

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/343198269>

A Method for Probability Matching Based on the Ensemble Maximum for Quantitative Precipitation Forecasts

Article in *Monthly Weather Review* · August 2020

DOI: 10.1175/MWR-D-20-0003.1

CITATIONS

0

READS

72

5 authors, including:



Xiaoshi Qiao

Nanjing University of Science and Technology

8 PUBLICATIONS 8 CITATIONS

[SEE PROFILE](#)



Shizhang Wang

Nanjing University of Information Science & Technology

13 PUBLICATIONS 43 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Key Technology and Application of Limited-area Numerical Weather Model in Subtropical Region [View project](#)



ensemble [View project](#)

A Method for Probability Matching Based on the Ensemble Maximum for Quantitative Precipitation Forecasts

XIAOSHI QIAO AND SHIZHANG WANG

Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Key Laboratory of Meteorological Disaster of Ministry of Education, Nanjing University of Information Science and Technology, Nanjing, China, and National Center for Atmospheric Research, Boulder, Colorado

CRAIG S. SCHWARTZ AND ZHIQUAN LIU

National Center for Atmospheric Research, Boulder, Colorado

JINZHONG MIN

Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Key Laboratory of Meteorological Disaster of Ministry of Education, Nanjing University of Information Science and Technology, Nanjing, China

(Manuscript received 4 January 2020, in final form 2 June 2020)

ABSTRACT

A probability matching (PM) product using the ensemble maximum (EnMax) as the basis for spatial reassignment was developed. This PM product was called the PM max and its localized version was called the local PM (LPM) max. Both products were generated from a 10-member ensemble with 3-km horizontal grid spacing and evaluated over 364 36-h forecasts in terms of the fractions skill score. Performances of the PM max and LPM max were compared to those of the traditional PM mean and LPM mean, which both used the ensemble mean (EnMean) as the basis for spatial reassignment. Compared to observations, the PM max typically outperformed the PM mean for precipitation rates $\geq 5 \text{ mm h}^{-1}$; this improvement was related to the EnMax, which had better spatial placement than the EnMean for heavy precipitation. However, the PM mean produced better forecasts than the PM max for lighter precipitation. It appears that the global reassignment used to produce the PM max was responsible for its poorer performance relative to the PM mean at light precipitation rates, as the LPM max was more skillful than the LPM mean at all thresholds. These results suggest promise for PM products based on the EnMax, especially for rare events and ensembles with insufficient spread.

1. Introduction

Ensemble quantitative precipitation forecasts (QPFs) quantify forecast uncertainties in numerical weather prediction (Seo et al. 2000; Mullen and Buizza 2001; Krajewski and Ciach 2003; Ciach et al. 2007) but often generate tens of ensemble members, which complicates the interpretation of ensemble forecasts. Moreover, there is difficulty in properly conveying prediction uncertainty and transitioning probabilistic guidance into risk management (Ansell 2013). It is also difficult for end-users unfamiliar with ensemble forecast information to ingest probabilistic guidance into their specific applications (Yang et al. 2019). Therefore, it is worth

designing deterministic products that gather information across all ensemble members for end-users.

A simple approach is to compute the ensemble mean (EnMean) (Leith 1974; Murphy 1988; Du et al. 1997; Speer and Leslie 1997), where the amount of precipitation at each grid point is the average of all members. This average highlights the common features of individual members but heavily dampens the highest intensities (Warner 2010). Furthermore, many studies have shown that the EnMean substantially underestimates observed heavy precipitation and increases the spatial coverage of light rainfall (Ebert 2001; Clark et al. 2008; Fang and Kuo 2013; Schwartz et al. 2014; Surcel et al. 2014; Hamill et al. 2017).

To overcome deficiencies of the EnMean, a probability matching (PM) method was proposed that blends

Corresponding author: Shizhang Wang, szwang@nuist.edu.cn

DOI: 10.1175/MWR-D-20-0003.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](https://www.ametsoc.org/PUBSReuseLicenses) (www.ametsoc.org/PUBSReuseLicenses).

one type of data that provides a better spatial representation with another type of data that has a greater accuracy of frequency distribution (Rosenfeld et al. 1993; Anagnostou et al. 1999). This technique was first applied to the EnMean in producing ensemble-based QPF products by Ebert (2001), who hypothesized that the most likely spatial placement of rainfall was given by the EnMean and the best probability density function (PDF) of precipitation rates was from individual ensemble members. Thus, Ebert (2001) introduced the “probability matched mean” (hereafter “PM mean”), a deterministic product where the EnMean precipitation is replaced by precipitation values from individual members through a spatial reassignment process based on the EnMean. The PM mean has been widely employed to produce ensemble-based QPF guidance (Clark et al. 2009, 2012; Novak et al. 2014; Huang and Luo 2017; Gowan et al. 2018) and has better average performance than individual ensemble members and the EnMean (e.g., Clark et al. 2009; Kong et al. 2009; Xue et al. 2011; Berenguer et al. 2012; Schwartz et al. 2014; Zhang 2018).

Traditionally, the PM mean has been computed by allowing all points in the computational domain to participate in reassignment, meaning gridpoint values from individual members could ultimately be reassigned to very different geographic locations in the PM mean. However, recently, this approach has been questioned because of concerns that incorporating spatially distant information within the PM mean may be inappropriate. For example, for a given grid point, Potvin et al. (2017) only allowed points within a 10-km radius to participate in reassignment and found this neighborhood PM mean was more representative of the ensemble than the traditional global approach. Clark (2017) also found similar issues when the PM mean was calculated over a large region, which risks assigning values to a completely different geographic or climatological area or mesoscale environment, leading to a failure of the PM mean to locally represent the ensemble members. Thus, Clark (2017) proposed the localized PM mean (LPM mean), which only allows points within a certain distance of each grid point to contribute to reassignment. Considering the steep computational cost of Clark (2017)’s point-by-point LPM mean algorithm, Snook et al. (2019) proposed an efficient “patchwise LPM mean” method, which applied the PM mean over a set of nonoverlapping local patches. Another development branch of the PM mean method was to linearly combine the PM mean and ensemble 90th percentile weighted by their precipitation amounts (Zhang 2018).

Despite the successful application and development of the PM mean at convection-allowing scales, using the EnMean as the basis for spatial placement sometimes

degraded performance of the PM mean. For example, Fang and Kuo (2013) showed that the PM mean may not be suitable for situations where rainfall spatial distributions are strongly influenced by topography. In addition, Surcel et al. (2014) determined that the PM mean had better skill than individual ensemble members because of reduced small-scale variability, as the EnMean and PM mean are smoother than individual members and thus filter out unpredictable small-scale information. Furthermore, Hamill et al. (2017) demonstrated that the diversity of ensemble positions could cause EnMean precipitation distributions to differ from the members, an undesirable situation. More broadly, EnMean precipitation fields depart from the model “attractor” (Ansell 2013; Schwartz et al. 2014) and may not always be the best representation of precipitation placement, especially for intense, localized phenomena like convection.

Consequently, this work considers whether a different field may better represent precipitation placement than the EnMean for PM purposes. Specifically, the ensemble maximum (EnMax) was selected as the basis for reassignment. Unlike the EnMean, which emphasizes agreement of ensemble members, the EnMax retains gridpoint maximum precipitation no matter the agreement among members, and thus depicts more small-scale precipitation variability. This characteristic of the EnMax avoids the overly smooth issue of the EnMean (Ansell 2013; Hollan and Ansell 2015). In addition, the EnMax can better capture extreme events and provides information about forecast “upper bounds” that can be informative for forecasters (Evans et al. 2014). Furthermore, while the EnMean emphasizes ensemble consensus and deemphasizes uncertainty, the EnMax reflects the diversity of ensemble members’ rainfall, implicitly retaining information about uncertainty, which may be beneficial for an ensemble with insufficient spread. However, whether using the EnMax to perform reassignment increases the ability of a PM product to capture high-magnitude events is unknown and will be investigated in this study.

Therefore, this work investigated whether using the EnMax as the basis for spatial reassignment to produce a “probability matched maximum” (hereafter, “PM max”) product can improve upon the PM mean. Additionally, since the LPM mean was reported to be better than the PM mean (Clark 2017), this work examines performances of the localized PM max (LPM max) and the LPM mean to determine whether using the EnMax can result in better forecast skill than the EnMean when the localized approach is adopted.

The methodology and data are described in sections 2 and 3, respectively, and characteristics of the EnMax and EnMean are discussed in section 4. PM and LPM

products are presented and verified in [section 5](#) before concluding in [section 6](#).

2. Methodology

The PM and LPM methods are briefly reviewed using the EnMean as an example; the corresponding EnMax product can be obtained by substituting the EnMax for EnMean. The EnMean and EnMax are determined by the average and maximum values of N ensemble members at each grid point, respectively.

a. The PM method

To provide a more realistic precipitation forecast than the EnMean, the PM mean combines the precipitation spatial distribution from the EnMean with the precipitation frequency distribution from the ensemble members ([Ebert 2001](#)). The procedure for obtaining the PM mean is summarized as follows:

- 1) Compute the EnMean at each of M grid points within the entire forecast domain;
- 2) Sort the M precipitation amounts of the EnMean from lowest to highest and store their ranks and corresponding gridpoint locations (i, j) in the array R_{mean} ;
- 3) Order the precipitation amounts of all N ensemble members for all grid points from lowest to highest and store every N th value in the array R_{mem} ;
- 4) Assign the m th (m is from 1 to M) value of R_{mem} to the m th (i, j) of R_{mean} (e.g., assign the *location* with the highest EnMean amount the highest value from the ensemble distribution). For any nonprecipitation grid points in the EnMean, the corresponding points in the PM mean are forced to zero.

This procedure can be applied to produce the PM max by replacing the EnMean in the above steps with the EnMax. Note that the EnMean, EnMax, PM mean, and PM max all have identical nonzero precipitation areas.

b. The localized PM method

The LPM mean ([Clark 2017](#)) is a modification of the PM mean that only allows points within a local neighborhood to participate in assigning a value at the m th point, contrasting the PM mean and PM max, which allow values from all points in the computational domain to potentially be reassigned to the m th point. The calculation of the LPM mean following [Clark \(2017\)](#) is summarized as follows:

- 1) Produce the EnMean as in the first step to produce the PM mean
- 2) For the m th grid point, sort the EnMean values from lowest to highest within a patch whose center is at the m th grid point, and store the rank R_{mean} of the center grid point (e.g., for a patch encompassing 100 points,

- $R_{\text{mean}} = 90$ means the value at the center grid point is greater than the values at 90 other points within the patch)
- 3) Sort the values of all N ensemble members from lowest to highest for grid points within the above patch
- 4) Replace the precipitation value at the m th grid point with the ensemble member value whose rank in the ensemble is R_{ens} . R_{ens} is calculated as

$$R_{\text{ens}} = \text{nint} \left[\frac{-NL(L - R_{\text{mean}})^\sigma}{L^\sigma} \right] + NL, \quad (1)$$

where L is the number of grid points within the patch, and $\sigma = 1.05$ is a coefficient that determines the deviation of R_{ens} from the simple linear rank. As with the traditional (unlocalized) PM mean, if the EnMean at the m th point is zero, the LPM mean is forced to zero at that point. According to [Clark \(2017\)](#), the expression in the function $\text{nint}[\]$ can be regarded as a normalized exponential, “where the exponential term R_{mean}^σ is normalized by L/L^σ [and] the function is ‘flipped’ [by taking the negative (i.e., the $-N$ term)] and ‘reversed’ by using $(N - R_{\text{mean}})^\sigma$ instead of R_{mean}^σ .” The introduction of σ was based on the finding that there were low biases when the linear calculation $R_{\text{ens}} = NR_{\text{mean}}$ was used.

Steps 2 to 4 are conducted for every grid point in the computational domain, and the LPM max can be produced by substituting the EnMax for the EnMean in the first two steps. The optimal value for σ was given by [Clark \(2017\)](#).

To generate LPM products, the patch radius in step 2 has to be prescribed. In the case of the patch radius approaching zero, both the LPM max and LPM mean converge to the EnMax. In contrast, when the radius is sufficiently large, the LPM max is more like the PM max and the LPM mean is more like the PM mean. Considering the above circumstances, the radius cannot be too large or too small, and we used a radius of 120 km, close to the optimal value obtained by [Clark \(2017\)](#).

To ensure the radius of 120 km was computationally affordable, we examined the computational cost to produce the LPM and PM products. In general, implementations of the PM and LPM approaches in this work have computational efficiencies comparable to [Clark \(2017\)](#). However, the actual time to produce PM and LPM products depends both on the programming language used to code the algorithm and other computational choices.

3. Data and metrics

a. Observation data

We produced PM and LPM products from hourly accumulated precipitation forecasts provided by the

National Center for Atmospheric Research (NCAR) convection-allowing ensemble (Schwartz et al. 2015, 2019). The ensemble forecasts had 10 members with 3-km horizontal grid spacing spanning the entire conterminous United States (CONUS), were integrated to 48 h, and were initialized at 0000 UTC from a continuously cycling, 80-member, 15-km ensemble adjustment Kalman filter (Anderson 2001, 2003) data assimilation system. All 3-km ensemble members used version 3.6.1 of the Advanced Research version of the Weather Research and Forecasting Model (WRF-ARW; Skamarock et al. 2008; Powers et al. 2017) and had common physical parameterizations. No cumulus parameterization was used. NCAR ensemble forecast output is available from NCAR's Research Data Archive (<https://rda.ucar.edu/datasets/ds300.0/>).

Gridded stage IV (ST4) observations (Lin and Mitchell 2005) from the National Centers for Environmental Prediction (NCEP) were used for precipitation verification (available from <https://www.emc.ncep.noaa.gov/mmb/ylin/pcpanl/stage4/>). All 1–36-h forecasts from the 2017 NCAR ensemble (364 cases, 1 January–30 December) were used for quantitative analysis. To compare to observations, the 3-km precipitation forecasts were interpolated to the ST4 grid using a distance-weighted interpolation: at each ST4 grid point, an interpolated forecast value was calculated using data at the four model grid points closest to each ST4 grid point weighted by the inverse of the distance between the model and ST4 grid locations. Both bilinear and budget interpolation methods (Accadia et al. 2003) were also tested and yielded nearly identical results as the distance-weighted approach. Following Schwartz et al. (2015), a verification domain spanning 30°–45°N, 82°–105°W was selected to ensure ST4 observations were robust and far from lateral boundaries. Although 48-h forecasts were available, this work only used hourly accumulated precipitation forecasts within the first 36 h to generate PM and LPM products.

b. Evaluation methods

To evaluate EnMean and EnMax precipitation placement, as well as their associated PM products, we computed fractions skill scores (FSSs; Roberts and Lean 2008), as is common for high-resolution QPF evaluation (Roberts and Lean 2008; Mittermaier and Roberts 2010; Schwartz and Liu 2014; Schwartz et al. 2014). The FSS is a neighborhood approach that measures spatial skill. A perfect forecast has an FSS of 1.0, while a no-skill forecast has an FSS = 0. FSSs can be computed for every precipitation forecast, but it is usually more informative to aggregate FSSs across many forecasts. So, we computed aggregate FSSs both for individual forecast hours and periods spanning multiple forecast hours. The hourly

scores show evolution of forecast performance, while aggregates over all 36 forecast hours succinctly evaluate the overall performance of the forecasts.

Although the EnMean and EnMax have identical areal coverages of nonzero precipitation, within precipitation areas event placement may substantially differ and the EnMean and EnMax have obviously different PDFs. Because reassignment in the PM procedure is based on *locations* of events within a reference field (i.e., EnMean or EnMax), to fully understand PM products it is necessary to understand spatial performance of the reference field.

Thus, we examined differences regarding spatial characteristics and skill between the EnMean and EnMax. To do so, we used precipitation percentile thresholds to account for the very different PDFs of the EnMean and EnMax, which allows for a robust examination of spatial skill without contamination from bias (e.g., Roberts and Lean 2008; Mittermaier and Roberts 2010; Schwartz and Liu 2014; Gowan et al. 2018). For the k th precipitation percentile threshold, $(100 - k)$ % of grid points have precipitation values larger than the physical threshold corresponding to the k th precipitation percentile threshold (e.g., if the physical threshold corresponding to the 95th precipitation percentile is 10.0 mm h^{-1} , 5% of points have precipitation $>10.0 \text{ mm h}^{-1}$). Precipitation percentile thresholds and their corresponding physical thresholds were computed for hourly accumulated precipitation across the verification domain following Schwartz and Liu (2014).

To determine the significance of the difference between any two products, the resampling procedure described by Hamill (1999) was used. The resampling was repeated 1000 times and the significance was determined by the 95% confidence interval. FSSs were computed separately for each season because of seasonal diversity regarding synoptic forcing that causes variations of precipitation amounts and predictability over the CONUS (Schwartz et al. 2019).

4. Spatial distributions and skill of the EnMean and EnMax

According to the average of one year statistics (Fig. 1), for a given precipitation percentile, the corresponding EnMax physical thresholds were always higher than the EnMean physical thresholds (i.e., EnMax precipitation was highest for a chosen precipitation percentile), as expected. EnMax precipitation percentiles were always much higher than those observed. For the 94th precipitation percentile threshold and below, EnMean precipitation percentiles were higher than observed precipitation percentiles (Figs. 1a–c), and for the 98th precipitation

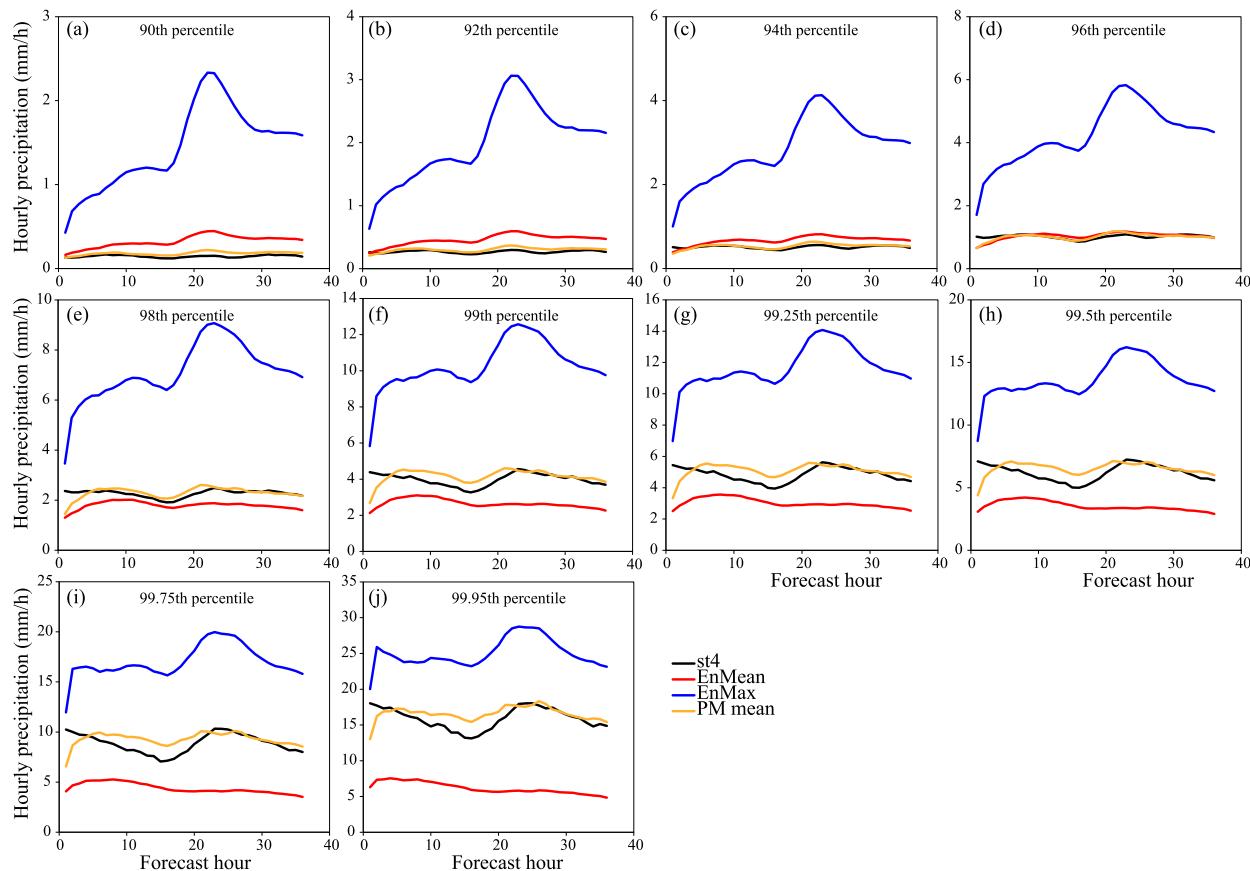


FIG. 1. The relationship between the hourly precipitation amount and the corresponding precipitation percentile as a function of forecast hour; data shown in plots are averaged over 364 cases and are computed within the verification domain.

percentile threshold and above, EnMean precipitation percentiles were lower than those observed (Figs. 1e–j), consistent with EnMean tendencies to overly smooth and decrease maximum magnitudes. However, PM mean precipitation percentiles, which reflect the ensemble PDF, well matched ST4 precipitation percentiles, indicating individual ensemble members captured the observed precipitation frequency.

Obviously, neither the EnMean nor EnMax magnitudes were appropriate compared to observations. However, for PM purposes, their magnitudes are irrelevant; only their spatial representations matter, and Fig. 1 clearly shows that precipitation percentile thresholds must be used to evaluate the relative spatial performance of EnMean and EnMax precipitation forecasts because of their very different PDFs.

a. Ensemble overlap and implications for EnMean and EnMax precipitation percentile distributions

Inspired by Fig. 8a of Hamill et al. (2017), we used a schematic diagram (Fig. 2) to demonstrate the EnMean and EnMax for a number of scenarios differing by the

overlap among individual ensemble members. Their figure showed a typical medium-range ensemble forecast where precipitation peaks of various ensemble members differed in positions and magnitudes, which forecasters might often see. In this work, five scenarios were designed with four members whose maximum precipitation amounts slightly differed. In each scenario, the largest amount corresponded to a precipitation percentile of 100% (the maximum precipitation value). Unlike the precipitation percentiles that were computed across the verification domain, the precipitation percentile values in Fig. 2 were computed within a one-dimensional domain with 80 grid points and were only computed for the EnMax and EnMean. There are some grid points with the same precipitation amounts; to assign different percentile values for those points, we determined the precipitation percentile values from large to small according to the locations from right to left (e.g., assuming precipitation values at grid points 35 and 55 were identical, the larger precipitation percentile was assigned to point 55). Note that only 4 members were used in this schematic diagram, and we refer to the

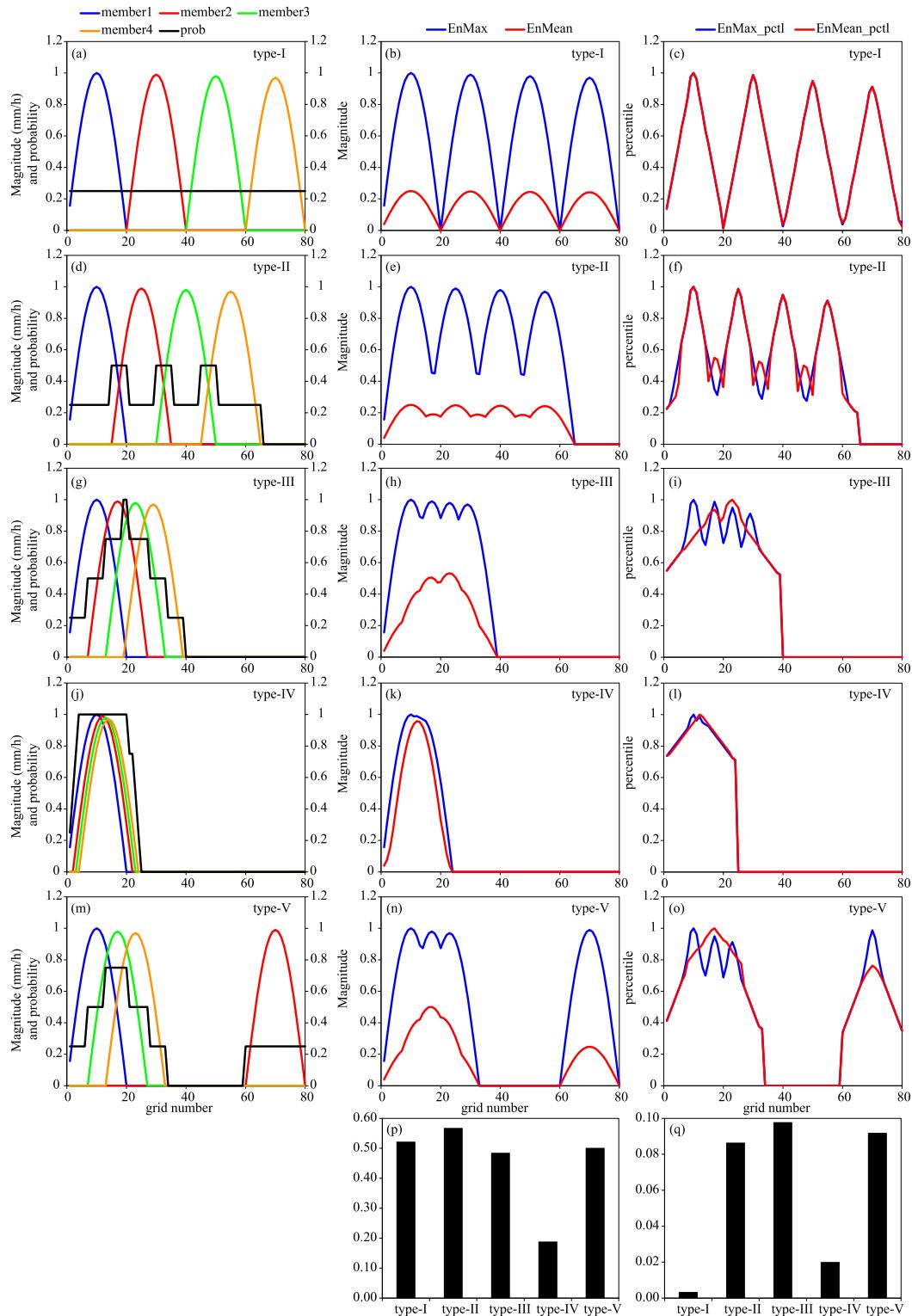


FIG. 2. (a)–(q) The schematic plots of (left) overlap types of members, (middle) the corresponding EnMax and EnMean, and (right) the precipitation percentiles of the EnMax and EnMean, where forecast probabilities at the 0.0 mm h^{-1} threshold are shown in black solid lines in the left panels. The root-mean-square differences (RMSD) between the EnMax and EnMean magnitudes and the corresponding differences of precipitation percentiles are shown in (p) and (q).

probability of 0% (no overlap) and 25% (two members overlap) as low probability, 50% and 75% as moderate/relatively high probability, and 100% (all members overlap) as very high probability.

Type-I (Fig. 2a) is an extreme example with no overlap between members, so at all locations the probability of precipitation $>0.0 \text{ mm h}^{-1}$ is approximately 25%. Although the EnMean and EnMax precipitation amount in type-I pronouncedly differ (Figs. 2b,p), the corresponding precipitation percentiles are very similar (Figs. 2c,q) because relative positions of low and high amounts within the EnMean and EnMax fields are identical. The tiny difference between the EnMean and EnMax precipitation percentiles was because of assigning successive ranks to equal values. In addition, the computational truncation is also attributable to the tiny difference. Type-IV (Fig. 2j) represents another extreme where members highly overlap, leading to a probability of 100% throughout most of the nonzero precipitation area. Given this strong overlap, the EnMean and EnMax are similar in terms of both their absolute magnitudes and precipitation percentiles (Figs. 2k,l,p,q).

These two extreme scenarios illustrate that precipitation percentile differences between the EnMean and EnMax are small when the probability throughout the nonzero precipitation area is very high (100%) or very low. Therefore, for these extreme situations, the similar EnMean and EnMax precipitation percentile distributions implies the PM mean and PM max may be similar and have comparable forecast skill.

In contrast, precipitation percentile differences between the EnMean and EnMax are larger for less extreme overlap situations that are likely more realistic. For type-II, where members slightly overlap (Fig. 2d), precipitation amount differences between the EnMean and EnMax are large (Figs. 2e,p) while precipitation percentile differences are similar except for when probability is >0.5 (Fig. 2f), where EnMax reaches a local minimum whereas EnMean has a subtle local maximum (Fig. 2e). For type-III and type-V that have larger but still moderate probabilities, even more precipitation percentile differences between the EnMean and EnMax appear (Figs. 2g–i, m–o). Type-III represents a situation where moderate overlap occurs throughout the nonzero precipitation area while type-V represents the coexistence of moderate overlap and an outlier. In both of these circumstances, spatial placement differences between the EnMean and EnMax are largest (Figs. 2i,o,q); there are three peaks of EnMax precipitation percentiles between grid points 0 and 30 with only one peak in the same area for EnMean precipitation percentiles (Figs. 2i,o).

Collective results in Fig. 2 indicate that spatial placements of the EnMean and EnMax substantially differ

when members moderately overlap, while differences are small when members rarely or fully overlap. Therefore, there may be little difference between the PM mean and PM max when members highly agree or disagree, but we can expect larger differences between them during more realistic situations when moderate overlap among members dominates.

b. Assessment of precipitation placement

FSSs for 1-h accumulated precipitation aggregated over all 36 h and the verification domain indicate FSSs increased with radius of influence (γ) and decreased with threshold (Fig. 3). For the 92nd precipitation percentile threshold and below, which represents light, stratiform precipitation (Figs. 1a,b), for most seasons and γ , the EnMean often had significantly higher FSSs than the EnMax (Figs. 3a,c,e,g). Conversely, in all seasons for most γ , the EnMax had significantly higher FSSs than the EnMean for the 96th precipitation percentile threshold and above (Fig. 3), corresponding to observed precipitation greater than approximately 1.0 mm h^{-1} (Figs. 1d–j) that includes convection. Significant differences were most noticeable at thresholds exceeding the 98th precipitation percentile in winter and spring (Figs. 3a–d), with smaller, but still often significant, differences in summer and autumn (Figs. 3e–h).

Further analysis revealed more precipitation spread in summer and autumn than in spring and winter (not shown), consistent with weak forcing, and hence, less predictability and lower probabilities. In terms of Figs. 2c and 2f, when low probabilities prevail, differences between EnMax and EnMean precipitation percentiles are small, consistent with smaller differences between the two products in the summer and autumn (Figs. 3e–h). In contrast, during winter and spring, stronger synoptic-scale forcing is associated with greater predictability, less spread, and higher probabilities, and according to Figs. 2g–i and 2m–o, for moderate ensemble member overlap, differences between EnMax and EnMean spatial placement in terms of precipitation percentiles become larger. Thus, the idealized scenarios are consistent with the largest differences between EnMean and EnMax FSSs occurring in winter and spring.

Overall, the EnMax performed comparably to or worse than the EnMean for lower precipitation percentile thresholds but clearly outperformed the EnMean at higher precipitation percentile thresholds, implying that the EnMax provided more accurate precipitation placement for smaller-scale, less common events, like convection. These findings suggest using the EnMax as the basis for reassignment in PM methods is appropriate, as the EnMax spatial representation of events was

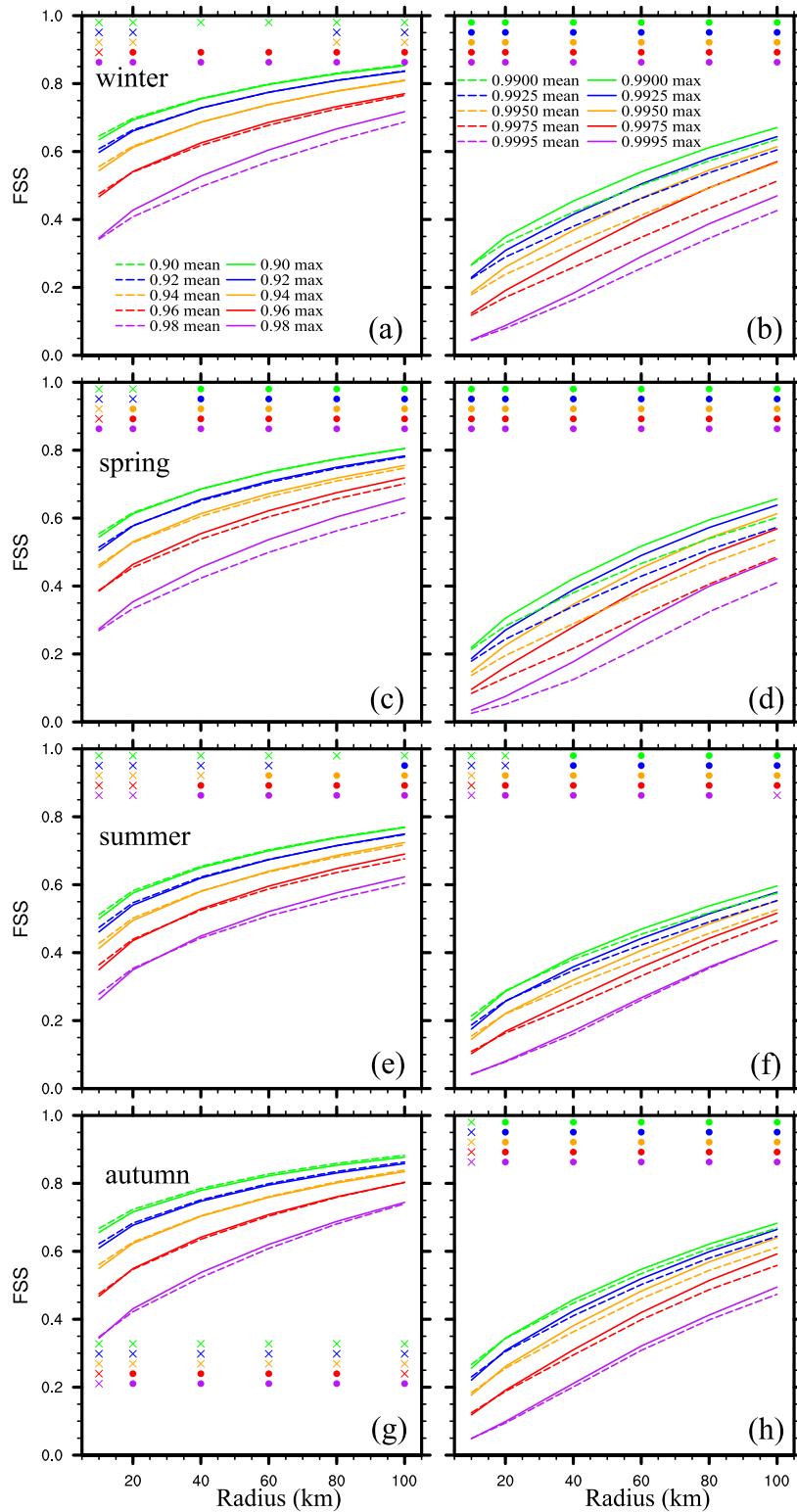


FIG. 3. The seasonal aggregated FSSs over 36 forecast hours as a function of radius of influence (km) for 1-h accumulated precipitation at (left) lower and (right) higher precipitation percentile thresholds. Seasons are defined as follows: (a),(b) winter (1 Jan–14 Mar and 15–30 Dec); (c),(d) spring (15 Mar–14 Jun); (e),(f) summer (15 Jun–14 Sep);

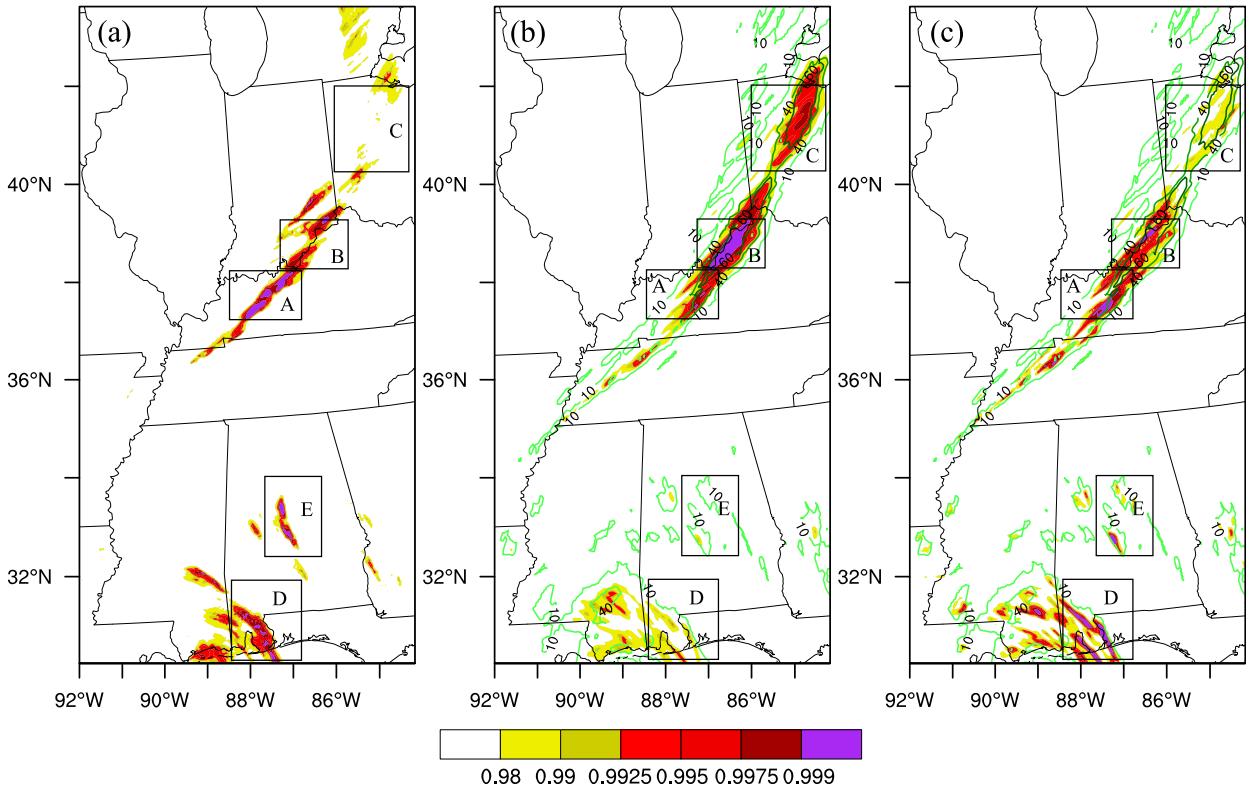


FIG. 4. Probabilities of 1-h accumulated precipitation (green contours of 10%, 40%, and 60%) at the 5 mm h^{-1} threshold and precipitation percentiles (filled colors) at 0300 UTC 8 Oct 2017 for the (a) ST4 observations, (b) EnMean, and (c) EnMax.

reasonable and better than the EnMean for most precipitation percentiles, including those encompassing moderate and heavy precipitation.

c. Example case: 8 October 2017

In addition to quantitative analysis, we performed a qualitative evaluation of precipitation percentiles (Fig. 4) to understand why EnMax FSSs were higher than EnMean FSSs at and above the 96th precipitation percentile threshold. The precipitation percentiles were computed for the ST4 observations, EnMax, and EnMean across the verification domain mentioned in section 3a. A frontal precipitation case that occurred in the evening of 8 October 2017 (UTC) and spanned several states was selected. The frontal precipitation band was accompanied by a tropical depression near the Gulf of Mexico coast. During this case period,

a shortwave trough within zonal 500-hPa flow progressed eastward, ahead of which warm and highly moist low-level southwest flow provided a favorable environment for a squall line, which moved through portions of Michigan, Ohio, Indiana, Kentucky, and Tennessee and produced 1-h rainfall amounts greater than 25 mm h^{-1} . During that evening, rainfall associated with the tropical depression was responsible for the maximum precipitation of approximately 29 mm h^{-1} .

Observed precipitation exceeding the 99.9th precipitation percentile threshold at 0300 UTC 9 October 2017 (purple areas) was mostly distributed in areas A, D, E, and the north border of area B (Fig. 4a). The highest precipitation percentile appeared in area E, corresponding to the maximum observed 1-h rainfall. In area C, observed precipitation at most grid points did not reach the 98th precipitation percentile threshold.

←

(g),(h) autumn (15 Sep–14 Dec). The EnMax and EnMean are shown with solid and dashed lines, respectively. Near the top of each plot are markers that represent FSS values of the EnMax being significantly higher than (solid circles), or significantly lower than (crosses), or comparable to (blank) those of the EnMean, where markers and the corresponding lines are common in colors.

Spatial distributions of both EnMean and EnMax precipitation percentiles (Figs. 4b,c) differed in details and degree to which they matched observed precipitation percentiles. The spatial placement of EnMean precipitation percentiles approximately followed the forecast probabilities (high precipitation percentiles were associated with high probabilities), while EnMax precipitation percentiles were not closely associated with forecast probabilities (e.g., area C), consistent with expectations that the EnMean and EnMax emphasize and deemphasize agreement among members, respectively.

However, the association between high EnMean precipitation percentiles and high probabilities was not typically favorable. For example, in area B, the EnMean produced substantial coverage of precipitation percentiles >0.9975 within the 60% probability area (Fig. 4b), but these coverages were much larger than those in the EnMax (Fig. 4c) and observations (Fig. 4a). Similar behavior occurred in area C, where EnMean precipitation percentiles were too high whereas EnMax precipitation percentiles were substantially lower and closer to those observed. In low forecast probability ($\sim 10\%$) areas (D and E), the EnMean missed the observed high precipitation percentiles, while they were better captured in the EnMax.

In general, EnMax precipitation percentiles were closest to those observed for this case. Furthermore, this case suggests that higher EnMax FSSs for precipitation percentile thresholds ≥ 0.96 (Fig. 3) may be partly attributable to improved EnMax performance in areas with relatively high probabilities, as suggested by Figs. 2g and 2m, and further explored in the next subsection.

d. Statistical relationship between precipitation percentiles and forecast probabilities

Figure 4 suggests the EnMean tended to have high precipitation percentiles in areas with high forecast probabilities and low precipitation percentiles in low forecast probability areas. Furthermore, for high precipitation percentiles, it appears that the EnMax precipitation percentile distribution (placement and local areal coverage) was more similar to that observed than the EnMean, although the EnMax was noisier than the EnMean. Moreover, as long as one member in the ensemble predicts relatively heavy precipitation at a grid point, the EnMax tends to assign a high precipitation percentile to that grid point, and this characteristic of the EnMax appears to benefit forecast skill for high precipitation percentile thresholds (Figs. 3 and 4).

To systematically determine whether situations as in Fig. 4 often occurred, we produced one year of aggregate statistics of the relationship between forecast probabilities of precipitation $>5.0 \text{ mm h}^{-1}$ and ST4

observations, EnMean, and EnMax precipitation percentiles (Fig. 5). Given that forecast probability was determined solely by the ensemble, the number of samples in each probability bin was identical for all products.

Overall, bounds of the interquartile ranges (IQRs; distance between the 25th and 75th percentiles¹ of the boxplot distribution) for the EnMean, EnMax, and observed precipitation percentiles increased with probability for all seasons, suggesting high precipitation percentiles (heavy precipitation) were associated with high probabilities (Fig. 5). However, compared to the EnMean, EnMax IQRs in most probability bins were broader, indicating ensemble probabilities were not as closely associated with high precipitation percentiles as in the EnMean, which had the narrowest IQRs. ST4 observations had much wider IQRs than the EnMax and EnMean, indicating a weaker relationship between observations and forecast probabilities, which seems sensible due to greater independence between observed and forecast quantities than between two forecast quantities (i.e., the EnMean/EnMax and ensemble probabilities). Similar results were obtained when using different probabilistic event definitions for the x axis of Fig. 5 (not shown).

Because high EnMean precipitation percentiles were strongly associated with high ensemble probabilities, it was natural to consider quality of the probabilistic forecasts, especially in regards to ensemble spread and placement of probabilistic events. So, reliability diagrams were constructed (Fig. 6), which revealed an overconfident ensemble, particularly in high probability bins, consistent with other studies indicating the NCAR ensemble was spread-deficient (Schwartz et al. 2015, 2019; Gowan et al. 2018). Thus, it appears high probability events were often incorrect, which was related to insufficient ensemble spread; rank histograms also documented insufficient spread (not shown) and the mean annual ratio of domain average ensemble spread to root-mean-square error was approximately 0.4 for 1-h precipitation forecasts. An example of insufficient ensemble spread in a high probability area is observed precipitation at a grid point $<1 \text{ mm h}^{-1}$ but all ensemble members predict precipitation $>10 \text{ mm h}^{-1}$ with values ranging from 11.0 to 20.0 mm h^{-1} , which corresponds to a probability of 100% at the 10 mm h^{-1} threshold. In this situation, the root-mean-square error is $>10 \text{ mm h}^{-1}$ (ensemble mean $> 10.0 \text{ mm h}^{-1}$ and observation $< 1.0 \text{ mm h}^{-1}$), but the ensemble spread is $<10.0 \text{ mm h}^{-1}$, which is insufficient. In other words, the above high probability appears in the wrong place and there is insufficient ensemble spread to

¹ These percentiles measure the distribution of the precipitation percentiles in each probability bin in Fig. 5.

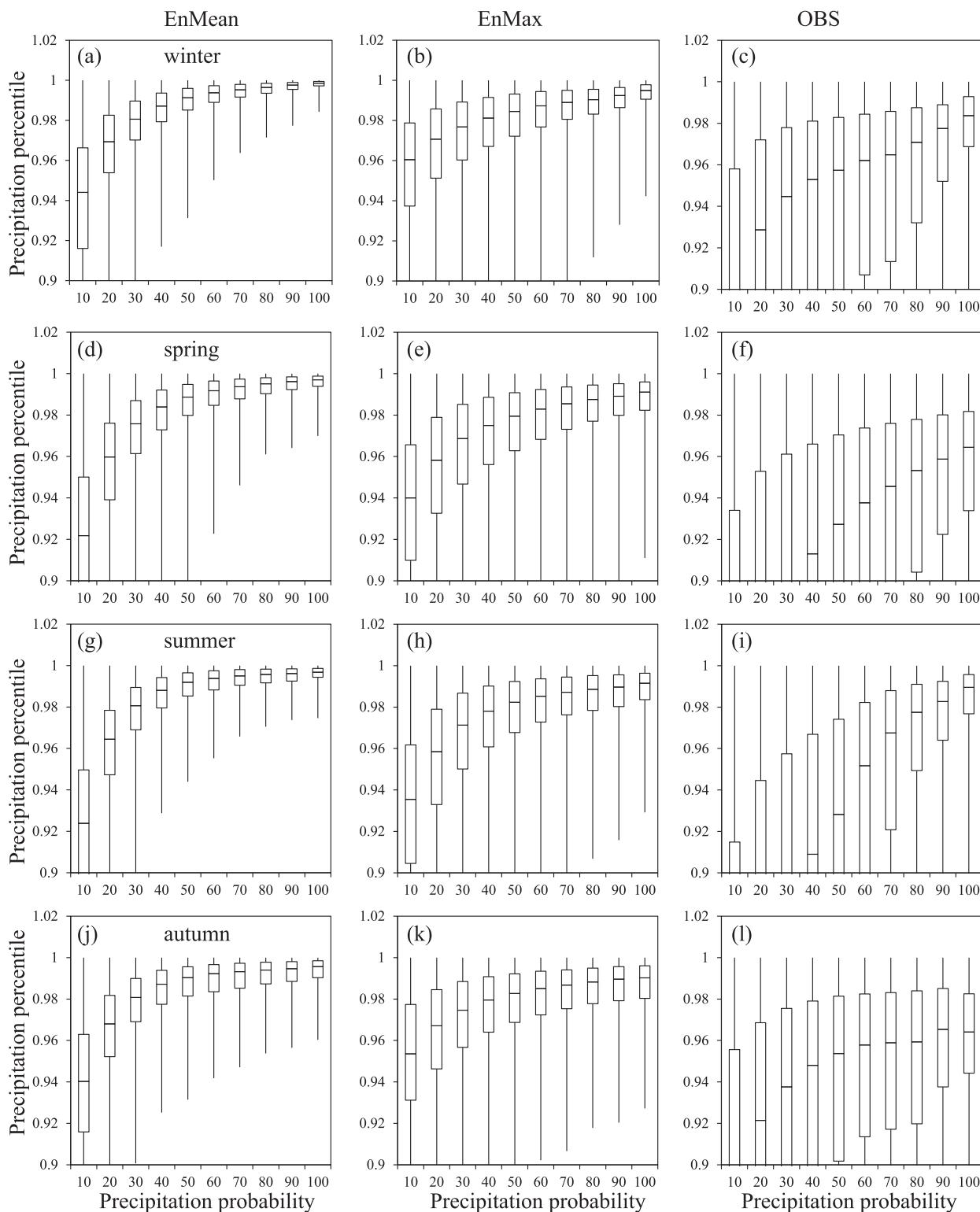


FIG. 5. The distribution of aggregated precipitation percentiles as a function of probabilities of 1-h accumulated precipitation at the 5.0 mm h⁻¹ threshold for the (a),(d),(g),(j) EnMean; (b),(e),(h),(k) EnMax; and (c),(f),(i),(l) ST4 observations. In each season and each probability bin, precipitation percentiles within the verification domain are gathered over all 36 forecast hours. Each boxplot depicts the distribution of the precipitation percentile of precipitation amount in the corresponding probability bin.

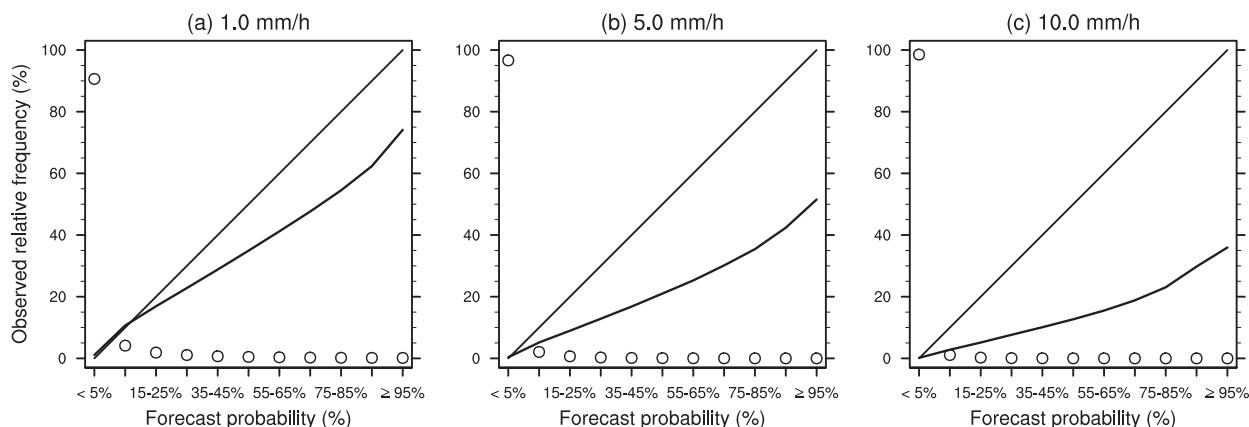


FIG. 6. Reliability diagrams for the (a) 1.0, (b) 5.0, and (c) 10.0 mm h^{-1} precipitation thresholds aggregated over all 364 hourly 1–36-h forecasts based on gridscale probabilities. The diagonal line represents perfect reliability, and the forecast frequencies (%) within each probability bin are shown as open circles.

encompass the observed precipitation value. So, because high EnMean precipitation percentiles were associated with high probabilities that were often misplaced (Fig. 5), high EnMean precipitation percentiles were also often misplaced. Overall, the relatively poor EnMean performance at most precipitation percentiles appears related to poor ensemble spread.

However, the EnMax is less tied to the probability distribution than the EnMean (Fig. 5) because it is less sensitive to ensemble spread. This EnMax characteristic appears beneficial for high precipitation percentile forecasts (Fig. 3) and justifies use of a spatial field for PM that does not highly follow the probability distribution of spread-deficient ensembles.

These findings are important to document because most high-resolution ensembles have poor reliability for precipitation in high probability bins (Duc et al. 2013; Schwartz et al. 2015; Hagelin et al. 2017; Schwartz et al. 2019). Thus, in spread-deficient ensembles where EnMean precipitation is strongly related to forecast probabilities (as in Fig. 5), the EnMax appears to better represent spatial placement of uncommon events. It is unclear whether similar advantages for the EnMax may occur in either well-calibrated ensembles or ensembles with too much spread.

5. Evaluation of the PM and LPM products

The above results indicate that the EnMax is a credible field to use as the basis for reassignment in the PM procedure. Next, we investigate whether using the EnMax translates into better PM and LPM products than when using the EnMean. Because the PM mean and PM max have identical precipitation frequencies and only differ regarding spatial placement, we used physical thresholds ranging from 1.0 to 20 mm h^{-1} to

compare performance of the PM mean and PM max. Although using physical thresholds does not control for bias in the FSS, because PM mean and PM max biases are identical, differences between their FSSs can be attributed to different spatial placements.

a. Forecast skill of the PM products

Figure 7 shows aggregate FSSs as a function of forecast hour for 1-h accumulated precipitation for $\gamma = 60$ km. At most forecast hours, FSSs for both PM products decreased as threshold increased. For thresholds ≥ 5 mm h^{-1} , the PM max usually outperformed the PM mean, especially in winter and spring (Figs. 7a,b), while in summer and autumn, benefits of the PM max were mainly for thresholds ≥ 10 mm h^{-1} (Figs. 7c,d). The generally better PM max performance compared to the PM mean for thresholds ≥ 5 mm h^{-1} was likely due to better spatial placement of precipitation in the EnMax at high precipitation percentiles (Fig. 3). Conversely, for the 1 mm h^{-1} threshold, the PM mean was better than the PM max at most forecast hours in all seasons (Fig. 7), consistent with poorer performance of the EnMax at lower precipitation percentiles (Figs. 3a,c,e,g).

Since the FSS is sensitive to neighborhood radius (γ), we investigated FSSs of the PM products with respect to γ (Fig. 8). As expected, FSSs of both PM products increased with γ . Higher FSSs were obtained by the PM max at thresholds ≥ 10 mm h^{-1} in all seasons for most γ . This advantage of the PM max was also evident for the 5 mm h^{-1} threshold in winter and spring for all γ and in summer for $\gamma \geq 60$ km.

For light precipitation (1 mm h^{-1}), the PM max was significantly worse than the PM mean in winter, summer, and autumn for all γ (Fig. 8), again consistent with poorer performance of the EnMax compared to the EnMean at low

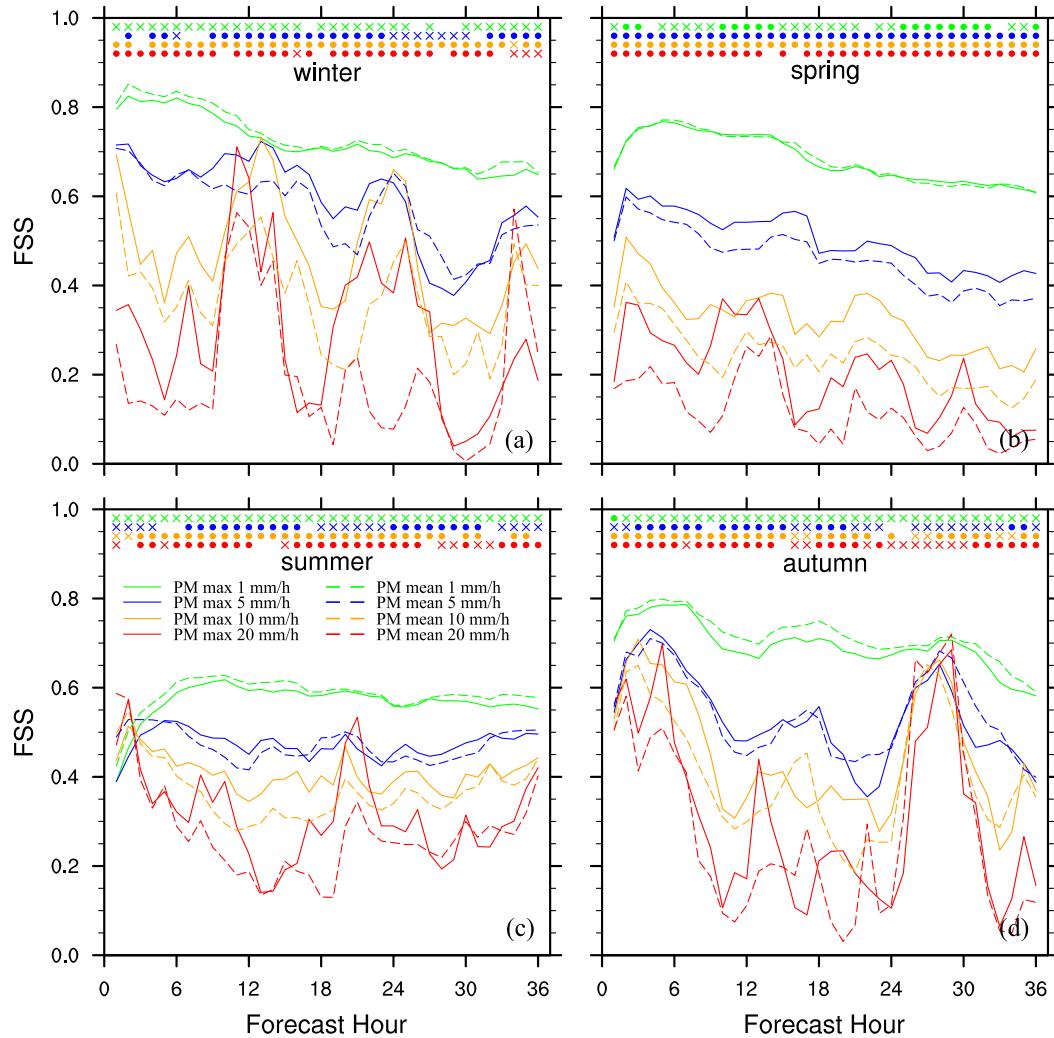


FIG. 7. Seasonal aggregated FSSs at a scale of 60 km as a function of forecast hour for 1-h accumulated precipitation at thresholds of 1 (green), 5 (blue), 10 (orange), and 20 (red) mm h^{-1} over the verification domain. Solid lines and dashed lines denote the PM max and PM mean, respectively. Near the top of each plot are markers that represent FSS values of the PM max being significantly higher than (solid circles), or significantly lower than (crosses), or comparable to (blank) those of the PM mean, where markers and the corresponding lines are common in colors.

precipitation percentiles. One possible cause of the PM max inferiority at the 1 mm h^{-1} threshold is that the PM mean well captures light, stratiform precipitation, which often falls over broad areas that can be reasonably represented by the smooth EnMean, while the PM max, based upon the noisier EnMax field, may not be beneficial for widespread, light precipitation events. Another plausible cause is that the global reassignment nature of the PM mean and PM max hurts the PM max for light precipitation, which is examined next by evaluating the LPM products.

b. Forecast skill of the LPM products

The LPM products were also evaluated with the FSS (Figs. 9 and 10). Compared to Figs. 7 and 8, the biggest

difference is that the LPM max outperformed the LPM mean for the 1 mm h^{-1} threshold in all seasons. This result implies that the PM method of reassigning precipitation amounts across the entire forecast domain caused the lower PM max FSSs compared to the PM mean for the 1 mm h^{-1} threshold.

In addition, for thresholds $\geq 5 \text{ mm h}^{-1}$, LPM max FSSs were usually significantly higher than LPM mean FSSs (Figs. 9 and 10), although differences between the LPM max and LPM mean were not as large as those between the PM max and PM mean. This result was expected because the localization aspect of the LPM products effectively means there are fewer options to where values can be re-assigned, and by definition, reflects a more local product.

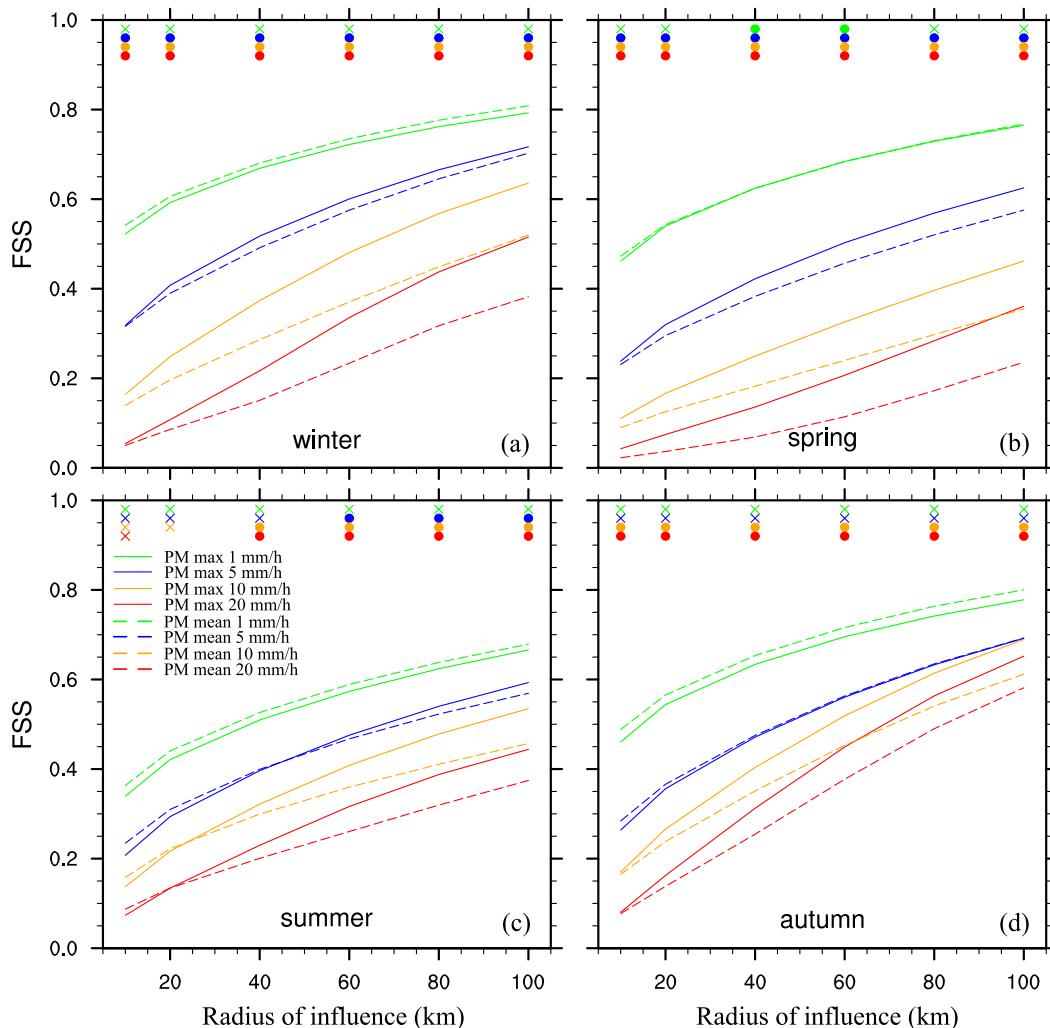


FIG. 8. As in Fig. 3, but for aggregated FSS values calculated over all 36 forecast hours in a season and as a function of the influence radius (km).

These findings are encouraging and suggest that using a better reference field as the basis for reassignment in LPM products is also beneficial. However, because the LPM max and LPM mean have different precipitation frequencies, biases could have impacted the FSSs. But, when FSSs were computed using precipitation percentile thresholds, LPM max still outperformed LPM mean for all seasons (not shown), implying that improved spatial placement was associated with higher LPM max FSSs in Figs. 9 and 10.

6. Conclusions

This work highlighted the importance of spatial placement in the PM method and designed and evaluated PM products based on the EnMax. All ensemble QPF products were generated from NCAR's 3-km convection-allowing ensemble over 364 cases. Spatial skill of the EnMean,

EnMax, PM mean, and PM max were evaluated with the FSS.

According to the results in sections 4 and 5, several conclusions can be drawn:

- 1) Precipitation placement of the EnMax matched ST4 observations better than the EnMean for precipitation greater than the 96th precipitation percentile threshold, suggesting the EnMax had more accurate placement for relatively uncommon events with high local magnitudes, like convection. However, the EnMean outperformed the EnMax for the lowest precipitation thresholds.
- 2) The distribution of EnMean precipitation percentiles closely followed the distribution of forecast probabilities, which often resulted in enlargement of local areal coverages of high precipitation percentiles (>0.9975) within high forecast probability areas

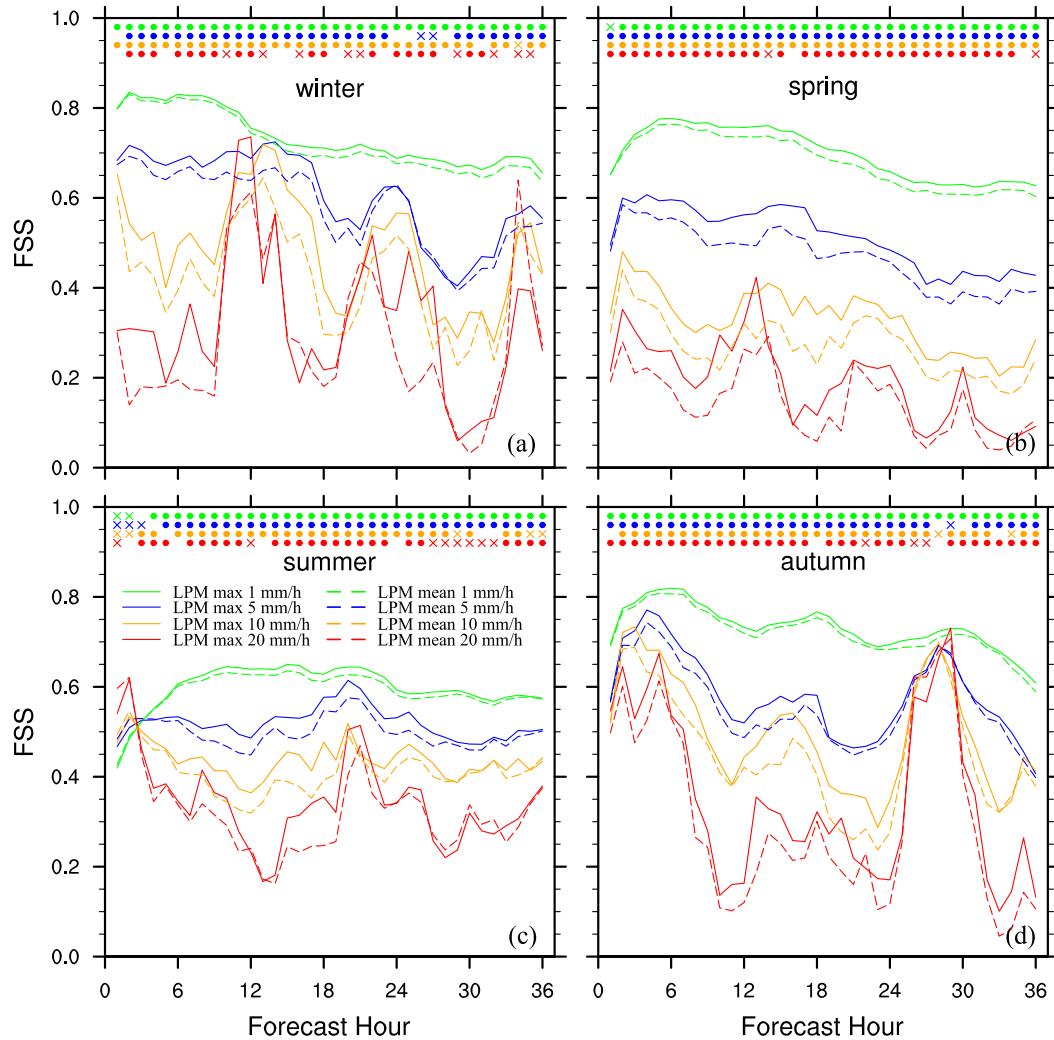


FIG. 9. As in Fig. 7, but for the LPM mean and LPM max.

(>60%) and fewer high precipitation percentiles in low probability areas (~10%). Given this EnMean correspondence between high precipitation percentiles and probabilities and the fact that the ensemble had insufficient spread and often misplaced areas of high probabilities, it appears that the spread deficiency of the ensemble contributed to the relatively poor EnMean placement. In contrast, the EnMax precipitation percentile distribution was not as closely tied to the forecast probability distribution, as the EnMax deemphasizes agreement among members. Therefore, using the EnMax as the basis for spatial placement may be attractive for spread-deficient high-resolution ensembles, particularly for rare events.

- 3) The PM max outperformed the PM mean for precipitation rates $\geq 5 \text{ mm h}^{-1}$ in winter and spring and precipitation rates $\geq 10 \text{ mm h}^{-1}$ in all seasons.

- However, the PM mean had higher FSSs for light precipitation (1 mm h^{-1} threshold). These results are consistent with the relative skill of the EnMean and EnMax and demonstrate that spatial skill of the reference field has a great impact on performance of PM products, because the only difference between the PM mean and PM max is the spatial placement.
- 4) The LPM max outperformed the LPM mean in all seasons at almost all thresholds and forecast hours, including at light precipitation thresholds. This finding suggests that the reassignment across the entirety of a large domain caused the lower PM max FSSs compared to the PM mean at the 1 mm h^{-1} threshold.

Our results are consistent with previous findings that precipitation displacement is the dominant source of QPF error (Ebert and McBride 2000; Cuo et al. 2011), especially for convective rainfall.

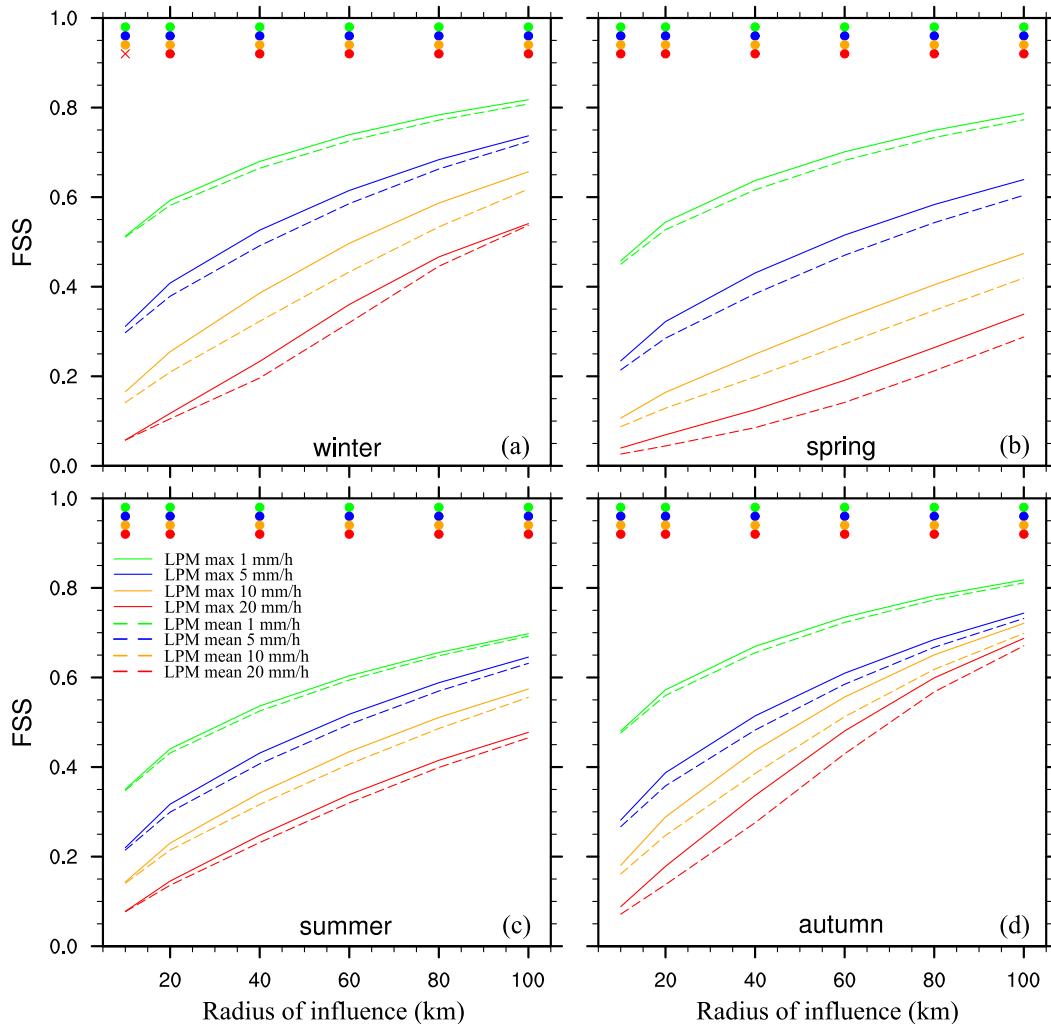


FIG. 10. As in Fig. 8, but for the LPM mean and LPM max.

Despite promise of the EnMax for PM applications, we acknowledge that the EnMax may not be the optimal spatial representation of precipitation, and other spatial reference fields may further improve performance of PM products. Thus, PM products based on the 80th and 90th precipitation percentiles of members' precipitation values (hereinafter PM80 and PM90) were also preliminarily examined, inspired by Zhang (2018), who utilized forecast distributions of the 90th precipitation percentile in a deterministic product derived from ensemble output. While performances of the PM80 and PM90 products were comparable to or better than PM mean performance at almost all thresholds, they performed worse than the PM max for thresholds $\geq 10 \text{ mm h}^{-1}$ but better than PM max for the 1 mm h^{-1} threshold, similar to the difference between PM mean and PM max. PM80 and PM90 also obtained higher FSSs for the 90th–98th precipitation percentile thresholds than PM mean,

implying that EnMax is not the only choice and the optimal spatial representation for PM methods needs further investigation.

In conclusion, the PM max appears to be a useful deterministic product derived from ensemble output, especially for ensembles with insufficient spread and for heavier precipitation events. However, because performance of PM and LPM products rely on ensemble forecast systems' representation of precipitation location, precipitation frequency, and member overlap, whether the PM max and LPM max are still more skillful than EnMean-associated PM products in other ensembles is worth studying.

Acknowledgments. Thanks to Glen Romine, Ryan Sobash, and Kate Fossell of the NCAR Ensemble team [NCAR/Mesoscale and Microscale Meteorology Laboratory (MMM)] for their efforts in running the

NCAR ensemble. This work was jointly sponsored by the National Key Research and Development Program of China (2017YFC1502103), the National Natural Science Foundation of China (41505089, 41875129, 41505090, 41430427, and 41805070), and National Key Research and Development Program of China (2018YFC1506404). We appreciate the constructive comments from three anonymous reviewers. NCAR is sponsored by the National Science Foundation.

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).
- Anagnostou, E. N., A. J. Negri, and R. F. Adler, 1999: A satellite infrared technique for diurnal rainfall variability studies. *J. Geophys. Res.*, **104**, 31 477–31 488, <https://doi.org/10.1029/1999JD900157>.
- Ancell, B. C., 2013: Nonlinear characteristics of ensemble perturbation evolution and their application to forecasting high-impact events. *Wea. Forecasting*, **28**, 1353–1365, <https://doi.org/10.1175/WAF-D-12-00090.1>.
- Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903, [https://doi.org/10.1175/1520-0493\(2001\)129<2884:AEAKFF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2).
- , 2003: A local least squares framework for ensemble filtering. *Mon. Wea. Rev.*, **131**, 634–642, [https://doi.org/10.1175/1520-0493\(2003\)131<0634:ALLSFF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<0634:ALLSFF>2.0.CO;2).
- 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 now-casting. *Mon. Wea. Rev.*, **140**, 2689–2705, <https://doi.org/10.1175/MWR-D-11-00181.1>.
- Ciach, G. J., W. F. Krajewski, and G. Villarini, 2007: Product-error-driven uncertainty model for probabilistic quantitative precipitation estimation with NEXRAD data. *J. Hydrometeorol.*, **8**, 1325–1347, <https://doi.org/10.1175/2007JHM814.1>.
- 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 Jr., and T.-C. Chen, 2008: Contributions of mixed physics versus perturbed initial/lateral boundary conditions to ensemble-based precipitation forecast skill. *Mon. Wea. Rev.*, **136**, 2140–2156, <https://doi.org/10.1175/2007MWR2029.1>.
- , —, M. Xue, and F. Kong, 2009: A comparison of precipitation forecast skill between small convection-allowing and large convection-parameterizing ensembles. *Wea. Forecasting*, **24**, 1121–1140, <https://doi.org/10.1175/2009WAF2222222.1>.
- , 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>.
- Cuo, L., T. C. Pagano, and Q. J. Wang, 2011: A review of quantitative precipitation forecasts and their use in short- to medium-range streamflow forecasting. *J. Hydrometeorol.*, **12**, 713–728, <https://doi.org/10.1175/2011JHM1347.1>.
- Du, J., S. L. Mullen, and F. Sanders, 1997: Short-range ensemble forecasting of quantitative precipitation. *Mon. Wea. Rev.*, **125**, 2427–2459, [https://doi.org/10.1175/1520-0493\(1997\)125<2427:SREFOQ>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<2427:SREFOQ>2.0.CO;2).
- Duc, L., K. Saito, and H. Seko, 2013: Spatial-temporal fractions verification for high-resolution ensemble forecasts. *Tellus*, **65A**, 18171, <https://doi.org/10.3402/tellusa.v65i0.18171>.
- Ebert, E., 2001: Ability of a poor man's ensemble to predict the probability and distribution of precipitation. *Mon. Wea. Rev.*, **129**, 2461–2480, [https://doi.org/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).
- , and J. McBride, 2000: Verification of precipitation in weather systems: Determination of systematic errors. *J. Hydrol.*, **239**, 179–202, [https://doi.org/10.1016/S0022-1694\(00\)00343-7](https://doi.org/10.1016/S0022-1694(00)00343-7).
- Evans, C., D. F. V. 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, <https://doi.org/10.1175/WAF-D-13-00064.1>.
- Fang, X., and Y.-H. Kuo, 2013: Improving ensemble-based quantitative precipitation forecasts for topography-enhanced typhoon heavy rainfall over Taiwan with a modified probability-matching technique. *Mon. Wea. Rev.*, **141**, 3908–3932, <https://doi.org/10.1175/MWR-D-13-00012.1>.
- Gowan, T. M., W. J. Steenburgh, and C. S. Schwartz, 2018: Validation of mountain precipitation forecasts from the convection-permitting NCAR ensemble and operational forecast systems over the western United States. *Wea. Forecasting*, **33**, 739–765, <https://doi.org/10.1175/WAF-D-17-0144.1>.
- Hagelin, S., J. Son, R. Swinbank, A. McCabe, N. Roberts, and W. Tennant, 2017: The Met Office convective-scale ensemble, MOGREPS-UK. *Quart. J. Roy. Meteor. Soc.*, **143**, 2846–2861, <https://doi.org/10.1002/qj.3135>.
- 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).
- , E. Engle, D. Myrick, M. Peroutka, C. Finan, and M. Scheuerer, 2017: The U.S. national blend of models for statistical post-processing of probability of precipitation and deterministic precipitation amount. *Mon. Wea. Rev.*, **145**, 3441–3463, <https://doi.org/10.1175/MWR-D-16-0331.1>.
- Hollan, M. A., and B. C. Ancell, 2015: Ensemble mean storm-scale performance in the presence of nonlinearity. *Mon. Wea. Rev.*, **143**, 5115–5133, <https://doi.org/10.1175/MWR-D-14-00417.1>.
- Huang, L., and Y. Luo, 2017: Evaluation of quantitative precipitation forecasts by TIGGE ensembles for south China during the presummer rainy season. *J. Geophys. Res. Atmos.*, **122**, 8494–8516, <https://doi.org/10.1002/2017JD026512>.
- Kong, F., and Coauthors, 2009: A real-time storm-scale ensemble forecast system: 2009 Spring Experiment. *23rd Conf. on Weather Analysis and Forecasting/19th Conf. on Numerical Weather Prediction*, Omaha, NB, Amer. Meteor. Soc., 16A.3, https://ams.confex.com/ams/23WAF19NWP/techprogram/paper_154118.htm.
- Krajewski, W. F., and G. J. Ciach, 2003: Towards probabilistic quantitative precipitation WSR-88D algorithms: Preliminary studies and problem formulation. NWS Office of Hydrologic Development Rep., 59 pp.
- Leith, C., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, **102**, 409–418, [https://doi.org/10.1175/1520-0493\(1974\)102<0409:TSOMCF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1974)102<0409:TSOMCF>2.0.CO;2).
- Lin, Y., 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>.
- Mittermaier, M., and N. Roberts, 2010: Intercomparison of spatial forecast verification methods: Identifying skillful spatial scales using the fractions skill score. *Wea. Forecasting*, **25**, 343–354, <https://doi.org/10.1175/2009WAF2222260.1>.
- Mullen, S. L., and R. Buizza, 2001: Quantitative precipitation forecasts over the United States by the ECMWF ensemble prediction system. *Mon. Wea. Rev.*, **129**, 638–663, [https://doi.org/10.1175/1520-0493\(2001\)129<0638:QPFOTU>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0638:QPFOTU>2.0.CO;2).
- Murphy, A. H., 1988: Skill scores based on the mean square error and their relationships to the correlation coefficient. *Mon. Wea. Rev.*, **116**, 2417–2424, [https://doi.org/10.1175/1520-0493\(1988\)116<2417:SSBOTM>2.0.CO;2](https://doi.org/10.1175/1520-0493(1988)116<2417:SSBOTM>2.0.CO;2).
- Novak, D. R., C. Bailey, K. F. Brill, P. Burke, W. A. Hogsett, R. Rausch, and M. Schichtel, 2014: Precipitation and temperature forecast performance at the Weather Prediction Center. *Wea. Forecasting*, **29**, 489–504, <https://doi.org/10.1175/WAF-D-13-00066.1>.
- Potvin, C. K., E. M. Murillo, M. L. Flora, and D. M. Wheatley, 2017: Sensitivity of supercell simulations to initial-condition resolution. *J. Atmos. Sci.*, **74**, 5–26, <https://doi.org/10.1175/JAS-D-16-0098.1>.
- Powers, J. G., and Coauthors, 2017: The Weather Research and Forecasting Model: Overview, system efforts, and future directions. *Bull. Amer. Meteor. Soc.*, **98**, 1717–1737, <https://doi.org/10.1175/BAMS-D-15-00308.1>.
- 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>.
- Rosenfeld, D., D. B. Wolff, and D. Atlas, 1993: General probability-matched relations between radar reflectivity and rain rate. *J. Appl. Meteor.*, **32**, 50–72, [https://doi.org/10.1175/1520-0450\(1993\)032<0050:GPMRBR>2.0.CO;2](https://doi.org/10.1175/1520-0450(1993)032<0050:GPMRBR>2.0.CO;2).
- Schwartz, C. S., and Z. Liu, 2014: Convection-permitting forecasts initialized with continuously cycling limited-area 3DVAR, ensemble Kalman filter, and “hybrid” variational–ensemble data assimilation systems. *Mon. Wea. Rev.*, **142**, 716–738, <https://doi.org/10.1175/MWR-D-13-00100.1>.
- , G. S. Romine, K. R. Smith, and M. L. Weisman, 2014: Characterizing and optimizing precipitation forecasts from a convection-permitting ensemble initialized by a mesoscale ensemble Kalman filter. *Wea. Forecasting*, **29**, 1295–1318, <https://doi.org/10.1175/WAF-D-13-00145.1>.
- , —, R. A. Sobash, K. R. Fossell, and M. L. Weisman, 2015: NCAR’s experimental real-time convection-allowing ensemble prediction system. *Wea. Forecasting*, **30**, 1645–1654, <https://doi.org/10.1175/WAF-D-15-0103.1>.
- , —, —, —, and —, 2019: NCAR’s real-time convection-allowing ensemble project. *Bull. Amer. Meteor. Soc.*, **100**, 321–343, <https://doi.org/10.1175/BAMS-D-17-0297.1>.
- Seo, D.-J., S. Perica, E. Welles, and J. Schaake, 2000: Simulation of precipitation fields from probabilistic quantitative precipitation forecast. *J. Hydrol.*, **239**, 203–229, [https://doi.org/10.1016/S0022-1694\(00\)00345-0](https://doi.org/10.1016/S0022-1694(00)00345-0).
- 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>.
- Snook, N., F. Kong, K. A. Brewster, M. Xue, K. W. Thomas, T. A. Supinie, S. Perfater, and B. Albright, 2019: Evaluation of convection-permitting precipitation forecast products using WRF, NMMB, and FV3 for the 2016–17 NOAA hydrometeorology testbed flash flood and intense rainfall experiments. *Wea. Forecasting*, **34**, 781–804, <https://doi.org/10.1175/WAF-D-18-0155.1>.
- Speer, M., and L. Leslie, 1997: An example of the utility of ensemble rainfall forecasting. *Aust. Meteor. Mag.*, **46**, 75–78.
- Surcel, M., I. Zawadzki, and M. Yau, 2014: On the filtering properties of ensemble averaging for storm-scale precipitation forecasts. *Mon. Wea. Rev.*, **142**, 1093–1105, <https://doi.org/10.1175/MWR-D-13-00134.1>.
- Warner, T. T., 2010: *Numerical Weather and Climate Prediction*. Cambridge University Press, 526 pp.
- Xue, M., and Coauthors, 2011: Realtime convection-permitting ensemble and convection-resolving deterministic forecasts of CAPS for the Hazardous Weather Testbed 2010 Spring Experiment. *24th Conf. on Weather and Forecasting/20th Conf. on Numerical Weather Prediction*, Seattle, WA, Amer. Meteor. Soc., 9A.2, <https://ams.confex.com/ams/91Annual/webprogram/Paper183227.html>.
- Yang, J., M. Astitha, and C. S. Schwartz, 2019: Assessment of storm wind speed prediction using gridded Bayesian regression applied to historical events with NCAR’s real-time ensemble forecast system. *J. Geophys. Res. Atmos.*, **124**, 9241–9261, <https://doi.org/10.1029/2018JD029590>.
- Zhang, X., 2018: Application of a convection-permitting ensemble prediction system to quantitative precipitation forecasts over southern China: Preliminary results during SCMREX. *Quart. J. Roy. Meteor. Soc.*, **144**, 2842–2862, <https://doi.org/10.1002/qj.3411>.