AN INVESTIGATION OF THE CONDITIONAL PRACTICAL PREDICTABILITY OF THE 31 MAY 2013 HEAVY-RAIN-PRODUCING MESOSCALE CONVECTIVE SYSTEM

by

Aidan Kuroski

A Thesis Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Master of Science

in Atmospheric Science

at

The University of Wisconsin-Milwaukee

August 2018

ABSTRACT

AN INVESTIGATION OF THE CONDITIONAL PRACTICAL PREDICTABILITY OF THE 31 MAY 2013 HEAVY-RAIN-PRODUCING MESOSCALE CONVECTIVE SYSTEM

by

The University of Wisconsin-Milwaukee, 2018
Under the Supervision of Professor Clark Evans

On 31 May 2013, strong thunderstorms initiated in west-central Oklahoma with one of the storms eventually creating a very strong tornado near El Reno, OK. The storms then grew upscale into a quasi-stationary mesoscale convective system that produced prolonged heavy rainfall that led to severe flooding across parts of Oklahoma, including the Oklahoma City metropolitan area. A 50-member ensemble of short range (0-24 h) forecasts was conducting using a set of initial conditions generated via cycled data assimilation to quantify event predictability and identify forecast sensitivities, primarily with CI and initial upscale growth. Both a composite and ensemble sensitivity analysis were performed using fractions skill score (FSS) as the partitioning forecast metric to measure these dependencies. Both analyses indicate that better forecasts are associated with an upper-level trough further to the west, stronger upper-level ridging to the east, stronger low-level convergence across much of Oklahoma, and a further west dryline and quasi-stationary front across northwest Oklahoma and Kansas. These conditions were associated with better early-stage MCS forecasts, which appears to be associated with better CI forecasts. These findings suggest that even slight differences in conditions leading to CI had large sensitivities in CI and the subsequent upscale growth.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to thank my advisor, Clark Evans, as I needed his help through most of these two years and I cannot thank him enough for sticking with me through everything and could not have done most of this without him. Secondly, I would like to thank my peers and friends in the mesoscale research group, Kevin and David, for all of their help with different periods of research and classwork. I would also like to thank my committee members Professors Paul Roebber and ??? for all of their helpful insight. Lastly, I would like to thank my family, closest friends, and classmates for all of their support throughout the years as well as all the memories.

1. Introduction

Shortly before 2200 UTC on 31 May 2013, a cluster of thunderstorms developed ahead of an upper-tropospheric trough and associated jet streak (Fig. 1), along a quasi-stationary surface frontal boundary (Fig. 2), and in an environment of strong surface-based instability and large westerly lower-middle tropospheric vertical wind shear (Fig. 3) in west-central Oklahoma. The strongest of these storms became a supercell thunderstorm and produced a large tornado near El Reno, Oklahoma. Subsequently, beginning just after 0000 UTC on 1 June, this storm grew upscale into an east-west oriented, quasi-stationary mesoscale convective system (MCS; Houze 2004) over central Oklahoma. Over the next 7 h, as the MCS slowly progressed eastward and developed backbuilding (Schumacher and Johnson 2005) characteristics, rainfall of over 175 mm was observed in much of central Oklahoma. This caused flash flooding in many areas, including Oklahoma City, resulting in twenty-two fatalities between the tornado (eight) and flash flooding (fourteen) portions of the event (NWS 2014; https://www.weather.gov/oun/events-20130531).

A couple of studies quantify the predictability of this event at very-short lead times (0-3 h). For instance, the initiation of this event in convection-allowing ensemble forecasts is sensitive to mesoscale uncertainty in lower- to middle-tropospheric wind, temperature, and moisture (Hitchcock et al. 2016), as manifest in the forecast representation of the frontal boundary that served as the triggering mechanism for deep, moist convection initiation (CI; e.g., Weckwerth and Parsons 2006). Assimilating rawinsonde observations collected in the pre- and near-storm environments during the Mesoscale Predictability Experiment (MPEX; Weisman et al. 2015, Trapp et al. 2016) improves CI location forecasts within 3 h of initiation, although this finding is not generalizable to other MPEX events (Hitchcock et al. 2016). Likewise, assimilating conventional radar data early in the event increases forecast skill at 0-3 h forecast lead times in a

separate ensemble of convection-allowing forecasts in which radar data are not assimilated (Yussouf et al. 2016). At longer leads (e.g., forecasts issued between 6-24 h prior to CI for this event), a cursory examination of operational Storm Prediction Center and convection-allowing model forecasts suggests that predictability of this event was relatively high, with minor uncertainty in the persistence of isolated CI preceding upscale growth and in the location (northern versus central Oklahoma) where upscale growth and backbuilding would later occur (not shown).

Atmospheric variability on the order of observational uncertainty in both thermodynamic and kinematic fields, particularly in the lower troposphere, can impact where, when, or if CI occurs (e.g., Crook 1996). Indeed, arguably "the most difficult prediction task is to foresee development of the first convective storm in an area" (Lilly 1990). CI predictability is fundamentally limited by an inability to accurately analyze or predict the atmospheric state across scales (e.g., Duda and Gallus 2013; Burghardt et al. 2014; Burlingame et al. 2017; Keclik et al. 2017). For instance, CI is sensitive to the meso-alpha- to synoptic-scale thermodynamic and kinematic environment, such as may be manifest through frontal boundary or dryline placement (e.g., Weisman et al. 2008, Duda and Gallus 2013, Hill et al. 2016, Burlingame et al. 2017, Keclik et al. 2017). However, the meso-beta- to microscales are also important for CI timing and location, such as may be manifest through local variability in planetary boundary layer (PBL) lifting and moistening (e.g., Crook 1996, Weckwerth 2000, Markowski et al. 2006, Weckwerth et al. 2008, Kain et al. 2013, Duda and Gallus 2013, Hill et al. 2016). Consequently, although convection-allowing models are able coarsely resolve the small-scale physical processes important for CI and thus provide skillful probabilistic CI forecasts, model-forecast CI appears to be a poor indicator of impending thunderstorm-related hazards such as organized MCSs (Kain et al. 2013).

Mesoscale convective systems can greatly impact society with heavy rain and damaging winds (Schumacher and Johnson 2005; Weisman 1992), yet like CI also have somewhat limited predictability (e.g., Wandishin et al. 2008, 2010, Melhauser and Zhang 2012). There is some suggestion that diabatic forcing associated with small-scale moist dynamics is of major importance to MCS predictability as a primary source of rapid error growth from small to large scales (e.g., Hohenegger et al. 2006, Hohenegger and Schär 2007, Melhauser and Zhang 2012). However, MCSs are also sensitive to small large-scale atmospheric initial uncertainties, with recent research suggesting that downscale growth of these uncertainties results in the aforementioned diabatically driven small-scale error growth (Durran and Weyn 2016, Weyn and Durran 2017). Uncertainties in near-surface fields (e.g., Hill et al. 2016), moisture and buoyancy profiles (e.g., Gilmore and Wicker 1998, Park 1999, Coniglio et al. 2007), and kinematic fields (e.g., Gilmore and Wicker 1998, Coniglio et al. 2007) are all known to contribute to limited MCS predictability.

Most MCS predictability research prescribes CI, such as through the use of a localized warm bubble near the surface in idealized simulations (Wandishin et al. 2008, 2010), or assumes CI in real-data simulations, each of which likely overstate MCS predictability. A limited number of studies examine MCS predictability when CI is uncertain. For example, model-forecast CI in west-central Oklahoma in the 31 May 2013 event discussed early is sensitive to horizontal grid spacing and planetary boundary layer parameterization (Schumacher 2015). For that specific model configuration, forecast skill for this event decreases with finer horizontal grid spacing, mainly due to the inability of the storms to initiate due to the development of a series of gravity waves that increases turbulent vertical mixing of moisture within the planetary boundary layer that affects dryline movement and lower-tropospheric moisture content (Schumacher 2015). Similarly, the 8 May 2009 “Super Derecho” case (Weisman et al. 2013; Evans et al. 2014) shows high practical

predictability for the MCS itself but minimal predictability for the associated mesovortex's intensity at lead times of 18-24 h (Grunzke and Evans 2017). Variability in the model-forecast occurrence of elevated CI over northwest Kansas during the observed MCS's initiation phase is identified as the primary distinguishing factor between more and less skillful forecasts of mesovortex intensity (Grunzke and Evans 2017). However, these studies consider only a portion of the predictability spectrum (initial condition uncertainty and a subset of potential sources of model error) for only two cases.

This research seeks to add to the limited conditional MCS predictability literature, focusing on quantifying the practical predictability and forecast sensitivities of a single case, the 31 May 2013 central Oklahoma event, without explicitly assuming or prescribing CI. It is hypothesized that the most important control upon MCS predictability for this event is atmospheric uncertainty in the pre-convective environment, particularly in proximity to the dryline and quasi-stationary surface frontal boundary along which CI occurs. Specifically, this research quantifies forecast sensitivities for the early-stage lifecycle (CI through upscale growth and initial backbuilding development) of the 31 May 2013 heavy-rain–producing MCS to atmospheric uncertainty within a 50-member convection-allowing ensemble, focusing on short forecast lead times (0-24 h). Consequently, the aim of this research is to improve the understanding of and ability to predict organized convective events, particularly high-impact cases that are difficult to forecast.

The remainder of this thesis is organized as follows. The experimental design, including specifics regarding the cycled analysis, free forecast, verification methods, composite generation, and sensitivity analysis, is documented in section 2. Ensemble assimilation quality and ensemble forecast sensitivities are quantified in section 3. Finally, a discussion, conclusions, and future work are presented in section 4.

2. Methodology

a. Cycled analysis

The configuration of the cycled ensemble data assimilation and forecast system follows closely to Grunzke and Evans (2017) and references therein. Briefly, cycled data assimilation is conducted using the Advanced Research version of the Weather Research and Forecasting model (WRF-ARW; Skamarock et al. 2008), version 3.9.1, coupled to the *lanai* release of the Data Assimilation Research Testbed (DART; Anderson and Collins 2007; Anderson et al. 2009) software. DART assimilates observations to update short-lead ensemble forecasts generated by WRF-ARW and generate ensemble initial conditions (ICs) using an ensemble adjustment Kalman filter. Cycling of the data assimilation system occurs every 6 h, with assimilation beginning at 1200 UTC 26 May 2013 and lasting until 1200 UTC 31 May 2013.

The cycled analysis uses 15-km horizontal grid spacing covering the conterminous United States and parts of Mexico, Canada, and the Atlantic and Pacific Oceans with 50 vertical levels extending to 50 hPa. Model parameterizations closely follow Romine et al. (2013, 2014) and Grunzke and Evans (2017) and include the Mellor-Yamada-Janjic (MYJ) planetary boundary layer scheme (Janjic 1994), the Rapid Radiative Transfer Model for Global Climate Models for both longwave and shortwave radiation (Iacono et al. 2008), the NOAH land surface model (Chen and Dudhia 2001), the Thompson microphysics scheme (Thompson et al. 2008), and the revised Tiedtke deep cumulus scheme (Tiedtke 1989; Zhang et al. 2011).

Initial and lateral boundary conditions (LBCs) at the beginning of the cycled analysis process are obtained from the 0.5° Global Forecast System (GFS) 0-h analysis and 6-h forecast from the 1200 UTC 26 May 2013 model cycle. These data are randomly perturbed with 50 random samples from the climatological background error covariance matrix embedded in the WRF-Var software

package (Barker et al. 2012) using the fixed covariance perturbation technique of Torn et al. (2006).

To decrease the effects of spurious correlations from sampling errors and to preserve ensemble spread (e.g., Torn 2010, Romine et al. 2013, Schwartz et al. 2014), following Grunzke and Evans (2017), adaptive Gaspari-Cohn localization (Gaspari and Cohn 1999; Anderson 2012), sampling error correction (Anderson 2012), and time- and space-varying adaptive prior inflation applied in model state space (Anderson 2009) are used. The Gaspari-Cohn localization uses a 5^{th}-order compactly supported polynomial and the adaptive prior inflation uses Bayesian statistics to update the spatially varying prior inflation factor. Table 1 provides information regarding the DART configuration.

Data are primarily obtained from the NOAA Meteorological Assimilation Data Ingest System and include Meteorological Terminal Aviation Routine (METAR), radiosonde/dropsonde, marine station (ship and buoy), aircraft, and profiler observations. Also assimilated are satellite-derived atmospheric motion vectors (AMVs) obtained from the UW-Madison Space Science and Engineering Center (Velden et al. 2005). Specific observation types assimilated from each dataset, including the assumed observation errors for each type, are provided in Table 2. Observation windows for all types are ±1 h. Following Romine et al. (2013), WRF-ARW state variables updated during cycled analysis include U, V, W, T, QVAPOR, QCLOUD, QRAIN, QNRAIN, QSNOW, QICE, QNICE, QGRAUP, H_DIABATIC, and REFL_10CM, as well as PH, MU, Q2, T2, V10, U10, and PSFC for vertical-coordinate transformations and/or use by forward operators.

Additional observational processing closely resembles that of Grunzke and Evans (2017). This includes a check to exclude observations when the observation height and model terrain height differ by more than 300 m (Ha and Snyder 2014). In addition, densely packed observations are

"superobbed," with distance thresholds for aircraft and satellite-derived observation superobbing set to ± 22.5 km and ± 25 hPa in the horizontal and vertical, respectively. Observational error for observations near the analysis domain's lateral boundaries is increased within the first five grid points inward of the boundaries, and any observations that fall outside of three standard deviations of the ensemble prior estimate are rejected. Assimilation quality is investigated by evaluating the bias, root-mean squared error (RMSE), and total ensemble spread (including both assumed observation error and ensemble spread; e.g., Dowell et al. 2004) of the prior and posterior analyses.

b. Free forecasts

Free forecasts start from the end of the cycled analysis period at 1200 UTC 31 May 2013 and run out to 36 h (0000 UTC 2 June 2013). Forecasts are conducted for all 50 ensemble members on a two-way nested domain, with a Δx = 3 km horizontal grid spacing domain (1501 x 951 grid points; 50 vertical levels extending up to 50 hPa) embedded in the Δx = 15 km analysis domain. The inner Δx = 3 km domain encompasses the conterminous United States. Figure 4 shows the outer and inner domains with an example of assimilated observations for the 1200 UTC 27 May 2013 analysis cycle. Ensemble forecasts are initialized with the posterior analyses from the last cycled analysis at 1200 UTC 31 May 2013. LBCs are obtained by perturbing 0.5° 1200 UTC 31 May 2013 GFS model cycle data every 3-h from the 0-h analysis through the 36-h forecast period using the fixed covariance technique of Torn et al. (2006). Physical parameterizations follow those used in the cycled analysis except for the deep cumulus parameterization, which is turned off on the 3-km domain under the assumption that deep cumulus is at least crudely resolved. This configuration largely follows previous studies that used WRF-ARW coupled with DART to generate initial conditions for convection-allowing model forecasts (e.g., Romine et al. 2013, 2014; Schwartz et al. 2014, 2015b,c; Torn and Romine 2015; Grunzke and Evans 2017).

c. Forecast verification

To assess forecast quality and inform subsequent composite and ensemble sensitivity analyses, fractions skill score (FSS; Roberts and Lean 2008, Schwartz et al. 2010) is used to verify individual ensemble member 6-h accumulated precipitation forecasts valid at 0400 UTC 1 June 2013, roughly encompassing the period of CI and initial upscale growth in central Oklahoma with the observed event. FSS is a neighborhood verification method that accounts for spatial displacements in forecast quantities at a given threshold and over a desired radius of influence. FSS has gained widespread use, especially within the last decade, to measure precipitation forecast skill (e.g., Roberts and Lean 2008, Schwartz et al. 2009, 2010, 2014, 2015, Romine et al. 2013, 2014, 2016, Schwartz 2016, Sobash et al. 2016a,b, Schwartz and Sobash 2017, Gowan et al. 2018) given its intrinsic neighborhood formulation (versus grid-point verification, where a small spatial offset between forecast and analysis fields results in both a miss and false alarm at their respective grid points).

The process of computing FSS, here done using the Model Evaluation Tools package (Fowler et al. 2010), first involves using linear interpolation to regrid both modeled and observed (in this study, one-hourly data from the NCEP Stage IV analysis, later combined into a six-hourly accumulation; Lin and Mitchell 2005) spatially varying accumulated precipitation data to a common 0.04° latitude/longitude grid encompassing all of Oklahoma except for the panhandle (Fig. 5). Next, a binary yes/no field for both observations and model forecasts is generated based on a 6-h (between 2200 31 May - 0400 UTC 1 June 2013) accumulation threshold of 25 mm. This period and accumulation threshold effectively delineate the region of supercell initiation and initial upscale growth into an MCS for the observed event (Fig. 5). Sensitivity in the results to the chosen evaluation region, accumulation threshold, and 6-h accumulation period is negligible for small

perturbations in each (not shown). Next, the fractions of observed and forecast grid points exceeding the 25-mm accumulation threshold within a neighborhood of radius 25 km are computed; an idealized example of this process is depicted in Fig. 6. Sensitivity in the results to small perturbations in the neighborhood size is also negligible (not shown).

The fractions Brier score (FBS) is then computed as:

$$\mathrm{FBS} = \frac{1}{N_v}\sum_{i=1}^{N_v}[\mathrm{NP}_{F(i)} - \mathrm{NP}_{O(i)}]^2$$

where NP is the neighborhood probability for the forecast (F) and observed (O) fraction fields, respectively, for each grid box (I; total N_v) considered. FBS ranges from 0 to 1, with 0 representing a perfect forecast. From this, the worst-possible FBS for the forecast and observed neighborhood fractions is defined as,

$$\mathrm{FBS}_{\mathrm{worst}} = \frac{1}{N_v}\left[\sum_{i=1}^{N_v}\mathrm{NP}_{F(i)}^2 - \sum_{i=1}^{N_v}\mathrm{NP}_{O(i)}^2\right].$$

The FSS is then computed, following Roberts (2005), as:

$$\mathrm{FSS} = 1 - \frac{\mathrm{FBS}}{\mathrm{FBS}_{\mathrm{worst}}},$$

where FSS represents the skill of a forecast compared to the observations in a way that FBS cannot because it no longer depends on the frequency of an event (Roberts and Lean 2008, Schwartz et al. 2010). FSS ranges from 0 to 1, with a score of 1 representing a perfect forecast. Note that FSS is computed separately for each ensemble member, to evaluate skill variation across the ensemble, rather than using an ensemble neighborhood probability over the entire ensemble, in which overall ensemble forecast skill would be quantified (Schwartz and Sobash 2017).

d. Analysis methods

Ensemble-member FSSs are sorted from best to worst (Fig. 7) for the purposes of identifying the ten-best and -worst forecasts to develop composites. The use of ten members per composite is

somewhat arbitrary, but the composite analyses are found to be somewhat insensitive to the precise members chosen for each composite (e.g., by swapping the 9th and 10th best/worst members with the next two members in the FSS ranking, respectively; not shown).

Separately, sensitivity in ensemble-member forecast skill, as measured using FSS, to the initial and forecast atmospheric state is assessed using ensemble sensitivity analysis (ESA; Ancell and Hakim 2007, Torn and Hakim 2008). ESA has been applied to ensemble forecasts at a wide range of lead times for a broad spectrum of meteorological phenomena spanning from the synoptic- to the convective-scale (e.g., Ancell and Hakim 2007, Torn and Hakim 2008, 2009, Torn 2010a,b, Bednarczyk and Ancell 2015, Torn and Romine 2015, Hill et al. 2016, Romine et al. 2016, Berman et al. 2017, Grunzke and Evans 2017, Torn et al. 2017).

ESA represents the slope of the linear regression between a forecast metric **J** and an analysis state variable **x**:

$$\frac{\partial J}{\partial x} = \frac{\mathrm{cov}(\mathbf{J},\mathbf{x})}{\mathrm{var}(\mathbf{x})},$$

where **J** represents a $1 \times M$ ensemble estimate of the forecast metric and **x** is an $N \times M$ ensemble estimate of the analysis state, with M = ensemble size (here, 50) and N = number of horizontal grid points in the model domain (here, 1501 x 951 = 1,427,451). Here, **J** is the ensemble-member FSS, and a multitude of atmospheric state fields from forecast hours 0 (1200 UTC 31 May 2013) to 16 (0400 UTC 1 June 2013) are considered for the analysis state variable **x**. Both **J** and **x** have their ensemble means removed prior to computing the sensitivity metric, following Ancell and Hakim (2007) and Torn and Hakim (2008). For interpretation, the sensitivity metric is normalized by the standard deviation of the analysis metric, such that a one standard deviation change in the analysis metric is associated with the displayed change in the forecast metric (here, an increase or decrease in FSS). Further, the sensitivity metric is only considered where the linear correlation coefficient

r between **J** and **x** is statistically significantly different than zero to at least 90% confidence as assessed using a Student's *t* test (given approximately normally distributed data; not shown).

Both composite analysis and ESA have strengths and weaknesses. For instance, ESA considers variation across the full ensemble whereas composite analysis only considers selected ensemble members, and variation across the members of a given composite is assumed to be negligible. That said, ESA assumes linear relationships across the ensemble, which may not be realistic given the importance of moist dynamics to nonlinear error growth on short time scales (Gilmour et al. 2001). Nevertheless, previous studies (e.g., Hill et al. 2016, Berman et al. 2017, Grunzke and Evans 2017) demonstrate that ESA can provide insight into forecast sensitivities and dynamical processes even in the presence of non-linearity.

3. Results

a. Assimilation performance

The cycled analysis is analyzed for performance to determine whether the posterior analysis is improved relative to the prior analysis and the extent to which the prior RMSE and total spread (ensemble spread plus assumed observation error) are consistent with each other. The latter of these measures is considered to determine whether the flow-dependent forecast error covariance statistics used in the assimilation process are both reliable and realistic (Houtekamer and Mitchell 1998, Dowell et al. 2004).

At the surface, here considered in terms of METAR observations, RMSE and total spread decrease substantially in the posterior relative to the prior analysis for all variables (Fig. 8). The posterior analysis is also associated with substantially reduced diurnal variation in RMSE and total spread as compared to the prior analysis. These results follow those of Romine et al. (2013) and Grunzke and Evans (2017), among others. In general, prior and posterior total spread for all

variables are slightly higher than their corresponding RMSE values at all analysis times, suggesting that minor analysis improvements might be obtained by reducing the assumed observation errors. Temperature is generally cool-biased, while specific humidity is generally moist-biased; although the posterior analysis reduces bias amplitudes, the signs of each bias persist (Fig. 8b,c). This is likely a result of the MYJ PBL parameterization scheme, which is known to have a cool and moist bias in warm-season, thunderstorm-supporting environments (e.g., Hu et al. 2010, Coniglio et al. 2013). The horizontal winds (Figs. 8d,e), on the other hand, exhibit a slight easterly and southerly bias in the prior that is eliminated for the zonal wind while considerably reduced for the meridional wind in their respective posterior analyses.

Above the surface, vertical profiles of selected variables from several observation platforms exhibit minimal change in RMSE, total spread, and bias between the posterior analyses from 1200 UTC 28 May 2013 and 1200 UTC 31 May 2013 (Fig. 9). However, the correspondence between total spread and RMSE is similar to that seen for the surface observation platforms; i.e., total spread is generally slightly larger than RMSE at each variable and altitude (c.f. Figs. 8-9). Posterior biases for both times in each variable are generally near zero except for the horizontal wind observations from radiosondes, for which bias is somewhat more variable with height (Figs. 9c-d).

The assimilation quality results closely mirror those of Grunzke and Evans (2017), which used a nearly identical analysis system. Although Grunzke and Evans (2017) considered a case from 2009, for which the corresponding GFS analysis at the completion of the cycled analysis period is likely degraded compared to that herein due to numerous model, data assimilation, and observation improvements since 2013, their analysis better fit observations than did the corresponding GFS analysis at the end of the cycled analysis period. Further investigation is needed, however, to determine if the same is true for this case.

b. Ensemble performance

With respect to CI, the better-performing ensemble members very closely follow the observed CI timing around 2200 UTC but are much different in terms of CI location, with modeled CI consistently occurring approximately 200 km to the south-southwest, near the Texas-Oklahoma border, of the observed CI event in central Oklahoma (Figs. 10a and 11). The worst-performing ensemble members have approximately the same initiation time but are 40 km even further south-southwest (Figs. 10b and 12), with these initial storms typically either dissipating in central Oklahoma before new convection initiated after 0400 UTC or remaining weak before intensifying and growing upscale after 0400 UTC. The CI timing errors are consistent with previous research (e.g., Kain et al. 2013, Duda and Gallus 2013, Burghardt et al. 2014, Burlingame et al. 2017, Keclik et al. 2017); however, the spatial errors are somewhat larger than those in previous studies (although several of them restricted their search radius for event verification to small distances than 200 km). Whereas the spatial error in model-predicted CI was large with small temporal error (Fig. 10), the opposite is somewhat seen with respect to MCS formation: spatial errors are smaller (models a bit further northeast, despite the large spatial CI errors) with larger temporal error across the ensemble (Fig. 11d and 13). Some members produce the MCS much earlier in the simulations while others produce it much later however it does seem in part dependent on how the initial cells fare upon initiation. Regardless, most ensemble members initiate an MCS in northeast-to-central Oklahoma at some point during the evening to mid-morning hours on 31 May – 1 June 2013 (not shown). This may be related to a low-level jet that may have provided the lift necessary to produce the MCS even if it may have been well after 0400 UTC. This will be expanded upon in future research where this part of the event may be of greater focus in order to better understand the entirety of this event.

Ensemble member forecasts exhibit a wide range of FSSs for the 25-mm threshold for the 2200 UTC 31 May - 0400 UTC 1 June 2013 period (Table 3; Fig. 7). Compared to previous studies, the mean of the ensemble-member FSSs of 0.35 for a 25-km neighborhood is higher than that for the same-sized neighborhood in Schwartz et al. (2010) when comparing similar rain rate thresholds. It is also higher than those for a larger 50 km neighborhood in other studies (Schwartz et al. 2014, Romine et al. 2013), albeit with different rain rate thresholds. It should be noted, however, that the cases considered by each of these studies are quite different, and no attempt is made to determine whether the case considered here is associated with intrinsically higher predictability. Future work is planned to compute the "target skill" metric of Roberts and Lean (2008) to evaluate the extent to which the ensemble forecasts in this study are truly skillful.

c. Composites

Composites of middle-tropospheric fields highlight small meso-to-synoptic-scale differences in feature amplitude and placement between the best- and worst-performing ensemble members. At 500 hPa, the primary feature of interest is a meso-alpha-scale area of higher geopotential height (Fig. 14c) and more anticyclonic flow (Figs. 14a-b) along the Iowa-Missouri border, immediately downstream of a central United States longwave trough, in better-performing ensemble members. This results in a larger horizontal geopotential height gradient, larger differential cyclonic absolute vorticity advection, and greater quasi-geostrophic forcing for lower-to-middle tropospheric ascent from western Kansas and Oklahoma into the central Mississippi River valley. The composites also indicate that a weak shortwave in the north-central High Plains on the western edge of the central United States longwave trough is located further southwest in the better-performing ensemble members (Fig. 14c). It is speculated that this helps to sustain initial convection in western

Oklahoma in the better-performing members (relative to the worst-performing members) through the local evening hours, as discussed in Section 3b.

In the lower troposphere, the lee cyclone along the quasi-stationary frontal boundary in western Oklahoma is stronger in the better-performing ensemble members (Fig. 15a). In conjunction with this, southerly flow impinging on the baroclinic zone ahead of the lee cyclone is faster in the better-performing ensemble members near the time of CI (Fig. 15b). This indicates stronger convergence and thus lower-tropospheric forcing for ascent along the quasi-stationary front. The kinematic field differences persist through the period of upscale growth and backbuilding development (Fig. 15c), potentially helping to support MCS maintenance in better-performing ensemble members through the local evening hours. As a result of the kinematic differences, horizontal gradients of moisture (Figs. 16-17) and instability (Fig. 18) along the quasi-stationary front are larger in magnitude and potentially displaced slightly westward in the better-performing ensemble members.

Together, the lower-to-middle tropospheric composite analyses near the time of observed CI imply stronger forcing for ascent in a favorable thermodynamic environment in western Oklahoma (Fig. 18) in the better-performing ensemble members. This is hypothesized to erode the convective inhibition over western Oklahoma (Fig. 18b) sufficiently so as to allow for CI and upscale growth to more readily occur in the better-performing ensemble members but not to the same extent in the worst-performing ensemble members.

d. Ensemble Sensitivity Analysis

On an overarching basis, the ESA shows similar features to the composites with regions of highest sensitivity in the same regions of greatest differences between the best and worst ensembles. In particular, 500 hPa meridional and zonal winds (Fig. 19 at 2200 UTC) showed the same southwestern shift in the trough axis over the central US supporting the idea of a stronger

ridge in the central US/mid-Mississippi Valley region with better FSSs. Similar findings continued with 850 hPa winds (Fig. 20 at 2200 UTC) that features converging winds (especially meridional) along the southwest-northeast axis of Oklahoma. This is seen from the surface to at least 700 hPa, likely indicative of the stationary front. The similarities continue with low-level moisture fields (Fig. 21 at 2200 UTC) in the vicinity of the dryline, including the dryline bulge region with seemingly a more western dryline, possibly stronger gradients, and a more pronounced and more northerly dryline bulge with better FSSs likely explaining the differences in CI location.

With the ESA a few additional features were investigated including 300 hPa winds, equivalent potential temperature and potential temperature. With the 300 hPa winds, the greatest sensitivities in the jet streak region for the zonal winds (Fig. 22a at 2200 UTC) and meridional winds (Fig. 22b at 2200 UTC). These unique sensitivities are likely a result of strength and location differences in the jet streak between ensemble members with the positively sensitive regions showing stronger winds overall in the northern part of the jet streak, indicative of perhaps a more expansive jet with better FSSs. This might thus lead to varying regions and/or degrees of divergence aloft that could impact CI but more likely the upscale growth.

Equivalent potential temperature at 700 hPa (Fig. 23a at 2200 UTC) shows that there may have been localized higher moisture further aloft in the region of CI for better-performing ensembles. This may have some significance particularly for CI because it may make the storm environment more susceptible to initiation given more low-level buoyancy despite that it might initially lower instability. However, outside of this localized environment, storms may be more likely to flourish while moving into a new environment. At 850 hPa, as well as at 925 hPa, shows similar features to those seen in the moisture fields as might be expected with mostly positive

sensitivities along the dryline region (Fig. 23b at 2200 UTC). This reinforces the idea that the dryline is further west and has a stronger gradient.

This is similarly seen with potential temperature as well, particularly at 700 hPa (Fig. 24a at 2200 UTC) though this may not be related to the dryline, though remains consistent with the implications made with previous fields. At 850 hPa (Fig. 24b at 2200 UTC), there are explicit differences in the potential temperature field perhaps showing influence from the lee cyclone and the region of converging winds across Oklahoma. These sensitivities may be followed closer to the surface at 925 hPa (not shown) and with the locations of these sensitivities there are some indications that CI may have been influenced by differences in potential temperature at the low levels in a similar manner to the moisture boundaries influence. These differences may not be directly related to the moisture boundary differences but perhaps suggest differences in the region are indicative of differences with temperature boundaries (cold/stationary front) located in the same region as the dryline.

4. Conclusions and Future Work

This research quantified the practical predictability and forecast sensitivities of the 31 May 2013 heavy-rain–producing MCS in central Oklahoma, particularly with respect to its sensitivity to CI prior to upscale growth. In both the composites and ESA, winds at 500 hPa showed sensitivity to the location of a 500 hPa trough, with better ensemble members likely indicating a more southwesterly trough axis, which would likely shift absolute vorticity further west pushing associated cyclonic vorticity advection further west as well. This might thus shift CI further west, however, it is worth noting that sensitivities at 500 hPa are mainly found in Kansas and north of CI location. Further aloft, negative sensitivities within a jet streak at 300 hPa seemingly indicates stronger zonal winds in the jet to the east associated with decreasing FSS suggesting that

downward motion linked to the right exit region of the jet might be stronger and/or more expansive that might inhibit convection and corresponding upscale growth in northern Oklahoma and eastern Kansas. At 500 hPa and 300 hPa direct correlations to CI are found and helps determine potential impacts on upscale growth and mode of convection.

As one might expect from regions closer to the surface or more generally within the PBL, largest differences are seen at the lower levels, at 850 hPa and below, though 700 hPa also hints at regions of interest with respect to convergence along the axis of the trough. The moisture parameters were of particular importance with the dryline likely being one of the main forcing mechanisms for CI with FSSs being very sensitive. At 850 hPa, the dryline is not immediately noticeable as there are multiple but the regions of higher moisture in the field were very sensitive with the composites and ESA indicating that the entire moisture field is further west with increasing FSSs with possibly higher moisture overall and stronger moisture gradient. The 850 hPa winds also indicate high positive sensitivity to meridional winds in Oklahoma and to a lesser degree the zonal winds that indicate stronger winds overall in this region in addition to a stronger gradient. This is indicative of stronger convergence in this area in a slightly different region than that indicated by the dryline possibly signifying the quasi-stationary/cold front. At lower levels the dryline becomes more apparent and reinforces the idea of higher FSSs associated with a more northwesterly dryline and a likely stronger gradient. The low-levels also indicate a slight dryline bulge in northern Texas with similar sensitivities. This would seemingly lead to this region being most likely for CI with CI potentially further west and slightly earlier associated with members with higher FSSs. However, it is likely that the mid-upper levels act in concert with the lower levels, that while differences are not as large, the impact of smaller differences plays a large role in the greater differences seen at lower levels.

Determining the best 6 h period for joint CI and MCS predictability FSS highlights variable skill metric values across the ensemble based on whether an earlier 6-h period including CI or a later 6-h period focusing on upscale growth and/or backbuilding development is considered. After evaluating the sensitivity via six 6-h periods that span over different parts of the event from 2200-0400 UTC increasing each hour until 0300-0900 UTC, along with one 11-h period from 2200-0900 UTC, the smallest FSSs are seen in the middle two 6-h periods. More specifically, the highest ensemble members' FSSs decrease until the middle two 6-h time periods (starting at 0000 UTC and 0100 UTC) and then increases again for the last two periods. In addition, the ensemble members with the highest FSS values for the earliest-considered periods generally had comparatively low FSS values for later periods, and vice versa: the ensemble members with the lowest FSS values for the earliest-considered periods generally had comparatively high FSS values for later periods. Based on the actual event having multiple modes (e.g., supercell-to-MCS followed by separate CI tied to the low-level jet), this suggests that the ensembles may struggle more during the middle of the event, which may be attributed to the modality of this event. This may be because precipitation is extremely sensitive to the spatial and temporal variability of these modes and the transition phases from mode to mode as the actual event would indicate discrete cells initially, shifting into an east-west oriented quasi-stationary MCS before back-building, where then the storm quickly starts pushing southeast. Further investigation is needed to quantify why event predictability varies as it does through the MCS's life cycle and the extent to which forecast sensitivities vary as a function of the chosen verification period.

Future work could be used to expand on much of the research here including further work with this event and using similar methods to study other events. One area of future work could be to further quantify contributions from model error, including with parameterization choices

(particularly the microphysical parameterization, which previous studies have demonstrated to exert a significant control on convective system cold-pool properties and thus longevity; e.g., Adams-Selin et al. 2013). Although this research did not investigate the changes in mode of the storms a great deal, this research suggests that storm mode may have had sensitivities based upon the variations in FSS while using different time periods and future work could be useful to determine why the changes in mode occurred and what lead to the spatial and temporal changes. Finally, a more in-depth look at these ensemble members and their differences is needed to better understand the potential triggers of upscale growth and back-building phases of this event.

List of Tables

Tables

Parameter	Value
Filter Type	Ensemble adjustment Kalman Filter
Ensemble members	50
Assimilation interval	6 h
Localization type	Gaspari-Cohn
Adaptive localization threshold	2000
Outlier threshold	3.0
Inflation type	Spatially-varying
Adaptive inflation initial mean	1.0
Adaptive inflation initial standard deviation	0.8
Sampling error correction	On
Cutoff	0.1
Horizontal localization half-width	635 km
Vertical localization half-width	8 km

Table 1. DART parameters used in the cycled analysis system.

Observation Platform	Variable
METAR/Marine	Temperature U-, V- winds Specific humidity Altimeter
Radiosonde/Dropsonde	Temperature U-, V- winds Specific humidity Surface altimeter
ACARS (aircraft) (±22.5 km, ±25 hPa)*	Temperature U-, V- winds Specific humidity
Profiler	U-, V- winds Pressure
Satellite Derived (±22.5 km, ±25 hPa)*	U-, V- winds

Table 2. Observation platforms and the corresponding quantities assimilated.

Best/Worst Ensemble Member FSSs

Best		Worst	
26	0.67	23	0.14
42	0.61	27	0.12
35	0.61	37	0.09
48	0.61	34	0.09
50	0.59	8	0.09
7	0.59	13	0.08
20	0.59	6	0.05
47	0.58	21	0.05
32	0.57	33	0.05
24	0.54	44	0.04

Table 3. List of the 10 best- and worst-performing ensemble members with their corresponding FSSs for the 6-hr period from 2200 UTC 31 May 2013 to 0400 UTC 1 June 2013.

List of Figures

Figure 1. Rapid Refresh model 0-h analysis at 1800 UTC 31 May 2013, showing 300 hPa wind speed (color shading in m s^{-1}), 850 hPa wind barbs (half barb = 5 kt, full barb = 10 kt, pennant = 50 kt; 1 kt = 0.5144 m s^{-1}), sea-level pressure (black contours every 4 hPa, with low pressure centers indicated by L), and surface-based convective available potential energy (gray shading in J kg^{-1}). Figure reproduced from Schumacher (2015), their Fig. 1.

Figure 2. Surface analysis at 2100 UTC 31 May 2013 over central Oklahoma, including standard and Oklahoma Mesonet surface observations (temperature at upper left in °C, dew point temperature at lower left in °C, sea-level pressure at upper right in hPa*10 with the leading 9 or 10 omitted, and wind barbs at center with half barb = 5 kt, barb = 10 kt, and pennant = 50 kt). Surface boundaries are manually analyzed, including a cold/stationary front, dry line, and pre-frontal trough. The red star marks the location of the observed and model vertical profiles depicted in Fig. 3. Figure reproduced from Schumacher (2015), their Fig. 2.

Figure 3. Skew *T*–ln *p* diagram depicting a mobile radiosonde observation collected by Colorado State University at 1920 UTC 31 May in the pre-frontal, pre-dryline airmass at the location in west-central Oklahoma marked by the red star (temperature trace in red, dew point temperature trace in green, wind barbs in black in the first column at right, and the associated hodograph in the upper inset in red). A profile from the same location in a Colorado State University real-time, convection-allowing 0000 UTC model forecast valid at 1930 UTC is also depicted (temperature trace in solid blue, dew point temperature trace in dashed blue, wind barbs in black in the second column at right, and the associated hodograph in the upper inset in blue). For the wind barbs, a half barb = 5 kt, full barb = 10 kt, and pennant = 50 kt. The parcel curve (dashed black) is for a parcel with the mean properties of the lowest 50 hPa in the observed sounding. Thermodynamic

quantities for both the observed (top) and model soundings (bottom) are shown at right. Figure reproduced from Schumacher (2015), their Fig. 6.

Figure 4. Observations assimilated at 1200 UTC 27 May 2013, with the outer and inner domains outlined by concentric thick black lines.

Figure 5. Precipitation (mm) observations from NCEP Stage IV analysis (Lin and Mitchell 2005) using the same coordinates used in the computation of the FSS for the same 6-h period of 2200 UTC 31 May to 0400 UTC 1 June.

Figure 6. Hypothetical example of forecast (left) and observed (right) fractions within a 5 x 5 neighborhood (outer black line) centered on a single grid point (inner black line). In this example, the center grid point did not forecast the field of interest (e.g., 6-h accumulated precipitation) at the threshold of interest, even though it was observed. In a traditional grid-point verification, this would be classified as a miss (0/1 forecast, 1/1 observation). However, for both forecast and observation, seven grid points within the twenty-five grid-point neighborhood met the threshold of interest. For fractions skill score, this would result in a "perfect" neighborhood forecast (7/25 forecast, 7/25 observation). Adapted from Roberts and Lean (2008; their Fig. 2) and Schwartz et al. (2010; their Fig. 11).

Figure 7. FSS for the 2200 UTC 31 May 2013 to 0400 UTC 1 June 2013 period over Oklahoma in order of decreasing skill, with the best 10 members to the left of the blue line and the worst 10 members to the right of the green line. The *x*-axis indicates the ensemble member number associated with a given FSS.

Figure 8: RMSE (green), total spread (ensemble spread plus assumed observation error; red), and bias (blue) of the prior (solid) and posterior (dashed) ensemble estimates for METAR (a) altimeter

(hPa), (b) temperature (K), (c) specific humidity (g kg^{-1}), (d) zonal wind (m s^{-1}), and (e) meridional wind (m s^{-1}) observations between 1800 UTC 26 May 2013 and 1200 UTC 31 May 2013.

Figure 9. RMSE (green), total spread (ensemble spread plus assumed observation error; red), and bias (blue) of the posterior analyses for 1200 UTC 28 May 2013 (solid) and 1200 UTC 31 May 2013 for radiosonde (a) specific humidity (g kg^{-1}), (b) temperature (K), (c) zonal wind (m s^{-1}), and (d) meridional wind (m s^{-1}); ACARS (e) specific humidity (g kg^{-1}), (f) temperature (K), (g) zonal wind (m s^{-1}), and (h) meridional wind (m s^{-1}); and AMV (i) zonal wind (m s^{-1}), and (j) meridional wind (m s^{-1}).

Figure 10. Reflectivity (dBZ, shaded per the color bar in each panel) shown near initiation at 2200 UTC 31 May 2013 for the (a) best ensemble member (26), (b) worst ensemble member (44), and (c) the actual case (http://www2.mmm.ucar.edu/imagearchive/). In panels (a) and (b), barbs indicate 10-m wind (half barb: 2.5 m s^{-1}, barb: 5 m s^{-1}, pennant: 25 m s^{-1}).

Figure 11. Reflectivity (dBZ, shaded) from ensemble member 26 (the best-performing member, in terms of FSS) at (a) 2200 UTC 31 May 2013, (b) 0100 UTC, (c) 0400 UTC, (d) 0700 UTC on 1 June 2013.

Figure 12. Reflectivity (dBZ, shaded) from ensemble member 44 (the worst-performing member, in terms of FSS) at (a) 0000 UTC and (b) 0300 UTC on 1 June 2013.

Figure 13. Reflectivity (dBZ) from the actual case (http://www2.mmm.ucar.edu/imagearchive/) at 0630 UTC on 1 June 2013.

Figure 14. Composite differences (shaded) and full-ensemble–mean (contoured) of 500 hPa (a) meridional wind (m s^{-1}), (b) zonal wind (m s^{-1}), and (c) geopotential height (m) at 2200 UTC 31 May 2013. Red (blue) indicates higher values in the average of the best (worst) 10 ensemble

members. Barbs in (a) and (b) indicate the full-ensemble–mean 500 hPa horizontal wind (half barb: 5 kt, full barb: 10 kt, pennant: 50 kt).

Figure 15. Composite differences (shaded) and full-ensemble–mean (contoured) of 850 hPa (a) zonal wind difference (m s^{-1}), (b) meridional wind difference (m s^{-1}) at 2200 UTC 31 May 2013, and (c) meridional wind difference (m s^{-1}) at 0500 UTC 1 June 2013. Barbs in each panel indicate the vector wind difference between the best- and worst-performing composite means (half barb: 5 kt, full barb: 10 kt, pennant: 50 kt).

Figure 16. Composite differences (shaded) and full-ensemble–mean (contoured) of (a) 850 hPa dew point temperature (°C) and 1000-500 hPa average relative humidity (%) at 2200 UTC 31 May 2013.

Figure 17. Composite differences (shaded) and full-ensemble–mean (contoured) of (a) 925 hPa dew point temperature (°C) and surface to 850 hPa average water vapor mixing ratio (g kg^{-1}) at 2200 UTC 31 May 2013.

Figure 18. Composite differences (shaded) and full-ensemble–mean (contoured) of (a) surface-based convective available potential energy (J kg^{-1}) and (b) surface-based convective inhibition (J kg^{-1}) at 2200 UTC 31 May 2013.

Figure 19. Sensitivity of 2200 UTC 31 May 2013 – 0400 UTC 1 June 2013 FSS for a 25-km radius neighborhood over the domain in Fig. 6 to 500 hPa (a) meridional wind (m s^{-1}) and (b) zonal winds (m s^{-1}) at 2200 UTC 31 May 2013.

Figure 20. Sensitivity of 2200 UTC 31 May 2013 – 0400 UTC 1 June 2013 FSS for a 25-km radius neighborhood over the domain in Fig. 6 to 850 hPa (a) meridional wind (m s^{-1}) and (b) zonal winds (m s^{-1}) at 2200 UTC 31 May 2013.

Figure 21. Sensitivity of 2200 UTC 31 May 2013 – 0400 UTC 1 June 2013 FSS for a 25-km radius neighborhood over the domain in Fig. 6 of (a) 850 hPa water vapor mixing ratio (g kg^{-1}), (b) 850 hPa dew point temperature (°C), (c) 925 hPa dew point temperature (°C), and (d) surface dew point temperature (°C) at 2200 UTC 31 May 2013.

Figure 22. Sensitivity of 2200 UTC 31 May 2013 – 0400 UTC 1 June 2013 FSS for a 25-km radius neighborhood over the domain in Fig. 3 of 300 hPa (a) zonal wind (m s^{-1}) and (b) meridional wind (m s^{-1}) at on 2200 UTC 31 May 2013. Orange indicates that a one standard deviation increase in the contoured field will result in the shaded increase in FSS, whereas blue indicates that a one standard deviation increase in the contoured field will result in the shaded decrease in FSS.

Figure 23. Sensitivity of 2200 UTC 31 May 2013 – 0400 UTC 1 June 2013 FSS for a 25-km radius neighborhood over the domain in Fig. 6 of equivalent potential temperature (K) at (a) 700 hPa and (b) 850 hPa on 2200 UTC 31 May 2013.

Figure 24. Sensitivity of 2200 UTC 31 May 2013 – 0400 UTC 1 June 2013 FSS for a 25-km radius neighborhood over the domain in Fig. 6 of potential temperature (K) at (a) 700 hPa and (b) 850 hPa on 2200 UTC 31 May 2013.

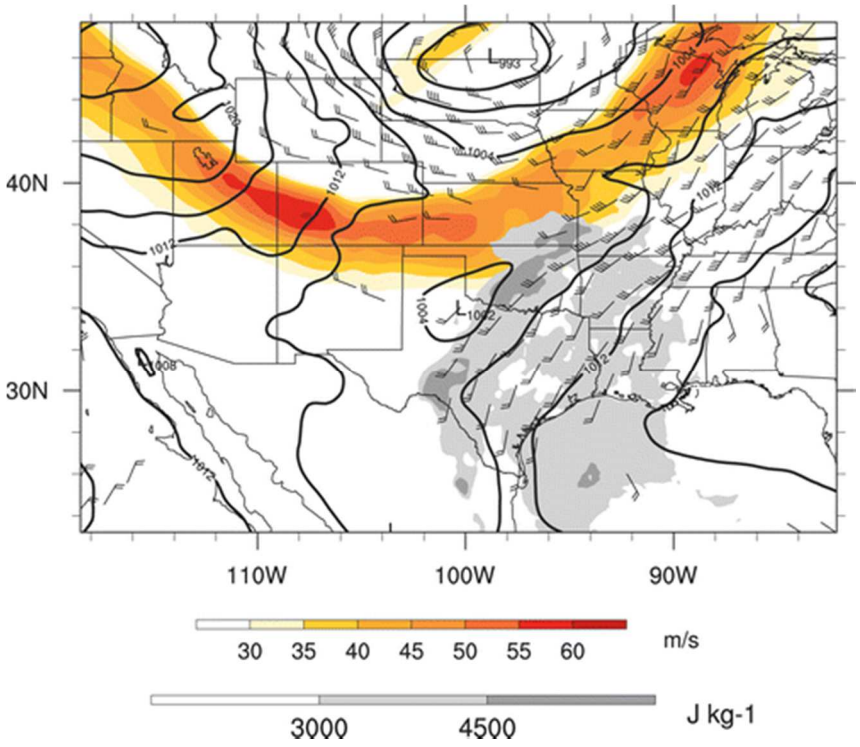


Figure 1. Rapid Refresh model 0-h analysis at 1800 UTC 31 May 2013, showing 300 hPa wind speed (color shading in m s^{-1}), 850 hPa wind barbs (half barb = 5 kt, full barb = 10 kt, pennant = 50 kt; 1 kt = 0.5144 m s^{-1}), sea-level pressure (black contours every 4 hPa, with low pressure centers indicated by L), and surface-based convective available potential energy (gray shading in J kg^{-1}). Figure reproduced from Schumacher (2015), their Fig. 1.

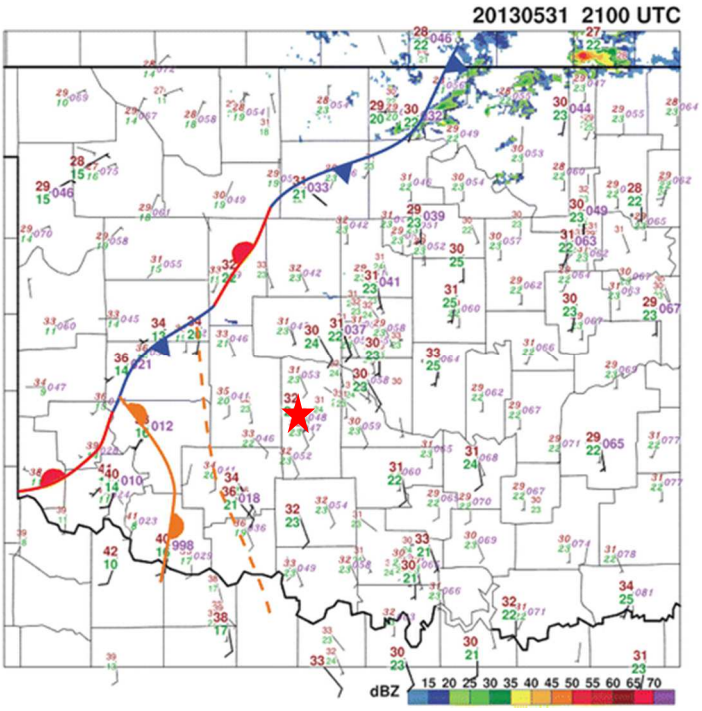


Figure 2. Surface analysis at 2100 UTC 31 May 2013 over central Oklahoma, including standard and Oklahoma Mesonet surface observations (temperature at upper left in °C, dew point temperature at lower left in °C, sea-level pressure at upper right in hPa*10 with the leading 9 or 10 omitted, and wind barbs at center with half barb = 5 kt, barb = 10 kt, and pennant = 50 kt). Surface boundaries are manually analyzed, including a cold/stationary front, dry line, and pre-frontal trough. The red star marks the location of the observed and model vertical profiles depicted in Fig. 3. Figure reproduced from Schumacher (2015), their Fig. 2.

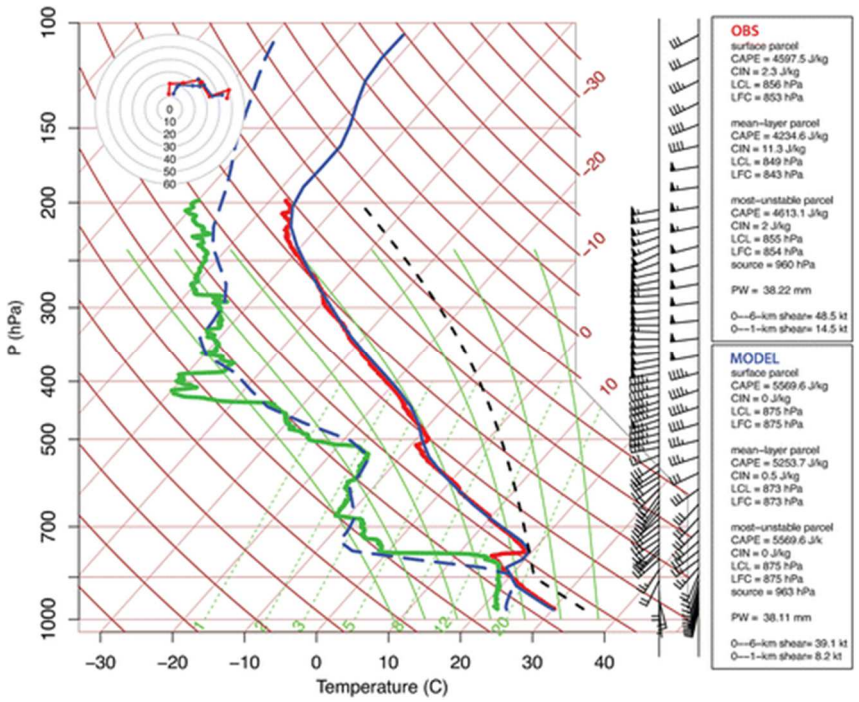


Figure 3. Skew *T*–ln *p* diagram depicting a mobile radiosonde observation collected by Colorado State University at 1920 UTC 31 May in the pre-frontal, pre-dryline airmass at the location in west-central Oklahoma marked by the red star (temperature trace in red, dew point temperature trace in green, wind barbs in black in the first column at right, and the associated hodograph in the upper inset in red). A profile from the same location in a Colorado State University real-time, convection-allowing 0000 UTC model forecast valid at 1930 UTC is also depicted (temperature trace in solid blue, dew point temperature trace in dashed blue, wind barbs in black in the second column at right, and the associated hodograph in the upper inset in blue). For the wind barbs, a half barb = 5 kt, full barb = 10 kt, and pennant = 50 kt. The parcel curve (dashed black) is for a parcel with the mean properties of the lowest 50 hPa in the observed sounding. Thermodynamic quantities for both the observed (top) and model soundings (bottom) are shown at right. Figure reproduced from Schumacher (2015), their Fig. 6.

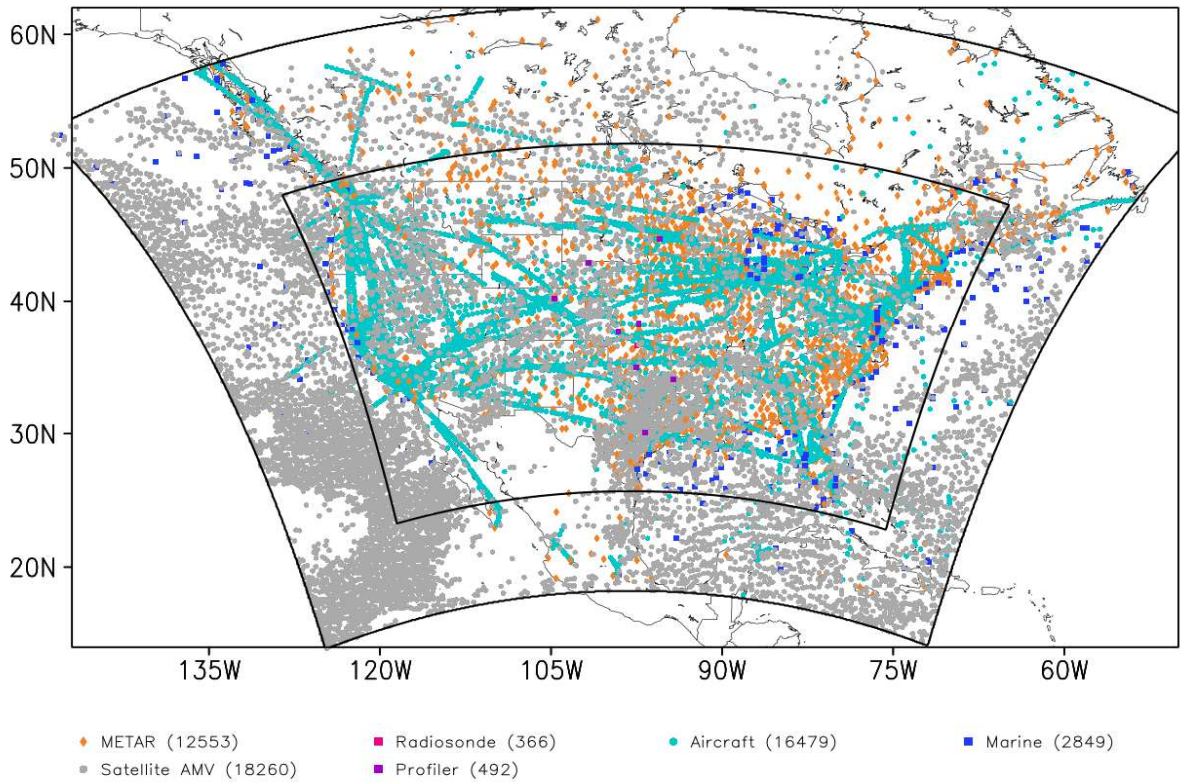


Figure 4. Observations assimilated at 1200 UTC 27 May 2013, with the outer and inner domains outlined by concentric thick black lines.

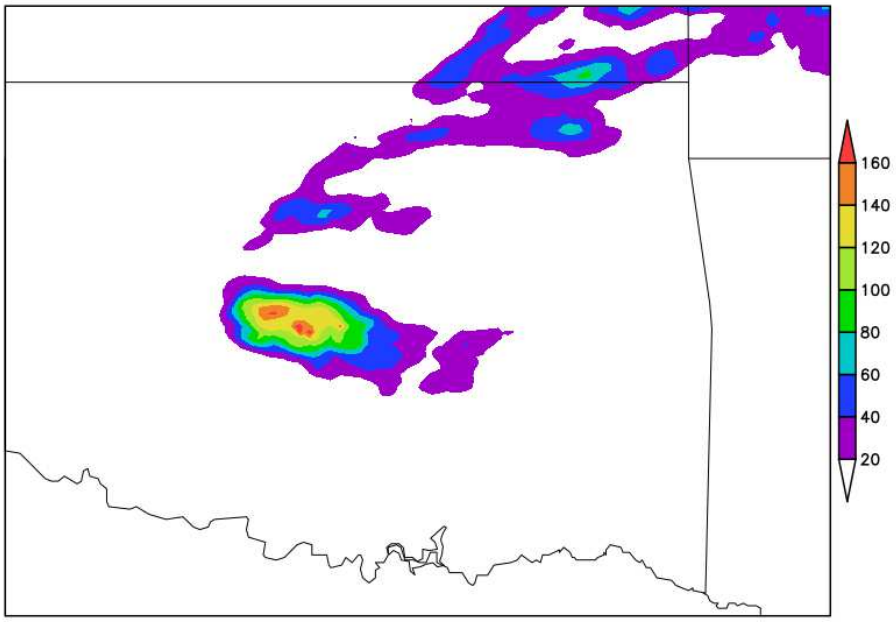


Figure 5. Observed precipitation (shaded; mm) from the NCEP Stage IV analysis (Lin and Mitchell 2005) using the same coordinates used in the computation of the FSS for the 6-h period between 2200 UTC 31 May 2013 and 0400 UTC 1 June 2013.

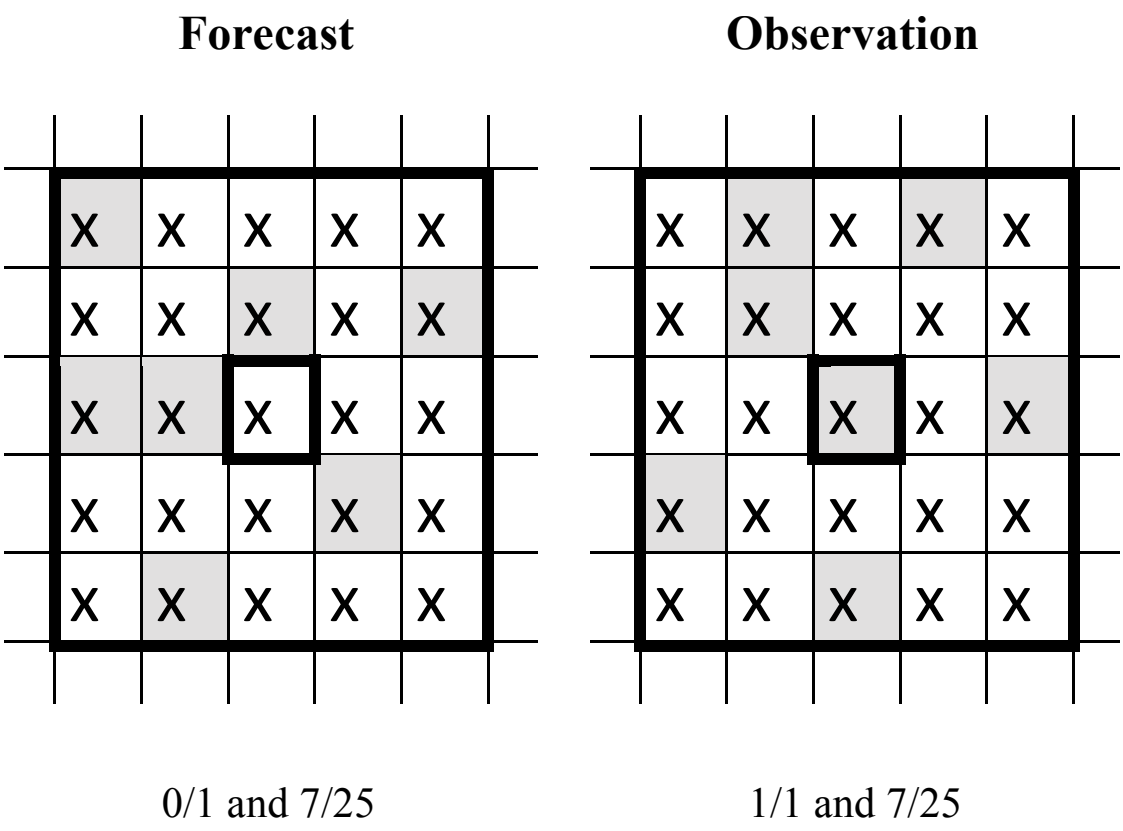


Figure 6. Hypothetical example of forecast (left) and observed (right) fractions within a 5 x 5 neighborhood (outer black line) centered on a single grid point (inner black line). In this example, the center grid point did not forecast the field of interest (e.g., 6-h accumulated precipitation) at the threshold of interest, even though it was observed. In a traditional grid-point verification, this would be classified as a miss (0/1 forecast, 1/1 observation). However, for both forecast and observation, seven grid points within the twenty-five grid-point neighborhood met the threshold of interest. For fractions skill score, this would result in a "perfect" neighborhood forecast (7/25 forecast, 7/25 observation). Adapted from Roberts and Lean (2008; their Fig. 2) and Schwartz et al. (2010; their Fig. 11).

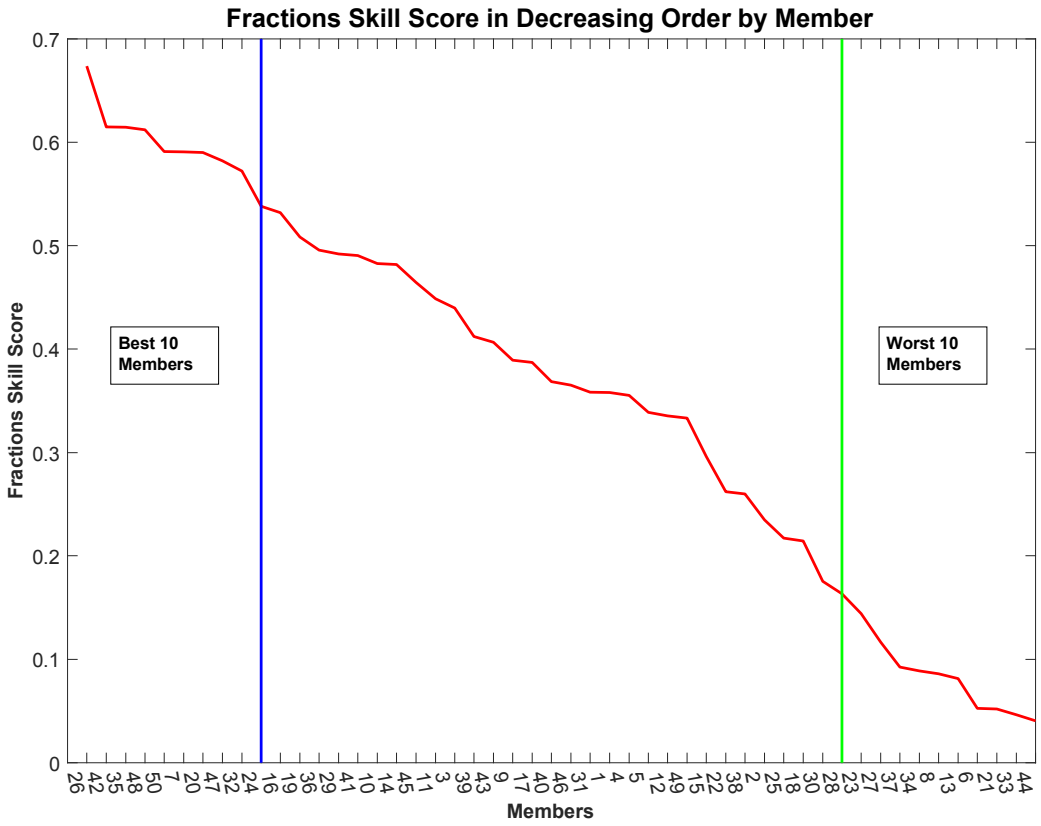


Figure 7. FSS for the 2200 UTC 31 May 2013 to 0400 UTC 1 June 2013 period over Oklahoma in order of decreasing skill, with the best 10 members to the left of the blue line and the worst 10 members to the right of the green line. The *x*-axis indicates the ensemble member number associated with a given FSS.

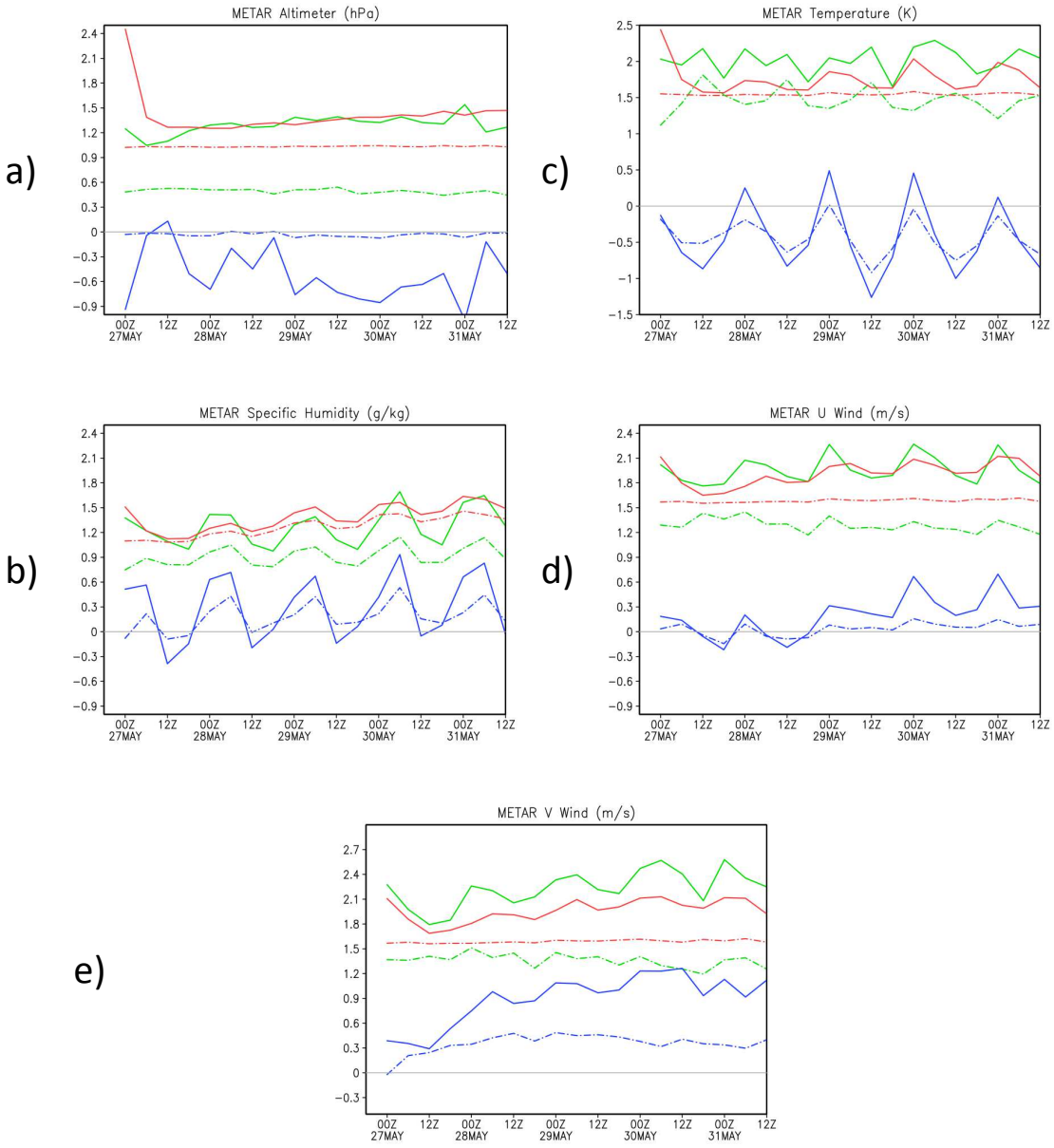
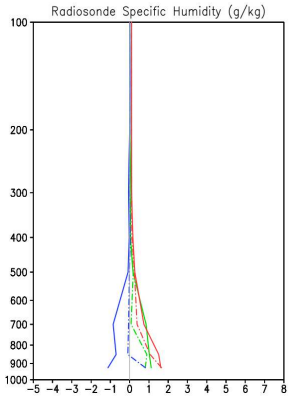
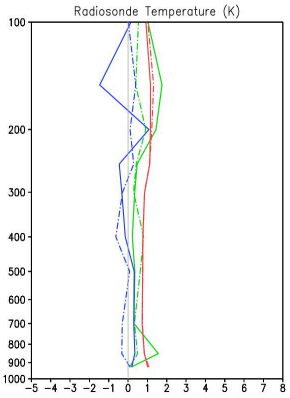


Figure 8: RMSE (green), total spread (ensemble spread plus assumed observation error; red), and bias (blue) of the prior (solid) and posterior (dashed) ensemble estimates for METAR (a) altimeter (hPa), (b) temperature (K), (c) specific humidity (g kg^{-1}), (d) zonal wind (m s^{-1}), and (e) meridional wind (m s^{-1}) observations between 1800 UTC 26 May 2013 and 1200 UTC 31 May 2013.

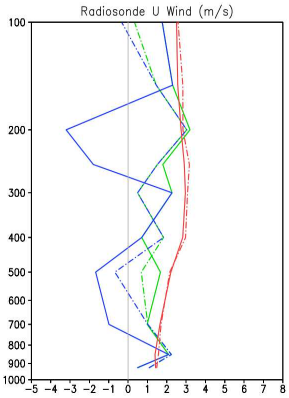
a)



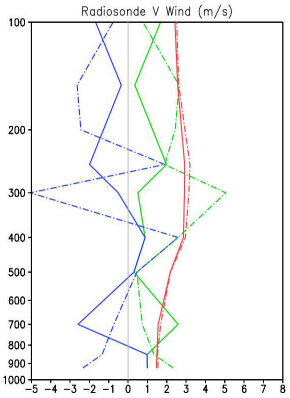
b)



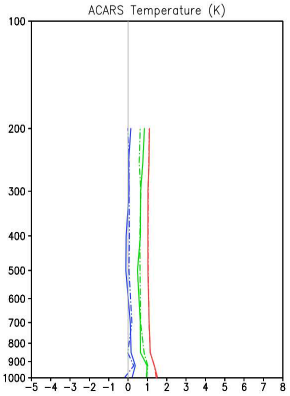
c)



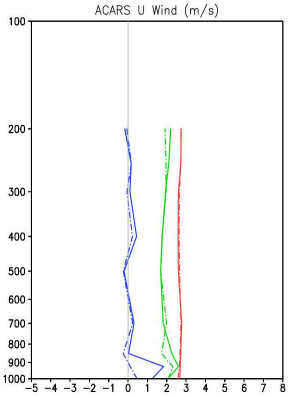
d)



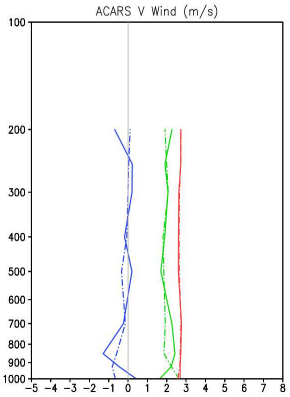
f)



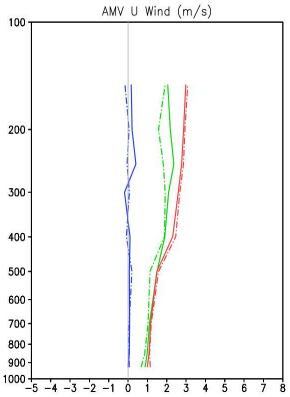
g)



h)



i)



j)

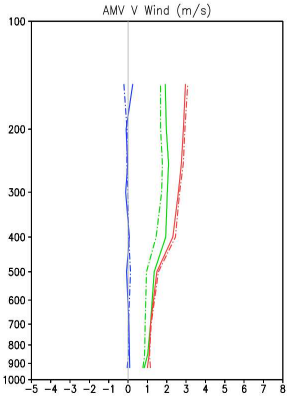


Figure 9. RMSE (green), total spread (ensemble spread plus assumed observation error; red), and bias (blue) of the posterior analyses for 1200 UTC 28 May 2013 (solid) and 1200 UTC 31 May 2013 for radiosonde (a) specific humidity (g kg^{-1}), (b) temperature (K), (c) zonal wind (m s^{-1}), and (d) meridional wind (m s^{-1}); ACARS (e) specific humidity (g kg^{-1}), (f) temperature (K), (g) zonal wind (m s^{-1}), and (h) meridional wind (m s^{-1}); and AMV (i) zonal wind (m s^{-1}), and (j) meridional wind (m s^{-1}).

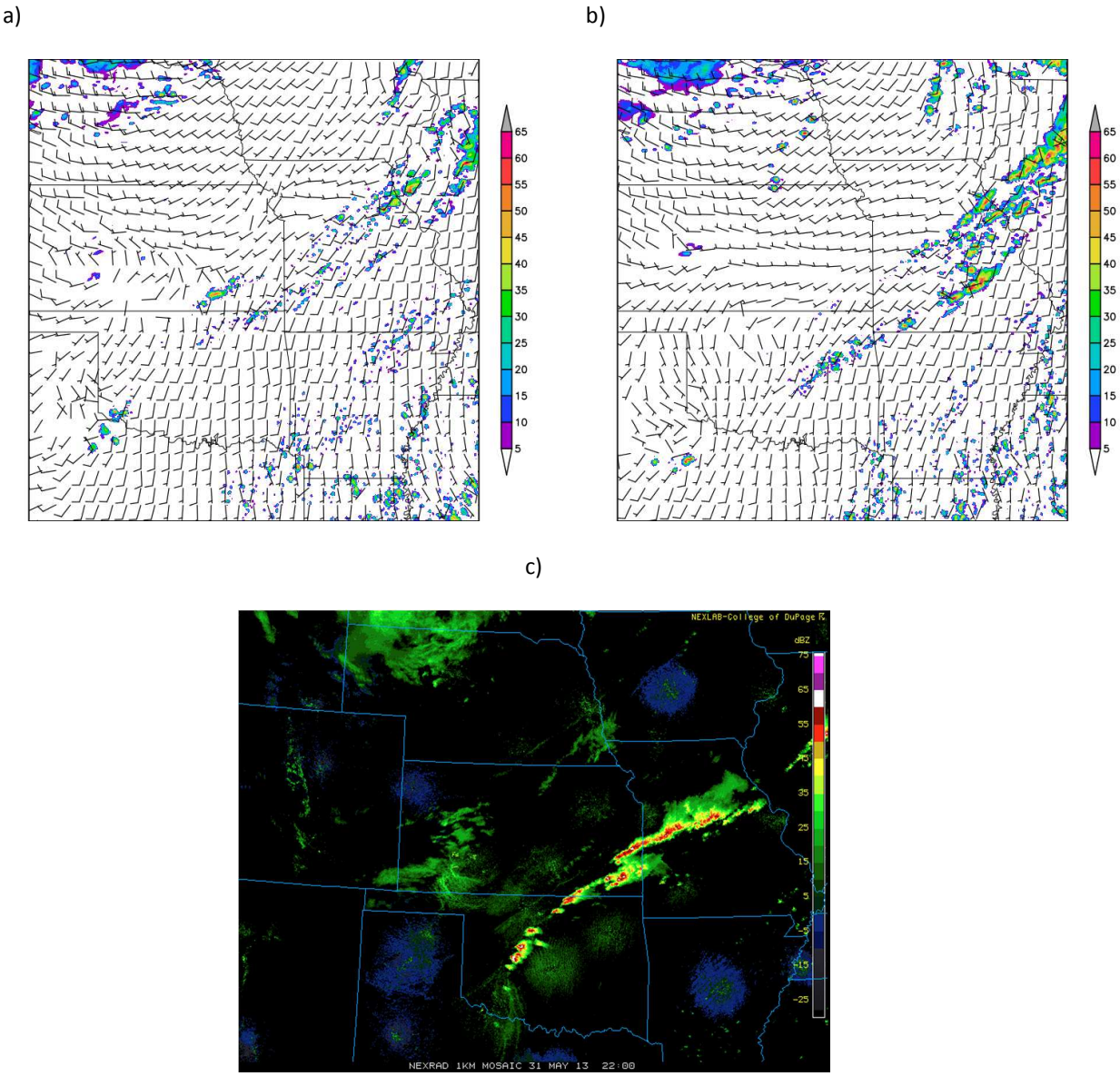


Figure 10. Reflectivity (dBZ, shaded per the color bar in each panel) shown near initiation at 2200 UTC 31 May 2013 for the (a) best ensemble member (26), (b) worst ensemble member (44), and (c) the actual case (http://www2.mmm.ucar.edu/imagearchive/). In panels (a) and (b), barbs indicate 10-m wind (half barb: 2.5 m s^{-1}, barb: 5 m s^{-1}, pennant: 25 m s^{-1}).

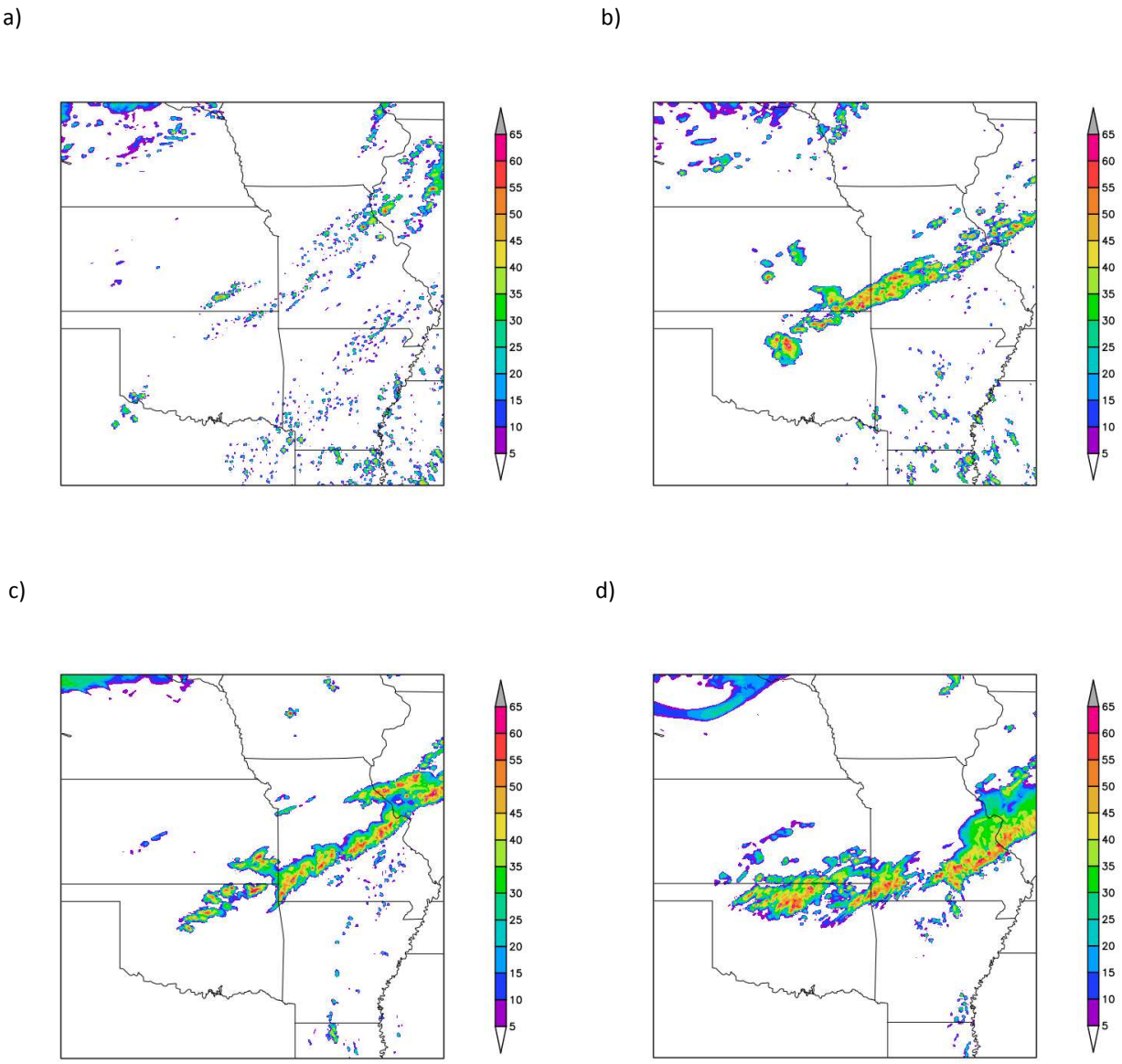


Figure 11. Reflectivity (dBZ, shaded) from ensemble member 26 (the best-performing member, in terms of FSS) at (a) 2200 UTC 31 May 2013, (b) 0100 UTC, (c) 0400 UTC, (d) 0700 UTC on 1 June 2013.

a)

b)

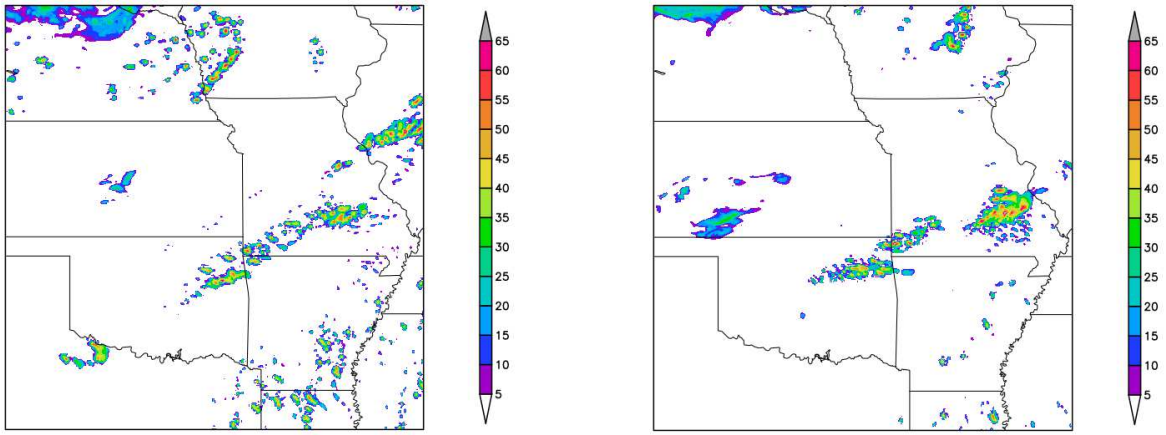


Figure 12. Reflectivity (dBZ, shaded) from ensemble member 44 (the worst-performing member, in terms of FSS) at (a) 0000 UTC and (b) 0300 UTC on 1 June 2013.

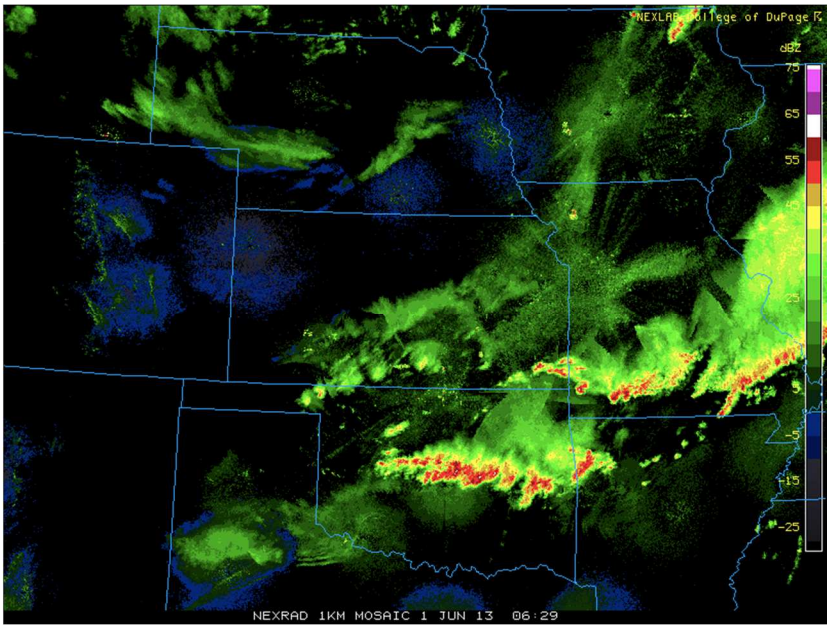


Figure 13. Reflectivity (dBZ) from the actual case (http://www2.mmm.ucar.edu/imagearchive/) at 0630 UTC on 1 June 2013.

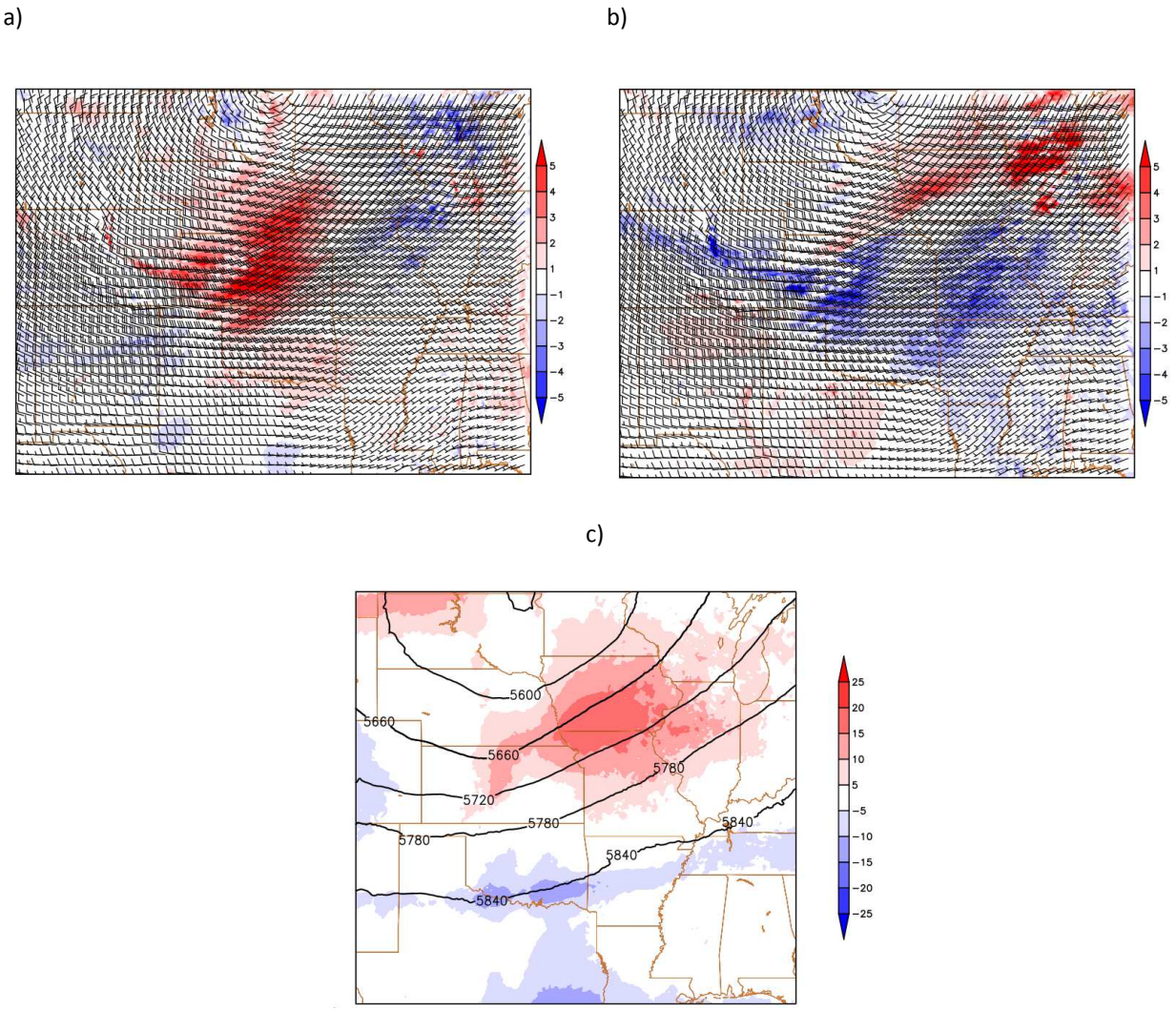


Figure 14. Composite differences (shaded) and full-ensemble–mean (contoured) of 500 hPa (a) meridional wind (m s^{-1}), (b) zonal wind (m s^{-1}), and (c) geopotential height (m) at 2200 UTC 31 May 2013. Red (blue) indicates higher values in the average of the best (worst) 10 ensemble members. Barbs in (a) and (b) indicate the full-ensemble–mean 500 hPa horizontal wind (half barb: 5 kt, full barb: 10 kt, pennant: 50 kt).

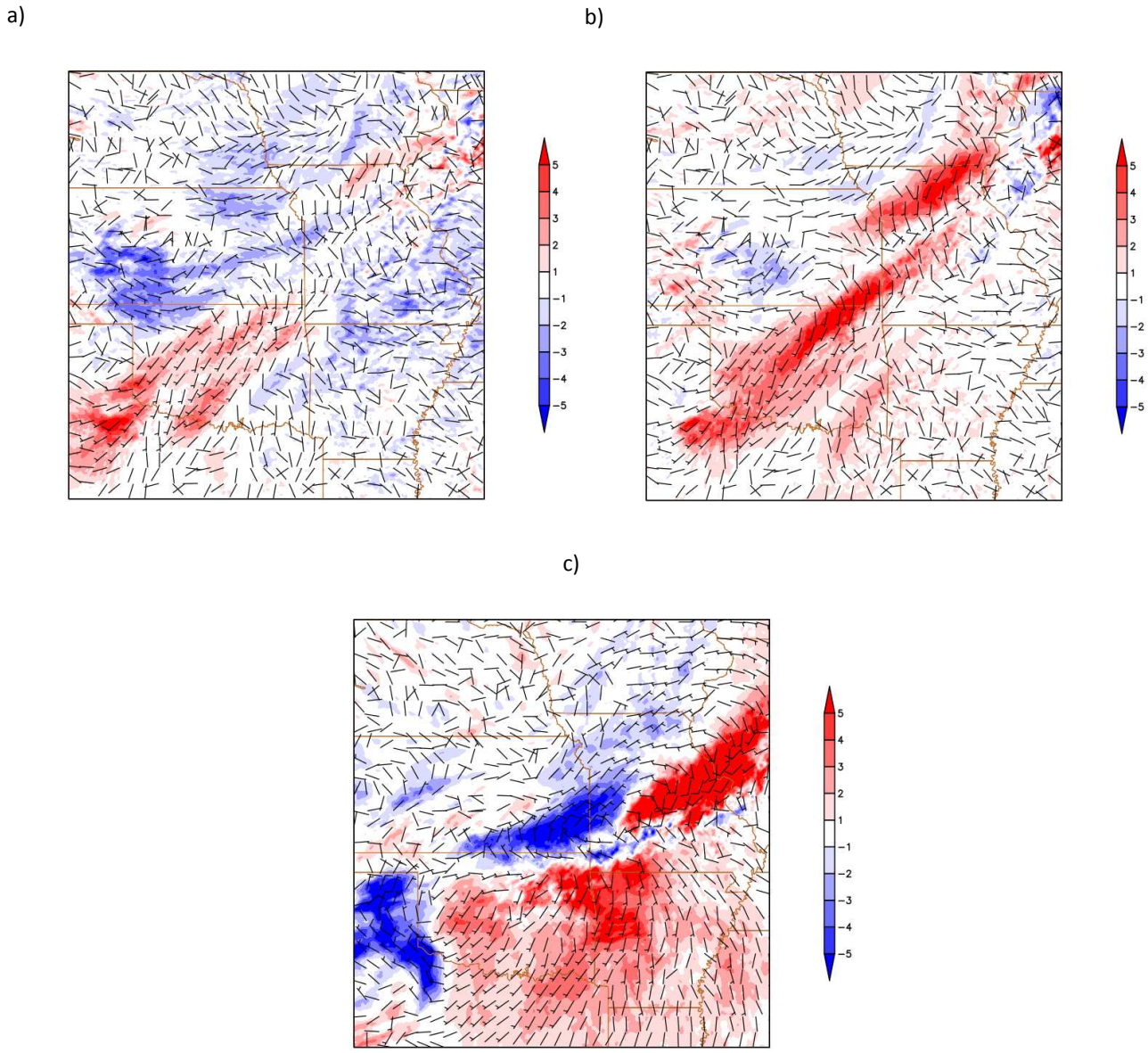


Figure 15. Composite differences (shaded) and full-ensemble–mean (contoured) of 850 hPa (a) zonal wind difference (m s^{-1}), (b) meridional wind difference (m s^{-1}) at 2200 UTC 31 May 2013, and (c) meridional wind difference (m s^{-1}) at 0500 UTC 1 June 2013. Barbs in each panel indicate the vector wind difference between the best- and worst-performing composite means (half barb: 5 kt, full barb: 10 kt, pennant: 50 kt).

a)

b)

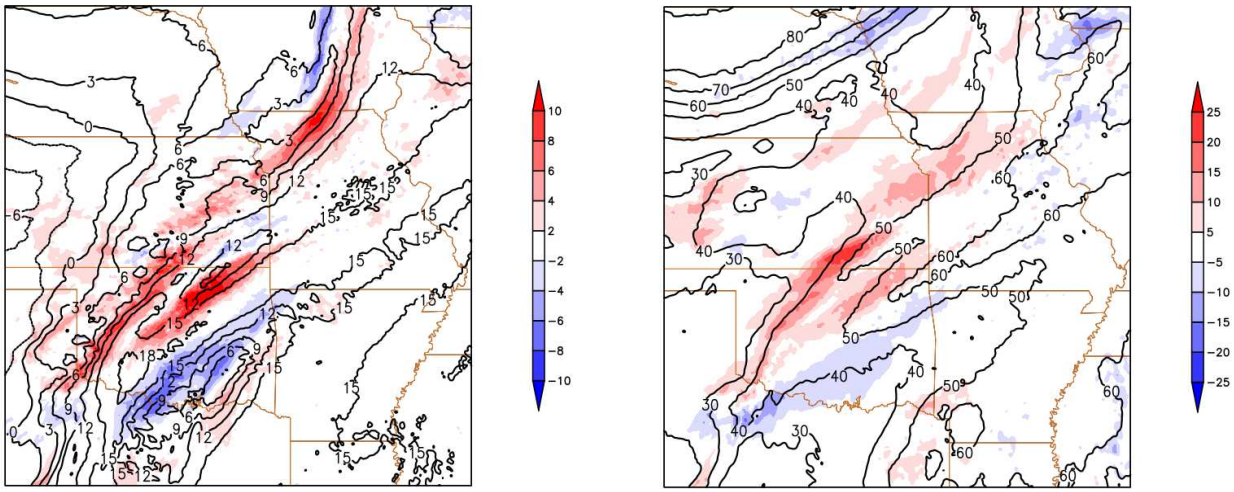


Figure 16. Composite differences (shaded) and full-ensemble–mean (contoured) of (a) 850 hPa dew point temperature (°C) and 1000-500 hPa average relative humidity (%) at 2200 UTC 31 May 2013.

a) b)

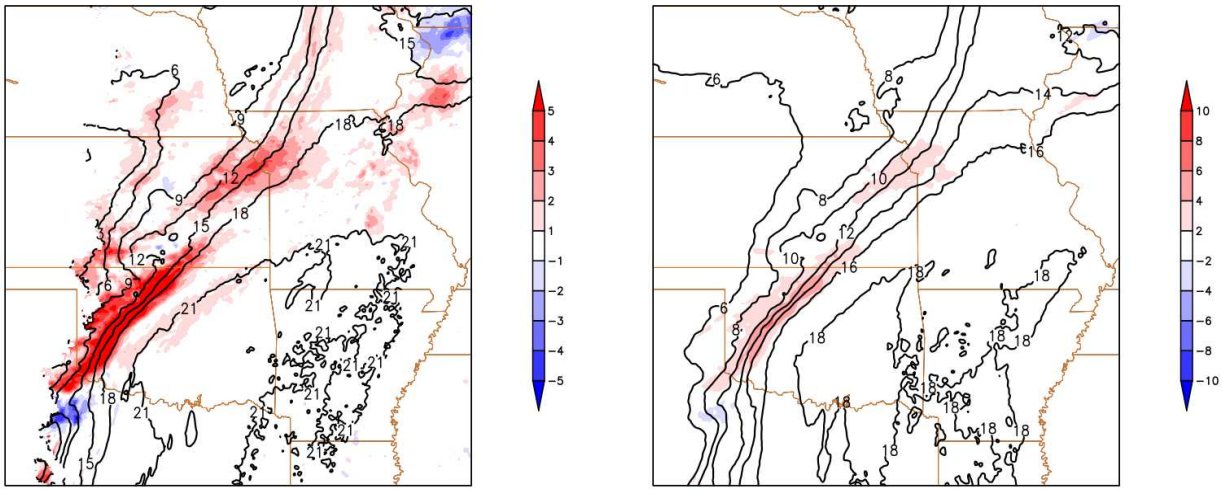


Figure 17. Composite differences (shaded) and full-ensemble–mean (contoured) of (a) 925 hPa dew point temperature (°C) and surface to 850 hPa average water vapor mixing ratio (g kg^{-1}) at 2200 UTC 31 May 2013.

a) b)

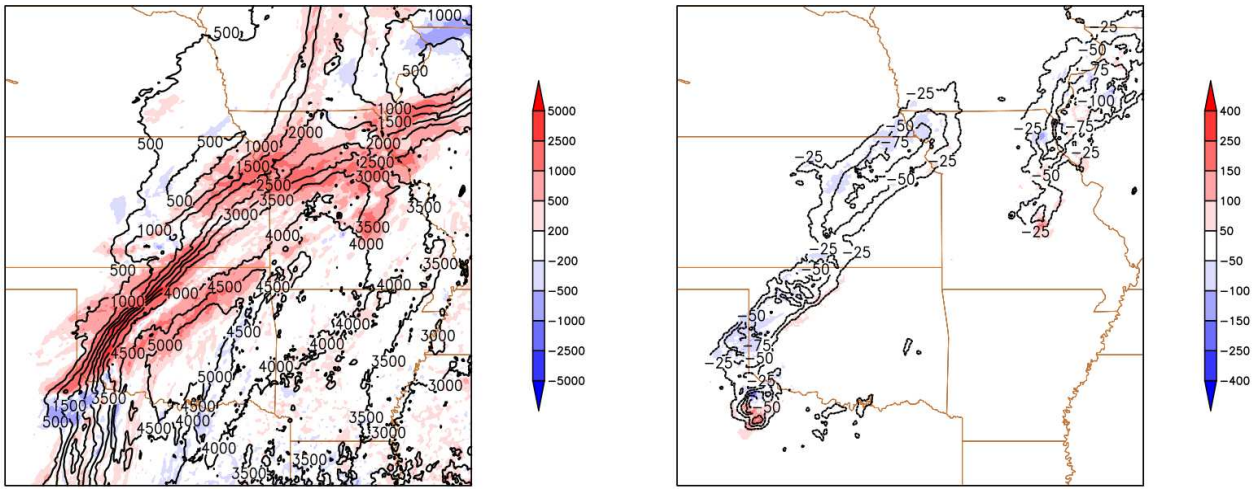
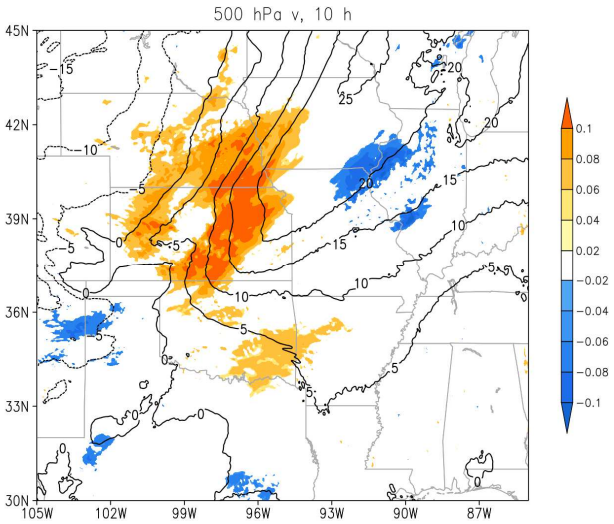


Figure 18. Composite differences (shaded) and full-ensemble–mean (contoured) of (a) surface-based convective available potential energy (J kg^{-1}) and (b) surface-based convective inhibition (J kg^{-1}) at 2200 UTC 31 May 2013.

a)



b)

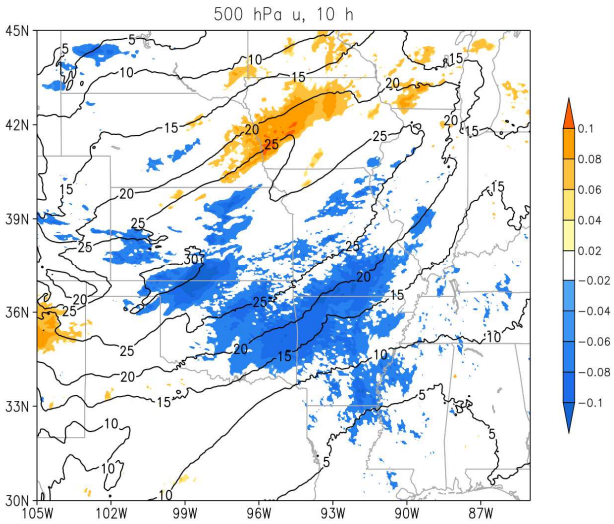
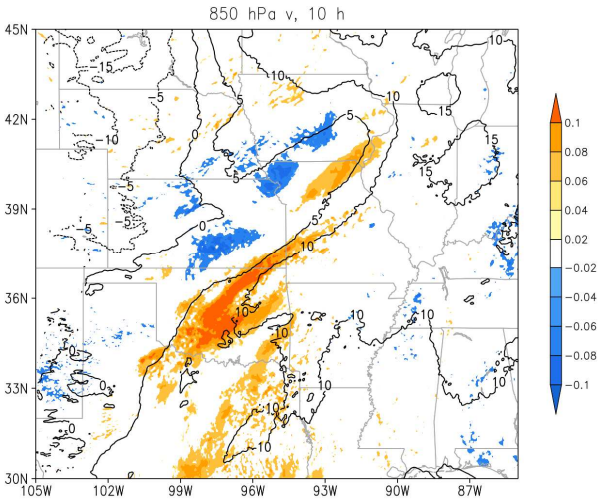


Figure 19. Sensitivity of 2200 UTC 31 May 2013 – 0400 UTC 1 June 2013 FSS for a 25-km radius neighborhood over the domain in Fig. 6 to 500 hPa (a) meridional wind (m s^{-1}) and (b) zonal winds (m s^{-1}) at 2200 UTC 31 May 2013.

a)



b)

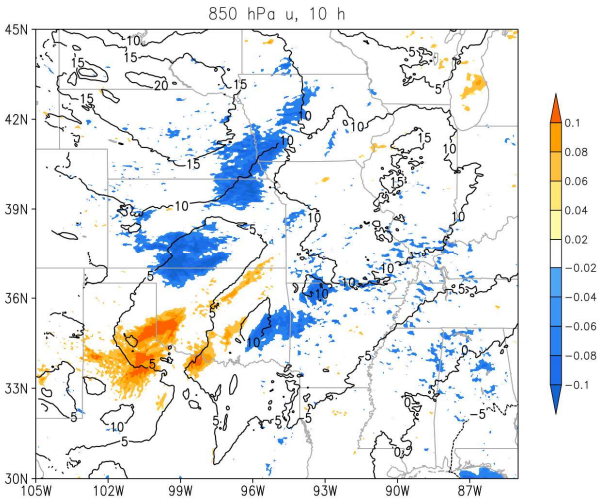


Figure 20. Sensitivity of 2200 UTC 31 May 2013 – 0400 UTC 1 June 2013 FSS for a 25-km radius neighborhood over the domain in Fig. 6 to 850 hPa (a) meridional wind (m s^{-1}) and (b) zonal winds (m s^{-1}) at 2200 UTC 31 May 2013.

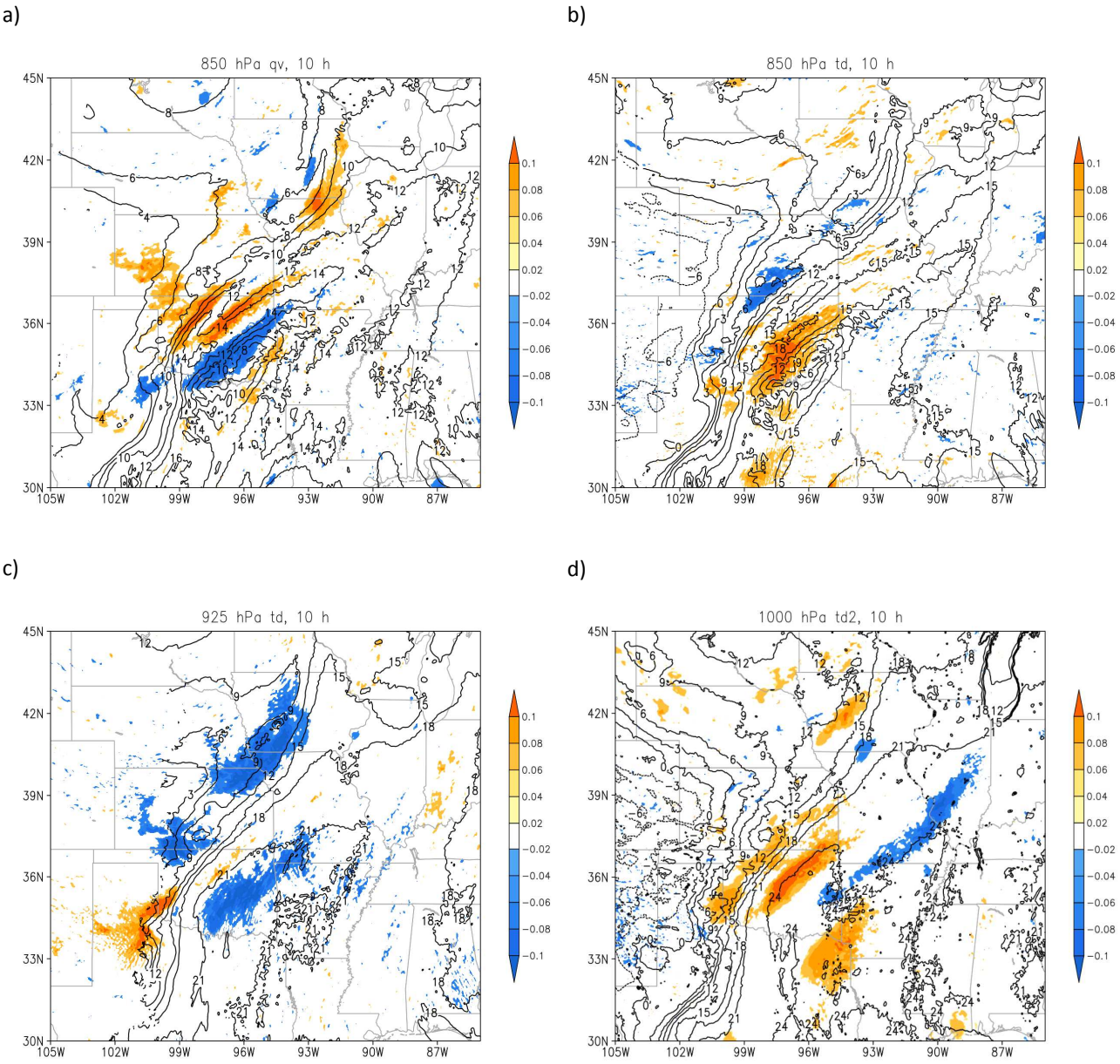


Figure 21. Sensitivity of 2200 UTC 31 May 2013 – 0400 UTC 1 June 2013 FSS for a 25-km radius neighborhood over the domain in Fig. 6 of (a) 850 hPa water vapor mixing ratio (g kg^{-1}), (b) 850 hPa dew point temperature (°C), (c) 925 hPa dew point temperature (°C), and (d) surface dew point temperature (°C) at 2200 UTC 31 May 2013.

a) b)

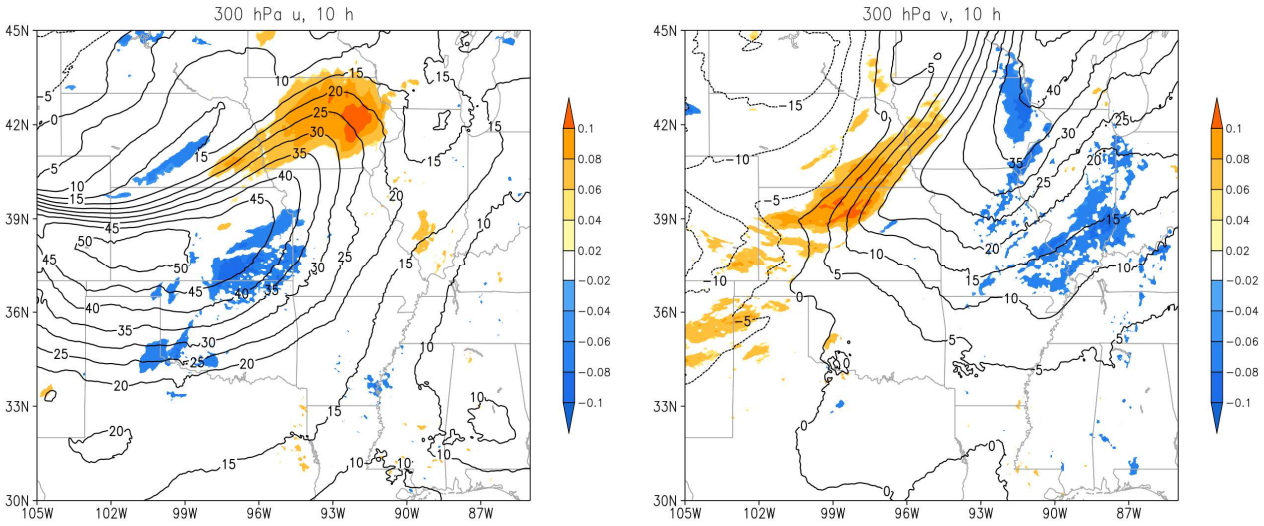
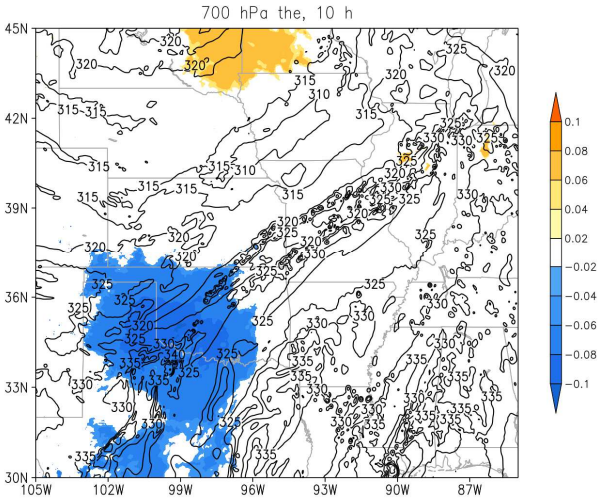


Figure 22. Sensitivity of 2200 UTC 31 May 2013 – 0400 UTC 1 June 2013 FSS for a 25-km radius neighborhood over the domain in Fig. 3 of 300 hPa (a) zonal wind (m s^{-1}) and (b) meridional wind (m s^{-1}) at on 2200 UTC 31 May 2013. Orange indicates that a one standard deviation increase in the contoured field will result in the shaded increase in FSS, whereas blue indicates that a one standard deviation increase in the contoured field will result in the shaded decrease in FSS.

a)



b)

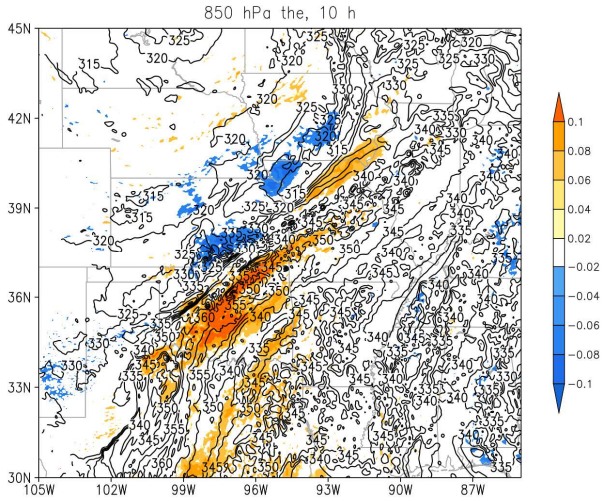
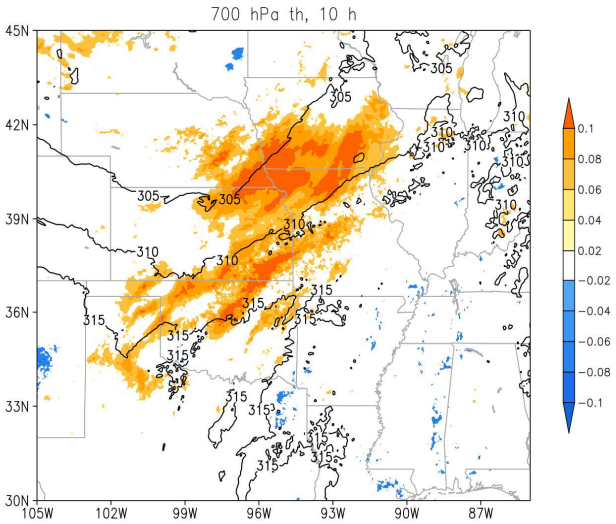


Figure 23. Sensitivity of 2200 UTC 31 May 2013 – 0400 UTC 1 June 2013 FSS for a 25-km radius neighborhood over the domain in Fig. 6 of equivalent potential temperature (K) at (a) 700 hPa and (b) 850 hPa on 2200 UTC 31 May 2013.

a)



b)

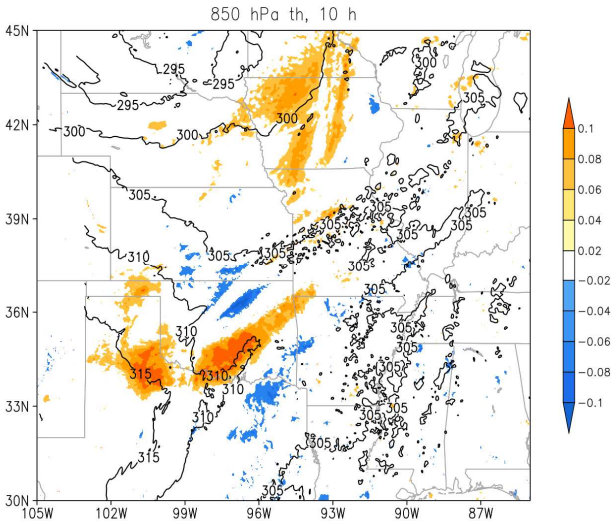


Figure 24. Sensitivity of 2200 UTC 31 May 2013 – 0400 UTC 1 June 2013 FSS for a 25-km radius neighborhood over the domain in Fig. 6 of potential temperature (K) at (a) 700 hPa and (b) 850 hPa on 2200 UTC 31 May 2013.

References

Adams-Selin, R. D., S. C. van den Heever, and R. H. Johnson, 2013: Sensitivity of bow-echo simulation to microphysical parameterizations. *Wea. Forecasting*, **28**, 1188–1209.

Anderson, J. L. and N. Collins, 2007: Scalable implementations of ensemble filter algorithms for data assimilation. *J. Atmos. Oceanic Technol.,* **24**, 1452–1463.

Anderson, J. L., T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Avellano, 2009: The data assimilation research testbed: A community facility. *Bull. Amer. Meteor. Soc.,* **90**, 1283–1296.

Anderson, J. L., 2009: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus A*, **61**, 72–83.

Anderson, J. L., 2012: Localization and sampling error correction in ensemble Kalman filter data assimilation. *Mon. Wea. Rev.,* **140**, 2359–2371.

Barker, D., and coauthors, 2012: The weather research and forecasting model's community variational/ensemble data assimilation system: WRFDA. *Bull. Amer. Meteor. Soc.,* **93**, 831–843.

Berman, J. D., R. D. Torn, G. S. Romine, and M. L. Weisman, 2017: Sensitivity of northern Great Plains convection forecasts to upstream and downstream forecast errors. *Mon. Wea. Rev.*, **145**, 2141–2163.

Burghardt, B. J., C. Evans, and P. J. Roebber, 2014: Assessing the predictability of convection initiation in the high plains using an object-based approach. *Wea. Forecasting,* **29**, 403–418.

Burlingame, B. M., C. Evans, and P. J. Roebber, 2017: The influence of PBL parameterization

on the practical predictability of convection initiation during the mesoscale predictability experiment (MPEX). *Wea. Forecasting,* **32**, 1161–1183.

Chen, F. and J. Dudhia, 2001: Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Mon. Wea. Rev.,* **129**, 569–585.

Chen, F. and J. Dudhia, 2001: Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part II: Preliminary model validation. *Mon. Wea. Rev.,* **129**, 587–604.

Coniglio, M. C., H. E. Brooks, S. J. Weiss, and S. F. Corfidi, 2007: Forecasting the maintenance of quasi-linear mesoscale convective systems. *Wea. Forecasting,* **22**, 556–570.

Crook, N. A., 1996: Sensitivity of moist convection forced by boundary layer processes to low-level thermodynamic fields. *Mon. Wea. Rev.,* **124**, 1767–1785.

Dowell, D. C., F. Zhang, L. J. Wicker, C. Snyder, and N. A. Crook, 2004: Wind and temperature retrievals in the 17 May 1981 Arcadia, Oklahoma, supercell: Ensemble Kalman filter experiments. *Mon. Wea. Rev.*, **132**, 1982–2005.

Duda, J. D. and W. A. Gallus, 2013: The impact of large-scale forcing on skill of simulated convective initiation and upscale evolution with convection-allowing grid spacings in the WRF. *Wea. Forecasting,* **28**, 994–1018.

Durran, D. R. and J. A. Weyn, 2016: Thunderstorms do not get butterflies. *Bull. Amer. Meteor. Soc.,* **97**, 237–243.

Evans, C., M. L. Weisman, and L. F. Bosart, 2014: Development of an intense, warm-core mesoscale vortex associated with the 8 May 2009 “super derecho” convective event. *J. Atmos. Sci.*, **71**, 1218-1240.

Fowler, T. L., T. Jensen, E. I. Tollerud, J. Halley Gotway, P. Oldenburg, and R. Bullock, 2010: New Model Evaluation Tools (MET) software capabilities for QPF verification. Preprints, *Third Int. Conf. on QPE/QPF and Hydrology, Nanjing, China, World Weather Research Programme*, 189–193.

Gaspari, G. and Cohn, S. E., 1999: Construction of correlation functions in two and three dimensions. *Q.J.R. Meteorol. Soc.*, **125**, 723–757.

Gilmore, M. S. and L. J. Wicker, 1998: The influence of midtropospheric dryness on supercell morphology and evolution. *Mon. Wea. Rev.,* **126**, 943–958.

Gilmour, I., L. A. Smith, and R. Buizza, 2001: Linear regime duration: Is 24 hours a long time in synoptic weather forecasting? *J. Atmos. Sci.*, **58**, 3525–3539.

Grunzke, C. T. and C. Evans, 2017: Predictability and dynamics of warm-core mesoscale vortex formation with the 8 May 2009 "super derecho" event. *Mon. Wea. Rev.*, **145**, 811–832.

Gowan, T. M., W. J. Steenburgh, and C. S. Schwartz, 2018: Validation of mountain precipitation forecasts from the convection-permitting NCAR ensemble and operational forecast systems over the western United States. *Wea. Forecasting*, in press.

Ha, S. and C. Snyder, 2014: Influence of surface observations in mesoscale data assimilation using an ensemble Kalman filter. *Mon. Wea. Rev.,* **142**, 1489–1508.

Hill, A. J., C. C. Weiss, and B. C. Ancell, 2016: Ensemble sensitivity analysis for mesoscale forecasts of dryline convection initiation. *Mon. Wea. Rev.,* **144**, 4161–4182.

Hu, X.-M., J. W. Nielsen-Gammon, and F. Zhang, 2010: Evaluation of three planetary boundary layer schemes in the WRF model. *J. Appl. Meteor. Climatol.*, **49**, 1831–1844.

Hitchcock, S. M., M. C. Coniglio, and K. H. Knopfmeier, 2016: Impact of MPEX upsonde

observations on ensemble analyses and forecasts of the 31 May 2013 convective event over Oklahoma. *Mon. Wea. Rev.,* **144**, 2889–2913.

Hohenegger, C., D. Lüthi, and C. Schär, 2006: Predictability mysteries in cloud-resolving models. *Mon. Wea. Rev.,* **134**, 2095–2107.

Hohenegger, C. and C. Schär, 2007: Predictability and error growth dynamics in cloud-resolving models. *J. Atmos. Sci.,* **64**, 4467–4478.

Houze, R. A. Jr., 2004: Mesoscale convective systems, *Rev. Geophys.*, **42**, RG4003.

Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models, *J. Geophys. Res.*, **113**, D13103.

Janjić, Z. I., 1994: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.,* **122**, 927–945.

Kain, J. S., and coauthors, 2013: A feasibility study for probabilistic convection initiation forecasts based on explicit numerical guidance. *Bull. Amer. Meteor. Soc.,* **94**,1213–1225.

Keclik, A. M., C. Evans, P. J. Roebber, and G. S. Romine, 2017: The influence of assimilated upstream, preconvective dropsonde observations on ensemble forecasts of convection initiation during the mesoscale predictability experiment. *Mon. Wea. Rev.,* **145**, 4747–4770.

Lilly, D. K., 1990: Numerical prediction of thunderstorms—has its time come?. *Q.J.R. Meteorol. Soc.*, **116**, 779–798.

Lin, Y. and K. E. Mitchell, 2005: The NCEP stage II/IV hourly precipitation analyses:

development and applications. *Preprints, 19th Conf. on Hydrology, American Meteorological Society*, San Diego, CA, 9-13 January 2005, Paper 1.2.

Madaus, L. E., and G. J. Hakim, 2016: Observable surface anomalies preceding simulated isolated convection initiation. *Mon. Wea. Rev.*, **144**, 2265–2284.

Markowski, P., C. Hannon, and E. Rasmussen, 2006: Observations of convection initiation "failure" from the 12 June 2002 IHOP deployment. *Mon. Wea. Rev.*, **134**, 375–405.

Melhauser, C. and F. Zhang, 2012: Practical and intrinsic predictability of severe and convective weather at the mesoscales. *J. Atmos. Sci.,* **69**, 3350–3371.

Park, S. K., 1999: Nonlinearity and predictability of convective rainfall associated with water vapor perturbations in a numerically simulated storm, *J. Geophys. Res.*, **104(D24)**, 31575–31587.

Roberts, N., 2005: An investigation of the ability of a storm scale configuration of the Met Office NWP model to predict flood producing rainfall. *UK Met Office Tech. Rep*. **455**, 80pp.

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

Romine, G. S., C. S. Schwartz, C. Snyder, J. L. Anderson, and M. L. Weisman, 2013: Model bias in a continuously cycled assimilation system and its influence on convection-permitting forecasts. *Mon. Wea. Rev.,* **141**, 1263–1284.

Romine, G. S., C. S. Schwartz, J. Berner, K. R. Fossell, C. Snyder, J. L. Anderson, and M. L. Weisman, 2014: Representing forecast error in a convection-permitting ensemble system. *Mon. Wea. Rev.,* **142**, 4519–4541.

Romine, G. S., C. S. Schwartz, R. D. Torn, and M. L. Weisman, 2016: Impact of assimilating

dropsonde observations from MPEX on ensemble forecasts of severe weather events. *Mon. Wea. Rev.*, **144**, 3799–3823.

Schumacher, R. S. and R. H. Johnson, 2005: Organization and environmental properties of extreme-rain-producing mesoscale convective systems. *Mon. Wea. Rev.,* **133**, 961–976.

Schumacher, R. S., 2015: Resolution dependence of initiation and upscale growth of deep convection in convection-allowing forecasts of the 31 May–1 June 2013 supercell and MCS. *Mon. Wea. Rev.,* **143**, 4331–4354.

Schwartz, C. S., and coauthors, 2009: Next-day convection-allowing WRF model guidance: a second look at 2-km versus 4-km grid spacing. *Mon. Wea. Rev.*, **137**, 3351–3372.

Schwartz, C. S., and coauthors, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Wea. Forecasting,* **25**, 263–280.

Schwartz, C. S., G. S. Romine, K. R. Smith, and M. L. Weisman, 2014: Characterizing and optimizing precipitation forecasts from a convection-permitting ensemble initialized by a mesoscale ensemble Kalman filter. *Wea. Forecasting,* **29**, 1295–1318.

Schwartz, C. S., G. S. Romine, R. A. Sobash, K. R. Fossell, and M. L. Weisman, 2015: NCAR's experimental real-time convection-allowing ensemble prediction system. *Wea. Forecasting,* **30**, 1645–1654.

Schwartz, C. S., G. S. Romine, M. L. Weisman, R. A. Sobash, K. R. Fossell, K. W. Manning, and S. B. Trier, 2015: A real-time convection-allowing ensemble prediction system initialized by mesoscale ensemble Kalman filter analyses. *Wea. Forecasting,* **30**, 1158–1181.

Schwartz, C. S., 2016: Improving large-domain convection-allowing forecasts with high-

resolution analyses and ensemble data assimilation. *Mon. Wea. Rev.*, **144**, 1777–1803.

Schwartz, C. S., G. S. Romine, K. R. Fossell, R. A. Sobash, and M. L. Weisman, 2017: Toward 1-km ensemble forecasts over large domains. *Mon. Wea. Rev.*, **145**, 2943–2969.

Schwartz, C. S. and R. A. Sobash, 2017: Generating probabilistic forecasts from convection-allowing ensembles using neighborhood approaches: a review and recommendations. *Mon. Wea. Rev.*, **145**, 3397–3418.

Skamarock, W. C., and coauthors, 2008: A description of the advanced research WRF version 3, pp. 1-113.

Sobash, R. A., C. S. Schwartz, G. S. Romine, K. R. Fossell, and M. L. Weisman, 2016: Severe weather prediction using storm surrogates from an ensemble forecasting system. *Wea. Forecasting*, **31**, 255–271.

Sobash, R. A., G. S. Romine, C. S. Schwartz, D. J. Gagne, and M. L. Weisman, 2016: Explicit forecasts of low-level rotation from convection-allowing models for next-day tornado prediction. *Wea. Forecasting*, **31**, 1591–1614.

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

Tiedtke, M., 1989: A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Mon. Wea. Rev.,* **117**, 1779–1800.

Torn, R. D., G. J. Hakim, and C. Snyder, 2006: Boundary conditions for limited-area ensemble Kalman filters. *Mon. Wea. Rev.,* **134**, 2490–2502.

Torn, R. D. and G. J. Hakim, 2008: Ensemble-based sensitivity analysis. *Mon. Wea. Rev.*, **136**, 663–677.

Torn, R. D. and G. J. Hakim, 2009: Initial condition sensitivity of western pacific extratropical transitions determined using ensemble-based sensitivity analysis. *Mon. Wea. Rev.*, **137**, 3388–3406.

Torn, R. D., 2010: Performance of a mesoscale ensemble Kalman filter (EnKF) during the NOAA high-resolution hurricane test. *Mon. Wea. Rev.,* **138**, 4375–4392.

Torn, R. D., 2010: Ensemble-based sensitivity analysis applied to African easterly waves. *Wea. Forecasting*, **25**, 61–78.

Torn, R. D., 2010: Diagnosis of the downstream ridging associated with extratropical transition using short-term ensemble forecasts. *J. Atmos. Sci.*, **67**, 817–833.

Torn, R. D. and G. S. Romine, 2015: Sensitivity of central Oklahoma convection forecasts to upstream potential vorticity anomalies during two strongly forced cases during MPEX. *Mon. Wea. Rev.,* **143**, 4064–4087.

Torn, R. D., G. S. Romine, and T. J. Galarneau, 2017: Sensitivity of dryline convection forecasts to upstream forecast errors for two weakly forced MPEX cases. *Mon. Wea. Rev.*, **145**, 1831–1852.

Trapp, R. J., D. J. Stensrud, M. C. Coniglio, R. S. Schumacher, M. E. Baldwin, S. Waugh, and D. T. Conlee, 2016: Mobile radiosonde deployments during the mesoscale predictability experiment (MPEX): Rapid and adaptive sampling of upscale convective feedbacks. *Bull. Amer. Meteor. Soc.,* **97**, 329–336.

Wandishin, M. S., D. J. Stensrud, S. L. Mullen, and L. J. Wicker, 2008: On the predictability of mesoscale convective systems: Two-dimensional simulations. *Wea. Forecasting,* **23**, 773–785.

Wandishin, M. S., D. J. Stensrud, S. L. Mullen, and L. J. Wicker, 2010: On the predictability of

mesoscale convective systems: Three-dimensional simulations. *Mon. Wea. Rev.,* **138**, 863–885.

Weckwerth, T. M., 2000: The effect of small-scale moisture variability on thunderstorm initiation. *Mon. Wea. Rev.,* **128**, 4017–4030.

Weckwerth, T. M. and D. B. Parsons, 2006: A review of convection initiation and motivation for IHOP_2002. *Mon. Wea. Rev.,* **134**, 5–22.

Weckwerth, T. M., H. V. Murphey, C. Flamant, J. Goldstein, and C. R. Pettet, 2008: An observational study of convection initiation on 12 June 2012 during IHOP_2002. *Mon. Wea. Rev.*, **136**, 2283–2304.

Weisman, M. L., 1992: The role of convectively generated rear-inflow jets in the evolution of long-lived mesoconvective systems. *J. Atmos. Sci.,* **49**, 1826–1847.

Weisman, M. L., C. Davis, W. Wang, K.W. Manning, and J. B. Klemp, 2008: Experiences with 0–36-h explicit convective forecasts with the WRF-ARW model. *Wea. Forecasting*, **23**, 407–437.

Weisman, M. L., C. Evans, and L. F. Bosart, 2013: The 8 May 2009 "super derecho": Analysis of a realtime explicit convective forecast. *Wea. Forecasting*, **28**, 863-892.

Weisman, M. L., and coauthors, 2015: The mesoscale predictability experiment (MPEX). *Bull. Amer. Meteor. Soc.,* **96**, 2127–2149.

Weyn, J. A. and D. R. Durran, 2017: The dependence of the predictability of mesoscale convective systems on the horizontal scale and amplitude of initial errors in idealized simulations. *J. Atmos. Sci.*, **74**, 2191–2210.

Velden, C., and coauthors, 2005: Recent innovations in deriving tropospheric winds from meteorological satellites. *Bull. Amer. Meteor. Soc.,* **86**, 205–224.

Yussouf, N., J. S. Kain, and A. J. Clark, 2016: Short-term probabilistic forecasts of the 31 May 2013 Oklahoma tornado and flash flood event using a continuous-update-cycle storm-scale ensemble system. *Wea. Forecasting,* **31**, 957–983.

Zhang, C., Y. Wang, and K. Hamilton, 2011: Improved representation of boundary layer clouds over the southeast Pacific in ARW-WRF using a modified Tiedtke cumulus parameterization scheme. *Mon. Wea. Rev.,* **139**, 3489–3513.