An Examination of WRF 3DVAR Radar Data Assimilation on Its Capability in Retrieving Unobserved Variables and Forecasting Precipitation through Observing System Simulation Experiments

SOICHIRO SUGIMOTO
Central Research Institute of Electric Power Industry, Chiba, Japan

N. ANDREW CROOK,* JUANZHEN SUN, AND QINGNONG XIAO
National Center for Atmospheric Research, Boulder, Colorado

DALE M. BARKER
Met Office, Exeter, Devon, United Kingdom, and National Center for Atmospheric Research, Boulder, Colorado

(Manuscript received 24 October 2008, in final form 5 April 2009)

ABSTRACT

The purpose of this study is to investigate the performance of 3DVAR radar data assimilation in terms of the retrievals of convective fields and their impact on subsequent quantitative precipitation forecasts (QPFs). An assimilation methodology based on the Weather Research and Forecasting (WRF) model three-dimensional variational data assimilation (3DVAR) and a cloud analysis scheme is described. Simulated data from 25 Weather Surveillance Radar-1988 Doppler (WSR-88D) radars are assimilated, and the potential benefits and limitations of the assimilation are quantitatively evaluated through observing system simulation experiments of a dryline that occurred over the southern Great Plains. Results indicate that the 3DVAR system is able to analyze certain mesoscale and convective-scale features through the incorporation of radar observations. The assimilation of all possible data (radial velocity and reflectivity factor data) results in the best performance on short-range precipitation forecasting. The wind retrieval by assimilating radial velocities is of primary importance in the 3DVAR framework and the storm case applied, and the use of multiple-Doppler observations improves the retrieval of the tangential wind component. The reflectivity factor assimilation is also beneficial especially for strong precipitation. It is demonstrated that the improved initial conditions through the 3DVAR analysis lead to improved skills on QPF.

1. Introduction

In recent years, there has been an increased demand for improved forecasting of severe convective weather and its associated hazards. Although the general skill of the numerical weather prediction is being steadily improved thanks to the increased model resolution and enhanced data assimilation systems, a very low skill for quantitative precipitation forecasts (QPFs) persists especially in the warm season (Fritsch and Carbone 2004). One of the causes for the low QPF skill is the inevitable model spinup associated with a poor specification of the initial state of the atmosphere at small scales. Such a deficiency often leads to an inaccurate simulation of the timing, location, and intensity of convective systems. Therefore, better initial conditions through assimilating mesoscale observations are considered vital for improving short-range QPF.

Among the existing mesoscale observation platforms, Doppler radar has a significant advantage in that it can make millions of measurements of precipitation and wind fields with a spatial resolution of a few kilometers and a temporal resolution of a few minutes. Such an unrivalled
feature of the spatial and temporal resolutions has great potential for the improvement of short-range QPF. Starting from the pioneering works by Sun and Crook (1997) and Sun and Crook (1998), the use of four-dimensional variational data assimilation (4DVAR) with radar data has been demonstrated to be promising for cloud-resolving models (Sun 2005; Kawabata et al. 2007; Sun and Zhang 2008). Recently, the potential of an ensemble Kalman filter (EnKF) technique (e.g., Snyder and Zhang 2003; Dowell et al. 2004; Zhang et al. 2004) has also been explored as another feasible methodology for assimilation of radar data.

In spite of the achievement in the above-mentioned 4D data assimilation research, there is a demand to investigate a methodology for assimilating radar data based on three-dimensional variational data assimilation (3DVAR) mainly because of its practicability in terms of computational efficiency. Many developmental efforts on 3DVAR have been put forth in operational centers as well as in research institutes. However, the efforts seem to be directed mainly toward large-scale and long-term forecasting in the past. The inclusion of 3DVAR data assimilation of radar data (radial velocity and reflectivity factor) into regional mesoscale models was not paid much attention until recently. Noticeable studies are found as follows. Lindskog et al. (2004) developed a method to assimilate Doppler radar wind data on the basis of the High-Resolution Limited-Area Model (HIRLAM) 3DVAR. Their method can handle radial wind superobservations that are spatially smoothed, or horizontal winds from velocity azimuth display (VAD) analysis. A benefit is found in 24-h wind forecasts compared with a case study in which no radar data is assimilated. Gao et al. (2004) developed a 3DVAR radar radial velocity assimilation algorithm for the Advanced Regional Prediction System (ARPS). It was used by Hu et al. (2006a,b) for a study to examine its impact on the prediction of a supercell thunderstorm. Xiao et al. (2005) developed a method for radial velocity observations on the basis of the Weather Research and Forecasting (WRF) 3DVAR system, and later the method was extended to include the assimilation of the reflectivity factor using a warm-rain partition technique (Xiao et al. 2007; Xiao and Sun 2007). The technique was tested with a few heavy rain events and found the improvement in short-range QPF.

It is commonly speculated that 3DVAR is not suitable for application in the convective scale because of its lack of convective-scale balance. The WRF 3DVAR scheme was designed with the assumption of hydrostatic and simplified (e.g., geostrophic) dynamics, so it may have limited ability in retrieving the unobserved variables by radar [i.e., the tangential (cross-beam) wind velocity, temperature, and microphysical variables]. However, this speculation has never been examined through controlled experiments such as Observation System Simulation Experiments (OSSEs). Although the aforementioned studies that used the WRF 3DVAR for radar data assimilation of real observations found the positive impact on precipitation forecasts, these studies did not explicitly examine the 3DVAR’s capability in retrieving the unobserved fields. The current study is designed to fill this gap through OSSEs by asking the following questions: 1) How capable is the 3DVAR technique in retrieving the unobserved cross-beam wind component as well as thermodynamical and microphysical variables by assimilating radar radial velocity and reflectivity? 2) If the 3DVAR technique has limited ability in retrieving the unobserved variables, will it still provide considerable impact on the improvement of QPF? 3) How important are the clear-air observations from the boundary layer to precipitation forecasts? The third question often arises when a radar network is incapable of observing the atmospheric boundary layer through clear-air reflectors (i.e., insects) due to either the lack of reflectors in certain region or the fact that a radar is operated in a mode that can better observe storm echoes. To answer these questions, we believe it is best to use the OSSEs because OSSEs provide a “true” measure not only for assessing the potential of a method but also for evaluating the sensitivity of the method. In this study, assimilation experiments are performed using the WRF 3DVAR radar data assimilation system originally developed by Xiao et al. (2005). A method to assimilate reflectivity factor data is developed, which is an extension of the initial work by Xiao et al. (2007).

This paper is organized in the following manner. Section 2 briefly describes the WRF 3DVAR system and its radar data assimilation methodology. In section 3, the configuration for WRF model simulations as well as OSSEs are described. Section 4 presents the results of the experiments. Conclusions and topics for future research are summarized in section 5.

2. WRF 3DVAR and its radar data assimilation methodology

a. Brief description of WRF 3DVAR

The WRF-Var is a variational data assimilation system designed and built for the WRF model including both 3DVAR and 4DVAR components. This system originated and evolved from the fifth-generation Pennsylvania State University–National Center for Atmospheric Research (PSU–NCAR) Mesoscale Model (MM5) 3DVAR that was described in Barker et al. (2004). Barker et al.
(2003) provided more details of 3DVAR’s technical aspect. Skamarock et al. (2005) described the latest improvements of both WRF model and WRF-Var. The basic interface of the variational system is now fully consistent with the WRF model, which enables the future application of 4DVAR. Some other significant progresses and additions are found in a supplemental technical note (Skamarock et al. 2005). Here, a few explanations are given to help understand the 3DVAR component of the WRF-Var system.

A prescribed cost function $J$ is defined to measure the difference between the model and observations:

$$J(x) = \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \frac{1}{2}(y - y^o)^T R^{-1}(y - y^o),$$

(1)

where $x$ is the vector of the analysis, $x_b$ the vector of the background (or first guess), $B$ is the background error covariance matrix, $y^o$ is the vector of the observation, $y$ is the vector of the model-derived observation transformed by the observation operator $H$ [i.e., $y = H(x)$], and $R$ is the observational and representative error covariance matrix. The basic goal of a variational system is to produce an optimal estimate of the atmospheric state at a given time through iteratively minimizing the cost function. The specification of the background and the observation errors is described in section 3.

In WRF-Var, the incremental formulation of the model-space cost function is adopted with preconditioning via a control variable transformation $x - x_b = U \nu$, where $\nu$ is the vector of the incremental control variable, and $U$ represents the transformation operator. The matrix $B$ is equivalent to $UU^T$ if $U$ is well designed. Preconditioning leads to efficient computation of $B^{-1}$ by choosing control variables so that their cross correlations are negligible. In this study, streamfunction, unbalanced velocity potential, unbalanced temperature, total water mixing ratio (the sum of water vapor, cloud water, and rainwater mixing ratios), and unbalanced surface pressure are used as control variables. The unbalanced control variables are defined as the difference between full and balanced (or correlated) components of the field. The balanced component is modeled via a regression analysis of the field using specific predictor fields [e.g., streamfunction; see Skamarock et al. (2005) and Wu et al. (2002) for further details]. The regression analysis results in statistical balances between the streamfunction and unbalanced velocity potential, unbalanced temperature, and unbalanced surface pressure, respectively.

Regarding the control variable transformation, a practical procedure is performed by a series of operations $U = U_p U_p U_p$, where $U_p$ is the physical variable transformation including the conversion of control variables to model variables and the transform via the statistical balances, $U_p$ is the vertical transformation via an empirical orthogonal function decomposition, and $U_b$ is the horizontal transformation by a recursive filter. Accordingly, three components are required to be estimated offline for dealing with $B^{-1}$ as follows:

Component 1: Regression coefficients to establish the balance relations between the streamfunction and the unbalanced control variables (used in $U_p$).

Component 2: Eigenvectors/eigenvalues (used in $U_p$).

Component 3: Recursive filter’s characteristic length scale for each control variable and for each vertical mode (used in $U_b$).

b. Assimilation of radial velocities

The observation operator for radial velocity $V_r$ from a Doppler radar observation is formulated with the 3D wind field $(u, v, w)$, the vertical fall speed of hydrometeor $V_t$ (currently related only to rainwater), and the distance $R$ between the location of a data point and the radar antenna:

$$V_r = \frac{1}{R}[(x_d - x_r)u + (y_d - y_r)v + (z_d - z_r)(w - V_t)],$$

(2)

$$R = \sqrt{(x_d - x_r)^2 + (y_d - y_r)^2 + (z_d - z_r)^2},$$

(3)

where $(x_d, y_d, z_d)$ represents the location of the data point and $(x_r, y_r, z_r)$ represents the location of the radar antenna. Note that $V_t$ is calculated from the rainwater mixing ratio with height correction following Sun and Crook (1997).

Model-derived radial velocities are calculated using this operator. To estimate the vertical velocity increments, a balance equation based on the Richardson equation (White 2002) is introduced in $U_p$. The derivation and details are described in Xiao et al. (2005).

c. Assimilation of reflectivity factor

1) OBSERVATION OPERATOR AND A PARTITIONING TECHNIQUE

The following equation is used as the observation operator for reflectivity factor (Sun and Crook 1997):

$$Z = 43.1 + 17.5 \log(\rho q_r),$$

(4)

where $Z$ is the reflectivity factor (dBZ), $\rho$ is the air density (kg m$^{-3}$), and $q_r$ is the rainwater mixing ratio (g kg$^{-1}$).
Such a relationship is derived assuming a Marshall–Palmer type of drop size distribution and no contribution of ice phases to the reflectivity factor.

In the OSSEs of this study, the radial velocity and reflectivity “observations” are taken from model simulations in the model space. Therefore, there is no need for an operator to transform the radial velocity and reflectivity factor from the observation space to the model space as required in a real data case.

As noted before, the total water mixing ratio \( q_t \) is used as a moisture control variable. Therefore, we need to introduce a microphysical scheme in \( \mathbf{U}_p \) operator to partition the increment of \( q_t \) into the increments of water vapor, cloud water, and rainwater mixing ratios. Four microphysical processes in a warm-rain regime are considered: condensation, autoconversion, accretion, and evaporation. In the condensation process, the isobaric process is assumed in the transformation between water vapor and cloud water. The same empirical equations as in Sun and Crook (1997) are used for the remaining processes. The reader is referred to Xiao et al. (2007) for more description of the partitioning process.

It is noteworthy that treatments for zero derivatives of the evaporation and the rainwater fall speed with respect to the rainwater mixing ratio are quite important for the stable minimization of the cost function. That is, a gradient with respect to the rainwater mixing ratio becomes very large for the empirical equations of the evaporation and the rainwater fall speed, when the rainwater mixing ratio is very small (close to zero). This restriction below the freezing level is due to the assumption of no contribution of ice phases to the reflectivity factor. Since the purpose of this cloud analysis is to turn on the switches for microphysical processes, it is only applied to extreme cases in which the background has no significant convection whereas convection is indicated by radar or vice versa.

There are three modifications to the background through the cloud analysis. The first modification is performed for the rainwater mixing ratio. If the observed rainwater mixing ratio on a model grid point is smaller than a threshold (e.g., \( 5.0 \times 10^{-5} \text{ kg kg}^{-1} \)), the rainwater mixing ratio in the background is replaced by the value of the observation. If the observation is over the threshold whereas the background is below, a blending is performed:

\[
q_r^{\text{new}} = BLq_r^{\text{bak}} + (1 - BL)q_r^{\text{obs}},
\]

where \( q_r^{\text{new}} \) represents the new rainwater mixing ratio, and \( q_r^{\text{bak}} \) and \( q_r^{\text{obs}} \) are the rainwater mixing ratios of the background and the observation, respectively. In this study, the value of BL is set to be 0.9.

After the rainwater first-guess field is obtained, the temperature field is then modified by an estimated latent heating associated with particle condensation. First, the following equation is applied to the model and observation rainwater mixing ratio data, respectively,

\[
\frac{\partial q_r}{\partial x} + v \frac{\partial q_r}{\partial y} + w \frac{\partial q_r}{\partial z} - \frac{\partial V}{\partial z} q_r = Q,
\]

to obtain the difference of \( Q \). Then, the differential latent heating is computed and inserted into the background to modify the temperature below the freezing level of each column in the background. Note that the restriction below the freezing level is due to the assumption of no contribution of ice phases to the reflectivity factor.

Finally, relative humidity (RH) and the cloud water mixing ratio \( q_c \) are modified. Such modifications are performed below the freezing level and above a level at a specified height [i.e., lifting condensation level (LCL) + 500 m height] of each column in the background. The 500-m height is added to the LCL height based on the preliminary sensitivity experiments and the basic concept that the application of the cloud analysis should be minimized only to switch on microphysical processes in

\(^1\) Note that in the real data case, the radar-derived rainwater mixing ratio needs to be interpolated to the model grid before the cloud analysis.
Latent heating from radar observations is used to check if the actual air is saturated or not. The RH is then modified in the following manner. If the air is saturated and the value of RH is smaller than LowRH (%), the value is reset to be LowRH. If the air is unsaturated and the value of RH is larger than HighRH (%), the value is reset to be HighRH. In this study, the values of LowRH and HighRH are set to be 70 and 90, respectively.

For cloud water $q_c$, its value is set to be zero if the air is unsaturated. Otherwise, $q_c$ is adaptively modified following the criteria listed below:

1. If $q_r > 1.0 \times 10^{-3}$ and $q_c < 5.0 \times 10^{-4}$, then $q_c = 5.0 \times 10^{-4}$. (The units are kg kg$^{-1}$.)
2. If $q_r > 7.0 \times 10^{-4}$ and $q_c < 3.5 \times 10^{-4}$, then $q_c = 3.5 \times 10^{-4}$.
3. If $q_r > 4.0 \times 10^{-4}$ and $q_c < 2.0 \times 10^{-4}$, then $q_c = 2.0 \times 10^{-4}$.
4. If $q_r > 1.0 \times 10^{-4}$ and $q_c < 5.0 \times 10^{-5}$, then $q_c = 5.0 \times 10^{-5}$.

In the above four cases, the values of $q_c$ thresholds are set to be QC ratio (%) (=50%) of $q_r$ thresholds.

We realize that these thresholds are quite arbitrary. However, considering the fact that the above cloud analysis is only used as the first guess for the 3DVAR assimilation, variations in the threshold’s specification should not have a significant impact on the final 3DVAR analysis. To confirm this statement, a number of preliminary experiments were conducted by using different threshold and parameter values. The results of experiments indicated the importance of rainwater blending. A value of 0.9 seems to be sufficient for the parameter BL. Thresholds for the modification of relative humidity (LowRH and HighRH) were relatively sensitive to the final 3DVAR analysis, suggesting that these thresholds should be carefully set not to modify too much. The above setting worked well for the objective of the cloud analysis. The retrieval was not sensitive to the threshold QC ratio for the modification of the cloud water. The impact of the cloud analysis was also evaluated by conducting two experiments: one with the cloud analysis and the other without. It was found that the cloud analysis improved the skill score of low precipitation amount at longer forecast range (figure not shown).

3. Description of OSSEs

a. WRF model configurations

The numerical model chosen for this study is the WRF-ARW model version 2.1 that is described in detail by Skamarock et al. (2005). The model configurations and parameterization schemes are given as follows. A single model domain is used with a horizontal grid spacing of 4 km and 36 full sigma levels in the vertical. The model top is located at the 50-hPa level. The geographic location for the grid can be seen in Fig. 1. The number of the horizontal grid points is $300 \times 300$. The model includes parameterizations of dynamical and physical processes that are important for mesoscale systems. No cumulus parameterization is applied so that convection is explicitly treated by a microphysical scheme. In this study, the WRF single-moment (WSM) microphysics 6-class scheme (Hong and Lim 2006) is used. This scheme accounts for the resolvable scale convection with explicit treatment of cloud water, rainwater, snow, cloud ice, and graupel. Other primary physics options include Yonsei University planetary boundary layer model (Hong et al. 2006), the Noah land surface model (Chen and Dudhia 2001), Rapid Radiative Transfer Model (RRTM) longwave radiation (Mlawer et al. 1997), and Dudhia shortwave radiation (Dudhia 1989).

The model is initialized using archived analyses from Eta Data Assimilation System on the Advanced Weather and Interactive Processing System (AWIPS) grid 212 provided by National Centers for Environmental Prediction (NCEP), from which the initial fields for model simulations are generated on the model grid by an interpolation using the WRF SI (standard initialization) routine. The NCEP AWIPS datasets are provided at 3-h intervals with a horizontal resolution of 40 km. The time-dependent lateral boundary conditions are produced by
linear interpolation in time. A time step of 15 s is used for model integration.

b. Setup of OSSEs

1) GENERAL DESCRIPTION

The OSSEs are performed for a severe storm case occurred over the southern Great Plains on 12–13 June 2002. Based on an analysis of this event by Wilson and Roberts (2006) using a variety of observations, the severe storms are associated with a dryline and an outflow boundary left behind from an earlier convective system.

The assimilation with WRF 3DVAR system requires both the background and observations. In our OSSEs, results from two different WRF simulations are used for the background and for deriving the observations, respectively. The only difference in the two WRF runs is the time of initialization. Figure 2 illustrates a time line for WRF runs and assimilation in OSSEs.

A “truth” field is first defined. The truth field is used for making radar observations (radial velocity and reflectivity factor) with the operators described in section 2, and to evaluate the performance of the assimilation experiments. The truth field is prepared from a simulation that is initialized at 1500 UTC 12 June 2002. This simulation will be referred to as the “truth run.” Next, the background is prepared from another simulation with the initial conditions at 1200 UTC 12 June 2002. The simulation will be referred to as the “background run.” The forecasts from the background run are also referred to as the forecasts without assimilation (“no-assimilation” case, see later). Finally, experiments with radar data assimilation are conducted and the 3DVAR analyses of the model variables are verified against those from the truth run.

2) SIMULATED OBSERVATIONS AND THEIR ERRORS

Radar observations for each radar site are simulated using results from the truth run. The actual radar locations of the Weather Surveillance Radar-1988 Doppler (WSR-88D) network are used, resulting in 25 radars within the model domain (Fig. 1). The calculation of the observations is done on each model grid point. That is, no geometric transformation between radar observation space and model space is considered.

To consider as realistic as possible the radar beam pattern, the radial coverage, and the minimum level of a detectable signal, observations are calculated only if all of the following conditions are met:

Condition 1—A model point is located within 200 km from a radar antenna.
Condition 2—A model point falls below the elevation angle of 20° from any of the 25 radar locations.
Condition 3—Radial velocity and reflectivity factor are obtained in stormy regions where reflectivity factor calculated from Eq. (4) is larger than 5 dBZ.

Note that in this study we assume that the radial velocities (clear-air echoes) can be obtained in nonstormy regions if a model point is located below 2 km above the ground, and within 100 km from a radar site. In the above definition, the precipitation mode of WSR-88D Volume Coverage Pattern (VCP) is basically considered. It should be also noted that the simulated radar data extracted have areas of multiple-Doppler coverage according to Fig. 1. For overlapping reflectivity observations, a data location check is performed in WRF-Var subroutine to avoid counting the same observations. If the distance between radar site and radar data is longer than another data, the reflectivity observation is not accounted for assimilation.

Errors in radial velocity observations are assumed to be random numbers with an unbiased normal distribution. The standard deviation is 1 m s\(^{-1}\). These errors are simulated using a random number generator based on the Box–Muller transformation (Box and Muller 1958). Then, the radial velocity observation is obtained by adding the error to the truth that is calculated using the fields from the truth run and the observation operator.
For the observations of the reflectivity factor, the same procedure is adopted except for the unit of dBZs.

3) ESTIMATION OF THE BACKGROUND ERROR STATISTICS

The performance of a variational system depends on the background error statistics, because the statistics contain important information about how an observation spreads in the model space and how the final analysis is physically balanced. The commonly used WRF 3DVAR background error statistics were derived for large-scale applications by the so-called National Meteorological Center [(NMC) now known as NCEP] method (Parrish and Derber 1992) using long (typically at least 3 months) time series of previous forecasts.

In this study, to estimate the NMC-based error statistics, a couple of 24-h forecasts (starting from 0000 UTC) and 12-h forecasts (starting from 1200 UTC) were performed every day for the period of a week (6–12 June 2002). Two forecasts valid at 0000 UTC are obtained during a week (7–13 June). The difference between the 24- and 12-h forecasts was then used to calculate the domain-averaged error statistics described in section 2a. Among the three components described in section 2a, eigenvectors–eigenvalues and length scales are rescaled in WRF 3DVAR to obtain more reasonable analyses. By trial-and-error we found that the rescaling factor of 1.0 for all variance scales and the factor of 0.3 for all length scales produced better analyses. The need of these rescaling factors is discussed by several authors. For example, it was indicated by Ingleby (2001) that the NMC-based background error variances seemed to be a little overestimated, and the spatial correlation scales tended to be large.

Most studies use a period of more than one month to compute the NMC-based background error statistics. In the current study, we use the one-week period because we have found that the regression coefficients tend to be too small for longer time window. Consequently, the impact on the temperature increment resulting from the radar data assimilation is very small. The general issue on how to generate the background error statistics for radar data assimilation is being studied by one of the coauthors of this paper. The results will be summarized in a separate paper.

In the OSSEs, it is assumed that we know the truth field, so that the three components of the background error covariance matrix are estimated by a statistical
analysis of the difference between the forecasts of the truth run and the background run valid at 0000 UTC 13 June 2002. Similar to the NMC-based background error statistics, we found that the rescaling factor of 0.9 for all variance scales and the factor of 0.4 for all length scales. We do not fully understand why tuning with these rescaling factors is still needed to obtain better results. Obviously, more studies are needed to find a rigorous methodology to determine the appropriate scaling factors for storm-scale data assimilation in the 3DVAR system.

We ran experiments using the error statistics derived by both methods and compared the results. No significant differences were found although the statistics based on the difference field between the truth and the background was slightly better. We then decided to use the OSSE-based background error statistics for the experiments that are presented in this paper in order to better demonstrate the sensitivity of the experiments.

c. Experimental design

As depicted in Fig. 2, all OSSEs are performed with a cold start. The experiments are divided into three sets. The first experiment is done to examine the performance of the WRF 3DVAR in terms of the retrieval of the 3D winds, temperature, and microphysical variables under an idealized situation in which the radial velocities are obtained on all model grid points within the assumed 200-km radar coverage range regardless of the value of the reflectivity factor. That is, the condition 3
defined above is ignored for the radial velocity observations. Note that the condition 3 is still used for the assimilation of the reflectivity factor. This benchmark experiment is referred to as case 0.

For the second set of experiments, radar data defined in section 3b(2) are applied. This set of experiments is conducted to examine the impact of the amount of data on the retrieval and subsequent forecast. These experiments are summarized below:

Case 1-1: Radial velocity and reflectivity data from all the radars in Fig. 1 are assimilated.
Case 1-2: Same as 1-1, but single-radar coverage data from the radars of KVNX, KTWX, KGLD, KAMA, and KDYX are assimilated.
Case 1-3: Same as 1-1, but only the radial velocity data are assimilated.
Case 1-4: Same as 1-1, but only the reflectivity data are assimilated.

In case 1-2, the selected five radars provide single-radar coverage of the storm with almost no overlapping observations (see Figs. 1 and 3), so that the benefit of dual-radar coverage can be examined. The experiment in case 0 and the experiments in case 1 assimilate data at 0000 UTC 13 June 2002 when the storm was developed into a mature stage. In the last set of experiments, the assimilation is performed at 2100 UTC 12 June 2002, when the storm is still in its earlier development stage. The purpose of this set of experiments is to test the impact on QPF at the early development stage and the impact of clear-air data when there are less precipitation returns and hence less amount of data. The last set of experiments is summarized below:
4. Results

a. The truth and the background runs

The forecasts of the hourly accumulated precipitation from 2100 UTC 12 June to 0100 UTC 13 June 2002 for the truth and the background runs are plotted in Fig. 3. The distributions of the NCEP stage IV analysis (Lin and Mitchell 2005) are also plotted for reference. The truth run, initialized at 1500 UTC without data assimilation, simulates the convective line with remarkable resemblance to the stage IV analysis. In contrast, the background run, initialized only 3 h earlier than the truth run, results in much weaker convection. In the truth field, convective cells have been initiated along a dryline at 2100 UTC. There is no significant convection, however, in the same region for the background run. Such a difference in the precipitation simulation between the two runs is striking given the fact that there is only a 3-h time lag of the model initialization, which is not uncommon in high-resolution simulations without data assimilation.

b. Retrieval experiments (case 0)

1) WIND RETRIEVAL

The quality of the 3DVAR analysis is evaluated by comparing the analysis increment "Analysis minus Background (A – B)" with the "Truth minus Background (T – B)" field. If the retrieval is successful, these two quantities should be close. The root-mean-square (RMS) error of the analysis is computed against the truth field, which is expected to reduce from that of the first guess/background.

Figure 4 shows a comparison of the west–east wind (u wind) component between T – B and A – B at the model level of 10. The vertical profiles of the RMS error in the analysis and the background are also plotted. The RMS error in the analysis is reduced significantly from that in the background. Comparing the A – B field (Fig. 4b) with the T – B field (Fig. 4a), it is clearly seen that the major features are retrieved. Similar results are observed for the north–south wind component (not shown). The retrieved vertical wind at the same level is shown in Fig. 5. The analysis misses many small-scale disturbances, resulting in the RMS error that only has a slight improvement over that of the first guess. This is
not surprising considering that the hydrostatic balance is implied in the 3DVAR formulation.

2) RETRIEVALS OF MICROPHYSICAL VARIABLES AND TEMPERATURE

Figures 6, 7, and 8 show comparisons between $T - B$ and $A - B$ at model levels of 5, 10, and 15 as well as the vertical profiles of the RMS error in the analysis and the background for mixing ratios of rainwater, water vapor, and cloud water, respectively. The error in the rainwater mixing ratio is reduced substantially. However, the errors in the water vapor and the cloud water are only reduced slightly. The water vapor variations within and surrounding the squall line are not retrieved (Fig. 7). Some clouds are obtained in the analysis, but most are missed. These results suggest that the 3DVAR has limited ability in retrieving the water vapor and the cloud water even with the adjustment of the background field through the cloud analysis scheme. However, we found the cloud analysis performed to modify the first-guess field improved the 3DVAR analysis. Other moisture observation platforms [e.g., satellite and Global Positioning System (GPS) ground station] are needed to improve the water vapor and cloud water analysis in 3DVAR.

Figure 9 shows the result of the temperature retrieval. Significant impact is found near the surface and in the upper-to-middle levels. The temperature increments in these levels result from the increments of the wind components, using a regression coefficient between streamfunction and unbalanced temperature. The positive impact shown in the temperature increment indicates that the estimated regression coefficients are apparently reasonable. Benefits are also found in the lower atmosphere. For example, the negative perturbation caused by the evaporation of rainwater is recovered well in $A - B$, resulting mainly from the assimilation of the reflectivity factor data.

The above analysis shows that the 3DVAR has certain abilities in the partial retrieval of the unobserved fields of wind, thermodynamics, and microphysics. Among all the model variables, the horizontal wind received the most benefit from the radar data assimilation through the correction of the radial component and the retrieval of the tangential component (shown by Fig. 11 and explained in section 4c). The potential temperature field also obtained significant error reduction via the assimilation. The microphysical variables, except for the rainwater field that is observed by radar, had the least impact.

c. Benefits of multiple-Doppler observations (case I-1 versus case I-2)

The above experiment uses all the radars within the analysis domain. From Fig. 1, it is seen that these radars
have dual- or multiple-Doppler coverage areas because the detectable ranges from the radars overlap in some regions. One may wonder what the retrieval capability of the 3DVAR is if there is only single-Doppler coverage in the domain. To answer that question, the experiment 1–2 is conducted to compare WRF 3DVAR’s ability in multiple and single Doppler radar retrieval.

Scatterplots in terms of the radial and tangential components of wind are displayed in Figs. 10 and 11, respectively, for the background, analysis by multiple-Doppler retrieval (case 1-1), and analysis by single-Doppler retrieval (case 1-2). The comparison is made in stormy regions within the area covered by the KVNX radar. The background or analysis is compared with the truth field in each scatterplot. Since the radial velocity is the observed variable, the plots in Fig. 10 actually represent the fitting of the background or analysis field to the “observations.” It is clearly shown that in both experiments the radial velocity calculated from the final analysis fits well to the radial velocity derived from the truth run. The fitting in both experiments is much improved from the background. However, a significant difference in the retrieved tangential wind component (cf. Figs. 11b,c) is shown between the two experiments. Comparing Figs. 11a,c, we found that the single-Doppler assimilation improves the tangential wind over the background, but the improvement is not significant, indicating that the single-Doppler assimilation has a limited ability in retrieving the tangential wind component. As is expected, the use of multiple radars (Fig. 11b) significantly improves the retrieval of the tangential wind.

d. Impact of assimilation on precipitation forecasting

In this section, we investigate the impact of 3DVAR data assimilation on short-range QPF. Figure 12 shows the forecasts of 1-h accumulated precipitation for the experiments 1-1, 1-3, and 1-4 along with the results from the truth and the background runs from 0100 to 0600 UTC 13 June 2002. The RMS error of the 1-h accumulated precipitation and the threat score for 1 mm threshold are shown in Fig. 13. First of all, forecasts of case 1-1 indicate that the evolution of convective system is forecasted much better, compared with the background run (i.e., “no-assimilation run”). This suggests that the assimilation of all possible data can achieve the most significant improvement in precipitation forecasting.

When either reflectivity factor or radial velocity is excluded (cases 1-3 and 1-4), the impact of data assimilation becomes smaller. Particularly, the impact of assimilating reflectivity factor is rather small in case 1-4; the precipitation forecasts do not have much difference from
those in the background run. That indicates the impact of improved temperature and microphysical fields rapidly diminish because of the inaccurate wind field. Therefore, in our 3DVAR framework and the storm case applied, the assimilation of radial velocities is of primary importance in improving precipitation forecasts within the time range of 6 h or so. By comparing cases 1-1 and 1-3 in Fig. 13, we also found that although the assimilation of reflectivity factor alone does not have much effect on the precipitation forecasts, its assimilation together with the radial velocity clearly demonstrates the positive impact.

Quantitative verification result shown in Fig. 13 supports the qualitative evaluation described above. Substantial improvement is found in lead times up to 6 h after the radar data assimilation except for case 1–4. In case 1–4 when only reflectivity factor data are assimilated, the threat score for 1-mm precipitation drops down rapidly after the lead time of 1 h and becomes lower than the no-assimilation experiment. The threat scores with respect to precipitation amount for the second hour forecasts are plotted in Fig. 14. Obviously, the 3DVAR data assimilation improves the forecasts of both weak and strong precipitation, and more benefit is shown for the strong precipitation by the additional assimilation of the reflectivity factor data.

From the above experiments, we recommend that a caution must be exercised in assimilating radar reflectivity alone without the radial velocity data and using techniques with less dynamical constraints such as 3DVAR and Newtonian nudging. Sun (2005) showed that the rainwater increment resulting from the assimilation of reflectivity tends to fall to the ground and causes
spurious convection without appropriate dynamical support from the assimilation of the radial velocity. Sun (2005) also demonstrated that when a 4DVAR technique was used the problem associated with the assimilation of reflectivity alone could be remedied by using a longer assimilation window to allow for the information propagating to other variables.

Figure 15 shows forecasts of 1-h accumulated precipitation for the experiment in case 2-1 and for the truth and the background runs from 2200 UTC 12 June to 0300 UTC 13 June 2002. Note that at the data assimilation time (2100 UTC) the convective system just began to develop. The background run does not simulate the system in the first few hours because of the need for model spinup. The experiment that includes radar data assimilation simulates convective systems that have
better agreement with the truth at every forecast hour. The above qualitative evaluation is confirmed by the threat scores shown in Fig. 16.

We ran another experiment that includes no clear-air data assimilation (case 2-2). By comparing the forecasts with case 2-1, it is found that the additional assimilation of clear-air echoes does not improve the forecast of the convective system significantly. The clear-air data from radar partially observes the boundary layer convergence (with only radial velocity) that is believed to play a critical role in convective initiation and therefore is expected to have a substantial impact on the prediction of...

Fig. 12. Forecasts of 1-h accumulated precipitation for (a) the truth run, (b) the background run (no-assimilation case), (c) case 1-1, (d) case 1-3, and (e) case 1-4.
convective systems. Our current study using 3DVAR proved the positive impact, but not as much as one would expect.

5. Summary

This study investigates the performance of radar data assimilation in terms of the retrievals of convective fields and their impact on short-range QPF. This investigation is done through OSSEs with WRF 3DVAR system and simulated observations. The radar data assimilation scheme that was developed by Xiao et al. (2005, 2007) is extended to improve the reflectivity data assimilation.

Results of the retrieval experiments indicate that radar data assimilation works reasonably well in recovering key features of convection. Although a more sophisticated approach with 4DVAR or EnKF may provide more capabilities in retrieving the unobserved fields at the convective scale and hence lead to better convective weather forecast, it was shown in this study that the assimilation of radial velocity and reflectivity using 3DVAR resulted in the partial retrieval of the information on wind, thermodynamics, and microphysics. Our study showed that the horizontal wind received the most benefit from the radar data assimilation. The potential temperature field also obtained significant error reduction via the assimilation. The microphysical variables, except for the rainwater field that is observed by radar, had the least impact. Our experiments also showed that although the single-Doppler assimilation fits the radial component equally well to the multiple-Doppler assimilation, the capability in retrieving the tangential wind component differed substantially. Although the single-Doppler observations showed certain ability in retrieving the tangential wind, the use of multiple-Doppler observations clearly improved the retrieval ability.

The improved initial conditions through radar data assimilation resulted in better QPF. In most of the experiments, the positive impact on the precipitation forecasts was found up to 6 h after assimilation. The only exception happened when only reflectivity factor data are assimilated. In this case, the impact of improved temperature and microphysical fields rapidly diminish because of the inaccurate wind field, so that the positive impact was found up to 1 h or so after assimilation. The best performance was achieved when all possible radar data (radial velocity and reflectivity factor data) were assimilated. The wind retrieval was of primary importance for triggering and maintaining convective activities in our 3DVAR framework and the storm case applied. Adding the reflectivity factor data to the assimilation (along with the radial velocity) showed improved skill of QPF. However, the assimilation of reflectivity alone had small and mixed positive–negative impact on QPF. The assimilation of clear-air data that are observed in nonstormy regions is also beneficial. However, the positive impact is not as much as one

![Figure 13](image1.png)

**FIG. 13.** Verification scores of 1-h accumulated precipitation: (a) RMS error and (b) threat score.

![Figure 14](image2.png)

**FIG. 14.** Threat scores of 1-h accumulated precipitation for 2-h forecasts.
would expect. We believe the limited ability of the 3DVAR in retrieving the unobserved variables could have limited the utility of the data.

This study demonstrates that high-resolution assimilation using a 3DVAR technique has the potential to improve QPF. Specifically, we have shown that the improvement is associated with 3DVAR's ability in partially retrieving the unobserved wind component and temperature field through the assimilation of the radar radial velocity and reflectivity. However, this study also revealed some issues of the WRF 3DVAR radar data assimilation scheme that require further investigation. First of all, the retrieval of water vapor and cloud water were difficult even with the cloud analysis scheme. Other moisture observation platforms such as satellite and GPS ground station would be needed to improve the water vapor and cloud water analysis in 3DVAR. The second issue stems from the moisture partition procedure with the assumption of the warm-rain processes in WRF 3DVAR. This procedure should be further examined with the extension to include cold rain processes. The third issue is how to design the rapid update cycles (RUC) for radar data assimilation using 3DVAR. In this current study, we conducted our experiments with cold start. Work is under way to develop the best RUC strategies that are suitable for high-resolution data assimilation and short-range QPF. The final issue is essential for any variational data assimilation system, which is how we can estimate the background error statistics for high-resolution data assimilation. Our preliminary experiment suggests that the use of the

![Fig. 15. Forecasts of 1-h accumulated precipitation for (a) the truth run, (b) the background run (no-assimilation case), and (c) case 2-1.](image)

![Fig. 16. Threat scores of 1-h accumulated precipitation.](image)
case-specific statistics result in the better performance compared with the use of the NMC-based statistics. A methodology to estimate sophisticated (e.g., flow dependent) and adaptive background error statistics is desired. An ensemble approach or the combined use of EnKF would be promising to overcome the shortcomings of the NMC-based approach.

Acknowledgments. This study was performed during the first author’s long-term visit at the National Center for Atmospheric Research (NCAR). This paper is dedicated to the memory of our coauthor, respected colleague, and good friend N. Andrew Crook, who died in February 2006. We would like to acknowledge the reviewers for their helpful comments, which improved this paper. The authors are also grateful to Yong-Run Guo for his help in successfully running the WRF 3DVAR system.

REFERENCES


Xiao, Q., and J. Sun, 2007: Multiple-radar data assimilation and short-range quantitative precipitation forecasting of a


