On the Impact of Additive Noise in Storm-Scale EnKF Experiments

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

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

(Manuscript received 6 October 2014, in final form 17 April 2015)

ABSTRACT

Storm-scale ensemble Kalman filter (EnKF) studies routinely use methods to accelerate the spinup of convective structures when assimilating convective-scale radar observations. This typically involves adding coherent perturbations into analyses at regular intervals in regions where radar observations indicate convection is ongoing. Significant uncertainty remains as to the most effective use of these perturbations, including appropriate perturbation magnitudes, spatial scales, fields, and smoothing kernels, as well as flexible strategies that can be applied across a spectrum of convective events with negligible a priori tuning. Here, several idealized experiments were performed to elucidate the impact and sensitivity of adding coherent perturbations into storm-scale analyses of convection. Through the use of toy experiments, it is demonstrated that various factors exhibit substantial influence on the postsmoothed perturbation magnitudes, making tuning challenging. Several OSSEs were performed to document the impact of these perturbations on the analyses, particularly thermodynamic analyses within convection. The repeated addition of coherent perturbations produced temperature and moisture biases that are most pronounced in analyses of the surface cold pool and aloft near the tropopause, and eventually lead to biases in the dynamic fields. In an attempt to reduce these biases and make the noise procedure more adaptive, reflectivity innovations were used to restrict the addition of noise to areas where these innovations are large. This produced analyses with reduced thermodynamic biases and RMSE values comparable to the best-performing experiment where the noise magnitudes were manually adjusted. The impact of these findings on previous and future convective-scale EnKF analyses and forecasts are discussed.

1. Introduction

The ensemble Kalman filter (EnKF) has become a popular data assimilation approach to generate analyses of convective storms (Snyder and Zhang 2003; Zhang et al. 2004; Dowell et al. 2004; Tong and Xue 2005; Aksoy et al. 2009, and many others). In the EnKF, observations are assimilated using error statistics derived from an ensemble of forecasts (Evensen 1994; Houtekamer and Mitchell 1998), producing an ensemble of analyses. The analysis ensemble is then either advanced to the next assimilation time (as in a cycled EnKF system) or is integrated into the future to produce an ensemble forecast. In a storm-scale EnKF system, analyses rely on high-resolution observations from radar networks (e.g., radial velocity and reflectivity from the WSR-88D network) to represent convective structures.

For optimal assimilation, the forecast ensemble in an EnKF system must account for all potential sources of errors (e.g., model error, observation error, forward operator error, sampling error, etc.). Since typical convective-scale errors in fields such as temperature, velocity, and moisture are not well known, producing an ensemble of convective storms that possesses an appropriate amount of spread and useful covariances is not straightforward (Snyder and Zhang 2003). Thus, ad hoc techniques have been developed to add convective-scale perturbations into the ensemble to accelerate the spinup

* Current affiliation: National Center for Atmospheric Research, Boulder, Colorado.

Corresponding author address: Dr. Ryan A. Sobash, NCAR/MMM, P.O. Box 3000, Boulder, CO 80307.
E-mail: sobash@ucar.edu

DOI: 10.1175/MWR-D-14-00323.1

© 2015 American Meteorological Society
of convection and maintain spread during the assimilation period. For example, in observing system simulation experiments (OSSEs), Snyder and Zhang (2003) and Tong and Xue (2005) added random Gaussian noise to their initial ensemble, while Zhang et al. (2004) added random Gaussian noise to create an ensemble, but added a warm bubble to initiate convection in the ensemble mean, which accelerated spinup of the storm. Caya et al. (2005) and Dowell et al. (2004) added smoothed perturbations in localized areas centered in the region of active convection for an OSSE and observed supercell, respectively.

Dowell and Wicker (2009) and DW09 produced analyses of an observed isolated tornadic supercell using the Caya et al. (2005) automated perturbation technique. While the studies referred to in the previous paragraph used perturbations to construct an initial ensemble, DW09 continued to add perturbations after each assimilation cycle to further accelerate the spinup of convection and to maintain ensemble spread. Limited ensemble spread is often encountered in convective-scale forecasts because of unrepresented error sources (Romine et al. 2014), which can cause the filter divergence as the ensemble drifts away from the observations. Other spread maintenance techniques, such as covariance inflation (e.g., Anderson 2009), have been used with some success in convective-scale EnKF experiments. DW09 performed several experiments using various initialization and inflation approaches and found additive noise to produce the best results.

Although it has been used with success in many storm-scale EnKF studies, the implementation of the additive noise approach in DW09 possesses some undesirable properties. First, it is sensitive to the values of several tunable parameters, including the magnitude of the noise that generates the perturbations, the length scales over which the perturbations are smoothed, and the fields that are perturbed. Running extensive tuning experiments by varying these parameters is expensive and will not be possible in future real-time storm-scale data assimilation systems (e.g., Stensrud et al. 2009). Second, the settings are often held fixed for the entire assimilation period, or decrease to smaller magnitudes at a time during the cycling period when it is assumed convection will be well initialized in the analyses. The latter approach was demonstrated to be beneficial in DW09, by having large magnitudes of noise during the spinup of convection and smaller values of noise to maintain spread after convection was initialized. This has worked well to produce EnKF analyses of isolated convective storms, but is problematic in more complex events where multiple areas of convection exist that may initiate at different times within the assimilation period. Third, the smoothing technique in DW09 produces smoothed perturbations with magnitudes that are larger than the prescribed standard deviations, with the final post-smoothing magnitudes being strongly dependent on model grid spacing and the smoothing length scales. Thus, it is difficult to precisely prescribe a desired amount of uncertainty in the perturbed ensemble fields.

Finally, there is evidence in previous work that adding random, smoothed, perturbations to model fields may have undesirable consequences on the resulting thermodynamic analyses in convective storms. For example, in the analyses of an isolated, tornadic, supercell produced in Dowell et al. (2011), larger (smaller) magnitudes of noise produced a larger (smaller) surface cold pool with larger (smaller) surface temperature deficits (cf. their Figs. 9b,c). Dowell et al. (2011) attributed this largely to model error, namely an underprediction of reflectivity by the microphysics scheme in the downshear region of the storm at low levels. This bias leads to positive reflectivity increments, and negative surface temperature increments, due to negative correlations between reflectivity and surface temperature. Their experiment with larger (smaller) spread due to differences in additive noise magnitudes produced larger (smaller) temperature increments at the surface, leading to a larger (smaller) cold pool. In another study that generated EnKF analyses of an observed tornadic supercell, Marquis et al. (2014) reduced the magnitudes of additive noise perturbations well below the range used in Dowell et al. (2011) and DW09 because of the presence of unrealistic warm and cold anomalies near the surface in the EnKF analyses when using larger noise magnitudes. This behavior may have been related to the previously mentioned sensitivity to model grid spacing.

These sensitivities were also observed in the OSSEs analyzed in Sobash and Stensrud (2013), where model error was not present. While not the primary focus of that study, the magnitude of the maximum cold pool temperature deficit in the analyses of a developing convective system was sensitive to the magnitude of additive noise in the same way as in Dowell et al. (2011). That is, large negative temperature errors were present near the surface compared to the truth simulation. Additionally, large positive temperature errors were present near the model tropopause that appeared to be equally sensitive to the magnitude of additive noise. Whether the predominant source of surface temperature biases in Dowell et al. (2011) is due to components of the data assimilation system (e.g., additive noise) or model biases is difficult to determine because of the lack of extensive thermodynamic observations available for validation.

Here, we investigate the role that smoothed noise perturbations play in the development of errors in EnKF analyses of convective storms, particularly errors associated
with the thermodynamic analyses. These errors are important because of the relationship between properties of the surface cold pool and the behavior and severity of the convective system (e.g., Rotunno et al. 1988; Marquis et al. 2012, 2014). Further, a more flexible method for introducing these perturbations into analyses is presented that reduces the amount of a priori tuning. In section 2, the experiment design is discussed, and in section 3 a toy experiment is conducted to reveal some of the behaviors of the DW09 smoothing technique. Sensitivity of the temperature analyses to additive noise is documented. Sections 4 and 5 describe several baseline and sensitivity OSSEs that document thermodynamic errors. Section 6 describes the impact of only applying perturbations where reflectivity innovations are large. The paper concludes with a summary and discussion of the results and implications for previous and future convective-scale EnKF experiments.

2. Methods

Much of the experimental design, including the truth simulation, generation of synthetic radar observations, and OSSEs, were performed in an identical fashion to those in Sobash and Stensrud (2013); thus, the details of these aspects of the experiments are only summarized in the following sections (further details can be found in sections 2a and 2b of that work).

a. Experiment design

Reflectivity, radial velocity, and clear-air reflectivity (i.e., locations where reflectivity < 10 dBZ) observations were extracted every 5 min from a convection-permitting, idealized WRF simulation of a convective system that is initially composed of five individual cells and grows upscale through cell mergers into a linear convective system. These observations were assimilated into OSSEs using the deterministic EnKF algorithm in the Data Assimilation Research Testbed software (DART; http://www.image.ucar.edu/DARtS/DART). Each perfect-model OSSE used a 50-member ensemble initialized with a set of horizontally homogeneous wind profiles produced by adding random, uncorrelated Gaussian noise to the horizontal wind components, potential temperature, and horizontal and vertical length scales; and $N$ is a three-dimensional field of Gaussian noise; $l_h$ and $l_v$ are model grid indices; and $f$ and $f'$ are the unperturbed and perturbed model fields, respectively. A new field of Gaussian noise is added to each ensemble member at each time the additive noise procedure is performed. Perturbations are added to the two horizontal wind components, potential temperature, and temperature fields. For the water vapor mixing ratio field is first transformed to dewpoint temperature, as in DW09, to avoid perturbing the non-negative water vapor mixing ratio field directly. Once the smoothed perturbations are applied, the model grid fields are transformed back to the water vapor mixing ratio.

b. Additive noise procedure

The implementation of the additive noise procedure introduced in DW09, and used earlier by Caya et al. (2005), is briefly reviewed in this section (refer to DW09, section 2c, for a more thorough discussion). Immediately prior to advancing the model to the next assimilation time, a three-dimensional field of random noise, sampled from a Gaussian distribution with zero mean and specified standard deviation, is added to specified model fields at grid points where the observed reflectivity exceeds a specified threshold, in this case 20 dBZ. Reflectivity observations are mapped to the nearest model grid points to determine the model grid points where noise is added. The field of noise is then smoothed and added to the model field as follows:

$$ f'(l,m,n) = f(l,m,n) + \sum_{i,j,k} N(i,j,k) \exp \left( -\frac{|X_i - X_l|}{l_h} - \frac{|Y_j - Y_m|}{l_h} - \frac{|Z_k - Z_n|}{l_v} \right), $$

where $N$ is a three-dimensional field of Gaussian noise; $l_h$ and $l_v$ are horizontal and vertical length scales; $l, m,$ and $n$ are model grid indices; and $f$ and $f'$ are the unperturbed and perturbed model fields, respectively. A new field of Gaussian noise is added to each ensemble member at each time the additive noise procedure is performed. Perturbations are added to the two horizontal wind components, potential temperature, and dewpoint temperature fields. For dewpoint, the water vapor mixing ratio field is first transformed to dewpoint temperature, as in DW09, to avoid perturbing the non-negative water vapor mixing ratio field directly. Once the smoothed perturbations are applied, the model grid fields are transformed back to the water vapor mixing ratio.

c. Verification metrics and analysis techniques

State-space diagnostics, including ensemble mean root-mean-squared error (RMSE) and consistency ratio (CR), are used to assess the analysis quality by directly comparing the ensemble mean analyses and forecasts to
the truth state. The RMSE of the ensemble mean prior forecast and posterior analyses is computed as

$$\text{RMSE}^{a,f} = \sqrt{\frac{1}{M} \sum_{i=1}^{M} (x_i - xa)^2},$$

(1)

while the CR, which compares the ensemble mean RMSE to the ensemble spread, is defined as the squared ensemble spread (i.e., ensemble variance) divided by the mean squared error [i.e., Eq. (1) squared]:

$$\text{CR}^{a,f} = \frac{1}{M} \sum_{i=1}^{M} \left[ \frac{1}{N-1} \sum_{j=1}^{N} (x_{ij} - xa)^2 \right] \left( \text{RMSE}^{a,f} \right)^2,$$

(2)

where in both equations the subscript $i$ is an index over all $M$ grid points where the true state reflectivity is greater than 20 dBZ; $t$, $a$, and $f$ represent the true, analysis (i.e., posterior), and forecast (i.e., prior) state, respectively; and $N$ is the number of ensemble members. State-space RMSE and CR are computed for the three velocity components, potential temperature, and the WRF single-moment 6-class (WSM6; Hong and Lim 2006) microphysical species (water vapor, cloud water, cloud ice, rain, snow, and graupel).

Prior minus truth (PR-TR) increments provide a sense of the accuracy of the state estimate throughout the model domain. These 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, while the time averages are taken across the 12 prior ensemble mean forecasts during the final hour of the assimilation period. Finally, PR-TR increments of the cold pool strength parameter $C$ at the final assimilation time were used as an integrated measure of the differences in the representation of the surface cold pool between the truth simulation and the OSSEs, using Eqs. (1) and (2) in Weisman (1992).

3. Behavior of DW09 additive noise technique

This section will illustrate the salient features of the DW09 additive noise smoothing procedure in a toy experiment to better interpret the results of the OSSEs. The smoothing kernel described in the previous section, as used in Caya et al. (2005) and DW09, was applied to a three-dimensional array of grid points with random Gaussian noise added at grid points within an ellipse placed at the center of the grid. This grid was assumed to have grid spacing of 3 km in the horizontal and 0.5 km in the vertical, similar to the resolution of the OSSEs. This ellipse was designed to mimic an area of contiguous reflectivity points within a storm that exceed the additive noise reflectivity threshold. This procedure was conducted for various combinations of $\sigma$ (the prescribed standard deviation of the Gaussian noise distribution), smoothing length scales, and ellipse sizes (Fig. 1). For each set of parameters tested, 250 realizations of Gaussian noise were smoothed to obtain reliable statistics on the maximum and minimum perturbations that resulted from the smoothing procedure. The 250 maximum and minimum perturbations for each realization of additive noise were plotted as histograms. The perturbation magnitudes that were produced during the OSSEs routinely fell within the distributions that were produced within this more idealized framework, providing confidence that this approach correctly represents the behavior of the additive noise as implemented in the OSSEs.

a. Impact of $\sigma$ and smoothing length scales

For the baseline case, the standard deviation of the zero-mean Gaussian distribution $\sigma$ from which noise was drawn was set to 1.0 and the horizontal (vertical) length scales were set to 4 km (2 km), typical values in storm-scale assimilation experiments. The ellipse size was constructed to produce approximately 5000 contiguous grid points where Gaussian noise was added (this number is very close to the number of grid points where noise is added during the middle cycles of the OSSEs). For this set of parameters, the mean of the distribution of maximum and minimum perturbations is approximately 10.0, with a spread of approximately 4.0 (Fig. 1c). Thus, $\sigma = 1.0$ leads to smoothed perturbations that range between 7.0 and 12.0. As $\sigma$ is decreased, the distribution of maximum and minimum perturbations shifts toward zero and the spread is reduced (Figs. 1a,b). For example, at $\sigma = 0.5$, the mean is approximately 5 with a spread of 1 (Fig. 1b). At $\sigma = 0.25$, the mean is further reduced to 2 with a spread of less than 1 (Fig. 1a). Thus, for a given smoothing length scale, the mean of the smoothed perturbations approximately doubles when $\sigma$ is doubled, although the spread also increases, with large outliers becoming more likely at larger values of $\sigma$.

Since the smoothing kernel is not normalized, increasing the smoothing length scales increases the weights at surrounding grid points, allowing nearby points to contribute more toward the post-smoothed values (Figs. 1d–f). For example, when $\sigma = 1.0$ and the smoothing length scales are increased to 9 km in the horizontal and 6 km in the vertical, the distribution of perturbation magnitudes becomes much broader with a mean approximately double that (10 versus 20) of the experiment with smaller smoothing
length scales (cf. Figs. 1c,f). This also holds for smaller values of additive noise.

b. Impact of the number of grid points

Another complicating factor is the number of grid points where noise is added. To reproduce this effect, the size of the ellipse was changed to produce a number of grid points that would be equivalent to doubling or halving the horizontal grid spacing. That is, compared to the baseline case, the ellipse was increased to produce 4 times the number of points where noise was added (\( \approx 20000 \) grid points), and the ellipse size was decreased to produce \( \frac{1}{4} \) of the number of points (\( \approx 1000 \) grid points). The increased (decreased) number of grid points has a similar effect to increasing (decreasing) \( \sigma \) or the smoothing parameters (Fig. 2). In the experiment with 4 (\( \frac{1}{4} \)) times the number of grid points where noise was added, the mean of the distribution of perturbation magnitudes shifts toward larger (smaller) values, although the spread of the distribution remains mostly unchanged (Figs. 2a–c). Thus, increasing or decreasing grid spacing also has a notable effect on the post-smoothed magnitudes of the perturbations. This effect is likely even more amplified in many cases since the vertical resolution is usually also increased when increasing the horizontal resolution, exacerbating the issues noted above.

This effect also occurs within a storm-scale assimilation experiment where the number of points is increasing because of the growth of a convective system, especially in cases where convection grows upscale and the number of points exceeding the reflectivity threshold where noise is added is increasing with time. In these cases, the post-smoothed perturbation magnitudes will tend to increase as well. Yet, it is more desirable to have smaller-magnitude perturbations later in the assimilation period after convection has become well established across the ensemble.

---

**Fig. 1.** Distributions of domain-wide maximum and minimum perturbation magnitudes for each of the 250 iterations of the additive noise smoothing procedure applied to a coherent region of points as described in the text for \( \sigma = \) (a),(d) 0.25; (b),(e) 0.5; and (c),(f) 1.0; and (a)–(c) \( l_h/l_v = 4/2 \) km, and (d)–(f) \( l_h/l_v = 9/6 \) km. The perturbation units are nondimensional and each bin is 0.5 nondimensional units wide. The distributions of the 250 maximums and minimums for each set of parameters are plotted together, resulting in two distributions per panel.
This is likely one reason why experiments in which $\sigma$ was forced to decrease at a given analysis time outperform experiments where $\sigma$ is constant.

Combined, these sets of experiments demonstrate some of the inherent difficulties of using the DW09 additive noise smoothing technique to add perturbations to storm-scale analyses. Because of the behavior of the smoothing procedure, one value of $\sigma$ may produce a range of perturbation fields based on the choices for the smoothing length scales and the number of grid points where noise is added, which depends on factors such as convective system size and intensity, model resolution, and distance from the radar platform.

4. Analysis sensitivity to $\sigma$

To illustrate various sensitivities that the convective-scale analyses have to the DW09 additive noise parameters and the behaviors of the smoothing procedure as described in the previous section, three OSSEs were performed with $\sigma = 0.25, 0.5,$ and $1.0$ K and $m$ s$^{-1}$ (EXP_NOISE0.25, EXP_NOISE0.5, and EXP_NOISE1.0, respectively; see Table 1 for experiment summary). The perturbation magnitudes were kept the same for both the thermodynamic and dynamic fields; in experiments where these were varied independently the behavior of the biases discussed later were qualitatively similar. The horizontal and vertical smoothing length scales were set to 9 and 6 km, respectively; thus, the resulting smoothed perturbation magnitudes that are added to the $U$, $V$, $T$, and $T_d$ fields should be drawn from a distribution that looks similar to Figs. 1d–f. An additional OSSE was conducted using $\sigma = 1.0$ K and $m$ s$^{-1}$ during the first 20 min of the assimilation period and $\sigma = 0.25$ K and $m$ s$^{-1}$ after 20 min (EXP_NOISE1.0–0.25; Table 1). This is akin to the strategy employed by DW09 in their experiments.

a. Analysis verification

During the first 40 min of the assimilation period, the errors in most of the state fields are smaller in EXP_NOISE1.0 than EXP_NOISE0.25 (Fig. 3). The errors for EXP_NOISE0.5 fall generally between EXP_NOISE1.0 and EXP_NOISE0.25 during this period (not shown). Here, perturbations with larger magnitudes are beneficial, even though the maximum and minimum perturbation magnitudes can end up being quite large.
(frequently >10 m s\(^{-1}\) and 10 K), and the prior state consistency ratios are well above 1 (Fig. 4). The ability of the radar observations to produce large increments due to the large background spread that is added through the additive noise procedure is likely beneficial since the initial estimate of the background errors in the ensemble is initially very poor.

The benefit of a larger value of \(\sigma\) in EXP_NOISE1.0 disappears after the first 40 min of assimilation, as errors begin to rapidly increase. The forecast errors in EXP_NOISE1.0 grow especially rapidly in \(U, V, T, T_d\), and QVAPOR, which are directly modified by the addition of perturbations (Fig. 3). By the end of the 90-min assimilation period, the prior RMSE for EXP_NOISE0.25 is approximately 1.1 K, while in EXP_NOISE1.0 the prior RMSE is approximately 2.7 K. This general behavior between EXP_NOISE1.0 and the two smaller noise experiments also exists in the other state fields not directly modified by the additive noise procedure (e.g., the five microphysical variables). By the end of the assimilation period, EXP_NOISE0.25 has the smallest RMSE among the three OSSEs.

Decreasing \(\sigma\) from 1.0 K and m s\(^{-1}\) to 0.25 K and m s\(^{-1}\) after 20 min, as in EXP_NOISE1.0-0.25, produces similar error to EXP_NOISE1.0 for the first 30 min. After, as the error in EXP_NOISE1.0 increases, the RMSE EXP_NOISE1.0-0.25 continues to decrease for most state fields (this is especially notable in the \(U\) RMSE). By the end of the period, the RMSEs of EXP_NOISE1.0-0.25 and EXP_NOISE0.25 are comparable.

For all four experiments, the consistency ratio varies substantially through the assimilation period (Fig. 4). In EXP_NOISE1.0, the consistency ratio for \(U, V, W,\) and \(T\) increases well above the optimal value of 1 during the first 40 min then decreases, while QVAPOR increases slightly, but drops below 1 by the end of the period. For the same fields in EXP_NOISE0.25, the consistency ratio gradually increases and stabilizes between 0.5 and 1.0 by the end of the period. The consistency ratios for the five microphysical fields are relatively more stable, yet different experiments are closer to 1 at different times. After a spike near 25 min, the consistency ratio in EXP_NOISE1.0-0.25 decreases and is close to the consistency ratio value for EXP_NOISE0.25 by the end of the period. These results demonstrate the challenges in tuning the additive noise settings given the variability that exists in the consistency ratio between different state fields at different points in the assimilation period. For example, temperature errors grow rapidly by the end of the period (Fig. 3), yet the consistency ratio approaches 1 at this time (Fig. 4). Understanding the nature of the thermodynamic errors should lead to a clearer interpretation of the behavior of these verification metrics.

### b. Thermodynamic analysis errors

The OSSEs performed in Sobash and Stensrud (2013) contained thermodynamic errors that affected the behavior of the simulated convective system. To study these errors in an analogous manner, the prior-minus-truth, line-averaged and time-averaged potential temperature increments were computed as in Sobash and Stensrud (2013). Averages were computed over the last hour of the assimilation period (i.e., 12 assimilation cycles) and across the convective line.

As the magnitude of noise is increased between EXP_NOISE0.25, EXP_NOISE0.5, and EXP_NOISE1.0, the average temperature errors increase substantially (Figs. 5a–c). Also, larger perturbation magnitudes result in a larger, more robust, convective system, as suggested by the outline of the cloud water mixing ratio in Fig. 5. As in Sobash and Stensrud (2013), the temperature errors are concentrated near the tropopause and in the

<table>
<thead>
<tr>
<th>Expt</th>
<th>Noise std dev (K and m s(^{-1}))</th>
<th>Perturbed fields</th>
<th>Reflectivity innovation threshold (dBZ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXP_NOISE0.25</td>
<td>0.25</td>
<td>(U, V, T, T_d)</td>
<td>None</td>
</tr>
<tr>
<td>EXP_NOISE0.5</td>
<td>0.5</td>
<td>(U, V, T, T_d)</td>
<td>None</td>
</tr>
<tr>
<td>EXP_NOISE1.0</td>
<td>1.0</td>
<td>(U, V, T, T_d)</td>
<td>None</td>
</tr>
<tr>
<td>EXP_NOISE1.0-0.25</td>
<td>1.0 ((&lt;=20) min)</td>
<td>(U, V, T, T_d)</td>
<td>None</td>
</tr>
<tr>
<td>EXP_NOISE1.0-0.25</td>
<td>0.25 ((&gt;20) min)</td>
<td>(U, V, T, T_d)</td>
<td>None</td>
</tr>
<tr>
<td>EXP_NO-T</td>
<td>1.0</td>
<td>(U, V, T_d)</td>
<td>None</td>
</tr>
<tr>
<td>EXP_NO-T-Td</td>
<td>1.0</td>
<td>(U, V)</td>
<td>None</td>
</tr>
<tr>
<td>EXP_NOISE0.25_INNOV</td>
<td>0.25</td>
<td>(U, V, T, T_d)</td>
<td>10</td>
</tr>
<tr>
<td>EXP_NOISE0.5_INNOV</td>
<td>0.5</td>
<td>(U, V, T, T_d)</td>
<td>10</td>
</tr>
<tr>
<td>EXP_NOISE1.0_INNOV</td>
<td>1.0</td>
<td>(U, V, T, T_d)</td>
<td>10</td>
</tr>
<tr>
<td>EXP_NOISE1.0_INNOV5</td>
<td>1.0</td>
<td>(U, V, T, T_d)</td>
<td>5</td>
</tr>
<tr>
<td>EXP_NOISE1.0_INNOV20</td>
<td>1.0</td>
<td>(U, V, T, T_d)</td>
<td>20</td>
</tr>
</tbody>
</table>
surface cold pool. At the tropopause, positive temperature errors exist (i.e., the prior ensemble mean forecast is warmer than the truth in these locations), the peak magnitude of which approximately doubles between EXP_NOISE0.25 and EXP_NOISE0.5, and increases even further in EXP_NOISE1.0. At the surface, negative temperature errors exist (i.e., the prior ensemble mean forecast is colder than the truth) with an overall pattern of behavior that is similar to the errors at the tropopause between the three experiments. Further, negative errors are present in the overshooting top region at the top of the updraft above the tropopause that increase in magnitude as $\sigma$ increases.

The spatial structures of the biases described above are also present in EXP_NOISE1.0–0.25 with maxima similar to those in EXP_NOISE1.0 (cf. Figs. 5b,d). Thus, reducing $\sigma$ after 20 min is beneficial when assessing analysis quality through RMSE, yet these biases still exist in EXP_NOISE1.0–0.25 and have magnitudes that fall between the extremes seen in EXP_NOISE1.0 and EXP_NOISE0.25.

In addition to temperature errors, moisture errors also exist within the analyses (Fig. 6). Negative dewpoint temperature errors exist below $\sim$3 km, while positive differences exist above this level. Combined with the temperature biases noted above, the surface thermodynamic analyses tend to be too cold and dry. Aloft, small positive differences in dewpoint are located near the tropopause coincident with the temperature maximum, although the small moisture magnitudes lead to much smaller differences in this layer. Similar to the temperature biases, these errors have similar structures between the OSSEs with magnitudes that are a function of $\sigma$ (not shown).

5. Sensitivity to thermodynamic perturbations

Three additional OSSEs were performed to test the impact of the thermodynamic perturbations on the analyses to further isolate the origin of the thermodynamic errors in the analyses and determine the skill of
experiments without any thermodynamic perturbations (Table 1). In all three experiments, perturbations were added to the $U$ and $V$ fields, as in the OSSEs described in section 4, but perturbations were not added to the temperature (EXP_NO-T) or water vapor (EXP_NO-Td) fields, or to both (EXP_NO-T-Td). For these experiments, the $\sigma$ is the same as in EXP_NOISE1.0. Even though this experiment had the largest error during the second half of the assimilation period, this value of $\sigma$ was chosen to see if these large errors were due primarily to the thermodynamic perturbations.

The benefit of perturbing the thermodynamic fields is evident during the first hour of the assimilation period (Fig. 7). In EXP_NO-T-Td, the RMSE for many of the state fields (e.g., $U$, $V$, QVAPOR) does not begin to decreasing until after 40 min, while the RMSE in EXP_NO-T and EXP_NO-Td has stabilized by 40 min. This is primarily due to issues in the data assimilation, as the update is frequently detrimental to the state in EXP_NO-T-Td (i.e., posterior RMSE greater than prior RMSE), even in state fields such as $U$ and $V$ that are being modified by the additive noise procedure. Thus, the thermodynamic perturbations introduced via additive noise are an important component that accelerates the development of meaningful covariances for both the dynamic and thermodynamic fields, especially in these experiments where the ensemble possesses no initial temperature uncertainty.

Among the other two experiments, EXP_NO-T generally outperforms EXP_NO-Td during the first 60 min, with the greatest differences in the velocity and temperature fields; the microphysical errors between these two experiments are similar. Between 60 min and the end of the assimilation period, the RMSE for all of the state fields in EXP_NO-T-Td decreases and becomes
smaller than the RMSE of the corresponding fields from EXP_NO-T and EXP_NO-T-Td, as the RMSEs in those experiments increase. As in the first 60 min, the RMSE for EXP_NO-Td is greater than EXP_NO-T for most of the state fields, and these differences increase during this period for some fields, most notably for the temperature field. Without the temperature perturbations, the PR-TR increments in EXP_NO-T have much smaller peak magnitudes compared to EXP_NOISE1.0 (cf. Figs. 5c and 8a), while the increments in EXP_NO-Td are similar to EXP_NOISE1.0 (cf. Figs. 5c and 8b). In both of these experiments, the increment structures are similar, with positive values near the tropopause and associated with the updraft, and negative values in the surface cold pool and in the overshooting top region. In EXP_NO-T-Td, positive increments exist near the surface as a result of the delayed development of the surface cold pool, while negative increments exist just below the tropopause (Fig. 8c).

This behavior is also seen in the cold pool strength parameter, C (Fig. 9). The size and magnitude of the cold pool is too large in EXP_NO-Td (Fig. 9a), similar to EXP_NOISE1.0 (not shown). The size is reduced in EXP_NO-T and compares well with the truth simulation (Fig. 9b), although pockets exist along the leading edge of the cold pool where the strength parameter is too large, indicating that the cold pool is too cold or deep in EXP_NO-T. In EXP_NO-T-Td, the cold pool size is smaller than in the truth simulation (Fig. 9c), as indicated by the large area along the westward edge with PR-TR values of C that are positive.

Thus, the behavior of these experiments differs within the assimilation period. Early in the period, the thermodynamic perturbations introduce beneficial spread into the thermodynamic fields since the initial ensemble is devoid of temperature or moisture perturbations. The continued addition of these same perturbations into the analysis later in the period appears to result in a degradation of analysis quality, as the biases described in the previous section grow in magnitude.

6. Additive noise using reflectivity innovations

a. Description of technique

The results from the previous two sections provide evidence that the patterns of behavior of the additive noise smoothing procedure may be contributing to the ineffective assimilation of radar observations and producing analyses biases that contribute to forecast error. First,
given the tendency for larger (smaller) values of noise to be beneficial early (later) in the assimilation period, it is desirable to construct a noise technique that adds larger perturbations at the start of the forecast period, but decreases in magnitude with time as convection becomes well established in the analyses. This will also counteract the effect that was described in section 3, where the post-smoothed perturbation magnitudes tend to increase during the assimilation period, as the convective system grows upscale. Yet, this is challenging to achieve unless one makes an arbitrary choice to reduce $\sigma$ at a given time within the assimilation period, as performed in DW09.

Further, in situations where multiple areas of convection are in various stages of development, it will be necessary to have a technique that produces larger (smaller) perturbations in areas of the domain possessing less (more) mature convection. As is presently implemented, the DW09 technique is incapable of targeting areas of convection to produce a spatially varying field of noise perturbations.

Several previous studies have used observation innovations to guide the placement of noise in ensemble assimilation systems (Dee 1995; Mitchell and Houtekamer 2000). In this case, using reflectivity observation innovations seems promising for guiding the placement of DW09 additive noise perturbations, since the procedure should be able to place larger noise perturbations where the ensemble is far away from the observations, indicating that the background estimate is poor, while reducing the amount of noise added in areas where the background estimate is of higher quality as estimated by the innovation statistics.

As an initial attempt to improve the implementation of DW09, the DW09 procedure was modified to add random noise at grid points only where the reflectivity innovation (i.e., the prior ensemble mean forecast reflectivity minus the reflectivity observations) exceeded a threshold, as well as requiring that the observed reflectivity exceeds 25 dBZ as in the original procedure. Innovation thresholds of 5, 10, and 20 dBZ were tested to assess the sensitivity to this parameter. The same DW09 smoothing procedure used in the previous experiments is then applied to this subset of grid points. Thus, the only modification to the original DW09 technique is the selection of grid points where noise is added, and not the smoothing procedure.

b. Experiments using reflectivity innovations

The same three OSSEs were performed as in section 4, but with the reflectivity innovations used in the selection of grid points (EXP_NOISE0.25_INNOV, EXP_NOISE0.5_INNOV, and EXP_NOISE1.0_INNOV; Table 1). In these three OSSEs, the innovation threshold was set to 10 dBZ. Using the innovations results in a much smaller number of grid points where noise is added compared to EXP_NOISE1.0, especially during the second half of the assimilation period (Fig. 10). For the first two cycles, the number of points is identical between EXP_NOISE1.0 and the experiments using the innovations, as the background ensemble does not possess any hydrometeors,
resulting in all points where reflectivity is greater than 25 dBZ also possessing innovations larger than 10 dBZ. By the third cycle, the set of points begins to diverge. In EXP_NOISE1.0_INNOV, noise is added in ~25% of the number of points compared to EXP_NOISE1.0 at the third analysis cycle; by 45 min into the assimilation period, this percentage drops to less than 2%. The reduction in the number of points between 10 and 45 min is more gradual for EXP_NOISE0.25_INNOV and EXP_NOISE0.5_INNOV since convection spins up more slowly in these experiments.

The number of grid points where noise is added increases slightly and reaches a secondary peak between 45 and 70 min in the experiments using innovations, while in EXP_NOISE1.0 the number grows throughout the period (Fig. 10). This peak coincides with a set of cell mergers, and the RMSE generally increases during this time. This increase in the number of large innovations implies the presence of larger background errors, in this case due to larger forecast uncertainty during the storm merger process.

The differences in the areas where noise is added between the two experiments leads to changes in the evolution of RMSE and ensemble spread. During the first 60 min, the RMSE decreases for all three experiments, with EXP_NOISE1.0_INNOV having the smallest error for all of the state fields (Fig. 11). It is during the first hour that the differences between the experiments are largest. By the end of the period, the RMSE of the three experiments begins to converge for most fields, although EXP_NOISE1.0_INNOV continues to possess slightly smaller RMSEs for the U field. Given the smaller areas of noise in the innovation experiments, the amount of spread added to the analyses is much less, leading to small consistency ratios (<0.5) in EXP_NOISE0.25_INNOV and EXP_NOISE0.5_INNOV for all state fields (Fig. 12). Higher consistency ratios are present in EXP_NOISE1.0_INNOV, with values generally >0.5 for U, V, T, and W, and near
or slightly below 0.5 for QVAPOR and the microphysical fields. These values tend to be more stable through the assimilation period compared to the corresponding non-innovation experiments (cf. Figs. 4 and 12).

During the entire 90-min period, EXP_NOISE1.0_INNOV consistently outperforms the other OSSEs. Compared to the original set of experiments, EXP_NOISE0.25 had the best performance at the end of the assimilation.
period, with a slightly smaller RMSE than EXP_NOISE0.5 and small PR-TR increments, but EXP_NOISE0.25 took longer to spin up than did either EXP_NOISE0.5 and EXP_NOISE1.0. Experiment EXP_NOISE1.0_INNOV possesses the best aspects of EXP_NOISE1.0 and EXP_NOISE0.25, namely a relatively fast spinup period, followed by a period where the RMSEs stabilize, similar to EXP_NOISE1.0–0.25. Compared to EXP_NOISE1.0, the line- and time-averaged PR-TR increment magnitudes are also much smaller in EXP_NOISE1.0_INNOV (Fig. 13a) and are also smaller than in EXP_NOISE1.0–0.25. A maximum still exists near the tropopause and a minimum is found near the surface, but the magnitudes are less than half of the magnitudes in EXP_NOISE1.0 (cf. Figs. 5c and 13a). The PR-TR differences in $C$ are also small, with the prior cold pool size matching fairly closely to the truth cold pool size with small PR-TR magnitudes (Fig. 13b).

Even though the prescribed values of $\sigma$ are the same between EXP_NOISE1.0 and EXP_NOISE1.0_INNOV (1.0 m s$^{-1}$ and 1.0 K), the resultant post-smoothed magnitudes are much less after the first few assimilation cycles in EXP_NOISE1.0_INNOV because of the smaller number of grid points where noise is added. Because of the behavior of the smoothing procedure described in section 3, this reduction of grid points effectively reduces the magnitudes of the smoothed perturbations, allowing the noise procedure to produce temporally variable post-smoothed perturbations through the assimilation period. This is beneficial, as it accelerates the spinup process, but does not possess the deleterious effects of large noise perturbations later in the period. Another contributing factor to the improved performance is that noise is only added in “targeted” locations where the background errors are larger. An experiment akin to EXP_NOISE1.0 where the number of points where noise was added was randomly reduced to the number in EXP_NOISE1.0_INNOV had a slightly larger RMSE for most of the assimilation period (not shown), supporting the value of using the innovations to target areas of larger uncertainty.

c. Sensitivity to innovation threshold

In another set of OSSEs, the innovation threshold was increased to 20 dBZ (EXP_NOISE1.0_INNOV20) and decreased to 5 dBZ (EXP_NOISE1.0_INNOV5). Changing the innovation threshold adjusts the number of points selected in the noise procedure. As a percentage of points used in the standard noise procedure, EXP_NOISE1.0_INNOV20, EXP_NOISE1.0_INNOV, and EXP_NOISE1.0_INNOV5 have 16%, 2%, and <1% of the grid points at the final analysis time, respectively.
Some differences in RMSE exist between the three experiments, mainly during the first 30 and last 30 min of the assimilation period (Fig. 14). For the dynamical fields, EXP_NOISE1.0_INNOV had a larger RMSE than either EXP_NOISE1.0_INNOV5 or EXP_NOISE1.0_INNOV20 during the spinup period, while the RMSEs were fairly similar for the other state fields. After 60 min, the RMSE for EXP_NOISE1.0_INNOV10 was either smaller or similar to the RMSEs for the other two experiments. The differences were largest for $U$ and $V$, with the final posterior analysis RMSEs approximately 1 m s$^{-1}$ and 0.5 m s$^{-1}$ smaller in EXP_NOISE1.0_INNOV, for those two fields respectively. The differences in the cold pool strength parameter at the final time were fairly small between the three experiments, as were the differences in the PR-TR increment cross sections (not shown).

The differences in the number of points where noise is added result in fairly large changes in spread, and thus consistency ratio (Fig. 15). As the innovation threshold increases and the number of points decreases, the consistency ratio decreases. For most of the assimilation period, EXP_NOISE1.0_INNOV20 possesses a consistency ratio below 0.5 for all fields. Experiment EXP_NOISE1.0_INNOV5 has a consistency ratio well above 1 for $U$, $V$, $W$, and $T$, although it is closest to 1 among the three experiments for the microphysical fields. Experiment EXP_NOISE1.0_INNOV falls in between these two extremes, producing ratios closer to 1 than the other two experiments for $U$, $V$, $W$, and $T$. Even though the consistency ratio is closer to 1 for the microphysical fields in EXP_NOISE1.0_INNOV, it appears that having a smaller consistency ratio for these fields, as in EXP_NOISE1.0_INNOV, is not detrimental to the analysis quality, and that having an appropriate value of spread for $U$, $V$, and $T$ fields may be of greater importance. Overall, these results suggest that 10 dBZ is an appropriate innovation threshold to use for the selection of points where additive noise is added in this set of OSSEs.
7. Summary

Additive noise of the type described by DW09 has been used routinely in convective-scale EnKF data assimilation experiments (e.g., Dowell et al. 2011; Dawson et al. 2012) to insert coherent convective-scale perturbations in analyses that are initially devoid of these structures. This accelerates the spinup of convection and provides a source of convective-scale spread. In the present work, the behavior of the DW09 additive noise procedure, and its impact on convective-scale analyses, was investigated through a series of toy experiments and OSSEs of a developing convective system (Sobash and Stensrud 2013). A summary of the key findings is provided below:

1) The DW09 additive noise smoothing procedure produces coherent perturbations with magnitudes that vary depending on the prescribed values of $s$ and the smoothing length scales, as well as the number of model grid points where noise is added. The latter factor is a function of convective system size, and thus mode and model grid spacing, as well as the number and density of the reflectivity observations. The resulting distribution of smoothed perturbation magnitudes that are added to the analyses can vary substantially depending on these parameters.

2) In a set of OSSEs, the experiment with the smallest ensemble mean analysis RMSE (EXP_NOISE1.0–0.25) had perturbations with relatively large magnitudes (1.0 K and m s$^{-1}$) added to the analyses between 0 and 20 min into the assimilation period, followed by smaller-magnitude perturbations (0.25 K and m s$^{-1}$) after 20 min, similar to the findings of DW09. OSSEs with constant $s$ through the period were either slower to spin up (EXP_NOISE0.25) or had rapidly growing error during the second half of the assimilation period (EXP_NOISE1.0).
3) Negative temperature and moisture biases within the surface cold pool and positive temperature errors in the tropopause existed to varying degrees within all the OSSEs. The magnitudes of these errors were strongly controlled by the magnitudes of the perturbations to the temperature and moisture fields. In an OSSE with the temperature perturbations removed, errors were reduced, but removing moisture perturbations had little effect on analysis RMSE and bias. Thus, these biases appear to be the result of the continual addition of thermodynamic perturbations by the additive noise procedure, beyond the period when these perturbations were beneficial to encourage convective development and increase spread in areas where the background analysis is devoid of convection.

4) An OSSE that used a reflectivity innovation threshold (10 dBZ) to select where noise was added produced ensemble mean analyses with equal or smaller RMSEs than the best-performing OSSE where $\sigma$ was tuned manually (i.e., EXP_NOISE1.0–0.25). This was primarily a result of a decrease in the number of model grid points where noise was added during the first hour, in contrast to the standard noise configuration where the number of points continually increased through the period. This produced smoothed perturbations with magnitudes that became smaller as the assimilation period progressed as a result of the behavior of the smoother. During a period of cell mergers, the number of points increased, demonstrating the ability of the technique to temporally adapt to the changing background uncertainty. The analyses in this experiment possessed smaller thermodynamic biases aloft and near the surface, as well as a surface cold pool that more closely resembled the truth cold pool in both magnitude and aerial extent.

8. Discussion

These results provide insight into the behavior of the DW09 smoother when applied to produce coherent convective-scale perturbations, document the impact of the perturbations on EnKF storm-scale analyses, and investigate the benefits of using reflectivity innovations to guide the insertion of coherent additive noise into these analyses. In this case, the innovation approach appears to be beneficial for several reasons. First, the post-smoothed perturbation magnitudes decrease with time as the number of points where noise is added decreases, which is a more flexible strategy for decreasing the perturbation magnitudes than selecting a transition time. Second, the innovations allow for a spatially variable field of noise, putting larger perturbations in areas where the ensemble mean is farther away from the observations. Third, the results appear to be less sensitive to the prescribed values of $\sigma$ than when using the innovations.

Yet, additional research is needed to develop a more unified, adaptive technique for producing storm-scale perturbations in EnKF experiments. Here, the smoothed perturbation magnitudes are still dependent on the number
of grid points where noise is added. Normalizing the smoothing kernel, in combination with a spatially and temporally variable field of noise magnitudes that are determined by reflectivity innovations, may lead to additional benefits; work is ongoing to develop such a technique. Also, further experiments are needed to clarify the physical source of the thermodynamic errors in order to understand how to construct better methods to perturb the convective and in-storm environments.

Using reflectivity innovations to guide the insertion of noise should demonstrate an even greater benefit in more realistic scenarios of convective development, where convection is developing in portions of the domain at different times. Designing an OSSE to test such a scenario and applying the innovation technique in real-data experiments are potential avenues for future research. Other initialization and spread maintenance methods (i.e., running in place, adaptive inflation) should be tested to either replace or be combined with the existing additive noise implementation (e.g., additive noise where innovations are large combined with adaptive inflation).

These results provide alternative interpretations for thermodynamic errors observed in other storm-scale EnKF studies (e.g., Dowell et al. 2011). The temperature and moisture biases produced by the additive noise technique are potentially present to some degree in previous studies that have used DW09 additive noise, although it is difficult to disentangle the contribution of additive noise to errors in real-data assimilation experiments, compared to other error sources, such as model error. While surface temperature analyses are routinely compared to surface observations within convection, biases aloft would be challenging to identify in real-data studies without high quality observations within this region of convection.

The presence of these biases can lead to a variety of issues during the data assimilation and forecast process.
For example, observations in the surface cold pool, such as 2-m temperature, or observations near the tropopause, such as cloud-top temperature, are likely to be ineffectively assimilated if these biases exist. As mentioned previously, the properties of the surface cold pool exert a strong influence on the behavior of the convective system, thus a biased cold pool will lead to errors in storm mode, speed, longevity, etc.

Acknowledgments. Funding for this work was provided by the Warn-on-Forecast project by NOAA/Office of Oceanic and Atmospheric Research under NOAA–University of Oklahoma Cooperative Agreement NA11OAR4320072, U.S. Department of Commerce. Discussions with David Dowell, Glen Romine, Jeff Anderson, Dave Stensrud, Chris Kerr, and Kent Knopfmeier improved various aspects of the work.

REFERENCES


