# Ensemble Probabilistic Forecasts of a Tornadic Mesoscale Convective System from Ensemble Kalman Filter Analyses using WSR-88D and CASA Radar Data

Nathan Snook[1,2], Ming Xue[1,2], and Youngsun Jung[2]

School of Meteorology[1] and Center for Analysis and Prediction of Storms[2]
University of Oklahoma, Norman OK 73072

Corresponding author address:
Nathan Snook
Center for Analysis and Prediction of Storms
University of Oklahoma,
120 David Boren Blvd, Room 5240, Norman OK 73072
nsnook@ou.edu

**Abstract**

This study examines the ability of a storm-scale numerical weather prediction (NWP) model to predict precipitation and mesovortices within a tornadic mesoscale convective system that occurred over Oklahoma on 8-9 May 2007, when the model is initialized from ensemble Kalman filter (EnKF) analyses including data from 4 CASA X-band and 5 WSR-88D S-band radars. Ensemble forecasts are performed and probabilistic forecast products generated, focusing on prediction of radar reflectivity (a proxy of quantitative precipitation) and mesovortices (an indication of tornado potential).

The ensemble assimilating data from both CASA and WSR-88D radars and using a mixed-microphysics ensemble during data assimilation produces the best probabilistic meso-vortex forecast. The use of multiple microphysics schemes within the ensemble aims to address at least partially the model physics uncertainty and effectively plays a role of flow-dependent inflation (in precipitation regions) during EnKF data assimilation. The ensemble predicts with high probability (approximately 0.65) the near-surface meso-vortex associated with the first of three reported tornadoes. Though a bias toward stronger precipitation is noted in the ensemble forecasts, all experiments produce skillful probabilistic forecasts of radar reflectivity on a 0-3 hour timescale as evaluated by multiple probabilistic verification metrics. These results suggest that both inclusion of CASA radar data and use of a mixed-microphysics ensemble during EnKF data assimilation positively impact the skill of 2-3 hour ensemble forecasts of mesovortices, despite having little impact on the quality of precipitation forecasts (analyzed in terms of predicted radar reflectivity), and are an important step toward an operational EnKF-based ensemble analysis and probabilistic forecast system to support convective-scale warn-on-forecast operations.

## 1. Introduction

Since the inception of explicit numerical weather prediction (NWP) of severe convective storms (Lilly 1990), assimilation of Doppler weather radar data has been shown to be critical and often effective for initializing such model predictions (e.g., Sun et al. 1991; Sun and Crook 1998; Xue et al. 2003; Hu et al. 2006). Recent studies have produced promising results assimilating Doppler radar data for convective-scale NWP in real-time and over large domains (e.g., Xue et al. 2008). The ensemble Kalman filter (EnKF) technique, initially developed by Evensen (1994, 2003), has been gaining popularity as an effective method of radar data assimilation for storm-scale NWP (e.g., Snyder and Zhang 2003; Dowell et al. 2004; Tong and Xue 2005b; Tong and Xue 2008). Though EnKF is more computationally expensive than the 3-dimensional variational method (3DVAR) widely used operationally, it provides flow-dependent multivariate background error covariances and cross-covariances that 3DVAR cannot. Such cross-covariances are essential for radar data assimilation, because most state variables are not directly observed (Tong and Xue 2005a, 2008). Additional comments on the relative merits of various radar data assimilation methods including 3DVAR, 4DVAR (4-D variational) and EnKF for convective storm analysis can be found in Tong and Xue (2005a). As available computational power increases, it will soon become feasible to run a real-time convective-scale forecast system which assimilates data via EnKF (Zhang et al. 2009) and produces convective-scale ensemble forecasts (e.g., Xue et al. 2008).

One of the major goals of the National Weather Service (NWS) in the coming decade is to develop a warn-on-forecast paradigm for convective-scale severe weather warnings (e.g., tornado, severe thunderstorm and flash flood warnings) in order to increase warning lead-time beyond what is possible with nowcasting techniques alone. Such an increase in warning lead-

1

time would allow entities such as hospitals and stadiums sufficient time to respond in the event of a warning (Stensrud et al. 2009). To achieve this goal, reliable short-term ($0-3$ hour) forecasts at the convective scale will be vital.

Due to the chaotic nature of the atmosphere, and inevitable errors in observations and NWP models, weather forecasts always contain uncertainty. No forecast is therefore complete without a description of its uncertainty (NRC 2006), which is often expressed in terms of forecast probability. Ensemble forecasting offers a practical way to provide a probabilistic forecast (Leith 1974). Global and regional ensemble forecasting has been operational for nearly two decades (e.g., Toth and Kalnay 1993; Houtekamer et al. 1996; Du et al. 2003; Bowler and Mylne 2009); by comparison, convective-scale ensemble forecasting is still in its infancy (Kong et al. 2006; Xue et al. 2011), though a few recent papers have sought to address the topic. Wandashin et al. (2010) use an idealized 2-dimensional model to investigate the predictability of ensemble forecasts of mesoscale convective systems, while Aksoy et al. (2010) used the WRF model to assimilate Doppler radar observations and perform idealized ensemble predictions of storms with both supercellular and linear convective modes. Stensrud and Gao (2010) assimilate radar data using 3DVAR and perform short-range ensemble forecasts of a supercell case. All three of these recent studies demonstrate the value of using ensembles for short-range convective NWP. Convective scale weather poses a greater prediction challenge due to its intermittent nature, smaller spatial and temporal scale, higher nonlinearity, and often due to incomplete observation coverage; these challenges increase the forecast uncertainty, making probabilistic forecasting even more crucial (Stensrud et al. 2009; Xue et al. 2011).

While it is possible to obtain probabilistic forecast products from a single deterministic forecast by examining the occurrence of an event at the surrounding grid points (Theis et al.

2

2005; Schwartz et al. 2009), the capability of such a method is limited by inherent biases and errors in the deterministic forecast used. One such source of error is microphysical parameterization; Snook and Xue (2008) and Dawson et al. (2009) found that the choice of microphysical scheme (and the parameter settings therein) strongly impact the mode and intensity of convection predicted. Furthermore, such probabilistic forecast products are inadequate in addressing uncertainty in the initial condition and dynamic error growth. Convective-scale errors generally grow very quickly (Lorenz 1969), greatly limiting the utility of a single deterministic forecast.

EnKF provides a set of analyses that, in principle, best characterize the analysis uncertainty, making them desirable initial conditions for ensemble forecasts. At the global scale, ensembles using EnKF analysis initial conditions have shown superior probabilistic forecasting performance compared to those using more traditional perturbation methods (Houtekamer et al. 2005; Hamill et al. 2011). EnKF methods have proven effective in generating dynamically consistent wind, temperature, and microphysical fields for convective storms when assimilating Doppler radar reflectivity and radial velocity data (e.g., Dowell et al. 2004; Tong 2006; Snook et al. 2011) but probabilistic forecasts at the convective scale using EnKF analyses has so far received limited attention. Zhang et al. (2010) is an example where convection-permitting-resolution ensemble forecasts of a tropical cyclone were initialized from global EnKF analyses.

This study details the results of ensemble forecasts produced from the EnKF analyses of the 8-9 May 2007 tornadic mesoscale convective system presented in Snook et al. (2011, SXJ11 hereafter), with the goal of evaluating the suitability of EnKF analyses of radar data for initializing an ensemble for the short-term convective-scale probabilistic forecast goals of "warn-on-forecast" (Stensrud et al. 2009). We use a neighborhood ensemble probability (NEP)

3

approach (Schwartz et al. 2010) to obtain probabilistic forecasts of radar reflectivity, and an object-based ensemble approach to obtain probabilistic forecasts of near-surface meso-vortices. The value of assimilating X-band radar data by the Engineering Research Center for Collaborative Adaptive Sensing of the Atmosphere (CASA) (McLaughlin et al. 2009) for improving the forecasts is evaluated, and the impact of microphysical parameterization during analysis and forecast periods is examined.

The remainder of this paper is organized as follows: Section 2 discusses the methodology used in the ensemble forecasts, as well as techniques used to generate probabilistic forecast products. Section 3 examines the results of ensemble and probabilistic forecasts in terms of simulated radar reflectivity and low-level mesovortices and verifies them against observations. Finally, section 4 contains a summary of analysis and forecast results, as well as concluding remarks.

## 2. Forecast Setup and Verification Methodology

Ensemble forecasts have become an indispensable part of convective-scale NWP in operational and research settings, providing valuable information about forecast sensitivity and uncertainty, as well as probabilistic forecast guidance (NRC 2006). In this study, 3-hour storm-scale ensemble forecasts are examined for the tornadic mesoscale convective system (MCS) that occurred over southwestern and central Oklahoma on 8-9 May 2007; this was the first tornadic event observed by the then newly-deployed CASA IP-1 (Integrated Project No. 1) radar network (McLaughlin et al. 2009; Brotzge 2010) . During this event, a pronounced line-end vortex (LEV) developed within the MCS and moved through much of southwest/central Oklahoma, passing directly over the CASA IP-1 network. The MCS spawned two confirmed EF-1 tornadoes and one

4

confirmed EF-0 tornado in central Oklahoma between 0354 UTC and 0443 UTC, just north of the CASA radar network. Additional details regarding the 8-9 May 2007 MCS can be found in SXJ11.

To quantify the impact of the new CASA radar data in storm-scale NWP, SXJ11 assimilated data from 4 CASA IP-1 radars and 5 WSR-88D radars using the ARPS ensemble Kalman filter (EnKF) system configured with full model physics and real terrain. The ensemble contained 40 members, and radar data are assimilated every 5 minutes for 1 hour. While the assimilation of WSR-88D data alone produced a reasonably accurate analysis of the convective system, assimilating CASA data in addition to WSR-88D data was found to improve the resulting analysis. Use of a mixed-microphysics ensemble was found to alleviate under-dispersion by increasing the ensemble spread.

In this study, the EnKF ensemble analyses produced by SXJ11, valid at 0200 of 9 May 2007, are used to initialize 3-hour ensemble forecasts. The model setup and naming convention used for these forecast experiments follow those in SXJ11. The computational domain has 256 × 256 × 40 grid points with a 2 km horizontal grid spacing and stretched vertical grid spacing (see Fig. 1). The analyses used as initial conditions were produced by assimilating radar data at 5-minute intervals for 1 hour, and provide an initial model state that closely matches the observations (SXJ11); the ongoing MCS is represented relatively well, as shown in comparison of the mean analyses of composite radar reflectivity in the ensemble to that observed by the WSR-88D network (Fig. 2). Results from three ensemble forecast experiments are presented here. The control experiment (hereafter "CNTL") assimilates both CASA and WSR-88D radar data and contains 40 ARPS ensemble members; 16 of these use the Lin ice microphysical scheme (Lin et al. 1983), 16 use the WRF single-moment 6-class (WSM6) ice microphysics

scheme (Hong and Lim 2006), and the remaining 8 members use the NWP explicit micropyhsics (NEM) scheme developed by Schultz (1995). The second experiment, NoMMP, assimilates the same data as CNTL but uses the Lin microphysics scheme in all 40 of its ensemble members. The third experiment, NoCASA, uses the same ensemble setup as CNTL but does not assimilate CASA data.

To isolate the impact of using a mixed-microphysics ensemble during the forecast period, two more ensemble forecast experiments are run in addition to the three mentioned above. Experiments CNTL_LIN and NoCASA_LIN are initialized from the CNTL and NoCASA initial conditions (respectively), but use a single-microphysics forecast ensemble consisting of 40 ARPS members using the Lin microphysics scheme, as in NoMMP. Though the number of microphysical species is the same in the Lin, WSM-6, and NEM microphysical schemes, the WSM-6 scheme uses graupel whereas the Lin and NEM schemes use hail. To solve this inconsistency, hail and graupel are treated as interchangeable when initializing a Lin forecast ensemble member from a WSM-6 member analysis in CNTL_LIN and NoCASA_LIN. A summary of all forecast experiments included in this study is provided in Table 1.

In all experiments, a reduced rain intercept parameter of $8 \times 10^5$ was used, consistent with Snook and Xue (2008), who found that reducing the rain intercept parameter yielded more realistic cold-pool structure. A timeline for these experiments is presented in Fig. 3. Lateral boundary conditions for all ensemble members are obtained from the NCEP NAM 6-hourly analyses and intervening 3 hour forecasts. Because of the relatively short forecast period and the use of a relatively large domain, the absence of boundary condition perturbations did not negatively impact ensemble data assimilation and prediction in the main regions of interest.

Forecast verification is performed for radar reflectivity at 0300, 0400, and 0500 UTC, and

for low-level mesovortices at 0400, 0420, and 0440 UTC; the latter times correspond closely to tornado reports received during this event at 0354, 0426, and 0443 UTC (see section 2 of SXJ11). Given the 2-km horizontal grid spacing used, tornado-scale circulations cannot be resolved, we therefore focus on prediction of resolvable low-level circulations linked to the observed tornadoes, rather than on the tornadoes themselves. With the 2-km horizontal grid spacing used, the mesovortices (Trapp and Weisman 2003) which were present in this case can be resolved.

Because convective cells are highly localized, even small displacement errors in a storm-scale forecast can result in very low objective skill scores when verified on a point-by-point basis, even though the forecast being scored may still be quite valuable to researchers and operational meteorologists (Baldwin et al. 2001; Schwartz et al. 2009). To alleviate this problem, Schwartz et al. (2009) applied a probabilistic "neighborhood" approach following Roberts and Lean (2008), where the probability of an event (e.g., radar reflectivity > 40 dBZ; rainfall rate > 2.0 cm/hr; hail mixing ratio > 0) at a grid point is determined by conditions at all grid points within a given radius of influence $r$ from that grid point; this collection of points comprises the neighborhood for the selected grid point. In this study, as was done in Schwartz et al. (2010), the neighborhood is extended to include all grid points on the same model level within the radius of influence from every ensemble member. Thus, drawing from Schwartz et al. (2009; 2010), the forecast probability ($P_i$) of an event at the $i^{th}$ grid point of the ensemble forecast is defined by:

$$P_i = \frac{1}{N}\sum_{j=1}^{N} B_j \qquad (1)$$

where $N$ is the unique collection of all points comprising the neighborhood for point $i$, and $B_j$ is

7

the binary probability at the $j^{th}$ grid point of the neighborhood, defined to be 1 if the event was observed at that grid point, and 0 if it was not. Schwartz et al. (2010) call $P_i$ neighborhood ensemble probability (NEP); further discussion can be found in that paper.

Mesovortices and other circulations cannot easily be treated as point variables, and are therefore not well-suited to a NEP method. Thus, for prediction of mesovortices, we instead apply an object-based ensemble verification approach. Significant low-level circulations are manually identified for each ensemble member in forecasts valid at 0400, 0420, and 0440 UTC. For a feature to count as a significant circulation, three criteria must be met: (1) vertical vorticity must exceed 0.02 s$^{-1}$ at 2 km above ground level; (2) discernable rotation must be present in the horizontal wind field at this level, and (3) the feature must be located within a convective region with radar reflectivity exceeding 30 dBZ. These criteria were chosen to discriminate true mesovortices from other phenomena, such as shear zones occurring along outflow boundaries. The binary probability $B_{im}$ for the $i^{th}$ model grid volume of the $m^{th}$ ensemble member is calculated such that $B_{im} = 1$ if the center of that grid volume lies within a predefined distance (in this study, 25 km) of a circulation center identified in the forecast for that ensemble member, and $B_{im} = 0$ otherwise. A purely ensemble-based probability results,

$$P_i = \frac{1}{N_{ens}} \sum_{m=1}^{N_{ens}} B_{im} \qquad (2)$$

where $N_{ens}$ is the number of members in the ensemble. In essence, $P_i$ from equation (2) can be viewed as the predicted probability of a strong near-surface vortex being present within 25 km of a given point; this forecast methodology is analogous to that used operationally by the Storm Prediction Center in their Day 1 tornado, hail, and wind outlook products, which forecast the probability of an event occurring within a 25 mile radius of a given point (Edwards et al. 2002).

**3. Forecast Results and Analysis**

We begin by presenting the results of ensemble and probabilistic forecasts of radar reflectivity (*Z*) for CNTL, CNTL_LIN and NoMMP. Radar reflectivity, closely linked to precipitation, is a field of meteorological interest that can be directly verified against WSR-88D radar observations over the entire area of the convective system. Ensemble and probabilistic forecasts of low-level vortices are then analyzed; near-surface mesovortices were closely co-located with the observed tornadoes in this event, as demonstrated by the proximity of rotational signatures in the KTLX storm-relative radial velocity observations (Fig. 4) to the tornadoes reported in association with them (SXJ11, their Fig. 1). Both the impact of assimilating CASA data and the use of a mixed-microphysics ensemble during the analysis and forecast periods are considered in evaluating the performance of the ensemble mesovortex forecasts.

*a. Probability-matched ensemble mean forecasts of radar reflectivity*

In ensemble forecasting, particularly at the convective scale, averaging the individual ensemble members to produce ensemble mean fields for precipitation-related variables (such as radar reflectivity) often leads to a smoother distribution with increased geographic extent and a low bias in intensity. To counteract the low-bias tendency in the ensemble mean, it is often desirable to reassign the values of precipitation-related fields using values from the component ensemble members used to compute the mean; the result of this process (Ebert 2001) is the probability-matched (PM) mean. The PM mean is a useful tool for forecasters, producing a single "best estimate" deterministic forecast; such ensemble-mean products for precipitation fields often outperform most or all of the ensemble members used to produce them (Ebert 2001).

9

PM mean $Z$ is calculated at approximately 2 km above the surface for CNTL (Fig. 5a-c), CNTL_LIN (Fig. 5d-f), and NoMMP (Fig. 5g-i) at 1, 2, and 3 hours of forecast time, and compared with $Z$ as observed by the WSR-88D radar network and interpolated to the model grid at the corresponding times (Fig. 5j-l).

In all experiments, the dominant convective mode is predicted with reasonable success (Fig. 5); the forecast ensembles predict a large mesoscale convective system with a trailing convective line in a similar location to the corresponding features in the WSR-88D observations (Fig. 5j-l). There are noticeable errors in the PM mean forecasts, however; in particular, the leading convective line, extending southeast from the region of the LEV in the observations (Fig. 5j-l) is not captured in the forecast ensemble. In addition, the southern portion of the trailing convective line decays too quickly in the ensemble forecast at 0400 UTC and beyond in all forecast experiments (Fig. 5b-c, f-g, h-i). Because of the prevailing south and south-southwesterly flow during this case, the prediction of the southern portion of the trailing convective line is likely to have been affected by the flow conditions at the southern boundary of the model domain at the later times; the role of data assimilation tends to be diminished near the inflow boundaries.

The Lin microphysics scheme, used in all 40 members of the NoMMP experiment during both the analysis and forecast periods, produces greater precipitation coverage than NEM and WSM schemes for this case (not shown). In addition, the reduced spread among NoMMP ensemble members as compared to the CNTL ensemble (SXJ11), results in much less variation in position (and thus higher ensemble mean values) in the trailing line, resulting in a more intense trailing line in the PM mean (Fig. 5g-i). In contrast, CNTL used a mixed-microphysics ensemble containing members from the WSM-6 and NEM microphysical schemes, both of

which favored solutions with fewer regions of light precipitation. Combined with greater variation among ensemble members, the result is a PM mean field with a weaker trailing line and less extensive regions of lighter precipitation in the northern portion of the system in CNTL (Fig. 5c, f, i).


*b.  Probabilistic predictions of radar reflectivity*

Fast error growth and high levels of uncertainty from various sources make probabilistic forecast products potentially very valuable at the convective scale. Using a NEP method, as described in section 2, $P[Z > 25$ dBZ$]$ (Fig. 6) and $P[Z > 40$ dBZ$]$ (Fig. 7) are calculated at 0300, 0400, and 0500 UTC for CNTL, CNTL_LIN, and NoMMP, and are compared with the corresponding reflectivity contour observed by the WSR-88D network.  A neighborhood radius of 5 km was used, resulting in a neighborhood consisting of 21 points in each ensemble member and an ensemble-wide neighborhood consisting of 840 forecast values across 40 members.

The observed 25 dBZ threshold (the thick contour in Fig. 6) encompasses large areas, including the entire region surrounding the LEV circulation and much of the convective line extending to the south and southwest; areas with a high NEP of reflectivity exceeding 25 dBZ closely match regions where observed reflectivity exceeds 25 dBZ over the northern portion of the system throughout the forecast period (Fig. 6), though the forecast precipitation region does not extend quite as far north and east as in observations at 0400 and 0500 UTC.  The presence of the trailing convective line to the south and southwest is strongly indicated in the probabilistic forecasts, particularly in the NoMMP ensemble (Fig. 6g-i), though the decay of the southern portion of the line discussed in section 3a is evident in all experiments.

Particularly at the 25 dBZ threshold, the spread among ensemble members in the single-

microphysics forecast ensemble of NoMMP (Fig. 6g-i) is considerably less than that in the mixed-microphysics ensemble of CNTL (Fig. 6a-c), and also somewhat less than in CNTL_LIN (Fig. 6d-f) which uses single Lin microphysics in the ensemble forecast but starting from the mixed-microphysics ensemble analyses, as suggested by the probability (also by direct examination of the spread fields, not shown). Both CNTL (Fig. 6a-c) and CNTL_LIN (Fig. 6d-f) exhibit greater variation among ensemble members; the difference is most prominent in the trailing convective line. Almost all members of NoMMP agree on a wide area of precipitation in excess of 25 dBZ in the trailing convective line at 0400 and 0500 UTC (Fig. 6h, i), giving rise to high probability there. In CNTL (Fig. 6b, c) many members predict the precipitation in the trailing line, but there is much greater variation in the location and extent of that precipitation than in NoMMP. Variation among ensemble members in CNTL_LIN (**Error! Reference source not found.**e, f) is greater than NoMMP, but far less than in CNTL. The greater coverage of light precipitation in members using the Lin scheme and the reduced variation among ensemble members as a result of the use of a single-microphysics ensemble result in a large region of very high probability of $Z > 25$ dBZ associated with the trailing line in NoMMP, particularly at 0400 and 0500 UTC (Fig. 6h, i). In general, use of a mixed-microphysics ensemble results in increased ensemble spread in a variety of model fields, particularly those closely related to precipitation processes, due to variation in hydrometeor type and distribution resulting from different treatment of microphysical processes. Increased spread alone does not necessarily improve the derived probabilistic forecast products, however, as we will consider later in this section through skill-score analyses. It does seem to help better account for model uncertainty with precipitation physics during the EnKF analysis.

The 40 dBZ threshold is exceeded only in small, localized areas in the WSR-88D

observations (Fig. 7), in individual convective elements within the core of the LEV and the leading convective line. Near the LEV (c.f. Fig. 5i-l) where observed radar reflectivity exceeds 40 dBZ, moderate to high NEP values are found (Fig. 7), although observed reflectivity exceeding 40 dBZ in the leading convective line is largely missed by the forecast ensembles. Particularly at later forecast times, all experiments strongly overestimate the geographic coverage of 40 dBZ echoes, due in part to spurious convection that develops near the CASA radar network (Fig. 7c, f, i). The causes of this spurious convection are considered in section 3c. The intensity of the trailing convective line extending southeast and south of the CASA domain into north-central Texas (Fig. 7) is also overestimated; moderate to high NEP values are present in areas where no reflectivity exceeding 40 dBZ was observed. The ensemble adequately forecasts the extent and location of regions of moderate precipitation within the system (Fig. 6), but a combination of position and intensity errors, and development of spurious convection limits forecast skill for more intense convective cores (Fig. 7).

To more quantitatively evaluate the skill of the NEP forecasts of radar reflectivity, we next examine their performance using the Relative Operating Characteristic (ROC) skill score (Mason and Graham 1999). The ROC Skill Score (RSS) is derived from the area under the ROC curve (Mason and Graham 1999), and is given by $RSS = 2A_{ROC} - 1$, where $A_{ROC}$ is the area under the ROC curve. The RSS has a maximum value of 1.0 for a perfect probabilistic forecast, and a minimum of -1.0, with scores at or below 0.0 indicating forecasts with no skill. The ROC measures hit and false alarm rates at varying probability thresholds (Mason and Graham 1999); thus RSS is a summary statistic providing information on the ability of the probabilistic forecast system to correctly discriminate between events and non-events. RSS is calculated for forecasts of reflectivity exceeding 25 dBZ at grid level 10, over the entire model domain (Fig. 8a), as well

13

as over the verification sub-domain (Fig. 8b) located within and downwind of the CASA radar network (see Fig. 1). In all experiments, the RSS is considerably higher over the verification sub-domain (Fig. 8b) than over the entire domain (Fig. 8a), because of error associated with the trailing convective line south and southwest of the CASA network.

All forecast ensembles show considerable skill in forecasting reflectivity exceeding 25 dBZ. The 1-hour ensemble forecasts (valid at 0300 UTC) show the greatest skill, with verification sub-domain RSSs ranging from 0.84 to 0.89. RSSs steadily decrease between 0300 and 0500 UTC (Fig. 8), with domain-wide values (Fig. 8a) remaining slightly lower than those over the verification sub-domain (Fig. 8b). The minimum RSS of 0.40, obtained for NoCASA_LIN on the full domain at 0500 UTC, corresponds to an area under the ROC curve of 0.70; a ROC area of 0.70 is often considered to be the lower bound for a skillful forecast (Buizza 1997; Kong et al. 2011), thus all experiments produced skillful 1, 2, and 3 hour NEP forecasts for regions of $Z$ exceeding 25 dBZ.

NoMMP has the highest domain-wide RSS (Fig. 8a) throughout the forecast period and the highest RSS within the verification sub-domain (Fig. 8b) at 0300 and 0400 UTC; at 0500 UTC, however, NoMMP actually has the lowest RSS in the sub-domain, while CNTL_LIN has the highest. For this case, ensemble members using Lin microphysics showed, on average, slightly better agreement with observed $Z$ than members using WSM6 or NEM microphysics. When ROC skill scores are calculated for three sub-ensembles within CNTL, grouped by the microphysics scheme used, the Lin sub-ensemble produces the highest RSS, while the NEM sub-ensemble produces the lowest (not shown). Similarly, within the verification sub-domain, RSSs decrease more slowly with time for CNTL_LIN and NoCASA_LIN than for their mixed-microphysics counterparts CNTL and NoCASA (Fig. 8b). The slightly improved RSS

14

performance in CNTL_LIN and NoCASA_LIN at later times appears to be in part due to slightly faster movement of the trailing convective line, in better agreement with observations. These results suggest a positive impact of using a mixed-microphysics ensemble during the assimilation period, but an improved forecast error trend when using the Lin scheme only during the subsequent ensemble forecast, at least for this case. The difference in the error trend is especially notable at longer forecast ranges when the impact of initial condition is reduced relative to the impact of forecast model. This is believed to be due to the relatively better performance of the Lin scheme, especially when compared to the simplified Schultz scheme.

When the ensemble forecasts are evaluated using the Brier score (Brier 1950) (not shown), another metric commonly used in probabilistic forecast evaluation, the conclusions drawn from verification using the RSS remain valid. Brier scores for CNTL_LIN and NoCASA_LIN deteriorate more slowly with time than those of their mixed-microphysics counterparts, CNTL and NoCASA, a result which matches the improved error trend of the single-microphysics forecast ensembles seen in verification using the RSS. Also, as with the RSS, NoMMP initially yields a good Brier score compared to the ensembles that used mixed-microphysics during data assimilation, but exhibits the fastest deterioration of Brier score during the forecast period (and, by 0500 UTC, the worst Brier scores of any experiment).

c. *Prominent biases and errors within the ensemble forecasts*

During the discussion of the NEP forecasts of radar reflectivity exceeding 40 dBZ (Fig. 7), we noted that the ensemble forecasts predict high probabilities of reflectivity exceeding 40 dBZ in regions where no 40 dBZ reflectivity values were observed, particularly at 0400 and 0500 UTC. To more closely examine biases in reflectivity within the ensemble forecasts, we

15

construct domain-wide histograms of radar reflectivity each hour from 0200 UTC to 0500 UTC in the ensemble mean and ensemble member forecasts of CNTL, and compare them to histograms constructed from radar reflectivity observed by the WSR-88D radar network and interpolated to the model grid. Bins are placed at intervals of 1 dBZ, and histograms constructed using the individual ensemble members are normalized by the size of the ensemble. The resulting histograms are shown in Fig. 9.

Two prominent biases are evident in the histograms of $Z$ for both the ensemble mean and the ensemble members: an under-prediction (low bias) of weak (<25 dBZ) precipitation regions, and an over-prediction (high bias) of areas of intense (>40 dBZ) precipitation (Fig. 9). At the initial time (0200 UTC), only the under-prediction of low $Z$ is present in the ensemble members and the ensemble mean; the histograms match observations well for $Z$ greater than approximately 25 dBZ (Fig. 9). By 0300 UTC both biases are present in both the ensemble mean and the ensemble members; the magnitude of these biases remains relatively constant from 0300 to 0500 UTC (Fig. 9f-h, j-l).

Both the low-bias in weak precipitation and the high bias in strong precipitation are smaller in magnitude for the ensemble mean than for the individual ensemble members at and after 0300 UTC (Fig. 9f-h, j-l). The primary source of these differences is smoothing and decreased intensity of $Z$ in the ensemble mean due to differences in the position of small-scale reflectivity features among individual ensemble members. While this smoothing leads to a distribution of $Z$ that agrees more closely with observations in this case, such smoothing is not universally desirable; for example, in the absence of the high bias in intense precipitation seen in the ensemble members in this case, the smoothing would degrade, not improve, the distribution of $Z$ in the ensemble mean.

16

In addition to the two biases identified above, a third bias is also evident in the histograms of Fig. 9: a low-bias in total histogram population (number of grid cell volumes having $\geq 10$ dBZ values) in the ensemble forecast compared to the observations. This bias is present in both the ensemble mean (Fig. 9e-h) and the individual ensemble members (Fig. 9i-l), but is more prominent in the ensemble members. Because a low-bias in total histogram population corresponds to a smaller spatial coverage, the low-bias is mitigated in the ensemble mean by the spatial smoothing inherent in the ensemble mean. This low bias results largely from the low bias of reflectivity of less than 30 dBZ in the ensemble, and appears to be a result of under-prediction of the extent of stratiform precipitation within the system (see Fig. 6).

The high bias in intense precipitation observed in the ensemble can be in part attributed to spurious convection that developed in and near the CASA radar network, as noted in section 3a. To provide additional insight into the source of this spurious convection, we calculate ensemble mean horizontal wind convergence fields at the first model level above ground (Fig. 10a, b) and show them together with potential temperature and horizontal winds at the same level (Fig. 10c, d), for CNTL at 0300 and 0400 UTC (Fig. 10). In the ensemble, the cold pool in the wake of the MCS (cf. Fig. 5) is relatively weak, with temperatures at and near the surface under the MCS only around 2 to 3 K cooler than the surrounding area (Fig. 10c, d). By comparison, many Oklahoma mesonet sites in southwestern Oklahoma recorded temperature drops of around 4 K during the passage of the MCS (not shown). In addition, particularly at 0300 UTC (Fig. 10a), the strongest convergence near the surface occurs behind the MCS, in and near the CASA domain (Fig. 10c), where strong near-surface southeasterly and easterly flows upwind of the CASA domain transitions to weaker, more directionally-varying flow within (and to the northwest of) the CASA domain (Fig. 10). The combination of low-level convergence and

17

relatively warm, buoyant air in the wake of the MCS seem to have contributed to the development of the spurious convection in this region in the model between 0300 and 0500 UTC.

*d.  Impact of microphysics on ensemble reflectivity forecasts*

To examine the impact of model microphysics during the ensemble forecasts, we calculate and compare RMS innovation of radar reflectivity ($Z$) for ensemble forecast experiments CNTL and CNTL_LIN.  As discussed in section 2, the CNTL and CNTL_LIN ensembles start from the same initial conditions, but CNTL uses a mixed-microphysics ensemble, while CNTL_LIN uses a single-microphysics ensemble consisting only of Lin microphysics members. Innovation is calculated using all grid volumes where the reflectivity observed by the WSR-88D radar network (interpolated to the model grid) exceeds 15 dBZ; this analysis technique is similar to that used in Aksoy et al. (2010).  Innovation values for each individual ensemble member, as well as the ensemble mean and probability-matched (PM) ensemble mean, are shown at 1, 2, and 3 hours of forecast time (0300, 0400, and 0500 UTC respectively) in Fig. 11.

Both CNTL and CNTL_LIN exhibit similar evolution of RMS innovation of $Z$ for the ensemble mean and PM mean forecasts.  The ensemble mean consistently outperforms the PM mean, as well as most individual ensemble members (Fig. 11).  This result is in agreement with Aksoy et al. (2010), whose ensemble mean RMS innovation of $Z$ was also lower than most of their ensemble members.  The relatively good performance of the ensemble mean can be attributed to the high bias in the reflectivity forecast (Fig. 9) discussed in section 3b, since position differences in the strongest reflectivity cores between individual members give an

18

ensemble mean $Z$ field that is smoother and contains lower reflectivity values than the individual members. The PM mean, though it contains the same spatial structure as the ensemble mean, has the same high bias as the individual ensemble members. Because the ensemble mean consistently outperforms the PM mean, we can conclude that it is the absence of the high bias, rather than improved spatial structure of the $Z$ field, that results in lower RMS innovation of $Z$ in the ensemble mean.

In the CNTL ensemble, the individual ensemble members exhibit a trimodal distribution, grouped into three clusters segregated by the microphysical scheme used in the model. The cluster of members using the Lin microphysical scheme has the lowest RMS innovation, followed by the cluster of members using WSM-6 microphysics, and the cluster using NEM microphysics has the highest RMS innovation values. The strong clustering in CNTL indicates that error associated with the microphysical parameterization is a dominant factor in model error in $Z$ during the ensemble forecast; in CNTL_LIN, where all members used the Lin microphysical scheme, the ensemble members have a unimodal distribution. Though the increased ensemble dispersion of CNTL is desirable, ensemble dispersion could also be increased using other methods, such as perturbed boundary conditions, perturbations within a microphysical scheme, and the use of different spread-maintenance techniques during the assimilation period; such methods will be a subject for future work. Because the microphysical parameterization appears to be a dominant source of error in the forecasts, using a more sophisticated 2- or 3-moment microphysical scheme might also be desirable to reduce overall RMS error in the ensemble; this too could be a subject for future study.

*e. Ensemble and probabilistic forecasts of mesovortices*

19

Forecast ensemble members of CNTL, NoMMP, and NoCASA were manually examined at 0400, 0420, and 0440 UTC to identify significant low-level vortices, as outlined in section 2. Such low-level vortices are considered objects for the purpose of verification. Equation (2) was then applied to generate forecasts of the probability of a significant (vertical vorticity $\zeta > 0.02$ s$^{-1}$) low-level vortex being present within 25 km of a point. The results of this analysis are presented in Fig. 12; the probabilities thus calculated are shown at 0400, 0420, and 0440 UTC for CNTL (Fig. 12a-c), NoMMP (Fig. 12d-f), and NoCASA (Fig. 12g-i), along with the locations of reported tornadoes (which were co-located with their parent mesovortices).

All three forecast ensembles predict near-surface vortices in the vicinity of the reported tornadic activity at 0400, 0420, and 0440 UTC (Fig. 12). At 0400 and 0420 UTC, the probability field of CNTL exhibits a tight concentration of the highest probability values close to the reported tornado location (Fig. 12a, b). At 0400 UTC, the location of the maximum near-surface vortex probability for CNTL of 0.65 (Fig. 12a) is located within 3 km of the reported tornado, and its probability field is tightly concentrated around the location of the observed tornado. In NoMMP and NoCASA (Fig. 12d, g) maximum probabilities are located approximately 10 and 20 km, respectively, from the reported tornado location. NoCASA and NoMMP forecast probabilities are only 0.43 and 0.35 (respectively) at the reported tornado location at 0400 UTC, and, especially in NoMMP (Fig. 12d), have relatively diffuse probability distributions with lower maxima. In NoMMP, this difference results from a wider spread of vortex location predictions seen within the single-microphysics ensemble. The only difference between NoCASA and CNTL, however, was the absence or presence of CASA radar data during the assimilation period, suggesting that assimilating CASA data positively impacts the prediction of low-level vortex features, yielding a higher maximum probability prediction at 0400 UTC in

20

the vicinity of the observed tornado. This finding is consistent with that of Schenkman et al (2011), where low-level radial velocity data from CASA radars were found to improve the LEV forecast.

The results at 0420 UTC (Fig. 12b, e, and h) are similar to those at 0400 UTC: CNTL (Fig. 12b) produces the highest maximum probability (between 0.4 and 0.5) and exhibits a probability field concentrated close to the reported tornado location. NoCASA (Fig. 12h) and NoMMP (Fig. 12e) predict lower probabilities (approximately 0.2 and 0.1 respectively) of a significant low-level vortex being present in the vicinity of the observed tornado.  The probability in the vicinity of the observed tornado is lower in all three experiments at 0420 UTC than at 0400 UTC.  At 0440 UTC (Fig. 12g-i), CNTL and NoCASA both feature probability fields with the highest predicted probabilities in close proximity to the observed tornado. Unlike at 0400 and 0420 UTC, the probability near the reported tornado location is actually slightly higher in NoCASA than in CNTL.  NoMMP performs poorly at 0440 UTC, (Fig. 12h), predicting generally low probability of significant vortices, with the highest probabilities located several tens of kilometers away from the reported tornado.  Because of the small size of the CASA domain and increasing impact of model factors (such as microphysical parameterization) as the forecast period proceeds, the greatest impact of CASA data from the initial condition of the ensemble would be expected at shorter forecast times, consistent with the results shown in Fig. 12.

One issue that arises when considering prediction of tornadic mesovortices is a question of the limits of predictability for these features.  Smaller features within atmospheric flows tend to exhibit error growth on shorter time scales (Lorenz 1969), and indeed, the tornadoes observed during this case formed and decayed with time scales of less than an hour.  The tornadic

21

mesovortices within the 8-9 May 2007 MCS, however, were strongly forced by larger-scale features within the MCS that spawned them. Because MCS's often exhibit predictability on a significantly longer time scale (Carbone et al. 2002), we believe that it is therefore reasonable to consider ensemble predictions of the tornadic mesovortices within the MCS at time scales of 2-3 hours because of this larger-scale forcing.

We further examine the role of microphysics in vortex prediction by plotting locations of low-level vortex centers in individual ensemble members at 0400 UTC, along with the ensemble mean vortex location and the location of the EF-1 Minco tornado (reported to have begun at 0354 UTC), for CNTL (Fig. 13a), NoMMP (Fig. 13b), and NoCASA (Fig. 13c). Each low-level vortex center is marked by a letter corresponding to the microphysical scheme used in the member that produced it ("L" for Lin, "W" for WSM-6, and "N" for NEM). CNTL (Fig. 13a) exhibits the closest clustering of predicted low-level vortices around the location of the observed tornado; 26 of 40 (65%) ensemble members predict a low-level vortex center within 25 km. These 26 members include 12 of the 16 Lin members and 11 of the 16 WSM-6 members, but only 3 of the 8 NEM members. In NoCASA (Fig. 13c) 17 of 40 (43%) members predict vortex centers within 25 km of the observed tornado location; these 17 members consist of 10 Lin, 5 WSM-6, and 2 NEM members. In CNTL and NoCASA, Lin ensemble members have the least error in vortex location, while NEM members have the most; many NEM members are outliers (see Fig. 13a, c), or fail to produce significant low-level vortices. In the NoMMP ensemble (Fig. 13b), predicted vortex locations vary substantially among members; only 14 members (35%) predict low-level vortices within 25 km of the observed tornado location, underscoring the positive impact of a mixed-microphysics forecast ensemble in mesovortex prediction.

Despite differences in the distribution of individual members, the forecast ensemble mean

low-level vortex locations (marked by "EM" in Fig. 13a-c) were similar in all three experiments. In each case the ensemble mean location was located slightly to the northwest of the observed tornado location, with the difference in position ranging from approximately 12 km in CNTL (Fig. 13a) to 18 km in NoCASA (Fig. 13c). This result underscores the importance of probabilistic information on forecast uncertainty; while CNTL produced a much sharper probabilistic forecast of the low-level vortex than NoMMP, ensemble mean vortex locations differed by only about 3 km, or 1.5 grid intervals (Fig. 13a, b) between the two experiments. Also, NoCASA forecasted a higher vortex probability with more individual members clustered around the true vortex location than NoMMP but its ensemble mean vortex location had a larger position error (Fig. 13b, c). Probabilistic information is important in assessing the reliability of these forecasts; an ensemble mean or single deterministic forecast cannot provide such information.

## 4. Discussion and Summary

In this paper, we perform 3-hour ensemble forecasts for the tornadic MCS of 8-9 May 2007 that occurred over northern Texas and southern and central Oklahoma, starting from ensemble Kalman filter analyses produced by assimilating CASA and WSR-88D radar data. We examine the impact of assimilating CASA (in addition to WSR-88D) radar data on the forecast, and the impact of the model microphysics during both the assimilation and forecast periods. The choice of microphysical scheme made a significant impact on forecast evolution for predictions of radar reflectivity and mesovortices. Ensemble members using the Lin or WSM-6 scheme performed considerably better in prediction of low-level mesovortex locations than members using the NEM microphysical scheme, while members using the Lin microphysical scheme were

found to perform best in terms of RMS innovation for prediction of radar reflectivity.

From the ensemble forecast experiments, probabilistic predictions for radar reflectivity and low-level circulations are obtained. A neighborhood ensemble probability approach (Schwartz et al. 2010) is applied to generate 60-180 minute probabilistic forecasts for radar reflectivity, and an object-oriented ensemble forecast approach is used to generate 120-160 minute probabilistic forecasts of low-level vortex location. Both the assimilation of CASA data and the use of a mixed-microphysics ensemble during the EnKF data assimilation cycles have a positive impact on forecasts of mesovortices.

All ensemble forecast experiments predict the dominant convective mode during the forecast period reasonably well, indicating a MCS with an embedded LEV, and a trailing convective line extending to the south, though the southern end of the convective line decays too quickly in the forecasts as compared to the observations. Two prominent biases are present in the ensemble member and ensemble mean forecasts of radar reflectivity: a low bias for light (< 25 dBZ) precipitation, and a high bias for intense (> 40 dBZ) precipitation. Microphysical parameterization is found to be a dominant factor in model error in reflectivity forecasts. When a single microphysics scheme is used, the innovation of radar reflectivity within the forecast ensemble exhibits a unimodal distribution, but when a mixed-microphysics scheme containing members of three different single-moment microphysics schemes is used, a trimodal distribution results, segregated by microphysical scheme.

The ensemble forecasts predict the location of tornadic mesovortices between 120 and 160 minutes of forecast time with varying degrees of success. The CNTL ensemble performs particularly well, predicting a localized region of high (maximum of 0.65) neighborhood ensemble probability of a significant low-level vortex in close proximity to the observed

mesovortex that spawned the 0354 UTC EF-1 Minco tornado. NoCASA and NoMMP predict comparatively diffuse probability fields with lower probability values.

The results of this study are encouraging with regard to development of a future warn-on-forecast severe weather warning system (Stensrud et al. 2009), demonstrating that, at least for the case studied, it is possible to provide useful probabilistic predictions of convective hazards, including areas of heavy precipitation and tornadic mesovortices, with lead times of 1 to 3 hours. Significant challenges do remain; the results of this study indicate that controlling model bias and maintaining appropriate ensemble spread will be particularly important in probabilistic predictions of this kind. In addition, the robustness of such probabilistic predictions still needs to be tested and calibrated over a large number of cases including a wide range of tornadic and non-tornadic convective storms. We also note that objective identification and verification of tornado-scale vortices is non-trivial (Potvin et al. 2009). At present, the high computational cost of performing such high-resolution predictions in real-time remains a barrier to operational implementation which will need to be addressed with algorithm and code optimizations and access to petascale computing systems. Further improvement in the prediction model and the inclusion of other available observations are also important; such work is ongoing with this and other cases. Despite the use of a single case in this study, we believe such careful case studies represent an essential first step towards full systematic testing and operational implementation.

**References**

Aksoy, A., D. C. Dowell, and C. Snyder, 2010: A multi-case comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part I: Short-range ensemble forecasts. *Monthly Weather Review*, In press.

Baldwin, M. E., S. Lakshmivarahan, and J. S. Kain, 2001: Verification of mesoscale features in NWP models. *Preprint, 9th Conf. Mesoscale Proc.*, Fort Lauderdale, FL, Amer. Meteor. Soc., 255-257.

Bowler, N. E. and K. R. Mylne, 2009: Ensemble transform Kalman filter perturbations for a regional ensemble prediction system. *Q. J. R. Meteorol. Soc.*, **135,** 757-766.

Brier, G. W., 1950: Verification of weather forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78,** 1-3.

Brotzge, J., K. Hondle, B. Philips, L. Lemon, E. Bass, D. Rude, and D. Andra, Jr., 2010: Evaluation of Distributed Collaborative Adaptive Sensing for detection of low-level circulations and implications for severe weather operations. *Wea. Forecasting*, **25,** 173-189.

Buizza, R., 1997: Potential Forecast Skill of Ensemble Prediction and Spread and Skill Distributions of the ECMWF Ensemble Prediction System. *Mon. Wea. Rev.*, **125,** 99-119.

Carbone, R. E., J. D. Tuttle, D. A. Ahijevych, and S. B. Trier, 2002: Inferences of Predictability Associated with Warm Season Precipitation Episodes. *J. Atmos. Sci.*, **59,** 2033-2056.

Dawson, D., 2009: Impacts of Single- and Multi-moment Microphysics on Numerical Simulations of Supercells and Tornadoes of the 3 May 1999 Oklahoma Tornado Outbreak, School of Meteorology, University of Oklahoma, 173.

Dowell, D. C., L. J. Wicker, and D. J. Stensrud, 2004: High-resolution analyses of the 8 May 2003 Oklahoma City storm. Part II: EnKF data assimilation and forecast experiments. *Preprints, 22nd Conf. on Severe Local Storms*, Hyannis, MA,, Amer. Meteor. Soc., CDROM, 12.5.

Du, J., G. DiMego, S. Tracton, and B. Zhou, 2003: NCEP short-range ensemble forecasting (SREF) sys-tem: Multi-IC, multi-model, and multi-physics approach. Research Activities in Atmospheric and Oceanic Modelling, J. Cote, Ed., CAS/JSC Working Group Numerical Experimentation (WGNE), Rep. 33, WMO Tech. Doc. 1161, 5.09-5.10.

Ebert, E. E., 2001: Ability of a poor man's ensemble to predict the probability and distribution of precipitation. *Mon. Wea. Rev.*, **129,** 2461-2480.

Edwards, R., S. F. Corfidi, R. L. Thompson, J. S. Evans, J. P. Craven, J. P. Racy, D. W. McCarthy, and M. D. Vescio, 2002: Storm Prediction Center Forecasting Issues Related to the 3 May 1999 Tornado Outbreak. *Weather and Forecasting*, **17,** 544-558.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99,** 10143-10162.

Evensen, G., 2003: The ensemble Kalman filter: Theoretical formulation and practical

implementation. *Ocean Dynamics*, **53,** 343-367.

Hamill, T. M., J. S. Whitaker, M. Fiorino, and S. G. Benjamin, 2011: Global ensemble predictions of 2009's tropical cyclones initialized with an ensemble Kalman filter. *Monthly Weather Review***,** In press.

Hong, S.-Y. and J.-O. J. Lim, 2006: The WRF single-moment 6-class microphysics scheme (WSM6). *J. Korean Meteor. Soc.*, **42,** 129-151.

Houtekamer, P. L., L. Lefaivre, J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124,** 1225-1242.

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. *Monthly Weather Review*, **133,** 604-620.

Hu, M., M. Xue, and K. Brewster, 2006: 3DVAR and cloud analysis with WSR-88D level-II data for the prediction of Fort Worth tornadic thunderstorms. Part I:  Cloud analysis and its impact. *Mon. Wea. Rev.*, **134,** 675-698.

Jung, Y., M. Xue, and M. Tong, 2011: Ensemble Kalman filter analyses of the 29-30 May 2004 Oklahoma tornadic thunderstorm using one- and two-moment bulk microphysics schemes. *Mon. Wea.   Rev.***,** Under review.

Kong, F., K. K. Droegemeier, and N. L. Hickmon, 2006: Multi-resolution ensemble forecasts of an observed tornadic thunderstorm system, Part I: Comparison of coarse and fine-grid experiments. *Mon. Wea. Rev.*, **134,** 807-833.

Kong, F., M. Xue, K. W. Thomas, Y. Wang, K. A. Brewster, X. Wang, J. Gao, S. J. Weiss, A. J. Clark, J. S. Kain, M. C. Coniglio, and J. Du, 2011: CAPS multi-model storm-scale ensemble forecast for the NOAA HWT 2010 Spring Experiment. *24th Conf. Wea. Forecasting/20th Conf. Num. Wea. Pred.*, Amer. Meteor. Soc., Paper 457.

Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon.   Wea. Rev.*, **102,** 409-418.

Lilly, D. K., 1990: Numerical prediction of thunderstorms - Has its time come? *Quart. J. Roy. Meteor. Soc.*, **116,** 779-798.

Lin, Y.-L., R. D. Farley, and H. D. Orville, 1983: Bulk parameterization of the snow field in a cloud model. *J. Climat. Appl. Meteor.*, **22,** 1065-1092.

Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21,** 289-307.

Mason, S. J. and N. E. Graham, 1999: Conditional probabilities, relative operating characteristics, and relative operating levels. *Wea. Forcasting*, **14,** 713-725.

McLaughlin, D., D. Pepyne, V. Chandrasekar, B. Philips, J. Kurose, M. Zink, K. Droegemeier, S. Cruz-Pol, F. Junyent, J. Brotzge, D. Westbrook, N. Bharadwaj, Y. Wang, E. Lyons, K. Hondl, Y. Liu, E. Knapp, M. Xue, A. Hopf, K. Kloesel, A. DeFonzo, P. Kollias, K. Brewster, R. Contreras, B. Dolan, T. Djaferis, E. Insanic, S. Frasier, and F. Carr, 2009: Short-wavelength technology and the potential for distributed networks of small radar systems. *Bull. Amer. Meteor. Soc.*, **90,** 1797-1817.

NRC, 2006: *Completing the Forecast: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts*. The National Academies Press, 112 pp.

Potvin, C. K., A. Shapiro, T.-Y. Yu, J. Gao, and M. Xue, 2009: Using a low-order model to detect and characterize tornadoes in multiple-Doppler radar data. *Mon. Wea. Rev.*, **137,** 1230-1249.

Roberts, N. M. and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev*, **136,** 78-97.

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

Schultz, P., 1995: An explicit cloud physics parameterization for operational numerical weather prediction. *Mon. Wea. Rev.*, **123,** 3331-3343.

Schwartz, C. S., J. S. Kain, D. R. Bright, S. J. Weiss, M. Xue, F. Kong, J. J. Levit, M. C. Coniglio, and M. S. Wandishin, 2009: Optimizing probabilistic high resolution ensemble guidance for hydrologic prediction. *Preprint, 23rd Conf. Hydrology*, Phoenix, AZ., Amer. Meteor. Soc., Paper 9.4.

Schwartz, C. S., J. S. Kain, S. J. Weiss, M. Xue, D. R. Bright, F. Kong, K. W. Thomas, J. J. Levit, M. C. Coniglio, and M. S. Wandishin, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Wea. Forecasting*, **25,** 263-280.

Snook, N. and M. Xue, 2008: Effects of microphysical drop size distribution on tornadogenesis in supercell thunderstorms. *Geophy. Res. Letters*, **35,** L24803, doi:10.1029/2008GL035866.

Snook, N., M. Xue, and J. Jung, 2011: Analysis of a tornadic meoscale convective vortex assimilating CASA X-band and WSR-88D radar data using an ensemble Kalman filter. *Mon. Wea. Rev.*, Conditionally accepted.

Snyder, C. and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **131,** 1663-1677.

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.*, Accepted.

Stensrud, D. J., M. Xue, L. J. Wicker, K. E. Kelleher, M. P. Foster, J. T. Schaefer, R. S. Schneider, S. G. Benjamin, S. S. Weygandt, J. T. Ferree, and J. P. Tuell, 2009: Convective-scale Warn on Forecast System:  A Vision for 2020. *Bull. Am. Meteor. Soc.*, **90,** 1487-1499.

Sun, J. and N. A. Crook, 1998: Dynamical and Microphysical Retrieval from Doppler Radar Observations Using a Cloud Model and Its Adjoint. Part II: Retrieval Experiments of an Observed Florida Convective Storm. *J. Atmos. Sci.*, **55,** 835-852.

Sun, J., D. W. Flicker, and D. K. Lilly, 1991: Recovery of three-dimensional wind and temperature fields from simulated single-Doppler radar data. *J. Atmos. Sci.*, **48,** 876-890.

Theis, S. E., A. Hense, and U. Damrath, 2005: Probabilistic precipitation forecasts from a deterministic model: A pragmatic approach. *Meteor. Appl.*, **12,** 257-268.

Tong, M., 2006: Ensemble Kalman filer assimilation of Doppler radar data for the initialization and prediction of convective storms, School of Meteorology, University of Oklahoma, 243.

Tong, M. and M. Xue, 2005a: Ensemble Kalman filter assimilation of Doppler radar data with a compressible nonhydrostatic model: OSS Experiments. *Mon. Wea. Rev.*, **133,** 1789-1807.

Tong, M. and M. Xue, 2005b: Simultaneous retrieval of microphysical parameters and atmospheric state variables with radar data and ensemble Kalman filter method. *Preprint, 17th Conf. Num. Wea. Pred.*, Washington DC, Amer. Meteor. Soc., CDROM P1.30.

Tong, M. and M. Xue, 2008: Simultaneous estimation of microphysical parameters and atmospheric state with radar data and ensemble square-root Kalman filter. Part I: Sensitivity analysis and parameter identifiability. *Mon. Wea. Rev.*, **136,** 1630–1648.

Toth, Z. and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74,** 2317-2330.

Trapp, R. J. and M. L. Weisman, 2003: Low-Level Mesovortices within Squall Lines and Bow Echoes. Part II: Their Genesis and Implications. *Monthly Weather Review*, **131,** 2804-2823.

Wandishin, M. S., D. J. Stensrud, S. L. Mullen, and L. J. Wicker, 2010: On the Predictability of Mesoscale Convective Systems: Three-Dimensional Simulations. *Monthly Weather Review*, **138,** 863-885.

Xue, M., D.-H. Wang, J.-D. Gao, K. Brewster, and K. K. Droegemeier, 2003: The Advanced Regional Prediction System (ARPS), storm-scale numerical weather prediction and data assimilation. *Meteor. Atmos. Physics*, **82,** 139-170.

Xue, M., F. Kong, K. W. Thomas, J. Gao, Y. Wang, K. Brewster, K. K. Droegemeier, J. Kain, S. Weiss, D. Bright, M. Coniglio, and J. Du, 2008: CAPS realtime storm-scale ensemble and high-resolution forecasts as part of the NOAA Hazardous Weather Testbed 2008 Spring Experiment. *24th Conf. Several Local Storms*, Savannah, GA, Ameri. Meteor. Soc., Paper 12.2.

Xue, M., F. Kong, K. W. Thomas, Y. Wang, K. Brewster, J. Gao, X. Wang, S. J. Weiss, A. J. Clark, J. S. Kain, M. C. Coniglio, J. Du, T. L. Jensen, and Y. H. Kuo, 2011: CAPS Realtime Storm Scale Ensemble and High Resolution Forecasts for the NOAA Hazardous Weather Testbed 2010 Spring Experiment. *24th Conf. Wea. Forecasting/20th Conf. Num. Wea. Pred.*, Amer. Meteor. Soc., Paper 9A.2.

Zhang, F., Y. Weng, J. A. Sippel, Z. Meng, and C. H. Bishop, 2009: Cloud-Resolving Hurricane Initialization and Prediction through Assimilation of Doppler Radar Observations with an Ensemble Kalman Filter. *Monthly Weather Review*, **137,** 2105-2125.

Zhang, F., Y. Weng, Y.-H. Kuo, J. S. Whitaker, and B. Xie, 2010: Predicting typhoon Morakot's catastrophic rainfall with a convection-permitting mesoscale ensemble system. *Weather and Forecasting*, **25,** 1816-1825.

**List of figures**

UTC, (e) 0400 UTC, and (f) 0500 UTC, and NoMMP at (g) 0300 UTC, (h) 0400 UTC, and (i) 0500 UTC.   The 25 dBZ radar reflectivity contours observed by the WSR-88D radars at the same time are in bold black contours.

Fig. 7. As Fig. 6, but for $P[Z > 40 \text{ dBZ}]$ and the 40 dBZ radar reflectivity contour.

Fig. 8. ROC skill score for 1-, 2-, and 3-hour forecasts of radar reflectivity at the 25 dBZ threshold on vertical grid level k = 10 (slightly more than 2 km above mean sea level) calculated over (a) the entire horizontal model domain and (b) the ETS verification subdomain as depicted by the black outline in Fig. 1.

Fig. 9. Domain-wide histograms of reflectivity intensity in ensemble forecast experiment CNTL during the 3-hour forecast period.   Shown are WSR-88D observations interpolated to the model grid (top row), the ensemble mean (center row), and individual ensemble members normalized by the size of the ensemble (bottom row).   The vertical axis indicates the number of model grid cell volumes (in thousands) containing reflectivity of a given intensity.

Fig. 10. Ensemble mean horizontal wind convergence ($s^{-1}$; shaded) at (a) 0300 and (b) 0400 UTC, as well as horizontal winds (barbs) and potential temperature (K; shaded) at (c) 0300 and (d) 0400 UTC in CNTL at the first model level above ground.

Fig. 11. RMS innovation of radar reflectivity during the forecast period for ensemble forecasts CNTL (left) and CNTL_LIN (right).   The ensemble mean forecast is indicated by the thick, solid black line, while the probability-matched ensemble mean forecast is indicated by the dashed black line.   The thin gray lines indicated innovation within individual ensemble members, with different shades of gray used for members using differing microphysical parameterization schemes.

Fig. 12. Ensemble-based forecast of P[ low-level circulation with vertical vorticity $\zeta > 0.02 \text{ s}^{-1}$

within 25 km of a point ] for CNTL at (a) 0300 UTC, (b) 0400 UTC, and (c) 0500 UTC;

NoMMP at (d) 0300 UTC, (e) 0400 UTC, and (f) 0500 UTC; and NoCASA at (g) 0300 UTC, (h)

0400 UTC, and (i) 0500 UTC.   The triangles indicate the reported locations of the tornadoes

reported at 0354 UTC (in panels (a), (d), and (g)), at 0426 UTC (in panels (b), (e), and (h)), and

at 0443 UTC (in panels (c), (f), and (i)).

Fig. 13. Locations of significant (vertical vorticity $\zeta > 0.02$ s$^{-1}$), discernible, low-level (grid level

10, approximately 2 km above the surface) mesovortex centers for all ensemble members at 0400

UTC in experiments (a) CNTL, (b) NoMMP, and (c) NoCASA.   Individual members are coded

by microphysical type; members using Lin microphysics   are marked "L", members using

WSM6 microphysics are marked "W", and members using NEM microphysics are marked "N".

The ensemble mean vortex location is marked "EM".   The location of the EF1 tornado reported

at 0354 UTC is marked by the black triangle.

Table 1.  Summary of forecast experiments.

| | CNTL | CNTL_LIN | NoCASA | NoCASA_LIN | NoMMP |
|---|---|---|---|---|---|
| WSR-88D Assimilated? | Yes | Yes | Yes | Yes | Yes |
| CASA Assimilated? | Yes | Yes | No | No | Yes |
| Mixed-Microphysics Ensemble? | Yes | During Data Assimilation Only | Yes | During Data Assimilation Only | No |
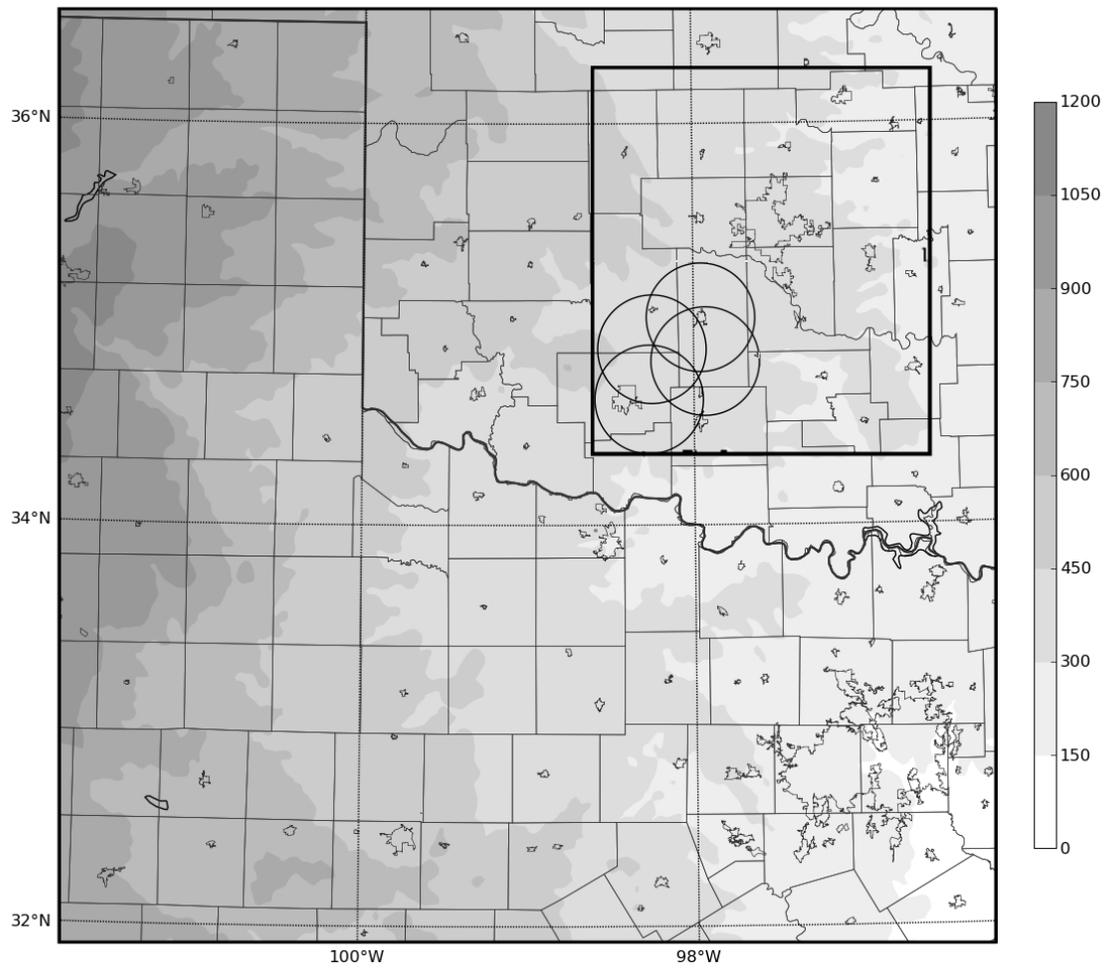
Fig. 1. Geographic extent of the model domain. Shading indicates surface elevation in meters above sea level. The forecast verification sub-domain used for skill score calculations in

Fig. 8b is indicated by the black box in the northeastern portion of the forecast domain.
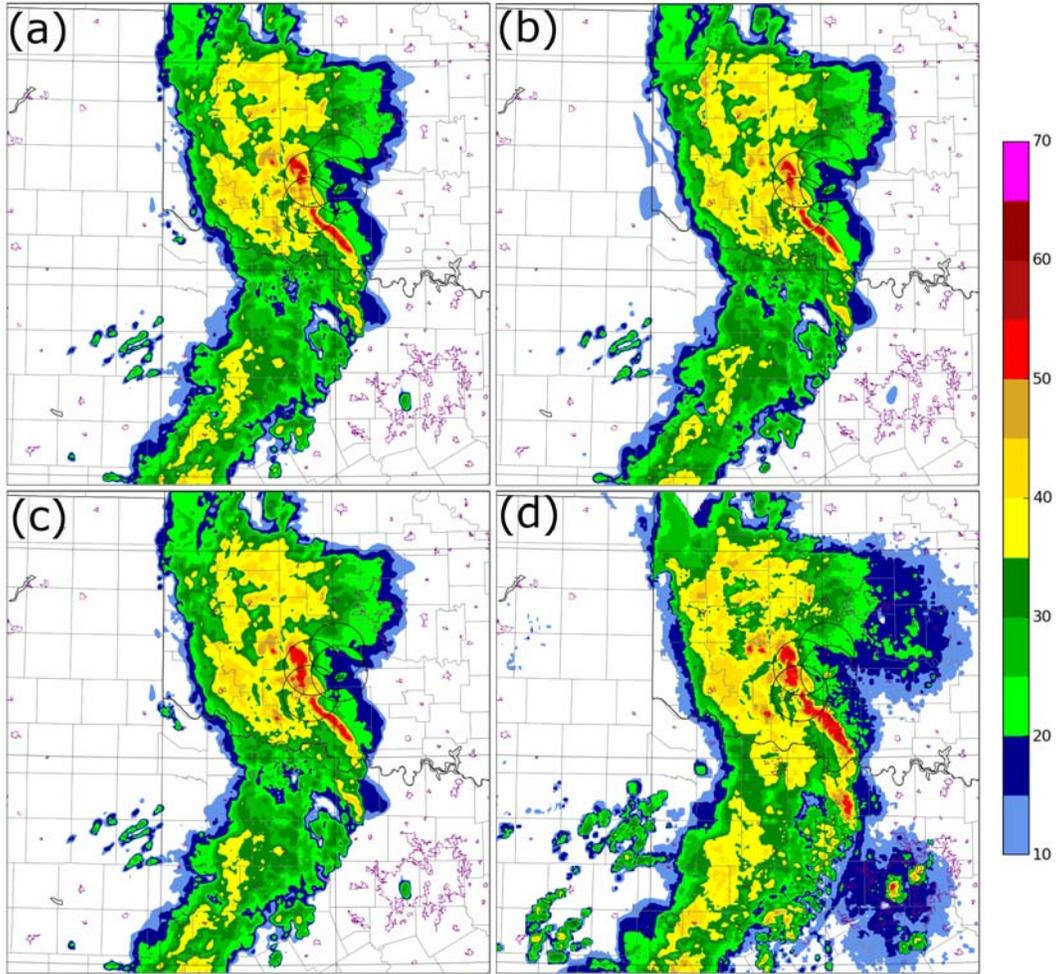
Fig. 2. Composite radar reflectivity (dBZ) of the ensemble mean analysis state from at 0200 UTC 9 May 2007 for (a) CNTL, (b) NoMMP, and (c) NoCASA, from which the deterministic forecasts in Fig. 4 were launched; also shown is (d) composite radar reflectivity (dBZ) observed by WSR-88D radars KAMA, KDYX, KFWS, KLBB, and KTLX at 0200 UTC 9 May 2007.   30 km CASA radar range rings are included for reference.   This figure is a reproduction of Fig. 4 of Snook et al. (2011).
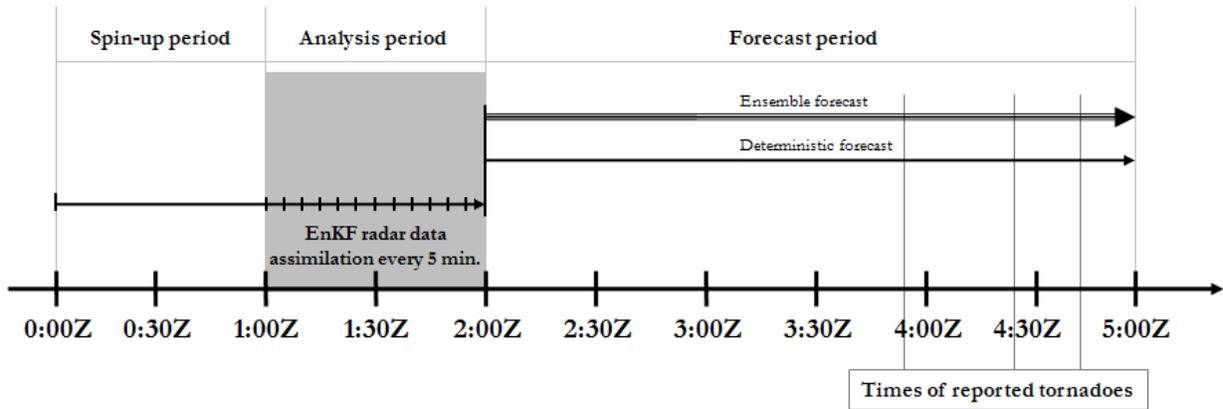
Fig. 3. Forecast diagram detailing the analysis and forecast periods for experiments CNTL, NoMMP, and NoCASA.
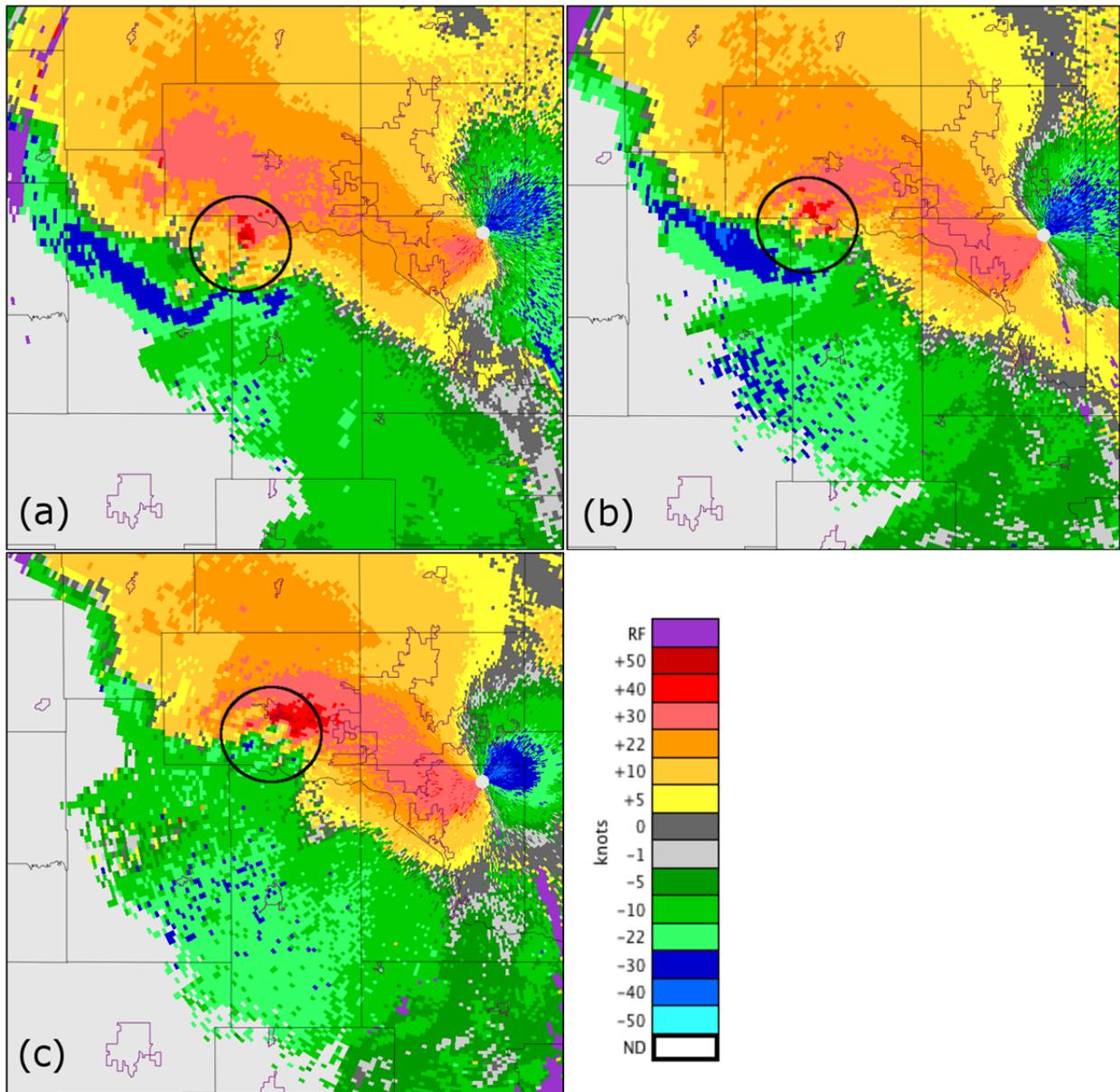
Fig. 4. Storm-relative radial velocity from observed level-III WSR-88D data for the 1.5 degree elevation scan of KTLX at (a) 0400 UTC, (b) 0420 UTC, and (c) 0440 UTC, 9 May 2007. Mesovortices associated with the reported tornadoes are circled. Urban and county boundaries are shown.
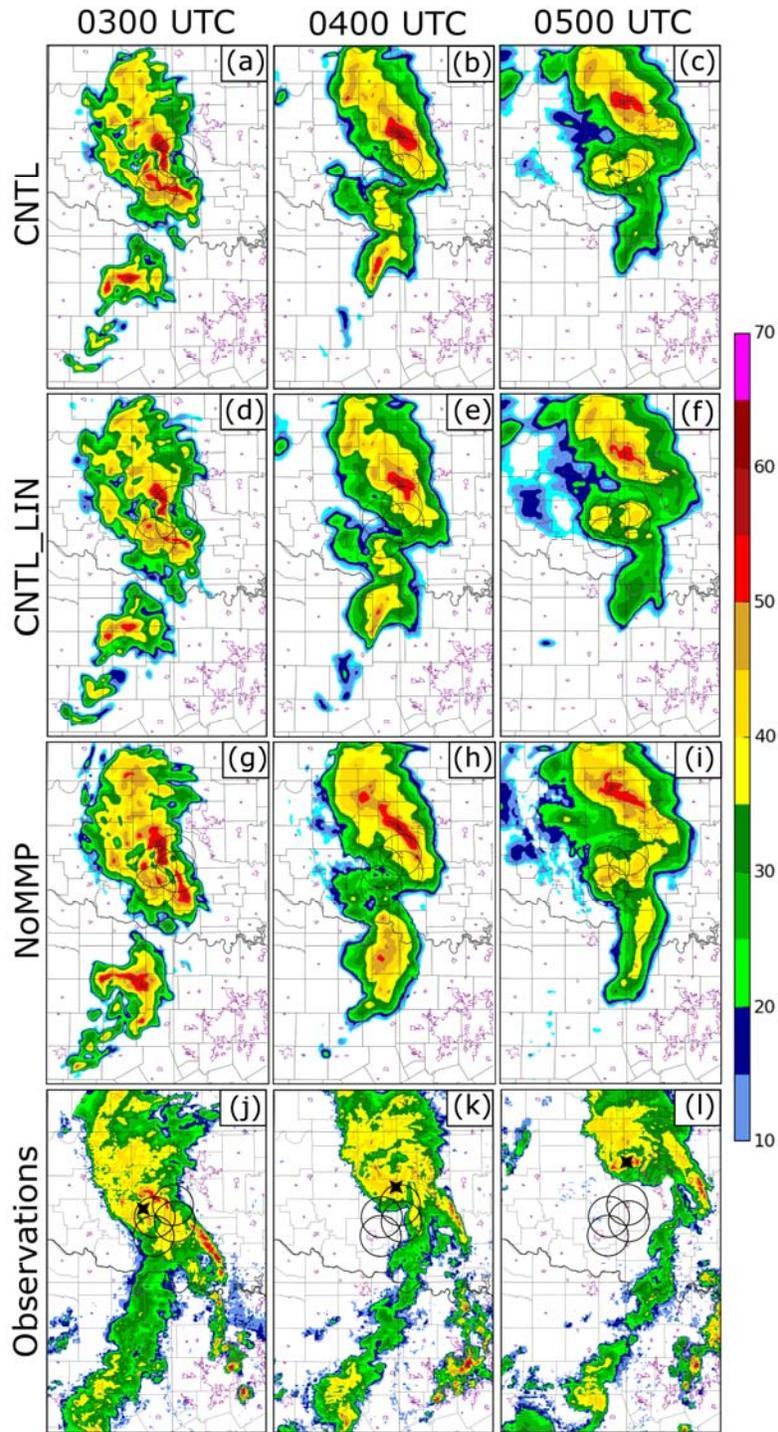
Fig. 5. Probability-matched ensemble mean forecasts of reflectivity (shaded) at model grid level 10 (about 2 km AGL) for (a-c) CNTL, (d-f) CNTL_LIN, and (g-i) NoMMP at 0300, 0400, and 0500 UTC. Also shown is (j-l) reflectivity observed by the WSR-88D network, interpolated to the same model grid level at 0300, 0400, and 0500 UTC. The center of the line-end vortex in the observations is indicated by the black marker in panels (j-l).
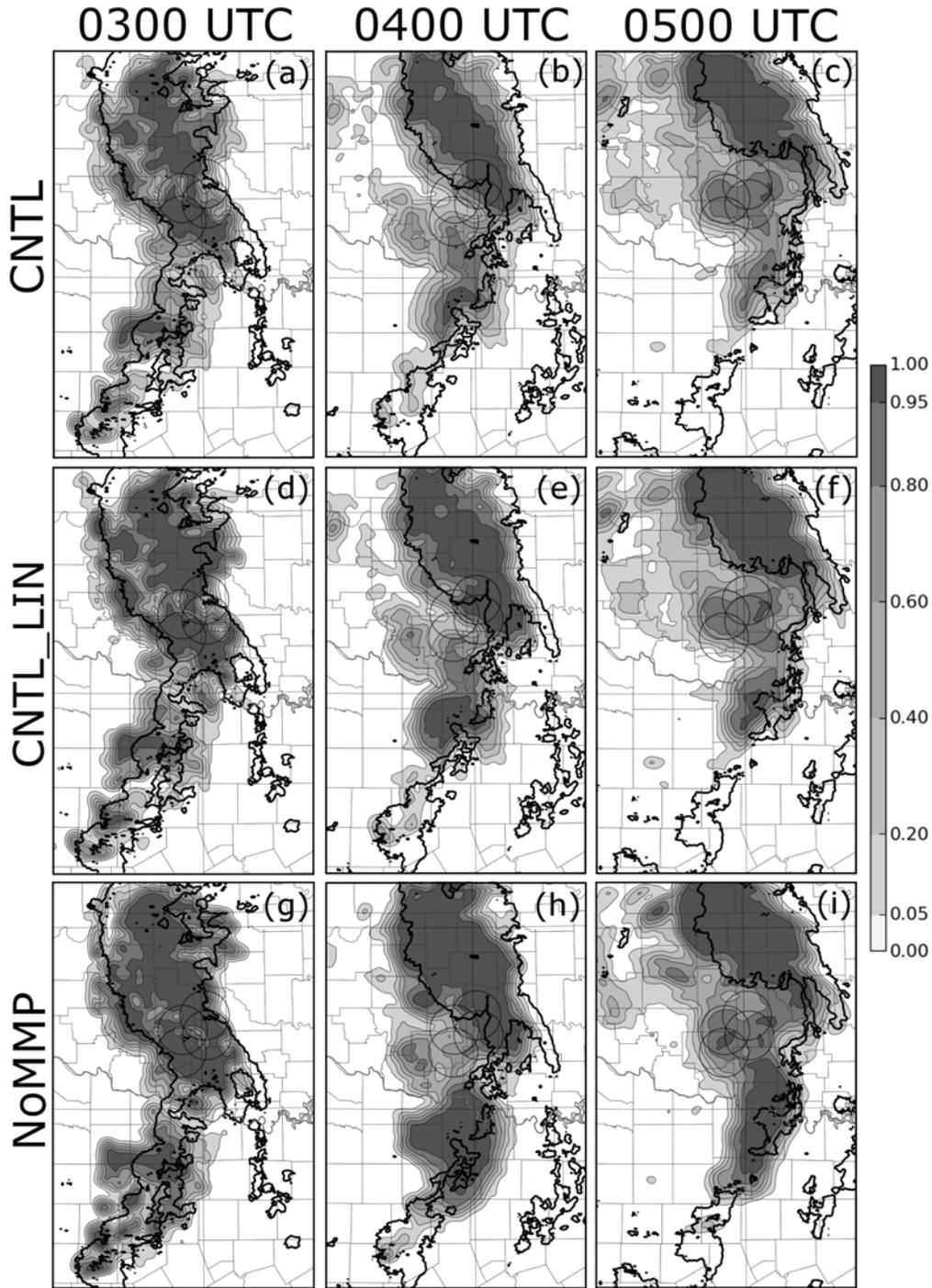
Fig. 6. Neighborhood ensemble probabilities (shaded) of forecast reflectivity exceeding 25 dBZ, P[ Z > 25 dBZ], at model grid level 10 (about 2 km AGL), in an area surrounding the CASA domain, for CNTL at (a) 0300 UTC, (b) 0400 UTC, and (c) 0500 UTC, CNTL_LIN at (d) 0300 UTC, (e) 0400 UTC, and (f) 0500 UTC, and NoMMP at (g) 0300 UTC, (h) 0400 UTC, and (i) 0500 UTC. The 25 dBZ radar reflectivity contours observed by the WSR-88D radars at the same time are in bold black contours.

41

Fig. 7. As Fig. 6, but for P[ Z > 40 dBZ ]  and the 40 dBZ radar reflectivity contour.

Fig. 8. ROC skill score for 1-, 2-, and 3-hour forecasts of radar reflectivity at the 25 dBZ threshold on vertical grid level k = 10 (slightly more than 2 km above mean sea level) calculated over (a) the entire horizontal model domain and (b) the ETS verification subdomain as depicted by the black outline in Fig. 1.

Fig. 9. Domain-wide histograms of reflectivity intensity in ensemble forecast experiment CNTL during the 3-hour forecast period. Shown are WSR-88D observations interpolated to the model grid (top row), the ensemble mean (center row), and individual ensemble members normalized by the size of the ensemble (bottom row). The vertical axis indicates the number of model grid cell volumes (in thousands) containing reflectivity of a given intensity.

Fig. 10. Ensemble mean horizontal wind convergence (s$^{-1}$; shaded) at (a) 0300 and (b) 0400 UTC, as well as horizontal winds (barbs) and potential temperature (K; shaded) at (c) 0300 and (d) 0400 UTC in CNTL at the first model level above ground.

Fig. 11. RMS innovation of radar reflectivity during the forecast period for ensemble forecasts CNTL (left) and CNTL_LIN (right). The ensemble mean forecast is indicated by the thick, solid black line, while the probability-matched ensemble mean forecast is indicated by the dashed black line. The thin gray lines indicated innovation within individual ensemble members, with different shades of gray used for members using differing microphysical parameterization schemes.

46

Fig. 12. Ensemble-based forecast of P[ low-level circulation with vertical vorticity $\zeta > 0.02$ s$^{-1}$ within 25 km of a point ] for CNTL at (a) 0300 UTC, (b) 0400 UTC, and (c) 0500 UTC; NoMMP at (d) 0300 UTC, (e) 0400 UTC, and (f) 0500 UTC; and NoCASA at (g) 0300 UTC, (h) 0400 UTC, and (i) 0500 UTC. The triangles indicate the reported locations of the tornadoes reported at 0354 UTC (in panels (a), (d), and (g)), at 0426 UTC (in panels (b), (e), and (h)), and at 0443 UTC (in panels (c), (f), and (i)).

Fig. 13. Locations of significant (vertical vorticity $\zeta > 0.02$ s$^{-1}$), discernible, low-level (grid level 10, approximately 2 km above the surface) mesovortex centers for all ensemble members at 0400 UTC in experiments (a) CNTL, (b) NoMMP, and (c) NoCASA.  Individual members are coded by microphysical type; members using Lin microphysics  are marked "L", members using WSM6 microphysics are marked "W", and members using NEM microphysics are marked "N". The ensemble mean vortex location is marked "EM".  The location of the EF1 tornado reported at 0354 UTC is marked by the black triangle.