Ensemble Kalman Filter Assimilation of Radar Observations of the 8 May 2003 Oklahoma City Supercell: Influences of Reflectivity Observations on Storm-Scale Analyses

DAVID C. DOWELL
Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, and NOAA/OAR National Severe Storms Laboratory, Norman, Oklahoma, and National Center for Atmospheric Research,* Boulder, Colorado

LOUIS J. WICKER
NOAA/OAR National Severe Storms Laboratory, Norman, Oklahoma

CHRIS SNYDER
National Center for Atmospheric Research,* Boulder, Colorado

(Manuscript received 31 March 2010, in final form 23 July 2010)

ABSTRACT

Ensemble Kalman filter (EnKF) techniques have been proposed for obtaining atmospheric state estimates on the scale of individual convective storms from radar and other observations, but tests of these methods with observations of real convective storms are still very limited. In the current study, radar observations of the 8 May 2003 Oklahoma City tornadic supercell thunderstorm were assimilated into the National Severe Storms Laboratory (NSSL) Collaborative Model for Multiscale Atmospheric Simulation (NCOMMAS) with an EnKF method. The cloud model employed 1-km horizontal grid spacing, a single-moment bulk precipitation-microphysics scheme, and a base state initialized with sounding data. A 50-member ensemble was produced by randomly perturbing base-state wind profiles and by regularly adding random local perturbations to the horizontal wind, temperature, and water vapor fields in and near observed precipitation.

In a reference experiment, only Doppler-velocity observations were assimilated into the NCOMMAS ensemble. Then, radar-reflectivity observations were assimilated together with Doppler-velocity observations in subsequent experiments. Influences that reflectivity observations have on storm-scale analyses were revealed through parameter-space experiments by varying observation availability, observation errors, ensemble spread, and choices for what model variables were updated when a reflectivity observation was assimilated. All experiments produced realistic storm-scale analyses that compared favorably with independent radar observations. Convective storms in the NCOMMAS ensemble developed more quickly when reflectivity observations and velocity observations were both assimilated rather than only velocity, presumably because the EnKF utilized covariances between reflectivity and unobserved model fields such as cloud water and vertical velocity in efficiently developing realistic storm features.

Recurring spatial patterns in the differences between predicted and observed reflectivity were noted particularly at low levels, downshear of the supercell’s updraft, in the anvil of moderate-to-light precipitation, where reflectivity in the model was typically lower than observed. Bias errors in the predicted rain mixing ratios and/or the size distributions that the bulk scheme associates with these mixing ratios are likely responsible for this reflectivity underprediction. When a reflectivity observation is assimilated, bias errors in the model fields associated with reflectivity (rain, snow, and hail–graupel) can be projected into other model variables through the ensemble covariances. In the current study, temperature analyses in the downshear anvil at low levels, where reflectivity was underpredicted, were very sensitive both to details of the assimilation algorithm and to ensemble spread in temperature. This strong sensitivity suggests low confidence in analyses of low-level cold pools obtained through reflectivity-data assimilation.

* The National Center for Atmospheric Research is sponsored by the National Science Foundation.
1. Introduction

The ensemble Kalman filter (EnKF) assimilation of radar observations into numerical cloud models has been investigated recently as a means for producing detailed analyses of convective storms and initializing weather prediction models (Snyder and Zhang 2003; Dowell et al. 2004b; Tong and Xue 2005; Caya et al. 2005; Gao and Xue 2008; Xu et al. 2008; Jung et al. 2008b; Dowell and Wicker 2009; Aksoy et al. 2009; Yussouf and Stensrud 2010; Lu and Xu 2009; Xue et al. 2009; Zhang et al. 2009). EnKF methods use the statistics of a forecast ensemble to estimate the background-error covariances needed for data assimilation (Evensen 1994; Houtekamer and Mitchell 1998). These methods are relatively easy to implement and parallelize (e.g., Anderson and Collins 2007), and they provide both an analysis (ensemble mean) and an uncertainty estimate (ensemble spread). Typical formulations of EnKF methods are considered optimal when the following assumptions are satisfied: 1) the observations are linearly related to the true model state, 2) observation and forecast errors have Gaussian distributions, 3) observation and forecast errors are unbiased, and 4) ensemble statistics represent true forecast errors (Evensen 1994; Houtekamer and Mitchell 1998; Anderson and Anderson 1999; Whitaker and Hamill 2002; Lorenc 2003; Snyder and Zhang 2003; Dowell et al. 2004b). Considering the localized, unbalanced, and rapidly evolving nature of convective storms, one might expect state estimation for these storms to be particularly challenging. Nevertheless, results from initial tests of EnKF methods for storm-scale radar-data assimilation have been encouraging, demonstrating that flow-dependent covariances estimated from ensembles are useful for analyzing both observed and unobserved fields in convective storms (Snyder and Zhang 2003).

Among many observation types provided by weather radars, Doppler velocity and reflectivity are the most commonly used observations for both operational and research applications. Doppler velocity is a power-weighted, volume-averaged measure of the component of scatterer motion away from or toward a radar (Doviak and Zrnic 1993). Typically, the scatterer motion is approximately equal to the air motion plus the scatterer fall velocity. Several studies have shown that using an EnKF method to assimilate Doppler-velocity observations into a numerical cloud model is a viable method for storm-scale state estimation. Snyder and Zhang (2003) and then Zhang et al. (2004) assimilated synthetic radial velocity observations into the Sun and Crock (1997) cloud model, with the intent of reproducing the supercell storm (Browning 1968) in a reference simulation from imperfect observations and a simplified initial state estimate. Accurate EnKF analyses were produced after roughly 6 volumes of synthetic radar data had been assimilated into the model over a 30-min period. As more radial velocity observations were assimilated, supercell storms in the EnKF analyses and reference simulation became more similar, and forecasts initialized from the ensemble-mean analyses were also improved.

For these first investigations of EnKF radar-data assimilation, Snyder and Zhang (2003) and Zhang et al. (2004) assumed that the forecast model was perfect and that the synthetic observations were unbiased and had known error characteristics. Following Snyder and Zhang, Dowell et al. (2004b) applied a similar EnKF method to a real supercell, for which model error is expected to be significant and observation-error characteristics are not known precisely. Despite these limitations, EnKF analyses obtained by assimilating single-Doppler-velocity observations into the Sun and Crock (1997) cloud model had realistic features in the wind field and compared favorably to independent observations, which included Doppler-velocity observations from a second radar, a dual-Doppler analysis, and in situ wind measurements from the surface to 400 m AGL from a tower.

Reflectivity is a measure of how effectively a volume of targets backscatters energy transmitted by a radar (Doviak and Zrnic 1993). Reflectivity depends on scatterer concentration, size, phase (liquid or ice), orientation, and other properties. Assimilating reflectivity observations of convective storms into cloud models would seem to be more problematic than assimilating Doppler-velocity observations for a number of reasons. First, model error is expected to be particularly large for reflectivity predicted by cloud models (Smith et al. 1975; Smith 1984; Gilmore et al. 2004b). Whereas simulated Doppler velocity is computed from the three wind components, which are controlled by grid-scale dynamics, simulated reflectivity is computed from hydrometeor fields, which are controlled by microphysical parameterizations and their inherent uncertainties. Second, even if the hydrometeor fields are predicted well, there is still uncertainty in how reflectivity is diagnosed from these fields. Third, reflectivity is nonlinearly related to the model state (Tong and Xue 2005), which means that the EnKF update is suboptimal (cf. conditions for optimality above). Fourth, unlike Doppler-velocity observation errors, reflectivity observation errors are affected significantly by attenuation and errors in radar calibration (e.g., Wilson and Brandes 1979; Ulbrich and Lee 1999).

---

1 Throughout the text, we commonly use the abbreviated term "reflectivity" to refer to the equivalent reflectivity factor (Doviak and Zrnic 1993) on the customary logarithmic scale (dBZ).
Our current expertise with EnKF assimilation of reflectivity observations into cloud models has been developed mainly from perfect-model experiments with synthetic radar observations. In their synthetic-data experiments for a simulated supercell, Caya et al. (2005) assimilated observations of rainwater mixing ratio and radial velocity together into the Sun and Crook (1997) cloud model. Since the model’s precipitation-microphysics scheme included only a single precipitation category (rain), assimilating observations of the rainwater mixing ratio was comparable to assimilating reflectivity observations, except that the observation operator is linear in the former case and nonlinear in the latter. After rainwater and velocity observations were assimilated for a few tens of minutes, the EnKF ensemble mean was an accurate representation of the model state in the reference simulation. EnKF performance in the Caya et al. (2005) experiments was comparable to that obtained by Snyder and Zhang (2003) experiments, in which only velocity observations were assimilated.

Unlike Caya et al. (2005), who used a model with a very simplified precipitation-microphysics parameterization, Tong and Xue (2005) used a cloud model [the Advanced Regional Prediction System (ARPS); Xue et al. (2000), (2001)] with a multiclass scheme that included ice. Reflectivity was diagnosed from the model’s rain, snow, and hail–graupel fields. Again in a perfect-model scenario, Tong and Xue (2005) demonstrated that wind, thermodynamic, and microphysical fields in a reference supercell simulation could all be reproduced accurately by assimilating synthetic Doppler-velocity and reflectivity observations for a few tens of minutes into the same cloud model that produced the reference simulation.

For reflectivity-data assimilation, it has proven helpful to distinguish between reflectivity observations of precipitation and reflectivity observations that have values so low that they indicate the absence of precipitation (no-precipitation reflectivity observations; Aksoy et al. 2009). Reflectivity observations in precipitation are potentially useful for analyzing inner storm details whereas the no-precipitation observations are helpful mainly for suppressing spurious convective storms that develop in ensemble members (Dowell et al. 2004a; Tong and Xue 2005; Aksoy et al. 2009). Tong and Xue (2005) and Aksoy et al. (2009) obtained the best analyses when both types of reflectivity observations (precipitation and no precipitation) were assimilated into the model, together with radial velocity observations.

Tong and Xue (2008a,b) demonstrated that reflectivity assimilation could be successful when the cloud model is initially slightly imperfect. Specifically, they considered uncertainty in the basic parameters in a single-moment precipitation-microphysics scheme. They demonstrated that reflectivity-data assimilation could be used to estimate simultaneously the atmospheric state and also a subset of the microphysics parameters (particle densities and/or intercept parameters), at least when the cloud model is otherwise perfect.

Experience with EnKF assimilation of reflectivity observations of real convective storms is quite limited. In a recent study, Aksoy et al. (2009) assimilated Doppler-velocity and reflectivity observations together into the Weather Research and Forecasting (WRF) model (Skamarock et al. 2005). This multi-case study included ordinary cells, convective lines, and supercells. Results were considered successful in that 1) assimilating observations consistently reduced prior rms fits to observations during the first 15–30 min of the assimilation period, 2) comparable results were obtained for the various convective storm types, and 3) model fields obtained by assimilating observations for 1 h appeared realistic. In another recent study, Lei et al. (2009) initialized ARPS forecasts of the supercell case that is featured in the current study by assimilating Doppler-velocity, reflectivity, and surface-mesonet observations into the model. Since reflectivity observations were assimilated together with other observations types in all experiments in the Aksoy et al. (2009) and Lei et al. (2009) studies, it is difficult to evaluate the influences that each observation type has individually on the analyses.

In our current study, we aim to provide more insight into reflectivity-data assimilation for real convective storms by describing our experiences with a single case in detail. The case we have selected is the 8 May 2003 Oklahoma City, Oklahoma, tornadic supercell (Fig. 1), which is being used as a test case in multifaceted studies (Burgess 2004; Dowell et al. 2004a; Hu and Xue 2007; Rasmine et al. 2008; Dowell and Wicker 2009; Lei et al. 2009). As described in section 2, Doppler-velocity and reflectivity observations of the Oklahoma City supercell are available from multiple radars, together with nearby surface observations and a sounding. To help us understand how reflectivity-data assimilation works for a more typical operational situation in which a storm is observed well by only one radar, we have chosen to assimilate observations from one radar into the cloud model and to use the other observations for verification. The method used to assimilate radar data into the National Severe Storms Laboratory (NSSL) Collaborative Model for Multiscale Atmospheric Simulation (NCOMMAS; Wicker and Skamarock 2002; Coniglio et al. 2006) is summarized in section 2 and the appendix. A reference storm-scale analysis is produced by assimilating only Doppler-velocity observations into the model, following the work of Snyder and Zhang (2003), Dowell et al.
In section 3, we describe the model’s ability to simulate reflectivity in the reference (Doppler velocity only) experiment and also when reflectivity observations are assimilated together with velocity observations. In section 4, we describe how reflectivity-data assimilation changes the storm-scale analyses overall, pointing out both advantages and disadvantages of assimilating such observations for this case. We close with a summary, speculation about error sources that affect reflectivity-data assimilation, and thoughts about future work (section 5). The current study is essentially a follow-up study to Tong and Xue (2005), except for real instead of synthetic observations, and a complementary study to Aksoy et al. (2009), which provides more detailed analysis of one case.

2. Description of the data assimilation experiments

a. Case overview

Since Romine et al. (2008) provide a detailed overview of the Oklahoma City supercell thunderstorm and its environment, we only provide a brief description here. The Oklahoma City storm was the southernmost of many tornadic supercells that formed over Oklahoma and Kansas on 8 May 2003. Initial radar echoes of what was to become the Oklahoma City supercell developed near a dryline in west-central Oklahoma at 2045 UTC. The Oklahoma City storm became more organized as it moved northeastward, developing a mesocyclone above 2 km AGL by 2120 UTC (Burgess 2004) and then a “hook echo” (Stout and Huff 1953; van Tassel 1955) in reflectivity near the surface at 2204 UTC (Fig. 1). Romine et al. (2008) suggest that a single tornado event occurred from 2206 to 2238 UTC, although others have classified this event as three tornadoes (Burgess 2004). Over its 27-km path, the tornado produced F2 damage (Fujita 1981) in Moore, Oklahoma, and F4 damage in Oklahoma City. Approximately 50 km south of the Oklahoma City storm’s path, other convective cells (southwest part of Fig. 1), which formed as early as 2015 UTC, failed to develop into severe thunderstorms.

A National Weather Service sounding from Norman, Oklahoma (from location “KOUN” in Fig. 1) provides an estimate of the environmental conditions of the Oklahoma City storm (Fig. 2, which illustrates the profiles after interpolation to the NCOMMAS grid). The rawinsonde was released roughly when the Oklahoma City tornado was dissipating 35 km northeast of the sounding site. Convective available potential energy (CAPE) and convective inhibition (CIN) for the raw sounding (Fig. 3 of Romine et al. 2008) are approximately 3800 and 220 J kg\(^{-1}\), respectively, for a parcel originating with the mean properties of the lowest 100 mb of the sounding. The bulk Richardson number (Weisman and Klemp 1984) is 18. Such environmental conditions support strong supercell storms in numerical cloud models (Weisman and Klemp 1984).

Observations from multiple radars document the life cycle of the Oklahoma City storm. The current study particularly utilizes data from the KOUN radar, a 10-cm radar in Norman that is essentially a Weather Surveillance Radar-88 Doppler (WSR-88D) upgraded to have polarimetric capability (Ryzhkov et al. 2005; Romine et al. 2008). The first echoes of the Oklahoma City storm appeared approximately 80 km west-southwest of the KOUN radar, and then the storm moved closer to and north of the radar (Fig. 1). The Moore–Oklahoma City tornado formed only 12 km north-northwest of KOUN. Since the KOUN radar has a 0.95° beamwidth, the beam was 200–1300 m wide at these ranges of 12–80 km. The azimuthal sampling interval was 1.0°, comparable to the beamwidth. Unfortunately, a power failure caused KOUN data loss between 2210 and 2225 UTC. The current study utilizes KOUN Doppler-velocity and reflectivity data before 2210 UTC, from storm formation until tornadogenesis. Polarimetric observations of the Oklahoma City storm are not used here but are described in detail by Romine et al. (2008).

The KOUN velocity and reflectivity data that were assimilated into NCOMMAS (section 2b) had been
edited manually, which involves removing ground clutter, range folding, and other spurious data and unfolding aliased Doppler velocities, and then gridded. Velocity and reflectivity data are available in precipitation and also in “clear air” within roughly 30 km of the radar (Fig. 1 of Dowell and Wicker 2009). The KOUN radar obtained complete volumes (14 different elevation angles) approximately every 6 min. Data from each elevation angle in each volume were objectively analyzed to grid points on the conical scan surfaces (Sun and Crook 2001; Dowell et al. 2004b; Dowell and Wicker 2009). The grid points are 2000 m apart in each horizontal direction, and the radius of influence for the Cressman (1959) objective analysis was 1000 m. The 2000-m grid spacing is roughly the same as the mean observation spacing for the more distant radar observations. We have confirmed for two of our data-assimilation experiments that using observations analyzed on a finer (1000 m) grid does not change qualitative inferences about storm-scale features and processes, so we use the coarser observations in order to reduce computational requirements. Certainly, other studies focusing on topics such as mesocyclogenesis and tornadogenesis could benefit from using observations at the original (higher) resolution.

KTLX WSR-88D (Fig. 1) observations of the Oklahoma City storm are also available, and we use these data for verification purposes. The ranges from KTLX to the storm features of interest are initially 110 km but decrease to 20 km during the period of study. Terminal Doppler weather radar observations, which provide higher-resolution information at low levels during the tornadic stage (Romine et al. 2008), are not used in the current study but could be used in future studies focusing on topics such as tornadogenesis. Additional data we use for verification are the surface observations from KOKC, an Automated Surface Observing System site (Fig. 1).

b. Data assimilation

The numerical model, ensemble design, and data-assimilation method are similar to those used by Dowell and Wicker (2009). Briefly, this study employs NCOMMAS (Wicker and Skamarock 2002; Coniglio et al. 2006), which is a nonhydrostatic, compressible model designed for simulating convective storms in an idealized framework, where “idealized” refers to a flat lower boundary, no surface fluxes, no radiative transfer, and a horizontally homogeneous base state. The model domain for these experiments is 100 km wide in both horizontal
directions, is 18 km tall, and follows the storm of interest by moving at $U = 14$ m s$^{-1}$ and $V = 8$ m s$^{-1}$, where $U$ and $V$ are the components of domain motion toward the east and north, respectively. The grid spacings are uniformly 1.0 km (0.5 km) in the horizontal (vertical). The grid is rather coarse for simulating convective storms (Bryan et al. 2003), but keeps the computation manageable for the numerous experiments that are required to demonstrate the sensitivities of the EnKF analyses to observation availability, ensemble design, and assimilation methodology. The model time step is 5 s. The wind, temperature, and water vapor profiles in the raw 0000 UTC 9 May 2003 Norman sounding (Fig. 3 of Romine et al. 2008) were interpolated to model grid levels$^2$ in order to initialize the base state in the model (Fig. 2). Evidence in other studies of this supercell case suggests that more realistic representation of the spatially- and temporally-varying mesoscale environment is important for obtaining realistic storm-scale forecasts for 1–2.5 h following the data-assimilation window (Hu and Xue 2007; Lei et al. 2009). However, since the focus of the current study is instead on identifying issues with reflectivity-data assimilation during the assimilation window rather than on multi-hour forecasts, we have opted for the simpler initialization of the model base state. Comparison with other available observations (not shown) suggests that the Norman sounding provides a reasonable approximation of mean environmental conditions.

We initialized a 50-member NCOMMAS ensemble at 2040 UTC and used additive noise (Dowell and Wicker 2009) to introduce variability among the members. This additive noise consisted of random, smooth, local perturbations (e.g., Fig. 3 of Dowell and Wicker 2009) added every 5 min to each ensemble member’s $u$ (westerly wind component), $v$ (southerly wind component), $\theta$ (potential temperature), and $q_v$ (water vapor) fields, in and near where reflectivity observations indicated precipitation. The horizontal and vertical length scales in the smoothing function in the additive-noise algorithm were 4 and 2 km, respectively (Caya et al. 2005). Dowell et al. (2004b), Caya et al. (2005), and Tong and Xue (2008b) have suggested that using smooth perturbations such as these rather than initializing the ensemble with gridpoint noise results in a more persistent ensemble spread. Also, potential problems with reflectivity-data assimilation early in the assimilation window (Caya et al. 2005; Tong and Xue 2005) associated with noisy ensemble covariances are mitigated (Tong and Xue 2008b).

Incipient convective storms developed in the ensemble members through the additive noise, model advance, and data assimilation, and then the additive noise helped maintain ensemble spread as the different realizations of storms in the ensemble matured. The additive-noise parameters for most experiments in our study (including “$V_r$ and $Z_{dB}$” and $V_r$-only in Table 1) are the same as those in the “0.25” experiment described by Dowell and Wicker (2009). The standard deviations, before smoothing, are 1.0 (0.25) m s$^{-1}$, 1.0 (0.25) m s$^{-1}$, 1.0 (0.25) K, and 1.0 (0.25) K before (after) 2100 UTC for $u$, $v$, $\theta$, and $T_d$ (dewpoint), respectively. Two sensitivity experiments (“less additive noise” and “more additive noise” in Table 1) used different magnitudes of additive noise, as described later in section 4.

The ensemble in our current study also includes perturbed base-state wind profiles for each member. The motivations for these wind-profile perturbations are the uncertainty in how well the base-state sounding represents the actual storm environment and the spatial and temporal variability expected in this environment. Aksoy et al. (2009) have shown that for storm-scale radar-data assimilation in cloud models with simplified storm environments, the increased ensemble spread resulting from the wind-profile perturbations can help improve prior and posterior fits of the ensemble mean to the Doppler-velocity observations. Here, a realization of random, Gaussian noise with mean 0 and standard deviation 2.0 m s$^{-1}$ (magnitude chosen through experimentation) was added to the base-state $u$ and $v$ wind components at each grid level in each ensemble member.

The EnKF method we used to assimilate observations into the NCOMMAS ensemble is commonly called an ensemble square root filter (EnSRF; Whitaker and Hamill 2002; cf. Dowell and Wicker 2009 and Dowell et al. 2004b). A primary focus of this study is to compare an experiment with Doppler-velocity and reflectivity observations assimilated together ($V_r$ and $Z_{dB}$ in Table 1) to a reference experiment with only Doppler velocity assimilated ($V_r$ only). Every 60 s from 2046 to 2210 UTC, Doppler-velocity observations from scans (elevation angles) close to the current model time were assimilated into the NCOMMAS ensemble. Since KOUN obtained complete radar volumes every 6 min, approximately
one-sixth of a data volume was assimilated at a time. In most experiments, reflectivity observations were also assimilated into the model. Observation-error standard deviations were assumed to be 2 m s$^{-1}$ and 2 dBZ for Doppler velocity and reflectivity, respectively (Dowell et al. 2004b; Dowell and Wicker 2009; Aksoy et al. 2009; Yussouf and Stensrud 2010). Unless otherwise noted, all model variables except the Exner function and the mixing coefficient (Dowell and Wicker 2009) were updated when each observation was assimilated. We used trilinear interpolation to compute point estimates of model fields at observation locations. The transformation from model fields to reflectivity is described later.

The EnKF employs covariance localization following Houtekamer and Mitchell (2001) and Hamill et al. (2001). The influence of each observation on the model state was restricted to a sphere with a radius of 6 km around the observation. The functional form for this localization was taken from the compactly supported fifth-order correlation function in Eq. (4.10) of Gaspari and Cohn (1999). The localization radius used here is in the range used in previous studies for ensembles of 50–100 members (Dowell et al. 2004b; Caya et al. 2005; Tong and Xue 2005; Aksoy et al. 2009; Dowell and Wicker 2009). Although storm-scale analysis sensitivity to localization radius for radar observations has been investigated in previous studies (e.g., Dowell et al. 2004b; Caya et al. 2005), we suggest that there is still room for improving the data-assimilation algorithm through more rigorous and/or adaptive methods for determining localization radii for Doppler-velocity observations, reflectivity observations in precipitation, and no-precipitation observations.

For the representation of cloud processes in NCOMMAS, we selected the Gilmore et al. (2004a) version of the Lin et al. (1983) precipitation-microphysics scheme, which is a single-moment bulk scheme for cloud modeling. This scheme includes two liquid and three ice categories: $q_r$ (cloud droplets), $q_c$ (rain), $q_h$ (cloud ice crystals), $q_s$ (snow), and $q_a$ (hail–graupel). The cloud-droplet and ice-crystal distributions in the model are monodispersed, whereas the rain, snow, and hail–graupel categories have inverse exponential distributions (Marshall and Palmer 1948). Single-moment schemes like the one used here are employed in the current generation of high-resolution, real-time models that produce convective storms explicitly (e.g., Kain et al. 2008).

The appendix describes how reflectivity corresponding to model fields, on the customary logarithmic scale (dBZ), was computed. A minimum threshold of 0 dBZ was used when mapping model fields to reflectivity and also applied to the reflectivity observations themselves. Reflectivity observations both within and outside precipitation regions were assimilated in the same manner, as in Tong and Xue (2005). (In contrast, Aksoy et al. 2009 used different localization radii for reflectivity observations in precipitation and no-precipitation observations.) Most of the no-precipitation KOUN reflectivity observations in our case are in the lowest 1–2 km AGL within 30 km of the radar (Fig. 1 of Dowell and Wicker 2009). Reflectivity observations below the radar’s noise threshold are not used in this study, even though they proved to be useful in the study by Aksoy et al. 2009.

## 3. Differences between observed and simulated reflectivity

Our first objective in presenting the results of radar-data assimilation experiments for the 8 May 2003 Oklahoma City supercell is to illustrate differences between the simulated and observed reflectivity fields, as a function of location and time. The reference experiment is the $V_r$-only experiment (Table 1; Figs. 3a,c,e,g; Fig. 4a; Figs. 5a–d), in which Doppler-velocity observations are assimilated and reflectivity observations are used only to select regions for additive noise. The Doppler-velocity observations strongly constrain the kinematic fields in the storm, whereas these observations weakly constrain the hydrometeor fields (and thus reflectivity) through covariances that develop

### Table 1. Summary of radar-data-assimilation experiments. Here $V_r$ and $Z_{dB}$ refer to Doppler velocity and reflectivity, respectively.

<table>
<thead>
<tr>
<th>Expt name</th>
<th>Assimilated obs</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_r$ and $Z_{dB}$</td>
<td>$V_r$ and $Z_{dB}$</td>
<td>As in $V_r$ and $Z_{dB}$, but the standard deviation of random noise, before smoothing, is 0.1 m s$^{-1}$, 0.1 m s$^{-1}$, 0.1 K, and 0.1 K after 2100 UTC for $u$, $v$, $\theta$, and $T_d$, respectively.</td>
</tr>
<tr>
<td>$V_r$ only</td>
<td>$V_r$</td>
<td>As in $V_r$, but the standard deviation of random noise, before smoothing, is 0.5 m s$^{-1}$, 0.5 m s$^{-1}$, 0.5 K, and 0.5 K after 2100 UTC for $u$, $v$, $\theta$, and $T_d$, respectively.</td>
</tr>
<tr>
<td>Less additive noise</td>
<td>$V_r$ and $Z_{dB}$</td>
<td>As in $V_r$ and $Z_{dB}$, but the standard deviation of random noise, before smoothing, is 0.5 m s$^{-1}$, 0.5 m s$^{-1}$, 0.1 K, and 0.1 K after 2100 UTC for $u$, $v$, $\theta$, and $T_d$, respectively.</td>
</tr>
<tr>
<td>More additive noise</td>
<td>$V_r$ and $Z_{dB}$</td>
<td>As in $V_r$, but the reflectivity observation error is assumed to be 5 instead of 2 dBZ.</td>
</tr>
<tr>
<td>5-dB obs error</td>
<td>$V_r$ and $Z_{dB}$</td>
<td>As in $V_r$, but $\theta$ is not updated when a reflectivity observation is assimilated.</td>
</tr>
<tr>
<td>No $\theta$ update</td>
<td>$V_r$ and $Z_{dB}$</td>
<td>As in $V_r$ and $Z_{dB}$, but only $q_r$, $q_c$, and $q_h$ are updated when a reflectivity observation is assimilated.</td>
</tr>
</tbody>
</table>
FIG. 3. Ensemble mean reflectivity (20- and 40-dB Z contours) at 4.25 km AGL in the (left) $V_r$-only experiment and (right) $V_r$ and $Z_{dB}$ experiment at two times: (a)–(d) 2118 and (e)–(h) 2202 UTC. In (a), (b), (e), and (f), observed (KOUN) reflectivity is shown (20- and 40-dBZ shading). In (c), (d), (g), and (h), ensemble-mean vertical velocity (shading at intervals of 4 m s$^{-1}$) is shown. Horizontal coordinates (km) are relative to KOUN.
with Doppler velocity (cf. Snyder and Zhang 2003; Dowell et al. 2004b; Tong and Xue 2005; W. Deierling and D. Dowell 2010, unpublished manuscript). By illustrating analyses at midlevels early (38 min after the model initialization) and late (82 min after the model initialization) in the assimilation window, Fig. 3 provides an overall sense of how the storms were simulated. At 2118 UTC, the $V_r$-only experiment had two large convective storms (Fig. 3a). Whereas the precipitation region in the southern storm actually had shrunk horizontally (Figs. 1 and 3a), the analysis shows a large precipitation region. Farther north at 2118 UTC, the $V_r$-only experiment had storms in the vicinity of the developing Oklahoma City storm cluster, but the largest precipitation echo (near $X = -35, Y = 10$ in Fig. 3a) was too far northeast, and only weak updrafts had developed within the southwestern part of the storm cluster (from $X = -50, Y = -10$ to $X = -40, Y = 0$ in Fig. 3c). Later, at 2202 UTC, there was better agreement between the observations and the $V_r$-only analysis. On the broad scale, there was good correspondence between the observed and simulated locations of the main 40-dBZ core (Fig. 3e), and a strong updraft was present in the analysis (near $X = -12.5, Y = 10$ in Fig. 3g) in the tornadic region of the Oklahoma City storm. However, there was some noticeable discrepancy between the observed and simulated reflectivity in the downshear region (northern part of Fig. 3c).

Many of the plots discussed in this section are of the innovation $d$:

$$d = y_o - H(x_f) \quad \text{or} \quad y_o - H(x_a),$$

where $y_o$ is the observation; $H$ is the observation operator, which maps the model state to the observation location and type (cf. the appendix); $x$ represents the entire model state; superscript $f$ indicates a forecast (prior) state; $a$ indicates an analysis (posterior) state; and an overbar indicates an ensemble mean. In this section, $H$ is the reflectivity observation operator; in section 4, $H$ is either the reflectivity or the Doppler-velocity observation operator. The opposite of $d$ (i.e., forecast minus observation instead of observation minus forecast) is plotted as a function of time and height (Fig. 4) and horizontal location for a particular time and radar elevation angle (Fig. 5). In Fig. 6, innovation statistics are plotted as a function of time (radar volume).

Figure 4 focuses on the maturing and tornadogenesis periods of storm evolution (Romine et al. 2008), starting 46 min after the NCOMMAS ensembles had been initialized, when the storm had a well-developed precipitation core both in the model and observations. Throughout the analysis period in the $V_r$-only experiment (Fig. 4a), the ensemble mean on average underpredicted reflectivity below 3 km AGL by as much as 8 to 14 dBZ.

To further illustrate spatial characteristics of how the predicted and observed reflectivity agree and disagree, we show results in observation space, at a low elevation angle ($1.5^\circ$), and at a time representative of the mature
FIG. 5. Reflectivity (dBZ) at 1.5° elevation angle at 2202 UTC 8 May 2003: (a) observed by KOUN; (b) prior ensemble mean in the $V_r$-only experiment; (c) prior ensemble mean minus observation, with observed 20- and 40-dBZ contours also shown; (d) posterior ensemble mean minus prior ensemble mean; and (e)–(h) as in (a)–(d), but for the $V_r$ and $Z_{dBH}$ experiment. The KOUN radar location is indicated by a black square.
storm phase (Figs. 5a,b). Some aspects of the observed reflectivity structure are predicted well, including the curved shape of the main (southernmost) high-reflectivity cell (from roughly \(X = -20, Y = 10\) to \(X = 10, Y = 25\)) and the locations of two left-flank cells (from \(X = -20, Y = 30\) to \(X = 5, Y = 35\) and from \(X = -10, Y = 50\) to \(X = 10, Y = 55\)). The underprediction that dominates the statistics in Fig. 4a occurs in the downshear anvil of the Oklahoma City storm (approximately \(0 < X < 30\) and \(15 < Y < 40\) in Fig. 5c). For the following reasons, we believe this underprediction is a symptom of model error more than other error sources. First, other studies suggest that single-moment bulk microphysics schemes such as the one used in this study poorly represent the areal extent and reflectivity distribution within downshear anvils at low levels in supercells (Dawson et al. 2010). Second, since the patterns in \(-d\) vary spatially and have large magnitudes that are both positive and negative, we cannot explain the patterns simply through a uniform observation bias (e.g., radar miscalibration). Third, since the precipitation in the downshear anvil at low levels is rain, for which there is less uncertainty in the reflectivity calculation than for ice–mixed-phase precipitation (cf. the appendix), we suggest that other error sources—specifically, errors in the predictions of hydrometeor fields—are more significant.

Figure 5c indicates overprediction of reflectivity (red colors) both on the southwest side of the Oklahoma City storm (along roughly \(X = -15\)) and in the northernmost cell on the left flank (along roughly \(Y = 50\)). This overprediction can be explained in at least three ways. First, it appears the rainfall is concentrated too close to the updraft; the “blob” of high reflectivity near \(X = -2, Y = 10\) in Fig. 5b is associated with \(q_r\) of nearly \(15 \text{ g kg}^{-1}\) beneath a midlevel core of heavy rain and hail–graupel (not shown). Other studies have noted that single-moment bulk microphysics schemes often produce precipitation that is too heavy close to the updraft (Gilmore et al. 2004b). Second, the simulated storm could be moving too slowly toward the northeast; red colors on the southwest side of the precipitation core and blue colors on the northeast side in Fig. 5c would be consistent with such a phase difference between simulated and observed storm locations. Higher up, the differences in the simulated and observed reflectivity patterns on the southwest side of the core (near \(X = -15, Y = 10\) in Fig. 3e) would also be consistent with this slower movement. Third, in the simulations of this case, we have noticed a tendency for the model to be too slow to weaken cells on the left flank.

At midlevels—roughly between 3 and 8 km AGL—there is better agreement between observed and predicted reflectivity (Fig. 4a). At upper levels—roughly between 8 and 12 km AGL—the interpretation is complicated because both model and/or observation-operator errors and changes in radar-data coverage seem to be important factors. After 2140 UTC, the high-reflectivity top of the Oklahoma City storm moves into the KOUN radar’s cone of silence, while the downshear anvil moves out of the cone of silence (not shown). The following differences between simulated and observed reflectivity at upper levels contribute to the patterns seen in Fig. 4: the maximum reflectivity values in storm tops, both in the main cell and in the left-flank cell, are generally overpredicted; the model is too slow to decrease the core
reflectivity at the top of the left-flank cell; the areal coverage of the upshear anvil of the main cell is underpredicted, and the magnitude of reflectivity in the downshear anvil is generally overpredicted (not shown).

Before examining the influences that reflectivity-data assimilation has on the storm-scale analyses in more detail in the following section, we briefly comment here on how the reflectivity innovations change when reflectivity observations are assimilated together with Doppler-velocity observations (i.e., $V_r$ and $Z_{dB}$ experiment; cf. Table 1). Relative to the $V_r$-only experiment, the $V_r$ and $Z_{dB}$ experiment (Figs. 4b and 6a) has generally closer prior and posterior fits of the ensemble mean to the reflectivity observations, as expected. For example, the shape of the high-reflectivity core of the main cell and the extent of its downshear anvil is better represented in the $V_r$, and $Z_{dB}$ experiment (cf. Figs. 3f and 3e; cf. Figs. 5f and 5e,b). Earlier in the assimilation window, the $V_r$ and $Z_{dB}$ experiment (Fig. 3b) better represented the shape of the main northern storm cluster and the smaller extent of the southern storm than did the $V_r$-only experiment (Fig. 3a). On average, each assimilation cycle produced much larger changes to the reflectivity field when both reflectivity and Doppler-velocity observations were assimilated (Figs. 5h and 6a) than when just Doppler-velocity observations were assimilated (Figs. 5d and 6a). This result supports the notion that hydrometeor fields are weakly constrained by Doppler-velocity observations and strongly constrained by reflectivity observations.

Some of the patterns seen before in the reflectivity innovations for the $V_r$-only experiment (Fig. 4a) remain in the $V_r$, and $Z_{dB}$ experiment (Fig. 4b), but with reduced magnitudes. The ensemble mean underpredicts reflectivity at low levels throughout experiment. Although there is significant spatial variability, the reflectivity-data assimilation increases the ensemble-mean reflectivity at low levels at many more locations than it decreases the reflectivity (e.g., Fig. 5h). As indicated in plots at the same elevation angle at other times (not shown), and as would be inferred from Figs. 4a,b, the reflectivity-data assimilation is repeatedly increasing the reflectivity at low levels during each cycle. The need to do so at low levels is perhaps increased by the reflectivity-data assimilation itself at midelevations. Between roughly 3 and 8 km AGL, increasing the fit to the reflectivity observations in the $V_r$ and $Z_{dB}$ experiment, relative to the $V_r$-only experiment, results in a net loss of hydrometeors and thus reflectivity over the layer (i.e., decrease in coverage of red colors and increase in coverage of blue colors in Fig. 4b relative to Fig. 4a). Thus, the amount of precipitation falling into the lowest levels is less in the $V_r$ and $Z_{dB}$ experiment than in the $V_r$-only experiment, and the reflectivity-data assimilation at low levels attempts to compensate. As discussed in the following section, recurring spatial patterns in analysis increments such as those identified here can be indicators of model or observation biases (Dell 2005).

4. Influences of reflectivity-data assimilation on storm-scale analyses

a. Observation-space diagnostics for the $V_r$-only and $V_r$, and $Z_{dB}$ experiments

We reveal the influences that reflectivity-data assimilation is having on the storm-scale analyses by examining observation- and model-space fields and by comparing multiple experiments. First, we use observation-space diagnostics (Dowell et al. 2004b; Dowell and Wicker 2009) to compare the $V_r$-only and $V_r$, and $Z_{dB}$ experiments. One diagnostic is the root-mean-square of the innovations:

$$\text{RMSI} = \sqrt{\langle d^2 \rangle},$$

(4.1)

where $\langle \cdot \rangle$ indicates an average over all observations in a radar volume (a period of roughly 6 min). Another diagnostic is the ensemble spread (standard deviation):

$$\text{spread} = \sqrt{\frac{1}{N-1} \sum_{n=1}^{N} [H(x_n) - \overline{H(x)}]^2},$$

(4.2)

where $N$ is the number of ensemble members (50 in these experiments), $n$ is an index that identifies a particular ensemble member, and $H$ is either the reflectivity or the Doppler-velocity observation operator. RMSI and spread statistics are shown for prior and posterior ensembles in Figs. 6 and 7.

Not surprisingly, assimilating reflectivity and Doppler-velocity observations together reduces both the prior and posterior RMSI for reflectivity, typically by 5–10 dBZ, relative to assimilating only velocity observations (Fig. 6a). At the same time, assimilating both observation types increases the prior and posterior RMSI (i.e., results in worse prior and posterior fits) for both the assimilated KOUN and not-assimilated (verification) KTLX Doppler-velocity observations (Figs. 6b,c). This result is somewhat surprising because the ensemble is being provided more information in the $V_r$ and $Z_{dB}$ than in the $V_r$-only experiment.

We suggest four possible reasons for worse fits to the velocity observations. One possibility is that assimilating both reflectivity and velocity observations decreases the ensemble spread too much, simply because
more (and perhaps also because different) observations are being assimilated than in the \( V_r \)-only experiment. Reduced prior ensemble spread during each assimilation cycle implies that observations have less influence on the posterior ensemble mean, all else being equal. However, results from another experiment (“more additive noise”; Table 1), described in more detail later, suggest that the reduced ensemble spread is not the most significant reason for worse fits to the velocity observations. In this more additive noise experiment, in which both velocity and reflectivity observations are assimilated, the ensemble spread is greater than in the \( V_r \) and \( Z_{dB} \) experiment and comparable to that in the \( V_r \)-only experiment, but the prior fits to the velocity observations are even worse than in the \( V_r \) and \( Z_{dB} \) and \( V_r \)-only experiments (not shown).

A second reason why the reflectivity-data assimilation might make the velocity analysis worse is model bias errors (Dee 2005). A third possibility is a biased reflectivity observation operator, particularly for ice and mixed-phase precipitation (cf. the appendix), and a fourth possibility is biased reflectivity observations (e.g., Ulbrich and Lee 1999). Figure 4 indicates that the mean innovation for reflectivity is nonzero (i.e., blue colors dominate over red colors), and diagnostics averaged over entire radar volumes confirm that \( \langle d \rangle \) is 2–8 dBZ in magnitude during the mature storm phase (not shown), suggesting bias in the model, observation operator, and/or observations. Bias in the reflectivity observations, particularly bias associated with radar-calibration errors, could possibly be significant (Sun and Crook 1997, 1998) but will not be addressed further here, since there seems to be insufficient information available to quantify a bias. Thus, the remaining discussion on errors in the assimilation will focus on evidence of bias in the model and/or reflectivity observation operator. In the analysis of differences between the predicted and observed reflectivity at low levels (section 3), we have already shown evidence of model errors that have regular patterns in a storm-relative sense. Such model-bias errors would lead to inconsistency between ensemble statistics and forecast errors, and thus the data assimilation would be expected to be suboptimal (Dee 2005). It is unclear exactly how reflectivity-data assimilation in the presence of model- and/or observation-bias errors leads to worse velocity analyses in our experiments, but the explanations could involve both how the velocity fields are affected directly by the reflectivity-data assimilation through the ensemble covariances and how errors in thermodynamic analyses resulting from the reflectivity-data assimilation project into the dynamics during the model integration.

An encouraging sign in Fig. 6c is that assimilating KOUN observations—whether just Doppler velocity or both velocity and reflectivity—results in a better posterior fit to the independent KTLX velocity observations; that is, analysis RMSI is less than forecast RMSI. RMSI changes between the forecasts and analyses are slight at some times, significant at other times. [Note: Storm evolution, changing radar viewing angles, and inconsistent overlap of data volumes from each radar (collected at different sampling frequencies) probably all contribute to the erratic nature of the RMSI curves in Fig. 6c.] The KOUN and KTLX radars viewed the updraft and mesocyclone region with between-beam angles increasing toward 90° as the storm approached the radars (Fig. 1), meaning that the KTLX radar was increasingly observing a component of the wind field that was unobserved by the KOUN radar.

**b. Analysis sensitivities to observation errors and limited updates of model fields**

To further reveal the influences that reflectivity-data assimilation has on the storm-scale analyses, we compare...
the results of assimilation-sensitivity experiments. The first sensitivity experiment ("5-dBZ obs error"; Table 1) is the same as \( V_r \) and \( Z_{\text{dB}} \) experiment, but that the assumed standard deviation of reflectivity-observation errors is 5 dBZ instead of 2 dBZ. That is, fit to the reflectivity observations is weaker. One motivation for this sensitivity experiment is to increase ensemble spread, thus allowing the velocity observations to have more influence on the analyses than in the \( V_r \) and \( Z_{\text{dB}} \) experiment. Another motivation is to reduce the influence of the possible error sources related to reflectivity that were mentioned previously: model error, biased observations, and/or a biased observation operator. Dowell et al. (2004a) assumed reflectivity-observation errors were 5 dBZ for their earlier experiments with the 8 May 2003 supercell, for the reasons just provided. Tong and Xue (2005) also assumed 5-dBZ observation errors for their perfect-model experiments with a simulated supercell, although no explanation was provided for using such a large error magnitude.

The observation-space statistics of the 5-dBZ observation error experiment (Fig. 7) differ from those of the \( V_r \) and \( Z_{\text{dB}} \) experiment (Fig. 6) as follows: the prior RMSI (spread) in KOUN reflectivity is increased by 8% (increased by 17%), the prior RMSI (spread) in KOUN velocity is decreased by 9% (increased by 6%), and the prior RMSI (spread) in KTLX velocity is decreased by 6% (increased by 4%). As expected, increasing the KOUN reflectivity observation error from 2 to 5 dBZ increases the observation-space ensemble spread. However, given the opposite trends of roughly the same percentage for RMSI in reflectivity and velocity, it is unclear whether the 5-dBZ observation error experiment would be considered more successful than the \( V_r \) and \( Z_{\text{dB}} \) experiment. Without further experimentation, it is difficult to determine if the increased fit to the velocity observations in the 5-dBZ observation error experiment is mainly a result of more ensemble spread, or if reduced influences of errors in the model, observations, and/or observation operator are also significant.

A second sensitivity experiment (update \( q_r \), \( q_s \), and \( q_h \); Table 1) is the same as the \( V_r \) and \( Z_{\text{dB}} \) experiment, but that when a reflectivity observation is assimilated, only the \( q_r \), \( q_s \), and \( q_h \) fields are updated; \( q_r \), \( q_s \), and \( q_h \) are the three prognostic variables in the reflectivity observation operator [cf. (A.4)–(A.7)]. One possible benefit of this approach is reducing the influence of model error; disregarding covariances between reflectivity observations and model fields not directly related to reflectivity would prevent model biases in \( q_r \), \( q_s \), and \( q_h \) from projecting into these other fields during the data assimilation. Another possible benefit is increased ensemble spread.

The statistics of the update \( q_r \), \( q_s \), \( q_h \) experiment (Fig. 7) differ from those of the \( V_r \) and \( Z_{\text{dB}} \) experiment (Fig. 6) as follows: the prior RMSI (spread) in KOUN reflectivity is increased by 4% (increased by 17%), the prior RMSI (spread) in KOUN velocity is decreased by 12% (increased by 16%), and the prior RMSI (spread) in KTLX velocity is decreased by 11% (increased by 20%). The update \( q_r \), \( q_s \), \( q_h \) experiment also has slightly lower prior RMSI for velocity observations (2% lower for KOUN and 3% lower for KTLX) and similar ensemble spread (within 2%) relative to the \( V_r \)-only experiment. According to these measures, we consider the update \( q_r \), \( q_s \), \( q_h \) experiment the "best" that was produced in this study. Further supporting this claim, evidence will be provided shortly that the update \( q_r \), \( q_s \), \( q_h \) experiment results in a better analysis near the surface than does the \( V_r \) and \( Z_{\text{dB}} \) experiment.

c. Model-space diagnostics for the \( V_r \)-only, \( V_r \) and \( Z_{\text{dB}} \), and update \( q_r \), \( q_s \), \( q_h \) experiments

We now evaluate results in model space (Fig. 8) for quantities integrated over the entire model domain: total masses of \( q_r \), \( q_r \), \( q_s \), \( q_h \), and \( q_i \); and total volume of updraft for a threshold \( w \geq 5 \text{ m s}^{-1} \). For the three prognostic variables in the reflectivity observation operator—\( q_r \), \( q_h \), and \( q_h \) (Fig. 8b), \( q_h \) (Fig. 8c), and \( q_h \) (Fig. 8d)—it is not surprising that the \( V_r \) and \( Z_{\text{dB}} \) and update \( q_r \), \( q_s \), \( q_h \) experiments, in which reflectivity observations are assimilated, have trends in the domain totals that are more similar to each other than to the trends for the \( V_r \)-only experiment, in which reflectivity observations are not assimilated. Relative to the \( V_r \)-only experiment, the \( V_r \) and \( Z_{\text{dB}} \) and update \( q_r \), \( q_s \), \( q_h \) experiments generally increase the amount of cloud droplets, decrease the amount of rain, decrease the amount of hail–graupel, and slightly decrease the amount of snow. Late in section 3, it was noted that on average, reflectivity-data assimilation decreases the reflectivity between 3 and 8 km AGL and increases the reflectivity below 3 km AGL (Fig. 4). The changes in rain and hail–graupel at midlevels resulting from the reflectivity-data assimilation dominate the domain-total statistics in Figs. 8b,c relative to the changes at low levels.

Before 2130 UTC, the \( V_r \) and \( Z_{\text{dB}} \) experiment develops high updraft volume much more quickly than the other experiments (Fig. 8f); consistent with this result, the \( V_r \) and \( Z_{\text{dB}} \) experiment also generally has more cloud (Fig. 8a). Figure 3 provides an example of these significant differences, showing much stronger updrafts in the developing Oklahoma City storm cluster (from \( X = -50, Y = -10 \) to \( X = -40, Y = 0 \)) in the \( V_r \) and \( Z_{\text{dB}} \) experiment (Fig. 3d) than in the \( V_r \)-only experiment (Fig. 3c). During this time in the storm’s history, the updraft was relatively far from the radar—roughly
50 km away. We speculate that the relative contributions of reflectivity- and velocity-data assimilation to the analyses could be range dependent. When a storm is close to a radar, the Doppler-velocity observations resolve convergent, divergent, and rotational properties of the updraft, whereas when a storm is far from a radar, the velocity observations could be too coarse to resolve the updraft. Therefore, at far ranges, where the Doppler-velocity observations provide less information, the covariances between reflectivity and unobserved fields, such as $w$ and $q_a$, could contribute proportionally more to the retrieval of the updraft and cloud. Further support of this speculation comes from discussions with our colleagues (particularly M. Coniglio 2009, personal communication) about data-assimilation results for other cases (not shown) in which the convective storms were 100 km from the radar. For these cases, adding reflectivity-data assimilation to the velocity-data assimilation also helped produce strong storms in the model more quickly. For three-dimensional variational data assimilation (3DVar) of Doppler-velocity observations and ‘cloud analysis’ based on reflectivity observations for convective storms that were roughly 50–200 km from the radar, Hu et al. (2006) also found that reflectivity observations were very important for supporting the development of strong storms during the data assimilation.

d. Cold-pool analyses

The remaining discussion in this section about how reflectivity-data assimilation affects storm-scale analyses focuses on the analyzed pools of cold air near the surface. Cold pools in supercell thunderstorms develop from the evaporation, melting, and sublimation of precipitation. Dowell et al. (2004b) discuss the importance of these cold pools for storm evolution and the difficulties in analyzing them from Doppler radar observations. In the current study, cold-pool analyses are very sensitive to some aspects of the data-assimilation system. Given the limited surface observations available for this case study, we cannot determine definitively which analysis is best, but we can rely on surface observations in the cold pool from one site—KOKC (Figs. 1 and 9a)—for model verification.

Temperature analyses at 250 m AGL at 2210 UTC in 6 different experiments (Fig. 9) illustrate how reflectivity-data assimilation affects the cold pool and, when reflectivity data are assimilated, how the analyses depend on ensemble spread. Both reflectivity and Doppler-velocity observations are assimilated into NCOMMAS in five of the six experiments (Figs. 9a–e), but only Doppler velocity is assimilated in the sixth (Fig. 9f). The six experiments produce a considerable variety of temperature
analyses, indicating that reflectivity-data assimilation can indeed have a significant influence on the low-level cold pool. Relative to the $V_r$-only experiment, experiments in which reflectivity observations are also assimilated have significantly different magnitudes of temperature perturbations in the cold pool (e.g., Figs. 9c vs. 9f), different storm-relative locations and horizontal extents of the cold air (e.g., Figs. 9a,d vs 9f), and different horizontal scales of structures in the cold pool (small-scale features in Figs. 9a–c vs smoother features in 9f).

A comparison of the $V_r$ and $Z_{dB}$ experiment and a “no-θ update” sensitivity experiment (Table 1) demonstrates how reflectivity-data assimilation contributes significantly to the cold-pool analysis through covariances between reflectivity and temperature. When reflectivity observations are assimilated, all model fields are updated in the $V_r$ and $Z_{dB}$ experiment, whereas all model fields except potential temperature are updated in the no θ update experiment. The $V_r$ and $Z_{dB}$ experiment (Fig. 9a) produces a much larger and stronger cold pool than the no-θ update experiment (Fig. 9d). At low levels, the correlation field between reflectivity and potential temperature is dominated by negative values (Fig. 10a), likely related to cooling by evaporation of rain. More rain implies higher reflectivity, and more rain also implies more evaporative cooling. Thus, a negative correlation between reflectivity and temperature develops at low levels. Higher up in the storm (e.g., at 8.25 km AGL in Fig. 10b), the situation is more complicated. Large areas of both positive and negative correlations develop, likely related to both heating and cooling by moist processes and strong vertical motions (heating by compression in downdrafts and cooling by expansion in updrafts).

At low levels, ensemble-mean reflectivity is on average too low throughout the assimilation period (Fig. 4). Therefore, on average, reflectivity-data assimilation repeatedly produces positive increments in $q_r$ and $q_h$ (and corresponding positive increments in the associated reflectivity) at low levels. At the same time, since the correlation with temperature is negative (Fig. 10a), the reflectivity-data assimilation repeatedly produces negative increments in temperature at low levels. Thus, much stronger and more widespread cold pools develop in the
experiments in which reflectivity observations are assimilated and in which potential temperature is updated during this data assimilation (Figs. 9a–c) than in the other experiments (Figs. 9d–f). The analyses are extrapolating, through the ensemble covariances, the bias in reflectivity into systematic changes in temperature. While the reflectivity-temperature covariances are physically plausible, there is no reason to expect them to account for relations of bias in these model variables.

Experiments with different magnitudes of ensemble spread (Table 1), controlled by the specified magnitudes of additive noise, further reveal how reflectivity-data assimilation affects the low-level cold pools. In most experiments, including $V_r$ and $Z_{db}$, the standard deviations of additive gridpoint noise, before smoothing, were 1.0 (0.25) m s$^{-1}$, 1.0 (0.25) m s$^{-1}$, 1.0 (0.25) K, and 1.0 (0.25) K before (after) 2100 UTC for $u$, $v$, $\theta$, and $T_d$, respectively (cf. Dowell and Wicker 2009). In a “less additive noise” sensitivity experiment, the standard deviations were decreased to 0.1 m s$^{-1}$, 0.1 m s$^{-1}$, 0.1 K, and 0.1 K after 2100 UTC for $u$, $v$, $\theta$, and $T_d$, respectively, and in a “more additive noise” experiment, the standard deviations were 0.5 m s$^{-1}$, 0.5 m s$^{-1}$, 0.5 K, and 0.5 K after 2100 UTC. The strength and coverage of the cold pool (Figs. 9a–c) depends greatly on the ensemble spread (Fig. 11) associated with these different magnitudes of additive noise. A region in which the cold pool is particularly sensitive to the amount of ensemble spread is the forward flank (downshear anvil) at low levels, roughly in the location indicated by the arrow in Fig. 9. As noted earlier, the model was generally underpredicting reflectivity in this region (blue colors in Figs. 5c,g, 0 $< X < 30$ and 15 $< Y < 40$). Thus, reflectivity-data assimilation repeatedly introduces positive rain and negative temperature increments in the downshear anvil near the surface. Since greater ensemble spread results in larger increments, the experiment with the greatest ensemble spread (more additive noise) has the strongest and most widespread cold pool (Fig. 9c).

For ensemble forecasting and radar-data assimilation for real supercells, it is not clear what the ideal amounts of ensemble spread are for temperature and other variables. We do not have widespread temperature observations available for verification and must instead rely on the radar observations to evaluate consistency between ensemble spread and forecast errors. Relative to the $V_r$ and $Z_{db}$ experiment, the less additive noise and more additive noise experiments changed the prior ensemble spread in reflectivity (velocity) by as much as $-24\%$ and $+47\%$ ($-29\%$ and $+49\%$), respectively (Fig. 11).

One could argue that the more additive noise experiment is best, since it results in prior consistency ratios, as defined by Dowell and Wicker (2009), closest to the desired value of 1.0 for KOUN and KTLX velocity (not shown). One could also argue that the $V_r$ and $Z_{db}$ and less additive noise experiments are better than the more additive noise experiment because they have lower values of prior RMSI for KOUN and KTLX velocity (not shown). Thus, the range of ensemble spread in these experiments seems well within the range of plausible uncertainty, but this variation in ensemble spread yielded large differences in ensemble-mean temperature (Figs. 9a–c). An analysis that depends so much on ensemble spread could be viewed as nonrobust, that is, not to be trusted.
As mentioned previously, with the few available observations we cannot determine which analysis of the cold pool in the 8 May 2003 Oklahoma City supercell is the best. Nevertheless, a comparison of the KOKC observations and the ensemble-mean analyses at the same location could suggest that the strong cold pool in the forward flank in the $V_r$ and $Z_{dB}$ experiment (Fig. 9a) is erroneous (Fig. 12a, which shows that the ensemble-mean temperature is too low at 2152 and 2159 UTC). In contrast, the update $q_r, q_s, q_h$ (Fig. 12b) and $V_r$-only (Fig. 12c) experiments more correctly capture the gradual cooling at the KOKC site before 2200 UTC as the storm approaches. Based on the limited observations from KOKC, we might conclude that the cold pool analyses in the no-$\theta$ update, update $q_r, q_s, q_h$, and $V_r$-only experiments (Figs. 9d–f) are the most plausible.

5. Summary and discussion

An ensemble Kalman filter was used to assimilate radar observations of the 8 May 2003 Oklahoma City tornadic supercell thunderstorm into a 50-member NCOMMAS ensemble. Variability among the ensemble members was introduced through random perturbations to the base-state wind profiles and local random perturbations added to the $u, v, \theta$, and $q_h$ fields where the observations indicated precipitation. A number of simplifications were employed for the simulations: an initial environment that was horizontally homogeneous, with profiles obtained from a National Weather Service sounding near the storm; a flat lower boundary; no surface fluxes; no radiative transfer; and a single-moment bulk (Lin–Gilmore) microphysics scheme.

In a reference experiment, only KOUN Doppler-velocity observations were assimilated into the NCOMMAS ensemble. Then, KOUN reflectivity observations together with velocity observations were assimilated in all other experiments. Sensitivity experiments were conducted to reveal how the reflectivity observations affected the storm-scale analyses, including experiments with different specified values of observation error, different amounts of ensemble spread (additive noise), and different choices for what model variables are updated by the EnKF when reflectivity observations are assimilated. Doppler-velocity observations from a WSR-88D (KTLX) were used for verification, in addition to the KOUN velocity and reflectivity observations. Results from all data-assimilation experiments were encouraging in that 1) realistic storm-scale structures were obtained in the analyses and 2) assimilating KOUN observations—either just Doppler velocity or both Doppler velocity and reflectivity—resulted in a better posterior than prior fit to independent KTLX observations.

Issues associated with reflectivity-data assimilation were explored for only one real-data case. Nevertheless, many of these issues should be relevant for any attempts to use ensemble Kalman filters to assimilate reflectivity observations of convective storms directly into numerical cloud models. Not surprisingly, assimilating reflectivity observations significantly affected the analyses of the hydrometeor variables in the reflectivity observation operator: rain, snow, and hail–graupel. Assimilating reflectivity observations also significantly affected the analyses of unobserved variables, including temperature, cloud, and vertical velocity. Large updraft volume developed more quickly in the experiments in which reflectivity and Doppler-velocity observations were assimilated together rather than just Doppler velocity, presumably through ensemble covariances between reflectivity and unobserved variables such as cloud and vertical velocity. We speculate that reflectivity observations are particularly useful for producing mature storms quickly in an ensemble when the storms are far from the radar, when Doppler-velocity observations are too coarse to resolve the updrafts.

Differences between simulated (ensemble mean) and observed reflectivity had regular patterns in a storm-relative
sense throughout the assimilation period. Biases in the reflectivity observations, the reflectivity observation operator, and the model hydrometeor predictions are candidate explanations for these patterns. We are unable to assess whether there are any biases in the reflectivity observations associated with radar miscalibration, but since the KOUN radar is a 10-cm radar, we expect that errors associated with attenuation should be small, much less than for research radars with shorter wavelengths (e.g., 5 or 3 cm). Since there is considerable uncertainty in how to calculate reflectivity from model hydrometeor fields, particularly for ice and mixed-phase precipitation, bias errors in the reflectivity observation operator are a plausible error source in these experiments. Other reflectivity observation operators that have been proposed for the Lin–Gilmore microphysics scheme (Jung et al. 2008a) would be worth trying in future work with this case. Model error is also a very likely error source in the current experiments. One noteworthy discrepancy between the predicted and observed reflectivity occurred in the downshear anvil at low levels, where reflectivity was consistently underpredicted by as much as 20 dBZ. Since rain was the precipitation type in this region, and since errors in the reflectivity observation operator are believed to be relatively small for rain, we suggest that errors in the model’s $q_r$ predictions and/or the associated drop-size distributions were responsible for the underpredicted reflectivity.

Throughout the data-assimilation period, reflectivity-data assimilation repeatedly added rain in the downshear anvil at low levels, where reflectivity was underpredicted. Simultaneously, through negative correlations with temperature, reflectivity-data assimilation repeatedly introduced negative increments in temperature in the same region. Based on the limited available surface observations, we would conclude that the forward-flank cold pools that developed in these experiments were too strong. In general, the operational observing network does not provide cold-pool observations on the scale of individual convective storms. Thus, since we cannot verify cold-pool analyses, we propose instead that robustness is a more practical quality measure for storm-scale analyses obtained by radar-data assimilation. If storm-scale temperature analyses were robust, then ensemble-mean cold-pool analyses would vary only slightly over the plausible range of ensemble spread. Unfortunately, for the case shown, when reflectivity data are assimilated, the temperature analyses depend greatly on ensemble spread, and thus the temperature analyses are not robust.

Bias errors in the hydrometeor predictions appear to be a major error source in these experiments. An option that was presented for reflectivity-data assimilation in this situation was only updating a subset of model variables, so that errors in the hydrometeor forecasts do not project directly into other variables during the data assimilation. An experiment in which only $q_r$, $q_s$, and $q_h$ were updated by the EnKF during the reflectivity-data assimilation resulted in the best verification scores for the experiment set in this study. A drawback was that not updating all model variables resulted in slower storm growth in the model (Fig. 8f), presumably because some useful covariances were being ignored.

A more attractive approach that could be explored for future work would be to improve the hydrometeor forecasts, with a more sophisticated microphysics scheme being one possible option. Double- and triple-moment schemes (Seifert and Beheng 2001; Milbrandt and Yau 2005; Morrison and Grabowski 2008) could mitigate some of the model errors that affect our results. Another possible approach is parameter estimation (Tong and Xue...
2008a,b), that is, using the observations to optimize a small number of parameters in the microphysics scheme. This approach could offer some benefit for real-data cases, but for this particular case, analyzed differences between predicted and observed reflectivity showed considerable spatial variability, suggesting that different parts of the storm could have significantly different optimal microphysics parameters.

Although beyond the scope of the current study, developing methods to satisfy better the conditions for EnKF optimality could be a worthwhile endeavor in future studies of radar-data assimilation for convective storms. For example, the observation operator for reflectivity is nonlinear (Tong and Xue 2005 and the appendix of the current paper) because the application of (A.8) results in a logarithmic relationship between $Z_{\text{DB}}$ and individual hydrometeor mixing ratios. Lorenc (2003) shows an example of how posterior distributions are suboptimal after an observation that is nonlinearly related to the model state is assimilated. We believe we see evidence of related problems in our results (not shown). One possible approach that could be considered for future studies would be to use an iterative method (Annan et al. 2005) to improve the posterior mean and variance when reflectivity observations are assimilated. Other suggestions for future work include 1) developing more rigorous and/or adaptive methods for choosing localization radii for reflectivity observations in precipitation and “no precipitation” observations; 2) including a more realistic mesoscale environment in the model, so that multi-hour forecasts after the data-assimilation window can provide an additional basis for evaluating experiments; 3) repeating experiments like those in the current study for other cases with a variety of storm modes; and 4) choosing cases that have more detailed observations, particularly thermodynamic observations, available for verification.

Acknowledgments. D. Dowell’s work on this project was supported by the National Science Foundation under Grant ATM-0333872 at the Cooperative Institute for Mesoscale Meteorological Studies and by the National Center for Atmospheric Research. C. Snyder’s work on this project was supported by the National Science Foundation under Grant ATM-0205655. Matt Gilmore developed the precipitation-microphysics scheme used in these experiments and answered questions about the code. Mike Coniglio, Ted Mansell, and Glen Romine contributed to the data-assimilation code development. Kevin Scharfenberg and Gordon Carrie provided the radar observations of the 8 May 2003 Oklahoma City supercell. Discussions with Altug Aksoy, Glen Romine, Ted Mansell, and Don Burgess were quite beneficial. Three anonymous reviewers provided helpful suggestions for improving the manuscript. Mary Haley provided valuable assistance with NCL graphics.

APPENDIX

Reflectivity Estimation from Model Fields

When the Rayleigh approximation (radar wavelength $\gg$ particle diameter) is valid, the reflectivity factor for spherical raindrops is

$$Z = \int_0^\infty n_r(D)D^6 \, dD,$$  
(A.1)

where $D$ is the raindrop diameter and $n_r(D)$ is the raindrop number concentration per unit diameter (Doviak and Zrnic 1993). For reflectivity associated with non-spherical raindrops, ice particles of any shape, and/or mixtures of particle types, the term “equivalent reflectivity factor” ($Z_e$) is used instead. Specifically, $Z_e$ is the reflectivity factor of a hypothetical target of spherical raindrops that would produce the same amount of backscattering as the observed target. To obtain the equivalent reflectivity factor for simulated storms, we sum the contributions from three hydrometeor categories:

$$Z_e = Z_r + Z_s + Z_h,$$  
(A.2)

where $Z_r$, $Z_s$, and $Z_h$ are the equivalent reflectivity factors associated with rain, snow, and hail–graupel, respectively. The reflectivity values associated with the other two categories—cloud droplets and cloud-ice crystals—are considered negligible and thus not included in (A.2).

The Lin–Gilmore bulk microphysics scheme assumes that raindrops have an inverse exponential size distribution (Gilmore et al. 2004a; Kessler 1969; Smith et al. 1975):

$$n_r(D) = n_{0r} \exp \left[ -D \left( \frac{\rho p n_{0r}}{q_r} \right)^{0.25} \right],$$  
(A.3)

where $n_{0r}$ (m$^{-4}$) is the intercept parameter, $\rho$ is the air density (kg m$^{-3}$), $p_r$ is the density of water (1000 kg m$^{-3}$), and $q_r$ and all other mixing ratios in this section are in units of kilograms per kilogram. This equation for the size distribution, together with the definition of the reflectivity factor in (A.1), leads to the following relationship between the model fields and the reflectivity factor associated with rain:

$$Z_r = \frac{7.2 \times 10^{20} (pq_r)^{1.75}}{\pi^{1.75} n_{0r}^{0.75} p_r^{1.75}},$$  
(A.4)
where $Z_r$ and all other equivalent reflectivity factors in this section are in mm$^6$ m$^{-3}$. For the intercept parameter used in the current experiments ($n_{gh} = 8 \times 10^4$ m$^{-4}$; Lin et al. 1983; Gilmore et al. 2004a), $Z_r = 3.63 \times 10^6 (\rho_q h)^{1.75}$.

Since ice hydrometeors have various shapes, orientations, and degrees of surface wetness, the relationship between reflectivity and ice is more uncertain than it is for rain (Smith et al. 1975; Smith 1984). As is the case for rain, snow in the Lin–Gilmore bulk scheme has an inverse exponential size distribution, with the diameter $D$ in this case referring to the size of the resulting water drop if the snowflake were to melt. In our experiments (and also in the experiments of Tong and Xue 2005), we assume that when the temperature is above (below) freezing, snow has a wet (dry) surface. Thus, following Smith et al. (1975) and Smith (1984), the relationship between reflectivity and the mixing ratio of snow is

$$Z_s = \frac{0.23 (\rho_s h^2) 7.2 \times 10^{20} (\rho_q h)^{1.75}}{\pi^{1.75} n_{0s}^{0.75} \rho_s^{1.75}}, \quad T \leq 0^\circ\mathrm{C} \text{ (dry snow),} \quad (A.5)$$

$$Z_s = \frac{7.2 \times 10^{20} (\rho_q h)^{1.75}}{\pi^{1.75} n_{0s}^{0.75} \rho_s^{1.75}}, \quad T > 0^\circ\mathrm{C} \text{ (wet snow),} \quad (A.6)$$

where the coefficient 0.23 in (A.5) is the square of the dielectric constant for ice divided by the square of the dielectric constant for water (0.21/0.93 = 0.23; Smith 1984), $\rho_s$ is the density of snow, and $n_{0s}$ is the intercept parameter for snow. For the chosen values $n_{0s} = 3 \times 10^6$ m$^{-4}$ and $\rho_s = 100$ kg m$^{-3}$ (Lin et al. 1983; Gilmore et al. 2004a), Eq. (A.5) and (A.6) become $Z_s = 9.80 \times 10^6 (\rho_q h)^{1.75}$ and $Z_s = 4.26 \times 10^{11} (\rho_q h)^{1.75}$, respectively.

The relationship between the hail–graupel mixing ratio and reflectivity also depends significantly on whether the surface is wet or dry (Smith et al. 1975). The relationship is complicated because evaporation of surface water and absorption of surface water into the interior can give graupel a dry surface even when the temperature is above freezing. We considered two simplifying assumptions for our experiments; one possibility is that the hail–graupel has a dry (soft) surface when the temperature is above (below) freezing, whereas another possibility is that the hail–graupel always has a dry surface. For the 8 May 2003 experiments, we found better agreement between observed and simulated reflectivity when the latter assumption was made. Thus, for all experiments discussed in this paper, hail–graupel is assumed to be dry, resulting in the following equation for reflectivity (Smith et al. 1975):

$$Z_h = \frac{0.23 (\rho_q h^2) 7.2 \times 10^{20} (\rho_q h)^{1.75}}{\pi^{1.75} n_{0h}^{0.75} \rho_h^{1.75}}, \quad (A.7)$$

where $\rho_h$ and $n_{0h}$ are the density and intercept parameters of hail–graupel, respectively. We use the default Lin et al. (1983) and Gilmore et al. (2004a) parameters: $\rho_h = 900$ kg m$^{-3}$ and $n_{0h} = 4 \times 10^4$ m$^{-4}$. In this case, (A.7) becomes $Z_h = 4.33 \times 10^{10} (\rho_q h)^{1.75}$. Although our reflectivity observation operators for rain and snow are very similar to those in Tong and Xue (2005), our observation operator for hail–graupel is different from theirs. Tong and Xue (2005) used the wet-hail form of the observation operator, and thus, all else being equal, they would obtain larger reflectivity values from hail–graupel.

The final step in computing reflectivity is to convert to the customary logarithmic scale:

$$Z_{\text{dB}} = 10 \log_{10} Z_r, \quad (A.8)$$

where the units of $Z_{\text{dB}}$ are dBZ. The logarithmic scale is convenient not only for graphical displays of the wide range of reflectivity values in observed precipitation, but also for representing random errors in reflectivity, which are commonly assumed to have a uniform error variance in dB (e.g., Wilson and Brandes 1979; Doviak and Zrnic 1993). A threshold for a minimum $Z_{\text{dB}}$ value has to be specified when mapping model hydrometeor fields to reflectivity because $Z_{\text{dB}}$ in (A.8) is undefined at locations where the hydrometeor mixing ratios are all zero. After some experimentation, we chose a minimum $Z_{\text{dB}}$ threshold of 0 dBZ for our experiments, as did Tong and Xue (2005). The minimum threshold is applied both to the observations before they are assimilated and to the values returned by the reflectivity observation operator.

REFERENCES


