Geophysical Research Letters

Key Points:
- Direct assimilation of radar data is implemented in operational GSI system and interfaced with the new stand-alone regional FV3 model
- The capabilities with GSI 3DVar, EnKF, and hybrid and pure En3DVar are tested with a case of Central Plains tornadic storm cluster
- Overall, EnKF and hybrid En3DVar perform better than pure En3DVar, with 3DVar being significantly worse

Supporting Information:
- Supporting Information S1

Correspondence to:
M. Xue,
mxue@ou.edu

Direct Assimilation of Radar Data With Ensemble Kalman Filter and Hybrid Ensemble-Variational Method in the National Weather Service Operational Data Assimilation System GSI for the Stand-Alone Regional FV3 Model at a Convection-Allowing Resolution

Chong-Chi Tong[1] iD, **Youngsun Jung**[1,2,3] iD, **Ming Xue**[1,2] iD, **and Chengsi Liu**[1]

[1]Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, OK, USA, [2]School of Meteorology, University of Oklahoma, Norman, OK, USA, [3]Now at Office of Science and Technology Integration, NWS, NOAA, Silver Spring, MD, USA

Abstract Capabilities to directly assimilate radar radial velocity (V_r) and reflectivity (Z) data are implemented within the operational GSI data assimilation (DA) framework and coupled with the new stand-alone regional (SAR) FV3 model. The effectiveness and performance of 3DVar, EnKF, and hybrid En3DVar methods are evaluated with a storm cluster over the U.S. Central Plains at 3-km grid spacing. During the DA cycles, 3DVar analyses show better fit to Z observations but fastest error growth, while EnKF and pure En3DVar lead to smaller forecast errors. For V_r, EnKF outperforms other methods in both analysis and forecast. Good correspondence with tornado reports is obtained by most experiments for probabilistic forecast of updraft helicity (UH), except for 3DVar which shows insufficient confidence in certain regions. Overall, EnKF and hybrid En3DVar show best forecast skills in terms of composite reflectivity and UH. Tests with more cases are needed to draw more general conclusions, however.

Plain Language Summary The Finite-Volume Cubed-Sphere Dynamical Core (FV3) was chosen to serve as the single dynamical core for forecasts at all scale by the National Weather Service in late 2016. A stand-alone regional (SAR) version of FV3 became available in early 2019 and is planned to replace the current operational 3-km grid spacing High-Resolution Rapid Refresh (HRRR) system for convection-allowing forecasting at NCEP. As a key component in convective-scale numerical weather forecast (NWP) initialization, effective assimilation of radar data is crucial for desirable performances of the upcoming SAR FV3 model. In this paper, direct radar data assimilation (DA) capabilities implemented within the operational GSI DA framework by Center for Analysis and Prediction of Storms (CAPS), distinguished from the indirect approach applied in current operational HRRR to assimilate reflectivity data and believed to be more beneficial in particular for hydrometeor analyses, are interfaced with SAR FV3 model and the performances of different DA methods implemented are examined through the case study of a Central Plains tornadic storm cluster. Overall, results show good improvement of the forecasts, in terms of evolution of storm structures and occurrences of storm-related severe hazards, which greatly encourage acceleration of operational adoption of such capabilities.

1. Introduction

Radar observations are indispensable for initializing convective-scale numerical weather prediction (NWP); many studies have reported positive impact of radar data on short-range convection-allowing models (CAMs) using various data assimilation (DA) methods. Radial velocity (V_r) data are commonly assimilated using three- or four-dimensional variational (3DVar or 4DVar, respectively) methods due to the relatively straightforward observation operator of V_r (e.g., Gao et al., 2004; Hu et al., 2006; Sun & Crook, 1997). Direct assimilation of reflectivity (Z) data is more difficult with 3DVar/4DVar because of the dependency of Z on multiple hydrometeor species and the high nonlinearity of its observation operator that can cause a number of problems with variational minimization (Gao & Stensrud, 2012; Sun & Crook, 1997). Some of

the earlier studies retrieve hydrometeor mixing ratios from reflectivity before assimilation, leading to indirect assimilation of Z (e.g., Sun & Crook, 1997; Wang et al., 2013).

Gao and Stensrud (2012) modify the Z operator to alleviate the non-uniqueness problem within 3DVar assimilation of Z. More recently, Liu et al. (2019) propose the use of temperature-dependent hydrometeor background error profiles to deal with the same problem. Alternative to the variational approach, the ensemble Kalman filter (EnKF; Evensen, 1994) gained popularity in recent years thanks to its ability to provide flow-dependent multivariate background error covariances (BECs) without requiring the development of an adjoint model (e.g., Snyder & Zhang, 2003; Tong & Xue, 2005). Despite of its many advantages, EnKF does suffer from rank deficiency of the ensemble BEC due to the limited ensemble size that can be practically employed. A hybrid algorithm, first proposed by Hamill and Snyder (2000), that utilizes a weighted average of ensemble flow-dependent and static BECs is found to improve results over pure ensemble BEC (e.g., Buehner et al., 2010; Clayton et al., 2013). The hybrid DA method was first applied to radar DA by, for example, Gao et al. (2013) and Kong et al. (2018).

To more efficiently utilize developmental resources and mutually leverage efforts in NWP at different scales, the prevailing trend is to develop and maintain a single DA and modeling framework and apply it to NWP at an operational center ranging from convective through global weather and climate scales. In late 2016, the GFDL Finite-Volume Cubed-Sphere Dynamical Core (FV3; Putman & Lin, 2007) was chosen to serve as the single dynamical core for forecasts at all scale by the National Weather Service (NWS). The global FV3 dynamic core was implemented into the operational Global Forecasting System (GFS) at NCEP in 2019, while a stand-alone regional (SAR) version of FV3 became available in early 2019. A Rapid Refresh Forecasting system based on SAR FV3 is to replace the current operational 3-km grid spacing High-Resolution Rapid Refresh (HRRR) system (Benjamin et al., 2016) for convection-allowing forecasting. Currently, DA of HRRR uses the hybrid En3DVar method based on GSI (Hu et al., 2017) but assimilates radar reflectivity using a semi-empirical cloud analysis/diabatic initialization procedure (Benjamin et al., 2004; Weygandt & Benjamin, 2007). It is highly desirable to assimilate radar data directly within the hybrid En3DVar framework, taking advantage of the ensemble-derived flow-dependent covariances that are especially important for the analysis of hydrometeor state variables (Tong & Xue, 2005). Furthermore, coupling the hybrid GSI system with the newly developed SAR FV3 and evaluating the performance of such a coupled system are needed toward future operational implementation.

In this paper, we present results testing recently developed radar DA capabilities within the GSI framework using EnKF and hybrid En3DVar algorithms when coupled with SAR FV3 model, using a convective storm case that occurred over the Central Great Plains of the United States. While some of the DA capabilities were tested with the WRF model first and the results are reported elsewhere, it is the first time that they are coupled and evaluated with SAR FV3. Studies comparing different DA algorithms within the same DA framework for radar data are also relatively few (Kong et al., 2018) so far. The rest of this paper is organized as follows. In section 2, the 30 April 2019 test case is briefly introduced. Section 3 provides details about the data, model, and design of DA experiments. Results of the experiments are presented and evaluated in section 4. Lastly, a summary is provided in section 5.

2. 30 April 2019 Case Overview

Beginning in the morning of 30 April 2019, a cluster of severe thunderstorms started developing along a NE–SW-oriented stationary front passing through central Oklahoma. Around 1700 UTC, a number of tornadic supercells were initiated ahead of the frontal boundary in a supercell supportive environment with the mixed layer convective available potential energy (MLCAPE) exceeding 2,000 J kg^{-1}. Other damages inflicted by the storms include large hail and strong winds. The storm cluster and supercells continuously moved northeastward and gradually grew upscale into a mesoscale convective system (MCS) in central Oklahoma by 2300 UTC. Most tornadoes were spawned by supercells in northeast Oklahoma between 2000 and 2200 UTC before they merged into the convective line. Once the MCS formed, tornadoes were reported mostly in southwest Missouri.

3. Data, Model, and Experiment Design

The regional FV3 domain is centered on the Oklahoma state and consists of 450×450 grid points with ~3-km grid spacing in the horizontal (see Figure S1a) and 63 vertical layers extending up to ~1 hPa. Model physics options include the Mellor-Yamada-Nakanishi-Niino (MYNN) planetary boundary layer parameterization (Mellor & Yamada, 1982; Nakanishi & Niino, 2004, 2006), Thompson microphysics scheme, and Noah land surface model. Forty ensemble forecasts are initialized at 1800 UTC 30 April 2019 from the Global Data Assimilation System (GDAS) EnKF analyses and forced at the lateral boundary by the 6-hr forecasts. One-hour spin up forecasts are run before the first DA is performed at 1900 UTC and the DA cycles end at 2000 UTC.

To evaluate the impact of radar data on the short-term forecasts using various DA algorithms, we designed four experiments named after the DA method used: 3DVar, EnKF, pure ensemble 3DVar (pEn3DVar), and hybrid ensemble 3DVar (hEn3DVar). pEn3DVar uses 100% flow-dependent BEC derived from ensemble forecasts, and hEn3DVar uses 75% ensemble BEC as in the operational HRRR. An additional experiment, which is a free forecast launched from the ensemble mean forecast at 1900 UTC, is named the control experiment (CNTL) and serves as a reference. All DA experiments are performed using the GSI system. The Multi-Radar Multi-Sensor (MRMS) 3D Z mosaic product at 33 height levels from 500 m to 19 km MSL (Zhang et al., 2005) and level-II V_r data from 30 WSR-88D radars are assimilated at 15-min intervals, while conventional data including sounding, ship, aircraft, wind profiler, and surface observation are assimilated at the first and last analysis cycles. Automatic velocity dealiasing (Brewster et al., 2005) is applied to V_r data prior to DA.

For EnKF, the Z observation operator after Jung et al. (2008, hereafter JZX08) is used to produce Z observation priors consistent with the Thompson microphysics scheme. Prognostic variables updated by EnKF include three wind components (u, v, and w), pressure (p), temperature (t), specific humidity (q_v), mixing ratios of rainwater (q_r), cloud water (q_c), cloud ice (q_i), snow (q_s), graupel (q_g), and the total number concentration of rainwater (q_{nr}). In horizontal direction, covariance localization radii of 300 and 12 km are used for conventional and radar data, respectively. Vertically, length scales in a unit of logarithmic p of 1.1 (conventional) and 0.7 (radar) are used. For radar DA, localization radii ranging from 4 to 18 km, usually associated with couple times the grid spacing, are commonly used (Potvin & Wicker, 2013; Snook et al., 2015); the use of such short radii limits analysis increments from radar data within or close to the precipitation regions where the convective scale structures are characterized.

For 3DVar and En3DVar, the adjoint of the Z operator for Thompson microphysics scheme is not yet available; therefore, that for single-moment (SM) Lin scheme (Lin et al., 1983) is used. Linearization of the nonlinear Z observation operator in variational DA often causes difficulties in the convergence of cost function in the minimization process (Wang et al., 2013). To overcome these difficulties, logarithmic mixing ratios of hydrometeors as the control variables (hereafter CVlogq) proposed by Carley (2012) is used. With CVlogq, one outer loop with 50 inner loop iterations is found sufficient during the minimization of cost function. A two-pass procedure is applied to assimilate conventional and radar observation using different decorrelation length scales (l) with the variational methods. The first pass assimilates conventional data with l of 82 km and 0.3 (log p unit) in horizontal and vertical directions, respectively, to update model state variables u, v, p, t, and q_v. In the second pass, radar data are assimilated with considerably shorter horizontal (vertical) l of 3.29 km (0.2) and variables u, v, w, p, t, q_v, q_r, q_s, and q_g (q_c and q_i for En3DVar) are updated. The two-pass procedure is an alternative approach for realization of observation type dependent localization radii, as utilized in Pan et al. (2014) and Zhu et al. (2013) for assimilating conventional data using GSI. Note that the total number concentration of rainwater (q_{nr}), a prognostic variable in the Thompson scheme, is not updated; background values are used. The above localization length scales are equivalent to those used in EnKF based on the e-folding distance relation (see equation 4 in Pan et al., 2014).

In all DA experiments, observation errors for V_r and Z are set to 3 m s^{-1} and 5 dBZ, respectively. Temperature-dependent profiles of static background errors for hydrometeors (Liu et al., 2019) are used. More details about the experiments are presented in Figure S1b and Table S1.

4. Results of Experiments

4.1. Evaluation of Analyses

Artificial lateral boundary conditions (LBCs) often trigger numerical noise that propagates into the area of interests in the limited area models (Baumhefner & Perkey, 1982). The current version of SAR FV3 does not include a lateral boundary nudging zone which exuberates this problem. To avoid this issue, we perform forecast evaluation in an inner domain consisting of 300×300 grid points covering most of storms that caused weather hazards (see Figure S1a). For each DA experiment, the root mean square innovations (RMSIs) are computed against Z and V_r observations for analyses/forecasts during the 1-hr DA window and shown in Figure 1. The RMSI for CNTL is also presented as a reference. The Z operator of JZX08 is used to compute forecast Z for all experiments, while the corresponding Z operator used in DA is used to produce analyzed Z for evaluating how well the analysis fits observations.

In Figure 1b, the 3DVar analyses fit the observed Z most closely throughout the entire DA period; this is primarily because the static background error variances for the hydrometeor variables are specified and cannot be too small. However, the error rapidly grows in forecast (Figure 1a), mainly because 3DVar cannot update state variables not directly involved in the Z operator when assimilating Z, resulting in forecast RMSIs comparable to those of CNTL. Therefore, 3DVar loses the benefit of Z DA quickly. The RMSIs of hEn3DVar analyses are slightly higher than those of 3DVar, but the forecast error increases more slowly. The EnKF and pEn3DVar show much larger analysis RMSIs of Z, but the error growth rates during forecast are much smaller, indicating that the analyses of state variables are in better balance benefiting from the flow-dependent error covariances. For EnKF, larger analysis RMSIs are mainly because of underestimation of Z intensity in stratiform regions (Figure 2m). On the contrary, CVlogq used in pEn3DVar may cause overestimation of analysis increments during the inverse log transformation (Liu et al., 2020) and lead to higher Z cores (Figure 2j). The mismatch between Z operators based on the Thompson and Lin microphysics schemes used during the forecast and DA, respectively, is another likely source of errors in the variational systems. Initial positive bias in forecast reflectivity decreases quickly and stays low in EnKF, but all other experiments present higher positive bias than CNTL (Figure 1c).

In terms of RMSIs of V_r (Figures 1e and 1f), 3DVar exhibits noticeably larger errors than the other three experiments throughout the DA window. EnKF outperforms pEn3DVar and hEn3DVar especially for forecast RMSIs, but the differences are not as large as in Z. All DA experiments show smaller forecast bias in V_r than CNTL, but the magnitudes are smaller than 1 m s^{-1} (Figure 1g).

4.2. Evaluation of Forecasts

Deterministic forecasts of the storms in terms of composite Z for different DA experiments are presented in Figure 2 and verified against MRMS observations (Figures 2a–2c) for up to 4 hr. The EnKF forecast is launched from the ensemble mean analysis at the final cycle.

During the first 2 hr into the forecast, the scattered storms in the western half of the verification domain gradually organize into a quasi-linear convective line (QLCS) L1 (Figure 2b), which extends from eastern Kansas to central Oklahoma. Among all experiments, only 3DVar exhibits significant reduction in storm coverage within the first 15 min (not shown), especially in the stratiform regions in Kansas and Missouri. Then, storms redevelop and recover most of the storm coverage by 0000 UTC. However, the leading convective line passing through Oklahoma in 3DVar is least organized throughout the entire forecast period compared to other experiments. All experiments overforecast radar echoes in Texas in general. On the other hand, another convection line L2 ahead of L1, which propagates slowly northeastward during this period, is relatively well maintained in all experiments. Except for 3DVar, the storm structures from other experiments in the 2-hr forecast are comparable.

By 0000 UTC, 4 hr into the forecast, convective line L1 forms a solid line while L2 appears as a broken line, both oriented NE–SW across Oklahoma. In the meantime, storms are continuously initiated at the prefrontal convergence zone near north Texas border and then move northwest toward convective line L1 and eventually merge into the line. In terms of the structures of L1 and L2, EnKF performs better than other experiments. L1 and L2 are slightly less organized in hEn3DVar and pEn3DVar, but overall, the predicted structures of convective lines are reasonably good. However, the L1 propagates eastward noticeably slower than observations. All experiments overpredict convective initiation in the northern Texas where individual
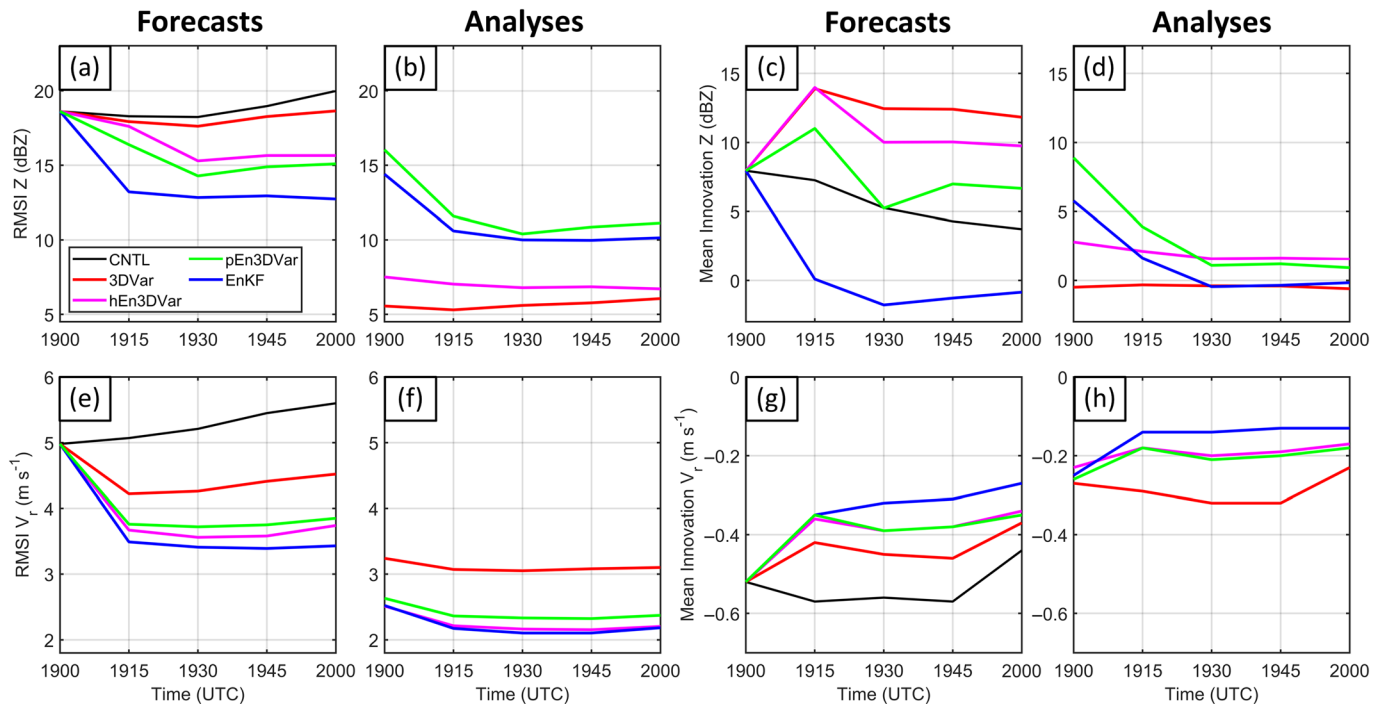
Figure 1. Observation space diagnostic statistics during the assimilation period for experiments 3DVar, hEn3DVar, pEn3DVar, EnKF, and the referential CNTL forecast. (a) Forecast and (b) analysis RMS innovation and (c) forecast and (d) analysis mean innovation for reflectivity. Panels (e), (f), (g), and (h) as in (a), (b), (c), and (d), but for radial velocity. For reflectivity, calculations are limited to locations where observed or forecast/analysis Z exceeds 15 dBZ.

cells move northeast fast. Combined with a slower propagation speed of L1, overpredicted storms in southeast Oklahoma and northern Texas develop into a north–south-oriented mesoscale convective line instead of supercell storms shown in observations.

It should be pointed out that the predicted convective cores in all experiments are more scattered and less organized than in observations. This can be attributed to the Thompson microphysics scheme used; the scheme takes into account the mixing ratio and median volume diameter of supercooled water when diagnosing graupel intercept parameter. This often leads to overestimation of reflectivity for graupel in the updraft above the freezing level (Johnson et al., 2016). Similar findings are also indicated in Skinner et al. (2018).

Equitable threat score (ETS) is the matrix that is most commonly used for precipitation-related forecast performance evaluation. Following Clark et al. (2010), we compute neighborhood-based ETSs for composite Z with a ~42-km radius (Figure 3a), which is consistent with the Storm Prediction Center (SPC) forecast products. The 25-dBZ threshold is applied to account for all precipitation exceeding light rain. This also prevents clear air from dominating the scores. Overall, there are distinct differences among the forecasts. Throughout the entire 4-hr forecast, EnKF generally produces the highest ETSs, while 3DVar scores the lowest almost throughout. hEn3DVar scores lower than both EnKF and pEn3DVar in the first hour and then gets very close to EnKF beyond 1 hr. The score of pEn3DVar falls below those of EnKF and hEn3DVar to be close to 3DVar after 3.5 hr. The low skill of 3DVar forecast is most likely due to the delayed rebuilding or misforecast of some major storms which is particularly significant in the early stage of the forecast. The scores of 3DVar and hEn3DVar show quick drop in the first 15 min, which is not seen in EnKF and pEn3DVar; such a behavior is most likely related to the use of inconsistent Z operator and the lack of cross-variable updating associated with the static BEC. The relatively poor performance of pEn3DVar after 2.5 hr is related to the significant increase in false alarm rate and decrease in hit rate (refer to Figure S2) in its forecast. Overall, EnKF and hEn3DVar show higher sustained ETS scores.

Contingency table-based verification indices including bias, probability of detection (POD), false alarm ratio (FAR), and critical success index (CSI) at 4 forecast hours are presented in a performance diagram
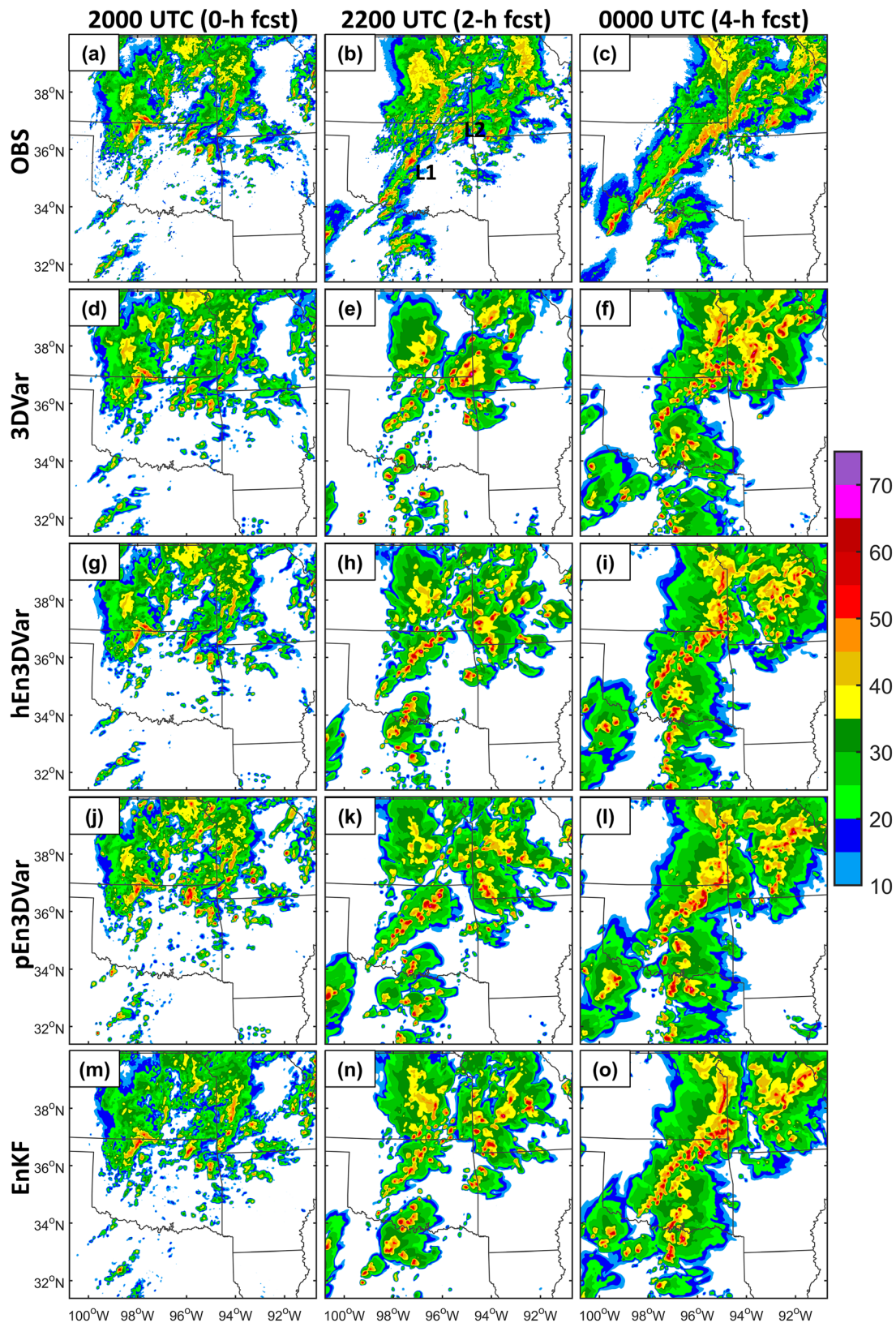
Figure 2. Composite reflectivity (dBZ) analyzed and predicted by experiments (d, e, and f) 3DVar, (g, h, and i) hEn3DVar, (j, k, and l) pEn3DVar, and (m, n, and o) EnKF, as compared to (a, b, and c) MRMS observation. Corresponding times are marked at the top of each column. Convective lines of interest are assigned with L1 and L2 in (b).
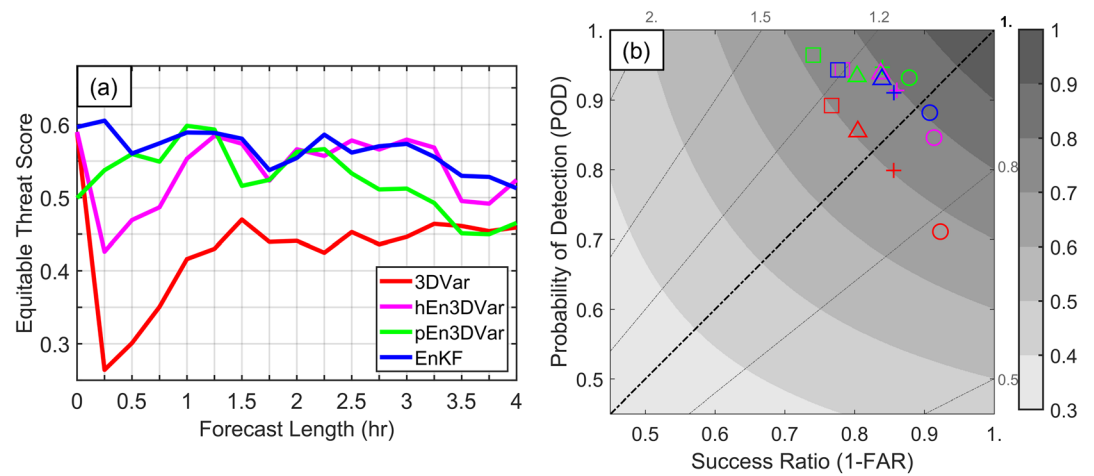
Figure 3. (a) Neighborhood ETSs of 4-hr forecast composite reflectivity at a 25-dBZ threshold for experiments 3DVar, hEn3DVar, pEn3DVar, and EnKF, verified at 15-min intervals. (b) Performance diagram of composite *Z* at 1- (circle), 2- (cross), 3- (triangle), and 4-hr (square) forecast for 3DVar (red), hEn3DVar (magenta), pEn3DVar (green), and EnKF (blue). Dashed lines and shading contours represent constant bias and CSI, respectively.
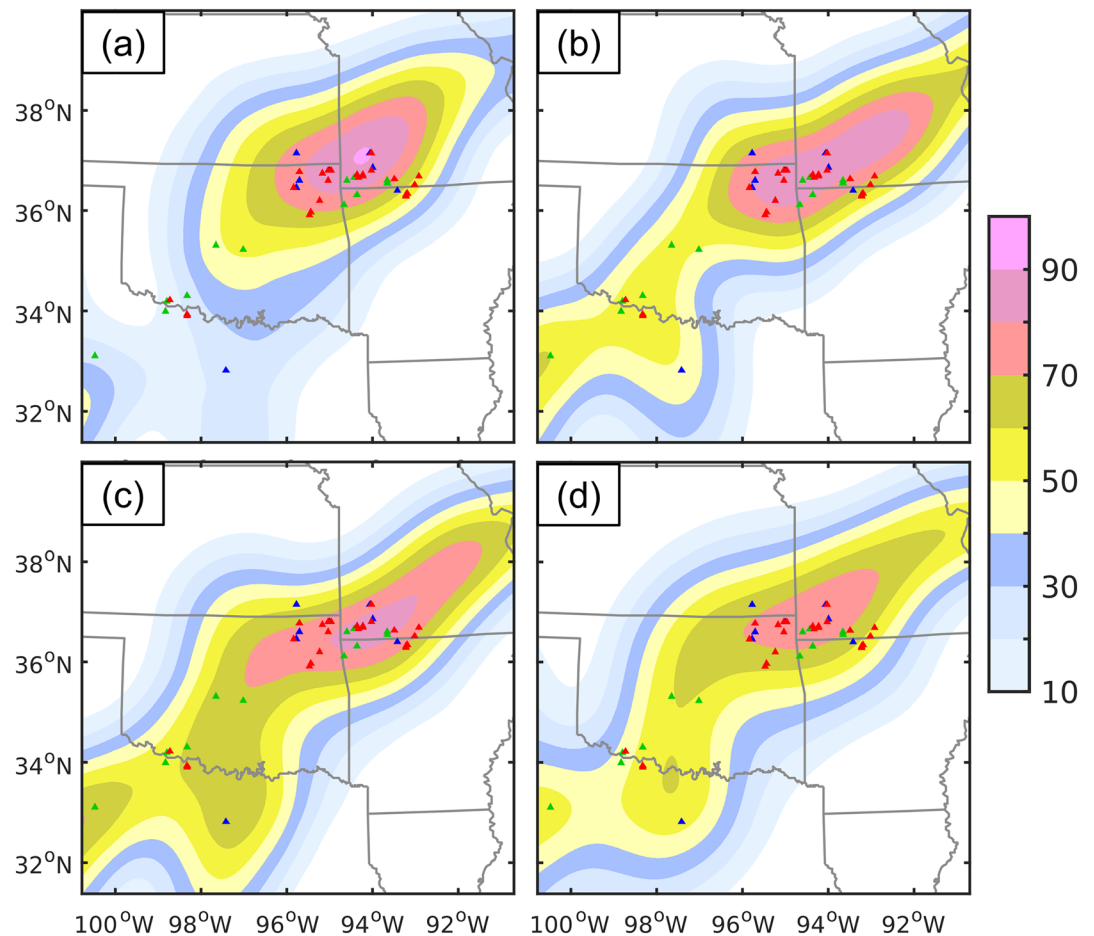

Figure 4. Neighborhood probability (%) of 4-hr maximum UH exceeding 75 m^{-2} s^{-2} in color shading for experiments (a) 3DVar, (b) hEn3DVar, and (c) pEn3DVar. (d) Neighborhood ensemble probability of 40-member 4-hr max UH exceeding 75 m^{-2} s^{-2} from EnKF. The SPC reports of tornado, wind, and hail during the same period are overlaid in red, blue, and green triangles, respectively.

(Roebber, 2009) in Figure 3b. The grouping of CSI generally matches the ETS results, indicating again the lowest (highest) performance of 3DVar (EnKF and hEn3DVar) for 4 verification hours. For almost all forecast lengths, pEn3DVar shows the greatest POD, followed by hEn3DVar and EnKF, which are very closely comparable especially from hour 2. In terms of the bias score (as divided by the diagonal dashed line), overforecasting is seen to increase with forecast length for all experiments and appears to be greatest with pEn3DVar. The findings in these quantitative matrices are generally consistent with earlier subjective evaluations based on Figure 2.

Lastly, severe storm prediction is examined in terms of the neighborhood probability (NP) of the 4-hr maximum 2- to 5-km updraft helicity (UH) shown in Figure 4. The probability fields are first generated with a UH threshold of 75 m^{-2} s^{-2} and a ~42-km neighborhood radius and then smoothed with a two-dimensional Gaussian filter (Brooks et al., 1998) with a ~81-km σ (Gaussian kernel). General correspondence is found between high NP and SPC severe weather reports. Certain location displacement, however, is presented in different DA experiments. Specifically, 3DVar (Figure 4a) exhibits a more marginal probability in forecasting the tornadoes near southwest Oklahoma border compared to other experiments. Generally speaking, the NP forecast of hEn3DVar (Figure 4b) appears to have the best match with the major group of reported events near the four state corners, and it shows the largest area of NP $\geq$ 80%. For both hEn3DVr and pEn3DVar, all observed tornadoes occurred within regions with NP $\geq$ 50%. The most extensive distribution of NP $\geq$ 30% in pEn3DVar (Figure 4c) can be linked to its greatest overforecast tendency.

Additionally, performance of the 40-member ensemble forecasts from EnKF is evaluated in terms of the neighborhood ensemble probability (NEP) (Schwartz et al., 2010) and is presented in Figure 4d. Overall, all observed tornadoes except for one occurred within regions with NEP $\geq$ 50%. However, given the additional dispersion among ensemble members, a lower peak in probability is found compared to the NP derived from single deterministic forecasts of other experiments (Figures 4a–4c). Nevertheless, the NEP of EnKF appears to be more discriminative than some of the other experiments.

5. Summary

Capabilities developed by CAPS to directly assimilate radar data in the GSI EnKF and En3DVar systems are interfaced with the new SAR FV3 model and evaluated with a tornado-spawning storm cluster over the Central Plains at ~3-km grid spacing; this effort aims to accelerate the adoption of more advanced DA methods for next-generation FV3-based operational regional forecasting systems. One-hour frequent cycling analysis is performed. Radar reflectivity and radial velocity data are assimilated every 15 min over a 1-hr DA window, while conventional data are assimilated hourly. Four DA experiments that use GSI 3DVar, EnKF, pure En3DVar (pEn3DVar), and hybrid En3DVar with 75% ensemble covariance (hEn3DVar) are conducted, with model physics similar to the operational HRRR system.

Results show that direct assimilation of radar data, regardless the methods used, significantly improves both analyses and forecasts during the 1-hr DA window, in terms of the RMSIs of both Z and V_r compared to the no-DA control forecast. Among all DA experiments, 3DVar and hEn3DVar produce significantly better fit of analyses to Z observations, while EnKF and pEn3DVar produce smaller forecast error growth rates between analysis cycles. The larger forecast error growth in 3DVar is most likely due to the lack of analysis updating to state variables not directly involved in the Z observation operator. In terms of V_r RMSI, 3DVar shows distinctly worst analyses and forecasts, while differences among other three experiments are relatively insignificant. Overall, within the DA window, EnKF produces the best fit to V_r observations in analyses and the smallest forecast error growth rate.

Forecasts up to 4 hr for the DA experiments are further examined. Subjective evaluation of the forecast storm evolutions and the quantitative ETS verifications of composite reflectivity together suggest superior performance of EnKF and pEn3DVar within the first 15-min forecast, given their better storm maintenance and increasing ETS. This suggests the value of flow-dependent BEC derived from the ensemble. Beyond the first hour, EnKF and hEn3DVar produce the best reflectivity forecasts. In terms of the performance diagram,

pEn3DVar shows the highest POD and also the highest positive bias, while the performances of EnKF and hEn3DVar are similar and all much better than 3DVar.

Lastly, severe weather prediction skill is evaluated in terms of the NP of 4-hr maximum 2- to 5-km UH. All DA experiments produce high NP (≥50%) over a cluster of severe weather reports near the northeast corner of Oklahoma with hEn3DVar producing the largest high NP (≥80%) area. Several severe weather reports near southwestern Oklahoma are also captured by EnKF, hEn3DVar, and pEn3DVar. Overall, the NP of hEn3DVar matches the severe weather distribution the best. The ensemble NP from EnKF has generally smaller value due to additional ensemble smoothing. In summary, this study demonstrates, for the first time according to our knowledge, positive impacts of directly assimilating radar data using GSI-based advanced DA methods on SAR FV3 model forecasts. This helps accelerate operational adoption of such capabilities.

Finally, we point out that while important, understanding the algorithm differences and their relative performance is not the primary goal of this paper; achieving such a goal requires more comprehensive experiments, more detailed analyses, and full length papers, as in, for example, Kong et al. (2018) and Pan et al. (2014; for conventional data). Given that the relative performances of the different DA algorithms observed in this paper, while mostly agreeing with expectations, are based on a single case, the robustness of the conclusions requires tests with more cases and performance statistics over larger samples. Such evaluations should be carried out in future studies.

Data Availability Statement

Reflectivity data used in this study are available online (ftp://ftp.caps.ou.edu/ctong_2020_grl).

References

Baumhefner, D. P., & Perkey, D. J. (1982). Evaluation of lateral boundary errors in a limited-domain model. *Tellus*, *34*(5), 409–428. https://doi.org/10.3402/tellusa.v34i5.10828

Benjamin, S. G., Devenyi, D., Weygandt, S. S., Brundage, K. J., Brown, J. M., Grell, G. A., et al. (2004). An hourly assimilation-forecast cycle: The RUC. *Monthly Weather Review*, *132*(2), 495–518. https://doi.org/10.1175/1520-0493(2004)132<0495:AHACTR>2.0.CO;2

Benjamin, S. G., Weygandt, S. S., Brown, J. M., Hu, M., Alexander, C. R., Smirnova, T. G., et al. (2016). A North American hourly assimilation and model forecast cycle: The rapid refresh. *Monthly Weather Review*, *144*(4), 1669–1694. https://doi.org/10.1175/MWR-D-15-0242.1

Brewster, K., Hu, M., Xue, M., & Gao, J. (2005). *Efficient assimilation of radar data at high resolution for short-range numerical weather prediction*. Paper presented at the Int. Symp. on Nowcasting Very Short Range Forecasting, Toulouse, France.

Brooks, H. E., Kay, M., & Hart, J. A. (1998). *Objective limits on forecasting skill of rare events*. Paper presented at the 19th Conf. on Severe Local Storms, Minneapolis, MN.

Buehner, M., Houtekamer, P. L., Charette, C., Mitchell, H. L., & He, B. (2010). Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: Description and single-observation experiments. *Monthly Weather Review*, *138*(5), 1550–1566. https://doi.org/10.1175/2009MWR3157.1

Carley, J. (2012). *Hybrid ensemble-3DVar radar data assimilation for the short-term prediction of convective storms* (Doctoral dissertation). West Lafayette, IN: Purdue University.

Clark, A. J., Gallus, W. A., & Weisman, M. L. (2010). Neighborhood-based verification of precipitation forecasts from convection-allowing NCAR WRF model simulations and the operational NAM. *Weather and Forecasting*, *25*(5), 1495–1509. https://doi.org/10.1175/2010WAF2222404.1

Clayton, A. M., Lorenc, A. C., & Barker, D. M. (2013). Operational implementation of a hybrid ensemble/4D-Var Global Data Assimilation System at the Met Office. *Quarterly Journal of the Royal Meteorological Society*, *139*(675), 1445–1461. https://doi.org/10.1002/qj.2054

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

Gao, J., & Stensrud, D. J. (2012). Assimilation of reflectivity data in a convective-scale, cycled 3DVAR framework with hydrometeor classification. *Journal of the Atmospheric Sciences*, *69*(3), 1054–1065. https://doi.org/10.1175/JAS-D-11-0162.1

Gao, J., Xue, M., Brewster, K., & Droegemeier, K. K. (2004). A three-dimensional variational data analysis method with recursive filter for Doppler radars. *Journal of Atmospheric and Oceanic Technology*, *21*(3), 457–469. https://doi.org/10.1175/1520-0426(2004)021<0457:ATVDAM>2.0.CO;2

Gao, J., Xue, M., & Stensrud, D. J. (2013). The development of a hybrid EnKF-3DVAR algorithm for storm-scale data assimilation. *Advances in Meteorology*, *2013*, 1–12. https://doi.org/10.1155/2013/512656

Hamill, T. M., & Snyder, C. (2000). A hybrid ensemble Kalman filter-3D variational analysis scheme. *Monthly Weather Review*, *128*(8), 2905–2919. https://doi.org/10.1175/1520-0493(2000)128<2905:AHEKFV>2.0.CO;2

Hu, M., Benjamin, S. G., Ladwig, T. T., Dowell, D. C., Weygandt, S. S., Alexander, C. R., & Whitaker, J. S. (2017). GSI three-dimensional ensemble–variational hybrid data assimilation using a global ensemble for the regional rapid refresh model. *Monthly Weather Review*, *145*(10), 4205–4225. https://doi.org/10.1175/MWR-D-16-0418.1

Hu, M., Xue, M., Gao, J., & Brewster, K. (2006). 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. *Monthly Weather Review*, *134*(2), 699–721. https://doi.org/10.1175/MWR3093.1

Johnson, M., Jung, Y., Dawson, D. T., & Xue, M. (2016). Comparison of simulated polarimetric signatures in idealized supercell storms using two-moment bulk microphysics schemes in WRF. *Monthly Weather Review*, *144*(3), 971–996. https://doi.org/10.1175/MWR-D-15-0233.1

Jung, Y., Zhang, G., & Xue, M. (2008). Assimilation of simulated polarimetric radar data for a convective storm using the ensemble Kalman filter. Part I: Observation operators for reflectivity and polarimetric variables. *Monthly Weather Review*, *136*(6), 2228–2245. https://doi.org/10.1175/2007MWR2083.1

Kong, R., Xue, M., & Liu, C. (2018). Development of a hybrid En3DVar data assimilation system and comparisons with 3DVar and EnKF for radar data assimilation with observing system simulation experiments. *Monthly Weather Review*, *146*(1), 175–198. https://doi.org/10.1175/MWR-D-17-0164.1

Lin, Y. L., Farley, R. D., & Orville, H. D. (1983). Bulk parameterization of the snow field in a cloud model. *Journal of Climate and Applied Meteorology*, *22*(6), 1065–1092. https://doi.org/10.1175/1520-0450(1983)022<1065:BPOTSF>2.0.CO;2

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

Liu, C., Xue, M., & Kong, R. (2020). Direct variational assimilation of radar reflectivity and radial velocity data: Issues with nonlinear reflectivity operator and solutions. *Monthly Weather Review*, *148*(4), 1483–1502. https://doi.org/10.1175/MWR-D-19-0149.1

Mellor, G. L., & Yamada, T. (1982). Development of a turbulence closure-model for geophysical fluid problems. *Reviews of Geophysics*, *20*(4), 851–875. https://doi.org/10.1029/RG020i004p00851

Nakanishi, M., & Niino, H. (2004). An improved Mellor-Yamada level-3 model with condensation physics: Its design and verification. *Boundary-Layer Meteorology*, *112*(1), 1–31. https://doi.org/10.1023/B:BOUN.0000020164.04146.98

Nakanishi, M., & Niino, H. (2006). An improved Mellor-Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog. *Boundary-Layer Meteorology*, *119*(2), 397–407. https://doi.org/10.1007/s10546-005-9030-8

Pan, Y., Zhu, K., Xue, M., Wang, X., Hu, M., Benjamin, S. G., et al. (2014). A GSI-based coupled EnSRF-En3DVar hybrid data assimilation system for the operational rapid refresh model: Tests at a reduced resolution. *Monthly Weather Review*, *142*(10), 3756–3780. https://doi.org/10.1175/MWR-D-13-00242.1

Potvin, C. K., & Wicker, L. J. (2013). Assessing ensemble forecasts of low-level supercell rotation within an OSSE framework. *Weather and Forecasting*, *28*(4), 940–960. https://doi.org/10.1175/WAF-D-12-00122.1

Putman, W. M., & Lin, S.-J. (2007). Finite-volume transport on various cubed-sphere grids. *Journal of Computational Physics*, *227*(1), 55–78. https://doi.org/10.1016/j.jcp.2007.07.022

Roebber, P. J. (2009). Visualizing multiple measures of forecast quality. *Weather and Forecasting*, *24*(2), 601–608. https://doi.org/10.1175/2008WAF2222159.1

Schwartz, C. S., Kain, J. S., Weiss, S. J., Xue, M., Bright, D. R., Kong, F., et al. (2010). Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Weather and Forecasting*, *25*(1), 263–280. https://doi.org/10.1175/2009WAF2222267.1

Skinner, P. S., Wheatley, D. M., Knopfmeier, K. H., Reinhart, A. E., Choate, J. J., Jones, T. A., et al. (2018). Objective-based verification of a prototype warn-on-forecast system. *Weather and Forecasting*, *33*, 1225–1250. https://doi.org/10.1175/WAF-D-18-0020.1

Snook, N., Xue, M., & Jung, Y. (2015). Multiscale EnKF assimilation of radar and conventional observations and ensemble forecasting for a tornadic mesoscale convective system. *Monthly Weather Review*, *143*(4), 1035–1057. https://doi.org/10.1175/MWR-D-13-00262.1

Snyder, C., & Zhang, F. (2003). Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Monthly Weather Review*, *131*, 1663–1677. https://doi.org/10.1175/2555.1

Sun, J., & Crook, N. A. (1997). Dynamical and microphysical retrieval from Doppler radar observations using a cloud model and its adjoint. Part I: Model development and simulated data experiments. *Journal of the Atmospheric Sciences*, *54*(12), 1642–1661. https://doi.org/10.1175/1520-0469(1997)054<1642:DAMRFD>2.0.CO;2

Tong, M., & Xue, M. (2005). Ensemble Kalman filter assimilation of Doppler radar data with a compressible nonhydrostatic model: OSS experiments. *Monthly Weather Review*, *133*(7), 1789–1807. https://doi.org/10.1175/MWR2898.1

Wang, H., Sun, J., Fan, S., & Huang, X.-Y. (2013). Indirect assimilation of radar reflectivity with WRF 3D-Var and its impact on prediction of four summertime convective events. *Journal of Applied Meteorology and Climatology*, *52*, 889–902. https://doi.org/10.1175/JAMC-D-12-0120.1

Weygandt, S. S., & Benjamin, S. (2007). *Radar reflectivity-based initialization of precipitation systems using a diabatic digital filter within the Rapid Update Cycle*. Paper presented at the 22nd Conf. on Weather Analysisand Forecasting/18th Conf. on Numerical Weather Prediction, Park City, UT.

Zhang, J., Howard, K., & Gourley, J. J. (2005). Constructing three-dimensional multiple-radar reflectivity mosaics: Examples of convective storms and stratiform rain echoes. *Journal of Atmospheric and Oceanic Technology*, *22*(1), 30–42. https://doi.org/10.1175/JTECH-1689.1

Zhu, K., Pan, Y., Xue, M., Wang, X., Whitaker, J. S., Benjamin, S. G., et al. (2013). A regional GSI-based ensemble Kalman filter data assimilation system for the rapid refresh configuration: Testing at reduced resolution. *Monthly Weather Review*, *141*(11), 4118–4139. https://doi.org/10.1175/MWR-D-13-00039.1