

# The Utility of the NOAA Reforecast Dataset for Quantitative Precipitation Forecasting over the Coastal Western United States

BENJAMIN J. MOORE

*Cooperative Institute for Research in the Environmental Sciences and NOAA/Earth System Research Laboratory, Boulder, Colorado*

THOMAS M. HAMILL

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

ELLEN M. SUKOVICH

*Cooperative Institute for Research in the Environmental Sciences and NOAA/Earth System Research Laboratory, Boulder, Colorado*

THOMAS WORKOFF and FAYE E. BARTHOLD

*I.M. Systems Group, Inc. and NOAA/NWS/NCEP Weather Prediction Center, College Park, Maryland*

(Manuscript received 23 April 2015; review completed 26 August 2015)

## ABSTRACT

The National Oceanic and Atmospheric Administration’s second-generation Global Ensemble Forecast System reforecast dataset is utilized to generate probabilistic quantitative precipitation forecasts (PQPFs) for the cool season during 1985–2011 over the West Coast of the United States using a rank analog post-processing technique. It is found that the rank analog technique produced large improvements in forecast skill and reliability over the raw uncalibrated ensemble. Forecasts for three heavy precipitation events associated with atmospheric rivers are examined to demonstrate the utility of the rank analog technique. The rank analog technique yielded increases in forecast skill for two of the three cases but decreases in forecast skill for the other case, indicating that in some instances the technique can perform poorly. Overall, the results of this study are consistent with the findings from the 2012 Atmospheric River Retrospective Forecasting Experiment, in which participants found the reforecast-based rank analog PQPFs to provide useful guidance for forecasting precipitation over the West Coast of the United States.

## 1. Introduction

The western coast of the United States (hereafter, the “West Coast”) receives a majority of its annual precipitation and extreme precipitation events during the cool season (e.g., Ralph and Dettinger 2012) in connection with landfalling extratropical cyclones developing near the terminus of the North Pacific jet stream. These cyclones frequently are accompanied by enhanced horizontal water vapor transport concentrated within long (>2000 km), narrow (<1000 km) corridors called atmospheric rivers (ARs; e.g., Newell et al. 1992; Zhu and Newell 1998; Ralph et al. 2004; Sodemann and Stohl 2013). When ARs make landfall

and impinge upon the mountainous terrain of the West Coast, heavy orographically forced precipitation can be produced (e.g., Neiman et al. 2008a,b, 2014; Smith et al. 2010; Ralph et al. 2011; Ralph and Dettinger 2012). This heavy precipitation can result not only in high-impact flooding (e.g., Ralph et al. 2006; Neiman et al. 2008b, 2011) but also in increased snowpack at high elevations (e.g., Guan et al. 2013) and significant contributions to regional water supplies (e.g., Dettinger et al. 2011, Dettinger 2013). Given these socioeconomic and environmental impacts, accurate and reliable forecasts of precipitation along the West Coast are of critical importance.

To address the forecasting challenges associated with ARs, the National Oceanic and Atmospheric Administration (NOAA) Hydrometeorological Testbed partnered with the National Centers for Environmental Prediction's (NCEP's) Weather Prediction Center (WPC) to conduct the Atmospheric River Retrospective Forecasting Experiment (ARRFEX) in September 2012. ARRFEX brought together model developers, researchers, and National Weather Service forecasters in a pseudo-operational setting to focus on identifying new and old techniques, guidance, and tools that could help improve forecasts of AR-related precipitation.

One of the major outcomes of ARRFEX was the success of probabilistic quantitative precipitation forecasts (PQPFs) generated using NOAA's second-generation Global Ensemble Forecast System (GEFS) reforecast dataset. The reforecast PQPF guidance, generated using a "rank analog" statistical post-processing technique (e.g., Hamill and Whitaker 2006; Hamill et al. 2013), was provided to participants as an experimental forecast product for use in the forecast process during ARRFEX. Participants concluded that the reforecast PQPF guidance was superior to operational ensemble guidance for AR events, particularly at 3–5-day lead times. The reforecast PQPFs consistently and accurately captured areas that were at risk for receiving heavy precipitation [defined as  $>3.0$  in ( $\sim 75$  mm) in 24 h]. Based on these results, WPC began to work with NOAA's Earth Systems Research Laboratory to establish data transfer of the reforecast PQPF products to WPC, allowing WPC forecasters to utilize the reforecast data on the operations floor in real-time. Although the reforecast PQPFs performed well during ARRFEX, the forecast evaluations in the experiment were subjective in nature. Further quantitative analysis is required to determine the benefit of reforecast PQPFs for West Coast precipitation forecasting.

In the current study, we apply a rank analog technique to the GEFS reforecast dataset to produce calibrated PQPFs of 24-h accumulated precipitation over the West Coast for the cool season (October–March) during 1985–2011. The primary objectives of this study are to 1) quantitatively assess whether the application of the rank analog technique to the GEFS reforecast dataset can yield significant improvements to precipitation forecast skill along the West Coast and 2) demonstrate the utility of analog-based PQPFs from the GEFS reforecast dataset for forecasting heavy precipitation associated with landfalling ARs.

## 2. Data and methods

### a. Reforecast and verification datasets

The NOAA second-generation GEFS reforecast dataset consists of an extensive (1985–present) archive of daily (initialized at 0000 UTC) 0–16-day ensemble forecasts generated with a model configuration consistent with the 2012–2014 operational version (9.0.1) of the NCEP GEFS model. The GEFS reforecasts are run at T254L42 resolution ( $\sim 40$  km at  $40^\circ\text{N}$ ) for the first seven days of the forecast and at T190L42 resolution ( $\sim 54$  km at  $40^\circ\text{N}$ ) thereafter, and comprise one control run and ten members with perturbed initial conditions. Refer to Hamill et al. (2013) for complete details regarding the GEFS reforecast dataset.<sup>1,2</sup>

As in Hamill and Whitaker (2006) and Hamill et al. (2013), we elected to use 24-h accumulated precipitation analyses (valid at 0000 UTC) from the North American Regional Reanalysis (NARR; Mesinger et al. 2006)—available for 1979–present on a 32-km Lambert conformal grid covering North America—for forecast calibration (i.e., training) and verification. The GEFS reforecast output was interpolated to the NARR grid. There are drawbacks to using the NARR precipitation analyses, which may limit the quality of the associated calibrated PQPFs. Specifically, these analyses have been shown to contain systematic errors (e.g., Bukovsky and Karoly 2007; West et al. 2007) and, at 32-km horizontal grid spacing, are unable to fully resolve the orographic modulation of precipitation in areas of complex terrain. Multi-sensor precipitation analyses, such as the NCEP Stage-IV or the Climatology-Calibrated Precipitation Analysis (Hou et al. 2014), available at a much higher horizontal grid spacing (e.g., 4 km), can provide more accurate precipitation measurements and can be useful for generating reforecast-based analog forecasts (Hamill et al. 2015). However, these analyses typically are only available for 2002–present, thus limiting the size of training samples with which to calibrate forecasts. The NARR was selected for this study as it provides continuous coverage over the West Coast for the entire period of the GEFS reforecast dataset and thus enables the use of large training samples to calibrate forecasts.

<sup>1</sup> The reforecast data can be accessed online at [esrl.noaa.gov/psd/forecasts/reforecast2/download.html](http://esrl.noaa.gov/psd/forecasts/reforecast2/download.html).

<sup>2</sup> Real-time reforecast-based deterministic and probabilistic forecast products can be accessed online at [esrl.noaa.gov/psd/forecasts/reforecast2/](http://esrl.noaa.gov/psd/forecasts/reforecast2/).

### *b. Rank analog technique*

Prior studies have applied analog techniques to the first-generation (e.g., Hamill et al. 2006; Hamill and Whitaker 2006) and second-generation (e.g., Hamill et al. 2013, 2015) GEFS reforecast datasets to produce calibrated short- and medium-range (i.e., 1–7 day) PQPFs over the conterminous United States, demonstrating a significant improvement in forecast skill over uncalibrated ensemble PQPFs. These analog techniques generally involve a search for analog dates on which the precipitation forecast was similar to the current precipitation forecast. An ensemble is then constructed from the observed or analyzed conditions (i.e., training data) on those analog dates. From this ensemble of observed conditions, probabilities of exceedance of various precipitation thresholds then can be computed. This type of analog technique can correct for systematic model errors and biases and can statistically downscale precipitation forecasts if the training data are at a higher spatial resolution than the forecast model. Refer to Hamill and Whitaker (2006) for further information regarding reforecast-based analog techniques.

In this study, we employed a rank analog technique (e.g., Hamill and Whitaker 2006; Hamill et al. 2013) to generate calibrated PQPFs. This technique involved the following steps: 1) for the date and lead time of the current forecast, a subset of ensemble-mean 24-h accumulated precipitation forecasts at the same lead time during 1985–2011 were loaded from the reforecast dataset for the month of the current forecast and the two surrounding months; 2) at each grid point this subset of ensemble-mean forecasts—including the current ensemble-mean forecast—was sorted, and each forecast was assigned a rank; 3) analogs were selected as the  $n$  dates with the closest forecast ranks to that of the current forecast; and 4) an ensemble of NARR 24-h accumulated precipitation analyses corresponding to the  $n$  analog dates was constructed, and PQPFs were computed as the relative frequency of an event in that ensemble. In step 3, the differences between the current forecast rank and the reforecast ranks were summed over a  $7 \times 7$  grid-point ( $224 \text{ km} \times 224 \text{ km}$ ) box centered on each grid point, and analogs were selected as the  $n$  cases with the lowest rank difference value. This procedure allowed for some degree of precipitation pattern matching.

As in Hamill et al. (2013), the number of analogs used ( $n$ ) was allowed to vary based on the rarity of the forecast event relative to the reforecast climatology.

For example, if an event was in the top 1% of the climatological distribution, only 30 analogs were used. If an event was very common, say in the bottom 50% of the climatological distribution, up to 200 analogs were used. This variation in the number of analogs is motivated by the results of Hamill and Whitaker (2006), which demonstrated that forecast skill for rare events often is improved when fewer analogs are used. As in Hamill et al. (2013), deterministic forecasts were produced from the rank analog output using Ebert's probability-matched mean approach ([www.cawcr.gov.au/staff/eee/etrap/probmatch.html](http://www.cawcr.gov.au/staff/eee/etrap/probmatch.html)).

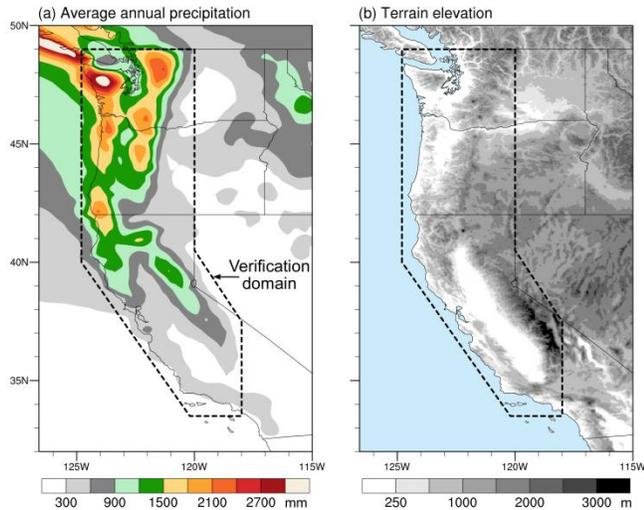
### *c. Verification*

To demonstrate the improvements of the rank analog technique over the raw, uncalibrated reforecast ensemble (hereafter referred to as the raw ensemble), the skill and reliability of PQPFs of 24-h accumulated precipitation generated using the rank analog technique and those generated from the raw ensemble relative frequencies were compared for October–March during 1985–2011. The ensemble relative frequencies were computed at each grid point by simply adding up the number of ensemble members in which precipitation exceeded a given threshold and then dividing by the total number of ensemble members (11 in this case). The skill of the PQPFs was assessed using the Brier skill score [BSS; Wilks 2011, his Eqs. (8.36) and (8.37)]. The BSS measures the skill of PQPFs relative to reference climatological forecasts, which were derived from monthly climatological distributions of 24-h precipitation from the NARR. A BSS  $>0$  indicates that a forecast has relative skill over the climatological forecast. The reliability of the PQPFs was assessed using reliability diagrams (Wilks 2011, section 8.44), which depict the level of correspondence between forecast probabilities and analyzed relative frequencies of a given event. In these diagrams, curves lying nearly along the diagonal indicate that forecasts are reliable, whereas curves lying far above (below) the diagonal indicate a tendency to underforecast (overforecast) precipitation. The BSS and reliability calculations were carried out for all grid points over land within the domain outlined in Fig. 1.

## **3. Analysis**

### *a. Brier skill score*

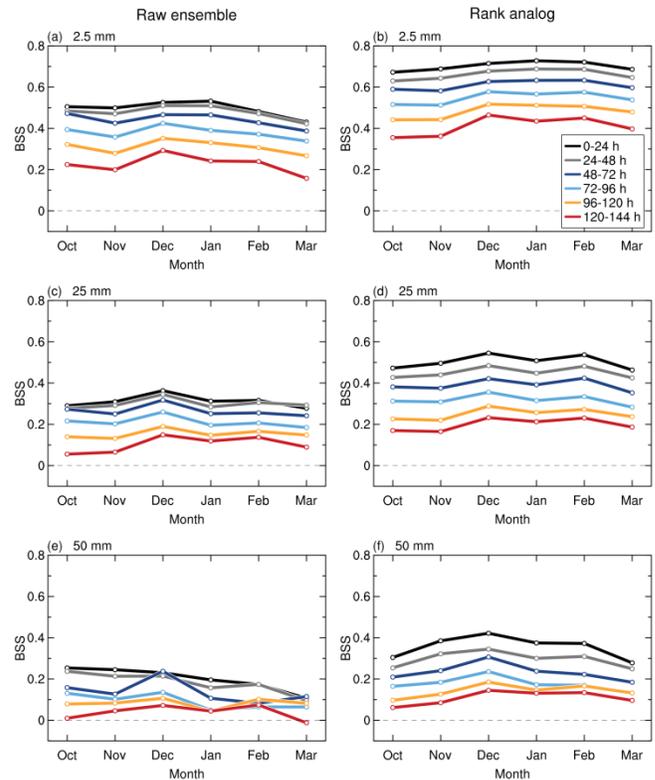
First, we compare the monthly BSS for PQPFs of 24-h accumulated precipitation from the raw ensemble



**Figure 1.** (a) Average annual precipitation accumulations (color shading in mm) for 1985–2011 from the NARR. (b) 2-arcminute terrain elevation (gray shading in m). The polygon in (a) and (b) outlines the domain (grid points over land only) for which verification statistics were calculated. *Click image for an external version; this applies to all figures hereafter.*

and the rank analog technique at 1–6-day lead times for the 2.5-, 25-, and 50-mm precipitation thresholds (Fig. 2). For both types of forecasts, skill generally decreased with increasing precipitation amount and with increasing lead time. For all months, lead times, and precipitation thresholds, the BSS values were considerably higher for the rank analog forecasts than for the raw ensemble forecasts, indicating that the rank analog technique yielded large improvements in forecast skill.

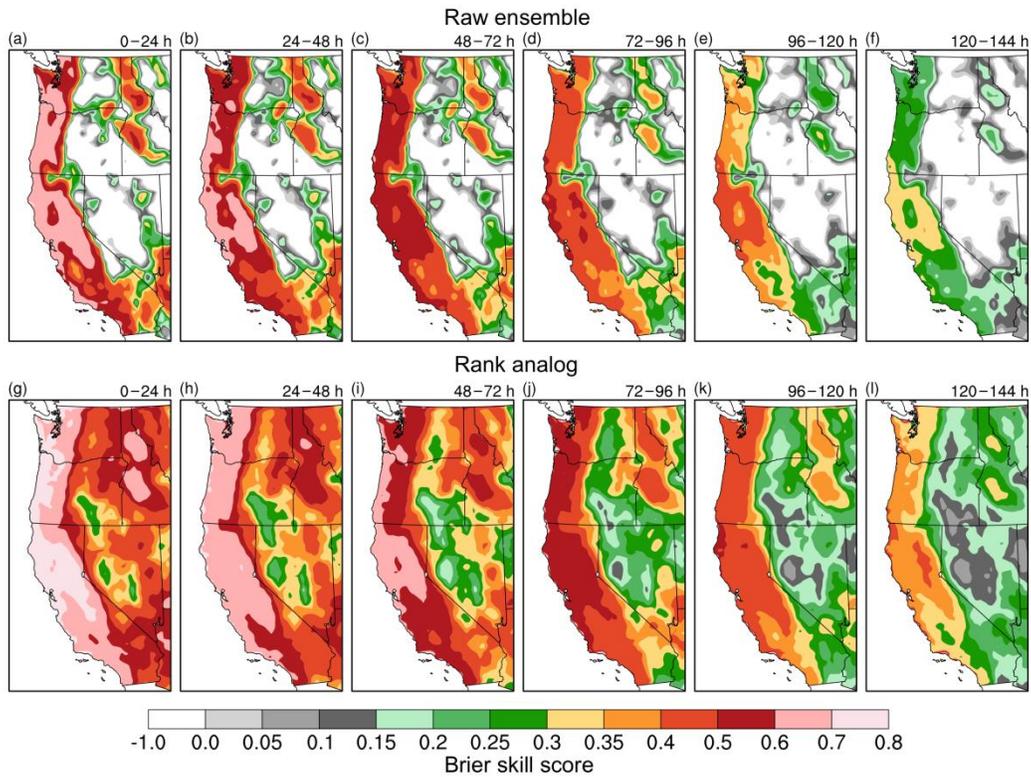
The geographic distribution of forecast skill is illustrated in maps of BSS for the 2.5-mm (Fig. 3) and 25-mm (Fig. 4) thresholds for raw ensemble and rank analog forecasts at 1–6-day lead times for all months during October–March. For both precipitation thresholds and at all lead times, the skill of the rank analog forecasts was greater than that of the raw ensemble forecasts at nearly all grid points. For the 25-mm threshold, there were areas, namely the climatologically dry Central Valley of California and the Mojave Desert (Fig. 1), where the raw ensemble was consistently unskillful ( $BSS < 0$ ) but where the rank analog technique consistently exhibited at least some skill (Fig. 4). For both types of forecasts, skill for the 25-mm threshold (Fig. 4) was substantially higher in climatologically wet mountainous regions relative to climatologically dry regions (Fig. 1); this likely relates to the fact that days with  $>25$  mm were rare or nonexistent in the dry regions (not shown).



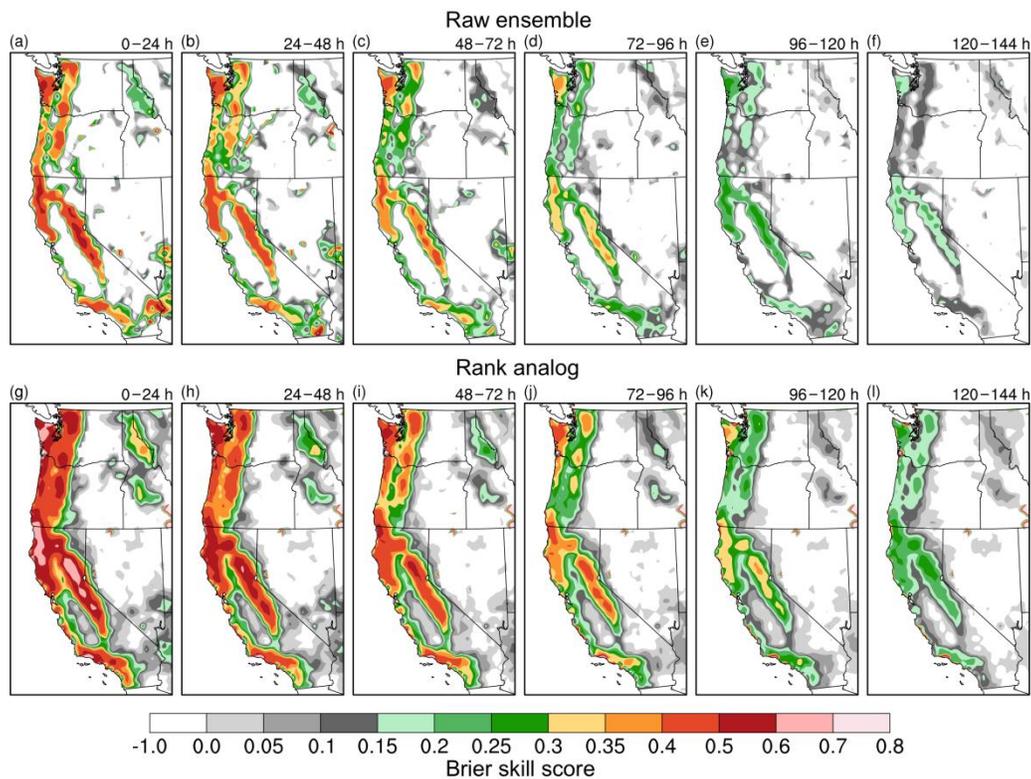
**Figure 2.** BSS as a function of month and forecast lead time for (left) the raw ensemble forecasts and (right) the rank analog forecasts of (a), (b)  $>2.5$  mm  $(24 \text{ h})^{-1}$ , (c), (d)  $>25$  mm  $(24 \text{ h})^{-1}$ , and (e), (f)  $>50$  mm  $(24 \text{ h})^{-1}$ . All grid points over land within the domain outlined in Fig. 1 were used for the BSS calculations.

### b. Reliability

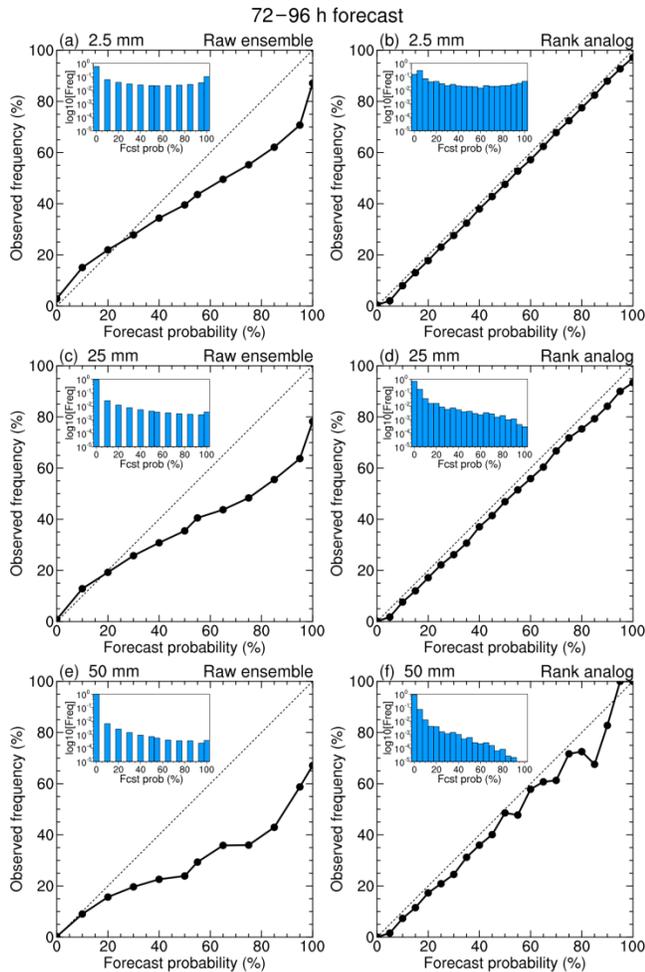
Reliability diagrams for 4-day PQPFs of 24-h accumulated precipitation from the raw ensemble and the rank analog technique for the 2.5-, 25-, and 50-mm precipitation thresholds are shown in Fig. 5. The reliability curves for the raw ensemble lie well below the diagonal for all precipitation thresholds (Figs. 5a,c,e), indicating an overforecasting (wet) bias and low reliability. Conversely, the reliability curves for the rank analog technique lie nearly along the diagonal for all thresholds (Figs. 5b,d,f), indicating substantial improvements in reliability over the raw ensemble. The raw ensemble produced more forecasts with high probabilities (i.e., greater sharpness) than the rank analog technique, particularly for the 25- and 50-mm thresholds (inset histograms in Figs. 5c–f). The reliability curves for the rank analog forecasts are situated just below the diagonal, indicating a slight overforecasting bias. For the 50-mm threshold (Fig. 5f), moderate deviations of the reliability curve for the rank analog technique from the diagonal are apparent at larger forecast probability bins, likely a consequence of the



**Figure 3.** Maps of BSS for forecasts of  $>2.5 \text{ mm (24 h)}^{-1}$  during October–March for (top) the raw ensemble and (bottom) the rank analog technique at (a), (g) 1-day, (b), (h) 2-day, (c), (i) 3-day, (d), (j) 4-day, (e), (k) 5-day, and (f), (l) 6-day lead times.



**Figure 4.** As in Fig. 3, except for forecasts of  $>25 \text{ mm (24 h)}^{-1}$ .

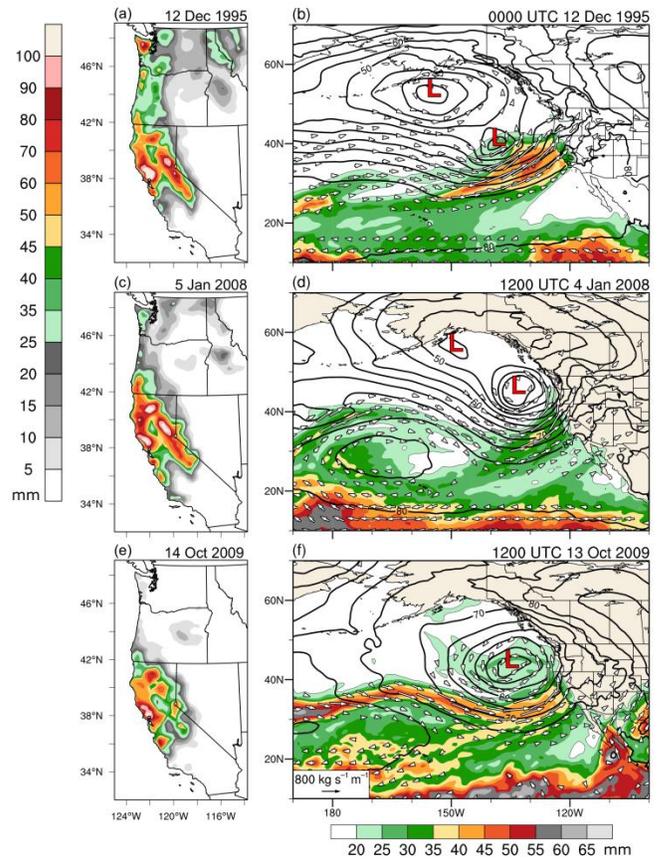


**Figure 5.** Reliability diagrams for 4-day (72–96-h) forecasts of (a), (b)  $>2.5 \text{ mm (24 h)}^{-1}$ , (c), (d)  $>25 \text{ mm (24 h)}^{-1}$ , and (e), (f)  $>50 \text{ mm (24 h)}^{-1}$  during October–March from (left) the raw ensemble and (right) the rank analog technique. The inset histograms show the frequency of usage for the probability bins (note the logarithmic scale on the vertical axis). All grid points over land within the domain outlined in Fig. 1 were used to construct these diagrams.

small sample sizes for those bins (inset histogram in Fig. 5f). Reliability diagrams for all other lead times between one and six days (not shown) exhibit generally similar characteristics to those in Fig. 5.

*c. Application of the rank analog technique for AR-related heavy precipitation events in California*

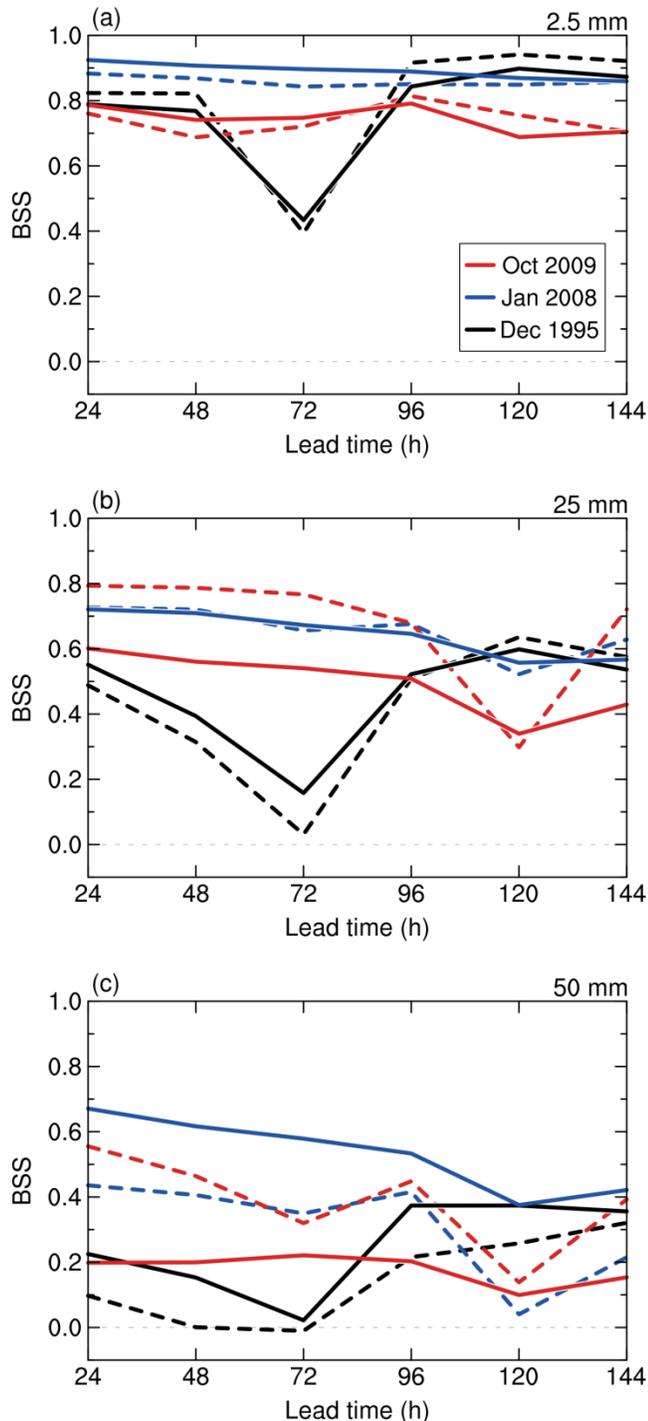
Three heavy precipitation events over northern California (Fig. 6), which occurred during 11–12 December 1995, 4–5 January 2008, and 13–14 October 2009, respectively, are examined to demonstrate the utility of rank analog PQPFs. For each of these events, a strong extratropical cyclone positioned off the West Coast established a corridor of enhanced water vapor



**Figure 6.** (left) Analyzed 24-h accumulated precipitation (shaded in mm according to the left color bar) from the NARR valid at 0000 UTC on (a) 12 December 1995, (c) 5 January 2008, and (e) 14 October 2009. (right) 925-hPa geopotential height (contoured in black every 5 dam; minima marked by “L” symbols), column-integrated water vapor (shaded in mm according to the lower color bar), and 1000–300-hPa vertically integrated water vapor transport vectors [ $\text{kg m}^{-1} \text{ s}^{-1}$ ; reference vector shown in panel (f)] from the 0.5° NCEP Climate Forecast System Reanalysis (Saha et al. 2010) at (b) 0000 UTC 12 December 1995, (d) 1200 UTC 4 January 2008, and (f) 1200 UTC 13 October 2009.

content and strong southwesterly water vapor transport—subjectively determined to be an AR—directed toward northern California (Figs. 6b,d,f). During these events, strong water vapor transport linked to the AR impinged upon northern California (Figs. 6b,d,f), resulting in widespread heavy precipitation (24-h accumulations  $>50 \text{ mm}$ ; maxima  $>100 \text{ mm}$ ) in the coastal mountains, the Mount Shasta region, and the northern Sierra Nevada (Figs. 6a,c,e).

Plots of the BSS for the 2.5-, 25-, and 50-mm thresholds at 1–6-day lead times are provided in Fig. 7 for the raw ensemble and the rank analog technique for the three cases. These plots indicate that, in some instances, the rank analog technique produced more



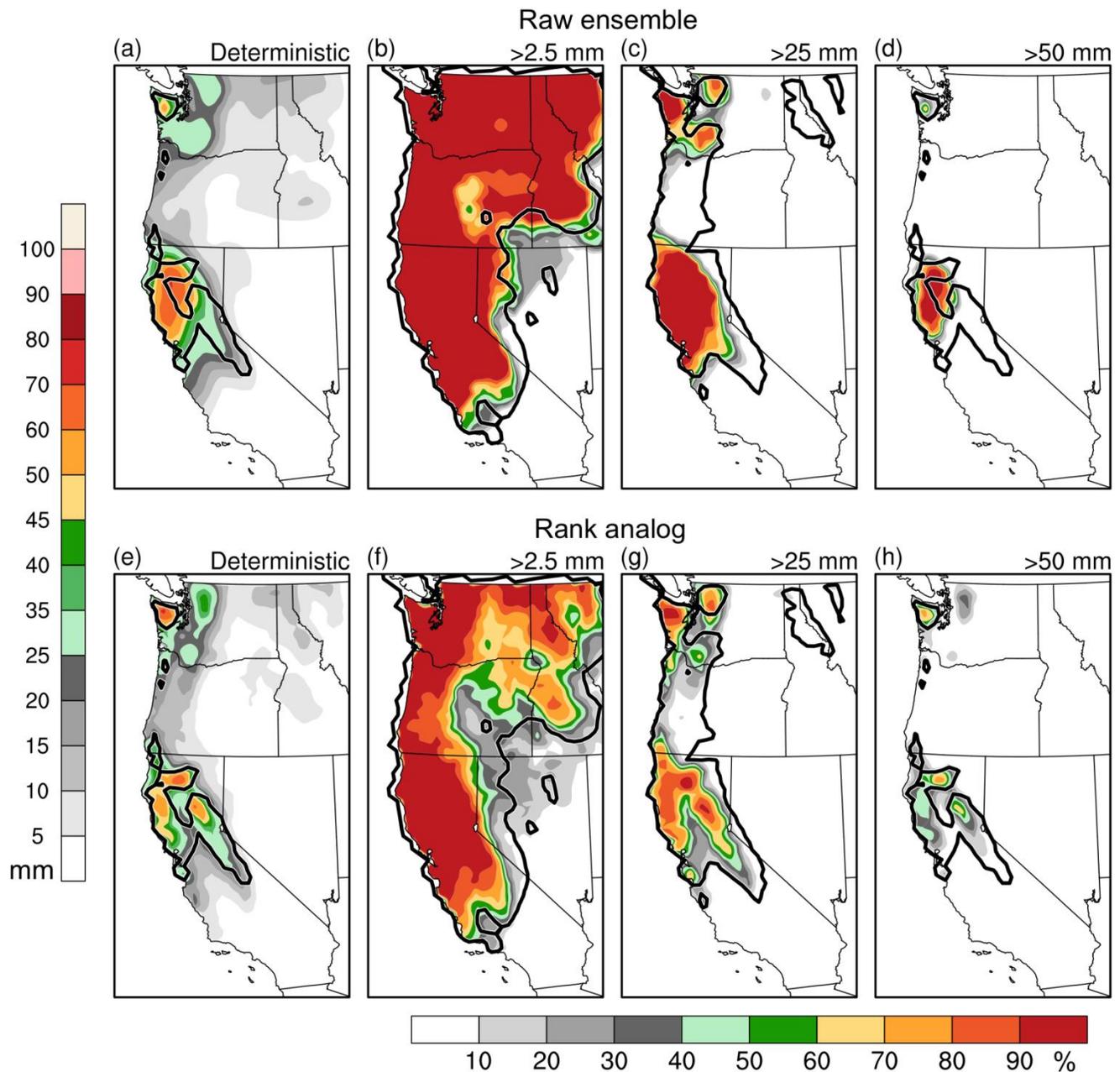
**Figure 7.** BSS as a function of forecast lead time for the raw ensemble forecasts (dashed lines) and the rank analog forecasts (solid lines) of (a)  $>2.5 \text{ mm } (24 \text{ h})^{-1}$ , (b)  $>25 \text{ mm } (24 \text{ h})^{-1}$ , and (c)  $>50 \text{ mm } (24 \text{ h})^{-1}$  valid at 0000 UTC 12 December 1995 (black), 0000 UTC 5 January 2008 (blue), and 0000 UTC 14 October 2009 (red). All grid points over land within the domain outlined in Fig. 1 were used for the BSS calculations.

skillful forecasts than the raw ensemble, while in other instances the rank analog technique produced less skillful forecasts. In general, the differences in skill between the two types of forecasts were most prominent for the 25- and 50-mm thresholds (Figs. 7b,c). In particular, the rank analog technique exhibited higher skill for the (i) 25-mm threshold at the 1–3-day lead time for the December 1995 event and (ii) 50-mm threshold at all lead times for both the December 1995 and January 2008 events—but exhibited lower skill for the 25- and 50-mm thresholds at all lead times for the October 2009 event.

Four-day (72–96-h) deterministic and probabilistic 24-h accumulated precipitation forecasts valid at 0000 UTC 12 December 1995, 0000 UTC 5 January 2008, and 0000 UTC 14 October 2009 for the three cases, respectively, are shown in Figs. 8–10 for the rank analog technique and the raw ensemble. In these forecasts, the spatial structure of the rank analog forecasts, relative to that of the raw ensemble forecasts, tended to more closely align with major orographic features (cf., Fig. 1b), with higher precipitation amounts and higher probabilities in mountainous regions relative to low-lying regions (Figs. 8e–h, 9e–h, and 10e–h). Moreover, the rank analog technique frequently resulted in decreases in forecast probability values and in precipitation amounts at locations with relatively high raw ensemble forecast probability values (Figs. 8b–d, 8f–h, 9b–d, 9f–h, 10b–d, and 10f–h) and precipitation amounts (Figs. 8a,e, 9a,e, and 10a,e), consistent with the tendency of the rank analog forecasts to be less sharp than the raw ensemble forecasts (Fig. 5).

For the December 1995 and January 2008 events, the raw ensemble forecasts featured areas of high ( $>70\%$ ) probability of  $>25 \text{ mm}$  and  $>50 \text{ mm}$  that overlapped with the areas in California where these thresholds were exceeded in the verifying analyses (Figs. 8c,d and 9c,d). However, these forecasts exhibited errors with respect to the spatial structure and placement of precipitation, most prominently for the 50-mm threshold. For the December 1995 case, the raw ensemble forecast failed to capture an axis of  $>50 \text{ mm}$  along the northern Sierra Nevada and erroneously indicated high probability of  $>50 \text{ mm}$  in the far northern Central Valley outside of the area enclosed by the analyzed 50-mm isohyet (Fig. 8d). For the January 2008 case, the raw ensemble forecast featured areas of high probability of  $>50 \text{ mm}$  along the northern coastal mountains and the Sierra Nevada that spread into the Central Valley outside of the area enclosed by the

Initialized: 0000 UTC 8 Dec 1995, valid: 0000 UTC 12 Dec 1995



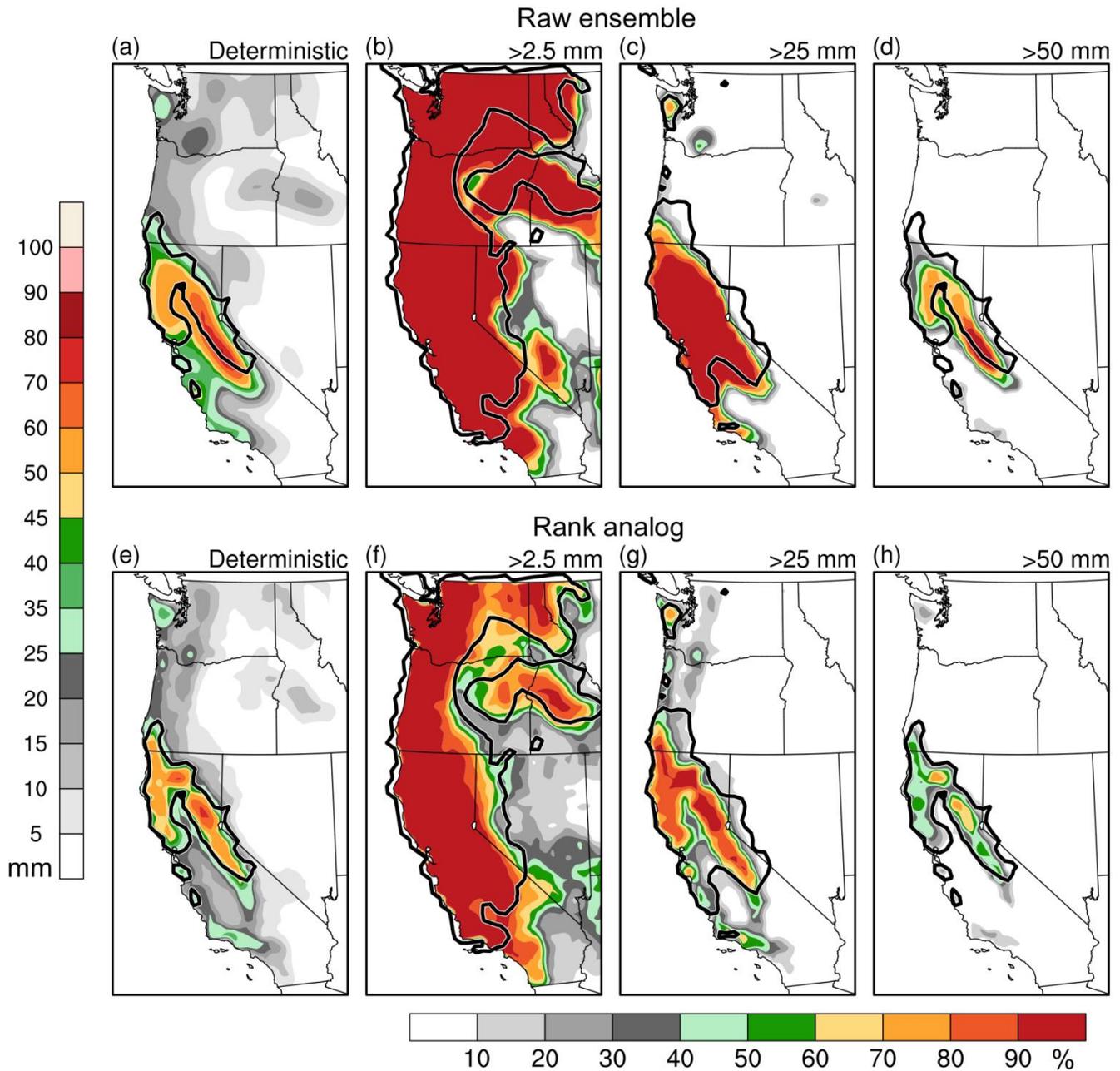
**Figure 8.** Four-day (72–96-h) 24-h accumulated precipitation forecasts from (top) the raw ensemble and (bottom) the rank analog technique initialized at 0000 UTC 8 December 1995 and valid at 0000 UTC 12 December 1995. Maps show (a), (e) deterministic forecasts (shaded in mm according to the left color bar) and probabilities (shaded in % according to the lower color bar) of (b), (f)  $>2.5 \text{ mm (24 h)}^{-1}$ , (c), (g)  $>25 \text{ mm (24 h)}^{-1}$ , and (d), (h)  $>50 \text{ mm (24 h)}^{-1}$ . The black contour in (b)–(d) and (f)–(h) denotes the NARR-analyzed isohyet for the threshold of interest. In (a), (e) the black contour denotes the NARR-analyzed 50-mm isohyet.

analyzed 50-mm isohyet (Fig. 9d). For both of these cases, the rank analog technique had the effect of partially correcting errors in the raw ensemble forecasts, yielding forecast precipitation amounts and probabilities that bore closer resemblance in spatial structure to the analyzed precipitation distributions

(Figs. 8 and 9), particularly for the 50-mm threshold. Accordingly, for the two cases, the rank analog technique exhibited higher skill than the raw ensemble for the 50-mm threshold at 4-day lead time (Fig. 7c).

For the October 2009 case, large precipitation amounts and high probabilities of  $>25 \text{ mm}$  and  $>50$

**Initialized: 0000 UTC 1 Jan 2008, valid: 0000 UTC 5 Jan 2008**



**Figure 9.** As in Fig. 8, except for forecasts initialized at 0000 UTC 1 January 2008 and valid at 0000 UTC 5 January 2008.

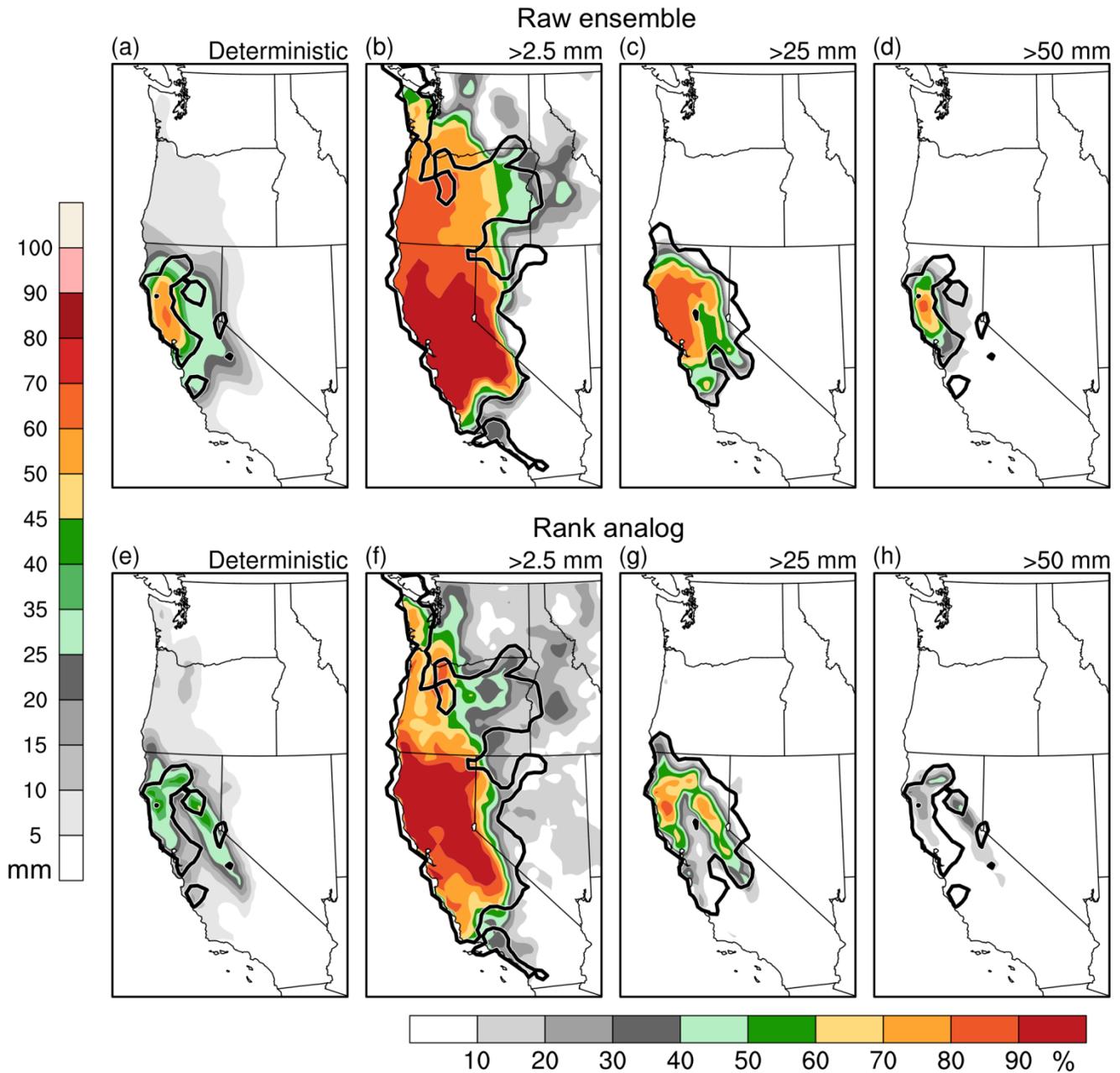
mm were forecast by the raw ensemble in approximately the correct location along the northern California coast (Figs. 10a,c,d). The rank analog technique yielded notable decreases in precipitation amounts and in probability values for the 25- and 50-mm thresholds along the northern California coast (Figs. 10e,g,h), corresponding to a decrease in forecast skill relative to the raw ensemble (Figs. 7b,c). This decrease in forecast skill suggests that there may have been few close

analogs of the October 2009 event in the reforecast dataset with similarly high precipitation amounts. As noted by Hamill and Whitaker (2006), this can sometimes be the case for extreme events.

**4. Summary and concluding remarks**

In this study, the NOAA second-generation GEFS reforecast dataset was used to generate PPFs for the

**Initialized: 0000 UTC 10 Oct 2009, valid: 0000 UTC 14 Oct 2009**



**Figure 10.** As in Fig. 8, except for forecasts initialized at 0000 UTC 10 October 2009 and valid at 0000 UTC 14 October 2009.

cool season (October–March) during 1985–2011 over the West Coast of the United States using a rank analog post-processing technique. For a given forecast, calibrated PQPFs were generated from an ensemble of NARR precipitation analyses corresponding to a set of close analog dates selected from the reforecast dataset. Forecast skill and reliability of PQPFs of 24-h accumulated precipitation from the raw reforecast ensemble and the rank analog technique were compared.

Consistent with the qualitative findings of ARRFEX, the results of this comparison indicated that the rank analog technique produced substantial improvements in forecast skill and reliability over the raw ensemble. Forecasts for three AR-related heavy precipitation cases were examined to demonstrate the utility of the rank analog technique. For two of these cases (11–12 December 1995 and 4–5 January 2008), the rank analog technique served to partially correct errors present

in the raw ensemble forecasts, yielding increases in forecast skill. Conversely, for the third case (13–14 October 2009) the rank analog technique resulted in decreases in forecast skill relative to the raw ensemble.

The results of this study indicate that, in aggregate, the rank analog technique applied to the GEFS reforecast dataset yields improved PQPFs compared to the raw ensemble guidance for the West Coast. Thus, forecasters may expect the rank analog technique to provide improved guidance in a majority of situations. However, as suggested by the results of the October 2009 case study, this technique may result in decreases in forecast quality for some cases, including some AR-related heavy precipitation cases. The overall skill and reliability of rank analog PQPFs specifically for AR-related heavy precipitation events remains unclear and deserves further study.

We close by noting that further research could be conducted to develop and test other reforecast post-processing techniques (e.g., logistic regression and distribution-fitting approaches) for generating improved PQPFs for the West Coast. Future research could assess whether the selection of analogs based upon other relevant variables, such as water vapor transport, column-integrated water vapor, or wind speed and direction, can further improve precipitation forecast skill. Last, the efficacy of high-resolution multi-sensor precipitation datasets—available for shorter time periods but at much higher resolution than the NARR—for generating calibrated PQPFs over the West Coast should be investigated.

*Acknowledgments.* This research was supported by NOAA's Hydrometeorological Testbed program, the United States Weather Research Program, and the Sonoma County Water Agency. We thank two anonymous reviewers whose comments helped to improve this manuscript.

## REFERENCES

- Bukovsky, M. S., and D. J. Karoly, 2007: A brief evaluation of precipitation from the North American Regional Reanalysis. *J. Hydrometeorol.*, **8**, 837–846, [CrossRef](#).
- Dettinger, M. D., 2013: Atmospheric rivers as drought busters on the U.S. West Coast. *J. Hydrometeorol.*, **14**, 1721–1732, [CrossRef](#).
- \_\_\_\_\_, F. M. Ralph, T. Das, P. J. Neiman, and D. R. Cayan, 2011: Atmospheric rivers, floods and the water resources of California. *Water*, **3**, 445–478, [CrossRef](#).
- Guan, B., N. P. Molotch, D. E. Waliser, E. J. Fetzer, and P. J. Neiman, 2013: The 2010/11 snow season in California's Sierra Nevada: Role of atmospheric rivers and modes of large-scale variability. *Water Resour. Res.*, **49**, 6731–6743, [CrossRef](#).
- Hamill, T. M., and J. S. Whitaker, 2006: Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory and application. *Mon. Wea. Rev.*, **134**, 3209–3229, [CrossRef](#).
- \_\_\_\_\_, \_\_\_\_\_, and S. L. Mullen, 2006: Reforecasts: An important dataset for improving weather predictions. *Bull. Amer. Meteor. Soc.*, **87**, 33–46, [CrossRef](#).
- \_\_\_\_\_, 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, [CrossRef](#).
- \_\_\_\_\_, M. Scheuerer, and G. T. Bates, 2015: Analog probabilistic precipitation forecasts using GEFS reforecasts and Climatology-Calibrated Precipitation Analyses. *Mon. Wea. Rev.*, **143**, 3300–3309, [CrossRef](#).
- 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, [CrossRef](#).
- Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343–360, [CrossRef](#).
- Neiman, P. J., F. M. Ralph, G. A. Wick, J. D. Lundquist, and M. D. Dettinger, 2008a: Meteorological characteristics and overland precipitation impacts of atmospheric rivers affecting the West Coast of North America based on eight years of SSM/I satellite observations. *J. Hydrometeorol.*, **9**, 22–47, [CrossRef](#).
- \_\_\_\_\_, \_\_\_\_\_, \_\_\_\_\_, Y.-H. Kuo, T.-K. Wee, Z. Ma, G. H. Taylor, and M. D. Dettinger, 2008b: Diagnosis of an intense atmospheric river impacting the Pacific Northwest: Storm summary and offshore vertical structure observed with COSMIC satellite retrievals. *Mon. Wea. Rev.*, **136**, 4398–4420, [CrossRef](#).
- \_\_\_\_\_, L. J. Schick, F. M. Ralph, M. Hughes, and G. A. Wick, 2011: Flooding in western Washington: The connection to atmospheric rivers. *J. Hydrometeorol.*, **12**, 1337–1358, [CrossRef](#).
- \_\_\_\_\_, F. M. Ralph, B. J. Moore, and R. J. Zamora, 2014: The regional influence of an intense Sierra barrier jet and landfalling atmospheric river on orographic precipitation in northern California: A case study. *J. Hydrometeorol.*, **15**, 1419–1439, [CrossRef](#).
- Newell, R. E., N. E. Newell, Y. Zhu, and C. Scott, 1992: Tropospheric rivers? – A pilot study. *Geophys. Res. Lett.*, **19**, 2401–2404, [CrossRef](#).
- Ralph, F. M., and M. D. Dettinger, 2012: Historical and national perspectives on extreme West Coast precipitation associated with atmospheric rivers during December 2010. *Bull. Amer. Meteor. Soc.*, **93**, 783–790, [CrossRef](#).

- \_\_\_\_\_, P. J. Neiman, and G. A. Wick, 2004: Satellite and CALJET aircraft observations of atmospheric rivers over the eastern North Pacific Ocean during the winter of 1997/98. *Mon. Wea. Rev.*, **132**, 1721–1745, [CrossRef](#).
- \_\_\_\_\_, \_\_\_\_\_, S. I. Gutman, M. D. Dettinger, D. R. Cayan, and A. B. White, 2006: Flooding on California's Russian River: Role of atmospheric rivers. *Geophys. Res. Lett.*, **33**, L13801, [CrossRef](#).
- \_\_\_\_\_, \_\_\_\_\_, G. N. Kiladis, K. Weickmann, D. W. Reynolds, 2011: A multiscale observational case study of a Pacific atmospheric river exhibiting tropical–extratropical connections and a mesoscale frontal wave. *Mon. Wea. Rev.*, **139**, 1169–1189, [CrossRef](#).
- Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1057, [CrossRef](#).
- Smith, B. L., S. E. Yuter, P. J. Neiman, and D. E. Kingsmill, 2010: Water vapor fluxes and orographic precipitation over northern California associated with a landfalling atmospheric river. *Mon. Wea. Rev.*, **138**, 74–100, [CrossRef](#).
- Sodemann, H., and A. Stohl, 2013: Moisture origin and meridional transport in atmospheric rivers and their association with multiple cyclones. *Mon. Wea. Rev.*, **141**, 2850–2868, [CrossRef](#).
- 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, [CrossRef](#).
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. Academic Press, 676 pp.
- Zhu, Y., and R. E. Newell, 1998: A proposed algorithm for moisture fluxes from atmospheric rivers. *Mon. Wea. Rev.*, **126**, 725–735, [CrossRef](#).