

# Comparisons of QPFs Derived from Single- and Multicore Convection-Allowing Ensembles

ADAM J. CLARK

*NOAA/OAR/National Severe Storms Laboratory, and School of Meteorology, University of Oklahoma,  
Norman, Oklahoma*

(Manuscript received 26 June 2019, in final form 23 September 2019)

## ABSTRACT

This study compares ensemble precipitation forecasts from 10-member, 3-km grid-spacing, CONUS domain single- and multicore ensembles that were a part of the 2016 Community Leveraged Unified Ensemble (CLUE) that was run for the 2016 NOAA Hazardous Weather Testbed Spring Forecasting Experiment. The main results are that a 10-member ARW ensemble was significantly more skillful than a 10-member NMMB ensemble, and a 10-member MIX ensemble (5 ARW and 5 NMMB members) performed about the same as the 10-member ARW ensemble. Skill was measured by area under the relative operating characteristic curve (AUC) and fractions skill score (FSS). Rank histograms in the ARW ensemble were flatter than the NMMB ensemble indicating that the envelope of ensemble members better encompassed observations (i.e., better reliability) in the ARW. Rank histograms in the MIX ensemble were similar to the ARW ensemble. In the context of NOAA's plans for a Unified Forecast System featuring a CAM ensemble with a single core, the results are positive and indicate that it should be possible to develop a single-core system that performs as well as or better than the current operational CAM ensemble, which is known as the High-Resolution Ensemble Forecast System (HREF). However, as new modeling applications are developed and incremental changes that move HREF toward a single-core system are made possible, more thorough testing and evaluation should be conducted.

## 1. Introduction and motivation

NOAA has embarked on a multiyear initiative to evolve the Next Generational Global Prediction System (NGGPS) to the Unified Forecast System (UFS). The UFS is envisioned to be a community model that leverages expertise from government, academic, and private partners, with applications spanning local to global domains and predictive time scales from subhourly to seasonal. The ultimate goal is to dramatically improve U.S. operational weather forecasting and advance the United States to world leadership in numerical modeling. The UFS will use a single dynamical core—NOAA's Finite-Volume on a Cubed Sphere (FV3) model (Putman and Lin 2007), which will simplify the National Centers for Environmental Prediction's production suite and better focus model development efforts since multiple model cores will no longer need to be maintained. The FY19–21 UFS Strategic Implementation Plan ([https://www.weather.gov/sti/stimodeling\\_nggps](https://www.weather.gov/sti/stimodeling_nggps)) provides further details.

With the push toward the UFS, the vision for a future operational, convection-allowing model (CAM) ensemble is a single-core system using FV3. However, this is contingent upon development of a computationally affordable CAM ensemble that provides similar or improved forecast guidance relative to the current operational system, which is the High-Resolution Ensemble Forecast System version 2.0 (HREF, hereafter; Roberts et al. 2019; Jirak et al. 2012). Matching or exceeding the effectiveness of HREF with a single-core system may prove to be quite challenging since HREF's configuration uses different model cores, physics parameterizations, initial and lateral boundary conditions, and includes time-lagged members, which results in optimal spread and very reliable forecast probabilities, even without calibration (e.g., Jirak et al. 2018).

Although HREF's configuration strategy has proven to be very effective, the degree to which specific aspects of the configuration contribute to ensemble spread and skill is very difficult, if not impossible, to quantify since multiple sources of ensemble diversity are simultaneously present. Thus, controlled experiments are needed to better understand forecast characteristics of CAM ensembles and identify the most important contributions

---

*Corresponding author:* Adam J. Clark, [adam.clark@noaa.gov](mailto:adam.clark@noaa.gov)

DOI: 10.1175/WAF-D-19-0128.1

For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy \(www.ametsoc.org/PUBSReuseLicenses\)](http://www.ametsoc.org/PUBSReuseLicenses).

to CAM ensemble spread and skill. This need was a primary motivation for the first Community Leveraged Unified Ensemble (CLUE) that was run during the 2016 NOAA Hazardous Weather Testbed Spring Forecasting Experiment (SFE).<sup>1</sup> The CLUE is a coordinated multi-agency effort to standardize CAM ensemble contributions from SFE collaborators so that more controlled experiments can be conducted to help guide decision making for the configuration of future operational CAM ensembles. Details on the CLUE are found in [Clark et al. \(2018\)](#).

This study focuses on a subset of 2016 CLUE members that enable an evaluation of single- versus multi-core CAM ensemble configurations. Prior research has demonstrated the effectiveness of multicore (or multimodel) ensembles for seasonal, medium and short-range forecasting applications (e.g., [Kirtman et al. 2014](#); [Candille 2009](#); and [Du et al. 2006, 2015](#), respectively), but there is little prior research testing multimodel CAM ensembles in controlled experiments. Examining the subset of 2016 CLUE members that will be evaluated herein, [Clark et al. \(2018\)](#) found slight advantages (although not statistically significant) in the multi- versus single-model strategy for severe weather and extreme precipitation, but these comparisons were conducted over 24-h time windows at relatively coarse spatial scales (severe weather proxies/reports and extreme precipitation values were remapped to 81-km grids). Thus, this study aims to analyze the impact of the multi- versus single-core strategy for CAM ensemble precipitation forecasts at higher spatial and temporal resolution. Precipitation is chosen to build on the results of [Clark et al. \(2018\)](#) and because it is a very important sensible weather field for which CAM ensembles exhibit large improvements in forecast skill relative to coarser, parameterized convection ensembles (e.g., [Clark et al. 2009](#); [Iyer et al. 2016](#)). The remainder of this study is organized as follows: [section 2](#) presents information on the model and observational datasets, [section 3](#) presents results, and [section 4](#) provides a summary and conclusions.

## 2. Data and methods

### a. Model specifications and observations

Three 10-member ensembles were compared. The first used the Advanced Research version of the Weather Research and Forecasting (WRF ARW) Model ([Skamarock et al. 2008](#)), the second used the Nonhydrostatic Multiscale

Model on the B Grid (NMMB; [Janjić and Gall 2012](#)), and the third used five of the ARW and five of the NMMB members (ARW, NMMB, and MIX, respectively). All members were initialized at 0000 UTC over a contiguous U.S. domain with 3-km horizontal grid spacing and forecasts to 36 h. These members were available for each day the 2016 SFE operated (24 days; 2 May–3 June, excluding weekends/holidays). The NMMB and ARW ensembles share the same set of initial and lateral boundary conditions, and each ensemble uses its own single set of physics. A control member in each ensemble used initial conditions (ICs) and lateral boundary conditions (LBCs; 3-h updates) from 12-km grid-spacing North American Mesoscale Forecast System (NAM) analyses and forecasts, respectively. The other members in each ensemble were initialized by combining NAM analyses with 3-h evolved perturbations of 2100 UTC initialized Short-Range Ensemble Forecast (SREF; [Du et al. 2015](#)) System members. Corresponding SREF member forecasts were used for LBCs.

ARW members used [Thompson et al. \(2004\)](#) microphysics, Rapid Radiative Transfer Model for general circulation models (RRTMG; [Iacono et al. 2008](#)) for short- and longwave radiation, 1680 grid points in the east–west direction, 1152 in the north–south direction, and 51 vertical levels with a model top of 50 hPa. NMMB members used Ferrier–Aligo ([Aligo et al. 2014](#)) microphysics, Rapid Radiative Transfer Model (RRTM; [Mlawer et al. 1997](#)) for short- and longwave radiation, 1568 grid points in the east–west direction, 1120 in the north–south direction, and 50 vertical levels with a model top of 50 hPa. Both ensembles used the Mellor–Yamada–Janjić (MYJ; [Mellor and Yamada 1982](#); [Janjić 2002](#)) boundary layer parameterization and the Noah land surface model ([Chen and Dudhia 2001](#)). Ensemble specifications for ARW and NMMB are summarized in [Tables 1 and 2](#), respectively. The MIX ensemble used the first five members of the ARW and NMMB ensembles (italicized in [Tables 1 and 2](#)). Thus, all the ICs/LBCs for MIX came from the ARW SREF members and the NAM.

Two alternative configurations of MIX were also tested since it is possible that different combinations of ICs/LBCs could impact the forecast skill and spread. One alternative configuration used the last five members of the ARW and NMMB ensembles, so that the ICs/LBCs were from one ARW and four NMMB SREF members. The other used the first five members of the ARW ensemble and the last five members of the NMMB ensemble, so that none of the members shared ICs/LBCs, and 4 of 5 ARW (NMMB) members were matched with ARW (NMMB) SREF members for ICs/LBCs. Although the tests revealed that differences between the MIX configurations were quite small, the skill scores

---

<sup>1</sup>For information on SFEs see [Kain et al. \(2003\)](#), [Clark et al. \(2012\)](#), and [Gallo et al. \(2017\)](#). The operations plan and results summary for the 2016 SFE can be found in [Clark et al. \(2016a\)](#) and [Clark et al. \(2016b\)](#), respectively.

TABLE 1. Specifications for the ARW members (as described in the 2016 SFE operations plan). The naming convention of the members “s-phys-norad” is short for single-physics and no radar data assimilation. The 12-km NAM analysis and forecast are referred to as NAMA and NAMf, respectively. In the IC column, the model names appended with “pert” refer to perturbations extracted from SREF members and the terms like p1 and n1 refer to positive and negative perturbations, respectively. Italicized members are included in MIX.

Member	IC	BC	Microphysics	LSM	PBL	Model
<i>s-phys-norad01</i>	<i>NAMA</i>	<i>NAMf</i>	<i>Thompson</i>	<i>Noah</i>	<i>MYJ</i>	<i>ARW</i>
<i>s-phys-norad02</i>	<i>NAMA+arw-p1_pert</i>	<i>arw-p1</i>	<i>Thompson</i>	<i>Noah</i>	<i>MYJ</i>	<i>ARW</i>
<i>s-phys-norad03</i>	<i>NAMA+arw-n1_pert</i>	<i>arw-n1</i>	<i>Thompson</i>	<i>Noah</i>	<i>MYJ</i>	<i>ARW</i>
<i>s-phys-norad04</i>	<i>NAMA+arw-p2_pert</i>	<i>arw-p2</i>	<i>Thompson</i>	<i>Noah</i>	<i>MYJ</i>	<i>ARW</i>
<i>s-phys-norad05</i>	<i>NAMA+arw-n2_pert</i>	<i>arw-n2</i>	<i>Thompson</i>	<i>Noah</i>	<i>MYJ</i>	<i>ARW</i>
s-phys-norad06	NAMA+arw-p3_pert	arw-p3	Thompson	Noah	MYJ	ARW
s-phys-norad07	NAMA+nmm-b-p1_pert	nmm-b-p1	Thompson	Noah	MYJ	ARW
s-phys-norad08	NAMA+nmm-b-n1_pert	nmm-b-n1	Thompson	Noah	MYJ	ARW
s-phys-norad09	NAMA+nmm-b-p2_pert	nmm-b-p2	Thompson	Noah	MYJ	ARW
s-phys-norad10	NAMA+nmm-b-n2_pert	nmm-b-n2	Thompson	Noah	MYJ	ARW

for the alternative configurations were slightly lower than the configuration with ICs/LBCs derived from ARW SREF members and the NAM, which justifies this choice for the MIX configuration. In subsequent analyses, results from these alternative MIX configurations are not shown.

For QPF verification, accumulated precipitation from NCEP’s radar-derived, gauged corrected Stage IV dataset (Lin and Mitchell 2005; Lin 2011; Nelson et al. 2016) with ~4.8-km grid spacing was used.

### b. Verification

Forecasts and observations were remapped to a common 4-km grid using a neighbor-budget interpolation (Accadia et al. 2003) for 3-h (12 time periods; 0–3, 3–6, . . . , 33–36 h) and 24-h (1 time period; 12–36 h) accumulation intervals. Verification metrics were computed over a masked area that covers most of the United States east of the Rocky Mountains (Fig. 1). The Intermountain West was avoided since precipitation estimates are not as reliable there (e.g., Smalley et al. 2014).

Separate sets of forecast probabilities for each ensemble were computed using the ratio of members that forecast above the thresholds of 0.10, 0.25, 0.50, 0.75,

1.00, 2.00, and 4.00 in. Additionally, a Gaussian kernel with  $\sigma$  of 0, 10, and 25 km was applied to smooth the probabilities resulting in three sets of forecasts in each ensemble at each time and threshold. The smoother was applied because probabilities on 4-km grids can be very noisy, and the smoothing adds additional spatial uncertainty and helps account for undersampling from small ensemble membership (e.g., Clark et al. 2011). Another set of bias-corrected probabilities was also computed using the precipitation quantile in each ensemble member corresponding to each verification threshold. For example, if the 1.00-in. threshold in the observations corresponded to a quantile of  $X$ , then I would find the forecast precipitation amount in each member at the quantile of  $X$ , and use that forecast precipitation amount as the threshold to compute the probabilities for precipitation  $\geq 1.00$  in. Thus, the bias-corrected forecast probabilities were always computed using ensemble members with a bias of 1.0. The bias correction was an important step because a high bias can sometimes artificially inflate verification metrics (e.g., Hamill 1999).

The area under the relative operating characteristic curve (AUC; Mason 1982), which plots the probability of detection versus the probability of false detection

TABLE 2. Specifications for the NMMB members (as described in the 2016 SFE operations plan). IC and BC columns are the same as in Table 1. Italicized members are included in MIX.

Member	IC	BC	Microphysics	LSM	PBL	Model
<i>nssl-nmm-b01</i>	<i>NAMA</i>	<i>NAMf</i>	<i>Ferrier–Aligo</i>	<i>Noah</i>	<i>MYJ</i>	<i>NMMB</i>
<i>nssl-nmm-b02</i>	<i>NAMA+arw-p1_pert</i>	<i>arw-p1</i>	<i>Ferrier–Aligo</i>	<i>Noah</i>	<i>MYJ</i>	<i>NMMB</i>
<i>nssl-nmm-b03</i>	<i>NAMA+arw-n1_pert</i>	<i>arw-n1</i>	<i>Ferrier–Aligo</i>	<i>Noah</i>	<i>MYJ</i>	<i>NMMB</i>
<i>nssl-nmm-b04</i>	<i>NAMA+arw-p2_pert</i>	<i>arw-p2</i>	<i>Ferrier–Aligo</i>	<i>Noah</i>	<i>MYJ</i>	<i>NMMB</i>
<i>nssl-nmm-b05</i>	<i>NAMA+arw-n2_pert</i>	<i>arw-n2</i>	<i>Ferrier–Aligo</i>	<i>Noah</i>	<i>MYJ</i>	<i>NMMB</i>
caps-nmm-b01	NAMA+arw-p3_pert	arw-p3	Ferrier–Aligo	Noah	MYJ	NMMB
caps-nmm-b02	NAMA+nmm-b-p1_pert	nmm-b-p1	Ferrier–Aligo	Noah	MYJ	NMMB
caps-nmm-b03	NAMA+nmm-b-n1_pert	nmm-b-n1	Ferrier–Aligo	Noah	MYJ	NMMB
caps-nmm-b04	NAMA+nmm-b-p2_pert	nmm-b-p2	Ferrier–Aligo	Noah	MYJ	NMMB
caps-nmm-b05	NAMA+nmm-b-n2_pert	nmm-b-n2	Ferrier–Aligo	Noah	MYJ	NMMB

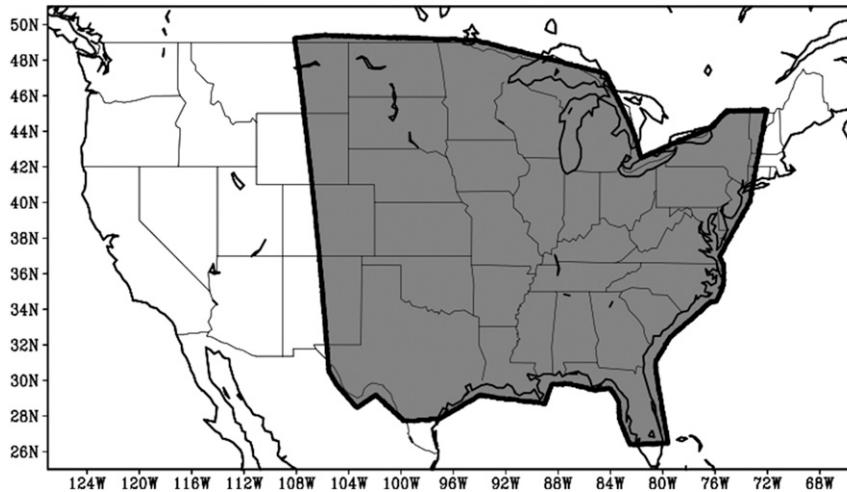


FIG. 1. Verification domain.

for a range of forecast probabilities (2% and 5%–95% in increments of 5% were used herein), was used to verify the probabilistic forecasts. The AUC measures the discriminating ability of a forecast system (AUC = 1.0 is perfect and  $\text{AUC} \leq 0.5$  indicates no skill). Additionally, the fractions skill score (FSS; Roberts and Lean 2008) was used to assess spatial placement. FSS is based on the difference in the fraction of forecast and observed grid points that exceed a specified threshold within a given radius of influence (ROI) and was computed using Eq. (3) in Loken et al. (2019). ROIs of 0, 12, 24, 40, and 60 km were used. To assess whether differences in AUC

were statistically significant, the resampling approach described by Hamill (1999) was used with resampling repeated 1000 times and  $\alpha = (0.05, 0.10)$ . For differences in FSS, paired  $t$  tests were used with  $\alpha = (0.05, 0.10)$ .

### 3. Results

For 24-h precipitation accumulation (i.e., 1200–1200 UTC), the AUC and FSS as a function of threshold are shown in Fig. 2 (note, only scores calculated using bias-corrected probabilities are shown, but results from the raw forecasts are qualitatively similar). For both AUC

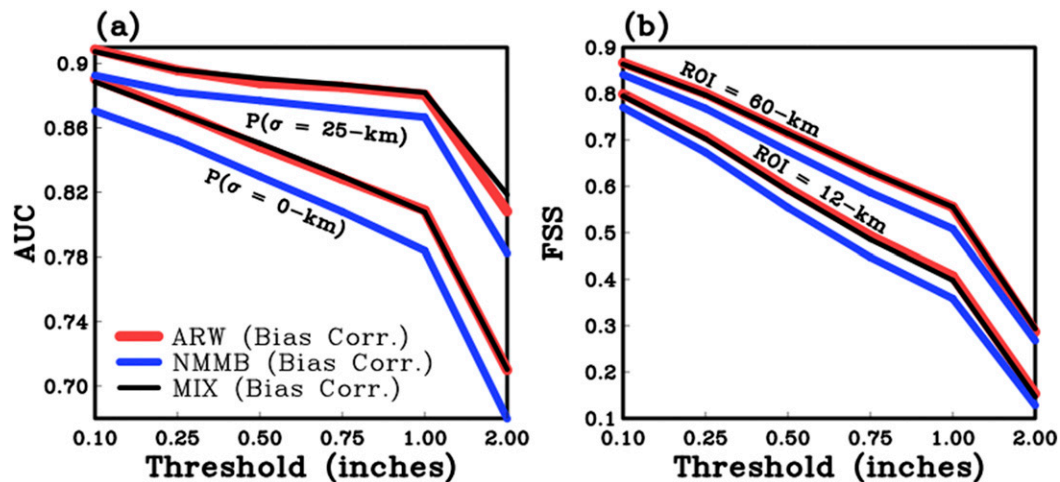


FIG. 2. (a) AUC for 24-h accumulated precipitation forecasts as a function of verification threshold from the bias-corrected ARW, NMMB, and MIX ensembles. The probabilities from one set of forecasts is smoothed using a Gaussian filter with  $\sigma = 25$  km [indicated by the label  $P(\sigma = 25\text{-km})$ ], while the other does not use any smoothing [indicated by the label  $P(\sigma = 0\text{-km})$ ]. A legend is provided at the bottom left of (a). (b) As in (a), but for FSS calculated using a radius of influence (ROI) of 12 and 60 km. Labels in (b) above each set of forecasts indicate the associated ROI.

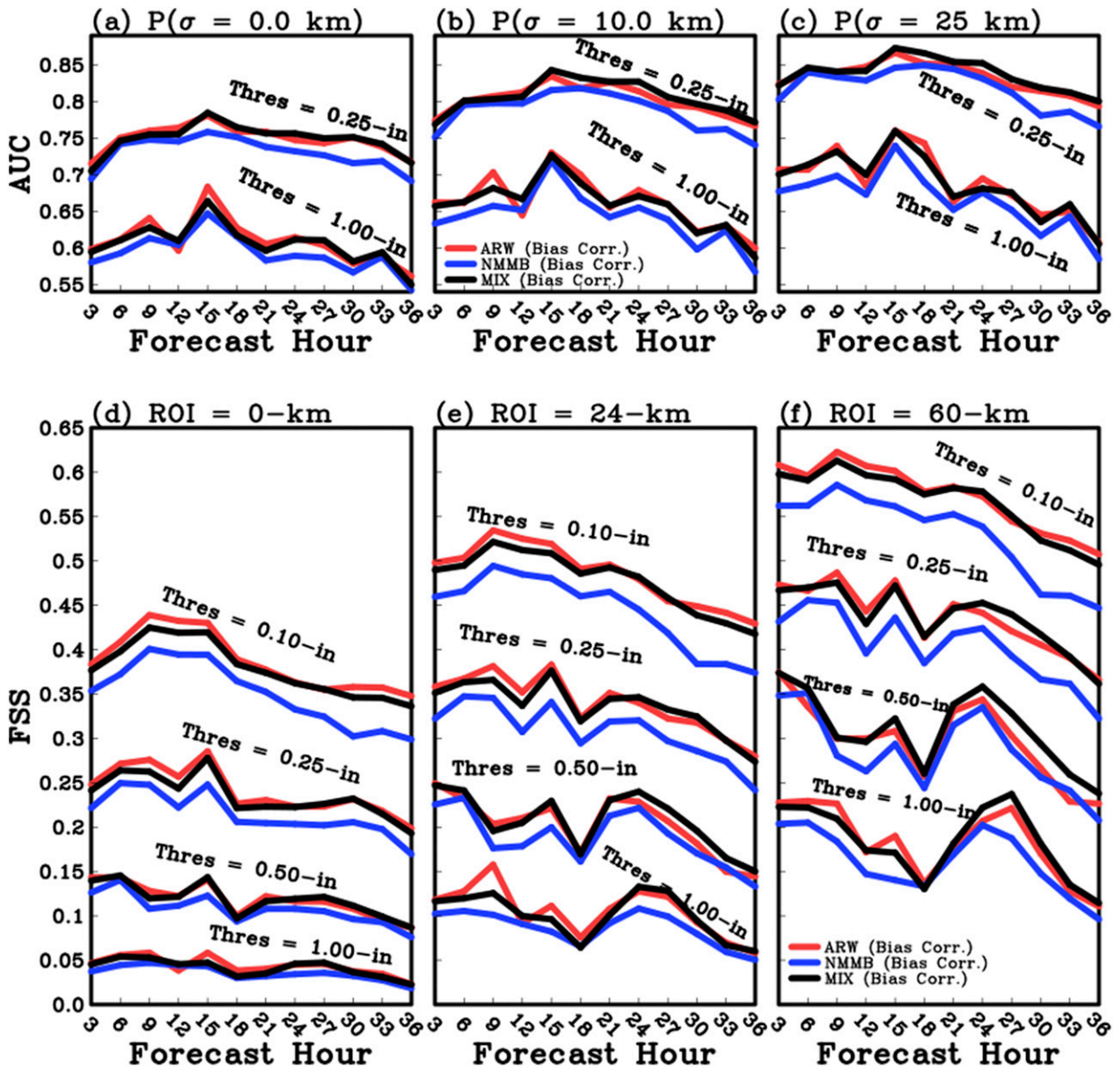


FIG. 3. (a) AUC at precipitation thresholds of 0.25 and 1.00 in. (the thresholds are indicated by labels in each panel) as a function of forecast hour for precipitation accumulated at 3-h intervals for bias-corrected ARW, NMMB, and MIX ensembles. (b),(c) As in (a), but the forecast probabilities are smoothed using a Gaussian filter with  $\sigma = 10$  km and  $\sigma = 25$  km, respectively. A legend is provided at the bottom of (b). (d)–(f) As in (a)–(c), but for FSS with ROI of 0, 24, and 60 km, respectively, and precipitation thresholds of 0.10, 0.25, 0.50, and 1.00 in. are displayed.

and FSS, the ARW and MIX perform very similarly, while the NMMB has noticeably lower skill. Although these results seem to favor the ARW and MIX over the NMMB, none of the differences were significant at  $\alpha = 0.10$ , which may result from small sample size.

For the shorter precipitation accumulation interval of 3 h (Fig. 3), ARW also appears to be superior to NMMB and about the same as MIX. However, for these shorter accumulation intervals, significance tests were performed

by treating each 3-h accumulation interval as an independent sample (i.e., each day had twelve 3-h periods resulting in 12 samples per day  $\times$  24 days = 288 samples). Though this may violate independence assumptions since errors at consecutive forecast intervals may have some correlation, past studies using this strategy such as Schwartz et al. (2015a,b, 2017), find the autocorrelations to be very small. Results from the significance testing are shown in Fig. 4. For both AUC and

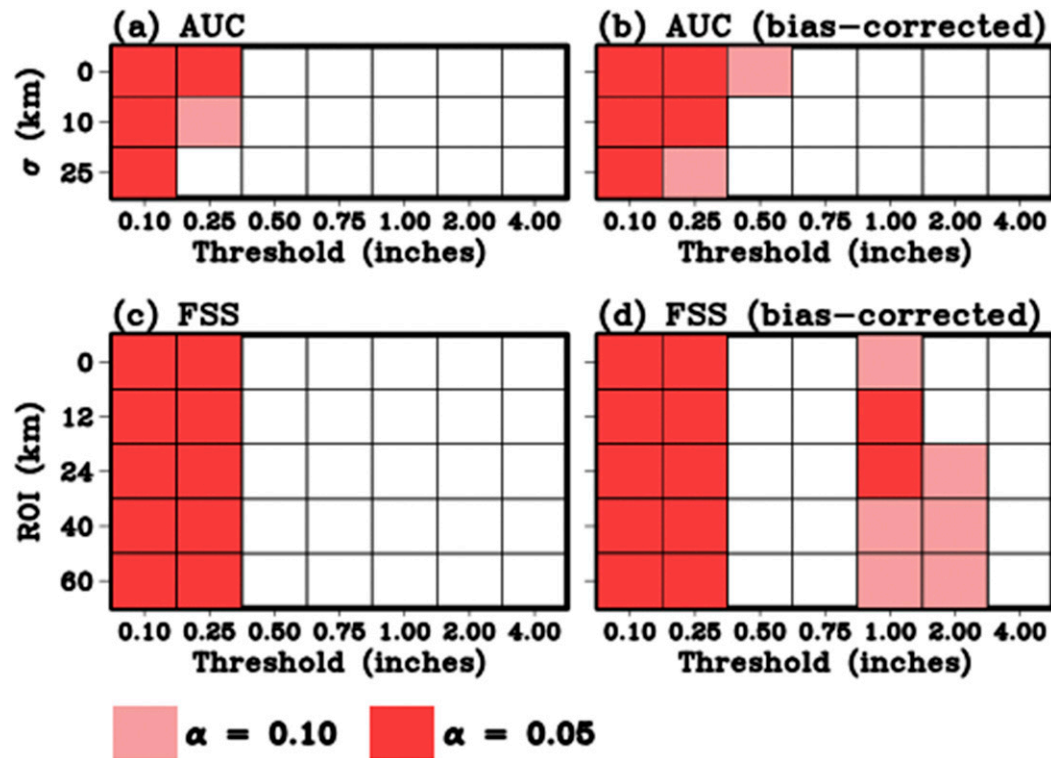


FIG. 4. Grid showing at which  $\sigma$  and precipitation threshold differences in AUC between the ARW and NMMB ensembles were statistically significant at  $\alpha = 0.05$  (red) and  $\alpha = 0.10$  (pink) for (a) raw forecasts, and (b) bias-corrected forecasts. (c),(d) As in (a) and (b), but for FSS at different ROIs.

FSS, ARW has significantly higher scores than NMMB mostly for the 0.10- and 0.25-in. thresholds (Fig. 4), however none of the differences between ARW and MIX were statistically significant (not shown).

To examine how each ensemble depicted the diurnal rainfall cycle, Hovmöller (or time–longitude) diagrams were constructed using longitudinally averaged ensemble-mean precipitation at each forecast hour, along with the Stage-IV observed precipitation (Figs. 5a–c). The ensemble means were computed using probability matching (e.g., Clark 2017). The Hovmöller diagrams show that, although ARW and NMMB depict the general characteristics of the diurnal cycle quite well, both have a westward displacement in the west to east propagating corridor of heaviest precipitation between forecast hours 3–15. In addition, NMMB has noticeably worse overprediction, especially during forecast hours 3–12 over the Central Plains. To quantify the quality of the diurnal cycle depiction in the members, spatial correlation coefficients were computed in time–longitude space for forecast hours 3–36 (Fig. 6d). The ARW correlations tended to be stronger than those of NMMB indicating a better diurnal cycle depiction in ARW.

Finally, to evaluate the spread for each ensemble, rank histograms (e.g., Hamill 2001) for 24-h precipitation

accumulations were constructed for both non-bias-corrected (Fig. 6a) and bias-corrected (Fig. 6b) forecasts. The bias-corrected rank histograms were constructed by simply using the precipitation quantiles instead of amounts. For the raw forecasts, ARW, NMMB, and MIX all have right skewness indicating overprediction, but the ARW and MIX are flatter than NMMB, which indicates that observations more frequently fall with the envelope of ensemble members (usually an indication of improved reliability). For the bias-corrected rank histograms, all three ensembles have the characteristic U shape indicating underdispersion, but ARW and MIX are noticeably flatter than NMMB, and MIX and ARW are very similar. In the alternative configuration of MIX in which all ICs/LBCs were unique, the MIX rank histograms were slightly flatter than ARW (not shown).

#### 4. Summary and conclusions

Much work is needed to meet the ambitious goal of a Unified Forecast System (UFS) to dramatically improve U.S. operational weather forecasting, especially for short-range, storm-scale applications. The CLUE initiative, which began in 2016 as part of the NOAA Hazardous

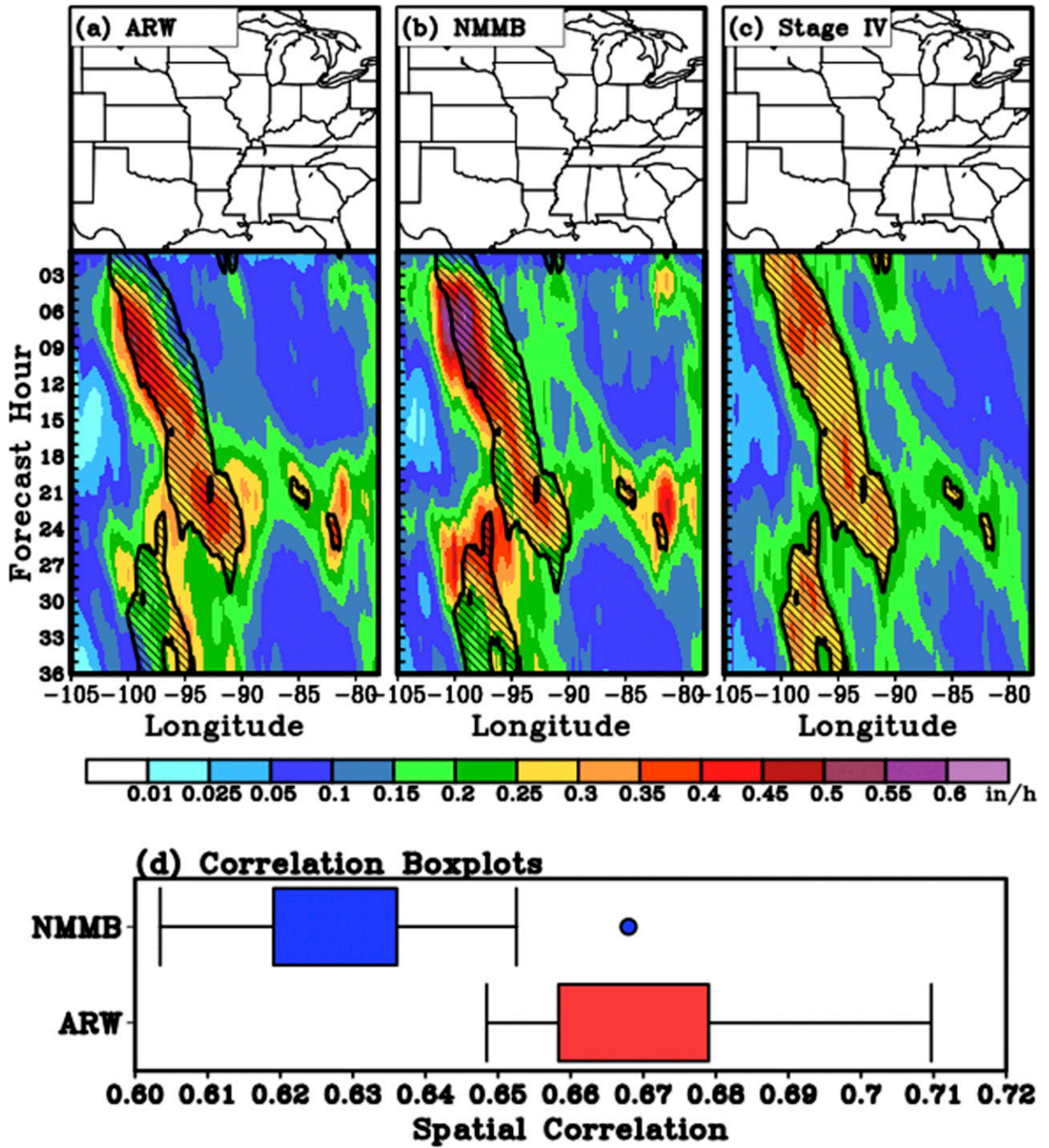


FIG. 5. Time–longitude (Hovmöller) diagrams of ensemble mean precipitation (computed using probability matching) averaged at each forecast hour over all 24 cases for (a) ARW, (b) NMMB, and (c) Stage IV observations. The hatched area in each panel indicates where the Stage IV precipitation was greater than  $0.25 \text{ in. h}^{-1}$ , and the maps above each panel indicate the area over which the Hovmöller diagrams were constructed. (d) Boxplots showing the distribution of spatial correlations between Stage IV observations and ARW and NMMB members computed in Hovmöller space for forecast hours 4–36 (i.e., each boxplot shows the distribution of spatial correlations from the 10 members comprising each ensemble).

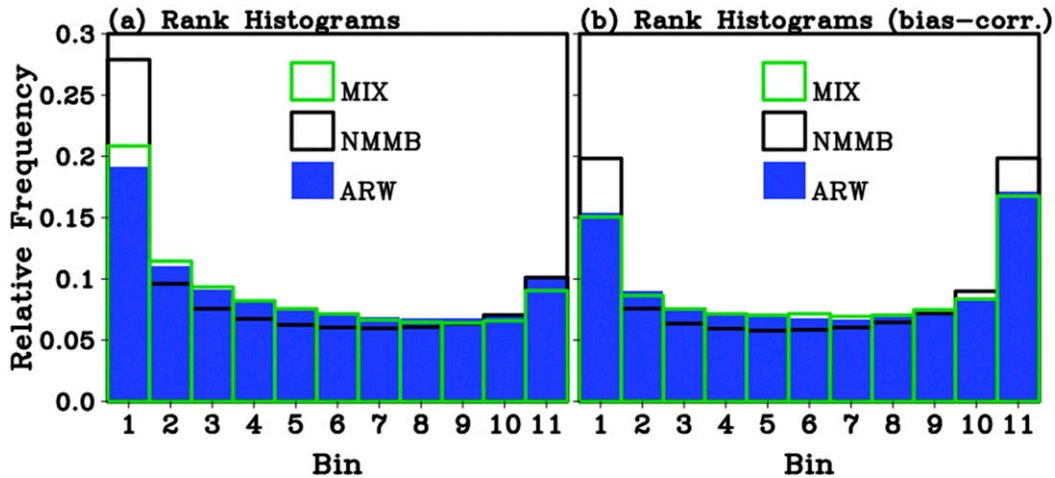


FIG. 6. Rank histograms calculated from 24-h accumulated precipitation for the ARW (blue shading), NMMB (black outline), and MIX (green outline) ensembles. The rank histograms in (a) use the raw precipitation amounts and those in (b) are bias corrected.

Weather Testbed Spring Forecasting Experiment, aims to help meet the goal of a UFS by providing evidence from controlled experiments to help guide decision making for configurations of future operational CAM ensembles. In this study, ensemble subsets from the 2016 CLUE are examined to evaluate the impact of multi- versus single-core CAM ensemble configuration strategies on ensemble precipitation forecasts. This is an important aspect of CAM ensemble configuration, since the vision for the UFS is a single-model system (based on FV3) to better focus model development efforts. In the summary that follows, be mindful that the scope of the validation was limited to precipitation during May and early June. Although this period does account for a disproportionately large amount of the annual precipitation for much of the verification domain, it is not known whether results would be similar if other seasons were included.

Results showed that, in general, for direct comparisons of precipitation forecasts between 10-member, single-core ARW and NMMB ensembles, the ARW ensemble is superior as measured by AUC and FSS. The MIX ensemble—comprised of five ARW and five NMMB members—performs about the same as the ARW. These results are in slight contrast to those of Clark et al. (2018), which showed that ARW performed better than NMMB for severe weather and extreme precipitation, but that MIX performed the best with scores that were slightly higher than ARW (although differences were not significant). The main difference between this study and Clark et al. (2018) is that forecasts at higher spatial and temporal resolution were examined herein. Thus, it appears that the relative skill of multi- and single-model configurations

can depend on the time and space scales at which verification is performed. Results from the rank histogram analysis were similar to the AUC and FSS results. Both the ARW and MIX ensembles had flatter, or less “U” shaped, rank histograms than the NMMB ensemble indicating that observations fell within the envelope of ARW and MIX ensemble members more frequently.

In the context of the UFS, extrapolating these results suggests that, in the short term, replacing NMMB members with ARW members within the HREF would make sense, since this should result in similar or improved forecast guidance relative to the current version of HREF, and it would contribute to a more unified production suite. However, it is important to emphasize that this would assume that the “replacement” ARW members have the same sources of nonmodel ensemble diversity (e.g., ICs/LBCs, time-lagging, physics) as the NMMB members they replace. In the longer term, the results suggest that a single-model CAM ensemble using FV3 must, *at the very least*, equal the skill of an ARW ensemble. However, even with equal skill, a combination of ARW and FV3 members could quite possibly perform better than an ARW ensemble. In this case, an FV3 ensemble would need to perform better than an ARW ensemble to be ideally suited for the UFS. Of course, it may be a stretch to extrapolate these results to this extent, and it is possible that results may be different for other forecast fields (e.g., temperature, dewpoint, winds). Furthermore, it is possible that as other means of introducing ensemble spread in a single-core system are developed, such as stochastic perturbations methods applied to uncertainties in physics parameterizations (e.g., Jankov et al. 2019 and references therein), additional skill will be realized. In any case, as CAM FV3 modeling applications are



developed and incremental changes that move HREF toward a FV3-based system are made possible, thorough testing and evaluation should continue to ensure that stakeholders like NOAA's National Weather Service receive the best ensemble guidance possible.

*Acknowledgments.* The ARW ensemble and half of the NMMB ensemble were run by Scott Dembek of the Cooperative Institute of Mesoscale Meteorological Studies (CIMMS) at the University of Oklahoma using high performance computing resources from the Texas Advanced Computing Center's (TACC) Lonestar5 System. The other half of the NMMB ensemble was run by the Center for Analysis and Prediction of Storms (CAPS) using TACC's Stampede system. AJC completed this work as part of regular duties at the federally funded NOAA/National Severe Storms Laboratory.

#### 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, [https://doi.org/10.1175/1520-0434\(2003\)018<0918:SOPFSS>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<0918:SOPFSS>2.0.CO;2).
- Aligo, E., B. Ferrier, J. Carley, E. Rogers, M. Pyle, S. J. Weiss, and I. L. Jirak, 2014: Modified microphysics for use in high-resolution NAM forecasts. *27th Conf. on Severe Local Storms*, Madison, WI, Amer. Meteor. Soc., 16A.1, <https://ams.confex.com/ams/27SLS/webprogram/Paper255732.html>.
- Candille, G., 2009: The multiensemble approach: The NAEFS example. *Mon. Wea. Rev.*, **137**, 1655–1665, <https://doi.org/10.1175/2008MWR2682.1>.
- Chen, F., and J. Dudhia, 2001: Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model description and implementation. *Mon. Wea. Rev.*, **129**, 569–585, [https://doi.org/10.1175/1520-0493\(2001\)129<0569:CAALSH>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2).
- Clark, A. J., 2017: Generation of ensemble mean precipitation forecasts from convection-allowing ensembles. *Wea. Forecasting*, **32**, 1569–1583, <https://doi.org/10.1175/WAF-D-16-0199.1>.
- , W. A. Gallus, 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, <https://doi.org/10.1175/2009WAF2222222.1>.
- , 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, <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, <https://doi.org/10.1175/BAMS-D-11-00040.1>.
- , and Coauthors, 2016a: Spring Forecasting Experiment 2016: Program overview and operations plan. NOAA/NSSL/SPC, 30 pp., [https://hwt.nssl.noaa.gov/sfe/2016/HWT\\_SFE2016\\_operations\\_plan\\_final.pdf](https://hwt.nssl.noaa.gov/sfe/2016/HWT_SFE2016_operations_plan_final.pdf).
- , and Coauthors, 2016b: Spring Forecasting Experiment 2016: Preliminary findings and results. NOAA/NSSL/SPC, 50 pp., [https://hwt.nssl.noaa.gov/Spring\\_2016/HWT\\_SFE\\_2016\\_preliminary\\_findings\\_final.pdf](https://hwt.nssl.noaa.gov/Spring_2016/HWT_SFE_2016_preliminary_findings_final.pdf).
- , and Coauthors, 2018: The Community Leveraged Unified Ensemble (CLUE) in the 2016 NOAA/Hazardous Weather Testbed Spring Forecasting Experiment. *Bull. Amer. Meteor. Soc.*, **99**, 1433–1448, <https://doi.org/10.1175/BAMS-D-16-0309.1>.
- Du, J., and Coauthors, 2006: New dimension of NCEP Short-Range Ensemble Forecasting (SREF) system: Inclusion of WRF members. Preprints, *WMO Expert Team Meeting on Ensemble Prediction Systems*, Exeter, United Kingdom, WMO, [http://www.wcrp-climate.org/WGNE/BlueBook/2006/individual-articles/05\\_Du\\_Jun\\_WMO06.pdf](http://www.wcrp-climate.org/WGNE/BlueBook/2006/individual-articles/05_Du_Jun_WMO06.pdf).
- , G. DiMego, B. Zhou, D. Jovic, B. Ferrier, and B. Yang, 2015: Regional ensemble forecast systems at NCEP. *27th Conf. on Weather Analysis and Forecasting/23rd Conf. on Numerical Weather Prediction*, Chicago, IL, Amer. Meteor. Soc., 2A.5.
- Gallo, B. T., and Coauthors, 2017: Breaking new ground in severe weather prediction: The 2015 NOAA/Hazardous Weather Testbed Spring Forecasting Experiment. *Wea. Forecasting*, **32**, 1541–1568, <https://doi.org/10.1175/WAF-D-16-0178.1>.
- Hamill, T. M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting*, **14**, 155–167, [https://doi.org/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).
- , 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, **129**, 550–560, [https://doi.org/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).
- Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res.*, **113**, D13103, <https://doi.org/10.1029/2008JD009944>.
- Iyer, E. R., A. J. Clark, M. Xue, and F. Kong, 2016: A comparison of 36–60-h precipitation forecasts from convection-allowing and convection-parameterizing ensembles. *Wea. Forecasting*, **31**, 647–661, <https://doi.org/10.1175/WAF-D-15-0143.1>.
- Janjić, Z. I., 2002: Nonsingular implementation of the Mellor–Yamada level 2.5 scheme in the NCEP Meso Model. NCEP Office Note 437, 61 pp., <https://www.emc.ncep.noaa.gov/officenotes/newernotes/on437.pdf>.
- , and R. Gall, 2012: Scientific documentation of the NCEP Nonhydrostatic Multiscale Model on the B Grid (NMMB). Part 1: Dynamics. NCAR/TN-489+STR, 75 pp., <https://doi.org/10.5065/D6WH2MZX>.
- Jankov, I., J. Beck, J. Wolff, M. Harrold, J. B. Olson, T. Smirnova, C. Alexander, and J. Berner, 2019: Stochastically perturbed parameterization in an HRRR-based ensemble. *Mon. Wea. Rev.*, **147**, 153–173, <https://doi.org/10.1175/MWR-D-18-0092.1>.
- Jirak, I. L., S. J. Weiss, and C. J. Melick, 2012: The SPC Storm-Scale Ensemble of Opportunity: Overview and results from the 2012 Hazardous Weather Testbed Spring Forecasting Experiment. *26th Conf. on Severe Local Storms*, Nashville, TN, Amer. Meteor. Soc., 137, <https://ams.confex.com/ams/26SLS/webprogram/Paper211729.html>.
- , A. J. Clark, B. Roberts, B. T. Gallo, and S. J. Weiss, 2018: Exploring the optimal configuration of the High-Resolution Ensemble Forecast System. *25th Conf. on Numerical Weather Prediction*, Denver, CO, Amer. Meteor. Soc., 14B.6, <https://ams.confex.com/ams/29WAF25NWP/webprogram/Paper345640.html>.
- Kain, J. S., P. R. Janish, S. J. Weiss, M. E. Baldwin, R. S. Schneider, and H. E. Brooks, 2003: Collaboration between forecasters and research scientists at the NSSL and SPC: The spring program. *Bull. Amer. Meteor. Soc.*, **84**, 1797–1806, <https://doi.org/10.1175/BAMS-84-12-1797>.

- Kirtman, B. P., and Coauthors, 2014: The North American Multi-model Ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.*, **95**, 585–601, <https://doi.org/10.1175/BAMS-D-12-00050.1>.
- Lin, Y., 2011: GCIPEOP surface: Precipitation NCEP/EMC 4KM Gridded Data (GRIB) Stage IV data, version 1.0. UCAR/NCAR–Earth Observing Laboratory, accessed 1 August 2018 <https://doi.org/10.5065/D6PG1QDD>.
- , and K. E. Mitchell, 2005: The NCEP stage II/IV hourly precipitation analyses: Development and applications. *19th Conf. on Hydrology*, San Diego, CA, Amer. Meteor. Soc., 1.2, <https://ams.confex.com/ams/pdfpapers/83847.pdf>.
- Loken, E. D., A. J. Clark, M. Xue, and F. Kong, 2019: Spread and skill in mixed- and single-physics convection-allowing ensembles. *Wea. Forecasting*, **34**, 305–330, <https://doi.org/10.1175/WAF-D-18-0078.1>.
- 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, <https://doi.org/10.1029/RG020i004p00851>.
- Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16 663–16 682, <https://doi.org/10.1029/97JD00237>.
- Nelson, B. R., O. P. Prat, D.-J. Seo, and E. Habib, 2016: Assessment and implications of NCEP Stage IV quantitative precipitation estimates for product inter-comparisons. *Wea. Forecasting*, **31**, 371–394, <https://doi.org/10.1175/WAF-D-14-00112.1>.
- Putman, W. M., and S.-J. Lin, 2007: Finite-volume transport on various cubed-sphere grids. *J. Comput. Phys.*, **227**, 55–78, <https://doi.org/10.1016/j.jcp.2007.07.022>.
- Roberts, B., I. L. Jirak, A. J. Clark, S. J. Weiss, and J. S. Kain, 2019: Postprocessing and visualization techniques for convection-allowing ensembles. *Bull. Amer. Meteor. Soc.*, **100**, 1245–1258, <https://doi.org/10.1175/BAMS-D-18-0041.1>.
- 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, <https://doi.org/10.1175/2007MWR2123.1>.
- Schwartz, C. S., G. S. Romine, R. A. Sobash, K. R. Fossell, and M. L. Weisman, 2015a: NCAR’s experimental real-time convection-allowing ensemble prediction system. *Wea. Forecasting*, **30**, 1645–1654, <https://doi.org/10.1175/WAF-D-15-0103.1>.
- , —, M. L. Weisman, R. A. Sobash, K. R. Fossell, K. W. Manning, and S. B. Trier, 2015b: A real-time convection-allowing ensemble prediction system initialized by mesoscale ensemble Kalman filter analyses. *Wea. Forecasting*, **30**, 1158–1181, <https://doi.org/10.1175/WAF-D-15-0013.1>.
- , —, K. R. Fossell, R. A. Sobash, and M. L. Weisman, 2017: Toward 1-km ensemble forecasts over large domains. *Mon. Wea. Rev.*, **145**, 2943–2969, <https://doi.org/10.1175/MWR-D-16-0410.1>.
- Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., <https://doi.org/10.5065/D68S4MVH>.
- Smalley, M., T. L’Ecuyer, M. Lebsock, and J. Haynes, 2014: A comparison of precipitation occurrence from the NCEP Stage IV QPE product and the CloudSat cloud profiling radar. *J. Hydrometeor.*, **15**, 444–458, <https://doi.org/10.1175/JHM-D-13-048.1>.
- 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, [https://doi.org/10.1175/1520-0493\(2004\)132<0519:EFOWPU>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<0519:EFOWPU>2.0.CO;2).