Ensemble Probabilistic Prediction of a Mesoscale Convective System and Associated Polarimetric Radar Variables using Single-Moment and Double-Moment Microphysics Schemes and EnKF Radar Data Assimilation

Bryan J. Putnam[1, 2, 3], Ming Xue[1, 2], Youngsun Jung[1], Nathan A. Snook[1], and Guifu Zhang[2, 3]

[1] Center for Analysis and Predication of Storms, Norman, Oklahoma
[2] Advanced Radar Research Center, Norman, Oklahoma
[3] School of Meteorology, University of Oklahoma, Norman, Oklahoma 73072

April 2016

Revised October 9, 2016

Submitted to Monthly Weather Review

Corresponding author address:
Ming Xue
Center for Analysis and Prediction of Storms
University of Oklahoma,
120 David Boren Blvd., Norman, OK 73072
mxue@ou.edu

Abstract

Ensemble-based probabilistic forecasts are performed for a mesoscale convective system (MCS) that occurred over Oklahoma on 8-9 May 2007, initialized from ensemble Kalman filter analyses using multi-network radar data and different microphysics schemes. Two experiments are conducted, using either a single-moment or double-moment microphysics scheme during the one-hour long assimilation period and in subsequent three-hour ensemble forecasts. Qualitative and quantitative verifications are performed on the ensemble forecasts, including probabilistic skill scores. The predicted dual-polarization (dual-pol) radar variables and their probabilistic forecasts are also evaluated against available dual-pol radar observations, and discussed in relation to predicted microphysical states and structures.

Evaluation of predicted reflectivity (Z) fields shows that the double-moment ensemble predicts the precipitation coverage of the leading convective line and stratiform precipitation regions of the MCS with higher probabilities throughout the forecast period compared to the single-moment ensemble. In terms of the simulated differential reflectivity (Z_{DR}) and specific differential phase (K_{DP}) fields, the double-moment ensemble compares more realistically to the observations and better distinguishes the stratiform and convective precipitation regions. Z_{DR} from individual ensemble members indicates better raindrop size-sorting along the leading convective line in the double-moment ensemble. Various commonly used ensemble forecast verification methods are examined for the prediction of dual-pol variables. The results demonstrate the challenges associated with verifying predicted dual-pol fields that can vary significantly in value over small distances. Several microphysics biases are noted with the help of simulated dual-pol variables, such as substantial over-prediction K_{DP} values in single-moment ensembles.

1. Introduction

A major focus in recent convective scale numerical weather prediction (NWP) research has been improving both the forecast initial conditions and the microphysics parameterizations that are important for convective-scale predictions; both areas address major challenges identified for the Warn-on-Forecast paradigm by Stensrud et al. (2013). Data assimilation (DA), which is an indispensable part of convective-scale NWP, aims to improve the forecast initial condition by optimally combining available observations and a background model state to produce the best possible estimate of the atmospheric state. One popular DA method for convective-scale NWP is the ensemble Kalman filter (EnKF, Evensen 1994; 2003), which uses an ensemble of forecasts to estimate the background error covariance. The application of EnKF methods for the assimilation of radar observations has produced successful results for a variety of real storm cases (e.g., Dowell et al. 2004; Dowell and Wicker 2009; Lei et al. 2009; Aksoy et al. 2009; Aksoy et al. 2010; Dowell et al. 2011; Snook et al. 2011; Dawson et al. 2012; Jung et al. 2012; Snook et al. 2012; Yussouf et al. 2013; Tanamachi et al. 2013; Putnam et al. 2014, hereafter P14; Wheatley et al. 2014; Snook et al. 2015; Yussouf et al. 2015).

Additionally, microphysics parameterization (MP) schemes are used in convective-scale NWP models for the explicit prediction of fields describing the type and amount of hydrometeors present within the simulated storms. Due to computational expense, most MP schemes treat the hydrometeor particle size distributions (PSDs) in a bulk form, as opposed to representing the PSDs using a spectral bin model (e.g., Khain et al. 2004), and the three-parameter gamma distribution is often assumed:

$$N(D)_x = N_{0x} D_x^{\alpha_x} e^{(-\Lambda_x D)} \quad , \qquad (1)$$

where $N(D)_x$ is the number of particles of hydrometeor species x with diameter D in a unit volume, and Λ_x, N_{0x}, and α_x are the slope, intercept, and shape parameters, respectively (Ulbrich

1983; Milbrandt and Yau 2005a). MP schemes are often characterized by the number of PSD moments that are explicitly predicted and used to derive the same number of PSD parameters. Single-moment schemes (SM) usually predict the third moment of the distribution, the hydrometeor mixing ratio (q_x), while specifying the intercept and shape parameters; double-moment (DM) schemes also predict the zeroth moment, the total number concentration (N_{tx}), so that both the slope and intercept parameters can be updated; triple-moment (TM) schemes predict the additional sixth moment of the distribution, often called radar reflectivity factor (z), and effectively allow the slope, intercept, and shape parameters of the gamma distribution to vary independently. The shape parameter is specified as a constant or diagnosed value in DM schemes.

The use of DM scheme for EnKF-based convective-scale NWP has been shown to improve storm structure and evolution during the analysis cycles as well as forecasts for both supercell and mesoscale convective system (MCS) cases. Dawson et al. (2009) showed that DM and triple-moment (TM) schemes produced better predictions of a supercell storm than a SM scheme. Xue et al. (2010) first successfully applied EnKF to the estimation of model states associated with a DM scheme using simulated radar observations of a supercell, while Jung et al. (2012) first successfully used a DM scheme for EnKF radar DA for a real supercell storm. For the 8 May 2003 Moore, Oklahoma supercell, Yussouf et al. (2013) found that both a fully DM scheme (which predicts total number concentration for graupel, N_{tg}) as well as a semi-DM scheme (which diagnoses intercept parameter for graupel, N_{0g}) produced more small graupel than a SM scheme; this graupel was advected farther downwind, forming a broader forward flank downdraft (FFD), in agreement with observations. For MCS cases, P14, and subsequently Wheatley et al. (2014), found that DM MP schemes improved the development of trailing stratiform precipitation compared to a SM scheme. A dramatic increase in the formation and detrainment of snow and ice from the leading convective towers rearward over the stratiform region resulted in much broader stratiform coverage.

Recently, simulated dual-polarization (dual-pol) radar variables have been used to evaluate microphysical states estimated through data assimilation and predicted by convective scale models for real cases, by comparing these variables to observations (Jung et al. 2012; Li and Mecikalski 2012; Dawson et al. 2014; P14; Posselt et al. 2015; Putnam et al. 2016). The dual-pol variables contain additional information on PSDs over reflectivity (Z), specifically information about the size, content, and diversity of hydrometeors present in the radar volume. For example, differential reflectivity (Z_{DR}) values are dependent on the horizontal-to-vertical axis ratio of hydrometeors; values are higher for large, oblate raindrops and low for dry, tumbling hail (Bringi and Chandrasekar 2001). Additionally, specific differential phase (K_{DP}) is sensitive to the amount of liquid water the radar pulse interacts with.

Dynamical and microphysical processes can lead to significant variation in hydrometeor PSDs over small spatial scales. For example, the size-sorting of hydrometeors associated storm-relative wind shear in the forward flank of supercells leads to a significant increase in the number of large raindrops in low-level rain PSDs that can be identified by an increase Z_{DR} values known as the Z_{DR} arc (Kumjian and Ryzhkov 2008; Kumjian and Ryzhkov 2012; Dawson et al. 2014). This signature is indistinguishable in the observed Z pattern. Jung et al. (2012), in an EnKF data assimilation study of a supercell storm that occurred on 29 May 2004 in central Oklahoma, showed that using a DM MP scheme (Milbrandt and Yau 2005b) allowed the model to replicate observed dual-pol signatures such as the Z_{DR} arc. P14 found that simulated Z_{DR} patterns in the final EnKF analysis of an MCS produced using a DM scheme better represented the distribution of large, oblate raindrops in the leading convective line and small to medium sized raindrops in the trailing stratiform region compared to an analysis produced using a SM scheme. The SM analysis failed to capture this distinction, overestimating raindrop size in the stratiform region.

P14, which considered DM schemes and simulated dual-pol variables, focused on the final EnKF analyses of the experiments and on deterministic forecasts of simulated *Z*. P14 paid particular attention to the improvement in the microphysical and dynamical aspects of the MCS when using the DM scheme, such as the hydrometeor distributions and cold pool, and did not consider forecasts of dual-pol variables in depth. The current study expands upon P14 by performing and examining ensemble forecasts of the 8-9 May 2007 MCS case in terms of both *Z* and dual-pol radar variables. Ensemble forecasts offer additional benefits compared to deterministic forecasts, including the ability to produce probabilistic forecasts for precipitation events instead of a binary hit or miss forecast. Ensemble forecasts are integral to the Warn-on-Forecast vision outlined in Stensrud et al. (2009), providing the basis for operational probabilistic prediction of hazards associated with severe convection in the near future. Probabilistic forecasts help account for the uncertainties related to both the initial condition and the prediction model (including the microphysics), so as to provide a means of measuring the level of confidence in the prediction.

One of the advantages of EnKF methods is that they inherently provide an ensemble of analyses suitable for initializing an ensemble of forecasts (Kalnay 2002). Analyses from well-tuned EnKF systems provide a good representation of flow-dependent background error that properly characterizes the analysis uncertainty (Kalnay et al. 2006). EnKF-initialized ensemble forecasts have been used to produce convective-scale probabilistic forecasts in several recent studies. For tornadic storms, probabilistic forecasts have focused on the low-level vorticity; Dawson et al. (2012) and Yussouf et al. (2013; 2015; 2016) showed that the ensemble probability of vorticity exceeding certain thresholds predicted the observed damage paths of tornadoes well in supercell cases, while Snook et al. (2012; 2015) obtained similarly successful results for an MCS case. Snook et al. (2012; 2015) also examined and demonstrated the benefits of using multiple SM MP schemes in EnKF ensembles for probabilistic forecasts of *Z*, while Yussouf et al. (2016) showed the assimilation of radar data in a continuous-update-cycle EnKF DA system provides significant improvement during the first three hours of probabilistic quantitative precipitation forecasts.

In previous convective-scale EnKF studies using DM MP schemes, little attention has been given to probabilistic prediction of simulated radar variables or quantitative probabilistic forecast skill scores of simulated radar variables. In particular, probabilistic forecasting of simulated dual-pol variables has never been reported in the formal literature as far as we know. Although Snook et al (2012; 2015) examined probabilistic prediction of *Z*, the studies were limited to the use of SM MP schemes, and they did not examine any of the dual-pol variables either. Dawson et al. (2012), Yussouf et al. (2013), and Wheatley et al. (2014) conducted ensemble forecasts using DM MP schemes, but they only examined individual member or ensemble mean forecasts, not probabilistic forecasts of *Z*. The more recent studies of Yussouf et al. (2015; 2016) showed that probabilistic forecasts of *Z* exceeding 40 dBZ based on the semi-DM Thompson (Thompson et al. 2004; Thompson et al. 2008) scheme for two tornadoes cases matched the locations of observed supercells well. However, no quantitative probabilistic forecast skill scores for Z were presented. Additionally, these preceding studies did not directly compare simulated radar variables on the elevation levels where observed data were taken, but such comparisons are more intuitive for operational forecasting purposes and therefore should be performed first. Putnam et al. (2016) simulated dual-pol variables from the CAPS storm-scale ensemble forecasts for Hazardous Weather Testbed Spring Experiment (Kong 2013) for several members that differed only in the use of MP schemes. The study emphasized the differences among the different MP schemes in their ability to simulate dual-pol radar signatures, but ensemble probabilistic forecasting of dual-pol radar variables was not investigated.

In this study, we examine two ensemble forecasts of an MCS produced using either mixed SM MP schemes or a DM MP scheme during both the EnKF DA and subsequent forecasts. We evaluate the simulated dual-pol variables both qualitatively and quantitatively. Neighborhood probabilities are calculated for both Z and the dual-pol variables from ensemble forecasts with both perturbed initial conditions and microphysics perturbations, and the probabilistic forecasting performances of the two ensembles are compared. Probabilistic forecasts of the dual-pol variables include additional physical meaning beyond what Z can show, including the connection between K_{DP} and rainfall rate, and the uncertainty such forecasts may contain. As pointed out earlier, probabilistic forecasts of dual-pol radar variables have never been examined before.

The remainder of this paper is organized as follows: Section 2 reviews the 8-9 May 2007 MCS case and the experiment design, and briefly summarizes the methods used in the SM and DM ensemble forecasts. In section 3, we assess the skills of the ensemble probabilistic forecasts obtained with the SM and DM schemes. Finally, section 4 summarizes the findings. The challenges associated with probabilistic forecasting and evaluation of highly localized dual-pol signatures are also discussed and some suggestions for future research are given.

2. Experimental case and method

The model, EnKF settings, and data sources used in this study are all inherited from P14. Two experiments are conducted using a SM and DM MP scheme, respectively, in which ensemble forecasts are initialized from the final EnKF analyses for the 8-9 May 2008 MCS. The SM ensemble (EXP_S) and the DM ensemble (EXP_D) use the same configuration during the EnKF analysis period as the corresponding control experiments EXP_S_M_3_5/EXP_S and

EXP_D_M_3_5/EXP_D from P14. A brief summary of the case and experiment settings is provided below.

a. System overview

On 8 May 2007, an MCS developed in western Texas and moved to the northeast into southwestern and central Oklahoma during the evening hours (approximately 0000-0500 UTC 9 May). During the day on 8 May, a positively-tilted upper level trough and seasonably warm, moist air at the surface led to the development of widespread convection over western Texas. The cool outflow from these storms helped to initiate additional convection and contributed to upscale growth over time as the storms became organized into a convective line. Ahead of the line, isolated supercell storms developed in northwest Texas and southwest Oklahoma. The developing MCS interacted with two of these storms, leading to the development and maintenance of a line end vortex (LEV) near the northern end of the MCS (P14, Schenkman et al. 2011). During the 0100 – 0500 UTC 9 May timeframe the system remained in the asymmetric stage of MCS development, with a broad area of leading stratiform precipitation, an intense leading convective line, and a trailing region of stratiform precipitation (Fig. 1, with term definitions based on Fritsch and Forbes (2001)). Widespread heavy rain was observed with this MCS, and four tornadoes were reported near the LEV (NWS 2012). For a more detailed discussion of the development, structure, and impacts of this MCS, we refer the reader to P14, Schenkman et al. (2011), and Snook et al. (2011).

b. Forecast model settings

The forecast model used is the Advanced Regional Prediction System (ARPS, Xue et al. 2000; Xue et al. 2001; Xue et al. 2003). ARPS is a fully-compressible, non-hydrostatic, three-dimensional atmospheric model suitable for convective-scale simulation and prediction. ARPS predicts the three-dimensional wind components (u,v,w), pressure (p), potential temperature (θ), water vapor mixing ratio (q_v), as well as the mixing ratios for cloud water (q_c), rain (q_r), snow (q_s), ice (q_i), graupel (q_g), and hail (q_h) for a SM MP scheme. For a DM MP scheme, the model also predicts the hydrometeor number concentrations (N_{tx}, where x refers to individual hydrometeor species). Additional parameterizations used include NASA Goddard Space Flight Center long- and shortwave radiation, 1.5-order turbulent kinetic energy (TKE)-based subgrid-scale turbulence closure and convective boundary layer parameterization schemes, and a two-layer land surface/soil-vegetation model. More details on the model physics can be found in Xue et al. (2001) . The model domain used consists of 259 × 259 grid points in the horizontal with a 2 km horizontal grid spacing and a stretched vertical grid using 53 vertical grid points with a minimum grid spacing of 100 m and average grid spacing of 500 m. The model terrain is interpolated to the 2 km grid from a 30 arcsecond high resolution USGS dataset.

The full experiment consists of a 1-hour (h) spin-up period, 1-h data assimilation period, and a 3-h ensemble forecast. During the spin-up period, a 1-h deterministic forecast on the 2 km model grid is initialized from the NCEP North American Mesoscale Model (NAM) analysis at 0000 UTC. The 3-h NAM forecast from 0000 UTC valid at 0300 UTC and the NAM analysis at 0600 UTC provide lateral boundary conditions during the forecast. At 0100 UTC, smoothed random perturbations are added to the 1-h spin-up forecast (Tong and Xue 2008, Snook et al. 2011) to initialize a 40-member ensemble for performing the EnKF data assimilation cycles. The first assimilation is performed at 0105 UTC and the last at 0200 UTC, with an assimilation cycle length of 5 minutes. Only radar data are assimilated. Further details on the data assimilation are given below. Following the assimilation period, the final ensemble analyses are used to initialize 3-h ensemble forecasts from 0200 UTC through 0500 UTC.

c. Data sources

As in Snook et al. (2011) and P14, Level-II Z and radial velocity (V_r) data from five WSR-88D S-band radars in Oklahoma and Texas are assimilated. These include KTLX of Twin Lakes, Oklahoma City, Oklahoma, KVNX of Vance Air Force Base, Oklahoma, KAMA of Amarillo, Texas, KLBB of Lubbock, Texas, and KDYX of Abilene, Texas. Together, these five radar sites provide full coverage of the MCS during the DA period. KFDR (Fredrick, Oklahoma) is also located near the MCS, but level-II data from KFDR are unavailable during the assimilation window. Z and V_r data are also assimilated from four experimental X-band radars maintained by the Engineering Research Center (ERC) for Collaborative and Adaptive Sensing of the Atmosphere (CASA, McLaughlin et al. 2009) in southwestern Oklahoma. These radars, KCYR (Cyril, Oklahoma); KSAO (Chickasha, Oklahoma); KLWE (Lawton, Oklahoma); and KRSP (Rush Springs, Oklahoma), provide additional low-level radar coverage over a portion of the MCS near the LEV. The National Severe Storms Laboratory's dual-pol S-band radar KOUN is used for verification. The locations of radars used in this study are marked in Fig. 1.

Radar observations are interpolated to the model grid horizontally, but are left at the height of the radar elevation scan in the vertical, following Xue et al. (2006). The observations are interpolated to the time of each assimilation cycle using the previous and subsequent volume scan. Quality control procedures, include despeckling, ground clutter removal, and velocity dealiasing, are applied to the radar data prior to assimilation. For the CASA X-band Z observations, attenuation correction is performed before the data are assimilated (Chandrasekar et al. 2004). The data quality control procedure used follows P14. Specifically, for KOUN, dual-pol variables are removed when $\rho_{HV} < 0.8$, which corresponds to non-meteorological echoes. K_{DP} is calculated by first unfolding and then smoothing the differential phase (Φ_{DP}) data using an averaging window with 9 gates when $Z > 40$ dBZ and 25 gates when $Z < 40$ dBZ. The least squares fit method of Ryzhkov and Zrnic (1996) is then used to calculate K_{DP} using the same threshold to determine the number of gates.

d. Ensemble Kalman filter settings

The EnKF algorithm used is an implementation of the ensemble square root filter (EnSRF) of Whitaker and Hamill (2002). As mentioned earlier, the ensemble is first initialized at 0100 UTC by adding random, smoothed, Gaussian perturbations to the 1-h spin-up forecast. Perturbations with a standard deviation of 2 m s^{-1} are added to u, v, and w and a standard deviation of 2 K to θ (using positive values only) across the entire model domain. Additional perturbations with a standard deviation of .001 kg kg^{-1} are added to the hydrometeor mixing ratios and water vapor but they are confined to regions of precipitation where Z is greater than 5 dBZ. The perturbations are smoothed following Tong and Xue (2008) and use a horizontal correlation length scale of 8 km and vertical scale of 5 km.

Processed Z and V_r data from the nine radars are assimilated every 5 minutes between 0105 UTC and 0200 UTC. This includes clear-air Z data from the WSR-88D radars, which Tong and Xue (2005) have shown helps to suppress development of spurious convection. Clear-air data from the CASA network are not used because of uncertainties associated with the X-band attenuation (Z values similar to those associated with clear air may be due to a completely attenuated signal). Assimilation of V_r is limited to regions where $Z > 20$ dBZ. The radar observation operator used is that of Jung et al. (2008), which is different from that used in Snook et al. (2011) and the same as that in P14. A horizontal and vertical covariance localization radius of 6 km is used for both Z and V_r based on the correlation function of Gaspari and Cohn (1999).

The observation error and covariance inflation methods used are the same as in P14. They were chosen based on preliminary experiments using various configurations. Radar observation error values of 5 dBZ for Z and 3 m s^{-1} for V_r are used. Multiplicative inflation (Anderson 2001) with a factor of 1.25 is applied to the prior ensemble for grid points where $Z > 20$ dBZ in order to maintain ensemble spread and produce a closer to optimal consistency ratio value (Dowell et al.

2004) throughout the assimilation period than could be achieved using lower values of observation error and other covariance inflation methods such as additive noise (Dowell and Wicker 2009) and relaxation to prior ensemble (Zhang et al. 2004).

e. Microphysics schemes used and their configurations

The two control experiments differ solely in terms of the microphysics scheme used. EXP_S uses a combination of three different SM MP schemes during both the assimilation period and the forecast. Using multiple MP schemes within the ensemble was shown to increase ensemble spread and improve root-mean-square innovation (RMSI) during the assimilation period by Snook et al. (2011). Of the 40 ensemble members, 16 use the Lin scheme (Lin et al. 1983), 16 use the WRF single-moment 6-class WSM6 scheme (Hong and Lim 2006), and 8 use the simplified NWP scheme (NEM) of Schultz 1995). Fewer NEM members are included because NEM member forecasts did not tend to perform as well as members using the other SM schemes. The intercept parameter used for rain (N_{0r}) is reduced by a factor of 10 from the typical value of 8 x 10^6 m^{-4} to 8 x 10^5 m^{-4} , following Snook and Xue (2008), who found that the reduced N_{0r} value led to a lower and more realistic evaporation rate and associated surface cold pool intensity.

The DM experiment, EXP_D, uses the Milbrandt and Yau (MY, 2005b) scheme. During the assimilation period, the shape parameters (α) for rain and hail vary inversely between 0.0 and 2.0 in 0.05 increments for each member to increase ensemble spread. All other hydrometeor categories use α = 0; furthermore α is set to 0 for all categories in the forecasts after 0200 UTC. As in Snook et al. (Snook et al. 2011) and P14, the graupel category of the MY scheme is turned off to more closely resemble the majority of members in EXP_S which did not predict graupel.

3. Results of experiments

In this section, ensemble forecast results from EXP_S and EXP_D are presented. The results are divided into two parts: (1) an evaluation of the overall forecast quality of the complete MCS using *Z* mosaics and (2) verification of simulated dual-pol variables against KOUN observations. Evaluations include qualitative discussion of system structure and feature placement, evaluation of probabilistic forecasts, as well as quantitative verification. We also discuss methods and challenges as they relate to dual-pol variables.

a. Ensemble forecasts of radar reflectivity

1) QUALITATIVE EVALUATION OF REFLECTIVITY MOSAICS

Ensemble forecasts of the MCS are evaluated at 1, 2, and 3-h of forecast time by verifying the probability matched ensemble mean (PMEM; Ebert 2001) forecasts of *Z* from EXP_S and EXP_D against constant height mosaics of observed *Z* at model level 10, which is approximately 2 km AGL (Fig. 2). Model level 10 is the lowest level where complete radar coverage of the MCS is available without gaps between radars. The mosaics of observed *Z* are created by combining observations from the five WSR-88D radars used during assimilation, with observations interpolated to the model grid as discussed above in section 2c. Where multiple radars observe a specific grid point, the maximum value of *Z* is used in the mosaics. The larger values are used because they are less likely to have been subject to resolution smearing and attenuation effects, although the latter is usually rather small. The PMEM is used instead of a regular ensemble mean because *Z* can vary greatly over small distances, leading to under-prediction of intensity and over-prediction of areal coverage when ensemble members with even slightly-displaced convective features are averaged. The PMEM ranks all *Z* values in the domain from highest to lowest for both the ensemble mean and the full ensemble, then reassigns values from the full ensemble probability density function of *Z* to the grid location with the same rank in the ensemble mean; this process helps mitigate the aforementioned biases introduced by taking the ensemble mean (Ebert 2001; Clark et al. 2009).

Unlike in P14, the simulated radar variables in the results in this manuscript use a different, more complex observation operator than was used for EnKF DA. This operator, outlined in Jung et al. (2010), uses a lookup table of scattering amplitudes for all hydrometeors calculated using the T-matrix method (Vivekanandan et al. 1991; Bringi and Chandrasekar 2001). This operator enables us to take into account Mie scattering for large ice particles, such as hail or graupel, and to use a new axis ratio for rain revised based on observations. The simpler operator used during EnKF DA, based on Jung et al. (2008), uses a fitted approximation to the T-matrix values for rain, and uses the Rayleigh approximation for ice species. The simpler operator is used during DA to reduce computational expense, while the more advanced operator is used for forecast verification because it allows for a more realistic comparison to observations. Specifically, this has a noticeable effect on Z_{DR}, reducing maximum values by more than 0.5 dB, which is beyond the estimated uncertainty of observed Z_{DR} of approximately 0.1-0.3 dB (Ryzhkov et al. 2005; Doviak and Zrnic 1993). The new operator also, more correctly, simulates lower Z values for both dry and wet hail beyond the typical uncertainty for Z observations, which is approximately 1-2 dBZ.

The PMEM of Z in EXP_S contains a region of anomalously high Z (>55 dB) centered near the LEV (see Fig. 1), and there is little distinction between regions of stratiform and convective precipitation (Fig. 2d-f). The intensity of the trailing stratiform precipitation is also over-forecast. On the other hand, the PMEM of Z in EXP_D (Fig. 2g-i) contains broader precipitation coverage in the leading stratiform region and a convective line with greater southern extent, though it does over-forecast Z intensity in the leading stratiform region. The ensemble spread of Z is lower in EXP_D than in EXP_S (not shown); only one MP scheme is used in EXP_D, leading to closer agreement among members and higher ensemble mean values (Snook et al. 2012). These results are similar to those obtained in deterministic forecasts of this case in P14, where the authors found the size sorting of smaller raindrops rearward in the leading convective line when using a DM scheme (absent in EXP_S) led to greater evaporative cooling and a stronger cold pool that helped maintain a more realistic MCS structure. They also found that the cold pool in EXP_S is disorganized, contributing to the development of spurious convection near the LEV. It should be noted that neither EXP_S nor EXP_D predict the small clusters of storms that develop in the southeast and southwest portion of the domain in the observations, likely in part because this convection developed mostly after the DA period.

2) PROBABILISTIC FORECASTS OF REFLECTIVITY

Uncertainty within the ensemble forecast due to, e.g., initial condition and model errors, can be considered by producing probabilistic forecasts of *Z* from the forecast ensemble. High-resolution, convection-permitting NWP forecasts are particularly sensitive to timing and location errors as forecast lead time increases due to the small spatial and temporal scales of convective storms (Lorenz 1969; Roberts 2008). To account for this sensitivity, we use the neighborhood ensemble probability (NEP) method (Ebert 2008; Roberts and Lean 2008; Schwartz et al. 2009), which, at each model grid point, produces a probabilistic forecast using a collection of nearby points in all ensemble members rather than relying solely on data from that single grid point in each member. In this way, NEP accounts for spatial uncertainty as well as uncertainty conferred by the ensemble. Appropriate specification of the neighborhood is important; in this study we use a circular neighborhood with a radius of 5 km, which is appropriate for the grid spacing used and convective features predicted (Snook et al. 2012; 2015). NEP is calculated for P[Z > 20 dBZ] (Fig. 3) and P[Z > 40 dBZ] (Fig. 4) at the same vertical level 10 as in Fig. 2. The 20 dBZ threshold is used to consider overall precipitation coverage in the MCS, including the stratiform regions, while the 40 dBZ threshold is chosen to focus on areas of heavy, convective precipitation. In Fig. 3 and Fig. 4, the observed *Z* contours for the corresponding threshold are also plotted.

As was noted in the PMEM forecasts, the NEP forecasts of *Z* for EXP_D exhibit improved precipitation structure and feature placement compared to EXP_S. At the 20 dBZ threshold, the region of high P[*Z* > 20 dBZ] in EXP_D (Fig. 3d-f) closely matches the observed region of 20+ dBZ *Z*, particularly in the leading stratiform region and leading convective line. In particular, EXP_D predicts broad area of very high probability (> 0.9) that closely matches the observed leading stratiform region in terms of position, shape, and motion throughout the forecast period. In contrast, EXP_S (Fig. 3a-c) exhibits high probability (> 0.8) for only about half of the observed region of 20+ dBZ *Z* during the first two hours of the forecast, and even less in the 3-h forecast. EXP_S also has a substantial region of moderately high probabilities (up to 0.8) to the west of the MCS where no precipitation is observed. Considering the individual SM microphysics schemes within EXP_S, the LIN members exhibit the best agreement with observations in terms of forecast coverage and intensity of *Z*; WSM6 members generally over-forecast the extent of the trailing stratiform region, while NEM members under-forecast the extent of both the trailing and leading stratiform regions (not shown). These results are consistent with those of Snook et al. (2012), which, using a similar ensemble with the same MP schemes, found that the RMS innovation of Lin members during the forecast period was lower than that of WSM6 and NEM members. In both EXP_D and EXP_S, low probabilities are predicted for the for the trailing stratiform precipitation region; overall, this region is the worst forecast portion of the MCS.

Although overall precipitation coverage (*Z* > 20 dBZ) is generally good for both cases, the P[Z > 40 dBZ] associated with heavier convective precipitation exhibits greater error. EXP_S has only a small overlap of low probabilities (0.05-0.2) with the observed 40 dBZ region in the 1-h forecast (Fig. 4a); EXP_D has greater overlap throughout the forecast period (Fig. 4d-f), but the predicted probabilities remain low. The convective line has a width of a few km and will be more susceptible to spatial error as forecast lead time increases compared to the stratiform regions, even with the consideration of a 5 km neighborhood. EXP_D also has higher probabilities in the convection near the LEV on the north end of the MCS. However, there are areas of high probability in the stratiform region as well, where EXP_D over-forecasts *Z* intensity. The over-forecast in intensity is in part due the height of the model grid used in Fig. 2-4 being at the bottom of the model melting layer, where high *Z* occurs due to the presence of large, oblate, and wet hydrometeors. MP schemes used thus far with the simulator have shown a tendency to delay melting until warmer temperatures at lower elevations below the 0º isotherm in the model, and compared to observations, due to overestimated evaporative cooling (Jung et al. 2008; Jung et al. 2010; Johnson et al. 2016). Because radar coverage is incomplete below this level, though, this issue is difficult to avoid. A modified melting model in the radar simulator that includes temperature information to help account for the delay is considered for future work. Previous studies have also shown that DM MP schemes can overestimate *Z* values compared to observations due to excessive size sorting (Kumjian and Ryzhkov 2012).

3) QUANTITATIVE EVALUATION OF REFLECTIVITY FORECASTS

Qualitative evaluations based on the PMEM (Fig. 2) show quite skillful forecasts in terms of Z but there are still apparent spatial errors that would adversely affect quantitative skill scores. The NEP of *Z* > 40 dBZ used to identify the leading convective line indicated how small spatial error can lead to lower *Z* probabilities. When considering features with small spatial scales, scores such as the equitable threat score, which consider hits, misses, and false alarms in a deterministic point by point framework, are susceptible to a ‘double-penalty’: a forecast with even a modest spatial displacement in a feature not only misses the observed feature but also produces a false alarm because the forecast feature is not coincident with any observed feature (Ebert and McBride 2000; Rossa et al. 2008; Mittermaier et al. 2013). Therefore, quantitative measures that consider the probability of an event within a neighborhood are considered.

The first metric considered is the area under the relative operating characteristic (ROC) curve (AUC, Mason 1982; Mason and Graham 1999). The AUC is a summary skill score that compares the probability of detection and the probability of false detection for a given event over a range of probability thresholds; in this case the event is *Z* exceeding a given threshold within a neighborhood with a 5 km radius. Possible AUC values range from 0.0 to 1.0, with 1.0 indicating a perfect forecast (no false alarms or misses). AUC values of 0.5 or below indicate that the forecast has no useful skill. AUC is calculated for EXP_S and EXP_D using *Z* thresholds ranging from 10 dBZ to 50 dBZ for 1, 2, and 3-h forecast times (Fig. 5), and a bootstrap procedure is used to resample the ensemble 1000 times to determine the 5^{th} to 95^{th} percentile range, which is shaded. Background shading is included to indicate the areas of useful forecast skill (green; AUC > 0.7), low skill (yellow; 0.5 < AUC < 0.7), and no skill (red; AUC < 0.5). Calculations are performed over the full experiment domain (Fig. 5a-c) as well as an Oklahoma subdomain positioned to cover the leading stratiform region and leading convective line, where both forecasts performed better compared to the trailing line (Fig. 5d-f).

Both experiments generally produce high AUC values, except for the very highest *Z* thresholds, associated with intense convective precipitation; confidence in AUC at these thresholds is low, however, because the sample size of *Z* exceeding these thresholds is quite small, and the regions in question are very small in spatial extent. AUC also, as expected, decreases with increasing forecast time. In general, EXP_D shows improvement over EXP_S in skill, especially for moderate *Z* thresholds representing the stratiform region in the later hours.

The AUC increases overall for both experiments when calculations are limited to the Oklahoma subdomain (Fig. 5d-f). In the 1-h forecast, AUC is similar in EXP_S and EXP_D, but in the 2 and 3-h forecasts, EXP_D outperforms EXP_S in terms of AUC at nearly all thresholds. In particular, EXP_D has an AUC value over 0.9 for thresholds of 20-25 dBZ throughout the forecast period over the Oklahoma subdomain (Fig. 5d-f), indicating a highly skillful forecast of general precipitation coverage of the leading stratiform region. EXP_D also exhibits useful skill (AUC > 0.7) for higher *Z* thresholds representing convective precipitation throughout the forecast period over the Oklahoma subdomain, suggesting that the poorer scores over the full domain are partially due to the overly quick dissipation of the trailing convective line and the newly-developed convection in the southern portion of the domain, while the leading convective line is generally well forecast.

Reliability and sharpness diagrams are examined next. A probabilistic forecast is considered reliable when the probability of an event forecast to occur closely corresponds to the rate at which the event actually occurs (Brown 2001). Reliability diagrams are calculated for P[Z > 20 dBZ] using a 5 km radius neighborhood at 1, 2, and 3-h forecast times (Fig. 6). In these reliability diagrams, perfect reliability is indicated by the one-to-one diagonal and the shaded region indicates a skillful forecast. Areas where the calculated reliability lies above the diagonal indicate that *Z* is under-forecast (forecast probability is lower than the observed frequency); conversely, areas below the diagonal indicate that *Z* is over-forecast (forecast probability is higher than the observed frequency). Sharpness diagrams, which are histograms of the calculated probability values, are shown in Fig. 7. An ideal forecast will have many values near 1.0 or 0.0, distinguishing sharply between events and non-events. Calculations are again performed over both the full domain and Oklahoma subdomain.

Overall, there is not much difference in the reliability of EXP_S and EXP_D either on the full domain or the Oklahoma subdomain. For the 1-h forecast time (Fig. 6a,d), both forecasts show good reliability, with the region of Z > 20 dBZ slightly under-forecast in EXP_S and slightly over-forecast in EXP_D. For the 2 and 3-h forecast times (Fig. 6b,c,e,f), precipitation coverage is generally over-forecast in both experiments. EXP_D does show greater sharpness than EXP_S, particularly over the Oklahoma domain (Fig. 7j-l). Both experiments have a large number of probabilities of 0.0 that represent the large areas where precipitation is not observed, but EXP_D has a much higher number of points with probabilities close to 1.0 where the ensemble predicts precipitation with very high confidence. As indicated by the AUC (Fig. 5) and the qualitative evaluation of NEP forecasts (Fig. 3), this region of very high confidence agrees well with observations in EXP_D, outperforming EXP_S. One concern is that the mixed MP scheme setup of EXP_S may lead to a decrease in sharpness compared to EXP_D because of higher spread in the ensemble. However, Snook et al. (2012) showed that the use of multiple SM MP schemes in an ensemble actually increased the sharpness of a forecast of mesovortices compared to an ensemble using only the SM LIN MP scheme.

b. Ensemble forecasts of polarimetric variables

1) QUALITATIVE EVALUATION OF PREDICTED POLARIMETRIC VARIABLES

The PMEM is calculated as in Fig. 2 for simulated Z, Z_{DR}, and K_{DP} as though the ensemble forecasts of EXP_S and EXP_D were observed by KOUN at 1-h (Fig. 8), 2-h (Fig. 9), and 3-h (Fig. 10) forecast times; KOUN observations at the corresponding times are provided for comparison. The simulated fields are shown at the 0.5° elevation; this choice of the lowest elevation is because dual-pol radar signatures tend to be the strongest at the low levels where size sorting effects (Dawson et al. 2014) and rain water species dominate. Also, the lower elevation is less affected by the melting layer. The difference in Z between the forecasts over the KOUN observing region is similar to the PMEM mosaics considered earlier (Fig. 2); EXP_D exhibits improved representation of the leading convective line and better coverage of the stratiform region compared to EXP_S, though it somewhat overestimates intensity due to the low model melting layer compared to the 0º isotherm and the excessive size-sorting seen in DM MP schemes.

There are two notable differences between EXP_D and EXP_S in terms of their forecast dual-pol fields. First, the areal coverage of high Z_{DR} values (Z_{DR} > 2.3dB), a threshold that distinguishes the convective region from the stratiform region in the observations, is over-forecast in EXP_S. The highest Z_{DR} values predicted by EXP_S are coincident with the poorly-organized region of intense convection within the system due to the monotonic relationship between the Z and Z_{DR} (e.g., Fig. 8e). The Z_{DR} values in EXP_D (Fig. 8f), while slightly higher than the observations (Fig. 8d), still show a similar general distribution of high and low Z_{DR} regions compared to the observations, indicating a distinct difference in maximum raindrop size between the convective and stratiform regions that is maintained throughout the entire forecast period. P14 found that these MCS features were maintained by an improved cold pool due to increased evaporative cooling from the advection of small raindrops rearward by the DM scheme.

The second notable difference is that the K_{DP} values in EXP_S are unrealistically high when compared to the observations, with values peaking at nearly 10° km^{-1}, suggesting that EXP_S greatly over-forecasts liquid water content in the convective precipitation. By comparison, K_{DP} in EXP_D is much closer to the observations. In fact, the q_r values near the surface in EXP_S associated with the maximum values of simulated K_{DP} are, on average, twice as high as those in EXP_D. The difference in K_{DP} values during the forecast in the current study is notable when compared to P14, where dual-pol variables were considered only for the analysis (not for forecasts). In the P14 analysis, K_{DP} values were generally quite similar between SM and DM experiments, with K_{DP} slightly underestimated compared to the observations due to a high hail bias. However, the Z_{DR} patterns differ between the SM and DM analyses in P14 and in the forecasts of EXP_S and EXP_D in the current study. These differences in Z_{DR} occur in the stratiform region. The region of non-zero K_{DP} values is mainly confined to the leading convective line. The unrealistically-high K_{DP} values in EXP_S occur by the first forecast hour. Rain development in the stratiform region of the MCS is heavily dependent on the transport of frozen hydrometeors in the mid and upper levels of the MCS from the convective to the stratiform region. There is very little hydrometeor transport from the convective line to stratiform region in the SM case (P14), and therefore there is a higher precipitation rate in the convective line. While the difference in K_{DP} values in the convective region between the SM and DM experiment is less substantial in the analysis, the improved development and maintenance of the MCS when using the DM scheme leads to improved representation of Z_{DR} and K_{DP} fields, compared to the observations, throughout the forecast (Fig. 8- Fig. 10).

The patterns in the dual-pol variables that reflect microphysical processes can be subtle; one such pattern is increased Z_{DR} along the leading convective line due to size sorting. Though the PMEM helps to alleviate some of the biases introduced by taking an ensemble mean, it can smear such high-detail patterns. For this reason, the best individual ensemble member from each experiment is examined in order to bring to light distinct pattern differences within the predicted dual-pol fields (Fig. 11). The 2-h forecast of EXP_S member 14 and EXP_D member 39 are chosen based upon a qualitative examination of the ensemble members that considers placement of system features, Z_{DR} patterns, and overall value range. The best EXP_S member contains precipitation extending southeastward where the observations have the leading convective line, but the intensity and extent is rather limited compared to the best EXP_D member. As expected, areas of high Z_{DR} coincide with areas of high Z in the EXP_S member. In the EXP_D member, however, high Z_{DR} values are located along the eastern/leading edge of the leading convective line. This Z_{DR} pattern is indicative of the size sorting of raindrops within the convective line, with smaller raindrops being advected rearward in the line while larger raindrops remain.

2) PROBABILISTIC FORECASTS OF POLARIMETRIC VARIABLES

In section 3a2, probabilistic forecasts were used to evaluate the ensemble forecast precipitation coverage of stratiform and convective precipitation, based on 20 and 40 dBZ Z thresholds, respectively. A distinct variation in the Z_{DR} values also occurs, with Z_{DR} increasing where larger raindrops are present along the leading edge of the convective line. To evaluate how well the two experiments forecast the high Z_{DR} signatures, the probability of Z_{DR} > 2.3 dB at the 1, 2, and 3-h forecast time is calculated (Fig. 12). The threshold of Z_{DR} = 2.3 dB is chosen based on the observed values in this case (Fig. 8d, 9d, 10d), and the observed Z_{DR} = 2.3 dB contour is shown as a thick black line. EXP_S has a broad expanse of relatively high probability of Z_{DR} > 2.3 dB over the stratiform region, a result consistent with the overall pattern of Z_{DR} in Fig. 8e, 9e, and 10e. This region of high P[Z_{DR} > 2.3 dB] is significantly displaced from the observed leading convective line. In EXP_D, there is some overlap of low to moderate probabilities of Z_{DR} > 2.3 dB with the observed 2.3 dB contour in the 1-h forecast, and some overlap of low probabilities at the 2 and 3-h forecasts. Though the regions of moderate P[Z_{DR} > 2.3 dB] in EXP_D do not exactly match the observed region of high Z_{DR}, the geographic distribution of higher probability follows a north-northwest to south-southeast orientation, similar to the observed leading convective line, and substantially improved compared to the more circular pattern found in EXP_S. The EXP_D probabilistic forecast of Z_{DR} thus has greater practical value, indicating moderate probability of an arc of larger raindrops relatively near the observed leading convective line, correctly indicating the existence and general direction of motion of the leading convective updrafts in the MCS within the ensemble forecast.

3) QUANTITATIVE VERIFICATION OF POLARIMETRIC VARIABLES

The same concerns for how small spatial errors can affect quantitative skill scores of *Z* discussed in section 3a3 are even greater when considering skill scores for predicted dual-pol variables. Dual-pol signatures follow patterns associated with microphysical processes that occur at very small scales, such as the size-sorting of raindrops along the leading convective line. With this potential limitation in mind, the AUC is calculated for Z_{DR} (0.0 to 2.7 dB) and K_{DP} (0.0 to 1.5 ° km^{-1}) for the 1, 2, and 3-h forecasts (Fig. 13) using a 5 km neighborhood radius as was done in Fig. 5 for Z. Both experiments have similar, skillful AUC values for predicting Z_{DR} at thresholds of 0.0 to 1.0 dB (Fig. 13a-c). For higher thresholds, the AUC for EXP_S indicates very poor skill, while EXP_D still produces a skillful forecast. AUC for Z_{DR} is better in EXP_D due to the lower Z_{DR} values throughout the leading stratiform region, which agree much more closely with observations than the forecast of EXP_S. The Z_{DR} associated with the leading convective line also has a good overlap with observed values in EXP_D. EXP_S outperforms EXP_D for the considered thresholds of K_{DP} due to erroneous broader coverage in EXP_S that overlaps the observations and the displacement error in EXP_D. K_{DP} coverage is significantly less than either Z or Z_{DR}; AUC is particularly sensitive to the probability of detection and therefore EXP_D scores are poorer. Additionally, the significant high bias in K_{DP} in EXP_S is not accounted for at these thresholds chosen based on observed values; the AUC threshold limit is set to 1.5 ° km^{-1} because few observations exceed this value. K_{DP} is poorer qualitatively in comparison to EXP_D, but limitations in the quantitative scores used lead to poor and misleading results.

Due to the large impact of spatial error on the quantitative skill scores for the dual-pol variables, other quantitative methods of evaluation not reliant on location are useful. Domain-wide histograms of the simulated dual-pol variables can be used to identify significant biases in the forecast. Histograms of the simulated values from all members of EXP_S and EXP_D as well as the observed values are plotted in Fig. 14. The values from EXP_S and EXP_D are normalized by the size of the ensemble for comparison to the observations. For observed Z, values associated with the widespread stratiform precipitation lead to a peak between about 30 to 35 dBZ throughout the experiment period (Fig. 14a-c). The EXP_D ensemble forecast *Z* values match the observed distribution in this range better than EXP_S during the first two forecast hours. Both experiments over-forecast the geographic extent of the convective precipitation, and over-forecast the intensity of Z in part due to a delayed model melting layer relative to the 0º isotherm and excessive size-sorting in EXP_D, leading to a higher number of $Z > 50$ dBZ values compared to the observations, though this high bias is slightly greater in EXP_S than in EXP_D in the 1- and 2-hour forecasts.

Differences between EXP_S and EXP_D are readily apparent in histograms of the predicted dual-pol values (Fig. 14d-i). Observed Z_{DR} values (Fig. 14d-f) peak at about 1.0 to 1.5 dB due to the broad coverage of moderately-sized raindrops in the leading stratiform region. EXP_D over-forecasts the coverage of the leading stratiform precipitation, leading to an overall high-bias in the Z_{DR} histogram, and slightly over-forecasts the location of the histogram peak in Z_{DR} values, but the overall histogram pattern is similar to that of the observations. EXP_S, on the other hand, has a uniform distribution of Z_{DR} values throughout the forecast period, with no evidence of the peak seen in the observations and in EXP_D, due to the lack of broad coverage of stratiform precipitation in EXP_S. EXP_S also has a larger number of very high values ($Z_{DR} >$ 3.0dB) resulting from the unorganized region of intense convection in the center of the system. Relatively little bias is noted in the K_{DP} histograms for EXP_D, producing histograms similar to that produced by to the dual-pol observations (Fig. 14g-i). EXP_S over-forecasts the total coverage of non-zero K_{DP} values, again suggesting a high-bias in liquid water content overall compared to the observations. This substantial high-bias in liquid water content in convective precipitation skews EXP_S towards high values, with grid volumes exhibiting $K_{DP} > 3.0\ ° \text{ km}^{-1}$, particularly in the 1-hour forecast (Fig. 14g).

4. Summary and conclusions

Ensemble forecasts initialized from cycled EnKF ensemble analyses are produced for a mesoscale convective system (MCS) that occurred over Oklahoma and northern Texas on 8-9 May 2007 using single-moment (SM, Lin et al. 1983) and double-moment (DM) microphysics (Milbrandt and Yau 2005b) schemes. Qualitative and quantitative probabilistic methods are used to examine the MCS structure and precipitation distribution for the SM (EXP_S) and DM (EXP_D) experiments. Additionally, predicted dual-polarization (dual-pol) radar variables and their probabilistic forecasts are also evaluated against available dual-pol radar observations, and discussed in connection with model predicted microphysical states and structures. The current study expands on the work of Putnam et al. (2014) which focused on the EnKF data assimilation and the deterministic forecasting aspects of the same two experiments that used SM and DM microphysics schemes, respectively. This paper focuses on ensemble probabilistic forecasting of reflectivity and the simulated dual-pol radar variables associated with the 8-9 May 2007 MCS.

Both qualitative and quantitative evaluations of the probabilistic forecasts show that EXP_D predicts the MCS with high confidence. EXP_D predicts the overall precipitation coverage of the system (considering a threshold region of Z > 20 dBZ) with very high probabilities throughout the forecast period, particularly for the stratiform precipitation region. EXP_S predicts similarly high probabilities for approximately half of this region and includes a large area of moderate probability of Z > 20 dBZ outside of the observed region. EXP_D has higher forecast skill, measured in terms of the area under the relative operating characteristic curve (AUC), for 2 and 3-h forecasts of the stratiform precipitation and leading convective line comprising the northern portion of the MCS. EXP_D also provides ensemble forecasts with greater sharpness, as well as one in which the highest precipitation probabilities match regions of observed precipitation at a higher frequency.

EXP_D better represents the microphysics-related features in the MCS throughout the forecast period. This is notable in terms of Z_{DR} values, where the dual-moment ensemble forecast shows a clear distinction between the convective and stratiform precipitation regions, similar to that seen in the final EnKF analysis in Putnam et al. (2014), which continues throughout the forecast period. Additionally, EXP_D implies more realistic liquid water content in the convective region than EXP_S, where unrealistically high K_{DP} values suggest the liquid water content has been over-forecast, associated with the unorganized system structure and precipitation development in the forecasts. When evaluating a DM ensemble, the increased computational expense cost of a DM over a SM scheme should also be considered. Previous studies have shown that DM schemes can increase computation time by 10-30% depending on the scheme used (Morrison et al. 2005; Milbrandt and McTaggart-Cowan 2008; Morrison and Gettelman 2008; Lim and Hong 2010). However, future increases in available computing resources will make the operational use of DM schemes increasingly feasible, and current operationally-oriented research projects, such as the 2016 Storm Scale Ensemble Forecasts (SSEF) produced by the Center for Analysis and Prediction of Storms (CAPS, Kong 2016), as part of the NOAA Hazardous Weather Testbed Spring Experiment, are already using DM MP schemes successfully in real time.

Producing meaningful probabilistic forecasts of dual-polarization (dual-pol) variables proves challenging. Dual-pol signatures are often produced by physical processes within convective systems with very small spatial scales; often less than a few kilometers. These small-scale structures are smeared when probabilistic forecasts are generated using neighborhood-methods or a probability-matched ensemble mean. When individual ensemble members are examined, though, EXP_D maintains a better quality in the predicted dual-pol fields compared to EXP_S, as in the final EnKF analyses noted in Putnam et al. (2014). There is a notable arc of high Z_{DR} along the leading convective line in the MCS, resulting from size-sorting processes that are not represented in EXP_S, where Z_{DR} shows a monotonic relationship with Z. Probabilistic forecasts of Z_{DR} for EXP_D, while not particularly accurate in matching the location of the observations, still indicate the presence of this arc of large raindrops along the leading convective line and the general direction and speed of motion of the convective updrafts of the line. The EXP_D Z_{DR} forecasts also show higher skill based on AUC calculations compared to EXP_S. For K_{DP}, the EXP_S forecasts show higher skill. However, this is due to spatial displacement in EXP_D and significant erroneous coverage of K_{DP} in EXP_S; the AUC is more sensitive to the probability of detection. The low probabilities and spatial displacement errors associated with both of these variables indicate how uncertain the forecast of intense, convective updrafts and excessive rainfall can be. Understanding and increasing the skill in predicting intense rainfall rates has the greatest broader impact on forecasting potential flash flood events.

This is the first study to consider explicit ensemble-based probabilistic forecasting of simulated dual-pol radar variables, and it highlights several challenges for future work. Even on high resolution grids capable of resolving microphysical patterns that occur on small spatial scales, quantitative verification scores for dual-pol signatures that usually have very small spatial scales (even compared to convective storms) suffer from a double penalty: forecasts of precipitation variables not only miss the location of the observations (a 'miss'), but also occur in a nearby location where the event was not observed (a 'false alarm'). This was also noted in a recent study evaluating storm-scale forecasts using different DM MP schemes (Putnam et al. 2016). Some probabilistic neighborhood-based metrics are used in this case, including the AUC, to help account for spatial errors, but the distance and orientation of patterns in the simulated variables still presents a challenge when using such methods. Scores for dual-pol variables, specifically K_{DP}, are poorer as the threshold considered increases, despite the neighborhood radius of 5 km used, due to both the small spatial scale of the patterns being considered and discrepancy in the range of forecast values versus observed values. Although using a larger neighborhood may alleviate to a larger extent the effect of spatial error, the probabilistic forecasts produced using progressively larger neighborhood radii will be more and more smoothed, losing the resolution necessary to capture small-scale features and negating their intended purpose. Additional methods of quantitatively evaluating dual-pol variables include histograms, which can provide information on general biases without considering spatial error. In such histograms produced for this case, high biases in the number of large drops and overall liquid water content, as suggested by the high biases in predicted K_{DP} values, are identified in EXP_S, likely due to the representation of convective precipitation within EXP_S. Possible future quantitative verification methods for dual-pol fields include object-based methods (e.g. Davis et al. 2006, Johnson et al. 2013, Zhu et al. 2015) that match similar storm features in observations to those in the forecasts to compare better dual-pol variable patterns; this, and other forecast evaluation methods for dual-pol fields, remains a promising area for future research endeavors. The methods used in this paper can be applied to storm-scale ensemble forecasts, such as the Storm Scale Ensemble Forecasts (SSEF) run as part of the NOAA Hazardous Weather Testbed Spring Experiment (e.g., Kong 2016), to evaluate similar issues over multiple cases. Putnam et al. (2016) represents the first effort in that direction. More studies evaluating and improving microphysics parameterizations and the dual-pol radar simulators are also needed (e.g., Johnson et al. 2016; Putnam et al. 2016).

Acknowledgements: This work was primarily supported by NSF grant AGS-1046171. The first and fifth author were also supported by NOAA grant NA11OAR4320072. The second and forth authors were also supported by NSF grant AGS-1261776. The third author is also partially supported by a research grant of “Development of a Polarimetric Radar Data Simulator for Local Forecasting Model” by the Korea Meteorological Administration. Computing was performed primarily on the Kraken system at the National Institute for Computational Sciences, part of the XSEDE resources. Critiques from Dr. Jason Milbrandt as well as two additional anonymous reviewers helped to improve the original manuscript.

References

Aksoy, A., D. C. Dowell, and C. Snyder, 2009: A multi-case comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part I: Storm-scale analysis. *Mon. Wea. Rev.*, **137,** 1805-1824.

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 II: Short-range ensemble forecasts. *Mon. Wea. Rev.*, **138,** 1273–1292.

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129,** 2884-2903.

Bringi, V. N. and V. Chandrasekar, 2001: *Polarimetric Doppler Weather Radar*. Cambridge, 636 pp.

Brown, B. G., 2001: Verification of Precipitation Forecasts: A Survey of Methodology Part II: Verification of Probability Forecasts at Points. National Center for Atmospheric Research.

Chandrasekar, V., S. Lim, N. Bharadwaj, W. Li, D. McLaughlin, V. N. Bringi, and E. Gorgucci, 2004: Principles of networked weather radar operation at attenuating frequencies. *Proc. Third European Conf. on Radar Meteorology and Hydrology*, 109-114.

Clark, A. J., W. A. Gallus, Jr., M. Xue, and F. Kong, 2009: A comparison of precipitation forecast skill between small convection-permitting and large convection-parameterizing ensembles. *Wea. Forecasting*, **24,** 1121-1140.

Davis, C., B. Brown, and R. Bullock, 2006: Object-Based Verification of Precipitation Forecasts. Part II: Application to Convective Rain Systems. *Mon. Wea. Rev.*, **134,** 1785-1795.

Dawson, D. T., II, M. Xue, and J. A. Milbrandt, 2009: High resolution real-data simulations of the 3 May 1999 tornadic storms with multi-moment microphysics**,** Under preparation.

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.

Dawson, D. T., II. , E. R. Mansell, Y. Jung, L. J. Wicker, M. R. Kumjian, and M. Xue, 2014: Low-level ZDR signatures in supercell forward flanks: The role of size sorting and melting of hail. *J. Atmos. Sci.*, **71,** 276-299.

Doviak, R. and D. Zrnic, 1993: *Doppler Radar and Weather Observations*. 2nd ed. Academic Press, 562 pp.

Dowell, D., 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.

Dowell, D. C. and L. J. Wicker, 2009: Additive noise for storm-scale ensemble data assimilation. *J. Atmos. Oceanic Technol.*, **26,** 911-927.

Dowell, D. C., L. J. Wicker, and C. Snyder, 2011: Ensemble Kalman filter assimilation of radar observations of the 8 May 2003 Oklahoma City supercell: Influence of reflectivity observations on storm-scale analysis. *Mon. Wea. Rev.*, **139,** 272–294.

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.

Ebert, E. E., 2008: Fuzzy verification of high-resolution gridded forecasts: A review and proposed framework. *Meteorological Applications*, **15,** 51-64.

Ebert, E. E. and J. L. McBride, 2000: Verification of precipitation in weather systems: determination of systematic errors. *Journal of Hydrology*, **239,** 179–202.

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.

Fritsch, J. M. and G. S. Forbes, 2001: Mesoscale convective systems. *Meteor. Monogr.*, **50,** 323-358.

Gaspari, G. and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125,** 723-757.

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.

Johnson, A., X. Wang, F. Kong, and M. Xue, 2013: Object-Based Evaluation of the Impact of Horizontal Grid Spacing on Convection-Allowing Forecasts. *Mon. Wea. Rev.*, **141,** 3413–3425.

Johnson, M., Y. Jung, D. Dawson, and M. Xue, 2016: Comparison of simulated polarimetric signatures in idealized supercell storms using two-moment bulk microphysics schemes in WRF. *Mon. Wea. Rev.*, **144,** 971-996.

Jung, Y., G. Zhang, and M. Xue, 2008: Assimilation of simulated polarimetric radar data for a convective storm using ensemble Kalman filter. Part I: Observation operators for reflectivity and polarimetric variables. *Mon. Wea. Rev.*, **136,** 2228-2245.

Jung, Y., M. Xue, and G. Zhang, 2010: Simulations of polarimetric radar signatures of supercell storm using a two-moment bulk microphysics scheme. *J. Appl. Meteor. Climatol.*, **49,** 146-163.

Jung, Y., M. Xue, and M. Tong, 2012: Ensemble Kalman filter analyses of the 29-30 May 2004 Oklahoma tornadic thunderstorm using one- and two-moment bulk microphysics schemes, with verification against polarimetric data. *Mon. Wea. Rev.*, **140,** 1457-1475.

Kalnay, E., 2002: *Atmospheric Modeling, Data Assimilation, and Predictability*. Cambridge University Press, 341 pp.

Kalnay, E., B. Hunt, E. Ott, and I. Szunyogh, 2006: Ensemble forecasting and data assimilation: two problems with the same solution? *Predictability of Weather and Climate*, T. Palmer and R. Hagedorn, Eds., Cambridge University Press, 734 pp.

Khain, A., A. Pokrovsky, M. Pinsky, A. Seifert, and V. Phillips, 2004: Simulation of Effects of Atmospheric Aerosols on Deep Turbulent Convective Clouds Using a Spectral Microphysics Mixed-Phase Cumulus Cloud Model. Part I: Model Description and Possible Applications. *J. Atmos. Sci.*, **61**.

Kong, F., 2013: 2013 CAPS Spring Forecast Experiment Program Plan, 24 pp.

Kong, F., 2016: 2016 CAPS spring experiment forecast experiment program plan, 29 pp.

Kumjian, M. R. and A. V. Ryzhkov, 2008: Polarimetric signatures in supercell thunderstorms. *J. Appl. Meteor. Climatol.*, **47,** 1940-1961.

Kumjian, M. R. and A. V. Ryzhkov, 2012: The Impact of Size Sorting on the Polarimetric Radar Variables. *J. Atmos. Sci.*, **69,** 2042–2060.

Lei, T., M. Xue, and T. Yu, 2009: Multi-scale analysis and prediction of the 8 May 2003 Oklahoma City tornadic supercell storm assimilating radar and surface network data using EnKF. *Extended abstract, 13th Conf. of IOAS-AOLS, AMS Meetings 2008*, Paper 6.4.

Li, X. and J. R. Mecikalski, 2012: Impact of the Dual-Polarization Doppler Radar Data on Two Convective Storms with a Warm-Rain Radar Forward Operator. *Mon. Wea. Rev.*, **140,** 2147–2167.

Lim, K.-S. S. and S.-Y. Hong, 2010: Development of an Effective Double-Moment Cloud Microphysics Scheme with Prognostic Cloud Condensation Nuclei (CCN) for Weather and Climate Models. *Mon. Wea. Rev.*, **138,** 1587-1612.

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: Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.*, **26,** 636-646.

Mason, I. B., 1982: A model for the assessment of weather forecasts. *Aust. Meteor. Mag.*, **30,** 291-303.

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.

Milbrandt, J. A. and M. K. Yau, 2005a: A multi-moment bulk microphysics parameterization. Part I: Aanlysis of the role of the spectral shape parameter. *J. Atmos. Sci.*, **62,** 3051-3064.

Milbrandt, J. A. and M. K. Yau, 2005b: A multi-moment bulk microphysics parameterization. Part II: A proposed three-moment closure and scheme description. *J. Atmos. Sci.*, **62,** 3065-3081.

Milbrandt, J. A. and R. McTaggart-Cowan, 2008: An efficient semi-double-moment bulk microphysics scheme. *15th International Conference on Clouds and Precipitation*, International Commission on Clouds and Precipitation (ICCP), P01.

Mittermaier, M., N. Roberts, and S. A. Thompson, 2013: A long-term assessment of precipitation forecast skill using the Fractions Skill Score. *Meteor. Appl.*, **20,** 176–186.

Morrison, H. and A. Gettelman, 2008: A new two-moment bulk stratiform cloud microphysics scheme in the community atmosphere model, version 3 (CAM3). Part 1: Description and numerical tests. *J. Climate*, **21,** 3642-3659.

Morrison, H., J. A. Curry, M. D. Shupe, and P. Zuidema, 2005: A new double-moment microphysics parameterization for application in cloud and climate models. Part II: Single-column modeling of arctic clouds. *J. Atmos. Sci.*, **62,** 1678-1693.

NWS, cited 2012: Storm Data and Unusual Weather Phenomena - May 2007. [Available online from http://www.srh.noaa.gov/media/oun/stormdata/oun200705.pdf.]

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.

Putnam, B. J., M. Xue, Y. Jung, N. Snook, and G. Zhang, 2014: The Analysis and Prediction of Microphysical States and Polarimetric Radar Variables in a Mesoscale Convective System Using Double-Moment Microphysics, Multinetwork Radar Data, and the Ensemble Kalman Filter. *Mon. Wea. Rev.*, **142,** 141–162.

Putnam, B. J., M. Xue, Y. Jung, G. Zhang, and F. Kong, 2016: Simulation of polarimetric radar variables from 2013 CAPS spring experiment storm scale ensemble forecasts and evaluation of microphysics schemes. *Mon. Wea. Rev.***,** accepted.

Roberts, N., 2008: Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model. *Meteor. Appl.*, **15,** 163-169.

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.

Rossa, A., P. Nurmi, and E. Ebert, 2008: Overview of methods for the verification of quantitative precipitation forecasts. *Precipitation: Advances in Measurement, Estimation and Prediction,* S. Michaelides, Ed., Springer, 540pp.

Ryzhkov, A. and D. S. Zrnic, 1996: Assessment of rainfall measurement that uses specific differential phase. *J. Appl. Meteor.*, **35,** 2080–2090.

Ryzhkov, A. V., S. E. Giangrande, V. M. Melnikov, and T. J. Schuur, 2005: Calibration issues of dual-polarization radar measurements. *J. Atmos. Ocean. Tech.*, **22,** 1138-1155.

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.

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 based on ensemble Kalman filter assimilation of CASA X-band and WSR-88D radar data. *Mon. Wea. Rev.*, **139,** 3446-3468.

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

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.

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.

Stensrud, D. J., L. J. Wicker, M. Xue, D. Dawson, N. Yussouf, D. Wheatley, T. E. Thompson, N. A. Snook, T. M. Smith, A. D. Schenkman, C. K. Potvin, E. R. Mansell, T. Lei, K. M. Kuhlman, Y. Jung, T. A. Jones, J. Gao, M. C. Coniglio, H. E. Brooks, and K. A. Brewster, 2013: Progress and challenges with Warn-on-Forecast. *Atmos. Res.*, **123,** 2-16.

Tanamachi, R. L., L. J. Wicker, D. C. Dowell, H. B. Bluestein, D. T. Dawson, II., and M. Xue, 2013: EnKF Assimilation of High-Resolution, Mobile Doppler Radar Data of the 4 May 2007 Greensburg, Kansas, Supercell into a Numerical Cloud Model. *Mon. Wea Rev.*, **141,** 625–648.

Thompson, G., R. M. Rasmussen, and K. Manning, 2004: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part I: Description and sensitivity analysis. *Mon. Wea. Rev.*, **132,** 519-542.

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

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

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.

Ulbrich, C. W., 1983: Natural variations in the analytical form of the raindrop size distributions. *J. Appl. Meteor.*, **22,** 1764-1775.

Vivekanandan, J., W. M. Adams, and V. N. Bringi, 1991: Rigorous Approach to Polarimetric Radar Modeling of Hydrometeor Orientation Distributions. *J. Appl. Meteor.*, **30,** 1053–1063.

Wheatley, D. M., N. Yussouf, and D. J. Stensrud, 2014: Ensemble Kalman Filter Analyses and Forecasts of a Severe Mesoscale Convective System Using Different Choices of Microphysics Schemes. *Monthly Weather Review*, **142,** 3243-3263.

Whitaker, J. S. and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130,** 1913-1924.

Xue, M., K. K. Droegemeier, and V. Wong, 2000: The Advanced Regional Prediction System (ARPS) - A multiscale nonhydrostatic atmospheric simulation and prediction tool. Part I: Model dynamics and verification. *Meteor. Atmos. Physics,* **75,** 161-193.

Xue, M., M. Tong, and K. K. Droegemeier, 2006: An OSSE framework based on the ensemble square-root Kalman filter for evaluating impact of data from radar networks on thunderstorm analysis and forecast. *J. Atmos. Ocean Tech.*, **23,** 46–66.

Xue, M., Y. Jung, and G. Zhang, 2010: State estimation of convective storms with a two-moment microphysics scheme and an ensemble Kalman filter: Experiments with simulated radar data *Q. J. Roy. Meteor. Soc*, **136,** 685-700.

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., K. K. Droegemeier, V. Wong, A. Shapiro, K. Brewster, F. Carr, D. Weber, Y. Liu, and D.-H. Wang, 2001: The Advanced Regional Prediction System (ARPS) - A multiscale nonhydrostatic atmospheric simulation and prediction tool. Part II: Model physics and applications. *Meteor. Atmos. Phy.*, **76,** 143-165.

Yussouf, N., J. Kain, and A. Clark, 2016: Short-term Probabilistic Forecasts of the 31 May 2013 Oklahoma Tornado and Flash Flood Event Using a Continuous-Update-Cycle Storm-scale Ensemble System. *Wea. Forecasting***,** in press.

Yussouf, N., E. R. Mansell, L. J. Wicker, D. M. Wheatley, and D. J. Stensrud, 2013: The Ensemble Kalman Filter Analyses and Forecasts of the 8 May 2003 Oklahoma City Tornado Supercell Storm using Single and Double Moment Microphysics Schemes. *Mon. Wea. Rev.*, **141,** 3388-3412.

Yussouf, N., D. C. Dowell, L. J. Wicker, K. H. 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.

Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observations on the convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132,** 1238-1253.

Zhu, K. F., Y. Yang, and M. Xue, 2015: Percentile-based neighborhood precipitation verification and its application to a landfalling tropical storm case with radar data assimilation. *Adv. Atmos. Sci.*, **32,** 1449-1459.

List of figures

Fig. 1. Mosaic of observed reflectivity (dBZ) from KAMA, KDYX, KFWS, KLBB, KTLX, and KVNX at 0200 UTC at about 2 km above ground. The locations of all radars assimilated are marked. Also, notable MCS features including the line end vortex (LEV), leading convective line, leading stratiform region, and trailing stratiform region are given. Reproduced from Putnam et al. (2014).

Fig. 2. Mosaics of observed reflectivity (dBZ) as in Fig. 1 from (a-c) 0300 UTC to 0500 UTC as well as probability matched ensemble mean reflectivity for (d) EXP_S and (g) EXP_D at 0300 UTC/1-h forecast, (e,h) 0400 UTC/2-h forecast, and (f,i) 0500 UTC/3-h forecast.

Fig. 3. Probability of reflectivity exceeding 20 dBZ at 2 km AGL for EXP_S at (a) 1-h, (b) 2-h, and (c) 3-h forecast times and (d-f) EXP_D. The thick black line outlines observed reflectivity exceeding 20 dBZ.

Fig. 4. Probability of reflectivity exceeding 40 dBZ at 2 km AGL for EXP_S at (a) 1-h, (b) 2-h, and (c) 3-h forecast times and (d-f) EXP_D. The thick black line outlines observed reflectivity exceeding 40 dBZ.

Fig. 5. Area under the relative operating characteristic curve (AUC) for EXP_S (red line and shading) and EXP_D (blue line and shading) at (a) 1-h, (b) 2-h, and (c) 3-h forecast times at 2 km AGL for the full experiment domain and also (d-f) a subdomain covering Oklahoma.

Fig. 6. Reliability diagrams calculated for reflectivity exceeding 20 dBZ for EXP_S (red line) and EXP_D (blue line) at (a) 1-h, (b) 2-h, and (c) 3-h forecast times at 2 km AGL for the full experiment domain and also (d-f) a subdomain covering Oklahoma.

Fig. 7. Sharpness diagrams calculated for reflectivity exceeding 20 dBZ for EXP_S (red) at (a) 1-h, (b) 2-h, and (c) 3-h forecast times and (d-f) EXP_D (blue) at 2 km AGL for the full experiment domain and also (g-l) a subdomain covering Oklahoma.

Fig. 8. (a) Observed reflectivity (dBZ) and simulated reflectivity from (b) EXP_S and (c) EXP_D at 0300 UTC/1-h forecast at a 0.5° tilt from KOUN, as well as (d-f) differential reflectivity (dB) and (g-i) specific differential phase (°km^{-1}).

Fig. 9. As in Fig. 8 but at 0400 UTC with 2-h forecast results.

Fig. 10. As in Fig. 8 but at 0500 UTC with 3-h forecast results.

Fig. 11. (a) Observed reflectivity (dBZ) and simulated reflectivity from (b) EXP_S member 14 and (c) EXP_D member 39 at 0400 UTC/2-h forecast at a 0.5° tilt from KOUN, as well as observed (d) and simulated (e-f) differential reflectivity (dB).

Fig. 12. Probability of differential reflectivity exceeding 2.3 dB at a 0.5° tilt from KOUN for EXP_S at (a) 1-h, (b) 2-h, and (c) 3-h forecast times and for (d-f) EXP_D.

Fig. 13. Area under the relative operating characteristic curve (AUC) for differential reflectivity (dB) for EXP_S (red line and shading) and EXP_D (blue line and shading) at (a) 1-h, (b) 2-h, and (c) 3-h forecast times at a 0.5° tilt as well as for (d-f) specific differential phase (° km^{-1}).

Fig. 14. Histograms of observed (black) KOUN and simulated reflectivity (dBZ) values from EXP_S (red) and EXP_D (blue) at (a) 0300, (b) 0400, and (c) 0500 UTC at a 0.5° tilt as well as (d-f) observed and simulated differential reflectivity (dB) values and (g-i) observed and simulated specific differential phase (° km^{-1}) values. The values in EXP_S and EXP_D are normalized by the size of each ensemble.

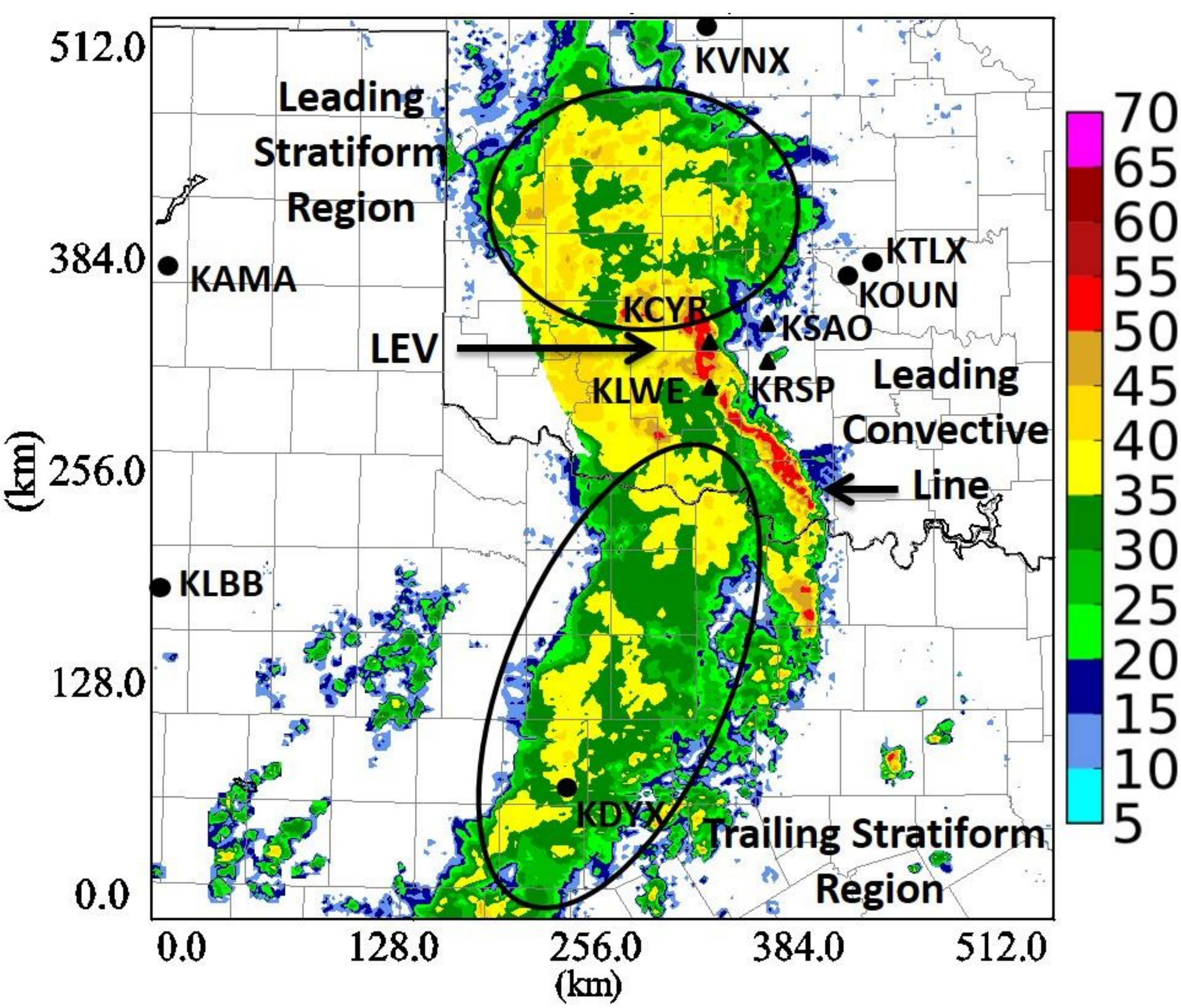

Fig. 1. Mosaic of observed reflectivity (dBZ) from KAMA, KDYX, KFWS, KLBB, KTLX, and KVNX at 0200 UTC at about 2 km above ground. The locations of all radars assimilated are marked. Also, notable MCS features including the line end vortex (LEV), leading convective line, leading stratiform region, and trailing stratiform region are given. Reproduced from Putnam et al. (2014).

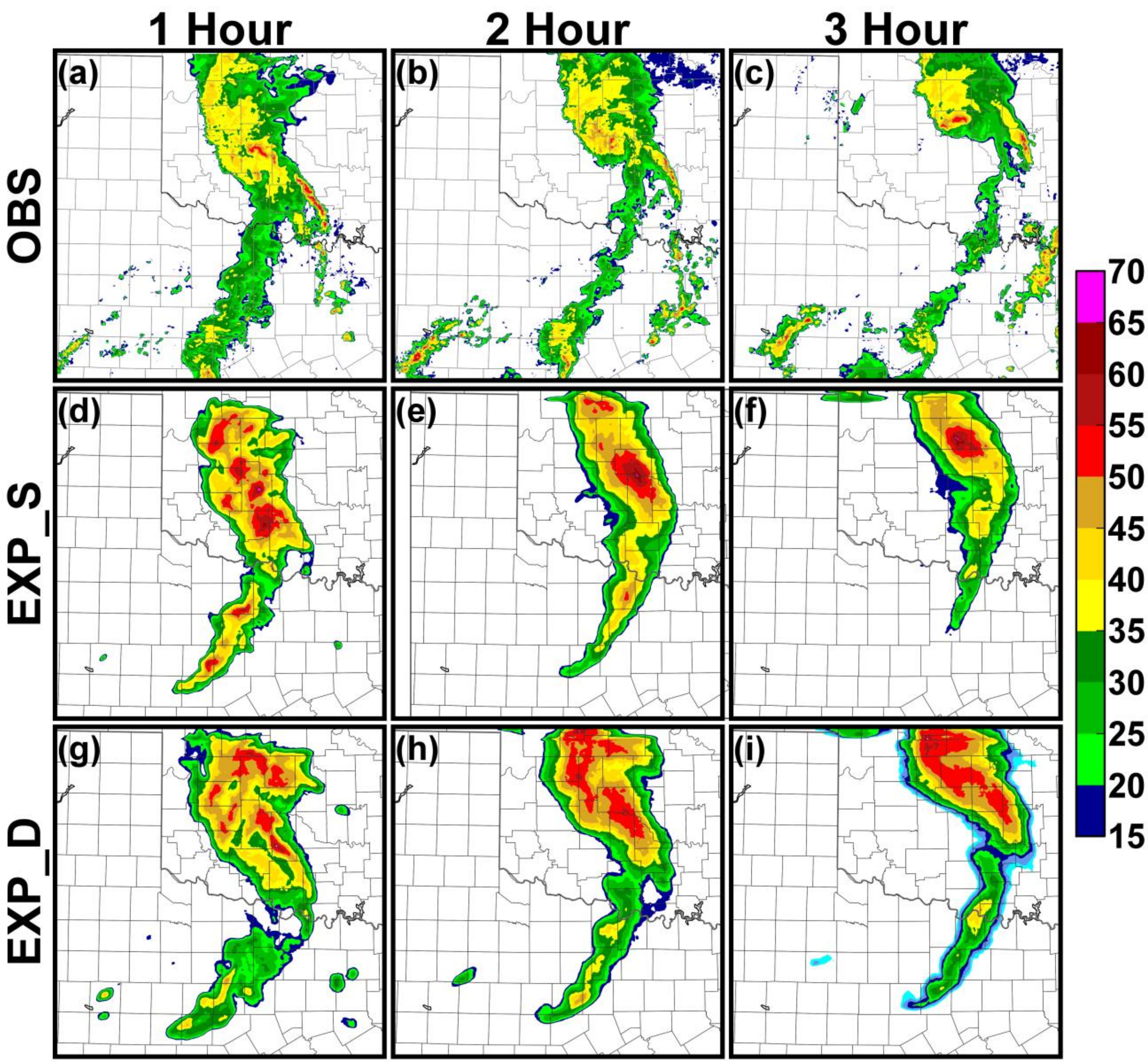

Fig. 2. Mosaics of observed reflectivity (dBZ) as in Fig. 1 from (a-c) 0300 UTC to 0500 UTC as well as probability matched ensemble mean reflectivity for (d) EXP_S and (g) EXP_D at 0300 UTC/1-h forecast, (e,h) 0400 UTC/2-h forecast, and (f,i) 0500 UTC/3-h forecast.

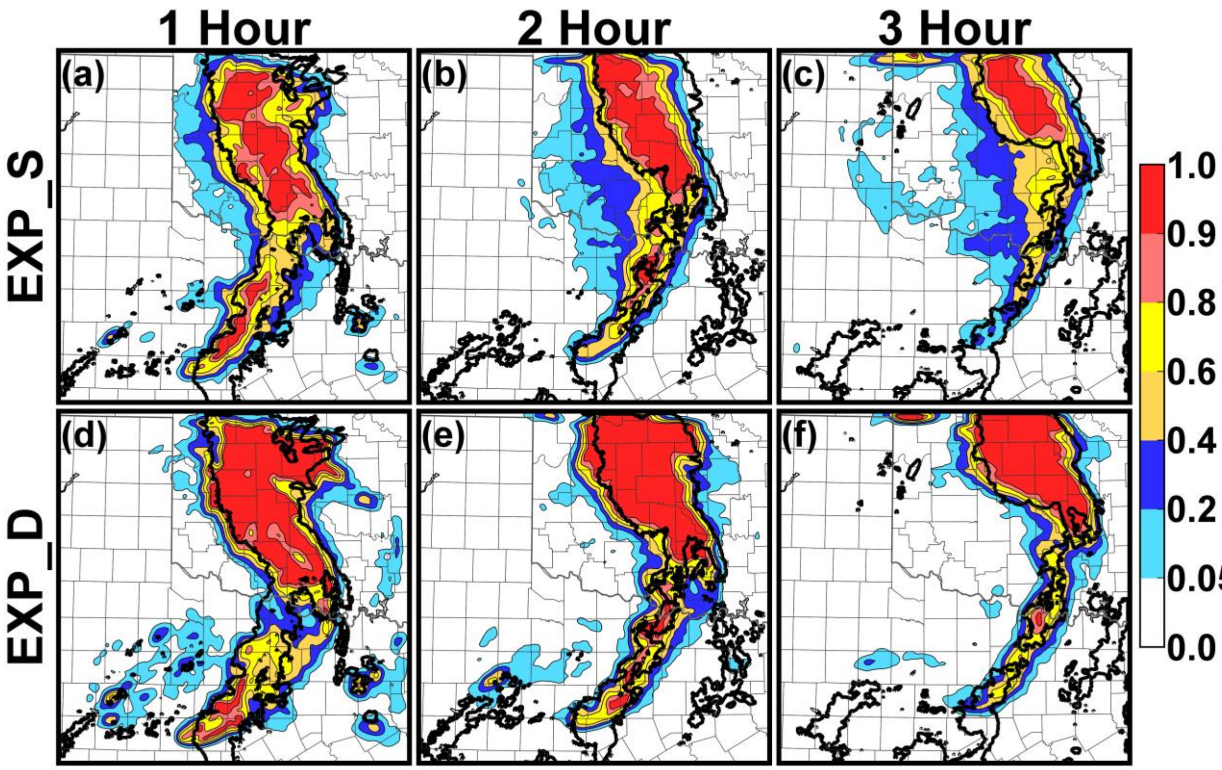

Fig. 3. Probability of reflectivity exceeding 20 dBZ at 2 km AGL for EXP_S at (a) 1-h, (b) 2-h, and (c) 3-h forecast times and (d-f) EXP_D. The thick black line outlines observed reflectivity exceeding 20 dBZ.

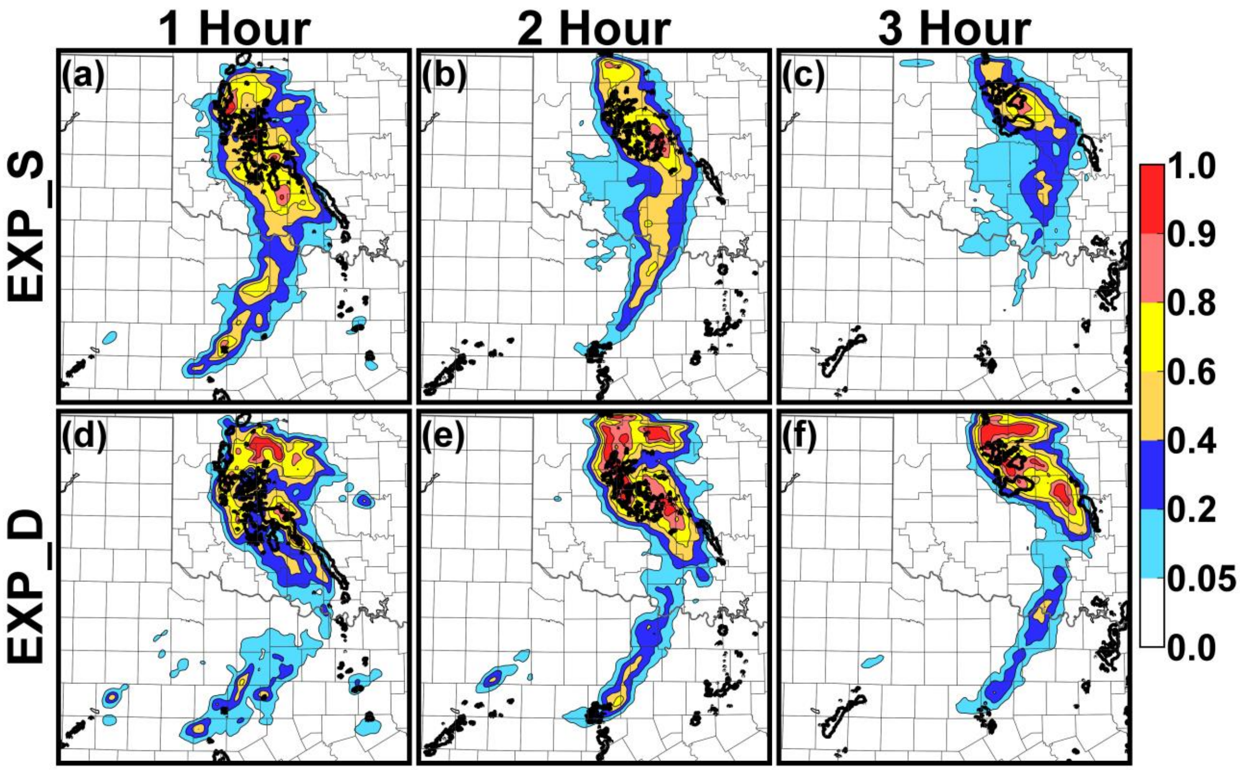

Fig. 4. Probability of reflectivity exceeding 40 dBZ at 2km AGL for EXP_S at (a) 1-h, (b) 2-h, and (c) 3-h forecast times and (d-f) EXP_D. The thick black line outlines observed reflectivity exceeding 40 dBZ.

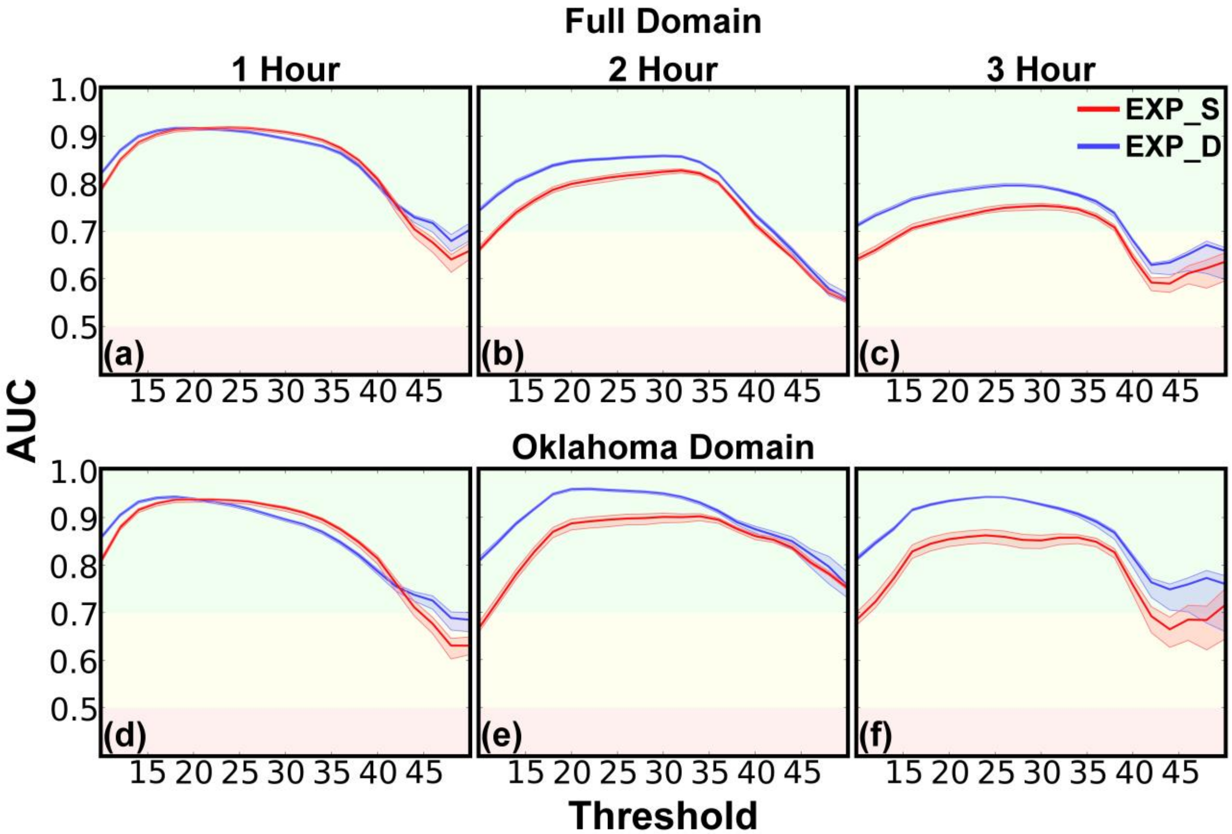

Fig. 5. Area under the relative operating characteristic curve (AUC) for EXP_S (red line and shading) and EXP_D (blue line and shading) at (a) 1-h, (b) 2-h, and (c) 3-h forecast times at 2 km AGL for the full experiment domain and also (d-f) a subdomain covering Oklahoma.

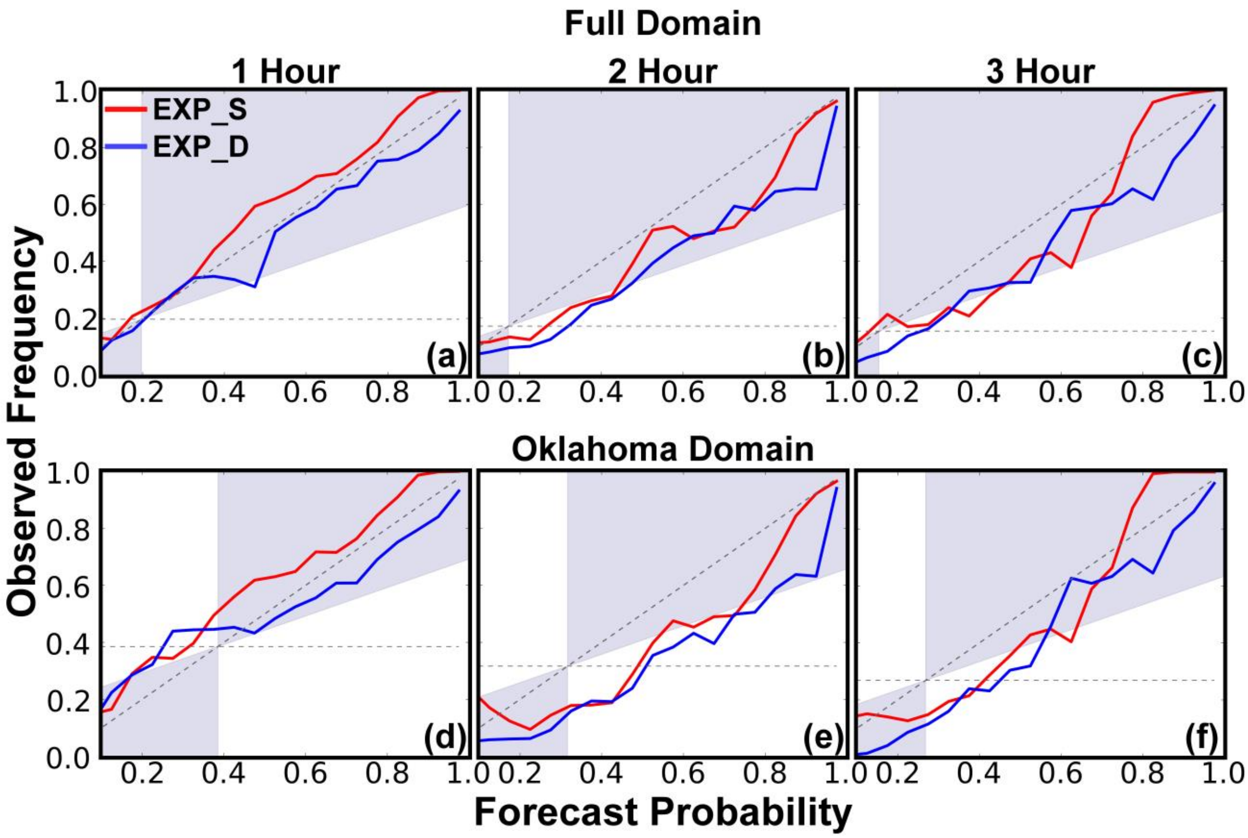

Fig. 6. Reliability diagrams calculated for reflectivity exceeding 20 dBZ for EXP_S (red line) and EXP_D (blue line) at (a) 1-h, (b) 2-h, and (c) 3-h forecast times at 2 km AGL for the full experiment domain and also (d-f) a subdomain covering Oklahoma.

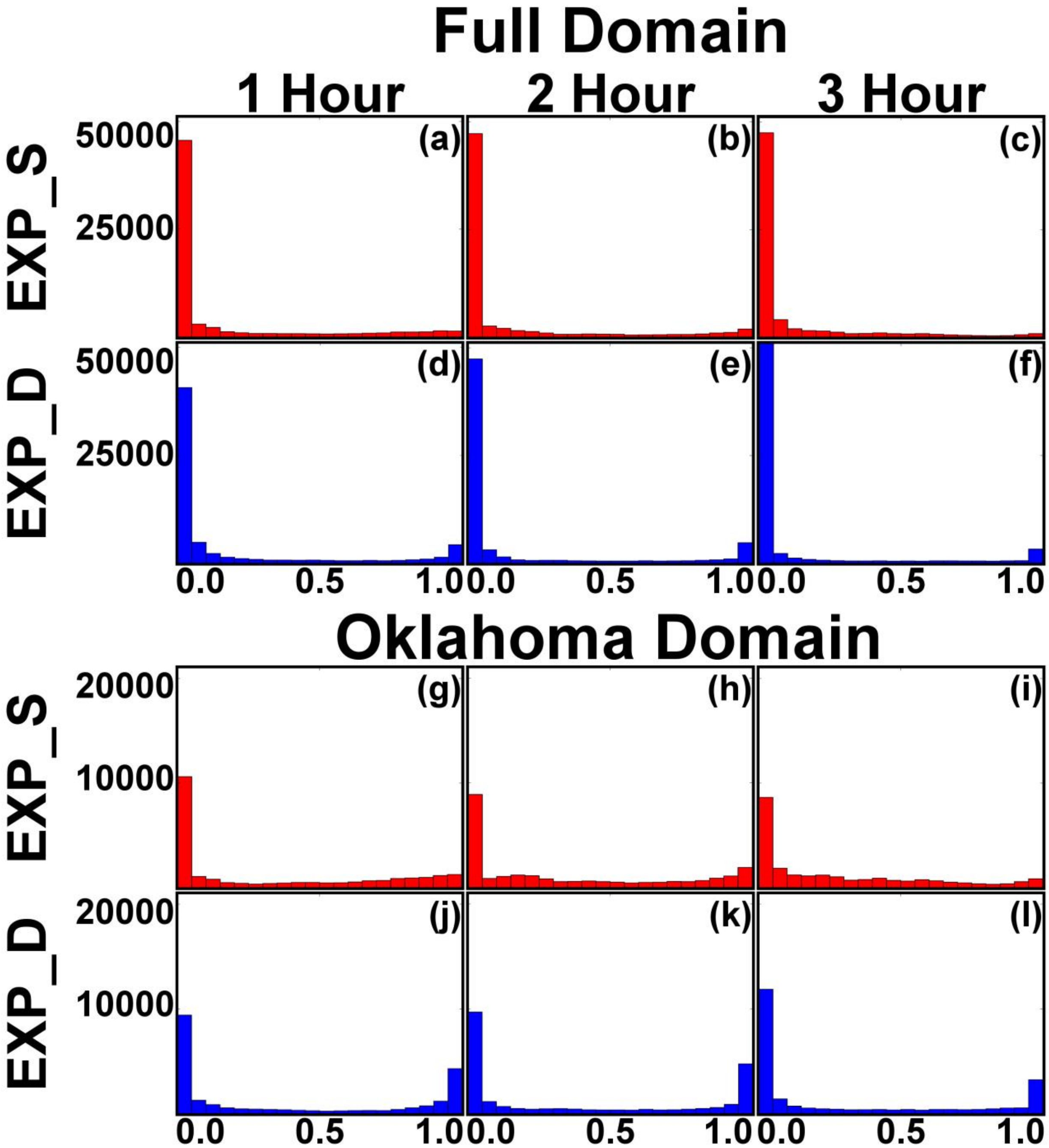

Fig. 7. Sharpness diagrams calculated for reflectivity exceeding 20 dBZ for EXP_S (red) at (a) 1-h, (b) 2-h, and (c) 3-h forecast times and (d-f) EXP_D (blue) at 2 km AGL for the full experiment domain and also (g-l) a subdomain covering Oklahoma.

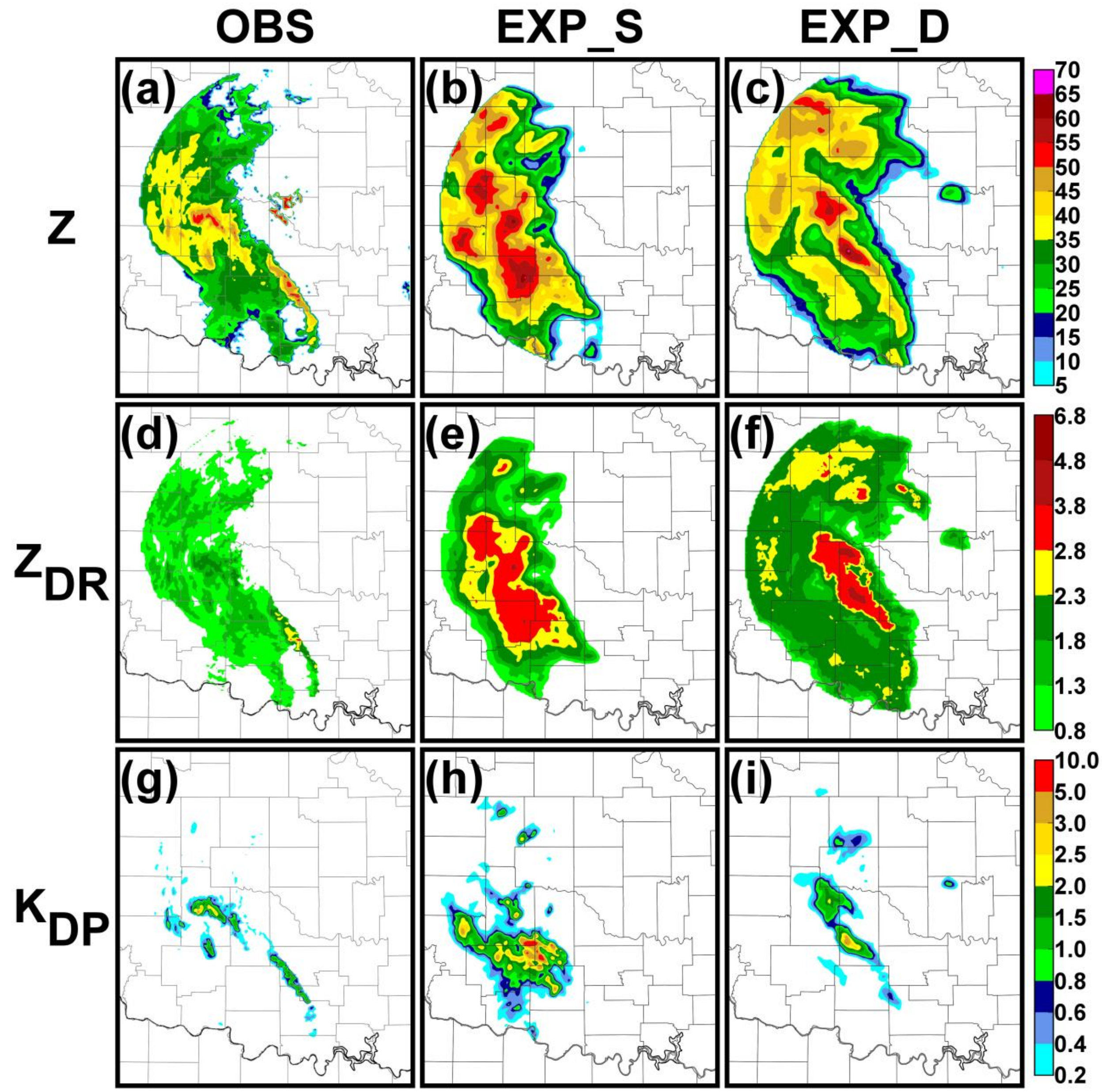

Fig. 8. (a) Observed reflectivity (dBZ) and simulated reflectivity from (b) EXP_S and (c) EXP_D at 0300 UTC/1-h forecast at a 0.5° tilt from KOUN, as well as (d-f) differential reflectivity (dB) and (g-i) specific differential phase (° km^{-1}).

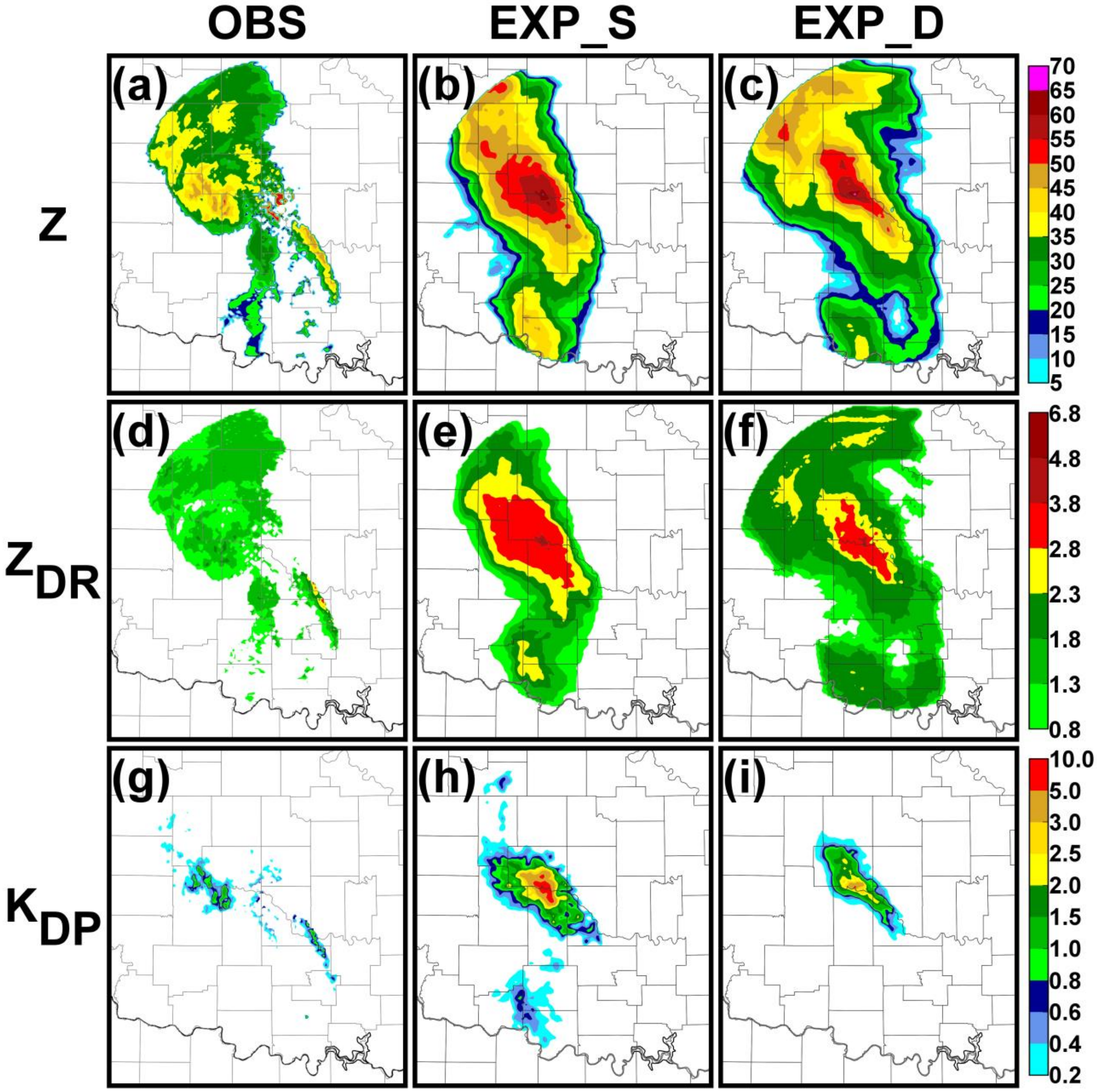

Fig. 9. As in Fig. 8 but at 0400 UTC with 2-h forecast results.

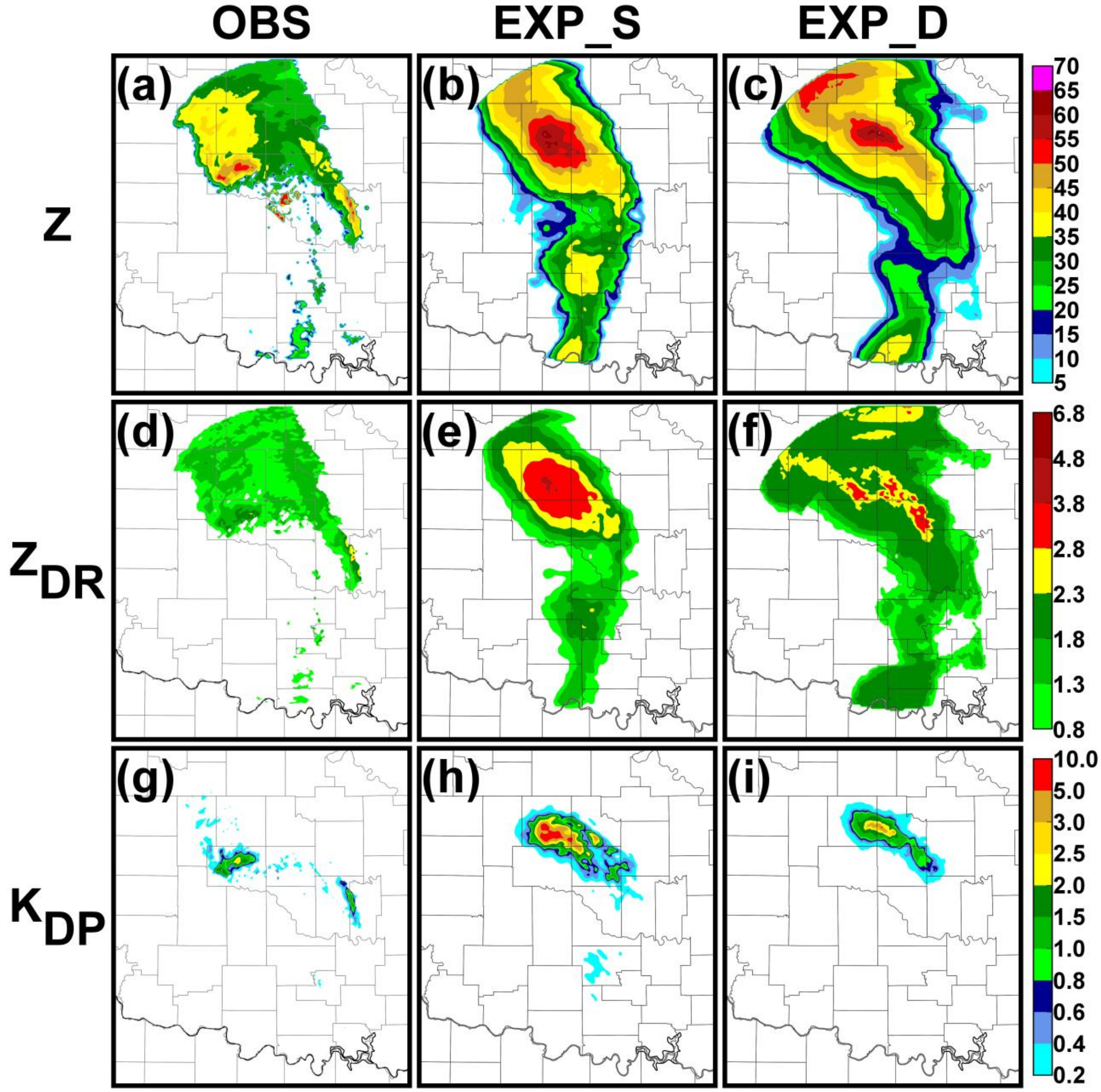

Fig. 10. As in Fig. 8 but at 0500 UTC with 3-h forecast results.

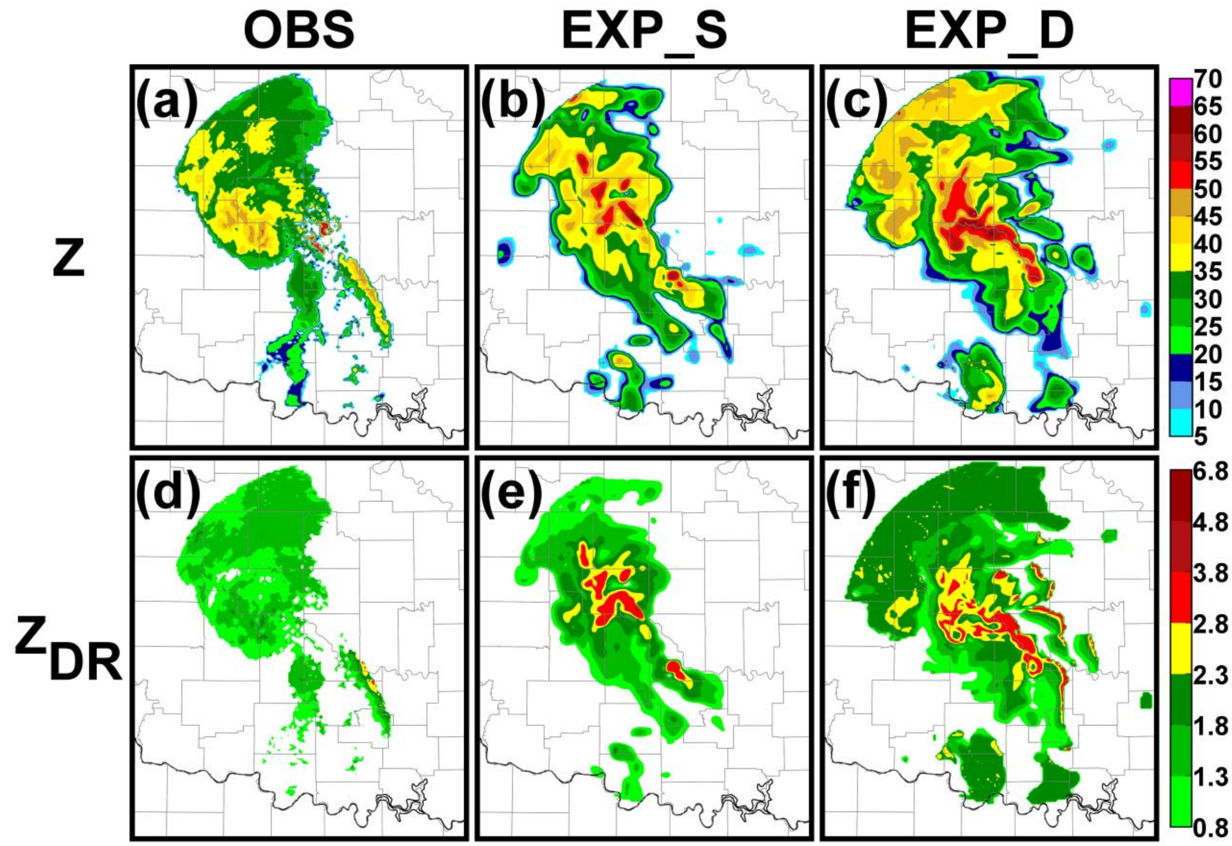

Fig. 11. (a) Observed reflectivity (dBZ) and simulated reflectivity from (b) EXP_S member 14 and (c) EXP_D member 39 at 0400 UTC/2-h forecast at a 0.5° tilt from KOUN, as well as observed (d) and simulated (e-f) differential reflectivity (dB).

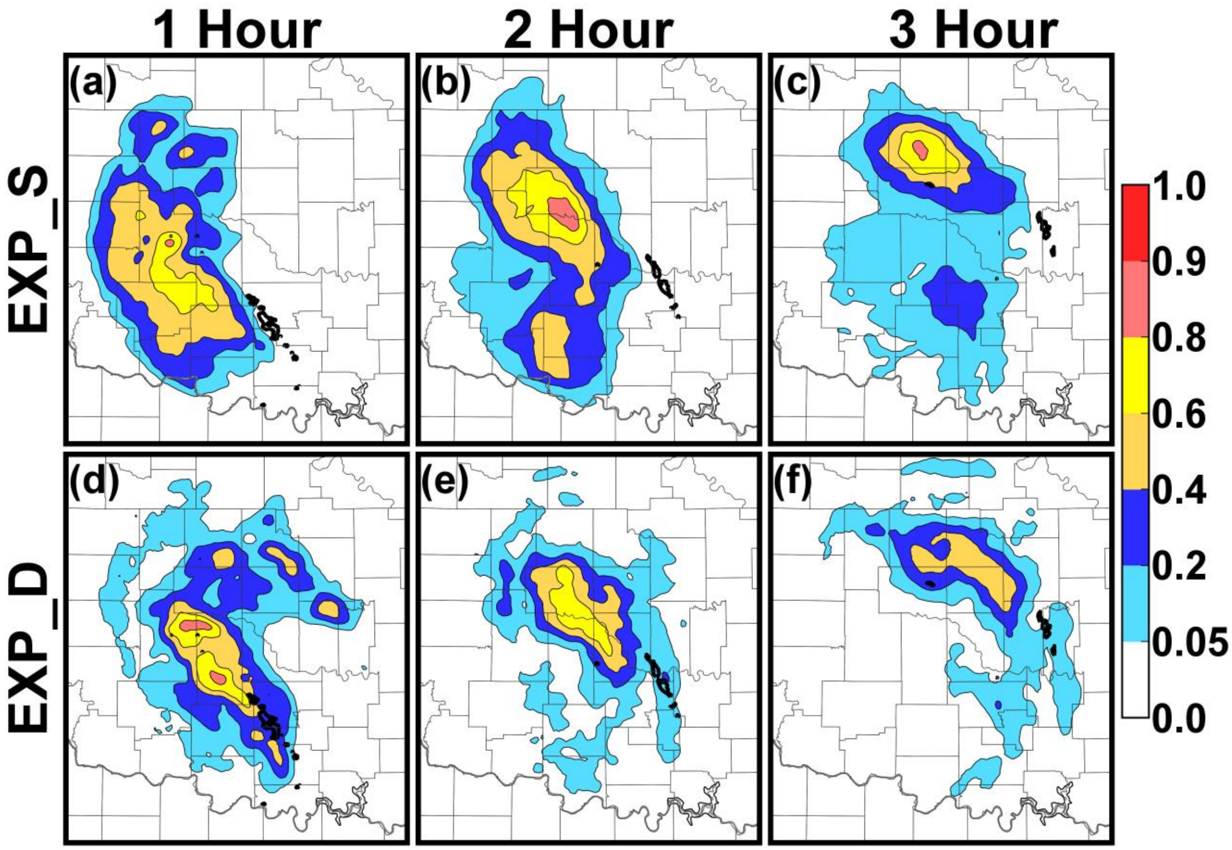

Fig. 12. Probability of differential reflectivity exceeding 2.3 dB at a 0.5° tilt from KOUN for EXP_S at (a) 1-h, (b) 2-h, and (c) 3-h forecast times and for (d-f) EXP_D.

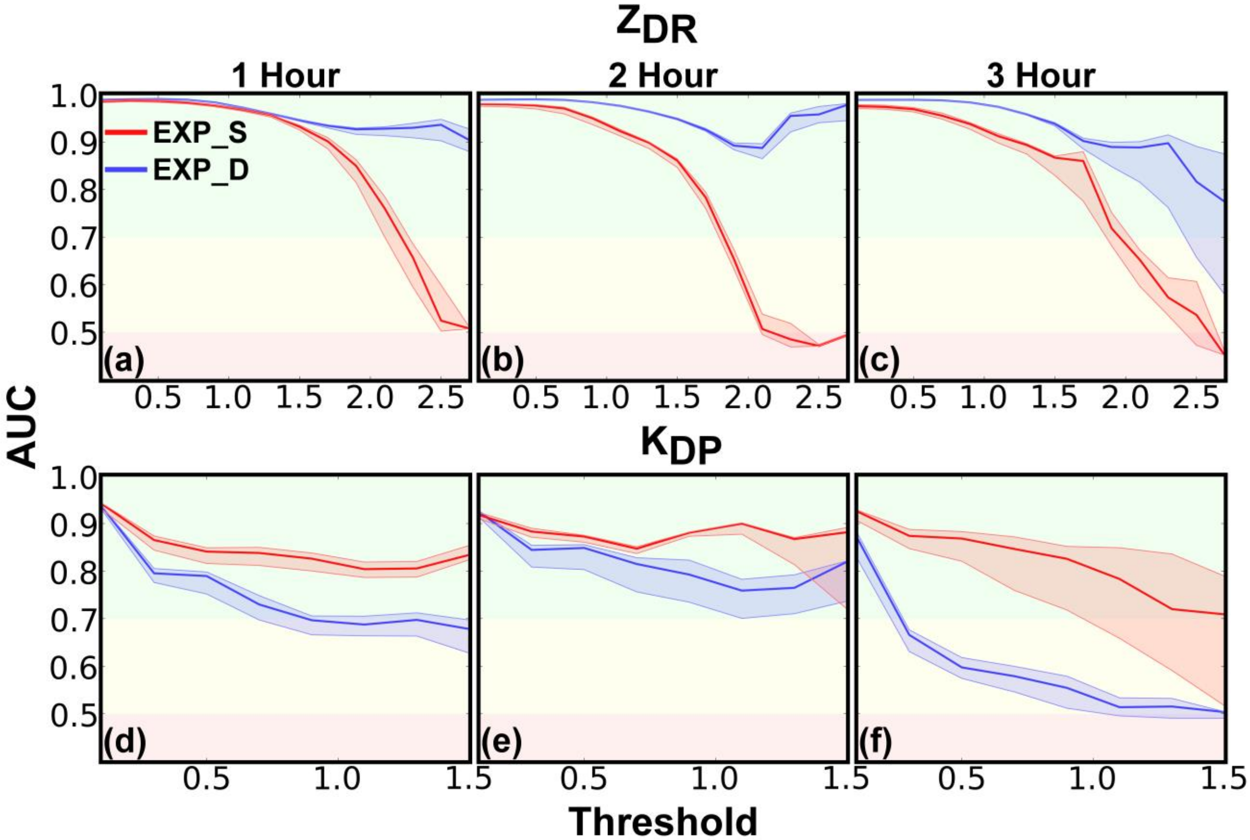

Fig. 13. Area under the relative operating characteristic curve (AUC) for differential reflectivity (dB) for EXP_S (red line and shading) and EXP_D (blue line and shading) at (a) 1-h, (b) 2-h, and (c) 3-h forecast times at a 0.5° tilt as well as for (d-f) specific differential phase (° km^{-1}).

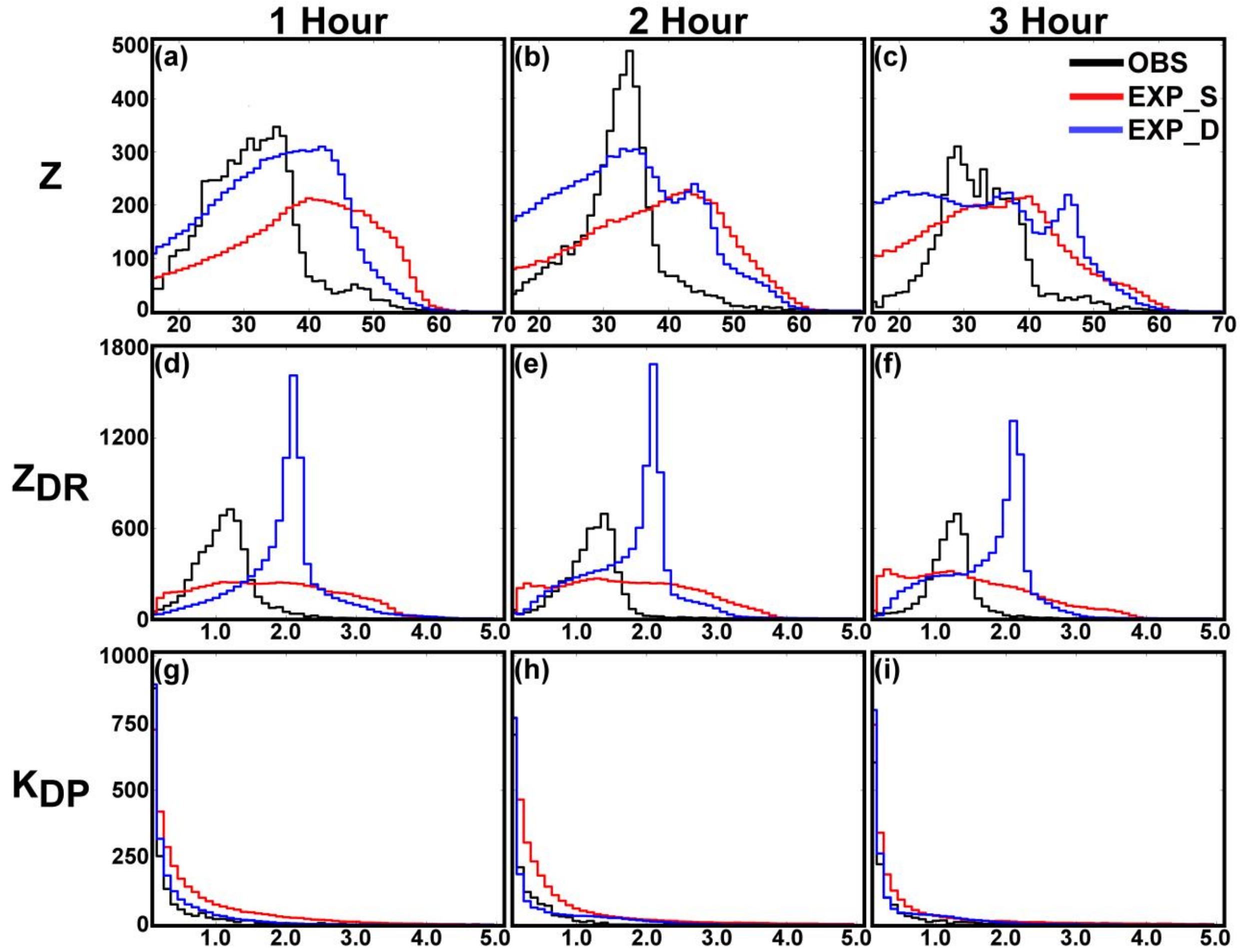

Fig. 14. Histograms of observed (black) KOUN and simulated reflectivity (dBZ) values from EXP_S (red) and EXP_D (blue) at (a) 0300, (b) 0400, and (c) 0500 UTC at a 0.5° tilt as well as (d-f) observed and simulated differential reflectivity (dB) values and (g-i) observed and simulated specific differential phase (° km^{-1}) values. The values in EXP_S and EXP_D are normalized by the size of each ensemble.