

Short-Term Probabilistic Forecasts of the 31 May 2013 Oklahoma Tornado and Flash Flood Event Using a Continuous-Update-Cycle Storm-Scale Ensemble System

NUSRAT YUSSOUF

Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, and NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma

JOHN S. KAIN AND ADAM J. CLARK

NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma

(Manuscript received 16 November 2015, in final form 6 April 2016)

ABSTRACT

A continuous-update-cycle storm-scale ensemble data assimilation (DA) and prediction system using the ARW model and DART software is used to generate retrospective 0–6-h ensemble forecasts of the 31 May 2013 tornado and flash flood event over central Oklahoma, with a focus on the prediction of heavy rainfall. Results indicate that the model-predicted probabilities of strong low-level mesocyclones correspond well with the locations of observed mesocyclones and with the observed damage track. The ensemble-mean quantitative precipitation forecast (QPF) from the radar DA experiments match NCEP's stage IV analyses reasonably well in terms of location and amount of rainfall, particularly during the 0–3-h forecast period. In contrast, significant displacement errors and lower rainfall totals are evident in a control experiment that withholds radar data during the DA. The ensemble-derived probabilistic QPF (PQPF) from the radar DA experiment is more skillful than the PQPF from the no_radar experiment, based on visual inspection and probabilistic verification metrics. A novel object-based storm-tracking algorithm provides additional insight, suggesting that explicit assimilation and 1–2-h prediction of the dominant supercell is remarkably skillful in the radar experiment. The skill in both experiments is substantially higher during the 0–3-h forecast period than in the 3–6-h period. Furthermore, the difference in skill between the two forecasts decreases sharply during the latter period, indicating that the impact of radar DA is greatest during early forecast hours. Overall, the results demonstrate the potential for a frequently updated, high-resolution ensemble system to extend probabilistic low-level mesocyclone and flash flood forecast lead times and improve accuracy of convective precipitation nowcasting.

1. Introduction

The vision for the National Oceanic and Atmospheric Administration (NOAA) Warn-on-Forecast (WoF; [Stensrud et al. 2009](#)) research and development project is to provide probabilistic guidance derived from storm-resolving numerical weather prediction (NWP) models to NOAA's National Weather Service (NWS) forecasters to aid in their issuance of warnings for severe and hazardous convective weather ([Stensrud et al. 2013](#)).

Severe weather is particularly dangerous when there are multiple weather threats (i.e., tornadoes, hailstorms, damaging windstorms, and flash floods), with perhaps the most deadly combination involving both tornadoes and flash floods since the lifesaving actions for these two hazards are contradictory ([Nielsen et al. 2015](#)). To provide probabilistic guidance regarding the evolution of these types of complex multiple severe weather threats, it is essential to develop a convective-scale, ensemble-based data assimilation and probabilistic forecast system that continuously assimilates Doppler radar and other available observations of ongoing convection into the NWP models ([Stensrud et al. 2009, 2013](#)). The goal of this study is to demonstrate the capability of a WoF-type system [i.e., a storm-scale ensemble Kalman filter

Corresponding author address: Dr. Nusrat Yussouf, NOAA/National Severe Storms Laboratory/National Weather Center, 120 David L. Boren Blvd., Norman, OK 73072.
E-mail: nusrat.yussouf@noaa.gov

(EnKF) based frequently updated data assimilation and forecast system] to handle multiple-weather-threat scenarios over the same area (i.e., both tornado and flash flood hazards).

Several recent studies have examined the ability of a WoF-type system to assimilate observed tornadic supercell storms and provide 0–1-h probabilistic numerical forecasts of strong low-level vertical vorticity and/or updraft helicity (both of which have been used as a proxy for tornadoes in convection-allowing models), and the results obtained are very encouraging (Dawson et al. 2012; Yussouf et al. 2013a,b, 2015; Potvin and Wicker 2013; Wheatley et al. 2015). For example, Yussouf et al. (2013b, 2015) illustrated that a WoF-type system was able to assimilate strong supercells and predict intense low-level mesocyclone tracks that align well with the locations of radar-derived rotation tracks associated with observed tornadic storms.

The success of these studies in demonstrating the potential of the analysis and forecast systems to provide useful 0–1-h guidance for predicting tornadoes suggests that this type of system also holds promise in providing guidance for the prediction of other high-impact severe convective weather events like extreme rainfall, high winds, and hailstorms (Stensrud et al. 2013). Application for extreme rainfall associated with severe storms is particularly intriguing because intense rainfall over a short amount of time can lead to flash flooding (Ashley and Ashley 2008), which, on average, causes more fatalities per year than either tornadoes or hurricanes (Barthold et al. 2015). The average storm-based flash flood warning lead time over the past ~10 years is approximately 1 h (<https://verification.nws.noaa.gov>). There have been significant advances in recent years in the use of hydrologic models to predict flash floods (Chen et al. 2013; Barthold et al. 2015 and references therein). However, these models are generally most skillful when they are initialized with observed precipitation totals [i.e., quantitative precipitation estimates (QPEs)], rather than quantitative precipitation forecasts (QPFs) from models because the forecasts tend to have errors in placement, amplitude, and timing that can be very detrimental to accurate flash flood prediction with hydrologic models. In spite of these errors, QPF fields have one large advantage over QPEs: they can be generated before a heavy rain event occurs. The challenge is to develop QPF and probabilistic QPF (PQPF) systems that are sufficiently accurate to provide useful input to hydrologic models and timely enough to allow these models to be initialized well before QPEs are available. When this challenge is met, hydrologic models can provide flash flood guidance to forecasters earlier than with the QPE-based paradigm, providing

forecasters with the tools to issue flash flood warnings with longer lead times and no loss in accuracy compared to current warnings.

QPF systems have improved over the last decade, due largely to the increasing availability of QPFs derived from convection-allowing models (CAMs) that are initialized by downscaling coarser-resolution operational NWP models (e.g., Kain et al. 2006, 2010; Weisman et al. 2008; Clark et al. 2009, 2010; Schwartz et al. 2009; Chang et al. 2012; Duc et al. 2013; Tang et al. 2013; Gagne et al. 2014). Further advances in QPFs and PQPFs have come from downscaling experimental mesoscale EnKF data assimilation systems (e.g., Jones and Stensrud 2012; Schumacher and Clark 2014; Romine et al. 2013; Schwartz and Liu 2014; Schwartz et al. 2014, 2015). However, when convection-allowing NWP models are initialized by downscaling from coarser-resolution analyses, it takes as long as 3–6 h for deep-convective processes to spin up and, when storms do develop, they often emerge with the previously mentioned errors in timing, amplitude, and placement, leaving little value in the short-term model guidance for driving hydrologic models. Thus, the downscaling approach is likely to have limited value for flash-flood warnings.

A recent study (Sun et al. 2014) reviewed the current progress and challenges of nowcasting (0–6 h) convective precipitation and illustrated that the assimilation of radar observations into high-resolution NWP models using a rapid-update-cycle strategy is essential for accurate numerical nowcasting of rainfall. The continuous assimilation of radar data and other routinely available conventional observations into the storm-scale model is needed to overcome the inherent model spinup issues during the first few hours into the forecasts (Sun et al. 2014 and references therein). This fundamental approach is also being used as part of the WoF effort. Therefore, in addition to focusing on improving the accuracy and forecast lead times for tornado-producing thunderstorms, it is important to test the applicability and robustness of the same WoF type system for other high-impact weather events like heavy rainfall and flash-flood-producing thunderstorms.

One hazardous convective weather outbreak in recent years that not only produced violent tornadoes but also caused heavy rainfall and flash flooding is the central Oklahoma event of 31 May 2013. Almost two-thirds of the total 22 fatalities on that day were due to flash flooding. While the violent tornado in El Reno, Oklahoma, garnered most of the media attention that evening, the back-building slow-moving convective system associated with the tornadic storm produced heavy rainfall over the area in the hours following the tornadoes. This type of event is particularly complicated from

the safety perspective since the recommended shelters for tornados and flash floods are often contradictory, for example, below ground if possible for violent tornados versus high ground for flash floods (Nielsen et al. 2015).

Yussouf et al. (2015) demonstrates the capability of the emerging WoF-type frequently updated system in predicting the \sim 1-h probabilistic forecasts of reflectivity and the mid- and low-level rotational characteristics of a severe tornado outbreak event. In this study, we investigate the utility of the same WoF-type system in predicting both the low-level mesocyclone associated with the tornadic storm and the heavy rainfall that follows from the 31 May 2013 event. Recognizing that there is rapid error growth at storm scale, we nonetheless extend the forecast length out to 6 h to evaluate whether the system can be used for nowcasting out this far. The objective is to examine the quality of the analyses and very short-range ensemble probabilistic forecasts of storm rotation and heavy rainfall and also to speculate on the potential utility of this analysis and forecast system in providing quantitative precipitation nowcasts to drive hydrologic models.

A brief overview of the 31 May 2013 tornado and flash flood event in Oklahoma is provided in section 2, followed by the experiment design of the WoF-type data assimilation and forecast system in section 3. The filter performance is discussed in section 4. Section 5 assesses the qualitative and quantitative results of the analyses and forecasts of severe weather associated with low-level rotation and precipitation forecasts. A final discussion is found in section 6.

2. Overview of the 31 May–1 June 2013 tornado and flash flood event over central Oklahoma

The 31 May–1 June 2013 tornado and flash flood over central Oklahoma was a unique severe weather event that not only produced one of the widest tornados on record but also the deadliest flash flood in the NWS Norman forecast area since 1934 (NWS WFO Norman 2015). An overview of the environmental conditions and severe weather event is provided by NOAA (2014), Wurman et al. (2014), Snyder and Bluestein (2014), and Bluestein et al. (2015). In the late afternoon hours (around 2130 UTC), a cluster of storms formed along a cold front/dryline in west-central Oklahoma and an associated severe weather episode started at 2235 UTC with an [enhanced Fujita (EF) scale] EF0 tornado in Kingfisher County (Figs. 1a–c; see Fig. 1g for county identification). Multiple tornados occurred on this day but the most intense and longest lived of all was the “El Reno tornado” that began southwest of El Reno, at around 2303 UTC and ended at around 2344 UTC after

carving a damage path approximately 25.7 km long and 4.2 km wide (Fig. 1g). This tornado was rated an EF3, and it injured 26 and killed 8 people, including several veteran storm chasers (Wurman et al. 2014). There were a total of 19 tornados over Oklahoma during the afternoon and evening hours on that day, 12 of which occurred in the NWS Norman Forecast area (NWS WFO Norman 2015).

Although the storm cell that spawned the El Reno tornado moved slowly eastward into eastern Oklahoma after 0000 UTC on 1 June, a sequence of new convective cells formed near the original initiation point and tracked eastward over the same area. Thus, this back-building storm system brought heavy rainfall to the Oklahoma City, Oklahoma (OKC), metropolitan area (Figs. 1d–f and 1h), resulting in significant flash flooding during the evening of 31 May and the early morning of 1 June. In all, 13 people were killed by flash floods, including 12 people in OKC, making this event the deadliest flash flood event in OKC history. There were about 23 high-water rescues and at least 100 000 homes and businesses lost power during the long-lived storm system, with the first flooding reported at 0100 UTC (NWS WFO Norman 2015). Several operational Weather Surveillance Radar-1988 Doppler (WSR-88D) radars recorded the life cycle of this severe weather event. The observations from these radars are assimilated continuously at 5-min intervals into the storm-scale ensemble to assess the capability of this prediction system to forecast the low-level rotation and intense rainfall associated with the storms.

3. Experimental design

a. Multiscale WRF ensemble system

The configuration of the ensemble data assimilation and forecast system is very similar to that used by Yussouf et al. (2015) and is based on the Advanced Research version of the Weather Research and Forecasting (ARW version 3.4.1; Skamarock et al. 2008) Model. A storm-scale domain with 3-km horizontal grid spacing is nested within a 15-km coarse-resolution grid (Fig. 2a) and covers Oklahoma and parts of surrounding (Fig. 2b) states. There are 51 vertical grid levels on both domains that extend from the surface to 10 hPa at the top. The ensemble is initialized at 0000 UTC 31 May 2013 with a 36-member multiphysics configuration using the analyses from National Centers for Environmental Prediction’s (NCEP) Global Ensemble Forecast System (GEFS; Toth et al. 2004; Wei et al. 2008). The different combinations of physics schemes among the ensemble members are the same as in Table 2 of Yussouf et al. (2015).

The ensemble adjustment Kalman filter (EAKF; Anderson 2001) from the Kodiak release branch (revision 5038) of the Data Assimilation Research Testbed

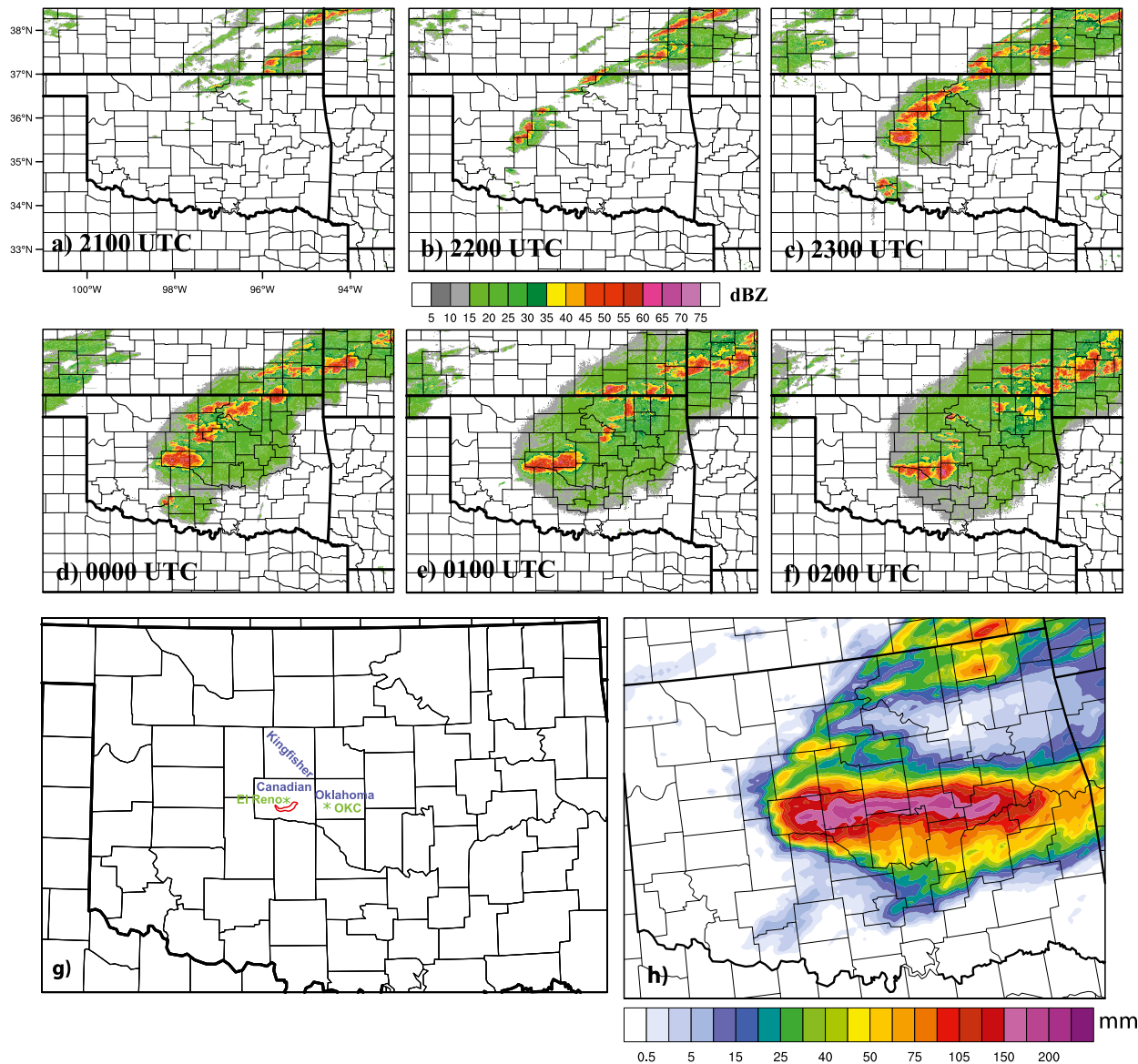


FIG. 1. Composite reflectivity (dBZ, from the NSSL NMQ system) at (a) 2100, (b) 2200, and (c) 2300 UTC 31 May 2013, and (d) 0000, (e) 0100, and (f) 0200 UTC 1 Jun 2013 over the region of interest. (g) El Reno tornado path and location of counties (in blue) and cities (in green) of interest, and (h) observed precipitation (in mm) over central OK valid from 0700 LT 31 May to 0700 LT 1 Jun from NCEP stage IV 24-h accumulated precipitation analysis.

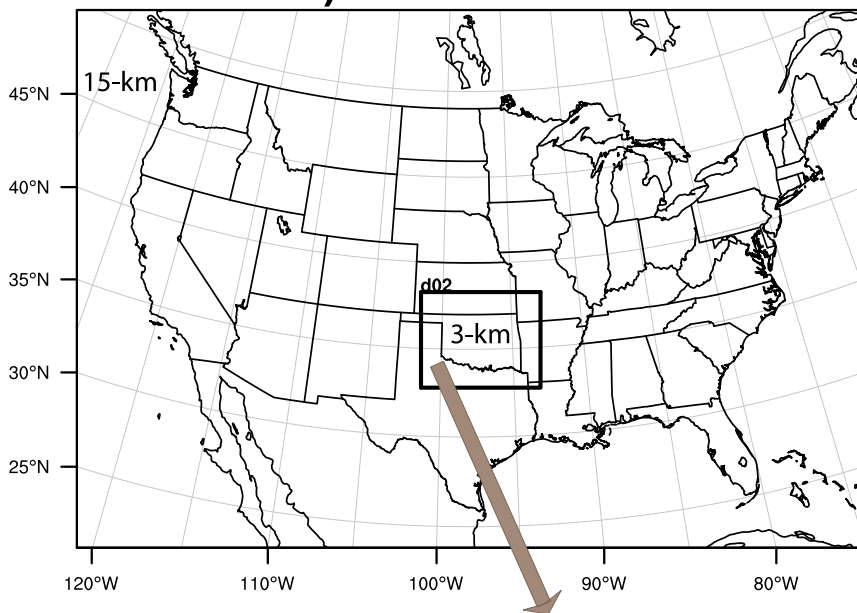
software system (DART; Anderson and Collins 2007; Anderson et al. 2009) is used as the data assimilation tool. DART is a community toolkit maintained by the Data Assimilation Research Section (DAReS) at the National Center for Atmospheric Research (NCAR; available online at <http://www.image.ucar.edu/DAReS/DART/>).

b. Mesoscale 1-h DART ensemble data assimilation and forecast system

The data sources for mesoscale observation assimilation include METARs, mesonet observations, marine

reports, rawinsondes, and aircraft and satellite-derived winds, all of which are available from the NOAA Meteorological Assimilation Data Ingest System (MADIS). The altimeter setting, temperature, dewpoint, and horizontal wind components are assimilated into the ensembles every 1 h from 0100 UTC 31 May to 1100 UTC 1 June 2013 (Fig. 3a) to create the mesoscale background fields. Both 15-km mesoscale and 3-km storm-scale grids are run simultaneously in a one-way nested setup for the storm-scale grid. The mesoscale ensemble provides the boundary conditions for the nested storm-scale ensemble. Additional

a) Multiscale Domain



b) Storm-scale Domain

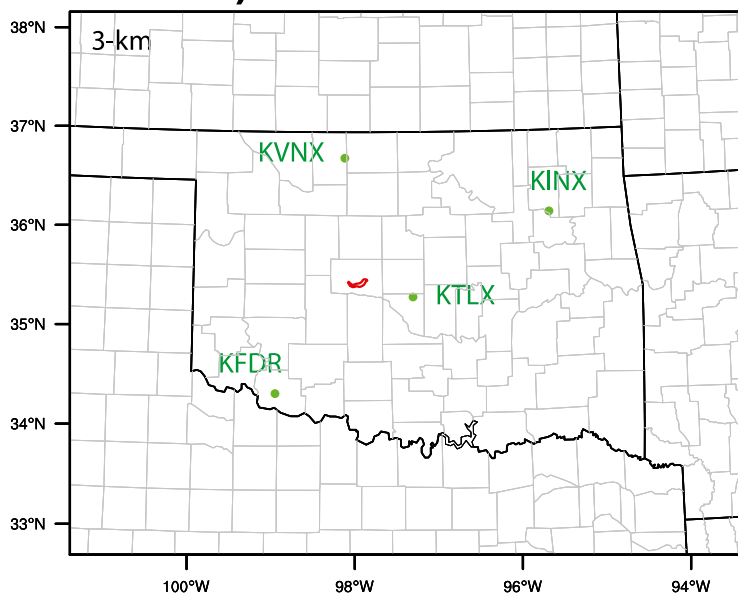


FIG. 2. (a) The multiscale domain with a 15-km horizontal grid-spacing mesoscale domain covering the continental United States, and the nested 3-km storm-scale domain centered over OK. (b) The storm-scale domain enlarged, with WSR-88D locations (green dots), and the NWS damage swath (in red) from the El Reno tornado (EF3).

details of the hourly updated mesoscale data assimilation system can be found in [Yussouf et al. \(2015\)](#).

c. DART storm-scale continuous 5-min data assimilation and forecast system

Radar observations are assimilated into the 3-km storm-scale ensemble starting at 2100 UTC 31 May

2013, the time when discontinuous lines of supercells start to initiate along the cold front/dryline over west-central Oklahoma. The 2100 UTC 31 May storm-scale model output from the hourly updated system is used as the background (prior) to assimilate radar, mesonet, METAR, radiosonde, and aircraft observations only on the storm-scale domain every 5 min for a 7-h period out

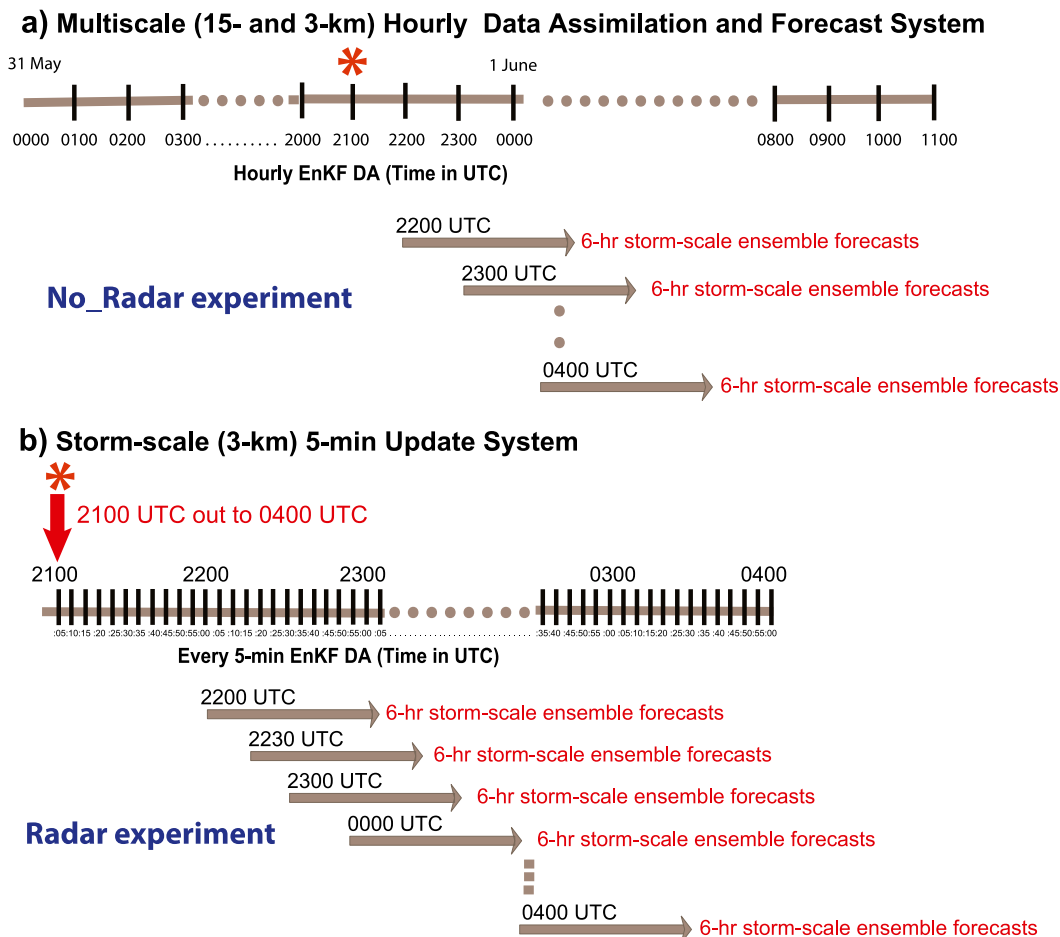


FIG. 3. (a) The timeline of the hourly multiscale data assimilation experiments and storm-scale forecasts every hour starting from 2200 UTC (no_radar) and (b) the timeline for the every 5-min storm-scale radar data assimilation experiment and ensemble forecasts starting from 2200 UTC (radar experiment).

to 0400 UTC the next day (Fig. 3b). The hourly updated mesoscale domain is used to provide the boundary conditions for the storm-scale 5-min update system. Reflectivity and radial velocity observations from four operational WSR-88Ds located in Oklahoma at Vance Air Force Base (KVNK), Twin Lakes (KTLX), Tulsa (KINX), and Frederick (KFDR) are assimilated (Fig. 2b), in addition to the other conventional observations. These data consist of the level II radar observations, obtained from the National Centers for Environmental Information (NCEI) comprising 14 scan angles [volume coverage pattern (VCP) 12 mode] and completing each full volume scan in approximately 4.5 min. The details of observation quality control and preprocessing can be found in Yussouf et al. (2015). The observation-error standard deviations are assumed to be 5 dBZ and 2 m s^{-1} for reflectivity and Doppler velocity, respectively.

The model state variables included in the state vector for the DART EAKF scheme to be updated include the

perturbation surface pressure of dry air, perturbation geopotential, perturbation potential temperature, three wind components, potential temperature tendency due to microphysics, water vapor, and all available hydrometeor fields from the semi-double-moment Thompson microphysics scheme. The model-diagnosed total surface pressure, reflectivity, fall-weighted velocity, 10-m u and v wind components, 2-m temperature, and water vapor also are included in the state vector. The radial velocity forward operator interpolates the fall-weighted velocity directly from the state vector to account for the terminal fall speed of hydrometeors [cf. Eq. (2) of Aksoy et al. (2009)]. The covariance localization for the radar observations is set to have a half-radius in the horizontal (vertical) of 9 km (3 km), as well as 180 km (3 km) for conventional and 60 km (3 km) for mesonet observations. During each assimilation cycle, temporally and spatially varying adaptive inflation (Anderson 2009) is applied to maintain the ensemble spread. Additional

spread is provided by applying the additive noise technique (Dowell and Wicker 2009) every 15 min of the assimilation cycle to each ensemble member at grid points where the observed reflectivity value is greater than 25 dBZ.

Two sets of 6-h storm-scale ensemble forecasts are initialized every hour starting from 2200 UTC (which is after 60 min of data assimilation) with WRF history output files every 5 min. One set of ensemble forecasts is generated from the hourly updated 3-km ensemble (Fig. 3a; referred to as the no_radar experiment hereafter) and the other set of ensemble forecasts are generated from the every 5-min update system (Fig. 3b; referred to as the radar experiment hereafter). The goal is to examine the capability of this prototype WoF system in forecasting low- to midlevel rotation and intense convective rainfalls.

d. Feature-based storm tracking

A feature-based storm-tracking algorithm is utilized to identify, track, and diagnose various characteristics of matching precipitation features from the forecasts and observations. This algorithm was utilized in VandenBerg et al. (2014) for tracking reflectivity features, and is implemented for the current application as follows.

The first step in the algorithm is to identify precipitation features, defined as contiguous regions of precipitation exceeding a specified threshold, at each 1-h output time (hourly accumulations). In this study, a threshold value of 6.35 mm (0.25 in.) is chosen because, when applied to the stage IV observations of precipitation (Baldwin and Mitchell 1997; Lin and Mitchell 2005) for this event, this threshold yields a coherent feature track associated with the areas of localized extreme rainfall areas that moved across central Oklahoma during the period of the experiment. However, a range of threshold values may be necessary to apply this algorithm to a broader range of cases.

To determine the track of a feature once it is identified from the hourly forecast history files, the algorithm associates the feature in subsequent future forecasts. If a single feature at one forecast output time overlaps with, or is adjacent to, a single feature at the next forecast output time, it is considered part of the same track. This is true when there is no merger or split. In addition, once a track is identified, no part of the track is considered part of any other track. The algorithm keeps searching through each subsequent forecast time once a track is identified. The search continues until the end of the track is obtained, or until the last forecast output time is reached. If more than one future feature overlaps with, or is adjacent to, a current feature, the algorithm assumes that a feature split has occurred, and the largest

future feature is considered a continuation of the present feature track, while the smaller features are considered the starting points for new tracks. Conversely, a storm merger is assumed to occur when more than two features at the present time overlap with, or are adjacent to, one feature at the subsequent future forecast time. In this scenario, the track of the largest feature continues to the future time, and the tracks of any smaller features at the present time are terminated.

To match the predicted (simulated) feature tracks to the observed track, the proportion of the observed feature at its start time overlapped by a selected predicted feature at the same time is computed. If the proportion is $\geq 10\%$, then the selected predicted feature is considered a match. The 10% criteria may seem somewhat lenient, but was chosen because larger values result in very few matched features in the no_radar experiments, which had trouble generating precipitation at the exact location of the observations during the first 0–1 h of the forecasts since radar data were not assimilated. Once matching features were identified, a series of feature attributes were computed for each matched feature at each forecast hour.

Although the observed feature exists for the entire experimental period, not all of the matching predicted features persist that long. Thus, attributes of the forecast features are computed until the feature terminates. The attributes calculated are 1) the overlap, or the proportion of the observed feature covered by the forecast feature; 2) the nonoverlap, or the proportion of the observed feature not covered by any forecast feature; 3) the feature displacement, computed as the distance between centroids of predicted and observed features, where centroids are computed using a simple average of the x and y grid coordinates comprising each feature; 4) the maximum feature intensity (i.e., the highest value of hourly precipitation within the feature); and 5) the feature size, or the number of grid points comprising each feature.

e. Objective verification of probabilistic precipitation forecasts

PQPFs from the ensemble prediction systems are verified using fractions skill scores (FSSs; Roberts 2005; Roberts and Lean 2008) and the area under the relative operating characteristic (ROC) curve (AUC; Mason 1982), applied over a range of rainfall thresholds, forecast hours, and forecast lengths. Both FSS and AUC have a range from 0 to 1, with a perfect skill score of 1 and a score of 0 meaning zero skill. While the FSS is a measure of the spatial skill of precipitation forecasts, the AUC measures the ability of the forecasts to distinguish between events and nonevents (or resolution). The FSS

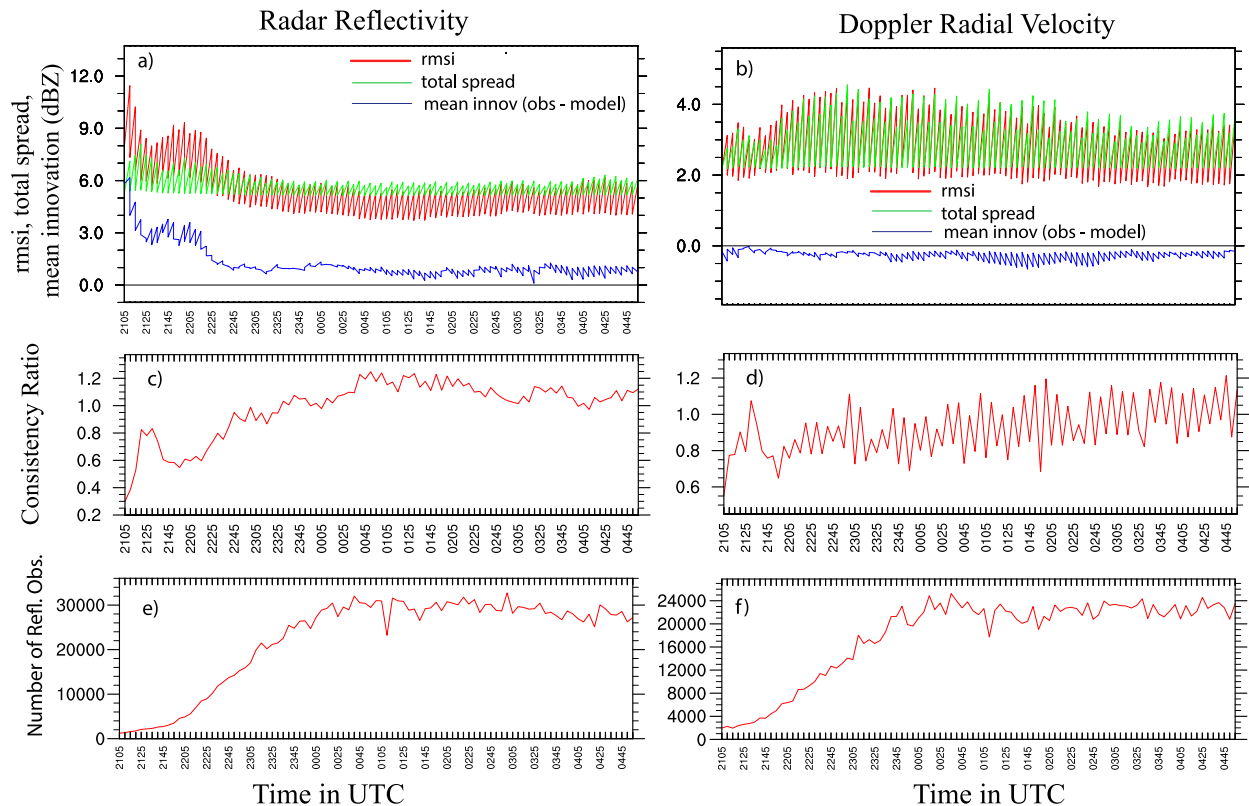


FIG. 4. Observation-space diagnostic statistics of (a),(b) rmsi, total ensemble spread, and mean innovations; (c),(d) consistency ratio; and (e),(f) number of observations assimilated for the assimilated reflectivity (dBZ) and Doppler velocity (m s^{-1}) observations, respectively, from the four radars during the 7-h every 5-min radar storm-scale data assimilation period. The reflectivity statistics are computed only where the assimilated observed reflectivity is greater than 10 dBZ. The sawtooth patterns are due to the plotted forecast and analysis statistics.

and AUC are calculated using a neighborhood approach (Duc et al. 2013; Schwartz et al. 2010, 2014, 2015; Snook et al. 2015) with a 12-km neighborhood radius. The neighborhood approach, FSS, and AUC calculations used in this study follow the method in Schwartz et al. (2010). The neighborhood method uses Eqs. (1), (3), and (4), and the FSS uses Eqs. (6)–(8) in Schwartz et al. (2010). The FSS and AUC are verified against NCEP's hourly stage IV multisensor rainfall estimates. To get the observations and the ensemble forecasts on a common grid for verification, the 3-km grid-spacing ensemble rainfall forecasts are remapped onto the ~ 4.7 -km stage IV grids using a neighbor-budget interpolation (e.g., Accadia et al. 2003). The regridded ensemble data are used for verification.

4. Filter performance

To evaluate the overall filter performance during the 7-h-long radar data assimilation period from the 3-km storm-scale radar experiment, a set of observation-space statistics are generated, as in Yussouf et al. (2015).

Specifically, mean innovation (observation – model), root-mean-square innovation (rmsi), total ensemble spread (standard deviation), and consistency ratio (from prior/background) are calculated for the assimilated reflectivity and radial velocity observations (Fig. 4) from the 5-min bins (Dowell et al. 2004; Dowell and Wicker 2009; Yussouf et al. 2013b, 2015).

The improvement in mean innovation for reflectivity is largest during the initial spinup of the storm in the ensemble. The mean innovation starts with a high value of 6 dBZ and decreases to 1 dBZ during the early ~ 45 min and remains within the range of 0.5–1.0 dBZ for the remaining assimilation period (Fig. 4a). The mean innovation is a measure of the forecasts/analyses bias (0 means no bias) and therefore indicates that the model underpredicts reflectivity during the assimilation period. For radial velocity, the mean innovation (Fig. 4b) is close to 0 with values that vary between 0 and -0.5 m s^{-1} for the entire assimilation period. The rmsi, which is a measure of the overall fit of the observations to the forecasts/analyses, starts with a higher value of 9–11.5 dBZ for reflectivity but the error decreases with

subsequent assimilation cycles during the first 1 h of data assimilation and becomes fairly stable (Fig. 4a). For radial velocity observations (Fig. 4b), the rmsi slightly increases after the initial ~45 min and decreases again during the later part of the assimilation period but, overall, remains stable for the entire assimilation period. For both observation types, the total spread and the rmsi are of comparable magnitude, indicating that the ensemble spread is representative of the forecast error for the 5-min time scale of the storm-scale NWP system.

The consistency ratio for reflectivity is smaller in early assimilation cycles, but increases with time and remains within the range of 1.0–1.2 (Fig. 4c). A consistency ratio of ~1.0 indicates that the prior ensemble variance is a good approximation of the forecast error variance for the assumed observation error. The consistency ratio for radial velocity observations starts with initial values of around 0.6, increasing rather quickly with subsequent assimilation cycles to values within a range of 0.8–1.2 (Fig. 4d). Importantly, the filter shows no sign of divergence during the 7-h-long continuous period of 5-min assimilation, indicating the robustness of the data assimilation system. The smaller number of observations (Figs. 4e,f) during the initial ~1 h (2100–2200 UTC) of the assimilation period is due to the smaller number of storm echoes (Figs. 1a,b) from the rapid initiation of the El Reno storm during that time period. The overall observation-space diagnostics (Fig. 4) suggest that the ensemble data assimilation system is fairly reasonable and stable.

5. Results and discussion

a. Ensemble probabilistic analyses and forecasts of reflectivity

A series of 1-, 2-, 3-, and 6-h ensemble probabilistic forecasts of reflectivity (greater than 40 dBZ at 2 km MSL) from the radar experiments are compared against the observed reflectivity obtained from the National Mosaic and Multi-Sensor QPE (NMQ) 3D radar reflectivity mosaic (Zhang et al. 2011) system (Fig. 5). The 1-km grid-spacing NMQ gridded reflectivity observations are thinned to 3-km grid spacing to match the storm-scale WRF grid. The results from the ensemble analyses reveal that the assimilation system is able to associate 100% probabilities with the dominant observed storms as early as only after 60 min of radar data assimilation at 2200 UTC (Fig. 5a). The analyses at latter times (Figs. 5f,k) also demonstrate the ability of the system to place the main supercells in the model at approximately the correct locations. The 1- and 2-h

ensemble forecasts generate high ensemble probabilities that correspond to the main reflectivity core reasonably well but with a small northeastward displacement error. The 3-h forecast probability values are comparatively lower than those from the earlier forecast hours with implied storm position errors farther north-northeast. This is due to the radar data assimilation placing the storms in the same location initially and then having the individual storms diverge as forecast lead time increases. The probabilities from the 6-h forecast clearly indicate that the forecast storms tend to move faster to the north-northeast than do the observed storms. This is a very common problem with storm-scale forecasts (Snook et al. 2015; Yussouf et al. 2015) and may be due to the generation of a cold pool that is too intense, which is likely associated with uncertainties related to the assimilation of reflectivity (Dowell et al. 2011; Yussouf et al. 2013b, 2015).

b. Ensemble probabilistic forecasts of low- and midlevel rotation of the El Reno tornadic supercell storm

The ability of the 5-min-update radar experiment to reproduce low- and midlevel rotations associated with the El Reno supercell thunderstorm are evaluated using the ensemble probabilistic forecasts initialized from the storm-scale analyses every 30 min during the 1-h period preceding tornadogenesis (Fig. 6). The forecast probability swath of vertical vorticity greater than 0.003 s^{-1} at 1 km AGL is used to represent low-level rotation, while the forecast probability swath of 2–5-km updraft helicity (UH; Kain et al. 2008; Clark et al. 2012, 2013) greater than $100 \text{ m}^2 \text{ s}^{-2}$ is used to represent midlevel rotation. At a given time, the forecast UH and vertical vorticity from each ensemble member at each grid point is compared with the specified threshold value, and the probability values are set to the fraction of ensemble members exceeding the threshold (Stensrud and Gao 2010; Dawson et al. 2012; Stensrud et al. 2013; Yussouf et al. 2013a,b, 2015). Figures 6a and 6d show 105-min forecasts initialized at 2200 UTC, which is 63 min before tornadogenesis. Figures 6c and 6f show 45-min forecasts initialized at 2300 UTC, which is just 3 min before tornadogenesis. Therefore, the forecast rotation swaths cover the entire duration of the El Reno tornado. The NWS-surveyed tornado damage path and the Warning Decision Support System–Integrated Information (WDSS-II; Lakshmanan et al. 2007) generated 0–2- and 2–5-km mesocyclone circulations (Miller et al. 2013) are used to compare with the model-generated vorticity swaths. Note that while these fields derived from observations are useful for validating model forecasts of severe storms, caution must be

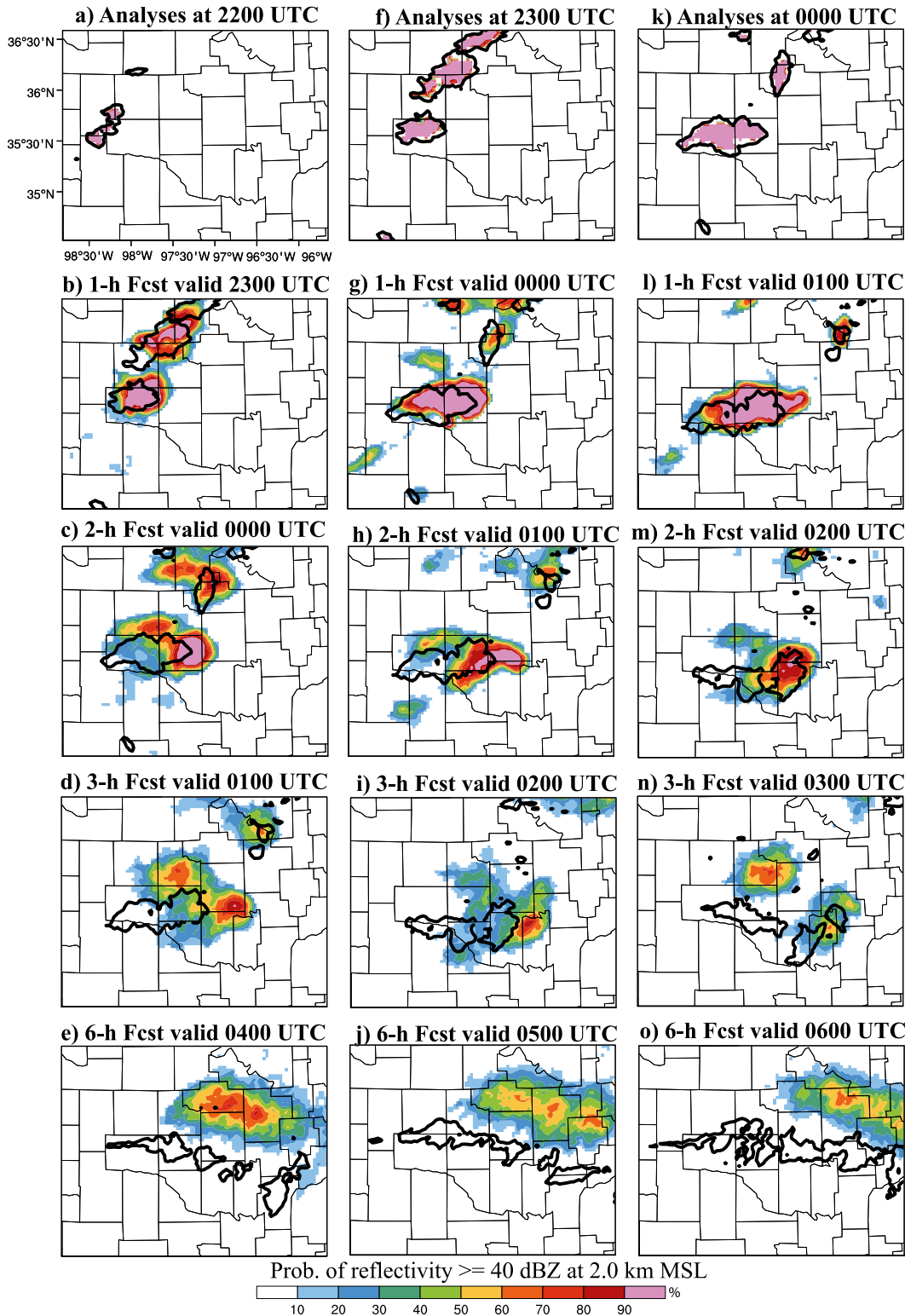


FIG. 5. The ensemble probability of reflectivity greater than 40 dBZ (colors, 10% increment) at 2 km MSL from the radar experiment at the (a),(f),(k) analyses time and then for the every (b),(g),(l) 1-, (c),(h),(m) 2-, (d),(i),(n) 3-, and (e),(j),(o) 6-h forecasts. The thick black contour is the observed 40-dBZ reflectivity contour. The portion of the domain shown here is over central OK.

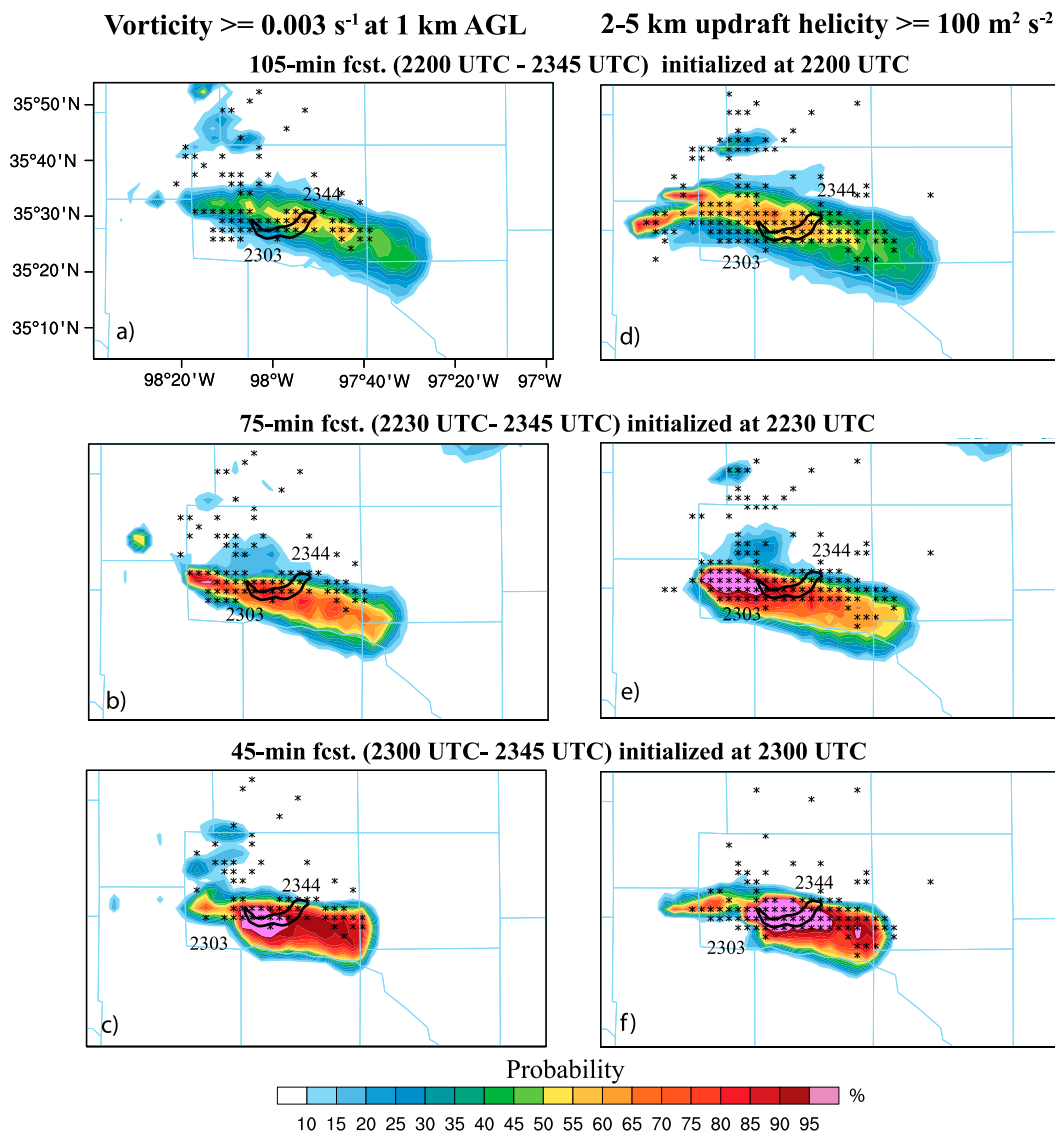


FIG. 6. Raw model gridpoint-based ensemble probability of (a)–(c) vorticity forecasts exceeding a threshold of 0.003 s^{-1} at 1 km AGL, and (d)–(f) 2–5-km UH exceeding a threshold of $100 \text{ m}^2 \text{ s}^{-2}$ from every 30-min radar analyses for the El Reno supercell. Overlaid in each panel is the NWS-observed tornado damage track (black outline) and the WDSS-II-generated radar-derived low-level (0–2 km AGL) in (a)–(c) and midlevel (2–5 km AGL) in (d)–(f) mesocyclone exceeding a threshold of 0.006 s^{-1} (black asterisk) during the indicated forecast periods.

exercised when doing so because the observed and predicted quantities are not the same and may not always correlate strongly.

In general, the forecasts of low-level vorticity and midlevel UH paint a consistent picture of the forecast scenario (Fig. 6). The 105-min forecast probability of vorticity (Fig. 6a) and UH (Fig. 6d) initialized at 2200 UTC indicates mesocyclone probabilities with maximum values of 65%, and the swath overlaps the radar-generated circulation locations. The midlevel rotation forecasts show higher-probability values with a wider swath than is found from the low-level rotation, as one

would expect. The maximum forecast probabilities are increased to 85% for the low-level vorticity (Fig. 6b) and 100% for midlevel UH (Fig. 6e) from the 2230 UTC analyses, which is after 30 min of continuous data assimilation. The swaths also are enhanced along the radar-observed rotation. The forecast probabilities in this area are consistently enhanced and increased with a continuous update cycle. The forecast rotation probabilities from 2300 UTC analyses, the time when the El Reno tornado is about to occur, reach values as high as 100% (Figs. 6c,f). Compared to the forecasts from the earlier lead times, the probabilities along the mesocyclone

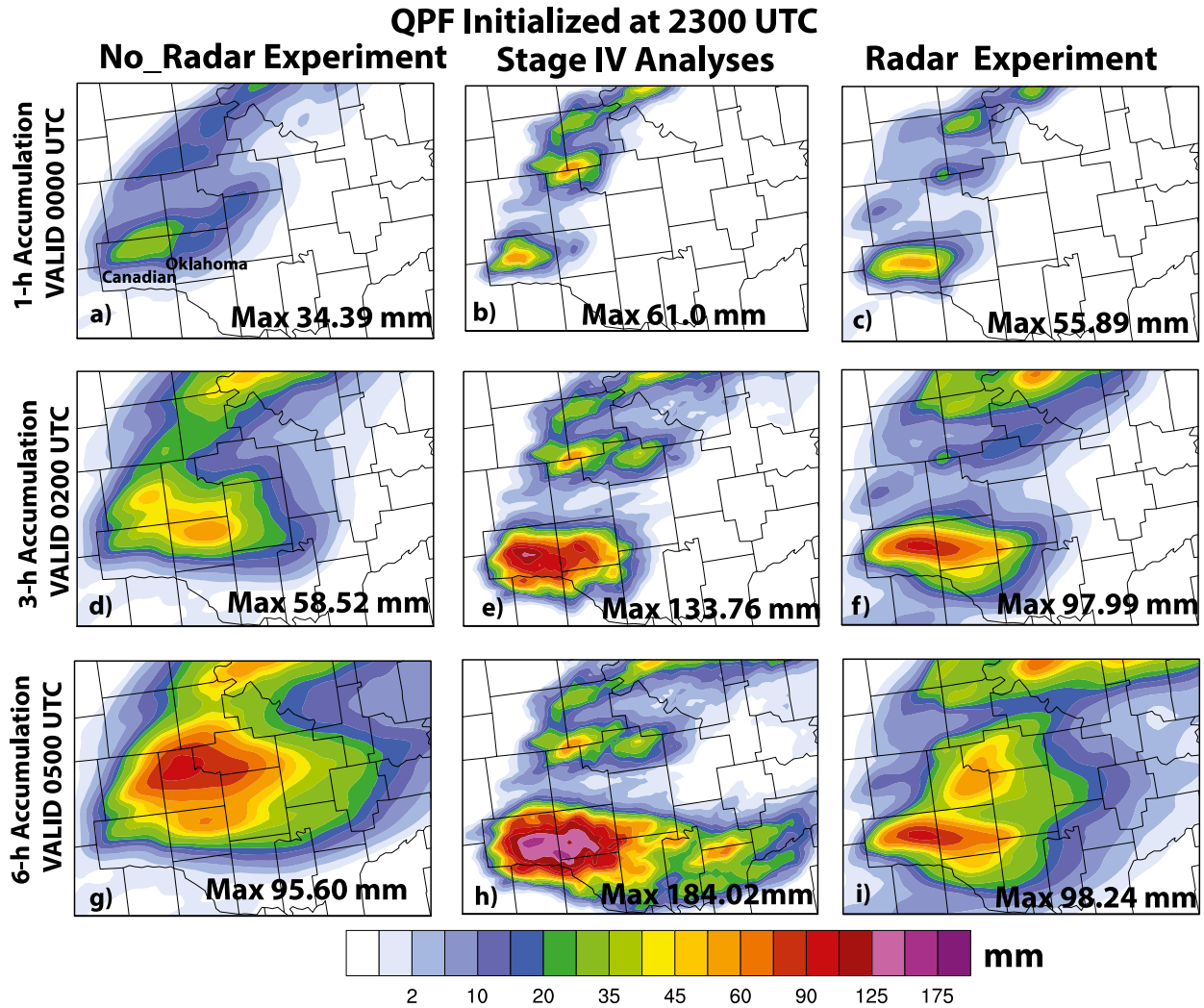


FIG. 7. Ensemble mean 1-, 3-, and 6-h accumulated precipitation forecasts initialized at 2300 UTC from the (a),(d),(g) no_radar and (c),(f),(i) radar experiments, as well as (b),(e),(h) NCEP stage IV accumulated precipitation analyses. The portion of the domain shown here is over central OK. The locations of Oklahoma and Canadian Counties are shown in (a).

track from the 2300 UTC forecast are higher and aligned better with the El Reno radar-derived rotations and NWS tornado damage tracks.

c. Quantitative precipitation forecasts

Numerical prediction of flash floods with long lead time may be significantly enhanced if NWP models can accurately predict the amount, location, and timing of heavy rainfall. Thus to evaluate the ability of the ensemble system to forecast accurate rainfall amounts, 0–1-, 0–3-, and 0–6-h ensemble-mean accumulated precipitation forecasts from the no_radar and radar experiments (initialized from 2300 and 0000 UTC storm-scale analyses) are presented here. As mentioned earlier, the NWS WFO in Norman received the first flooding report for this event at 0100 UTC. For this study, the

focus is on Canadian and Oklahoma Counties (see Fig. 7a), where the loss of life and property was greatest during this event.

During the 2300–0000 UTC period, the stage IV analyses indicated heavy rainfall centered on west-central Canadian County, with a peak magnitude of ~61 mm (Fig. 7b). The 0–1-h ensemble-mean forecast from the radar experiment shows good agreement with these observations, but with a somewhat broader swath of heavy rainfall, a slight eastward displacement toward the eastern part of the county, and a peak value of ~56 mm (Fig. 7c). By comparison, the corresponding forecast from the no_radar experiment produces a much more diffuse rainfall pattern with clear northeastward displacement and a peak value of only ~35 mm (Fig. 7a). These results are consistent with those of Sun et al.

(2014), indicating that the continuous 5-min radar data assimilation helps with the initial model spinup issue.

In terms of the perceived relative skill of forecasts from the radar and no_radar experiments, the 0–3-h rainfall forecasts from the 2300 UTC initialization time yield similar results. Heavy rainfall covers much of Canadian and Oklahoma Counties in the radar experiment, in reasonably good agreement with observations (cf. Figs. 7e,f), while the heavier rainfall in the no_radar experiment is displaced to the north and east (cf. Figs. 7d–f). Both experimental forecasts underestimate the peak value of the observed very heavy rainfall during this period (~134 mm), with maximum values of ~98 and ~59 mm coming from the radar and no_radar experiments, respectively.

Visual inspection of the 0–6- (Fig. 7g) and 0–3-h (Fig. 7d) accumulations from the no_radar experiment indicates that heavy rainfall was generated in this experiment during the 3–6-h period, suggesting that, after the first few hours into the forecast, convection in the no_radar ensemble reaches an intensity that is comparable to storms that were observed on this day. However, the simulated activity remains displaced to the northeast relative to observations (cf. Figs. 7g,h). A similar comparison of the 0–3- and 0–6-h rainfall patterns from the radar experiment (Figs. 7i,f, respectively) reveals that the 3–6-h forecast yielded the heaviest rain to the northeast of Canadian and Oklahoma Counties in this experiment as well. Yet, observations show that very heavy rainfall continued in these counties during this time period (cf. Figs. 7h,e). Thus, it appears that the positive impact of radar data assimilation decreases rapidly after the 0–3-h forecast period, as suggested by Kain et al. (2010) and others.

One of the challenges of evaluating ensemble prediction systems for this type of event is to find concise, but revealing ways of diagnosing the contributions from different ensemble members. The results shown above indicate that the ensemble-mean precipitation field is potentially useful in this case, but this field can diminish the sharpness of the coverage, amplitude, and configuration of fields like accumulated precipitation. One way of restoring the amplitude of the precipitation field is to use the probability-matched mean (Ebert 2001), but this representation does little to reveal the error characteristics of individual ensemble members and how these characteristics translate into errors in the field of mean values. Here, we use the “feature-based storm tracking” method (described in section 3d) introduced by VandenBerg et al. (2014), which is conceptually similar to Carley et al. (2011), to track and characterize dominant precipitation features in the ensemble members, revealing important feature-specific

error characteristics related to amplitude, displacement, movement, and coverage.

The feature-specific results for the radar experiment at 2300 UTC initialization indicate that the dominant precipitation feature in most of the ensemble members has an eastward displacement error by the 1-h time, and the mean displacement in this direction increases with time through the first 3 h of the ensemble forecast (Figs. 8a,c). During the 3–6-h forecast period, the magnitude of the displacement errors continues to grow, but interpretation of output from the object-tracking algorithm becomes more ambiguous, likely due to differences in the splitting, merging, decay, etc. of the features in the different datasets. These characteristics are corroborated by measures of the degree of overlap of predicted and observed features, as well as their individual sizes (Figs. 8b,e, respectively). Each member has a high degree of overlap with observations 1 h into the forecast, with a slow drop-off through 2 h, and a more rapid decline thereafter (Fig. 8b). In terms of peak intensity, observations fall squarely within the envelope of predicted values through the 2-h time period, but in general, the predicted features appear to decay after that time while observed precipitation rates continue to increase (Fig. 8d). The dominant forecast features are uniformly larger than the matching observed features at the 1-h forecast time, but the former tend to shrink after 2–3 h, with more than half of them completely dissipating by hour 5, in contrast to observations (Fig. 8e). The picture that emerges from these feature-specific results is much more nuanced than, but nonetheless consistent with, the QPF fields presented in Fig. 7.

Application of the object-tracking algorithm to the no_radar output is more challenging because, in the absence of radar-data assimilation, less than half of the ensemble members generate features that meet the criteria for direct correspondence (matching) with the dominant observed precipitation feature (Fig. 9). For those members whose output contains matching features, a relatively large northeastward displacement is evident even after the first hour (Figs. 9a,c). A few members produce matching objects that overlap significantly with the dominant observed feature, but the degree of overlap decreases sharply with time through the 3-h forecast (Fig. 9b). In terms of peak intensity, the matching objects bracket the observations quite well during the first hour, but the simulated storms fail to match the modest increase in observed peak rainfall thereafter (Fig. 9d). The matching objects are generally somewhat larger than the dominant observed feature at the 1-h time, but they come into better agreement at later times (Fig. 9e).

Focusing now on the 0000 UTC initialization time and corresponding observations (Fig. 10), one can see

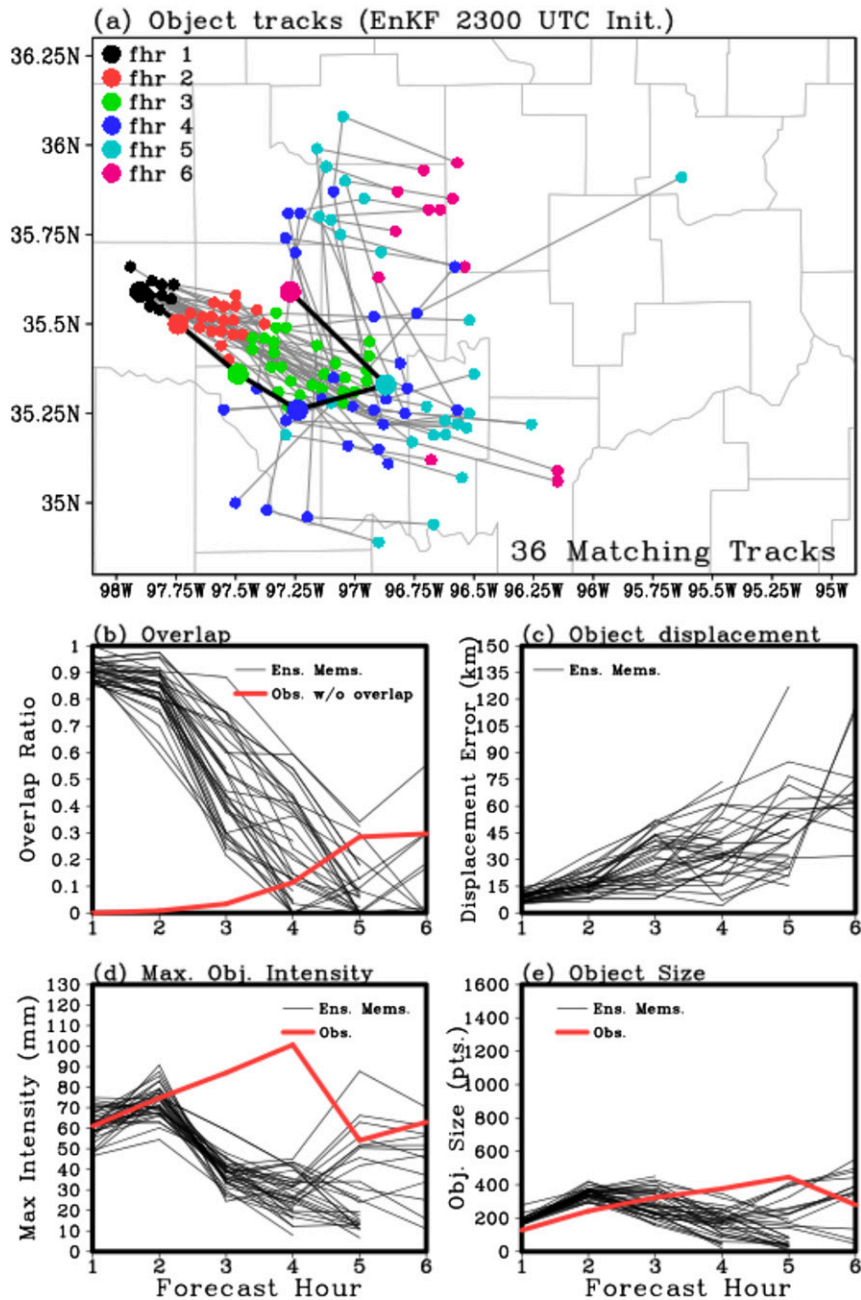


FIG. 8. (a) The observed precipitation object track (6.35-mm criteria) over central OK starting at 2300 UTC 31 May 2013 is indicated by the thick black line. Larger circles denote the centroid locations of the objects comprising the observed track, where the colors denote the specific forecast hour as indicated by the legend at the top left. The gray lines and smaller circles show the same thing, but they are for matching forecast object tracks from the radar experiment initialized at 2300 UTC. (b) Black lines indicate the proportion of the observed object overlapped by each radar ensemble member at each forecast hour. The red line indicates the proportion of the observed object not overlapped by any of the forecast objects. (c) Displacement error at each forecast hour for matching forecast objects in each radar ensemble member. (d) As in (c), but for maximum object intensity, and the thick red line is for the observed object track. (e) As in (d), but for object size.

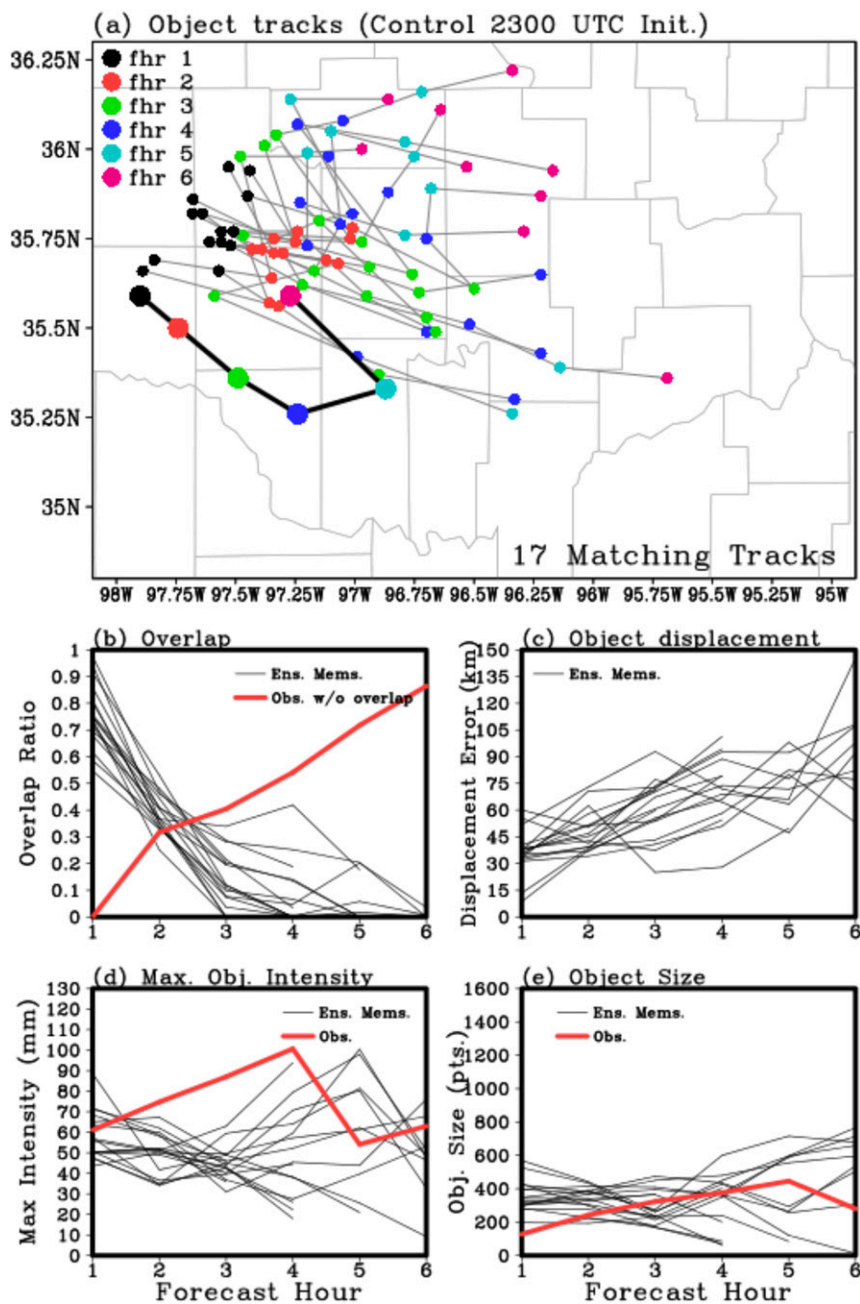


FIG. 9. As in Fig. 8, but for the no_radar experiment.

that the heaviest rain fell in a band from west-central Canadian County to east-central Oklahoma County during the 0000–0100 UTC period (Fig. 10b). The ensemble-mean QPF from the radar experiment reproduces this rainfall pattern quite well (cf. Figs. 10b,c), but appears to generate too much rainfall near the center of this band and perhaps not enough on the eastern and western ends of it. However, these errors are minimal compared to those from the corresponding prediction from the no_radar experiment, which significantly

underpredicts the maximum rainfall amount in this first hour and suffers a downstream (relative to the midlevel flow) displacement error (cf. Figs. 10a–c).

For the longer accumulation periods, the no_radar experiment again produces heavy rainfall (Figs. 10d,g) although the heavier amounts remain displaced to the north and east by a county or so (cf. Figs. 10d,e and 10g,h). The maximum accumulation in the no_radar experiment almost doubles from the 0–3- to the 0–6-h period (cf. Figs. 10d,g), suggesting that the extreme event is

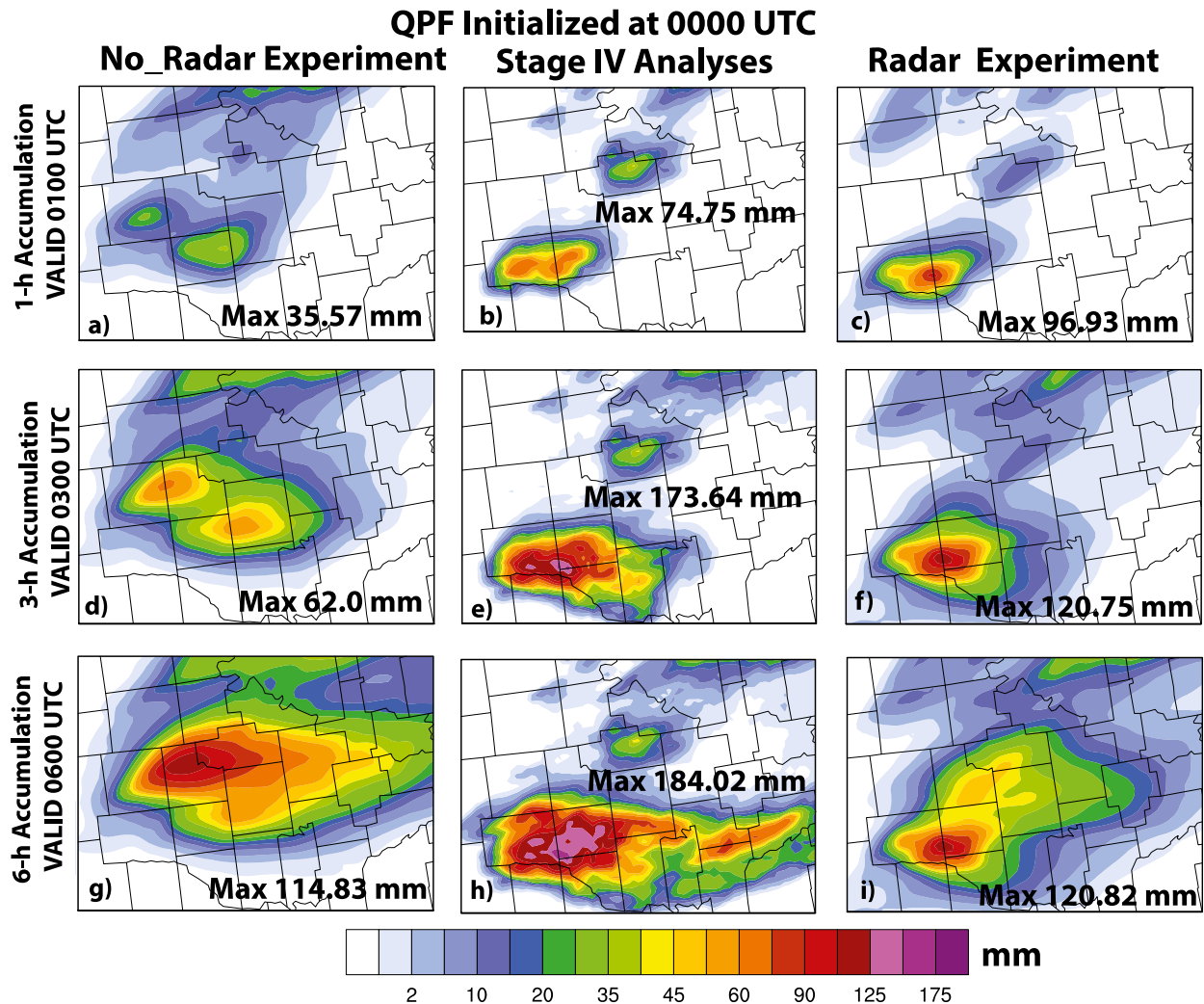


FIG. 10. Ensemble mean 1-, 3-, and 6-h accumulated precipitation forecasts initialized at 0000 UTC from the (a),(d),(g) no_radar and (c),(f),(i) radar experiments, as well as (b),(e),(h) NCEP stage IV accumulated precipitation analyses during the same period. The portion of the domain shown here is over central OK.

displaced in both time and space with this prediction system. Interestingly, the 0–6-h heavy-rainfall feature in this experiment has a configuration and shape that closely resemble corresponding observations (cf. Figs. 10g,h). Meanwhile, the maximum ensemble-mean accumulated rainfall from the radar experiment remains anchored where it emerged during the first forecast hour (cf. Figs. 10c,d,i), again suggesting that the added value associated with radar-data assimilation declines rapidly with model integration time.

The feature-tracking diagnostics in the radar experiment reveal that the tracks of the dominant feature in each of the ensemble members match the observations quite well through 2 h but, generally, diverge significantly thereafter (Figs. 11a,c)—likewise with the sizes in each realization of this feature (Fig. 11e). The degrees of

overlap and the maximum rainfall rates both drop sharply after the first hour of the forecast (Figs. 11b,d). In contrast, in the no_radar experiment, the displacement errors are relatively large even after 1 h and the degree of spatial overlap with the observed feature is relatively small (cf. Figs. 12a–c to Figs. 11a–c). On the other hand, the peak intensity of the ensemble members in the no_radar experiment tends to lag the observations (the spinup issue), but the envelope of maximum-intensity solutions is quite consistent with a good probabilistic representation of the observations during the 3–6-h period (Fig. 12d), and the envelope of feature-size solutions contains the observed feature size during most of the 0–6-h period (Fig. 12e). In general, comparing Figs. 11 and 12, it is quite clear that the assimilation of radar data yields an initial condition that is characterized by relatively much

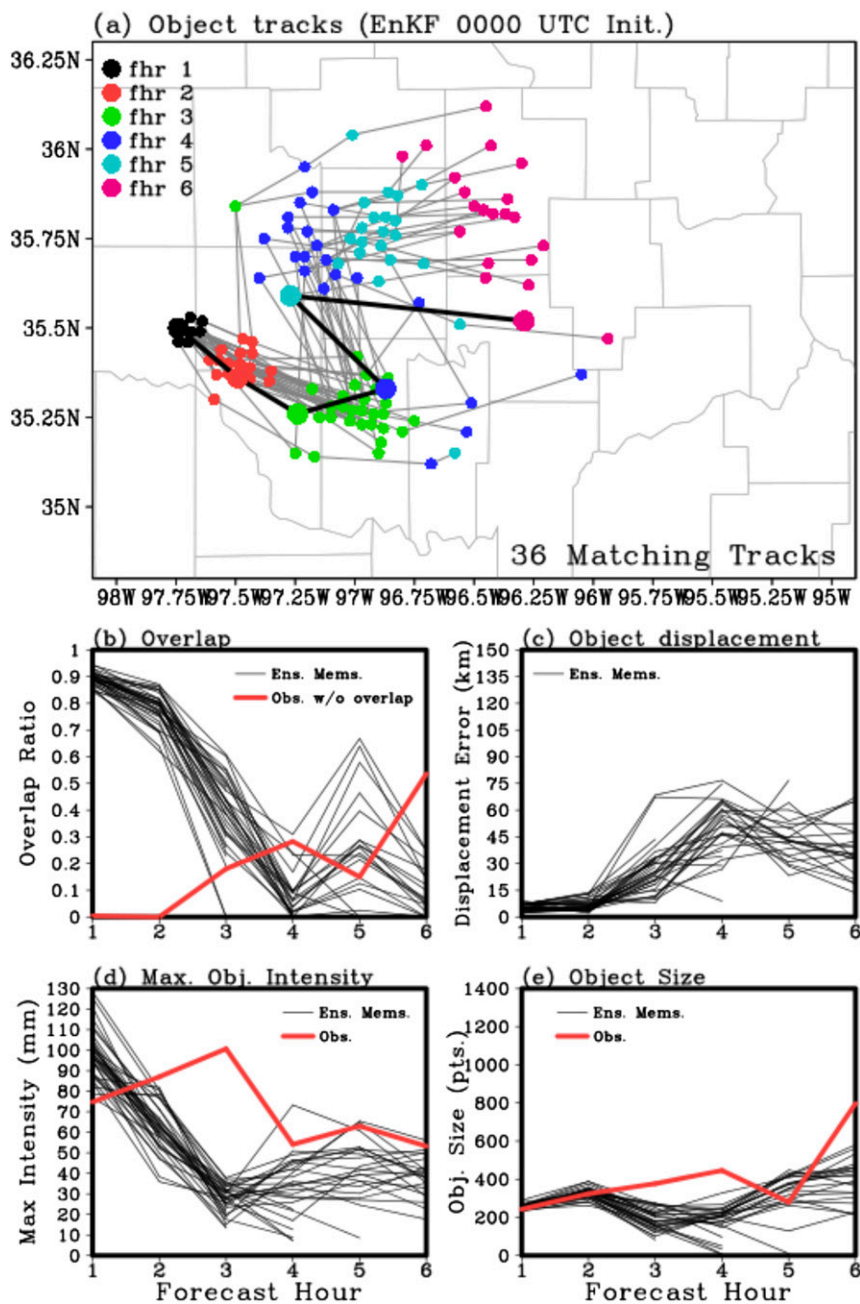


FIG. 11. As in Fig. 8, but for forecasts initialized at 0000 UTC 1 Jun 2013.

sharper ensemble definition of the deep-convective storm that dominated this event, but the added value of this convective-scale definition is not necessarily carried much beyond the 0–3-h forecast time.

d. Probabilistic quantitative precipitation forecasts

The ensemble-based probability of 0–3- and 3–6-h accumulated rainfall (greater than 25 mm) are compared with the observed stage IV 25-mm rainfall amounts from the forecasts initialized during the 2300 and 0000 UTC analyses

(Figs. 13 and 14, respectively). The 0–3-h PQPF from the radar experiment (Figs. 13b and 14b) is able to associate 100% probabilities with the observed location reasonably well. The 0–3-h PQPF from the no_radar experiment generates comparatively lower probabilities of the dominant precipitation core, and its probability field is displaced to the north relative to the observations (Figs. 13a and 14a). Not surprisingly, the 3–6-h probability fields are lower in magnitude than the 0–3-h fields for both experiments, and the northward displacement errors are also more pronounced.

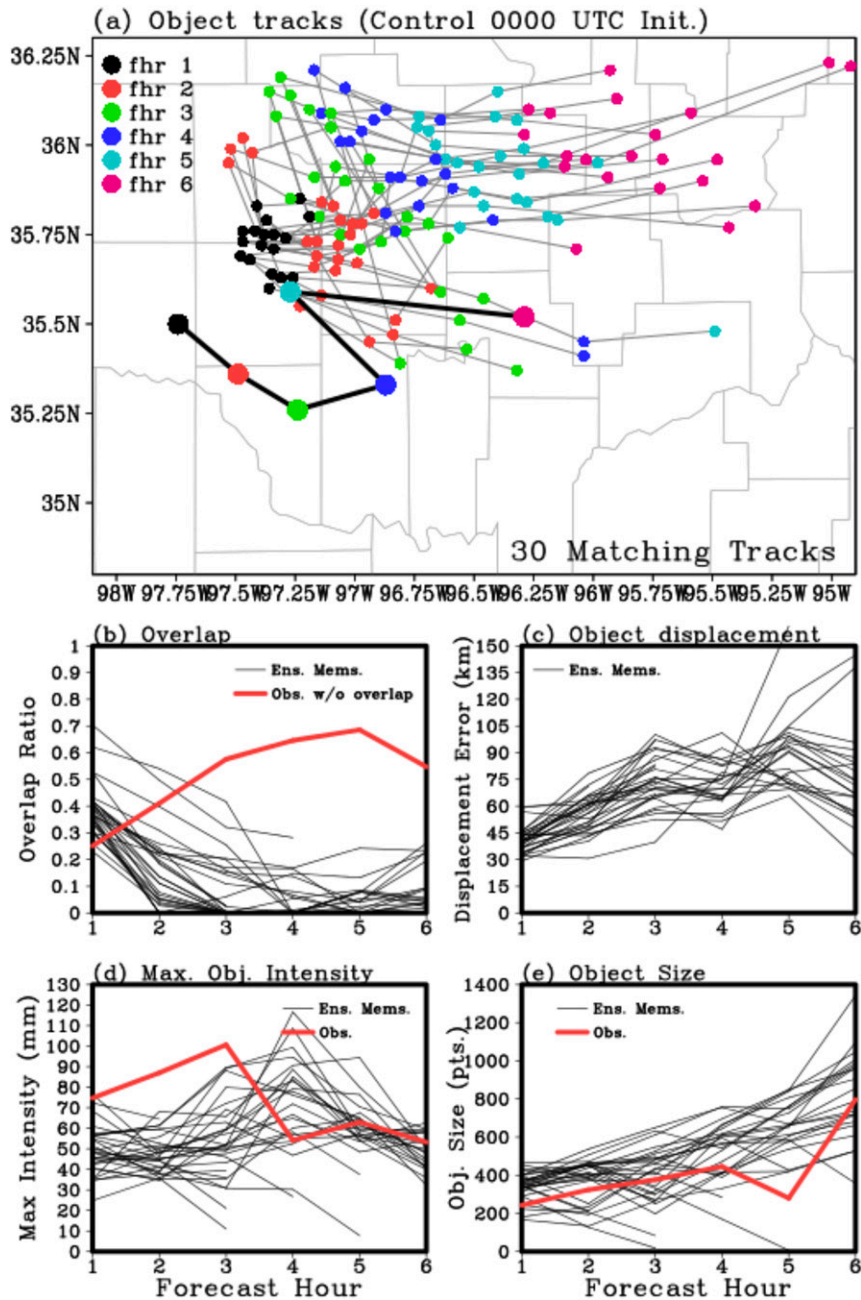


FIG. 12. As in Fig. 9, but for forecasts initialized at 0000 UTC 1 Jun 2013 from the no_radar experiment.

Overall, the results indicate that the radar experiment corresponds to the stage IV analyses much more closely than does the no_radar experiment during the 0–3-h forecast period. The spatial uncertainty is greatly reduced by the explicit initialization of storms using radar data in this experiment. The implications of this impact are significant because flash flood forecasting can depend very heavily on the accurate prediction of the specific location of the heavy rainfall. Heavy rainfall is more likely to result in flash

flooding when it occurs in a flash-flood-prone area or basin. Small errors in predicting the location of heavy precipitation can lead to much bigger errors in predictions from hydrologic models and/or human forecasters. Therefore, the rapid-update-cycle ensemble NWP system used here, including the assimilation of radar data, has the potential to improve forecasts of flash flood events substantially compared to current operational forecast systems. This assessment is corroborated by applying

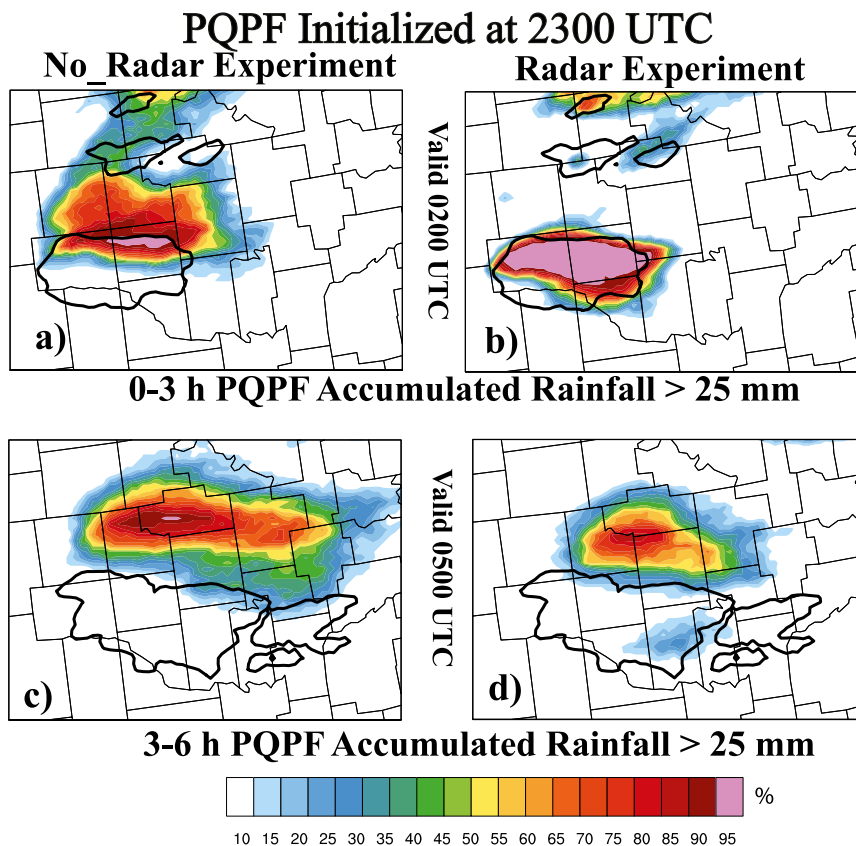


FIG. 13. The ensemble probability of rainfall greater than 25 mm (colors, 5% increment) from 0–3- and 3–6-h accumulated precipitation forecasts initialized at 2300 UTC. The thick black contour is the stage IV 25-mm precipitation contour. The portion of the domain shown here is over central OK.

objective verification metrics to the probabilistic forecasts, focusing on the domain shown in Figs. 13 and 14.

1) FRACTIONS SKILL SCORES

The FSSs from the radar experiment are considerably higher than those from the no_radar experiment for both 15 and 30 mm h⁻¹ threshold values during the initial 0–3-h forecast period. The differences in skill between the two experiments are as high as 0.82 (Fig. 15c) for the initial 1-h forecast, but the differences decrease as forecast lead time increases (Fig. 15). The FSSs from both experiments are lower for the higher 30 mm h⁻¹ threshold value compared to the 15 mm h⁻¹ threshold, indicating that accurate prediction of higher precipitation amounts is more challenging.

This challenge is revealed in more detail by examining the FSSs as a function of increasing precipitation threshold for the 0–3- and 3–6-h periods (Fig. 16). The 0–3-h FSSs from the radar experiment are consistently higher than 0.80 for the lowest (5 mm) threshold during the initial 0–3-h period (Figs. 16a–c), decreasing to lower values for

higher thresholds. The 3–6-h FSSs are much lower at all thresholds. The decrease in FSS with increasing threshold value indicates, once again, that the ensemble prediction system tends to be less skillful at predicting the specific location of intense rainfall features. This tendency is likely due to the fact that higher precipitation thresholds highlight smaller “features,” which tends to reduce the degree of overlap between the observed and predicted features for given errors in displacement and amplitude (see, e.g., Baldwin and Kain 2006).

2) AREA UNDER THE RELATIVE OPERATING CHARACTERISTICS CURVE

Consistent with the FSSs, the forecasts initialized from 2300, 0000, and 0100 UTC analyses as a function of forecast hours (Fig. 17) and increasing rainfall thresholds (Fig. 18) indicate that the ROC areas from the radar experiment are generally higher compared to the no_radar experiment. ROC areas greater than 0.5 indicate successful discriminating ability (Schwartz et al. 2015 and references therein), with a value of 1.0 indicating a

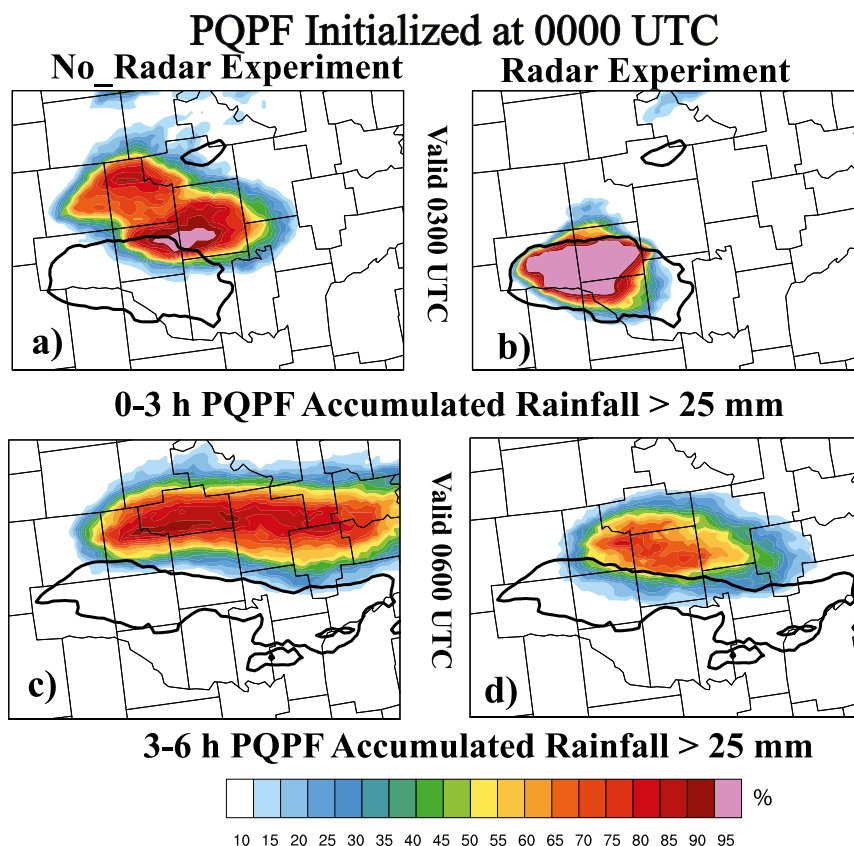


FIG. 14. As in Fig. 13, but initialized at 0000 UTC.

perfect score. Both experiments yield relatively high ROC areas early in the forecast (Fig. 17). The ROC areas from the radar experiment start with values approaching 1.0 (Figs. 17c,f) and generally remain greater than ~ 0.5 at all forecast lead times for both 15 and 30 mm h^{-1} thresholds, indicating skillful discriminating ability. The ROC areas from the no_radar experiment are greater than ~ 0.5 earlier in the forecasts hours but the value drops to less than 0.5 at later times. The 0–3-h rainfall accumulations from the radar experiment demonstrate high ROC areas for all rainfall thresholds with values above 0.9 (Figs. 18a,b), indicating skillful discriminating ability of both lighter and heavier precipitations amounts. The ROC areas are smaller for the later 3–6-h accumulation period (Figs. 18d–f). The differences in ROC areas from the two experiments are very small for heavier rainfall amounts.

6. Summary and conclusions

NOAA's WoF program is designed to provide NWS forecasters with the probabilistic guidance needed to substantially improve short-term forecasts and warnings for convective weather threats such as tornadoes, large

hail, strong winds, and flash floods. This study focuses on an event with two distinct threats: tornado and heavy convective rainfall, which is often a precursor to flash flooding. It examines the 31 May 2013 severe weather event over central Oklahoma, in which 22 people were killed—8 by a tornado and 14 by flash floods associated with the same storm complex.

Retrospective short-term (0–6 h) probabilistic ensemble forecasts of this event are generated. Specifically, two sets of 0–6-h ensemble-forecast experiments are conducted at storm scale. The first set is initialized from the hourly updated prediction system that assimilates routinely available observations from METARs, marine sources, mesonet reports, ACARS, and radiosonde and satellite-derived winds. The other set is initialized from a system that is updated every 5 min and assimilates both WSR-88D radar data and the set of routinely available observations described above. The goal is to quantify the benefits of assimilating conventional and radar observations in initializing short-term rainfall forecasts, while validating the forecast system by comparing predictions of low- and midlevel storm rotations to observed rotation tracks and the path of a large tornado that hit El Reno, Oklahoma.

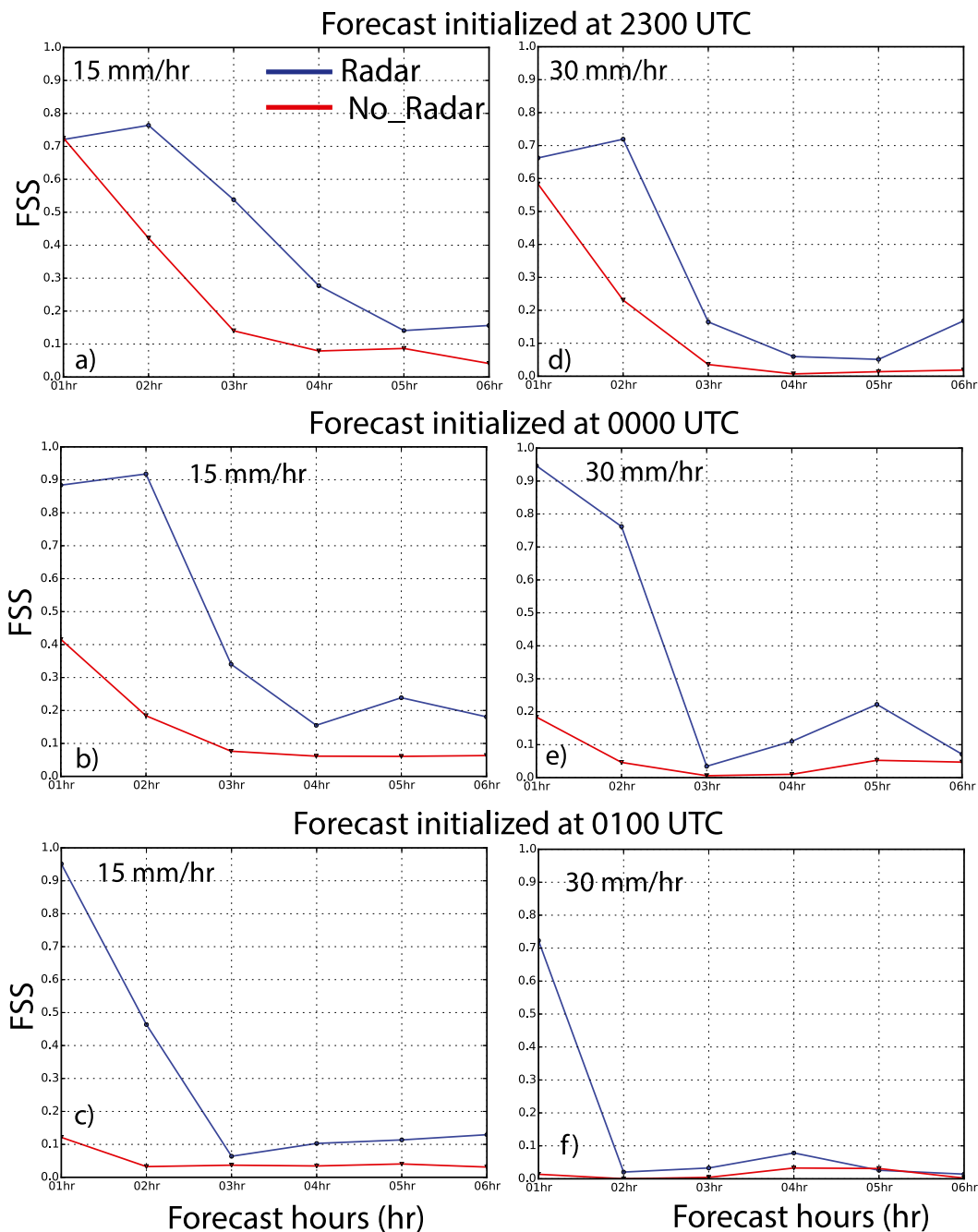


FIG. 15. FSS as a function of forecast hour for (a)–(c) 15 and (d)–(f) 30 mm h⁻¹ rainfall threshold values from the ensemble forecasts initialized at 2300, 0000, and 0100 UTC. The blue line is from the radar experiment, and the red line is from the no_radar experiment.

Observation-space diagnostic statistics reveal that the ensemble Kalman filter shows no sign of forecast divergence during the 7-h assimilation period, indicating robustness of the data assimilation system. For both radial velocity and reflectivity observations, the rmsi and ensemble spread are of comparable magnitude, indicating that the 5-min forecast error is representative of the ensemble spread.

The 0–6-h probabilistic forecasts of reflectivity greater than 40 dBZ from the radar experiment show that the ensemble is successful at associating high-reflectivity probabilities with the dominant observed storms in the analyses. The reflectivity probability fields have high amplitude and are well aligned with the locations of the observed storms early in the forecast period, but the

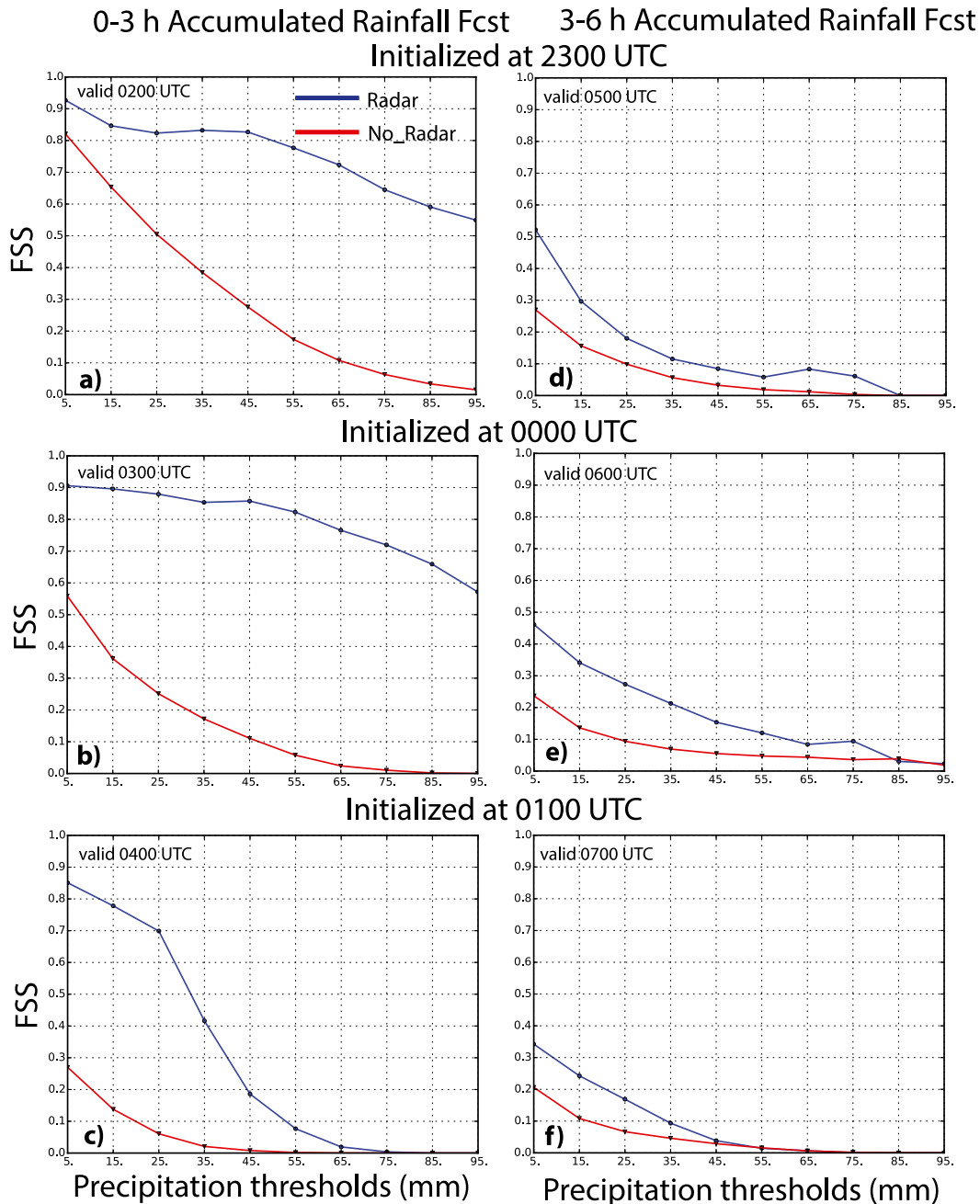


FIG. 16. FSS as a function of precipitation threshold (mm) for 0–3- and 3–6-h accumulated precipitation ensemble forecasts initialized from (a),(d) 2300, (b),(e) 0000, and (c),(f) 0100 UTC analyses. The blue line is from the radar experiment and the red line is from the no_radar experiment.

probability values gradually decrease as the forecast hour increases, and the probability field is clearly displaced to the north and east relative to the observations by the end of the period. The biases associated with the motion of simulated storms are likely due to the model error, but this is the subject of ongoing investigation.

The radar experiment was also successful at initializing the supercell responsible for the El Reno tornado and predicting a high probability of strong low-level rotation along a path that corresponded reasonably well to the observed rotation tracks associated with this storm and the associated deadly tornado. Overall, there is a consistent and gradual increase in the probability of

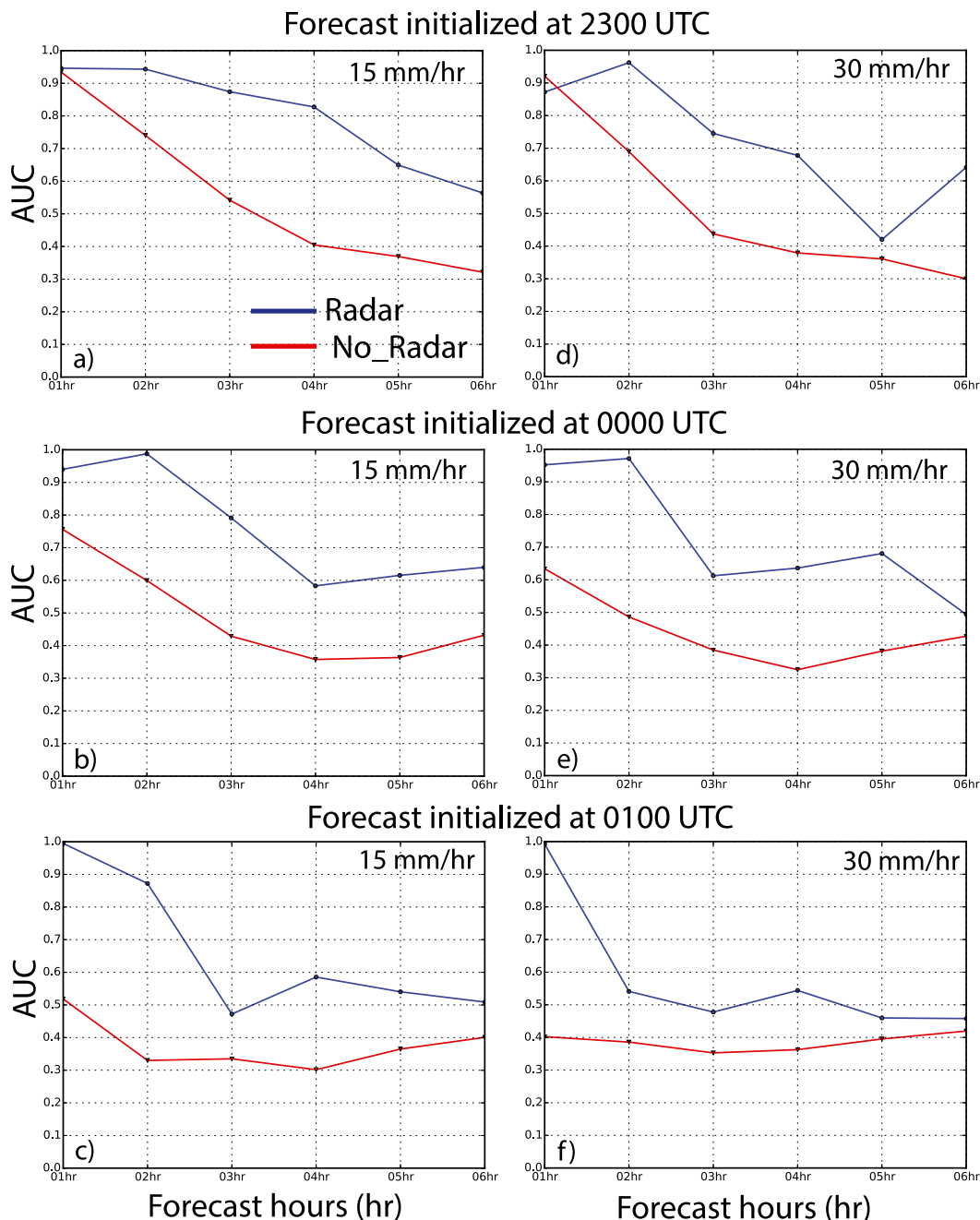


FIG. 17. AUC as a function of forecast hours for (a)–(c) 15 and (d)–(f) 30 mm h⁻¹ rainfall thresholds from the ensemble forecasts initialized at 2300, 0000, and 0100 UTC. The blue line is from the radar experiment, and the red line is from the no_radar experiment.

low-level rotation with each successively later initialization leading up to the time of tornadogenesis.

The ensemble mean QPFs from the radar experiment correspond quite well to the stage IV precipitation analyses for the first 1–2 h of the forecasts. In contrast, the no_radar experiment generates noticeably lower rainfall rates during the early forecast hours.

Beyond the 0–3-h forecast period, both experiments produce heavy rainfall but they also suffer from significant downstream (relative to the midlevel steering flow) displacement errors for the heaviest highest rainfall rates.

A consistent picture is painted by the probability forecasts that the ensemble systems provide. During the

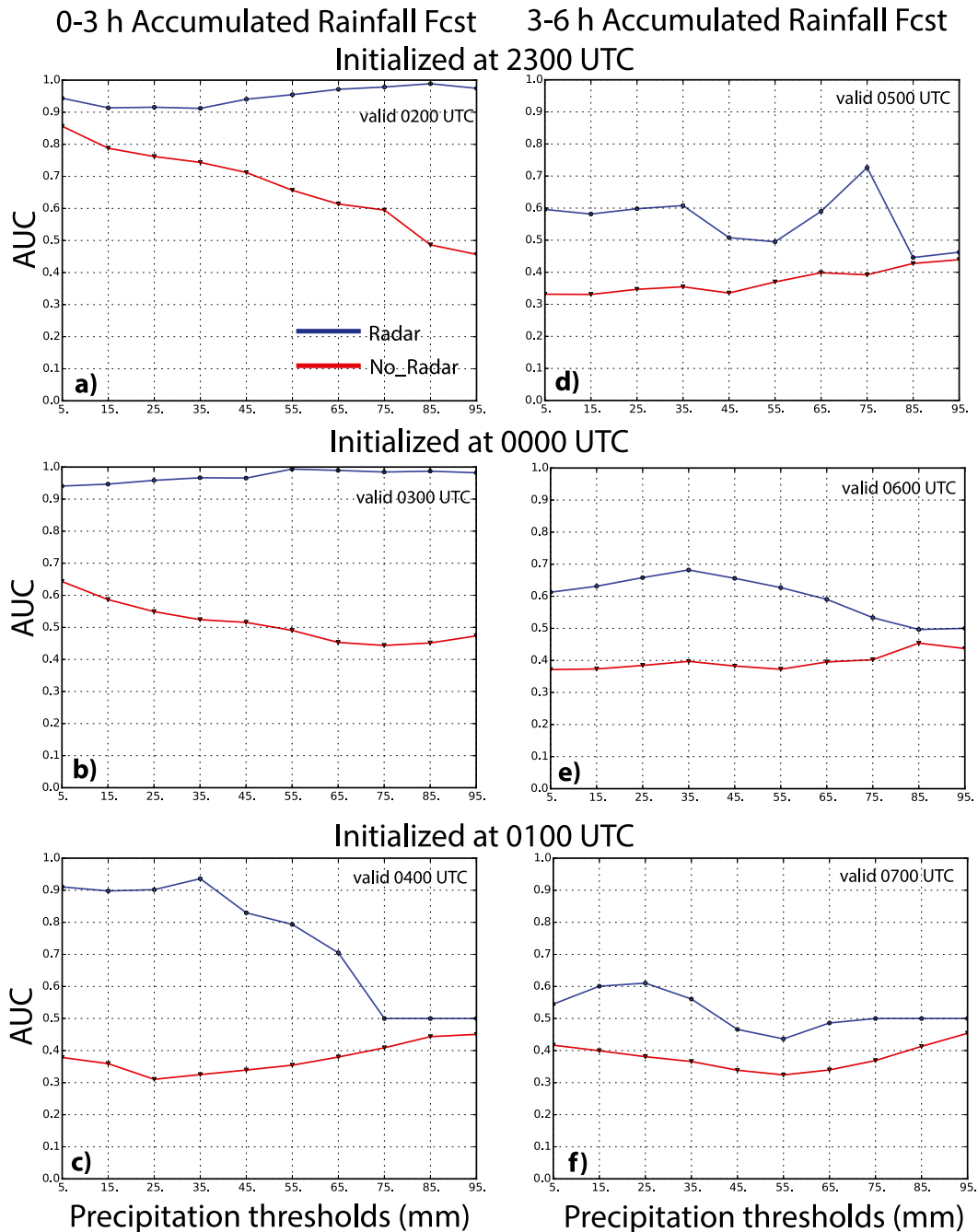


FIG. 18. AUC as a function of precipitation thresholds (mm) for 3- and 6-h accumulated precipitation ensemble forecasts initialized from (a),(d) 2300, (b),(e) 0000, and (c),(f) 0100 UTC. The blue line is from the radar experiment, and the red line is from the no_radar experiment.

first 1–2 h of each relevant forecast period, the radar experiment generates high probabilities of heavy precipitation very close to where high rainfall rates actually occurred. In contrast, the probability fields from the no_radar experiment imply that there are significant model errors in the timing and placement of heavy precipitation during this same forecast period. Both

experiments produce large areas with greater than 50% probability of heavy precipitation during the 3–6-h period, but the probability fields are displaced downstream from the areas where the heaviest precipitation was observed. These subjective assessments of the probability forecasts are corroborated using FSSs and AUC as objective metrics.

Additional insight is provided by applying a feature-tracking algorithm that identifies, characterizes, and tracks precipitation features over time. This algorithm provides important information about the representation of individual storms in the ensemble members and how this representation is translated into the prediction parameters of the ensemble. The results indicate that the range of solutions in the radar experiment captures the maximum rainfall rates of the El Reno storm remarkably well for the first 2 h of key forecast periods. The size of this dominant feature is generally captured well during this forecast period too, while position/track errors grow, but at a slow rate during this period. Beyond the 2-h time period, errors grow more rapidly and the range of predicted solutions, in terms of the different storm-feature attributes, becomes less likely to encompass observed storm attributes in later forecast hours. The higher disparity between the simulated and observed storm attributes is generally true for the no_radar experiment from the start of the forecast period.

In summary, the results presented here reveal that the type of continuous-update-cycle storm-scale ensemble system that is being developed for the WoF project shows great promise for 1–2-h predictions of intense convective rainfall. This result is particularly significant because it suggests that this type of prediction system could be very helpful for increasing lead times for forecasts and warnings of flash floods. Specifically, a WoF-type system could be used to initialize hydrologic models with realistic rainfall timing, intensity, and location well before a severe rainfall event occurs. In turn, these models could provide critically important flash flood guidance for forecasters 1–2 h earlier than the current paradigm allows because the models would be driven by heavy rainfall forecasts rather than detection, consistent with the fundamental basis of the WoF initiative.

To explore the robustness of this system beyond this single case study, experiments with multiple heavy rainfall and flash flood events will be examined in future studies. Needless to say, continued improvements in all aspects of the WoF system must be made to extend the skillful forecast lead time beyond the first 2 h. For example, microphysical and planetary boundary layer parameterizations appear to be significant sources of error that can severely limit the practical predictability of storm-scale model forecasts (Houtekamer et al. 2005), even when initial conditions are generated with fidelity using sophisticated radar data assimilation schemes. Designing storm-scale ensembles that better represent typical forecast errors several hours in advance also is crucial (Yussouf et al. 2015). Storm-scale data assimilation is essentially a retrieval problem; most of the state variables are severely underconstrained by the

observations (here, reflectivity and radial velocity). Additional storm-scale observations have recently become available from the WSR-88D radar network in the form of dual-polarization variables. A major challenge now is to understand how to best assimilate those dual-polarization observations, which opens up an active area of research in the storm-scale data assimilation community (Jung et al. 2008a,b; Posselt et al. 2015).

Acknowledgments. The computing for this project was performed at the University of Oklahoma (OU) Supercomputing Center for Education and Research (OSCER). Thanks to Chris Karstens for helping with the WSR-88D radar data quality control. Thanks to Carrie Langston for the MRMS reflectivity data. Local computer assistance is provided by Brett Morrow, Steven Fletcher, and Robert Coggins. Partial funding for this research was provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA–University of Oklahoma Cooperative Agreement NA11OAR4320072, U.S. Department of Commerce.

REFERENCES

- Accadia, C., S. Mariani, M. Casaioli, A. Lavagnini, and A. Speranza, 2003: Sensitivity of precipitation forecast skill scores to bilinear interpolation and a simple nearest-neighbor average method on high-resolution verification grids. *Wea. Forecasting*, **18**, 918–932, doi:10.1175/2003-018<0918:SOPFSS>2.0.CO;2.
- Aksoy, A., D. Dowell, and C. Snyder, 2009: A multicaser comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part I: Storm-scale analyses. *Mon. Wea. Rev.*, **137**, 1805–1824, doi:10.1175/2008MWR2691.1.
- Anderson, J. L., 2001: An ensemble adjustment filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903, doi:10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.
- , 2009: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus*, **61A**, 72–83, doi:10.1111/j.1600-0870.2008.00361.x.
- , and N. Collins, 2007: Scalable implementations of ensemble filter algorithms for data assimilation. *J. Atmos. Oceanic Technol.*, **24**, 1452–1463, doi:10.1175/JTECH2049.1.
- , T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Avellano, 2009: The Data Assimilation Research Testbed: A community facility. *Bull. Amer. Meteor. Soc.*, **90**, 1283–1296, doi:10.1175/2009BAMS2618.1.
- Ashley, S. T., and W. S. Ashley, 2008: Flood fatalities in the United States. *J. Appl. Meteor. Climatol.*, **47**, 805–818, doi:10.1175/2007JAMC1611.1.
- Baldwin, M. E., and K. E. Mitchell, 1997: The NCEP hourly multisensory U.S. precipitation analysis for operations and GCIIP research. Preprints, *13th Conf. on Hydrology*, Long Beach, CA, Amer. Meteor. Soc., 54–55.
- , and J. S. Kain, 2006: Sensitivity of several performance measures to displacement error, bias, and event frequency. *Wea. Forecasting*, **21**, 636–648, doi:10.1175/WAF933.1.
- Barthold, F. E., T. E. Workoff, B. A. Cosgrove, J. J. Gourley, D. R. Novak, and K. M. Mahoney, 2015: Improving flash flood

- forecasts: The HMT-WPC flash flood and intense rainfall experiment. *Bull. Amer. Meteor. Soc.*, **96**, 1859–1866, doi:10.1175/BAMS-D-14-00201.1.
- Bluestein, H. B., J. C. Snyder, and J. B. Houser, 2015: A multiscale overview of the El Reno, Oklahoma, tornadic supercell of 31 May 2013. *Wea. Forecasting*, **30**, 525–552, doi:10.1175/WAF-D-14-00152.1.
- Carley, J. R., B. R. J. Schwedler, M. E. Baldwin, R. J. Trapp, J. Kwiatkowski, J. Logsdon, and S. J. Weiss, 2011: A proposed model-based methodology for feature-specific prediction for high-impact weather. *Wea. Forecasting*, **26**, 243–249, doi:10.1175/WAF-D-10-05008.1.
- Chang, H., H. Yuan, and P. Lin, 2012: Short-range (0–12 h) PQPFs from time-lagged multimodel ensembles using LAPS. *Mon. Wea. Rev.*, **140**, 1496–1516, doi:10.1175/MWR-D-11-00085.1.
- Chen, H., D. Yang, Y. Hong, J. Gourley, and Y. Zhang, 2013: Hydrological data assimilation with the ensemble square-root-filter: Use of streamflow observations to update model states for real-time flash flood forecasting. *Adv. Water Resour.*, **59**, 209–220, doi:10.1016/j.advwatres.2013.06.010.
- Clark, A. J., W. A. Gallus Jr., M. Xue, and F. Kong, 2009: A comparison of precipitation forecast skill between small convection-allowing and large convection parameterizing ensembles. *Wea. Forecasting*, **24**, 1121–1140, doi:10.1175/2009WAF222222.1.
- , —, and M. L. Weisman, 2010: Neighborhood-based verification of precipitation forecasts from convection-allowing NCAR WRF model simulations and the operational NAM. *Wea. Forecasting*, **25**, 1495–1509, doi:10.1175/2010WAF2222404.1.
- , J. S. Kain, P. T. Marsh, J. Correia, M. Xue, and F. Kong, 2012: Forecasting tornado pathlengths using a three-dimensional object identification algorithm applied to convection-allowing forecasts. *Wea. Forecasting*, **27**, 1090–1113, doi:10.1175/WAF-D-11-00147.1.
- , J. Gao, P. T. Marsh, T. Smith, J. S. Kain, J. Correia Jr., M. Xue, and F. Kong, 2013: Tornado pathlength forecasts from 2010 to 2011 using ensemble updraft helicity. *Wea. Forecasting*, **28**, 387–407, doi:10.1175/WAF-D-12-00038.1.
- Dawson, D. T., II, L. J. Wicker, E. R. Mansell, and R. L. Tanamachi, 2012: Impact of the environmental low-level wind profile on ensemble forecasts of the 4 May 2007 Greensburg, Kansas, tornadic storm and associated mesocyclones. *Mon. Wea. Rev.*, **140**, 696–716, doi:10.1175/MWR-D-11-00008.1.
- Dowell, D. C., and L. J. Wicker, 2009: Additive noise for storm-scale ensemble forecasting and data assimilation. *J. Atmos. Oceanic Technol.*, **26**, 911–927, doi:10.1175/2008JTECHA1156.1.
- , F. Zhang, L. J. Wicker, C. Snyder, and N. A. Crook, 2004: Wind and temperature retrievals in the 17 May 1981 Arcadia, Oklahoma, supercell: Ensemble Kalman filter experiments. *Mon. Wea. Rev.*, **132**, 1982–2005, doi:10.1175/1520-0493(2004)132<1982:WATRIT>2.0.CO;2.
- , L. J. Wicker, and C. Snyder, 2011: Ensemble Kalman filter assimilation of radar observations of the 8 May 2003 Oklahoma City supercell: Influences of reflectivity observations on storm-scale analyses. *Mon. Wea. Rev.*, **139**, 272–294, doi:10.1175/2010MWR3438.1.
- Duc, L., K. Saito, and H. Seko, 2013: Spatial-temporal fractions verification for high resolution ensemble forecasts. *Tellus*, **65A**, 18171, doi:10.3402/tellusa.v65i0.18171.
- Ebert, E. E., 2001: Ability of a poor man's ensemble to predict the probability and 337 distribution of precipitation. *Mon. Wea. Rev.*, **129**, 2461–2480, doi:10.1175/1520-0493(2001)129<2461:AOAPMS>2.0.CO;2.
- Gagne, D. J., II, A. McGovern, and M. Xue, 2014: Machine learning enhancement of storm-scale ensemble probabilistic quantitative precipitation forecasts. *Wea. Forecasting*, **29**, 1024–1043, doi:10.1175/WAF-D-13-00108.1.
- Houtekamer, P. L., H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2005: Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Wea. Rev.*, **133**, 604–620, doi:10.1175/MWR-2864.1.
- Jones, T. A., and D. J. Stensrud, 2012: Assimilating AIRS temperature and mixing ratio profiles using an ensemble Kalman filter approach for convective-scale forecasts. *Wea. Forecasting*, **27**, 541–564, doi:10.1175/WAF-D-11-00090.1.
- Jung, Y., G. Zhang, and M. Xue, 2008a: Assimilation of simulated polarimetric radar data for a convective storm using the ensemble Kalman filter. Part I: Observation operators for reflectivity and polarimetric variables. *Mon. Wea. Rev.*, **136**, 2228–2245, doi:10.1175/2007MWR2083.1.
- , M. Xue, G. Zhang, and J. Straka, 2008b: Assimilation of simulated polarimetric radar data for a convective storm using the ensemble Kalman filter. Part II: Impact of polarimetric data on storm analysis. *Mon. Wea. Rev.*, **136**, 2246–2260, doi:10.1175/2007MWR2288.1.
- Kain, J. S., S. J. Weiss, J. J. Levit, M. E. Baldwin, and D. R. Bright, 2006: Examination of convection-allowing configurations of the WRF model for the prediction of severe convective weather: The SPC/NSSL Spring Program 2004. *Wea. Forecasting*, **21**, 167–181, doi:10.1175/WAF906.1.
- , and Coauthors, 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Wea. Forecasting*, **23**, 931–952, doi:10.1175/WAF2007106.1.
- , and Coauthors, 2010: Assessing advances in the assimilation of radar data within a collaborative forecasting–research environment. *Wea. Forecasting*, **25**, 1510–1521, doi:10.1175/2010WAF2222405.1.
- Lakshmanan, V., T. Smith, G. Stumpf, and K. Hondl, 2007: The Warning Decision Support System–Integrated Information. *Wea. Forecasting*, **22**, 596–612, doi:10.1175/WAF1009.1.
- Lin, Y., and K. Mitchell, 2005: The NCEP Stage II/IV hourly precipitation analyses: Development and applications. Preprint, *19th Conf. on Hydrology*, San Diego, CA, Amer. Meteor. Soc., 1.2. [Available online at <https://ams.confex.com/ams/pdfpapers/83847.pdf>.]
- Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.*, **30**, 291–303.
- Miller, M. L., V. Lakshmanan, and T. Smith, 2013: An automated method for depicting mesocyclone paths and intensities. *Wea. Forecasting*, **28**, 570–585, doi:10.1175/WAF-D-12-00065.1.
- Nielsen, E. R., G. R. Herman, R. C. Tournay, J. M. Peters, and R. S. Schumacher, 2015: Double impact: When both tornadoes and flash floods threaten the same place at the same time. *Wea. Forecasting*, **30**, 1673–1693, doi:10.1175/WAF-D-15-0084.1.
- NOAA, 2014: Service assessment: May 2013 Oklahoma tornadoes and flash flooding. National Weather Service, 42 pp. + appendices. [Available online at http://www.nws.noaa.gov/om/assessments/pdfs/13oklahoma_tornadoes.pdf.]
- NWS WFO Norman, 2015: The May 31–June 1, 2013 tornado and flash flooding event. [Available online at <http://www.srh.noaa.gov/oun/?n=events-20130531>.]
- Posselt, D. J., X. Li, S. A. Tushaus, and J. R. Mecikalski, 2015: Assimilation of dual-polarization radar observations in mixed- and ice-phase regions of convective storms: Information content and forward model errors. *Mon. Wea. Rev.*, **143**, 2611–2636, doi:10.1175/MWR-D-14-00347.1.

- Potvin, C. K., and L. J. Wicker, 2013: Assessing ensemble forecasts of low-level supercell rotation within an OSSE framework. *Wea. Forecasting*, **28**, 940–960, doi:10.1175/WAF-D-12-00122.1.
- Roberts, N. M., 2005: An investigation of the ability of a storm scale configuration of the Met Office NWP model to predict flood-producing rainfall. Met Office Tech. Rep. 455, 80 pp. [Available online at http://research.metoffice.gov.uk/research/nwp/publications/papers/technical_reports/2005/FRTR455/FRTR455.pdf.]
- , and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, doi:10.1175/2007MWR2123.1.
- Romine, G. S., C. S. Schwartz, C. Snyder, J. L. Anderson, and M. L. Weisman, 2013: Model bias in a continuously cycled assimilation system and its influence on convection-permitting forecasts. *Mon. Wea. Rev.*, **141**, 1263–1284, doi:10.1175/MWR-D-12-00112.1.
- Schumacher, R. S., and A. J. Clark, 2014: Evaluation of ensemble configurations for the analysis and prediction of heavy-rain-producing mesoscale convective systems. *Mon. Wea. Rev.*, **142**, 4108–4138, doi:10.1175/MWR-D-13-00357.1.
- Schwartz, C. S., and Z. Liu, 2014: Convection-permitting forecasts initialized with continuously cycling limited-area 3DVAR, ensemble Kalman filter, and “hybrid” variational-ensemble data assimilation systems. *Mon. Wea. Rev.*, **142**, 716–738, doi:10.1175/MWR-D-13-00100.1.
- , and Coauthors, 2009: Next-day convection-allowing WRF model guidance: A second look at 2-km versus 4-km grid spacing. *Mon. Wea. Rev.*, **137**, 3351–3372, doi:10.1175/2009MWR2924.1.
- , and Coauthors, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Wea. Forecasting*, **25**, 263–280, doi:10.1175/2009WAF2222267.1.
- , G. S. Romine, K. R. Smith, and M. L. Weisman, 2014: Characterizing and optimizing precipitation forecasts from a convection-permitting ensemble initialized by a mesoscale ensemble Kalman filter. *Wea. Forecasting*, **29**, 1295–1318, doi:10.1175/WAF-D-13-00145.1.
- , —, M. L. Weisman, R. Sobash, K. Fossell, K. Manning, and S. Trier, 2015: A real-time convection-allowing ensemble prediction system initialized by mesoscale ensemble Kalman filter analyses. *Wea. Forecasting*, **30**, 1158–1181, doi:10.1175/WAF-D-15-0013.1.
- Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp, doi:10.5065/D68S4MVH.
- Snook, N., M. Xue, and Y. Jung, 2015: Multiscale EnKF assimilation of radar and conventional observations and ensemble forecasting for a tornadic mesoscale convective system. *Mon. Wea. Rev.*, **143**, 1035–1057, doi:10.1175/MWR-D-13-00262.1.
- Snyder, J. C., and H. B. Bluestein, 2014: Some considerations for the use of high-resolution mobile radar data in tornado intensity determination. *Wea. Forecasting*, **29**, 799–827, doi:10.1175/WAF-D-14-00026.1.
- Stensrud, D. J., and J. Gao, 2010: Importance of horizontally inhomogeneous environmental initial conditions to ensemble storm-scale radar data assimilation and very short-range forecasts. *Mon. Wea. Rev.*, **138**, 1250–1272, doi:10.1175/2009MWR3027.1.
- , and Coauthors, 2009: Convective-scale warn-on-forecast system: A vision for 2020. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1499, doi:10.1175/2009BAMS2795.1.
- , and Coauthors, 2013: Progress and challenges with Warn-on-Forecast. *Atmos. Res.*, **123**, 2–16, doi:10.1016/j.atmosres.2012.04.004.
- Sun, J., and Coauthors, 2014: Use of NWP for nowcasting convective precipitation: Recent progress and challenges. *Bull. Amer. Meteor. Soc.*, **95**, 409–426, doi:10.1175/BAMS-D-11-00263.1.
- Tang, Y., H. W. Lean, and J. Bornemann, 2013: The benefits of the Met Office variable resolution NWP model for forecasting convection. *Meteor. Appl.*, **20**, 417–426, doi:10.1002/met.1300.
- Toth, Z., Y. Zhu, and R. Wobus, 2004: March 2004 upgrades of the NCEP global ensemble forecast system. NOAA/NCEP/EMC. [Available online at http://www.emc.ncep.noaa.gov/gmb/ens/ens_imp_news.html.]
- VandenBerg, M. A., M. C. Coniglio, and A. J. Clark, 2014: Comparison of next-day convection-allowing forecasts of storm motion on 1- and 4-km grids. *Wea. Forecasting*, **29**, 878–893, doi:10.1175/WAF-D-14-00011.1.
- Wei, M., Z. Toth, R. Wobus, and Y. Zhu, 2008: Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system. *Tellus*, **60A**, 62–79, doi:10.1111/j.1600-0870.2007.00273.x.
- Weisman, M. L., C. A. Davis, W. Wang, K. W. Manning, and J. B. Klemp, 2008: Experiences with 0–36-h explicit convective forecasts with the WRF-ARW model. *Wea. Forecasting*, **23**, 407–437, doi:10.1175/2007WAF2007005.1.
- 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, doi:10.1175/WAF-D-15-0043.1.
- Wurman, J., K. Kosiba, P. Robinson, and T. Marshall, 2014: The role of multiple-vortex tornado structure in causing storm researcher fatalities. *Bull. Amer. Meteor. Soc.*, **95**, 31–45, doi:10.1175/BAMS-D-13-00221.1.
- Yussouf, N., J. Gao, D. J. Stensrud, and G. Ge, 2013a: The impact of mesoscale environmental uncertainty on the prediction of a tornadic supercell storm using ensemble data assimilation approach. *Adv. Meteor.*, **2013**, 1–15, doi:10.1155/2013/731647.
- , E. R. Mansell, L. J. Wicker, D. M. Wheatley, and D. J. Stensrud, 2013b: The ensemble Kalman filter analyses and forecasts of the 8 May 2003 Oklahoma City tornadic supercell storm using single- and double-moment microphysics schemes. *Mon. Wea. Rev.*, **141**, 3388–3412, doi:10.1175/MWR-D-12-00237.1.
- , D. C. Dowell, L. J. Wicker, K. Knopfmeier, and D. M. Wheatley, 2015: Storm-scale data assimilation and ensemble forecasts for the 27 April 2011 severe weather outbreak in Alabama. *Mon. Wea. Rev.*, **143**, 3044–3066, doi:10.1175/MWR-D-14-00268.1.
- Zhang, J., and Coauthors, 2011: National Mosaic and Multi-sensor QPE (NMQ) system: Description, results, and future plans. *Bull. Amer. Meteor. Soc.*, **92**, 1321–1338, doi:10.1175/2011BAMS-D-11-00047.1.