The Impact of Covariance Localization for Radar Data on EnKF Analyses of a Developing MCS: Observing System Simulation Experiments

RYAN A. SOBASH
School of Meteorology, and Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma

DAVID J. STENSRUD
NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma

(Manuscript received 18 July 2012, in final form 23 May 2013)

ABSTRACT

Several observing system simulation experiments (OSSEs) were performed to assess the impact of covariance localization of radar data on ensemble Kalman filter (EnKF) analyses of a developing convective system. Simulated Weather Surveillance Radar-1988 Doppler (WSR-88D) observations were extracted from a truth simulation and assimilated into experiments with localization cutoff choices of 6, 12, and 18 km in the horizontal and 3, 6, and 12 km in the vertical. Overall, increasing the horizontal localization and decreasing the vertical localization produced analyses with the smallest RMSE for most of the state variables. The convective mode of the analyzed system had an impact on the localization results. During cell mergers, larger horizontal localization improved the results. Prior state correlations between the observations and state variables were used to construct reverse cumulative density functions (RCDFs) to identify the correlation length scales for various observation-state pairs. The OSSE with the smallest RMSE employed localization cutoff values that were similar to the horizontal and vertical length scales of the prior state correlations, especially for observation-state correlations above 0.6. Vertical correlations were restricted to state points closer to the observations than in the horizontal, as determined by the RCDFs. Further, the microphysical state variables were correlated with the reflectivity observations on smaller scales than the three-dimensional wind field and radial velocity observations. The ramifications of these findings on localization choices in convective-scale EnKF experiments that assimilate radar data are discussed.

1. Introduction

During the past decade, the ensemble Kalman filter (EnKF; Evensen 1994), a Monte Carlo approximation to the linear Kalman filter (Kalman 1960), has emerged as a data assimilation technique capable of merging observations with a first-guess model state to provide analyses of atmospheric features (Houtekamer and Mitchell 1998, hereafter HM98; Houtekamer and Mitchell 2001, hereafter HM01). More recently, the feasibility of the EnKF data assimilation technique for convective storm analyses has been demonstrated, first in observing system simulation experiments (OSSEs; e.g., Snyder and Zhang 2003; Caya et al. 2005; Tong and Xue 2005) and later in experiments using radar data collected within observed convective storms (e.g., Dowell et al. 2004; Aksoy et al. 2009; Dowell et al. 2011; Dawson et al. 2012).

Insufficient ensemble size leads to degraded background error covariance statistics due to sampling error, resulting in a suboptimal data assimilation update that is propagated forward by the ensemble advance. For convective-scale EnKF applications, having a large enough ensemble of model simulations (>100 members) to mitigate sampling errors is computationally prohibitive, thus these errors are inevitable for such experiments. A ubiquitous treatment to reduce these errors is to confine the impact of an observation to nearby model grid points. Errors in the sample correlation are greatest where the true correlations are small (Fisher 1915), thus the true correlation structure is more robustly measured at grid points near the observation. Thus, using localization implies that distant estimates of nonzero correlations...
are more vulnerable to sampling error and should be ignored.

Localization is commonly implemented by weighting the observation-state covariance estimates from the ensemble using a function that becomes zero beyond a specified cutoff distance. This was first applied using a quasigeostrophic model by HM98 and further refined to utilize a Schur (elementwise) product for filtering noisy covariances. This has several benefits, including reducing computational burden by only updating a subset of model state points, and increasing the rank of the background error covariance matrix (Hamill et al. 2001).

Improper localization can introduce imbalance into the analysis, thus optimally choosing localization is essential for numerical weather prediction (Greymbush et al. 2011). Flow-dependent error covariance structures vary in time and space; adaptive localization techniques that permit these structures (e.g., Bishop and Hodyss 2009a,b) could provide more benefit than the fixed weighting schemes described above, but have not yet been applied to convective-scale EnKF.

All previous convective-scale EnKF studies have implemented some manner of covariance localization (Table 1). When implementing localization, the two adjustable parameters are the weighting function and the cutoff distance $r$. Early studies used a boxcar weighting function (e.g., HM98), while later studies (e.g., HM01) have implemented a weighting function that approximates a Gaussian function (Gaspari and Cohn 1999). Often, tuning is used to select an optimal $r$ for the weighting function, although the specific sensitivities often go undocumented or a cutoff is chosen following recommendations in previous work. Ultimately, the optimal localization cutoff depends on the ensemble size, state variable, observation type, density, location, and the dynamics of the observed system. Many of these characteristics differ among the previous studies, preventing any robust conclusions about the effects of localization on the analyses, especially those that utilize WSR-88D datasets.

Early studies using OSSEs chose $r$ between 4 and 8 km, with equivalent values for the horizontal and vertical cutoffs (i.e., the region of influence around an observation is a sphere). For example, Snyder and Zhang (2003), in the first documented convective-scale perfect-model OSSE of a supercell, used $r = 4$ km for a 50-member ensemble with a boxcar weighting function. While not the focus of their work, their results showed some sensitivity to the cutoff, with large $r$ values (>$20$ km) being detrimental to the analyses and, little sensitivity for $r$ values between 2 and 6 km. Using the GC weighting function improved their results over the boxcar function, and thus later studies have almost exclusively used the GC function for localization.

Caya et al. (2005) further examined the analysis error of differing localization cutoffs. The smallest errors occurred using a value of 7.3 km for $r$ in their 100-member OSSEs, but they noted that a smaller $r$ ($\sim 5.5$ km) should be used for a 50-member ensemble. Tong and Xue (2005) also used a 100-member ensemble, but with $r = 8$ km. In a real-data study, Dowell et al. (2004) used $r = 6$ km, and concluded that cutoff radii between 4 and 10 km produced the best results, while $r = 2$ km was detrimental to the analysis. Aksoy et al. (2009) chose $r = 5$ km (4 km) horizontal (vertical) cutoff, and found little sensitivity to a larger horizontal cutoff radius, but noted larger sensitivity to the vertical localization cutoff. Later studies generally settled on $r = 6$ km (Dowell and Wicker 2009; Dowell et al. 2011; Dong et al. 2011), although Dawson et al. (2012) used a larger horizontal cutoff ($r = 12$ km) than all previous studies. In other

Table 1. Summary of covariance localization choices in a selection of previous convective-scale EnKF studies. Radar observation spacing of “radar scans” indicates observations are assimilated at model grid points in the horizontal, but kept on the radar sweep surfaces in the vertical. “Model grid” indicates observations are located at model grid points in both the horizontal (H) and vertical directions (V).

<table>
<thead>
<tr>
<th>Reference</th>
<th>Ensemble size</th>
<th>Analysis grid (km)</th>
<th>Radar observation spacing</th>
<th>Localization cutoff (r, km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Snyder and Zhang (2003)</td>
<td>50</td>
<td>2</td>
<td>Model grid</td>
<td>H: 4; V: 4</td>
</tr>
<tr>
<td>Dowell et al. (2004)</td>
<td>50</td>
<td>2</td>
<td>Radar scans</td>
<td>H: 6; V: 6</td>
</tr>
<tr>
<td>Tong and Xue (2005)</td>
<td>100</td>
<td>2</td>
<td>Model grid</td>
<td>H: 3; V: 3</td>
</tr>
<tr>
<td>Caya et al. (2005)</td>
<td>100</td>
<td>2</td>
<td>Radar scans</td>
<td>H: 5; V: 4</td>
</tr>
<tr>
<td>Aksoy et al. (2009)</td>
<td>50</td>
<td>2</td>
<td>Radar scans</td>
<td>H: 6; V: 6</td>
</tr>
<tr>
<td>Dowell and Wicker (2009)</td>
<td>50</td>
<td>1</td>
<td>Radar scans</td>
<td>H: 5; V: 6</td>
</tr>
<tr>
<td>Dowell et al. (2011)</td>
<td>50</td>
<td>1</td>
<td>Radar scans</td>
<td>H: 6; V: 6</td>
</tr>
<tr>
<td>Dong et al. (2011)</td>
<td>50</td>
<td>2</td>
<td>Radar scans</td>
<td>H: 6; V: 6</td>
</tr>
<tr>
<td>Dawson et al. (2012)</td>
<td>30</td>
<td>1</td>
<td>Radar scans</td>
<td>H: 12; V: 6</td>
</tr>
</tbody>
</table>

1 Here, the cutoff distance $r$ is the distance beyond which the localization weights are zero. Some literature defines the term “cutoff” to mean a half-width, that is, $0.5r$. 

real-data studies, attempts were made to extract information about multiple scales of motion from radar observations by using varying localization cutoffs for different groups of observations (Zhang et al. 2009). While this was employed to produce analyses of a hurricane using dense radar observations, the technique is also potentially useful for convective storm analyses.

In an attempt to isolate the impacts of changes to covariance localization on convective-scale EnKF analyses, OSSEs of a developing mesoscale convective system (MCS) were conducted using a variety of cutoff distances encompassing the range of values used in previous work. MCSs are understudied in the convective-scale EnKF literature, with supercells garnering most of the attention [all the referenced studies in Table 1 have focused on supercells, with the exception of Aksoy et al. (2009), which included a supercell and squall-line case]. In addition to testing the sensitivity of the EnKF analyses to the localization cutoff, the results herein could provide evidence of differences in localization between convective modes (e.g., isolated supercells versus squall lines).

To capture such differences, the present OSSEs simulate the process of upscale growth from initially isolated cells into a linear squall line during a 150-min period (much longer than any previous OSSE or real-data convective-scale EnKF experiment). The main thrust of this work aims to understand the impacts of changes to the error of the state fields and by doing so, gaining an understanding of the role of localization in producing accurate convective-scale analyses across a spectrum of convective modes. Further, the proper choice of localization is related to the covariance structures present within a simulated ensemble of convective storms. Little attention has been paid in the convective storms literature to understanding these structures; the analyses herein relate the covariance structures to the localization choices.

Section 2 describes the experimental design and sections 3 and 4 examine the impacts of the localization cutoff on analysis error, and the analysis increments respectively. Section 5 examines the correlation length scales between the observations and state, while section 6 examines short-term ensemble forecasts spawned from the final EnKF analyses. Section 7 contains conclusions and a synthesis of the results.

2. Methods

a. Truth simulation

The Advanced Research Weather Research and Forecasting Model (ARW-WRF) (Skamarock et al. 2005) V3.2.1 was used to create a truth simulation of a developing MCS. The model was initialized using a thermodynamic profile derived from 28 proximity soundings from strong, weakly forced MCS events (Coniglio et al. 2006; Fig. 1). Coniglio et al. (2006) tested a variety of wind shears in the environmental wind profile; the wind shear used herein produced a strong, long-lived, upright-tilted squall line by the end of the simulation. This shear profile (Fig. 1) contains 20 m s$^{-1}$ of 0–5-km westerly shear, sufficient for long-lived squall lines in model simulations (Weisman and Rotunno 2004). Convection was triggered with five +3 K bubbles, with random perturbations added along the bubble edges to initiate three-dimensional motions. The horizontal grid spacing is 3 km and spanned 396 km in both horizontal directions (133 grid points), and 16 km in the vertical (40 grid points). These dimensions, larger than most EnKF convective-scale experiments, were required to retain the system within the domain during the 3-h period. The ARW-WRF was integrated as a cloud model with WRF single-moment 6-class microphysics scheme (WSM6; Hong and Lim 2006), open-boundary conditions, no terrain, surface physics, or boundary layer physics. The Coriolis force was permitted to act on the perturbation wind, as is customary in convective-scale cloud modeling. A prognostic 1.5-order turbulence kinetic energy closure is used to compute the horizontal and vertical eddy viscosities necessary to represent the effects of subgrid-scale turbulence.

b. Simulated radar observations

Synthetic radial velocity and reflectivity observations were extracted from the truth simulation every 5 min assuming a radar placement at the center of the domain. A severe weather WSR-88D scanning strategy composed of 14 scan elevations (between 0.5$^\circ$ and 19.5$^\circ$) was used. The observations were extracted from each grid column within the range of the radar (230 km), at the intersection of the grid columns and the 14 scan surfaces. This technique for extracting simulated radar observations was also applied in Tong and Xue (2005) and serves to reduce the number of observations and produce a grid-ded observation dataset. The observations were computed by trilinearly interpolating the required model state fields to the location of each simulated radar bin. Observation errors were added to each synthetic observation by performing a random draw from a Gaussian with zero mean and standard deviation of 2 m s$^{-1}$ (2 dBZ) for radial velocity (reflectivity). Reflectivity observations of less than 10 dBZ were considered “clear air” observations for the purposes of this study and the corresponding radial velocity observation at these locations was removed. All other observations are considered “precipitation” observations.
c. Data assimilation and OSSEs

The assimilation experiments were conducted with the EnKF implementation within the Data Assimilation Research Testbed (DART) software (www.image.ucar.edu/DARes/DART). Specifically, the sequential parallel version (Anderson and Collins 2007) of the deterministic ensemble adjustment Kalman filter (Anderson 2001) was used for the data assimilation update. The 50-member ensemble was initialized with different horizontally homogeneous wind profiles (identical thermodynamic profiles), each produced by adding random, uncorrelated Gaussian noise to the wind profile (both $U$ and $V$) of the truth sounding below 11 km AGL. The ensemble generation technique used herein is consistent with other OSSEs and real-data experiments that use homogenous environments for initial conditions (e.g., Dowell et al. 2004). The initial ensemble was the same for each OSSE. The OSSEs were initialized 30 min after the beginning of the truth simulation when appreciable amounts of hydrometeors develop within the simulated storm. The synthetic observations were assimilated every 5 min for 150 min. To increase ensemble spread, additive, smoothed, perturbations were added to the wind, temperature, and dewpoint fields where the observed (i.e., simulated) reflectivity is greater than 20 dBZ immediately prior to the ensemble advance (Dowell and Wicker 2009) throughout the length of the assimilation period. The additive noise smoothing procedure prevents any detrimental effects due to discontinuities on the edges of the convective system. The magnitudes of additive noise (Table 2) were chosen using values from previous work (e.g., Dowell and Wicker 2009). The additive noise magnitudes were not tuned for each experiment, yet the results herein appear to be robust. Comments about additive noise sensitivities are found in section 4c.

The OSSEs use the same model configuration as the truth simulation (i.e., they are perfect-model experiments). In addition, the experiments are identical except for the localization cutoffs $r$. Values of 6, 12, and 18 km...
are selected for \( r_h \) for three experiments to assess the impacts of the horizontal localization (\( r_v = 6 \) km), while in another set of experiments, \( r_v \) is varied between 3, 6, and 12 km while \( r_h \) is held constant at 12 km (Table 2). The localization choices for the clear-air observations remain constant across all experiments; those observations are thinned to 6 km in the horizontal with \( r = 12 \) km. These clear-air observations have been shown to improve convective-scale EnKF analyses by suppressing spurious convection (Aksoy et al. 2009).

### d. Verification metrics and analysis techniques

The fidelity of the analyses to the truth simulation is gauged by domain-averaged gridpoint-based verification metrics, as is common in EnKF analysis verification (e.g., Dowell et al. 2004). In an OSSE, the state estimate can be directly compared to the true state at all state points. Herein, these “state space” diagnostics (as opposed to “observation space” diagnostics that are computed at the observation locations) are the primary metrics used in this study to assess the overall quality of an OSSE. Root-mean-square error (RMSE) of the analyses and 5-min forecasts is computed, defined as

\[
\text{RMSE}^{a,f} = \sqrt{\frac{1}{M} \sum_{i=1}^{M} (x_i^t - \bar{x}^{a,f})^2},
\]

where the subscript \( i \) is an index over all \( M \) grid points where the true state reflectivity is greater than 20 dBZ; and \( t, a, \) and \( f \) represent the true, analysis (i.e., posterior), and forecast (i.e., prior) state, respectively. The use of clear-air observations prevents the development of spurious convection outside the truth state convection (this was confirmed by visually inspecting the ensemble for areas of convection outside the truth state), justifying the use of the truth state for defining the points over which RMSE is computed. State-space RMSE is computed for the three velocity components, temperature, and the WSM6 microphysical species (water vapor, cloud water, cloud ice, rain, snow, and graupel). In addition, the average bias is computed, defined as

\[
\text{BIAS}^{a,f} = \frac{1}{M} \sum_{i=1}^{M} (x_i^t - \bar{x}^{a,f}).
\]

This is analogous to the mean innovation in observation space. The consistency ratio, a comparison of the ensemble mean RMSE to the ensemble spread, is also computed to ensure the ensemble contains an appropriate amount of spread. This ratio is defined as

\[
\text{CR}^{a,f} = \frac{1}{M} \sum_{i=1}^{M} \left[ \frac{1}{N-1} \sum_{j=1}^{N} (x_i^{a,f} - \bar{x}^{a,f})^2 \right] / \left[ \text{RMSE}^{a,f} \right]^2.
\]

where the numerator is the squared ensemble spread (i.e., ensemble variance) and the denominator is the mean squared error [i.e., Eq. (1) squared]. Here \( N \) represents the number of ensemble members. As in Eqs. (1) and (2), the spread computation for CR is averaged over \( M \) grid points where the true state reflectivity is greater than 20 dBZ.

The prior analyses were compared to both the posterior and truth analyses to analyze the spatial structures of the prior analysis errors and the data assimilation increments. While posterior-prior (PO-PR) increments provide a sense of the accuracy of the state estimate throughout the model domain. Both sets of increments were averaged in time and space (along the convective line) over select periods within the OSSE to observe the systematic behavior of each set of increments. The line averages are computed on a domain 200 km across that is translated each output time with the leading edge of the storm’s cold pool.

### 3. State-space RMSE

#### a. Truth simulation

The 3-h truth simulation encompasses the development of five initially isolated cells into a quasi-linear convective system. The homogeneous environment used to initialize the model contains a vertical wind profile...
that produces splitting cells after approximately 30 min (Fig. 2a). By 90 min, the eight pairs of left and right cell splits merge to produce four cells, while the left (right) split of the northernmost (southernmost) cells remain isolated (Fig. 2b). These six cells proceed to merge into a linear convective system by the end of the simulation as the cells’ cold pools merge, forcing convection along the upshear edge (Fig. 2c). The initially isolated cells are located approximately 50–80 km west of the radar at 30 min. By 90 min, the line of cells is collocated with the longitude of the radar, and by the end of the simulation, the squall line is 30–60 km east of the radar.

The process simulated by the model is representative of upscale growth observed in nature, when isolated multicells or supercells in close proximity merge into a convective line after several hours (this is a distinctly different process compared to convection forced along linear boundaries). By 3 h, a well-developed cold pool is established with a vertically upright line of convection along the leading edge of the cold pool. While not used in this study, extending the simulation further (1–2 h) results in a mature bow echo system.

b. LOC6V6 experiment

The LOC6V6 experiment is conducted using $r = 6$ km, a common choice in previous convective-scale data assimilation studies (e.g., Dowell et al. 2004; Dong et al. 2011). The horizontal grid spacing of the observations is 3 km, matching the model grid spacing. Observations extracted from the truth simulation at $T + 30$ min are assimilated at the initial time in LOC6V6. No temperature bubbles are used in the assimilation experiments to promote convective circulations. Thus, the model state takes approximately 20–30 min for the increments to stabilize as convection initially develops. Statistics are shown from this point forward (Fig. 3). Both the prior and posterior RMSE are plotted, hence, the appearance of the “sawtooth” pattern. The RMSE decreases gradually for all state fields through the first 60 min of assimilation. This decrease is purely due to the assimilation of simulated radar data; the RMSE is not reduced because of the choice of boundary conditions, since open boundary conditions were used, or the choice of initial conditions, which is void of convection.

After 60 min, the RMSE increases until approximately 90 min; this 30-min period coincides with the first set of cell mergers (shaded in Fig. 3). Afterward, the RMSE again stabilizes, although an increase in RMSE is evident in some state fields after 110 min (especially $V$, QVAPOR, and QRAIN). The increase in RMSE between 110 and 150 min (shaded in Fig. 3) is likely related to the second set of mergers that occur as the cells and associated cold pools merge near the end of the assimilation period. As the storm passes the radar location between 60 and 120 min, portions of the storm become unobserved due to the “cone of silence” above the radar. The effect of this on the RMSE statistics (and other analyses) is minimal since the unobserved portions of the storm are small compared to the size of the convective system. In addition, the increase in RMSE
between 60 and 90 min occurs when the main convective cores are well observed and only a segment of the anvil is unobserved. Thus, the increase in RMSE is thus most likely due to the cell merger process and not gaps in observations.

Comparing the velocity fields, the $V$ field has the smallest RMSE compared to $U$ and $W$ throughout the assimilation period. This may simply be a reflection of a smaller overall $V$ component of the horizontal wind compared to the $U$ component, since the system is dominated by momentum in the west–east direction. While the horizontal wind is directly related to radial velocity, several of the other state fields (e.g., $T$ and QVAPOR) are only changed through cross correlations and the evolution of the dynamics during the model advance. For $T$ and QVAPOR, most of the RMSE change occurs during the model integration. Among the other microphysical fields, QGRAUP has the largest RMSE, likely due to larger overall mixing ratios compared to QRAIN and QSNOW. In general, QGRAUP has a larger degree of variability in these experiments, which is evident in the slightly larger RMSE values.

c. Larger horizontal localization experiments

LOC12V6 and LOC18V6 both produce analyses with less error than LOC6V6 during the entire assimilation period (Fig. 3). During the first 60 min, the LOC12V6 and LOC18V6 state stabilizes to lower RMSE values more quickly than LOC6V6, but the LOC12V6 or LOC18V6 analyses are generally of similar quality during this period. Between 60 and 90 min, an increase in RMSE is also evident in the LOC12V6 and LOC18V6 analyses. Yet during this period, the RMSE of the LOC12V6 analyses increases more rapidly, thus LOC18V6 produces more accurate analyses. For most of the state fields, this result persists throughout the remainder of the assimilation period, although for some (e.g., QCLOUD, $T$, QRAIN), LOC12V6 and LOC18V6 produce comparable results near the end of the period. These results suggest that the underlying convective mode contributes to the ability of the EnKF system to produce accurate analyses. Specifically, in these experiments, a larger horizontal localization appears to be especially
beneficial during the first period of cell mergers. This is also true for the vertical localization, as will be discussed in a following section.

The differences in produce differences in posterior ensemble spread in the three experiments. As the localization cutoff is increased, state points are impacted by more observations, decreasing the posterior ensemble spread. The average prior state-space consistency ratio was computed to gauge the suitability of the spread within each experiment, similar to previous work (e.g., Snyder and Zhang 2003). This consistency ratio will be \( N/(N+1) \), where \( N \) is the ensemble size, if the ensemble variance is equivalent to the error of the ensemble mean in these experiments (Murphy 1988). This optimal ratio is approximately 0.98 for the 50-member ensemble used herein. The consistency ratio within the storm (at \( Z > 20 \text{ dBZ} \)) increases during the first 50-min for each experiment and state field (Fig. 4) as the RMSE decreases. Afterward, the consistency ratio gradually decreases or remains steady through the end of the assimilation period as the RMSE remains steady or increases.

As expected, the state fields altered by additive noise before the model advance (\( U, V, T, \text{QVAPOR} \)) have larger consistency ratios than the other state fields. For example, the \( U, V, \) and \( T \) consistency ratios are greater than the optimal ratio for the first half of the assimilation period in all three experiments, with values closer to the optimal ratio during the second half. The spread is likely too large for \( U, V, \) and \( T \) between 20 and 80 min, although this may be desirable to give more weight to the observations as the covariances develop in the ensemble. It is clear that efforts to implement adaptive spread adjustment techniques (e.g., Anderson 2009) in convective-scale EnKF studies will be necessary in the future.

d. Prior analysis vertical profiles

Vertical profiles of time-averaged RMSE and bias [Eqs. (1) and (2)] are shown in Fig. 5. The biggest differences between the three experiments are found in the velocity fields; \( U, V, \) and \( W \) LOC18V6 RMSE are approximately 1 m s\(^{-1}\) smaller than LOC6V6 RMSE between 2 and 12 km. The differences are similar between
the experiments for $U$, but for $V$ the differences increase with height. For all three experiments, the largest $U$, $V$, and $W$ errors occur between 3 and 7 km, coincident with the strongest vertical velocities (not shown). Only small differences exist between the three $T$ RMSE profiles. LOC18V6 has consistently smaller RMSE for QRAIN, QSNOW, and QGRAUP throughout the profile. The behavior of the biases between the three experiments is not consistent among the state fields. The existence of some of these prior analysis biases, particularly the large
negative bias that exists in the surface cold pool for \( T \) and QVAPOR will be discussed further in section 4.

e. Sensitivity to vertical localization

The LOC12V6 vertical localization cutoff length \( r_y \) was changed in two additional experiments to assess the analysis sensitivity to changes in only the vertical localization. In these experiments, \( r_y \) was changed to 3 and 12 km, while the horizontal localization remained constant at 12 km (LOC12V3 and LOC12V12). LOC12V6 from the previous section serves as a control experiment in this set of experiments.

During the first 60 min of assimilation, smaller values of \( r_y \) have either no appreciable impact or are slightly detrimental to the analyses (Fig. 6). The larger vertical localization appears to improve the spinup time for some of the fields during the first 40 min (e.g., \( W \), QCLOUD). By 40 min, the differences in RMSE between the three experiments are minimal. Between 60 and 100 min, a noticeable difference in the RMSE between the three experiments emerges, with LOC12V3 having the smallest RMSE for most state fields (Fig. 6). As described earlier, an initial set of cell mergers occurs during this time period. The smaller vertical localization length scale in LOC12V3 appears to improve the state and covariance estimate during the time of cell mergers and interactions, and this benefit continues through the end of the assimilation period in most of the state fields.

In LOC12V3, many of the fields have assimilation increments that improve the state estimate. For example, assimilation decreases both the \( T \) and \( V \) analysis RMSE during the final six assimilation cycles (mean increments of approximately \(-0.1 \) K and \(-0.2 \) m s\(^{-1}\), respectively). In LOC12V6 and LOC12V12, the \( T \) and \( V \) analysis RMSE increases or remains unchanged after assimilation during these six cycles (mean increments of approximately \(+0.2 \) K and \(+0.3 \) m s\(^{-1}\), respectively). This is evident in other fields (e.g., QCLOUD, QVAPOR, QRAIN) as well, particularly during the second half of the assimilation period. Overall, the LOC12V3 experiment produces the smallest mean analysis RMSE.

4. Line-averaged, time-averaged increments

a. \( W \) and \( T \) increment patterns

Line-averaged, time-averaged (90–150 min) PO-PR and PR-TR updraft increments can elucidate systematic adjustments to the prior mean model state that are occurring within the convective system through the
Figure 7. Line-time-averaged (a)–(c) prior-truth and (d)–(f) posterior-prior vertical velocity increments (shaded) from LOC6V6, LOC12V6, and LOC18V6. The line averages are taken between grid point 20 and 110 in the y direction and the time averages include analyses between 90 and 150 min. Also shown is the region where the combined cloud water and ice is greater than $10^{-4} \text{g kg}^{-1}$ (gray contour). The average position of the leading edge of the surface cold pool is also shown (gray dashed line).

assimilation period. Line averages were taken in the y direction from the northernmost to southernmost point along the convective line (from grid points 20 to 110).

A negative–positive increment pair straddles the prior mean updraft W location in all of the experiments (Fig. 7). These adjustments indicate the location of the maximum updraft is, on average, displaced ahead of the truth updraft maximum. This is also evident in the PO-PR increments in the temperature field (Fig. 8). The negative updraft increments are collocated with negative temperature increments. At the surface below the location of the PO-PR updraft increments, positive PO-PR temperature increments exist. The assimilation, through the covariances between the radar observations and temperature field, is attempting to adjust the location of the cold pool by warming the leading edge by 0.5–1 K. These W and T PO-PR increments indicate the surface cold pool location is displaced eastward of the cold pool placement in the truth simulation, resulting in a low and midlevel updraft that is also biased eastward.

Compared to the truth simulation, not only is the assimilated convective system displaced eastward compared to the truth run, but the prior mean state surface cold pool is 1–3 K colder in all assimilation experiments. If the simulated cold pool propagates as a density current, then a colder cold pool will produce a simulated convective system with a faster forward speed (Wakimoto...
1982) and will lead to the convective system in the assimilation experiments being slightly ahead of the system in the truth simulation. This relationship suggests that inaccuracies in cold pool temperature are producing the displacement error in the assimilated convective systems. The assimilation acts to correct this displacement error by adjusting the placement of the updraft and cold pool edge. Unfortunately, because of the lack of observations below 1 km (≈10% of the total radar observations are located below 1 km), any adjustments to the cold pool must be accomplished exclusively through covariances between observations above the cold pool and state locations within the cold pool. Results suggest that the assimilation is unable to adequately adjust these fields, as the cold pool temperature and position errors remain throughout the assimilation period. Although, as will be discussed in the following section, changes in localization lessen the impact of this bias. While more observations below 1 km would improve the perfect-model results herein, in real cases with parameterized boundary layers, observations may still be unable to produce accurate analyses due to model error.

In addition to the errors in the midlevel updraft region, positive temperature errors exist in a band near the model tropopause (~12 km) and negative temperature errors exist above the tropopause in the overshooting top. The assimilation attempts to correct the latter errors by adjusting the temperature upward in the mean posterior state (i.e., positive PO-PR T increments exist). Data assimilation is unable to correct the former temperature errors as evidenced by negligible PO-PR T increments. In both locations, observations are extremely sparse (e.g., <40 radar observations exist between above 12 km over the entire domain at each assimilation time), so the
assimilation increments in the overshooting top region are likely dominated by cross correlations that extend between observations in the updraft to the temperature field above the updraft. These correlations are unable to correct the errors near the tropopause, thus, the errors persist and grow through the assimilation period. Errors in the horizontal wind field also exist at this level (not shown). The wind errors support the notion that these errors are primarily due to observational limitations, since radial velocity observations, if sufficiently dense, should be able to correct the wind errors in the upper-level outflow.

b. Differences between localization experiments

The patterns noted above are present in varying degrees in all of the assimilation experiments (Figs. 7–8). In general, the spatial extent and magnitude of the PO-PR and PR-TR increments are the smallest in LOC18V6. For the T field, the PR-TR increments within the surface cold pool are smallest in LOC18V6. This is likely a contributing factor to the smaller W increments in the same experiment. In all three experiments, covariances develop between the assimilated radar observations and the state temperature field within the cold pool. The cold pool is a region that is inadequately sampled, so it is promising that the covariances are partly able to make appropriate corrections to the temperature errors near the surface, especially along the leading edge of the cold pool. The PO-PR increments are somewhat larger near the surface in LOC6V6, yet the magnitude of the positive increments in this region is similar between the experiments (since the vertical localization is not changing between these three experiments, the regions of the storm that rely on vertical covariances for corrections should be similar). The T errors near the tropopause are smallest in LOC18V6, although the PO-PR increments along the tropopause are similar between the three experiments.

c. Cold pool development sensitivity to additive noise

The convective system speed bias that develops in all of the experiments is a result of the prior state cold pool being colder than the truth simulation. In real-data experiments, cold pool biases often are attributed to errors in the model microphysics scheme (e.g., Dowell et al. 2011). In a perfect-model OSSE, the same microphysics scheme is used in the truth and assimilation experiment, thus, it is hypothesized that the errors are due to some component of the data assimilation process (e.g., spurious covariances between reflectivity and surface temperature) or additive noise. Dowell et al. (2011) demonstrated substantial sensitivity to the surface cold pool when the magnitude of additive noise was varied (cf. their Figs. 9b,c). Experiments were conducted to test the sensitivity of the cold pool speed bias to the additive noise parameters. Smaller additive noise settings marginally improved the cold pool temperature bias, but large errors remained at several points along the surface gust front (not shown). A more systematic examination of the effects of additive noise on the present results is left to future work.

5. Observation-state correlations

The results in the previous sections can be further interpreted by examining the covariance structures present within the OSSEs. Several previous studies have analyzed the covariance structures within convective storms using various techniques (e.g., Michel et al. 2011), primarily to develop covariance models for 3D-VAR applications. Herein, we apply a brute-force method by manually computing correlations and analyzing the spatial scales over which substantial correlations exist. Correlations are computed between the assimilated observations (radial velocity and reflectivity within the storm) and prior state fields (at the state locations where reflectivity is greater than 20 dBZ). For the present analyses, correlations are computed between radial velocity and the U, V, and W state fields, as well as reflectivity and the QRAIN, QGRAUP, and QSNOW state fields. These observation-state pairs were chosen because of their direct relationship to each other, although meaningful correlations exist between other observation-state pairs (e.g., reflectivity and temperature). The prior state ensemble correlations drive the data assimilation update, and should provide insight into the characteristic spatial scales of the correlation field between the radar observations and state variables.

Observation-state distances and correlations are used to compute reverse cumulative density functions (RCDFs) that show the number of pairs greater than a given distance. These give an aggregate sense of the percentage of observations that fall at various distances away from the state points (or, alternatively, the percentage of state points that fall at a given distance away from the observation location). RCDFs were produced for the observation-state pairs with correlations greater than 0.4, 0.6, and 0.8. These thresholds were chosen to represent the length scales for mid- and high-end correlations that are not significantly affected by sampling noise.

a. 3D correlation structure

Figures 9a–c shows the RCDFs from the LOC12V12 60-min prior analysis for all six observation-state pairs. Correlations greater than 0.4 exist for all observation-state pairs at distances up to 20 km (Fig. 9a), while those greater than 0.6 and 0.8 are generally restricted closer to
the observation location (Figs. 9b,c). In the latter case, the number decreases quickly as the observation-state distance increases. For the microphysical fields, roughly half of the observation-state pairs with correlations greater than 0.4 are within \(6.5\) km (Fig. 9a). For the dynamic fields, this occurs at distances between 8 and 10 km. Thus, the radial velocity correlations with the dynamic fields \((U, V, W)\) extend to farther distances than the reflectivity correlations with the microphysical fields. This is also true for the larger observation-state correlations, although the RV-U correlations tend to behave similarly to the microphysical observation-state pairs.

Some insight into the performance of the various localization experiments can potentially be gleaned from these results. In LOC6V6, observations are not permitted to update the state at distances beyond 6 km. Yet, only 20%–25% (40%) of the state points that are correlated at \(>0.4\) with the radial velocity (reflectivity) observations lie within this radius (Fig. 9a; these results are consistent across LOC6V6 and LOC12V6). This may have proven to be detrimental to the LOC6V6 analyses, especially in data-sparse regions. Increasing \(r\) to 12 km, as in LOC12V6, captures 70%–75% of the state points that are correlated at \(>0.4\) with the radial velocity observations (while capturing all of the state points with correlations \(>0.8\)). For reflectivity observations, roughly 90% (100%) of the state points with correlations greater than 0.4 (0.8) are within 12 km. While the localization choices in LOC18V6 permit the inclusion of an even greater fraction (nearly 100%) of the observation-state pairs, a larger fraction of noisier covariances less than 0.4 are also included (not shown). Thus, the improvement of the LOC12V6 analysis RMSE over LOC6V6 and LOC18V6 may be due to a closer match between the localization choice and the spatial scales present within the observation-state correlation field.

By the end of the assimilation period (150 min) the length scales of most observation-state pairs have increased compared to the correlations present at 90 min across all three experiments (not shown). This is hypothesized to be the result of upscale growth during the second half of the assimilation period. This will be examined further in section 5c.

b. Horizontal versus vertical correlation structure

To isolate the correlation spatial scales between the horizontal and vertical, the RCDFs in Figs. 9a–c were reconstructed based on the observation-state pair horizontal and vertical distance. The horizontal scales have similar structures to those shown in Figs. 9a–c; thus, they are not shown. The vertical correlation scales are much
smaller, especially at correlation thresholds of 0.6 and 0.8 (Figs. 9e,f). The length scales are similar for five of the six observation-state pairs, where approximately 90% (98%) of the pairs have vertical separation distances of less than 2 km for correlations greater than 0.6 (0.8). For the radial velocity-W pair, the length scales are longer at all correlation thresholds. Only 50% (70%) of these pairs have vertical separation distances less than 2 km for correlations greater than 0.6 (0.8). These findings suggest why LOC12V3 had the smallest RMSE between the three vertical localization experiments.

A 3-km vertical cutoff contains a large majority of the observation-state pairs with correlations greater than 0.6, while reducing the number of observation-state pairs that are dominated by sampling noise.

c. Comparison with computed correlation length scales

Correlation length scales of the forecast errors from the OSSEs were computed using a technique described by Belo Pereira and Berre (2006, hereafter PB06). While the RCDFs provide some insight into the spatial structure of the observation-state correlations, the PB06 technique computes a local horizontal correlation length scale for each model grid point for each state field. A domain median length scale was computed for all model grid points where reflectivity was greater than 20 dBZ for each assimilation period beginning with the 30-min prior ensemble analyses. The PB06 technique tends to produce large length scales along the edges of the simulated convection. These length scales were judged to be an artifact of the large horizontal derivatives of the state fields in these areas. The domain median, instead of average, length scale was computed to reduce the overall influence of the length scales at these grid points.

Some agreement exists between the RCDF and PB06 length scales. First, the length scales associated with the microphysical fields from both methods are generally smaller than the dynamic fields (Fig. 10a) over the entire assimilation period. In addition, the changes in length scale during the assimilation period as the convection grows upscale (discussed in section 5a) are also observed using the PB06 length scales. For instance, between 30 and 90 min, the PB06 length scales for $U$, $V$, and $T$ increase by 300–500 m. During this period, a more rapid increase occurs with QGRAUP and QSNOW between 30 and 60 min, followed by a period of little change for 60 min.

Differences are also evident between the two techniques. For example, the RCDFs produced using correlations larger than 0.4 indicate that the RV-W correlation length scale is larger than the RV-U and RV-V (Fig. 9a), while the $W$ field PB06 length scale is smaller than the $U$ and $V$ length scales. In addition, separation is noted among the PB06 microphysical length scales while the RCDFs show no evidence of any difference between the three fields. Specifically, the QRAIN PB06 length scale is the smallest of the seven fields for almost the entire assimilation period. This length scale evolves differently than the other six fields; instead of increasing or remaining steady during the 30–90-min period, it gradually decreases during the entire 120-min period. During the final 30 min of the assimilation period, the PB06 length scales begin to decrease for all state fields except for.
QRAIN. There is no indication of this behavior using the RCDFs.

The median vertical profiles of PB06 length scale were created for the same seven state fields for the final assimilation period (Fig. 10b). The $U$ and $V$ length scales are maximized near the surface and aloft (above 8 km) with a minimum near 3 km. The $W$ length scale is smallest near the surface and is largest above 8 km. The QGRAUP and QSNOW length scales gradually increase from the midlevels (they are not defined below this level) to the top of the storm. The length scale is smaller in the midlevels than $U$, $V$, and $W$, but larger than these three fields above 11 km. The QRAIN length scale is fairly constant below 5 km and is 1–2 km smaller than the $U$ and $V$ length scales near the surface.

The PB06 length scales appear to have some substantial variability within the convective storms simulated in these experiments. For example, the $W$ length scale at 6.3 km from the final prior analysis is plotted in Fig. 11 (only points where the reflectivity is greater than 20 dBZ are shown). The length scale varies in the across-line direction by as much as 5 km, with larger length scales toward the front of the line and smaller length scales toward the rear. The PB06 length scales are more consistent in the along-line direction. While potentially revealing, further analysis of the spatial variability of the forecast error length scales is left to future work.

6. Ensemble forecasts

As another measure of the proper localization cutoff length, 60-min ensemble forecasts are initialized from the final 50 model states at the end of the 150-min assimilation interval. While analyses with smaller initial error should produce better forecasts, imbalance in the analysis also impacts the development of forecast error. Imbalance can be introduced because of suboptimal localization cutoffs, which could potentially hamper the ability to initialize forecasts with these analyses (Greybush et al. 2011). The 60-min ensemble forecasts were produced for LOC6V6, LOC12V6, LOC12V3, and LOC12V12. Ensemble mean state-space RMSE was used as a measure of the skill of the forecasts; the truth simulation was run for an addition 60 min to use as verification for the ensemble mean forecasts.

For most of the experiments and state fields, rapid error growth occurs during the first 15–30 min (Fig. 12). For some fields, RMSE continues to increase through the end of the 60-min forecast period (e.g., $U$, QRAIN) while for others the RMSE is steady (e.g., QVAPOR), or even decreases in the 30–60-min period (e.g., $W$, QCLOUD). The model states become progressively more similar among the experiments during the forecast period. For example, the range of RMSE values from the ensemble mean $U$ analysis at the initialization time is roughly 2 m s$^{-1}$, while the error differences in the 60-min forecast of $U$ are less than 1 m s$^{-1}$.

For most of the forecast period, the RMSE evolves similarly in time between the experiments, with the forecast error a function of the initial analysis error (i.e., the experiments with smaller 60-min forecast RMSE have smaller initial RMSE). For example, LOC12V3 generally produces the most accurate analyses, and the 60-min forecasts benefit from this more accurate initial state.

While the sensitivity of the forecasts to localization is unique to this study, previous studies (e.g., Aksoy et al. 2010; Dawson et al. 2012) that examined convective-scale ensemble forecasts using the EnKF have found similar results in terms of error growth, although their statistics were computed in observation space. A significant difference in this work is the apparent decrease in error evident in several state fields as described above. This may be due to the simplistic view of convective-scale forecast quality used herein; other more advanced verification metrics (e.g., Schwartz et al. 2010) could provide a different view of these findings.
7. Summary and discussion

This work describes the results of several OSSEs to assess the impact of covariance localization for radar data on EnKF analyses of a developing convective system. Sensitivity to both horizontal and vertical localization was assessed via RMSE statistics, assimilation increments, and prior-truth increment structures. To interpret the results further, the prior correlations between the assimilated radar observations and state variables within the convective storm were computed and analyzed both in aggregate (RCDFs) and on a local scale (PB06 length scales). Further, ensemble forecasts were produced from the final EnKF analyses to provide another measure of the impact of different localization options on the analyses.

The results herein suggest that when assimilating radar data with the temporal and spatial resolution typically provided by a WSR-88D, a larger horizontal (~12–18 km), but smaller vertical (~3 km) localization cutoff (using a typical Gaspari–Cohn localization function) than previous studies could be beneficial, especially in cases where the convective evolution is dominated by thunderstorms interactions, mergers, and/or upscale growth into a linear convective system, as was the case in the present simulations. These findings run counter to the localization choices used in all previous convective-scale EnKF literature, especially the vertical localization cutoff, where the vertical length scales of the covariance structures for radar observations were seemingly limited. When considering the choice of localization, factors such as ensemble size and inflation should be considered. For example, a larger localization length scale would likely be beneficial when employing a larger ensemble (HM98).

Some discrepancies between the previous work and this present work exist that are worth mentioning. For example, Tong and Xue (2005) noted that $r = 6$ km produced the best results after decreasing the ensemble size from 100 to 40 members, which is similar to the

![Figure 12: RMSE for ensemble mean forecasts initialized with the final posterior ensemble from LOC6V6, LOC12V12, LOC12V3, and LOC12V6 for 10 state fields. The last 12 analysis cycles are shown (unshaded), in addition to the 60-min ensemble forecasts (shaded).](image-url)
ensemble size used herein. One difference between Tong and Xue (2005) and this work is the model grid and observation density (2 vs 3 km). Another difference is the mode of the simulated convective system. Previous convective-scale EnKF studies employing both OSSEs and real-data experiments, primarily have focused on producing analyses of an individual thunderstorm cell (e.g., tornadic supercells). While these studies have laid the foundation for convective-scale EnKF and proved its viability, the scenarios depicted in these studies are only a subset of the large variety of convective modes observed in nature. Thus, more work is needed to demonstrate how those results generalize to other types of convective systems and forms of severe weather.

The results of the current experiments were obtained with the assumption of a perfect model. In reality, model error (e.g., due to microphysical parameterizations) plays a significant role in producing accurate analyses and will likely modulate the exact results shown herein. Specifically, model errors can produce erroneous covariance structures, which could change the results reported herein. Future work will emphasize the impacts of localization on both imperfect-model experiments and real-data experiments to gain more understanding how the present results are changed by the inclusion of model error.

The results herein reiterate the difficulties with analyzing the surface cold pool, even in perfect-model OSSEs. Producing accurate analyses of the cold pool structure near the surface is essential in order to produce accurate forecasts of the speed and structure of cold-pool-driven convective storms whose behavior and severity are largely controlled by the thermodynamic properties of the surface cold pool, even more so than other convective modes (e.g., supercells). Yet, a paucity of observations near the surface, sensitivities to assimilation parameters such as additive noise, and microphysical scheme errors (in real-data experiments) will make this a challenging task.

Herein, covariance localization was achieved by applying an isotropic weighting function to all ensemble covariances, based on the observation distance from the state location being updated. While this type of localization is easy to implement and has proven to be useful, the present results indicate a need for more flexible, adaptive localization options that vary based on location, state variable type, time, and dynamical considerations. For example, during a time when cell mergers and interactions were occurring the OSSE using the largest localization radius produced the most accurate analyses. In addition, the covariance length scales increase as the overall system grows upscale and the covariances exhibit substantial spatial variability. Furthermore, in the present experiments, radial velocity observations are related to the dynamic state variables of $U$, $V$, and $W$ on slightly larger length scales than reflectivity observations are related to the microphysical variables. This result is consistent with several studies that have developed heterogeneous background error covariance models for 3D-VAR applications (e.g., Michel et al. 2011). While the present results may provide guidance for Gaspari–Cohn localization choices, they also underscore the need for techniques that determine the meaningful prior ensemble covariances dynamically during the data assimilation step (e.g., Bishop and Hodyss 2009a,b; Chen and Oliver 2009).

Acknowledgments. Much of this work was developed during the first author’s participation in the Graduate Student Visitor Program at the National Center for Atmospheric Research. Discussions with Chris Snyder, Jeff Anderson, Glen Romine, and David Dowell were extremely beneficial and greatly improved the work. Nancy Collins and Tim Hoar offered valuable technical assistance with the DART software package. Stimulating conversations with Louis Wicker, Ted Mansell, and Mike Coniglio were particularly helpful. Also, the comments of three anonymous reviewers greatly improved the manuscript. Funding was provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA–University of Oklahoma Cooperative Agreement NA11OAR4320072, U.S. Department of Commerce.

REFERENCES


