Comparing the Assimilation of Radar Reflectivity Using the Direct GSI-Based Ensemble–Variational (EnVar) and Indirect Cloud Analysis Methods in Convection-Allowing Forecasts over the Continental United States

JEFFREY D. DUDA, XUGUANG WANG, AND YONGMING WANG

School of Meteorology, University of Oklahoma, Norman, Oklahoma

JACOB R. CARLEY

NOAA/NCEP/Environmental Modeling Center, College Park, Maryland

ABSTRACT

Two methods for assimilating radar reflectivity into deterministic convection-allowing forecasts were compared: an operationally used, computationally less expensive cloud analysis (CA) scheme and a relatively more expensive, but rigorous, ensemble Kalman filter–variational hybrid method (EnVar). These methods were implemented in the Nonhydrostatic Multiscale Model on the B-grid and were tested on 10 cases featuring high-impact deep convective storms and heavy precipitation. A variety of traditional, neighborhood-based, and features-based verification metrics support that the EnVar produced superior free forecasts compared to the CA procedure, with statistically significant differences extending up to 9 h into the forecast. Despite being inferior, the CA scheme was able to provide benefit compared to not assimilating radar reflectivity at all, but limited to the first few forecast hours. While the EnVar is able to partially suppress spurious convection by assimilating 0-dBZ reflectivity observations directly, the CA is not designed to reduce or remove hydrometeors. As a result, the CA struggles more with suppression of spurious convection in the first-guess field, which resulted in high-frequency biases and poor forecast evolution, as illustrated in a few case studies. Additionally, while the EnVar uses flow-dependent ensemble covariances to update hydrometeors, thermodynamic, and dynamic variables simultaneously when the reflectivity is assimilated, the CA relies on a radar reflectivity-derived latent heating rate that is applied during a separate digital filter initialization (DFI) procedure to introduce deep convective storms into the model, and the results of CA are shown to be sensitive to the window length used in the DFI.

1. Introduction

Accurate initialization of convection-allowing numerical weather prediction (NWP) models is particularly critical for short-range forecasting. While accurate initial conditions are important for good forecasts at all scales, the deficiency in hydrometeor content and convective-scale circulations at initialization are unique issues for higher-resolution models, which can explicitly represent deep moist convection. These issues are particularly important not only because of the ability of NWP models to represent individual convective storms on the model grid using a microphysics parameterization scheme, but also because such high-impact, small-scale features can be easily missed or subsampled by modern observing systems compared to larger-scale features (e.g., jet streams and air masses) that are comparatively well forecast. If high-impact phenomena like convective storms are missing in the initial conditions, subsequent short-range forecasts are likely to also lack these phenomena, which may result in poor forecasts with societally significant impacts.

One of the most useful quantitative observational tools for measuring the water content of thunderstorms is microwave radar. One advantage of using radar data to sample convective storms is that the particular wavelengths used in Weather Surveillance Radar–1988 Doppler (WSR-88D; https://www.weather.gov/iwx/wsr_88d) radars enable the radar beam to penetrate convective storms so that the internal storm structure can be sampled.

Corresponding author: Xuguang Wang, xuguang.wang@ou.edu

Therefore, radar reflectivity and velocity are not just useful, they are critical measurements for use in data assimilation (DA) in convection-allowing NWP models.

A variety of techniques have been developed to assimilate radar data into a convection-allowing model [see Sun (2005) for a review]. Among the most popular are the three-dimensional variational (3DVar; Xiao and Sun 2007; Sugimoto et al. 2009; Schenkman et al. 2011; Gao and Stensrud 2012), four-dimensional variational (4DVar; Caya et al. 2005; Wang et al. 2013), and ensemble Kalman filter (EnKF; Caya et al. 2005; Tong and Xue 2005; Aksoy et al. 2009, 2010; Dowell et al. 2011; Snook et al. 2011; Yussouf and Stensrud 2012; Johnson et al. 2015) methods. In many prior studies only radial velocity was assimilated due to simplicity or to isolate the impact on convective-scale forecasts to just a single observation type (e.g., Snyder and Zhang 2003; Zhang et al. 2004), but assimilating reflectivity has been shown to add value as well (e.g., Dowell et al. 2011; Dawson et al. 2015).

Recently, hybrid ensemble–variational (EnVar) methods have been developed that combine the best aspects of both a stand-alone ensemble-based method and a stand-alone variational method (Lorenc 2003; Buehner 2005; Wang et al. 2007; Wang 2010). Experiments with either global or coarse-resolution regional models assimilating real observations have exhibited improved forecasts (e.g., Wang et al. 2008a,b; Wang et al. 2013; Wang and Lei 2014; Pan et al. 2014). Studies experimenting with the EnVar approach to assimilate real radar observations at convection-allowing grid spacings are more limited (e.g., Carley 2012; Wang and Wang 2017). A recent study by Wang and Wang (2017) identified and discussed an issue of using the EnVar approach to directly assimilate radar reflectivity observations. The issue is attributed to the nonlinearity of the reflectivity operator and its tangent linear and adjoint of the nonlinear operator used by the variational method. They proposed a method to directly assimilate reflectivity observations in the EnVar approach; in this method, the reflectivity is directly added as a state variable. Therefore, no tangent linear or adjoint of the reflectivity operator is needed. They implemented this approach in the Gridpoint Statistical Interpolation (GSI)-based EnVar system and conducted experiments using the 8 May 2003 Oklahoma City tornadic supercell case. It was found that the new approach of directly assimilating reflectivity more realistically maintained the supercell during the entire 1-h forecast compared to other approaches.

Latent heat nudging (e.g., Leuenberger and Rossa 2007) is also a common approach to ingesting radar reflectivity observations into a model due to being computationally efficient relative to other more sophisticated techniques owing to absence of a need for a convection-allowing ensemble for scale-appropriate covariances. On the same branch of thought as latent heat nudging, a "cloud analysis" type of assimilation scheme—designed specifically to insert hydrometeors into the model grid while also inserting latent heating for dynamical balance—has also been developed and found to be useful in convection-allowing forecasts of severe storms compared to not ingesting reflectivity at all (e.g., Hu et al. 2006a,b; Hu and Xue 2007; Sugimoto et al. 2009; Schenkman et al. 2011; Snook et al. 2011). The cloud analysis and/or latent heat adjustment scheme is currently operational in mesoscale and high-resolution model forecasts such as the High-Resolution Rapid Refresh (HRRR; Benjamin et al. 2016; https://rapidrefresh.noaa.gov/hrrr/) and the 3-km North American Mesoscale Forecast System (NAM) CONUS nest and other operational centers in Europe (Gustafsson et al. 2018). In contrast to this type of latent heat nudging/cloud analysis scheme, other operational centers around the world instead convert measured radar reflectivity into relative humidity "pseudo-observations," which are then assimilated using a 3D- or 4DVar scheme. Compared to the simpler cloud analysis scheme detailed in section 2c(2), the direct reflectivity assimilation in the EnVar system is more rigorous since it incorporates ensemble-based flow-dependent error covariances and is therefore hypothesized to produce an analysis that yields a superior forecast relative to that produced by the cloud analysis scheme. This is consistent with prior work that has demonstrated the effectiveness of ensemble-derived background error covariances in the assimilation of radar observations to produce physically coherent storm-scale analyses (e.g., Snyder and Zhang 2003; Dowell et al. 2011; Marquis et al. 2014; Johnson et al. 2015).

In this paper convection-allowing NWP forecasts using an EnVar approach are compared to those using the cloud analysis scheme to study and understand the difference between a formal, more rigorous, but also more expensive radar DA technique and a less sophisticated but also less expensive technique. Given the cloud analysis scheme is currently operational within the National Weather Service models, and similar approaches are operational at other operational centers, this study also offers an opportunity to compare an experimental method with the existing operational method of assimilating radar reflectivity as a means of correcting convective-scale errors in convection-allowing NWP forecasts before the experimental method is considered for large-scale tests for future operational usage. The operational version of the cloud analysis scheme and a large CONUS-sized model domain close to that is used for operations are adopted to address the

TABLE 1. List of selected cases and radar DA windows. A brief description of the location and convective mode are also included.

Case	Radar DA window (UTC)	Location	Morphology
16 May 2015	2200–2300	Plains	Supercell outbreak
25 May 2015	1200–1300	Southern plains	Squall line
26 Jun 2015	0300–0400	Midwest	Mixed clusters of cells and lines
14 Jul 2015	1800–1900	Great Lakes and southeast United States	Two severe squall lines
11 Sep 2015	0000–0100	Central plains	Asymmetric MCS with MCV
22 May 2016	2200–2300	Plains	Supercell outbreak
17 Jun 2016	1900–2000	Midwest and southeast United States	Squall line and nonlinear convection
5–6 Jul 2016	0000–0100	Midwest	Long-lived, well-formed MCS
6–7 Jul 2016	2300–0000	Plains and Midwest	Large trailing-stratiform MCS
10 Jul 2016	0300–0400	Upper Midwest	Large MCS with MCV

above-posed scientific questions. A similar study was conducted by Bick et al. (2016) and revealed that assimilating reflectivity data using a sophisticated DA scheme can provide analyses and forecasts of convective precipitation that are competitive with an existing simple scheme used operationally. While the current study corroborates many of the broader findings of Bick et al. (2016), this study differs in that the test was conducted over a wider variety of synoptic, geographical, and seasonal regimes, and in the specific details of the cloud analysis and direct reflectivity assimilation through EnVar.

2. Experiment setup

a. Case selection

The two methods of assimilating radar reflectivity were compared using a collection of cases occurring in 2015–16. Cases were selected based on the following characteristics: variety of convective morphology, seasonality, diurnal cycle timing, geographical location, strength of large-scale forcing, and convective coverage. These characteristics were selected to test the robustness of the results over a variety of regimes, and cases with higher convective coverage were preferred to increase the amount of reflectivity to be assimilated. Cases featuring well-organized convective events, such as mesoscale convective systems and outbreaks of supercell thunderstorms, were particularly favored. A total of 10 cases were analyzed (Table 1). The temporal spacing of cases was also preferred to maximize the independence of samples.

b. NWP model

The nonhydrostatic multiscale model on the B-grid (NMMB; Janjić 2003; Janjić and Gall 2012) was used to conduct the experiments. The model and DA configuration used here very closely resembled that in the operational version 4 of the North American Mesoscale

Forecast System (NAM; Rogers et al. 2017) as of early 2018, with modifications described below to accommodate compute resources and experiment design. The CONUS-covering model domain is on a rotated latitude–longitude grid with a center point at 38.7°N, 97.0°W, or about 50 km southeast of Salina, Kansas, and has a size of $1568 \times 1120 \times 50$ grid points. The horizontal grid spacing is 0.0268°, which translates to about 1.9–2.6 km in the midlatitudes. The model vertical coordinate is hybrid sigma–pressure; the hybrid coordinate interface is at 300 hPa. The model top was at 50 hPa. Model output was regridded to a Lambert conformal grid with a grid spacing of 3 km for plotting and verification.

A diabatic digital filter initialization (DFI) process was used to initialize all model simulations following DA; in this case the twice DFI flavor was used (Peckham et al. 2016), as in the operational NAM. For the EnVar simulations the half-window length was 2 min, whereas it was 10 min for the cloud analysis simulations, which also matches that in the operational NAM. Initial tests revealed that with a 2-min half-window, insufficient vertical motions were produced to support hydrometeors injected by the cloud analysis, even when accounting for the latent heating. Johnson et al. (2015) noted similar issues when assimilating reflectivity using a 3DVar scheme. Sufficient vertical motion was generated using the 10-min window, but at the sacrifice of hydrometeor coverage and small-scale information (section 4). Initial testing revealed the EnVar simulations were not sensitive to the choice of half-window length, and 2 min was chosen for computational efficiency.

c. System description, methods, and experimental design

1) SYSTEM DESCRIPTION

The system was designed as follows (see also Fig. 1). The GSI-based EnVar DA system was constructed with
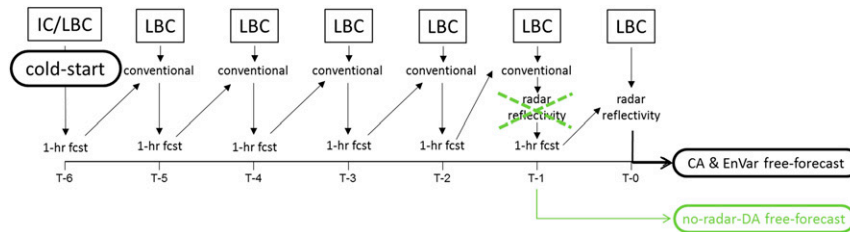
FIG. 1. Flowchart of experimental procedure. The green section denotes processes that were relevant to the ''no-radar-DA'' forecasts, where the radar reflectivity assimilation that occurs at $T - 1$ h in the CA and EnVar forecasts is skipped and the free forecast begins at $T - 1$ h.

40 ensemble members updated using the GSI-based EnKF and one control member updated using the EnVar system as in Wang et al. (2013; their Fig. 1b, and also described below). Both the GSI-based EnKF and GSI-based EnVar are extended to include direct radar reflectivity and radar radial velocity assimilation (Johnson et al. 2015; Wang and Wang 2017) and both were also used for conventional data assimilation. In this study, the initial conditions and lateral boundary conditions (ICs and LBCs) for the EnVar control were interpolated from the Global Forecast System (GFS) control analysis and its subsequent forecast. Perturbed ensemble member ICs and LBCs came from NCEP Global Ensemble Forecast System (GEFS; Zhou et al. 2017) and Short-Range Ensemble Forecast system (SREF; Du et al. 2015) members. Upon cycling, the full-resolution NMMB ensemble member fields initialized from its own GSI EnKF were used to calculate the flow-dependent background error covariances. This differs from the DA cycling in the operational NAM in which GFS ensemble members (much coarser; T574 resolution) provided the background error covariances for the EnVar DA. The members were cold started at $T - 6$ h (relative to the start of the free forecast). Conventional observations (i.e., surface observations from METARs, ship and ocean buoys, radiosondes, aircraft, and mesonet) were assimilated each hour for five cycles starting at $T - 5$ h, with the final conventional DA occurring at $T - 1$ h. After the final conventional DA cycle, only radar reflectivity was assimilated at $T - 1$ h and $T - 0$ h using the analysis after the final conventional DA as the first guess. The EnVar and cloud analysis methods diverged at this point, each using their respective methods to assimilate the radar reflectivity data through the final 1-h cycle. The choice to apply two reflectivity assimilation cycles was based on the desire to balance between revealing instantaneous differences and differences accumulated over multiple cycles of assimilation between the two methods.

The sole source of assimilated radar reflectivity was the three-dimensional reflectivity data produced through the Multi-Radar Multi-Sensor (MRMS) project (http://www.nssl.noaa.gov/projects/mrms; Lakshmanan and Humphrey 2014; Lakshmanan et al. 2006). MRMS reflectivity data are available on 33 vertical levels (250-m spacing between 500 and 3000 m MSL, every 500 m from 3000–9000 m MSL, and every 1000 m from 9000–19 000 m MSL) and have a grid spacing of 0.01° latitude–longitude. The exact temporal availability varies depending on the scanning times of the component WSR-88D radars that are mosaicked to produce the final product, but data are generally available every 2.5 min. MRMS data were interpolated to the NMMB model grid before being assimilated, but were not thinned, and were altered as described below. Following Aksoy et al. (2009), Yussouf et al. (2013), and Wang and Wang (2017), the observed reflectivity lower than or equal to 5 dBZ are set to 0 dBZ. Reflectivity observations greater than 10 dBZ and equal to 0 dBZ were assimilated with an observation error of 5 dBZ. Observations within 5 and 10 dBZ were not assimilated as the reflectivity within these thresholds likely is not able to distinguish precipitation from, for example, ground clutter, birds, and insects, especially when the observation is near the radar site (Aksoy et al. 2009). Consistent with prior studies involving the assimilation of radar reflectivity, we employ the assimilation of 0-dBZ observations as a means to suppress spurious convection (Tong and Xue 2005; Aksoy et al. 2009; Dowell et al. 2011). Similar treatment of small reflectivity values is also applied to the control and ensemble first guess to be consistent with the treatment of the observations. Equivalent horizontal and vertical localization scales were applied to the EnVar and EnKF here; for the assimilation of conventional (radar) observations, cutoff distances of 500 (12) km in the horizontal and of 1.1 (0.55) scale heights in the vertical were used. To account for the deficiency of the spread of the first guess ensemble in the EnKF, the relaxation to prior spread (RTPS; Whitaker and Hamill 2012) inflation method with a coefficient of 0.95 was applied to relax the posterior ensemble perturbation spread back to 95% of the prior ensemble spread.

Free forecasts were launched six hours after the cold start in each case. They were integrated 18 h to cover the maturity and dissipation of the convective events of interest. All NWP simulations, including the control and 40-member ensemble background forecasts, and the free forecasts, were run on the same computational domain using the same convection allowing grid spacing (1.9–2.6 km).

2) ENVAR METHOD FOR DIRECT RADAR REFLECTIVITY ASSIMILATION

The EnVar method follows Wang and Wang (2017) in which the reflectivity observations are directly assimilated as part of the variational update, rather than in a separate step as in the current operational cloud analysis. Wang and Wang (2017) discussed three approaches to directly assimilate reflectivity in EnVar. The first approach uses hydrometeor mixing ratio as the state variable and a nonlinear reflectivity operator to relate the hydrometeor mixing ratio to the reflectivity. Therefore, the tangent linear and adjoint of the nonlinear operator are used during the EnVar minimization. It was found that when the hydrometeor mixing ratio was small, the large values of the tangent linear of the reflectivity led to large differences of the cost function gradients among different observed variables. Such large differences prevent efficient convergence of the cost function and therefore led to underestimations of the hydrometeor increments. Though the second approach— where the logarithm-transformed hydrometeor mixing ratio is used as state variable—can fix this cost function gradient issue, this approach causes additional issues associated with the forward and backward transform between the logarithm and nonlogarithm spaces. Such issues led to anomalously large hydrometeor increments. Wang and Wang (2017) also showed the underestimation and overestimation of the hydrometeor increments are exacerbated by the tangent linear assumption of the nonlinear reflectivity operator itself.

In this paper, the third approach proposed by Wang and Wang (2017) was adopted. In this method, the reflectivity is directly included as a state variable. Therefore, the observation operator for reflectivity is an identity matrix, thus avoiding the need for an explicit nonlinear operator to transform the prognostic hydrometeors into reflectivity during the variational minimization; the associated tangent linear and adjoint formulation of the reflectivity operator are unnecessary and do not exist in the EnVar cost function. This approach fixes the cost function gradient issue seen in the first approach. This approach of assimilating radar reflectivity is considered more rigorous when compared to the cloud analysis approach since

the microphysical state variables and other dynamic and thermodynamic variables are updated more consistently using the cross-variable correlations in the background ensemble covariance matrix. Since reflectivity in particular tends to be more strongly correlated with intense vertical motion compared to other basic model fields such as the wind and temperature fields, this cross-variable correlation should promote convective development in areas where real-world reflectivity corresponding to existing convection is sampled in the absence of hydrometeors in the forecast model. In comparison, the cloud analysis inserts hydrometeors into the model grid based on empirical relationships between reflectivity and hydrometeors in addition to inserting a separate latent heating to introduce convective-scale motions consistent with observed deep convection. While the EnVar method can reduce spurious convection by assimilating 0-dBZ reflectivity, the cloud analysis was generally not configured in a way to do so, consistent with how it was implemented in the operational NAM.

3) CLOUD ANALYSIS METHOD

The cloud analysis is heavily detailed in Hu et al. (2006a,b), although the version used here was modified for the operational NAM and is slightly different. Briefly, the cloud analysis inserts condensate, including hydrometeors, into the model grid by inverting the radar reflectivity equations associated with the microphysics scheme in the model, the Ferrier–Aligo scheme (Aligo et al. 2018). Therefore, both rain and snow mixing ratios were directly altered by the observation reflectivity. The cloud analysis also inserts cloud water and ice using a choice of schemes—either designed for stratiform or cumulus clouds. To remain consistent with the NAM configuration, the former option was chosen. Finally, a latent heating scheme is used to reinforce the presence of deep convection and to prevent injected condensate from immediately evaporating/sublimating, which would otherwise render the hydrometeor retrieval useless.

A detailed version of the cloud analysis procedure follows. The cloud analysis is capable of ingesting not only radar reflectivity, but also lightning data, METAR surface observations including ceiling heights, and satellite data. However, in this study, the focus is to understand the difference of the two methods in assimilating only the radar reflectivity observations. To make a fair comparison, only radar reflectivity was assimilated. The procedures to assimilate reflectivity using the cloud analysis follow closely the operational version in NAM. In grid boxes in which ingested reflectivity exceeded 35 dBZ, cloud material (water and ice) were eligible to be altered. Otherwise nothing was done to

add cloud material at that grid point. Hydrometeor content is not impacted by this particular setting. Cloud material was retrieved from the background fields using the in-cloud saturation vapor mixing ratio; if the in-cloud temperature was above $-5°C$ (268.13 K) all available water was assigned to the liquid phase, whereas where in-cloud temperature was below $-10°C$ (263.15 K) all available water was assigned to the ice phase. Between $-10°$ and $-5°C$ the available water was partitioned between the liquid and solid phases according to the relative distance along that 5°C window, with more water going to the liquid phase when the in-cloud temperature was closer to $-5°C$ and vice versa. An arbitrary and fixed cloud autoconversion rate was used to determine the total amount of available water.

Rain and snow mixing ratios were retrieved by directly inverting the reflectivity equations for those species consistent with the Ferrier–Aligo microphysics scheme. Hydrometeors were first retrieved only where reflectivity observations exceed 0 dBZ (a threshold employed to discriminate between precipitating, >0 dBZ, and nonprecipitating, ≤0 dBZ, echoes). Final hydrometeor mixing ratios were then determined by comparing retrieved values of the sum of rain and snow to the background values at the level of the maximum observed reflectivity. The following rules are used if the retrieved mixing ratio sum was less than the background sum:

- Calculate RATIO, the ratio of retrieved hydrometeor sum to the background hydrometeor sum (a lower bound of 0.8 is enforced).
- For rain: if the background mixing ratio is greater than $1.2\,\mathrm{g\,kg}^{-1}$ then the analyzed mixing ratio is set to the maximum of the following two quantities: 1) retrieved mixing ratio (using an upper bound of $3.0\,\mathrm{g\,kg}^{-1}$), 2) RATIO × the background mixing ratio. Otherwise the background mixing ratio is used. In other words, with some restrictions, if there is nonzero rain retrieved from the observed reflectivity and it is less than the background value, the background value is reduced, potentially to as low as the retrieved value.
- For snow: same as rain, except an initial threshold mixing ratio of $1.0\,\mathrm{g\,kg}^{-1}$ is used instead of $1.2\,\mathrm{g\,kg}^{-1}$.

Otherwise (the retrieved sum was larger than the background sum), the retrieved mixing ratios were used unless the background mixing ratio was less than $3.0\,\mathrm{g\,kg}^{-1}$ and the retrieved mixing ratio was greater than $3.0\,\mathrm{g\,kg}^{-1}$ in which case the analyzed mixing ratio value was set to $3.0\,\mathrm{g\,kg}^{-1}$. If a state of "zero hydrometeors" was retrieved (i.e., valid reflectivity observations of < 0 dBZ) then the background hydrometeor mixing ratios were reduced by 20%.

Vapor mixing ratio was adjusted after all condensate species have been adjusted. A series of rules are used to adjust the final vapor mixing ratio:

- If a layer was cloud free
  - If liquid condensate was present and the layer was more than 100 mb above the surface, RH was reduced to 80%.
  - If there was cloud immediately above or below the layer and the layer itself was subsaturated, 70% of the saturation deficit was added.
- If there was cloud present in a layer, add 50% of the deficit between the background mixing ratio and that at 102% RH.

The temperature tendency was calculated assuming an arbitrary amount of added hydrometeor content associated with the observed reflectivity value and an assumed storm lifetime time scale, which is linked to the DFI half-window length. The final temperature tendency has an upper bound of $2.5 \times 10^{-3}\,\mathrm{K\,s}^{-1}$. The temperature tendency was only applied during the DFI.

4) EXPERIMENTAL DESIGN

Three experiments were designed in this study. The EnVar and cloud analysis introduced in sections 2c(2) and 2c(3) were employed for direct and indirect reflectivity assimilation experiments, respectively. As a benchmark comparison for the overall utility of radar DA, a separate "no-radar-DA" forecast was initialized at $T - 1$ h using the same postconventional-DA analysis used by the EnVar and cloud analysis methods. The no-radar-DA forecast was integrated for 19 h to cover the same valid time range as the other two methods. All forecasts used twice DFI; for the EnVar and cloud analysis methods, DFI was applied at the start of the final background cycle as well as at the start of the free forecast. For the no-radar-DA forecast, the DFI was applied only at the beginning of the free forecast (at $T - 1$ h). The half-window length of the DFI in the no-radar-DA forecast was 10 min, matching that in the cloud analysis experiment.

d. Verification

A single domain with bounds at 110°–80°W longitude and 25°–50°N latitude, or roughly the eastern two-thirds of the CONUS was used for verification, where areal coverage of observation used in this study are nearly complete. Since high-resolution, convection-allowing NWP model forecasts are capable of at least partially resolving convective-scale circulations, they often produce forecasts with considerable, realistic detail. Owing to the highly structured nature of these fields (e.g., simulated radar reflectivity) traditional verification metrics

TABLE 2. Relevant settings in MODE.

Setting[a]	Composite reflectivity[b]	Precipitation[c]
Convolution radius	5 grid points (15 km)	8.33 grid points (25 km)
Convolution threshold	30 dBZ	5.0 mm
Fuzzy engine weights		
Centroid distance	5.0	2.0
Boundary distance	4.0	4.0
Convex hull distance	0.0	0.5
Angle difference	3.0	2.0
Area ratio	5.0	3.0
Complexity ratio	0.5	0.5
Intersection over area ratio	2.0	1.0
Percentile intensity ratio (percentile value)	0.5 (95th)	0.5 (75th)
Matching interest threshold	0.7	0.7

[a] See section 14.3.2 of the MET v6.0 User's Guide at https://dtcenter.org/met/users/docs/users_guide/MET_Users_Guide_v6.0.pdf, for a description of the settings.
[b] Composite reflectivity objects with an area less than 16 grid squares were removed from consideration.
[c] Precipitation objects with an area less than 36 grid squares were removed from consideration.

like bias and RMSE generally do not offer an adequate measure of forecast quality compared to subjective judgment (Davis et al. 2006, Wolff et al. 2014). Other verification methods are applicable (Gilleland et al. 2009), and two such methods used here are based upon neighborhood and features-based methods. Circular neighborhoods of radius 50 km (Johnson and Wang 2012; Schwartz and Liu 2014; Duda et al. 2014) were applied and used in the fractions skill score (FSS; Roberts and Lean 2008) and the neighborhood equitable threat score (NETS; Clark et al. 2010; Schwartz 2017).

For the features-based method, the Method of Object-based Diagnostic Evaluation (MODE; Davis et al. 2006) was used. MODE creates objects from a pair of input gridded datasets and compares the features-based attributes between pairs of objects in each input grid. Typically, a forecast and observation dataset valid at the same time are input, but any two gridded datasets can be compared, including two forecasts to each other. MODE has been used in prior studies to verify convection-allowing forecasts (e.g., Johnson and Wang 2013; Duda and Gallus 2013; Cai and Dumais 2015) and is an ideal utility for extracting features-based information from gridded fields, including parameters useful for describing storm cells and precipitation regions such as centroid location, area, and shape. MODE attempts to match corresponding objects in the two fields by using the attribute values between pairs of objects to calculate an interest value between 0.0 and 1.0, with 1.0 implying

perfect correspondence between the two objects and 0.0 implying no meaningful correspondence. The interest value is sensitive to the configuration settings, so a judicious choice of settings is necessary to obtain meaningful results. There are configuration settings for defining objects and for weighting the object attributes when calculating the interest value, which allows the user strong control over which aspects of the forecast to emphasize when evaluating them. Here, MODE was configured to best identify smaller-scale features like individual storms and small bands of precipitation, thus allowing a high degree of object detail to be measured (Table 2). A possible disadvantage to these settings is that pairs of larger objects (corresponding to squall lines and their associated precipitation, for example) may be given low interest values due to differences in small-scale details internal to the objects when the objects themselves are similar on coarser scales. However, if verification of storm-scale detail is desired, such a disadvantage is not likely to severely impair analysis of MODE output.

Two fields were verified in this study: simulated composite reflectivity (column maximum) and 1-h accumulated precipitation. Both observation datasets were obtained from the MRMS project. Statistical significance testing was applied to the standard and neighborhood verification metrics using the bootstrapping method of Hamill (1999) and using 1000 resamples. A two-tailed test using a significance level of 0.95 was applied, meaning differences more extreme than the 97.5th or 2.5th percentile of the null distribution were considered significant.

3. Results

Both subjective and objective verification (traditional and features-based) support the claim that the EnVar forecasts are improved over those initialized from the cloud analysis (hereafter, CA), and also that the cloud analysis offers limited improvement over no radar data assimilation for the first few hours (no-radar-DA). Verification was also performed on two groups of five cases each separated by strength of forcing. Little difference in verification outcomes was seen between the two strength-of-forcing groups, likely a result of both small sample size and overlap of moderately forced cases existing in both groups (albeit different cases). One minor difference was noted between the CA and no-radar-DA forecasts in that there was a larger separation between the two for neighborhood-based metrics over the first 3–6 forecast hours in the weakly forced cases. The cause seemed to be poorer performance of the no-radar-DA forecast, as the magnitude of the

scores themselves did not differ appreciably for the CA forecasts between the strength-of-forcing groups.

a. Quantitative verification

1) TRADITIONAL METRICS

Examination of both mean error (not shown) and frequency bias (Fig. 2) suggests a tendency for all the CA, EnVar, and no-radar-DA forecasts to contain too much precipitation and reflectivity in the early forecast hours (through the first 3–9 forecast hours depending on field and threshold). With the notable exception of forecast hours 6–8 at the 0.254 mm h^{-1} precipitation threshold in the CA forecast (Fig. 2b), both the CA and EnVar forecasts generally failed to outperform the no-radar-DA forecasts in terms of frequency bias. In fact, the no-radar-DA forecast often statistically significantly outperformed the CA forecast in terms of convective coverage during the first 6 h of the forecast. The difference between no-radar-DA and EnVar is often statistically insignificant for this period (not shown). While the no-radar-DA forecasts had one extra hour of spinup, both radar DA methods struggle to sufficiently suppress spurious convection (section 3b). But suppression of spurious convection through data assimilation cannot be the only factor contributing to the positive coverage bias since the frequency biases also increase from forecast hour 0 to 2 in the no-radar-DA free forecast, as well. This behavior could represent the need to better constrain the storm environment; there could also be some model shock despite the use of a 2-min DFI window that is factoring into the apparent spinup in the first couple forecast hours.

At early lead times, frequency biases for 1-h precipitation largely follow the pattern as in composite reflectivity. However, there is a difference in character of the frequency biases for the EnVar forecasts at the 0.254 and 2.54 mm h^{-1} thresholds compared to at the 20- and 30-dBZ reflectivity thresholds. In the former, the frequency bias is around 0.9 at forecast hour 0, jumping above 1.0 afterward before decreasing, whereas in the latter, the frequency biases are around 1.4–1.6 at forecast hour 0 and drop quickly afterward. This may be explained by the way the EnVar method handles thunderstorm anvils, which are readily delineated in composite reflectivity (but not precipitation). The EnVar technique maintained anvil reflectivity through the DA cycles, whereas it tended to be lost through the DFI in the cloud analysis procedure (see section 4). At light thresholds such as 20 dBZ, the EnVar reflectivity frequency bias is statistically significantly better than CA for the first 7 h except hour 0. Because of the switching of sign (frequency bias > 1.0 indicates too

much coverage and vice versa), EnVar precipitation forecasts at light thresholds such as 0.254 mm h^{-1} are only statistically better than CA in the first 2–4 forecast hours except at forecast hour 0, whereas the CA forecasts were better from forecast hours 5–8. There were fewer statistically significant differences between the CA and EnVar forecasts at the higher thresholds for both reflectivity and precipitation forecasts.

Both the frequency biases and mean errors tend to decrease over the length of the forecast, switching sign around forecast hours 6–9, suggesting the forecasts are not able to sustain observed convection sufficiently long into the forecast. All forecasts tend to behave very similarly especially after about forecast hour 12 (Fig. 2). This behavior suggests a transition from initial condition error to model error dominance as the forecast draws on.

2) NEIGHBORHOOD METRICS

FSS values are shown in Fig. 3. There is some noise in the scores due to the use of only 10 cases. However, the tendency for the EnVar forecast to outperform the CA forecast is clear. The EnVar forecast is significantly superior through almost the entire 18 forecast hours for simulated reflectivity except the middle of the forecast period. A similar trend is found for the precipitation forecasts. For both the simulated reflectivity and precipitation forecasts, the improvement of EnVar is more persistent over time at the heaviest precipitation threshold. There is also a difference in the character of the FSSs in the first few forecast hours in both fields at the light and moderate thresholds. The EnVar forecast shows the largest improvement at hour 0. There is a decrease from a maximum FSS value in the EnVar forecast from hour 0, whereas FSSs start from a low value at hour 0 and increase in the CA forecast after hour 0, but reach a lower peak value than in the EnVar forecast at around forecast hour 5 before decreasing as the forecast ages. The large improvement in the EnVar forecasts at the early lead times suggests the superior update of hydrometeors through direct assimilation of reflectivity observations by the EnVar. The low FSS values at hour 0 in the CA forecasts suggest the update of hydrometeors is not as effective using this method. The increase in FSS in the first few hours of the CA forecasts is consistent with its reliance on model integration to further spinup the hydrometeor fields. The persistent improvement by EnVar beyond the first several hours is also consistent with expectations that the EnVar method of assimilating reflectivity adjusts more than just the hydrometeor fields; it also adjusts the thermodynamic and dynamical state in a consistent manner via the ensemble cross-variable correlations, which allows for benefits to persist longer into the forecast. While other fields are
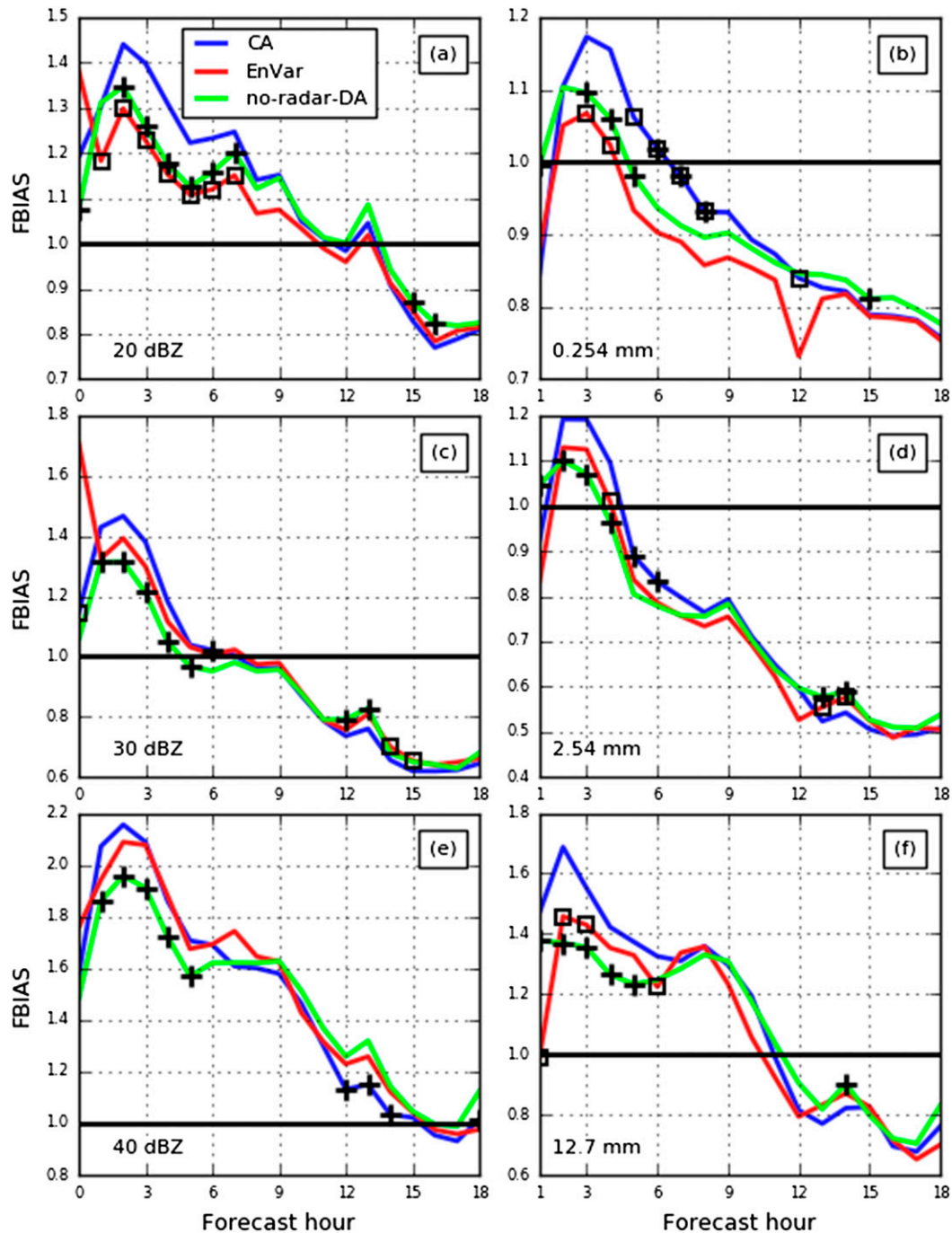
FIG. 2. Frequency bias over all cases for (a),(c),(e) composite reflectivity at 20, 30, and 40 dBZ, and (b),(d),(f) 1-h precipitation at 0.254, 2.54, and 12.7 mm. Statistically significant differences between pairs of forecasts are indicated using squares for the CA-vs-EnVar and plus signs for the no-radar-DA-vs-CA comparisons. The symbols mark the superior forecast. The thick black line delineates a perfect frequency bias value of 1.0.

also adjusted via the additional DFI step in the cloud analysis procedure, the results suggest the cross-variable update by the EnVar is more effective. These hypotheses are further examined through subjective verification discussed in section 3b.

The CA forecast performed significantly better than the no-radar-DA forecast up to forecast hour 5 at the 20-dBZ and $0.254\,\mathrm{mm\,h^{-1}}$ simulated reflectivity and precipitation thresholds (Figs. 3c,d). At higher thresholds in both fields, the CA forecast was only better than the
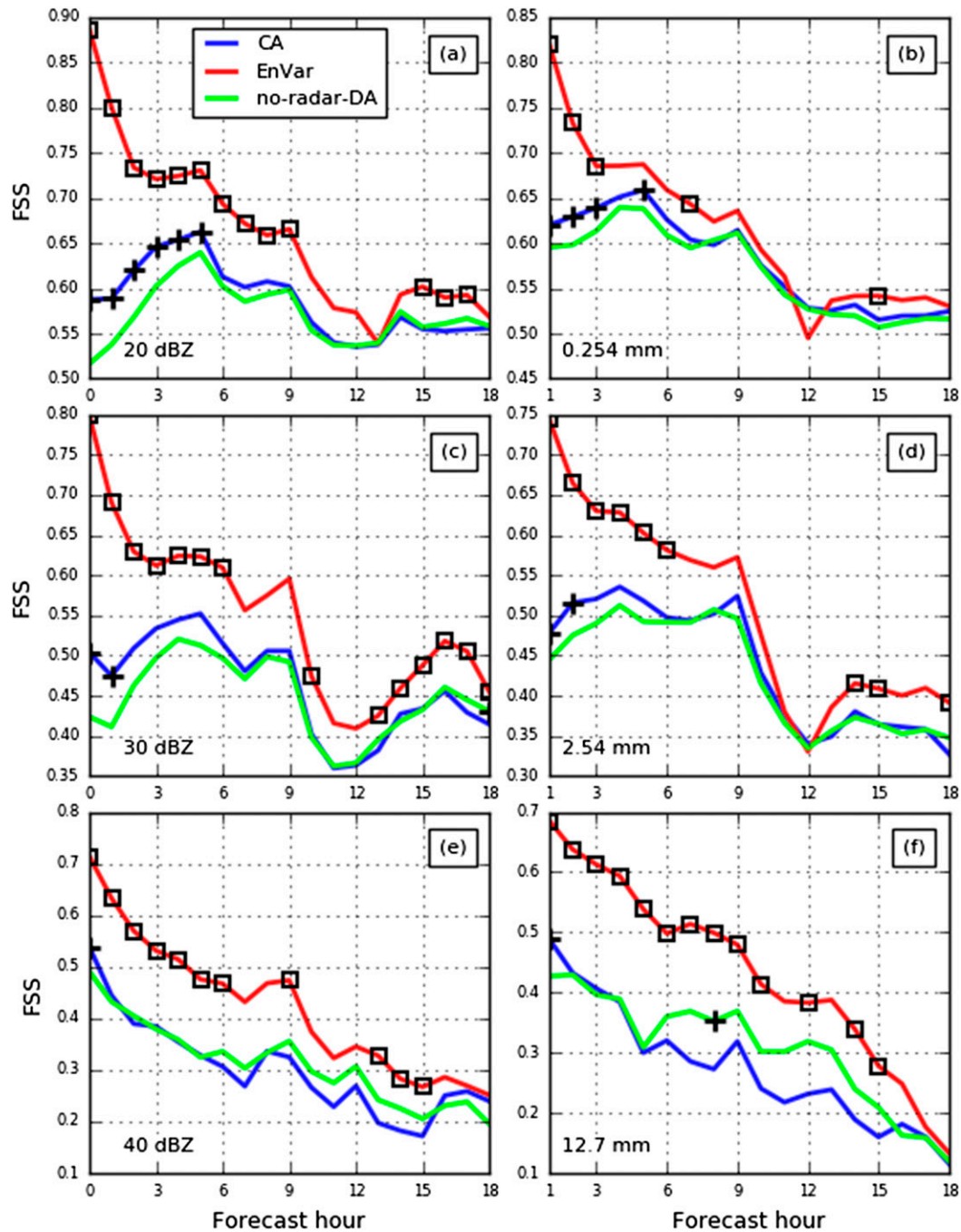
FIG. 3. As in Fig. 2, but for FSS.

no-radar-DA forecast for up to 2 h (Fig. 3f), indicating little, if any, forecast improvement by this method over not assimilating reflectivity, except perhaps at analysis time. Beyond that, the two forecasts performed similarly, although there was a tendency for the CA forecasts to be inferior at the 12.7 mm h^{-1} precipitation threshold during forecast hours 6–15 (Fig. 3f).

NETSs are shown in Fig. 4. While NETS and FSS use a similar methodology (examining a sense of spatial closeness between the forecast and verifying field), they assess different aspects of the forecast; the FSS is more strict, requiring a similar match in the areal coverage of the event between the forecast and observations for a good score, whereas the NETS is more permissive, only

requiring a single event in the forecast to occur any-where within the neighborhood of a grid point in the observation field to constitute a hit [Schwartz (2017) and Schwartz and Sobash (2017) provide thorough commentary on this distinction]. The relative performance among the EnVar, CA, and no-radar-DA forecasts measured by NETS are in general similar to the FSS. The EnVar forecast is significantly better than the CA forecast for the first 8–9 forecast hours in both fields and at a few extra hours at later lead times. The difference in NETS between the CA and no-radar-DA forecasts is only significant in the first few forecast hours, or only at forecast hour 1 in the case of precipitation the 12.7 mm h^{-1} threshold (Fig. 4f). This means the CA does offer an improvement in the early stages of the forecast, but the effects wear off quickly, as scores are very similar to the no-radar-DA experiment after about forecast hour 4.

3) FEATURES-BASED METRICS

A number of custom metrics were calculated from MODE output. One advantage of using features-based metrics to assess the quality of high-resolution forecasts is the inherent image matching characteristic: unimportant smaller-scale aspects of the forecast and observation fields are smoothed and blurred so that strict gridpoint-to-gridpoint quantitative criteria are not as important as in traditional metrics. Therefore, features-based verification techniques should give a better representation of what a human evaluation would produce. Since the most important objects to focus on in this study include individual thunderstorms (supercells) and intense convective segments (small line segments or squall lines), a features-based technique should provide a more representative assessment of forecast quality than the metrics previously presented.

One useful features-based metric is the object-based threat score (OTS; Johnson and Wang 2013), which ranges from 0 to 1 (perfect). The OTS combines the area of intersection of forecast–observation object pairs with the strength of the correspondence between those pairs. Nonintersecting forecast–observation object pairs are also accounted for. OTSs for both reflectivity and precipitation are shown in Fig. 5. No significance testing was performed on OTS. The OTSs generally agree with the traditional and neighborhood metrics in that the EnVar forecasts are better than the CA forecasts, especially at the analysis time (forecast hour 0), although there are forecast times when this is not true, such as forecast hour 5 and 10 for composite reflectivity. For 1-h precipitation, the EnVar OTS values are always higher than those of the CA forecast, although the differences are occasionally very small. The OTS values are rather noisy, likely due not only to the use of 10 cases, but also due to

low convective activity in parts of the domain and at some forecast times, especially later forecast hours. As the cases focused generally on mesoscale convective systems or supercell outbreaks, by late in the forecast those convective systems had generally matured and were dissipating, thus leaving little convective activity in the model domain. The OTS is sensitive to the total number and size of objects created by MODE. Although the differences in the OTS values may not be statistically significant, the generally larger OTSs of the EnVar forecasts reinforce the results measured by the neighborhood-based skill scores. OTS differences between the CA and no-radar-DA forecasts are generally small, especially for 1-h precipitation. However, there is some indication that simulated reflectivity is better in the CA forecast at hours 0 and 1.

Forecast quality can also be examined via the phase diagram shown in Fig. 6. The phase diagram compares the total intersection area (summed over all forecast–observation objects in all cases) to the total symmetric difference area, which is the sum of both forecast false alarm objects and missed observation objects. In a perfect forecast, the intersection area is maximized and the symmetric difference area is 0.0, so improved forecasts lay closer to the lower right in the diagram. One major observation that is apparent is that the total symmetric difference area is generally at least one order of magnitude larger than the total intersection area, implying a tendency for many inaccurate forecast objects (either falsely predicted or missed). Given the frequency biases larger than 1.0, especially during the first several hours, much of the symmetric difference is due to forecast false alarms from spurious convection in the initial condition. The improved reflectivity assimilation in the EnVar method is clearly shown by the outlier point at forecast hour 0. After that, the curves for the different forecast methods are much closer and converge after about forecast hour 10. Before then, the EnVar forecasts have more intersection area and about the same amount of symmetric difference area as the other two forecasts, with the CA method having slightly more intersection area than the no-radar-DA forecasts but also slightly more symmetric difference area (again, due to a larger coverage bias). For 1-h precipitation, the superiority of the EnVar forecasts is much more apparent and persists longer into the forecast, whereas the CA and no-radar-DA forecasts are very similar to each other, consistent with the OTS results.

Finally, as a measure of strictly location accuracy, the average centroid distance between matched forecast–observation object pairs is shown in Fig. 7. Matched pairs had an interest value of 0.7 or greater,
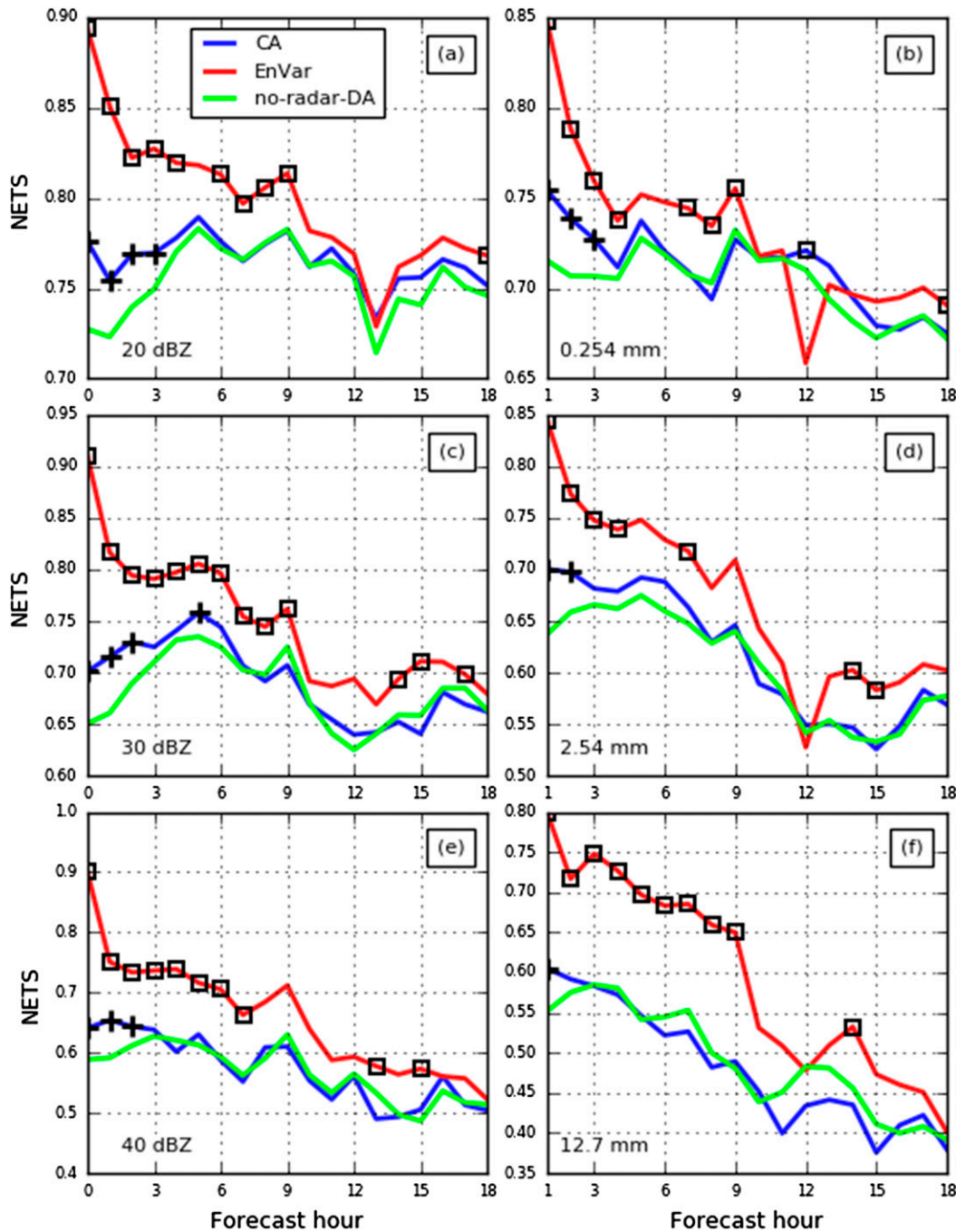
FIG. 4. As in Fig. 2, but for NETS.

and a given forecast or observation object was only counted once, as in the OTS computation. Given the large total weight for distance attributes (Table 2), it is unlikely that an unmatched forecast–object pair would have a small centroid distance, so discounting unmatched objects should not impact the mean centroid distance meaningfully. Much like the other

features-based scores, the mean centroid distances are noisy after the first few forecast hours when the number of eligible matches has decreased, consistent with the decrease in object counts. However, similar tendencies as with other verification metrics are also seen in the first few hours for mean centroid distances; the EnVar forecasts have lower distance error for both simulated
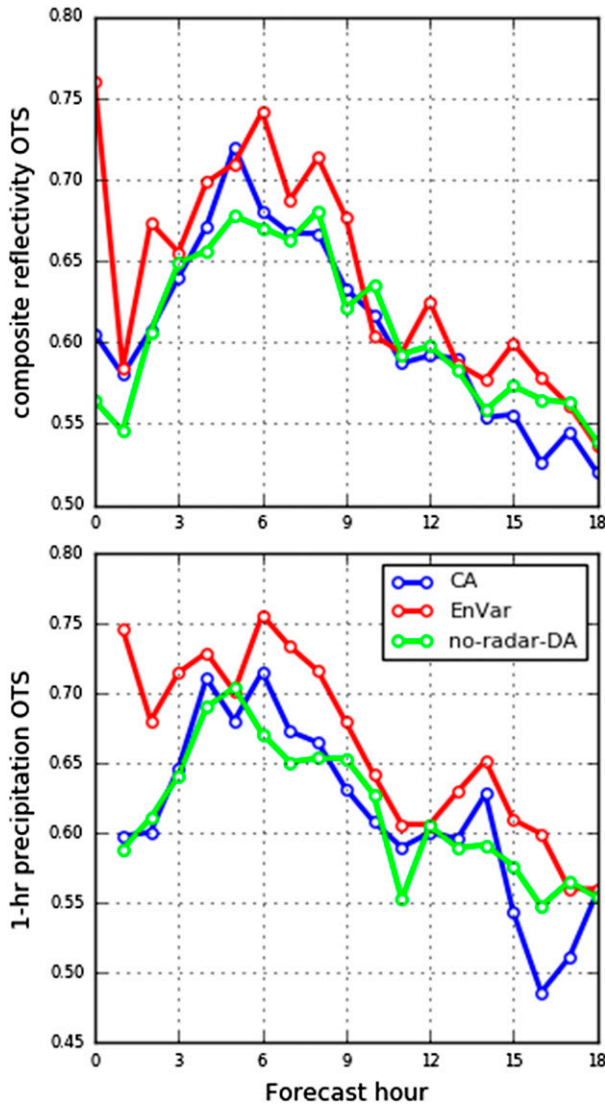
FIG. 5. Object-based threat score integrated over all cases for (top) composite reflectivity and (bottom) 1-h precipitation.
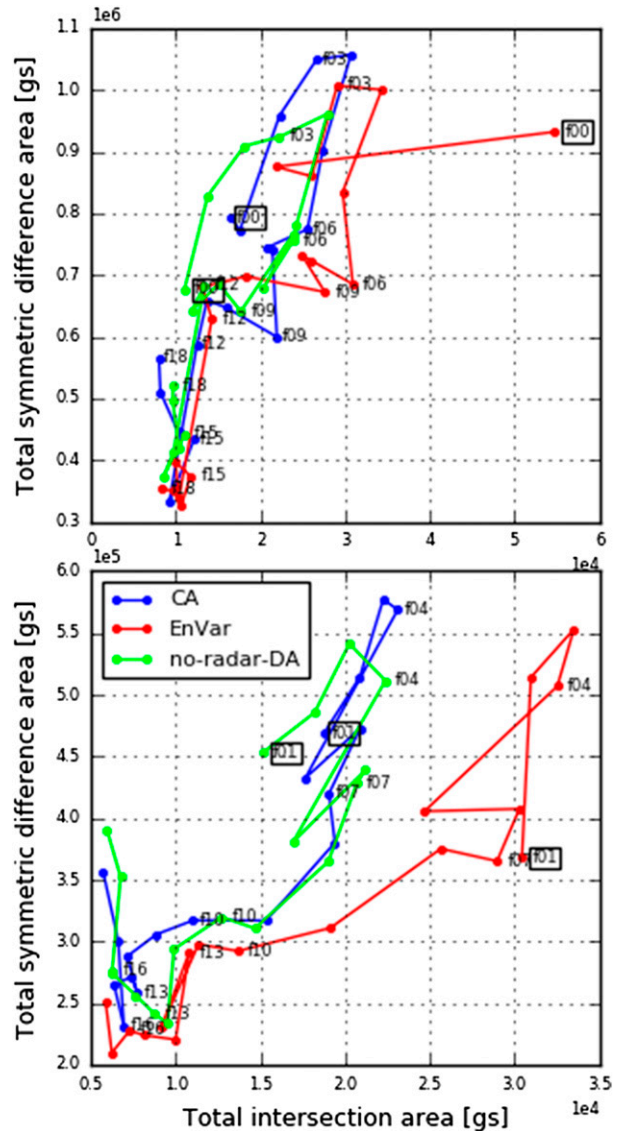

FIG. 6. MODE object area phase diagram for (top) composite reflectivity and (bottom) 1-h precipitation. Units of area are grid squares (gs) (given $\Delta x = \Delta y = 3$ km, then 1 gs = 9 km^2) and areas are integrated over all cases. Data points for select forecast hours are marked for each DA method. The scientific notation exponents for the vertical axis are located on the upper left of each panel, while those for the horizontal axis are located on the bottom right of each panel.

reflectivity and precipitation, and CA forecasts have lower distance error than no-radar-DA forecasts. However, these differences are generally absent after the first forecast hour. All forecasts tended to produce too many forecast objects (i.e., storms), especially up to forecast hour 5, again consistent with the frequency biases.

b. Subjective evaluation through case studies

Strong evidence supporting the EnVar method provides superior forecasts of composite reflectivity and 1-h precipitation for the 10 cases was shown in section 3a. Specific attributes responsible for the improved analyses and forecasts are discussed below, as is a notable weakness in the cloud analysis procedure. In general, it was found that the cloud analysis struggles not only to

suppress spurious convection more so than does the EnVar procedure (which itself could be better), but it also does not adequately assimilate convection missing from the first guess. It was also found that through the inherent cross-variable update, other dynamic and thermodynamic variables are more consistently updated by the EnVar although only reflectivity is assimilated. Although the additional DFI step in the cloud analysis procedure can to some extent help to adjust other
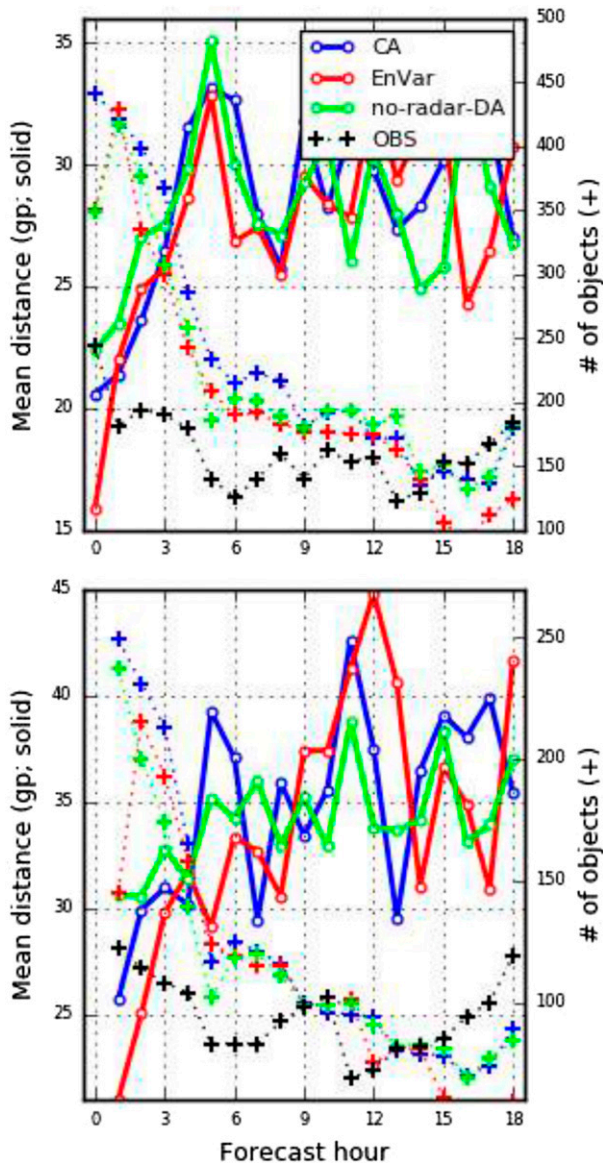
FIG. 7. Mean distance between centroids of matched objects (solid with dots) and number of objects (dotted lines with plus symbols) for the various methods (forecast objects for CA, EnVar, and no-radar-DA, and observation objects for OBS) for (top) composite reflectivity and (bottom) 1-h precipitation. Mean centroid distance units are in model domain grid points (gp) ($\Delta x = 3$ km gp^{-1}).

variables, a lot of weak and moderate assimilated reflectivity coverage is nonetheless lost during the DFI procedure. But reducing the DFI half-window length to maintain more of the simulated reflectivity results in insufficient vertical motion (no updrafts) to support maintenance of assimilated reflectivity once the model begins integrating. It was also determined that the quality of the analysis after the final cycle of conventional observation assimilation (referred to in this

section simply as the "background") largely controlled the quality of the CA and no-radar-DA forecasts, and to a lesser extent the EnVar forecast.

Robust deep convection in the area of observed intense convection was largely absent from the background in 5 of the 10 cases studied. Specific cases that illustrate these issues are presented below.

1) 6 JULY 2016

This case featured two organized convective systems; one longer-lived mesoalpha-scale convective system featuring a leading-line/trailing-stratiform morphology across the northern plains and Midwest (Schumacher and Johnson 2005) and a shorter-lived mesobeta-scale convective cluster over the lower Ohio River valley (not shown). The plains MCS evolved slowly upscale during peak diurnal heating from a widely scattered set of supercells and line segments over the northern High Plains while the smaller MCS developed in the wake of a previous MCS. The plains MCS was strongly forced by a midlevel shortwave trough prompting a surface trough in the lee of the Rocky Mountains and associated low-level warm-air advection, whereas the Ohio River valley MCS formed amid weak low-level warm-air advection.

The background (valid at 2300 UTC 6 July) included significant spurious convection across eastern Nebraska and across Iowa among other areas (Fig. 8). The cloud analysis made little to no changes to this convection (cf. Figs. 9a,b and Fig. 8), whereas the EnVar was able to reduce the intensity of the convection through assimilation of 0-dBZ reflectivity (cf. Figs. 9c,d and Fig. 8). However, well-established convective circulations, although reduced, still remain even in the EnVar analysis, which illustrates that complete suppression of spurious convection remains a challenging problem in storm-scale DA. Because of the incorrectly favorable mesoscale environment for sustained convection in this area, the storms persisted and even reintensified in both background cycles, although the convection was arguably more intense and widespread in the CA first guess for the final radar DA cycle at 0000 UTC 7 July compared to the EnVar first guess (not shown). However, the EnVar reduced the extent of the convection, whereas the cloud analysis did essentially nothing to suppress this convection, and it carried into the free forecast initial condition.

The spurious convection in the initial condition significantly disrupted the thermodynamic environment in the CA free forecast. A large area with low instability and significant inhibition (not shown) developed in eastern Nebraska and was advected northwestward by low-level winds through the inflow region of where
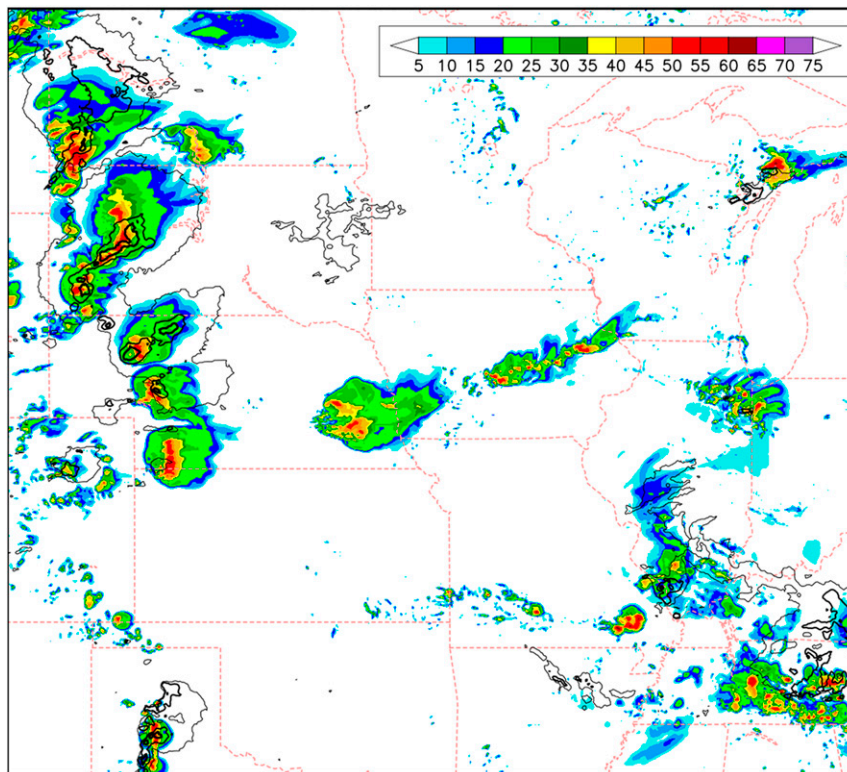
FIG. 8. Composite reflectivity (dBZ) in the background (analysis after final cycle of conventional DA) at 2300 UTC 6 Jul 2016. The light, medium, and thick black contours delineate observed composite reflectivity at 10, 30, and 50 dBZ, respectively.

observed storms would form shortly after initialization of the free forecast. Even though much of the observed convection across the High Plains was present in the final radar DA analysis for the CA experiment, this convection moved through the area that had been rendered thermodynamically unfavorable for sustained convection in the first few hours of the free forecast and failed to evolve correctly. The resulting 7-h forecast (around the time of maturity of the observed MCS) contained unacceptable mesoscale errors in the morphology of the MCS compared to the EnVar free forecast, which successfully predicted the location, size, and shape of the observed system (Fig. 10), even though there was also a small region that was also rendered thermodynamically unfavorable for convection in the EnVar experiment as well. The EnVar did not completely suppress this spurious convection, which is why the instability gap was still present. But it was not nearly as large as that in the CA experiment and evidently was not enough to interfere with the upscale growth of the existing cells on the High Plains into the observed MCS. Objective verification for this case also showed the EnVar forecast has better scores than the CA forecast (not shown), suggesting the less extent of

the spurious convections of the EnVar gets translated into a better skill score.

2) 17 JUNE 2016

This case also featured two MCSs, both weakly forced, and in different regions. One occurred across the southeastern United States—in northern Mississippi, Alabama, and Georgia—and pushed southward. The other developed shortly later in Minnesota and moved in an unusual direction—to the southwest. Other weakly forced and disorganized convection developed after the radar DA period in this forecast over the High Plains of northwestern Kansas and southwestern Nebraska but expanded and persisted long enough to interact with the midwestern MCS (not shown).

The background (valid at 1900 UTC) successfully captured the southeastern U.S. squall line, whereas it almost completely missed the early stages of the burgeoning MCS in Minnesota (Fig. 11). The background also contained widely scattered spurious storms across the plains states. As in the 6 July 2016 case, the EnVar was able to partially suppress these storms, whereas the CA did essentially nothing to suppress them (Fig. 12). However, the mesoscale environment was strongly supportive of
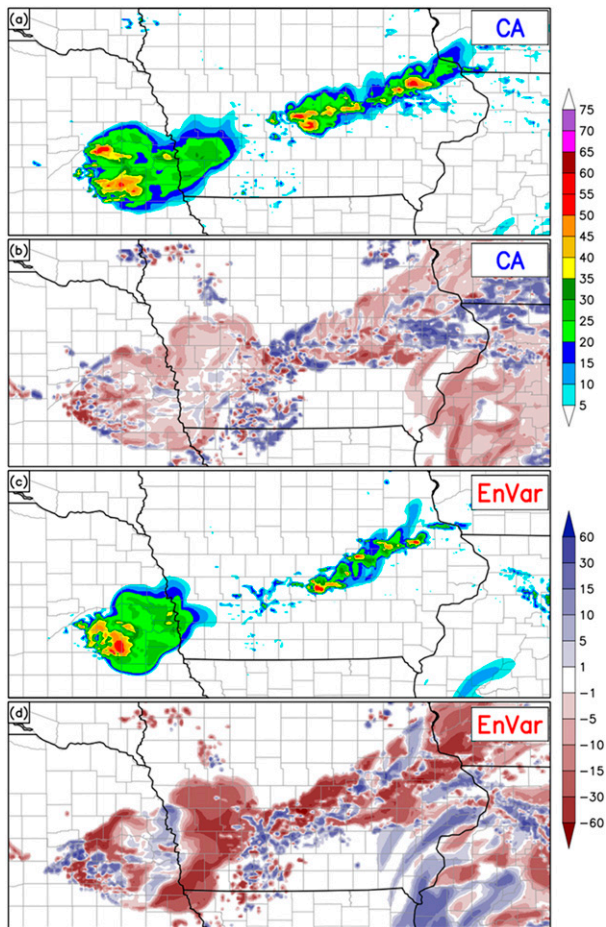
FIG. 9. (a),(c) Composite reflectivity analysis after DFI and (b),(d) change in composite reflectivity due to radar DA and DFI procedure for the (a),(b) cloud analysis and (c),(d) EnVar methods. All units are dBZ. Valid 2300 UTC 6 Jul 2016.
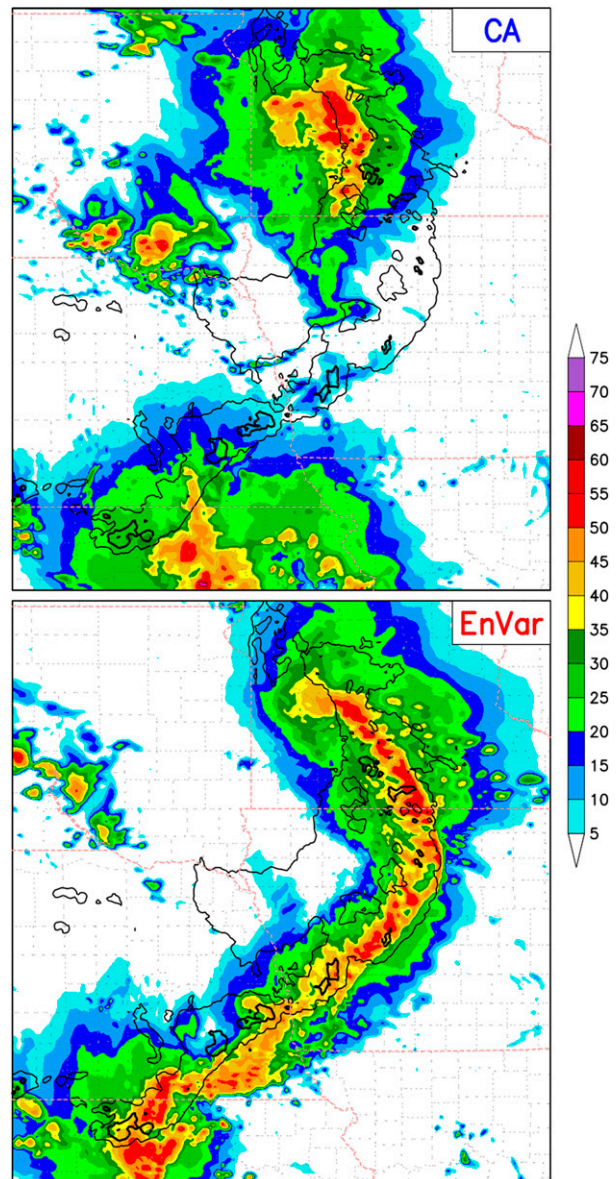

FIG. 10. Forecast composite reflectivity (color shades; dBZ) and observed composite reflectivity (thin and thick black contours at 30 and 50 dBZ, respectively) at 0700 UTC 7 Jul 2016; a 7-h forecast.

sustained convective development. Storms reintensified and additional storms developed during the 1-h background forecast leading up to the final radar-DA cycle (valid at 2000 UTC). However, the convection was more widespread in the CA first guess than in the EnVar first guess. And again, these storms were not suppressed and were carried through the final analysis and into the free forecast in both experiments, but the situation was significantly worse in the CA experiment.

In contrast to the CA (Figs. 13a,b), the EnVar performed better to add the missed hydrometeors (Fig. 13c). Moreover, spotty intense vertical motion accompanied the added simulated reflectivity (Fig. 13d) so that much of the newly assimilated reflectivity persisted into the free forecast, improving it. This evolution is an illustration of the causes of the superior performance of the EnVar. Not only does the direct assimilation of reflectivity result in a better simulated reflectivity

analysis, but the cross-variable correlations from the ensemble component allow unobserved state variables to be updated in a dynamically consistent manner to the reflectivity update, thus providing for a supportive environment for maintenance of convection rather than inserting hydrometeors with accompanying empirically derived latent heating into the model as the cloud analysis does. Therefore, the more accurate initial condition provided by the EnVar tends to lead to a more accurate EnVar free forecast compared to that from the CA.
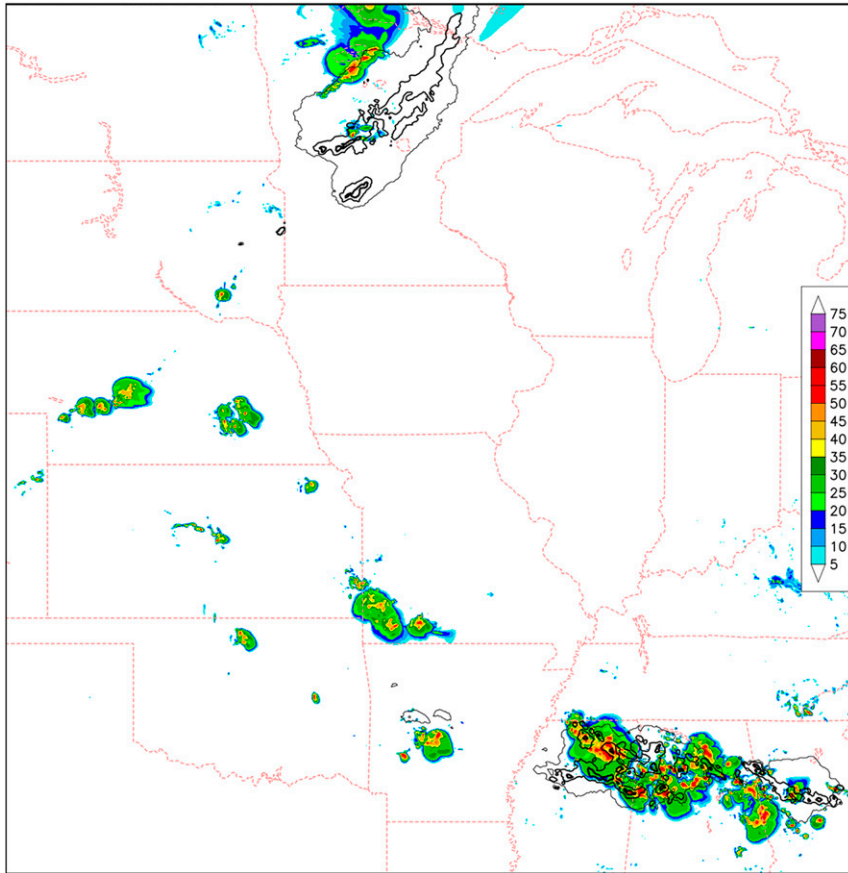
FIG. 11. As in Fig. 8, but valid at 1900 UTC 17 Jun 2016.

4. DFI half-window length

The relative inability of the CA-plus-DFI procedure to maintain newly assimilated reflectivity is illustrated by the forecast of the MCS over Minnesota. The analysis using a 2-min half-window length resulted in a pleasing final analysis of simulated reflectivity (Fig. 13a), but with insufficient vertical motion (Fig. 13b). Note that the spots of strong upward motion in Fig. 13b occurred with simulated reflectivity that was present in the first guess (outlined by the brown polygons in Figs. 13a,b). As a result of the negligible upward motion, hydrometeors simply sedimented upon integration of the free forecast, evinced by a rapid loss of simulated reflectivity coverage within the first 30 min of the free forecast (not shown). In contrast, using the 10-min half-window length in the DFI, deep convection is disappointingly scant (Fig. 14e), but for a similar reason as in the analysis using a 2-min half-window; instead of sedimenting during the free forecast, hydrometeors sedimented during the forward integration of the DFI. With a half-window length of 10 min, sedimentation occurs for a longer

time so that fewer hydrometers and hence less simulated reflectivity coverage were present at the end of the DFI integration, when the final analysis is valid. The other difference between the analyses using a 2- and 10-min DFI half-window length is in the magnitude and coverage of upward motion in the final analysis; substantially more upward motion is present in the final analysis using a 10-min half-window length (Fig. 14f). Additional CA experiments using DFI half-window lengths of 5 and 20 min were conducted to address the sensitivity of simulated reflectivity retention and vertical motion generation to DFI half-window length (Fig. 14). Overall simulated reflectivity coverage, especially in the region where simulated reflectivity coverage was absent in the first guess, gradually decreases with increasing DFI half-window length. However, there is a corresponding increase in vertical motion magnitudes with increasing DFI half-window length as well (Figs. 14b,d,f,h). A judicious choice of DFI half-window length is clearly needed to balance these two competing behaviors in storm-scale forecasts or else the analysis may be degraded.
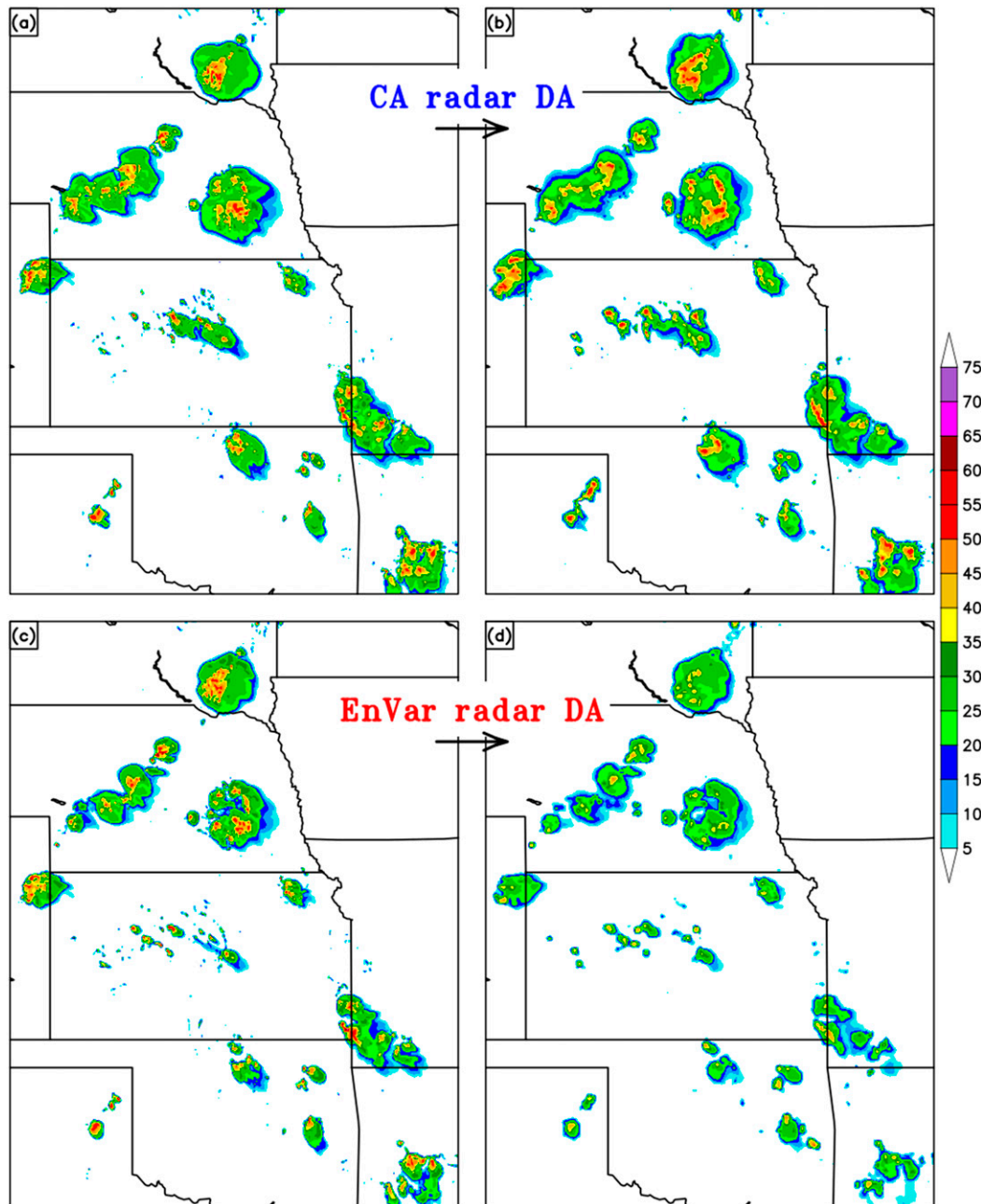
FIG. 12. NMMB composite reflectivity (dBZ) at 2000 UTC 17 Jun 2016; (a),(c), first guess; (b),(d) post-DFI analysis from the (a),(b) cloud analysis and (c),(d) EnVar method. Note that nearly all of the forecast reflectivity is spurious.

5. Summary and conclusions

Two methods for assimilating radar reflectivity observations into convection-allowing deterministic NWP forecasts were compared using the NMMB model. The cloud analysis method is being used operationally in convection-allowing prediction systems such as the HRRR and NAM, although it should be noted that the operational HRRR (version 3 as of late 2018) does not use a DFI. The cloud analysis inserts deep moist convection into the model by retrieving hydrometeor content from observed radar reflectivity, and then derives a latent heating rate which is applied through a DFI to generate convective circulations, restore dynamical balance, and filter unwanted and noisy small-scale features. The cloud analysis is relatively computationally cheap. Recently the GSI-based EnVar system was further developed to directly assimilate radar reflectivity
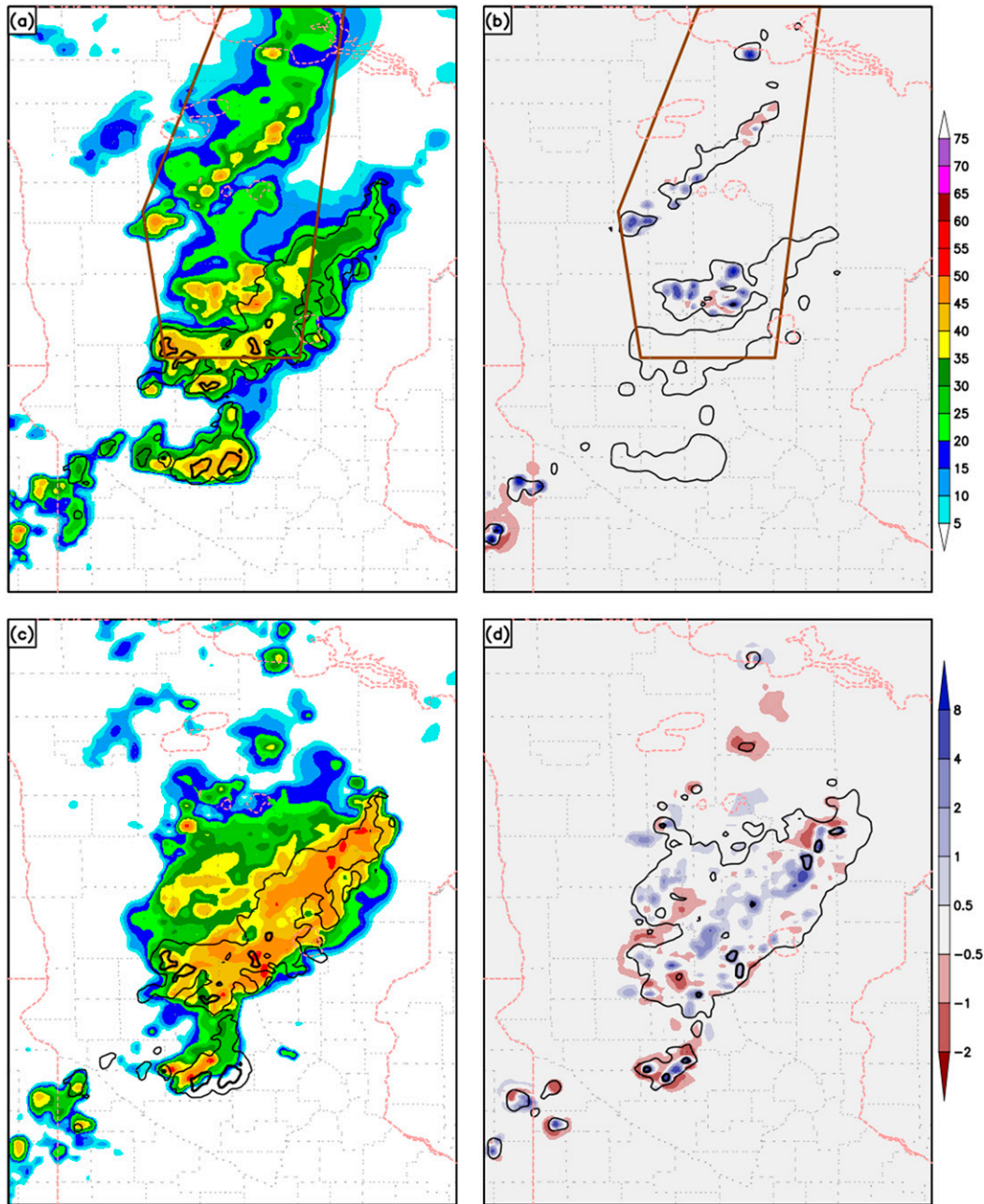
FIG. 13. (a),(c) Composite reflectivity (dBZ; shaded) with observed composite reflectivity in thin and thick black contours at 30 and 50 dBZ, respectively, and (b),(d) mean vertical velocity (m s^{-1}) between model levels 10 and 30 with forecast composite reflectivity in solid contours; 0-h forecast (post-DFI) valid 2000 UTC 17 Jun 2016 from the (top) CA and (bottom) EnVar methods. The brown polygon outlines the approximate area covered by reflectivity in the first guess for the CA method.

observations (Wang and Wang 2017). In this method the radar reflectivity is directly assimilated by minimizing the EnVar cost function. A method without involving the tangent linear and adjoint of the reflectivity operator was proposed and implemented. The ensemble covariance is provided by the GSI-based EnKF

from the NMMB model which is also extended for radar observations (Johnson et al. 2015). The two techniques were compared over 10 cases showcasing various convective regimes, regions within the United States, times of year, and strengths of large-scale forcing.
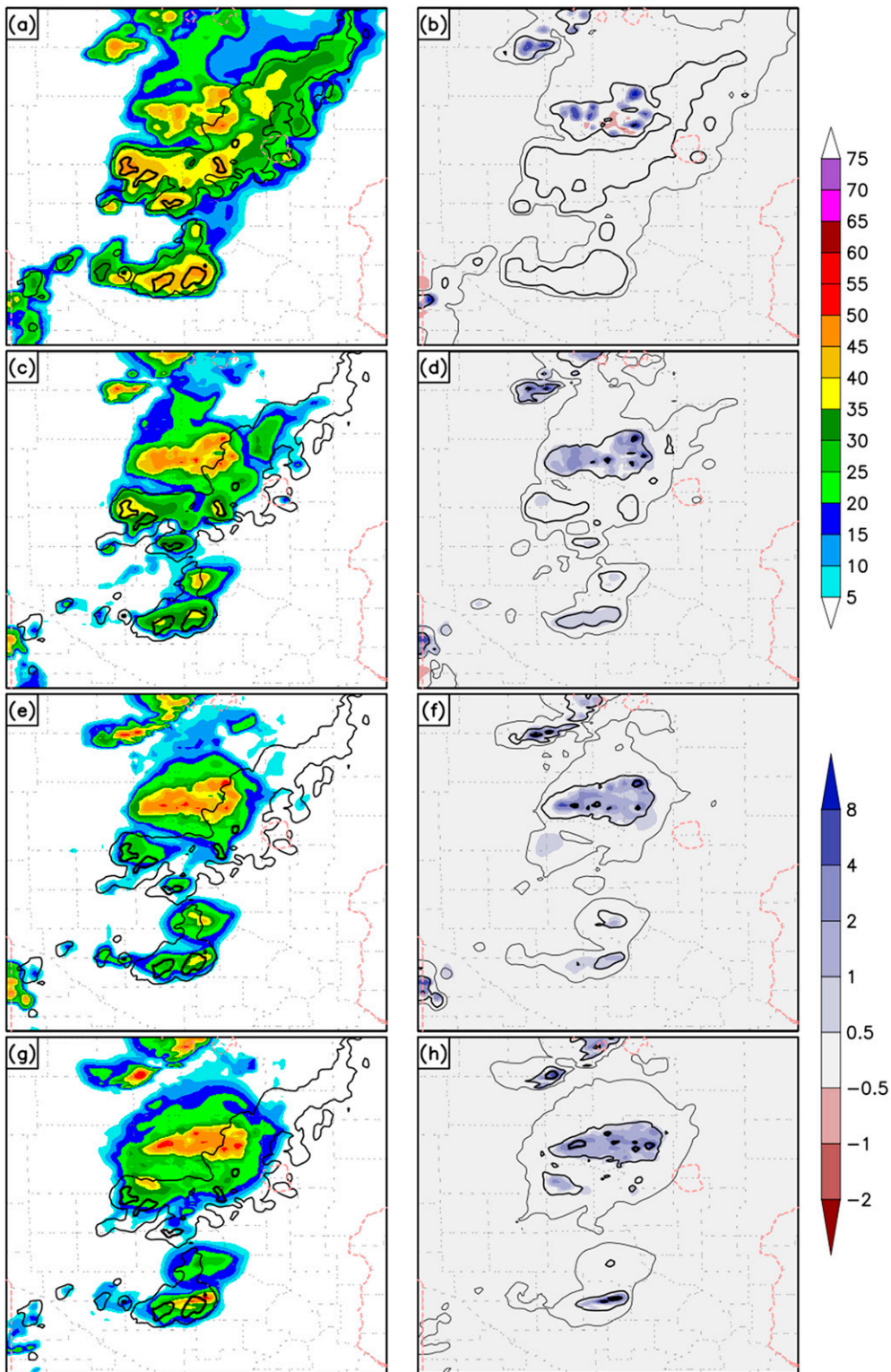
FIG. 14. (left) Forecast composite reflectivity (dBZ; shaded) with observed composite reflectivity contours (30 and 50 dBZ) and (right) forecast composite reflectivity contours (10, 30, and 50 dBZ) with mean vertical velocity (m s^{-1}; shaded) between model levels 10 and 30 in the final analysis after DFI from the cloud analysis procedure, valid at 2000 UTC 17 Jun 2016. The DFI half-window lengths are (a),(b) 2, (c),(d) 5, (e),(f) 10, and (g),(h) 20 min.

A variety of verification metrics, including neighborhood- and features-based methods, were used and demonstrated that the EnVar method is superior to the cloud analysis method, with statistically significant differences not only at the analysis time of the free forecast, but 6–9 h into the free forecast as well. The cloud analysis was found to improve convective-scale forecasts over no radar reflectivity assimilation, but only for the first few hours.

The reasons for the differences in performance between the EnVar and cloud analysis techniques were illustrated using case studies. It was shown that the EnVar method was able to more efficiently suppress spurious convection in the first guess than the CA method, even though neither method is capable of completely suppressing the entire spectrum of spurious convective-scale circulations. The tendency for overprediction of storms early in the forecast was shown by frequency biases exceeding 1.0 and forecast storm counts exceeding observed storm counts by a large margin. The case studies also illustrated the ability of the EnVar method to simultaneously update fields that are not directly assimilated. In particular, the assimilation of reflectivity and the correlation between simulated reflectivity and vertical motion enabled the EnVar to update the vertical motion field in a dynamically consistent manner, thus allowing assimilated reflectivity to persist upon integration of the model. In the cloud analysis procedure, instead a DFI is required to generate corresponding vertical velocity and adjust other related fields. The case studies also illustrated the need for a balance between the length of the DFI half-window and the amount of reflectivity/hydrometeor content retained in the cloud analysis method: a short window favors retention of hydrometeors in the analysis but is not sufficient for convective updrafts to develop, and thus newly assimilated simulated reflectivity quickly dissipates; whereas the opposite is true for a longer window.

In this study, the configuration of the cloud analysis to assimilate radar reflectivity matches that of the operational cloud analysis implementation in the NAM except for only assimilating radar reflectivity (and no other observations). This work represents an early study of understanding of the difference between EnVar direct reflectivity assimilation and an indirect cloud analysis approach in near-operational settings. Additional tuning of the EnVar method would still be needed in the future before being considered for operational implementation. For example, the localization scales for both the conventional and radar observations are used to account for the system errors associated with the sampling error and misrepresentation of model error (Houtekamer et al. 2005; Whitaker et al. 2008); variance inflation is employed to address the underestimated background-error variances (Whitaker and Hamill 2012);

radar DA frequency and time window should be tuned to sufficiently absorb the high spatiotemporal radar information for initializing deep moist convection. As research in the area of radar data assimilation proceeds, it is likely that latent heat nudging schemes such as the cloud analysis will eventually be dropped in favor of the sophisticated ensemble-based approaches such as the EnVar method, especially as computer technology allows these more expensive methods to become feasible for operational forecasting purposes where forecast latency is a significant constraint.

Further advancement of methods is still needed for direct radar reflectivity assimilation. The treatments for model states and observations associated with small reflectivity values and the bounded nature of hydrometeor mixing ratios deviate from the assumption of Gaussianity undertaken by the EnVar approach.[1] Other methods considering higher-order moments (e.g., Hodyss 2012), mixed lognormal-Gaussian distributions (e.g., Fletcher and Jones 2014), or using the nonparametric distribution approach (e.g., Poterjoy 2016) are left for future studies. In addition, the impacts of handling sampling errors associated with a small ensemble size and methods to further evaluate and improve the ensemble covariance for convective-scale data assimilation may be further investigated in future studies. In addition, how to treat ensemble deficiencies associated with the underrepresentation of model errors at the convective scale require further study. Zeng et al. (2018) compared various approaches to facilitate sampling model errors for convective-scale data assimilation. Methods other than the RTPS should be investigated in future studies.

In future studies of convection-allowing forecasts of convective storms, it may be useful to verify other types of radar fields. Reflectivity is the most widely available, but it is a three-dimensional field, and due to issues with volume coverage patterns, it is essentially impossible to sample the entire tropospheric reflectivity field. We chose to verify composite reflectivity here for simplicity and universality. However, other observed radar products are available. One such product available from MRMS is called "hybrid-scan reflectivity" and represents reflectivity at the lowest level above ground that the radars can reasonably sample. It could be useful to compare this field to two-dimensional model reflectivity on model surfaces since model vertical coordinates are also generally terrain-following, as is the hybrid-scan

---

[1] Like earlier studies using EnKF to assimilate radar reflectivity, sometimes negative hydrometeor values are produced during the assimilation due to the Gaussian assumption adopted. These values are reset to zeros before being used to initialize the subsequent forecast.

reflectivity product. Low-level reflectivity represents low-level storm structure better than does composite reflectivity, and typically the near-surface reflectivity is more interesting and critical to forecast since internal storm structures, such as hook echoes in supercell storms and gust fronts in squall lines, can both be represented in low-level reflectivity fields. One advantage to using composite reflectivity, however, is that it delineates thunderstorm anvil structure, which is important for verifying sophisticated microphysics schemes like those commonly used in convection-allowing NWP forecasts.

REFERENCES

Aksoy, A., D. C. Dowell, and C. Snyder, 2009: A multicase comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part I: Storm-scale analyses. *Mon. Wea. Rev.*, **137**, 1805–1824, https://doi.org/10.1175/2008MWR2691.1.

——, ——, and ——, 2010: A multicase comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part II: Short-range ensemble forecasts. *Mon. Wea. Rev.*, **138**, 1273–1292, https://doi.org/10.1175/2009MWR3086.1.

Aligo, E. A., B. Ferrier, and J. R. Carley, 2018: Modified NAM microphysics for forecasts of deep convective storms. *Mon. Wea. Rev.*, **146**, 4115–4153, https://doi.org/10.1175/MWR-D-17-0277.1.

Benjamin, S. G., and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, https://doi.org/10.1175/MWR-D-15-0242.1.

Bick, T., and Coauthors, 2016: Assimilation of 3D radar reflectivities with an ensemble Kalman filter on the convective scale. *Quart. J. Roy. Meteor. Soc.*, **142**, 1490–1504, https://doi.org/10.1002/qj.2751.

Buehner, M., 2005: Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting. *Quart. J. Roy. Meteor. Soc.*, **131**, 1013–1043, https://doi.org/10.1256/qj.04.15.

Cai, H., and R. E. Dumais, 2015: Object-based evaluation of a numerical weather prediction model's performance through forecast storm characteristic analysis. *Wea. Forecasting*, **30**, 1451–1468, https://doi.org/10.1175/WAF-D-15-0008.1.

Carley, J. R., 2012: Hybrid ensemble-3DVar radar data assimilation for the short-term prediction of convective storms. Ph.D. dissertation, Department of Earth, Atmospheric, and Planetary Sciences, Purdue University, 205 pp.

Caya, A., J. Sun, and C. Snyder, 2005: A comparison between the 4DVAR and the ensemble Kalman filter techniques for radar data assimilation. *Mon. Wea. Rev.*, **133**, 3081–3094, https://doi.org/10.1175/MWR3021.1.

Clark, A. J., W. A. Gallus, and M. L. Weisman, 2010: Neighborhood-based verification of precipitation forecasts from convection-allowing NCAR WRF Model simulations and the operational NAM. *Wea. Forecasting*, **25**, 1495–1509, https://doi.org/10.1175/2010WAF2222404.1.

Davis, C., B. Brown, and R. Bullock, 2006: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, **134**, 1772–1784, https://doi.org/10.1175/MWR3145.1.

Dawson, D. T., M. Xue, J. A. Milbrandt, and A. Shapiro, 2015: Sensitivity of real-data simulations of the 3 May 1999 Oklahoma City tornadic supercell and associated tornadoes to multimoment microphysics. Part I: Storm- and tornado-scale numerical forecasts. *Mon. Wea. Rev.*, **143**, 2241–2265, https://doi.org/10.1175/MWR-D-14-00279.1.

Dowell, D. C., L. J. Wicker, and C. Snyder, 2011: Ensemble Kalman filter assimilation of radar observations of the 8 May 2003 Oklahoma City supercell: Influences of reflectivity observations of storm-scale analyses. *Mon. Wea. Rev.*, **139**, 272–294, https://doi.org/10.1175/2010MWR3438.1.

Du, J., G. DiMego, D. Jovic, B. Ferrier, B. Yang, and B. Zhou, 2015: Short Range Ensemble Forecast (SREF) system at NCEP: Recent development and future transition. *27th Conf. on Weather Analysis and Forecasting/23rd Conf. on Numerical Weather Prediction*, Chicago, IL, Amer. Meteor. Soc., 2A.5, https://ams.confex.com/ams/27WAF23NWP/webprogram/Paper273421.html.

Duda, J. D., and W. A. Gallus, 2013: The impact of large-scale forcing on skill of simulated convective initiation and upscale evolution with convection-allowing grid spacings in WRF. *Wea. Forecasting*, **28**, 994–1018, https://doi.org/10.1175/WAF-D-13-00005.1.

——, X. Wang, F. Kong, and M. Xue, 2014: Using varied microphysics to account for uncertainty in warm-season QPF in a convection-allowing ensemble. *Mon. Wea. Rev.*, **142**, 2198–2219, https://doi.org/10.1175/MWR-D-13-00297.1.

Fletcher, S. J., and A. S. Jones, 2014: Multiplicative and additive incremental variational data assimilation for mixed lognormal–Gaussian errors. *Mon. Wea. Rev.*, **142**, 2521–2544, https://doi.org/10.1175/MWR-D-13-00136.1.

Gao, J., and D. J. Stensrud, 2012: Assimilation of reflectivity data in a convective-scale, cycled 3DVAR framework with hydrometeor classification. *J. Atmos. Sci.*, **69**, 1054–1065, https://doi.org/10.1175/JAS-D-11-0162.1.

Gilleland, E., D. Ahijevych, B. G. Brown, B. Casati, and E. E. Ebert, 2009: Intercomparison of spatial forecast verification methods. *Wea. Forecasting*, **24**, 1416–1430, https://doi.org/10.1175/2009WAF2222269.1.

Gustafsson, N., and Coauthors, 2018: Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres. *Quart. J. Roy. Meteor. Soc.*, **144**, 1218–1256, https://doi.org/10.1002/qj.3179.

Hamill, T. M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting*, **14**, 155–167, https://doi.org/10.1175/1520-0434(1999)014<0155:HTFENP>2.0.CO;2.

Hodyss, D., 2012: Accounting for skewness in ensemble data assimilation. *Mon. Wea. Rev.*, **140**, 2346–2358, https://doi.org/10.1175/MWR-D-11-00198.1.

Houtekamer, P. L., H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2005: Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Wea. Rev.*, **133**, 604–620, https://doi.org/10.1175/MWR-2864.1.

Hu, M., and M. Xue, 2007: Impact of configurations of rapid intermittent assimilation of WSR-88D radar data for the 8 May 2003 Oklahoma City tornadic thunderstorm case. *Mon. Wea. Rev.*, **135**, 507–525, https://doi.org/10.1175/MWR3313.1.

——, ——, and K. Brewster, 2006a: 3DVAR and cloud analysis with WSR-88D Level-II data for the prediction of the Fort Worth, Texas, tornadic thunderstorms. Part I: Cloud analysis and its impact. *Mon. Wea. Rev.*, **134**, 675–698, https://doi.org/10.1175/MWR3092.1.

——, ——, J. Gao, and K. Brewster, 2006b: 3DVAR and cloud analysis with WSR-88D Level-II data for the prediction of the Fort Worth, Texas, tornadic thunderstorms. Part II: Impact of radial velocity analysis via 3DVAR. *Mon. Wea. Rev.*, **134**, 699–721, https://doi.org/10.1175/MWR3093.1.

Janjić, Z. I., 2003: A nonhydrostatic model based on a new approach. *Meteor. Atmos. Phys.*, **82**, 271–285, https://doi.org/10.1007/s00703-001-0587-6.

——, and R. Gall, 2012: Scientific documentation of the NCEP nonhydrostatic multiscale model on the B grid (NMMB). Part I: Dynamics. NCAR Tech. Note NCAR/TN-489+STR, 75 pp., https://doi.org/10.5065/D6WH2MZX.

Johnson, A., and X. Wang, 2012: Verification and calibration of neighborhood and object-based probabilistic precipitation forecasts from a multimodel convection-allowing ensemble. *Mon. Wea. Rev.*, **140**, 3054–3077, https://doi.org/10.1175/MWR-D-11-00356.1.

——, and ——, 2013: Object-based evaluation of a storm-scale ensemble during the 2009 NOAA Hazardous Weather Testbed Spring Experiment. *Mon. Wea. Rev.*, **141**, 1079–1098, https://doi.org/10.1175/MWR-D-12-00140.1.

——, ——, J. R. Carley, L. J. Wicker, and C. Karstens, 2015: A comparison of multiscale GSI-based EnKF and 3DVar data assimilation using radar and conventional observations for midlatitude convective-scale precipitation forecasts. *Mon. Wea. Rev.*, **143**, 3087–3108, https://doi.org/10.1175/MWR-D-14-00345.1.

Lakshmanan, V., and T. W. Humphrey, 2014: A MapReduce technique to mosaic continental-scale weather radar data in real-time. *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, **7**, 721–732, https://doi.org/10.1109/JSTARS.2013.2282040.

——, T. Smith, K. Hondl, G. J. Stumpf, and A. Witt, 2006: A real-time, three-dimensional, rapidly updating, heterogeneous radar merger technique for reflectivity, velocity and derived products. *Wea. Forecasting*, **21**, 802–823, https://doi.org/10.1175/WAF942.1.

Leuenberger, D., and A. Rossa, 2007: Revisiting the latent heat nudging scheme for the rainfall assimilation of a simulated convective storm. *Meteor. Atmos. Phys.*, **98**, 195–215, https://doi.org/10.1007/s00703-007-0260-9.

Lorenc, A. C., 2003: The potential of the ensemble Kalman filter for NWP—A comparison with 4D-VAR. *Quart. J. Roy. Meteor. Soc.*, **129**, 3183–3203, https://doi.org/10.1256/qj.02.132.

Marquis, J., Y. Richardson, P. Markowski, D. Dowell, J. Wurman, K. Kosiba, P. Robinson, and G. Romine, 2014: An investigation of the Goshen County, Wyoming, tornadic supercell of 5 June 2009 using EnKF assimilation of mobile mesonet and radar observations collected during VORTEX2. Part I: Experiment design and verification of the EnKF analyses. *Mon. Wea. Rev.*, **142**, 530–554, https://doi.org/10.1175/MWR-D-13-00007.1.

Pan, Y., K. Zhu, M. Xue, X. Wang, M. Hu, S. G. Benjamin, S. S. Weygandt, and J. S. Whitaker, 2014: A GSI-based coupled EnSRF–En3DVar hybrid data assimilation system for the operational Rapid Refresh model: Tests at a reduced resolution. *Mon. Wea. Rev.*, **142**, 3756–3780, https://doi.org/10.1175/MWR-D-13-00242.1.

Peckham, S. E., T. G. Smirnova, S. G. Benjamin, J. M. Brown, and J. S. Kenyon, 2016: Implementation of a digital filter initialization in the WRF Model and its application in the Rapid Refresh. *Mon. Wea. Rev.*, **144**, 99–106, https://doi.org/10.1175/MWR-D-15-0219.1.

Poterjoy, J., 2016: A localized particle filter for high-dimensional nonlinear systems. *Mon. Wea. Rev.*, **144**, 59–76, https://doi.org/10.1175/MWR-D-15-0163.1.

Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, https://doi.org/10.1175/2007MWR2123.1.

Rogers, E., and Coauthors, 2017: Upgrades to the NCEP North American Mesoscale (NAM) system. *Working Group on Numerical Experimentation Blue Book*, 2 pp., http://wmc.meteoinfo.ru/bluebook/uploads/2017/docs/05_Rogers_Eric_mesoscale_modeling.pdf.

Schenkman, A. D., M. Xue, A. Shapiro, K. Brewster, and J. Gao, 2011: The analysis and prediction of the 8–9 May 2007 Oklahoma tornadic mesoscale convective system by assimilating WSR-88D and CASA radar data using 3DVAR. *Mon. Wea. Rev.*, **139**, 224–246, https://doi.org/10.1175/2010MWR3336.1.

Schumacher, R. S., and R. H. Johnson, 2005: Organization and environmental properties of extreme-rain-producing mesoscale convective systems. *Mon. Wea. Rev.*, **133**, 961–976, https://doi.org/10.1175/MWR2899.1.

Schwartz, C. S., 2017: A comparison of methods used to populate neighborhood-based contingency tables for high-resolution forecast verification. *Wea. Forecasting*, **32**, 733–741, https://doi.org/10.1175/WAF-D-16-0187.1.

——, and Z. Liu, 2014: Convection-permitting forecasts initialized with continuously cycling limited-area 3DVAR, ensemble Kalman filter, and "hybrid" variational–ensemble data assimilation systems. *Mon. Wea. Rev.*, **142**, 716–738, https://doi.org/10.1175/MWR-D-13-00100.1.

——, and R. A. Sobash, 2017: Generating probabilistic forecasts from convection-allowing ensembles using neighborhood approaches: A review and recommendations. *Mon. Wea. Rev.*, **145**, 3397–3418, https://doi.org/10.1175/MWR-D-16-0400.1.

Snook, N., M. Xue, and Y. Jung, 2011: Analysis of a tornadic mesoscale convective vortex based on ensemble Kalman filter assimilation of CASA X-band and WSR-88D radar data. *Mon. Wea. Rev.*, **139**, 3446–3468, https://doi.org/10.1175/MWR-D-10-05053.1.

Snyder, C., and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **131**, 1663–1677, https://doi.org/10.1175/2555.1.

Sugimoto, S., N. A. Crook, J. Sun, Q. Xiao, and D. M. Barker, 2009: An examination of WRF 3DVAR radar data assimilation on its capability in retrieving unobserved variables and forecasting precipitation through observing system simulation

experiments. *Mon. Wea. Rev.*, **137**, 4011–4029, https://doi.org/10.1175/2009MWR2839.1.

Sun, J., 2005: Convective-scale assimilation of radar data: Progress and challenges. *Quart. J. Roy. Meteor. Soc.*, **131**, 3439–3463, https://doi.org/10.1256/qj.05.149.

Tong, M., and M. Xue, 2005: Ensemble Kalman filter assimilation of Doppler radar data with a compressible nonhydrostatic model: OSS experiments. *Mon. Wea. Rev.*, **133**, 1789–1807, https://doi.org/10.1175/MWR2898.1.

Wang, X., 2010: Incorporating ensemble covariance in the gridpoint statistical interpolation variational minimization: A mathematical framework. *Mon. Wea. Rev.*, **138**, 2990–2995, https://doi.org/10.1175/2010MWR3245.1.

——, and T. Lei, 2014: GSI-based four-dimensional ensemble-variational (4DEnsVar) data assimilation: Formulation and single-resolution experiments with real data for NCEP Global Forecast System. *Mon. Wea. Rev.*, **142**, 3303–3325, https://doi.org/10.1175/MWR-D-13-00303.1.

——, and X. Wang, 2017: Direct assimilation of radar reflectivity without tangent linear and adjoint of the nonlinear observation operator in the GSI-based EnVar system: Methodology and experiment with the 8 May 2003 Oklahoma City tornadic supercell. *Mon. Wea. Rev.*, **145**, 1447–1471, https://doi.org/10.1175/MWR-D-16-0231.1.

——, C. Snyder, and T. M. Hamill, 2007: On the theoretical equivalence of differently proposed ensemble-3DVAR hybrid analysis schemes. *Mon. Wea. Rev.*, **135**, 222–227, https://doi.org/10.1175/MWR3282.1.

——, D. M. Barker, C. Snyder, and T. M. Hamill, 2008a: A hybrid ETKF–3DVAR data assimilation scheme for the WRF Model. Part I: Observing system simulation experiment. *Mon. Wea. Rev.*, **136**, 5116–5131, https://doi.org/10.1175/2008MWR2444.1.

——, ——, ——, and ——, 2008b: A hybrid ETKF–3DVAR data assimilation scheme for the WRF Model. Part II: Real observation experiments. *Mon. Wea. Rev.*, **136**, 5132–5147, https://doi.org/10.1175/2008MWR2445.1.

——, D. Parrish, D. Kleist, and J. Whitaker, 2013: GSI 3DVar-based ensemble–variational hybrid data assimilation for NCEP Global Forecast System: Single-resolution experiments. *Mon. Wea. Rev.*, **141**, 4098–4117, https://doi.org/10.1175/MWR-D-12-00141.1.

Whitaker, J. S., and T. M. Hamill, 2012: Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Wea. Rev.*, **140**, 3078–3089, https://doi.org/10.1175/MWR-D-11-00276.1.

——, ——, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP global forecast system. *Mon. Wea. Rev.*, **136**, 463–482, https://doi.org/10.1175/2007MWR2018.1.

Wolff, J. K., M. Harrold, T. Fowler, J. H. Gotway, L. Nance, and B. G. Brown, 2014: Beyond the basics: Evaluating model-based precipitation forecasts using traditional, spatial, and object-based methods. *Wea. Forecasting*, **29**, 1451–1472, https://doi.org/10.1175/WAF-D-13-00135.1.

Xiao, Q., and J. Sun, 2007: Multiple-radar data assimilation and short-range quantitative precipitation forecasting of a squall line observed during IHOP_2002. *Mon. Wea. Rev.*, **135**, 3381–3404, https://doi.org/10.1175/MWR3471.1.

Yussouf, N., and D. J. Stensrud, 2012: Comparison of single-parameter and multiparameter ensembles for assimilation of radar observations using the ensemble Kalman filter. *Mon. Wea. Rev.*, **140**, 562–586, https://doi.org/10.1175/MWR-D-10-05074.1.

——, E. R. Mansell, L. J. Wicker, D. M. Wheatley, and D.J. Stensrud, 2013: The ensemble Kalman filter analyses and forecasts of the 8 May 2003 Oklahoma City tornadic supercell storm using single- and double-moment microphysics schemes. *Mon. Wea. Rev.*, **141**, 3388–3412, https://doi.org/10.1175/MWR-D-12-00237.1.

Zeng, Y., T. Janjić, A. de Lozar, U. Blahak, H. Reich, C. Keil, and A. Siefert, 2018: Representation of model error in convective-scale data assimilation: Additive noise, relaxation methods, and combinations. *J. Adv. Model. Earth Syst.*, **10**, 2889–2911, https://doi.org/10.1029/2018MS001375.

Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238–1253, https://doi.org/10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2.

Zhou, X., Y. Zhu, D. Hou, Y. Luo, J. Peng, and R. Wobus, 2017: Performance of the new NCEP global ensemble forecast system in a parallel experiment. *Wea. Forecasting*, **32**, 1989–2004, https://doi.org/10.1175/WAF-D-17-0023.1.