Analog Probabilistic Precipitation Forecasts Using GEFS Reforecasts and Climatology-Calibrated Precipitation Analyses*

THOMAS M. HAMILL

NOAA/Earth System Research Laboratory/Physical Sciences Division, Boulder, Colorado

MICHAEL SCHEUERER AND GARY T. BATES

Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado

(Manuscript received 31 December 2014, in final form 11 May 2015)

ABSTRACT

Analog postprocessing methods have previously been applied using precipitation reforecasts and analyses to improve probabilistic forecast skill and reliability. A modification to a previously documented analog procedure is described here that produces highly skillful, statistically reliable precipitation forecast guidance at $1/8^\circ$ grid spacing. These experimental probabilistic forecast products are available via the web in near-real time.

The main changes to the previously documented analog algorithm were as follows: (i) use of a shorter duration (2002–13), but smaller grid spacing, higher-quality time series of precipitation analyses for training and forecast verification (i.e., the Climatology-Calibrated Precipitation Analysis); (ii) increased training sample size using data from 19 supplemental locations, chosen for their similar precipitation analysis climatologies and terrain characteristics; (iii) selection of analog dates for a particular grid point based on the similarity of forecast characteristics at that grid point rather than similarity in a neighborhood around that grid point; (iv) using an analog rather than a rank-analog approach; (v) varying the number of analogs used to estimate probabilities from a smaller number (50) for shorter-lead forecasts to a larger number (200) for longer-lead events; and (vi) spatial Savitzky–Golay smoothing of the probability fields. Special procedures were also applied near coasts and country boundaries to deal with data unavailability outside of the United States while smoothing.

The resulting forecasts are much more skillful and reliable than raw ensemble guidance across a range of event thresholds. The forecasts are not nearly as sharp, however. The use of the supplemental locations is shown to especially improve the skill of short-term forecasts during the winter.

1. Introduction

Previous studies have shown that probabilistic forecasts of precipitation can be significantly improved by postprocessing with reforecasts (e.g., Hamill et al. 2006, hereafter H06; Hamill et al. 2013, hereafter H13; Hamill and Whitaker 2006, hereafter HW06). The real-time forecast is adjusted using a long time series of past

forecasts and associated precipitation analyses. Appealing for its simplicity was the “analog” procedure used in these studies.¹ For a given location, dates in the past were identified that had reforecasts similar to today’s forecast. An ensemble was formed from the

* Supplemental information related to this paper is available at the Journals Online website: <http://dx.doi.org/10.1175/MWR-D-15-0004.s1>.

Corresponding author address: Dr. Thomas M. Hamill, NOAA/Earth System Research Laboratory/Physical Sciences Division, R/PSD 1, 325 Broadway, Boulder, CO 80305.
E-mail: tom.hamill@noaa.gov

¹ Analog procedures have been demonstrated in the past (see, e.g., Toth 1989; Van den Dool 1989; Livezey et al. 1994; Zorita and von Storch 1999; Sievers et al. 2000). There is an important distinction between the analog approach discussed here and the analog approach in these citations; here we find past dates where the forecast was similar to today’s forecast, and then select the observed on those dates. In these other articles, dates are identified where analyses or observations were similar to today’s forecast. These older analog procedures are more akin to “perfect-prog” approaches (Wilks 2011), while the current analog procedure is more akin to a model output statistics (ibid.) approach. Both can implicitly apply statistical downscaling.

observed or analyzed precipitation amounts on the dates of the chosen analogs, and probabilities were estimated from the ensemble relative frequency. Maps of precipitation probabilities were constructed by repeating the procedure across the model grid points.

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

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

This article briefly describes modifications to previously documented analog forecast procedures. What adjustments will allow it to provide improved probabilistic forecasts while using a shorter time series of analyses? We describe a series of changes to the analog algorithm and show that the resulting analog probabilistic forecasts are skillful, somewhat more sharp, and reliable. Since the statistically postprocessed guidance provides a significant improvement over probabilities from the raw Global Ensemble Forecast System (GEFS) forecast data, we are also making experimental web-based guidance available in near-real time during the next few years (<http://www.esrl.noaa.gov/psd/forecasts/reforecast2/ccpa/index.html>).

2. Methods and data

a. Reforecast data, observational data, and verification methods

In this study we considered 12-hourly accumulated precipitation forecasts during the 2002–13 period for lead times up to +8 days. Precipitation analyses were obtained on a $\sim 1/8^\circ$ grid from the CCPA dataset of Hou et al. (2014). Probabilistic forecasts were produced at this $\sim 1/8^\circ$ resolution over the CONUS. All of the forecast data used in this project were obtained from the second-generation GEFS reforecast dataset, described in H13. Ensemble-mean precipitation and total-column ensemble-mean precipitable water were used in the analog procedure. GEFS data were extracted (for precipitation) on the GEFS's native Gaussian grid at $\sim 1/2^\circ$ resolution in an area surrounding the CONUS. Precipitable water forecasts, which were archived on a 1° grid, were interpolated to the native Gaussian grid before input to the analog procedure.

Forecasts were cross validated; for example, 2002 forecasts were trained using 2003–13 data. For the production of forecasts in a given month, the training data used that month and the surrounding two months (e.g., January forecasts were trained with December–January–February data). Was the use of future data in the cross-validation procedure a source of unrealistic skill of these forecasts? As shown in Baxter et al. (2014, their Fig. 5), the interannual variability of skill in the Southeast United States was larger than the systematic changes from 2002 to 2013. This suggests that the use of future forecasts in the cross-validation procedure probably did not result in a large overestimation of forecast skill for the earlier years.

One of the controls against which the new method was compared were the raw event probabilities generated from the 11-member GEFS reforecast ensemble, bilinearly interpolated to the $1/8^\circ$ grid.

Verification methods included reliability diagrams and Brier skill scores computed in the conventional way [Wilks (2011), his Eqs. (8.36) and (8.37); Hamill and Juras (2006)], with climatology providing the reference probabilistic forecasts. Maps of Brier skill scores were also generated for the CONUS. These were produced by accumulating the probabilistic forecasts' and climatological forecasts' average of squared error at that grid point across all years and all months prior to the calculation of skill. Because of the extremely large sample size, confidence intervals for the skill differences (very small; see HW06) were not included on the plots.

b. Rank-analog forecast procedure as a control

A revised "rank analog" approach served as another standard of comparison for the newer, somewhat more involved analog methodology described in section 2c

below. For the most part, the rank-analog approach was a hybrid of the techniques that have previously been shown to work well, described in sections 3b(6) and 3b(8) in [HW06](#). This control rank-analog methodology was further updated in the following respects:

- As with the rank-analog algorithm of [HW06](#), the rank of the forecast for a particular date of interest and set of grid points was compared against the ranks of sorted forecasts at the same set of grid points for each date in the training dataset. In evaluating which forecasts were closest to today's forecast, the difference between forecasts was calculated as 70% of the absolute difference of the precipitation forecast ranks and 30% of the absolute difference in precipitable water forecast ranks averaged over the set of grid points, following [HW06](#) (other predictors were tested in that article as well). As shown therein, with the exception of warm-season probability of precipitation, there was minimal sensitivity to the chosen weight between precipitation and precipitable water. A more precise definition of the forecast difference is as follows: let S be the set of grid points in a region surrounding the current grid point of interest. Let t_c be the current date, and let t be another date from the set of dates T whose forecast data will be compared against the forecast at t_c . As indicated previously, by cross validation $t_c \notin T$. Define $\text{rpr}_s^{t_c}$ as the rank of the current forecast precipitation amount at time t_c and at grid point s from a combined set with the training data at s . Similarly, $\text{rpw}_s^{t_c}$ is the associated rank of the current total-column precipitable water forecast. Then the difference in ranks for date t was calculated as

$$d_t = \sum_{s=1}^S [0.7 \times |\text{rpr}_s^{t_c} - \text{rpr}_s^t| + 0.3 \times |\text{rpw}_s^{t_c} - \text{rpw}_s^t|]. \quad (1)$$

The chosen date t was then simply the date in T that had the minimum difference. Once this date was selected, it was omitted from further consideration.

- The size of the search region for pattern matching of forecasts was allowed to vary with forecast lead time, inspired by the results of testing the method described in section 3b(9) in [HW06](#). Specifically, let t_e denote the end of the forecast precipitation accumulation period in hours, and let δ denote the box width in units of numbers of grid points on the $\sim 1/2^\circ$ Gaussian grid. If $t_e \leq 48$, then $\delta = 5$; if $48 < t_e \leq 96$, then $\delta = 7$; if $96 < t_e \leq 132$, then $\delta = 9$; and if $132 < t_e$, then $\delta = 11$.
- The number of analogs used in the generation of probabilities was allowed to vary as a function of the forecast lead time and how unusual the precipitation forecast in question was, measured in terms of its

percentile relative to the climatological distribution of forecasts (q_f). Let n_a be the number of analogs used. If the end period for the forecast precipitation was >48 h, then when $q_f < 0.75$, $n_a = 100$; when $0.75 \leq q_f < 0.9$, $n_a = 75$; when $0.9 \leq q_f < 0.95$, $n_a = 50$; and when $q_f > 0.95$, $n_a = 25$. If the end period for the forecast ≤ 48 h, then when $q_f < 0.75$, $n_a = 50$; when $0.75 \leq q_f < 0.9$, $n_a = 40$; when $0.9 \leq q_f < 0.95$, $n_a = 30$; and when $q_f > 0.95$, $n_a = 20$. This dependence of analog size on forecast lead time and unusualness of the forecast with respect to the climatology was inspired by the results of Fig. 7 in [H06](#) and the associated discussion. This showed that fewer analogs provided the best skill for shorter lead times and for heavy-precipitation events; more analogs were desirable at longer leads and for more common light- or no-precipitation events. The values do not correspond exactly with the optimal values from [H06](#) in part because the length of the training dataset was somewhat shorter here (11 years with cross validation).

- To deal with the possibility of ties (e.g., no precipitation in the forecast) and to permit a diverse set of analog dates in such a circumstance, a very small random number was added to each difference in rank before sorting. In most circumstances, though, ties are very unlikely because of the precipitable water forecasts, which are unlikely to have ties even if precipitation forecasts do.

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

We now describe an update to the basic analog (hereafter, simply “analog”) procedure described in section 3a (3) of [HW06](#). This revised procedure was evaluated here against the rank-analog procedure described in [section 2b](#), and was used in the generation of our real-time web graphics. The following modifications were made:

- Analogues were chosen not by finding a forecast pattern match in an area surrounding the analysis grid point of interest, but rather by using only the forecast data specifically at a grid point, as in [Delle Monache et al. \(2013\)](#). With this modification, data from other supplemental grid points, described below, could be used as additional training samples. In large part, the reason for not using a rank analog with a pattern match over an area was computational efficiency; with many extra supplemental locations under consideration, matching forecasts at points was much faster than matching forecasts over regions encompassing many grid points.
- The number of analogs used in the computation of the probabilities varied with forecast lead time. The number of analogs was defined as follows: if the end period t_e for the forecast precipitation was ≤ 24 h, then $n_a = 50$; if

FIG. 1. Illustration of the location of supplemental locations and their dependence on the analyzed precipitation climatology. Colors denote the 95th percentile of the analysis distribution for the month of January, based on 2002–13 CCPA data. Supplemental data locations are also shown. The larger symbols indicate sample locations where supplemental data are sought, and the smaller symbols indicate the chosen supplemental locations.

$24 < t_e \leq 48$ h, $n_a = 75$; if $48 \leq t_e < 96$ h, $n_a = 100$; if $96 \leq t_e < 120$ h, $n_a = 150$; and if $t_e \geq 120$ h, $n_a = 200$.

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

- In the selection of analog dates, the interpolated forecast for a particular date of interest and analysis grid point (i, j) was compared against interpolated forecasts at (i, j) for each date in the training dataset. In evaluating which forecasts were closest to today's forecast, the difference between forecasts was calculated as 70% of the absolute difference of the precipitation forecasts and 30% of the absolute difference in precipitable water forecasts. That is, let $\text{pr}_{i,j}^{\text{tc}}$ be the forecast precipitation amount at the grid point (i, j) and the current date, and $\text{pw}_{i,j}^{\text{tc}}$ be total-column precipitable water. Then the difference d_t at a different date t was

$$d_t = |0.7 \times (\text{pr}_{i,j}^{\text{tc}} - \text{pr}_{i,j}^t) + 0.3 \times (\text{pw}_{i,j}^{\text{tc}} - \text{pw}_{i,j}^t)|. \quad (2)$$

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

- The interpolated forecast for a particular date of interest and grid point (i, j) was also compared against interpolated forecasts at 19 other supplemental locations (i_s, j_s) on other dates. When a the closest match was found to occur with data at one of these supplemental locations, then the analysis from this supplemental location on this date was used as an analog member. That supplemental member and date were then excluded from further consideration. The 19 supplemental locations were determined for each grid point based upon the similarity of the observed climatology and the similarity of terrain characteristics. There were also constraints on a minimum distance between supplemental locations and a penalty for distance between points. The specific methodology of defining supplemental locations is described in appendix A in the online supplemental material. An example of the selected supplemental locations and their relation to the local climatology is shown in Fig. 1.
- Once probability forecasts were generated from the ensemble of analyzed states on the dates of the selected forecast analogs, the probability forecasts were smoothed using a 2D Savitzky–Golay smoother with a window size of nine grid points and using a third-order polynomial.

The details of this smoother are also described in appendix A in the online supplemental material.

Which of the changes above were significant and which were more minor?

Not considering supplemental locations, the use of the analog with point data versus the rank analog with surrounding-area data decreased skill somewhat (not shown). However, the inclusion of supplemental training data had an even bigger positive impact and provided overall the largest impact on skill and reliability. The variable number of analog members with forecast lead produced a smaller improvement relative to using the same number at all leads. The smoothing did not affect the reliability or skill much, but the resulting forecasts were much more visually appealing. Appendix A in the online supplemental material provides an example of the before versus after smoothing difference.

3. Results

Figures 2 and 3 show Brier skill scores as a function of forecast lead time for the >1 mm $(12\text{ h})^{-1}$ event and the >25 mm $(12\text{ h})^{-1}$, respectively. Skill scores for other event thresholds are presented in appendix B in the online supplemental material. While both rank-analog and analog forecasts provided a significant improvement with respect to the raw guidance, the skills of the warm-season forecasts at shorter leads from the newer analog method for the >1 -mm event were slightly lower skill than those of the rank-analog method. This was likely because the >1 -mm event was not an especially rare event at most locations, so the increased sample size with the new analog method did not compensate for the other relative advantages of using a rank-analog rather than a straight-analog approach. Considering the skill for the >25 -mm event in Fig. 3, the new analog procedure did provide a skill improvement, especially for shorter-lead forecasts during the cool season. In these circumstances, the day +2 analog forecasts with supplemental locations were more skillful than the day +1 rank-analog forecasts, and both were notably higher in skill than the raw ensemble. Why was there greater improvement of heavy precipitation forecasts with the new analog procedure in winter? Though not confirmed, we hypothesize that in winter there was higher intrinsic skill of the forecasts than in summer due to the different phenomena driving precipitation with their different space and time scales: synoptic-scale ascent in mid-latitude winter cyclones and thunderstorms during the summer season. Further, in wintertime, there were larger fluctuations of the probabilities about their long-term climatological mean with meaningful signal. Thus,

the additional samples helped refine the estimates of $O|F$, the conditional distribution of observations given the forecast [HW06, their Eq. (3)], thereby improving the probabilistic forecast, despite the lack of pattern matching used in the rank-analog approach.

Figure 4 shows maps of Brier skill scores for the >1 -mm event at the 60–72-h lead time. There was little difference between the two analog forecasts, consistent with Fig. 2. Both were more skillful than the raw ensemble, which had BSS < 0 over a significant percentage of the country, in part due to sampling error (Richardson 2001), but mostly due to systematic errors and suboptimal treatment of model uncertainty in the GEFS. Skill for all methods was largest in mountainous areas along the U.S. West Coast, with the predictable phenomena of the flow from midlatitude cyclones impinging upon the stationary topography. Figure 5 shows maps of skill for the >25 -mm event at the 60–72-h lead time. There appeared to be a general improvement in skill across the country for the analog with supplemental locations. Again, raw ensembles were notably unskillful across drier regions of the United States but competitive in a few select locations in the Sierra Nevada range. Maps for other forecast lead times and thresholds are provided in appendix B in the online supplemental material.

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

4. Discussion and conclusions

This article has demonstrated an improved method for postprocessing that provides dramatically improved guidance of probabilistic precipitation when paired with a reforecast dataset of sufficient length and precipitation analyses of sufficient quality. This article

Brier skill scores, > 1mm

FIG. 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, and (e) skill difference, rank analog minus raw.

provides additional evidence to support the assertion that the regular production of weather reforecasts will help with the objective definition of relatively heavy precipitation event probabilities.

Though the use of supplemental locations was shown to provide significant improvement to heavy precipitation forecast calibration, our examination of possible methods for choosing the location and number of

Brier skill scores, > 25mm

FIG. 3. As in Fig. 2, but for the $>25 \text{ mm (12 h)}^{-1}$ event. The climatology is computed separately for each month and each $\sim 1/8^\circ$ gridpoint location.

supplemental location data was far from systematic. The methods for the selection of these locations deserve further study.

This method may provide a useful benchmark for comparison of other methods. Whereas the analog method here has been shown to work well with larger

reforecast datasets, these are not always available. We anticipate subsequent studies will compare the efficacy of analog methods with respect to other (e.g., parametric) postprocessing methods, including when using much smaller training sample sizes. In this way we hope to understand whether the choice of a preferred

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

(a) Analog forecast (with 20 supplemental locations)

(b) Rank analog forecast, no supplemental locations

(c) Raw 11-member ensemble forecast

FIG. 4. Maps of yearly 60–72-h forecast Brier skill scores, for probabilistic forecasts of the $>1 \text{ mm (12 h)}^{-1}$ event, generated from (a) analog forecasts with 19 supplemental locations, (b) rank-analog forecast with no supplemental locations, and (c) 11-member raw ensemble.

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

FIG. 5. As in Fig. 4, but for the $>25 \text{ mm (12 h)}^{-1}$ event.

Reliability for 060-072-h, > 25mm

FIG. 6. Reliability diagrams for the >25-mm event for 60–72-h forecasts. (a) Analog forecasts with 19 supplemental locations, (b) rank-analog forecast with no supplemental locations, and (c) 11-member raw ensemble.

postprocessing algorithm is robust from small to large training sample sizes.

Acknowledgments. This research was supported by a NOAA U.S. Weather Program grant as well as funding from the National Weather Service Sandy Supplemental project. The reforecast dataset was computed at the U.S. Department of Energy's (DOE) National Energy Research Computing Center, a DOE Office of Science user facility.

REFERENCES

- Baxter, M. A., G. M. Lackmann, K. M. Mahoney, T. E. Workoff, and T. M. Hamill, 2014: Verification of precipitation reforecasts over the Southeast United States. *Wea. Forecasting*, **29**, 1199–1207, doi:[10.1175/WAF-D-14-00055.1](https://doi.org/10.1175/WAF-D-14-00055.1).
- Bukovsky, M. S., and D. J. Karoly, 2007: A brief evaluation of precipitation from the North American Regional Reanalysis. *J. Hydrometeorol.*, **8**, 837–846, doi:[10.1175/JHM595.1](https://doi.org/10.1175/JHM595.1).
- Delle Monache, L., F. A. Eckel, D. L. Rife, B. Nagarajan, and K. Searight, 2013: Probabilistic weather prediction with an analog ensemble. *Mon. Wea. Rev.*, **141**, 3498–3516, doi:[10.1175/MWR-D-12-00281.1](https://doi.org/10.1175/MWR-D-12-00281.1).
- Hamill, T. M., and J. Juras, 2006: Measuring forecast skill: Is it real skill or is it the varying climatology? *Quart. J. Roy. Meteor. Soc.*, **132**, 2905–2923, doi:[10.1256/qj.06.25](https://doi.org/10.1256/qj.06.25).
- , and J. S. Whitaker, 2006: Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory and application. *Mon. Wea. Rev.*, **134**, 3209–3229, doi:[10.1175/MWR3237.1](https://doi.org/10.1175/MWR3237.1).
- , —, and S. L. Mullen, 2006: Reforecasts: An important dataset for improving weather predictions. *Bull. Amer. Meteor. Soc.*, **87**, 33–46, doi:[10.1175/BAMS-87-1-33](https://doi.org/10.1175/BAMS-87-1-33).
- , G. T. Bates, J. S. Whitaker, D. R. Murray, M. Fiorino, T. J. Galarneau Jr., Y. Zhu, and W. Lapenta, 2013: NOAA's second-generation global medium-range ensemble reforecast dataset. *Bull. Amer. Meteor. Soc.*, **94**, 1553–1565, doi:[10.1175/BAMS-D-12-00014.1](https://doi.org/10.1175/BAMS-D-12-00014.1).
- Hou, D., and Coauthors, 2014: Climatology-calibrated precipitation analysis at fine scales: Statistical adjustment of Stage IV toward CPC gauge-based analysis. *J. Hydrometeorol.*, **15**, 2542–2557, doi:[10.1175/JHM-D-11-0140.1](https://doi.org/10.1175/JHM-D-11-0140.1).
- Livezey, R. E., A. G. Barnston, G. V. Gruza, and E. Y. Ran'kova, 1994: Comparative skill of two analog seasonal temperature prediction systems: Objective selection of predictors. *J. Climate*, **7**, 608–615, doi:[10.1175/1520-0442\(1994\)007<0608:CSOTAS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1994)007<0608:CSOTAS>2.0.CO;2).
- Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343–360, doi:[10.1175/BAMS-87-3-343](https://doi.org/10.1175/BAMS-87-3-343).
- Richardson, D. L., 2001: Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size. *Quart. J. Roy. Meteor. Soc.*, **127**, 2473–2489, doi:[10.1002/qj.4971275715](https://doi.org/10.1002/qj.4971275715).
- Sievers, O., K. Fraedrich, and C. C. Raible, 2000: Self-adapting analog ensemble predictions of tropical cyclone tracks. *Wea. Forecasting*, **15**, 623–629, doi:[10.1175/1520-0434\(2000\)015<0623:SAAEPO>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0623:SAAEPO>2.0.CO;2).
- Toth, Z., 1989: Long-range weather forecasting using an analog approach. *J. Climate*, **2**, 594–607, doi:[10.1175/1520-0442\(1989\)002<0594:LRWFUA>2.0.CO;2](https://doi.org/10.1175/1520-0442(1989)002<0594:LRWFUA>2.0.CO;2).
- Van den Dool, H. M., 1989: A new look at weather forecasting through analogues. *Mon. Wea. Rev.*, **117**, 2230–2247, doi:[10.1175/1520-0493\(1989\)117<2230:ANLAWF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1989)117<2230:ANLAWF>2.0.CO;2).
- West, G. L., W. J. Steenburgh, and W. Y. Y. Chen, 2007: Spurious grid-scale precipitation in the North American Regional Reanalysis. *Mon. Wea. Rev.*, **135**, 2168–2184, doi:[10.1175/MWR3375.1](https://doi.org/10.1175/MWR3375.1).
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. Academic Press, 676 pp.
- Zorita, E., and H. von Storch, 1999: The analog method as a simple statistical downscaling technique: Comparison with more complicated methods. *J. Climate*, **12**, 2474–2489, doi:[10.1175/1520-0442\(1999\)012<2474:TAMAAS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<2474:TAMAAS>2.0.CO;2).