Further Evaluation of Probabilistic Convective Precipitation Forecasts Using the QPF–PoP Neighborhood Relationship

MICHAEL C. KOCHASIC
NOAA/NWS/Weather Forecast Office, Sacramento, California

WILLIAM A. GALLUS JR.
Department of Geological and Atmospheric Sciences, Iowa State University, Ames, Iowa

CHRISTOPHER J. SCHAFFER
NOAA/NWS/Southeast River Forecast Center, Peachtree City, Georgia

(Manuscript received 30 December 2016, in final form 16 May 2017)

ABSTRACT

A neighborhood postprocessing approach that relates quantitative precipitation forecasts (QPF) to probability of precipitation (PoP) forecasts applied to a single model run was found by Schaffer et al. to be as good as traditional ensemble-based approaches using 10 members in 30-h forecasts of convective precipitation. The present study evaluates if PoP forecasts derived from additional variations of the approach can improve PoP forecasts further compared with previous methods. Ensemble forecasts from the Center for Analysis and Prediction of Storms (CAPS) are used for neighborhood tests comparing a single model run and a traditional ensemble. In the first test, PoP forecasts for different combinations of training and testing datasets using a single model member with 4-km grid spacing are compared against those obtained with a 10-member traditional ensemble. Overall, forecasts for the neighborhood approach with just one member are only slightly less accurate to those using a more traditional neighborhood approach with the ensemble. PoP forecasts improve when using older data for training and newer data for testing. Assessments of the sensitivity of the neighborhood PoPs suggest that thinning of the horizontal grid at fine grid spacing is an effective way of maintaining the accuracy of PoP forecasts while reducing computational expenses. In an additional test, the diurnal variation of the forecast is examined on a day-by-day basis, showing good agreement between the two approaches for all but a few cases during 2008.

1. Introduction

Numerous probability of precipitation (PoP) forecast generation approaches exist, with the simplest approach considering the agreement between members of an ensemble prediction system (EPS). If there are 10 ensemble members, for example, with two showing precipitation exceeding a specified threshold amount, then the PoP for that threshold is 20%. This approach is referred to as the uncalibrated traditional approach (denoted Uncali_trad hereafter) in Schaffer et al. (2011, hereafter SGS11). Another postprocessing approach applies a point-based EPS calibration using observations (referred to in SGS11 as the calibrated traditional approach, or Cali_trad), which improves PoP forecast accuracy over the Uncali_trad approach (Hamill and Colucci 1997; Hamill and Whitaker 2006).

Another approach to PoP creation is precipitation binning in a model forecast, which creates a reliability-based calibration (Atger 2001; Gallus and Segal 2004, hereafter GS04; Wilks 2006; Gallus et al. 2007). This approach is applied by placing the forecast into a bin based on a quantitative precipitation forecast (QPF) range, and then determining the observed frequency for forecasts in that bin during a training period. Once computed, the observed frequency is then used as the PoP forecast (Zhu et al. 1996). Although skill scores for warm season convective QPFs are slowly improving (Gallus 2002; Ralph et al. 2005; Schwartz et al. 2009; Mariani et al. 2015; see http://www.emc.ncep.noaa.gov/mmb/ylin/pcpverif/scores/ for current scores), errors...
associated with convective warm season QPFs can be highly nonlinear (Hohenegger and Schär 2007) and generally remain large (more so in measures like the equitable threat score than in bias). To help address this issue, placing model precipitation into bins allows for model forecasts to identify the grid points where atmospheric processes are more likely to result in at least some precipitation, while acknowledging that the specific amounts are probably not accurate. GS04 applied precipitation binning to a single deterministic forecast, and the resulting PoP forecasts were better able to correctly identify areas that observed at least some precipitation rather than correctly identify the precipitation amounts for these areas (referred to as the GS approach hereafter). Using the GS approach, Gallus et al. (2007) found the same QPF–PoP relationship existed when using the coarser grid spacing of the former operational Eta and Aviation Models over a longer time period. Similar conclusions were reached by Ruiz et al. (2009).

More recent approaches to forecasting PoPs have used a neighborhood, or an area surrounding a grid point (Theis et al. 2005; Roberts and Lean 2008; Ebert 2009; Schwartz et al. 2010; SGS11; Johnson and Wang 2012, hereafter JW12; Ruiz and Saulo 2012; among others). These studies demonstrated that the neighborhood of points around a grid point could be considered to establish a “virtual ensemble” for which an ensemble PoP forecast could be obtained from a single deterministic model run. In SGS11, skill scores in 30-h simulations were comparable or superior to the 10-member Cali_trad approach when SGS11 used the neighborhood approach combined with the GS approach (referred to in SGS11 and hereafter as Ave_nbh). Similar results were obtained also in JW12 as well as in Ruiz and Saulo (2012). Many of these approaches likely improve skill in part because they filter out unpredictable detail, which can reduce random errors (Surecel et al. 2014).

The persistent trend toward increased horizontal grid spacing and improved physics in model simulations of warm season convection makes the traditional EPS potentially computationally expensive in operational use. Although operational convection-allowing EPSs are feasible and in fact are being used at some large operational centers (Barthold et al. 2015), smaller governmental forecast offices or private weather forecasting firms may not have the computing resources for convection-allowing EPSs. Therefore, any postprocessing approach of a single model run that provides PoP forecasts that are competitive with traditional PoP approaches based on an ensemble is extremely attractive as a result of its computational cost effectiveness, and further comparisons of Cali_trad and Ave_nbh are of unique merit. However, as pointed out by JW12 and Johnson et al. (2014), conclusions obtained in studies using coarser grid spacing are not necessarily valid for simulations using relatively fine horizontal grid spacing, especially considering differences in how convection is handled (parameterized vs explicitly resolved).

The goal of the present study is to further evaluate the Ave_nbh approach from SGS11, focusing on 4-km horizontal grid output, particularly to determine if for PoP forecasts derived from additional datasets, the Ave_nbh approach variations can provide even better PoP forecasts for spring convection than those shown in the 20-km verification completed by SGS11. Specific details about the variations of the Ave_nbh approach that is tested are described in section 2. Results using Brier scores (BSs) are presented in section 3. A comprehensive discussion of the model results and a conceptual evaluation is included in section 4, with conclusions presented in section 5.

2. Methodology

2a. General information

Weather Research and Forecasting (WRF) Model runs that were generated as part of the NOAA Hazardous Weather Testbed (HWT) Spring Experiments and distributed by the Center for the Analysis and Prediction of Storms (CAPS) are used for the present study, as in SGS11. Additional precipitation output from 2010 is added to the 2007 and 2008 output used in SGS11. SGS11 primarily averaged the native CAPS 4-km data onto a 20-km grid for testing, with limited analysis using 4-km output. The present research exclusively focuses on 4-km output with a total of 76 cases used: 20 from 2007, 29 from 2008, and 27 from 2010.

The CAPS high-resolution ensemble members differed in each of the experiment years, although the WRF-ARW model dynamic core was used in each year (Skamarock et al. 2008). The 2007 simulations were initialized at 2100 UTC for 33 h (Kong et al. 2007). The ensemble initial conditions consisted of a mixture of bred perturbations coming from the 2100 UTC Short Range Ensemble Forecast (SREF) perturbed members and physics variations (grid-scale microphysics, land surface, and PBL physics), along with a control run. The initial conditions and lateral boundary conditions are described in Kong et al. (2007). The spatial extent of the model domain was approximately 3000 km × 2500 km (Fig. 1 in Kong et al. 2007).

The 2008 and 2010 CAPS experiments additionally used available WSR-88D data that were assimilated...
through ARPS three-dimensional variational data assimilation (3DVAR) into the model ensemble members. Different initial perturbations and physics schemes were used (Xue et al. 2008; Kong et al. 2010). The 2008 and 2010 CAPS datasets were initialized at 0000 UTC and ran for 30h. Although the domains in all years covered the eastern 7/3 of the continental United States, the exact size differed (2008 covered 3600 km × 2700 km; 2010 covered 3400 km × 2700 km). The 2010 CAPS experiment had an expanded dataset in comparison to 2007 and 2008 with the ensemble members increasing from 10 to 26. The main improvement for the 2010 CAPS experiment was using an updated dynamic core [WRF version 3.1.1 (WRFV3.1.1)] compared with 2007 and 2008, which used WRFV2.2 (Kong et al. 2010; Xue et al. 2008). The National Centers for Environmental Prediction (NCEP) Stage IV precipitation observation dataset (Baldwin and Mitchell 1997) was used for verification.

Because the 2007 CAPS dataset was initialized at a different time than the 2008 and 2010 datasets, the first 3 h of the 2007 dataset are not included in our evaluations. Because of the slightly different domain sizes, a common subdomain from SGS11 is used in the present study for testing and verification that covers an area of 1980 km × 1840 km (Fig. 1, from SGS11). As in SGS11, the Stage IV and forecast data are interpolated onto a common Midwest subdomain 4-km grid (Fig. 1) using NCEP procedures that conserve the amount of precipitation within the subdomain (Fig. 1). The number of ensemble members for 2010 is larger than for 2007 and 2008, but the approach in the current study uses only one member to generate the neighborhood-calibrated PoP forecasts from each model year and then tests the neighborhood-calibrated PoP results on one model member from a different year. A static approach in which a fixed PoP from a dataset that does not change (as opposed to a dynamic dataset that changes based on the current forecast date) is used with the creation of the static PoP table during the training process (see Ruiz et al. (2009) for elaboration). The PoP table is created once during the training process, and the same set of PoP forecasts is used for reference and is assigned to all future forecasts, which results in the PoP forecasts for the current study being considered static. The PoP forecasts are not calculated for every test forecast dataset, which would be considered dynamic. The Ave_nbh approach used a single model member from the 2007, 2008, and 2010 datasets; the cn member is used for 2007, c0 for 2008, and s4cn for 2010.

Some studies (SGS11; Ruiz and Saulo 2012) found advantages to combining the neighborhood and traditional ensemble approaches compared with using the neighborhood approach alone. In SGS11, a third approach used a neighborhood on a traditional ensemble to calculate PoP results, and the fourth approach averaged the PoP forecasts produced by numerous post-processing methods to create a “superensemble.” These methods proposed by SGS11 showed that BSs could be improved relative to both the Ave_nbh and Cali_trad simulation by combining Ave_nbh with a traditional ensemble. However, SGS11 also noted that for large neighborhood sizes and numerous model forecasts, the forecast scenarios would become very large and could have a negative impact on the efficiency and reliability of the approach. An advantage of using Ave_nbh alone (relative to Cali_trad) is in the increased efficiency, making additional investigations into Ave_nbh worthwhile.

Following SGS11, a neighborhood of points (see Fig. 2) of various sizes is used to create 2D PoP tables (see Table 1) by determining the 1) average QPF within the neighborhood of points and 2) the fraction of points in the neighborhood with QPF surpassing a threshold of interest. The average neighborhood QPF for each grid point is placed into a precipitation bin. As in GS04, seven separate QPF bins are used for each neighborhood size: <0.01, 0.01–0.05, 0.05–0.10, 0.10–0.25, 0.25–0.50, 0.50–1.00, and >1.0 in. After the average neighborhood QPF from each case of data is computed, success ratios [referred to as the correct alarm ratio in SGS11, also called the success ratio; see Roebber (2009)] are calculated for each of the seven bin ranges as $h(i, f)$, where $h$ is the number of “hit” points where the observed precipitation exceeded a specified threshold (0.01, 0.10, and 0.25 in. are used in this study, as in SGS11) when the average neighborhood QPF fell within one of the seven QPF bins. The variable $f$ is the total number of grid points with forecast precipitation within a bin range (Table 1), as well as the same neighborhood fraction of points, regardless of whether it surpassed a given threshold. Once the “hit” points $h$ are divided by the total number of points $f$ for each bin, the success ratio becomes the PoP forecast for a particular observed QPF threshold and neighborhood percent agreement of points exceeding the observed threshold for one of the seven QPF bins. This two-parameter approach (observed QPF threshold and neighborhood percent agreement of observed threshold exceedance) calibrates a PoP forecast, which is calculated as just a fraction of the grid points with precipitation above an observed threshold and the mean neighborhood precipitation within the neighborhood. Two-dimensional PoP tables provide the discrete joint probability for a given training year and neighborhood size using Ave_nbh based on two parameters: 1) the forecasted precipitation amount
within a bin (the GS approach) and 2) the percentage of neighborhood points forecasting exceedance of the precipitation above a threshold amount. In summary, 2D PoP tables are created following these steps:

1) For each of the seven forecast QPF bins, two categories of precipitation are recorded (precipitation occurred with an amount at or above an observed threshold, or it did not).

2) Within each of the seven forecast QPF bins and for each possible neighborhood percentage agreement (at how many points in the neighborhood was precipitation forecast in the QPF bin), the number of hit points (those forecast points where QPF occurred at or above an observed threshold, or \( h \)) is divided by the total number of forecasts (bin points, or \( f \)).

3) The success ratio (\( h/f \)) becomes the PoP forecast given the forecast QPF threshold amount and the fraction of all points in the neighborhood forecast to have precipitation.

For example, consider the QPF in a conceptual 3 × 3 neighborhood (Fig. 2, adapted from SGS11). In this example, six of the nine points within the neighborhood have QPFs greater than or equal to 0.01 in. In addition, the average QPF of approximately 0.06 in. falls within the 0.05–0.10-in. bin. When the training process is completed, a PoP table (Table 1) is produced. When these tables are compiled for numerous days and time periods within a dataset, there are typically a number of neighborhoods that report similar combinations of the two parameters of grid points that exceed a certain number of points (agreement) and average precipitation in a QPF bin. Referring to Table 1, the PoP that would be forecast based on this example scenario is 43.8%, which is the fraction of these similar forecasts that actually had precipitation above the specified threshold in the training dataset for that combination of parameters.

GS04 found that if the mean forecast precipitation within the neighborhood is high, the chance of receiving no precipitation is reduced. Conversely, if just one neighborhood point has precipitation, and just a small amount, one may expect the precipitation chances to be less likely than if all neighborhood points agreed to have a small amount of precipitation. If one member was chosen from an EPS that consistently performed poorly when having a single parameter neighborhood applied.
to it (similar to Theis et al. 2005), the resulting neighborhood PoP forecast would also perform poorly. Single-parameter approaches are more susceptible to errors derived from the original model forecast, whereas a two-parameter approach (such as QPF binning) provides a calibration that can increase accuracy. Therefore, it is more useful for accurate PoP forecasts to use the two-parameter Ave_nbh approach.

The Ave_nbh method is compared with the point-based Cali_trad approach, as was done in SGS11 (not shown), as well as the Cali_trad approach using a neighborhood (called Cali_trad_nbh hereafter), with both using a traditional EPS with application of a calibration to generate PoP forecasts. The Cali_trad_nbh approach is used as the primary benchmark for comparison here since studies like Schwartz et al. (2010) have shown that neighborhood techniques applied to ensembles improve the skill of the PoP forecasts compared to calibration at just a point, as was done in the Cali_trad approach used in SGS11. The calibration is applied by training with observed data to incorporate bias corrections into the PoP forecasts, which leads to forecast improvements. The traditional EPS in this study uses the 10 available CAPS model members for the 2007, 2008, and 2010 CAPS datasets.

Specifically in the present study, after Cali_trad PoP values are computed at every grid point, the Cali_trad_nbh method is developed by assigning to each point in the forecast domain a PoP value averaged from all of the points over the neighborhood size that results in the best skill (where the neighborhood size varied). It is acknowledged that there are multiple ways that a neighborhood technique could be developed from a traditional EPS. For the present study, this forecast merely serves as a benchmark to which Ave_nbh can be compared.

Note that the notation TRvsTE, which will be used hereafter, refers to the time intervals (either one 6-h period or a 30-h period consisting of five 6-h periods) used for training (TR) and testing (TE). For example, 30vs6 indicates that all five 6-h periods (from each 30-h model integration) are used for the training process and a specific 6-h period is used for testing. The TR and TE notation will also refer to which datasets were used for training and testing. For example, the 2008 data used for TR and TE using the 2010 dataset will be depicted as 08TR10TE.

Neighborhood sizes at 4 km for this study are selected to have the same areal extent as the sizes used for the 20-km Ave_nbh approach used in SGS11, which evaluated neighborhoods as large as $17 \times 17$. The neighborhood size tested in the current study, therefore, calculates the Ave_nbh from the $5 \times 5$ to the $85 \times 85$ neighborhood sizes at 4 km ($17 \times 17$ at 20 km and $85 \times 85$ at 4 km, both of which have a search radius of 170 km). The neighborhood size that minimizes the computed BS in the TR data is used in the TE computations for both Ave_nbh and Cali_trad_nbh. The results that follow are valid for the best-performing neighborhood, unless otherwise mentioned.

### Evaluations conducted in the study

Although static TR is likely to be performed separately from the weather prediction model simulations in an operational environment, it would still be useful for reducing the computational costs associated with this process. SGS11 cautioned about high computational costs for TR over the 4-km horizontal grid. To resolve this issue, thinning of the Ave_nbh approach in the current study is applied to both the TR and TE data in order to speed up the computational time needed to make the Ave_nbh approach more attractive for operational forecasting. The same neighborhood sizes at 4 km mentioned earlier (from $5 \times 5$ through $85 \times 85$ points) were used for thinning. Several thinning levels are examined using the Ave_nbh approach and demonstrate that using a dataset that includes only every fifth grid point in the neighborhood square maintains the

### Table 1. The 2D PoP table (PoP %) created from the thinned Ave_nbh approach using the QPF bin range (top row) and the number of points in a $3 \times 3$ neighborhood to meet or exceed precipitation above a given threshold of 0.01 in. (left column).

<table>
<thead>
<tr>
<th></th>
<th>&lt;0.01</th>
<th>0.01–0.05</th>
<th>0.05–0.10</th>
<th>0.10–0.25</th>
<th>0.25–0.50</th>
<th>0.50–1.0</th>
<th>&gt;1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10.6</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>1</td>
<td>38.1</td>
<td>26.4</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2</td>
<td>38.5</td>
<td>34.6</td>
<td>9.5</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>3</td>
<td>41.4</td>
<td>40.7</td>
<td>36.8</td>
<td>25.0</td>
<td>100.0</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>4</td>
<td>40.1</td>
<td>43.3</td>
<td>34.2</td>
<td>27.1</td>
<td>0.0</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>5</td>
<td>36.5</td>
<td>43.2</td>
<td>39.7</td>
<td>29.9</td>
<td>15.0</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>6</td>
<td>31.1</td>
<td>44.5</td>
<td>43.8</td>
<td>42.5</td>
<td>37.9</td>
<td>0.0</td>
<td>—</td>
</tr>
<tr>
<td>7</td>
<td>0.0</td>
<td>46.3</td>
<td>46.7</td>
<td>44.7</td>
<td>35.2</td>
<td>17.6</td>
<td>—</td>
</tr>
<tr>
<td>8</td>
<td>—</td>
<td>47.9</td>
<td>48.4</td>
<td>47.4</td>
<td>42.6</td>
<td>34.8</td>
<td>42.9</td>
</tr>
<tr>
<td>9</td>
<td>—</td>
<td>55.8</td>
<td>62.4</td>
<td>70.0</td>
<td>78.4</td>
<td>83.1</td>
<td>88.3</td>
</tr>
</tbody>
</table>
general accuracy with slight degradation of the computed BS (less than 0.0050 worsening in most cases), while the computational resources decline sharply (e.g., when TE with an \(85 \times 85\) neighborhood size is used on the 2008 data employing a virtualized Linux workstation with 98 GB of memory and 16 Intel cores, thinning reduces the CPU time needed to 9% of the level required without thinning for this study). Note that the spatial area for the neighborhood sizes \((N \times N)\) corresponding to the 4-km output is unaffected by the thinning. Even using a large \(85 \times 85\) neighborhood size for the 4-km thinned neighborhood, PoP forecasts are generated in about 6 min of wall-clock time on the virtualized Linux workstation described previously. This is much faster than running another NWP ensemble member at the same resolution, which would take several hours on the same Linux workstation. The Ave_nbh approach was limited in the present study to a minimum size of \(5 \times 5\) points for the neighborhood calculations. As demonstrated by the cross-validation TE in this study, PoP tables do not need to be reproduced again when using the PoP results on newer model data, making the technique even more advantageous in an operational setting.

Two primary tests are performed using this thinned version of the Ave_nbh approach. The first examines the impacts of the use of different years of data for TR and TE with the newer 2010 dataset, including the 2007 and 2008 datasets used in SGS11. These tests investigate the sensitivity to the size of the TR datasets by combining the 2007 and 2008 datasets for TR and using the 2010 output for TE (0708TR10TE). Precipitation accumulation for the TR and TE datasets is considered separately for each 6-h period instead of using only the 30-h averages, as was done in SGS11. Tabular presentation of the results is adopted to help in the comparison of the relevant differences between the two approaches. Although not shown in the tables, the Ave_nbh and Cali_trad_nbh approaches are also compared to Uncali_trad to see if improvement was obtained.

The second test involves a day-by-day comparison of the thinned Ave_nbh and Cali_trad_nbh approaches to gain better insight into the impact of forecast difficulty on the performance of the two techniques in a daily forecast that would occur in an operational environment. All cases were averaged together in the results in SGS11, which did not allow for daily comparison between the Ave_nbh and Cali_trad_nbh approaches. Diurnal trends in the BSs of the techniques are also noted.

The results are presented for precipitation thresholds of 0.01 and 0.25 in. per 6-h period. The 0.01-in. threshold delineates the area with any precipitation; as such, it may provide guidance on various forecasting issues, whereas the 0.25-in. threshold delineates moderate precipitation usually forecasted with an acceptable level of accuracy.

c. Verification indices used in the study

For each variation of the Ave_nbh approach above, Brier scores (Brier 1950) are computed using

\[
BS = \frac{1}{n} \sum_{k=1}^{n} (p_k - o_k)^2, \tag{1}
\]

where \(p_k\) is the PoP forecast for the \(k\) forecast of the total forecasts \(n\) and \(o_k\) is the observed PoP to occur from the observations (either 0 or 1) for each forecast scenario. BSs can also be decomposed, as outlined in Murphy (1973), Wilks (2006), and Schaffer et al. (2011), into three components, reliability, resolution, and uncertainty, using

\[
BS = \frac{1}{n} \sum_{i=1}^{I} N_i(p_i - \bar{o}_i)^2 - \frac{1}{n} \sum_{i=1}^{I} N_i(\bar{p}_i - \bar{o})^2 + \bar{o}(1 - \bar{o}), \tag{2}
\]

where

\[
\bar{o}_i = \frac{1}{N_i} \sum_{k \in N_i} o_k \quad \text{and} \quad \tag{3}
\]

\[
\bar{o} = \frac{1}{n} \sum_{i=1}^{I} N_i \bar{o}_i, \tag{4}
\]

where \(N_i\) is the number of forecasts in the \(i\)th forecast category, \(n\) is the total number of forecasts, and \(I\) is the number of bins. Smaller BSs are considered to be more accurate. Since the main objective of the present BS computation is the comparison of Ave_nbh with Cali_trad_nbh for the same combinations of the TR and TE datasets and periods of the day, as well as for the same simulations and domains, BS is a reasonable choice as a metric for this purpose. Because errors are distributed normally, differences are tested for statistical significance at the 95% confidence levels using paired Student’s \(t\) tests.

3. Results

a. Impacts of different TR and TE datasets

The 07TR, 08TR, and 10TR datasets on a 4-km horizontal grid are used, and TE is compared against another year of data to investigate the sensitivity of BSs to the different TR and TE datasets. In addition, for one test, the 07TR and 08TR datasets are combined to determine how BSs would behave if the sample size of the TR forecast was increased greatly. The BS results shown
in Tables 2 and 3 are only for the 6vs6 TR/TE method because scores are similar compared with the 30vs6 TR/TE method, and \( p \) values between 0.4 and 0.5 show that there is not a statistically significant difference between the methods at the 95% confidence interval.

Using the thinning technique to reduce computational expenses, the 0708TR dataset performed better on 10TE as compared to the older datasets 07TE and 08TE for both the Ave_nbh and Cali_trad_nbh approaches (Tables 2 and 3). The 0708TR dataset has a slightly lower average BS for the Ave_nbh approach for the 10TE dataset than when compared with 08TR10TE, but the 0708TR and 08TR datasets perform very similarly. The Cali_trad_nbh approach has similar BSs when the 0708TR dataset is used compared with the 08TR dataset alone on 10TE. BSs do not show a statistically significant (\( p > 0.05 \)) improvement for either Ave_nbh or Cali_trad_nbh when the 0708TR dataset is compared with using the 08TR dataset alone on 10TE. BSs do not show a statistically significant difference between the methods at the 95% confidence interval.

Using the thinning technique to reduce computational expenses, the 0708TR dataset performed better on 10TE as compared to the older datasets 07TE and 08TE for both the Ave_nbh and Cali_trad_nbh approaches (Tables 2 and 3). The 0708TR dataset has a slightly lower average BS for the Ave_nbh approach for the 10TE dataset than when compared with 08TR10TE, but the 0708TR and 08TR datasets perform very similarly. The Cali_trad_nbh approach has similar BSs when the 0708TR dataset is used compared with the 08TR dataset alone on 10TE. BSs do not show a statistically significant (\( p > 0.05 \)) improvement for either Ave_nbh or Cali_trad_nbh when the 0708TR dataset is compared with using the 08TR dataset alone on 10TE. BSs do not show a statistically significant difference between the methods at the 95% confidence interval.

Using the thinning technique to reduce computational expenses, the 0708TR dataset performed better on 10TE as compared to the older datasets 07TE and 08TE for both the Ave_nbh and Cali_trad_nbh approaches (Tables 2 and 3). The 0708TR dataset has a slightly lower average BS for the Ave_nbh approach for the 10TE dataset than when compared with 08TR10TE, but the 0708TR and 08TR datasets perform very similarly. The Cali_trad_nbh approach has similar BSs when the 0708TR dataset is used compared with the 08TR dataset alone on 10TE. BSs do not show a statistically significant (\( p > 0.05 \)) improvement for either Ave_nbh or Cali_trad_nbh when the 0708TR dataset is compared with using the 08TR dataset alone on 10TE. BSs do not show a statistically significant difference between the methods at the 95% confidence interval.

Using the thinning technique to reduce computational expenses, the 0708TR dataset performed better on 10TE as compared to the older datasets 07TE and 08TE for both the Ave_nbh and Cali_trad_nbh approaches (Tables 2 and 3). The 0708TR dataset has a slightly lower average BS for the Ave_nbh approach for the 10TE dataset than when compared with 08TR10TE, but the 0708TR and 08TR datasets perform very similarly. The Cali_trad_nbh approach has similar BSs when the 0708TR dataset is used compared with the 08TR dataset alone on 10TE. BSs do not show a statistically significant (\( p > 0.05 \)) improvement for either Ave_nbh or Cali_trad_nbh when the 0708TR dataset is compared with using the 08TR dataset alone on 10TE. BSs do not show a statistically significant difference between the methods at the 95% confidence interval.

Using the thinning technique to reduce computational expenses, the 0708TR dataset performed better on 10TE as compared to the older datasets 07TE and 08TE for both the Ave_nbh and Cali_trad_nbh approaches (Tables 2 and 3). The 0708TR dataset has a slightly lower average BS for the Ave_nbh approach for the 10TE dataset than when compared with 08TR10TE, but the 0708TR and 08TR datasets perform very similarly. The Cali_trad_nbh approach has similar BSs when the 0708TR dataset is used compared with the 08TR dataset alone on 10TE. BSs do not show a statistically significant (\( p > 0.05 \)) improvement for either Ave_nbh or Cali_trad_nbh when the 0708TR dataset is compared with using the 08TR dataset alone on 10TE. BSs do not show a statistically significant difference between the methods at the 95% confidence interval.

Using the thinning technique to reduce computational expenses, the 0708TR dataset performed better on 10TE as compared to the older datasets 07TE and 08TE for both the Ave_nbh and Cali_trad_nbh approaches (Tables 2 and 3). The 0708TR dataset has a slightly lower average BS for the Ave_nbh approach for the 10TE dataset than when compared with 08TR10TE, but the 0708TR and 08TR datasets perform very similarly. The Cali_trad_nbh approach has similar BSs when the 0708TR dataset is used compared with the 08TR dataset alone on 10TE. BSs do not show a statistically significant (\( p > 0.05 \)) improvement for either Ave_nbh or Cali_trad_nbh when the 0708TR dataset is compared with using the 08TR dataset alone on 10TE. BSs do not show a statistically significant difference between the methods at the 95% confidence interval.

Using the thinning technique to reduce computational expenses, the 0708TR dataset performed better on 10TE as compared to the older datasets 07TE and 08TE for both the Ave_nbh and Cali_trad_nbh approaches (Tables 2 and 3). The 0708TR dataset has a slightly lower average BS for the Ave_nbh approach for the 10TE dataset than when compared with 08TR10TE, but the 0708TR and 08TR datasets perform very similarly. The Cali_trad_nbh approach has similar BSs when the 0708TR dataset is used compared with the 08TR dataset alone on 10TE. BSs do not show a statistically significant (\( p > 0.05 \)) improvement for either Ave_nbh or Cali_trad_nbh when the 0708TR dataset is compared with using the 08TR dataset alone on 10TE. BSs do not show a statistically significant difference between the methods at the 95% confidence interval.

Using the thinning technique to reduce computational expenses, the 0708TR dataset performed better on 10TE as compared to the older datasets 07TE and 08TE for both the Ave_nbh and Cali_trad_nbh approaches (Tables 2 and 3). The 0708TR dataset has a slightly lower average BS for the Ave_nbh approach for the 10TE dataset than when compared with 08TR10TE, but the 0708TR and 08TR datasets perform very similarly. The Cali_trad_nbh approach has similar BSs when the 0708TR dataset is used compared with the 08TR dataset alone on 10TE. BSs do not show a statistically significant (\( p > 0.05 \)) improvement for either Ave_nbh or Cali_trad_nbh when the 0708TR dataset is compared with using the 08TR dataset alone on 10TE. BSs do not show a statistically significant difference between the methods at the 95% confidence interval.
and the lowest accuracy in the 18–24-h period for the 0.01-in. threshold, which likely reflects lower accuracy when convective precipitation is more active in the afternoon and higher accuracy when convection is less active in the morning hours (not shown).

BSs at the 0.25-in. threshold are lower (better) than those computed at the 0.01-in. threshold. This trend likely reflects the fact that more of the domain is dry when the threshold is raised, and correct “no events” are easier to forecast. This can be noted across the forecast domain as uncertainty values decrease, along with BSs, as the threshold increases (not shown). For example, the uncertainty term for the first 6-h period for the 10TE dataset is 0.1345 at the 0.01-in. threshold whereas it is 0.0399 at the 0.25-in. threshold (similar results can be seen in Table 3 in SGS11). The optimal neighborhood size for Ave_nbh resulting in the best BS using the TR dataset for 4-km grid spacing is shown in Table 4, while the optimal size for Cali_trad_nbh resulting in the best BS is shown in Table 5. Optimal sizes are much larger for Ave_nbh than for Cali_trad_nbh. Once the neighborhood size increases beyond the optimal neighborhood size for which the best BS is yielded, BSs continue to worsen with larger neighborhood sizes in both the Ave_nbh and Cali_trad_nbh approaches.

Roberts and Lean (2008) found that fractions skill scores improved as the neighborhood size increased until an asymptote was reached. However, the present study finds that at scales larger than the optimal neighborhood size, BSs worsen. This difference in results from Roberts and Lean (2008) is likely due to differences in the methodology used in each study. For example, Roberts and Lean (2008) used a verification strategy that incorporated a neighborhood approach on radar data. In contrast, the current study does not use a neighborhood on the Stage IV data used for verification. The Stage IV dataset uses a combination of radar and observed gauge readings. The current study only applies a neighborhood technique to the forecast dataset. Therefore, as the neighborhood size increases, some details of the forecasted convective systems are lost as QPFs are smoothed from areas where either 1) no precipitation is occurring or 2) the precipitation characteristics are different from those of convection (e.g., light stratiform rain versus intense convective rain). The increased smoothing with scale in the forecasts and the lack of smoothing in the observations likely explains why an optimal neighborhood size exists.

When an older dataset (e.g., 08TR) is used for TR and a newer dataset (e.g., 10TE) for TE, BSs (not shown) are found to improve over BSs from TR on an older dataset and TE on the same dataset (i.e., 08TR08TE). This finding indicates that as models improve with time, the importance of updating the TR dataset may be reduced in the operational forecasting environment. Instead, the same static PoP tables may be used with even better results as high-resolution forecast models improve over time. The impact of TR on newer data (e.g., 10TR) and TE on an older dataset (e.g., 08TE) is also studied. Resulting BSs show that a newer

### Table 3. As in Table 2, but for the 0.25-in. threshold.

<table>
<thead>
<tr>
<th>Period (h)</th>
<th>Ave_nbh</th>
<th>Cali_trad_nbh</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–6</td>
<td>0.0385</td>
<td>0.0376</td>
</tr>
<tr>
<td>6–12</td>
<td>0.0395</td>
<td>0.0396</td>
</tr>
<tr>
<td>12–18</td>
<td>0.0334</td>
<td>0.0347</td>
</tr>
<tr>
<td>18–24</td>
<td>0.0475</td>
<td>0.0482</td>
</tr>
<tr>
<td>24–30</td>
<td>0.0348</td>
<td>0.0364</td>
</tr>
<tr>
<td>Avg</td>
<td>0.0356</td>
<td>0.0344</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Period (h)</th>
<th>Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–6</td>
<td>0.0058</td>
</tr>
<tr>
<td>6–12</td>
<td>0.0050</td>
</tr>
<tr>
<td>12–18</td>
<td>0.0017</td>
</tr>
<tr>
<td>18–24</td>
<td>0.0012</td>
</tr>
<tr>
<td>24–30</td>
<td>0.0004</td>
</tr>
<tr>
<td>Avg</td>
<td>0.0028</td>
</tr>
</tbody>
</table>
and for the same period at the 0.25-in. threshold in Fig. 4. The simulations behaved similarly in terms of reliability (0000–0600 UTC) for the 0.01-in. threshold (Fig. 3) to Cali_trad_nbh. To better understand how the re-

method, but the Ave_nbh approach was not compared to Cali_trad approach (not shown), but not as well as Cali_trad_nbh. 

SGS11 compared reliability curves for different ap-

approaches and determined that the GS approach showed better reliability than the point-based Cali_trad approach (not shown), but not as well as Cali_trad_nbh. 

TABLE 5. The 4-km grid-spacing optimal Cali_trad_nbh size for which the best BS is computed for the TR dataset years for the 6vs6 test for thresholds of 0.01 in. (Table 2) and 0.25 in. (Table 3). 

<table>
<thead>
<tr>
<th>Period (h)</th>
<th>TR year</th>
<th>TE year</th>
</tr>
</thead>
<tbody>
<tr>
<td>6–12</td>
<td>0.01 in.</td>
<td>0.01 in.</td>
</tr>
<tr>
<td>12–18</td>
<td>0.01 in.</td>
<td>0.01 in.</td>
</tr>
<tr>
<td>18–24</td>
<td>0.01 in.</td>
<td>0.01 in.</td>
</tr>
<tr>
<td>24–30</td>
<td>0.01 in.</td>
<td>0.01 in.</td>
</tr>
</tbody>
</table>

An additional insight into the potential computer resources saved by using the Ave_nbh approach versus Cali_trad_nbh is provided by evaluating how many members of Cali_trad_nbh are needed (i.e., what is the minimum subensemble size \( n_{se} \) needed) to provide a BS that is better than that obtained by Ave_nbh. The evaluation is done for each of the seven combinations of TR and TE years and five forecast periods given in the previous BS tables. The subensembles are generated by using alternating top–bottom selection of members listed in Table 4 in SGS11; for example, for \( n_{se} = 4 \), the members of the subensembles are 1, 10, 2, and 9. This

TE dataset is more important than a newer TR dataset. Although the BSs using the thinning approach were slightly worse compared with using a nonthinning approach with Ave_nbh (not shown), the differences between the BSs were very small and not statistically significant \( (p > 0.05) \), whereas the increased computational efficiency was substantial. The Ave_nbh approach performed better than the point-based Cali_trad approach (not shown), but not as well as Cali_trad_nbh. 

TABLE 5. The 4-km grid-spacing optimal Cali_trad_nbh size for which the best BS is computed for the TR dataset years for the 6vs6 test for thresholds of 0.01 in. (Table 2) and 0.25 in. (Table 3).

<table>
<thead>
<tr>
<th>Period (h)</th>
<th>TR year</th>
<th>TE year</th>
</tr>
</thead>
<tbody>
<tr>
<td>6–12</td>
<td>0.01 in.</td>
<td>0.01 in.</td>
</tr>
<tr>
<td>12–18</td>
<td>0.01 in.</td>
<td>0.01 in.</td>
</tr>
<tr>
<td>18–24</td>
<td>0.01 in.</td>
<td>0.01 in.</td>
</tr>
<tr>
<td>24–30</td>
<td>0.01 in.</td>
<td>0.01 in.</td>
</tr>
</tbody>
</table>

An additional insight into the potential computer resources saved by using the Ave_nbh approach versus Cali_trad_nbh is provided by evaluating how many members of Cali_trad_nbh are needed (i.e., what is the minimum subensemble size \( n_{se} \) needed) to provide a BS that is better than that obtained by Ave_nbh. The evaluation is done for each of the seven combinations of TR and TE years and five forecast periods given in the previous BS tables. The subensembles are generated by using alternating top–bottom selection of members listed in Table 4 in SGS11; for example, for \( n_{se} = 4 \), the members of the subensembles are 1, 10, 2, and 9. This
analysis for the 4-km simulations and the 0.01- and 0.25-in. thresholds (Table 6) shows that the number of ensemble members required to result in a more accurate PoP forecast than the Ave_nbh approach increases with lead time, which may indicate that greater filtering of unpredictable scales is needed for longer lead times (Surcel et al. 2014). The number of ensemble members needed to result in more accurate PoP forecasts may also be influenced by the diurnal cycle noted earlier in the resulting BSs, where a greater number of ensemble members is needed during diurnal periods with more convection. The Cali_trad_nbh 10-member ensemble requires more ensemble members to obtain a better BS compared with that obtained by the single-model Ave_nbh approach when using the newer 2010 dataset for TE. The Cali_trad_nbh method most often performs better than the Ave_nbh approach, albeit slightly, when the older datasets 07TE and 08TE are used; however, Cali_trad_nbh consistently requires multiple members to produce a better BS.

Table 7 shows the number of times out of all of the different TR and TE combinations for which Ave_nbh performs better than various subensembles of Cali_trad_nbh. Ultimately, the Cali_trad_nbh method outperforms Ave_nbh for both thresholds, with Cali_trad_nbh having the lowest BS 31 times compared to 4 for Ave_nbh. However, when the point-based Cali_trad approach was used (not shown), the Cali_trad method had the lowest BS at the 0.01-in. (0.25 in.) threshold 18 times (14 times) compared with 17 times (21 times) for Ave_nbh. Although the two methods are similar in terms of the BSs, it is important to note that the Ave_nbh method only uses one model for computations whereas the point-based Cali_trad and Cali_trad_nbh methods require substantially more ensemble members to compete with Ave_nbh. For both the 0.01- and 0.25-in. thresholds, Cali_trad_nbh often requires more than one member.
to obtain a better BS than the Ave_nbh approach (Table 7). Therefore, Ave_nbh has a computational advantage over Cali_trad_nbh while yielding similar BSs that do not differ in a statistically significant way ($p > 0.05$) when the same combinations of TR and TE datasets are used.

b. Impact of increased TR sample size using multiple years

To examine the sensitivity of BS to the sample size more precisely while eliminating possible BS changes due to different yearly configurations of the CAPS ensemble for 4-km grid spacing, BSs are computed using randomly chosen subsets (roughly 25%, 50%, 75%, and 100%) of the 49 cases in the 0708TR dataset for the Ave_nbh approach using thinning. The 10TE dataset is used for the 0–6-h forecast period for the 0.01-in. threshold (Fig. 5). TR and TE are completed over the 0–6-h forecast period for 6vs6. Using 49 days (100% of the sample) in the 0–6-h forecast period results in the lowest BS at the $45 \times 45$ Ave_nbh size, with small monotonic worsening in BS occurring for most neighborhoods as the sample size is decreased. However, the $45 \times 45$ size corresponds to the $9 \times 9$ neighborhood size at the 20-km horizontal grid spacing that was tested in SGS11. At 20 km, the $13 \times 13$ neighborhood size was found to be optimal (SGS11). Thus, the Ave_nbh at 4 km in the present study yields a smaller optimal spatial neighborhood area. The minimal BS impact of sample size in the static TR of the dataset in the current study supports the finding of JW12, where little to no forecast improvement was found when using a dynamic TR approach based either on 10 or 25 days for TR. Although there is some small monotonic BS improvement with an increase in sample size, the results are not considered to be statistically significant ($p$ values $> 0.05$).

c. Daily variations of BS

Further insight into the differences between the Ave_nbh and Cali_trad_nbh approaches should include daily comparisons of the difference in BS, as would be done in an operational forecasting environment, particularly for

![Figure 4](InlineFigure4.png)

*Fig. 4.* As in Fig. 3, but for the 0.25-in. threshold.
events that have a significant impact for users of the forecasts. A cursory examination of spatial variations in PoP for available cases subjectively indicates that differences are quite small, although the Ave_nbh results imply more sharpness (not shown).

Daily changes in the BSs for each 6-h period for each date for the Ave_nbh and Cali_trad_nbh approaches at the 0.01-in. (Fig. 6) and 0.25-in. (Fig. 7) thresholds for the 4-km output for three of the seven possible TR/TE dataset combinations show that the two approaches behave similarly. A perfunctory look at synoptic and mesoscale patterns in comparison to the CAPS model forecasts is done to determine how well the models handle different types of weather patterns (not shown).

On days with low predictability (or where model forecasts have difficulty with the weather pattern of the forecast period), neither approach seems to have an advantage over the other. One exception occurs for 08TE. On a few days when the BS is relatively poor with the Cali_trad_nbh approach, the Ave_nbh technique earns noticeably worse BSs (see Fig. 6a; case 0526). It appears that the WRF Model improvements realized by using 2010 data may have eliminated such behavior in Ave_nbh.

The BSs from the Cali_trad_nbh approach are subtracted from the BSs of the Ave_nbh approach to show the differences between the two methods (Fig. 8). Therefore, negative differences indicate that Ave_nbh is the better approach whereas positive differences indicate that Cali_trad_nbh is the better approach. The differences between Ave_nbh and Cali_trad_nbh for these three TR/TE dataset combinations are usually small, supporting subjective impressions comparing PoP maps generated by the two approaches. Differences in BSs between the Ave_nbh and Cali_trad_nbh approaches are also examined for all seven combinations of TR and TE, but the differences between the BSs appear to be sensitive only to the TE dataset used. Therefore, the three examples shown in Fig. 8 similarly convey the trends among all seven combinations. The results imply that during more highly predictable precipitation regimes, both techniques perform well. The Cali_trad_nbh and Ave_nbh approaches also appear to yield similar BSs during less predictable regimes. Although the techniques generally produce similar PoP forecasts, a few outlier periods appear for the 08TE dataset using Ave_nbh. In this limited set of events, where Cali_trad_nbh noticeably outperforms Ave_nbh, equitable threat scores (not shown) for the 6-h periods for the simulation used in Ave_nbh are often relatively low. It would seem possible that in the event of an especially challenging forecast, it would be more likely that Ave_nbh could produce poorer PoP forecasts than Cali_trad_nbh, since Ave_nbh, as used here, makes use of only a single deterministic run. The Cali_trad_nbh approach, because it makes use of an ensemble of runs, may be less impacted by the likelihood that several forecasts would be poor. In these rare events, it is possible that the use of neighboring points in Ave_nbh is insufficient to alleviate the problems in a very poor model forecast. As forecasting models continue to improve, it is likely that these outlier points would become even rarer, as suggested by the results for 10TE.

4. Discussion

a. Model results

In the following discussion, we further evaluate the model results and the potential benefits from using the specific Ave_nbh approach with thinning and we also provide some assessment of the results obtained based on comparisons with other relevant studies. The two tests conducted in this study add insight into the use of the GS approach employed in GS04, Gallus et al. (2007), and SGS11.
The first test investigates the impacts of using different TR and TE datasets making use of 2007, 2008, and 2010 CAPS WRF model ensemble output with 4-km grid spacing. Although the 2010 CAPS dataset included improvements over the 2007 and 2008 datasets, the 10TR dataset does not lead to BS improvements on 07TE or 08TE. An increase in the TR dataset sample size leads to only slightly better BSs, which implies limited sensitivity to sample size. BSs are more sensitive to the TE dataset than the TR dataset used.

Statistical significance is not found between 6vs6 and 30vs6 for all thresholds across the 6-h time periods ($p > 0.05$). Using a 6-h time period for TR and TE over the same 6-h time period is of computational advantage because of the reduction in the time required for PoP creation compared to using all five 6-h periods for TR. The Cali_trad_nbh approach generally has better BSs than the Ave_nbh method for both thresholds. However, for a given TE and TR dataset, BSs are similar and differences in BSs between the Ave_nbh and Cali_trad_nbh approaches are not statistically significant ($p > 0.05$).

A second test reveals that the behavior of the different TE approaches generally is similar on a day-by-day basis. Both the Cali_trad_nbh and the Ave_nbh approaches appear to improve the BSs compared to that of the Uncali_trad approach (not shown) during days when forecasts are difficult (BS is poor). Lesser impacts are

**FIG. 5.** BSs for different percentages (see legend at top right) of the 49 cases used for the 0708TR10TE thinned dataset for the 0.01-in. threshold during the 0000–0600 UTC period (6vs6) with 4-km output.
present on days when Uncali Tradable is already providing relatively good BSs. Similar to previous findings, statistical significance was found when comparing the BSs of different TR/TE years \((p < 0.05)\), and statistical significance was not found when comparing different TR datasets on the same TE year \((p > 0.05)\).

Different variations of the Ave_nbh approach could improve the PoP forecast accuracy in certain situations. The optimal neighborhood size varies depending on the threshold and time period. Generally, the optimal neighborhood areal size occurs over regions of about 180 km × 180 km. Duda and Gallus (2013) found that the average errors in initiation are about 105 km, whereas the optimal neighborhood sizes are larger in this study. Because the optimal neighborhood size in the present study is larger than the average error in initiation location, this result suggests that location errors may become larger as systems mature and grow upscale (Surcel et al. 2014).

As the horizontal grid resolution is enhanced, the resulting increase in the number of neighborhood grid points is likely to generate a computational burden in needed resources for TR and TE in the Ave_nbh approach. This study finds that a thinning technique requires considerably fewer computational resources while improving the reliability term (but worsening the resolution term) of the decomposed BS of Ave_nbh (not shown), as well as providing some generally slight improvement in the BS compared to the point-based Cali Tradable approach when using the newer 2010 dataset.
However, the Cali_trad_nbh approach had better BSs compared to Ave_nbh in most instances. It appears that the thinning is optimized when the remaining domain grid points provide a reasonable representation of the precipitation within their subneighborhoods (Snell et al. 2000). Although the Ave_nbh approach is not a substitute for a neighborhood approach using a full ensemble (like Cali_trad_nbh), the PoP forecasts yielded from such an approach can provide comparable forecast accuracy with the use of far fewer computer resources.

b. Conceptual evaluation of the computer resource advantage of the neighborhood approach

To evaluate the potential benefit of using the neighborhood postprocessing of a single model prediction to obtain PoP guidance compared to running a traditional ensemble, the following conceptual bulk approach is useful. Without loss of generality, the 4-km grid spacing Ave_nbh and Cali_trad_nbh approaches are adopted. Assuming that the full capacity of a computer is 1 CPU (considered to be a normalized CPU unit of time, denoted hereafter CPUN, that is entirely consumed in operational use) for the simulation of one member of a traditional ensemble, for 10 members in a traditional ensemble, and given $D_x$, $D_y$ in the $x$ and $y$ directions, respectively, and $\Delta t$ (time step) values, the total CPUN needed would be 10. Conversely, for the Ave_nbh approach based on a single ensemble member, the CPUN needed would only be 1. Hence, nine CPUN are left for future improvements in the single-model run supplying information needed by the Ave_nbh approach (e.g., reducing $\Delta x$, and $\Delta y$ by a factor of 2 and as a result the
CFL criteria \( \Delta t \) as well). The potential CPUN saved allows for refinements, which are likely to produce some forecast improvement. This differs from the traditional calibrated 10-member ensemble, which needs 10 CPU while using unrefined \( \Delta x \) and \( \Delta y \). Further, the small higher-resolution ensemble could be postprocessed in a way that could improve a single output consensus QPF. The Ave_nbh approach could use this new consensus QPF to perhaps produce better PoP forecasts.

Another advantage of using more dynamic ensemble members would be a potential smaller Ave_nbh size that would be needed to compute more accurate PoP forecasts given the consensus QPF already filtering out unpredictable precipitation behavior (Seed 2003). Using a smaller neighborhood size would increase the computational speed. Another potential advantage of using a dynamic ensemble in combination with the Ave_nbh approach would be in retaining the details of the QPFs where ensemble members agree, including forecast areas influenced by topographic effects. Therefore, the Ave_nbh approach with thinning for PoP creation may increase its own value in operational forecasting.

5. Conclusions

The paper by Theis et al. (2005) triggered increased research activity in using neighborhood approaches for warm season precipitation prediction. However, most of the research has been oriented toward using the neighborhood approach to boost the traditional ensemble approach, which is essentially free of real-world constraints on computer resources.

The present paper provides an additional evaluation of the Ave_nbh approach used in SGS11 with enhanced
data and the thinning refinement application. The Ave_nbh approach based on a single-model run and using a thinning technique has been found to perform relatively similarly at 6-h intervals for each case in the CAPS datasets to a calibrated traditional approach using a neighborhood (Cali_trad_nbh) based on the CAPS EPS composed of 10 members. It is important to note that this conclusion is obtained while the computer resources needed for Cali_trad_nbh are roughly 10 times those needed for the Ave_nbh approach. Evaluations suggest that for equal computational resource allocation, competitive BSs are associated with the Ave_nbh approach compared with the Cali_trad_nbh method evaluated in the present study. With the computational resources savings and the speed of PoP forecast generation, the Ave_nbh approach can be used efficiently to provide accurate PoP forecasts in an operational forecast environment.

Acknowledgments. This research was funded in part by National Science Foundation Grants ATM-0848200 and ATM-1222383 with funds from the American Recovery and Reinvestment Act of 2009. The authors thank Moti Segal for substantial input into this research, Michael Greve for help creating map analyses, and Daryl Herzmann for computational assistance. The reviews from three anonymous reviewers were appreciated.

REFERENCES


