Improved simulation of meteorological systems over India: Impact of assimilating satellite observations using 3DVar and EnKF

A thesis submitted
in partial fulfillment for the degree of

Doctor of Philosophy

by

DHANYA M.

Department of Earth and Space Sciences

INDIAN INSTITUTE OF SPACE SCIENCE AND TECHNOLOGY
Thiruvananthapuram - 695 547

April 2016

CERTIFICATE

This is to certify that the thesis entitled **Improved simulation of meteorological systems over India : Impact of assimilating satellite observations using 3DVar and EnKF** submitted by **Dhanya M.** 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 my 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.

Dr. A. Chandrasekar
Supervisor
Department of Earth and Space Sciences

Thiruvananthapuram
April 2016

Counter signature of HOD with seal

DECLARATION

I declare that this thesis entitled **Improved simulation of meteorological systems over India : Impact of assimilating satellite observations using 3DVar and EnKF** 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. A.Chandrasekar**, 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 Dhanya M.

April 2016 (SC11D005)

ACKNOWLEDGEMENTS

Life is a journey with twists and turns, ups and downs. After five years filled with moments of agony and ecstasy, anguish and elation, my journey has reached a significant juncture. There are many, whose support and encouragement made this possible.

First, I would like to express my sincere gratitude to my advisor, Dr.A.Chandrasekar for his guidance, always being accessible inspite of being extremely busy, giving constructive advice that encouraged me to grow as a researcher. I have also learnt a lot from his example of hard word, patience and humility. Furthermore, I would like to thank him for the many conversations we've had on academic as well as non-academic matters which have been a positive influence on my personal life also.

All my doctoral committee members have provided support and encouragement during my PhD research. I would like to acknowledge the valuable suggestions and feedback provided by my committee members Dr. Anandmayee Tej, Dr. P.K.Pal, Dr.S.K.Dash, Dr.Amit Kesarkar, Dr. Raju K. George, Dr. M.V.Ramana and Dr. Sarita Vig during my research work.

I would like to thank our director, Dr. K.S.Dasgupta for providing me this opportunity to be a part of this institute. I would also like to thank him for enabling the use of our high performance computing system for my research. Thanks are due to all the faculty members of IIST for their support and goodwill.I would like to thank Dr. Gorthi R. K. S. S. Manyam for his support and encouragement.

I also take this opportunity to thank Prof. A. Jayaraman, director, National Atmospheric Research Laboratory, for permitting the use of their high performance computing system. A big thanks is due to Dr. Amit Kesarkar, for providing access to this system as well as for all his support, advices and encouragements. I wholeheartedly thank all the members of the WCRG group at NARL and especially Ms. Jyoti Bhate, Dr. Vikas Singh, Rajasree, Umakanth and Harish.

I would also like to thank the Computer system group at IIST for their help with the computing facilities. I would also like to thank all members of library and administrative staff at IIST.

We always look upon the ones who have walked our way before, for valuable insights. I am thankful to my seniors, Dr. Pankaj Kumar Sinha and Dr. Govindankutty for helping me through this journey. I should thank Dr.Govindankutty especially, for his valuable support right from the beginning of my PhD work, till the very end. He has been my immediate point of consultation in any research-related queries throughout this period.

All throughout my life at IIST, my fellow PhD students, my friends, have been a wonderful source of comfort. I would like to thank all of them for their companionship, especially, Deepak, Arun, Bharath, Reneesh, Gopakumar to name a few. A special thanks to my friends in the hostel, in particular, Haripadmam, Sarika, Sarah, Remya, Kavitha, Dhanya, Aneesha, Jalaja, Harsha, Rashmi, Ruchi and Preethi for all the impromptu discussions on everything under the sun and for all the warm, wonderful memories.

I am blessed to have some wonderful friends, who, even though are physically far, gave me support and encouragement. I would like to thank all my MTech classmates for being such a wonderful bunch of friends. Thanks to them for being there and giving me valuable insights on PhD as well as life in general. I am also thankful to Chandrasekar who has been walking along the same path at the same time and for all 'PhD-horror stories' we shared. It is my pleasure to acknowledge the wonderful discussions I had with Dr. Anand Narayanan, on other aspects of life and am thankful to him for providing those views of life outside research. I am very thankful to my best friends Smitha and Mathew, for telling me repeatedly that even though the PhD feels hard like mountain-climbing, the view at the top will be worth it. There are absolutely no words to express my thanks to Dr.Rajmohan Kombiyil, my former colleague and friend who has encouraged, inspired and at times pushed me hard to excel at what I do and supported me throughout the many ups and downs of PhD.

Last, but definitely not the least, let me thank my family - my parents for their love, encouragement, faith and support they have given me. All my extended family of grandparents, aunts and uncles and cousins have also been wonderfully supportive

and I am truly blessed to be among them.

The whole PhD experience has been a wonderful journey filled with the whole spectrum of human emotions. I am blessed to have traveled along this path in the company of so many wonderful people, all of whom I could not mention here. And, for all those blessings, I thank God.

Dhanya M.

ABSTRACT

The main objective of this thesis study is to investigate the impact of assimilating satellite observations for simulating various mesoscale weather features over the Indian region. The influence of background error covariance in determining the impact of satellite observations is the main focus of the present thesis study.

The impact of assimilating humidity information from MeghaTropiques SAPHIR (Sounder for Probing Vertical Profiles of Humidity) radiances in simulating three tropical cyclones that formed over the Bay of Bengal is investigated in chapter 3 of this thesis. The three dimensional Variational (3DVar) assimilation technique is used in the Weather Research and Forecasting (WRF) model. SAPHIR radiances have moderate positive impact in the simulation of various cyclone features in terms of minimum sea level pressure, maximum wind speed, area average temperature anomaly, relative vorticity as well as accumulated rainfall.

WRF 3DVar utilizes a static background error covariance matrix ($\mathbf{B}$) which is estimated using the National Meteorological Center (NMC) method. The various balance relations between the model variables are expressed through regression relations in this method. The formulation of the above-mentioned regression relations influence the background error correlations in 3DVar system. This significantly impacts the spread of observation information between the various grid points as well as model variables. This thesis study explores the two formulations - cv5 and cv6 - available in the WRF 3DVar system. The influence of the formulation of $\mathbf{B}$ is investigated in chapter 4 by performing assimilation of conventional as well as satellite radiance observations from the Advanced microwave sounding unit (AMSU-A) using cv5 and cv6 options. Three case studies involving the simulation of monsoon depressions over the Indian region are performed comparing the impacts of cv5 and cv6 options. It is seen that the moisture field is maximum impacted by the cv6 formulation of $\mathbf{B}$. Model simulations of horizontal wind divergence, moisture convergence, temperature anomaly, relative humidity as well as rainfall are seen to be moderately influenced by the choice of formulation of $\mathbf{B}$.

The impact of using cv5 and cv6 options in the 3DVar assimilation of SAPHIR radiances is investigated in the Chapter 5 , by considering the simulation of three tropical cyclone cases over Bay of Bengal. The simulation of tropical cyclones are also impacted in a moderate way by the choice of B matrix used.

3DVar techniques have the disadvantage of utilizing a static B. The evolution of background errors, are, hence not represented adequately in 3DVar assimilation technique. Flow-dependent B using ensembles provides a method to alleviate this disadvantage. Chapter 6 explores the use of ensembles to provide $\mathbf{B}$. Ensemble Kalman filter technique is used to assimilate satellite wind observations from Oceansat-2 satellite in the simulation of two heavy rainfalls during the north east monsoon as well as the simulation of a monsoon depression. It is seen that the use of EnKF method in assimilating observations provide positive impact on the simulation of these mesoscale weather phenomena as compared with assimilation using 3DVar method. The present thesis study, thus, indicates the significant role of $\mathbf{B}$ in determining the impact of satellite data assimilation over Indian region.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

Introduction

The fundamental philosophy of weather prediction owes its origin to the Norwegian scientist Vilhelm Bjerknes, who, in 1904, proposed the following procedure for numerical model-based weather forecasting. The above procedure has two important steps, namely, (i) a diagnostic step where the initial state of the atmosphere is determined using observations and (ii) a prognostic step in which physical laws as expressed as a system of coupled nonlinear partial differential equations are used to estimate the change of the state with time (Lynch, 2008). The first practical step to implement the above was initiated by Lewis Fry Richardson in the early twentieth century. Richardson's valiant attempt at predicting weather however turned out to be unsuccessful. Successful modern numerical weather prediction became a reality with the advent of the first digital computer during 1950s. Since the realization of the first barotropic model forecast made in early 1950s, the subsequent numerical weather prediction models have shown considerable improvement as a result of increased computational power, improved physics of the model, increased availability of data especially satellite data over sparsely observed regions and the use of more accurate data assimilation methods (Kalnay, 2003). Shuman (1989) has opined that the procedures followed by Richardson were quite similar to those that are in operational use today. However, the deficiencies in the knowledge of modeling the atmospheric physics component, numerical instabilities as well as lack of requisite computational power resulted in the failure of Richardson's experiment.

The numerical weather prediction is essentially an initial value problem. Hence the accurate specification of the initial state of the atmospheric system is a very important requirement for a successful weather forecast. Data assimilation is a robust technique designed to determine this initial state of the atmosphere, called the 'analysis'. Estimation of the analysis through the assimilation technique is a complex and non-trivial problem due to the following reasons. The fundamental problem is due to the under-determined nature of estimating the initial state of the atmosphere from its observations. The typical degrees of freedom (the number of values required to

prescribe the initial state of the atmosphere which depends on the number of three dimensional grid points and the number of dependent variables of the atmosphere) of a numerical model is $\sim 10^7$ whereas the total available atmospheric observations are two orders lower than this. Furthermore, the observations of the atmosphere do not sample the system in a homogeneous way, both in space and time. Also, different types of observations have different patterns of error structures. In addition, many important sets of observations like satellite radiances and observations from Doppler weather radar (DWR) such as radial velocity and reflectivity do not manifest as model variables. Data assimilation systems provide the statistically most likely state of the atmosphere, taking into account all the information from atmospheric observations as well as a 'prior' or 'background information' about the atmospheric system. The 'prior' or 'background' or 'first guess' information of the atmosphere typically is taken to be the output of a short numerical model forecast itself.

Data assimilation usually proceeds in a sequential manner. A short atmospheric model forecast defines the apriori first guess available for specifying background information about the atmospheric system. The available observations of the atmosphere are used to provide a correction to this background state to generate the "analysis" at any time. From the analysis, the model is integrated forward in time further. The above atmospheric model forecast provides the background information at the next analysis time. The above sequence continues over several assimilation cycles. The information provided by the first guess as well as observations of the atmosphere are therefore significant for obtaining the best possible analysis. Both the above sources of atmospheric information include associated errors of the observations and the background . Hence a statistical approach is necessary to blend them in a physically consistent manner. Several such blending (data assimilation) techniques exist for this purpose which are briefly reviewed in the next section.

1.1 Data assimilation techniques

The earliest available method for quantitative numerical estimation is known as the 'objective analysis'. Panofsky (1949) employed interpolation of observations to model grid positions using curve fitting methods and applied suitable 'weights' de-

pending on the accuracy of each observation. The above method was further improved by Gilchrist and Cressman, who in 1954, proposed restricting the polynomial fit of observations within a local region of influence. In addition to assigning weights to observations based on subjective analyses and incorporating geostrophic constraints, they proposed the following two suggestions. The suggestions are the following. (i) include automatic quality check of observations and (ii) incorporate a previous estimate of the state of the system in the analysis procedure (Daley, 1993).

The successive corrections method (SCM) was developed by Bergthórsson and Döös (1955). In the SCM method, the analysis increments at each grid points were estimated as weighted linear combination of observation increments in a region of influence. The observations weights were inversely proportional to the distance between observation and model grid points. The analysis increments were added to the background information to generate an analysis of the state. An iterative variation of this method was suggested by Gandin in 1959. The methods of optimal interpolation (OI) provided for the weights to observations relative to their errors (Gandin and Hardin, 1965). OI method was operationally employed in many forecasting centers during 1970s and 1980s.

Another popular technique used for producing optimal estimates of atmospheric state is nudging or Newtonian relaxation. This method adjusts dynamical variables of models using the observations. A term proportional to the difference between model variable and observation is included in the dynamical prognostic equations. The nudging procedure "adjusts" or "relaxes" the model state close to the observations.

One of the major advancements in data assimilation methods was the introduction of variational technique (Sasaki, 1958). In three dimensional variational (3DVar) technique, a cost function proportional to the square of the distance between the analysis and both the background and the observations is minimized. Minimization procedures like conjugate gradient method or quasi-newton methods are utilised in 3DVar. This has the further advantage that observations like satellite radiances and DWR observations such as radial velocity and reflectivity which do not manifest as model variables can be assimilated. The above variational technique has been extended to four dimensions to give the four dimensional variational technique (4DVar). In 4DVar, the cost function includes the distance of the analysis from observations over a time interval.

The 4DVar method is computationally expensive.

The utilisation of Kalman filter algorithm (Kalman, 1960) provides distinct advantages among the several data assimilation methods. Broadly, the Kalman filter algorithm proceeds in two steps. In the first step, estimates of the current state variables along with their uncertainties are generated. These estimates are updated in the second step when observations are available, by prescribing weights based on their errors. This is a recursive algorithm and is known to be an optimal filter for linear systems. However, for a nonlinear system like a weather model, the Kalman filter equations are not directly applicable since the error covariances cannot be exactly determined. The use of Kalman filter equations for the purposes of data assimilation in the atmospheric system received widespread and immediate acceptance with the introduction of the Ensemble Kalman filter (EnKF) (Evensen, 1994; Burgers et al., 1998; Evensen, 2003) method. In the EnKF algorithm, the error covariances for background and observations are estimated using ensembles which can be obtained by perturbing model and observation states. Several variants of EnKF analysis algorithms exist in the literature. These fall into two broad categories: stochastic (Houtekamer and Mitchell, 1998) and deterministic, based on whether perturbed observations are used to obtain the ensemble analysis. Deterministic filters can be further categorized as ensemble adjustment Kalman filter (EAKF); (Anderson, 2001), ensemble transform Kalman filter (ETKF) ; (Bishop et al., 2001), ensemble square root filter (EnSRF) (Whitaker and Hamill, 2002)), and local ensemble transform Kalman filter (LETKF) (Hunt et al., 2007). As noted by Tippett et al. (2003), the first three of the above mentioned deterministic filters yield analysis ensembles with identical means and covariances when implemented from the same forecast ensemble.

All the assimilation techniques mentioned here combines a background information with the observation information by taking into account the error statistics of both of them. The weight given to each of these components is, in general, inversely proportional to their errors.

The variational and EnKF based assimilation techniques have been extensively employed in the recent decades in global and regional assimilation systems. Together with better observational capabilities through the use of satellite sensors, the use of sophisticated assimilation algorithms have resulted in marked improvements in nu-

merical forecasting of the weather. In this context, analysing the impact of such observations on the improved numerical forecast would help in better utilising the existing observations as well as in providing guidelines for setting up of future observing systems. Hence the observation impact studies as undertaken in this thesis which quantify the impact of satellite observations on weather simulations are highly relevant.

1.2 Impact of assimilating satellite observations using 3DVar

It is known that the assimilation of satellite observations using the 3DVar technique has significant impact on the simulation of mesoscale weather systems. A number of studies have demonstrated the positive impact of satellite observations in the simulation of weather phenomena over the Indian region by assimilating them using 3DVar technique in Weather Research and Forecasting (WRF) model. A brief review of such studies are given here.

Monsoon depressions are one of the most significant weather systems that form over the Indian region. These low pressure systems embedded in the monsoon trough during the south west monsoon (June-September) provide copious amount of rainfall over this region. Hence accurate simulation of monsoon depressions are considered very important. Variational assimilation of satellite observations can contribute to improved simulation of monsoon depressions since these systems form over ocean surfaces which are conventionally sparsely observed. Govindankutty and Chandrasekar (2010) demonstrated the positive impact of assimilating Moderate Resolution Imaging Spectroradiometer (MODIS) retrieved temperature and humidity profiles in simulating the dynamic and thermodynamic features of monsoon depressions over India. Sinha and Chandrasekar (2010) showed that assimilation of ocean surface winds from Quick Scatterometer (QuickSCAT) improves the simulation of mesoscale features of the monsoon depressions. The positive impact of assimilation of Advanced TIROS Vertical Sounder (ATOVS) temperature and humidity profiles and Spectral sensor microwave imager (SSM/I) total precipitable water (TPW) on the simulation of a monsoon depression was shown by Govindankutty and Chandrasekar (2011).

Several studies have demonstrated the considerable positive impacts of assimilating satellite observations on the simulation of tropical cyclones over the Indian region. Singh et al. (2008a) and Osuri et al. (2012) showed the positive influence of Special Sensor Microwave Imager (SSM/I) and Quick Scatterometer (QuikSCAT) observations in simulating tropical cyclone features. Singh et al. (2011) investigated the impact of the Quick Scatterometer (QuikSCAT) ocean surface winds, Special Sensor Microwave/Imager (SSM/I)-derived Total Precipitable Water (TPW), and Meteosat-7-derived Atmospheric Motion Vectors (AMVs) on the track and intensity prediction of tropical cyclones over the North Indian Ocean. Sowjanya et al, 2013 established that assimilation of total precipitable water and surface wind data retrieved from the Special Sensor for Microwave Imaginary (SSM/I) sensors improved the simulation of rainfall by the WRF model over the Indian monsoon region. Kumar et al. (2013a) evaluated the impact of scatterometer and radiometer data on tropical cyclone (Phet) prediction by assimilating Oceansat-2 scatterometer (OSCAT) winds, wind speed and precipitable water derived from the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI) in the WRF model. Srinivas et al. (2012a) showed the positive impact of assimilating the Quick Scatterometer (QuikSCAT) ocean surface winds on the simulations of intensity and track positions of the tropical cyclones Fanoos and Nargis. Yesubabu et al. (2014a) investigated the improvements in simulating the various features of two tropical cyclones Jal and Thane by assimilating Oceansat-2 winds and MODIS temperature/humidity profiles in the WRF model. Assimilation of Kalpana-1 AMVs showed positive impact in 12-h wind forecast over the tropical region in the upper troposphere as well as in rainfall forecast over India in Kaur et al, 2015.

The above mentioned studies utilized observations retrieved from satellite radiances using inversion algorithms. Variational techniques can directly ingest satellite radiance measurements through the use of a radiative transfer model. Theoretically, the assimilation of satellite radiances are expected to yield more positive impact as compared to the assimilation of retrieved meteorological information. Singh et al. (2012c) showed that assimilation of Advanced Microwave Sounding Unit-A (AMSU-A) radiances does improve tropical cyclone simulations. The positive impact of assimilation of assimilation of Kalpana Very High Resolution Radiometer (VHRR) channel 3 clear-sky radiances in the simulation of rainfall over India was demonstrated by

Singh et al. (2010). The impact of directly assimilating the Advanced Television and Infrared Observation Satellite Operational Vertical Sounder (ATOVS) radiance in the WRF model was investigated by Singh et al. (2012b). Singh et al. (2012a) demontrated that the assimilation of AIRS measurements has a significant impact on WRF analysis and short range forecasts over India. Greeshma et al. (2015) showed that assimilation of AMSU radiances and Atmospheric Motion Vectors (AMV) improved the intensity and track predictions of tropical cyclones. Routray et al. (2016) evaluated the improvement in simulation of tropical cyclones over the Bay of Bengal through the assimilation of satellite radiances

All the above mentioned studies highlight the importance of assimilating satellite observations in WRF model using 3DVar technique for improved simulation of mesoscale features over the Indian region.

One of the significant factors that affect the impact of observations is the background error covariance. However, not many studies have investigated the above impact "in-depth" for the simulation of mesoscale weather features over the Indian domain. Rakesh and Goswami (2011a, b) , showed that using region-specific background error covariances instead of global error covariances did improve the simulation of tropical cyclones over the North Indian Ocean (NIO) region. The significance of utilizing region-specific BEC was further reiterated by Routray et al (2014) who utilized the region specific BEC for simulating monsoon depressions. One of the main components in BEC modeling is the so-called "control variable transform" (CVT). CVT converts the numerical weather model variables to control variables of a data assimilation system as well as imposes approximate balance relations. A few studies have examined the effect of tuning the length scales in the CVTs in estimating BECs. In the default region-specific specification of BEC (called CV5 option) in WRF 3DVar system, balance relationships between the stream function with velocity potential, temperature, and surface pressure are defined. All previous studies over the Indian region assimilating meteorological observations using WRF 3DVar have used the CV5 option for specifying the BEC in WRF. However, WRF 3DVar system supports a modified version of BEC wherein the humidity variable also has multivariate contributions in the balance relations. This newer option named CV6 option has not been used for assimilation of observations for limited area weather models over India.

However, the background error covariances in 3DVar is still isotropic and homogeneous, and hence, the flow dependent "errors of the day" do not get represented in it. Ensemble based data assimilation techniques inherently incorporate flow dependent error structures in assimilation. There are several studies available in the literature which have investigated the impact of assimilating satellite observations using ensemble based data assimilation methods in the WRF model. A brief overview of such studies are given here. However, utilising EnKF based assimilation techniques over India in a regional model context is a largely unexplored topic in the available literature at present.

1.3 Previous studies using EnKF for assimilating satellite observations

Torn and Hakim (2009) investigated the impact of assimilating conventional in situ observations, reconnaissance dropsondes, including data taken during the Hurricane Rainband and Intensity Exchange Experiment (RAINEX) on the simulation of tropical cyclones. Schwartz et al. (2012) showed the positive impact of assimilating microwave radiances with ensemble adjustment Kalman filter (EAKF) in the simulation of the track, intensity, and precipitation forecasts of Typhoon Morakot (2009). The improvement in simulation of tropical cyclones by assimilating radiance observations from the Advanced Microwave Sounding Unit-A (AMSU-A) using the EAKF algorithm is demonstrated by Liu et al. (2012). The study by Jones and Stensrud (2012) assesses the impact of assimilating the temperature and mixing ratio profiles derived from the Atmospheric Infrared Sounder (AIRS) instrument on board the Aqua satellite in WRF model using EAKF algorithm. Jones et al. (2013) explored the potential of assimilating water vapor-sensitive satellite infrared brightness temperatures and Doppler radar reflectivity and radial velocity observations with EnKF technique using an observing system simulation experiment. The impact of assimilating Quick Scatterometer (QuikSCAT) ocean surface wind vectors and surface mesonet observations on the simulation of tropical cyclones was investigated by Zhang and Pu (2014). Wu et al. (2015a) investigated the various aspects of assimilating satellite-derived atmospheric motion vectors (AMVs) using the ensemble adjustment Kalman filter.

The ensemble based assimilation techniques yield superior analyses when compared with 3DVar analyses, in principle. Several studies have compared the two assimilation techniques and arrived at similar conclusions. Meng and Zhang (2008a) and Meng and Zhang (2008b) demonstrated that EnKF generally outperforms 3DVar through the use of flow-dependent error covariances as well as due to its ensemble-based state estimation. Comparison of EnKF and 3DVar assimilation performances in assimilating near-surface observations through observing system simulation experiments by Pu et al. (2013) also showed that EnKF assimilation system outperforms 3DVar in producing more realistic analyses. Dhanya and Chandrasekar (2014) show that using EnKF technique for assimilating Oceansat-2 ocean surface winds in WRF model has improved the simulation of heavy rainfall event over India, when compared with that by 3DVar technique.

The main difference between 3DVar and EnKF assimilation systems is in the way the background error covariances are specified in them. The specification of background error is significant in assimilation since it helps spread the observation information in both spatial and temporal dimensions in the model. However, the model error cannot be exactly calculated because the true state of the system is not known. Furthermore, typically the background error covariance matrix has a large number of elements that cannot be exactly represented in a computing environment. Hence this matrix has to be estimated and several techniques exist for the purpose. Bannister (2008a) and Bannister (2008c) have demonstrated the significance of background error information in a data assimilation system and have reviewed the techniques used in its estimation. 3DVar technique used in WRF model employ the National Meteorological Center (NMC) method (Parrish and Derber (1992)) for estimating the background error. This technique uses statistics obtained by calculating the forecast differences between 24hr and 12hr forecasts of the model for a period of typically one month to estimate the model error. However, this method assumes that the background error is homogeneous and isotropic and hence does not include the error structures in the model associated with the atmospheric flow pattern of the particular day considered. Use of ensembles to estimate the flow-dependent background errors is a method that can overcome this deficiency. The ensembles, in principle represent the probability distribution function (pdf) of the system. However, in practise, the pdf of the system is not fully represented because the pdf is accurate only when the number of ensem-

ble members become infinitely large. Since this is not practically possible to achieve, the background error covariance matrix generated by the ensemble techniques is in general rank-deficient. This results in sampling errors and manifestations of spurious correlations as well as filter divergence in an ensemble based assimilation system (Ehrendorfer (2007)).Hybrid data assimilation systems utilize the ensemble-derived error covariances along with the static background error covariances to provide for an improved analysis than what is possible using 3DVar alone (Wang et al. (2008a,b); Li et al. (2012); Feifei and Jinzhong (2014); Mizzi et al. (2015)).

In a data assimilation system, the background information provided by a short numerical forecast forms an essential component determining the analysis accuracy. Better estimates of BECs help in improving the positive impact of assimilating meteorological observations. Therefore, investigations for improving the existing methods for estimating BECs are important, especially for limited-area weather models. Studies that explore and utilize modifications in the existing formulation of BEC are very much essential for improving the analyses in the case of 3DVar system. However, over the Indian monsoon domain, detailed investigations on the specification of BEC in terms of its different formulations are still lacking. Furthermore, improving the operational analyses through the utilization of flow-dependent error covariances using EnKF based assimilation systems also is extremely important. The generation and maintenance of an ensemble with adequate spread is a difficult problem for a limited area model due to its dependence on boundary conditions. However, the above is still a largely unexplored research area in the data assimilation literature available for the Indian region. The present thesis aims to address this 'gap' in the field of data assimilation research in the country. However, there are some other gap areas in the field of data assimilation research that do not come under the purview of the present study. The currently operational data assimilation systems all assume Gaussianity of the underlying probability distribution function. However, in reality, the distributions are nonlinear due to the inherent nonlinearities in the model. While efforts for incorporating non-Gaussianity in geophysical data assimilation through the use of techiques like particle filters are well-underway (Bocquet et al. (2010),**?**), such research results are still in the nascent stages only. This is a potential area for future research studies.

1.4 Objectives of the thesis

Accurate representation of moisture information in the initial condition provided to a numerical weather model helps to improve the accuracy of its forecast. The above assumes greater significance over the Indian region where the generation and evolution of mesoscale weather features depend heavily on the moist convective processes. Hence assimilation of moisture information from satellite observations is of paramount importance. The present thesis first explores the impact of assimilation of humidity information from MeghaTropiques satellite for simulating three tropical cyclones over the Bay of Bengal. Accuracy of analysis also depends to a great extent on the correct formulation for the specification of background error covariance in a data assimilation system. The effect of multivariate formulation of the moisture control variable in the WRF model's background error formulation is investigated in this study. The thesis explores the impact of assimilating satellite radiances using this newer formulation of background error covariances in the simulation of mesoscale weather features over the Indian domain. Another major objective of the present study is the implementation of ensemble Kalman filter (EnKF) based data assimilation for a mesoscale model and investigating the improved impact of assimilating satellite observations using EnKF vis.a.vis the 3DVar system in WRF model.

The thesis is structured as follows.

Chapter 2 provides an overview of the WRF model used in this study. The various components of the model are detailed. A brief description of the data assimilation methods along with the observations utilized in this thesis are also provided in this chapter. Furthermore, the techniques used for validating the model analysis and forecast fields are also given. The impact of assimilating radiances from MeghaTropiques satellite's SAPHIR sensor in the simulation of three tropical cyclones is investigated in chapter 3. The 3DVar technique is used for assimilation in this chapter. The WRF 3DVar system supports various options to specify the background error covariance matrix. Chapter 4 provides an overview of two of the formulations (cv5 and cv6 options) used and compares them with each other. Impact of assimilating conventional as well as radiance observations from AMSU-A using these two formulations are also investigated. The simulation of three monsoon depressions are considered here to compare the effect of these two formulations. The influence of the above-mentioned

background error formulation on the impact of SAPHIR radiances is investigated in chapter 5 with regard to three tropical cyclone cases. Background error covariances generated using both cv5 and cv6 options in WRF 3DVar are static in nature. Flow-dependent background error covariances can be generated using model forecast ensembles. Chapter 6 explores some of the techniques used for generating ensembles using WRF model. This chapter also investigates the improvement in the simulation of heavy rainfall events over India on using the EnKF assimilation technique as compared with that using the 3DVar technique. Lastly, the chapter 7 provides for a general discussion of the results of this thesis and the main conclusions of the present thesis work are summarized.

CHAPTER 2

Data and Methodology

The present chapter describes the model used in this study along with the assimilation techniques used. The chapter also details the observations utilised for assimilation as well as verification of model forecasts. The chapter concludes with the description of different verification measures used to validate the model forecasts.

2.1 Model Overview

The Weather Research and Forecasting (WRF) model is a community mesoscale non-hydrostatic numerical weather prediction system. It can be used to model the weather phenomena for operational as well as research purposes. It has two dynamical cores - the ARW (Advanced Research WRF) core and the NMM (Nonhydrostatic Mesoscale Model) core. The model supports numerous physics options which can be used in a broad spectrum of applications. The schematic diagram representing the various model components are shown in Fig.2.1

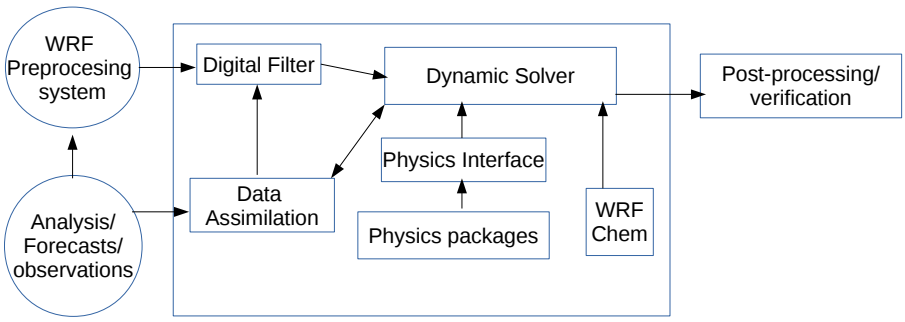

Figure 2.1: WRF model system components(Skamarock et al., 2008)

2.1.1 WRF pre-processing system

The WRF pre-processing system (WPS) consists of three parts. (i) the simulation domain is specified and the various terrestrial data sets are interpolated to this domain. (ii) the various input meteorological fields, typically taken from a global model's analysis or forecast or observation data, are read and re-written to an intermediate format. (iii) the data in the intermediate format are interpolated to the simulation domain. The output of WPS is used to generate initial conditions as well as boundary conditions for solving the model equations.

2.1.2 ARW Dynamical core

The WRF model's dynamic solver uses the Euler equations written in flux form. A terrain following hydrostatic pressure vertical coordinate denoted by η is used in the model and is defined as

$$\eta = (p_h - p_{ht})/\mu \;\; where, \mu = p_{hs} - p_{ht}$$

p_h is the hydrostatic component of pressure. Its value at the top boundary is denoted by p_{ht} while p_{hs} denotes the value at the surface.

The flux form of Euler equations used are given by,

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

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

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

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

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

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

Here, $\mathbf{V} = \mu\mathbf{v} = (U, V, W)$ where $\mathbf{v} = (u, v, w)$, the covariant velocities in the two horizontal and vertical directions, respectively, while $\omega = \dot{\eta}$. θ is the potential

temperature and $\Theta = \mu\theta$. $\Omega = \mu\dot{\eta}$, $\phi = gz$ which is the geopotential and p is the pressure (Skamarock et al., 2008).

The model equations Eqn.2.1-2.6 are discretized using Arakawa C-grid in the horizontal. A time-split integration scheme is used where the meteorologically significant slow modes are integrated using the third order Runge-Kutta scheme and the fast or high frequency acoustic modes are integrated over smaller time steps. The model supports nesting with both one-way and two-way nesting options. A dynamical filter is supported by the WRF model system , which can be used to remove initial model imbalances.

Inclusion of moisture

In the WRF ARW model, the moist Euler equations are formulated by coupling the dry air mass to the prognostic variables and retaining the conservation equation for dry air. The vertical coordinate is defined as,

$$\eta = (p_{dh} - p_{dht})/\mu_d \tag{2.7}$$

where μ_d represents the mass of the dry air in the column and p_{dh} and p_{dht} represent the hydrostatic pressure of the dry atmosphere and the hydrostatic pressure at the top of the dry atmosphere. The coupled variables are defined as

$\mathbf{V} = \mu_d \mathbf{v}$, $\Omega = \mu_d \dot{\eta}$, $\Theta = \mu_d \theta$

And the moist Euler equations are defined as,

$$\partial_t U + (\nabla.\mathbf{V}u) + \mu_d \alpha \partial_x p - \mu_d(\alpha/\alpha_d)\partial_\eta p \partial_x \phi = F_u \tag{2.8}$$

$$\partial_t V + (\nabla.\mathbf{V}u) + \mu_d \alpha \partial_y p - \mu_d(\alpha/\alpha_d)\partial_\eta p \partial_y \phi = F_V \tag{2.9}$$

$$\partial_t W + (\nabla.\mathbf{V}w) - g[(\alpha/\alpha_d)\partial_{eta} p - \mu_d] = F_W \tag{2.10}$$

$$\partial_t \Theta + (\nabla.\mathbf{V}\theta) = F_\Theta \tag{2.11}$$

$$\partial_t \mu_d + (\nabla.V) = 0 \tag{2.12}$$

$$\partial_t \phi + \mu^{-1}[(V.\nabla\phi) - gW] = 0 \tag{2.13}$$

$$\partial_t Q_m + (\nabla.\mathbf{V}q_m) = F_{Qm} \tag{2.14}$$

with the diagnostic equation for dry inverse density given by,

$$\partial_\eta \phi = -\alpha_d \mu_d \tag{2.15}$$

along with the diagnostic relation for the full pressure (vapor plus dry air)

$$p = p_0 (R_d \theta_m / p_0 \alpha_d)^\gamma \tag{2.16}$$

Here, α_d is the inverse density of the dry air and α is the inverse density taking into account the full parcel density. $\alpha = \alpha_d (1 + q_v + q_c + q_r + q_i + ...)^{-1}$ where q_v is the water vapor mixing ratio, q_c is the mixing ratio for cloud, q_r is the mixing ratio of rain and q_i is the mixing ratio for ice.

Furthermore, $\theta_m = \theta(1 + (\frac{R_v}{R_d})q_v)$ and $Q_m = \mu_d q_m$; $q_m = q_v, q_c, q_i ...$

2.1.3 Physics schemes

The WRF model supports simulation of meteorological phenomena over a wide range of space and time-scales. Those physical processes that cannot be explicitly resolved in a numerical model, either due to their small-scale or complexity are represented using various parameterization schemes. Physics schemes for microphysics, cumulus convection, long wave radiation, short wave radiation, boundary layer turbulence, surface layer and land-surface processes are included in the model. The various physics schemes are utilised by the dynamic solver through an interface which calls the required physics packages. The physics packages compute tendencies for the velocity components (un-staggered), potential temperature, and moisture field.

Fig. 2.2 shows a schematic representation of the interactions between various parameterization schemes in the WRF model.

The interactions between the different physics schemes happen via the various model state variables and their tendencies. The tendencies for radiation, surface, planetary boundary layer and cumulus schemes are calculated in the first Runge-Kutta step while the microphysics is computed after the last Runge-Kutta step. The various physics schemes available in WRF model are the following

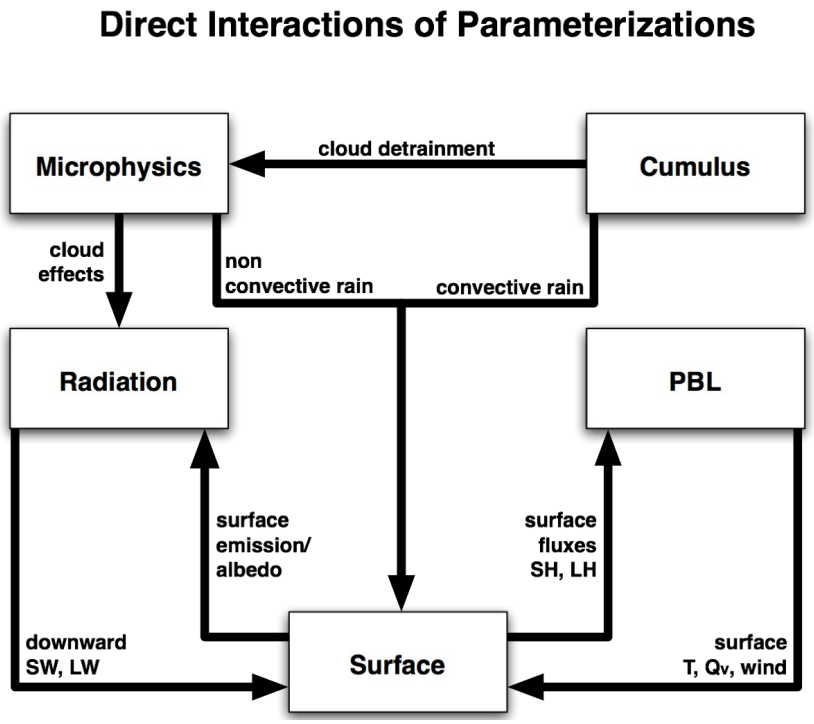

Figure 2.2: Schematic representation of the interactions between various parameterization schemes in WRF model

Microphysics schemes

Microphysics parameterization schemes attempt to account for sub-grid scale updrafts, clouds and precipitation processes. These include processes controlling formation and growth of cloud droplets and ice crystals, as well as their fall out as precipitation. The microphysics scheme updates the atmospheric state at the end of the model time step. WRF model supports many microphysics schemes and the microphysics schemes supported by WRF-ARW model version 3.5 are listed out in the Table 2.1 . These schemes mainly differ from one another in the manner in which the mixing ratios, namely, the cloud water mixing ratio (q_c), rain water mixing ratio (q_r), ice mixing ratio (q_i), snow mixing ratio (q_s), graupel mixing ratio (q_g), total ice mixing ratio (q_t), mixing ratio for hail (q_h) and the various processes are modelled. Table 2.1 also indicates the various mass variables that are handled by each of the schemes of WRF model.

Table 2.1: Microphysics parametrization options available in WRF-ARW model

sl.no.	Scheme	Mass variables
1	Kessler scheme	q_c, q_v, q_r
2	Lin et al. scheme	$q_c, q_r, q_v, q_i, q_s, q_g$
3	WRF Single-Moment 3-class scheme (WSM-3)	$q_c/q_i, q_r/q_s, q_v$
4	WRF Single-Moment 5-class scheme (WSM-5)	q_c, q_r, q_v, q_i, q_s
5	Eta microphysics	q_c, q_r, q_v, q_t
6	WRF Single-Moment 6-class scheme (WSM-6)	$q_c, q_r, q_v, q_i, q_s, q_g$
7	Goddard microphysics scheme	$q_c, q_r, q_v, q_i, q_s, q_g$
8	New Thompson et al. scheme	$q_c, q_r, q_v, q_i, q_s, q_g$
9	Milbrandt-Yau Double-Moment 7-class scheme	$q_c, q_r, q_i, q_s, q_g, q_h$
10	Morrison double-moment scheme	q_c, q_r, q_i, q_s, q_g
11	WRF Double-Moment 5-class scheme	q_c, q_r, q_i, q_s
12	WRF Double-Moment 6-class scheme	q_c, q_r, q_i, q_s, q_g
13	Stony Brook University (Y. Lin) scheme	q_c, q_r, q_i, q_s
14	NSSL 2-moment scheme	$q_c, q_r, q_i, q_s, q_g, q_h$
15	NSSL 2-momemt + CCN scheme	$q_c, q_r, q_i, q_s, q_g, q_h$
16	NSSL 1-moment scheme	$q_c, q_r, q_i, q_s, q_g, q_h$
17	NSSL 1-momlfo scheme	q_c, q_r, q_i, q_s, q_g

Cumulus schemes

Arakawa (2004) defines cumulus parameterization as "the problem of formulating the statistical effects of moist convection to obtain a closed system for predicting weather and climate" generally. In this sense, the cumulus parametrization handles the unresolved components of moist convection in a discrete model where the resolved components influence and are in turn influenced by the unresolved components. Cu-

mulus parametrization schemes in the WRF model represent the effects of sub-grid-scale processes associated with convective and shallow clouds. The vertical fluxes due to updrafts and downdrafts that are unresolved by the model and the associated compensating motion outside the clouds are handled by the cumulus physics schemes. Several cumulus parametrization schemes exist that aim to resolve the sub-grid scale effects of the convective and/or shallow clouds in the model. They also represent vertical fluxes due to unresolved updrafts and downdrafts and compensating motion outside the clouds. The cumulus schemes available in WRF ARW model version 3.5.1 are listed below.

i Kain-Fritsch (KF) Scheme (Kain, 2004): The KF scheme is a mass flux convective scheme. It determines the existence of instability, estimates whether cloud growth is possible via an existing instability and also the properties of any convective clouds that are formed. The KF scheme has three major components , namely , (i) trigger function, (ii) mass flux formulation, and (iii) closure assumption. The trigger function determines the location and time of occurrence of deep convection. A steady state entraining/detraining plume model is used to represent the convective updrafts in the KF scheme. The evaporation of condensate generated within the updraft determines the convective downdrafts. The updraft and downdraft processes determines the mass flux in a convective system in the KF scheme. The closure of KF scheme depends on the convective available potential energy (CAPE) for an entraining air parcel. In KF scheme the mass fluxes are adjusted through updrafts, downdrafts and environmental mass flux so that at least 90% of CAPE is removed.(Kain, 2004).

ii Betts-Miller-Janjic (BMJ) Scheme : The BMJ scheme is a convective adjustment scheme .In such a scheme, the temperature and moisture fields are adjusted (relaxed) towards observed quasi-equilibrium thermodynamic structures. The basic shapes of the empirical quasi-equilibrium thermodynamic profiles are based on observations. The scheme treats both the deep as well as shallow convection. The Betts-Miller scheme (Betts and Miller, 1986), as modified by Janjic (Janjic, 1994) includes a parameter called cloud efficiency which depends on the precipitation, entropy change and the temperature of the cloud. The relaxation time of the scheme is determined using the cloud efficiency (Janjic, 1994).

iii Grell Devenyi ensemble scheme (GD) scheme (Grell and Dévényi, 2002) : The GD scheme is an ensemble based cumulus scheme, in which multiple cumulus schemes with small variants are run within a single grid box. The linear weight average of these ensembles of cumulus scheme is taken and is given as feed back to the model. The framework of this scheme is based on the simple convective parameterization developed by Grell (1993). The WRF model includes a newer scheme called Grell 3D scheme which is an improved version of GD scheme which can be used for high resolution model simulations in addition to coarser resolutions also. The Grell 3D scheme provides options to spread subsidence to neighboring grid points also which makes it suitable for smaller grid sizes where the subsidence may not occur within the same grid column as the updraft. Furthermore, the WRF model includes another variant of the GD scheme, called the Grell-Freitas (GF) scheme (Grell and Freitas, 2014) which provides for cloud-aerosol interactions also, and can be useful for nearly cloud-scale parametrizations. (Grell and Dévényi, 2002)

iv Simplified Arakawa-Schubert (SAS) scheme (Pan and Wu, 1995): This is based on Arakawa and Schubert (1974) scheme with the simplifications by Grell (1993) and with a saturated downdraft. In this scheme, the atmospheric temperature and moisture fields are adjusted using a mass flux concept. The simplified scheme assumes one cloud type with detrainment only from its top. Modified version of the SAS scheme that accommodates shallow convection by modifying cloud-base mass flux, entrainment, and detrainment specifications (Han and Pan, 2011) is also available in the WRF model.

v Tiedtke scheme (Tiedtke, 1989) : This scheme also uses a mass-flux approach. An ensemble of clouds is represented using one-dimensional bulk model here. Various types of convection are considered and the penetrative and midlevel convection are assumed to be maintained by large-scale moisture convergence while the shallow convection is maintained by the supply of moisture through surface evaporation.

vi Zhang-McFarlane scheme (Zhang and McFarlane, 1995) : This is also a mass-flux based scheme which uses a closure condition that CAPE is consumed at an exponential rate by cumulus convection with characteristic time scale. This is a cumulus scheme suitable for use in general circulation models.

Atmospheric radiation

The radiation schemes provides for the total radiative flux at any given location. The surface downward longwave (LW) and shortwave (SW) radiation as well as atmospheric heating due to radiative flux divergence are taken into account by the radiation parameterization schemes. Shortwave radiation encompasses visible and surrounding wavelengths and takes into account, the processes like absorption, reflection, and scattering in the atmosphere and at surfaces. The longwave radiation accounts for the infrared and thermal radiation absorbed and emitted by the atmospheric trace gases and the surface.

The most commonly used radiation parametrization in WRF model is the 'Rapid Radiative Transfer Model (RRTM)'. RRTM is a spectral-band scheme which calculates the fluxes and cooling rates for the long wave spectral region (10-300 cm^1) for an arbitrary clear atmosphere. The fluxes are calculated using a correlated-k method from the line-by-line radiative transfer model (LBLRTM). This model takes into account water vapor, carbon dioxide, ozone, methane, nitrous oxide and the common halocarbon (Mlawer et al., 1997)

The Dudhia scheme is most commonly employed for shortwave radiation parametrization in the WRF model. This scheme integrates the downward solar flux and accounts for clear sky scattering, water vapor absorption and cloud albedo and absorption (Dudhia, 1989).

Planetary Boundary Layer

The planetary boundary layer (PBL) scheme provides for the vertical sub-grid-scale fluxes due to eddy transports in the whole atmospheric column. The PBL schemes provide atmospheric tendencies of temperature, moisture (including clouds), and horizontal momentum in the entire atmospheric column by determining the flux profiles within the well-mixed boundary layer and the stable layer. The representation of lower-tropospheric thermodynamic and kinematic structures via the PBL parametrization helps in improving the mesoscale weather forecasts. The parametrization of unresolved fluxes of temperature, momentum, and moisture in the PBL requires a closure scheme for obtaining turbulent fluxes from the mean quantities. In schemes

following local closure, the turbulent fluxes at each grid points is estimated from the atmospheric variables and/or their gradients at that grid point. Nonlocal fluxes which are transported over large distances through eddies in convective conditions are modeled using non-local closure PBL schemes. The most popular among the various PBL schemes available with the WRF model are the Yonsei University (YSU) scheme and the Mellor-Yamada-Janjic (MYJ) scheme.

The YSU scheme (Hong et al. (2006)) uses a non-local closure approach whereas MYJ scheme (Janjic (1994)) is typically a local closure scheme. YSU scheme is a modified version of medium range forecast (MRF) scheme (Hong and Pan, 1996). The MRF scheme is modified by treating the entrainment processes at the top of the PBL explicitly. This is done through the addition of an asymptotic entrainment flux term to the turbulence diffusion equation. MYJ scheme is an implementation of Mellor Yamada level 2.5 model. In the MYJ scheme, the eddy diffusion coefficients are determined from prognostically calculated turbulent kinetic energy (TKE). An upper limit to the master length scale is imposed using TKE as well as buoyancy and shear of the driving flow. In the unstable range, the requirement that the TKE production be non-singular in the growing turbulence case is used to derive the function for the upper limit. In the stable regime, the condition that the ratio of the variance of the vertical velocity deviation and TKE cannot be smaller than that corresponding to the regime of vanishing turbulence is used to impose the upper limits.

Land Surface Model

The land surface models (LSM) provide heat and moisture fluxes over land points and sea-ice points. This is done by utilising the atmospheric information from the surface layer scheme, radiative forcing from the radiation scheme, and precipitation forcing from the microphysics and convective schemes as well as internal information on the land's state variables and land-surface properties. Currently, in the WRF model all the available LSMs are one dimensional and these do not provide for interaction between neighboring grid points. The thermal diffusion scheme (TD) is a simple 5 layer soil temperature model and the layer and the soil layers are centered at 1, 2, 4, 8, and 16 cm respectively from the top to the bottom. The temperature at the bottom layer and below is fixed at a deep layer average. In this scheme, the soil moisture is fixed

with land use and a season dependent constant and this model not represent adequately the vegetation processes. The unified NOAH LSM has 4 soil layers centered at 10, 30, 60, 100 cm from the top to the bottom. The major advantage of NOAH LSM is that it is consistent with time dependent soil fields provided by analysis datasets. NOAH LSM takes into account, root zone, evaporation, soil drainage, run off and the vegetation processes. Despite the availability of other options for LSM in the WRF model, the above mentioned schemes are those that are commonly used for WRF model simulations over the Indian region. Table 2.2 lists all available the physics options with the WRF model. For brevity, detailed discussion of the other physics options are not provided.

The calculation of surface heat and moisture fluxes by the land-surface models as well as the calculation of surface stress by the planetary boundary layer scheme are utilized in the surface layer scheme in the model that provides for the friction velocities and exchange coefficients. The surface layer scheme provides stability-dependent information about the surface layer and not the tendencies and hence every surface layer option in the model is linked an associated boundary-layer scheme option.

Furthermore, the WRF model system supports the simulation of emission, transport, mixing, and chemical transformation of trace gases and aerosols through the WRF-Chem which is the WRF atmospheric chemistry model. The WRF-Chem model is used for investigation of regional-scale air quality, field program analysis, and cloud-scale interactions between clouds and chemistry. The WRF-Chem component is not utilized in the present thesis study.

Table 2.2: List of Physics parametrization schemes available in WRF model

Cumulus	**S.W Radation**	**LW Radiation**	**PBL**	**LSM**
Kain-Fritsch	Dudhia	RRTM	YSU	5-layer thermal diffusion
Betts-Miller-Janjic	Goddard	CAM	MYJ	NOAH LSM
Grell-Freitas	CAM	RRTMG	GFS	RUC LSM
Old Simplied ArakawaSchubert	RRTMG	New Goddard	QNSE	Pleim-Xiu LSM
Grell-3	New Goddard	FLG	MYNN2	NOAH-MP LSM
Tiedtke	FLG	Held-Suarez	MYNN3	SSiB LSM
Zhang-McFarlane	GFDL	GFDL	ACM2	CLM4
New SAS			BouLac	
Grell-Devenyi			UW	
Old Kain-Fritsch			TEMF	
			GBM	
			MRF	

2.2 Data assimilation techniques

The fundamental problem of data assimilation is to deduce the best possible estimate (called the analysis) of the atmospheric state given the observations and background information along with their respective error information.

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

Consider a vector of observations $\mathbf{y} \in \mathbb{R}^m$. An observation operator $H(\mathbf{x})$ provides a mapping from the model's state space to the observation space. In general the observation operator can be non-linear also. $\mathbf{y} = H(\mathbf{x}^t) + \epsilon^o$ where ϵ^o denotes the error in observations. It is assumed that the observations are unbiased ($\langle \epsilon^o \rangle = 0$) and that the observational and background errors are uncorrelated.The observational error covariance is given by $\mathbf{R} = \langle \epsilon^o \epsilon^{oT} \rangle \in \mathbb{R}^{m \times m}$.

There are several data assimilation techniques available for estimating the analysis. In this thesis study, two assimilation techniques namely, the three dimensional variational (3DVar) technique and Ensemble Kalman Filter (EnKF) technique are utilized.

2.2.1 Three dimensional variational technique

In the three dimensional variational (3DVar) technique, the best analysis is obtained by minimising a cost function given by,

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^b) + (\mathbf{y} - H(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y} - H(\mathbf{x})) \tag{2.17}$$

Eqn.4.1 is minimized by assigning the gradient of cost function to zero. ie, $\nabla_x J(x) = 0$ where,

$$\nabla_{\mathbf{x}} J(\mathbf{x}) = \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) - \mathbf{H}^T \mathbf{R}^{-1} (\mathbf{y} - H(\mathbf{x})) \tag{2.18}$$

Here, $\mathbf{H}$ is the Jacobian of H. Linearising H around the background $\mathbf{x}^b$ using

Taylor expansion, substituting in Eqn.2.18 and equating to zero and rearranging, we get the analysis,

$$\mathbf{x}^a = \mathbf{x}^b + (\mathbf{B}^{-1} + \mathbf{H^T R^{-1} H}^{-1})\mathbf{H^T R^{-1}}(\mathbf{y} - H(\mathbf{x})) \tag{2.19}$$

2.2.2 Ensemble Kalman Filter

The Ensemble Kalman filter (EnKF) is an improvement over the original Kalman filter algorithm where the probability distribution function (pdf) of the prior state is represented via an ensemble of states $\mathbf{X} = [\mathbf{x_1}, \mathbf{x_1}, ...\mathbf{x_N}]$ where $\mathbf{x_1}, \mathbf{x_1}, ...\mathbf{x_N}$ represent N ensemble members , each with a state space of dimension n. Hence, $\mathbf{X}$ is an $n \times N$ matrix whose columns are the ensemble members.

In the original formulation of EnKF (Evensen, 2003; Burgers et al., 1998), the observation vector is also perturbed, generating an observation ensemble $D = [\mathbf{d_1}, \mathbf{d_2}...\mathbf{d_N}$. Here, each perturbed observation vector $\mathbf{d}$ is obtained by adding a random vector from the m-dimensional normal distribution $N(0, R)$ to the observation state vector $\mathbf{y}$.

The EnKF algorithm defines the analysis step consisting of update to each model state $\mathbf{x_i}$ as given by,

$$\mathbf{x_i}^a = \mathbf{x_i}^b + \mathbf{K}(\mathbf{d_i} - H\mathbf{x_i}) \tag{2.20}$$

This results in an ensemble of analysis states.

The Kalman gain $\mathbf{K}$ is defined as

$$\mathbf{K} = P^f H^T (H P^f H^T + \mathbf{R})^{-1} \tag{2.21}$$

Here, P^f is the model forecast error covariance $P^f = \overline{(\mathbf{x} - (\overline{\mathbf{x}}))(\mathbf{x} - (\overline{\mathbf{x}})^T)}$. The overbar denotes an expectation value.

The mean of the analysis ensembles is taken as the best estimate.

This analysis step is followed by the forecast step, where the analysed states are integrated through the model equations, like in a standard Kalman filter.

2.3 Data Used

2.3.1 Data for initialising the WRF model

The WRF model, being a limited area model, requires meteorological fields for initial and boundary conditions. Typically, analyses and forecast fields from a global model is utilised for the purpose. Here, the forecast fields from National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) is used to provide the initial and boundary conditions for the WRF model. GFS is a coupled global spectral model consisting of an atmosphere model, an ocean model, a land/soil model, and a sea ice model. Model output obtained at a horizontal resolution of 0.5° and available every 6 hours are used in this study.

2.3.2 Observations used

Conventional surface and upper air observations

The observations available from NCEP ADP global upper air and surface weather observations are utilized for assimilation as well as verification here. These include land surface, marine surface, radiosonde, pibal and aircraft reports together with satellite wind data. These data are available in PREPBUFR format from http://rda.ucar.edu/datasets/ds337.0/.

MeghaTropiques SAPHIR radiances

The MeghaTropiques satellite, is an Indo-French joint satellite mission and carries the following four instruments: Microwave Analysis and Detection of Rain and Atmospheric Structure (MADRAS), a microwave imager operating in the frequency range from 18 to 157 GHz for measuring rain, atmospheric water vapor content, liquid water content, and ocean surface wind speed; SAPHIR (Sounder for Probing Vertical Profiles of Humidity) is a six-channel microwave sounder operating at 183 GHz measuring vertical profiles of atmospheric humidity over land and ocean; ScaRab (Scanner for Radiation Budget), operating in the optical region for estimating Earth radiation budget over tropical convective region and Radio Occultation Sensor for Vertical Pro-

filing of Temperature and Humidity.

The chief objective of the MeghaTropiques satellite mission is to investigate the water cycle and the tropical convection life cycle. The satellite was launched on 12th October 2011 and has been placed in a circular orbit at 800 km with a 20° inclination, covering the tropical belt between 23° N and 23°S. One of the important features of the above satellite is that it is able to observe tropical region upto five times a day. SAPHIR instrument aboard the MeghaTropiques satellite consists of cross-track, scanning microwave radiometer for atmospheric moisture sounding. The satellite has six channels around the 183.31 GHz water vapor absorption line. It is expected that by utilizing SAPHIR radiance, an improved water vapour vertical profile can be deduced. The satellite has a footprint of 10 km horizontal resolution at the nadir (Aguttes et al., 2000; Brogniez et al., 2013; Desbois et al., 2003).

AMSU Radiances

The Advanced microwave sounding unit (AMSU) is a multi-channel microwave radiometer. AMSU-A is flown aboard the NOAA satellites (NOAA-15, 16,17,18) as well as the NASA Aqua Earth science satellite and the EUMETSAT MetOp series, all polar-orbiting satellites in sun-synchronous orbits. AMSU-A has 15 channels between 23.8 and 89 GHz, and is used primarily for measuring atmospheric temperatures . AMSU has a ground resolution near nadir of 45 km.

Ocean surface winds from Oceansat-2 scatterometer

Oceansat-2 is an Indian Space Research Organisation's (ISRO) satellite which is dedicated to ocean research and was launched in September 2009. It carried a 13.5-GHz Ku-band microwave scatterometer (OSCAT), in addition to an Ocean Colour Monitor and a Radio Occultation Sounder for Atmospheric Studies (ROSA). OSCAT has a swath of 1800 km with a ground resolution cell of size 50 $\times$ 50 km and was designed to retrieve the horizontal wind speed and direction over the ocean surface. The instrument provides wind speed measurements of 4-24ms^{-1} with an accuracy of $2ms^{-1}$ and direction with an accuracy of 20^0. Oceansat-2 satellite provided ocean surface data during November 2009 to February 2014 after which it became dysfunc-

tional.

Rainfall observations from TRMM

The Tropical Rainfall Measuring Mission (TRMM) a joint mission between NASA and the Japan Aerospace Exploration (JAXA) Agency to study rainfall for weather and climate research, was launched in November 1997 and provided data till April 15, 2015. TRMM carried the following instruments: a 3-sensor rainfall suite (precipitation radar, TRMM microwave imager, Visible and infrared scanner) and two related instruments (Clouds and the Earth's Radiant Energy Sensor and Lightning Imaging Sensor). The Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) has provided combined precipitation estimates from multiple satellites, as well as gauge analyses where feasible with a horizontal resolution of $0.25^o \times 0.25^o$ (Adler et al., 1993).

2.4 Validation methodology

Quantitative verification of model forecasts are performed using various commonly employed verification measures in this thesis study. The commonly employed verification measures used in this thesis study are given below.

(i) Root mean square error (RMSE) :- the RMSE for a predicted quantity x_f with respect to its observed value x_o value is given by

$$RMSE(x) = \sqrt{\frac{1}{N}\sum(x_f - x_o)^2} \tag{2.22}$$

where N is the total number of observations.

(ii) Forecast impact parameter (FI) (Wilks, 2011) for any model parameter is defined based on its RMSE and is given by,

$$FI = \left(1 - \frac{RMSE_E}{RMSE_C}\right) \times 100 \tag{2.23}$$

where $RMSE_E$ refers to the RMSE of the assimilation experiment, $RMSE_C$ refers

Table 2.3: Contingency table utilized for calculating skill scores for model rainfall with respect to TRMM rainfall observations

TRMM Rainfall			
Model forecast		***Yes***	***No***
	Yes	a	b
	No	c	d

to the RMSE of the control (no-assimilation) experiment.

(iii) Improvement parameter : A positive η implies that the difference between a control forecast of a model parameter and the observation is more than that between the parameter forecast in an assimilation experiment and the observation. A positive value of the improvement parameter indicates that the assimilation experiment has forecasted a value closer to observations than the control forecast.

$$\eta = |observation - control| - |observation - experiment| \tag{2.24}$$

(iv)Variou skill scores for the rainfall forecast: For quantitative verification of the rainfall forecast skill, various skill scores like equitable threat score (ETS) , Bias score , False Alarm Ratio (FAR) and Probability of Detection (POD) are utilized. The above mentioned skill scores are calculated using the contingency Table 2.3 considering whether a forecast occurs (yes) or not (no) (Wilks, 2011). While 'a' is called a hit (both model and observations have predicted a rainfall event of a certain threshold), the situation 'c' is called a 'miss'. Situation 'b' is called as 'false alarm' (model simulates rainfall event of a certain threshold that is not observed), and 'd' refers to a situation where both model and observation agree on a non-rainfall event.

ETS estimates how well the observed event is forecast, discounting the situation where the correct rainfall forecasts occur due to pure chance.

$$ETS = \frac{(a - a_r)}{(a + b + c - a_r)} \tag{2.25}$$

where a_r represents the expected number of correct forecast in a random forecast

and is defined as

$$a_r = \frac{((a+b)(a+c))}{(a+b+c+d)} \tag{2.26}$$

Bias score estimates the ratio of frequency of forecast events to the frequency of observed events, indicating whether there is over or under prediction by the model. The FAR gives the fraction of false alarms (model simulated rainfall that is not observed) and POD gives the fraction of correctly forecast events.

$$Bias = \frac{a+b}{a+c} \tag{2.27}$$

POD gives the fraction of correctly forecast events defined as,

$$POD = \frac{a}{a+c} \tag{2.28}$$

False alarm ratio (FAR) is used in conjunction with POD and is defined as,

$$FAR = \frac{b}{a+b} \tag{2.29}$$

CHAPTER 3

Impact of 3DVar assimilation of MeghaTropiques SAPHIR radiances on the simulation of tropical cyclones over Bay of Bengal

3.1 Introduction

Accurate prediction of track and intensity changes of tropical cyclones is essential to minimize the destruction and damage it causes to millions of people. Since tropical cyclones form and develop over the oceans, observations over oceans become crucial for their accurate numerical simulation. However, oceans are sparsely observed by conventional observing systems. This chapter explores the impact of assimilating MeghaTropiques SAPHIR radiances in simulating three tropical cyclones over the Bay of Bengal using the WRF model.

Many previous studies like (Langland et al., 2009; Chen, 2007; Singh et al., 2008b; Srinivas et al., 2012b; Wu et al., 2014, 2015b; Yesubabu et al., 2014b) etc have shown that assimilating satellite observations did improve the simulation of tropical cyclone features.

There are two distinct ways in which microwave satellite observations are assimilated in a model. The first is to use the satellite-observed radiances directly in the assimilation system by utilising a radiative transfer model (RTM) to simulate the radiance values corresponding to the model variables. The second method is to utilise a retrieval algorithm to invert the observed radiances to atmospheric variables, the latter can then be directly assimilated in the model without the need for an RTM. However, the second method has some inherent drawbacks since, the inversion algorithm entails approximations, which results in errors. Furthermore, the assumption of non-correlated observation errors is more true in assimilating the radiance observations directly (Derber and Wu, 1998; McNally et al., 2000; Eyre et al., 1993). Hence, it is better to assimilate radiance observations directly in a numerical model.

Unlike in the case of infra-red and visible frequency radiations, clouds and aerosols are essentially transparent for the microwave frequencies and allow the latter to pass through. Hence, microwave satellite sensors are vital in providing observations over tropical cyclones. Singh et al. (2013) have assessed the quality of radiances measured by SAPHIR and have assimilated these observations into the Weather Research and Forecasting (WRF) model. They showed that the assimilation of SAPHIR radiances resulted in improvements in the simulation of fields such as moisture, temperature, winds and precipitation etc. The present study aims to investigate the impact of assimilating SAPHIR observations in the simulation of tropical cyclones over the Bay of Bengal. The genesis and life cycle of a tropical cyclone is significantly dependent on the convective processes over oceans. The inclusion of humidity information over the oceanic regions can, hence, significantly impact on the simulation of a cyclone by a numerical model. The Advanced Research Weather Research and Forecasting (WRF ARW) model and its three dimensional variational (3DVar) assimilation technique have been utilised in this study. Three tropical cyclones that formed over the Bay of Bengal have been considered in this study to investigate on the impact of atmospheric humidity measurements from SAPHIR on the model simulation of the tropical cyclones.

The chapter describes the numerical experiments performed, the cases investigated and discussion on the impact of SAPHIR radiance assimilation in simulating the tropical cyclone cases that are investigated.

3.2 Model configuration

Fig.3.1 shows the model domain used in this study. The WRF ARW model is configured in a two-way nested domain. The outer domain has 200 grid cells in the east-west direction and 170 grid cells in the north-south direction with a horizontal resolution of 27 km. The inner domain has a horizontal resolution of 9 km with 325$\times$ 274 grid cells in the east-west and north-south directions respectively. Both domains have 30 levels in the vertical. Both the domains use Kain-Fritsch cumulus scheme with Yonsei University Scheme for planetary boundary layer, RRTM scheme for long wave radiation , Dudhia scheme for short wave radiation, Kessler scheme for microphysics

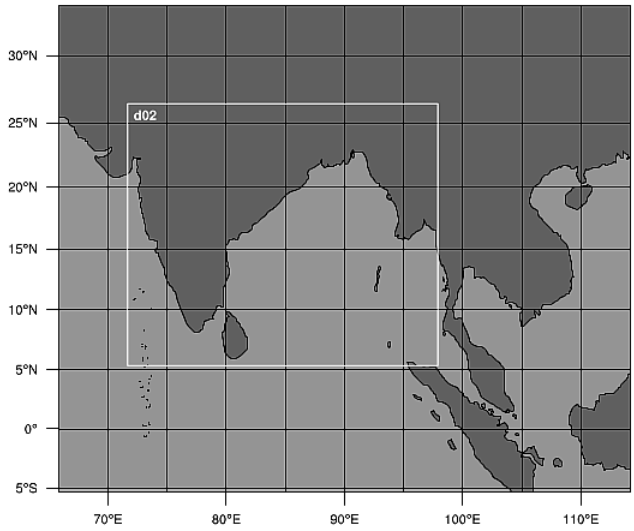

Figure 3.1: Model Domain used in the experiment

and NOAH land surface model for land surface parameterisation.

3.3 Observations assimilated

The global surface and upper air observations available from NCEP include land surface, marine surface, radiosonde, pibal and aircraft reports from the Global Telecommunications System (GTS), profiler and US radar derived winds, SSM/I oceanic winds and TCW retrievals, and satellite wind data from the National Environmental Satellite Data and Information Service (NESDIS) (NCEP, 2008). These observations are assimilated in this study along with the radiance observations from SAPHIR sensor of MeghaTropiques satellite. Fig.5.2 shows the distribution of observations that are typically available over the domain, used for assimilation in this study.

3.4 Assimilation Methodology

The available observations are assimilated using the three dimensional variational (3DVar) technique here. The background error covariance ($\mathbf{B}$) associated with the model forecasts over the experimental domain has been estimated using the NMC method which approximates the background error using model forecasts which have different lead times (24 hours and 12 hours in the regional) valid at the same time. The

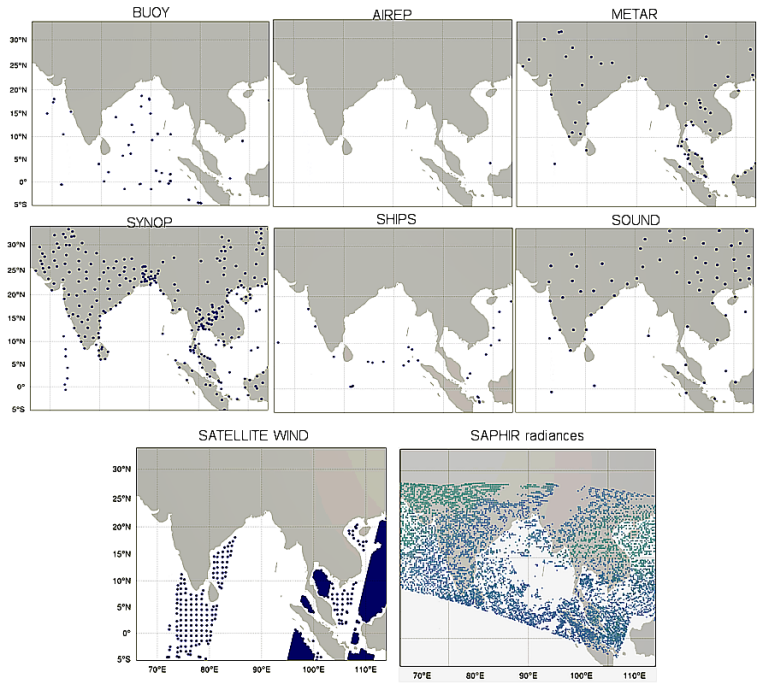

Figure 3.2: Distribution of typically available observations over the domain used for assimilation

statistics for the above calculation are obtained from model simulations over a month long period.

The WRFDA system, by default, supports the assimilation of radiances from several satellite sensors directly. However, the assimilation of observations from the SAPHIR sensor aboard the MeghaTropiques satellite is not supported as such in the WRF model. Hence, modifications are made in the programming modules associated with assimilating observations from the new sensor (SAPHIR) in the WRF 3DVar system.

Direct assimilation of satellite radiances in a numerical weather model requires a radiative transfer model (RTM) which simulates the model-equivalent brightness temperatures. WRF varitaional assimilation system supports the use of two RTMs , namely the (i) Radiative Transfer for Television and Infrared Observation Satellite (TOVS; RTTOV) developed and maintained by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT), and the U.S. Joint Center for Satellite Data Assimilation (JCSDA) CRTM (Barker et al., 2012) and (ii) Community Radiative Transfer Model (CRTM) developed at the Joint Center for Satellite Data Assimilation (JCSDA). The RTTOV model (Saunders et al., 2005) is used in this study

for simulating radiances from model variables.

The observations are subjected to quality control check before assimilation. Meteorological observations available from NCEP in prepbufr format are already quality-controlled for assimilation into the various NCEP analyses (Keyser, 2013). Furthermore, a gross quality check is performed by WRF 3DVar system and observations that differ from the model's first guess by more than five times the observation error standard deviation are excluded from assimilation. Quality check has also been performed for the MeghaTropiques SAPHIR radiances. The precipitation affected radiance observations have not been considered in this study using the method used by Singh et al. (2013). In this study, the observation errors for SAPHIR radiances are assumed to be 3K for all the channels. In addition, the observations from channel S6 only over land are used for assimilation as the observations from this channel have impact from land surface. Since the observational errors are assumed to be uncorrelated in the horizontal, observation thinning is also performed on SAPHIR radiances.

In addition to quality check, the radiance observations are also required to undergo a bias correction since the radiance measurements and the RTMs are prone to contain biases (systematic errors). The biases in the observations depend on platform, instrument, channel, scan angle, and atmospheric conditions. The bias correction coefficients can be estimated either offline (Harris and Kelly, 2001) or using a variational bias correction (VarBC; (Dee, 2005), (Auligné et al., 2007)). This study uses the VarBC scheme for removing the SAPHIR radiance biases.

Cyclone Cases Investigated

3.5 Cyclone Thane (25-31 December, 2011)

3.5.1 Synoptic Conditions

The very severe cyclonic storm Thane formed as a tropical disturbance over the southeast Bay of Bengal on 24^{th} December 2011. The above system concentrated into a depression on 25^{th} December 2011 and lay over 8.5°N, 88.5°E. Moving northeast,

the depression strengthened into deep depression by 00Z 26 December 2011. Further moving northwards, the system became a tropical cyclone by 18Z of 26^{th} December 2011 and lay centred near 11°N, 87.5°E. The cyclone moved north-westwards over the southeast Bay of Bengal and lay over 12°N, 87°E on 27 December 2011. It then moved westwards and intensified into a severe cyclonic storm by 28^{th} December 2011. By 12Z of the same day, the system further intensified into a very severe cyclonic storm. Further moving west, the very severe cyclonic storm crossed north Tamil Nadu coast around 00Z 30^{th} December 2011. The system moved westwards and weakened into a severe cyclonic storm by 03Z of the same day. Moving further westwards , the system further weakened into a deep depression by 06Z of 30th December 2011. Further moving westsouthwestwards, it further weakened and dissipated into a low pressure area by 31st December 2011.

3.5.2 Experimental Details

The WRF model is configured in the manner that is described in section 3.2. The model is initialized using forecast fields from GFS model of $0.5^0 \times 0.5^0$ horizontal resolution. From the initial conditions valid at 18Z of 24^{th} December 2011, WRF model is integrated for a period of 6 hours for model spin up. Thereafter, two numerical simulation experiments are performed. (i) ctrl run - A control run where the surface and upper air observations together with satellite winds from NCEP are assimilated at 00Z, 06Z, 12Z and 18Z of 25^{th} December 2011. From the analysis at 18Z of 25^{th} December 2011 a free forecast is performed without any further assimilation of observations till the cyclone experienced landfall on 30^{th} December 2011 (ii) saph run - Here, along with NCEP surface, upper air and satellite wind observations, radiance observations from SAPHIR are also assimilated at 00Z, 06Z, 12Z and 18Z of 25^{th} December 2011. A free forecast run is performed from 18Z 25 December 2011 to 00Z 30^{th} December 2011 without any assimilation of observations during the free forecast period.

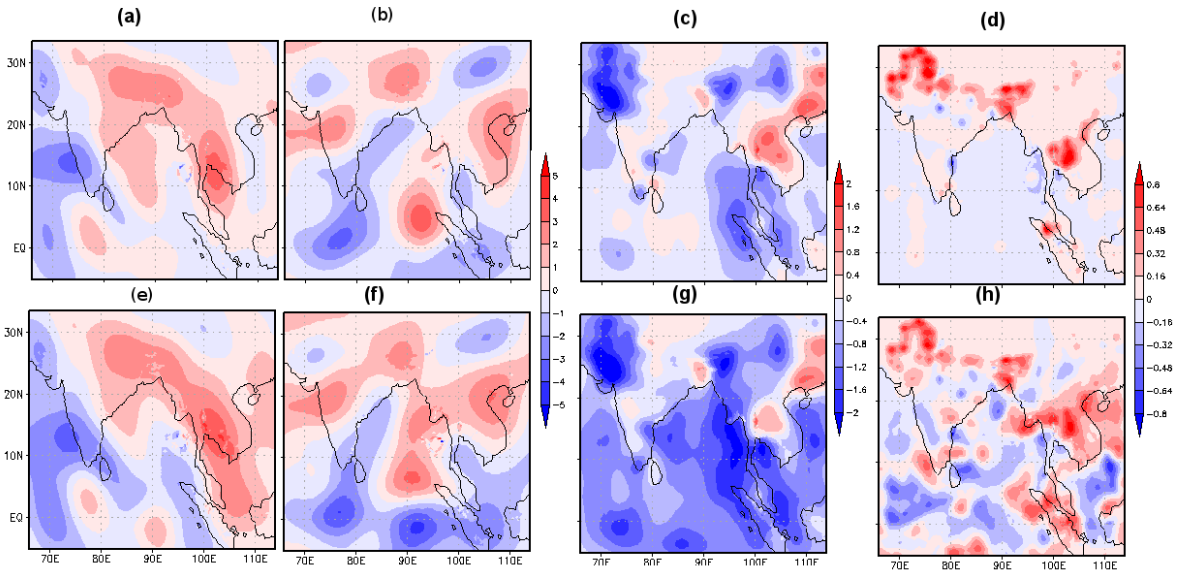

Figure 3.3: Analysis increment for ctrl (a-d) and saph runs (e-h) for zonal wind (a,e), meridional wind (b,f), temperature (c,g) and water vapor mixing ratio (d,h) at model level 1 for the cyclone Thane

3.5.3 Results and discussion

Fig.3.3 indicates the analysis increment for ctrl (a-d) and saph runs (e-h) for zonal (u) wind (a,e) in ms^{-1}, meridional (v) wind (b,f) in ms^{-1}, temperature (c,g) in K and water vapor mixing ratio (q; d,h) in gkg^{-1} at model level 1 for the cyclone Thane. It is clear that the SAPHIR radiances do have influence on the analysis fields. In comparison with the ctrl analysis, the assimilation of SAPHIR humidity information has resulted in a small reduction in u-wind over the Bay of Bengal region while the temperature at model level 1 has been slightly lowered due to SAPHIR assimilation. The assimilation of radiances from SAPHIR has also modified the analysis field of q over relatively larger number of grid points as compared with the ctrl run.

Mean sea level pressure and wind

Fig.3.4 depicts the mean sea level pressure (slp) and lower tropospheric wind at 850hPa at (+00hr forecast), 24 hour forecast (+24hr) and at 48 hour forecast (+48hr) simulated by the ctrl and saph runs. From Fig.3.4(a-c) it can be seen that the initial position and structure of the slp and wind fields in saph runs are relatively closer

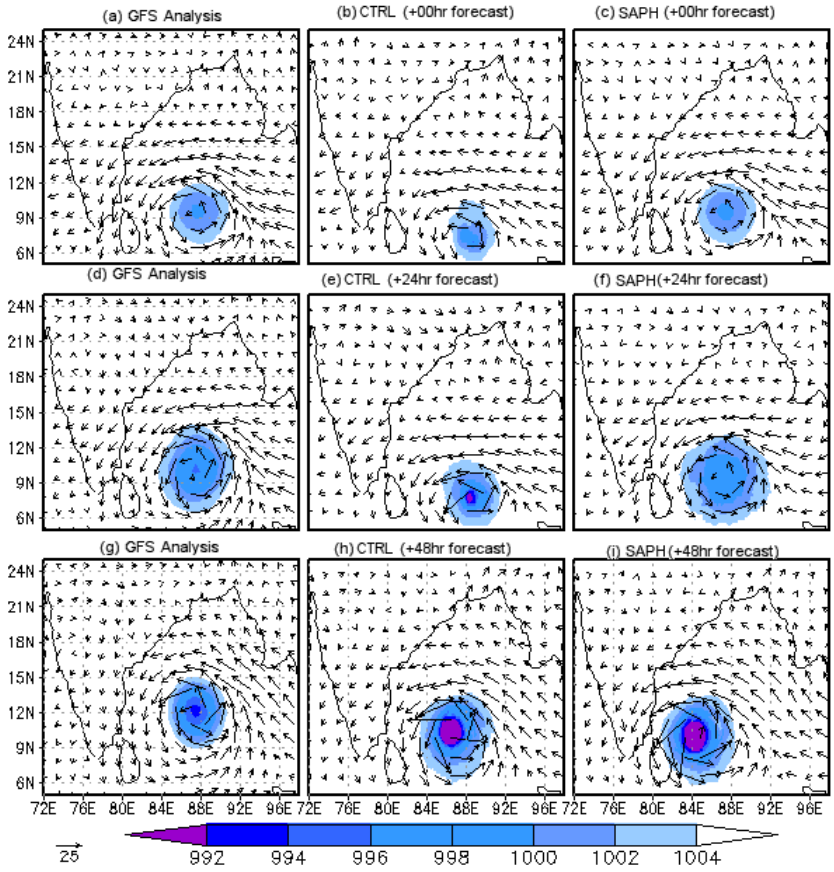

Figure 3.4: Mean sea level pressure and 850hPa wind from GFS analysis (a,d,g), CTRL run (b,e,h) and 3DVar run (c,f,i) at +00hr, +24hr and +48hr forecasts for cyclone Thane

to the GFS analysis as compared with the ctrl simulation. After 24 hours of model integration, the cyclonic system is seen to have intensified more in the ctrl run as compared with the GFS analysis. However, the assimilation of SAPHIR radiances in the saph run has resulted in the system reaching an intensity comparable with GFS run in the 24hr simulation as seen in Fig.3.4(d). The 48 hour model forecast of slp and wind (Fig.3.4(h-i)) as compared with the GFS analysis shown in Fig.3.4(g) indicates that the model simulations have resulted in greater intensification of the cyclone than seen in the GFS analysis. There are errors in the simulation of both position and intensity of the cyclone by both ctrl and saph runs, with the saph runs showing the system relatively closer to land than seen in ctrl run as well as GFS analysis.

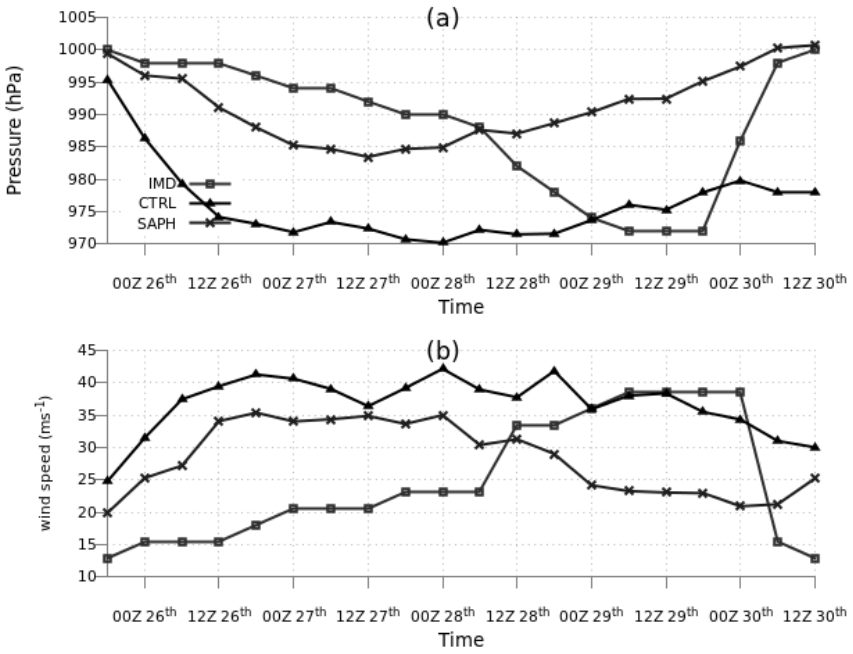

Figure 3.5: Time series of minimum sea level pressure in hPa (a) and maximum wind speed in ms^{-1} (b) for cyclone thane

Fig.3.5a shows the time series of minimum sea level pressure simulated by the CTRL and 3DVar runs as compared with the IMD observations for cyclone Thane during the free forecast period. The initial value of minimum slp in saph run is closer to that observed by IMD observations. The ctrl run has simulated a more intense system at the initial forecast hours. Furthermore, the observed intensification of the cyclone is not faithfully reproduced by the saph run after 42 hours of model integration (ie. after 12Z 27 December 2011). The time series of maximum wind speed simulated by the CTRL and 3DVar runs along with the observations from IMD are shown in

Fig.3.5b. As seen in the case of simulation of minimum slp, the maximum wind speeds simulated by the model in the saph run is closer to observation than that simulated in the ctrl run during initial forecast hours. Both the ctrl and saph runs have simulated stronger winds than the observed. The observed winds strengthen especially after 00Z 28December 2011, corresponding to the intensification of the system. However, the cyclone simulated by the models reveal no intensification of wind speed, after this time. This may be because of the fact that assimilation of SAPHIR observations is not attempted over the entire duration of the simulation period. The model errors do manifest during the free forecast period and this may cause increase in the errors in the cyclone simulation. Assimilating observations throughout the life cycle of the cyclone system may possibly contribute to simulating the cyclonic features more accurately by the WRF model.

Vertical profiles of average temperature anomaly and relative vorticity

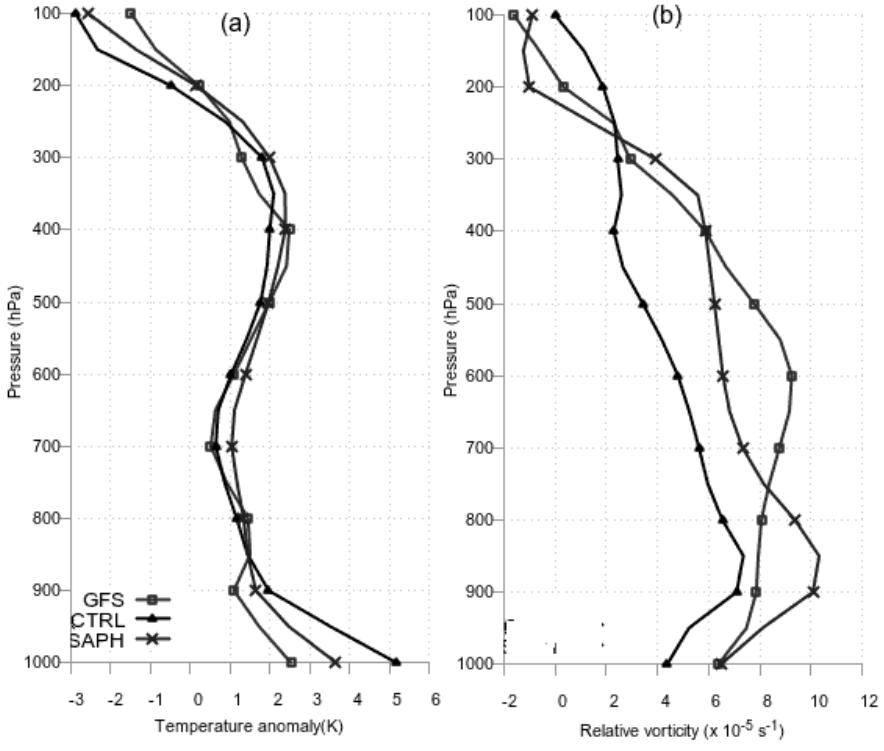

Figure 3.6: Area averaged and time averaged temperature anomaly (a) and relative vorticity (b) in a $3^\circ \times 3^\circ$ box around the cyclone centre for cyclone Thane

To investigate the impact of SAPHIR radiance assimilation on the vertical thermo-

dynamic structure of the tropical cyclones, an area averaged and time averaged temperature anomaly is analysed. For this, a $3^\circ \times 3^\circ$ box is considered around the cyclone centre. The temperature anomaly is defined as the difference of the average temperature in this box and that from the outside environment. This temperature anomaly is averaged over the first 48 hours of the free forecast period. The average temperature anomaly for GFS analysis, CTRL run and 3DVar run are shown in Fig.3.6a. The trend in the average temperature anomaly profile in gfs analysis is reasonably reproduced by both the model runs. However, the temperature anomaly values simulated by the saph run are closer to GFS analysis in the lower model levels as compared with the ctrl simulation. Fig. 3.6b indicates the area averaged and time averaged relative vorticity profile in a $3^\circ \times 3^\circ$ box around the cyclone centre. Here also, the first 48 hours of the model free forecast is considered for time averaging based on inference of trend of maximum wind speed and minimum sea level pressure as seen in Fig.3.5. The relative vorticity profile (Fig.3.6b) shows that the saph run has simulated a stronger vortex in the lower levels as compatred with the ctrl as well as GFS analysis. However, the profile of relative vorticity over the higher levels in saph run is closer to the profile from GFS analysis.

Rainfall

Fig.3.7 shows the spatial distribution of 24 hour accumulated rainfall for day 1 and day 2 forecasts from CTRL run (Fig.3.7(b,e)) and 3DVar run(Fig.3.7(c,f)) along with observed rainfall from TRMM (Fig.3.7(a,c)). The maximum rainfall observed is over the oceans. While there are location and intensity errors in the simulation of maximum rainfall in both the CTRL and 3DVar runs, the 3DVar run simulates the accumulated rainfall field with lower location and intensity errors than the CTRL run.

A quantitative measure of improvement in rainfall simulation due to assimilation of SAPHIR radiances is obtained by calculating statistical skill scores of 48hr accumulated precipitation simulated by the model. The skill scores are calculated with respect to TRMM rainfall observations. Fig.3.8 shows the equitable threat score (ETS)(Fig.3.8a), bias score (Fig.3.8b), false alarm ratio (FAR) (Fig.3.8c), and probability of detection (./images/chapter3) (Fig.3.8d). 3DVar run has higher ETS scores, higher ./images/chapter3 scores and lower FAR scores in all the higher rainfall thresh-

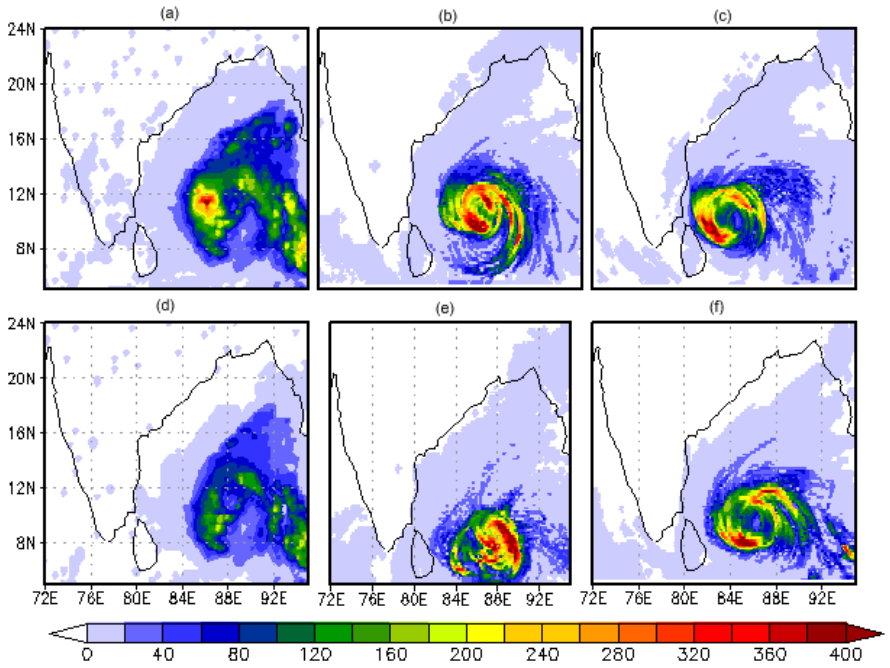

Figure 3.7: 24 hour accumulated precipitation (in mm) from TRMM observations (a,d), CTRL run (b,e) and 3DVar run (c,f) forecasts for day 1(d-f) and day 2(a-c) for cyclone Thane

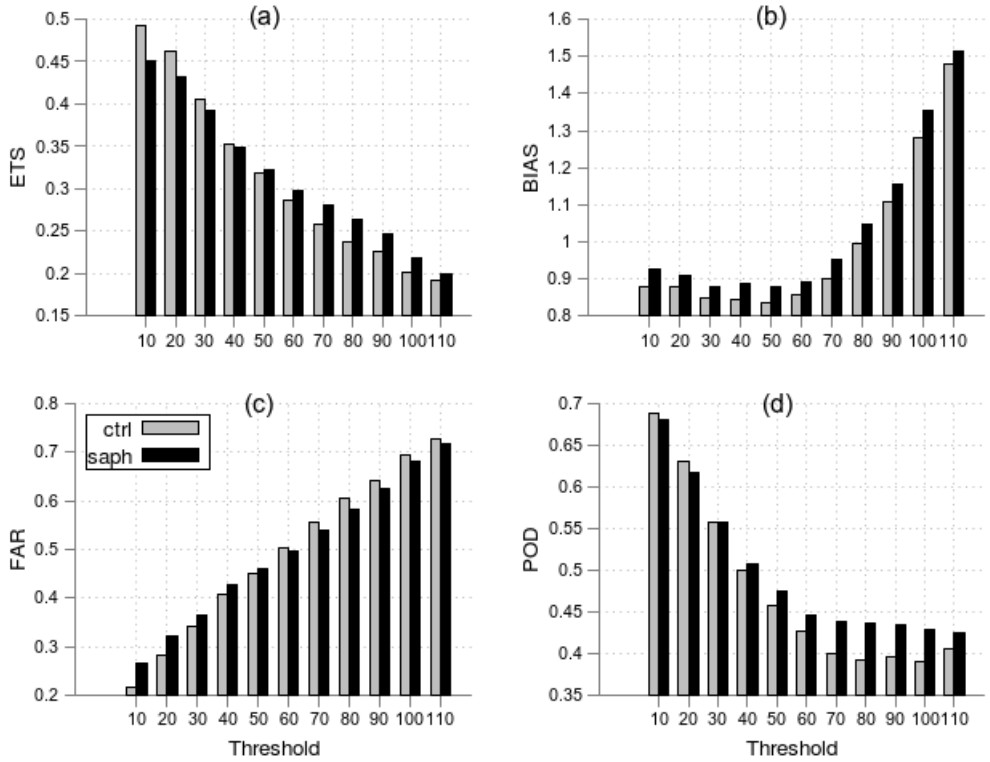

Figure 3.8: Equitable threat score (a), bias (b), false alarm ratio(c) and probability of detection (d) of 48hr accumulated rainfall estimated with respect to TRMM observations for cyclone Thane

olds . Bias values are similar in both CTRL and 3DVar runs. The above result indicates that the assimilation of SAPHIR radiances has a moderate positive impact on the heavy rainfall simulation.

3.6 Cyclone Nilam (28 October - 01 November 2012)

3.6.1 Synoptic Conditions

A depression was formed over southeast and adjoining southwest Bay of Bengal on 28^{th} October 2012. Moving westward, the system deepened into a deep depression by 29^{th} October 2012 near 9°N, 83.0°E. The deep depression moved further west and intensified into cyclonic storm Nilam. The storm then moved northnorthwestwards, crossed north Tamilnadu coast near Mahabalipuram on 31^{st} October 2012. After experiencing landfall, the cyclonic storm Nilam moved west-northwestwards and weakened gradually into a deep depression and then into a depression over south Interior Karnataka in the morning of 01^{st} November 2012.

3.6.2 Experimental Details

For the simulation of cyclone Nilam, the GFS model forecasts of $0.5^0 \times 0.5^0$ horizontal resolution are used to initialize the WRF model. From the GFS forecasts fields, the model has been integrated for a 6-hour spin up from 18Z 27^{th} October 2012. Subsequently two numerical experiments are performed. (i) ctrl run - surface and upper air observations from NCEP as well as satellite winds are assimilated in the control run at 00Z, 06Z, 12Z and 18Z of 28^{th} October 2012. (ii) saph run - Assimilations are performed in a cyclic manner at 00Z, 06Z, 12Z and 18Z of 28^{th} October 2012 by ingesting SAPHIR radiances also to the observations assimilated for the ctrl run. For both ctrl and saphir runs, the WRF model has been integrated in a free forecast mode with no further assimilation of observations, till the cyclone experienced land fall. The model domains' horizontal resolution, number of grid cells in the E-W and N-S directions as well as the number of vertical levels for the simulation of cyclone Nilam are the same as utilized for cyclone Thane. The physics options utilized here are also the same as used for cyclone Thane.

3.6.3 Results and discussion

Mean sea level pressure and wind

The spatial structure of mean SLP and 850hPa wind vectors from GFS analysis, ctrl run and 3DVar run are shown in Fig.3.9. Fig.3.9(a-c) shows the SLP and wind fields at the start of the free forecast. Both ctrl and saph runs simulate a relatively stronger vortex closer to SriLanka than seen in the GFS analysis. After 24 hours, the saph run has simulated a cyclonic system with a larger horizontal extent than in the GFS system. However, by 48 hours of model forecast, the saph run has simulated a spurious vortex over the southern tip of India.

The time series of minimum SLP and maximum wind speed for Nilam cyclone is shown in Fig.3.10a and Fig.3.10b respectively. As inferred from the spatial plots for slp and wind Fig.3.9, the time series plot also reveals that the model has simulated a stronger cyclonic system than what is observed in terms of minimum slp and maximum wind speed. Both ctrl and saph runs have however, over estimated the strength of the cyclonic system. However, the saph run has simulated a cyclonic system having both minimum slp and maximum wind speed closer to observations as compared to the ctrl run.

Vertical profiles of average temperature anomaly and relative vorticity

The time averaged and area averaged air temperature anomaly profile and relative vorticity in a $3^{\circ} \times 3^{\circ}$ box around the cyclone centre for the first 48 hours simulated for cyclone Nilam are shown in Fig.3.11(a) and (b). The temperature anomaly (Fig.3.11(a)) indicates that the model has simulated a stronger warm core at the surface in both ctrl and saph runs, as compared with the GFS analysis. The warm core at 300 hPa is more pronounced in the ctrl simulation as compared with the saph run as well as GFS analysis. Also, the relative vorticity profile is also stronger in the ctrl run as compared with the saph run at all levels.

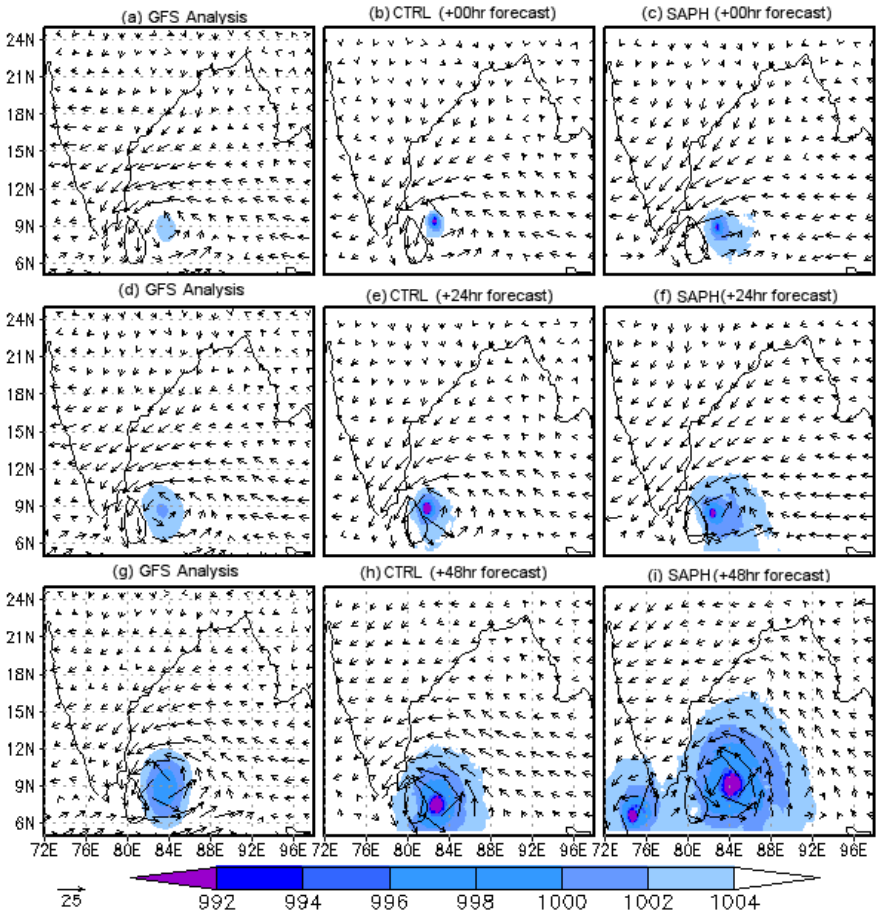

Figure 3.9: Mean sea level pressure and 850hPa wind from GFS analysis (a,d,g), CTRL run (b,e,h) and 3DVar run (c,f,i) at +00hr, +24hr and +48hr forecasts for cyclone Nilam

Rainfall

Fig. 3.12 shows the 24hr accumulated rainfall for day 1 (Fig.3.12(a-c)) and day 2 (Fig.3.12(d-f)). Accumulated rainfall for ctrl run is shown in Fig. 3.12(b,e), saph run in Fig. 3.12(c,f) and TRMM observations in Fig. 3.12(a,d) for Nilam cyclone. The intensity of heavy rainfall in the saph run for day 1 forecast is closer to TRMM observations as compared with the ctrl run. In the day 2 forecast, the location error of the maximum rainfall is lower in the saph run as compared with the ctrl run. However, the saph run has simulated relatively intense rainfall in day 2 forecast near the south west coast of India as compared with the ctrl run and TRMM observations. The above-mentioned anomalous rainfall pattern is possibly associated with the spurious cyclonic vortex simulated by the saph run as seen in Fig.3.9(g-i).

The quantitative verification of the first 48 hr accumulated rainfall as simulated by

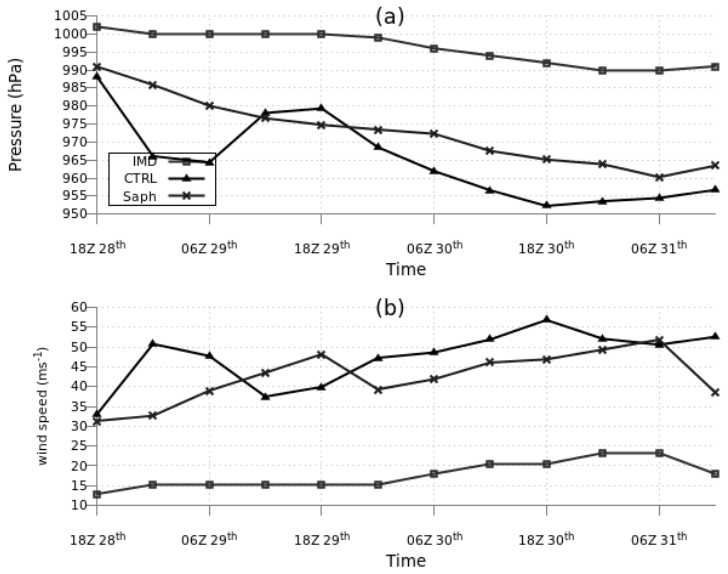

Figure 3.10: Time series of minimum sea level pressure in hPa (a) and maximum wind speed in ms^{-1} (b) for cyclone Nilam

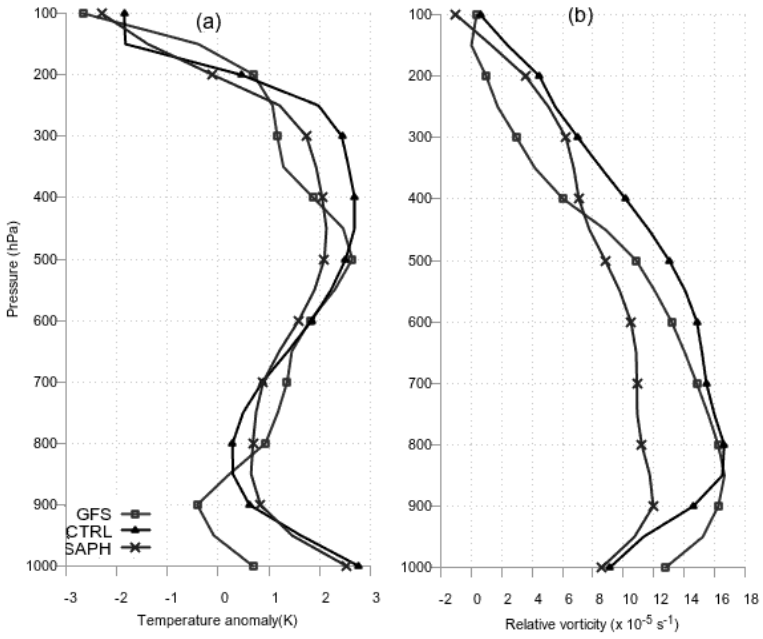

Figure 3.11: Area averaged and time averaged temperature anomaly (a) and relative vorticity (b) in a $3^{\circ}\times3^{\circ}$ box around the cyclone centre for cyclone Nilam during the first 48 hours of free forecast

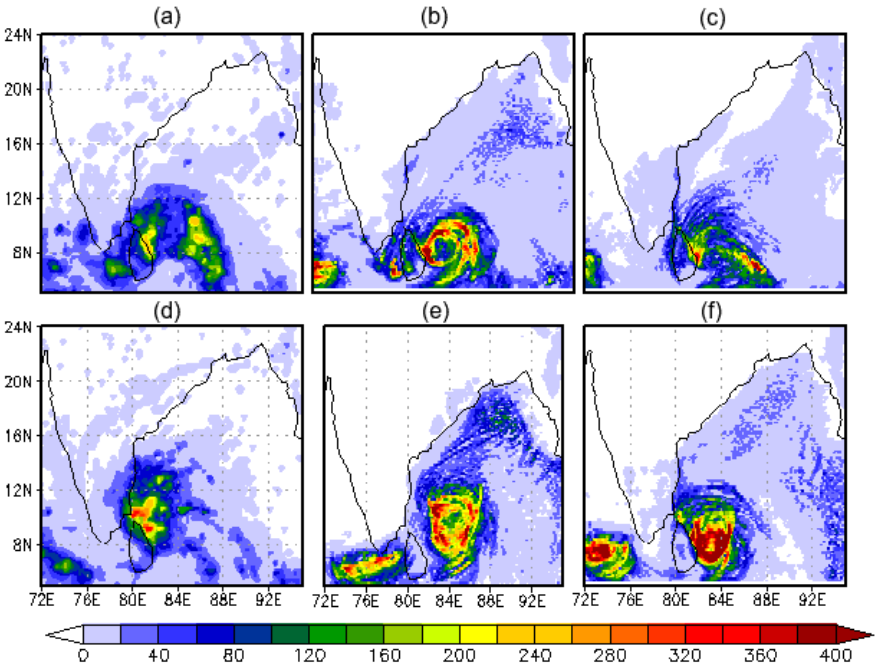

Figure 3.12: 24 hour accumulated precipitation (in mm) from TRMM observations (a,d), ctrl run (b,e) and saph run (c,f) forecasts for cyclone Nilam

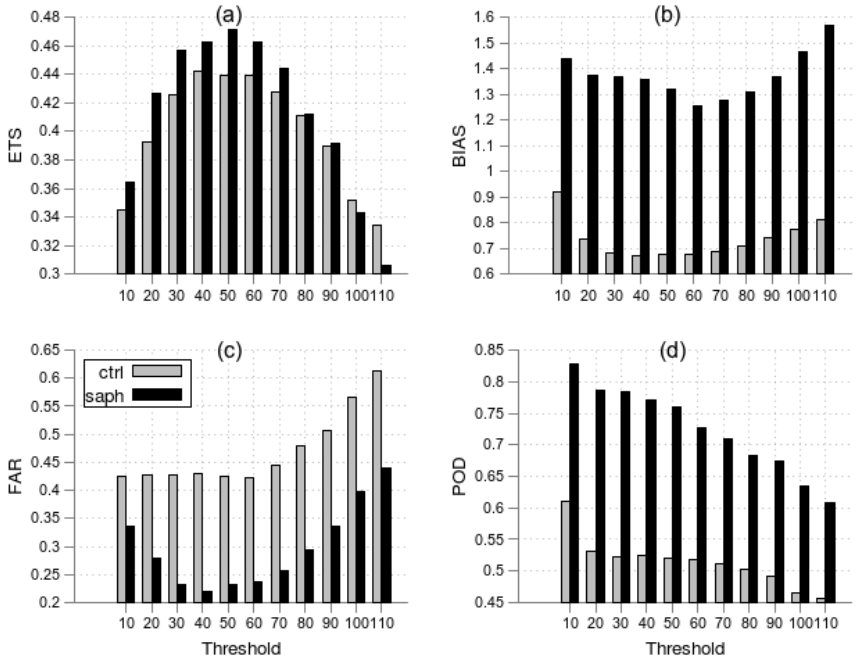

Figure 3.13: Equitable threat score (a), bias (b), false alarm ratio(c) and probability of detection (d) of 48hr accumulated rainfall estimated with respect to TRMM observations for cyclone Nilam

the model with respect to TRMM observations (Fig.3.13) shows that assimilation of SAPHIR radiances have improved the rainfall simulation in quantitative terms considerably in terms of higher ETS and ./images/chapter3 values and lower FAR values for all the three cyclones. However, the bias score is higher in the saph run as compared to ctrl run indicating that the radiance assimilation also results in over prediction of heavy rainfall.

3.7 Cyclone Phailin (8-14 October 2013)

3.7.1 Synoptic Conditions

The very severe cyclonic storm Phailin originated from the remnant of a cyclonic system from the South China Sea. The cyclonic circulation manifested as a low pressure area over Tenasserim coast on 6th October 2013. The system was over north Andaman Sea as a well marked low pressure area on 7th October 2013 and intensified to a depression over the north Andaman Sea on 8th October 2013 with its centre near 12.00°N, 96.00°E. The system intensified into a deep depression on 9th October 2013 near 13.00°N and 93.50°E. Moving west-northwestwards, the system crossed Andaman Islands and moved over east central Bay of Bengal intensifying into a cyclonic storm. The cyclonic system further intensified to a severe cyclonic storm on 10th October 2013. The system further intensified and experienced landfall on 12th October 2013 near Andhra Pradesh and Odisha coast. The system weakened subsequently and by October 14th, 2013 the depression had become a well marked low pressure system.

3.7.2 Experimental Details

As in the previous two cyclone cases, for the cyclone Phailin also, the WRF model is initialized from GFS forecasts and a spin up of 6 hours is performed. For cyclone Phailin, the model spin up is from 18Z 7th October 2013 to 00Z 8^{th} October 2013. Subsequently, two numerical experiments are performed. (i) ctrl run , which is a control run where conventional observations (surface and upper air) along with satellite winds are assimilated at 00Z, 06Z, 12Z and 18Z of 8^{th} October 2013. Thereafter the

model is integrated from 18Z 8^{th} October 2013 to 12^{th} October 2013 in a free forecast mode without assimilation of observations. (ii) saph run which is configured the same way as the ctrl run except that SAPHIR radiances are also included in the assimilation cycles in addition to conventional observations and satellite wind observations. The model domain, horizontal resolution, number of grid cells in E-W and N-S directions and the number of vertical levels for cyclone Phailin are the same as that employed for cyclone Thane. The physics options utilized here are also the same as used for cyclone Thane.

3.7.3 Results and Discussion

Mean sea level pressure and wind

Fig.3.14 shows the mean sea level pressure and 850hPa wind from GFS analysis, ctrl run and saph run at the start of the free forecast (Fig.3.14a-c), valid at 24 hour forecast (Fig.3.14 d-f) and at 48 hour forecast (Fig.3.14g-i). The above figures indicate that the WRF model simulates a more intense cyclonic system as compared with the GFS analysis in both ctrl run as well as saph run. The initial and 24 hour cyclonic systems are similar to each other in the ctrl and saph runs in terms of location of minimum sea level pressure. However, the saph run has simulated an anomalous low pressure system which is not seen in the GFS analysis close to the Indian east coast in the 48 hour simulation.

The time series of minimum sea level pressure (Fig.3.15a) and maximum wind speed (Fig.3.15b) over the domain indicates that relatively stronger winds as compared with the ctrl run and observations are simulated in day-1 of the forecast for cyclone Phailin due to the assimilation of SAPHIR radiances. SAPHIR provides humidity information at high spatial resolution even in the presence of non-precipitating clouds. The improved moisture content of the atmosphere due to SAPHIR observations can impact the low level moisture convergence associated with the cyclonic system, which will definitely impact the vertical ascending motion of moist air and the release of the latent heat of condensation in the model simulation. This can lead to further intensification of the cyclonic system through the CISK (Conditional Instability of Second Kind) mechanism. For cyclone Phailin, the time series of maximum

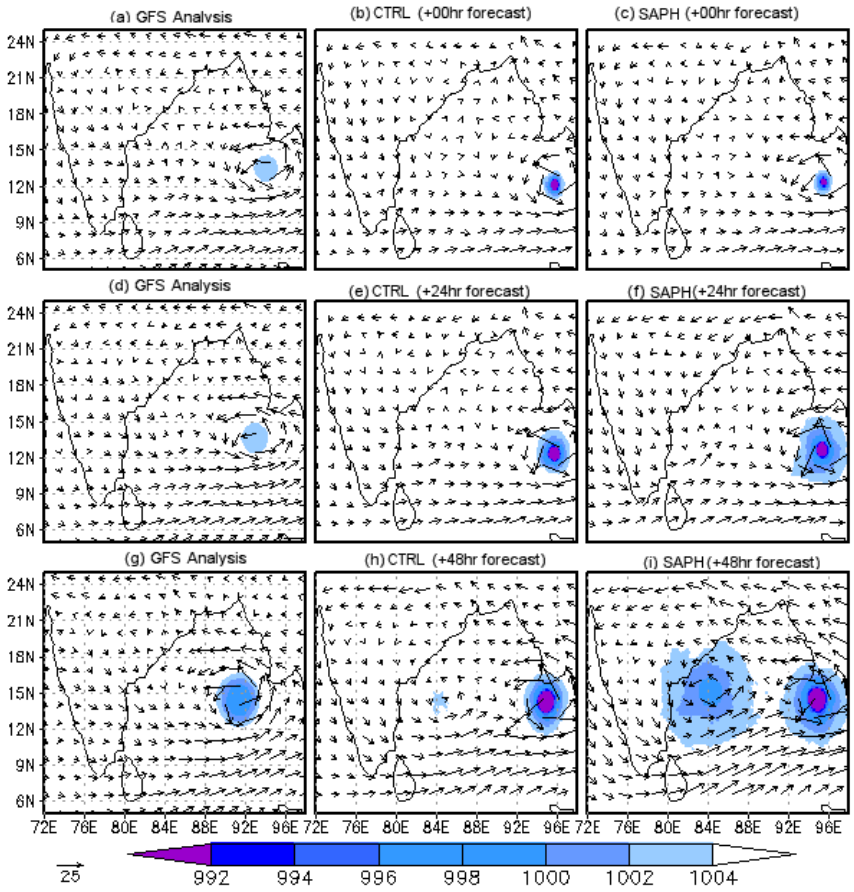

Figure 3.14: Mean sea level pressure and 850hPa wind from GFS analysis (a,d,g), ctrl run (b,e,h) and saph run (c,f,i) at +00hr, +24hr and +48hr forecasts for cyclone Phailin

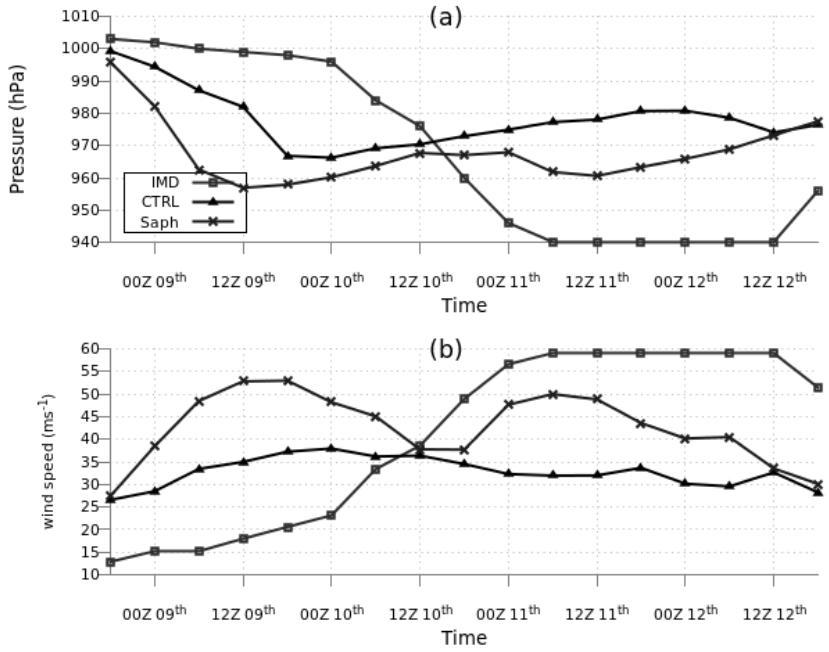

Figure 3.15: Time series of minimum sea level pressure in hPa (a) and maximum wind speed in ms^{-1} (b) for cyclone Phailin

wind speed indicates that the rapid observed strengthening of the cyclone system after 48 hours of integration is not captured by the model. The above inadequacy in the model simulation is also seen in the time series of minimum SLP.

Vertical profiles of average temperature anomaly and relative vorticity

The time averaged and area averaged temperature anomaly profile and relative vorticity in a $3^{\circ} \times 3^{\circ}$ box around the cyclone centre for the first 48 hours simulated for cyclone Phailin is shown in Fig.3.16(a) and (b). The assimilation of SAPHIR radiances has resulted in a stronger cyclone than is simulated in the ctrl run, which can be inferred from the higher values of temperature anomaly as well as larger relative vorticity at all model levels in the saph run as compared with the ctrl run as well as GFS analysis. Cyclone Phailin is characterised by stronger maximum wind speeds and lower minimum SLP for saph experiment for the first 30 hours of free forecast as compared to ctrl and observation (refer Fig. 3.15a and Fig.3.15b). Stronger pressure gradients at the surface would lead to greater moisture convergence and larger vertical ascending motion. These would invariably lead to higher relative vorticity through Ekman layer dynamics and the above effect is seen in Fig.3.16a for cyclone Phailin. Larger horizontal wind speed for cyclone Phailin in the saph run can lead to establish-

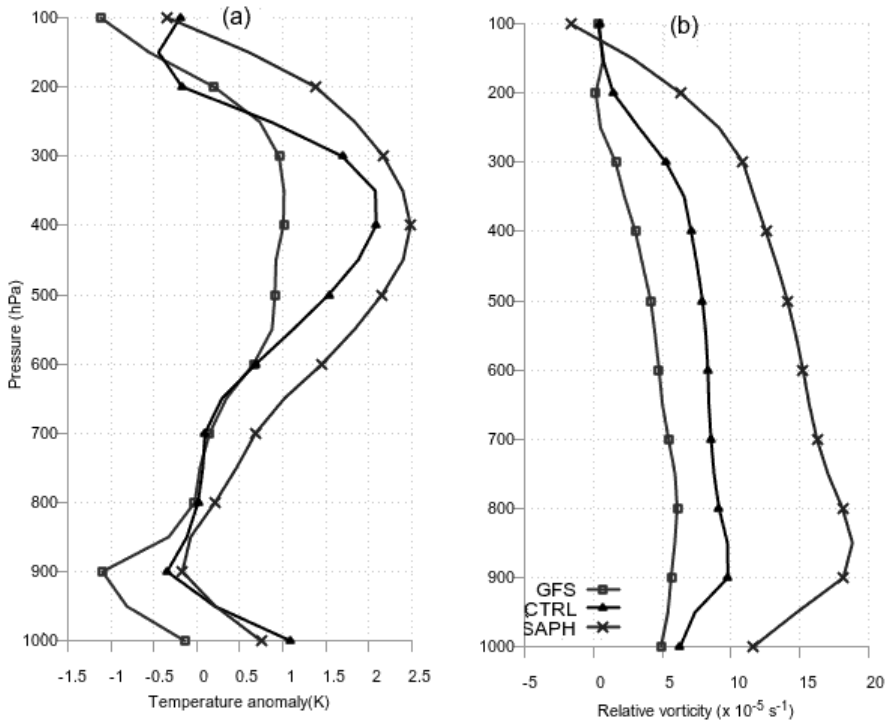

Figure 3.16: Area averaged and time averaged temperature anomaly (a) and relative vorticity (b) in a $3^\circ \times 3^\circ$ box around the cyclone centre for cyclone Phailin

ment of super gradient wind speeds which would entail larger subsiding motion over the centre of the tropical cyclone contributing to stronger warm core over the cyclone centre. The above effect is seen in Fig.3.16b for the cyclone Phailin.

Rainfall

The spatial distribution of 24 hour accumulated rainfall from day1 and day2 forecasts of ctrl and 3DVar runs are compared with TRMM observations in Fig.3.17. In the day 1 forecast, the ctrl run has not simulated the observed spatial pattern of rainfall over the Bay of Bengal region whereas the saph run is able to simulate rainfall over this region. However, the saph run has simulated more intense rainfall than is observed. In the day 2 forecast also, saph run has overestimated rainfall intensity both over the east coast as well as the west coast of India. This may be a manifestation of the anomalous low pressure system that is simulated by the saph run (refer Fig.3.14(i)).

Fig.3.18 shows the quantitative verification scores of 48 hour accumulated rainfall from the ctrl and saph runs calculated with respect to the TRMM observations.

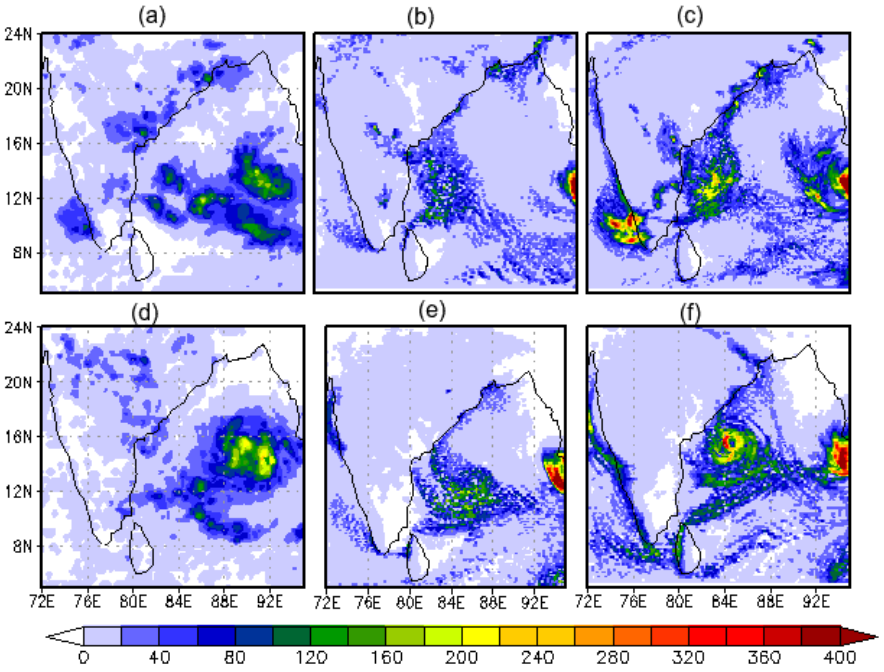

Figure 3.17: 24 hour accumulated precipitation (in mm) from TRMM observations (a,d), ctrl run (b,e) and saph run (c,f) forecasts for day 1(a-c) and day 2(d-f) forecasts for cyclone Phailin

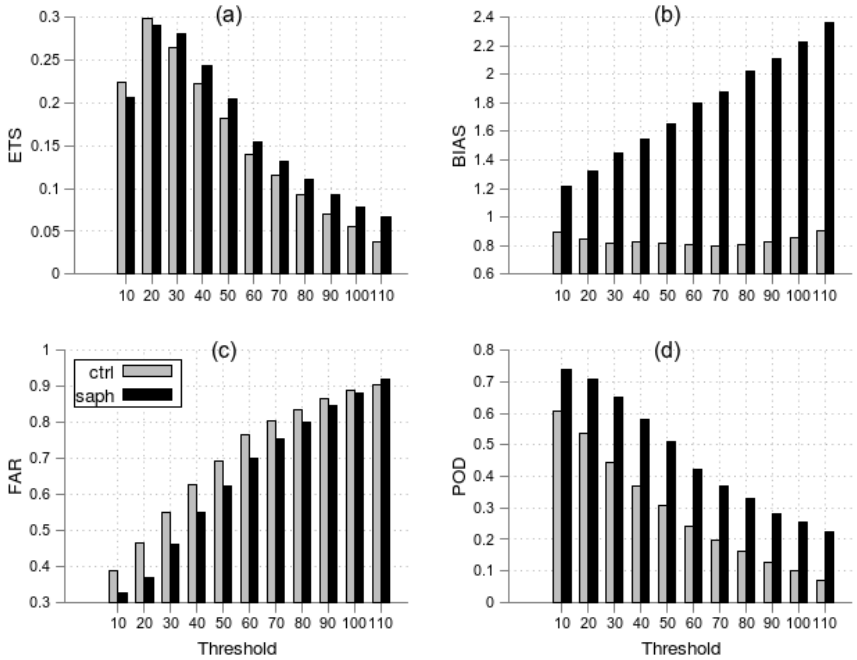

Figure 3.18: Equitable threat score (a), bias (b), false alarm ratio(c) and probability of detection (d) of 48hr accumulated rainfall estimated with respect to TRMM observations for cyclone Phailin

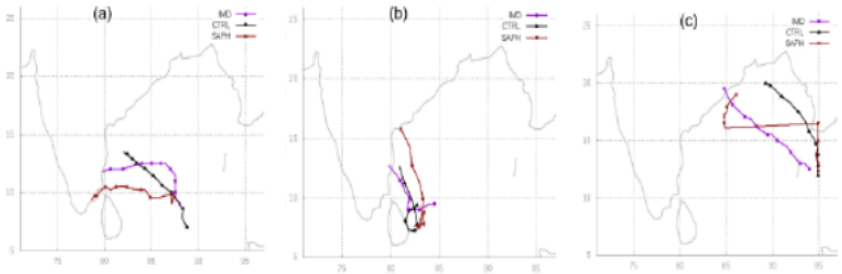

Figure 3.19: Cyclone track from CTRL, SAPH and IMD for cyclone Thane (a) , Nilam (b) and Phailin (c)

The saph runs has higher ETS (Fig.3.18(a)), lower FAR (Fig.3.18(c)) and higher ./images/chapter3 (Fig.3.18(d)). However, the bias is high in the saph run (Fig.3.18(b)), which agrees with the previous inference that the saph run over-predicts the precipitation as compared with the ctrl run.

3.7.4 Track simulated for the three cyclones

The track errors for the three tropical cyclones with respect to IMD track observations are given in Table 3.1. The Table3.1 shows that the error in the initial position of the cyclonic center is lower in the saph run as compared with the ctrl run for Thane and Nilam cyclones. However, the track error in the saph run as compared with to the ctrl run increases with forecast time. For cyclone Phailin, the results indicate lower track error in the last couple of days before occurrence of landfall for the saph run as compared to the ctrl run. Fig. 3.19 show the cyclone tracks for the model runs ctrl and saph runs together with IMD track observed for all the tropical cyclones investigated in this chapter.

3.8 Summary

The study of impact of assimilation of radiance observations from MeghaTropiques SAPHIR sensor is performed in this chapter. The impact of this assimilation on the simulation of three tropical cyclones namely Thane, Nilam and Phailin is investigated. It is seen that the SAPHIR humidity observations do influence the model simulation. When compared with a control (ctrl) run which assimilates conventional as well as

Table 3.1: Track Error(in km) for the cyclones Thane, Nilam and Phailin calculated with respect to IMD track data

	Thane		**Nilam**		**Phailin**	
Hours	CTRL	3DVar	CTRL	3DVar	CTRL	3DVar
00	87.6	65.7	198.6	180.8	104.5	104.3
06	89.3	37.1	233.2	179.6	160	160
12	98.1	46.1	167.5	86.7	215.6	215.6
18	26.5	199.2	79.5	47.2	271.2	253.2
24	37.5	269.4	58.9	250.9	326.8	308.8
30	109.3	352.3	100.0	232.5	382.4	373.4
36	99.0	350.9	103.3	309.1	466.6	493.6
42	88.6	376.6	94.5	336.8	495.2	549.2
48	86.9	402.1	45.3	297.8	496.8	604.9
54	76.1	400.6	63.3	311.3	563.0	400.4
60	74.3	426.1	46.9	281.0	504.2	387.2
66	72.6	424.9	126.1	307.3	471.9	320.4
72	62.0	432.2			441.7	242.6
78	40.5	393.7			416.3	160.0
84	107.2	319.4			451.3	71.0
90	167.0	243.0			470.9	29.8
96	209.2	193.4			497.3	146.3
102	242.6	146.0				

satellite wind observations, additional ingesting of SAPHIR radiances along with the conventional and satellite wind observations is seen to have a moderate positive impact of the model simulation of cyclone features. The humidity information obtained from SAPHIR has impacted all the model fields such as temperature, maximum wind speed, minimum sea level pressure and accumulated rainfall in a moderate manner.

Atmospheric humidity is a highly significant parameter that influences numerical simulation of mesoscale weather features of tropical cyclones. Assimilation of humidity information from SAPHIR observations has resulted in an improved analysis of moisture variable in the case of the three tropical cyclones investigated in this chapter. The representation of low level moisture convergence is improved by the assimilation of SAPHIR observations. This has resulted in an improved representation of latent heat release in the model simulation. Since a better simulation of latent heat of condensation influences the convective processes simulated by the model, the intensification of the tropical cyclones due to the CISK (Conditional Instability of Second Kind) mechanism is better simulated in the SAPHIR run as compared with the control run. The improved representation of latent heat flux results in an improved simulation of the intensity and the dynamical characteristcs of the cyclone. Hence, the simulation of physical and dynamical features of the tropical cyclones investigated in this chapter are improved due to the assimilation of SAPHIR radiances.

CHAPTER 4

Impact of formulation of background error covariance matrix on 3DVar assimilation

4.1 Introduction

The optimal analysis, obtained by combining observations with background or apriori information in a data assimilation scheme, is derived from statistical estimation theory. The above-mentioned two sources of information are weighted according to their respective error covariances. The observation error covariance matrix, $\mathbf{R}$ is determined by instrument errors and representation errors, the latter arising from the effect of unresolved scales in a model as well as errors that arise from the specification of observation operator that maps model space variables to observation space. However, the background error cannot be calculated exactly as we do not have the information regarding the true state of the system. Furthermore, in the case of a numerical weather prediction model, the background state has degrees of freedom of $\sim 10^7$. Hence the background error covariance matrix ($\mathbf{B}$) has a dimension of $\sim 10^7 \times 10^7$, which is impossible to explicitly calculate in a practical data assimilation system. Hence it is clear that one has to invoke methods which provide for estimation of 'background error covariance' ($\mathbf{B}$).

There are three main methods that are used for estimating $\mathbf{B}$. They are (i) using innovation statistics (Hollingsworth and Lönnberg, 1986) Usage of innovation (observation-minus-background) statistics has the limitation that this method requires a good quality, homogeneous observation network. Furthermore, this method provides estimates of background error for observed quantities only.

(ii) the National Meteorological Center (NMC) method (Parrish and Derber, 1992) The NMC method approximates the background error statistics using the difference between two model forecasts of different lead times, valid at the same time. The most popularly used method for generating the $\mathbf{B}$ is the NMC method despite having its own shortcomings. It is known that the NMC method underestimates the variance

of background error over data-sparse regions. Also, the time duration of forecasts used to generate $\mathbf{B}$ are typically 12 to 48 hours. Since the background fields used in the assimilation system would typically be of a shorter duration, the covariances of estimated background error could become broader than the actual background error (Fisher, 2003).The popularity of the method however arises from the fact that this method yields global statistics for $\mathbf{B}$ for model variables at all levels. Also it is less expensive to implement the NMC method in an operational weather prediction environment since the forecasts required to generate the differences would be already available.

(iii) the analysis-ensemble method (Fisher, 2003) The analysis system is run several times using perturbed inputs in the analysis-ensemble method. The difference between background fields in the various runs provides a surrogate of background errors in this method. The chief limitation with this method is due to the fact that generating and maintaining an ensemble which can adequately represent the prior distribution is not an easy task. This is especially true in the case of a limited area model whose boundary conditions also significantly influence the model ensemble.

The NMC method has been implemented to estimate the $\mathbf{B}$ in Weather Research and Forecasting (WRF) model's variational data assimilation system (WRFDA) (Barker et al., 2004). This method employs a control variable transform (CVT) in which the model variables are converted to the control variables of the data assimilation system. This CVT renders the $\mathbf{B}$ matrix diagonal and also imposes balance relations within the variables. The manner in which the CVT is specified can impact the data assimilation results. In WRFDA3.5.1, there are three choices for control variable transforms, namely, cv3, cv5 and cv6 options. In cv3 option, the control variables are described in physical space while the control variables are defined in the eigen vector space in cv5 and cv6 options. While cv3 option uses a vertical recursive filter to model the vertical covariance, both cv5 and cv6 options use an empirical orthogonal function (EOF) approach to represent the vertical covariance. Previous studies(Routray et al., 2014) have shown that the simulation of mesoscale weather phenomena like monsoon depressions over India can be improved using domain-specific $\mathbf{B}$ statistics calculated using cv5 option rather than utilizing cv3 option in WRF.

4.2 Background error covariance in the WRF model

The 3DVar analysis is obtained by minimizing a cost function defined as

$$J(\mathbf{x}) = J^b + J^o = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + (\mathbf{y} - H(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y} - H(\mathbf{x})) \quad (4.1)$$

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

The calculation of the background term J_b requires $\sim O(n^2)$ calculations for a system with n degrees of freedom. For a numerical weather model with typically 10^7 degrees of freedom, the direct calculation of this term is not possible. To reduce the computational cost , J_b is calculated in terms of control variables $\mathbf{v}$ defined via the relation $\mathbf{x}' = \mathbf{U}\mathbf{v}$ where $\mathbf{x}'$ denotes the analysis increment, $\mathbf{x}' = \mathbf{x} - \mathbf{x}^b$. Using the incremental formulation (Courtier et al., 1994; Barker et al., 2004) and the control variables, equation 4.1 can be rewritten as

$$J(\mathbf{v}) = J^b + J^o = \frac{1}{2}\mathbf{v}^T\mathbf{v} + \frac{1}{2}(\mathbf{y} - H\mathbf{U}\mathbf{v})^T \mathbf{R}^{-1}(\mathbf{y} - H\mathbf{U}\mathbf{v}) \quad (4.2)$$

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

i.e, $\mathbf{x}' = \mathbf{U}_h\mathbf{U}_v\mathbf{U}_p\mathbf{v}$

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

2004). A tuning factor is applied to the length scale, to reflect the actual correlation length scales in a domain.

In the vertical transform $\mathbf{U}_v$, an empirical orthogonal function (EOF) decomposition is performed on the vertical component of the background error $\mathbf{B}_v$. The analysis increments are projected onto the eigen vector space and the eigen values specify the relative weights of increments in the calculation of cost function. The parameter transform $\mathbf{U}_p$ is applied so that the errors in the control variables are not correlated with each other and the $\mathbf{B}$ matrix is rendered block-diagonal. The increments in model variables u (zonal wind) , v (meridional wind), T (temperature), p (pressure) and humidity (q) are converted to new set of variables such as stream function (ψ), velocity potential(χ), temperature (T), surface pressure (ps) and relative humidity(rh). WRF-Var system provides the balance relations between the new set of variables using regression relations. After the "balanced part" of analysis variables is estimated, the "unbalanced part" is determined by subtracting the former from the full fields. Hence, while some fields are analysed in full , for some other variables the unbalanced parts are included in the analysis system. The control variable options cv5 and cv6 differ from each other in the specification of the balance relations between these control variables.

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

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

$$\chi_u(i,j,k) = \chi(i,j,k) - \alpha_{\psi\chi}(i,j,k)\psi(i,j,k) \tag{4.3a}$$

$$T_u(i,j,k) = T(i,j,k) - \sum_{l=1}^{N_k} \alpha_{\psi T}(i,j,k,l)\psi(i,j,l) \tag{4.3b}$$

$$ps_u(i,j) = ps(i,j) - \sum_{l=1}^{N_k} \alpha_{\psi ps}(i,j,l)\psi(i,j,l) \tag{4.3c}$$

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

The relations specified by equations 4.3 indicate the manner in which the various analysis control variables are related in WRFVar system. Here, the balanced part of velocity potential is related with the stream function alone and hence the possible relations that the divergent component of wind could have with other variables are not properly considered. Similarly, in cv5, temperature and surface pressure are not related with each other as well as with the moisture variable. Relative humidity field is not influenced by any of the model variables like temperature or wind, as per the relations in equations 4.3.

The balance relations defined in cv6 option, on the other hand, includes additional correlations between model variables. The cv6 option has the following balance relations.

$$\chi_u(i,j,k) = \chi(i,j,k) - \alpha_{\psi\chi}(i,j,k)\psi(i,j,k) \tag{4.4a}$$

$$T_u(i,j,k) = T(i,j,k) - \sum_{l=1}^{N_k} \alpha_{\psi T}(i,j,k,l)\psi(i,j,l) - \sum_{l=1}^{N_k} \alpha_{\chi_u T}(i,j,k,l)\chi_u(i,j,l) \tag{4.4b}$$

$$ps_u(i,j) = ps(i,j) - \sum_{l=1}^{N_k} \alpha_{\psi ps}(i,j,l)\psi(i,j,l) - \sum_{l=1}^{N_k} \alpha_{\chi_u ps}(i,j,l)\chi_u(i,j,l) \tag{4.4c}$$

$$rh_u(i,j,k) = rh(i,j,k) - \sum_{l=1}^{N_k} \alpha_{\psi rh}(i,j,k,l)\psi(i,j,l) - \sum_{l=1}^{N_k} \alpha_{\chi_u rh}(i,j,k,l)\chi_u(i,j,l) - \sum_{l=1}^{N_k} \alpha_{T_u rh}(i,j,k,l)T_u(i,j,l) - \sum_{l=1}^{N_k} \alpha_{ps_u rh}(i,j,k)ps_u(i,j) \tag{4.4d}$$

Here, additional correlation coefficients connect model variables in more ways than are available in cv5 option. For example, the balanced part of temperature and surface pressure are now correlated with the unbalanced velocity potential also. Hence, temperature and surface pressure are influenced by the divergent component of wind in the cv6 option unlike in the cv5 option. Similarly, additional correlations defined

in the moisture variable makes the moisture analysis multivariate in nature in the cv6 option (Chen et al., 2013).

In both cv5 and cv6 options, the estimation of background error covariances is accomplished through the following five stages

(i) Calculation of standard perturbations from forecast differences. In NMC method, this is calculated as $\mathbf{x}' = \mathbf{x}_{T2} - \mathbf{x}_{T1}$ where $\mathbf{x}_{T2}$ and $\mathbf{x}_{T1}$ are forecast difference times (e.g. 48hr, 24hr for global, 24hr, 12hr for regional).

(ii) Time/bin mean for each variable/level is removed so that zero-mean fields are obtained

(iii) Regression analysis is performed and the various correlations are determined between the control variable fields. The unbalanced components of the fields are calculated.

(iv) The vertical component of control variable transform is applied

(v) Recursive filter is used to provide horizontal correlations

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

4.3 Comparison of cv5 and cv6

The characteristic features of $\mathbf{B}$ generated by cv5 and cv6 options are compared here to determine in what all features they significantly differ. In the WRFVar system, the vertical transform is determined by the eigen vectors and the corresponding eigen values represent the variance. During assimilation of observations, the analysis increment is weighted by the eigen values. The horizontal length scale of the recursive filter determines the horizontal spread of the information in the EOF space. The information spreads vertically due to the transformation from the EOF space to the physical space (Descombes et al., 2015). Fig.4.1 show the first five eigenvectors for the control variables for cv5 Fig.4.1(a-d) and cv6 Fig.4.1(e-h). Fig.4.2(a,b) shows the asscoiated eigen values of unbalanced temperature (T_u) corresponding to cv5 and cv6 options and fig.4.2(c,d) shows the eigen values for relative humidity for cv5 and cv6. Here,

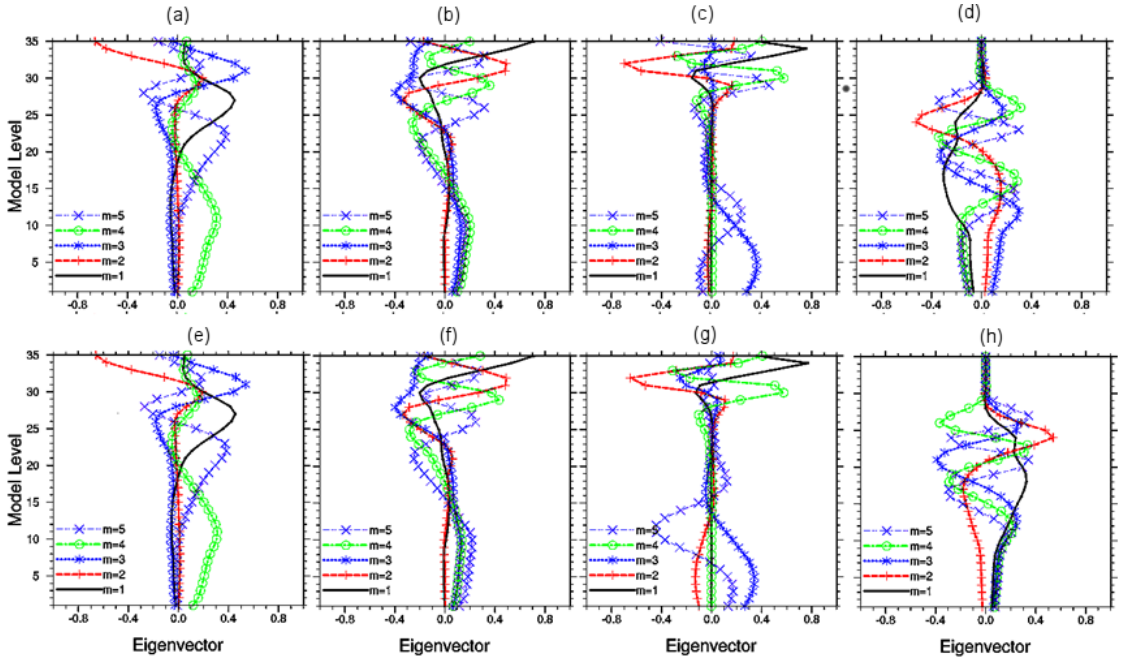

Figure 4.1: First five eigenvectors from the EOF decomposition on the vertical component of $\mathbf{B}$ for ψ (a,e), χ_u (b,f), T_u (c,g) and rh (d,h) for cv5 option (a-d) and cv6 option (e-h)

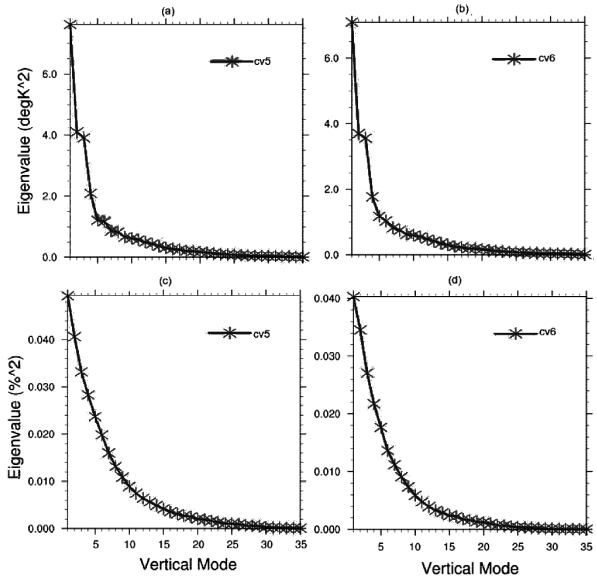

Figure 4.2: Eigen values for unbalanced temperature (a,b) and relative humidity (c,d) for cv5 (left panel) and cv6 (right panel)

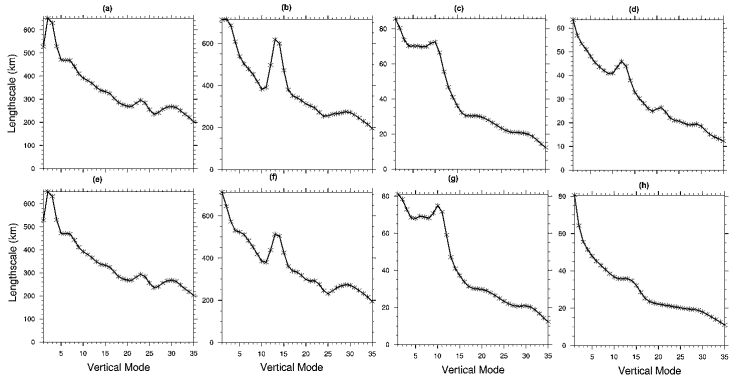

Figure 4.3: Horizontal length scales for each vertical mode for $\mathbf{B}$ for ψ (a,e), χ_u (b,f), T_u (c,g) and rh (d,h) in cv5 (a-d) and cv6 (e-h)

the vertical mode number of the corresponding EOF is shown in x-axis. y-axis gives the eigen value for the associated EOF mode. The eigen values corresponding to these two analysis control variables are shown here, since the same for ψ and χ_u in both cv5 and cv6 options are same and hence their eigen values do not differ. However, as shown here , for T_u and rh the eigen values are lower in cv6 option as compared to that in cv5 option over all vertical modes. The horizontal length scales for each vertical mode for $\mathbf{B}$ for all the control variables are shown in Fig.4.3. The figure shows that for smaller vertical modes the rh in cv6 option has a larger lengthscale than that in cv5 option, indicating a greater horizontal spread for relative humidity information in cv6 as compared with cv5. The above results do provide evidence that the characteristic features of $\mathbf{B}$ generated using cv6 option do differ from the cv5 option and the major difference between the two is in the moisture variable as is to be expected.

4.3.1 Single Observation Experiment

The significance of $\mathbf{B}$ in spreading the information in vertical and horizontal directions in space can be explained using the single observation experiment.To understand this, let us consider the expression of analysis found from the Best Linear Unbiased Estimator (BLUE (Kalnay, 2003)) ,

$$\mathbf{x}^a - \mathbf{x}^b = \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}\{\mathbf{y}^o - H(\mathbf{x}^b)\} \tag{4.5}$$

The analysis obtained by BLUE is the same as that from eqn.(2.19) when H is

linear. Even if H is nonlinear, the above provides a useful estimate when

$$H(\mathbf{x}^b + \delta\mathbf{x}) \approx H(\mathbf{x}^b) + \mathbf{H}\delta(\mathbf{x})$$

Consider only k_{th} element of the state $\mathbf{x}$, is observed at one location. In this case, $\mathbf{y} = y$ and $\mathbf{R} = \sigma^2$. Also, H and $\mathbf{H}$ become row vectors of zeroes apart from their k^{th} element, which becomes 1. Then, from eqn.(4.5), analysis increment at element l is ((Bannister, 2008b),

$$\mathbf{x}_l^a - \mathbf{x}_l^b = B_{lk}\frac{\mathbf{y} - \mathbf{x}_k^b}{B_{kk} + \sigma^2} \tag{4.6}$$

That is, the analysis increment is proportional to the matrix elements B_{lk}. This is called a structure function.

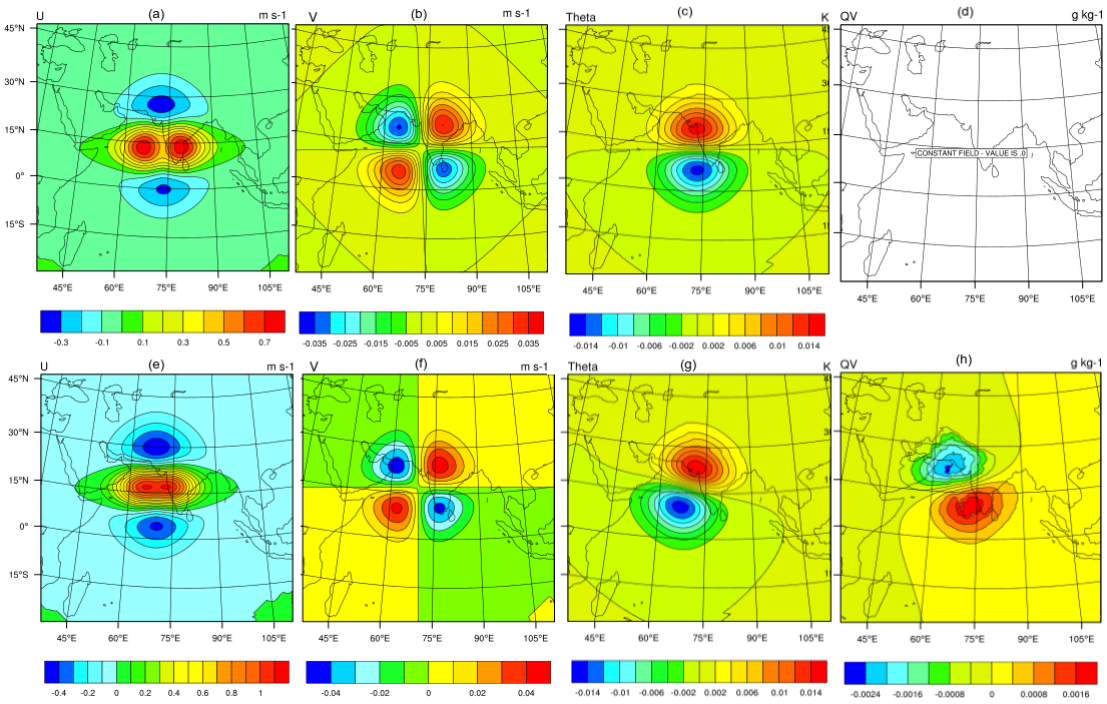

Figure 4.4: Analysis increment for zonal wind (a, e), meridional wind (b, f), potential temperature (c, g) and water vapour mixing ratio (d, h) when a single u-wind observation is assimilated at the middle of the domain at model level 19. Fig(a-d) is for cv5 run and Fig(e-h) is for cv6 run

Fig. 4.4 shows the analysis increments in u, v, θ and q when a single u-wind observation is assimilated at the middle of the domain at model sigma level 19. u-wind observation which differs from the background with a magnitude of $1ms^{-1}$ is assim-

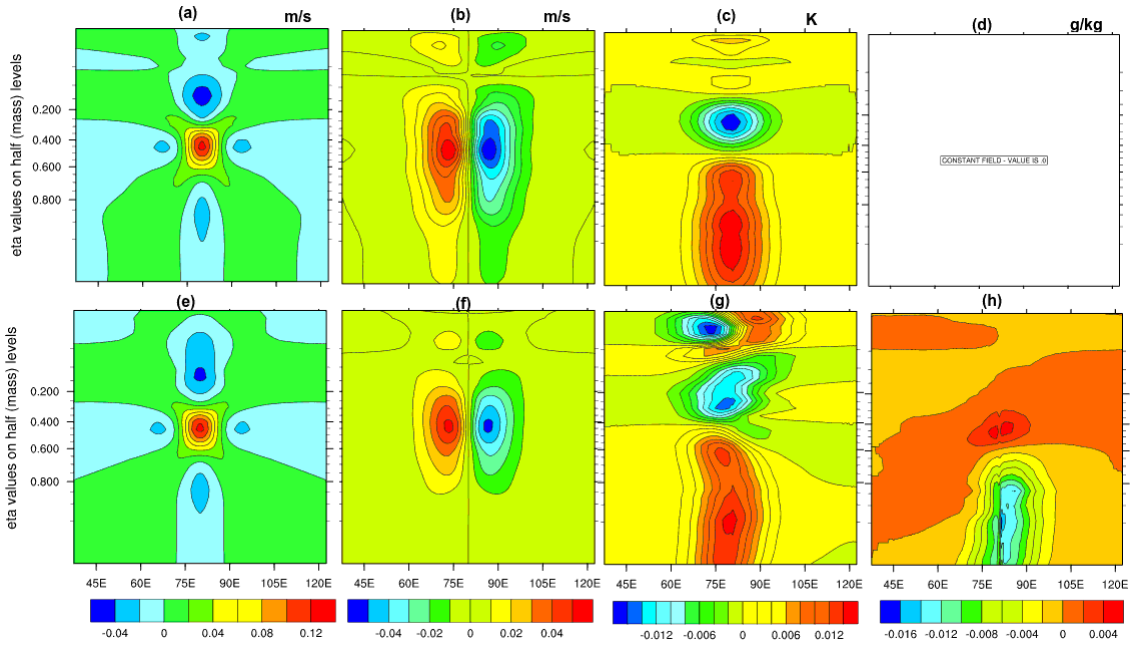

Figure 4.5: Analysis increment in the x-z plane for zonal wind (a, e), meridional wind (b, f), potential temperature (c, g) and water vapour mixing ratio (d, h) when a single u-wind observation is assimilated at the middle of the domain at model level 19. Fig(a-d) is for cv5 run and Fig(e-h) is for cv6 run

ilated using both cv5 and cv6 options of $\mathbf{B}$ matrix. Fig 4.4a-d provides the analysis increment in u, v, θ and q respectively for cv5 option while Fig 4.4e-h shows the same for cv6 options. The analysis increment patterns for cv5 option are similar to that obtained by Routray et al. (2014) for regional BE option with a double maxima over the ocean region in the C(u,u) pattern (Fig.4.4 a and e). The inclusion of additional correlation functions in cv6 has resulted in non-vanishing increment in the moisture field due to assimilation of wind information. Furthermore, the additional correlation coefficient $\alpha_{\chi_u T}$ in the expression for T_u has resulted in a changed temperature field (Fig4.4 (g)) for the cv6 option. Also, the location of maximum impact of the observation is shifted in cv6 run as compared to the cv5 run. Since the covariance matrix is symmetric, assimilation of a single temperature observation will result in a relative rotation of increment pattern in the u-field (Chen et al., 2013). Hence, the results of the single observation experiments reveal that the difference in formulation of control variables has resulted in different analysis increment structures for the cv6 option as compared to the cv5 option.

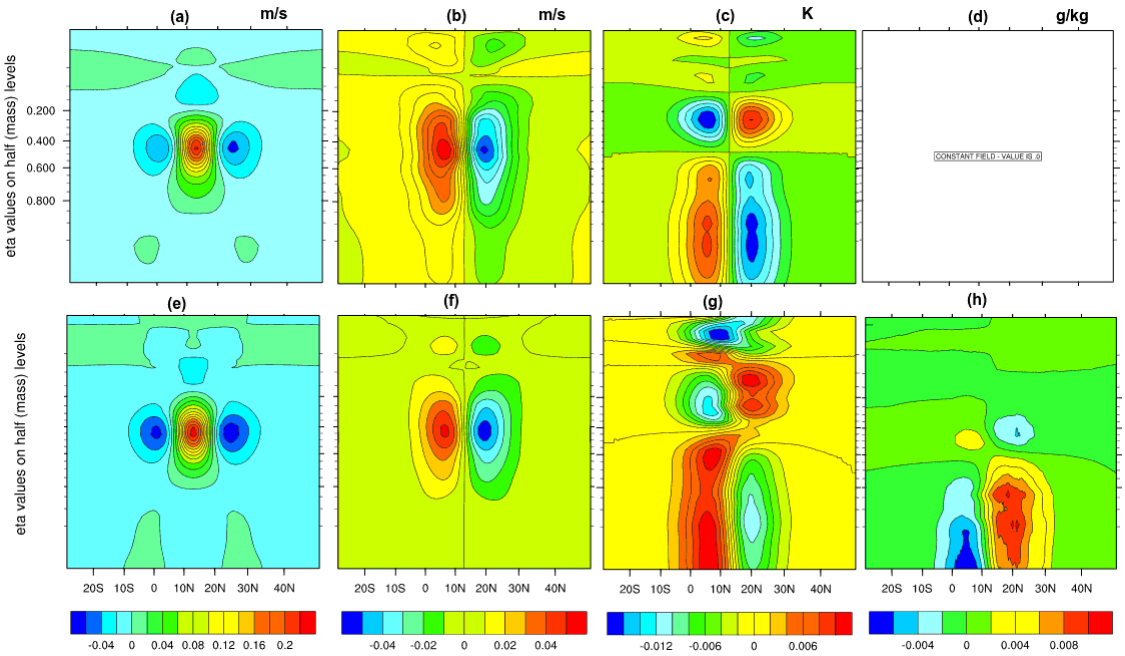

Figure 4.6: Analysis increment in the y-z plane for zonal wind (a, e), meridional wind (b, f), potential temperature (c, g) and water vapour mixing ratio (d, h) when a single u-wind observation is assimilated at the middle of the domain at model level 19. Fig(a-d) is for cv5 run and Fig(e-h) is for cv6 run

Fig.4.5 shows the x-z cross section of the analysis increments at the centre longitude of the domain where the single u-wind observation is assimilated. Fig.4.5a-d provides the analysis increment in u, v, θ and q, respectively for cv5 option while Fig 4.5e-h shows the same for the cv6 option. The increments in all the fields show differences between cv5 and cv6 options. The most prominent difference is in the mixing ratio field where the assimilation of u-wind observation has not modified the background field in cv5 option. However, the assimilation of u-wind pseudo-observation has resulted in a modified mixing ratio field at the lower model levels significantly, close to the location of the observation. Also, the pattern of increment is not symmetrical about the location of observation, as in the case of increments in u-wind and v-wind fields, but is tilted slightly westwards. Similarly, the pattern of increments in θ also show a tilt in the cv6 option as compared to the cv5 option. Furthermore, u-wind observation has impacted the potential temperature field at a higher model level in the cv6 option as compared with the cv5 option.

The analysis increments in a y-z cross section at the longitude of the pseudo-

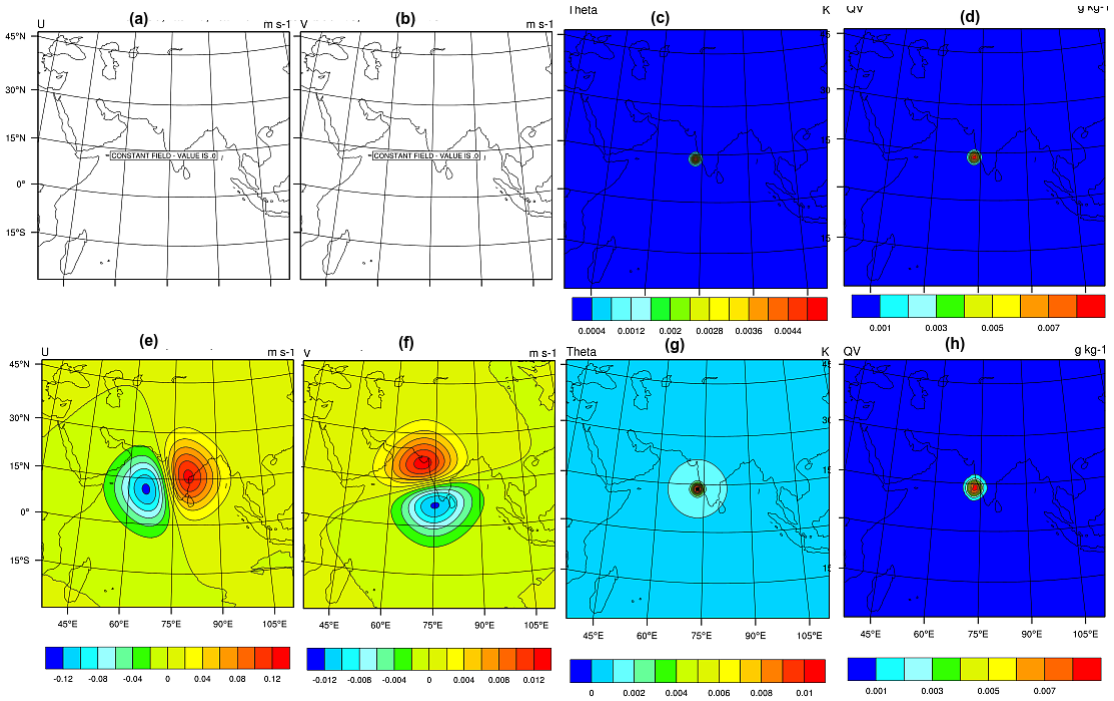

Figure 4.7: Analysis increment in the for zonal wind (a, e), meridional wind (b, f), potential temperature (c, g) and water vapour mixing ratio (d, h) when a single q observation is assimilated at the middle of the domain at model level 19. Fig(a-d) is for cv5 run and Fig(e-h) is for cv6 run

observation is shown in Fig.4.6. Here also, the major differences between the cv5 and cv6 options are in the θ and q fields. Both cv5 and cv6 options show symmetrical increment structures at the observation longitude for θ. However, the pattern of increments show that the analysis increments in θ has higher magnitude in the cv6 option as compared to the cv5 option. The tilt in the increment pattern of cv6 option with respect to that in cv5 option is seen here also. The increment pattern of mixing ratio field in cv6 option indicates that the maximum of analysis increment has been shifted southwards from the observation location.

To verify the differences between cv5 and cv6, further, a single q pseudo-observation is assimilated in a fashion similar to the single u-wind assimilation experiment. Here, the observation error is fixed as 0.001 $kgkg^{-1}$ and the observation is assumed to have a magnitude which differs from the background value by 0.001 $kgkg^{-1}$. The resulting analysis increments shown in Fig.4.7 affirm the differences between cv5 and cv6 options. Since $\mathbf{B}$ is symmetric, it is not surprising that the assimilation of q-observation

has not influenced the wind fields in cv5 option similar to u-wind observation not influencing the q field in cv5 option. Furthermore, the q-observation has had more impact on the θ field in cv6 option as compared with cv5 option. This can be inferred from the magnitude of the analysis increments of θ, the same being one order higher in cv6 case as compared with the cv5 case.

4.4 Impact of cv formulation on forecasts of monsoon depressions

4.4.1 Case Description

Case 1- Depression (29May - 31May 2013)

Under the influence of a cyclonic circulation over northwest Bay of Bengal, off Odisha - West Bengal coasts, a low pressure area formed over the North Bay of Bengal and neighbourhood. The system intensified into a Depression and was centered near 21^0 N, 89.5^0 E, at 0300 UTC of 29th May 2013. Moving north-northwestwards, the depression's center was near 21.7^0 N, 88.8^0 E, at 1200 UTC of 29th May 2013. The depression crossed the West Bengal coast near 21.8^0 N, 88.7^0 E between 1330 and 1430 UTC on the same day. The above mentioned depression caused heavy to very heavy rainfall during its passage over the Indian region.

Case 2 - Land Depression (20July - 24July 2014)

A land depression manifested with its center over the northeastern parts of Odisha and adjacent areas of Gangetic West Bengal on 21st July 2014 morning. The depression moved west-northwestwards and weakened into a well-marked low pressure area over west Madhya Pradesh and neighbourhood on 23rd July 2014.

Case 3 - Depression (29July - 01August 2013)

A low pressure area manifested with its center over the northeast Bay of Bengal and adjacent Bangladesh and coastal areas of West Bengal on 29 July 2013. The system

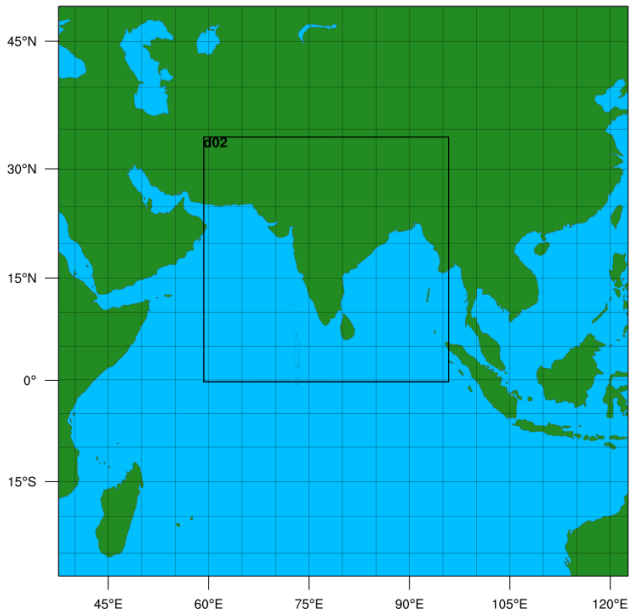

Figure 4.8: Model Domains used in the study

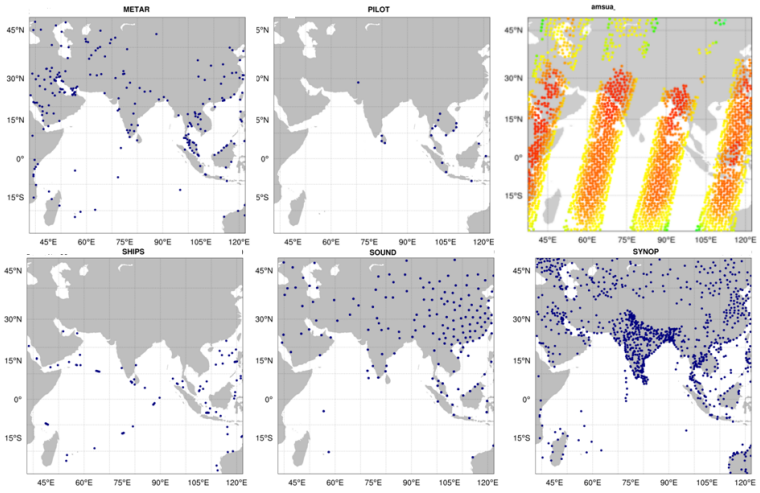

Figure 4.9: Observations available typically over the domain.

intensified into a Depression with its center near 21^0 N, 88^0 E. The depression moved northwestwards subsequently and crossed Odisha coast at 1200 UTC of 30th July. The depression further moved westwards and weakened into a low pressure system by August 1st 2013.

4.5 Experimental

The Advanced Research WRF (ARW) model version 3.5.1 is configured with two-way nested domains as shown in Fig.4.8. The outer domain has 350 $\times$ 350 grid

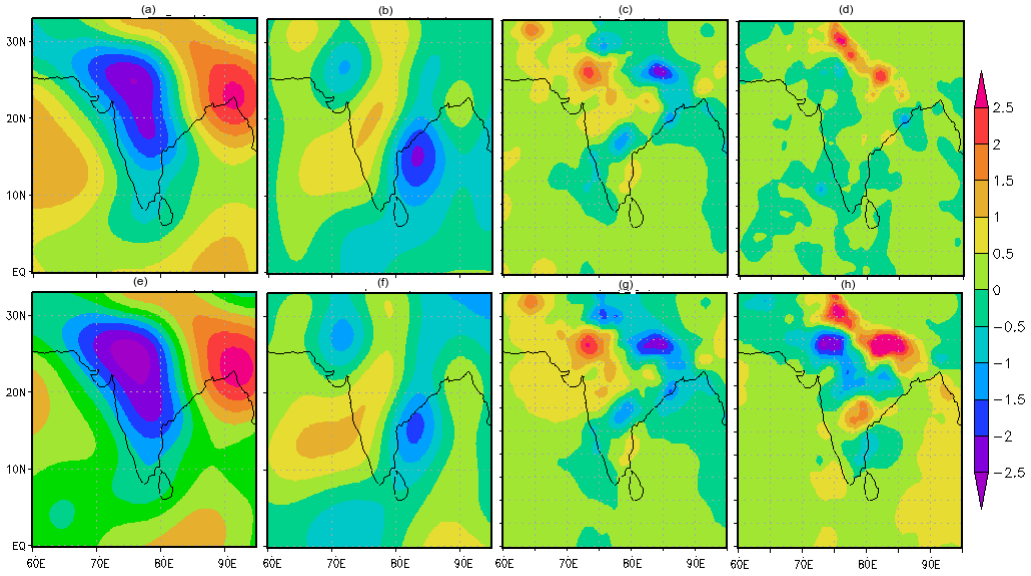

Figure 4.10: Analysis increment at model level 1 in zonal wind (a,e) meridional wind (b,f), temperature(c,g) and water vapour mixing ratio (d,h) for cv5 option(a-d) and cv6 option(e-h) for the depression case 1 due to assimilation of all available observations

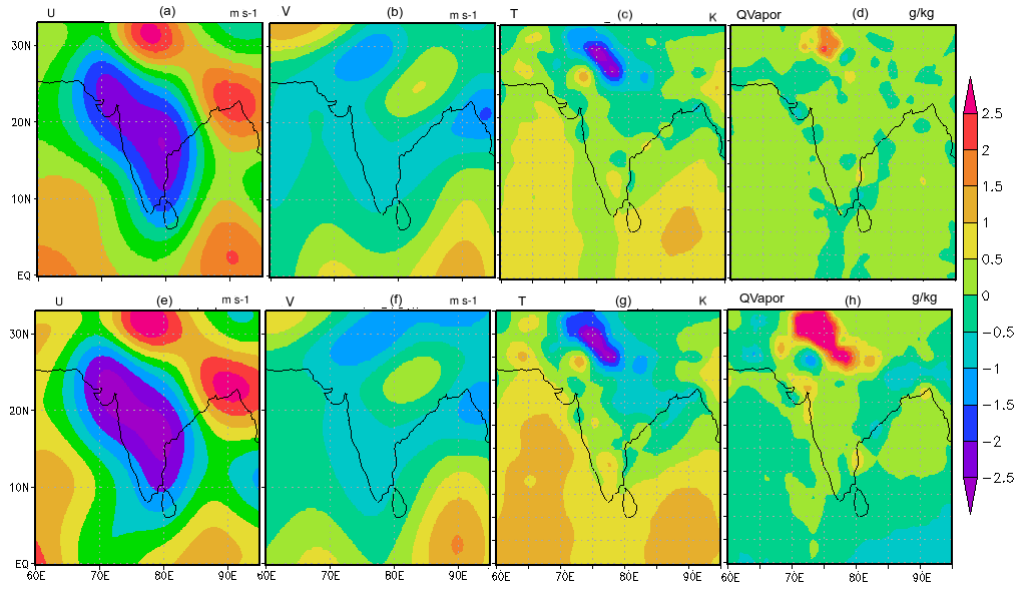

Figure 4.11: Analysis increment at model level 1 in zonal wind (a,e) meridional wind (b,f), temperature(c,g) and water vapour mixing ratio (d,h) for cv5 option(a-d) and cv6 option(e-h) for the depression case 2 due to assimilation of all available observations

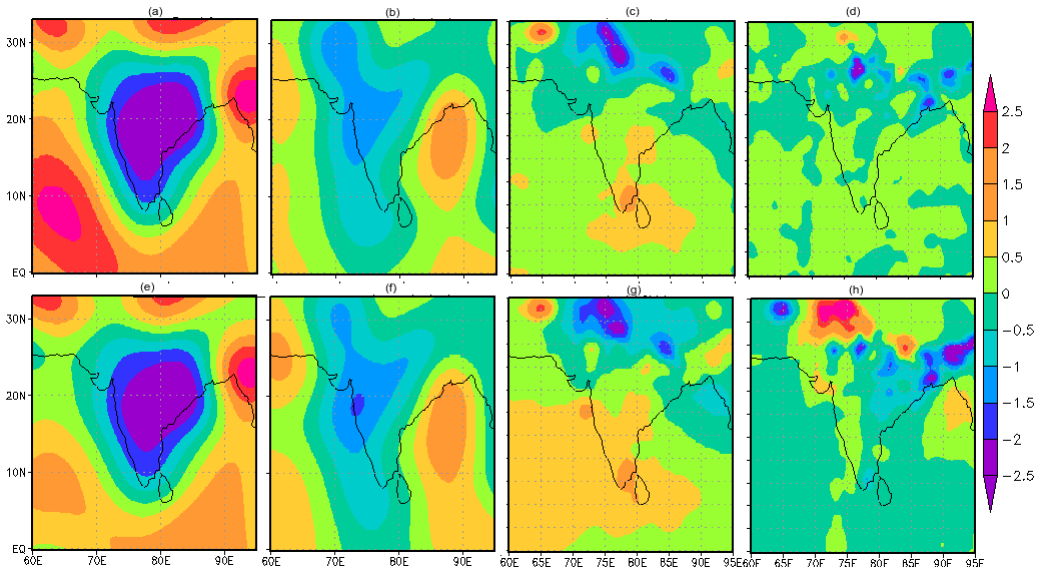

Figure 4.12: Analysis increment at model level 1 in zonal wind (a,e) meridional wind (b,f), temperature(c,g) and water vapour mixing ratio (d,h) for cv5 option(a-d) and cv6 option(e-h) for the depression case 3 due to assimilation of all available observations

cells in the east-west and north-south directions with a horizontal resolution of 27 km while the inner domain is configured with 451×451 grid points with a horizontal resolution of 9 km. There are 36 vertical levels in both the domains. This study uses the Kain-Fritsch scheme for cumulus, RRTM for longwave radiation, Dudhia scheme for shortwave radiation, Unified NOAH land surface model for land surface, YSU scheme for planetary boundary layer and WRF single-moment 3-class simple ice scheme for microphysics parametrization. All the results discussed in this study are from the inner, higher resolution domain.

The initial and boundary conditions for the model are obtained from the National Centre for Environmental Prediction (NCEP) Global Forecast System (GFS) forecast fields of horizontal resolution $0.5^0 \times 0.5^0$. Data assimilation is performed in the outer domain only for this study. For estimating the $\mathbf{B}$ matrix using the NMC method, 12hr and 24 hr forecasts are generated for a period of one month during the 2013 summer monsoon period (01June 2013 to 01 July 2013) for the outer domain. In WRF 3DVar system, the forecast differences of 24hr and 12hr forecast times are typically used for calculating background error using NMC method in the regional domain (Skamarock et al., 2008). The forecast perturbations generated using the differences of these forecasts are used for calculating BE covariances using both cv5

and cv6 options. The conventional surface and upper air observations from the Research Data Archive, Computational and Information Systems Laboratory (CISL) of NCAR (http://rda.ucar.edu/datasets/ds337.0) are utilized here for assimilation. Also, Advanced Microwave Sounding Unit-A (AMSU-A) radiances (Baker et al., 2005; Singh et al., 2012c) are assimilated in this study (shown in Fig.4.9. In all three depression cases investigated here, WRF model is integrated for 12hr without any assimilation for model spin-up. Using this model forecast as background, 3DVar analysis is performed using both cv5 and cv6 options at one time followed by free forecast (no assimilation of observations) for the next 48 hours . Fig.4.9 shows the all observations typically available in the domain at the assimilation time. In order to bring out clearly the effect of data assimilation, a control (CTRL) run is performed with no assimilation of observations.

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

4.5.1 Results and Discussion

Fig.4.10-4.12, shows the analysis increment for cv5 and cv6 options at the model level 1 for zonal wind (a and e), meridional wind ((b and f), temperature(c and g) and water vapor mixing ratio (d and h) with all observations assimilated for the three depression cases investigated here. Assimilation of all available observations (conventional and AMSU-A radiances) has yielded different impacts in cv5 and cv6 cases as seen from the differing analysis increment patterns. The increments in u-wind shows a larger positive region over the oceans in the cv5 run as compared with the cv6 run. This indicates that the strength of the zonal wind has increased due to assimilation in the cv5 run as compared with cv6 run. While the increment patterns in meridional winds and temperature at the surface level are fairly similar, assimilation of all obser-

vations has impacted the water vapour mixing ratio differently in cv5 and cv6 cases. Additional correlation information in the $\mathbf{B}$ matrix formulation has resulted in larger increments in cv6 option as compared with cv5 option. The above results indicate that differing formulation of $\mathbf{B}$ matrix indeed impacts the assimilation of observations in WRF 3DVar system.

Table 4.1: Track error in km of the three depression cases with respect to IMD data

	Case 1			Case 2			Case 3		
Forecast hour	CTRL	cv5	cv6	CTRL	cv5	cv6	CTRL	cv5	cv6
00	414.9	73.14	82.2	137.1	110.0	110.0	230.9	41.3	139.6
06	320.4	113.5	122.5	74.9	19.8	22.9	422.1	494.4	512.5
12	165.3	123.9	132.9	257.4	239.3	293.6	360.3	360.3	378.3
18	120.8	159.3	168.4	368.6	434.5	440.8	449.9	391.5	499.9
24	118.04	253.2	271.3	515.9	390.7	506.9	675.9	594.5	543.0
30	156.9	331.7	313.6	499.2	219.1	508.3	741.9	588.2	543.0
36	170.2	390.1	335.9	419.2	15.2	545.8	616.7	499.2	472.1
42	222.81	545.6	464.3	178.1	96.8	241.3	812.0	260.8	703.6
48	29.4	601.9	229.7	470.0	261.4	406.1	813.4	786.3	668.9

The impact of assimilation on the structure and movement of the depressions is investigated by estimating the depression track errors for a 48hr forecast period with respect to India Meteorological Department (IMD) track reports. The track errors for CTRL, cv5 and cv6 runs are shown in Table 4.1. In general both cv5 and cv6 runs perform better than CTRL run in terms of lower track errors. Here, cv5 run has lower track errors as compared to cv6 run. From Fig.4.10-4.12 we can infer that the analysis increments between cv5 and cv6 options are significantly different for the watervapor mixing ratio only. Therefore, it may be unrealistic to expect that the simulation of mean sea level pressure pattern and center of depression and the hence the track errors will be improved by utilising the cv6 option instead of the cv5 option.

Monsoon depressions are known to have cold core lows at the lower levels and warm core highs at the upper levels. It is known that systems with cold core at lower levels typically intensify with increasing height indicating that the monsoon

depression has higher intensity at lower to mid-troposphere (Godbole, 1977; Sikka, 1977). To investigate how the various experiments have simulated this observed thermal structure of depressions, the profile of the temperature anomaly has been calculated over the depression centre at the initial time. For this a $3^o \times 3^o$ box has been considered around the depression centre. The difference of average temperature in this box with that outside it is calculated. Fig. 4.13(a,b,c) shows the vertical profile of the temperature anomaly at the analysis time at the depression centre for the case1, case2 and case3 depressions respectively. All assimilation runs have simulated the thermal structure of the depression well with respect to the high resolution analysis as compared with the CTRL run, in all the three cases. However, as compared with the ECMWF analysis, the simulated temperature anomalies have stronger cold cores at lower levels. For the depression case1, the ECMWF analysis reveals a system whose cold core weakens with height at the lower levels indicating a weakening of the system's intensity with height at the lower levels. This is at variance with the known thermodynamic structure of the monsoon depressions. In all the three depression cases, cv6 runs have simulated a relatively stronger cold core at the lower levels, as compared to CTRL run, cv5 run as well as the ECMWF and high resolution analyses. This indicates that the thermodynamic structure of the monsoon depressions are modified by the cv6 analysis.

Fig.4.14a-4.14c shows the vertical profile of moisture divergence at the centre of the depression cases at the initial time of the forecast (+0hr forecast). Fig. 4.14a shows that cv6 run has contributed to higher moisture convergence values that are closer to the high resolution analysis in the lower levels at the depression centre as compared with both the CTRL and cv5 runs for first depression case. In the second depression case shown in Fig.4.14(b), cv6 run shows higher moisture convergence value close to the surface and relatively higher divergence values in the upper levels as compared with cv5 run. Fig. 4.14(c), which shows the third depression case indicates that surface moisture convergence values of cv6 are closer to the high resolution analysis. The cv6 run simulated moisture convergence is closer to the ECMWF analysis as compared to cv5 run at low levels for the first depression case.

Fig. 4.15(a-c) shows the vertical profile of relative vorticity at the centre of the depression for all the depression cases at analysis time of the cv5/cv6 sensitivity experiments (+0hr forecast). For the second and third depression cases, at lower levels,

the relative vorticity values for the cv6 run are closer to the high resolution analysis as compared to cv5 and CTRL runs. The difference in relative vorticity profiles between cv5 and cv6 options are not marked for all the depression cases. However, CTRL run simulated relative vorticity profile for the first depression are very far from the analysis values at all levels. ECMWF analysis of the relative vorticity profile for the first depression case also shows a weaker vortex at all the levels.

Vertical profiles of horizontal divergence over the depression centre at analysis time of the cv5/cv6 sensitivity experiments (+0hr forecast) for all the three depression cases are shown in Fig. 4.16(a-c) respectively. For the first and third depression cases considered here, the cv6 option has simulated stronger convergence at the lower levels than the cv5 option. In the first case the simulated profile of horizontal divergence is closer to the ECMWF analysis.

Fig 4.17(a-c) shows the vertical profiles of relative humidity at analysis time of the cv5/cv6 sensitivity experiments over the depression centre for all the three depression cases. For the first depression case, the CTRL run simulates relatively drier atmosphere till the mid troposphere while the assimilation runs are closer to the high resolution analysis and are more moist. ECMWF analysis also indicates a drier atmosphere. This is consistent with the warmer core at around 900hPa present in the ECMWF analysis. The ECMWF analysis also shows large relative humidity values for all depressions at the upper levels. For the first and second depression cases, cv6 option simulates higher values of relative humidity at lower levels as compared to the high resolution analysis as well as CTRL and cv5 runs. In the case of the third depression, despite having lower relative humidity values at the initial time at lower levels, relative humidity simulated by cv6 run is closer to the high resolution analysis and the ECMWF analysis as compared to cv5 and CTRL runs, especially in the middle levels.

The differences between the vertical profiles of say, a meteorological variable at the center of depression between the cv5 and cv6 runs with respect to the CTRL run at each model level at the beginning of the free forecast period are calculated. To estimate whether the assimilation with the cv6 option has significantly altered the model output fields, a student's t test (Wilks, 2011) has been performed by considering these differences. It is found that the differences in the vertical profiles of moisture convergence and relative vorticity between the reference (CTRL) run and the cv5/cv6

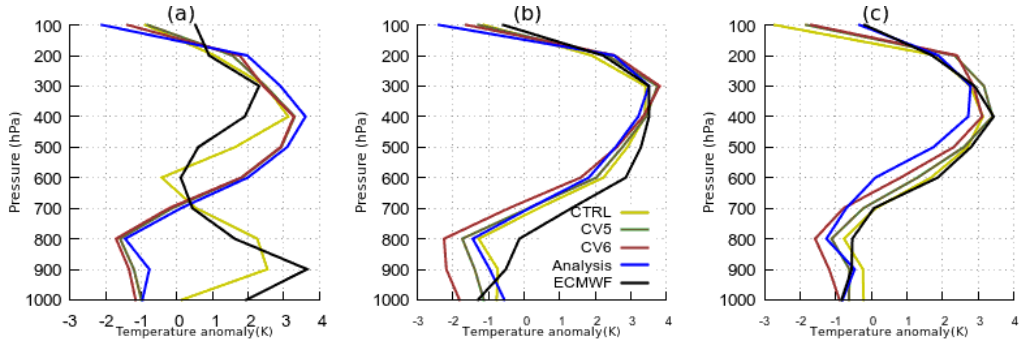

Figure 4.13: Vertical Profiles of temperature anomaly (a,b,c) at the analysis time of the cv5/cv6 sensitivity experiments over the depression centre for case1 (a), case2 (b) and case3(c) depressions

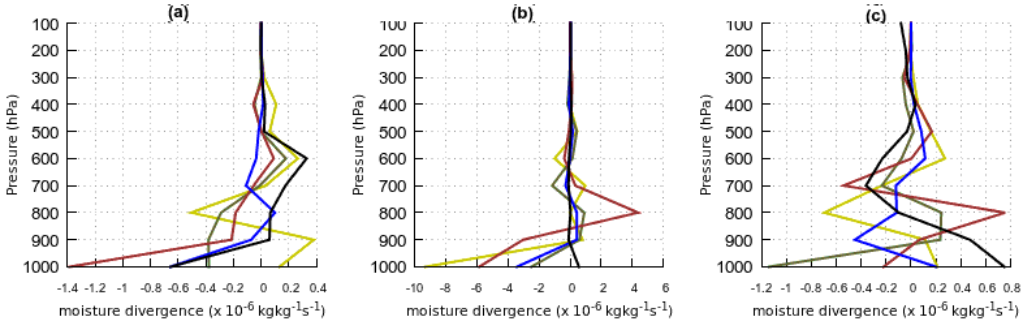

Figure 4.14: Vertical Profiles of moisture divergence (a,b,c) at the analysis time of the cv5/cv6 sensitivity experiments over the depression centre for case1 (a), case2 (b) and case3(c) depressions

analyses are significant at 95% confidence level. The differences in relative humidity are found to be 95% significant over the first fifteen model levels only. However, the divergence profiles are found to be statistically significant at 65% level only.

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

(i) Stronger cold core low at lower levels (upto 800hPa) for all the three depressions

(ii) Stronger low level moisture convergence for two of the three cases at initial time

(iii) Stronger low level horizontal wind convergence for two of the three depressions

(iv) Higher values of relative humidity at lower levels in two out of three depressions.

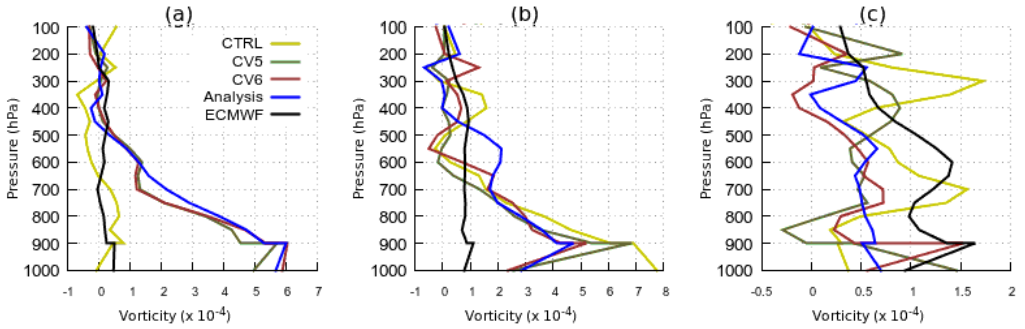

Figure 4.15: Vertical Profiles of relative vorticity (a,b,c) at the analysis time of the cv5/cv6 sensitivity experiments over the depression centre for case1 (a), case2 (b) and case3(c) depressions

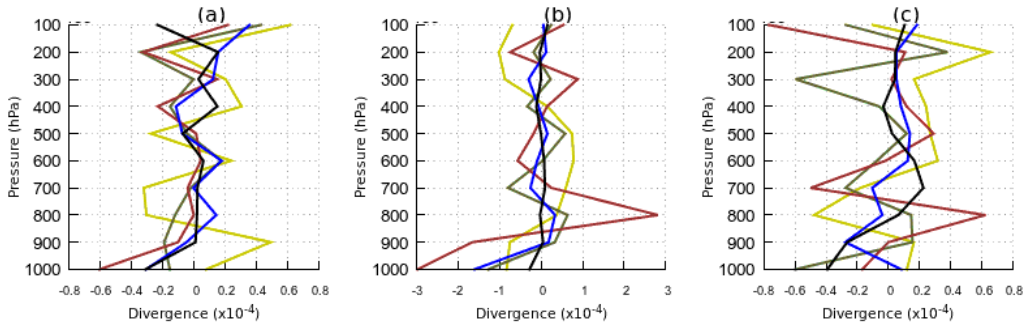

Figure 4.16: Vertical Profiles of horizontal divergence (a,b,c) at the analysis time of the cv5/cv6 sensitivity experiments over the depression centre for case1 (a), case2 (b) and case3(c) depressions

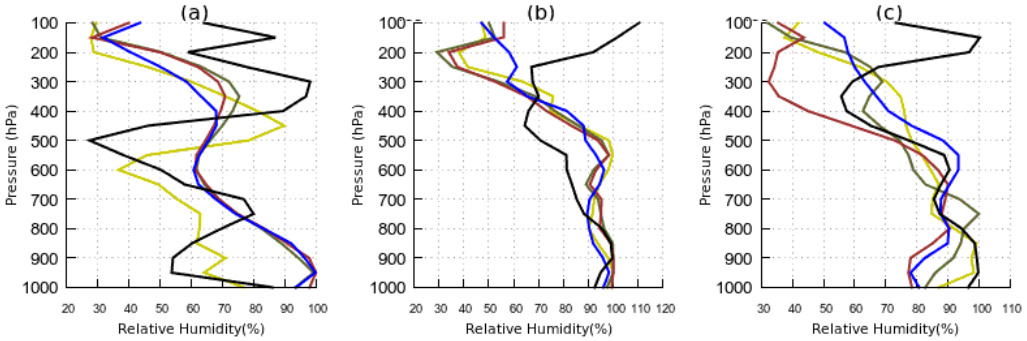

Figure 4.17: Vertical Profiles of relative humidity (a,b,c) at the analysis time of the cv5/cv6 sensitivity experiments over the depression centre for case1 (a), case2 (b) and case3(c) depressions

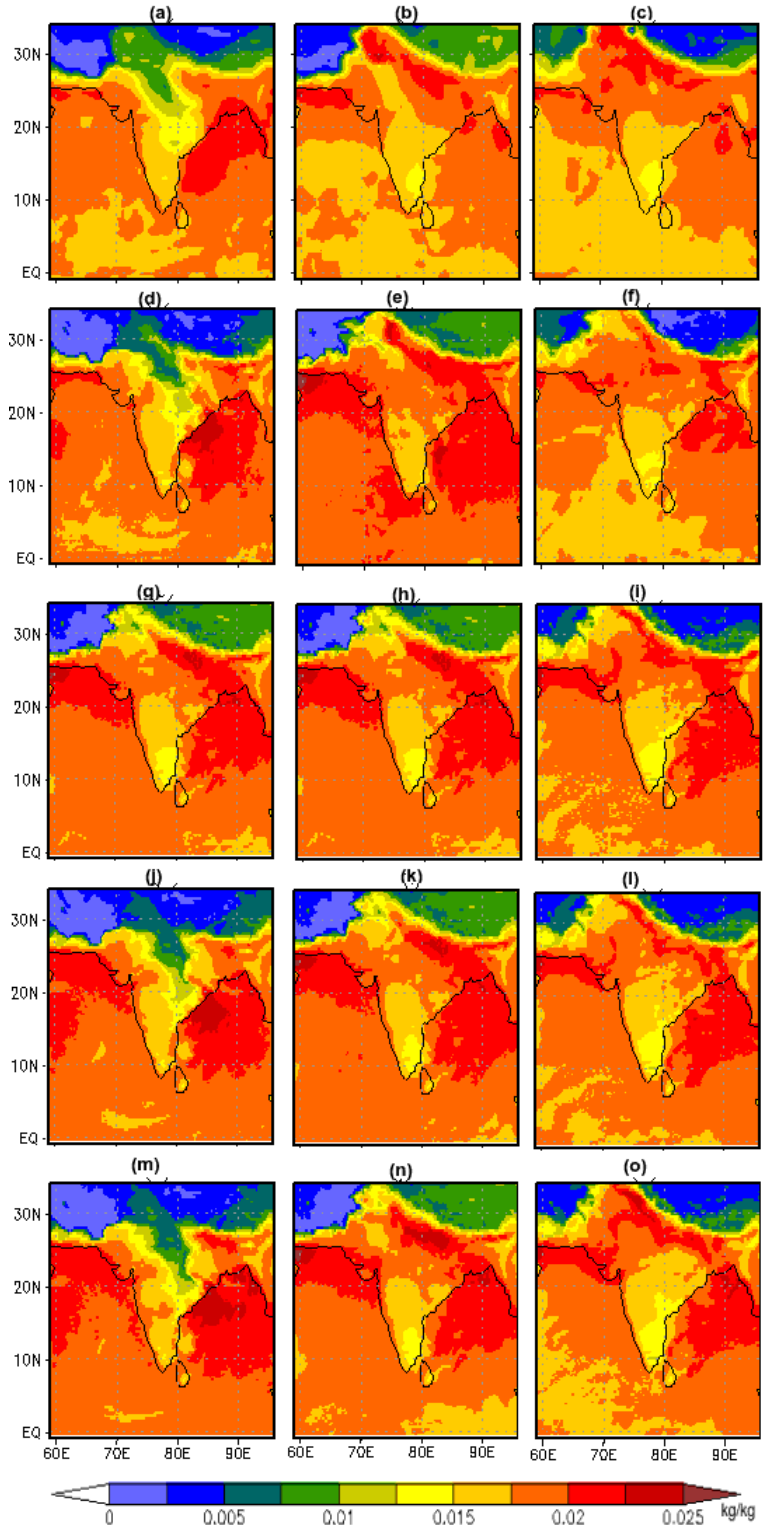

Figure 4.18: Water vapour mixing ratio at level 1 compared with the ECMWF analysis at the analysis time of the cv5/cv6 sensitivity experiments at level 1 for case 1(left panel) , case2(middle panel), case3(right panel) depressions for ECMWF analysis (a-c), high resolution analysis (d-f), CTRL run (g-i), cv5 run(j-l) and cv6 run(m-o)

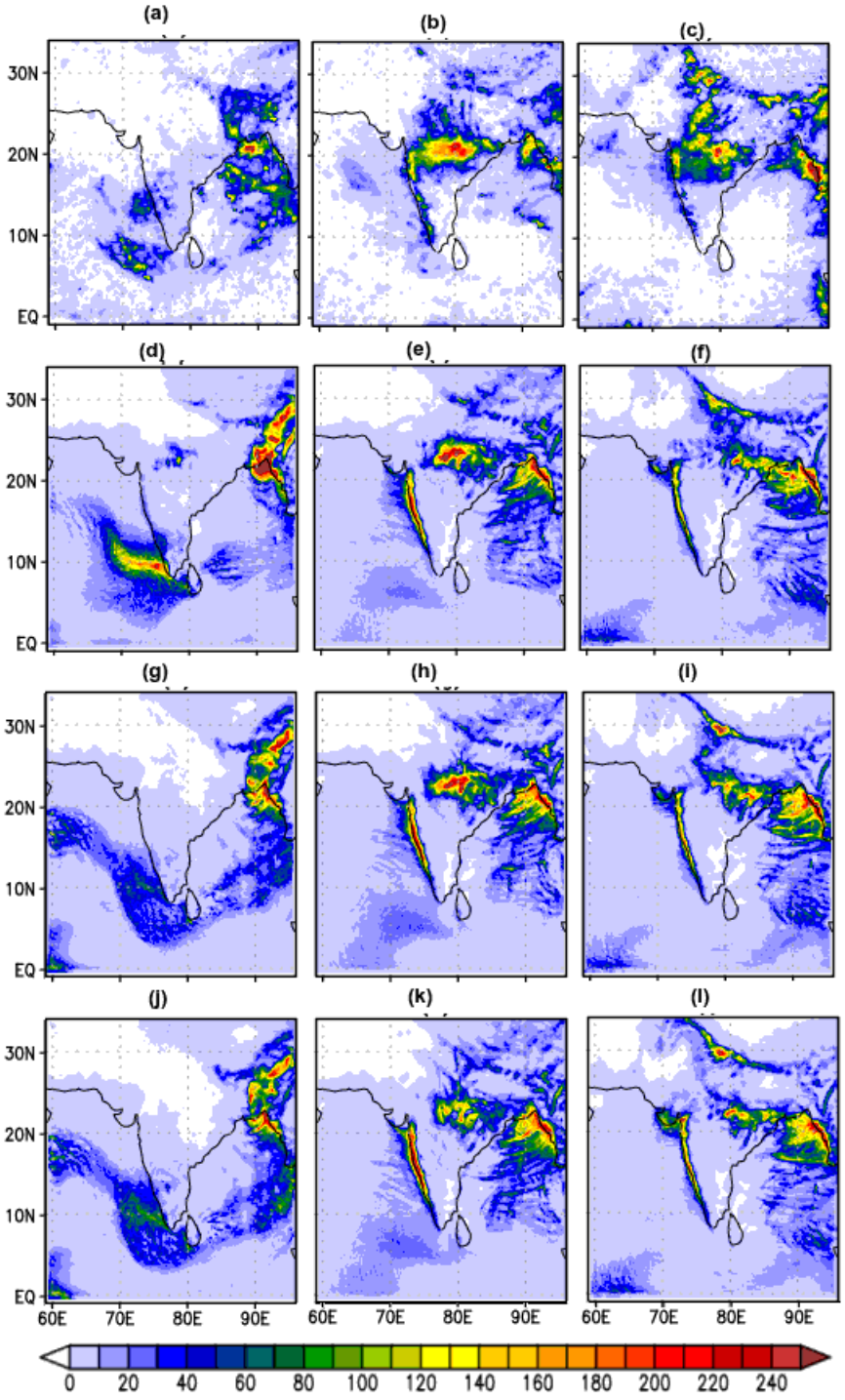

Figure 4.19: 48hr accumulated precipitation for case 1 (left panel), case2(middle panel) and case3(right panel) depressions from TRMM (a-c), CTRL run(d-f), cv5 run(g-i) and cv6 run(j-l)

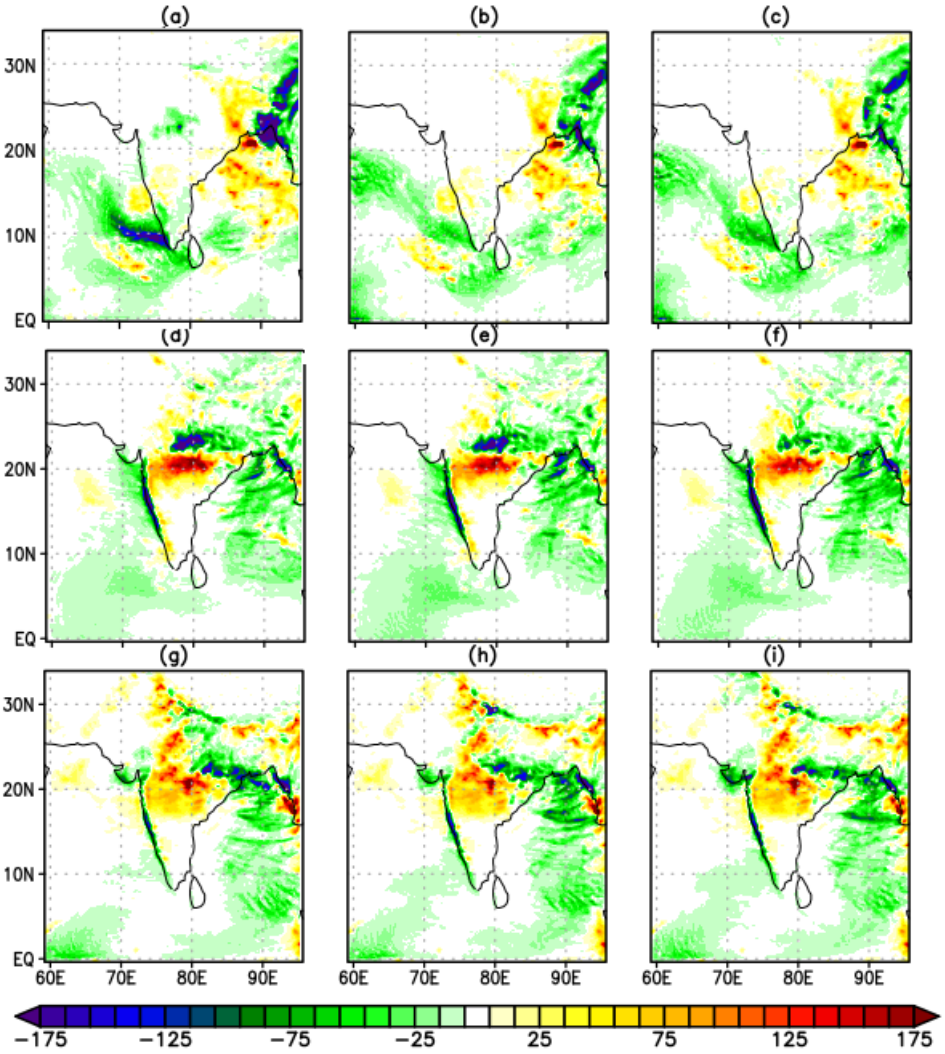

Figure 4.20: Differences in 48hr accumulated precipitation (in mm) from TRMM rainfall observation for case 1(a-c) , case2(d-f), case3(g-i) for CTRL (a,d,g), cv5(b,e,h), cv6(c,f,i) runs

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

Fig.4.18 shows the surface level water vapour mixing ratio of CTRL run(g-i), cv5(j-l) and cv6 run(m-o) compared with the ECMWF analysis(a-c) and the high resolution analysis(d-f) for all the three depressions. In general, all the model simulations indicate more moisture content than the ECMWF analysis. The cv6 run has simulated moderately higher mixing ratio values as compared with CTRL and cv5 runs over the land region in two out of the three cases considered. With more moisture content being simulated over land, both in horizontal and vertical directions along with larger horizontal moisture convergence, the cv6 run is expected to simulate rainfall better than the cv5 and CTRL runs.

The model simulated 48hr accumulated rainfall is compared with TRMM rainfall observations to analyse the impact of $\mathbf{B}$ formulation on rainfall simulation in Fig.4.19. The accumulated precipitation from TRMM (a-c), CTRL run (d-f), cv5 run(g-i) and cv6 run (j-l) for the three depressions considered here are shown in Fig.4.19(a-l). The left panel (Fig.4.19(a,d,g,j)) shows the first depression case, the middle panel (Fig.4.19(b,e,h,k)) shows the second depression case and the right panel(Fig.4.19(c,f,i,l)) shows the third depression case, considered in this study. For all the three cases, accumulated rainfall values indicate heavy rainfall over the Indian region due to the monsoon depressions. In the first depression case, the model simulates excess rainfall over the Head Bay of Bengal and north eastern regions of India. However, the TRMM observations show that the location of maximum precipitation is over the Head Bay of Bengal. Furthermore, the CTRL run simulates erroneous rainfall maximum off the west coast of India, which is not seen in the TRMM observations as well as in the assimilated runs. All the model runs show similar spatial pattern of precipitation for the second depression case. However the intensity of rainfall is different in the cv6 run as compared with the CTRL and cv5 runs. For the third depression case, the observed rainfall maximum over oceans is reproduced by all the model runs in terms of location and intensity. However, both cv5 and cv6 runs simulate rainfall closer to TRMM ob-

servations as compared to the CTRL run. The differences of 48hr accumulated rainfall simulated by the three model runs with respect to the TRMM observation are shown in Fig.4.20. For all the depression cases, the maximum differences between TRMM observation and model simulations in CTRL, cv5 and cv6 runs occur over the regions where TRMM observations show maximum accumulated rainfall.

Table 4.2: Location and intensity of maximum precipitation(cm) as obtained from TRMM for all the three depressions together with locational and magnitude error of maximum precipitation for CTRL, cv5 and cv6 runs

	TRMM		CTRL		cv5		cv6	
	Location of max. rainfall	magnitude (cm) of max. rainfall	location error(km) of max precipitation	magnitude error (cm) of max precipitation	location error(km) of max precipitation	magnitude error (cm) of max precipitation	location error(km) of max precipitation	magnitude error (cm) of max precipitation
Depression 1	89.13^0E, 20.88^0N	35.7	1173	-11.8	257	-4.9	146	-4.4
Depression 2	80^0E, 21.25^0N	24.1	218	-21.6	211	-11.4	183	-2.1
Depression 3	79.65^0E, 20.88^0N	29.9	1325	-13.5	937	-15.2	943	-13.3

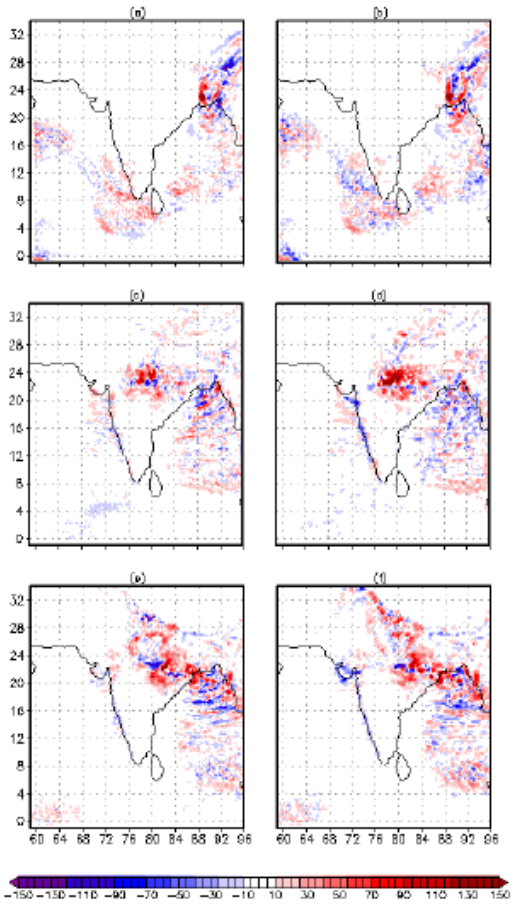

Figure 4.21: Improvement parameter of 48 hr accumulated rainfall for case 1 depression (a-b) case 2 depression (c-d) and case 3 depression (e-f) for CV5 run (a,c,e) and CV6 run (b,d,f)

Table 4.2 gives the location and magnitude of maximum 48hr accumulated rainfall from TRMM observations as well as the location error (in km) and magnitude error (in cm) in the model simulations of the same for CTRL, cv5 and cv6 runs. The results show that in all the depression cases considered here, the model simulated location and intensity of maximum accumulated rainfall has errors when compared with TRMM observations. It is also found that cv6 run has lower error in the location as well as magnitude of maximum rainfall when compared with both CTRL and cv5 runs for all the three depression cases.

Fig.4.21 indicates the improvement parameter (η) of 48 hour model simulated rainfall in the CV5 and CV6 runs. Postive value of IP indicates that the assimilation run has improved the simulation of rainfall as compared to the control run. Here, the IP of rainfall for CV5 and CV6 runs in the three depression cases are compared. In the three

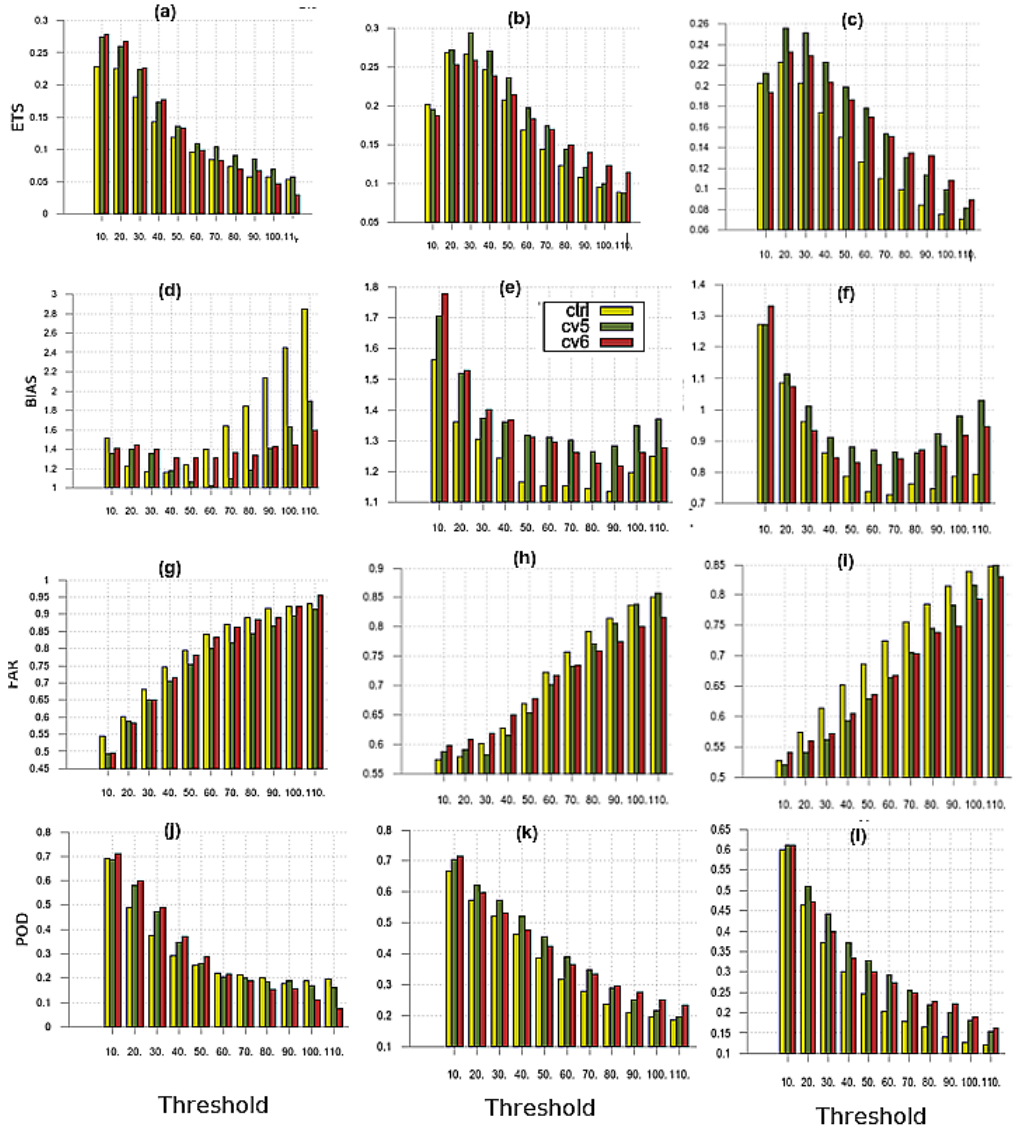

Figure 4.22: Skill scores of 48hr accumulated precipitation in mm w.r.to TRMM for case 1 in left panel (a,d,g,j), case2 in middle panel (b,e,h,k) and case3 in right panel (c,f,i,l) depressions. Equitable threat score is shown in Fig. a-c . Bias scores in Fig.d-f , False alarm ratio in Fig.g-i and Probability of Detection in Fig.j-l

cases, CV6 run produces a larger value of IP over the land, in general, as compared with CV5 run. This indicates that the CV6 run has in general improved the rainfall simulation over a larger number of grid points over land, as compared with the CV5 run. However, over the west coast of India as well as over the ocean, there are more number of grid points having negative values for IP in the CV6 run as compared to CV5 run.

For quantitative verification of rainfall forecast skill, various skill scores like equitable threat score (ETS) , Bias score , False Alarm Ratio (FAR) and Probability of Detection (POD) have been calculated for 48hr accumulated rainfall with respect to TRMM observations for all the three depression cases and are shown in Fig4.22. Towards this, model simulated 48hr accumulated rainfall were regridded to the resolution of TRMM rainfall observations. The skill scores are calculated using the contingency table considering whether a forecast occurs or not (Wilks, 2011). The ETS estimates how well the observed event is forecast, taking into account the correct forecasts that can occur by chance. Bias score estimates the ratio of frequency of forecast events to the frequency of observed events, indicating whether there is over or under prediction by the model. The FAR gives the fraction of false alarms (model simulated rainfall that is not observed) and POD gives the fraction of correctly forecast events. In general, assimilated runs perform better than CTRL run in all rainfall thresholds with higher ETS values, lower false alarm ratio and higher probability of detection in all the three cases. In the second and third depression cases, cv6 runs have higher ETS score, lower bias , lower FAR and larger POD as compared with cv5 and CTRL runs in the higher threshold regions. The higher ETS score, higher POD and lower FAR scores for all cases in the high rainfall thresholds, also supports the earlier inference (refer Table 4.2) that cv6 runs perform better than the CTRL and cv5 runs for higher intensity rainfall.

4.6 Summary

This chapter explores the two formulations of $\mathbf{B}$ matrix, namely, cv5 and cv6 options in the WRF 3DVar assimilation system. It is seen that the formulation of $\mathbf{B}$ matrix do significantly impact the analysis. In the cv5 formulation, the background

error correlations between humidity and other variables are not included. Inclusion of the unbalanced velocity potential related correlations and humidity related correlations leads to multivariate analysis of huimdity in the cv6 formulation. Improved analysis of humidity is significant in improving the simulation of convective systems. Furthermore, temperature and surface pressure are influenced by the divergent component of wind in the cv6 option unlike in the cv5 option. The correlations specified in the background error formulation results in spread of observation information among the various model variables. The total number of humidity observations available over a convectively active region may be less. Furthermore, the quality of the available humidity observations will also be less over such regions. Hence the multivariate correlations in the humidity variable will ensure that the information from other observations improve the quality of humidity analysis for the simulation of convective systems.

The impact of assimilation of conventional and satellite observations using the above two methods are compared in the simulation of three cases of monsoon depressions. The formulation of $\mathbf{B}$ matrix does have an impact on the analysis fields. The moisture field is seen to be impacted the most by using the cv6 option as compared with the cv5 option. The cv6 option has simulated a stronger cold core at lower model levels as compared with the cv5 run for the depression cases considered here. Vertical profiles of moisture convergence indicate stronger moisture convergence values for two of the depressions at low levels in cv6 run as compared with the cv5 run. Higher relative vorticity and horizontal divergence values also are seen at the lower levels in cv6 option for two cases. Furthermore, in two of the three depression cases considered here, skill scores estimated for 48hr rainfall forecasts are better in cv6 runs for the higher threshold levels as compared with the cv5 run and CTRL run. The location and magnitude of maximum 48hr accumulated rainfall is also better simulated in cv6 run when compared with cv5 run as well as CTRL run.

CHAPTER 5

Impact of 3DVar assimilation of SAPHIR radiances using multivariate background error covariances in the simulation of tropical cyclones over Bay of Bengal

5.1 Introduction

The cv6 option in WRF 3DVar includes additional correlations between mixing ratio field with other fields like temperature, divergent wind and surface pressure. Utilizing the above-mentioned cv6 option should lead to improved representation of moisture field in the analysis. Formation and development of tropical cyclones greatly depend on convection-driven physical processes. Hence, one would expect that the improved representation of moisture correlations in the 3DVar assimilation system would result in better analysis of initial fields ; and hence would contribute to improved simulation of tropical cyclone features.

The above hypothesis is evaluated in this study by employing the cv5 and cv6 options for $\mathbf{B}$ matrix in the 3DVar assimilation of MeghaTropiques SAPHIR radiances using the WRF model. Three tropical cyclones that formed over the Bay of Bengal in the recent years are selected in this chapter to investigate the influence of $\mathbf{B}$ matrix on the analysis fields as well as on the forecasts of the WRF model.

5.2 Model configuration

Fig.5.1 shows the model domain used in this study. The WRF ARW model is configured in a two-way nested domain. The outer domain has 350 grid cells in the east-west direction and 350 grid cells in the north-south direction with a horizontal resolution of 27 km. The inner domain has a horizontal resolution of 9 km with 451$\times$ 451 grid cells in the east-west and north-south directions respectively. Both domains have 36 levels in the vertical. Both the domains use Kain-Fritsch cumulus

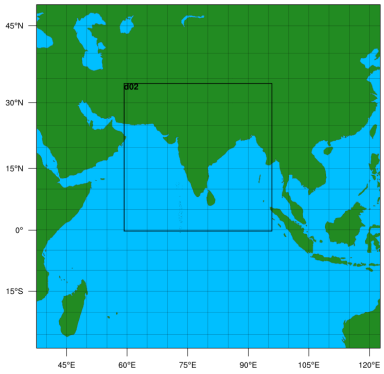

Figure 5.1: Model Domain used in this chapter

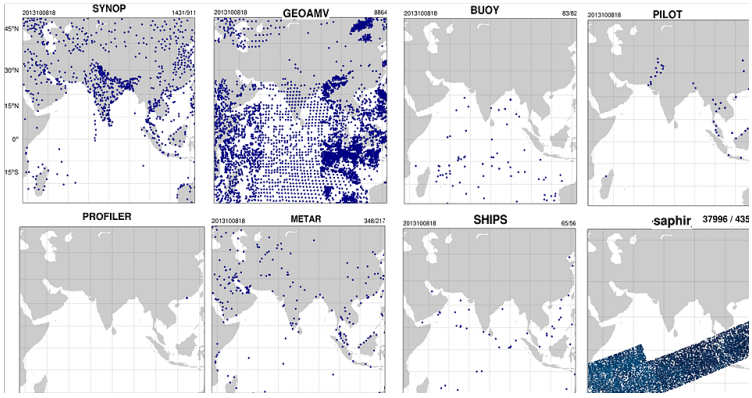

Figure 5.2: Distribution of typically available observations over the domain used for assimilation in this chapter

scheme with Yonsei University Scheme for planetary boundary layer, RRTM scheme for long wave radiation, Dudhia scheme for short wave radiation, WSM scheme for microphysics and Noah land surface model for land surface parametrisation.

5.3 Observations assimilated

The global surface and upper air observations available from NCEP include land surface, marine surface, radiosonde, pibal and aircraft reports from the Global Telecommunications System (GTS) and satellite winds (NCEP, 2008). These observations are assimilated in this study along with the radiance observations from SAPHIR sensor of the MeghaTropiques satellite. Fig.5.2 depicts the distribution of observations that are typically available over the domain, used for assimilation in this study.

Cyclone Cases Investigated

5.4 Cyclone Nilam (28 October - 01 November 2012)

5.4.1 Synoptic Conditions

A depression was formed over southeast and adjoining southwest Bay of Bengal near 9.50^{o}N, 86.00^{o}E at around 06Z of 28th October 2012. Moving westward, it deepened into a deep depression by 29th October 2012 near 9°N, 83.0°E. The depression moved further west and intensified into cyclonic storm Nilam. The cyclone then moved northnorthwestwards, crossed north Tamilnadu coast near Mahabalipuram on 31st October 2012. After experiencing landfall, the cyclonic storm, Nilam moved west-northwestwards and weakened gradually into a deep depression and then into a depression over south Interior Karnataka in the morning hours of 01st November 2012.

5.4.2 Experimental Details

In order to investigate the impact of formulation of $\mathbf{B}$ matrix on the simulation of cyclone Nilam, three numerical simulations have been conducted. These are called the control (CTRL), CV5 run and CV6 run. In the CTRL run, model is integrated from 12Z of 28th October 2012 to 31st October 2012 without assimilating any observations. In both the CV5 and CV6 runs, the model has been initialized at 12Z 28th October 2012 and underwent 6 hours of spin up integration till 18Z 28th October 2012. All available surface and upper air observations, satellite winds along with SAPHIR radiances are assimilated at 18Z 28th October 2012 using the 3DVar technique. Subsequently in both the CV5 and CV6 runs, the model is integrated without assimilating any observations till the tropical cyclone experienced landfall. In the CV5 run, the $\mathbf{B}$ matrix is estimated using the cv5 option and in CV6 run, it is estimated using the cv6 option.

5.4.3 Results and discussion

Fig.5.3 depicts the analysis increments at model level one at 18Z, 28th October 2012 after assimilating the conventional and satellite observations. The increments in the fields of zonal wind ,u (a,e), meridional wind,v (b,f) , temperature, T (c,g), water vapor mixing ratio q (d,h) for the CV5 run is shown in the top panel (a-d) and the same for the CV6 run is depicted in the bottom panel(e-h) . The results indicate that the differences between CV5 run and CV6 run in the analysis fields for T and q are most pronounced. Specifically, the analysis increments between the CV5 and CV6 options differ from each other mainly in the q-field (Fig.5.3(d,h)) . The results of the first model level shown in the Fig.5.3 indicates larger moisture increments in the CV6 analysis as compared with the CV5 analysis over the domain of interest. Considering the importance of moisture field in the evolution of intensity changes associated with tropical cyclones, it is hoped that employing the cv5 and cv6 options would lead to differences in the simulation of intensities associated with cyclone Nilam.

The vertical profile of moisture divergence field at the cyclone center at +0hr, +24hr and +48 hr forecasts of cyclone Nilam for the CTRL, CV5 and CV6 runs are shown in Fig.5.4 and are compared with the GFS analysis]. There are very little differences in the moisture convergence field at the cyclone center at the initial time between the CV5 and CV6 simulations. The large low level convergence indicated in the GFS analysis is not reproduced by the model runs. However, the results of the 24 hr simulation shows that stronger convergence values at the lower levels are seen in the CV5 simulation as compared with the CV6 run. The CV5 run produces more moisture convergence at the lower model levels as compared with the GFS analysis as well as CV6 run in the 24 hour simulation.

The horizontal divergence field at the cyclone center at +0hr, +24hr and +48 hr forecasts of cyclone Nilam for the CTRL, CV5, CV6 runs and GFS analysis are compared in Fig.5.5. At the initial time, as in the case of moisture divergence field, there is very little differences in the horizontal divergence fields between the CV5 and CV6 runs. Both CV5 and CV6 runs produce a divergence/convergence profile closer to the GFS analysis as compared with the CTRL run. However, the 24 and 48 hour forecasts indicate that the model simulations using the cv5 and cv6 options do differ in the simulation of horizontal divergence fields. The CV5 run has simulated greater

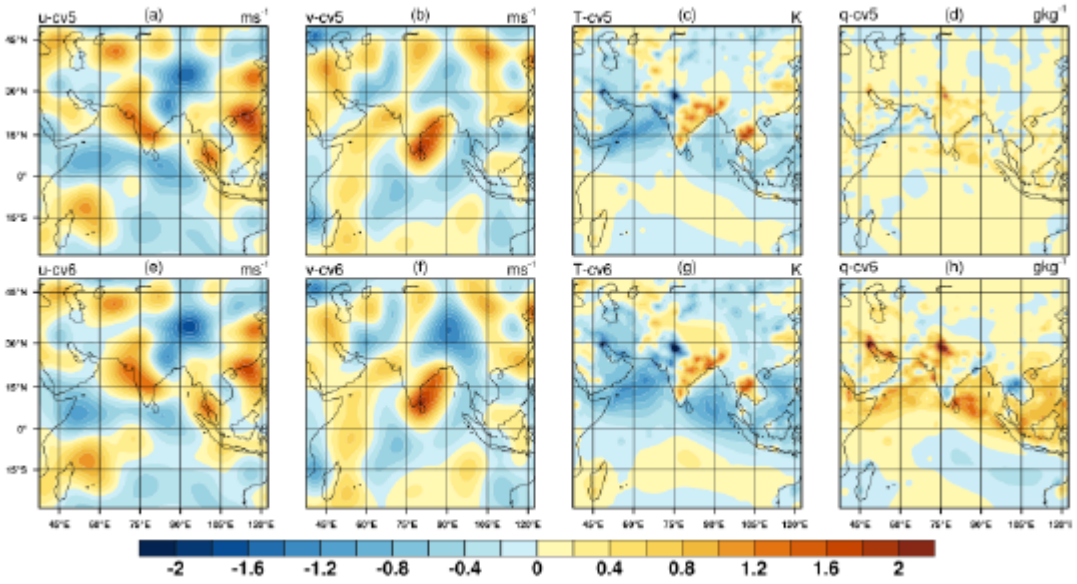

Figure 5.3: Analysis increment at model level 1 for zonal wind(a,e), meridional wind,v (b,f) , temperature, T (c,g), water vapor mixing ratio q (d,h) for CV5 run (a-d) and CV6 run (e-h) for cyclone Nilam

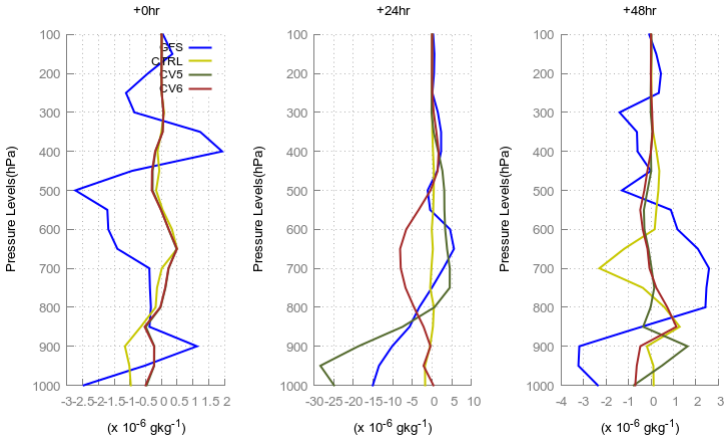

Figure 5.4: Moisture divergence at the cyclone center at +0hr, +24hr and +48 hr forecasts of cyclone Nilam for CTRL, CV5 and CV6 runs and GFS analysis

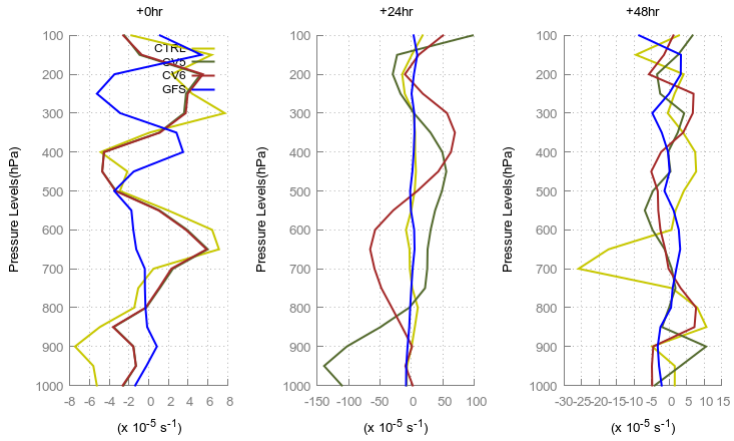

Figure 5.5: Horizontal divergence at the cyclone center at +0hr, +24hr and +48 hr forecasts of cyclone Nilam for CTRL, CV5 and CV6 runs and GFS analysis

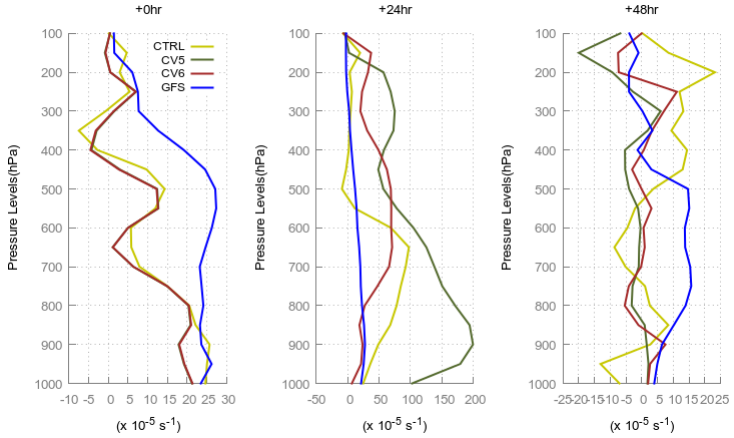

Figure 5.6: Relative vorticity at the cyclone center at +0hr, +24hr and +48 hr forecasts of cyclone Nilam for CTRL, CV5 and CV6 runs and GFS analysis

convergence/divergence at lower/upper model levels as compared with the CTRL and CV6 run as well as the GFS analysis in the 24 hour forecast.

Fig.5.6 depicts the vertical profiles of relative vorticity at the center of the cyclone Nilam at +0hr, +24hr and +48 hr forecasts for CTRL, CV5 and CV6 runs as well as GFS analysis. Here also, at +0hr of forecast there is no discernible differences in the profiles simulated in the CV5 and CV6 runs and all the model runs have simulated lower values of relative vorticity at the cyclone center as compared with the GFS analysis. However, the +24hr forecast shows that the relative vorticity at the cyclone center is more intense in the CV5 run at lower model levels as compared with that

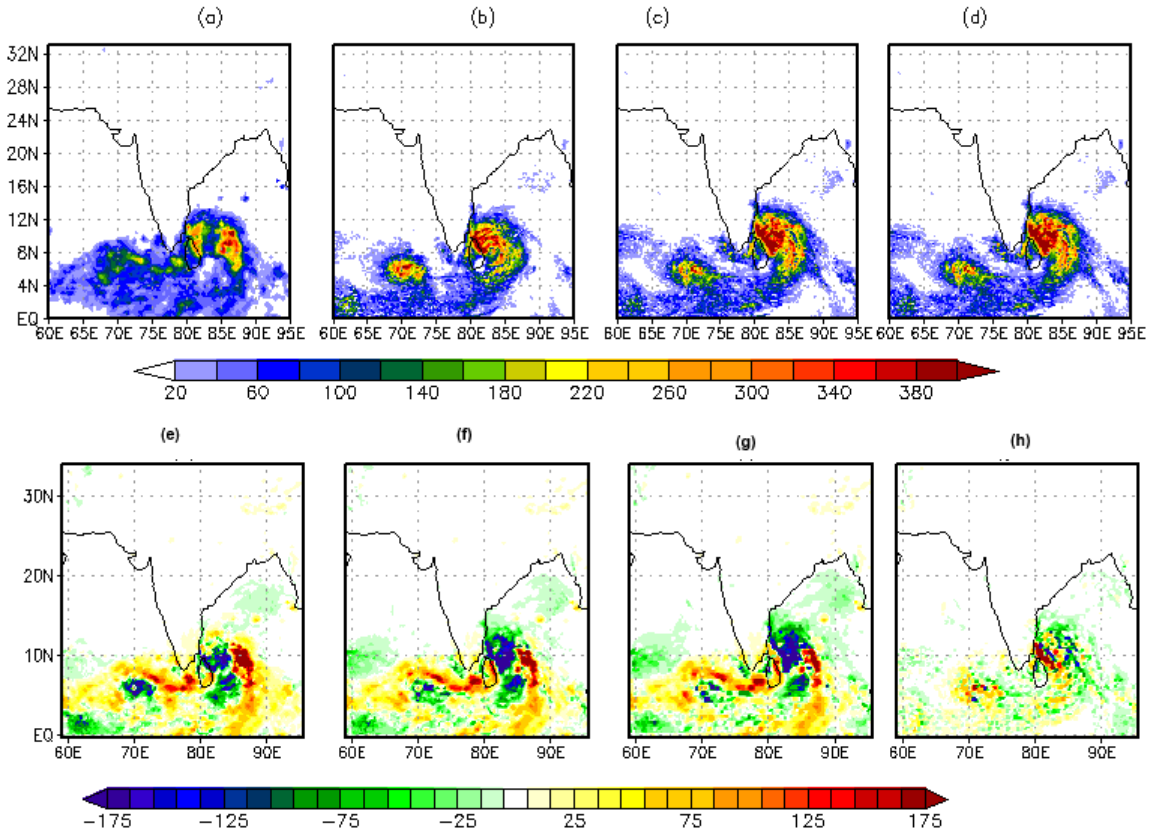

Figure 5.7: 48 hour accumulated precipitation observed from TRMM (a), simulated by CTRL (b), CV5 run (c) and CV6 run (d) for cyclone Nilam. The difference in accumulated rainfall of CTRL run from TRMM observations is shown in (e), difference between CV5 and TRMM in (f) and difference between CV6 and TRMM in (g) and the difference between CV5 and CV6 is shown in (h) for cyclone Nilam

in the CV6 runs. In the 24 hour forecast, CV5 run has simulated a stronger relative vorticity as compared with CTRL, CV5 runs as well as the GFS analysis.

Fig.5.7(a-d) shows the 48 hr accumulated rainfall as observed by TRMM (a) and simulated by CTRL (b), CV5 run (c) and CV6 run (d). The spatial extent of observed rainfall is simulated reasonably well by the model simulations. However, the model over-estimates the precipitation amounts as compared with the TRMM observations. Assimilation of SAPHIR radiances have resulted in the model simulating excess rainfall in both the CV5 and CV6 simulations. However, the difference between the 48 hour accumulated rainfall in the CV5 and CV6 simulations (Fig.5.7c and 5.7d) rainfall

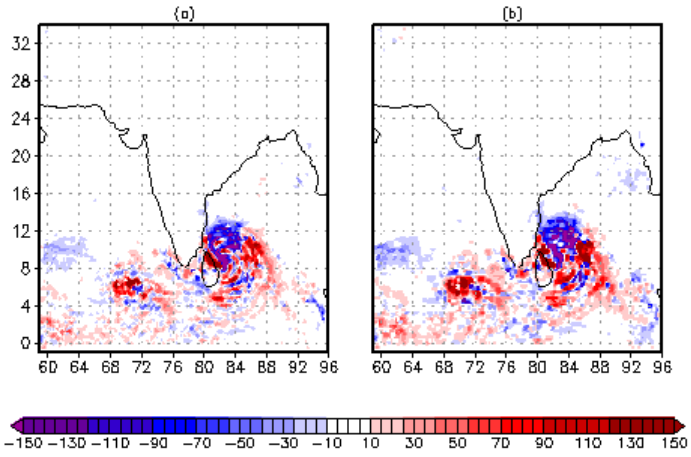

Figure 5.8: Improvement parameter of 48 hour accumulated rainfall simulated by CV5 (a) and CV6 (b) runs for cyclone Nilam.

is not easily discerned. Fig.5.7(e-g) shows the difference between the TRMM rainfall observations and simulated 48 hour accumulated rainfall by CTRL (fig.5.7(e)), CV5 run (fig.5.7(f)) and CV6 run (fig.5.7(g)). The difference between the 48hr accumulated rainfall simulated in the CV5 and CV6 runs in Fig.5.7(h) and the same indicates that the differences in the formulation of the $\mathbf{B}$ matrix does have an influence on the simulation of model rainfall for the case of cyclone Nilam.

Fig.5.8 indicates the improvement parameter (IP) of 48 hour accumulated rainfall simulated by CV5 (a) and CV6 (b) runs for cyclone Nilam. In general, both CV5 and CV6 assimilation runs have negative IP over the north-west region of the cyclonic system. Also, both CV5 and CV6 runs indicate that IP is positive for a larger number of grid points indicating that the assimilation of SAPHIR radiances have positive impact on the simulation of rainfall over more model grid points. The CV6 run has more number of grid points with positive IP as compared with the CV5 run in the case of cyclone Nilam.

The quantitative verification of rainfall simulation is performed by calculating the various skill scores with respect to the 48hr accumulated TRMM observations for cyclone Nilam. Fig.5.9 (a) shows the ETS, Fig.5.9 (b) shows Bias , Fig.5.9 (c) the FAR and Fig.5.9 (d) shows the POD for cyclone Nilam for CTRL, CV5 and CV6 runs. The skill scores also indicate that the rainfall simulated by CV5 and CV6 runs are indeed different. CV6 run provides for higher ETS, higher POD and higher bias as compared with the CV5 run.

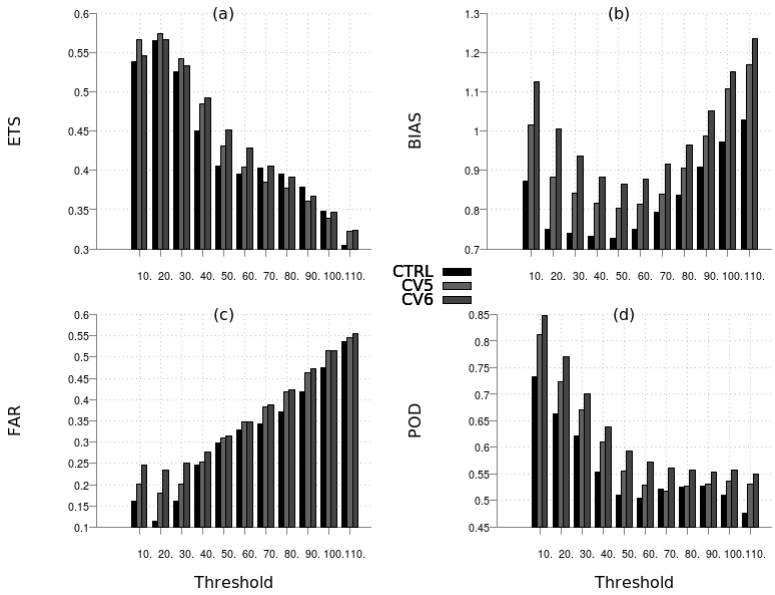

Figure 5.9: ETS (a), Bias (b), FAR (c) and POD (d) of CV5 and CV6 model simulations estimated with respect to TRMM observations for cyclone Nilam.

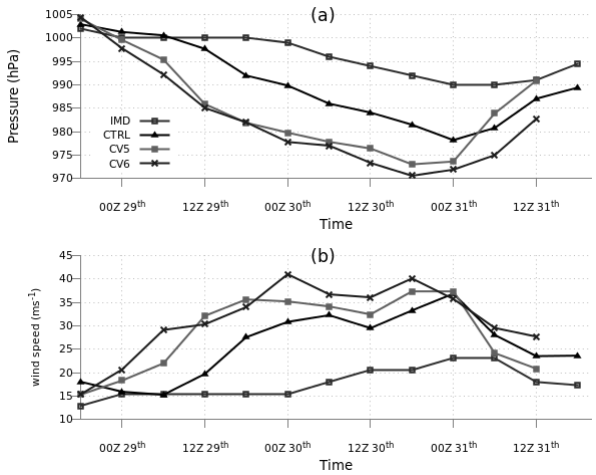

Figure 5.10: Time series of minimum slp (a) and maximum wind speed (b) of cyclone Nilam for CTRL, CV5 and CV6 runs compared with IMD observations

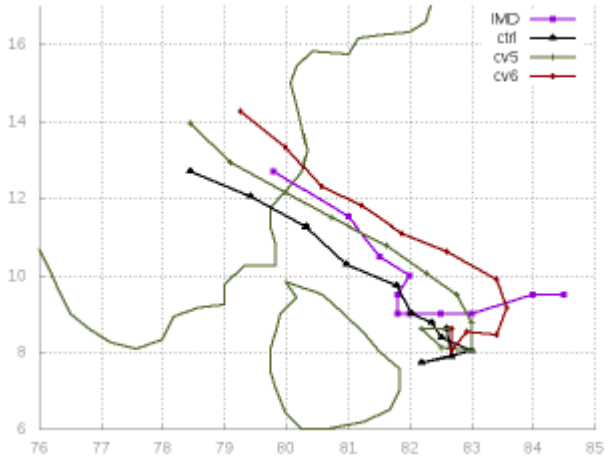

Figure 5.11: Track of cyclone Nilam from CTRL, CV5 and CV6 runs compared with IMD observations

The time series of minimum slp and maximum wind speed simulated by CTRL, CV5 and CV6 runs are compared with the IMD observations in Fig.5.10 (a) and Fig.5.10(b), respectively for cyclone Nilam. The overall trend of intensification as observed is reasonably well simulated by the model in CTRL, CV5 and CV6 runs. However, both the assimilated runs (CV5 and CV6) indicate simulation of a more intense cyclonic system as compared with the observations. The simulation of the CTRL run however is relatively closer to the observations than the assimilated run. The reasons for the simulation of an excessively intense cyclonic system (as compared to observations) in the assimilated run are difficult to establish.The utilization of cv6 option has resulted in a slightly more intense cyclone than the CV5 simulation in terms of the minimum slp and maximum wind speeds.

Fig.5.11 shows the simulated tracks of cyclone Nilam, by the CTRL, CV5 and CV6 runs as compared with the IMD observations. Both the CV5 and CV6 runs have reduced the location error of the cyclone as compared to the CTRL simulation. However, the cyclone Nilam experiences landfall in the CV5 and CV6 simulations at an earlier time than the observed time of landfall.

Fig.5.12 indicates that during the intial forecast hours CV6 run has the least track error for cyclone Nilam. However, after 12 hours of model simulation, CV6 run has simulated the location of the cyclonic system with larger error with respect to IMD observations as compared with both CTRL and CV5 runs.

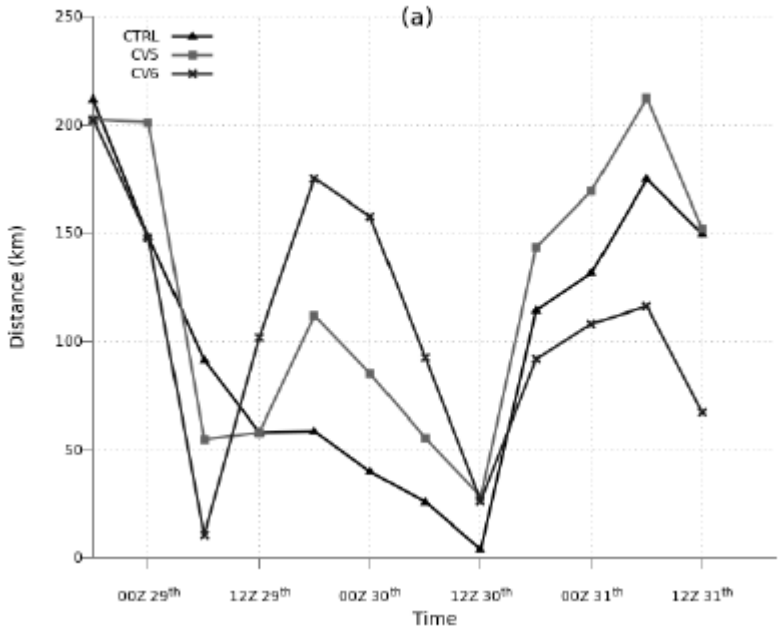

Figure 5.12: Track error of cyclone Nilam from CTRL, CV5 and CV6 runs as compared with IMD observations

5.5 Cyclone Phailin (8-14 October 2013)

5.5.1 Synoptic Conditions

Cyclone Phailin happens to be the strongest cyclone to hit Indian coasts in recent times after the devastating Orissa super cyclone that formed in 1999. The system originated from a remnant cyclonic system over the South China Sea. The cyclonic circulation observed over the Tenasserim coast on 6th October 2013 became a well-marked low pressure over north Andaman sea on the next day. By 03UTC , 8th October 2013, the system had intensified into a depression. As a result of favourable atmospheric conditions like upper level divergence, increased values of low level convergence and low level relative vorticity as well as low to moderate vertical wind shear, along with the favourable SST values and Madden Jullian oscillation (MJO) index, the system further intensified into a deep depression by 00Z 09th October 2013. Moving west-northwestwards, the system crossed Andaman islands and moving over the east central Bay of Bengal, intensified into a cyclonic storm by 12Z 09th October 2013. The cyclonic storm underwent rapid intensification and became a severe cyclonic storm by 03Z of 10th October 2013 and a very severe cyclonic storm by 06Z of the same day.

The associated wind speed of the cyclonic storm Phailin experienced an increase from 55 knots on 03Z 10th October 2013 to 115 knots on 03Z 11th October 2013. Cyclone Phailin further moved northwestwards and crossed Andhra Pradesh and Odisha coast near Gopalpur (19.20^oN and 84.90^oE) around 1700 UTC on 12 October 2013. The system weakened gradually into a severe cyclonic storm by 03Z of 13 October 2013 and a deep depression by 18Z on 13October 2013. The system further weakened to become a well marked low by 09Z of 14 October 2013

5.5.2 Experimental Details

Three numerical experiments are performed to investigate the impact of cv6 option in the simulation of cyclone Phailin. They are (i) a control (CTRL) run where no assimilation of observations is performed. The model is integrated from 12Z of 08th October 2013 to 18Z of 12th October 2013 utilizing NCEP GFS model forecasts for initialization. (ii) CV5 run - the WRF model is initialized using NCEP GFS model forecasts at 12Z , 08th October 2013 and a spin-up of the model is performed from 12Z 08th October 2013 to 18Z 08th October 2013. All conventional surface and upper air observations including satellite wind observations, along with SAPHIR radiances are assimilated using the WRF 3DVar technique with the cv5 option for background error covariance at 18Z 08th October 2013 in the CV5 run. Subsequently, a free forecast of the model is performed from 18Z of 08th October 2013 to 18Z of 12th October 2013 in the above CV5 run. CV6 run is exactly similar to the CV5 run except that cv6 option is utilized instead of cv5 option in background error covariance.

5.5.3 Results and Discussion

Fig.5.13 shows the analysis increments at model level one at 18Z , 08th October 2013 after assimilating all the observations. The top panel shows the increments in the fields of zonal wind ,u (a,e), meridional wind,v (b,f) , temperature, T (c,g), water vapor mixing ratio q (d,h) for the CV5 run and the bottom panel shows the same for the CV6 run. It can be seen that the variation in the $\mathbf{B}$ matrix has an impact on the analysis of all the fields. However, the differences between CV5 run and CV6 run in the analysis fields for T and q are most pronounced. The analysis increments in the

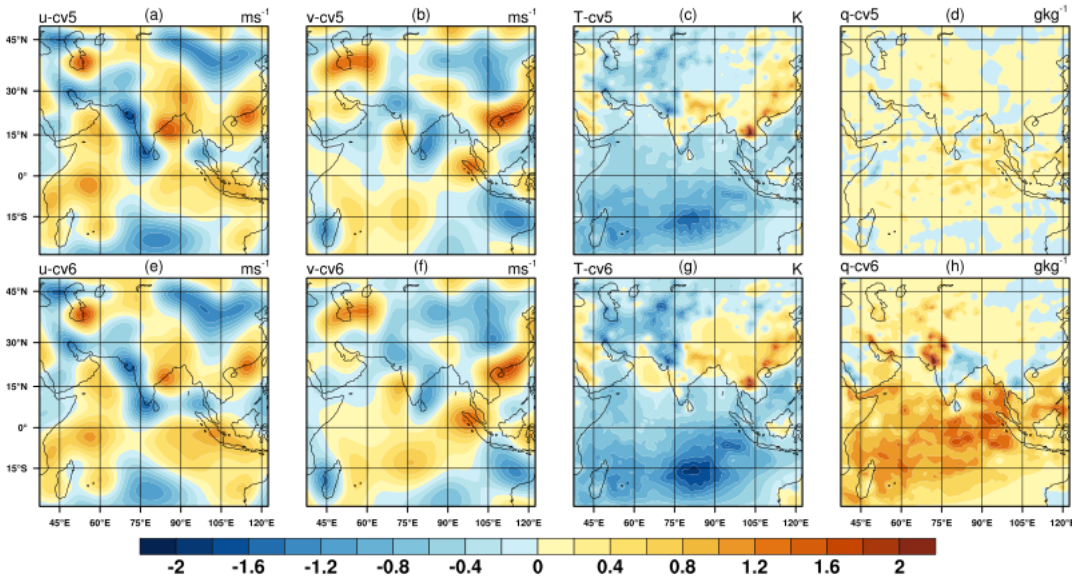

Figure 5.13: Analysis increment at model level 1 for zonal wind(a,e), meridional wind,v (b,f) , temperature, T (c,g), water vapor mixing ratio q (d,h) for CV5 run (a-d) and CV6 run (e-h) for cyclone Phailin

q-field (Fig.5.13(d,h)) indicates that the CV6 option has yielded non-zero moisture analysis increments at the lowest model level , especially over the oceans as compared with the CV5 run.

The vertical profile of moisture divergence at the cyclone center at +0hr, +24hr and +48 hr forecasts of cyclone Phailin for CTRL, CV5 and CV6 runs are shown in Fig.5.14. The moisture convergence simulated by the CV6 run at has a lower value at the lower model levels as compared with the CV5 run and GFS analysis in general in this case. Lower moisture convergence at the cyclone centre could result in reduced vertical motion and affect further intensification of the cyclone in the CV6 run. Since the analysis fields of water vapor mixing ratio show an increment in the CV6 run as compared with the CV5 run, the reduction in moisture convergence could indicate a reduced value of horizontal convergent wind fields. This implies that the modification in the formulation of $\mathbf{B}$ matrix has impacted the divergent component of wind flow.

The horizontal divergence at the cyclone center at +0hr, +24hr and +48 hr forecasts of cyclone Phailin for CTRL, CV5 and CV6 runs are compared in Fig.5.15 with GFS analysis. The initial profile indicates that the CV6 run has simulated lower values of horizontal convergence at the lowest model levels as compared with the CV5 run and GFS analysis. In the subsequent forecast hours also (+24 and +48 hours) shown

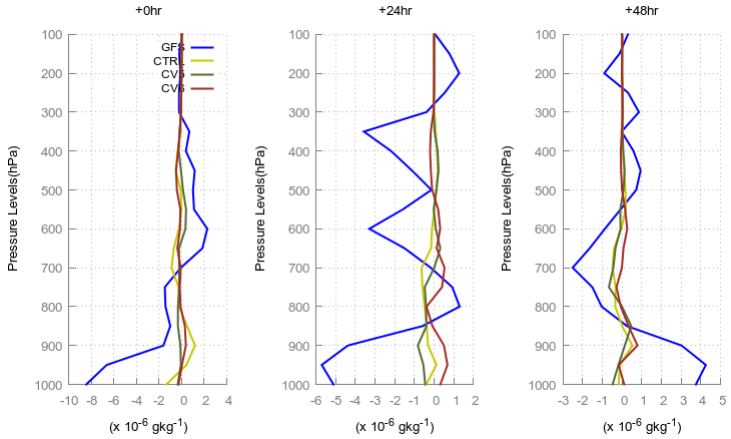

Figure 5.14: Moisture divergence at the cyclone center at +0hr, +24hr and +48 hr forecasts of cyclone Phailin for CTRL, CV5 and CV6 runs compared with GFS analysis

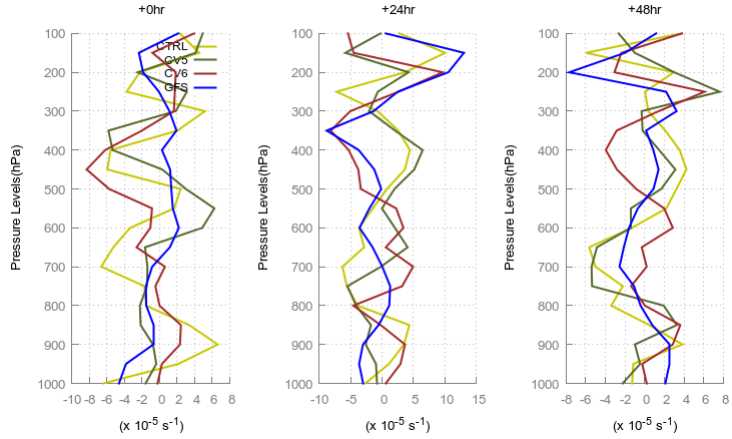

Figure 5.15: Horizontal divergence at the cyclone center at +0hr, +24hr and +48 hr forecasts of cyclone Phailin for CTRL, CV5 and CV6 runs compared with GFS analysis

in Fig.5.15, the horizontal convergence at lower model levels is not as marked in the CV6 run as in the CV5 run and in the GFS analysis.

The Fig.5.16 depicts the vertical profiles of relative vorticity at the center of the cyclone Phailin at +0hr, +24hr and +48 hr forecasts for GFS analysis, CTRL, CV5 and CV6 runs. The relative vorticity profiles simulated by the CV6 run indicates a stronger vortex as compared to CV5 run as well as GFS analysis in the initial forecast hour. The +24 hr and +48 hr forecast profiles of relative vorticity indicates that the vortex simulated in the two assimilated model forecasts (CV5 and CV6) do differ from each other significantly.

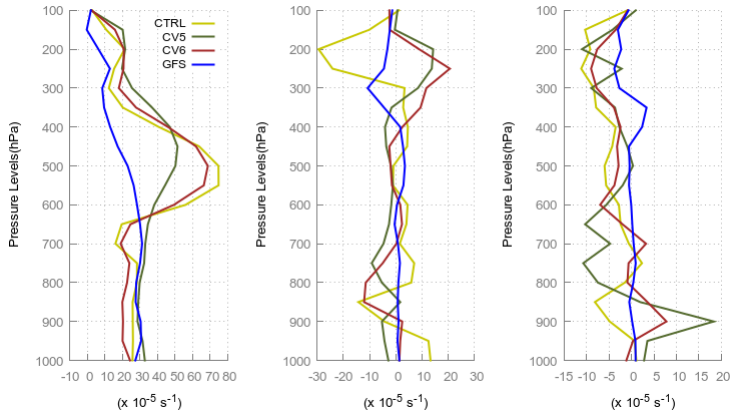

Figure 5.16: Relative vorticity at the cyclone center at +0hr, +24hr and +48 hr forecasts of cyclone Phailin for CTRL, CV5 and CV6 runs compared with GFS analysis

Fig.5.17(a-d) shows the 48 hr accumulated rainfall as observed by TRMM (a) and simulated by CTRL (b), CV5 run (c) and CV6 run (d). Fig.5.17(e-g) shows the difference between TRMM observations and simulated accumulated rainfall by CTRL (Fig.5.17(e)), CV5 run (Fig.5.17(f)), (Fig.5.17(g)) and the difference between CV5 and CV6 run in (Fig.5.17(h)). The figure indicates that the simulation of precipitation associated with Phailin is impacted by the formulation of control variables in the background error covariance matrix. The difference between the rainfall simulated by the CV5 and CV6 runs appear to be associated with the location of position of the cyclonic system. This could be because of the convective processes that occur in the cyclone and the dependence of these processes on the moisture analysis.

Fig.5.18 indicates the improvement parameter (IP) of 48 hour accumulated rainfall simulated by CV5 (a) and CV6 (b) runs for cyclone Phailin. It is seen that there are a significant number of grid points where the assimilation runs , CV5 and CV6 , have positive IP value. However, there are a large number of grid points where the IP is negative in both CV5 and CV6 runs. These corresponds to the ocean where the spatial plot of accumulated rainfall (Fig.5.17) indicates that the model runs are predicting excessive precipitation as compared with the TRMM observations. As compared with the CV5 run, CV6 run has produced a marginally more number of grid points with positive IP values for cyclone Phailin.

The quantitative verification of rainfall simulation is performed by calculating the various skill scores with respect to the TRMM observations. Fig.5.19 (a) shows the

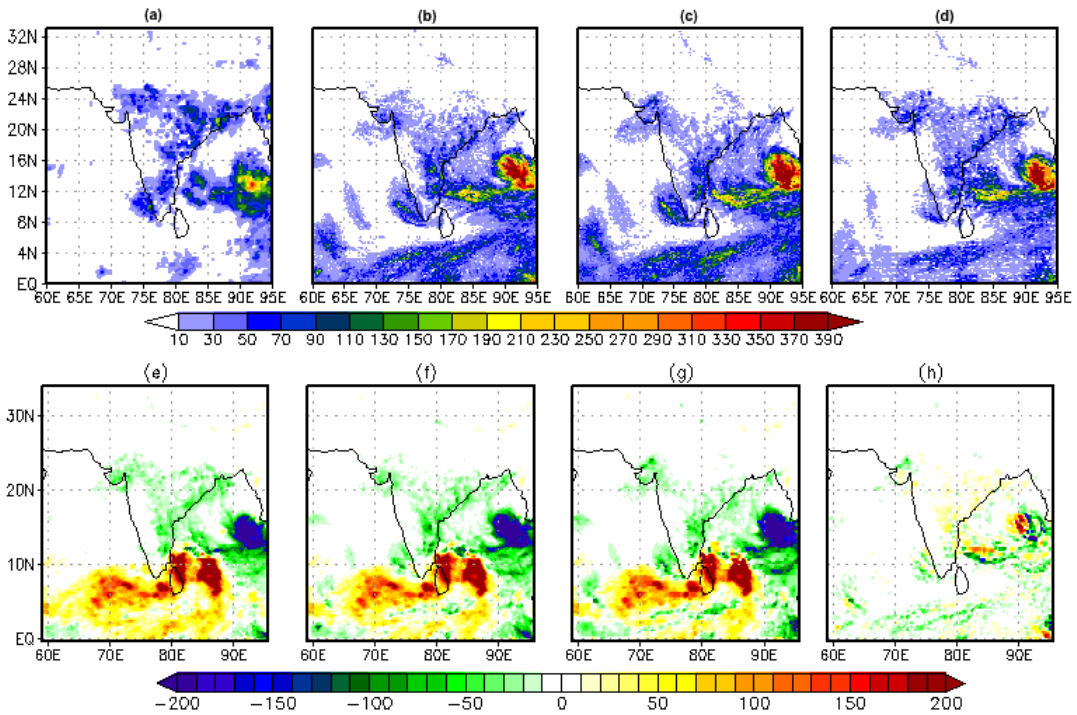

Figure 5.17: 48 hour accumulated precipitation observed from TRMM (a), simulated by CTRL (b), CV5 run (c) and CV6 run (d) for cyclone Phailin. The difference in accumulated rainfall of CTRL run from TRMM observations is shown in (e), difference between CV5 and TRMM in (f) , between CV6 and TRMM in (g) and the difference between CV5 and CV6 is shown in (h)

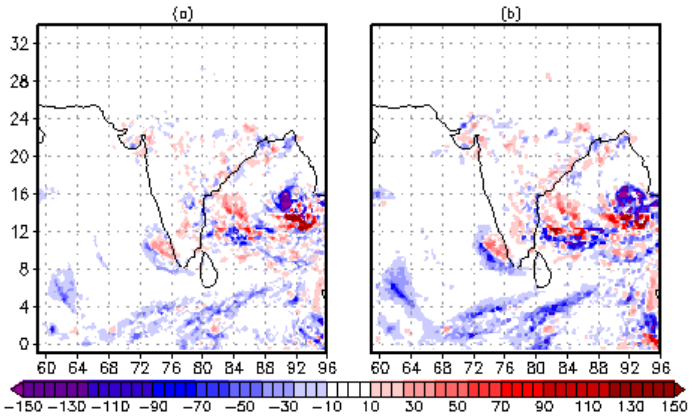

Figure 5.18: Improvement parameter of 48 hour accumulated rainfall simulated by CV5 (a) and CV6 (b) runs for cyclone Phailin.

ETS, Fig.5.19 (b) shows Bias , Fig.5.19 (c) the FAR and Fig.5.19 (d) shows the POD for cyclone Phailin. The skill scores indicate that the assimilation of SAPHIR radiances using the cv6 option has resulted in improvement in ETS score especially at

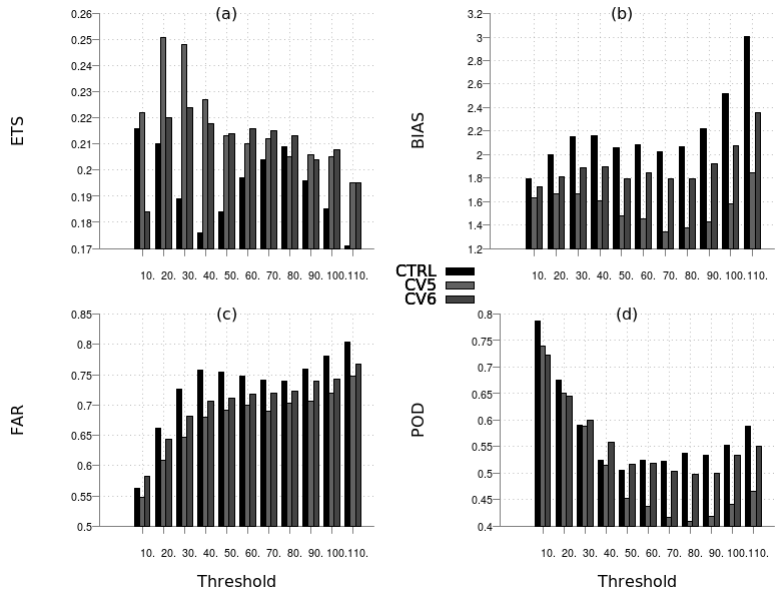

Figure 5.19: ETS (a), Bias (b), FAR (c) and POD (d) of CV5 and CV6 model simulations estimated with respect to TRMM observations for cyclone Phailin.

higher rainfall threshold values as compared with the CV5 run. POD also shows improved values in the CV6 run as compared with the CV5 run. However, the number of false alarms is higher in the CV6 run as compared with the CV5 run.

The time series of minimum slp and maximum wind speed simulated by the CTRL, CV5 and CV6 runs are compared with the IMD observations in Fig.5.20 (a) and Fig.5.20(b) respectively. The model simulations follow somewhat the evolution of the observed minimum slp and observed maximum wind speed. However, the assimilation of SAPHIR radiances has intensified the cyclonic system during day 1 of the simulation of minimum slp and maximum wind speed. It is to be noted that CV6 run and CV5 runs have assumed similar values for minimum slp and maximum wind speed. The formulation of $\mathbf{B}$ matrix has not impacted the simulation of maximum wind speed of cyclone Phailin in a significant manner. However, the observed intensification of the cyclonic system as manifested in Fig.5.20 is not however captured by the model simulations.

The Fig.5.21 shows the simulated tracks of cyclone Phailin, by the CTRL, CV5 and CV6 runs as compared with the IMD observations. All the model runs reproduce the general northwestward movement of the observed Phailin track

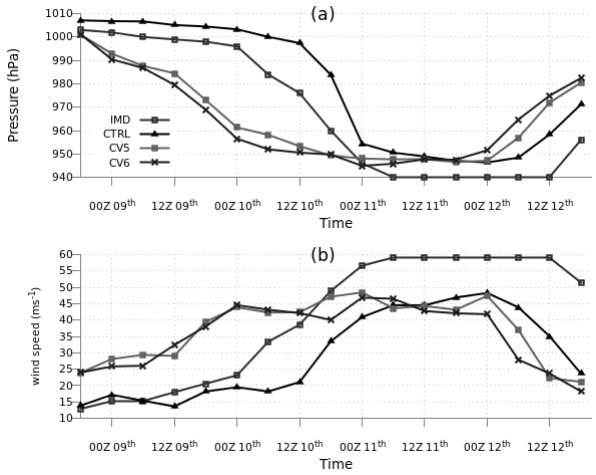

Figure 5.20: Time series of minimum slp (a) and maximum wind speed (b) of cyclone Phailin for CTRL, CV5 and CV6 runs for cyclone Phailin

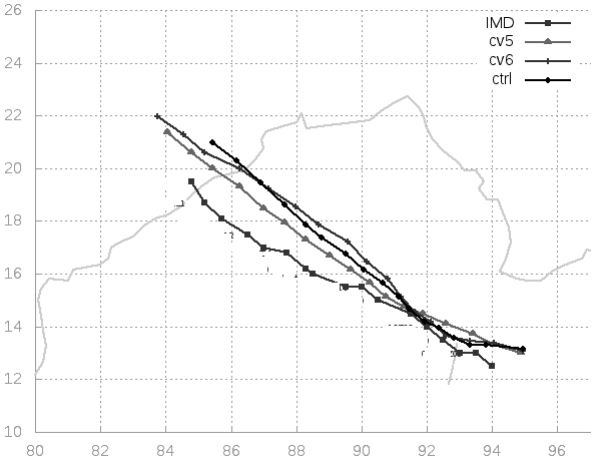

Figure 5.21: Track of cyclone Phailin from CV5 and CV6 runs compared with observations

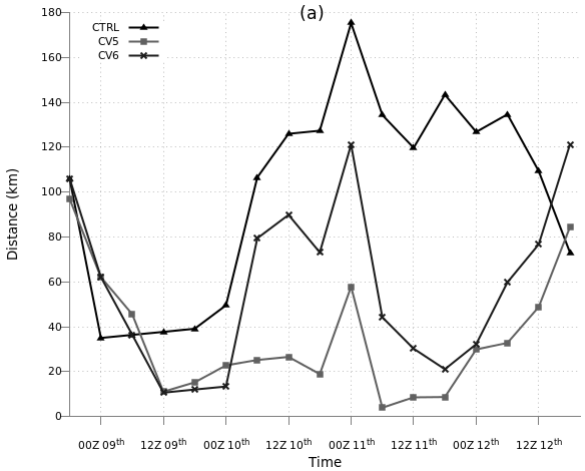

Figure 5.22: Track error of cyclone Phailin from CV5 and CV6 runs as compared with observations

Fig.5.22 depicts the time series track error of the three model simulations with respect to IMD observations. It is seen that at all the forecast hours, CTRL run simulated cyclone center location with the maximum error, and the CV5 run has the least error in cyclone center location.

5.6 Cyclone Thane (25-31 December, 2011)

5.6.1 Synoptic Conditions

The very severe cyclonic storm Thane formed as a tropical disturbance over the southeast Bay of Bengal on 24^{th} December 2011. The above system concentrated into a depression on 25^{th} December 2011 and lay over 8.5°N, 88.5°E. Moving northeast, the depression strengthened into deep depression by 00Z 26 December 2011. Further moving northwards, the system became a tropical cyclone by 18Z of 26^{th} December 2011 and lay centred near 11°N, 87.5°E. The cyclone moved north-westwards over the southeast Bay of Bengal and lay over 12°N, 87°E on 27 December 2011. The cyclone then moved westwards and intensified into a severe cyclonic storm by 28^{th} December 2011. By 12Z of the same day, the system further intensified into a very severe cyclonic storm. Further moving west, the very severe cyclonic storm crossed north Tamil Nadu coast around 00Z 30^{th} December 2011. The system moved westwards and weakened into a severe cyclonic storm by 03Z of the same day. Moving further westwards , the system further weakened into a deep depression by 06Z of 30th December 2011. Further moving westsouthwestwards, the deep depression further weakened into a low pressure area by the early hours of the next day.

5.6.2 Experimental Details

The following three experiments have been performed namely (i) control (CTRL) run - forecast fields from NCEP GFS model are used to initialize WRF model. The CTRL simulation is performed from 12Z 25th December 2011 while the system was in a depression stage to 30th December 2011 , when the cyclone made landfall. No assimilation of observations is performed in the CTRL run. (ii) CV5 run - WRF model is initialized at 12Z 25 December 2011 and a spin up of the model is performed till

18Z 25 December 2011. All available surface and upper air observations along with SAPHIR radiances are assimilated at 18Z 25 December 2011. There after the model is integrated in a free forecast mode without further assimilation of observations, till the cyclone experienced landfall. For the CV5 run, the 3DVar assimilation is performed using the cv5 option of $\mathbf{B}$ matrix. (iii) CV6 run - This numerical experiment is configured in the same way as the CV5 run, except for the utilization of the cv6 option instead of cv5 option in estimating the $\mathbf{B}$ matrix for the 3DVar assimilation.

5.6.3 Results and discussion

Analysis increments in the first model level in zonal wind (u), meridional wind(v), temperature (T) and water vapour mixing ratio (q) in the CV5 and CV6 assimilations at 18Z 25 December 2011 is shown in Fig.5.23. The analysis increments in u an v are more or less similar in the two assimilation runs. However, the model fields T and q are impacted mostly by the utilization of cv6 formulation of $\mathbf{B}$ matrix. Larger positive increments are seen over Arabian Sea and Bay of Bengal in the T and q fields at model level 1 in the assimilation run with cv6 option.

The vertical profile of moisture divergence at the cyclone center at +0hr, +24hr and +48 hr forecasts of cyclone Thane for CTRL, CV5 and CV6 runs are shown in Fig.5.24 along with GFS analysis. The moisture divergence field at the centre of the cyclone at the initial forecast time has similar profile for CV5 and CV6 runs. Here, the model simulations have not faithfully reproduced the more stronger moisture convergence at lower levels as indicated in the GFS analysis. However, the vertical profiles of moisture divergence differ between CV5 and CV6 runs at the +24 hour forecast. CV6 run is seen to produce more moisture convergence at lower model levels in the 24 hour model forecast as compared with the CV5 run. However, the convergence at lower levels simulated by the CV6 run is smaller than indicated by the GFS analysis

The horizontal divergence at the cyclone center at +0hr, +24hr and +48 hr forecasts of cyclone Thane for CTRL, CV5 and CV6 runs are compared in Fig.5.25 along with GFS analysis. At the initial time, at the centre of the cyclone, CV5 and CV6 runs simulate similar vertical profiles for horizontal divergence. All the three model runs produce more horizontal convergence/divergence as compared with the GFS analysis.

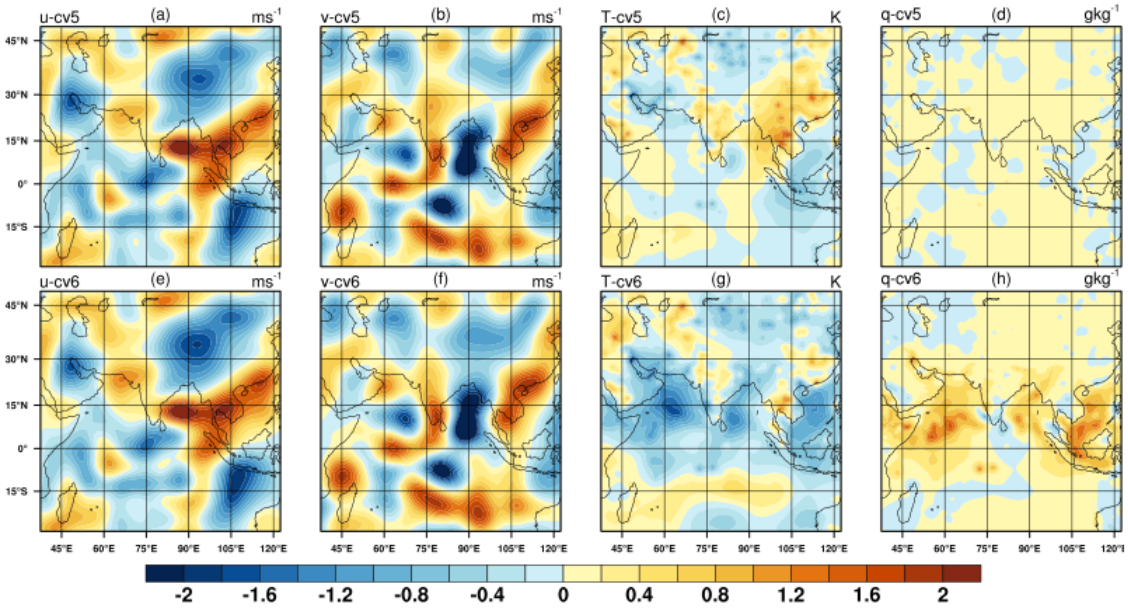

Figure 5.23: Analysis increment at model level 1 for zonal wind(a,e), meridional wind,v (b,f) , temperature, T (c,g), water vapor mixing ratio q (d,h) for CV5 run (a-d) and CV6 run (e-h) for cyclone Thane

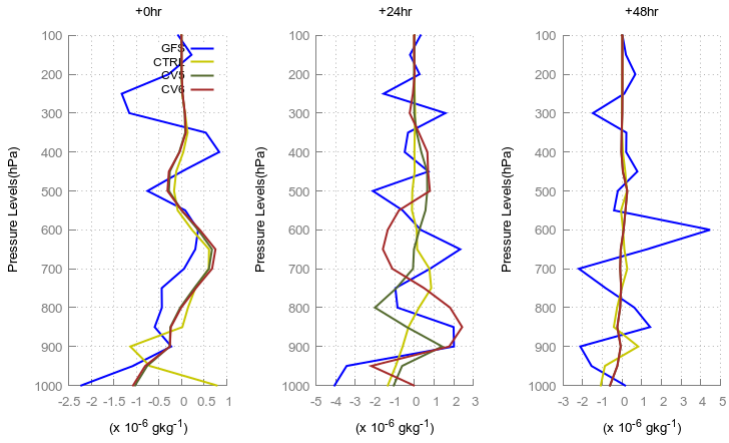

Figure 5.24: Moisture divergence at the cyclone center at +0hr, +24hr and +48 hr forecasts of cyclone Thane for CTRL, CV5, CV6 runs and GFS analysis

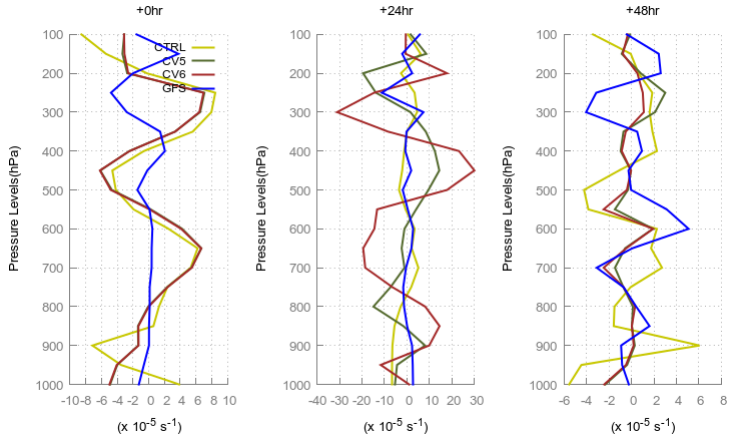

Figure 5.25: Horizontal divergence at the cyclone center at +0hr, +24hr and +48 hr forecasts of cyclone Thane for CTRL, CV5, CV6 runs and GFS analysis

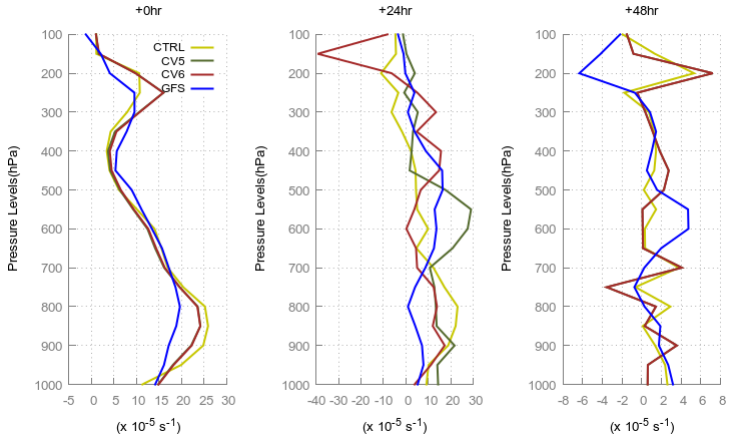

Figure 5.26: Relative vorticity at the cyclone center at +0hr, +24hr and +48 hr forecasts of cyclone Thane for CTRL, CV5, CV6 runs and GFS analysis

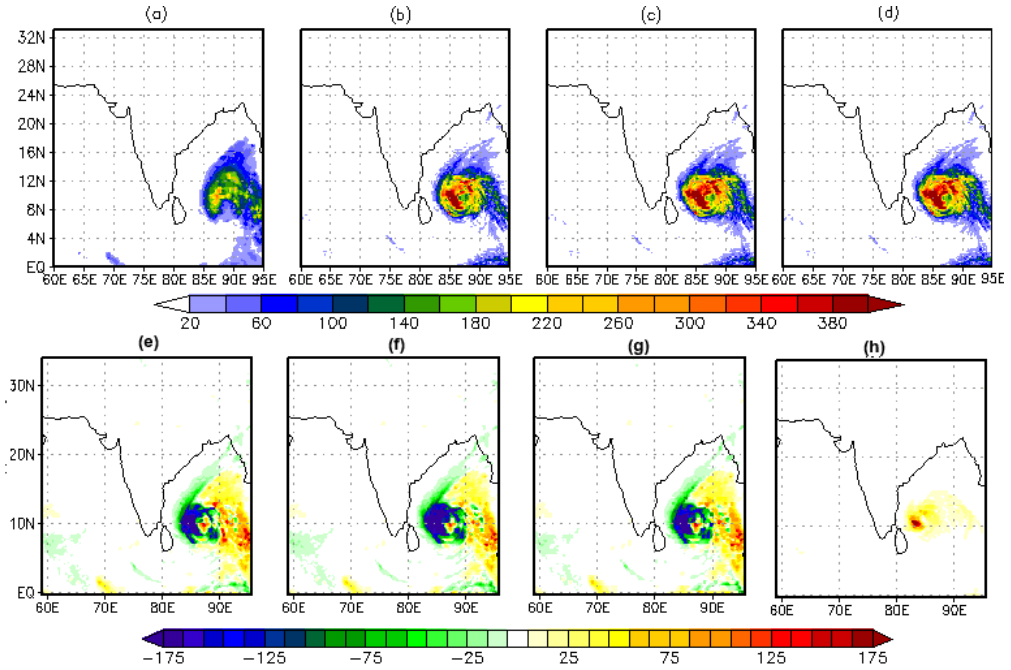

Figure 5.27: 48 hour accumulated precipitation observed from TRMM (a), simulated by CTRL (b), CV5 run (c) and CV6 run (d) for cyclone Thane. The difference in accumulated rainfall of CTRL run from TRMM observations is shown in (e), difference between CV5 and TRMM in (f) , between CV6 and TRMM in (g) and the difference between CV5 and CV6 is shown in (h)

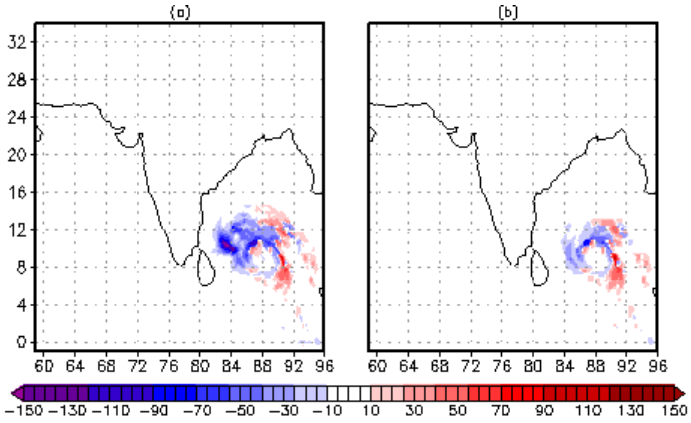

Figure 5.28: Improvement parameter of 48 hour accumulated rainfall simulated by CV5 (a) and CV6 (b) runs for cyclone Thane.

However, the CV6 run simulates more convergence/divergence as compared with the CV5, CTRL run and the GFS analysis at +24 hour model forecast. At +48 hour of the model forecast also, CV5 and CV6 runs simulate similar moisture divergence profiles.

Fig.5.26 depicts the vertical profiles of relative vorticity at the center of the cyclone

Thane at +0hr, +24hr and +48 hr forecasts for GFS analysis, CTRL, CV5 and CV6 runs. The initial profile of relative vorticity at the cyclone centre is also similar for CV5 and CV6 runs and both CV5 and CV6 runs as well as the CTRL run has simulated a larger relative vorticity as compared with the GFS analysis. The +24 forecast results however do show the differences between the two simulations. Both CV5 and CV6 runs have simulated larger vorticity than the GFS analysis in the 24hr forecast also. The CV5 run indicates stronger relative vorticity at the lower model levels as compared with the CV6 and CTRL runs. CV6 runs indicate a relative vorticity profile closer to the GFS analysis as compared with the CV5 run and CTRL run.

Fig.5.27(a-d) shows the 48 hr accumulated rainfall as observed by TRMM (a) and simulated by CTRL (b), CV5 run (c) and CV6 run (d). Fig.5.27(e-g) shows the difference between TRMM observations and simulated accumulated rainfall by CTRL (Fig.5.27(e)), CV5 run (Fig.5.27(f)), (Fig.5.27(g)) and the difference between CV5 and CV6 run in (Fig.5.27(h)). The spatial pattern of simulated 48 hour accumulated rainfall is similar in both CV5 and CV6 runs. Both these runs simulate larger amounts of rainfall as compared with the TRMM observations. The difference plot Fig.5.27(h) indicates that assimilation of observations using the cv5 option has resulted in heavier rainfall as compared with that using the cv6 option for the cyclone Thane during the 48 hours of model simulation.

Fig.5.28 indicates the improvement parameter (IP) of 48 hour accumulated rainfall simulated by CV5 (a) and CV6 (b) runs for cyclone Thane. The north-western region of the cyclonic system is seen to have large number of grid points with negative IP values in both CV5 and CV6 runs. The spatial plot of accumulated rainfall (Fig.5.27) indicates that the model simulations are producing excessive precipitation over this region. However, in the CV6 run, the number of grid points where the IP is negative is lower as compared with the CV5 run, indicating an improvement in rainfall simulation over these grid points when the cv6 option is utilized.

The quantitative verification of rainfall simulation is performed by calculating the various skill scores with respect to the TRMM observations. Fig.5.29 (a) shows the ETS, Fig.5.29 (b) shows Bias , Fig.5.29 (c) the FAR and Fig.5.29 (d) shows the POD for cyclone Thane. ETS is similar for both CV5 and CV6 runs with CV6 run indicating only slightly larger ETS values as compared with CV5 run. Bias for the model is

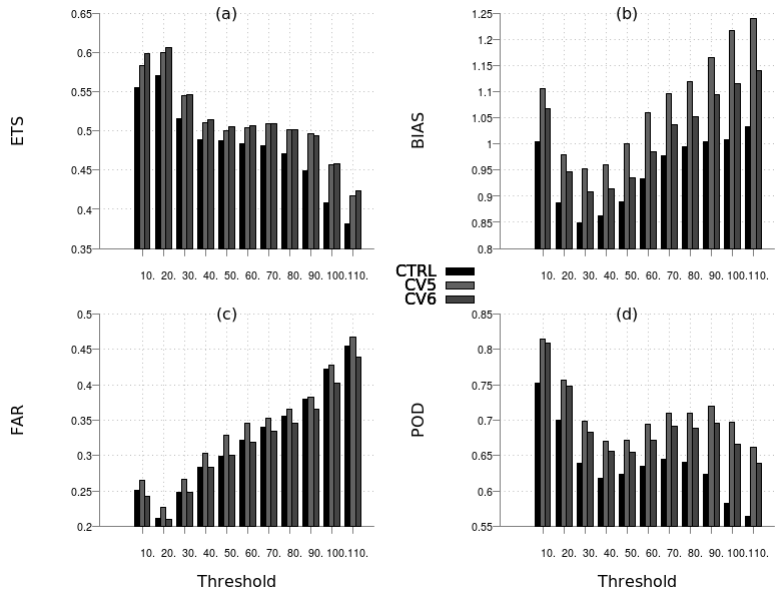

Figure 5.29: ETS (a), Bias (b), FAR (c) and POD (d) of CV5 and CV6 model simulations estimated with respect to TRMM observations for cyclone Thane.

larger at the higher rainfall thresholds with the CV5 run showing the largest values of bias, thereby indicating excessive amounts of rainfall as compared with TRMM observations in the higher rainfall thresholds. FAR is lower in the CV6 run at all thresholds. However, CV6 runs indicate lower probability of detection at all rainfall thresholds when compared with the CV5 run.

The time series of minimum slp and maximum wind speed simulated by the CV5 and CV6 runs are compared with the IMD observations in Fig.5.30 (a) and Fig.5.30(b) respectively. As compared with the IMD observations as well as the CTRL run, both assimilated runs (CV5 and CV6) have simulated a stronger cyclonic system in terms of minimum sea level pressure and maximum wind speed attained at the initial forecast hours. However, after about 60 hours of model forecast, at 12Z 28th December 2011, the observations show an intensification of the system in terms of minimum sea level pressure and maximum wind speed. However, this intensification is not reproduced by all the model forecasts. All the three model simulations indicate a weakening of the system at and after 60 hours of forecast.

Fig.5.31 shows the simulated tracks of cyclone Thane, by the CTRL, CV5 and CV6 runs as compared with the IMD observations. The cyclone track simulations by all the three model runs are fairly close to each other with the CTRL run simulating a

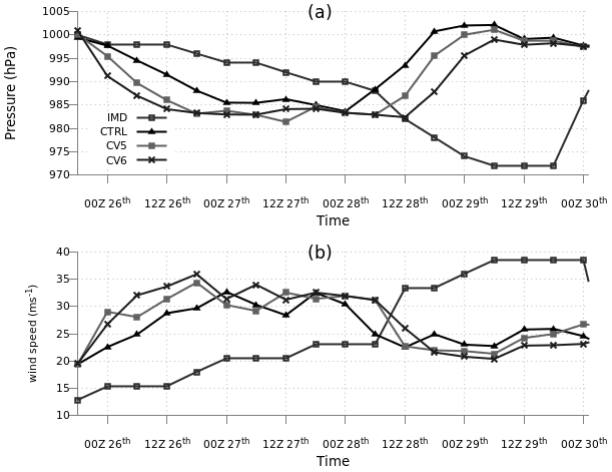

Figure 5.30: Time series of minimum slp (a) and maximum wind speed (b) of cyclone Thane for CTRL, CV5, CV6 runs compared with IMD observations for cyclone Thane

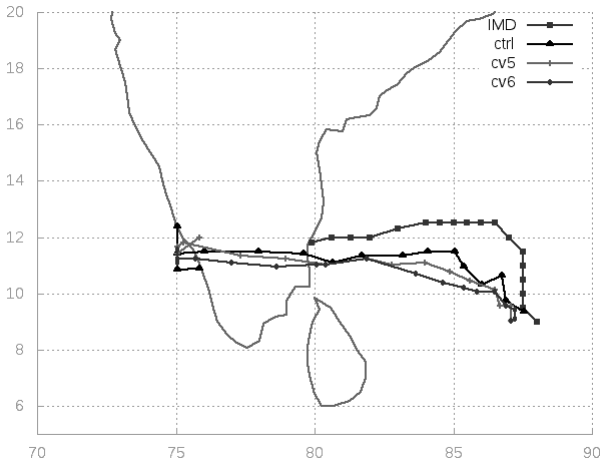

Figure 5.31: Track of cyclone Thane from CTRL, CV5 and CV6 runs compared with IMD observations

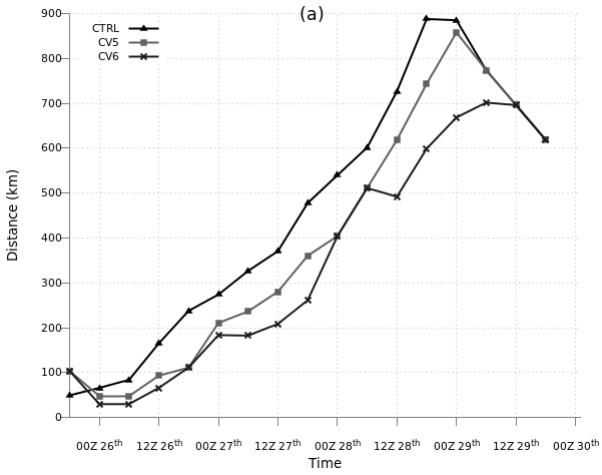

Figure 5.32: Time series of track error of cyclone Thane from CTRL, CV5 and CV6 runs compared with IMD observations

track somewhat more closer to the IMD observations as compared with the CV5 and CV6 runs. In all the model runs, the cyclone experiences landfall earlier than that is observed.

Fig.5.32 shows the time series of track error in CTRL, CV5 and CV6 runs with respect to IMD observations. Fig.5.32 indicates that at the initial time, the locational error in cyclone centre is more in the CV5 and CV6 runs as compared with the CTRL. However, after the initial forecast hour, the CV6 run has simulated the track of cyclone Thane with lowest location errors.

5.7 Summary

The results of this chapter clearly indicate that the impact of assimilation of SAPHIR radiances are indeed influenced by the formulation of $\mathbf{B}$ matrix in the WRF 3DVar system. In general, the three cyclones considered here are seen to be more intense when the cv6 option is used , in terms of minimum sea level pressure and maximum wind speed. Moisture fields are seen to be impacted the most in the model simulations. This has resulted in higher ETS, lower FAR and higher POD in the 48hr accumulated rainfall forecasts in two out of three cyclones considered here. The differences in the model forecasts between the CV5 and CV6 runs are seen to be moderate in this study, even though the analysis of the moisture field differs between the two. The reason for this is the fact that cv5 and cv6 only differ in the representation of correlations

between the control variables in the 3DVar analysis, and both these options generate background error covariances which are static and homogeneous in nature.

The results from the previous chapter (chapter 4) indicate that the formulation of background error covariance matrix does impact the analysis as well as forecasts of mesoscale weather systems over the Indian domain. Monsoon depressions and tropical cyclones have different thermodynamical as well as dynamical characteristics. For example, the central region of monsoon depressions are characterised by cold core at the lower levels and warm core at the upper levels as opposed to tropical cyclones which have a warm core at the middle and upper atmopsheric levels. Also, while tropical cyclone weakens in intensity with height from the surface, the monsoon depression intensifies with height from the surface. Furthermore, monsoon depressions are relatively larger systems in terms of their horizontal extent as compared with the tropical cyclonic systems. However, the formation and further evolution of both these low pressure systems are highly dependent on atmospheric moisture content. Hence better analysis of humidity variable in a numerical model would result in better simulation of monsoon depressions as well as tropical cyclones.

The cv6 formulation of background error covariance in the WRF model provides for an improved representation of moisture in the initial condition for simulation. The correlations among the control variables in the cv6 formulation include relations of humidity variable with other atmospheric variables. In contrast, the cv5 formulation incorporates only univariate humidity analysis. The multivariate humidity analysis in cv6 formulation is seen to have a positive impact on the simulation of both monsoon depressions as well as tropical cyclones.

CHAPTER 6

Impact of assimilating satellite observations using Ensemble Kalman Filter over the Indian Region

6.1 Introduction

The Ensemble Kalman filter (EnKF) has been used for assimilating meteorological observations in many studies world over. However, such studies over the Indian region in a regional model framework have not been performed. The major difference between 3DVar and EnKF is in incorporating the effects of flow-dependent $\mathbf{B}$ matrix in the latter. The current study details the implementation of EnKF in WRF model over the Indian region and investigates the impact of assimilating satellite wind observations using EnKF. Furthermore, a comparative impact study is performed by assimilating the same observations using the 3DVar technique. For the initial two cases investigated, a relatively small domain is utilized, considering the higher computational cost involved in the implementation of EnKF. Subsequently, a larger domain is also utilized for investigating an additional case study.

6.2 Generating ensembles

The numerical forecast of any weather system includes inherent uncertainties. These uncertainties have two causes

(i) Errors in initial conditions : Being a chaotic system, the evolution of the atmospheric state is very sensitive to the prescribed initial conditions of the model. The errors in the initial conditions manifest due to the inherent inadequacy of required number of atmospheric observations as well as instrumental errors associated with atmospheric observations.

(ii) Errors in the forecast model : A numerical weather model is used as a replica of the real atmospheric structure and evolution. All weather models are not es-

sentially perfect. Furthermore, several weather phenomena manifest at scales that cannot be explicitly resolved by the numerical model. Due to the above, the average statistical effects of the such processes need to be parametrized. The parametrization schemes are not perfect and hence can introduce errors in the model. Furthermore, numerical approximations utilized to solve the model equations are an additional source of model errors.

6.2.1 Ensemble generation by random perturbations to initial and boundary conditions (initial-condition ensemble)

In the method of ensemble generation by random perturbations to initial and boundary conditions, all the ensemble members have the same configuration. In this method, WRF model's initial and boundary fields are perturbed using spatially correlated, Gaussian noise. The spatial and multivariate covariances of the above noise follows the background error covariance of WRF 3DVar (Torn et al., 2006). The main advantage of this method is its relative simplicity as compared with other methods such as bred vectors or singular vectors for generating perturbations. Storage of large amounts of data is also not required in this method. However, this method may have some disadvantages since the state-dependent information is not used for generating perturbations (Torn et al., 2006). In this method, a set of random control variables with zero mean and standard deviation of one unit, following a normal distribution, are generated. The analysis increments obtained by the 3DVar minimization for these random control variables are transformed back to the model space using an empirical orthogonal function (EOF) transform (for vertical correlations), a recursive filter (for horizontal correlations) and a variable transform. The zonal and meridional components of horizontal wind, potential temperature and water vapour mixing ratio are perturbed in this method (Zhang et al., 2009). The perturbations thus generated are added to the GFS 6-hour forecast fields to generate an initial ensemble. Here, both initial as well as boundary conditions are perturbed using the same methodology.

6.2.2 Ensemble generation by perturbing model parametrization schemes (multiphysics ensemble)

In the method of ensemble generation using perturbation of model parametrization schemes, each of the ensemble members is configured with a different parametrization combination. However, the same initial and boundary conditions are used to integrate each of these ensemble members. This method addresses the uncertainties in the model itself which arise because of the imperfections in the representation of the various atmospheric processes by the model. Table 6.1 depicts the number of ensemble members which are configured using a particular parameterization scheme.

Table 6.1: Model configuration for the multiphysics ensemble depicting the number of ensemble members (N) for each physics scheme used

N	**Cumulus scheme**	**N**	**Microphysics scheme**	**N**	**PBL scheme**
8	Kain-Fritsch	10	Kessler scheme	20	YSU scheme
8	Betts-Miller-Janjic	10	Lin et al. scheme	20	Mellor-Yamada-Janjic scheme
8	Grell-Freitas (GF) scheme	10	WRF Single-Moment 3-class scheme		
8	Grell 3D scheme	10	New Thompson et al. scheme		
8	Simplified Arakawa-Schubert scheme				

Stensrud et al. (2000) has opined that the initial-condition ensemble and the multiphysics ensemble have similar mean square error values. However, it is well known that the multiphysics ensemble generates larger variance in the temperature, specific humidity, geopotential height, and wind component fields as compared to the initial-condition ensemble method. The variance in the multiphysics ensemble is seen to be

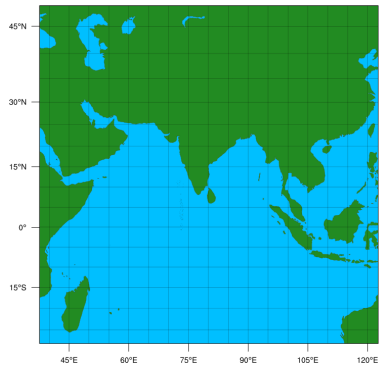

Figure 6.1: Model domain

two to six times larger than that in initial-condition ensemble during the first 12 hours of model simulation (Stensrud et al., 2000). Hence, in a data assimilation scenario where the typical assimilation cycles are performed at a time interval which normally is not more than 12 hours, employing multiphysics ensemble method for generating ensembles can be advantageous.

6.2.3 Ensemble generation using stochastic kinetic energy backscatter (skeb ensemble)

In this method of ensemble generation using the stochastic kinetic energy backscatter (skeb), the error associated with the interaction between the resolved and unresolved components of a numerical weather model is the basis for ensemble generation. The motivation behind this method of ensemble generation is the idea that there is a flow of energy from the unresolved subgrid-scales to the explicitly resolved scales in a model (Shutts, 2005). In this method, using a prescribed power spectrum and a decorrelation time parameter, perturbations are introduced to the rotational component of horizontal wind and potential temperature tendency equations. Berner et al. (2011) provides the theoretical basis of the skeb scheme in the WRF model.

An ensemble ideally represents the flow-dependent evolution of uncertainties during a numerical weather forecast. The uncertainties in an ensemble is quantified by estimating the spread in the ensemble, or the dispersion among the ensemble members. To compare the spread of ensembles generated using the above three methods, a 40-member ensemble is generated using these methods. Towards this end, the input fields for the WRF model obtained from the GFS forecast fields are perturbed at 00Z

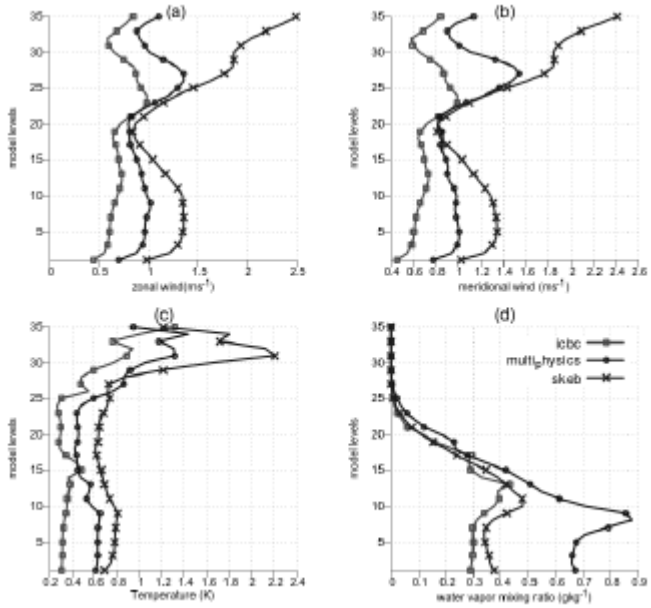

Figure 6.2: (a-d) indicates the vertical profiles of domain averaged spread in 40-member ensembles generated using the three perturbation methods for zonal wind (u), meridional wind(v), temperature (T) and water vapor mixing ratio (q)

of 01 July 2013. This is performed in a single domain of 27km horizontal resolution with 36 vertical levels shown in Fig.6.1 . From these perturbed states, ensemble members are integrated forward in time for a period of 24 hours. The domain-averaged vertical profiles of 12 hr spread of the 40 member ensembles generated using the three methods are shown in Fig.6.2 for zonal wind,u (a), meridional wind,v (b), temperature, T (c) and water vapor mixing ratio, q (d). It is seen from Fig6.2. that all the three methods generate ensembles with somewhat similar vertical structures associated with the 12hr forecast uncertainty. In general, all the three methods provide for larger spread in the upper model levels as compared with the lower levels for u, v and T. However, the extent of spread is quite different between the three methods. The icbc ensemble method has the least spread in all the variables at all model levels. For the wind and temperature fields, the skeb method generates maximum spread among the three methods. However, the maximum uncertainty in the moisture field is found in the multiphysics ensemble method which indicates that the uncertainties associated with the parametrisation schemes provides for larger influence in the uncertainties of the moisture field.

6.3 Ensemble Kalman Filter

The main disadvantage in employing the 3DVar techniques for data assimilation is due to the static background error covariance employed in this method. However, the forecast errors associated with the evolution of the atmospheric system are not reflected in the estimation of $\mathbf{B}$ matrix in 3DVar method. EnKF method inherently provides flow-dependent $\mathbf{B}$ matrix through the use of forecast ensembles. This has the advantage of not requiring the adjoint of the forecast model as in the case of four-dimensional variational technique (4DVar). The 4DVar method provides flow-dependent $\mathbf{B}$ implicitly whereas in EnKF method, the flow-dependency is included directly through the ensemble statistics. In addition, the analysis error covariance is also provided by ensemble based methods. Also it is conceptually simpler to implement an ensemble based assimilation technique. This is because the EnKF method only requires a forecast model and an observation operator to estimate the state of the system unlike a variational technique which requires linearized models and forward operators, their adjoints, and additional prior knowledge about the covariance between different model components (Kalnay et al., 2007).

The EnKF technique is implemented here using the utilities provided by the Data Assimilation Research Testbed (DART) (Anderson et al., 2009). DART is a community, open-source facility developed and maintained at NCAR. It provides software tools for data assimilation.

The background error covariance in an EnKF analysis depend on the forecast ensemble used for assimilation. To illustrate this, a single observation assimilation experiment is performed here. A single temperature observation which differs from the mean background by a magnitude of 1K is assimilated at 10^oN, 80^oE at the surface. The analysis increment at model level 1 in the u,v, T and q fields after assimilating the above pseudo observation is evaluated for three different ensemble member sizes and the results are provided in fig. 6.3. The location of the observation is shown as a black dot in Fig.6.3(g). The results (Fig6.3) indicates the extent to which the background error covariance values are modified through the use of ensembles. Since the analysis increment in a single observation experiment is proportional to a column of the background error covariance, the results as depicted in Fig. 6.3 essentially shows the variation in the correlation between temperature at surface and other model variables

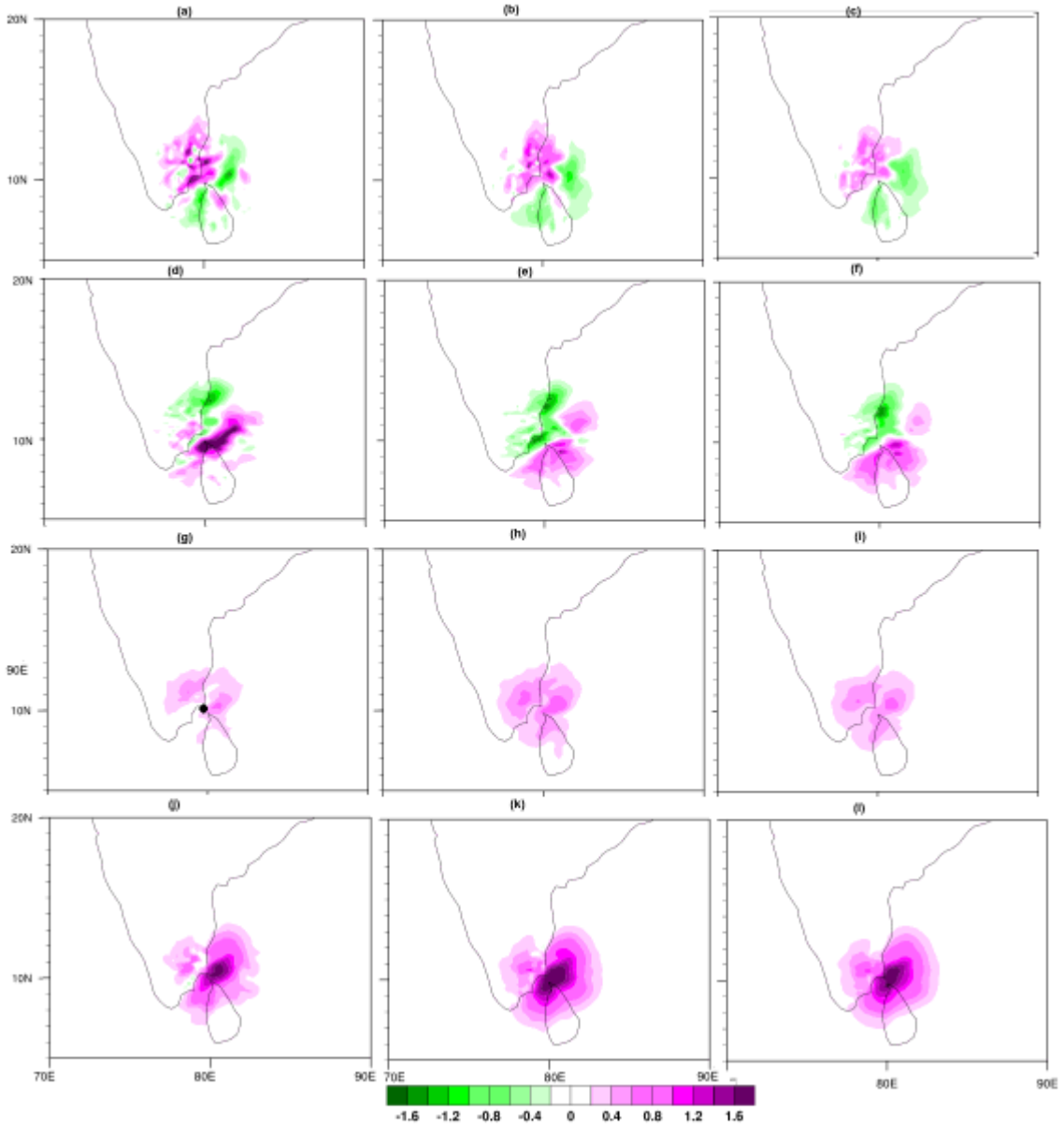

Figure 6.3: Analysis increments in u(a-c), v(d-f), T(g-i) and q(j-l) fields for ensemble sizes 5 (a,d,g,j), 15(b,e,h,k) and 30(c,f,i,l) when a single pseudo observation of surface temperature is assimilated.

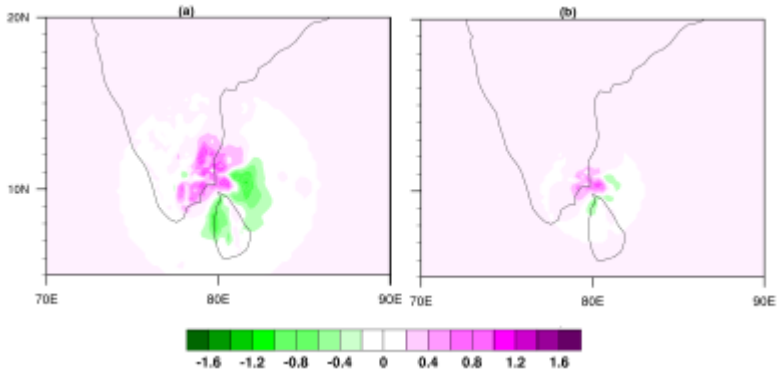

Figure 6.4: Analysis increments in u with covariance cutoff of 600 km (a) and 300 km (b) when a single pseudo observation of surface temperature is assimilated.

with the increase of number of ensemble members. With the increase in the number of ensemble members, the representation of error correlations between the various model fields at different model levels improves. It is clear from the above result that smaller number of ensemble members lead to spurious correlation between the various model fields.

The model sub-space has to be sampled adequately by the ensemble members to properly depict the statistical representation of the atmospheric system for the EnKF assimilation (Oke et al., 2007). Since a general numerical weather model has state space with a typical size $\sim 10^7$, in principle, the ensemble that represents the above system should be very large. However, in practise, a typical ensemble filter, has a maximum of around 100 ensemble members. Even employing 100 ensemble members entails huge computational costs when we consider the cost involved in generating and maintaining the ensemble over several analysis cycles. If the number of ensemble members is small, this results in inadequate representation of the system it samples. Such under-sampling leads to a reduced-rank estimate of the background error covariance matrix (Hamill et al., 2001) .

Undersampling results in what it termed as inbreeding (Houtekamer and Mitchell, 1998) in an ensemble filter. In general, the sampling errors in the ensemble contributes to the underestimation of the forecast errors. Since the analysis errors are expected to be lower than the forecast errors (Furrer and Bengtsson, 2007), the analysis errors are also incorrectly estimated. Inbreeding refers to the situation where the analysis error is systematically underestimated due to assimilation of observations. In the case

where the EnKF utilises the perturbed observation ensemble, more errors associated with the sampling of observational ensemble can also add to this problem. As a result of inbreeding, filter divergence can occur (Hamill et al., 2001). If the forecast error covariances are incorrectly estimated to be larger than what they are, the background will not be properly weighed by the filter and the observations will be given greater weightage. Conversely, if forecast error covariances are very low, then higher weightage will be given to the background than the observations by the filter. Since the analysis error is expected to be lower than the forecast error, the ensemble members incorrectly converge after a few assimilation cycles. Furthermore, spurious long-range correlations between the model variables can manifest due to inbreeding. Due to the above, grid point values can be erroneously influenced by observations that are physically remote.

Covariance inflation and covariance localization (Hamill et al., 2001) have been used to circumvent the issue of undersampling. Covariance inflation is one method of correcting an underestimation in the forecast error covariance matrix by inflating, for each ensemble member, the deviation of the background error from the ensemble mean by a percentage (Anderson and Anderson, 1999). The choice of inflation factor used cannot be universally determined as it depends on the numerical model used, the type of ensemble filter used as well as the system dynamics. Covariance inflation does not address the issue of spurious correlations, which is taken care by covariance localization method (Houtekamer and Mitchell, 1998; Hamill et al., 2001) . In the covariance localization method, the long range correlations are cut-off at a specified distance. This is accomplished by applying a Schur product between the forecast error covariance matrix and a non-zero correlation function in a small (local) region. It is assumed that the Schur product outside the small local region is zero. Usually, the fifth order piecewise rational function as defined in Gaspari and Cohn (1999) is used as the correlation function for implementing localisation in the EnKF method. A correlation lengthscale needs to be specified, ensuring that the correlations drop off to zero at a distance larger than twice this lengthscale.

Fig.6.4 shows the increment in u at model level 1 after assimilating a pseudo temperature observation in an ensemble of size 30, with two correlation lengthscales specified. Fig.6.4(a) shows the analysis increment when a length of 600 km is chosen as localization distance. Here, the influence of the observation reduces to about half at

300 km and becomes zero at 600 km. In Fig.6.4(b) the length at which the impact of observation is zero is chosen to be 300km. The above results shown in the Fig.6.4 provides for the effects of covariance localization on analysis fields in EnKF.

6.4 Impact of assimilating Oceansat-2 satellite ocean surface wind vectors using EnKF in simulating heavy rainfall over Indian peninsula

Several studies have shown that assimilation of ocean surface winds from scatterometers can improve forecasts of a numerical model. Some of the recent studies like (Govindankutty and Chandrasekar, 2010; Sinha and Chandrasekar, 2010; Osuri et al., 2012; Singh et al., 2011) have shown that assimilating satellite data in WRF model using variational techniques have improved the simulations of various weather systems like tropical cyclones and depressions over the Indian region. The objective of the current study is to assimilate ocean surface winds retrieved from the scatterometer aboard Oceansat-2 satellite. Oceansat-2 satellite launched by Indian Space Research Organisation (ISRO) in September 2009 carries three payloads (i) a 13.5 GHz Ku-band scatterometer (OSCAT) (ii) an Ocean Color Monitor and (iii) Radio Occultation Sounder for Atmospheric Studies. The wide swath of 1800 km helps Oceansat-2 to cover 90% of the global oceans in 24 hours. OSCAT is designed to retrieve wind speed and direction over the ocean surface. It has a similar configuration and operating mechanism as that of NASA's QuikSCAT. Mission goals of OSCAT require an accuracy of $2ms^{-1}$ in speed and 20^o in direction for winds within 4-24ms^{-1}. Several recent studies (Mathew et al., 2012; Kumar et al., 2013b; Sathiyamoorthy et al., 2012) have shown that this requirement has been met and OSCAT provides good quality wind measurements for use in numerical weather forecasting. Singh et al. (2008a) have shown that assimilation of OSCAT winds in WRF model has positive impact on the forecasts of surface winds, midtropospheric relative humidity, temperature and precipitation. It has been shown by Kumar et al. (2013a) that assimilation of OSACT winds improves track prediction for a tropical cyclone with degradation in the forecasts of intensity and evolution characteristics. Both these studies assimilated observations using WRF's 3DVar technique. There are no studies available in the literature

that have investigated the impact of assimilating OSCAT winds using EnKF method over the Indian region.

Studies like (Meng and Zhang, 2007, 2008a,b; Whitaker et al., 2008; Zhang et al., 2010; Torn, 2010; Zhang et al., 2011) etc have indicated that improved results are obtained by using ensemble based assimilation method as compared to the more popular three dimensional variational techniques (3DVar). This is because, ensemble based assimilation methods provide flow-dependent multivariate background error covariance and cross-covariance unlike 3DVar. This is possible as the former employ the Monte-Carlo based estimation of error statistics. A large ensemble of model states are integrated forwards in time using the dynamic equations of the model and the various moments of probability density function are calculated from this. Such a calculation of error evolution is not possible in 3DVar. It has been shown by (Buehner et al., 2010a,b; Miyoshi et al., 2010; Zhang et al., 2011) that the performance of EnKF is comparable to that of 4DVar. However, EnKF is simpler than 4DVar as the latter's requirement of adjoint of observation operators do not arise in the former. Further, EnKF's analysis code does not depend on the prediction model as in the case of 4DVar. Importantly, information regarding uncertainty in both analysis and observations can be obtained in an ensemble based technique unlike in a variational technique.

6.4.1 Case 1 Heavy rainfall event over the Indian peninsular region during November 2009

Case Investigated

A low pressure area was formed around the south west Bay of Bengal and neighbourhood on 14th November 2009. The trough from this low pressure region extended over west central Bay across Sri Lanka and southwest Bay of Bengal. The system persisted on 15th November and moved towards north west and became less marked on 16th November 2009. This system strengthened the northeast monsoon flow and resulted in heavy rainfall in the peninsular region, especially over the southern state of Tamil Nadu. A trough of low pressure over Lakshadweep islands in the Arabian sea was also seen on 14th November 2009. The above system persisted on 15th November and moved eastward and is seen over east Arabian sea. Fig.6.5(a-d) shows the mean

sea level pressure (MSLP) and associated lower tropospheric wind flow at 850hPa during the evolution of this system over the model domain, as obtained from the NCEP analysis fields at 06Z and 18Z of 14th November, 12Z on 15th November and 06Z on 16th November, respectively.

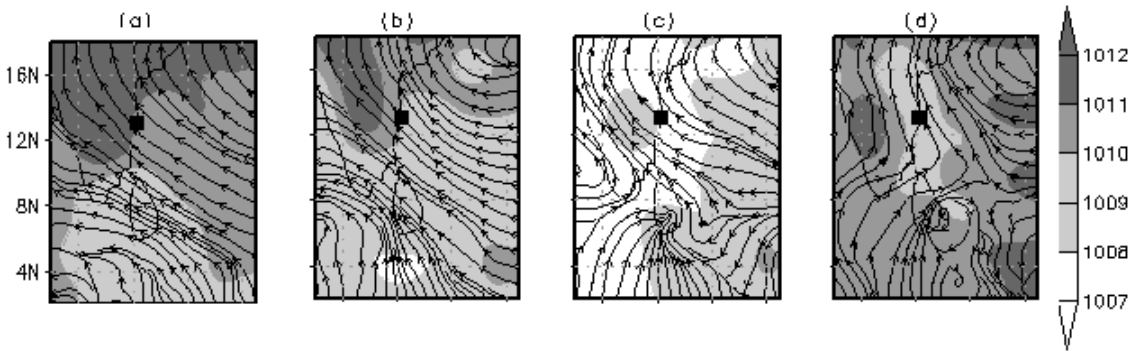

Figure 6.5: MSLP and 850hPa streamline pattern from NCEP analysis at 06Z (a) and 18Z of 14th November (b) , 12Z of 15th November (c) and 06Z of 16th November (d) for the first case study

The MSLP field from NCEP analysis shows clearly the existence of low pressure over South West Bay of Bengal, south of Sri Lanka on 14November 2009 (Fig.6.5a,b). The analysis shows that the above low pressure area has intensified slightly on 15th November 2009 moving northwards and is now lying over Sri Lanka (Fig.6.5c). The analysis further reveals that the low pressure has weakened considerably on 16th November 2009 (Fig.6.5d). The associated streamline pattern at 850hPa from the analysis reveals a cyclonic circulation which extends upto the southern peninsula on 14th November 2009. The centre of the above cyclonic circulation has moved northwest wards on 15th November 2009 and has clearly weakened on 16th November 2009 as shown in the analysis.

Model Details

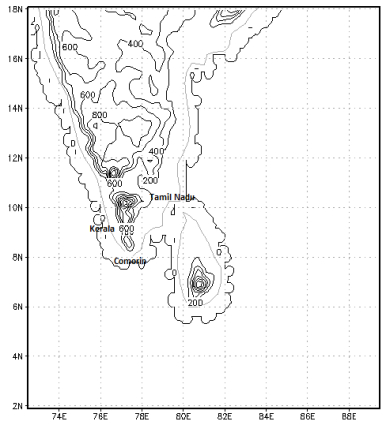

Figure 6.6: Model domain used.

For this case study, the WRF model is configured with a small domain so that computational expenses are somewhat modest. WRF model has a single domain of 15 km horizontal resolution with 120 x120 grid cells in the north-south and east-west directions and has 36 levels in the vertical (shown in Fig.6.6). The various parametrization schemes used in this study are Kain-Fritsch scheme for cumulus, WRF Double-Moment 5-class (WSM5) scheme for microphysics, Rapid Radiative Transfer Model (RRTM) for longwave radiation, Dudhia scheme for shortwave radiation, Yonsei University (YSU) scheme for boundary layer and thermal diffusion scheme for surface physics.

Experimental Details

Three experiments are performed to analyse the impact of assimilation of OSCAT winds on the simulation of this heavy rainfall event, namely (i) control (CTRL) run, (ii) assimilation run using 3DVar (3DVar run) and (iii) assimilation run using EnKF (EnKF run). The initial and boundary conditions for WRF model are obtained from National Centre for Environmental Prediction (NCEP) Global Forecast System (GFS) fields in all the three experiments. No observations are assimilated in the CTRL run. The model is integrated from 18Z 12th November 2009 to 06Z 16th November 2009 in the CTRL run. The 3DVar and EnKF runs, are started from 18Z 12th November 2009. OSCAT wind observations are assimilated thrice after this initial spin up - at 06Z and 18Z 13th November 2009 and 06Z 14th November 2009. From the assimilated state

at 06Z 14th November 2009, the model is integrated for 48 hours without further assimilation of OSCAT winds, in a free forecast mode. The EnKF run is identical to the 3DVar run, except that EnKF method is employed. The model simulated fields during the free forecast run (during 06Z 14th November 2009 to 06Z 16th November 2009) for all the three experiments are compared with each other and validated with observations and analyses to investigate the impact of assimilation. The EnKF experiment is configured with 64 ensemble members and uses a prior inflation of 1.2. The ensemble is generated by perturbing the model's initial and boundary conditions at 18Z 12th November 2009. The ensemble members are integrated till 06Z 13th November 2009 so that the error evolution of the model forecast is captured.

Scatterometer data quality is known to be affected by rain. Due to this, prior to data assimilation, rain-flagged OSCAT wind observations are not included in both assimilation runs.

Results

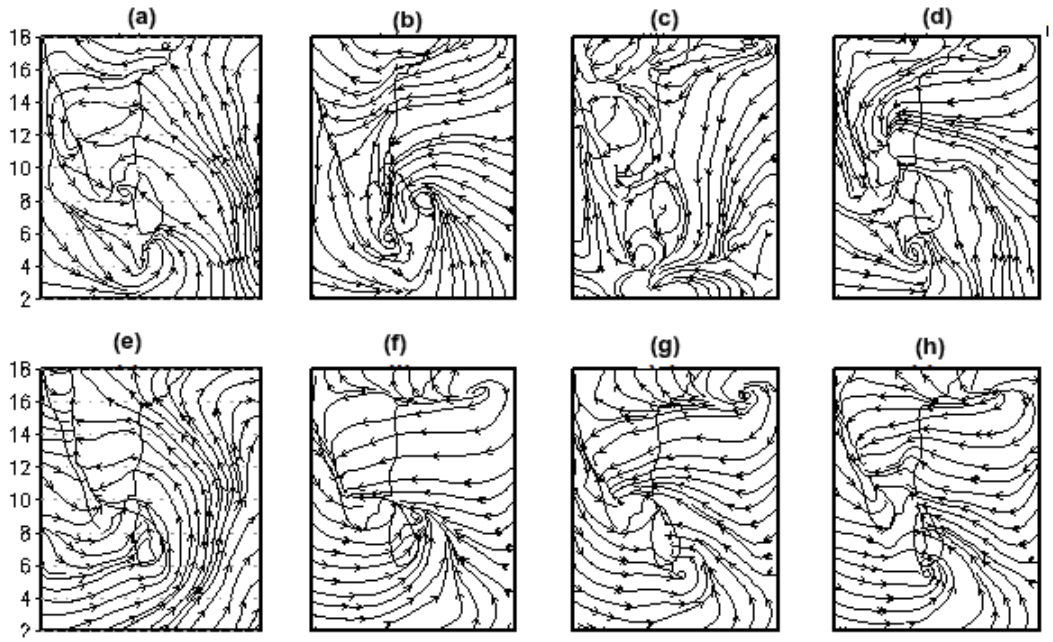

Figure 6.7: 850hPa moisture streamline (qu,qv) patterns at 12 hours of forecast from the NCEP analysis, CTRL, 3DVar and EnKF runs (a-d) and at 30 hours of forecast (e-h) for the first case study.

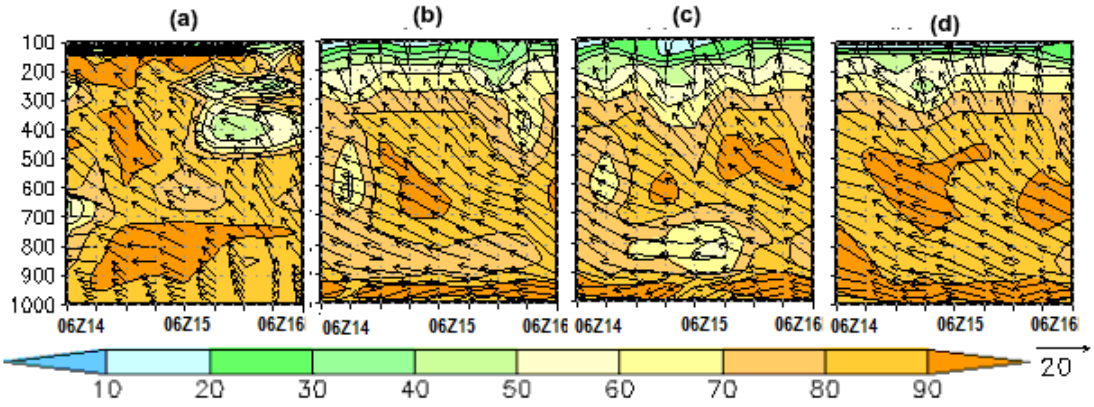

Figure 6.8: Height-time section of the wind speeds and relative humidity over Chennai from analysis (a) and the model runs (b-d) for the first case study.

The 850hPa lower tropospheric moisture flux at 12 hours of forecast , i.e. at 18Z 14th November (Fig 6.7b-d) and at 30hours of forecast , i.e. at 12Z 15th November 2009 (Fig.6.7f-h) as simulated by the model (CTRL, 3DVar and EnKF) are compared with that from NCEP analysis fields in Fig 6.7a and 6.7e. While the CTRL results are shown in Fig 6.7b, 6.7f, the 3DVar and EnKF results are shown in Fig 6.7c,6.7g and Fig 6.7d,6.7h respectively. While the analysis reveals that the lower tropospheric moisture flux convergence is south of Sri Lanka at 12 hours of the forecast (Fig 6.7a), the moisture flux convergence moves over northwards and is over Sri Lanka at 30 hours of forecast (Fig 6.7e). The CTRL and EnKF runs at 12hours of forecast have simulated lower tropospheric moisture convergence as manifested through associated cyclonic circulation (Fig 6.7b, 6.7d). The 3DVar run at 12 hour forecast simulates an anti-cyclonic circulation (Fig 6.7c) which is not observed in the analysis. However, all the three model runs have simulated lower tropospheric moisture convergence at 30hours of forecast (Fig 6.7f-h) over Sri Lanka as well as over the west coast of India. The analysis, however, at 30 hours of forecast shows low level moist convergence over the west coast of India (Fig 6.7e).

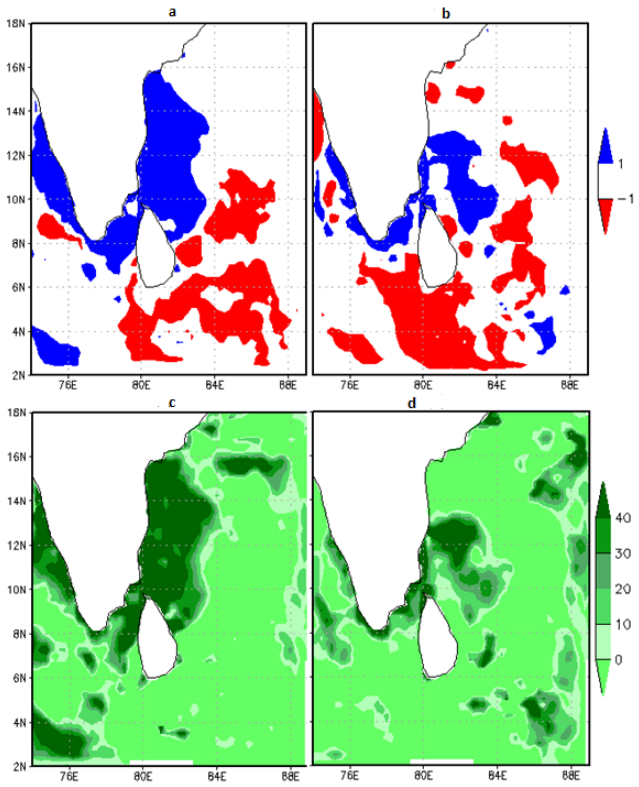

Figure 6.9: Spatial distribution of improvement parameter and forecast impact parameter in the simulation of wind magnitude with respect to QuikSCAT observations for 3DVar run (a),(c), EnKF run(b),(d) for the first case study

Fig 6.8(a-d) shows the time-height variation of relative humidity and wind vectors over a location (Chennai) on the east coast of peninsular India during the free forecast period. Chennai is chosen since it receives most of its rain during the northeast monsoon ; it received 120 mm in 48 hours during this case study event. The NCEP analysis reveals persistent south-easterlies over Chennai (Fig 6.8a), with the southerly component increasing with height as well as time. Also, the wind magnitude increases with height in the analysis (Fig.6.8a). All the three model runs faithfully reproduce the above easterly wind pattern (Fig6.8b-6.8d) seen during the northeast monsoon season. The NCEP analysis reveals the existence of very high relative humidity (over 80%) over Chennai during the entire two-day period (14-16 November 2009), except for a small patch of dry air around 700 hPa on day1 and during the last 12 hours over 300 hPa (Fig6.8a). All the three model runs manifest relatively drier air at and above 300 hPa (Fig 6.8b-6.8d). However, the 3DVar run simulates excessive dry air on day 1 of the forecast over the entire troposphere, which is also manifested in the lower tro-

posphere after 24 hours of forecast. (Fig 6.8c). A possible reason for the above feature in the 3DVar run is due to the lower tropospheric anticyclonic circulation simulated on day 1 of the forecast (Fig.6.7c) , a feature not observed in the analysis.

The model forecasts of the surface wind surface magnitude are compared with QuikSCAT gridded wind observations for validation since it is more meaningful to obtain an independent validation of the results of the model simulations with observations which have not been assimilated. Also, QuikSCAT winds are available at a finer horizontal resolution of 25 km. The quantitative evaluation of the simulated surface wind magnitude with respect to QuikSCAT observations are performed using Improvement parameter (η) and Forecast impact parameter (F.I).

Positive η values indicate positive impact of assimilation. The spatial distribution of improvement parameter for the 3DVar and EnKF runs are given in Fig.6.9a and 6.9b. It is seen that η is positive over large regions (51.8%) in the 3DVar run (Fig6.9a) indicating positive impact on the simulated wind magnitudes. However, the EnKF run shows a slightly lower percentage (40.8%) of grid points where η is positive. It is somewhat surprising to note that the simulation of surface winds over the ocean using EnKF method has resulted in lower number of positive values of η and F.I as compared to the 3DVar run.

A positive FI indicates that the model forecast with assimilation compares better with observations than the forecast without assimilation. Fig.6.9c and 6.9d shows the spatial plot of FI for 3DVar and EnKF runs. The result of the 3DVar run has large percentage of grid points where FI is positive (51.7%) while the same for the EnKF run is slightly smaller (42.7%).

The evaluations of the simulated rainfall for the CTRL, 3DVar and EnKF runs are performed by comparing with Tropical Rainfall Measurement Mission (TRMM) (Fig6.10a) and India Meteorological Department (IMD) gridded rainfall (Fig6.10b) observations and are shown in Fig.6.10c, 6.10e, 6.10h for 48hr accumulated rainfall forecasts. The TRMM observations indicate that the south eastern Arabian sea (off the west coast of Kerala) as well the region off the east coast of Tamil Nadu experience very good rainfall. The IMD gridded rainfall data is available over a coarser resolution (0.5^ox0.5^o) as compared with TRMM (0.25^ox0.25^o) and is available only over the Indian land region. Despite the above differences, the intensity and the spa-

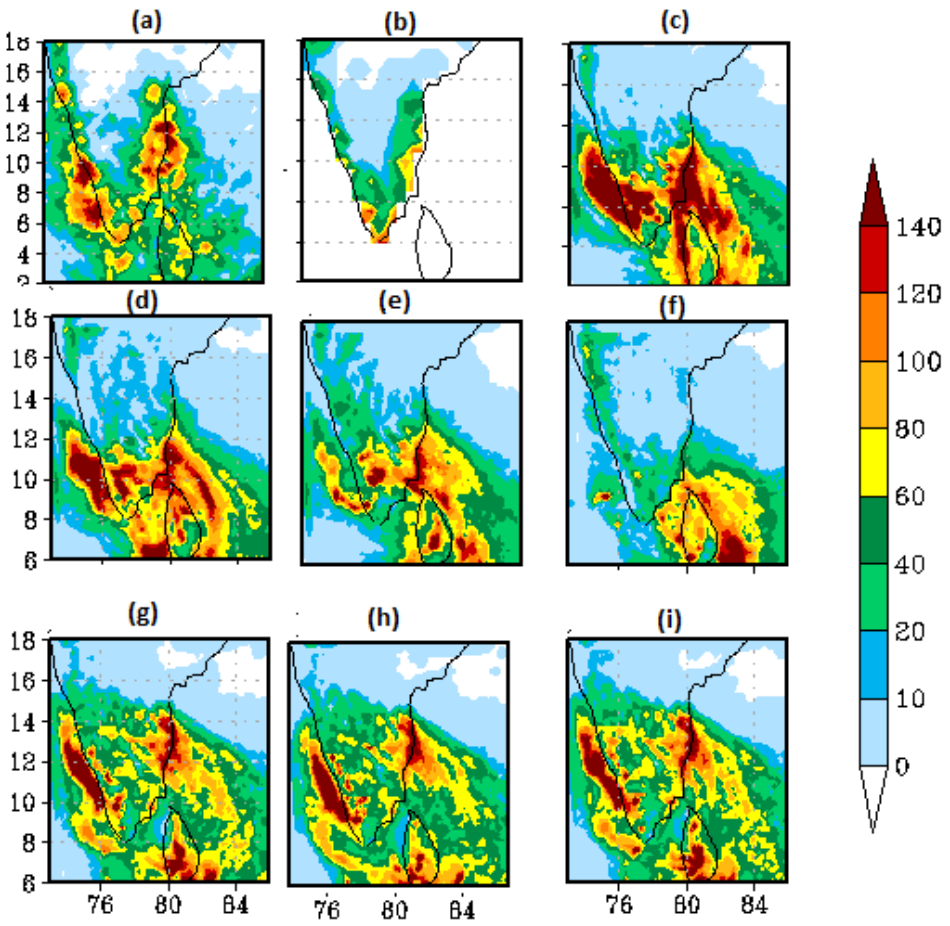

Figure 6.10: 48hr accumulated rainfall from TRMM(a), IMD(b), CTRL run(c), 3DVar run (d-f) with lenghtscale 0.5,1.0 and 1.5 and EnKF run (g-i) with cutoff radius 50km,100km and 150km for the first case study.

tial distribution of the observed rainfall matches well between the TRMM and IMD observations. The CTRL run (Fig 6.10c) simulates excessive rainfall off the coasts of Kerala and Tamil Nadu. Further, the CTRL run simulates heavy rain southeast of Sri Lanka which is not observed in TRMM.

The results of the 48hr rainfall simulation for the 3DVar and EnKF runs are shown in Fig 6.10e and 6.10h respectively. It is clear that the assimilation of OSCAT winds have succeeded in reducing the excessive rainfall simulated by the CTRL run. However, the rainfall simulated by the 3DVar run is severely underestimated over the west coast. Also, the location of the precipitation maxima over the east coast of India as seen in TRMM is displaced southward in the 3DVar run. A possible reason for the above behaviour can be due to the anomalous anticyclonic circulation together with

much drier air as simulated by the 3DVar run on day 1 of the forecast. The EnKF run, in addition to reducing the excessive rainfall simulated by the CTRL run, also correctly reproduces the location of the rainfall maximum over the east coast of India. Furthermore, the EnKF run also appears to reproduce the spatial distribution of rainfall maximum over the west coast of India. It is true that the EnKF run has simulated more rainfall in terms of intensity over North Indian ocean as compared to TRMM and other model runs.

Although the poor simulation of the 3DVar run has been attributed to the anomalous anticyclonic circulation, it is worthwhile to investigate whether the above underestimation of rainfall prediction by the 3DVar run can be corrected through tuning of parameters which eventually determine the "background error covariance ($\mathbf{B}$ matrix)". One of such parameters is the scaling factor for length scale of the recursive filter utilized in the $\mathbf{B}$ matrix calculations. Towards this end, the 3DVar experiment has been rerun by changing the scale factor of the length scale from 1 (Fig6.10e) to values such as 0.5(Fig6.10d) and 1.5 (Fig 6.10f). Similar studies such as Guo et al. (2006),Ha and Lee (2012) have been undertaken in the literature to study the sensitivity of scaling factor of length scale. Increasing the scaling factor of length scale to 1.5 yields a simulation where heavy rain is only observed over Indian Ocean with very little rain over the southern peninsula (Fig 6.10f). Decreasing the scaling factor of length scale to 0.5 yields a simulation closer to the CTRL simulation.

A similar sensitivity study to investigate the effect of "localization" on the model simulation using EnKF technique has been undertaken. While the cutoff radius Gaspari and Cohn (1999) (a measure of localization) is chosen as 100km in Fig6.10h, the same is changed to 50km (Fig 6.10g) and 150km (Fig 6.10i) respectively. The results of the above sensitivity studies have not produced any discernible changes in the simulated precipitation of the EnKF run (refer Fig6.10g-6.10i). To sum up, it is clear that the best possible rainfall results for 3DVar and EnKF runs are associated with Fig.6.10e and 6.10h, respectively and hence these are utilised in the subsequent analyses and discussions.

In addition to the comparison of qualitative spatial distribution of rainfall, Fig6.11 a-d provide quantitative skill score measures of 48hour simulated precipitation such as are Equitable threat score (ETS), Bias, Probability of detection (POD) and False alarm

ratio (FAR) for CTRL, 3DVar and EnKF runs with respect to TRMM observations.

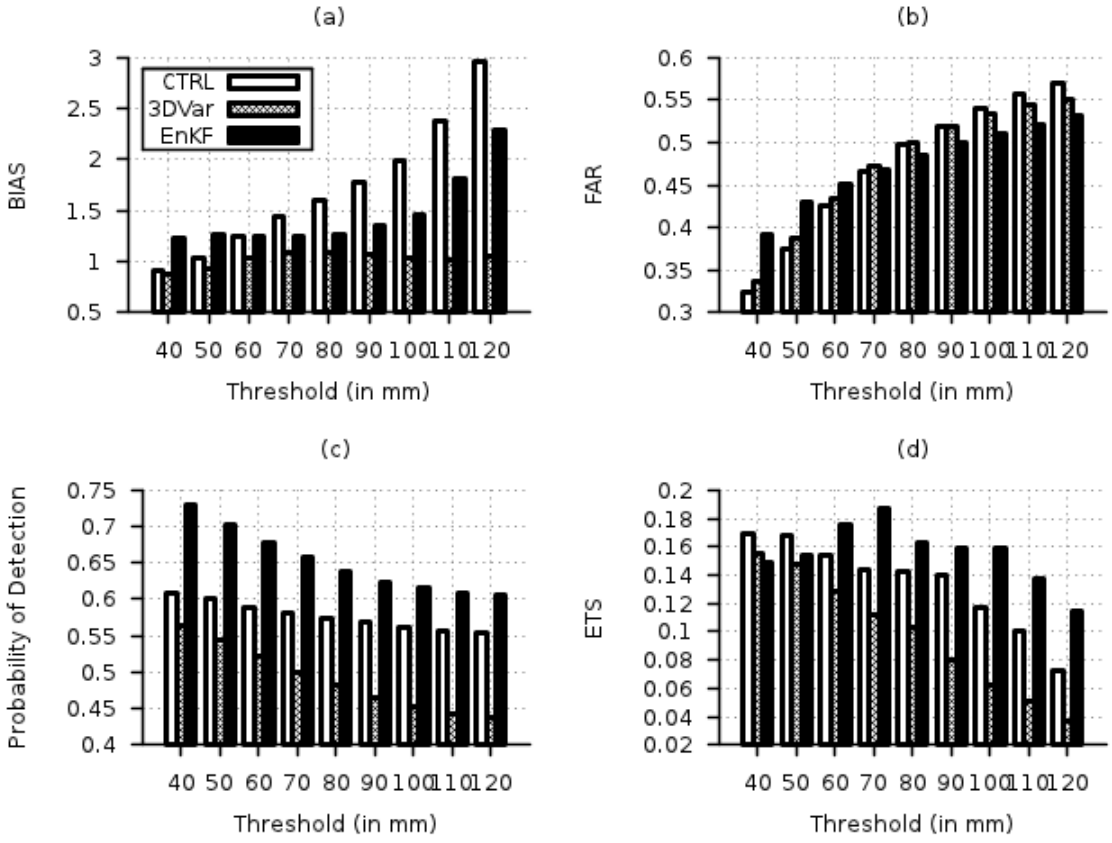

Figure 6.11: (a) Bias score, (b) False alarm ratio, (c) Probability of Detection and (d) Equitable threat score of the model runs for 48hour accumulated precipitation with respect to TRMM for different rainfall thresholds for the first case study

Bias score is a measure of the over prediction/under prediction of the model rainfall forecast with respect to TRMM observations. Fig 6.11a depicts the bias for CTRL, 3DVar and EnKF runs for various rainfall thresholds. While the CTRL run overestimates very significantly the rainfall for higher thresholds of rain, the 3DVar run underestimates rainfall for lower threshold values and is relatively unbiased for higher rainfall values. The EnKF run, however, over predicts rainfall as compared to the 3DVar run for all thresholds. FAR is a measure of the number of false alarms (rainfall simulated when there are no observations to indicate the presence of rainfall) for a model forecast. While the FAR for moderate to heavy rainfall thresholds indicate that EnKF run has lower false alarm ratio as compared to CTRL and 3DVar runs, the same is not true for low rainfall thresholds(Fig6.11b) .

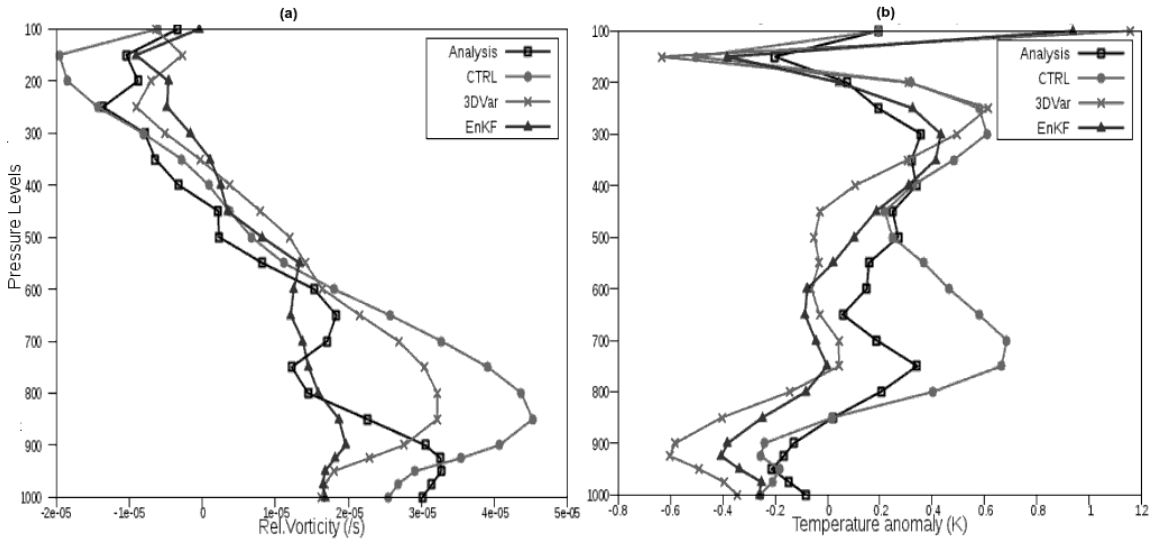

Figure 6.12: Area averaged and time averaged temperature anomaly (a) and relative vorticity (b) over 3^ox^o box around the centre of low pressure system from analysis and the three model runs for the first case study

POD is a measure of the skill of a model to predict rainfall with respect to TRMM observations including the cases of correct predictions which happen due to mere chance. The POD for 3DVar run (Fig.6.11c) is consistently lower as compared with the EnKF and CTRL runs indicating poor rainfall prediction by the 3DVar run for this case. The POD values for EnKF run are the highest for all rainfall thresholds among the three model run results. The ETS is a measure of the skill of a model to predict rainfall without including the cases of correct predictions which occur by mere chance. The poor performance of prediction of rainfall by the 3DVar in this case as seen in Fig 6.10e is reflected in the low ETS values for the 3Dvar run as well as compared to the CTRL and EnKF runs (Fig6.11d). As with POD, the ETS values for the EnKF are the highest for most of the rainfall thresholds.

The location of the minimum SLP over the entire domain at different times (every 6 hours) is determined from the GFS analysis fields as well as the three model runs. A $3^o \times 3^o$ box is considered around the centre of the minSLP region. The 48hour time averaged (06Z 14th November - 06Z 16th November 2009) and box averaged vertical profiles of this observed (GFS analysis) relative vorticity is compared with the model runs in Fig 6.12a. The relative vorticity as seen in NCEP analysis at the

lower levels are not reproduced by all the model forecasts. Both the assimilation runs have simulated vorticity profiles that are relatively closer to analysis as compared to CTRL run. The CTRL run has, however, simulated large values of cyclonic relative vorticity in the lower levels (4.5 $\times 10^{-5} s^{-1}$) with a maximum around 850 hPa. The EnKF run simulated cyclonic vorticity is closer to GFS analysis at heights above 800 hPa although they differ at lower levels. Also, the height at which the maximum relative vorticity observed in the EnKF run is closer to the GFS analysis and these are quite different for the CTRL and 3DVar runs. In order to quantify the departure of the time averaged and area averaged relative vorticity of model runs from the analysis, the rmse of the above is calculated with respect to the GFS analysis summed over all the vertical levels. and are as follows ; CTRL run: 2.15 $\times 10^{-5} s^{-1}$, 3DVar run :1.35$\times 10^{-5} s^{-1}$ and EnKF run : 7.22 x $\times 10^{-6} s^{-1}$, indicating clearly the closeness of the EnKF run to the analysis.

The area averaged and box averaged temperature anomaly is calculated using the same $3^o \times 3^o$ box over the minSLP region. The difference between the average temperature within this box with that over the entire domain at different pressure levels is considered as the temperature anomaly. Fig. 6.12b shows that at all the levels, the simulated temperature anomaly from the EnKF run is closer to the analysis than that from the 3DVar and the CTRL runs. The rmse of the 48 hour time averaged area averaged temperature anomaly of the model runs averaged over the vertical levels with respect to the GFS analysis for CTRL run is 0.51 K, 3DVar run is 0.63 K and EnKF run is 0.43 K. Clearly, the EnKF run has simulated the temperature structure closest to the GFS analysis among the three model runs.

6.4.2 Case 2 : Depression over Comorin region during November 2011

Case Investigated

Under the influence of an active ITCZ, a cyclonic circulation formed over southwest Bay of Bengal off Sri Lanka and south Tamil Nadu coast on 23rd November 2011 and manifested as a low pressure area over southwest Bay of Bengal and adjoining Sri Lanka on 24th November 2011. The above system moved west-northwestwards and

became well marked on 25th November 2011 and concentrated into a depression at 0300 UTC of 26th November 2011 over the Comorian area. During the intensification of the above mentioned low pressure system to depression, heavy rainfall was obtained over the peninsular region.

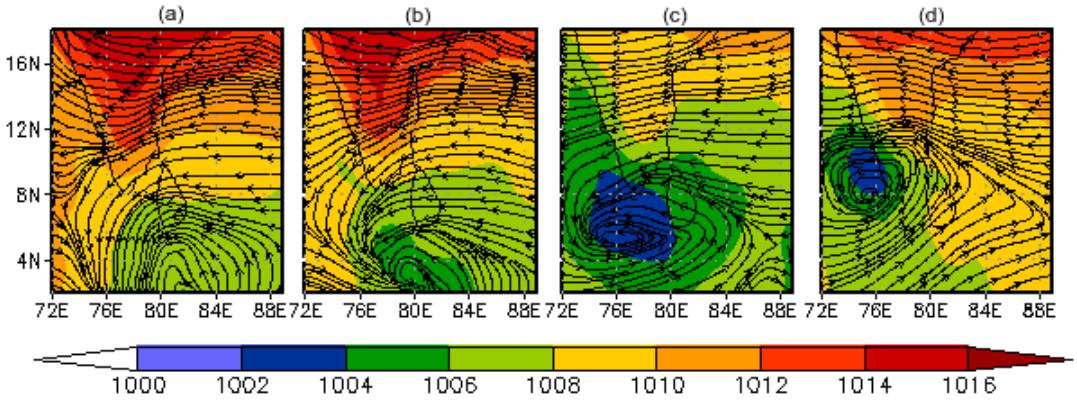

Figure 6.13: MSLP and 850hPa streamline pattern from NCEP analysis at 06Z (a) and 18Z of 14th November (b) , 12Z of 15th November (c) and 06Z of 16th November (d) for the second case study

The MSLP field from NCEP analysis shows clearly the existence of low pressure near Sri Lanka on 24November 2011 (Fig.6.13a,6.13b). The analysis shows that the above low pressure area has intensified slightly on 25th November 2011 moving northwestwards and is now lying over the Comorin area (Fig.6.13c) and (Fig.6.13d). The associated streamline pattern at 850hPa from the analysis reveals a cyclonic circulation on 24th November 2011.

Experimental

The WRF model is configured in the same manner in terms of model domain, horizontal resolution, number of vertical levels and various parametrization schemes utilized for case study 1 considered in the previous section in this chapter. Three experiments are performed to analyse the impact of assimilation of OSCAT winds on the simulation of this heavy rainfall event, namely (i) control (CTRL) run, (ii) assimilation run using 3DVar (3DVar run) and (iii) assimilation run using EnKF (EnKF run). The initial and boundary conditions for WRF model are obtained from NCEP GFS fields in all the three experiments. No observations are assimilated in the CTRL run.

The model is integrated from 18Z 22November 2011 to 06Z 26November 2011 in the CTRL run. The 3DVar and EnKF runs, are started from 18Z 22th November 2011. OSCAT wind observations are assimilated thrice after this initial spin up - at 06Z and 18Z 23th November 2011 and 06Z 24th November 2011. From the assimilated state at 06Z 24th November 2011, the model is integrated for 48 hours without further assimilation of OSCAT winds, in a free forecast mode. The EnKF run is identical to the 3DVar run, except that EnKF method is employed. The ensemble filter used here uses the same inflation and localization values that have been used for the heavy rainfall case 1 considered. The model simulated fields during the free forecast run (during 06Z 24th November 2011 to 06Z 26th November 2011) for all the three experiments are compared with each other and validated with observations and analyses to investigate the impact of assimilation. Here also, rain-flagged OSCAT wind observations are not included in the assimilation runs. This case study also has used the same configuration of EnKF system as in the first case study.

Results

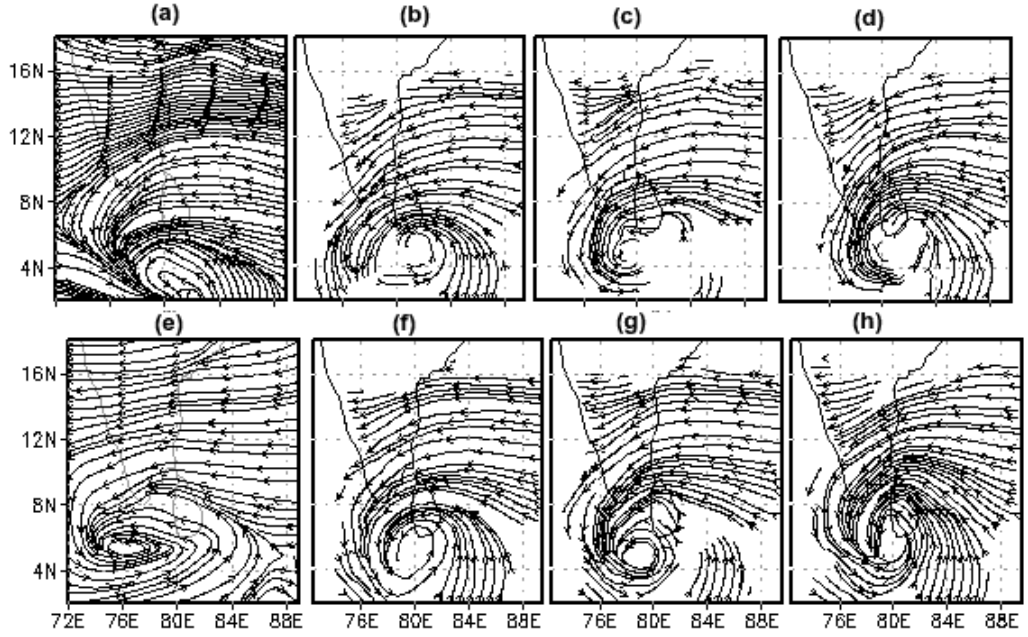

Figure 6.14: 850hPa moisture streamline (qu,qv) patterns at 12 hours of forecast from the NCEP analysis, CTRL, 3DVar and EnKF runs (a-d) and at 30 hours of forecast (e-h) for the second case study.

The 850hPa lower tropospheric moisture flux at 12 hours of forecast , i.e. at 18Z 24th November (Fig 6.7b-6.7d) and at 30hours of forecast , i.e. at 12Z 25th November 2011 (Fig.6.7f-6.7h) as simulated by the model are compared with that from NCEP analysis fields in Fig 6.7a and 6.7e. In the model runs, the moisture convergence seems to be centered near the south of SriLanka where as the NCEP analysis shows this feature further south. The 30hr model forecast shown in Fig6.7f-6.7h also indicates that the model has simulated more moisture flux nearer to the Indian peninsular region as compared with the NCEP analysis as shown in Fig. 6.7e.

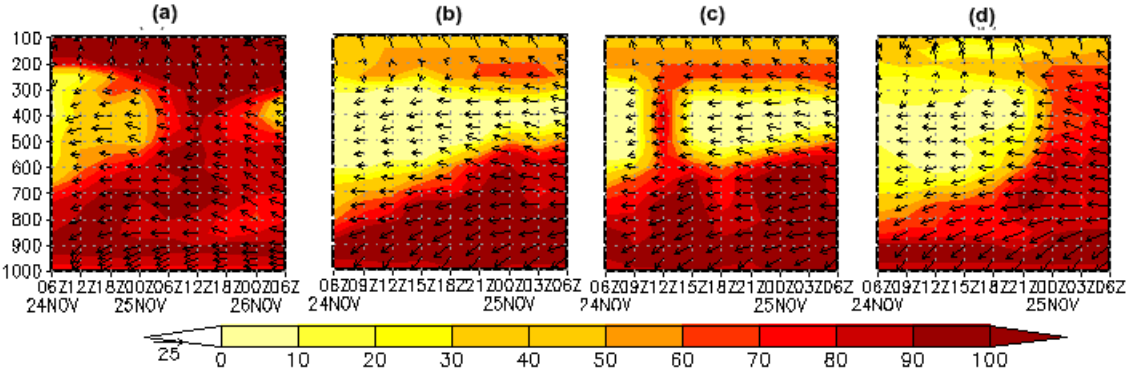

Figure 6.15: Height-time section of the wind speeds and relative humidity over Chennai from analysis (a) and the model runs (b-d) for the second case study.

Fig 6.15(a-d) shows the time-height variation of relative humidity and wind vectors over Chennai on the east coast of peninsular India during the free forecast period. The figure indicates that the three model runs given by Fig. 6.15(b-d) are able to reproduce the easterly wind pattern observed over Chennai during the North East monsoon period. The NCEP analysis of relative humidity field (Fig.6.15a) shows that the humidity is in general high over Chennai during this period. However, the CTRL and 3DVar model runs (Fig.6.15b,c) have simulated a drier atmosphere above around 600hPa at almost all times during the free forecast period. In EnKF simulation also, the levels above 600 hPa are drier during the day 1 forecast. However, during the day 2 forecast, the EnKF run has simulated more moisture in terms of relative humidity than that is simulated by the CTRL and 3DVar runs.

The moisture flux pattern and the relative humidity fields indicate that the EnKF run has simulated more moisture convergence and higher humidity values over the land

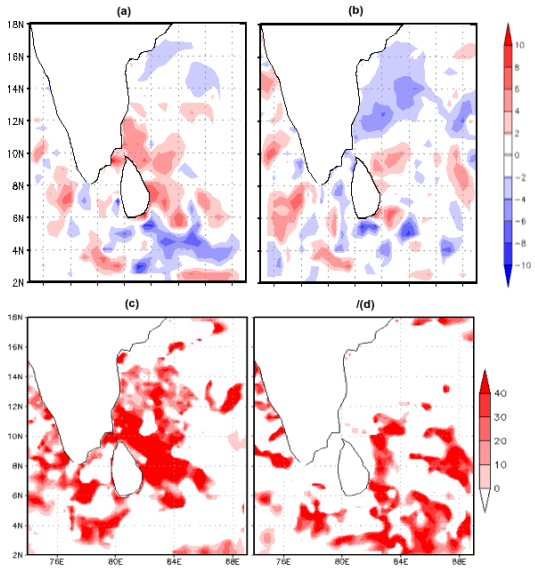

Figure 6.16: Spatial distribution of improvement parameter and forecast impact parameter in the simulation of wind magnitude with respect to ASCAT observations for 3DVar run (a),(c) and EnKF run(b),(d) for the second case study

region. This can impact the rainfall simulation and can result in heavier precipitation in the EnKF run over the Indian peninsula as compared with the CTRL and 3DVar runs.

Positive η values indicate positive impact of assimilation. The spatial distribution of improvement parameter for the 3DVar and EnKF runs are given in Fig.6.16a and 6.16b. It is seen that η is positive over large regions in the 3DVar run (Fig6.9a) indicating positive impact on the simulated wind magnitudes. However, the EnKF run

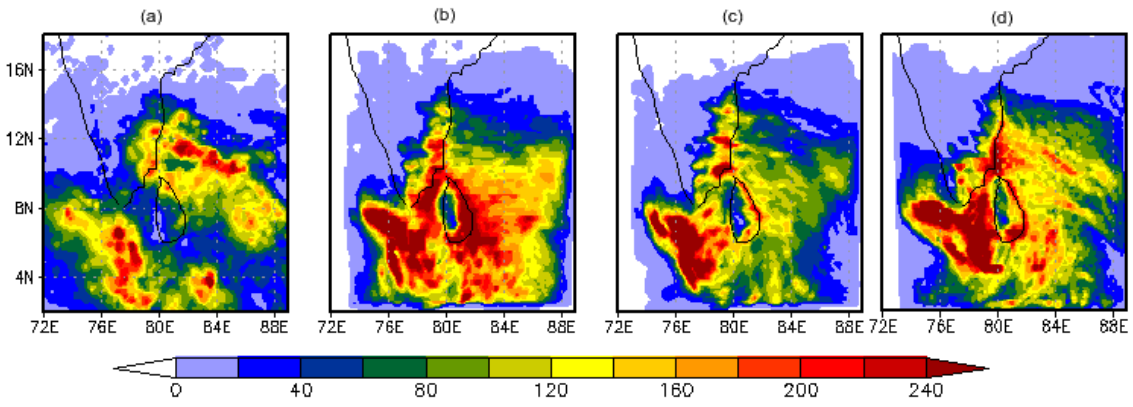

Figure 6.17: 48hr accumulated rainfall from TRMM(a), CTRL run(b), 3DVar run (c) and EnKF run (d) for the second case study

shows a slightly lower percentage of grid points where η is positive.

A positive FI indicates that the model forecast with assimilation compares better with observations than the forecast without assimilation. Fig.6.9c and 6.9d shows the spatial plot of FI for 3DVar and EnKF runs. The result of the 3DVar run has larger percentage of grid points where FI is positive as compared with the EnKF run.

Towards this end, 48 hour accumulated rainfall simulated by the CTRL, 3DVar and EnKF runs, are compared with the TRMM rainfall observations and are shown in Fig.6.17. The figure shows that the 3DVar run has reduced the extent of spurious heavy rainfall simulated by the CTRL run. It is also clear that EnKF run has simulated excessive rainfall as compared to 3DVar run. Furthermore, there are errors in the location and intensity of heavy rainfall simulated by both the 3DVar and EnKF runs. The 3DVar run has simulated heavy rainfall over a somewhat limited spatial extent as compared with TRMM observations. However, the spatial pattern of rainfall simulated by the EnKF run is close to that observed by TRMM. However, the EnKF run has simulated rainfall with more intensity than that observed by TRMM.

The quantitative verification of skill of rainfall forecast by the CTRL, 3DVar and EnKF runs are shown in Fig.6.18. The 48 hour accumulated rainfall simulated by the CTRL, 3DVar and EnKF runs are compared with the TRMM rainfall observations. The ETS score (Fig.6.18a) shows that EnKF runs are more skillful in simulating rainfall over all the thresholds as compared with the CTRL as well as 3DVar runs. Furthermore, the number of false alarms are lower and the probability of detection values are higher in EnKF run as compared with both CTRL and 3DVar runs over all the rainfall thresholds. However, the EnKF run exhibits large bias as compared with both CTRL and 3DVar runs. The above result is consistent with the inference that EnKF run is simulating heavier rainfall, a feature discerned from the spatial plot of accumulated rainfall simulation.

The vertical profiles of temperature anomaly and relative vorticity in a $3^o \times 3^o$ box around the depression centre is shown in Fig.6.19 (a) and (b), respectively. The vertical profile of temperature anomaly from GFS indicates a warm core system in lower and upper troposphere. The CTRL and 3DVar runs have simulated smaller temperature anomaly at the lower levels as compared with the GFS analysis. The thermal core structure simulated by the EnKF run is relatively more closer to that from the GFS

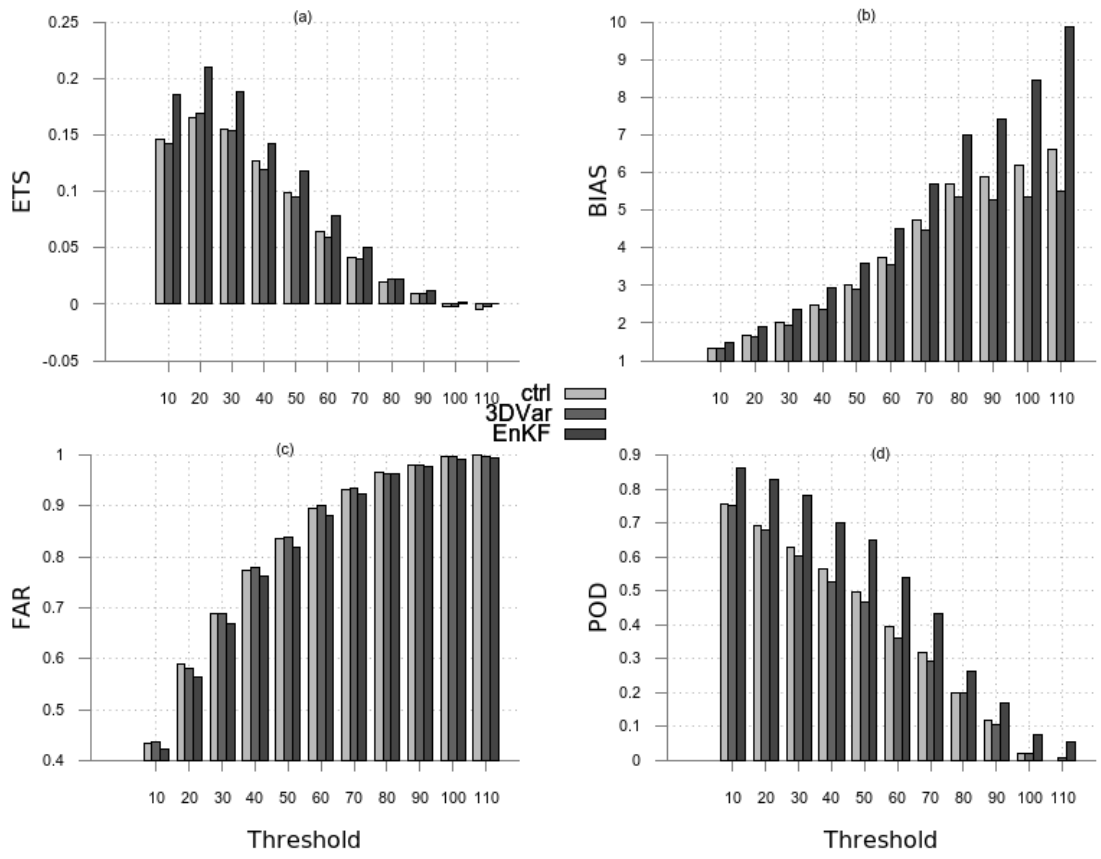

Figure 6.18: ETS (a), Bias (b), FAR (c) and POD (d) for CTRl, 3DVar and EnKF runs for 48 hr accumulated rainfall with respect to TRMM rainfall observations for the second case study

analysis at lower levels as compared with the CTRL and 3DVar runs. However, the warm core maximum at around 300 hPa seen in the GFS analysis is better captured by the 3DVar run as compared with the EnKF run. The vertical profile of relative vorticity around the center of the system is shown in Fig.6.19(b). For this case, the 3DVar run has simulated a higher relative vorticity throughout the troposphere as compared to the CTRL and EnKF runs as well with the GFS analysis.

6.4.3 Case 3 Depression over the Bay of Bengal during May 2013)

Case Investigated

Under the influence of a cyclonic circulation over northwest Bay of Bengal off Odisha - West Bengal coasts, a low pressure area formed over the North Bay of Bengal and neighbourhood. The low pressure system intensified into a Depression and

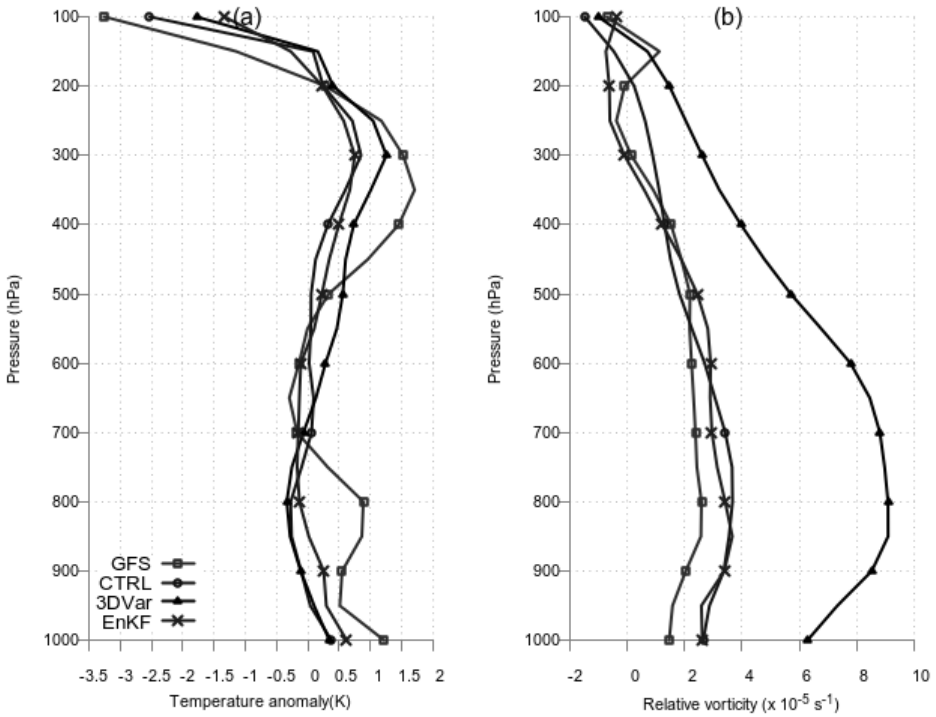

Figure 6.19: Vertical profiles of temperature anomaly (a) and relative vorticity (b) in a $3^o \times 3^o$ box around the depression centre for the second case study

was centered near 21^0 N, 89.5^0 E, at 0300 UTC of 29th May 2013. Moving north-northwestwards, the depression was located near 21.7^0 N, 88.8^0 E, at 1200 UTC of 29th May 2013 and crossed the West Bengal coast near 21.8^0 N, 88.7^0 E between 1330 and 1430 UTC on the same day. The above mentioned depression caused heavy to very heavy rainfall during its passage over the Indian region.

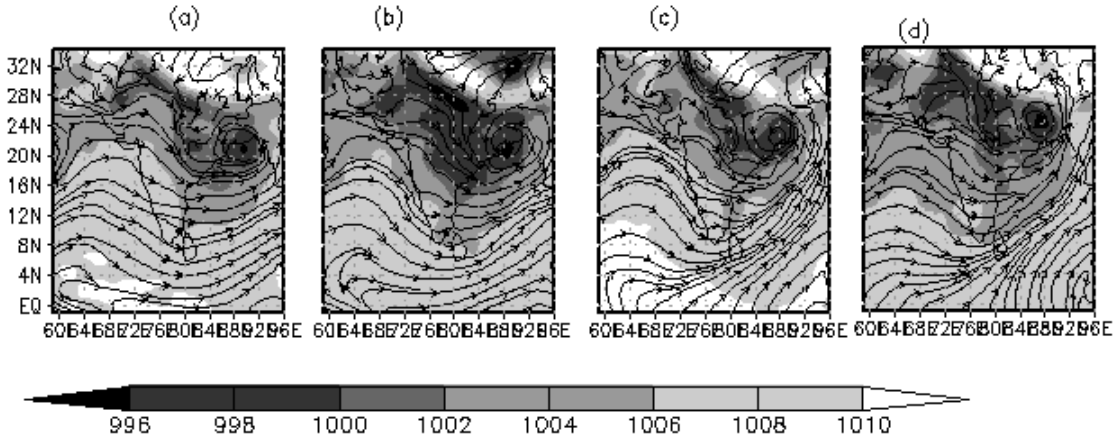

Figure 6.20: MSLP and 850hPa streamline pattern from NCEP analysis at 00Z (a) and 06Z of 29 May 2013 (b) , 18Z 30 May 2013 (c) and 00Z of 31 May 2013 (d) for the third case study

The MSLP field from NCEP analysis depicts the depression embedded in the monsoon trough (Fig.6.20a-d). The analysis shows that north-westward movement of the depression centre during 29May - 31May2013.

Experimental Details

The two heavy rainfall cases investigated in this chapter have been performed with a limited computational requirement by employing a relatively smaller domain to lower the computational costs involved since the EnKF utilizes large number of ensemble forecasts. However, the use of a relatively smaller domain has the limitation that the number of assimilated observations is low. Furthermore, the domain over peninsular India considered in the previous two case studies in this chapter do not include many conventional observations within it. Moreover, a relatively smaller domain might impact the spread among the ensemble members in the EnKF run due to closer boundaries. Hence it is envisaged that an experiment be performed using EnKF method which makes use of a relatively larger domain (shown in Fig.6.1) . The above domain has 350 $\times$ 350 grid points in the E-W and N-S directions with a horizontal resolution of 27km and has 36 levels in the vertical. The same parametrisation schemes in the previous case studies reported in this chapter are used here for the model integration.

The ensembles are generated using the multiphysics ensemble approach wherein each of the ensemble members has a unique physical parametrization combination. This ensures sufficient spread in the ensemble. An ensemble with 40 ensemble members are used in this third case study. A lower number of ensemble members appears to be adequate for this case study since the ensemble members are generated using the multiphysics scheme. This ensures that the ensemble members are sufficiently different from each other and thus provide adequate spread. Furthermore, to minimize the impact of lower number of ensemble members, a covariance inflation of 1.2 is also employed in this case study. The conventional surface and upper air observations as well as satellite winds over the oceans available from NCEP are used in this case study for assimilation.

The various experiments are configured as follows - a control (CTRL) run is performed without ingesting any observations from 12Z 28th May 2013 to 00Z 31 May 2013. Two assimilation runs are also performed, namely, a 3DVar run and an EnKF run. Both these model runs are performed in a similar manner. WRF model is intialized from GFS model forecasts at 12Z 28th May 2013. A 12 hr model spin up is performed without assimilating any observations. Thereafter, conventional observations and satellite winds are assimilated thrice - at 00Z 28th May 2013, 12Z 28th May 2013 and 00Z 29th May 2013. From the 3DVar as well as EnKF analyses at 00Z 29th May 2013, a 48 hour model forecast is performed without any further assimilation of observations.

Results

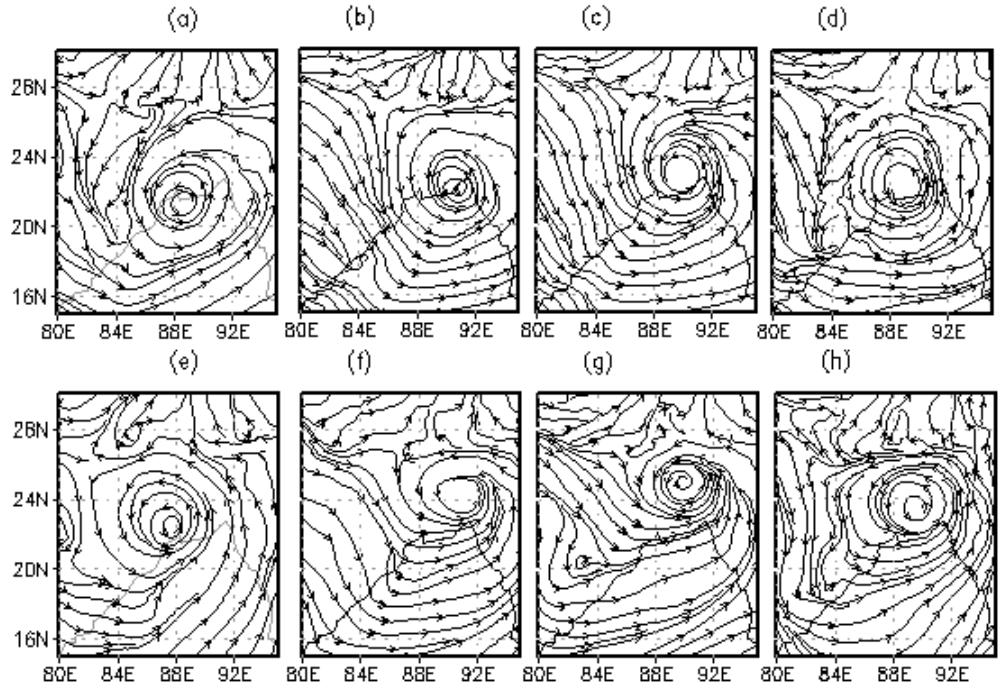

Figure 6.21: 850hPa moisture streamline (qu,qv) patterns at 12 hours of forecast from the NCEP analysis, CTRL, 3DVar and EnKF runs (a-d) and at 30 hours of forecast (e-h) for the third case study.

The 850hPa lower tropospheric moisture flux at 12 hours of forecast , i.e. at 12Z 29th May 2013 (Fig 6.7b-6.7d) and at 30hours of forecast , i.e. at 18Z 30th May 2013 (Fig.6.7f-6.7h) as simulated by the model are compared with that from NCEP analysis fields in Fig 6.7a and 6.7e. The Fig..6.7 indicates that bu 12 hours of model forecast, the moisture convergence pattern has moved more inland as compared with the analysis. The Fig.6.7f-6.7h) indicates that the 3DVar run has simulated more lower level moisture convergence as compared with the EnKF and CTRL runs as well as the NCEP analysis.

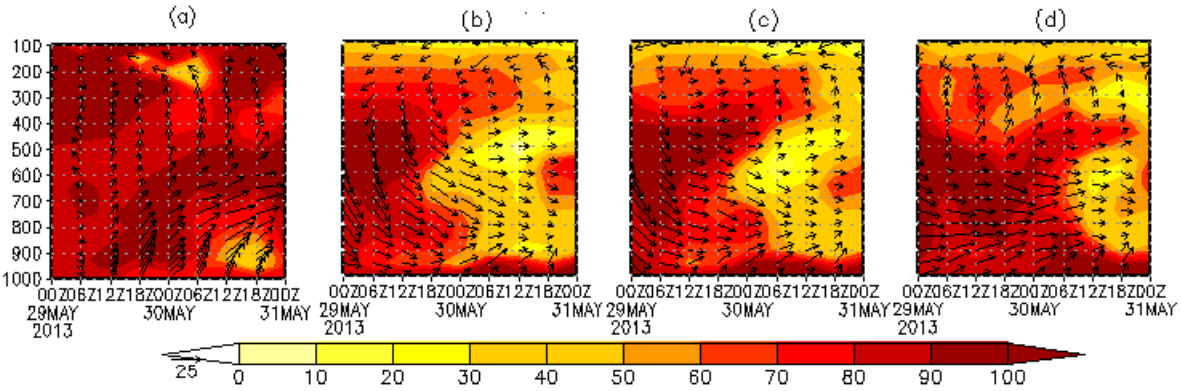

Figure 6.22: Height-time section of the wind speeds and relative humidity over the location of maximum observed precipitation from analysis (a) and the model runs (b-d) for the third case study.

Fig 6.22(a-d) shows the time-height variation of relative humidity and wind vectors over $20.75^oN, 89.1^oE$ on the where the maximum amount of rainfall has been observed during the free forecast period. The figure indicates that the three model runs have simulated a drier model upper levels over this location after about 24 hours of forecast as compared to the NCEP analysis. However, the EnKF run has simulated more relative humidity as compared to 3DVar and CTRL runs.

The track of the depression simulated by the three model runs are compared with the IMD best track data. From Fig.6.23 it is clear that EnKF assimilation of conventional and satellite wind observations has improved the simulation of location of depression centre at the initial forecast hours , as compared with the CTRL and 3DVar simulations. After the first eighteen hours of forecast, the EnKF simulated track does deviate farther from the IMD observation. It is quite possible that deviation with respect to observed track manifest after the initiation of free forecast ; larger the time after the initiation of free forecast, larger is the deviation. Additional assimilation cycles encompassing the free forecast period may possibly provide for a better track simulation by the EnKF.

The 48 hour accumulated rainfall observed by TRMM is shown in Fig.6.24(a). This is compared with the rainfall simulated by the CTRL run (Fig.6.24(b)), 3DVar run (Fig.6.24(c)) and the EnKF run (Fig.6.24(d)). The maximum intensity of rainfall is seen over the Head Bay region in the TRMM rainfall observations. There is rain over the west coast of India also. It is clear that the CTRL run has simulated

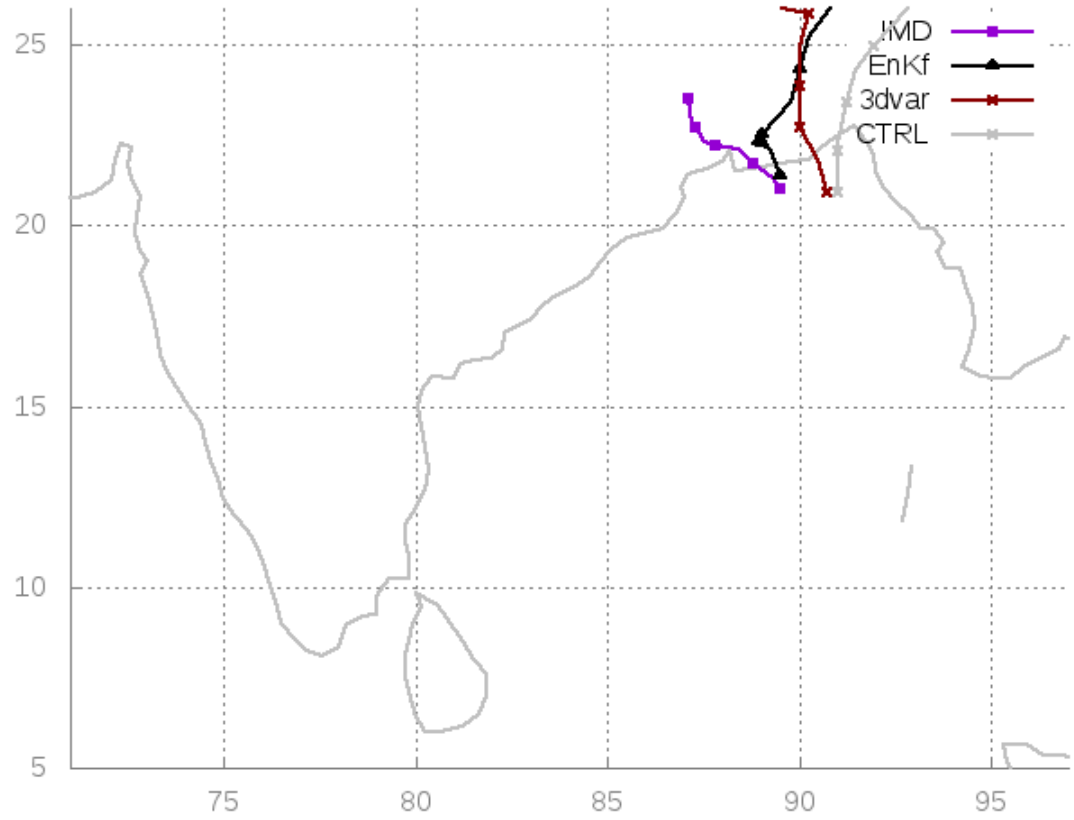

Figure 6.23: Depression track simulated by CTRL, 3DVar and EnKF runs compared with the IMD observation for the third case study

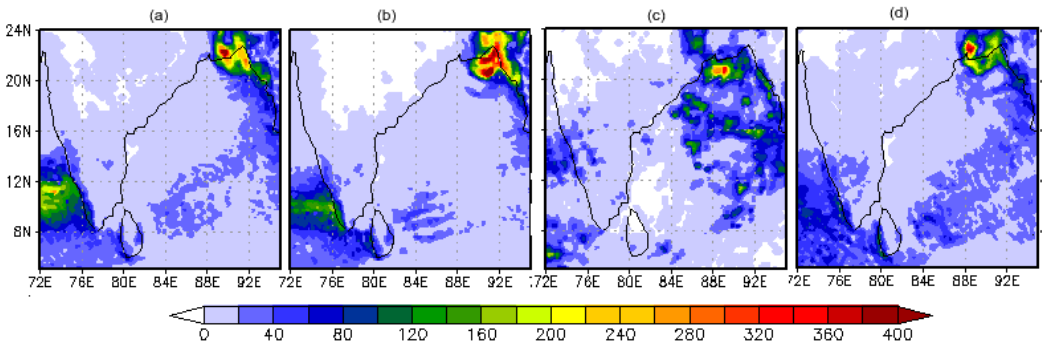

Figure 6.24: 48 hour accumulated precipitation simulated by CTRL (b), 3DVar (c) and EnKF (d) compared with TRMM observations (a)for the third case study

excess rainfall over the Head Bay region. Also, the spatial extend of rainfall over the west coast of India is seen to be lower in the CTRL run. Furthermore, CTRL run has produced relatively smaller spatial extend of heavy rainfall over the Head Bay as compared with the EnKF simulation. However, there is error in the location of maximum precipitation in the CTRL run. Furthermore, the rainfall over the Indian west coast in not reproduced by the 3DVar assimilation run. EnKF assimilation run, however has simulated larger extent of heavy rainfall over the Bay of Bengal as compared to the CTRL run. The location of maximum precipitation in the EnKF run is also closer to the TRMM observation. Furthermore, the spatial extent of rainfall over the west coast of India as well as over the Bay of Bengal region are better simulated by the EnKF run.

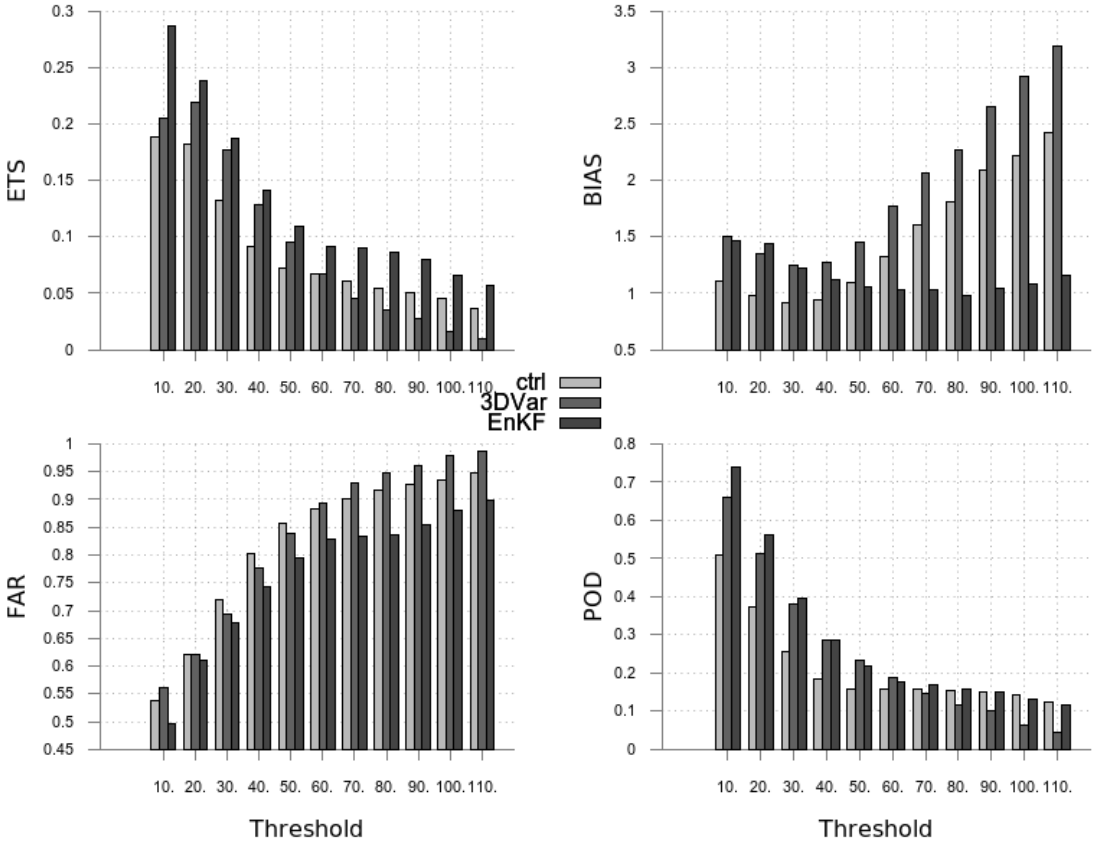

Figure 6.25: ETS (a), Bias (b), Far (c) and POD (d) of 48 hr accumulated rainfall calculated with respect to TRMM observations for the third case study

The improved simulation of accumulated rainfall in the EnKF run can be further substantiated by utiliszing the skill scores of rainfall simulations as shown in Fig.6.25. The ETS (a), Bias(b), FAR(c) and POD(d) in Fig.6.25 shows that EnKF assimilation

has resulted in better simulation of accumulated rainfall in terms of higher ETS, lower Bias, lower FAR and higher POD as compared with both the CTRL and 3DVar runs.

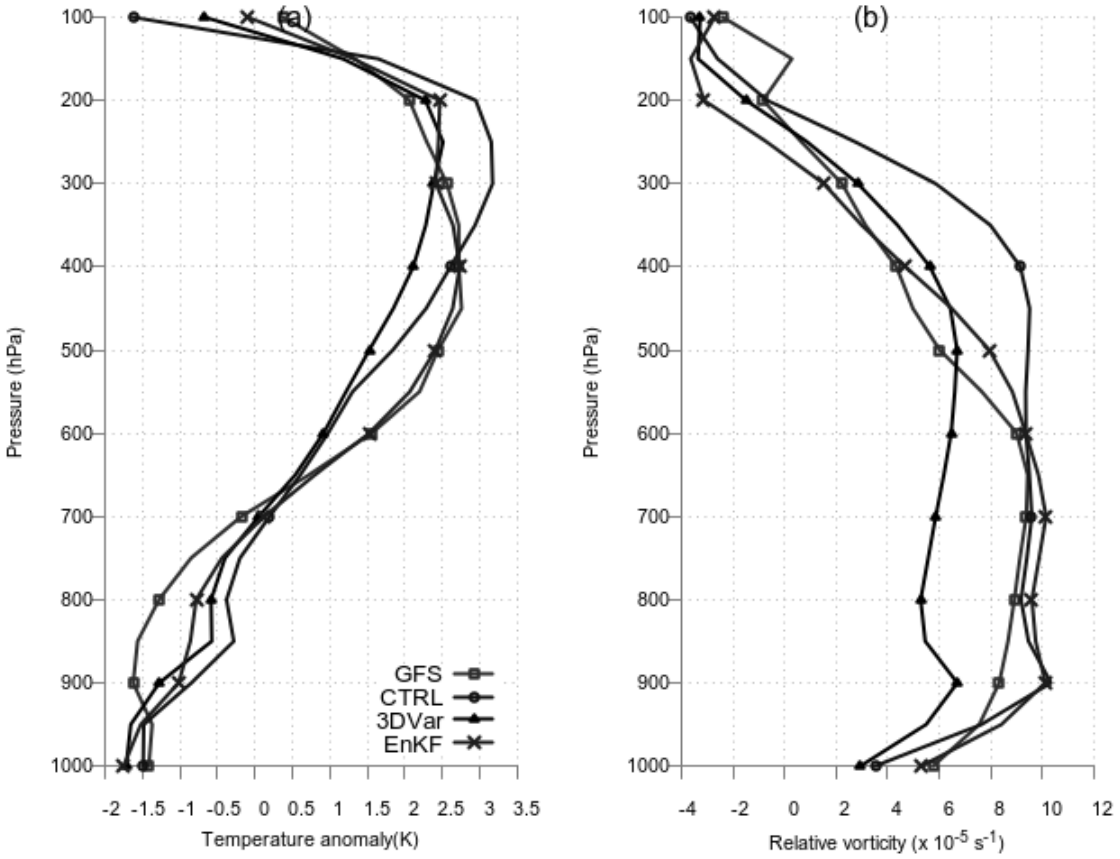

Figure 6.26: vertical profile of temperature anomaly (a) and relative vorticity (b) in a $3^o \times 3^o$ box around the depression centre for the third case study

The thermal structure of the monsoon depression is depicted using the temperature anomaly in a $3^o \times 3^o$ box around the depression centre for CTRL, 3DVar and EnKF runs in Fig.6.26(a). These are compared with the temperature anomaly profile from GFS analysis. Fig.6.26(a) shows that the thermal structure of the depression simulated by the EnKF run is more closer to GFS analysis that that simulated by the 3DVar and CTRL runs. Similarly, the vertical profile of relative vorticity simulated by EnKF run is closer to that from the GFS analysis as compared with the simulations from both CTRL and 3DVar runs (Fig.6.26(b)).

6.4.4 Summary and Conclusions

The EnKF assimilation technique is known to provide better analyses and hence improved prediction as compared with the 3DVar assimilation method in both global and regional models. However studies involving the use of EnKF assimilation technique for mesoscale weather systems over the Indian region is absent and hence the need for such an investigation as undertaken in this study. It is seen that the EnKF technique does provide improved analyses and hence, improved forecasts as compared with 3DVar assimilation. This is demonstrated by simulating three case studies of heavy rainfall events over the Indian region. The first two case studies investigated here are associated with the North east monsoon season while the third case corresponds to heavy rainfall event during a monsoon depression. The results of all the three case studies indicate that the EnKF assimilation run has provided for improved simulation of rainfall. Also the EnKF run has provided for better simulations of wind, moisture as well as temperature structures associated with the above mentioned weather events. The major reason for the improvements in the simulation associated with the EnKF run is the use of flow dependent $\mathbf{B}$ matrix in the EnKF assimilation system.

3DVar system assumes the correlations among the control variables in the $\mathbf{B}$ matrix to be homogeneous and isotropic in the horizontal. This assumption is invalid as the correlation structures in the background error covariances will be spatially heterogeneous depending on the synoptic situation that is simulated. The ensemble derived covariances capture these covariance structures better than the 3DVar background error covariances in the convectively active systems. Hence, in an EnKF analysis, the spread of observation information depends on the actual flow patterns of the day and this ensures improved analysis as compared to the 3DVar analysis.

Further studies using EnKF method are necessary over the Indian region. One of the limitations in the use of EnKF systems is the need for larger computational resources as compared with the 3DVar assimilation system. Further research is necessary to improve the existing EnKF methods and to deduce better methods over the EnKF system. The currently emerging hybrid data assimilation techniques which combines flow dependent $\mathbf{B}$ matrix from an ensemble along with the 3DVar technique is one of the possible options available for future research.

CHAPTER 7

Conclusions

It is well known that the observations obtained from the various satellite-based sensors play a significant role in improving the initial conditions for numerical weather models. The present thesis explores this and further investigates how the impact is affected through the modification of background error covariances in 3DVar. Furthermore, the improved impact of satellite data assimilation using the Ensemble Kalman filter vis.a.vis 3DVar is also investigated here.

The impact of assimilation of satellite radiances from MeghaTropiques satellite's SAPHIR sensor in the simulation of three tropical cyclones over the Bay of Bengal has been investigated. The results indicate that the SAPHIR radiances have a positive impact on the simulation of features of all the three tropical cyclone . The assimilation has been performed using the WRF 3DVar system. All the model variables like temperature, wind, sea level pressure as well as the rainfall are better simulated due to the humidity information assimilated from the SAPHIR observations. This is attributed to the fact that since the formation and development of cyclones are influenced by moist convective processes in the atmosphere, inclusion of moisture observations does improve the simulation of such processes and thereby leads to better simulation of tropical cyclones.

One of the major factors that influence the data assimilation system is the background error covariance matrix ($\mathbf{B}$). Since the background error covariance matrix is a huge matrix and cannot be determined exactly, $\mathbf{B}$ matrix can be estimated using different methods. Hence, the representation of $\mathbf{B}$ matrix in an assimilation system has scope for further improvement. Improved representation of $\mathbf{B}$ matrix will, in principle result in better analysis since $\mathbf{B}$ determines how effectively the observations correct the background information. This thesis investigates the impacts of two different methods in which $\mathbf{B}$ is represented in WRF 3DVar system. The default option used in regional assimilation applications is called as cv5 option. The results of cv5 option are compared with the more recent and supposedly of specification of $\mathbf{B}$, which is termed as the cv6 option.

Three monsoon depression cases are chosen to investigate the impact of using these two $\mathbf{B}$ options in the cost function of the 3DVar system. Assimilation of conventional surface and upper air observations along with satellite wind and radiance observations from AMSU-A reveal that the formulation of $\mathbf{B}$ does have an impact on the analysis fields. The moisture field understandably is seen to have the maximum impact while employing the cv6 option which has impacted the model forecast fields like water vapor mixing ratio as well as precipitation. The improvement of the skill scores based on 48 hour accumulated precipitation in two out of three depression cases considered in this study are in the higher rainfall thresholds. The results provide positive encouragement for further investigations as well as provides a means for improving heavy rainfall forecasts using the WRF model.

Since the cv6 option is found to impact the analysis as well as the forecast fields in WRF 3DVar system, further investigations of its influence on the simulation of three tropical cyclones that formed over the Indian region have been undertaken. All the three tropical cyclones Thane, Nilam and Phailin have been simulated by using cv5 option as well as cv6 option for assimilation. In this study, in addition to the surface, upper air observations and satellite wind observations, SAPHIR radiances are also utilized for assimilation. One of the chief reasons for using the SAPHIR radiances is that they provide high resolution humidity information for a microwave sensor from a satellite which provides 5 to 6 passes over the tropical region. Since the cv6 option impacts the moisture field most, the analysis would have a better representation of moisture fields if the same were obtained from a high resolution microwave sensor which provides humidity observations with higher repetitivity. Since tropical cyclones depend on moisture convective processes for their genesis and growth, the cv6 option has been hypothesized to provide for better tropical cyclone simulations. Based on the above, it is found that the analysis fields show a higher moisture content in the lower model levels while using the cv6 option. Due to the above, the choice of formulation of $\mathbf{B}$ matrix also impacts the forecast fields of the model in a moderate manner.

The comparison of assimilation using cv5 and cv6 formulations lead to the following conclusions. (i) the form of correlation functions in the background error covariance matrix do influence the impact of 3DVar analysis. This is beacuse the modification of the background field by the observation information depends on the structure of model error correlations in the horizontal as well as vertical directions in

the control variable space (ii) the use of cv6 formulation for BEC impacts the model's moisture field significantly. This is due to the provision for multivariate nature of the moisture field in the control variable transforms in the cv6 formulation as compared to the cv5 formulation (iii) the modified moisture analysis positively influences the model simulation of heavy rainfall associated with the monsoon depressions. The results of this study also clearly indicates that the simulation of tropical cyclones is also influenced by the choice of BEC option. Two out of three cyclones investigated in this study are found to be more intense in the cv6 formulation in terms of minimum sea level pressure and maximum wind speed as compared to the cv5 formulation. The above results are due to the fact that the background error covariance does provide for correlations of the humidity field with other fields like wind, temperature and surface pressure observations. Hence the observations of the above-mentioned fields do influence the moisture field and vice-versa.

The importance of formulation of $\mathbf{B}$ in a data assimilation system naturally leads one to employ methods that allow for improved, flow-dependent $\mathbf{B}$ statistics. The major disadvantage in the 3DVar technique is in its static and more or less isotropic nature of $\mathbf{B}$ matrix. It is known that the flow dependence can be incorporated in $\mathbf{B}$ through employing ensembles using EnKf method.

Ensembles of atmospheric state variables are generated by perturbing the model state variables (initial and boundary conditions) as well by through the utilization of different model physics schemes. The ensembles thus generated represent the evolution of model probability density function and hence it is expected that the background error estimated using the ensembles better represent the error evolution of the model system. The Ensemble Kalman filter (EnKF) routines available with Data Assimilation Research Testbed (DART) are used here to assimilate the ocean surface wind observations from Oceansat-2 scatterometer. The flow dependent $\mathbf{B}$ matrix used in EnKF results in better analysis which inturn provide improved forecasts. The model simulations of heavy rainfall events over peninsular Indian region using EnKF assimilation provides improved rainfall prediction , as compared with assimilation using 3DVar technique. Apart from the simulation of two heavy rainfall events, a case study of a monsoon depression has also been investigated using EnKF method. The results of the study provide proof of improvement in the simulation due to the flow dependent $\mathbf{B}$ statistics.

The present thesis addresses the main objectives of the study by utilizing a few case studies. It is known that while one cannot derive a broad conclusion based on the results of a few case studies alone, it is quite possible that the above results do provide for a positive inference . In the same, the results of the present thesis indicates that assimilating satellite radiance observations does improve the simulation of mesoscale weather systems in a limited area model. Furthermore, it can also be surmised that the use of an improved estimate of background error covariance can further enhance the positive impact of satellite observations on the simulation of weather phenomena. There is a significant factor of uncertainty associated with a limited area model where the boundary conditions also influence the background error, unlike in the case of a global model. Error evolution in the simulation of a mesoscale weather phenomenon also involve uncertainties that are different from those associated with a large scale system. Hence, better estimations of errors associated with the background flow can markedly influence the forecast of a limited area model like WRF.

However, the present study suffers from the following shortcomings. The thesis has focussed on the impact of formulation of background error covariances on assimilation of satellite observations. The above objective has been investigated by taking up a few case studies which are not adequate enough for drawing broad and general conclusions for the impact of formultion of background error covariance in the simulation of meteorological systems over India. Furthermore, the thesis has not systematically investigated one single weather system nor it has critically examined the impact of ingesting one type of observation for limited area models over India, which does not provide for generalization of the conclusions. Also, the perturbations for estimating background error covariance in the NMC method can be computed using an ensemble of model forecast, which has not been attempted in the present study . This study has not critically examined the sensitivity of various parameters such as (i) number of ensemble members , (ii) covariance inflation, (iii) localization etc while evaluating the impact of assimilating observations using EnKF method. The various ways of creating ensembles like multiphysics methods, perturbing initial and boundary conditions, utilizing SKEB scheme etc have been mentioned in the thesis. However, a critical examination of the impact of the above on the assimilation of observations has not been conducted. Furthermore, the impact of assimilating satellite radiances using EnKF has not been invetsigated in the present study.

The research queries that motivated the investigation carried out in this thesis can be further explored using the following avenues. Many other variants of the EnKF method such as the Ensemble Transform Kalman filter (ETKF) and Ensemble Adjustment Kalman filter (EAKF) methods have been in use in recent times. It is important to investigate on whether methods such as ETKF and EAKF perform better over the Indian region as compared to the EnKF method. Another aspect of research work possible as an extension of this study is in the use of satellite radiances in the EnKF system, especially radiances from sensors like SAPHIR from MeghaTropiques satellite which has no been addressed in the present study. A further interesting aspect could be to examine the impact of a 4DVar assimilation system over the Indian region and its benefits vis.a.vis the EnKF method.

REFERENCES

1. Adler, R. F., Negri, A. J., Keehn, P. R., and Hakkarinen, I. M. (1993). Estimation of monthly rainfall over japan and surrounding waters from a combination of low-orbit microwave and geosynchronous ir data. *Journal of Applied Meteorology*, 32(2):335–356.

2. Aguttes, J. P., Schrive, J., Goldstein, C., Rouze, M., and Raju, G. (2000). Megha-tropiques, a satellite for studying the water cycle and energy exchanges in the tropiques. In *Geoscience and Remote Sensing Symposium, 2000. Proceedings. IGARSS 2000. IEEE 2000 International*, volume 7, pages 3042–3044. IEEE.

3. Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., and Avellano, A. (2009). The data assimilation research testbed: A community facility. *Bulletin of the American Meteorological Society*, 90(9):1283–1296.

4. Anderson, J. L. (2001). An ensemble adjustment kalman filter for data assimilation. *Monthly weather review*, 129(12):2884–2903.

5. Anderson, J. L. and Anderson, S. L. (1999). A monte carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, 127(12):2741–2758.

6. Arakawa, A. (2004). The cumulus parameterization problem: Past, present, and future. *Journal of Climate*, 17(13):2493–2525.

7. Arakawa, A. and Schubert, W. H. (1974). Interaction of a cumulus cloud ensemble with the large-scale environment, part i. *Journal of the Atmospheric Sciences*, 31(3):674–701.

8. Auligné, T., McNally, A., and Dee, D. (2007). Adaptive bias correction for satellite data in a numerical weather prediction system. *Quarterly Journal of the Royal Meteorological Society*, 133(624):631–642.

9. Baker, N. L., Hogan, T., Campbell, W., Pauley, R., and Swadley, S. (2005). The impact of amsu-a radiance assimilation in the us navy's operational global atmospheric prediction system (nogaps). Technical report, DTIC Document.

10. Bannister, R. N. (2008a). A review of forecast error covariance statistics in atmospheric variational data assimilation. i: Characteristics and measurements of forecast error covariances. *Quarterly Journal of the Royal Meteorological Society*, 134(637):1951–1970.

11. Bannister, R. N. (2008b). A review of forecast error covariance statistics in atmospheric variational data assimilation. i: Characteristics and measurements of forecast error covariances. *Quarterly Journal of the Royal Meteorological Society*, 134(637):1951–1970.

12. Bannister, R. N. (2008c). A review of forecast error covariance statistics in atmospheric variational data assimilation. ii: Modelling the forecast error covariance statistics. *Quarterly Journal of the Royal Meteorological Society*, 134(637):1971–1996.

13. Barker, D., Huang, X.-Y., Liu, Z., Auligné, T., Zhang, X., Rugg, S., Ajjaji, R., Bourgeois, A., Bray, J., Chen, Y., et al. (2012). The weather research and forecasting model's community variational/ensemble data assimilation system: Wrfda. *Bulletin of the American Meteorological Society*, 93(6):831–843.

14. Barker, D. M., Huang, W., Guo, Y.-R., Bourgeois, A., and Xiao, Q. (2004). A three-dimensional variational data assimilation system for mm5: Implementation and initial results. *Monthly Weather Review*, 132(4):897–914.

15. Bergthórsson, P. and Döös, B. R. (1955). Numerical weather map analysis1. *Tellus*, 7(3):329–340.

16. Berner, J., Ha, S.-Y., Hacker, J., Fournier, A., and Snyder, C. (2011). Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations. *Monthly Weather Review*, 139(6):1972–1995.

17. Betts, A. and Miller, M. (1986). A new convective adjustment scheme. part ii: Single column tests using gate wave, bomex, atex and arctic air-mass data sets. *Quarterly Journal of the Royal Meteorological Society*, 112(473):693–709.

18. Bishop, C. H., Etherton, B. J., and Majumdar, S. J. (2001). Adaptive sampling with the ensemble transform kalman filter. part i: Theoretical aspects. *Monthly weather review*, 129(3):420–436.

19. Bocquet, M., Pires, C. A., and Wu, L. (2010). Beyond gaussian statistical modeling in geophysical data assimilation. *Monthly Weather Review*, 138(8):2997–3023.

20. Brogniez, H., Kirstetter, P.-E., and Eymard, L. (2013). Expected improvements in the atmospheric humidity profile retrieval using the megha-tropiques microwave payload. *Quarterly Journal of the Royal Meteorological Society*, 139(673):842–851.

21. Buehner, M., Houtekamer, P., Charette, C., Mitchell, H. L., and He, B. (2010a). Intercomparison of variational data assimilation and the ensemble kalman filter for global deterministic nwp. part i: Description and single-observation experiments. *Monthly Weather Review*, 138(5):1550–1566.

22. Buehner, M., Houtekamer, P., Charette, C., Mitchell, H. L., and He, B. (2010b). Intercomparison of variational data assimilation and the ensemble kalman filter for global deterministic nwp. part ii: One-month experiments with real observations. *Monthly Weather Review*, 138(5):1567–1586.

23. Burgers, G., Jan van Leeuwen, P., and Evensen, G. (1998). Analysis scheme in the ensemble kalman filter. *Monthly weather review*, 126(6):1719–1724.

24. Chen, S.-H. (2007). The impact of assimilating ssm/i and quikscat satellite winds on hurricane isidore simulations. *Monthly weather review*, 135(2):549–566.

25. Chen, Y., Rizvi, S. R., Huang, X.-Y., Min, J., and Zhang, X. (2013). Balance characteristics of multivariate background error covariances and their impact on analyses and forecasts in tropical and arctic regions. *Meteorology and Atmospheric Physics*, 121(1-2):79–98.

26. Courtier, P., Thépaut, J.-N., and Hollingsworth, A. (1994). A strategy for operational implementation of 4d-var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, 120(519):1367–1387.

27. Daley, R. (1993). *Atmospheric data analysis*. Number 2. Cambridge university press.

28. Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., et al. (2011). The era-interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656):553–597.

29. Dee, D. P. (2005). Bias and data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3323–3343.

30. Derber, J. C. and Wu, W.-S. (1998). The use of tovs cloud-cleared radiances in the ncep ssi analysis system. *Monthly Weather Review*, 126(8):2287–2299.

31. Desbois, M., Roca, R., Eymard, L., Viltard, N., Viollier, M., Srinivasan, J., and Narayanan, S. (2003). The megha-tropiques mission. In *Third International Asia-Pacific Environmental Remote Sensing Remote Sensing of the Atmosphere, Ocean, Environment, and Space*, pages 172–183. International Society for Optics and Photonics.

32. Descombes, G., Auligné, T., Vandenberghe, F., Barker, D., and Barré, J. (2015). Generalized background error covariance matrix model (gen_be v2. 0). *Geoscientific Model Development*, 8(3):669–696.

33. Dhanya, M. and Chandrasekar, A. (2014). Improved rainfall simulation by assimilating oceansat-2 surface winds using ensemble kalman filter for a heavy rainfall event over south india. *Geoscience and Remote Sensing, IEEE Transactions on*, 52(12):7721–7726.

34. Dudhia, J. (1989). Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *Journal of the Atmospheric Sciences*, 46(20):3077–3107.

35. Ehrendorfer, M. (2007). A review of issues in ensemble-based kalman filtering. *Meteorologische Zeitschrift*, 16(6):795–818.

36. Evensen, G. (1994). Inverse methods and data assimilation in nonlinear ocean models. *Physica D: Nonlinear Phenomena*, 77(1-3):108–129.

37. Evensen, G. (2003). The ensemble kalman filter: Theoretical formulation and practical implementation. *Ocean dynamics*, 53(4):343–367.

38. Eyre, J., Kelly, G., McNally, A., Andersson, E., and Persson, A. (1993). Assimilation of tovs radiance information through one-dimensional variational analysis. *Quarterly Journal of the Royal Meteorological Society*, 119(514):1427–1463.

39. Feifei, S. and Jinzhong, M. (2014). Assimilating amsu-a radiance data with the wrf hybrid en3dvar system for track predictions of typhoon megi (2010). *ADVANCES IN ATMOSPHERIC SCIENCES*, page 112.

40. Fisher, M. (2003). Background error covariance modelling. In *Seminar on Recent Development in Data Assimilation for Atmosphere and Ocean*, pages 45–63.

41. Furrer, R. and Bengtsson, T. (2007). Estimation of high-dimensional prior and posterior covariance matrices in kalman filter variants. *Journal of Multivariate Analysis*, 98(2):227–255.

42. Gandin, L. S. and Hardin, R. (1965). *Objective analysis of meteorological fields*, volume 242. Israel program for scientific translations Jerusalem.

43. Gaspari, G. and Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125(554):723–757.

44. Godbole, R. V. (1977). The composite structure of the monsoon depression. *Tellus*, 29(1):25–40.

45. Govindankutty, M. and Chandrasekar, A. (2010). Effect of 3dvar assimilation of modis temperature and humidity profiles on the dynamic and thermodynamic features of three monsoon depressions over the bay of bengal. *Meteorology and Atmospheric Physics*, 107(1-2):65–79.

46. Govindankutty, M. and Chandrasekar, A. (2011). Impact of assimilation of atovs temperature and humidity and ssm/i total precipitable water on the simulation of a monsoon depression. *Natural hazards*, 59(3):1647–1669.

47. Greeshma, M., Srinivas, C., Yesubabu, V., Naidu, C., Baskaran, R., and Venkatraman, B. (2015). Impact of local data assimilation on tropical cyclone predictions over the bay of bengal using the arw model. *Ann. Geophys*, 33:805–828.

48. Grell, G. A. (1993). Prognostic evaluation of assumptions used by cumulus parameterizations. *Monthly Weather Review*, 121(3):764–787.

49. Grell, G. A. and Dévényi, D. (2002). A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophysical Research Letters*, 29(14).

50. Grell, G. A. and Freitas, S. R. (2014). A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. *Atmospheric Chemistry and Physics*, 14(10):5233–5250.

51. Guo, Y., Lin, H., Ma, X., Huang, X., Terng, C., and Kuo, Y. (2006). Impact of wrf-var (3dvar) background error statistics on typhoon analysis and forecast. In *Seventh WRF Users' Workshop*.

52. Ha, J.-H. and Lee, D.-K. (2012). Effect of length scale tuning of background error in wrf-3dvar system on assimilation of high-resolution surface data for heavy rainfall simulation. *Advances in Atmospheric Sciences*, 29(6):1142–1158.

53. Hamill, T. M., Whitaker, J. S., and Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble kalman filter. *Monthly Weather Review*, 129(11):2776–2790.

54. Han, J. and Pan, H.-L. (2011). Revision of convection and vertical diffusion schemes in the ncep global forecast system. *Weather and Forecasting*, 26(4):520–533.

55. Harris, B. and Kelly, G. (2001). A satellite radiance-bias correction scheme for data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 127(574):1453–1468.

56. Hayden, C. M. and Purser, R. J. (1995). Recursive filter objective analysis of meteorological fields: Applications to nesdis operational processing. *Journal of Applied Meteorology*, 34(1):3–15.

57. Hollingsworth, A. and Lönnberg, P. (1986). The statistical structure of short-range forecast errors as determined from radiosonde data. part i: The wind field. *Tellus A*, 38(2):111–136.

58. Hong, S.-Y., Noh, Y., and Dudhia, J. (2006). A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*, 134(9):2318–2341.

59. Hong, S.-Y. and Pan, H.-L. (1996). Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Monthly weather review*, 124(10):2322–2339.

60. Houtekamer, P. L. and Mitchell, H. L. (1998). Data assimilation using an ensemble kalman filter technique. *Monthly Weather Review*, 126(3):796–811.

61. Hunt, B. R., Kostelich, E. J., and Szunyogh, I. (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform kalman filter. *Physica D: Nonlinear Phenomena*, 230(1):112–126.

62. Janjic, Z. I. (1994). The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Monthly Weather Review*, 122(5):927–945.

63. Jones, T. A., Otkin, J. A., Stensrud, D. J., and Knopfmeier, K. (2013). Assimilation of satellite infrared radiances and doppler radar observations during a cool season observing system simulation experiment. *Monthly Weather Review*, 141(10):3273–3299.

64. Jones, T. A. and Stensrud, D. J. (2012). Assimilating airs temperature and mixing ratio profiles using an ensemble kalman filter approach for convective-scale forecasts. *Weather and Forecasting*, 27(3):541–564.

65. Kain, J. S. (2004). The kain-fritsch convective parameterization: an update. *Journal of Applied Meteorology*, 43(1):170–181.

66. Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, 82(1):35–45.

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

68. Kalnay, E., Li, H., Miyoshi, T., YANG, S.-C., and BALLABRERA-POY, J. (2007). 4-d-var or ensemble kalman filter? *Tellus A*, 59(5):758–773.

69. Keyser, D. (2013). prepbufr processing at ncep.

70. Kumar, P., Harish Kumar, K., and Pal, P. K. (2013a). Impact of oceansat-2 scatterometer winds and tmi observations on phet cyclone simulation. *Geoscience and Remote Sensing, IEEE Transactions on*, 51(6):3774–3779.

71. Kumar, R., Chakraborty, A., Parekh, A., Sikhakolli, R., Gohil, B. S., and Kiran Kumar, A. (2013b). Evaluation of oceansat-2-derived ocean surface winds using observations from global buoys and other scatterometers. *Geoscience and Remote Sensing, IEEE Transactions on*, 51(5):2571–2576.

72. Langland, R. H., Velden, C., Pauley, P. M., and Berger, H. (2009). Impact of satellite-derived rapid-scan wind observations on numerical model forecasts of hurricane katrina. *Monthly Weather Review*, 137(5):1615–1622.

73. Li, Y., Wang, X., and Xue, M. (2012). Assimilation of radar radial velocity data with the wrf hybrid ensemble–3dvar system for the prediction of hurricane ike (2008). *Monthly Weather Review*, 140(11):3507–3524.

74. Liu, Z., Schwartz, C. S., Snyder, C., and Ha, S.-Y. (2012). Impact of assimilating amsu-a radiances on forecasts of 2008 atlantic tropical cyclones initialized with a limited-area ensemble kalman filter. *Monthly Weather Review*, 140(12):4017–4034.

75. Lynch, P. (2008). The origins of computer weather prediction and climate modeling. *Journal of Computational Physics*, 227(7):3431–3444.

76. Mathew, T., Chakraborty, A., Sarkar, A., and Kumar, R. (2012). Comparison of oceanic winds measured by space-borne scatterometers and altimeters. *Remote Sensing Letters*, 3(8):715–720.

77. McNally, A., Derber, J., Wu, W., and Katz, B. (2000). The use of tovs level-1b radiances in the ncep ssi analysis system. *Quarterly Journal of the Royal Meteorological Society*, 126(563):689–724.

78. Meng, Z. and Zhang, F. (2007). Tests of an ensemble kalman filter for mesoscale and regional-scale data assimilation. part ii: Imperfect model experiments. *Monthly weather review*, 135(4):1403–1423.

79. Meng, Z. and Zhang, F. (2008a). Tests of an ensemble kalman filter for mesoscale and regional-scale data assimilation. part iii: Comparison with 3dvar in a real-data case study. *Monthly Weather Review*, 136(2):522–540.

80. Meng, Z. and Zhang, F. (2008b). Tests of an ensemble kalman filter for mesoscale and regional-scale data assimilation. part iv: Comparison with 3dvar in a month-long experiment. *Monthly Weather Review*, 136(10):3671–3682.

81. Miyoshi, T., Sato, Y., and Kadowaki, T. (2010). Ensemble kalman filter and 4d-var intercomparison with the japanese operational global analysis and prediction system. *Monthly Weather Review*, 138(7):2846–2866.

82. Mizzi, A. P. et al. (2015). Impact of assimilating radiances with the wrfda etkf/3dvar hybrid system on prediction of two typhoons in 2012. *Journal of Meteorological Research*, 29(1):28–40.

83. Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., and Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: Rrtm, a validated correlated-k model for the longwave. *Journal of Geophysical Research: Atmospheres*, 102(D14):16663–16682.

84. NCEP (2008). "national centers for environmental prediction, national weather service, noaa, u.s. department of commerce, ncep adp global upper air and surface weather observations (prepbufr format), may 1997 - continuing".

85. Oke, P. R., Sakov, P., and Corney, S. P. (2007). Impacts of localisation in the enkf and enoi: experiments with a small model. *Ocean Dynamics*, 57(1):32–45.

86. Osuri, K. K., Mohanty, U., Routray, A., and Mohapatra, M. (2012). The impact of satellite-derived wind data assimilation on track, intensity and structure of tropical cyclones over the north indian ocean. *International journal of remote sensing*, 33(5):1627–1652.

87. Pan, H. and Wu, W.-S. (1995). Implementing a mass flux convection parameterization package for the nmc medium-range forecast model. *NMC office note*, 409(40):20–233.

88. Panofsky, R. (1949). Objective weather-map analysis. *Journal of Meteorology*, 6(6):386–392.

89. Parrish, D. F. and Derber, J. C. (1992). The national meteorological center's spectral statistical-interpolation analysis system. *Monthly Weather Review*, 120(8):1747–1763.

90. Pu, Z., Zhang, H., and Anderson, J. (2013). Ensemble kalman filter assimilation of near-surface observations over complex terrain: comparison with 3dvar for short-range forecasts. *Tellus A*, 65.

91. Purser, R. J., Wu, W.-S., Parrish, D. F., and Roberts, N. M. (2003). Numerical aspects of the application of recursive filters to variational statistical analysis. part i: Spatially homogeneous and isotropic gaussian covariances. *Monthly Weather Review*, 131(8):1524–1535.

92. Routray, A., Kar, S., Mali, P., and Sowjanya, K. (2014). Simulation of monsoon depressions using wrf-var: Impact of different background error statistics and lateral boundary conditions. *Monthly Weather Review*, 142(10):3586–3613.

93. Routray, A., Mohanty, U., Osuri, K. K., Kar, S., and Niyogi, D. (2016). Impact of satellite radiance data on simulations of bay of bengal tropical cyclones using the wrf-3dvar modeling system. *IEEE Transactions on Geoscience and Remote Sensing*, 54(4):2285–2303.

94. Sasaki, Y. (1958). A fundamental study of the numerical prediction based on the variational principle. *Journal of Meteorological Society of Japan*, 36(3):77–88.

95. Sathiyamoorthy, V., Sikhakolli, R., Gohil, B., and Pal, P. (2012). Intra-seasonal variability in oceansat-2 scatterometer sea-surface winds over the indian summer monsoon region. *Meteorology and Atmospheric Physics*, 117(3-4):145–152.

96. Saunders, R., English, S., Francis, P., Rayer, P., Brunel, P., Kelly, G., Bauer, P., Salmond, D., and Dent, D. (2005). Rttov-8 the latest update to the rttov model. In *Proc. International TOVS Study Conference XIV, Beijing, China*, pages 25–31.

97. Schwartz, C. S., Liu, Z., Chen, Y., and Huang, X.-Y. (2012). Impact of assimilating microwave radiances with a limited-area ensemble data assimilation system on forecasts of typhoon morakot. *Weather and Forecasting*, 27(2):424–437.

98. Shuman, F. G. (1989). History of numerical weather prediction at the national meteorological center. *Weather and Forecasting*, 4(3):286–296.

99. Shutts, G. (2005). A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society*, 131(612):3079–3102.

100. Sikka, D. (1977). Some aspects of the life history, structure and movement of monsoon depressions. *pure and applied geophysics*, 115(5-6):1501–1529.

101. Singh, R., Kishtawal, C., Ojha, S. P., and Pal, P. (2012a). Impact of assimilation of atmospheric infrared sounder (airs) radiances and retrievals in the wrf 3d-var assimilation system. *Journal of Geophysical Research: Atmospheres*, 117(D11).

102. Singh, R., Kishtawal, C., and Pal, P. (2012b). Impact of atovs radiance on the analysis and forecasts of a mesoscale model over the indian region during the 2008 summer monsoon. *Pure and applied geophysics*, 169(3):425–445.

103. Singh, R., Kishtawal, C., Pal, P., and Joshi, P. (2011). Assimilation of the multisatellite data into the wrf model for track and intensity simulation of the indian ocean tropical cyclones. *Meteorology and atmospheric physics*, 111(3-4):103–119.

104. Singh, R., Kishtawal, C., Pal, P., and Joshi, P. (2012c). Improved tropical cyclone forecasts over north indian ocean with direct assimilation of amsu-a radiances. *Meteorology and Atmospheric Physics*, 115(1-2):15–34.

105. Singh, R., Ojha, S. P., Kishtawal, C., and Pal, P. (2013). Quality assessment and assimilation of megha-tropiques saphir radiances into wrf assimilation system. *Journal of Geophysical Research: Atmospheres*, 118(13):6957–6969.

106. Singh, R., Pal, P., and Joshi, P. (2010). Assimilation of kalpana very high resolution radiometer water vapor channel radiances into a mesoscale model. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 115(D18).

107. Singh, R., Pal, P., Kishtawal, C., and Joshi, P. (2008a). The impact of variational assimilation of ssm/i and quikscat satellite observations on the numerical simulation of indian ocean tropical cyclones. *Weather and Forecasting*, 23(3):460–476.

108. Singh, R., Pal, P., Kishtawal, C., and Joshi, P. (2008b). The impact of variational assimilation of ssm/i and quikscat satellite observations on the numerical simulation of indian ocean tropical cyclones. *Weather and Forecasting*, 23(3):460–476.

109. Sinha, P. and Chandrasekar, A. (2010). Improvement of mesoscale forecasts of monsoon depressions through assimilation of quikscat wind data: two case studies over india. *Open Atmospheric Science Journal*, 4:160–177.

110. Skamarock, W., Klemp, J., Dudhia, J., Gill, D., Barker, D., Duda, M., Huang, X.-y., and Wang, W. (2008). A description of the advanced research wrf version 3. ncar technical note ncar/tn-475+str. Technical report, NCAR.

111. Srinivas, C., Yesubabu, V., Prasad, K. H., Venkatraman, B., and Ramakrishna, S. (2012a). Numerical simulation of cyclonic storms fanoos, nargis with assimilation of conventional and satellite observations using 3-dvar. *Natural hazards*, 63(2):867–889.

112. Srinivas, C., Yesubabu, V., Prasad, K. H., Venkatraman, B., and Ramakrishna, S. (2012b). Numerical simulation of cyclonic storms fanoos, nargis with assimilation of conventional and satellite observations using 3-dvar. *Natural hazards*, 63(2):867–889.

113. Stensrud, D. J., Bao, J.-W., and Warner, T. T. (2000). Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Monthly Weather Review*, 128(7):2077–2107.

114. Tiedtke, M. (1989). A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Monthly Weather Review*, 117(8):1779–1800.

115. Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M., and Whitaker, J. S. (2003). Ensemble square root filters*. *Monthly Weather Review*, 131(7):1485–1490.

116. Torn, R. D. (2010). Performance of a mesoscale ensemble kalman filter (enkf) during the noaa high-resolution hurricane test. *Monthly Weather Review*, 138(12):4375–4392.

117. Torn, R. D. and Hakim, G. J. (2009). Ensemble data assimilation applied to rainex observations of hurricane katrina (2005). *Monthly Weather Review*, 137(9):2817–2829.

118. Torn, R. D., Hakim, G. J., and Snyder, C. (2006). Boundary conditions for limited-area ensemble kalman filters. *Monthly weather review*, 134(9):2490–2502.

119. Wang, X., Barker, D. M., Snyder, C., and Hamill, T. M. (2008a). A hybrid etkf-3dvar data assimilation scheme for the wrf model. part i: Observing system simulation experiment. *Monthly Weather Review*, 136(12):5116–5131.

120. Wang, X., Barker, D. M., Snyder, C., and Hamill, T. M. (2008b). A hybrid etkf-3dvar data assimilation scheme for the wrf model. part ii: Real observation experiments. *Monthly Weather Review*, 136(12):5132–5147.

121. Whitaker, J. S. and Hamill, T. M. (2002). Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, 130(7):1913–1924.

122. Whitaker, J. S., Hamill, T. M., Wei, X., Song, Y., and Toth, Z. (2008). Ensemble data assimilation with the ncep global forecast system. *Monthly Weather Review*, 136(2):463–482.

123. Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences*, volume 100. Academic press.

124. Wu, T.-C., Liu, H., Majumdar, S. J., Velden, C. S., and Anderson, J. L. (2014). Influence of assimilating satellite-derived atmospheric motion vector observations on numerical analyses and forecasts of tropical cyclone track and intensity. *Monthly Weather Review*, 142(1):49–71.

125. Wu, T.-C., Velden, C. S., Majumdar, S. J., Liu, H., and Anderson, J. L. (2015a). Understanding the influence of assimilating subsets of enhanced atmospheric motion vectors on numerical analyses and forecasts of tropical cyclone track and intensity with an ensemble kalman filter. *Monthly Weather Review*, 143(7):2506–2531.

126. Wu, T.-C., Velden, C. S., Majumdar, S. J., Liu, H., and Anderson, J. L. (2015b). Understanding the influence of assimilating subsets of enhanced atmospheric motion vectors on numerical analyses and forecasts of tropical cyclone track and intensity with an ensemble kalman filter. *Monthly Weather Review*, 143(7):2506–2531.

127. Yesubabu, V., Srinivas, C., Hariprasad, K., and Baskaran, R. (2014a). A study on the impact of observation assimilation on the numerical simulation of tropical cyclones jal and thane using 3dvar. *Pure and Applied Geophysics*, 171(8):2023–2042.

128. Yesubabu, V., Srinivas, C., Hariprasad, K., and Baskaran, R. (2014b). A study on the impact of observation assimilation on the numerical simulation of tropical cyclones jal and thane using 3dvar. *Pure and Applied Geophysics*, 171(8):2023–2042.

129. Zhang, F., Weng, Y., Kuo, Y.-H., Whitaker, J. S., and Xie, B. (2010). Predicting typhoon morakot's catastrophic rainfall with a convection-permitting mesoscale ensemble system. *Weather and Forecasting*, 25(6):1816–1825.

130. Zhang, F., Weng, Y., Sippel, J. A., Meng, Z., and Bishop, C. H. (2009). Cloud-resolving hurricane initialization and prediction through assimilation of doppler radar observations with an ensemble kalman filter. *Monthly Weather Review*, 137(7):2105–2125.

131. Zhang, G. J. and McFarlane, N. A. (1995). Sensitivity of climate simulations to the parameterization of cumulus convection in the canadian climate centre general circulation model. *Atmosphere-ocean*, 33(3):407–446.

132. Zhang, H. and Pu, Z. (2014). Influence of assimilating surface observations on numerical prediction of landfalls of hurricane katrina (2005) with an ensemble kalman filter. *Monthly Weather Review*, 142(8):2915–2934.

133. Zhang, M., Zhang, F., Huang, X.-Y., and 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. *Monthly Weather Review*, 139(2):566–572.

APPENDIX A

WRF ARW model and WRF 3DVar system

The WRF model is a fully compressible and nonhydrostatic model (with a run-time hydrostatic option). Its vertical coordinate is a terrain-following hydrostatic pressure coordinate. The grid staggering is the Arakawa C-grid. The model uses the Runge-Kutta 2nd and 3rd order time integration schemes, and 2nd to 6th order advection schemes in both the horizontal and vertical. It uses a time-split small step for acoustic and gravity-wave modes. The dynamics conserves scalar variables.

The WRF model code contains an initialization program, a numerical integration program (wrf.exe), a program to do one-way nesting (ndown.exe), and a program to do tropical storm bogussing (tc.exe). The WRF model supports a variety of capabilities for simulating weather phenomena along various scales. The WRF model is an open-source, community weather model which can be downloaded from http://www2.mmm.ucar.edu/wrf/users/model.html.

The flowchart A.1 illustrates the component programs of the WRF Modeling System. The WRF model can be run with either idealized initialization or real-data initialization. The function of the WRF Preprocessing System (WPS) is to define WRF grid, generate map, elevation and land information for WRF, take real-data analyses/-forecasts from another model, and interpolate the data to the WRF grid. WRF-Var is used to assimilate observations into model initial conditions. The standard output from WPS, real, and WRF model is in netCDF format can be displayed using the graphic tools like NCAR Graphics NCL, GrADS etc.

A.2 shows the major steps that are performed during WRF 3DVar assimilation.

In the 3DVar system, the following steps are performed to obtain the analysis (ref. http://www.mmm.ucar.edu/wrf/users/wrfda/tutorial.html) .

- Reads in the namelist
- Set up the first guess or background field
- Reads the background error statistics and extracts the necessary quantities like

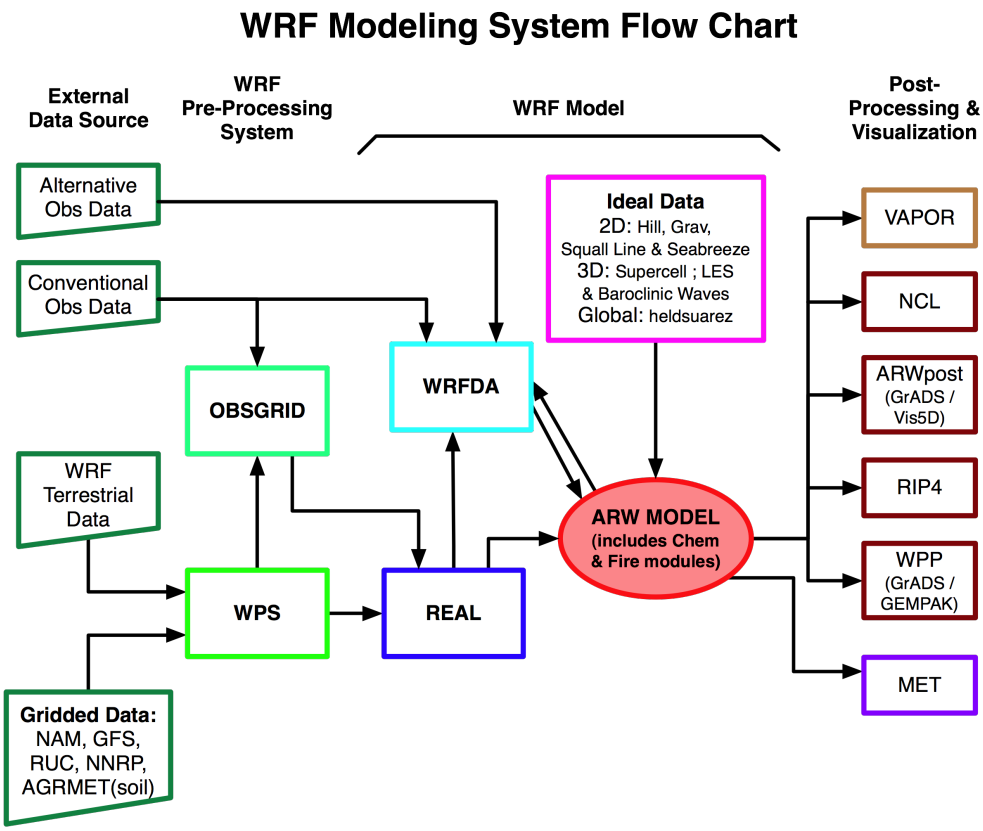

Figure A.1: WRF Modeling System (from WRF ARW userguide)

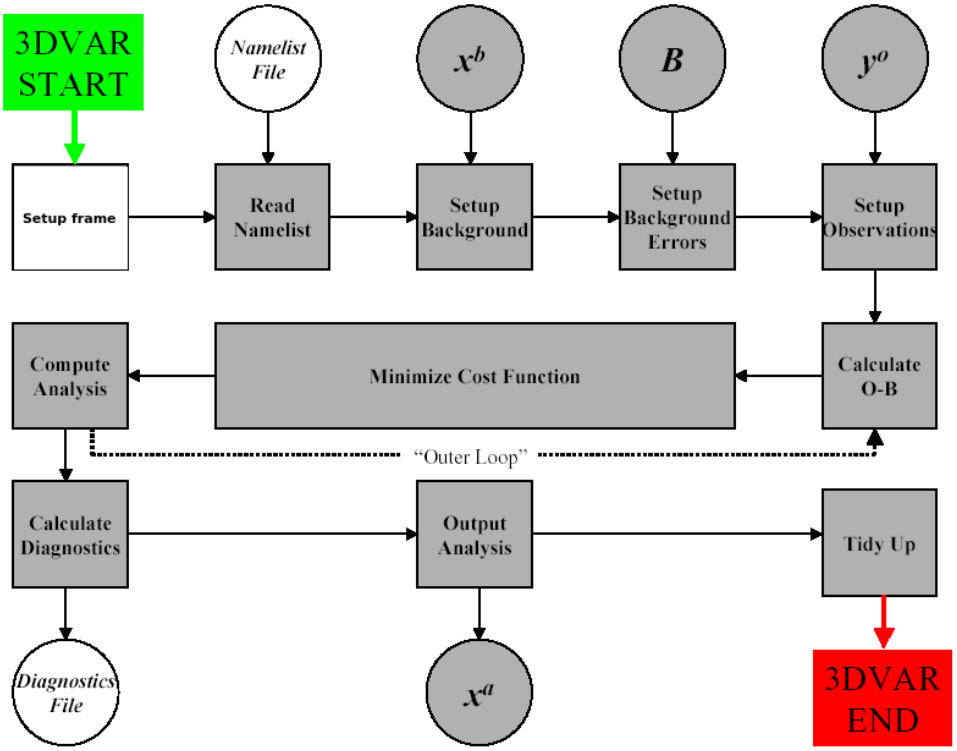

Figure A.2: Main steps during 3DVar assimilation

eigenvectors, eigenvalues, lengthscales, regression coefficients, etc

- Meteorological observations are read
  - Observation input in ASCII format for WRF-Var is supplied through observation preprocessor, OBSPROC. Observation input in BUFR format is read directly and do not pass through OBSPROC
  - The observations outside the model domain are excluded
  - The observations undergo quality control and are thinned
- Calculates "model equivalent" B of observation O utilizing the observation operator
- Computes observation minus first guess (O-B) value
- Minimize the 3DVar cost-function using conjugate gradient method
- The converged control variable is converted to model space analysis increments
- Analysis is computed by adding the analysis increments to first guess fields
- Various diagnostics are computed
- Boundary conditions for the model are updated

A.1 Modifications to the WRF 3DVar code for assimilating MeghaTropiques SAPHIR radiance

The WRF 3DVar system (version 3.6.1) does not support the input of MeghaTropiques SAPHIR radiance out-of-the-box. The WRF 3DVar source code is split into subdirectories containing logically distinct algorithms. Each subdirectory is identified with a particular Fortran90 module file i.e. all the routines within the subdirectory are "Fortran90 INCLUDEd" in a single module file with the same name as the subdirectory. The following modules were added the WRF 3DVar code to read the MeghaTropiques SAPHIR radiance in BUFR format and perform quality control.

1. Code for reading the SAPHIR radiance in BUFR format

```
subroutine da_read_obs_bufrsaphir (obstype,iv, infile)

!
   ------------------------------------------------------------------------


!  Purpose: read in NCEP bufr saphir 1b data to innovation structure
!
!   METHOD: use F90 sequential data structure to avoid reading file
   twice
!            so that da_scan_bufrsaphir is not necessary any more.
!            1. read file radiance data in sequential data structure
!            2. do gross QC check
!            3. assign sequential data structure to innovation
   structure
!               and deallocate sequential data structure
!
   ------------------------------------------------------------------------


implicit none

character(5)       ,  intent (in)    :: obstype
character(20)      ,  intent (in)    :: infile
type (iv_type)     ,  intent (inout) :: iv

#ifdef BUFR

integer           :: iost
integer(i_kind), allocatable :: nread(:)

integer(i_kind),parameter:: n1bhdr=15
integer(i_kind),parameter:: maxinfo=12
integer(i_kind),parameter:: maxchanl=100

logical saphir,hirs2,mhs,msu
logical outside, outside_all, iuse
integer :: inst

character(10) date
```

```
character(8) subset,subfgn
character(80) hdr1b
integer(i_kind) ihh,i,j,k,ifov,idd,ireadmg,ireadsb
integer(i_kind) iret,idate,im,iy,nchan
integer :: num_bufr(7), numbufr, ibufr
character(20) :: filename

! thinning variables
integer(i_kind) itt,itx,iobs,iout
real(r_kind) terrain,timedif,crit,dist
real(r_kind) dlon_earth,dlat_earth

real(r_kind) tbmin,tbmax, tbbad
real(r_kind) panglr,rato
! real(r_kind) rmask
real(r_kind) step,start

real(r_double),dimension(maxinfo+maxchanl):: data1b8
real(r_double),dimension(n1bhdr):: bfr1bhdr

! Instrument triplet, follow the convension of RTTOV
integer   :: platform_id, satellite_id, sensor_id

! pixel information
integer   ::  year,month,day,hour,minute,second  ! observation time
real*8    ::  obs_time
real      ::  rlat, rlon                         ! lat/lon in
   degrees   for Anfovs
real      ::  satzen, satazi, solzen ,solazi      ! scan angles for
    Anfovs
integer   ::  landsea_mask
real      ::  srf_height
! channels' bright temperature
real , allocatable ::   tb_inv(:)                 ! bright
   temperatures
!  end type bright_temperature

type (datalink_type), pointer    :: head, p, current, prev

integer                          ::  ifgat
```

```
type(info_type)                 ::  info
type(model_loc_type)            ::  loc

data hdr1b /'SAID FOVN YEAR MNTH DAYS HOUR MINU SECO CLAT CLON SAZA
   SOZA HOLS LSQL SOLAZI'/
!  data hdr1b /'FOVN YEAR MNTH DAYS HOUR MINU SECO CLAT CLON SAZA
   SOZA HOLS LSQL SLNM BEARAZ'/

data tbmin,tbmax,tbbad / 50.0_r_kind, 550.0_r_kind, -9.99e11_r_kind /
integer :: num_saphir_local, num_saphir_file, num_saphir_global,
   num_saphir_selected
integer :: num_saphir_thinned, num_saphir_used, num_saphir_used_tmp
integer :: lnbufr
integer :: n
integer(i_kind), allocatable :: ptotal(:)
real , allocatable :: in(:), out(:)
logical :: found, head_found

call da_trace_entry("da_read_obs_bufrsaphir")

! Initialize variables

nchan = 20
allocate(nread(1:rtminit_nsensor))
allocate(ptotal(0:num_fgat_time))
nread(1:rtminit_nsensor) = 0
ptotal(0:num_fgat_time) = 0

! Set various variables depending on type of data to be read



step   = 1.80_r_kind
start  = -49.5_r_kind
nchan=6
rato=1.1363987_r_kind

platform_id  = 20
satellite_id = 1
saphir=     obstype == 'saphir'
!   subfgn='NC021043'
```

```
subfgn='NC021057'
sensor_id     =  34


allocate (tb_inv(nchan))

num_saphir_file      = 0    ! number of obs in file
num_saphir_global    = 0    ! number of obs within whole domain
num_saphir_local     = 0    ! number of obs within tile
num_saphir_thinned   = 0    ! number of obs rejected by thinning
num_saphir_used      = 0    ! number of obs entered into innovation
   computation
num_saphir_selected = 0    ! number of obs limited for debuging
iobs = 0                 ! for thinning, argument is inout

! 0.0  Open unit to satellite bufr file and read file header
!---------------------------------------------------------------------

num_bufr(:)=0
numbufr=0
if (num_fgat_time>1) then
do i=1,7
call da_get_unit(lnbufr)
write(filename,fmt='(A,2I1,A)') trim(infile),0,i,'.bufr'
open(unit   = lnbufr, FILE   = trim(filename),iostat =  iost, form =
   'unformatted', STATUS = 'OLD')
if (iost == 0) then
numbufr=numbufr+1
num_bufr(numbufr)=i
else
close (lnbufr)
end if
call da_free_unit(lnbufr)
end do
else
numbufr=1
end if

if (numbufr==0) numbufr=1
```

```
bufrfile:  do ibufr=1,numbufr
if (num_fgat_time==1) then
filename=trim(infile)//'.bufr'
else
if ((numbufr ==1) .and. (num_bufr(ibufr) == 0)) then
filename=trim(infile)//'.bufr'
else
write(filename,fmt='(A,2I1,A)') trim(infile),0,num_bufr(ibufr),'.bufr
    ,
end if
end if

! We want to use specific unit number for bufr data, so we can
    control the endian format in environment.
lnbufr = 99

open(unit=lnbufr,file=trim(filename),form='unformatted', &
iostat = iost, status = 'old')
if (iost /= 0) then
call da_warning(__FILE__,__LINE__, &
(/"Cannot open file "//infile/))
call da_trace_exit("da_read_obs_bufrsaphir")
return
end if

call openbf(lnbufr,'IN',lnbufr)
call datelen(10)
call readmg(lnbufr,subset,idate,iret)
if (subset /= subfgn) then
call closbf(lnbufr)
close(lnbufr)
message(1)='The file title does not match the data subset'
write(unit=message(2),fmt=*) &
'infile=', lnbufr, infile,' subset=', subset, ' subfgn=',subfgn
call da_error(__FILE__,__LINE__,message(1:2))
end if

iy=0
im=0
idd=0
```

```
ihh=0
write(unit=date,fmt='(i10)') idate
read(unit=date,fmt='(i4,3i2)') iy,im,idd,ihh
write(unit=stdout,fmt=*) &
'Bufr file date is ',iy,im,idd,ihh,infile

! Loop to read bufr file and assign information
to a sequential structure
!
```

```
if ( ibufr == 1 ) then
allocate (head)
!  allocate ( head % tb_inv (1:nchan) )
nullify  ( head % next )
p => head
endif

!   if (saphir_start > 1) then
!      write (unit=stdout,fmt='(A,I6)') "   Skipping saphir obs
   before", saphir_start
!   end if

obs: do while (ireadmg(lnbufr,subset,idate)==0 .and. subset==subfgn)
do while (ireadsb(lnbufr)==0)

! 1.0      Read header record and data record

call ufbint(lnbufr,bfr1bhdr,n1bhdr,1,iret,hdr1b)
call ufbrep(lnbufr,data1b8,1,nchan,iret,'TMBR')

! check if observation outside range

num_saphir_file = num_saphir_file + 1

! 2.0      Extract observation location and other required information
!     QC1:  judge if data is in the domain, read next record if not
!
```

```
rlat = bfr1bhdr(bufr_lat)
rlon = bfr1bhdr(bufr_lat)
if (rlon < 0.0) rlon = rlon+360.0

info%lat  =  bfr1bhdr(bufr_lat)

info%lon  =  bfr1bhdr(bufr_lon)
call da_llxy (info, loc, outside, outside_all)

ifov = nint(bfr1bhdr(bufr_ifov))

! QC2: limb pixel rejected (not implemented)

! 3.2     Extract date information.

year   = bfr1bhdr(bufr_year)
month  = bfr1bhdr(bufr_month)
day    = bfr1bhdr(bufr_day)
hour   = bfr1bhdr(bufr_hour)
minute = bfr1bhdr(bufr_minute)
second = bfr1bhdr(bufr_second)

write(unit=info%date_char, fmt='(i4,a,i2.2,a,i2.2,a,i2.2,a,i2.2,a,i2
   .2)')  &
year, '-', month, '-', day, '_', hour, ':', minute, ':', second

! QC3: time consistency check with the background date

if (year <= 99) then
if (year < 78) then
year = year + 2000
else
year = year + 1900
end if
end if

call da_get_julian_time(year,month,day,hour,minute,obs_time)

!c today
```

```
!if (obs_time < time_slots(0) .or. &
!c today
!            obs_time >= time_slots(num_fgat_time)) cycle

! 3.2.1 determine FGAT index ifgat

do ifgat=1,num_fgat_time
if (obs_time >= time_slots(ifgat-1) .and. &
obs_time  < time_slots(ifgat)) exit
end do

! 3.3 Find wrfvar instrument index from RTTOV instrument triplet
!     go to next data if id is not in the lists

inst = 0
do i = 1, rtminit_nsensor
if (platform_id  == rtminit_platform(i) &
.and. satellite_id == rtminit_satid(i)    &
.and. sensor_id    == rtminit_sensor(i)) then
inst = i
exit
end if
end do
if (inst == 0) cycle

! 3.4 extract satellite and solar angle

panglr=(start+float(ifov-1)*step)*deg2rad
if (hirs2 .or. msu) then
satzen = asin(rato*sin(panglr))*rad2deg
satzen = abs(satzen)
else
satzen = bfr1bhdr(bufr_satzen) !*deg2rad   ! local zenith angle
satzen = abs(satzen)

solzen = bfr1bhdr(bufr_solzen)              ! solar zenith angle
solazi = bfr1bhdr(bufr_solazi)              !RTTOV9_3

num_saphir_global = num_saphir_global + 1
ptotal(ifgat) = ptotal(ifgat) + 1
```

```
num_saphir_selected = num_saphir_selected + 1

if (outside) cycle ! No good for this PE
num_saphir_local = num_saphir_local + 1

!  Make Thinning
!  Map obs to thinning grid
!--------------------------------------------------------------------
if (thinning) then
dlat_earth = info%lat
dlon_earth = info%lon
if (dlon_earth<zero) dlon_earth = dlon_earth+r360
if (dlon_earth>=r360) dlon_earth = dlon_earth-r360
dlat_earth = dlat_earth*deg2rad
dlon_earth = dlon_earth*deg2rad
timedif = 0.0 !2.0_r_kind*abs(tdiff)          ! range:  0 to 6
terrain = 0.01_r_kind*abs(bfr1bhdr(13))
crit = 1.0 !0.01_r_kind+terrain + timedif !+ 10.0_r_kind*float(iskip)
call map2grids(inst,ifgat,dlat_earth,dlon_earth,crit,iobs,itx,1,itt,
   iout,iuse)
if (.not. iuse) then
num_saphir_thinned=num_saphir_thinned+1
cycle
end if
end if

num_saphir_used = num_saphir_used + 1
nread(inst) = nread(inst) + 1

! 3.5 extract surface information
srf_height = bfr1bhdr(bufr_station_height)          ! station height
if (srf_height < 8888.0 .AND. srf_height > -416.0) then
else
srf_height = 0.0
endif

landsea_mask = nint(bfr1bhdr(bufr_landsea_mask))  ! 0:land ; 1:sea (
   same as RTTOV)
info%elv = srf_height
```

```
! 3.6 extract channel bright temperature

tb_inv(1:nchan) = data1b8(1:nchan)
do k = 1, nchan
if ( tb_inv(k) < tbmin .or. tb_inv(k) > tbmax) &
tb_inv(k) = missing_r
end do
if ( all(tb_inv<0.0) ) then
num_saphir_local = num_saphir_local -1
num_saphir_used = num_saphir_used - 1
nread(inst) = nread(inst) - 1
cycle
end if

! 4.0   assign information to sequential radiance structure
!
   ------------------------------------------------------------------------

allocate (p % tb_inv (1:nchan))
p%info              = info
p%loc               = loc
p%landsea_mask      = landsea_mask
p%scanpos           = ifov
p%satzen            = satzen
p%satazi            = satazi
p%solzen            = solzen
p%tb_inv(1:nchan)   = tb_inv(1:nchan)
p%sensor_index      = inst
p%ifgat             = ifgat
!RTTOV9_3
p%solazi            = solazi
!end of RTTOV9_3
allocate (p%next)   ! add next data

p => p%next
nullify (p%next)
end do
end do obs
```

```
call closbf(lnbufr)
close(lnbufr)
end do bufrfile

if (thinning .and. num_saphir_global > 0 ) then

#ifdef DM_PARALLEL

! Get minimum crit and associated processor index.
j = 0
do ifgat = 1, num_fgat_time
do n = 1, iv%num_inst
j = j + thinning_grid(n,ifgat)%itxmax
end do
end do

allocate ( in  (j) )
allocate ( out (j) )

j = 0
do ifgat = 1, num_fgat_time
do n = 1, iv%num_inst
do i = 1, thinning_grid(n,ifgat)%itxmax
j = j + 1
in(j) = thinning_grid(n,ifgat)%score_crit(i)
end do
end do
end do
call mpi_reduce(in, out, j, true_mpi_real, mpi_min, root, comm, ierr)

call wrf_dm_bcast_real (out, j)

j = 0
do ifgat = 1, num_fgat_time
do n = 1, iv%num_inst
do i = 1, thinning_grid(n,ifgat)%itxmax
j = j + 1
if ( ABS(out(j)-thinning_grid(n,ifgat)%score_crit(i)) > 1.0E-10 )
   thinning_grid(n,ifgat)%ibest_obs(i)  = 0
end do
```

```
end do
end do

deallocate( in  )
deallocate( out )

#endif

! Delete the nodes which being thinning out
p => head
prev => head
head_found = .false.
num_saphir_used_tmp = num_saphir_used
do j = 1, num_saphir_used_tmp
n = p%sensor_index
ifgat = p%ifgat
found = .false.

do i = 1, thinning_grid(n,ifgat)%itxmax
if ( thinning_grid(n,ifgat)%ibest_obs(i) == j .and. thinning_grid(n,
    ifgat)%score_crit(i) < 9.99e6_r_kind ) then
found = .true.
exit
endif
end do

! free current data
if ( .not. found ) then
current => p
p => p%next
if ( head_found ) then
prev%next => p
else
head => p
prev => p
endif
deallocate ( current % tb_inv )
deallocate ( current )
num_saphir_thinned = num_saphir_thinned + 1
num_saphir_used = num_saphir_used - 1
```

```
nread(n) = nread(n) - 1
continue
endif

if ( found .and. head_found ) then
prev => p
p => p%next
continue
endif

if ( found .and. .not. head_found ) then
head_found = .true.
head => p
prev => p
p => p%next
endif

end do

endif  ! End of thinning

iv%total_rad_pixel   = iv%total_rad_pixel   + num_saphir_used
iv%total_rad_channel = iv%total_rad_channel + num_saphir_used*nchan

iv%info(radiance)%nlocal = iv%info(radiance)%nlocal + num_saphir_used
iv%info(radiance)%ntotal = iv%info(radiance)%ntotal +
   num_saphir_global

do i = 1, num_fgat_time
ptotal(i) = ptotal(i) + ptotal(i-1)
iv%info(radiance)%ptotal(i) = iv%info(radiance)%ptotal(i) + ptotal(i)
end do
if ( iv%info(radiance)%ptotal(num_fgat_time) /= iv%info(radiance)%
   ntotal ) then
write(unit=message(1),fmt='(A,I10,A,I10)') &
"Number of ntotal:",iv%info(radiance)%ntotal," is different from the
   sum of ptotal:", iv%info(radiance)%ptotal(num_fgat_time)
call da_warning(__FILE__,__LINE__,message(1:1))
endif
```

```
write(unit=stdout,fmt='(a)') 'num_saphir_file num_saphir_global
    num_saphir_local num_saphir_used num_saphir_thinned'
write(unit=stdout,fmt='(5i10)') num_saphir_file,num_saphir_global,
    num_saphir_local,num_saphir_used,num_saphir_thinned

deallocate(tb_inv)

!  5.0 allocate innovation radiance structure
!---------------------------------------------------------------------

do i = 1, iv%num_inst
if (nread(i) < 1) cycle
iv%instid(i)%num_rad = nread(i)
iv%instid(i)%info%nlocal = nread(i)
write(UNIT=stdout,FMT='(a,i3,2x,a,3x,i10)') &
'Allocating space for radiance innov structure', &
i, iv%instid(i)%rttovid_string, iv%instid(i)%num_rad

call da_allocate_rad_iv(i,nchan,iv)

end do

!  6.0 assign sequential structure to innovation structure
!---------------------------------------------------------------------
nread(1:rtminit_nsensor) = 0
p => head
! do while ( associated(p) )

do n = 1, num_saphir_used
i = p%sensor_index
nread(i) = nread(i) + 1

call da_initialize_rad_iv (i, nread(i), iv, p)

current => p
p => p%next

! free current data
deallocate ( current % tb_inv )
deallocate ( current )
```

```
end do

deallocate ( p )

deallocate (nread)
deallocate (ptotal)

call da_trace_exit("da_read_obs_bufrsaphir")
#else
call da_error(__FILE__,__LINE__,(/"Needs to be compiled with a BUFR
   library"/))
#endif

end subroutine da_read_obs_bufrsaphir
```

2. Code for performing quality control for SAPHIR radiances

```
subroutine da_qc_saphir (it, i, nchan, ob, iv)

! Purpose: perform quality control for saphir data.

implicit none
integer, intent(in)             :: it         ! outer loop count
integer, intent(in)             :: i          ! sensor index.
integer, intent(in)             :: nchan      ! number of channel
type (y_type), intent(in)       :: ob         ! Observation structure
   .
type (iv_type), intent(inout)   :: iv         ! O-B structure.

! local variables

integer    :: n,scanpos,k,isflg,ios,fgat_rad_unit
real       :: si,si1,si2,si3,si4,si5
logical    :: lmix
integer    :: ngood(nchan),nrej(nchan),nrej_omb_abs(nchan), &
nrej_omb_std(nchan),       &
nrej_mixsurface,nrej_windowchanl, nrej_si,     &
nrej_clw,nrej_topo, num_proc_domain,  &
nrej_limb

character(len=30)  :: filename
```

```
if (trace_use) call da_trace_entry("da_qc_saphir")
ngood(:)        = 0
nrej(:)         = 0
nrej_omb_abs(:) = 0
nrej_omb_std(:) = 0
nrej_mixsurface = 0
nrej_windowchanl= 0
nrej_si         = 0
nrej_clw        = 0
nrej_topo       = 0
nrej_limb       = 0
num_proc_domain = 0

do n= iv%instid(i)%info%n1,iv%instid(i)%info%n2
if (iv%instid(i)%info%proc_domain(1,n)) &
num_proc_domain = num_proc_domain + 1

! 0.0  initialise QC flags by assuming good obs
iv%instid(i)%tb_qc(:,n) = qc_good
if (crtm_cloud) go to 2508
! a.  reject all channels over mixture surface type
isflg = iv%instid(i)%isflg(n)
lmix  = (isflg==4) .or. (isflg==5) .or. (isflg==6) .or. (isflg==7)
if (lmix) then
iv%instid(i)%tb_qc(:,n)  =  qc_bad
if (iv%instid(i)%info%proc_domain(1,n)) &
nrej_mixsurface = nrej_mixsurface + 1
end if
!
  ------------------------------------------------------------------------------------------



!   Following is added by SAC/ISRO Ahmedabad
!  b.  reject channels 6 over land/sea-ice/snow
if (isflg > 0) then
iv%instid(i)%tb_qc(6,n)    = qc_bad
if (iv%instid(i)%info%proc_domain(1,n)) &
nrej_windowchanl = nrej_windowchanl + 1
if (only_sea_rad) iv%instid(i)%tb_qc(:,n)    = qc_bad
end if
```

```
! d. check cloud/precipitation, By SAC/ISRO Ahmedabad
if (ob%instid(i)%tb(2,n) > 0.0 .and. &
ob%instid(i)%tb(5,n) > 0.0) then
si1 = ob%instid(i)%tb(2,n) - ob%instid(i)%tb(3,n)
si2 = ob%instid(i)%tb(2,n) - ob%instid(i)%tb(5,n)
si3 = ob%instid(i)%tb(3,n) - ob%instid(i)%tb(5,n)
si4 = ob%instid(i)%tb(2,n) - ob%instid(i)%tb(1,n)
si5 = ob%instid(i)%tb(6,n) - ob%instid(i)%tb(4,n)
if (si1 .gt. 0 .or. si2 .gt. 0 .or. si3 .gt. 0 .or. si5 .le. 4)then
iv%instid(i)%tb_qc(:,n) = qc_bad
iv%instid(i)%cloud_flag(:,n) = qc_bad
if (iv%instid(i)%info%proc_domain(1,n)) &
nrej_si = nrej_si + 1
end if
end if
!-----------------------------------------------------------------

if (iv%instid(i)%clwp(n) >= 0.2) then
iv%instid(i)%tb_qc(:,n) = qc_bad
iv%instid(i)%cloud_flag(:,n) = qc_bad
if (iv%instid(i)%info%proc_domain(1,n)) &
nrej_clw = nrej_clw + 1
end if

! g. check iuse (pre-rejected channels by .info files)
!-----------------------------------------------------------
do k = 1, nchan
if (satinfo(i)%iuse(k) .eq. -1) &
iv%instid(i)%tb_qc(k,n) = qc_bad
end do

2508      continue
! f. check innovation
!-----------------------------------------------------------------
do k = 1, nchan
! absolute departure check
if (.not. crtm_cloud) then
if (abs(iv%instid(i)%tb_inv(k,n)) > 15.0) then
iv%instid(i)%tb_qc(k,n)  = qc_bad
```

```
if (iv%instid(i)%info%proc_domain(1,n)) &
nrej_omb_abs(k) = nrej_omb_abs(k) + 1
end if
end if
! relative departure check
if (use_error_factor_rad) then
iv%instid(i)%tb_error(k,n) =  &
satinfo(i)%error_std(k)*satinfo(i)%error_factor(k)
else
iv%instid(i)%tb_error(k,n) = satinfo(i)%error_std(k)
end if
if (.not. crtm_cloud) then
if (abs(iv%instid(i)%tb_inv(k,n)) > 3.0*iv%instid(i)%tb_error(k,n))
   then
iv%instid(i)%tb_qc(k,n)  = qc_bad
if (iv%instid(i)%info%proc_domain(1,n)) &
nrej_omb_std(k) = nrej_omb_std(k) + 1
end if
! final QC decision
if (iv%instid(i)%tb_qc(k,n) == qc_bad) then
iv%instid(i)%tb_error(k,n) = 500.0
if (iv%instid(i)%info%proc_domain(1,n)) &
nrej(k) = nrej(k) + 1
else
if (iv%instid(i)%info%proc_domain(1,n)) &
ngood(k) = ngood(k) + 1
end if
end if
end do ! chan
end do ! end loop pixel
! Do inter-processor communication to gather statistics.
call da_proc_sum_int (num_proc_domain)
call da_proc_sum_int (nrej_mixsurface)
call da_proc_sum_int (nrej_windowchanl)
call da_proc_sum_int (nrej_si)
call da_proc_sum_int (nrej_clw)
call da_proc_sum_int (nrej_topo)
call da_proc_sum_int (nrej_limb)
call da_proc_sum_ints (nrej_omb_abs(:))
call da_proc_sum_ints (nrej_omb_std(:))
```

```
call da_proc_sum_ints (nrej(:))
call da_proc_sum_ints (ngood(:))
if (rootproc) then
if (num_fgat_time > 1) then
write(filename,'(i2.2,a,i2.2)') it, '_qcstat_'//trim(iv%instid(i)%
   rttovid_string)//'_',iv%time
else
write(filename,'(i2.2,a)') it, '_qcstat_'//trim(iv%instid(i)%
   rttovid_string)
end if
call da_get_unit(fgat_rad_unit)
open(fgat_rad_unit,file=trim(filename),form='formatted',iostat=ios)
if (ios /= 0) then
write(unit=message(1),fmt='(A,A)') 'error opening the output file ',
   filename
call da_error(__FILE__,__LINE__,message(1:1))
end if
write(fgat_rad_unit, fmt='(/a/)') &

'Quality Control Statistics for '//iv%instid(i)%rttovid_string
write(fgat_rad_unit,'(a20,i7)') ' num_proc_domain  = ',
   num_proc_domain
write(fgat_rad_unit,'(a20,i7)') ' nrej_mixsurface  = ',
   nrej_mixsurface
write(fgat_rad_unit,'(a20,i7)') ' nrej_windowchanl = ',
   nrej_windowchanl
write(fgat_rad_unit,'(a20,i7)') ' nrej_si          = ', nrej_si
write(fgat_rad_unit,'(a20,i7)') ' nrej_clw         = ', nrej_clw
write(fgat_rad_unit,'(a20,i7)') ' nrej_topo        = ', nrej_topo
write(fgat_rad_unit,'(a20,i7)') ' nrej_limb        = ', nrej_limb
write(fgat_rad_unit,'(a20)')    ' nrej_omb_abs(:)  = '
write(fgat_rad_unit,'(10i7)')     nrej_omb_abs(:)
write(fgat_rad_unit,'(a20)')    ' nrej_omb_std(:)  = '
write(fgat_rad_unit,'(10i7)')     nrej_omb_std(:)
write(fgat_rad_unit,'(a20)')    ' nrej(:)          = '
write(fgat_rad_unit,'(10i7)')     nrej(:)
write(fgat_rad_unit,'(a20)')    ' ngood(:)         = '
write(fgat_rad_unit,'(10i7)')     ngood(:)
close(fgat_rad_unit)
call da_free_unit(fgat_rad_unit)
```

```
end if
if (trace_use) call da_trace_exit("da_qc_saphir")
end subroutine da_qc_saphir
```

APPENDIX B

DART EnKF system

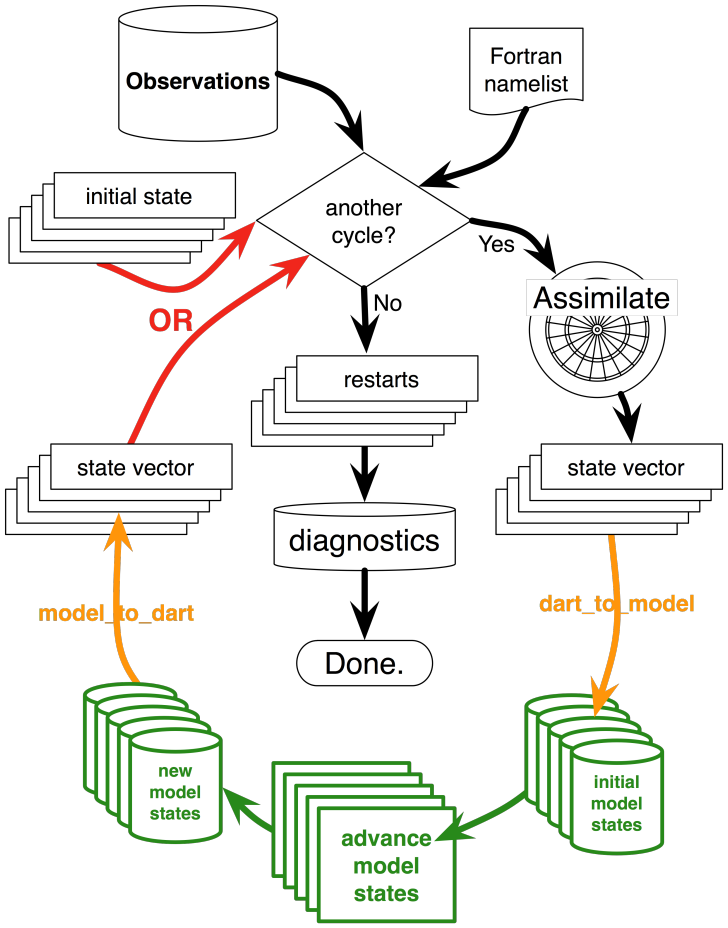

Figure B.1: Schematic of ensemble data assimilation

Fig. B.1 describes the schematic of ensemble data assimilation in the Data Assimilation Research Testbed (DART). A Fortran executable named 'filter' reads a namelist, an initial state for the ensemble, and a file containing observations and goes to work. Given the observations and an initial state, 'filter' assimilates the observations and then determines how far to advance the model (using information from the namelist and the observation file). 'filter' forks a shell script to the system and it is this shell script that is responsible for three things: 1) for converting the DART state vectors and 'advance to time' to the format required by the underlying model, 2) advancing the model, and 3) converting the model output into a form suitable for 'filter'. The model advances each ensemble member and the model output is converted to the input format expected by 'filter'. The shell script finishes and signals 'filter' to continue. We are now back

at the beginning and the cycle continues as long as there are observations to assimilate or until the control information in the Fortran namelist is met. When that happens, a set of restart files is written (suitable to continue an experiment with more observations) and diagnostic files are written. These diagnostic files allow for the exploration of the assimilation before and after each assimilation step and for exploration of the assimilation in 'observation space'; each real observation is paired with the estimates of the observation from all of the ensemble members. Minimally, the ensemble mean estimate of the observation and the ensemble spread of the estimates is recorded.

DART employs a modular programming approach to apply an Ensemble Kalman Filter which nudges the underlying models toward a state that is more consistent with information from a set of observations. Models may be swapped in and out, as can different algorithms in the Ensemble Kalman Filter. The method requires running multiple instances of a model to generate an ensemble of states. A forward operator appropriate for the type of observation being assimilated is applied to each of the states to generate the model's estimate of the observation. DART source code, which is an open-source community software avaiable from http://www.image.ucar.edu/DAReS/DART, has been used to perform EnKF assimilation in the present study.

LIST OF PAPERS BASED ON THESIS

Papers in Refereed International Journals

(i) Dhanya M. and A. Chandrasekar (2016). Impact of variational assimilation using multivariate background error covariances on the simulation of monsoon depressions over India., In *Annales Geophysicae*, 34(2): 187-201.

(ii) Dhanya M. and A.Chandrasekar (2014) Improved Rainfall Simulation by Assimilating Oceansat-2 Surface Winds Using Ensemble Kalman Filter for a Heavy Rainfall Event over South India. *Geoscience and Remote Sensing, IEEE Transactions on*, 52(12): 7721–7726.

(iii) Dhanya M. , Deepak Gopalakrishnan , A.Chandrasekar, Sanjeev Kumar Singh, V.S. Prasad. Impact of assimilating MeghaTropiques SAPHIR radiances in the simulation of tropical cyclones over Bay of Bengal using WRF model. Accepted in the *International Journal of Remote Sensing*.

Presentations in Conferences

(i) Dhanya M. , Deepak Gopalakrishnan , A.Chandrasekar, Sanjeev Kumar Singh, V.S. Prasad. Impact of Megha-Tropiques SAPHIR radiance assimilation on the simulation of tropical cyclones over Bay of Bengal. Accepted in the *10th SPIE Asia-Pacific Remote Sensing Symposium*, NewDelhi, India, April 4-7 2016.

(ii) Dhanya M. and A. Chandrasekar (2016). Impact of using multivariate background error covariance in the three-dimensional variational technique in simulating monsoon depressions over India, In *National Space Science Symposium*, Thiruvavanthapuram, Kerala, India, February 9-12, 2016.

(iii) Dhanya M. and A. Chandrasekar (2012). The impact of variational assimilation of Oceansat-2 wind observations on the numerical simulation of heavy rainfall

events during northeast Indian monsoon , In *39th COSPAR Scientific Assembly 2012*, Mysore, Karnataka, India, 14-22 July 2012.

(iv) Dhanya M. and A. Chandrasekar (2012) Impact of assimilating Oceansat2 and Quikscat wind observations on the simulation of weather systems during north east monsoon over India , In *IITM Golden jubilee International Conference on Opportunities and Challenges in Monsoon prediction in Changing Climate (OCHAMP)*, Pune, India, 21-25 February, 2012.