A Real-Time Convection-Allowing Ensemble Prediction System Initialized by Mesoscale Ensemble Kalman Filter Analyses

CRAIG S. SCHWARTZ, GLEN S. ROMINE, MORRIS L. WEISMAN, RYAN A. SOBASH, KATHRYN R. FOSSELL, KEVIN W. MANNING, AND STANLEY B. TRIER

National Center for Atmospheric Research,* Boulder, Colorado

(Manuscript received 21 January 2015, in final form 12 May 2015)

ABSTRACT

In May and June 2013, the National Center for Atmospheric Research produced real-time 48-h convection-allowing ensemble forecasts at 3-km horizontal grid spacing using the Weather Research and Forecasting (WRF) Model in support of the Mesoscale Predictability Experiment field program. The ensemble forecasts were initialized twice daily at 0000 and 1200 UTC from analysis members of a continuously cycling, limited-area, mesoscale (15 km) ensemble Kalman filter (EnKF) data assimilation system and evaluated with a focus on precipitation and severe weather guidance. Deterministic WRF Model forecasts initialized from GFS analyses were also examined. Subjectively, the ensemble forecasts often produced areas of intense convection over regions where severe weather was observed. Objective statistics confirmed these subjective impressions and indicated that the ensemble was skillful at predicting precipitation and severe weather events. Forecasts initialized at 1200 UTC were more skillful regarding precipitation and severe weather placement than forecasts initialized 12 h earlier at 0000 UTC, and the ensemble forecasts were typically more skillful than GFS-initialized forecasts. At times, 0000 UTC GFS-initialized forecasts had temporal distributions of domain-average rainfall closer to observations than EnKF-initialized forecasts. However, particularly when GFS analyses initialized WRF Model forecasts, 1200 UTC forecasts produced more rainfall during the first diurnal maximum than 0000 UTC forecasts. This behavior was mostly attributed to WRF Model initialization of clouds and moist physical processes. The success of these real-time ensemble forecasts demonstrates the feasibility of using limited-area continuously cycling EnKFs as a method to initialize convection-allowing ensemble forecasts, and future real-time high-resolution ensemble development leveraging EnKFs seems justified.

1. Introduction

Initial conditions (ICs) for deterministic limited-area, high-resolution, convection-allowing numerical weather prediction (NWP) models are often provided by down-scaled analyses from external, larger-domain, coarser-resolution NWP models (e.g., Done et al. 2004; Kain et al. 2006, 2008; Weisman et al. 2008). Similarly, convection-allowing ensemble forecast initialization has typically relied on external models in some capacity (e.g., Hohenegger et al. 2008; Kong et al. 2008, 2009; Gebhardt et al. 2011; Hanley et al. 2011, 2013; Vié et al. 2011; Peralta et al. 2012; Duc et al. 2013; Jirak et al. 2014). This reliance on external models has been successful, as convection-allowing NWP models typically produce better precipitation forecasts than those with parameterized convection (e.g., Done et al. 2004; Kain et al. 2006; Lean et al. 2008; Roberts and Lean 2008; Weisman et al. 2008; Clark et al. 2009, 2010; Schwartz et al. 2009; Duc et al. 2013) and can provide valuable guidance regarding severe weather forecasting (e.g., Kain et al. 2008, 2010, 2013; Sobash et al. 2011; Clark et al. 2012, 2013; Weisman et al. 2013).

Alternatively, limited-area convection-allowing NWP model ICs can be produced via continuously cycling data assimilation (DA), where the NWP model state is periodically updated by assimilating real observations. Thus, continuous cycling can yield a limited-area modeling system where the role of external models is relegated to solely providing lateral boundary conditions (LBCs), and LBC impacts can largely be
reduced by placing the meteorological region of interest sufficiently far from the model’s lateral boundaries (e.g., Warner et al. 1997; Romine et al. 2014). In principle, any DA method can be employed in a continuously cycling DA system. However, for convective applications, ensemble Kalman filters (EnKFs; Evensen 1994; Burgers et al. 1998; Houtekamer and Mitchell 1998) have become popular because they incorporate time-evolving forecast error statistics and produce analysis ensembles that can initialize ensemble forecasts, which are vital for convective-scale forecasting as a result of inherent small-scale unpredictability (e.g., Lorenz 1969). Furthermore, as EnKFs permit a seamless integration of ensemble DA and forecasting, they obviate the need for ad hoc ensemble initialization approaches, such as those employing external models.

While many studies have focused on storm-scale EnKF DA of radar observations for individual cases [e.g., Snyder and Zhang 2003; Zhang et al. 2004; Dowell et al. 2004; Putnam et al. (2014), and references therein], only more recently have mesoscale EnKF DA systems been employed to initialize convection-allowing ensemble forecasts over meso-α- to synoptic-scale regions. For example, several case studies (e.g., Jones and Stensrud 2012; Melhauser and Zhang 2012; Jones et al. 2013, 2015; Schumacher and Clark 2014) and a few works spanning month-long periods (Romine et al. 2014; Schwartz et al. 2014) have used EnKF analysis/forecast systems with 15–36-km horizontal grid spacing to initialize convection-allowing ensemble forecasts by downscaling EnKF analysis ensembles onto computational domains with 3- or 4-km horizontal grid lengths. Collectively, these studies demonstrated that meso-scale EnKFs can initialize skillful convection-allowing ensemble forecasts, and, in fact, Schumacher and Clark (2014) found that EnKF-initialized convection-allowing ensemble rainfall forecasts were less biased and sometimes more skillful than those produced when convection-allowing ensemble ICs were derived from external models.

Given these encouraging results, in May and June 2013, in support of the Mesoscale Predictability Experiment (MPEX) field program (Weisman et al. 2015), the National Center for Atmospheric Research (NCAR) produced 48-h real-time ensemble forecasts with 3-km horizontal grid spacing that were initialized from analysis members of a continuously cycling 15-km EnKF DA system. These 3-km ensemble forecasts were initialized twice daily at 0000 and 1200 UTC, employed the Advanced Research version of the Weather Research and Forecasting (WRF) Model (Skamarock et al. 2008), and covered much of the conterminous United States (CONUS). We believe these modeling efforts marked the first time a continuously cycling EnKF was used to initialize real-time convection-allowing ensemble forecasts. This paper documents multiple aspects of the real-time 3-km ensemble forecasts. Specifically, we quantitatively verify ensemble forecasts of precipitation and severe weather “surrogates” (e.g., Sobash et al. 2011). We also present several forecast examples to subjectively illustrate successes and failures. Moreover, we noticed very different rainfall bias characteristics depending on whether forecasts were initialized at 0000 or 1200 UTC, and we investigate the cause of these behaviors.

Section 2 describes the forecast model configurations and initialization approach. Forecast examples are presented in section 3 and quantitative verification of precipitation and simulated severe weather forecasts are presented in sections 4 and 5, respectively. Section 6 explores the reasons for different precipitation bias characteristics regarding the 0000 and 1200 UTC forecasts before we conclude in section 7.

2. Methodology

a. Model configurations

The real-time forecast model configurations largely followed those of Romine et al. (2014) and Schwartz et al. (2014). Specifically, all weather forecasts were produced by version 3.3.1 of the nonhydrostatic WRF Model, over a nested computational domain spanning the CONUS and adjacent areas (Fig. 1). The horizontal grid spacing was 15 km (415 × 325 grid points) in the outer domain and 3 km (1046 × 871 grid points) in the nest. Both domains had 40 vertical levels and a 50-hPa top. The time step was 75 s in the 15-km domain and 18.75 s in the 3-km nest. When the 3-km nest was included in WRF Model forecasts, two-way feedback linked the domains. Table 1 lists the WRF Model physical parameterizations that were shared by all ensemble members. The 15- and 3-km domains used the same physics schemes, as well as positive-definite moisture advection (Skamarock and Weisman 2009), except that cumulus parameterization was turned off in the 3-km domain.

b. Generation of ensembles and experimental design

A 50-member ensemble adjustment Kalman filter (EAKF; Anderson 2001, 2003) from the Data Assimilation Research Testbed (DART; Anderson et al. 2009) software was used to initialize ensemble forecasts. In summary, the EAKF system was run at 15-km horizontal grid spacing, assimilated many surface and upper-air
observations, employed vertical and horizontal localization to reduce spurious correlations due to sampling errors, and used prior (before assimilation) adaptive inflation (Anderson 2009) to preserve ensemble spread. Table 2 notes some EAKF settings and lists the observations that were assimilated. More details regarding the EAKF can be found in Romine et al. (2013, 2014).

The initial 15-km ensemble was produced by taking Gaussian random draws with zero mean and covariances from global background error covariances provided by the WRF DA (Barker et al. 2012) system and adding them to the 1800 UTC 30 April 2013 Global Forecast System (GFS) analysis interpolated onto the 15-km WRF Model domain (Barker 2005; Torn et al. 2006). Perturbed LBCs for the EnKF system were produced similarly. The randomly produced ensemble at 1800 UTC 30 April served as ICs for 6-h WRF Model forecasts, and the ensemble valid at 0000 UTC 1 May was the prior ensemble for the first analysis. The 0000 UTC 1 May analysis ensemble served as ICs for 6-h WRF Model forecasts, and the second EAKF analysis occurred at 0600 UTC 1 May. This cyclic analysis/forecast pattern with a 6-h period continued until 1200 UTC 16 June.

The EnKF DA system and associated WRF Model forecasts were run solely over the 15-km domain. From 15 May to 15 June (32 days total), each 0000 and 1200 UTC EnKF analysis initialized an ensemble of 48-h

![FIG. 1. Computational domain. Objective verification only occurred in the speckled region of the 3-km domain.](image)

<table>
<thead>
<tr>
<th>Physical parameterization</th>
<th>WRF Model option</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microphysics</td>
<td>Thompson</td>
<td>Thompson et al. (2008)</td>
</tr>
<tr>
<td>Long- and shortwave radiation</td>
<td>Rapid Radiative Transfer Model for Global Climate Models (RRTMG) with ozone and aerosol climatologies</td>
<td>Mlawer et al. (1997); Iacono et al. (2008); Tegen et al. (1997)</td>
</tr>
<tr>
<td>Land surface model</td>
<td>Noah</td>
<td>Chen and Dudhia (2001)</td>
</tr>
<tr>
<td>Cumulus parameterization</td>
<td>Tiedtke</td>
<td>Tiedtke (1989); Zhang et al. (2011)</td>
</tr>
</tbody>
</table>
WRF Model forecasts containing the nested 3-km domain\(^1\) (Fig. 1). The 3-km ensembles were initialized by interpolating 15-km EnKF analyses onto the 3-km grid. In real time, EnKF analyses initialized either 10- or 30-member 3-km ensemble forecasts. The choice to produce 10- or 30-member ensemble forecasts was guided by the prospects for MPEX field program operations, with 30-member forecasts usually produced when operations were deemed likely. However, after MPEX concluded, 30-member 3-km ensemble forecasts were retrospectively produced for those cases where only 10-member ensemble forecasts were generated in real time, and all analyses herein investigate 30-member ensembles. Schwartz et al. (2014) showed that 50-member ensemble forecasts of precipitation had comparable reliability, resolution, and skill as forecasts initialized from randomly chosen 30-member subensembles that were fairly insensitive to which 30 members (of 50) were selected. Thus, we believe our 30-member forecasts well represented the 50-member ensemble DA system. The 48-h ensemble forecasts employed perturbed LBCs produced identically as those described above for the cycling EnKF DA system and in Romine et al. (2014).

Additionally, mainly to determine if certain ensemble forecast characteristics were attributable to EnKF ICs or were symptomatic of WRF Model biases, 0000 and 1200 UTC operational 0.5\(^\circ\) GFS analyses were interpolated onto the nested computational domain and used to initialize deterministic 48-h WRF Model forecasts between 15 May and 15 June. During this period, the GFS model was initialized with a continuously cycling “hybrid” variational–ensemble DA system that blended flow-dependent and static background error covariances (e.g., Wang et al. 2013) and assimilated many observations not assimilated by the limited-area EnKF, such as satellite radiances. Furthermore, standard WRF Model preprocessing does not incorporate GFS hydrometeor fields into WRF Model ICs.\(^2\) Thus, GFS-initialized WRF Model forecasts did not contain initial hydrometeors, whereas the continuously cycled EnKF ICs contained nonzero hydrometeor fields.

Next, we present a subjective evaluation of selected forecasts.

3. Ensemble forecast examples

Varied weather regimes during the MPEX period, ranging from weakly forced convective events to strongly forced severe weather outbreaks, offered many opportunities to test the performance of the 3-km ensemble. Several cases are discussed that reveal both successes and failures. The 15 and 31 May 2013 events detailed in sections 3a and 3b were chosen because of their major societal impacts, while section 3c summarizes several additional cases representing forecasts across a broad spectrum of storm modes and geographical regions.

We focus our subjective evaluation on forecasts of simulated reflectivity and probabilities of hourly maximum 2–5-km updraft helicity\(^3\) (UH; Kain et al. 2008, 2010) accumulated between 1800 and 0600 UTC (12 h), although 6-h probabilities accumulated between 2100 and 0300 UTC yielded similar conclusions. UH has been documented as a useful diagnostic for assessing the intensity of simulated

---

\(^1\) At 0000 and 1200 UTC, two sets of ensemble forecasts were produced. The first set was for DA purposes and consisted of a 6-h 50-member ensemble forecast with the 3-km nest removed to advance the ensemble to the next analysis time. The second set was the ensemble of 48-h forecasts with the 3-km nest included. Thus, although there was two-way feedback during the 48-h forecasts (section 2a), 3-km forecasts did not impact the cycling EnKF DA system.

\(^2\) Hydrometeor partitioning in the WRF Model is a function of microphysics and vertical velocity, which is a function of horizontal grid length. Thus, unless the WRF Model configuration has the same resolution and microphysics as the model providing the ICs (here, the GFS), it is preferable that the WRF Model produce its own hydrometeors based on the relative humidity ICs (J. Bresch 2015, personal communication). It is standard practice that WRF Model forecasts begin from zero hydrometeor fields when an external model provides the ICs.

\(^3\) Updraft helicity is the integral of the product of vertical velocity and vertical vorticity between two layers.
convection in NWP model output (e.g., Kain et al. 2008; Clark et al. 2013) and is used as a surrogate for the prediction of subgrid-scale severe weather hazards (e.g., hail, damaging wind, and tornadoes).

a. 15 May 2013

The 500-hPa synoptic pattern at 1200 UTC 15 May was characterized by a weak cutoff low over the Texas Panhandle (Fig. 2a) with a moderately unstable air mass over most of central and eastern Texas and southern Oklahoma (not shown). Satellite water vapor at 1200 UTC 15 May (Fig. 2b) showed enhanced moisture and cloudiness from north-central Texas through central Oklahoma, spreading westward into the Texas Panhandle, north and west of the cutoff low. A dry slot was evident to the northwest of the cutoff low, with additional dry air wrapping around the southeastern side of the upper low. Further, a surface cold front was located across northern Kansas.

Convection initiated over south-central Oklahoma by 1700 UTC and slowly intensified into a well-organized asymmetric mesoscale convective system (MCS) over south-central Oklahoma by 2200 UTC, including a bow-shaped leading line with a cyclonic mesoscale vortex at its northern end. Isolated supercells developed south of the MCS in northern Texas by 0000 UTC 16 May (Fig. 3a) and were responsible for several strong tornadoes near Granbury, Texas.

A snapshot of the 12-h ensemble column-maximum reflectivity forecast valid at 0000 UTC 16 May (Figs. 3b–i) suggested convection over central and northeastern Texas, with some members indicating a dominant bow-shaped organization (Figs. 3b,g) while other members predicted supercell-like structures (Figs. 3d,i). However, lesser activity was forecast northward into Oklahoma, where the observed MCS was located, and spurious convection was forecast in eastern Kansas along the cold front (convection did develop along the cold front after 0000 UTC 16 May, but it was farther north and more isolated than predicted by most ensemble members). Nonetheless, higher probabilities of reflectivity >45 dBZ (Fig. 4a) and updraft helicity >75 and 150 m² s⁻² (Fig. 4b) were generally collocated with those areas that experienced severe weather. However, the ensemble missed some severe weather events in southeastern Oklahoma and produced false alarms in southwestern Texas (Fig. 4b).

b. 31 May 2013

At 1200 UTC 31 May (Fig. 5), there was a broad 500-hPa trough over the central plains with a deep, occluded low over the northern plains. The jet stream and associated mid- to upper-tropospheric front extended from Nevada through Colorado and eastward through Kansas and Missouri. A surface low pressure center was located in South Dakota, and a trailing cold front extended through the southern plains. An additional weak surface low pressure center was located in northwestern Oklahoma, from which a dryline extended southward. Ahead of the cold front and dryline, the air mass was very unstable and had strong vertical wind shear. Water vapor imagery (Fig. 5b) showed
enhanced moisture over Colorado, suggestive of an upper-level disturbance.

Convection initiated around 2000 UTC in southeastern Kansas and subsequently back-built southwestward to central Oklahoma by 2200 UTC. The southern end of this line remained just west of Oklahoma City, Oklahoma, for several hours, wherein a significantly tornadic supercell storm was spawned near El Reno, Oklahoma, around 2300 UTC near the cold front–dryline intersection. The observed radar reflectivity at 0000 UTC 1 June (Fig. 6a) depicts the El Reno storm at the southern end of the line, with the line extending northeastward through Missouri.

Nearly all ensemble members forecast convection at 0000 UTC 1 June from central Oklahoma through southwestern Missouri and Illinois (Figs. 6b–i and 7a). The ensemble also predicted convection south of Oklahoma City, and at 0000 UTC 1 June, a small storm was present in south-central Oklahoma (Fig. 6a). UH probabilities (Fig. 7b) correctly highlighted central Oklahoma for supercell activity, but several storms were missed in eastern Oklahoma.
c. Other cases

Figure 8 shows probabilistic forecasts of UH exceeding 75 and 150 m\(^2\) s\(^{-2}\) for four additional cases selected because of their geographical and storm mode diversity. While there were sometimes phase errors (e.g., Fig. 8b in northwestern Arkansas), false alarms (e.g., Fig. 8b in northwestern Oklahoma, Fig. 8c in central Kansas and northeastern Wyoming, and Fig. 8d in north-central Illinois), and missed events (e.g., Fig. 8a in western

![Figures](https://example.com/figs)
Illinois, Fig. 8b in Iowa, and Fig. 8c in South Dakota), at many locations, areas of relatively high ensemble UH probabilities corresponded with those areas that experienced severe weather.

The next two sections present objective verification that confirms subjective impressions of a skillful, but imperfect, ensemble.

4. Objective precipitation verification

The 3-km precipitation forecasts were verified using both deterministic and probabilistic techniques. Gridded stage IV (ST4) data (Lin and Mitchell 2005) from the National Centers for Environmental Prediction (NCEP) were used as “truth.” Objective verification was performed over a fixed domain encompassing most of the central CONUS (speckled region in Fig. 1), distant from lateral boundaries. We first describe general precipitation characteristics before assessing aspects related to forecast skill.

a. General precipitation characteristics

Total accumulated precipitation over the verification domain, calculated on the native grids, aggregated hourly over all forecasts, and normalized by 32 times the total number of grid points over the verification domain,
is shown in Fig. 9. After the spinup period, at most times, the models overpredicted rainfall compared to ST4 analyses, commensurate with many studies reporting high rainfall biases with convection-allowing models over the CONUS (e.g., Clark et al. 2010; Schwartz et al. 2010, 2014; Johnson et al. 2013; Romine et al. 2013). GFS-initialized forecasts were slower to spin up rainfall compared to EnKF-initialized ensemble forecasts because the GFS-initialized WRF Model forecasts lacked initial hydrometeors (section 2b), but 0000 (1200) UTC GFS-initialized forecasts had values closer to ST4 observations during the first (second) diurnal peak than did ensemble members initialized at the same time.

However, the most noteworthy differences between the ensemble and GFS-initialized forecasts and forecasts initialized at 0000 and 1200 UTC involved the first diurnal cycle. Around the first diurnal peak, 0000 UTC GFS-initialized forecasts overpredicted the mean total rainfall by ~12% compared to ST4 observations, but 1200 UTC GFS-initialized forecasts overpredicted total rainfall by ~45% during this same period. A similar pattern was noted with the 0000 and 1200 UTC ensemble forecasts, except differences were smaller than those between the corresponding GFS-initialized forecasts. During the second diurnal maximum, forecasts initialized at 1200 UTC produced less average rainfall compared to the first diurnal peak that was at most ~21% higher than ST4 values.

Figure 10 shows the fractional occurrence of various events, defined as the precipitation exceedance of selected accumulation thresholds $q$ (e.g., $q = 1.0 \text{ mm h}^{-1}$) over the verification domain, computed on native grids, and aggregated hourly over all forecasts. Areal coverages were broadly consistent with total precipitation (Fig. 9) and revealed that at all thresholds, forecasts initialized at 1200 UTC produced substantially greater coverages during the first diurnal peak than forecasts initialized 12 h earlier. Again, differences between forecasts initialized at 0000 and 1200 UTC were largest for GFS-initialized forecasts, with particularly pronounced differences at the 5.0 and 10.0 mm h$^{-1}$ thresholds (Figs. 10c,d). The biases were largest for higher thresholds, similar to previous findings over the CONUS (e.g., Weisman et al. 2008; Schwartz et al. 2009, 2010; Clark et al. 2010; Schwartz and Liu 2014). We further examine the excessive precipitation during the first diurnal peak in section 6.

Finally, we note that spatial patterns of total forecast rainfall over the entire experimental period agreed well with corresponding ST4 observations (not shown).
b. Objective probabilistic verification

Areas under the relative operating characteristic (ROC) curve and fractions skill scores (FSSs; Roberts and Lean 2008) were calculated for a range of precipitation thresholds. These metrics require that the observations and model be on a common grid, so the raw 3-km rainfall forecasts were interpolated onto the ST4 grid (\(\sim 4.7\)-km grid spacing) using a budget interpolation method (e.g., Accadia et al. 2003). FSSs are commonly used to verify high-resolution forecasts and the ROC area is regularly employed to verify ensemble forecasts (Roberts and Lean 2008; Clark et al. 2010; Schwartz et al. 2010, 2014; Duc et al. 2013; Romine et al. 2013, 2014; Schumacher and Clark 2014).

We wish to compare the relative quality of 0000 and 1200 UTC forecasts for fixed valid times, but the different precipitation bias characteristics of the 0000 and 1200 UTC forecasts (i.e., section 4a) potentially muddled this comparison. Thus, precipitation biases were effectively removed using a probability matching approach described by Clark et al. (2010). Specifically, the ST4 rainfall amounts were ranked from highest to lowest. Then, separately for each ensemble member and the deterministic GFS-initialized forecasts, the grid point containing the highest precipitation amount was replaced with the highest ST4 value, the grid point with the second-highest precipitation amount was replaced with the second-highest ST4 value, and so on. At

FIG. 8. Probability of UH exceeding 75 (\%; color fill) and 150 m² s⁻² (\%; red contours, with a 10\% contour interval beginning at 10\%) within \(\sim 40\) km of each grid point accumulated for 6–18-h forecasts initialized at 1200 UTC (a) 19 May, (b) 20 May, (c) 27 May, and (d) 12 Jun. If the UH event occurred at any grid point within \(\sim 40\) km of the \(i\)th grid point between the 6th and 18th forecast hours for any ensemble member, the probability at \(i\) was nonzero. SPC storm reports over the corresponding 12-h periods are overlaid in yellow symbols.
grid points where the model had values of zero precipitation, the bias-corrected field was also forced to zero to preserve precipitation area. All ROC and FSS precipitation statistics were computed using these bias-corrected rainfall fields.

Moreover, as high-resolution NWP models are not skillful at the grid scale, a neighborhood approach was employed to postprocess rainfall fields before performing verification. Specifically, for the ensemble forecasts, the neighborhood ensemble probability (NEP; Schwartz et al. 2010, 2014; Johnson and Wang 2012; Duc et al. 2013) was computed by averaging the point-based ensemble mean (EP) over a radius of influence $r$ (km). The EP at the $i$th grid point (EP$_i$) is simply the number of ensemble members with precipitation $\geq q$ at grid point $i$ divided by the ensemble size (30), and the NEP at the $i$th grid point (NEP$_i$) is found by averaging EP$_i$ over the $N_b$ grid boxes within the $r$ (km) neighborhood of $i$. ROC areas and FSSs were computed for values of $r = 5, 25, 50, 75, 100, 125,$ and $150$ km.

We also computed ROC areas and FSSs for GFS-initialized forecasts using a neighborhood approach to first transform the deterministic precipitation forecasts into fractional fields by counting the number of points within $r$ (km) of the $i$th grid point containing accumulated precipitation $\geq q$ and dividing by $N_b$ (e.g., Theis et al. 2005; Roberts and Lean 2008). These forecast fractions can be interpreted exactly as the EP fields described above and were used in ROC and FSS calculations. Initially, it may seem inappropriate to compare NEPs with probabilistic fields derived from deterministic forecasts, because in the latter, a neighborhood approach was used to create probabilities, while in the former, a neighborhood approach effectively smoothed an already-present probabilistic field. However, Schwartz et al. (2010) noted the NEP can also be calculated by transforming each individual ensemble member’s deterministic forecast into a fractional field and then averaging the fractional fields over all ensemble members. Viewed this way, the only additional step for computing NEPs is averaging, which is intrinsic to ensemble prediction systems. Thus, we see little problem comparing FSSs and ROC areas derived from initially deterministic and probabilistic forecasts.

The NEP should not be confused with the traditional point-based ensemble mean. Rather, the NEP is the mean of individual members’ fractional fields and can itself be interpreted as a probabilistic forecast. As different members’ errors tend to cancel in the averaging process that yields the NEP (Leith 1974), NEP fields are more skillful than individual ensemble members’ forecasts (e.g., Schwartz et al. 2010, 2014; Duc et al. 2013). Therefore, to focus on the highest-quality ensemble guidance and to encourage examination of the ensemble as a whole (as summarized by the NEP), we do not examine skill of individual members’ forecasts.

Statistical significance for FSSs and ROC areas was determined by a bootstrap technique applied to differences between pairs of experiments (e.g., Hamill 1999; Schwartz and Liu 2014; Wolff et al. 2014). For each forecast hour, paired resamples of daily statistics were randomly drawn from all forecast cases, and aggregate FSS and ROC area differences between two experiments were calculated. This procedure was repeated 1000 times to estimate bounds of 90% confidence intervals (CIs). If zero was not included within bounds of the 90% CIs, then differences were statistically significant at the 95% level or higher.

1) ROC AREAS

To produce the ROC, probabilistic forecast thresholds $p$ of $0\%$, $5\%$, $15\%$, $25\%$, ..., $85\%$, $95\%$, $100\%$ were chosen. Then, a series of $2 \times 2$ contingency tables (Table 3) was populated for each $p$ for different precipitation thresholds. Denoting ST4 precipitation at grid point $i$ as $O_i$ and the NEP of precipitation $\geq q$ at grid point $i$ as NEP$_{i,q}$, the $i$th grid point fell into category $a$ if NEP$_{i,q} \geq p$ and $O_i \geq q$, $b$ if NEP$_{i,q} \geq p$ and $O_i < q$, $c$ if NEP$_{i,q} < p$ and $O_i \geq q$, and $d$ if NEP$_{i,q} < p$ and $O_i < q$. 

![Figure 9](https://example.com/figure9.png)

**FIG. 9.** Total accumulated precipitation over the verification domain aggregated hourly over all 32 forecasts, computed on native grids, and normalized by 32 times the number of grid points over the verification domain.
Using the elements in Table 3, the probability of detection \( \text{POD} = \frac{a}{a + c} \) and probability of false detection \( \text{POFD} = \frac{b}{b + d} \) were computed for each probability threshold, and the ROC was formed by plotting the POFD against the POD over the range of \( \rho \). The area under this curve (the ROC area) was used to summarize the ROC and computed using a trapezoidal approximation. ROC areas \( >0.5 \) indicate successful discriminating ability.

ROC areas computed with \( r = 50 \text{ km} \) were \( >0.5 \) for all thresholds and times (Fig. 11), indicating the ensemble and GFS-initialized forecasts had skillful discriminating ability. Even though the ensemble forecasts initialized at 1200 UTC had poor precipitation biases during the first diurnal maximum, at most times, ensemble forecasts initialized at 1200 UTC had higher ROC areas than ensemble forecasts initialized 12 h earlier at 0000 UTC for fixed valid times. Similarly, after the spinup period,
1200 UTC GFS-initialized forecasts had higher ROC areas than forecasts initialized from GFS analyses 12 h earlier. Differences between the 0000 and 1200 UTC forecasts were often statistically significant at the 95% level, particularly for lighter rainfall thresholds. Ensemble forecasts consistently had higher ROC areas than GFS-initialized forecasts, although there was some ambiguity for $q = 10.0 \text{ mm h}^{-1}$.

Findings were generally similar using different $r$, except that differences between GFS-initialized and ensemble forecasts were smaller as $r$ increased to 100 km and above (not shown), which suggests GFS-initialized forecasts had larger small-scale errors than EnKF-initialized forecasts, perhaps due to its coarser ICs. Further, ROC areas were $>0.5$ for $r < 50$ km, including $r = 0$ km (i.e., the grid scale), for all forecasts, times, and thresholds.

2) FRACTIONS SKILL SCORES

The FSS measures spatial skill and can be used to verify any probabilistic (fractional) field. The FSS requires transforming the observations (i.e., ST4 analyses) into “observed fractions,” which is achieved in the same manner as producing forecast fractions from deterministic model fields. Forecast and observed fractions are then directly compared to produce the FSS (Roberts and Lean 2008). The FSS ranges from 0 to 1, with a perfect forecast attaining $\text{FSS} = 1$ and $\text{FSS} = 0$ indicating no skill.

Consistent with ROC areas, for fixed valid times, forecasts initialized at 1200 UTC usually had higher aggregate FSSs for all $r$ than forecasts initialized 12 h earlier (shown for $r = 50$ km in Fig. 12). Moreover, for a fixed initialization time, ensemble forecasts were typically more skillful than deterministic GFS-initialized forecasts. Again, for $r = 100$ km and larger, there were smaller differences between the GFS-initialized and ensemble forecasts, particularly for forecasts initialized at 1200 UTC (not shown).

3) DISCUSSION

For a fixed valid time, forecasts initialized at 1200 UTC produced better precipitation forecasts than those initialized 12 h earlier. These results agree with preliminary findings from the 2013 NOAA Hazardous Weather Testbed Spring Experiment, where similar conclusions were drawn from ad hoc ensembles essentially composed of various operational or experimental deterministic convection-allowing models (Jirak et al. 2014). Usually, probabilistic ensemble forecasts were more skillful than deterministic GFS-initialized forecasts. We note that the ensemble precipitation forecasts had insufficient spread, similar to Schwartz et al. (2014) and Romine et al. (2014), and forecasts of precipitation initialized at 1200 UTC had comparable reliability (i.e., calibration) to those initialized 12 h earlier (not shown).

Forecasts initialized at 0000 UTC were also compared to those initialized 12 h earlier at 1200 UTC. For the ensemble, results were as before: more recently initialized 0000 UTC forecasts were clearly better than previous 1200 UTC forecasts. However, for GFS-initialized forecasts, differences between 0000 UTC and previously initialized 1200 UTC forecasts were smaller, and sometimes, particularly for $q \geq 5.0 \text{ mm h}^{-1}$, older forecasts were better than newer ones, even after the spinup (not shown). Newer 0000 UTC GFS-initialized forecasts may have been disadvantaged compared to previous 1200 UTC forecasts, because its long spinup meant impacts from ongoing precipitation between ~0000 and 0600 UTC (such as locations of outflow boundaries) may not have been properly represented.

The next section applies similar objective verification approaches to evaluate surrogate severe weather guidance.

5. Objective verification of surrogate severe weather guidance

Forecasts of hourly maximum UH (see section 3) were verified following Sobash et al. (2011). As in Sobash et al. (2011), UH verification was restricted to grid points where both vertical velocity and vorticity were $>0$, which is meant to correspond to cyclonically rotating updrafts.

Verification of UH forecasts occurred on an ~80-km grid, which was chosen to approximately match the scale of Storm Prediction Center (SPC) probabilistic severe weather hazards forecasts (Sobash et al. 2011) and reduce documented biases with observed severe weather reports (Weiss and Vescio 1998). Quality-controlled observed reports of hail with $\geq 2.54 \text{ cm}$ diameter, wind gusts of $\geq 50$ knots ($1 \text{ kt} = 0.51 \text{ m s}^{-1}$), and tornadoes published by the SPC (retrieved from http://www.spc.noaa.gov/wcm/) were mapped to this ~80-km grid, and an 80-km grid box was flagged as containing an observed storm if its bounds contained any storm report. Similarly, model output of hourly maximum UH was superimposed on the same ~80-km grid. Various UH

<table>
<thead>
<tr>
<th>Forecast</th>
<th>Yes</th>
<th>No</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>$a$</td>
<td>$b$</td>
<td>$a+b$</td>
<td>$a+d$</td>
</tr>
<tr>
<td>$a+c$</td>
<td>$b+d$</td>
<td>$a+c$</td>
<td>$b+d$</td>
<td>$a+b$</td>
</tr>
</tbody>
</table>
thresholds $q_{UH}$ were chosen to define forecast events, and if any 3-km model grid point collocated with a particular 80-km grid box contained $UH \geq q_{UH}$, the 80-km box was flagged as containing a forecast event.

Probabilistic (fractional) UH forecasts were derived using kernel-density estimation with a Gaussian smoother, as described in Sobash et al. (2011) and Hitchens et al. (2013). Let $B_i$ denote the $i$th grid point on the ~80-km verifying grid, where $B_i = 1$ if $UH \geq q_{UH}$ at the $i$th grid point and $B_i = 0$ otherwise. Then, the forecast fraction at the $i$th grid point $f_i$ for UH was

$$f_i = \sum_{n=1}^{N_y} \frac{1}{2\pi \sigma^2} \exp \left[ -\frac{1}{2} \frac{(d_{i-n})^2}{\sigma^2} \right] B_n,$$

where $d_{i-n}$ is the physical distance (i.e., km) from grid-point $i$ to the $n$th of $N_y$ grid points within the verification domain and $\sigma$ (km) controls the spatially isotropic smoothing kernel’s weights and is analogous to $r$. Larger

---

**Fig. 11.** Area under the ROC curve as a function of forecast hour for (a) 0.25, (b) 1.0, (c) 5.0, and (d) 10.0 mm h$^{-1}$ precipitation thresholds using a 50-km radius of influence and aggregated over all 32 forecasts. Horizontal lines are "zero lines" about which 90% bootstrap CIs based on pairwise differences between forecasts initialized at 0000 and 1200 UTC are shown (1200 minus 0000 UTC). The red CIs are for the GFS-initialized forecasts and blue CIs are for EnKF-initialized forecasts. If the CI does not contain zero, then the differences were statistically significant at the 95% level or higher.
values of $\sigma$ represent larger smoothing distances, and fractions were computed for $\sigma$ ranging from 60 to 200 km (inclusive) in intervals of 20 km. Strictly, although $f_i$ is a function of all points within the verification domain, points where $d_{i,n} > 4\sigma$ contribute very little to $f_i$ and points where $B_n = 0$ do not contribute to $f_i$.

In Sobash et al. (2011), fractional UH fields were computed for 24-h forecast periods, but here, they were computed hourly for both the simulated and observed reports.

Within the ensemble context, fractional UH fields were computed separately for each member using Eq. (1), and the NEP for UH was found by averaging over all members’ $f_i$ fields. Hourly NEP UH forecasts were compared to hourly fractional observed storm report fields to compute FSSs. Bootstrap CIs based on differences between two distributions were computed identically as with precipitation verification.

The method of producing fractional observed and simulated severe weather reports, initially proposed by Sobash et al. (2011), differs from the creation of fractional precipitation forecasts (section 4b) owing to the application of a different weighting kernel in the smoothing procedure. Using the Gaussian kernel permits a distance-dependent weighting function where neighboring points closer to the $i$th grid point are weighted more than points farther from $i$, whereas the neighborhood approach described in section 4b employs a uniform weighting kernel, where all points within the neighborhood of $i$ are weighted equally. Both

Fig. 12. As in Fig. 11, but for aggregate FSSs computed with a 50-km radius of influence.
methods achieve the same end result of producing probabilistic fields from deterministic forecasts, and testing revealed that the Gaussian and uniform weight kernels yielded very similar probabilistic fields when analogous values of $\sigma$ and $r$ were used, such that our results were insensitive to the kernel. While future work may attempt to unify the application of neighborhood methods to precipitation and UH, here, our respective approaches of creating fractional UH and precipitation fields follow previously used techniques for each observation analysis type.

There is uncertainty regarding what model UH values provide a meaningful correspondence to observed storms. Further, simulated UH magnitudes are sensitive to horizontal grid length (e.g., Adlerman and Droegemeier 2002; Kain et al. 2008). Therefore, a range of $q_{\text{UH}}$ values was examined. Clearly, as $q_{\text{UH}}$ increased, there were fewer simulated severe reports over the verification domain (Fig. 13). Similar to precipitation, at all UH thresholds during the first diurnal maximum, 1200 UTC forecasts usually had more UH events than forecasts initialized 12 h earlier, which was most pronounced at the 25 m$^2$s$^{-2}$ threshold for GFS-initialized forecasts (Figs. 13c,d). Thus, not only did 1200 UTC forecasts produce more rain during the first diurnal peak, there were also more storms with high UH. While 0000 UTC forecasts produced approximately the same number of simulated storms during both diurnal peaks (Figs. 13a,c), 1200 UTC forecasts produced more simulated reports during the first diurnal maximum than the second (Figs. 13b,d), which reflects the overproduction of precipitation during the first diurnal peak.

The threshold where the number of simulated reports agreed most closely with the observed storm report frequency varied temporally but typically was between 75 and 100 m$^2$s$^{-2}$ (Fig. 13). Thus, since UH fields cannot be debiased analogously to the precipitation fields (section 4b), we focus on FSSs computed for $q_{\text{UH}} = 75$ m$^2$s$^{-2}$, which often had the smallest bias, although FSSs for $q_{\text{UH}} = 100$ m$^2$s$^{-2}$ were qualitatively similar.
FSSs exhibited a strong diurnal signal that was tied to the climatological UH distribution (Fig. 14). At $\sigma = 60$ and 200 km, for a fixed initialization time, ensemble forecasts usually had higher FSSs than GFS-initialized forecasts. In fact, during the first diurnal peak, 24-h ensemble forecasts initialized at 0000 UTC had similar or higher FSSs compared to 12-h GFS-initialized forecasts, indicating the ensemble forecasts provided the same or greater skill than deterministic GFS-initialized forecasts but with 12-h additional lead time.

During the first diurnal maximum, 1200 UTC forecasts typically had higher FSSs compared to previous 0000 UTC forecasts for smaller neighborhoods (e.g., $\sigma = 60$ km), with some instances of statistical significance at the 95% level (Fig. 14a). However, as $\sigma$ increased past 100 km, differences between 0000 and 1200 UTC FSSs during the first diurnal cycle gradually decreased (not shown), such that by $\sigma = 200$ km, 0000 and 1200 UTC forecasts had similar FSSs during the first diurnal maximum (Fig. 14b). These findings indicate that differences between 0000 and 1200 UTC forecasts during the first diurnal maximum were greatest at smaller scales. Overall, UH FSS patterns were similar to those for precipitation and indicate that the benefits of more recently initialized 1200 UTC forecasts extend to severe weather guidance.

6. Investigating excessive precipitation during the first diurnal maximum

In section 4a, we showed that forecasts initialized at 1200 UTC produced more rainfall during the first diurnal maximum than forecasts initialized 12 h earlier. To better understand these discrepancies, we further examined thermodynamic differences between the 0000 and 1200 UTC forecast sets and performed some sensitivity experiments. While we examined aspects of both the 0000 and 1200 UTC GFS- and EnKF-initialized forecasts, we focused slightly more on the GFS-initialized forecasts because their differences during the first diurnal peak were most pronounced.

a. Thermodynamic differences between 0000 and 1200 UTC GFS-initialized forecasts

The first few hours of GFS-initialized WRF Model forecasts featured little precipitation (e.g., Fig. 9) and cloud cover. Thus, 1200 UTC GFS-initialized WRF Model forecasts had more initial solar insolation than 12-h forecasts initialized from 0000 UTC GFS analyses (which had ongoing clouds and precipitation at 1200 UTC). This difference was manifested by increased relative outgoing longwave radiation (OLR) between ~1300 and 1800 UTC (Fig. 15a), and increased surface heating (Fig. 15b) on average over the verification domain. Accordingly, 1200 UTC GFS-initialized forecasts had greater mean latent heat flux (LHFX; Fig. 16a) compared to previously initialized 0000 UTC forecasts between ~1300 and 2200 UTC on average over the verification region.

Also, because 1200 UTC GFS-initialized forecasts had no initial clouds, there was no mechanism for water vapor removal via precipitation at 1200 UTC, whereas the 0000 UTC forecasts were removing atmospheric
Figure 16b shows the mean difference of hourly total accumulated precipitation and hourly change in total column precipitable water (PW) between the 0000 and 1200 UTC GFS-initialized forecasts (1200 minus 0000 UTC) over the verification domain. Initially, because of spinup, 1200 UTC forecasts lagged 0000 UTC forecasts regarding precipitation production. However, by ~1900 UTC, 1200 and 0000 UTC forecasts produced similar hourly precipitation, and thereafter, 1200 UTC forecasts produced more hourly rainfall (Fig. 16b). Commensurately, PW accumulated faster in the 1200 UTC forecasts until differences in hourly precipitation became similar, after which, 1200 UTC forecasts lost PW compared to 0000 UTC forecasts, presumably because of greater moisture removal via precipitation (Fig. 16b). Regarding absolute values, PW increased more in the 1200 UTC forecasts after ~1300 UTC than in 0000 UTC forecasts (Fig. 15c). Spatially, at 1200 UTC, GFS analyses were moister from central Texas through the upper Midwest compared to 12-h WRF Model forecasts (Fig. 17a), and with time, forecasts initialized at 1200 UTC became progressively moister than forecasts initialized at 0000 UTC in most areas (Figs. 17b-d).

Furthermore, 12-h WRF Model forecasts and 1200 UTC GFS analyses had different water vapor distributions, with 1200 UTC GFS analyses containing substantially more near-surface moisture than 12-h WRF Model forecasts (Fig. 16c). As boundary layer mixing developed, relative dewpoint differences between the 0000 and 1200 UTC forecasts were maximized progressively later moving from 2 m, to 925, 850, and 700 hPa (Fig. 16c). The additional vertical moistening in 1200 UTC forecasts likely contributed to enhanced potential for and production of precipitation by lessening dry air entrainment. Compared to 0000 UTC forecasts, greater low-level warming (e.g., Fig. 15b) and boundary layer moistening in the 1200 UTC forecasts yielded relatively higher mixed-layer convective available potential energy (MLCAPE) from ~1300 to 0200 UTC (Fig. 15d).

b. Thermodynamic differences between 0000 and 1200 UTC EnKF-initialized forecasts

Unlike GFS-initialized forecasts, EnKF-initialitzed forecasts rapidly spun up precipitation (e.g., Fig. 9). Consequently, 1200 UTC EnKF-initialized forecasts had similar or lower OLR values than 0000 UTC.
ensemble forecasts valid at common times (Fig. 18a), contrasting the behavior of GFS-initialized forecasts (Fig. 15a). Surface temperature differences between 0000 and 1200 UTC EnKF-initialized forecasts were also smaller than those of GFS-initialized forecasts (cf. Figs. 15b and 18b).

However, 1200 UTC EnKF mean analyses had noticeably more PW and MLCAPE than 12-h ensemble forecasts, and the magnitudes of these initial differences generally persisted through the first diurnal cycle (Figs. 18c,d). Conversely, while 1200 UTC GFS analyses had similar PW and MLCAPE as 12-h GFS-initialized forecasts, differences developed and amplified with time (Figs. 15c,d and 16a), likely as a result of increased solar insolation as a result of a long spinup (section 6a). Thus, spinup characteristics may have strongly influenced the evolution of thermodynamic fields during the daytime heating hours.

c. Auxiliary experiments

Based on the above analysis, we hypothesized that three factors may have contributed to the excessive rainfall from 1200 UTC GFS-initialized WRF Model forecasts during the first diurnal peak: 1) enhanced LHFX, and therefore, moisture, due to lack of initial clouds and precipitation; 2) a near-surface moist pool that gradually led to boundary layer moistening; and 3) lack of water vapor removal via precipitation. However, it appeared that greater moisture and instability in 1200 UTC EnKF ICs may have been the sole contributor to enhanced 1200 UTC ensemble rainfall during the first diurnal peak. Thus, several auxiliary experiments were conducted to test some of our hypotheses.

First, we initialized new 48-h WRF Model forecasts from 1200 UTC GFS analyses, except we replaced the water vapor fields with those from 12-h deterministic WRF Model forecasts initialized from 0000 UTC GFS analyses. This procedure substantially reduced the excess precipitation during the first diurnal peak and improved precipitation spinup (cf. dashed red and solid orange lines in Fig. 19a), which indicates that 1200 UTC GFS moisture analyses prominently contributed to the large precipitation bias during the first diurnal peak. Furthermore, in addition to replacing GFS analysis water vapor fields with those from 12-h WRF Model forecasts, we also inserted 12-h WRF Model forecast hydrometeor fields into 1200 UTC GFS analyses and initialized another set of 48-h WRF Model forecasts. Performing this extra step further improved precipitation spinup and reduced precipitation during the first diurnal maxima (green line in Fig. 19a) by lessening the accumulation of MLCAPE during the first few hours of integration (not shown). Nonetheless, despite these improvements, a considerable precipitation overshoot compared to 0000 UTC GFS-initialized forecasts still occurred during the first diurnal peak.

To examine the ensemble forecasts, we also initialized new 1200 UTC WRF Model forecasts from EnKF member 1’s analyses, except we replaced the water vapor and hydrometeor fields with those from member 1’s 12-h forecasts initialized from previous 0000 UTC analyses. This new experiment produced less precipitation during the first diurnal cycle compared to member 1’s original 1200 UTC forecasts (cf. the solid green and dashed red lines in Fig. 19b). Therefore, it appears that excessive IC moisture caused the 1200 UTC ensemble

---

4 Although we stress the importance of examining the ensemble as a whole, it was computationally impossible to perform similar experiments for all members, and we only produced 24-h forecasts given limited resources.
forecasts to produce more rainfall than 0000 UTC ensemble forecasts during the first diurnal peak. It is beyond the scope of this paper to determine why 1200 UTC EnKF analyses were moister than 12-h EnKF-initialized forecasts, although this topic deserves attention.

We do not believe these behaviors have been previously documented for convection-allowing forecasts over the CONUS. It is possible that similar problems may exist in other regions where WRF Model initialization occurs within ~12 h of the diurnal precipitation maximum and we urge further analysis into this issue, particularly for configurations with longer precipitation spinup times.

7. Summary and conclusions

NCAR produced real-time 3-km ensemble forecasts in May and June 2013 in support of the MPEX field program. The ensemble forecasts were initialized by analysis members of a continuously cycling mesoscale EnKF DA system. Forecasts were initialized twice daily at 0000 and 1200 UTC for 32 days. Verification efforts focused on evaluating 30-member ensemble forecasts using a variety of methods, with a focus on precipitation, severe weather guidance, and assessing the relative benefit of forecasts initialized at 1200 UTC compared to those initialized at 0000 UTC. Deterministic 3-km WRF
Model forecasts initialized from GFS analyses were also examined.

The ensemble forecasts were often subjectively good, despite some errors regarding convective timing, mode, and placement. These subjective impressions were corroborated by objective statistics, such as ROC areas and FSSs, which indicated skill at precipitation and severe weather prediction. Ensemble forecasts initialized...
at 1200 UTC were better than those initialized 12 h earlier for fixed valid times, and for fixed initialization times, ensemble forecasts were usually more skillful in terms of precipitation and severe weather placement than GFS-initialized forecasts, particularly at smaller scales.

During the first diurnal precipitation maximum, WRF Model forecasts initialized at 1200 UTC produced substantially more precipitation than forecasts initialized 12 h earlier at 0000 UTC. The same behavior was noted for EnKF- and GFS-initialized forecasts, although the behavior was more pronounced for GFS-initialized forecasts and consistent with excessive low-level moisture and absence of initial hydrometeors in 1200 UTC GFS-initialized WRF Model forecasts. The 1200 UTC EnKF-initialized forecasts appeared to produce too much rainfall because their ICs were too moist.

Overall, these NWP modeling efforts during MPEX demonstrated the real-time feasibility of continuously cycling, limited-area, EnKF-based convection-allowing ensemble prediction systems. Despite some shortcomings, results indicated that these forecasts were ultimately successful, and they provided valuable real-time guidance to MPEX forecasters during field program operations (e.g., Weisman et al. 2015). Inspired by this successful real-time implementation, NCAR plans to further develop and experiment with real-time convection-allowing ensemble forecast systems in upcoming years.

Acknowledgments. This work was partially funded by the Short Term Explicit Prediction Program (STEP). We would like to acknowledge high-performance computing support from Yellowstone (ark:/85065/d7wd3xhc) provided by NCAR’s Computational and Information Systems Laboratory. Three anonymous reviewers provided constructive comments that improved this paper.

REFERENCES


Putnam, B. J., M. Xue, Y. Jung, N. A. Snook, and G. Zhang, 2014: The analysis and prediction of microphysical states and polarimetric variables in a mesoscale convective system using


