THE DEVELOPMENT OF A HYBRID ENKF-3DVAR ALGORITHM FOR STORM-SCALE DATA ASSIMILATION

Jidong Gao[1,3], Ming Xue[1,2] and David J. Stensrud[3]

[1]*Center for Analysis and Prediction of Storms,* [2]*School of Meteorology, University of Oklahoma,*
[3]*NOAA/National Severe Storm Laboratory, Norman OK 73072*

1. Introduction

Many studies have been performed during the past decade aiming to use Doppler radar data for initializing cloud-resolving models. At the Center for Analysis and Prediction of Storms (CAPS), the ARPS Data Analysis System (ADAS, Brewster 2003a, b), with incremental analysis updating capabilities, was developed as the first step towards assimilating radar data and other conventional and remotely-sensed data into a nonhydrostatic NWP model, the ARPS (Xue et al. 2000; 2001). A limitation of ADAS is that observations that differ from the analysis variables cannot be directly analyzed and are handled only by generation of pseudo-observations (e.g., via single-Doppler velocity retrievals; Weygandt et al. 2002a, b). Examples of such observations include radar radial velocity and reflectivity, GPS precipitable water, and satellite radiances. All of these data are very important to storm-scale data assimilation. To more effectively use these data, advanced DA techniques, including variational and ensemble Kalman filter methods are needed.

In the context of large-scale hydrostatic flows, the 3DVAR method of analysis has reached a considerable state of maturity at operational NWP centers (Derber et al. 1991; Parrish and Derber 1992; Andersson et al. 1998; Courtier et al. 1998; Rabier et al. 1998; Wu et al. 2002, Barker et al. 2004). Non-conventional data, including those from satellite and radar, can be directly analyzed by 3DVAR. In current operational 3DVAR systems, analysis increments are often divided into balanced and unbalanced parts, with the balanced parts often linked together via statistically derived balance relations. For example, in the NCEP system (e.g., Wu et al. 2002, Purser et al. 2003a, b), the temperature increment is a sum of the unbalanced and balanced parts, with the latter computed from the stream-function increment via a balance operator reflecting geostrophy. This balance operator is a linear regression derived from historical data. Unfortunately, because such regressions are mainly consistent with geostrophy and the hydrostatic relationship, they are not entirely appropriate for storm-scale data assimilation. Furthermore, the background error covariances derived from the commonly used NMC method (Parrish and Derber 1992) usually do not reflect local weather situations that are highly intermittent in space and time. For these reasons, large-scale 3DVAR schemes cannot be directly extended to the meso- and convective scale nonhydrostatic flows.

A 3DVAR method, which uses mass continuity equations and other appropriate model equations as weak constraints, has been proposed and developed in recent years (Gao et al. 2004; Ge and Gao 2007). This method is mainly designed to assimilate radar data into convective scale nonhydrostatic flows (Hu et al. 2006, 2007; Ge et al. 2010; Stensrud and Gao 2010). It has been used to perform realtime 4-km multi-model convection-allowing ensemble and 1-km convection-resolving deterministic forecasts for the NOAA Hazardous Weather Testbed (HWT)'s Spring experiments since 2008 (Xue et al. 2008; Kong et al. 2009). The unique feature of these experiments is that Level-II radial velocity and reflectivity data from over 120 operational WSR-88D radars are analyzed using the ARPS 3DVAR program. The analyses serve as the control initial conditions onto which perturbations are added for ensemble runs. The use of more appropriate equation constraints couples different model variables, e.g., the three components of wind field. Xiao et al. (2005) used a 3DVAR method developed for WRF model to assimilate Doppler radar observation into WRF model. However, the truly flow-dependent background error covariances were also not included in the WRF 3DVAR system which was mainly designed for large mesoscale system.

Compared to 3DVAR, the more advanced 4DVAR technique incorporates the full prediction model into the assimilation system and implicitly includes the effects of flow-dependent error covariances through involvement of the forward model and backward model. In recent years, the 4DVAR technique has enjoyed significant success at several operational NWP centers, including ECMWF, Meteo-France, Meteorological Service of Canada, and Japan

Corresponding author current address: Dr. Jidong Gao, NOAA/National Severe Storms Laboratory, NWC, 120, David, L. Boren Blvd., Norman, OK 73072. *jidong.gao@noaa.gov*

Meteorological Agency (JMA), mainly in global NWP systems (Rabier et al. 2005). Some research has also been done with storm-scale radar data assimilation using the 4DVAR method, notably by Sun and Crook at NCAR (1997, 1998, 2001, 2005). Despite some encouraging results, 4DVAR for convective-scale applications has been limited to simple microphysics in almost all cases because the strong nonlinearity is difficult to handle in the minimization process. Honda and Koizumi (2006) recently reported difficulties, including slow convergence, when including complex ice microphysics within the inner loop of their 4DVAR system for a nonhydrostatic model at JMA. The high computational cost of 4DVAR is another hindrance for high-resolution NWP over large domains though this can be alleviated with the increasing power of new generations of computers.

The ensemble Kalman filter (EnKF) is an emerging advanced data assimilation method that shares many of the advantages of 4DVAR but appears to be less sensitive to nonlinearity. It has gained considerable popularity in recent years in meteorology and oceanography since first proposed by Evensen (1994). The method estimates flow-dependent background error covariance from an ensemble of nonlinear model forecasts, which is much cheaper than trying to integrate the prediction equation for the error covariance as in the classical or extended Kalman filter. Many studies have since appeared that explore new EnKF variations and the performance of these methods for various applications (Bishop et al. 2001; Burgers *et al.* 1998; Hamill and Synder 2000; Houtekamer and Mitchell 1998; Whitaker and Hamill 2002).

For convective storms, very encouraging results have been obtained in recent studies using the ensemble Kalman filter method in analyzing wind, temperature, moisture fields and even microphysics variables from radar data for convective storms (Snyder and Zhang 2003; Zhang et al. 2004; Tong and Xue 2005; Alkoy et al. 2009; Dowell et al. 2004, 2010; Yussouf and Stensrud 2010). One of the advantages of the EnKF method over variational methods is that it can dynamically evolve the background error covariances throughout the assimilation cycles. However, for the storm scale, the overall computational cost of ensemble-based assimilation methods is quite significant because of the need for running an ensemble of forecasts and analyses of nontrivial sizes (usually a few dozen to a few hundred), especially when high-density radar data are involved and when the ensemble of all forecasts is run at high resolution over a big domain. One of the major sources of errors with ensemble-based DA is rank deficiency or sampling error as a result of relatively small ensemble sizes. This problem may be even more severe with storm-scale data assimilation because the total degrees of freedom of the system may be significantly larger than the practical ensemble size. The commonly utilized remedy to the rank deficiency problem is covariance localization by a Schur product introduced by Houtekamer and Mitchell (2001). This solution, however, prevents the use of distant correlations that are physically meaningful. Further, the modification to the spatial covariances within a cut-off radius by a Schur product also introduces imbalances, and the effect is more substantial when the localization is more restrictive (Mitchell et al. 2002). A larger ensemble helps improve the background error covariance estimation, but increases the computational cost. To alleviate the above problems, Gao and Xue (2008) proposed a dual-resolution (DR) EnKF data assimilation strategy in which an ensemble of forecasts is run at lower resolution (LR) to provide the background error covariance estimation for both an ensemble of LR analyses and a single Higher Resolution (HR) analysis. The DR strategy has two benefits. One is that the computational cost of the overall EnKF analysis can be greatly decreased. The second is that the problem of rank deficiency, or sampling error, can be reduced by increasing the ensemble size because of the low cost of LR runs.

As discussed above, the 3DVAR method proposed by Gao et al. (2004) and Ge and Gao (2007) is attractive for convective scale because of its computational efficiency and use of simplified model equations as constraints without resorting to 4DVAR. Because the model equations are not applied as strong constraints, problems with nonlinearity of the system may be reduced, and the convergence to a global minimum might be achieved more easily than in 4DVAR. However, the major shortcoming is that the background error covariances are stationary and isotropic, and error covariances related to model equations can not be simply defined. For convective-scale weather, considering the nature of radar data, flow-dependent background error covariances, such as derived from EnKF method, are sorely needed. To blend advanced features of both the variational and EnKF methods, and to overcome their respective shortcomings, a hybrid EnKF-based 3DVAR framework is needed. For large-scale data assimilation, such an approach was initially demonstrated for a quasi-geostrophic system by Hamill and Snyder (2000) and further suggested by Lorenc (2003), Buehner (2005), and Zupanski (2005) with different formulations. Wang et al. (2007) showed that the formulations proposed by Hamill and Synder (2000), Lorenc (2003) and Buehner (2005), though different in their implementations and computational cost, are all mathematically equivalent. In the hybrid method, the variational framework is used to conduct data

assimilation within which the flow-dependent ensemble covariances are effectively incorporated to estimate the background error covariance. Further case studies have demonstrated the potential advantages of the hybrid method over both standalone variational method and EnKF method with small ensemble size (Wang et al. 2008a,b; Buehner et al. 2010a,b). However, whether these hybrid methods can be directly extended to the convective scale has not been explored so far. The purpose of this paper is to demonstrate the potential usefulness of the hybrid EnKF-3DVAR method to convective scale data assimilation, especially with radar data.

In section 2, we introduce the hybrid EnKF-3DVAR system while section 3 describes the DA experiment design. Experiment results and quantitative performance are assessed in section 4. We conclude in section 5 with a summary and outlook for future work.

2. The Hybrid EnKF-3DVAR scheme

In the current implementation of the hybrid method for convective scale, the ensemble covariance is incorporated in the variational framework through the extended control variable method (Lorenc 2003; Buehner 2005; Wang et al. 2007), and the ensemble covariance localization is conducted in the model state variable space (i.e., model space localization), and preconditioning is performed with respect to the background term. A convenient approach, initially suggestion by Lorenc (2003), is to combine the ensemble-derived and static covariance matrices through augmentation of the state vector, from **v** to (**v**, **w**). The cost function can be written as,

$$J = \frac{1}{2}\mathbf{v}^T\mathbf{v} + \frac{1}{2}\mathbf{w}^T\mathbf{w} + \frac{1}{2}\left[H\left(\overline{\mathbf{x}}^b + \Delta\mathbf{x}\right) - \mathbf{y}^o\right]^T \mathbf{R}^{-1}\left[H\left(\overline{\mathbf{x}}^b + \Delta\mathbf{x}\right) - \mathbf{y}^o\right] + J_c \qquad (1)$$

$$\Delta\mathbf{x} = \Delta\mathbf{x}_1 + \Delta\mathbf{x}_2 = \beta_1\left(\mathbf{B}\right)^{1/2}\mathbf{v} + \beta_2\left(\mathbf{P}\right)^{1/2}\mathbf{w}, \qquad (2)$$

Where, $\Delta\mathbf{x}$ is the analysis increment, **B** is the full-rank 3DVAR covariance matrix and is modeled in our system using recursive filters (Purser 2003a, b). **P** is the covariance matrix derived from an ensemble of forecasts. The **v** is the control variable defined in associated with **B**, and **w** is the augmented control vector associated with **P**. By using control variables **v** and **w**, instead of $\Delta\mathbf{x}_1$, $\Delta\mathbf{x}_2$ in Eq. (2), the minimization procedure is preconditioned by $\left(\mathbf{B}\right)^{1/2}$, $\left(\mathbf{P}\right)^{1/2}$ respectively. This technique is first proposed by Derber and Rosati (1989). The definition of $\left(\mathbf{B}\right)^{1/2}$ is the same as Gao et al. (2004). If no covariance localization is applied, $\mathbf{P}^{1/2}$ is simply a rectangular matrix whose columns are ensemble perturbations divided by square root of N-1, where N is the ensemble size. The covariance localization of the ensemble covariance in a variational system with preconditioning is discussed in Lorenc (2003), Buehner (2005) and Wang et al. (2007). The procedure and cost of doing so were also discussed in these papers. For computational efficiency, we also use recursive filter to do covariance localization, as suggested in Wang et al. (2008a).

In Eq. (2), there are two factors β_1 and β_2 that define the weights placed on the static background-error covariance and the ensemble covariance. To conserve the total background-error variance, β_1 and β_2 are constrained by,

$$\beta_1^2 + \beta_1^2 = 1 \qquad (3)$$

A similar constraint was applied in Hamill and Snyder (2000). This approach for combining two matrices to form hybrid covariances provides flexibility since it allows for weighing the relative contributions of two covariance matrices. When β_1 = 1, the analysis is back to a 3DVAR-alone analysis scheme, and when β_2 = 1, then the analysis is mathematically equivalent to a EnKF-alone scheme, we can called is an En-3DVAR scheme. Though the dimension of the control variables is increased, the form of the background term of the cost function remains unchanged or similar so that codes from an existing 3DVAR system can readily be utilized (Lorenc 2003).

In our hybrid EnKF-3DVAR system, the simplified model equation constraints, such as mass continuity equation, can still be used, as hinted by Hamill and Synder (2000) and Lorenc (2003). Though the simplified covaviances related to these constraints can be determined by numerical experiments, they are specified as constants through the data assimilation cycles. Spatial structure is not allowed because it is difficult to know the error statistics for these model equations. Thus the role of these constraints may not be effective. For example, the simplified covariance for the mass continuity constraint may vary significantly in the area where a thunderstorm exists compared to a relatively clear area.

In addition, the hybrid system will assimilate both radar reflectivity data and radial velocity data. Within this system, flow-dependent background-error covariances, in particular cross-covariances between microphysical and dynamic variables will be derived and utilized.

The single-resolution EnKF system developed in Gao and Xue (2008) will be used for generating perturbations for ensemble members. It is an ensemble square-root Kalman filter algorithm of Whitaker and Hamill (2002), which also was used by Snyder and

Zhang (2003), Dowell et al. (2004) and Tong and Xue (2005). The basic update equation of the Kalman filter is,

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{P}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{P}\mathbf{H}^T)^{-1}\left[\mathbf{y}^o - H(\mathbf{x}^b)\right]^T \quad (4)$$

As defined earlier, **x** is the state vector we seek to estimate, and superscripts *a* and *b* refer to the analysis (posteriori estimate) and the background (prior estimate), respectively, and $\mathbf{y}^o$ is the observation vector. *H* is the forward observation operator, and **H** is its linearized version. **R** the covariance matrices for the observation and **P** is the ensemble covariance matrices defined earlier. This EnKF approach involves the use of forecast and analysis ensembles. First, an ensemble of ARPS models is integrated forward for the length of analysis cycle or until the next observation is available, so as to yield an ensemble of forecasts. Then, the EnKF analyses are done serially, one observation at a time. For a given observation, the background error covariance is calculated from the ensemble, and is used to do the analysis. This is repeated until all observations are done at a given time.

For the hybrid method, an extra model integration for the length of the analysis cycle is needed to produce a control forecast and analysis cycle. The EnKF analyses using Eq. (4) are performed to update analysis perturbations for each ensemble members. Then, the cost function (1) is minimized, to obtain optimal analyses of control vectors **v** and **w**, and the optimal analysis increment, $\Delta\mathbf{x}$, is derived from (2). The ensemble mean analysis is partially or completely replaced using the high-resolution 3DVAR analysis and ensemble perturbation using EnKF. Finally, the initial conditions for ensemble and one control member of forecast are produced. The above steps are repeated to carry out another data assimilation cycle.

Basically, we perform in parallel a set of EnKF analyses to produce ensemble forecast and statistics and a single 3DVAR analysis that utilizes error statistics derived from the ensemble system. The single 3DVAR analysis may feed back into the ensemble system by providing a better analysis to completely or partially replace the ensemble mean (Fig. 1).

3. Model and experimental Design

a) Prediction model and truth simulation for OSSEs

We test our hybrid EnKF-3DVAR algorithm and compare its results with those of 3DVAR-alone and EnKF-alone schemes, using simulated data from a classic supercell storm of May 20, 1977 Del City, Oklahoma (Ray *et al.* 1981). The ARPS prediction model is used in a 3D cloud model mode and the prognostic variables include three velocity components *u*, *v*, *w*, perturbation potential temperature θ', pressure *p*, and six categories of water substances, *i.e.*, water vapor specific humidity q_v, and mixing ratios for cloud water q_c, rainwater q_r, cloud ice q_i, snow q_s and hail q_h. The microphysical processes are parameterized using the three-category ice scheme of Lin *et al.* (1983). More details on the model can be found in Xue *et al.* (2000; 2001).

For our experiments, the model domain is 57×57×16 km^3. The horizontal grid spacing is 1 km and the vertical resolution is 500 m. The truth simulation run is initialized from a modified real sounding plus a 4 K ellipsoidal thermal bubble centered at x =48, y =16 and z =1.5 km, with radii of 10 km in x and y and 1.5 km in the z direction. Open conditions are used at the lateral boundaries. The length of simulation is 2 hours. A constant wind of u = 3 ms^{-1} and v =14 ms^{-1} is subtracted from the observed sounding to keep the primary storm cell near the center of model grid. The evolution of the simulated storms is similar to those documented in Xue *et al.* (2001). During the truth simulation, the initial convective cell strengthens over the first 30 min. The strength of the cell then decreases over the next 30 min or so, which is associated with the splitting of the cell in two at around 55 min. The right moving (relative to the storm motion vector which is towards north-northeast) cell tends to dominate the system and its updraft reaches a peak value of over 40 m s^{-1} at 90 min. The initial cloud starts to form at about 10 min, and rainwater forms at about 15 min. Ice phase fields appear at about 20 min. A similar truth simulation was also used in Gao et al. (2001), Tong and Xue (2005), Gao and Xue (2008).

b) Simulation of radar observations

As in Snyder and Zhang (2003) and Tong and Xue (2005), the simulated radial velocity observations are assumed to be available on the grid points. The simulated radial velocity, v_r, is calculated from,

$$v_r = u\sin\phi\cos\mu + v\cos\phi co\,s\,\mu + w\sin\mu \quad (5)$$

where μ is the elevation angle and ϕ is the azimuth angle of radar beams, and *u, v* and *w* are the model-simulated velocities interpolated to the scalar points of the staggered model grid. Random errors drawn from a normal distribution with zero mean and a standard deviation of 1 m s^{-1} are added to the simulated data.
Since v_r is sampled directly from the model velocity fields, hydrometeor sedimentation is not involved. The ground-based radar is located at the southwest corner of the computational domain, *i.e.,* at the origin

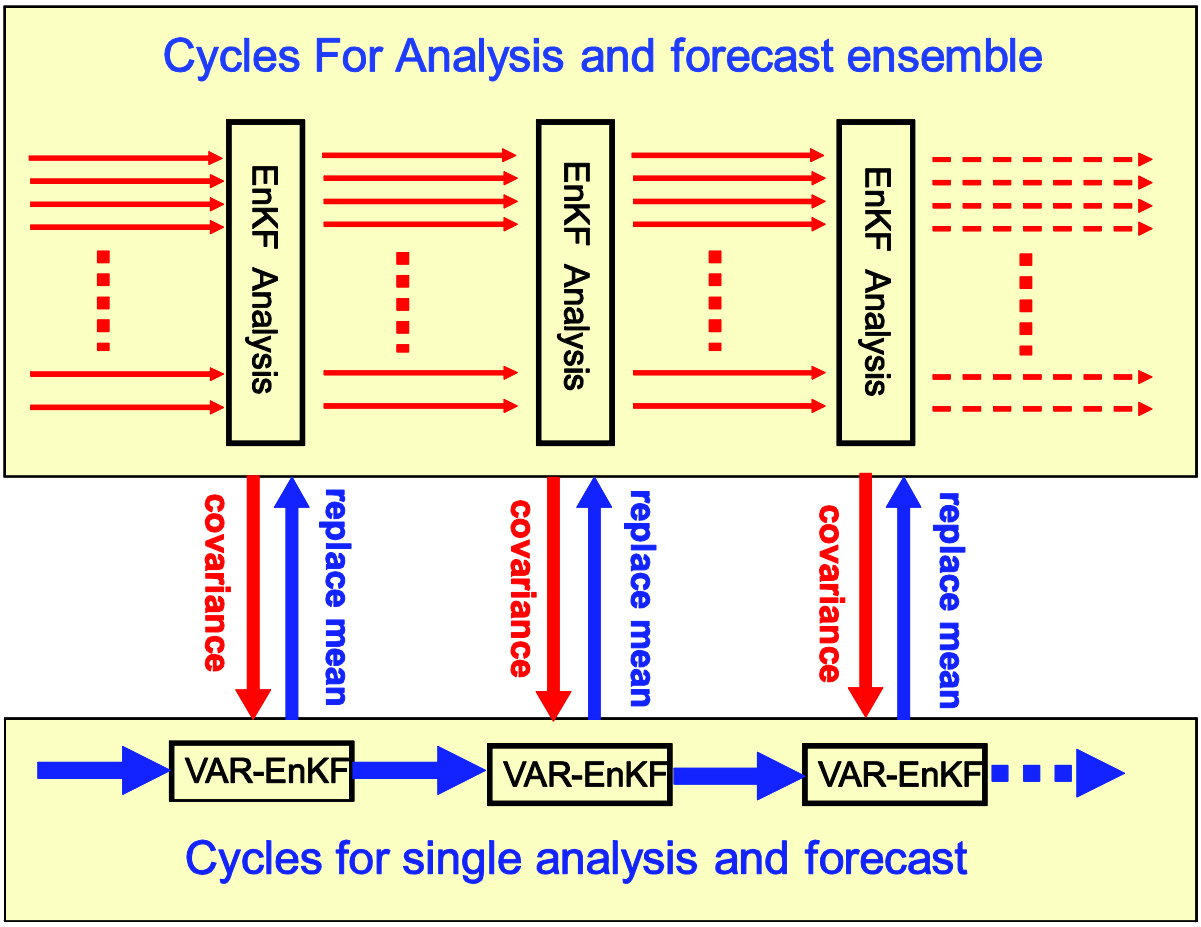

Fig. 1. Illustration of cycle used in a hybrid EnKF-3DVAR analysis scheme.

of the *x*-*y* coordinates. The simulated reflectivity observations are calculated based on Smith et al. (1975) and Ferrier (1994). For reflectivity, random errors drawn from a normal distribution with zero mean and a standard deviation of 3 dBZ are added to the simulated data. The radial velocity data are assimilated and are only available where the truth reflectivity is greater than zero in the analysis domain. We also use the data at every other grid point of the 1 km truth simulation grid in all three directions.

c) Design of assimilation experiments

We start the initial ensemble forecast at 20 min of the model integration time when the storm cell is well developed. To initialize the ensemble members, random noise is first added to the initially horizontally homogeneous first guess defined using the environmental sounding. A 2-D five-point smoother is applied to the resultant fields, similar to a method used by Zupanski et al. (2006). The random noise is sampled from Gaussian distributions with zero mean and standard deviations of 5 m s^{-1} for *u*, *v*, and *w*, and 3 K for potential temperature. These perturbation variances are somewhat larger than those used in Tong and Xue (2005) but the standard deviation of the final perturbations is not necessarily larger because of the smoothing. Other variables, including the microphysical variables, are not perturbed at the initial time. The radial and reflectivity observations are simulated and assimilated in 5 min cycle in all three data assimilation schemes. The first analysis is performed at 20 min and 20 ensemble members are used unless otherwise noted. To localize covariances during the analysis, Eq. (4.10) of Gaspari and Cohn (1999) is used when calculating the background error matrix $\mathbf{PH}^{\mathbf{T}}$, as suggested by Houtekamer and Mitchell (2001). A cut-off radius of 8 km is used in most of our experiments.

We perform two set of experiments. The first group of experiments is performed to compare the performances of three different schemes when observations from a single Doppler radar are used. The second group of experiments will be performed when observations from two Doppler radars are used. For comparison purposes, all three methods (3DVAR-alone, EnKF-alone and Hybrid EnKF-3DVAR) are performed with 16 data assimilation cycles and each cycle has a 5 minute interval. Sensitivity experiments with different combinations of weights and different physical scheme are being tested and have not been finished as the time of this writing.

4. Results

As stated above, the first group of experiments is performed with radial velocity and reflectivity from a single radar. Figure 2 shows the final assimilation results after 16 assimilation cycles with 5 minutes each. The low-level flow, reflectivity patterns, and the strength of the cold pool from both EnKF-alone and hybrid EnKF-3DVAR agree very well with the simulated truth (Fig. 2a) and are better than the result using 3DVAR-alone method (Fig. 2b), although this 3DVAR-alone can also establish the storm structures reasonably well (Fig. 2b). The most obvious difference is the reflectivity field in the center of model domain. The area of reflectivity values greater than 55 dBZ is over extended in a peanut-shaped region for the 3DVAR-alone method. The spread of potential temperature is little bit far to the south-south-west direction in the south-west corner (Fig. 2b). But the strength of the cold pool in the 3DVAR-alone method indicated by minimum potential temperature reaches to -7.30^{o}, which is closer to the truth simulation (-7.28^{o}) than both EnKF-alone method and hybrid method.

The rms errors of the analyzed fields with data from a single radar are shown in Figure 3. The rms error calculation is limited to the regions where the truth reflectivity exceeds 10 dB*Z*. Figure 3 shows that the rms errors for model variables *u, v*, *w*, *θ*, *qv*, and reflectivity *Z* (derived from the hydrometeor mixing ratios) generally decrease with the cycles in all three experiments. The errors for 3DVAR-alone decrease more slowly and remain at a higher level at the end of assimilation cycles than those for other two ensemble based method for most of model variables. For example, the rms error of *w* is close to 3 m s^{-1} at 100 min for 3DVAR-alone method, while that in EnKF and hybrid EnKF-3DVAR is close to 1.3 m s^{-1}. The rms errors of q_v for 3DVAR is 0.4 g/kg, and that in EnKF and hybrid EnKF-3DVAR is below 0.2 k/kg. While these differences are significant, the error levels of late assimilation period for EnKF and hybrid EnKF-3DVAR are unrealistically low due to the perfect model assumption. For real data cases where model error exists, the analysis errors are most likely to be much larger (see, e.g., Dowell et al. 2004, 2010). For systems containing discrete intense updrafts, the rms error tends to exaggerate errors because of small spatial displacement and/or structure discrepancies, such as those seen in Fig. 2. So the results for 3DVAR-alone are not unreasonable. It should be noted that for most of model variables, the performances of EnKF and hybrid method are very close, with EnKF a little bit better. Interestingly, the differences among the rms errors for *Z* in different experiments are smallest (Fig. 3f). The rms error of *Z* is decreased to about 5 dB*Z* for almost all three experiments. The variation of rms errors is volatile for 3DVAR method, especially near the very beginning of the assimilation. The method can bring down the errors from about 40 dBZ to 10 dBZ in two data assimilation cycles, but the errors quickly increase to above 20 dBZ after a 5-minute model integration step. The rms errors for the EnKF method decrease more smoothly throughout the data assimilation cycles because its statistical feature. Perhaps the advantage of hybrid method looks more obvious for reflectivity. It fits the observed reflectivity more closely than other two methods. Though the evolution of rms errors is also volatile for the first 10 minutes, it quickly settles down and its rms errors are the lowest among all three methods.

The second group of experiments is performed with radar data from two simulated Doppler radars. Figure 4 shows the final assimilation results after 16 assimilation cycles. As expected, the low-level flow, reflectivity patterns, and the strength of the cold pool look much better (Fig. 4b) for 3DVAR-alone method (Fig. 4b) especially for reflectivity field. The pattern for potential temperature is improved when compared with the single radar experiment (Fig. 2b), but still not as good as the truth simulation (Fig. 4a), and that for EnKF-alone (Fig. 4c) and hybrid EnKF-3DVAR (Fig. 4d). So with more data used, the results for 3DVAR-alone are quickly improved. Again, the most obvious improvement is for reflectivity field in the center of model domain. The area for reflectivity values larger than 55 dBZ is more similar to the shape of truth simulation. The storm structure for all three methods is well established by the end of data assimilation at 100 minutes of reference model assimilation time. The variation of rms errors for the analyzed fields using data from two radars is shown in Fig. 5. It is not surprising that the rms errors for model *u, v*, are much improved for the 3DVAR-alone method. For the first several data assimilation cycles, the errors for 3DVAR method are the lowest. With more cycles, the errors for the hybrid method become the lowest among three methods. For most of variables (except for potential temperature), the errors for 3DVAR-alone decrease

more quickly than that for other two methods for the first several data assimilation cycles, but remain at higher levels at later DA cycles. The variation of rms errors is less volatile with two radars than that with a single radar for 3DVAR method. The other features are quite similar to the cases when data from a single radar are used.

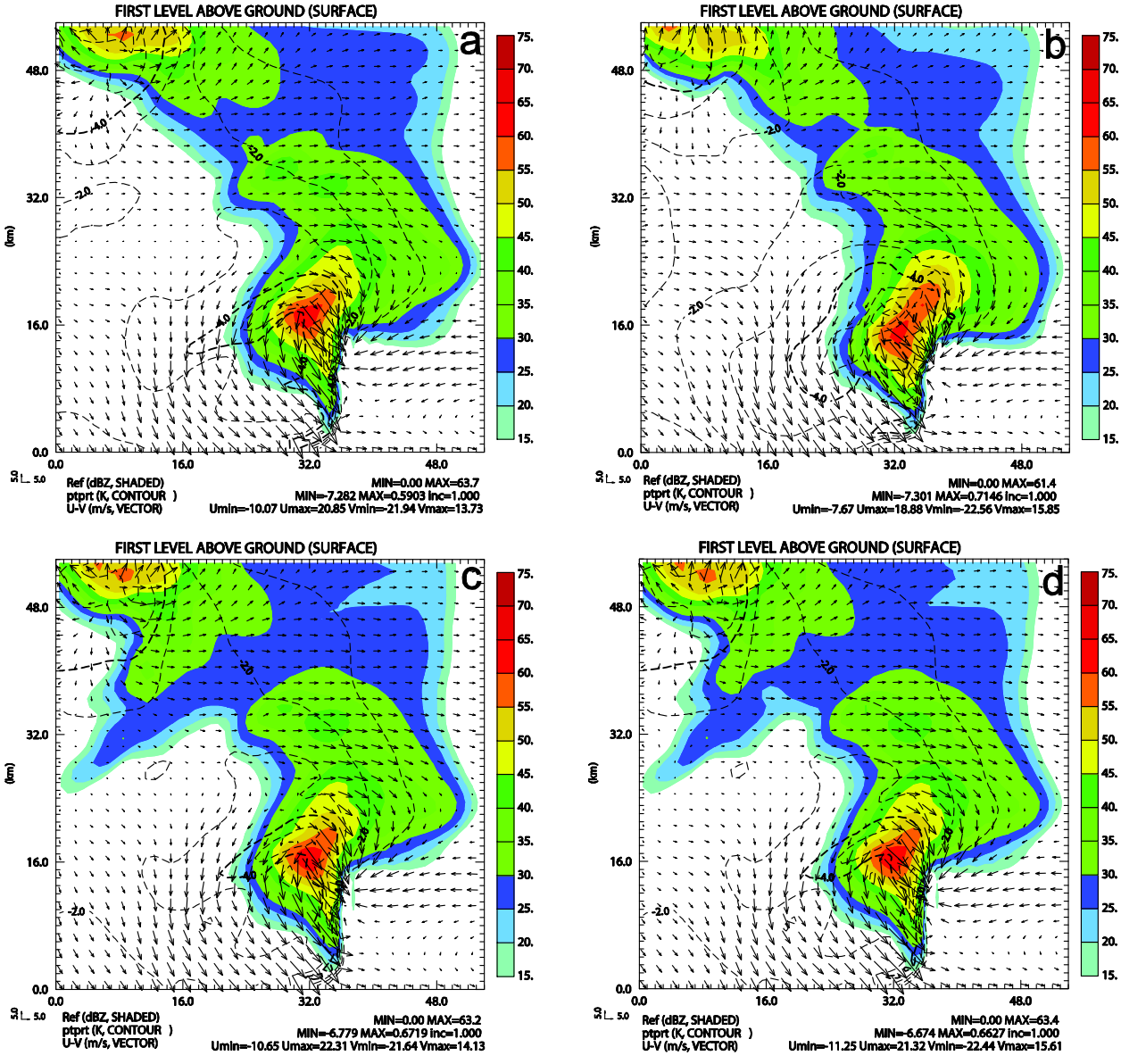

Fig. 2. Horizontal winds (vectors; m s^{-1}), perturbation potential temperature (contours at 1-K intervals), and simulated reflectivity (shaded contours; dBZ) at 250 m AGL for (a) the truth simulation; (b) the 3DVAR-alone analysis; (c) the EnKF-alone analysis; and (d) the hybrid EnKF-3DVAR analysis for the single radar experiment. The time shown is at 100 min (the end of data assimilation cycles). Wind vectors are shown every 2 km.

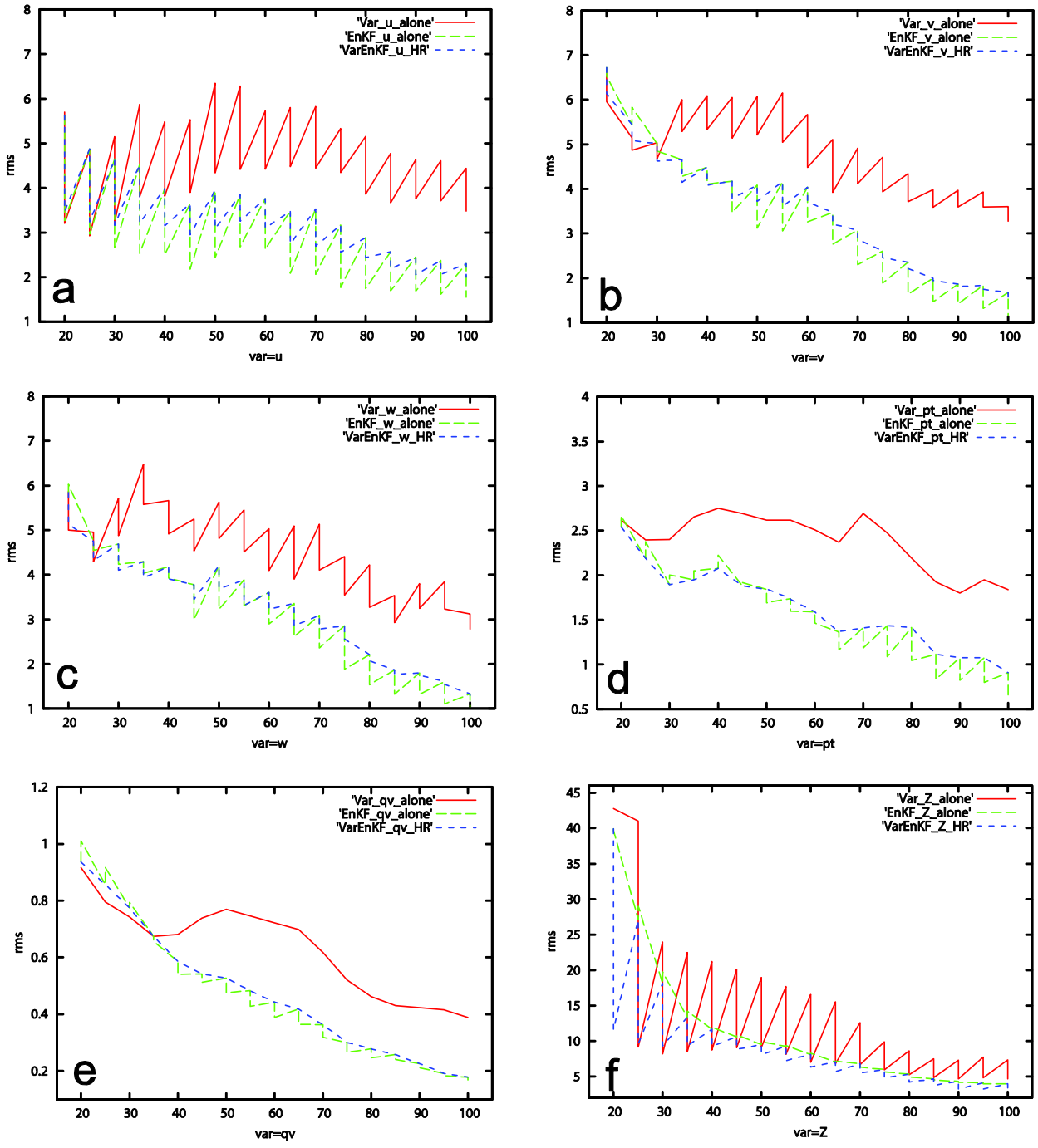

Fig. 3. The rms errors of the analysis and forecast for the 3DVAR-alone, EnKF-alone, high-resolution hybrid EnKF-3DVAR, and dual-resolution hybrid methods averaged over points at which the reflectivity is greater than 10 dBZ.for (a) u-wind component, (b) v-wind component, (c) vertical wind speed, (d) potential temperature, (e) water vapor mixing ratio, and (f) reflectivity.

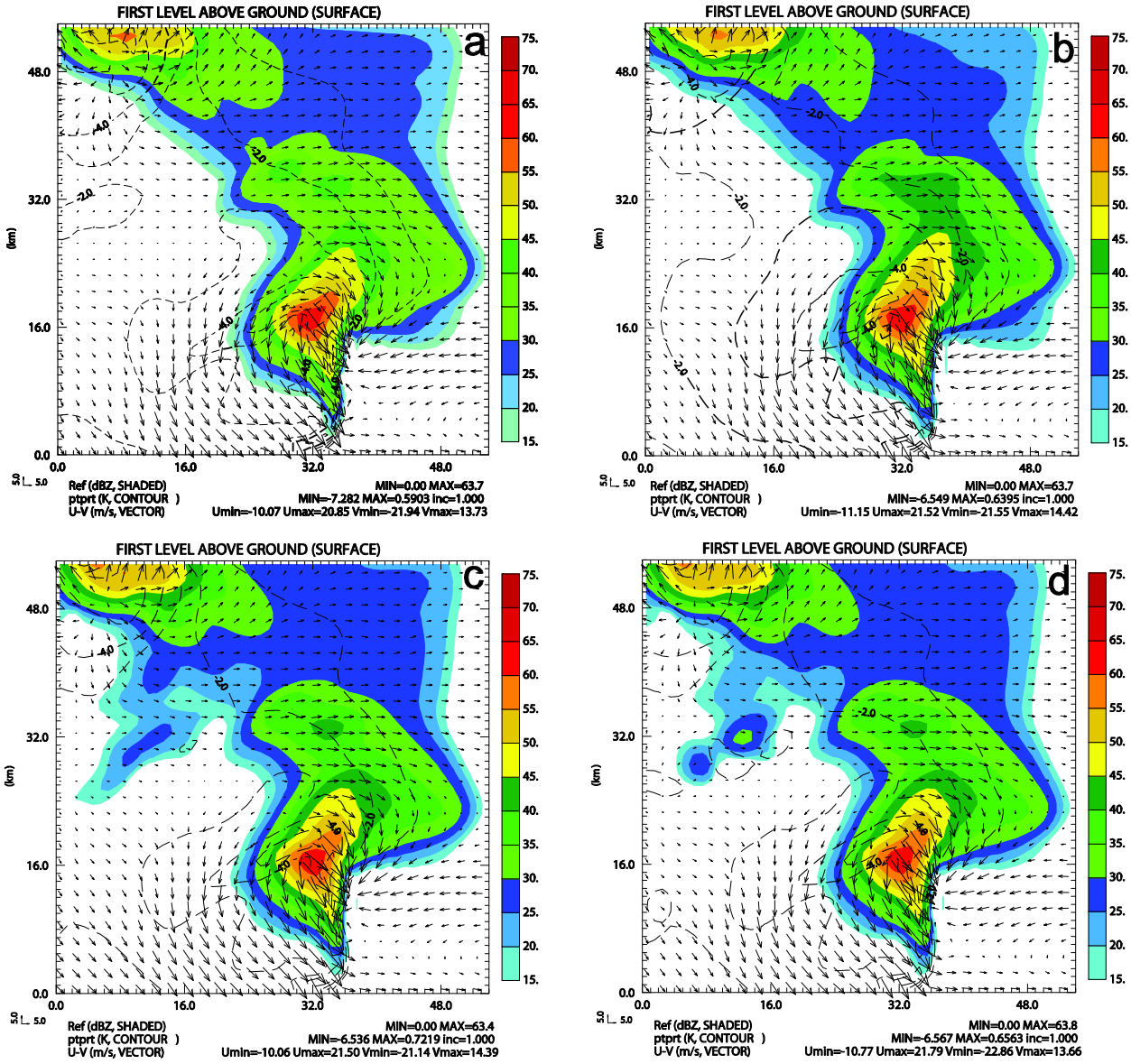

Fig. 4. Same as Fig. 2, but for the experiment with two radars

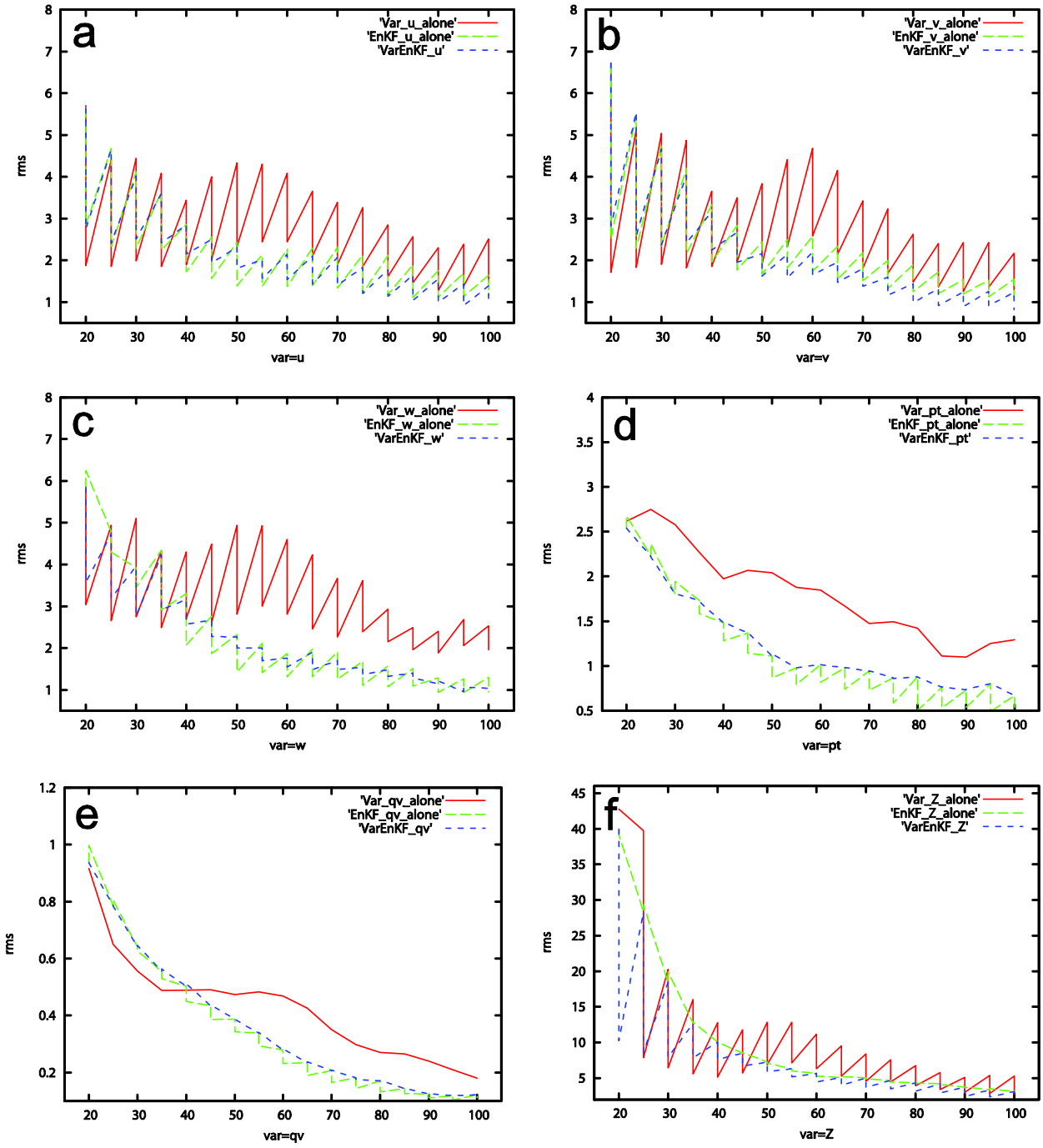

Fig. 5. Same as Fig. 3, but for the experiment with two radars.

5. Summary and future work

A preliminary version of a hybrid EnKF-3DVAR data assimilation system has been developed based on existing 3DVAR and ensemble Kalman filter (EnKF) programs using the ARPS model grid. The algorithm uses the extended control variable approach to combine the static and ensemble-derived flow-dependent forecast error covariances (Lorenc 2003; Buehner 2005; Wang et al. 2007). In the hybrid method, the relative weights assigned to the static and flow-dependent error covariances can be tuned, and the tuning can be case and scale dependent.

The method is applied to the assimilation of simulated radar data for a supercell storm. Two groups of experiments are performed with different amount of radar data. Results obtained using 3DVAR-alone (with static covariance entirely), Hybrid EnKF-3DVAR, and the standard EnKF-alone are compared. When data from a single radar are used, the EnKF-alone method provides the best results for model dynamic variables, but hybrid method provides the best results for hydrometeor related variables in term of rms errors. Though the storm structures can be established reasonably well using the 3DVAR-alone method, the rms errors are generally worse than other two methods. When data from two radars are used, the rms errors for hybrid method are the smallest for most of the model variables. With two radars, the results from the 3DVAR-alone are closer to that of the EnKF-alone method. For this idealized case, more weights for flow-dependent error covariances generally have better results. Also our tests indicate that the hybrid scheme can reduce the storm spin-up time because it fits the observations for the beginning of the cycles.

Our future studies will include answering a number of keys questions within the hybrid EnKF-3DVAR framework just described. They include 1) What is the optimal choice of the relative weight of the static and flow-dependent covariances, at the storm scale for radar data assimilation? 2) What is the benefit of equation constraints in the hybrid system and how is the error covariance for these equation constraints estimated from ensemble forecasts? 3) What is the optimal combination of ensemble size and grid spacing for a specific computational cost? 4) How does the overall performance of the proposed method compare with 3DVAR-alone and EnKF-alone methods when model error appears? 5) What is the general cost of the proposed method compared with the other two methods? The answers to these questions at storm-scale are likely to be quite different from those for large-scale flows and/or conventional observation networks where state variables are usually directly observed but the observations tend to be spare compared to model grid spacing. Sensitivity experiments will be performed to answer these questions in the near future.

References

Aksoy, A., D. C. Dowell, and C. Snyder, 2009: A multi-case comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part I: Storm-scale analyses. *Mon. Wea. Rev.,* **137,** 1805-1824.

Andersson, E., and coauthors, 1998: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). III: Experimental results. *Quart. J. Roy. Meteor. Soc*., **124**, 1831-1860.

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

Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129,** 420-436.

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

Brewster, K. A., 2003a: Phase-correcting data assimilation and application to storm-scale numerical weather prediction. Part I: Method description and simulation testing. *Mon. Wea. Rev.*, **131,** 480-492.

Brewster, K. A., 2003b: Phase-correcting data assimilation and application to storm scale numerical weather prediction. Part II: Application to a Severe Storm Outbreak. *Mon. Wea. Rev.*, **131,** 493-507.

Buehner, M., 2005: Ensemble-derived stationary and flow-dependent background-error covariances: evaluation in a quasi-operational NWP setting. *Quart. J. Roy. Meteor. Soc.*, **131**, 1013-1043.

Buehner, M., P. L. Houtekamer, C. Charette, H. L. Mitchell, B. He, 2010a: Intercomparison of Variational Data Assimilation and the Ensemble Kalman Filter for Global Deterministic NWP. Part I: Description and Single-Observation Experiments. *Mon. Wea. Rev.,* **138**, 1550-1566.

Buehner, M., P. L. Houtekamer, C. Charette, H. L. Mitchell, B. He, 2010b: Intercomparison of Variational Data Assimilation and the Ensemble Kalman Filter for Global Deterministic NWP. Part II: One-Month Experiments with Real Observations. *Mon. Wea. Rev*., **138**, 1567-1586.

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

Courtier, P., E. Andersson, W. Heckley, J. Pailleux, D.

Vasiljevic, M. Hamrud, A. Hollingsworth, F. Rabier, M. Fisher, 1998: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. *Quart. J. Roy. Meteor. Soc*., **124**, 1783-1807.

Dowell, D., F. Zhang, L. J. Wicker, C. Snyder, and N. A. Crook, 2004: Wind and temperature retrievals in the 17 May 1981 Arcadia, Oklahoma supercell: Ensemble Kalman filter experiments. *Mon. Wea. Rev.*, **132**, 1982-2005.

Dowell, D., L. J. Wicker, C. Snyder, 2010: Ensemble Kalman Filter Assimilation of Radar Observations of the 8 May 2003 Oklahoma City Supercell: Influences of Reflectivity Observations on Storm-Scale Analyses. *Mon. Wea. Rev. (Accepted).*

Derber, J. C., and A. Rosati, 1989: A global oceanic data assimilation system. *J. Phys. Oceanogr.*, **19**, 1333-1347.

Derber, J. C., D. F. Parrish and S. D. Lord, 1991: The new global operational analysis system at the National Meteorological Center. *Wea. And Forecasting*, **6**, 538-547.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99 (C5),** 10143-10162.

Ferrier, B. S., 1994: A double-moment multiple-phase four-class bulk ice scheme. Part I: Description. *J. Atmos. Sci.,* **51,** 249–280.

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., M. Xue, A. Shapiro, Q. Xu, and K. K. Droegemeier, 2001: Three-dimensional simple adjoint velocity retrievals from single Doppler radar. *J. Atmos. Ocean Technol.*, **18**, 26-38.

Gao, J., M. Xue, K. Brewster, and K. K. Droegemeier 2004: A three-dimensional variational data assimilation method with recursive filter for single-Doppler radar, *J. Atmos. Oceanic. Technol.* **21**, 457-469.

Gao, J. and M. Xue, 2008: An efficient dual-resolution approach for ensemble data assimilation and tests with assimilated Doppler radar data. *Mon. Wea. Rev.* **136**, 945-963.

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., and J. Gao, 2007: Latest development of 3DVAR system for ARPS and its application to a tornadic supercell storm, *22nd Conference on Weather Analysis and Forecasting/18th Conference on Numerical Weather Prediction*, on-line publication, 2B.6.

Ge G., J. Gao, K. A. Brewster, and M. Xue 2010: Effects of beam broadening and earth curvature in radar data assimilation. *J. Atmos. Oceanic. Technol.* **27**, 617-636.

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

Honda, Y., and K. Koizumi, 2006: The impact of the assimilation of precipitation data and Radar reflectivity with a pre-operational 4DVAR for the JMA nonhydrostatic model, *10th Symp. Integrated Obs. Assim. Sys. Atmos., Oceans, Land Surface (IOAS-AOLS)*, CDROM P2.1.

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

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

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.

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.

Kong F., K. K. Droegemeier, and N. L. Hickmon, 2006: Multiresolution ensemble forecasts of an observed tornadic thunderstorm system. Part I: Comparison of coarse and fine-grid experiments. *Mon. Wea. Rev.*, **134**, 807–833.

Kong, F. M. Xue, K. W. Thomas, J. Gao, Y. Wang, K. Brewster, K. K. Droegemeier, J. Kain, S. Weiss, D. Bright, M. Coniglio, and J. Du, 2009: A realtime storm-scale ensemble forecast system: 2009 spring experiment. 23nd Conf. Wea. Anal. Forecasting/19th Conf. Num. Wea. Pred., Omaha, NB, Amer. Meteor. Soc., Paper 16A.3.

Lin, Y., P. S. Ray, and K. W. Johnson, 1993: Initialization of a modeled convective storm using Doppler radar-derived fields. *Mon. Wea. Rev*., **121**, 2757-2775.

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

Mitchell, H. L., P. L. Houtekamer and G. Pellerin. 2002: Ensemble Size, Balance, and Model-Error Representation in an Ensemble Kalman Filter**.** *Mon. Wea. Rev.* **130**, 2791–2808.

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

Purser, J., W-S Wu, D. Parrish and N. M. Roberts, 2003a: Numerical aspects of the application of recursive filter to variational statistical analysis. Part I: spatially homogeneous and isotropic Gaussian covariances. *Mon. Wea. Rev.*, **131,** 1524-1535**.**

Purser, R. J., W.-S. Wu, D. F. Parrish, and N. M. Roberts, 2003b: Numerical aspects of the application of recursive filters to variational statistical analysis. Part I: Spatially homogeneous and isotropic Gaussian covariances. *Mon. Wea. Rev.*, **131**, 1524-1535.

Rabier, F., A. McNally, E. Andersson, P. Courtier, P. Unden, J. Eyre, A. Hollingsworth, F. Bouttier, 1998: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). II: Structure functions. *Quart. J. Roy. Meteor. Soc*., **124**, 1809-1829.

Rabier, F., 2005: Overview of data assimilation developments in numerical weather prediction centers. *Abstract, 4th WMO Int. Symp. Assimilation Obs. Meteor. Ocean.*, Prague, Czech Republic.

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.

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.

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

Stensrud D. J. and J. Gao, 2010: Importance of horizontally inhomogeneous environmental initial conditions to ensemble storm-scale radar data assimilation and very short range forecasts. *Mon. Wea. Rev.*, **138**, 1250-1272.

Sun, J. and N. A. Crook, 1997: Dynamical and microphysical retrieval from Doppler radar observations using a cloud model and its adjoint. Part I: Model development and simulated data experiments. *J. Atmos. Sci.,* **54**, 1642-1661.

Sun, J. and N. A. Crook, 1998: Dynamical and microphysical retrieval from Doppler radar observations using a cloud model and its adjoint. Part II: retrieval experiments of an observed Florida convective storm. *J. Atmos. Sci*., **55**, 835-852.

Sun, J. and N. A. Crook, 2001: Real-time low-level wind and temperature analysis using single WSR-88D data. *Wea. and Forecasting*., **16**, 117-132.

Sun, J., 2005: Initialization and numerical forecasting of a supercell storm observed during STEPS**.** *Mon. Wea. Rev.***, 133,** 793-813**.**

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.

Wang, X., C. Snyder, and T. M. Hamill, 2007: On the theoretical equivalence of differently proposed ensemble/3D-Var hybrid analysis schemes. *Mon. Wea. Rev*. **135**, 222-227.

Wang X., D. M. Barker, C. Snyder, T. M. Hamill, 2008a: A Hybrid ETKF–3DVAR Data Assimilation Scheme for the WRF Model. Part I: Observing System Simulation Experiment. *Mon. Wea. Rev.*, **136**, 5116-5131.

Wang X., D. M. Barker, C. Snyder, T. M. Hamill, 2008b: A Hybrid ETKF–3DVAR Data Assimilation Scheme for the WRF Model. Part II: Real Observation Experiments. *Mon. Wea. Rev.*, **136**, 5116-5131.

Weygandt, S.S., A. Shapiro and K.K. Droegemeier, 2002a: Retrieval of initial forecast fields from single-Doppler observations of a supercell thunderstorm. Part I: Single-Doppler velocity retrieval. *Mon. Wea. Rev,* **130**, 433-453.

Weygandt, S.S., A. Shapiro and K.K. Droegemeier, 2002b: Retrieval of initial forecast fields from single-Doppler observations of a supercell thunderstorm. Part II: Thermodynamic retrieval and numerical prediction. *Mon. Wea. Rev,* **130**, 454-476.

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

Wu, W.-S., R. J. Purser, and D. F. Parrish, 2002: Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.,* **130**, 2905-2916.

Xiao, Q., Kuo, Y., Sun, J., Lee, W., Lim, E., Guo, Y., Barker, D. M., 2005: Assimilation of Doppler radar observations with a regional 3D-VAR system: impact of Doppler velocities on forecasts of a heavy rainfall case. *Journal of Applied Meteorology*, **44**, 768-788.

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. Physics*, **75,** 161-193.

Xue, M., and Coauthors, 2001: The Advanced Regional Prediction System (ARPS) - A multiscale nonhydrostatic atmospheric simulation and prediction tool. Part II: Model physics and applications. *Meteor. Atmos. Physics*, **76**, 143-165.

Xue, M., F. Kong, K. Thomas, J. Gao, Y. Wang, K. Brewster, K. Droegemeier, J. S. Kain, S. J. Weiss, D. Bright, M. C. Coniglio, and J. Du, 2008: CAPS Realtime Storm-scale Ensemble and High-resolution Forecasts as Part of the NOAA Hazardous Weather Testbed 2008 Spring Experiment. *24th Conference on Severe Local Storms*, AMS, Savannah, GA, AMS.

Yussouf, N., and D. J. Stensrud, 2010: Impact of high temporal frequency phased array radar data to storm-scale ensemble data assimilation using observation system simulation experiments. *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.

Zupanski, M., 2005: Maximum likelihood ensemble filter: Theoretical aspects**.** *Mon. Wea. Rev.* **133**, 1710–1726.