

1  
2  
3 **Analog Probabilistic Precipitation Forecasts Using GEFS Reforecasts**  
4 **and Climatology-Calibrated Precipitation Analyses**  
5

6  
7 Thomas M. Hamill,<sup>1</sup> Michael Scheuerer,<sup>2</sup> and Gary T. Bates<sup>2</sup>  
8

9 <sup>1</sup> *NOAA Earth System Research Lab, Physical Sciences Division, Boulder, Colorado*

10  
11 <sup>2</sup> *CIRES, University of Colorado, Boulder, Colorado*  
12

13  
14  
15 Submitted to *Monthly Weather Review*

16  
17 as an expedited contribution  
18

19  
20  
21 revised  
22

23 20 March 2015  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35

36 Corresponding author:

37 Dr. Thomas M. Hamill

38 NOAA Earth System Research Lab

39 Physical Sciences Division

40 R/PSD 1, 325 Broadway

41 Boulder, CO 80305

42 [Tom.Hamill@noaa.gov](mailto:Tom.Hamill@noaa.gov)

43 Phone: (303) 497-3060

44 Telefax: (303) 497-6449  
45

46

47 ABSTRACT

48

49           Analog post-processing methods have previously been applied using  
50 precipitation reforecasts and analyses to improve probabilistic forecast skill and  
51 reliability. A modification to a previously documented analog procedure is  
52 described here that produces highly skillful, statistically reliable precipitation  
53 forecast guidance at a 1/8<sup>th</sup>-degree grid spacing. These experimental probabilistic  
54 forecast products are available via the web in near real-time.

55           The main changes to the previously documented analog algorithm were as  
56 follows: (a) use of a shorter duration (2002-2013) but smaller grid spacing, higher-  
57 quality time series of precipitation analyses for training and forecast verification,  
58 the Climatology Calibrated Precipitation Analysis; (b) increased training sample size  
59 using data from 19 supplemental locations, chosen for their similar precipitation  
60 analysis climatologies and terrain characteristics; (c) selection of analog dates for a  
61 particular grid point based on the similarity of forecast characteristics at that grid  
62 point rather than similarity in a neighborhood around that grid point; (d) using an  
63 analog rather than a rank-analog approach; (e) varying the number of analogs used  
64 to estimate probabilities from a smaller number (50) for shorter-lead forecasts to a  
65 larger number (200) for longer-lead events; (f) spatial Savitzky-Golay smoothing of  
66 the probability fields. Special procedures were also applied near coasts and country  
67 boundaries to deal with data unavailability outside of the US while smoothing.

68           The resulting forecasts are much more skillful and reliable than raw  
69 ensemble guidance across a range of event thresholds. The forecasts are not nearly

70 as sharp, however. The use of the supplemental locations is shown to especially  
71 improve the skill of short-term forecasts during the winter.

72 **1. Introduction.**

73 Previous studies have shown that probabilistic forecasts of precipitation can  
74 be significantly improved by post-processing with reforecasts (e.g., Hamill et al.  
75 2006, hereafter H06; Hamill et al. 2013, hereafter H13; Hamill and Whitaker 2006,  
76 hereafter HW06). The real-time forecast is adjusted using a long time series of past  
77 forecasts and associated precipitation analyses. Appealing for its simplicity was the  
78 “analog” procedure used in these studies. For a given location, dates in the past  
79 were identified that had reforecasts similar to today’s forecast. An ensemble was  
80 formed from the observed or analyzed precipitation amounts on the dates of the  
81 chosen analogs, and probabilities were estimated from the ensemble relative  
82 frequency. Maps of precipitation probabilities were constructed by repeating the  
83 procedure across the model grid points.

84 A challenge with analog procedures used in these previous studies was their  
85 inability to find many close-matching forecasts when today’s precipitation forecast  
86 amount was especially large, even with a long training data set. The method as  
87 previously documented used the data surrounding grid point of interest but did not  
88 supplement the training data set with observation and forecast data centered on  
89 other locations. The benefit of this location-specific approach was that if the  
90 model’s systematic errors varied greatly with location, it corrected for these, as  
91 shown in H06. One disadvantage was that if there were not many prior forecasts  
92 with similarly extreme precipitation, then the selected analogs were biased toward  
93 precipitation forecasts with less extreme forecast values and typically lighter

94 analyzed precipitation. Consequently, the forecast procedure did not often produce  
95 high probabilities of extreme events.

96 Another possible disadvantage of the forecast products demonstrated in  
97 these previous studies was that the associated precipitation analyses were in each  
98 case from the North American Regional Reanalysis (NARR, Mesinger et al. 2006).  
99 Several studies have identified deficiencies with this data set (e.g., West et al. 2007,  
100 Bukovsky and Karoly 2009). We have also noted a significant dry bias in the NARR  
101 over the northern Great Plains during the winter season. There is now an  
102 alternative data set that covers the contiguous US (CONUS) and that utilizes both  
103 gauge and adjusted radar-reflectivity data, the Climatology-Calibrated Precipitation  
104 Analysis (CCPA; Hou et al. 2014). Data is available from 2002-current. While this  
105 time period is shorter than the 1985-current time span of the most recent reforecast  
106 (H13), the availability of higher-resolution, more accurate precipitation analysis  
107 data has led us to consider whether useful products could be generated with this  
108 data set.

109 This article briefly describes modifications to previously documented analog  
110 forecast procedures. What adjustments will allow it to provide improved  
111 probabilistic forecasts while using a shorter time series of analyses? We describe a  
112 series of changes to the analog algorithm and show that the resulting analog  
113 probabilistic forecasts are skillful, somewhat more sharp, and reliable. Since the  
114 statistically post-processed guidance provide a significant improvement over  
115 probabilities from the raw Global Ensemble Forecast System (GEFS) forecast data,  
116 we are also making experimental web-based guidance available in near real time

117 during the next few years; this guidance can be obtained from  
118 <http://www.esrl.noaa.gov/psd/forecasts/reforecast2/ccpa/index.html>.

119

## 120 2. **Methods and data.**

### 121 a. *Reforecast data, observational data, and verification methods.*

122 In this study we considered 12-hourly accumulated precipitation forecasts  
123 during the 2002 to 2013 period for lead times up to +8 days. Precipitation analyses  
124 were obtained on a  $\sim 1/8$ -degree grid from the CCPA data set of Hou et al. (2014).  
125 Probabilistic forecasts were produced at this  $\sim 1/8$ -degree resolution over the  
126 CONUS. All of the forecast data used in this project were obtained from the second-  
127 generation GEFS reforecast data set, described in H13. Ensemble-mean  
128 precipitation and total-column ensemble-mean precipitable water were used in the  
129 analog procedure. GEFS data was extracted (for precipitation) on the GEFS's native  
130 Gaussian grid at  $\sim 1/2$ -degree resolution in an area surrounding the CONUS.  
131 Precipitable-water forecasts, which were archived on a 1-degree grid, were  
132 interpolated to the native Gaussian grid before input to the analog procedure.

133 Forecasts were cross validated; for example, 2002 forecasts were trained  
134 using 2003-2013 data. For the production of forecasts in a given month, the training  
135 data used that month and the surrounding two months, e.g., January forecasts were  
136 trained with December-January-February data. Was the use of future data in the  
137 cross-validation procedure a source of unrealistic skill of these forecasts? As shown  
138 in Baxter et al. (2014, Fig. 5 therein), the inter-annual variability of skill in the  
139 southeast US was larger than the systematic changes from 2002 to 2013. This

140 suggests that the use of future forecasts in the cross-validation procedure probably  
141 did not result in a large over-estimation of forecast skill for the earlier years.

142         One of the controls against which the new method was compared were the  
143 raw event probabilities generated from the 11-member GEFS reforecast ensemble,  
144 bi-linearly interpolated to the 1/8-degree grid.

145         Verification methods included reliability diagrams and Brier Skill Scores  
146 computed in the conventional way (Wilks 2006, eqs. 7.34 and 7.35, Hamill and Juras  
147 2006), with climatology providing the reference probabilistic forecasts. Maps of  
148 Brier Skill Scores were also generated for the CONUS. These were produced by  
149 accumulating the probabilistic forecasts' and climatological forecasts' average of  
150 squared error at that grid point across all years and all months prior to the  
151 calculation of skill. Because of the extremely large sample size, confidence intervals  
152 for the skill differences (very small; see HW06) were not included on the plots.

153

154 *b. Rank analog forecast procedure as a control.*

155         A revised "rank-analog" approach served as another standard of comparison  
156 for the newer, somewhat more involved analog methodology described in section  
157 2.c below. For the most part, the rank-analog approach was a hybrid of the  
158 techniques that have previously been shown to work well, described in sections  
159 3.b.6 and 3.b.8 of HW06. This control rank-analog methodology was further  
160 updated in the following respects:

161         • As with the rank-analog algorithm of HW06, the rank of the forecast for a  
162 particular date of interest and set of grid points was compared against the ranks of

163 sorted forecasts at the same set of grid points for each date in the training data set.  
 164 In evaluating which forecasts were closest to today's forecast, the difference  
 165 between forecasts was calculated as 70% of the absolute difference of the  
 166 precipitation forecast ranks and 30% of the absolute difference in precipitable  
 167 water forecast ranks averaged over the set of grid points, following HW06. As  
 168 shown therein, with the exception of warm-season probability of precipitation,  
 169 there was minimal sensitivity to the chosen weight between precipitation and  
 170 precipitable water. A more precise definition of the forecast difference is as follows:  
 171 let  $S$  be the set of grid points in a region surrounding the current grid point of  
 172 interest. Let  $tc$  be the current date, and let  $t$  be another date from the set of dates  $T$   
 173 whose forecast data will be compared against the forecast at  $tc$ . As indicated  
 174 previously, by cross validation  $tc \notin T$ . Define  $rpr_s^{tc}$  as the rank of the current  
 175 forecast precipitation amount at time  $tc$  and at grid point  $s$  from a combined set with  
 176 the training data at  $s$ . Similarly,  $rpw_s^{tc}$  is the associated rank of the current total-  
 177 column precipitable water forecast. Then the difference in ranks for date  $t$  was  
 178 calculated as

$$179 \quad d_t = \sum_{s=1}^S \{ |0.7 \times (rpr_s^{tc} - rpr_s^t)| + |0.3 \times (rpw_s^{tc} - rpw_s^t)| \}. \quad (1)$$

180 The chosen date  $t$  was then simply the date in  $T$  that had the minimum difference.

181 Once this date was selected, it was omitted from further consideration.

- 182 • The size of the search region for pattern matching of forecasts was  
 183 allowed to vary with forecast lead time, inspired by the results of testing the method  
 184 described in 3.b.9 of HW06. Specifically, let  $t_e$  denote the end of the forecast  
 185 precipitation accumulation period in hours, and let  $\delta$  denote the box width in units

186 of numbers of grid points on the  $\sim 1/2$ -degree Gaussian grid. If  $t_e \leq 48$ , then  $\delta=5$ ; if  
187  $48 < t_e \leq 96$ , then  $\delta=7$ ; if  $96 < t_e \leq 132$ , then  $\delta=9$ ; if  $132 < t_e$ , then  $\delta=11$ .

188         • The number of analogs used in the generation of probabilities was  
189 allowed to vary as a function of the forecast lead time and how unusual was the  
190 precipitation forecast in question, measured in terms of its percentile relative to the  
191 climatological distribution of forecasts ( $q_f$ ). Let  $n_a$  be the number of analogs used. If  
192 the end period for the forecast precipitation was  $> 48$  h, then when  $q_f < 0.75$ ,  $n_a=100$ ;  
193 when  $0.75 \leq q_f < 0.9$ ,  $n_a=75$ ; when  $0.9 \leq q_f < 0.95$ ,  $n_a=50$ ; when  $q_f > 0.95$ ,  $n_a=25$ . If the  
194 end period for the forecast  $\leq 48$  h, then when  $q_f < 0.75$ ,  $n_a=50$ ; when  $0.75 \leq q_f < 0.9$ ,  
195  $n_a=40$ ; when  $0.9 \leq q_f < 0.95$ ,  $n_a=30$ ; when  $q_f > 0.95$ ,  $n_a=20$ . This dependence of analog  
196 size on forecast lead time and unusualness of the forecast with respect to the  
197 climatology was inspired by the results of Fig. 7 and associated discussion in H06.  
198 This showed that fewer analogs provided the best skill for shorter lead times and for  
199 heavy-precipitation events; more analogs were desirable at longer leads and for  
200 more common light- or no-precipitation events. The values do not correspond  
201 exactly with the optimal values from H06 in part because the length of the training  
202 data set was somewhat shorter here (11 years with cross validation).

203

204 *c. New analog procedure with additional training data from supplemental locations.*

205         We now describe an update to the basic analog (hereafter, simply “analog”)  
206 procedure described in section 3.a.3 of HW06. This revised procedure was  
207 evaluated here against the rank-analog procedure described in section 2.b, and was

208 used in the generation of our real-time web graphics. The following modifications  
209 were made:

210       • Analogs were chosen not by finding a forecast pattern match in an area  
211 surrounding the analysis grid point of interest, but rather by using only the forecast  
212 data specifically at a grid point, as in Delle Monache et al. (2013). With this  
213 modification, data from other supplemental grid points, described below, could be  
214 used as additional training samples. In large part, the reason for not using a rank  
215 analog with a pattern match over an area was computational efficiency; with many  
216 extra supplemental locations under consideration, matching forecasts at points was  
217 much faster than matching forecasts over regions encompassing many grid points.

218       • The number of analogs used in the computation of the probabilities  
219 varied with forecast lead time. The number of analogs was defined as follows: if the  
220 end period  $t_e$  for the forecast precipitation was  $\leq 24$  h, then  $n_a=50$ ; if  $24 < t_e \leq 48$  h,  
221  $n_a=75$ ; if  $48 \leq t_e < 96$  h,  $n_a=100$ ; if  $96 \leq t_e < 120$  h,  $n_a=150$ ; if  $t_e \geq 120$  h,  $n_a=200$ .

222 However, unlike the rank-analog method described above, the number of analogs  
223 was not allowed to vary based on the unusualness of today's forecast; it was judged  
224 that ample training data was available in most situations, given the extra data from  
225 the 19 supplemental locations.

226       • In the selection of analog dates, the interpolated forecast for a particular  
227 date of interest and analysis grid point (i,j) was compared against interpolated  
228 forecasts at (i,j) for each date in the training data set. In evaluating which forecasts  
229 were closest to today's forecast, the difference between forecasts was calculated as  
230 70% of the absolute difference of the precipitation forecasts and 30% of the

231 absolute difference in precipitable water forecasts. That is, let  $pr_{i,j}^{tc}$  be the forecast  
232 precipitation amount at the grid point (i,j) and the current date, and  $pw_{i,j}^{tc}$  be total-  
233 column precipitable water. Then the difference  $d_t$  at a different date  $t$  was

$$234 \quad d_t = | 0.7 \times (pr_{i,j}^{tc} - pr_{i,j}^t) + 0.3 \times (pw_{i,j}^{tc} - pw_{i,j}^t) |. \quad (2)$$

235 Note that here the ranks of the precipitation values were not compared, as in the  
236 prior algorithm, but rather the raw forecasts values.

237       • The interpolated forecast for a particular date of interest and grid point  
238 (i,j) was also compared against interpolated forecasts at 19 other supplemental  
239 locations ( $i_s, j_s$ ) on other dates. When a the closest match was found to occur with  
240 data at one of these supplemental locations, then the analysis from this  
241 supplemental location on this date was used as an analog member. That  
242 supplemental member and date were then excluded from further consideration.  
243 The 19 supplemental locations were determined for each grid point based upon the  
244 similarity of the observed climatology and the similarity of terrain characteristics.  
245 There were also constraints on a minimum distance between supplemental  
246 locations and a penalty for distance between points. The specific methodology of  
247 defining supplemental locations is described in the online appendix A. An example  
248 of the selected supplemental locations and their relation to the local climatology is  
249 shown in Fig. 1.

250       • Once probability forecasts were generated from the ensemble of analyzed  
251 states on the dates of the selected forecast analogs, the probability forecasts were  
252 smoothed using a 2-D Savitzky-Golay smoother with a window size of 9 grid points

253 and using a third-order polynomial. The details of this smoother are also described  
254 in the online appendix A.

255 Which of the changes above were significant and which were more minor?  
256 Not considering supplemental locations, the use of the analog with point data vs. the  
257 rank analog with surrounding-area data decreased skill somewhat (not shown).  
258 However, the inclusion of supplemental training data had an even bigger positive  
259 impact and provided overall the largest impact on skill and reliability. The variable  
260 number of analog members with forecast lead produced a smaller improvement  
261 relative to using the same number at all leads. The smoothing did not affect the  
262 reliability or skill much, but the resulting forecasts were much more visually  
263 appealing. Online Appendix A provides an example of the before vs. after smoothing  
264 difference.

### 265 3. Results.

266 Figures 2 and 3 show Brier Skill Scores as a function of forecast lead time for  
267 the  $> 1 \text{ mm (12 h)}^{-1}$  event and the  $> 25 \text{ mm (12 h)}^{-1}$ , respectively. Skill scores for  
268 other event thresholds are presented in online appendix B. While both rank analog  
269 and analog forecasts provided a significant improvement with respect to the raw  
270 guidance, the skills of the warm-season forecasts at shorter leads from the newer  
271 analog method for the  $> 1 \text{ mm}$  event were slightly lower skill than those of the rank-  
272 analog method. This was likely because the  $> 1 \text{ mm}$  event was not an especially  
273 rare event at most locations, so the increased sample size with the new analog  
274 method did not compensate for the other relative advantages of using a rank-analog  
275 rather than a straight analog approach. Considering the skill for the  $> 25 \text{ mm}$  event

276 in Fig. 3, the new analog procedure did provided a skill improvement, especially for  
277 shorter-lead forecasts during the cool season. In these circumstances, the day +2  
278 analog forecasts with supplemental locations were more skillful than the day +1  
279 rank analog forecasts, and both were notably higher in skill than the raw ensemble.  
280 Why was there greater improvement of heavy precipitation forecasts with the new  
281 analog procedure in winter? Though not confirmed, we hypothesize that in winter  
282 there was higher intrinsic skill of the forecasts than in summer, due to the different  
283 phenomena driving precipitation with their different space and time scales:  
284 synoptic-scale ascent in mid-latitude winter cyclones, thunderstorms during the  
285 summer season. Further, in wintertime, there were larger fluctuations of the  
286 probabilities about their long-term climatological mean with meaningful signal.  
287 Thus the additional samples helped refine the estimates of  $O|F$ , the conditional  
288 distribution of observations given the forecast (HW06, eq. 3), thereby improving the  
289 probabilistic forecast, despite the lack of pattern matching used in the rank-analog  
290 approach.

291 Figure 4 shows maps of Brier skill scores for the > 1 mm event at the 60-72-h  
292 lead time. There was little difference between the two analog forecasts, consistent  
293 with Fig. 2. Both were more skillful than the raw ensemble, which had BSS < 0 over  
294 a significant percentage of the country, in part due to sampling error (Richardson  
295 2001) but mostly due to systematic errors and sub-optimal treatment of model  
296 uncertainty in the GEFS. Skill for all methods was largest in mountainous areas  
297 along the US West Coast, with the predictable phenomena of the flow from mid-  
298 latitude cyclones impinging upon the stationary topography. Figure 5 shows maps

299 of skill for the > 25 mm event at the 60-72-h lead time. There appeared to be a  
300 general improvement in skill across the country for the analog with supplemental  
301 locations. Again, raw ensembles were notably unskillful across drier regions of the  
302 US but competitive in a few select locations in the Sierra Nevada mountain range.  
303 Maps for other forecast lead times and thresholds are provided in online Appendix B.

304       The resulting post-processed forecast guidance was consistently reliable, too.  
305 Figure 6 provides reliability diagrams for the three methods for > 25 mm and 60-72  
306 h forecast leads; again, see appendix B for more diagrams at other leads and event  
307 thresholds. Both analog methods were quite reliable, though the analog with  
308 supplemental locations had somewhat more forecasts issuing high-probabilities  
309 (greater sharpness). Both analog methods were much less sharp than the raw  
310 forecast guidance but more reliable. Why was the analog method with  
311 supplemental locations sharper? This was because the extra training sample size  
312 permitted the identification of closer analogs than with the rank-analog approach.  
313 As noted in HW06, a general challenge with the analog or rank-analog forecasts  
314 (therein without supplemental location data) of extreme events was their inability  
315 to find many forecasts dates with amounts that were similar in magnitude.

316

#### 317 **4. Discussion and conclusions**

318       This article has demonstrated an improved method for post-processing that  
319 provides dramatically improved guidance of probabilistic precipitation when paired  
320 with a reforecast data set of sufficient length and precipitation analyses of sufficient  
321 quality. This article provides additional evidence to support the assertion that the

322 regular production of weather reforecasts will help with the objective definition of  
323 high-impact event probabilities.

324         Though the use of supplemental locations was shown to provide significant  
325 improvement to heavy precipitation forecast calibration, our examination of  
326 possible methods for choosing the location and number of supplemental location  
327 data was far from systematic. The methods for the selection of these locations  
328 deserves further study.

329         This method may provide a useful benchmark for comparison of other  
330 methods. Whereas the analog method here has been shown to work well with  
331 larger reforecast data sets, these are not always available. We anticipate  
332 subsequent studies will compare the efficacy of analog methods with respect to  
333 other (e.g., parametric) post-processing methods, including when using much  
334 smaller training sample sizes. In this way we hope to understand whether the  
335 choice of a preferred post-processing algorithm is robust from small to large  
336 training sample sizes.

337

338 **Acknowledgments:**  
339

340         This research was supported by a NOAA US Weather Program grant as well  
341 as funding from the National Weather Service Sandy Supplemental project. The  
342 reforecast data set was computed at the US Department of Energy's (DOE) National  
343 Energy Research Computing Center, a DOE Office of Science user facility.

344

345 **References**

346

347 Baxter, M. A., G. M. Lackmann, K. M. Mahoney, T. E. Workoff, and T. M. Hamill,

348 2014: [Verification of precipitation reforecasts over the Southeast United](#)

349 [States](#). *Wea. Forecasting*, **29**, 1199-1207.

350 Bukovsky, M. S., and D. J. Karoly, 2009: A brief evaluation of precipitation from the

351 North American Regional Reanalysis. *Journal Hydrometeor.*, **8**, 837-846.

352 Delle Monache, L., F. A. Eckel, D. L. Rife, B. Nagarajan, and K. Searight, 2013:

353 Probabilistic weather prediction with an analog ensemble. *Mon. Wea. Rev.*,

354 **141**, 3498–3516. doi: <http://dx.doi.org/10.1175/MWR-D-12-00281.1>

355 Hamill, T. M., J. S. Whitaker, and S. L. Mullen, 2006: [Reforecasts, an important dataset](#)

356 [for improving weather predictions](#). *Bull. Amer. Meteor. Soc.*, **87**,33-46.

357 Hamill, T. M., and J. S. Whitaker, 2006: [Probabilistic quantitative precipitation](#)

358 [forecasts based on reforecast analogs: theory and application](#) *Mon. Wea. Rev.*,

359 **134**, 3209-3229.

360 Hamill, T. M., and J. Juras, 2006: [Measuring forecast skill: is it real skill or is it the](#)

361 [varying climatology?](#) *Quart. J. Royal Meteor. Soc.*, **132**, 2905-2923.

362 Hamill, T. M., G. T. Bates, J. S. Whitaker, D. R. Murray, M. Fiorino, T. J. Galarneau, Jr., Y.

363 Zhu, and W. Lapenta, 2013: [NOAA's second-generation global medium-range](#)

364 [ensemble reforecast data set](#). *Bull Amer. Meteor. Soc.*, **94**, 1553-1565.

365 Hou, D., M. Charles, Y. Luo, Z. Toth, Y. Zhu, R. Krzysztofowicz, Y. Lin, P. Xie, D.-J. Seo,

366 M. Pena, and B. Cui, 2014: Climatology-calibrated precipitation analysis at

367 fine scales: statistical adjustment of Stage IV toward CPC gauge-based

368 analysis. *J. Hydrometeor*, **15**, 2542–2557. doi:  
369 <http://dx.doi.org/10.1175/JHM-D-11-0140.1>

370 Lin, Y., and K. E. Mitchell, 2005: The NCEP Stage II/IV hourly precipitation analyses:  
371 Development and applications. *19<sup>th</sup> Conf. on Hydrology*, San Diego, CA, Amer.  
372 Meteor. Soc., 1.2. Available online at  
373 [https://ams.confex.com/ams/Annual2005/techprogram/paper\\_83847.htm](https://ams.confex.com/ams/Annual2005/techprogram/paper_83847.htm) .

374 Mesinger, F., and others, 2006: North American regional reanalysis. *Bull. Amer.*  
375 *Meteor. Soc.*, **87**, 343-360.

376 Richardson, D. L., 2001: Measures of skill and value of ensemble prediction systems,  
377 their interrelationship and the effect of ensemble size. *Quart. J. Royal Meteor.*  
378 *Soc.*, **127**, 2473-2489.

379 West, G. L., W. J. Steenburgh, and W. Y. Y. Chen, 2007: Spurious grid-scale  
380 precipitation in the North American Regional Reanalysis. *Mon. Wea. Rev.*,  
381 **153**, 2168-2184.

382

383 **Figure captions**

384

385 **Figure 1.** Illustration of the location of supplemental locations and their  
386 dependence on the analyzed precipitation climatology. Colors denote the 95<sup>th</sup>  
387 percentile of the analysis distribution for the month of January, based on 2002-2013  
388 CCPA data. Supplemental data locations are also shown. The larger symbols indicate  
389 sample locations where supplemental data is sought, and the smaller symbols  
390 indicate the chosen supplemental locations.

391 **Figure 2:** Brier skill scores for the  $> 1 \text{ mm (12h)}^{-1}$  event over a range of lead times  
392 as a function of the month of the year. (a) Skills of forecasts from the new analog  
393 method with 19 supplemental locations; (b) skills of forecasts from the older rank-  
394 analog method for comparison; (c) skills of forecasts from the 11-member raw  
395 ensemble guidance.

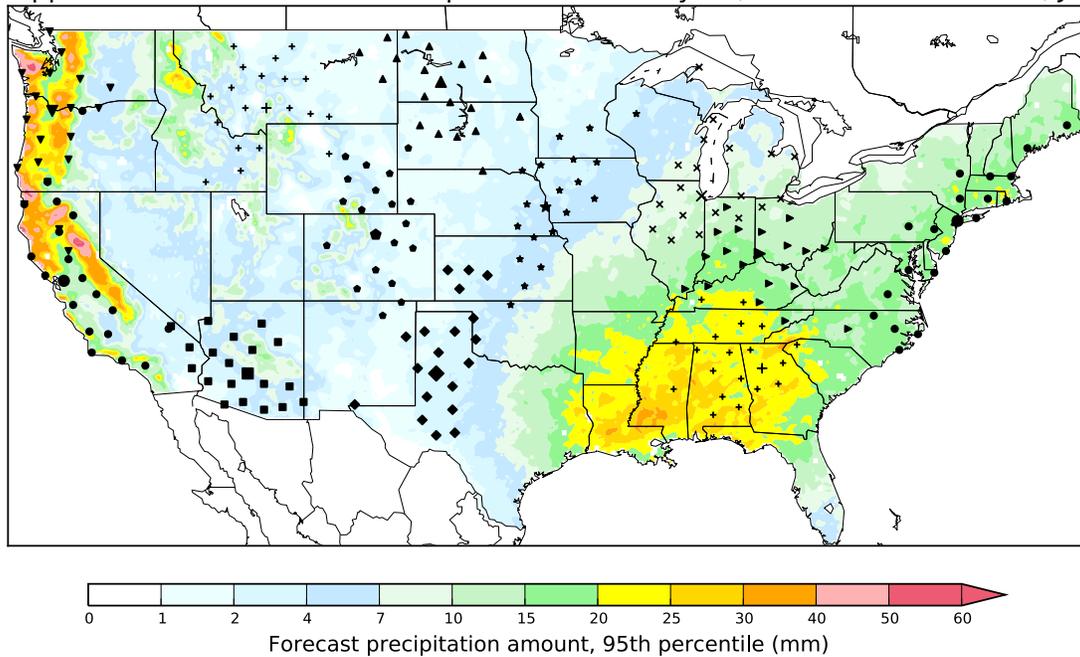
396 **Figure 3:** As in Fig. 2, but for the  $> 25 \text{ mm (12h)}^{-1}$  event. The climatology is  
397 computed separately for each month and each  $\sim 1/8$ -degree grid point location.

398 **Figure 4:** Maps of yearly 60-72 h forecast Brier Skill Scores, for probabilistic  
399 forecasts of the  $> 1 \text{ mm (12 h)}^{-1}$  event, generated from (a) analog forecasts with 19  
400 supplemental locations, (b) rank analog forecast with no supplemental locations,  
401 and (c) 11-member raw ensemble.

402 **Figure 5:** As in Fig. 4, but for  $> 25 \text{ mm}$  event.

403 **Figure 6:** Reliability diagrams for the  $> 25 \text{ mm}$  event for 60- to 72-h forecasts. (a)  
404 analog forecasts with 19 supplemental locations, (b) rank analog forecast with no  
405 supplemental locations, and (c) 11-member raw ensemble.

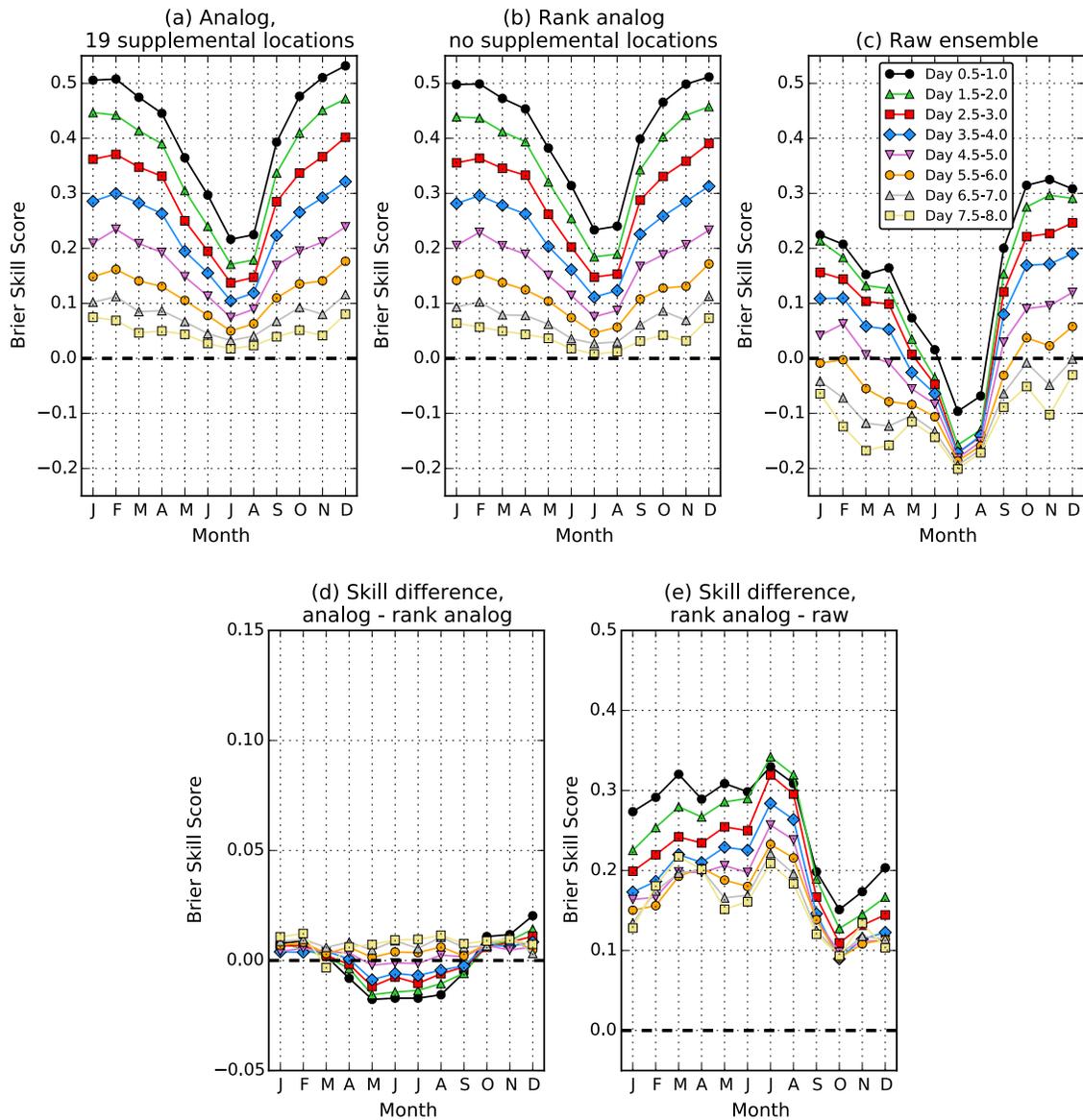
Supplemental locations and 95th percentile of analyses, 024 to 048-h forecast, Jan



406  
407  
408  
409  
410  
411  
412  
413  
414

**Figure 1.** Illustration of the location of supplemental locations and their dependence on the analyzed precipitation climatology. Colors denote the 95<sup>th</sup> percentile of the analysis distribution for the month of January, based on 2002-2013 CCPA data. Supplemental data locations are also shown. The larger symbols indicate sample locations where supplemental data is sought, and the smaller symbols indicate the chosen supplemental locations.

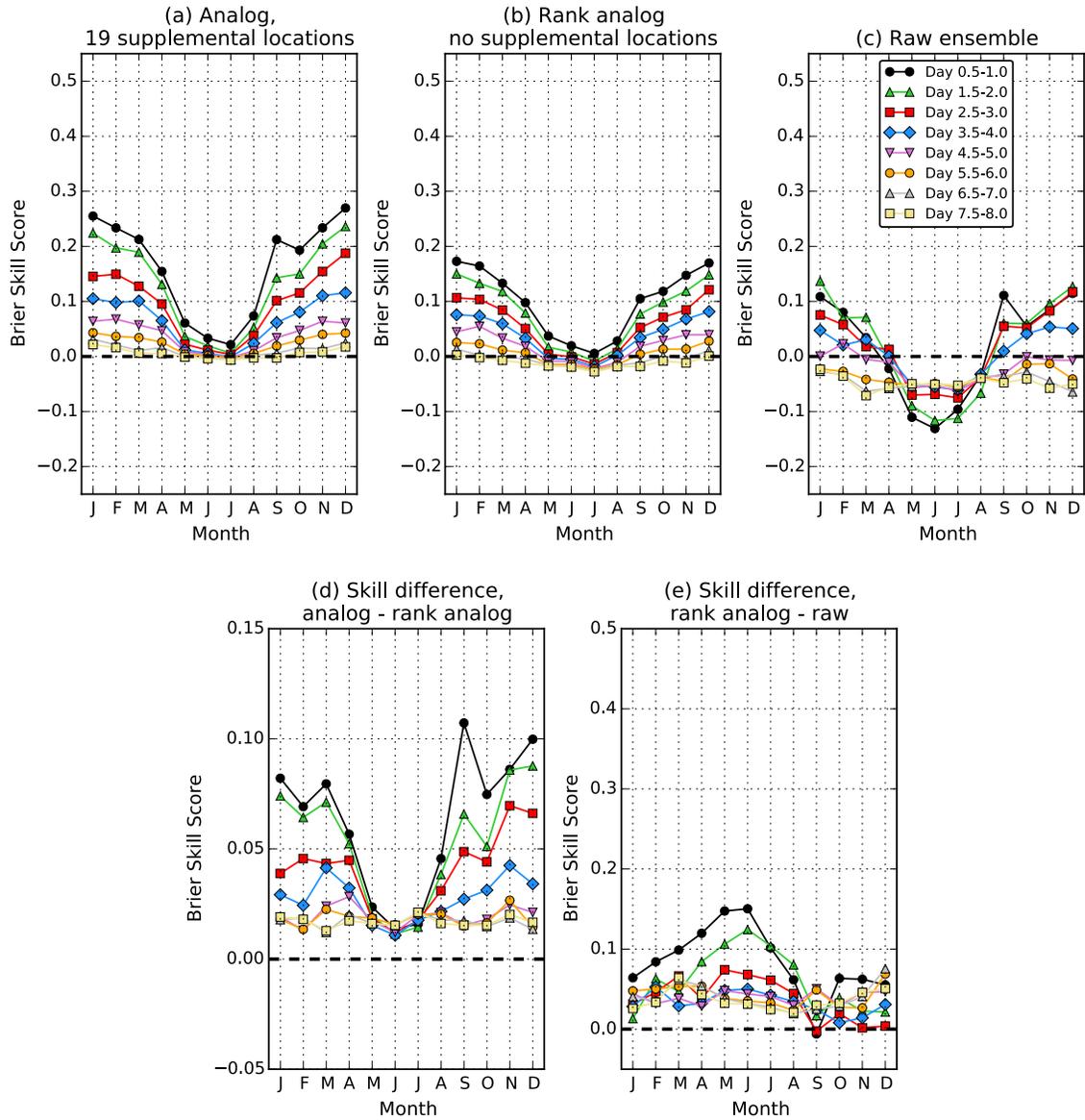
Brier skill scores, > 1mm



416  
 417  
 418  
 419  
 420  
 421  
 422  
 423

**Figure 2:** Brier skill scores for the  $> 1 \text{ mm (12 h)}^{-1}$  event over a range of lead times as a function of the month of the year. (a) Skills of forecasts from the new analog method with 19 supplemental locations; (b) skills of forecasts from the older rank-analog method for comparison; (c) skills of forecasts from the 11-member raw ensemble guidance. (d) skill difference, analog minus rank analog; (e) skill difference, rank analog minus raw.

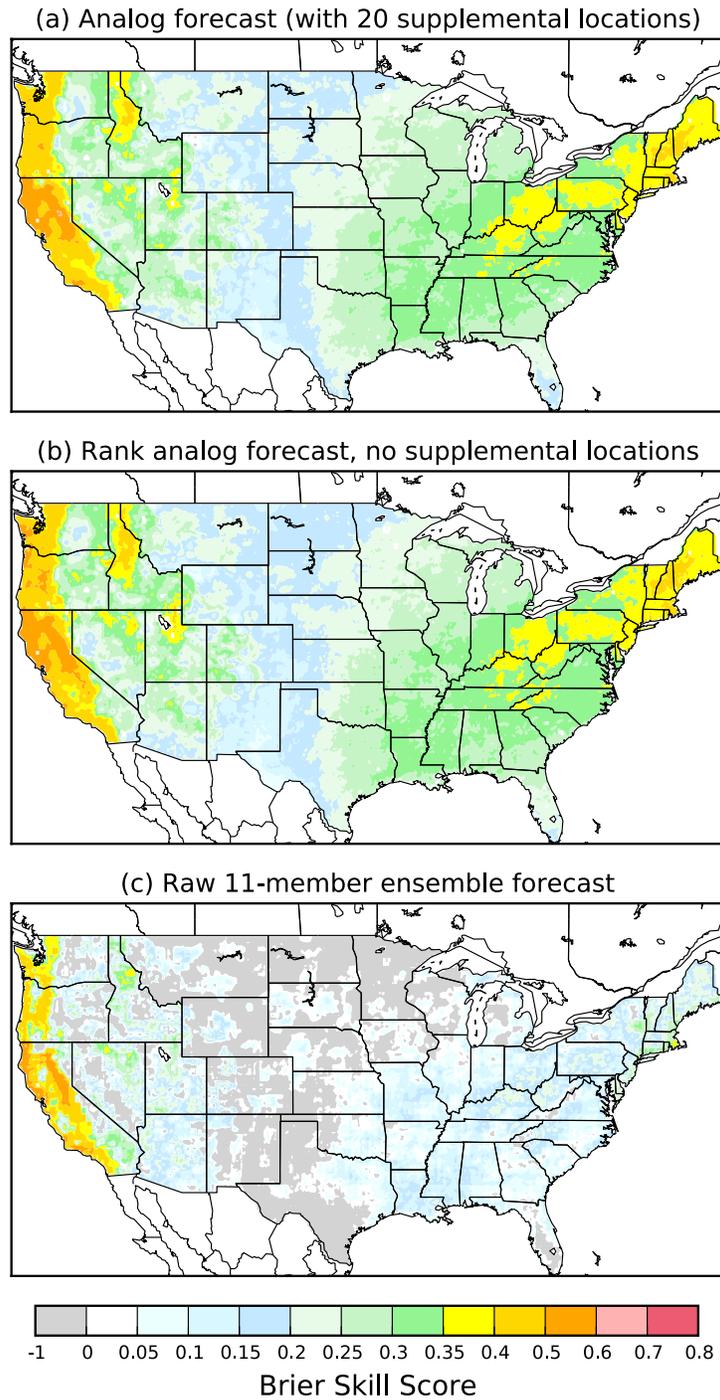
Brier skill scores, > 25mm



424  
425  
426  
427  
428

**Figure 3:** As in Fig. 2, but for the event of greater than > 25 mm (12 h)<sup>-1</sup>.

Brier Skill Scores for 060 to 072-h forecasts, > 1mm event



429

430

431 **Figure 4:** Maps of yearly 60-72 h forecast Brier Skill Scores, for probabilistic

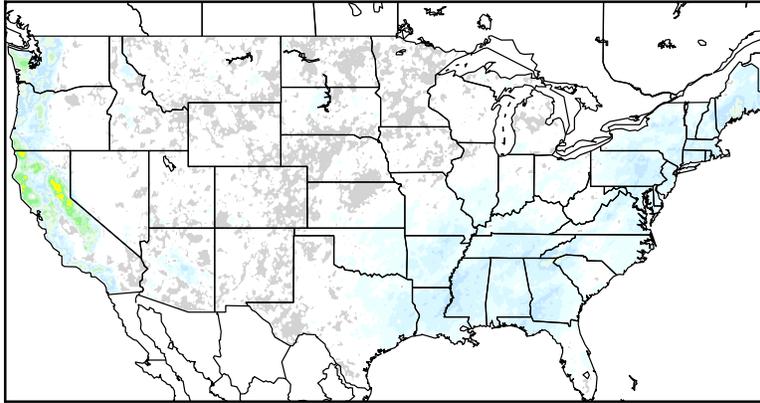
432 forecasts of the > 1 mm 12 h<sup>-1</sup> event, generated from (a) analog forecasts with 19

433 supplemental locations, (b) rank analog forecast with no supplemental locations,

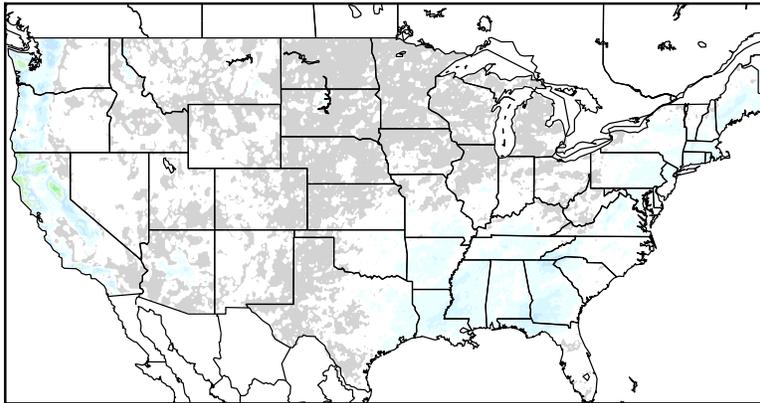
434 and (c) 11-member raw ensemble.

Brier Skill Scores for 060 to 072-h forecasts, > 25mm event

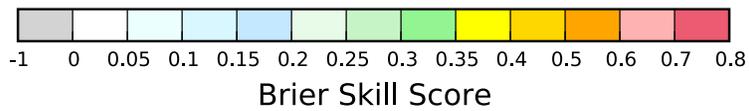
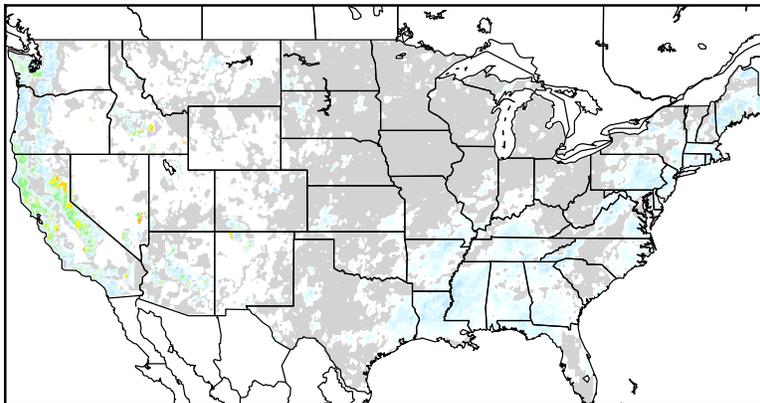
(a) Analog forecast (with 20 supplemental locations)



(b) Rank analog forecast, no supplemental locations



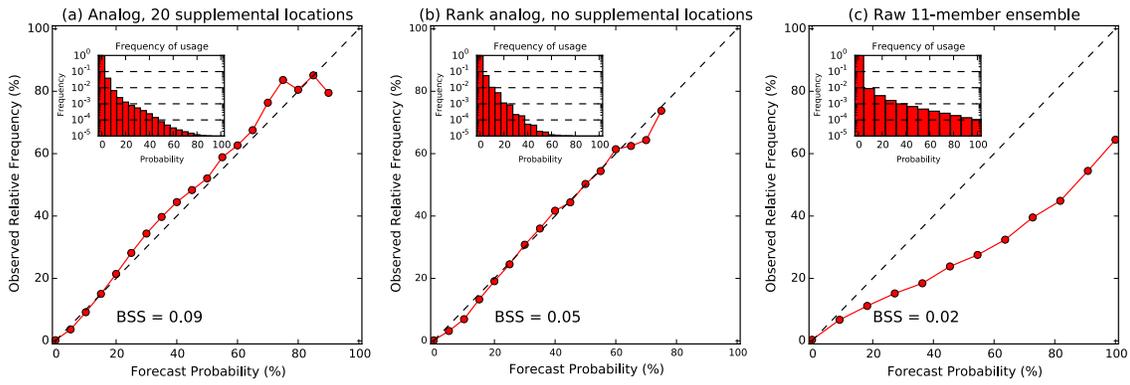
(c) Raw 11-member ensemble forecast



435  
436  
437  
438

**Figure 5:** As in Fig. 4, but for > 25 mm (12 h)<sup>-1</sup> event.

Reliability for 060-072-h, > 25mm



440  
 441  
 442  
 443  
 444  
 445  
 446  
 447

**Figure 6:** Reliability diagrams for the > 25 mm (12 h)<sup>-1</sup> event for 60- to 72-h forecasts. (a) analog forecasts with 19 supplemental locations, (b) rank analog forecast with no supplemental locations, and (c) 11-member raw ensemble.