

A Comparison of 36–60-h Precipitation Forecasts from Convection-Allowing and Convection-Parameterizing Ensembles

ESWAR R. IYER

School of Meteorology, University of Oklahoma, Norman, Oklahoma

ADAM J. CLARK

Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, and NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma

MING XUE

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

(Manuscript received 23 October 2015, in final form 21 January 2016)

ABSTRACT

Previous studies examining convection-allowing models (CAMs), as well as NOAA/Hazardous Weather Testbed Spring Forecasting Experiments (SFEs) have typically emphasized “day 1” (12–36 h) forecast guidance. These studies find a distinct advantage in CAMs relative to models that parameterize convection, especially for fields strongly tied to convection like precipitation. During the 2014 SFE, “day 2” (36–60 h) forecast products from a CAM ensemble provided by the Center for Analysis and Prediction of Storms (CAPS) at the University of Oklahoma were examined. Quantitative precipitation forecasts (QPFs) from the CAPS ensemble, known as the Storm Scale Ensemble Forecast (SSEF) system, are compared to NCEP’s operational Short Range Ensemble Forecast (SREF) system, which provides lateral boundary conditions for the SSEF, to see if the CAM ensemble outperforms the SREF through forecast hours 36–60. Equitable threat scores (ETSs) were computed for precipitation thresholds ranging from 0.10 to 0.75 in. for each SSEF and SREF member, as well as ensemble means, for 3-h accumulation periods. The ETS difference between the SSEF and SREF peaked during hours 36–42. Probabilistic forecasts were evaluated using the area under the receiver operating characteristic curve (ROC area). The SSEF had higher values of ROC area, especially at thresholds ≥ 0.50 in. Additionally, time–longitude diagrams of diurnally averaged rainfall were constructed for each SSEF/SREF ensemble member. Spatial correlation coefficients between forecasts and observations in time–longitude space indicated that the SSEF depicted the diurnal cycle much better than the SREF, which underforecasted precipitation with a peak that had a 3-h phase lag. A minority of SREF members performed well.

1. Introduction

Historically, warm season quantitative precipitation forecasts (QPFs) have been especially challenging for numerical weather prediction (NWP) models. While

NWP forecasts for fields such as 500-hPa heights have improved, the skill of warm season QPFs has exhibited little change over time (e.g., Fritsch et al. 1998). Improvements in warm season QPFs would not only help with predictions of hazards such as flash floods (e.g., Vasiloff et al. 2007), which have accounted for roughly 2500 fatalities in the United States over the past half century (NOAA 2015), but would also benefit agriculture (e.g., through improved irrigation management), transportation industries, government, and emergency

Corresponding author address: Eswar R. Iyer, School of Meteorology, University of Oklahoma, NWC, 120 David L. Boren Blvd., Norman, OK 73072.
E-mail: eswar.iyer@noaa.gov

management (e.g., Sukovich et al. 2014). Recognizing the gap in skill for QPFs relative to other variables, Roebber et al. (2004) discussed a number of avenues for closing this gap including increasing model forecasts to sufficient resolution to explicitly depict convection and utilizing ensembles to depict the high degree of forecast uncertainty often associated with convection. However, due to computational limitations, it has only been very recently that operational models with sufficient resolution to explicitly depict convection [hereafter referred to as convection-allowing models (CAMs)¹] have become available, and assessing and improving their capabilities is a rich area of research.

Recent work has shown that CAMs provide advantages relative to models that parameterize convection for several aspects of QPF. The advantages include an improved depiction of the diurnal precipitation cycle (Clark et al. 2007, 2009; Weisman et al. 2008; Berenguer et al. 2012) and better representation of the observed convective mode (e.g., Done et al. 2004; Kain et al. 2006). Additionally, Clark et al. (2009) found a distinct advantage using objective verification metrics in a small-membership CAM-based ensemble relative to a much larger convection-parameterizing ensemble. Furthermore, Roberts and Lean (2008), Schwartz et al. (2009), and Clark et al. (2010) also show improved precipitation forecasts in CAMs relative to coarser models, but illustrate that in some cases to see the improvements, spatial scales larger than the model grid spacing need to be considered using neighborhood-based objective metrics.

Other recent work has examined how CAM guidance is perceived relative to convection-parameterizing guidance by forecasters in simulated operational forecasting environments. For example, during the 2010 and 2011 NOAA/Hazardous Weather Testbed (HWT) Spring Forecasting Experiments (SFEs), QPF products from a CAM-based ensemble were compared to guidance from the operational Short Range Ensemble Forecast (SREF) system by a group of participants led by Weather Prediction Center (WPC; formerly the Hydrometeorological Prediction Center) forecasters. The WPC-led group found such an advantage in the CAM-based ensemble that they viewed the guidance as “transformational” to warm season QPFs (Clark et al. 2012). Evans et al. (2014) also conducted an experiment in a simulated forecasting environment, finding that forecasters felt CAM-based guidance added perceived value in QPFs relative to operational models that parameterized convection for an extreme heavy rainfall event related to a tropical storm.

From the aforementioned studies, it has become clear that CAM-based guidance provides important gains in QPF relative to convection-parameterizing guidance, but little work has been done to extend these comparisons past the day 1 (12–36 h) forecast period. For example, until April 2014, NOAA/HWT SFEs have only focused on CAM-based guidance extending to 36 h. However, starting in 2014, the Center for Analysis and Prediction of Storms (CAPS) ran the Storm Scale Ensemble Forecast (SSEF) system to 60 h to test the performance through the day 2 period. At such long forecast lead times, infiltration of lateral boundary conditions (LBCs) begins to have a relatively large influence on the forecasts; thus, whether a CAM-based ensemble can still maintain its advantage warrants further investigation. Therefore, the main purpose of this study is to objectively analyze these forecasts to evaluate whether the advantages relative to convection-parameterizing guidance translate to these longer lead times. For this purpose, the 60-h SSEF system forecasts are compared to those from the 16-km grid-spacing SREF ensemble using a variety of objective metrics for evaluating deterministic and probabilistic forecasts.

The remainder of the study is organized as follows. Section 2 presents information on datasets and methods, section 3 presents results along with graphs of these metrics, and section 4 provides a summary and conclusions.

2. Data and methodology

a. SSEF and SREF ensemble description

Forecast precipitation for 3-h periods was examined from the convection-allowing SSEF and the convection-parameterizing SREF, which had 4- and 16-km grid spacing configurations, respectively. The SSEF system was provided to support the 2014 NOAA/HWT Spring Forecasting Experiment and consists of model integrations conducted from 21 April through 6 June. The 30 days of model integrations that were used for this study are as follows: 24–29 April; 1–2, 5–9, 12–16, 19–23, and 26–30 May; and 2–3 June. The SREF system 2100 UTC initializations were used, and the SSEF was initialized 3 h later at 0000 UTC (Kong et al. 2014). Because CAPS did not run the SSEF ensemble members during most weekend days during April–June, 30 days during the 2014 SFE period had the full datasets available from both ensembles. Observed precipitation data were derived from the NCEP stage IV dataset (Baldwin and Mitchell 1997), which was on a 4-km grid.

The SREF ensemble data (Du et al. 2014) were available on a 32-km grid from NCEP’s archives (available upon request). At the time, the 21-member ensemble consisted

¹ It is generally believed that a maximum of 4-km grid spacing is required for CAMs to adequately resolve the bulk circulations within organized convective systems (e.g., Weisman et al. 1997).

TABLE 1. Model specifications for all 21 operational SREF members, divided into three models (NMMB, NMM, and ARW) of seven members each (one control, six perturbed). Boldfaced text indicates a member that supplied LBCs to the SSEF. NDAS refers to the NAM Data Assimilation System, GFS refers to the Global Forecast System, and RAP is the Rapid Refresh model. BV stands for breeding vector, and blend refers to a blend of BV and ETR. FER refers to the Ferrier microphysics scheme. GFDL refers to the Geophysical Fluid Dynamics Laboratory radiation scheme. References are provided in the text.

Member	IC	IC perturbation	Convective scheme	Microphysics	PBL	Radiation (LW and SW)	Land surface
nmm_b_ctl	NDAS	BV	BMJ	FER	MYJ	GFDL	Noah
nmm_b_n1	NDAS	BV	BMJ	FER	MYJ	GFDL	Noah
nmm_b_p1	NDAS	BV	BMJ	FER	MYJ	GFDL	Noah
nmm_b_n2	NDAS	BV	SAS	GFS	GFS	GFDL	Noah
nmm_b_p2	NDAS	BV	SAS	GFS	GFS	GFDL	Noah
nmm_b_n3	NDAS	BV	BMJ	WSM6	MYJ	GFDL	Noah
nmm_b_p3	NDAS	BV	BMJ	WSM6	MYJ	GFDL	Noah
nmm_ctl	GFS	Blend	BMJ	FER	MYJ	GFDL	Noah
nmm_n1	GFS	Blend	BMJ	FER	MYJ	GFDL	Noah
nmm_p1	GFS	Blend	BMJ	FER	MYJ	GFDL	Noah
nmm_n2	GFS	Blend	SAS	FER	MYJ	GFDL	Noah
nmm_p2	GFS	Blend	SAS	FER	MYJ	GFDL	Noah
nmm_n3	GFS	Blend	KF	FER	MYJ	GFDL	Noah
nmm_p3	GFS	Blend	KF	FER	MYJ	GFDL	Noah
em_ctl	RAP	ETR	KF	FER	MYJ	GFDL	Noah
em_n1	RAP	ETR	KF	FER	MYJ	GFDL	Noah
em_p1	RAP	ETR	KF	FER	MYJ	GFDL	Noah
em_n2	RAP	ETR	BMJ	FER	MYJ	GFDL	Noah
em_p2	RAP	ETR	BMJ	FER	MYJ	GFDL	Noah
em_n3	RAP	ETR	BMJ	FER	MYJ	GFDL	Noah
em_p3	RAP	ETR	BMJ	FER	MYJ	GFDL	Noah

of 7 members from the Nonhydrostatic Multiscale Model on the B grid (NMMB; Janjić 2005, 2010; Janjić and Black 2007; Janjić et al. 2011; Janjić and Gall 2012), 7 members from the WRF Nonhydrostatic Multiscale Model (NMM; Janjić 2003), and 7 members from the Advanced Research WRF Model (ARW; Skamarock et al. 2008). Each set of seven members had one control member, three positive perturbations, and three negative perturbations. To generate these perturbations, the NMMB members used a breeding cycle (e.g., Toth and Kalnay 1997) initialized at 2100 UTC to create perturbations, which are added and subtracted from the control, creating six different perturbed analyses. The ARW members used the ensemble transform with rescaling methodology (ETR; Ma et al. 2014) to generate perturbations, while the NMM members used a blend of ETR and breeding.

Physics parameterizations in the SREF system consisted of the MYJ planetary boundary layer scheme (Mellor and Yamada 1982; Janjić 1990, 2002), the MRF planetary boundary layer scheme (Troen and Mahrt 1986; Hong and Pan 1996), and the Noah land surface model (Ek et al. 2003). Surface layer parameterizations consisted of the MYJ surface layer scheme (Mellor and Yamada 1982; Janjić 2002) and the Monin–Obukhov scheme with the Janjić Eta Model (Monin

and Obukhov 1954; Paulson 1970; Dyer and Hicks 1970; Webb 1970; Janjić 2002). Radiation schemes consisted of the GFDL shortwave (SW; Lacis and Hansen 1974) and GFDL longwave (LW) schemes (Fels and Schwarzkopf 1975; Schwarzkopf and Fels 1991). Microphysical parameterizations consisted of the scheme used in the GFS (Zhao and Carr 1997), the Ferrier scheme (Ferrier et al. 2002), and the WSM6 scheme (Ferrier et al. 2002; Hong and Lim 2006). The convection parameterization schemes consisted of the Kain–Fritsch (KF; Kain and Fritsch 1998), the Betts–Mellor–Janjić (BMJ; Betts and Albrecht 1987; Janjić 2002), and the simplified Arakawa–Schubert (SAS; Arakawa 2004). Full specifications of the SREF are given in Table 1.

The SSEF system was generated using the WRF Model (Skamarock et al. 2008) run by CAPS for the 2014 NOAA/HWT Spring Forecasting Experiment (Kong et al. 2014). During 2014, the SSEF system had 20 members with 4-km grid spacing that were initialized on weekdays at 0000 UTC and integrated for 60 h over a continental United States (CONUS) domain from late April to the beginning of June. Initial condition (IC) analysis background and LBCs (3-h updates) for the control member were taken from the NAM analyses and forecasts, respectively. Radial velocity and

TABLE 2. Configurations for 12 out of the 20 SSEF members. The abbreviations in the table are described in the text.

Member	IC	LBC	Microphysics	PBL
arw_cn	00z ARPSa	0000 UTC NAMf	Thompson	MYJ
arw_m3	arw_cn+em_p1_pert	2100 UTC SREF em_p1	Morrison	YSU
arw_m4	arw_cn+em_n2_pert	2100 UTC SREF em_n2	Thompson	QNSE
arw_m5	arw_cn+nmm_p1_pert	2100 UTC SREF nmm_p1	Morrison	MYNN
arw_m6	arw_cn+nmb_n1_pert	2100 UTC SREF nmb_n1	MY2	MYJ
arw_m7	arw_cn-nmb_p1_pert	2100 UTC SREF nmb_p1	WDM6	YSU
arw_m8	arw_cn+em_n1_pert	2100 UTC SREF em_n1	WDM6	QNSE
arw_m9	arw_cn-em_p2_pert	2100 UTC SREF em_p2	MY2	MYNN
arw_m10	arw_cn-nmb_n3_pert	2100 UTC SREF nmb_n3	Morrison	YSU
arw_m11	arw_cn-nmb_p3_pert	2100 UTC SREF nmb_p3	Thompson	YSU
arw_m12	arw_cn-em_p3_pert	2100 UTC SREF em_p3	Thompson	MYNN
arw_m13	arw_cn-nmm_p2_pert	2100 UTC SREF nmm_p2	Morrison	QNSE

reflectivity data from up to 140 Weather Surveillance Radar-1988 Dopplers (WSR-88Ds) and other high-resolution observations were assimilated into the ICs using the ARPS three-dimensional variational data assimilation (3DVAR; Xue et al. 2003; Gao et al. 2004) data and cloud analysis system (Xue et al. 2003; Hu et al. 2006; Gao and Xue 2008). IC perturbations were derived from evolved (through 3 h) perturbations of 2100 UTC initialized members of the SREF system and added to the control member ICs. For each perturbed member, the forecast of the SREF member used for the IC perturbations was also used for the LBCs. For the purposes of this study, only the 12 members comprised of the control member and the 11 members with IC/LBC perturbations were utilized. The other eight SSEF members were run with the same ICs/LBCs as the control with different physics parameterizations to study physics sensitivities.

Because a subset of the 21 SREF member forecasts is used as the LBCs for the SSEF members (except in the control), the two systems are inherently linked to each other. With the LBCs infiltrating into much of the domain interior, particularly by the latter half of the 60-h forecasts, we are essentially testing whether—given similar driving data (i.e., the LBCs)—the convection-allowing grid spacing can still provide an advantage. Table 2 shows detailed specifications of the 12 SSEF members used for this study. Nine out of the 12 members used the Noah land surface model (Ek et al. 2003), as was the case in the SREF ensemble. PBL schemes include the MYJ, Mellor–Yamada–Nakanishi–Niino (MYNN; Nakanishi 2000, 2001; Nakanishi and Niino 2004, 2006), Yonsei University (YSU; Noh et al. 2003), and quasi-normal scale elimination (QNSE; Sukoriansky et al. 2005) schemes. The microphysical parameterization consisted of the Thompson scheme (Thompson et al. 2004), Morrison scheme (Morrison et al. 2005), the WRF double-moment 6-class scheme (WDM6; Morrison

et al. 2005; Lim and Hong 2010), and the double-moment Milbrandt and Yau scheme (MY2; Milbrandt and Yau 2005).

Before any verification metrics were computed, the SREF, SSEF, and 3-h observed NCEP stage IV precipitation data were interpolated onto the 32-km grid of the SREF using a neighbor budget interpolation (e.g., Accadia et al. 2003). In addition, the SREF has a larger domain than the SSEF, which only includes the CONUS. So, a mask was used to consider only points within the SSEF domain, east of the Rocky Mountains over land, and only in the United States (Fig. 1). This is due to the relative lack of reliable WSR-88D observations over the mountains, water, and outside the United States. The area of analysis is displayed in Fig. 1.

b. Forecast evaluation metrics

The first metric that was used to evaluate the precipitation forecasts from each ensemble was the equitable threat score (ETS; Schaefer 1990). ETS measures the fraction of observed and/or forecast events that were correctly predicted, adjusted for hits associated with random chance. The ETS was calculated using contingency table elements computed from every grid point in the 32-km grid-spacing analysis domain for each ensemble member every 3 h as follows: $ETS = (H - H_{cha}) / (H + FA + M + H_{cha})$, where H represents a hit (model correctly forecasted precipitation to exceed a certain threshold); H_{cha} represents the number of hits expected by chance; FA represents the number of false alarms (model forecasted precipitation to exceed a certain threshold, but the observed precipitation did not exceed that threshold); and M represents a miss (model did not forecast precipitation to exceed a certain threshold, but the observed precipitation did exceed that threshold). An ETS of 1 is perfect, and a score below zero represents no forecast skill.

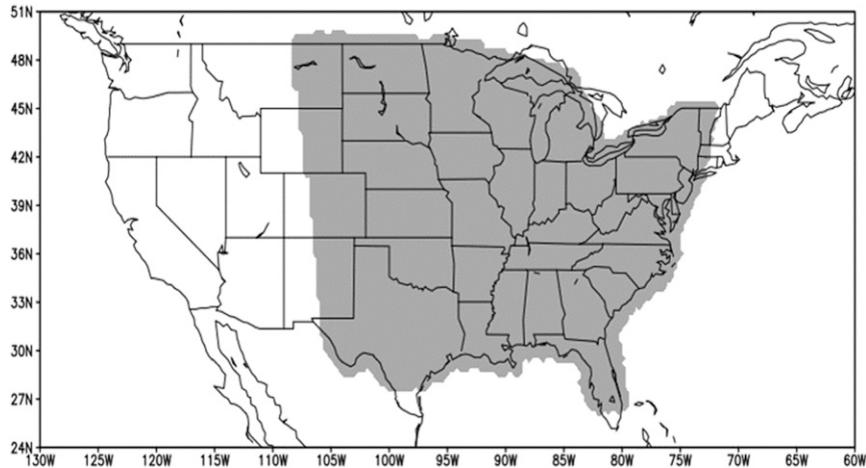


FIG. 1. Analysis domain (in gray) for the SSEF ensemble. The SSEF domain consists of the entire CONUS (not the oceans, Canada, and Mexico).

In addition to computing the ETS for individual ensemble members, ETSs were also computed for ensemble mean precipitation forecasts. Ensemble means were computed using the probability matching technique (Ebert 2001). This technique assumes that the best spatial representation of the precipitation field is given by the ensemble mean and that the best probability density function (PDF) of rain rates is given by the ensemble member QPFs for all n ensemble members. To compute the probability matched mean, the precipitation forecasts from the ensemble members for every grid point are ranked in order from largest to smallest, keeping every n th value. The precipitation forecasts from the ensemble mean forecast are similarly ranked from largest to smallest, keeping every value. Then, the grid point with the highest value in the ensemble mean QPF field is reassigned to the highest QPF value in the ensemble member QPF distribution. Next, the grid point with the second highest value in the ensemble mean QPF field is reassigned to the second highest value in the ensemble member QPF distribution. This process is then repeated for all of the rankings, ending with the lowest. ETSs were then calculated from these probability matched means for both the SSEF and SREF for forecast hours 3–60.

Finally, hypothesis testing was conducted to evaluate whether the SSEF forecasts were significantly more accurate than the SREF forecasts. The hypothesis testing was conducted for all 20 accumulation periods spanning 60 forecast hours using the resampling method of Hamill (1999). To apply this method, the test statistic used to look at the difference in accuracy of the 3-h precipitation forecast ending at hour hr (where hr is a given forecast hour) is $ETS_{SSEF_{hr}} - ETS_{SREF_{hr}}$. The null hypothesis

H_0 is $ETS_{SSEF_{hr}} - ETS_{SREF_{hr}} = 0.00$. The alternative hypothesis H_a is $ETS_{SSEF_{hr}} - ETS_{SREF_{hr}} \neq 0.00$. The significance level used was $\alpha = 0.05$ and resampling was done 1000 times for each hypothesis test. In addition, the forecasts were corrected for bias before the resampling hypothesis tests were conducted. This was done for each threshold by calculating the average bias of the two ensemble means, and then finding the precipitation threshold at which the bias of each ensemble equals the average bias from the original precipitation threshold. More detailed information about the resampling method can be found in Hamill (1999).

In addition to the deterministic forecasts, probabilistic quantitative precipitation forecasts (PQPFs) for both the SSEF and SREF were generated for all five precipitation thresholds for 3-h QPFs. Probabilities were computed using the ratio of members that exceeded the specified threshold to the total number of members. The probabilistic forecasts were evaluated using the area under the receiver operating characteristics curve (ROC area; Mason 1982), which measures the ability to discriminate between events (exceedances of specified threshold) and nonevents (failure to exceed a specified threshold). It is calculated by computing the area under a curve constructed by plotting the probability of detection (POD) versus the probability of false detection (POFD). The area under the curve is computed using the trapezoidal method (Wandishin et al. 2001) employing the probabilities 0.05–0.95, in increments of 0.05. Values of the ROC area range from 0 to 1, with a value of 0.5 indicating no forecast skill and values above 0.5 indicating positive forecast skill. Similar to ETS, statistical significance was tested for the ROC areas using the resampling method (Hamill 1999).

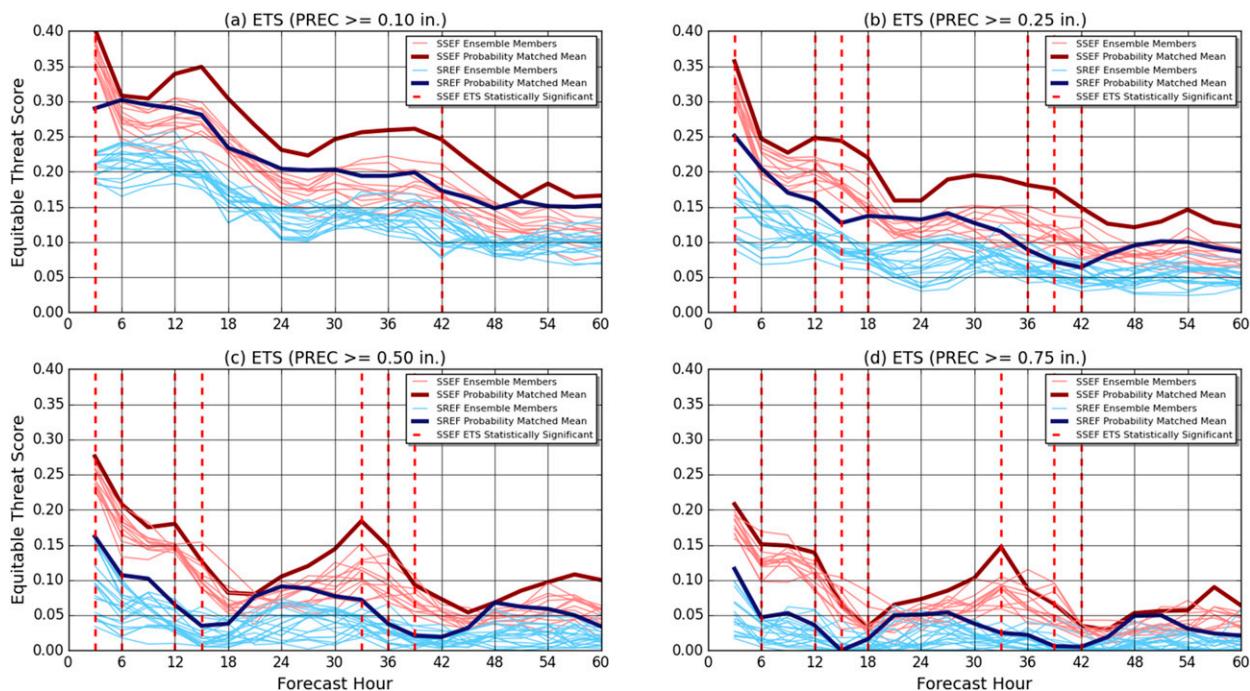


FIG. 2. The 3-h ETS at each forecast hour for the (a) 0.10-, (b) 0.25-, (c) 0.50-, and (d) 0.75-in. precipitation thresholds. Hours with significant differences between ensemble means are indicated by red, dashed, vertical lines.

To analyze the diurnal precipitation cycle, the 3-h QPFs were averaged over each forecast hour for each ensemble member, the probability matched means, and the stage IV observations. Latitudinal averages of forecast and observed 3-h precipitation were then computed and plotted in time–longitude space (i.e., Hovmoeller diagrams) for each ensemble member and the means. A Hovmoeller diagram depicting the difference between the model forecast and the observed precipitation was also constructed for each ensemble member. Spatial correlation coefficients between the forecasts and observations were computed for each 24-h forecast period (hours 12–36 and 36–60) for each ensemble member in order to quantify how well the model forecast precipitation corresponded to the observed diurnal cycle. This method is similar to that used in Clark et al. (2007, 2009).

3. Results

a. ETSs

Figures 2a–d depict the ETSs from the 0.10- to the 0.75-in. threshold for each of the ensemble members as a function of forecast hour. Generally, ETSs were fairly low, with values above 0.2 only existing at the lower thresholds and at forecast hours 3–12, mainly in the SSEF. However, these low values are consistent with the results from previous work that focused on the day 1 period (e.g.,

Clark et al. 2007). The SSEF outperformed the SREF during the day 1 period up to hour 36, with differences in ETS of around 0.05. ETSs in SSEF members had a broad maximum near 1200 UTC (forecast hours 12 and 36) and a broad minimum around 0000 UTC (forecast hour 24). There was a pronounced diurnal cycle in the SSEF member ETSs, especially for thresholds ≥ 0.50 in., likely associated with morning mesoscale convective system (MCS) activity leading to the peaks in ETSs. These peaks were likely due to the SSEF system members being able to explicitly depict large organized convective systems and their associated precipitation. This diurnal cycle in the ETSs was not as pronounced in the SREF ensemble, likely because of its inability to depict these types of convective systems.

The SSEF continued to have ETSs of 0.02–0.05 points higher than the SREF in the day 2 forecast period from hours 36 to 60, with a pronounced diurnal cycle. The difference in ETS between the SSEF and SREF ensembles was greater in the 36–42-h forecast period when compared to the later periods. But, a definite benefit still exists all the way out to forecast hour 60, as the mean ETS from the SSEF is always higher than the mean ETS from the SREF.

Significant differences between the ETSs of the probability matched means (indicated by red, dashed, vertical lines in Fig. 2) were more frequent at higher thresholds and generally occurred during the first 18 h of the forecasts and between hours 33 and 42. It is possible

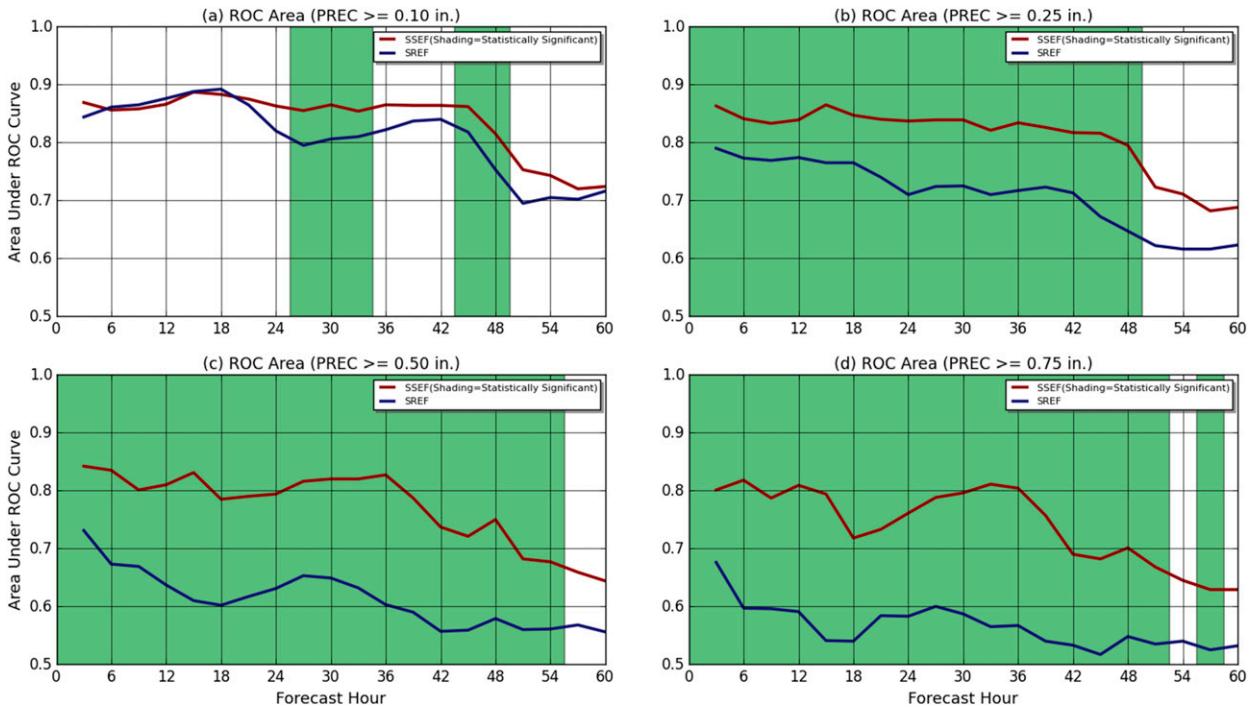


FIG. 3. The 3-h ROC area at each forecast hour for the (a) 0.10-, (b) 0.25-, (c) 0.50-, and (d) 0.75-in. precipitation thresholds. Hours with significant differences between ensemble PQPFs are indicated by green shading.

that a larger sample size or the consideration of larger spatial scales (e.g., Clark et al. 2010) would result in more times with significant differences.

b. Area under the ROC curve

Figures 3a–d depict the area under the ROC curve for all 60 forecast hours for the 0.10-, 0.25-, 0.50-, and 0.75-in. thresholds. The green shading represents statistically significant hours in favor of the SSEF. At the 0.10-in. threshold, both the SSEF and SREF have a similar amount of skill up to forecast hour 24. After that, the SSEF has a slightly greater ROC area, even during the day 2 period, with the SSEF ROC area being significantly higher at hours 42–48. At thresholds of 0.25 in. or greater, the SSEF outperforms the SREF by a much wider margin. The SSEF has significantly higher ROC area values up to hour 48 for thresholds of ≥ 0.25 in. At the 0.25-in. threshold, both ensembles have positive skill, but the SSEF consistently has a ROC area of about 0.1 greater than the SREF, including during the entire day 2 period, although hours 48–60 are not significant. The gap widens further, although the forecast skill starts to decrease at the 0.50-in. threshold. The SREF has almost no skill in the day 2 period, while the SSEF has some positive skill all the way out to forecast hour 60. During the day 2 period at the 0.75-in. threshold, the SREF has essentially no skill, while the

SSEF still has a distinct positive amount of forecast skill, although the ROC area values are lower than values at the lower thresholds. Similar to the ETSs, ROC area values for all four thresholds show a distinct advantage for the SSEF during the day 2 period. Unlike the ETS results, there was not a more pronounced difference in ROC area values for forecast hours 36–42 versus the remainder of the day 2 period. For both ensembles, the ROC area values decreased sharply after hour 48, with only a fraction of those hours being statistically significant for the SSEF. When compared to the ETS results, there were many more hours with significant differences in favor of the SSEF using ROC area. The lower differences for ROC area likely occur because the ETS evaluates single deterministic forecasts from the individual members or the ensemble mean and requires grid-point matches. Thus, if the forecast is wrong at the grid point, it is penalized. In contrast, the ROC area evaluates probabilistic forecasts that incorporate information from all ensemble members. Thus, if only one or two members of the ensemble have a correct forecast at a grid point, the forecast is only partially penalized and receives some credit for being correct. Thus, because of the inherent uncertainty associated with longer-range, high-resolution precipitation forecasts, it is easier for a superior ensemble system to

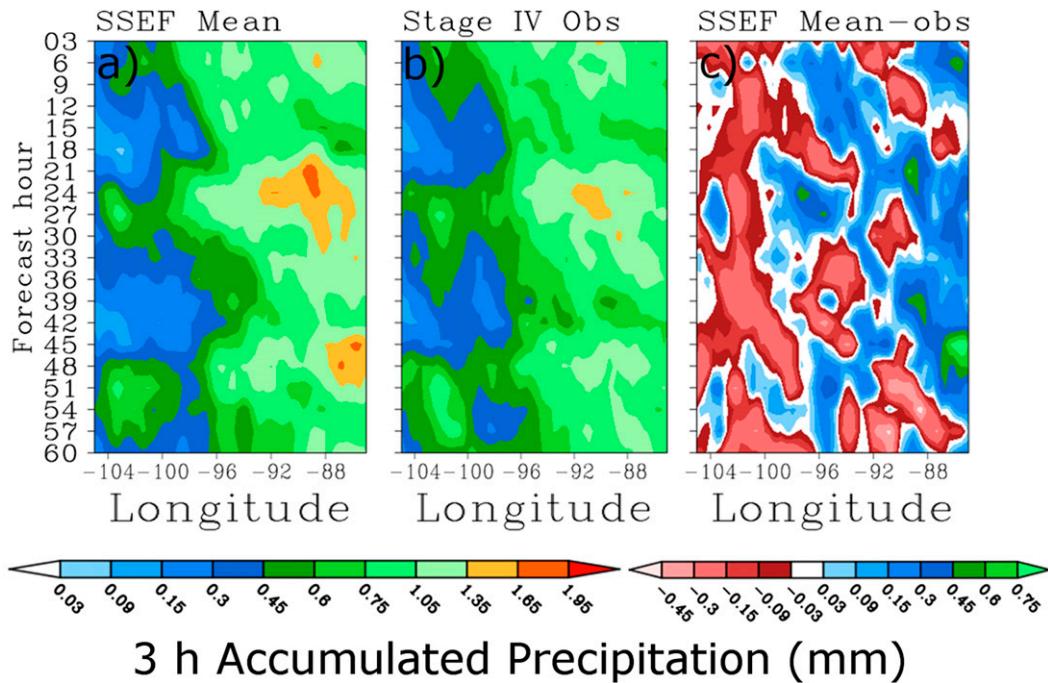


FIG. 4. Latitudinal average Hovmoeller diagram of 3-h accumulated precipitation of the (a) SSEF probability matched ensemble mean, (b) stage IV observed 3-h accumulated precipitation, and (c) difference between the forecast and observed precipitation. The color bar on the left is the legend for (a) and (b), while the color bar on the right is the legend for (c).

receive more credit using probabilistic measures that account for forecast uncertainty.

c. Hovmoeller diagrams and related metrics

Hovmoeller diagrams were created for each SSEF and SREF ensemble member, but only the probability matched means and selected members are displayed in Figs. 4 and 5. The SSEF ensemble mean is a representative depiction of the latitudinal average forecast precipitation field of all SSEF members, as the variation between members was small (not shown). There was more member-to-member variability in the SREF ensemble, but Fig. 5 is fairly representative of what the Hovmoeller diagrams of most of the SREF ensemble members looked like, with the exception of some NMM and NMMB members. These exceptions will be addressed later in this paper. In general, the SSEF better represented the observed precipitation, although there was slight overforecasting evident in the eastern areas. During the day 2 period, a diurnal cycle was clearly evident in the SSEF, with both forecast and observed precipitation maxima occurring around forecast hour 48. In the SREF, it can be seen that much of the precipitation is smoothed out, especially in the day 2 period. Furthermore, there is a 3-h phase lag relative to observations evident in the SREF mean

Hovmoeller diagram. This phase lag was observed with most of the SREF ensemble members, but not all of them. No phase lag was observed in any of the SSEF members.

As noted earlier in this section, there were a few SREF members that performed noticeably better than the SSEF ensemble mean during both the day 1 and day 2 periods. In total, four members came from the NMM model and two were from the NMMB model. Figure 6 is a Hovmoeller diagram of the nmm_n3, one of the six better-performing SREF members. As can be seen, there is no phase lag with the nmm_n3 SREF member, and a coherent diurnal cycle is evident even during the day 2 period. The other five better-performing SREF members had similar-looking Hovmoeller diagrams, with well-defined diurnal cycles (not shown). These six better-performing SREF members can be seen by looking at the results of the computed spatial correlation coefficients for the day 2 period, displayed in Fig. 7, along with the other SSEF/SREF members. The SSEF members had much higher values than most of the SREF members, but the six SREF members indicated in the plot had slightly higher coefficients than the SSEF members. In addition, the better-performing SREF members were able to accurately depict precipitation amounts across the area of analysis without a phase lag

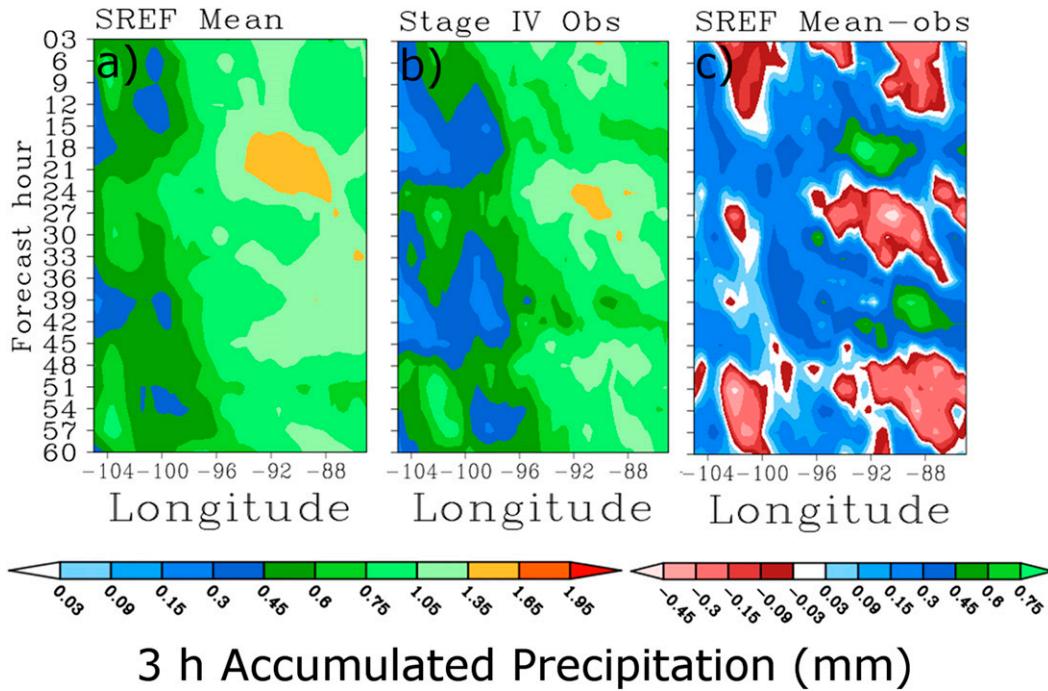


FIG. 5. As in Fig. 4 but for the SREF probability matched ensemble mean.

(not shown). Four of the six better-performing SREF members used the SAS convective parameterization scheme, but it is still not certain exactly why these six SREF members performed better.

Domain-averaged precipitation amounts are shown in Fig. 8 for all of the SREF and SREF members in red and blue, respectively, along with observations in black. The phase lag can easily be seen, as the maxima

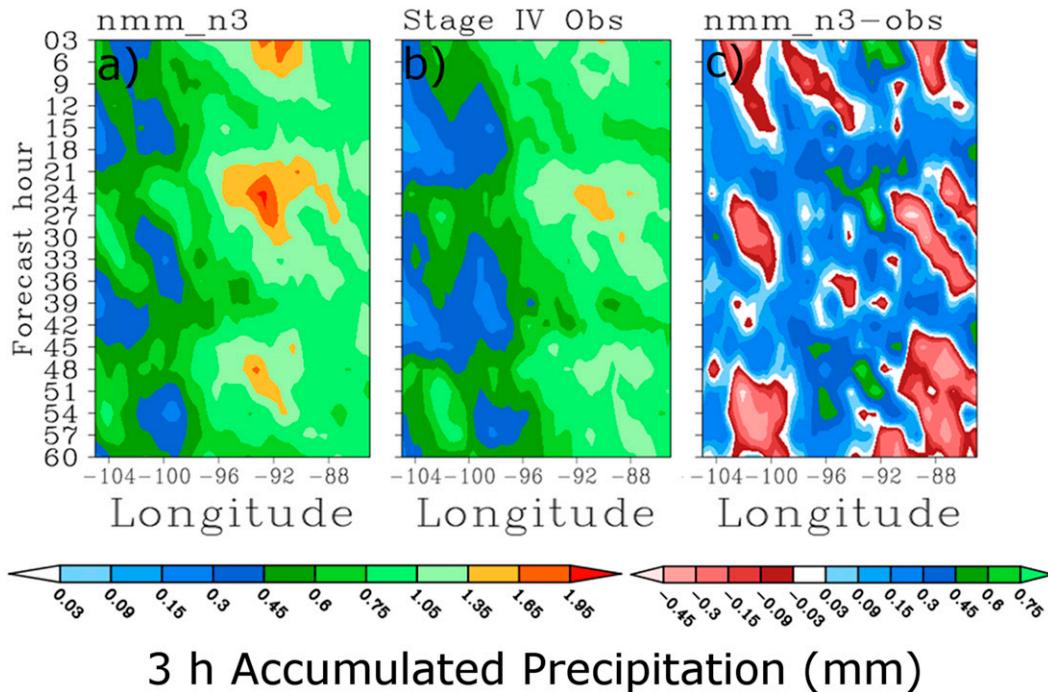


FIG. 6. As in Fig. 4 but for the nmm_n3 SREF member.

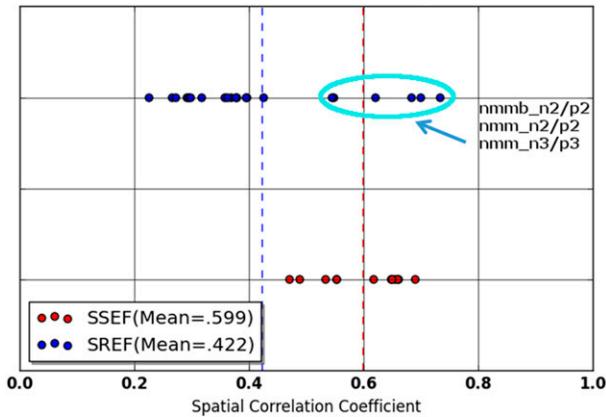


FIG. 7. Spatial correlation coefficients computed for each SSEF and SREF ensemble member for the day 2 forecast period. Red dots denote individual SSEF members, while blue dots denote individual SREF members. The red and blue dashed lines denote the mean spatial correlation coefficients of the SSEF and SREF ensembles, respectively. The six blue dots that are circled represent the six better-performing SREF members, which are listed in the plot.

in the SREF forecast precipitation are at hours 21 and 45, while the maxima in the SSEF forecasts and stage IV observations are both at hours 24 and 48. In addition, most of the SREF members greatly underforecasted the precipitation, during the both the day 1 and day 2 periods. Although the SSEF overforecast throughout the 60 h, the SSEF forecast and observed precipitation results match up fairly well, even in the day 2 period.

d. Individual cases

The ROC area for both ensembles was computed for each of the 30 cases to examine the distribution of ROC area differences between the SSEF and SREF during the day 2 forecast period. Figure 9 shows this distribution of ROC area differences (SSEF – SREF) for forecast hours 51 and 57. For both forecast hours, the median difference is around 0.1, but during forecast hour 57 there is a larger spread, with differences ranging from -0.164 to 0.498 . For both forecast hours, in about 15% of the cases, the SREF performed slightly better during the day 2 period. In a few instances, especially when large convective systems were present, the SSEF far outperformed the SREF during the day 2 period. Selected forecasts and observations from two of these cases are displayed in Fig. 10. Figures 10a and 10b show the 51-h forecast PQPFs valid at 0300 UTC on 29 April for the 0.50-in. threshold (color shading), ROC area, and 3-h observed precipitation greater than 0.50 in. (overlaid in red stippling on PQPF forecast maps). Figures 10c and 10d display the same parameters as Figs. 10a and 10b

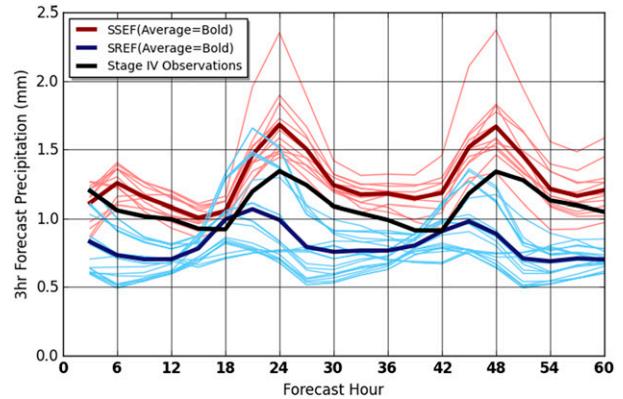


FIG. 8. Graph of the domain-averaged 3-h forecast precipitation from the SSEF members, SREF members, and stage IV observations.

except they display the 57-h forecast valid at 0900 UTC on 4 June.

There were a few differences in the synoptic setup and severe weather frequencies between the two cases. The 28–29 April case featured a vertically stacked closed low from the surface to around 300 hPa over southeast Nebraska and 50–80-kt (where $1 \text{ kt} = 0.51 \text{ m s}^{-1}$) 500-hPa winds over the region of interest displayed in Figs. 10a,b. At the surface, temperatures were only in the upper 70°s and lower 80°s F across Alabama and Mississippi, with dewpoints in the mid- to upper 60°s F. A weak cold front was advancing from west to east across the southeastern United States. Over 100 tornadoes were reported during this severe weather outbreak across the southeast, although none were rated higher on the Enhanced Fujita (EF) scale than EF2 (not shown). The SREF failed to depict where the heaviest precipitation would fall during the selected 3-h period on 29 April. None of the area with observed 3-h precipitation ≥ 0.50 in. had PQPFs greater than 0.2 for the 0.50-in. threshold. In addition, the SREF did not pick up on the southern part of the observed line of thunderstorms, having zero probabilities for a large area of Alabama and Mississippi at 0300 UTC on 29 April. On the other hand, the SSEF was able to predict the location of the main line of thunderstorms extending from eastern Mississippi to central Tennessee. Even the southern end of observed 3-h precipitation ≥ 0.50 in. corresponded fairly well to the southern end of the SSEF PQPFs for the forecast initialized 51 h before the event.

In contrast to the 29 April case, there were no well-defined upper-level features during the 3–4 June case across the plains and the midwestern United States. Zonal flow was observed in the upper levels, but there were 40–50-kt winds at 500 hPa over Nebraska, Iowa, and northern Missouri. Unlike the 28–29 April

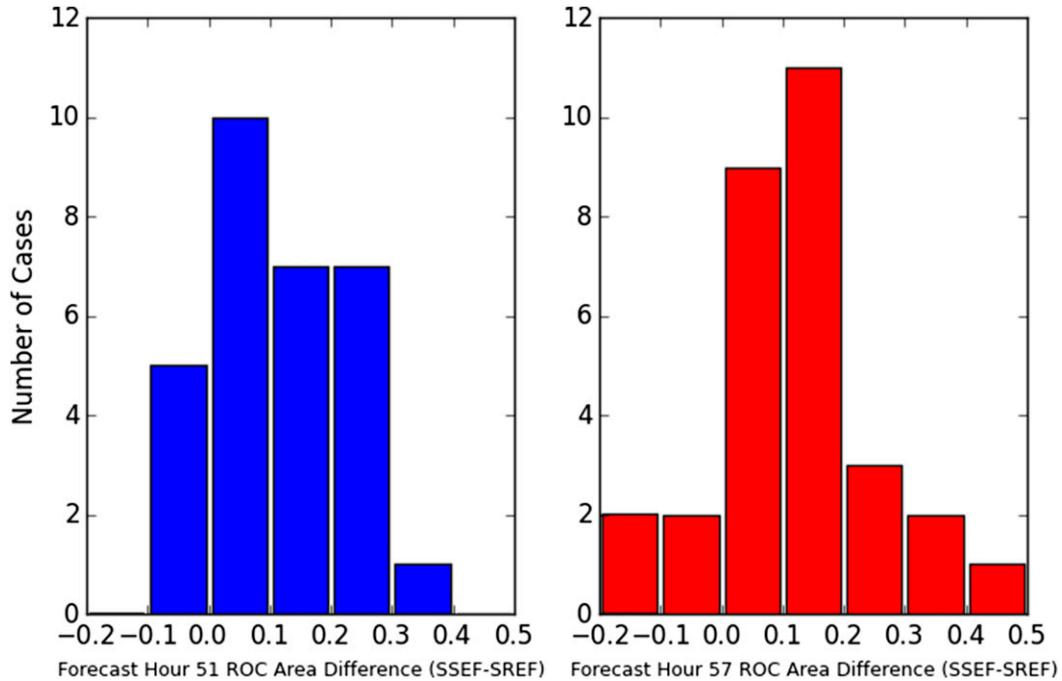


FIG. 9. Differences in ROC area for all of the cases for forecast hours (a) 51 and (b) 57.

outbreak, a large temperature gradient existed across the area as a front extended from west to east across Nebraska and Iowa. South of the front, temperatures rose to around 90°F on 3 June and dewpoints were around 70°F, while north of the front, it was only in the 60's F. Thunderstorms developed during the day in Nebraska and Iowa on 3 June and then moved into northern Missouri during the early morning hours on 4 June. Again, the SREF did not accurately predict where the bulk of the precipitation fell. It did not pick up on the area of northern Missouri and central Illinois where 3-h precipitation amounts were ≥ 0.50 in. In addition, the SREF also had a large area of false alarms in Iowa where PQPFs of 0.4–0.5 were forecasted, but no 3-h precipitation observations of ≥ 0.50 in. were observed. Conversely, the SSEF was able to predict the spatial extent of the precipitation associated with an MCS and had fewer false alarms in Iowa, as forecast probabilities were only in the 0.1–0.2 range. In the area where precipitation ≥ 0.50 in. was observed in the June case, the SSEF had probabilities of greater than 0.6 that verified. The SSEF was also able to accurately depict the eastern extension of the MCS into Illinois where the SREF had zero probability of 3-h precipitation ≥ 0.50 in.

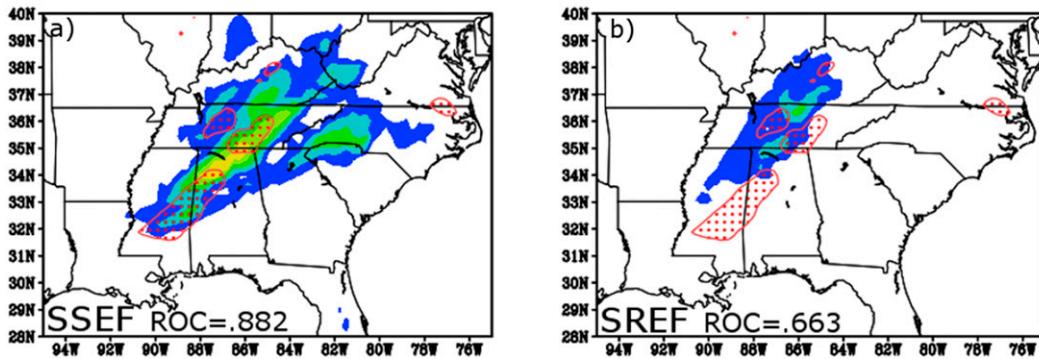
In both cases, the SREF failed to depict where the heaviest precipitation would fall during the selected 3-h period as none of the area with observed 3-h

precipitation ≥ 0.50 in. had PQPFs greater than 0.2 for the 0.50-in. threshold. In addition, the SREF did not pick up on the southern part of the observed line of thunderstorms in both the April and June cases, having zero probabilities for large areas where precipitation amounts of ≥ 0.50 in. were observed. The ROC area differences of 0.20–0.35 during these cases along with the forecasts and observations clearly show that the CAM SSEF ensemble had a definite forecast advantage during the day 2 period, with the advantage most recognizable when large-scale convection was present.

4. Summary and discussion

The 3-h QPFs during the day 2 forecast period from the 32-km grid-spacing SREF ensemble were compared to QPFs from the 4-km grid-spacing CAM SSEF ensemble. The forecasts were initialized at 2100 and 0000 UTC on 30 days during the 2014 NOAA/HWT Spring Forecasting Experiment from 24 April to 3 June. Some of the SREF members provide lateral boundary conditions for the SSEF. The goal of this study was to see whether the SSEF CAM ensemble QPFs outperform those from the convection-parameterizing SREF ensemble during the day 2 (36–60 h) forecast period, as previous work has found a distinct advantage in QPFs in CAMs relative to models that parameterize convection during the day 1 (12–36 h) period. The analysis was done

51 Hour 3 h PQPF Forecast Valid 0300 UTC 29 April 2014



57 Hour 3 h PQPF Forecast Valid 0900 UTC 4 June 2014

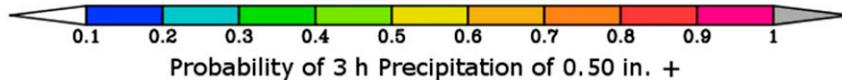
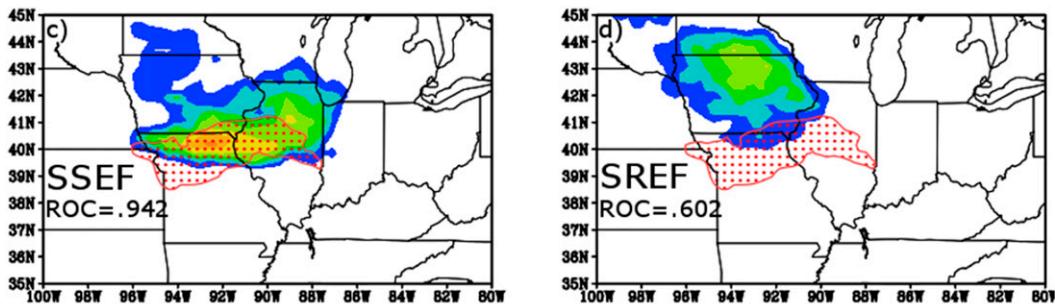


FIG. 10. The 3-h PQPF forecasts for the 0.50-in. threshold valid from 0000 to 0300 UTC 29 Apr and from 0600 to 0900 UTC 4 Jun, in color, overlaid with 3-h observed precipitation of ≥ 0.50 in. in red stippling for the (a) SSEF on 29 Apr, (b) SREF on 29 Apr, (c) SSEF on 4 Jun, and (d) SREF on 4 Jun.

by computing ETSS for both ensembles, using spatial correlation coefficients and Hovmoeller diagrams to examine the diurnal precipitation cycle, evaluating PQPFs by computing the ROC area, implementing hypothesis testing on the ETSS and ROC area, and examining two specific cases. Results are summarized below.

ETSS computed for the 0.10–0.75-in. thresholds were fairly low (≤ 0.2), but the ETSS from the SSEF were consistently 0.02–0.05 points higher during the day 2 period, with the advantage peaking during forecast hours 36–42, likely associated with morning MCS activity. Hypothesis testing supported this, as significant differences were observed during the 36–42-h forecast period at many of the thresholds. The ROC area showed an even more pronounced advantage in the SSEF, especially at the higher thresholds. While the ROC area of the SSEF was only a few hundredths of a point higher at the 0.10-in. threshold, the average difference in

ROC area was about 0.10–0.15 points higher at the higher thresholds. In addition, forecast hours 36–48 had significant differences at all thresholds ≥ 0.25 in. Hovmoeller diagrams for each ensemble member of latitudinally averaged QPFs compared with a diagram of the observed precipitation (in time–longitude space) showed that the SSEF overforecasted precipitation, but it modeled the diurnal cycle well, as evidenced by much higher mean spatial correlation coefficients than the SREF. Most of the SREF members had a 3-h phase lag in the QPFs. However, there were six SREF members that modeled the diurnal cycle well and had higher spatial correlation coefficients than the SSEF members, which was an unexpected result. Four of these six better-performing SREF members used the SAS convective parameterization scheme. However, there would need to be more investigation to determine why these six SREF members performed

significantly better than the other SREF members. Day 2 QPFs at the 0.50-in. threshold from two severe weather cases (28–29 April and 3–4 June) were examined in order to see how the SSEF and SREF forecasts would appear to an operational forecaster during a high-impact event. The SSEF greatly outperformed the SREF during the two selected cases, with ROC areas of 0.2–0.35 points higher for the selected forecast hours during these events.

After examining all of the results, it is clear that objective evaluation metrics, like the ETS, are in favor of the SSEF into the day 2 period. At all five thresholds examined, the difference in ETS is more pronounced in the early morning time frame (when some hours have statistically significant differences), when larger-scale convective systems tend to play more of a role in the forecast. The SSEF has the ability to explicitly depict these features, while the SREF cannot. Even out to hour 60, the ETS difference in favor of the SSEF ranges from about 0.02 to 0.05, with the larger differences at the higher thresholds. At thresholds higher than 0.50 in., the SREF showed essentially no skill during the day 2 period. Also, the ETS scores of the SSEF gradually decreased as the threshold increased. However, there was still some amount of skill even at the 1.00-in. threshold.

For the ROC area, similar to ETSs, the advantage of the SSEF was more pronounced as the precipitation threshold increased. This advantage was pronounced during the day 2 period as well, where the SREF showed no skill in terms of ROC area at thresholds greater than 0.50 in. The main difference in the ROC area results versus those of the ETS was that there was no pronounced maximum in performance during the early morning hours. Instead, there was a consistent difference between the SSEF and SREF during the day 2 period, with virtually all of the hours having statistically significant differences up to forecast hour 48 (at thresholds ≥ 0.25 in.).

In general, the SSEF depicts a more coherent diurnal cycle than the SREF over the domain of analysis. The SREF ensemble mean does not represent the diurnal cycle well, mainly because of its inability to explicitly depict convection, especially during the day 2 period. In addition, the SREF ensemble mean significantly underforecasted precipitation in the eastern part of the domain. The only result in favor of the SREF was those six NMM/NMMB members that slightly outperformed the SSEF in terms of spatial correlation coefficients. This was an unexpected result.

Future work should continue to examine the forecast range at which convection-allowing guidance demonstrates increased QPF skill relative to coarser modeling systems. The NOAA/HWT Spring Forecasting Experiments serve as an ideal place for these tests, since both

researchers and operational forecasters can be involved in the evaluations. The preliminary results from this study suggest that there would be at least noticeable benefits to extending an operational convection-allowing ensemble to at least 60 h.

Acknowledgments. This work was made possible by a Presidential Early Career Award for Scientists and Engineers (PECASE). Additional support was provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA–University of Oklahoma Cooperative Agreement NA11OAR4320072, U.S. Department of Commerce. CAPS SSEF forecasts were supported by the NOAA Collaborative Science, Technology, and Applied Research (CSTAR) program with supplementary support from NSF Grant AGS-0802888, using computing resources at the National Science Foundation XSEDE National Institute of Computational Science (NICS) at the University of Tennessee.

REFERENCES

- Accadia, C., S. Mariani, M. Casaioli, A. Lavagnini, and A. Speranza, 2003: Sensitivity of precipitation forecast skill scores to bilinear interpolation and a simple nearest-neighbor average method on high-resolution verification grids. *Wea. Forecasting*, **18**, 918–932, doi:10.1175/1520-0434(2003)018<0918:SOPFSS>2.0.CO;2.
- Arakawa, A., 2004: The cumulus parameterization problem: Past, present, and future. *J. Climate*, **17**, 2493–2525, doi:10.1175/1520-0442(2004)017<2493:RATCPP>2.0.CO;2.
- Baldwin, M., and K. Mitchell, 1997: The NCEP hourly multi-sensor U. S. precipitation analysis for operations and GCIP research. Preprints, *13th Conf. on Hydrology*, Long Beach, CA, Amer. Meteor. Soc., 54–55.
- Berenguer, M., M. Surcel, I. Zawadzki, M. Xue, and F. Kong, 2012: The diurnal cycle of precipitation from continental radar mosaics and numerical weather prediction models. Part II: Intercomparison among numerical models and with nowcasting. *Mon. Wea. Rev.*, **140**, 2689–2705, doi:10.1175/MWR-D-11-00181.1.
- Betts, A. K., and B. A. Albrecht, 1987: Conserved variable analysis of the convective boundary layer thermodynamic structure over the tropical oceans. *J. Atmos. Sci.*, **44**, 83–99, doi:10.1175/1520-0469(1987)044<0083:CVAOTC>2.0.CO;2.
- Clark, A. J., W. A. Gallus Jr., and T. C. Chen, 2007: Comparison of the diurnal precipitation cycle in convection-resolving and non-convection-resolving mesoscale models. *Mon. Wea. Rev.*, **135**, 3456–3473, doi:10.1175/MWR3467.1.
- , —, 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:10.1175/2009WAF2222222.1.
- , —, —, and —, 2010: Convection-allowing and convection-parameterizing ensemble forecasts of a mesoscale convective vortex and associated severe weather environment. *Wea. Forecasting*, **25**, 1052–1081, doi:10.1175/2010WAF2222390.1.
- , 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).
- Done, J., C. A. Davis, and M. Weisman, 2004: The next generation of NWP: Explicit forecasts of convection using the Weather Research and Forecasting (WRF) model. *Atmos. Sci. Lett.*, **5**, 110–117, doi:[10.1002/asl.72](https://doi.org/10.1002/asl.72).
- Du, J., G. DiMego, B. Zhou, D. Jovic, B. Ferrier, B. Yang, and S. Benjamin, 2014: NCEP regional ensembles: Evolving toward hourly-updated convection-allowing scale and storm-scale predictions within a unified regional modeling system. *22nd Conf. on Numerical Weather Prediction/26th Conf. on Weather Analysis and Forecasting*, Atlanta, GA, Amer. Meteor. Soc., J1.4. [Available online at <https://ams.confex.com/ams/94Annual/webprogram/Manuscript/Paper239030/NCEP%20Regional%20Ensembles.pdf>.]
- Dyer, A. J., and B. B. Hicks, 1970: Flux-gradient relationships in the constant flux layer. *Quart. J. Roy. Meteor. Soc.*, **96**, 715–721, doi:[10.1002/qj.49709641012](https://doi.org/10.1002/qj.49709641012).
- 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:AOAPMS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2461:AOAPMS>2.0.CO;2).
- Ek, M., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of Noah land-surface model advances in the NCEP operational mesoscale Eta model. *J. Geophys. Res.*, **108**, 8851, doi:[10.1029/2002JD003296](https://doi.org/10.1029/2002JD003296).
- Evans, C., D. F. Van Dyke, and T. Lericos, 2014: How do forecasters utilize output from a convection-permitting ensemble forecast system? Case study of a high-impact precipitation event. *Wea. Forecasting*, **29**, 466–486, doi:[10.1175/WAF-D-13-00064.1](https://doi.org/10.1175/WAF-D-13-00064.1).
- Fels, S. B., and M. D. Schwarzkopf, 1975: The simplified exchange approximation: A new method for radiative transfer calculations. *J. Atmos. Sci.*, **32**, 1475–1488, doi:[10.1175/1520-0469\(1975\)032<1475:TSEAAAN>2.0.CO;2](https://doi.org/10.1175/1520-0469(1975)032<1475:TSEAAAN>2.0.CO;2).
- Ferrier, B. S., Y. Lin, T. Black, E. Rogers, and G. DiMego, 2002: Implementation of a new grid-scale cloud and precipitation scheme in the NCEP Eta model. Preprints, *15th Conf. on Numerical Weather Prediction*, San Antonio, TX, Amer. Meteor. Soc., 10.1. [Available online at https://ams.confex.com/ams/SLS_WAF_NWP/techprogram/paper_47241.htm.]
- Fritsch, J. M., and Coauthors, 1998: Quantitative precipitation forecasting: Report of the Eighth Prospectus Development Team, U.S. Weather Research Program. *Bull. Amer. Meteor. Soc.*, **79**, 285–299, doi:[10.1175/1520-0477\(1998\)079<0285:QPFROT>2.0.CO;2](https://doi.org/10.1175/1520-0477(1998)079<0285:QPFROT>2.0.CO;2).
- Gao, J., and M. Xue, 2008: An efficient dual-resolution approach for ensemble data assimilation and tests with simulated Doppler radar data. *Mon. Wea. Rev.*, **136**, 945–963, doi:[10.1175/2007MWR2120.1](https://doi.org/10.1175/2007MWR2120.1).
- , —, 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).
- Hamill, T. M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting*, **14**, 155–167, doi:[10.1175/1520-0434\(1999\)014<0155:HTFENP>2.0.CO;2](https://doi.org/10.1175/1520-0434(1999)014<0155:HTFENP>2.0.CO;2).
- Hong, S. Y., and H. L. Pan, 1996: Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Mon. Wea. Rev.*, **124**, 2322–2339, doi:[10.1175/1520-0493\(1996\)124<2322:NBLVDI>2.0.CO;2](https://doi.org/10.1175/1520-0493(1996)124<2322:NBLVDI>2.0.CO;2).
- , and J. J. Lim, 2006: The WRF single-moment 6-class microphysics scheme (WSM6). *J. Korean Meteor. Soc.*, **42**, 129–151.
- Hu, M., M. Xue, and K. Brewster, 2006: 3DVAR and cloud analysis with WSR-88D level-II data for the prediction of the Fort Worth, Texas, 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. I., 1990: The step-mountain coordinate: Physical package. *Mon. Wea. Rev.*, **118**, 1429–1443, doi:[10.1175/1520-0493\(1990\)118<1429:TSMCPP>2.0.CO;2](https://doi.org/10.1175/1520-0493(1990)118<1429:TSMCPP>2.0.CO;2).
- , 2002: Nonsingular implementation of the Mellor-Yamada level 2.5 scheme in the NCEP Meso model. NCEP Office Note 437, NOAA/NWS, 61 pp. [Available online at http://www2.mmm.ucar.edu/wrf/users/phys_refs/SURFACE_LAYER/eta_part4.pdf.]
- , 2003: A nonhydrostatic model based on a new approach. *Meteor. Atmos. Phys.*, **82**, 271–285, doi:[10.1007/s00703-001-0587-6](https://doi.org/10.1007/s00703-001-0587-6).
- , 2005: A unified model approach from meso to global scales. *Geophysical Research Abstracts*, Vol. 7, Abstract 05582. [Available online at <http://www.cosis.net/abstracts/EGU05/05582/EGU05-J-05582.pdf>.]
- , 2010: Recent advances in global nonhydrostatic modeling at NCEP. *Proc. Workshop on Non-hydrostatic Modelling*, ECMWF, Reading, United Kingdom. [Available online at http://nwmstest.ecmwf.int/newsevents/meetings/workshops/2010/Non_hydrostatic_Modelling_presentations/Janjic.pdf.]
- , and T. Black, 2007: An ESMF unified model for a broad range of spatial and temporal scales. *Geophysical Research Abstracts*, Vol. 9, Abstract 05025. [Available online at <http://meetings.copernicus.org/www.cosis.net/abstracts/EGU2007/05025/EGU2007-J-05025.pdf>.]
- , and R. Gall, 2012: Scientific documentation of the NCEP Nonhydrostatic Multiscale Model on the B Grid (NMMB). Part 1: Dynamics. NCAR Tech. Note NCAR/TN-489+STR, 75 pp. [Available online at <http://nldr.library.ucar.edu/repository/assets/technotes/TECH-NOTE-000-000-000-857.pdf>.]
- , T. Janjić, and R. Vasic, 2011: A class of conservative fourth-order advection schemes and impact of enhanced formal accuracy on extended-range forecasts. *Mon. Wea. Rev.*, **139**, 1556–1568, doi:[10.1175/2010MWR3448.1](https://doi.org/10.1175/2010MWR3448.1).
- Kain, J. S., and J. M. Fritsch, 1998: Multiscale convective overturning in mesoscale convective systems: Reconciling observations, simulations, and theory. *Mon. Wea. Rev.*, **126**, 2254–2273, doi:[10.1175/1520-0493\(1998\)126<2254:MCOIMC>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<2254:MCOIMC>2.0.CO;2).
- , S. J. Weiss, J. J. Levit, M. E. Baldwin, and D. R. Bright, 2006: Examination of convection-allowing configurations of the WRF Model for the prediction of severe convective weather: The SPC/NSSL Spring Program 2004. *Wea. Forecasting*, **21**, 167–181, doi:[10.1175/WAF906.1](https://doi.org/10.1175/WAF906.1).
- Kong, F., and Coauthors, 2014: An overview of CAPS storm-scale ensemble forecast for the 2014 NOAA HWT Spring Forecasting Experiment. *27th Conf. on Severe Local Storms*, Madison, WI, Amer. Meteor. Soc., 43. [Available online at <https://ams.confex.com/ams/27SLS/webprogram/Paper255958.html>.]
- Lacis, A. A., and J. E. Hansen, 1974: A parameterization for the absorption of solar radiation in the earth's atmosphere. *J. Atmos. Sci.*, **31**, 118–133, doi:[10.1175/1520-0469\(1974\)031<0118:APFTAO>2.0.CO;2](https://doi.org/10.1175/1520-0469(1974)031<0118:APFTAO>2.0.CO;2).
- Lim, K. 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).
- Ma, J., Y. Zhu, D. Hou, X. Zhou, and M. Peña, 2014: Ensemble transform with 3D rescaling initialization method. *Mon. Wea. Rev.*, **142**, 4053–4073, doi:[10.1175/MWR-D-13-00367.1](https://doi.org/10.1175/MWR-D-13-00367.1).

- 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.*, **20**, 851–875, doi:10.1029/RG020i004p00851.
- Milbrandt, J. A., and M. K. Yau, 2005: A multimoment bulk microphysics parameterization. Part I: Analysis of the role of the spectral shape parameter. *J. Atmos. Sci.*, **62**, 3051–3064, doi:10.1175/JAS3534.1.
- Monin, A. S., and A. M. Obukhov, 1954: Basic laws of turbulent mixing in the surface layer of the atmosphere. *Tr. Akad. Nauk SSSR Geofiz. Inst.*, **24**, 163–187.
- Morrison, H., 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.
- Nakanishi, M., 2000: Large-eddy simulation of radiation fog. *Bound.-Layer Meteor.*, **94**, 461–493, doi:10.1023/A:1002490423389.
- , 2001: Improvement of the Mellor–Yamada turbulence closure model based on large-eddy simulation data. *Bound.-Layer Meteor.*, **99**, 349–378, doi:10.1023/A:1018915827400.
- , and H. Niino, 2004: An improved Mellor–Yamada level-3 model with condensation physics: Its design and verification. *Bound.-Layer Meteor.*, **112**, 1–31, doi:10.1023/B:BOUN.0000020164.04146.98.
- , and —, 2006: An improved Mellor–Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog. *Bound.-Layer Meteor.*, **119**, 397–407, doi:10.1007/s10546-005-9030-8.
- NOAA, 2015: 74 year list of severe weather fatalities. Office of Climate, Water, and Weather Services, NWS, 2 pp. [Available online at http://www.nws.noaa.gov/om/hazstats/resources/weather_fatalities.pdf.]
- Noh, Y., W. G. Cheon, and S. Raasch, 2003: The role of preconditioning in the evolution of open-ocean deep convection. *J. Phys. Oceanogr.*, **33**, 1145–1166, doi:10.1175/1520-0485(2003)033<1145:TROPIT>2.0.CO;2.
- Paulson, C. A., 1970: The mathematical representation of wind speed and temperature profiles in the unstable atmospheric surface layer. *J. Appl. Meteor.*, **9**, 857–861, doi:10.1175/1520-0450(1970)009<0857:TMROWS>2.0.CO;2.
- Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, doi:10.1175/2007MWR2123.1.
- Roebber, P. J., D. M. Schultz, B. A. Colle, and D. J. Stensrud, 2004: Toward improved prediction: High-resolution and ensemble modeling systems in operations. *Wea. Forecasting*, **19**, 936–949, doi:10.1175/1520-0434(2004)019<0936:TIPHAE>2.0.CO;2.
- Schaefer, J. T., 1990: The critical success index as an indicator of warning skill. *Wea. Forecasting*, **5**, 570–575, doi:10.1175/1520-0434(1990)005<0570:TCSIAA>2.0.CO;2.
- Schwartz, C. S., and Coauthors, 2009: Next-day convection-allowing WRF model guidance: A second look at 2-km versus 4-km grid spacing. *Mon. Wea. Rev.*, **137**, 3351–3372, doi:10.1175/2009MWR2924.1.
- Schwarzkopf, M. D., and S. B. Fels, 1991: The simplified exchange method revisited: An accurate, rapid method for computation of infrared cooling rates and fluxes. *J. Geophys. Res.*, **96**, 9075–9096, doi:10.1029/89JD01598.
- Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., doi:10.5065/D684SMVH.
- Sukoriansky, S., B. Galperin, and V. Perov, 2005: Application of a new spectral theory of stable stratified turbulence to the atmospheric boundary layer over sea ice. *Bound.-Layer Meteor.*, **117**, 231–257, doi:10.1007/s10546-004-6848-4.
- Sukovich, E. M., F. M. Ralph, F. E. Barthold, D. W. Reynolds, and D. R. Novak, 2014: Extreme quantitative precipitation forecast performance at the Weather Prediction Center from 2001 to 2011. *Wea. Forecasting*, **29**, 894–911, doi:10.1175/WAF-D-13-00061.1.
- Thompson, G., R. M. Rasmussen, and K. Manning, 2004: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part I: Description and sensitivity analysis. *Mon. Wea. Rev.*, **132**, 519–542, doi:10.1175/1520-0493(2004)132<0519:EFOWPU>2.0.CO;2.
- Toth, Z., and E. Kalnay, 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.
- Troen, I. B., and L. Mahrt, 1986: A simple model of the atmospheric boundary layer; sensitivity to surface evaporation. *Bound.-Layer Meteor.*, **37**, 129–148, doi:10.1007/BF00122760.
- Vasiloff, S. V., and Coauthors, 2007: Improving QPE and very short term QPF: An initiative for a community-wide integrated approach. *Bull. Amer. Meteor. Soc.*, **88**, 1899–1911, doi:10.1175/BAMS-88-12-1899.
- Wandishin, M. S., S. L. Mullen, D. J. Stensrud, and H. E. Brooks, 2001: Evaluation of a short-range multimodel ensemble system. *Mon. Wea. Rev.*, **129**, 729–747, doi:10.1175/1520-0493(2001)129<0729:EOASRM>2.0.CO;2.
- Webb, E. K., 1970: Profile relationships: The log-linear range, and extension to strong stability. *Quart. J. Roy. Meteor. Soc.*, **96**, 67–90, doi:10.1002/qj.49709640708.
- Weisman, M. L., W. C. Skamarock, and J. B. Klemp, 1997: The resolution dependence of explicitly modeled convective systems. *Mon. Wea. Rev.*, **125**, 527–548, doi:10.1175/1520-0493(1997)125<0527:TRDOEM>2.0.CO;2.
- , C. Davis, W. Wang, K. W. Manning, and J. B. Klemp, 2008: Experiences with 0–36-h explicit convective forecasts with the WRF-ARW model. *Wea. Forecasting*, **23**, 407–437, doi:10.1175/2007WAF2007005.1.
- Xue, M., D. Wang, J. 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.
- Zhao, Q.-Y., and F. H. Carr, 1997: A prognostic cloud scheme for NWP models. *Mon. Wea. Rev.*, **125**, 1931–1953, doi:10.1175/1520-0493(1997)125<1931:APCSFO>2.0.CO;2.