Object-Based Verification of GSI EnKF and Hybrid En3DVar Radar Data Assimilation and Convection-Allowing Forecasts within a Warn-on-Forecast Framework

LIANGLYU CHEN,[a,b] CHENGSI LIU,[b] YOUNGSUN JUNG,[b] PATRICK SKINNER,[c,d] MING XUE,[b,d] AND RONG KONG[b]

[a] *Chongqing Institute of Meteorological Sciences, Chongqing, China*
[b] *Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma*
[c] *Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma*
[d] *School of Meteorology, University of Oklahoma, Norman, Oklahoma*

(Manuscript received 28 September 2020, in final form 7 February 2022)

ABSTRACT: The Center for Analysis and Prediction of Storms has recently developed capabilities to directly assimilate radar reflectivity and radial velocity data within the GSI-based ensemble Kalman filter (EnKF) and hybrid ensemble three-dimensional variational (En3DVar) system for initializing convective-scale forecasts. To assess the performance of EnKF and hybrid En3DVar with different hybrid weights (with 100%, 20%, and 0% of static background error covariance corresponding to pure 3DVar, hybrid En3DVar, and pure En3DVar) for assimilating radar data in a Warn-on-Forecast framework, a set of data assimilation and forecast experiments using the WRF Model are conducted for six convective storm cases of May 2017. Using an object-based verification approach, forecast objects of composite reflectivity and 30-min updraft helicity swaths are verified against reflectivity and rotation track objects in Multi-Radar Multi-Sensor data on space and time scales typical of National Weather Service warnings. Forecasts initialized by En3DVar or the best performing EnKF ensemble member produce the highest object-based verification scores, while forecasts from 3DVar and the worst EnKF member produce the lowest scores. Averaged across six cases, hybrid En3DVar using 20% static background error covariance does not improve forecasts over pure En3DVar, although improvements are seen in some individual cases. The false alarm ratios of EnKF members for both composite reflectivity and updraft helicity at the initial time are lower than those from variational methods, suggesting that EnKF analysis reduces spurious reflectivity and mesocyclone objects more effectively.

KEYWORDS: Short-range prediction; Data assimilation; Numerical weather prediction/forecasting

1. Introduction

Increasing severe thunderstorm, flash flood, and tornado warning lead times is critical to reducing the loss of life, injury, and economic costs of high-impact weather. To extend warning lead time, NOAA's Warn-on-Forecast (WoF) program is tasked with producing probabilistic hazard guidance based on convection-allowing/resolving ensemble forecasts of numerical weather prediction (NWP) models (Stensrud et al. 2009, 2013). Recently, various prototype WoF systems (e.g., Snook et al. 2012, 2016; Yussouf et al. 2013; Wheatley et al. 2015; Jones et al. 2016; Johnson et al. 2017; Lawson et al. 2018; Labriola et al. 2019; Stratman et al. 2020) have demonstrated capabilities for producing skillful forecasts of severe weather hazards, including tornadoes, severe hail, and flash floods.

The WoF program includes two main aspects: the production of a rapidly updating assimilation system to incorporate various data such as radar, satellite and conventional observations, and the ability to produce short-range ensemble forecasts from these analyses to predict hazardous convective weather (Stensrud et al. 2013). In such a system, radar data are an important data source as it can provide three-dimensional observations of the internal structure of convective storms with high spatial and temporal resolution. Several prototype Warn-on-Forecast systems (e.g., Snook et al. 2012, 2016; Yussouf et al. 2013; Wheatley et al. 2015) used an ensemble Kalman filter approach (EnKF; Evensen 1994, 2003; Houtekamer and Zhang 2016) to assimilate radar data and initialize the convection-allowing model (CAM) forecasts. This approach is attractive at convective scale because flow-dependent error covariance statistics derived from the forecast ensemble can be used to update unobserved model state variables. Recently, EnKF techniques have been used to assimilate radar and satellite data into the National Severe Storm Laboratory (NSSL) Warn-on-Forecast System (WoFS) and provide real-time forecasts of convective storm hazards (Wheatley et al. 2015; Jones et al. 2016). Object-based verification of WoFS short-term thunderstorm forecasts across cases from 2016 to 2019 has provided evidence that the system can produce skillful short-term predictions of thunderstorms and mesocyclones at the scale of a typical National Weather Service (NWS) warning (Skinner et al. 2016; 2018; Flora et al. 2019). Real-time forecast output is also produced by the NSSL hybrid WoF analysis and forecast system (WoF-AFS; Wang et al. 2019). WoF-AFS includes two components: an ensemble analysis forecast system based on EnKF data assimilation (DA) and a deterministic analysis and forecast system based on hybrid ensemble three-dimensional variational (En3DVar) DA with a 50% weight of static background error covariance (BEC) and a 50% weight of ensemble-derived BEC.

Though EnKF is attractive for convective-scale DA and can effectively assimilate radar observations in the presence

Corresponding author: Chengsi Liu, cliu@ou.edu

DOI: 10.1175/WAF-D-20-0180.1

of complex microphysics (e.g., Tong and Xue 2005), insufficient ensemble spread and rank-deficient ensemble BEC due to a small ensemble size can cause suboptimal performance. Using a hybrid method that combines high-rank, static variational BEC and low-rank, flow-dependent ensemble BEC (Hamill and Snyder 2000) has demonstrated advantages over pure EnKF for large-scale NWP (Etherton and Bishop 2004; Wang et al. 2007; Buehner et al. 2010a,b, 2013; Clayton et al. 2013), and has also shown promise for convective-scale radar DA (Gao et al. 2013; Kong et al. 2018; Wang et al. 2019). However, when directly assimilating radar reflectivity within a variational DA framework, various issues arise owing to the high nonlinearity of the reflectivity observation operator (Sun and Crook 1997; Carley 2012; Wang and Wang 2017; Liu et al. 2019).

Recently, capabilities to directly assimilate radar reflectivity and radial velocity data were added to the Gridpoint Statistical Interpolation (GSI; Kleist et al. 2009) hybrid En3DVar system by the Center for Analysis and Prediction of Storms (CAPS), including special treatments to deal with a number of issues related to reflectivity assimilation. These treatments include the use of temperature-dependent background error profiles for hydrometeor control variables to ensure physical partitioning of radar-observed precipitation information among liquid and ice hydrometeors, as originally developed and tested in the ARPS 3DVar system (Liu et al. 2019). Also included is the ability to use of logarithmic hydrometeor mixing ratios (CVlogq; Liu et al. 2020) or a general power transformation of the mixing ratios (CVpq; Chen et al. 2021) instead of the mixing ratios themselves (CVq) as the control variables. The use of CVlogq or CVpq are found to be necessary for reasonably fast convergence of cost function minimization when assimilating reflectivity data, and to ensure proper realization of the radial velocity assimilation impact. Several other algorithmic treatments (e.g., adding a lower limit on the hydrometeor mixing ratios and assimilating radial velocity data in a separate pass) were also implemented within GSI En3DVar to avoid spurious analysis increments or convergence problems (Liu et al. 2020), and these features are used where desirable in this study. In addition to GSI En3DVar, GSI EnKF is also run to provide ensemble perturbations to the hybrid En3DVar (Pan et al. 2014). To ensure consistency with the microphysics scheme used in the WRF forecast model, we implemented a reflectivity observation operator consistent with the Thompson microphysics scheme within the GSI EnKF and hybrid En3DVar systems.

In this paper, we present results testing the radar DA capabilities of the GSI EnKF and hybrid En3DVar for six severe weather cases from May 2017, using configurations similar to the prototype WoFS system. In particular, we use an object-based verification system recently developed for evaluating WoFS guidance (Skinner et al. 2018) to compare predictions of composite reflectivity and updraft helicity (UH) swaths (Kain et al. 2008). Four DA and forecast experiments are run for each case, where the DA method uses EnKF or hybrid En3DVar with different hybrid weights. These experiments are named EnKF, EnVar (with 0% of static BEC), Hybrid (20% of static BEC), and 3DVar (100% of static BEC), respectively. While some earlier studies have examined the performance of similar algorithms (e.g., Johnson et al. 2017; Kong et al. 2018; Wang et al. 2019), this study marks the first time that the algorithms implemented within the same DA framework are intercompared for real cases for prototype WoFS configurations (Skinner et al. 2018).

The rest of this paper is organized as follows. A brief overview of the cases and experiment configurations are presented in sections 2 and 3 provides a description of the object-based verification methods. The verification results of composite reflectivity and updraft helicity forecasts are discussed in section 4. Section 5 summarizes the results of this study and discusses potential future work.

2. Case overview and experiment configuration

a. Cases overview

The six selected cases in May 2017 produced numerous severe weather warnings and reports. Figure 1 shows the domain and the Multi-Radar Multi-Sensor (MRMS; Smith et al. 2016) composite reflectivity at 2300 UTC (1700 central standard time). Isolated supercells occurred on 16 and 18 May, resulting in the largest numbers of tornado and severe hail reports (Figs. 1b,d). More isolated, discrete thunderstorms with fewer local storm reports occurred on 9 and 23 May (Figs. 1a,e). The 17 and 27 May cases (Figs. 1c,f) produced mesoscale convective systems with extensive damaging wind reports. Event dates, the number of severe weather reports, and brief descriptions are provided in Table 1.

b. Prediction model settings

The forecast model used is the Advanced Research version of WRF (ARW-WRF), version 3.8.1 (Skamarock et al. 2008), and has the same domain and physics configuration as the 2017 WoFS (Skinner et al. 2018) except that the Thompson microphysics scheme (Thompson et al. 2008) is used in place of the NSSL two-moment scheme (Mansell et al. 2010). The domain is roughly centered over severe weather events in each case and has 250 × 250 grid points in the horizontal with 3-km grid spacing, and 51 vertical levels. The Noah land surface model (Chen and Dudhia 2001), Yonsei University (YSU) planetary boundary layer (PBL) scheme (Hong et al. 2006), and the RRTMG shortwave and longwave radiation schemes (Iacono et al. 2008) are used for the deterministic forecasts of variational experiments. Multiple radiation and PBL schemes are used in the EnKF ensemble forecasts as described in Table 2, so the DA ensemble in this study is an initial conditions/lateral boundary conditions (IC/LBC) and model physics perturbed ensemble.

c. DA experiment configurations

In this study, we use GSI EnKF and hybrid En3DVar with the radar DA capabilities developed by CAPS. The EnKF DA configurations (Table 3) are based on those of GSI EnKF-initialized Storm Scale Ensemble Forecasts run by CAPS during the 2019 Hazardous Weather Testbed Spring Forecast Experiment (Clark et al. 2020), except that radar

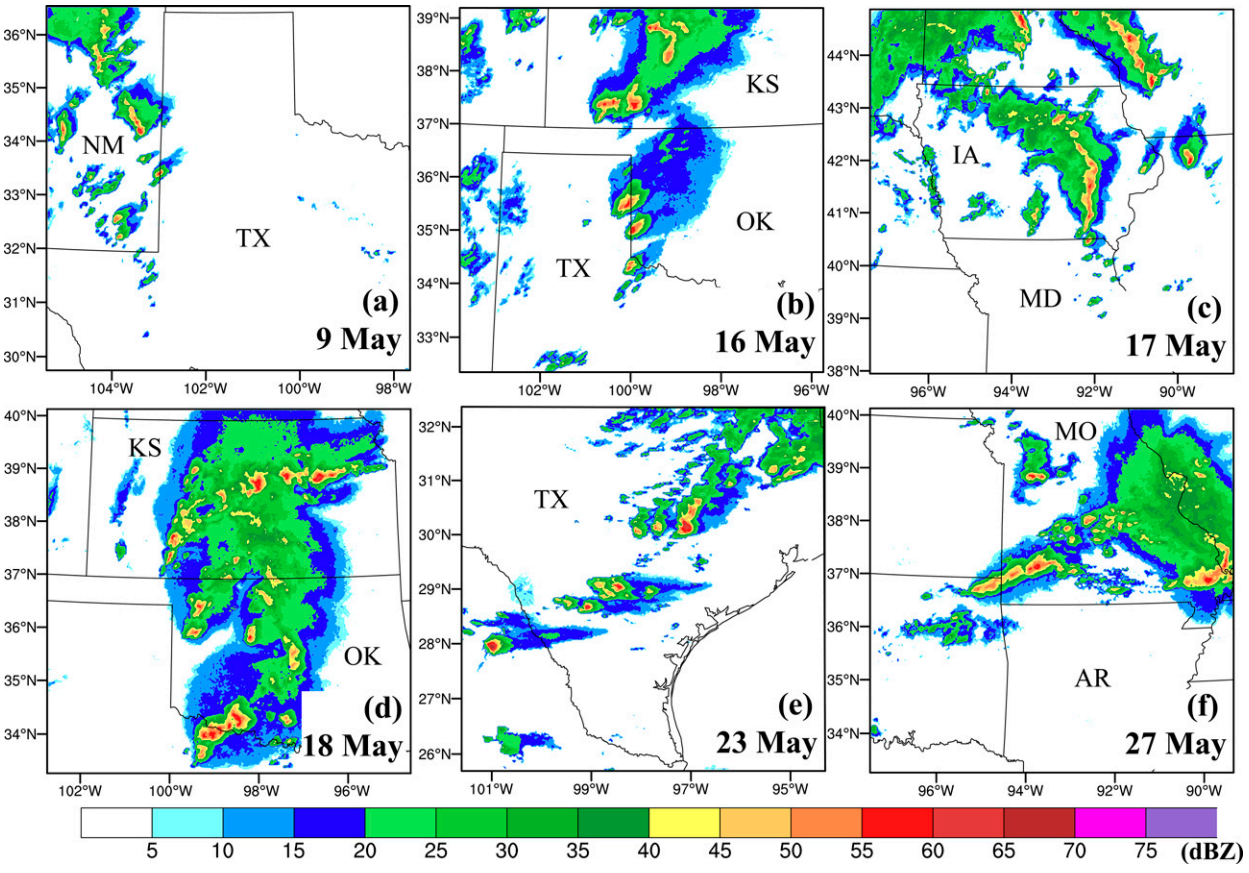

FIG. 1. MRMS observed column maximum Z (dBZ) valid at 2300 UTC for six cases.

data at every grid column are used here (no data thinning in the horizontal). These EnKF configurations have been tested and recommended by Labriola et al. (2020). Effectively the same localization radii are used for both EnKF and hybrid En3DVar, where the cutoff radii of EnKF are converted to equivalent recursive filter decorrelation length scales that are used for ensemble covariance localization (Pan et al. 2014). For the static BEC in hybrid En3DVar, the horizontal and vertical decorrelation scales are 1.6432 km and −0.0548 scale height, respectively. The negative value of scale height in GSI occurs because the background error statistics are initially read in at their original sigma levels and interpolated to log(sigma) coordinates on the analysis vertical sigma levels. Similar to Kong et al. (2018) and Wang et al. (2019), the EnKF system is one-way coupled to hybrid En3DVar, i.e., the hybrid En3DVar uses ensemble perturbations from the EnKF system, but does not feed back to the EnKF system with its deterministic analysis. One-way coupling keeps the EnKF DA cycles independent of the hybrid system, so that the performance of pure EnKF can be fairly compared with that of hybrid En3DVar.

A flowchart showing the DA cycles and forecasts for the EnKF and hybrid En3DVar experiments is shown in Fig. 2. Starting at 1800 UTC, gridded MRMS radar reflectivity data

TABLE 1. Event description and Storm Prediction Center (SPC) archived tornado, hail, and high-wind local storm reports from 1800 to 0500 UTC within the model domain for each case.

Date	Event description	No. of tornado reports	No. of hail reports	No. of high-wind reports
9 May 2017	Thunderstorms in eastern NM/TX Panhandle	2	12	1
16 May 2017	Discrete tornadic thunderstorms in KS/OK/TX	20	67	3
17 May 2017	Derecho in IA/MN	11	40	160
18 May 2017	Tornadic supercells in OK/KS	15	86	29
23 May 2017	High wind producing mixed mode convection in TX	0	12	8
27 May 2017	Mixed mode convection in AR/MO/OK	8	71	92

TABLE 2. Physics parameterization options for the 1–18 ensemble members of the EnKF experiments; members 19–36 repeat these 18 sets of physics options.

Member	Shortwave radiation	Longwave radiation	Planetary boundary layer
1	Dudhia	RRTM	YSU
2	RRTMG	RRTMG	YSU
3	Dudhia	RRTM	MYJ
4	RRTMG	RRTMG	MYJ
5	Dudhia	RRTM	MYNN
6	RRTMG	RRTMG	MYNN
7	Dudhia	RRTM	YSU
8	RRTMG	RRTMG	YSU
9	Dudhia	RRTM	MYJ
10	RRTMG	RRTMG	MYJ
11	Dudhia	RRTM	MYNN
12	RRTMG	RRTMG	MYNN
13	Dudhia	RRTM	YSU
14	RRTMG	RRTMG	YSU
15	Dudhia	RRTM	MYJ
16	RRTMG	RRTMG	MYJ
17	Dudhia	RRTM	MYNN
18	RRTMG	RRTMG	MYNN

and Level-II radial velocity data are assimilated every 15 min until 0000 UTC, with conventional observations assimilated hourly. The 3-h-long ensemble (deterministic) forecasts are launched every 30 min from the EnKF (3DVar, EnVar, and Hybrid) analyses at 1930–0000 UTC (i.e., 10 total forecasts for each case). Following Skinner et al. (2018), the ensemble initial forecast backgrounds (priors) are taken from the hourly-DA-cycled High-Resolution (i.e., 3 km) Rapid Refresh Ensemble (HRRRE; Dowell et al. 2016) 1-h forecasts (valid at 1800 UTC) initialized at 1700 UTC. The LBCs are taken from 9-member HRRRE forecasts issued at 1500 UTC and used four times to fill out the 36-member ensemble LBC. The initial deterministic forecast for the variational DA experiments starts from the ensemble mean of EnKF initial conditions at 1800 UTC. For radial velocity data, an automatic quality control procedure including velocity dealiasing developed by CAPS (Brewster et al. 2005) is applied.

TABLE 3. EnKF experiment configuration.

Configuration parameters	Expt configuration
Conventional localization radius (horizontal/vertical)	300 km/0.4 scale height
Radar localization radius (horizontal/vertical)	12 km/0.7 scale height
Observation errors (reflectivity/radial velocity)	6 dBZ/3 m s^{-1}
Inflation	99% relaxation-to-prior spread (RTPS)

3. Object-based verification methodology

The object-based verification system developed by Skinner et al. (2018) for WoFS is used to evaluate composite reflectivity and 30-min UH swath forecasts initialized by the various DA methods. The predicted composite reflectivity and UH swaths are used as proxies for the skill of experiments in forecasting convective storms and mesocyclones, respectively.

Considering that the dataset and model configurations used here generally follow Skinner et al. (2018), the same thresholds as Skinner et al. (2018) are adopted for object identification and matching. MRMS composite reflectivity data and cyclonic azimuthal wind shear data (Newman et al. 2013) with customized quality control (Skinner et al. 2018) are used as the observational verification dataset for composite reflectivity and UH forecasts, respectively. However, owing to differences between the forecast and verification fields (e.g., UH and azimuthal shear), different thresholds need to be defined for the model forecast and MRMS products. Skinner et al. (2018) constructed model climatologies of WoFS forecasts and corresponding MRMS observations for all real-time cases in 2017. Based on these climatologies, matched percentile values (e.g., the 99.95th percentile for UH swaths and azimuthal shear tracks) of MRMS observations and WoFS forecasts are identified as thresholds. Following Skinner et al. (2018), the forecast composite reflectivity objects are identified using a threshold of 45 dBZ and the UHs in the 2–5-km layers above ground level (AGL) use a threshold of 65.8 m^2 s^{-2}. The corresponding thresholds of MRMS composite reflectivity and

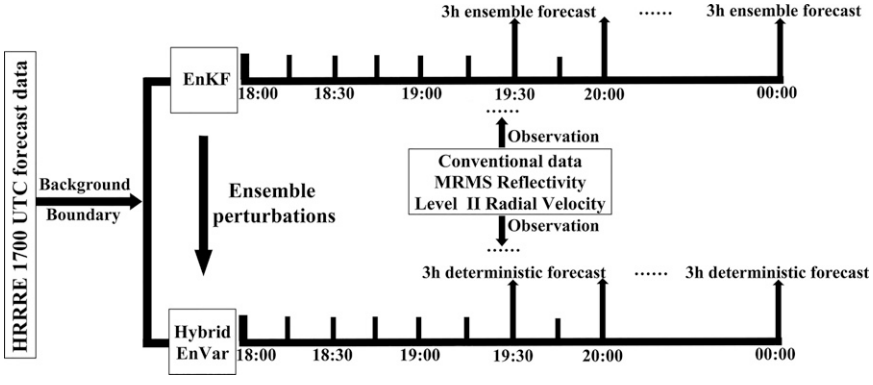

FIG. 2. The flowchart of the 6-h cycled assimilation and forecast experiments.

TABLE 4. The standard contingency table for verification scores.

	Observed (yes)	Observed (no)
Forecast (yes)	NA	NB
Forecast (no)	NC	ND

cyclonic azimuthal wind shear data in 2–5-km layers AGL are 41.2 dBZ and 0.0045 s^{-1}. It is noted that the thresholds from Skinner et al. (2018) will not correspond to the same percentiles in a distribution of values from the current study as it uses a different sample of cases and different system configurations, including a different microphysical parameterization. However, Skinner et al. (2018) found that while small changes to the object identification thresholds changed the absolute value of object-based verification scores, relative differences between two forecast datasets remained similar.

The object-based verification method is performed in three dimensions (2D area and time), forecast and verification objects are matched through a total interest score calculated by the centroid and minimum (defined as the smallest distance between any points in corresponding objects) spatial displacement and time displacement between object pairs (Davis et al. 2006; Skinner et al. 2016):

$$\mathrm{TI} = \left\{\frac{1}{2}\left[\frac{(\mathrm{cd}_{\max} - \mathrm{cd})}{\mathrm{cd}_{\max}}\right] + \frac{1}{2}\left[\frac{(\mathrm{md}_{\max} - \mathrm{md})}{\mathrm{md}_{\max}}\right]\right\}\left[\frac{(t_{\max} - t)}{t_{\max}}\right], \quad (1)$$

where TI is the total interest score; and cd, md, and t are the centroid distance, minimum distance, and time difference between an object pair, respectively. The max subscript indicates the maximum allowable threshold for object matching, which is set to 40 km for distance displacement and 25 min for time displacement. The matched pairs require a total interest score greater than 0.2, which is the same value used in Skinner et al. (2018).

Matched object pairs, unmatched forecast objects, and unmatched verification objects are classified as "hits," "false alarms," and "misses," respectively. Based on the standard contingency table (Table 4), verification scores such as probability of detection (POD), frequency bias (bias), false alarm ratio (FAR), and critical success index (CSI) are used to quantify the forecast skills, which can be calculated as follows:

$$\mathrm{POD} = \frac{\mathrm{NA}}{\mathrm{NA} + \mathrm{NC}}, \quad (2)$$

$$\mathrm{FAR} = \frac{\mathrm{NB}}{\mathrm{NA} + \mathrm{NB}}, \quad (3)$$

$$\mathrm{bias} = \frac{\mathrm{NA} + \mathrm{NB}}{\mathrm{NA} + \mathrm{NC}}, \quad (4)$$

$$\mathrm{CSI} = \frac{\mathrm{NA}}{\mathrm{NA} + \mathrm{NB} + \mathrm{NC}}, \quad (5)$$

where NA, NB, and NC represent "hits," "false alarms," and "misses," respectively. POD is the ratio of correct forecast objects to the total number of observed objects; FAR is the ratio of unmatched forecast objects to the total number of forecast objects; Bias is the ratio of the total number of forecast objects to the total number of observed objects; CSI is the ratio of correct forecast objects to the total number of forecast and observed objects. A perfect forecast means POD and CSI (FAR) values are 1.0 (0.0) while a useless forecast means POD and CSI (FAR) values are 0.0 (1.0); A bias value greater (less) than 1.0 presents overprediction (underprediction). A complete description of the object identification and matching methodology is provided in Skinner et al. (2018).

4. Object-based verification results

a. *Verification results of composite reflectivity*

The POD, bias, FAR, and CSI averaged across all 3-h forecasts launched between 1930 and 0000 UTC for all six cases are calculated by aggregating all object hits, misses, and false alarms. To qualitatively compare the forecast skill of the ensemble forecasts initialized from EnKF analyses with those of deterministic forecasts initialized from variational analyses, the minimum (EnKF_min), maximum (EnKF_max), and mean (EnKF_mean) values of the metrics from all ensemble members are calculated at each available forecast time. It is noted that the highest or lowest forecast skills EnKF_max or EnKF_min at any given forecast time are not necessarily from the same EnKF member.

Time series of object-based POD, bias, FAR, and CSI for composite reflectivity forecasts from EnKF, 3DVar, EnVar, and Hybrid experiments are plotted in Fig. 3, with 90% confidence intervals calculated every 15 min to determine statistical significance. In general, the POD for all experiments decreases with increasing forecast time after 30 min (Fig. 3a), although in the first 15 min there is a noticeable increase in the PODs of 3DVar and hybrid.

There is generally an overprediction bias at all forecast times for all experiments, and the bias is most pronounced in EnVar (Fig. 3b). During the first 30 min, the overprediction bias results in relatively high POD and FAR, which rapidly increases with forecast time (Figs. 3b,c), resulting in a decrease in CSI (Fig. 3d). The decrease in CSI during the first 30 min of forecast lead time is most pronounced in EnKF forecasts, suggesting rapid initial forecast error growth, likely due to development of spurious convection and/or over-intensification of convection.

To determine why the forecast error for each experiment rapidly increases during the first 30 min, the balance of model forecast states is evaluated by calculating the mean absolute surface pressure tendency (MASPT) (Hu and Xue 2007; Pan and Wang 2019) across all cases at every 5 min during the 3-h forecast. MASPT, which is averaged by the domain (for EnKF experiment, it is further averaged by the ensemble members), is defined by

$$\mathrm{MASPT} = \frac{1}{nx \times ny}\sum_{i=1}^{nx}\sum_{j-1}^{ny}\left|\frac{\partial P_s}{\partial t}\right|_{i,j}, \quad (6)$$

where P_s is the surface pressure forecast, and nx and ny are the dimensions of the domain. The MASPT value for all

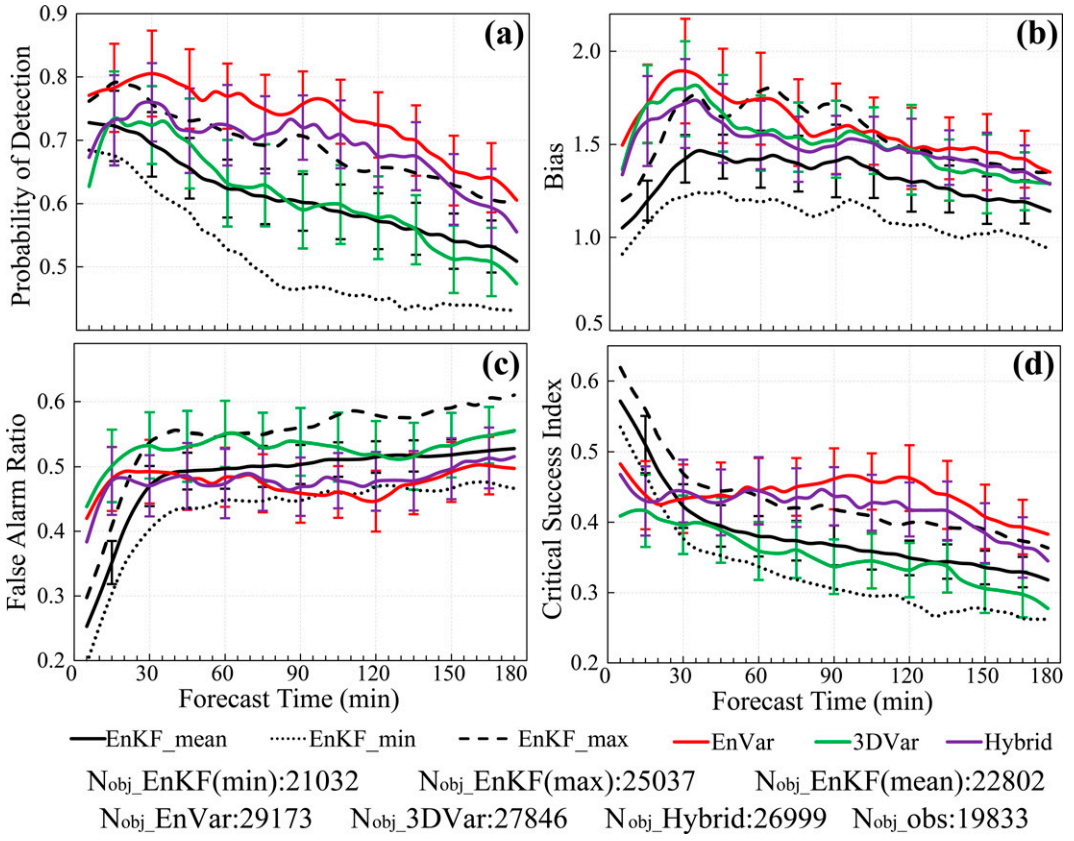

FIG. 3. Time series of object-based (a) POD, (b) bias, (c) FAR, and (d) CSI for composite reflectivity forecasts from EnVar (red), 3DVar (green), and Hybrid (purple) experiments, along with the total number of objects summed across all forecast valid times from each experiment with the observation count annotated at the bottom of the figure. For the EnKF experiments, the minimum (EnKF_min), maximum (EnKF_max), and mean (EnKF_mean) values of these metrics from all ensemble members at each available forecast time are provided. The colored bars with 15-min intervals represent the 90% confidence interval based upon one-sample *t* test.

experiments are initially large (~40 Pa min^{-1} for 3DVar, EnVar and Hybrid, and ~17 Pa min^{-1} for EnKF) but decrease with time to ~5 Pa min^{-1} as the model reduces the imbalance from DA (Fig. 4). The rapid decrease of MASPT during the first 30 min is related to the shock of DA and adjustment of the model state, which is known as the model spinup process, and can result in development of spurious reflectivity echoes, consistent with the rapid increase of FAR values during the first 30 min in Fig. 3c. In addition, the smallest (largest) MASPT occurs in EnKF (3DVar) experiments and indicates the EnKF (3DVar) analyses are more balanced (unbalanced) than others, which are again consistent with the evolution of FAR values in Fig. 3c.

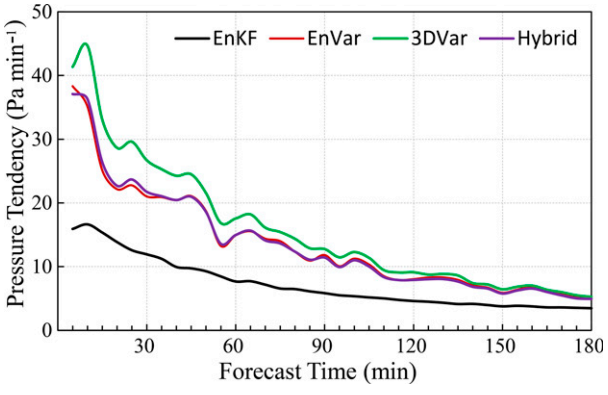

FIG. 4. The mean absolute surface pressure tendency during the 3-h forecast for the EnKF (black), EnVar (red), 3DVar (green), and Hybrid (purple) experiments.

For deterministic forecasts, Fig. 3 shows that EnVar (3DVar) has the highest (lowest) POD and the lowest (highest) FAR (comparing to the EnKF_mean, not the max or min), resulting in the highest (lowest) CSI. The POD and CSI of Hybrid forecasts fall between that of EnVar and 3DVar, while the bias is somewhat lower than EnVar, indicating the static background error covariance included in Hybrid does not generally improve the skill of thunderstorm object forecasts but can reduce the overprediction bias. The POD and CSI differences from 60 to 180 min of forecast time between EnVar (Hybrid) and 3DVar are statistically significant since their 90% CIs almost do not overlap, while the difference between EnVar and Hybrid is not statistically significant. For

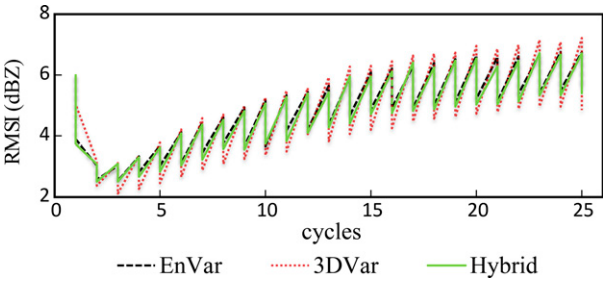

FIG. 5. The root-mean-square innovations (RMSIs) against reflectivity observations for the analyses/forecasts during the 6-h cycled data assimilation window with a 15-min interval for the EnVar, 3DVar, and Hybrid experiments.

the bias and FAR, the differences for all experiments are not statistically significant.

The root-mean-square innovations (RMSIs) of the background and analyzed reflectivity from DA cycling experiments are calculated for all grid points where the observed or analyzed reflectivity is greater than 15 dB*Z* (Fig. 5). During the analysis cycles, 3DVar always has a better fit of the analyzed reflectivity to observations than EnVar and Hybrid because the static background error variance in 3DVar tends to be larger than that estimated from the ensemble. However, the RMSIs of the 3DVar 15-min background forecast are larger than those of EnVar and Hybrid; the lack of balances among analyzed state variables with 3DVar, as suggested by Fig. 4, is suspected to be a main cause. Overall, the change in cycle-over-cycle RMSI is about the same between Hybrid and EnVar.

Figure 6 shows the RMSIs during the six hours of DA cycles for the background forecasts and analyses against surface observations. For wind, temperature and surface pressure, 3DVar has faster error growth than EnVar and Hybrid forecasts, likely because 3DVar cannot update these variables at upper levels using radar data owing to a lack of cross-variable covariance. For relative humidity, Hybrid underperforms EnVar and 3Dvar initially, but outperforms the other methods after two hours of DA cycling.

Among these six cases, a classic tornadic supercell case that occurred over Texas, Oklahoma, and Kansas 16–17 May 2017 is selected to compare the performance of EnKF, hybrid En3DVar, En3DVar, and 3DVar in greater detail (e.g., Wang et al. 2019; Chen et al. 2021). Figure 7 shows examples of the composite reflectivity object distributions for EnKF, EnVar, 3DVar, and Hybrid forecasts initialized at 2000 UTC 16 May 2017. The spread of ensemble forecast reflectivity objects in EnKF (Figs. 7a1–a4) encompasses observed objects. This spread results in clear differences between the minimum and maximum verification scores across ensemble members (e.g., the POD at 90-min forecast ranges from 0.29 to 1.0).

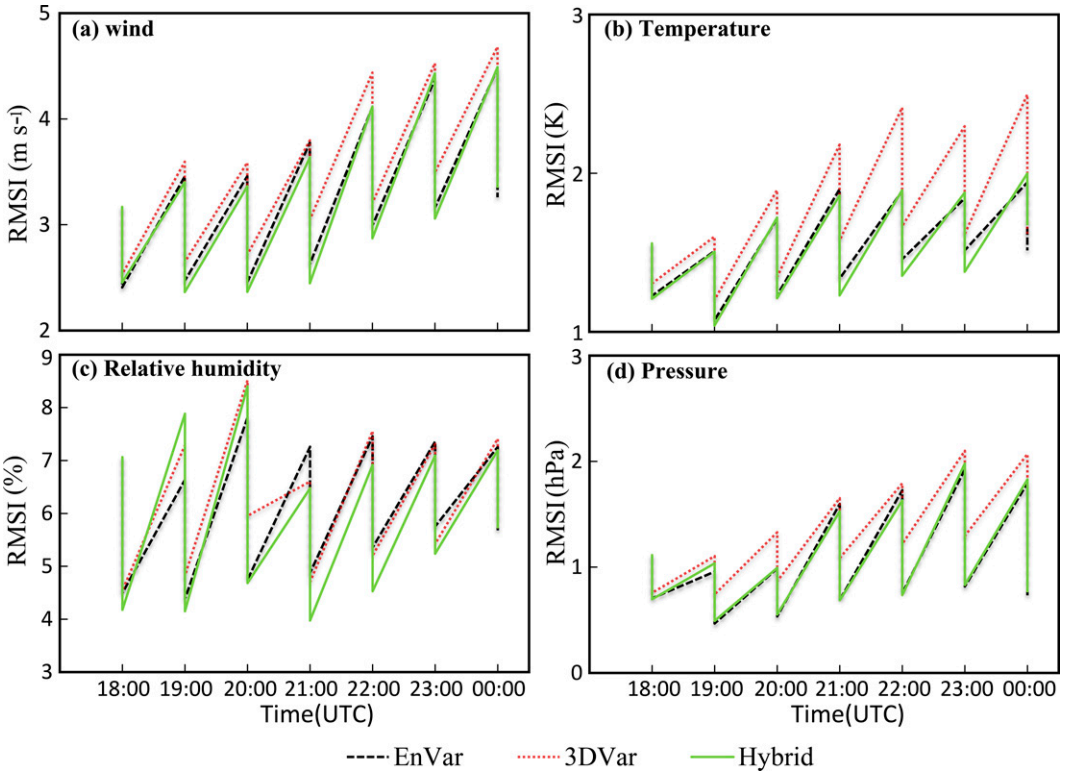

FIG. 6. The root-mean-square innovations (RMSIs) against surface (a) wind, (b) temperature, (c) relative humidity, and (d) pressure observations for the analyses/forecasts during the 6-h cycled data assimilation window with a 1-h interval for the EnVar, 3DVar, and Hybrid experiments.

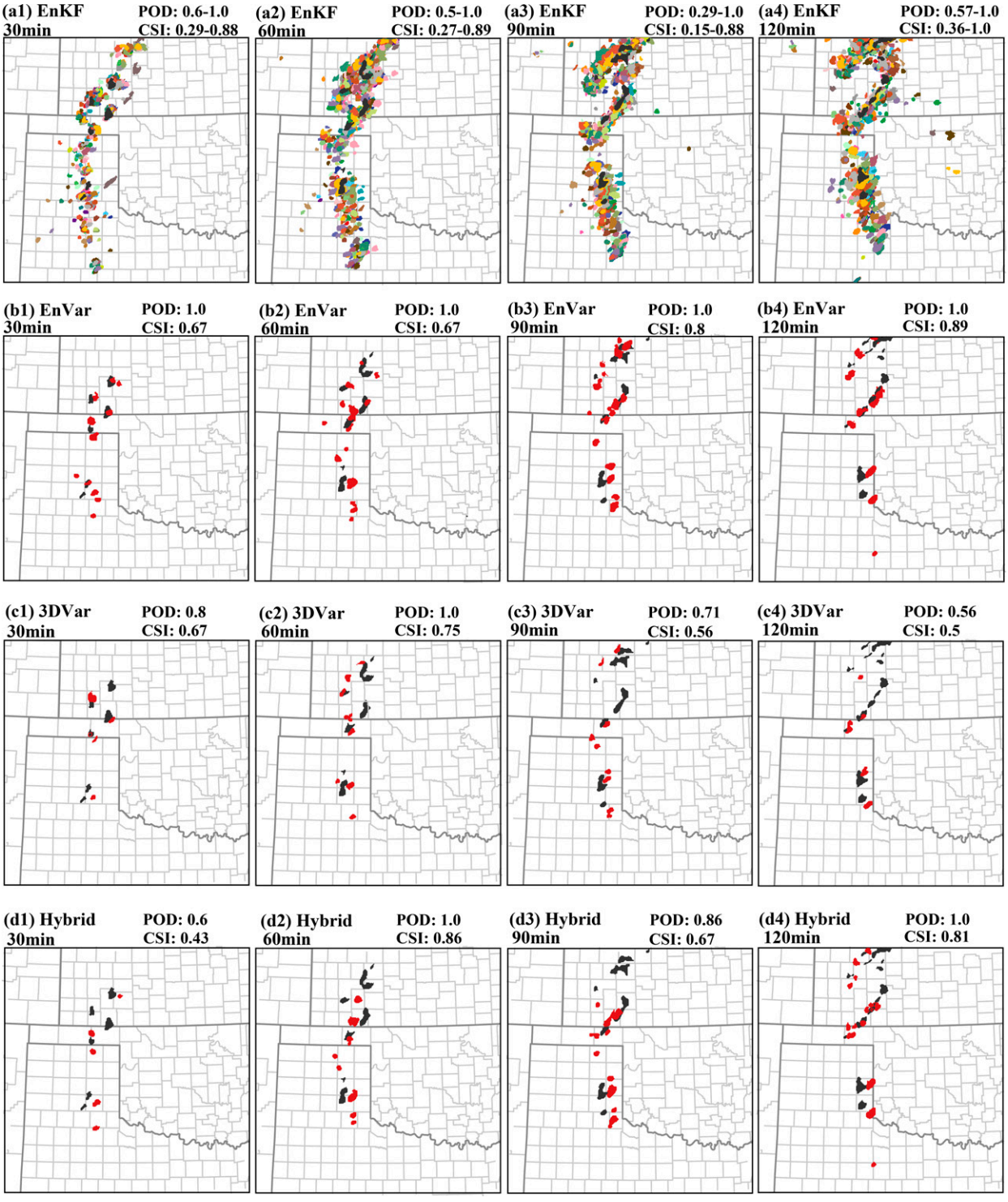

FIG. 7. Paintball plots of 30-, 60-, 90-, and 120-min composite reflectivity forecast objects initialized at 2000 UTC 16 May 2017 for (a1)–(a4) EnKF, (b1)–(b4) EnVar, (c1)–(c4) 3DVar, and (d1)–(d4) Hybrid experiments compared to the observed objects (black shading). Different colored shading in (a1)–(a4) indicates the forecast objects from each ensemble member of EnKF forecasts, and red shading in (b1)–(b4), (c1)–(c4), and (d1)–(d4), respectively, indicates forecast objects from EnVar, 3DVar, and Hybrid experiments. The POD, FAR, bias, and CSI scores of the EnVar, 3DVar, and Hybrid experiments, along with the minimum and maximum of these scores from all ensemble members of the EnKF experiments, are provided at the top right of each panel.

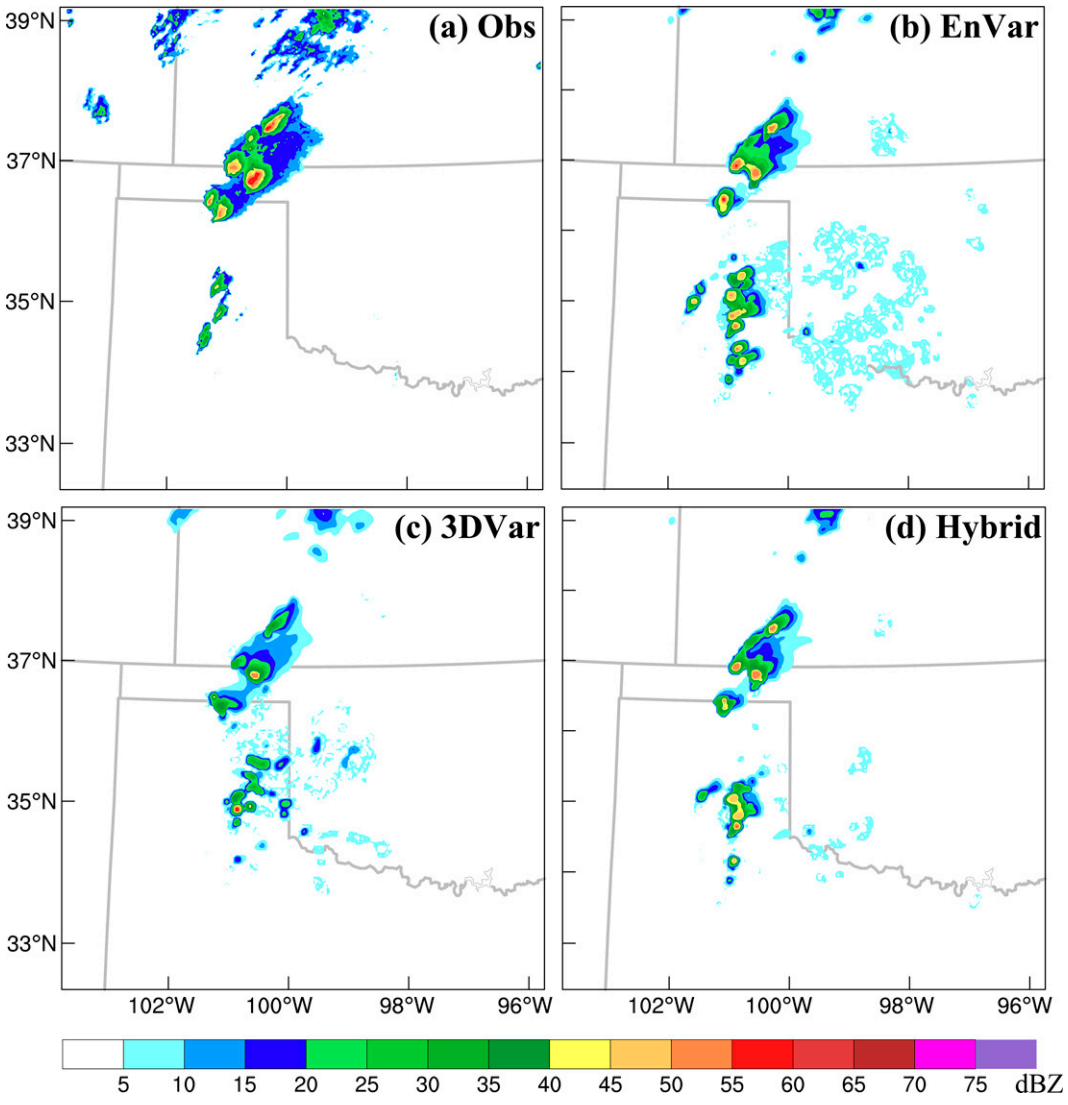

FIG. 8. (a) Composite reflectivity observations and the analyzed composite reflectivity at 2000 UTC 16 May 2017 for the (b) EnVar, (c) 3DVar, and (d) Hybrid experiments.

Most reflectivity objects are well predicted in the EnVar experiment (Figs. 7b1–b4). Though position errors are present in the forecasts of two supercells in Texas, the PODs of EnVar at all forecast times are 1.0 because position errors of forecast objects remain within prescribed displacement thresholds for object matching. Overall, the verification scores of EnVar are close to those of the best ensemble member in EnKF. Compared to the EnVar experiment, the forecast storms in the 3DVar experiment are not as well initialized and several observed objects are missing along the convective line in southwest Kansas at the 90- and 120-min forecast times. These misses are attributable to the inability of 3DVar to update state variables not directly involved in the observation operators, resulting in much lower POD and CSI than EnVar (e.g., the POD of EnVar at 90 min is 1.0 but only 0.56 for 3DVar). Compared to EnVar and 3DVar, the CSI of Hybrid is the highest (0.86) at 60 min but falls between them for other forecast times, indicating the static BEC is not always helpful for hybrid En3DVar.

Since variation in the initial conditions leads to differing meteorological behavior in the forecasts, the physical field analyses from DA experiments are additionally examined. Figure 8 shows the observed and analyzed composite reflectivity at 2000 UTC 16 May 2017 from the EnVar, 3DVar, and Hybrid experiment. In the observations, several convective storms are present near the Oklahoma Panhandle and in southwest Kansas, while a few other storms have recently initiated in the southern Texas Panhandle along a dryline. The intensity and coverage of analyzed reflectivity in the EnVar experiment best matches observations, with the exception of some spurious echoes in Texas Panhandle. The 3DVar analysis underpredicts the storm near the Oklahoma Panhandle

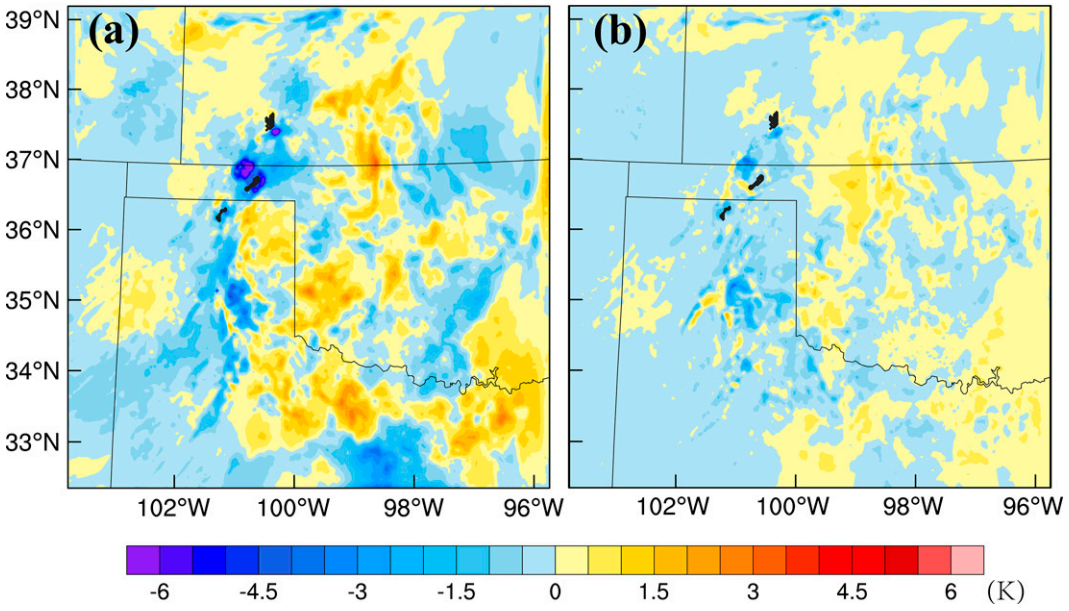

FIG. 9. The difference of analyzed air temperature (T) at the lowest level above surface at 2000 UTC 16 May 2017 between (a) EnVar and 3DVar and (b) EnVar and Hybrid experiments and paintball plots (black shaded) of 2–5-km updraft helicity observed objects.

and into southwest Kansas but does predict fewer spurious echoes in the Texas Panhandle. The analyzed reflectivity of Hybrid falls between EnVar and 3DVar, indicating that using static background error covariance can reduce the intensity of storms.

Because of the lack of cross-covariance in the static background error covariance, 3DVar assimilation can only update the variables involved in the observation operators (e.g., only update hydrometeors when assimilating reflectivity), which results in greater imbalance in analyzed state variables and rapid error growth in the forecast (Figs. 5 and 6). In contrast, EnVar and Hybrid assimilation contain ensemble estimated cross-variable covariances and update the full set of state variables, resulting in less imbalance in the analyses and improved forecasts.

To more comprehensively examine the impact of the cross covariance on the analysis variables that are not included in radar reflectivity operator, we compare analyzed surface temperature, vertical velocity, and cloud/ice mixing ratios from the DA experiments. Figure 9 shows the difference in temperature at the lowest model level between EnVar and 3DVar/Hybrid. It is clear that the temperature field near the observed storm in the Oklahoma Panhandle is cooler in EnVar and Hybrid compared to 3DVar, which indicates a stronger convective cold pool in the analysis and potentially results in a better forecast of the supercell. A vertical cross section of vertical velocity (through the maximum values in the supercell in the Oklahoma Panhandle) and cloud/ice mixing ratios from EnVar, Hybrid, and 3DVar are compared in Fig. 10. The updraft, cloud, and ice mixing ratios in the 3DVar analysis are weaker than those from EnVar and Hybrid, indicating a poor analysis of the storm intensity.

Overall, the reason that EnVar and Hybrid outperform 3DVar for the supercell in Oklahoma Panhandle is likely attributable to using ensemble-based cross covariances that produce analyzed state variables that are more physically consistent with each other, which promotes maintenance of the analyzed storm. However, similar improvements in forecasts of storms in the Texas Panhandle are not observed. This indicates that the ensemble-based cross covariance between reflectivity and other state variables is not improving analyses and resulting forecasts of these storms. The lack of benefit from the ensemble-based cross covariances in this region may result from relatively low reflectivities in developing storms (Fig. 8a) compared to other storms in the domain. It is noted that all variational experiments do not suppress spurious convection well. The limited suppression occurs because the variational DA method can struggle with minimization convergence related to the high nonlinearity of the reflectivity observation operator, making the suppression of weak, spurious echoes in background forecast of variational experiment less effective.

Even though many observed objects are well matched (e.g., the POD of EnVar, 3DVar, Hybrid, and the best EnKF member are 1.0 at 60-min forecast in Fig. 7), it does not mean that the objects are perfectly predicted. To more comprehensively characterize the performance quality, scatterplots of the centroid displacements of matched objects for 60-min forecast of composite reflectivity initialized at 2000 UTC 16 May 2017 for EnKF, EnVar, 3DVar, and Hybrid experiments are aggregated in Fig. 11, and the maximum (max) intensity and the area are aggregated in Fig. 12.

For EnKF, large variation in the centroid displacement of matched objects is present (Fig. 11). For variational DA

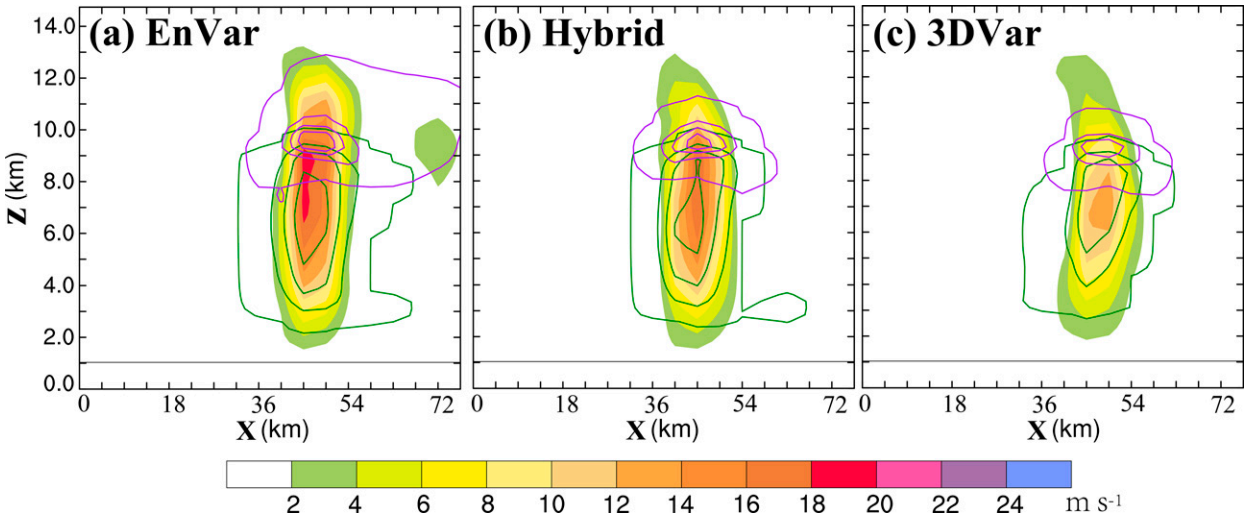

FIG. 10. Vertical cross sections of the analyzed vertical velocity field through their maximum values in the model domain (m s^{-1}; shaded), cloud mixing ratio (g kg^{-1}; green contour, from 0.01 to 10.01 with interval of 1.0 g kg^{-1}), and ice mixing ratio (g kg^{-1}; purple contour, from 0.01 to 9.01 with interval of 0.3 g kg^{-1}) at 2000 UTC 16 May 2017 for the EnVar, Hybrid, and 3DVar experiments.

experiments, EnVar has the most matched forecast objects. Nearly half of objects in EnVar present north and eastward bias in centroid displacement for nearly half of objects, indicating a positive bias in storm speed similar to the results of Skinner et al. (2018) and Flora et al. (2019). As shown in Fig. 12, much higher (smaller) maximum composite reflectivity (object size) values are produced by all experiments, with most objects exhibiting values between 55 and 65 dB*Z* (15 and 50 km^2), compared to 45–61 dB*Z* (15–135 km^2) for observation objects. The discrepancies in simulated object size and intensity compared to observations reflect biases in reflectivity calculation using the Thompson microphysics scheme and smoothing of MRMS observations in interpolation and are similar to biases in Skinner et al. (2018). Compared to EnVar, the lack of large object development in 3DVar and Hybrid is presented.

Performance diagrams (Roebber 2009) relating POD, FAR, CSI, and frequency bias of 60-min composite reflectivity

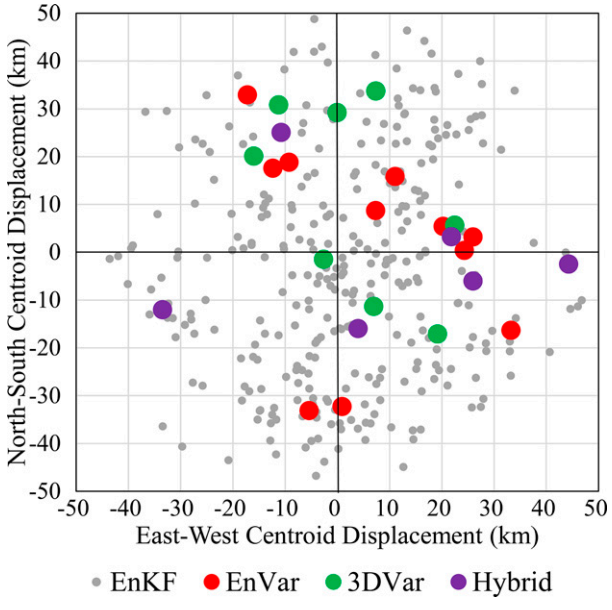

FIG. 11. Scatterplots of the east–west and north–south centroid displacement (km) of matched objects for 60-min forecasts of composite reflectivity (dB*Z*) initialized at 2000 UTC 16 May 2017 for the EnKF, EnVar, 3DVar, and Hybrid experiments.

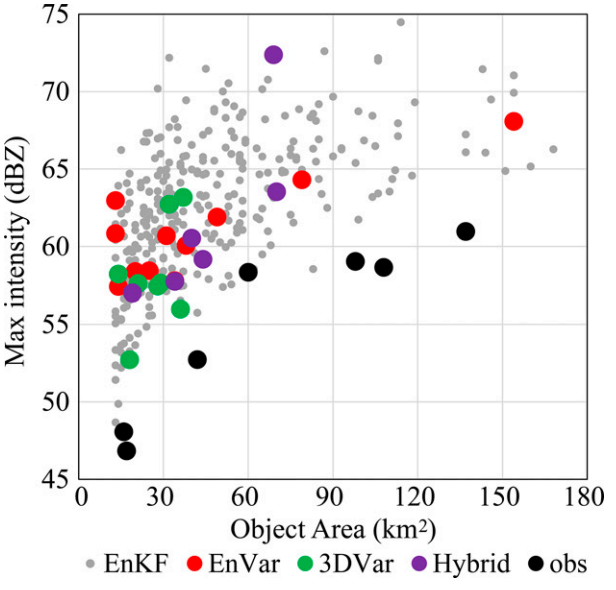

FIG. 12. Scatterplots of the object area (km^2) and maximum intensity (dB*Z*) of matched objects for 60-min forecasts of composite reflectivity (dB*Z*) initialized at 2000 UTC 16 May 2017 for the EnKF, EnVar, 3DVar, and Hybrid experiments and the observed objects at 2100 UTC 16 May 2017.

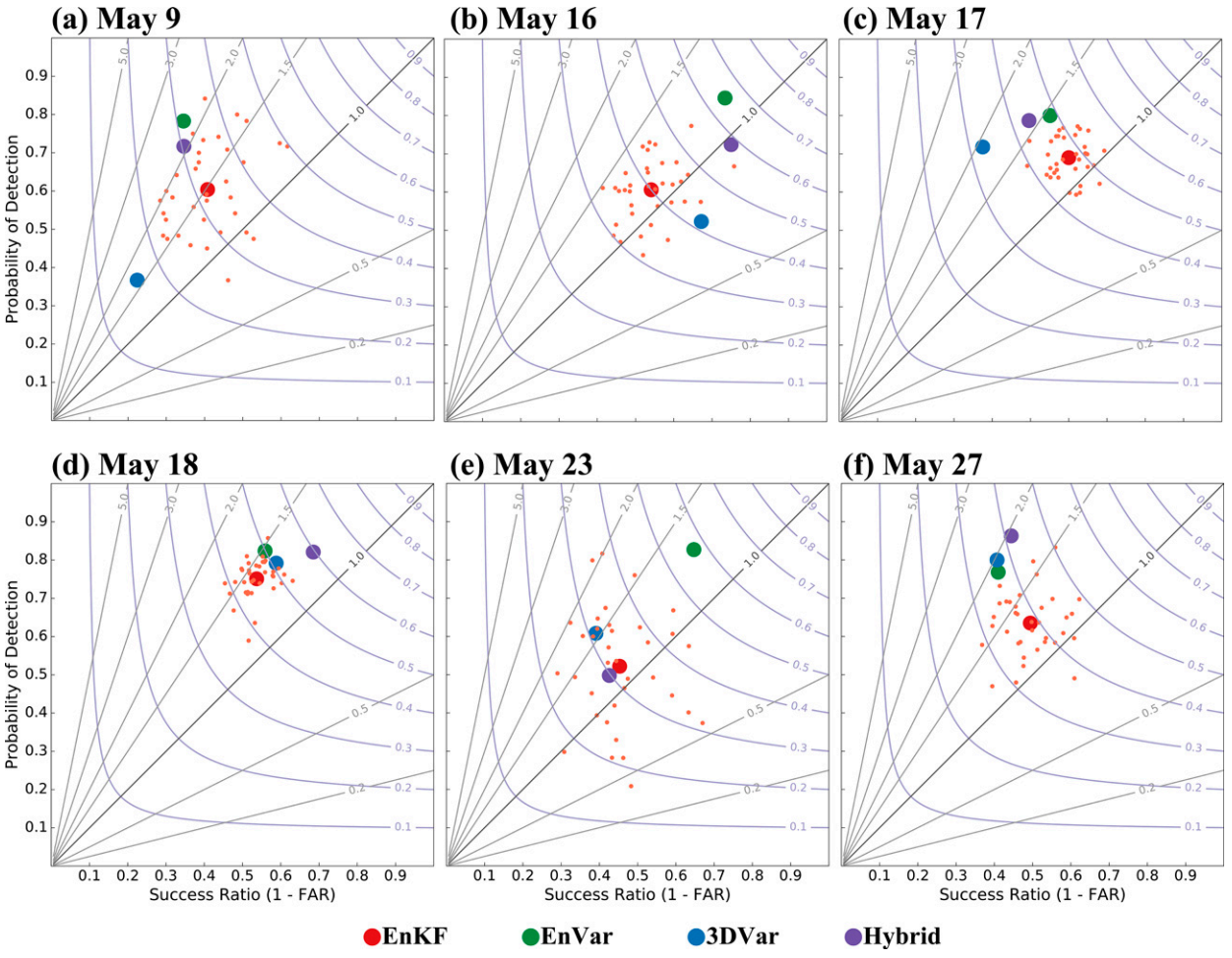

FIG. 13. Performance diagrams of 60-min composite reflectivity forecasts from each case for the EnKF, EnVar, 3DVar, and Hybrid experiments. The curved and diagonal lines represent CSI and bias, respectively. Small circles in EnKF plots represent scores of individual ensemble members, and large filled circles represent the ensemble mean of scores from each case.

forecasts for each case from EnKF, EnVar, 3DVar, and Hybrid experiments are shown in Fig. 13. An overprediction bias is present for all DA techniques and cases, with the exception of 3DVar forecast on 16 May 2017. For these outliers, there are several missing forecast objects in the convective line in southwest Kansas (Fig. 7), resulting in frequency biases below 1. For each case, large spread in object-based skill scores is apparent in the EnKF experiments (e.g., the POD and FAR values for the 9 May case range over approximately 0.35–0.85 and 0.4–0.7, respectively). For the 16 and 23 May cases, the CSIs of EnVar are higher than any individual ensemble member of EnKF. For the deterministic forecasts initialized from variational analyses (i.e., EnVar, Hybrid, and 3DVar), four cases (9, 16, 17, 23 May, Figs. 13a,b,c,e) show substantial improvement in EnVar compared to 3DVar, with comparable performances in the 18 and 27 May cases (Figs. 13d,f), in which CSI in Hybrid forecasts are somewhat higher than EnVar and 3DVar. Because there is no discernable, consistent distinction in the meteorological regimes of these two cases compared to the other four, sampling uncertainty may play a role here. This result suggests that static BEC in Hybrid may be helpful when the performance of 3DVar is at least comparable to EnVar. Otherwise, the performance of Hybrid may be degraded due to inclusion of static BEC.

To examine temporal variation in forecast skill, object contingency-table elements were aggregated across all available cases for each hourly forecast initialization time. Figure 14 shows the 3-h time series of object-based POD, bias, FAR, and CSI for composite reflectivity forecasts of each experiment aggregated for each forecast initialization hour from 2000 to 0000 UTC (as indicated by the horizontal axis labels). In general, for all experiments, POD and CSI increase while FAR decreases with successive initialization times between 2000 and 2100 UTC, and then POD and CSI decrease while FAR increases with successive initialization times between 2200 and 0000 UTC. Overall, EnVar and EnKF_max produce relatively high POD and CSI while EnKF_min produces the lowest values. At earlier initialization times, the FAR of EnKF forecasts are much lower than deterministic forecasts in the initial 30 min of the forecast, which is likely attributable

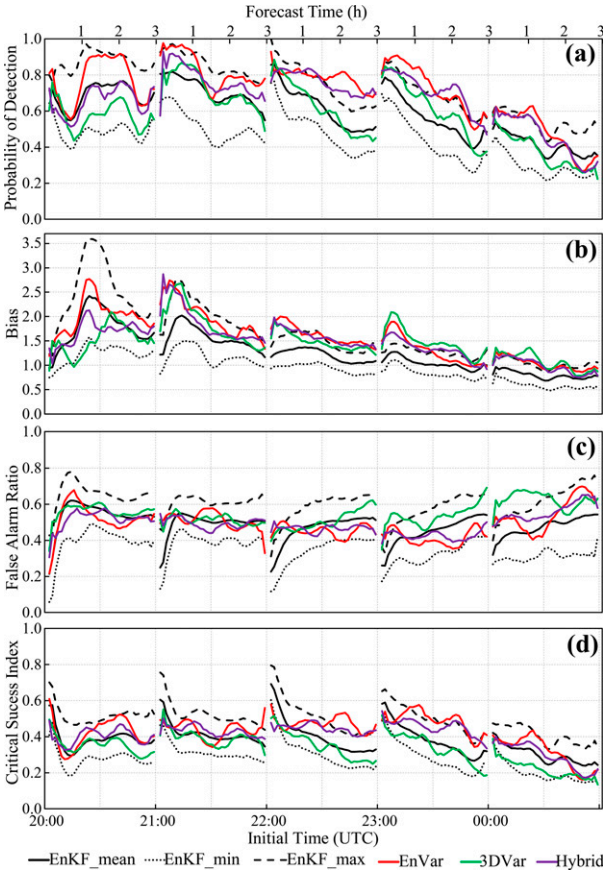

FIG. 14. The 3-h time series of object-based (a) POD, (b) bias, (c) FAR, and (d) CSI for composite reflectivity forecasts from EnVar (red), 3DVar (green), and Hybrid (purple) experiments aggregated for each forecast initialization time from 2000 to 0000 UTC (as indicated by the horizontal axis labels), along with the minimum (EnKF_min), maximum (EnKF_max), and mean (EnKF_mean) values of these metrics from all EnKF ensemble members at each available forecast time.

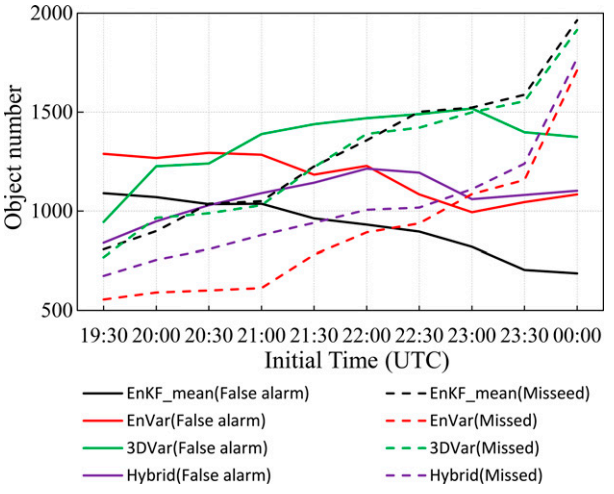

FIG. 15. The total number of false alarm and missed objects for composite reflectivity forecasts from EnVar (red), 3DVar (green), and Hybrid (purple) experiments aggregated across all three forecast hours for each forecast initialization time from 1930 to 0000 UTC (as indicated by the horizontal axis labels), along with the mean (EnKF_mean; black) of these metrics from all EnKF ensemble members at each available forecast time.

to the EnKF analysis being more balanced (Fig. 4) and development of fewer spurious reflectivity objects. It is noted that the EnVar FAR at the first 30 min for the 2000 UTC forecast are much lower than those of forecasts at later initial times, suggesting EnVar does not effectively suppress spurious echoes in the longer DA cycles compared to EnKF.

The POD, FAR, and CSI of 3DVar forecasts are comparable with those of the EnKF_mean at earlier initialization times; however, beginning in the 2200 UTC forecast, more spurious reflectivity objects are predicted by 3DVar, resulting in increased FAR and lower CSI than the other experiments in later forecasts. Because of the lack of cross covariances, 3DVar is generally not very effective in suppressing spurious convection in the background forecast (Tong and Xue 2005).

Positive biases are pronounced at the earlier initialization times but decrease with successive initialization times and become generally unbiased by 0000 UTC. Since reducing bias can be caused by either increasing misses or reducing false alarms, we examine the total number of false alarm and missed objects from EnKF_mean, EnVar, 3DVar, and Hybrid experiments aggregated for each forecast initialization time from 1930 to 0000 UTC (Fig. 15). It is shown that the missed objects substantially increase with successive initialization times because more observation objects are observed in the late afternoon but not predicted by the forecast, likely owing to erroneous prediction of convective initiation and delays for any DA method in spinning up developing thunderstorms (e.g., Tong and Xue 2005; Yussouf and Stensrud 2010; Stratman et al. 2020). In addition, in later DA cycles, the storm forecast can be susceptible to degradation owing to the lateral boundary condition.

b. *Verification results of updraft helicity*

Time series of object-based POD, bias, FAR, and CSI for 30-min swaths of predicted 2–5-km UH are shown in Fig. 16. As in Skinner et al. (2018), object-based verification scores for UH forecasts are generally lower than composite reflectivity forecasts, which is attributable to both predictability differences between mesocyclones and their parent thunderstorms as well as resolution limitations with 3-km horizontal grid spacing. The PODs of EnVar, 3DVar, Hybrid, and EnKF_mean decrease from approximately 0.3–0.45 to 0.25–0.35 by 165 min of lead time. Driven by a decrease of POD coupled with an increase of FAR, CSI values decrease from approximately 0.15–0.35 to 0.1–0.2. The relatively low scores in the first 15 min are related to forecast spin up of dynamical fields (vertical motion and vorticity) on the CAM grid from the relatively smooth analyses of wind fields. In contrast to the positive bias in composite reflectivity forecasts, the UH track forecasts for forecast times beyond 30–60 min are nearly unbiased. The bias

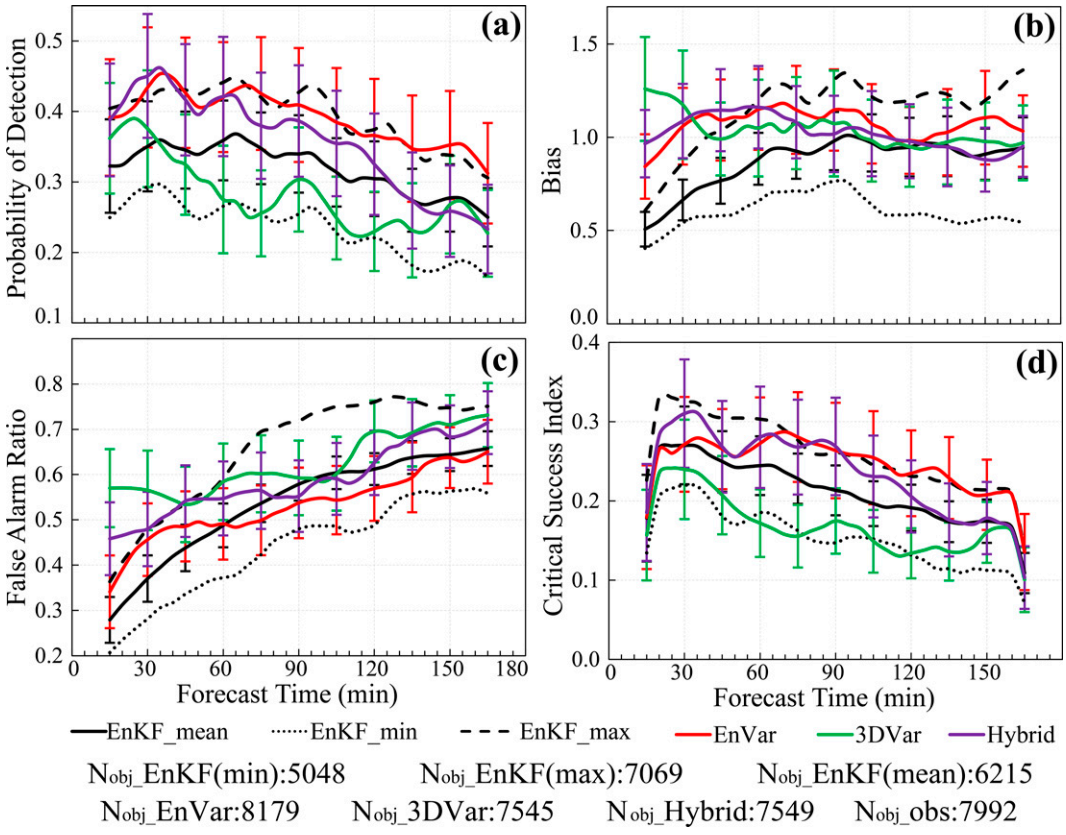

FIG. 16. As in Fig. 3, but for 2–5-km 30-min updraft helicity forecasts.

of EnKF_mean is low (about 0.5) during the first 90 min of the forecast and coupled with a low FAR (about 0.25); however, in the second half of the forecast, the FAR increases quickly, resulting in the forecast becoming nearly unbiased. Similarly to the composite reflectivity forecasts, the POD and CSI of rotation track forecasts from EnVar are comparable with the EnKF_max and values for 3DVar are similar to the EnKF_min. Hybrid values are comparable with EnVar during the first 2 h of the forecast and then become similar to EnKF_mean during the last hour. It is noted that the POD for 3DVar is similar to other experiments early in the forecasts, which indicates that 3DVar generally captures ongoing mesocyclones at the start of the forecast. The drop in POD and CSI (and increase in FAR) are then likely caused by the 3DVar-initialized forecasts dissipating UH swaths too early and initiating falsely predicted tracks.

To further compare the accuracy of 2–5-km 30-min UH forecasts from different experiments, the object distributions from forecasts initialized at 2000 UTC 16 May 2017 are plotted in Fig. 17. Compared to reflectivity, EnKF produces much fewer forecast objects for UH (Figs. 17a1–a4). At 30 and 60 min of lead time in EnVar, the UH objects in Kansas and Texas are missing (Figs. 17b1,b2) but they are accurately predicted at 90 and 120 min (Figs. 17b3,b4). 3DVar fails to accurately predict any observed rotation track objects during the first 2 h forecast time and fails to predict any UH tracks prior to the 90–120-min period (Figs. 17c1–c4), despite a substantial fraction of composite reflectivity objects are well predicted (Figs. 7c1–c4). The underpredicted rotation in the 3DVar experiment (Figs. 17c1–c4) may result from an inability to update background state variables through cross covariance with radar observations, such as vertical velocity and vorticity. Therefore, it is likely that the initial conditions do not support the rapid development of mesocyclones. The Hybrid forecasts generally perform better than 3DVar, but somewhat worse than EnVar for UH tracks near the border of Kansas and Oklahoma; when the ensemble-derived background error covariance is well behaved, the univariate static BEC used in this study may not be beneficial.

It is noted that there are many fewer false alarm UH forecast objects in Fig. 17 compared to composite reflectivity false alarms in Fig. 7, which suggests that many of the composite reflectivity false alarm objects are not rotating. To quantify this, we compared the forecast composite reflectivity objects to the UH objects using Eq. (1) and the results with the interest value are cross-referenced between forecast and observed composite reflectivity objects to determine which false alarm objects contain mesocyclones. The total and nonrotating false alarm object number for the composite reflectivity forecast aggregated for all cases and the nonrotating false alarm object ratios for all experiments are shown in Table 5. It is shown that nearly two-thirds of all false alarm composite reflectivity objects are nonrotating, suggesting that the model forecast overpredicts storms

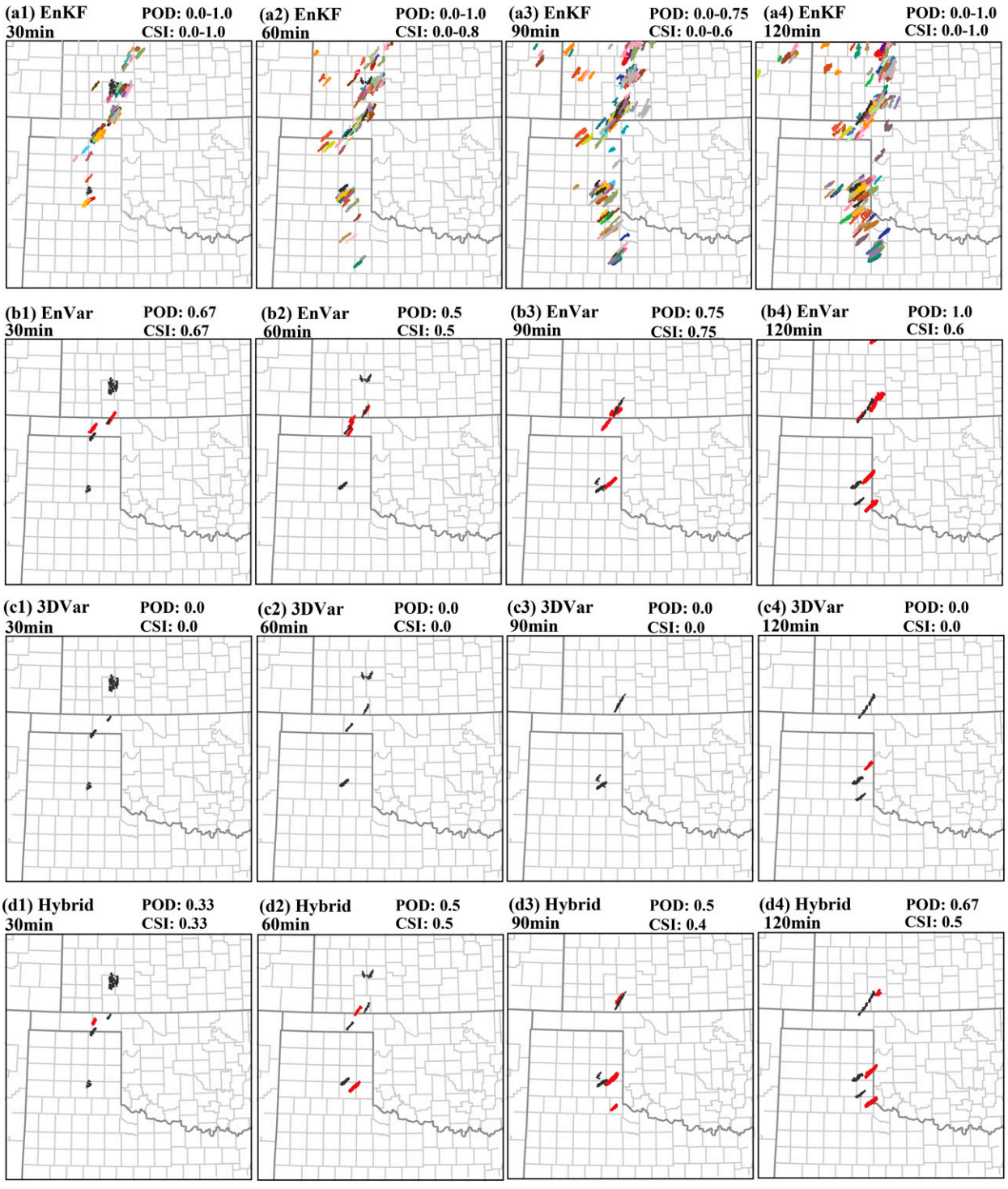

FIG. 17. As in Fig. 7, but for 2–5-km 30-min updraft helicity forecast and observed objects.

but a substantial fraction of those spurious storms do not contain mesocyclones and therefore, are likely not severe.

Performance diagrams of 60-min 2–5-km UH forecasts for each case from the four experiments are shown in Fig. 18. The object-based verification scores of 3DVar are much lower than those of En3DVar in four of the six cases, and are similar for 18 and 27 May. Compared to EnVar, a pattern similar to the composite reflectivity forecast is present in Hybrid CSI values, with higher scores on 9 and 18 May and similar or lower scores for the other four cases.

TABLE 5. The total and nonrotating false alarm object numbers for the composite reflectivity forecasts aggregated for all cases and nonrotating false alarm object ratios for the EnKF, EnVar, 3DVar, and hybrid experiments.

	EnKF_mean	EnVar	3DVar	Hybrid
False alarm object number	9190	11 330	13 269	10 605
Nonrotating false alarm object number	6370	7879	8629	3469
Nonrotating false alarm object ratio	69.3%	69.5%	65.1%	67.1%

Figure 19 shows the 3-h time series of object-based POD, bias, FAR, and CSI for 2–5-km UH forecasts for experiments aggregated for each forecast initialization time from 2000 to 0000 UTC (as indicated by the horizontal axis labels). Overall, the POD and CSI are much lower than those of composite reflectivity forecasts for all initialization times, reflecting the more challenging nature of mesocyclone prediction. At earlier initialization times, the EnKF bias is generally lower than that of variational methods, combined with lower POD, result in lower EnKF FAR. It is noted that the averaged bias for the 3DVar UH track forecasts initialized at 2000 UTC is actually at or above 1.0, while that of the 16 May 2017 case is essentially 0 (Fig. 17), indicating that the result for 3DVar in particular for the 16 May 2017 case is not representative of the average behavior.

5. Summary and conclusions

Capabilities to directly assimilate radar reflectivity and radial velocity data within the GSI EnKF and hybrid En3DVar systems were recently developed by CAPS and tested herein on six severe weather cases within a Warn-on-Forecast framework. The hybrid En3DVar method with different hybrid weights of 100%, 20%, and 0% for the static background error covariance, yielding, respectively, the pure 3DVar, hybrid En3DVar, and pure En3DVar algorithms, effectively, were intercompared as well as with EnKF. Specifically, object-based verification methods were used to evaluate forecasts of composite reflectivity and 30-min 2–5-km UH swaths against corresponding reflectivity and rotation track objects in Multi-Radar

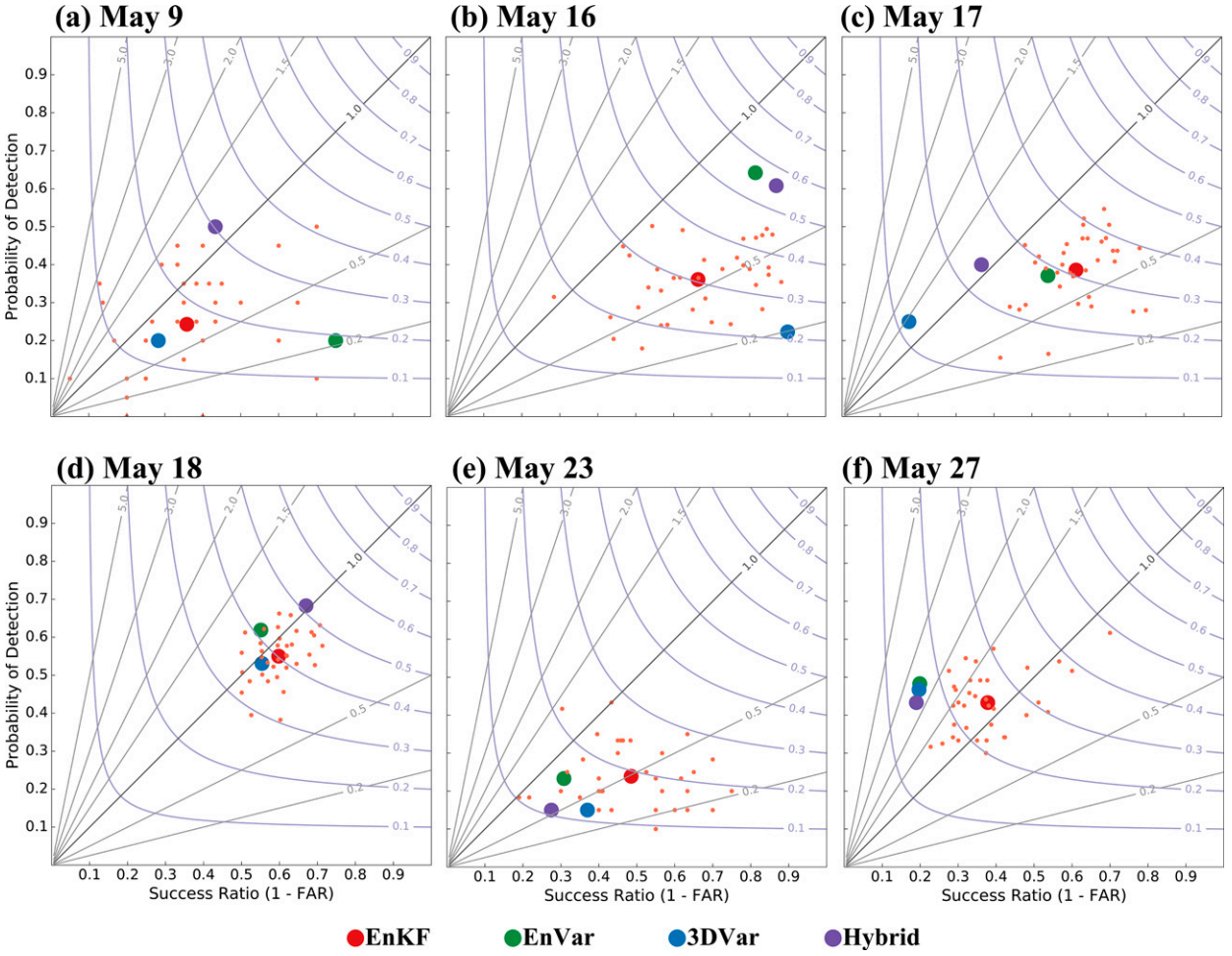

FIG. 18. As in Fig. 13, but for 2–5-km 30-min updraft helicity forecasts.

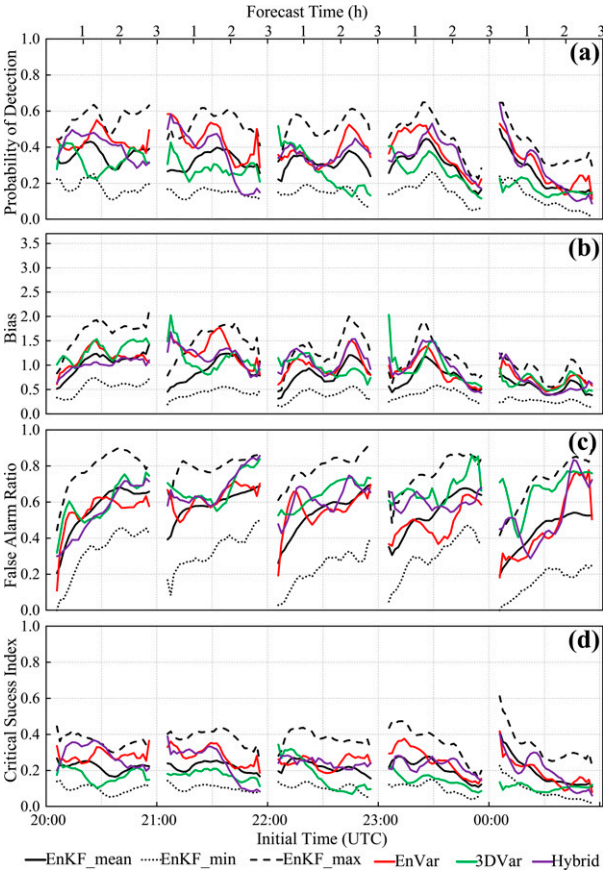

FIG. 19. As in Fig. 14, but for 2–5-km 30-min updraft helicity forecasts.

Multi-Sensor data on space and time scales typical of the National Weather Service severe weather warnings.

Even though 60 forecast samples and a 90% confidence interval are used for testing statistical significance in this study, 10 initializations per case are derived from the same cycle of DA, so that the samples are not completely independent. Consideration of more cases and use of a higher confidence interval will more robustly quantify significant differences in forecast quality resulting from different DA schemes. Such more systematic evaluations are warranted to help make decisions on eventual adoption of DA algorithms for operational WoF system. A summary of the conclusions is given below:

1) Despite variations across individual cases, the verification scores of EnVar in all six cases are generally similar to the best member of EnKF, with hybrid En3DVar scoring slightly lower than EnVar. 3DVar forecasts generally score lower than the mean score of EnKF members for composite reflectivity and close to the worst EnKF member for UH. This is not surprising because of the lack of information contained in the flow-dependent ensemble covariance in 3DVar that can help correct errors in state variables not directly linked to radar observables (Tong and Xue 2005). Even though the hybrid En3DVar with a 20% weight for static BEC generally underperforms pure En3DVar in terms of verification scores, it does produce highest verification scores in two individual cases (18 and 27 May), suggesting the potential for combining static and flow-dependent background error covariance to improve the analyses and subsequent forecasts. In this study, the static BEC does not contain cross-variable covariance, and was specified empirically. Better statistically derived static BEC may lead to better hybrid results, which will require further study. A potentially important benefit of including static BEC is when all ensemble members miss observed convection; in such a case, due to the complete lack of ensemble spread in the hydrometeor state variables, radar reflectivity observations at these locations will not be able to adjust hydrometeors with EnVar.
2) The FAR of EnKF analyses of both composite reflectivity and UH are lower than those of the deterministic analyses, indicating that EnKF suppresses spurious reflectivity and mesocyclone objects more effectively. Within EnKF, the ensemble cross covariances have been found to be effective in suppressing spurious background storms when assimilating clear-air reflectivity (Tong and Xue 2005). In addition, EnVar consistently has a higher POD and frequency bias than EnKF for both reflectivity and UH objects, as well as higher MASPT, suggesting EnVar more strongly overpredicts convective storm and mesocyclone occurrence.
3) In general, the PODs (FARs) for all reflectivity forecasts are greater (less) than 0.6 (0.5) for the first 60 min of lead time, then decrease (increase) with increasing forecast time, indicating that each method of cycled radar DA is able to accurately initialize preexisting thunderstorms.
4) The POD and CSI for mesocyclone forecasts are lower than those for composite reflectivity forecasts, indicating the systems are able to more accurately predict thunderstorm precipitation than mesocyclones.

For the hybrid DA method, the hybrid weight of static BEC and flow-dependent ensemble BECs is critical for the analysis. In this study, the weight of 20% for the static BEC may not be optimal. For radar DA, hybrid EnVar can benefit more from a smaller weight for static BEC, such as 5% (Kong et al. 2018). More effort is needed to determine an optimal hybrid weight for EnVar in convective-scale DA.

EnKF and hybrid EnVar in this study are one-way coupled, where EnKF perturbations are used by the hybrid En3DVar, but the hybrid analysis is not used to recenter the ensemble mean analysis of EnKF. Since the EnVar performance in this study is close to the best member of EnKF, future work will consider two-way coupling in which EnKF analyses may be improved by recentering the ensemble mean onto EnVar analysis. Two-way coupling may additionally benefit EnVar, since the EnKF perturbations may be improved. We chose one-way coupling in this study to keep the EnKF cycles independent from the hybrid DA cycles for easier comparison.

Acknowledgments. This research was supported by the Technology Innovation and Application Development Special

Program of Chongqing (cstc2019jscx-tjsb0007), the operational and technical program of Chongqing Meteorological Bureau (YWJSGG-202114), NOAA Warn-on-Forecast (WoF) Grant NA16OAR4320115, and NOAA Joint Technology Transfer Initiative (JTTI) Grant NA18OAR4590385. Dr. Kent H. Knopfmeier of NSSL is thanked for providing the ensemble initial conditions and lateral boundary conditions data for the selected cases. Computational resources of the Oklahoma Supercomputing Center for Research and Education (OSCER), and the NSF XSEDE Supercomputing Centers were used.

REFERENCES

Brewster, K., M. Hu, M. Xue, and J. Gao, 2005: Efficient assimilation of radar data at high resolution for short-range numerical weather prediction. *World Weather Research Program Symp. on Nowcasting and Very Short-Range Forecasting, WSN05*, Tolouse, France, WMO, 3.06.

Buehner, M., P. L. Houtekamer, C. Charette, H. L. Mitchell, and B. He, 2010a: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: Description and single-observation experiments. *Mon. Wea. Rev.*, **138**, 1550–1566, https://doi.org/10.1175/2009MWR3157.1.

——, ——, ——, ——, and ——, 2010b: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: One-month experiments with real observations. *Mon. Wea. Rev.*, **138**, 1567–1586, https://doi.org/10.1175/2009MWR3158.1.

——, J. Morneau, and C. Charette, 2013: Four-dimensional ensemble-variational data assimilation for global deterministic weather prediction. *Nonlinear Processes Geophys.*, **20**, 669–682, https://doi.org/10.5194/npg-20-669-2013.

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.

Chen, F., and J. Dudhia, 2001: Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Mon. Wea. Rev.*, **129**, 569–585, https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2.

Chen, L., C. Liu, M. Xue, R. Kong, and Y. Jung, 2021: Use of power transform mixing ratios as hydrometeor control variables for direct assimilation of radar reflectivity in GSI En3D-Var and tests with five convective storm cases. *Mon. Wea. Rev.*, **149**, 645–659, https://doi.org/10.1175/MWR-D-20-0149.1.

Clark, A., and Coauthors, 2020: A real-time, simulated forecasting experiment for advancing the prediction of hazardous convective weather. *Bull. Amer. Meteor. Soc.*, **101**, E2022–E2024, https://doi.org/10.1175/BAMS-D-19-0298.1.

Clayton, A. M., A. C. Lorenc, and D. M. Barker, 2013: Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Quart. J. Roy. Meteor. Soc.*, **139**, 1445–1461, https://doi.org/10.1002/qj.2054.

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

Dowell, D., and Coauthors, 2016: Development of a High-Resolution Rapid Refresh Ensemble (HRRRE) for severe weather forecasting. *28th Conf. on Severe Local Storms*, Portland, OR, Amer. Meteor. Soc., 8B.2, https://ams.confex.com/ams/28SLS/webprogram/Paper301555.html.

Etherton, B. J., and C. H. Bishop, 2004: Resilience of hybrid ensemble/3DVAR analysis schemes to model error and ensemble covariance error. *Mon. Wea. Rev.*, **132**, 1065–1080, https://doi.org/10.1175/1520-0493(2004)132<1065:ROHDAS>2.0.CO;2.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10 143–10 162, https://doi.org/10.1029/94JC00572.

——, 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367, https://doi.org/10.1007/s10236-003-0036-9.

Flora, M. L., P. S. Skinner, C. K. Potvin, A. E. Reinhart, T. A. Jones, N. Yussouf, and K. H. Knopfmeier, 2019: Object-based verification of short-term, storm-scale probabilistic mesocyclone guidance from an experimental Warn-on-Forecast system. *Wea. Forecasting*, **34**, 1721–1739, https://doi.org/10.1175/WAF-D-19-0094.1.

Gao, J. D., M. Xue, and D. J. Stensrud, 2013: The development of a hybrid EnKF-3DVAR algorithm for storm-scale data assimilation. *Adv. Meteor.*, **2013**, 512656, https://doi.org/10.1155/2013/512656.

Hamill, T. M., and C. Snyder, 2000: A hybrid ensemble Kalman filter–3D variational analysis scheme. *Mon. Wea. Rev.*, **128**, 2905–2919, https://doi.org/10.1175/1520-0493(2000)128<2905:AHEKFV>2.0.CO;2.

Hong, S.-Y., Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318–2341, https://doi.org/10.1175/MWR3199.1.

Houtekamer, P. L., and F. Zhang, 2016: Review of the ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **144**, 4489–4532, https://doi.org/10.1175/MWR-D-15-0440.1.

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.

Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res.*, **113**, D13103, https://doi.org/10.1029/2008JD009944.

Johnson, A., X. Wang, and S. Degelia, 2017: Design and implementation of a GSI-based convection-allowing ensemble based data assimilation and forecast system for the PECAN field experiment. Part II: Overview and evaluation of real-time system. *Wea. Forecasting*, **32**, 1227–1251, https://doi.org/10.1175/WAF-D-16-0201.1.

Jones, T. A., K. Knopfmeier, D. Wheatley, G. Creager, P. Minnis, and R. Palikondo, 2016: Storm-scale data assimilation and ensemble forecasting with the NSSL Experimental Warn-on-Forecast System. Part II: Combined radar and satellite data experiments. *Wea. Forecasting*, **31**, 297–327, https://doi.org/10.1175/WAF-D-15-0107.1.

Kain, J. S., and Coauthors, 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Wea. Forecasting*, **23**, 931–952, https://doi.org/10.1175/WAF2007106.1.

Kleist, D. T., D. F. Parrish, J. C. Derber, R. Treadon, W.-S. Wu, and S. Lord, 2009: Introduction of the GSI into the NCEP

global data assimilation system. *Wea. Forecasting*, **24**, 1691–1705, https://doi.org/10.1175/2009WAF2222201.1.

Kong, R., M. Xue, and C. Liu, 2018: Development of a hybrid En3DVar data assimilation system and comparisons with 3DVar and EnKF for radar data assimilation with observing system simulation experiments. *Mon. Wea. Rev.*, **146**, 175–198, https://doi.org/10.1175/MWR-D-17-0164.1.

Labriola, J., N. Snook, Y. Jung, and M. Xue, 2019: Explicit ensemble prediction of hail in 19 May 2013 Oklahoma City thunderstorms and analysis of hail growth processes with several multimoment microphysics schemes. *Mon. Wea. Rev.*, **147**, 1193–1213, https://doi.org/10.1175/MWR-D-18-0266.1.

——, ——, ——, and ——, 2020: Evaluating ensemble Kalman filter analyses of severe hailstorms on 8 May 2017 in Colorado: Effects of state variable updating and multimoment microphysics schemes on state variable cross covariances. *Mon. Wea. Rev.*, **148**, 2365–2389, https://doi.org/10.1175/MWR-D-19-0300.1.

Lawson, J. R., J. S. Kain, N. Yussouf, D. C. Dowell, D. M. Wheatley, K. H. Knopfmeier, and T. A. Jones, 2018: Advancing from convection-allowing NWP to warn-on-forecast: Evidence of progress. *Wea. Forecasting*, **33**, 599–607, https://doi.org/10.1175/WAF-D-17-0145.1.

Liu, C., M. Xue, and R. Kong, 2019: Direct assimilation of radar reflectivity data using 3DVAR: Treatment of hydrometeor background errors and OSSE tests. *Mon. Wea. Rev.*, **147**, 17–29, https://doi.org/10.1175/MWR-D-18-0033.1.

——, ——, and ——, 2020: Direct variational assimilation of radar reflectivity and radial velocity data: Issues with nonlinear reflectivity operator and solutions. *Mon. Wea. Rev.*, **148**, 1483–1502, https://doi.org/10.1175/MWR-D-19-0149.1.

Mansell, E. R., C. Ziegler, and E. Bruning, 2010: Simulated electrification of a small thunderstorm with two-moment bulk microphysics. *J. Atmos. Sci.*, **67**, 171–194, https://doi.org/10.1175/2009JAS2965.1.

Newman, J. F., V. Lakshmanan, P. L. Heinselman, M. B. Richman, and T. M. Smith, 2013: Range-correcting azimuthal shear in Doppler radar data. *Wea. Forecasting*, **28**, 194–211, https://doi.org/10.1175/WAF-D-11-00154.1.

Pan, Y., and M. Wang, 2019: Impact of the assimilation frequency of radar data with the ARPS 3DVar and cloud analysis system on forecasts of a squall line in southern China. *Adv. Atmos. Sci.*, **36**, 160–172, https://doi.org/10.1007/s00376-018-8087-5.

——, 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.

Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601–608, https://doi.org/10.1175/2008WAF2222159.1.

Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., https://doi.org/10.5065/D68S4MVH.

Skinner, P. S., L. Wicker, D. M. Wheatley, and K. H. Knopfmeier, 2016: Application of two spatial verification methods to ensemble forecasts of low-level rotation. *Wea. Forecasting*, **31**, 713–735, https://doi.org/10.1175/WAF-D-15-0129.1.

——, and Coauthors, 2018: Object-based verification of a prototype Warn-on-Forecast system. *Wea. Forecasting*, **33**, 1225–1250, https://doi.org/10.1175/WAF-D-18-0020.1.

Smith, T. M., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) severe weather and aviation products: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 1617–1630, https://doi.org/10.1175/BAMS-D-14-00173.1.

Snook, N., M. Xue, and Y. Jung, 2012: Ensemble probabilistic forecasts of a tornadic mesoscale convective system from ensemble Kalman filter analyses using WSR-88D and CASA radar data. *Mon. Wea. Rev.*, **140**, 2126–2146, https://doi.org/10.1175/MWR-D-11-00117.1.

——, Y. Jung, J. Brotzge, B. Putnam, and M. Xue, 2016: Prediction and ensemble forecast verification of hail in the supercell storms of 20 May 2013. *Wea. Forecasting*, **31**, 811–825, https://doi.org/10.1175/WAF-D-15-0152.1.

Stensrud, D. J., and Coauthors, 2009: Convective-scale Warn-on-Forecast system: A vision for 2020. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1500, https://doi.org/10.1175/2009BAMS2795.1.

——, and Coauthors, 2013: Progress and challenges with Warn-on-Forecast. *Atmos. Res.*, **123**, 2–16, https://doi.org/10.1016/j.atmosres.2012.04.004.

Stratman, D. R., N. Yussouf, Y. Jung, T. A. Supine, M. Xue, P. S. Skinner, and B. J. Putnam, 2020: Optimal temporal frequency of NSSL phased array radar observations for an experimental Warn-on-Forecast system. *Wea. Forecasting*, **35**, 193–214, https://doi.org/10.1175/WAF-D-19-0165.1.

Sun, J., and N. A. Crook, 1997: Dynamical and microphysical retrieval from Doppler radar observations using a cloud model and its adjoint. Part I: Model development and simulated data experiments. *J. Atmos. Sci.*, **54**, 1642–1661, https://doi.org/10.1175/1520-0469(1997)054<1642:DAMRFD>2.0.CO;2.

Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115, https://doi.org/10.1175/2008MWR2387.1.

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. G., T. A. Hamill, J. S. Whitaker, and C. H. Bishop, 2007: A comparison of hybrid ensemble transform Kalman filter-optimum interpolation and ensemble square root filter analysis schemes. *Mon. Wea. Rev.*, **135**, 1055–1076, https://doi.org/10.1175/MWR3307.1.

Wang, Y., 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.

——, J. Gao, P. S. Skinner, K. Knopfmeier, T. Jones, G. Creager, P. L. Heiselman, and L. J. Wicker, 2019: Test of a weather-adaptive dual-resolution hybrid Warn-on-Forecast analysis and forecast system for several severe weather events. *Wea. Forecasting*, **34**, 1807–1827, https://doi.org/10.1175/WAF-D-19-0071.1.

Wheatley, D. M., K. H. Knopfmeier, T. A. Jones, and G. J. Creager, 2015: Storm-scale data assimilation and ensemble forecasting with the NSSL Experimental Warn-on-Forecast System. Part I: Radar data experiments. *Wea. Forecasting*, **30**, 1795–1817, https://doi.org/10.1175/WAF-D-15-0043.1.

Yussouf, N., and D. J. Stensrud, 2010: Impact of phased-array radar observations over a short assimilation period: Observing system simulation experiments using an ensemble

Kalman filter. *Mon. Wea. Rev.*, **138**, 517–538, https://doi.org/10.1175/2009MWR2925.1.

——, 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 thunderstorm storm using single- and double-moment microphysics schemes. *Mon. Wea. Rev.*, **141**, 3388–3412, https://doi.org/10.1175/MWR-D-12-00237.1.