TESTS OF AN ENSEMBLE KALMAN FILTER FOR MESOSCALE AND REGIONAL-SCALE DATA ASSIMILATION

A Dissertation

by

ZHIYONG MENG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2007

Major Subject: Atmospheric Sciences

TESTS OF AN ENSEMBLE KALMAN FILTER FOR MESOSCALE AND REGIONAL-SCALE DATA ASSIMILATION

A Dissertation

by

ZHIYONG MENG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee, Fuqing Zhang
Committee Members, John Nielsen-Gammon
Courtney Shumacher
Robert Hetland
Christopher A. Davis
Head of Department, Richard E. Orville

May 2007

Major Subject: Atmospheric Sciences

ABSTRACT

Tests of an Ensemble Kalman Filter for Mesoscale and Regional-Scale Data Assimilation. (May 2007)

Zhiyong Meng, B.S., Beijing University;

M.S., Chinese Academy of Meteorological Sciences

Chair of Advisory Committee: Dr. Fuqing Zhang

This dissertation examines the performance of an ensemble Kalman filter (EnKF) implemented in a mesoscale model in increasingly realistic contexts from under a perfect model assumption and in the presence of significant model error with synthetic observations to real-world data assimilation in comparison to the three-dimensional variational (3DVar) method via both case study and month-long experiments. The EnKF is shown to be promising for future application in operational data assimilation practice.

The EnKF with synthetic observations, which is implemented in the mesoscale model MM5, is very effective in keeping the analysis close to the truth under the perfect model assumption. The EnKF is most effective in reducing larger-scale errors but less effective in reducing errors at smaller, marginally resolvable scales. In the presence of significant model errors from physical parameterization schemes, the EnKF performs reasonably well though sometimes it can be significantly degraded compared to its performance under the perfect model assumption. Using a combination of different physical parameterization schemes in the ensemble (the so-called “multi-scheme”

ensemble) can significantly improve filter performance due to the resulting better background error covariance and a smaller ensemble bias. The EnKF performs differently for different flow regimes possibly due to scale- and flow-dependent error growth dynamics and predictability.

Real-data (including soundings, profilers and surface observations) are assimilated by directly comparing the EnKF and 3DVar and both are implemented in the Weather Research and Forecasting model. A case study and month-long experiments show that the EnKF is efficient in tracking observations in terms of both prior forecast and posterior analysis. The EnKF performs consistently better than 3DVar for the time period of interest due to the benefit of the EnKF from both using ensemble mean for state estimation and using a flow-dependent background error covariance. Proper covariance inflation and using a multi-scheme ensemble can significantly improve the EnKF performance. Using a multi-scheme ensemble results in larger improvement in thermodynamic variables than in other variables. The 3DVar system can benefit substantially from using a short-term ensemble mean for state estimate. Noticeable improvement is also achieved in 3DVar by including some flow dependence in its background error covariance.

DEDICATION

To my beloved husband Shuanzhu

and

to our lovely daughter Ruoqi

ACKNOWLEDGMENTS

This dissertation is the result of three and half years of work whereby I have been supported by many people, without whom I would never have been able to finish it.

First of all, I would like to extend my deepest gratitude to the chair of my advisory committee, Dr. Fuqing Zhang, for his enthusiastic support during the whole process of my doctoral study and his inspiring and encouraging way to guide me toward a deeper understanding of not only knowledge, but also how to be a serious scientist. His expertise, enthusiasm and integral view on research and his mission for providing 'only high-quality work and not less' have made a deep impression on me. I would also like to thank Dr. John Nielsen-Gammon, Dr. Courtney Shumacher, Dr. Robert Hetland, and Dr. Christopher A. Davis for serving on my advisory committee and giving me valuable comments on my research work.

My gratitude also goes to other collegues and staffs in Department of Atmospheric Sciences of Texas A&M University (TAMU) who have in their own way added to my graduate education and life through classroom and personal interaction. I am also very grateful to the Mesoscale and Microscale Meteorological Division of the National Center for Atmopsheric Reseach (NCAR) for providing me student visitorship and support in June 2004, during which I learned lots of knowledge helpful to this research. Special thanks go to Dr. Chris Snyder of NCAR for his valuable comments on this work, Ryan Torn (University of Washington), Yongrun Guo (NCAR), and Yonghui Weng (China Meteorological Administration) for their technical help, Jason Sippel (TAMU) and Dr.

Altug Aksoy (NCAR) for their kind and valuable comments on both scientific and paper writing aspects.

Last but not least, I am greatly indebted to my husband, my daughter and my family for their understanding, patience and support during the entire period of my study.

This research is sponsored by the NSF grant ATM0205599 and by the Office of Navy Research under grant N000140410471.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF FIGURES

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

1. Motivation and objectives

The ensemble Kalman filter (EnKF) data assimilation technique, which estimates background error covariance through an ensemble of short-term forecasts, has been recently applied to various dynamical systems and realistic situations. Since being introduced by Geir Evensen in 1994, the EnKF has drawn increasing attention from the data assimilation community and has become one of the leading candidates for future operational data assimilation techniques. While the application of the EnKF has become quasi-operational for large-scale global models, its use for mesoscale systems has only started rather recently. The EnKF has the potential for success in the mesoscale data assimilation field because of its use of flow-dependent background error covariances. Variational data assimilation, an alternative to the EnKF, often has difficulties in capturing detailed structures of mesoscale systems due to its use of isotropic and homogeneous background error covariance. The performance of EnKF implemented in large-scale models has been shown better or at least comparable to the performance of variational technique with real observations (Houtekamer et al. 2005; Whitaker et al. 2006). However, the performance of the EnKF implemented in mesoscale models has not been compared directly to that of variational method.

This dissertation follows the style of *Monthly Weather Review.*

This doctoral study seeks to examine the performance of an EnKF for different mesoscale weather systems using perfect model and imperfect-model Observation System Simulation Experiments (OSSE) and real data assimilation. The EnKF is directly compared to the three dimensional variational method (3DVar) with in-situ observations. This work has two main objectives:

(1) Evaluate the performance of the mesoscale ensemble-based data assimilation system with synthetic observations for both perfect and imperfect models (i.e., the ensemble members and the truth are propagated with the same or a different model) and ensemble scenarios, at different scales and under different flow regimes, and introduce a possible model error treatment method.

(2) Assimilate in-situ observations from the Bow Echo and Mesoscale Convective Vortex Experiment (BAMEX); compare the performance of the EnKF to 3DVar.

2. Background

One goal of atmospheric science is to make weather or climate forecasts as accurate as possible. To this end meteorologists have developed different kinds of forecast methods, such as persistence, trends, climatology, analogy and numerical weather prediction (NWP). Numerical prediction has become increasingly reliable with the improvement of computer capabilities, model accuracy, observation availability and data assimilation techniques during the past 50 years. However, numerical weather prediction can never be error free because of two main reasons. One is model uncertainty due to the discretization, parameterization of sub-grid physical processes, or boundary conditions.

The other is from the initial condition error due to insufficient observations, measurement error and representativeness error. The accuracy of numerical weather prediction is very sensitive to initial fields (Thompson 1957); the atmosphere is chaotic, and initial small scale errors can evolve into large scale errors (Lorenz 1963). An effective way to improve forecast accuracy is to decrease the initial error; this is the goal of data assimilation. In fact, several major historical improvements in numerical weather prediction skill are the result of better data assimilation.

a. Data assimilation

Data assimilation can be described as the process through which all the available information (including observations and a numerical model) and their associated uncertainties are used in order to estimate as accurately as possible the state of the atmospheric flow and its associated uncertainties (Talagrand 1997). During the past 15 years, the data assimilation field has witnessed an explosion of studies, resulting in rather different empirical and statistical methods. Successive corrections and nudging are two empirical methods that can be useful for small-scale observations when there are no available statistics. Other techniques like optimum interpolation (OI), Kalman filter (KF), ensemble Kalman filter (EnKF), 3DVar, and 4DVar are based on statistical estimation theory. OI, KF, EnKF and 3DVar are carried out sequentially; namely whenever a model reaches an instant when an observation is available, the model forecast fields are updated by the observation. After that, the updated fields are propagated forward in time by the model until the next observation. In these approaches,

only past and present observations are assimilated. A different method, 4DVar, minimizes a penalty function that measures the distance between the observations and the model forecast over a time interval. The model is run forward while first calculating the mismatches between the forecast and all measurements in the interval. An adjoint model is then run backward, combining all the mismatches in the time interval to update the initial field. This procedure is then repeated to minimize the penalty function. 3DVar is a simplified form without the time dimension and with all observations valid at the same time. 3DVar and 4DVar have advantages such as the ease of assimilating indirect observations (not model variables) and assimilating data globally. However, they are flawed by ignoring model error, using an isotropic and static background error covariance, and not being able to provide the uncertainty of the analysis. In contrast, these problems can be easily addressed in the Kalman filter.

b. The Kalman filter

The Kalman filter is a linear, recursive estimator that produces the minimum variance estimate in a least squares sense under the assumption of white and Gaussian noise processes. The history of the Kalman filter can be traced back to the 17th century. In 1795, 19-year-old Karl F. Gauss invented the least squares method for astronomical estimation (Sorenson 1970). 150 years later, in 1941, Kolmogorov (1962) and Wiener (1949) independently developed linear minimum mean-square estimation, which is actually a probabilistic version of the least-squares method. Ten years later, J.W. Follin (1955) suggested a recursive approach for the estimation field. Based on the above

works, in 1960, Rudolph E. Kalman published a discrete-time recursive mean-square filter and predicted its great potential application in various fields.

The Kalman filter consists of two steps: the forecast step and the analysis step.

In the forecast step, the state and the associated uncertainty are propagated forward in time through

$$\mathbf{x}_t^b = M\mathbf{x}_{t-1}^a, \tag{1.1}$$

$$\mathbf{P}_t^b = \mathbf{M}\mathbf{P}_{t-1}^a\mathbf{M}^T + \mathbf{Q}, \tag{1.2}$$

in which M is the linear dynamic model, $\mathbf{M}$ is the tangent linear dynamic model, $\mathbf{M}^T$ is the adjoint of the tangent linear model, $\mathbf{Q}$ is model error covariance matrix, $\mathbf{x}_{t-1}^a$ and $\mathbf{P}_{t-1}^a$ are the analysis state vector and its error covariance at time t-1, and $\mathbf{x}_t^b$ and $\mathbf{P}_t^b$ are the background state vector and its error covariance matrix at time t whenever observations are available.

The analysis step is performed at time t by

$$\mathbf{x}_t^a = \mathbf{x}_t^b + \mathbf{K}(\mathbf{y}^o - H\mathbf{x}_t^b), \tag{1.3}$$

$$\begin{aligned}\mathbf{P}_t^a &= (\mathbf{I}-\mathbf{K}\mathbf{H})\mathbf{P}_t^b(\mathbf{I}-\mathbf{K}\mathbf{H})^T + \mathbf{K}\mathbf{R}\mathbf{K}^T \\ &= (\mathbf{I}-\mathbf{K}\mathbf{H})\mathbf{P}_t^b,\end{aligned} \tag{1.4}$$

$$\mathbf{K} = \mathbf{P}_t^b\mathbf{H}^T(\mathbf{H}\mathbf{P}_t^b\mathbf{H}^T + \mathbf{R})^{-1}, \tag{1.5}$$

where $\mathbf{x}_t^a$ is the updated state vector obtained through adding a weighted difference between observations $\mathbf{y}^0$ and the background to the background, $\mathbf{P}_t^a$ is the analysis error covariance matrix, $\mathbf{K}$ is the Kalman gain, which indicates how much to weight observations relative to the background and how to spread their impact to other points

and other variables to get a minimum variance analysis (detailed derivation is provided in Appendix), H is the observation operator mapping from model space to observation space, $\mathbf{H}$ is the tangent linear of the observation operator, and $\mathbf{R}$ is the observation error covariance.

c. The ensemble Kalman filter

Though the Kalman filter can provide flow-dependent background error covariance, it is very difficult to apply to atmospheric data assimilation due to the large dimension (10^6-10^8) and nonlinearity of atmospheric models. The Kalman filter also requires tangent linear and adjoint models, which makes its implementation even harder. To circumvent these problems, Evensen invented the ensemble Kalman filter in 1994 by approximating $\mathbf{P}^b$ with a short term ensemble forecast of size N through

$$\mathbf{P}^b = \overline{(\mathbf{x}^b - \overline{\mathbf{x}^b})(\mathbf{x}^b - \overline{\mathbf{x}^b})^T}, \tag{1.6}$$

in which the overbar indicates an ensemble mean.

The EnKF is equivalent to the Kalman filter for linear error growth and infinite ensemble size. There are several appealing features of the EnKF: (1.1) The background error covariance is flow-dependent, which reflects the error of the day; (1.2) The model and observation operator can be nonlinear; (1.3) It provides not only the best estimation, but also the associated uncertainty; (1.4) There is no need to code a tangent linear or adjoint model; (1.5) The ensemble members can be run simultaneously, and it is thus easy to parallelize. Evensen's pioneering work together with his further studies (Evensen and van Leeuwen 1996; Evensen 1997) demonstrates that the EnKF can adequately track

the true state and shows great potential in larger primitive-equation models.

Since its first introduction in 1994, the ensemble Kalman filter has drawn increasing attention from the data assimilation community and has become one of two leading candidates (the other is 4DVar) to be employed in future operational data assimilation practice (Evensen 2003; Lorenc 2003; Hamill 2005).

1) ISSUES INVOLVING THE ENKF

While the EnKF has those appealing features as described above, it has its own issues. First, the EnKF tends to systematically underestimate analysis error. The EnKF uses the ensemble mean and covariance as the best estimation of the state and its associated uncertainty. The ensemble mean is updated by

$$\overline{\mathbf{x}^{\mathrm{a}}} = \overline{\mathbf{x}^{\mathrm{b}}} + \mathbf{K}(\mathbf{y}^{\mathrm{o}} - H\overline{\mathbf{x}^{\mathrm{b}}}), \tag{1.7}$$

The covariance is updated through updating the perturbation of individual members from the ensemble mean by

$$\begin{aligned} \mathbf{x}_i'^{\mathrm{a}} &= \mathbf{x}_i'^{\mathrm{b}} + \mathbf{K}(\mathbf{y}'^{\mathrm{o}} - \mathbf{H}\mathbf{x}_i'^{\mathrm{b}}) \\ &= \mathbf{x}_i'^{\mathrm{b}} - \mathbf{K}\mathbf{H}\mathbf{x}_i'^{\mathrm{b}}, \end{aligned} \tag{1.8}$$

and the analysis covariance $\mathbf{P}^{\mathrm{a}}$ is obtained by

$$\begin{aligned} \mathbf{P}^{\mathrm{a}} &= \overline{x'^{a}(x'^{a})^{T}} \\ &= (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^{\mathrm{b}}(\mathbf{I} - \mathbf{K}\mathbf{H})^{\mathrm{T}} \end{aligned} \tag{1.9}$$

Comparing (1.9) with standard Kalman filter analysis covariance (1.4), it is easy to see that the second term $\mathbf{K}\mathbf{R}\mathbf{K}^{\mathrm{T}}$ in (1.4) is missing. The absence of this positive term will cause the EnKF systematically underestimate the analysis uncertainty, thus giving more

weight to the background and resulting in filter divergence when the analysis deviates from the truth.

Two main approaches have been introduced to address the covariance underestimation problem. One is to treat the observation as a random variable by adding random noise consistent with observational error to the observation (Houtekamer and Mitchell 1998; Burgers et al. 1998; Anderson 2001). However, Whitaker and Hamill (2002) noted that the addition of noise to the observations can reduce the accuracy of the analysis error covariance estimate due to an increase of the sampling error. In the meantime, the probability of underestimating the analysis-error covariance is also increased because of the appearance of forecast-observation covariance due to observation perturbation. In this regard, Pham (2001) put forward a method of forcing the background-observation error covariance to zero while perturbing the observation. However, the computation cost is very large when applying this method to high-degree-of-freedom systems. Whitaker and Hamill (2002) introduced the ensemble square-root filter (EnSRF) as a better way to deal with the underestimation of analysis uncertainty. Instead of perturbing observations, they choose to use a different Kalman gain $\mathbf{K}'$ to update the perturbation of individual members

$$\mathbf{x}_i'^{a} = \mathbf{x}_i'^{b} - \mathbf{K}'\mathbf{H}\mathbf{x}_i'^{b}. \tag{1.10}$$

The value of $\mathbf{K}'$ is chosen to be consistent with the analysis covariance of the standard Kalman filter by letting

$$(\mathbf{I} - \mathbf{K}'\mathbf{H})\mathbf{P}^{b}(\mathbf{I} - \mathbf{K}'\mathbf{H})^{T} = (\mathbf{I} - \mathbf{K}'\mathbf{H})\mathbf{P}^{b} \tag{1.11}$$

If $\mathbf{K}'$ is proportional to $\mathbf{K}$ by $\mathbf{K}' = \alpha\mathbf{K}$, then $\mathbf{K}'$ can be substituted in (1.11) to give

$$\frac{\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}}{\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}+\mathbf{R}}\alpha^2-2\alpha+1=0. \tag{1.12}$$

The solution of this equation for α is

$$\alpha=\left(1+\sqrt{\frac{\mathbf{R}}{\mathbf{H}\mathbf{P}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}+\mathbf{R}}}\right)^{-1}. \tag{1.13}$$

It was shown that EnSRF performs better than perturbing the observations (Whitaker and Hamill 2002).

Another issue still under active research is related to determination of optimal ensemble size. Theoretically, the ensemble size should be O(300) to get a mean with a 95% chance of lying in the interval of 0.25 standard deviation from the truth (Leeuwen 1999). However, in practice, 100 members are sufficient to get reasonable results (Evensen 1994; Houtekamer and Mitchell 1998). The ensemble size can be even smaller for a smaller scale model. For example, Snyder and Zhang (2003) used only 40 members for a cloud model application.

A finite ensemble size can result in systematic underestimation of analysis error covariance; this is related to the so-called “in-breeding” problem (Houtekamer and Mitchell 1998; van Leeuwen 1999). The reason for this underestimation is that the analysis error covariance is nonlinearly dependent on background error covariance. As a result, the decrease of analysis variance associated with large background variance is larger than that associated with smaller background variance (Whitaker and Hamill 2002; van Leeuwen 1999).

To address this issue, several remedies have been introduced. Houtekamer and

Mitchell (1998) run a pair of ensembles in parallel with one ensemble being updated with the statistics calculated from the other. Meanwhile, Anderson and Anderson (1999) put forward an inflation method by globally multiplying the covariance by a value slightly larger than one. Studies show that more inflation is necessary for small ensemble sizes than for larger ones (Hamill et al. 2001). The inflation method is problematic in data sparse areas because the covariance continuously increases and ultimately results in analysis failure. Zhang et al. (2004) put forward a covariance relaxation method in which the analysis deviation from the mean is replaced by a weighted average between the analysis perturbation and prior forecast perturbation. This approach performs better than the standard inflation method (Anderson and Anderson 1999) because the covariance is inflated only where updating takes place.

Houtekamer and Mitchell (1998) noted that a limited ensemble size can result in spurious correlation between distant points. In another study (Houtekamer and Mitchell 2001), they experimented with a covariance localization method using a "Schur product". Using this method, they multiplied the ensemble-based covariance estimates element by element with a distance-dependent correlation function that varied from one at the observation point to zero at specific radius of influence. This covariance localization significantly improved the analysis result.

Also in 2001, Hamill et al. provided a statistically based rationale for the covariance localization method and examined the contribution of covariance localization in cases of different data resolution and different ensemble size. Larger benefit was seen when data was sparse and the ensemble size was small. When covariance localization was applied

to a small-sized ensemble, analysis error comparable to that obtained with a large ensemble but without localization could be attained.

Another benefit of covariance localization is that it can somewhat alleviate the rank deficiency problem associated with the EnKF. This problem arises because the calculation of matrix $\mathbf{HP}^{b}\mathbf{H}^{T} + \mathbf{R}$ with an ensemble is singular and cannot be inversed if the ensemble size is smaller than observation number. Localization can alleviate the rank deficiency issue by decreasing the number of observations to be assimilated.

Since the corrections to the background field are created totally via statistically estimated covariance from an ensemble forecast without a geostrophic balance constraint, imbalance could become an issue. Besides, imbalance could also be caused by covariance localization.

2) APPLICATION IN LARGE SCALE SYSTEMS

While many recent studies have shown the EnKF to outperform 3DVar when using the perfect model assumption (Hamill and Snyder 2000; Whitaker and Hamill 2002; Anderson 2001), this assumption must be dropped in the real world where model errors are unavoidable. As mentioned before, model error is typically caused by inadequate parameterization of sub-grid physical process, numerical inaccuracy, truncation error or other random errors. The presence of model error can often result in large bias of the ensemble mean and too little ensemble spread, which can ultimately cause the ensemble forecast to fail. Studies show that including model error can lead to more realistic spread of the forecast solution (Houtekamer et al. 1996; Houtekamer and Lefaivre 1997).

However, model errors, especially those at the mesoscale, are generally hard to identify and portray due to the chaotic nature and flow-dependent characteristics of the atmosphere and the lack of observations for verification (e.g., Orrell et al. 2001; Orrell 2002; Simmons and Hollingsworth 2002; Stensrud et al. 2000).

There have been several different approaches to include model errors in ensemble forecasting. One is to use different forecast models (e.g., Evans et al. 2000; Krishnamurti et al. 2000) or different physical parameterization schemes (e.g., Stensrud et al. 2000). This method is relatively simple to implement though sometimes *ad hoc* in nature. Another way is to apply statistical adjustment to ensemble forecasts (Hamill and Whitaker 2005). The use of stochastic forecast models and/or stochastic physical parameterizations has also been examined (e.g., Palmer 2001; Grell and Denvenyi 2002).

Since the EnKF uses ensemble forecasts to estimate background error covariance, model error is unavoidable in assimilating real time observations. Many recent studies discuss the explicit treatment of model error in the context of ensemble-based data assimilation (Mitchell and Houtekamer 2002; Hansen 2002; Keppenne and Rienecker 2003; Hamill and Whitaker 2005; Houtekamer et al. 2005). The most common of these treatments is the covariance inflation method proposed in Anderson and Anderson (1999). This method parameterizes model error by expanding the background forecast members' deviation about the ensemble mean. Keppenne and Rienecker (2003) obtained encouraging results using an oceanic general circulation model and real data through covariance inflation. The inflation method is also used in Whitaker et al. (2004) for the reanalysis of the past atmospheric state using a long series of available surface pressure

observations; that study demonstrated that the EnKF performs better than the 3DVar method. Most recently, the additive error method has also been proposed as alternative to covariance inflation (Hamill and Whitaker 2005; Houtekamer et al. 2005). Hamill and Whitaker (2005) systematically examined the performance of the perturbation inflation and additive error methods in the treatment of model truncation error due to lack of interaction with smaller scale motions. Their results demonstrate that the performance of certain additive error methods is more satisfactory and might outperform the simulated 3DVar method (which was implemented similarly to the EnKF except that the covariance was stationary). In the meantime, Houtekamer et al. (2005) investigated the potential quality of the EnKF in a realistic environment by assimilating a complete set of real observations into a medium-resolution primitive-equation model that includes a complete set of physical parameterizations. Similar to Hamill and Whitaker (2005), they parameterized the model error by adding noise consistent in structure with a 3DVar background error covariance model. It was shown that the EnKF performed similarly to the 3DVar method implemented in the same forecast system.

3) APPLICATION IN SMALL SCALE SYSTEMS

In contrast to the extensive studies in the context of global or large scale systems, as reviewed above, applications of the EnKF to smaller scales have only recently begun. Preliminary results show that a cloud-model-based EnKF with the ensemble size of 50 can converge the analysis toward the true state through assimilating synthetic radar radial velocity (Snyder and Zhang 2003; Zhang et al. 2004) and outperform 4DVar in

later cycles (Caya et al. 2005). The efficiency of the EnKF at cloud scale is further confirmed with a more complex compressible model (Tong and Xue 2005) through assimilating both simulated radial velocity and reflectivity. With the same EnKF system as Snyder and Zhang (2003), Dowell et al. (2004) assimilated real radar radial velocity and reflectivity for a supercell thunderstorm and accurately determined the locations of the main updraft and mesocyclone. They also noted that the EnKF did little to correct the errors in the cold pools; this is possibly due to errors in the moist convective parameterization of the model or lack of thermal observations.

4) APPLICATIONS IN MESOSCALE SYSTEMS

Compared to the large-scale applications, there are fewer mesoscale applications of the EnKF in literature. One probable reason is the simultaneous and complicated scale interaction in mesoscale models. To investigate the possibility of applying the EnKF to a mesoscale system, Zhang (2005) examined the dynamics and structure of mesoscale error covariance based on a 20-member short term ensemble forecast with the Pennsylvania State University-National Center for Atmospheric Research (PSU/NCAR) non-hydrostatic mesoscale model MM5. It was found that the initially largely uncorrelated, mostly random errors evolve into strong, coherent structures over regions of active moist convection. The error covariance is highly anisotropic, flow-dependent and ultimately determined by the underlying governing dynamics and the associated error growth. This result suggests that the spatial and cross covariance estimated from the short-term ensemble forecast has the potential to spread observational information

non-uniformly to both observed and unobserved variables.

Several other studies have shown promising results with the EnKF. For example, an EnKF with an ensemble size of 50 using a sea-breeze model (Aksoy et al. 2005) was shown to be effective under the perfect model assumption. Lately, the next generation Weather Research and Forecasting (WRF) model has shown promise for mesoscale prediction, and a WRF-based EnKF with 90 members and simulated observations indicates that the flow-dependent filter corrections are accurate in both scale and amplitude (Dirren et al. 2006). In another study, simulated surface data is successfully assimilated with an EnKF using a planetary boundary layer (PBL) model (Hacker and Snyder 2005). The significant improvement in temperature, wind and moisture field in the PBL suggests the potential for surface observation assimilation to improve numerical simulations and forecasts of convective development and cyclogenesis where PBL preconditioning is important. In the same way, Fujita et al (2005) assimilated real hourly surface observations into a MM5-based EnKF with the ensemble size of 25. They found that when both initial and physical uncertainties are taken into account, the EnKF can considerably improve the lower tropospheric structure and the placement and intensity of convection.

3. Methodology

The EnKF employed in this work is an ensemble square root filter following the algorithm of Snyder and Zhang (2003), in which the observations are ingested sequentially instead of assimilating the observation vector as a whole in one step. The

state vector updated with one observation acts as the first guess of the next observation under the assumption that observation errors are independent among one another. This reduces the Kalman gain matrix to a column vector, thus largely simplifying the calculation. The following is a detailed description of this method

For a single scalar observation y^o and ensemble size N, the calculation of Kalman gain $\mathbf{K}$ (1.5) is significantly simplified. $\mathbf{P}^b\mathbf{H}^T$ in (1.5) becomes a column vector of dimension N_x (the dimension of state vector) denoted by $\mathbf{c}$

$$\begin{aligned}\mathbf{c} &= \mathbf{P}^b\mathbf{H}^T \\ &= \frac{1}{\text{N-1}}\sum_{n=1}^{N}(\mathbf{x}_n^b - \overline{\mathbf{x}^b})(\mathbf{H}\mathbf{x}_n^b - \overline{\mathbf{H}\mathbf{x}^b})^T\end{aligned} \tag{1.14}$$

$\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R}$ in (1.5) reduced to a scalar, which can be denoted by d as

$$\begin{aligned}d &= \mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R} \\ &= \frac{1}{\text{N-1}}\sum_{n=1}^{N}(\mathbf{H}\mathbf{x}_n^b - \overline{\mathbf{H}\mathbf{x}^b})(\mathbf{H}\mathbf{x}_n^b - \overline{\mathbf{H}\mathbf{x}^b})^T + \varepsilon^2\end{aligned} \tag{1.15}$$

Then the ensemble mean is updated by

$$\overline{\mathbf{x}^a} = \overline{\mathbf{x}^b} + \mathbf{c}(y^o - H\overline{\mathbf{x}^b})/d \tag{1.16}$$

The perturbation of the *n*th member from the mean is updated by

$$\mathbf{x}_n^a - \overline{\mathbf{x}^a} = \left[\mathbf{1} - \alpha(\mathbf{c}/d)\mathbf{H}\right](\mathbf{x}_n^b - \overline{\mathbf{x}^b}) \tag{1.17}$$

where $\alpha = \left[1 + \sqrt{\varepsilon^2/d}\right]^{-1}$, which is the coefficient to get the new Kalman gain to update the covariance in the EnSRF. By repeating the analysis step for each datum, the updated ensemble will be obtained finally.

As described in section 1, this work is planned to achieve two goals. One is to implement and evaluate a mesoscale, ensemble based data assimilation technique using synthetic surface and sounding observations typical of conventional temporal and spatial distribution in different scenarios, scales, and flow regimes. The other is to assimilate real-data observation to compare the EnKF to 3DVar.

To achieve the first goal, extensive experiments are carried out with a MM5-based EnKF under the perfect model assumption for the 'surprise' snowstorm of 24-26 January 2000 in Chapter II. Different values for ensemble size, cut-off radius of influence and covariance relaxation are used in various experiments. In Chapter III, the EnKF is evaluated in the presence of significant model error from parameterization of subscale physical processes and with different ensemble generation methods for both the snowstorm case and the MCV event. The flow and scale dependence of the EnKF is examined, and possible ways to treat model error are put forward.

To achieve the second goal, the same EnKF is implemented in WRF to assimilate BAMEX field observations including radiosondes, surface observations, wind profilers for the MCV event in Chapter IV. The performance of the EnKF is compared to WRF-3DVar. Following the MCV case study, Chapter V presents result of a month-long comparison between the performance of the EnKF and 3DVar to make the conclusion more general through assimilating 12 hourly sounding observations. Finally, a brief summary and discussion are given in Chapter VI.

Chapters II-V are written in the form of papers for submission to various professional journals. They are interconnected and follow a smooth flow though each of

the papers stands independently on its own merits.

CHAPTER II

OBSERVING SYSTEM SIMULATION EXPERIMENTS WITH PERFECT MODEL*

1. Introduction

The ensemble-based data assimilation method (ensemble Kalman filter or EnKF; Evensen 1994), which uses short-term ensemble forecasts to estimate the flow-dependent background error covariance, has recently been implemented in various atmospheric and oceanic models. These models vary from idealized examples based on simplified equation sets to those based on the complete, primitive equations with assimilation of real observations (Houtekamer and Mitchell 1998, 2001; Hamill and Snyder 2000; Keppenne 2000; Anderson 2001; Mitchell et al. 2002; Keppenne and Rienecker 2002; Whitaker and Hamill 2002; Zhang and Anderson 2003; Snyder and Zhang 2003; Houtekamer et al. 2004; Whitaker et al. 2004; Dowell et al. 2004; Zhang et al. 2004). These experimental studies demonstrated the feasibility and effectiveness of the EnKF for different scales and flows of interest and the advantages of using the EnKF over existing data assimilation schemes, which assume stationary, isotropic background error covariance. This present study seeks to exploit the potential of using the EnKF to

* Reprinted with permission from "Tests of an Ensemble Kalman Filter for Mesoscale and Regional-Scale Data Assimilation. Part I: Perfect Model Experiments" by Fuqing Zhang, Zhiyong Meng, and Altug Aksoy, 2006, *Monthly Weather Review*, 134, 722-736, Copyright [2006] by American Meteorological Society.

assimilate simulated sounding and surface observations for mesoscale numerical weather prediction systems, which often include dynamics and interactions among convective, meso- and sub-synoptic scales.

Recently, short-term ensemble forecasts generated with different sets of initial perturbations were used to examine the dynamics and structure of mesoscale error covariance of the 24-25 January 2000 surprise snowstorm (Zhang 2005). In the ensemble forecast initiated with rescaled random perturbations, initial errors grow from smaller-scale, largely unbalanced and uncorrelated perturbations to larger-scale, quasi-balanced disturbances within 12-24 h. Comparable ensemble spread is found in ensemble forecasts initialized with balanced random perturbations or with grid-point random perturbations. In all ensemble forecasts, the error growth is maximized in the vicinity of the strongest mean PV gradient and over the area of active moist convection, consistent with the lower predictability in these regions (Zhang et al. 2002; 2003). Consequently, the initially largely uncorrelated, mostly random errors evolve into strong coherent structures with spatial correlation not only within individual variables (auto-covariance) but also between different forecast variables (cross covariance), especially over the region of strong cyclogenesis and along the upper-level front. The error covariance is highly anisotropic. Dramatic differences in magnitude, structure, and sign are found between covariance estimated from the same set of ensemble forecasts but verified at different times. The structure of the mesoscale error covariance is ultimately determined by the underlying governing dynamics and the associated error growth.

The spatial and cross covariance estimated from the short-term ensemble forecast

has the potential to spread observational information non-uniformly to both observed and unobserved variables at different vertical layers (e.g., from the upper troposphere to the surface and vice versa) with a horizontal radius of influence potentially greater than 1000 km. The flow-dependent nature of the error growth dynamics and the covariance structure further indicates potential benefit of using anisotropic and flow-dependent representations of background error covariance for mesoscale and regional-scale data assimilation.

The current chapter seeks to examine the significance and the effectiveness of the error covariance estimated from the short-term ensemble forecasts for mesoscale and regional-scale data assimilation for the same event as in Zhang (2005). Section 2 introduces the forecast model and the formulation and configuration of the EnKF. The truth simulation and the reference forecast ensemble are presented in section 3. Performance of the control EnKF experiment is examined in section 4. Forecast error growth from ensembles with and without the EnKF is discussed in section 5. The sensitivity experiments to EnKF configuration, data coverage, frequency and uncertainty of observations are presented in section 6. Summary and conclusions are presented in section 7. The impacts of model error and ensemble initiation on the filter performance will be explored in Chapter III.

2. Forecast model and EnKF

The study uses the Pennsylvania State University-National Center for Atmospheric Research (NCAR) non-hydrostatic mesoscale model MM5 (Dudhia 1993). The model

domain has 190x120 horizontal grid points with 30-km grid spacing and covers the continental United States (Fig. 2.1). There are 27 layers in the terrain-following vertical coordinate with model top at 100 hPa and vertical spacing smallest within the boundary layer. The model has a total of 10 prognostic variables including three Cartesian velocity components (u, v, w), pressure perturbation (p'), temperature (T), and mixing ratios for water vapor (q), cloud water (q_c), rainwater (q_r), cloud ice (q_i) and graupel (q_g). The physical parameterization schemes used in this study are the Grell cumulus parameterization scheme (Grell 1993), the Mellor-Yamada Planetary Boundary Layer (PBL) scheme (Mellor and Yamada 1982) and the Reisener microphysics scheme with graupel (Reisener et al. 1998). The state dimension of the forecast model is ~10^7. Observations are taken only from the shaded area in Fig. 2.1 and only state vectors in this inner box are updated and analyzed.

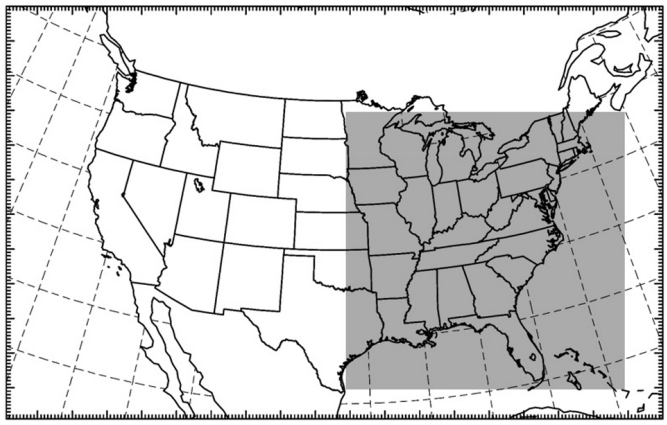

FIG. 2.1. Map of model domain. Only data in shaded areas are assimilated and analyzed.

The EnKF was first proposed for geophysical applications by Evensen (1994). The

implementation of the EnKF used in the current study follows closely that of Snyder and Zhang (2003). As in standard Kalman filter,

$$\mathbf{x^a = x^f + K\,(y - Hx^f)}\ , \tag{2.1}$$

where $\mathbf{x^f}$ represents the prior estimate or first guess, $\mathbf{x^a}$ is the posterior estimate or analysis, $\mathbf{y}$ is the observation vector, $\mathbf{H}$ is the observation operator that returns observed variables given the state, and $\mathbf{K}$ is the so-called Kalman gain matrix defined as:

$$\mathbf{K = P^f H^T\,(HP^f H^T + R)^{-1}}\ , \tag{2.2}$$

where $\mathbf{P^f}$ and $\mathbf{R}$ represent the background and observational error covariance, respectively. In the EnKF, the flow-dependent $\mathbf{P^f}$ is estimated through an ensemble of short-range forecasts. Observations are taken sequentially with uncorrelated observations errors. Further background on the EnKF can be found in Snyder and Zhang (2003) and references therein.

One modification to Snyder and Zhang (2003) is to adopt the covariance relaxation method proposed in Zhang et al. (2004) in which a modified analysis deviation $(\mathbf{x}^a_{new})'$ is computed by "relaxing" or weighting $(\mathbf{x}^f)'$ and $(\mathbf{x}^a)'$:

$$(\mathbf{x}^a_{new})' = (1 - \alpha)\,(\mathbf{x}^a)' + \alpha\,(\mathbf{x}^f)'\ , \tag{2.3}$$

where deviations from the mean are denoted by primes, and α =0.5 is used in this study. The modified analysis deviations are then used as initial conditions for the ensemble forecasts to the next assimilation time. Since the analysis (posterior) deviation $(\mathbf{x}^a)'$ is smaller than the forecast (prior) deviation $(\mathbf{x}^f)'$, reflecting the reduction of uncertainty after assimilating observations, the use of (2.3) will overestimate (inflate) the uncertainty

in the analysis, as an alternative to the covariance inflation used by Anderson (2001). It is worth noting that, both the covariance relaxation using Eq. (2.3) and the covariance inflation of Anderson (2001) are ad-hoc ways of dealing with the tendency of the spread of a small ensemble to underestimate the true error of the ensemble mean. Another alternative is to use an EnKF configuration with a pair of ensembles (Houtekamer and Mitchell 1998; Houtekamer et al. 2005), which does not require any 'adjustable' inflation or relaxation parameters but at the cost of essentially doubling the ensemble size.

In addition, a covariance localization method using the Gaspari and Cohn (1999) compactly supported fifth-order correlation function is performed in the full three-dimensional physical space. The covariance is set to be zero if the total grid-point distance [$r = (r_x + r_y + r_\sigma)^{0.5}$, where r_x, r_y and r_σ are distance in number of grid points in the x, y and σ (or z) directions, respectively] is greater than 30, equivalent to a horizontal distance of 900 km.

3. The truth simulation and the reference forecast ensemble

The truth simulation and the reference forecast ensemble are produced by randomly perturbing the reference analysis at 00 UTC 24 January 2000. The reference analysis is generated using the NCAR-National Center for Environmental Prediction (NCEP) reanalysis. The perturbations used are randomly sampled from the background error covariance of the MM5 three-dimensional variational (3DVar) data assimilation system (Barker et al. 2004). A set of random control vectors (including streamfuncton, velocity potential, unbalanced pressure, and relative humidity) with a normal distribution (zero

mean and unit standard deviation) is created and transformed back to model space via an empirical orthogonal function (EOF), a recursive filter and geostrophic balance constraint (Barker et al. 2003; their page 58-59). The domain-averaged standard deviations (STD) of the derived perturbations are approximately 1 ms^{-1} for wind components u and v, 0.5 K for temperature T, 0.4 hPa for pressure perturbation p' and 0.2 g/Kg for specific humidity q. Other prognostic variables (vertical wind w, mixing ratios of cloud water q_c, rain water q_r, snow q_s and graupel q_g) are not perturbed in the MM5 3DVar system used here. These perturbations are then added to the reference analysis at 0000 UTC 24 January 2000 to generate a 40-member reference forecast ensemble that is integrated for 36 h with boundary conditions provided by the NCAR-NCEP reanalysis updated every 12 h. The use of the 3DVar background error covariance to generate the initial ensemble for the EnKF can also be found in Houtekamer et al. (2004).

The truth simulation is generated in the same manner (i.e., the same model and the same initial uncertainties) as one of the members of the reference forecast ensemble but with a different realization of random perturbations. The truth simulation is used to generate observations and is also used as the reference to evaluate the performance of the EnKF. Only state vectors and observations in the shaded region of Fig. 2.1, an area of 2400 by 2400 km, are analyzed and assimilated. Selection of the shaded areas (instead of the total model domain) as the analysis domain is to minimize the impact of using the same boundary conditions for the integration of both the truth and the reference forecast ensemble. Over the 36-h integration, state variables inside the shaded domain have little

influence from the model lateral boundary conditions.

Figure 2.2 shows the mean sea level pressure (MSLP) and model-derived reflectivity at the 12-, 24-, and 36-h forecast times from the truth simulation (upper panels) and the reference forecast ensemble mean (lower panels). Corresponding geopotential heights, potential vorticity (PV) and vector winds at 300 hPa are displayed in Fig. 2.3. The truth simulation is chosen from 50 different random realizations to compare most favorably to observations of this event in terms of the location and strength of the surface cyclone (Figs. 3a,b of Zhang et al. 2002, hereafter also ZSR02) and 300-hPa short-wave trough (Fig. 2 of ZSR02) and the onshore precipitation band (Figs. 3c,d of ZSR02).

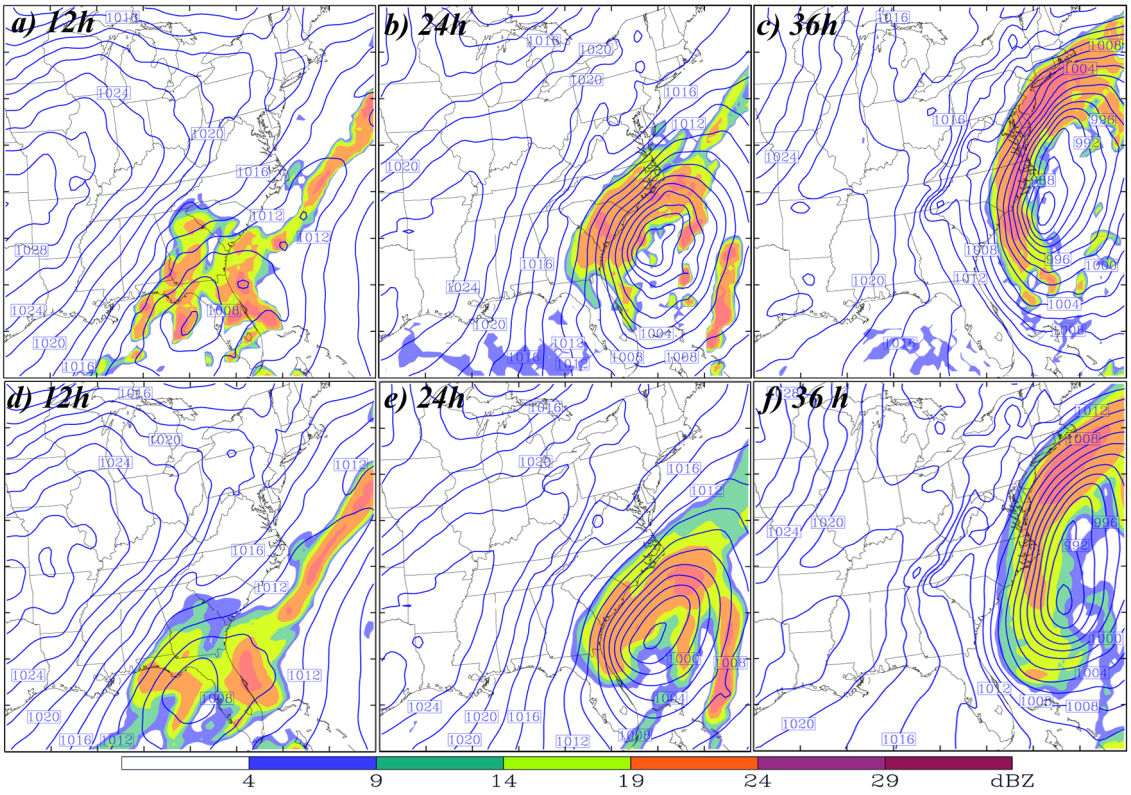

FIG. 2.2. The mean sea-level pressure (MSLP, every 2 hPa) and model-derived reflectivity at the 12-, 24-, and 36-h forecast times from the truth simulation (panels a-c) and the reference forecast ensemble mean (panels d-f).

After 12 h of simulation, the reference forecast ensemble mean, which is used as the first guess in the following EnKF experiments, has noticeable difference from the truth simulation in all fields. In addition to a ~1.5-hPa weaker surface incipient cyclone (differences of wind vectors and MSLP are shown in Fig. 2.4a), the incipient inland precipitation from the Gulf coast across Georgia to South Carolina in the reference forecast ensemble mean is much weaker (Fig. 2.2a vs. Fig. 2.2d). Moreover, the reference forecast ensemble mean of the 300-hPa short-wave PV trough is slightly but systematically shifted to the east (Figs. 2.3a, d, and 2.5a).

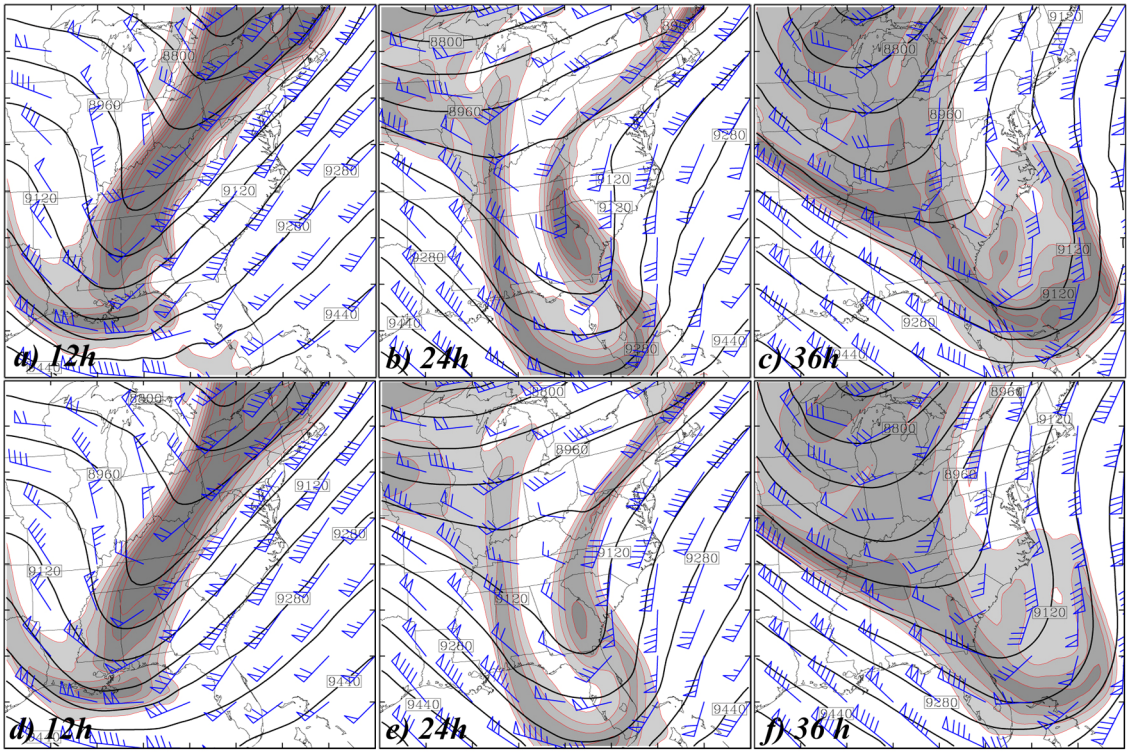

FIG. 2.3. As in Fig. 2.2 but for the geopotential heights (every 80 m), potential vorticity (shaded, every 1 PVU) and vector winds (full barb 5 ms^{-1}) at 300 hPa.

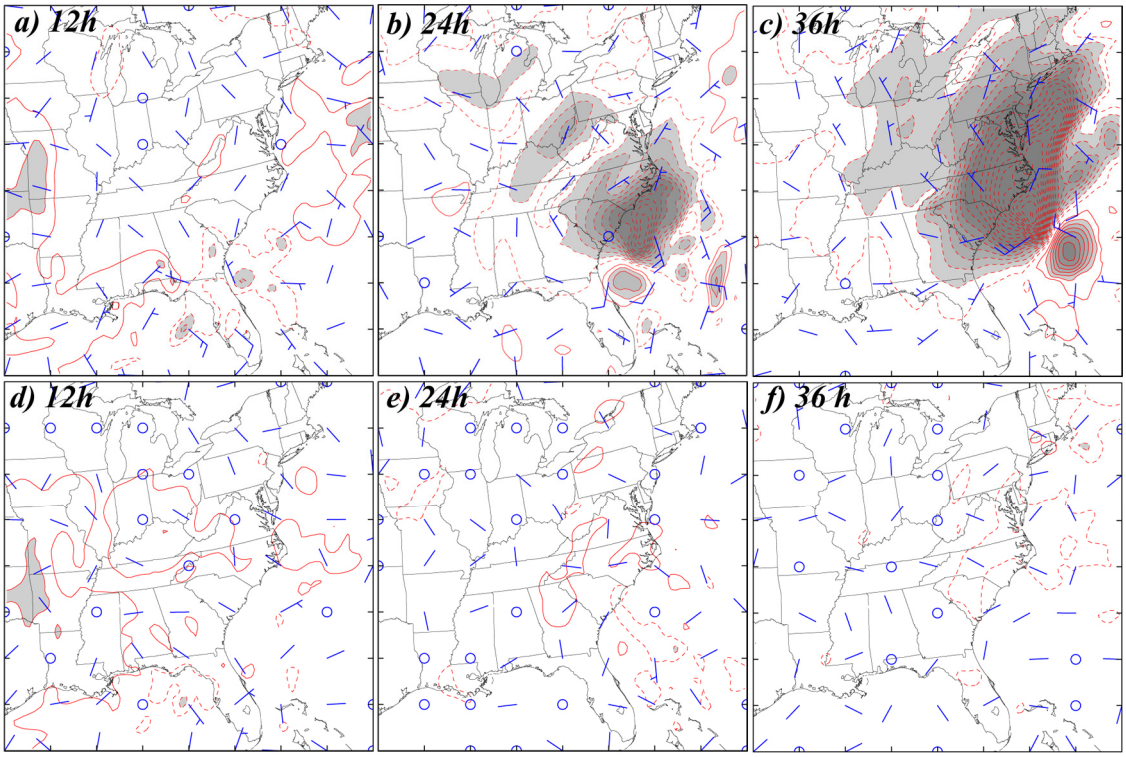

FIG. 2.4. Differences of wind vectors (full barbs 5 ms^{-1}) and MSLP (every 0.5 hPa) between the truth simulation and the reference forecast ensemble mean at (a) 12, (b) 24, and (c) 36 h, and between the truth simulation and the EnKF mean analyses at (d) 12, (e) 24 and (f) 36 h.

At 24 h, the maximum differences of MSLP and winds associated with the surface cyclone between the ensemble mean and the truth simulation are as large as 5 hPa and 12.5 ms^{-1}, respectively (Figs. 2.2b, e, and 2.4b). Moreover, the reference forecast ensemble mean (Fig. 2.2e) also misses the strong inland precipitation across the Carolinas seen in the reference run (Fig. 2.2b) and radar observations (Fig. 3a of ZSR02). Associated with a systematic eastward shift of the upper-level PV trough (fronts) in the ensemble mean forecast, the maximum PV and wind differences at 300-hPa reached an amplitude of 2.5 PVU and 22.5 ms^{-1}, respectively (Figs. 2.3b,e, and 2.5b). Growth of maximum difference along fronts is consistent with the error evolution

in the quasi-geostrophic model examined by Snyder et al. (2003). After 36 h of simulation, due to the strong diabatic destruction of the upper-level PV as the cyclone reaches its peak intensity, the maximum PV difference at 300 hPa (Fig. 2.5c) is slightly smaller than that at 24 h (Fig. 2.5b). Nevertheless, the maximum MSLP difference between the truth simulation and ensemble mean is as high as 8.5 hPa in addition to the even stronger dislocation of the surface cyclone and precipitation band (Figs. 2.2c, f and 2.4c).

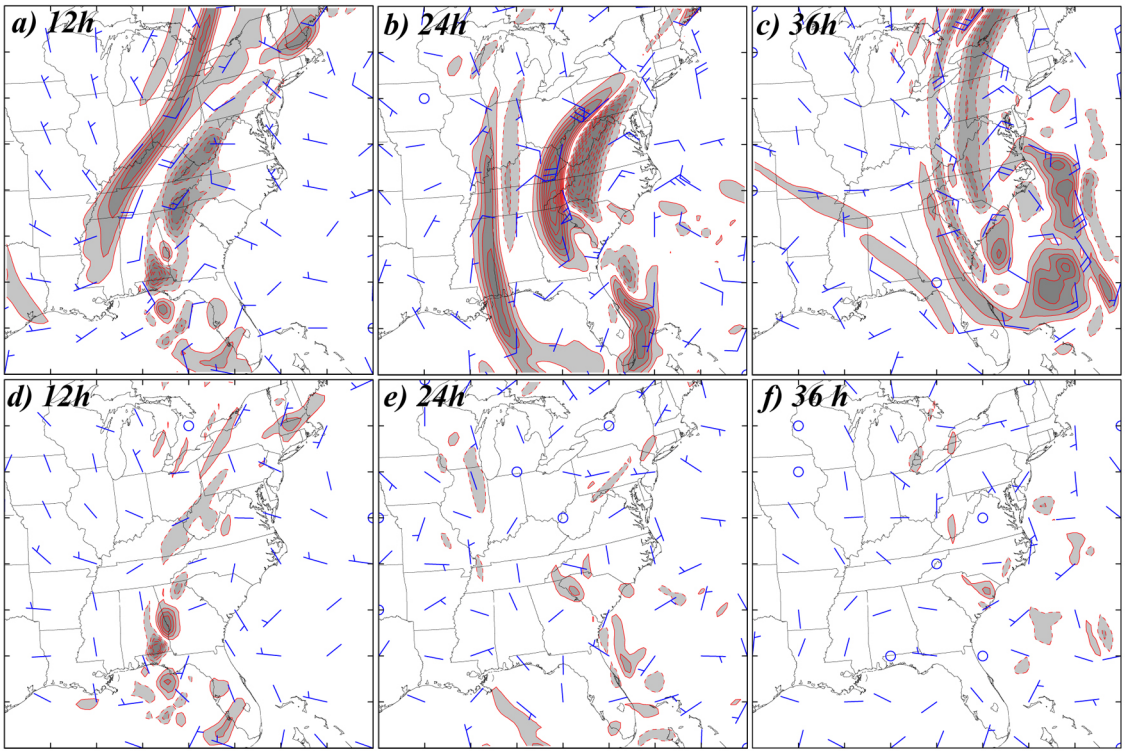

FIG. 2.5 As in Fig. 2.4 but for the difference of 300 hPa potential vorticity (every 0.5 PVU) and winds (full barb 5 ms^{-1}).

The evolution of the forecast error growth revealed from the difference between the truth simulation and the reference forecast ensemble mean can be best summarized in

terms of difference total energy (DTE):

$$\mathrm{DTE}=0.5(u'u'+v'v' + kT'T') \tag{2.4}$$

where primes denote difference between any two simulations and $k=C_p/T_r$ (C_p=1004.7 J kg^{-1} K^{-1} and the reference temperature T_r = 270K). The horizontal distribution of the (vertically-averaged) root-mean (RM) of DTE (RM-DTE) at 12, 24, and 36 h are displayed in Figs. 2.6a-c. The initial RM-DTE from the random initialization of the ensemble forecast using the MM5 3DVar method is ~1.2 ms^{-1} and is nearly constant across the domain (not shown). By 24 and 36 h, it has become greater than 4 ms^{-1} all across the Atlantic Coast with maxima of ~16 ms^{-1}. Consistent with Figs. 2.2-2.5 and Zhang (2005), the maximum error growth occurs near the surface cyclone, the upper-level short-wave trough and associated fronts and moist processes (Figs. 2.2-2.3).

Throughout the study, the reference forecast ensemble is used as a benchmark for the performance of EnKF and the evolution of the analysis error. It is also regarded as the worst-case scenario in which no observations are assimilated.

4. The control EnKF experiment

In the control EnKF experiment with a 40-member ensemble (CNTL), simulated sounding and surface wind and temperature observations are taken from the truth simulation. Typical of the standard sounding and surface observational network over the continental United State, the sounding observations are spaced 300 km apart horizontally and at every sigma level; the surface observations are spaced every 60 km apart and are available at the lowest model level. We impose independent, Gaussian random

observation errors of zero mean and variance of 2.0 ms^{-1} for u and v, and 1.0 K for T. Sounding and surface observations are assimilated every 12 and 3 h, respectively. The forecast model is assumed to be perfect; That is, the same numerical model produces the forecasts and the truth simulation from which observations are taken. We begin assimilating observations at 12 h using the 12-h short-term reference forecast ensemble as the first guess and to estimate the background error covariance.

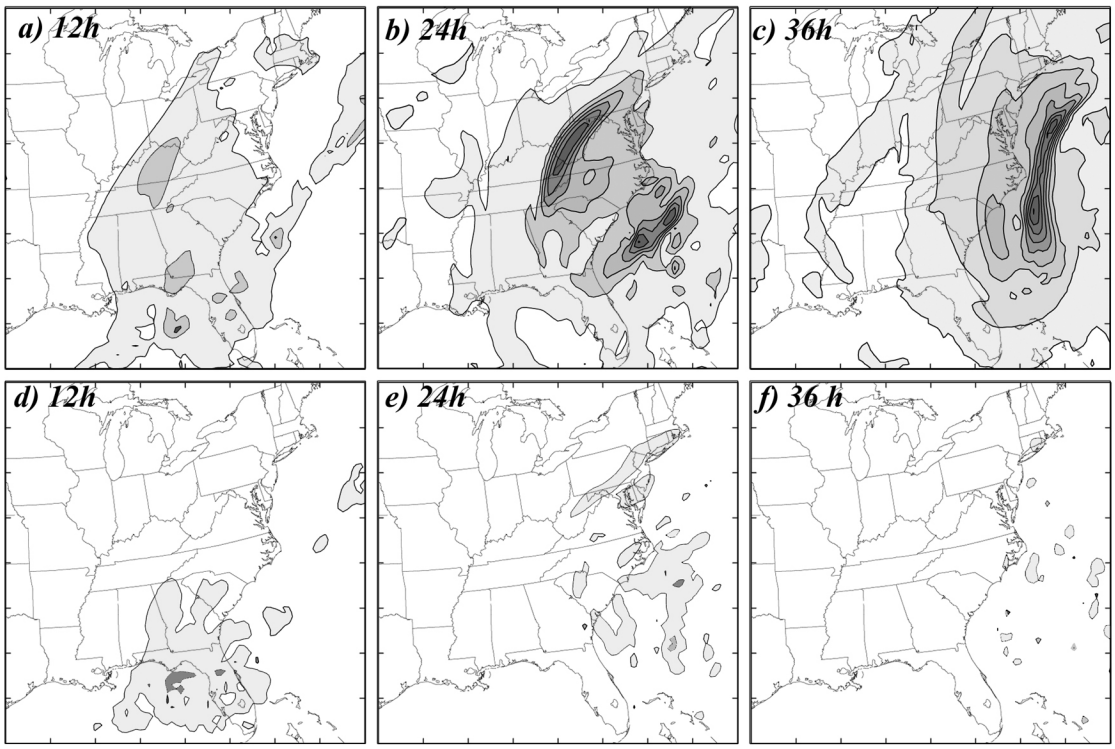

FIG. 2.6. As in Fig. 2.4 but for for the root-mean of column-averaged DTE (RM-DTE; every 2 m/s).

Differences in MSLP and surface winds between the ensemble mean analysis after the EnKF assimilation (EnKF analysis) and the truth simulation at 12-, 24- and 36-h are displayed in Figs. 2.4d-f. At 12 h, after the first cycle of assimilating both the surface

and sounding observations, there is only marginal overall reduction of MSLP error (compared to the reference forecast ensemble) but errors in surface winds are significantly reduced (Fig. 2.4d vs. Fig. 2.4a). At 300 hPa, not only the errors in the winds are reduced by ~30%, errors in PV (as signature of balanced dynamics) are also significantly reduced (Fig. 2.5d vs. Fig. 2.5a). The overall improvement after the EnKF assimilation across the domain is clearly seen in the horizontal distribution of the (column-averaged) RM-DTE in Fig. 2.6d. Compared to the RM-DTE of the mean forecast error of the reference ensemble at this time (Fig. 2.6a), we can see that the improvement is more pronounced in the vicinity of the upper-level short-wave trough than near the surface low, consistent with Figs. 2.4d and 2.5d.

At 24 h, after assimilating five sets of surface observations (every 3 h) and two sets of sounding observations (every 12 h), the EnKF analyses of the surface winds and MSLP and the 300-hPa winds and PV (not shown) approach those in the truth simulation (Figs. 2.2b, 2.3b). More specifically, the maximum analysis errors in surface winds and MSLP are ~2.5 ms^{-1} and 1 hPa, respectively (Fig. 2.4e), which represent 60-80% reduction of the ensemble mean forecast error without the EnKF (Fig. 2.4b). A similar or even larger degree of improvement can also be seen in the analysis error distribution at 300 hPa (Fig. 2.5e vs. Fig. 2.5b). The two local maxima of RM-DTE associated respectively with the upper-level front and the surface low in the forecast (Fig. 2.6b) are no longer noticeable in the DTE of the EnKF analysis (Fig. 2.6e).

Error reduction in both observed and unobserved (or derived) variables continues through 36 h, with more surface and sounding observations assimilated (Figs. 2.4f, 2.5f,

and 2.6f). Most strikingly, compared to the 8.5-hPa MSLP forecast error without EnKF (Fig. 2.4c), the EnKF analysis of MSLP has become nearly indistinguishable from that of the truth simulation. There are only a few small areas with the MSLP error greater than 1 hPa (Fig. 2.4f).

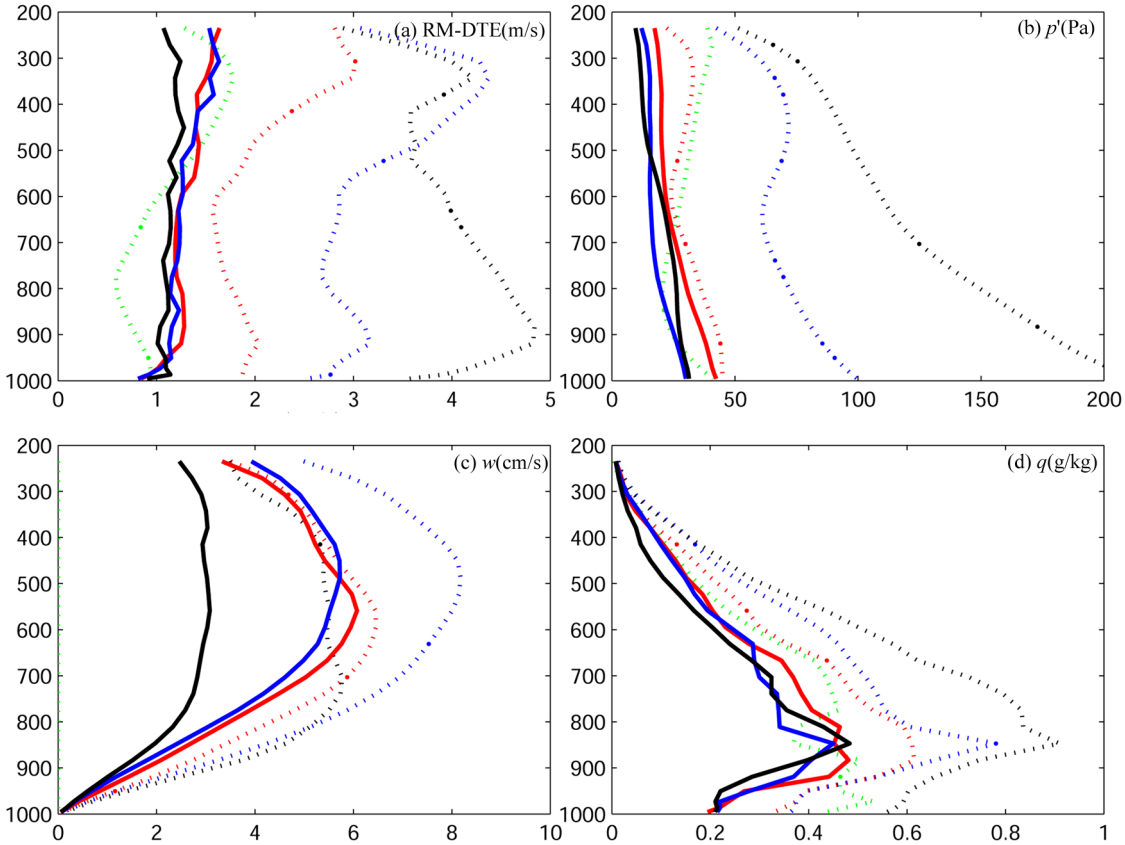

FIG. 2.7. Vertical distribution of the (horizontally-averaged) mean analysis errors of the control EnKF experiment (solid) and the (horizontally-averaged) mean forecast errors of the reference ensemble forecast (dotted) for (a) RM-DTE, (b) p', (c) w, and (d) q valid at 12 h (red), 24 h (blue), and 36 h (black). Errors in the initial ensemble are denoted with dotted green curves.

The vertical distributions of horizontally averaged analysis and forecast errors in terms of RM-DTE, root-mean-square errors of p', w and q at different times are shown in Fig. 2.7. The vertical distribution is plotted only up to 200hPa because differences

among different lines are not significant in layers near the model top. For the RM-DTE (Fig. 2.7a), the forecast error of the reference ensemble gradually grows into a distinct double-peak structure over the 36 h forecast, becoming maximum in the upper and lower troposphere, respectively. The primary peak in the upper troposphere is consistent with the forecast-error statistics of operational ensemble prediction systems (e.g., Molteni et al. 1996) as well as in simplified dry systems (Hamill et al. 2002; 2003). The secondary peak in the lower troposphere is likely due to the lower-level fronts associated with strong moist processes. On the other hand, the analysis error exhibits nearly the same amplitude vertically throughout the troposphere, implying that the largest improvement occurs where the reference forecast ensemble has the largest forecast errors.

For the pressure perturbation p' (Fig. 2.7b), the largest forecast error occurs near the surface. Consistently, through continuous analysis and forecast cycles, the most error reduction occurs in the lower troposphere. For the vertical velocity field (Fig. 2.7c), the reference forecast ensemble mean error peaks at 400-500 hPa layer. Unlike the RM-DTE or p', the forecast error in w follows closely the strength of w in the truth simulation (as an index of the intensity of the background cyclogenesis): the strongest forecast error occurs at ~24 h when there is strongest vertical motion in the truth simulation (not shown); there is an apparent decay of forecast error at 36 h when the surface cyclone has matured and begins to decay. Compared to the reference forecast ensemble mean, the overall error reduction for w at 36 h is 30-40%. Error reduction comes not only from direct EnKF analyses at any given time but also from a better first guess due to the improvement in unobserved variables.

The ensemble forecast error for the moisture field q peaks at 800-900 hPa in association with the abundance of lower-level background moisture as well as moist convection (Fig. 2.7d). The peak forecast error is approximately constant in the EnKF analysis with an overall error reduction of ~50% compared to the reference forecast ensemble.

The performance of the EnKF in this control experiment is best summarized in Fig. 2.8, which shows the evolution of the domain-averaged root-mean-square (RMS) errors in the EnKF analyses of the six prognostic variables (*u, v, T, p', w, q*), the corresponding standard deviation (STD) of the analysis ensemble, and the RMS errors of the reference forecast ensemble. Compared to the reference forecast ensemble, over the 24-h assimilation period, the overall error reduction for the observed variables *u, v, T* is ~60-80%. The overall analysis quality of all variables stays fairly constant throughout the EnKF, indicating that at later times, the error growth during the short-term (3-h) ensemble forecast will be approximately equal to the reduction of analysis error through assimilation of new observations. The final domain-averaged RMS error after 24-h assimilation is ~1.0-1.5 ms^{-1} for winds and ~1.0 K for temperature, which is less than or at most comparable to the specified observational errors. The unobserved variable *p'* has the biggest overall improvement with the 36-h analysis error being only one-sixth of the forecast error. Nearly 50% overall error reduction is observed in the moisture field. Again, there is relatively small (30-40%) overall improvement in the vertical velocity field.

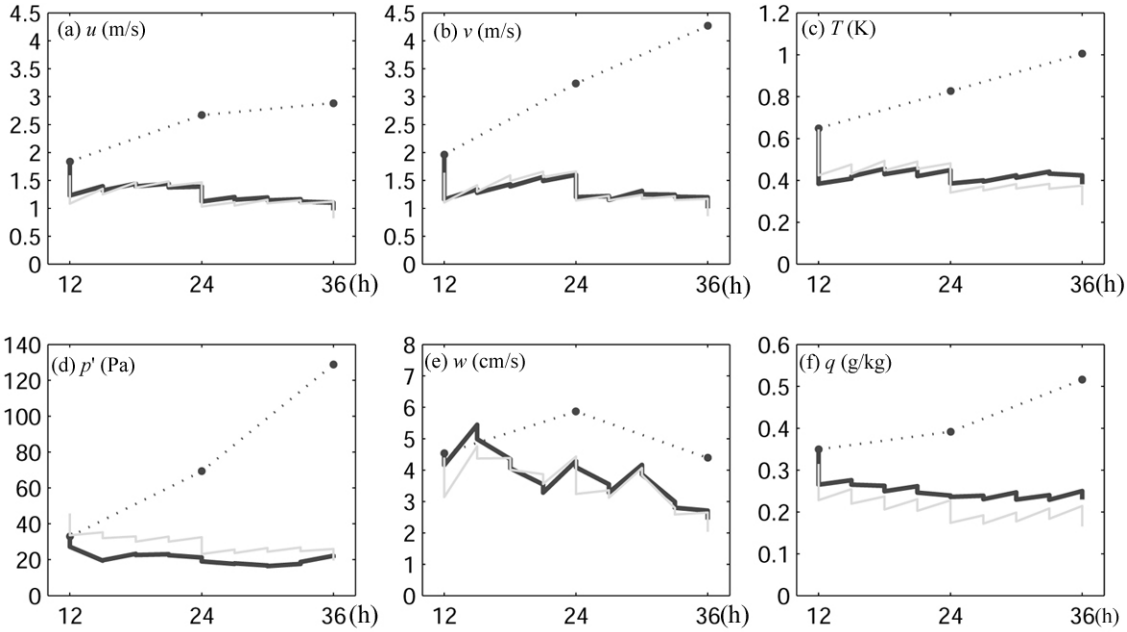

FIG. 2.8. Evolution of the domain-averaged root-mean-square errors of the EnKF analysis (solid black) with respect to the truth simulation for six prognostic variables (u, v, T, p', w, q), the corresponding standard deviation (gray) of analysis ensemble spread, and the root-mean-square errors of the reference ensemble forecast (dotted black, computed every 12 h only).

The difference in the degree of error reduction among different variables is also examined through the comparison of the power spectra of analysis and forecast errors of the reference forecast ensemble and CNTL at different times (Fig. 2.9). The vertical velocity and moisture fields, for which the EnKF assimilation is the least effective, have the most error energy in smaller scales. The pressure field, which has the most power at larger scales, in general enjoys the biggest error reduction. As a result of stronger error reduction at larger scales, power spectra in all the variables (except for *w*) become increasingly flattened at smaller and smaller wave numbers ("whitening"; Hamill et al. 2002; Daley and Menard 1993) through the EnKF assimilation (Fig. 2.9). In essence, the EnKF is very efficient in reducing errors at larger scales but less effective in reducing errors at smaller, marginally resolvable scales. The EnKF analyses of other water

substances associated with clouds, which have the strongest smaller-scale variations, do not have apparent improvement relative to the reference forecast (not shown), suggesting the accurate estimation of clouds with the current EnKF is not yet possible, at least for the current filter configuration and model resolution with parameterized moist convection.

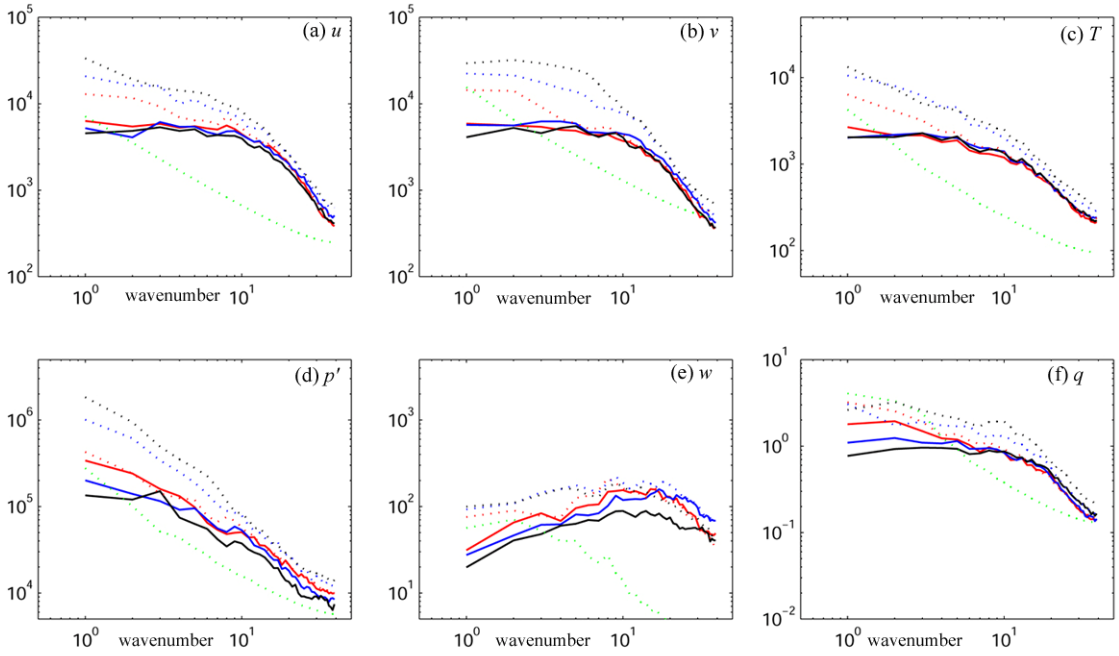

FIG. 2.9. Power spectrum analysis of the EnKF analysis errors (solid) and the reference ensemble forecast mean errors (dotted) at 12h (red), 24h (blue) and 36h (black) for six prognostic variables (u, v, T, p', w, q). The minimum (maximum) wavenumber of 1 (40) corresponds to a horizontal wavelength of 2400 (60) km. Error spectra in the initial ensemble are denoted with dotted green curves.

In an examination of spectral characteristics of Kalman filter systems, Daley and Menard (1993, their Fig. 2) showed that the Kalman filter has a much larger impact on the large scales than the smaller scales. Because the uncorrelated observational error is projected equally to all scales while the model-error spectrum is red, the observations are

considered to be more accurate with respect to the background for the large scales than they are for the small scales. As shown in Daley (1991, his Fig. 5.9), this is strictly applicable to univariate analysis in which the larger scales have the most error energy. The scale-dependent filter performance in the multivariate analysis is much more complex (Daley 1991, his Fig. 5.10), probably even more so for those unobserved variables using the flow-dependent background error covariance at the mesoscale for the current study. Besides the possible mechanisms discussed by Daley (1991, his book section 5.4-5.5), the scale- and variable-dependent filter performance may also be due to faster error saturation (thus shorter predictability) resulting in poorer estimate of the prior guess and background error covariance at the smaller, marginally resolvable scales. It could also arise from observations that are too sparse to provide sufficient information for analysis at smaller scales while larger scales are influenced (corrected) by observations of similar (comparable) horizontal resolutions.

5. Forecast experiments with the EnKF analysis

To evaluate the performance of short-range ensemble forecasts with improved analyses and to examine the forecast error growth dynamics after the EnKF assimilation, two 40-member ensemble forecast experiments ("EF12H" and "EF24H") are performed with the analyses from CNTL at 12 and 24 h as initial conditions. The horizontal distribution of the (vertically averaged) RM-DTE from the 12- and 24-h integration of EF12H and the 12-h integration of EF24H are shown in Fig. 2.10. For EF12H, which starts from the EnKF analysis cycle that assimilated only the observations at 12 h, there

are noticeably smaller ensemble mean errors in RM-DTE at both 24 and 36 h compared to the reference forecast ensemble (Figs. 2.10a-b vs. Figs. 2.6b-c). Even smaller RM-DTE error is found in the 12-h ensemble forecast by EF24H, which starts with the EnKF analysis at 24 h (after a 12-h assimilation period; Fig. 2.10c). Compared to a maximum RM-DTE error of ~16 ms^{-1} just off the Atlantic coast in the reference forecast ensemble (Fig. 2.6c), the maximum RM-DTE error for the 12-h forecast of EF24H is merely ~ 6 ms^{-1} (Fig. 2.10c).

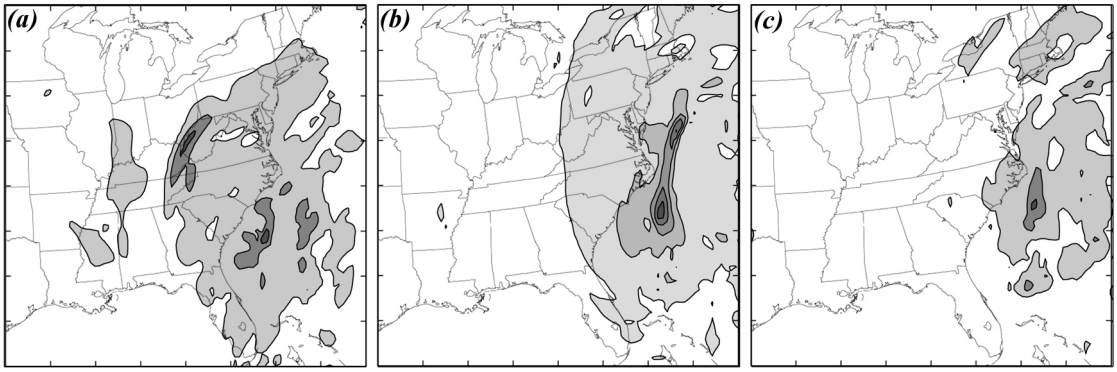

FIG. 2.10. As in Fig. 2.6 but from ensemble forecasts for (a) EF12H at 24 h, (b) EF12H at 36 h, and (c) EF24H at 36 h.

Evolution of the forecast errors of the six prognostic variables from these two forecast experiments (EF12H and EF24H) and the reference forecast ensemble as well as the analysis errors from CNTL is plotted in Fig. 2.11. Again, compared to the reference forecast ensemble, the positive effect of improved initial conditions using the EnKF analysis can be seen in both forecasts verified at 36 h in all prognostic variables shown. It is also seen that, with a longer assimilation period and thus more data being assimilated, the mean forecast error verified at 36 h of EF24H is considerably smaller than that of EF12H and the reference forecast ensemble.

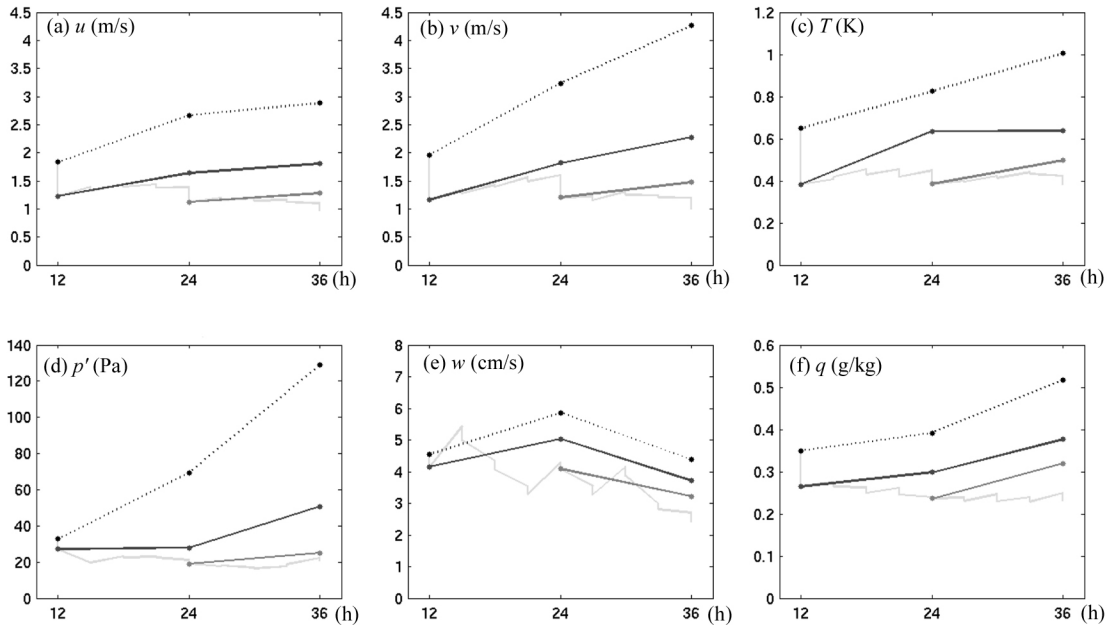

FIG. 2.11. Time evolution of domain-averaged root-mean-square errors from the reference ensemble forecast (dotted), EF12H (solid black; from 12 h), EF24H (solid black; from 24 h) and the control EnKF analysis (gray) for six prognostic variables (u, v, T, p', w, q).

6. Sensitivity experiments

a. Ensemble size, covariance relaxation and localization

Difference in error spectral distribution and error growth dynamics among the different state variables will potentially result in inconsistencies between the analysis/forecast error and ensemble spread between different variables if the same localization or error inflation/relaxation is used for all state variables. For the control EnKF experiment (CNTL), the domain-averaged standard deviations (ensemble spread) of the ensemble forecast and EnKF analysis (Fig. 2.8) stay very close to the RMS errors of the ensemble forecasts and EnKF analyses for all variables. There is no obvious filter divergence, which would be indicated by the growth of the ratio of ensemble mean error to ensemble spread. The (domain averaged) RM-DTE also agrees reasonably well with

the STD of the analyses when an ensemble with a smaller ensemble size of 20 ("CNTL20") is used to estimate the background error covariance (Fig. 2.12a). The difference of the analysis accuracy (in terms of RM-DTE) between CNTL and CNTL20 is rather insignificant (~0.1-0.2 ms^{-1}) throughout the assimilation. On the other hand, though much less accurate than CNTL and CNTL20, a 10-member ensemble EnKF experiment still performs well albeit with a significantly larger ratio of RMS error to STD (not shown).

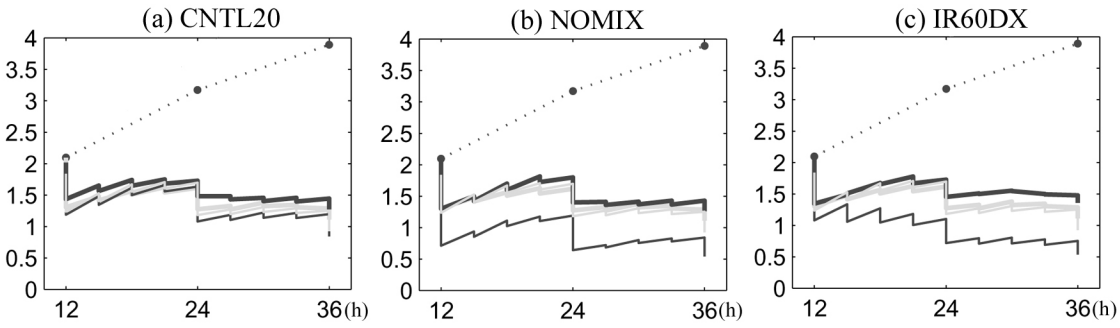

FIG. 2.12. Time evolution of the domain-averaged RM-DTE (m/s, black, thick) and standard deviation (m/s, black, thin) from sensitivity experiments (a) CNTL20, (b) NOMIX and (c) IR60DX. RM-DTE and standard deviation from the control experiment and the RM-DTE from the reference ensemble forecast are also displayed in thick gray, thin gray and dotted curves, respectively.

We observed that even though the pressure perturbation has the maximum overall improvement, the analysis error from CNTL20 after the EnKF assimilation at 12 h is greater than the forecast error at this time (not shown). The degradation occurred only in the lower troposphere for the first assimilation cycle. To test whether the degradation is systematic, we examined three additional experiments: the control experiment with 40 members, an experiment similar to CNTL20 but with different random realizations, and an experiment similar to CNTL20 but with a different truth (discussed in section 6c). In

all three experiments, we observed that the degradation of the pressure analysis did not occur, suggesting that the covariance between the observed variables and the pressure field may be unrepresentative of the true forecast error in pressure perturbations at this time when the ensemble has only 20 members.

The impact of the covariance relaxation can be clearly seen in a 40-member EnKF experiment similar to CNTL but without the application of covariance relaxation ("NOMIX") , which has larger overall RM-DTE and poorer agreement between RM-DTE and STD (Fig. 2.12b). The deficiency in the ensemble spread becomes even more severe when a 20-member ensemble is used (not shown). Thus, the application of the covariance relaxation from Zhang et al. (2004) helps prevent filter divergence that occurs when small ensembles are used. We also tested different implementations of the covariance inflation method (e.g., applying the covariance inflation either before or after the EnKF analysis) used in Anderson (2001) using 40-member ensembles, with the inflation factors from 1.05 to 1.5, either applied before or after the EnKF analysis. None of these additional experiments (not shown) exhibited satisfactory performances comparable to that from the control experiment.

The performance of the EnKF assimilation is also very sensitive to covariance localization. Apparent degradation of EnKF performance and the lack of ensemble spread compared to analysis error (possible filter divergence) are seen in the EnKF experiment ("IR60DX", Fig. 2.12c) in which the three-dimensional distance used in the Schur-product is set to be too large (60 rather than 30 grid points in the CNTL, which is equivalent to 1800 km vs. 900 km in terms of purely horizontal distance). When a 450-

km cut-off radius of influence is used in another EnKF experiment (not shown), the overall performance is similar to CNTL but the ensemble spread is bigger than the analysis error throughout the assimilation period. Similar sensitivity was also reported in the EnKF experiments in Houtekamer and Mitchell (2001, their Fig. 4). Since the best value for the radius of influence is not known *a priori*, the current EnKF configuration may unavoidably need to be "tuned" for different weather systems for best performance.

These sensitivity experiments demonstrate that the ratio of the RMS error of the ensemble forecast and EnKF analysis to the STD of ensemble variance, as a common index of filter divergence, changes with ensemble size, the cut-off radius of influence, and covariance inflation coefficient. Bigger ensemble size, smaller cut-off radius and proper covariance relaxation may lead to larger ensemble spread, potentially preventing severe filter divergence.

b. Observation quality and availability

The ensemble Kalman filter combines information from the initial estimate, the dynamics of the forecast model and the observations to get the best estimate and the associated uncertainty. The quality and availability (coverage, resolution and accuracy) of sounding and surface observations is different from case to case which could impact the ability to estimate the true state. In this subsection, various possible observational scenarios are tested using the EnKF, some of which follow closely those of Zhang et al. (2004). We use a 20-member EnKF with the same truth simulation and the same initial ensemble as those in CNTL20 for all the sensitivity experiments investigated in this

subsection since the difference between CNTL and CNTL20 is rather insignificant (Fig. 2.12a).

Experiment "HALFERR" ("TWICEERR") differs from CNTL20 in that the observational errors of the observed variables (u, v and T) are reduced (increased) to half (twice) of those used in CNTL20. The RMS of the DTE in HALFERR (TWICEERR), albeit slightly (<5%) smaller (larger), shows very similar convergence towards the reference solution in comparison to that of CNTL20 (gray curves; Figs. 2.13a-b). These two experiments demonstrate that, as long as the observational errors are uncorrelated, assimilation with the ensemble filter is rather insensitive to the observational accuracy given the typical range of observational errors for sounding and surface observations, consistent with those convective scale experiments in Zhang et al. (2004).

Experiment "UONLY" differs from CNTL20 in that only the zonal wind from the sounding observations is assimilated, which is similar to a case if we use radar radial velocity instead of sounding observations. Again, the EnKF analysis converges well towards the truth simulation over the 24-h assimilation; the RM-DTE at 36 h is only ~10-20% larger than that in the CNTL20 (Fig. 2.13c). In another experiment similar to CNTL20 but with the addition of pressure perturbation and moisture observations in the soundings, there is no significant improvement in the EnKF analysis compared to CNTL20 for all prognostic variables including p' and q (not shown). Filter performance is also nearly unchanged when the horizontal spacing of the sounding network changed from 300 km in CNTL20 to 450 km in the experiment "SND450KM" (Fig. 2.13d).

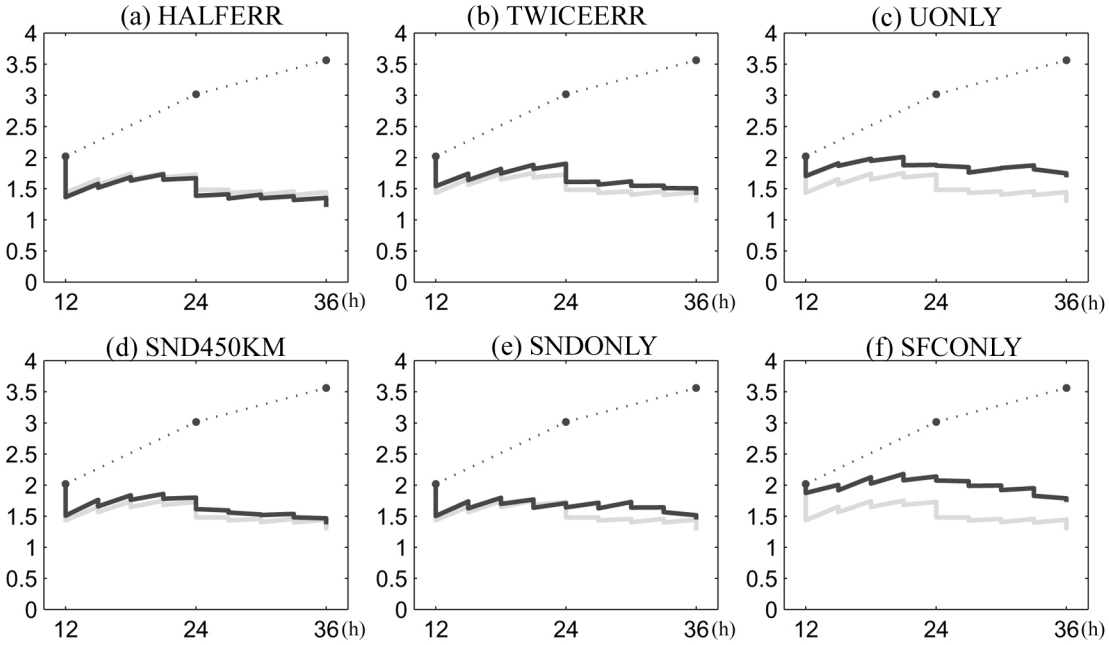

FIG. 2.13. Time evolution of the domain-averaged RM-DTE (m/s, black, thick) from sensitivity experiments (a) HALFERR, (b) TWICEERR, (c) UONLY, (d) SND450KM, (e) SNDONLY and (f) SFCONLY. RM-DTE from experiment CNTL20 and the 20-member reference ensemble forecast are also displayed in gray and dotted curves, respectively.

Experiments "SNDONLY" and "SFCONLY" differ from CNTL20 in that only either sounding or surface observations are assimilated every 3 h. For the first 12-h assimilation period of SNDONLY, the analysis follows closely that of the CNTL20 but the loss of surface observations cannot be corrected by more frequent sounding observations for the final 12-h assimilation period (Fig. 2.13e). Consistent with Whitaker et al. (2004), it is very encouraging to notice that the analysis still has significant smaller error than the reference forecast when only surface observations are assimilated (Fig. 2.13f), even though the advantage of sounding observations is clearly seen when compared to CNTL20.

c. Different truth simulations

We also performed several additional experiments with the same set of initial ensembles as in CNTL20 but using different realizations of the 3DVar perturbations to generate the truth simulation. Quantitatively similar performance (to the CNTL20) has been achieved in all of these EnKF experiments (not shown). Another experiment with the same truth simulation as in CNTL20 but a different set of 20 ensemble members behaves in a similar manner (not shown).

7. Summary and discussions

Through various observing system simulation experiments, this study exploits the potential of using the ensemble Kalman filter (EnKF), which estimates error covariances through an ensemble of short-term forecasts, for mesoscale and regional-scale data assimilation. The EnKF is implemented in a non-hydrostatic, mesoscale model (MM5) to assimilate simulated sounding and surface observations derived from truth simulations of the "surprise" snowstorm of January 2000. This is an explosive east coast cyclogenesis event with strong error growth at all scales as a result of interactions between convective-, mesoscale, and subsynoptic-scale dynamics.

It is found that the EnKF is very effective in keeping the analysis close to the truth simulation. In the control experiment (CNTL), a 24-h continuous EnKF assimilation of sounding and surface observations with realistic temporal and spatial resolutions can have an error reduction of as much as 80% for horizontal winds and temperature, 85% for pressure perturbation, and 45% for water vapor mixing ratio in comparison to the

reference forecast ensemble.

Error growth characteristics in the ensemble forecast with and without the EnKF, including the scale, structure and evolution of the forecast and analysis errors of different variables are also examined. It is found the EnKF is most effective in reducing larger-scale errors but less effective in reducing errors at smaller, marginally resolvable scales. This is consistent with the analysis of spectral characteristics of Kalman filter systems by Daley (1991) and Daley and Menard (1993). The scale-dependent error reduction may also be due to the faster error saturation (thus shorter predictability) and thus poorer quality of the prior estimate and background error covariance at the smaller, marginally resolvable scales. It could also arise from observational information that is insufficient to allow for a good estimate at smaller scales. There are also apparent improvements in the forecast initiated with EnKF analysis. Since error grows at all scales but saturates quicker at smaller scales, the overall error growth in the ensemble forecasts may largely comes from larger scales.

Error growth characteristics and the quality of initial estimate and background error covariance also differ greatly from variable to variable, resulting in different degrees of error reduction for different variables. The EnKF is least effective on the vertical motion and moisture fields which have more energy in smaller scales while pressure perturbation in general enjoys the biggest error reduction because it has the strongest larger-scale component among all variables. Different error growth from different variables also results in inconsistency between the analysis error and ensemble spread of different variables when the same localization or error inflation/relaxation is used for all

variables.

It is also found that the ratio of the root-mean-square analysis/forecast error to the standard deviation of the ensemble, as a common index of filter divergence, changes with ensemble size, the cut-off radius of influence (localization), and covariance relaxation (inflation) coefficient. Consistent with past studies, it is found that bigger ensemble size, smaller cut-off radius and the implementation of the variance relaxation method all lead to larger ensemble spread and potentially prevent from filter divergence.

Various experiments are also performed to test the sensitivity of the EnKF to the number of observed variables and the density and accuracy of sounding and surface observations. The EnKF is found to be quite resilient in most of the realistic observational scenarios tested.

The above conclusions on the mesoscale data assimilation with the EnKF are drawn from observation system simulation experiments under the perfect model assumption. Such a strong EnKF performance should not be readily expected in real-world situations where the forecast model unavoidably has errors and the initial ensemble statistics may be far from perfect (refer to Houtekamer et al. 2005). The EnKF performance under various imperfect-model scenarios will be explored in next chapter.

CHAPTER III

OBSERVING SYSTEM SIMULATION EXPERIMENTS WITH IMPERFECT MODEL*

1. Introduction

In the past few years, ensemble-based data assimilation has drawn increasing attention from the data assimilation community due to its prevailing advantages such as flow-dependent background error covariance, ease of implementation and its use of a fully nonlinear model. Since first proposed by Evensen (1994), ensemble-based data assimilation has been implemented in various scale numerical models of different realistic scenarios (Houtekamer and Mitchell 1998; Hamill and Snyder 2000; Keppenne 2000; Anderson 2001; Mitchell et al. 2002; Keppenne and Rienecker 2003; Zhang and Anderson 2003; Snyder and Zhang 2003; Houtekamer et al. 2005; Whitaker et al. 2004; Dowell et al. 2004; Zhang et al. 2004, 2006a; Tong and Xue 2005; Aksoy et al. 2005). More background on ensemble-based data assimilation can be found in recent reviews of Evensen (2003), Lorenc (2003) and Hamill (2006).

While Hamill and Snyder (2000), Whitaker and Hamill (2002), and Anderson (2001) showed that using an ensemble Kalman filter (EnKF) in the context of a perfect model (i.e., both the truth and ensemble propagate with the same model) can significantly

---

reduce error and outperform competitive data assimilation methods such as 3DVar, the perfect model assumption must be dropped in real-world studies where model error is caused by inadequate parameterization of subgrid physical processes, numerical inaccuracy, truncation error and other random errors. The presence of model error can often result in both a large bias of the ensemble mean and too little spread and can ultimately cause the ensemble forecast to fail. Fortunately, studies (e.g., Houtekamer et al. 1996; Houtekamer and Lefaivre 1997) show that including model error in an ensemble can lead to a more realistic spread of the forecast solution. Despite this, model error, especially at the mesoscale, is generally hard to identify and to deal with due to the chaotic nature of the atmosphere, its flow-dependent characteristics, and the lack of sufficiently dense observations for verification (e.g., Orrell et al. 2001; Orrell 2002; Simmons and Hollingsworth 2002; Stensrud et al. 2000).

There have been several different approaches for including model error in ensemble forecasts. One popular (yet at hoc) approach involves the use of different forecast models (e.g., Evans et al. 2000; Krishnamurti et al. 2000) or different physical parameterization schemes (e.g., Stensrud et al. 2000). Other ways to include model error are to apply statistical adjustment to ensemble forecasts (Hamill and Whitaker 2005) or to use stochastic forecast models and/or stochastic physical parameterizations (e.g., Palmer 2001; Grell and Denvenyi 2002).

Mitchell et al. (2002), Hansen (2002), Keppenne and Rienecker (2003), Hamill and Whitaker (2005), and Houtekamer et al. (2005) have all discussed explicit treatment of model error in ensemble-based data assimilation. For example, Keppenne and Rienecker

(2003) obtained encouraging results using covariance inflation (first proposed in Anderson 2001) with an oceanic general circulation model and real data. In a study that showed ensemble data assimilation can outperform 3DVar, Whitaker et al. (2004) also used the covariance inflation method to reanalyze the past atmospheric state using a long series of available surface pressure observations. Despite these successes, covariance inflation can cause a model to become unstable due to excessive spread in data-sparse regions (Hamill and Whitaker, 2005). The additive error method (Hamill and Whitaker 2005; Houtekamer et al. 2005) and the covariance relaxation method of Zhang et al. (2004) have recently been proposed as alternatives to covariance inflation. The performance of certain additive error methods was found to be superior to covariance inflation for the treatment of model truncation error caused by lack of interaction with smaller scale motions, and additive error methods might outperform a simulated 3DVar method (Hamill and Whitaker 2005). Meanwhile, Houtekamer et al. (2005) used a medium-resolution, primitive equation model with physical parameterizations and similarly parameterized model error by adding noise consistent in structure with 3DVar background error covariance. The EnKF performed similarly to the 3DVar method implemented in the same forecast system.

Most of the aforementioned studies that included an explicit treatment of model error used global models. To the best of our knowledge, the impacts of model error on ensemble data assimilation with a mesoscale model have rarely been addressed in the literature. Applications of an EnKF to the mesoscale have only recently begun with simulated observations (Snyder and Zhang 2003; Zhang et al. 2004; Tong and Xue 2005;

Caya et al. 2005; Aksoy et al. 2005, 2006a, b; Zhang et al. 2006a, hereafter also referred to as Chapter II) and with real data (Dowell et al. 2004; Dirren et al. 2006). In Chapter II, the authors examined the performance of an EnKF implemented in a mesoscale model through various observing system simulation experiments (OSSEs) assuming a perfect model. It is found that the EnKF with 40 members works very effectively in keeping the analysis close to the truth simulation. The result that most error reduction comes from large scales is consistent with Daley and Menard (1993). Furthermore, the EnKF performance differs among variables; it is least effective for vertical motion and moisture due to their relatively strong smaller scale power, and it is most effective for pressure because of its relatively strong larger scale power.

This chapter examines the performance of the same EnKF as that in Chapter II in the presence of significant model error due mainly to physical parameterizations. Past studies (e.g., Stensrud et al. 2000) suggested that a considerable part of model error comes from parameterization of sub-scale physical processes. The 'surprise' snowstorm of 24-26 January 2000 that was examined in Chapter II is also examined here.

In the next section, we describe the methodology, experimental design, and ensemble and model configurations. A synoptic overview and the control experiment results are described in section 3. Section 4 demonstrates the sensitivity of the EnKF to model error due to physical parameterizations. The EnKF performance in another case with a distinguishably different flow regime [the long-lived warm-season mesoscale convective vortex (MCV) event that occurred on 10-13 June 2003] is then examined in section 5 to address the impact of flow-dependent predictability. Finally, section 6 gives

our conclusions and discussions.

2. Methodology and experimental design

Unless otherwise specified, the EnKF system used here is the same as that employed in section 2 of Chapter II. It is a square-root EnKF with 40 ensemble members that uses covariance relaxation (Zhang et al. 2004, their Eq.5 where α=0.5) to inflate the background error covariance. The Gaspari and Cohn (1999) fifth-order correlation function with a radius of influence of 30 grid points (i.e., 900 km in horizontal directions and 30 sigma levels in vertical domain) is used for covariance localization.

The third version of the PSU-NCAR mesoscale model MM5 (Dudhia 1993) is used herein with 190×120 horizontal grid points and 30-km grid spacing to cover the continental United States (Fig. 3.1; a slightly newer update of MM5 version 3 is used here than was used in Chapter II). The model setup also includes 27 layers in the terrain-following vertical coordinate with the model top at 100 hPa, and a smaller vertical spacing within the boundary layer. NCAR-NCEP reanalysis data are used to create the initial and boundary conditions.

Various experiments are performed with different model configurations (Table 3.1) to explore the sensitivity of the EnKF to the uncertainties in physical parameterizations. Serving as a benchmark, the control experiment "CNTL" is performed under the assumption of a perfect forecast model in the same manner as in Chapter II (Section 3 of Zhang et al. 2006a) and it utilizes the Grell cumulus parameterization scheme, Reisner microphysics scheme with graupel and Mellor-Yamada (ETA) planetary boundary layer

(PBL) scheme [refer to Grell et al. (1994) and Wang and Seaman (1997) for a description of different parameterization schemes]. Aside from the control experiment, sensitivity experiments are conducted with different cumulus parameterizations ("KFens", "KF2ens", "BMens", "KUOens", "Multi1", "Multi2", "KF3ens", "Multi3", and "Multi4") and are described in Table 3.1 and Table 3.2.

TABLE 3.1. Model configurations of various experiments.

Experiments	**Physical parameterization schemes**		
	cumulus parameterization	PBL	microphysics
CNTL	Grell	ETA	Graupel（Reisner）
KFens	Kain-Fritsch (KF)	ETA	Graupel（Reisner）
KF2ens	Kain-Fritsch 2 (KF2)	ETA	Graupel（Reisner）
BMens	Betts-Miller (BM)	ETA	Graupel（Reisner）
KUOens	Anthes-KUO (KUO)	ETA	Graupel（Reisner）
Multi1	Grell, BM, KUO, KF	ETA	Graupel（Reisner）
Multi2	KF2, BM, KUO, KF	ETA	Graupel（Reisner）
KF3ens	Kain-Fritsch (KF)	MRF	Graupel（GSFC）
Multi3	KF2, BM, KUO, KF	MRF	Graupel（GSFC）
Multi4	Refer to Table 3.2		

TABLE 3.2. Model configuration of experiment "Multi4".

No. of ensemble members using cumulus scheme	**No. of ensemble members Using Cloud microphysics scheme**	**No. of ensemble members using PBL scheme**
10 Grell	5 Graupel（Reisner）	3 ETA 2 MRF
	5 Graupel（GSFC）	2 ETA 3 MRF
10 KUO	5 Graupel（Reisner）	3 ETA 2 MRF
	5 Graupel（GSFC）	2 ETA 3 MRF
10 KF	5 Graupel（Reisner）	3 ETA 2 MRF
	5 Graupel（GSFC）	2 ETA 3 MRF
10 BM	5 Graupel（Reisner）	2 ETA 3 MRF
	5 Graupel（GSFC）	3 ETA 2 MRF

The initial conditions for both the truth simulation and the ensemble are generated

with the MM5 3DVar method (Barker et al. 2004; also Chapter II). The perturbation standard deviations are approximately 1 m s^{-1} for wind components u and v, 0.5 K for temperature T, 0.4 hPa for pressure perturbation p', and 0.2 g/kg for water vapor mixing ratio q. Other prognostic variables (vertical velocity w, mixing ratios for cloud water q_c, rain wate q_r, snow q_s and graupel q_g) are not perturbed. The 3DVar perturbations are added to the NCEP reanalysis at 0000 UTC 24 January 2000 to form an initial ensemble which is then integrated for 12 h to develop a realistic, flow-dependent error covariance structure before the first data is assimilated. A relaxation inflow-outflow boundary condition is adopted for both the truth simulation and the ensemble.

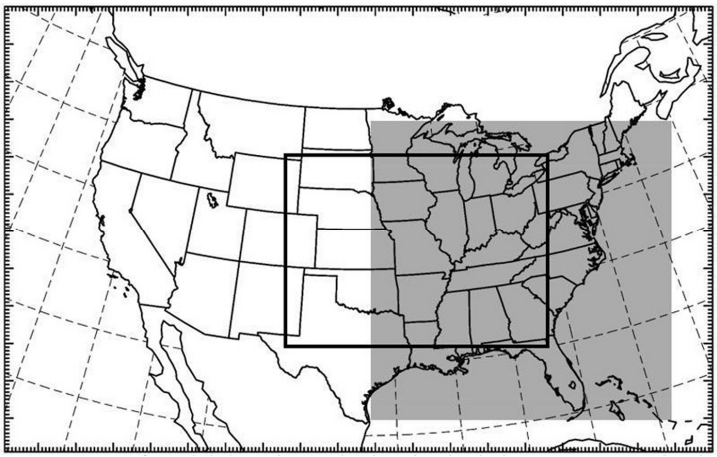

FIG. 3.1. Map of the model domain. Observations are extracted only from the area inside the shaded (solid) box for the snowstorm (MCV) case.

As in Chapter II, the tendencies in lateral boundaries are not perturbed. Instead, the analysis step of the EnKF is implemented only upon an inner area far from the inflow boundary as shown by the shaded box in Fig. 3.1. Since the reference ensemble forecast has no apparent decrease of variance in the inner (assimilation) domain, the lack of

boundary perturbations is assumed to have minimal effects upon the experiments. Also, examination of both the EnKF analyses and subsequent ensemble forecasts reveals no apparent inconsistencies at and near the boundaries of the inner (assimilation) domain.

Simulated soundings and surface observations of u, v, and T are extracted from within the assimilation domains of the truth simulations. The soundings are spaced every 300 km horizontally and sounding observations are taken at every sigma level. Surface observations are spaced every 60 km at the lowest model level (approximately 36 m above the ground). Assimilation of real surface observations could be more problematic due to the representative error, strong gradients and fluxes near the surface. We assume that all observations have independent Gaussian errors with zero mean and a standard deviation of 2 m s^{-1} for u and v, and 1 K for T. Sounding and surface observations are assimilated every 12 h and 3 h, respectively. Starting from the 12^{th} hour into the integration, data assimilation continues for 24 hours. Only the state variables inside the shaded box are updated and analyzed.

3. Overview of the event and the control experiment

a. Synoptic overview

The case that we investigate is an intense winter storm that occurred during 24-26 January 2000 off the southeastern coast of the United States and brought heavy snowfall from the Carolinas through the Washington D.C. area and into New England. Snow associated with this storm fell across North Carolina and the Raleigh-Durham area reported a record snowfall total of over 50 cm (Zhang et al. 2002). The system

developed as an upper-level short wave embedded in a broad synoptic trough over the eastern United States moved southeastward across the southeast states. A 300 hPa low formed around 0000 UTC 25 January near the coasts of Georgia and South Carolina and moved northward along the coast. The upper level low reached southeastern North Carolina by 1200 UTC 25 January (Zhang et al. 2002, Fig. 2), and the storm produced the most intense snowfall in this area. The minimum mean sea level pressure (MSLP) associated with the surface cyclone rapidly dropped from 1005 hPa at 1200 UTC 24 January to 983 hPa at 1200 UTC 25 January. The surface low then gradually weakened as it followed the northward-moving upper low along the coast.

b. The control EnKF experiment

The control EnKF experiment ("CNTL") utilizes the perfect model scenario of Chapter II in which the truth and the ensemble are simulated with the same forecast model physics configuration. The truth simulation is the ensemble member that most accurately simulates the observed location and intensity of the surface and 300-hPa cyclones and simulated reflectivity (Fig. 3.2). The reference ensemble forecast, which uses the same initial conditions as the ensemble in CNTL but is a pure ensemble forecast without any data assimilation, demonstrates rapid error growth in terms of both MSLP and surface wind forecast error (Figs. 3.3a-b) and in terms of the square root of column-averaged (mean) difference total energy (RM-DTE) (Figs. 3.3d-e). The DTE is defined as:

$$\mathrm{DTE} = 0.5(u'u' + v'v' + \mathrm{k}T'T'), \tag{3.1}$$

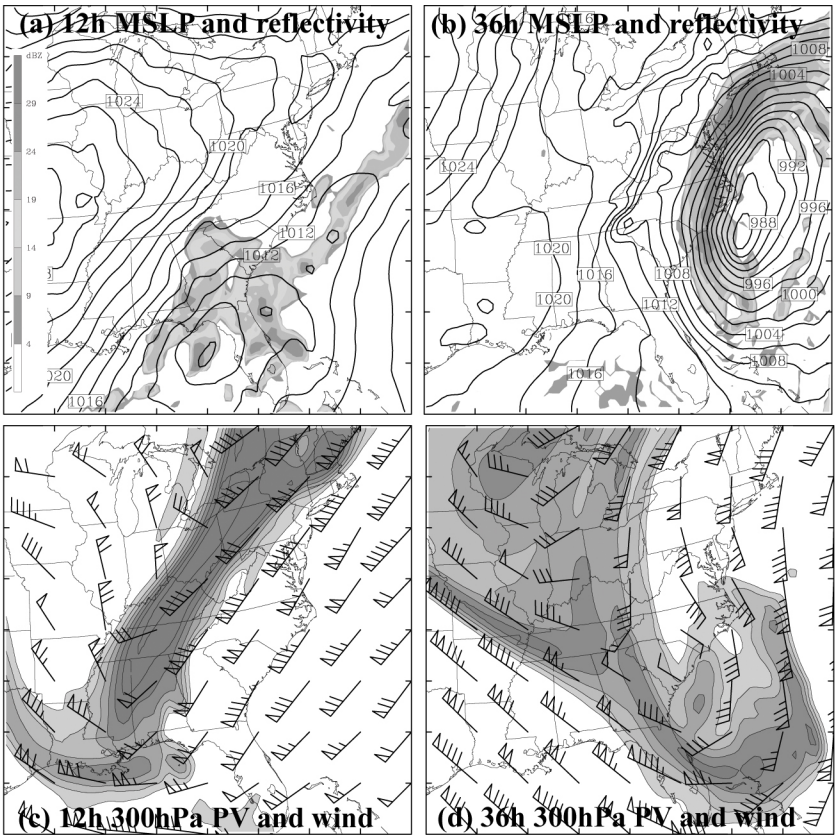

FIG. 3.2. The MSLP (every 2 hPa) and simulated reflectivity (shaded) valid at (a) 12 h and (b) 36 h from the truth simulation for the snowstorm case. (c-d) are as (a-b) but for the potential vorticity (every 1 PVU) and wind vectors (full barb 5 m s-1) at 300 hPa.

where the prime denotes the difference between the truth and the ensemble mean or between any two realizations, $k = C_p/T_r$, C_p = 1004.7 J kg^{-1} K^{-1} and the reference temperature T_r = 270 K. Large increases can be seen in the maximum forecast error of different variables from 12 h to 36 h. For example, the error increases from 1.5 hPa to 8.5 hPa with MSLP (Figs. 3.3a-b, respectively), from 2.5 m s^{-1} to 12.5 m s^{-1} with the surface wind, and from 1.2 m s^{-1} to 16 m s^{-1} for the column-averaged RM-DTE (Figs. 3.3d-e). Note that large errors generally occur near the surface cyclone, the upper-level short-wave trough, and the associated fronts and moist processes. These results are

consistent with Chapter II, Zhang et al. (2002, 2003) and Zhang (2005).

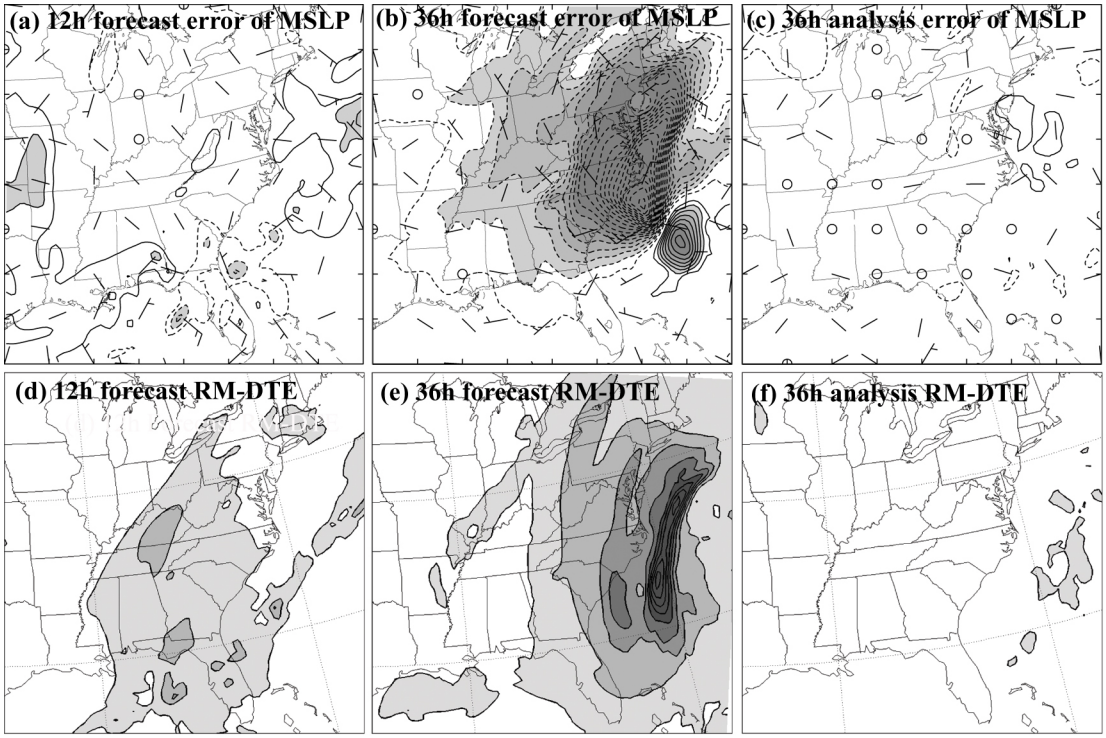

FIG. 3.3. Forecast errors of surface wind vectors (full bard 5 m s-1) and MSLP (every 0.5 hPa) at (a) 12 h and (b) 36 h for the snowstorm case. The analysis error of the same fields at 36 h is in (c). (d-f) are as (a-c) but for the column-averaged RM-DTE (every 2 m s-1).

After 24 hours of assimilation with the EnKF, the analysis error (defined as the difference between the posterior ensemble mean and the truth) decreases significantly for all variables of interest. The EnKF analysis of MSLP (Fig. 3.3c) and the column-averaged RM-DTE (Fig. 3.3f) are almost indistinguishable from those of the truth simulation.

Relative error reduction ("RER", as in Chapter II) will be used to verify the performance of the EnKF. RER is defined as

$$\mathrm{RER}=\frac{E_f - E_a}{E_f}\times 100\% \tag{3.2}$$

where E_f denotes the root-mean-square error of the reference ensemble forecast of an arbitrary variable in an experiment, and E_a denotes the root-mean-square error of the corresponding analysis of the same variable. As shown in the time evolution of forecast and analysis error for different variables including *u, v, T, p', w* and *q* (Fig. 2.8), the EnKF reduces the analysis error by as much as 85% for pressure perturbation, 80% for horizontal wind and temperature, 45% for water vapor mixing ratio and 30% for vertical velocity. The largest improvement is obtained when both sounding and surface observations are assimilated (Chapter II).

The effectiveness of the EnKF in "CNTL" is shown in Figs. 3.4 and 3.5. For example, there is no apparent filter divergence because the ensemble spread (dotted line in Fig. 3.4a) and analysis errors (solid thick dark-gray line in Fig. 3.4a) are quite close to each other. Also, compared to that in the reference ensemble forecast without data assimilation (dashed thick dark-gray line in Fig. 3.4a), RM-DTE is reduced by 73% (to 1.1 m s^{-1}) after the 24-h assimilation period (solid thick dark-gray line in Fig. 3.4a). In fact, the RM-DTE value after the assimilation period is less than that of typical observation errors. The vertical profile of horizontally-averaged RM-DTE at 36 h (Fig. 3.4d, thick dark-gray lines) suggests that the largest improvement occurs where the reference ensemble forecast has the largest error. Moreover, the power spectrum analysis of DTE at 36 h (Fig. 3.5a, solid thick dark-gray line for analysis and dotted line for reference forecast) demonstrates that the EnKF is very efficient at decreasing the error at larger scales where the covariance is most reliable. The EnKF less effectively reduces error at smaller, marginally resolvable scales. This is possibly due to the poor

representation of background error covariance, faster error growth at smaller scales, and/or insufficient observation information (Chapter II).

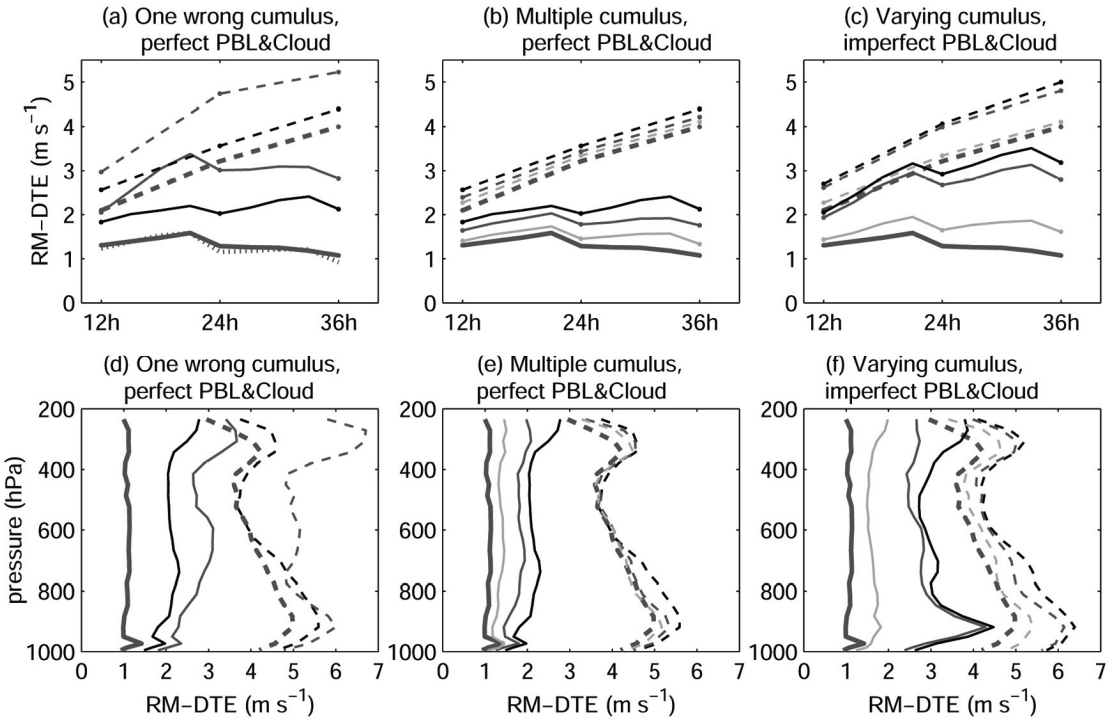

FIG. 3.4. (a-c) Time evolution of domain-averaged RM-DTE for different experiments and (d-f) the vertical distribution of horizontally-averaged RM-DTE of the EnKF analysis (solid lines) and corresponding reference forecast (dashed lines) for the snowstorm case at 36 h. (a) and (d) are for one wrong cumulus parameterization scheme and perfect PBL and microphysics. This includes experiments "KFens" (thin black), "KUOens" (thin dark-gray), and "CNTL" (thick dark-gray). The dotted line in (a) denotes the standard deviation of the EnKF analysis ensemble in "CNTL" in terms of RM-DTE. (b) and (e) are similar to (a) and (d) but for multiple cumulus schemes, including "Multi1" (thin gray), "Multi2" (thin dark-gray), "KFens" (black), and "CNTL" (thick dark-gray). (c) and (f) are similar to (a) and (d) but for varying cumulus and imperfect PBL and microphysics including experiments "KF3" (black), "Multi3" (thin dark-gray), "Multi4" (thin gray), and "CNTL" (thick dark-gray).

4. Sensitivity to model error in physical parameterizations

As mentioned in the introduction, model error can result in bias of the ensemble mean and insufficient ensemble spread due to its smaller projection onto the correct error growth direction. In numerical models, those processes that cannot be explicitly resolved

have to be approximated through different parameterization schemes that are major sources of model error. To test the performance of the EnKF in the presence of model error caused by physical parameterization schemes, we assume that the Grell cumulus scheme, the ETA PBL and the Reisner microphysics with graupel, which are employed to generate the truth simulation, are perfect. The ensemble forecast in the sensitivity experiments is then performed with either one or multiple parameterization schemes that differ from the truth simulation.

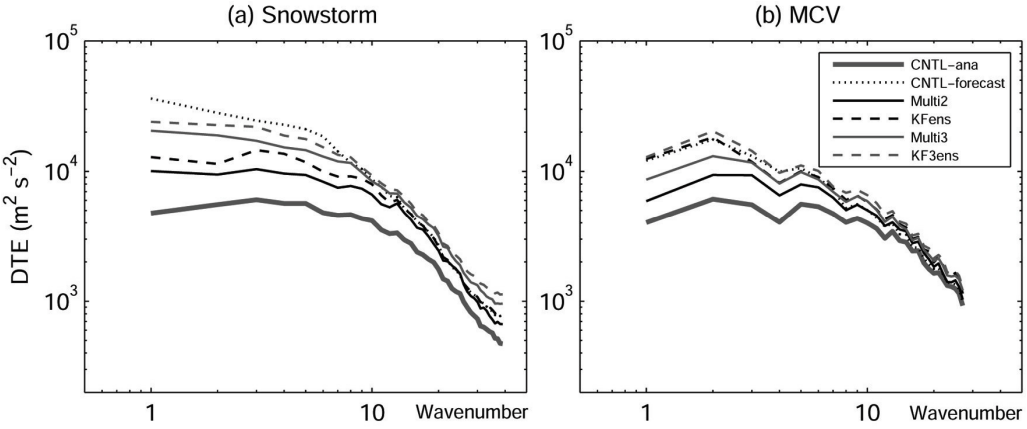

FIG. 3.5. Power spectrum of DTE for (a) the snowstorm at 36 h and (b) the MCV event at 48 h. The minimum (maximum) wave number 1 (40) in (a) and 1 (28) in (b) correspond to a horizontal wavelength of 2400 (60) km in (a) and 1680 (60) km in (b).

a. Impact of cumulus parameterization under perfect PBL and microphysics schemes

Cumulus parameterization, the problem of formulating the statistical effects of moist convection to obtain a closed system for predicting weather and climate (Arakawa 2004), has greater uncertainty than any other aspect of mesoscale numerical prediction (Molinari and Dudek 1992). Cumulus parameterization generally improves precipitation forecasts when it is utilized in a global/synoptic-scale model with a grid spacing of about

100 km or larger (Molinari and Corsetti 1985). Problems arise when the grid spacing reduces to below 50 km and initially irresolvable clouds turn into resolvable mesoscale circulations at later times. The lack of a power gap between cloud scale and mesoscale renders the conceptual basis of cumulus parameterization ill-posed for smaller grid spacings (Cotton and Anthes 1989).

Cumulus parameterization schemes generally contain convective initiation, a closure assumption and a cloud model. Different approaches to these three elements form different parameterization schemes. For example, seven cumulus parameterization schemes are available with MM5 (refer to Grell et al. 1994 for descriptions of individual schemes). Here we choose two convective adjustment methods that do not explicitly formulate the convective process [the Anthes-Kuo scheme ("KUO") and the Betts-Miller scheme ("BM")] and three mass flux methods [the original Kain-Fritch scheme ("KF"), the revised Kain-Fritsch scheme with shallow convection ("KF2"), and the Grell scheme] that include a cloud model to directly simulate the convective process.

1) THE USE OF A SINGLE BUT WRONG CUMULUS PARAMETERIZATION (SINGLE-SCHEME ENSEMBLE)

Four experiments named "KUOens", "KFens", "KF2ens", and "BMens" are executed to evaluate the EnKF performance with the use of a single wrong cumulus parameterization scheme in the ensemble forecast ("wrong" implies a difference from rather than inferiority to the scheme used for the truth). In these experiments, the truth simulation is generated using the Grell scheme (as in "CNTL"), but each ensemble

forecast for the EnKF uses one of the four different cumulus parameterization schemes [i.e., KUO, KF, KF2, and BM (Table 3.1)]. Because different physical parameterizations rely on significantly different underlying assumptions, the use of any scheme in the ensemble other than that used to generate the truth simulation will unavoidably incur model error if moist convection is present. This is also true when using any single-scheme ensemble to assimilate real-world observations.

To simplify subsequent discussions, we define 'bias' as the difference between the ensemble means of the reference ensemble forecast with the perfect physics and the one with imperfect scheme(s) in terms of root-mean difference total energy (RM-DTE, defined in Eq. 3.1). The biases of the four ensemble means (Figs. 3.6a and 3.6c) are found to be significantly different from each other (the mean sampling error in bias estimation is less than 0.1 K for temperature and less than 0.2 m s^{-1} for U and V). "KFens" (dashed black in Fig. 3.6a) and "KUOens" (dashed gray) respectively have the smallest and largest biases, while those of "BMens" and "KF2ens" (not shown) are between the two extremes. This suggests that the Kain-Fritsch and Grell (truth) schemes, which are significantly different from the other two convective adjustment schemes, perform similarly to one another in the winter season. Also, the magnitude of their bias is different throughout the column with two primary vertical peaks (dashed black and gray lines in Fig. 3.6c). These peaks are located at around 850 hPa and 300 hPa and are associated with moist convection and upper level fronts, respectively.

The spectral analysis of bias for the above experiments indicates that the power of the bias is higher at large scales and noticeably differences are observed among different

schemes (not shown). For example, “BMens” exhibits a similar bias to “KUOens” at large scales but has a relatively smaller bias at smaller scales. The bias of “KFens” is consistently smaller at all scales than that in “KUOens” and “BMens”. Moreover, differences are also observed in the domain-averaged reference ensemble spread at 36 h (Fig. 3.6b), with the smallest spread in “KUOens” due to its smaller spread at lower levels (dashed gray line in Fig. 3.6d). The aforementioned differences in the error growth structure will have significant impacts on the performance of the EnKF.

Figures 3.7 and 3.4a demonstrate degraded EnKF performance in the single wrong cumulus parameterization experiments (their ~50% error reduction is significantly less than the 73% error reduction in “CNTL”). The decreased performance is possibly a result of the worsened error covariance structure and bias of the ensemble mean. In general, the larger the mean bias of the reference forecast (model error) or the smaller the ensemble spread, the larger the degradation of the EnKF performance. This is demonstrated among the four single-scheme experiments, where “KUOens” shows the least improvement (46%), while “KFens”, “KF2ens”, and “BMens” show error reductions of 52%, 48% and 55%, respectively (black bins in Fig. 3.7a). Similarly, the absolute analysis error measured in terms of the domain-averaged RM-DTE after the 24-h EnKF assimilation is 2.8, 2.0, 2.1, and 2.3 m s^{-1} for “KUOens”, “BMens”, “KFens”, and “KF2ens” (black bins in Fig. 3.7b), respectively. This analysis error is comparable in magnitude to the observational error specified. In addition, most of the error reduction comes from larger scales and the maximum error decrease is obtained in the lower troposphere (Fig. 3.4d).

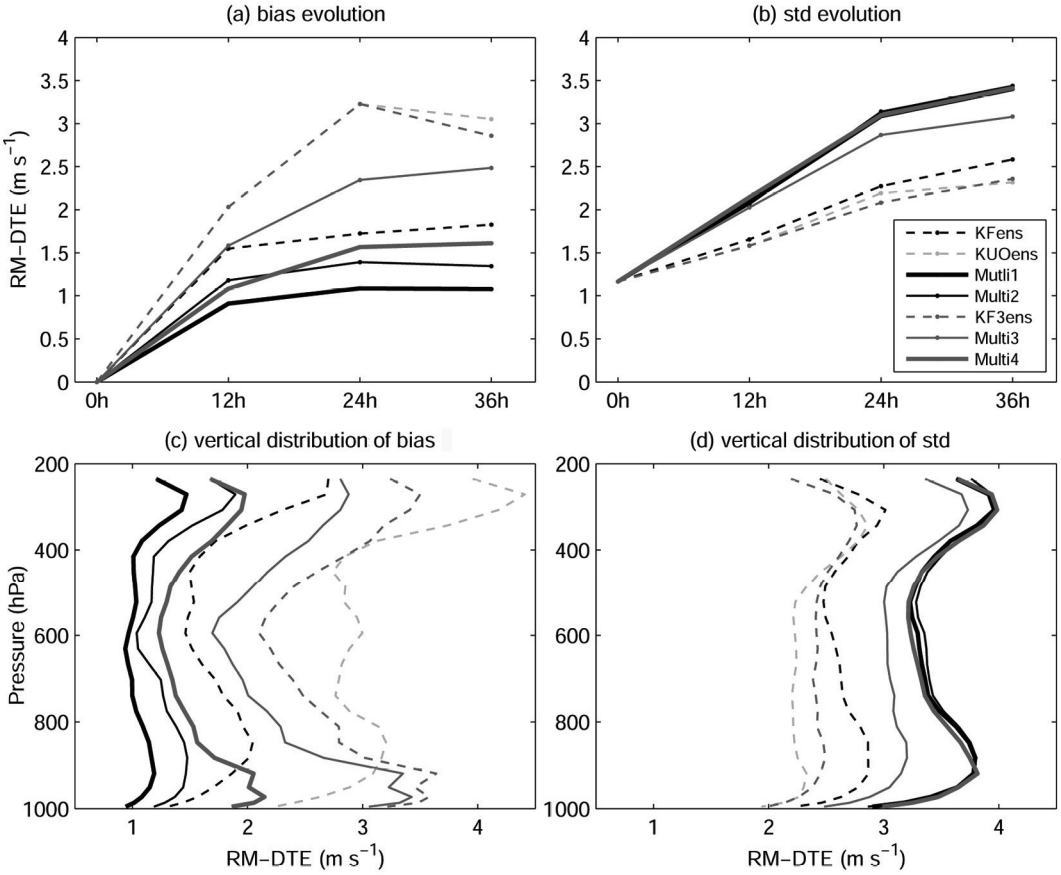

FIG. 3.6. Time evolution of (a) the bias (the root-mean-square difference between the imperfect experiments'reference ensemble mean and the "CNTL" reference ensemble mean) in terms of RM-DTE and (b) the corresponding reference ensemble spreads (standard deviation or std) of RM-DTE for the snowstorm case. (c-d) are as in (a-b) but for the vertical distribution at 36 h. The dashed lines denote one-scheme ensembles with black for "KFens", gray for "KUOens", and dark-gray for "KF3ens". The solid lines represent multi-scheme ensembles including "Multi1" (thick black), "Multi2" (thin black), "Multi3" (thin dark-gray), and "Multi4" (thick dark-gray).

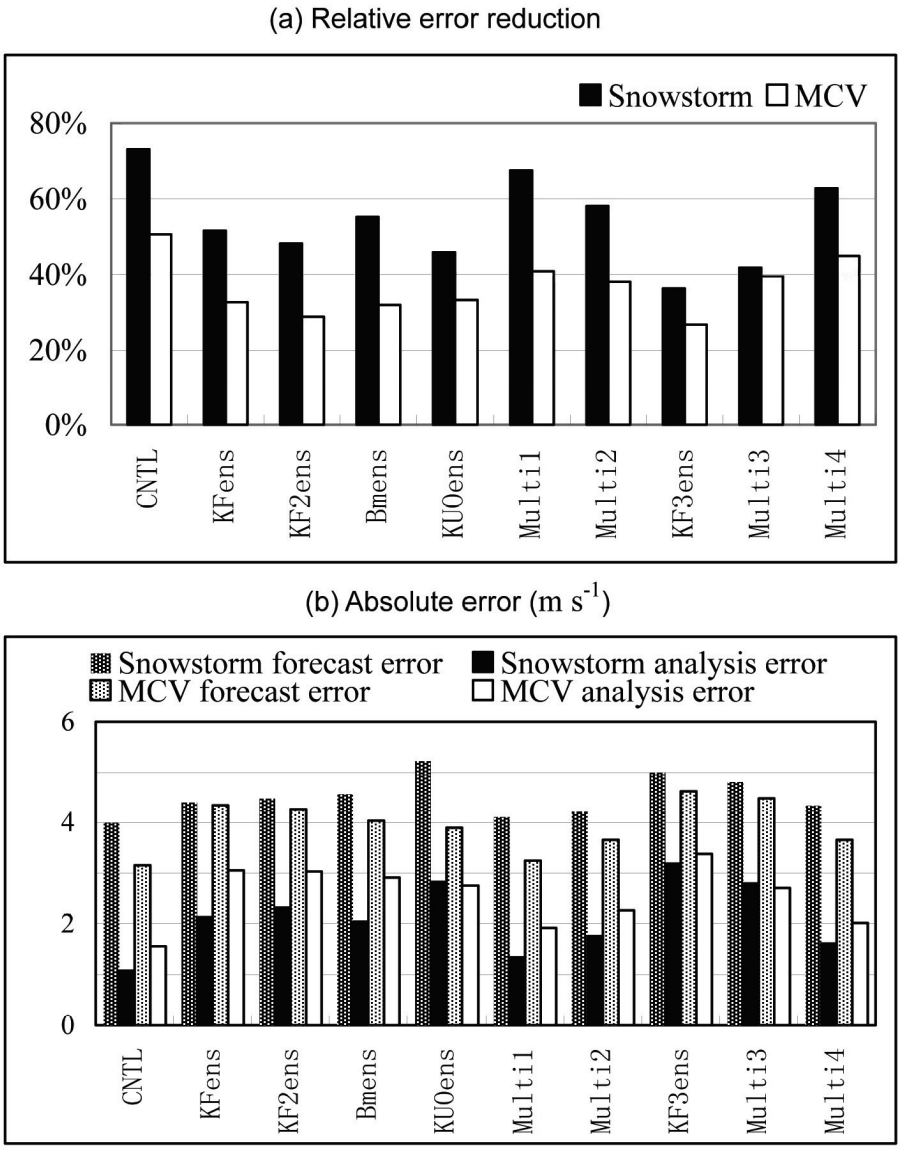

FIG. 3.7. (a) Relative error reduction and (b) absolute forecast/analysis errors (m s-1) in terms of domain-averaged RM-DTE at the final analysis time for the snowstorm case at 36 h (black bins) and the MCV case at 48 h (white bins). The experiments are labeled on the x-coordinate.

2) THE USE OF MULTIPLE CUMULUS PARAMETERIZATION SCHEMES (MULTI-SCHEME ENSEMBLE)

In practice, it is hard to determine *a priori* which cumulus parameterization scheme is the most suitable to predict certain kinds of weather systems in different flow regimes. For example, the above single-scheme experiments demonstrate that model error due to the use of a single wrong cumulus scheme can degrade the EnKF performance to different degrees. Also, Wang and Seaman (1997) compared the performance of four different cumulus parameterizations (i.e., the KUO, BM, KF and Grell schemes) in MM5 and showed that none of them demonstrates consistently better results than others.

A very natural treatment to account for model error from cumulus parameterization is thus to integrate an ensemble using a combination of different cumulus parameterization schemes (Stensrud, et al. 2000; Grell and Devenyi 2002). Through the use of different closure assumptions, cloud models and convection triggering mechanisms, a multi-scheme ensemble may provide a better estimate of the background error covariance by including both initial condition and model uncertainties. In this context, experiment "Multi1" (Table 3.1) is constructed by adopting four different cumulus parameterization schemes including the Grell, KF, BM and KUO schemes in the ensemble forecast (which implies that part of the cumulus parameterizations used in the multi-scheme ensemble are perfect). These four schemes are each assigned to a 10-member subset of the 40-member ensemble. Our use of a multi-scheme ensemble was motivated by a recent study using real-data EnKF experiments (Fujita et al. 2005).

The reference ensemble forecast of "Multi1" shown in the solid thick black lines of

Fig. 3.6 has significantly smaller bias (solid thick black line in Fig. 3.6a) and bigger spread (solid thick black line in Fig. 3.6b) at each level than do any of the single-scheme ensembles (Figs. 3.6c-d). As expected, the multi-scheme ensemble contributes to larger error reduction than do the single-wrong-scheme ensembles in the EnKF data assimilation. The domain-averaged RM-DTE and the vertical distribution of horizontally-averaged RM-DTE after the 24-h data assimilation are plotted in Figs. 3.4b and 3.4e (thin gray lines). For direct comparison, "KFens" (which has average performance) is repeated here to represent the single-wrong-scheme experiments. Compared to the 52% improvement in "KFens", nearly 67% error reduction is achieved in "Multi1". The 1.3 m s^{-1} RM-DTE in "Multi1" is also smaller than any of the single-wrong-scheme experiments (Fig. 3.7b). Again, the largest improvement occurs in the lower troposphere (Fig. 3.4e).

Because a quarter of the ensemble members in "Multi1" still use a perfect (the Grell) scheme, which is unrealistic, experiment "Multi2" replaces the Grell scheme in "Multi1" with the KF2 so that all cumulus schemes used in the ensemble are different from the truth (and thus imperfect, see Table 3.1). The reference ensemble forecast bias in "Multi2" (solid thin black line in Figs. 3.6a and 3.6c) is systematically larger than that in "Multi1" but smaller than the bias in "KFens". The relative error reduction in "Multi2" is about 58% and its absolute RM-DTE is 1.8 m s^{-1} at 36 h. Though it reduces error less than "Multi1", "Multi2" systematically outperforms any of the single-scheme experiments (Fig. 3.7a). Compared to "KFens" (dashed black line in Fig. 3.5a), most of the improvement in "Multi2" comes from larger scales (solid thin black line in Fig.

3.5a). The horizontal distribution of column-averaged RM-DTE also shows consistent improvement over "KFens", and the greatest error reduction is in the vicinity of the surface cyclone (Figs. 3.8a-c).

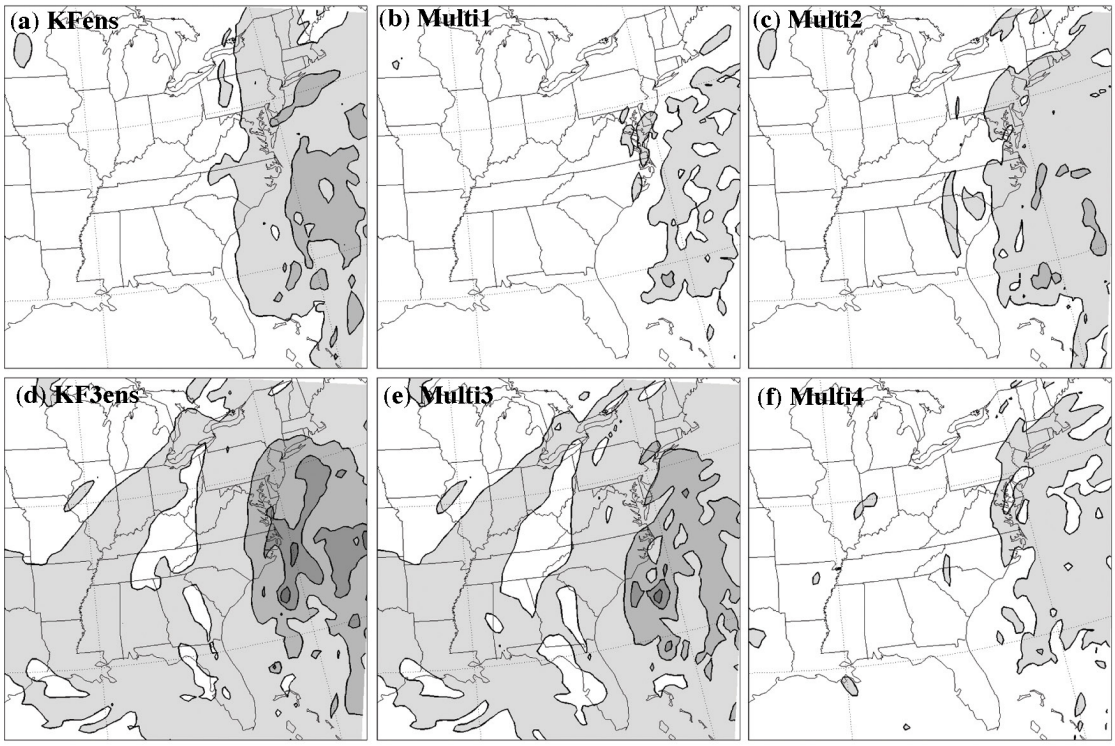

FIG. 3.8. Horizontal distribution of column-averaged RM-DTE (every 2 m s-1) at 36 h for the snowstorm case for (a) "KFens", (b) "Multi1", (c) "Multi2", (d) "KF3ens", (e) "Multi3", and (f) "Multi4", respectively.

b. Impact of cumulus parameterization under imperfect PBL and microphysics schemes

Not only does forecast error come from cumulus parameterization, but it also comes from parameterization of other subgrid-scale processes such as microphysics and planetary boundary layer ("PBL") processes. This subsection explores the impact of model error from cumulus parameterization with imperfect PBL and microphysics

schemes.

To account for the possibility of error from parameterization of multiple subgrid-scale processes, the ensemble in experiment "KF3ens" uses all imperfect schemes including the KF cumulus scheme, the MRF PBL scheme and the Goddard microphysics scheme with Graupel. This ensemble performs significantly worse than any aforementioned experiment and exhibits relative error reduction of only 36% and absolute analysis error of 3.2 m s^{-1} (thin black lines in Figs. 3.4c, f and 3.7). With additional model error from PBL and cloud microphysics, the reference ensemble of "KF3ens" has a large bias but a small spread (the largest bias is in the lower levels among all experiments as shown in dashed dark-gray line in Fig. 3.6).

Experiment "Multi3" expands on "KF3ens" by using the same combination of four (imperfect) cumulus parameterization schemes (i.e., KF, KF2, BM and KUO) as "Multi2" and the same imperfect PBL and microphysics schemes as "KF3ens" (Table 3.1). It is found that in the presence of model error from PBL and microphysics parameterizations, the use of the multiple-cumulus-scheme ensemble also helps to decrease the bias and increase the spread significantly at all levels (solid thin dark-gray line in Fig. 3.6) compared to "KF3ens". Consequently, the EnKF performs better in "Multi3" than in "KF3ens" by reducing the relative error by 42% and the absolute analysis error to 2.8 m s^{-1} at 36 h (Figs. 3.7, 3.4c). Also, most of the improvement occurs at large scales (solid thin dark-gray line in Fig. 3.5a) and at middle to upper levels (solid thin dark-gray line in Fig. 3.4f).

Experiment "Multi4" accounts for the possibility that some schemes may be nearly

perfect under certain flow regimes since all parameterization schemes are developed to represent real physical processes. To do this, “Multi4” uses a combination of different cumulus, PBL and microphysics schemes, each of which includes some of the same schemes as in the truth (Table 3.2). Specifically, each 10-member subset of “Multi1” is further divided into four subsets. Among the 10 members of each subset, five use the Reisner-graupel microphysics scheme while the other five adopt the GSFC-graupel scheme. The five-member subsets using the Reisner-graupel scheme are further divided into two groups of 3 and 2 members employing the ETA and MRF PBL schemes, respectively. The other five members with the GSFC-graupel scheme are treated similarly except that the two PBL schemes are switched between the 3- and 2-member groups. This particular configuration is used to make sure that any of the three categories of the physical parameterization schemes are evenly distributed among the 40 ensemble members.

The reference ensemble forecast (without the EnKF assimilation) of “Multi4” (solid thick dark-gray line in Fig. 3.6) has smaller bias and larger spread than those of both “KF3ens” and “Multi3” during the whole integration period. Figure 3.7 also shows that “Multi4” performs better than nearly all other imperfect-model experiments (except “Multi1”, which also includes the same schemes as in the truth). The relative error reduction for “Multi4” is 63%, and absolute analysis error of 1.6 m s^{-1} is observed in this experiment. This reduction is evident in both the domain average (Figs. 3.8d-f) and vertical distribution (solid thin gray line in Fig. 3.4f) of RM-DTE. Though they might be optimistic, experiments “Multi1” and “Multi4” exemplify a situation when part of the

parameterization schemes happens to perfectly describe the real physical process.

The large differences observed between "KFens" and "KF3ens" and between "Multi2" and "Multi3" demonstrate that the use of imperfect PBL and microphysics schemes (in addition to imperfect cumulus parameterizations) can significantly degrade the EnKF performance (Fig. 3.5a). However, due to the limited availability of microphysics and PBL parameterization schemes in MM5, we can not examine the impact of using multi-scheme ensembles in which none of the schemes in PBL or microphysics parameterizations is perfect (this is partially due to limited choices of usable PBL or microphysics schemes in MM5).

c. Comparison of error covariance between single- and multi-scheme ensembles

This subsection further investigates the reasons why the EnKF performs better with a multi-scheme ensemble than with a single-wrong-scheme ensemble. For example, the previous subsections showed that while the EnKF is quite effective at reducing the analysis error in the presence of significant model uncertainties, the analysis error in the imperfect-model experiments is noticeably larger than that of "CNTL". This indicates that the EnKF performance can be degraded to different extents with different physical parameterizations (Fig. 3.7). Such difference in the EnKF performance might be due to the ensemble mean error ("bias") and/or insufficient ensemble spread resulting from the use of an imperfect model.

The horizontal distributions in Figs. 3.9a,b show that the reference ensemble forecast of "Multi2" has a significantly larger standard deviation of column-averaged

RM-DTE than does "KFens" at 24 h. Similar large-scale, balanced features were observed evolving from initially uncorrelated, small-scale, unbalanced errors in a period of 12-24 h in a previous study of this snowstorm (Zhang 2005). The maximum error growth in the disturbances is associated with the upper trough and the surface cyclone and is collocated with the strongest PV gradient. The spectral analysis of the ensemble spread also shows a much larger difference between "Multi2" and "KFens" at larger scales (i.e, wavenumber <10 or wavelength >240 km) than at smaller scales (not shown). The differences between balanced disturbances of "Multi2" and "KFens" have implications when using the EnKF because the EnKF is most effective at correcting errors at larger scales (as shown in Chapter II).

To further illustrate the differences between the large-scale error structures of "KFens" and "Multi2", the cross-covariance between U and T at 300 hPa at 24 h is also examined for each ensemble (Figs. 3.9c, d). While "Multi2" and "KFens" exhibit similar covariance structures with increased covariance in the vicinity of strong PV, the magnitude of the covariance in "Multi2" is noticeably larger due to its relatively larger ensemble spread (Figs. 3.9a, b). When the ensemble spread is significantly smaller than the error of the ensemble mean, increase of the ensemble spread could improve the performance of the EnKF. A larger spread in the multi-scheme ensembles may increase the likelihood of keeping the truth within the uncertainties spanned by the imperfect ensemble, and a large covariance has the potential to propagate observational information more efficiently between variables. This is consistent with Fujita et al. (2005), a recent real-data study that partially motivated the use of multi-scheme

ensembles in the current study.

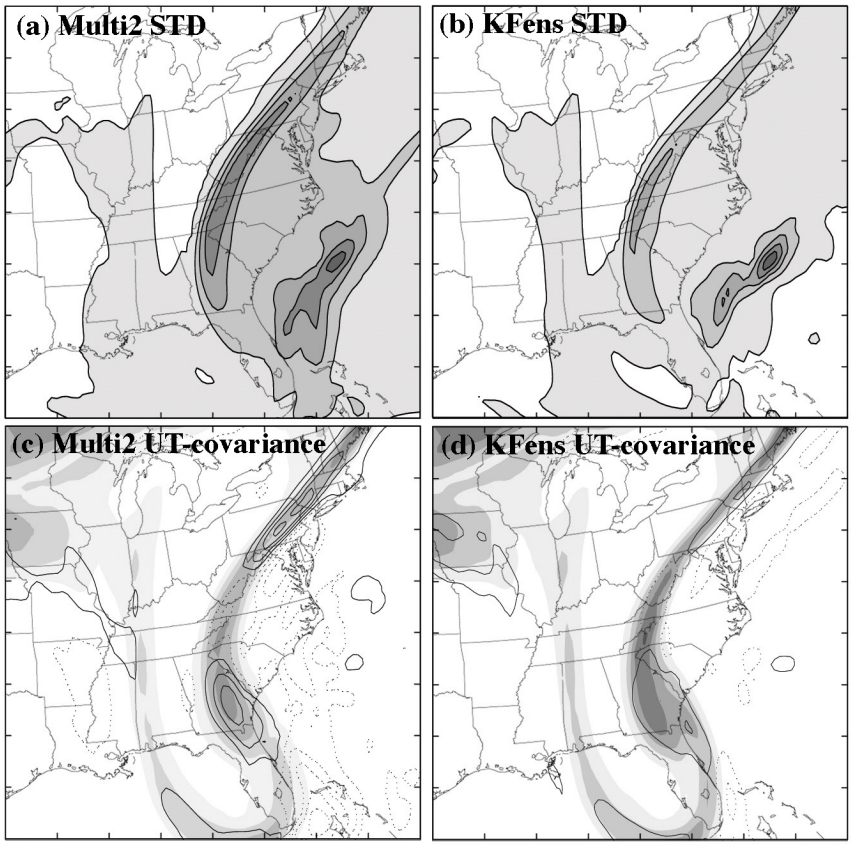

FIG. 3.9. Horizontal distribution of the standard deviation of column-averaged RM-DTE (every 2 m s-1) for (a) "Multi2" and (b) "KFens" at 24 h for the snowstorm case. (c-d) are as (a-b) but for the covariance between U and T on 300hPa (every 2 K m s-1; negative, dotted). The shading in (c-d) is PV at 300hPa every 1 PVU.

In order to understand whether or not the covariance structure developed in one of the above ensembles (i.e., "KFens" or "Multi2") is systematically better than the covariance structure of the other, four "static" EnKF experiments ("Pmulti2-Mmulti2", "Pkf-Mmulti2", "Pmulti2-Mkf", "Pkf-Mkf") are conducted. These experiments are "static" in the sense that observations are assimilated at only one selected time without subsequent forecast and analysis cycles. The naming convention is as follows: "M…"

refers to the reference ensemble mean, "P…" refers to perturbations/deviations from the mean and '…' refers to the experiments in previous subsections. For example, "Pmulti2-Mmulti2" and "Pkf-Mkf" use the (unaltered) reference ensemble forecast of "Multi2" and "KFens", respectively, to estimate the background error covariance for the EnKF. "Pkf-Mmulti2" and "Pmulti2-Mkf" are performed by switching the ensemble means of "Multi2" and "KFens" so that the perturbations of "Multi2" are added to the mean of "KFens", and the perturbations of "KFens' are added to the mean of "Multi2". Because any two experiments formed using the same ensemble mean have the same forecast error (e.g., "Pkf-Mkf" and "PMulti2-Mkf"), the quality of the covariance structure associated with each ensemble can be ascertained by the differences in error between the same two experiments after the assimilation cycle (i.e., the analysis error).

The results in Table 3.3 show that a systematically smaller analysis error can be achieved by using the background error covariance estimated from the multi-scheme ensemble ("Multi2") rather than the single-wrong-scheme ensemble ("KFens"). Similar results are also obtained for "KF3ens" and "Multi3" (see Table 3.3) and for different reference forecast times (not shown). Using a multi-scheme ensemble is also found to be beneficial in a warm-season MCV event for both the continuously evolving and static EnKF assimilation experiments (detailed in section 5).

While "KFens" also has the problem that its ensemble spread (solid thin gray line in Fig. 3.10a) is noticeably smaller than its analysis error (solid thick gray line in Fig. 3.10a), the potential for filter divergence with this ensemble may be alleviated with covariance inflation. Experiment "KFens_0.7" is conducted by changing the weighting

coefficient α in the relaxation method (Zhang et al. 2004; their Eq. 5) from 0.5 to 0.7 to give more weight to prior perturbations. The use of a larger weight for the prior estimate as an alternative for covariance inflation (e.g., Anderson 2001) consequently leads to systematically larger ensemble spreads (though still insufficient, solid thin black line in Fig. 3.10a) and slightly improved the EnKF performance over 24 hours of data assimilation (solid thick black line in Fig. 3.10a).

TABLE 3.3. Domain-averaged RM-DTE for one-time data assimilation experiments valid at 36 (48) h for the snowstorm (MCV) case which switch perturbations between the single scheme "KFens" and the multi-scheme EnKF experiments. EF means the reference ensemble forecast.

Ensemble mean	**Experiments**	**RM-DTE (m s^{-1})**	
		Snowstorm	**MCV**
Multi2	EF of Multi2	4.22	3.66
	Pmulti2-Mmulti2	2.50	2.43
	Pkf-Mmulti2	2.64	2.62
KFens	EF of KFens	4.39	4.34
	Pmulti2-Mkf	2.49	2.47
	Pkf-Mkf	2.79	3.08
Multi3	EF of Multi3	4.80	4.47
	Pmulti3-Mmulti3	3.12	2.85
	Pkf3-Mmulti3	3.31	3.09
KF3ens	EF of KFens	5.00	4.61
	Pmulti3-Mkf3	3.10	2.86
	Pkf3-Mkf3	3.44	3.33

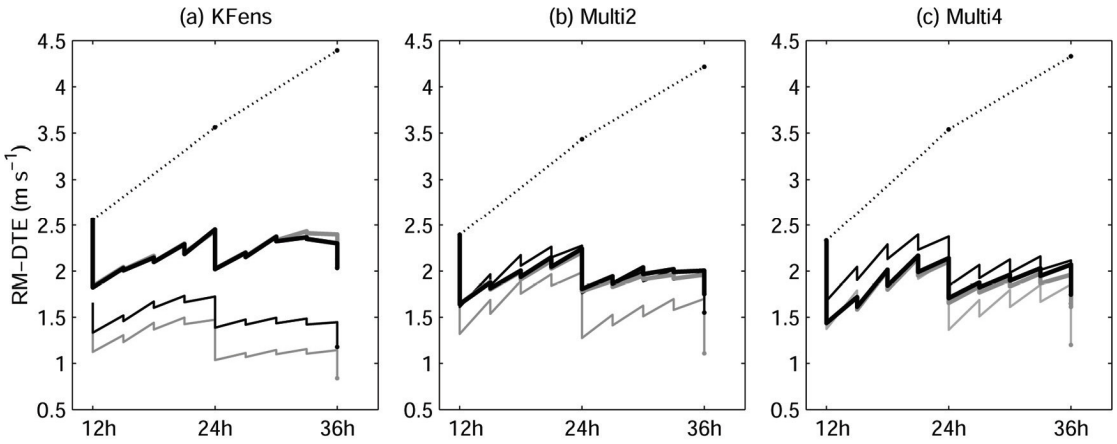

FIG. 3.10. The domain-averaged RM-DTE (thick solid lines) and analysis ensemble spread of RM-DTE (thin solid lines) with different weights (α) of prior perturbations in the covariance inflation (mixing) method for experiments (a) "KFens", (b) "Multi2", and (c) "Multi4" for the snowstorm case. The black lines are for $\alpha = 0.7$, gray lines for $\alpha = 0.5$. The reference ensemble forecast errors are also plotted in dotted lines.

When covariance inflation is applied to other ensembles for which the ensemble spread is not too small, the results are worsened somewhat. For example, when the relaxation coefficient in "Multi2" is modified from 0.5 to 0.7 in experiment "Multi2_0.7", the analysis ensemble spread (solid thin black line in Fig. 3.10b) quickly becomes larger than the analysis error ("over-inflation") and the EnKF performance worsens (solid thick black line in Fig. 3.10b). The ensemble spread eventually gets closer to or slightly smaller than the error and draws the analysis error back to that of "Multi2" at 36 h. This negative impact of "over-inflation" is more apparent when the relaxation coefficient changes from 0.5 to 0.7 in "Multi4" (Fig. 3.10c) because the initial spread is already comparable to the error. The larger ensemble spread results in consistently larger errors during the whole period.

d. Other experiments

Various experiments using the conventional covariance inflation of Anderson (2001) and additive error method of Hamill and Whitaker (2005) are also performed to account for model error from physical parameterizations. None of these experiments with different covariance inflation factors or different additive error gives acceptable EnKF performance (not shown). The traditional inflation leads to spuriously large ensemble spread in data sparse areas. For the additive error experiments, the additive error covariance sampled from the differences between different cumulus parameterization schemes (at different times) fails to increase the ensemble spread in desired regions where there is active parameterized convection at analysis times. This result is in strong contrast to the success of using similar additive error methods to account for model truncation error (Hamill and Whitaker 2005) that is likely to be less flow-dependent.

5. Impact of flow-dependent error growth dynamics

In this section, we investigate the performance of the EnKF for a vastly different flow regime than in previous sections. Since weather systems under different flow regimes may have different error growth dynamics and mesoscale predictability, and the EnKF performance is significantly scale- and dynamic-dependent (Chapter II), the EnKF is likely to behave differently in different regimes. The particular case examined is a long-lived warm-season MCV event that occurred on 10-13 June 2003. A recent study (Hawblitzel, et al. 2006) shows that the predictability of this MCV event is very limited due to its extreme sensitivity to convection. This result is not surprising given that past

studies (e.g., Wang and Seaman 1997; Zhang et al. 2006b) suggest that model error, especially that from cumulus parameterization, can be more detrimental to warm-season forecasts than to winter events.

a. Overview of the MCV event and the EnKF configuration

This MCV event occurred during an intense observation period (IOP8) of the Bow Echo and Mesoscale Convective Vortex Experiment (BAMEX) conducted from 18 May to 7 July 2003 over the central United States. At 0000 UTC 10 June, a disturbance embedded in the subtropical jet triggered convection over eastern New Mexico and western Texas. An MCV developed from the remnants of this convection over central Okalahoma at 0600 UTC 11 June, and matured by 1800 UTC 11 June as it traveled northeastward to Missouri and Arkansas. The MCV transitioned into an extratropical baroclinic system after 0000 UTC 12 June.

The EnKF configuration is the same as for the winter snowstorm event except that a 15-point (450-km) radius of influence is used here due to the relatively smaller scale of the weather system. The assimilated data and the updated grid points are constrained to lie within the solid box of Fig. 3.1. Because of the longevity of the MCV, a 36-h data assimilation is performed from 1200 UTC 10 June to 0000 UTC 12 June. The assimilation follows a 12-h ensemble forecast that starts at 0000 UTC 10 June. Employing the same method used for the winter case, synthetic soundings are assimilated at 12-h intervals and synthetic surface observations are assimilated every 3 hours. The ensemble member with the 48-h forecast being closest to the observed MCV

is adopted as the truth from which the observations are extracted (Fig. 3.11).

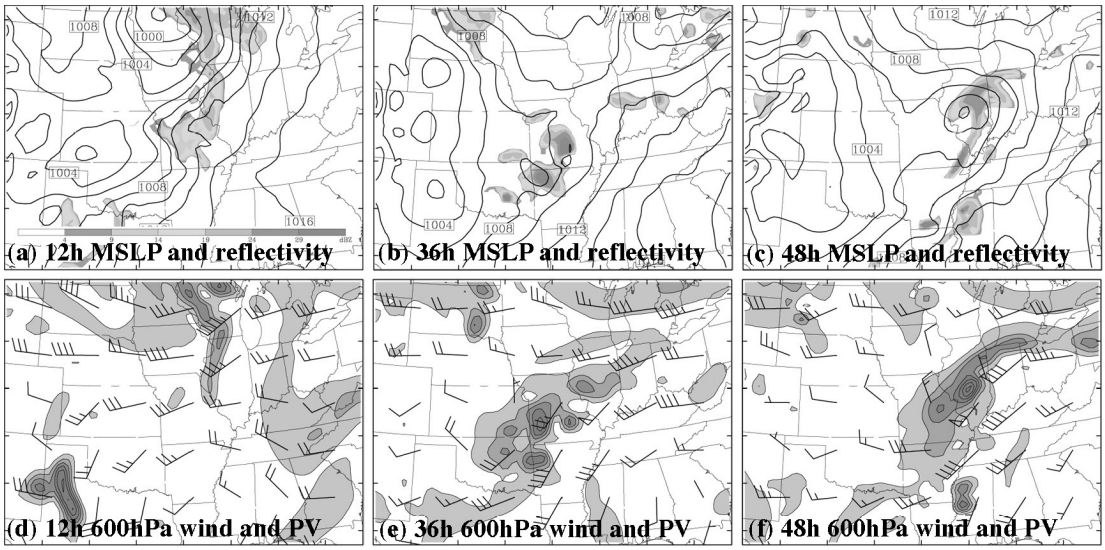

FIG. 3.11. The MSLP (every 2 hPa) and simulated reflectivity (shaded) valid at (a) 12 h, (b) 36 h, and (c) 48 h and the potential vorticity (every 1 PVU) and wind vectors (full barb 5 m s-1) at 600 hPa valid at (d) 12 h, (e) 36 h, and (f) 48 h from the truth simulation for the MCV case.

b. The control EnKF experiment for the MCV event

The control experiment for this MCV event, which is also conducted under a perfect model assumption using the Grell scheme (as in the snowstorm simulation), reveals that the largest errors are strongly associated with the MCV dynamics. The reference ensemble forecast error in terms of both the MSLP and the surface wind at 12 and 48 h and the column-averaged RM-DTE are shown in Fig. 3.12. Comparison of Fig. 3.12 with Fig. 3.3 reveals that the overall error amplitude in this MCV event at 36 h (as well as 48 h) is significantly smaller than that in the snowstorm event. Spectral analysis of the reference ensemble forecast error shows that the MCV event has smaller error at all scales, especially at large scale (dotted line in Fig. 3.5b) compared to the snowstorm event (dotted line in Fig. 3.5a). The smaller-scale error in

the MCV event initially grows faster and quickly saturates while the larger-scale error grows slowly.

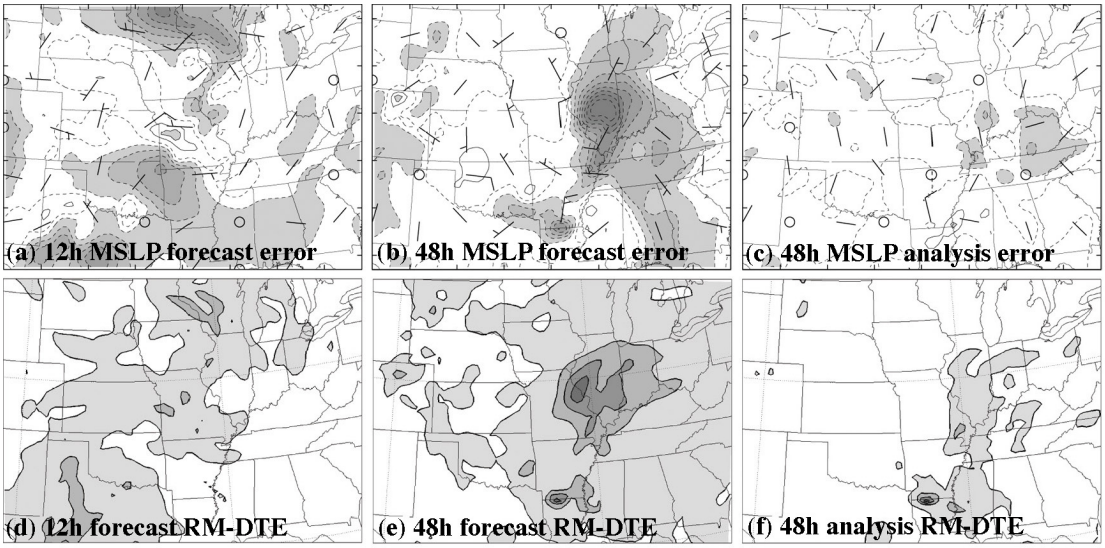

FIG. 3.12. Forecast errors of surface wind vectors (full bard 5 m s-1) and MSLP (every 0.5 hPa) at (a) 12 h and (b) 48 h for the MCV case and (c) analysis error of the same fields at 48 h. (d-f) are as (a-c) but for the column-averaged RM-DTE (every 2 m s-1).

Despite the apparent difference in error, spectral composition, and growth rate between the MCV event and the snowstorm event, the control EnKF ("CNTL") performs reasonably well for the MCV event. After the 36-h data assimilation in "CNTL", the maximum MSLP error is reduced from 4 to 1 hPa while the area of error larger than 0.5 hPa also decreases significantly (Fig. 3.12c). Error reduction in the surface wind field is also apparent as the maximum error value reduces from approximately 7.5 to 5 m s^{-1} (Fig. 3.12c). Significant error reduction is also exhibited in column-averaged RM-DTE for the entire assimilation domain, especially where the MCV is located. Furthermore, the maximum RM-DTE value decreases from 8 to 2 m s^{-1} (Fig. 3.12f). At 600 hPa, the

maximum PV error reduces from 2.5 to 1 PVU and the maximum velocity error decreases from 10 to 2.5 m s^{-1} in the vicinity of the MCV (not shown).

Figure 3.13 shows that the evolution of domain-averaged root-mean-square analysis and forecast error and the analysis ensemble spread for *u, v, T, p', w, q* for the "CNTL" of the MCV event are similar to those of the snowstorm case (see Fig. 2.8). As with the winter case, the ratio of the analysis error to the ensemble spread is very close to 1.0 (except for *p'* and *w*), suggesting no apparent filter divergence for the warm-season event. After the 36-h data assimilation, the relative error reduction of the observed variables *u, v,* and *T* is about 40-60%. Pressure perturbation (*p'*) still has the

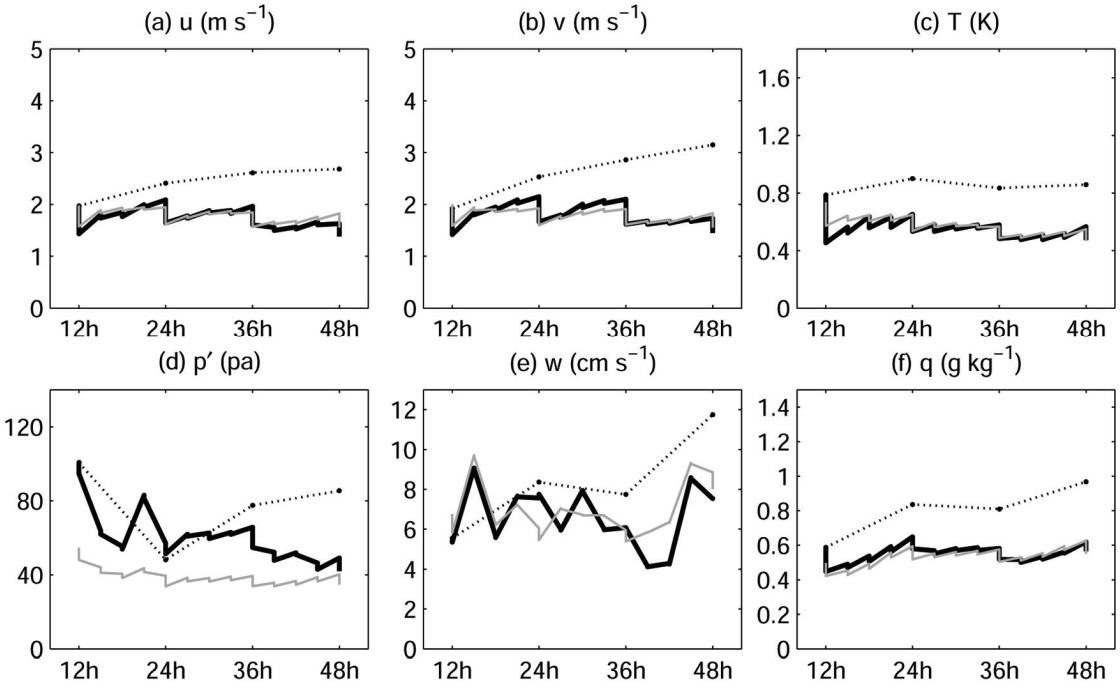

FIG. 3.13. Time evolution of the domain-averaged root-mean-square errors of (a) u, (b) v, (c) T, (d) p', (e) w, and (f) q for the EnKF analysis (solid black) and the reference ensemble forecast (dotted black, computed every 12 h) of "CNTL" in the MCV case. The gray lines are the standard deviation of analysis ensemble.

largest error reduction of about 60%, but its reduction is still less than that with the snowstorm event. Also, about 40% improvement is obtained in the moisture field. Again, the least improvement (about 37%) is observed with vertical velocity. In terms of column-averaged RM-DTE, the overall error reduction at 48 h is about 51% (Fig. 3.7a and thick dark-gray lines in Fig. 3.14a). As with the snowstorm case, most of the error reduction comes from larger scales (solid thick dark-gray line in Fig. 3.5b) and is maximized at lower levels (thick dark-gray lines in Fig. 3.14d). Both the analysis error

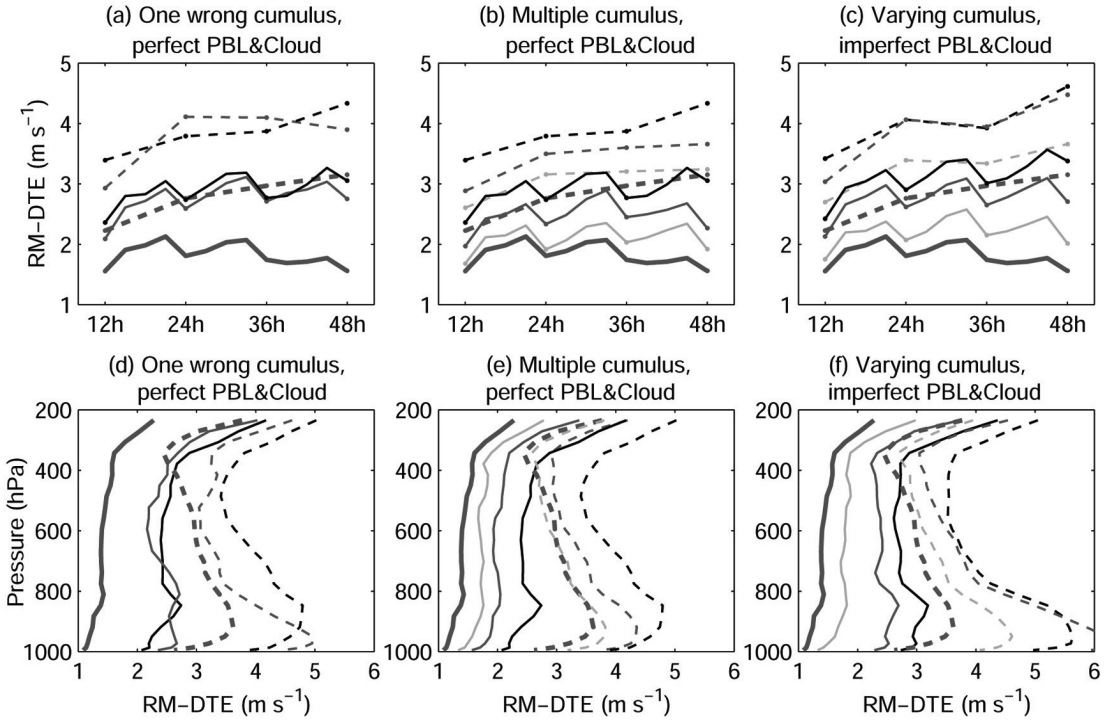

FIG. 3.14. As in Fig. 3.4 but for the MCV case with (d-f) valid at 48 h.

after the control EnKF assimilation (solid thick dark-gray line in Fig. 3.5b) and the reference ensemble forecast error has a flatter spectrum than does the snowstorm event error (solid thick dark-gray line in Fig. 3.5a). Since the EnKF is less effective for small,

marginally resolvable scales (Chapter II), the overall relative error reduction for the MCV event is smaller than that for the snowstorm event.

c. Impact of model error for the MCV event

The difference between the forecast ensemble mean in the control experiment and various sensitivity experiments using different physical parameterization schemes ("bias") in this warm season case evolves differently from that in the winter case (Fig. 3.15 vs. Fig. 3.6). The multiple-cumulus-scheme ensemble biases are much closer to each other than are those in the snowstorm case. The largest bias after 48 hours of integration is observed in "KF2ens" (not shown) and the smallest bias is observed with "KUOens" (dashed gray in Fig. 3.15). The biases of "KFens" (dashed black in Fig. 3.15) and "BMens" (not shown) fall between the two extremes. The ensemble spreads of these experiments are also quite close to each other (dashed lines in Fig. 3.15b). The vertical profiles of the biases (dashed lines in Fig. 3.15c) and spreads (dashed lines in Fig. 3.15d) exhibit a two-peak pattern similar to the winter case. The higher upper peaks in the MCV case than that in winter case are due to the higher tropopause and upper-level fronts in the summer. The 950 hPa bias peaks in the MCV case are at a slightly lower altitude and are stronger in magnitude than the ~900 hPa bias peaks of the snowstorm case (Fig. 3.15b). However, the lower peaks of the ensemble spread of the MCV case are at similar altitudes to those in the snowstorm case (around 900hPa, dashed black and gray lines in Fig. 3.15d and Fig. 3.6d).

When the EnKF is used with the above ensembles, the error reduction is smaller

than with the snowstorm case and the filter performance is very similar among the experiments “KFens”, “BMens”, “KF2ens” and “KUOens” (Figs. 3.7 and 3.14a, d). These similarities are not surprising given the similarity between reference ensembles. One possible culprit for the roughly similar results is the observed fast error saturation.

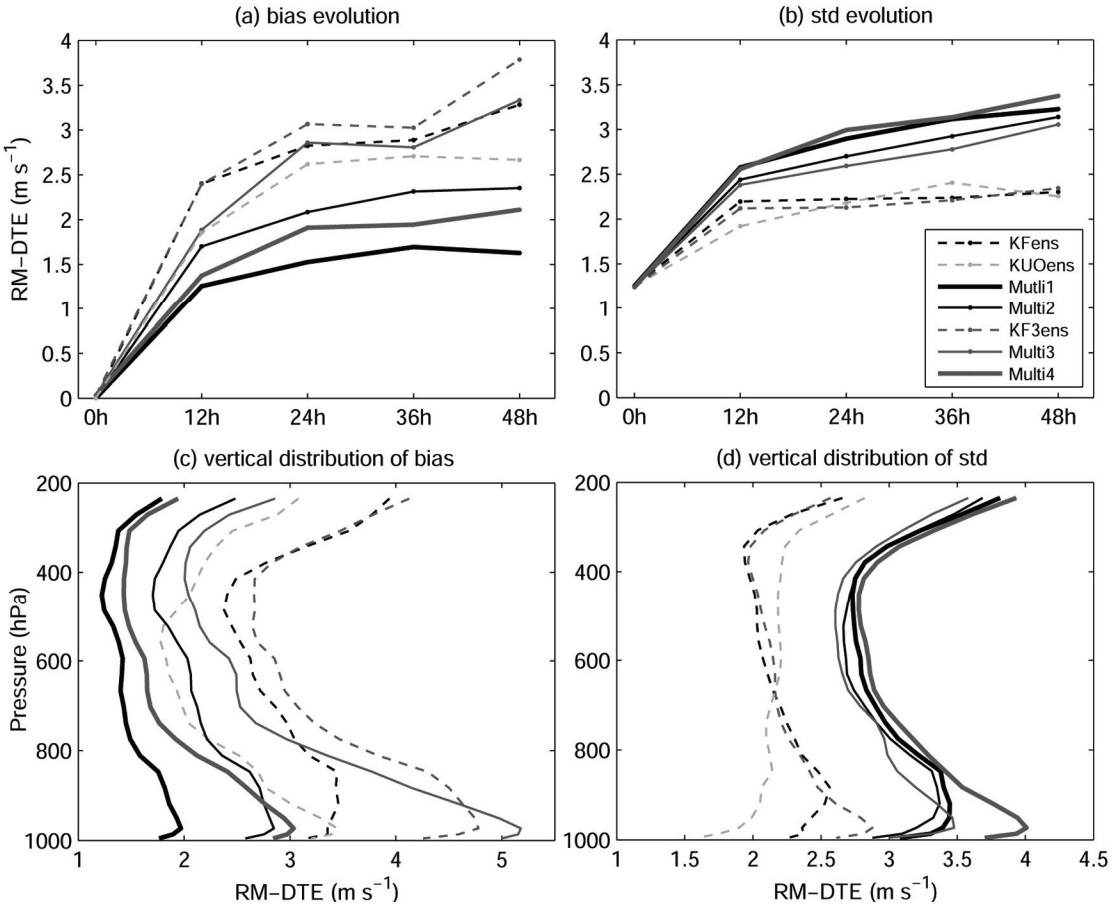

FIG. 3.15. As in Fig. 3.6 but for the MCV case with (c) and (d) valid at 48 h.

As with the snowstorm event, experiments using multi-scheme ensembles for this MCV event show improvement over those using single-scheme ensembles. “Multi1” and “Multi2”, the perfect PBL and microphysics multi-scheme experiments, have smaller

bias and larger spread than “KFens” (Figs. 3.15a, b); this result is similar to that of the snowstorm case. A systematically larger bias is observed for all experiments in the MCV case than in the winter case, and this suggests a larger impact of physical parameterizations for the warm-season event. The covariance between U and T at 600 hPa after a 36-h integration is larger in “Multi2” than in “KFens”, but it is weaker in both experiments when compared to the covariance in the winter case. One possible reason for this is the low predictability of smaller-scale convective activity. After 36-h data assimilation, the relative error reduction for “KFens”,”Multi2” and “Multi1” is 33%, 38% and 41%, respectively, and the absolute error is respectively 3.0, 2.3 and 1.9 m s^{-1} (Figs. 3.14b and 3.7). There is thus consistent improvement when a multi-scheme ensemble is adopted. Power spectrum analysis also shows that the improvement of “Multi2” over “KFens” comes mainly from the large scales (Fig. 3.5b).

Similar improvement in multi-scheme ensembles over single-wrong-scheme ensembles is also observed under imperfect PBL and microphysics parameterizations in “KF3ens”, “Multi3” and “Multi4”. Vertical distribution of the ensemble spread shows that the lower peaks of the spreads of these three experiments are at slightly lower levels and are larger than the corresponding peaks in the snowstorm case. This indicates that PBL processes may have a larger impact on error growth in the MCV than the snowstorm case (Fig. 3.15d vs. Fig. 3.6d). The EnKF result shows significant improvement of “Multi3” over “KF3ens” (Fig. 3.14c) at large scales (Fig. 3.5b) and on each level (Fig. 3.14f), suggesting the multi-cumulus-ensemble can decrease PBL error more than the winter case where very small differences are seen at lower levels between

'KF3ens" and "Multi3". Similarly, "Multi4" consistently reduces error during the whole period at all levels (gray line in Figs. 3.14c and f).

Experiments in this MCV event further demonstrate that a multi-scheme ensemble is capable of providing better estimation of the background error covariance than a single-wrong-scheme ensemble. The significance of improving error covariance by using a multi-scheme ensemble is also demonstrated through static EnKF experiments by switching the means of the reference ensemble forecast for "KFens" and "Multi2" and for "KF3ens" and "Multi3" (Table 3.3) in a similar way to that discussed in section 4c for the snowstorm event.

6. Conclusions and discussions

Through various observing system simulation experiments, the performance of an ensemble Kalman filter is explored in the presence of significant model error caused by physical parameterization. The EnKF is implemented in the mesoscale model MM5 to assimilate synthetic sounding and surface data derived from the truth simulations at typical temporal and spatial resolutions for the cold-season snowstorm event that occurred on 24-26 January 2000 and the warm-season MCV event that occurred on 10-13 June 2003.

Results show that although the performance of the EnKF is degraded by different degrees when a perfect model is not used, the EnKF can work fairly well in different kinds of imperfect scenario experiments. A 36-67% overall relative error reduction (improvement over the reference ensemble forecast) is found in each imperfect scenario

for the snowstorm event. In both the perfect and imperfect scenarios, most of the error reduction comes from larger scales and it is maximized in the lower troposphere.

The performance of the EnKF was tested and found to be very sensitive to model error introduced by different cumulus parameterizations. Sensitivity experiments herein used ensembles with either single or multiple imperfect cumulus parameterizations with and without model error from PBL and microphysics. The results demonstrate that using a combination of different cumulus parameterization schemes can significantly improve the EnKF performance over experiments using a single inaccurate parameterization scheme. Our results suggest that the improvement comes from a smaller bias and from a better background error covariance structure developed from the multi-scheme ensemble. This is consistent with a recent real-data EnKF study of Fujita et al. (2005). Model uncertainties from PBL and microphysics processes also have significant impacts on the EnKF performance.

The EnKF performance depends strongly on the scales and dynamics of the flow of interest. Comparison of the EnKF performance in the two events with distinguishably different flow regimes exemplifies the impacts of flow-dependent predictability. It is found that the EnKF behaves consistently in corresponding experiments examining the two events, but the relative error reduction over the reference ensemble forecast is 10-15% smaller in the warm season event. The growth of reference ensemble forecast error is much slower in the MCV event than in the snowstorm case. Slower error growth and the relatively smaller scale of the MCV circulation may be responsible for a smaller error reduction and also for less bias when using different cumulus parameterization

schemes in the ensemble forecast (Fig. 3.7). Impact of PBL and microphysics processes seems to be more significant for the warm-season case than for the winter case.

As a pretest for assimilating real data, this study is aimed at examining the impact of model error on an ensemble-based mesoscale data assimilation system. Apart from the errors explored here, there are other sources of uncertainty such as those from ensemble initialization, truncation error, lateral boundary and surface processes. In real data assimilation, model error could potentially be more detrimental than considered in this study. We not only need to understand the impact of various model errors on the EnKF, but we also need to design effective ways to treat them such as with parameterization of model error (e.g., Hamill and Whitaker 2005) and simultaneous estimation of parametric model error (e.g., Aksoy et al. 2006a,b).

CHAPTER IV

COMPARISON WITH 3DVAR IN A REAL-DATA CASE STUDY

1. Introduction

Since the first application of an ensemble Kalman filter (EnKF) (Evensen 1994) in the atmospheric sciences field (Houtekamer and Mitchell 1998), the EnKF has been widely examined with different models at different scales and to different realistic extents (Houtekamer and Mitchell 1998; Hamill and Snyder 2000; Anderson 2001; Whitaker and Hamill 2002; Mitchell et al. 2002; Snyder and Zhang 2003; Zhang and Anderson 2003; Zhang et al. 2004; Aksoy et al. 2005; Houtekamer et al. 2005; Tong and Xue 2005; Aksoy et al. 2006 a, b; Zhang et al. 2006a; Dirren et al. 2006; Meng and Zhang 2006; Whitaker et al. 2006). See Evensen (2003), Lorenc (2003) and Hamill (2006) for recent reviews. Studies with simulated observations demonstrate its success at decreasing forecast error and its better performance relative to variational data assimilation methods such as 3DVar in a large scale model (Hamill and Whitaker 2005) and 4DVar in a convective scale model (Caya et al. 2005).

One of the difficulties in real-world application of ensemble-based data assimilation techniques is the proper representation of model error (Zhang and Snyder 2006). Recent progress has been made to account for model error by using additive or multiplicative covariance inflation (Hamill and Whitaker 2005; Houtekamer et al. 2005; Barker 2005). For example, encouraging results in real-data applications have been obtained in an ocean general circulation model (Keppenne and Rienecker 2002), and in global

(Whitaker et al. 2004; Houtekamer et al. 2005; Whitaker et al. 2006) and limited-area atmospheric models (Dowell et al. 2004; Barker 2005; Fujita et al. 2005). The performance of the EnKF implemented in global models with real data has been shown to be better than (Whitaker et al. 2004, 2006) or at least comparable to (Houtekamer et al. 2005) the 3DVar method. However, direct comparison between an EnKF and 3DVar in limited-area models (which will a focus of this chapter) has not been seen in the literature.

In Chapter II, the performance of an EnKF implemented in the NCAR/Penn State mesoscale model MM5 was examined with the perfect-model assumption by assimilating synthetic sounding and surface observations with typical temporal and spatial resolutions (Zhang et al. 2006a; hereafter referred to as Chapter II). It was found that the EnKF with 40 members works very effectively in keeping the analysis close to the truth simulation. Most error reduction comes from the large scale, which is consistent with Daley and Menard (1993) (though different mechanisms are involved in their study). Furthermore, the EnKF performs differently for different variables; it is the least effective for vertical motion and moisture due to their relatively strong smaller-scale components, but it is the most effective in reducing the error in pressure (and also very effective, but to a lesser degree, for horizontal winds and temperature) because of its relatively strong larger scale error energy (Chapter II).

Subsequently in Chapter III, the performance of the EnKF was investigated in the presence of model error due to imperfect sub-grid physical parameterization schemes (Meng and Zhang 2006; hereafter referred to as Chapter III). The result shows that the

EnKF still performs reasonably well, though its performance can sometimes be significantly degraded by the presence of model error. It was found that using different combinations of different physical parameterization schemes in different ensemble members can significantly improve filter performance due to the resulting better background error covariance and mean estimation. Different performance of this EnKF system was also observed for different flow regimes (Chapter III).

As a natural extension of these two observing system simulation experiment (OSSE) studies under both perfect and imperfect model assumptions, the current chapter implements the same EnKF in the Weather Research and Forecasting model (WRF) for real-world data assimilation. Its performance is directly compared with the WRF-3DVar data assimilation system, complementary to similar comparisons performed with global models (Houtekamer et al. 2005; Whitaker et al. 2006). In the next section, a brief introduction is given on the methodology, including the model, the EnKF and the 3DVar method. The synoptic overview of the MCV event and the observations to be assimilated are described in section 3. Section 4 compares the performance of the EnKF and the 3DVar method. Sensitivities of both methods to the background error covariance are examined in section 5. A brief summary and discussion are given in section 6.

2. Methodology

a. The mesoscale model

The Advanced Research WRF (ARW) (instead of MM5) is used in this chapter considering that WRF is the next-generation mesoscale numerical weather prediction

system designed to serve both operational forecasting and atmospheric research needs. WRF has been used to provide real-time forecasts to aid the planning of the Bow Echo and MCV EXperiment (BAMEX) in summer of 2003, during which the MCV case examined in this chapter occurred. The ability of WRF to correctly capture the timing, location and mode of mesoscale convective systems in the majority of cases suggests its potential utility in this study (Done et al. 2003).

WRF is a fully compressible, nonhydrostatic mesoscale model (Skamarock et al. 2005). The vertical coordinate follows the terrain using hydrostatic pressure, and the model uses an Arakawa-C grid. Prognostic variables are column mass of dry air (μ), velocity (u, v, and w), potential temperature (θ) and geopotential height (ϕ).

In this work, two domains with one-way nesting are used. The coarse domain covers the contiguous United States with 64x45 grid points and a grid spacing of 90 km, and the inner domain covers the central United States with 76x61 grid points and a grid spacing of 30 km (Fig. 4.1a). Both model domains have 27 vertical layers, and the model top is set at 100 hPa. Unless otherwise specified, the physical parameterization schemes include the Grell-Devenyi cumulus scheme, WSM 6-class microphysics with graupel, and the YSU scheme for planetary boundary layer (PBL) processes [see Skamarock et al. (2005) for references to different schemes]. NCEP global final (FNL) analyses are used to create initial and boundary conditions. Data assimilation and verification are only performed in inner domain.

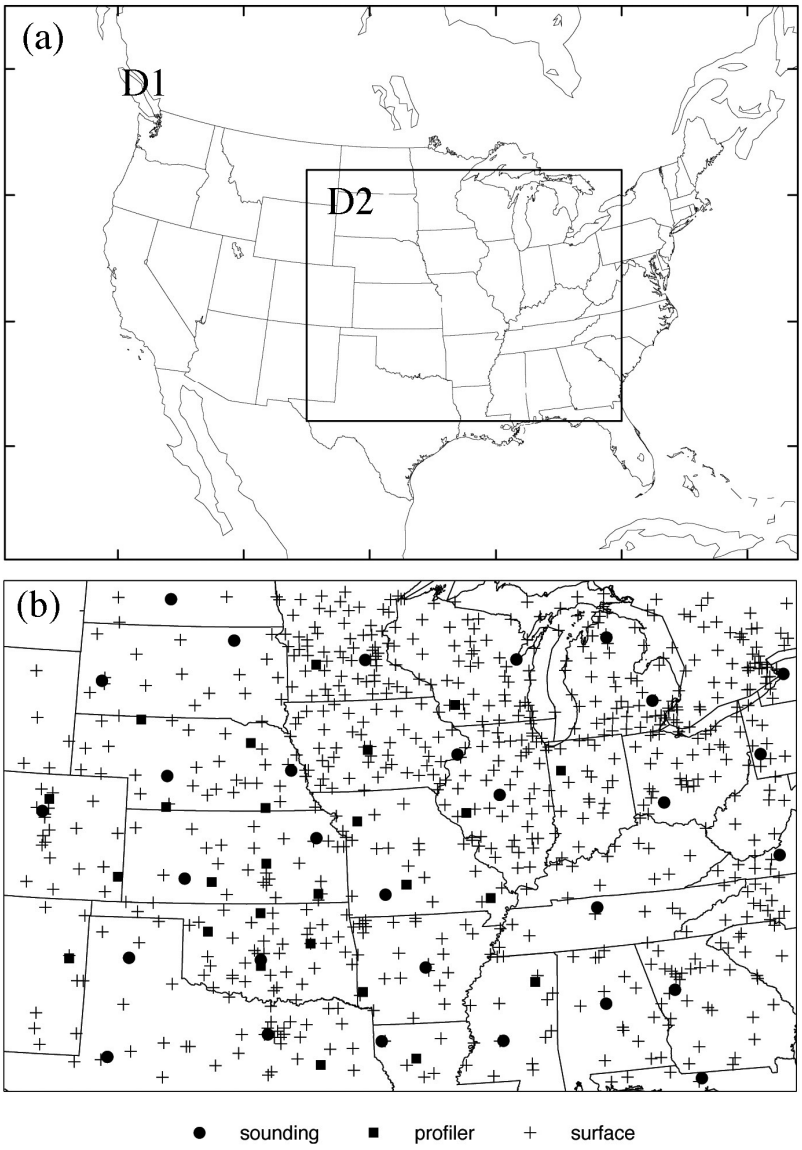

FIG. 4.1. (a) Map of the model domain and (b) distribution of assimilated observations in domain 2.

b. WRF-3DVar

The WRF-3DVar method used here was developed primarily at NCAR, and it is now operational at the Air Force Weather Agency (Barker et al. 2004). Its configuration is based on an incremental formulation, producing a multivariate analysis in the model

space. Its incremental cost function is minimized in a preconditioned control variable space where the errors of different control variables are largely uncorrelated.

WRF-3DVar has several background error statistic (BES) options for control variables ("cv"). The control variables for commonly used "cv3" and "cv5" are streamfuntion, unbalanced velocity potential, unbalanced surface pressure, unbalanced temperature, and "pseudo" relative humidity. The WRF-3DVar background error covariance can be estimated via the so-called "NMC" (an acronym of National Meteorological Center) method (Parrish and Derber 1992), which uses the statistics of differences between at least one month of 24- and 12-h forecasts valid at the same time. The "cv3" option is created with NMC method based on NCEP/GFS (Global Forecast System) forecast differences. Its vertical covariance is formulated in grid-space. The "cv5" option is created via NMC method based on regional model forecasts and its vertical covariance is formulated in eigenvector-space (Skamarock et al. 2005). Alternatively deviations of individual members from the mean of a short-term ensemble can be used, which may improve the representation of smaller-scale features (Lee and Barker 2005). Sensitivity of the performance of 3DVar to different representations of background error statistics will be examined in section 5.

After minimization in control space, the control vector is then transformed back to model space via an empirical orthogonal function (EOF) transform and a recursive filter. Wind increments are then calculated from the streamfunction and velocity potential. Balanced mass increments are obtained through linearized geostrophic and cyclostrophic mass-wind balance equations (Barker et al. 2004).

c. The EnKF

The EnKF is the same as that in Chapter II and Chapter III except for being implemented in the WRF model. It uses covariance relaxation (Zhang et al. 2004, their Eq.5) to inflate the background error covariance. Different from the standard inflation method (Anderson 2001) in which all points in the prior field are inflated, this relaxation method only inflates the covariance at updated points via a weighted average between the prior perturbation (denoted by superscript *f*) and the posterior perturbation (denoted by superscript *a*) as follows:

$$(\mathbf{x}^a_{new})' = (1-\alpha)(\mathbf{x}^a)' + \alpha(\mathbf{x}^f)' \tag{4.1}$$

The weighting coefficient, α, is set to 0.5 in the OSSE studies of Chapter II and Chapter III. Considering that prior error in real-data application may be larger due to the unavoidable imperfectness of the forecast model, a value of 0.7 is used here unless otherwise specified. The Gaspari and Cohn (1999) fifth-order correlation function with a radius of influence of 30 (10) grid points [i.e., 900 km (300km)] for soundings and profilers (surface observations) in horizontal directions and 15 sigma levels in vertical directions is used for covariance localization. Although the optimum ensemble size to estimate the forecast uncertainty is still under active research, 40 members are used herein. This is both affordable and reasonable based on previous studies (e.g., Houtekamer and Mitchell 2001; Anderson 2001; Snyder and Zhang 2003; Zhang 2005; Chapters II&III).

d. Ensemble initial and boundary perturbations

The initial ensemble is generated with the WRF-3DVar (Barker et al. 2004) using the BES of option cv3. To create a largely balanced perturbation, we first generate a set of random control vectors with a normal distribution (zero mean and unit standard deviation). Then the control increment vector is transformed back to model space via an EOF transform and a recursive filter. The balances, scales and correlations of the resulting model-space horizontal wind components, potential temperature, and mixing ratio for water vapor are defined via the climatological background error covariance cv3. Other prognostic variables such as vertical velocity (w) and mixing ratios for cloud water (q_c), rain water (q_r), snow (q_s) and graupel (q_g) are not perturbed. The perturbation standard deviations thus generated are approximately 2 m s^{-1} for horizontal wind components (u and v), 1 K for potential temperature (T), 1 hPa for pressure perturbation (p'), and 0.5 g/kg for water vapor mixing ratio (q). The 3DVar perturbations are added to the NCEP reanalysis to form an initial ensemble, which is then integrated for 12 h to develop an approximately realistic, flow-dependent background error covariance structure before the first observation is assimilated. Similar methods, using 3DVar to generate the initial ensemble for the EnKF, are also employed in Houtekamer et al. (2005) and Barker (2005).

The most natural way to perturb lateral boundary conditions for a limited area model is to use a global ensemble forecast with a correct size and resolution (which is usually unavailable) (Chessa et al. 2004). Torn et al. (2006) examined several alternative boundary perturbation methods and concluded that the error originating from using

different methods is limited to near the edges of the domain. In this chapter, the boundary is perturbed in the same manner as with the initial ensemble. Specifically, 6-hourly FNL analyses during the entire integration time-window are perturbed by randomly sampling the cv3 background error covariance of WRF-3DVar. Then each perturbed FNL analysis is used to generate a perturbed boundary condition for individual member of the initial ensemble.

3. Overview of the MCV event and observations to be assimilated

a. The MCV event in BAMEX

The case of interest is an MCV event that occurred during an intense observation period (IOP8) of the Bow Echo and Mesoscale Convective Vortex Experiment (BAMEX) conducted from 18 May to 7 July 2003 over the central United States (Davis et al. 2004). This event exhibited typical environmental features common to long-lived MCVs such as weak shear and moderate environmental instability (Hawblitzel et al. 2006; Davis and Trier 2006; Trier and Davis 2006; Trier et al. 2006). At 0000 UTC 10 June 2003, a disturbance embedded in the subtropical jet triggered convection over eastern New Mexico and western Texas. An MCV developed from the remnants of this convection over central Okalahoma at 0600 UTC 11 June 2003 and matured by 1800 UTC 11 June (with a bow echo occurring at western Tennessee) as it traveled northeastward to Missouri and Arkansas (Fig. 4.2). Its circulation was about 400 km wide and possessed a well-defined PV maximum around 600 hPa with a cold anomaly below the circulation and a warm anomaly above it. This MCV seemed to help initiate

widespread convection on its downshear side at 1600 UTC and on the north side of the circulation at 2200 UTC. Most convection died by 0000 UTC 12 June, and the MCV then transitioned into an extratropical baroclinic system (Hawbliztel et al. 2006).

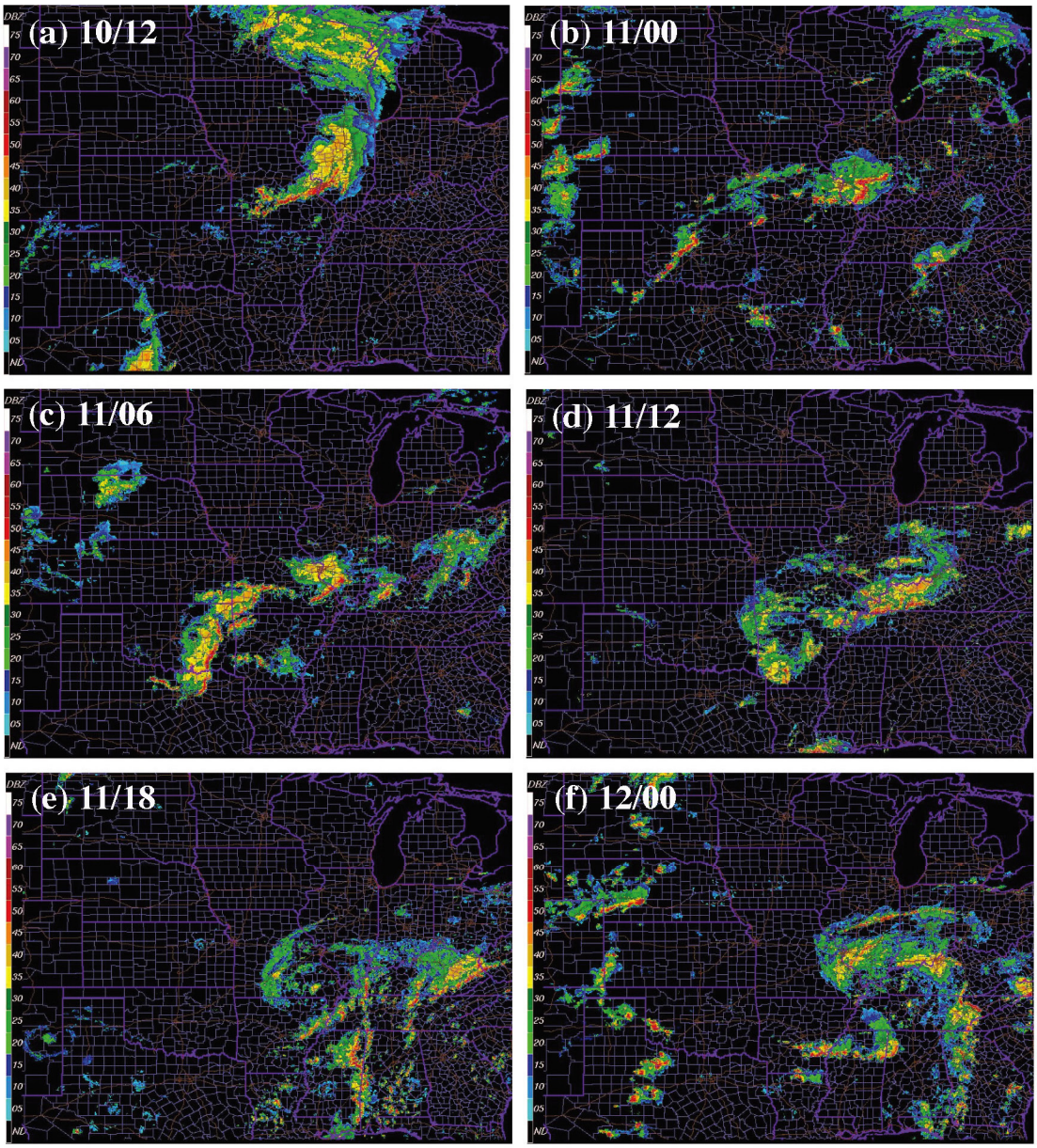

FIG. 4.2. Observed radar echoes of the MCV event at (a) 1200 UTC June 10, 2003, (b) 0000 UTC, (c) 0600 UTC, (d) 1200 UTC, (e) 1800 UTC, June 11, 2003, and (f) 0000 UTC June 12, 2003.

b. Observations to be assimilated

The observations to be assimilated in this chapter include sounding, surface, and wind profiler observations located in domain 2 (Fig. 4.1b). Wind profilers (denoted by filled squares) are distributed mainly in central United States to fill the gap between radiosondes. They have previously proven to be effective in improving short-range (3-12h) forecasts (Benjamin et al. 2004). Data thinning is performed here on the profiler observations so that the vertical resolution is similar to that of typical soundings.

Here we use a quality control method similar to a procedure in Barker (2005) that guarantees that 3DVar and the EnKF assimilate exactly the same observations. First, the final analyses ("FNL") of global forecast system of National Centers for Environmental Prediction are interpolated to the model grid at 6-h intervals. Then the model is integrated for 6 hours starting from the 6-hourly initial condition. The interpolated 6-hourly FNL analyses together with WRF forecasts in between are used as the first guess for a pre-run of WRF-3DVar to assimilate observations generated by the observation preprocessor of WRF-3DVar. The data processed at hourly intervals that have been ingested by this pre-run will then be assimilated by the following 3DVar and EnKF experiments. The output file containing the ingested observations transforms the original wind speed, wind direction and relative humidity into horizontal wind components (u and v) and mixing ratio of water vapor (q). Consequently, the assimilated variables become horizontal wind components (u and v), temperature (T), and mixing ratio of water vapor (q) for soundings, surface pressure (ps), u, v, T, and q for surface data, and u

and *v* for wind profiler data. In this way, the observations first go through basic quality control such as range, domain, persistency, extreme-value and horizontal consistency checks. Then in the 3DVar pre-run, an "errormax" quality control is performed by dropping the observations whose absolute difference from the first guess is more than five times larger than the corresponding observation error.

The observation error of sounding and wind profiler is what is defaulted in officially issued WRF-3DVar system (Table 4.1). The observation error of surface data is 2 K for temperature, 10% for relative humidity, and 100 Pa for surface pressure (default in the formally-released WRF-3DVar system version 2.1).

TABLE 4.1. Vertical distribution of observation error for sounding and profiler observations.

Pressure (hPa)		100	150	200	250	300	400	500	700	850	1000
Wind (m/s)	sounding	2.7	3.0	3.3	3.3	3.3	2.8	2.3	1.4	1.1	1.1
	profiler	2.8	2.8	3.0	3.1	3.2	3.0	2.8	2.2	2.2	2.2
Temperature (K)		1	1	1	1	1	1	1	1	1	1
RH (%)		10	10	10	10	10	10	10	10	10	15

All the results in this study including both prior (forecast before data assimilation) and posterior (analysis after data assimilation) are verified against soundings that have passed our quality control procedure at nine standard pressure levels, i.e., 925, 850, 700, 500, 400, 300, 250, 200 and 150 hPa.

4. Comparison between the EnKF and 3DVar

a. The reference forecasts

With the initial and boundary conditions interpolated from the FNL analysis, a 36-h

reference deterministic forecast ("DF") is conducted starting from 1200 UTC 10 June 2003 without assimilating any observations. As shown in Fig. 4.3a, the simulated MCV (denoted by *X*) moves much faster than the observed one (denoted by *L*), resulting in a position error of about 400 km. The simulated radar reflectivity is more localized and not well organized compared to the observed radar echoes (Fig. 4.2f).

To assess the benefit of the EnKF from both the data assimilation algorithm itself and the utilization of the ensemble forecast for state estimate, a 36-h reference ensemble forecast is also performed with the same model configuration for all members as that in the DF but with the addition of initial and boundary perturbations described in section 2d (hereafter also referred to as "EF"). Relative to the DF, the mean of EF shows slightly smaller error in the 36-h surface location of the simulated MCV (Fig. 4.3b). The magnitude of its simulated reflectivity is lower due to the ensemble averaging.

The error of the reference EF is smaller than that of DF in terms of column-averaged RM-DTE and root-mean square (RMS) error (RMSE) of q verified against sounding observations (Figs. 4.4 a-d). The DTE is defined as in Zhang (2005):

$$\mathrm{DTE} = 0.5(u'u' + v'v' + \mathrm{k}T'T'), \tag{4.2}$$

where the prime denotes the difference between the observations and the verified fields, $\mathrm{k} = C_p/T_r$, $C_p = 1004.7$ J kg^{-1} K^{-1} and the reference temperature $T_r = 290$ K. Fig. 4.4 also shows that the largest errors in both reference forecasts occur around the MCV.

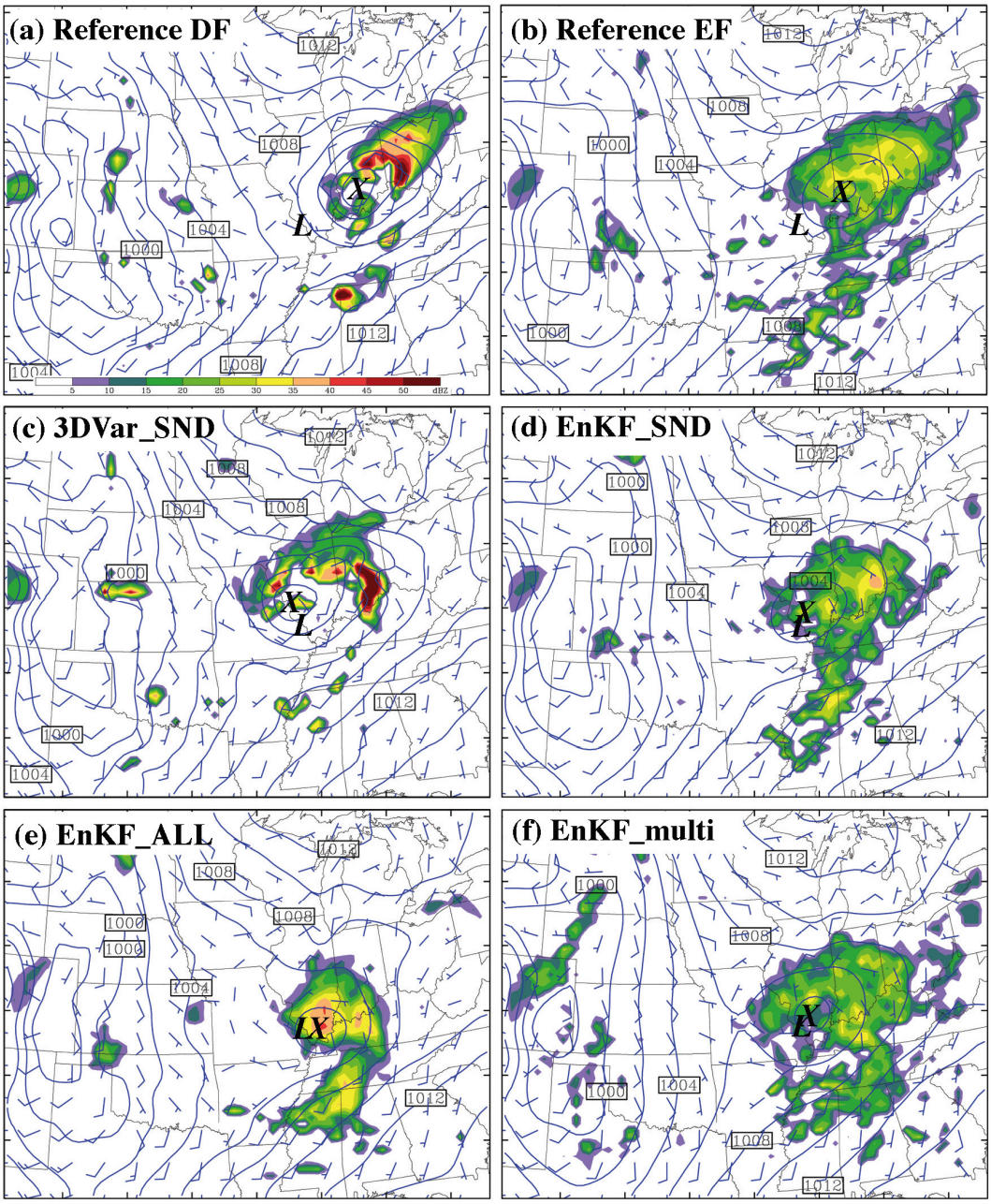

FIG. 4.3. The MSLP (every 2 hPa), 10-m wind vectors (full barb 5 m/s) and simulated reflectivity (shaded) valid at 36 h (0000 UTC 12 June) of the reference forecast (a) DF and (b) EF, and the prior forecast of (c) 3DVar_SND, (d) EnKF_SND, (e) EnKF_ALL and (f) EnKF_multi. The big X and L respectively denote the simulated and observed MCV centers at surface.

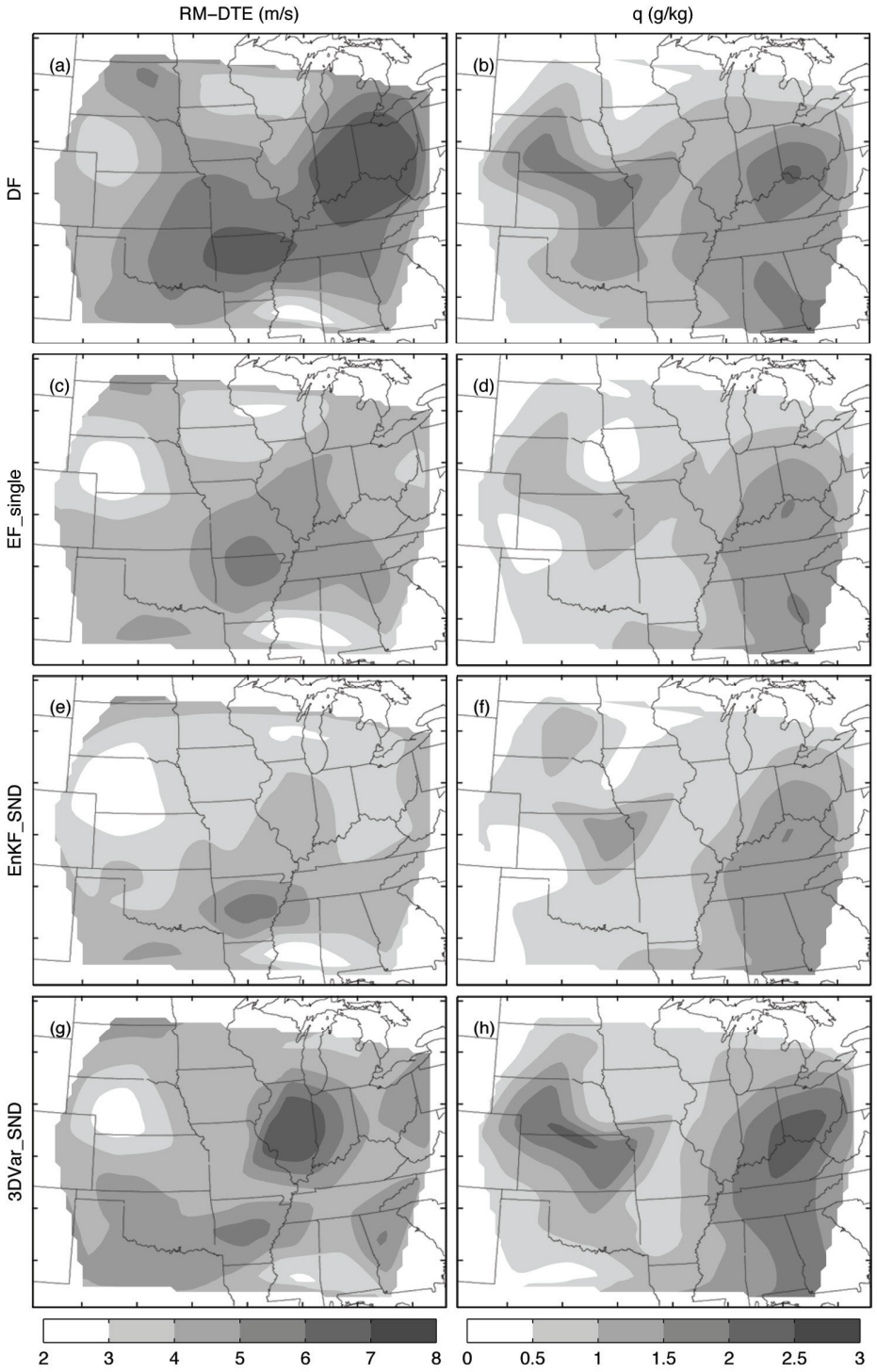

FIG. 4.4. Horizontal distributions of column-averaged RM-DTE and RMSE of q valid at 36 h for the reference forecast (a-b) DF and (c-d) EF, the prior forecast of (e-f) EnKF_SND, and (g-h) 3DVar_SND.

b. Experiments with different types of observations

Three types of observations including sounding, surface and wind profiler data are assimilated separately via both the EnKF and 3DVar in this section. The experiments are named "EnKF_SND" and "3DVar_SND" for sounding assimilation, "EnKF_SFC" and "3DVar_SFC" for surface assimilation, and "EnKF_PFL" and "3DVar_PFL" for profiler assimilation.

1) SOUNDING DATA

The time-evolutions of domain-averaged RMSE of u, v, T, q are shown in Fig. 4.5. The red (green) dashed line is the error of reference ensemble (deterministic) forecast; the red (green) solid line is the prior error of EnKF_SND (3DVar_SND). The dotted-dashed lines are corresponding posterior error. Blue lines denote the result of another experiment to be described in section 5, in which model error from physical parameterization is accounted for using multiple physical parameterization schemes in the ensemble. The results show that the EnKF very efficiently draws the analysis (red dotted-dashed line) close to observations for the 30 sounding sites within the inner (assimilation and verification) domain where u, v, T, q and ps are assimilated every 12 h. At each data assimilation time, the posterior RMSE (red dotted-dashed line) is smaller than the prior RMSE (red solid line) by up to 50%. However, since the verification of the analysis (posterior) uses the same sounding observations as those assimilated, it is more appropriate to judge the performance of the data assimilation via the short-term forecast (prior estimate) initialized with the posterior analysis from the previous

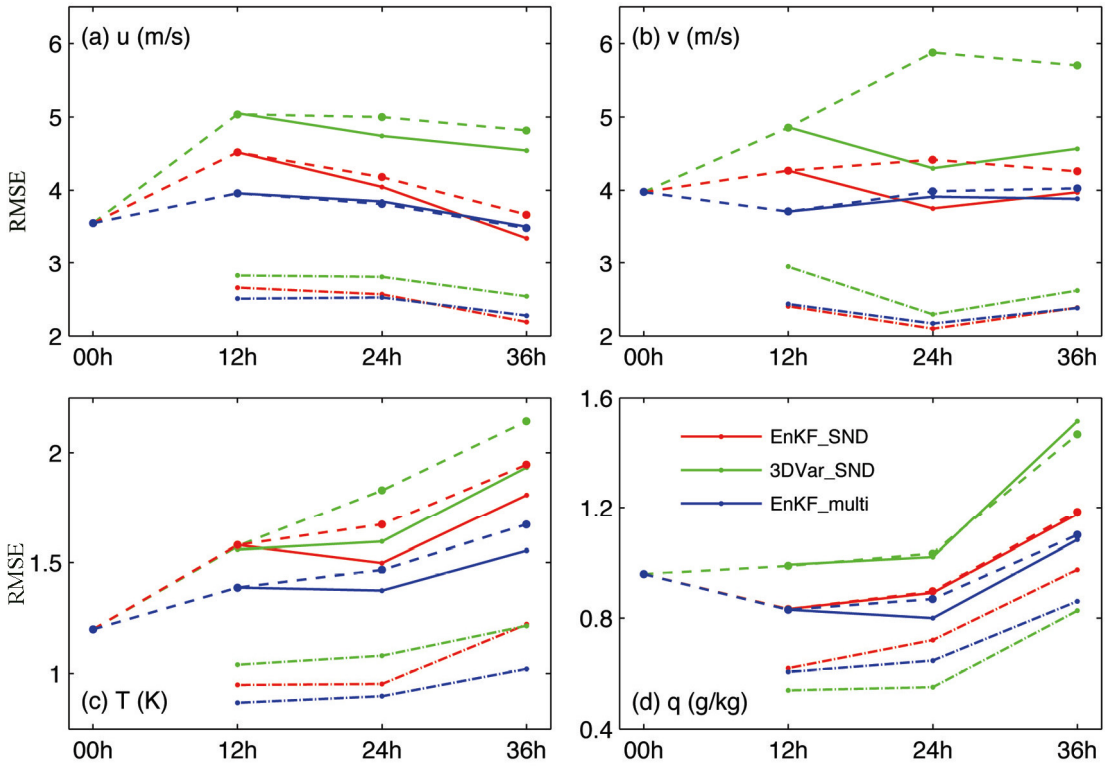

FIG. 4.5. Time evolution of domain-averaged RMSE of (a) u, (b) v, (c) T, and (d) q for EnKF_SND (red), 3DVar_SND (green), and EnKF_multi (blue). The solid lines denote the prior RMSE and the dot-dashed lines the posterior RMSE. Also plotted are the RMSE of the reference forecast DF (green dashed), EF (red dashed), and EF_multi (blue dashed).

assimilation cycle. The EnKF prior estimate (red solid line) at 36 h also tracks the observations better than both the reference DF (green dashed line) and EF (red dashed line). Similar to the result obtained in OSSE studies (Chapters II&III), larger improvements are observed in variables with higher power at large scales such as *u, v* and *T*. After two cycles at 36 h, the prior RMSEs of *u*, *v*, and *T* are respectively 3.3 m/s, 4.0 m/s, and 1.8 K. The RMSE of *q* grows faster than other variables during the subsequent integrations, likely due to its higher spectral power in smaller scales (Chapter II).

In comparison to EnKF_SND, apparently worse performance is seen in

3DVar_SND. Though its prior RMSE at 36 h (green solid line) is smaller than that of the reference DF (green dashed line), it is much larger than that of EnKF_SND and even slightly larger than the RMSE of the reference EF for all *u*, *v*, *T*, and *q*. The posterior errors (green dot-dashed line) of *u* and *v* at 36h are also noticeably larger than those of EnKF_SND (The reason for the smaller posterior error of 3DVar_SND in q than that of EnKF_SND is under examination). The errors grow much faster in the 3DVar than in the EnKF during the 12-h forecasts initiated from the posterior analyses. This result is consistent with another work showing a better performance of a global model based EnKF relative to 3DVar by assimilating the NCEP operational observation network except for satellite data (Whitaker et al. 2006).

The better performance of the EnKF relative to 3DVar can also be seen in the vertical distribution of domain-averaged prior RMSEs of *u*, *v*, *T* and *q* at 36 h (Fig. 4.6). The errors of 3DVar (green solid line) are generally larger than those of the EnKF (red solid line) in each layer with local maxima near the tropopause and the surface. Most of the error reduction of the EnKF relative to the reference EF (red dashed lines) comes from the lower troposphere. Examination of the horizontal distribution of the column-averaged prior error shows that the EnKF performs significantly better than 3DVar over the MCV area although errors in southern portion of domain persist (Figs. 4.4e-h).

EnKF_SND also draws the simulated MCV closer to the observed location (Fig. 4.3d) than the reference forecasts and 3DVar (Figs. 4.3a-c). The simulated reflectivity is slightly stronger and better organized (especially to the south of the surface center) than that in the reference EF. Compared to the reference DF (Fig. 4.3a), 3DVar_SND (Fig.

4.3c) also better simulates the position of the MCV and associated reflectivity.

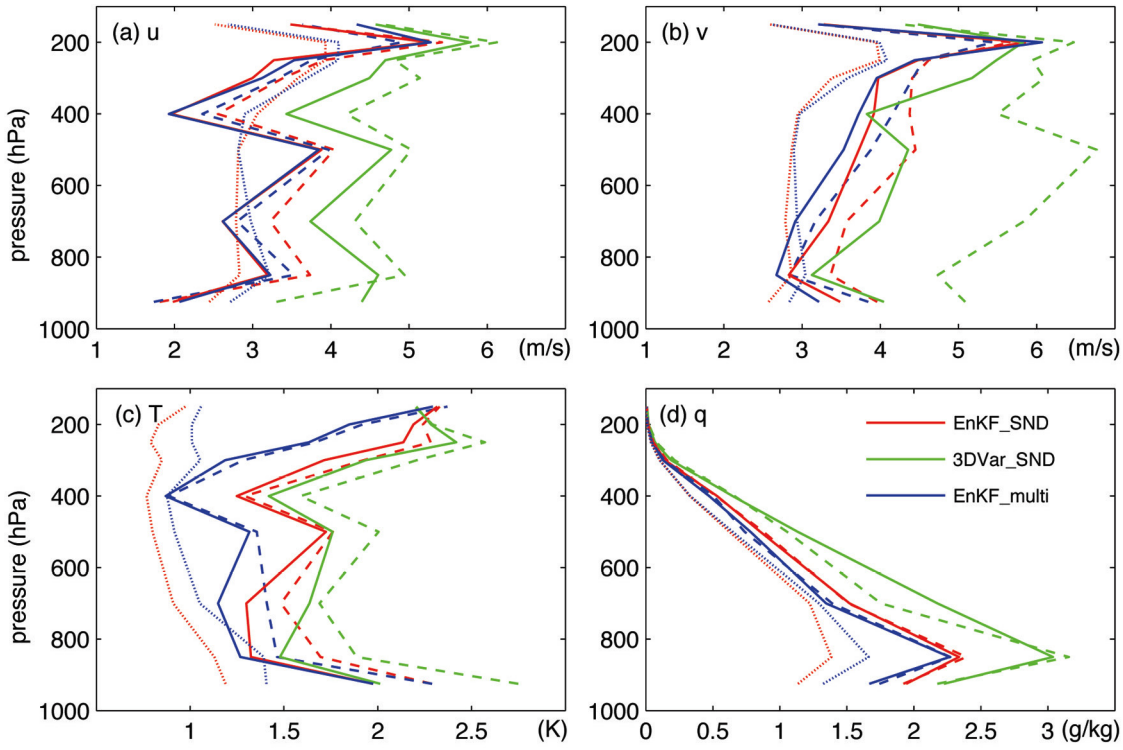

FIG. 4.6. Vertical distributions of horizontally-averaged prior RMSE of (a) u, (b) v, (c) T, and (d) q for EnKF_SND (red solid), 3DVar_SND (green solid), and EnKF_multi (blue solid). Also plotted are the horizontally-averaged prior ensemble spread (STD) of EnKF_SND (red dotted) and EnKF_multi (blue dotted) together with the horizontally-averaged RMSE of the reference forecast DF (green dashed), EF (red dashed), and EF_multi (blue dashed).

The domain-averaged RM-DTE of forecasts and analyses of all experiments at 36 h (Fig. 4.7) clearly shows that EnKF_SND performs better than 3DVar_SND. Also shown is that EnKF_SND forecasts perform better than reference EF and DF, while 3DVar_SND forecasts perform better than reference DF but worse than reference EF. The RM-DTE of EnKF_SND is 4.38 m/s while that of 3DVar_SND is 5.19 m/s. Results from other experiments also shown in Fig. 4.7 will be discussed in subsequent sections.

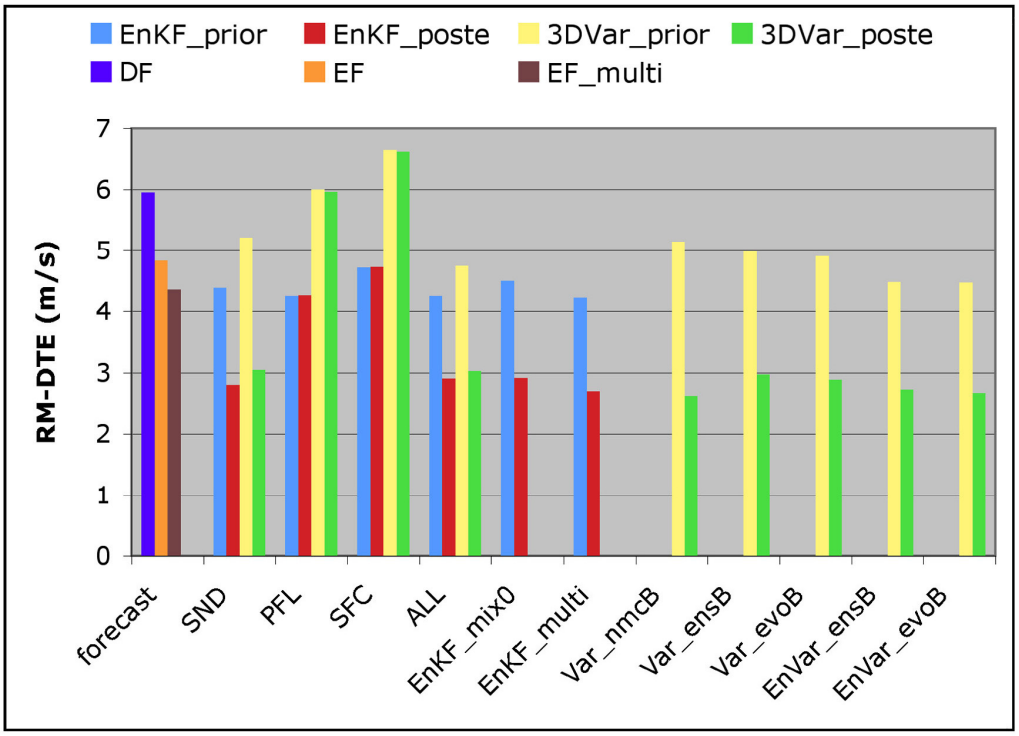

FIG. 4.7. A summary of domain-averaged RM-DTE prior forecast error valid at 36 h for all experiments.

2) WIND PROFILER

While the 28 profilers in the inner domain with *u* and *v* observations take observations as frequent as every six minutes (Fig. 4.1b), profiler data assimilation was only performed every 3 h. Testing showed no benefit (in terms of analysis error) in assimilating profiler data more often than every three hours.

Compared to EnKF_SND, EnKF_PFL has a slightly smaller prior RM-DTE at 36 h (4.25 m/s vs. 4.38 m/s in Fig. 4.7), likely because the previous assimilation cycle was only 3 h prior in EnKF_PFL (whereas it was 12 h prior in EnKF_SND). The posterior error is almost the same as the prior error, but it is significantly larger than that of

EnKF_SND since the same sounding observations (assimilated in EnKF_SND) are used to verify the results in all cases. There is apparently larger improvement in *v* (Fig. 4.8b). It is worth noting that though temperature is not an observed variable, it is also improved relative to the reference EF below about 500 hPa through the flow-dependent background error covariance of the EnKF (blue solid line in Fig. 4.8c).

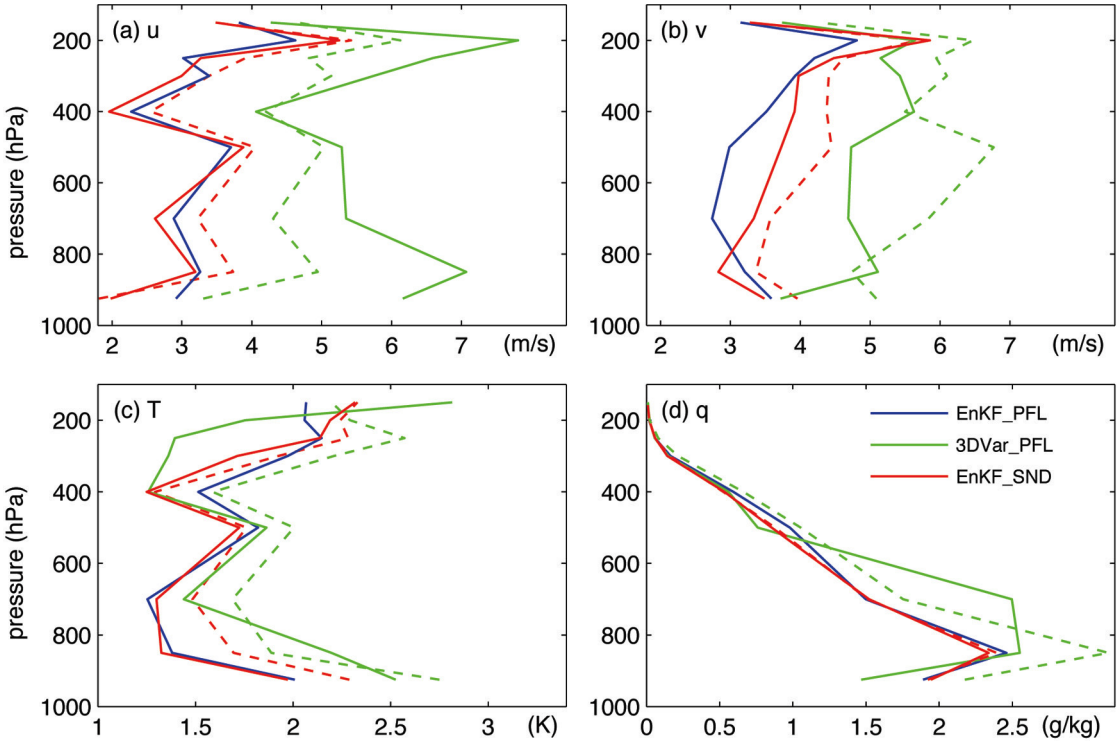

FIG. 4.8. Vertical distributions of horizontally-averaged RMSE of (a) u, (b) v, (c) T, and (d) q at 36 h for the prior forecast of EnKF_PFL (blue solid), 3DVar_PFL (green solid), EnKF_SND (red solid), the RMSE of the reference forecast DF (green dashed) and EF (red dashed).

Figure 4.8 shows that 3DVar_PFL (green solid line) shows a generally larger prior error than EnKF_PFL (blue solid line). It performs worse than EnKF_PFL at every layer for *u*, *v* and *q* and lower layers for *T*. The prior RM-DTE of 3DVar_PFL at 36 h is 1.75 m/s larger than that of EnKF_PFL (Fig. 4.7).

3) SURFACE DATA

Compared to the sounding and profiler data assimilation results, even though there are many more surface stations in the inner domain, much less improvement is seen when only surface data is assimilated every 6 h in experiments EnKF_SFC and 3DVar_SFC (Fig. 4.9). Only a small error reduction is seen in terms of the prior RM-DTE in EnKF_SFC relative to the reference EF (Fig. 4.7). As expected, error reduction in EnKF_SFC (blue solid line in Fig. 4.9) versus EF is mainly in the lower troposphere. Though less improvement is obtained in this case, EnKF_SFC still outperforms 3DVar_SFC (green solid line in Fig. 4.9). The prior RM-DTE of EnKF_SFC at 36 h is 1.94 m/s less than that of the 3DVar_SFC.

4) SOUNDING, PROFILER AND SURFACE DATA

The results of experiments "EnKF_ALL" and "3DVar_ALL", which assimilate all three types observations at the aforementioned frequencies, show that utilization of all data types results in generally better performance than assimilating any individual source of observations (in terms of the prior error at 36 h; see Fig. 4.10 and Fig. 4.7). The RM-DTE in EnKF-All at 36 h is 4.25 m/s, which is 0.58 m/s less than that of the reference EF (Fig. 4.7). The most significant improvement comes from the middle troposphere esp. in v (Fig. 4.10). The position of the MCV and its associated reflectivity are also noticeably improved (Fig. 4.3e). Again, the EnKF outperforms 3DVar for all u, v, T and q (Fig. 4.8). The prior RM-DTE of 3DVar_ALL at 36 h is 4.74 m/s, which is 0.49 m/s

larger than that of EnKF_ALL but the difference is smaller than that between EnKF_SND and 3DVar_SND when less data are assimilated (Fig. 4.7).

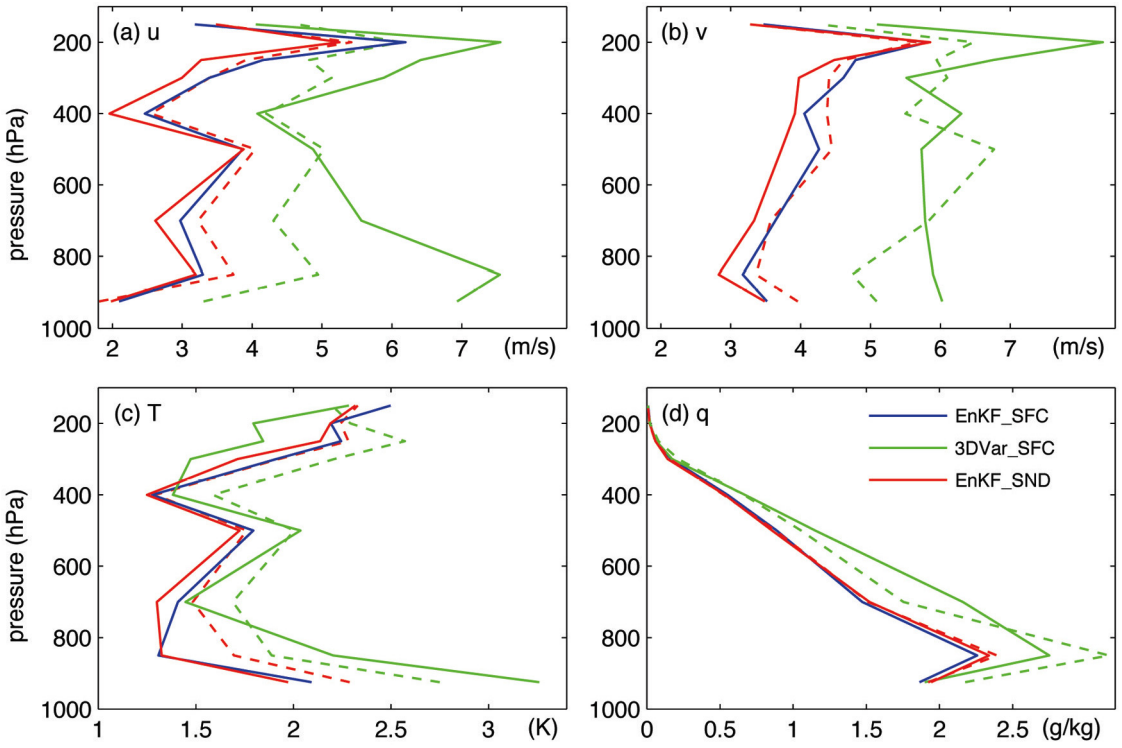

FIG. 4.9. As in Fig. 4.7 except for EnKF_SFC (blue solid), 3DVar_SFC (green solid), and EnKF_SND (red solid).

5. Sensitivity to background error covariance

Background error covariance, which determines how and where to spread the observed information to other points and other variables, plays a very important role in the performance of any data assimilation method. This section investigates the sensitivity of both the EnKF and 3DVar to different treatments of background error covariance; for simplicity, only soundings are assimilated every 12 h using EnKF_SND and 3DVar_SND as benchmarks for comparison.

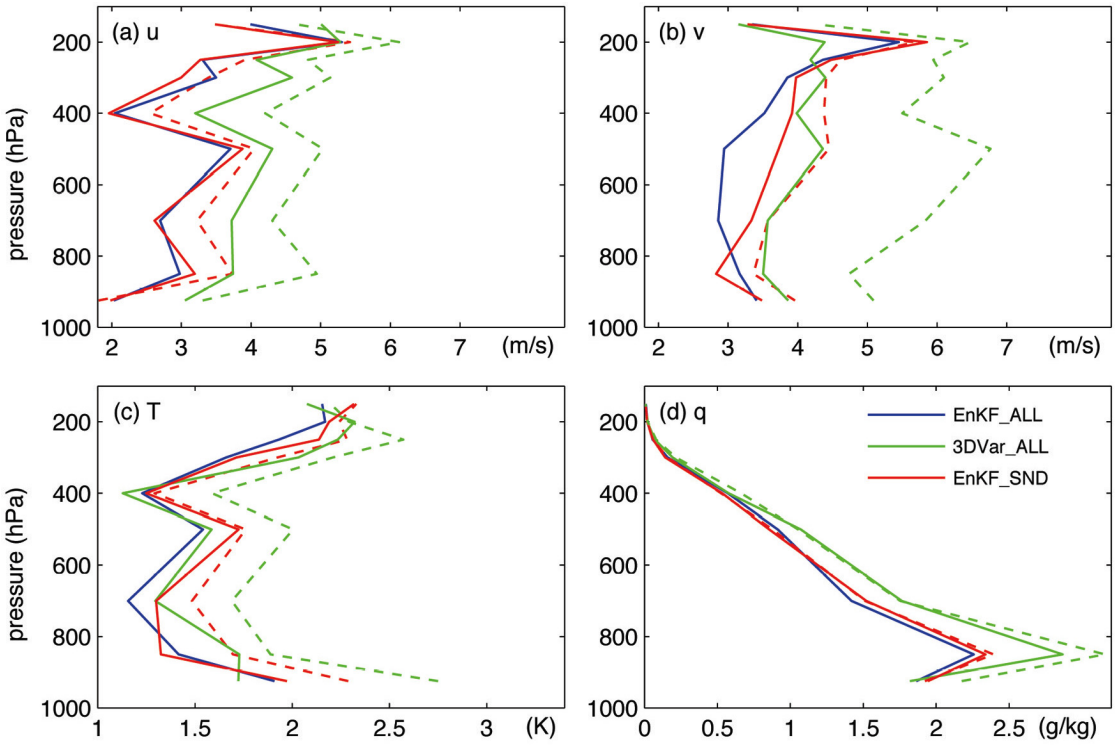

FIG. 4.10 As in Fig. 4.7 except for EnKF_ALL (blue solid), 3DVar_ALL (green solid), and EnKF_SND (red solid).

a. Model error treatments in the EnKF

1) MULTI-SCHEME ENSEMBLE

The impact of using a multi-scheme ensemble is investigated here to test the OSSE result of Chapter III that the EnKF performance can be improved through the use of different physical parameterization schemes in different ensemble members to account for model physics uncertainty. Experiment "EnKF_multi" uses a combination of three cumulus schemes (Kain-Fritch, Betts-Miller, Grell-Devenyi), three PBL schemes (YSU, ETA, and MRF), and three cloud-physics schemes (Lin et al., Thompson et al., and WSM 6-class graupel) which are distributed among nearly the same number of ensemble

members (Table 4.2). A reference to different parameterization schemes can be found in Skamarock et al. (2005). To differentiate the benefit of using the multi-scheme ensemble for better background error covariance from better prior estimate (or ensemble mean), another reference forecast "EF_multi" is performed with the same multi-scheme configuration as that in EnKF_multi but without assimilating any observations.

TABLE 4.2. Model configuration of the experiment with multi-scheme ensemble Mix0.7_multi.

Number of members using cumulus scheme	Number of members using microphysics scheme	Number of members using PBL scheme
13 Kain-Fritsch	4 Lin et al.	1 YSU 2 ETA 1 MRF
	4 Thompson et al.	1 YSU 2 ETA 1 MRF
	5 WSM 6-class graupel	2 YSU 2 ETA 1 MRF
13 Betts-Miller	4 Lin et al.	1 YSU 2 ETA 1 MRF
	4 Thompson et al.	1 YSU 2 ETA 1 MRF
	5 WSM 6-class graupel	2 YSU 2 ETA 1 MRF
14 Grell-Devenyi	4 Lin et al.	1 YSU 2 ETA 1 MRF
	5 Thompson et al.	2 YSU 2 ETA 1 MRF
	5 WSM 6-class graupel	2 YSU 2 ETA 1 MRF

The result shows that both the prior and posterior errors at 36 h of v, T and q in EnKF_multi are decreased more than those in EnKF_SND (Fig. 4.5). The error reduction of 0.16 m/s comes from both the prior estimate using the multi-scheme ensemble forecast (blue dashed line in Fig. 4.5) and the EnKF with better background error covariance (blue solid in Fig. 4.5). The largest improvement is observed in T, which should be more closely related to the uncertainty in the physical parameterization schemes. EnKF_multi has apparently larger ensemble spread for all variables shown (blue dotted line in Fig. 4.6) compared to that of EnKF_SND (red dotted line in Fig.

4.6). At the surface (Fig. 4.3f), the simulated position is similar to that of EnKF_SND (Fig. 4.3d). The associated simulated reflectivity (Fig. 4.3f) expands farther to the southeast and northeast and gets closer to the observed radar echo (Fig. 4.2f) in area but weaker in intensity.

2) COVARIANCE INFLATION

Due to sampling and model error, the EnKF may underestimate the analysis uncertainty, which could lead to filter divergence if untreated. Two common ways to cope with this problem are the multiplicative covariance inflation method (Anderson 2001) and the additive covariance inflation method (Hamill and Whitaker 2005; Houtekamer et al. 2005). However, both inflation methods may lead to excessive ensemble spread in data-sparse regions (Zhang et al. 2004).

As in Zhang et al. (2004) and described in section 2c, this chapter uses the covariance relaxation method (Eq. 4.1) to inflate the background error covariance but avoid excessive spread. Three experiments with α of 0., 0.5 and 0.7, referred to as EnKF_mix0, EnKF_mix0.5, and EnKF_SND, respectively, are performed to test the EnKF sensitivity to the relaxation coefficient. Result shows the relaxation (mixing) between the prior and posterior perturbations can draw the subsequent prior estimate closer to observations, and a larger relaxation coefficient is necessary for this real-data application than the value of 0.5 used in the OSSEs (Chapters II&III). Overall, EnKF_SND (red solid line in Fig. 4.11) results in slightly smaller prior RMSE than EnKF_mix0.5 (not shown). They both outperform EnKF_mix0 with no relaxation (green

solid line in Fig. 4.11).

Figure 4.11 also shows a consistent improvement from EnKF_mix0 to EnKF_SND and then to EnKF_multi (solid lines) possibly due to the correspondingly increasing ensemble spread as shown in dotted lines. This result indicates that proper covariance inflation or relaxation may improve the performance of the EnKF.

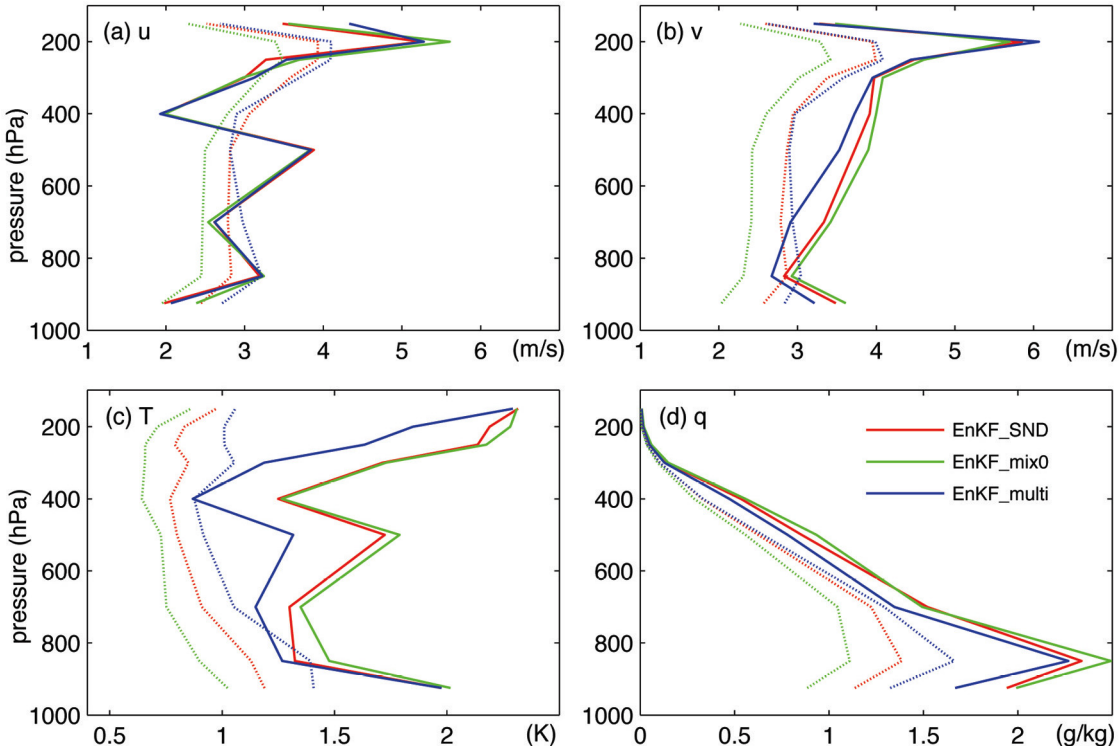

FIG. 4.11. As in Fig. 4.7 except for EnKF_mix0 (green lines), EnKF_SND (red lines), and EnKF_multi (blue lines).

b. Sensitivity to background error covariance in 3DVar

As introduced in section 2b, the WRF-3DVar default background error covariance (the cv3 option) is calculated using a month-long GFS global model forecast via the

"NMC" method. This could potentially be problematic due to error statistics calculated from a different model and a different grid size. In this section, sensitivity of the 3DVar to different background errors are examined in two ways (Table 4.3): one is the standard3DVar system using a newly generated BES but a single forecast for the prior estimate; the other is ensemble-based 3DVar, in which a short-term ensemble forecast is utilized to derive the prior estimate while the BES is either fixed or estimated with the same short-term ensemble (and thus has some flow dependency).

TABLE 4.3. Experiment designs on the sensitivity of 3DVar to different BESs.

Groups	Experiment	BES employed (the cv5 option)	Prior and posterior estimate
Standard 3DVar	Var_nmcB	Fixed BES via the "NMC" method	Standard 3DVar
	Var_ensB	Fixed BES via the first 12-h ensemble forecast (i.e., first 12h of EF)	Standard 3DVar
	Var_evoB	Evolving BES via preceding 12-h ensemble forecast	Standard 3DVar
Ensemble -based 3DVar	EnVar_ensB	Fixed BES as in Var_ensB; standard 3DVar analysis for each member	Ensemble mean of the forecast and analysis ensemble
	EnVar_evoB	Evolving BES via preceding 12-h ensemble forecast; standard 3DVar analysis for each member	Ensemble mean of the forecast and analysis ensemble

1) STANDARD 3DVAR EXPERIMENTS

Using the cv5 option in WRF 3DVar and the "NMC" method, experiment "Var_nmcB" generates the BES from 60 pairs of month-long (May 2003) 24-h and 12-h forecast differences valid at the same time (every 12 h) with the WRF model configured in the same manner as with the reference DF. This method is expected to provide more

reasonable BES for scale, balance and correlation structures than the one interpolated from the defaulted BES that was derived from a coarse-resolution global model (Lee and Barker 2005). However, in this case, the results of 3DVar_SND and Var_nmcB are nearly indistinguishable in terms of the prior error at 36 h (Figs. 4.7 and 4.12).

Another way to generate the BES is to use a short-term ensemble forecast (Lorenc 2003; Lee and Barker 2006). Experiment "Var_ensB" utilizes a 40-member 12-h ensemble forecast valid at 0000 UTC June 11 to generate a BES that is fixed for subsequent assimilation cycles. In this case, the prior RM-DTE at 36 h is slightly smaller than that of 3DVar_SND, which is consistent with other studies (e.g., Lee and Barker 2006). Vertical distributions of the prior RMSE show the improvement comes mainly from u and v (purple solid line in Fig. 4.12). The simulated reflectivity (Fig. 4.13 a) is similar to that of 3DVar_SND (Fig. 4.3c).

Experiment "Var_evoB" utilizes a 40-member 12-h ensemble to generate a different BES at every assimilation time (instead of fixed as in Var_ensB). This BES has some flow-dependency but is still isotropic and mostly homogeneous. Its difference from the fixed BES lies in the time-variant correlation length scale and variance. Relative to 3DVar_ensB, including some flow-dependency in the BES results in only slight further improvement in terms of the prior error at 36 h (Figs. 4.7 and 4.12). Again, the surface structure of the MCV is similar to that in the previous standard 3DVar experiments (Fig. 4.13b).

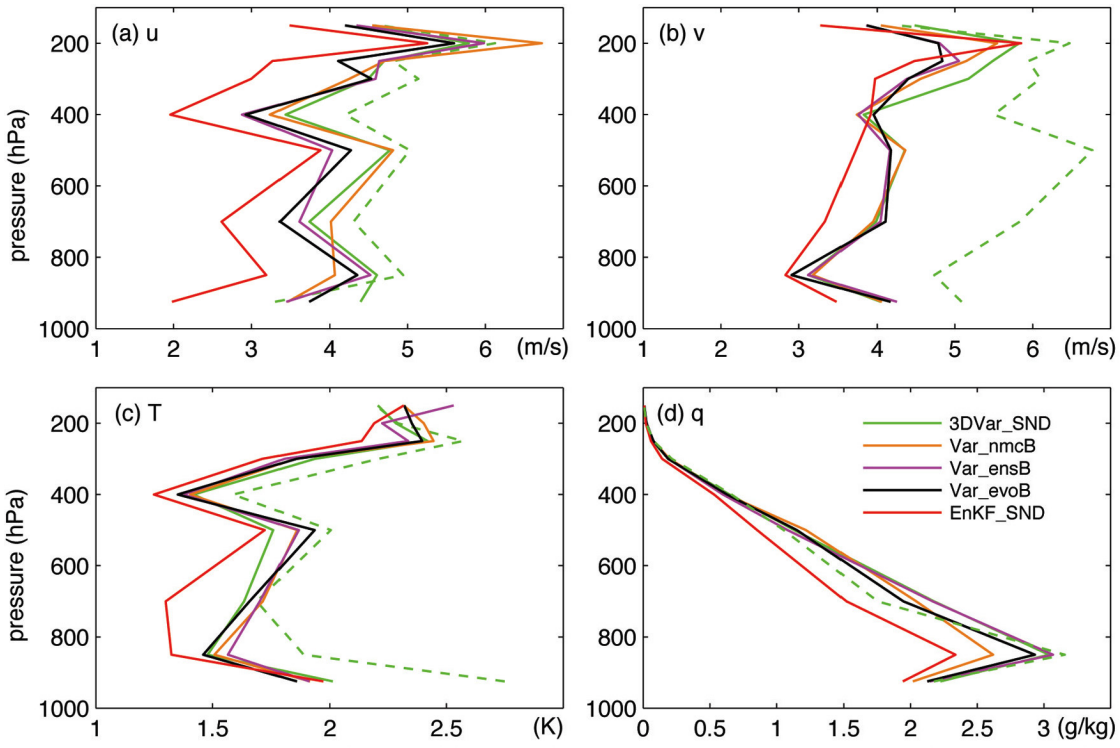

FIG. 4.12. Vertical distribution of horizontally-averaged prior RMSE of (a) u, (b) v, (c) T, and (d) q at 36 h in Var_nmcB (brown solid), Var_ensB (purple solid), and Var_evoB (black solid) in comparison to EnKF_SND (red solid). Also plotted are the RMSE of the reference forecast DF (green dashed) and the prior RMSE of 3DVar_SND (green solid).

2) ENSEMBLE-BASED 3DVAR EXPERIMENTS

Two ensemble-based 3DVar experiments, namely "EnVar_ensB" and "EnVar_evoB" (Table 4.3), are performed in this section to further examine the impacts of flow-dependent BES and using an ensemble forecast mean to perform prior estimate. EnVar_ensB uses the mean of an ensemble initiated by 3DVar with the cv3 BES option to perform the state estimation. Each member goes through the forecast and analysis cycling independently with the same fixed BES as that in Var_ensB. The prior and posterior estimate at each assimilation time are just the ensemble mean of the forecast and analysis ensembles. In EnVar_evoB, BES is generated from the forecast ensemble at

each assimilation time and it is thus time-variant with some flow dependency involved.

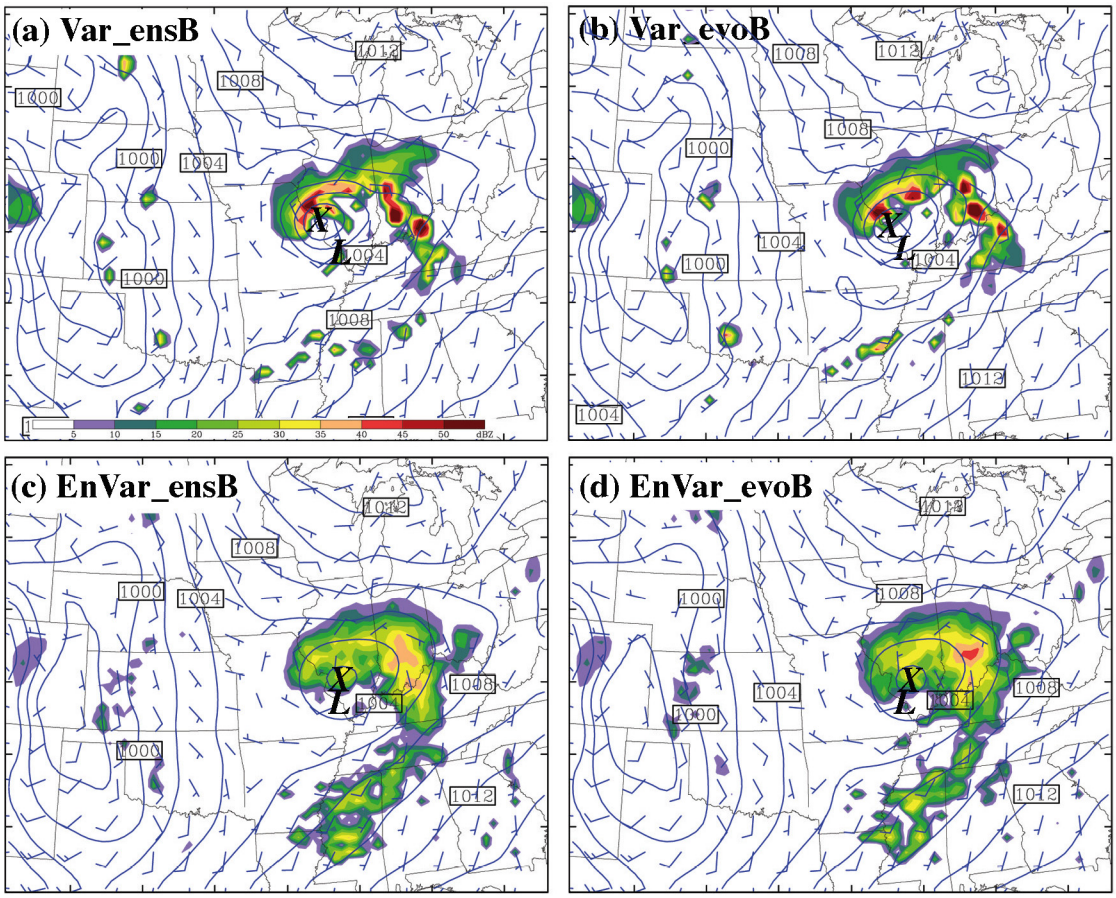

FIG. 4.13. As in Fig. 4.2 except for (a) Var_ensB, (b) Var_evoB, (c) EnVar_ensB, and (d) EnVar_evoB.

The result shows that the prior error of EnVar_ensB at 36 h (Figs. 4.7 and 4.14) is comparable to that of EnKF_SND, and it is significantly smaller than that of the standard 3DVar experiments. The surface location of the MCV and associated reflectivity pattern also apparently improve (Fig. 4.13c), becoming similar to that in EnKF_SND (Fig. 4.3d). This result suggests that the use of an ensemble mean can significantly improve the prior estimate and thus subsequent analyses. The improvement of EnVar_ensB over Var_ensB (0.5 m/s) is apparently larger than that of Var_evoB over Var_ensB (0.08 m/s),

which indicates that using an ensemble mean to estimate the prior estimate and/or the posterior analysis is an important benefit of the EnKF.

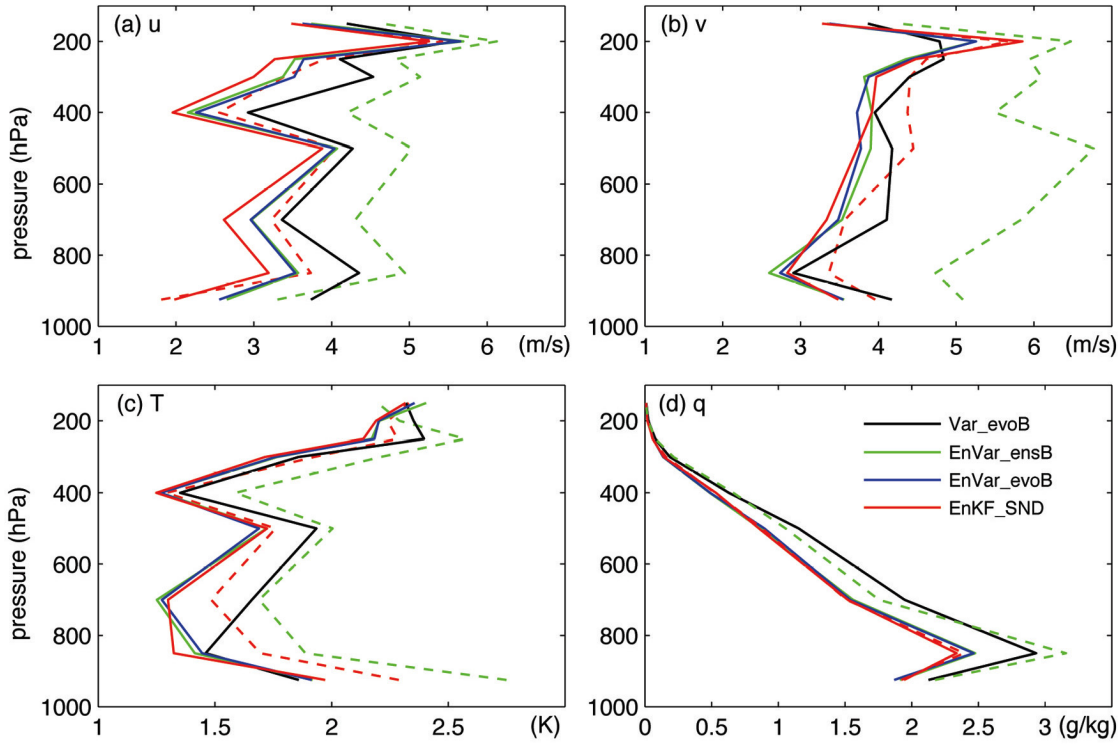

FIG. 4.14. Vertical distribution of horizontally-averaged prior RMSE of (a) u, (b) v, (c) T, and (d) q at 36 h in Var_evoB (black solid), EnVar_ensB (green solid), and EnVar_evoB (blue solid) in comparison to EnKF_SND (red solid). Also plotted are the RMSE of the reference forecast DF (green dashed) and EF (red dashed).

In EnVar_evoB there is only slight further improvement in *v* (blue solid in Fig. 4.14) relative to EnVar_ensB. The overall prior RM-DTE at 36 h in these two ensemble-based 3DVar experiments is very similar. The intensity of simulated reflectivity associated with the MCV is slightly improved (Fig. 4.13d). Similar to the comparison of Var_ensB and Var_evoB, the utilization of a time-variant BES in EnVar_evoB does not result in significant improvement over EnVar_ensB. Finally, the prior RM-DTE at 36 h of EnVar_evoB (4.47 m/s) is smaller than that of Var_evoB (4.9 m/s) (Fig. 4.7). This

further confirms the benefit of using an ensemble to perform state estimation in ensemble-based data assimilation methods.

6. Summary and discussions

Through assimilating conventional observations including sounding, wind profiler and surface data, the performance of a WRF-based EnKF is examined in comparison to WRF-3DVar for the MCV event of 10-12 June 2003. Results show that the EnKF generally outperforms 3DVar for this particular MCV case (in terms of the forecast RMSE initialized with the posterior analysis from the previous assimilation cycle).

The impact of different data types on the performance of both data assimilation methods varies. The prior RM-DTE is respectively 4.38, 4.25, and 4.71 m/s for sounding, profiler, and surface data assimilation experiments. This is smaller than the RM-DTE of the reference forecasts (which is 5.96 m/s for the reference deterministic forecast and 4.83 m/s for the reference ensemble forecast). Similar to the OSSEs (Chapters II&III), larger improvements are seen in u, v and T than in q. In each experiment with a single observation source, the EnKF shows consistently better results than 3DVar. The prior RM-DTE at 36 h of the EnKF is about 0.5-1.5 m/s smaller than that of 3DVar. The error growth rate of the forecast initiated from the 3DVar analysis appears to be significantly larger than that from the EnKF analysis, possibly because of inappropriate error correction by the 3DVar at smaller scales where error usually grows faster. Both the EnKF and 3DVar perform generally better when sounding, profiler and surface observations are assimilated together than separately. The EnKF again performs

better than 3DVar though the difference becomes smaller than assimilating only sounding observations, suggesting that further reduction of the difference between the two methods may be expected when more observations, such as those from satellite remote sensing, are assimilated.

A series of experiments are implemented to examine the sensitivity of both the EnKF and 3DVar to different background error covariance via sounding data assimilation. Examination of the impact of weighting coefficients in the covariance relaxation procedure of the EnKF shows that a value between 0.5 and 0.7 gives apparently better results than without relaxation. Similar to the OSSE experiments (Chapter III), using different combinations of physical parameterization schemes in different ensemble members can significantly improve the EnKF performance, especially for the thermodynamic variables.

The sensitivity of 3DVar to various background error covariance methods is examined. Methods examined include the default cv3 BES option, option cv5 generated with a month-long 24- and 12-h WRF forecast differences of May 2003, and option cv5 generated with a 40-member 12-h forecast ensemble. Result shows that there are no significant performance differences among the different standard 3DVar experiments. Slight improvement is observed when the background error statistics are generated at each data assimilation time with a 12-h forecast ensemble relative to the experiments with fixed background error statistics.

To examine the role of state estimation with an ensemble mean, two ensemble-based 3DVar experiments with fixed and time-variant background error statistics are

conducted. The two experiments perform similarly to each other but apparently better than the standard 3DVar experiments. Their performance is comparable to that of the EnKF, likely due to a better prior estimate via the use of an ensemble mean. Results also demonstrate that the utilization of an ensemble rather than a deterministic forecast results in larger improvement than that obtained when one chooses a time-variant background error covariance over a fixed one. This suggests that the EnKF outperforms 3DVar through not only its flow-dependent background error covariance but also its ensemble-based state estimation.

To make a more general comparison between the EnKF and 3DVar, month-long experiments are performed in June 2003 and will be presented in Chapter V.

CHAPTER V

COMPARISON WITH 3DVAR IN A MONTH-LONG EXPERIMENT

1. Introduction

The Ensemble Kalman filter (EnKF) (Evensen 1994), which estimates the background error covariance with a short-term ensemble forecast, is drawing increasing attention in data assimilation community (Evensen 2003; Lorenc 2003; and Hamill 2006). Progress is being made lately toward implementing the **e**nsemble-based data assimilation into weather prediction models with real observations (Dowell et al. 2004, Fujita et al. 2005; Hacker and Snyder 2005; Hakim and Torn 2006; Houtekamer et al. 2005; Whitaker et al. 2004, 2006).

Application of the EnKF in real-world data assimilation with large-scale models has made large progresses. Houtekamer et al. (2005) implemented the EnKF into a large-scale model that includes a standard operational set of physical parameterization by assimilating real observations from a fairly complete observational network. The 6-hourly data assimilation cycles were conducted for about half month with the model error treated by additive random perturbations sampled from the background error statistics of the three dimensional variational data assimilation method (3DVar). The performance of the EnKF is shown comparable to that of the 3DVar system. Whitaker et al. (2004) and Compo et al. (2006) demonstrated that the EnKF is well suited to historical reanalysis problems and outperforms 3DVar. Whitaker et al. (2006) implemented the EnKF in a reduced-resolution version of the Global Forecast System

(GFS) at the National Centers for Environmental Prediction (NCEP) with all of the real-time observations used by the operational data assimilation system of NCEP except satellite radiance. Their 40-day 6-hourly data assimilation experiments show that the EnKF works better than the operational 3DVar system configured in the same way. The 48-h forecast from the ensemble data assimilation system was as accurate as the 24-h forecast from the 3DVar system. They also compared three different model error treatment methods of covariance inflation via additive (Hamill and Whitaker 2005; Houtekamer et al. 2006), multiplicative (Anderson 2001) and relaxation to the prior covariance (Zhang et al. 2004) techniques. Their result demonstrated that the additive covariance inflation method works the best for the GFS model during the time period chosen.

Real-data application of the EnKF has also been performed recently with mesoscale models. Barker (2005) implemented the EnKF in a polar numerical weather prediction model: the Antarctic Mesoscale Prediction System (AMPS, which is based on the fifth-generation Pennsylvania State University-National Center for Atmospheric Research (PSU-NCAR) Mesoscale Model (MM5)). Model error was accounted for with a multiplicative covariance inflation method. A two-week long 12-hourly cycling EnKF/AMPS assimilation/forecast system was carried out using real conventional observations from the surface, rawinsondes and satellite retrievals. The system worked reasonably well; however, the performance of the EnKF was not compared to other data assimilation methods.

In previous chapters, a mesoscale ensemble-based data assimilation system has been

shown promising in increasingly realistic environments from using simulated observations under a perfect model assumption (Zhang et al. 2006a, hereafter referred to as Chapter II) and in the presence of significant model error (Meng and Zhang 2006a, hereafter referred to as Chapter III) to a mesoscale convective vortex (MCV) case study of real-world data assimilation in Chapter IV. As a natural extension of the case study of real-world data assimilation with the EnKF implemented in WRF (Weather Research and Forecasting model), a month-long sounding data assimilation experiment is performed in this chapter to examine the behavior of the EnKF in more general manner in comparison to the WRF-3DVar. Model error is treated using the "multi-scheme" ensemble and relaxation method for covariance inflation (Zhang et al. 2004).

The rest of this chapter is presented as follows. The methodology is briefly introduced in section 2. Section 3 shows the result of reference forecasts without doing data assimilation. The consistency of the EnKF is examined in section 4. Section 5 compares the performance of the EnKF with 3DVar. Then the benefit of using the multi-scheme ensemble is demonstrated in section 6 via comparing the result of the multi-scheme to a single-scheme experiment. The impact of using different background error statistics in the 3DVar is also examined in the same section. Finally, the conclusion and discussion is given in section 7.

2. Methodology

a. The model

The same version of the Advanced Research WRF (ARW) (Skamarock et al. 2005)

with two domains and one-way nesting is used in this study as in Chapter IV. The coarse domain covers the contiguous United States with 45 × 64 grid points and a grid spacing of 90 km (Fig. 5.1), which is the same as that in Chapter IV. The inner domain is extended from covering only the central United States in Chapter IV (the dashed box in Fig.5.1) to encompassing most of the mainland U.S. with 115 × 82 grid points and a grid spacing of 30 km (solid box in Fig. 5.1). Both model domains have 27 vertical layers, and the model top is set at 100 hPa. The physical parameterization schemes include Grell-Devenyi cumulus scheme, WSM 6-class microphysics with graupel, and YSU scheme for planetary boundary layer process [see Skamarock et al. (2005) for references to different schemes]. NCEP FNL analysis data are used to create the initial and boundary conditions.

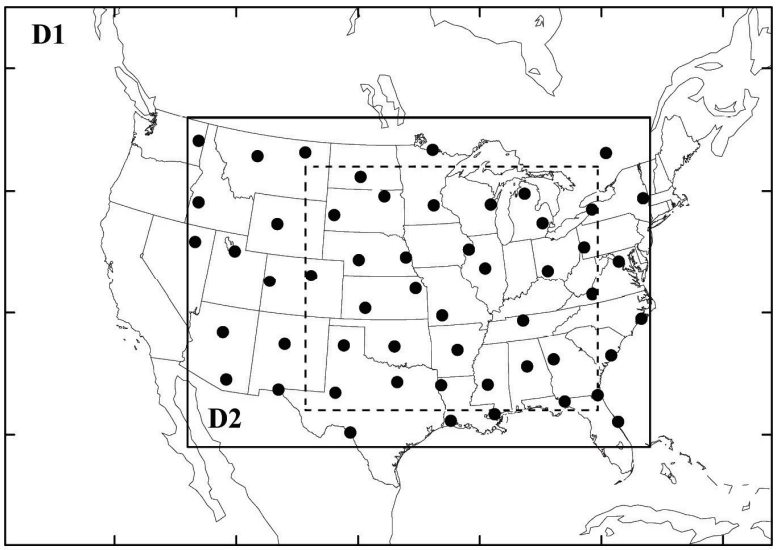

FIG. 5.1. Map of model domain. The solid circles denote the sounding observations to be assimilated. The dashed box is where the verification is done (the same as the inner domain used in the MCV case in Chapter IV).

b. The EnKF and 3DVar

The EnKF and the 3DVar method used in this study are the same as that in Chapter IV (Whitaker and Hamill 2002; Snyder and Zhang 2003; Chapter II; Chapter III; Barker et al. 2004). A detailed description can be found in sections 2b and 2c of Chapter IV. The EnKF uses an ensemble size of 40 and the covariance relaxation method (Zhang et al. 2004, their Eq.5) to inflate the background error covariance with a relaxation coefficient of 0.7. The standard 3DVar method with the background error statistics (BES) of option cv3 is compared to the EnKF in this chapter. Data assimilation is performed only in domain 2.

1) DATA TO BE ASSIMILATED

For this month-long application, only sounding observations are assimilated at a 12-h cycling because soundings work generally best compared to assimilating other data sources based on results of Chapter IV. The Gaspari and Cohn (1999) fifth-order correlation function with a radius of influence of 30 grid points (i.e., 900 km) in the horizontal direction and 15 sigma levels in the vertical direction are used for covariance localization.

The sounding observations of June 2003 go through similar quality control procedure as described in Chapter IV. The quality control contains two steps including basic quality control through the 3DVar preprocessor and so-called “errormax” checking through the prerun of 3DVar to guarantee 3DVar and the EnKF use exactly the same observations. First, 12-hourly sounding observations are processed by the observation

preprocessor of the WRF-3DVar for basic quality control such as range, domain, persistency, extreme-value, and buddy (horizontal consistency) checks. Then the FNL analysis fields at 12-h intervals are interpolated to the WRF model grid and used as the first guess in a pre-run of the WRF-3DVar to assimilate the observations that has gone through the basic quality control procedure. In the prerun, an "errormax" quality control is performed by dropping the observations whose absolute differences from the first guess are more than 5 times larger than the corresponding observation errors. The observations that pass the second quality control procedure will then be assimilated in the following 3DVar and EnKF experiments. In the output file containing the ingested observations, the original wind speed, wind direction and relative humidity are transformed into wind components u, v and mixing ratio of water vapor q. Consequently, the assimilated variables become wind components u, v, temperature T, and mixing ratio of water vapor q from soundings. The assumed observation error of soundings is given in Table 4.1. There are 53 soundings being assimilated within the inner domain; they are shown as solid circles in Fig. 5.1.

2) ENSEMBLE INITIAL AND BOUNDARY CONDITIONS

The initial ensemble is generated with the WRF-3DVar as described in section 2b of Chapter IV (Barker et al. 2004; Barker 2005) at 0000 UTC, 01 June 2003. The initial ensemble is integrated for 12 h to develop an approximately realistic, flow-dependent error covariance structure before the first data is assimilated. The boundary condition for each member is created with the perturbed initial condition and a perturbed FNL analysis

12 h after the initial time.

To provide a reasonably perturbed boundary condition for the inner domain in this month-long data assimilation and forecast cycling, the model is run in the following way. At the initial time, both the coarse and inner domains are interpolated from FNL analysis. After 12-h of the integration, sounding observations are assimilated to update the prior ensemble of inner domain. For the subsequent 12-h forecast, the initial and boundary conditions of coarse domain are regenerated via the 3DVar method based on the FNL reanalysis data at that time, while the initial condition of inner domain are just the analysis created by data assimilation. This procedure is cycled untill the end of the month. This method approximately provides a global ensemble forecast in a model configuration consistent with the inner domain. It can also prevent the model from drifting from the real atmosphere.

c. Verification metrics

Verification metrics used in this study are similar to that in Houtekamer et al. (2005) and Whitaker et al. (2006). All experiments are verified against soundings that have passed the quality control procedure, at nine standard pressure levels: 925, 850, 700, 500, 400, 300, 250, 200 and 150 hPa. To simplify the description, the root-mean-squared (RMS) fit of the verified field to sounding observations is referred to as RMSE, and the root-mean-squared ensemble standard deviation as RMS ensemble spread. The “mean” denotes an average over all verifying observations. In addition to RMSE of individual variables, RM-DTE, which combines errors of u, v and T, is also used for verification.

The DTE is defined as:

$$\mathrm{DTE} = 0.5(u'u' + v'v' + \mathrm{k}T'T'), \tag{5.1}$$

where the prime denotes the difference between the observations and the verified fields, $k = C_p/T_r$, $C_p = 1004.7$ J kg^{-1} K^{-1} and the reference temperature $T_r = 290$ K.

Another metric is bias, which is defined here as the mean of the innovation over all observations. Innovation at an observation location is the difference between the observation and the prior forecast interpolated to the observation location. To be consistent with Chapter IV, all the experiments are verified in the BAMEX field experiment domain (dashed box in Fig. 5.1, same as the inner domain in Chapter IV).

Beside the absolute RMSE and RM-DTE, we also use relative error reduction to describe the relative performance of data assimilation for different variables and different experiments. The relative improvement of A over B is defined as:

$$(\mathrm{RMSE}^B - \mathrm{RMSE}^A) / \mathrm{RMSE}^B \times 100\% . \tag{5.2}$$

3. The reference forecast

Reference forecasts are performed in both ensemble and deterministic modes, which are referred to as “EF” and “DF” respectively. The Grell-Devenyi cumulus scheme, WSM 6-class microphysics with graupel, and YSU scheme for planetary boundary layer process are used in the EF and DF runs**.** The time window is from 0000 UTC 01 June 2003 to 1200 UTC 30 June 2003.

The time evolution of 12-hourly forecast RMSE of EF and DF (Fig. 5.2) demonstrates that the model remains stable, showing no continuous error growth

throughout the month, likely due to the way we provide the boundary conditions for inner domain through updating the coarse domain with FNL analysis field every 12 hours. During the whole month, there are four apparent error peaks around 3, 12, 23, and 29 June. The horizontal distribution of RMSE of DF shows that these error peaks are in part associated with outbreaks of major convections (Davis et al. 2004, their Table 2).

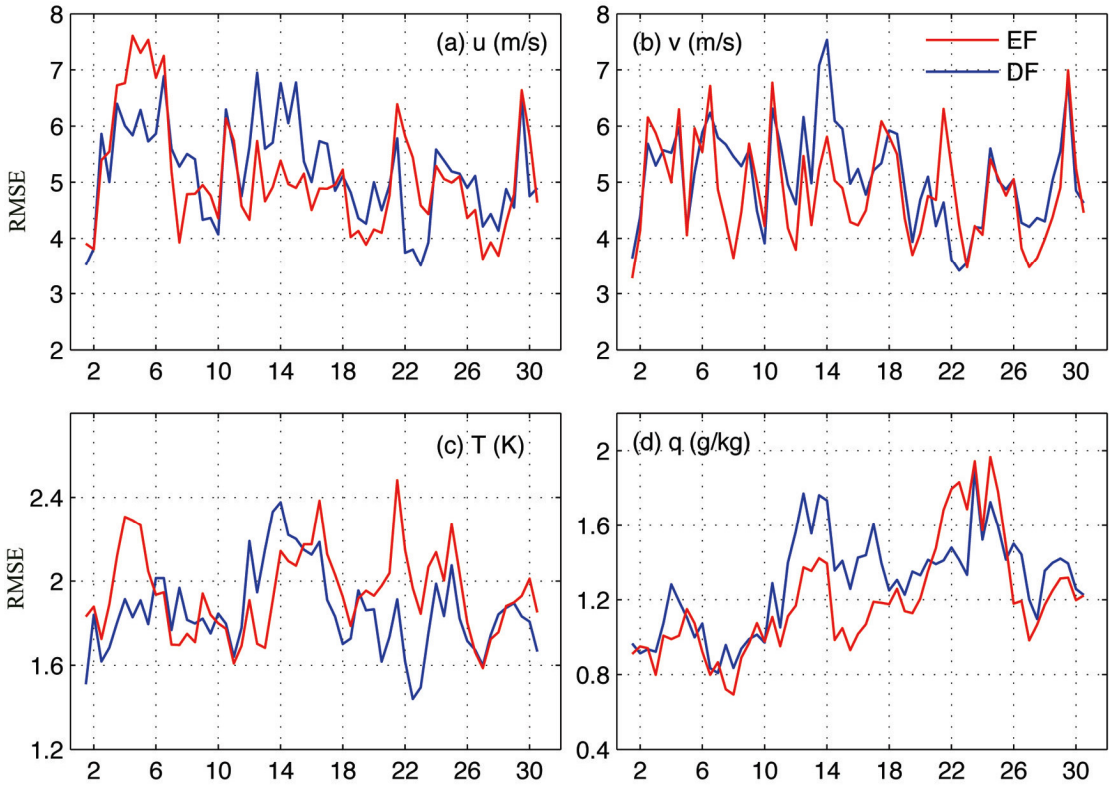

FIG. 5.2. Time evolution of domain-averaged RMSE of (a) u, (b) v, (c) T, and (d) q for the reference forecasts EF (red) and DF (blue).

It is surprising that the EF does not show consistently better forecast than DF during the month (Fig. 5.2). The EF only apparently outperforms DF during 11-16 June. However, the month-averaged error (Fig. 5.3 and Fig. 5.4) of EF is slightly smaller than

that of DF. As expected, the bias of DF is somewhat smaller than that of EF (dashed lines in Fig. 5.3).

4. Ensemble consistency of the EnKF

Due to the better performance of utilizing a multi-scheme ensemble relative to a single-scheme one as shown in Chapter IV, the control EnKF experiment for this month-long test is performed with a multi-scheme ensemble (hereafter also referred to as "EnKF_multi"). The configuration of the multi-scheme ensemble is the same as that described in Table 4.3.

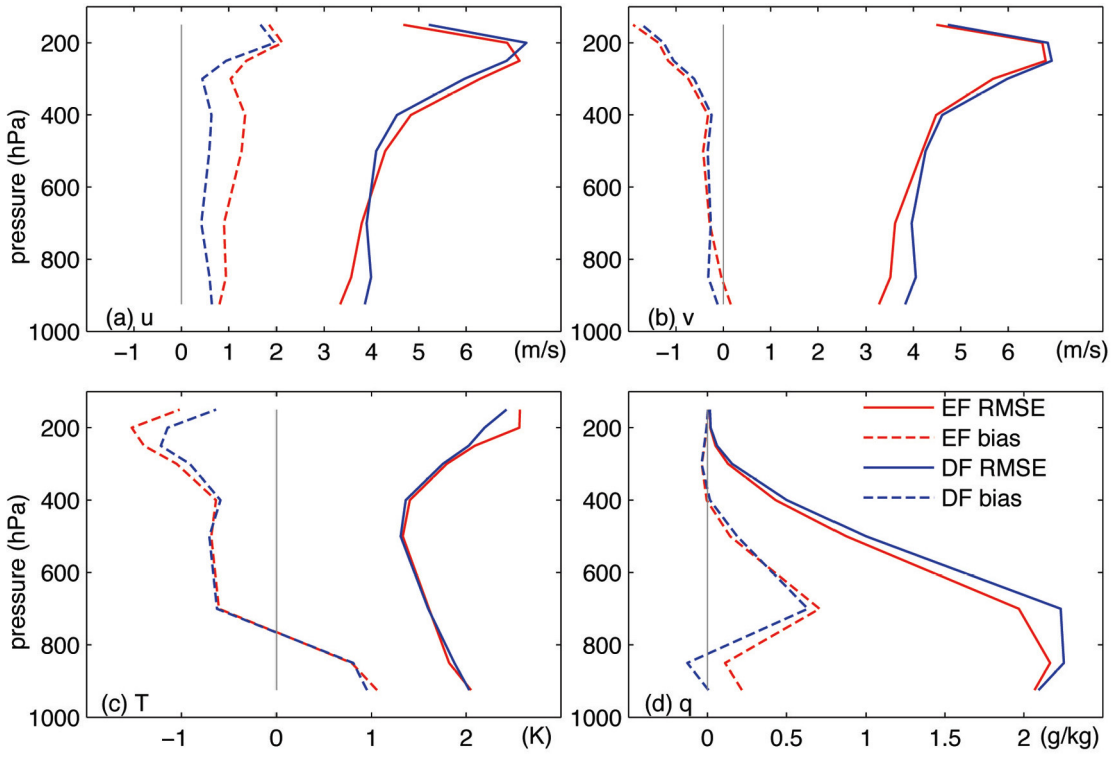

FIG. 5.3. Vertical distribution of month-averaged RMSE (solid) and bias (dashed) of (a) u, (b) v, (c) T, and (d) q for reference forecasts EF(red) and DF (blue). Zero line is also plotted in gray solid for reference.

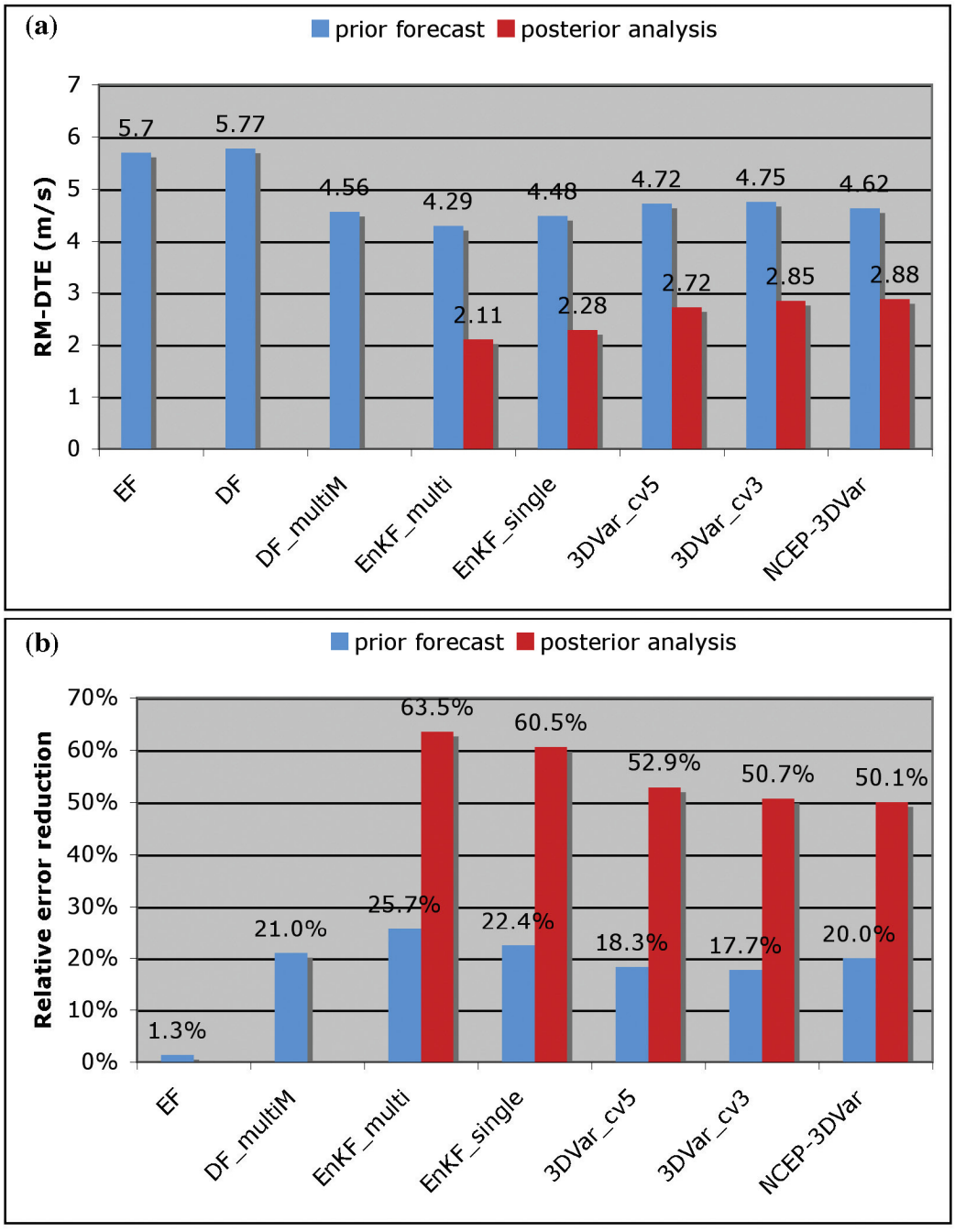

FIG. 5.4. (a) Absolute and (b) relative error reduction with respect to DF of different experiments in terms of RM-DTE.

The time window of EnKF_multi is the same as EF, from 0000 UTC 01 June 2003 untill 1200 UTC 30 June 2003. The first observation is assimilated 12 h into the integration at 1200 UTC 01 June 2003. The 12-hourly data assimilation and forecast cycle is conducted untill 1200 UTC 30 June 2003.

A common criterion to assess the optimality of an ensemble is a reasonable agreement between the innovation covariance and the sum of the background error covariance and observation error covariance assuming the observation error is independent of forecast error (Dee 1995; Houtekamer et al. 2005; Whitaker et al. 2006):

$$\left\langle (\mathbf{y}^0 - \mathbf{H}\bar{\mathbf{x}}^b)(\mathbf{y}^0 - \mathbf{H}\bar{\mathbf{x}}^b)^T \right\rangle = \mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R} \tag{5.3}$$

Similar to Houtekamer et al. (2005) and Whitaker et al. (2006), here only the diagonals of both sides of equation (5.3) are considered. In this case, if the innovation variance, which is actually the prior RMSE, is similar to the sum of background and observation error variance, which is named the "predicted RMSE" (Houtekamer et al. 2005), the prior ensemble can be regarded as consistent with the verifying observation. The discrepancy between the left and right hand sides are usually attributed to deficiency in model error parameterization (Whitaker et al. 2006).

In this experiment, the month-averaged predicted RMSE (red dot-dashed lines in Fig. 5.5) is generally larger than the observed RMSE (red solid lines in Fig. 5.5), especially below 300 hPa. The maximum difference between the observed and predicted RMSE is 1 m/s for *u* and *v* and 0.4 K for *T*. Smaller predicted RMSE is observed above 300hPa in *T* with a maximum magnitude of about 0.8 K.

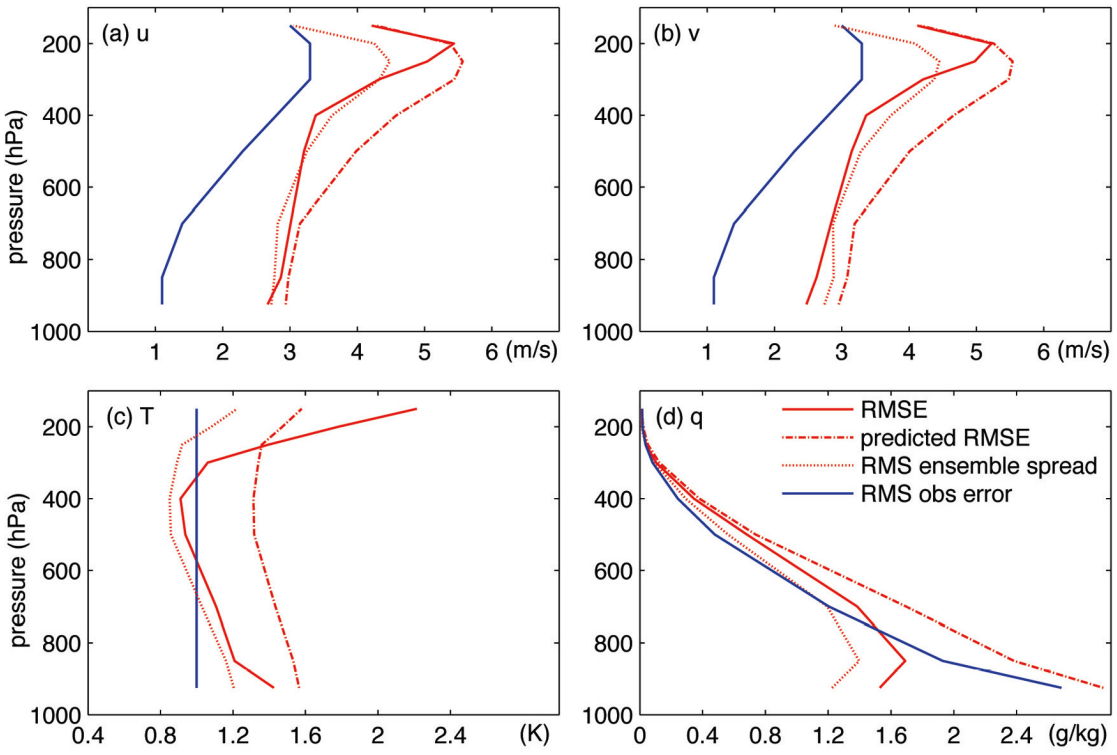

FIG. 5.5. Vertical distribution of the month-averaged prior RMSE (red solid), predicted prior RMSE (red dot-dashed), RMS observation error (blue solid), and RMS prior ensemble spread (red dotted) of (a) u, (b) v, (c) T and (d) q for EnKF_multi.

This result is comparable to other large-scale studies (Houtekamer et al. 2005; Whitaker et al. 2006). Houtekamer et al. (2005) demonstrates a closer match between both sides of Eq. 5.2 especially in the lower troposphere probably because they used the tuned 3DVar covariance to represent the model error. In Whitaker et al. (2006), larger deficiencies in the ensemble spread are observed at upper and lower levels. The deficiency at upper levels in our study is similar in magnitude to that observed in Whitaker et al. (2006). The relatively larger predicted RMSE in this study could be caused by slight overestimation of both the prior ensemble spread and the observation error, the latter being what is default in the formally released WRF-3DVar system.

Another way to assess the realism of an ensemble is by rank histograms (Hamill

2001; Hou et al. 2001). A rank histogram describes the extent to which an ensemble encompasses the verifying data by ranking the verifying data in the sorted ensemble. It describes the relative frequency of verifying observation falling into categories formed by a sorted ensemble. Suppose we have a forecast ensemble with the ensemble size of N for an observation, we first need to sort the N members by their values from the smallest to the largest. If the verifying data is smaller than the smallest value, it falls in category 1. If the observation is larger than the largest value, it falls in category N+1. If the observation is equal to or larger than the n^{th} value and smaller than the $(n+1)^{th}$ value, it falls in category n+1. The N^{th} category includes the situation when the observation is equal to the N^{th} value. The times that an observation falls in each category is then accumulated for all observations. Dividing the accumulated times of each category by the total number of the observations gives the relative frequency distribution, which is the so-called rank histogram.

The reliability of an ensemble can be diagnosed by the shape of its rank histogram. A flat shape implies that the observation could be just like a random member of the ensemble, and consequently the ensemble is reliable. A U shape suggests the ensemble spread is insufficient while a reversed U shape indicates an overestimated ensemble spread. According to Hou et al. (2001), a histogram can be regarded as fairly flat if the adjusted missing-rate is lower than 10%. The missing-rate is the sum of the relative frequencies of the two extreme (the first and last) categories. The adjusted missing-rate is defined by Zhu et al. (1996) by subtracting the expected missing-rate [2/(N+1), N is the ensemble size] from the missing-rate to give a generalized metric for the realism of

an ensemble with different sizes. The histogram can also be used to examine the bias of an ensemble. If an ensemble has a positive (negative) bias, the relative frequency of the verifying observation will be shifted toward higher (lower) categories.

The histogram created with all of the observations within the verifying domain during the month (Fig. 5.6) shows the prior ensemble of EnKF_multi is generally reliable. The rank histogram of q is based on the observations lower than 300 hPa because the mixing ratio of water vapor beyond 300 hPa is very small and the ensemble spread is so small that the rank histogram there could be meaningless. It is clear that u and v have reasonable ensemble spread, while T and q are slightly insufficient in their ensemble spread (Fig. 5.6), which is consistent with the smaller ensemble spread (red dotted lines in Fig. 5.5) relative to the corresponding RMSE. The expected missing-rate is 5% for an ensemble size of 40. The adjusted missing-rates are thus less than 10% for all u, v, T and q. Consequently, it is safe to say that the forecast ensemble in our study is generally reasonable. Note that even though the predicated RMSE is larger than the observed RMSE, there is still ensemble-spread deficiency in T and q. This result suggests that a matching between predicted and observed RMSE does not necessarily imply the optimality of the forecast ensemble in real-world data assimilation. This hypothesis will be further confirmed by the comparison between EnKF_multi and an EnKF experiment with single-scheme ensemble (same as that with reference EF) in section 6a.

5. Comparison between the EnKF and 3DVar

In comparison to EnKF_multi, one standard 3DVar (using one model instead of an ensemble) is performed using newly generated background error statistics of cv5 (hereafter also referred to as "3DVar_cv5") via the "NMC" method (Parrish and Derber 1992) which estimates the background error covariance using the statistics of differences between at least one month-long 24- and 12-h WRF-based forecast valid at the same time. 3DVar_cv5 is started from the same time as EnKF_multi. The first data are also assimilated after 12 h of integration.

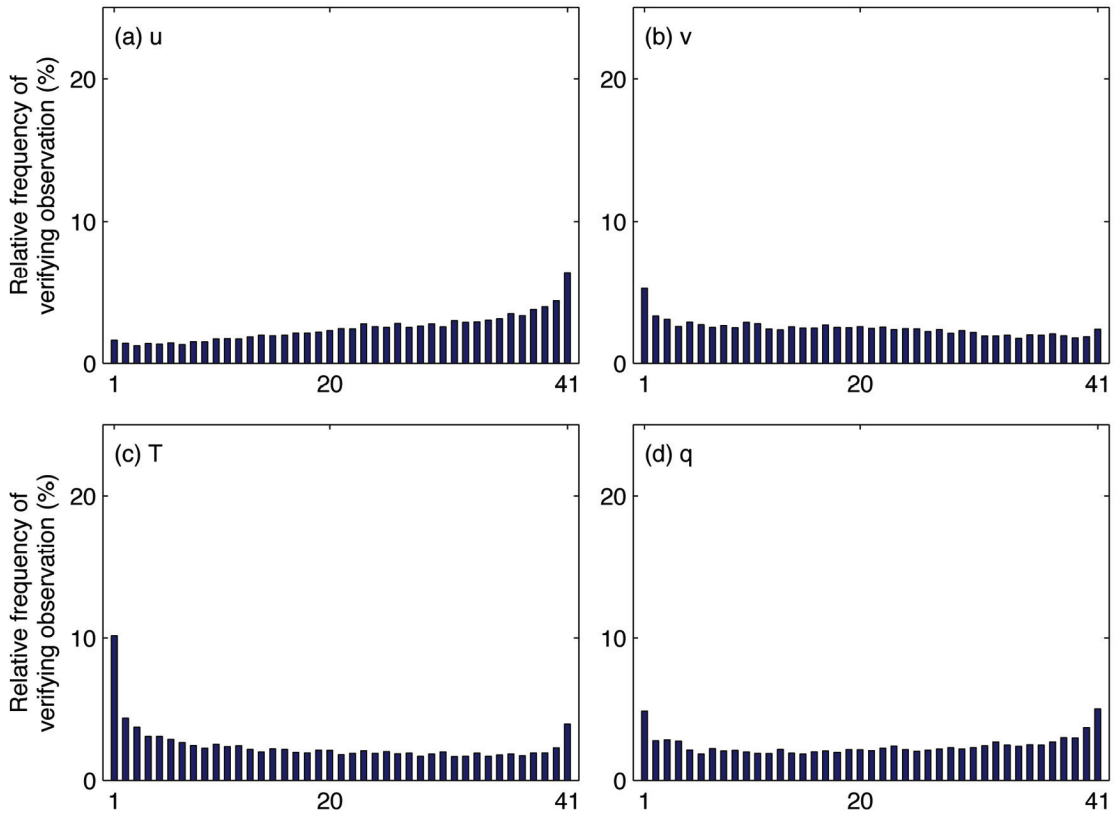

FIG. 5.6. Rank histogram for (a) u, (b) v, (c) T, and (d) q of the prior ensemble in EnKF_multi.

Time evolution of prior (solid lines in Fig. 5.7) and posterior (dot-dashed lines in Fig. 5.7) RMSE for EnKF_multi (red lines in Fig. 5.7) and 3DVar_cv5 (blue lines in Fig. 5.7) shows that both methods can decrease the prior error significantly. A striking feature is EnKF_multi performs generally better than 3DVar_cv5 in terms of the RMSE of both prior forecast and posterior analysis except for the comparable posterior RMSE in q. Similar to what was observed in previous chapters, the improvement in analysis is larger in u, v, and T than that in q.

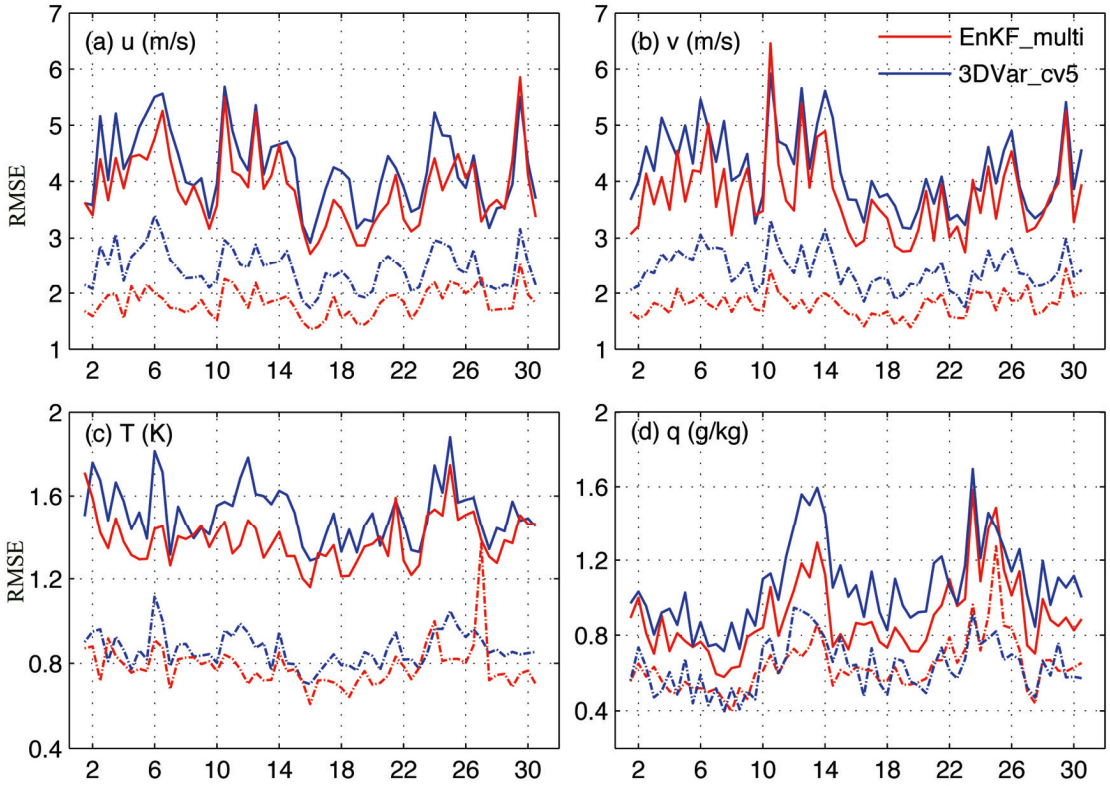

FIG. 5.7. Time evolution of domain-averaged RMSE of the 12 hourly prior forecast (solid) and posterior analysis (dot-dashed) of (a) u, (b) v, (c) T, and (d) q for EnKF_multi (red) and 3DVar_cv5 (blue).

The better performance of EnKF_multi over 3DVar_cv5 can be clearly seen on the month-averaged prior RMSE (solid lines in Fig. 5.8). EnKF_multi (red solid lines in Fig.

5.8) consistently outperforms 3DVar_cv5 (blue solid lines in Fig. 5.8) for all variables in each layer. The prior RM-DTE of EnKF_multi and 3DVar_cv5 are 4.29 m/s and 4.72 m/s respectively, which are both significantly smaller than the 5.7 m/s of EF and 5.77 m/s of DF (Fig. 5.4a). The improvement of EnKF_multi (3DVar_cv5) over DF is 25.7% (18.3%) (Fig. 5.4b). EnKF_multi outperforms 3DVar_cv5 by 9% in terms of prior RM-DTE. The improvements are different in different variables. The domain-averaged prior RMSE of EnKF_multi for u, v, T and q are 8%, 10%, 8% and 16% smaller than that of 3DVar_cv5 (Fig. 5.8). Larger improvements in q than u, v and T are likely due to reasonable correction in previous posterior estimates of q by the flow-dependent background covariance of the EnKF.

To differentiate the benefit of the EnKF from prior estimates with an ensemble mean and the flow-dependent background error covariance, "EnKF_multiM" is performed by running a one-model 12-h forecast (instead of a 12-h ensemble forecast) initiated from the mean of each posterior analysis ensemble of EnKF_multi. It is found that its forecast RMSE is between that of EnKF_multi and 3DVar_cv5 but noticeably closer to the latter (Fig. 5.4, red dashed line in Fig. 5.8, and Fig. 5.9). This result further confirms that improvement of the EnKF over 3DVar comes from both the prior estimate with an ensemble mean and the flow-dependent background error covariance but largely from the former.

The performance of EnKF_multiM and 3DVar_cv5 are also compared to that of NCEP/GFS data assimilation system (which is actually a 3DVar method) in "NCEP_3DVar". 12 hourly FNL analysis data are used as the initial field to perform 12-

h one-model forecast with the same configuration as that in EnKF_MultiM. Result shows NCEP_3DVar and EnKF_multiM have similar performance and both outperform 3DVar_cv5 (Fig. 5.9). As shown in month-averaged forecast RMSE (Fig. 5.4 and green solid in Fig. 5.8), NCEP_3DVar has generally larger error than EnKF_multi, slightly larger than EnKF_multiM and generally smaller than 3DVar_cv5.

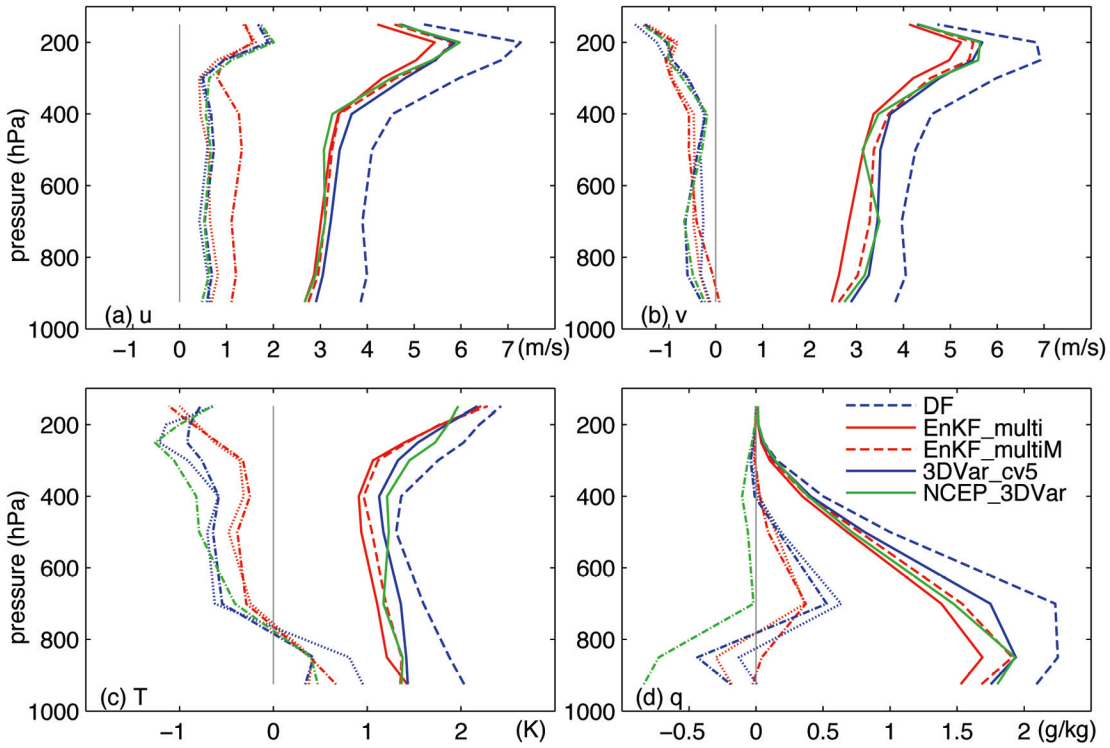

FIG. 5.8. Vertical distribution of the month-averaged prior RMSE (solid) and bias (dot-dashed) of (a) u,(b) v, (c) T, and (d) q for EnKF_multi (red), 3DVar_cv5 (blue), and NCEP_3DVar (green). The RMSE and bias for EnKF_multiM are respectively plotted in red dashed and red dotted lines. The reference forecast RMSE (blue dashed) and bias (blue dotted) for DF are also plotted for comparison. Zero line is plotted in gray solid for reference.

The bias of prior forecast (dot-dashed lines in Fig. 5.8) is significantly smaller than corresponding RMSE for both EnKF_multi and 3DVar_cv5. *U* has consistently positive

bias, which is also inferred by the slight shifting of the verifying observation toward higher categories (Fig. 5.6a). *V* has negative bias (Fig. 5.8b), which is consistent with the slight shift of the verifying observation toward lower categories (Fig. 5.6b). T has negative (positive) bias above (below) 800 hPa with the maximum bias around 200 hPa. The negative bias of T in most part of the troposphere is also reflected in the slight shift of verifying observation toward lower categories in Fig. 5.6c. For q, 3DVar_cv5 shows mixing results while EnKF_multi shows positive bias below 300 hPa which is also reflected from the rank histogram (Fig. 5.6d) from the slight shift of the verifying observation towards higher categories.

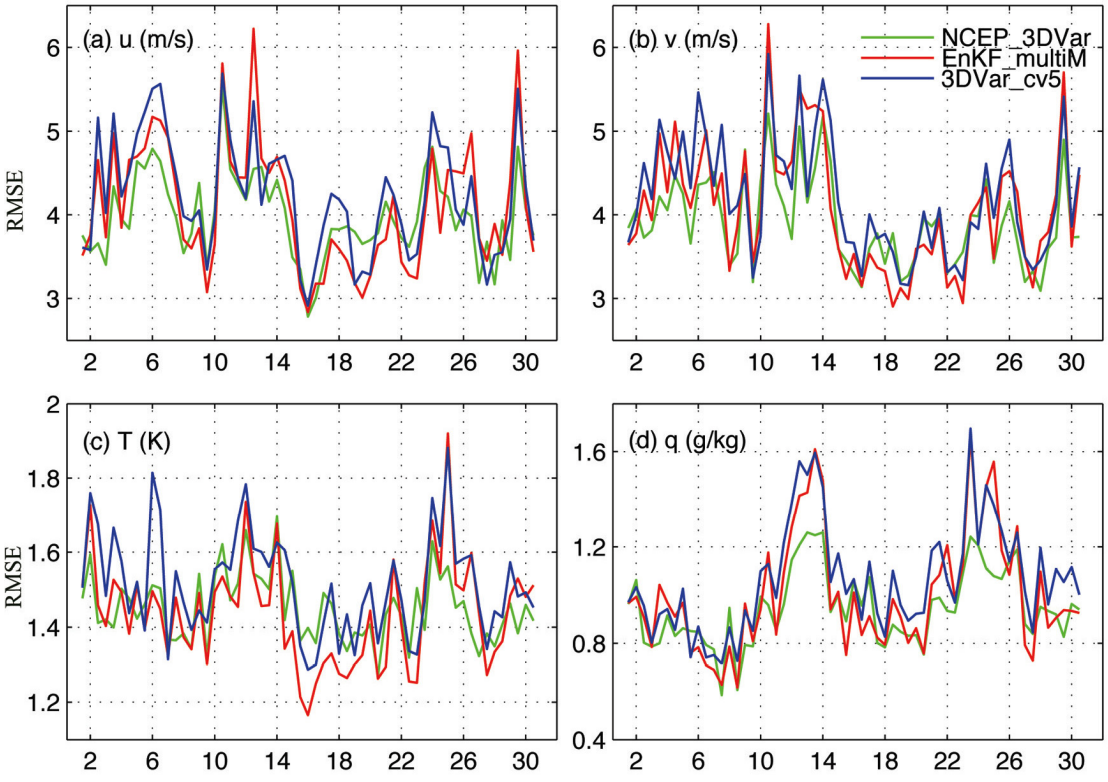

FIG. 5.9. Time evolution of domain-averaged RMSE of (a) u, (b) v, (c) T, and (d) q for the prior forecast of 3DVar_cv5 (blue), the forecast of EnKF_multiM (red) and NCEP_3DVar (green solid).

The existence of bias could affect the performance of the EnKF. Bias is usually caused by the deficiency in model error parameterization (Whitaker et al. 2006). Consequently, underestimation of the ensemble spread can be partially explained by its bias. For example, large bias in *T* (red dot-dashed in Fig. 5.8c) around the tropopause is likely associated with its smaller predicted RMSE (red dot-dashed in Fig. 5.5c) relative to the observed one. The existence of bias could partially interpret the reasonable behavior of an ensemble with an overestimated predicted RMSE.

The posterior RM-DTE of EnKF_multi is apparently smaller than that of 3DVar_cv5 and NCEP_3DVar (Fig. 5.4a), the latter two are similar to each other. EnKF_multi outperforms 3DVar_cv5 in *u* (Fig. 5.10a) and *v* (Fig. 5.10b) by about 25%. This might be caused by the larger ensemble spread relative to the observation error shown in Figs. 5.5a and 5.5b. The improvement in *T* is about 9%. The two methods perform similarly in *q* probably due to the insufficient ensemble spread of EnKF_multi (Figs. 5.5d, and 5.6d). The posterior bias (dot-dashed in Fig. 5.10) is apparently smaller than that of the prior (dotted in Fig. 5.10) for both methods. The posterior bias of EnKF_multi is consistently smaller than that of 3DVar_cv5.

6. Sensitivity to background error covariance

a. Comparison between multi- and single-scheme ensembles in the EnKF

"EnKF_single" is performed in this section with the same physical schemes as that in the EF to compare with EnKF_multi. EnKF_single shows a larger RMSE (blue vs. red solid lines in Fig. 5.11), smaller RMS ensemble spread (dotted lines in Fig. 5.11) and

consequently smaller predicted RMSE (dot-dashed lines in Fig. 5.11) for all *u*, *v*, *T* and *q* in each layer than EnKF_multi. This result further confirms that the closer distance between the observed and predicted RMSE as shown in EnKF_single does not necessarily imply a smaller prior RMSE as indicated in section 4c for this month-long experiment. The relatively larger predicted RMSE partially contributed by the larger RMS ensemble spread could help to decrease the prior RMSE.

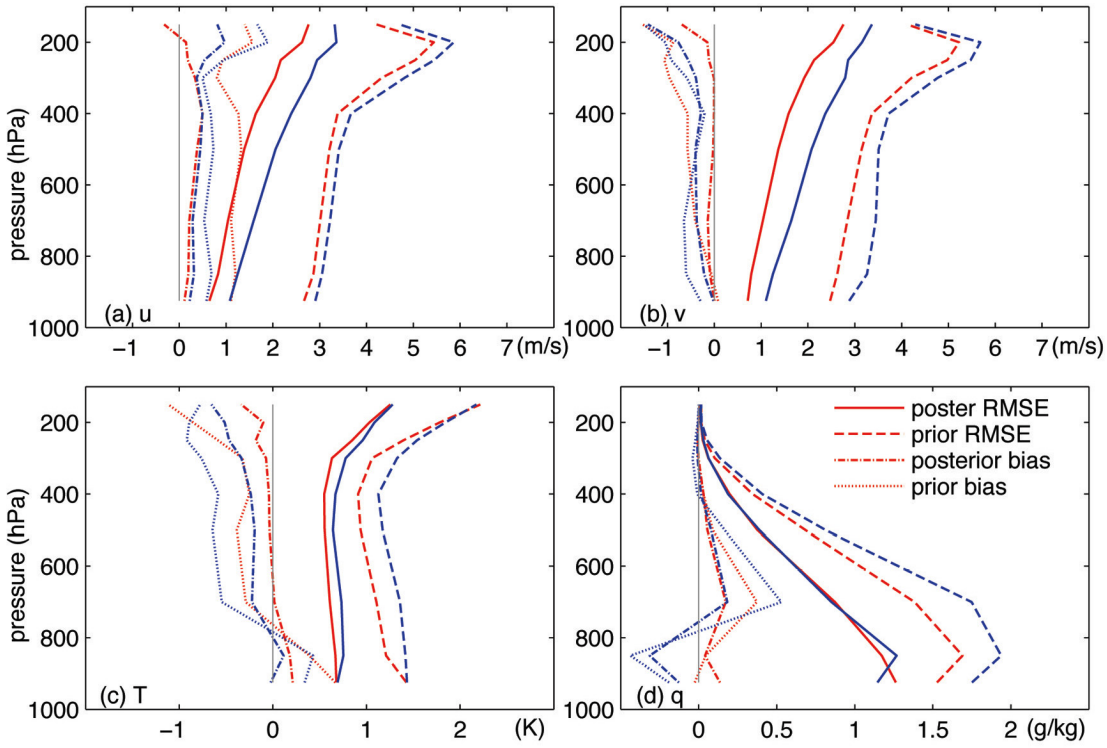

FIG. 5.10. Vertical distribution of the month-averaged posterior RMSE (solid) and bias (dot-dashed) of (a) u, (b) v, (c) T, and (d) q for EnKF_multi (red) and 3DVar_cv5 (blue). The prior RMSE (dashed) and their bias (dotted) for EnKF_multi (red) and 3DVar_cv5 (blue) are also plotted for comparison. Zero line is plotted in gray solid for reference.

EnKF_single increases the prior RM-DTE by 0.2 m/s (Fig. 5.4) relative to EnKF_multi. It works generally worse for all variables especially for *T* and *q* (blue vs. red solid lines in Fig. 5.12). The prior RMSEs of EnKF_multi are respectively 3%, 3%,

8% and 6% smaller than those of EnKF_single for *u*, *v*, *T* and *q*. The larger improvement of EnKF_multi in *T* and *q* could be partially explained by the consistent decrease of bias from EF, to EnKF_single and then to EnKF_multi (dashed lines in Figs. 5.12c and 5.12d). The prior bias of *u* and *v* are comparable among the three experiments. Better performance due to the multi-scheme ensemble relative to the single-scheme ensemble can also be seen in terms of the posterior RMSE (dot-dashed lines in Fig. 5.12) and the corresponding bias (dotted lines in Fig. 5.12). Similar to what is seen in prior error, there are also larger improvement in *T* (12%) and *q* (10%) than in *u* (6%) and *v* (6%) in terms of posterior analysis of EnKF_multi relative to EnKF_single.

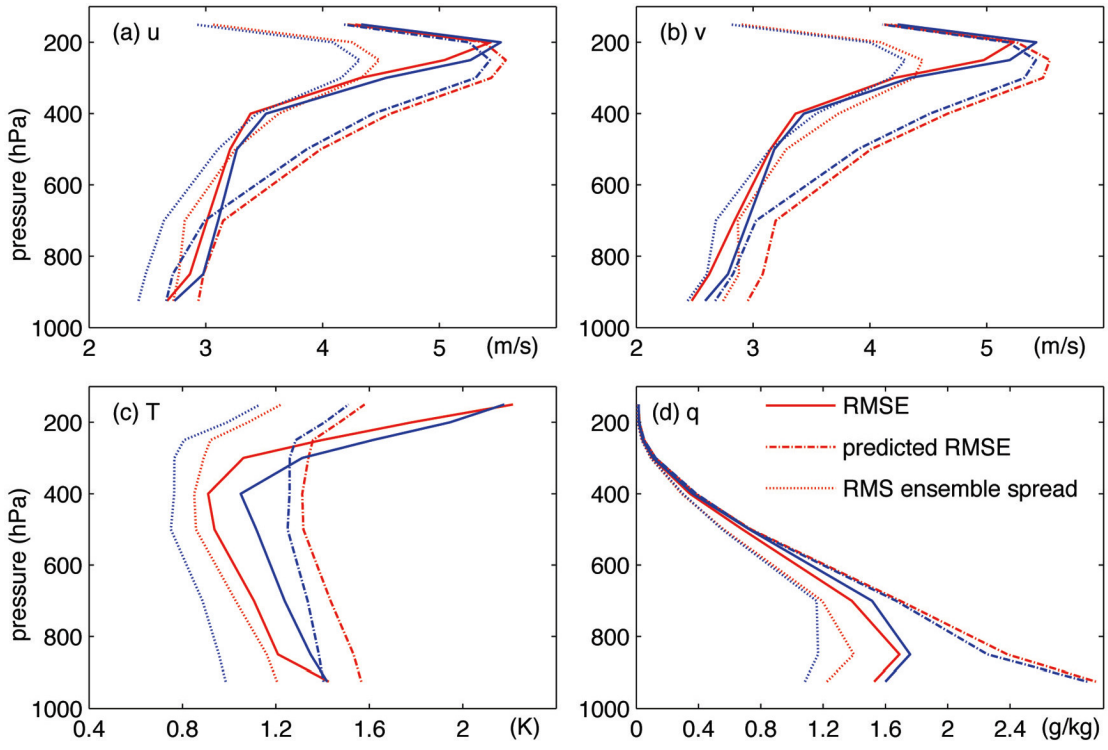

FIG. 5.11. Vertical distribution of the month-averaged RMSE (solid), predicted RMSE (dot-dashed), and RMS ensemble spread (dotted) of the prior forecast for (a) u, (b) v, (c) T, and (d) q in EnKF_multi (red) and EnKF_single (blue).

b. Comparison between cv5 and cv3 in 3DVar

Similar to what is seen in Chapter IV, 3DVar_cv5 performs similarly to that with default BES option of cv3 (hereafter referred to also as "3DVar_cv3") (red vs. blue in Fig. 5.13). 3DVar_cv5 works slightly better than 3DVar_cv3 in terms of overall prior RM-DTE (Fig. 5.4). The biases of these two 3DVar experiments (dashed lines in Fig. 5.13) are also similar to each other.

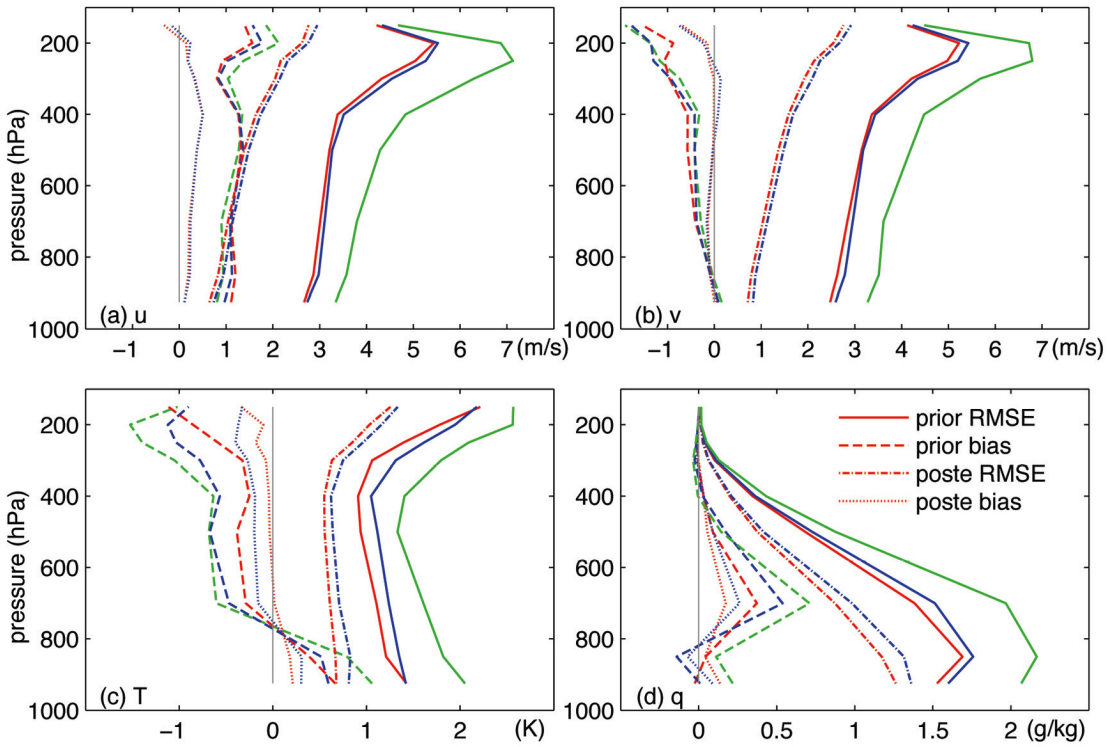

FIG. 5.12. Vertical distribution of month-averaged RMSE (solid for prior forecast and dot-dashed for posterior analysis) and bias (dashed for prior forecast and dotted for posterior analysis) of (a) u, (b) v, (c) T, and (d) q for EnKF_multi (red) and EnKF_single (blue). The RMSE (green solid) and bias (green dashed) for reference forecast EF are also plotted for comparison. Zero line is plotted in gray solid for reference.

c. Comparison between different initiating times for the EnKF

To further investigate the stability of the EnKF system, a five-day experiment

"EnKF_multi_5days" is performed starting from 1200 UTC 10 June to 0000 UTC 15 June with the same multi-scheme configuration as in EnKF_multi. The result shows that the experiments starting from different times are generally comparable to each other (blue vs. red lines in Fig. 5.14). Both experiments are similar to the MCV case study (green lines in Fig. 5.14) in terms of prior RMSE, while the posterior RMSE of the MCV case study are noticeably larger.

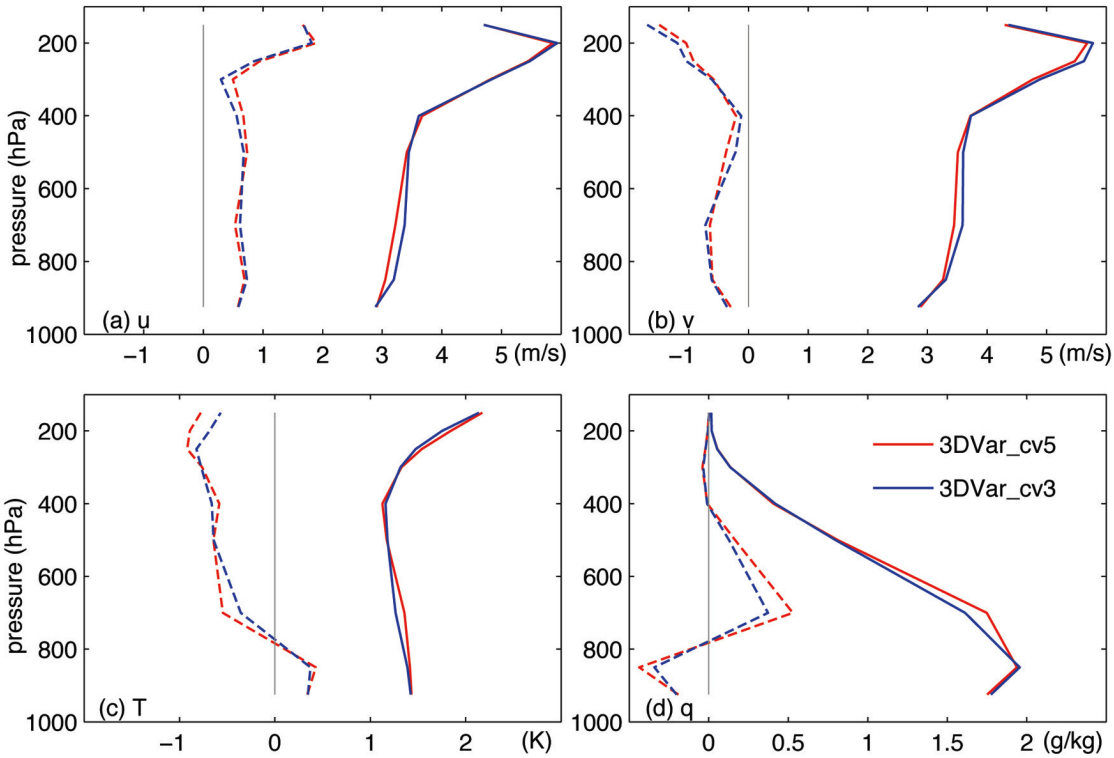

FIG. 5.13. Vertical distribution of month-averaged prior RMSE (solid) and bias (dashed) of (a)u, (b)v, (c) T, and (d) q for 3DVar_cv5 (red) and 3DVar_cv3 (blue). Zero line is plotted in gray solid for reference.

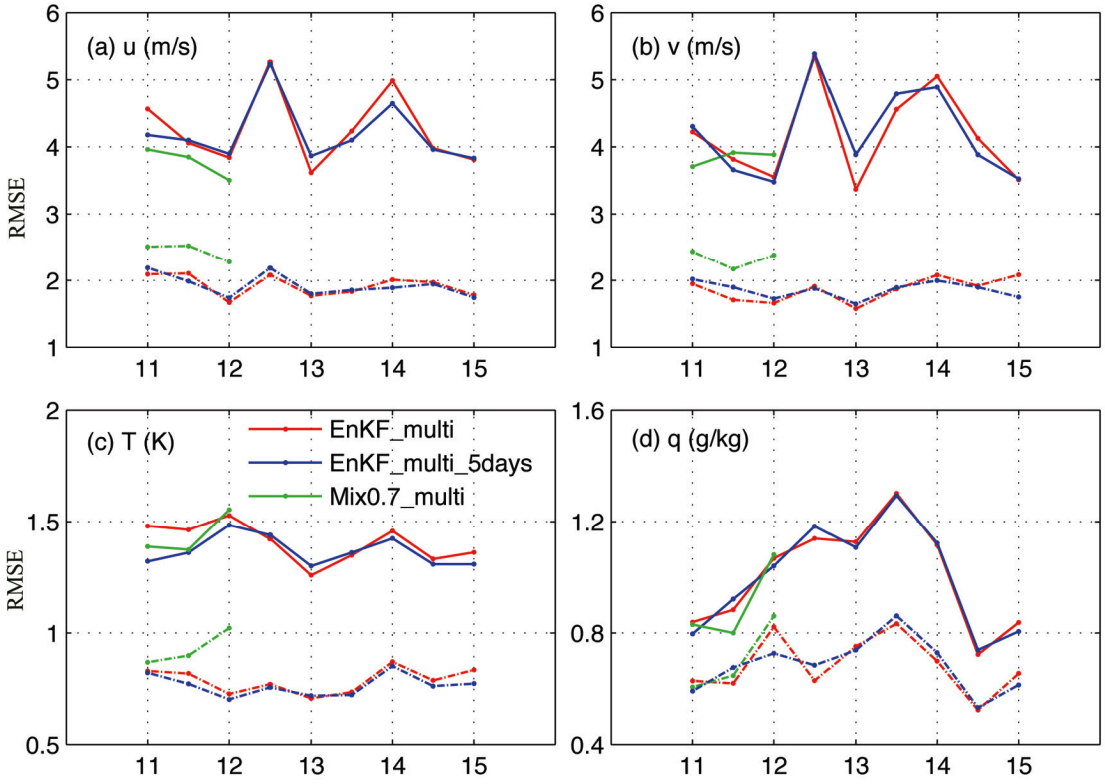

FIG. 5.14. Time evolution of domain-averaged RMSE of prior forecast (solid) and posterior analysis (dot-dashed) of (a) u, (b) v, (c) T, and (d) q for EnKF_multi_5days (blue), Mix0.7_multi from the MCV case study in Chapter IV (green), and EnKF_multi valid in the same time period (red).

Similar performance between EnKF_multi and EnKF_multi_5days is also demonstrated in the horizontal distribution of 10-m wind vector, mean sea level pressure and simulated radar reflectivity at 0000 UTC 12 June 2003 (Figs. 5.15d and 5.15c). The pattern of the simulated radar reflectivity of these two experiments are slightly better than that of Mix0.7_multi in Chapter IV (Fig. 5.15b) and fairly close to the observed radar echo (Fig. 5.15a). This is likely due to the updating of initial fields for the coarse domain every 12 hours in the experiments of this chapter, while in the Mix0.7_multi, only the boundary condition of the coarse domain is updated at 12-h interval. Result of this section also indicates that 12-h pre-run is long enough to provide a reasonable

background error covariance structure.

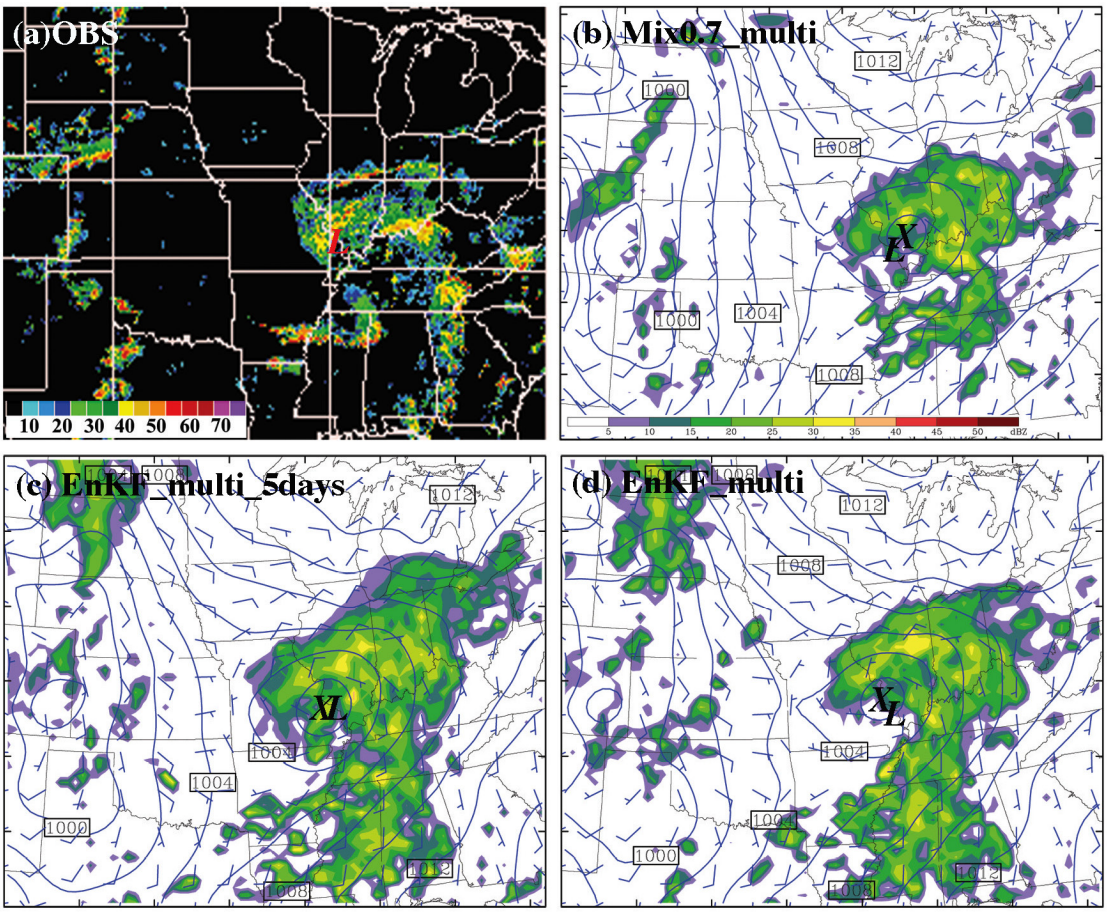

FIG. 5.15. (a) The observed radar echo, the prior forecast of MSLP (every 2 hPa), 10-m wind vector (full barb 5 m/s) and simulated reflectivity (shaded) valid at 0000 UTC June 12, 2003 of (b) Mix0.7_Multi in Chapter IV, (c) EnKF_multi_5days, and (d) EnKF_multi. The big ***X*** and ***L*** respectively denote the simulated and observed MCV centers at surface.

7. Summary and discussion

Month-long performance of a WRF-based EnKF is examined in comparison to WRF-3DVar by conducting 12 hourly in-situ sounding data in June 2003. Multi-scheme ensemble is utilized to account for model error in the EnKF. 3DVar uses a newly generated background error statistics via "NMC" method.

Stable performance of both the EnKF and 3DVar are achieved by providing an approximate global ensemble forecast for the data assimilation domain. Specifically, a new initial ensemble is created for the coarse domain in 12-h interval based on the GFS analysis field, while the inner domain is continuously updated by the observational data through forecast and data assimilation cycling.

The result shows that the prior ensemble of the EnKF is generally reasonable. In our case, the sum of RMS ensemble spread and observation error does not have to be comparable to the observed RMSE to achieve a smaller forecast error.

It is found that the EnKF works efficiently in drawing both the forecast and analysis close to the observations. Relative to the reference forecast without data assimilation, 26% (64%) error reduction is achieved in the prior forecast (posterior analysis). The EnKF works apparently better than 3DVar for all variables in each layer of the troposphere for this particular time period. The EnKF outperforms 3DVar by about 9% (25%) in terms of prior forecast (posterior analysis) RM-DTE. Larger improvements are observed in q than in *u, v and T*. Compared to the prior RMSE, the corresponding bias is fairly small in magnitude. It is also found that the EnKF seems to benefit more from the ensemble-based prior estimate than from using a flow-dependent background error covariance.

The benefit of using multi-scheme ensemble is demonstrated clearly due to a resulting reasonable inflation of prior ensemble spread. Apparent smaller prior errors are observed relative to the single-scheme ensemble in all variables in each layer, especially for T and q. The better performance in T and q could also be partially explained by their

smaller bias resulted from using multi-schemes. Similar to Chapter IV, the performance of the 3DVar with background error covariance of cv3 and cv5 are similar to each other.

Similar error statistics and MCV structures are observed for a particular period of time between experiments starting at different times. This result further confirms the stability of the EnKF system, suggesting that long-term experiment is not very sensitive to the initial condition and 12-h pre run is long enough to provide a reasonable background error covariance structure.

CHAPTER VI

SUMMARY AND DISCUSSION

This doctoral study demonstrates the feasibility of a mesocale ensemble Kalman filter (EnKF) with both simulated and real-world observations.

Through observing system simulation experiments (OSSEs), the performance of the EnKF under the perfect model assumption in which the truth simulation is produced with the same model and same initial uncertainties as those of the ensemble is examined (Zhang et al. 2006a). The EnKF is implemented in a non-hydrostatic, mesoscale model (MM5) to assimilate simulated sounding and surface observations derived from simulations of the "surprise" snowstorm of January 2000. It is found that the EnKF is very effective in keeping the analysis close to the truth simulation under the perfect model assumption.

Power spectrum analyses of OSSEs show that the EnKF is most effective in reducing larger-scale errors but less effective in reducing errors at smaller, marginally resolvable scales. A 24-h continuous EnKF assimilation of sounding and surface observations of typical temporal and spatial resolutions is found to reduce the error by as much as 80% (compared to a 24-h forecast without data assimilation) for both observed and unobserved variables including zonal and meridional winds, temperature, and pressure. However, it is observed to be relatively less efficient in correcting errors in the vertical velocity and moisture fields that have larger power at smaller-scales. The analysis domain-averaged root-mean-square error after 24-h assimilation is ~1.0-1.5 ms^{-1}

for winds and ~1.0 K for temperature, which is comparable to or less than specified observational errors. Various sensitivity experiments demonstrated that the EnKF is quite successful in all realistic observational scenarios tested.

The performance of the EnKF is also examined in the presence of significant model errors from using imperfect physical parameterizations by assimilating synthetic sounding and surface observations with typical temporal and spatial resolutions for the snowstorm case of January 2000 and a mesoscale convective vortex (MCV) event of 10-12 June 2003 (Meng and Zhang 2006a).

The significance of model error in both warm- and cold-season events is demonstrated when the use of different cumulus parameterization schemes within different ensembles results in significantly different forecasts in terms of both ensemble mean and spread. Nevertheless, the EnKF performed reasonably well in most experiments with the imperfect model assumption (though its performance can sometimes be significantly degraded). Similar to what is observed under perfect model assumption, most analysis error reduction comes from larger scales. Results show that using a combination of different physical parameterization schemes in the ensemble forecast (so called "multi-scheme ensemble") can significantly improve filter performance. A multi-scheme ensemble has the potential to provide better background error covariance estimation and a smaller ensemble bias.

There are noticeable differences in the performance of the EnKF for different flow regimes. In the imperfect scenarios considered, the improvement over the reference ensembles (pure ensemble forecasts without data assimilation) after 24 hours of

assimilation for the winter snowstorm event ranges from 36% to 67%. This is higher than the 26% to 45% improvement noted after 36 hours of assimilation for the warm-season MCV event. Scale- and flow-dependent error growth dynamics and predictability are possible causes for the differences in improvement. Compared to the power spectrum analyses for the snowstorm, it is found that forecast errors and ensemble spreads in the warm-season MCV event have relatively smaller power at larger scales and an overall smaller growth rate.

As a natural extension of previous OSSE studies, the performance of the EnKF is investigated then by assimilating real-data observations for the MCV event of 10-12 June 2003 (Meng and Zhang 2006b). Direct comparison between the EnKF and a three-dimensional variational (3DVar) data assimilation system, both implemented in the Weather Research and Forecasting model (WRF), is carried out.

It is found that the EnKF performs consistently better than the 3DVar method when assimilating either individual or multiple data sources (i.e., sounding, surface and wind profiler) for this MCV event. The EnKF outperforms 3DVar likely due to both using an ensemble mean for state estimates and using a flow-dependent background error covariance. Background error covariance plays an important role in the performance of both the EnKF and the 3DVar system. Proper covariance inflation and using different combinations of physical parameterization schemes in different ensemble members (the so-called “multi-scheme” ensemble) can significantly improve the EnKF performance for this particular MCV case. The 3DVar system can benefit substantially from using short-term ensembles to improve the prior estimate (with the ensemble mean).

Noticeable improvement is also achieved by including some flow dependence in the background error covariance of 3DVar.

To make more general comparison between the EnKF and 3DVar, a month-long test is performed in June 2003. It is found that the EnKF generally outperforms the 3DVar method for the time period of interest. Relative to the reference forecast without data assimilation, 26% (64%) error reduction can be achieved in the prior forecast (posterior analysis) of the EnKF. The prior forecast error of the EnKF is smaller than that of 3DVar by 9%. Utilization of a multi-scheme ensemble can apparently decrease the error especially in temperature and moisture. Similar error features are observed for a particular period of time with different lead times, which indicates that the 12-h pre-forecast is long enough to develop a reasonable background error covariance structure.

Result of this study demonstrates great potential of the EnKF for future use in operational data assimilation practice. On the other hand, though the EnKF has been shown promising in different realistic scenarios, there are many open questions remaining to answer (Zhang et al. 2006). Above all, model error is the most difficult issue to deal with. In this study, despite we account for the model error with a multi-scheme ensemble, background error covariance inflation and localization, there is still noticeable deficiency in ensemble spread as shown by the U shape of rank histogram of the forecast ensemble. Another issue is observation error. What we use in this study is based on climatological statistics and somewhat simple. The representative part of observation error, which is model-dependent, is not considered. Besides, only synoptic, sounding and profiler observations are assimilated in this work. Problems could occur in

the assimilation of high-resolution remote-sensed data such as radar and satellite radiance. In those cases, the higher spatial resolution may invalidate the assumption of independent observation error and make the EnKF ill posed in principle. Though the EnKF is regarded as one of the candidates (the other is four dimensional variational data assimilation method (4DDA)) for future operational data assimilation method, considering various issues involved in the EnKF algorithm, a better solution could be coupling the EnKF with the variational method to accomplish the best possible weather forecast.

APPENDIX

DERIVATION OF THE KALMAN FILTER IN LEAST SQUARES SENSE

Notations needed to do this derivation are as follows:

M : dynamic model

$\mathbf{M}$: tangent linear of the dynamic model

$\mathbf{y}$: observation vector

$\mathbf{x}^{b}$: background state vector

$\mathbf{x}^{a}$: analysis state vector

$\mathbf{x}^{t}$: truth

H : observation operator

$\mathbf{H}$: tangent linear of observation operator mapping from model space to observation space

$\boldsymbol{\eta}$: model error $M(\mathbf{x}_{t-1}^{t}) - \mathbf{x}_{t}^{t}$

$\boldsymbol{\varepsilon}$: observation error $\mathbf{y} - H(\mathbf{x}^{t})$

$\boldsymbol{\delta}$: background error $\mathbf{x}^{b} - \mathbf{x}^{t}$

$\boldsymbol{\mu}$: analysis error $\mathbf{x}^{a} - \mathbf{x}^{t}$

$\mathbf{R}$: observation error covariance matrix $\left\langle \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{T} \right\rangle$

(<> means expectation, superscript T denotes transpose)

$\mathbf{P}^{b}$: background error covariance matrix $\left\langle \boldsymbol{\delta\delta}^{T} \right\rangle$

$\mathbf{P}^{a}$: analysis error covariance matrix $\left\langle \boldsymbol{\mu\mu}^{T} \right\rangle$

$\mathbf{Q}$: model error covariance matrix $\left\langle \boldsymbol{\eta\eta}^{T} \right\rangle$

Kalman filter has following basic assumptions:

(1) Linearized model, namely the variations of model prediction in the vicinity of the forecast state is linear, which can be expressed as $M(\mathbf{x}_{t}) - M(\mathbf{x}_{t-1}) = \mathbf{M}(\mathbf{x}_{t} - \mathbf{x}_{t-1})$.

(2) Linearized observation operator, namely the variations of observation operator in the vicinity of the background state is linear, which can be expressed as $H(\mathbf{x}) - H(\mathbf{x}^{b}) = \mathbf{H}(\mathbf{x} - \mathbf{x}^{b})$.

(3) Model error is white and has Gaussian distribution with zero mean.

(4) Observation error is white and has Gaussian distribution with zero mean.

(5) Background error has Gaussian distribution with zero mean.

(6) Model error, observation error and background error are mutually independent.

(7) $\mathbf{Q}$, $\mathbf{R}$ and $\mathbf{P}^{b}$ are positive definite.

Kalman filter contains two steps: the forecast step and the analysis step.

In the forecast step, suppose we have the analysis $\mathbf{x}_{t-1}^{a}$ and its associated error covariance $\mathbf{P}_{t-1}^{a}$, the subscript indicates time step of t-1, we propagate both with the dynamic model M to time step t.

$$\mathbf{x}_{t}^{b} = M(\mathbf{x}_{t-1}^{a}). \tag{A.1}$$

The truth at time t can be expressed as

$$\mathbf{x}_t^t = M(\mathbf{x}_{t-1}^t) - \boldsymbol{\eta}. \tag{A.2}$$

Then the background error can be calculated by

$$\begin{aligned} \boldsymbol{\delta}_t &= \mathbf{x}_t^b - \mathbf{x}_t^t \\ &= M(\mathbf{x}_{t-1}^a) - M(\mathbf{x}_{t-1}^t) + \boldsymbol{\eta} \\ &= M(\mathbf{x}_{t-1}^a - \mathbf{x}_{t-1}^t) + \boldsymbol{\eta} \\ &= \mathbf{M}(\boldsymbol{\mu}_{t-1}) + \boldsymbol{\eta}. \end{aligned} \tag{A.3}$$

The background error covariance is then given by

$$\begin{aligned} \mathbf{P}_t^b &= \left\langle \boldsymbol{\delta}_t \boldsymbol{\delta}_t^T \right\rangle \\ &= \left\langle (\mathbf{M}\boldsymbol{\mu}_{t-1} + \boldsymbol{\eta})[\mathbf{M}\boldsymbol{\mu}_{t-1} + \boldsymbol{\eta}]^T \right\rangle \\ &= \left\langle \mathbf{M}\boldsymbol{\mu}_{t-1}\boldsymbol{\mu}_{t-1}^T\mathbf{M}^T + \mathbf{M}\boldsymbol{\mu}_{t-1}\boldsymbol{\eta}^T + \boldsymbol{\eta}\boldsymbol{\mu}_{t-1}^T\mathbf{M}^T + \boldsymbol{\eta}\boldsymbol{\eta}^T \right\rangle \\ &= \mathbf{M}\left\langle \boldsymbol{\mu}_{t-1}\boldsymbol{\mu}_{t-1}^T \right\rangle \mathbf{M}^T + \left\langle \boldsymbol{\eta}\boldsymbol{\eta}^T \right\rangle \\ &= \mathbf{M}\mathbf{P}_{t-1}^b\mathbf{M}^T + \mathbf{Q}. \end{aligned} \tag{A.4}$$

With $\mathbf{x}_t^b$ and $\mathbf{P}_t^b$ in hand, we can then perform analysis with observation $\mathbf{y}_t$ by seeking the analysis at time t through a linear combination of $\mathbf{x}_t^b$ and $\mathbf{y}_t$ as

$$\mathbf{x}_t^a = \mathbf{K}_1\mathbf{x}_t^b + \mathbf{K}\mathbf{y}_t. \tag{A.5}$$

Since the observation and background state vector are assumed unbiased, so the analysis is consequently unbiased, which means $\langle \boldsymbol{\mu}_t \rangle = 0$.

$$
\begin{aligned}
\langle \boldsymbol{\mu}_t \rangle &= \left\langle \mathbf{x}_t^a - \mathbf{x}_t^t \right\rangle \\
&= \left\langle \mathbf{K}_1 \mathbf{x}_t^b + \mathbf{K}\mathbf{y}_t - \mathbf{x}_t^t \right\rangle \\
&= \left\langle \mathbf{K}_1 (\mathbf{x}_t^t + \boldsymbol{\delta}_t) + \mathbf{K}(H(\mathbf{x}_t^t) + \boldsymbol{\varepsilon}) - \mathbf{x}_t^t \right\rangle \\
&= \left\langle (\mathbf{K}_1 + \mathbf{K}H - \mathbf{I})\mathbf{x}_t^t + \mathbf{K}_1 \boldsymbol{\delta}_t + \mathbf{K}\boldsymbol{\varepsilon} \right\rangle \\
&= \left\langle (\mathbf{K}_1 + \mathbf{K}H - \mathbf{I})\mathbf{x}_t^t \right\rangle \\
&= (\mathbf{K}_1 + \mathbf{K}H - \mathbf{I}) \left\langle \mathbf{x}_t^t \right\rangle \\
&= 0,
\end{aligned}
\tag{A.7}
$$

which indicates $\mathbf{K}_1 + \mathbf{K}H - \mathbf{I} = 0$, or $\mathbf{K}_1 = \mathbf{I} - \mathbf{K}H$. Substitute $\mathbf{K}_1$ to (A.5) we get the expression of analysis as

$$
\mathbf{x}_t^a = \mathbf{x}_t^b + \mathbf{K}[\mathbf{y}_t - H(\mathbf{x}_t^b)]. \tag{A.8}
$$

Then, using the independent relation between observation error and model error, the analysis covariance can be calculated as

$$
\begin{aligned}
\mathrm{P}_t^a &= \left\langle \boldsymbol{\mu}_t \boldsymbol{\mu}_t^{\mathrm{T}} \right\rangle \\
&= \left\langle (\mathbf{x}_t^a - \mathbf{x}_t^t)(\mathbf{x}_t^a - \mathbf{x}_t^t)^{\mathrm{T}} \right\rangle \\
&= \left\langle \{\mathbf{x}_t^b + \mathbf{K}[\mathbf{y}_t - H(\mathbf{x}_t^b)] - \mathbf{x}_t^t\} \{\mathbf{x}_t^b + \mathbf{K}[\mathbf{y}_t - H(\mathbf{x}_t^b)] - \mathbf{x}_t^t\}^{\mathrm{T}} \right\rangle \\
&= \left\langle \{\mathbf{x}_t^b - \mathbf{x}_t^t + \mathbf{K}[H(\mathbf{x}_t^t) + \boldsymbol{\varepsilon}] - \mathbf{K}H(\mathbf{x}_t^t + \boldsymbol{\delta}_t)]\} \{\mathbf{x}_t^b - \mathbf{x}_t^t + \mathbf{K}[H(\mathbf{x}_t^t) + \boldsymbol{\varepsilon}] - \mathbf{K}H(\mathbf{x}_t^t + \boldsymbol{\delta}_t)]\}^{\mathrm{T}} \right\rangle \\
&= \left\langle \{\boldsymbol{\delta}_t - \mathbf{K}[H(\mathbf{x}_t^t + \boldsymbol{\delta}_t) - H(\mathbf{x}_t^t)] + \mathbf{K}\boldsymbol{\varepsilon}\} \{\boldsymbol{\delta}_t - \mathbf{K}[H(\mathbf{x}_t^t + \boldsymbol{\delta}_t) - H(\mathbf{x}_t^t)] + \mathbf{K}\boldsymbol{\varepsilon}\}^{\mathrm{T}} \right\rangle \\
&= \left\langle \{\boldsymbol{\delta}_t - \mathbf{K}\mathbf{H}\boldsymbol{\delta}_t + \mathbf{K}\boldsymbol{\varepsilon}\} \{\boldsymbol{\delta}_t - \mathbf{K}\mathbf{H}\boldsymbol{\delta}_t + \mathbf{K}\boldsymbol{\varepsilon}\}^{\mathrm{T}} \right\rangle \\
&= \left\langle [(\mathbf{I} - \mathbf{K}\mathbf{H})\boldsymbol{\delta}_t + \mathbf{K}\boldsymbol{\varepsilon}][(\mathbf{I} - \mathbf{K}\mathbf{H})\boldsymbol{\delta}_t + \mathbf{K}\boldsymbol{\varepsilon}]^{\mathrm{T}} \right\rangle \\
&= \left\langle (\mathbf{I} - \mathbf{K}\mathbf{H})\boldsymbol{\delta}_t \boldsymbol{\delta}_t^{\mathrm{T}} (\mathbf{I} - \mathbf{K}\mathbf{H})^{\mathrm{T}} + \mathbf{K}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}} \right\rangle \\
&= (\mathbf{I} - \mathbf{K}\mathbf{H}) \left\langle \boldsymbol{\delta}_t \boldsymbol{\delta}_t^{\mathrm{T}} \right\rangle (\mathbf{I} - \mathbf{K}\mathbf{H})^{\mathrm{T}} + \mathbf{K} \left\langle \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}} \right\rangle \mathbf{K}^{\mathrm{T}} \\
&= (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}_t^b (\mathbf{I} - \mathbf{K}\mathbf{H})^{\mathrm{T}} + \mathbf{K}\mathbf{R}\mathbf{K}^{\mathrm{T}}.
\end{aligned}
\tag{A.9}
$$

Then we need to find a $\mathbf{K}$ that minimizes the variance of the analysis by letting

$$\frac{\partial}{\partial \mathbf{K}} trace(\mathbf{P}_t^a) = \frac{\partial}{\partial \mathbf{K}} trace[(\mathbf{I}-\mathbf{KH})\mathbf{P}_t^b(\mathbf{I}-\mathbf{KH})^T + \mathbf{KRK}^T] = 0. \quad (A.10)$$

$$\frac{\partial\{trace[(\mathbf{I}-\mathbf{KH})\mathbf{P}_t^b(\mathbf{I}-\mathbf{KH})^T]\}}{\partial(\mathbf{I}-\mathbf{KH})} \frac{\partial(\mathbf{I}-\mathbf{KH})}{\partial \mathbf{K}} + \frac{\partial\{trace(\mathbf{KRK}^T)\}}{\partial \mathbf{K}} = 0. \quad (A.11)$$

By doing matrix differentiation, we will end up with

$$-2(\mathbf{I}-\mathbf{KH})\mathbf{P}_t^b\mathbf{H}^T + 2\mathbf{KR} = 0. \quad (A.12)$$

The solution for K is then given by

$$\mathbf{K} = \mathbf{P}_t^b\mathbf{H}^T(\mathbf{HP}_t^b\mathbf{H}^T + \mathbf{R})^{-1}. \quad (A.13)$$

By plugging $\mathbf{K}$ into (A.8) and (A.9), we can get the analysis and the associated error covariance at time t. By repeating the whole procedure forward in time, observations can be assimilated sequentially to attain an improved evolution of the system.

REFERENCES

Aksoy A., F. Zhang, J. W. Nielsen-Gammon, 2006a: Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea-breeze model. *Mon. Wea. Rev.*, **134**, 2951-2970.

Aksoy A., F. Zhang, J. W. Nielsen-Gammon, 2006b: Ensemble-based simultaneous state and parameter estimation with MM5. *Geophys. Res. Letters*, **33**, L12801, doi:10.1029/2006GL026186.

Aksoy A., F. Zhang, J. W. Nielsen-Gammon, and C. C. Epifanio, 2005: Ensemble-based data assimilation for thermally forced circulations. *J. Geophys. Res.*, **110**, D16105, doi:10.1029/2004JD005718.

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884-2903.

Anderson, J. L., and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741-2758.

Arakawa, A., 2004: The cumulus parameterization problem: past, present and future. *J. Climate*, **17**, 2493-2525.

Barker, D. M., W. Huang, Y.-R. Guo, A.J. Bourgeois, and Q. N. Xiao, 2004: A three-dimensional variational data assimilation system for MM5: Implementation and initial results. *Mon. Wea. Rev.*, **132**, 897-914.

Barker, D., W. Huang, Y.-R. Guo and A. J. Bourgeois, 2003: A three-dimensional variational (3DVAR) data assimilation system for use with MM5. *NCAR Technical Note,* NCAR/TN-453+STR, 68pp.

Barker, D. M., 2005: Southern high-latitude ensemble data assimilation in the Antarctic mesoscale prediction system. *Mon. Wea. Rev.*, **133**, 3431-3449.

Benjamin, S., B. E. Schwartz, E. J. Szoke, and S. E. Koch, 2004: The value of wind profiler data in U.S. weather forecasting: case studies. *Bull. Amer. Meteor. Soc.*, **85,** 1871-1886**.**

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719-1724.

Caya, A., J. Sun, and C. Snyder, 2005: A comparison between the 4DVAR and the ensemble Kalman filter techniques for radar data assimilation. *Mon. Wea. Rev.*, **133**, 3081–3094.

Chessa, P. A., G. Ficca, M. Marrocu, and R. Buizza, 2004: Application of a limited-area short-range ensemble forecast system to a case of heavy rainfall in the Mediterranean region. *Wea. Forecasting*, **19,** 566–581

Compo, G. P., J. S. Whitaker, and P. D. Sardeshmukh, 2006: Feasibility of a 100 year reanalysis using only surface pressure data. *Bull. Amer. Meteor. Soc.*, **87**, 175-190.

Cotton, W. R., and R. A. Anthes, 1989*: Storm and Cloud Dynamics*. Academic Press, 883pp.

Daley, R. 1991: *Atmospheric Data Analysis*. Cambridge University Press, 457 pp.

Daley, R., and R. Menard, 1993: Spectral characteristics of Kalman filter systems for atmospheric data assimilation. *Mon. Wea. Rev*., **121**, 1554-1565.

Davis, C. A., and S. B. Trier, 2006: Mesoscale convective vortices observed during BAMEX. Part I: Kinematic and thermodynamic structure. *Mon. Wea. Rev.,* in review.

Davis, C. A., and co-authors, 2004: The Bow-Echo and MCV Experiment (BAMEX): observations and opportunities. *Bull. Amer. Meteor. Soc.*, **85**, 1075-1093.

Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.*, **123**, 1128–1144.

Dirren, S., R. D. Torn, and G. J. Hakim, 2006: A data assimilation case study using a limited-area ensemble Kalman filter. *Mon. Wea. Rev.*, in press.

Done, J., C. Davis, and M. Weisman, 2003: The next generation of NWP: Explicit forecasts of convection using Weather Research and Forecasting (WRF) model. *Atmos. Sci. Lett.*, DOI: 10.1002/asl.72.

Dowell, D. C., F. Zhang, L. J. Wicker, C. Snyder, and N. A. Crook, 2004: Wind and temperature retrievals in the 17 May 1981 Arcadia, Oklahoma, supercell: Ensemble

Kalman filter experiments. *Mon. Wea. Rev.*, **132**, 1982–2005.

Dudhia, J., 1993: A nonhydrostatic version of the Penn State–NCAR mesoscale model: Validation tests and simulation of an Atlantic cyclone and cold front. *Mon. Wea. Rev.*, **121**, 1493-1513.

Evans, R. E., M. S. J. Harrison, R. J. Graham, K.R. Mylne, 2000: Joint medium-range ensembles from the Met. Office and ECMWF systems. *Mon. Wea. Rev.*, **128**, 3104-3127.

Evensen, G., 2003: The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics,* **53**, 343-367.

Evensen, G., 1997: Advanced data assimilation for strongly nonlinear dynamics. *Mon. Wea. Rev.*, **125**, 1342-1354.

Evensen, G., and P. J. van Leeuwen, 1996: Assimilation of Geosat altimeter data for the Agulhas Current using the ensemble Kalman filter with a quasigeostrophic model. *Mon. Wea. Rev.*, **124**, 85-96.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10143-10162.

Fujita, T., D. J. Stensrud, and D. C. Dowell, 2005: Surface data assimilation using an ensemble Kalman filter approach with initial condition and model physics uncertainties. *Preprint of the 11th AMS Conference on Mesoscale Processes*, **1M3**, Albuquerque, New Mexico, 24-29 October 2005.

Gaspari, G. and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723-757.

Grell, G. A., 1993: Prognostic evaluation of assumptions used by cumulus parameterizations. *Mon. Wea. Rev.*, **121**, 5-31.

Grell, G. A., and D. Devenyi, 2002: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophys. Res. Letters*, **29,** 38-1 – 38-4.

Grell, G. A., J. Dudhia and D. R. Stauffer, 1994: A description of the fifth-generation Penn State / NCAR mesoscale model (MM5). *NCAR Technical Note*, **NCAR/TN-**

398+STR, 117pp.

Hacker, J. P., and C. Snyder, 2005: Ensemble Kalman filter assimilation of fixed screen-height observations in a parameterized PBL. *Mon. Wea. Rev.,* **133**, 3260–3275.

Hakim, G. J., and R. D. Torn, 2006: Ensemble Synoptic Analysis. *Fred Sanders Monograph*, American Meteorological Society, accepted.

Hamill, T. M., 2006: Ensemble-based atmospheric data assimilation. Chapter 6 of *Predictability of Weather and Climate*, Cambridge Press, 124-156.

Hamill, T. M., and J. S. Whitaker, 2005: Accounting for the error due to unresolved scales in ensmeble data assimilation: a comparison of different approaches. *Mon. Wea. Rev.,* **133,** 3132-3147.

Hamill, T. M., C. Snyder, and J. S. Whitaker, 2003: Ensemble forecasts and the properties of flow-dependent analysis-error covariance singular vectors. *Mon. Wea. Rev,* **130**, 1741-1758.

Hamill, T. M., C. Snyder, and R. E. Morss, 2002: Analysis-error statistics of a quasi-geostrophic model using 3-dimensional variational assimilation. *Mon. Wea. Rev,* **130**, 2777-2791.

Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev*., **129**, 550-560.

Hamill, T. M., and C. Snyder, 2000: A hybrid ensemble Kalman filter-3D variational analysis scheme. *Mon. Wea. Rev.,* **128**, 2905-2919.

Hansen, J. A., 2002: Accounting for model error in ensemble-based state estimation and forecasting. *Mon. Wea. Rev.,* **130**, 2373-2391.

Hawblitzel, D. P., F. Zhang, Z. Meng, and C. A. Davis, 2006: Probabilistic evaluation of the dynamics and predictability of a meoscale convective vortex event of 10-13 June 2003. *Mon. Wea. Rev.,* in press.

Houtekamer, P. L., H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Speak, and B. Hansen, 2005: Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Wea. Rev.,* **133**, 604-620.

Houtekamer, P. L., and H. L. Mitchell, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.,* **129**, 123-137.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.,* **126**, 796-811.

Houtekamer, P. L., and L. Lefaivre, 1997: Using ensemble forecasts for model validation. *Mon. Wea. Rev.,* **125**, 2416-2426.

Houtekamer, P. L., L. Lefaivre, J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.,* **124**, 1225-1242.

Keppenne C. L., 2000: Data assimilation into a primitive-equation model with a parallel Ensemble Kalman Filter. *Mon. Wea. Rev.*, **128**, 1971-1981

Keppenne C. L, and M. M. Rienecker, 2003: Assimilation of temperature into an isopycnal ocean general circulation model using a parallel Ensemble Kalman Filter. *J. Mar. Sys.,* **40-41**, 363-380.

Keppenne, C. L. and M. M. Rienecker, 2002: Initial testing of a massively parallel ensemble Kalman filter with the Poseidon isopycnal ocean general circulation model. *Mon. Wea. Rev.*, **130**, 2951-2965.

Kolmogorov, A. N., 1962: Interpolation and extrapolation of stationary random sequences, translated by W. Doyle and J. Selin, Rept. RM-3090-PR, RAND Corp., Santa Monica, Calif.

Krishnamurti, T. N., C. M. Krishtawal, Z. Zhang, T. LaRow, D. Bachiochi, E. Williford, S. Gadgil, and S. Surendran, 2000: Multimodel ensemble forecasts for weather and seasonal climate. *J. Climate,* **13**, 4196-4216.

Lee M. S., and D. M. Barker, 2005: Background error statistics in 3DVAR derived from WRF ensembles. *WRF/MM5 users' workshop*, 3.20, Boulder, 27-30 June 2005.

Lorenc, A. C.，2003: The potential of the ensemble Kalman filter for NWP-a comparison with 4Dvar. *Quart. J. Roy. Meteor. Soc.*, **129**, 3183-3203.

Lorenz, E. N., 1963: Deterministic non-periodic flow, *J. Atmos. Sci.*, **20**, 130-141.

Mellor, G. L., and T. Yamada, 1982: Development of a turbulence closure model for

geophysical fluid problems. *Rev. Geophys. Space Phys.*, **20**, 851-875.

Meng, Z., and F. Zhang, 2006a: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation, Part II: Imperfect model experiments. *Mon. Wea. Rev.* in press.

Meng, Z., and F. Zhang, 2006b: Test of an ensemble-Kalman filter for mesoscale and regional-scale data assimilation. Part III: Comparison with 3DVar for a real-data case study. *Mon. Wea. Rev.,* in review.

Mitchell H. L., and P. L. Houtekamer, 2002: Ensemble size, balance, and model-error representation in an ensemble Kalman filter. *Mon. Wea. Rev.,* **130**, 2791-2808.

Mitchell H. L., and P. L. Houtekamer, and G. Pellerin, 2002: Ensemble size, balance, and model-error representation in an ensemble Kalman filter. *Mon. Wea. Rev.,* **130**, 2791-2808.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.,* **122**, 73–119.

Molinari, J., and T. Corsetti, 1985: Incorporation of cloud-scale and mesoscale downdrafts into a cumulus parameterization: Results of one- and three-dimensional integrations. *Mon. Wea. Rev.,* **113**, 485-501.

Molinari, J., and M. Dudek, 1992: Parameterization of convective precipitaion in mesoscale numerical models: a critical review. *Mon. Wea. Rev.,* **120**, 326-344.

Orrell, D., 2002: Role of the metric in forecast error growth: how chaotic is the weather? *Tellus*, **54A**, 350-362.

Orrell, D., L. A. Smith, J. Barkmeijer, and T. N. Palmer, 2001: Model error in weather forecasting. *Nonlin. Proc. Geophys.*, **8**, 357-371.

Palmer, T. N., 2001: A nonlinear dynamical perspective on model error: a proposal for non-local stochastic-dynamic parameterization in weather and climate prediction models. *Quart. J. Roy. Meteor. Soc.*, **127**, 279-304.

Parrish, D. F., J. C. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747-1763.

Pham, D. T., 2001: Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Wea. Rev.,* **129**, 1194-1207.

Reisener, J., R. J. Rasmussen, and R. T. Bruintjes, 1998: Explicit forecasting of supercooled liquid water in winter storms using the MM5 mesoscale model. *Quart. J. Roy. Meteor. Soc.*, **124B**, 1071-1107.

Simmons, A. J., and A. Hollingsworth, 2002: Some aspects of the improvement in skill of numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **128**, 647-677.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers, 2005: A description of the advanced research WRF version 2. *NCAR Technical Note*, NCAR/TN-468+STR.

Snyder, C., and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **131**, 1663-1677.

Snyder, C., T. M. Hamill, and S. B. Trier, 2003: Linear evolution of error covariances in a quasigeostrophic model. *Mon. Wea. Rev.,* **131**, 189–205.

Sorenson H. W., 1970: Least-squares estimation: from Gauss to Kalman, *IEEE Spectrum*, **7**, 63-68.

Stensrud, D. J., J.-W. Bao and T. T. Warner, 2000: Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Wea. Rev.,* **128**, 2077-2107.

Talagrand, O., 1997: Assimilation of observations, an introduction. *J. Meteor. Soc. Japan*, **75**, 191-209.

Tong, M., and M. Xue, 2005: Ensemble Kalman filter assimilation of Doppler Radar data with a compressible nonhydrostatic model: OSS experiments. *Mon. Wea. Rev.,* **133**, 1789-1807.

Torn, R. D., G. J. Hakim, and C. Snyder, 2006: Boundary conditions for limited-area ensemble Kalman filters. *Mon. Wea. Rev.*, **134,** 2490--2502.

Trier, B. Stanley, and Chris A. Davis, 2006: Mesoscale convective vortices observed during BAMEX. Part II: Influences on secondary deep convection. *Mon. Wea. Rev.,* in review.

Trier, B. Stanley, C. A. Davis, D. A. Ahijevych, M. L. Weisman, and G. H. Bryan, 2006: Mechanisms supporting long-lived episodes of propagating nocturnal convection within a 7-day WRF model simulation. *J. Atmos. Sci.*, in review.

Van Leeuwen, P. J., 1999: Comment on "Data assimilation using an ensemble Kalman filter techniques". *Mon. Wea. Rev.*, **127**, 1374-1377.

Wang W., and, N. L. Seaman, 1997: A comparison study of convective parameterization schemes in a mesoscale model. *Mon. Wea. Rev.,* **125**, 252-278.

Whitaker, J. S., T. M. Hamill, X. Wei, Y. Song, and Z. Toth, 2006: Ensemble data assimilation with the NCEP global forecast system, *Mon. Wea. Rev*., in review.

Whitaker, J. S., G. P. Compo, X. Wei, and T. M. Hamill, 2004: Reanalysis without radiosondes using ensemble data assimilation. *Mon. Wea. Rev.,* **132**, 1190-1200.

Whitaker J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.,* **130**, 1913-1924.

Whitaker, J. S., T. M. Hamill, X. Wei, Y. Song, and Z. Toth, 2006: Ensemble data assimilation with the NCEP global forecast system. *Mon. Wea. Rev*., in review.

Wiener, N., 1949: *The Extrapolation, Interpolation and Smoothing of Stationary Time Series*, New York: Wiley.

Xiao Q., J. Sun, W. Lee, E. Lim, Yong-Run. Guo, D. M. Barker, Ying-Hwa Guo, 2003: Assimilation of Doppler Radar Observations with a Regional 3DVar System: A Heavy Rainfall Case Study. *31st International Conference on Radar Meteorology*, Seattle, WA, 5-12 August 2003.

Ying-Hwa Kuo, 2003: Mesoscale numerical weather prediction, presented at the *Harold D. Orville Symposium*, 26 April 2003, Rapid City, SD.

Zhang, F., 2005: Dynamics and structure of mesoscale error covariance of a winter cyclone estimated through short-range ensemble forecasts. *Mon. Wea. Rev.*, **133**, 2876-2893.

Zhang, F. and C. Snyder, 2006: Ensemble-based data assimilation. *Bull. Amer. Meteor. Soc.*, in press.

Zhang, F., Z. Meng, and A. Aksoy, 2006a: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation, Part I: Perfect model experiments. *Mon. Wea. Rev.*, **134**, 722-736.

Zhang, F., A. Odins, and J. W. Nielsen-Gammon, 2006b: Mesoscale predictability of an extreme warm-season precipitation event. *Wea. Forecasting*, **21**, 149-166.

Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238-1253.

Zhang, F., C. Snyder, and R. Rotunno, 2003: Effects of moist convection on mesoscale predictability. *J. Atmos. Sci.*, **60**, 1173-1185.

Zhang, F., C. Snyder, and R. Rotunno, 2002: Mesoscale predictability of the 'surprise' snowstorm of 24-25 January 2000. *Mon. Wea. Rev.,* **130**, 1617-1632.

Zhang, S., and J. L. Anderson, 2003: Impact of spatially and temporally varying estimates of error covariance on assimilation in a simple atmospheric model. *Tellus*, **55A**, 126-147.

Zhu, Y., G. Iyengar, Z. Toth, M. S. Tracton, and T. Marchok, 1996: Objective evaluation of the NCEP global ensemble forecasting system. Preprints, *15th Conf. on Weather Analysis and Forecasting,* Norfolk, VA, Amer. Meteor. Soc., J79–J82.

VITA

Name: Zhiyong Meng

Address: RM 1204 Eller O&M BLDG, TAMU 3150, College Station, TX, 77843-3150

Email Address: zmeng@tamu.edu

Education

Texas A&M University, Atmospheric Sciences, Ph.D., 2007
Chinese Academy of Meteorological Sciences, Atmospheric Sciences, M.S., 1994
Beijing University, Atmospheric Sciences, B.S., 1991

Professional Experiences

2004	Student visitor, National Center for Atmospheric Research
2003–2007	Graduate Research/Teaching Assistant, Texas A&M University
2001–2003	Research associate, Chinese Academy of Meteorological Sciences
1995	Visiting scholar on numerical prediction of tropical cyclones at the Japan Meteorological Agency, Tokyo, Japan
1994–2001	Research assistant, Chinese Academy of Meteorological Sciences
1991–1994	Graduate research assistant, Chinese Academy of Meteorological Sciences

Selected Publications

Meng, Z, and F. Zhang, 2006: Test 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*, in review.

Meng, Z, and F. Zhang, 2006: Test of an ensemble-Kalman filter for mesoscale and regional-scale data assimilation. Part II: Imperfect-model experiments. *Monthly Weather Review*, in press.

Hawbliztel, D., F. Zhang, Z. Meng and C. A. Davis, 2006: Probabilistic evaluation of the dynamics and predictability of mesoscale convective vortex event of 10-13 June 2003. *Monthly Weather Review*, in press.

Zhang, F., Z. Meng and A. Aksoy, 2006: Test of an ensemble-Kalman filter for mesoscale and regional scale data assimilation. Part I: Perfect-model experiments. *Monthly Weather Review*, 134, 722-736.