A Comparison of Multiscale GSI-Based EnKF and 3DVar Data Assimilation Using Radar and Conventional Observations for Midlatitude Convective-Scale Precipitation Forecasts

AARON JOHNSON
School of Meteorology, and Cooperative Institute for Mesoscale Meteorological Studies, and Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma

XUGUANG WANG
School of Meteorology, University of Oklahoma, Norman, Oklahoma

JACOB R. CARLEY
I.M. Systems Group, Inc., and NOAA/NWS/NCEP, Environmental Modeling Center, College Park, Maryland

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

CHRISTOPHER KARSTENS
Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma

(Manuscript received 20 October 2014, in final form 12 February 2015)

ABSTRACT

A GSI-based data assimilation (DA) system, including three-dimensional variational assimilation (3DVar) and ensemble Kalman filter (EnKF), is extended to the multiscale assimilation of both meso- and synoptic-scale observation networks and convective-scale radar reflectivity and velocity observations. EnKF and 3DVar are systematically compared in this multiscale context to better understand the impacts of differences between the DA techniques on the analyses at multiple scales and the subsequent convective-scale precipitation forecasts.

Averaged over 10 diverse cases, 8-h precipitation forecasts initialized using GSI-based EnKF are more skillful than those using GSI-based 3DVar, both with and without storm-scale radar DA. The advantage from radar DA persists for ~5 h using EnKF, but only ~1 h using 3DVar.

A case study of an upscale growing MCS is also examined. The better EnKF-initialized forecast is attributed to more accurate analyses of both the mesoscale environment and the storm-scale features. The mesoscale location and structure of a warm front is more accurately analyzed using EnKF than 3DVar. Furthermore, storms in the EnKF multiscale analysis are maintained during the subsequent forecast period. However, storms in the 3DVar multiscale analysis are not maintained and generate excessive cold pools. Therefore, while the EnKF forecast with radar DA persists better than the forecast without radar DA throughout the forecast period, the 3DVar forecast quality is degraded by radar DA after the first hour. Diagnostics revealed that the inferior analysis at mesoscales and storm scales for the 3DVar is primarily attributed to the lack of flow dependence and cross-variable correlation, respectively, in the 3DVar static background error covariance.

1. Introduction

The accuracy of convective-scale precipitation forecasts depends not only on convective-scale processes, but also on the synoptic-scale and mesoscale environment.

Corresponding author address: Dr. Xuguang Wang, School of Meteorology, University of Oklahoma, 120 David L. Boren Blvd., Norman, OK 73072.
E-mail: xuguang.wang@ou.edu

DOI: 10.1175/MWR-D-14-00345.1

© 2015 American Meteorological Society
and interactions across multiple scales (e.g., Lorenz 1969; Perkey and Maddox 1985; Zhang et al. 2007; Rotunno and Snyder 2008). A unique challenge for storm-scale\textsuperscript{1} data assimilation (DA) is therefore to properly estimate the atmospheric state on a broad range of spatial scales. Observations resolving synoptic-, meso-, and convective-scale features are thus needed. The forecast model must also permit convective-scale motions over a large enough domain to resolve synoptic-scale features.

Early studies assimilating convective-scale radar data used a homogeneous ambient environment derived from a representative atmospheric sounding (e.g., Snyder and Zhang 2003; Dowell et al. 2004, 2011; Caya et al. 2005; Aksoy et al. 2009, 2010). Some studies used coarser-resolution model analyses and forecasts as background fields for storm-scale DA, often adopting different independent DA methods for each (Xiao and Sun 2007; Stephan et al. 2008; Zhao et al. 2008; Dixon et al. 2009; Snook et al. 2011; Caron 2013; Brousseau et al. 2014; Chang et al. 2014; Simonin et al. 2014; Wheatley et al. 2014). A few studies have applied three-dimensional variational (3DVar) techniques iteratively with successively smaller length scales of background error covariance to generate multiscale analyses (e.g., Crook and Sun 2002; Dixon et al. 2009; Xie et al. 2011, Schenkman et al. 2011). Studies have also started to use ensemble-based DA methods such as the ensemble Kalman filter (EnKF; Evensen 2003; Hamill 2006) to provide multiscale analyses by assimilating observations resolving synoptic to convective-scale features (Zhang et al. 2009; Yussouf et al. 2013; Thompson 2014; Sobash and Wicker 2014). In such studies, the same ensemble-based DA system is used to analyze both the storm-scale features and the meso- to synoptic-scale environment.

One advantage of ensemble-based DA over 3DVar is the potential for more accurate spatial and cross-variable correlations between model state and observed variables. This is made possible by the flow-dependent ensemble-based background error covariance. There has not yet been a systematic comparison of the variational and ensemble-based methods in the context of multiscale DA where scales ranging from several to thousands of kilometers are resolved by both the model and the observations. Past studies comparing 3DVar with ensemble-based DA focus either on large-scale (i.e., relatively coarse resolution requiring cumulus parameterization) or convective-scale (e.g., radar observation) DA alone. For example, mesoscale to global-scale (grid spacing tens to thousands of kilometers) studies have shown EnKF and other ensemble-based DA to produce both more accurate analyses and forecasts than 3DVar, especially in data-sparse regions (e.g., Meng and Zhang 2008, 2011; Whitaker et al. 2008, 2009; Wang et al. 2008a,b; Wang 2011; X. Wang et al. 2013; Yang et al. 2009; Buehner et al. 2010; M. Zhang et al. 2011; Schwartz and Liu 2014). Convective-scale (grid spacing of several kilometers) studies comparing 3DVar and ensemble-based DA have been more limited in terms of the amount of studies, the types of radar observations assimilated, and the diversity of cases studied. For example, Potvin et al. (2013) compared 3DVar and EnKF supercell analyses with radial wind observations. Li et al. (2012) compared tropical cyclone forecasts initialized from 3DVar and hybrid ensemble–3DVar analyses with radial wind observations. Caya et al. (2005) compared 4DVar and EnKF for convective-scale radar DA in a perfect model OSSE framework for an isolated supercell case. Carley (2012) showed that forecasts initialized from hybrid ensemble–3DVar analyses outperformed those from 3DVar for convective-scale radar DA with a single case that featured upscale growth of supercells into a multicellular mode. Carley (2012) used the Non-hydrostatic Multiscale Model on the B grid (NMMB) and the ensemble transform technique, rather than EnKF, to obtain the ensemble part of the hybrid background error covariance. Different from these early studies, the present study compares 3DVar and EnKF in the multiscale scenario and uses 10 diverse midlatitude convection cases.

A Gridpoint Statistical Interpolation (GSI)-based hybrid EnKF-3DVar system has been implemented operationally as part of the Global Forecast System (GFS) at the National Centers for Environmental Prediction (NCEP). The newly implemented hybrid system improved both global forecast and hurricane forecast applications (Hamill et al. 2011; X. Wang et al. 2013; Wang and Lei 2014). The GSI-based hybrid system has also been integrated with other regional modeling systems such as the North American Mesoscale Forecast System (NAM) using the NMMB, the Rapid Refresh (RAP) using the Advanced Research version of the Weather Research and Forecasting (WRF) Model (ARW), and the Hurricane WRF (HWRF). These regional NCEP operational models currently use the ensemble covariance from the GFS in their hybrid DA systems. This study extends the GSI-based 3DVar and EnKF systems by further developing the convective-scale radar DA capability for the ARW model to enable multiscale DA. Since the processing of observations, quality control,
and observation forward operators are all unified under the GSI framework, the GSI-based system allows a clean and direct comparison for understanding the 3DVar and EnKF in multiscale DA scenarios. Such understanding will contribute to the development of a GSI-based hybrid system for multiscale DA, which will be the focus of future studies.

The purpose of the present study is to better understand the differences between EnKF and 3DVar analyses in the context of multiscale DA, and the impact of such differences on subsequent convection-permitting precipitation forecasts. Here, multiscale DA refers to the assimilation of observations from networks that have been designed to sample different scales of motion ranging from synoptic-scale rawinsonde observations to convective-scale radar observations. The cases selected for systematic evaluation represent convective organization on scales ranging from discrete cellular convection to supercells to organized mesoscale convective systems (MCSs). The cases also include forcing mechanisms on a range of scales such as synoptic-scale waves and fronts, mesoscale features such as drylines and storm-scale features such as cold pools. A case study of an upscale growing MCS is also evaluated in greater detail to further understand the systematic differences.

In section 2, the case study and forecast periods, and details of the GSI-based 3DVar and EnKF DA systems, are described. The systematic results over 10 cases are presented in section 3 and the results for the case study are presented in section 4. Section 5 contains a brief summary and discussion of conclusions.

2. Methods

a. Extension of GSI-based 3DVar and EnKF for direct assimilation of radar observations

Each of the GSI-based 3DVar and EnKF systems are extended to have the capability to directly assimilate radar observations. This capability is needed to explore the differences between 3DVar and EnKF in the multiscale data assimilation scenario. In this subsection, the extension for GSI 3DVar is first described followed by the extension for GSI-based EnKF.

GSI-based 3DVar combines the first-guess background forecast and assimilated observations by variational minimization of a cost function [e.g., Eq. (2.1) of Wu et al. 2002]. The cost function includes penalty terms for the difference between the analysis and the observations, relative to the observation error covariance, and for the difference between the analysis and the background forecast, relative to the background error covariance. In 3DVar, the background error covariance is predefined and quasistatic. 3DVar therefore requires specification of a static background error covariance, which affects how the observation information is spread out into the analysis. For the current GSI-based 3DVar, only the radar radial wind can be assimilated during variational minimization. To extend GSI-based 3DVar to directly assimilate radar reflectivity observations, additional control variables, forward observation operator, and background error statistics are developed. In this study, WRF single moment 6-class (WSM6; Hong and Lim 2006) microphysics is used. The new control variables added in GSI variational minimization are the mixing ratios of rain, snow, and graupel (i.e., hail) hydrometeors. The logarithm is first applied to these hydrometeor mixing ratio control variables to reduce non-Gaussianity of the error statistics and to minimize the errors associated with the linearization and adjoint of the reflectivity observation operator, Carley (2012) also extended the GSI variational minimization with hydrometeor control variables for the Ferrier microphysics scheme (Ferrier et al. 2002, 2011) and conducted experiments with the NMMB model.

An observation forward operator consistent with WSM6 is introduced into the GSI-based 3DVar system for reflectivity DA. The observation operator defines the relationship between model control variables (hydrometeors) and the observed quantity (reflectivity). The reflectivity is a function of the rain, snow, and graupel hydrometeor mixing ratios that also depend on the background temperature. The observation operator follows Dowell et al. (2011) except that, in addition to snow, graupel is also classified as either wet or dry based on the background temperature as in Tong and Xue (2005). The analysis increment is divided among rain, snow, and graupel based on the relative magnitude of the background error variance for the different hydrometeor types (defined below). The contribution to reflectivity from rain and snow/graupel is therefore set to zero if the background temperature is less than −5°C or greater than 5°C, respectively. This step eliminates the possibility of unrealistic increments that add snow/graupel at very warm levels or rain at very cold levels (Gao and Stensrud 2012).

One of the challenges of 3DVar for radar reflectivity assimilation is the specification of an appropriate static background error covariance model. The background errors of hydrometeor variables can be highly correlated with errors in other model variables (Michel et al. 2011). However, including such cross-variable correlations in the static background error covariance model in a computationally efficient manner during variational minimization remains a challenge. Therefore, cross-variable correlations between hydrometeor and other control variable errors have typically been neglected in...
the static covariance in previous variational assimilation of reflectivity observations (e.g., Caya et al. 2005; Carley 2012; H. Wang et al. 2013). This approach is also used in the present study. Future work will explore the inclusion of such cross-variable correlations associated with reflectivity assimilation through the use of ensemble covariance in the variational framework, following a similar approach as the hybrid ensemble–variational DA method (e.g., Wang et al. 2008a; Wang 2010).

Here, the static 3DVar background error covariance for reflectivity observation assimilation is defined as follows. The amplitudes of the background error variances for the hydrometeor mixing ratios are defined as a function of height and hydrometeor type. The height dependence of the background error variance for each hydrometeor type is the variance of 5-min ensemble forecasts initialized at 0000 UTC 20 May 2010, averaged over the convectively active region. The amplitude is further tuned to obtain analyses that are subjectively similar to the observations when used to assimilate reflectivity observations in test cases. The horizontal correlation length scale of the hydrometeor background errors is determined by tuning the horizontal correlation length scale of the static background error covariance used by the mesoscale NCEP NAM model to assimilate conventional water vapor observations. The e-folding distance for the mesoscale observations is about 215 km (i.e., equivalent to a cutoff radius of 600 km; Pan et al. 2014) in the horizontal and about 4 km (equivalent to a cutoff radius of 11 km) in the vertical. The horizontal correlation length scale is reduced by a factor of 20 to ∼11-km e-folding distance and the vertical correlation length scale is reduced by a factor of 4 to about 1-km e-folding distance. These values are also chosen to minimize objective and subjective errors during test cases. The static background error covariance for radar radial wind assimilation is also tuned by reducing the spatial scale using the same factors as for reflectivity.

The GSI-based EnKF is based on the ensemble square root filter (EnSRF) of Whitaker and Hamill (2002). Like the GSI-based 3DVar, the GSI software performs the observation quality control and applies the observation operators to the model first-guess fields. The GSI-based EnKF has the option to take into account the four-dimensional ensemble covariance within the assimilation window to assimilate asynchronous observations. This EnKF code has been efficiently parallelized following Anderson and Collins (2007) and directly interfaced with the GSI by using the GSI’s observation operators, preprocessing, and quality control for operationally assimilated data. In the GSI-based EnKF currently implemented operationally for NCEP GFS, it does not contain radar data assimilation.

In this study, the GSI-based EnKF is further extended to include the assimilation of radar data through three new developments. First, the ensemble observation priors are extended to include both radar radial wind and reflectivity. This is accomplished by applying the GSI observation operators developed for GSI 3DVar on the first-guess forecast ensembles and ingesting the resulting observation priors (i.e., model variables converted to observation space) into EnKF. Second, the option to include rain, snow, and graupel hydrometeor mixing ratios as state variables is added to the EnKF code. Third, the EnKF WRF interface is extended to read the ensemble first guess and update the ensemble analysis of these new state variables.

### b. Forecast period and case study overview

Retrospective forecasts during the convectively active month of May 2010 are used to provide a robust evaluation of the impact of the differences in multiscale GSI-based EnKF and 3DVar. Since not all days during this period contained widespread convection in the central United States at and after 0000 UTC when the forecasts are initialized, 10 of the most convectively active days at 0000 UTC are selected (Table 1). Diverse events are used to obtain robust averaged results. The events include many examples of both discrete cellular convection and organized MCSs. The events also include diverse forcing and organizing mechanisms such as strong upper-level shortwaves and surface cold fronts, slow-moving or stationary frontal zones with relatively weak large-scale ascent, and convective storm outflows. Since the focus of each convective episode was in a slightly different geographic location for each day, the center of the model domain was relocated for each day (Table 1).

<table>
<thead>
<tr>
<th>Analysis time and date</th>
<th>Domain latitude (°N)</th>
<th>Domain longitude (°W)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0000 UTC 11 May 2010</td>
<td>37.5</td>
<td>98.0</td>
</tr>
<tr>
<td>0000 UTC 13 May 2010</td>
<td>35.5</td>
<td>100.0</td>
</tr>
<tr>
<td>0000 UTC 15 May 2010</td>
<td>32.0</td>
<td>101.5</td>
</tr>
<tr>
<td>0000 UTC 17 May 2010</td>
<td>34.0</td>
<td>95.0</td>
</tr>
<tr>
<td>0000 UTC 18 May 2010</td>
<td>29.0</td>
<td>101.0</td>
</tr>
<tr>
<td>0000 UTC 19 May 2010</td>
<td>36.0</td>
<td>102.5</td>
</tr>
<tr>
<td>0000 UTC 20 May 2010</td>
<td>35.5</td>
<td>100.0</td>
</tr>
<tr>
<td>0000 UTC 21 May 2010</td>
<td>34.0</td>
<td>94.5</td>
</tr>
<tr>
<td>0000 UTC 25 May 2010</td>
<td>37.5</td>
<td>102.5</td>
</tr>
<tr>
<td>0000 UTC 26 May 2010</td>
<td>37.5</td>
<td>102.5</td>
</tr>
</tbody>
</table>

In addition to the systematic results, the 0000 UTC 20 May 2010 case study is selected for more detailed and qualitative evaluation of the differences between the
GSI-based EnKF and 3DVar. This event is selected because the convective cells present at 0000 UTC grew upscale into an MCS during the forecast period. Forecasts for such events are particularly sensitive to both the storm-scale analysis and larger-scale environment because of the multiscale nature of such convective systems (e.g., Perkey and Maddox 1985) and upscale growth of small errors for this event (Johnson et al. 2014). The synoptic-scale environment was characterized by a broad slow-moving upper-level trough with an embedded shortwave approaching the southern plains around 0000 UTC. At the surface a dryline–warm-front intersection provided a focus for supercell development in northern Oklahoma between 2100 and 0000 UTC. These storms then grew upscale into an MCS as they propagated eastward along the warm-front boundary over the next several hours (Fig. 1). For more information on the synoptic-scale environment for this event, please see Johnson et al. (2014).

c. Assimilated observations and quality control for radar data

Since both the 3DVar and EnKF techniques are GSI-based, the same observations and same quality control methods are used for both techniques. The synoptic–mesoscale observation data are obtained from the Climate Forecast System Reanalysis project at the NOAA Operational Model Archive and Distribution System (NOMADS). These observations include rawinsonde, surface station, surface mesonet, Aircraft Communications Addressing and Reporting System (ACARS), and NOAA wind profiler observations (Fig. 2).

Radar observations of reflectivity and radial velocity are obtained from the NEXRAD level 2 data archived at the National Climatic Data Center (NCDC) and quality controlled using the Warning Decision Support System–Integrated Information (WDSSII; Lakshmanan et al. 2007b) software (www.wdssii.org). For the reflectivity data, the neural-network-based w2qcnn utility within WDSSII (Lakshmanan et al. 2007a, 2010) is used to remove nonmeteorological echoes such as anomalous propagation and biological echoes (e.g., insects). The reflectivity data are then thresholded at 5 dBZ such that all data less than or equal to 5 dBZ are considered “no precipitation” observations (Aksoy et al. 2009). Velocity data are then dealiaised using a two-dimensional dealiasing algorithm (Jing and Wiener 1993). Velocity data are also thresholded based on the reflectivity data, such that the velocity data are omitted where the reflectivity is less than or equal to 5 dBZ (Aksoy et al. 2009). Additional gross error checks are also performed within GSI. Velocity observations are rejected if the difference from the background value is greater than 30 m s\(^{-1}\), which is only likely to occur in cases of extreme aliasing that was missed during preprocessing, although a very poor background forecast could also lead to such differences. Reflectivity observations are not rejected based on the observation–background difference because very large differences may not indicate bad observation data, especially during the early DA cycles. The observation error of radar velocity and reflectivity is assumed to be 2 m s\(^{-1}\) and 5 dBZ, respectively.

d. Model and DA system configuration

The ARW, version 3.2, is used for all simulations (Skamarock and Klemp 2007). An outer domain is configured with 12-km grid spacing and 50 vertical terrain-following levels over a 326 \(\times\) 259 gridpoint domain (Fig. 2). The physics configuration includes the Mellor–Yamada–Janjić PBL scheme with eta surface exchange parameterization (Janjić 1994; 2001), Noah land surface model (Ek et al. 2003), Rapid Radiative Transfer Model for general circulation models
longwave (Iacono et al. 2008) and Goddard shortwave radiation (Tao et al. 2003), WSM6 microphysics (Hong and Lim 2006), and Grell-3 cumulus parameterization (Grell and Dévényi 2002). An inner nested domain is configured similarly, except with 4-km grid spacing over a 346 × 277 gridpoint domain (Fig. 2) and no cumulus parameterization.

An important consideration for cycled multiscale DA with observations and models that resolve many different scales of motion is the choice of cycling frequency. Peña and Kalnay (2004) suggest that the cycling frequency can be consistent with either the larger-scale baroclinically driven or smaller-scale convective-instability-driven modes of error growth, but not both. This is because the atmosphere contains fast-growing (i.e., moist convective instability) errors that saturate at a lower amplitude than slower-growing (i.e., synoptic-scale baroclinic instability) errors. Here, the storm-scale and synoptic–mesoscale observations are assimilated with different cycling frequencies, chosen to correspond to the approximate error growth rates of features well observed by each. A 3-h cycling interval is used to assimilate the synoptic–mesoscale observations on the outer domain, and a 5-min cycling interval is used to
assimilate the storm-scale radar observations on the inner domain (Fig. 3). Different cycling intervals for outer-domain DA and inner-domain radar DA were also used in Yussouf et al. (2013).

Synoptic/mesoscale observations within the 3-h window centered on the analysis time are assimilated. For both 3DVar and EnKF, first guess at appropriate time (FGAT) is used by outputting first-guess fields at 30-min increments until 1.5 h after the analysis time. The first-guess fields are then compared to the observations at the observation time using linear interpolation in time. In addition, for EnKF, asynchronous assimilation (Sakov et al. 2010) is adopted, where the ensemble covariance between 4D observation priors valid at observation times and priors in model state space valid at the analysis time is calculated.

The second-to-last outer-domain analysis at 2100 UTC is downscaled to the inner domain using the WRF ndown utility to initialize the first inner-domain 5-min forecast for the inner-domain DA cycles (Fig. 3). For each member, the forecast from the 2100 UTC outer-domain analysis provides the inner-domain lateral boundary conditions (LBCs). Radar observations within 5-min windows are assimilated every 5 min (synchronously) until the final analysis time at 0000 UTC. Before the final analysis of radar observations at 0000 UTC, the synoptic–mesoscale observations are also assimilated on the inner domain using the same parameters (localization, inflation, etc.) used to assimilate synoptic–mesoscale observations on the outer domain. This final step reduces inconsistency between the outer- and inner-domain analyses of the synoptic–mesoscale environment. In this study, “mesoscale analysis” refers to the analysis generated by the outer-domain DA, which assimilates observations that only resolve mesoscale and larger features using a 12-km grid. The “multiscale” analysis refers to the result of the inner-domain DA, which includes further storm-scale radar DA using the mesoscale analysis as a background field.

For the 3DVar experiments, specification of a static background error covariance is needed. In this study, the outer domain has comparable resolution to the operational NAM. The background error covariance from the regional NCEP NAM (Hu et al. 2011) is therefore adopted for nonradar observations assimilated on the outer domain. The background error covariance for assimilation of radar observations is constructed and tuned as described above in section 2a.

For the EnKF experiments, each forecast step contains a 40-member ensemble forecast and each analysis step provides a 40-member analysis ensemble. The 12-km grid ensemble at the very beginning of the DA cycles is created by adding random perturbations to the operational NCEP GFS analysis. These perturbations are drawn so that their covariance is equal to the static covariance in the WRF 3DVar (Wang et al. 2008a). The same method is used to perturb the 12-km outer-domain LBCs. The outer-domain ensemble provides the initial and lateral boundary conditions for the inner-domain ensemble.

The general EnKF–EnSRF theory and equations have been described in many papers [e.g., Eqs. (1)–(5) of Whitaker and Hamill 2002; Whitaker et al. 2008]. One challenge in applying EnKF is the treatment of system errors associated with the sampling errors and misrepresentation of model errors. Covariance localization and inflation are commonly used to treat such deficiencies. Optimal methods and parameters for covariance localization and inflation are application dependent. The details of the methods used for the present study are outlined below.

Two methods of posterior covariance inflation are used starting from the values found for Whitaker and Hamill (2012). The parameters are tuned to minimize the first-guess errors during the 20 May case study. The
first-guess errors for the other cases are similar in magnitude to the 20 May case used for tuning, suggesting that the parameters are also generally appropriate for other cases. The first inflation method is a height-dependent multiplicative inflation that is applied uniformly across the domain to all ensemble perturbations. This multiplicative inflation is intended to account for model errors that are not represented in the ensemble (Whitaker and Hamill 2012), such as the errors associated with physical parameterization. The amount of multiplicative inflation at the surface is 15% every 3 h on the outer domain or, equivalently, \( \approx 0.4\% \) every 5 min on the inner domain. The multiplicative inflation smoothly tapers to \( \approx 9\% \) at 200 mb (1 mb = 1 hPa) and \( \approx 3\% \) at the 50-mb model top to avoid excessive spread near the model top, similar to Zhu et al. (2013). The second inflation method is relaxation to prior spread (RTPS; Whitaker and Hamill 2012), which inflate the posterior ensemble spread to a fraction, \( \alpha \), of the prior ensemble spread. The RTPS accounts for excessive spread reduction during the assimilation of observations resulting from sampling errors in the ensemble approximation of the Kalman gain. Thus the RTPS inflation is greatest where there are many observations and is absent where there are no observations. For our experiments, a value for \( \alpha \) of 0.95 is used to inflate the posterior ensemble spread to 95% of the prior ensemble spread for both the inner domain and outer domain. On the inner domain, the average consistency ratio of observation first guesses (i.e., ensemble spread divided by ensemble mean error) for wind, temperature, and water vapor is 0.87 at the end of DA, indicating a reasonably well-tuned system. On the outer domain, the consistency ratios are stable and, similar to other studies (e.g., Aksoy et al. 2009; Dowell et al. 2011; Sobash and Stensrud 2013), indicate some ensemble underdispersion with values \( \approx 0.75 \) for both radial velocity and reflectivity. However, tuning tests show slight degradation of accuracy when the inflation is further increased (not shown). While the diagnostic statistics during DA were somewhat sensitive to these inflation parameters, the sensitivity of the overall results to these parameters was much less than the sensitivity to the DA technique (i.e., EnKF versus 3DVar). Therefore more complex adaptive inflation methods (e.g., Anderson 2007) were not explored.

Covariance localization is used to minimize the impact of sampling errors in the ensemble covariance, which are greatest where the actual correlation is small, such as at large distances (Sobash and Stensrud 2013). The covariance localization is applied using the Gaspari and Cohn (1999) function with a cutoff radius also tuned to minimize the first-guess errors during the 20 May case study. For the assimilation of synoptic–mesoscale observations on the outer domain, the localization is also height and variable dependent (Zhu et al. 2013). For the synoptic–mesoscale observation assimilation, the horizontal localization is set to 700 km, similar to the \( \approx 600\text{-km} \) cutoff radius for 3DVar, at the surface and increases by a factor of 1.5 at the model top (Fig. 4). Vertical localization increases from 0.275 to 0.55 scale height (natural log of pressure) for temperature and moisture and increases from 0.55 to 1.1 for wind (Fig. 4). For the inner-domain storm-scale DA, constant covariance localization length scales are used. For the inner-domain DA, tests of horizontal covariance localization showed very little sensitivity of the forecast to the cutoff radius. Among the range of radii tested, a value of 20 km (i.e., 5 grid points) showed a slight improvement over other values such as 16 and 12 km, although the sensitivity of the results to the covariance localization was also generally small. The chosen value of 20 km is comparable to the 18-km radius that was found to work well in Sobash and Stensrud (2013) for a similar 50-member ensemble. However, this value is somewhat larger than that used in many early studies of EnKF with radar data. Such early studies focused primarily on isolated supercells, whereas Sobash and Stensrud (2013) suggest that the larger localization
radius is beneficial for other convective modes, such as cell mergers and MCS cases. The grid spacing of 4 km also necessitates a larger radius than used in past studies with (1–3)-km grid spacing. The vertical localization for radar reflectivity and velocity assimilation is 1.1 in scale height units.

3. Results aggregated over 10 forecasts

In addition to providing a background environment consistent with the storm-scale features to be assimilated, the mesoscale environment also interacts with the storm-scale features during the DA and forecast periods to affect the overall forecast evolution (e.g., Perkey and Maddox 1985). Therefore, the mesoscale analyses are first evaluated to distinguish the impacts of the synoptic–mesoscale environment differences on the subsequent forecasts in section 3a, followed by evaluation of the multiscale analyses with radar DA in section 3b.

a. Mesoscale analysis evaluation

Since the true atmospheric state is unknown and approximated by the DA analyses, the quality of such analyses is evaluated based on the similarity of a subsequent forecast to independent observations. For example, the first-guess errors of the short-term forecasts during the DA period are commonly used to evaluate DA systems. Here, the first-guess errors are averaged only over the last 5 cycles to allow the EnKF to spin up reasonable estimates of the background error covariance. This approach also emphasizes the end of the DA period, which is consistent with the focus on the final 0000 UTC analysis time. The first-guess errors during the mesoscale DA on the 12-km grid are generally smaller for EnKF than for 3DVar (Fig. 5), with the exception of temperature between 500 and 700 mb and above 100 mb. The difference is fairly uniform with height for wind (Fig. 5b) and is most pronounced at low levels for temperature and moisture (Figs. 5a,c). The differences in first-guess error are generally statistically significant at most levels for wind and moisture (Figs. 5b,c), but are more limited for temperature (Fig. 5a). This limited temperature advantage may be due to systematic model biases in temperature at those levels. A more pronounced warm bias for EnKF than 3DVar was noted both between 500 and 700 mb and above 100 mb (not shown). A WRF Model temperature bias near the model top (i.e., above 100 mb) was also documented by Wee et al. (2012). EnKF may not correct these systematic biases as well as 3DVar because such biases would be common to all members and therefore lead to an underestimate of the background error based on the ensemble variance. While further covariance inflation may improve EnKF performance at such levels, there is an overall negative effect of additional inflation when all levels and variables are considered. The smaller first-guess errors using GSI-based EnKF suggest that the synoptic–mesoscale environment is analyzed more accurately than using GSI-based 3DVar.

Free forecasts are also run out to 8-h lead time on the 4-km grid by interpolating the mesoscale analyses and hourly forecasts to the 4-km grid using the WRF ndown

Fig. 5. RMSE of outer-domain first guess (i.e., 3-h forecast) from GSI-EnKF and -3DVar of (a) temperature, (b) wind, and (c) water vapor mixing ratio observations averaged over the inner-domain region, excluding the first 3 DA cycles while the ensemble covariance spins up (i.e., last 5 cycles only) and averaged over all 10 cases. Markers indicate a significant difference between the two lines at the 90% level (crosses) or 95% level (asterisks). Statistical significance is determined using permutation resampling (Hamill 1999).
utility to generate the initial and lateral boundary conditions, respectively. The EnKF-initialized forecasts are initialized from the ensemble mean analysis. Precipitation forecasts are verified against radar-derived quantitative precipitation estimates from the National Severe Storms Laboratory Q2 product (J. Zhang et al. 2011). The better mesoscale analyses for EnKF than 3DVar are also reflected in the equitable threat score (ETS; Wilks 2011) of 8-h precipitation forecasts on the 4-km grid without radar DA (Fig. 6; dashed lines). The results discussed herein are similar using the Hiedke skill score (Wilks 2011) and neighborhood probability Brier skill score (not shown). The precipitation forecasts initialized from the mesoscale analyses show the contribution of the synoptic–mesoscale environment analysis to the forecast skill differences. Although the 3DVar precipitation forecasts are more skillful than the EnKF precipitation forecasts for the first couple of hours without radar DA (i.e., mesoscale/synoptic DA only), the EnKF forecasts become more skillful starting at about forecast hour 3 (Fig. 6; dashed lines). The differences in skill are consistent with past studies showing 3DVar to fit to observations better than EnKF at the analysis and very short forecast times but with faster error growth during the forecast for 3DVar (e.g., Wang et al. 2008b; Li et al. 2012). The initially closer fit to observations for 3DVar may be important for more quickly spinning up the initial storms. The difference in skill during the first 2 h may also be related to a smoothing of features in the ensemble mean used for the EnKF background. A smoothing of features associated with focused convergence can slow the spinup of corresponding convection. The better performance of the EnKF forecasts at later lead times after the short spinup period suggests that the larger-scale environment is more supportive of the actual convective evolution in the EnKF analyses than the 3DVar analyses.

b. Multiscale analysis evaluation

The mesoscale analyses drive the inner-domain radar DA, which adds storm-scale features to provide the final multiscale analyses. The short lead time (5 min) first-guess forecasts during radar DA are used to evaluate the analysis of storm-scale features. First-guess errors are averaged over the last part of the DA period (here, 12 cycles or 1 h) to avoid the initial spinup period and focus on the final analysis time. The 4-km-domain EnKF first-guess errors are significantly smaller than the 3DVar first-guess errors for radial velocity at all levels (Fig. 7a) and for reflectivity at many levels (Fig. 7b). The exceptions for reflectivity include the near-surface level and the 500–600-mb level (Fig. 7b). Therefore the storm-scale details on average are also better analyzed with GSI-based EnKF than GSI-based 3DVar. Longer (8 h) forecasts are used to evaluate the combined influence of the multiscale analysis of storm-scale and mesoscale features. The 8-h forecasts initialized from the multiscale analyses are also more skillful using GSI-based EnKF than using GSI-based 3DVar. The difference is generally statistically significant except for the highest threshold at several times, likely because of the higher threshold being a rarer event (Fig. 6; solid lines). The more pronounced difference in skill indicates that the storm-scale part of the analysis further improves the relative performance of EnKF over 3DVar. Thus, all scales of motion contributing to the convective
evolution, not only the mesoscale environment, are better analyzed with GSI-based EnKF than GSI-based 3DVar. At the later lead times the skill of forecasts initialized from the multiscale analyses (Fig. 6; solid lines) is generally similar to the skill of forecasts initialized from the mesoscale analyses (Fig. 6; dashed lines). This shows that at later lead times the differences in the mesoscale environment contribute increasingly to the difference in forecast skill. This contrasts with the dominant impact of the storm-scale analysis at early lead times.

The impact of storm-scale radar DA on the precipitation forecasts, compared to the mesoscale analyses without radar DA, is also notably different for EnKF and 3DVar. The impact of the better storm-scale analyses for EnKF is to increase the skill of the precipitation forecast for \( \sim (4-5) \) h (Fig. 6). However, for 3DVar the precipitation forecast skill is only improved by the multiscale analysis during the first hour and is then degraded compared to the mesoscale analysis at later times (Fig. 6; black dashed versus black solid). This shows that the multiscale EnKF analyses lead to forecasts that more realistically maintain the storm-scale features and their interaction with the larger-scale environment, compared to the multiscale 3DVar analyses.

4. 20 May 2010 case study

A case study is used to better understand the systematic differences between the GSI-based EnKF and
3DVar multiscale analyses and subsequent forecasts. This case of an upscale growing MCS is selected for evaluation of multiscale analyses because of strong sensitivity to analysis perturbations on all spatial scales (Johnson et al. 2014). The mesoscale analyses are again evaluated first to distinguish the impact of differences in the synoptic–mesoscale environment from the storm-scale analysis impacts.

a. Mesoscale analysis evaluation

The mesoscale analyses in this case lead to generally smaller first-guess errors during the outer-domain DA period for GSI-based EnKF than for GSI-based 3DVar (Fig. 8). The EnKF advantage is seen for all variables at most levels, with the exception of temperature near the model top (Fig. 8a). These exceptions are likely related to the WRF temperature bias at such levels, as mentioned in section 3. This case is thus representative of the systematic results in that the GSI-based EnKF provides a better analysis of the synoptic–mesoscale environment than the GSI-based 3DVar.

The impact of a more accurately analyzed synoptic–mesoscale environment on the precipitation forecasts is seen in the neighborhood probability (NP; Schwartz et al. 2010) forecasts initialized from the mesoscale analyses (Fig. 9). The 3DVar and EnKF forecasts both develop an MCS in about the same location as observed during the second forecast hour (Figs. 9b,j). However, for both the 3DVAR (Fig. 9i) and EnKF (Fig. 9a) forecasts, the MCS takes more than an hour to “spin up,” resulting in underprediction of precipitation during the first hour and a slight westward displacement of the NP maximum relative to the observed MCS at later forecast times (e.g., Figs. 9c,k). Furthermore, the MCS in the 3DVar forecast starts to dissipate several hours too early (Figs. 9n–p). During this time spurious convection develops in southeastern Oklahoma and becomes dominant as it moves into southwestern Arkansas (Figs. 9j–p). The spurious precipitation develops along a northwest–southeast-oriented warm front in southeastern Oklahoma. The development and dominance of this spurious precipitation in the 3DVar forecast is a result of the poorer synoptic–mesoscale environment analysis for 3DVar. For example, the warm front—including location, wind shift, and enhanced surface temperature gradient—is much better analyzed by the EnKF analysis than the 3DVar analysis. The difference is still present in the final multiscale analysis (Fig. 10), showing the importance of the mesoscale environment analysis.

Subjective evaluation of the mesoscale analyses throughout the DA period was conducted to better

Fig. 9. Neighborhood probability forecast of hourly accumulated precipitation > 12.7 mm h\(^{-1}\) (gray shaded) and observation red contour of 12.7 mm h\(^{-1}\). Forecasts are initialized from downscaled outer-domain (a)–(h) EnKF and (i)–(p) 3DVar analyses.
understand how the differences between the 3DVar and EnKF methods contributed to the above differences in the final mesoscale analyses. The southwestward displacement of the warm front analyzed by 3DVar, compared to the warm front analyzed by EnKF, first appears in the 1500 UTC background forecast (Figs. 11a,b; thick black line). Between 1200 and 1500 UTC, low-level clouds developed north of the warm front in the 3DVAR forecast only (not shown). The clouds prevented surface warming and impeded the northward advancement of the warm front during subsequent cycles. This is evident at 1500 UTC by the cooler temperatures north of the warm-front wind shift and the more pronounced temperature gradient in the 3DVar background field (Fig. 11a), compared to the EnKF background field (Fig. 11b). The difference is particularly pronounced along the Arkansas–Oklahoma border (Figs. 11e,f; area in red circle). The observation innovations (observation minus background) in this area are generally positive by several kelvins for 3DVar (Fig. 11e). However, the resulting 3DVar increment (analysis minus background) is only positive in southern Arkansas and is neutral and even negative along the Arkansas–Oklahoma border despite the positive nearby innovations (Fig. 11c). This shows that the 3DVar increment did not correct the background errors in the warm-front location and temperature gradient in this area.

The inconsistency between the observation innovation and analysis increment along the Arkansas–Oklahoma border for 3DVar is attributed to the static background error covariance. In particular, although the background error covariance length scales represent the average spatial correlation of background errors, the spatial correlation does not reflect the shape and spatial extent of the relevant mesoscale features for this particular case, such as cold pools and warm fronts. Both 3DVar and EnKF are warmer than the observations in northern and central Oklahoma (Figs. 11e,f). This background forecast error is a result of inadequate resolution of an MCS and associated cold pool (Figs. 11a,b; thick purple line) in northern Oklahoma on the outer-domain 12-km grid. Both 3DVar and EnKF also have negative innovations in southeastern Oklahoma just south of the warm front (Figs. 11e,f). The corresponding EnKF increments are focused along the east–west-oriented temperature gradient in northern Oklahoma, effectively enhancing and shifting southward the cold-pool boundary (Fig. 11d). The EnKF increments are also elongated along the wind shift and temperature gradient in southeastern Oklahoma associated with the warm front (Fig. 11d). In contrast to the flow-dependent shape and localized spatial scale of the EnKF increments, the 3DVar surface temperature increments are isotropic and too large in scale for the mesoscale cold pool and warm front. As a result, the 3DVar increment in Oklahoma does not show spatial structure corresponding to these features (Fig. 11c). Because of the relatively large number of observations in central Oklahoma, the impact of the negative increment extends to the Arkansas–Oklahoma border (Fig. 11e). This limits the ability of the sparse observations along the Arkansas–Oklahoma border to adequately correct the too-cold background forecast in this area. Thus, the 3DVar increments do not sufficiently correct the error in the location of the warm front that spurious precipitation develops along in the subsequent forecast.

In summary, the EnKF with flow-dependent background error covariance provides more physically reasonable analysis increments to correct the mesoscale first guess than 3DVar. As a result, a more accurate
synoptic–mesoscale analysis is produced by EnKF at the end of the DA period (i.e., 0000 UTC).

b. Multiscale analysis evaluation

The mesoscale analyses evaluated in the previous subsection provide the background for the inner-domain storm-scale radar DA. In this section, the differences between GSI-based 3DVar and EnKF for the resulting multiscale analyses are evaluated. Also representative of the systematic results, the storm-scale radar DA for this case shows consistently smaller reflectivity and velocity first-guess errors for EnKF than for 3DVar (Fig. 12).

FIG. 11. The 1500 UTC 20 May 2010 outer-domain analysis plotted over the region of the inner domain for (a) 3DVar background temperature (color shading; °F) and wind (barbs; kt); (b) as in (a), but for EnKF. (c) 3DVar increment (analysis minus background) for temperature (shading) with first-guess temperature contours and wind barbs overlaid; (d) as in (c), but for EnKF. (e) 3DVar surface temperature innovations (observation minus background; K; scale on left); (f) as in (e), but for EnKF. The thick black and purple lines in (a) and (b) represent approximate locations of the warm-front and cold-pool boundaries, respectively. The red ellipse in (e) and (f) highlights an area where the too-cold 3DVar background forecast was not correct by the 3DVar data assimilation.
This indicates a better analysis of the storm-scale features for EnKF than 3DVar, in addition to the synoptic–mesoscale features evaluated above.

Forecasts initialized from the multiscale analysis (Fig. 13) further reveal differences in the impact of the storm-scale analysis on the subsequent precipitation forecasts, compared to the mesoscale analysis (Fig. 9). The storm-scale radar DA results in an improved forecast during the first hour for both EnKF and 3DVar because of the reduced spinup time for the MCS (Figs. 9a,i). The reduced spinup time is a result of the convective systems already being present at the initialization time. In addition to allowing the initial storms to spin up, the assimilation of radar observations is also expected to further improve the forecasts. To isolate these two factors, an additional experiment for this MCS case was conducted by downscaling the outer-domain analyses at 2100 UTC, and assimilating the mesoscale observations on the inner domain at 0000 UTC without any radar DA (not shown). This experiment eliminates the factor due to the spinup time difference. Allowing the initial storms to spin up in this additional experiment improves upon the forecast initialized directly from the mesoscale analysis (not shown). However, the forecast initialized from the multiscale analysis is more skillful than the forecast in this additional experiment (not shown). This shows that the advantage of the radar DA is not just limited to reducing the impacts of the spinup time.

For EnKF, the subjective improvement resulting from storm-scale radar DA persists through the 8-h forecast period (Figs. 13a–h). However, for 3DVar the initial storms are not maintained in the forecast and the spurious precipitation in eastern Oklahoma and western Arkansas is further enhanced (Figs. 13i–p). Further diagnostics show that the enhancement of spurious convection is due primarily to convergence resulting from the cold pools emanating from the convection in central Oklahoma (Fig. 10). At later lead times the subjective differences between the forecasts initialized with multiscale analyses (Figs. 13h,p) are similar to the differences between the forecasts initialized with mesoscale analyses (Figs. 9h,p). This shows the increasing impact of the mesoscale analysis at later lead times, compared to the impact of the storm-scale analysis at earlier lead times. The overall result for 3DVar is that after an initial improvement during the first hour, the forecast is similar or even degraded by the radar DA, especially at forecast hours 2–6. This contrasts with EnKF, which shows a subjectively improved forecast at all lead times resulting from the radar DA. This result is also consistent with the systematic impacts of the storm-scale radar DA for 3DVar discussed in the previous section (e.g., Fig. 6).

The generation of excessive and unrealistic cold pools in the 3DVar analysis is illustrated by the early evolution of the initial supercell in western Oklahoma at ~2100 UTC. Both the EnKF and 3DVar techniques are able to increase reflectivity associated with this storm at 2105 UTC that is missing in the first guess (Figs. 14a,b and 15a,b). However, only the EnKF analysis also adjusts other fields such as vertical velocity, temperature, and
humidity to create a weak updraft and more saturated environment corresponding to the added reflectivity (Figs. 14a–d). As a result, a buoyant updraft develops and maintains the convection during the subsequent cycles for EnKF (e.g., Figs. 15e–h). However, for 3DVar the added reflectivity does not correspond to increased humidity in the vicinity of the increased reflectivity (Fig. 15b). Therefore, a weak downdraft forms during the subsequent cycle as a result of precipitation loading and evaporative cooling (Fig. 15g). During subsequent cycles the subsidence results in column stabilization with net warming above ~3 km due to adiabatic descent and net cooling of up to 1–2 K within just 10 min at lower levels where the evaporation of precipitation dominates (Figs. 15c,g,k). For example, the surface between 10 and 15 grid points in the horizontal direction cools from ~293–294 K in the 2105 UTC background to <292 K in the 2115 UTC analysis, with even stronger cooling just above the surface at ~2–3 km (Fig. 15).

Additional experiments also showed that the comparison between EnKF and 3DVar for this case is not particularly sensitive to the length of the radar DA period. For example, instead of doing 3 h of radar DA initialized from the downscaled 2100 UTC outer-domain analyses, 2 h of free forecast were followed by 1 h of radar DA in a sensitivity test (not shown). The results were not sensitive to this difference in system configuration.

In summary, the excessive storm-scale cold pools in the 3DVar multiscale analysis are a result of the lack of coherent cross-variable correlation in the static background error covariance for storm-scale reflectivity assimilation. When hydrometeors are added to unsaturated locations, without corresponding consistent increments to the dynamic and thermodynamic variables, much of the added hydrometeors may evaporate, creating or enhancing evaporative cooling and downdrafts. The storm-scale differences between the EnKF and 3DVar analyses dominate the subsequent precipitation forecasts for several hours. At later lead times [e.g., ~5–8 h], the mesoscale differences discussed in the previous subsection have a greater impact on the precipitation forecasts than the storm-scale differences.

The impact of the lack of cross-variable correlations while assimilating reflectivity data with 3DVar is further evaluated for the 20 May case study in an additional experiment (denoted as 3DVar_RH experiment, Figs. 13q–x). For this experiment, the relative humidity (RH) is set to 100% at every grid point that is at least 1.5 km above ground level and has analysis reflectivity greater than 20 dBZ (Caumont et al. 2010). This experiment improves forecast skill during the first several hours, compared to 3DVar without the RH adjustment (Fig. 13). The improvement is a result of preventing the evaporation of the analyzed hydrometeors and subsequent generation of excessive cold pools. However, the forecast initialized from the 3DVar with RH adjustment analysis is still not as good as the EnKF-initialized forecast at all times (Fig. 13). The inferior performance of
3DVar_RH to the EnKF also demonstrates that, in addition to the inherent cross-variable correlations, the flow-dependence aspect of the EnKF background error covariance also contributes to the superior performance of EnKF multiscale analyses over the 3DVar.

5. Summary and conclusions

The accuracy of storm-scale precipitation forecasts depends not only on processes at the storm scale but also on the mesoscale and synoptic-scale environment supporting them. Therefore, accurate forecasts for convective scales require DA systems to properly estimate the atmospheric state on multiple scales. To perform multiscale data assimilation, the GSI-based DA system, including both 3DVar and EnKF, is extended to directly assimilate radar observations, in addition to the capability to assimilate synoptic–mesoscale observations. In this study, the newly extended multiscale GSI-based DA system is used to compare 3DVar and EnKF in the context of multiscale DA where scales ranging from convective scales to synoptic scales are resolved by both the model and the observations. The purpose of such a comparison is to facilitate understanding of how the differences among DA techniques lead to analysis differences at different scales and their subsequent impact on storm-scale precipitation forecasts.

The comparison of GSI-based EnKF and 3DVar is performed systematically over 10 diverse convectively
active cases in the central United States. The goal is a robust evaluation of the differences between the EnKF and 3DVar techniques for producing analyses at multiple scales. The mesoscale analyses obtained from assimilation of synoptic–mesoscale observations on the outer domain provide estimates of the synoptic–mesoscale environment for the storm-scale radar DA. Multiscale analyses result from the further storm-scale assimilation of radar observations on the inner domain. Comparison of forecasts initialized from the mesoscale and multiscale analyses differentiates the impacts of the different spatial scales on the subsequent precipitation forecasts.

The convection-permitting precipitation forecasts initialized from the multiscale analyses are more skillful with GSI-based EnKF than GSI-based 3DVar for two reasons. First, precipitation forecasts initialized from the mesoscale analyses become more skillful with EnKF than 3DVar after about the 3-h forecast time. This suggests that the synoptic–mesoscale environment is more accurately analyzed by EnKF than 3DVar. Second, the improved forecast skill at early lead times resulting from the further inner-domain storm-scale radar DA lasts about 5 h for EnKF and only 1 h for 3DVar. This suggests that the analysis of storm-scale features is also more accurate using EnKF than 3DVar. After the first hour the forecast initialized from the 3DVar analysis is actually degraded by the storm-scale DA (Fig. 6). The greater benefit of the storm-scale DA at early lead times for EnKF, together with the forecast degradation and inferior synoptic–mesoscale environment at later lead times for 3DVar, explains the systematically better forecasts initialized from the GSI-based EnKF multiscale analyses compared to the GSI-based 3DVar multiscale analyses.
A case study of upscale growth of cellular convection into an MCS during the evening/overnight hours of 19/20 May 2010 is used to qualitatively understand the systematic differences. Convection-permitting forecasts initialized from the EnKF mesoscale analysis show a subjectively better MCS forecast than forecasts initialized from the 3DVar mesoscale analysis. The forecast difference is largely due to a difference in the warmfront analysis including location, temperature gradient, and convergence. These differences in the mesoscale analysis occur because of the lack of flow dependence in the static 3DVar background error covariance. Storm-scale radar DA on top of the mesoscale DA (i.e., multiscale DA) alleviates underforecasting of precipitation during the first hour for both 3DVar and EnKF. However, only the EnKF forecast properly maintains the initial storms, leading to a subjective improvement over the forecast initialized from the mesoscale analysis throughout the forecast period. The initial 3DVar storms quickly collapse and generate unrealistically strong cold pools as a result of the lack of cross-variable correlations in the static background error covariance for hydrometeors. Further diagnostics revealed that the reflectivity observations assimilated with 3DVar are successfully able to correct errors in the precipitation hydrometeor fields. However, corresponding increments to vertical velocity, temperature, and humidity are not obtained. This results in substantial evaporation of hydrometeors added to regions that are subsaturated in the first-guess background field. The evaporative cooling generates excessive cold pools during both the DA and forecast periods for 3DVar. Both the storm-scale and mesoscale analysis differences contribute to the better EnKF forecast initialized from the multiscale analysis. Consistent with the systematic results, the storm-scale analysis dominates the precipitation forecast at early lead times, while the synoptic–mesoscale environment analysis dominates at later times. These results support the hypothesis that skillful convective-scale precipitation forecasts require effective multiscale DA methods.

Further development of the static background error covariance for radar reflectivity DA in 3DVar is clearly needed. It is expected that the 3DVar results can be improved if an appropriate and efficient method of addressing the lack of cross-variable correlation in the 3DVar background error covariance is identified. This study also suggests that if ensemble estimates of the background error covariance are affordable, then using ensemble-based covariance in variational radar DA systems provides a straightforward solution. This method is commonly referred to as hybrid DA (Wang et al. 2008a,b; X. Wang et al. 2013). Static covariance constructed with the simple method in this study even has some useful aspects for reflectivity DA. For example, the initial ensemble downscaled from mesoscale analyses may have very small or zero variance of hydrometeors, limiting the impact of the assimilated reflectivity observations. While studies such as Dowell et al. (2011) have alleviated this issue by adding random noise where observations indicate precipitation should be occurring, making use of the 3DVar static covariance model provides an alternative method (Carley 2012). Compared to EnKF, the static covariance model more quickly and effectively adds reflectivity that is completely absent from the first-guess field. This is evident in larger RMS first-guess errors for reflectivity during the first few forecast cycles for EnKF than for 3DVar (not shown). Furthermore, EnKF can take several cycles for physically reasonable cross-variable correlations with the hydrometeors to develop in the flow-dependent background error covariance (Tong and Xue 2005). As also noted in Caya et al. (2005), this spinup time motivates additional research on hybrid methods to take advantage of both the reduced spinup time of a static background error covariance and the improved forecast performance of the EnKF flow-dependent background error covariance.

Acknowledgments. The work is primarily supported by NSF Awards AGS-1046081 and AGS-1359703. Radar data QC developed by Chris Karstens, fifth author of the paper, was supported by the NOAA/Office of Oceanic and Atmospheric Research under NOAA–University of Oklahoma Cooperative Agreement NA11OAR4320072, U.S. Department of Commerce. Some of the computing for this project was performed at the OU Supercomputing Center for Education and Research (OSCER) at the University of Oklahoma (OU). David Akin, Senior System Administrator at OSCER, provided valuable contributions to debugging and running some of the software on OSCER machines. This work also used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation Grant ACI-1053575.

REFERENCES


——, 2001: Nonsingular implementation of the Mellor-Yamada level 2.5 scheme in the NCEP Meso Model. NCEP Office Note 437, 61 pp.


