

# **Multi-Scale EnKF Assimilation of Radar and Conventional Observations and Ensemble Forecasting for a Tornadic Mesoscale Convective System**

Nathan Snook<sup>1</sup>, Ming Xue<sup>1,2</sup>, and Youngsun Jung<sup>1</sup>

Center for Analysis and Prediction of Storms<sup>1</sup> and School of Meteorology<sup>2</sup>  
University of Oklahoma, Norman OK 73072

Submitted to Monthly Weather Review

August 2013  
Revised December 2013

Corresponding author address:  
Nathan Snook  
Center for Analysis and Prediction of Storms  
University of Oklahoma,  
120 David Boren Blvd, Room 4230, Norman OK 73072  
nsnook@ou.edu

## **Abstract**

In recent studies, the authors have successfully demonstrated the ability of an ensemble Kalman filter (EnKF), assimilating real radar observations, to produce skillful analyses and subsequent ensemble-based probabilistic forecasts for a tornadic mesoscale convective system (MCS) that occurred over Oklahoma and Texas on 9 May 2007. The current study expands upon this prior work, performing experiments for this case on a larger domain using a nested-grid EnKF which accounts for mesoscale uncertainties through the initial ensemble and lateral boundary condition perturbations. In these new experiments, conventional observations (including surface, wind profiler, and upper-air observations) are assimilated in addition to the WSR-88D and CASA radar data used in the previous studies, better representing meso- and convective-scale features. The relative impacts of conventional and radar data on analyses and forecasts are examined, and biases within the ensemble are investigated.

Compared to prior results, the radar-assimilating experiments accounting for mesoscale uncertainties produce superior forecasts based on both subjective and objective verification metrics. The new experiments produce a substantially-improved forecast, including better representation of the convective lines of the MCS. Assimilation of radar data substantially improves the ensemble precipitation forecast. Assimilation of conventional data together with radar observations substantially improves the forecast of near-surface mesovortices within the MCS, improves forecasts of surface temperature and dewpoint, and imparts a slight but noticeable improvement to short-term precipitation forecasts. Furthermore, ensemble analyses and forecasts are found to be sensitive to the localization radius applied to conventional data within the EnKF.

## 1 **1. Introduction**

2           The ensemble Kalman filter (EnKF), first developed by Evensen (1994, 2003), has been  
3 successfully applied to atmospheric data assimilation (DA) using both simulated and real data from  
4 a variety of observation platforms, for models ranging from global to convective storm scales  
5 (Houtekamer and Mitchell 1998; Hamill and Snyder 2000; Anderson 2001; Whitaker and Hamill  
6 2002; Snyder and Zhang 2003; Dowell et al. 2004; Zhang et al. 2004; Dirren et al. 2007; Tong and  
7 Xue 2008a; Xue et al. 2010; Dawson et al. 2011; Snook et al. 2011; Jung et al. 2012; Yussouf and  
8 Stensrud 2012; Yussouf et al. 2013). Though EnKF is rather expensive in terms of computation,  
9 requiring an ensemble of forecasts (typically using several dozen members), it provides flow-  
10 dependent multivariate background error covariances that less computationally-intensive 3-  
11 dimensional variational (3DVAR) methods cannot. Cross-covariances produced by the EnKF  
12 system are very valuable, especially for convective-scale DA, because state variables that are not  
13 directly observed can be retrieved (Tong and Xue 2005, 2008a). Further discussion of DA  
14 techniques commonly used for assimilation of weather observations, including 3DVAR, 4-  
15 dimensional variational methods (4DVAR), and EnKF can be found in Tong and Xue (2005).

16 Analysis ensembles generated using EnKF are generally well-suited as initial conditions for  
17 convective-scale ensemble forecasts. EnKF assimilation of Doppler radar data has proven to be  
18 effective in retrieving wind, temperature, and microphysical fields at the convective scale (e.g.,  
19 Dowell et al. 2004; Tong 2006; Snook et al. 2011; Jung et al. 2012; Putnam et al. 2013).

20 Furthermore, EnKF analyses, in principle, also characterize the analysis uncertainty; this is a  
21 particularly desirable quality in the ensemble forecast initial conditions. Forecast ensembles  
22 initialized from EnKF analyses have been shown to produce superior probabilistic predictions

23 compared to ensembles initialized using traditional perturbation methods (Houtekamer et al. 2005;  
24 Hamill and Whitaker 2010). EnKF analyses have been successfully applied to ensemble forecasts  
25 of convective systems, including supercell thunderstorms (e.g. Aksoy et al. 2009; Aksoy et al. 2010;  
26 Dawson et al. 2011) and mesoscale convective systems (e.g. Snook et al. 2012; Putnam et al. 2013),  
27 as well as tropical cyclones (e.g. Wu et al. 2010; Aksoy et al. 2012; Aksoy et al. 2013). As available  
28 computational power increases, it will become increasingly feasible to run a real-time convective-  
29 scale ensemble forecast system (e.g., Xue et al. 2008) incorporating EnKF DA (e.g. Snook et al.  
30 2012), as envisioned in the “warn-on-forecast” paradigm being developed by the National Weather  
31 Service (Stensrud et al. 2009). In Snook et al. (2011, hereafter SXJ11), an ensemble square-  
32 root Kalman filter (EnSRF) (Whitaker and Hamill 2002) is used together with the Advanced  
33 Regional Prediction System (ARPS; Xue et al. 2000, 2001) atmospheric model (Tong and Xue  
34 2005; Xue et al. 2006; Tong and Xue 2008b) to assimilate radar reflectivity and radial velocity  
35 observations from multiple WSR-88D (Crum et al. 1993) S-band radars, and from the X-band  
36 radars deployed by the Center for Collaborative Adaptive Sensing of the Atmosphere (CASA)  
37 (McLaughlin et al. 2009), for a tornadic mesoscale convective system (MCS) that occurred over  
38 Texas and Oklahoma on 9 May 2007. The 40-member ensemble mean analysis of SXJ11 produces  
39 model storms whose geographic extent, convective mode, and intensity agree well with the radar  
40 observations. Furthermore, SXJ11 finds that assimilation of CASA radar data improves the  
41 representation of near-surface circulations and cold pool structure. Ensemble forecasts initialized  
42 from the ensemble analyses of SXJ11 are subsequently examined in Snook et al. (2012, hereafter  
43 SXJ12). The forecast ensembles of SXJ12 produce skillful 0-3 hour probabilistic forecasts for radar  
44 reflectivity and 2-hour probabilistic forecasts of the presence and location of the tornadic  
45 mesovortex embedded within the MCS with probability maxima localized within several tens of

46 kilometers of the observed tornadic mesovortex. The EnKF experiments of SXJ11 and SXJ12 did  
47 not assimilate any conventional observations, nor did they include any mesoscale perturbations in  
48 the initial ensemble or any perturbations to the lateral boundary conditions. The ensemble forecasts  
49 of SXJ12 exhibited substantial high-biases in heavy precipitation, as well as the development of  
50 spurious convection near the CASA radar network later in the forecast period. SXJ12 also noted  
51 that the trailing convective line of the MCS dissipated too quickly near the southern model domain  
52 boundary in their forecast ensembles.

53 This study builds upon and extends the work of SXJ11 and SXJ12, addressing the  
54 shortcomings of those studies through an improved ensemble DA and forecast framework. The new  
55 experiments investigate the assimilation of both radar and conventional observations, including  
56 surface observations at five-minute intervals from the Oklahoma Mesonet. In their analysis and  
57 forecast study of the same case, Schenkman et al. (2011) found that assimilating Oklahoma Mesonet  
58 observations via 3DVAR significantly improved the near-surface wind field within the model.  
59 Furthermore, the current study uses lateral boundary conditions that include mesoscale  
60 perturbations on the outer grid; the boundary conditions for the inner nest are interpolated from the  
61 outer-nest ensemble members. Recent studies (e.g. Jung et al. 2012; Yussouf et al. 2013) have  
62 shown promising results for storm-scale data assimilation using similar ensemble designs. The  
63 geographic extent of inner-nest domain is also doubled in both horizontal directions compared to  
64 SXJ12, reducing the potential negative impacts of boundary conditions.

65 This study will examine the relative and combined impacts of radar and conventional  
66 observations, assimilated using an EnKF, on the ensemble analyses and subsequent ensemble  
67 forecasts of the 9 May 2007 MCS. The remainder of this paper is organized as follows: Section 2  
68 discusses the data assimilated, the ensemble DA, and the forecast experiments and methods.

69 Section 3 discusses the results of the experiments, focusing on improvements from prior work,  
70 impacts of assimilating radar and/or conventional data, and ensemble forecast verification. Analyses  
71 and forecasts of radar reflectivity (as a proxy for precipitation), mesovortices (an indicator of  
72 tornado potential), surface temperature, and surface dewpoint are verified against radar and  
73 Oklahoma Mesonet observations, and forecast sensitivity to the assimilation configuration of  
74 conventional observations is considered. Finally, section 4 contains a summary with concluding  
75 remarks.

## 76 **2. Experiment setup and verification methodology**

77 Similar to SXJ11 and SXJ12, EnKF analyses and 3-hour storm-scale ensemble forecasts are  
78 generated for the tornadic MCS that occurred over Oklahoma and Texas on 8-9 May 2007. During  
79 this event, a line-end vortex (LEV) developed near the northern end of the MCS. This LEV moved  
80 through southwestern and central Oklahoma, producing two confirmed EF-1 tornadoes and one  
81 confirmed EF-0 tornado in central Oklahoma between 0354 UTC and 0443 UTC. For additional  
82 details regarding the structure, evolution, and timing of the 8-9 May 2007 MCS we refer the reader  
83 to SXJ11.

84 SXJ11 and SXJ12 sought to assess the impact of two factors on their ensemble analyses and  
85 forecasts: (1) the assimilation of CASA X-band radar observations, and (2) the use of a mixed-  
86 microphysics ensemble as a means to mitigate ensemble under-dispersion. While the assimilation  
87 of WSR-88D data alone produced a reasonable analysis of the convective system, SXJ11 found that  
88 assimilating CASA X-band data in addition to WSR-88D data improved the resulting analysis,  
89 particularly with regard to the representation of near-surface circulations. Use of a mixed-  
90 microphysics ensemble was found to alleviate under-dispersion by increasing the ensemble spread.

91 SXJ12, which carried out ensemble forecasts initialized from the EnKF analyses of SXJ11, found  
92 that both assimilation of CASA data and the use of a mixed-microphysics ensemble improved 2-  
93 hour forecasts of the tornadic mesovortex embedded within the MCS.

94 SXJ11 and SXJ12 used a single DA and forecast domain with a 2 km horizontal grid  
95 spacing; only radar data were assimilated. Lateral boundary conditions were provided by the NCEP  
96 NAM 6-hourly analyses and intervening 3-h forecasts. The initial ensemble was created by adding  
97 random perturbations with 6-km spatial de-correlation scales to a 1-hour spinup forecast on the 2-  
98 km grid initialized from the 0000 UTC, 9 May 2007 NCEP NAM analysis. While SXJ11 and SXJ12  
99 produced encouraging analyses and forecasts, several deficiencies exist with their setup. Only  
100 storm-scale perturbations were used; no mesoscale perturbations were applied. Previous storm-  
101 scale data assimilation studies (e.g. Aksoy et al. 2009) suggest that proper structure in mesoscale  
102 uncertainty is highly important in obtaining good analyses and forecasts. Furthermore, the single  
103 DA domain used the same lateral boundary condition for all members, reducing ensemble spread  
104 near the upwind lateral domain boundaries and contributing to under-dispersion in the ensemble  
105 analyses and forecasts. The geographic extent of the SXJ11/12 domain was also rather limited,  
106 which caused detrimental interaction between the simulated MCS and the southern domain  
107 boundary. SXJ11/12 also did not assimilate surface observations; SXJ12 found that convergence in  
108 the near-surface flow in the model contributed to the development of spurious convection in the  
109 forecast ensemble.

110 To improve upon the results of SXJ11 and SXJ12, several enhancements are considered in  
111 this study. Most prominently, two grids are used: an outer  $300 \times 300 \times 40$  grid with 6 km  
112 horizontal spacing, and an inner  $512 \times 512 \times 40$  grid with 2 km horizontal spacing—the extent of the  
113 2 km domain is substantially expanded from that of SXJ11/12 (Fig. 1). Data assimilated on the

114 inner grid include both radar and conventional observations, including surface observations at five  
115 minute intervals from the Oklahoma Mesonet. Schenkman et al. (2011) found that frequent  
116 assimilation of Oklahoma Mesonet observations using 3DVAR substantially improved the near-  
117 surface flow for a short-term deterministic forecast of this MCS.

118         Uncertainties in the storm environment are taken into account by adding mesoscale  
119 perturbations to the initial ensemble, and by introducing perturbations to the lateral boundary  
120 conditions. Member-by-member one-way nesting is applied from the outer- to the inner-domain—  
121 both ensembles contain 40 members. On both the outer and inner domains, a single-moment ice  
122 microphysics scheme based upon Lin et al (1983) is used, with a rain intercept parameter of  $8.0 \times 10^5$   
123  $\text{m}^{-4}$ ; this value is reduced from the default value following the results of Snook and Xue (2008). All  
124 other model settings, including terrain, radiation, surface physics, and turbulence closure, follow  
125 those of SXJ12.

126         As in SXJ11, we use the ARPS EnSRF DA system (Xue et al. 2006; Tong and Xue 2008b).  
127 The outer grid (Fig. 1) forecast is first initialized at 1800 UTC on 8 May 2007 from the 8 May 2007  
128 NCEP 1800 UTC North American Mesoscale Model (NAM) analysis, and a single, 3-hour pre-  
129 forecast is performed from this initial condition (Fig. 2). At 2100 UTC, an ensemble of 40 members  
130 is created by adding smoothed, random, Gaussian, mesoscale perturbations to the deterministic  
131 forecast with de-correlation scales of 36 and 7.2 km in the horizontal and vertical, respectively,  
132 using the method of Tong and Xue (2008a). Perturbations are added to the horizontal wind ( $u$ ,  $v$ )  
133 with a mean standard deviation of  $2 \text{ m s}^{-1}$ , to the potential temperature ( $\theta$ ) using positive  
134 perturbations only with a mean standard deviation of 1 K, and to the mixing ratio of water vapor  
135 ( $q_v$ ) with a mean standard deviation of 10% of the  $q_v$  value at the given grid point.

136         On the outer domain, conventional observations are assimilated, including Automatic

137 Surface Observation System (ASOS) and Automatic Weather Observing System (AWOS)  
138 observations, Oklahoma Mesonet observations, wind profiler data, and upper air observations  
139 (including soundings at 0000 UTC, 9 May 2007); these data are assimilated hourly from 2200 UTC,  
140 8 May 2007 to 0100 UTC, 9 May 2007. Assumed observation errors used vary by observation type  
141 as indicated in Table 1. To help maintain ensemble spread during the DA on the 6 km grid, we  
142 apply multiplicative covariance inflation to the prior ensemble over the entire domain with an  
143 inflation factor of 1.03 (Anderson and Anderson 1999; Tong and Xue 2005). In addition, we also  
144 apply the relaxation technique of Zhang et al. (2004) with a coefficient of 0.5. Finally, 4-hour  
145 ensemble forecasts are performed from the 0100 UTC ensemble analyses on the outer grid,  
146 producing forecasts until 0500 UTC; these forecasts are used to provide ensemble lateral boundary  
147 conditions for the inner nest forecast ensembles.

148         The 2 km inner-grid EnKF DA experiments are initialized from the outer grid ensemble  
149 analyses at 0100 UTC via spatial interpolation. Lateral boundary conditions for inner-grid ensemble  
150 members are from the forecasts of corresponding outer grid members at 15 minute intervals. The  
151 inner-nest experiments assimilate data every 5 minutes from 0105 through 0200 UTC; the data  
152 assimilated include conventional data as described above, as well as radar reflectivity and radial  
153 velocity from WSR-88D and CASA radars. For radar data, observation error standard deviations  
154 are assumed to be  $2 \text{ m s}^{-1}$  for radial velocity and 3 dBZ for radar reflectivity; these values are  
155 increased from the  $1 \text{ m s}^{-1}$  and 2 dBZ used in SXJ11 following Jung et al. (2012). The observation  
156 operators used to map the model state to reflectivity and radial velocity observations follow Jung et  
157 al. (2008). As in Xue et al. (2006), a Gaussian power-gain function following Wood and Brown  
158 (1997) is used in the forward operator to sample radar data on the radar elevation angles. The  
159 horizontal and vertical covariance localization radii for radar data is set to 6 km. For conventional

160 data, the vertical localization radius is set to 6 km, and horizontal localization radii of 300 km for  
161 surface observations and 800 km for wind profiler and upper-air observations are used in the  
162 primary set of experiments; smaller localization radii are tested in sensitivity experiments on the  
163 inner nest. The locations of assimilated conventional observations and radar sites on the inner nest  
164 are shown in Fig. 3.

165 Four primary experiments are run on the inner grid to investigate the impacts of radar and  
166 conventional weather observations, assimilated individually or in combination, on the ensemble  
167 analyses and forecasts. These four experiments are summarized in Table 2. In experiment  
168 RADCONV, data from the WSR-88D and CASA network radars are assimilated, along with  
169 conventional observations. In experiment RAD, only radar data are assimilated; similarly, in  
170 experiment CONV radar data are omitted and only conventional data are assimilated. Finally, a  
171 control experiment (CNTL) is performed in which no data of any kind are assimilated on the inner  
172 domain—the CNTL forecast ensemble is allowed to run freely from the initial ensemble states at  
173 0100 UTC.

### 174 **3. Results**

175 In evaluating the results of the forecast experiments, we will first briefly consider the  
176 performance of the experiments in this study relative to SXJ12, before moving on to the primary  
177 focus: the nature and extent of the individual and combined impacts of assimilated conventional and  
178 radar observations on the ensemble forecasts. Forecast verification is performed hourly between  
179 0300 and 0500 UTC for radar reflectivity, surface dewpoint, and surface temperature. In addition,  
180 forecasts of low-level mesovortices are produced and verified at 0400 UTC, at which time a  
181 pronounced tornadic mesovortex was present in the observations (SXJ12). Radar reflectivity is

182 chosen for verification because it serves as a proxy for precipitation and can be readily verified  
183 against WSR-88D observations spanning the full extent of the MCS. A neighborhood ensemble  
184 probability (NEP) (Schwartz et al. 2010) method with a 5 km neighborhood radius is used for  
185 verification of radar reflectivity forecasts, and object-based probabilistic verification following the  
186 methodology of SXJ12 is used for the verification of mesovortex forecasts. The radar observations  
187 used for forecast verification are obtained by interpolating full radar reflectivity volumes to the  
188 model grid to produce a gridded radar reflectivity mosaic. Verification of surface temperature and  
189 dewpoint is performed by direct comparison of the forecast ensemble mean against Oklahoma  
190 Mesonet observations.

191

192 *a) Comparison to prior results of SXJ11 and SXJ12*

193 Compared to SXJ12, there are two primary differences in the ensemble design in this study:  
194 the use of (1) a nested-grid EnKF with mesoscale initial perturbations evolved on the outer nest,  
195 allowing the ensemble to take into account uncertainties on multiple scales, and (2) a much larger  
196 inner-domain to allow for the assimilation of more observations on the high-resolution inner nest,  
197 and reduce the influence of the relatively coarse lateral boundary conditions, which are now  
198 perturbed. Experiment RAD can be considered analogous to the NoMMP experiment of SXJ11 and  
199 SXJ12 in terms of the data assimilated and the model configuration—both use the Lin  
200 microphysical scheme (Lin et al. 1983) for all ensemble members and assimilate only radar  
201 observations. It should be noted, however, that data from WSR-88D radar at Fort Worth, Texas  
202 (KFWS) are assimilated in RAD but not in the experiments of SXJ12. When ensemble forecasts  
203 were produced omitting KFWS radar data, the ensemble forecasts obtained were very similar to  
204 those presented in this study, with the exception of a transient area of spurious convection in

205 northeastern Texas, outside the domain of SXJ12 (not shown). Also, while no conventional data are  
206 assimilated in RAD, the experiment still benefits indirectly from conventional observations  
207 assimilated on the outer nest that provides the initial ensemble and the lateral boundary conditions  
208 for RAD.

209         Qualitatively, RAD produces an ensemble forecast similar to that of NoMMP from SXJ12.  
210 Both ensembles predict the northeastward motion of the MCS and its embedded line-end vortex  
211 with reasonable accuracy. In NEP forecasts of radar reflectivity exceeding 25 dBZ ( $P[Z > 25$   
212 dBZ]), a threshold corresponding to light-to-moderate rainfall, greater variation within the RAD  
213 ensemble is evident, as indicated by larger areas of low to moderate probability in RAD (Fig. 4d-f).  
214 Experiment NoMMP of SXJ12 has relatively little variation among members, indicated by a  
215 forecast dominated by regions of either very high or near-zero probability (Fig. 4a-c). The  
216 increased variability in RAD is likely a result of increased ensemble spread imparted by the initial  
217 and lateral boundary condition ensembles from the outer grid. Though the sharpness seen in  
218 NoMMP of SXJ12 can be a useful trait in a probabilistic forecast, this is only the case when high  
219 confidence in the forecast outcome is justified. The sharpness and the corresponding high  
220 confidence do not reflect the relatively large position error at 0500 UTC (Fig. 4c).

221         Looking at the forecast structure in detail, RAD (Fig. 4d-f) outperforms SXJ12's NoMMP  
222 (Fig. 4a-c) late in the forecast period in two regions: in the convective lines south of the line-end  
223 vortex, and near the CASA radar network in southwest and south-central Oklahoma. In NoMMP,  
224 convection in the southern portion of the trailing convective line (as outlined by the black contours  
225 in the southern  $\frac{1}{4}$  of domain in Fig. 4c) dissipates relatively quickly in most ensemble members; by  
226 0500 UTC (Fig. 4c), the southward extent of the trailing convective line is greatly underestimated in  
227 the NEP forecast (NEP is mostly zero there). In contrast, RAD maintains low to moderate

228 probability in the southern portion of the trailing convective line throughout the forecast period  
229 (Fig. 4d-f). Since the bulk of the trailing convective line was located far from the additional data  
230 from KFWS during the DA period (between 0100 and 0200 UTC), the greatest contribution to the  
231 improved representation of the trailing convective line in RAD is likely from the expanded inner  
232 domain (see Fig. 1). SXJ12 speculated that interaction with the southern domain boundary was the  
233 cause of the deterioration of the trailing convective line within their forecast ensemble; the  
234 improved representation of the line in RAD supports this theory.

235 In addition to improvement in the trailing line, the NEP forecast of radar reflectivity  
236 exceeding 25 dBZ ( $P[Z > 25 \text{ dBZ}]$ ) in RAD also indicates the presence of the leading convective  
237 line at 0400 and 0500 UTC (Fig. 4e, f); this line is absent in NoMMP (Fig. 4b, c). Also, overall  
238 storm motion, as inferred from the evolution of the NEP forecasts of radar reflectivity (Fig. 4), is  
239 forecasted well in RAD, whereas the system moves too slowly in NoMMP. Finally, SXJ12 noted  
240 spurious convection occurring near the CASA radar network (see Fig. 3) after 0300 UTC in all of  
241 their experiments as a result of near-surface convergence in this region in the model. Spurious  
242 convection is absent in this region in RAD (Fig. 4e-f).

243 One commonly-used measure of the skill of probabilistic forecasts is the area under the  
244 relative operating characteristic curve (AUC) (Mason 1982); this skill score measures the ability of  
245 a probabilistic forecast to correctly differentiate between events and non-events, with higher values  
246 indicating greater skill. In Fig. 5, AUC is shown for NEP forecasts of  $P[Z > 25 \text{ dBZ}]$  over a  
247 subdomain in the region of the LEV, identical to that used in SXJ12 (see Fig. 3) for RAD and for  
248 two experiments from SXJ12 (SXJ12-NoMMP, and their best-performing experiment, SXJ12-  
249 CNTL). AUC for RADCONV, which will be discussed in section 3b, is also included for  
250 comparison. Though the SXJ12 experiments slightly outperform RAD at 0300 UTC, RAD

251 outperforms the SXJ12 experiments at 0400 and 0500 UTC. Furthermore, the decrease in forecast  
252 skill with time is slower in RAD than in the SXJ12 experiments. This slower decline in forecast  
253 skill, along with the overall better performance of RAD late in the forecast period, suggests the  
254 positive impact of accounting for mesoscale uncertainties and assimilating conventional data on the  
255 outer grid on the reflectivity forecast.

256 As in NoMMP, a substantial high bias in heavy precipitation (indicated by areas of radar  
257 reflectivity exceeding 40 dBZ) is present throughout the forecast period in RAD. This bias is  
258 evident in NEP forecasts of  $P[Z > 40 \text{ dBZ}]$  (Fig. 6), where the extent of moderate-to-high  
259 probability in the forecast ensemble is far greater than the observed coverage of radar reflectivity  
260 exceeding 40 dBZ, particularly in the region of the trailing convective line. This type of bias is  
261 noted in all experiments assimilating radar data, and will be discussed in greater detail in section 3d.

#### 262 *b) Impact of data sources during the analysis period*

263 Experiments RAD, CONV, and RADCONV focus on the relative impact of conventional  
264 and radar data sources assimilated by the EnKF. RAD (CONV) evaluates the impact of radar  
265 (conventional) data while RADCONV assesses the combined impact of radar and conventional  
266 observations. CNTL provides a basis for comparison. We note, however, that all experiments  
267 (including CNTL) benefit from hourly EnKF DA of conventional observations on the outer grid  
268 prior to 0100 UTC. All radar-assimilating experiments discussed in this study use both WSR-88D  
269 and CASA observations. The specific impact of CASA X-band radar data will not be considered in  
270 this paper—such experiments were performed, and they yielded results largely consistent with the  
271 findings of SXJ11 and SXJ12.

272 The assimilation of radar data and the assimilation of conventional observations each have  
273 positive impacts on the ensemble forecasts and analyses during the DA period. Compared to the

274 CNTL ensemble, the radar-assimilating ensembles (RADCONV and RAD) exhibit substantially  
275 lower root-mean-square (RMS) innovation, as expected, in both radial velocity (Fig. 7) and radar  
276 reflectivity (Fig. 8) within the observational area of four WSR-88D radars close enough to observe  
277 the MCS. Furthermore, the probability-matched (Ebert 2001) ensemble mean radar reflectivity  
278 field of RADCONV and RAD (Fig. 9a, b) is structurally much closer to the observed radar  
279 reflectivity field (Fig. 9e) than that of CNTL (Fig. 9d). The ensemble spread in the radar-  
280 assimilating experiments is quickly reduced, both in terms of radial velocity (Fig. 7) and radar  
281 reflectivity (Fig. 8). Despite the use of multiplicative covariance inflation to maintain spread, the  
282 low spread in RADCONV and RAD indicates that the ensembles quickly become under-dispersive;  
283 such under-dispersion has often been noted in convective-scale ensembles that assimilate radar  
284 observations (e.g. Aksoy et al. 2009; Dowell and Wicker 2009; Jung et al. 2012; Yussouf et al.  
285 2013).

286 Assimilation of conventional data alone in CONV results in substantially reduced RMS  
287 innovation of Z, compared to CNTL, against the observations of the KDYX and KFWS radars (Fig.  
288 8a, b). CONV performs similarly to CNTL against KTLX and KVNK observations (Fig. 8c, d).  
289 KDYX and KFWS primarily observe the trailing stratiform precipitation and trailing convective  
290 line between 0100 and 0200 UTC, while KTLX and KVNK mainly observe the leading portion of  
291 the MCS. Both the CNTL and CONV ensembles contain a large area of spurious precipitation in  
292 northern Oklahoma and southern Kansas, located within the observation areas of KTLX and KVNK  
293 (Fig. 9c-e). Assimilation of conventional observations alone could not suppress this region of  
294 spurious convection. RMS innovations of  $V_r$  in CONV are generally lower than or similar to those  
295 of CNTL (Fig. 7). The 0200 UTC probability-matched ensemble mean of radar reflectivity in  
296 CONV (Fig. 9c) shows improved representation of heavy precipitation in central and south-central

297 Oklahoma compared to CNTL (Fig. 9d), as well as hinting at the trailing stratiform region, but also  
298 contains quite a bit of spurious precipitation, particularly to the east of the observed MCS (Fig. 9c,  
299 e). By contrast, in RAD and RADCONV, where radar data are assimilated alone or alongside  
300 conventional data, the probability-matched ensemble mean reflectivity at 0200 UTC (Fig. 9a)  
301 closely matches the structure of the observations (though the predicted intensity is 5-10 dBZ lower  
302 than observed over much of the MCS), and the spurious precipitation regions seen in CONV (Fig.  
303 9c) are absent. Previous studies (e.g., Tong and Xue 2005) have shown the importance of  
304 assimilating radar data in clear-air regions in suppressing spurious precipitation during EnKF DA,  
305 consistent with the current results.

306 *c) Impact of data sources on ensemble precipitation forecasts*

307 NEP forecasts of  $P[Z > 25 \text{ dBZ}]$  at 0300, 0400, and 0500 UTC (Fig. 10), are generally  
308 skillful, particularly for the radar-assimilating experiments RADCONV (Fig. 10a-c) and RAD (Fig.  
309 10d-f). RADCONV and RAD both predict regions of high  $P[Z > 25 \text{ dBZ}]$  which closely match the  
310 region of precipitation exceeding 25 dBZ observed by the WSR-88D radar network, both in shape  
311 and in extent, particularly at 0300 and 0400 UTC. Decay of the southern portion of the trailing line  
312 is observed in many ensemble members at 0400 and 0500 UTC in RADCONV and RAD, though  
313 not to as great an extent as in SXJ12. The motion of the precipitation regions exceeding 25 dBZ in  
314 RADCONV and RAD matches well with the observed system (Fig. 10a-f).

315 In CONV and CNTL, where no radar data are assimilated, NEP forecasts of  $P[Z > 25 \text{ dBZ}]$   
316 are less accurate than those in RADCONV and RAD. The region of highest probability in CONV  
317 and CNTL is located in a west-southwest to east-northeast oriented streak in southern Kansas, near  
318 and just beyond the northern end of the observed region of precipitation exceeding 25 dBZ (Fig.  
319 10g-l). Assimilation of conventional observations in CONV (Fig. 10g-i) results in an improved

320 representation of the leading portion of the MCS in the NEP forecast of  $P[Z > 25 \text{ dBZ}]$  compared to  
321 CNTL (Fig. 10j-l), as well as increased values of  $P[Z > 25 \text{ dBZ}]$  in central Oklahoma near the LEV,  
322 particularly at 0400 and 0500 UTC (Fig. 10h, i). CONV also, however, contains some low-  
323 moderate values of  $P[Z > 25 \text{ dBZ}]$  in the southwestern and southeastern portions of the forecast  
324 domain, away from any observed precipitation exceeding 25 dBZ—since no radar data were  
325 assimilated, this spurious convection could not be effectively suppressed. Overall, assimilation of  
326 conventional data alone improved the ensemble precipitation forecast modestly, but not nearly as  
327 much as assimilating radar observations.

328         The threshold of 25 dBZ is chosen to focus on all precipitation exceeding a light to moderate  
329 intensity. Depending upon the desired forecast focus, however, a lower threshold may be used to  
330 include light precipitation in the NEP forecast, or a higher threshold may be chosen in order to focus  
331 exclusively on convective cores. To examine the impact of data sources on probabilistic forecasts  
332 of reflectivity with varying thresholds, AUC calculated over a subdomain encompassing the general  
333 region observed by the Oklahoma Mesonet (the red box in Fig. 3) is presented hourly between 0200  
334 and 0500 UTC in Fig. 11 for all experiments for NEP reflectivity forecasts with thresholds varying  
335 from 10 to 50 dBZ. For each experiment, the 5<sup>th</sup> to 95<sup>th</sup> percentile range (a 90 % confidence  
336 interval) is also shown; to generate confidence intervals, a bootstrap method is used to produce  
337 1000 randomly re-sampled 40 member ensembles to evaluate the statistical significance of the  
338 differences between experiments.

339         For the radar-assimilating experiments (RAD and RADCONV), AUC in the 0200 UTC  
340 analysis (Fig. 11a) is very high (close to 1) for thresholds between 10 and 30 dBZ, declining to  
341 between 0.8 and 0.9 for higher thresholds, indicating a highly-skillful ensemble analysis over the  
342 Oklahoma subdomain. Progressing through the forecast period, AUC generally decreases in RAD

343 and RADCONV; the highest values (and thus most skillful forecasts) are for thresholds between 15  
344 and 35 dBZ. AUCs for thresholds above 40 dBZ decline quickly; this is to be expected, since at  
345 these thresholds only very intense reflectivity cores are being considered, and forecast skill is highly  
346 sensitive to displacement errors of these small, intense cores. Furthermore, the sample size at  
347 thresholds above 40 dBZ is quite small, reducing the confidence of AUC values at these thresholds.  
348 Also, though RAD has a slightly higher AUC than RADCONV at 0200 UTC (Fig. 11a), particularly  
349 for higher thresholds, RADCONV actually outperforms RAD for all thresholds at 0300 UTC (Fig.  
350 11b), and shows similar or better performance at 0400 and 0500 UTC (Fig. 11c, d). Though RAD  
351 produces a better initial fit to the radar observations (note that a tighter fit of analysis to  
352 observations assimilated does not necessarily mean better analysis), the addition of conventional  
353 data in RADCONV results in more skillful 1- and 2-hour forecasts. In both RAD and RADCONV,  
354 the 5<sup>th</sup> to 95<sup>th</sup> percentile range is quite small at most thresholds, indicating relatively low spread  
355 within the ensemble. As noted earlier, under-dispersion within the ensemble is a common issue  
356 when assimilating radar observations (Aksoy et al. 2009; Dowell and Wicker 2009; Jung et al. 2012;  
357 Yussouf et al. 2013).

358         When conventional data are assimilated alone in CONV, the 0200 UTC ensemble analysis of  
359 radar reflectivity has a substantially higher AUC than control experiment CNTL, but a substantially  
360 lower AUC than the radar-assimilating experiments. AUC in CONV actually increases at high  
361 thresholds between 0200 and 0300 UTC (Fig. 11a, b), even outperforming RAD and RADCONV  
362 due to good placement of heavy convective cores in south-central Oklahoma and less over-  
363 prediction of very intense rainfall cores (not shown), before declining at all thresholds between  
364 0300 and 0500 UTC (Fig. 11b-d). Positive impact of conventional data in CONV (compared to  
365 CNTL) remains evident at 0300 UTC (Fig. 11b), but cannot be discerned at later times because of

366 the relatively high AUC of CNTL. Although the overall structure of MCS predicted by CNTL is  
367 poor (Fig. 10), the relatively low false-alarm rate coupled with decent precipitation placement (Fig.  
368 10k, l) leads to a deceptively high AUC score for high thresholds at later forecast hours.

369 Another method of assessing the skill of a probabilistic forecast is the reliability diagram,  
370 which compares the observed relative frequency of an event to the forecast probability. Because the  
371 reliability diagram is conditioned on the ensemble forecast, while AUC is conditioned on the  
372 observations, these two metrics complement one another and give a more complete assessment of  
373 forecast skill. Reliability diagrams, calculated over the Oklahoma verification subdomain (see Fig.  
374 3) using forecast probability bins with a width of 0.05, are plotted for NEP forecasts of  $P[Z > 25$   
375  $\text{dBZ}]$  for all experiments in Fig. 12 to complement the analysis of AUC using the same subdomain  
376 presented in Fig. 11. In an ideal forecast, the observed frequency would be equal to the forecast  
377 probability, resulting in a straight reliability curve oriented along the 45-degree diagonal. The  
378 region below the diagonal indicates over-forecasting of the event, while the area above the diagonal  
379 indicates under-forecasting. Sharpness diagrams are also presented in Fig. 12, indicating the number  
380 of model grid points falling into each probability bin, and thus the overall distribution of  
381 probabilities in the forecast. Since the verification subdomain extends well outside of the MCS,  
382 these curves contain many zero values.

383 In the ensemble analyses at 0200 UTC (Fig. 12a), the radar-assimilating experiments under-  
384 estimate the coverage of  $Z > 25 \text{ dBZ}$ , while CNTL and CONV substantially over-estimate it. In all  
385 of the data-assimilating experiments (RADCONV, RAD, and CONV), however, there is an overall  
386 monotonic increase in observed frequency as forecast probability increases, which is a desirable  
387 trait. Reliability for the 0200 UTC analysis of CNTL, by contrast, shows no distinct pattern.

388 During the forecast period (Fig. 12b-d), there is a general trend toward over-prediction of  $Z$

389 > 25 dBZ in the data-assimilating experiments. RADCONV shows good reliability at 0300 UTC  
390 (Fig. 12b), remaining near the diagonal except at the highest forecast probabilities, while RAD and  
391 CONV substantially over-predict  $Z > 25$  dBZ for forecast probabilities above 0.4. At 0400 and  
392 0500 UTC, RADCONV, RAD, and CONV show similar behaviors in terms of forecast reliability,  
393 with relatively good reliability at low forecast probabilities and no significant reduction in reliability  
394 with time. The greater over-prediction of  $Z > 25$  dBZ in CONV early in the forecast period (Fig.  
395 12a, b) can largely be attributed to the increased incidence of spurious precipitation regions (Fig.  
396 10g, h). The tendency toward greater over-prediction of  $Z > 25$  dBZ with time noted in RAD and  
397 RADCONV is similar to that seen in Clark et al. (2009) in their convection-allowing ensemble with  
398 4 km horizontal grid spacing.

399 At all hourly forecast times, the radar-assimilating experiments (RAD and RADCONV)  
400 produce significantly more extreme probability values (near 0 or 1) than either CONV or CNTL  
401 (Fig. 12, right-hand side). The assimilation of radar data in these experiments results in strong  
402 agreement among the ensemble members in the structure of the MCS in the 0200 UTC analysis,  
403 while greater spread remains in the CONV and CNTL ensembles (see Fig. 8). The MCS evolves  
404 similarly in many RAD and RADCONV members during the forecast period (see Fig. 10a-f),  
405 causing this sharpness to persist throughout the forecast period.

#### 406 *d) Impact of data sources on precipitation forecast bias*

407 As in SXJ12, domain-wide histograms of radar reflectivity (Fig. 13) reveal persistent biases  
408 in the ensemble forecasts. The forecast histograms are obtained by counting occurrences of radar  
409 reflectivity values in each member separately. The resulting total in each bin is then divided by the  
410 number of members in the ensemble, so that the number of occurrences can be compared directly to  
411 the same quantity for the gridded radar observations. In three of the experiments (RADCONV,

412 RAD, and CNTL), there is a low bias for light precipitation (15-25 dBZ) which is most prevalent  
413 later in the forecast period at 0400 and 0500 UTC. In RAD and RADCONV there is an abundance  
414 of moderately-intense precipitation (25-40 dBZ), resulting in a slight high bias for 30 dBZ and  
415 above. These biases are similar to those found in SXJ12, where assimilation of radar data resulted  
416 in over-prediction of convective regions and under-prediction of light precipitation in stratiform  
417 precipitation regions; this behavior is also present in radar-assimilating experiments RAD and  
418 RADCONV. The high bias in moderate precipitation is absent in CNTL, suggesting that this bias is  
419 induced by the assimilation of radar data. It is possible that some of the bias in the radar-  
420 assimilating forecasts may be due to the interaction of assimilated radar data with the single-  
421 moment microphysics scheme used in the ensembles. In a related study (Putnam et al. 2013) the  
422 high biases in moderately-intense precipitation fields are reduced when a more sophisticated, two-  
423 moment scheme is used for this case.

424 At 0200 UTC, CONV exhibits a substantial high bias for both light and moderate  
425 precipitation (15-40 dBZ) mostly due to overestimation of the extent of the precipitation area. By  
426 0300 UTC, the high bias in  $Z < 20$  dBZ has disappeared, but the high bias in moderately-intense  
427 precipitation (20-35 dBZ) remains through the rest of the forecast period. While CONV shows very  
428 different bias behavior compared to CNTL, RADCONV and RAD have very similar bias behavior  
429 throughout the forecast period. From these results we can conclude that, at least for this case,  
430 conventional data have a different impact upon the forecast bias of  $Z$  when they are assimilated  
431 alone; radar data, when they are assimilated, appear to be the dominant factor with regard to the bias  
432 of  $Z$  within the forecast ensemble.

#### 433 *e) Impact of data sources on mesovortex prediction*

434 Object-based ensemble forecasts of the probability of low-level mesovortices within 25 km

435 of a point are calculated using the two-hour ensemble forecasts valid at 0400 UTC. The  
436 methodology and criteria used to perform these forecasts follow that of SXJ12, and we refer the  
437 reader to SXJ12 for further details regarding the probability calculation. At 0400 UTC, a tornadic  
438 mesovortex was present west-southwest of the Oklahoma City metropolitan area, indicated by the  
439 black triangle in each panel of Fig. 14. All four ensemble forecast experiments (RADCONV, RAD,  
440 CONV, and CNTL) indicate a probability of at least 0.1 of a mesovortex being present in close  
441 proximity to the observed tornadic mesovortex; probability near the observed vortex location is  
442 highest in RADCONV, and lowest in CONV and CNTL.

443 RADCONV produces the best probabilistic mesovortex forecast, with a region of moderate  
444 probability (maximum  $> 0.5$ ) concentrated near the observed mesovortex location (Fig. 14a). RAD  
445 predicts a wider region of relatively low probability (maximum  $\approx 0.3$ ), centered 20-30 km northwest  
446 of the observed mesovortex location, as well as a lobe of probability between 0.05 and 0.20  
447 extending into northwestern Oklahoma (Fig. 14b). CNTL, which did not benefit from any  
448 assimilated observations on the inner grid, predicts a small region of low probability (maximum  $<$   
449 0.3), also centered 30-40 km northwest of the observed mesovortex location (Fig. 14d), suggesting  
450 that at least some of the information needed to correctly predict the mesovortex in this MCS is  
451 captured in the outer-nest ensemble providing the initial and lateral boundary conditions for CNTL.

452 The better mesovortex prediction of RADCONV compared to RAD supports the findings of  
453 Schenkman et al. (2010), who showed that the assimilation of Oklahoma Mesonet, CASA, and  
454 WSR-88D observations for this case using a 3DVAR and cloud analysis system yielded a better  
455 prediction of the low-level wind field and the tornadic mesovortex than when assimilating radar  
456 data alone. We note, however, that assimilation of conventional data alone does not improve the  
457 probabilistic mesovortex forecast over CNTL. In short, for this case, assimilation of conventional

458 data only results in an improved mesovortex forecast when radar data are also assimilated.  
459 Assimilation of radar data, which have relatively complete volumetric coverage throughout the  
460 MCS, is necessary to produce a good initial representation of the storm within the model.  
461 Assimilating conventional observations alone (which are far coarser than the radar observations and  
462 most abundant at the surface) cannot substantially improve the storm-scale ensemble forecast or  
463 impart accurate 3-dimensional storm structure in this case.

464 *f) Verification of surface temperature and dew point*

465 The ability of radars to provide complete volumetric coverage is generally limited near the  
466 surface, since the curvature of the earth prevents radars from observing the near-surface region  
467 beyond a few tens of kilometers from the radar site. This limitation motivates the assimilation of  
468 conventional observations alongside radar, particularly when relatively dense surface observations  
469 such as those from the Oklahoma Mesonet are available. To assess the skill of the ensemble  
470 forecasts near the surface, surface temperature and dewpoint are verified against Oklahoma  
471 Mesonet observations at 0400 UTC (2 hours of forecast time) in Fig. 15 and Fig. 16, respectively.

472 In all forecast experiments, the ensemble mean surface temperature is colder than observed  
473 near the LEV (located near the Oklahoma City metropolitan area at 0400 UTC; c.f. Fig. 14), and  
474 warmer than observed to the east of the MCS and in the vicinity of the trailing convective line in  
475 southern Oklahoma (Fig. 15). The cold bias near the LEV is greatest in RAD (Fig. 15b), exceeding  
476 3 C in places, and least in CONV (Fig. 15c). RADCONV (Fig. 15a) and RAD (Fig. 15b) show  
477 similar patterns of temperature forecast error, but the magnitude of the cold bias is reduced in  
478 RADCONV compared to RAD.

479 When 2-hour surface dewpoint forecasts are verified (Fig. 16), a prominent dry bias is  
480 present in the forecast ensembles over much of central and northwest Oklahoma. The dry bias is

481 most intense in RAD (Fig. 16b), extending over nearly all of Oklahoma, and exceeding 5 C in  
482 northwest Oklahoma. In CONV and CNTL (Fig. 16c, d), which did not assimilate radar data, the  
483 dry bias is less intense, and more limited in extent. The overall bias is smallest in CONV.  
484 RADCONV (Fig. 16a) has a greatly reduced dry bias compared to RAD (Fig. 16b). The reduced  
485 error in the surface temperature and surface dewpoint forecasts of RADCONV in comparison to  
486 RAD suggest that the conventional observations assimilated in RADCONV impart a substantial  
487 improvement to the surface thermodynamic and moisture fields of the ensemble that is maintained  
488 through the subsequent forecast period.

489 To more closely examine the ensemble behavior of surface fields, ensemble temperature and  
490 dewpoint fields are interpolated to the location of three Oklahoma mesonet sites at five minute  
491 intervals and compared against five-minute observations from the corresponding sites in Fig. 17 and  
492 Fig. 18. The Oklahoma Mesonet sites chosen are marked in Fig. 1 and include Marena (MARE),  
493 located in the northern stratiform region of the MCS; Norman (NRMN), which observes the passage  
494 of the leading convective line during the forecast period; and Grandfield (GRA2), which observed  
495 the post-storm environment during much of the forecast period. In addition to the ensemble mean,  
496 Fig. 17 and Fig. 18 also show the 5<sup>th</sup> to 95<sup>th</sup> percentile range within the ensemble.

497 All experiments produce cold biases at NRMN (Fig. 17e-h) and warm biases at GRA2 (Fig.  
498 17i-l) throughout the analysis and forecast period, while producing values near those observed at  
499 MARE (Fig. 17a-d). The ensembles of RADCONV and CONV (which assimilate conventional  
500 data) show more similar trends in temperature to the observed values than those of RAD and CNTL,  
501 particularly at GRA2 (Fig. 17i-l). For dewpoint, the ensembles show dry biases at MARE and  
502 NRMN (Fig. 18a-h) throughout most of the assimilation and forecast period; this bias is greatest in  
503 RAD. RAD and RADCONV also show a dry bias at GRA2 (Fig. 18i-l) during much of the forecast

504 period; again, this bias is larger in RAD. For both temperature (Fig. 17) and dewpoint (Fig. 18),  
505 RAD has much greater spread within the ensemble, as evidenced by the very large 5<sup>th</sup>-95<sup>th</sup>  
506 percentile range in RAD compared to RADCONV and CNTL. The ensemble of RAD had a few  
507 ensemble members that produced very dry air near the surface over central and southwestern  
508 Oklahoma (not shown), contributing to the very large spread in the ensemble.

509         Because RAD does not assimilate surface observations, it relies on radar observations alone  
510 to adjust the surface thermodynamic fields. Though there is sufficient near surface radar coverage  
511 over much of the domain, the fields produced in RAD based on the influence of radar covariance  
512 structure near the surface do not result in good forecasts in all ensemble members. When  
513 conventional data are assimilated alongside radar data in RADCONV, the bias of the ensemble is  
514 reduced and the ensemble spread in the surface fields is reduced to levels similar to CONV. We  
515 note, however that because the ensemble has a horizontal resolution of only 2 km and a minimum  
516 vertical spacing of 25 m, accurate prediction of surface fields may be beyond the capability of the  
517 ensemble, regardless of the data used. Yussouf et al. (2013) also noted relatively poor agreement  
518 between their storm-scale ensemble using 2-km horizontal grid-spacing and individual surface  
519 observation timeseries.

#### 520 *g) Impact of localization radius for conventional observations*

521         When using an EnKF that assimilates observations from multiple networks, the spatial  
522 covariance localization radii usually need to be tuned. The use of different, optimized radii for  
523 different observation types has been found to improve analyses and forecasts (Dong et al. 2011; Zhu  
524 et al. 2013). In addition to the four primary experiments presented above, several sensitivity  
525 experiments were performed using different covariance localization radii for conventional data  
526 assimilated in RADCONV and CONV. Horizontal localization radii from 100 to 300 km were

527 tested for surface data, and radii from 150 to 800 km were tested for upper-air observations; these  
528 ranges are based upon prior experimentation by the authors. Though all radii tested resulted in  
529 qualitatively similar 0200 UTC analyses (not shown), substantial differences were noted during  
530 subsequent ensemble forecasts. Because the forecast trend is monotonic as radii increase, we only  
531 show results using the smallest and largest radii tested.

532         When only conventional data are assimilated, the quality of NEP forecasts of  $P[Z > 25 \text{ dBZ}]$   
533 is better for larger localization radii. When a localization radius of 300 km is used for surface data  
534 and 800 km for upper-air data, as in CONV, regions of high probability are confined to a relatively  
535 tight north-south line slightly to the east of the observed MCS location (Fig. 19a-c), and NEP near  
536 the LEV is high (in many places,  $> 0.95$ ). By comparison, when reduced localization radii of 100  
537 km for surface data and 150 km for upper air data are used (Fig. 19d-f), the linear structure of the  
538 MCS is not as evident, particularly at 0500 UTC (Fig. 19f), and NEP is lower in the vicinity of the  
539 LEV. The two distinct precipitation regions in the simulation using reduced localization radii are  
540 more similar to those of CNTL. When radar data are assimilated alongside conventional data, the  
541 positive impact of the assimilated observations is less when a smaller localization radius is used for  
542 conventional observations (not shown).

543         The tornadic mesovortex that was ongoing at 0400 UTC is also better predicted when larger  
544 localization radii are used to assimilate conventional data. Mesovortex probability predictions for  
545 RADCONV (Fig. 20a) and CONV (Fig. 20c), which use 300 km and 800 km localization radii for  
546 surface and upper-air data, respectively, compare favorably to their counterparts using reduced  
547 localization radii for conventional DA (Fig. 20b, d). In particular, when radar data are assimilated  
548 alongside conventional data (Fig. 20a-b), the probability at the observed mesovortex location is  
549 increased, the probability field is more concentrated, and spurious detections are reduced using the

550 larger localization radii (Fig. 20a). The seemingly optimal 300 km localization radius used for the  
551 surface data appears rather large compared to the mean station spacing of the Oklahoma Mesonet,  
552 which is about 30 km. Dong et al. (2011) suggested optimal localization radii that are slightly larger  
553 than the mean station spacing of individual observation networks. We theorize that, since  
554 conventional observations were rather sparse above the surface and outside of the region covered by  
555 the much denser Oklahoma Mesonet, expanding the influence of conventional observations helps  
556 spread their positive impact upstream into regions with sparse data coverage during DA, thus  
557 increasing their ability to positively influence the ensemble forecast.

#### 558 **4. Discussion and summary**

559 The relative impacts of various data sources, and their effective assimilation, are important  
560 issues in convective-scale weather forecasting. At convective scales, Doppler radars provide  
561 temporally and spatially dense observations of radar reflectivity ( $Z$ ) and radial velocity ( $V_r$ ).  
562 Conventional observations, including ASOS, AWOS, mesonet, wind profiler, and upper-air  
563 observations, offer sparser coverage than radar, but provide valuable information close to the  
564 surface and in clear-air regions that radar typically cannot.

565 In this study, we examined the individual and combined impacts of assimilating radar and/or  
566 conventional observations, using an ensemble square-root Kalman filter (EnSRF), upon ensemble  
567 analyses and forecasts of a tornadic mesoscale convective system (MCS), focusing on prediction of  
568 radar reflectivity (a proxy for precipitation) and low-level mesovortices (a proxy for tornado  
569 potential). Either conventional observations, radar observations, both, or neither are assimilated  
570 using a storm-scale EnKF nested within a mesoscale ensemble; 3-hour ensemble forecasts are  
571 performed from the final ensemble analyses. Conventional and radar observations were both found

572 to improve the analyses and forecasts, but in different ways. We also compared the results of this  
573 study to those of our previous study on the same case using a simpler EnKF DA setup and without  
574 conventional observations (SXJ11, and SXJ12).

575 Compared to SXJ12, the ensemble forecast of the current study that also assimilates radar  
576 data only produced a more skillful 2-3 hour probabilistic forecast of  $P[Z > 25 \text{ dBZ}]$ , and comparable  
577 2-hour forecasts of near-surface mesovortices. The skill of reflectivity forecasts declined more  
578 slowly in the current study than in SXJ12, an improvement largely attributable to the nested-  
579 ensemble procedure used in this study. Furthermore, the representation of the trailing convective  
580 line of the MCS was improved compared to SXJ12, particularly later in the forecast period. This  
581 line is located in the southern portion of the model domain, in a region close to the southern  
582 boundary of the 2 km grid; thus predictions of the trailing line also benefited from the enlarged 2  
583 km grid and the presumably improved southern boundary conditions provided by the outer grid  
584 ensemble.

585 All ensemble forecasts that assimilated conventional and/or radar data outperformed the  
586 control experiment (which assimilated neither) in terms of radar reflectivity probabilistic forecasts.  
587 The most skillful forecast, in terms of the area under the relative operating characteristic (ROC)  
588 curve (AUC), was produced by the ensemble assimilating both radar and conventional observations,  
589 though the result of assimilating radar data alone is almost as good. Assimilation of conventional  
590 observations alone resulted in modest improvement over the control experiment, with the greatest  
591 impact in the ensemble analysis and during the first two hours of the ensemble forecast. In  
592 comparison, the positive impact of assimilating radar observations is retained throughout the 3-hour  
593 forecast period. Assimilating radar data imparted the most skill to probabilistic reflectivity forecasts  
594 for thresholds that include light and/or moderate precipitation (i.e., 10-40 dBZ).

595           When radar and conventional data were assimilated together, the resulting probabilistic  
596 forecasts of  $Z$  closely resemble those obtained by assimilating radar data alone, in terms of  
597 structure, bias, and skill, although the positive impact of conventional data is evident in AUC skill  
598 assessments in the 1- and 2-hour ensemble forecasts. Though conventional observations are  
599 relatively sparse, particularly above the surface, and primarily measure quantities only indirectly  
600 related to precipitation (such as temperature, humidity, and wind velocity), they provide valuable  
601 information on the near-surface temperature, moisture, and wind fields, all of which can help  
602 improve the accuracy of forecasts of convective storms.

603           Probabilistic forecasts in all experiments over-predicted the spatial extent of precipitation. A  
604 high-bias in heavy precipitation and a low-bias in light precipitation ( $> 20$  dBZ) were noted in the  
605 radar-assimilating ensembles. The high-bias in heavy precipitation was absent in the ensemble  
606 assimilating conventional data only, but a strong high-bias in light to moderate precipitation was  
607 noted. When radar and conventional data were assimilated together, the bias behavior was similar  
608 to the ensemble assimilating radar data alone.

609           All four forecast ensembles predicted a discernible ( $> 0.1$ ) probability of a near-surface  
610 mesovortex being present near the observed mesovortex location in a 2-hour forecast. The  
611 experiment assimilating both radar and conventional observations produced the most accurate  
612 prediction, with a region of moderately-high (maximum  $> 0.5$ ) probability tightly concentrated near  
613 and just to the northwest of the observed mesovortex. The forecast in the ensemble assimilating  
614 radar data alone yielded a more dispersed region of low-to-moderate (maximum  $< 0.4$ ) probability  
615 near the observed mesovortex and to the north and west. Both radar-assimilating forecasts  
616 outperformed the other two ensembles (which assimilated conventional data only or no data at all).  
617 While conventional observations can help to improve the near-surface wind field, as found by

618 Schenkman et al. (2011), it appears that they are only beneficial if the ensemble already has a  
619 reasonably accurate representation of the MCS, which for this case requires the assimilation of radar  
620 observations. In short, for prediction of the mesovortex in this case, conventional data provided a  
621 strong benefit, but only when assimilated alongside radar observations. This is consistent with the  
622 findings of the observing system simulation experiments in Dong et al. (2011).

623 Adding conventional observations resulted in substantial improvements to 2-hour ensemble-  
624 mean forecasts of surface dewpoint and temperature. Forecast error, verified against Oklahoma  
625 Mesonet observations, were reduced somewhat over the region near the MCS for 2 m temperature  
626 when both radar and conventional data are assimilated compared to when just radar data are used.  
627 In the ensemble assimilating radar data alone, a substantial dry bias was present in the 2-hour  
628 surface dewpoint forecast—this bias was greatly reduced in the ensemble assimilating both radar  
629 and conventional data. The lowest biases in those variables are found when conventional data were  
630 assimilated alone. These reductions in bias show the value of surface observations—particularly  
631 surface observations with high temporal and spatial resolution—for convective scale forecasts.

632 The choice of horizontal localization radius is found to be important for assimilation of  
633 conventional observations. Localization radii of 100 to 300 km were tested for surface  
634 observations, and 150 to 800 km for upper air observations. Larger radii were found to produce  
635 better ensemble forecasts for precipitation and mesovortices. We theorize that, for this case, since  
636 the bulk of conventional data was confined to the area covered by the Oklahoma Mesonet, using  
637 larger localization radii helped spread the information in the mesonet surface observations  
638 upstream, allowing its benefit to persist longer in the ensemble forecasts. Sobash and Stensrud (2012)  
639 found that convective-scale EnKF DA and forecasts of an MCS, obtained using an OSSE framework, were  
640 sensitive to the covariance localization used for radar observations. While we did not vary covariance

641 localization radii for radar observations in this study, this topic has been actively investigated through  
642 numerical experiments at CAPS. At 1-2 km grid spacing, a localization radius of 6 km has been found to  
643 work well for radar observations (Xue et al. 2006).

644         Though the experiments presented in this paper represent a substantial improvement from  
645 prior studies, we note that some shortcomings remain. The experiments presented in this study use  
646 a single moment microphysical scheme; for this case, Putnam et al. (2013) found that using a dual-  
647 moment microphysical scheme resulted in improved representation of the trailing convective line,  
648 and better representation of dual-polarimetric radar signatures in emulated radar data obtained from  
649 model forecasts. Though the use of a dual-moment microphysical scheme increases the  
650 computational cost and complexity of DA and NWP, it offers a promising means of improving  
651 analysis and forecast quality.

652         Looking forward toward the implementation of real-time ensemble-forecast-based severe  
653 weather warnings, as envisioned in the Warn-on-Forecast paradigm (Stensrud et al. 2009), the  
654 relative benefit of various data sources should be considered together with their associated  
655 assimilation costs. As efficient parallel EnKF algorithms suitable for dense observations are  
656 developed (e.g., Wang et al. 2013), real-time implementation of such systems is increasingly within  
657 reach. Data assimilation experiments, possibly run in a quasi-operational realtime environment,  
658 could provide opportunities to examine the impact of multi-scale data sources on the forecasting of  
659 a large number of cases, leading to more robust conclusions on the impacts. Such studies should be  
660 pursued in the future.

661

662 *Acknowledgements:* This work was primarily supported by NSF grant EEC-0313747 as part of ERC  
663 CASA program. The second author was also supported by NSF grants AGS-0802888, AGS-

664 1046171, AGS-0738370, AGS-0608168, AGS-0750790, AGS-0941491, and OCI-0905040. The  
665 authors would like to acknowledge Keith Brewster and Kevin Thomas for their assistance in data  
666 preparation, as well as Altug Aksoy and two anonymous reviewers for their constructive comments  
667 which helped to improve this manuscript. Computing was performed primarily at the Kraken  
668 supercomputer of the National Institute for Computational Sciences (NICS).

## References

- Aksoy, A., D. C. Dowell, and C. Snyder, 2009: A Multicase Comparative Assessment of the Ensemble Kalman Filter for Assimilation of Radar Observations. Part I: Storm-Scale Analyses. *Mon. Wea. Rev.*, **137**, 1805-1824.
- Aksoy, A., D. C. Dowell and C. Snyder, 2010: A Multicase Comparative Assessment of the Ensemble Kalman Filter for Assimilation of Radar Observations. Part II: Short-Range Ensemble Forecasts. *Mon. Wea. Rev.*, **138**, 1273-1292.
- Aksoy, A., S. Lorsolo, T. Vukicevic, K. J. Sellwood, S. D. Aberson, and F. Zhang, 2012: The HWRF Hurricane Ensemble Data Assimilation System (HEDAS) for High-Resolution Data: The Impact of Airborne Doppler Radar Observations in an OSSE. *Mon. Wea. Rev.*, **140**, 1843-1862.
- Aksoy, A., S. D. Aberson, T. Vukicevic, K. J. Sellwood, S. Lorsolo, and X. Zhang, 2013: Assimilation of High-Resolution Tropical Cyclone Observations with an Ensemble Kalman Filter Using NOAA/AOML/HRD's HEDAS: Evaluation of the 2008–11 Vortex-Scale Analyses. *Mon. Wea. Rev.*, **141**, 1842-1865.
- Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884-2903.
- Anderson, J. L. and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741-2758.
- 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.
- Crum, T. D., R. L. Alberty, and D. W. Burgess, 1993: Recording, archiving, and using WSR-88D

- data. *Bull. Amer. Meteor. Soc.*, **74**, 645-653.
- Dawson, D. T., L. J. Wicker, E. R. Mansell, and R. L. Tanamachi, 2011: Impact of the Environmental Low-Level Wind Profile on Ensemble Forecasts of the 4 May 2007 Greensburg, Kansas, Tornadoic Storm and Associated Mesocyclones. *Mon. Wea. Rev.*, **140**, 696-716.
- Dirren, S., R. D. Torn, and G. J. Hakim, 2007: A Data Assimilation Case Study Using a Limited-Area Ensemble Kalman Filter. *Mon. Wea. Rev.*, **135**, 1455-1473.
- Dong, J., M. Xue, and K. K. Droegemeier, 2011: The analysis and impact of simulated high-resolution surface observations in addition to radar data for convective storms with an ensemble Kalman filter. *Meteor. Atmos. Phy.*, **112**, 41-61.
- Dowell, D. C., and L. J. Wicker, 2009: Additive noise for stormscale ensemble forecasting and data assimilation. *J. Atmos. Oceanic Technol.*, **26**, 911-927.
- 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.
- 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.
- 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. and C. Snyder, 2000: A hybrid ensemble Kalman filter - 3D variational analysis scheme. *Mon. Wea. Rev.*, **128**, 2905-2919.

- Hamill, T. M. and J. S. Whitaker, 2010: What Constrains Spread Growth in Forecasts Initialized from Ensemble Kalman Filters? *Mon. Wea. Rev.*, **139**, 117-131.
- Houtekamer, P. L. and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796-811.
- Houtekamer, P. L., et al., 2005: Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Wea. Rev.*, **133**, 604-620.
- 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 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 radar data. *Mon. Wea. Rev.*, **140**, 1457–1475.
- 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.
- Mason, I. B., 1982: A model for the assessment of weather forecasts. *Aust. Meteor. Mag.*, **30**, 291-303.
- McLaughlin, D., et al., 2009: Short-wavelength technology and the potential for distributed networks of small radar systems. *Bull. Amer. Meteor. Soc.*, **90**, 1797-1817.
- Putnam, B. J., M. Xue, Y. Jung, N. Snook, and G. Zhang, (2013): The analysis and prediction of microphysical states and polarimetric variables in a mesoscale convective system using double-moment microphysics, multi-network radar data, and the ensemble Kalman filter. *Mon. Wea. Rev.* Accepted.
- Schenkman, A., et al., 2011: Impact of CASA radar and Oklahoma mesonet data assimilation on the

- analysis and prediction of tornadic mesovortices in a MCS. *Mon. Wea. Rev.*, **139**, 3422-3445.
- Schenkman, A. D., et al., 2010: 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.
- Schwartz, C. S., et al., 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. *Geophys. Res. Lett.*, **35**, L24803, doi:10.1029/2008GL035866.
- Snook, N., M. Xue, and Y. Jung, 2011: Analysis of a Tornadic Mesoscale 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.*, **140**, 2126-2146.
- Sobash, R. A, and D. J. Stensrud, 2012: The Impact of Covariance Localization for Radar Data on EnKF Analyses of a Developing MCS: Observing System Simulation Experiments. *Mon. Wea. Rev.*, **141**, 3691-3709.
- 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., et al., 2009: Convective-scale Warn on Forecast System: A Vision for 2020. *Bull. Am. Meteor. Soc.*, **90**, 1487-1499.

- Tong, M., 2006: Ensemble Kalman filter 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, 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, 2008a: 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.
- Tong, M. and M. Xue, 2008b: Simultaneous estimation of microphysical parameters and atmospheric state with radar data and ensemble square-root Kalman filter. Part II: Parameter estimation experiments. *Mon. Wea. Rev.*, **136**, 1649–1668.
- Wang, Y., Y. Jung, T. A. Supinie, and M. Xue, 2013: A hybrid MPI/OpenMP parallel algorithm and performance analysis for an ensemble square root filter suitable for dense observations. *J. Atmos. Ocean. Tech.*, **30**, 1382-1397.
- Whitaker, J. S. and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913-1924.
- Wood, V. T. and R. A. Brown, 1997: Effects of radar sampling on single-Doppler velocity signatures of mesocyclones and tornadoes. *Wea. Forecast.*, **12**, 928-938.
- Wu, Chun-Chieh, Guo-Yuan Lien, Jan-Huey Chen, Fuqing Zhang, 2010: Assimilation of Tropical Cyclone Track and Structure Based on the Ensemble Kalman Filter (EnKF). *J. Atmos. Sci.*, **67**, 3806–3822.
- Xue, M., K. K. Droegemeier, and V. Wong, 2000: The Advanced Regional Prediction System (ARPS)--A multiscale non-hydrostatic atmospheric simulation and prediction tool. Part I: Model dynamics and verification. *Meteor. Atmos. Phys.*, **75**, 161-193.

- Xue, M., et al., 2001: The Advanced Regional Prediction System (ARPS)--A multiscale nonhydrostatic atmospheric simulation and prediction tool. Part II: Model physics and applications. *Meteor. Atmos. Phys.*, **76**, 143-165.
- Xue, M., et al., 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., 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.
- Yussouf, N., and D. J. Stensrud, 2012: Comparison of Single-Parameter and Multiparameter Ensembles for Assimilation of Radar Observations Using the Ensemble Kalman Filter. *Mon. Wea. Rev.*, **140**, 562-586.
- 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 Tornadoic Supercell Storm Using Single- and Double-Moment Microphysics Schemes. *Mon. Wea. Rev.*, **141**, 3388-3411.
- 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., et al., 2013: A regional GSI-based ensemble Kalman filter data assimilation system for the

### List of figures

Fig. 1. Geographic extent of the outer model domain (6 km horizontal grid spacing) and the nested inner domain (2 km horizontal grid spacing). Surface elevation (in meters above mean sea level) is plotted for reference. Also shown for reference is the smaller 2-km domain used in SXJ11 and SXJ12. The three black dots in western and central Oklahoma indicate the locations of Oklahoma Mesonet stations MARE (Marena), NRMN (Norman), and GRA2 (Grandfield) used for timeseries verification.

Fig. 2. Flow diagram for forecast experiments. The outer nest forecast is initialized at 1800 UTC on 8 May 2007 via interpolation from the 1800 UTC NAM analysis; 6-hourly NAM analyses and the intervening 3-hour forecasts are used as boundary conditions for the outer nest. The inner nest is initialized at 0100 UTC on 9 May 2007 using the outer nest ensemble for initial and boundary conditions.

Fig. 3. Observations assimilated using EnKF on the inner nested grid (2 km grid spacing). The dashed circles and large solid circles indicate 50 and 150 km radius range rings, respectively, for WSR-88D radar sites used. Small, thin circles indicate 30 km range rings for CASA X-band radar sites used, black triangles indicate ASOS and AWOS surface station sites, squares indicate Oklahoma Mesonet station sites, and diamonds indicate wind profiler sites. The red box indicates the Oklahoma verification subdomain; the smaller green box indicates the verification subdomain used in SXJ12.

Fig. 4. Neighborhood ensemble probabilities (shaded) of reflectivity exceeding 25 dBZ,  $P[Z > 25 \text{ dBZ}]$ , at model grid level 10 (approximately 2 km above the surface) for experiment NoMMP of SXJ12 at (a) 0300 UTC, (b) 0400 UTC, and (c) 0500 UTC, and RAD at (d) 0300 UTC, (e) 0400 UTC, and (f) 0500 UTC. The bold black line in each panel indicates the location of the 25 dBZ radar reflectivity contour observed by the WSR-88D radar network. The leading convective line, trailing convective line, and stratiform region of the MCS are indicated in panel (a).

Fig. 5. Area under the ROC curve (AUC) for RADCONV, RAD, and two experiments from SXJ12 (CNTL and NoMMP) at 0300, 0400, and 0500 UTC 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 the verification subdomain of SXJ12 (the green box in Fig. 2).

Fig. 6. As Fig. 4, but for  $P[Z > 40 \text{ dBZ}]$  and the 40 dBZ radar reflectivity contour.

Fig. 7. Average root-mean-square (RMS) innovation (solid lines) of ensemble mean and the ensemble spread (dotted lines) of radial velocity ( $\text{m s}^{-1}$ ) over the observation region of four WSR-88D radars within the model domain from 0110 to 0200 UTC for all experiments. Calculations are limited to locations where observed and/or model (ensemble mean) reflectivity exceeds 15 dBZ.

Fig. 8. As Fig. 7 but for radar reflectivity (dBZ) instead of radial velocity.

Fig. 9. Probability-matched ensemble mean reflectivity at model grid level 10 (approximately 2 km above the surface) for the 0200 UTC ensemble analyses of (a) RADCONV, (b) RAD, (c) CONV, and (d) CNTL. Also shown is (e) observed reflectivity at 0200 from the WSR-88D network, interpolated to the model grid.

Fig. 10. Neighborhood ensemble probabilities (shaded) of radar reflectivity exceeding 25 dBZ,  $P[Z > 25 \text{ dBZ}]$ , at model grid level 10 (approximately 2 km above the surface) for (a-c) RADCONV,

(d-f) RAD, (g-i) CONV, and (j-l) CNTL at 0300, 0400, and 0500 UTC. The region of radar reflectivity exceeding 25 dBZ observed by the WSR-88D radar network at the corresponding time is outlined by a bold black contour. Urban boundaries are shown in purple.

Fig. 11. Area under the relative operating characteristic (ROC) curve (AUC) (solid, bold lines) for all experiments at (a) 0200, (b) 0300, (c) 0400, and (d) 0500 UTC for forecasts of radar reflectivity at vertical grid level  $k = 10$  (slightly more than 2 km above mean sea level) exceeding threshold values ranging from 10 to 50 dBZ at intervals of 2 dBZ. Also shown are 90% confidence intervals calculated using a 1000-member bootstrap to resample the ensemble (shaded regions). Calculations are performed over the Oklahoma verification subdomain (the red box in Fig. 2). In each panel, the green region indicates AUC values associated with an operationally-useful forecast ( $AUC > 0.7$ ). The red region indicates forecasts with no skill ( $AUC < 0.5$ ).

Fig. 12. Reliability and sharpness diagrams for NEP forecasts of  $P[Z > 25 \text{ dBZ}]$  for all experiments at (a) 0300 UTC, (b) 0400 UTC, and (c) 0500 UTC calculated over the Oklahoma verification subdomain (the red box in Fig. 2). Forecast probability bins are spaced at intervals of 0.05.

Fig. 13. Hourly, domain-wide histograms of forecast radar reflectivity for all ensemble forecast experiments, compared to WSR-88D observed radar reflectivity interpolated to the ensemble forecast grid (bottom row). Bins are placed every 1 dBZ. The vertical axis indicates the number of model grid volumes within each bin, normalized by the size of the forecast ensemble.

Fig. 14. Ensemble-based probability of a significant near-surface mesovortex occurring within 25 km of a point (shaded) at 0400 UTC for (a) RADCONV, (b) RAD, (c) CONV, and (d) CNTL. The location of the observed tornadic mesovortex (located within the line-end vortex of the MCS) at 0400 UTC is indicated by the black triangle in each panel. Urban boundaries are shown in purple.

Fig. 15. Contours of the difference between ensemble mean surface (2 m) temperature forecast and

objectively-analyzed Oklahoma Mesonet observations at 0400 UTC for (a) RADCONV, (b) RAD, (c) CONV, and (d) CNTL. Red areas indicate an ensemble mean temperature warmer than observed by the mesonet, while blue areas indicate an ensemble mean colder than observations.

Fig. 16. Contours of the difference between ensemble mean surface (2 m) dewpoint forecast and objectively-analyzed Oklahoma Mesonet observations at 0400 UTC for (a) RADCONV, (b) RAD, (c) CONV, and (d) CNTL. Green areas indicate an ensemble mean dewpoint higher than observed by the mesonet, while brown areas indicate an ensemble mean dewpoint lower than observations.

Fig. 17. Observed surface (2 m) temperature (F) (solid black lines) at 5 minute intervals between 0105 UTC and 0500 UTC from Oklahoma Mesonet sites at (a-d) Marena (MARE), (e-h) Norman (NRMN), and (i-l) Grandfield (GRA2). Also shown are ensemble mean (thick colored lines) and 5<sup>th</sup> to 95<sup>th</sup> percentile range (shaded colored regions) for experiments RADCONV (purple; (a), (i), and (j)); RAD (blue; (b), (f), and (j)); CONV (red; (c), (g), and (k)); and CNTL (gray; (d), (h), and (l)) interpolated to the location of the corresponding Oklahoma Mesonet sites. The vertical dotted line in each panel denotes the end of the DA period at 0200 UTC.

Fig. 18. As Fig. 17, but for observed surface (2 m) dewpoint (solid black lines) and ensemble mean (thick colored lines) and the 5<sup>th</sup> to 95<sup>th</sup> percentile range within the ensemble (shaded colored regions) of dewpoint interpolated to the corresponding Oklahoma Mesonet site.

Fig. 19. Neighborhood ensemble probability (shaded) of radar reflectivity exceeding 25 dBZ,  $P[Z > 25 \text{ dBZ}]$ , at model grid level 10 (approximately 2 km above the surface) for (a-c) CONV, and (d-f) a variant of CONV using reduced localization radii for conventional data assimilation, at 0300, 0400, and 0500 UTC. The region of radar reflectivity exceeding 25 dBZ observed by the WSR-88D radar network at the corresponding time is outlined by a bold black contour. Urban boundaries are shown in purple. The localization radius used to assimilate surface and upper-air data is noted in

each panel.

Fig. 20. Ensemble-based probability of a significant near-surface mesovortex occurring within 25 km of a point (shaded) at 0400 UTC for (a) RADCONV, (b) a variant of RADCONV using reduced localization radii for conventional data assimilation, (c) CONV, and (d) a variant of CONV using reduced localization radii for conventional data assimilation. The location of the observed tornadic mesovortex at 0400 UTC is indicated by the black triangle in each panel. Urban boundaries are shown in purple. The localization radius used to assimilate surface, upper-air, and (if used) radar observations, is noted in each panel.

Table 1. Assumed observation error magnitude for conventional observations.

<b>Data Type</b>	<b>u (ms<sup>-1</sup>)</b>	<b>v (ms<sup>-1</sup>)</b>	<b>Temperature (K)</b>	<b>Dewpoint (K)</b>	<b>Pressure (hPa)</b>
<b>Surface</b>	1.5	1.5	1.5	2.0	2.0
<b>Upper-air</b>	2.5	2.5	1.2	2.0	0.6
<b>Profiler</b>	2.5	2.5			

Table 2. Summary of experiments.

Experiment Name	Radar Data Used?		Conventional Data Used?	
	Outer Domain	Inner Domain	Outer Domain	Inner Domain
<b>RADCONV</b>	NO	YES	YES	YES
<b>RAD</b>	NO	YES	YES	NO
<b>CONV</b>	NO	NO	YES	YES
<b>CNTL</b>	NO	NO	YES	NO

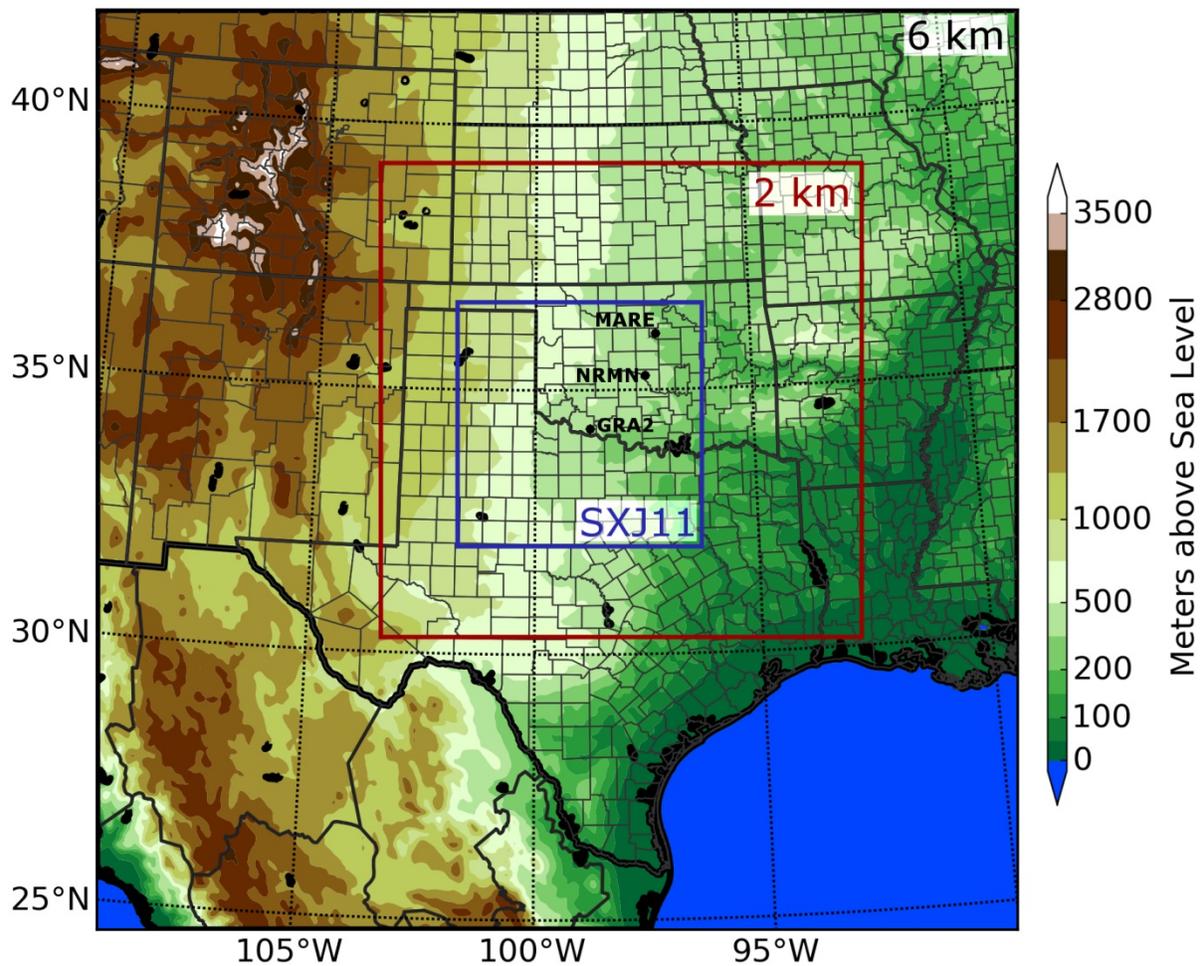


Fig. 1. Geographic extent of the outer model domain (6 km horizontal grid spacing) and the nested inner domain (2 km horizontal grid spacing). Surface elevation (in meters above mean sea level) is plotted for reference. Also shown for reference is the smaller 2-km domain used in SXJ11 and SXJ12. The three black dots in western and central Oklahoma indicate the locations of Oklahoma Mesonet stations MARE (Marena), NRMN (Norman), and GRA2 (Grandfield) used for timeseries verification.

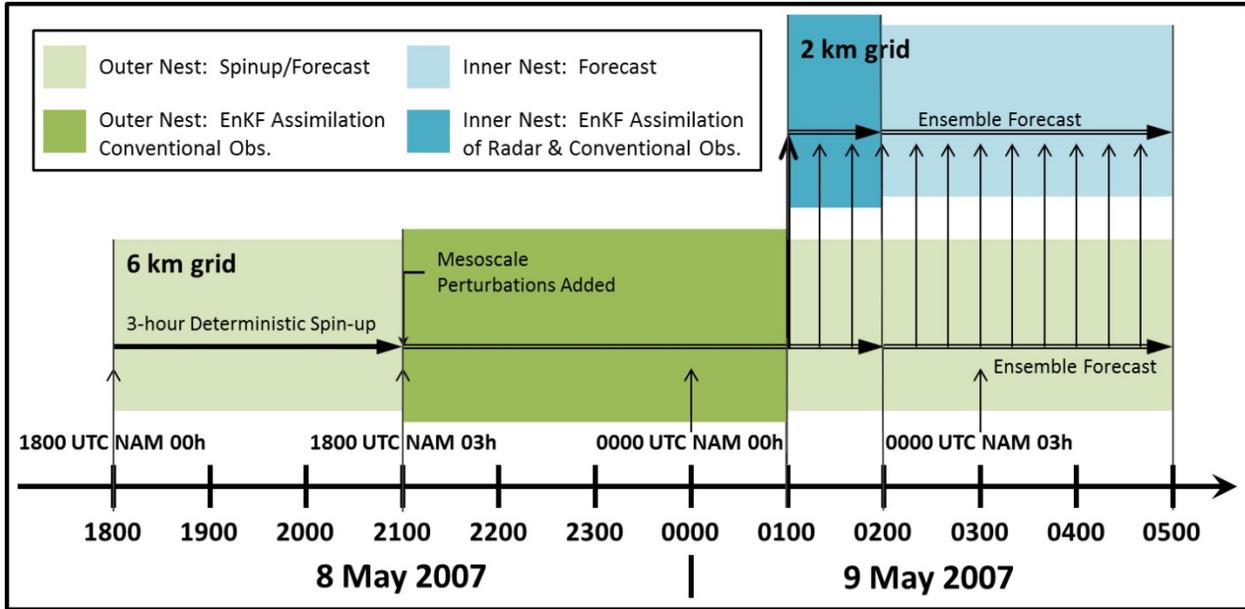


Fig. 2. Flow diagram for forecast experiments. The outer nest forecast is initialized at 1800 UTC on 8 May 2007 via interpolation from the 1800 UTC NAM analysis; 6-hourly NAM analyses and the intervening 3-hour forecasts are used as boundary conditions for the outer nest. The inner nest is initialized at 0100 UTC on 9 May 2007 using the outer nest ensemble for initial and boundary conditions.

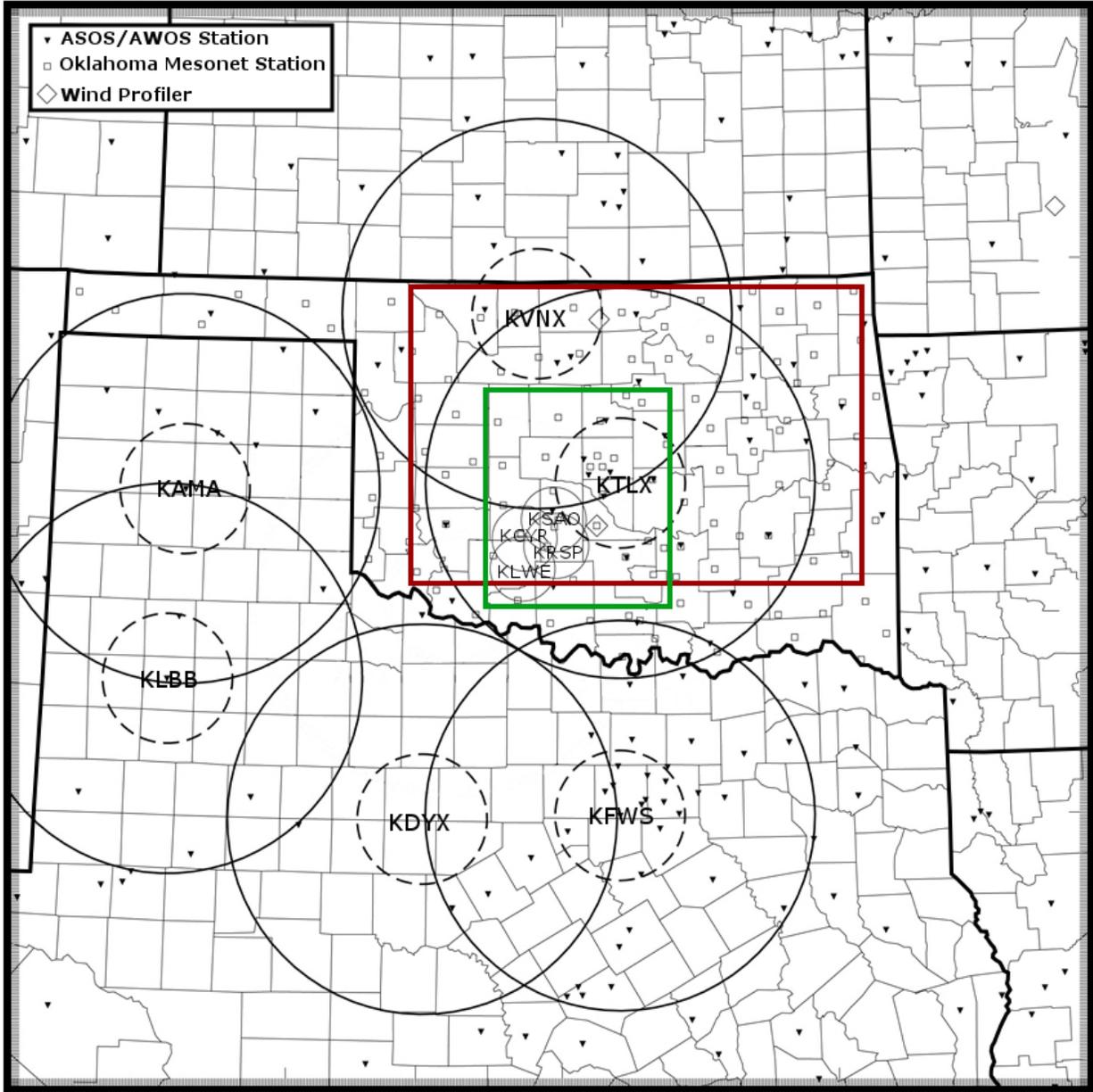


Fig. 3. Observations assimilated using EnKF on the inner nested grid (2 km grid spacing). The dashed circles and large solid circles indicate 50 and 150 km radius range rings, respectively, for WSR-88D radar sites used. Small, thin circles indicate 30 km range rings for CASA X-band radar sites used, black triangles indicate ASOS and AWOS surface station sites, squares indicate Oklahoma Mesonet station sites, and diamonds indicate wind profiler sites. The red box indicates the Oklahoma verification subdomain; the smaller green box indicates the verification subdomain used in SXJ12.

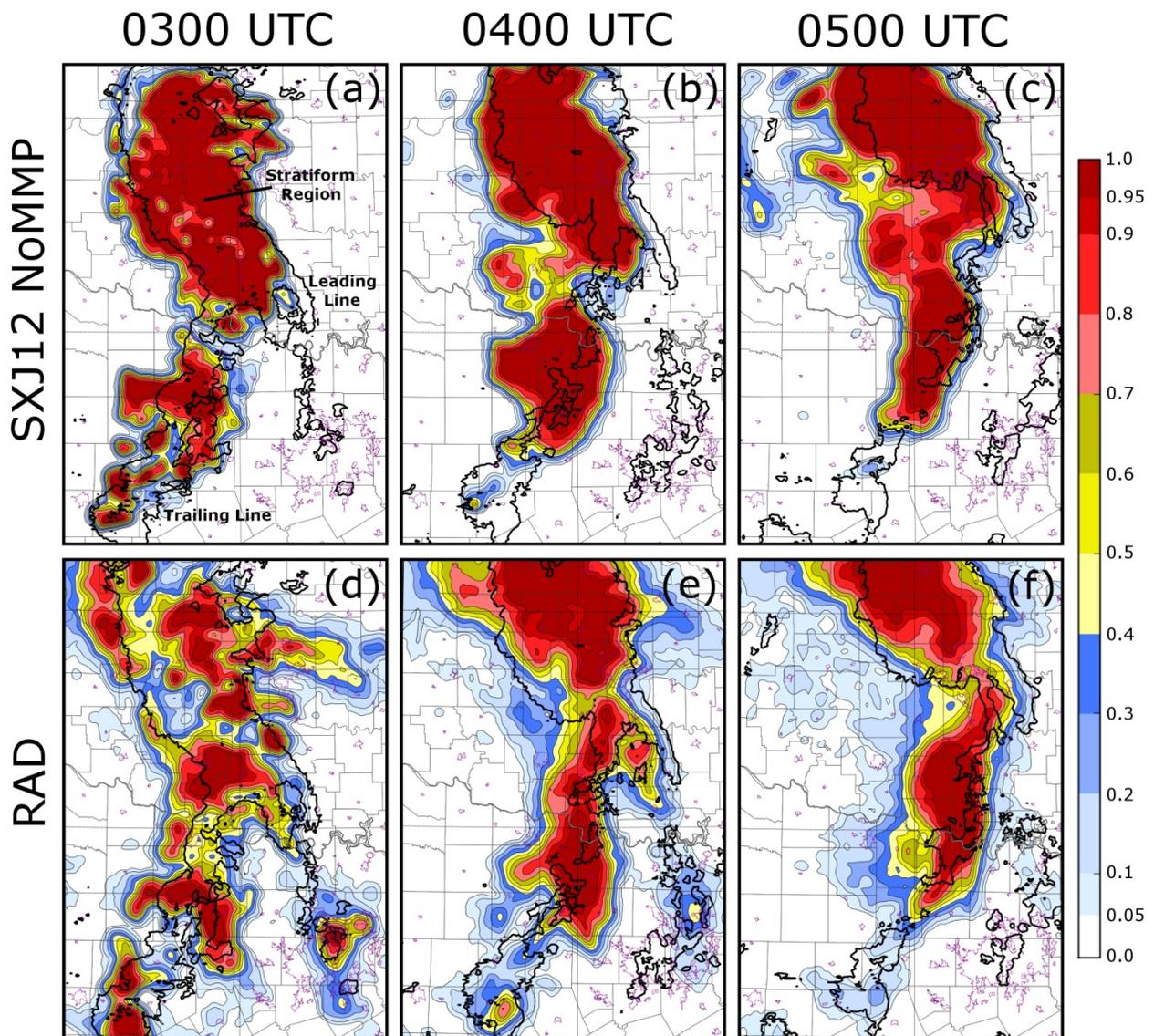


Fig. 4. Neighborhood ensemble probabilities (shaded) of reflectivity exceeding 25 dBZ,  $P[Z > 25 \text{ dBZ}]$ , at model grid level 10 (approximately 2 km above the surface) for experiment NoMMP of SXJ12 at (a) 0300 UTC, (b) 0400 UTC, and (c) 0500 UTC, and RAD at (d) 0300 UTC, (e) 0400 UTC, and (f) 0500 UTC. The bold black line in each panel indicates the location of the 25 dBZ radar reflectivity contour observed by the WSR-88D radar network. The leading convective line, trailing convective line, and stratiform region of the MCS are indicated in panel (a).

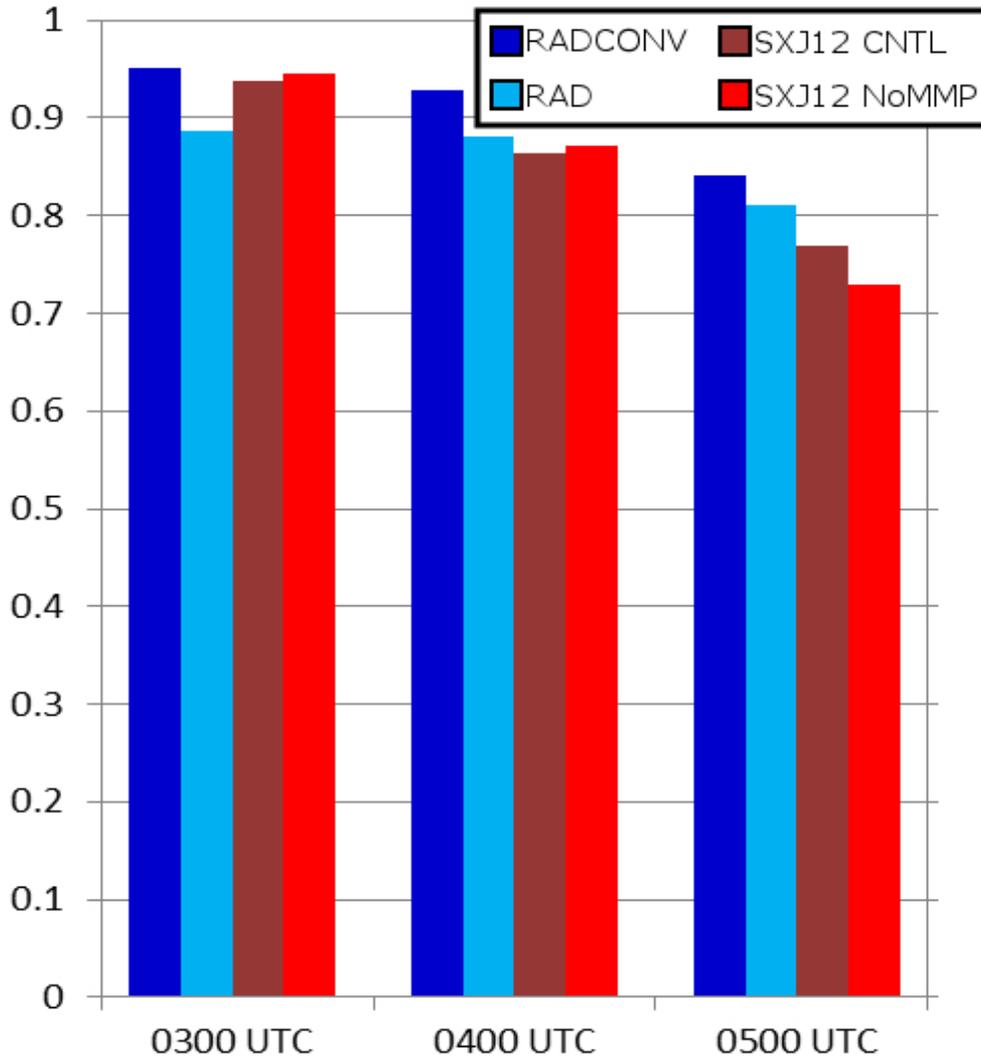


Fig. 5. Area under the ROC curve (AUC) for RADCONV, RAD, and two experiments from SXJ12 (CNTL and NoMMP) at 0300, 0400, and 0500 UTC for 1-, 2-, and 3-hour forecasts of radar reflectivity at the 25 dBZ threshold on vertical grid level  $k = 10$  (slightly more than 2 km above mean sea level) calculated over the verification subdomain of SXJ12 (the green box in Fig. 2).

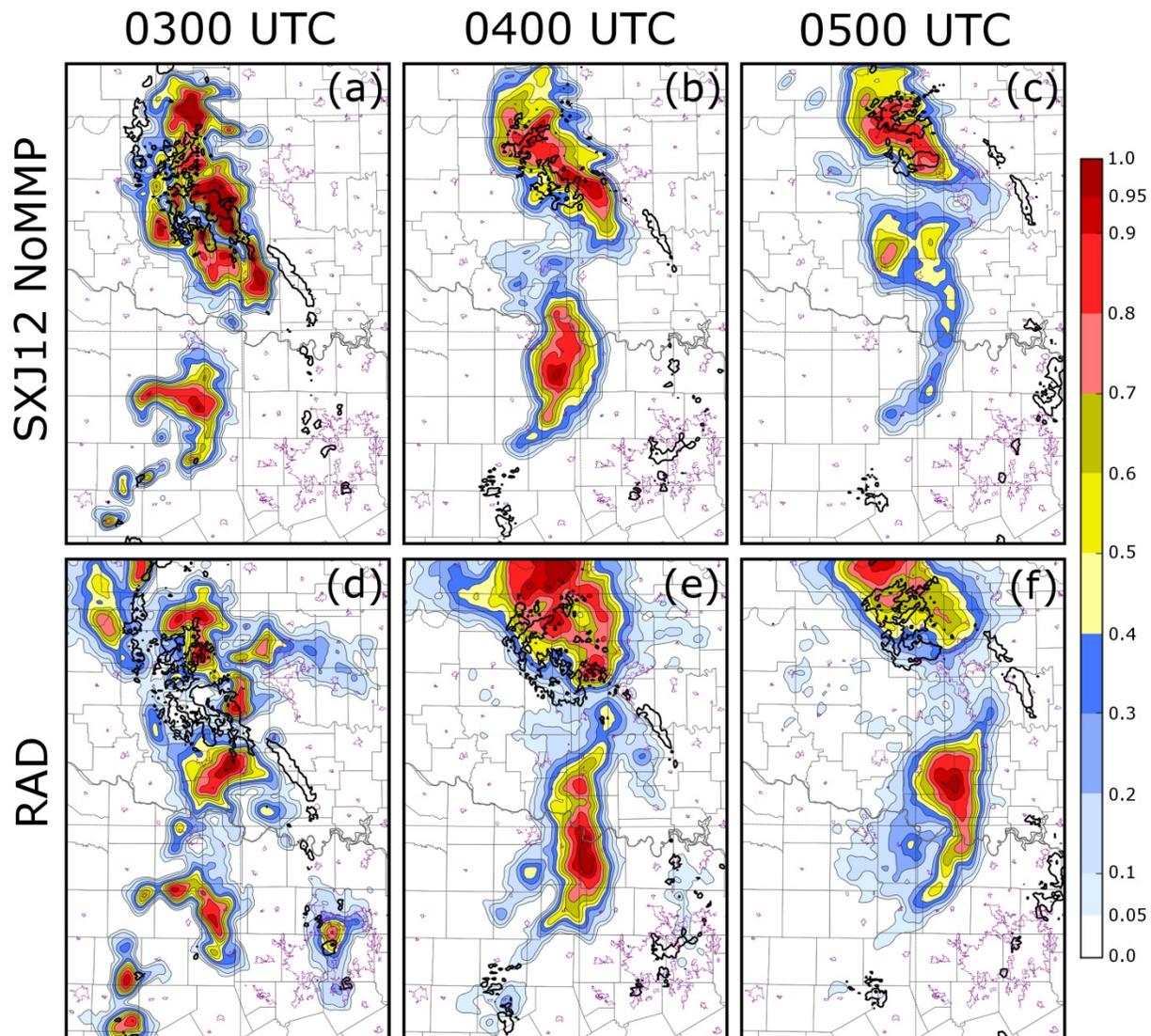


Fig. 6. As Fig. 4, but for  $P[Z > 40 \text{ dBZ}]$  and the 40 dBZ radar reflectivity contour.

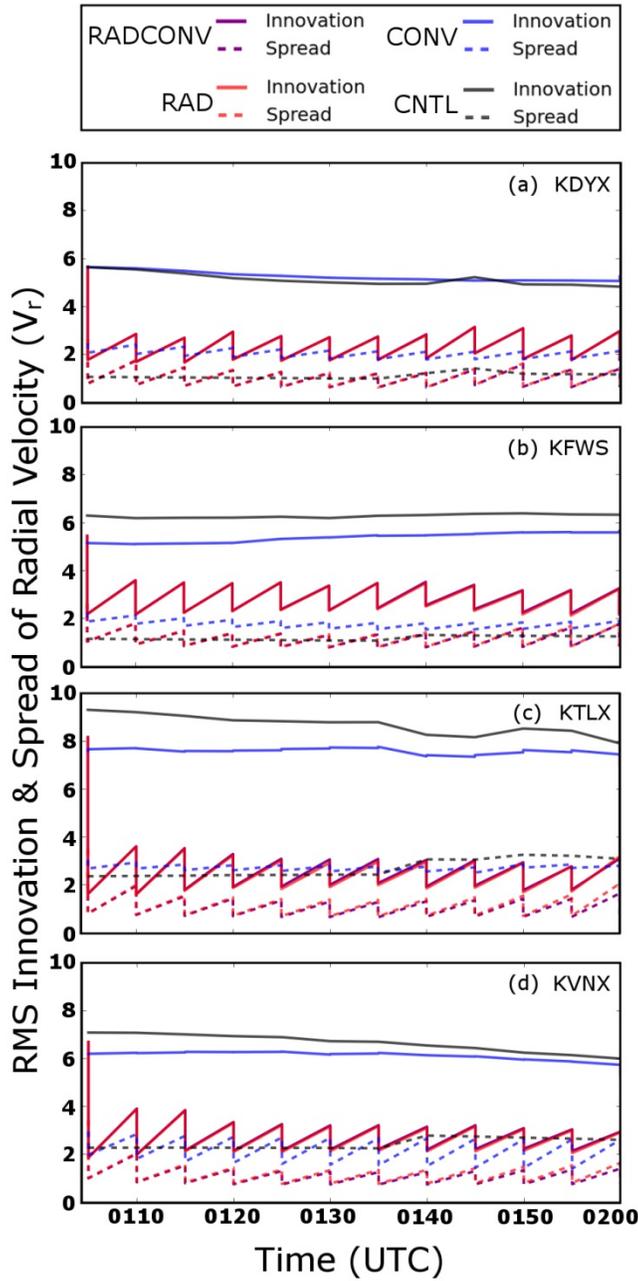


Fig. 7. Average root-mean-square (RMS) innovation (solid lines) of ensemble mean and the ensemble spread (dotted lines) of radial velocity ( $\text{m s}^{-1}$ ) over the observation region of four WSR-88D radars within the model domain from 0110 to 0200 UTC for all experiments. Calculations are limited to locations where observed and/or model (ensemble mean) reflectivity exceeds 15 dBZ.

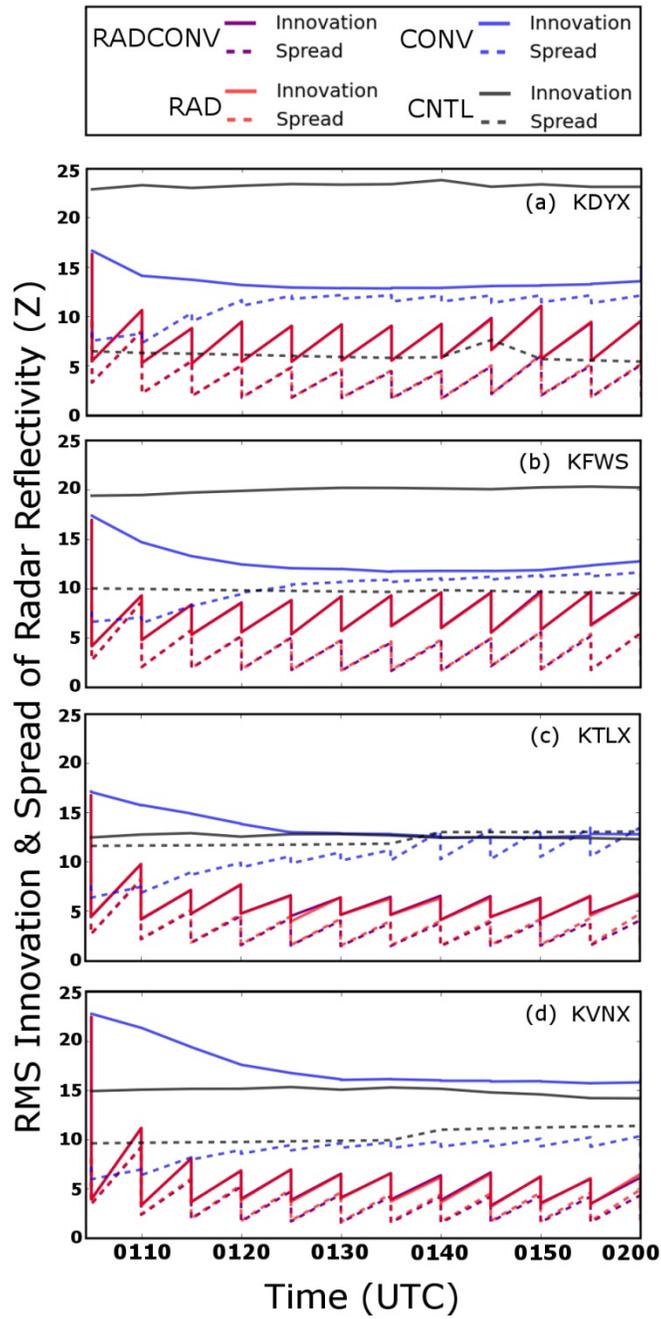


Fig. 8. As Fig. 7 but for radar reflectivity (dBZ) instead of radial velocity.

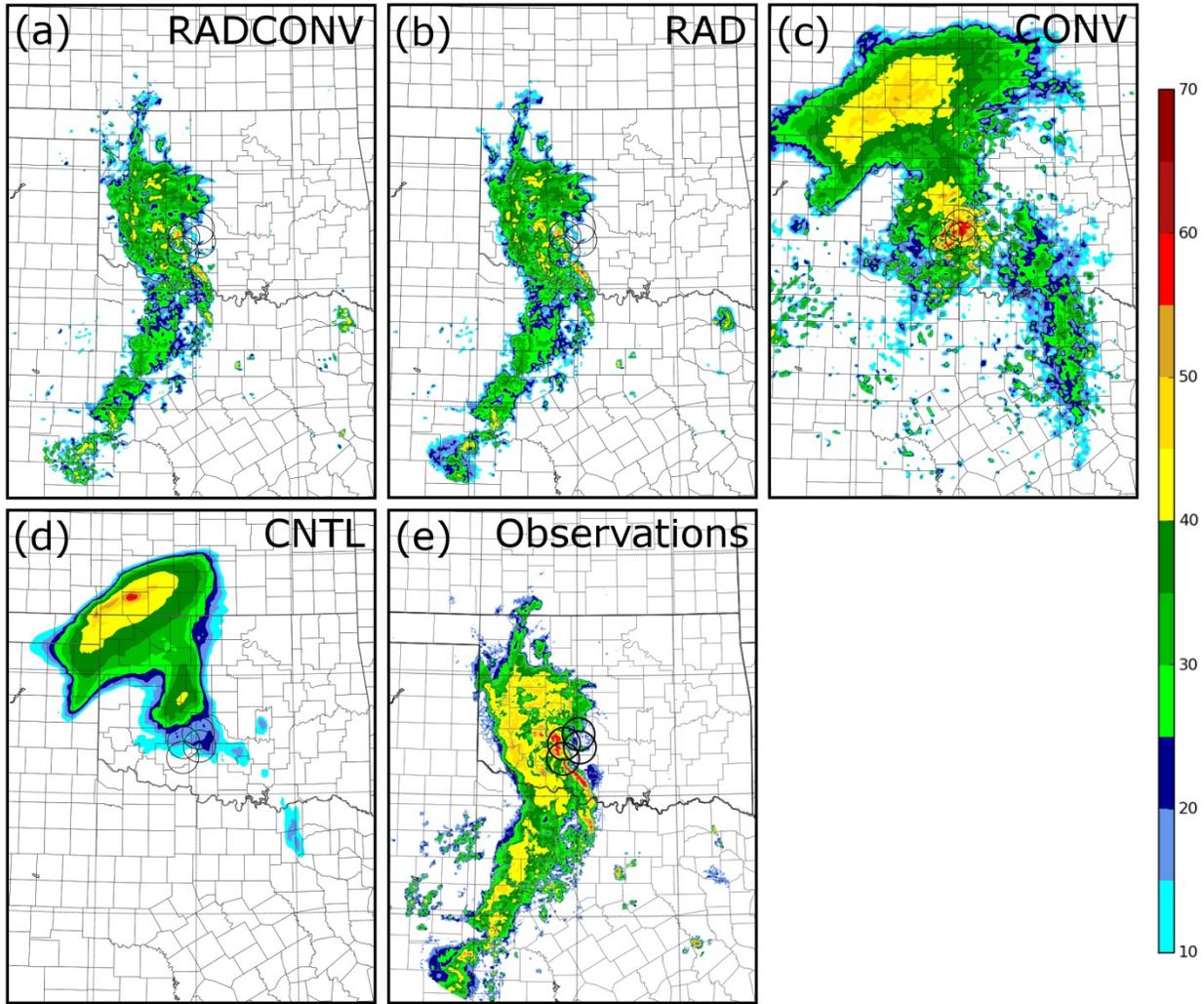


Fig. 9. Probability-matched ensemble mean reflectivity at model grid level 10 (approximately 2 km above the surface) for the 0200 UTC ensemble analyses of (a) RADCONV, (b) RAD, (c) CONV, and (d) CNTL. Also shown is (e) observed reflectivity at 0200 from the WSR-88D network, interpolated to the model grid.

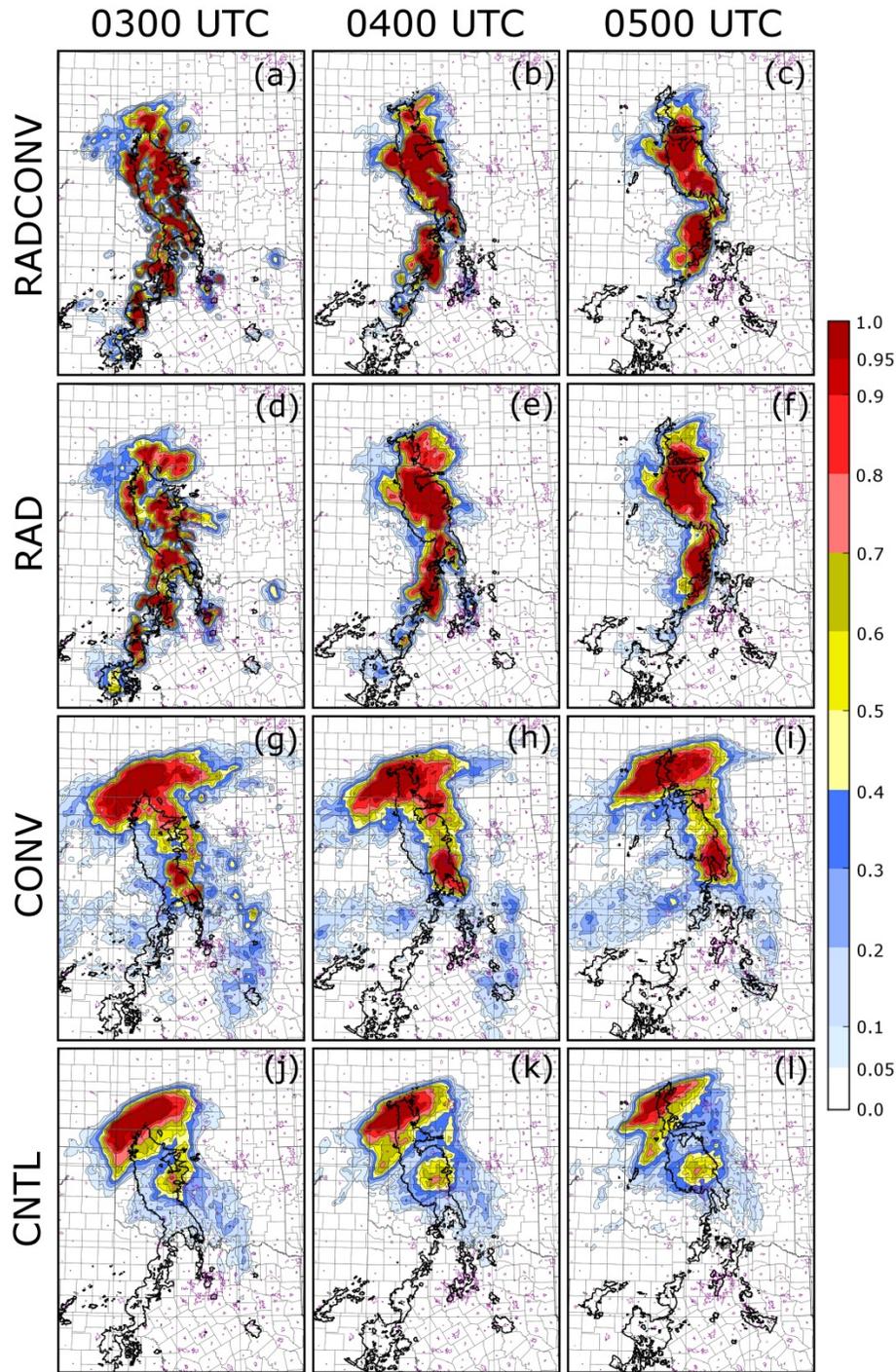


Fig. 10. Neighborhood ensemble probabilities (shaded) of radar reflectivity exceeding 25 dBZ,  $P[Z > 25 \text{ dBZ}]$ , at model grid level 10 (approximately 2 km above the surface) for (a-c) RADCONV, (d-f) RAD, (g-i) CONV, and (j-l) CNTL at 0300, 0400, and 0500 UTC. The region of radar reflectivity exceeding 25 dBZ observed by the WSR-88D radar network at the corresponding time is outlined by a bold black contour. Urban boundaries are shown in purple.

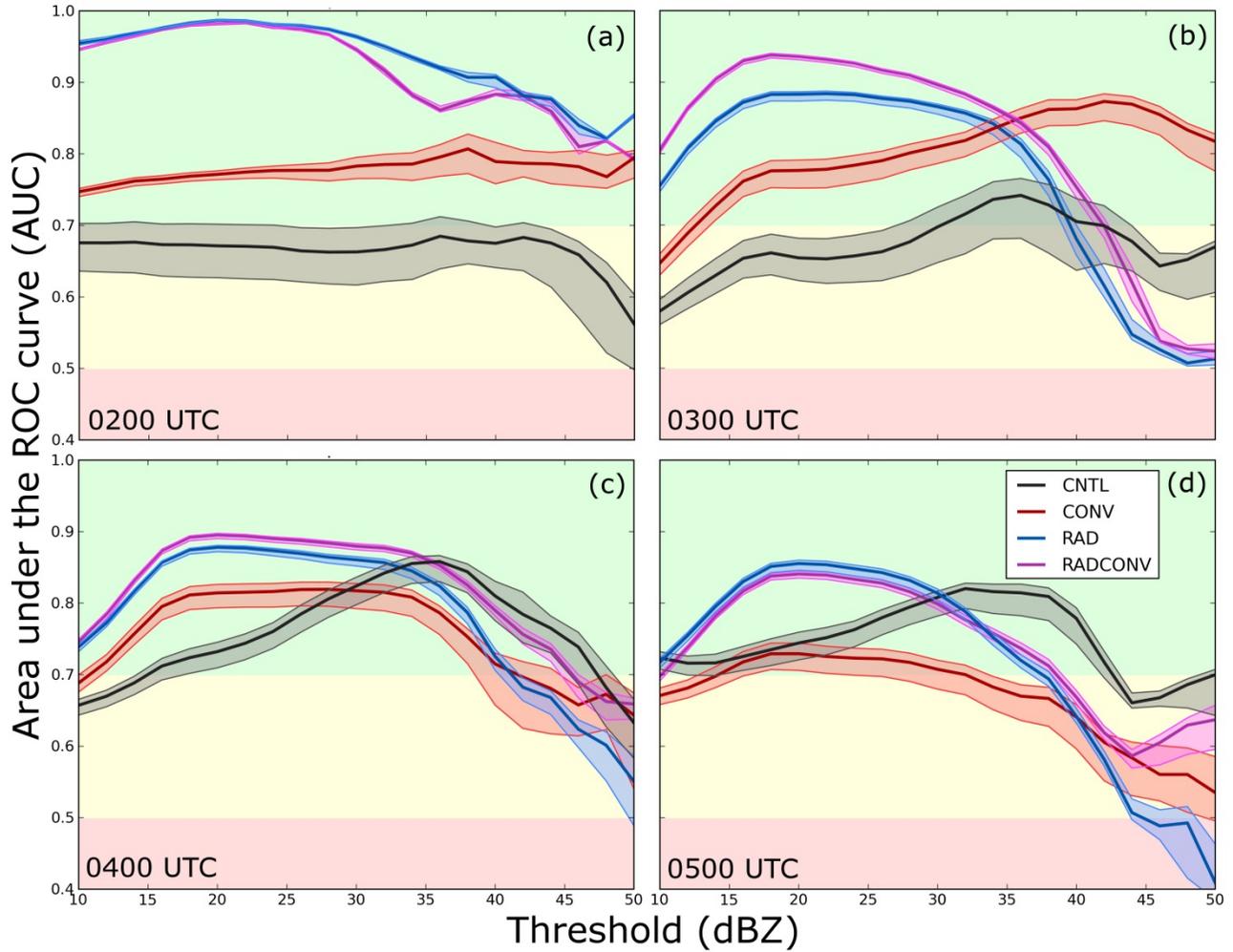


Fig. 11. Area under the relative operating characteristic (ROC) curve (AUC) (solid, bold lines) for all experiments for the analyses at (a) 0200 UTC, and forecasts at (b) 0300, (c) 0400, and (d) 0500 UTC for forecasts of radar reflectivity at vertical grid level  $k = 10$  (slightly more than 2 km above mean sea level) exceeding threshold values ranging from 10 to 50 dBZ at intervals of 2 dBZ. Also shown are 90% confidence intervals calculated using a 1000-member bootstrap to resample the ensemble (shaded regions). Calculations are performed over the Oklahoma verification subdomain (the red box in Fig. 2). In each panel, the green region indicates AUC values associated with an operationally-useful forecast ( $AUC > 0.7$ ). The red region indicates forecasts with no skill ( $AUC < 0.5$ ).

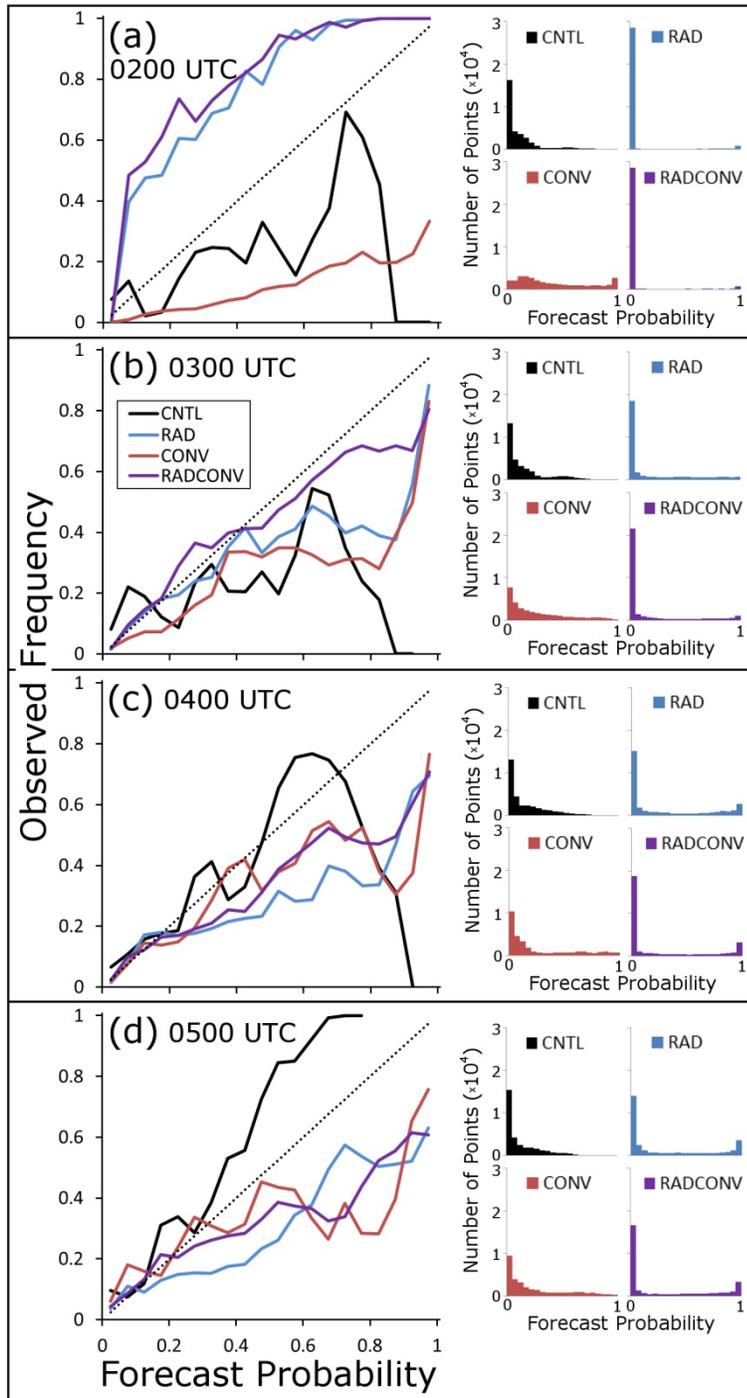


Fig. 12. Reliability and sharpness diagrams for NEP forecasts of  $P[Z > 25 \text{ dBZ}]$  for all experiments at (a) 0200 UTC, (b) 0300 UTC, (c) 0400 UTC, and (d) 0500 UTC calculated over the Oklahoma verification subdomain (the red box in Fig. 2). Forecast probability bins are spaced at intervals of 0.05.

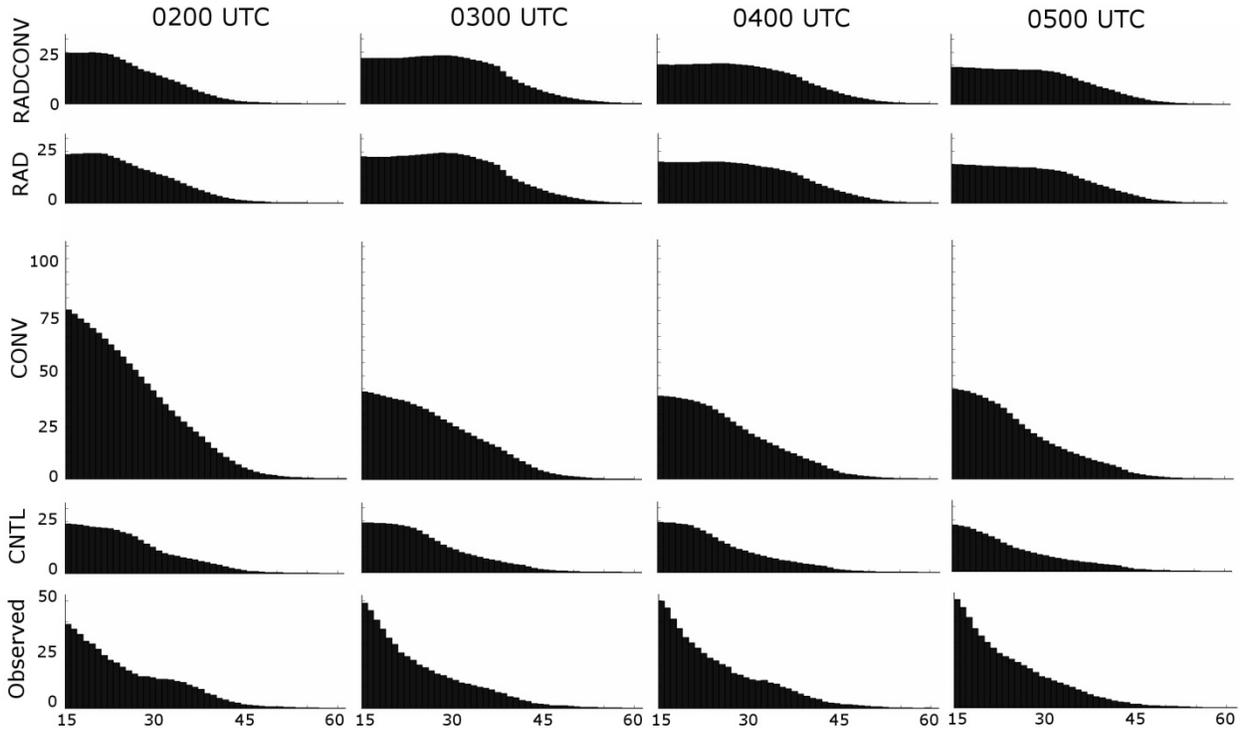


Fig. 13. Hourly, domain-wide histograms of forecast radar reflectivity for all ensemble forecast experiments, compared to WSR-88D observed radar reflectivity interpolated to the ensemble forecast grid (bottom row). Bins are placed every 1 dBZ. The vertical axis indicates the number of model grid volumes within each bin, normalized by the size of the forecast ensemble.

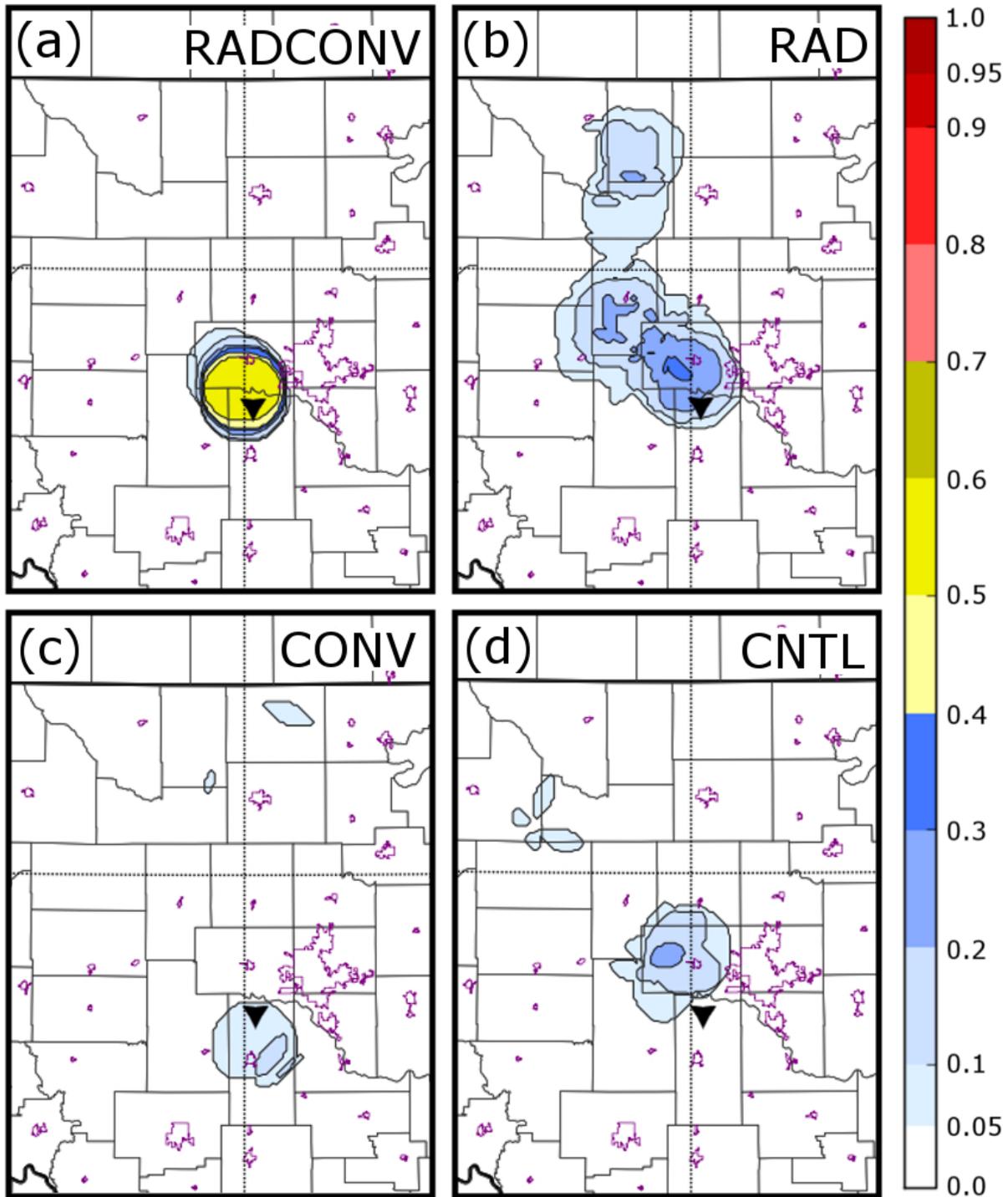


Fig. 14. Ensemble-based probability of a significant near-surface mesovortex occurring within 25 km of a point (shaded) at 0400 UTC for (a) RADCONV, (b) RAD, (c) CONV, and (d) CNTL. The location of the observed tornadic mesovortex (located within the line-end vortex of the MCS) at 0400 UTC is indicated by the black triangle in each panel. Urban boundaries are shown in purple.

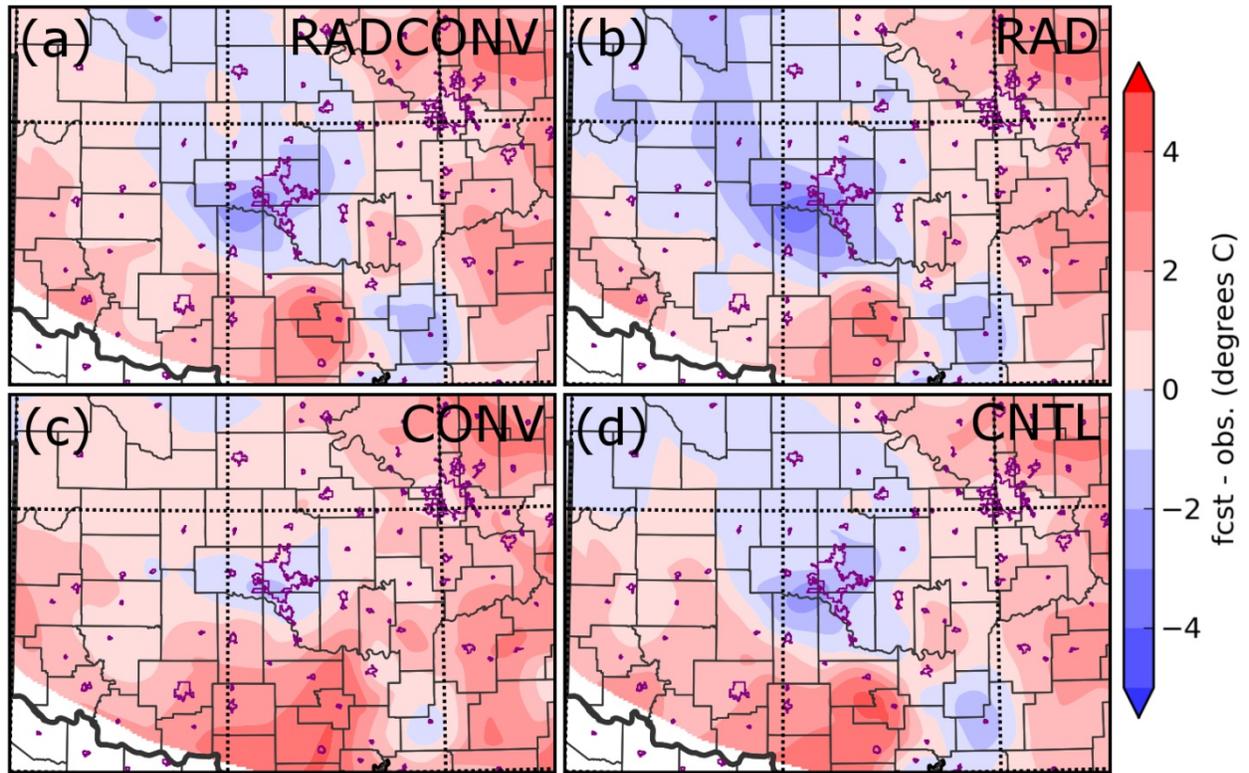


Fig. 15. Contours of the difference between ensemble mean surface (2 m) temperature forecast and Oklahoma Mesonet observations at 0400 UTC for (a) RADCONV, (b) RAD, (c) CONV, and (d) CNTL. Red areas indicate an ensemble mean temperature warmer than observed by the mesonet, while blue areas indicate an ensemble mean colder than observations.

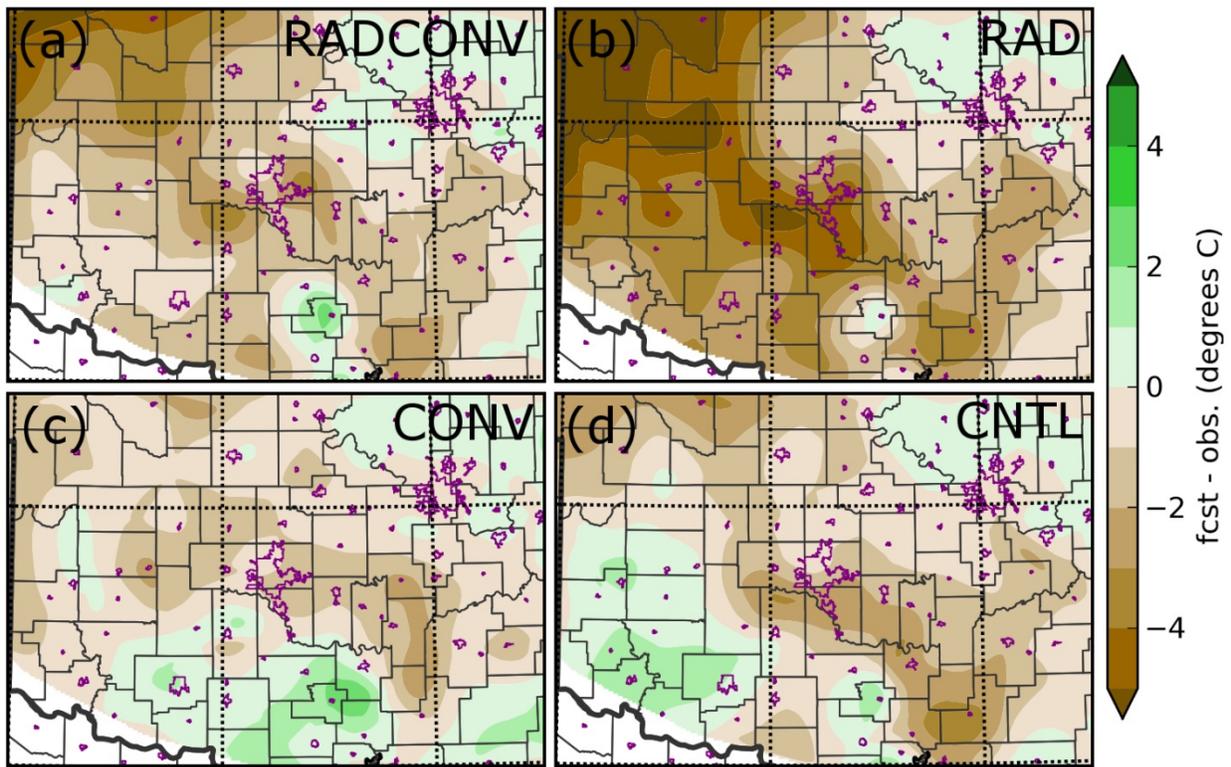


Fig. 16. Contours of the difference between ensemble mean surface (2 m) dewpoint forecast and Oklahoma Mesonet observations at 0400 UTC for (a) RADCONV, (b) RAD, (c) CONV, and (d) CNTL. Green areas indicate an ensemble mean dewpoint higher than observed by the mesonet, while brown areas indicate an ensemble mean dewpoint lower than observations.

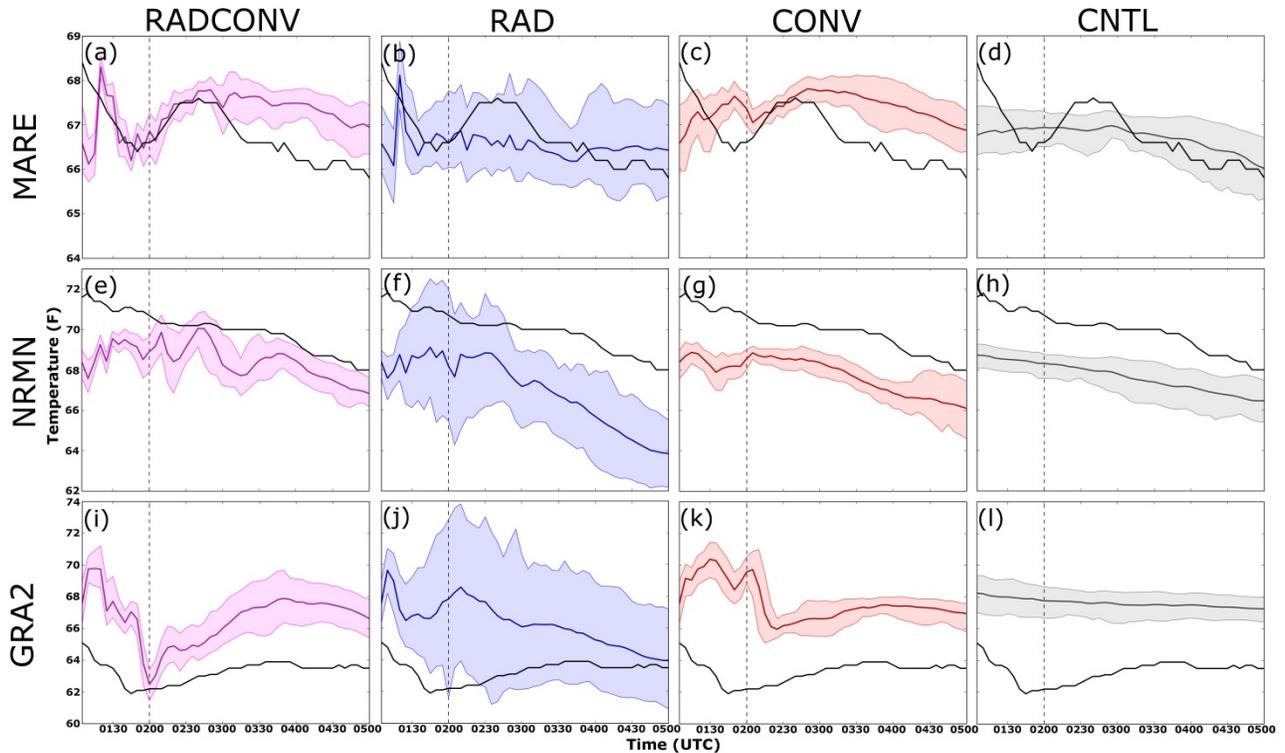


Fig. 17. Observed surface (2 m) temperature (F) (solid black lines) at 5 minute intervals between 0105 UTC and 0500 UTC from Oklahoma Mesonet sites at (a-d) Marena (MARE), (e-h) Norman (NRMN), and (i-l) Grandfield (GRA2). Also shown are ensemble mean (thick colored lines) and 5<sup>th</sup> to 95<sup>th</sup> percentile range (shaded colored regions) for experiments RADCONV (purple; (a), (i), and (j)); RAD (blue; (b), (f), and (j)); CONV (red; (c), (g), and (k)); and CNTL (gray; (d), (h), and (l)) interpolated to the location of the corresponding Oklahoma Mesonet sites. The vertical dotted line in each panel denotes the end of the DA period at 0200 UTC.

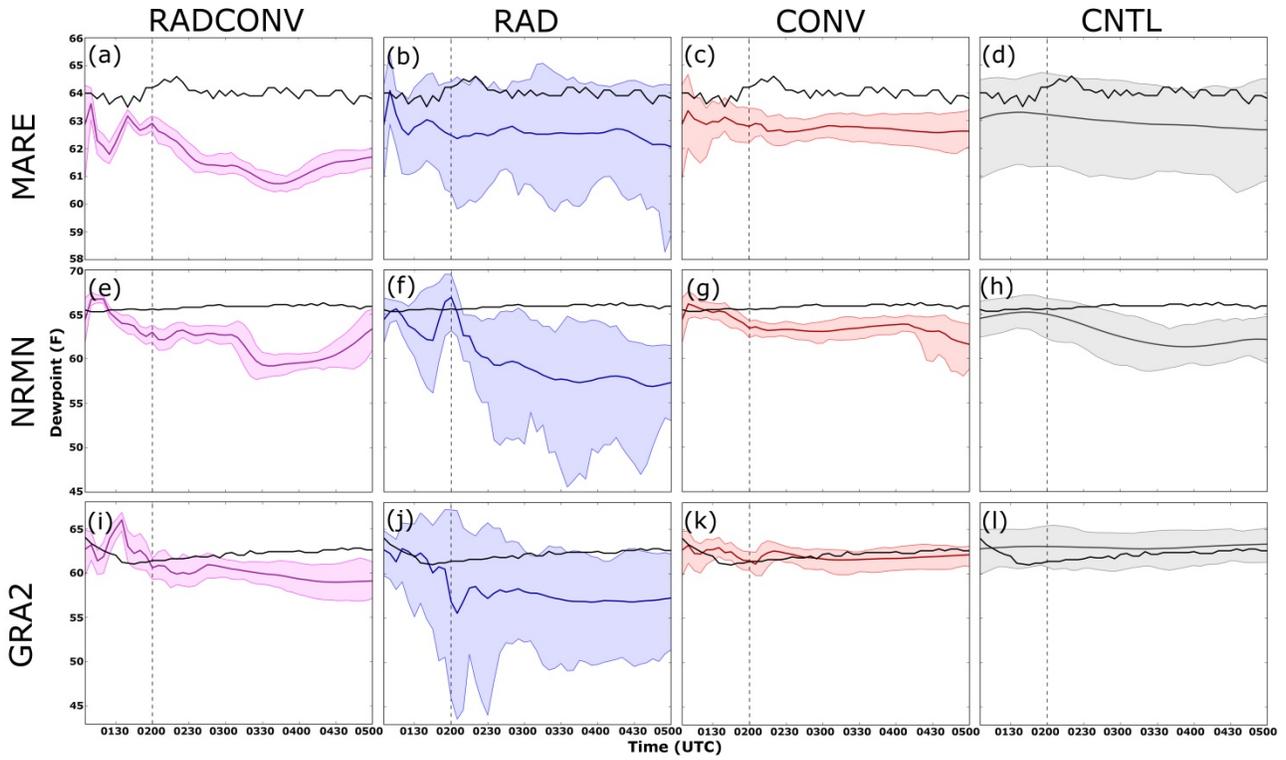


Fig. 18. As Fig. 17, but for observed surface (2 m) dewpoint (solid black lines) and ensemble mean (thick colored lines) and the 5<sup>th</sup> to 95<sup>th</sup> percentile range within the ensemble (shaded colored regions) of dewpoint interpolated to the corresponding Oklahoma Mesonet site.

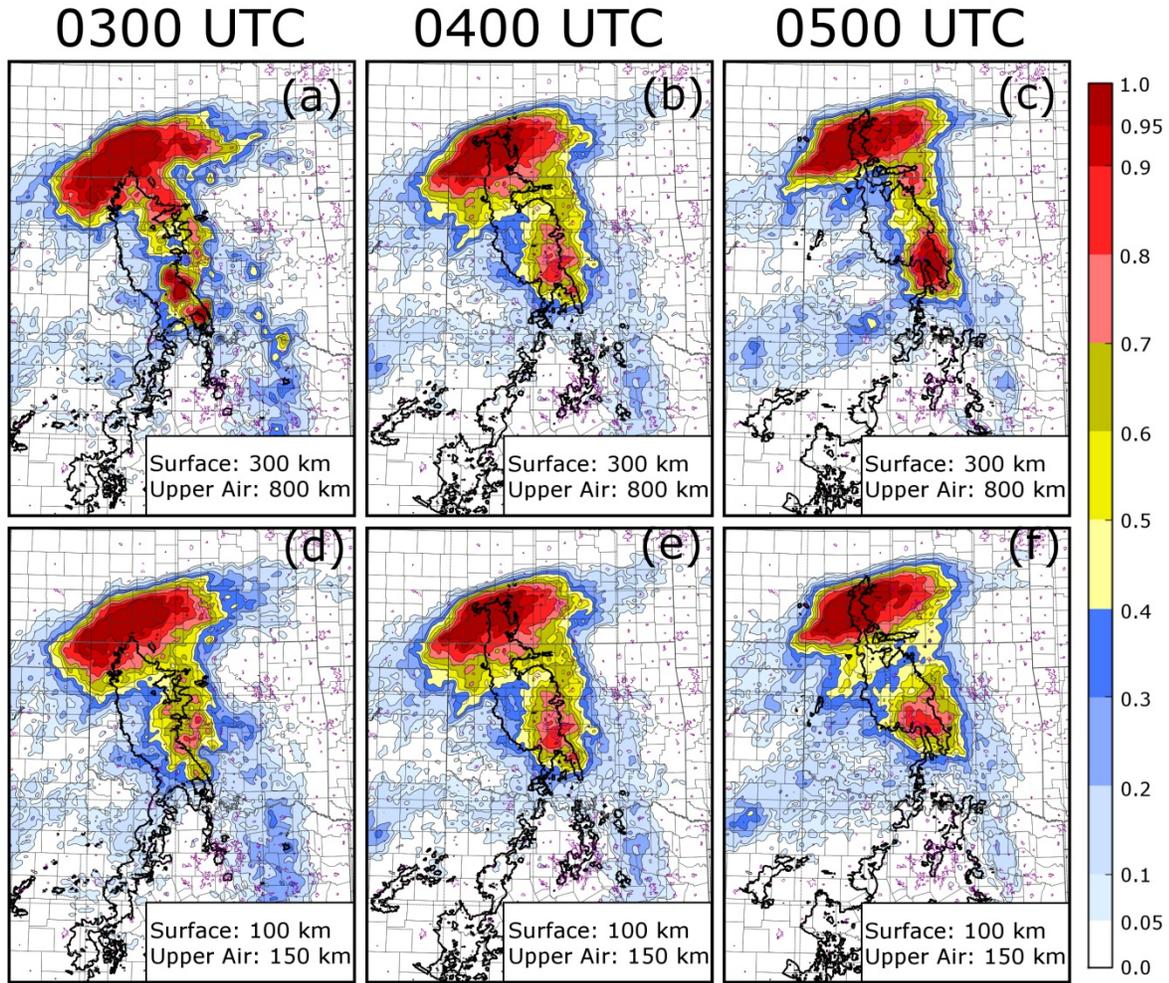


Fig. 19. Neighborhood ensemble probability (shaded) of radar reflectivity exceeding 25 dBZ,  $P[Z > 25 \text{ dBZ}]$ , at model grid level 10 (approximately 2 km above the surface) for (a-c) CONV, and (d-f) a variant of CONV using reduced localization radii for conventional data assimilation, at 0300, 0400, and 0500 UTC. The region of radar reflectivity exceeding 25 dBZ observed by the WSR-88D radar network at the corresponding time is outlined by a bold black contour. Urban boundaries are shown in purple. The localization radius used to assimilate surface and upper-air data is noted in each panel.

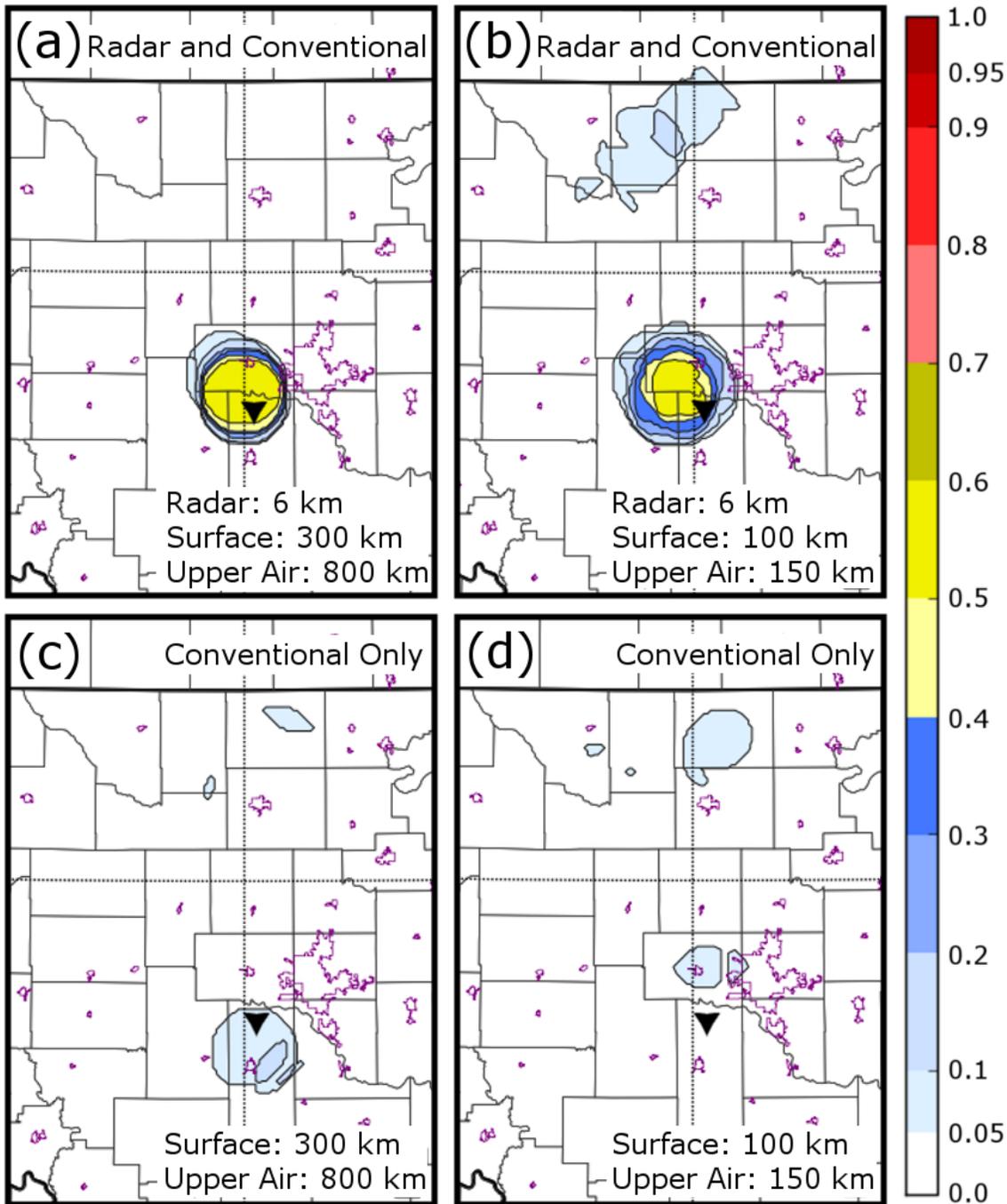


Fig. 20. Ensemble-based probability of a significant near-surface mesovortex occurring within 25 km of a point (shaded) at 0400 UTC for (a) RADCONV, (b) a variant of RADCONV using reduced localization radii for conventional data assimilation, (c) CONV, and (d) a variant of CONV using reduced localization radii for conventional data assimilation. The location of the observed tornadic mesovortex at 0400 UTC is indicated by the black triangle in each panel. Urban boundaries are shown in purple. The localization radius used to assimilate surface, upper-air, and (if used) radar observations, is noted in each panel.