UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

DEVELOPMENT OF ARPS LETKF WITH FOUR DIMENSIONAL EXTENSION AND INTERCOMPARISON WITH ARPS ENSRF

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

GANG ZHAO
Norman, Oklahoma
2013

DEVELOPMENT OF ARPS LETKF WITH FOUR DIMENSIONAL EXTENSION
AND INTERCOMPARISON WITH ARPS ENSRF

A DISSERTATION APPROVED FOR THE
SCHOOL OF METEOROLOGY

BY

Dr. Ming Xue, Chair

Dr. Kelvin Droegemeier

Dr. Lance Leslie

Dr. Robert Palmer

Dr. S. Lakshimivarahan

Dr. Xuguang Wang

This dissertation is dedicated to my family.

Acknowledgements

It is my honor and pleasure to express my gratitude to a large number of people who helped me complete this dissertation and Ph.D. study at all levels.

First of all, I wholeheartedly thank my advisor, Dr. Ming Xue, for his wise supervision and continual support over the years. With great patience and encouragement, not only did he give me the freedom to work on the ideas that interested me — even if they were impractical or naïve sometimes — but also he was always ready to help me out of the difficult problems. He kept me grounded with his excellent academic guidance and profound physical insight. I also wish to thank Drs. Kelvin Droegemeier, Lance Leslie, Robert Palmer, Lakshmivarahan and Xuguang Wang for serving on my doctoral advisory committee, and for their professional suggestions on my work and encouragement to me at many steps through my research.

I also thank my colleagues in the Center for Analysis and Prediction of Storms (CAPS) for their help and support. Especially my great thankfulness goes to the ARPS EnKF research group. Their works were the solid foundation on which my research was laid. Particularly, I am grateful to Drs. Youngsun Jun, Shizhang Wang, Chengsi Liu, Tim Supinie, Nathan Snook and Yunheng Wang for very helpful and productive discussions. Youngsun and Shizhang especially helped me to understand the ARPS-EnSRF system and provided many good suggestions. Thanks to Yunheng for the professional support on using ARPS system, code development and parallelization. Thanks Scott Hill and Chris Cook for providing the healthy computing environment in CAPS. My sincere thanks are also due to the staff of our School of Meteorology: Celia Jones, Marcia Pallutto, Christie Upchurch, Nancy Campbell, Becky Steely, Mark

Laufersweiler, and the staff of CAPS: Debra Farmer and Eileen Hasselwander. Their help provided a cheerful and helpful research environment.

Although I could not list out all my friends in detail due to the space limit, I still want to extend my great thankfulness to my officemates and all the friends from school and church. Your friendship and support will be cherished through my life. I would like to show my special thanks to Dr. Ming Hu for recommending me to my advisor, Dr. Ming Xue, and his helpful suggestions whenever I need them in my work.

I am particularly indebted to Dr. Daphne LaDue for her patient and careful proof-reading of the manuscript of my dissertation. Her comments and corrections helped improve the clarity and quality of the manuscript significantly. I take responsibility for any errors that remain.

Lastly and most importantly, my family has always been the harbor for me to rest and get refilled with love, encouragement and confidence. My parents in China always support me and understand me with unconditional love and lifetime encouragement: My parents-in-law helped me out by taking care of my son for 2 years. To my wife, Ying Su, thank you for your understanding and support to me through my research, and especially thank you for your unselfish sacrifice and great patience to allow me to work almost every evening and for accommodating my seemingly endless student life style. My son, Matthew, you are the most wonderful gift given to me from God. You made my life bright and continue to fill my life with joy.

The research work contained within this dissertation was primarily supported by funding from the following resources: NSF Grants OCI-0905040 and EEC-0313747, NOAA Grant NA080AR4320904. I am also thankful for the financial support from the

Linked Environment for Atmospheric Discovery (LEAD) through NSF grant ATM-0331594, which supported my initial research work for the first two and a half years. Most of the experiments in this work were performed on the Linux cluster of CAPS at the University of Oklahoma (OU).

Table of Contents

List of Tables

List of Figures

Abstract

The local ensemble transform Kalman filter (LETKF) has not been applied the storm-scale radar data assimilation (DA) problem until very recently. In this work, LETKF and its four-dimensional version, 4D-LETKF, are implemented within the Advanced Regional Prediction System (ARPS) ensemble DA framework. This LETKF system is first tested by assimilating simulated radar observations of a supercell storm through observing system simulation experiments (OSSEs).

The performance of the LETKF for radar DA is compared with the ensemble square root filter (EnSRF) in the ARPS DA system in the OSSEs. With the same optimal covariance localization radii, the performances of the EnSRF and LETKF are close to each other when their assimilation cycles reach the stable stage. In the earlier spin-up stage, when the state estimation and ensemble covariance are poor, the results of assimilating radial wind observations are also very close. When the radar reflectivity observations are assimilated, however, EnSRF outperforms LETKF during the spin-up stage. The differences are believed to be related to the nonlinearity of the reflectivity observation operator. The linearization for the nonlinear observation operator is implicitly made at different places in the serial EnSRF and simultaneous LETKF algorithm. This difference in effective linearization leads to systematic differences in the analyses of EnSRF and LETKF. This conclusion was further verified through several carefully designed experiments where the above differences are deliberately removed.

The performance of 4D- LETKF is then assessed and compared to the regular 3D-LETKF when analyzing radar batches spanning different time intervals. With the

4D-LETKF algorithm, high frequency observations distributed over time can be analyzed simultaneously without introducing observation timing error. Compared to 3D-LETKF with short, 1 minute intermittent assimilation cycles, 4D-LETKF can produce analyses of similar quality after a few DA cycles while reducing computational cost significantly by using fewer cycles. When using 3 minute or longer larger data batch, 4D-LETKF performs better than 3D-LETKF, mainly due to the observation timing error in 3D-LETKF. Considering the flow-dependence in error covariance, the effects of time localization and adaptive localization with a hierarchical filter (HF) are also tested within 4D-LETKF, respectively. The results show that both time localization and hybrid adaptive and non-adaptive localization can improve the performance of 4D-LETKF, especially when larger data batch length is large.

As an effort to enhance the balance in the analysis, divergence equation and pressure tendency equation are implemented as equation constraints in the ARPS-LETKF framework and tested through OSSEs. The analysis of pressure is improved, while the overall quality of analyses does not change much.

Chapter 1: Introduction and Overview

1.1 Background and Motivation

Mathematically numerical weather prediction (NWP) is an initial-boundary value problem (Lorenc 1995). The quality of initial conditions, which is an estimation of the atmospheric state, is vitally important to NWP. Data assimilation (DA), which is the process of finding the model representation of atmospheric state that is most consistent with the current and past observations (Lorenc 1995), plays key role in NWP. Variational DA (VarDA) and ensemble Kalman filter (EnKF) are two dominant methods used in modern DA. EnKF is not as well-developed as VarDA, which has been used in most operational NWP centers around the world. However, since it was first introduced into atmosphere and ocean data assimilation by Evensen (1994), the EnKF method is gaining popularity. EnKF and its variants, such as the ensemble square root filter (EnSRF) (Whitaker and Hamill 2002), ensemble adjustment Kalman filter (EAKF) (Anderson 2001), and local ensemble transform Kalman filter (LETKF) (Hunt et al. 2007), have been intensively studied and widely applied to atmospheric DA from large to small scales in the most recent decade.

In a typical convective-scale DA system, usually the cloud microphysical process in the forecast model is strongly nonlinear, even non-differential in some situations; radar observations, as the main existing routine observations which are suitable for resolving convective-scale storms, are related to the model variables through nonlinear observation operators, especially the radar reflectivity. Due to the high cost in the implementation and maintenance of adjoint model, and the difficulty in developing the adjoint code for nonlinear and non-differentiable formulas in the model and/or observation operators, VarDA, especially the four dimensional variational (4DVar) method, is limited to relatively simple applications and model settings when applied to convective-scale DA. EnKF, as an ensemble extension to the classical Kalman filter, applies the Monte Carlo method to estimate the forecast error statistics (Evensen 1994;2003). Both the fully nonlinear forecast model and nonlinear observation operators could be completely included in EnKF methods without the need of the adjoint code. With the advantage of explicit estimation of the error covariance in the EnKF analysis, error covariance in EnKF is not only flow-dependent, but also continuously improved through the EnKF DA cycles. Then a reasonable, balanced, and flow-dependent error covariance structure could be built up for a convective-scale DA system, which could not be gained in the VarDA. For its simplicity and ease of implementation, the EnSRF, as a variant of EnKF, has been applied to the convective-scale data assimilation of radar data in the last decade (Snyder and Zhang 2003; Xue et al. 2006; Jung et al. 2008a; Jung et al. 2008b; Xue et al. 2010; Yussouf and Stensrud 2010; Dowell et al. 2011; Snook et al. 2011; Jung et al. 2012). These studies used both observation system simulation experiments (OSSE) and real cases to show the promise of the EnSRF method in a convective-scale DA.

The high computational cost of EnKF is still a serious concern for real-time applications. The high resolution of the model and high density of radar data, which are typical in a real convective-scale DA practice, require efficient, massive parallelization of the EnKF system. Anderson and Collins (2007) proposed a parallel algorithm for EnSRF; Wang et al. (2013c) developed a different efficiently parallel version of EnSRF for the ARPS model on massive parallel computers. However, due to the sequential nature of EnSRF algorithm, the implementation of parallel EnSRF is still sophisticated in practice.

Compared to EnSRF, The local ensemble transform Kalman filter (LETKF), designed to be a parallelizing filter, is an attractive alternative algorithm to overcome the high computational cost problem. The LETKF algorithm (Hunt et al. 2007; Miyoshi and Yamane 2007) was proposed as an efficient upgrade to local ensemble Kalman filter (LEKF) (Ott et al. 2004). LETKF updates the analysis of each grid point independently by assimilating the observations in the local region centered on each grid point simultaneously. As each grid point could be analyzed in parallel, LETKF has significant computational efficiency in modern super-parallel computers. A number of studies have shown the promise of the LETKF with wide applications including global, regional, and mesoscale NWP models (Szunyogh et al. 2008; Seko et al. 2011; Miyoshi and Kunii 2012).

In this study, LETKF is developed with a real, convective-scale NWP model — the ARPS model — and inter-compared with ARPS-EnSRF (Tong 2006) using simulated radar observations. The ARPS-EnSRF had been assessed in both of OSSE cases (Xue et al. 2006) and real data cases (Snook et al. 2011; Jung et al. 2012), and could analyze the radar observations successfully. It is used as a reference in this work. The first purpose of this work is to assess the capability of LETKF to assimilate radar data into a real, convective-scale NWP model.

The second purpose is to compare the performance of EnSRF and LETKF with radar observations, especially the nonlinear observation (e.g., reflectivity). Holland and Wang (2013), and Greybush et al. (2011) made the comparison of EnSRF and LETKF by considering the effects of B-localization vs. R-localization, and sequential vs. simultaneous observation processing. It should be noticed that their studies compared the two schemes of EnSRF and LETKF with linear observation operators and large-scale models. This work investigated the effect of nonlinear observations on EnSRF with sequential observation processing and LETKF with simultaneous processing in the convective-scale DA.

Another advantage of the LETKF algorithm is that it is easy and straightforward to extend to its four dimensional formula, compared to the EnSRF algorithm. The 4D algorithm could simultaneously analyze the observations frequently collected over time, at the times when they were observed, and with fewer analysis cycles. Wang et al. (2013a) showed the significant advantages of 4DEnSRF with the WRF model in a supercell storm OSSE. The third purpose of this work was to implement 4D-LETKF with the ARPS model. Furthermore, the hierarchical filter (Anderson 2007), was applied to the 4D-LETKF algorithm as an adaptive localization method.

In addition, due to the covariance localization, sampling error, and model error, the degradation and/or violation of physical and dynamical balance constraints occur in the EnKF analysis. A method, which can enforce the broken constraints or exploit the additional constraints, must be applied to the current EnKF algorithm in order to improve the filtering performance. In order to understand the feasibility and capability of constrained-LETKF in convective scale DA problem, the fourth purpose of this research was to set up a constrained ARPS-LETKF system, and test this constrained-LETKF system in the convective-scale DA experiments with simulated radar observations.

1.2 Overview

The goals of this study are to:

1) develop an atmospheric data assimilation system based on the Local Ensemble Transform Kalman filter (LETKF) with a convective-scale NWP model, ARPS, and test whether it could effectively utilize the Doppler radar data,
2) inter-compare the existing ARSP-EnSRF DA system with the newly-developed ARPS-LETKF system with simulated radar data in an OSSE of a convective storm event,
3) extend the LETKF to its 4D version, and compare 4D-LETKF with regular 3D-LETKF by assimilating the simulated asynchronous radar observations,
4) develop a constrained-LETKF system based on this ARPS-LETKF system.

Chapter 2 first describes the ARPS-EnKF DA systems and OSSE experiment design, including a brief introduction of the model and truth simulation, observation simulation, and ARPS-EnSRF system. Then the formulation and implementation of ARPS-LETKF is described. This newly-developed ARPS-LETKF system is then assessed through data assimilation cycle experiments with radar observations. The ARPS-LETKF and ARSP-EnSRF are then inter-compared through the data assimilation experiments with radar radial wind and reflectivity observation. They performed differently in the beginning spin-up cycles when using radar reflectivity observations. The effect of the nonlinear observation operator is investigated next. Finally, five additional experiments are used to verify the nonlinear effect.

Chapter 3 first introduces the background and motivations for using the 4D-LETKF algorithm to more fully utilize the frequency observations distributed over time. Then this 4DLETKF algorithm and its implementation are described, and its difference to the 4DEnSRF algorithm is also discussed. Its performance is then tested and inter-compared to regular 3D-LETKF in OSSEs with simulated radar data.

In Chapter 4, the background and motivations for EnKF with constraints are discussed first, then how the constraints are incorporated as "imperfect measurement" to the ARPS-LETKF framework is described next. The divergence constraint and pressure tendency constraint are then introduced. The impacts of the two constraints are inspected through the data assimilation experiments with radar observations.

Chapter 5 summarizes the body of work, and outlines a future study where the current ARPS-LETKF system could be extended.

Chapter 2: Development of ARPS LETKF and Inter-comparison with ARPS EnSRF: Effects of Nonlinear Observation Operator

In this chapter, the ARPS-EnKF DA systems and OSSE experiment design, including the brief introduction of the model and truth simulation, observation simulation, are first described. Then the formulation and implementation of ARPS-LETKF are then discussed. The ARPS-LETKF system is assessed through data assimilation cycle experiments with simulated radar radial wind and reflectivity observations. After that the ARPS-LETKF and ARSP-EnSRF are inter-compared also through the data assimilation experiment with simulated radar observations. Difference in performances is found during the earlier DA cycles when using reflectivity observations and the effects of nonlinear reflectivity observation operator are investigated. Five additional experiments are designed to test the nonlinear effects.

2.1 ARPS EnKF Data Assimilation Systems and Experiment Design

2.1.1. The prediction model and truth simulation

In this work, the test and evaluation of the newly-developed ARPS-LETKF system as well as the inter-comparison with ARPS-EnSRF, are based on simulated data from a classic supercell storm case that occurred on 20 May 1977 in Del City, Oklahoma (Ray et al. 1981). Such simulation experiments are commonly referred to as observing system simulation experiments (OSSEs). Similar OSSE experiments had been used in previous work with ARPS-EnSRF in Tong and Xue (2005), Xue et al. (2006) and Tong and Xue (2008b). The configurations of the forecast model and truth simulation are mostly inherited from Tong and Xue (2005). The forecast model used is the Advanced Regional Prediction System (ARPS) (Xue et al. 2000; Xue et al. 2001; Xue et al. 2003). In this study, the ARPS model is used in a 3D cloud model mode and the prognostic variables include three velocity components, u, v, w; potential temperature, θ; pressure, p; and six categories of water substances: water vapor specific humidity, q_v; mixing ratios for cloud water, q_c ; rainwater, q_r ; cloud ice, q_i; snow, q_s ; and hail, q_h. The LFO83 ice microphysics and 1.5-order turbulent kinetic energy (TKE)-based subgrid-scale turbulence schemes are the only physics options included. More details on the model can be found in Xue et al. (2000) and Xue et al. (2001).

For all experiments, the physical domain is 64 km × 64 km × 16 km. The model grid comprises 35 × 35 ×35 grid points (including points that facilitate the specification of boundary conditions), with grid intervals of 2 km in the horizontal direction and 0.5 km in the vertical. The truth simulation or nature run is initialized from a modified real sounding as used in Xue et al. (2001). A 4-K ellipsoidal thermal bubble centered at x = 24, y = 16, and z = 1.5 km, with radii of 10 km in the x and y directions and 1.5 km in the z direction is used to initiate the storm. The horizontal position of the warm bubble is a little bit different to the position used in Tong and Xue (2005); in this study the bubble is centered at x = 48 km. The length of simulation is up to 3 h. Despite the differences in resolutions, the evolution of the simulated storms is very similar to those documented in Xue et al. (2001). More information about the truth simulation could be found in Tong and Xue (2005).

2.1.2. Simulation of Radar obervations

One WSR-88D radar is assumed in this OSSE study. The simulation of the WSR-88D radar data are similar to the simulation used in Xue et al. (2006). For the WSR-88D radar, standard precipitation-mode parameters are assumed, including 10-cm wavelength, 1° beamwidth, and a total of 14 elevations with the lowest elevation at 0.5° and the highest at 19.5°. More details could be found in Xue et al. (2006).

Different from the earlier OSSE studies of Tong and Xue (2005), in this work it is assumed that the simulated observations are available in the plan position indicator (PPI) planes or the elevation levels of the radar rather than at the model grid points. We do assume that the radar observations are already interpolated to the Cartesian coordinate in the horizontal directions; in other words, the observations are found in the vertical columns through the model scalar points. The main purpose of interpolation is to make the data distribution more uniform in the horizontal so as to avoid the cost of assimilating an excessive number of data at the close ranges of the radar. The effects of the curvature of the earth and the beam bending resulting from the vertical change of refractivity are taken into account by using the simple 4/3 effective earth radius model discussed in Doviak and Zrnic (1993); it is mainly the earth curvature effect that prevents the radars from seeing the lower atmosphere far away.

Because the observations are not at the model grid points, a forward observation operator is needed to bring the data from the model vertical levels to the radar elevation levels. This is achieved by using a simplified radar emulator that does the power-gain-based sampling only in the vertical direction. More information about the vertical sampling technique could be found in Xue et al. (2006). With the vertical sampling technique, the gridpoint values of the simulated radial velocity and reflectivity could be interpolated onto radar elevation levels.

The gridpoint values of simulated radar observations are first calculated with the observation operators as used in Tong and Xue (2005). The simulated radial velocity, V_r, is calculated from,

$$V_r = u\cos\phi_g \sin\gamma_g + v\cos\phi_g \cos\gamma_g + w\sin\phi_g + random \quad error \ . \tag{2.1}$$

where ϕ_g is the elevation angle and γ_g is the azimuth angle of radar beam that goes through the given grid point; and *u*, *v*, and *w* are the model-simulated velocities interpolated to the scalar points of the staggered model grid. The random error is drawn from a normal distribution with zero mean and a standard deviation of $1\,ms^{-1}$. Because V_r is sampled directly from velocity fields, the effect of hydrometeor sedimentation does not come into play.

The simulated logarithmic reflectivity factor (referred to simply as reflectivity in most parts of this paper; in dBZ) is estimated from the following equations:

$$Z = 10\log_{10}\left(\frac{Z_e}{1mm^6m^{-3}}\right) + simulated \quad error \ . \tag{2.2}$$

The equivalent reflectivity factor, Z_e, is made up of three components,

$$Z_e = Z_{er} + Z_{es} + Z_{eh} \ . \tag{2.3}$$

where Z_{er}, Z_{es}, and Z_{eh} are contributions from rainwater, snow, and hail. Information about the error model used for the simulated error for reflectivity can be found in Jung et al. (2008b).

The rain component of the reflectivity is calculated, based on Smith et al. (1975), from

$$Z_{er} = \frac{10^{18} \times 720 \times (\rho q_r)^{1.75}}{\pi^{1.75} N_r^{0.75} \rho_r^{1.75}} \ . \tag{2.4}$$

If the temperature is less than 0°C, then the component of the reflectivity is, for dry snow,

$$Z_{es} = \frac{10^{18} \times 720 K_i^2 \rho_s^{0.25} \times (\rho q_s)^{1.75}}{\pi^{1.75} K_r^2 N_s^{0.75} \rho_i^2} \ . \tag{2.5}$$

Wet snow, which occurs at temperatures higher than 0°C, is treated in a similar way as rainwater, and the equivalent reflectivity factor-mixing ratio relation is

$$Z_{es} = \frac{10^{18} \times 720 \times (\rho q_s)^{1.75}}{\pi^{1.75} N_s^{0.75} \rho_s^{1.75}} \ . \tag{2.6}$$

For hail, the wet hail formulation of Smith et al. (1975) is used

$$Z_{eh} = \left(\frac{10^{18} \times 720}{\pi^{1.75} N_h^{0.75} \rho_h^{1.75}} \right)^{0.95} (\rho q_h)^{1.6625} \ . \tag{2.7}$$

The above equations define the observation operator, denoted as H, in the following sessions. When creating simulated observations for our OSS experiments, a lower limit of 1 $\mathrm{mm^6 m^{-3}}$ is imposed on the equivalent reflectivity factor to yield a lower limit of 0 dBZ for the logarithmic reflectivity factor.

The ground-based radar is located at the southwest corner of the computational domain, that is, at the origin of the *x*–*y* coordinates (Figure 2.1). Figure 2.1 shows the truth simulation (top panels), simulated radar reflectivity (middle panels) and radial wind (bottom panels). The left panels are at *t*=20min, right panels are at *t*=60min. The radar location is marked with a red dot in the simulated reflectivity plot. Note that the truth wind and reflectivity are plotted at the model levels, but the simulated observations are plotted at the radar elevation angle=10^{o}. In the simulated reflectivity plots (middle panels), the red circle indicates the places where is 4 km high above ground on the radar scan tilt with elevation angle=10°, the blue circle is for places where is 6 km high above ground.

For data sampling and assimilation, we assume that the observation operators, given by the above equations, are perfect. As with most atmospheric data assimilation systems, the prediction model is also assumed to be perfect; that is, no model error is explicitly taken into account.

From the equation (2.1)-(2.7), it is obviously that, without considering the effect of terminal velocity of hydrometers, the observation operator for radial wind is linear, but the observation operators for reflectivity factors are strongly nonlinear. Because of its strong non-linearity, the error distribution of the reflectivity data is likely non-Gaussian. One of the goals of this study is to inter-compare the performances of EnSRF and LETKF with assimilation of the nonlinear radar reflectivity observations.

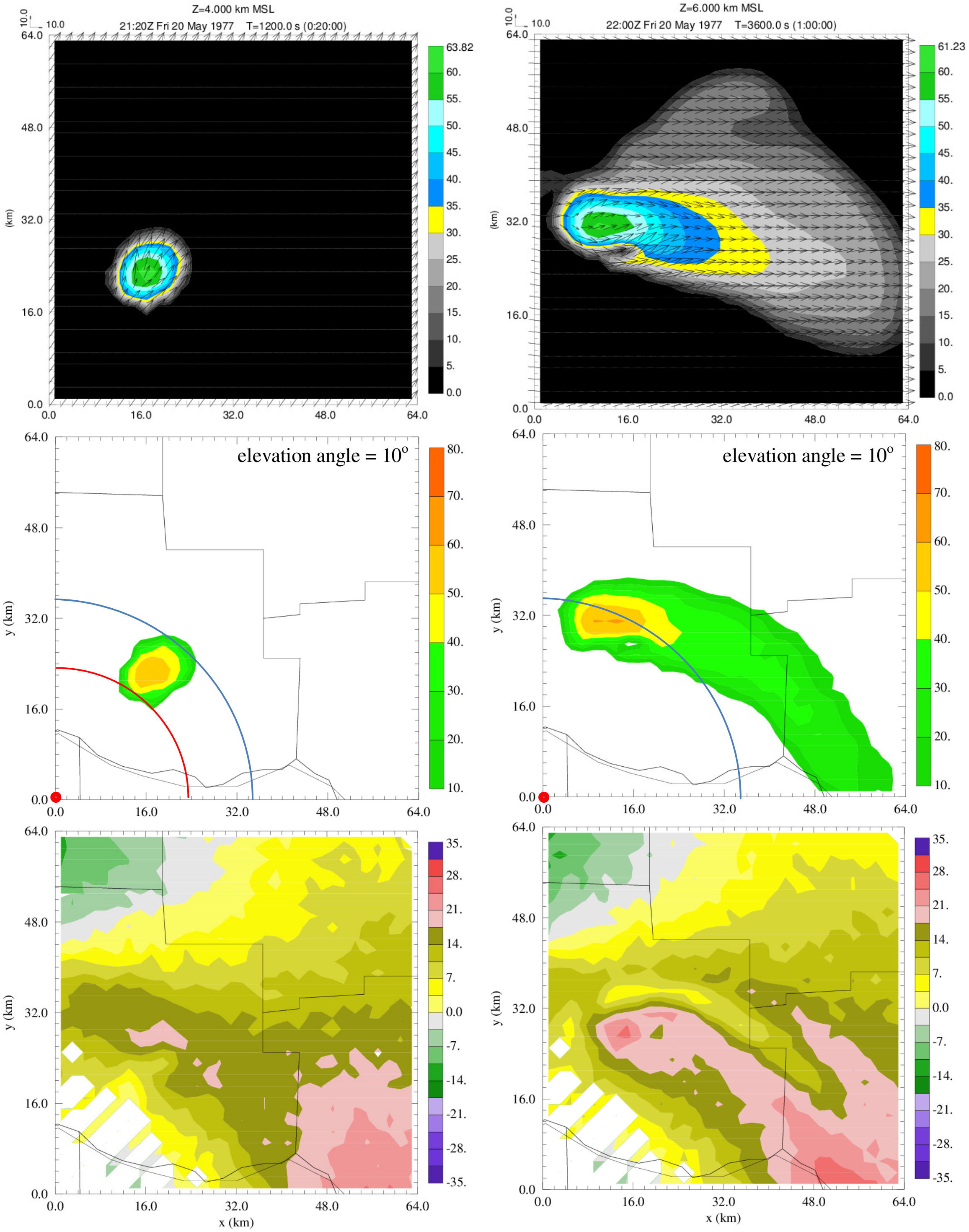

Figure 2. 1 Truth simulations (top panels) and simulated radar observations at elevation angle = 10^o (middle panels for reflectivity; bottom panels for radial wind) at t=20min (left panels) and t=60min (right panels).

2.1.3. ARPS-EnSRF scheme

The current ARPS-EnSRF system (Xue et al. 2006) is primarily based on the EnSRF algorithm of Whitaker and Hamill (2002). The system includes capabilities for parameter estimation (Tong and Xue 2008b;2008a), dual-polarimetric radar data assimilation (Jung 2008; Jung et al. 2008a; Jung et al. 2008b), simultaneous reflectivity attenuation correction (Xue et al. 2009), and the ability to handle a variety of data sources (Dong et al. 2011). Additionally, it has been coupled with a double-moment microphysics scheme (Xue et al. 2010; Jung et al. 2012). Recently, an efficient parallelization of ARPS-EnSRF system had been implemented by Wang et al. (2013c). A full description of the filter can be found in Tong (2006), Xue et al. (2006) and Wang et al. (2013c).

2.1.4. Formulation and Implementation of ARPS-LETKF

The LETKF was proposed by Hunt et al. (2007) as an efficient upgrade of local Ensemble Kalman filter (LEKF) (Ott et al. 2004; Szunyogh et al. 2005). LETKF updates the analysis of each grid point independently by assimilating the observations in the local region centered on each grid point simultaneously(Hunt et al. 2007; Miyoshi and Yamane 2007). Because each grid point can be analyzed in parallel, LETKF has significant computational efficiency in modern super-parallel computers. Compared with its predecessor LEKF, solving the analysis equation in the subspace spanned by ensemble members lowers the computational cost of LETKF, and using a symmetric square root matrix algorithm helps to make the final analysis smoother and increase the accuracy (Hunt et al. 2007; Szunyogh et al. 2008).

Formulation of LETKF

In brief, LETKF first determines the information of the entire first guess and the observations, separates the entire fields into local patches around each grid points. Then LETKF updates the central grid point of each local patch using the localized information in parallel, finally collecting the analysis at every grid point to make up the global analysis ensemble (Liu et al. 2007; Miyoshi and Yamane 2007). The LETKF analysis equations used in ARPS-LETKF are documented here, and are the same as in Hunt et al. (2007):

$$\overline{w}^a = \widetilde{P}^a (Y^b)^T R^{-1} (y^o - \overline{y}^b) \ . \tag{2.8}$$

$$\widetilde{P}^a = \left[(K-1)I + (Y^b)^T R^{-1} Y^b\right]^{-1} \ . \tag{2.9}$$

$\overline{w}^a$ represents the analysis increment in subspace *S* spanned by ensemble member. It is a K_{mem}–dimensional column vector (K_{mem} is the ensemble size.). $\widetilde{P}^a$ is a $K_{mem} \times K_{mem}$ matrix, and it represents the analysis error covariance in space *S*. And y^o is the observation vector, $\overline{y}^b$ is ensemble observation prior mean vector, Y^b is the ensemble observation prior perturbation matrix, and R is the observation error covariance matrix, usually it is assumed to be a diagonal matrix with off-diagonal elements as zero. In model space, the analysis mean and covariance are then as follows:

$$\overline{x}^a = \overline{x}^b + X^b \overline{w}^a \ . \tag{2.10}$$

$$P^a = X^b \widetilde{P}^a \left(X^b\right)^T \ . \tag{2.11}$$

LETKF would find the analysis ensemble with mean and covariance equal to $\overline{x}^a$ and P^a in (2.10) and (2.11). So the LETKF algorithm first solves equations (2.8) and (2.9) in the ensemble space *S*, then converts the solutions back to model space with

(2.10) and (2.11). One of the advantages of finding the solutions in ensemble space first is that it could reduce the dimension of the problem (either of model dimension of observation dimension) to the low dimension of sub-space S spanned by the ensemble. As we know, the sample/ensemble covariance matrices (either P^b or P^a) have serious rank deficiency problems because the model dimension is much larger than the dimension of the ensemble. Solving $\tilde{P}^a$ in the ensemble space, instead of solving P^a in the model space, alleviates the rank deficiency problem. In addition, the inverse of the matrix is much easier in ensemble space than in observation space when the observation dimension is much larger than the ensemble size, which is common in the practice of DA problems with high dense and frequent radar observations. On the other hand, since the solutions are found in ensemble space, the covariance in model or observations space is not represented explicitly, and the regular covariance localization (e.g, B-localization) cannot be applied in LETKF directly. The localization in LETKF will be discussed later.

In LETKF the symmetric square root is adopted to choose the analysis ensemble perturbations as

$$X^a = X^b W^a \ . \tag{2.12}$$

where the Ensemble Transform Matrix is

$$W^a = \left[(K_{mem} - 1)\tilde{P}^a\right]^{1/2} \ . \tag{2.13}$$

where W^a is the symmetric square root of $\tilde{P}^a$, it is also a $K_{mem} \times K_{mem}$ matrix. Equation (2.12) shows that the ensemble analysis perturbations are just the linear combination of the ensemble analysis perturbations; the weight used for the linear combination is defined as W^a in equation (2.13). Its columns represent the analysis ensemble perturbations in ensemble space *S*.

Finally, combined with equation (2.10) and (2.11), the *i*-th analysis ensemble member is as following

$$x^{a(i)} = \bar{x}^b + X^b w^{a(i)} \ . \tag{2.14}$$

where $\{w^{a(i)}\}$ is the columns of the resulting matrix by adding $\overline{w}^a$ to each of the columns of W^a . $\{w^{a(i)}\}$ represents the weight vectors that specify what linear combinations of background ensemble perturbations to add to the background mean to obtain the analysis member in model space.

The LETKF is implemented in ARPS with the following steps, also described in Hunt et al. (2007). These steps are documented here:

Step 1. Apply observation operator $H_{[g]}$ to each background ensemble member $x^{b(i)}_{[g]}$ to form the global background observation ensemble $\{y^{b(i)}_{[g]}\}$. Average $\{y^{b(i)}_{[g]}\}$ to get the ensemble mean $\bar{y}^b_{[g]}$. Subtract $\bar{y}^b_{[g]}$ from $\{y^{b(i)}_{[g]}\}$, get the ensemble perturbations $Y^b_{[g]}$.

$$y^{b(i)}_{[g]} = H_{[g]}(x^{b(i)}_{[g]})$$

$$\bar{y}^b_{[g]} = \bar{y}^{b(i)}_{[g]} = \overline{\left(H_{[g]}(x^{b(i)}_{[g]})\right)}$$

$$y^{b(i)}_{[g]} = y^{b(i)}_{[g]} - \bar{y}^b_{[g]}$$

Step 2. Average $\{x^{b(i)}_{[g]}\}$ to get the ensemble mean $\bar{x}^b_{[g]}$. Subtract $\bar{x}^b_{[g]}$ from $\{x^{b(i)}_{[g]}\}$, get the background ensemble perturbations $X^b_{[g]}$.

$$\bar{x}^b_{[g]} == \overline{\{x^{b(i)}_{[g]}\}}$$

$$x_{[g]}^{b(i)} = x_{[g]}^{b(i)} - \overline{x}_{[g]}^{b}$$

Step 3. Decompose the global fields into local patches around each grid point.

$Y_{[g]}^{b} \rightarrow Y_{[l]}^{b}$, $\overline{y}_{[g]}^{b} \rightarrow \overline{y}_{[l]}^{b}$, $X_{[g]}^{b} \rightarrow X_{[l]}^{b}$ and $\overline{x}_{[g]}^{b} \rightarrow \overline{x}_{[l]}^{b}$

Also the observation error covariance $R_{[g]} \rightarrow R_{[l]}$

Note: Step 4–8 are done in each local patch independently, and could be implemented in parallel. Hereafter the subscript $[l]$, which represents the local patch, is omitted.

Step 4. Compute $k_{mem} \times l_{obs}$ matrix C

$$C = \left(Y^{b}\right)^{T} R^{-1}$$

If R is diagonal (observation errors are uncorrelated), it is easy to get its inverse R^{-1}; If R is block diagonal, then solve the linear system $RC^{T} = Y^{b}$ to get matrix C. The covariance R-localization (which is discussed later) could also be applied in this step.

Step 5. Calculate $k_{mem} \times k_{mem}$ matrix $\widetilde{P}^{a}$

$$\widetilde{P}^{a} = \left[(k_{mem} - 1)I/\rho + CY^{b}\right]^{-1}$$

Method 1. Calculate the inverse of $k_{mem} \times k_{mem}$ symmetric matrix $\left[(k_{mem} - 1)I/\rho + CY^{b}\right]$; (use subroutine ***rs*** in EISPACK Fortran library .)

Method 2. Solve the eigen-space of symmetric matrix $\left[(k_{mem} - 1)I/\rho + CY^{b}\right]$ or $\left[CY^{b}\right]$

$$CY^{b} = V\Lambda V^{T} \text{ and } \left[(K-1)I/\rho + CY^{b}\right] = V\Lambda^{*}V^{T}$$

Then

$$\widetilde{P}^{a} = V\Lambda^{*^{-1}}V^{T}, \text{ where } \Lambda^{*} = (K-1)I/\rho + \Lambda$$

Subroutines ssyev/ssyevx/ssyevr (or corresponding double precision subroutines) in Linear Algebra PACKage (LAPACK) is used to solve the eigenspace problem.

Note: check whether all eigenvalues are positive, or at least semi-positive. Negative or very tiny eigenvalues will be assigned to be 0, and the corresponding efficient rank of matrix is reduced.)

Step 6. Calculate $K_{mem} \times K_{mem}$ matrix W^a :

$$W^a = \left[(K-1)\tilde{P}^a\right]^{1/2} = V\left[(K-1)\Lambda^*\right]^{1/2} V^T$$

Step 7. Calculate K_{mem}-dimensional vector $\overline{w}^a = \tilde{P}^a C(y^o - \overline{y}^b)$

$$\overline{w}^a_{k_{mem}\times 1} = \tilde{P}^a_{k_{mem}\times k_{mem}} C_{k_{mem}\times l_{obs}} (y^o - \overline{y}^b)_{l_{obs}\times 1} = \tilde{P}^a_{K_{mem}\times K_{mem}} C_{K_{mem}\times l_{obs}} \delta y^o_{l_{obs}\times 1}$$

Add $\overline{w}^a$ to each column of W^a, to form a resulting matrix, whose columns are $\{w^{a(i)}\}$

$$\{w^{a(i)}\} = \{w^{a(i)}\} + \overline{w}^a$$

Step 8. Update the analysis ensemble $X^a = \{x^{a(i)}\}$ in the local patch, or just update the center grid points:

$$x^{a(i)} = X^b w^{a(i)} + \overline{x}^b, \ i=1,\ 2,\ 3, \ldots \ Kmem.$$

Step 9. Repeat step 4-8 for each grid points. After that, combine the analysis ensemble of every grid points to get the global analysis ensemble.

Note: notice that steps 4–7 only depend on the information of observations used in the local patch. Different local patches, which use the same observations subset, could share the matrices $\tilde{P}^a$ and $\{w^{a(i)}\}$.

In the ARPS-LETKF system, as in the ARPS-EnSRF, the analysis variables contained in state vector x include the gridpoint values of *u, v, w,* θ*, p,* q_v, q_c, q_r, q_i, q_s and q_h. The observation vector y^o contains the radar radial velocity V_r and reflectivity *Z*. The observation operator *H* is defined by Eqs. (2.1)-(2.7) in the simulation of radar observation session.

Localization in ARPS-LETKF

The forecast error covariance matrix plays a vital role in modern DA systems. The accuracy of the error covariance matrix has direct influence on the quality of the analysis. In the EnKF framework, the forecast error matrix is approximately estimated by using the sample covariance of the ensemble forecast. Due to the limited ensemble size, the sampling error will produce spurious correlations between distant locations in the background covariance matrix (Whitaker and Hamill 2002; Hunt et al. 2007; Greybush et al. 2011). Ensemble covariance that are deteriorated by the spurious correlations lead to an analysis with degraded quality, such as the analysis in a place that was influenced by an observation which is far way in an essentially random manner(Hunt et al. 2007).

Localization is a technique by which the impact of observations from distant regions upon an analysis is suppressed(Greybush et al. 2011).

$$K_{gain} = \frac{P^b H^T}{HP^b H^T + R} = \frac{X^b (HX^b)^T}{(HX^b)(HX^b)^T + R} \tag{2.15}$$

Equation (2.15) is the formula of ensemble Kalman gain. In EnSRF, because the observations are analyzed serially one by one, the denominator is a scalar. The numerator is a vector representing the covariance between the analyzed model state variable and the observation variable. To suppress the spurious correlation existing in the covariance $X^b(HX^b)^T$, the localization is applied to the each element in the covariance vector $X^b(HX^b)^T$ by the Shur product,

$$K_{gain} = \frac{(\rho_B \circ P^b)H^T}{HP^bH^T + R} = \frac{\rho_B \circ (X^b(HX^b)^T)}{(HX^b)(HX^b)^T + R} \tag{2.16}$$

ρ_B is the localization weight function, which is a distance-dependent function, " $\circ$ " represents the Shur product operator. So in EnSRF, the localization by a distance-dependent function is performed upon the covariance between a model grid point and the observation(Whitaker and Hamill 2002). This is called B-localization(Greybush et al. 2011) because it is applied to the P^b matrix (which is also called B matrix). In ARPS-EnSRF, the distance-dependent function used for B-localization is the 5th-order correlation function(Gaspari and Cohn 1999). In this function, the specific distance, beyond which the function value is 0, is called localization ***cut-off radius*** in this work. However, as mentioned before, in LEKTF the analysis update is computed mainly in the ensemble–spanned space, and the background error covariance matrix is not explicitly used, so B-localization is not suitable for LETKF (Greybush et al. 2011).

In ARPS-LETKF, the localization includes two steps(Miyoshi and Yamane 2007). The first step is to select the observations only from a region surrounding the location of the analysis grid. This step is intrinsically contained in LETKF at step three of the LETKF algorithm. The data selection method is used in Houtekamer and Mitchell (1998). LEKF and LETKF are named with "Local" just for this local analysis approach. In addition to providing better results, this localization approach allows the LETKF analysis to be done more efficiently as a parallel computation (Hunt et al. 2007). However, in this way, the weighting function of the localization is the Heaviside function that weights 1 inside and 0 outside the domain (Miyoshi and Yamane 2007); the abrupt localization cutoff can result in a noisy analysis (Greybush et al. 2011). So Hunt et al. (2007) proposed a "smoothed localization", called R-localization, also called "observation localization" in Nerger et al. (2012), by multiplying the observational error variance (which is the element of diagonal matrix *R*) by the inverse of a localization weighting function.

$$K_{gain} = \frac{P^b H^T}{HP^b H^T + \rho_R^{-1} \circ R} \tag{2.17}$$

Equation (2.17) shows how the localization weight is applied to the Kalman gain used in LETKF. This weight, also called localization function, decays smoothly from one to zero as the distance of the observations from the analysis grid point increases. By weighting the observation error variance in the range from its original value to infinitely large, the influence of observations is localized. Since the weight is applied to the observation error variance matrix, it is called "R-localization". This R-localization is used as the second step for localization in ARPS-LETKF. Figure 2.2 depicts the localization scheme used in LETKF. The 5-point star represents the location of the current analysis variable, and the circle surrounding the analysis grid (dashed line) is the domain for observation selection. The radius of the circle also works as the localization cut-off radius, meaning that only the green dots are selected for local analysis; the hollow dots beyond the circle are excluded. Finally, R-localization weights for the green dots are based on the distances from observation to the analysis grid.

In addition, in order to improve the performance of R-localization, Nerger et al. (2012) introduced a ***regulated R-localization*** scheme. Their experiments demonstrate that the regulated localization can lead to a significant reduction of estimation errors as well as increased stability of the assimilation process. Although this regulated R-localization is proved to have identical effect to B-localization when using the same localization cut-off radius only in the single observation case, it is also implemented in our ARPS-LETKF and tested when inter-comparing with ARPS-EnSRF in the data assimilation experiment with radar observations. The effect of R-localization, regulated R-localization, and B-localization on the performance will be compared in this work. The (2.18) is the regulated R-localization weight $\tilde{\rho}_R$ in Nerger et al. (2012).

$$\tilde{\rho}_R = \frac{1}{1+(1-\rho_B)\times\frac{HP^bH^T}{R}}\times\rho_B \tag{2.18}$$

The same 5th order correlation function of Gaspari and Cohn (1999) is used as the localization function for R-localization function ρ_R, regulated R-localization weight function $\tilde{\rho}_R$ and B-localization function ρ_B. The optimal cut-off radii for different localization schemes are going to be found through comparing the performances of the DA experiments with different cut-off radii. Notice that in all of the DA experiments in this work, the horizontal radius and vertical radius are the same, meaning our localization function is isotropic, although it could be anisotropic.

Adaptive localization approaches, such as the hierarchical filter (HF) of Anderson (2007), will be applied to 4D-LETKF in Chapter 3.

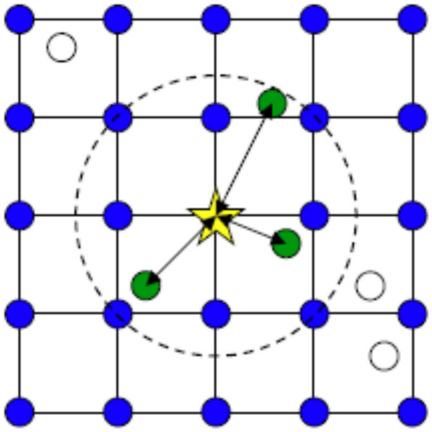

Figure 2. 2 Localization scheme: The star is the localization of current analysis variable, the circle is the domain for observation selection, and the radius of the circle is the cut-off radius for R-localization. Only green dots inside the circle would be used in the local analysis for the variable at the star point. The R-localization weight for each green dot is based on the distance to the star point. This figure was originally from Miyoshi (2010).

Issue related to the staggered model grid

In the ARPS system, model variables are staggered on an Arakawa C-grid (see Figure 2.3), with scalars (including perturbation potential temperature θ', perturbation pressure p', cloud water q_c, rainwater q_r, water vapor q_v , cloud ice q_i, snow q_s, and hail q_h) defined at the center of the grid boxes and the normal velocity components (u, v, w) defined on the corresponding box faces. The coordinate variables x, y and z are also staggered and are defined at the u, v and w points, respectively (Xue et al. 1995). So for u, v, w and the scalars with the same grid indices, the distances from the same observation to them are different. As we know, the data selection and localization weight are both distance-dependent, so the local observation search and localization weights have to be repeated for u, v, w and the scalars, and correspondingly the weight matrices also have to be re-calculated for them respectively. ARPS-EnSRF also makes the analysis for u, v, w and the scalar variables on their own grid points respectively, but the computational costs are different. For EnSRF, it is to search the nearby analysis variables located on regularly-gridded points which are easily organized and found with their indices; for LETKF, it is to search the nearby observations located on irregularly-distributed points. Furthermore, in EnSRF, the calculation of the update is simple and fast because it is only a scalar operation, not a matrix or vector manipulation; in LETKF, the complicated matrix manipulations (such as SVD, matrix multiplication) are relatively expensive computations. Table 2.1 shows the computational cost for EnSRF and LETKF with (u, v, w) exactly analyzed on their stagger. It shows that the computational cost of LETKF is usually larger than the cost of EnSRF when they both are the serial version. To save the extra computational time of an LETKF analysis for $u/v/w$, the ensemble and mean of the background for $u/v/w$ are interpolated from their own grid points to the scalar grid points before the analysis, therefore $u/v/w$ can be analyzed with all the other scalar variables together. After the analysis, the updated $u/v/w$ have to be interpolated back from scalar grid points back to their own grid points. Three different interpolation schemes are implemented in the ARPS-LETKF system. The first interpolates the full analysis of $u/v/$w back, so it is called Full-Interpolation; the second interpolates the analysis incrementally back and then adds the interpolated analysis increment to the background to get the full analysis at their own grid points, so it is called Incremental Interpolation; the third interpolates the analysis weight matrix $\boldsymbol{w}^{\boldsymbol{a}}$ (in equation (2.13) and (2.14)) to the $u/v/w$ grid points, then uses the interpolated weight matrix to update the $u/v/w$ at their own grid points, so it is called Weight-matrix interpolation. The third scheme is similar to the LETKF with the weight interpolation used in Yang et al. (2009). It is essentially also an incremental interpolation scheme.

Table 2.1 also shows the computational cost of LETKF without any interpolation scheme, LETKF with incremental interpolation and with weight matrix interpolation. It shows that the interpolation scheme could save the time significantly to one third of the time of LETKF without interpolation. To compare the EnSRF and LETKF accurately, the LETKF analysis for staggered variables (*u/v/w*) is still made based on their own grid points without interpolation scheme, same as in EnSRF.

Table 2. 1 Compuational costs of EnSRF and LETKF

Experiment	Total CPU Time (sec)	Total CPU Time w/o I/O	Note
EnSRF	44.3	34.5	*u/v/w* is exactly analyzed on the staggered C-grid
LETKF without Interpolation	105.9	95.5	*u/v/w* is ***exactly*** analyzed on the staggered C-grid
LETKF with incremental interpolation	41.4	31.0	*u/v/w* analyzed at scalar grid
LETKF with W-interpolation	41.8	32.1	W update matrix interpolation

It should be pointed out that, any interpolation scheme unavoidably introduces the interpolation error into the data. Through the single observation test, Figure 2.6 shows that the interpolation with full variables could introduce a noticeable interpolation error in the area where the observation has no influence, while the incremental interpolation and weight interpolation have the interpolation error limited to the vicinity of the observation. Figure 2.28 shows that the analysis performance of LETKF with both incremental interpolation and weight interpolation is very close to the LETKF without interpolation throughout all of the DA cycles. The RMSEs of the analysis from the LETKF with full variable interpolation is higher than the other three schemes during the later DA cycles. Therefore, the incremental interpolation and weight interpolation could save computational time without degradation of the analysis performance.

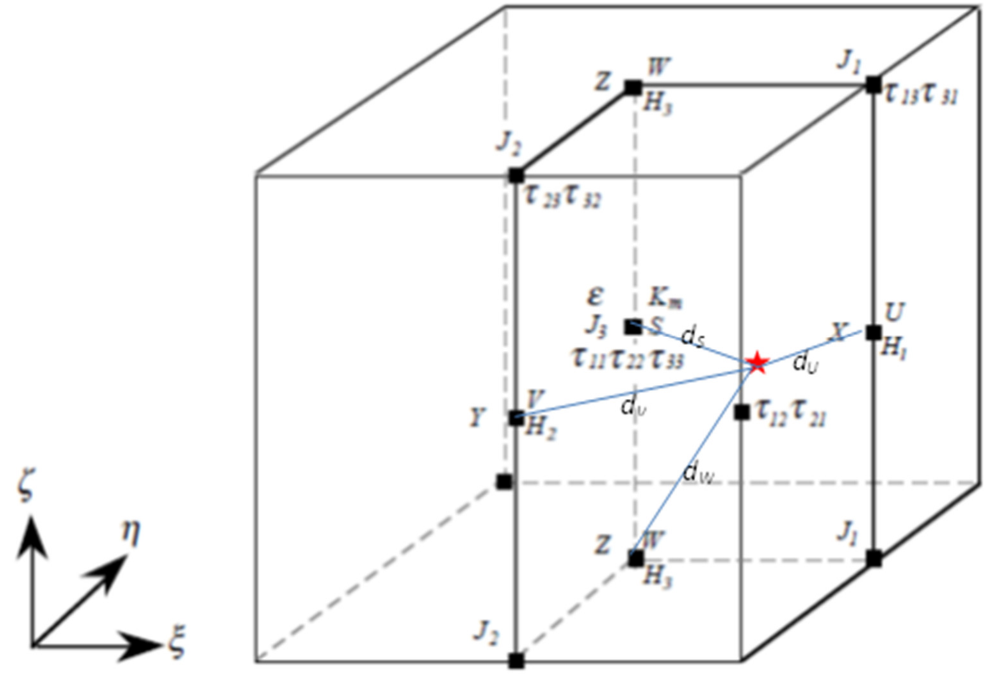

Figure 2. 3 Arakawa C Grid in ARPS-LETKF. The five-point asterisk is an observation, and *dS*, *dU*, *dV* and *dW* represent the distances from the observation to scalar grid point, U grid point, V grid point, and W grid point. This figure is originally from figure 6.1 in ARPS User Guide by Xue et al. (1995)

OpenMP Parallelization of the current LETKF code

LETKF updates the analysis of each grid point independently by assimilating the observations in the local region centered on each grid point simultaneously. It is designed in nature as a parallelizing filter. This makes it easy and straightforward to parallelize the LETKF code, compared to the serial EnSRF algorithm. In our current LETKF analysis code, the analysis loop over the analysis grid points are optimized and parallelized with OpenMP directives. The procedures inside the parallel analysis loop include the steps (4)-(8) described in the earlier section. In the following, letkf_main is the time cost involving the parallelized analysis loop over the analysis grid points. Therefore it directly represents the parallel performance. It shows the good scalability of LETKF algorithm. The time gap between Total Time without I/O and letkf_main is mainly related to the calculation of ensemble observation priors, which has not been parallelized yet. In Figure 2.4, all the LETKF-OpenMP run with the incremental interpolation scheme. Notice that the total CPU time 105.9 sec for serial LETKF analysis without interpolation, whereas the OpenMP parallel LETKF analysis with incremental interpolation and 16 OpenMP threads uses less than 20 sec.

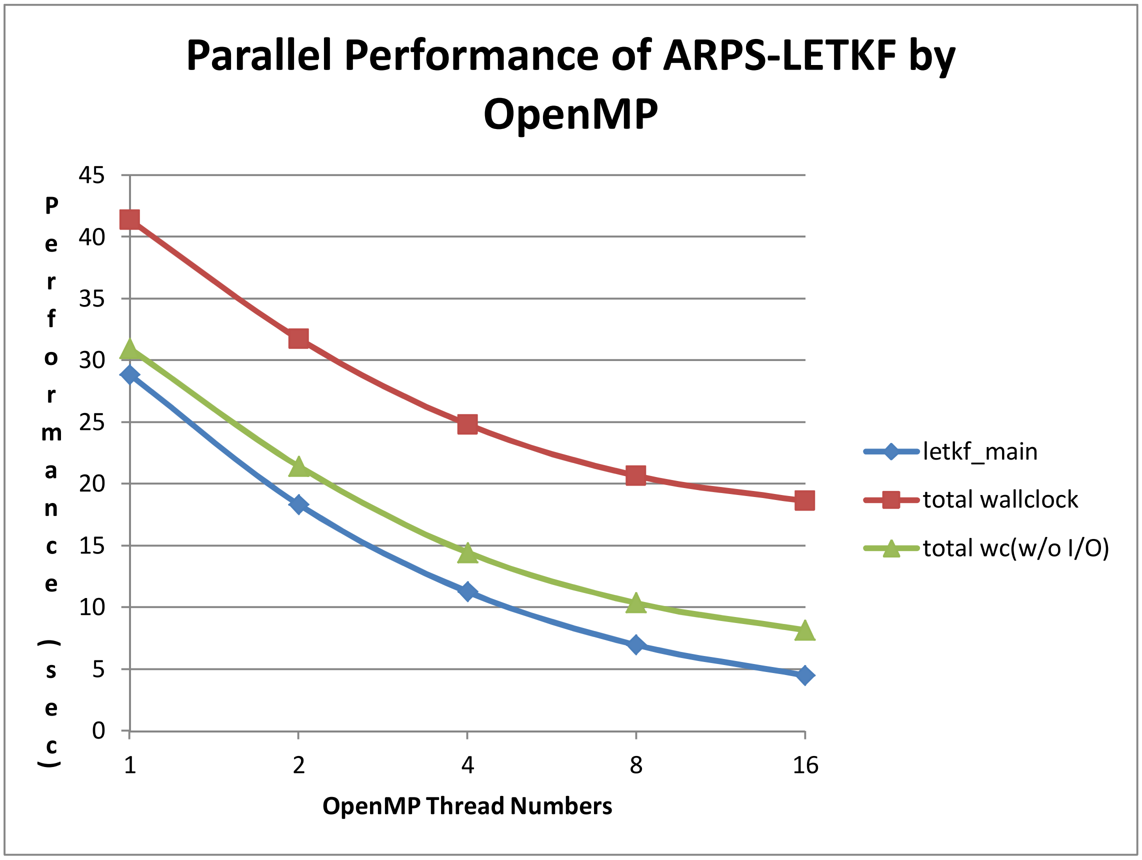

Figure 2. 4 Parallel Performance of LETKF with OpenMP

2.1.5. Ensemble Data assimilation procedure

Following Xue et al. (2006),Tong and Xue (2008b), Tong and Xue (2008a) and Jung et al. (2008a), initial ensemble members are initialized at t = 20 min of model time by adding spatially smoothed perturbations to the initially horizontally homogeneous first guess defined by the Del City, Oklahoma, sounding. The details about the adding spatially smoothed perturbations could be found in Tong and Xue (2008b). The standard deviations of the perturbations added to each variable are 2 ms^{-1} for u, v, and w; 2 K for θ ; and 0.6 g kg^{-1} for the mixing ratios of hydrometeors (q_v, q_c, q_r, q_i, q_s, and q_h). The perturbations are added to the velocity components, potential temperature, and specific humidity in the entire domain, excluding grids composing the lateral boundaries. For the mixing ratios, the perturbations are added only to the grid points located within 6 km horizontally and 2 km vertically from the observed precipitation. Negative values of mixing ratios after the perturbations are added are reset to zero. The pressure variable is not perturbed. These configurations are the same as in Tong and Xue (2008b).

The first assimilation of simulated observations is performed at 25 min of model time and the analyses are repeated every 5 min until 100 min. The EnSRF and LETKF both use 40 ensemble members, a covariance inflation factor of 12%, different to 15% used in Jung et al. (2008b), and a covariance localization radius of 6 km, and the inflation is applied to only to the area influenced by radar observations where the reflectivity > 5dBZ. For more detailed information on the configuration of the assimilation experiment, see Xue et al. (2006),Tong and Xue (2008b), Tong and Xue (2008a) and Jung et al. (2008a).

2.2 Assimilation of simulated radar observations with ARPS-LETKF system and Inter-comparison with ARPS-EnSRF

In this section, the newly-developed ARPS-LETKF system is tested and evaluated in the data assimilation cycle experiments with the simulated radar observations. The performance of ARPS-LETKF is inter-compared with the performance of the existing ARPS-EnSRF. In the scenario of OSSE, the Root Mean Square (RMS) Error of ensemble-mean forecast and analysis is used to show the performance of a DA system. And the Root Mean Squared (RMS) Difference is also used to show how well the ensemble-mean forecast and analysis fit the observations. In this work, RMS Error (RMSE) and RMS Difference (RMSD) are averaged over points at which the true reflectivity is greater than 10 dBZ. Therefore, the verification area is basically within the simulated storm.

2.2.1. Evaluation of ARPS-LETKF with assimilation of Radar Observations

Single Observation Test

First we use the experiments with a single radar observation to test the effect of regulated R-Localization. The single observation is a radar reflectivity observation, which is located at x=39. 000 m, y=23000.0 m and z=3286.461 m (on the 7^{th} scan tilt). The observation value is 67.22541 dBZ with observation innovation of 24.84894 dBZ. Nerger et al. (2012) showed that, in the case of single observation, the B-localization and regulated R-localization should have identical effect. In this single observation analysis, the analysis of EnSRF with B-localization, LETKF with R-localization and regulated R-localization are compared. They use same localization radius=6 km. Figure 2.5 shows that the influences of B-localization and R-localization on the analysis are obviously different, while the results of B-localization and regulated R-localization are almost same. Their differences are in the magnitude of 10^{-7}. So considering the computational accuracy, these two localization schemes are numerically same when analyzing single observation.

Next, the same single observation analysis experiment is used to compare the three interpolation schemes. Figure 2.6 shows that the differences between analysis of LETKF with incremental interpolation and LETKF without interpolation are small and just limited near the observation. The result of LETKF with weight matrix is similar; it shows even smaller interpolation errors. However, for the LETKF with full variable interpolation errors are found throughout the domain, not only in the vicinity of the single observation. Therefore, in the data sparse area, or when observations have small influence (e.g., when analysis is very close to observations and/or truth in the later DA cycles), the interpolation error caused by the full variable interpolation may turn out to be fairly important and degrade the analysis performance. This is the reason why it is observed that the analysis performance of LETKF with full variable interpolation is not as good as the performances from LETKF without interpolation, LETKF with incremental interpolation, and LETKF with weight interpolation.

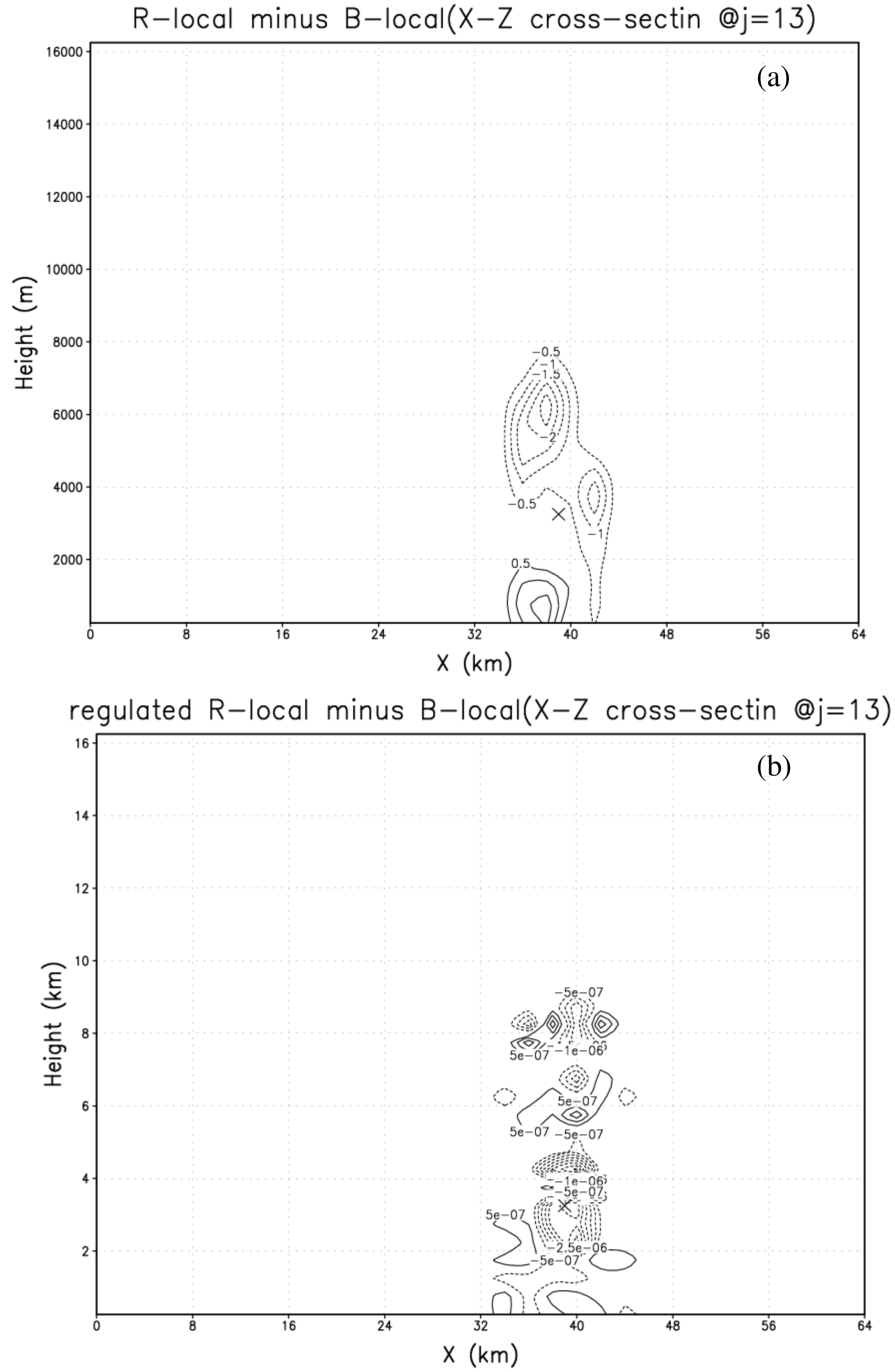

Figure 2. 5 The X-Z cross-section of the difference between analysis of EnSRF with B-localization and the analysis of (a) LETKF with R-localization, (b) LETKF with regulated R-localization in the case of single observation. The observation position is marked with “X” in the figure.

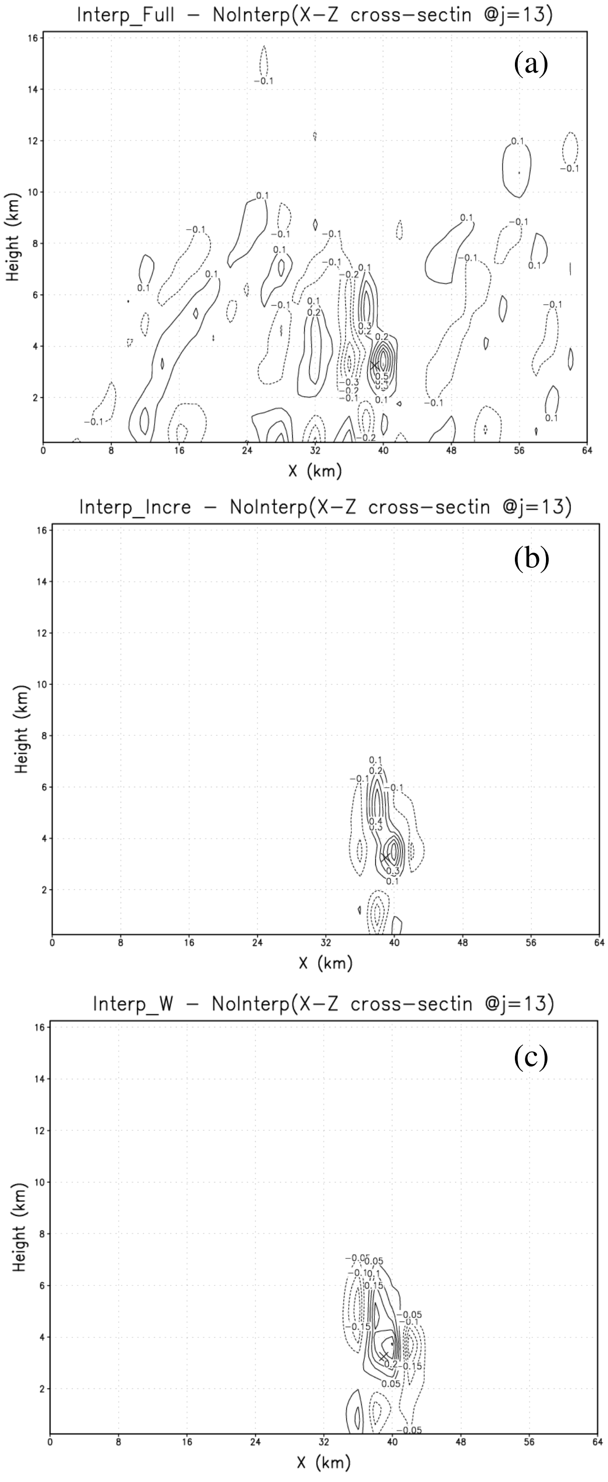

Figure 2. 6 The X-Z cross-section of the difference between analysis of LETKF without interpolation and the analysis of (a) LETKF with full interpolation, (b) LETKF with incremental interpolation, and (c) LETKF with weight interpolation, in the case of a single observation. The observation position is marked with "X" in the figure.

Greybush et al. (2011) and Holland and Wang (2013) indicated that in their work, the R-localization and B-localization used a different optimal scale length to reach the comparable performances. To compare the performances of ARPS-LETKF and ARPS-EnSRF in our convective-scale DA cycle experiment with radar observations, the optimal localization cut-off radius should be found for LETKF and EnSRF first. Through these data assimilation experiments with radar observations, the performances of ARPS-LETKF are evaluated.

Experiment I: Optimal cut-off radius for R-localization in ARPS-LETKF

In Experiment I, the ARPS-LETKF runs with different cut-off radii for R-localization (including 6000m, 5000m, 4000m, 3500m and 2000m) in the DA cycles with radar observations. Figure 2.7 shows the ensemble-mean forecast and analysis RMS Errors (RMSE) of all analysis variables for experiments of LETKF with different cut-off radii of R-localization.

In Figure 2.7, at each analysis step, the RMSEs for all analysis variables decrease quickly by assimilating the radar observations. Although at the model forecast step the RMSEs may increase, the RMSEs still get reduced down to a very low and relatively steady level after a few assimilation cycles. Especially in the first several analysis cycles, the RMSEs are reduced significantly. The period of beginning several DA cycles is called the “spin-up” stage of a DA system. During the spin-up period, there are the significant adjustments of model state toward the observations and true state. In this case, the spin-up stage begins at t=25 min and usually lasts for 4~6 analysis cycles until t=50~60 mins. The period of spin-up is about 35 minutes long. Then after the spin-up, the RMSEs lower to relatively low and steady levels. The significant reduction of RMSEs indicates that this newly-developed ARPS-LETKF system could effectively assimilate the radar observations and produce a good analysis which fits the true state very well after a few DA cycles in this OSSE case under the perfect model assumption. These results are consistent with the performances from ARPS-EnSRF (Tong 2006).

In Figure 2.7, the LETKF running with a cut-off radius equal to 3500 m made the lowest RMSE in both the spin-up stage and the final steady stage for almost every variable. The performances with a cut-off radius of 4000 m are very close to the performances of 3500 m. The performances from longer radii (like 6000 m, 5000 m) or shorter radii (like 2000 m) are not as good as the performances from 3500 m. So in this OSSE case 3500 m is the optimal cut-off radius for R-localization in ARPS-LETKF.

Experiment II: Optimal cut-off radius for regulated R-localization in ARPS-LETKF

Experiment II is similar to Experiment I but for ARPS-LETKF running with different cut-off radii (including 600 0m, 4000 m, and 8000 m) for the regulated R-localization. In the DA cycles the same radar data are used as in Experiment I, including the radar radial wind and reflectivity data. By comparing the RMSEs in Figure 2.8, the performance with cut-off radius equal to 6000 m appears to be the best. Thus the optimal cut-off radius for a regulated R-localization in the ARPS-LETKF is 6000m for this OSSE case.

Experiment III: Optimal cut-off radius for B-localization in ARPS-EnSRF

Experiment III is also similar to Experiment I but for ARPS-EnSRF running with different cut-off radii (including 6000 m, 4000 m, and 8000 m) for the B-localization. In the DA cycles the same radar data are used, including radar radial wind and reflectivity data. Comparing the RMSEs in Figure 2.9, the performance with cut-off radius equal to 6000 m appears to be the best. Thus the optimal cut-off radius for B-localization in the ARPS-EnSRF is 6000 m for this OSSE case.

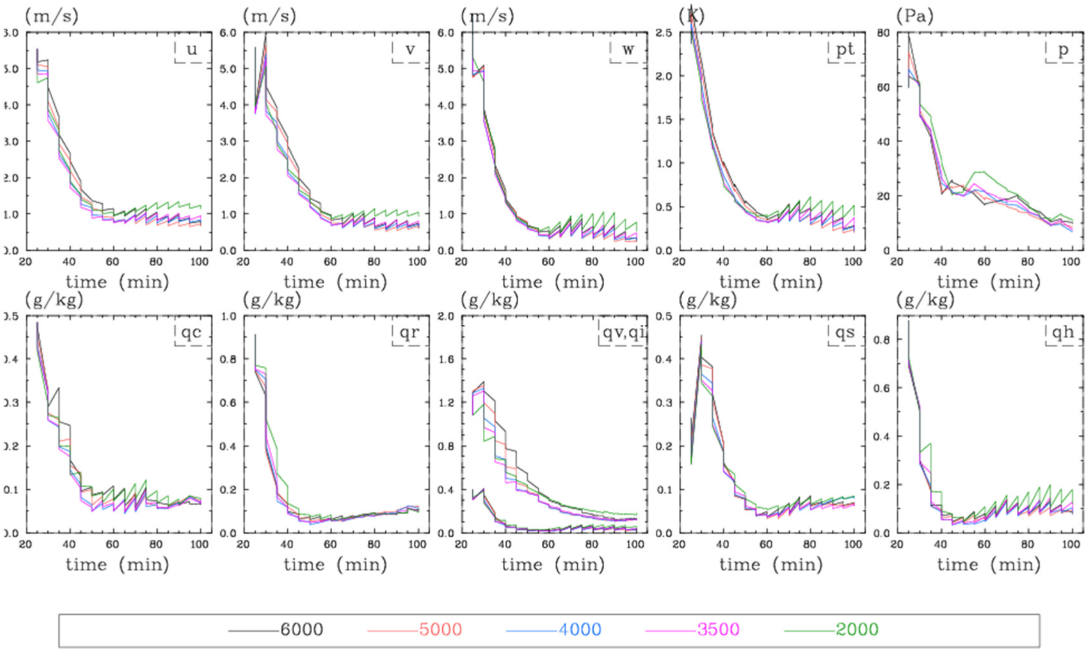

Figure 2. **7** The ensemble-mean forecast and analysis Root Mean Squared Errors (RMSE) averaged over points at which the true reflectivity is greater than 10 dBZ for: u, v, w and perturbation potential temperature θ', perturbation pressure p', cloud water q_c, rainwater q_r, water vapor q_v (the curves with larger values), cloud ice q_i (the curves with lower values), snow q_s, and hail q_h, for experiments of R-localization in ARPS-LETKF with different cut-off radius: 6000 m (black), 5000 m (red), 4000 m(blue), 3500 m (pink) and 2000 m (green).

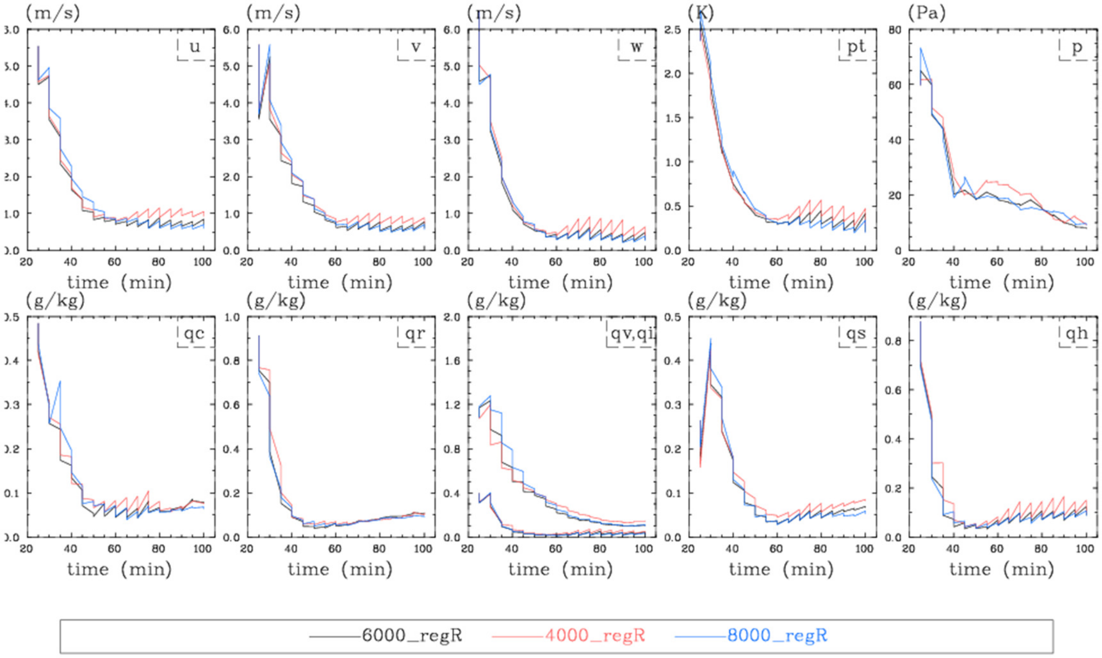

Figure 2. 8 RMSEs, same as figure 2.7, but for experiments of the regulated R-localization in ARPS-LETKF with different cut-off radius: 6000 m (black), 4000 m (red) and 8000 m (blue).

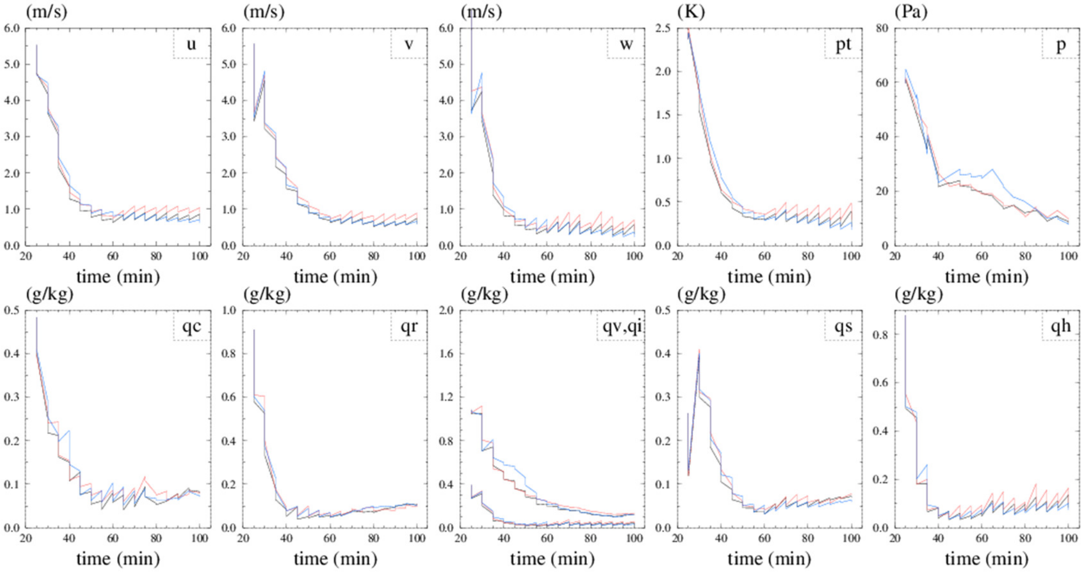

Figure 2. 9 RMSEs, same as figure 2.7, but for experiments of the B-localization in ARPS-EnSRF with different cut-off radii: 6000m (black), 4000m (red) and 8000m (blue).

2.2.2. Inter-comparison of EnSRF and LETKF with Simulated radar observations

In Experiment I and II, the performances of ARPS-LETKF were evaluated. In an OSSE case and with perfect model assumption, ARPS-LETKF could assimilate the radar observation effectively and make good analysis after a few DA cycles, which fit the truth simulations. Based upon the experiment I, I, and III, the optimal cut-off radius for the B-localization in ARPS-EnSRF is 6000 m, the optimal radius for R-localization in ARPS-LETKF is 3500m, and ARPS-LETKF with regulated R-localization is 6000 m. Now, with their optimal localization cut-off radii, the performance of the ARPS-EnSRF and ARPS-LETKF are inter-compared in the experiments of data assimilation with radar observations. The impacts of nonlinear radar observations (e.g., reflectivity) on EnSRF and LETKF are the main focus of this inter-comparison.

Experiment IV: Assimilation of both of radar radial wind V_R and reflectivity Z

In Experiment IV, radar radial wind V_R and reflectivity Z are both assimilated by the EnSRF and LETKF systems. The RMSEs of EnSRF with B-localization, LETKF with R-localization, and LETKF with regulated R-localization are plotted in Figure 2.10.

Figure 2.10 shows that after the spin-up stage, the RMS errors of EnSRF, LETKF with R-localization, and LETKF with regulated R-localization all get significantly reduced and converge to similar levels. This indicates that with their optimal cut-off radii, EnSRF with B-localization, LETKF with R-localization, and LETKF with regulated R-localization may perform similarly. The optimal cut-off radius for R-localization (3.5 km in this case) is shorter than the radius for B-localization (6 km in this case). This result is consistent with the result in Holland and Wang (2012) and Greybush et al. 2011. Shorter optimal radii for R-localization in LETKF leads to fewer data selections and less data influence than for B-localization in EnSRF. This would make the inter-comparison of EnSRF and LETKF unequal. However, with very similar performances, the regulated R-localization has the same optimal cut-off radius (6km in this case) as B-localization has. Although the regulated R-localization is exactly equivalent to B-localization only for single observation case (Nerger et al. 2012; Holland and Wang 2013), these results indicate that the effect of the regulated R-localization could be coarsely equivalent to B-localization in the case of multiple observations.

Furthermore, Figure 2.10 shows that the performance of LETKF with regulated R-localization is a little bit better than with R-localization in both of the spin-up stage and the final steady stage. By using same localization cut-off radius in regulated R-localization and B-localization, the differences in data selection and data influence are reduced to minimum; meanwhile, the performances of the two localization schemes remain very close. So the inter-comparison of EnSRF with B-localization and LETKF with regulated R-localization is more equitable. Hereafter, only the LETKF with regulated R-localization will be inter-compared with EnSRF with B-localization. A cut-off radius of 6 km is used for both schemes.

In Figure 2.10, the significant differences between RMSEs of EnSRF and LETKF occur in the spin-up stage. In the beginning cycles, the RMS errors of LETKF with either R-localization or regulated R-localization are obviously higher than RMS errors of EnSRF. This indicates that in the spin-up stage, the EnSRF outperforms

LETKF, so that EnSRF reaches the final steady stage earlier than LETKF by 1~2 DA cycles. The differences between their performances are significant and systematic. After the spin-up, the differences between the performances from EnSRF and LETKF become trivial and non-systematic.

This significant and biased difference between EnSRF and LETKF in beginning DA cycles also could be observed in the ensemble-mean analyses fields of EnSRF and LETKF. Figure 2.11, Figure 2.12, and Figure 2.13 show the true state and ensemble mean analysis from EnSRF and LETKF at t=25min (first analysis), 35mins (in the middle of spin-up) and 60mins (final stable stage). The first DA cycle is at t=25min, so the analysis of EnSRF and LETKF are based on same observations and same ensemble background. However, Figure 2.11 shows that the strength of update and reflectivity in EnSRF analysis are obviously stronger than in LETKF analysis, and the EnSRF analysis also fits the true state better than LETKF. In Figure 2.12, t= 35min after three DA cycles, but the model is still in the spin-up period, and the differences between EnSRF and LETKF are not as big as in the first analysis step. EnSRF analysis is still a little bit better than LEKTF analysis. For example, in the panel of velocity below 4km, the updraft in LETKF is still weaker than in EnSRF, which leads to a little bit weaker reflectivity around that level in LETKF analysis. At t=60min, in Figure 2.13, there is almost no difference between EnSRF and LETK, and they both are very close to the true state. So the results from Figure 2.11, Figure 2.12, and Figure 2.13 are consistent with the results from Figure 2.10. The EnSRF outperforms LETKF significantly in the spin-up stage.

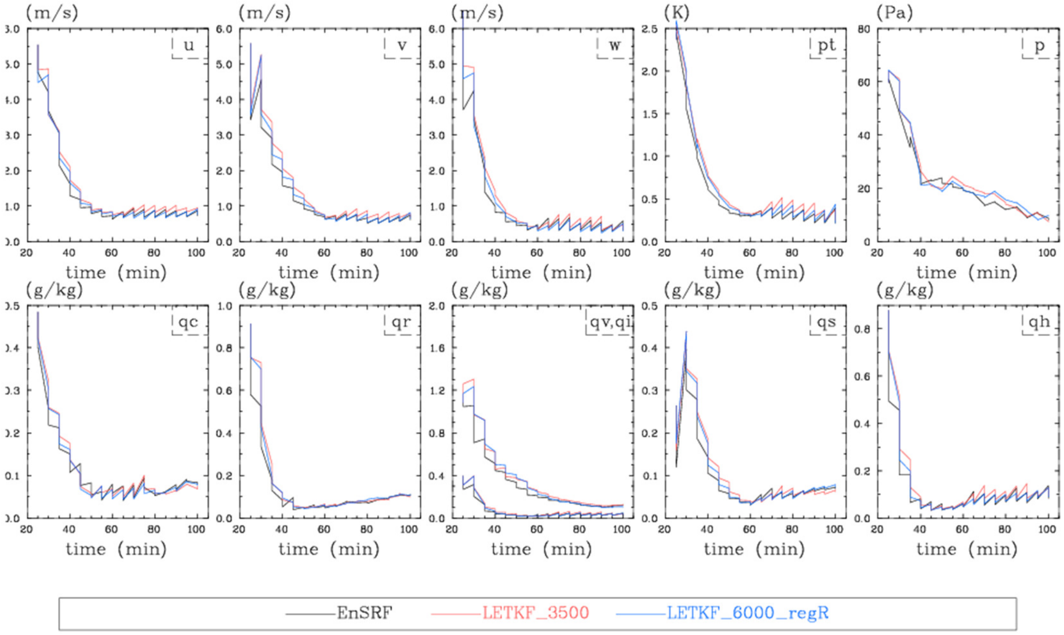

Figure 2. 10 Ensemble-mean forecast and analysis RMS errors for EnSRF with B-localization cut-off radius=6 km (black), LETKF with R-localization cutoff radius = 3.5 km (red) and with regulated R-localization cut-off radius=6 km(blue).

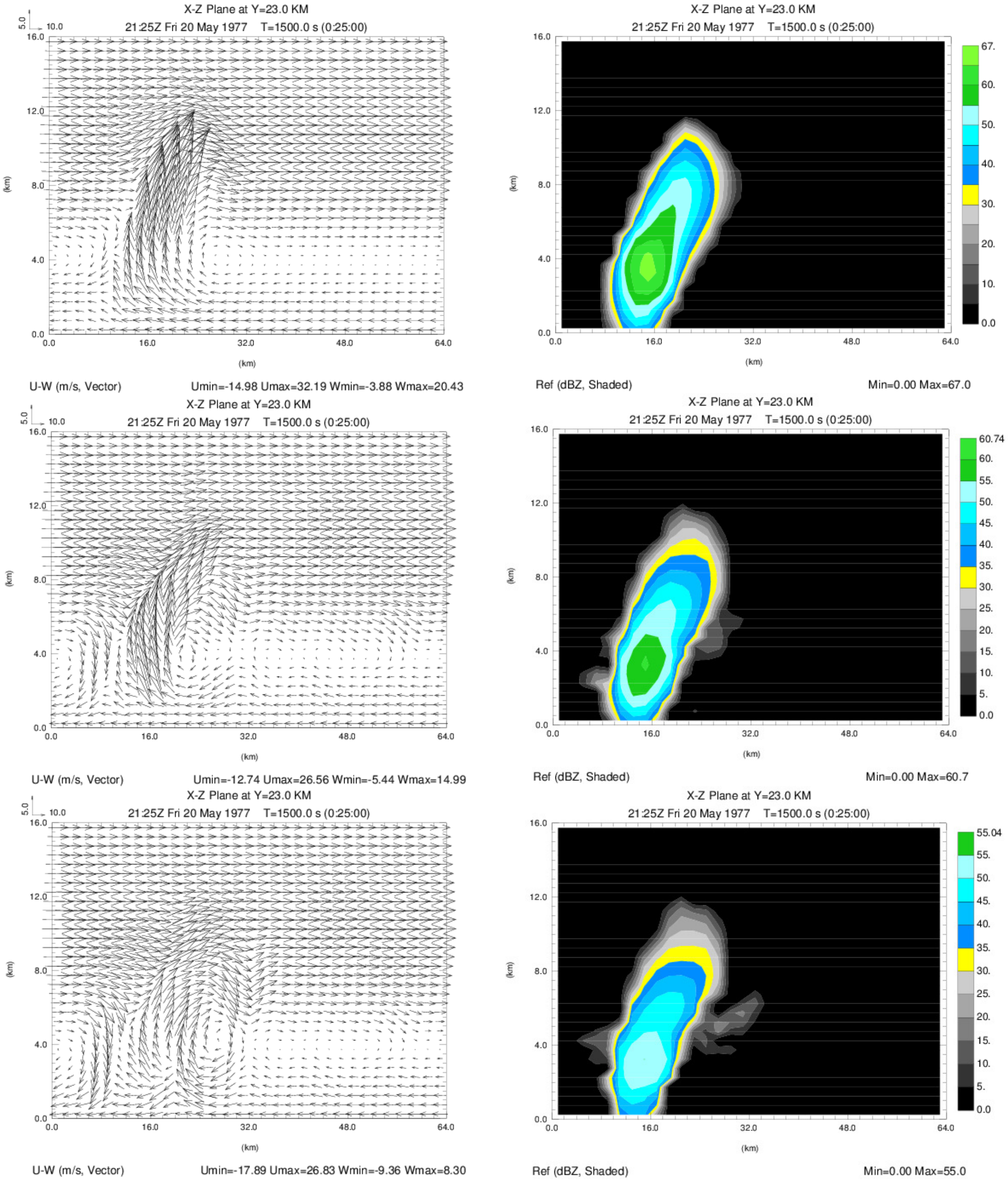

Figure 2. 11 Velocity (left panels) and reflectivity (right panels) on the X-Z slice at y=23km and t=25min (first analysis). Top panels (True), middle panels (EnSRF analysis mean), bottom panels (LETKF analysis mean).

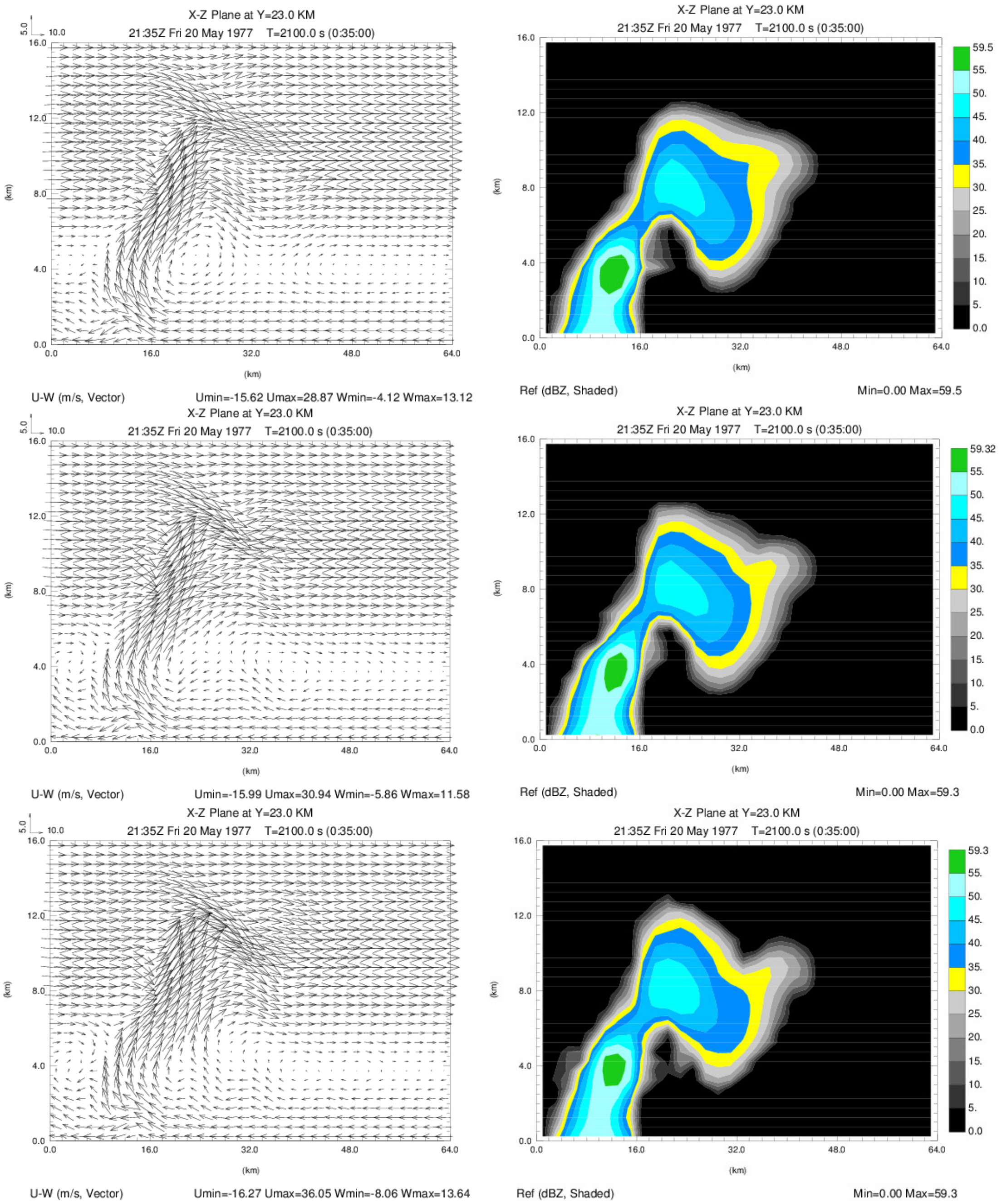

Figure 2. 12 Same as Figure 2.11, but for t= 35min.

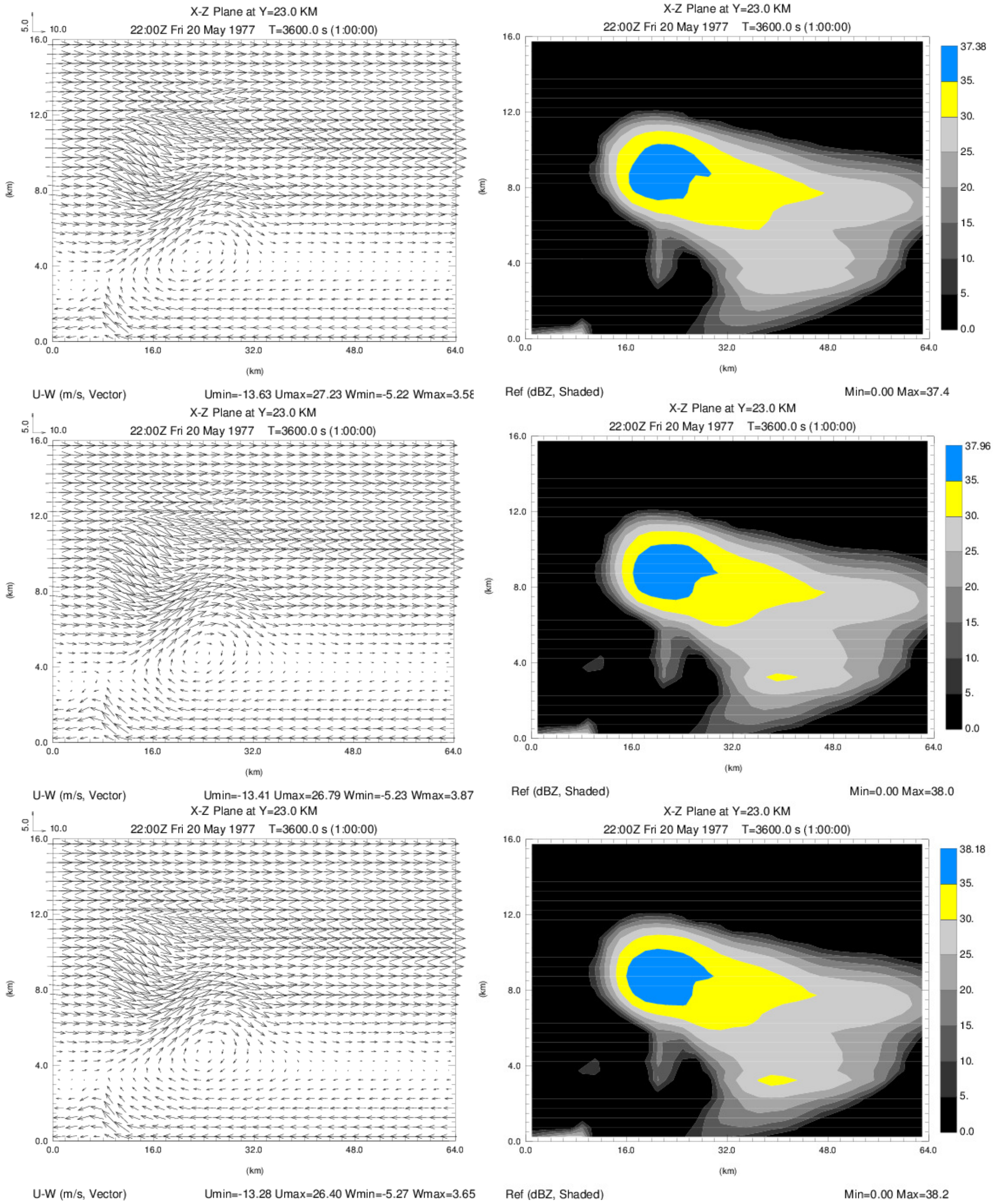

Figure 2. 13 Same as Figure 2.11, but for t= 60min

Experiment V: Assimilation of only radar radial wind V_R

In Experiment V, the LETKF and EnSRF are inter-compared through a data assimilation with only radar radial wind V_R The results from EnSRF with assimilation of both V_R and Z observations are used as a reference in this experiment. In Figure 2.14, the RMS errors from EnSRF and LETKF with only V_R are very close to each other from the beginning cycles to the final cycles for all analysis variables. It indicates that if only radar radial wind V_R is assimilated, EnSRF and LETKF have very similar performance in spin-up and final stable stages.

It should be noticed that, in the later spin-up stage (t=35min~50min), the RMS errors of LETKF are a little bit smaller than RMSEs of EnSRF for some variables, such as the wind components, u, v and w. Thus the LETKF outperforms EnSRF during the later spin-up stage when only radial wind is analyzed. The cause of this phenomenon might be the *effect of covariance localization on the balance* in the analysis of LETKF. Holland and Wang (2013) and Greybush et al. (2011) pointed out the R-localization causes less imbalance in the analysis than B-localization when same scale length is used. Although the imbalance caused by localization in the analysis already occurs from the first analysis step of the spin-up stage, the analysis error performance at the beginning step could not reflect it directly. The effect of imbalance has to be shown out through several model forecast steps. This is why LETKF outperforms EnSRF in the later spin-up stage. After the spin-up stage, the performance of data assimilation system is getting stable. In the stable stage the observations could not make the impact on the analysis as much as in the spin-up stage. Correspondingly the effect of localization on the imbalance also becomes insignificant. After 6~7 forecast steps, the model state in the final stable stage are in good balance, so the differences caused by between the performances of EnSRF and LETKF become trivial in the final steady stage.

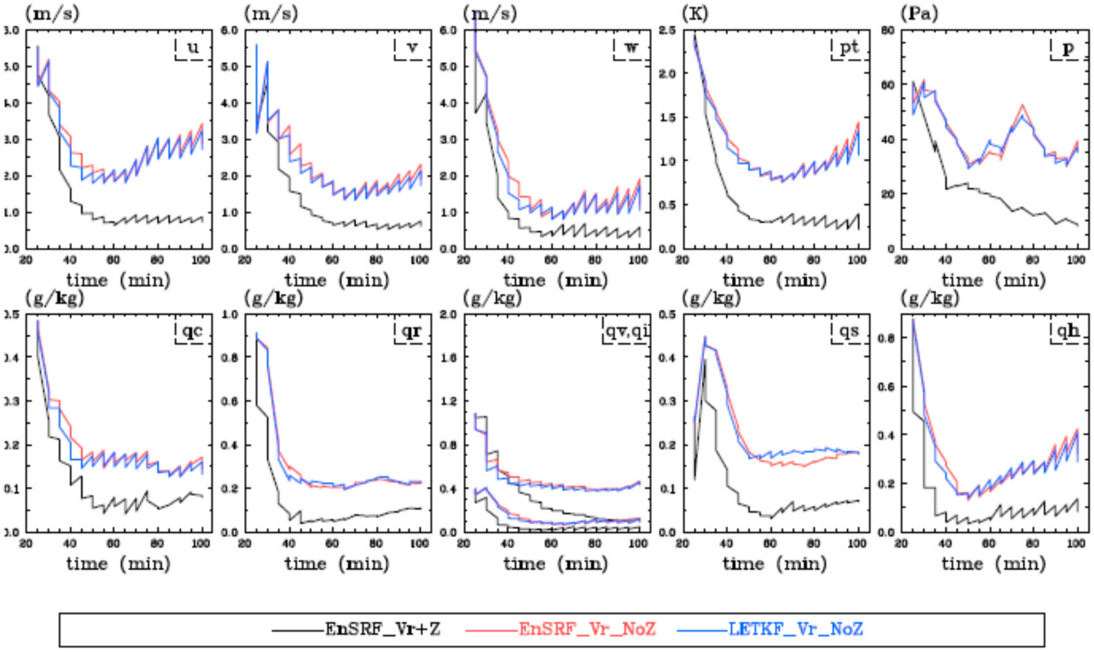

Figure 2. 14 Ensemble-mean forecast and analysis RMS Errors for experiments of EnSRF with only radial wind V_R (red), LETKF with only radial wind V_R (blue), and EnSRF with radial wind V_R and reflectivity Z (black, used as reference and control).

Experiment VI: Assimilation of only radar reflectivity

In Experiment VI, the LETKF and EnSRF are inter-compared through the data assimilation cycles with only radar reflectivity Z. The results from EnSRF with both V_R and Z data are used as reference and control run. In Figure 2.15, at the first several analysis steps, the reduction of the analysis RMS Errors of EnSRF is more significant than the reduction of analysis RMS Errors of LETKF for almost variables, such as u, w, p, q_r, q_s, q_h, q_v and q_i. Then after the spin-up stage, the RMS errors of LETKF and EnSRF are getting closer. Figure 2.16 shows the ensemble-mean analysis and forecast Root Mean Square Differences (RMSDs) to the observations for EnSRF and LETKF. In the middle panel for reflectivity observation Z, it is obvious that, in the first 5~6 analysis steps, the RMSDs of EnSRF get reduced more significantly than RMSDs of LETKF. Both of Figure 2.15 and Figure 2.16 indicate that analysis of EnSRF fits much better to true state and observation than analysis of LETKF in the spin-up stage when only radar reflectivity data is assimilated. EnSRF outperforms LETKF in the spin-up stage when reflectivity observations are assimilated. This is consistent to the results of the experiment IV in that both of radial wind and reflectivity are assimilated.

It should be also noticed that, in Figure 2.15, for some variables, such as w, q_r, q_s and q_h, although in the analysis step the analysis RMSEs of EnSRF decrease more significantly than LETKF, the forecast RMSEs of EnSRF increase also more significantly in the forecast step. Similarly in the middle panel of Figure 2.16 for reflectivity observation Z, the forecast RMSD of EnSRF jumps up back to the level very close to RMSDs of LETKF in the forecast step. It seems that if only reflectivity data is analyzed, EnSRF analysis is better than LETKF analysis, however, this advantage of EnSRF over LETKF is weakened in the forecast. The reason is not clear yet. Our understanding to this phenomenon is related the balance and adjustment in the convective-scale. In the convective-scale adjustment processing, the wind fields dominate over the mass and/or moisture fields. If new information (like reflectivity) only for the mass fields or moisture fields are assimilated in the analysis step so that mainly the moisture fields are updated, then through the adjustment process in which the wind fields dominate, the considerable part of the update to mass or moisture fields might be lost in the forecast step. So for the EnSRF analysis, although it is better and closer to the observations then LETKF analysis, so in EnSRF analysis there are more update for moisture fields by the reflectivity observations than in LETKF analysis, but there is no update from wind data to support the update in moisture field, and the B-localization in EnSRF might cause more imbalance in the analysis than LETKF, then EnSRF loses its advantage in the forecast. On the contrary, if the wind observations are assimilated together, then the assimilation of wind observation data could help to remain the new information in the mass and moisture fields. So in the results of the experiments with both radar radial wind and reflectivity data assimilated together, we could see the EnSRF outperforms LETKF in the spin-up stage consistently in both of analysis steps and forecast steps, not just only in analysis step like in the experiments of only reflectivity data.

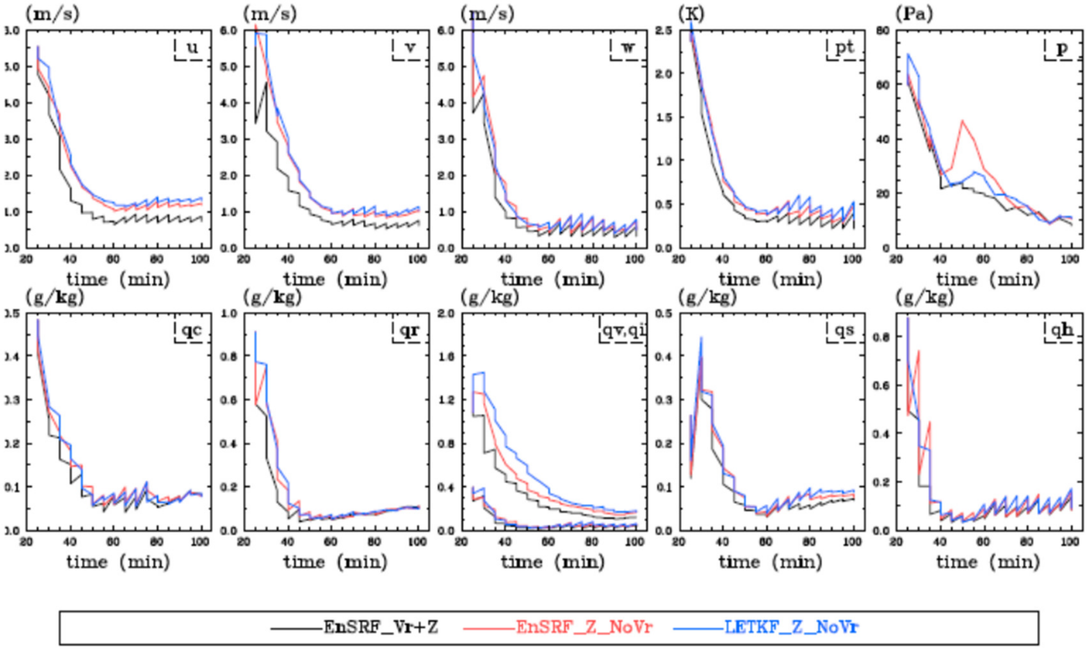

Figure 2. 15 Ensemble-mean Forecast and analysis RMS Errors for experiments of EnSRF with only reflectivity (red), LETKF with only reflectivity (blue), and EnSRF with radial wind and reflectivity (black, used as reference and control)

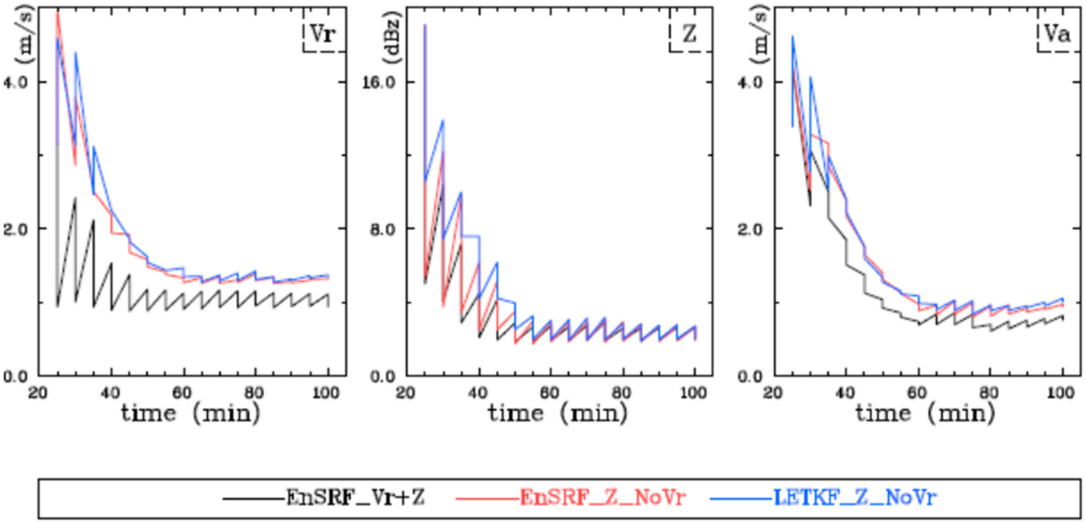

Figure 2. 16 Ensemble-mean Forecast and analysis RMS Differences (RMSDs) with radar the observations (radial wind , reflectivity and cross-beam wind) for experiments of EnSRF with only reflectivity (red), LETKF with only reflectivity (blue), and EnSRF with radial wind and reflectivity (black, used as reference and control)

As a brief summary for the inter-comparison experiments of ARPS-LETKF and EnSRF, EnSRF and LETKF could make comparable performances with their optimal localization radii. And the less imbalance in LETKF analysis may help LETKF performs better in the spin-up stage when using linear radial wind data. Those results are consistent with the previous research work on the inter-comparison of EnSRF and LETKF (Greybush et al. 2011; Holland and Wang 2013). However, the result that EnSRF outperforms LETKF significantly in the spin-up stage has not been reported and discussed in the previous research work. Considering the strong nonlinearity in the observation operator for radar reflectivity, the significant and systematic differences in performances of EnSRF and LETKF in the spin-up stage should be related to the effect of nonlinear observation operator on the EnKF algorithm. In the next session, it will be inspected.

2.3 Effect of Non-linear forward observation operator

The results in last session show that in the spin-up stage of DA cycles, EnSRF outperforms LETKF significantly when the nonlinear radar reflectivity observations are assimilated. In this session, we are going to investigate how the nonlinear observations effect on the analyses of EnSRF and LETKF and how it helps EnSRF output LETKF in the beginning cycles.

First let us look at how the nonlinear observation forward operator is involved when calculating the forecast error covariance used in Kalman gain of EnKF algorithm (Lorenc 2003b)

$$\mathrm{P}^f\mathrm{H}^T \approx \overline{\left(x^f - \overline{x^f}\right)\left(H\left(x^f\right) - \overline{H\left(x^f\right)}\right)^T} \tag{2.19}$$

$$\mathrm{HP}^f\mathrm{H}^T \approx \overline{\left(H\left(x^f\right)-\overline{H\left(x^f\right)}\right)\left(H\left(x^f\right)-\overline{H\left(x^f\right)}\right)^T} \tag{2.20}$$

As can be seen, the nonlinear observation operator H is directly used in the R.H.S. of the equations (2.19) and (2.20) to approximately calculate the estimates of the error covariance (which are in linear form) in the L.H.S, without needing its linearized operator H and the adjoint operator H^T. Although the linearization is not used directly in EnKF algorithm, the approximation in equations (2.19) and (2.20) is still called "*implicit linearization*" in (Lorenc 2003b). To understand how this linearization is made implicitly in equation (2.19) and (2.20), these 2 equations are derived out in detail as the following:

First applying the Taylor expansion to the nonlinear operator $H(x)$, the ensemble form of $H(x)$ is also called "ensemble observation prior"

$$H(x^f)=H(\overline{x^f})+\mathrm{H}(\overline{x^f})\mathrm{z}^f+O(\left\|z^f\right\|^2)\approx H(\overline{x^f})+\mathrm{H}(\overline{x^f})\mathrm{z}^f \tag{2.21}$$

where $z^f = x^f - \overline{x^f}$ is the ensemble perturbations, and the ensemble covariance is $P^f = Z^f Z^{fT}$. Applying the ensemble mean operator to the ensemble observation prior, or say, equation (2.21), with $\overline{z^f} = 0$, and ignoring the 2nd order and higher order terms, then

$$\begin{aligned}\overline{H(x^f)} &= \overline{H(\overline{x^f})+\mathrm{H}(\overline{x^f})\mathrm{z}^f+O(\left\|z^f\right\|^2)} = \\ &= H(\overline{x^f})+\mathrm{H}(\overline{x^f})\overline{\mathrm{z}^f}+\overline{O(\left\|z^f\right\|^2)}\approx H(\overline{x^f})\end{aligned} \tag{2.22}$$

With equation (2.21) and (2.22), the ensemble perturbations in observation space is as following

$$\overline{H\left(x^f\right)-\overline{H\left(x^f\right)}} = H(\overline{x^f}) + \mathrm{H}(\overline{x^f})\mathrm{z}^f + O(\left\|z^f\right\|^2) - H(\overline{x^f}) \approx \mathrm{H}(\overline{x^f})\mathrm{z}^f \tag{2.23}$$

Applying equation (2.23) into the R.H.S of equation (2.19) and (2.20),

$$\begin{aligned} &\overline{\left(x^f-\overline{x^f}\right)\left(H\left(x^f\right)-\overline{H\left(x^f\right)}\right)^T} \approx \overline{z^f\left(\mathrm{H}(\overline{x^f})\mathrm{z}^f\right)^T} \\ &= \overline{z^f z^{fT}\mathrm{H}(\overline{x^f})^T} = \mathrm{P}^f\mathrm{H}(\overline{x^f})^T \end{aligned} \tag{2.24}$$

$$\begin{aligned} &\overline{\left(H\left(x^f\right)-\overline{H\left(x^f\right)}\right)\left(H\left(x^f\right)-\overline{H\left(x^f\right)}\right)^T} \\ &\approx \overline{\left(\mathrm{H}(\overline{x^f})\mathrm{z}^f\right)\left(\mathrm{H}(\overline{x^f})\mathrm{z}^f\right)^T} \\ &= \overline{\mathrm{H}(\overline{x^f})z^f z^{fT}\mathrm{H}(\overline{x^f})^T} \\ &= \mathrm{H}(\overline{x^f})\mathrm{P}^f\mathrm{H}(\overline{x^f})^T \end{aligned} \tag{2.25}$$

Equations (2.24) and (2.25) are same as (2.19) and (2.20). The derivation above obviously show that the validity of using the full nonlinear operator directly to approximately estimate the error covariances in linearized form is based on the linearization of the nonlinear operator represented by equation (2.21). So although only full nonlinear operator is needed in EnKF algorithm, the linearization is actually calculated in equation (2.19) and (2.20) implicitly (Lorenc 2003b). Equation (2.21) indicates that this implicit linearization is made at the place of ensemble mean $\overline{x^f}$, denoted as the explicitly linearized observation operator $\mathrm{H}(\overline{x^f})$. And in this sense, EnKF is an implicitly linearized form of Kalman filter (KF) as an extension to the nonlinear observation scenario.

EnSRF and LETKF both adopt the implicit linearization for the nonlinear observation operator in their analysis algorithms. By inspecting the linearization in sequential observation processing of EnSRF and the simultaneous processing of

LETKF, it is found that they are choosing different places to linearly approximate the nonlinear relation between the state and observations priors. For the simultaneous observation processing in LETKF, the implicit linearization is made to fit over the entire prior ensemble for all nonlinear observations just for once before the whole analysis starts. For EnSRF, its sequential algorithm in nature requires the implicit linearization is re-made over the intermediate ensemble, not the entire prior ensemble used by the linearization in LETKF. This fact was also pointed out by Anderson and Collins (2007). For the nonlinear operator $H(x)$, its linearized operator $\mathrm{H}(x)$ itself is also a function of model state x. Then the linearization for same nonlinear observations but over different ensemble definitely generates the different linearization, or say, different linearized operator $\mathrm{H}(\overline{x^f})$. Even though this linearized operator is implicit for EnKF, the implicit linearization is different for EnSRF and LETKF indeed. Since the linearized operator $\mathrm{H}(\overline{x^f})$ is implicitly involved in the EnKF analysis algorithm, such as the calculation of Kalman gain and observation innovation, then the different linearization used by EnSRF and LETKF leads to the different final analyses for EnSRF and LETKF.

It should be also noticed that, for the sequential processing in EnSRF, the intermediate ensemble is the current ensemble analysis with the new information from the observations previously analyzed, so the intermediate ensemble statistically are the more accurate and better ensemble estimate of the model state than the entire prior ensemble used in LETKF. As we know, the linearization is valid and accurate locally around the place the linearization is made. So for EnKF, which essentially is the implicit linearized form of Kalman filter as an extension in the nonlinear observation problem, its analysis has only local optimality around the place where the linearization is made, not like the global optimality of the KF analysis for linear system. So if the linearization is re-made about a better model state which is closer to the observations or true state, then statistically this re-made linearization should help EnKF find a better analysis with local optimality around that better model state. In this sense, EnSRF, in which the implicit linearization is re-made about the inter-mediate ensemble, is superior to LETKF, in which the linearization is made about prior ensemble. So EnSRF could outperform LETKF. In fact, for the similar reason, Lorenc (2003a) claimed that "These implicit linearization (in EnKF) are fitted over the entire prior ensemble; the EnKF is inferior in this regard to 4D-Var which linearizes explicitly about the current best estimate." This is also one of the motivations to the iterative EnSRF (Sakov et al. 2012; Wang et al. 2013b) and RIP-LETKF (Yang et al. 2012).

According the discussion above, the linearization in EnSRF is sequentially re-made over the inter-mediate ensemble, which is different to, and more accurate and better than the entire prior ensemble used by linearization in LETKF. This leads to the different linearization for EnSRF and LETKF, and also makes LETKF inferior to EnSRF, then finally causes the biased differences in the analyses from EnSRF and LETKF. This is the effect of nonlinear observation forward operator on the EnKF analysis.

If the linearization for EnSRF and LETKF are more different, then correspondingly the differences in the final analyses from EnSRF and LETKF are more significant. To answer the question – why the differences between LETKF and EnSRF are significant in spin-up stage only and then turn to trivial, we should inspect the factors which contribute to the differences of linearization for EnSRF and LETKF. For the simplicity and also without losing generality, a simple 0-D system is used here for this purpose.

This model only has 1 model variable x and 2 independent nonlinear observations y_1 and y_2. Let x_b be the prior background. In LETKF, the linearized operators for the 2 nonlinear observations are both based on x_b, denoted as $\mathrm{H}_1(x_b)$ for y_1 and $\mathrm{H}_2(x_b)$ for y_2. In EnSRF, without losing the generality, assuming y_1 is analyzed first, the linearized operator for y_1 is also based on the prior background x_b, denoted as $\mathrm{H}_1(x_b)$ for y_1. Obviously for observation y_1, there is no difference in the linearization for LETKF and EnSRF. When observation y_2 is analyzed, the linearized operator for y_2 is re-made on the intermediate background (denoted as x_b^*) updated by observation y_1, denoted as $\mathrm{H}_2(x_b^*)$. The difference between $\mathrm{H}_2(x_b^*)$ (for EnSRF) and $\mathrm{H}_2(x_b)$ (for LETKF) is the difference of the linearizations for same nonlinear observation y_2. By Taylor expanding $\mathrm{H}_2(x_b^*)$ at x_b and ignoring the higher order terms,

$$\mathrm{H}(x_b^*) \approx \mathrm{H}(x_b) + \mathrm{H}'(x_b)(x_b^* - x_b) \tag{2.26}$$

Re-written (2.26), and the difference between the linearization for observations y_2 used in EnSRF (denoted as $\mathrm{H}_2(x_b^*)$) and the linearization for y_2 used in LETKF (denoted as $\mathrm{H}_2(x_b)$) is

$$\mathrm{H}(x_b^*) - \mathrm{H}(x_b) = \mathrm{H}'(x_b)(x_b^* - x_b) \tag{2.27}$$

Note that $\mathrm{H}(x)$ is the first-order derivative function of the nonlinear operator $H(x)$, $\mathrm{H}'(x)$ is the first-order derivative function of the linearized operator $\mathrm{H}(x)$, also is the second-order derivative function of the nonlinear operator $H(x)$.

The equation (2.27) shows that the difference between these two linearization for same observation is determined by two factors, $\mathrm{H}'(x_b)$ and $(x_b^* - x_b)$. The factor $\mathrm{H}'(x_b)$ represents the nonlinearity in the nonlinear operator . It is called *real nonlinearity* in Jazwinski (1970) because $\mathrm{H}'(x_b) \equiv 0$ if $H(x)$ is a linear operator. Another factor $(x_b^* - x_b)$ is the analysis increment due to the analysis of earlier processed observation in the sequential processing scheme, and it represents the differences between the places where the linearization is made in sequential and simultaneous processing schemes respectively. Although this factor is different to the *induced nonlinearity* in Jazwinski (1970), it is still called as *induced nonlinearity* in this dissertation for its function as an "inducer" and "amplifier" to stimulate the effect of *real nonlinearity* $\mathrm{H}'(x_b)$. These two factors work together to contribute to the significance of the effect of the nonlinear observation operator.

For this simple 0-D system, the *induced nonlinearity* $(x_b^* - x_b)$ is equal to the analysis increment by the analysis of 1st observation,

$$\begin{aligned} x_b^* - x_b &= \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}\left(y_1 - H_1\left(x_b\right)\right) \\ &= \frac{1}{1 + (\sigma_o^2 / \sigma_b^2)}\left(y_1 - H_1\left(x_b\right)\right) \end{aligned} \tag{2.28}$$

The equation (2.28) shows that the *induced nonlinearity* $(x_b^* - x_b)$ depends on σ_o^2 / σ_b^2 and $y_1 - H_1\left(x_b\right)$. Term σ_o^2 / σ_b^2 is the ratio of observation error variance σ_o^2 and background error variance σ_b^2, and term $y_1 - H_1\left(x_b\right)$ is the observation innovation for y_1.

Although the equations (2.27) and (2.28) are derived for a simple 0-D system, the *real* and *induced nonlinearity* in these 2 equations are useful concepts in the general EnKF algorithm when using the nonlinear observations. Especially equation (2.28) indicates that the 2 terms, σ_o^2 / σ_b^2 and $y_1 - H_1\left(x_b\right)$, could determine the significance of the effect of nonlinear observation operator through the induced nonlinearity. Another fact in equation (2.28) is that when the observation noise is much bigger than the forecast error, then the large noise could effectively “cover” the neglected nonlinearities (Jazwinski 1970).

This *induced nonlinearity* could be used to explain the question -- why the differences between the analyses from EnSRF and LETKF are significant only in the spin-up stage of the DA cycles. In the beginning cycles, the quality of the initial ensemble is poor, so that statistically the background error variance σ_b^2 is much larger than the observation error variance σ_o^2, then the ratio σ_o^2 / σ_b^2 is very small. For the same reason, the background x_b is fairly far away from the true. Correspondingly the observation prior $\mathrm{H}_1(x_b)$ departs from observation y_1 significantly. With equation (2.28), the *induced nonlinearity* $(x_b^* - x_b)$ becomes important. It leads to the significant differences in the analyses from EnSRF and LETKF in the beginning DA cycles. However, after the spin-up stage, the quality of the background ensemble are getting good enough, σ_b^2 could be equal to or even smaller than σ_o^2, so the ratio σ_o^2 / σ_b^2

increases. And the observation prior $\mathrm{H}_1(x_b)$ based on the better background is also much closer to the observations y_1, the observation innovation is small. Then *induced nonlinearity* $(x_b^* - x_b)$ is small, so the effect of nonlinear observation operator becomes trivial. So there is no significant difference between the analyses from EnSRF and LETKF in the final stable stage.

2.4 Sensitivity experiments for verification of the effect of non-linear observation forward operator

In this section, based on the theory about the effect of nonlinear observation discussed in section 2.3, several additional sensitivity experiments were designed to further test and verify this theory.

Experiment VII: Assimilation of the *Rarefied Reflectivity* Z data only

In this experiment, EnSRF and LETKF were inter-compared with a rarefied radar reflectivity dataset. The original simulated reflectivity data are located horizontally on the model grid points. The horizontal distance between 2 adjacent observations is equal to the horizontal grid spacing, 2 km. By selecting the reflectivity observations every 3 grid points from the original dataset, the horizontal distance between 2 observations in this thin reflectivity dataset is 6 km, which is just equal to the cut-off radius used by the covariance localization in EnSRF and LETKF. When the sparse reflectivity data is analyzed in EnSRF, the impact of each observation could not spread out to the grid points where the other observations are located due to the localization. Then for the sequential processing in EnSRF, the earlier analyzed observations could not influence the re-linearization for the following processed observations. So for both of the EnSRF with the sequential processing and LETKF with simultaneous processing, the implicit linearization for each non-linear observation is still made on the entire prior ensemble. This experiment is designed to remove the induced nonlinearity by rarefying the data with localization cut-off radius, correspondingly to void the effect of nonlinear observation operator, even though the real nonlinearity still exists. It is expected that the differences of the performances from LETKF and EnSRF with this thin reflectivity data be very small and trivial from the beginning DA cycles. It has to be noticed that, in this experiment, the original dataset is rarefied only within the horizontal plane, so the impact of observation still could spread out in the vertical column. So we still could see some differences of the analysis between EnSRF and LETKF.

In Figure 2.17, the RMS Errors of EnSRF and LETKF with the thin reflectivity data are very close to each other from the beginning cycles. In the middle panel of Figure 2.18, differences between the RMSDs of EnSRF and LETKF to the reflectivity observations are still observed, it is because the observations are not rarefied in the vertical column so that the effect of nonlinear observation still works but is limited only in the vertical direction. So the differences are much smaller than the differences between the RMSDs of EnSRF and LETKF with the original dense data set showed in the Figure 2.16. So the results of this experiment fulfill our expectation for this experiment.

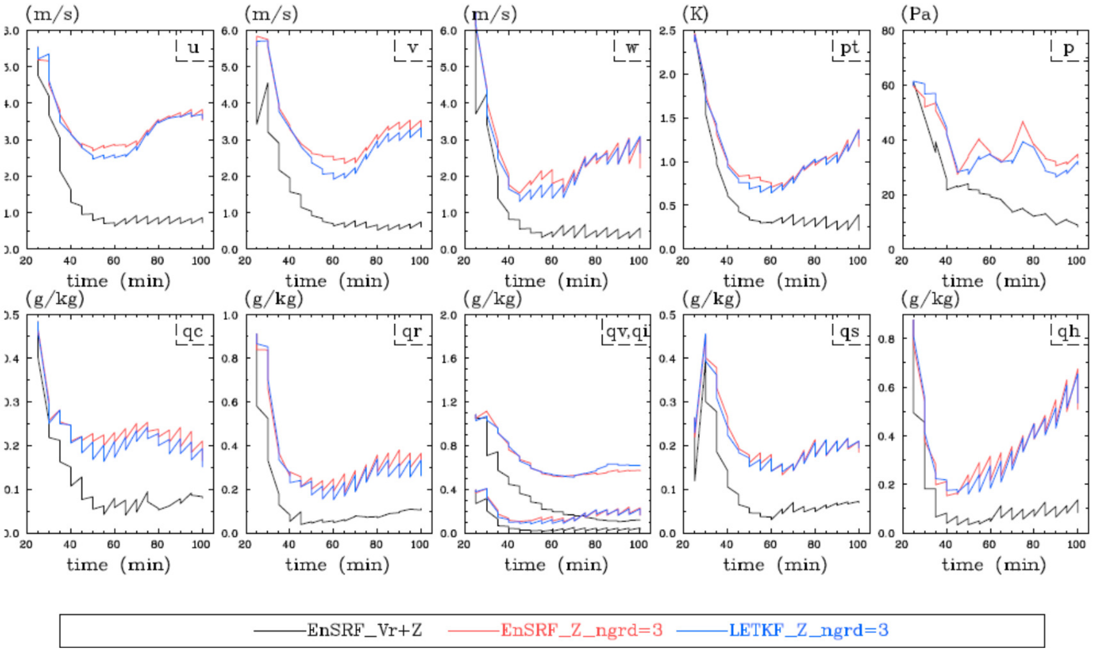

Figure 2. 17 Same as Figure 2.15, RMS Errors but for experiments of EnSRF with only rarified reflectivity data (red), LETKF with only rarefied reflectivity data (blue), and EnSRF with original dense radial wind and reflectivity data (black, used as reference and control)

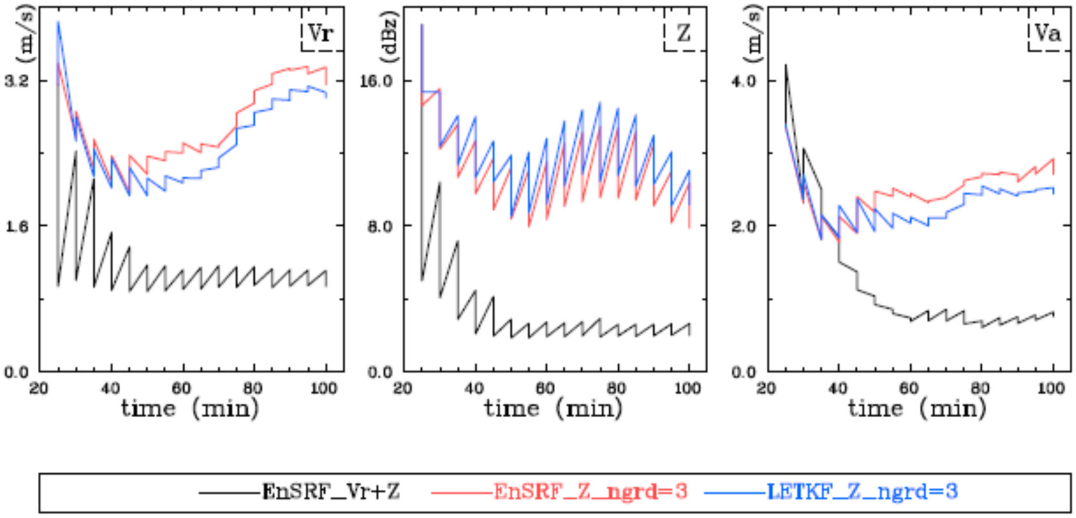

Figure 2. 18 Same as Figure 2.16, RMS Differences but for experiments of EnSRF with only rarified reflectivity data (red), LETKF with only rarefied reflectivity data (blue), and EnSRF with original dense radial wind and reflectivity data (black, used as reference and control)

Experiment VIII: LETKF Assimilating reflectivity data in two sequential batches

In this experiment, the original reflectivity dataset is split into two smaller datasets. Then LETKF processes the two smaller datasets in two analysis batches sequentially. In each analysis batch, the observations in the smaller dataset are still processed simultaneously. This experiment is designed to make the LETKF process the data in the way of partial sequence. Because for the 2nd analysis batch, the ensemble prior are already updated by the analysis from 1st analysis batch, based on this better ensemble prior, the implicit linearization used for the nonlinear observations in 2nd dataset should be more accurate than the linearization in LETKF which processes all data in 1 batch. It is expected the performance of the LETKF with 2 batches is better than the performance of LETKF with 1 batch, and is closer to the performance of EnSRF.

Figure 2.19 shows that for the analysis variables, u, v, w, q_s, q_r, q_h, and q_v, the RMS errors of LETKF processing the data in 2 analysis batches are improved obviously compared to the LETKF with 1 analysis batch, and is closer to the RMS Errors of EnSRF from the beginning DA cycles. The RMS Differences in Figure 2.20 also indicates that the analysis of LETKF with 2 batches is fitting the observations better than the analysis of LETKF with 1 batch. The results in this experiment also support our expectations.

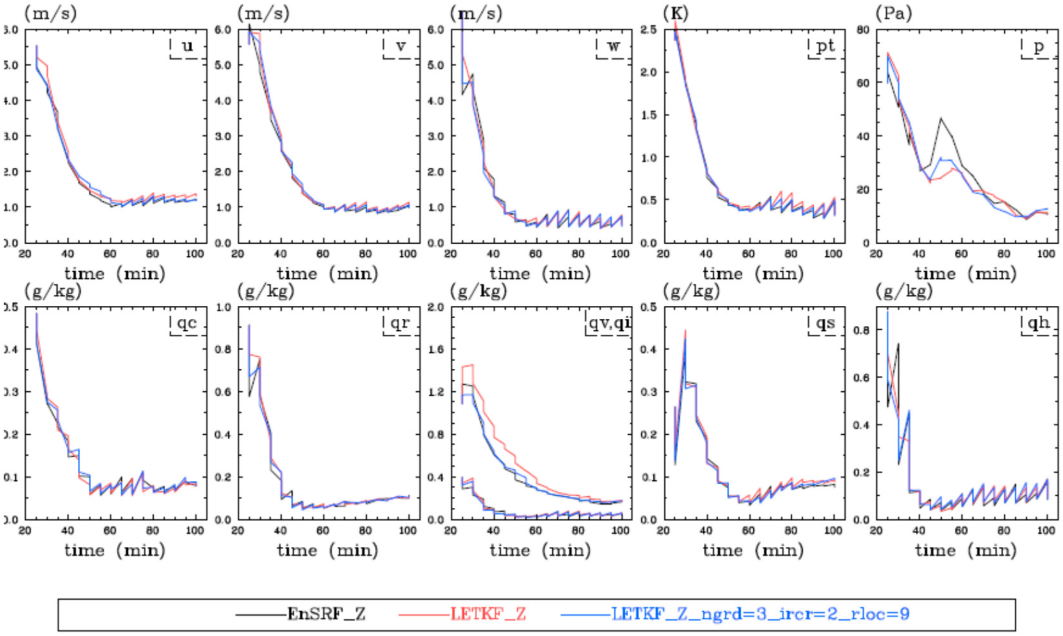

Figure 2. 19 Same as Figure 2.15, RMS Errors but for experiments of EnSRF with only reflectivity data (black), LETKF processing the reflectivity data all in 1 batch (red), and LETKF processing the reflectivity data in 2 batches sequentially (blue)

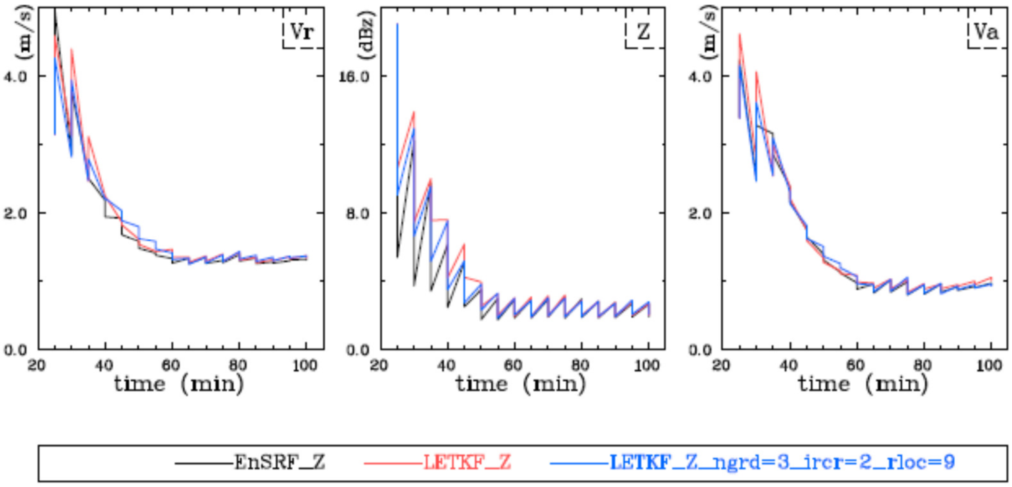

Figure 2. 20 Same as figure 2.16, RMS Differences but for experiments of EnSRF with only reflectivity data (black), LETKF processing the reflectivity data all in 1 batch (red), and LETKF processing the reflectivity data in 2 batches sequentially (blue)

Experiment IX: Assimilation with artificial observation $Z_{lin} = q_r + q_s + q_h$

As described in section 2.1.2, the analysis variables q_r, q_s, q_h are related to the radar reflectivity through the nonlinear observation forward operator. In this experiment, the q_r, q_s, q_h are related to an artificial observations through a linear observation forward operator, $Z_{lin} = q_r + q_s + q_h$. This artificial observation is called linear reflectivity here. Then LETKF and EnSRF are compared by assimilating this linear artificial observation. It is expected that the performances from LEKTF and EnSRF should be very close from the beginning DA cycles, just like the results of the experiments with radial wind data.

In Figure 2.21, for all variables, the differences between the RMS Errors of LETKF and EnSRF are very small and trivial in the spin-up stage. It also should be noticed that, there are some differences between EnSRF and LETKF in the later final stage. And there are some big spikes in the RMS errors for all the analysis variables in the first DA cycle. The reasons for these phenomena are not clear to us yet. It may be related to the imbalance introduced by the different localization scheme. In Figure 2.22, both of the analyses from LETKF and EnSRF fit the real reflectivity observation (in middle panel) and the artificial linear reflectivity observation (in the right panel) very well in the same way. There is no significant difference between the RMS Differences of EnSRF and LETKF observed in Figure 2.22. So the results in this experiment meet our expectation, too.

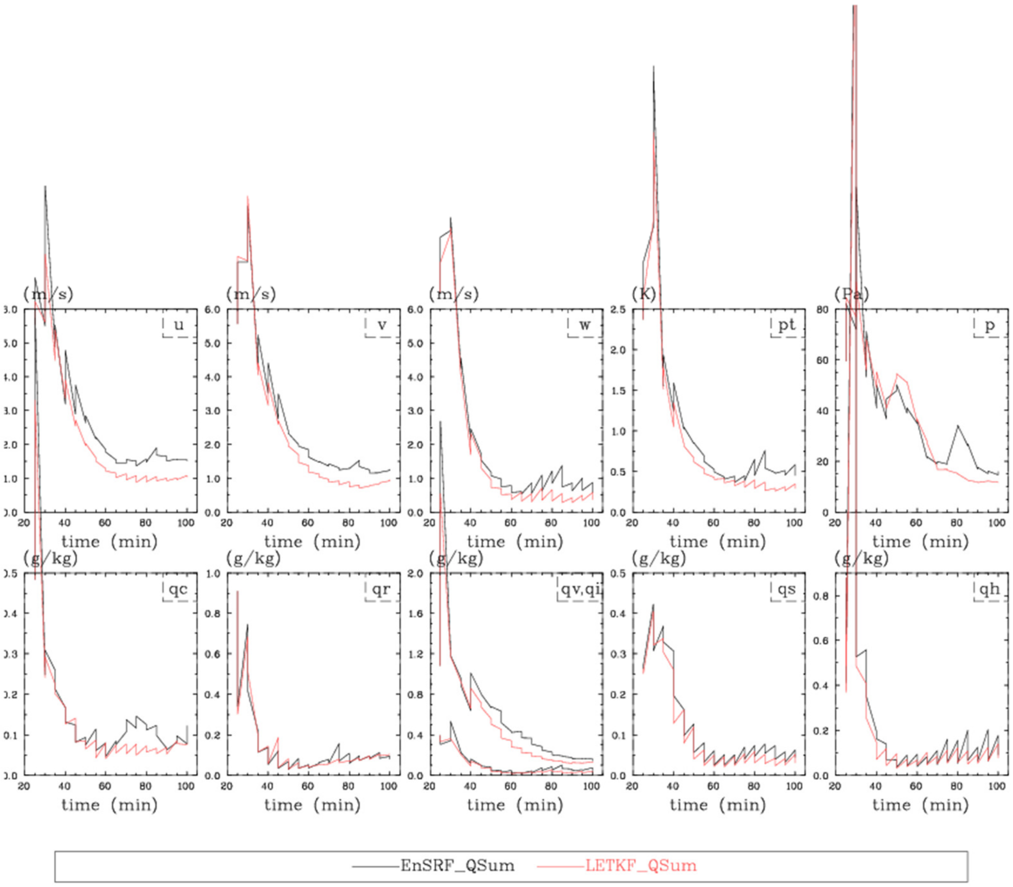

Figure 2. 21 Same as Figure 2.15, RMS Errors but for experiments of EnSRF with only the artificial observations (black), LETKF with only the artificial observations (red)

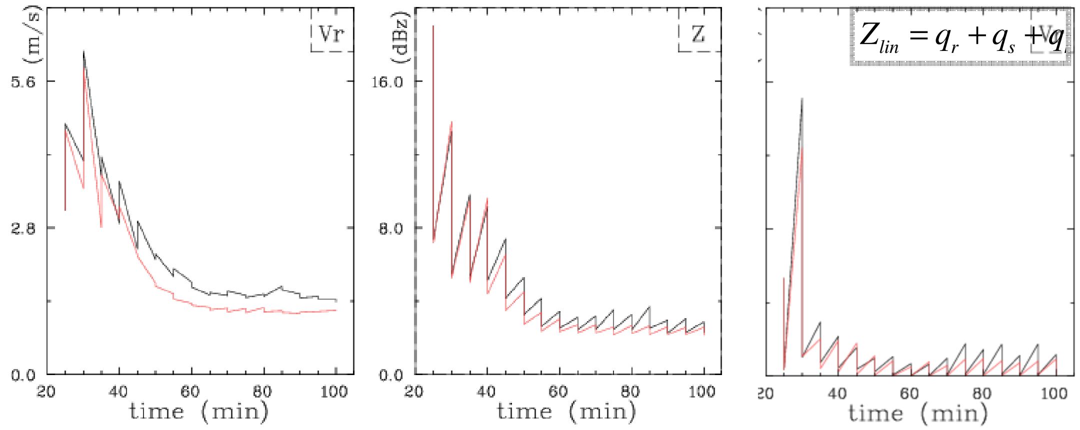

Figure 2. 22 Same as Figure 2.16, the RMS Differences but for experiments of EnSRF with only the artificial observations (black), LETKF with only the artificial observations (red). Note that the right panel is the analysis and forecast RMS Differences to the artificial observations, not to the cross-beam wind in Figure 2.16.

Experiment X: Iterative 3D-LETKF

In this experiment, the radar observations are iteratively used by LETKF. This is different to the experiment VIII, in which the observations are split into two datasets and processed in two batches sequentially, but every observation is analyzed for once only. In this experiment, every observation is analyzed for twice. The purpose of this experiment is similar to the experiment VIII. First analysis provides a better background for second analysis. So the implicit linearization in the second analysis is more accurate. Then the second analysis could fit the observations and the true state more closely. Essentially, this iterative analysis is just an application of the Newton-Raphson method for nonlinear equations into the nonlinear DA problem (Lorenc 1988). The approach of the iterative analysis is already widely used in Variational DA schemes (Courtier et al. 1994; Lorenc 1997; Courtier et al. 1998). And recently it has been also adopted into the EnKF framework, such as the RIP-LETKF (Yang et al. 2012), iterative EnSRF (Sakov et al. 2012; Wang et al. 2013b).

In Figure 2.23, for analysis variables q_r, q_s, q_h, the analysis RMS errors of the iterative LETKF is smaller than the analysis RMS errors of the regular LETKF in the beginning DA cycles. However, for the other variables, there is only trivial change, or even a little bit negative impact by the iterative analysis. In Figure 2.24, it also shows that the iterative analysis could more closely fit the observations. But after the forecast step, the forecast RMS Differences to the observations jump back to levels which is similar to the forecast RMSDs of regular LETKF. In the middle panel of Figure 2.24, the advantage of iterative LETKF only shows significantly in the beginning 2 DA cycles, then iterative LETKF lost the advantage and performed in the way similar to the regular LETKF in the following DA cycles.

The results in this experiment indicate that the iterative analysis could help to retrieve more information from the nonlinear observations and improve the analysis for the variables which are related to the nonlinear observations directly in the beginning of the spin-up stage. So in this sense, the results of this experiment satisfy our expectation.

It also should be noticed that the performances from the iterative LETKF are only improved in the beginning DA cycles compared to the non-iterative LETKF. And the non-iterative EnSRF even still outperforms the iterative LETKF. One of the reasons might be the imbalance caused by the over-fitting in the iterative analysis. The other reason could be the correlation between the background error and the observation error for the second analysis. EnKF algorithm is based on the assumption that the background error and observation error are un-correlated. Then with the violation of this assumption in iterative analysis, the sub-optimality in EnKF analysis is degraded.

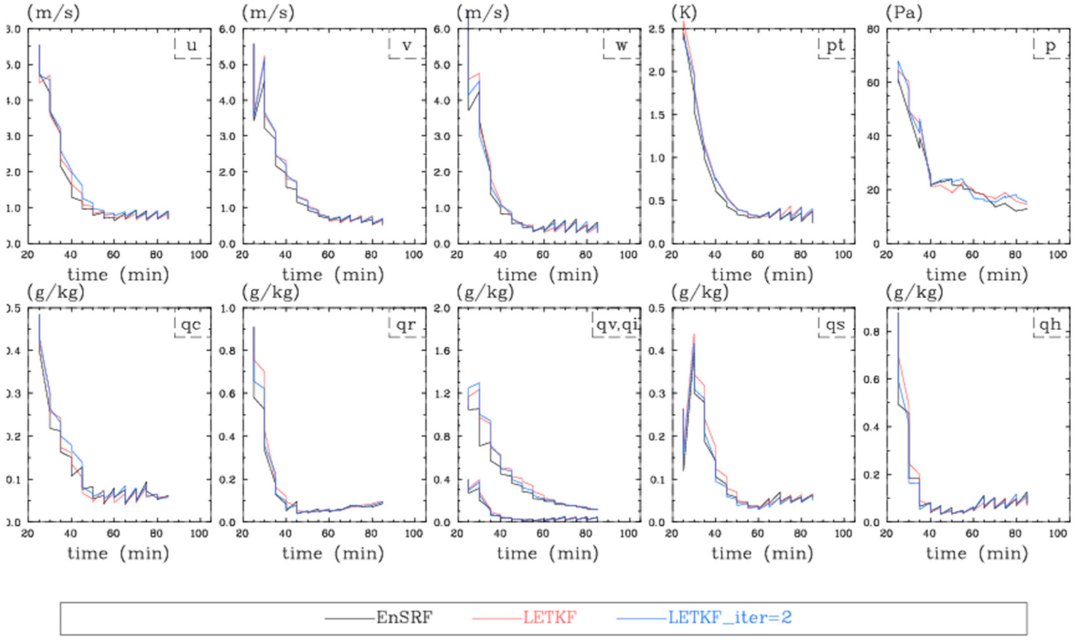

Figure 2. 23 Same as Figure 2.15, RMS Errors but for experiments of EnSRF (black), LETKF (red) and iterative LETKF (blue)

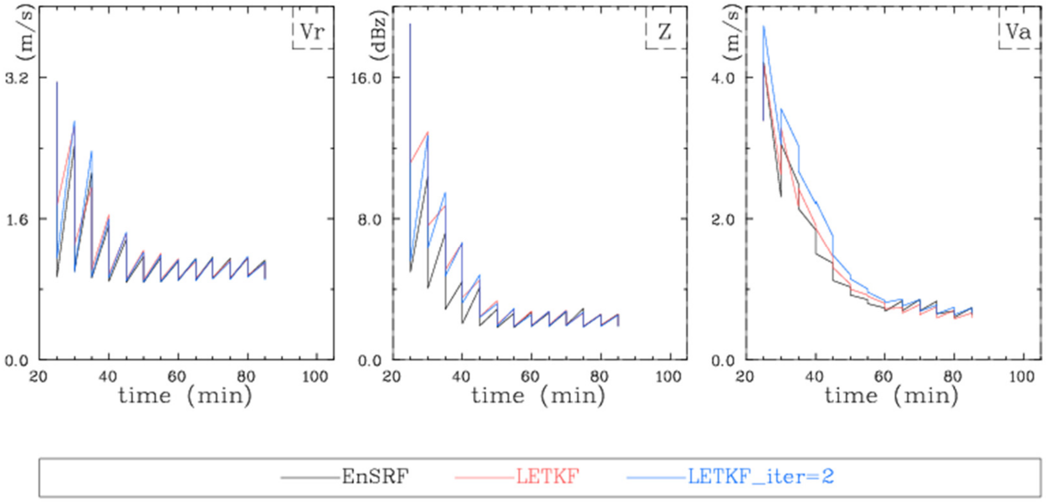

Figure 2. 24 Same as Figure 2.16, RMS Differences but for experiments of EnSRF (black), LETKF (red) and iterative LETKF (blue)

Experiment XI: OSSE with a new supercell added to the south of the original single supercell at t = 50 min

In this OSSE, a new supercell was added to the south of the original single supercell at t = 50 min. The purpose is to simulate a multi-cell storm case. In this multi-cell situation, the successively-emerging new cells could keep the analysis field not close to the observations through the data assimilation cycles. Then effect of nonlinear observation would keep working through the whole data assimilation cycles. So EnSRF might outperform LETKF throughout most of the DA cycles. This is not like in the single super-cell case, in the later DA cycles the difference of the performance from EnSRF and LETKF are very small due to the quality-improved analysis.

To add a super-cell, we have to use a different OSSE configuration with larger domain and a fast-moving storm. The similar OSSE configuration for a fast-moving storm in a large domain is used in Chapter 3, and the detail of this OSSE configuration could be found in Chapter 3. Figure 2.26 shows the composite reflectivity with 2 cells at t = 50min. The origin of coordinate is still at the left bottom corner of the domain. The radar location is at (0, 90km). Here we just describe how the second super-cell was added to the south of original super-cell. Figure 2.25 illustrates how to generate the truth simulation with the second super-cell. First, the truth simulation with original single supercell is made with the initial bubble centered at (45 km, 18 km, 1.5 km) at t = 0 min. This truth simulation is called “Truth-1” hereafter. At t = 25 min this original first warm bubble already develops into a super-cell and move northwards. Then the forecast of Truth-1 at t=25min is used as the base field for the 2nd super-cell. A new warm bubble, which is same as the first warm bubble, is placed at the same initial location as (45km, 18km, 1.5km) in the forecast field of Truth-1 valid at t=25min. And this forecast field with 2nd warm bubble is used as the initial conditions to make the truth simulation of 2 super-cells. This truth simulation is called Truth-2. Then the final truth simulation used for this experiment is a hybrid of Truth-1 and truth-2. Before t=50min, the truth simulation is from Truth-1, from t=50min, the truth simulation switches from Truth-1 to Truth-2. This combined truth simulation is called Truth-2cells. And the simulated radar observations are also based on this Truth-2cells. Figure 2.26 shows the composite reflectivity of the Truth-2cells at t = 50min. There are 2 super-cells, the original first super-cell and the 2nd "newly-emerging" super-cell to the south of first super-cell.

Then the analysis performances from EnSRF and LETKF are compared with the simulated radar observations based on the Truth-2cells. Figure 2.27 shows the RMSEs of EnSRF and LETKF in OSSE with 2 supercells. The RMSEs of EnSRF and LETKF in OSSE with 1 supercell is also plot as the control experiment and a reference. Before t=50min, EnSRF outperforms LETKF due to the effect of nonlinear observations. And after a few cycles, both of analyses of LETKF and EnSRF are improved and close to the truth and observations. The effect of nonlinear observations turns to trivial, and the differences of the performances of EnSRF and LETKF turns to insignificant at t=50min. At t=50mins, due the newly-emerging 2nd supercell, the analyses of LETKF and EnSRF are both "knocked" away from the new truth and observations with 2 supercells. So it is observed that there are jumps in the RMSEs of LETKF and EnSRF at t=50min. Then the effect of nonlinear observations is re-enhanced and turns to relatively remarkable again than before t=50min. Correspondingly it is observed that the RMSE of EnSRF get reduced faster than the RMSE of LETKF after t=50min. EnSRF outperforms LETKF after t=50min again in the first several analysis cycles due to re-enforced effect of nonlinear observations. This result matches our expectation to this experiment and supports our theory about the effect of nonlinear observations.

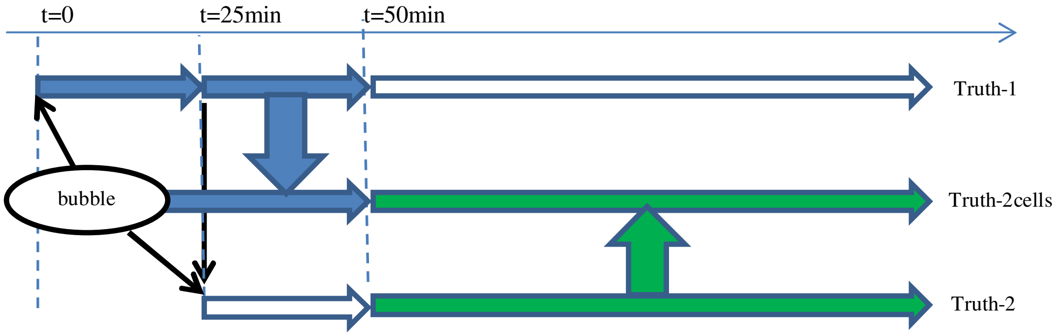

Figure 2. 25 schematic flow chart of the simulation with 2nd supercell

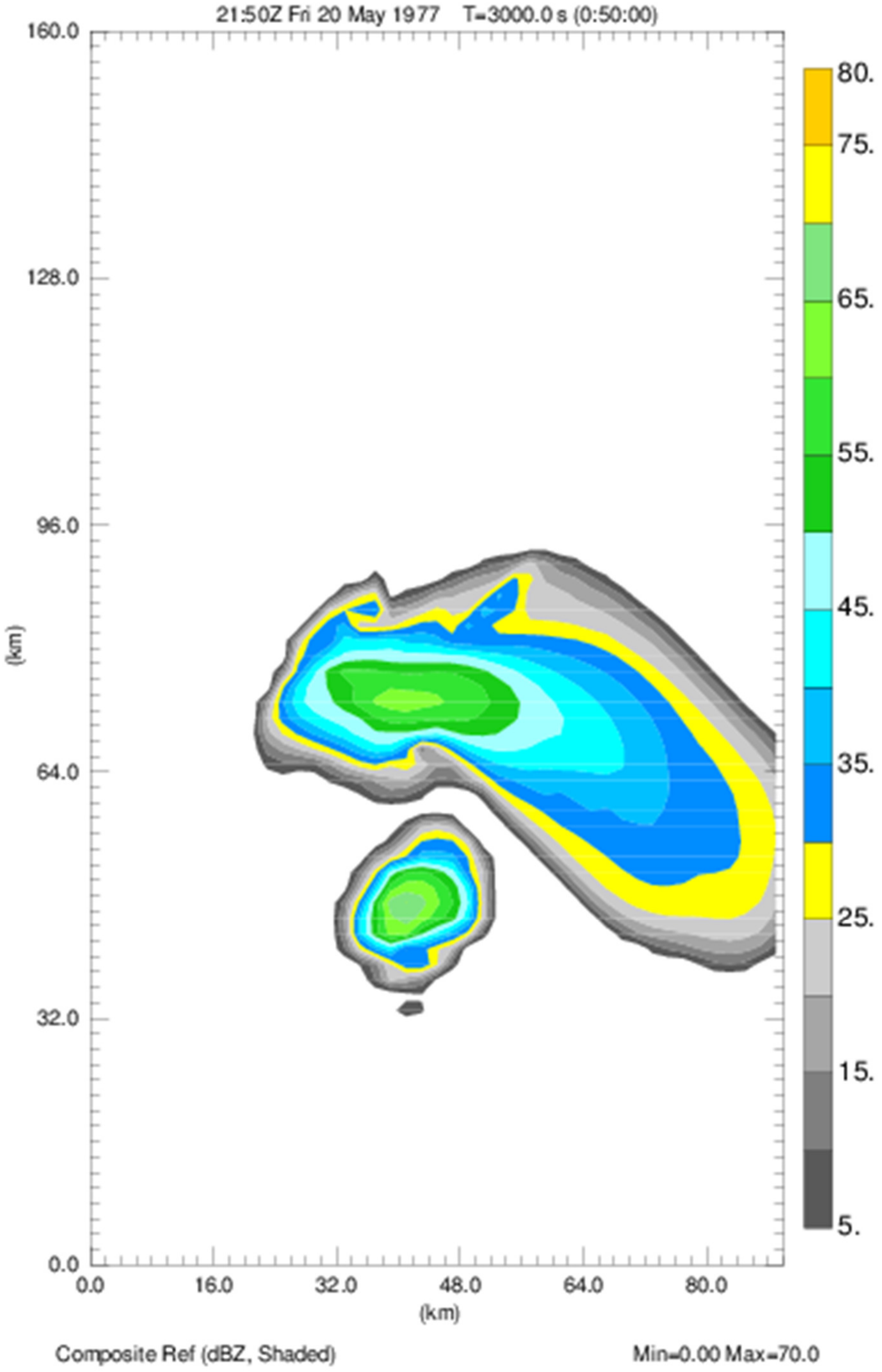

Figure 2. 26 the composite reflectivity with a new beginning supercell "added" to the south of original supercell at t=50min (3000sec)

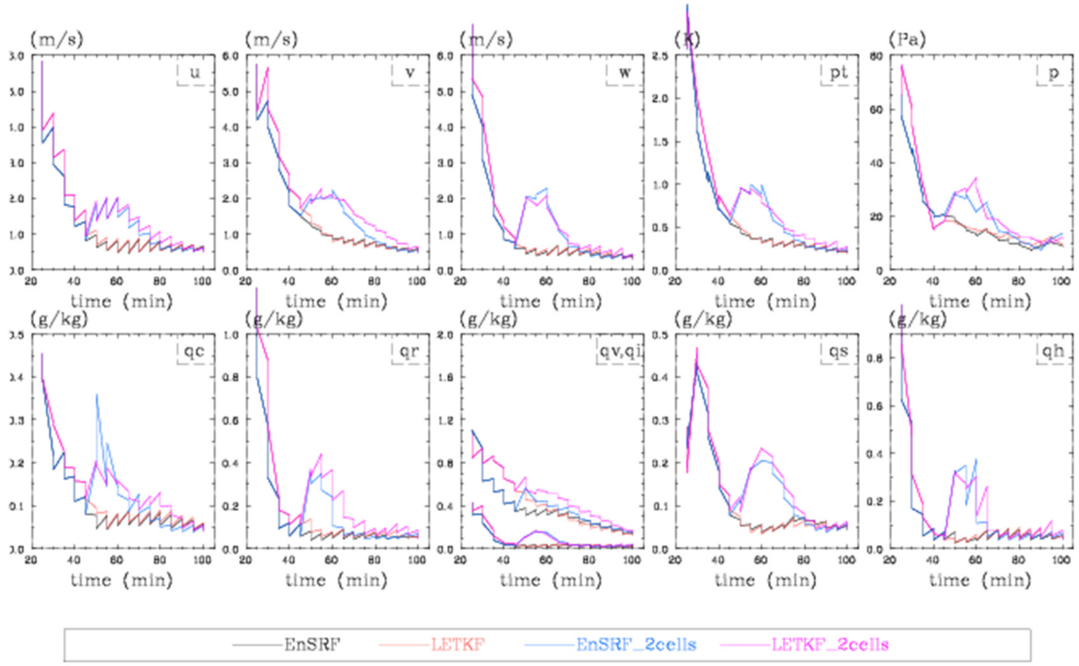

Figure 2. 27 RMSEs of EnSRF(Black) and LETKF(red) with single supercell, EnSRF(blue) and LETKF(purple) with a new supercell added to the south of original single spuercell at t-50min.

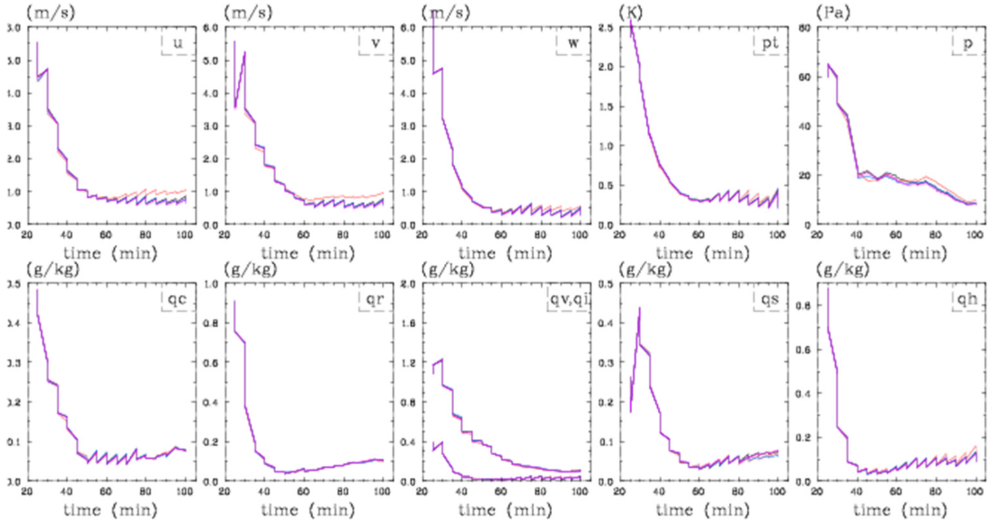

Figure 2. 28 RMSEs of LETKF without interpolation (black), with full variable interpolation (red), with incremental interpolation (blue) and with weight interpolation (purple).

As a summary for this section, to test and verity the theory about the effect of nonlinear observation operator, 5 sensitivity experiments (Experiment VII, VIII, IX, X and XI) are designed. In Experiment VII, the induced non-linearity is removed by rarefication of the data; in Experiment IX, the real non-linearity is removed by using artificial linear reflectivity observations. Then in these 2 experiments, the effect of nonlinear observation operator could not work. So the performances from EnSRF and LETKF are very close from the beginning DA cycles. In Experiment VIII, the data are analyzed in 2 batches by LETKF. In Experiment X the data are analyzed in the iterative approach by LETKF. The analysis of 1st batch in Exp. VIII and the first time analysis in Exp. X could provide a better prior ensemble for the second time analysis. Then the implicit linearization in second time analysis of LETKF is made over a better ensemble, which could be statistically closer to the intermediate ensemble used by the re-made linearization in EnSRF. In this sense, the effect of induced nonlinearity is weakened in both of Exp. VIII and X. So the performances of LETKF in these 2 experiments are improved and getting closer to the performances from EnSRF, at least in the beginning DA cycles. In experiment XI with newly-emerging 2nd supercell, the effect of nonlinear observations is re-enhanced by the new supercell, so EnSRF outperforms LETKF again in the later analysis cycles. The theory about the effect of nonlinear observation forward operator is verified in all of these 5 sensitivity experiments.

2.5 Summary and discussions

The Local Ensemble Transform Kalman filter (LETKF) is developed with the Advanced Regional Prediction System (ARPS). This newly-developed ARPS-LETKF data assimilation system is assessed by assimilating the simulated radar observations and inter-compared with the existing ARPS-EnSRF system. Since LETKF algorithm is designed to be a parallelizing filter (Miyoshi and Kunii 2012), this ARPS-LETKF provides a frame for future research of huge DA problem on massive parallel computers, but also an additional inter-comparison test-bed to the ARPS-EnSRF system in convective-scale data assimilation with radar data.

In the experiments of ARPS-LETKF data assimilation with simulated radar observations, root-mean-square-error (RMSE) decreases significantly and quickly in the beginning cycles and then reaches a very low and steady level. The results of ARPS-LETKF with radar observations are consistent to the results of ARPS-EnSRF. So this newly-developed ARPS-LETKF system could effectively assimilate the radar observations and produce the good analysis which fits the true state very well after a few DA cycles under the perfect model assumption.

The experiments for comparison of different localization schemes show that with their optimal localization cut-off radii respectively, the performances from EnSRF with B-localization, LETKF with R-localization and with regulated R-localization are very similar after the spin-up stage. The optimal cut-off radius for R-localization is shorter than the radius for B-localization. This result is consistent with the result in Greybush et al. (2011) and Holland and Wang (2013). But the regulated R-localization and B-localization use the same optimal cut-off radius when using multiple observations. Similar performances between regulated R-localization and B-localization indicate that the effect of regulated R-localization could be coarsely equivalent to B-localization for the case of the multiple observations. By using same localization cut-off radius, the differences in data selection and data influence of regulated R-localization and B-localization are reduced to minimum. So the inter-comparison of EnSRF with B-localization and LETKF with regulated R-localization is more equitable.

The experiments for the inter-comparison of EnSRF and LETKF with radar observations show that, when either the radial wind observations only, or the reflectivity observations only, or both of them, are assimilated, after the beginning spin-up stage, the performances from LETKF and EnSRF with their optimal cut-off radius are very close in the final steady stage. In the spin-up stage, when only the radial wind data are analyzed, LETKF outperforms EnSRF in the later DA cycles of the spin-up stage, but the differences between EnSRF and LETKF are insignificant and turn to trivial in the final steady stage. When only reflectivity data are analyzed, EnSRF analysis outperforms LETKF analysis significantly in the analysis step, but in the forecast step the EnSRF tends to lose its advantage in the error performance. When both of radial wind and reflectivity data are assimilated, the EnSRF outperforms significantly LETKF consistently in both of analysis and forecast steps, so that EnSRF reaches the final steady stage earlier than LETKF by 1~2 DA cycles. The differences in error performances from EnSRF and LETKF are significant and systematic in the spin-up stage, then turn to trivial in the final stale stage. Considering the strong nonlinearity in the observation operator for radar reflectivity, the results that EnSRF outperforms significantly LETKF in the spin-up stage should be related to the effect of nonlinear observation operator on the EnKF algorithm.

By inspecting the implicit linearization in EnSRF and LETKF, it is found that EnSRF and LETKF choose the different places to make the implicit linearization. The linearization in EnSRF is sequentially re-made over the inter-mediate ensemble, which is different to, and more accurate and better than the entire prior ensemble used by linearization in LETKF. This leads to the different linearization for EnSRF and LETKF, and also makes LETKF inferior to EnSRF, then finally causes systematic/biased differences in the analyses from EnSRF and LETKF. This is the effect of nonlinear observation forward operator on the EnKF analysis. The real nonlinearity and induced nonlinearity are responsible for the significance of the effect of the nonlinear observation operator. Especially the evolution of induced linearity with time determines the temporal change of differences between the analyses from LETKF and EnSRF. In the spin-up stage, due to the poor quality of initial ensemble, the significant induced nonlinearity makes the nonlinear effect significant, then leads to significant differences between LETKF and EnSRF. But in the later steady stage, prior ensemble is improved through several DA cycles and closer to the observations, then induced nonlinearity and the effect of nonlinear observation turn to trivial.

Five additional experiments are designed to test the theory about the effect of nonlinear observation operator, especially the effect of the induced nonlinearity. In 2 experiments, by using artificial linear reflectivity observations to remove the real non-linearity, or by rarefying the nonlinear observations to remove the induced non-linearity, the nonlinear effect stops working, then the performances from EnSRF and LETKF are very close from the beginning DA cycles. In another 2 experiments, by analyzing observations twice with iterative LEKTF, or by splitting the observations into 2 datasets and processing them in 2 batches sequentially, the effect of induced non-linearity is weakened, then the performances of LETKF in these 2 experiments are improved and getting closer to the performances of EnSRF, at least in the beginning DA cycles. And in the fifth experiment, a new cell was added to south of original supercell at t=50min when the analyses of EnSRF and LETKF are both very close to the truth and observations. Then analysis is “knocked” away by the new cell from the truth and observation. So the effect of nonlinear observation is enhanced, then EnSRF outperforms LETKF in the next a few analysis cycles again. All the results from these additional experiments meet our expectations. The theory about the effect of nonlinear observation forward operator got verified in these 5 sensitivity experiments.

Through the investigation of the reason for causing the different performances from EnSRF and LETKF when assimilating the nonlinear observations, it is realized that the implicit linearization is an important concept in EnKF algorithm. When the implicit linearization is applied to different assimilation patterns, i.e. the sequential vs. simultaneous observation processing (Holland and Wang 2013), EnKF might behave differently, e.g., EnSRF could outperform LETKF at the spin-up stage. And in the sense of implicit linearization, EnKF is an implicit linearized form of Kalman filter as an extension into the nonlinear observation problem. So it also has the limitations as a linearization approximation to the nonlinear problem, such as the local optimality, not global optimality. And in the KF framework, the estimates for the mean and covariance are not enough to describe the problem with non-Gaussian PDF caused by nonlinearity (Lorenc 2003b). So although in EnKF the nonlinear operator could be directly used by means of implicit linearization, how to effectively assimilating the strong nonlinear observations, such as radar reflectivity, polarimetric radar measurements and satellite radiances, are still very challenging for EnKF and its variants. Iterative analysis is an approach for EnKF when dealing with the nonlinear problems.

The less imbalance in LETKF analysis was also observed in our experiments, especially when linear observations are assimilated. The effect of localization on the imbalance in analysis was discussed in Greybush et al. (2011) and Holland and Wang (2013) with their large scale models. Localization should also introduce the imbalance to analysis in convective scale. But the balance in convective-scale is somehow complicated and not dominant in most time, not like the geostrophical balance in large scale. So how to evaluate the effect of localization on balance in convective scale is a concerned issue and might be worth to be investigated in the future. Using the approximate form of some model equations or diagnostic relation as a constraint in the framework of EnKF could be a way for balance issue. It will be discussed in chapter 4.

Chapter 3: Four Dimensional Extension of ARPS LETKF

3.1 Motivations for 4D extension of ARPS LETKF

Convective-scale weather systems, such as super-cell thunderstorms, usually evolve quickly and have short lifespans, ranging from 30 minutes to a few hours. Because of this, observations with high density in space and time are very useful for convective-scale NWP. Modern weather radars provide high frequency, nearly continuous observations. How to effectively and efficiently exploit the high frequency observations to How to use the data assimilation system to effectively utilize as much as possible from these frequent observations to create the initial conditions for a convective-scale NWP model is still challenging. Usually there are three strategies to utilize such frequent data. Figure 3.1 shows the schematic flow chart of the data utilization in these strategies.

The first strategy is to bundle observations into many small groups, and then analyze each small bundle at frequent intervals. This is the so-called intermittent assimilation cycle (Hu and Xue 2007; Dowell and Wicker 2009; Wang et al. 2013a). This strategy processes the frequent data sequentially in a very short time interval (e.g., 1 minute in figure 1 (a)). The analysis and forecast are both made with high frequency. It could introduce imbalances into the analysis and forecast field. The frequent switch between analysis and forecast also involves more input and output of DA system and correspondingly increase the computational cost.

The second strategy is to group the frequent data into larger data batches, meaning a longer time window (e.g,, 5 minutes in figure 1(b)). Then the data in this larger batch is processed in one step of analysis, saving computation cost when compared to serial analysis of small batches in first strategy. In this second strategy, all the data in the larger batches are distributed in time, but are all utilized as valid at the same single time — the analysis time. Obviously this strategy has observation timing error, which could raise serious problems when weather systems are evolving quickly.

To solve the large timing error problem in the second strategy, as well as save on computational costs, the third strategy uses a 4D analysis algorithm that could assimilate the observations collected over a longer time window simultaneously like in the second strategy, but use these asynchronous observations at their observation time (e.g, 5minutes in figure 1(c)). The 4DVar is a 4D assimilation algorithm that does this by searching the model trajectory which best fits the asynchronous observations distributed over a specific time window. Based on the EnKF framework, several variations of the 4D EnKF algorithms have been proposed, including the 4D-LEnKF of Hunt et al. (2004) and its computational-efficient upgrade version — 4D-LETKF proposed by Hunt et al. (2007) Hunt, Kostelich et al. (2007). Sakov et al. (2010) also proposed a more general form of Asynchronous EnKF (AEnKF).

Wang et al. (2013a) developed the asynchronous 4DEnSRF system and applied it to the convective-scale radar DA problem with the WRF model and tested it with simulated radar data in an OSSE of a fast-evolving storm. Their results show that the 4DEnSRF outperforms the EnSRF when the cycle length is more than 1 min in the presence of observation timing error. For long cycle lengths 4DEnSRF improves the analysis by utilizing more data, whereas the EnSRF performs well only when data far away from the analysis time are discarded. Their results show that the 4DEnSRF algorithm better utilized observations collected over time than the regular EnSRF algorithm.

Compared to the algorithm of the EnSRF, it is simple to upgrade the current LETKF framework to a 4D version. In this work, the ARPS-4DLETKF system was implemented on the current LETKF framework and tested with simulated radar data in OSSEs of a convective-scale storm.

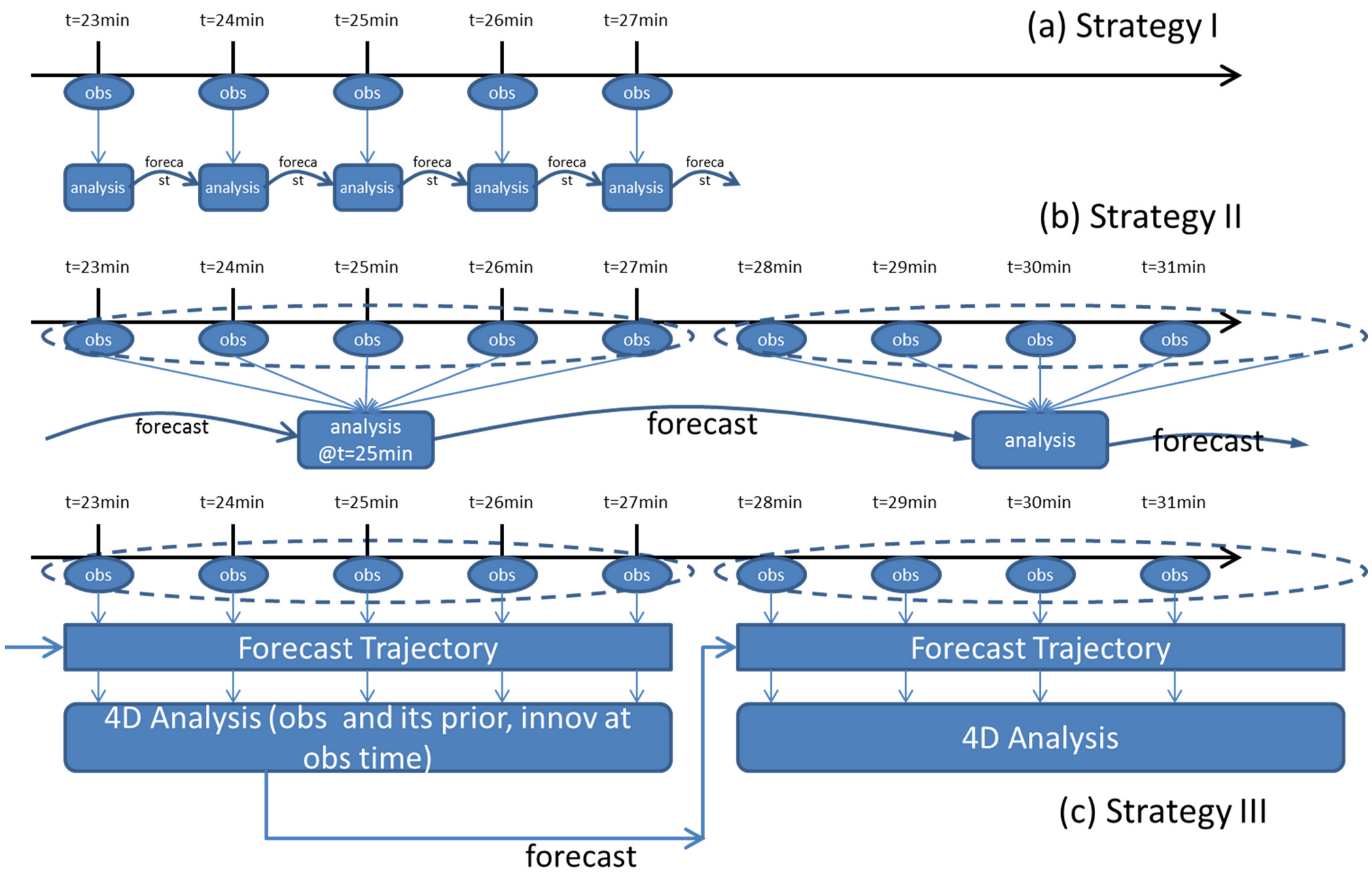

Figure 3. 1 Schematic flow chart for 3 strategies to utilize the asynchronous observations distributed in time. (a) Strategy I: frequent intermittent assimilation (SE1m in Table 3.1); (b) Strategy II: regular 3D-LETKF with longer time window (SE3m, SE5m and SE10m in Table 3.1); (c) Strategy III: 4D-LETKF (4D3m, 4D5m and 4D10m in Table 3.1);

3.2 Methodology

3.2.1 Algorithm and Implementation

The regular 3D-LETKF algorithm uses a linear combination of ensemble model states that best fit the observations at the analysis time. In 4D-LETKF, instead of searching for the best linear combination of ensemble model state to fit the synchronous observations only at analysis time, a search is made for the linear combination of ensemble model trajectory which best fits the asynchronous observations through the data time window. From 3D to 4D-LETKF, the same algorithm is used to find the weight W for the best linear combination. The difference is that all of the array/matrix observation variables and observation-involved variables need to extend to include the same type of variables but from different observation times. The form of algorithms for 4D and 3D is almost the same.

The following equations are same as the equations (2.8) and (2.9) in section 2.1.4 of Chapter 2

$$\overline{w}^a = \tilde{P}^a (Y^b)^T R^{-1} (y^o - \overline{y}^b) \tag{3.1}$$

$$\tilde{P}^a = \left[(K-1)I + (Y^b)^T R^{-1} Y^b \right]^{-1} \tag{3.2}$$

When (4.1) and (4.2) extend to the 4D algorithm, the observation vector y^o, ensemble observation prior mean vector $\overline{y}^b$, ensemble observation prior perturbation matrix Y^b and the observation error covariance matrix R will be the combined vectors or matrices by concatenating their corresponding variables at different observation times. It should be noted that when calculating the ensemble observation prior variables, observation operators must be applied to the ensemble background at the observation time. In the 3D algorithm the operators are applied to the ensemble background only at the analysis time. With the combined observation vector y^o , combined ensemble observation prior mean vector $\overline{y}^b$, combined ensemble observation prior perturbation matrix Y^b , and the combined observation error covariance matrix R , equations (3.1) and (3.2) keep the same formulas when extending from the 3D to the 4D algorithm. The derivation of the 4D LETKf algorithm can be found in Harlim and Hunt (2007) and Hunt et al. (2007).

Since the algorithm is almost the same, the implementation of 4D-LETKF is particularly simple. The modification to the original 3D-LETKF system is to pre-calculate the ensemble observation priors from the ensemble background states at the observation times, instead of at analysis time as would be done in the 3D algorithm.

It should be noted that, compared to the 4DEnSRF algorithm, 4DLETKF, just like 3DLETKF, simultaneously processes the observations. The observation priors are only calculated once based on the entire ensemble prior at different observation times before the analysis. In the 4DEnSRF and EnSRF, the observations are processed serially, and the observation priors have to be updated after one observation is analyzed. In EnSRF, the model state is updated first by analysis of one observation, then observation operators are applied to the updated model state to update the observation prior. In 4DEnSRF, the computational cost will be very expensive if the model state is updated at all the observation times before the observation prior is updated for all observation times within the time window. Then, based on some linear assumptions, the observation prior can be updated directly by the analyzed observation in 4DEnSRF, without updating the ensemble model state and mean at observation times. The detail could be found in Sakov et al. (2010) and Wang et al. (2013a).

It is also worth noting that, in 4DLETKF, the weight ***W*** is time-independent: it could be applied to the ensemble model state at any time within the assimilation time window. This ***W*** is used to make the optimal trajectory fit during the time window. Allowing this time-independent W to be applied to the ensemble model state at different times requires the linearity assumption that the evolution of ensemble perturbations is linear (Evensen 2003; Hunt et al. 2007; Sakov et al. 2010). So the length of the time window should not be very long for strong nonlinear dynamic systems, such as convective-scale weather. Otherwise the trajectories may diverge enough that the linear combination of the ensemble model trajectories will not approximate the real model trajectory (Hunt et al. 2007).

3.2.2 Localization

As previously stated, the ensemble observation priors in the 4D-LETKF algorithm include background information at different observation times, meaning that $\overline{w}^a$ and $\tilde{P}^a$ in (3.1) and (3.2) involve the ensemble priors at different times. The ensemble covariance in ensemble space, $\tilde{P}^a$, implicitly includes the asynchronous covariance, both spatial and temporal. Therefore, in 4DLETKF, the localization has to be implemented not only in space, but also in time. In this 4D-LETKF, the regulated R-localization is used as the non-adaptive localization scheme, which is similar to the regulated R-localization used in regular LETKF in Chapter 2. To consider the influence of time localization, the time localization weight ρ_t is multiplied by the spatial localization weight ρ_s. This means the non-adaptive localization weight $\rho_N = \rho_t \rho_s$ is same as used in Wang et al. (2013a). The spatial and temporal localization weights are both calculated by using the fifth-order correlation function of Gaspari and Cohn (1999) with the user-specified cut-off radius for space and time respectively.

The hierarchical filter of Anderson (2007) is applied into our 4DLETKF system as the adaptive localization scheme. The adaptive localization weight is defined as the regression confidence factor (RCF), and is estimated by the hierarchical filter (HF). In Anderson (2007),

$$\beta = \frac{\sigma_{x,y}}{\sigma_{y,y}} \tag{3.3}$$

β is the regression coefficient used in an EnKF-like algorithm to update the model state, $\sigma_{x,y}$ is the sample covariance between analysis variable x and observation variable y, and $\sigma_{y,y}$ is the sample variance of observation variable. β works like the Kalman gain. The HF estimates the confidence, or uncertainty of β by using the resampling method. This confidence factor of the regression coefficient (RCF) is then used as the localization weight for the regression coefficient β itself when updating the model state.

It should be pointed out that the regression coefficient β is mainly determined by the sample correlation between the analysis variable x and the observation variable y. The confidence factor, or uncertainty of β is basically similar to the uncertainty of the correlation coefficient. The textbook formula for the standard error of the correlation coefficient is as following,

$$\sigma_{corr} = \frac{1-\hat{\rho}_{corr}^2}{\sqrt{n-3}} \tag{3.4}$$

where σ_{corr} is the standard error of the correlation coefficient ρ_{corr}, *n* is the sample size, and $\hat{\rho}_{corr}$ is the estimate of the correlation coefficient. If using the standard error of correlation to represent the uncertainty of the correlation, (3.4) indicates that large correlation tends to have small standard errors. Correspondingly, the large value of β will have less uncertainty, meaning they will be given more confidence. In other words, large values of β will be assigned a large regression confidence factor (RCF), ρ_a, i.e. the localization weight. The HF method is essentially a method where, if there is strong sample correlation between the observation variable and the state variable, a larger weight will be assigned to the corresponding element in the Kalman gain by the Shur product. Therefore, the influence of observations is applied to this model state variable. If the sample correlation between the observation and model state variables are small, then a very small weight will be given to the corresponding element in the Kalman gain. There would be only trivial influence from the observation to the model state variable.

As mentioned before, β is a variant of regular Kalmann gain. The RCF is designed as a B-localization weight. If it is applied to the R-localization, it can be regulated in the same way as the regulated R-localization in Chapter 2 (see equation (2.18)). Furthermore, the formula of the hybrid nonadaptive and adaptive localization is defined as following,

$$\rho = (1-c)\rho_N + c\rho_A \tag{4.5}$$

where *c* is the weight for adaptive localization part. *c*=0.5 was also used in Wang et al. (2013a).

3.3 OSSE system and data assimilation experiment design

3.3.1. Truth Simulation

This 4D-LETKF system was tested in the OSSE with 20 May 1977 Del City, Oklahoama, supercell storm (Ray et al. 1981), which is also used for the OSSE in Chapter 2. The ARPS model is used in both the simulation and data assimilation cycles. Two truth simulations were created: one for a slow-moving storm system in the small domain (referred to as SMS hearafter), and one for a fast-moving storm system in the large domain (referred to as FML hearafter). These similar truth simulations had been used in Xue et al. (2006). Both simulations are initialized with same initial bubble in the same way as in Chapter 2.

The SMS simulation is the same in all aspects as the truth simulation used in the OSSE in Chapter 2, except that the center location of the initial bubble is moved from x=24 km (used in Chapter 2) to x=48 km. The model configuration, domain configuration, etc., are as same as used in Chapter 2. The model resolution has a horizontal grid spacing of 2km and vertical grid spacing of 500 m; the 3D grid point network is 35×35×35. Horizontally, it covers a 64kmx64km physical domain. Figure 3.2 shows the composite reflectivity of the truth simulation for SMS at t=1800 sec, 3600 sec, and 5400 sec.

The FML simulation uses the same model resolution as the SMS. To accommodate the fast-moving storm, the 3D grid point network is 48×83×35. The coverage in the vertical direction is the same as in SMS simulation: 500m. The horizontal physical domain is 90km×160km. And the center location of the initial bubble is at x=45km, y=18km. Note that z=1.5km which is the same as used in SMS, as well as in Chapter 2. The turbulence mixing scheme in the FML simulation is the Smagorinsky mixing coefficient. This is different than the 1.5-order turbulent kinetic energy (TKE) scheme used in the SMS simulation and in Chapter 2. To make the storm move fast, a constant wind of u=3ms^{-1} and v=14ms^{-1} is added to the sounding data used in SMS simulation. Figure 3.3 shows the composite reflectivity of the truth simulation for SMS at t=1800 sec, 3600 sec, and 5400 sec.

3.3.2. Simulation of Radar Observation

Similar to the simulation of radar data in Chapter 2, one WSR-88D radar is involved in this OSSE study. For the WSR-88D radar, standard precipitation-mode parameters are assumed, including 10-cm wavelength, 1° beamwidth, and a total of 14 elevations, with the lowest elevation at 0.5° and the highest at 19.5°. Also similar to the observation simulation in Chapter 2, it is also assumed that the simulated observations are available in the plan position indicator (PPI) planes, or the elevation levels of the radar, rather than at the model grid points. We do assume that the radar observations are already interpolated to the Cartesian coordinate in the horizontal directions; in other words, the observations are found in the vertical columns through the model scalar points. The observation operators are also same as describe in Chapter 2. In the OSSE for the SMS, the ground-based radar is located at the x-y coordinate origin, while in the OSSE for the FML, the radar is located at x=0 km and y=80 km. In Figure 3.2 and Figure 3.3, the red dot indicates the location of radar for OSSE of SMS and FML respectively.

In this work, we use the same volume scan strategy as used in Wang et al. (2013a). The volume scan strategy is described as following (cited from Wang et al. (2013a)):

> *Each volume scan spans 5 min. Following Yussouf and Stensrud (2010), the lowest 12 elevations of observations are collected at a rate of three elevations per minute and the upper two elevations are collected during the last minute of each volume scan with observations stored in data files minute-by-minute. Data in each file are assumed to be observed simultaneously and these 1 min data files are referred to as raw data files.*

The first raw data file here is at t=23 min of the model time, not at t=21 min as done in Wang et al. (2013a).

The data collection strategy in this work is different to the strategy used in Chapter 2. In Chapter 2 the full volume scans are assumed to have been collected simultaneously, i.e. all the radar observations on the 14 radar scan tilts are observed at same time, which is also the analysis time. Then after the interval of 5 min, another simultaneous full volume scan is made and the data are used for next analysis time.

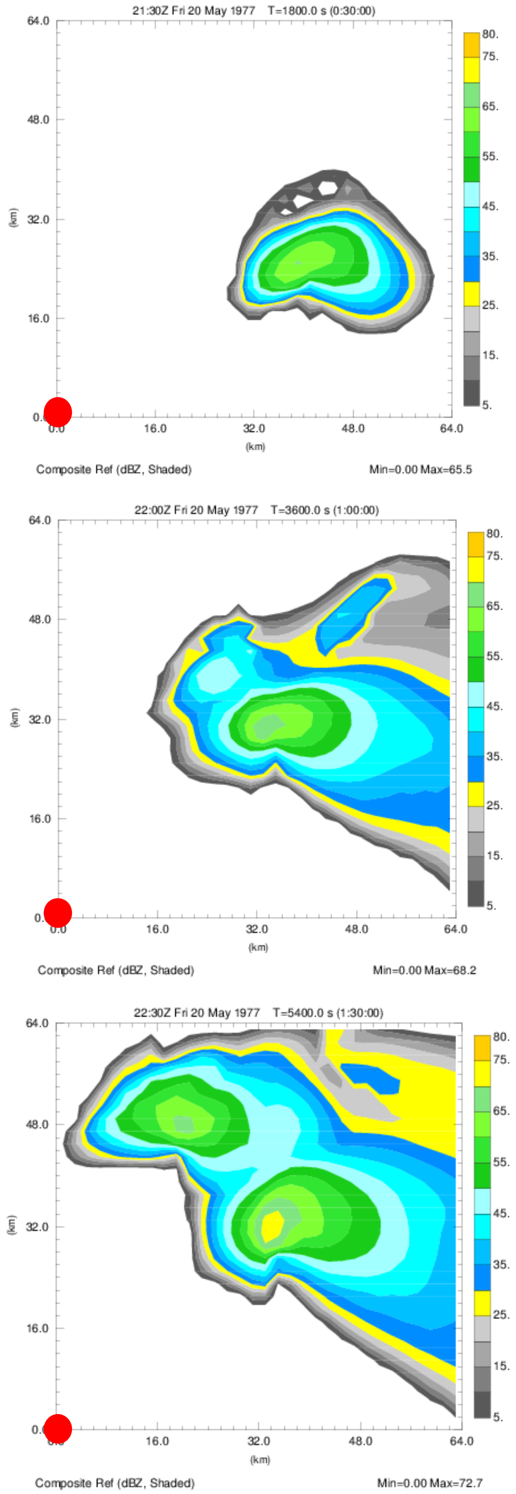

Figure 3. 2 composite reflectivity of truth simulation for SMS at t=30min (top), 60min (middle), and 90min (bottom)

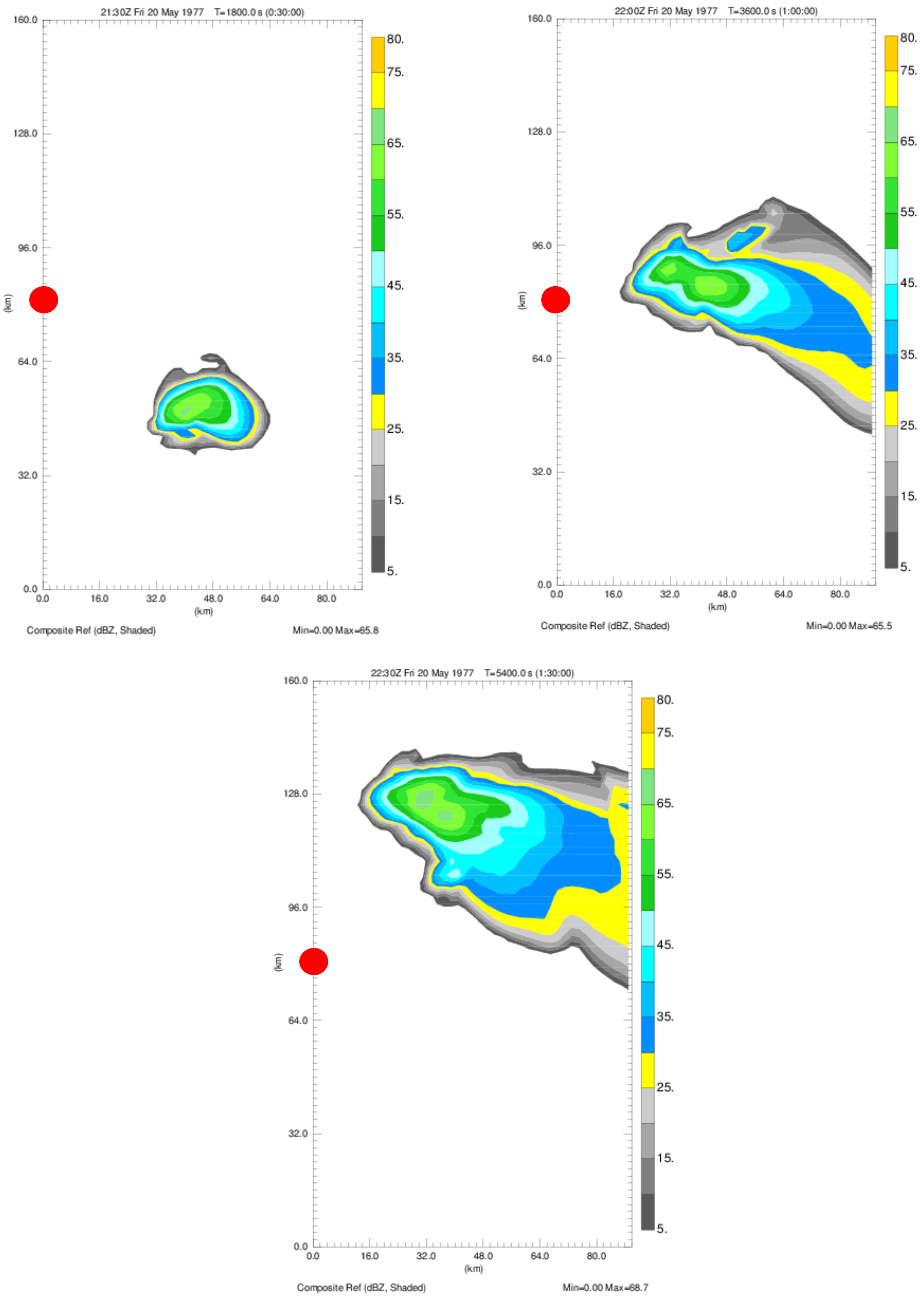

Figure 3. 3 Ccomposite reflectivity of truth simulation for FML at t=30min (top left), 60min (top right), and 90min (bottom).

4.3.3 Data Assimilation Experiment Design

The 4DLETKF and regular 3DLETKF both use 40 ensemble members, initialized in the same way as described in Chapter 2. Initial ensemble members are initialized at t = 20 min of model time by adding spatially smoothed perturbations to the initially horizontally homogeneous first guess defined by the Del City, Oklahoma, sounding. The standard deviations of the perturbations added to each variable are 2 ms^{-1} for u, v, and w; 2 K for θ'; and 0.6 g kg^{-1} for the mixing ratios of hydrometeors (q_v, q_c, q_r, q_i, q_s, and q_h). The perturbations are added to the velocity components, potential temperature, and specific humidity in the entire domain excluding grids composing the lateral boundaries. For the mixing ratios, the perturbations are added only to the grid points located within 6 km horizontally and 2 km vertically from the observed precipitation. Negative values of mixing ratios after the perturbations are added are reset to zero. The pressure variable is not perturbed.

Different to the multiplicative covariance inflation scheme used in Chapter 2, the relaxation inflation scheme (Zhang et al. 2004) is adopted here. The weight for the ensemble background is set to 0.5 as in Wang et al. (2013a). The spatial localization cut-off radius is 6km in both of horizontal and vertical directions. For the calculation of RCF of the hierarchical filter used as adaptive localization, 40 ensemble members are divided into five groups of eight members each, which is different to the eight groups of five members each used in Wang et al. (2013a),; our earlier test showed that this group strategy was best for our case.

The main purpose of this portion of the work is to compare the performance of the two analysis schemes (4D-LETKF and 3D-LETKF) when using the asynchronous observations distributed over a time window. The experiments were designed for this purpose. Table 3.1 lists out the experiment s. The rules for naming the experiments are similar to the rules used in Wang et al. (2013a) Wang, Xue et al. (2013).

Table 3. 1 Experiment names for 4DLETKF and 3DLETKF

Experiment name	**Analysis scheme**	**Analysis start time**	**Data Batches (min)**
SE1m	3DLETKF	23	1
SE3m	3DLETKF	24	3
4D3mw	4DLETKF	24	3
SE5m	3DLETKF	25	5
4D5mw	4DLETKF	25	5
SE10m	3DLETKF	27	10
4D10mw	4DLETKF	27	10

As in Wang et al. (2013a), SE in the experiment name means the regular 3DLETKF scheme is used; 4D represents the 4DLETKF scheme. The ‘nm’ or ‘nmw’ in the experiment names indicates the data batch length (also known as time window length) with ‘n’ indicating minutes. The observation data in the ‘nm’ data batch are collected over n minutes centered at the analysis time. In our radar data simulation, each raw data file includes 1 min batch data; 1 min is the shortest interval in our OSSE. So SE1m represents the first strategy to group the high frequency data into small batches and analyze these short batches of data serially with frequent intervals. SEnm (n=3, 5, 10) represents the second strategy to analyze the asynchronous observations in longer batches, and assume all the observations to be valid at the analysis time, while 4Dnmw (n= 3, 5, 10) represents the third strategy to use the asynchronous data distributed over a time at their valid observation time.

Considering the flow-dependency in the covariance localization, hierarchical filter was implemented in this 4D-LETKF as the adaptive localization scheme, and also tested in the OSSEs of SMS and FML. Table 4.2 lists the experiments to test the adaptive localization scheme with the hierarchical filter.

Table 3. 2 Experiment names used in tests of the hierarchical filter

Experiment name	
NonAL	Non-adaptive localization, i.e. the regulated R-localization including the spatial and temporal localization.
AL	Adaptive localization with hierarchical filter (HF)
RAL	Regulated adaptive localization with HF
HRAL	Hybrid of regulated adaptive localization and regulated R-localization. The weight of adaptive localization is 0.5
NoRL	No R-localization and adaptive localization, only the data selection in the cut-off radius.

3.4 Experiment Results and Discussion

The 4DLETKF and regular 3DLETKF are compared in the OSSE case. As in Chapter 2, the Root Mean Square (RMS) Error of ensemble-mean forecast and analysis is used to show the performance of each DA system.

3.4.1 3DLETKF with intermittent assimilation cycle (1 minute window) vs. 4DLETKF with varying data window length

First is a comparison of the performance of the 3DLETKF with 1min data batches to the 4DLETKF with varying data batch length. Figure 3.4 shows the evolution of the RMSEs from SE1m, 4D3mw, 4D5mw and 4D10mw in the OSSE of FML. And Figure 3.5 shows the evolution of ensemble spread of SE1m, 4D3mw, 4D5mw and 4D10mw in the OSSE of FML. The analysis cycle of 3DLETKF with frequent intervals (SE1m) outperforms 4DLETKF in the early cycles. The main reason for the outperformance of SE1m is that the temporal covariance structure of 4DLETKF did not gain information from observations as much as the 3DLETKF in the beginning cycles. 4DLETKF has a relatively poor quality covariance structure. After several beginning analysis cycles, the model state and covariance structure improve and the RMSEs of 4DLETKF with 3min and 5min data batches reduce down to the same level as the RMSEs of 3DLETKF with 1min data batches. Also note that there are more jumps in the RMSE of perturbation pressure from the 3DLETKF with 1min data batches. That indicates an imbalance due to the impact of high intermittent analysis and very short forecast interval.

An inter-comparison of the performance of the 4DLETKF with varying batch lengths is shown in Figure 3.4. The 4DLETKF with shorter data batches outperforms

4DLETKF with longer data batches. The advantage of 3min over 5min is not very significant because they both reach the same performance level after a few cycles. The 4DLETFK with 10min data batches, however, performs poorly compared to 3min and 5min data batches. The final RMSEs of 4D10mw remain about 2–3 times of RMSEs of 4D3mw and 4D5mw. The reason the shorter data batches have a shorter spin-up period may be related to the poor quality temporal covariance structure. After several beginning analysis cycles, the amount of observations utilized by 4DLETKF with different time windows (e.g., 3min and 5min) should be same — or at least very similar — but the poor quality temporal covariance could do more harm to the 4D analysis with long window (5min) than to the short window (3min). It would prevent the 4DLETKF with the long window from gaining as much information as the 4DLETKF with a short window, making the 4D analysis algorithm less efficient than the 4D analysis with short window in the early cycles. As long as the time window is not too long, the 4DLETKF with longer window (e.g., 5min in this case) still could have similar performance after the spin-up cycles. The relatively poor performance of 4DLETKF with 10min data batches (4D10mw) in the final cycles may be due to the time window being too long for this fast-moving storm. The linearity assumption required by 4D EnKF algorithm might break up, which would lead to failure of the 4D10mw. Considering the additional computational cost and possible imbalance in the analysis related to the shorter data batch (e.g., 1 min), and the possible break-up of the linearity assumption by the longer window (e.g., 10 min), the 5 min window could be a good balance between performance and computational cost.

3.4.2 4DLETKF vs. 3DLETKF with longer DA time window (3 minutes and longer)

Next, the performance of the 4DLETKF is compared to the 3DLETKF with data batches longer than 1 min. Figure 3.6, Figure 3.7 and Figure 3.8 show the RMSEs from 4DLETKF and 3DLETKF with 3min, 5min and 10min data batches in the OSSE of FML, respectively. The 3DLETKF clearly performs poorly compared to the 4DLETKF throughout all DA cycles in the OSSE of fast-moving storm. In final analysis cycles, the RMSE of 3DLETKF is still as high as about twice that of 4DLETKF. The main reason for this is the observation timing error in the 3DLETKF, which treats all the asynchronous data be valid at the analysis time. Figure 3.9 shows the 4DLETKF vs. 3DLETKF with 5 min data batches in OSSE of SMS. For the slow-moving storm case, 4DLETKF still performs better than 3DLETKF with 5min data batches, though the differences are small. In the final analysis cycles, the performance from 4D and 3DLETKF are comparable. This indicates that the significance of the observation timing error is also related to the nonlinearity and evolution of the weather system.

3.4.3 Effect of Time Localization

Statistical analyses with longer time windows tend to have more spurious temporal correlations. In this work the time localizations are mainly applied to the experiments with longer data batches, such as SE10m and 4D10mw. This part of the analysis examines the effect of time localization. Figure 3.10 shows the RMSEs of SE10m without and with time localization in the FML (fast-moving storm) case. The cut-off radius for time localization is 600 sec (10 min). The figure shows that the time localization could improve the performance to some extent. Note that for SE10m, the 10 min data batches already include two full volume scans of radar data. Although these data are asynchronous they are all considered as valid at same time, which could cause fairly large observation timing errors, especially in this fast-moving storm case. The time localization could suppress the harmful influence of the data that have large timing errors, since those data are usually not close to the analysis time; time localization could improve the analysis. For the 3DLETKF with a longer time window, time localization works in a way that is similar to simply discarding observations that are not close to the analysis time. Wang et al. (2013a) point out that the inclusion of additional data with larger timing error can actually harm the analysis of regular EnSRF with longer data batches, such as 10 min.

Time localization is then applied to 4DLETKF with 10 min data batches for the OSSE of the FML (fast-moving storm) case. Figure 3.11 shows the RMSEs of 4D10mw, 4D10mw_timeloc600 (cut-off radius for time localization is 10min) and 4D5mw_timeloc300 (cut-off radius for time localization is 5min.). The result is interesting. In most analysis cycles, the performance from analyses localized by a 10min cut-off radius is better than performance without time localization. On the other hand, the performance from analyses localized by a 5min cut-off radius is worse than 4DLETKF without time localization. This indicates that the time localization could improve the analysis, but could also degrade the analysis. Its effect depends on the cut-off radius. Finding an optimal value for the time cut-off radius could be challenging. Even if found, the flow-dependent or case-dependent temporal covariance structure might make the radius useless. Applying an adaptive localization scheme in our 4D-LETKF system may be appropriate, especially when considering time localization.

3.4.4 Effect of Adaptive Localization

The hierarchical filter of Anderson (2007) was applied as the adaptive localization scheme into our 4DLETKF system. Figure 3.12 shows the RMSEs of 4DLETKF with 5min data batches in the experiments listed in Table 3.2; these are for the slow-moving storm case. The performance with pure adaptive localization (AL) is better than the performance with only data selection but without R-localization applied to the selected observation. This indicates that HF could effectively remove or suppress the spurious correlation and improve the analysis performance. Also note that in the early cycles the performance of AL is not as good as the regulated R-localization (NonAL). The performance of AL, however, is eventually improved through the analysis cycles. In the final cycles, its performance is comparable to the performance of NonAL. In this case, the pure Adaptive Localization is not as efficient as the Non-adaptive localization. One possible reason for this phenomenon is that the hierarchical filter is a type of statistical resampling method. Statistics in resampling methods are dependent upon the quality of the original sample. Thus, in the early cycles, the covariance structure is poor among the ensembles, the results from HF are also poor, and HF is not as efficient as the non-adaptive localization in the early cycles.

As mentioned in section 3.2, the RCF, localization weight, and estimated in HF are designed for EnKF and its variants using B-localization. This localization weight is regulated in the same way as the R-localization in Chapter 2, and is called ‘RAL”. Figure 3.12 shows that with the regulated localization weight, the performance of RAL is better than the unregulated AL, as expected. Furthermore, the hybrid regulated non-adaptive and regulated adaptive localization result in the best performance.

Finally, the adaptive localization is tested with a longer window (10min) for the fast-moving storm case. Figure 3.13 shows the RMSEs of the experiments with different localization schemes. The performance of the hybrid localization method is the best, and its advantage being most significant in the final cycles. The performance of pure adaptive localization is as good as the non-adaptive localization in the early cycles, but becomes poor in the final cycles. The reason is unknown yet.

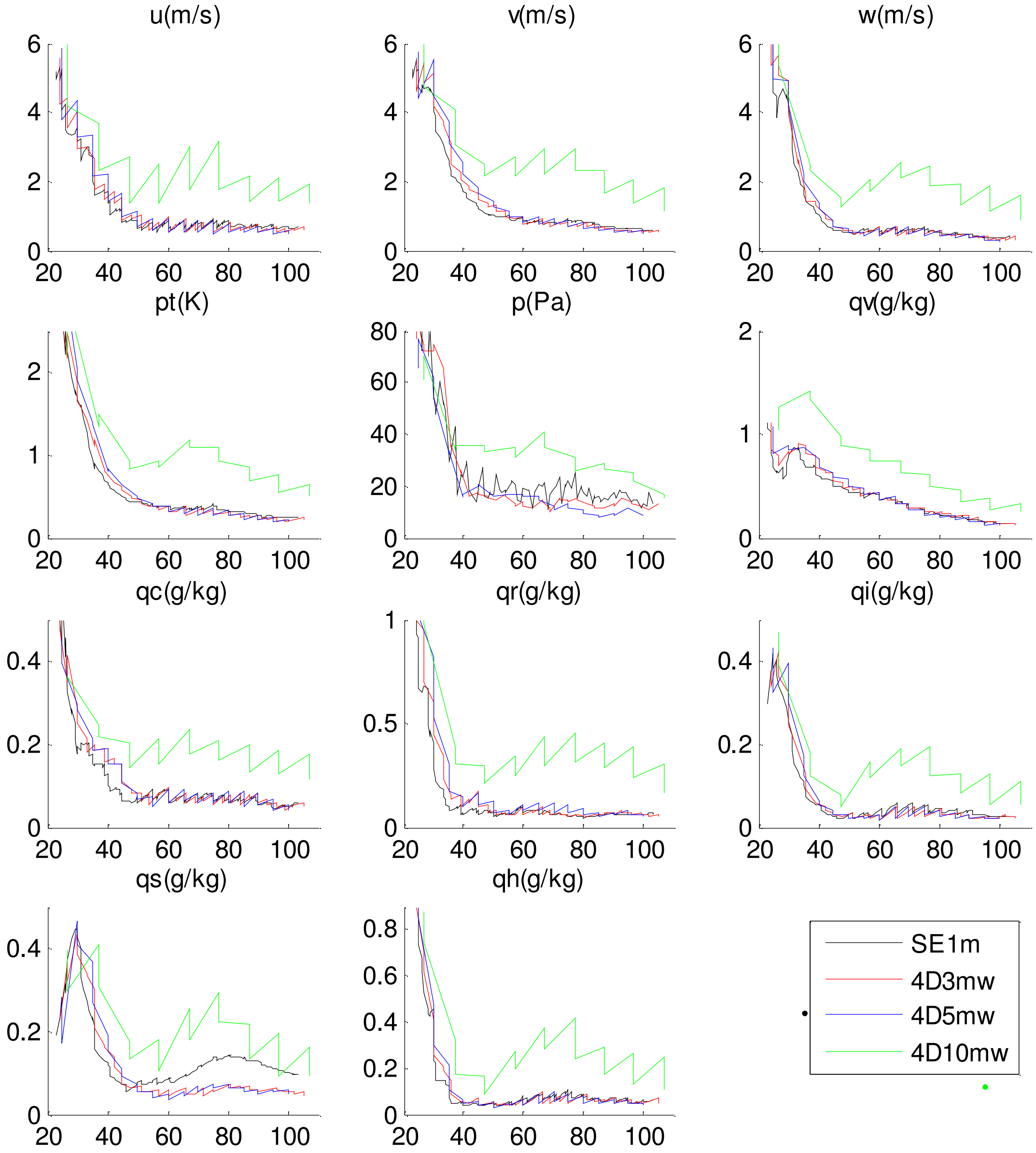

Figure 3. 4 RMSE of SE1m, 4D3mw, 4D5mw, and 4D10mw in FML (fast-moving storm OSSE)

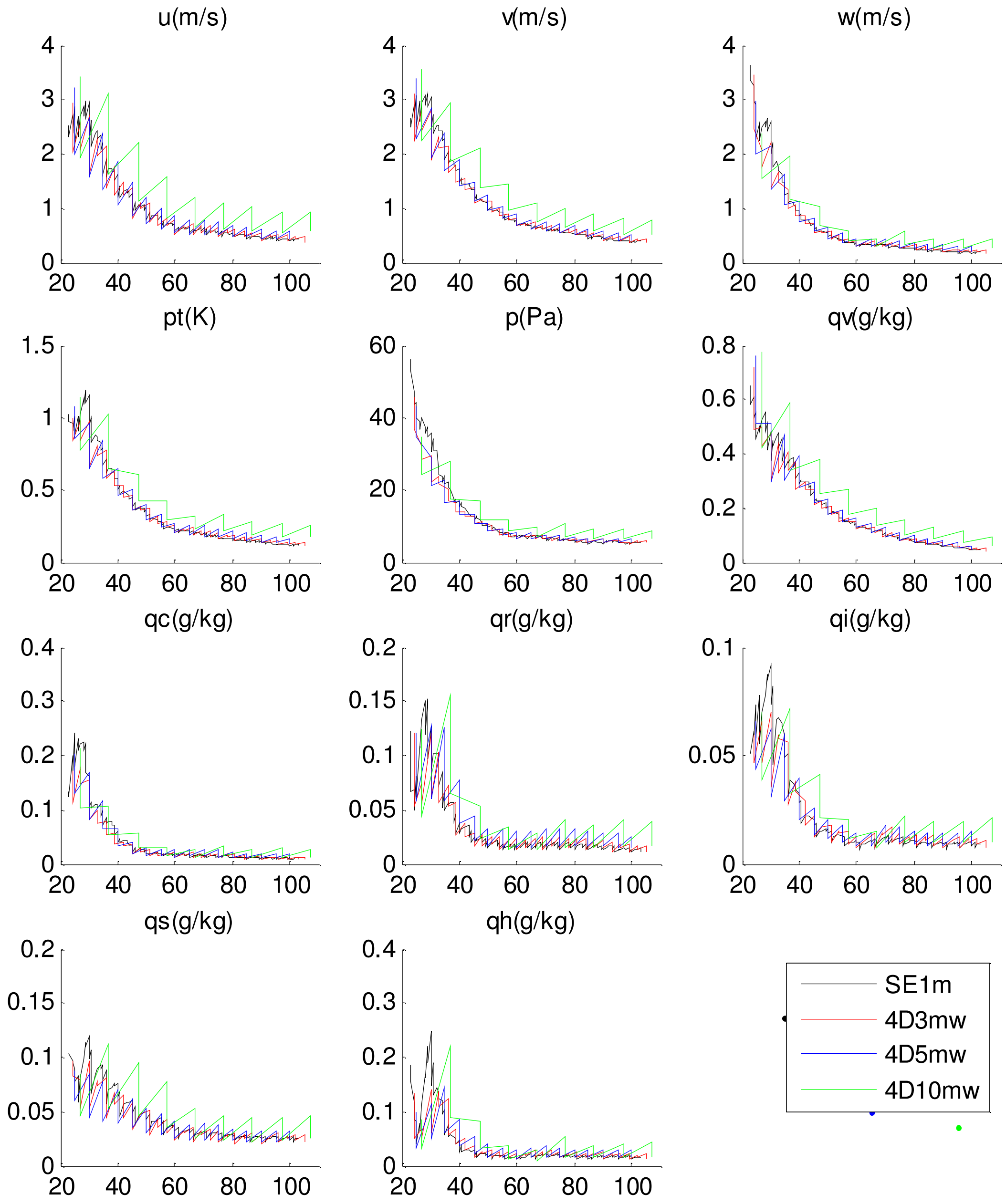

Figure 3. 5 Ensemble spread of SE1m, 4D3mw, 4D5mw, and 4D10mw in FML

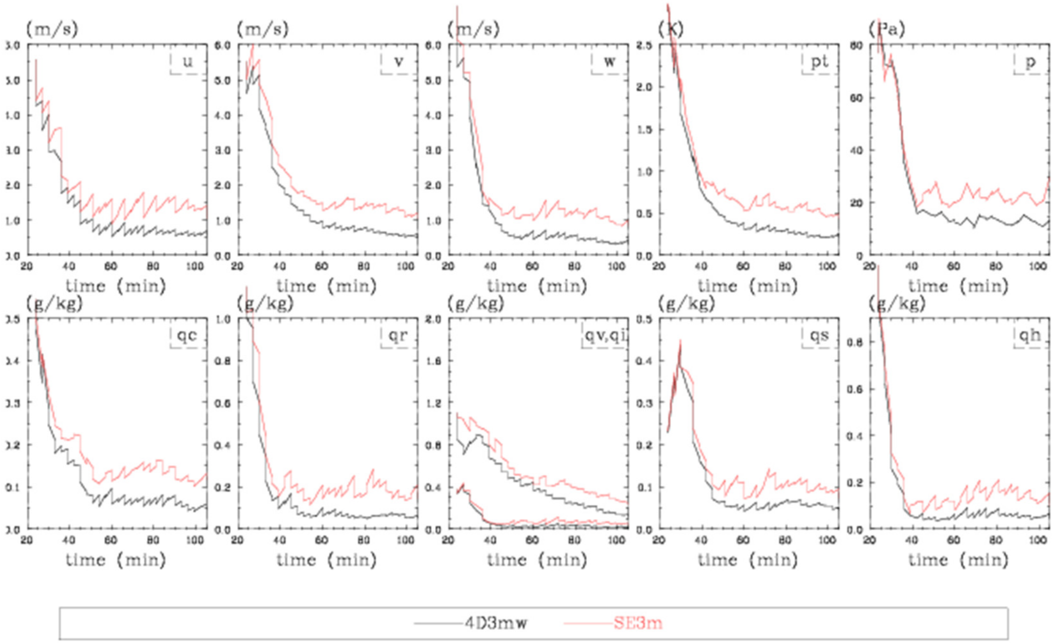

Figure 3. 6 Ensemble spreads of 4D3mw and SE3m in FML (fast-moving storm OSSE)

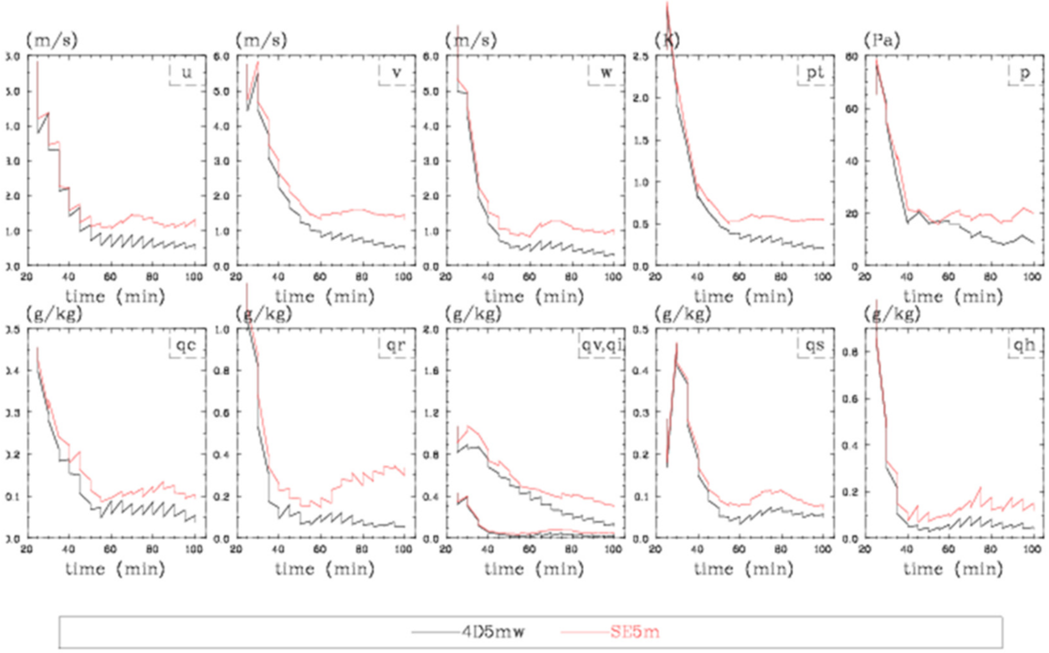

Figure 3. 7 RMSEs of 4D5mw and SE5m in FML (fast-moving storm OSSE)

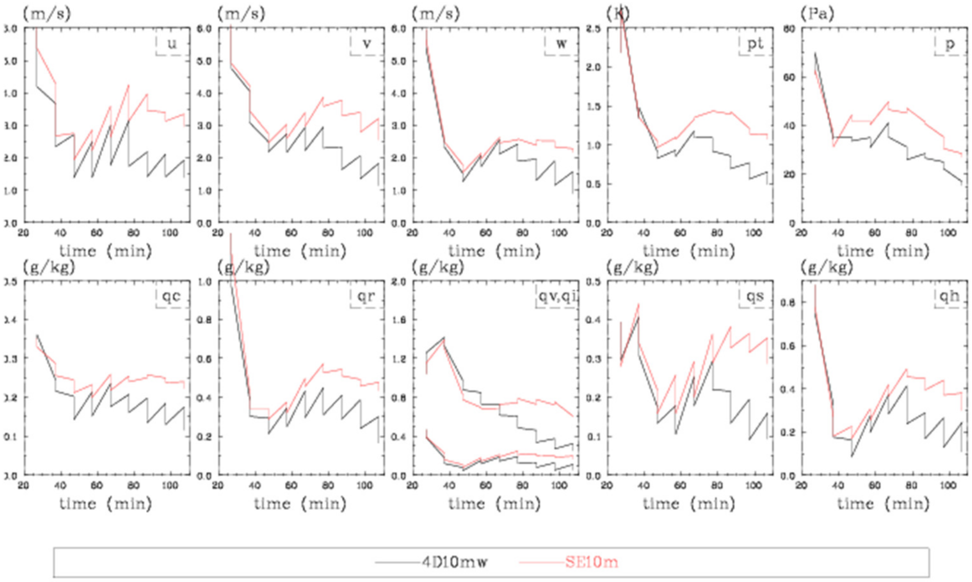

Figure 3. 8 RMSEs of 4D5mw and SE5m in SMS (slowt-moving storm OSSE)

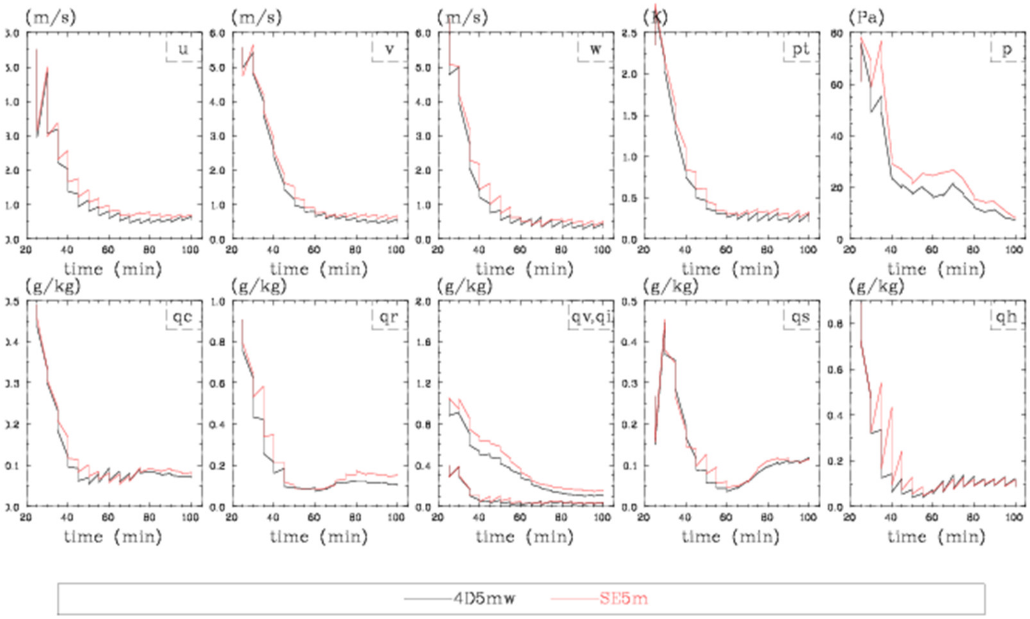

Figure 3. 9 RMSEs of 4D5mw and SE5m in SMS (slow-moving storm OSSE)

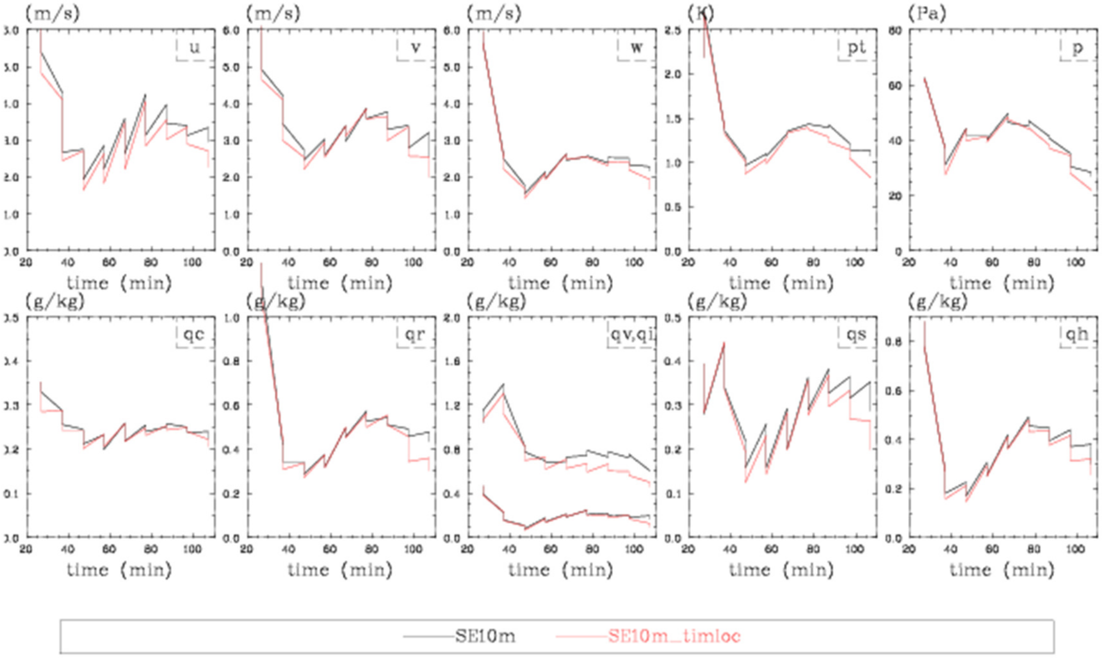

Figure 3. 10 RMSE of SE10m and SE10m with time localization in FML (cut-off radius for time localization is 10 min)

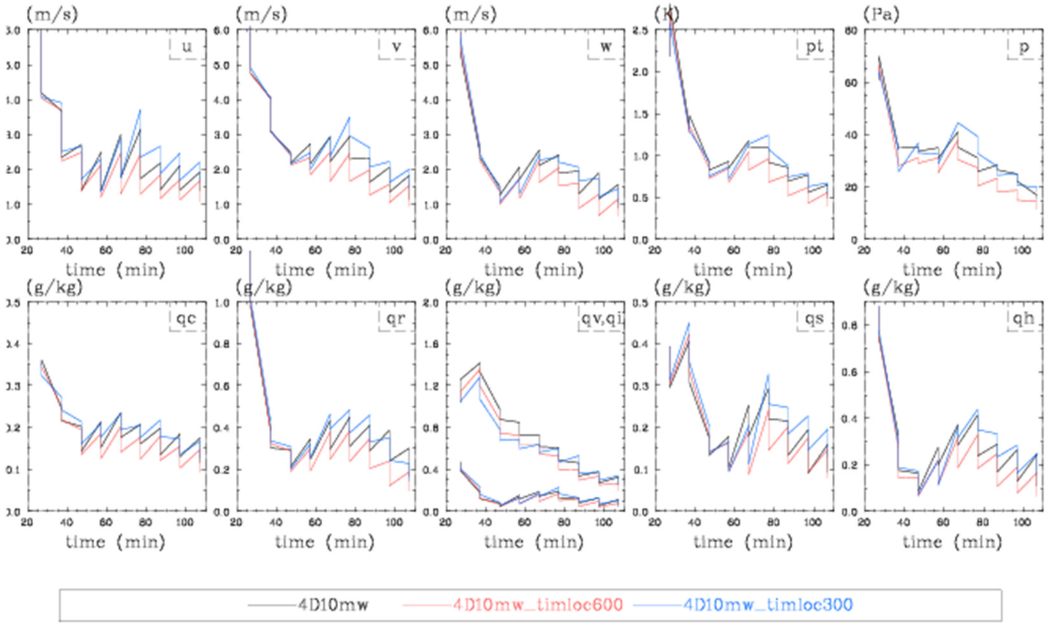

Figure 3. 11 RMSE of 4D10mw with no time location, 4D10mw with a cut-off radius of 10 min (4D10mw_timeloc600), and 4D5mw with a cut-off radius of 5 min (4D5mw_timeloc300).

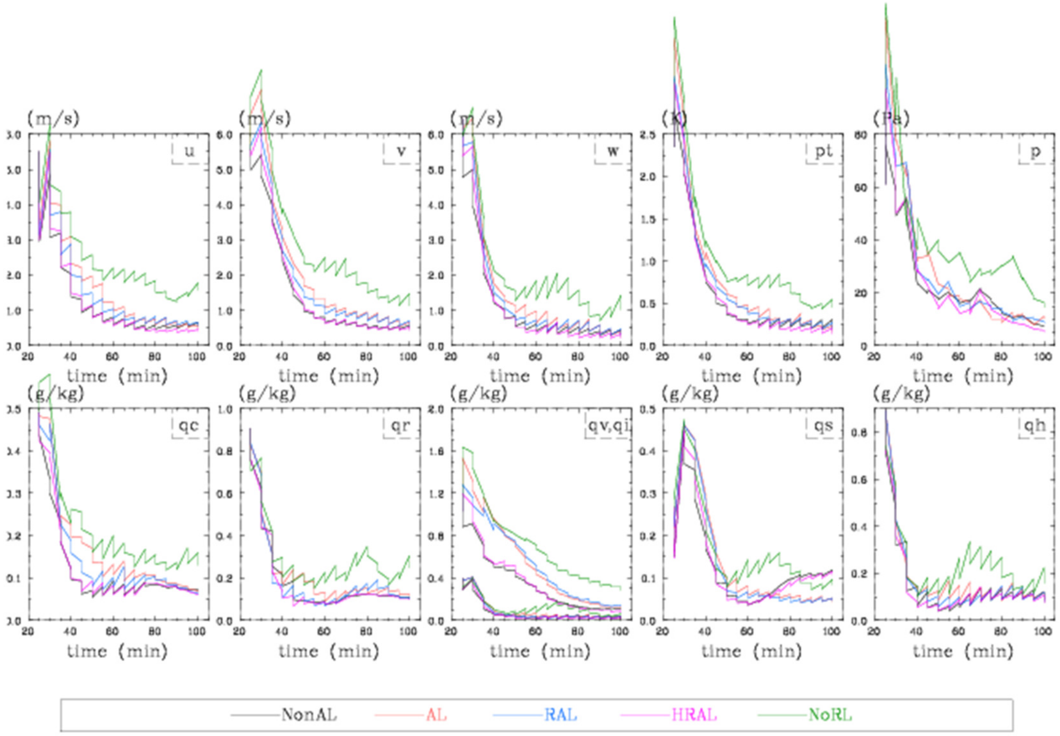

Figure 3. 12 RMSE of 4D5mw of SMS without adaptive localization (black curve), 4D5mw with adaptive localization (AL, in red), 4D5mw with regulated adaptive localization (RAL, in blue), 4D5mw with hybrid localization (HRAL, purple), and 4d5mw without localization (NoRL, green) for SMS OSSE.

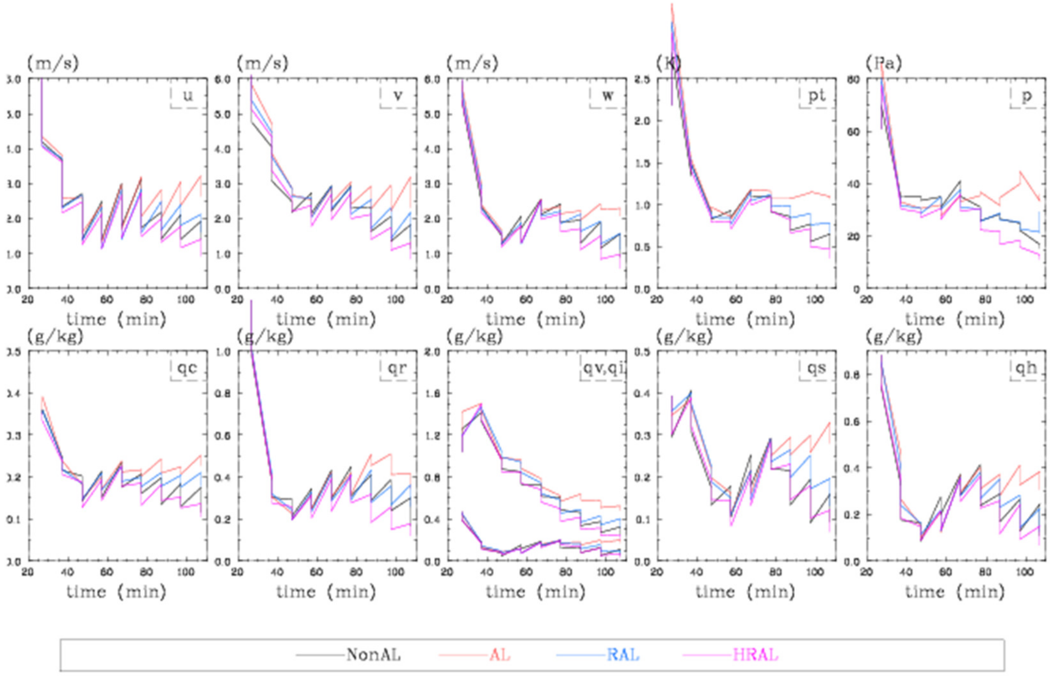

Figure 3. 13 RMSE of 4D10mw without adaptive localization (NonAL, black), 4D10mw with adaptive localization (AL, red), 4D10mw with regulated adaptive localization (RAL, blue) and 4D10mw with hybrid localization, for FML OSSE**.**

3.5 Summary and Conclusions

In this work, the newly developed ARPS-LETKF system was extended to its 4D version. With this 4D-LETKF algorithm, high frequency observations (such as radar observations) distributed over a time window could be analyzed simultaneously without timing error. This 4D-LETKF system is assessed by analyzing simulated radar observations in two OSSEs, with a fast-moving and a slow-moving supercell thunderstorm, respectively. Its performance was compared with the regular 3D-LETKF with different time window lengths.

The results indicate that, compared to the regular 3D-LETKF (which takes the asynchronous data in a time window assigns it the same analysis time), 4D-LETKF shows better performance. This was mainly due to observation timing errors in the former method. Compared to the 3D-LETKF, which utilizes the observations with very short time windows, 4D-LETKF could reach the same performance after a few DA cycles and save computational cost by fewer analysis cycles.

The effect of time localization was also inspected with long data batches. Time localizations improved the performance of 3DLETKF with long data batches by suppressing the harmful influence of data with large timing errors. The time localization also helped improve the performance for 4DLETKF with long data batches. Finding the optimal cut-off radius for time localization, however, could be difficult. The hierarchical filter (HF) was adopted as an adaptive localization scheme in this 4D-LETKF system for the sake of the flow-dependent covariance, especially the temporal covariance. The hybrid localization of HF and regular non-adaptive localization improved the performance of 4D-LETKF, especially with the longer time window.

Chapter 4: The Constrained LETKF

4.1 Background and Motivations

Modern atmospheric data assimilation is a sophisticated procedure to combine all the available information, such as observations and model forecast, as well as the error information of the data, to produce the best estimate of the model state, which is also called analysis. Because this analysis would be provided to the model forecast as the initial condition, the analysis is not only required to fit the observations well, but also required to fulfill the physical constraints used in forecast model as much as possible. Dynamical balance is one important constraint among those physical constraints. If the analysis variables are in poor balance, then the imbalance in the initial condition would cause unreal and computational oscillations in the beginning of the forecast and the model needs some time to adjust to the balance (a.k.a. spin-up). The unreal computational oscillations contaminate the forecast state in the spin-up period. Sometimes they even could ruin the forecast at final in case of serious imbalance. So the balance is also one of the essential features of analysis. Besides the constraint of dynamical balance, there are the other constraints, such as non-negative mixing ratio, no super-saturation mixing ratio. To some extent, the data assimilation problem in NWP is an estimation problem constrained explicitly or implicitly by the dynamical and physical laws and limitations.

In the framework of EnKF, the balance constraint is represented in the background covariance. In theory, since the ensemble background fields are the solutions of forecast model, they should be in dynamical balance, and if the ensemble size is ideally large enough, then the background covariance sampled from this ideal ensemble, should be optimal and could represent the balance constraint perfectly. From the point view of ETKF algorithm, the analysis ensemble are the linear combination of background ensemble, so analysis ensemble should be in balance since background ensemble are in balance. However, in practice, the background covariance is not optimal and the balance constraint is correspondingly degraded. First reason is that, the forecast ensemble might not be well-balanced in cases that, e.g., the ensemble forecast begins from a poor-balanced initial ensemble so that forecast ensemble is still in the spin-up stage. Secondly, due to the limit ensemble size in real NWP problem, the sample covariance could not represent the balance completely and clearly, the balance implied in the sample covariance might be distorted by the sampling noise. Furthermore, the localization could deteriorate the balance since it changes the covariance structure (Kepert 2009; Greybush et al. 2011). Besides the diminishment in the balance constraint, there might be some additional information or limitations about the forecast model that the EnKF does not incorporate. For example, the model error, non-linearity in the model and observation operator and the non-Gaussian PDF could downgrade the EnKF and lead to violation of the other constraints, e.g., the negative mixing ratio in the analysis sometimes. Then in practice, the dynamical balance and other physical constraints in analysis required by the forecast model could not be assured by EnKF. So the degradation and/or violation of the constraints in EnKF analysis motivate us to find the methods to modify the current EnKF framework to enforce the broken constraints or exploit the additional constraints and improve the filtering performance (Simon 2010).

Compared to KF and EnKF, the Variational Data assimilation problem is a constrained minimization of the cost function. In its nature it is straightforward and easy to add a new constraint term into the cost function of VarDA (Gao et al. 1999; Gao et al. 2004; Hu et al. 2006; Ge et al. 2012). And this is one of the advantages of Variational Data Assimilation (Lorenc 2003a;2003b; Gustafsson 2007). Although Kalman filter (KF) was developed originally from the unconstrained optimal state estimation problem, there are many research works to re-think about the optimal estimate subject to the additional constraints. Quite a few different constrained Kalman filtering approaches have been developed in the engineering applications.

Simon (2010) made a comprehensive survey on the Kalman filtering with state constraints. It covers linear and nonlinear systems, linear and nonlinear constraints, and equality and inequality constraints. For linear constraints, there are various ways including model reduction, perfect measurements, estimate projection, gain projection, probability density function (PDF) truncation and system projection. Under certain conditions, all these approaches result in the same state estimation. In these approaches, perfect measurements and estimation projection are more frequently applied in different area. In the approach of the perfect measurements, state equality constraints can be treated as perfect measurements with zero measurement noise, i.e. the observation error is 0. So its analysis equation is same as unconstrained KF, but the observation data space is augmented with the error-free constraints and the observation operators are also augmented. One disadvantage of this approach is its singular observation error covariance which might cause numerical problems such as ill-conditioning in the error covariance estimate. But in this method the equality constraints could be assimilated with regular observations simultaneously, so computationally it is more efficient. Phale and Oliver (2011) extended this method into the classical EnKF framework with an inequality constraint for the DA problem with oil reservoir model. Another approach, the estimation projection, is to project the unconstrained estimate of Kalman filtering onto the constraint surface. This method has no issue of singular error covariance and can be easily extended to the inequality constraints. So the estimate projection approach is a general form a class of the constrained Kalman filters. Basically this method is a two-step method which analyzes the observations in step 1and then assimilates the constraints in step2. This method was extended into constrained EnKF system for state estimation of hydrological models by Wang et al. (2009), and for estimating the terrestrial water budget by Pan and Wood (2006).

However, all these approaches above are suitable for the "hard constraints", which are the constraints that are required to be satisfied exactly and rigorously. In practice, there also are the cases that constraints are heuristic, or the constraint equations have some uncertainty. So these constraints are only required to be approximately satisfied and are called "soft constraint", also called as "weak constraint". One of the ways to apply the weak constraints in KF is to modify the perfect measurements by adding small non-zero observation errors, which will be called "imperfect measurements" approach in this study. Johns and Mandel (2008) proposed a two-stage EnKF that imposes the smoothness as a penalized constraint. In the second stage, the smoothing step consists of another application of the same EnKF code with the smoothness constraint as an independent observation with the specified error. In this 2-stage EnKF the constraint is used as the observation with noise, so it is essentially an extension of the “imperfect measurements” approach to the EnKF. In next session, it will be showed that this “imperfect measurements” approach is essentially similar to the approach of introducing additional constraint term into the cost function in 3DVar DA system.

What we are interested is the EnKF subject to the additional constraints within the context of NWP. There are a few approaches and works in the atmospheric ensemble data assimilation area. For example, for the inequality constraint like non-negative mixing ratio, i.e. $q_{cloud,rain,snow,ice,hail,etc.} \geq 0$, setting the updated $q = 0$ directly if this constraint is violated. This method is simple and does not consider the corresponding influence on the other adjacent variables. So the balance might be broken. This method is essentially connected to the PDF truncation approach. In Wang et al. (2009), this approach is called “naïve method”.

Considering the imbalance in ensemble background covariance, Liu (2005) proposed a full constrained EnKF and a semi-constrained EnKF to incorporate the dynamical balance equation into the covariance by separation of analysis variables into balanced and unbalanced parts. This approach had been successfully used in 3DVar community for large scale and even some meso-scale DA problem. But it is not suitable for convective-scale. Because there is no dominant and definite balance equation either in analytical form or in statistical form found for convective-scale yet, so it is not easy to split the variables into balanced and unbalanced parts.

Compared to the incorporation of constraints into the ensemble covariance of EnKF (Liu 2005), the perfect/imperfect measurements and the estimate projection are easy and flexible, could deal with linear and nonlinear constraints, equality and nonequality constraints. And these approaches had been extended to EnKF framework in area of hydrological models (Pan and Wood 2006; Wang et al. 2009), oil reservoir model (Phale and Oliver 2011) and even a fire model (Johns and Mandel 2008). It is worth to try this method for the convective-scale EnKF problem.

In this study, our goals are to extend the imperfect measurements approach to the newly-developed ARPS-LETKF system to set up a constrained-LETKF system, test this constrained-LETKF system in the convective-scale DA cycling experiment, and gain the understanding of the feasibility and capability of constrained-LETKF in convective scale DA problem. In this work, the divergence equation and pressure tendency equation are incorporated as “soft constraints” into the ARPS-LETKF framework. Why is divergence equation chosen in this work? First, the divergence equation is an equity constraint. The inequity constraint, e.g., the non-negative mixing ratio, $q_{cloud,rain,snow,ice,hail,etc.} \geq 0$,would be applied only for the member which violates the inequity constraint. Not every member would be analyzed simultaneously with the inequity constraint. So the inequity constraint might be easy for the classical EnKF, which analyzes each member individually. For the equity constraint, it is used by all the members simultaneously so that it is easy to be introduced into the deterministic EnKF algorithm, such as EnSRF and LETKF. The second reason is the divergence equation has been already successfully used as a weak constraint in ARPS-3DVar framework for the convective-scale DA problem (Gao et al. 1999; Gao et al. 2004; Hu et al. 2006; Ge et al. 2012). Their works showed that the divergence constraint could improve the balance in the field and help 3DVar DA system get better performance. Kalnay et al.

(2007a) and Kalnay et al. (2007b) also suggested that the divergence could be tested as a constraint first in the EnKF framework.

4.2 ARPS-LETKF constrained by divergence equation and pressure tendency equation

By using the imperfect measurements approach, the divergence equation and pressure tendency equation are treated as the special types of observation to build up a weak-constrained LETKF system. It is similar to the two-stage EnKF with a weak constraint for smoothness (Johns and Mandel 2008). But the difference is that the weal constraints are assimilated with the other regular observations simultaneously in our LETKF, not sequentially in 2 steps.

Since the divergence equation had been used as a weak constraint successfully in ARPS-3DVar, so at first let us look at the ARPS-3DVar with divergence constraint. The following formula is the cost function used in ARPS-3DVar (Gao et al. 2004).

$$\begin{aligned} J(x) &= J_B + J_O + J_C \\ &= \frac{1}{2}(x-x_b)^T B^{-1}(x-x_b) + \frac{1}{2}\left[H(x)-y_o\right]^T R^{-1}\left[H(x)-y_o\right] + \frac{1}{2}\lambda_c D^2 \end{aligned} \tag{4.1}$$

where J_O is observation term, J_B is the background term, and J_C is the divergence constraint term. The weight coefficient, λ_C, controls the relative importance of this constraint term in the cost function. They all are in quadratic form. J_C could be re-written as the following form

$$J_C = \frac{1}{2}(D-0)^T \left(\lambda_c^{-1}\right)^{-1} (D-0) \tag{4.2}$$

Comparing the observation term to the above re-written constraint term, the constraint term could be taken as a special type of observation term, in which the observation is divergence at model grid point and the observation value is always 0. The constraint weight, λ_C, could be viewed as the inverse of error variance of this special observation.

In the sense of taking the divergence constraint as a special observation, the weak-constrained LETKF shares the same spirit as the 3DVar with divergence constraint. However, it should be noticed that there are some differences between the 3DVar and LETKF with weak constraints. In the constrained LETKF scheme, the constraints used as special type of observation are selected locally for the local analysis at each grid point. Furthermore the influence of constraint is tampered with R-localization. So the constraints make local impact, just like the regular observations used in LETKF. This is different to the constraints applied in ARPS-3DVar, where those constraints work globally. Another difference is that in LETKF, the constraint could have impact on the analysis mean as well as on the analysis covariance structure. And the impact on covariance could spread though the ensemble forecast. In 3DVar, the constraint only has influence on the analysis. The background covariance would not be changed. This difference could make considerable differences through the data assimilation cycle experiment.

Since the divergence is treated as the same “observations” as the regular observations, so the analysis equations in this constrained LETKF scheme are almost same as in regular LETKF scheme. One difference to the unconstrained LETKF is the augmentation of total observation space due to this “new special observation". Another difference is the new forward operator for this new observation.

Table 4. 1 Differences of LETKF and 3DVar both with weak constraints

	LETKF	3DVar
Localization	Due to data selection and R-localization, the impact of constraint is local.	No localization. The impact of constraint is global.
Background Covariance	Impact from constraint on ensemble mean as well as on the ensemble background covariance	No impact from the constraint on the background covariance. Only impact on estimate of state

Now let us look at the forward observation operator of divergence, which is just the diagnostic equation of divergence used in ARPS (ARPS User Guide, Xue et al. (1995)).

$$D = Div(x) = \frac{1}{\sqrt{G}}\left[\frac{\partial u^*}{\partial \xi} + \frac{\partial v^*}{\partial \eta} + \frac{\partial W^{c^*}}{\partial \zeta}\right] \qquad (4.3)$$

Now, let us take a look at the divergence observation by applying this divergence observation operator to the model fields.

Figure 4.1 shows the divergence of background and analysis from an unconstrained ARPS-LETKF DA system. In Figure 4.1 panel (a) shows the vertical profiles of the divergence at grid point (grid index i=16, j =16) for true field (blue curve), ensemble forecast mean (green curve) and ensemble analysis mean (red curve) at t=25mins. The analysis is unconstrained LETKF analysis with assimilation of the radar observations; Panel (b) is similar as (a), but shows the vertical profiles of the horizontal average of the square of divergence; Panel (c) is same as (a) but for each ensemble members; and panel (d) is same as (b) but for each ensemble members. Note that the vertical coordinate is the index of vertical levels, not the height. The larger is the index number, the higher is the level.

In Figure 4.1, it could be found that the divergences in the background and analysis are small. And apparently the divergence of background field is much closer to 0 and is about 2~3 orders of magnitude smaller than the divergence of analysis field. Since the background is the model forecast, it should be well-balanced in model dynamics. The balance in analysis field should be broken because of the assimilation of the new observations. So the Figure 4.1 told us that the divergence is very close to 0 in balanced field. And the knowledge of atmospheric dynamics told us that the divergence in the well-balanced field should be very small and very close to 0. This is the reason why the divergence equation is used as constraint of balance. It is also noticed that even in the balanced field the divergence is really tiny value, but not as same as 0. So divergence constraint is not an accurate constraint and is not strictly fulfilled by the real NWP model. And very small divergence is the necessary condition for balance, not the sufficient condition. So the divergence constraint is a weak constraint. And actually this divergence constraint could impose a weak anelastic mass continuity constraint in the analyzed wind field (Gao et al. 2004). So Figure 4.1does indicate that divergence with value of 0 could be used as an imperfect observation into the EnKF algorithm.

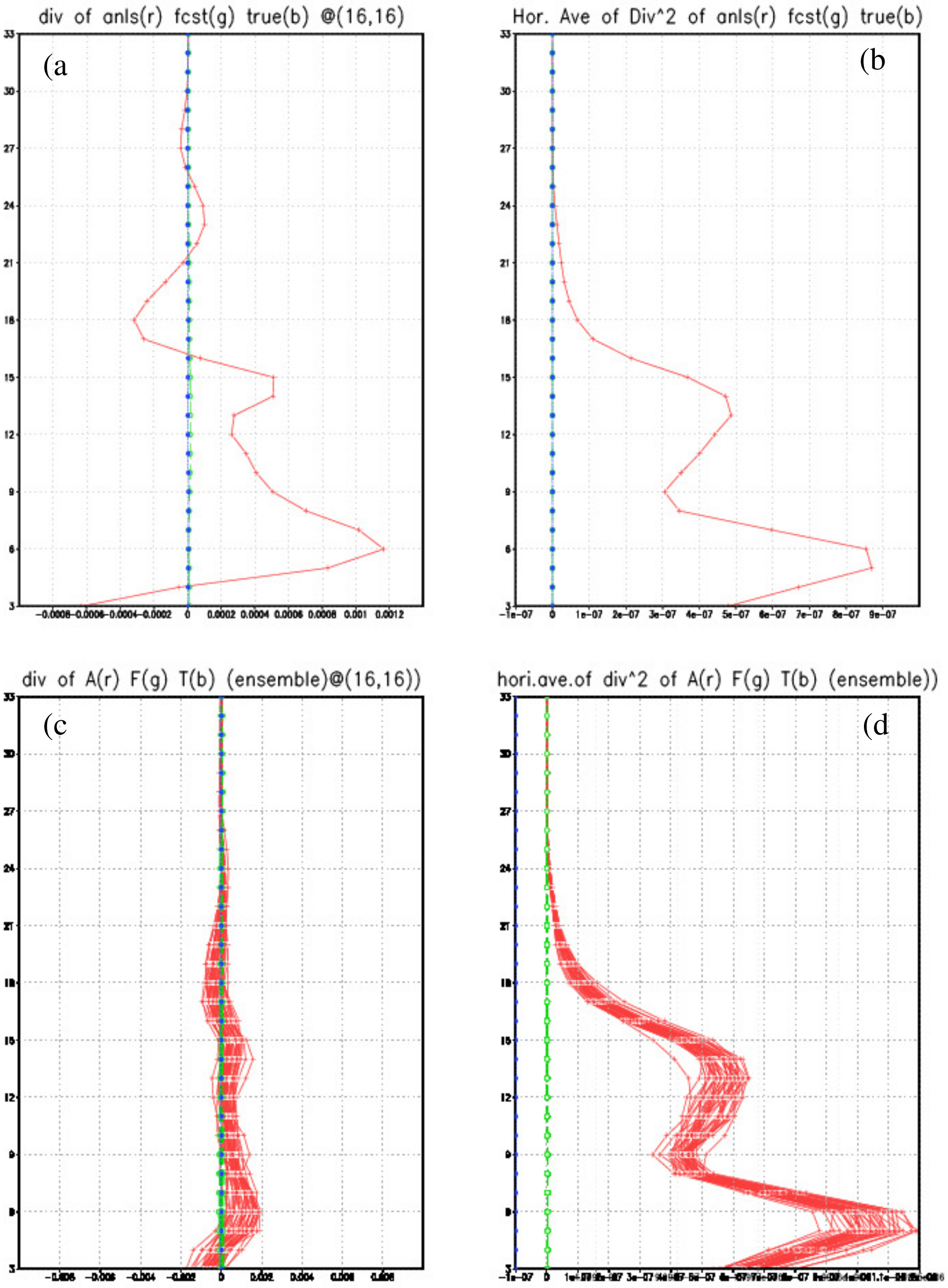

Figure 4. 1 Divergence in true (blue), forecast (green) and analysis (red) fields (the analysis is from unconstrained LETKF analysis with radar observations at t=25min)

The pressure tendency constraint is used in the same way as the divergence tendency in constrained LETKF scheme. It is also used as the observation of pressure tendency at model grid points and the observation value is 0. The forward observation operator of pressure tendency is just the diagnostic equation of pressure tendency used in ARPS (ARPS User Guide, Xue et al. (1995)), which is as following

$$\begin{aligned} p-tendency = \frac{\partial}{\partial t}(J_3 p') = &-\left[(J_3 u)\frac{\partial p'}{\partial \xi} + (J_3 v)\frac{\partial p'}{\partial \eta} + (J_3 W^c)\frac{\partial p'}{\partial \zeta}\right] + J_3 \bar{\rho} g w \\ &-\bar{\rho} c_s^2 \left[\frac{\partial (J_3 u)}{\partial \xi} + \frac{\partial (J_3 v)}{\partial \eta} + \frac{\partial (J_3 W^c)}{\partial \zeta}\right] \\ &+ J_3 \bar{\rho} c_s^2 \left[\frac{1}{\theta}\frac{d\theta}{dt} - \frac{1}{E}\frac{dE}{dt}\right] \end{aligned} \qquad (4.4)$$

where $\bar{\rho}$ is the mean air density at a given horizontal level, $E \equiv 1 + 0.61 q_v + q_{liquid+ice}$. Note that the divergence and p-tendency are both "observed" and defined on scalar grid points, even the kinetic fields are defined on the staggered grid points.

In the pressure tendency equation RHS terms are the advection of perturbation pressure p' (which is non-linear term), the advection of base-state pressure $\bar{p}$ (which is linear term), the divergence term (linear term) and the diabetic terms. The diabatic terms are usually small and are therefore neglected in ARPS as well as in this work. And the divergence term is usually the dominant term for most meteorological applications. So in this sense, the divergence equation is an approximation to the pressure tendency equation. Furthermore, the pressure p' is directly responsible for the mass balance in the system, and the divergence constraint is a weak anelastic mass continuity constraint in the analyzed wind field. So these two constraints both impose mass balance constraint on the analysis field. Then it is expected that these two constraints might produce the similar effect. This is one of the reasons that why we use pressure tendency constraint here along with divergence constraint. They could be used to cross-validate against each other. By the way, it is suggested that these 2 constraints should not be used in constrained LETKF simultaneously. And note that this pressure tendency equation constraint is different with the Diagnostic Pressure Equation Constraint (DPEC) used in ARPS-3DVar (Ge et al. 2012). In their work, the diabatic terms and turbulence effect are included.

The divergence equation is linear. Pressure tendency equation is dominant by divergence term, so it is weak nonlinear. Then according to the results in chapter 2, simultaneous assimilation of constraints and regular observation in 1 step, or sequential assimilation in 2 steps should have similar performances when using linear observations. This is why we did not use the 2-step EnKF scheme. For the same reason, applying these two weak constraints to EnSRF scheme and LETKF scheme also should make no significant and systematic differences. Furthermore, because the observation space is augmented by the new special types of observations, the computational cost of constrained EnSRF and constrained LETKF both increase. As EnSRF processes the observations sequentially, so the computational cost of EnSRF is more sensitive to the dimension of observation space than the computational cost of LETKF. Since the performance of EnSRF and LETKF both constrained by these two linear/quasi-linear constraints is expected to be similar, and LETKF could take advantage over EnSRF in the sense of the computational cost, so only the constrained LETKF is implemented in ARPS-EnKF system at this time.

4.3 Experiments and Results

4.3.1. Experiment of LETKF Analysis with only divergence constraint

First, an experiment of LETKF analysis with only the divergence constraint but no any other observations is made. This experiment runs only for one analysis step. The purpose is to investigate the effect of assimilating the divergence constraint as an observation. Since the divergence constraint is used as the observation with value 0 and the divergence in background is tiny but not 0, so the divergence of the final analysis should be reduced and get closer to 0 than divergence of background statistically. The variables being plotted in the 4 panels of Figure 4.2 are same as of in the panels of Figure 4.1. In Figure 4.2 panel (a), it shows that, from level 9 and above, the divergence of ensemble analysis mean (red curve) is closer to 0 than the divergence of ensemble forecast mean (green curve). But from level 8 and below, the divergence of analysis mean is larger than the divergence of forecast mean. In panel (b) and (d) the similar fact could be observed. It indicates that the divergence constraint in LETKF could reduce the divergence in middle and high levels. So this divergence constraint could make reasonable and expected effect in the LETKF framework at least in middle and higher levels. But why does it not work the same way in the lower levels? The reason is unknown yet to us now. It will be discussed in summary session of this chapter.

Note that in this experiment the regular observations (like radar observation) are not used. The reason is that the increase of divergence due to the assimilation of regular observations is much more significant than the reduction of divergence due to the assimilation of divergence constraint.

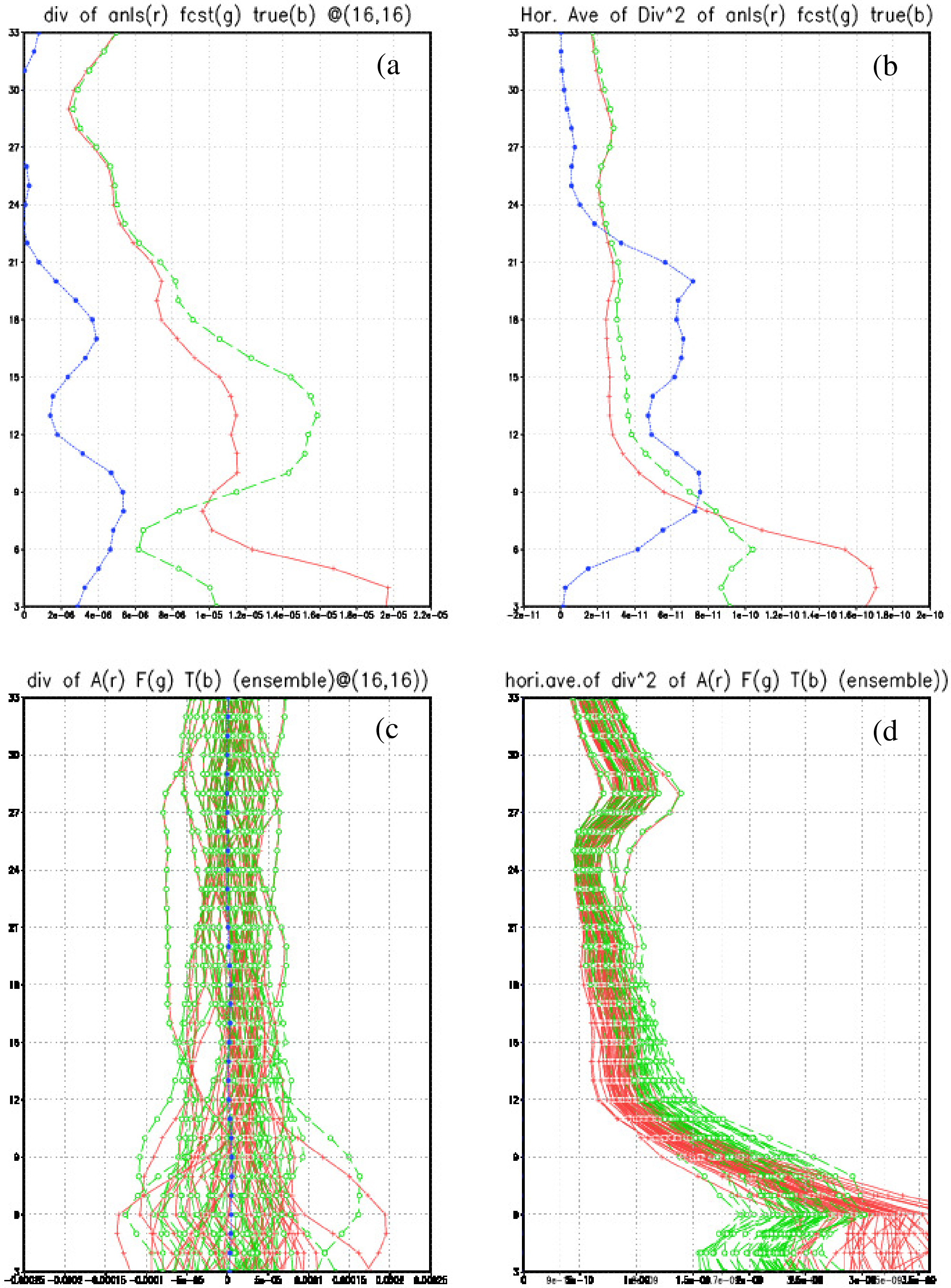

Figure 4. 2 Divergence in background (green) and analysis (red) fields , same as Figre 4.1 but Analysis is from LETKF analysis with only the divergence constraint, with error std. = 0.001kg m^{-3} s^{-1})

4.3.2. Impact of divergence constraint with various constraint weight

In the experiment above, the effect of divergence constraint on analysis is examined without analyzing the regular observations. Now, the impact of divergence constraint on the performance of ARPS-LETKF is investigated through a set of experiments of data assimilation with radar observations. In this set of experiments, the different constraint weights are used. The ARPS-LETKF system configuration, the radar observation data set are same as used in chapter 2. The constraint weight coefficients are listed in Table 4.2. The Figure 4.3 shows the RMSE of the set of experiments listed in Table 4.2. And Figure 4.4 shows the ensemble spread for the same experiment set in Figure 4.3.

Table 4. 2 Weight coefficients used in experiments with divergence constraint

Experiment Name	Constraint weight coefficient	Error stddev.
LETKF	0	infinite
Div1E-3	1E+6	1E-3
Div2E-4	2.5×1E+7	2E-4
Div1E-4	1E+8	1E-4
Div1E-5	1E+10	1E-5

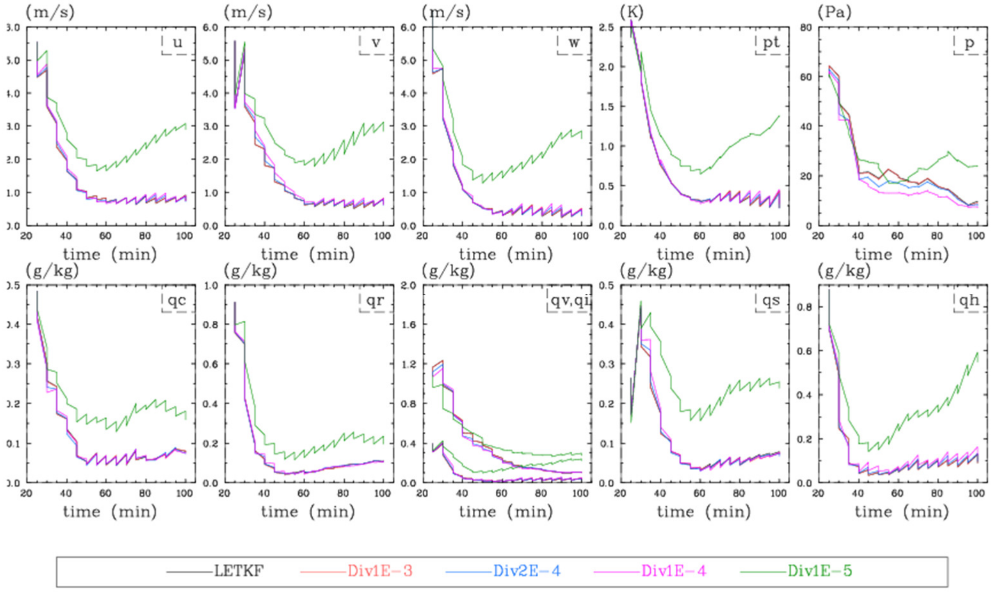

Figure 4. 3 The ensemble-mean forecast and analysis Root Mean Squared Errors (RMSE) averaged over points at which the true reflectivity is greater than 10 dBZ for: u, v, w and perturbation potential temperature θ', perturbation pressure p', cloud water qc, rainwater qr, water vapor qv (the curves with larger values), cloud ice qi (the curves with lower values), snow qs, and hail qh, for experiments Unconstrained LETKF (black), LETKF with divergence constraint err.std.=0.001(red), with divergence constraint err.std.=0.0002(blue), err.std.=0.0001(pink) and err.std.=0.00001(green).

In Figure 4.3, it shows that the impact of constraint depends on the magnitude of the weight constraint coefficient. When the weight is small (like the error stddev is 1.0e-3), the influence of constraint is very trivial. By varying the weight in the reasonable range (like the error stddev. is around 1.0E-4), the obvious positive impact of divergence constraint could be observed from the RMSE of perturbation pressure p'. Even in the beginning of the cycles, the improvement in RMSE of p' could be noticeable. After 2~3 cycles, the further reduction in RMSE of p' due to the divergence constraint becomes considerable. When the constraint weight is bigger but still in reasonable range, the improvement is more significant. For the other model variables, the influence of constraint is very small. When increasing the weight, a little bit increase in the RMSE could be seen on some variables, such as wind component . But this negative impact is still very insignificant. When weight is too big and beyond the reasonable range (like the erro stddev is 1.0e-5), filter divergence comes into existence.

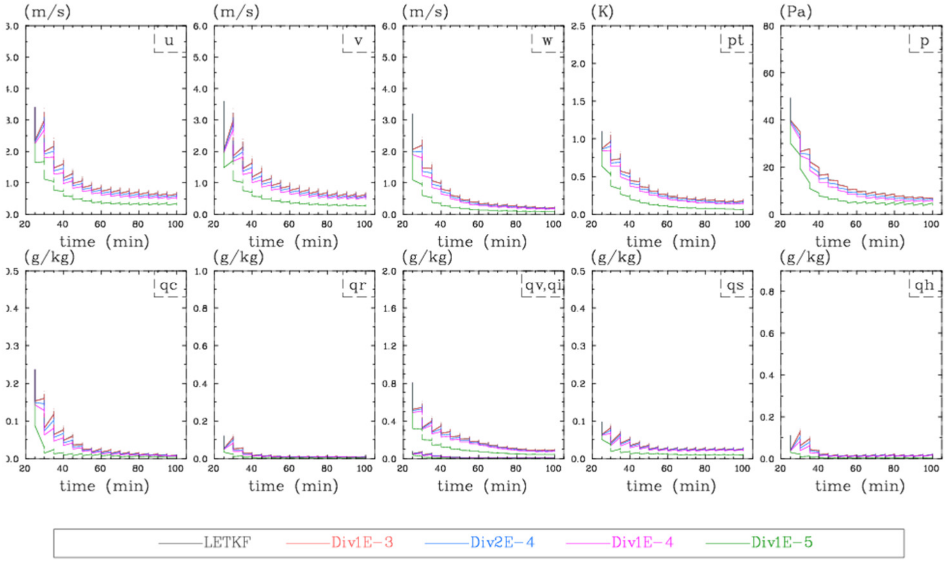

Figure 4. 4 Same as Figure 4.3 but the Ensemble Spread for experiments with various weights for divergence constraint

Looking at the ensemble spread in Figure 4.4, it could be found that the ensemble spread of the constrained LETKF is smaller than of the unconstrained LETKF. When constraint weight increases, the reduction in spread is more considerable. This might explain the slight negative impact of the constraint on other variables except for perturbation pressure p'. Especially when large weight is applied, the ensemble spread is too small to be consistent with RMSE. Then the filter diverges. In theory, more observations could improve the uncertainty in the analysis. Since the divergence constraint is used as observations, it might further reduce the ensemble spread to some extent compared to the unconstrained LETKF. In practice, the bigger covariance inflation could be applied to compensate this extra reduction of ensemble spread caused by the constraint.

Figure 4.8 shows the pressure perturbation field at t=60min on X-Z cross-section at y = 23 km from Truth simulation, LETKF analysis with divergence constraint and LETKF analysis without constraint. It could be noticed that the results of LETKF with constraint are closer to the true state. The significant differences of pressure perturbations between LETKF with and without constraint are in the lower levels, especially in the near ground levels. The values of the pressure perturbations in LETKF with constraint near the surface are much smaller and closer to the true values. Especially in the left-bottom corner area, there is an area with negative values in the true and LETKF with constraint, but in LETKF without constraint, that is an area with large positive values.

4.3.3. Impact of pressure tendency constraint with various weight

Similar as the investigation on the impact of divergence constraint discussed above, the impact of pressure tendency constraint on the performance of ARPS-LETKF is also examined through a set of experiments of data assimilation with radar observations. In this set of experiments, the different constraint weights are used. The ARPS-LETKF system configuration, the radar observation data set are same as used in chapter 2. The constraint weight coefficients are listed in Table 4.3. The Figure 4.5 shows the RMSEs of the experiment set listed in Table 4.3. And Figure 4.6 shows the ensemble spread for the same experiment set listed in Table 4.3.

Table 4. 3 Weight coefficients used in experiments with pressure tendency constraint

Experiment Name	Constraint weight coefficient	Error stddev.
LETKF (no constraint)	0	infinite
Ptnd1.0	1.0	1.0
Ptnd10.0	0.01	10.0
Ptnd50.0	0.0004	50.0

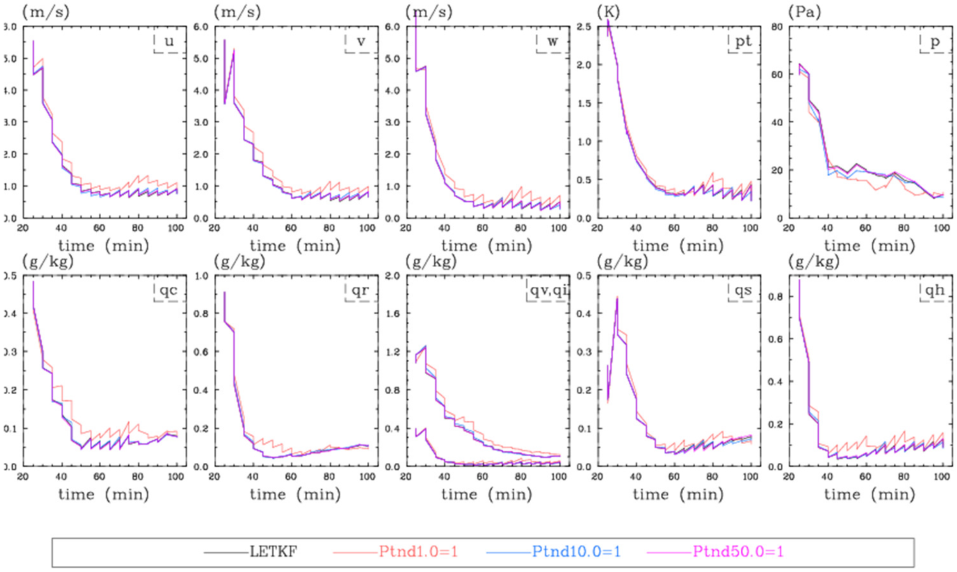

Figure 4. 5 RMSE for: u, v, w , θ', p', q_c, q_r, q_v (the curves with larger values), q_i (the curves with lower values), q_s, and q_h, for experiments Unconstrained LETKF (black), LETKF with pressure tendency constraint err.std.=1.0 Pascal s^{-1} (red), err.std.=10.0 (blue) and err.std.=50.0 (pink).

In Figure 4.5, with smaller constraint weight (like 0.0004), the impact of constraint is very tiny. When the weight is increased to 0.01, the positive impact on the RMSE of the perturbation pressure p' could be obviously observed. When using larger weight (like 1.0), the positive impact on p' turns to more significant, but a little bit of negative impact on the other model variables could be noticed also.

In Figure 4.6, the impact of constraint is to reduce the ensemble spread further compared to unconstrained LETKF. With larger constraint weight, the ensemble spread decreases more significantly.

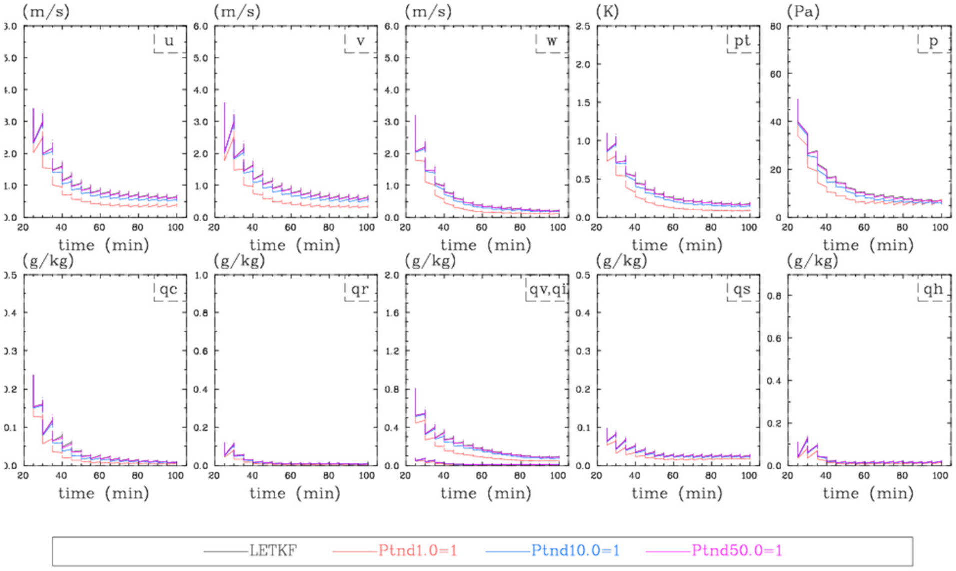

Figure 4. 6 Same as Figure 4.5 but the Ensemble Spread for experiments with various weights for pressure tendency constraint.

So the impact of pressure tendency constraint is similar to the impact of divergence constraint. Either of the two constraints always tends to reduce the ensemble spread. And too large constraint weight could lead to filter divergence. With proper constraint weight, both of the two constraints could have considerable positive impact on RMSE of the perturbation pressure p' , and slight negative impact on RMSEs of the other model variables. In Figure 4.7, the RMSE of unconstrained LETKF, LETKF constrained with divergence constraint (error stddev.=1.0e-4) and LETKF constrained with pressure tendency constraint (error stddev. =10.0) are plotted to show the similarity of these two constraints. As discussed in section 4.2, according to the equations of these two constraints, it is expected that they should behave in similar way. And the results in these experiments meet the expectations.

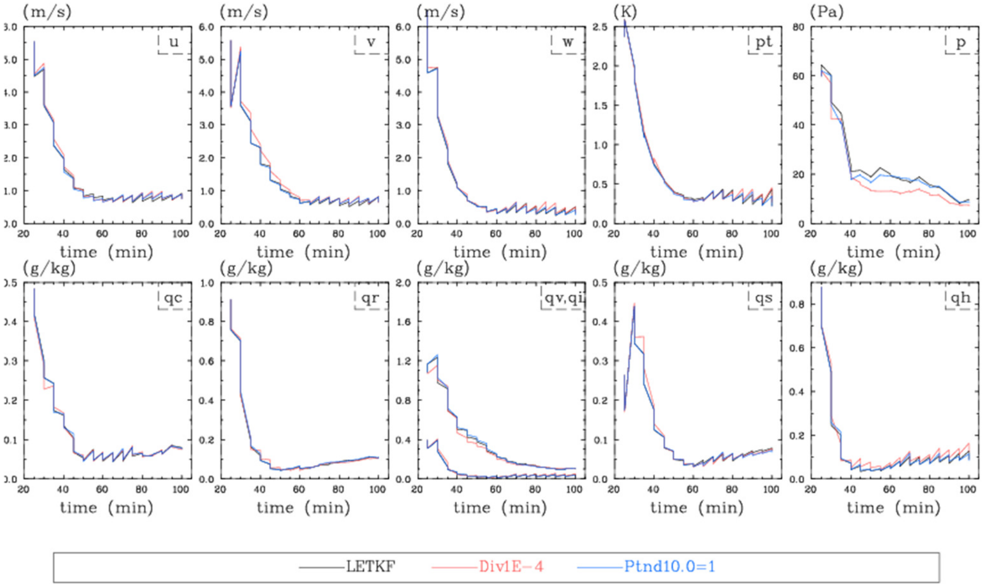

Figure 4. 7 RMSE of experiments of LETKF unconstrained (black curve), LETKF with divergence constraint err. std. = 0.0001 (red curve), and with pressure tendency constraint err.std. = 10.0 (blue curve).

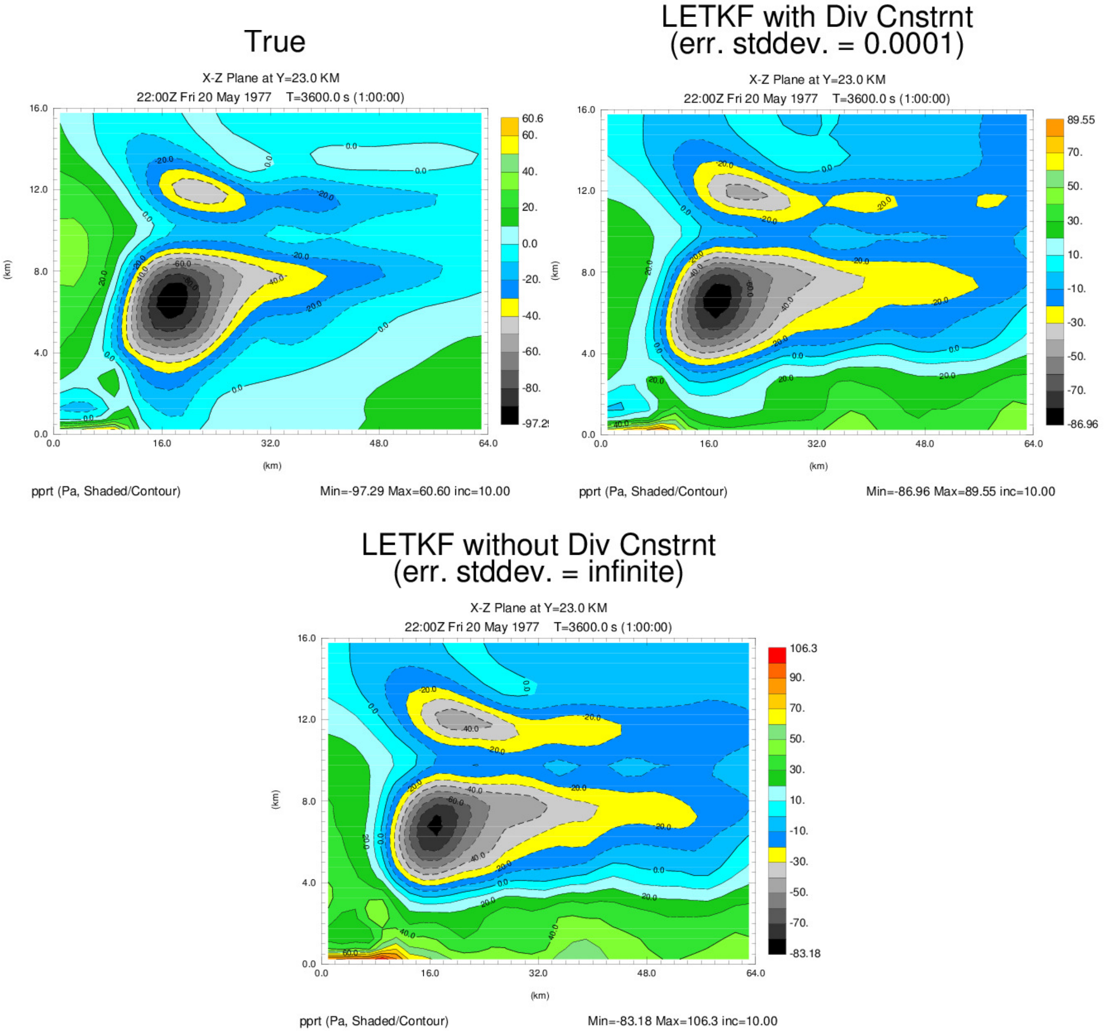

Figure 4. 8 Pressure Perturbation field (p' in unit of Pa) at t=60min on a X-Z cross-section at y=23km for True simulation (top left), LETKF with divergence constraint (top right) and LETKF without constraint (bottom)

4.4. Summary and discussions

Based on newly-developed ARPS-LETKF, a constrained ARPS-LEKTF system, that imposes the divergence equation and pressure tendency equation as weak constraints to the analysis, is set up. The two constraints, divergence and pressure tendency, are used to improve the balance in analysis field. In the constrained LETKF scheme, these two constraints are treated as special “observations” with specified error as the weighting coefficient of the constraints. The LETKF analysis algorithm does not change, except that the observation space is augmented by the observations of the divergence and pressure tendency, and the observation forward operators are also augmented by the divergence equation and pressure tendency equation.

Then this constrained ARPS-LETKF system was tested in the DA cycling experiments with radar observations in the same system configuration as in chapter 3. The results show that these two constraints have similar effects on the performance of LETKF. With proper constraint weight, both of the two constraints show two different impacts: the significant positive impact on RMSE of the perturbation pressure , and very slight negative impact on RMSEs of the other model variables. Since the perturbation pressure and these two constraints are closely related through the mass and dynamic balance, so the significant positive impact on pressure showed in our experiments is reasonable and expected. The slight negative impact might be because the constraints could reduce the ensemble spread further than the unconstrained LETKF. And too big constraint weight could lead to filter divergence due to the significant reduction in ensemble spread.

Ge et al. (2012) used both of divergence constraint and Diagnostic Pressure Equation Constraint (DPEC) in ARPS-3DVar framework. In their work the effect of the constraints could be seen in almost each analysis variable and the improvement appears from the beginning cycles of DA. In the constrained LETKF, the impacts of these 2 constraints are not significant on all the variables except for pressure p'. These different impacts of constraints from ARPS-3DVar and Constrained ARPS-LETKF might be explained as the following: the background covariance in 3DVar has a stationary but flow-independent structure. On the contrary, the background covariance in EnKF is sampled from the ensemble forecast and it is flow-dependent. For convective scale DA problem, the background covariance in EnKF is better than in 3DVar. Correspondingly the balance constraint implied in EnKF is also better. Then when additional balance constraints (such as divergence constraint) are applied to both of EnKF and 3DVar, the 3DVar could get more benefits and improvements. Especially during the DA cycle with 3DVar, the fixed and potentially not-well-balanced covariance is repeatedly used. But for the EnKF, the balance in covariance should be getting better and better through the DA cycles at least in the scenario of OSSE with perfect model.

On the other side, since the effect of constraints in constrained LETKF still could be observed significantly on perturbation pressure, and as we know, pressure is a variable sensitive to the balance in the wind field. The improvement in analysis of pressure p' indicates that the ensemble covariance in EnKF is not very well-balanced yet due to the sampling error, model error or other reasons. So these constraints still could make some improvements in the EnKF framework.

This work shows the feasibility and potential capability of the constrained-LETKF scheme in a real convective-scale NWP DA system. In the future work, we will deeply investigate the impact of these 2 constraints on the balance. For example, how is the covariance structure influenced by the constraint? This impact of constraint on background covariance is not available in the 3DVar framework. And since the divergence and pressure tendency constraints could improve the analysis of perturbation pressure, if this means positive impact on the balance in the analysis, then this analysis should be used as the initial condition to do the 3~6 hrs forecast to examine the impact of constraint on forecast. Furthermore, within this constrained LETKF framework, other types of constraints could be considered in the future, such as the non-negative mixing-ratio (q) as an inequity constraint.

There are some other issues which also need be concerned in the future work. One issue is that when only the divergence constraint is assimilated in the analysis, the divergence could be reduced in the middle and higher levels as expected, but not reduced in the lower levels. This is not reasonable. According to the estimation theory, the analysis should fit the observation statistically better than the background. Our understanding to this phenomenon is that, if this is not due to the code bug, it might be because the divergence constraint is improper in the lower levels of atmosphere. In the current model configuration, level 8 is about 2.5~3km high. So the layer from level 1 to level 8 covers the planetary boundary layer. As we know, the flow in the planetary boundary layer is strong divergent in most time. The zero-divergence constraint might not be a good balance constraint there and it is incompatible to the balance property in the ensemble covariance for the lower levels. If this zero-divergence constraint is forced to be assimilated in these lower layers, it might cause unreal and computational imbalance in the analysis field and lead to the unexpected increase of divergence in the analysis. If this understanding is right, then this zero-divergence constraint should not be used for lower level atmosphere. And the following questions would be raised: Is the flow in boundary layer totally unbalanced? Then it might be not necessary to apply any balance constraint there. Or is the boundary flow in any other kind of balance sometimes? is there any analytical equation or statistical equality which could describe it and be used as a constraint in DA problem?

Another issue is that since the constraint is treated as observation, is it necessary to consider about the possible correlation of the error of these special observations in the LETKF algorithm? Usually in VarDA framework it is assumed to be uncorrelated since it is used only as an additional constraint term to force the analysis to be more balance, not real observation. But if the error of constraint are correlated but used as if they are un-correlated, then the impact of constraint would be over-emphasized or distorted in the analysis, and may lead to the unreasonable reduction in the ensemble spread.

Chapter 5: Summary and Future Plan

5.1 General Summary

In convective-scale data assimilation (DA), radar reflectivity observations, model dynamics, and microphysical processes are strongly nonlinear. This non-linearity limits the feasibility of 4DVar for convective-scale data assimilation due to the difficulties in developing and maintaining of the full model adjoint code and the associated convergence issues with variational cost function minimization.

In the EnKF algorithm, the nonlinear model and observation forward operators could be directly used without any linearization. The EnKF is easy to implement in convective-scale DA. With the advantage of explicit estimation of the error covariance in the EnKF analysis, error covariance in EnKF is not only flow-dependent, but also continuously improved through the DA cycles. This ensemble-based error covariance can help to build up a reasonable, balanced and flow-dependent error covariance structure for a convective-scale DA system, which could not be gained in the 4DVarDA framework. So The EnKF and its variants have been widely applied into connective-scale DA with radar observations. Due to the high resolution of models and high density of observations in convective-scale DA, the large computational costs demand a massive parallel implementation of the EnKF algorithm. Compared to the sequential observation processing in the nature of the EnSRF algorithm, LETKF was designed as a parallelizing filter with its simultaneous observation processing algorithm. LETKF is more suitable to be used as a frame for future studies of large DA problems on massive parallel computers.

This work first developed the ARPS LETKF DA system and tested it through observing system simulation experiments (OSSEs) with simulated radar observations. Then this newly-developed ARPS-LETKF system was inter-compared with the existing ARPS-EnSRF system in OSSEs also. Furthermore, this ARPS LETKF system was extended to its four dimensional version, 4D-LETKF, and compared with regular 3D-LETKF. Finally, the constrained ARPS-LETKF was built up based on the ARPS-LETKF framework. Divergence equation and pressure tendency equation were used as two equation constraints, for the purpose of reinforcement of the balance in LETKF analysis.

The results of the experiments of ARPS-LETKF with simulated radar observations show that this newly-developed system can effectively assimilate radar observations and produce good analyses which fit the true state very well after a few DA cycles under the perfect model assumption. By using the incremental interpolation or weight interpolation scheme in ARPS-LETKF, the computational costs were reduced significantly without loss of analysis performance. This LETKF algorithm was then parallelized with OpenMP directives. The computational performance of LETKF with OpenMP shows the capability of LETKF to be used as parallelizing filter for large data assimilation problems in the future work.

Through the experiments for the inter-comparison of EnSRF and LETKF with simulated radar radial wind and reflectivity observations, it is found that the performances of EnSRF with B-localization, LETKF with R-localization, and LETKF with regulated R-localization are very similar after the spin-up stage, with their optimal localization cut-off radii respectively. R-localization uses a shorter optimal cut-off radius, while the regulated R-localization and B-localization use the same optimal radius. It is also found that, in the spin-up stage, when using linear radial wind data, the smaller imbalance in the LETKF analysis may help the LETKF perform a little bit better. These results are consistent with previous research on the inter-comparison of EnSRF and LETKF (Greybush et al. 2011; Holland and Wang 2013).

However, when reflectivity data are analyzed, EnSRF analysis outperformed LETKF analysis systematically in the earlier analysis steps, so that EnSRF reached the final steady stage earlier than LETKF by 1~2 DA steps. The differences in the performances from EnSRF and LETKF are major and systematic in the spin-up stage, and then become trivial in the final steady stage. These results have not been reported or discussed in previous research.

By inspecting the implicit linearization in EnSRF and LETKF, it was found that EnSRF and LETKF make the implicit linearization at different places. The linearization in EnSRF is sequentially re-made over the intermediate ensemble, which is different to, and more accurate than the entire prior ensemble used by linearization in LETKF. In this regard, the EnSRF is superior to LETKF. The differences in linearization lead to the systematic/biased differences in the analyses from EnSRF and LETKF. This is the effect of the nonlinear observation forward operator on the EnKF analysis. Further inspection shows that the real nonlinearity and induced nonlinearity are responsible for the significance of the effect of the nonlinear observation operator. In the spin-up stage, due to the poor quality of the initial ensemble, the significant induced nonlinearity makes the nonlinear effect significant, then leads to significant differences between LETKF and EnSRF. But in the later steady stage, the background ensemble is improved through several analysis steps and is closer to the observations. Then the induced nonlinearity turns to small, and the effect of nonlinear observation become trivial. Five additional experiments were designed to test the theory about the effects of the nonlinear observation operator. All the results from these additional experiments supported the above theory about the effect of the nonlinear observation forward operator.

Then the ARPS-LETKF system was extended to its 4D version. With this 4D-LETKF algorithm, high frequency observations distributed over time could be analyzed simultaneously without timing error. This 4D-LETKF system was tested and compared with the regular 3D-LETKF by analyzing the simulated radar observations. The results indicated that,

compared to 3D-LETKF with intermittent assimilation cycles (e.g., 1 minute), 4D-LETKF with longer time windows (e.g., 3 ~ 5 minutes) can produce similar analyses after a few DA cycles while reducing computational cost significantly by using fewer cycles. When both of 4D and 3D LETKF using 3 minute or longer data batch, 4D-LETKF performs better than 3D-LETKF, mainly due to the observation timing error in 3D-LETKF.

The effect of time localization was also inspected. It could improve the performance of 3D-LETKF with longer data batch (like 10 minutes) by suppressing the harmful influence of data with large timing error. Time localization also could help to improve the performance for 4D LETKF with longer data batches by removing the negative influence from the observations observed at the time which is not close to the analysis time.

The hierarchical filter (HF) was then adopted as an adaptive localization scheme in this 4D-LETKF system. The results showed that the hybrid localization of HF and regular non-adaptive localization could improve the performance of 4D-LETKF, especially when 4D-LETKF using longer time window.

The last portion of this research was to build up a constrained ARPS-LEKTF system that imposed the divergence equation and pressure tendency equation as weak constraints to the analysis. In the constrained LETKF scheme, these two constraints were treated as special “observations” with specified error as the weighting coefficient of the constraints. The LETKF analysis algorithm did not change, except that the observation space was augmented by the observations of the divergence and pressure tendency, and the observation forward operators were also augmented with the divergence equation and pressure tendency equation.

Then this constrained ARPS-LETKF system was tested in the DA cycling experiments with simulated radar observations. These two constraints showed similar effects on the performance of LETKF. With a proper constraint weight, both of the constraints show a significant improvement on the analysis of the pressure perturbation. Because the pressure perturbations and these two constraints are closely related through the mass and dynamic balance, this improvement in analysis of pressure indicates that this constrained LETKF could help to enhance the balance in LETKF analysis. It also should be pointed out that the overall quality of analyses for the other variables does not change much. It suggests deeper inspections into the constrained LETKF in convective-scale DA in the future work.

5.2 Future Plan

To use this ARPS-LETKF system in the large dimension DA problem, this LETKF analysis scheme should be parallelized with high efficiency. Current ARPS-LETKF code was already parallelized with OpenMP. But it could not easily handle the problem with gigantic dimension and massive amount of observations. And OpenMP parallel code is not very efficient on the popular distributed parallel computational environment, such as the clusters of computer, the emerging cloud computing platform. So this ARPS-LETKF will be parallelized in a way of hybrid of MPI and OpenMP soon.

LETKF-RIP (Run-In-Place), as an iterative algorithm of 4D-LETKF, had been applied to several different DA problems (Kalnay and Yang 2010; Yang et al. 2012) . In the convective-scale DA, Wang et al. (2013b) developed the iterative 4DEnSRF with WRF model and tested it in OSSEs. Their research work shows the advantages of the iterative approach for 4DEnSRF, which includes the better analysis for nonlinear problems and reducing the period of spin-up. So a tentative future work is to implement the iterative version based on the current ARPS-4DLETKF system, and then inter-compare it with the existing iterative ARPS-4DEnSRF.

References

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

Anderson, J. L., 2007: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D: Nonlinear Phenomena*, **230,** 99-111.

Anderson, J. L. and N. Collins, 2007: Scalable implementations of ensemble filter algorithms for data assimilation. *J. Atmos. Ocean. Technol.* , **24,** 1452-1463.

Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120,** 1367-1387.

Courtier, P., E. Andersson, E. Keckley, J. Pailleux, D. Vasiljevic, M. Hamurd, A. Hollingsworth, F. Rabier, and M. Fisher, 1998: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: formulation. *Quart. J. Roy. Met. Soc.*, **124,** 1783-1808.

Dong, J., M. Xue, and K. K. Droegemeier, 2011: The analysis and impact of simulated high-resolution surface observations in addition to radar data for convective storms with an ensemble Kalman filter. *Meteor. Atmos. Phy.*, **112,** 41-61.

Doviak, R. and D. Zrnic, 1993: *Doppler Radar and Weather Observations*. 2nd ed. Academic Press, 562 pp.

Dowell, D. C. and L. J. Wicker, 2009: Additive noise for storm-scale ensemble data assimilation. *J. Atmos. Oceanic Technol.*, **26,** 911-927.

Dowell, D. C., L. J. Wicker, and C. Snyder, 2011: Ensemble Kalman filter assimilation of radar observations of the 8 May 2003 Oklahoma City supercell: Influence of reflectivity observations on storm-scale analysis. *Mon. Wea. Rev.*, **138,** 1152-1171.

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.

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

Gao, J.-D., M. Xue, A. Shapiro, and K. K. Droegemeier, 1999: A variational method for the analysis of three-dimensional wind fields from two Doppler radars. *Mon. Wea. Rev.*, **127,** 2128-2142.

Gao, J.-D., M. Xue, K. Brewster, and K. K. Droegemeier, 2004: A three-dimensional variational data analysis method with recursive filter for Doppler radars. *J. Atmos. Ocean. Technol.*, **21,** 457-469.

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

Ge, G., J. Gao, M. Xue, and K. K. Droegemeier, 2012: Diagnostic pressure equation as a weak constraint in a storm-scale three dimensional variational radar data assimilation system. *J. Atmos. Ocean. Tech.*, **29,** 1075-1092.

Greybush, S. J., E. Kalnay, T. Miyoshi, K. Ide, and B. R. Hunt, 2011: Balance and Ensemble Kalman Filter Localization Techniques. *Mon. Wea. Rev.*, **139,** 511 - 522.

Gustafsson, N., 2007: Discussion on ‘4D-Var or EnKF?’. *Tellus*, **59,** 774-777.

Harlim, J. and B. R. Hunt, 2007: Four-dimensional local ensemble transform Kalman filter: numerical experiments with a global circulation model. *Tellus A*, **59,** 731-748.

Holland, B. and X. Wang, 2013: Effects of sequential or simultaneous assimilation of observations and localization methods on the performance of the ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, **139,** 758-770.

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

Hu, M. and M. Xue, 2007: Impact of configurations of rapid intermittent assimilation of WSR-88D radar data for the 8 May 2003 Oklahoma City tornadic thunderstorm case. *Mon. Wea. Rev.*, **135,** 507–525.

Hu, M., M. Xue, J. Gao, and K. Brewster, 2006: 3DVAR and cloud analysis with WSR-88D level-II data for the prediction of Fort Worth tornadic thunderstorms. Part II: Impact of radial velocity analysis via 3DVAR. *Mon. Wea. Rev.*, **134,** 699-721.

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

Hunt, B. R., E. Kalnay, E. J. Kostelich, E. Ott, D. J. Patil, T. Sauer, I. Szunyogh, J. A. York, and A. V. Zimin, 2004: Four-dimensional ensemble Kalman filtering. *Tellus*, **56A,** 273-277.

Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory*. *Academic Press*, 376 pp.

Johns, C. J. and J. Mandel, 2008: A two-stage ensemble Kalman filter for smooth data assimilation. *Environmental and Ecological Statistics*, **15,** 101-110.

Jung, Y., 2008: State and Parameter Estimation using Polarimetric Radar Data and an Ensemble Kalman Filter (Ph. D. Dissertation), School of Meteorology, University of Oklahoma, 209.

Jung, Y., G. Zhang, and M. Xue, 2008a: Assimilation of simulated polarimetric radar data for a convective storm using ensemble Kalman filter. Part I: Observation operators for reflectivity and polarimetric variables. *Mon. Wea. Rev.*, **136,** 2228-2245.

Jung, Y., M. Xue, and M. Tong, 2012: Ensemble Kalman filter analyses of the 29-30 May 2004 Oklahoma tornadic thunderstorm using one- and two-moment bulk microphysics schemes, with verification against polarimetric data. *Mon. Wea. Rev.*, **140,** 1457-1475.

Jung, Y., M. Xue, G. Zhang, and J. Straka, 2008b: Assimilation of simulated polarimetric radar data for a convective storm using ensemble Kalman filter. Part II: Impact of polarimetric data on storm analysis. *Mon. Wea. Rev.*, **136,** 2246-2260.

Kalnay, E. and S. C. Yang, 2010: Accelerating the spin-up of ensemble Kalman filtering. *Q J Roy Meteor Soc*, **136,** 1644-1651.

Kalnay, E., H. Li, T. Miyoshi, S.-C. Yang, and J. BALLABRERA-POY, 2007a: Response to the discussion on “4-D-Var or EnKF?” by Nils Gustafsson. *Tellus*, **57,** 778-780.

Kalnay, E., H. Li, M. Takemsa, S.-C. Yang, and J. Ballabrera-Poy, 2007b: 4DVAR or ensemble Kalman filter. *Tellus*, **59A,** 758-773.

Kepert, J. D., 2009: Covariance localisation and balance in an Ensemble Kalman Filter. *Q. J. R. Meteorol. Soc.*, **135,** 1157-1176.

Liu, C., 2005: Design, Developement and Research of an Ensemble Kalman filter Data Assimilation Scheme (Thesis for Master Degree in Chinese), Chinese Academy of Meteorological Science, Chinese Academy of Meteorological Science, pp69.

Liu, J., E. J. Fertig, H. Li, E. Kalnay, B. R. Hunt, E. J. Kostelich, I. Szunyogh, and R. Todling, 2007: Comparison between Local Ensemble Transform Kalman Filter and PSAS in the NASA finite volume GCM: Perfect model experiments. *Atmospheric and Oceanic Physics*, http://arxiv.org/abs/physics/0703066v1.

Lorenc, A., 1995: Atmospheric Data Assimilation. *the Second WMO Symposium on Assimilation of Observations in Meteorology and Oceanography*, Tokyo, Japan.

Lorenc, A., 2003a: Modelling of error covariances by 4DVAR data assimilation. *Qart. J. Roy. Meteor. Soc.*, **129,** 3167-3182.

Lorenc, A., 2003b: The potential of the ensemble Kalman filter for NWP - a comparison with 4D-Var. *Quart. J. Roy. Meteor. Soc.*, **129,** 3183-3204.

Lorenc, A. C., 1988: Optimal nonlinear objective analysis. *Quart. J. Roy. Meteor. Soc.*, **114,** 205-240.

Lorenc, A. C., 1997: Development of an operational variational assimilation scheme. *J. Met Soc Japan*, **75,** 339-346.

Miyoshi, T., 2010: An adaptive covariance localization method with the LETKF (presentation). *American Meteorological Society Annual Meeting 2010*, Atlanta, GA.

Miyoshi, T. and S. Yamane, 2007: Local Ensemble Transform Kalman Filtering with an AGCM at a T159/L48 Resolution. *Mon. Wea. Rev.*, **135,** 3841-3861.

Miyoshi, T. and M. Kunii, 2012: The Local Ensemble Transform Kalman Filter with the Weather Research and Forecasting Model: Experiments with Real Observations. *Pure and Applied Geophysics*, **169,** 321 - 333.

Nerger, L., T. Janjic, J. Schroter, and W. Hiller, 2012: A regulated localization scheme for ensemble-based Kalman filters. *Q. J. R. Meteorol. Soc.*, **138,** 802-812.

Ott, E., B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. J. Patil, and J. A. Yorke, 2004: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus A*, **56,** 415-428.

Pan, M. and E. F. Wood, 2006: Data Assimilation for Estimating the Terrestrial Water Budget Using a Constrained Ensemble Kalman Filter. *J. Hydrometeor*, **7,** 534-547.

Phale, H. A. and D. S. Oliver, 2011: Data Assimilation Using the Constrainted Ensemble Kalman Filter. *Society of Petroleum Engineers Annual Technical Conference and Exhibition.*

Ray, P. S., B. Johnson, K. W. Johnson, J. S. Bradberry, J. J. Stephens, K. K. Wagner, R. B. Wilhelmson, and J. B. Klemp, 1981: The morphology of severe tornadic storms on 20 May 1977. *J. Atmos. Sci.*, **38,** 1643-1663.

Sakov, P., G. Evensen, and L. Bertino, 2010: Asynchronous data assimilation with the EnKF. *Tellus*, **62,** 24-29.

Sakov, P., D. S. Oliver, and L. Bertino, 2012: An Iterative EnKF for Strongly Nonlinear Systems. *Mon. Wea. Rev.*, **140,** 1988-2004.

Seko, H., T. Miyoshi, Y. Shoji, and K. Saito, 2011: Data assimilation experiments of precipitable water vapour using the LETKF system: intense rainfall event over Japan 28 July 2008. *Tellus*, **63A,** 402-414.

Simon, D., 2010: Kalman Filtering with State Constraints: a Survey of Linear and Nonlinear Algorithms. *IET Control Theory and Applications*, **4,** 1303-1318.

Smith, P. L., Jr., C. G. Myers, and H. D. Orville, 1975: Radar reflectivity factor calculations in numerical cloud models using bulk parameterization of precipitation processes. *J. Appl. Meteor.*, **14,** 1156-1165.

Snook, N., M. Xue, and J. Jung, 2011: Analysis of a tornadic meoscale convective vortex based on ensemble Kalman filter assimilation of CASA X-band and WSR-88D radar data. *Mon. Wea. Rev.*, **139,** 3446-3468.

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

Szunyogh, I., E. J. Kostelich, G. Gyarmati, E. Kalnay, B. R. Hunt, E. Ott, and E. Satterfield, 2008: A local ensemble tranform Kalman filter data assimilation system for the NCEP global model. *Tellus*, **60A,** 113-130.

Szunyogh, I., E. J. Kostelich, G. Gyarmati, D. J. Patil, B. R. Hunt, E. Kalnay, E. Ott, and J. Yorke, 2005: Assessing a local ensemble Kalman filter: perfect model experiments with the National Centers for Environmental Prediction global model. *Tellus*, **57A,** 528-545.

Tong, M., 2006: Ensemble Kalman filer assimilation of Doppler radar data for the initialization and prediction of convective storms (Ph.D. Dissertation), School of Meteorology, University of Oklahoma, 243.

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.

Tong, M. and M. Xue, 2008a: Simultaneous estimation of microphysical parameters and atmospheric state with radar data and ensemble square-root Kalman filter. Part II: Parameter estimation experiments. *Mon. Wea. Rev.*, **136,** 1649–1668.

Tong, M. and M. Xue, 2008b: Simultaneous estimation of microphysical parameters and atmospheric state with radar data and ensemble square-root Kalman filter. Part I: Sensitivity analysis and parameter identifiability. *Mon. Wea. Rev.*, **136,** 1630–1648.

Wang, D., Y. Chen, and X. Cai, 2009: State and parameter estimation of hydrologic models using the constrained ensemble Kalman filter. *Water Resour. Res.*, **45**.

Wang, S., M. Xue, and J. Min, 2013a: A four-dimensional asynchronous ensemble square-root filter (4DEnSRF) algorithm and tests with simulated radar data. *Quarterly Journal of the Royal Meteorological Society*, **139,** 805-819.

Wang, S., M. Xue, A. D. Schenkman, and J. Min, 2013b: An iterative ensemble square root filter and tests with simulated radar data for storm-scale data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **139,** 1888-1903.

Wang, Y., Y. Jung, T. A. Supinie, and M. Xue, 2013c: A Hybrid MPI–OpenMP Parallel Algorithm and Performance Analysis for an Ensemble Square Root Filter Designed for

Multiscale Observations. *Journal of Atmospheric and Oceanic Technology*, **30,** 1382-1397.

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

Xue, M., K. K. Droegemeier, and V. Wong, 2000: The Advanced Regional Prediction System (ARPS) - A multiscale nonhydrostatic atmospheric simulation and prediction tool. Part I: Model dynamics and verification. *Meteor. Atmos. Phy.*, **75,** 161-193.

Xue, M., M. Tong, and K. K. Droegemeier, 2006: An OSSE framework based on the ensemble square-root Kalman filter for evaluating impact of data from radar networks on thunderstorm analysis and forecast. *J. Atmos. Ocean Tech.*, **23,** 46–66.

Xue, M., M. Tong, and G. Zhang, 2009: Simultaneous state estimation and attenuation correction for thunderstorms with radar data using an ensemble Kalman filter: Tests with simulated data. *Q. J. Roy. Meteor. Soc.*, **135,** 1409-1423.

Xue, M., Y. Jung, and G. Zhang, 2010: State estimation of convective storms with a two-moment microphysics scheme and an ensemble Kalman filter: Experiments with simulated radar data *Q. J. Roy. Meteor. Soc*, **136,** 685-700.

Xue, M., K. K. Droegemeier, V. Wong, A. Shapiro, and K. Brewster, 1995: *ARPS Version 4.0 User's Guide*. [Available at http://www.caps.ou.edu/ARPS], 380 pp.

Xue, M., D.-H. Wang, J.-D. Gao, K. Brewster, and K. K. Droegemeier, 2003: The Advanced Regional Prediction System (ARPS), storm-scale numerical weather prediction and data assimilation. *Meteor. Atmos. Phy.*, **82,** 139-170.

Xue, M., K. K. Droegemeier, V. Wong, A. Shapiro, K. Brewster, F. Carr, D. Weber, Y. Liu, and D. Wang, 2001: The Advanced Regional Prediction System (ARPS) - A multi-scale nonhydrostatic atmospheric simulation and prediction tool. Part II: Model physics and applications. *Meteor. Atmos. Phys.*, **76,** 143-165.

Yang, S.-C., E. Kalnay, and B. Hunt, 2012: Handling nonlinearity in an Ensemble Kalman Filter: Experiments with the three-variable Lorenz model. *Mon. Wea. Rev.*, 10.1175/mwr-d-11-00313.1.

Yang, S.-C., E. Kalnay, B. Hunt, and N. E. Bowler, 2009: Weight interpolation for efficient data assimilation with the Local Ensemble Transform Kalman Filter. *Q. J. R. Meteorol. Soc.*, **135,** 251-262.

Yussouf, N. and D. J. Stensrud, 2010: Impact of phased-array radar observations over a short assimilation period: Observing system simulation experiments using an ensemble Kalman filter. *Mon. Wea. Rev.*, **138,** 517-538.

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