

The analysis and impact of simulated high-resolution surface observations in addition to radar data for convective storms with an ensemble Kalman filter

Jili Dong · Ming Xue · Kelvin Droegemeier

Received: 9 October 2010 / Accepted: 17 February 2011 / Published online: 2 March 2011
© Springer-Verlag 2011

Abstract Observing system simulation experiments are performed using an ensemble Kalman filter to investigate the impact of surface observations in addition to radar data on convective storm analysis and forecasting. A multi-scale procedure is used in which different covariance localization radii are used for radar and surface observations. When the radar is far enough away from the main storm so that the low level data coverage is poor, a clear positive impact of surface observations is achieved when the network spacing is 20 km or smaller. The impact of surface data increases quasi-linearly with decreasing surface network spacing until the spacing is close to the grid interval of the truth simulation. The impact of surface data is sustained or even amplified during subsequent forecasts when their impact on the analysis is significant. When microphysics-related model error is introduced, the impact of surface data is reduced but still evidently positive, and the impact also increases with network density. Through dynamic flow-dependent background error covariance, the surface observations not only correct near-surface errors, but also errors at the mid- and upper levels. State variables different from observed are also positively impacted by the observations in the analysis.

1 Introduction

Successful numerical weather prediction (NWP) depends greatly on accurate initial conditions obtained with data assimilation. For convective storms, radar is the primary observational platform used. However, radars usually do not observe down to the ground level, because of the non-zero elevation of the lowest scans and due to the earth curvature effect. This problem becomes worse when the storm is located far from the radar. For example, when a storm is 100 km away, the center of the lowest 0.5° elevation beam of a WSR-88D radar is more than 2 km above the ground. Even the lower edge of the half power beam does not reach the ground (see, e.g., Fig. 2 of Xue et al. 2006, XTD06 hereafter). Yet, for convective systems, the low-level flows and cold pool are critically important in storm development and evolution. At the ground level, observations from automated meteorological stations and sometimes from mesoscale observing networks are available. Effective assimilation of these observations, in combination with radar data, has the potential to significantly improve storm analysis and forecasts. These data are also valuable for defining the low-level environment of the convective storms, where radar has very limited observing capabilities. The resolution of the surface network, however, is usually low compared with the scales of convective features; thus their quantitative impact when radar is also present is not necessarily clear. Further, the downdrafts are primarily responsible for the cold pool, but it is not clear how well the storm can retain the information contained in surface observations taken in the downdraft region.

For the prediction of convective initiation and its later evolution, Liu and Xue (2008) showed that the strength of the analyzed cold pool is sensitive to the vertical correlation length scale specified in the ARPS (Xue et al. 2000)

Responsible editor: F. Mesinger.

J. Dong (✉) · M. Xue · K. Droegemeier
Center for Analysis and Prediction of Storms and School
of Meteorology, National Weather Center,
University of Oklahoma, Suite 2500,
120 David L. Boren Blvd, Norman, OK 73072, USA
e-mail: jldong@ou.edu

Data Analysis System (ADAS, Brewster 1996), when analyzing high-resolution surface observations. Similar sensitivity was found by Dawson and Xue (2006). ADAS is based on the Bratseth successive correction method (Bratseth 1986) that theoretically converges to the optimal interpolation (OI) solution. The background error covariance in ADAS is modeled using a Gaussian correlation model and allows for the use of different spatial correlation scales in different analysis passes. In ADAS, the vertical correlation length can be specified in terms of geometric height or potential temperature. The latter allows more vertical spreading of observation information in less stable condition, but the correlation scale is still empirical. Using ADAS and observations from both routine and special mesoscale surface networks that were gathered by the 2002 International H₂O Project (IHOP_2002, Weckwerth et al. 2004), Liu and Xue (2008) demonstrated significant positive impact of hourly surface data over a 6-h period on convective initiation forecast in the ARPS model. Similar results were obtained by Xue and Martin (2006) for another convective initiation case from IHOP_2002.

To effectively assimilate surface observations, flow-dependent background error covariance is needed. The ensemble Kalman filter method (EnKF, Evensen 2003) is one method that estimates and evolves flow-dependent background error covariance using ensemble forecasts through assimilation cycles. There are several studies demonstrating that EnKF can provide comparable or better performance to three-dimensional variational data assimilation method (3DVar) which uses a static background error covariance (Whitaker et al. 2004; Compo et al. 2006; Whitaker et al. 2007; Meng and Zhang 2008).

The assimilation of surface observations for the planetary boundary layer using EnKF has been examined recently in simple column model settings with simulated and real observations (Hacker and Snyder 2005; Hacker and Rostkier-Edelstein 2007). Simulated surface observations have also been tested with a mesoscale model for the case of a synoptic-scale winter cyclone by Zhang et al. (2006) and Meng and Zhang (2007). Fujita et al. (2007) examined the performance of EnKF for synoptic to mesoscale flows with real surface observations. These studies generally address the situations where the atmospheric boundary layer is strongly influenced by the land surface processes; they do not address the specific situation where the thunderstorm downdraft and cold pool play an important role.

Assimilation of high-resolution surface observations, from, e.g., a high-density mesonet, for the initialization of explicit convective storms, however, is limited to the study of Zhang et al. (2004), which found positive impact of simulated surface observations on model grids at 10 km spacing when radar data were artificially limited to levels

above 4 km. Given the focus on the impact of radar observations in their study, the impact of surface data was only addressed briefly. Their study also made the simplifying assumptions that the radar observations are available at model grid points; one of the surface observation types considered is the liquid–water potential temperature which is not directly measured. Only warm rain microphysics was used. These simplifying assumptions make it impossible to address the issue of radar distance from the storm, which in many cases is the most important factor determining the low-level data coverage. The assumption that the radar data are available at the model grid points also makes it impossible to include the effect of beam spreading which affects data coverage. The lack of low-level radar data coverage is also suspected to be an important source of error in the EnKF analysis of thunderstorms as shown in the studies of Dowell et al. (2004) and Tong (2006).

We set out to perform a systematic OSSE (Observing System Simulation Experiment, Lord et al. 1997) study on the impact of assimilating high-resolution observations from hypothetical surface networks, in addition to observations from a radar located at different distances from the main storm, trying to complete a systematic study on surface data impact in convective scale storm analysis. A realistic radar simulator is used, providing realistic data coverage. Surface observing networks of different mean spacing are examined, and the impact of surface data is interpreted with the help of ensemble error covariance structures and our understanding of convective storm dynamics. Both perfect and imperfect forecasting model scenarios are considered in the OSSEs; in the latter case model error is introduced by using different microphysics schemes.

The rest of this paper is organized as follows: Section 2 describes the numerical model used, the design of the observational network and the EnKF algorithm used. Experiments assimilating surface observations with EnKF are discussed in Sect. 3 and sensitivity tests are presented and analyzed in Sect. 4. The impact of this assimilation on subsequent forecasts is discussed in Sect. 5. Section 6 further examines the surface data impact in the presence of model error, and Sect. 7 gives a summary and conclusions.

2 Model and experiment settings

We take the OSSE approach in this study, partly because it is impossible or impractical to perform a complete set of experiments with real observations that would require, for example, observing the same real storm many different times using the same radar from different distances, and the real surface observing networks often have limited density. In the case of real thunderstorms, full measurement on the complete atmospheric state is not available, making the

quality of analysis difficult to judge. With OSSEs, a model simulation serves as the ‘truth’ or ‘nature run’ for all experiments. Realistic observations can be simulated from this model atmosphere using a radar simulator and assuming certain surface network characteristics. With OSSEs, observation configurations that are currently unavailable can be tested. This study builds upon and extends the earlier OSSE studies of Tong and Xue (2005, TX05 hereafter), XTD06 and Tong and Xue (2008a), which focused on radar data.

2.1 Description of simulation and assimilating model

The Advanced Regional Prediction System (ARPS, Xue et al. 2000; Xue et al. 2001; Xue et al. 2003) is used as the simulation and prediction model in this study. ARPS is a compressible nonhydrostatic model that predicts velocity components u , v , w , potential temperature θ , pressure p and mixing ratios for water vapor, cloud water, rainwater, cloud ice, snow and hail/graupel (q_v , q_c , q_r , q_i , q_s , and q_h , respectively) associated with ice microphysics schemes in the model. These variables make up the state vector that is estimated or analyzed using EnKF. The model also predicts the turbulence kinetic energy (TKE), which is used in a 1.5-order subgrid-scale turbulence closure scheme. TKE is not updated by the EnKF during assimilation as in our previous studies (e.g., TX05). The ARPS includes several single-moment ice microphysics packages. The Lin et al. (1983, LFO83 hereafter) scheme is the default and is used in the truth simulation and perfect model OSSEs. A modified version of the LFO83 scheme (Gilmore et al. 2004a, LFO04 hereafter), the Schultz simplified ice scheme (Schultz 1995, Schultz hereafter), the WRF 6-category Single-Moment Microphysics scheme (Hong and Lim 2006, WSM6 hereafter), and their combinations are used in the imperfect-model OSSEs.

2.2 Truth storm simulation

As in TX05 and XTD06, the 20 May 1977 Del City, Oklahoma, supercell storm (Ray et al. 1981) simulated by the ARPS is used as the truth for the OSSEs. A horizontal resolution of 2 km is chosen with a grid of $67 \times 67 \times 35$ points in the x , y and z directions, respectively, giving a *physical* domain of $128 \times 128 \times 16$ km. Radar and surface observations are assumed to be available only within a 64×64 km² subdomain located at the center of the model grid. The main storm cells remained within this subdomain throughout assimilation and prediction, and this observation domain has the same size as that used by TX05. The larger domain used here is to avoid occasional problems arising from perturbations created by the open lateral boundary condition. In the vertical, a grid stretching

scheme based on a hyperbolic tangent function is used and the vertical grid spacing is 200 m near the ground and increases to 800 m at the model top.

An initial thermal bubble with a maximum potential perturbation of 4 K is centered at $x = 80$ km, $y = 48$ km and $z = 1.5$ km to initialize convection in the truth simulation. The radii of the bubble are 10 km in the horizontal and 1.5 km in the vertical. Open lateral and free-slip top and bottom boundary conditions are used in both simulation and assimilation. A constant wind of $u = 3$ m s⁻¹ and $v = 14$ m s⁻¹ is subtracted from the original sounding to keep the main storm (right mover) near the center of the domain. These configurations are the same as those used in TX05, except for the larger computational domain and the use of vertical grid stretching; the stretched grid gives better vertical resolutions near the surface.

The bubble-triggered storm updraft reaches its full updraft intensity within 30 min. At around 60 min, the supercell starts to split into two, with one right mover and one left mover (Fig. 1a). By 120 min, the left mover exits the northwest corner of the central subdomain (Fig. 2a). Additional details on the general evolution of the simulated storm can be found in TX05. The simulated model state is output every 5 min for creating simulated observations and for analysis/forecast verification.

2.3 Simulation of observations

A hypothetical WSR-88D radar with 1° beamwidth is positioned to the southwest of the storm. For the first set of experiments, the radar is located at $x = -68$ km and $y = -68$ km (the coordinate origin is located at the southwest corner of the 128×128 km model domain); this is about 185 km from the domain center; approximately where the right-moving cell is located. At this distance, the earth curvature effect combined with beam bending based on the $\frac{3}{4}$ earth radius model (Doviak and Zrnic 1993) places the lower edge and the center of the half-power beam of 0.5° elevation at 1.98 and 3.60 km above ground, respectively. In another word, there is no direct radar data coverage below 1.98 km level at all at this distance. The vertical beamwidth as well as the spacing between two consecutive beams is about 3.8 km at this distance; for this reason, proper vertical beam pattern weighting is important. Two other sets of experiments assume that the radar is located 115 or 45 km southwest of the main storm (see Table 1).

As in XTD06, radar data are assumed to be available on the elevation levels in the vertical and already interpolated to the model grid columns in the horizontal. The radar emulator as used in XTD06 is applied to bring the data from the model grid points to the radar elevation levels with vertical beam pattern weighting. The standard WSR-

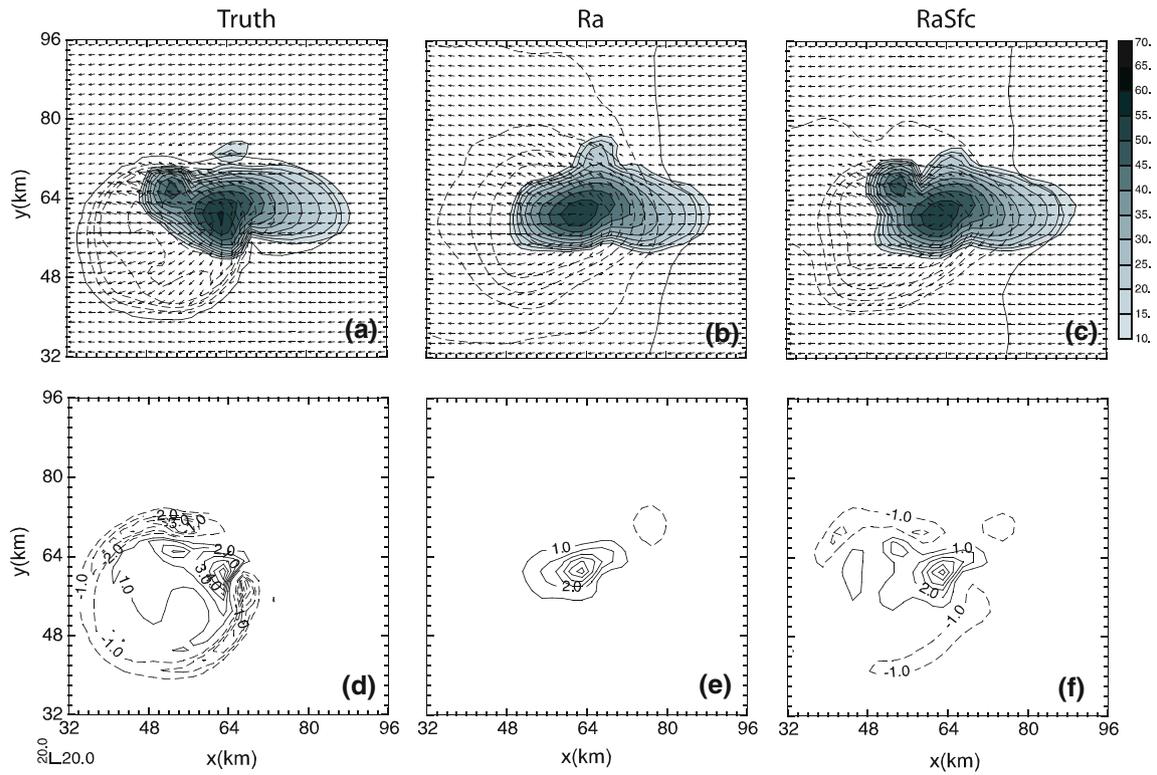


Fig. 1 Perturbation wind vectors, simulated reflectivity Z (dBZ) and perturbation potential temperature θ' (every 1 K) (upper panel), and divergence fields ($\times 1,000 \text{ s}^{-1}$) (lower panel) at $z = 100 \text{ m}$ (first model level above ground) for truth (a, d), and experiments Ra (b, e) and RaSfc (c, f) at 60 min of model time

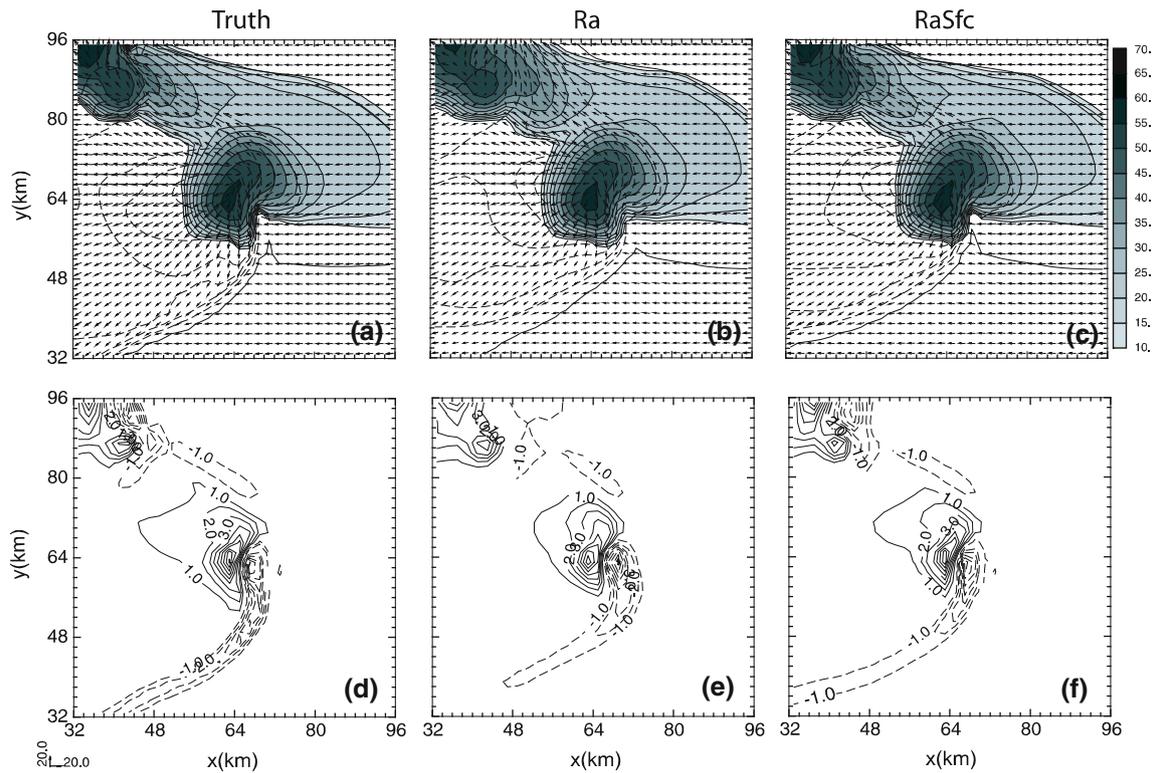


Fig. 2 As Fig. 1, but at 120 min

Table 1 List of OSSE experiments examining the impact of surface observation data of different spacings and for different radar distances under the scenario of perfect and imperfect model

Experiment	Obs. Type	Sfc. Obs. Spacing (km)	Radar Dist. (km)	Remark/aspects tested
Ra/Cntl	Radar only	N.A.	185	Control exp.
RaSfc	Radar + Sfc All	20	185	Cntl sfc obs spacing
RaSfcUV	Radar + Sfc V	20	185	Individual sfc obs
RaSfcT	Radar + Sfc T	20	185	Individual sfc obs
RaSfcP	Radar + Sfc Pres	20	185	Individual sfc obs
RaSfcQv	Radar + Sfc Qv	20	185	Individual sfc obs
RaSfcS32	Radar + Sfc All	32	185	Sfc obs spacing
RaSfcS16	Radar + Sfc All	16	185	Sfc obs spacing
RaSfcS12	Radar + Sfc All	12	185	Sfc obs spacing
RaSfcS6	Radar + Sfc All	6	185	Sfc obs spacing
RaSfcS4	Radar + Sfc All	4	185	Sfc obs spacing
RaSfcS2	Radar + Sfc All	2	185	Sfc obs spacing
RaD115	Radar only	N.A.	115	Radar Dist
RaSfcD115	Radar + Sfc All	20	115	Radar Dist
RaSfcD115S6	Radar + Sfc All	6	115	Radar Dist and Sfc Obs spacing
RaD45	Radar only	N.A.	45	Radar Dist
RaSfcD45	Radar + Sfc All	20	45	Radar Dist
RaSfcD45S6	Radar + Sfc All	6	45	Radar Dist and Sfc Obs spacing
RaNr0	Radar only	20	185	Imperfect model with $10 \times N_{r0}$
RaLFO	Radar only	20	185	Imperfect model with LFO84
RaSchultz	Radar only	20	185	Imperfect model with Schultz
RaWSM6	Radar only	20	185	Imperfect model with WSM6
RaSfcNr0	Radar + Sfc All	20	185	Imperfect model with $10 \times N_{r0}$
RaSfcLFO	Radar + Sfc All	20	185	Imperfect model with LFO84
RaSfcSchultz	Radar + Sfc All	20	185	Imperfect model with Schultz
RaSfcWSM6	Radar + Sfc All	20	185	Imperfect model with WSM6
RaSfcNr0	Radar + Sfc All	6	185	Imperfect model with $10 \times N_{r0}$
RaSfcS6LFO	Radar + Sfc All	6	185	Imperfect model with LFO84
RaSfcS6Schultz	Radar + Sfc All	6	185	Imperfect model with Schultz
RaSfcS6WSM6	Radar + Sfc All	6	185	Imperfect model with WSM6

88D precipitation scanning mode is assumed (see Fig. 2 of XTD06). Radial velocity (V_r) is simulated as in XTD06, and the same V_r observation operator is used in the EnKF assimilation. The reflectivity formula described in TX05 is used for simulation and assimilation. This formula returns the reflectivity (Z) in dBZ from the mixing ratios of rain, snow and hail/graupel. More details of the radar emulator can be found in XTD06.

Gaussian-distributed random errors with zero mean and standard deviations of 1 m s^{-1} and 3 dBZ are added to simulated V_r and Z , respectively. The 3 dBZ standard deviation of Z error is smaller than the 5 dBZ used in TX05 and XTD06, but is suggested to be more appropriate by Xue et al. (2007); the value is also used by Tong and Xue (2008a). Using the larger error value results in slightly worse analyses.

Surface stations of a hypothetical mesoscale observing network are located inside the $64 \times 64 \text{ km}$ interior sub-domain with station spacings of about 20 km in the control and other directly related experiments, resulting in a total of 9 surface stations. To simulate a network whose stations are not on the grid point while keeping the network more or less uniform, the stations are placed randomly within $400 \times 400 \text{ m}$ square boxes that are centered on the grid points 20 km apart. The EnKF code is general; it does not require a uniform distribution of stations.

The observed variables at these stations include the horizontal wind components, the temperature, pressure and water vapor mixing ratio. The standard deviation of the zero-mean Gaussian errors added to the simulated surface observations are: 1 m s^{-1} for wind components, 1 K for temperature, 1 hPa for pressure and 1 g kg^{-1} for water

vapor mixing ratio. In sensitivity experiments, different network densities are tested (see Table 1).

Both radar and surface observations are assumed to be available every 5 min. The latter is actually true with the Oklahoma Mesonet while the WSR-88D radar volume scan interval in precipitation mode is typically 5 min.

2.4 The EnKF algorithm

The EnKF algorithm used in this study is based on the serial ensemble square root filter (EnSRF) of Whitaker and Hamill (2002). In this study, we refer to the algorithm as EnKF in general. The implementation follows XTD06 exactly, except for the addition of surface observations, which are assimilated after the radar data (results obtained assimilating surface observations first are very similar).

The state variables analyzed include u , v , w , θ , p , q_v , q_c , q_r , q_i , q_s and q_h . To initialize the first ensemble forecast cycle, random initial perturbations as used in XTD06 are added to a horizontally homogenous first guess defined using the May 20 sounding of the truth simulation. The random perturbations are drawn from a Gaussian distribution with zero mean and standard deviations of 2 m s^{-1} for u , v , w , 2 K for θ and 0.6 g kg^{-1} for q_v and all microphysical variables. Perturbations for all except for the microphysical variables are added in the entire subdomain. The perturbations for the latter are added only in the region where radar echo is present at 20 min, the start time of the first assimilation cycle. Reflectivity data in both the precipitation and clear air regions (negative Z values are set to zero) are used. Radial velocity data are only used in regions where Z is greater than or equal to 10 dBZ. We note here that because the storm environment is initialized with a perfect model sounding and there is no land surface process in the model, there is no error in the storm environment except for that introduced by the initial ensemble perturbations. The benefit of the surface observations is expected only where storm-induced disturbance exists.

Covariance localization (Houtekamer and Mitchell 2001) is used to limit the spatial influence of observations and reduce sampling error. A Schur product is applied with a smooth 5th-order distance-dependent function (Eq. 4.10 of Gaspari and Cohn 1999) multiplying the calculated background error covariance. For radar observations, a 6 km localization radius is chosen in all directions to ensure the best results when only radar data is used (XTD06). For surface observations, when the mean station spacing (station spacing hereafter for short) is 20 km, a horizontal localization radius of 30 km and a vertical radius of 6 km are found to be optimal through experimentation. When the station spacing changes from 2 km through 32 km in sensitivity experiments, different optimal horizontal localization radii ranging between 6 km and

36 km are chosen, again based on experimentation. Further discussions can be found in Sect. 4. Forty ensemble members are used for all experiments, as in XTD06.

To avoid the filter divergence problem caused by underestimation of covariance due to small sample size and/or model error, covariance inflation is used in all experiments following the procedure of TX05. For cases using radar observations only, covariance inflation is only applied in and near (within 6 km of) the region where observed reflectivity exceeds 10 dBZ. This procedure is applied to avoid amplifying spurious cells in precipitation-free regions (XTD06). If a uniform inflation is applied everywhere in the model domain but there is not much data to constrain the ensemble spread (i.e., in the precipitation free region that is also free of spurious convection), the ensemble spread can grow too large, causing unphysical and unstable solutions. In experiments with only radar observations, a 5% inflation factor is applied for perfect-model experiments and 15% is used for imperfect-model experiments (see Sect. 6). For experiments including surface observations, additional covariance inflation is applied at the lower levels within the cutoff radius of the surface observations, in the entire subdomain, using the same inflation factor (at pointed influenced by both radar and surface data, covariance inflation is done only once). Therefore, covariance inflation in regions not directly reachable by observations is avoided. This avoids excessive increase in the ensemble spread where the ensemble is not constrained by any observations.

2.5 Assimilation experiments

A complete list of experiments can be found in Table 1. The control experiment, Cntl or Ra (for radar data only), uses only data from a radar located 185 km from the domain center where the main storm is located; this case serves as the baseline. Other experiments add data from a surface network of different station spacing and test radar distances of 185, 115 or 45 km, respectively. Some experiments assimilate a particular surface observation variable. In the experiment names, 'Ra' denotes the use of radar data, 'Sfc' indicates the use of surface observations, 'D' followed by a number indicates the radar distance and 'S' followed by a number indicates the station spacing. For example, RaSfcD115S6 means that both radar and surface data are used and the radar is located 115 km from the domain center and the station spacing is 6 km. Additional characters, such as T in RaSfcT, indicate which variables in the surface data are assimilated. The results of these experiments are reported in the next three sections, focusing on the impact of additional surface observations. The experiments with the default radar distance of 185 km and default mean station spacing of 20 km have

abbreviated names. Experiments discussed in Sect. 6 include model error. In all experiments, the first assimilation cycle starts at 20 min and the first EnKF analysis occurs at 25 min.

3 Impact of surface observations

3.1 Results of experiments Ra and RaSfc

The control experiment Ra (Table 1) is first examined, together with RaSfc. Figures 1 and 2 show the surface fields of the truth simulation and the ensemble mean analyses of experiments Ra and RaSfc, at 60 and 120 min of model time, respectively. RaSfc adds data from a surface network of 20 km mean spacing with all observed variables (u , v , T , p and q_v) assimilated. The ensemble mean analysis fields are plotted. At 60 min, or after eight 5-min analysis cycles, precipitation and cold pool associated with the storm are established in both Ra and RaSfc in general (Fig. 1b, c) but Ra does not capture the cell splitting or the left moving cell at this time. The cold pool is too broad in Ra based on the -1 K perturbation potential temperature, θ' , contour (Fig. 1b) but is much closer to the truth in RaSfc (Fig. 1c). Ra completely misses the surface convergence center associated with the main updraft while the divergence center underneath the precipitation core is too weak and does not have the right pattern (compare Fig. 1e with d). In comparison, RaSfc does a much better job capturing the convergence pattern along the gust fronts on the south and north side and reproduces the magnitude and pattern of the main divergence center much better (Fig. 1f). It does significantly underestimate the convergence associated with the main updraft. In both cases, the analyzed fields are not very accurate quantitatively at this time.

At 90 min (not shown), the right and left moving cells as seen from the low-level Z are now reasonably well captured in both cases. However, the convergence field is still not accurately analyzed, especially along the gust fronts and in Ra. By 120 min (Fig. 2), the hook echo structure of the major storm is reproduced well in both Ra and RaSfc (Fig. 2b, c) but we see more differences in the analyzed surface divergence fields. Both the intensity and pattern of the divergence in RaSfc are reproduced accurately (Fig. 2f) but the divergence in Ra is generally broader and weaker. The tail of the divergence band at the southwest end of the gust front is missing in Ra. Also mostly missing is the convergence center associated with the left mover near the northwest corner of the plotted domain (Fig. 2e).

There are also clear differences between the analyses of gust front. The temperature gradients along the gust fronts are noticeably weaker in Ra (Fig. 2b) than in RaSfc (Fig. 2c), and there is also larger position error of the gust

fronts in Ra, especially near the western and northern domain boundaries. In general, the low-level analysis of RaSfc can be considered very good at this time.

The subjective comparison between Ra and RaSfc indicates that the surface observations have a noticeable positive impact on the analysis of low-level features, including the cold pool and associated flows, even though the network has a much coarser resolution (~ 20 km) than would be needed to resolve the sharp gradient along the gust fronts. Our results agree with Zhang et al.'s (2004) study, which found the addition of surface observations retrieves the low-level cold pool better with a denser surface network of 10 km spacing. The impact on the precipitation field seems smaller, presumably because of that the hydrometeors that descend to the ground level are well captured by the radar data above. Without surface data, we have to rely on the model alone to establish the cold pool through microphysical and dynamic processes. As the cold air spreads away from the precipitation region, no more radar data (except for the clear air information contained in the zero reflectivity values) are available.

The root mean square (*rms*) error of the ensemble mean analysis calculated against the truth is used to further quantify EnKF performance and the impact of surface data. The evolution of *rms* error with time is shown in Fig. 3 for Ra and RaSfc. As in TX05, the errors were calculated at grid points where observed Z exceeds 10 dBZ. The errors decrease during the first few cycles rapidly, from the initially very high level associated with the poor sounding-based initial guess. Errors starting from 40 min, or after four analysis cycles, are shown here. The relative error ratio (RER) averaged over the last ten analysis cycles ending at 120 min is also calculated to evaluate the impact of surface observations. It is defined as

$$\text{RER} = \frac{1}{10} \sum_{i=n}^{n-9} \frac{E_{\text{Ra+Sfc}}^i}{E_{\text{Ra}}^i},$$

where $E_{\text{Ra+Sfc}}^i$ is the analysis *rms* error of a given state variable at the i th cycle when using both radar and surface observations and E_{Ra}^i denotes the corresponding error using radar data only. The averaging over the last ten analysis cycles removes temporal fluctuations and provides a more reliable measure of the analysis accuracy. The RERs of various state variables can be combined further to form RERs for several major categories, such as that for the wind, which averages the RERs of u , v and w and that for microphysical variables, which averages the RERs of 5 microphysics variables. The total RER (TRER) averages the RERs of all 11 state variables and is used as the main indicator quantifying the improvement by surface observations. The RERs or TRERs of all experiments are

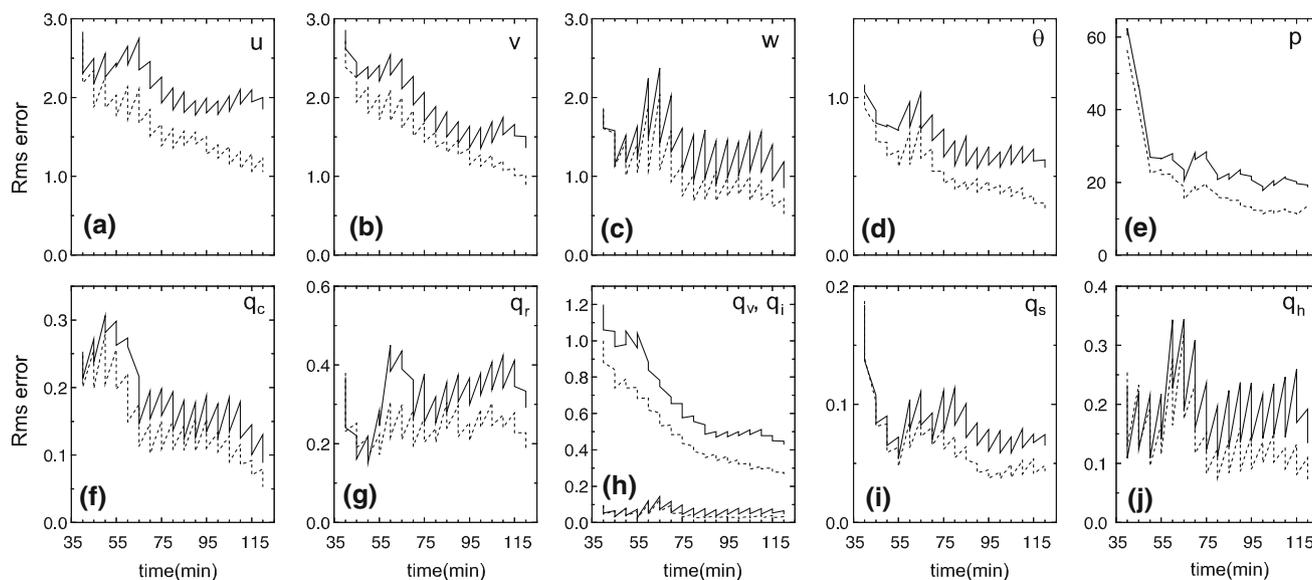


Fig. 3 The *rms* error of ensemble mean forecasts and analyses plotted against time for Ra (*solid*) and RaSfc (*dotted*) for **a** u (m s^{-1}), **b** v (m s^{-1}), **c** w (m s^{-1}), **d** potential temperature θ (K), **e** pressure p (Pa), **f** q_c (g kg^{-1}), **g** q_r (g kg^{-1}), **h** (*upper curves*) q_v and (*lower curves*) q_i (g kg^{-1}), **i** q_s (g kg^{-1}), and **j** q_h (g kg^{-1}).

The sharp reductions in the error at the analysis times are due to analysis updates. See TX05 for further explanations of this type of plots

summarized in Table 2. A similarly defined error ratio has been used in Xu et al. (2008).

At the end of assimilation, or 120 min, the analysis errors in most fields are clearly lower in RaSfc than in Ra (Fig. 3). The percent improvement due to surface data is the largest in u and θ , with the errors in RaSfc being only about 50–60% those of Ra. The absolute analysis *rms* error reduction in RaSfc is 0.8 m s^{-1} for u and 0.27 K for θ . Other variables also display improvements to various

extents. The relative *rms* error in RaSfc is 65% of Ra for v , 61% for w , 73% for p and 60% for q_v . Even for microphysical variables which are not directly observed by surface observations, significant improvements are also found: the *rms* error is 60% of Ra for q_c , 62% for q_r , 48% for q_i , 62% for q_s and 54% for q_h . This amounts to a 30–50% reduction in errors in various model fields due to the assimilation of surface observations. We note that there are temporary increases in the *rms* errors near 65 min; this also occurs in our earlier studies for this storm (TX05 and XTD06) and we believe it is related to the cell splitting at this time. The RER of RaSfc also shows evident improvement in the last ten analysis cycles (Table 2). The reduction in the total relative error ($1 - \text{TRER}$) is 32% with the help of surface observations.

To examine the impact of surface data at different levels, vertical profiles of *rms* errors obtained by horizontal averaging over the precipitation regions are plotted in Fig. 4 for 120 min. The *rms* errors for u , v , θ , p , q_r and q_v are the largest at the low levels and near the surface in Ra, primarily because of the lack of low-level radar data. The largest improvements in RaSfc for these variables are also found at the surface. For example, the *rms* error reductions are 2.6 m s^{-1} for u , 0.8 K for θ and 12 Pa for p at the surface, reflecting the better analysis of the surface cold pool and the associated fields with the use of surface observations. For w and all four microphysical variables shown (q_c , q_r , q_s and q_h), the largest *rms* error reduction is generally found at the levels where the corresponding Ra errors are the largest. For example, q_h error is reduced from

Table 2 Relative *rms* error ratio (RER) of u , v and w , for microphysical variables, θ , q_v and all variables (total) for listed experiments

Experiment	uvw	Micro	θ	q_v	Total
RaSfc	0.72	0.67	0.67	0.67	0.68
RaSfcUV	0.80	0.76	0.73	0.76	0.77
RaSfcT	0.84	0.75	0.80	0.86	0.80
RaSfcP	0.93	0.90	0.91	0.93	0.91
RaSfcQv	0.86	0.79	0.77	0.80	0.81
RaSfcS32	0.85	0.83	0.78	0.80	0.83
RaSfcS16	0.66	0.64	0.62	0.61	0.64
RaSfcS12	0.61	0.62	0.55	0.56	0.60
RaSfcS6	0.50	0.55	0.48	0.43	0.51
RaSfcS4	0.47	0.53	0.45	0.39	0.48
RaSfcS2	0.47	0.56	0.45	0.38	0.49
RaSfcD115	0.71	0.71	0.72	0.68	0.70
RaSfcD115S6	0.52	0.61	0.56	0.50	0.56
RaSfcD45	0.97	1.04	0.96	0.95	0.99
RaSfcD45S6	0.85	0.96	0.89	0.82	0.89

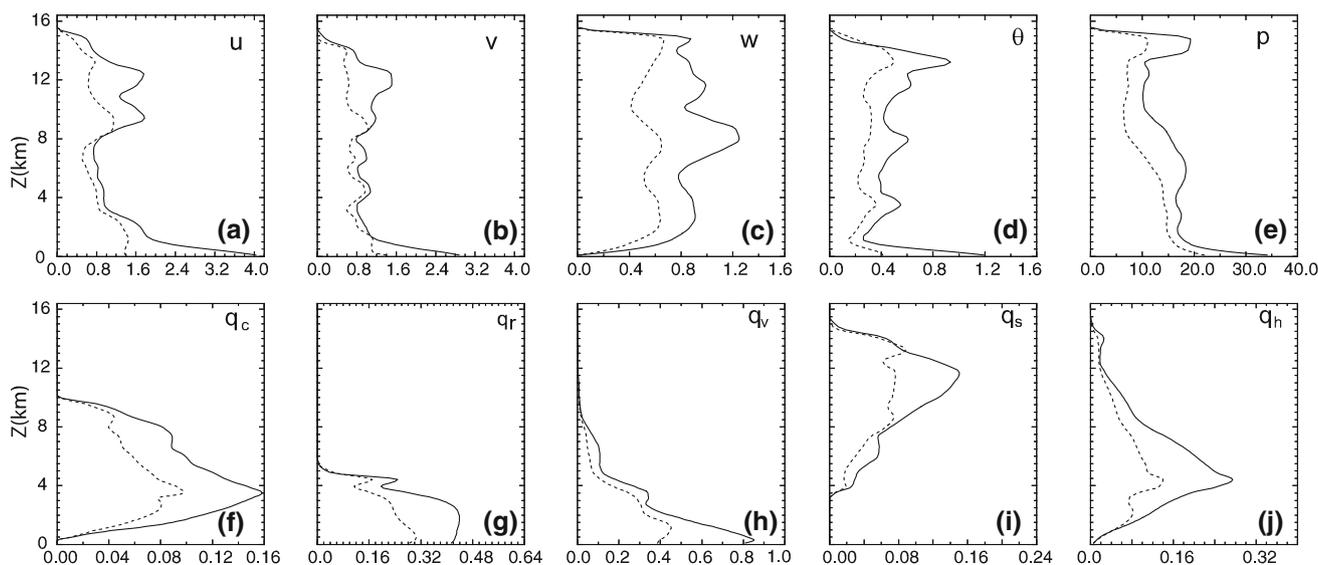


Fig. 4 The *rms* error profiles of the ensemble mean analyses of Ra (*solid*) and RaSfc (*dotted*), for **a** u (m s^{-1}), **b** v (m s^{-1}), **c** w (m s^{-1}), **d** θ (K), **e** p (Pa), **f** q_c (g kg^{-1}), **g** q_r (g kg^{-1}), **h** q_v (g kg^{-1}) and q_i (g kg^{-1}), **i** q_s (g kg^{-1}), and **j** q_h (g kg^{-1}) at 120 min

about 0.28 g kg^{-1} to about 0.14 g kg^{-1} near the 4 km level, and the maximum reduction in q_s error, about 0.08 g kg^{-1} , occurs close to 12 km. The error reduction in w is between 0.3 and 0.7 m s^{-1} between the levels of 2 and 13 km. This general improvement in the storm analysis through the addition of surface data, even at the upper levels away from the ground, is very encouraging.

3.2 Background error correlation structure

To gain insight into how the surface data affect the model state, we examine a few fields of ensemble-derived background error correlation (ρ) between a hypothetical surface observation (called the observation prior, not the observation itself, in data assimilation terminology) and a state variable. The background error covariance estimated from the ensemble members determines how observation information is spread in space and across variables and thereby helps improve the analysis of variables not directly observed. Accurate error correlations should reflect the physical structure of the storm.

We choose two locations of possible surface observations at 75 min (Fig. 5) to calculate the spatial forecast error correlations. The first location, O1, is in the inflow region of the storm, at (73, 65) km, and the second location, O2, is in the outflow region within the cold pool, at (57, 59) km (Fig. 5a). The surface inflow through O1 feeds the convergence center marked as CC and the outflow through O2 originates from the surface divergence center marked as DC in Fig. 5a. As in TX05, we calculate the correlation coefficients between an observation prior (which is linked to the state via an observation operator) at a given

observation location, and a state variable at all other grid points. When these two variables are the same, we are calculating the spatial auto-correlation coefficient, and when they are different, we obtain a spatial cross-correlation. Such correlation information, after being localized, is used directly in the EnKF algorithm for analysis update.

We choose two vertical cross sections, along lines A-B and C-D, as shown in Fig. 5a. The horizontal wind vector is projected onto each of these two lines (positive towards B and D, respectively), and we calculate the correlation between this wind component (referred to as U) with other state variables. The projected wind component U instead of actually assimilated wind components is chosen just for convenience of discussion.

We first present the ρ field at the surface, between U at location O2 (wind along line C-D) and u at the grid points (Fig. 5a), and between surface temperature at O2 and u at the grid points (Fig. 5b). Positive U - u correlations are found to occupy a large part of the cold pool outflow region and extend out to 20–30 km from O2 (Fig. 5a) and more so in the generally westward direction. This positive correlation indicates that when U is smaller (more negative) at O2, u is smaller (more negative), which is physically reasonable because these U and u are located within the same outflow region; a stronger outflow observed at O2 suggests stronger outflows at locations with positive correlations. To the northeast of O2, correlation contours are more closely packed, and the correlation coefficient becomes negative immediately beyond the divergence center (but before going into the inflow region). This demonstrates that the outflow on the other side of DC is also stronger when that on this side of divergence center near O2 is stronger.

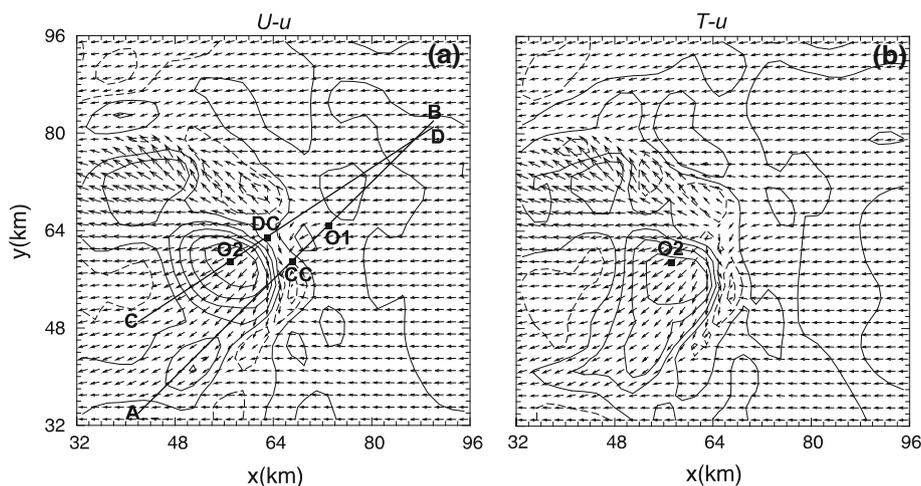


Fig. 5 The surface truth wind vectors, together with forecast error correlation coefficients estimated from the forecast ensemble at 75 min for experiment RaSfc. Error correlation between **a** a hypothetical surface wind observation U (wind along C-D line, see text for reference) at station O2 ($x = 57$ km, $y = 59$ km) and u at the grid points, and **b** between a hypothetical surface temperature observation

at station O2 and u . Solid (dashed) contours represent positive (negative) correlations at intervals 0.2; zero contours are omitted. CC and DC in **a** mark the low-level convergence and divergence centers, respectively, and lines A-B and C-D indicate location of the vertical cross sections shown in Fig. 6

The correlation between surface temperature T at O2 and u at the grid points shows a similar pattern as that between U and u (Fig. 5b), which shows that a stronger colder cold pool corresponds to stronger outflows, with more negative u west of DC and less positive u east of DC. When such cross-correlation is estimated with reasonable accuracy, surface temperature observations can be used effectively to update the flow field, and vice versa. The generally well-defined correlation patterns up to 30 km from the observation location suggest that the localization radius of 30 km used for the current surface network is appropriate.

In the upper panels of Fig. 6, the correlations between surface observations at O1 and other state variables in the vertical cross-section along line A-B, through O1 and CC, are shown (in contours), together with the corresponding state variable fields (shaded contours). For the correlation between T at O1 and vertical velocity w , the most prominent feature is a deep column of positive correlation coinciding with the main updraft (Fig. 6a), suggesting the presence of a stronger updraft when surface T in the low-level inflow region is higher. A similar pattern is found between U at O1 and w , suggesting a stronger updraft when the surface inflow towards CC is stronger (Fig. 6d); again, this is physically consistent with the expected storm dynamics. The negative correlation with potential temperature, θ , in the updraft region indicates enhancement of the warm core structure by the low-level inflow (Fig. 6b). The correlation of U with u is negative in the lower part of the troposphere and positive above (Fig. 6c), suggesting that a stronger easterly component at the surface contributes to

more westward tilt of the updraft at the low levels while at the upper levels this enhanced easterly component may turn into an enhanced westerly component as the updraft air parcels rise and overturn.

The corresponding correlation fields in the C-D cross section are plotted in the lower panels of Fig. 6. In this case, the cross section passes through the surface divergence center (DC) with the observation station located inside the cold pool (O2). Again, we find similar correlation patterns between surface T and U with w (Fig. 6e, h). The strong positive correlations between U , T in the downdraft outflow region and w in the updraft region, extending all the way from the surface to about 9 km, indicate a stronger downdraft (or weaker updraft) when temperature at O2 inside the cold pool is lower, or when the outflow near the surface is stronger. The positive correlation between U and θ in the downdraft region (Fig. 6f) also reflects the relation between the outflow strength and downdraft temperature. The correlation between U and u in this cross section changes sign at the middle troposphere (Fig. 6g), similar to what is observed in the A-B cross section (Fig. 6c).

The physically meaningful and dynamically consistent correlations of surface temperature and wind with the state variables presented here, taken in both the inflow and outflow regions, indicate that the ensemble system is able to properly estimate the spatial auto- and cross-correlations. Such flow-dependent error correlation information is valuable for optimally utilizing surface observations. The 6 km covariance localization radius that we use in the vertical allows the surface observations to directly

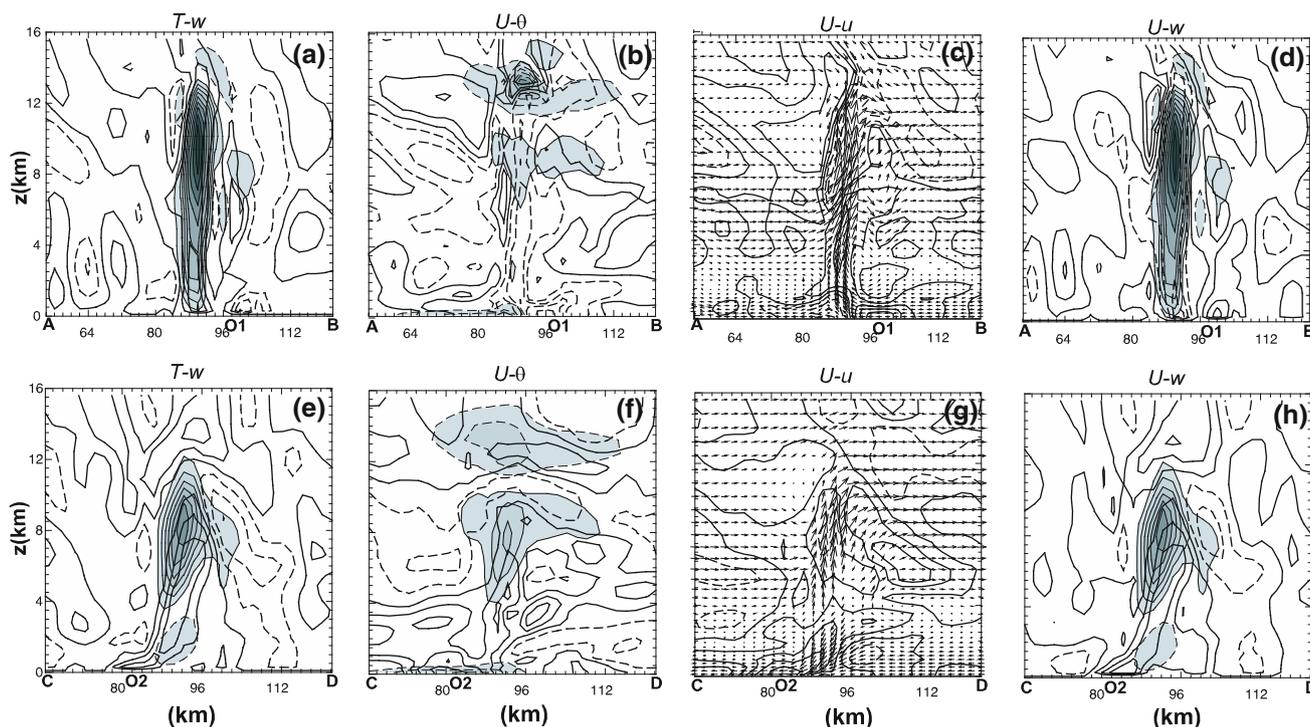


Fig. 6 Forecast error correlation coefficients estimated from the forecast ensemble at 75 min for experiment RaSfc, in the vertical cross section along line A-B (upper panels) and along C-D (lower panels) in Fig. 5. Error correlations (upper panels) between **a** hypothetical surface temperature observation T at O1 ($x = 73$ km, $y = 65$ km) and w , **b** hypothetical surface wind observation U at station O1 and θ , **c** U and u , **d** U and w . Error correlations (lower

panels) between, **e** surface T at station O2 ($x = 57$ km, $y = 59$ km) and w , **f** U at O2 and θ , **g** U and u , and **h** U and w . Thick solid (dashed) contours represent positive (negative) correlations at intervals 0.2. Shaded with thin contours shows the truth field of w in **a**, **d**, **e** and **h** with interval 2.5 m s^{-1} , and perturbation θ' in **b** and **f** with interval 2 K. Wind vectors in **c** and **g** show the truth perturbation wind field

influence state variables up to 6 km above the surface. Of course, the impact of surface observations is not limited to the lowest 6 km (see Fig. 4), because of information propagation in time and space by the prediction model.

There is also some correlation patterns with medium magnitude contained in Fig. 6 which are far away from surface stations and appear not understandable with our current knowledge about supercell thunderstorm dynamics. We believe some of these patterns might be spurious due to the sampling error caused by the small ensemble number. The application of covariance localization helps to alleviate the sampling error by limiting the impact of surface observations to nearby state variables. The cutoff radius of 30 km in horizontal and 6 km in vertical prevent most of the spurious covariance structure affecting the analysis.

4 Sensitivity experiments

In this section, we will further examine the impact of different observation variables or types and the impact of surface data for different network densities and/or radar distances.

4.1 Impact of surface measurement type

In this sub-section, the surface measurement types or variables, including wind, temperature, pressure and water vapor, are assessed individually for their impact. The experiments are listed as RaSfcUV, RaSfcT, RaSfcP and RaSfcQv in Table 1 and are the same as RaSfc except that only the listed surface observation variables are included.

The relative analysis error plots for these four experiments are shown in Fig. 7. The relative error is defined as the ratio of the *rms* error of the individual sensitivity experiment to that of the chosen reference, which in this case is Ra (c.f., Fig. 3). The plots show that surface wind observations have the largest impact while pressure observations have the least impact; the relative *rms* errors of RaSfcP are close to 1 at the end of assimilation cycles. In this case, the RERs or TRER are 93–90% (Table 2), indicating only 7–10% error reduction when only additional pressure observations are assimilated. For microphysical variables, temperature observations result in the smallest RER (Table 2). None of these individual measurements were able to produce as large an impact as RaSfc

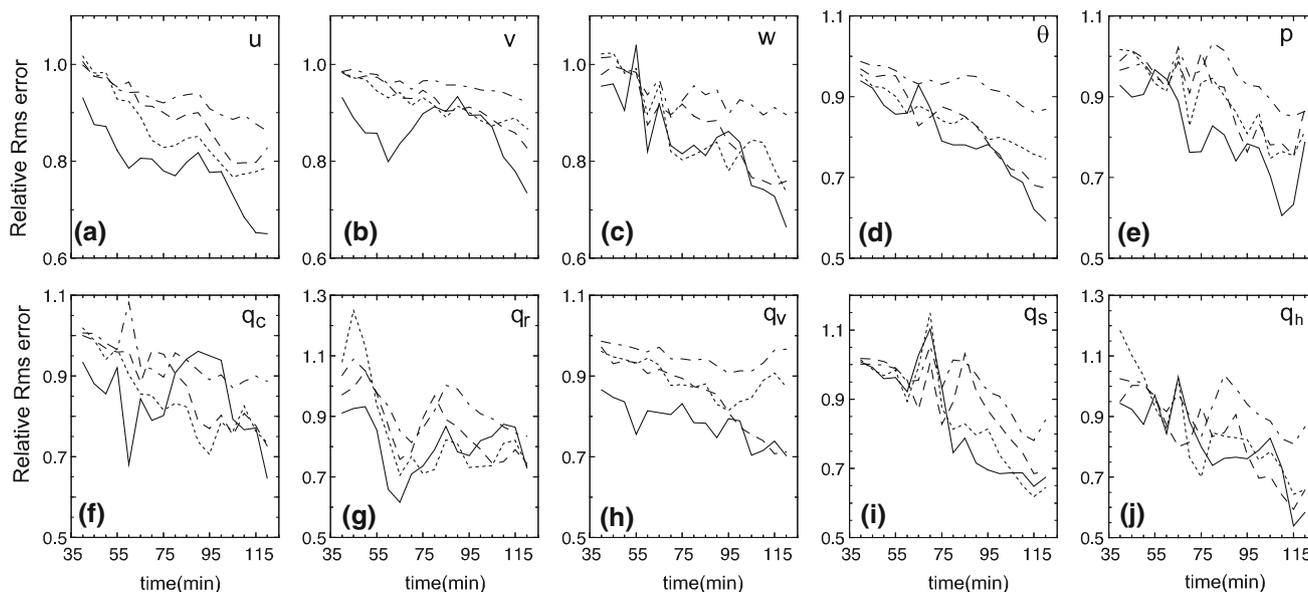


Fig. 7 Analysis *rms* errors relative to those of Ra for RaSfcUV (solid), RaSfcT (dotted), RaSfcP (dash-dotted) and RaSfcQv (dashed), as a function of analysis time

which assimilates all surface variables; it achieves a TRER of 68%.

It is noted that the largest improvements for state variables θ , p and q_v are also obtained when assimilating surface winds rather than direct measurements of these variables themselves. This seems to be at least partly because *rms* errors for θ , p and q_v in Ra are already much smaller than the standard deviations of the corresponding surface observation errors (Fig. 3), which are 1 K for T , 100 Pa for p and 1 g kg^{-1} for q_v . Such error-containing observations have a limited ability to further reduce the errors of the corresponding state variables (however, keep in mind that local errors in θ , p and q_v , especially underneath the active storms, are larger than those of precipitation-domain mean shown in Fig. 3, so the benefit of analyzing surface data should still be achievable). For u and v , the analysis errors from Ra are still relatively large (Fig. 3) compared with the surface wind observation errors; therefore, there is more room for direct positive impact by the wind observations. When wind fields are improved, other variables benefit too. In general, θ benefits from surface observations most among all state variables; this is consistent with the fact the most important features at the surface are related to the cold pool. The improvement in analyzed cold pool temperature comes not only from thermodynamic observations, but even more so from surface wind observations, which directly improve the analysis of convergence and divergence patterns that are closely linked to updraft and downdraft intensities as well as gust front positions.

It is expected that the impact of surface observations is the largest when all of the measured variables are assimilated, since in this case every individual measurement shows benefit. The total impact using all measurements together is smaller than the sum of individual impacts, in terms of the percentage error reduction.

Zhang et al. (2004) show that liquid–water potential temperature observations have a larger impact than winds in retrieving the cold pool, which is different from our conclusion. Since liquid–water potential temperature is typically not measured, we cannot directly compare our results with theirs. Simplistic radar data coverage assumption, the use of V_r data only and the simple microphysics used in Zhang et al. (2004) further hinder a direct comparison in details.

In summary, the surface wind observations are found to have the largest impact on the analysis of wind and temperature fields for the chosen supercell and for the typical observation errors assumed. The potential temperature fields benefit not only from thermodynamic observations but also from wind observations. Such inter-connections are reflected in the background error covariance derived from the forecast ensemble, which drives the multivariate EnKF analysis. Wind observations are used to update all state variables, not just wind components themselves.

4.2 Varying surface network spacing

A series of experiments are conducted to examine the impact of surface network spacing on the analysis.

Additional experiments with network spacing ranging from 32 km to 2 km are performed (Tables 1, 2). In these experiments, surface observations are still approximately uniformly distributed in the central subdomain with 4, 16, 25, 100, 256 and 961 surface stations in RaSfcS32, RaSfcS16, RaSfcS12, RaSfcS6, RaSfcS4 and RaSfcS2, respectively.

For each network spacing, experiments are conducted to determine the optimal horizontal covariance localization radius. The optimal localization radius is plotted against the network spacing in Fig. 8a. As the spacing decreases, the optimal radius also decreases. When a denser network is employed to resolve smaller scale features, reducing the

radius has the effect of keeping the influence of each observation more localized. The sampling error in the ensemble filter can also be decreased when a smaller localization radius is used (Anderson et al. 2005).

We first examine the cases of medium- and high-density observation networks with 12 and 6 km grid spacings. In RaSfcS12, for u , v , θ' , p' , q_v , q_c , q_r and q_s , the relative analysis *rms* errors decrease faster and reach lower levels than those in RaSfc by 90 min (Fig. 9). After 90 min, as the left-moving cell propagates quickly towards the northwest corner of the subdomain, the *rms* errors start to grow in RaSfcS12 and become close to those of RaSfc for some variables by the end of the assimilation period; this corresponds to the time when the number of surface observations (which are confined to the subdomain) covering the left-moving cell decreases.

When the number of surface stations is increased to 100 with a spacing of 6 km in RaSfcS6, the *rms* errors of the variables directly observed (u , v , θ , p , and q_v) become consistently smaller than in RaSfc and RaSfcS12. For q_c , q_h , q_r and w , this decreased spacing of 6 km leads to an improvement at most analysis times, but a slight degradation in the last couple of cycles as compared with the cases using coarser-resolution observing networks. Even so, the improvement over the radar-only case, Ra, is still significant after the last cycle, with a 10% to 45% reduction in error for the variables. In the last 2–3 cycles, the analysis error is already small enough, especially at the surface, which makes improvement from surface observations more difficult and might be the reason why the error of the dense network from last few cycles approach or surpass those of the sparse ones. The degradation of the analysis with increasing network density in some later cycles for some of the indirectly observed variables also reflects the nonlinear nature of the storm-scale data assimilation problem. It is also possible that the ensemble spread during the final few cycles is too small. Future studies should examine more effective ways of adaptively adjusting the inflation rather than using ad-hoc methods. Relative to the generally large improvement in other variables, and during other cycles, the amount and extent of the degradation are small.

The TRERs of the above experiments are plotted against the observation spacing in Fig. 8b. The TRER decreases roughly linearly as network spacing decreases (network resolution increases), until the network spacing is close to the grid interval of truth simulation. Given that the number of surface observations is much smaller than the number of radar observations (which is kept fixed), such a strong dependence of analysis accuracy on surface network density is rather interesting; it suggests that the impact of the surface network on the analysis of convective-scale features becomes increasingly large as the density of the network increases.

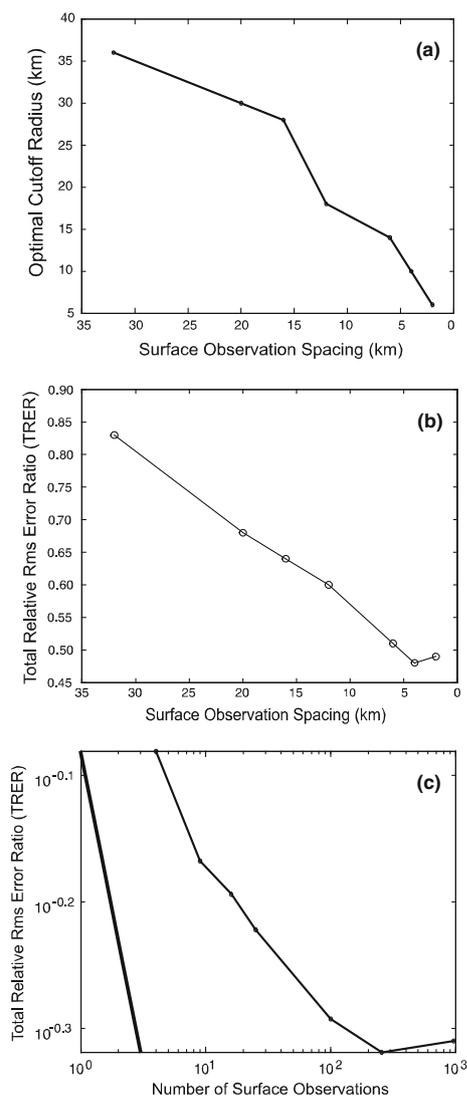


Fig. 8 **a** The optimal horizontal covariance localization radius for surface data as a function of mean surface network spacing, **b** the total *rms* error ratio (TRER) as a function of the mean surface network spacing, and **c** the TRER as a function of the number of surface observations, plotted in a logarithmic space. The thick straight line in **c** represents the $-1/2$ slope

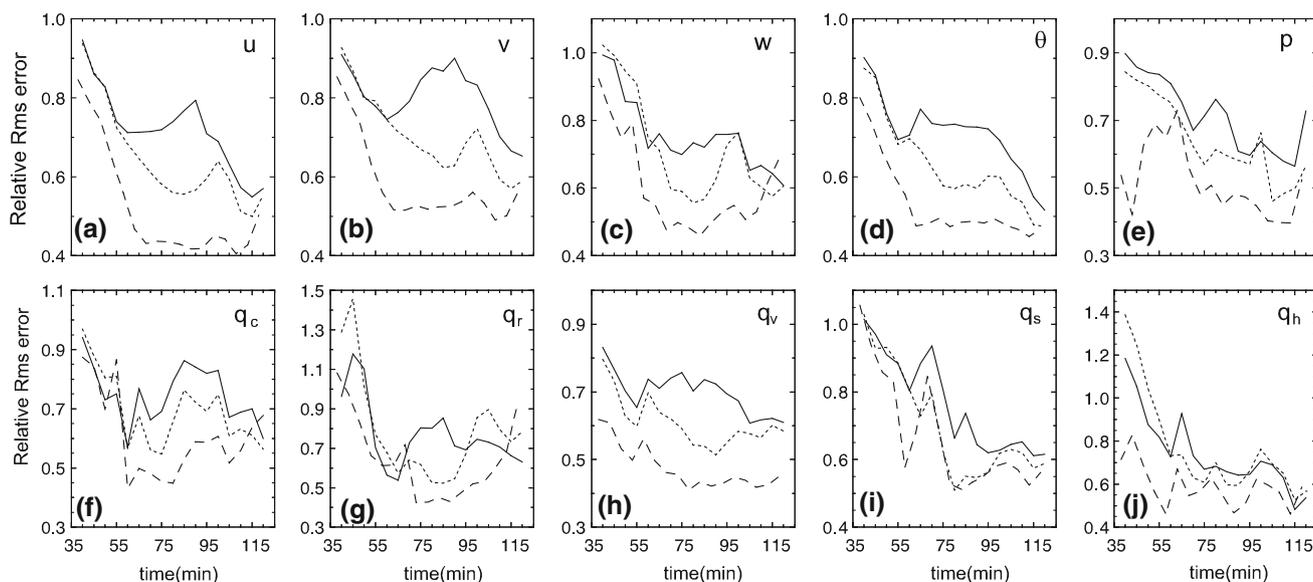


Fig. 9 As Fig. 7 but for RaSfc (*solid*), RaSfcS12 (*dotted*) and RaSfcS6 (*dashed*). All are errors relative to those of Ra

Since Fig. 8b shows that $\text{TRER} \propto d$, where d is the mean network spacing, it follows that $\text{TRER} \propto A n^{-1/2}$ and $\log(\text{TRER}) \propto -1/2 \log(n)$, where A is the area covered by the network and n is the number of observation stations. Morss et al. (2001) point out that such a power-law behavior is expected of a network where the number of observations is much greater than the number of degrees of freedom in the analysis. This condition is obviously not met in our case, however, even if we include radar observations.

In Fig. 8c, we plot the TRERs against the number of surface observations in logarithmic space; the thick straight line represents a slope of $-1/2$. It can be seen that for relatively coarse network resolutions, the error reduction rate is close to the $-1/2$ power law and for higher resolutions, the rate of error reduction is slower than the power law suggests. Morss et al. (2001) and Anderson et al. (2005) also found steep error reduction when the observations are relatively sparse (but not too sparse); for relatively dense networks the error reduction is close to or somewhat shallower than the power law indicates. Our case is more complicated because of the constant presence of much denser radar observations assimilated before surface data. In addition of a different index to measure error reduction (we use relative error ratio to examine surface data impact, instead of absolute analysis error), it is not appropriate to compare our results with those previous researches directly. However, it is interesting to notice that in our case the general trend in the error curve is similar to those two studies. It suggests that there might be commonality between single- and multi-source data assimilation in the study of surface observation impact. Further investigation is needed in future research.

The above experiments clearly demonstrate the benefit of higher surface network densities for storm-scale analysis. For real storms that may contain even more small-scale structures than the simulated truth storm does, additional benefit could potentially be achieved with higher density networks.

4.3 Varying radar distance

In the next set of experiments, the radar is placed closer to the storm, at a distance of 115 km in RaD115, RaSfcD115 and RaSfcD115S6 and a distance of 45 km in RaSfcD45 and RaSfcD45S6 (see Table 1). Figure 10 shows the analysis *rms* errors of RaD115 relative to those of Ra, as well as those of RaSfcD115 and RaSfcD115S6 relative to those of RaD115. With the radar at a closer distance of 115 km, RaD115 is consistently able to produce better analyses than Ra, especially for microphysical variables (which are directly linked to Z measurements). The final error levels in RaD115 are lower for all variables. Relative *rms* errors are even smaller during the intermediate cycles than in the final analysis. Relative error reductions reach 10–20% compared with Ra for most variables, with the largest error reduction around 75% at 65 min for q_r . In fact, the error peaks found in many variables in Ra (Fig. 3) and other experiments of the same radar distance (not directly shown) at around 60 min are mostly gone, resulting in significantly smaller relative errors around this time (Fig. 10). The closer radar location provides more low-level data coverage and better vertical resolution at the location of storm. When the radar is located at 115 km from the main storm, the impact of additional surface

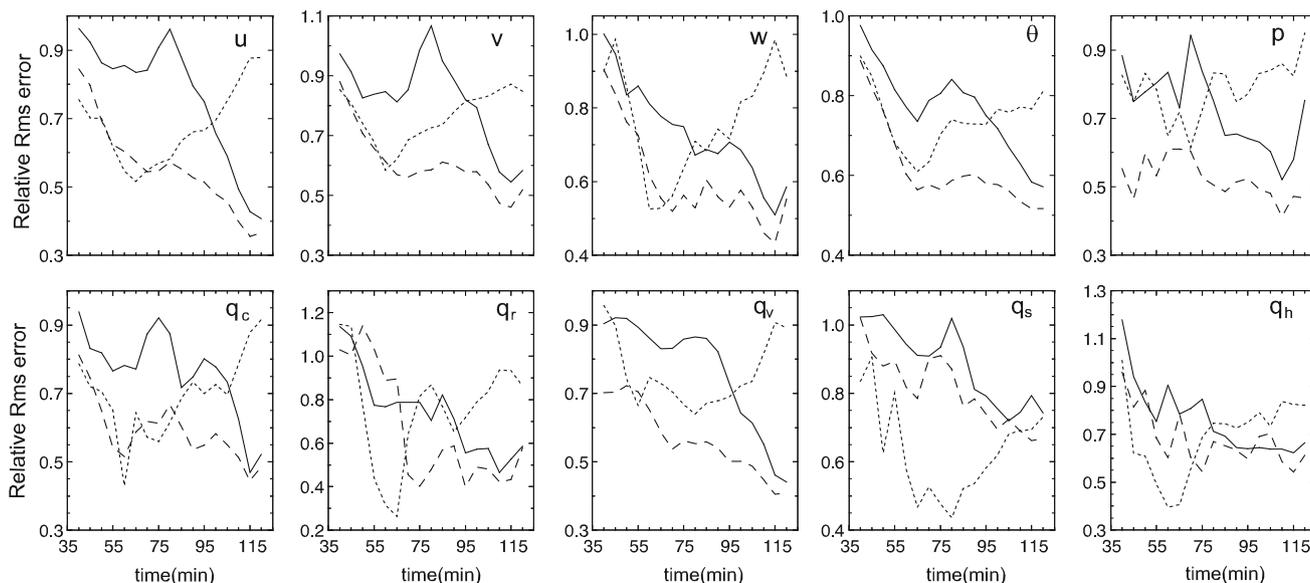


Fig. 10 Analysis rms errors of RaD115 relative to those of Ra (dotted), of RaSfcD115 relative to those of RaD115 (solid), and of RaSfcD115S6 relative to those of RaD115 (dashed)

observations with a mean spacing of 20 km is smaller during the early assimilation cycles (RaSfcD115 vs. RaD115), compared with the 185 km case (RaSfc vs. Ra), but the impact increases during the later cycles (compare the solid curves in Figs. 9 and 10 noting the difference in vertical axis scales), partly because of the error increase in the case using no surface data (RaD115). The TRER of 70% is slightly larger but still close to that of RaSfc (68%).

RaSfcD115S6, with a network spacing of 6 km, shows a much more pronounced impact of the surface data, starting from the early cycles (Fig. 10). After 90 min, both 20 and 6 km networks provide significant positive impacts, with the 6 km network providing the most. Unlike the case with the radar distance of 185 km (see Sect. 4.2), we observe no degradation in the analyses of microphysical variables as the spacing decreases in this case whereas the errors of some variables (w , q_r and q_s) with 6 km network spacing in the final few cycles still approach close to the 20 km case. It is not clear whether the degradation will happen if more assimilation cycles are performed after 120 min. In spite of the small error increase of some variables in last 1–2 cycles, the improvement from the dense network is still significant.

When the radar is located much closer to the storm center, at a distance of about 45 km, radar observations alone give very good analyses; there is not much room left for further improvement by surface observations of 20 km spacing, with the achieved TRER being 99%. The assimilation of surface observations at a spacing of 6 km in RaSfcD45S6 improves the results over those of RaD45, with an error reduction of 11% (TRER = 0.89, Table 2).

5 Impact of surface observations on forecast

As mentioned earlier, data assimilation aims to provide accurate initial conditions for the following forecasts. It has been shown that assimilation of surface observations helps to improve the analysis, which is used as the initial condition for the subsequent forecasts. To further examine the impact of surface observations on the ensuing forecast, we perform 1-h forecasts from the ensemble mean analyses of Ra, RaSfc and RaSfcS6, at 60, 90 and 120 min. We plot the average relative error ratios between the forecasts with and without surface data for groups of state variables, with station spacing of either 20 or 6 km. The groups of variables are wind components u , v and w , potential temperature and pressure and the microphysical variables.

Figure 11 shows that when the forecasts start at 60 min of model time and when the station spacing is 20 km, the RER (thick black lines) of the wind components grows quickly from the initial value of about 0.7 to near 1 at 100 min and exceeds 1 after 100 min. When the station spacing is 6 km, the corresponding RER (thick dashed lines) grows from about 0.56 at 60 min to near 0.75 by 90 min and remains approximately level afterward. The RERs for the microphysical variables have very similar trends (Fig. 11c) while those for θ and p show that the benefit of surface data is sustained for the entire period when 20 km station spacing is used, and even increases with time for 6 km station spacing (Fig. 11b).

It is clear from Fig. 12 that the 60-min forecast of RaSfc valid at 120 min is rather poor (Fig. 12c) and is similar to that of Ra (Fig. 12b); the reflectivity associated with the

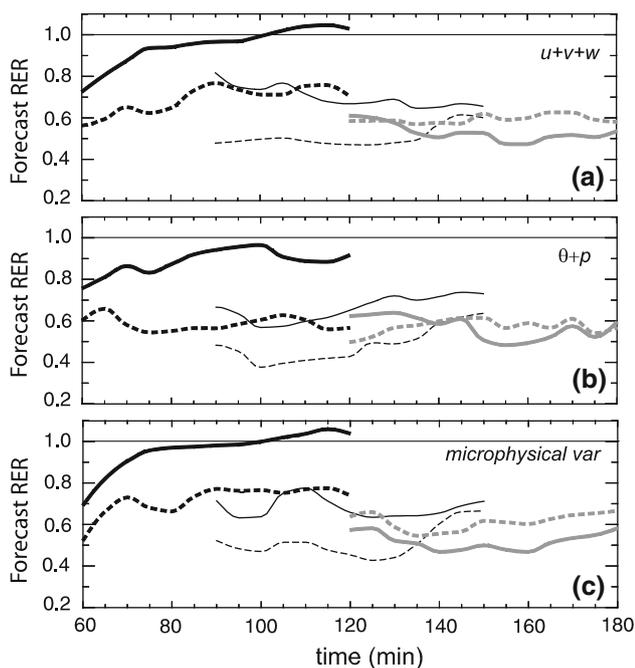


Fig. 11 The average relative *rms* error ratios of the u , v and w components (*upper panel*), θ and p (*middle panel*), and moisture and microphysical variables (*lower panel*), for 60-min-long forecasts starting from ensemble-mean analyses at 60 min (*thick black lines*), 90 min (*thin black lines*), and 120 min (*thick gray lines*). The *solid curves* are for the forecast errors starting from the analyses of RaSfc (with 20 km station spacing) relative to the corresponding errors of Ra (radar only), and the *dashed lines* are for the forecast errors of RaSfcS6 (6 km station spacing) relative to those of Ra

main cell near the center is completely missing in both cases. The corresponding forecast in RaSfcS6 is much closer to the truth (Fig. 12d vs. a); because the surface data are assimilated for a relatively short period of time (8 cycles over 40 min), the benefit of a denser network is much greater. The benefit of the 20 km network is effectively lost after a forecast period of 40 min; the relatively poor initial condition at 60 min appears to be the reason. For a denser 6 km network the benefit is sustained for a forecast period of 60 min and even increases slightly with time for thermodynamic variables (Fig. 11b).

When surface observations are assimilated until 90 min of model time, significant benefit of surface observations is sustained throughout the hour-long forecast period (thin black lines in Fig. 11), for both network densities. In this case, the RERs for the wind components and the microphysical variables decrease with time for much of the time period. Overall, the RERs are lower than in the previous case, where a shorter assimilation period was used, and the RERs for the 6 km network are lower than those of the 20 km network in both cases throughout the forecast. The difference due to network density is generally larger when the forecasts are initialized at the earlier time of 60 min.

This suggests that when surface observations are assimilated within fewer cycles, there is more benefit from having a denser network. These conclusions are also supported by the low-level forecast fields at 150 min (Fig. 13). The forecast of Ra (Fig. 13b) is the poorest among the three and that of RaSfcS6 (Fig. 13d) is the best. The most noticeable differences are due to the left-mover near the northwestern corner of the plotting domain. For the main cell near the center, the difference is smaller, especially when compared with the previous case where the forecasts start from 60 min (c.f., Fig. 12).

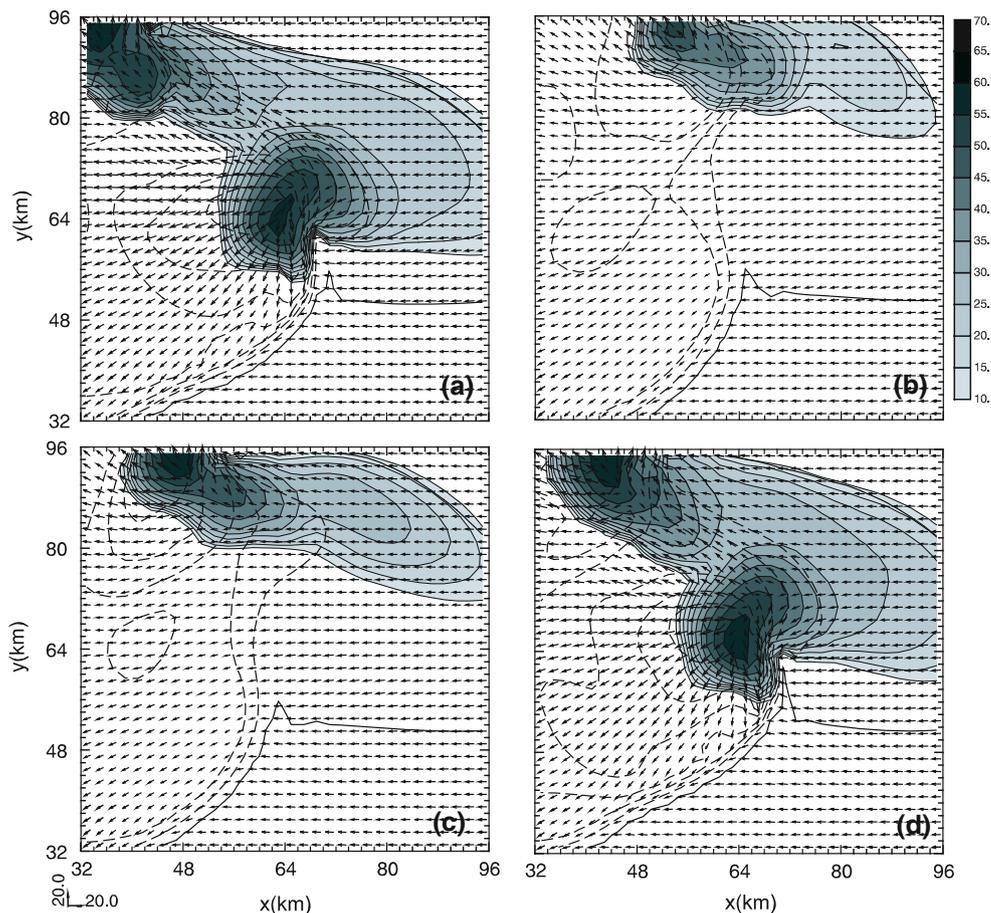
When the assimilation cycles continue until 120 min, the analysis RERs of the 6 km (RaSfcS6) and 20 km (RaSfc) cases (gray lines in Fig. 11) become closer. In fact, the analysis RERs for the wind components (Fig. 11a) are so close that after 10 min of forecast time the RER of the 20 km case (thick gray curve) becomes lower than that of the 6 km case (this behavior is related to nonlinear error growth). The actual analysis errors at this time are already very small, even for the case using a 20 km network spacing (c.f., Fig. 3); therefore, further improvement due to higher network density becomes negligible. For temperature and pressure, the benefit of the denser network lasts a little longer (Fig. 11b) while the benefit is not seen in terms of the microphysical variables throughout the hour-long forecast (Fig. 11c). This suggests that when radar data coverage on a storm is available for a sufficiently long time, a 20-km resolution surface network is enough.

6 Impact of surface data in the presence of model error

Model error can significantly impact the behavior of the EnKF analysis. For real-world problems, model error is inevitable. Model error can originate from physical parameterizations, model numerics and resolution and simplifying assumptions made within the dynamic equations. For short-range (\sim a few hours) simulation/prediction of convective storms, the microphysical parameterization (MP) is often the largest source of error or uncertainty (e.g., Gilmore et al. 2004b; Tong and Xue 2008b) and efforts to alleviate such uncertainty can be made through parameter estimation (Tong and Xue 2008a). In this section, we simulate the potential MP error by using different MP scheme(s) in the assimilating model than the one used in the truth simulation. Such an approach has been used with cumulus parameterization schemes in the OSSEs of Meng and Zhang (2007) for larger scale applications. Our main goal here is to assess the impact of surface data in the presence of a dominant model error.

We first assess the impact of model error in the radar-only case. Experiments RaLFO04, RaSchultz and RaW-SM6 are the same as Ra except for the use of the LFO04,

Fig. 12 Perturbation wind vectors, Z (dBZ, shaded) and θ' (K, contours) fields, valid at 120 min, from the truth **a**, and 1-h forecasts starting from 60-min ensemble-mean analyses of Ra **b**, RaSfc **c** and RaSfcS6 **d**



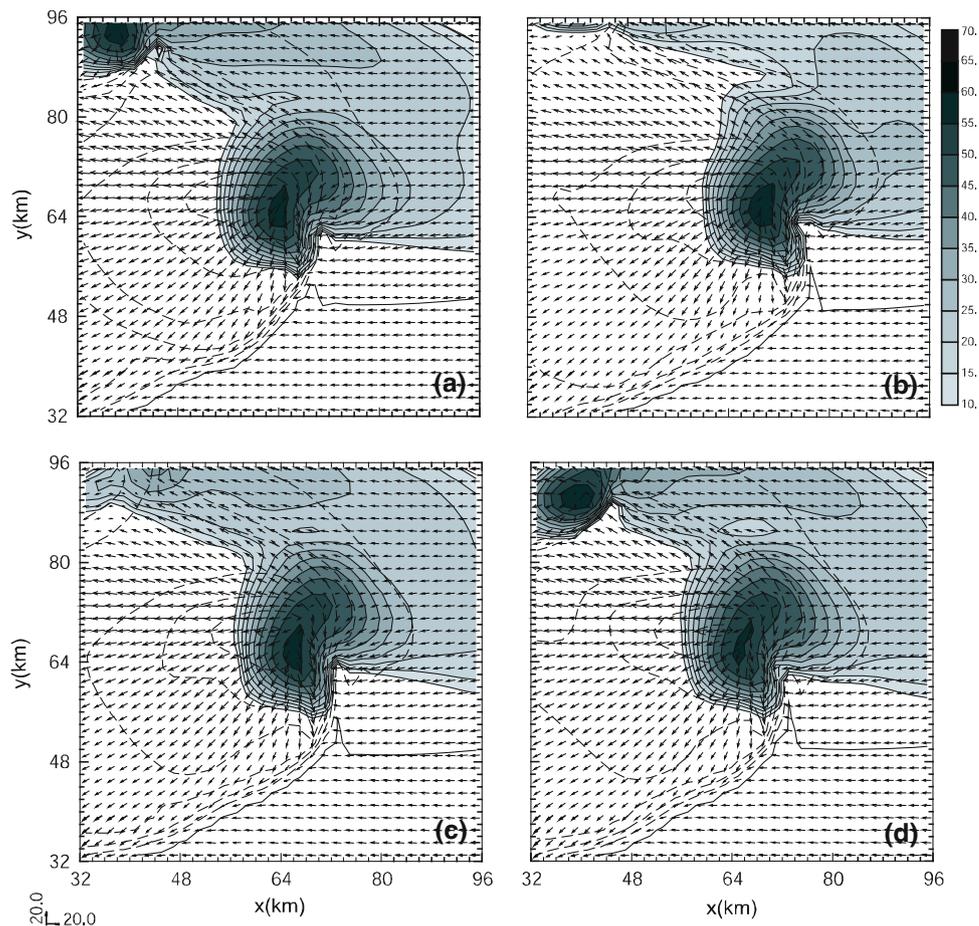
Schultz and WSM6 (c.f., Sect. 2a) MP schemes, respectively, instead of the truth MP scheme of LFO83. Experiment RaMulti uses LFO04, Schultz and WSM6 in 13, 13 and 14 of the 40 ensemble members and such a multi-physics approach has been found effective by Meng and Zhang (2007) and Fujita et al. (2007) in increasing the ensemble spread and reducing analysis error. Another experiment, RaNr0, uses LFO83 with the rain water intercept parameter, N_{r0} , increased by a factor of 10. This increases the number of rain drops and reduces their average sizes, leading to more rain evaporation and stronger cold pools. In the presence of model error, the covariance inflation coefficient is increased to 15%.

Figure 14a shows the TRERs of the radar-only experiments with MP error relative to the perfect-model experiment Ra. In all cases the TRER is larger than 1, indicating degradation of the analysis due to model error. Among them, RaWSM6 has the largest TRER and RaNr0 has the smallest TRER. For RaNr0 the error is increased by about 27%, while for RaWSM6 the error is larger than that of the perfect-model case by about a factor of 3. For real cases, we do not know which MP scheme is more accurate; one way to account for such uncertainty is to use multiple MP

schemes in the ensemble, helping to better sample the uncertainty. Figure 14a shows that the TRER of RaMulti is smaller than those of RaWSM6 and RaSchultz but slightly larger than that of RaLFO04. Considering that LFO04 is a ‘close relative’ of LFO83 that is used in the truth simulation, the fact that RaMulti performs close to RaLFO04 suggests that the use of multiple MP schemes is rather effective.

In the presence of MP error, it is shown that with a network of 20 km spacing surface observations still show positive impact on the storm analyses (Fig. 14b), but the magnitude of impact is reduced. The error reduction is the largest for RaSfcNr0; being about 20% with a TRER of about 80%. RaSfcNr0 has the smallest MP error since only the value of the rain intercept parameter is incorrect. The TRERs of all 20-km experiments using a single wrong microphysics scheme are close to 95%, indicating rather small positive impact from surface data. Although larger model errors (using a single wrong MP scheme) can create more room for state error correction, the model error can also lead to less accurate covariance estimates on which the data assimilation depends and hence to smaller impacts from the observations than RaSfcNr0. When multiple MP

Fig. 13 As Fig. 12, but for fields valid at 150 min, of truth **a** and 1-h forecasts starting from 90-min ensemble-mean analyses of Ra **b**, RaSfc **c** and RaSfcS6 **d**



schemes are used in RaSfcMulti, the TRER is about 88%, smaller than the cases using a single wrong MP scheme. We also notice that the microphysical state variables are improved least by the surface observations among these experiments (not shown). A denser surface network of 6 km spacing is shown to produce larger positive impacts (Fig. 14b), with the error reduction in the multi-scheme case being about 25%, while the error reduction in the case using the wrong rain intercept parameter value is about 40%. The relative impacts among different schemes are similar to those seen in the 20 km case.

The larger impact from using multiple MP schemes than using single “wrong” MP schemes is partly due to the increase of the spread (Fig. 15). Spread is averaged over the last ten assimilation cycles to provide the relative spread increase. The average relative spread increase from using multiple MP schemes is always positive for all of the state variables. Generally, there are larger relative spread increases in the microphysical variables than in others. The relative spread increase of q_s can be as large as 600%.

We note here that in this study, the environment of the convective storms is assumed perfect (apart from the noise introduced when initializing the initial forecast ensemble), being based on a perfect sounding. The surface data are used to help analyze the perturbations associated with the convective storms; near the surface the cold pool and associated features are profoundly influenced by the microphysics (Snook and Xue 2008); for this reason, microphysics error significantly reduces the ability of surface observations in improving state estimation.

7 Summary and conclusions

In this paper, an ensemble square root Kalman filter is used to assimilate simulated radar and surface network observations for a supercell storm, with the goal of examining the impact of additional surface observations on the storm analysis and forecast. Realistic low-level radar data coverage is simulated with a radar emulator using realistic beam pattern weighting.

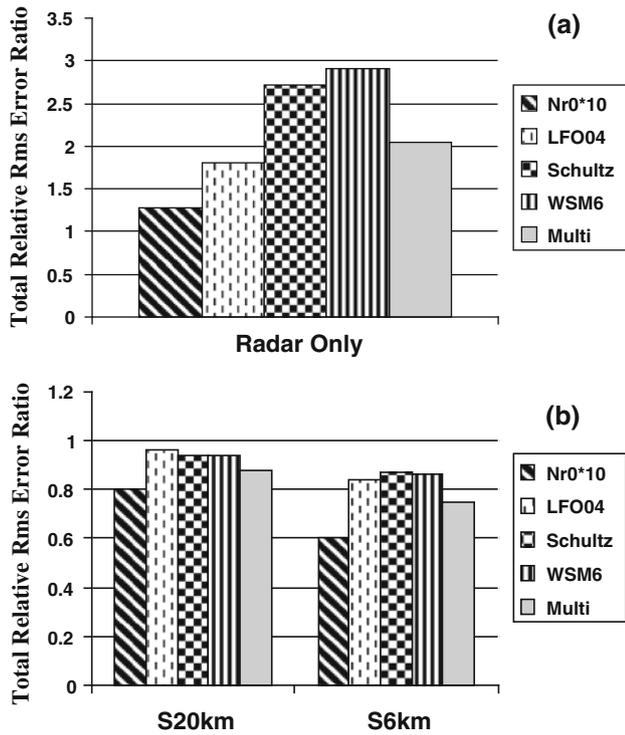


Fig. 14 **a** Total relative *rms* error ratios (TRERs) between imperfect and perfect model experiments for radar data only experiments RaNr0, RaLFO04, RaSchultz, and RaMulti, and **b** total relative *rms* error ratios (TRERs) for experiments with and without surface data when the model is imperfect

It is shown that the assimilation of mesonet-like surface observations can significantly improve convective storm analysis when compared with the cases using radar

observations only, particularly when the radar is at sufficient distance from the storm to have poor low-level coverage. Surface observations help fill the low-level data gap and improve the storm analysis even when the resolution of these observations is much coarser (e.g., at a 20 km spacing) than the radar data. The *rms* error reductions due to surface observations reach 30–50% in such a case when the assimilation model is perfect. Surface observations improve analyses of the cold pool, consistent with Zhang et al. (2004)’s study using simplified settings. The convergence and divergence associated with the cold pool are also retrieved better in terms of pattern, magnitude and position while the precipitation fields are improved less because hydrometeors at higher levels are only indirectly related to surface observations and are better captured by radar.

It is shown that surface observations not only help correct the near-surface errors, but also improve the analyses of state variables at the mid- and upper levels. The greatest improvement from surface observations usually occurs at levels where analysis errors obtained without using surface data are the largest. The flow-dependent background error covariance estimated from the ensemble and the dynamical interactions realized through the prediction model are believed to play an important role. The background error correlations estimated from the ensemble exhibit physically reasonable structures and confirm the ability of the surface observations to properly influence all state variables at other grid points.

In general, the better analysis obtained using both radar and surface observations, used as the initial condition for

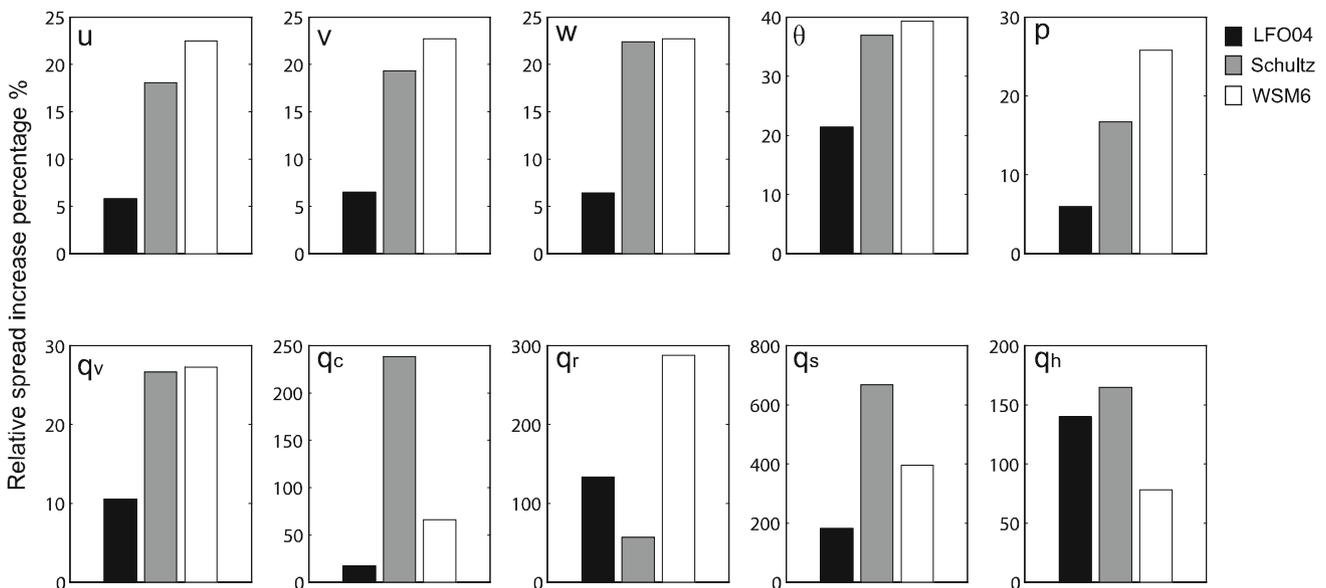


Fig. 15 Relative spread increase percentage of RaSfcS6Multi to RasfcS6LFO, RasfcS6Schultz and RasfcS6WSM6. Spreads are averaged over the last ten assimilation cycles

the following forecast, also improves the subsequent storm forecast. When the forecast starts at an earlier time, when fewer assimilation cycles have been taken and when the cold pool occupies a smaller area, a higher density surface network results in a much greater positive impact. When many analysis cycles are performed, the 20-km network is able to produce a similar level of positive impact as the 6-km network, and the impact lasts for at least 1 h in both cases during the forecast. In some cases, the improvement in the forecast even increases with time.

When assimilating the surface observation types individually, using typical observation errors, wind observations show the largest positive impact, reflecting the important role of low-level flows in the storm development, maintenance and propagation. Temperature observations are found to have the second largest impact and pressure observations are found to have the least impact. Part of the reason for the small impact of pressure observations may be due to the relatively large error specified for the pressure observations (1 hPa standard deviation) as compared with the error level of analyzed pressure achieved using radar data alone. The largest positive impact is achieved when all surface observation types are assimilated together.

Different covariance localization radii are used for radar and surface observations in the EnKF analysis. The optimal radii for surface observations, found through experimentation, generally decrease as network spacing decreases and are generally larger than those of radar data; such a ‘multi-scale’ analysis technique seems to be effective when dealing with data from networks with very different resolutions. An argument in support of such an approach is that observation networks are often designed to capture their respective target flow scales (e.g., a rawinsonde network is designed to primarily capture synoptic-scale structures). A practical consideration is the desire to produce relatively smooth analyses given the observation networks, and to avoid ‘bull’s eyes’ in the analyses. Similar multi-scale techniques have been used in the context of 3DVAR (e.g., Hu et al. 2006) and successive correction methods (e.g., Xue and Martin 2006).

The *rms* error relative to the radar-only case is found to decrease linearly as mean surface network spacing decreases until the spacing is close to the grid interval of truth simulation. When plotted against the total number of surface stations, this relative error is approximately proportional to the inverse of the square root of the number of stations when the network is relatively coarse, similar to the behavior found by Morss et al. (2001).

When the radar is located at a closer distance of 115 km, the radar-only analysis is significantly improved over that of 185 km case. In this case, the surface observations still show consistent positive impact and this impact increases with increasing network density. When the radar is located

only 45 km from the storm, a station spacing of 6 km is needed to achieve a noticeable impact because the radar-only analysis is already very good.

The impact of surface observations is also examined in the presence of microphysics-related model error. Such model error is found to reduce the relative impact of surface observations. When multiple microphysics schemes are used in the forecast ensemble, a surface network of 20 km (6 km) spacing produces error reduction of about 10% (25%).

The impact of surface observations does depend on the quality of analysis obtained using radar data only. It also depends on the general quality of state estimation, which affects the estimated error covariance between the surface observations and unobserved state variables. Attempts to analyze convective storms using surface data only were unsuccessful. The impact of surface observations may be larger when error also exists in the storm environment, which can affect the forecast even more; this is a topic for future study. Finally, we note that while this study may share some of the common limitations of OSSEs, the results obtained do give us valuable insights that would be difficult, if not impossible, to obtain with real data.

Acknowledgments We thank Dr. Mingjing Tong for her assistance and helpful discussions on the ARPS EnKF assimilation system, and Nathan Snook for proofreading the manuscript. This work was primarily supported by NSF grants AGS-0331594, AGS-0802888 and OCI-0905040. Ming Xue was also supported by NSF grants AGS-0750790, AGS-0608168 and EEC-0313747. Computations were performed at the Pittsburgh Supercomputing Center (PSC) and Oklahoma Supercomputing Center for Research and Education (OSCER).

References

- Anderson JL, Wyman B, Zhang S, Hoar T (2005) Assimilation of surface pressure observations using an ensemble filter in an idealized global atmospheric prediction system. *J Atmos Sci* 62:2925–2938
- Bratseth AM (1986) Statistical interpolation by means of successive corrections. *Tellus* 38A:439–447
- Brewster K (1996) Application of a Bratseth analysis scheme including Doppler radar data. In: Preprints, 15th Conf Wea Anal Forecasting, Norfolk, VA, Amer Meteor Soc, pp 92–95
- Compo GP, Whitaker JS, Sardeshmukh PD (2006) Feasibility of a 100 year reanalysis using only surface pressure data. *Bull Am Meteor Soc* 87:175–190
- Dawson DT II, Xue M (2006) Numerical forecasts of the 15–16 June 2002 Southern Plains severe MCS: impact of mesoscale data and cloud analysis. *Mon Wea Rev* 134:1607–1629
- Doviak R, Zrníc D (1993) Doppler radar and weather observations, 2nd edn. Academic Press, San Diego
- Dowell D, Zhang F, Wicker LJ, Snyder C, Crook NA (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 (2003) The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn* 53:343–367

- Fujita T, Stensrud DJ, Dowell DC (2007) Surface data assimilation using an ensemble Kalman filter approach with initial condition and model physics uncertainties. *Mon Wea Rev* 135:1846–1868
- Gaspari G, Cohn SE (1999) Construction of correlation functions in two and three dimensions. *Quart J Roy Meteor Soc* 125:723–757
- Gilmore MS, Straka JM, Rasmussen EN (2004a) Precipitation and evolution sensitivity in simulated deep convective storms: comparisons between liquid-only and simple ice and liquid phase microphysics. *Mon Wea Rev* 132:1897–1916
- Gilmore MS, Straka JM, Rasmussen EN (2004b) Precipitation uncertainty due to variations in precipitation particle parameters within a simple microphysics scheme. *Mon Wea Rev* 132:2610–2627
- Hacker JP, Rostkier-Edelstein D (2007) PBL state estimation with surface observations, a column model, and an ensemble filter. *Mon Wea Rev* 135:2958–2972
- Hacker JP, Snyder C (2005) Ensemble Kalman filter assimilation of fixed screen-height observations in a parameterized PBL. *Mon Wea Rev* 133:3260–3275
- Hong SY, Lim JOJ (2006) The WRF single-moment 6-class microphysics scheme (WSM6). *J Korean Meteor Soc* 42:129–151
- Houtekamer PL, Mitchell HL (2001) A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon Wea Rev* 129:123–137
- Hu M, Xue M, Gao J, Brewster K (2006) 3DVAR and cloud analysis with WSR-88D level-II data for the prediction of Fort Worth tornadic thunderstorms. Part II: Impact of radial velocity analysis via 3DVAR. *Mon Wea Rev* 134:699–721
- Lin YL, Farley RD, Orville HD (1983) Bulk parameterization of the snow field in a cloud model. *J Clim Appl Meteor* 22:1065–1092
- Liu H, Xue M (2008) Prediction of convective initiation and storm evolution on 12 June 2002 during IHOP. Part I: control simulation and sensitivity experiments. *Mon Wea Rev* 136:2261–2283
- Lord SJ, Kalnay E, Daley R, Emmitt GD, Atlas R (1997) Using OSSEs in the design of the future generation of integrated observing systems. In: 1st symposium on integrated observation systems, Preprint volume. Amer Meteor Soc Long Beach, pp 45–47
- Meng Z, Zhang F (2007) Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part II: Imperfect model experiments. *Mon Wea Rev* 135:1403–1423
- Meng Z, Zhang F (2008) Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part IV: comparison with 3DVAR in a month-long experiment. *Mon Wea Rev* 136:3671–3682
- Morss RE, Emanuel KA, Snyder C (2001) Idealized adaptive observation strategies for improving numerical weather prediction. *J Atmos Sci* 58:210–232
- Ray PS, Johnson B, Johnson KW, Bradberry JS, Stephens JJ, Wagner KK, Wilhelmson RB, Klemp JB (1981) The morphology of severe tornadic storms on 20 May 1977. *J Atmos Sci* 38:1643–1663
- Schultz P (1995) An explicit cloud physics parameterization for operational numerical weather prediction. *Mon Wea Rev* 123:3331–3343
- Snook N, Xue M (2008) Effects of microphysical drop size distribution on tornadogenesis in supercell thunderstorms. *Geophys Res Lett* 35:L24803. doi:[10.1029/2008GL035866](https://doi.org/10.1029/2008GL035866)
- Tong M (2006) Ensemble Kalman filter assimilation of Doppler radar data for the initialization and prediction of convective storms. Dissertation, University of Oklahoma
- Tong M, Xue M (2005) Ensemble Kalman filter assimilation of Doppler radar data with a compressible nonhydrostatic model: OSS experiments. *Mon Wea Rev* 133:1789–1807
- Tong M, Xue M (2008a) Simultaneous estimation of microphysical parameters and atmospheric state with radar data and ensemble square-root Kalman filter. Part II: parameter estimation experiments. *Mon Wea Rev* 136:1649–1668
- Tong M, Xue M (2008b) Simultaneous estimation of microphysical parameters and atmospheric state with radar data and ensemble square-root Kalman filter. Part I: sensitivity analysis and parameter identifiability. *Mon Wea Rev* 136:1630–1648
- Weckwerth TM, Parsons DB, Koch SE, Moore JA, LeMone MA, Demoz BB, Flamant C, Geerts B, Wang J, Feltz WF (2004) An overview of the International H2O Project (IHOP_2002) and some preliminary highlights. *Bull Am Meteor Soc* 85:253–277
- Whitaker JS, Hamill TM (2002) Ensemble data assimilation without perturbed observations. *Mon Wea Rev* 130:1913–1924
- Whitaker JS, Compo GP, Wei X, Hamill TM (2004) Reanalysis without Radiosondes using ensemble data assimilation. *Mon Wea Rev* 132:1190–1200
- Whitaker JS, Hamill TM, Wei X, Song Y, Toth Z (2007) Ensemble data assimilation with the NCEP global forecast system. *Mon Wea Rev* 136:463–482
- Xu Q, Lu H, Gao S, Xue M, Tong M (2008) Time-expanded sampling for ensemble Kalman filter: assimilation experiments with simulated radar observations. *Mon Wea Rev* 136:2651–2667
- Xue M, Martin WJ (2006) A high-resolution modeling study of the 24 May 2002 case during IHOP. Part I: numerical simulation and general evolution of the dryline and convection. *Mon Wea Rev* 134:149–171
- Xue M, Droegemeier KK, Wong V (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
- Xue M, Droegemeier KK, Wong V, Shapiro A, Brewster K, Carr F, Weber D, Liu Y, Wang DH (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
- Xue M, Wang DH, Gao JD, Brewster K, Droegemeier KK (2003) The Advanced Regional Prediction System (ARPS), storm-scale numerical weather prediction and data assimilation. *Meteor Atmos Phys* 82:139–170
- Xue M, Tong M, Droegemeier KK (2006) An OSSE framework based on the ensemble square-root Kalman filter for evaluating impact of data from radar networks on thunderstorm analysis and forecast. *J Atmos Ocean Tech* 23:46–66
- Xue M, Jung Y, Zhang G (2007) Error modeling of simulated reflectivity observations for ensemble Kalman filter data assimilation of convective storms. *Geophys Res Lett* 34:L10802. doi:[10.1029/2007GL029945](https://doi.org/10.1029/2007GL029945)
- Zhang F, Snyder C, Sun J (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 F, Meng Z, Aksoy A (2006) Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part I: perfect model experiments. *Mon Wea Rev* 134:722–736