimilation of radar radial velocity data with the WRF hybrid ensemble--3DVAR system for the prediction of Hurricane Ike (2008). Mon Weather Rev 140:3507–3524

53. Lim AHN, Jung JA, Nebuda SE, et al (2019) Tropical cyclone forecasts impact assessment from the assimilation of hourly visible, shortwave, and clear-air water vapor atmospheric motion vectors in HWRF. Weather Forecast 34:177–198

54. Lim K-SS, Hong S-Y (2010) Development of an effective double-moment cloud microphysics scheme with prognostic cloud condensation nuclei (CCN) for weather and climate models. Mon Weather Rev 138:1587–1612

55. Lorenc AC (2003) The potential of the ensemble Kalman filter for NWP—a comparison with 4D-Var. Q J R Meteorol Soc A J Atmos Sci Appl Meteorol Phys Oceanogr 129:3183–3203

56. Lorenc AC (1986) Analysis methods for numerical weather prediction. Q J R Meteorol Soc 112:1177–1194

57. Lorenc AC, Ballard SP, Bell RS, et al (2000) The Met. Office global three-dimensional variational data assimilation scheme. Q J R Meteorol Soc 126:2991–3012

58. Lu X, Wang X, Li Y, et al (2017) GSI-based ensemble-variational hybrid data assimilation for HWRF for hurricane initialization and prediction: impact of various error covariances for airborne radar observation assimilation. Q J R Meteorol Soc 143:223–239

59. Malakar P, Kesarkar AP, Bhate J, Deshamukhya A (2020a) Appraisal of data assimilation techniques for dynamical downscaling of the structure and intensity of tropical cyclones. Earth Sp Sci 7:e2019EA000945

60. Malakar P, Kesarkar AP, Bhate JN, et al (2020b) Comparison of reanalysis data sets to comprehend the evolution of tropical cyclones over North Indian Ocean. Earth Sp Sci 7:e2019EA000978

61. Mazzarella V, Maiello I, Capozzi V, et al (2017) Comparison between 3D-Var and 4D-Var data assimilation methods for the simulation of a heavy rainfall case in central Italy. Adv Sci Res 14:271–278

62. Mlawer EJ, Taubman SJ, Brown PD, et al (1997) Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. J Geophys Res

Atmos 102:16663–16682

63. Mounika K, Kutty G, Gorthi SSRK, others (2018) Consistent Robust and Recursive Estimation of Atmospheric Motion Vectors From Satellite Images. IEEE Trans Geosci Remote Sens 57:1538–1544

64. Ooyama K (1969) Numerical simulation of the life cycle of tropical cyclones. J Atmos Sci 26:3–40

65. Pai DS, Sridhar L, Badwaik MR, Rajeevan M (2015) Analysis of the daily rainfall events over India using a new long period (1901--2010) high resolution (0.25$\times$ 0.25) gridded rainfall data set. Clim Dyn 45:755–776

66. Parrish DF, Derber JC (1992) The National Meteorological Center's spectral statistical-interpolation analysis system. Mon Weather Rev 120:1747–1763

67. Prasad SK, Mohanty UC, Routray A, et al (2014) Impact of Doppler weather radar data on thunderstorm simulation during STORM pilot phase$\diamondsuit$2009. Nat hazards 74:1403–1427

68. Prasad VS, Johny CJ, Sodhi JS (2016) Impact of 3D Var GSI-ENKF hybrid data assimilation system. J Earth Syst Sci 125:1509–1521

69. Rajeevan M, Bhate J, Jaswal AK (2008) Analysis of variability and trends of extreme rainfall events over India using 104 years of gridded daily rainfall data. Geophys Res Lett 35:

70. Rakesh V, Goswami P (2011) Impact of background error statistics on forecasting of tropical cyclones over the north Indian Ocean. J Geophys Res Atmos 116:

71. Rakesh V, Prashant G, others (2015) Impact of data assimilation on high-resolution rainfall forecasts: a spatial, seasonal, and category analysis. J Geophys Res Atmos 120:359–377

72. Rakesh V, Singh R, Pal PK, Joshi PC (2009) Impacts of satellite-observed winds and total precipitable water on WRF short-range forecasts over the Indian region during the 2006 summer monsoon. Weather Forecast 24:1706–1731

73. Rani SI, Das Gupta M (2014) An inter-comparison of K alpana-1 and M eteosat-7

atmospheric motion vectors against radiosonde winds and NWP forecasts during monsoon 2011. Meteorol Appl 21:820–830

74. Routray A, Mohanty UC, Das AK, Sam N V (2005) Study of heavy rainfall event over the west-coast of India using analysis nudging in MM5 during ARMEX-I. Mausam 56:107–120

75. Routray A, Mohanty UC, Osuri KK, et al (2016) Impact of satellite radiance data on simulations of Bay of Bengal tropical cyclones using the WRF-3DVAR modeling system. IEEE Trans Geosci Remote Sens 54:2285–2303

76. Routray A, Mohanty UC, Osuri KK, Prasad SK (2013) Improvement of monsoon depressions forecast with assimilation of Indian DWR data using WRF-3DVAR analysis system. Pure Appl Geophys 170:2329–2350

77. Routray A, Mohanty UC, Rizvi SRH, et al (2010) Impact of Doppler weather radar data on numerical forecast of Indian monsoon depressions. Q J R Meteorol Soc 136:1836–1850

78. Sankhala DK, Deb SK, Sathiyamoorthy V (2019) INSAT-3D low-level atmospheric motion vectors: Capability to capture Indian summer monsoon intra-seasonal variability. J Earth Syst Sci 128:31

79. Sankhala DK, Deb SK, Sharma SK, Lal S (2020) Inter-comparison of INSAT-3D atmospheric motion vectors height with cloud-base height from a Ceilometer. Int J Remote Sens 41:2946–2961

80. Sawada M, Ma Z, Mehra A, et al (2019) Impacts of Assimilating High-Resolution Atmospheric Motion Vectors Derived from Himawari-8 on Tropical Cyclone Forecast in HWRF. Mon Weather Rev 147:3721–3740

81. Sawada M, Ma Z, Mehra A, et al (2020) Assimilation of Himawari-8 Rapid-Scan Atmospheric Motion Vectors on Tropical Cyclone in HWRF System. Atmosphere (Basel) 11:601

82. Schaefer JT (1990) The critical success index as an indicator of warning skill. Weather Forecast 5:570–575

83. Shen F, Min J (2015) Assimilating AMSU-A radiance data with the WRF Hybrid En3DVAR system for track predictions of Typhoon Megi (2010). Adv Atmos Sci 32:1231–1243

84. Shen F, Min J, Xu D (2016) Assimilation of radar radial velocity data with the WRF Hybrid ETKF–3DVAR system for the prediction of Hurricane Ike (2008). Atmos Res 169:127–138. https://doi.org/https://doi.org/10.1016/j.atmosres.2015.09.019

85. Skamarock C, Klemp B, Dudhia J, et al (2019) A Description of the Advanced Research WRF Model Version 4

86. Skamarock WC, Klemp JB, Dudhia J, et al (2008) 2005: A description of the advanced research WRF version 2. In: NCAR Tech. Note

87. Soden BJ, Velden CS, Tuleya RE (2001) The impact of satellite winds on experimental GFDL hurricane model forecasts. Mon Weather Rev 129:835–852

88. Speer MS, Leslie LM (2002) The prediction of two cases of severe convection: implications for forecast guidance. Meteorol Atmos Phys 80:165–175

89. Srinivas C V, Mohan GM, Yesubabu V, et al (2017) Data Assimilation Experiments with ARW--3DVAR for Tropical Cyclone Extreme Weather Predictions Over Bay of Bengal. In: Tropical Cyclone Activity over the North Indian Ocean. Springer, pp 317–336

90. Srivastava K, Bhardwaj R, Bhowmik SKR (2011) Assimilation of Indian Doppler Weather Radar observations for simulation of mesoscale features of a land-falling cyclone. Nat Hazards 59:1339–1355

91. Stull RB (1988) An introduction to boundary layer meteorology. Springer Science \& Business Media

92. Sun J, Wang H Radar Data Assimilation with WRF 4D-Var. Part II: Comparison with 3D-Var for a Squall Line over the U.S. Great Plains. Mon Weather Rev 141:2245–2264. https://doi.org/10.1175/MWR-D-12-00169.1

93. Sun J, Zhang Y (2008) Analysis and prediction of a squall line observed during IHOP using

multiple WSR-88D observations. Mon Weather Rev 136:2364–2388

94. Thiruvengadam P, Indu J, Ghosh S (2020) Significance of 4DVAR Radar Data Assimilation in Weather Research and Forecast Model-Based Nowcasting System. J Geophys Res Atmos 125:e2019JD031369. https://doi.org/https://doi.org/10.1029/2019JD031369

95. Tong C-C, Jung Y, Xue M, Liu C (2020) Direct Assimilation of Radar Data With Ensemble Kalman Filter and Hybrid Ensemble-Variational Method in the National Weather Service Operational Data Assimilation System GSI for the Stand-Alone Regional FV3 Model at a Convection-Allowing Resolution. Geophys Res Lett 47:e2020GL090179

96. Torn RD, Hakim GJ (2009) Ensemble data assimilation applied to RAINEX observations of Hurricane Katrina (2005). Mon Weather Rev 137:2817–2829

97. Tyagi A (2007) Thunderstorm climatology over Indian region. Mausam 58:189

98. Velden CS, Hayden CM, Paul Menzel W, et al (1992) The impact of satellite-derived winds on numerical hurricane track forecasting. Weather Forecast 7:107–118

99. Wang H, Sun J, Fan S, Huang X-Y (2013a) Indirect assimilation of radar reflectivity with WRF 3D-Var and its impact on prediction of four summertime convective events. J Appl Meteorol Climatol 52:889–902

100. Wang H, Sun J, Zhang X, et al (2013b) Radar data assimilation with WRF 4D-Var. Part I: System development and preliminary testing. Mon Weather Rev 141:2224–2244

101. Wang X (2010) Incorporating ensemble covariance in the gridpoint statistical interpolation variational minimization: A mathematical framework. Mon Weather Rev 138:2990–2995

102. Wang X (2011) Application of the WRF hybrid ETKF--3DVAR data assimilation system for hurricane track forecasts. Weather Forecast 26:868–884

103. Wang X, Barker DM, Snyder C, Hamill TM (2008) A hybrid ETKF--3DVAR data assimilation scheme for the WRF model. Part I: Observing system simulation experiment.

Mon Weather Rev 136:5116–5131

104. Wang, X., Barker, D. M., Snyder, C., Hamill, T. M. (2008b). A hybrid ETKF–3DVAR data assimilation scheme for the WRF model. Part II: Real observation experiments. Mon Weather Rev 136:5132-5147.

105. Wang X, Parrish D, Kleist D, Whitaker J (2013c) GSI 3DVar-based ensemble--variational hybrid data assimilation for NCEP Global Forecast System: Single-resolution experiments. Mon Weather Rev 141:4098–4117

106. Wang X, Snyder C, Hamill TM (2007) On the theoretical equivalence of differently proposed ensemble--3DVAR hybrid analysis schemes. Mon Weather Rev 135:222–227

107. Weisman ML, Skamarock WC, Klemp JB (1997) The resolution dependence of explicitly modeled convective systems. Mon Weather Rev 125:527–548

108. Westermayer A, Groenemeijer P, Pistotnik G, et al (2017) Identification of favorable environments for thunderstorms in reanalysis data. Meteorol Zeitschrift 26:59–70

109. Wu T-C, Liu H, Majumdar SJ, et al (2014) Influence of assimilating satellite-derived atmospheric motion vector observations on numerical analyses and forecasts of tropical cyclone track and intensity. Mon Weather Rev 142:49–71

110. Xiao Q, Kuo Y-H, Sun J, et al Assimilation of Doppler Radar Observations with a Regional 3DVAR System: Impact of Doppler Velocities on Forecasts of a Heavy Rainfall Case. J Appl Meteorol 44:768–788. https://doi.org/10.1175/JAM2248.1

111. Xu D, Min J, Shen F, et al (2016) Assimilation of MWHS radiance data from the FY-3B satellite with the WRF Hybrid-3DVAR system for the forecasting of binary typhoons. J Adv Model Earth Syst 8:1014–1028

112. Yang J, Dua KQ, Wu JK, et al (2014) Effect of Data Assimilation Using WRF-3DVAR for Heavy Rain Prediction on the Northeastern Edge of the Tibetan Plateau. Advances in Meteorology, 2015, Article ID 294589

113. Zhang F, Weng Y, Sippel JA, et al (2009) Cloud-resolving hurricane initialization and

prediction through assimilation of Doppler radar observations with an ensemble Kalman filter. Mon Weather Rev 137:2105–2125

114. Zhang F, Yang Y, Wang C, et al (2015) The Effects of Assimilating Conventional and ATOVS Data on Forecasted Near-Surface Wind with WRF-3DVAR. Mon Weather Rev 143:153–164. https://doi.org/10.1175/MWR-D-14-00038.1

115. Zhang M, Zhang F, Huang X-Y, Zhang X (2011) Intercomparison of an Ensemble Kalman Filter with Three- and Four-Dimensional Variational Data Assimilation Methods in a Limited-Area Model over the Month of June 2003. Mon Weather Rev 139:566–572. https://doi.org/10.1175/2010MWR3610.1

116. Zhang S, Pu Z, Velden C (2018) Impact of enhanced atmospheric motion vectors on HWRF hurricane analyses and forecasts with different data assimilation configurations. Mon Weather Rev 146:1549–1569

## LIST OF PAPERS BASED ON THESIS

1. **Gogoi, R.B**., Kutty, G. & Boroghain, A. (2021) Intercomparison of the impact of INSAT-3D atmospheric motion vectors in 3DVAR and hybrid ensemble-3DVAR data assimilation systems during Indian summer monsoon. **Theoretical and Applied Climatology** (2021). https://doi.org/10.1007/s00704-021-03649-2

2. **Gogoi, R. B.**, Kutty, G., Rakesh, V., & Borgohain, A. (2020). Comparison of the Performance of Hybrid ETKF-3DVAR and 3DVAR Data Assimilation Systems on Short-Range Forecasts during Indian Summer Monsoon Season in a Limited-Area Model. **Pure and Applied Geophysics, 177(10), 5007-5026.**

3. Kutty, G., **Gogoi, R.,** Rakesh, V., & Pateria, M. (2020). Comparison of the performance of HYBRID ETKF-3DVAR and 3DVAR data assimilation scheme on the forecast of tropical cyclones formed over the Bay of Bengal. **Journal of Earth System Science, 129(1), 1-14.**

4. **Gogoi, R.B**., Kutty, G. & Boroghain, A. (2021) Intercomparison of the Impact of INSAT-3D Atmospheric Motion Vectors in 3DVAR and Hybrid Ensemble-3DVAR Data Assimilation Systems on the Forecast of Tropical Cyclones formed over the Bay of Bengal. **Modeling Earth Systems and Environment, 1-11.**

5. Indirect assimilation of Doppler Weather Radar Reflectivity for Simulation of Thunderstorms over North-eastern Region of India using WRF model ( **In preparation**)

## Conferences

1. Rohith Muralee dharan, **Rekha Bharali Gogoi**, and Govindan Kutty, 2016: Impact of Flow-Dependent Error Covariance in the Variational Data Assimilation System for the Forecast of Two Tropical Cyclones Over Bay of Bengal. National Symposium on Tropical Meteorology: Climate change and Coastal vulnerability- **Tropmet-2016**

2. **Rekha Bharali Gogoi** and Govindan Kutty: Impact of flow - dependent ensemble error covariance in the three – dimensional variational data assimilation method in a limited area model over the month of July 2013, Oral Presentation, **Tropmet-2018**