

Using Varied Microphysics to Account for Uncertainty in Warm-Season QPF in a Convection-Allowing Ensemble

JEFFREY D. DUDA AND XUGUANG WANG

School of Meteorology and Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma

FANYOU KONG

Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma

MING XUE

School of Meteorology and Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma

(Manuscript received 22 September 2013, in final form 14 January 2014)

ABSTRACT

Two approaches for accounting for errors in quantitative precipitation forecasts (QPFs) due to uncertainty in the microphysics (MP) parameterization in a convection-allowing ensemble are examined. They include mixed MP (MMP) composed mostly of double-moment schemes and perturbing parameters within the Weather Research and Forecasting single-moment 6-class microphysics scheme (WSM6) MP scheme (PPMP). Thirty-five cases of real-time storm-scale ensemble forecasts produced by the Center for Analysis and Prediction of Storms during the NOAA Hazardous Weather Testbed 2011 Spring Experiment were examined.

The MMP ensemble had better fractions Brier scores (FBSs) for most lead times and thresholds, but the PPMP ensemble had better relative operating characteristic (ROC) scores for higher precipitation thresholds. The pooled ensemble formed by randomly drawing five members from the MMP and PPMP ensembles was no more skillful than the more accurate of the MMP and PPMP ensembles. Significant positive impact was found when the two were combined to form a larger ensemble.

The QPF and the systematic behaviors of derived microphysical variables were also examined. The skill of the QPF among different members depended on the thresholds, verification metrics, and forecast lead times. The profiles of microphysics variables from the double-moment schemes contained more variation in the vertical than those from the single-moment members. Among the double-moment schemes, WDM6 produced the smallest raindrops and very large number concentrations. Among the PPMP members, the behaviors were found to be consistent with the prescribed intercept parameters. The perturbed intercept parameters used in the PPMP ensemble fell within the range of values retrieved from the double-moment schemes.

1. Introduction

Numerical weather prediction (NWP) using an ensemble of forecasts has become increasingly popular since the early 1990s when such centers as the National Centers for Environmental Prediction in the United States and the European Centre for Medium-Range Weather Forecasts began running global ensembles on

the synoptic scale for forecasts ranging up to 16 days (e.g., Toth and Kalnay 1993; Molteni et al. 1996). The purpose of using an ensemble instead of a single deterministic forecast was to account for the uncertainty in the model forecasts.

Early work on ensemble forecasting first focused on addressing the uncertainty in the initial conditions (Toth and Kalnay 1993, 1997; Buizza and Palmer 1995; Molteni et al. 1996; Houtekamer et al. 1996; Wang and Bishop 2003; Wang et al. 2004). However, not long afterward, researchers began accounting for uncertainty in a forecast due to model error as well (Du et al. 1997; Stensrud et al. 2000; Hou et al. 2001). Global- and regional-scale ensemble modeling has since been extensively studied

Corresponding author address: Jeff Duda, School of Meteorology and Center for Analysis and Prediction of Storms, University of Oklahoma, Rm. 2500, 120 David Boren Blvd., Norman, OK 73072.
E-mail: jeffduda319@gmail.com

(Du et al. 1997; Wang and Bishop 2003; Palmer et al. 2005; Wei et al. 2008); techniques to appropriately generate initial condition perturbations and model perturbations have resulted in large-scale and mesoscale ensembles that are increasingly skillful and statistically reliable. In contrast, higher-resolution, convective-scale ensemble forecasting remains in a relative infancy. Although research related to convective-scale ensemble prediction has increased in the last few years (e.g., Kong et al. 2006, 2007a,b; Xue et al. 2007; Mittermaier 2007; Clark et al. 2008, 2010; Schwartz et al. 2010; Vie et al. 2011; Johnson et al. 2011a,b; Johnson and Wang 2012, 2013; Johnson et al. 2013, 2014; Ropnack et al. 2013; Caron 2013), many questions addressing the optimal design of convection-allowing model ensembles remain. Furthermore, computer power has improved to the point where limited-area simulations with resolutions on the order of 1 km have become feasible, even for the continental U.S. domain (e.g., Xue et al. 2009; Xue et al. 2013; Johnson et al. 2013). Therefore, experimental and semi-operational models now run at these resolutions and can resolve smaller-scale features. Perhaps the most actively researched of all such features as it pertains to operational NWP are those that produce precipitation. Specifically, convective storms and mesoscale convective systems, which are dominantly warm-season phenomena, are now explicitly simulated without cumulus parameterization.

Warm-season quantitative precipitation forecast (QPF) skill has historically been poor compared to that from the cold season (Ralph et al. 2005). Additionally, growth rates of forecast errors for convective scales can be highly nonlinear (Hohenegger and Schär 2007). Therefore, extensive research that optimizes convective-scale ensemble model forecasting to improve warm-season probabilistic quantitative precipitation forecasting is needed.

There are many ways to address model error in an ensemble. One particular way is to represent error stochastically (Buizza et al. 1999; Berner et al. 2009). Another way is by using multiple models and varied physics options within a given model (e.g., Doblas-Reyes et al. 2000; Hou et al. 2001; Ebert 2001; Hagedorn et al. 2005; Xue et al. 2007, 2008, 2009, 2011; Kong et al. 2007b, 2010, 2011; Candille 2009; Berner et al. 2011; Hacker et al. 2011; Charron et al. 2010; Johnson et al. 2011a,b; Johnson and Wang 2012). Since convection is explicitly represented by microphysics parameterizations in high-resolution NWP simulations, warm-season QPF can be very sensitive to uncertainties in microphysics parameterizations. Earlier studies have investigated the sensitivity of convective-scale features like supercell thunderstorms and squall lines to the complexity of, and the values of parameters used in, particle size distributions in microphysics schemes (Gilmore et al. 2004; Snook and

Xue 2008; Tong and Xue 2008; Morrison et al. 2009; Dawson et al. 2010; Putnam et al. 2013). Great sensitivity of accumulated precipitation and cold pool size and intensity to parameters that describe rain, graupel, and hail size distributions has been found. It has also been found that simulations using multimoment microphysics (MP) schemes produce more realistic reflectivity structures and more realistic stratiform precipitation regions in squall lines. However, most of these studies were conducted within an idealized and/or deterministic framework. There are also a large number of studies (e.g., Eckel and Mass 2005; Xue et al. 2009; Clark et al. 2008, 2009, 2010; Schwartz et al. 2010; Hacker et al. 2011; Berner et al. 2011; Clark et al. 2011; Johnson et al. 2011a,b; Johnson and Wang 2012, 2013) that account for model physics uncertainty in an ensemble by varying the MP schemes. However, the use of mixed microphysics in these studies is usually embedded within other varied physics such as convection, boundary layer, and radiation parameterizations, thus masking the contribution of the varied microphysics to the total benefit of the ensemble.

In this study we examine the use of varied microphysics in a convection-allowing forecast within an ensemble and real-data framework. Two approaches to addressing model error from uncertainties in the microphysics are tested. In one approach the values of some parameters within a single MP scheme are varied. This way one can address uncertainty by sampling the distribution of possible values of parameters significant to cloud and precipitation physics. The perturbed parameters in this study include the intercept parameters for precipitation particle size distributions (PSDs) and graupel density. The resultant ensemble is hereinafter denoted as perturbed parameter microphysics (PPMP). A single-moment MP scheme is used for testing this approach. These parameters were chosen based on earlier studies that examined the uncertainty ranges of PSD-related parameters and showed great sensitivities of modeled storm dynamics and precipitation forecasts to these parameters (Gilmore et al. 2004; Snook and Xue 2008; Tong and Xue 2008; Jung et al. 2010; Yussouf and Stensrud 2012). In the second approach separate MP parameterizations are used. Not only can this approach address uncertainty in parameter values, but it can also address uncertainty in the microphysical processes within the parameterization schemes. The resultant ensemble is hereinafter denoted as mixed microphysics (MMP). Most of the MP parameterizations used in the MMP ensemble predict two moments of at least one of the precipitating species. Therefore, this experiment also enables an investigation into whether an ensemble using various sophisticated MP parameterizations is superior to one using simpler and more computationally efficient MP parameterizations

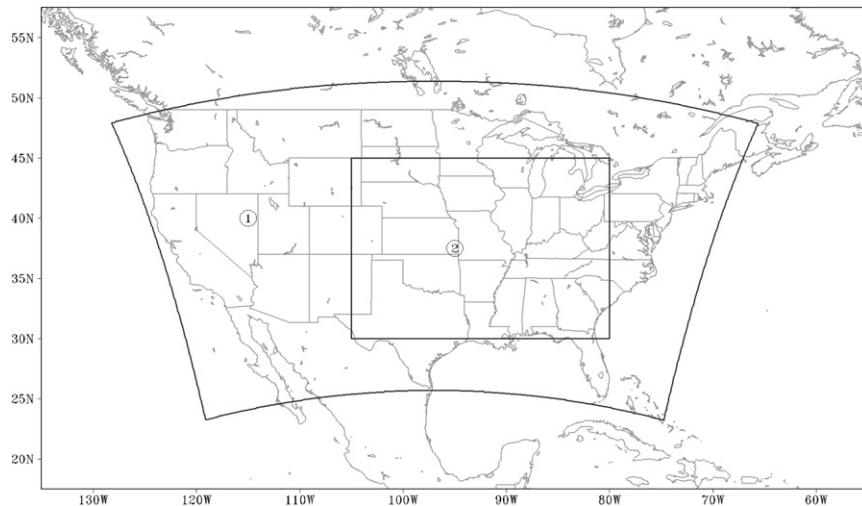


FIG. 1. Model domain (circled “1”) and verification domain (circled “2”) used in this study.

with perturbed parameters. This research presents a step in the investigation of how to best sample the microphysics errors in a convection-allowing ensemble for warm-season QPF.

The purpose of this paper is twofold. One purpose is to examine the effectiveness of different approaches to accounting for model microphysics error in convective-scale probabilistic QPF (PQPF) via conducting various verifications. The mixed results from a comparison of the MMP and PPMP ensembles as shown later motivate investigation of the combination of the approaches; the combination of the MMP and PPMP ensembles formed a third ensemble, called the pooled ensemble. The second purpose is to examine and compare the systematic behaviors of the various microphysical variables over a broad range of cases for the MP schemes that compose the MMP and PPMP ensembles in convective-allowing forecasts to facilitate ensemble design in the future.

2. Experimental setup

a. Model description

Version 3.2.1 of the Weather Research and Forecasting Model (WRF), with the Advanced Research dynamic core (ARW; Skamarock et al. 2008), was used as the NWP model. WRF-ARW was the primary NWP model used in the Storm-Scale Ensemble Forecast system (SSEF) conducted by the Center for Analysis and Prediction of Storms and used in the National Oceanic and Atmospheric Administration’s (NOAA) Hazardous Weather Testbed (HWT) 2011 Spring Experiment (Kong et al. 2011). The 2011 Spring Experiment extended from late April to early June, which included 35 cases [generally

one case each weekday from 27 April to 10 June; Kong et al. (2011)]. Forecasts were initialized at 0000 UTC and ran for 36 h over the contiguous U.S. model domain (see Fig. 1).

Ten members were designed to investigate which approach to addressing model error due to microphysics uncertainty results in a more skillful ensemble. All other aspects of the model configuration, including initial and lateral boundary conditions, as well as other physics parameterizations, were identical among the 10 members. The members used six different MP schemes including Thompson (Reisner et al. 1998; Thompson et al. 2008), Ferrier (Ferrier et al. 2002, 2013; Ferrier 2013), Morrison (Morrison et al. 2009), Milbrandt–Yau (MY; Morrison and Milbrandt 2011), and WRF single-moment 6-class microphysics scheme (WSM6; Hong and Lim 2006), and WRF double-moment 6-class microphysics scheme (WDM6; Lim and Hong 2010). Relevant characteristics and parameters of the members and MP schemes are given in Table 1, and a description of the differences between the members is given in the appendix. The MMP ensemble contained six members that used the Thompson, Ferrier, MY, Morrison, WDM6, and WSM6 schemes. The PPMP ensemble contained five members, each of which used the WSM6 scheme but with different values of the rain and graupel intercept parameter of the respective PSDs and different graupel densities. The graupel density ranges between values typical of graupel and hail, while the intercept parameters take values within observed uncertain ranges [see discussions in Tong and Xue (2008) on the uncertain ranges of values]. The WSM6 scheme was a member of both ensembles. To investigate whether a combination of the above two approaches to addressing microphysics error is superior to

TABLE 1. Description of the 10 SSEF members used in this study. The first six members (in boldface) compose the MMP ensemble, whereas the bottom five members (in italics) compose the PPMP ensemble.

Member name and scheme (perturbation)	No. of classes ^a	Double-moment species	N_{0r} (m^{-4})	N_{0g} (m^{-4})	ρ_g (kg m^{-3})	α_r	α_g
Thompson	6	Rain, ice	—	Variable	400	0	0
Ferrier+^b	3	—	8×10^6	—	—	—	—
Morrison	6	Rain, graupel, snow, ice	—	—	400	0	0
Milbrandt–Yau^c	7	All but vapor	—	—	400	2	0
WDM6	6	Rain, cloud water	—	4×10^6	500	1	0
WSM6 (control)	6	—	8×10^6	4×10^6	500	0	0
<i>WSM6 (M1)</i>	6	—	8×10^6	4×10^4	913	0	0
<i>WSM6 (M2)</i>	6	—	8×10^7	4×10^6	500	0	0
<i>WSM6 (M3)</i>	6	—	8×10^5	4×10^2	913	0	0
<i>WSM6 (M4)</i>	6	—	8×10^5	4×10^3	913	0	0

^a Water classes (species) include vapor, liquid cloud drops, frozen cloud ice particles, rain, snow, graupel, and hail. Hail was the least common species included.

^b The Ferrier+ scheme treats all condensate species as one.

^c The Milbrandt–Yau scheme adds hail as a separate class with $\rho_h = 900 \text{ kg m}^{-3}$ and $\alpha_h = 3$.

either approach alone, a third ensemble—the “pooled” ensemble—was composed of members from the PPMP and MMP ensembles.

The initial conditions were generated by the Advanced Regional Prediction System’s three-dimensional variational data assimilation and cloud analysis system (Gao et al. 2004; Xue et al. 2003; Hu et al. 2006) using 0000 UTC operational 12-km North American Mesoscale Model (NAM) analyses as the background field. The data assimilated include Weather Surveillance Radar-1988 Doppler radial velocity and reflectivity, surface pressure, horizontal wind, temperature, and specific humidity from the Oklahoma Mesonet, surface aviation observations, and wind profiler data. The boundary conditions were provided by 12-km NAM model forecasts initialized at 0000 UTC. The Noah land surface model (Ek et al. 2003) and the Mellor–Yamada–Janjić (MYJ) boundary layer parameterization (Mellor and Yamada 1982; Janjić 2002) were used for all 10 members.

b. Verification methodology

Verification was performed over a portion of the model domain containing the central and southern United States (Fig. 1). This domain was chosen because the environment in this region during the late spring and early summer supports organized and intense convection with tight spatial gradients. Therefore, the verifications in this region are a representative evaluation of the skill of convective-scale warm-season QPF. A variety of verification metrics were used, including gridpoint-based, contingency-table-based, and neighborhood-based metrics. These metrics are introduced as they are encountered. Probabilistic verification was computed using uncalibrated model output. Therefore, probability of precipitation (PoP) was computed as the proportion of members in

which a specified 1-h precipitation amount threshold¹ was exceeded. PoP values across the verification domain constituted the PQPF verified in this study.

Verifying precipitation data were provided by the National Mosaic and Multi-sensor quantitative precipitation estimation (QPE) project (NMQ) of the National Severe Storms Laboratory (Zhang et al. 2011). The output fields are available at a horizontal resolution of 0.01° . Rainfall rate estimates are produced every 2.5 min and are integrated to produce QPEs for a variety of accumulation intervals, including 1 h. The QPEs are the result of a combination of radar-estimated rainfall and rain-gauge measurements. The radar rainfall estimates are calculated using a variety of Z–R relationships applied to the hybrid scan reflectivity field (essentially, the reflectivity at the lowest height above ground). Gauge correction uses inverse-distance weighting with corrections for gauge density and quality control measures to ignore measurements from gauges with anomalously large errors. The NMQ QPEs were regressed to the verification domain using bilinear interpolation. Verifications were performed on 1-h accumulated precipitation fields. The NMQ QPEs have been used in earlier studies to verify storm-scale precipitation forecasts (e.g., Johnson and Wang 2012, 2013; Johnson et al. 2013, 2014).

The MMP, PPMP, and pooled ensembles contained 6, 5, and 10 members, respectively. Earlier studies have shown that verification scores can be sensitive to ensemble size (Richardson 2001). The purpose of the study is to investigate different methods of accounting for microphysics scheme errors in a storm-scale ensemble. To

¹ The thresholds tested included 0.254, 2.54, 6.35, 12.7, and 25.4 mm (0.01, 0.10, 0.25, 0.50, and 1.00 in.).

minimize the impact of different ensemble sizes on this purpose, a single ensemble size was used for verification. Since the PPMP ensemble had the smallest size at five members, five members were selected randomly from each of the MMP and pooled ensembles to constitute a resample. One hundred such resamples were obtained for each of the MMP and pooled ensembles. The verification scores were averaged over these resamples for the MMP and pooled ensembles before being compared with the scores from the PPMP ensemble. The choice of 100 resamples follows earlier studies using bootstrap resampling techniques in statistical significance tests (e.g., Wang and Bishop 2005). In addition, each five-member resample from the pooled ensemble was selected such that each resample is randomly composed of either two or three members from the MMP and PPMP ensembles to ensure each of the MMP and PPMP ensembles were equally represented in the resample of the pooled ensemble.

3. Verification of quantitative precipitation forecasts

a. Verification of individual members

The existence of particularly skillful or unskillful members could impact the overall ensemble performance (Ebert 2001). Therefore, individual members were first verified. Figure 2 shows the average hourly precipitation forecast errors. A periodic trend, featuring local maxima around 0000 UTC (0- and 24-h forecasts), was observed. This trend has been commonly observed in the SSEF in the past (e.g., Clark et al. 2009; Johnson and Wang 2012) and in other verification scores in this study. The relative maxima around forecast hours 0 and 24 identify “convection maxima,” or diurnal peaks in convective activity around 0000 UTC, which correspond to 1900 or 2000 local time in the central and eastern United States. The convection maximum around forecast hour 0 will hereinafter be referred to as the first convection maximum, while that around forecast hour 24 will be called the second convection maximum. The convection minima, on the other hand, refer to the reduced convective activity centered near forecast hours 12 and 36.

Significant variation among members accompanied the mean precipitation errors in the first few forecast hours (i.e., f01²–f04) as each member adjusted from the initial conditions that contained cloud and hydrometeors created by the cloud analysis with radar reflectivity data assimilation (Fig. 2). At the first convection

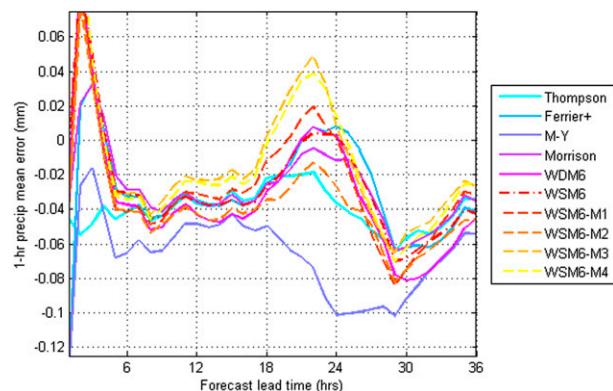


FIG. 2. Domain- and case-average mean error of 1-h QPF for individual members.

maximum, the Thompson and MY schemes produced too little precipitation whereas all other schemes produced too much. The overproduction was very large in the WDM6, WSM6 (control), and WSM6-M1 schemes. All members produced too little precipitation during the convection minima. During the second convection maximum large variation was found in the mean errors as some members produced too much precipitation (all from the PPMP ensemble) whereas others produced too little. In particular, the WSM6-M3 and WSM6-M4 members had a larger positive bias in the f18–f24 period while the MY scheme had a much larger negative bias in the f18–f30 period.

The frequency biases (the ratio of the number of grid points at which forecast precipitation exceeded a threshold to the number of grid points at which precipitation was observed to have exceeded that threshold) of the WSM6-M3 and WSM6-M4 schemes were not appreciably higher than those of other members at f18–f24 for lower thresholds (Fig. 3), but the frequency bias increased significantly for the higher thresholds, indicating those members overforecast heavy precipitation episodes. The overall negative error for the MY scheme is consistent with its frequency biases, which were almost exclusively below 1.0, indicating a consistent trend to underforecast precipitation at all thresholds. The increased variability in frequency biases among the different schemes at high thresholds is related to the counts in each cell of the 2×2 contingency table. It is rare for 12.7 or 25.4 mm of precipitation to fall in 1 h, occurring only within the cores of stronger thunderstorms. Therefore, the contingency table has small numbers of hits, false alarms, and misses at these thresholds—a few thousandths of a percent of the total number of grid points at which verification was performed. Since frequency bias depends on the number of hits, false alarms, and misses, seemingly large differences in frequency bias can appear between different schemes.

² We use f01 to mean forecast hour 1. This symbolism will be used throughout the remainder of the paper.

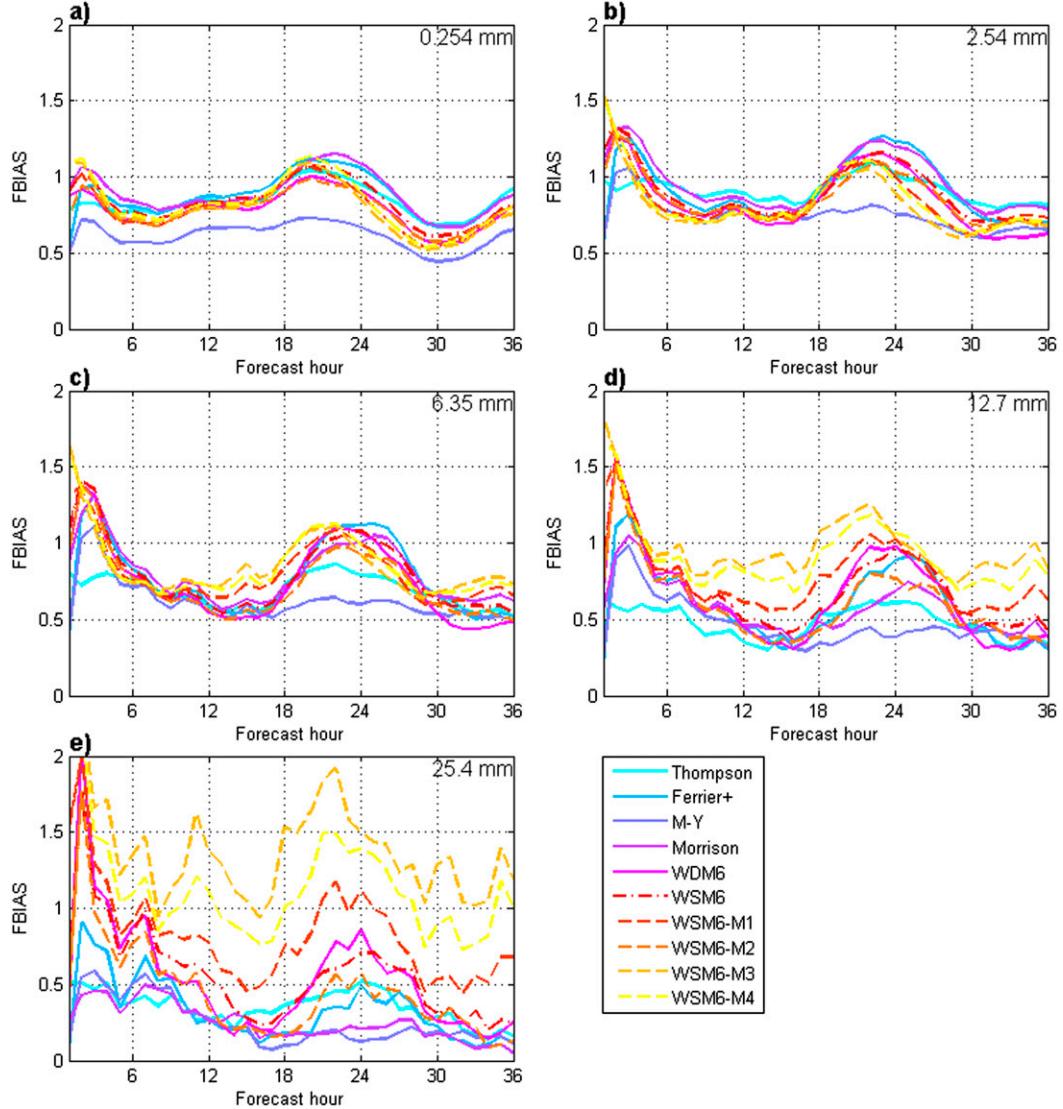


FIG. 3. As in Fig. 2, but for frequency biases at the various thresholds indicated.

Regarding equitable threat scores (ETS), a large decrease in skill from very high values within the first one or two forecast hours was observed (Fig. 4). This higher skill at early forecast hours can be attributed to the initialization of preexisting convection within the initial conditions from the assimilation of radar data (e.g., Xue et al. 2008, 2013), the effects of which wore off quickly. The benefits to assimilating radar data for very short-range convective-scale QPF on the order of a few hours have been observed in previous SSEF forecasts (e.g., Johnson and Wang 2012; Kong et al. 2011; Xue et al. 2008, 2009, 2011, 2013). ETS were very similar among the members with a few exceptions; the Thompson and Morrison schemes had higher ETS than the other members for most forecast hours at the lightest threshold. The

improved performance of those schemes dwindled with increasing threshold. No member appeared to perform worse except for the WDM6 scheme around f30. ETS for the PPMP members were tightly clustered, and no single member stood out except for around f06, when the WSM6-M3 and WSM6-M4 schemes were slightly inferior at moderate thresholds. ETS were highly variable at the highest thresholds, owing to the decreased numbers of hits, false alarms, and misses in the contingency table due to limited samples.

Verification metrics based on gridpoint values such as the ETS may give a misleading interpretation of the skill of high-resolution precipitation forecasts due to the small horizontal scale of features compared to the horizontal scale of acceptable spatial errors (Baldwin et al.

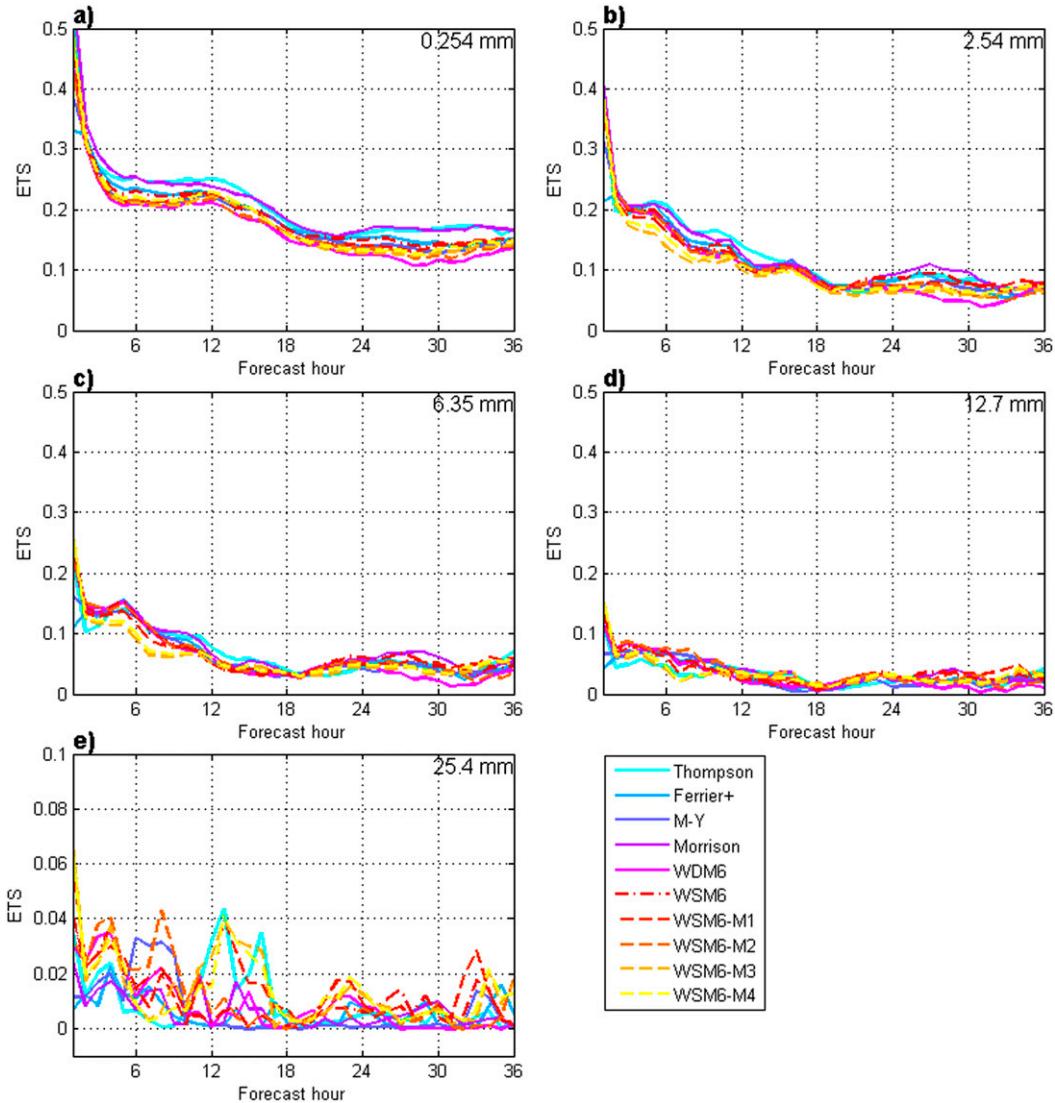


FIG. 4. As in Fig. 3, but for ETS.

2001). For that reason it is more appropriate to consider neighborhood-based verification metrics. One such measure of neighborhood-based verification is applied to the Brier score and is referred to as the fractions Brier score (FBS; Roberts and Lean 2008; Schwartz et al. 2010). Despite being a probabilistic verification metric, the FBS can be computed for deterministic forecasts by constructing probabilistic forecasts based on spatial neighborhoods (Theis et al. 2005). A square neighborhood of radius 48 km (12 grid points) was used for all verifications in this study applied to a neighborhood (Johnson and Wang 2012), including those for both deterministic and ensemble forecasts. Thus, FBS is a mean square difference between the forecast PoP in a spatial neighborhood and the proportion of observed precipitation events in

a neighborhood around a given point. A diurnal cycle was present in the FBSs, and the highest FBSs (poorest performance) occurred at the second convection maximum for low thresholds and at the first convection maximum for the highest thresholds (Fig. 5). The Thompson and MY schemes were commonly the best for a given threshold and forecast lead time. However, the WSM6-M3 and WSM6-M4 schemes had better FBSs around the second convection maximum at the 2.54-mm threshold. The WDM6 scheme commonly had the worst FBSs around the second convection maximum, except at the highest threshold, indicating consistently poor performance of that member around that time of forecast. Among the PPMP members, the WSM6-M2 scheme typically had the worst FBS around the convection maxima, while the

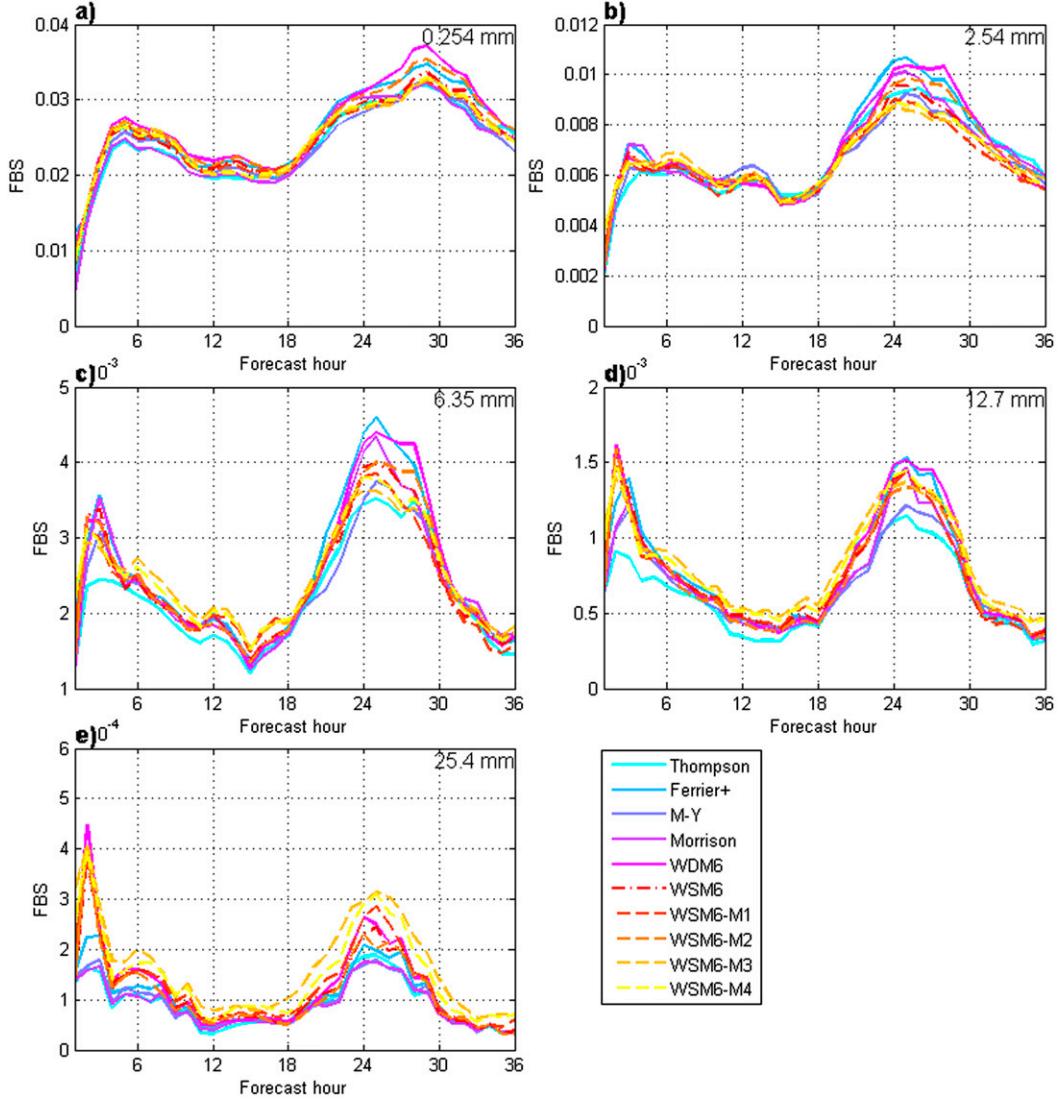


FIG. 5. As in Fig. 3, but for FBS.

WSM6-M3 scheme frequently had the worst FBS away from the convection maxima.

The area under the relative operating characteristic curve (ROC; Mason 1982) for each member was computed using neighborhood forecasts verified against single-point observations. The ROC areas indicate that the MY scheme was commonly the worst-performing scheme, especially around the second convection maximum (Fig. 6). The Thompson and Morrison schemes had larger (better) ROC areas for the lower thresholds, but the WSM6-M3 and WSM6-M4 schemes had larger areas at the highest thresholds and for later forecast hours. For the lower thresholds the envelope of scores from the MMP ensemble members tended to contain those of the PPMP ensemble members, implying more

variability in skill within the MMP ensemble than within the PPMP ensemble. At higher thresholds, however, the ROC areas of the PPMP ensemble members tended to be larger than most of those of the MMP ensemble members.

The metrics discussed here indicate that members of the MMP ensemble commonly underpredicted both the amount and areal coverage of precipitation, especially at high thresholds. In contrast, members of the PPMP ensemble were generally less biased. However, these PPMP members were also generally less skillful, with a few exceptions. Superior skill was frequently noted with the Thompson and Morrison schemes, in agreement with Clark et al. (2012), and inferior performance was commonly seen with the WDM6 scheme, although the

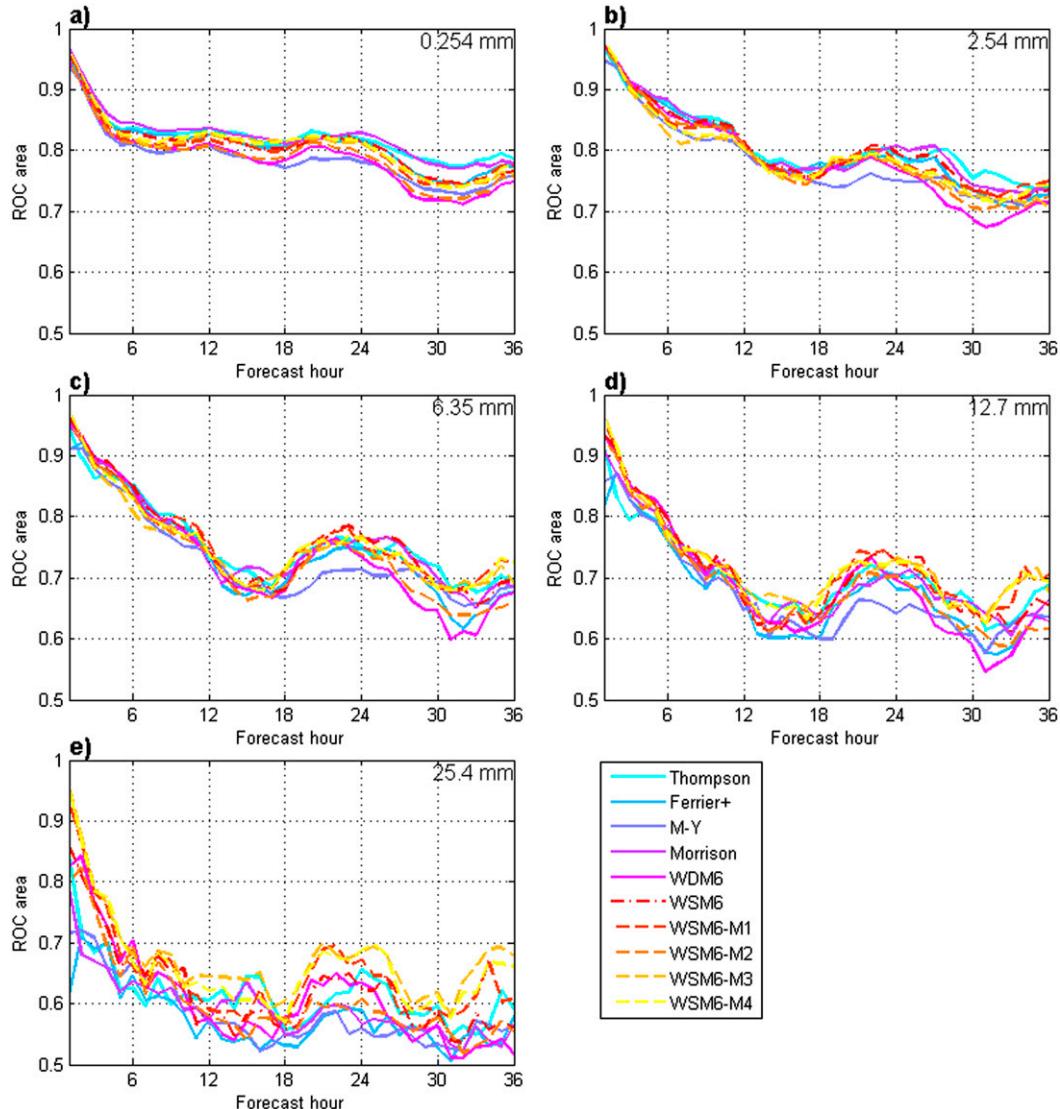


FIG. 6. As in Fig. 3, but for the area under the ROC curve.

ordering of skill was somewhat a function of metric, forecast lead time, and threshold. The WSM6 scheme frequently ranked near the middle of all schemes and near the middle of the WSM6 perturbation schemes.

b. Verification of the MMP, PPMP, and the pooled ensembles

All of the following discussion refers to the mean values of verification scores from the five-member resamples except where otherwise noted. The significance of the difference was determined using the standard deviation of the scores from the resamples, similar to the method used in Wang and Bishop (2005). Here, means differing by three standard deviations or more are declared to represent ensembles having statistically significantly

different skill. The larger value of the standard deviations between the ensembles compared was used. After about f04, the magnitude of the mean error of the ensembles was around one order of magnitude less than the season-average rain amounts per grid point, which ranged from 0.10 to 0.30 mm (Fig. 7b). This was especially true during f12–f27. For early and late forecast hours the errors were about of the same order of magnitude as the average precipitation amount at each grid point, which indicates that the bias at these forecast hours could be significant. The ensembles were biased low during f04–f18 and f24–f36 (Fig. 7a). The bias was variable among different ensembles during f18–f25 and in the first 3 h of the forecast. For the first 3 h the ensembles produced too much precipitation, with the PPMP ensemble producing the most

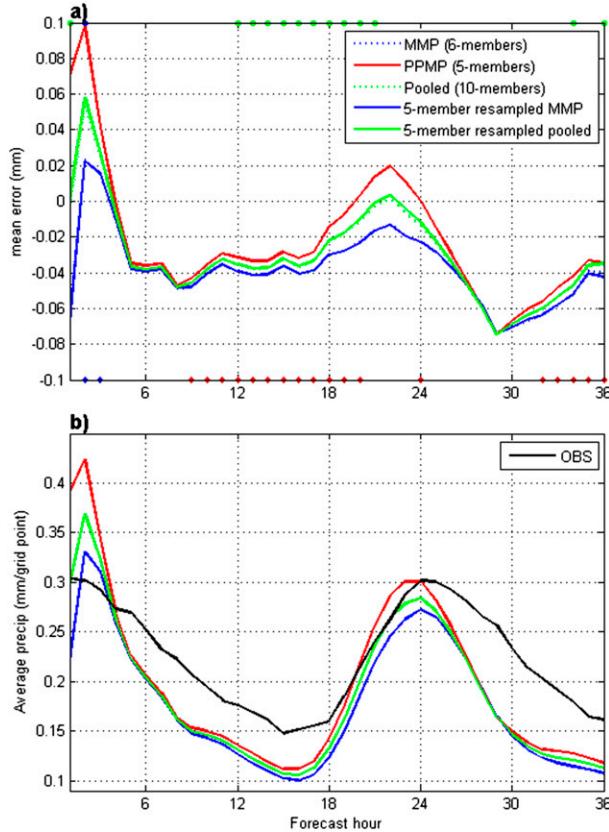


FIG. 7. (a) Domain- and case-average mean error of 1-h ensemble mean QPF. Diamonds across the bottom indicate forecast hours at which the difference between the MMP and PPMP ensembles was significant, with the color indicating which ensemble was superior, whereas filled circles across the top indicate forecast hours at which the difference between the MMP and pooled ensembles was significant. (b) Domain-average 1-h ensemble mean QPF and observed QPF (black line).

precipitation and the MMP ensemble being the least positively biased. During f21–f23, the PPMP and MMP ensembles had similar absolute biases, with the MMP ensemble biased low and the PPMP ensemble biased high. During f18–20 and f24–25, the PPMP ensemble was less negatively biased than the MMP ensemble. The exceptionally low bias of the MY scheme is the leading cause of this low bias in the MMP ensemble, as the remaining members of the MMP ensemble were less biased during that period. The bias for the pooled ensemble was, as expected, between that of the MMP and PPMP ensembles. Around the second convection maximum the pooled ensemble had a nearly zero mean error. The bias for the full 10-member pooled ensemble was nearly identical to that of the resampled pooled ensemble. The same was true for the full six-member MMP ensemble compared to the resampled MMP ensemble. The frequency biases were consistent with the mean

errors (not shown), regardless of precipitation accumulation thresholds.

The reliability of the ensembles was evaluated through examination of rank histograms (Hamill 2001). Easily recognizable signatures that represent the dispersion and bias characteristics of ensembles can be illustrated by rank histograms. Note that to reduce the impact of the ensemble size on evaluating the flatness of the rank histogram, resampled ensembles with an ensemble size of five were used. The underdispersive nature of all ensembles is apparent (Fig. 8). The underdispersion was generally worse during the first 10 forecast hours and better around the second convection maximum. The left skew of the histograms is consistent across all forecast hours with varying degrees of magnitude. Also, while it has been shown that a composite of signals may result in rank histograms that give a misleading impression of the reliability of an ensemble (Hamill 2001), rank histograms computed from a few localized subsets and a scattered subset of the domain (counts were obtained from grid points widely separated in space) revealed the same signatures as those using all data points. Therefore, the aggregated signatures are unlikely to be the result of a combination of high bias in some regions of the domain and low bias in other regions. Although all ensembles are underdispersive, a way of quantifying the extent of underdispersion ext is by the proportion of values that fall within the outer bins of the histogram normalized by the proportion of values expected to be located in each rank for a flat histogram (value in top-right corner of each panel in Fig. 8). The values of ext indicate that the MMP and PPMP ensembles were underdispersive to about the same extent, although the value of ext indicates that the MMP ensemble was slightly better than the PPMP ensemble. The value of ext for the pooled ensemble is in between that of MMP and PPMP. Note that in this study, the ensemble examined only accounted for the micro-physics scheme errors. Early work analyzing the SSEF that included perturbations to represent other sources of forecast errors has shown flatter histograms [cf. Fig. 8a of Clark et al. (2009) and Fig. 15 of Xue et al. (2011)].

We now consider the neighborhood-based FBS. A diurnal cycle was observed with the FBSs. For the highest two thresholds the highest (worst) scores occurred during the first convection maximum, whereas for the remaining thresholds the worst scores occurred around the second convection maximum (Fig. 9). This behavior is identical to that seen in the FBSs for the individual members (section 3a). The best FBSs occurring at f01–f02 were likely due to the improved initial conditions due to the assimilation of radar data. At most thresholds and forecast lead times the MMP ensemble performed better than the PPMP ensemble. The biggest exception was around

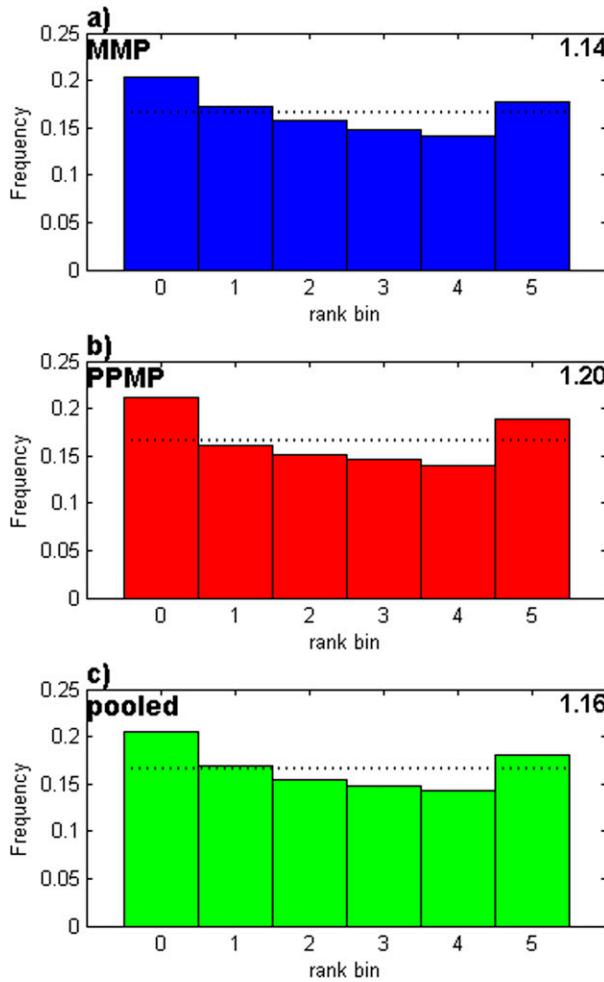


FIG. 8. Rank histograms for 1-h QPF integrated over all forecast hours for the (a) MMP, (b) PPMP, and (c) pooled ensembles. The parameter ext , defined as the ratio of the proportion of values in the extreme ranks to the proportion of values in the extreme ranks of a flat histogram, is displayed at the top right of each panel. A rank histogram of a perfectly dispersive ensemble is indicated by the dotted black line in each panel.

the second convection maximum for the 2.54- and 6.35-mm thresholds, when the skill of the MMP ensemble was equal to that of the PPMP ensemble. The FBSs of the pooled ensemble were generally smaller than those of the PPMP and nearly identical to those of the MMP ensemble. The largest difference between the MMP and pooled ensembles occurred around the convection maxima at the highest two thresholds where the MMP ensemble had statistically significantly better FBSs than the PPMP ensemble. The FBSs of the full 10-member pooled ensemble were generally similar to those of the resampled pooled ensemble. However, the full 10-member pooled ensemble was generally better around the second convection maximum for the lower thresholds, thus showing

the positive impact from using a larger ensemble during periods of intense convective activity at longer lead times.

Probabilistic forecasts derived from spatial neighborhoods are further evaluated by computing the areas under ROC curves. Scores decreased only slightly with increasing threshold and forecast hour (Fig. 10), and all scores were greater than the no-skill value of 0.5. The spread among resampled scores was small. This is probably due to the larger number of samples used to compute scores for each neighborhood (several thousand for neighborhood verification). At the lowest two thresholds the MMP ensemble was generally more skillful than the PPMP ensemble. At the highest two thresholds the PPMP ensemble was more skillful than the MMP ensemble. At the low thresholds the ROC area of the pooled ensemble was nearly identical to that of the MMP ensemble. At the high thresholds the ROC area of the pooled ensemble was between that of the MMP and PPMP ensembles with better scores than the MMP for most lead times. These results are consistent with the ROC area for the individual members (section 3a). The ROC area for the full 10-member pooled ensemble was larger than that for the other ensembles at all forecast hours and precipitation thresholds, suggesting the positive impact of using a larger ensemble. This result is consistent with the findings of Clark et al. (2011).

The ensemble verification presented here suggests that the MMP ensemble was more skillful than the PPMP ensemble according to FBSs for most lead times and thresholds and less skillful than the PPMP at high thresholds according to ROC areas. While this result may seem contradictory, each metric assesses different aspects of the ensemble PQPF. The FBS measures the mean squared error of the neighborhood PQPF relative to the neighborhood coverage of observed precipitation, whereas the ROC measures the ability of the ensemble to distinguish between “yes” and “no” forecasts for precipitation exceeding a threshold. The pooled ensemble did not outperform the MMP or PPMP ensemble except for at a small number of thresholds, forecast hours, and metrics.

The resampling allowed for an examination of the impact of ensemble size on the quality of the forecasts. Differences between the full six-member and the five-member resampled MMP ensembles were generally imperceptible. However, when differences were found, it was generally the full six-member MMP ensemble that was superior. There were larger, more noticeable differences between the full 10-member and 5-member resampled pooled ensembles than between the full and resampled MMP ensembles. The full 10-member pooled ensemble was always better than the resampled pooled ensemble.

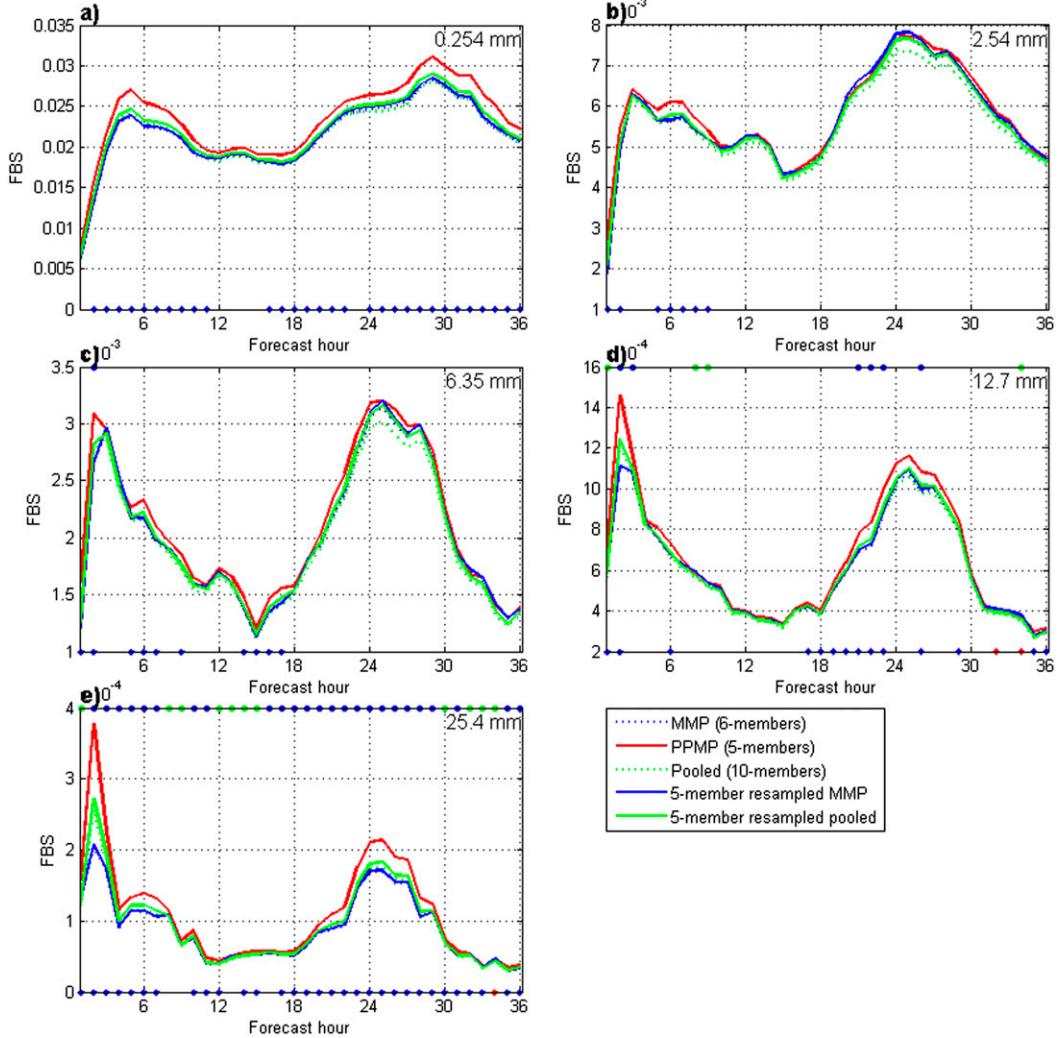


FIG. 9. FBSs for 1-h ensemble mean QPF for the various thresholds indicated in the top right of each panel. Significant differences are indicated as in Fig. 7.

The preceding discussion implies that the relative skill of using a combination of mixed MP schemes and perturbing parameters within a single MP scheme to sample the microphysics errors can be dependent on methods of verification and the precipitation thresholds. Using a combination of the two methods generally is no more skillful than using either approach separately. Ensemble size seems to have as large an impact on the quality of the forecasts as the method used to construct the ensembles.

4. A comparison of the microphysical parameters

Microphysics parameters such as the intercept parameters for PSDs and particle density have been shown to impact the simulation of convective systems. Specifically,

Gilmore et al. (2004), Snook and Xue (2008), Dawson et al. (2010), and Morrison and Milbrandt (2011) have shown that the dynamics and precipitation patterns in supercells, in particular, are sensitive to the values of the rain and hail intercept parameters, graupel and hail density, and the number of predicted moments of the PSD for precipitating species in bulk microphysics schemes. Motivated by these earlier studies, two methods to account for microphysics scheme errors in convection-allowing ensembles were tested within the Center for Analysis and Prediction of Storms (CAPS) SSEF system in 2011. In the PPMP ensemble, the rain and graupel intercept parameters and the graupel density were perturbed within a single microphysics scheme. Alternatively, the MMP ensemble contained a variety of microphysics schemes. To facilitate developing a strategy to optimally account for

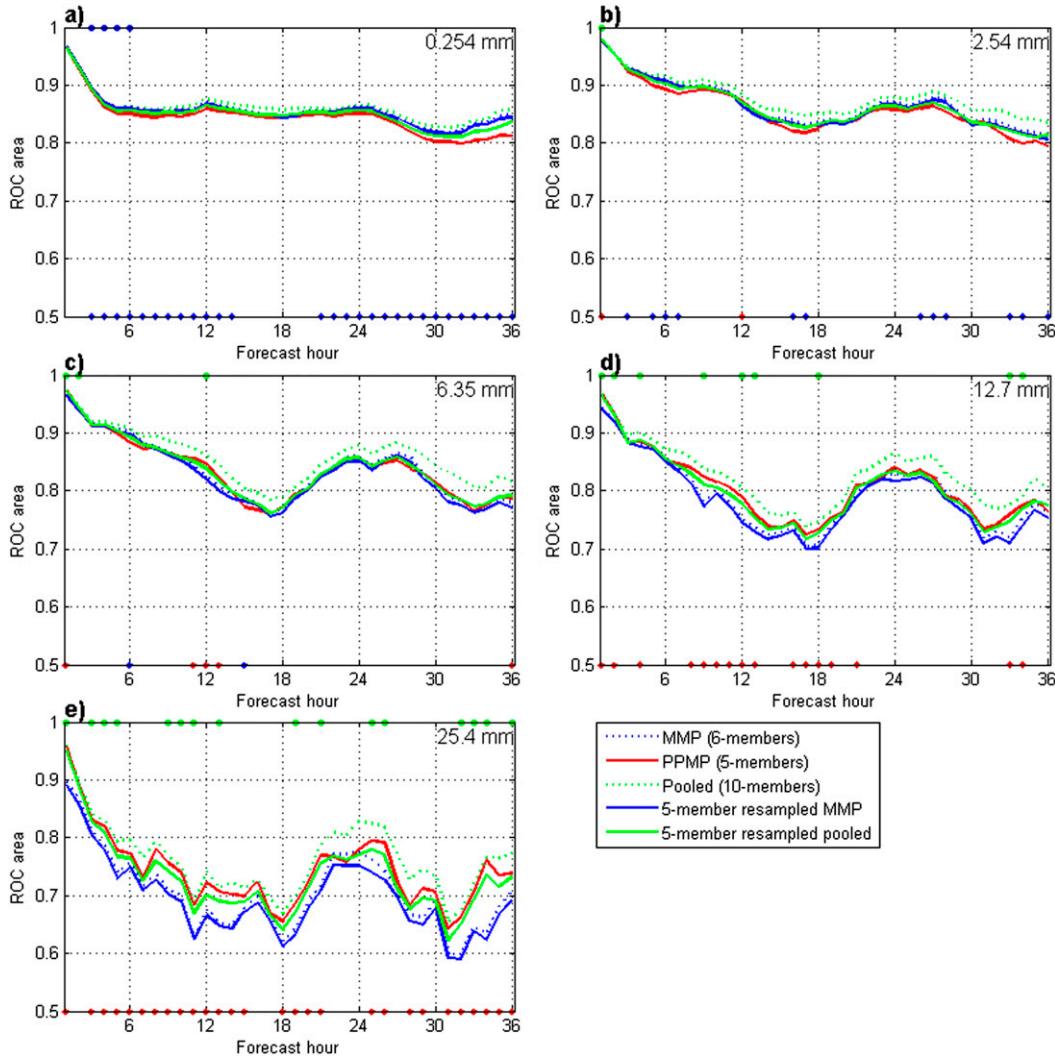


FIG. 10. As in Fig. 9, but for the area under the ROC curve.

microphysics scheme errors in the ensemble, in addition to evaluating the ensembles using objective verification scores, we examined the systematic behaviors of various microphysical variables from the PPMP and MMP ensemble members. The following analysis is meant to answer the following questions: How did the prescribed intercept parameters for rain and graupel in the WSM6 schemes compare to those retrieved from the double-moment microphysics schemes? How different are the microphysics parameters between single and double-moment schemes? How did the WDM6 and WSM6 schemes in particular compare?

Using the method discussed in the appendix, various parameters in these MP schemes were retrieved from the forecast moments' fields. Parameters were not retrieved from the Ferrier+ scheme due to significant differences between that scheme and the others regarding the

treatment of the water species. Numerous additional assumptions regarding the rain and graupel PSDs would be required to retrieve parameters from the Ferrier+ scheme. Since the graupel intercept parameter is diagnosed from the mixing ratio internally in the Thompson scheme, the graupel intercept parameter was not retrieved from the Thompson scheme. We chose specifically to examine the mean mass particle diameter, particle surface area, number concentration, mixing ratio, and intercept parameter for the rain and graupel PSDs. The former two parameters serve as proxies for the tendencies for rain evaporation and cold pool formation (Dawson et al. 2010), which can have a large impact on the organization and lifetime of convective systems (Rotunno et al. 1988; Snook and Xue 2008).

Vertical profiles of horizontally averaged microphysical parameters for rain are shown in Fig. 11, while those

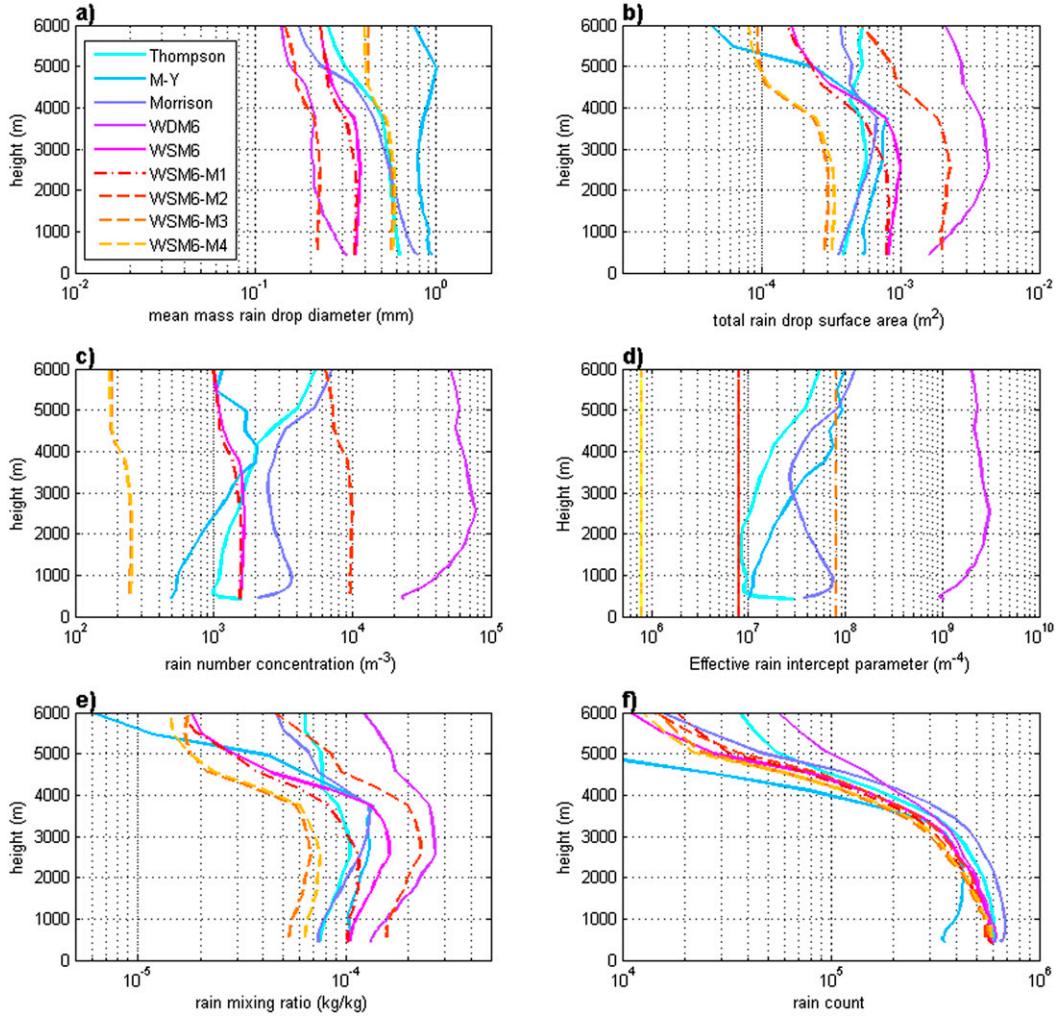


FIG. 11. Vertical profiles of area-average microphysics parameters: (a) mean mass raindrop diameter (mm), (b) raindrop surface area (m^2), (c) rain number concentration (drops per cubic meter), (d) normalized rain intercept parameter (m^{-4}), (e) rain mixing ratio ($kg\ kg^{-1}$), and (f) number of grid points at which the rain mixing ratio exceeded $10^{-6}\ kg\ kg^{-1}$.

for graupel are shown in Fig. 12. We only consider rain and graupel since the PPMP ensemble only perturbed the parameters associated with these two species. The average was computed over only those grid boxes at which the mixing ratio exceeded $10^{-6}\ kg\ kg^{-1}$, a general threshold used to distinguish precipitating from non-precipitating grid boxes (the counts can be seen in Figs. 11f and 12f). The profiles were also averaged over all forecast hours. For rain the discussion will focus generally below 3 km, a common height of the freezing level during the experiment, where rainfall is the dominant precipitation type. On the other hand, the discussion for graupel will focus on the lowest 10 km.

As pointed out earlier, the prescribed values of intercept parameters in the PPMP members were chosen

based on values reported in the literature (Snook and Xue 2008; Tong and Xue 2008; Yussouf and Stensrud 2012). Figure 11d shows that the effective rain intercept parameters sampled by the MMP members generally laid in a similar range as those prescribed to the members of the PPMP scheme, although that for the WDM6 scheme is anomalously high, and the $8 \times 10^5\ m^{-4}$ value used for the PPMP members seems somewhat low. Other microphysical variables such as raindrop size, total raindrop surface area, rain number concentration, and mixing ratio sampled by the MMP members also in general laid in the same range as those in the PPMP ensemble (Fig. 11). The Morrison scheme was the only scheme to give reliable results for the graupel intercept parameter (Fig. 12d), and the values from that scheme line up nicely

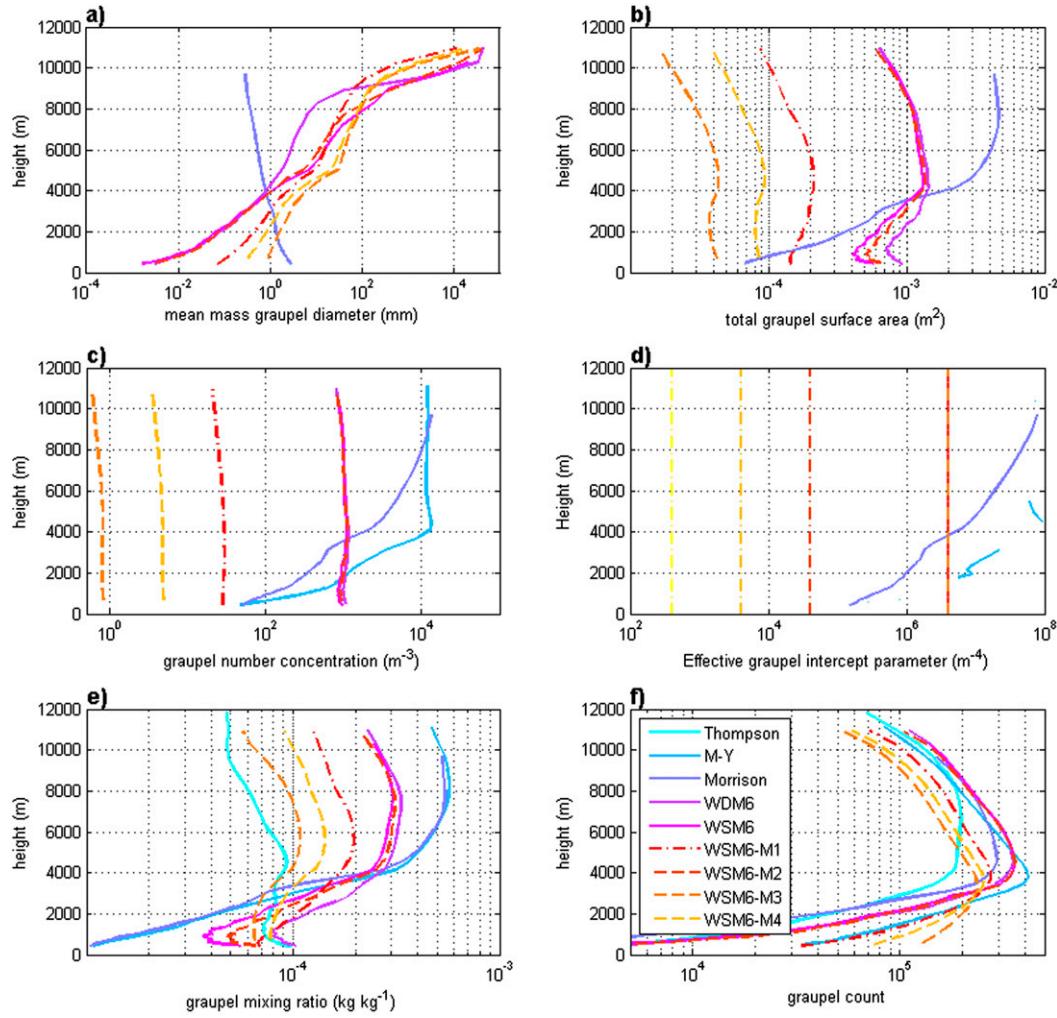


FIG. 12. As in Fig. 11, but for graupel.

with the values from the unperturbed WSM6 scheme ($4 \times 10^6 \text{ m}^{-4}$) except for the vertical variation of nearly three orders of magnitude. Except for the Morrison scheme, which exhibited a different pattern of behavior than the other schemes, the WSM6 schemes in the PPMP ensemble tended to produce less of the various graupel quantities examined, including surface area, number concentration, and mixing ratio (Fig. 12). The Thompson scheme produced the least graupel mixing ratio, however (Fig. 12e). The normalized graupel intercept parameters retrieved from the Morrison scheme most closely resemble those of the control value of $4 \times 10^6 \text{ m}^{-4}$ (Fig. 12d), suggesting that the perturbed values of the graupel intercept parameters may be too low.

Among the single-moment members that composed the PPMP ensemble, the behavior was found to be consistent with the prescribed intercept parameters. WSM6-M2 was designed to represent small raindrops. For a given

mixing ratio, mean drop size decreases with an increasing intercept parameter. WSM6-M2 was the only scheme for which the rain intercept parameter was perturbed high compared to the control value (Table 1). While the rain mixing ratios are not identical between the single-moment members, differences of approximately one to two orders of magnitude are required for this rule to be violated. That was not the case at any level or time in this experiment (Fig. 11e). Similarly, the larger raindrop sizes for WSM6-M3 and WSM6-M4 compared to those of the WSM6-M2 and WSM6 (control) schemes are also consistent given that the rain intercept parameters for WSM6-M3 and WSM6-M4 are lower than they are for WSM6 (control) and WSM6-M2. Given the similarity in mixing ratio, it is also sensible that WSM6-M2 also had the largest number concentration of rain and the largest total raindrop surface area, while WSM6-M3 and WSM6-M4 had smaller number concentrations and surface areas. The vertical profiles

of rain number concentration and raindrop surface area were very similar between pairs of members (e.g., M3 compared to M4 and control compared to M1) that used the same set of intercept values because these parameters are derived from the only free moment, the mixing ratio. Slight differences in the mixing ratios correspond to slight differences in these other retrieved parameters. Similar arguments apply to the retrieved graupel variables. WSM6-M3 and WSM6-M4 were meant to contain large hail-like particles. Therefore, it makes sense that these schemes have the largest graupel (Fig. 12a), smallest number concentration (Fig. 12c), and smallest surface area (Fig. 12b). Similarly, WSM6 and WSM6-M2 contained smaller and more numerous graupel particles. Interestingly, the graupel mixing ratios above about 3 km were ranked according to prescribed graupel intercept parameters for the WSM6 schemes.

The profiles from the double-moment schemes contained more variation in the vertical than those from the single-moment scheme members (Figs. 11 and 12). This behavior is likely the result of the ability of the double-moment schemes to account for size sorting of rain and graupel and other processes causing size distribution to change, since they contain separate predictive equations for the additional moment of number concentration (see, e.g., Reisner et al. 1998; Milbrandt and Yau 2005; Morrison et al. 2005; Lim and Hong 2010). The overall increase in mean mass diameter of raindrops toward the surface in the double-moment schemes is an indication that smaller raindrops are evaporating during their descent from the midlevels; some completely evaporate before reaching the ground. This is consistent with the decrease in the number concentration and surface area toward the surface, although there is some anomalous behavior seen in the Thompson scheme, where the number concentration and intercept parameter increase in the lowest two to three model levels (Figs. 11c,d). The rain mixing ratio generally decreases from a height of about 3 km toward the ground. The general decrease supports the notion that rain is evaporating in unsaturated down-drafts or due to entrainment of unsaturated environmental air. The graupel mixing ratio and number concentration also decrease below about 3 km (except for Thompson; see Fig. 12e), as graupel melts below the freezing level. The Morrison scheme is the only one that predicts the number concentration for graupel and that gave reliable results. In the Morrison scheme, only larger particles are left below the freezing level as the smaller particles completely melt (the lavender trace in Fig. 12a). Among the double-moment schemes considered, WDM6 produced the smallest raindrops and very large rain number concentrations compared to the other schemes. This behavior of WDM6 has been noted and modifications to the

scheme to reduce the number concentration are under way by the scheme developers (K.-S. Lim 2012, personal communication). The MY scheme tended to produce larger raindrops and a larger rain surface area than did the Morrison scheme. This is opposite of what was found in Morrison and Milbrandt (2011) regarding comparisons between those schemes. However, the versions of the schemes used in that study are slightly different from those used in this study. Namely, the shape parameters of the DSD differ [$\alpha_r = 0$ was used in the MY scheme of Morrison and Milbrandt (2011), whereas $\alpha_r = 2$ was used in this study]. However, the graupel parameters examined in this study are identical to those in Morrison and Milbrandt (2011), and the similar mixing ratios between the MY and Morrison schemes in this study disagree with the results of Morrison and Milbrandt (2011), which found that the MY scheme tended to produce more graupel. Also, above 3 km, the MY and Morrison schemes produced a higher graupel mixing ratio (Fig. 12e) than the single-moment schemes for graupel.

The design of the WDM6 and WSM6 schemes makes them more similar to each other than any other pair of schemes in this study (not including the WSM6 schemes with perturbed parameters). The major differences between the WDM6 and WSM6 schemes include the number of moments predicted for rain, the shape parameter of the rain DSD, and the addition of cloud condensation nuclei as a prognostic variable in the WDM6 scheme. There are also minor differences in how some micro-physics processes like accretion rates are parameterized. The retrieved rain variables for WDM6 and WSM6 were quite different, however. The WDM6 scheme produced the smallest raindrops overall, corresponding to the largest drop concentration and surface areas as well. It also produced a higher rain mixing ratio and slightly greater radar reflectivity (not shown), but similar domain-accumulated precipitation (also not shown). Since the graupel distributions had the same shape parameter, differences in the graupel variables were reduced compared to those in the rain variables. The WSM6 scheme tended to produce a lower graupel mixing ratio near the surface (Fig. 12e). This is likely related to how some of the graupel and rain processes were parameterized. It is beyond the scope of the study to isolate the specific impact from incorporating the major differences one at a time. However, the results herein are generally in agreement with those obtained in Lim and Hong (2010) in which the same versions of both schemes were compared.

5. Summary and conclusions

Two approaches to accounting for the uncertainty of the MP parameterization in warm-season QPF in a

convection-allowing ensemble were examined. The two approaches include a set of completely different MP parameterizations (MMP) and a set of different numerical values of a set of parameters within a single MP scheme (PPMP). The combination of the two approaches was also tested. These approaches were implemented within the storm-scale ensemble framework of the Center of Analysis and Prediction of Storms at the University of Oklahoma, and tested during the NOAA HWT 2011 Spring Experiment (27 April–10 June 2011). An ensemble of convection-allowing WRF-ARW simulations was run in real time at 4-km resolution over the contiguous United States for 35 cases during the period. The two ensembles tested—MMP and PPMP—along with an ensemble composed of members from both ensembles, had 5, 6, and 10 members, respectively (1 member belonged to both ensembles). To minimize the impact of ensemble size on the comparison between approaches to accounting for microphysics errors, scores for the MMP and pooled ensembles were obtained from an average of 100 five-member resamples.

It was found that, in general, the MMP ensemble was more skillful than the PPMP ensemble (based on WSM6) with variations dependent on the metric chosen, the forecast lead time, and the precipitation threshold. The MMP ensemble was more skillful for FBSs for most lead times and thresholds, but the PPMP ensemble was more skillful for high thresholds when ROC areas were examined. The rank histograms of the MMP ensemble were slightly flatter than those of the PPMP and pooled ensembles. The MMP ensemble was also somewhat more biased, but mostly due to one member. Also, the combined approach where both sophisticated MP schemes and perturbed parameters within a simpler single-moment MP scheme were used was no more skillful than the better of the MMP and PPMP ensembles. Since all PPMP members were based on WSM6, the skill of those members was limited by the skill of WSM6. Therefore, the results obtained here may not apply to perturbed parameter ensembles generated by perturbing other microphysics schemes. The similarity in the performances of the MMP and PPMP ensembles was sensible, however, since the WSM6 scheme frequently ranked in the middle of the microphysics schemes considered and near the middle of the WSM6 perturbation schemes, as shown in section 3a.

The QPF from individual members was also examined. Verification scores varied more among members of the MMP ensemble than among members of the PPMP ensemble for low and moderate thresholds. Large variation was found in the bias during the first and second convection maxima. The WSM6-M3 and WSM6-M4 schemes showed the largest positive bias in the mean errors and in

the frequency biases at the highest two thresholds during the second convection maximum. The MY scheme showed the largest negative bias. PPMP members were generally less biased, but were also less skillful at all but the highest thresholds. The skill of individual members depended on the thresholds and methods of verification. For example, for ETS, the Thompson and Morrison schemes were the most skillful at the lightest two thresholds whereas the WDM6 scheme was the least skillful at low and moderate thresholds at later lead times. For FBS, the Thompson scheme showed the highest skill at moderate thresholds, and both Thompson and Morrison showed the highest skill at the highest threshold. The Ferrier+ and WDM6 schemes were the least skillful at moderate thresholds at the second convection maximum. For ROC areas, the WSM6-M3 and WSM6-M4 schemes were the best at the highest thresholds. The MY scheme was the least skillful at the second convection maximum.

In addition to evaluating the ensembles using objective verification scores, the systematic behaviors of the members using retrieved microphysics parameters relevant to this study were examined. The range of the retrieved rain intercept parameters from the double-moment members is, in general, consistent with the range of perturbed rain intercept parameters for the PPMP members. The retrieved graupel intercept parameter values suggest the perturbed graupel intercept values used in the PPMP members were too low. The behavior of the PPMP members was found to be consistent with the prescribed values of the rain and graupel intercept parameters and graupel density. The pattern of behavior was more variable for the double-moment members, and the causes of differences in retrieved parameters among these MP schemes are more complex than differences among the single-moment members owing to the differences in parameterizations of individual processes within each scheme. The WDM6 scheme tended to produce a very large number of small raindrops, whereas the MY scheme tended to produce very large raindrops.

Although resampling of the pooled ensemble with an equal size as the PPMP and MMP ensembles was no more skillful than the better of the PPMP and MMP ensembles, the full 10-member pooled ensemble showed better skill than other ensembles tested for nearly all lead times, thresholds, and metrics. This result brings up an interesting point in that increasing ensemble size had a significant impact in the skill of the ensembles tested (see also Clark et al. 2011; Ebert 2001). One easy way to populate the ensemble is to perturb the parameters for a given physical parameterization scheme such as the PPMP.

This study represents one step toward developing a strategy for the optimal design of a convection-allowing

ensemble prediction system. Ideally, members of an ensemble should represent random draws from the probability distribution of truth for a given forecast system and each member should be equally possible to represent the truth. The design of the PPMP and MMP ensembles may violate such rules. The parameters used in the PPMP ensemble were not randomly perturbed. The MMP ensemble is composed of different microphysics schemes, with each scheme having its own bias. In addition, only the uncertainty in QPF associated with varied micro-physics was tested, but uncertainty in QPF from model error can also come from errors in other physics parameterizations such as the boundary layer, radiation, and land surface. Since only QPF was examined, it cannot be determined whether either approach is more skillful at predicting other meteorological fields such as surface temperature, relative humidity, wind speed, or geopotential height at other levels in the troposphere. Thus, future studies should also examine the impacts of varied microphysics on other fields. The specific conclusions may also be dependent on the specific choices of the MP schemes used in the ensemble, and to the way parameters are perturbed. Future studies should investigate ways to use variations in these parameterizations to better design convection-allowing model ensembles, as their use will be more strongly desired in the future for operational weather forecasting purposes.

Acknowledgments. This research was primarily supported by NSF Grants AGS-1046081, AGS-0802888, and OCI-0905040. The CAPS storm-scale ensemble members analyzed in this study were also produced under the support of NOAA's CSTAR program (NA17RJ1227). Computing resources at the National Science Foundation XSEDE National Institute of Computational Science (NICS) at the University of Tennessee and those at the OU Supercomputing Center for Education and Research (OSCER) were used for this study. Ming Xue was also supported by NSF Grants AGS-0941491 and AGS-1046171. Dan Dawson is acknowledged for discussions during the initial stages of the work. This manuscript was improved by the suggestions of two anonymous reviewers.

APPENDIX

Micrometeorology Vocabulary

Except for the Ferrier+ scheme, all of the MP schemes used in this experiment use similar assumptions for the PSDs of the various classes. They all use the following model:

$$n_x(D_x) = N_{0x} D_x^{\alpha_x} e^{-\lambda_x D_x}, \quad (\text{A1})$$

where subscript x refers to one of the water species (r , rain; g , graupel; I , cloud ice; s , snow; h , hail; and c , cloud water; the subscript will be dropped hereinafter), n is the volumetric concentration of spherical particles of diameter D to $D + \Delta D$, N_0 is the intercept parameter, λ is the slope parameter, and α is the shape parameter. Note that when $\alpha = 0$, this reduces to the inverse-exponential distribution. The p th moment of the PSD is given by

$$M(p) = \int_0^\infty D^p n(D) dD = \frac{N_0}{\lambda^{\alpha+p+1}} \Gamma(\alpha + p + 1), \quad (\text{A2})$$

where Γ is the gamma function and $\Gamma(Z) = (Z - 1)!$ when Z is an integer. Three moments would be required to fully characterize this PSD without making any assumptions. Typically, the three moments used include the mixing ratio, which is proportional to the third moment of the PSD; the number concentration, which is the zeroth moment of the PSD; and the reflectivity factor, which is the sixth moment, for each species. Each moment requires its own set of conservation equations for each species, which increases the complexity of the scheme and also increases the computation time. However, use of a greater number of moments has been shown to increase the realism of simulations of squall lines at convection-allowing resolutions (Morrison et al. 2009). Four schemes in this study—Thompson, Morrison, MY, and WDM6—use two moments for some species. All are double moment for rain, and two (Morrison and MY) are double moment for graupel as well. Species mixing ratio and number concentration are the two moments used in these schemes. While the other schemes are only single moment, other moments can be derived using manipulations of (A1) and (A2). For instance, the number concentration can be determined by

$$N_t = \left\{ [N_0 \Gamma(\alpha + 1)]^d \left[\frac{q \rho_a}{c} \frac{\Gamma(\alpha + 1)}{\Gamma(\alpha + 4)} \right]^{\alpha+1} \right\}^{1/(\alpha+d+1)}, \quad (\text{A3})$$

where ρ_a is air density and c and d are related through the mass–diameter relationship, where $m(D) = cD^d$ and q is the species mixing ratio. In this paper we retrieved the intercept parameter for the double-moment schemes using (A3). However, the meaning of the intercept parameter becomes nonphysical when the shape parameter is nonzero. Therefore, the “normalized” intercept parameter, which is the intercept parameter for an exponential distribution with the same liquid water content and mean volume diameter (Testud et al. 2001), is used

instead. The total surface area of particles is proportional to $M(2)$:

$$\text{Surface area} = \pi M(2) = \frac{\pi N_t(\alpha + 2)(\alpha + 1)D_m^2}{\left[\frac{\Gamma(\alpha + 4)}{\Gamma(\alpha + 1)}\right]^{2/3}}, \quad (\text{A4})$$

where it has been assumed that α is an integer and $d = 3$, which is true for both rain and graupel, and D_m is the mean mass diameter, the size that splits the PSD into two equal halves, is given by

$$\bar{D} = \left(\frac{\rho_a}{c_x} \frac{q}{N_t} \right)^{1/3}. \quad (\text{A5})$$

REFERENCES

- Baldwin, M. E., S. Lakshmivarahan, and J. S. Kain, 2001: Verification of mesoscale features in NWP models. Preprints, *Ninth Conf. on Mesoscale Processes*, Ft. Lauderdale, FL, Amer. Meteor. Soc., 8.3. [Available online at <https://ams.confex.com/ams/pdpapers/23364.pdf>.]
- Berner, J., G. Shutts, M. Leutbecher, and T. Palmer, 2009: A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. *J. Atmos. Sci.*, **66**, 603–626, doi:[10.1175/2008JAS2677.1](https://doi.org/10.1175/2008JAS2677.1).
- , S.-Y. Ha, J. P. Hacker, A. Fournier, and C. Snyder, 2011: Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations. *Mon. Wea. Rev.*, **139**, 1972–1995, doi:[10.1175/2010MWR3595.1](https://doi.org/10.1175/2010MWR3595.1).
- Buizza, R., and T. N. Palmer, 1995: The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.*, **52**, 1434–1456, doi:[10.1175/1520-0469\(1995\)052<1434:TSVSOT>2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052<1434:TSVSOT>2.0.CO;2).
- , M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125B**, 2887–2908, doi:[10.1002/qj.49712556006](https://doi.org/10.1002/qj.49712556006).
- Candille, G., 2009: The multiensemble approach: The NAEFS example. *Mon. Wea. Rev.*, **137**, 1655–1665, doi:[10.1175/2008MWR2682.1](https://doi.org/10.1175/2008MWR2682.1).
- Caron, J.-F., 2013: Mismatching perturbations at the lateral boundaries in limited-area ensemble forecasting: A case study. *Mon. Wea. Rev.*, **141**, 356–374, doi:[10.1175/MWR-D-12-00051.1](https://doi.org/10.1175/MWR-D-12-00051.1).
- Charron, M., G. Pellerin, L. Spacel, P. L. Houtekamer, N. Gagnon, H. L. Mitchell, and L. Michelin, 2010: Toward random sampling of model error in the Canadian ensemble prediction system. *Mon. Wea. Rev.*, **138**, 1877–1901, doi:[10.1175/2009MWR3187.1](https://doi.org/10.1175/2009MWR3187.1).
- Clark, A. J., W. A. Gallus Jr., and T. C. Chen, 2008: Contributions of mixed physics versus perturbed initial/lateral boundary conditions to ensemble-based precipitation forecast skill. *Mon. Wea. Rev.*, **136**, 2140–2156, doi:[10.1175/2007MWR2029.1](https://doi.org/10.1175/2007MWR2029.1).
- , —, M. Xue, and F. Kong, 2009: A comparison of precipitation forecast skill between small convection-allowing and large convection parameterizing ensembles. *Wea. Forecasting*, **24**, 1121–1140, doi:[10.1175/2009WAF222222.1](https://doi.org/10.1175/2009WAF222222.1).
- , W. A. Gallus, M. Xue, and F. Kong, 2010: Growth of spread in convection-allowing and convection-parameterizing ensembles. *Wea. Forecasting*, **25**, 594–612, doi:[10.1175/2009WAF2222318.1](https://doi.org/10.1175/2009WAF2222318.1).
- , and Coauthors, 2011: Probabilistic precipitation forecast skill as a function of ensemble size and spatial scale in a convection-allowing ensemble. *Mon. Wea. Rev.*, **139**, 1410–1418, doi:[10.1175/2010MWR3624.1](https://doi.org/10.1175/2010MWR3624.1).
- , and Coauthors, 2012: An overview of the 2010 Hazardous Weather Testbed Experimental Forecast Program Spring Experiment. *Bull. Amer. Meteor. Soc.*, **93**, 55–74, doi:[10.1175/BAMS-D-11-00040.1](https://doi.org/10.1175/BAMS-D-11-00040.1).
- Dawson, D. T., M. Xue, J. A. Milbrandt, and M. K. Yau, 2010: Comparison of evaporation and cold pool development between single-moment and multimoment bulk microphysics schemes in idealized simulations of tornadic thunderstorms. *Mon. Wea. Rev.*, **138**, 1152–1171, doi:[10.1175/2009MWR2956.1](https://doi.org/10.1175/2009MWR2956.1).
- Doblas-Reyes, F. J., M. Déqué, and J.-P. Piedelievre, 2000: Multi-model spread and probabilistic seasonal forecasts in PROVOST. *Quart. J. Roy. Meteor. Soc.*, **126**, 2069–2807, doi:[10.1002/qj.49712656705](https://doi.org/10.1002/qj.49712656705).
- Du, J., S. L. Millen, and F. Sanders, 1997: Short-range ensemble forecasting of quantitative precipitation. *Mon. Wea. Rev.*, **125**, 2427–2459, doi:[10.1175/1520-0493\(1997\)125<2427:SREFOQ>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<2427:SREFOQ>2.0.CO;2).
- Ebert, E. E., 2001: Ability of a poor man's ensemble to predict the probability and distribution of precipitation. *Mon. Wea. Rev.*, **129**, 2461–2480, doi:[10.1175/1520-0493\(2001\)129<2461:AOPAMS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2461:AOPAMS>2.0.CO;2).
- Eckel, F. A., and C. F. Mass, 2005: Aspects of effective mesoscale, short-range ensemble forecasting. *Wea. Forecasting*, **20**, 328–350, doi:[10.1175/WAF843.1](https://doi.org/10.1175/WAF843.1).
- Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarplay, 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta Model. *J. Geophys. Res.*, **108**, 8851, doi:[10.1029/2002JD003296](https://doi.org/10.1029/2002JD003296).
- Ferrier, B. S., cited 2013: The cloud and precipitation scheme in the operational NCEP (regional) models: Description and system integration issues. [Available online at http://www.emc.ncep.noaa.gov/mmb/bf/presentations/Stony_Brook_3-1-05.ppt.]
- , Y. Jin, Y. Lin, T. Black, E. Rogers, and G. DiMego, 2002: Implementation of a new grid-scale cloud and precipitation scheme in an NCEP eta model. Preprints, *19th Conf. on Weather Analysis and Forecasting/15th Conf. on Numerical Weather Prediction*, San Antonio, TX, Amer. Meteor. Soc., 10.1. [Available online at <https://ams.confex.com/ams/pdpapers/47241.pdf>.]
- , and Coauthors, cited 2013: Changes to the NCEP meso eta analysis and forecast system: Modified cloud micro-physics, assimilation of GOES cloud-top pressure, and assimilation of NEXRAD 88D radial wind velocity data. NCEP Technical Procedures Bull. [Available online at <http://www.emc.ncep.noaa.gov/mmb/tpb.spring03/tpb.htm>.]
- Gao, J.-D., M. Xue, K. Brewster, and K. K. Droegemeier, 2004: A three-dimensional variational data analysis method with recursive filter for Doppler radars. *J. Atmos. Oceanic Technol.*, **21**, 457–469, doi:[10.1175/1520-0426\(2004\)021<0457:ATVDAM>2.0.CO;2](https://doi.org/10.1175/1520-0426(2004)021<0457:ATVDAM>2.0.CO;2).
- Gilmore, M. S., J. M. Straka, and E. N. Rasmussen, 2004: Precipitation uncertainty due to variations in precipitation particle parameters within a simple microphysics scheme. *Mon. Wea. Rev.*, **132**, 2610–2627, doi:[10.1175/MWR2810.1](https://doi.org/10.1175/MWR2810.1).
- Hacker, J. P., and Coauthors, 2011: The U.S. Air Force Weather Agency's mesoscale ensemble: Scientific description and performance results. *Tellus*, **63A**, 1–17, doi:[10.1111/j.1600-0870.2010.00497.x](https://doi.org/10.1111/j.1600-0870.2010.00497.x).

- Hagedorn, R., F. J. Doblas-Reyes, and T. N. Palmer, 2005: The rationale behind the success of multi-model ensembles in seasonal forecasting—I. Basic concept. *Tellus*, **57A**, 219–233, doi:[10.1111/j.1600-0870.2005.00103.x](https://doi.org/10.1111/j.1600-0870.2005.00103.x).
- Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, **129**, 550–560, doi:[10.1175/1520-0493\(2001\)129<0550:IORHFV>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2).
- , and S. J. Colucci, 1997: Verification of Eta-RSM short range ensemble forecasts. *Mon. Wea. Rev.*, **125**, 1312–1327, doi:[10.1175/1520-0493\(1997\)125<1312:VOERSR>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<1312:VOERSR>2.0.CO;2).
- Hohenegger, C., and C. Schär, 2007: Atmospheric predictability at synoptic versus cloud-resolving scales. *Bull. Amer. Meteor. Soc.*, **88**, 1783–1793, doi:[10.1175/BAMS-88-11-1783](https://doi.org/10.1175/BAMS-88-11-1783).
- Hong, S.-Y., and J.-O. J. Lim, 2006: The WRF single-moment 6-class microphysics scheme (WSM6). *J. Korean Meteor. Soc.*, **42**, 129–151.
- Hou, D., E. Kalnay, and K. K. Droegemeier, 2001: Objective verification of the SAMEX '98 ensemble forecasts. *Mon. Wea. Rev.*, **129**, 73–91, doi:[10.1175/1520-0493\(2001\)129<0073:OVOTSE>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0073:OVOTSE>2.0.CO;2).
- Houtekamer, P. L., L. Lefavre, J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**, 1225–1242, doi:[10.1175/1520-0493\(1996\)124<1225:ASSATE>2.0.CO;2](https://doi.org/10.1175/1520-0493(1996)124<1225:ASSATE>2.0.CO;2).
- Hu, M., M. Xue, and K. Brewster, 2006: 3DVAR and cloud analysis with WSR-88D level-II data for the prediction of Fort Worth tornadic thunderstorms. Part I: Cloud analysis and its impact. *Mon. Wea. Rev.*, **134**, 675–698, doi:[10.1175/MWR3092.1](https://doi.org/10.1175/MWR3092.1).
- Janjić, Z., 2002: Nonsingular implementation of the Mellor–Yamada_level 2.5 scheme in the NCEP Meso model. NCEP Office Note 437, 61 pp. [Available online at <http://www.emc.ncep.noaa.gov/officenotes/newernotes/on437.pdf>.]
- Johnson, A., and X. Wang, 2012: Verification and calibration of neighborhood and object-based probabilistic precipitation forecasts from a multi-model convection-allowing ensemble. *Mon. Wea. Rev.*, **140**, 3054–3077, doi:[10.1175/MWR-D-11-00356.1](https://doi.org/10.1175/MWR-D-11-00356.1).
- , and —, 2013: Object-based evaluation of a storm scale ensemble during the 2009 NOAA Hazardous Weather Testbed Spring Experiment. *Mon. Wea. Rev.*, **141**, 1079–1098, doi:[10.1175/MWR-D-12-00140.1](https://doi.org/10.1175/MWR-D-12-00140.1).
- , —, —, and —, 2011a: Hierarchical cluster analysis of a convection-allowing ensemble during the Hazardous Weather Testbed 2009 Spring Experiment. Part I: Development of the object-oriented cluster analysis method for precipitation fields. *Mon. Wea. Rev.*, **139**, 3673–3693, doi:[10.1175/MWR-D-11-00015.1](https://doi.org/10.1175/MWR-D-11-00015.1).
- , —, —, and —, 2011b: Hierarchical cluster analysis of a convection-allowing ensemble during the Hazardous Weather Testbed 2009 Spring Experiment. Part II: Ensemble clustering over the whole experiment period. *Mon. Wea. Rev.*, **139**, 3694–3710, doi:[10.1175/MWR-D-11-00016.1](https://doi.org/10.1175/MWR-D-11-00016.1).
- , —, —, and —, 2013: Object-based evaluation of the impact of horizontal grid spacing on convection-allowing forecasts. *Mon. Wea. Rev.*, **141**, 3413–3425, doi:[10.1175/MWR-D-13-00027.1](https://doi.org/10.1175/MWR-D-13-00027.1).
- , and Coauthors, 2014: Multiscale characteristics and evolution of perturbations for warm season convection-allowing precipitation forecasts: Dependence on background flow and method of perturbation. *Mon. Wea. Rev.*, **142**, 1053–1073, doi:[10.1175/MWR-D-13-00204.1](https://doi.org/10.1175/MWR-D-13-00204.1).
- Jung, Y., M. Xue, and G. Zhang, 2010: Simulations of polarimetric radar signatures of a supercell storm using a two-moment bulk microphysics scheme. *J. Appl. Meteor. Climatol.*, **49**, 146–163, doi:[10.1175/2009JAMC2178.1](https://doi.org/10.1175/2009JAMC2178.1).
- Kong, F., K. K. Droegemeier, and N. L. Hickmon, 2006: Multi-resolution ensemble forecasts of an observed tornadic thunderstorm system. Part I: Comparison of coarse and fine-grid experiments. *Mon. Wea. Rev.*, **134**, 807–833, doi:[10.1175/MWR3097.1](https://doi.org/10.1175/MWR3097.1).
- , —, and —, 2007a: Multiresolution ensemble forecasts of an observed tornadic thunderstorm system. Part II: Storm-scale experiments. *Mon. Wea. Rev.*, **135**, 759–782, doi:[10.1175/MWR3323.1](https://doi.org/10.1175/MWR3323.1).
- , —, and Coauthors, 2007b: Preliminary analysis on the real-time storm-scale ensemble forecasts produced as a part of the NOAA Hazardous Weather Testbed 2007 Spring Experiment. *22nd Conf. on Weather Analysis and Forecasting/18th Conf. on Numerical Weather Prediction*, Salt Lake City, UT, Amer. Meteor. Soc., 3B.2. [Available online at <https://ams.confex.com/ams/pdffiles/124667.pdf>.]
- , —, and Coauthors, 2010: Evaluation of CAPS multi-model storm-scale ensemble forecast for the NOAA HWT 2010 Spring Experiment. *25th Conf. on Severe Local Storms*, Denver, CO, Amer. Meteor. Soc., P4.18. [Available online at <https://ams.confex.com/ams/pdffiles/175822.pdf>.]
- , —, and Coauthors, 2011: Storm-scale ensemble forecasting for the NOAA hazardous weather testbed. *Extended Abstracts, Sixth European Conf. on Severe Storms*, Palma de Mallorca, Spain, European Severe Storms Laboratory. [Available online at <http://www.essl.org/ECSS/2011/programme/abstracts/171.pdf>.]
- Lim, K.-S. S., and S.-Y. Hong, 2010: Development of an effective double-moment cloud microphysics scheme with prognostic cloud condensation nuclei (CCN) for weather and climate models. *Mon. Wea. Rev.*, **138**, 1587–1612, doi:[10.1175/2009MWR2968.1](https://doi.org/10.1175/2009MWR2968.1).
- Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.*, **30**, 291–303.
- Mellor, G. L., and T. Yamada, 1982: Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys. Space Phys.*, **20**, 851–875, doi:[10.1029/RG020i004p00851](https://doi.org/10.1029/RG020i004p00851).
- Milbrandt, J. A., and M. K. Yau, 2005: A multimoment bulk microphysics parameterization. Part II: A proposed three-moment closure and scheme description. *J. Atmos. Sci.*, **62**, 3065–3081, doi:[10.1175/JAS3535.1](https://doi.org/10.1175/JAS3535.1).
- Mittermaier, M. P., 2007: Improving short-range high-resolution model precipitation forecast skill using time-lagged ensembles. *Quart. J. Roy. Meteor. Soc.*, **133**, 1487–1500, doi:[10.1002/qj.135](https://doi.org/10.1002/qj.135).
- Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF Ensemble Prediction System: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119, doi:[10.1002/qj.49712252905](https://doi.org/10.1002/qj.49712252905).
- Morrison, H., and J. Milbrandt, 2011: Comparison of two-moment bulk microphysics schemes in idealized supercell thunderstorm simulations. *Mon. Wea. Rev.*, **139**, 1103–1130, doi:[10.1175/2010MWR3433.1](https://doi.org/10.1175/2010MWR3433.1).
- , G. Thompson, and V. Tatarkii, 2009: Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes. *Mon. Wea. Rev.*, **137**, 991–1007, doi:[10.1175/2008MWR2556.1](https://doi.org/10.1175/2008MWR2556.1).
- , J. A. Curry, and V. I. Khvorostyanov, 2005: A new double-moment microphysics parameterization for application in cloud and climate models. Part I: Description. *J. Atmos. Sci.*, **62**, 1665–1677, doi:[10.1175/JAS3446.1](https://doi.org/10.1175/JAS3446.1).

- Palmer, T. N., G. J. Shutts, R. Hagedorn, F. J. Doblas-Reyes, T. Jung, and M. Leutbecher, 2005: Representing model uncertainty in weather and climate prediction. *Annu. Rev. Earth Planet. Sci.*, **33**, 163–193, doi:[10.1146/annurev.earth.33.092203.122552](https://doi.org/10.1146/annurev.earth.33.092203.122552).
- Putnam, B. J., M. Xue, Y. Jung, N. A. Snook, and G. Zhang, 2013: The analysis and prediction of microphysical states and polarimetric variables in a mesoscale convective system using double-moment microphysics, multinetwork radar data, and the ensemble Kalman filter. *Mon. Wea. Rev.*, **142**, 141–162, doi:[10.1175/MWR-D-13-00042.1](https://doi.org/10.1175/MWR-D-13-00042.1).
- Ralph, F. M., and Coauthors, 2005: Improving short-term (0–48 h) cool-season quantitative precipitation forecasts: Recommendations from a USWRP workshop. *Bull. Amer. Meteor. Soc.*, **86**, 1619–1632, doi:[10.1175/BAMS-86-11-1619](https://doi.org/10.1175/BAMS-86-11-1619).
- Reisner, J., R. M. Rasmussen, and R. T. Bruintjes, 1998: Explicit forecasting of supercooled liquid water in winter storms using the MM5 mesoscale model. *Quart. J. Roy. Meteor. Soc.*, **124B**, 1071–1107, doi:[10.1002/qj.49712454804](https://doi.org/10.1002/qj.49712454804).
- Richardson, D. S., 2001: Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size. *Quart. J. Roy. Meteor. Soc.*, **127**, 2473–2489, doi:[10.1002/qj.49712757715](https://doi.org/10.1002/qj.49712757715).
- Roberts, N. M., and H. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, doi:[10.1175/2007MWR2123.1](https://doi.org/10.1175/2007MWR2123.1).
- Ropnick, A., A. Hense, C. Gebhardt, and D. Majewski, 2013: Bayesian model verification of NWP ensemble forecasts. *Mon. Wea. Rev.*, **141**, 375–387, doi:[10.1175/MWR-D-11-00350.1](https://doi.org/10.1175/MWR-D-11-00350.1).
- Rotunno, R., J. B. Klemp, and M. L. Weisman, 1988: A theory for strong, long-lived squall lines. *J. Atmos. Sci.*, **45**, 463–485, doi:[10.1175/1520-0469\(1988\)045<0463:ATFSLL>2.0.CO;2](https://doi.org/10.1175/1520-0469(1988)045<0463:ATFSLL>2.0.CO;2).
- 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, doi:[10.1175/2009WAF2222671](https://doi.org/10.1175/2009WAF2222671).
- Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp. [Available online at http://www.mmm.ucar.edu/wrf/users/docs/arw_v3.pdf.]
- Snook, N., and M. Xue, 2008: Effects of microphysical drop size distribution on tornadogenesis in supercell thunderstorms. *Geophys. Res. Lett.*, **35**, L24803, doi:[10.1029/2008GL035866](https://doi.org/10.1029/2008GL035866).
- Stensrud, D. J., J.-W. Bao, and T. T. Warner, 2000: Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Wea. Rev.*, **128**, 2077–2107, doi:[10.1175/1520-0493\(2000\)128<2077:UICAMP>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<2077:UICAMP>2.0.CO;2).
- Testud, J., S. Oury, R. A. Black, P. Amayenc, and X. Dou, 2001: The concept of “normalized” distribution to describe raindrop spectra: A tool for cloud physics and cloud remote sensing. *J. Appl. Meteor.*, **40**, 1118–1140, doi:[10.1175/1520-0450\(2001\)040<1118:TCNDT>2.0.CO;2](https://doi.org/10.1175/1520-0450(2001)040<1118:TCNDT>2.0.CO;2).
- Theis, S. E., A. Hense, and U. Damrath, 2005: Probabilistic precipitation forecasts from a deterministic model: A pragmatic approach. *Meteor. Appl.*, **12**, 257–268, doi:[10.1017/S1350482705001763](https://doi.org/10.1017/S1350482705001763).
- 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, doi:[10.1175/2008MWR2387.1](https://doi.org/10.1175/2008MWR2387.1).
- Tong, M., and M. Xue, 2008: Simultaneous estimation of microphysical parameters and atmospheric state with simulated radar data and ensemble square root Kalman filter. Part I: Sensitivity analysis and parameter identifiability. *Mon. Wea. Rev.*, **136**, 1630–1648, doi:[10.1175/2007MWR2070.1](https://doi.org/10.1175/2007MWR2070.1).
- Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330, doi:[10.1175/1520-0477\(1993\)074<2317:EFANTG>2.0.CO;2](https://doi.org/10.1175/1520-0477(1993)074<2317:EFANTG>2.0.CO;2).
- , and —, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319, doi:[10.1175/1520-0493\(1997\)125<3297:EFANAT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<3297:EFANAT>2.0.CO;2).
- Vie, B., O. Nuissier, and V. Ducrocq, 2011: Cloud-resolving ensemble simulations of Mediterranean heavy precipitating events: Uncertainty on initial conditions and lateral boundary conditions. *Mon. Wea. Rev.*, **139**, 403–423, doi:[10.1175/2010MWR3487.1](https://doi.org/10.1175/2010MWR3487.1).
- Wang, X., and C. H. Bishop, 2003: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.*, **60**, 1140–1158, doi:[10.1175/1520-0469\(2003\)060<1140:ACOBAE>2.0.CO;2](https://doi.org/10.1175/1520-0469(2003)060<1140:ACOBAE>2.0.CO;2).
- , and —, 2005: Improvement of ensemble reliability with a new dressing kernel. *Quart. J. Roy. Meteor. Soc.*, **131**, 965–986, doi:[10.1256/qj.04.120](https://doi.org/10.1256/qj.04.120).
- , —, and Simon J. Julier, 2004: Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble? *Mon. Wea. Rev.*, **132**, 1590–1605, doi:[10.1175/1520-0493\(2004\)132<1590:WIBAEQ>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1590:WIBAEQ>2.0.CO;2).
- Wei, M., Z. Toth, R. Wobus, and Y. Zhu, 2008: Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system. *Tellus*, **60A**, 62–79.
- Xue, M., D.-H. Wang, J.-D. Gao, K. Brewster, and K. K. Droegemeier, 2003: The Advanced Regional Prediction System (ARPS), storm-scale numerical weather prediction and data assimilation. *Meteor. Atmos. Phys.*, **82**, 139–170, doi:[10.1007/s00703-001-0595-6](https://doi.org/10.1007/s00703-001-0595-6).
- , and Coauthors, 2007: CAPS realtime storm-scale ensemble and high-resolution forecasts as part of the NOAA Hazardous Weather Testbed 2007 Spring Experiment. *22nd Conf. on Weather Analysis and Forecasting/18th Conf. on Numerical Weather Prediction*, Park City, UT, Amer. Meteor. Soc., 3B.1. [Available online at <https://ams.confex.com/ams/pdffiles/124587.pdf>.]
- , and Coauthors, 2008: CAPS realtime storm-scale ensemble and high-resolution forecasts as part of the NOAA Hazardous Weather Testbed 2008 Spring Experiment. *24th Conf. Several Local Storms*, Savannah, GA, Amer. Meteor. Soc., 12.2. [Available online at <https://ams.confex.com/ams/pdffiles/142036.pdf>.]
- , and Coauthors, 2009: CAPS realtime multi-model convection-allowing ensemble and 1-km convection-resolving forecasts for the NOAA Hazardous Weather Testbed 2009 Spring Experiment. *23rd Conf. on Weather Analysis and Forecasting/19th Conf. on Numerical Weather Prediction*, Omaha, NE, Amer. Meteor. Soc., 16A.2. [Available online at <https://ams.confex.com/ams/pdffiles/154323.pdf>.]
- , and Coauthors, 2011: Real-time convection-permitting ensemble and convection-resolving deterministic forecasts of CAPS for the Hazardous Weather Testbed 2010 Spring Experiment. *25th Conf. on Weather Analysis and Forecasting/19th Conf. on Numerical Weather Prediction*, Seattle, WA, Amer.

- Meteor. Soc., 9A.2. [Available online at <https://ams.confex.com/ams/91Annual/webprogram/Paper183227.html>.]
- , F. Kong, K. A. Brewster, K. W. Thomas, J. Gao, Y. Wang, and K. K. Droegemeier, 2013: Prediction of convective storms at convection-resolving 1 km resolution over continental United States with radar data assimilation: An example case of 26 May 2008 and precipitation forecasts from spring 2009. *Adv. Meteor.*, **2013**, 259052, doi:[10.1155/2013/259052](https://doi.org/10.1155/2013/259052).
- Yussouf, N., and D. J. Stensrud, 2012: Comparison of single-parameter and multiparameter ensembles for assimilation of radar observations using the ensemble Kalman filter. *Mon. Wea. Rev.*, **140**, 562–586, doi:[10.1175/MWR-D-10-05074.1](https://doi.org/10.1175/MWR-D-10-05074.1).
- Zhang, J., and Coauthors, 2011: National Mosaic and Multi-sensor QPE (NMQ) system: Description, results, and future plans. *Bull. Amer. Meteor. Soc.*, **92**, 1321–1338, doi:[10.1175/2011BAMS-D-11-00047.1](https://doi.org/10.1175/2011BAMS-D-11-00047.1).