An OSSE Framework Based on the Ensemble Square Root Kalman Filter for Evaluating the Impact of Data from Radar Networks on Thunderstorm Analysis and Forecasting

MING XUE, MINGJING TONG, AND KELVIN K. DROEGEMEIER

School of Meteorology and Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma

ABSTRACT

A framework for Observing System Simulation Experiments (OSSEs) based on the ensemble square root Kalman filter (EnSRF) technique for assimilating data from more than one radar network is described. The system is tested by assimilating simulated radial velocity and reflectivity data from a Weather Surveillance Radar-1988 Doppler (WSR-88D) radar and a network of four low-cost radars planned for the Oklahoma test bed by the new National Science Foundation (NSF) Engineering Research Center for Collaborative Adaptive Sensing of the Atmosphere (CASA). Such networks are meant to adaptively probe the lower atmosphere that is often missed by the existing WSR-88D radar network, so as to improve the detection of low-level hazardous weather events and to provide more complete data for the initialization of numerical weather prediction models.

Different from earlier OSSE work with ensemble Kalman filters, the radar data are sampled on the radar elevation levels and a more realistic forward operator based on the Gaussian power-gain function is used. A stretched vertical grid with high vertical resolution near the ground allows for a better examination of the impact of low-level data. Furthermore, the impacts of storm propagation and higher-volume scan frequencies up to one volume scan per minute on the quality of analysis are examined, using a domain of a sufficient size. The generally good analysis compared to earlier work indicates that the filter can effectively handle the non-uniform-resolution data on the radar elevation levels.

The assimilation of additional data from a well-positioned (relative to the storm) CASA radar improves the analysis of a supercell storm system that uses data from one WSR-88D radar alone; and the improvement is most significant at the low levels. When data from a single CASA radar are assimilated and when the radar does not provide full coverage of the storm system, significant errors develop in the analysis that cannot be effectively corrected. The combination of three CASA radars produces analyses of similar quality as the combination of one WSR-88D radar and one well-positioned CASA radar.

The most significant effect of storm propagation speed appears to be on the data coverage, which in turn affects the analysis quality. It is generally true that the more observations, the better the analysis. The results of the EnSRF assimilation are not very sensitive to the propagation speed. The quality of analysis can be improved by employing faster volume scans. The sensitivity of the EnSRF analysis to the volume scan interval is however much less than that of traditional velocity and thermodynamic retrieval schemes, suggesting the superiority of the EnSRF method compared to traditional methods. The very frequent update of the model state by the filter, even at 1-min intervals, does not show any negative effect, indicating that the analyzed fields are well balanced.

---

1. Introduction

A new National Science Foundation Engineering Research Center, the Center for Collaborative Adaptive Sensing of the Atmosphere (CASA), was established in

2003 to develop innovative observing systems for high-resolution sensing of the lower atmosphere. The development of low-cost, high-spatial density (also short range), and dynamically adaptive networks of Doppler radars with polarimetric capabilities is one of the first goals. Such networks are to probe the lower atmosphere that is often missed by the existing operational Weather Surveillance Radar-1988 Doppler (WSR-88D) radar network, so as to significantly improve the detection of low-level hazardous weather events such as tor-

Corresponding author address: Dr. Ming Xue, School of Meteorology, University of Oklahoma, SEC 1310, 100 E. Boyd, Norman, OK 73019.
E-mail: mxue@ou.edu

nadoes, and to provide more complete data for the initialization of storm-scale numerical weather prediction (NWP) models. A future upgrade to electronically steered phased-array antennas will permit even more dynamic scan strategies and collaborations among the network radars.

To help with the design and optimal operation of the first CASA radar test bed to be deployed in southwestern Oklahoma, and to examine the potential impact of the data from the test bed radars on storm-scale weather prediction through the assimilation of these data into the model initial conditions, an ensemble Kalman filter (EnKF)–based Observing Simulation System Experiment (OSSE) system is established based on a nonhydrostatic mesoscale weather prediction model. The system, when coupled with realistic radar emulators, permits a variety of OSSEs examining various issues related to optimal use of radar configurations and scan strategies.

The EnKF technique for data assimilation was first introduced by Evensen (1994). It has gained enormous popularity in recent years because of its simple formulation and relative ease of implementation compared to, for example, the four-dimensional variational data assimilation (4DVAR) method. Further, its computational requirements are comparable to sophisticated methods such as the 4DVAR method (Le Dimet and Talagrand 1986; Talagrand and Courtier 1987) and the representer method (Bennett 1992). The method has found applications in a number of fields, including meteorology and oceanography (see Evensen 2003 for a review). In the standard EnKF formulation, observations are treated as random variables that are subject to added perturbations (Burgers et al. 1998; Houtekamer and Mitchell 1998, 2001; Evensen 2003) so that the analysis error covariance is consistent with that of the traditional Kalman filter. Deterministic methods have been developed more recently to avoid sampling errors associated with the use of perturbed observations or to address the adaptive observational network design problem; these methods include the ensemble square root filter (EnSRF; Whitaker and Hamill 2002; Tippett et al. 2003), ensemble adjustment filter (EAKF; Anderson 2001), and ensemble transform Kalman filter (ETKF; Bishop et al. 2001), all of which belong to the broader class of square root filters (Tippett et al. 2003).

In general, the EnKF and related methods are designed to make possible the computation of flow-dependent error statistics. Rather than solving the equation for the time evolution of the probability density function of model state as in the traditional Kalman filter, EnKF methods apply the Monte Carlo method to estimate the forecast error statistics. An ensemble of model states is integrated forward in time using the dynamic equations, the moments of the probability density function are then calculated from this ensemble for different times (Evensen 1994, 2003).

Most recently, EnKF and its variation, the EnSRF, have been applied to the assimilation of simulated Doppler radar data for modeled convective storms (Snyder and Zhang 2003, hereafter SZ03; Zhang et al. 2004, hereafter ZSS04; Tong and Xue 2005, hereafter TX05) and to the assimilation of real radar data by Dowell et al. (2004). Encouraging results are obtained in these studies in analyzing the state variables for convective storms, even though none of these state variables are directly observed by the radar. SZ03 and ZSS04 assimilated only radial velocity data while Dowell et al. (2004) include reflectivity data, but their use is limited to the update of rainwater mixing ratio only. All three studies used an anelastic cloud model with warm-rain microphysics. Very recently, Caya et al. (2005) show through OSSEs that the similarly configured EnSRF and 4DVAR methods perform comparably. The 4DVAR method is able to produce better analyses within a 10–15-min assimilation window, while the EnSRF method does a better job after more assimilation cycles, when more volume scans are assimilated.

In TX05, the perturbed-observation EnKF method is used with a general-purpose compressible model that includes a multiclass ice microphysics parameterization. Different from SZ03 and ZSS04, both radial velocity and reflectivity data are assimilated and used to update all state variables. The study demonstrates, through OSSEs, the ability of EnKF in retrieving, from single-Doppler radar data, multiple microphysical species associated with a multiclass ice microphysics scheme, and in accurately retrieving the wind and thermodynamic variables as well. In general, the filter is able, after a number of assimilation cycles, to establish the model storm not present in the initial guess. The best results are obtained when both radial velocity and reflectivity data are used, even though the observation operator for reflectivity is highly nonlinear.

In this paper, an EnSRF (Whitaker and Hamill 2002)-based system is developed as an extension of the EnKF system of TX05 (because the EnSRF belongs to the general class of ensemble-based Kalman filters, we also use EnKF to refer to the filter used in this study). The system is used to assimilate simulated single- or multiple-radar data from the WSR-88D and/or the CASA networks. The single WSR-88D radar is located at Oklahoma City while the CASA radars are located in the Oklahoma test bed near Chickasha and Lawton, about 80–100 km southwest of Oklahoma City. Details on the assimilation system, the assumed configurations

of CASA radars, the simulation of observations, and the design of the OSSEs will be given in section 2. In section 3, the results of the OSSEs are reported that examine the value added by one well-positioned (relative to the storm location) CASA radar to the existing WSR-88D radar network, and the effectiveness in assimilating data from one or multiple CASA radars alone. The evaluation is performed for both quasi-stationary and fast-moving storm systems and for different volume scan frequencies. A general summary is given in section 4, together with some general discussions.

2. The OSSE system and EnKF analysis procedure

a. The prediction model and truth simulations

While our OSSE system is general and can be applied to any convective systems, in this paper, we (as in TX05) test our system and evaluate the data impact with the 20 May 1977 Del City, Oklahoma, supercell storm (Ray et al. 1981). Both simulation and analysis use the Advanced Regional Prediction System (ARPS) model (Xue et al. 2000, 2001, 2003), a fully compressible and nonhydrostatic atmospheric prediction system. The ARPS contains 12 prognostic state variables, including wind components u, v, w, potential temperature θ, pressure p, the mixing ratios for water vapor q_v, cloud water q_c, rainwater q_r, cloud ice q_i, snow q_s, and hail q_h, plus the turbulent kinetic energy (TKE) used by the 1.5-order subgrid-scale turbulent closure scheme. 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).

Different from TX05, a smaller horizontal grid spacing of 1.5 km is used in this study. Further, to better resolve the lower atmosphere, a vertically stretched grid with a minimum vertical resolution of 100 m is used, instead of the uniform 500-m vertical resolution that is used by TX05, SZ03, and ZSS04. The vertically stretched grid is defined using the hyperbolic tangent function, which is one of the built-in options in the ARPS (Xue et al. 1995). The model domain is 16 km deep with 40 physical layers. Two domain sizes are used (see Fig. 1); the smaller one has 47 × 47 horizontal points, which, excluding two boundary points, is 66 km on each side. The larger domain, needed to accommodate fast-moving storms, has 55 × 103 horizontal points and is 78 km × 150 km in physical size. The small domain is centered at 34.8°N and 98.1°W and the large domain is centered at 34.75°N and 98.11°W, and both use Lambert projection. The true latitudes of projec-
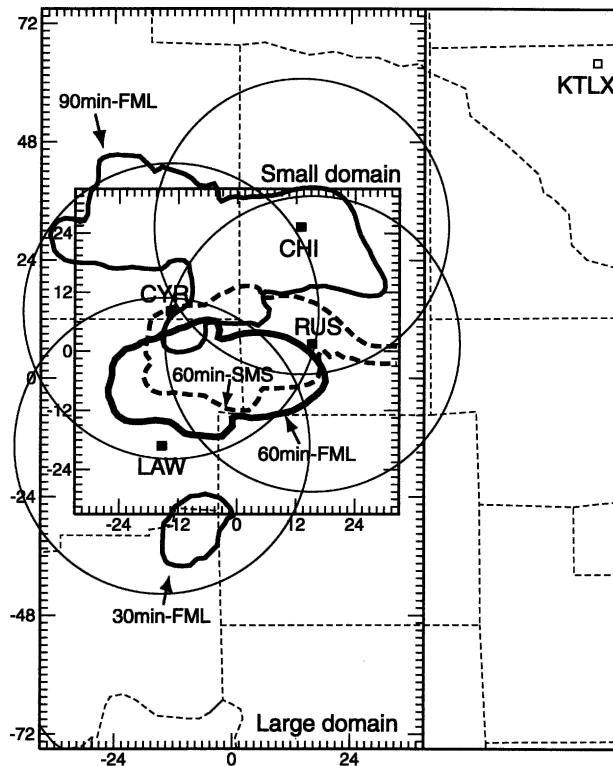

FIG. 1. A map with county borders showing the locations of the first four planned CASA Oklahoma test bed radars, near Chickasha (CHI), Rush Springs (RUS), Lawton (LAW), and Cyris (CRY), together with the Twin Lake (KTLX) WSR-88D radar near Oklahoma City. The assumed 30-km maximum range of CASA radars are shown by the low-level 30-km range circles, and the two (large and small) analysis and forecast domains are shown as square and rectangular boxes with axis tick marks and labels. The origin of both domains is set at the center of the domains. Also plotted are the 10-dBZ low-level (50 m AGL) reflectivity contours of truth simulations FML (thick solid contours, with that for 60 min the thickest) and SMS (thick dashed), at the times labeled in the figure. The 10-dBZ contours for SML are similarly located as those of FML at 60 min, and are therefore not shown.

tion are 30° and 60°N and the true longitude circle goes through the center of each grid. These domains roughly center on the four-radar Oklahoma CASA test bed.

The truth simulations or nature runs are initialized in the same way as in TX05. A sounding of 3300 J kg^{-1} CAPE (see Xue et al. 2001 for a skew T–log$p$ plot of the sounding) is used to define the environmental condition, and a 4-K ellipsoidal thermal bubble is used to initiate the storm. Three truth simulations were created; one for a slow-moving storm system with a quasi-stationary right-moving cell in the small domain (referred to as SMS), one for a fast-moving system in the large domain (FML), and one for a slow-moving system in the large domain (SML).

The center location of the bubble is at x = 16, 0, and

9 km, and $y = -12$, -62, and -12 km, respectively, for SMS, FML, and SML. In the vertical, the bubble is centered at $z = 1.5$ km. The origin of horizontal coordinates is located at the domain center for both grids. The initial location of the bubble for FML is chosen so that over most of the assimilation period, much of the storm system remains within the coverage of the four-radar CASA network (see Fig. 1). The choice for SML is such that at 60 min, the storm system is located at roughly the same location as that in FML so as to facilitate more direct comparisons.

A radiation condition is applied at the lateral boundaries. The top boundary is a rigid lid combined with a wave-absorbing layer; the lower boundary is free slip. For the experiments with a slow-moving supercell system, a constant wind of $u = 3$ m s^{-1} and $v = 14$ m s^{-1} is subtracted from the original sounding to keep the right-moving cell near the center of model grid, as is done in TX05. For the experiments with a fast-moving storm system, the original wind profile is used.

The time evolutions of the simulated storm intensities and structures are very similar. During each simulation, the strength of the initial cell that develops out of the initial bubble increases quickly over the first 20 min then decreases over the next 30 min because of the splitting of the cell into two at around 55 min (see Fig. 5). The right-moving (relative to the storm motion vector toward the north-northeast) cell tends to dominate the system; the updraft reaches a peak value of 56 m s^{-1} at 101 min. The left-moving cell starts to split again at 90 min. The initial cloud started to form at about 10 min, and rainwater formed shortly thereafter. Significant ice phase fields appeared at about 15 min. The general evolution of the storm is similar to that documented in Xue et al. (2001).

b. Simulation of radar observations

One WSR-88D and up to four CASA test bed radars are involved in this OSSE study. For the WSR-88D radar, standard precipitation-mode parameters are assumed, including 10-cm wavelength, 1° beamwidth, and a total of 14 elevations with the lowest elevation at 0.5° and the highest at 19.5°. The radial resolution is 250 m for radial velocity and 1 km for reflectivity (the radial resolution does not matter in this study because data are sampled in Cartesian coordinates in the horizontal, see later); the maximum range is assumed to be 230 km (which is sufficient to cover the entire computational domain). For the first four CASA radars to be installed in the Oklahoma test bed, whose primary goal is for the detection of low-level hazardous weather including tornadoes, and for improving weather prediction, the wavelength will be 3 cm (X band) and the beamwidth

will be about 2°. The maximum range will be 30 km. Because the CASA radars will be dynamically configurable in real time in response to weather situation and user needs, their scanning strategies will remain flexible. For the purpose of this study, we assume a total of 10 elevations at 2° increments, with the elevation of the center of the lowest beam at 1°. The impact of a variety of scanning strategies, including the vertical data coverage, on the quality of storm analysis will be the subject of future study.

The locations of the initial four Oklahoma test bed radars to be installed as early as fall 2005 are near Chickasha, Rush Springs, Lawton, and Cyril in Oklahoma, about 90 km to the southwest of the Twin Lakes (KTLX) WSR-88D radar near Oklahoma City. The locations of the four CASA radars are shown in Fig. 1, together with their assumed 30-km low-level range circles.

Different from the earlier OSSE studies of SZ03, ZSS04, and TX05, we assume that the simulated observations are available in the plan position indicator (PPI) planes or the elevation levels of the radar rather than at the model grid points. We do assume that the radar observations are already interpolated to the Cartesian coordinate in the horizontal directions; in other words, the observations are found in the vertical columns through the model scalar points. This assumption is reasonable because a horizontal interpolation to bring real radar data to the vertical columns is usually done before assimilating the data (e.g., in the 4DVAR work of Sun and Crook 2001). The main purpose of interpolation is to make the data distribution more uniform in the horizontal so as to avoid the cost of assimilating an excessive number of data at the close ranges of the radar. Still, we plan to examine in the future the effect of such horizontal interpolation by comparing with the analyses using data in native radar coordinates.

The effects of the curvature of the earth and the beam bending resulting from the vertical change of refractivity are taken into account by using the simple 4/3 effective earth radius model discussed in Doviak and Zrnic (1993); it is mainly the earth curvature effect that prevents the radars from seeing the lower atmosphere far away. The radar beams of one WSR-88D radar and one CASA radar located 90 km apart are illustrated in Fig. 2, together with the coverage by these beams up to their half-power width.

Because the observations are not at the model grid points, a forward observation operator is needed to bring the data from the model vertical levels to the radar elevation levels. This is achieved by using a simplified radar emulator that does power-gain-based sampling only in the vertical direction,
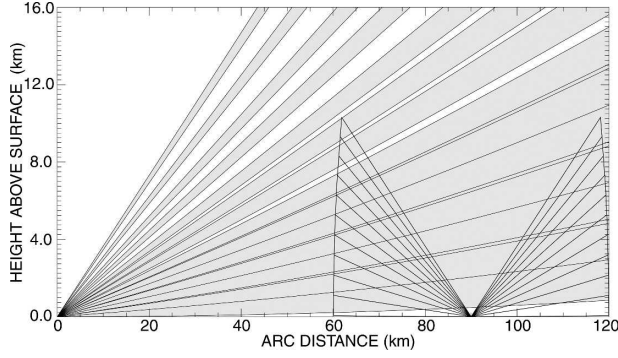
FIG. 2. The radial beams of a WSR-88D radar located at $x = 0$ and of an assumed CASA radar located 90 km away. The 1° beams of the WSR-88D radar are shaded, and are based on the WSR-88D volume coverage pattern 11 (VCP 11) in precipitation mode. The CASA radar has 2° beams with 2° elevation increments and its maximum range is assumed to be 30 km. There is slight overlapping of several lower-elevation beams for the WSR-88D radar.

$$\eta_e = \frac{\sum G \eta_g \Delta z}{\sum G \Delta z}, \tag{1}$$

where η_e and η_g are, respectively, the elevation level and gridpoint values of either radial velocity (V_r) or reflectivity (Z); Δz is the depth of the layer in which grid point value η_g is found. Equation (1) neglects the effect of reflectivity weighting for radial velocity observations. The power-gain function G is assumed to be Gaussian and has the form of

$$G = \exp\left[-4 \ln 4 \left(\frac{\phi_g - \phi_0}{\phi_w}\right)^2\right], \tag{2}$$

following Wood and Brown (1997), where ϕ_w is the beamwidth and is equal to 1° and 2° for WSR-88D and CASA radars, respectively. The elevation angle for the grid point value is ϕ_g and ϕ_0 the elevation at the beam center. A simpler and less accurate version of Eq. (2) has been used by Sun and Crook (2001). The power returns from sidelobes are not considered in Eq. (2).

For radial velocity, the gridpoint values involved in the numerator of Eq. (1) are first calculated from

$$V_{rg} = u \cos\phi_g \sin\gamma_g + v \cos\phi_g \cos\gamma_g + w \sin\phi_g, \tag{3}$$

where ϕ_g is the elevation angle (as defined earlier) and γ_g is the azimuth angle of the radar beam that goes through the given grid point; and u, v, and w are the model velocities interpolated to the scalar point of a staggered model grid. Subscript g denotes the gridpoint value. After V_r observations are sampled from the gridpoint values according to Eq. (1), random errors drawn

from a normal distribution with zero mean and standard deviation of 1 m s^{-1} are added as simulated observation errors. Because V_r is sampled directly from velocity fields, the effect of hydrometeor sedimentation does not come into play. When the effect is simulated, incorrect estimation of the terminal velocity will introduce errors into the assimilation system; this will be considered in our future work.

The simulated reflectivity Z in dBZ is calculated from the mixing ratios of rainwater, snow, and hail hydrometeors, using the same formulations as in TX05. Formally, it is given by

$$Z = Z(q_r, q_s, q_h). \tag{4}$$

The formulations mostly follow those of Smith et al. (1975) and are consistent with the ARPS ice microphysics. As with V_r, Z is first calculated at the grid points within the beamwidth then is transferred to the beam elevations using Eq. (1). Random errors of zero mean and standard deviations of 5 dBZ are then added to the simulated reflectivity data. The effect of attenuation is currently not considered but will be in the future.

When Eq. (1) is used to create simulated observations, no data are collected when no grid level is found within the beamwidth. This is equivalent to saying that data are discarded by the assimilation when the forward observation operator involves no grid levels within the beamwidth. For our vertically stretched analysis grids with high resolutions at the low levels, this does not happen often; when it does happen, the radar beam is usually very narrow and the atmosphere is, therefore, already well sampled. Our procedure is thus a natural way of thinning the data. For data sampling and for assimilation, we assume that the observation operators as well as the prediction model are perfect, an assumption that is used in all earlier radar OSSE studies. Model error will be an issue for future study.

c. The EnSRF data assimilation procedure

The EnSRF system used in this study is similar to that described in Tong and Xue (2005), but with several important differences. The first is that we now use an EnSRF instead of the perturbed observation method as discussed earlier. Relatively small differences between the two methods were reported in TX05 where 100 ensemble members were used. In this study, 40 members are used in all experiments and the EnSRF method is expected to perform better (Whitaker and Hamill 2002).

Following Whitaker and Hamill (2002), the serial EnSRF algorithm for analyzing uncorrelated observations, one after another, is summarized here. With the

serial analysis, the observations are analyzed one at a time. Therefore, the observation error covariance matrix $\mathbf{R}$ reduces to a scalar; so does matrix $\mathbf{HP}^b\mathbf{H}^T$, which is the background error covariance between observation points. The analysis equations for ensemble mean state $\bar{x}$ and the ensemble deviation from the mean x_i' are, respectively,

$$\bar{x}^a = \bar{x}^b + \mathbf{K}[y^o - H(\bar{x}^b)], \tag{5}$$

$$x_i'^a = \beta(\mathbf{I} - \alpha\mathbf{KH})x_i'^b, \tag{6}$$

where

$$\mathbf{K} = \mathbf{P}^b\mathbf{H}^T(\mathbf{HP}^b\mathbf{H}^T + \mathbf{R})^{-1} \tag{7}$$

is the Kalman gain matrix, $\mathbf{P}^b$ is the background or prior error covariance matrix, and $\mathbf{H}$ is the linearized version of the observation operator, H, that projects state variable x to observation y^o. Here, the superscripts, a, b, and o denote analysis, background, and observation, respectively. The ensemble mean analysis $\bar{x}^a$ is obtained first from Eq. (5); the deviation from the mean of the ith ensemble member is then given by Eq. (6), in which β is a covariance inflation factor that is usually slightly larger than 1, and

$$\alpha = [1 + \sqrt{\mathbf{R}(\mathbf{HP}^b\mathbf{H}^T + \mathbf{R})^{-1}}]^{-1}. \tag{8}$$

Equation (8) is only valid for single observation analysis and therefore both the numerator and denominator inside the square root are scalars and the evaluation of α is easy. In the above, the background error covariances $\mathbf{P}^b\mathbf{H}^T$ and $\mathbf{HP}^b\mathbf{H}^T$ are estimated from the background ensemble, according to

$$\mathbf{P}^b\mathbf{H}^T = \frac{1}{N-1}\sum_i^N (x_i^b - \bar{x}^b)[H(x_i^b) - \overline{H(x^b)}]^T, \tag{9}$$

$$\mathbf{HP}^b\mathbf{H}^T = \frac{1}{N-1}\sum_i^N [H(x_i^b) - \overline{H(x^b)}][H(x_i^b) - \overline{H(x^b)}]^T, \tag{10}$$

where N is the ensemble size, and H is the observation operator, which can be nonlinear (and is in this study). For a single observation, $\mathbf{P}^b\mathbf{H}^T$ is a vector having the length of vector x and $\mathbf{HP}^b\mathbf{H}^T$ is a scalar. In practice, because of covariance localization, all elements in $\mathbf{P}^b\mathbf{H}^T$ are not calculated; those outside the influence range of a given observation are assumed to be zero. After the analysis of one observation is completed, the analysis becomes the new background (x^a becomes x^b) for the next observation and the analysis is repeated. After all observations at a given time are analyzed, an ensemble of forecasts proceeds from the analysis ensemble until

the time of new observation(s); at that time the analysis cycle is repeated.

In our system, the analysis variables contained in state vector x include the gridpoint values of u, v, w, θ, p, q_v, q_c, q_r, q_i, q_s, and q_h. In this study, the Smagorinsky subgrid-scale turbulence option is used; therefore TKE is not needed. The observation vector y^o contains radar radial velocity V_r and reflectivity Z. The observation operator H is defined by Eqs. (1), (3), and (4), with the detailed formation of (4) given in TX05.

As in TX05, we start the initial ensemble forecast at 20 min of model time when the first storm cell developing out of an initial bubble reaches peak intensity. The ensemble is initialized by adding random perturbations to a horizontally homogeneous ensemble mean defined by the environmental sounding, and the random perturbations are sampled from Gaussian distributions with zero mean and standard deviations of 3 m s^{-1} for u, v, and w, 3 K for potential temperature θ, and 0.5 g kg^{-1} for water vapor mixing ratio q_v. The pressure and microphysical variables are not perturbed. Similar to TX05, we apply the initial perturbations to the entire domain except for the outermost five rings of grid points near the lateral boundaries. We do not perturb u, v, θ, and q_v at the first grid level about ground either; doing so was found to introduce pressure noise into the analysis.

For the standard set of experiments, observations are assimilated every 5 min. Here, the CASA radars are assumed to be operating in the traditional full volume scan mode and initially at a volume scan interval of 5 min. In other sets of experiments, CASA radar data at volume scan intervals of 1 or 2.5 min are assimilated and the impact of faster scanning is examined. For all experiments, the first analysis is performed at 25 min.

As mentioned earlier, 40 ensemble members are used. As in TX05, a covariance localization procedure following Houtekamer and Mitchell (2001) is employed that applies a Schur product of the background error covariance calculated from the ensemble and a correlation function with local support. This method multiplies each element of the background error matrix $\mathbf{PH}^T$ with a weight computed from the correlation function given by Eq. (4.10) of Gaspari and Cohn (1999). The weight decreases gradually from 1 at the observation point to zero at an effective cutoff radius and remains zero beyond. An effective cutoff radius of 6 km in all directions is used in this study rather than 8 km as in TX05, which used 100 ensemble members and a 2-km horizontal resolution. The best choice of cutoff radius is found through numerical experimentation.

Covariance inflation is necessary because of the typical underestimation of background error covariances

TABLE 1. List of small-domain assimilation experiments.

Experiment	Radars assimilated and configuration
TLX	KTLX radar data only, small domain, storm-motion subtracted, standard assimilation procedure
TLXCYR	As TLX, but with both KTLX and Cyril radars
TLXCYR3km	As TLXCYR, but with Cyril radar data limited to below 3-km level
CASA1S	As TLX, but with one CASA radar at Rush Springs
CASA3S	As CASA1S, but with three CASA radars at Cyril, Rush Spring, and Lawton

from the limited-sized ensemble. The procedure we use is based on that of Anderson (2001) with an important modification; instead of being applied everywhere, covariance inflation is limited to the grid points that are directly influenced during the analysis update by the observations found within the precipitation (where observed $Z > 10$ dBZ) regions. This modification is necessary to avoid amplifying spurious cells in precipitation-free regions. Snyder and Zhang (2003) also noted this problem and proposed an alternative inflation procedure, which was, however, found by Caya et al. (2005) to degrade their analysis. The inflation factor β we use is 1.07.

In TX05, it is found that the best analysis is obtained when, during the first four analysis cycles, reflectivity (Z) data are used to update only the variables that are directly related to reflectivity via the observation operator, that is, the mixing ratios of rainwater and snow, because of initially unreliable estimate of cross covariances with Z. In this study, we found that further delay in the use of Z to update indirectly related variables by five more minutes gives even better analyses. Therefore, in all experiments, the update of all variables when assimilating Z starts from the fifth analysis, that is, at 45 min. The problems with updating Z in the initial cycles are also noted by Caya et al. (2005). Unless noted otherwise, all reflectivity data, including data with no reflectivity, within the radar range are assimilated, while for radial velocity, only data in regions where $Z \geq 10$ dBZ are used. This was found to work best in TX05.

3. Assimilation of WSR-88D and CASA radar data and their impact

a. Impact of single CASA radar in addition to one WSR-88D radar

As stated in the introduction, one of the key problems with the existing national WSR-88D radar network is that the typical radar spacings of a few hundred kilometers preclude the observations of the lowest kilometers of the atmosphere at many locations because of the earth curvature effect and the nonzero elevation of the lowest beams. The problem is even worse in mountainous regions because of terrain blocking. Important low-level features such as tornados, cold pools, gust fronts, and downbursts are often missed by the network, and it is well established that low-level cold pools are very important for the support and maintenance of convective systems. The proposed CASA-type dense, inexpensive radar networks promise to significantly improve the low-level coverage.

In this section, we compare the analyses of the TLX-CYR, TLXCRY3km, and TLX experiments (Table 1). In TLXCYR, data from the CASA radar at the Cyril site are assimilated in addition to those from the KTLX radar. Experiment TLX used only KTLX data, while TLXCYR3km is the same as TLXCYR except that Cyril radar data that are assimilated are limited to the lowest 3 km. This latter case is likely to be more realistic because there is an initial plan to collect only data below the 3-km level by the CASA radars.

As can be seen from Fig. 2 at a 90-km range, the center of the lowest radar beam of KTLX is over 1 km above the ground, and the lower edge of the 1°-wide beam is about 500 m above ground, implying that the atmosphere below 500 m is not illuminated by the radar beam, and hence is not observed (this problem will be worse when the radar is farther away). The addition of CASA radars fills such gaps while at the same time may increase the data resolution in the covered region. In certain regions, the CASA radars may even provide dual- or multiple-Doppler wind coverage; in our case, such coverage is very limited partly because of the positioning of the limited precipitation regions (see Fig. 1). The total numbers of V_r and Z observations available and assimilated at selected times in each experiment will be given later (see Table 3).

The curves of the root-mean-square (rms) errors for the state variables of experiments TLXCYR and TLXCYR3km, calculated in the regions where true reflectivity exceeds 10 dBZ, are plotted in Fig. 3 together with those of TLX. It is immediately clear through comparison that the additional CASA radar provides a consistently positive impact on the analyses of essentially all variables. Most of the impact comes from the data
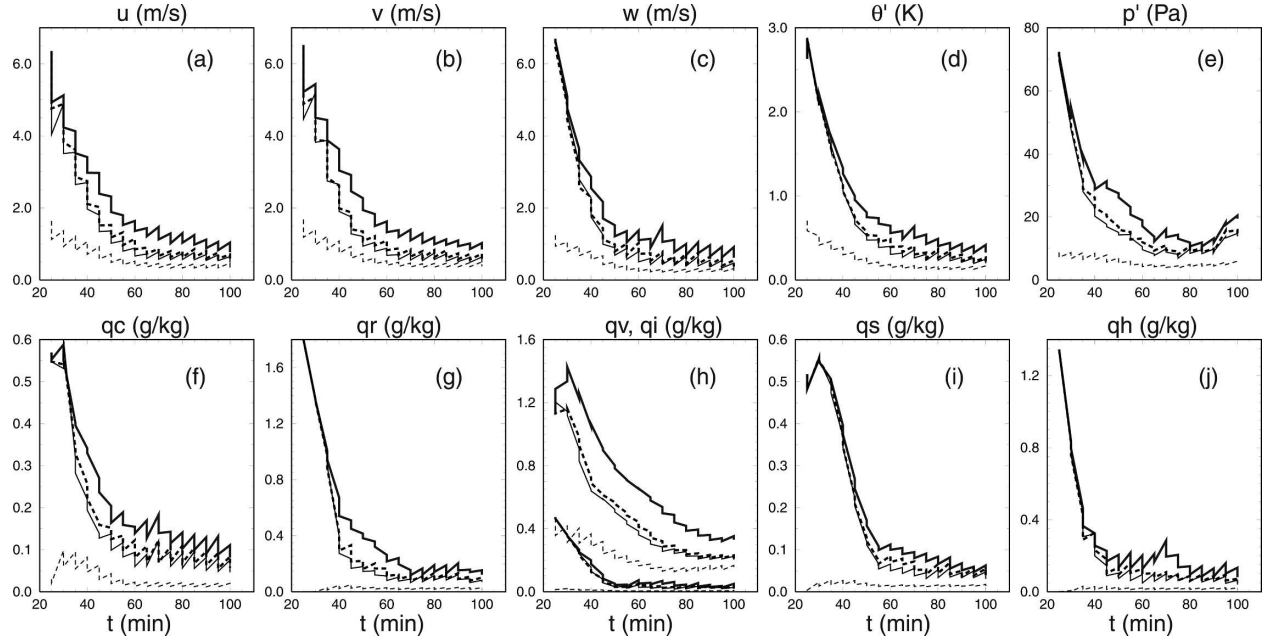
FIG. 3. The rms errors of ensemble mean forecast and analysis, averaged over points at which the true reflectivity is greater than 10 dBZ for (a) u, (b) v, (c) w, (d) θ', (e) p', (f) q_c, (g) q_r, (h) q_v (the curves with larger values) and q_i (the curves with lower values), (i) q_s, and (j) q_h, for experiments TLX (thick solid), TLXCYR (thin solid), and TLXCYR3km (thick dashed), and the ensemble spread of TLXCYR (thin dashed). Units are shown in the plots. The drop of the error curves at specific times corresponds to the reduction of error by analysis.

below 3 km because the errors from TLXCYR and TLXCYR3km are very close but are lower than those of TLX. The improvement is largest before 60 min when errors are reduced significantly by the analysis cycles. The decrease in the rms errors resulting from the addition of the Cyril radar is more than 0.5 m s^{-1} in the wind components, making the analysis errors in the later cycles significantly lower than the typical observational error of 1 m s^{-1} in V_r data. The reduction in the rms error of q_v is about 0.25 g kg^{-1} at around 50 min and the error is reduced from 0.3 to 0.2 g kg^{-1} at the end of the assimilation (100 min) for q_v, which are quite significant improvements. Improvements in the analyses of other variables are also clearly evident according to Fig. 3. The filter is very well behaved overall; less increase in error is found in later cycles than in TX05, which was attributed to the nonsteady evolution of the storm system.

In Fig. 3, the spread, in terms of standard deviation, of the ensemble forecasts before the analysis are also plotted for TLXCYR. It is clear that for all variables, the ensemble spread underestimates the corresponding forecast error, although the discrepancy generally decreases with the assimilation cycles, indicating convergence of the filter. Similar behaviors are also observed in earlier OSSE studies with radar data (Snyder and Zhang 2003; Tong and Xue 2005).

To examine the impact of additional CASA radar data at the low levels on the analyses at different levels, we plot in Fig. 4 the vertical profiles of rms errors for TLX and TLXCYR3km. For these profiles, the errors are calculated for the entire horizontal domain. It can be seen that the largest differences in the rms errors are found at the low levels for most variables. For u, the rms difference at the surface is about 1 m s^{-1} at both times. For θ', the surface rms error differences are between 0.35 and 0.4 K at the two times. The difference in the rms errors between TLX and TLXCYR3km is particularly large in the q_r field, with the difference at the surface being 0.07 g kg^{-1} at 60 min, which is reduced to about 0.04 by 80 min. The larger errors in TLX reflect the poorer analysis of the low-level cold pool, which is further driven by the poorer precipitation analysis (see Fig. 5). The error difference is generally larger at the earlier times (e.g., at 60 min); the difference decreases with the cycles as the analysis of TLX is also improved with the assimilation of more data and with the buildup of the storm.

Figure 5 shows the low-level flow and the simulated reflectivity, together with the cold pool as revealed by negative θ perturbations. It is clear that the analyzed cold pool and the precipitation region are too small in area coverage in both experiments at the earlier times and are even more so in TLX (Figs. 5d and 5g). The
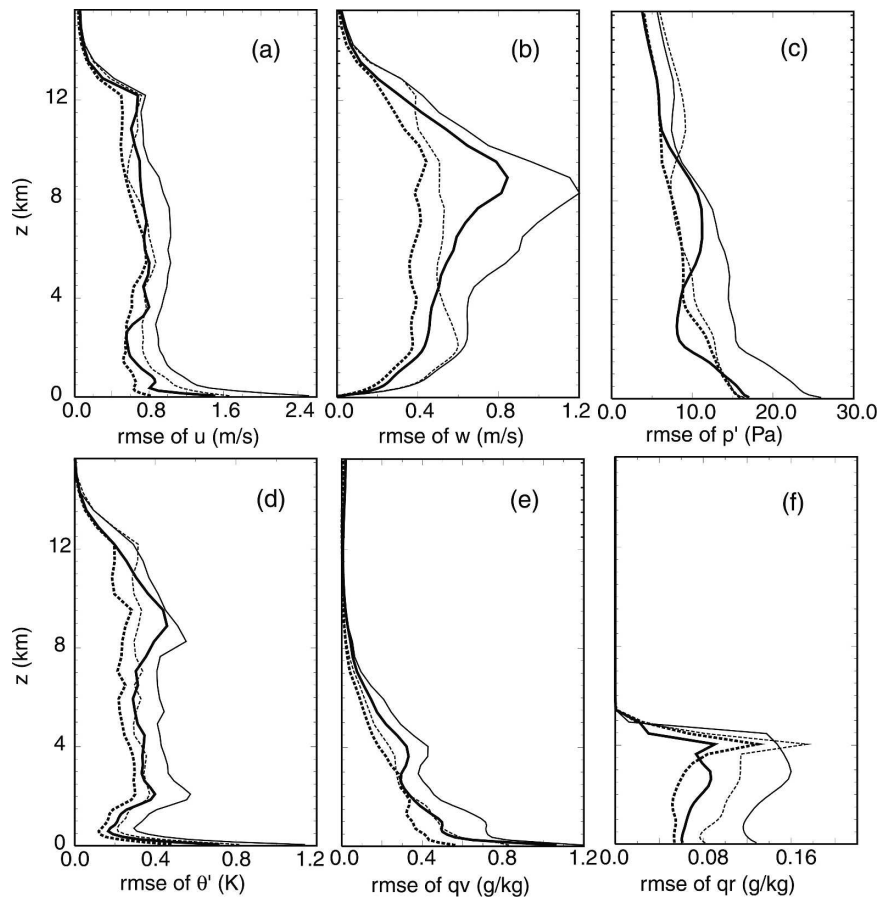
F
IG
. 4. Vertical profiles of rms errors of EnKF analyses (averaged over the entire horizontal domain) from experiments TLX (thin lines) and TLXCYR3km (thick lines) at 60 (solid) and 80 (dashed) min for variables (a) u, (b) w, (c) p', (d) θ', (e) q_v, and (f) q_r.

reflectivity shows a better hook pattern in TLXCYR at 40 min (Fig. 5g). The cold pool expands with time in both cases, but that in TLX never reaches the south boundary of the plotted domain by 100 min as the real one and TLXCYR do. In fact, by 100 min, the structure and location of the cold-pool boundary (or gust front) in TLXCYR agree very well with the truth (Fig. 5i and Fig. 5c), while the agreement in the reflectivity with the truth is very good in both cases at 100 min. At higher levels where the KTLX radar provides good coverage, the differences between the two experiments are smaller though still identifiable, with the updraft structure in TLXCYR generally agreeing better with the truth (not shown).

To further examine the impact of Cyril radar, we plot in Fig. 6 low-level vertical cross sections at 40 and 60 min through the cold pool along lines A–A′ and B–B′ indicated in Fig. 5. The cold pool is indicated by the contours of negative θ', which are plotted together with the winds projected to the plane of cross section. It can

be seen by 40 min, or after four analysis cycles, that the gust-front updraft is reasonably captured in TLXCYR (Fig. 6e) but is much weaker in TLX (Fig. 6c). The latter is because of the absence of a cold pool and the associated gust-front lifting. The low-level winds also differ from the truth significantly more in TLX. After four more analysis cycles, at 60 min, the cold pool becomes much better established in both cases, though that in TLX remains noticeably narrower while the gust-front locations and cold-pool width in TLXCYR are almost identical to the true ones (Fig. 6f versus Fig. 6b). The gust-front strength measured in terms of the horizontal θ' gradient is very good for TLXCYR (Fig. 6f) but is slightly weaker for TLX (Fig. 6d). The above results clearly indicate that the additional data from the Cyril radar are very helpful, especially during the earlier cycles, in establishing accurate low-level precipitation and cold-pool structures as well as the associated winds in the model. At the later times, TLX is able to establish a reasonably good cold pool as the upper-level
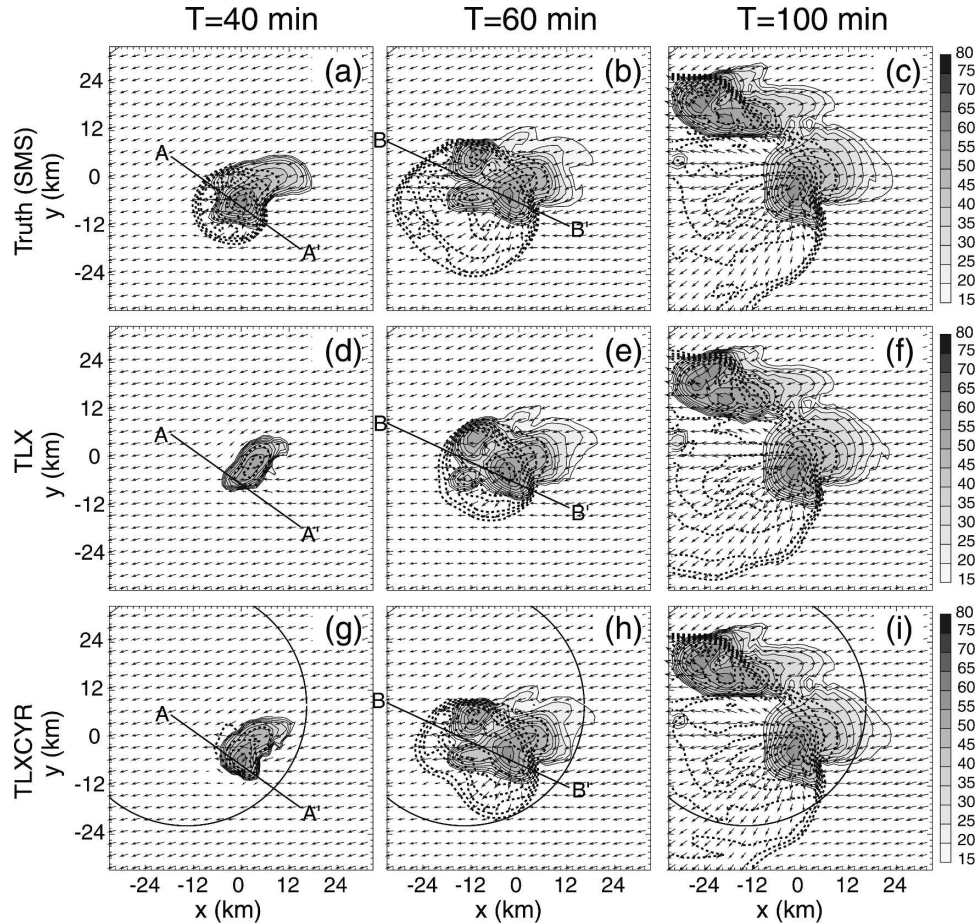
FIG. 5. Horizontal winds (vectors; m s^{-1}), perturbation potential temperature (thick dashed lines at 1-K intervals), and simulated reflectivity (thin solid contours with shading at intervals of 5 dBZ, starting from 15 dBZ) in a subdomain at $z = 50$ m AGL for (a)–(c) truth simulation in the small domain (SMS); (d)–(f) analyses from experiments TLX, and (g)–(i)TLXCYR at $t = 40$, 60, and 100 min during the assimilation period. Lines A–A′ and B–B′ in the plots indicate the locations of vertical cross sections to be shown in Fig. 6. The arcs in the plots are part of the 30-km-range circles of the CASA radars whose data are assimilated by the corresponding experiment(s).

analysis improves, even though the lowest levels are not directly observed. A relatively early start time of the assimilation relative to the life cycle of the true storm might have contributed to this, as did the perfect model that was used.

The impact of vertical radar data coverage is also examined by Zhang et al. (2004), in which simulated radial velocity data are assumed to be available at the grid points. Experiments in which radar data were available only above 2 or 4 km were compared with their control experiment. Similar conclusions on the role of the low-level data were obtained.

b. Assimilation of CASA radar data alone

In this section, we examine the ability of the CASA radar(s) alone to produce a good analysis of the super-cell storm system. Results from experiments CASA1S and CASA3S will be shown (Table 1), which use the same analysis grid and procedure as the previous two experiments. In CASA1S, data from single radar at Rush Springs are used (see Fig. 1), while in CASA3S data from three radars at Cyril, Rush Springs, and Lawton are used. The trailing "S" in the names denotes the small analysis domain.

Figure 7 shows that when three CASA radars that provide a good spatial coverage of the storm system are used (see Fig. 1, Fig. 8, and Table 3), the quality of analysis is close to that of TLXCYR (whose error curves are reproduced in thin black lines in Fig. 7); in fact, for most variables, the error curves of CASA3S are between those of TLX and TLXCYR (cf. Fig. 3). For w, p', q_c, q_r, and q_h, the errors of CASA3S are
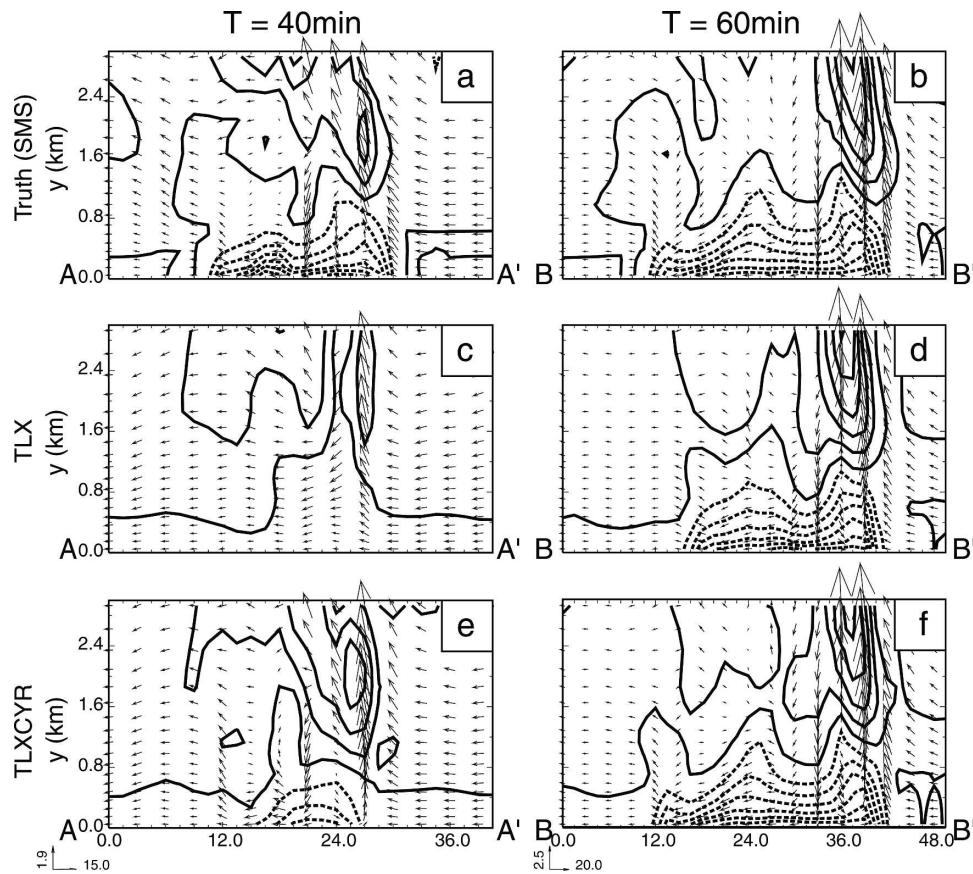
FIG. 6. Vertical cross sections along lines A–A′ (left) and B–B′ (right) in Fig. 5, showing the analysis perturbation potential temperature (θ') contours at 1-K intervals, and the wind vectors projected to the cross section for (a), (b) truth simulation SMS, and (c), (d) experiments TLX and TLCYR (e), (f) at 40 (left) and 60 (right) min during the assimilation period.

actually smaller than those of TLXCYR before 45 min. The wind analysis errors are reduced to below 1 m s^{-1} after 50 min. Similar conclusions can be drawn from the surface plots shown in Fig. 8. The analysis of CASA3S can therefore be considered very good.

The errors of CASA1S are, however, consistently larger at all times (Fig. 7), and the rms errors start to increase significantly after 70 min, reaching above 2 m s^{-1} in the winds by 100 min. The later increase is because of the lack of spatial coverage on the left-moving cells from 60 min onward (Fig. 8b and Fig. 8c) by the single Rush Springs radar. The lack of coverage in the western portion of the analysis domain is also responsible for the inability of the filter to suppress spurious precipitation persistent in this part of domain (Fig. 8b). At 100 min, the low-level cold pool extends too far north (Fig. 8c), partly because of the merger with earlier spurious precipitation near the western boundary (Fig. 8b). Because of the lack of coverage, the midlevel updraft core associated with the left-moving

cell is poorly analyzed by 100 min, while that in CASA3S is very close to the truth (not shown).

The above results show that when three CASA radars work together to provide lower- and midlevel coverage of the storm system during the assimilation period, the EnKF analysis is almost as good as that from one well-positioned CASA radar plus one WSR-88D radar, even though the CASA radars do not reach as high as the WSR-88D radar (cf. Fig. 2). When only one CASA radar is available and when a portion of the storm system moves out of the radar coverage, the quality of analysis deteriorates significantly. Spurious precipitation developed in part of the domain that could not be corrected by the analysis because of the lack of observations there.

c. Effect of storm motion

In all of our OSSEs reported so far, and in those of TX05 and ZSS04, a mean storm motion speed is first subtracted from the environmental sounding to keep
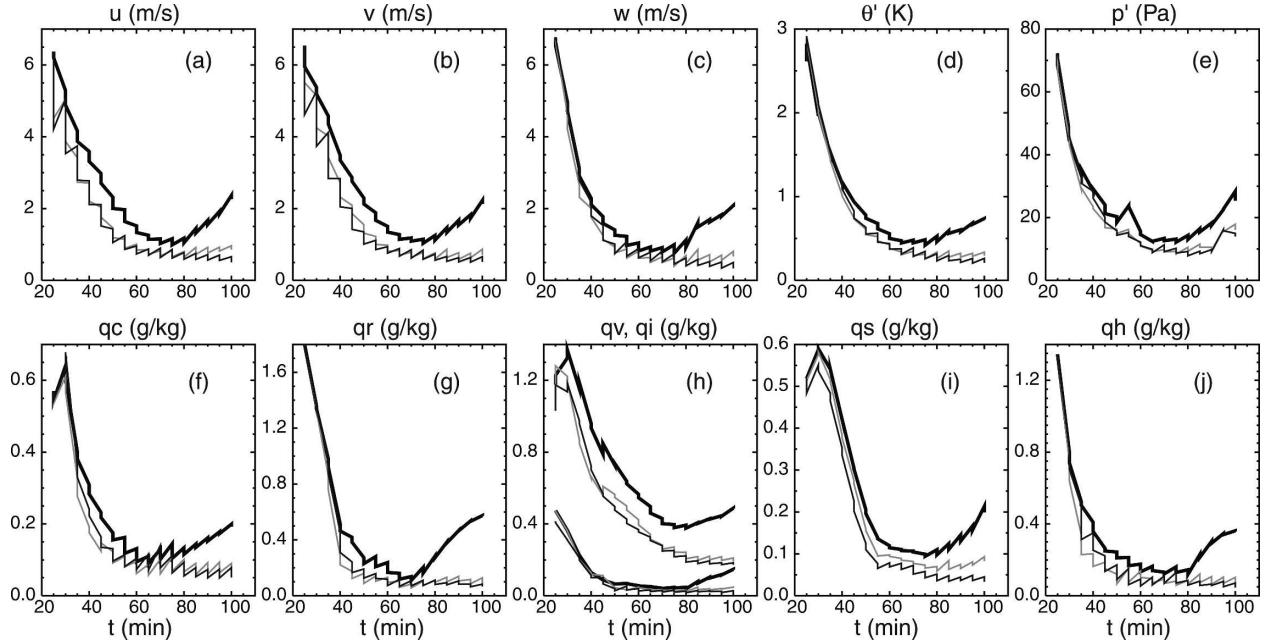
FIG. 7. As in Fig. 3, but for experiments CASA1S (thick black), CASA3S (thin gray), and TLXCYR (thin black).

the main storm cell quasi stationary relative to the model grid. Doing so effectively reduces the local time tendency of the model state and may have helped improve the quality of the analysis. The use of a moving reference frame that follows the storm system is known to improve single-Doppler wind analysis (Gal-Chen 1982; Zhang and Gal-Chen 1996; Liou and Luo 2001);

traditional techniques that retrieve thermodynamic fields from the Doppler wind analyses (Gal-Chen 1978; Gal-Chen and Kropfli 1984) are also sensitive to the accuracy of time tendency estimate (Sun and Crook 1996). For general NWP applications, a moving reference frame is not possible.

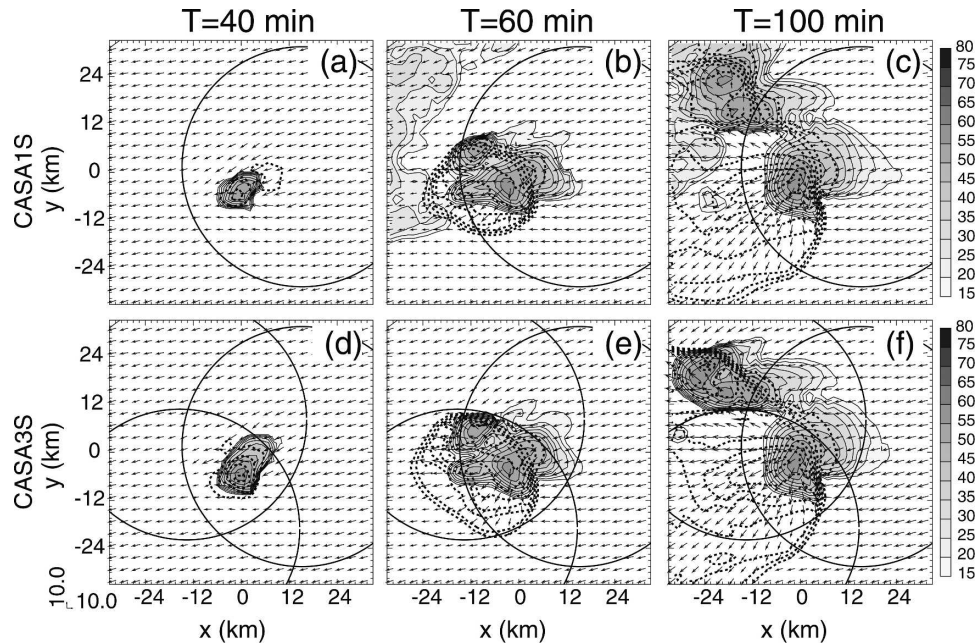In this section, we examine the effect of storm motion

FIG. 8. As in Fig. 5, but for the analyses from experiments (a)–(c) CASA1S and (d)–(f) CASA3S.
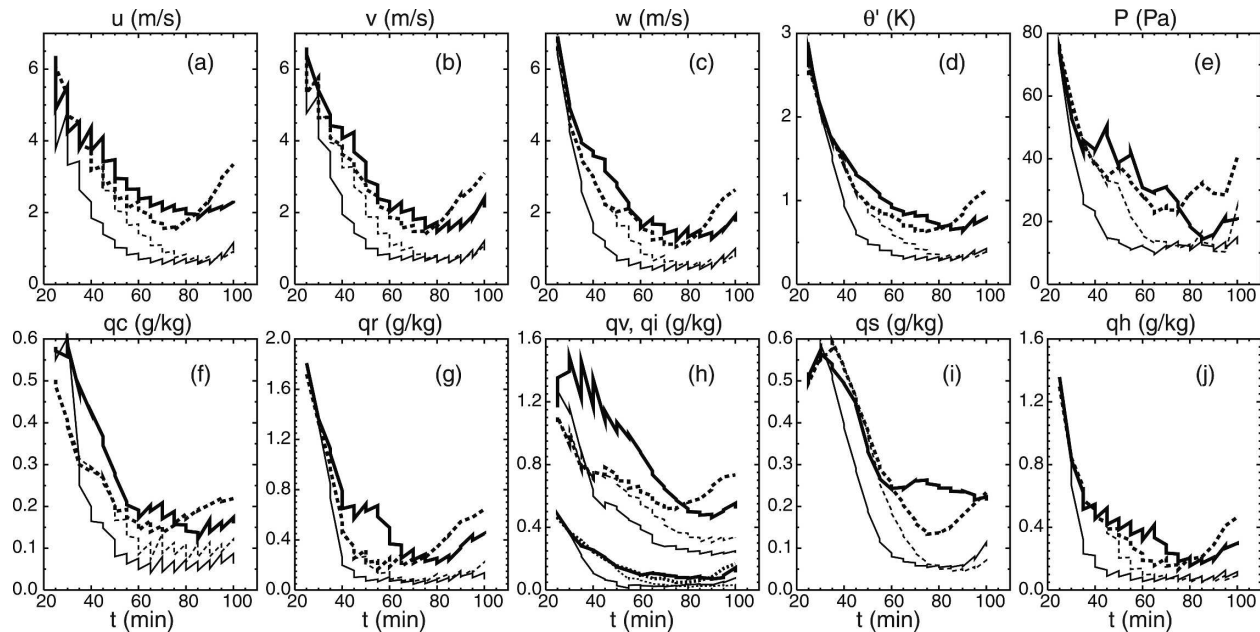
FIG. 9. As in Fig. 3, but for experiments CASA1L (thick solid), CASA4L (thin solid), CASA1LM (thick dashed), and CASA4LM (thin dashed).

speed on the quality of EnKF analysis by comparing experiments that do and do not subtract the storm motion from the sounding. A larger grid, as shown in Fig. 1, is used to contain within the domain the fast-moving storm cells for the entire assimilation period. The truth simulations used for the slow- and fast-moving experiments are SML and FML, respectively.

Figure 9 shows the forecast and analysis errors in the model state variables during the assimilation period for experiments CASA1L, CASA4L, CASA1LM, and CASA4LM (see Table 2). The former two contain a slow-moving storm system while the latter two contain a fast-moving one. In CASA1L and CASA1LM, a single CASA radar located at Lawton is assimilated, while in CASA4L and CASA4LM all four CASA radars are assimilated. Lawton radar provides the best coverage of the fast-moving storm system during the early cycles. All four experiments use the same large domain.

It can be seen from the rms error plots (Fig. 9) that the analyses using four radars are consistently better than those using one radar. Between 70 and 80 min when the errors are generally the smallest, the difference in the wind analysis errors is larger than 1 m s^{-1}, and the difference in θ errors is about 0.5 K between the one- and four-radar cases. In CASA1LM, the errors in all variables increase rapidly after 75 min from a level that is below that of CASA1L, when the precipitation regions of both left and right movers propagate or are

about to propagate out of the range of the Lawton radar (cf. Fig. 11e and Fig. 11f). The same happens to the left movers in CASA1L but at a later time (Fig. 10f). At earlier times, spurious precipitation also exists in a significant portion of the model domain in both single-radar cases that is not corrected (Fig. 10e and Fig. 11e). As in CASA1S, such spurious precipitation is mainly caused by the initial random perturbations used to start the initial ensemble. Because the rms errors shown in Fig. 9 are calculated over the regions where true reflectivity exceeds 10 dBZ, most of the errors resulting from the spurious precipitation are not even reflected in the error plots.

The CASA4L case with a slow-moving storm system and four radars produces the best analysis (Fig. 9), although after 80 min that of CASA4LM becomes very close for almost all variables. The significant difference in the amount of V_r data before 60 min (see Table 3) appears to be the main cause for the difference in the analysis quality between CASA4L and CASA4LM, although the effect of the different local rates of change with these storms may have also played a role.

The surface analysis fields in Fig. 10 and Fig. 11 help us better understand the error evolutions shown in Fig. 9. As suggested earlier, in general, the data coverage appears to be the most significant factor that affects the quality of storm analysis. A slower-moving storm system tends to remain within the radar network longer. Once a storm or a portion of it moves out of the range

TABLE 2. List of large-domain assimilation experiments.

Experiment	Radar(s) used	Storm motion	Volume scan and assimilation interval (min)
CASA1L	Single Lawton radar	Storm motion subtracted	5
CASA4L	All four CASA radars	Storm motion subtracted	5
CASA1LM	Single Lawton radar	Fast moving	5
CASA4LM	All four CASA radars	Fast moving	5
CASA1LF2	Single Lawton radar	Storm motion subtracted	2.5
CASA4LF2	All four CASA radars	Storm motion subtracted	2.5
CASA1LMF2	Single Lawton radar	Fast moving	2.5
CASA4LMF2	All four CASA radars	Fast moving	2.5
CASA1LF1	Single Lawton radar	Storm motion subtracted	1
CASA4LF1	All four CASA radars	Storm motion subtracted	1
CASA1LMF1	Single Lawton radar	Fast moving	1
CASA4LMF1	All four CASA radars	Fast moving	1

of radar network, the model state error growth can no longer be controlled and the analysis deteriorates. The lack of data coverage in the entire analysis domain also negatively impacts the overall analysis because some spurious cells can be not suppressed.

d. Impact of volume scan interval

Another factor that can affect the accuracy of local time tendency estimation is the radar volume scan interval (VSI). Faster scanning tends to give a better estimate of the tendency.[1] Furthermore, the frozen turbulence assumption (Taylor 1938) made in certain single-Doppler velocity retrieval (SDVR) techniques becomes more valid between two scans of a shorter time interval. The mean winds determined from successive volume scans based on the principle of tracking quasi-conserved quantities (e.g., Qiu and Xu 1992; Shapiro et al. 1995) can definitely benefit from higher-frequency data. Shapiro et al. (2003) find that with a SDVR scheme based on a Lagrangian form of the radial component of the equation of motion, the wind retrieval error statistics are substantially improved because the volume scan intervals decrease from 5 min down to 1 min, using real Doppler-on-Wheels mobile radar data. ZSS04 find, however, with their EnKF system, that the analysis is only marginally better during the first few assimilation cycles when 2- instead of 5-min volume scan data are assimilated, and the difference becomes minimal during later cycles. The main storm in their case was quasi stationary, however.

The CASA radars will be designed to operate with a

---

[1] There are limits to the improvement one can expect from scanning more rapidly. The quality of radar measurements deteriorates significantly when the pulse returns used to obtain the averaged values become temporally correlated.

variety of scan strategies that would respond in real time to user needs. When necessary, the radar can perform sector or even spatially or temporally interleaved scans at short time intervals. The increased scan frequency is, however, usually associated with a shortened dwell time of the radar signal on a particular sampling volume, and therefore results in reduced accuracy of the radar measurements. It is important to better understand the impact of scan frequency on the quality of thunderstorm analysis, so as to help design the control system of the network and to optimize the system operations. In this section, we attempt to address some of these issues by comparing analyses from the WSR-88D standard 5-min-interval data and data collected at 2.5- and 1-min intervals. To be fair, the analysis starts at 25 min in all cases. Further, we will examine the impact of VSI for both slow- and fast-moving storms. As in the previous subsection, truth simulations SML and FML are used.

The four large-domain experiments in the previous subsection are repeated by assimilating 2.5- and 1-min volume scan data instead (see Table 2). The rms error curves during the assimilation period for eight fast-scan cases are plotted in Figs. 12–15, together with the corresponding 5-min scan cases. Figure 12 shows that for the single-radar slow-moving storm cases, the analysis and forecast errors are clearly smaller for the 2.5- and 1-min VSIs, and the 1-min VSI gives the smallest errors until 80 min when the analysis becomes ineffective as storm cells move out of the radar coverage. For u, the 1-min volume scan is able to reduce the rms errors to about 1 m s^{-1} while those in the 5-min case remain above 2 m s^{-1}, except for a short period of time around 84 min (Fig. 12a). The error in q_s is also significantly lower in the 1-min case. It can be said that for this single-radar case, faster volume scanning is clearly beneficial. In addition, the errors in the model state is
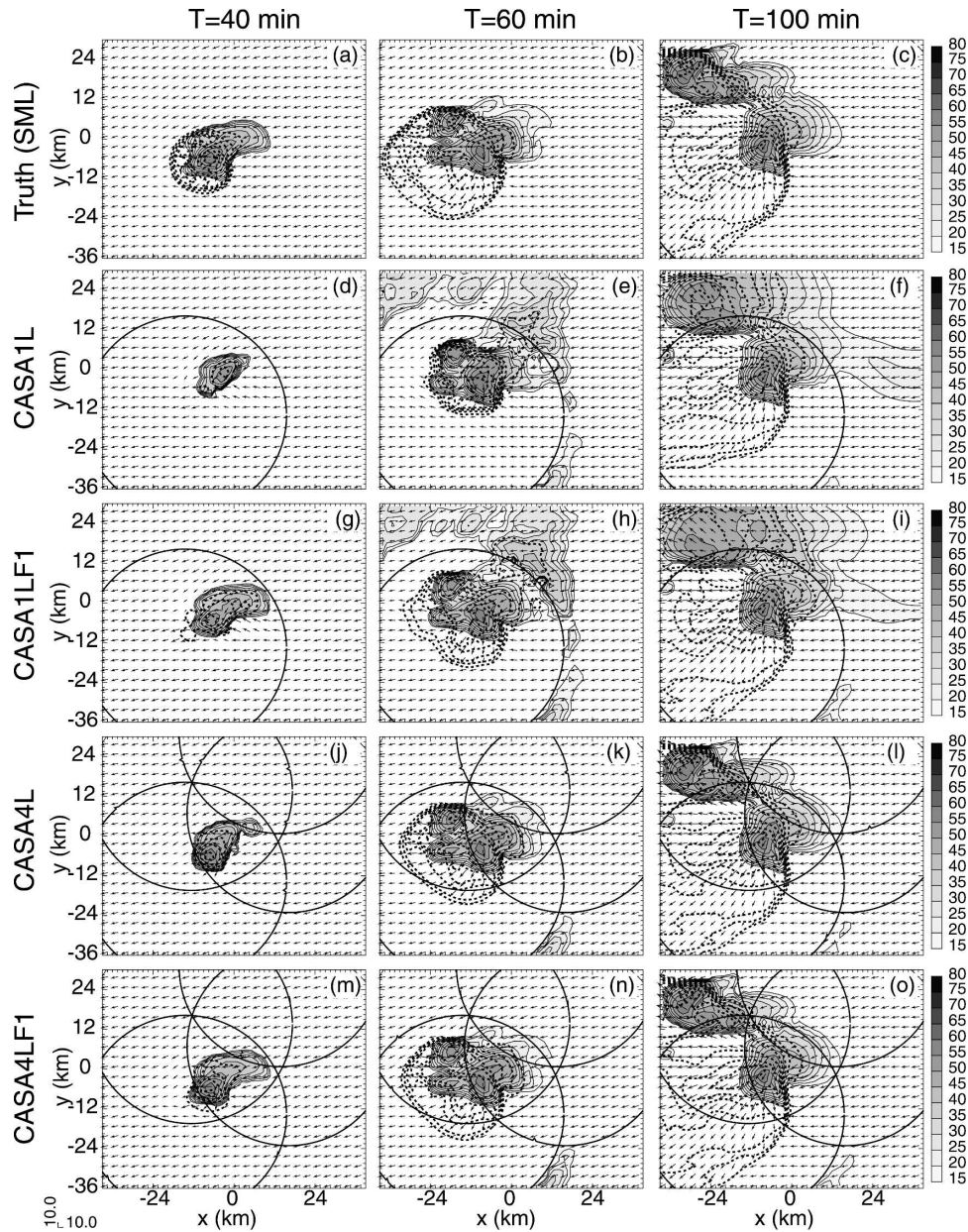
FIG. 10. Horizontal wind (vectors; m s^{-1}), θ' (thick dashed contours at 1-K intervals), and simulated reflectivity (thin solid contours with shading at intervals of 5 dBZ, starting from 15 dBZ) at $z = 50$ m AGL for (a)–(c) truth simulation of slow-moving storm in a large domain (SML); analyses from experiments (d)–(f) CASA1L, (g)–(i) CASA1LF1, (j)–(l) CASA4L, and (m)–(o) CASA4LF1 at $t = 40$, 60, and 100 min during the assimilation period.

steadily reduced in all cases by 1-min analysis cycles during which the forecast errors apparently have little chance to grow; in addition, the system does not seem to suffer from any possible shock introduced by very frequent analysis updates, indicating that the analyzed fields are well balanced.

The errors for the single-radar fast-moving storm cases are shown in Fig. 13. Somewhat counterintu-

itively, the shorter VSI did not have as much an impact on the analysis as in the corresponding slow-moving cases (cf. Fig. 12), although the positive impact does exist before the cells move out of radar coverage. This is mainly because the 5-min VSI analysis of CASA1LM was already very good compared to CASA1L in terms of the rms errors (cf. Fig. 9), therefore there is less room for further improvement.
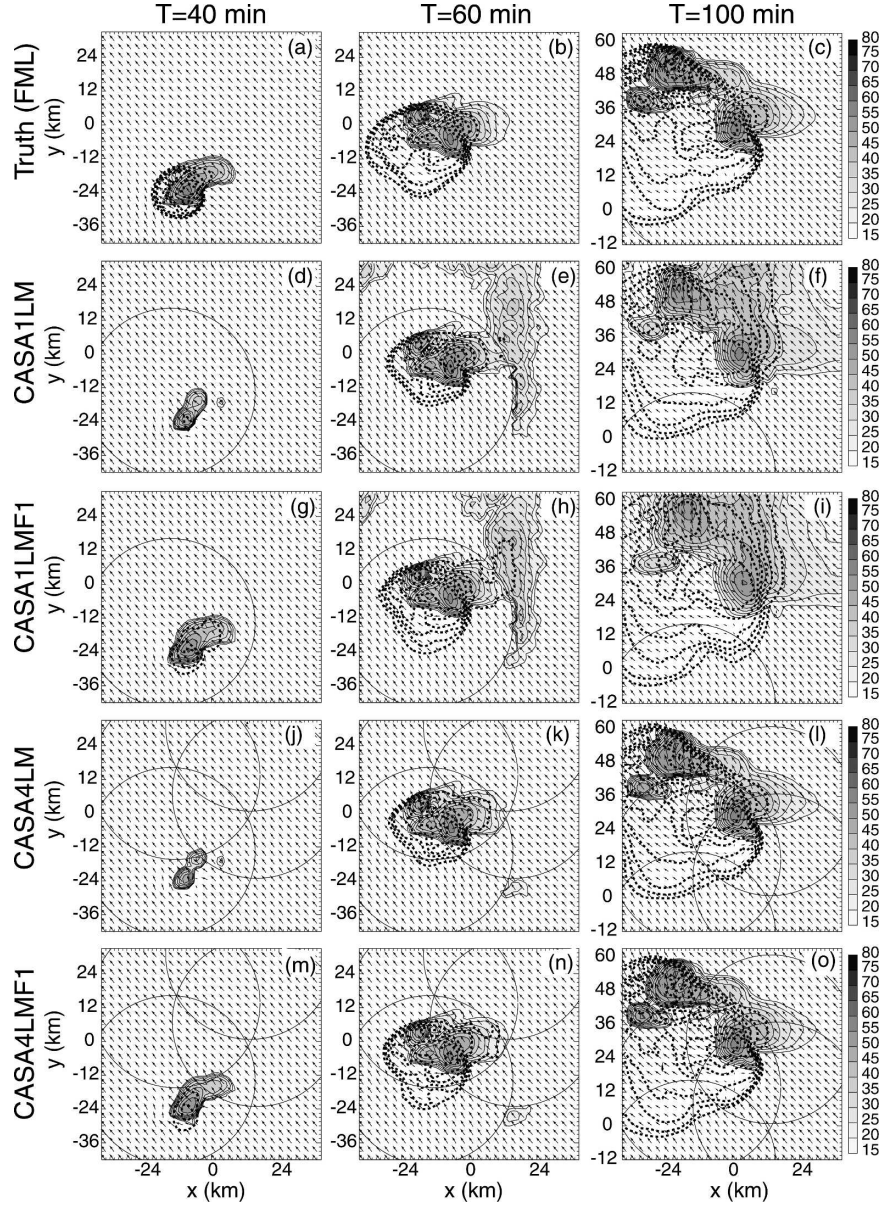
FIG. 11. As in Fig. 10, but for (a)–(c) truth simulation FML, and analyses from experiments (d)–(f) CASA1LM, (g)–(i) CASA1LMF1, (j)–(l) CASA4LM, and (m)–(o) CASA4LMF1. Note the change in the *y*-axis labels across the columns.

The fact that the 5-min VSI analyses of CASA1LM are better than those of CASA1L before the storm system in CASA1LM propagates out of the radar range around 80 min is believed to be because of the better data coverage in the case of CASA1LM before 60 min (Table 3). At 30 min, CASA1LM had about 7% more V_r data, while at 40 and 50 min, it had about 30% more than CASA1L. As pointed out earlier, because the "truth" storms are not the same in the slow- and fast-moving cases, direct comparisons of the errors and the relative impact of data should be viewed with caution.

For the four-radar cases, the relative improvement resulting from the shorter VSI is even less, again because the analyses with 5-min VSI were already very good (Fig. 14 and Fig. 15). Still, the 1-min VSI gives the best analysis in general and the improvement from 5- to 2.5-min VSI is slightly more than that from 2.5 to 1 min. Comparing Fig. 14 and Fig. 15, we can see that the impact of shorter VSI is larger in the fast-moving storm cases for four-radar experiments, and this is most evident in the error curves of w, p', and q_h.

A conclusion can be drawn that faster volume scan-

TABLE 3. Number of radar radial velocity observations for each experiment at selected times during the assimilation cycles.

Experiments	Time						
	30 min	40 min	50 min	60 min	70 min	80 min	90 min
TLX	1121	2106	3043	3754	4272	5049	6177
TLXCYR	1789	3359	5093	6508	7677	9284	11 024
TLXCYR3km	1389	2681	4082	5297	6187	7531	9013
CASA1S	711	2254	3837	4886	4836	5600	6187
CASA3S	1796	4203	6945	8938	9698	11 321	12 408
CASA1L, CASA1LF1/2	1091	1665	2242	2739	3120	3197	2859
CASA4L, CASA4LF1/2	3012	6514	10 164	13 568	14 398	16 347	18 431
CASA1LM, CASA1LMF1/2	1175	2240	2876	2722	2030	828	179
CASA4LM, CASA4LMF1/2	1194	3728	8107	12 751	15 763	16 361	15 231

ning does improve the quality of analysis, especially during the earlier cycles, but the sensitivity of the analysis to VSI is much smaller than the traditional retrieval techniques, and the sensitivity decreases as the length of the assimilation period increases, but generally increases when the storm evolves faster.[2] However, because of the already very good analysis from the 5- and 2.5-min data, the much increased cost of collecting and assimilating 1-min data does not appear justified, at least for the purpose of analyzing a supercell storm system when using a perfect model. This conclusion may or may not hold when the model resolution is increased so as to allow for the resolution of more finescale details that evolve at much shorter time scales and/or when model errors are present that cause faster forecast error growth and/or poorer analysis. The former is not a major issue right now because operational NWP resolution is not expected to exceed that tested here any time soon, while the model error is an important issue for our planned future research.

4. Summary and discussion

In this paper, a radar data assimilation system based on the ensemble square root Kalman filter is described and tested with simulated data. In particular, the impact of data from a standard WSR-88D network radar and those from a network of closely spaced low-cost short-range radars planned for the Oklahoma test bed by the Center for Collaborative Adaptive Sensing of the Atmosphere (CASA; a new NSF Engineering Research Center), are examined for the analysis of a supercell storm. Differing from earlier OSSE studies using an ensemble Kalman filter, the radar data in this study are sampled on the radar elevation levels instead of at the

model grid points, so vertical data resolution varies with range. A realistic Gaussian power-gain weighting function is used to sample the simulated data and as the forward operator during the assimilation. The excellent analysis results indicate that the ensemble-based square root Kalman filter can effectively handle the non-uniform-resolution data.

It is shown that the assimilation of data from one CASA radar, in addition to data from a WSR-88D radar located about 90 km away, improves the analysis. Such improvement is most significant at the low levels where the WSR-88D radar does not observe. The experiments also show that when a single CASA radar is assimilated and when the radar does not provide full coverage of the storm system, significant errors can develop in the analysis that cannot be effectively corrected. The combination of several CASA radars effectively eliminates the problem when complete coverage is available. When three CASA radars work together to provide lower and midlevel coverage of the storm system, the analysis is almost as good as that from one well-positioned CASA radar plus one WSR-88D radar, even though the CASA radars do not measure as high as does the WSR-88D radar.

The impact of storm propagation speed on the quality of analysis is also examined. In general, the effect of storm propagation on the data coverage by the small radar network appears to be the most important factor, and the analysis is generally not very sensitive to the propagation speed otherwise. The quality of analysis can be improved by employing faster volume scans, especially for less well analyzed storms with slower volume scans, but the sensitivity of the EnSRF analysis of convective storms to the volume scan interval is much less than that of more traditional single-Doppler velocity and thermodynamic retrieval schemes. In fact, very good analyses can be obtained even with the WSR-88D standard 5-min volume scan interval. For this reason, more versatile scan strategies may be developed for and

---

[2] This is true for the four CASA radar cases (CASA4L and CASA4LM). The counterintuitive behaviors with the single CASA radar cases (CASA1L and CASA1LM) have been discussed earlier.
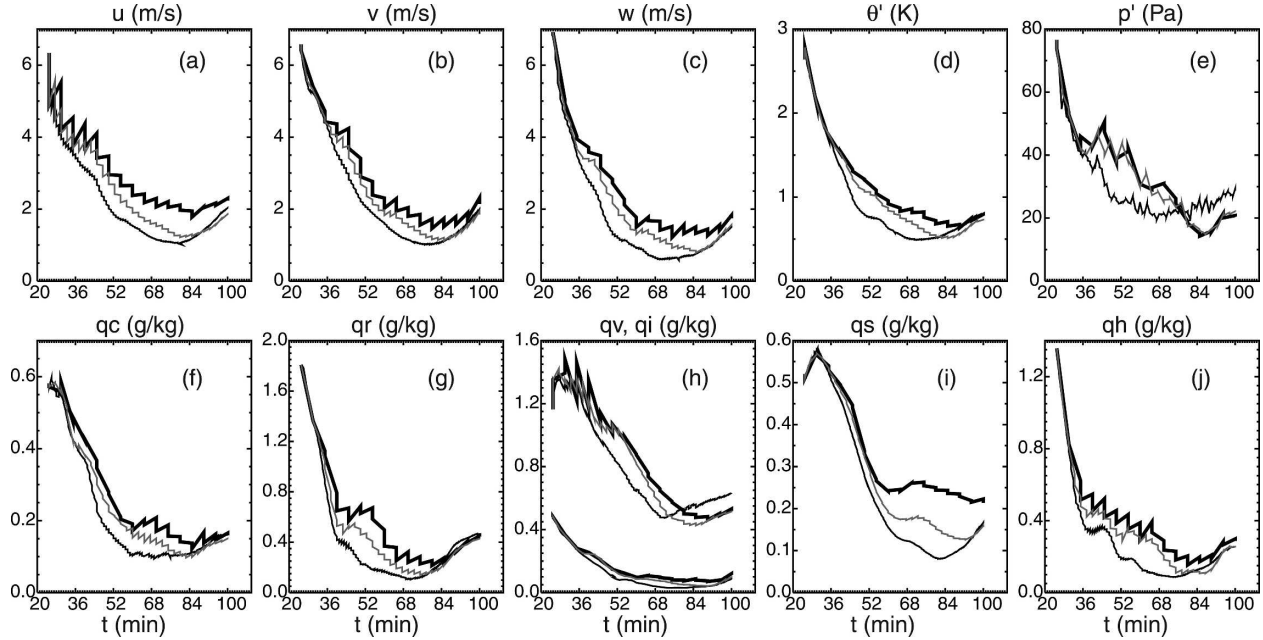
FIG. 12. As in Fig. 3, but for experiments CASA1L (thick solid), CASA1LF2 (gray), and CASA1LF1 (thin solid).

employed by the CASA radars in response to user needs. For example, complete volume scans can be made by the radar network every 5 min in short periods while in between, the radars can be doing sector scans that focus on active local features such as a tornado or a microburst.

The data coverage appears to be the most significant factor that affects the quality of storm analysis. A general rule that appears to be valid is that the more observations, the better the analysis. A slower-moving storm system tends to remain within the relatively small radar network longer. Once a storm or a portion of it moves out of the range of radar network, the model error growth can no longer be controlled and the analy-
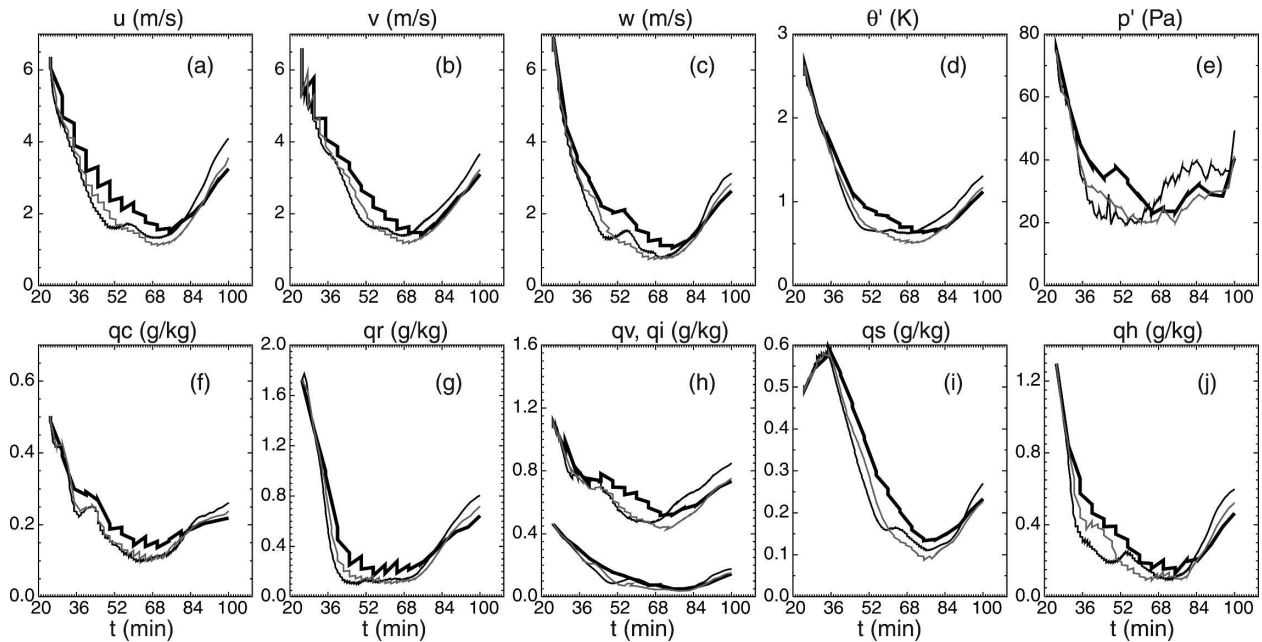

FIG. 13. As in Fig. 3, but for experiments CASA1LM (thick solid), CASA1LMF2 (gray), and CASA1LMF1 (thin solid).
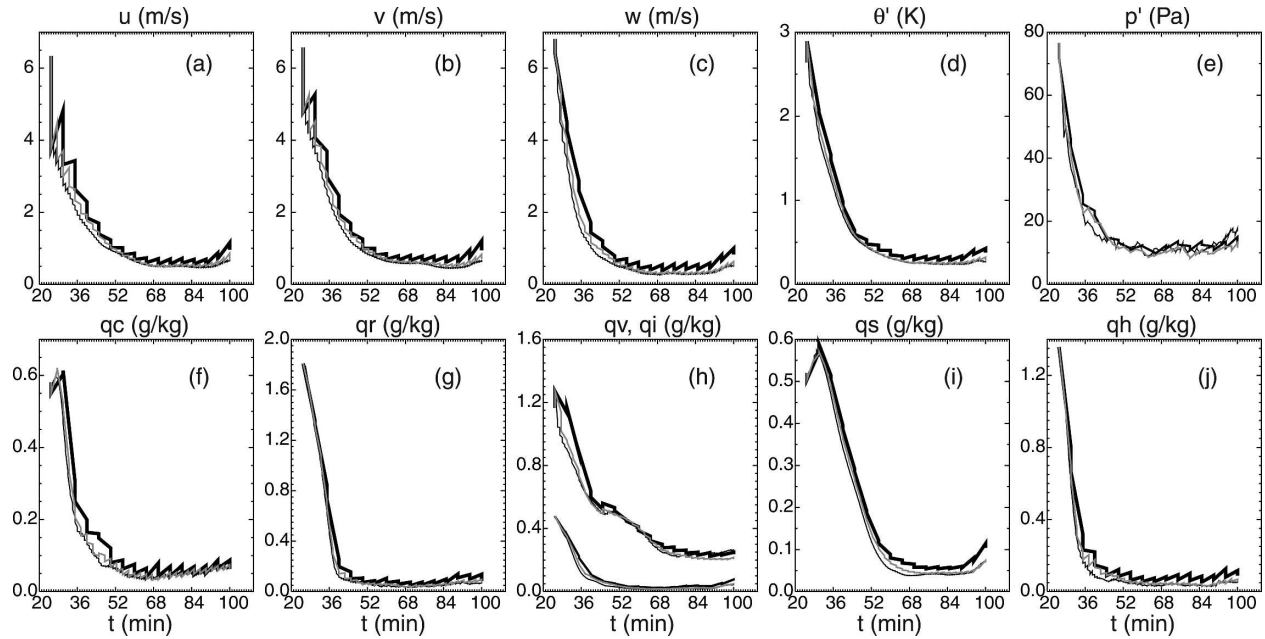
FIG. 14. As in Fig. 3, but for experiments CASA4L (thick solid), CASA4LF2 (gray), and CASA4LF1 (thin solid).

sis deteriorates. The lack of data coverage in the entire analysis domain also negatively impacts the overall analysis because some spurious cells cannot be suppressed.

In Snyder and Zhang (2003), a considerable variability was found in the analysis quality for different realizations of the initial random perturbations used to ini-

tialize their 50-member forecast ensemble. In Tong and Xue (2005), much less variability was found, using the classical EnKF method with perturbed observations but with 100 ensemble members. In this paper, we performed 10 versions of experiment TLX, starting from different realizations of initial random perturbations, and found the rms errors of analysis to be generally
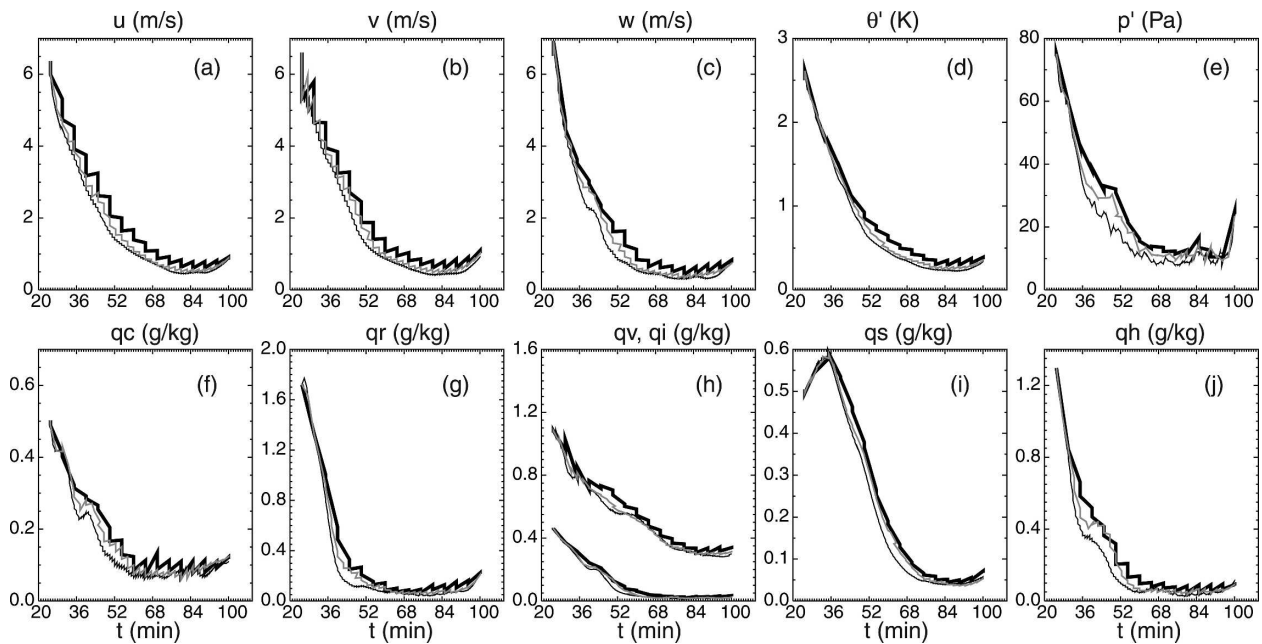

FIG. 15. As in Fig. 3, but for experiments CASA4LM (thick solid), CASA4LMF2 (gray), and CASA4LMF1 (thin solid).

within a few percent of each other (not shown), suggesting that the analyses are not sensitive to the different realizations, as long as the same initialization procedure is used.

Near the end, we point out that the CASA radar-related issues examined in this study are only a few of many. The dynamic adaptive systems in the CASA radar networks promise to establish a new paradigm for the sensing of the atmosphere, and the impact of data collected using a variety of possible scanning modes remain to be studied in a more systematic way. This study represents only the first step in this direction.

We also point out that the value of obtaining a good estimate of the state of the atmosphere is to produce accurate weather predictions. Limited by space, we confine our current study to the impact of CASA radar data on the analysis only. The impact of the data on model predictions will be a subject of future studies, which will also examine the growth of errors during the forecast period. We also caution the readers that the results obtained in this study are based on a perfect prediction model and perfect forward observational operators. In the presence of model errors, as is the case with real data, it will be much more difficult, if at all possible, to obtain analyses that match the accuracy obtained here. The impact of additional or even redundant observations, such as those provided by multiple CASA radars scanning at high frequencies, is expected to produce a larger impact than observed here.

Including model errors in ensemble Kalman filter data assimilation systems is a very important area of research, especially given the fact that real data applications of this method remain very limited. For the convective scale, when significant model bias and/or initial condition error exist, it is possible that all members of an ensemble fail to predict any storm. In such a case, the background error estimate will be in serious error, causing the filter to fail. These are some of the issues that one has to deal with when working with real data.

REFERENCES

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

Bennett, A. F., 1992: *Inverse Methods in Physical Oceanography.* Cambridge University Press, 346 pp.

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

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

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

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.

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), 10 143–10 162.

——, 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.,* **53,** 343–367.

Gal-Chen, T., 1978: A method for the initialization of the anelastic equations: Implications for matching models with observations. *Mon. Wea. Rev.,* **106,** 587–606.

——, 1982: Errors in fixed and moving frame of references: Applications for convectional and Doppler radar analysis. *J. Atmos. Sci.,* **39,** 2279–2300.

——, and R. A. Kropfli, 1984: Buoyancy and pressure perturbation derived from dual-Doppler radar observations of the planetary boundary layer: Applications for matching models with observations. *J. Atmos. Sci.,* **41,** 3007–3020.

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

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

——, and ——, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.,* **129,** 123–137.

Le Dimet, F. X., and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus,* **38A,** 97–110.

Lin, Y.-L., R. D. Farley, and H. D. Orville, 1983: Bulk parameterization of the snow field in a cloud model. *J. Climate Appl. Meteor.,* **22,** 1065–1092.

Liou, Y. C., and I. S. Luo, 2001: An investigation of the moving-frame single-Doppler wind retrieval technique using Taiwan area mesoscale experiment low-level data. *J. Appl. Meteor.,* **40,** 1900–1917.

Qiu, C.-J., and Q. Xu, 1992: A simple adjoint method of wind analysis for single-Doppler data. *J. Atmos. Oceanic Technol.,* **9,** 588–598.

Ray, P. S., B. C. 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.

Shapiro, A., S. Ellis, and J. Shaw, 1995: Single-Doppler radar retrievals with Phoenix II data: Clear air and microburst wind

retrievals in the planetary boundary layer. *J. Atmos. Sci.,* **52,** 1265–1287.

——, P. Robinson, J. Wurman, and J. Gao, 2003: Single-Doppler velocity retrieval with rapid-scan radar data. *J. Atmos. Oceanic Technol.,* **20,** 1758–1775.

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.

Sun, J., and N. A. Crook, 1996: Comparison of thermodynamic retrieval by the adjoint method with the traditional retrieval method. *Mon. Wea. Rev.,* **124,** 308–324.

——, and ——, 2001: Real-time low-level wind and temperature analysis using single WSR-88D data. *Wea. Forecasting,* **16,** 117–132.

Talagrand, O., and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. Part I: Theory. *Quart. J. Roy. Meteor. Soc.,* **113,** 1311–1328.

Taylor, G. I., 1938: The spectrum of turbulence. *Proc. Roy. Soc. London,* **164A,** 476–490.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.,* **131,** 1485–1490.

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.

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

Wood, V. T., and R. A. Brown, 1997: Effects of radar sampling on single-Doppler velicity signatures of mesocyclones and tornadoes. *Wea. Forecasting,* **12,** 928–938.

Xue, M., K. K. Droegemeier, V. Wong, A. Shapiro, and K. Brewster, 1995: ARPS version 4.0 user's guide. CAPS, University of Oklahoma, 380 pp. [Available online at http://www. caps.ou.edu/ARPS.]

——, ——, and ——, 2000: The Advanced Regional Prediction System (ARPS)—A multiscale nonhydrostatic atmospheric simulation and prediction tool. Part I: Model dynamics and verification. *Meteor. Atmos. Phys.,* **75,** 161–193.

——, 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. Phys.,* **76,** 143–165.

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

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.

Zhang, J., and T. Gal-Chen, 1996: Single Doppler wind retrieval in the moving frame of reference. *J. Atmos. Sci.,* **53,** 2609–2623.