



Assessment of Extreme Quantitative Precipitation Forecasts and Development of Regional Extreme Event Thresholds Using Data from HMT-2006 and COOP Observers

F. M. RALPH,* E. SUKOVICH,*⁺ D. REYNOLDS,[#] M. DETTINGER,[@] S. WEAGLE,&
W. CLARK,*⁺ AND P. J. NEIMAN*

* NOAA/Earth System Research Laboratory, Boulder, Colorado

⁺ CIRES, University of Colorado, Boulder, Colorado

[#] NOAA/NWS/Monterey WFO, Monterey, California

[@] U.S. Geological Survey, Scripps Institution of Oceanography, La Jolla, California

& NOAA/NWS/Portland WFO, and Northwest River Forecast Center, Portland, Oregon

(Manuscript received 4 November 2009, in final form 19 April 2010)

ABSTRACT

Extreme precipitation events, and the quantitative precipitation forecasts (QPFs) associated with them, are examined. The study uses data from the Hydrometeorology Testbed (HMT), which conducted its first field study in California during the 2005/06 cool season. National Weather Service River Forecast Center (NWS RFC) gridded QPFs for 24-h periods at 24-h (day 1), 48-h (day 2), and 72-h (day 3) forecast lead times plus 24-h quantitative precipitation estimates (QPEs) from sites in California (CA) and Oregon–Washington (OR–WA) are used. During the 172-day period studied, some sites received more than 254 cm (100 in.) of precipitation. The winter season produced many extreme precipitation events, including 90 instances when a site received more than 7.6 cm (3.0 in.) of precipitation in 24 h (i.e., an “event”) and 17 events that exceeded 12.7 cm (24 h)^{−1} [5.0 in. (24 h)^{−1}]. For the 90 extreme events {>7.6 cm (24 h)^{−1} [3.0 in. (24 h)^{−1}]}, almost 90% of all the 270 QPFs (days 1–3) were biased low, increasingly so with greater lead time. Of the 17 observed events exceeding 12.7 cm (24 h)^{−1} [5.0 in. (24 h)^{−1}], only 1 of those events was predicted to be that extreme. Almost all of the extreme events correlated with the presence of atmospheric river conditions. Total seasonal QPF biases for all events {i.e., ≥0.025 cm (24 h)^{−1} [0.01 in. (24 h)^{−1}] were sensitive to local geography and were generally biased low in the California–Nevada River Forecast Center (CNRFC) region and high in the Northwest River Forecast Center (NWRFC) domain. The low bias in CA QPFs improved with shorter forecast lead time and worsened for extreme events. Differences were also noted between the CNRFC and NWRFC in terms of QPF and the frequency of extreme events. A key finding from this study is that there were more precipitation events >7.6 cm (24 h)^{−1} [3.0 in. (24 h)^{−1}] in CA than in OR–WA. Examination of 422 Cooperative Observer Program (COOP) sites in the NWRFC domain and 400 in the CNRFC domain found that the thresholds for the top 1% and top 0.1% of precipitation events were 7.6 cm (24 h)^{−1} [3.0 in. (24 h)^{−1}] and 14.2 cm (24 h)^{−1} [5.6 in. (24 h)^{−1}] or greater for the CNRFC and only 5.1 cm (24 h)^{−1} [2.0 in. (24 h)^{−1}] and 9.4 cm (24 h)^{−1} [3.7 in. (24 h)^{−1}] for the NWRFC, respectively. Similar analyses for all NWS RFCs showed that the threshold for the top 1% of events varies from ~3.8 cm (24 h)^{−1} [1.5 in. (24 h)^{−1}] in the Colorado Basin River Forecast Center (CBRFC) to ~5.1 cm (24 h)^{−1} [3.0 in. (24 h)^{−1}] in the northern tier of RFCs and ~7.6 cm (24 h)^{−1} [3.0 in. (24 h)^{−1}] in both the southern tier and the CNRFC. It is recommended that NWS QPF performance in the future be assessed for extreme events using these thresholds.

1. Introduction

One of the greatest challenges in meteorology is the prediction of precipitation, particularly the accurate

prediction of extreme precipitation events (i.e., events with large precipitation amounts). Recent surveys of public use of forecast information (Lazo et al. 2009) have documented that precipitation prediction (e.g., the location, timing, and amount of precipitation) is the most heavily utilized part of standard forecasts. This general public demand for precipitation forecasts is echoed by the needs of many specific forecast user communities, such as

Corresponding author address: F. M. Ralph, NOAA/ESRL/PSD, 325 Broadway, Boulder, CO 80305-3328.
E-mail: marty.ralph@noaa.gov

transportation, water resources and flood control, emergency management, and many others. These needs were highlighted in the U.S. Weather Research Program (USWRP)-sponsored community planning report (Ralph et al. 2005b), in which experts in hydrology, transportation, and emergency management described the linkages of quantitative precipitation forecasts (QPFs) to their specific community's needs. The major recommendation of this report was to establish a Hydrometeorology Testbed (HMT) to accelerate research and improvements in QPFs—particularly extreme QPFs—through better physical understanding, observations, numerical modeling, and decision support systems—all of which are critical elements of the operational forecast process for QPFs and are important for future climate services related to extreme events (e.g., National Weather Service 1999; Antolik 2000; Morss and Ralph 2007). This paper follows a long history of examining a particular period and region to document QPF performance, including the early paper by Bosart (1980) that documented errors in the operational forecast model of that era.

A key driver of HMT is the recognition that the current metric and precipitation threshold used to assess forecast skill are inadequate for many users, particularly for large precipitation events. Currently, the official National Oceanic and Atmospheric Administration (NOAA)'s National Weather Service (NWS) performance measure for QPFs is the threat score [also known as the Critical Success Index (CSI)] of 2.5 cm (1 in.) or greater in 24 h $\{\geq 2.5 \text{ cm (24 h)}^{-1} [1.0 \text{ in. (24 h)}^{-1}]; \text{Anthes 1983; Olson et al. 1995}\}$, where the threat score is a measure of the degree of coincidence between forecast and observed events for a specified accumulated precipitation threshold.

While NOAA's NWS has tracked this measure for more than 50 yr and the threat score does provide a foundation for gauging related forecast improvements (e.g., Reynolds 2003; Charba et al. 2003), the threat score metric does not adequately address the user's needs for accurate forecasts of extreme precipitation events that are the most critical to many key users. This limitation is related to the fact that the threat score is highly sensitive to correctly forecasted events and penalizes for missed and falsely predicted events. Thus, for extreme events, which tend to occur less frequently and over smaller areas than weaker precipitation events, the threat score tends to zero. In addition, the threat score does not distinguish the source of forecast error, making it difficult to identify causes of missed and/or falsely predicted events.

A leading example of the demand to address extreme precipitation events comes from the hydrology–reservoir operations, flood control, and emergency management communities. Currently, all 13 of NOAA's NWS River Forecast Centers (RFCs) in the United States ingest

precipitation (both observed and predicted) in their hydrologic models to predict streamflow and runoff in their domains of responsibility. On the basis of these inputs (and others), the RFCs' hydrologic models calculate the total volume runoff and timing over a basin area, such that streamflow at certain points along the river can be predicted (Fread et al. 1995). Therefore, the QPFs directly affect hydrologic–reservoir operations, flood control, and emergency management communities. This is clearly illustrated in the area around Sacramento, California, where a major reservoir and levee system protect the city from potentially catastrophic urban flood inundation (National Research Council 1995, 1999). This risk is analogous to that faced by New Orleans, Louisiana, prior to the Hurricane Katrina disaster in 2005, although the Sacramento meteorology, hydrology, and infrastructure differ substantially. Specifically, extreme precipitation events in the Sacramento region consist of land-falling extratropical cyclones with embedded atmospheric rivers (ARs) that concentrate water vapor transport into California's Sierra Nevada, creating major orographic precipitation (e.g., Pandey et al. 1999; Neiman et al. 2002; Dettinger et al. 2004; Ralph et al. 2005a, 2006). When an extratropical storm has strong water vapor transport and stalls, it can produce more than 76.2 cm (30.0 in.) of precipitation in 3 days (Neiman et al. 2008a). While Folsom Dam, just upstream of Sacramento, provides significant flood protection, five storms since the dam was built in 1955 have exceeded its flood design, and a major winter storm in 1997 threatened Sacramento with severe flooding (National Research Council 1999). These risks, and the recent Katrina disaster, have led to a major focus on related flood risks in California and have motivated the passage of major bond issues to help mitigate those risks. While most of the investment is in levees and dam improvements, an innovative approach of improving QPFs, such that reservoir water can be released before the landfall of major winter storms, is also under development. A crucial step in the flood-control process involves the reservoir operator at Folsom Dam (U.S. Bureau of Reclamation) making decisions about how much water to release from the dam. A major factor compounding the risks is the potential that future climate conditions will reduce snowpack and move the peak runoff season into the earlier part of the spring season, when flooding events are more likely to occur than they are today.

Since 2005, HMT has focused its efforts in the western United States to address flood risks. Findings from HMT will be implemented by the state of California's Department of Water Resources (DWR) through the Enhanced Flood Response and Emergency Preparedness (EFREP) program, a joint effort between the DWR,

NOAA, and the Scripps Institution of Oceanography. An important step is to evaluate forecast performance for extreme precipitation events.

The key purpose of this article is to explore QPF verification for extreme precipitation events using information from the very wet winter season studied in HMT-2006 and from Cooperative Observer Program (COOP) daily precipitation observations. This paper analyzes 2005/06 cool-season data from the California–Nevada River Forecast Center (CNRFC) and the Northwest River Forecast Center (NWRFC) for the HMT-California and Oregon–Washington (OR–WA) regions, respectively, to assess QPF performance of extreme precipitation events. Using not only the threat score but also the probability of detection (POD), false-alarm rate (FAR), mean absolute error (MAE), and bias, QPF performance is analyzed for these regions. Results of this QPF assessment indicate that extreme events vary in magnitude by region; thus, this paper expands upon these results to identify extreme daily precipitation thresholds that correspond to the same frequency of occurrence in each NWS RFC nationally based on ~30 yr of rain gauge statistics. On the basis of the importance of extreme precipitation, the challenge of predicting it accurately, and the regional variations of what defines an “extreme” precipitation event, this paper recommends that unique regional precipitation thresholds be used in the future to assess extreme precipitation QPF performance. These regionally relevant QPF performance statistics can then be combined to evaluate the forecast performance of extreme precipitation events nationally.

2. Data and methodology overview

Data from the November 2005 to April 2006 cool season were analyzed for 41 sites along the West Coast. Using QPF and quantitative precipitation estimate (QPE) data provided by the NWS CNRFC and NWRFC, 41 sites spanning a study area from the Washington–Canada border through the Sacramento, California, region were examined (Fig. 1; Table 1). The 2005/06 cool season was selected mainly because of significant precipitation events that were observed in the CNRFC region and the generally wet cool season on average. Overall, 17 sites were selected and analyzed in the CNRFC region and 24 sites were studied in the NWRFC region. All precipitation amounts (observed and predicted) examined in this paper include English units in parentheses, since precipitation forecasts in the United States are commonly issued in units of inches.

a. QPF products

The NWS RFC QPF products verified in this study are the result of a multistep process. Initially, the Hydrometeorological Prediction Center (HPC) issues

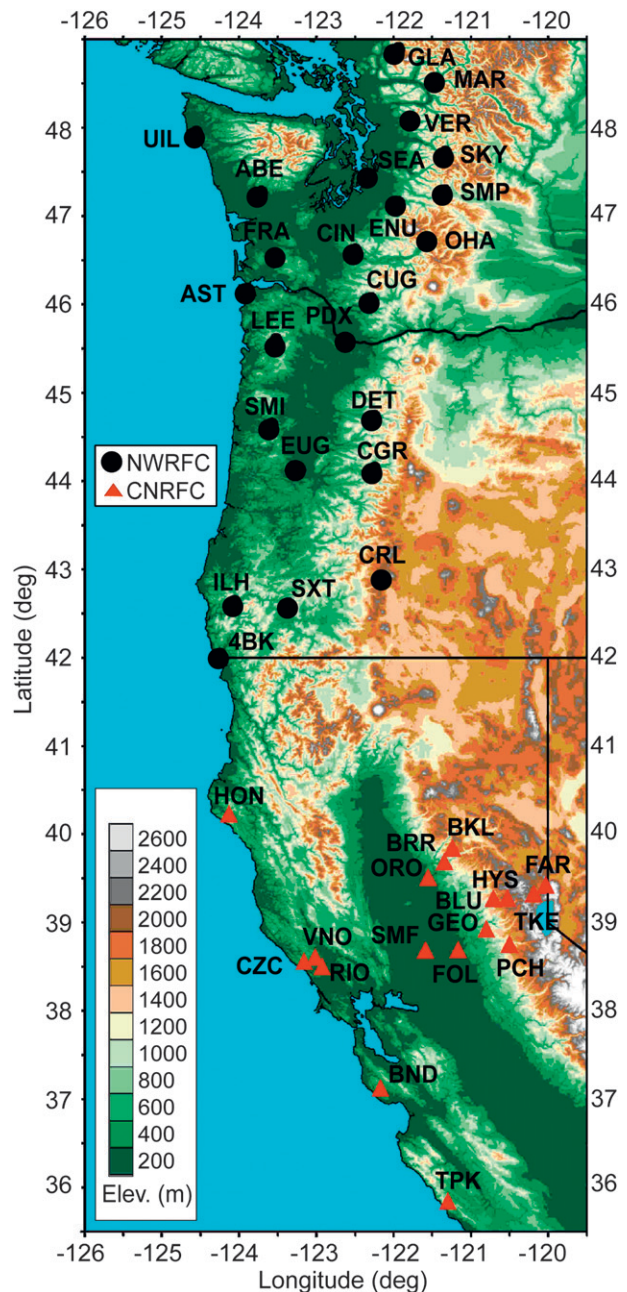


FIG. 1. Terrain base map of the West Coast with the 17 selected CNRFC verification sites indicated by red triangles and the 24 selected NWRFC verification sites indicated by black circles. Sites are identified by three letters (see Table 1 for abbreviations).

national QPFs of the total amount of liquid precipitation expected during a specified period for given locations. Isohyets of expected precipitation amounts of 0.025, 0.64, 1.3, 2.5, and 3.8 cm (0.01, 0.25, 0.50, 1, and 1.50 in., respectively) and greater are drawn for three consecutive (days 1–3) 24-h intervals and then gridded at 32-km resolution. These forecasts are issued twice daily and are

TABLE 1. The 41 RFC verification sites.

RFC	Site name	ID	Altitude (m)	Latitude (°)	Longitude (°)	Total precipitation (cm)	Total precipitation (in.)	Geographic region
CN	Ben Lomand, CA	BND	111	37.1	122.2	172.3	67.8	Coastal valley
CN	Rio Nido, CA	RIO	30	38.5	122.9	147.8	58.2	Coastal valley
CN	Cazadero, CA	CZC	475	38.6	123.2	187.2	73.7	Coastal mountains
CN	Honeydew, CA	HON	110	40.2	124.1	304.9	120.1	Coastal mountains
CN	Three Peaks, CA	TPK	1021	35.9	121.3	61.1	24.1	Coastal mountains
CN	Venado, CA	VNO	384	38.6	123.0	197.5	77.8	Coastal mountains
CN	Folsom, CA	FOL	129	38.7	121.2	63.2	24.9	Central Valley
CN	Oroville, CA	ORO	274	39.5	121.6	113.4	44.7	Central Valley
CN	Sacramento, CA	SMF	10	38.7	121.6	48.2	19.0	Central Valley
CN	Brush Creek, CA	BRR	1085	39.7	121.4	265.0	104.3	Sierra foothills
CN	Georgetown, CA	GEO	991	38.9	120.8	181.4	71.4	Sierra foothills
CN	Pacific House, CA	PCH	1049	38.8	120.5	175.3	69.0	Sierra foothills
CN	Bucks Lake, CA	BKL	1665	39.9	121.2	266.4	104.9	Sierra Mountains
CN	Blue Canyon, CA	BLU	1609	39.3	120.7	247.1	97.3	Sierra Mountains
CN	Huysink, CA	HYS	2073	39.3	120.5	229.6	90.4	Sierra Mountains
CN	Farad, CA	FAR	1571	39.4	120.0	72.7	28.6	Sierra lee
CN	Truckee, CA	TKE	1835	39.3	120.2	99.1	39.0	Sierra lee
NW	Brookings, OR	4BK	14	42.0	124.3	197.4	77.7	Coast
NW	Astoria, OR	AST	3	46.2	123.9	148.5	58.5	Coast
NW	Quillayute, WA	UIL	59	47.9	124.6	169.3	66.7	Coast
NW	Frances, WA	FRA	70	46.5	123.5	173.7	68.4	Coastal mountains
NW	Agness, OR	ILH	106	42.6	124.1	229.4	90.3	Coastal mountains
NW	Lees Camp, OR	LEE	200	45.6	123.5	247.4	97.4	Coastal mountains
NW	Summit, OR	SMI	227	44.6	123.6	142.6	56.1	Coastal mountains
NW	Aberdeen, WA	ABE	133	47.3	123.7	199.2	78.4	Inland valley
NW	Eugene, OR	EUG	108	44.1	123.2	94.8	37.3	Inland valley
NW	Portland, OR	PDX	6	45.6	122.6	73.8	29.1	Inland valley
NW	Seattle, WA	SEA	137	47.4	122.3	73.3	28.8	Inland valley
NW	Cougar Dam, OR	CGR	377	44.1	122.3	162.3	63.9	Cascade foothills
NW	Cinebar, WA	CIN	317	46.6	122.5	124.8	49.1	Cascade foothills
NW	Cougar, WA	CUG	165	46.1	122.3	175.5	69.1	Cascade foothills
NW	Detroit Dam, OR	DET	479	44.7	122.3	189.0	74.4	Cascade foothills
NW	Enumclaw, WA	ENU	399	47.1	121.9	102.4	40.3	Cascade foothills
NW	Glacier, WA	GLA	285	48.9	121.9	103.6	40.8	Cascade foothills
NW	Marblemount, WA	MAR	106	48.5	121.5	135.8	53.5	Cascade foothills
NW	Packwood, WA	OHA	594	46.7	121.6	147.8	58.2	Cascade foothills
NW	Skykomish, WA	SKY	314	47.7	121.3	143.9	56.7	Cascade foothills
NW	Verlot, WA	VER	297	48.1	121.8	208.5	82.1	Cascade foothills
NW	Crater Lake, OR	CRL	1974	42.9	122.1	161.8	63.7	Cascades
NW	Stampede Pass, WA	SMP	1207	47.3	121.3	136.6	53.8	Cascades
NW	Sexton Summit, OR	SXT	1168	42.6	123.4	125.8	49.5	Cascades

based upon current surface observations and upper-air analyses, radar data, satellite estimates, and guidance from the NWS Global Forecast System (GFS), North American Mesoscale model (NAM), European Centre for Medium-Range Weather Forecasts (ECMWF), Met Office (UKMO), and the Meteorological Service of Canada.

The HPC gridded-point QPF values are forwarded to the NWS RFCs, where the Hydrometeorological Analysis and Support (HAS) meteorologist reviews the RFC regional gridded-point values for a specified number of sites throughout the RFC region and modifies these points when appropriate to the local terrain, regional climate,

and sensitivity of hydrologic models to precipitation. The adjusted QPF point values are then input into the Parameter-elevation Regressions on Independent Slopes Model (PRISM; Daly et al. 1994)-based Mountain Mapper software to interpolate the QPF data points to a 4-km grid. PRISM is a hybrid statistical-geographical approach to mapping climate data between data points. Between gauge locations, a double linear interpolation technique is applied to a triangulated irregular network (TIN) to vary the bias for each grid (Daly et al. 1994). The result is a unique bias for each 4 km × 4 km grid. Using these products, it is possible to capture localized precipitation patterns that are often missed in the coarser national product grids.

Finally, the RFC issues a routine QPF product each day in Standard Hydrometeorological Exchange Format (SHEF) designed for ingestion into the RFC's hydrologic models. At the CNRFC, these 4-km gridded QPF data are converted to basin mean areal precipitation (MAP) and serve as input to their NWS River Forecast System (RFS) hydrologic models. In contrast, at the NWRFC, the QPF point data from the 4-km grid are themselves input to their NWS RFS models. In this study, the 6- to 72-h intervals of QPF data are all based on the 1200 UTC forecast. The 6-h QPF and QPE precipitation values were then accumulated to create 24-h daily totals starting at 1200 UTC each day. The resulting precipitation amounts are identified by their end times, which for the QPF values occur 24, 48, or 72 h (day 1, day 2, and day 3, respectively) after 1200 UTC of the day the forecast was made.

b. QPE verification data

Six-hourly QPE data used to verify the QPF products were obtained from the CNRFC and NWRFC, as each RFC performs daily quality control of gauge data throughout their respective hydrologic service areas. These estimates are based on precipitation amounts observed at hundreds of stations throughout each domain. When gauge data are missing from one or more of these sites because of various potential sources of error (e.g., technical difficulties, undercatch due to strong winds, and overrepresentation of precipitation due to snow melting into a heated tipping-bucket rain gauge), the QPE value is estimated using PRISM-based Mountain Mapper software to interpolate to the missing sites using precipitation amounts observed at nearby quality-controlled sites. Similar to the QPF process, Mountain Mapper and PRISM provide the capability to use climatological relationships between sites to derive a more accurate estimate of precipitation for missing sites in complex terrain. Rain gauge data determined to be "good" and estimated data for missing sites are then compiled and combined into the final 4 km \times 4-km gridded QPE product, with preference given toward quality-controlled site data over Mountain Mapper estimates when discrepancies occur.

c. COOP dataset

Daily accumulated precipitation totals were obtained from the historical Summary of the Day (SOD; National Weather Service 1989 and updates thereto) observations from cooperative weather stations across the United States, obtained from the National Climatic Data Center (NCDC). Daily data from 1950 to 2007 were considered from 6020 stations. Missing data were excluded, as were accumulations from multiple days that were reported as a single total. All records from stations within the

boundaries of each RFC region were analyzed together to assess large-area exceedance frequencies for daily precipitation. It should be noted that the measurement periods are not uniform from one site to the next. Some sites are several decades long, while others are shorter. Also, because the COOP data are point measurements and the time of day they represent varies from site to site, the COOP data are not used here for direct verification of gridded QPF.

d. QPF performance measures

To assess extreme event QPFs, the POD, FAR, CSI, MAE, and bias were analyzed for precipitation events during the 2005/06 cool season. In this study, an event is defined as a 24-h period when the predicted and/or observed accumulated precipitation at a verification site matches or exceeds a specified precipitation threshold. Thus, if both the QPF and observed precipitation are greater than the precipitation threshold, then a "hit" occurs. If the QPF is less than and the observed precipitation is greater than the threshold, then a "miss" occurs, and if the QPF exceeds the specified precipitation threshold but the observed precipitation is below the threshold, then a "false alarm" occurs. Within this framework, the POD represents the fraction of observed events that were correctly predicted (hits) to the total number of events that were observed (hits and misses), while FAR is the fraction of events predicted but did not occur (false alarms) to the total number of events predicted. The CSI, also referred to as the threat score, is the fraction of predicted events that occurred (hits) to all possible events (hits, misses, and false alarms). All three metrics range from 0 to 1, with 1 being perfect POD and CSI scores, and 0 being a perfect FAR score. To better determine the accuracy of QPFs (i.e., how close forecasts or predictions are to the eventual outcomes), the MAE, which is the average of the absolute errors, is calculated. The MAE is computed conditionally for QPE values greater than a specified threshold. To better understand how the QPFs overforecast or underforecast precipitation, the bias is calculated from the ratio of the predicted events to observed events, either in frequency of occurrence or precipitation amount.

3. Analysis of the 2005/06 cool season

a. Frequency of observed precipitation events

Over the 172 days between 5 November 2005 and 25 April 2006, the observed 24-h precipitation amounts from the 17 sites in the CNRFC region totaled 2832.4 cm (1115.1 in.) [or 166.6 cm (65.6 in.) on average per site], and the total 24-h observed precipitation amount for the

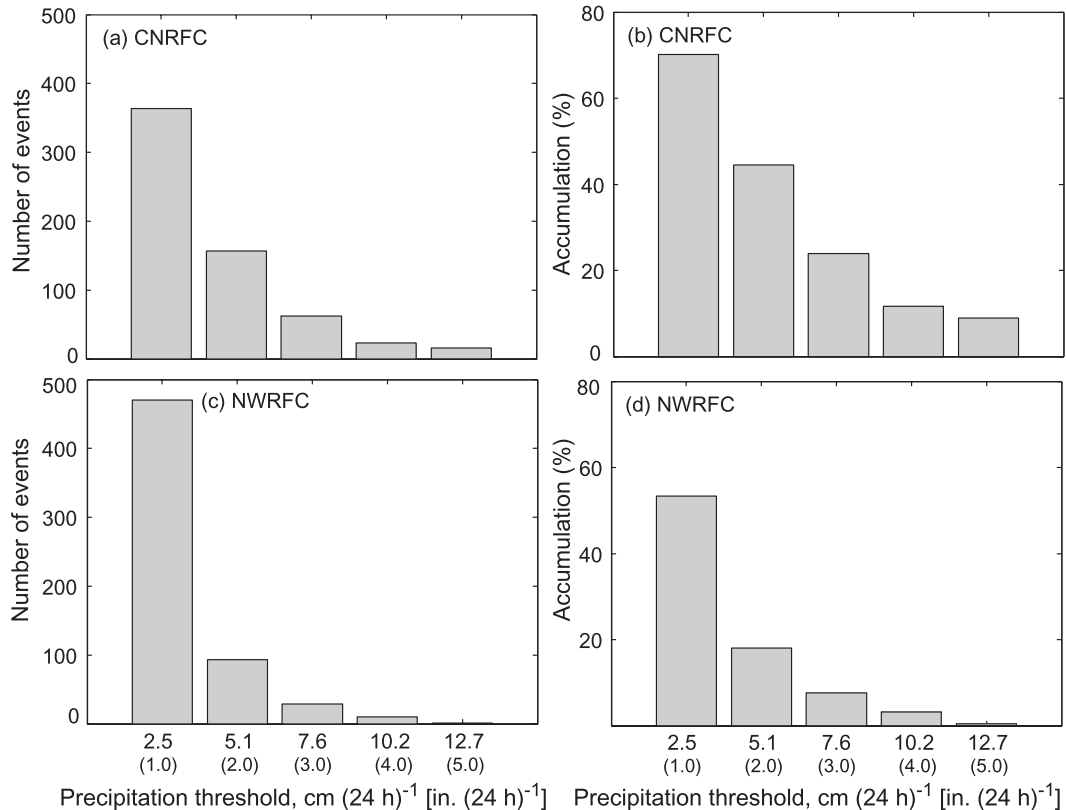


FIG. 2. Number of 24-h precipitation events exceeding 2.5, 5.1, 7.6, 10.2, and 12.7 cm (1.0, 2.0, 3.0, 4.0, and 5.0 in., respectively) between 5 Nov 2005 and 26 Apr 2006 for the (a) 17 CNRFC and (c) 24 NWRFC verification sites. Percent of total seasonal precipitation for (b) 17 CNRFC and (d) 24 NWRFC verification sites, exceeding 24-h precipitation amounts of 2.5, 5.1, 7.6, 10.2, and 12.7 cm (1.0, 2.0, 3.0, 4.0, and 5.0 in., respectively).

24 sites in the NWRFC was 3667.0 cm (1443.7 in.) [or 152.7 cm (60.1 in.) per site]. Table 1 lists each verification site in the CNRFC and NWRFC regions and the total accumulated precipitation at each site for this period. Throughout the CNRFC region, there were 1140 site days wherein ≥ 0.25 cm (24 h) $^{-1}$ [0.1 in. (24 h) $^{-1}$] accumulated precipitation was observed. Figure 2a shows the number of precipitation events exceeding specified precipitation thresholds {e.g., 62 events exceeded 7.6 cm (24 h) $^{-1}$ [3.0 in. (24 h) $^{-1}$] and 16 events exceeded 12.7 cm (24 h) $^{-1}$ [5.0 in. (24 h) $^{-1}$]} in the CNRFC region. Similarly, Fig. 2b shows the cumulative fraction of total precipitation observed at the 17 CNRFC sites as a function of daily total rainfall {e.g., roughly 50% of the total precipitation for the season occurred in events that produced >5.1 cm (24 h) $^{-1}$ [2 in. (24 h) $^{-1}$]}]. In comparison, the NWRFC region had 2130 site days with ≥ 0.25 cm (24 h) $^{-1}$ [0.1 in. (24 h) $^{-1}$] observed precipitation (Fig. 2c); however, most of the precipitation accumulated over the 2005/06 season occurred in the lowest precipitation thresholds (Fig. 2d). Overall, the CNRFC sites were typically wetter than the NWRFC sites during

the 2005/06 cool season, with the per-site accumulation larger for all precipitation thresholds and with the NWRFC having smaller precipitation events and fewer extreme precipitation events. For example, in the CNRFC domain, events with >7.6 cm (24 h) $^{-1}$ [3.0 in. (24 h) $^{-1}$] precipitation accounted for roughly 23% of the CNRFC domain precipitation total, whereas for the NWRFC >7.6 cm (24 h) $^{-1}$ [3.0 in. (24 h) $^{-1}$] events accounted for roughly 7% of the NWRFC domain precipitation total.

b. QPF performance

To assess the performance of the QPFs for day 1, day 2, and day 3 forecast lead times, in terms of total seasonal accumulation bias, the total forecasted precipitation and the total observed precipitation at each site in the CNRFC was calculated and plotted in Fig. 3a. With the exception of the site with approximately 177.8 cm (70 in.) total precipitation observed, the total day 1 QPFs at each site generally aligned with the observed total precipitation amounts over the entire cool season. The total calculated bias (predicted precipitation/observed accumulated

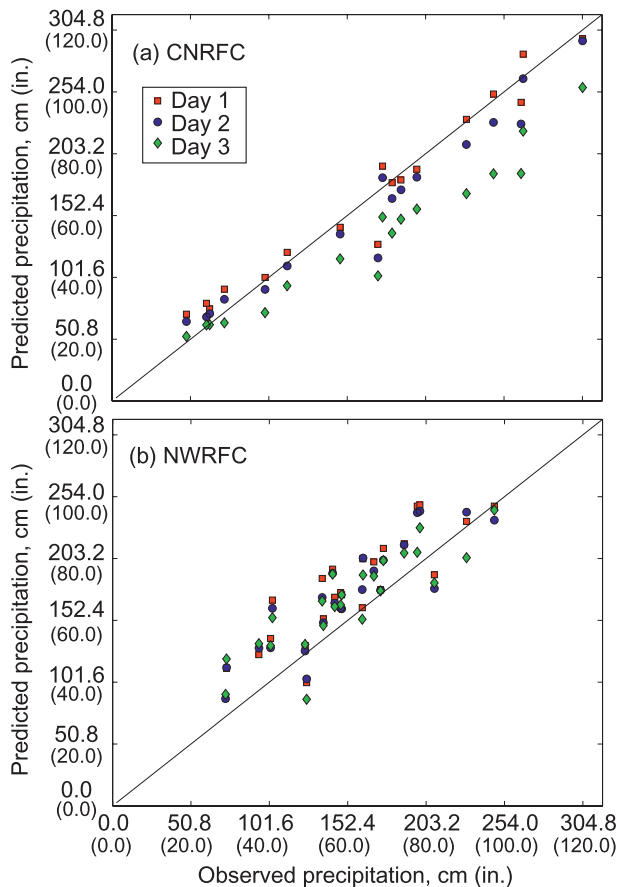


FIG. 3. Scatterplot of observed vs predicted total precipitation accumulations in cm (in.) between 5 Nov 2005 and 26 Apr 2006 for (a) 17 sites in the CNRFC region and (b) 24 sites in the NWRFC region. The QPFs are for day 1 (red squares), day 2 (blue circles), and day 3 (green diamonds).

precipitation) for all the sites of 1.01 reflects this alignment. However, as the forecast lead time increased, the total calculated biases for days 2 and 3 over the season decreased to 0.95 and 0.79, respectively. In other words, for the CNRFC day 1 QPF, the overall bias for seasonal accumulation was nearly perfect, while for day 2 and day 3 the QPFs were progressively underforecast. In contrast, the NWRFC site seasonal biases were generally positive, indicating a trend toward overprediction, at almost all sites (Fig. 3b). Overall, for the 24 sites in the NWRFC region, the seasonal biases were 1.16 (day 1), 1.14 (day 2), and 1.10 (day 3), indicating that the NWRFC QPFs overforecast for all three forecast intervals, with the day 1 QPF most overpredicted.

It is useful to test if these seasonal biases vary with respect to the topography and between the two RFC regions. Seasonal accumulated precipitation amounts varied from less than 63.5 cm (25 in.) measured at SMF (refer to

Table 1 for abbreviations), in an inland valley, to greater than 304.8 cm (120 in.) measured at HON (Fig. 4a), a coastal mountain site. Of the 17 CNRFC sites, 7 sites (RIO, BND, CZC, HON, VNO, GEO, and PCH) had all three forecast days underpredicting the total precipitation accumulation over the season. These represent the coastal mountains and valleys, and 2 of the 3 Sierra foothills sites. Meanwhile, only three sites (TPK, SMF, FOL), two of which were located in the Central Valley, had all three forecast days overpredicting the total accumulation. It should be noted that the TPK site is most likely anomalous from the other coastal sites because of its location farther south and for the tendency for northerly storm tracks during the 2005/06 winter season, which brought much more precipitation to northern California and less to central and Southern California.

The CNRFC overpredicted precipitation in both the Central Valley and the Sierra lee for day 1 QPFs and underpredicted the coastal valley precipitation for all forecast lead times. The Central Valley sites had mean biases of 1.20 and 1.11 for days 1 and 2 QPFs, respectively. The Sierra lee sites had a mean bias of 1.13 for day 1. The coastal valley had mean biases of 0.85, 0.79, and 0.69 for days 1–3, respectively.

In the NWRFC region, the sites were more challenging to categorize by geographic region because of the more complex topography of the Pacific Northwest. However, while the NWRFC QPFs tended to overpredict for all sites, it particularly overpredicted for the interior valley sites (Fig. 4b). The biases for these sites range from 1.35 to 1.43 for days 1–3 QPFs. This overprediction by the NWRFC for the inland valley sites is similar to that observed for QPFs in the Central Valley by the CNRFC (Fig. 4a). The trend of both the CNRFC and NWRFC to overforecast in the valleys and mountain lee regions reflects the influences of the topography on QPFs. The orographic nature of the precipitation strongly influences the QPF errors by providing a non-time-varying forcing, which allows QPFs in mountainous areas to have greater skill than in other regions (e.g., Central Valley and inland valley) without such strong forcing.

To better illustrate the range and daily variation of forecasts by site over the 2005/06 season, Fig. 5 shows box plots (Hoaglin et al. 1983) of the difference between the QPFs and QPEs for verification sites representative of key geographic regions in the CNRFC and NWRFC regions for the day 1 lead time. To highlight moderate to extreme events, only dates when the QPE or QPF exceeded $2.5 \text{ cm (24 h)}^{-1}$ [$1.0 \text{ in. (24 h)}^{-1}$] at a given site are included. Figure 5 shows the following: the range of daily errors is relatively large at some sites (e.g., BND

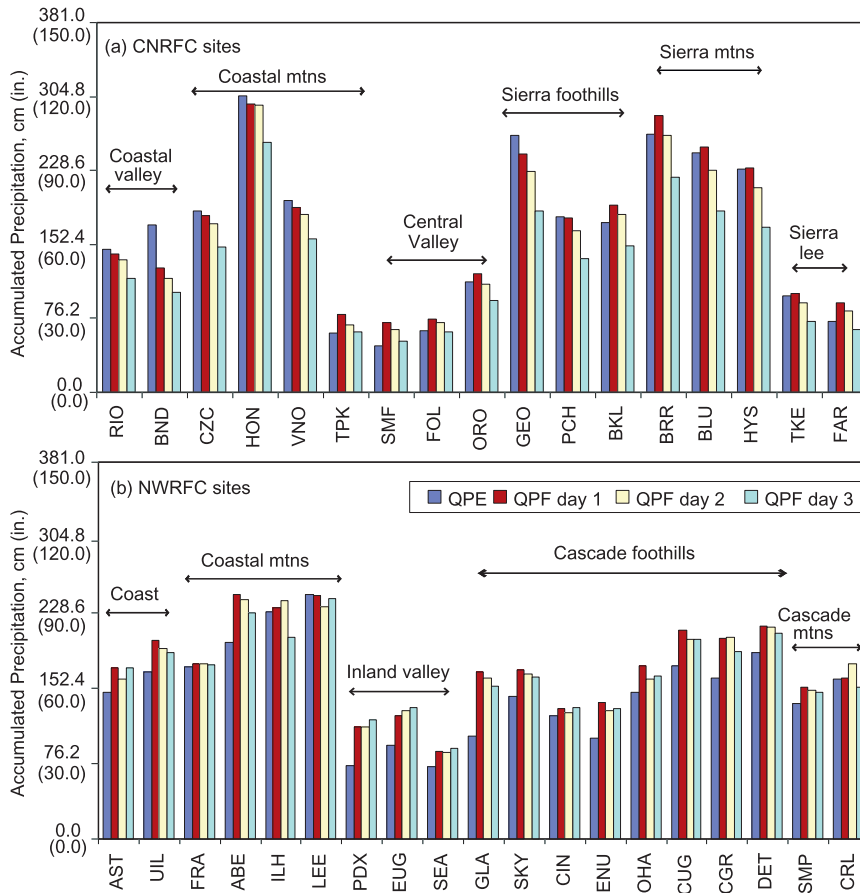


FIG. 4. Total observed and predicted accumulated precipitation amounts for days 1–3 QPF by verification site. (a) The 17 CNRFC verification sites are categorized geographically as coastal valley, coastal mountains, Central Valley, Sierra foothills, Sierra mountains and Sierra lee. (b) Of the 24 NWRFC verification sites, 19 sites are categorized geographically as coastal, coastal mountains, inland valley, Cascade foothills, and Cascades.

and HON in the CNRFC and ABE in the NWRFC); the errors mostly represent underprediction in the CNRFC and overprediction in the NWRFC; and the range of errors can vary significantly from one geographic situation to another (e.g., moderate and extreme precipitation events were mostly underpredicted for the Sierra lee, coastal valley, and coastal mountain sites, while they were often overpredicted for the Central Valley and inland valley).

To assess the QPFs in terms of predicting moderate and extreme events specifically, the POD, FAR and CSI scores were examined for observed and predicted precipitation events exceeding specified precipitation thresholds. Figure 6a shows the POD and FAR scores for days 1–3 QPFs in the CNRFC and Fig. 6b shows the CSI scores. On the basis of the CSI scores (Fig. 6b), it can be seen that the performance of the CNRFC QPFs decreases both for larger precipitation thresholds and for greater forecast lead times. While neither of these trends is surprising, it is

important to quantify them, and it is a key step toward deeper study. Closer examination of the POD and FAR scores (Fig. 6a) shows that the POD score decreases quickly from 0.98 at $0.025 \text{ cm (24 h)}^{-1}$ [$0.01 \text{ in. (24 h)}^{-1}$] to 0.06 at $12.7 \text{ cm (24 h)}^{-1}$ [$5.0 \text{ in. (24 h)}^{-1}$], while the FAR scores have a slower rate of increase from 0.15 at $0.025 \text{ cm (24 h)}^{-1}$ [$0.01 \text{ in. (24 h)}^{-1}$] to 0.50 at $12.7 \text{ cm (24 h)}^{-1}$ [$5.0 \text{ in. (24 h)}^{-1}$]. This implies that the CNRFC QPFs tend to be more conservative, especially for larger precipitation events, with QPF values more likely to be less than the observed 24-h accumulated precipitation amounts rather than falsely overpredicted. In comparison, Figs. 6c and 6d show the POD, FAR, and CSI scores for the NWRFC sites for days 1–3 QPFs. Similar to the CNRFC, the NWRFC forecast scores show that the POD decreases with increased precipitation threshold; however, unlike the CNRFC, the NWRFC FAR increases rapidly, indicating that not only does the NWRFC tend to have more forecast misses and less hits at larger precipitation

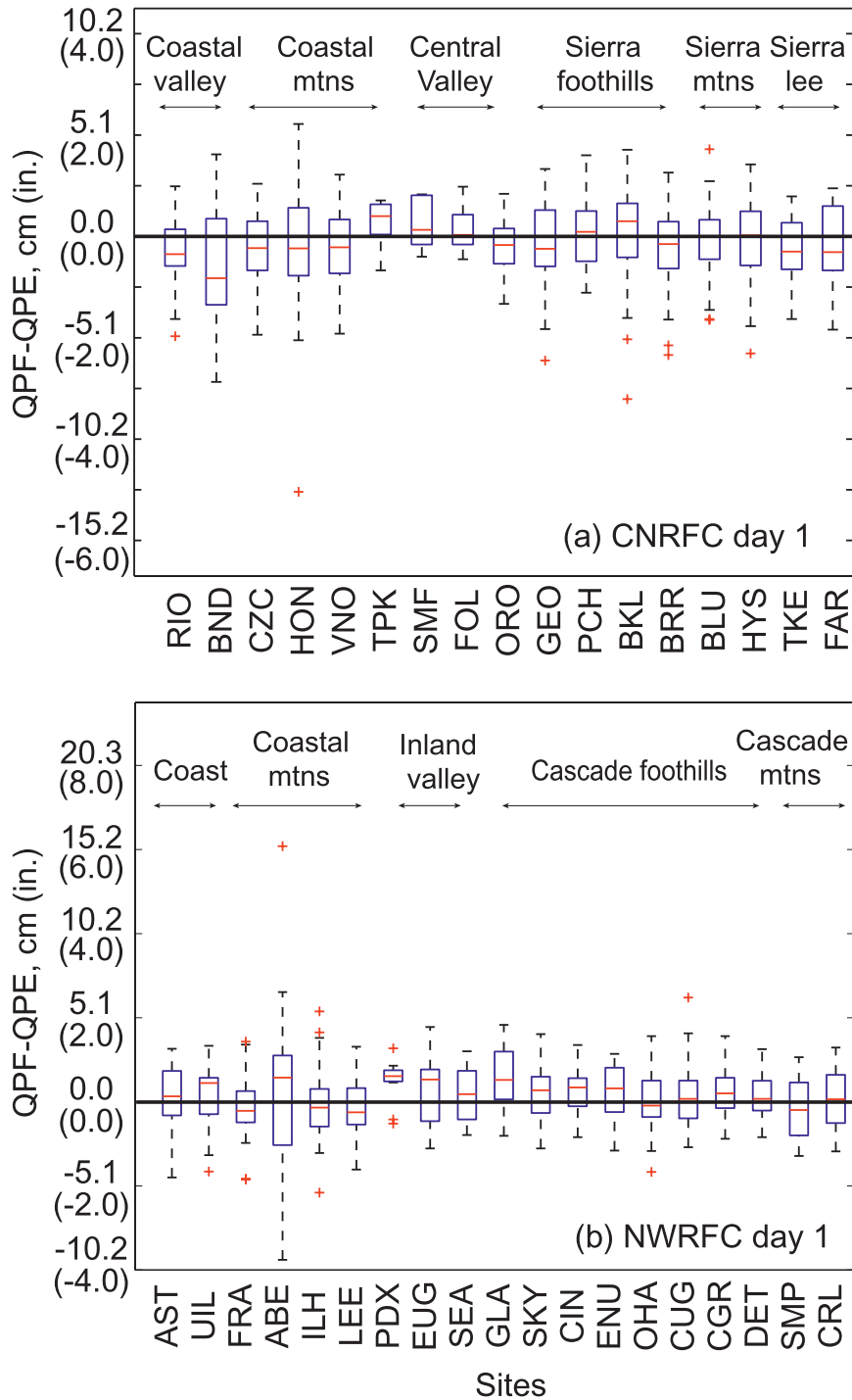


FIG. 5. Box (and whisker) plots of the daily forecast error (i.e., QPF – QPE) for verification sites representative of key geographic regions in the (a) CNRFC and (b) NWRFC regions for the day 1 lead time are shown. Only dates when the QPE or QPF exceeded $2.5 \text{ cm (24 h)}^{-1}$ [$1 \text{ in. (24 h)}^{-1}$] at a given site are included. The bottom and top edges of each box show the upper and lower quartiles (i.e., 75th and 25th percentile values) of the data. The whiskers extend from the ends of each box to 1.5 times the interquartile range, or to the most extreme value if the data point is less than that (Hoaglin et al. 1983). Data points that exceed 1.5 times the interquartile range are designated as outliers and are indicated by red plus symbols. The red line through each box shows the median value at each site.

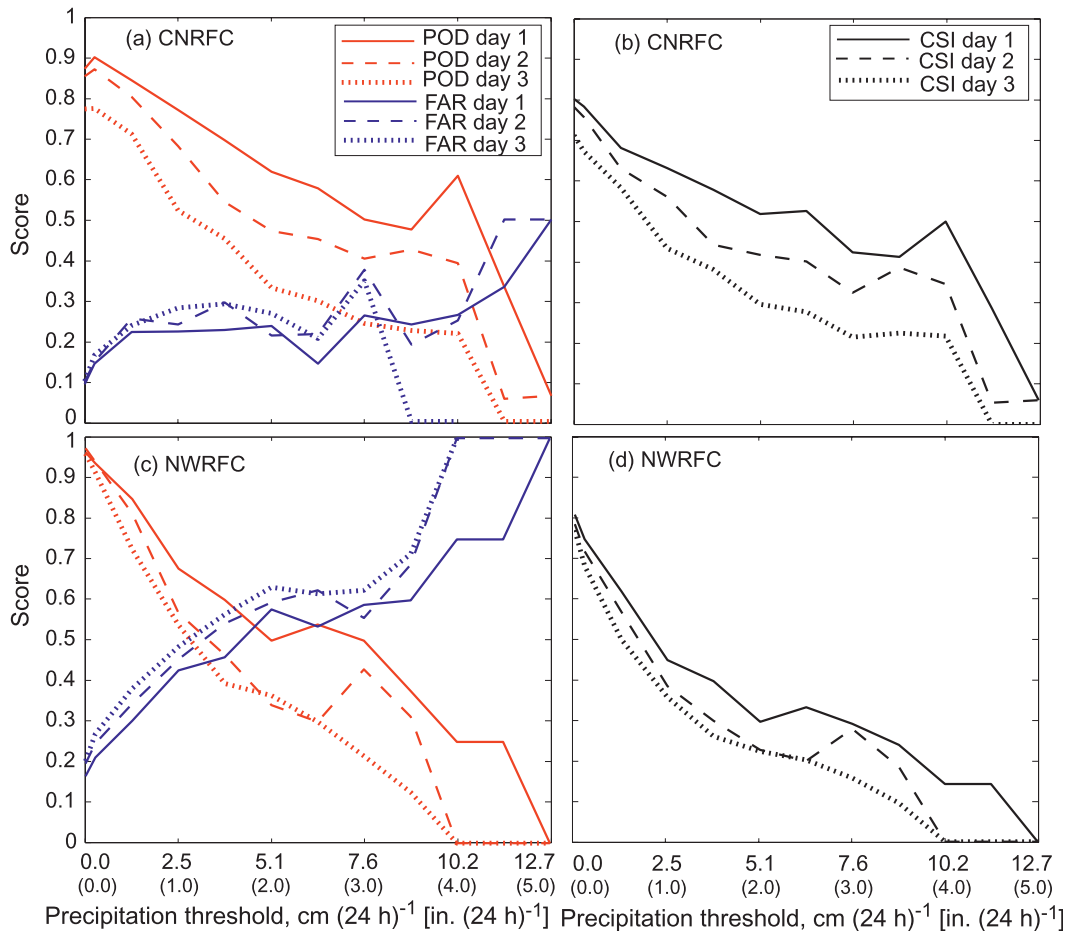


FIG. 6. The (a) POD (red) and FAR (blue) values as a function of event precipitation threshold {cm (24 h)⁻¹ [in. (24 h)⁻¹]} for the CNRFC and (b) CSI (black) values vs event threshold (inches) for the CNRFC verification sites. (c) POD (red) and FAR (blue) values as a function of precipitation threshold {cm (24 h)⁻¹ [in. (24 h)⁻¹]} for the NWRFC and (d) CSI (black) values by event precipitation threshold {cm (24 h)⁻¹ [in. (24 h)⁻¹]} for the NWRFC verification sites. Solid, dashed, and dotted lines indicate the days 1–3 forecast intervals, respectively.

thresholds but the NWRFC QPFs also tend to have more false alarms. This is further supported by calculating the MAE for each precipitation threshold. Figure 7 shows the MAE calculated for QPE events exceeding the precipitation threshold for the CNRFC and the NWRFC. As stated previously, the MAE is a good measure of accuracy. Here both RFC regions indicate that the MAE increases with larger precipitation threshold, indicating that the accuracy of the QPFs decreases with larger precipitation amounts. It also shows that the MAE increases as the lead time increases for each RFC. For example, at a precipitation threshold of 10.2 cm (24 h)⁻¹ [4 in. (24 h)⁻¹], the CNRFC day 1 MAE is about 5.1 cm (24 h)⁻¹ [2 in. (24 h)⁻¹], which is roughly half of the 10.2 cm (24 h)⁻¹ [4 in. (24 h)⁻¹] threshold precipitation value, while the CNRFC day-3 MAE is about 7.6 cm (24 h)⁻¹ [3 in. (24 h)⁻¹], which is three-fourths of that threshold precipitation.

c. Extreme events

The issue of forecasting extreme events is further documented in Fig. 8, which shows the POD, FAR, and CSI scores of both the CNRFC and NWRFC day 1 QPFs for the entire season. In this particular cool season, the POD for >7.6 cm (24 h)⁻¹ [3.0 in. (24 h)⁻¹] was 0.50 at the CNRFC and 0.50 at the NWRFC, while the FAR was ~0.3 at the CNRFC and ~0.6 at the NWRFC. Thus, the CSI for >7.6 cm (24 h)⁻¹ [3.0 in. (24 h)⁻¹] was higher at the CNRFC (0.42), relative to the NWRFC (0.29). However, we know from Fig. 2 that the CNRFC had more events exceeding 7.6 cm (24 h)⁻¹ [3.0 in. (24 h)⁻¹] than NWRFC. Within the CNRFC domain, there were 62 events that exceeded 7.6 cm (24 h)⁻¹ [3.0 in. (24 h)⁻¹; Fig. 9a], while there were only 28 events in the NWRFC domain (Fig. 9b). Further, it is also useful

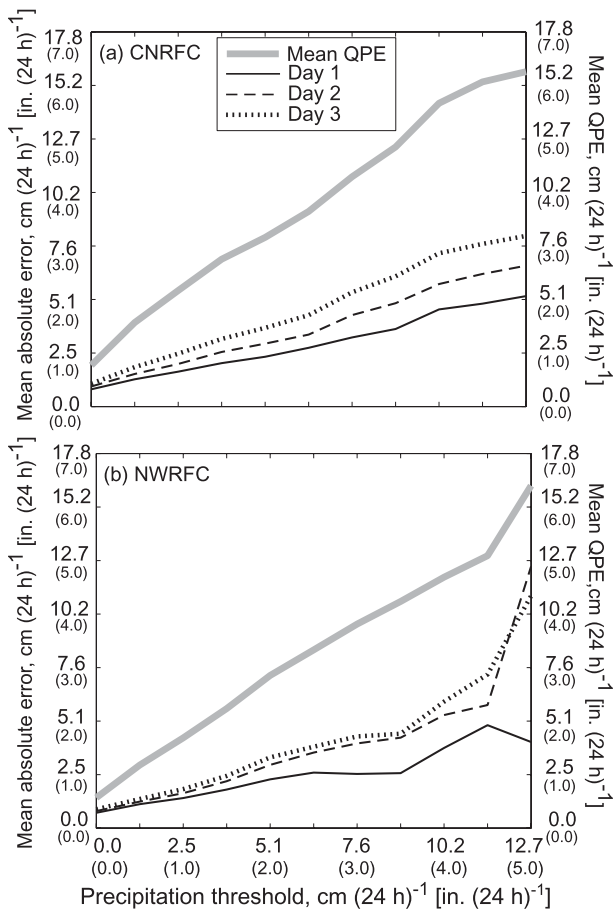


FIG. 7. The MAE $\{\text{cm (24 h)}^{-1}$ [in. $(24 \text{ h})^{-1}$]\} for events exceeding the 24-h precipitation thresholds [centimeters (inches)] by forecast lead time (days 1–3) for the (a) CNRFC and (b) NWRFC. For reference, the mean QPE $\{\text{cm (24 h)}^{-1}$ [in. $(24 \text{ h})^{-1}$]\} exceeding each precipitation threshold is also included.

to note that of the 62 events in the CNRFC, 16 of those events surpassed $12.7 \text{ cm (24 h)}^{-1}$ [$5.0 \text{ in. (24 h)}^{-1}$], while the NWRFC only had 1 such event.

Figure 9a shows that of the 16 observed site events in the CNRFC region that exceeded $12.7 \text{ cm (24 h)}^{-1}$ [$5.0 \text{ in. (24 h)}^{-1}$], only 2 events were predicted. Closer examination of the data showed that of the 2 events predicted 24 h ahead of time (i.e., day 1), only 1 of the predicted events actually occurred; the other predicted event was a false alarm. The results are the same for the 2 predicted events on day 2: 1 event occurred and the other was a false alarm. Thus, of the 16 site events that exceeded $12.7 \text{ cm (24 h)}^{-1}$ [$5.0 \text{ in. (24 h)}^{-1}$], only 1 site event was predicted. Similar examination of the NWRFC data showed that each event predicted to be greater than $12.7 \text{ cm (24 h)}^{-1}$ [$5.0 \text{ in. (24 h)}^{-1}$] was a false alarm. It is intriguing to note that an analysis of model-generated QPFs in the OR–WA area, performed

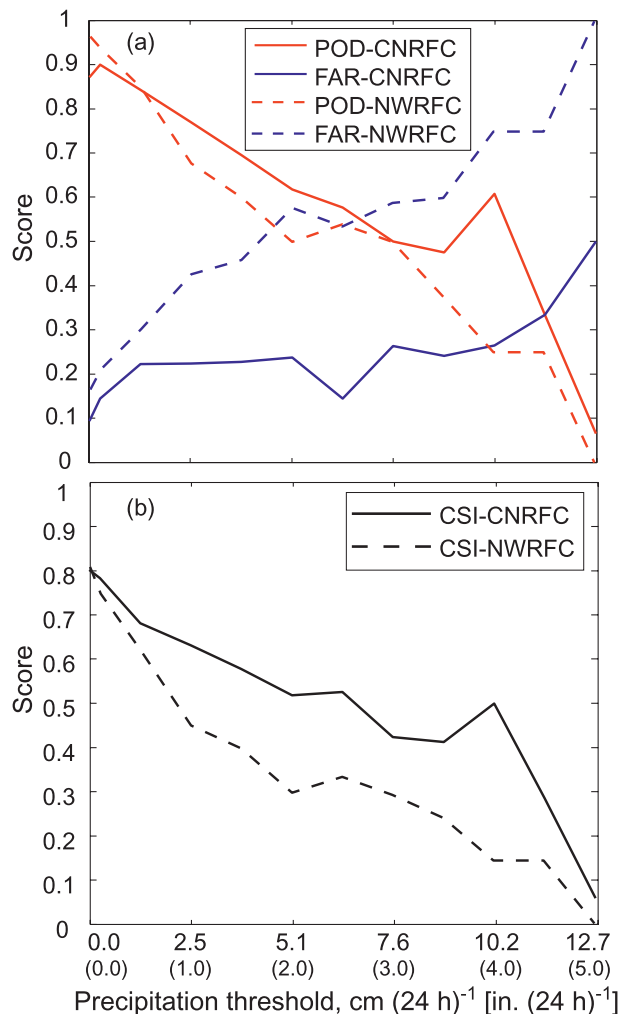


FIG. 8. Comparison of day 1 (a) POD (red) and FAR (blue) values as a function of precipitation threshold in cm (in.) for the CNRFC (solid) and NWRFC (dashed) verification sites. (b) CSI (black) values vs event threshold in cm (in.) for the CNRFC verification sites (solid) and NWRFC verification sites (dashed) for the day 1 forecast interval.

as part of a study of the effect of new satellite data on QPF [Constellation Observing System for Meteorology Ionosphere and Climate (COSMIC); Z. Ma et al. 2010, unpublished manuscript], found that 1 occurrence of $>12.7 \text{ cm (24 h)}^{-1}$ [$5.0 \text{ in. (24 h)}^{-1}$] in that storm (November 2006) was predicted out of the 8 events that were observed. Additionally, the study found that 12 of 18 occurrences of $>7.6 \text{ cm (24 h)}^{-1}$ [$3.0 \text{ in. (24 h)}^{-1}$] were predicted, which is nearly identical to the 42 of 62 events in the season studied here. The similarities suggest the results of this paper may be more broadly applicable than for just the one winter studied. Clearly, the skill in predicting the less-frequent-but-larger-precipitation

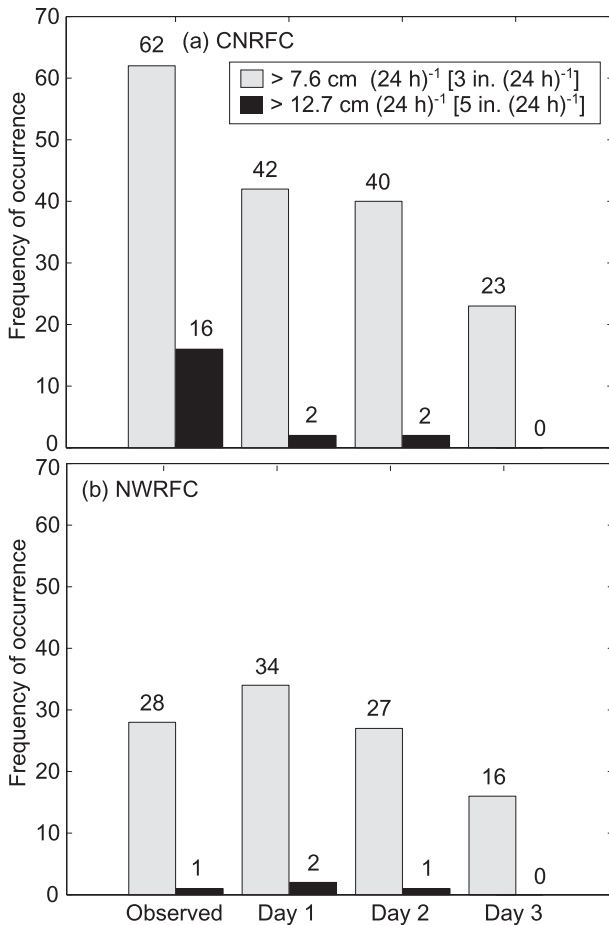


FIG. 9. Number of events (observed or predicted) exceeding 7.6 cm (24 h)⁻¹ [3 in. (24 h)⁻¹] and 12.7 cm (24 h)⁻¹ [5 in. (24 h)⁻¹] of accumulated precipitation for the (a) CNRFC and (b) NWRFC verification sites for the 2005/06 cool season.

events is much lower than the prediction of weaker precipitation events.

To further illustrate the nature and prediction of the events exceeding 7.6 cm (24 h)⁻¹ [3.0 in. (24 h)⁻¹], the spatial distribution of these observed events is shown in Fig. 10 (see Table 2 for details of specific events). As is to be expected, these events occurred primarily in the mountains, and for this winter, at least, they were focused in California. The QPFs for these events are shown in Fig. 11 at verification sites in the CNRFC and NWRFC domains. There is a clear tendency to underpredict precipitation amounts on these days in both RFC domains. Similarly, there is a trend for this underprediction to increase with greater forecast lead times. For events that exceed 7.6 cm (24 h)⁻¹ [3.0 in. (24 h)⁻¹], the average biases for the CNRFC domain were calculated to be 0.88 for day 1, 0.82 for day 2, and 0.60 for day 3, while the NWRFC domain had average biases of 1.47 for day 1, 1.58 for day 2, and 1.22 for day 3. The CNRFC

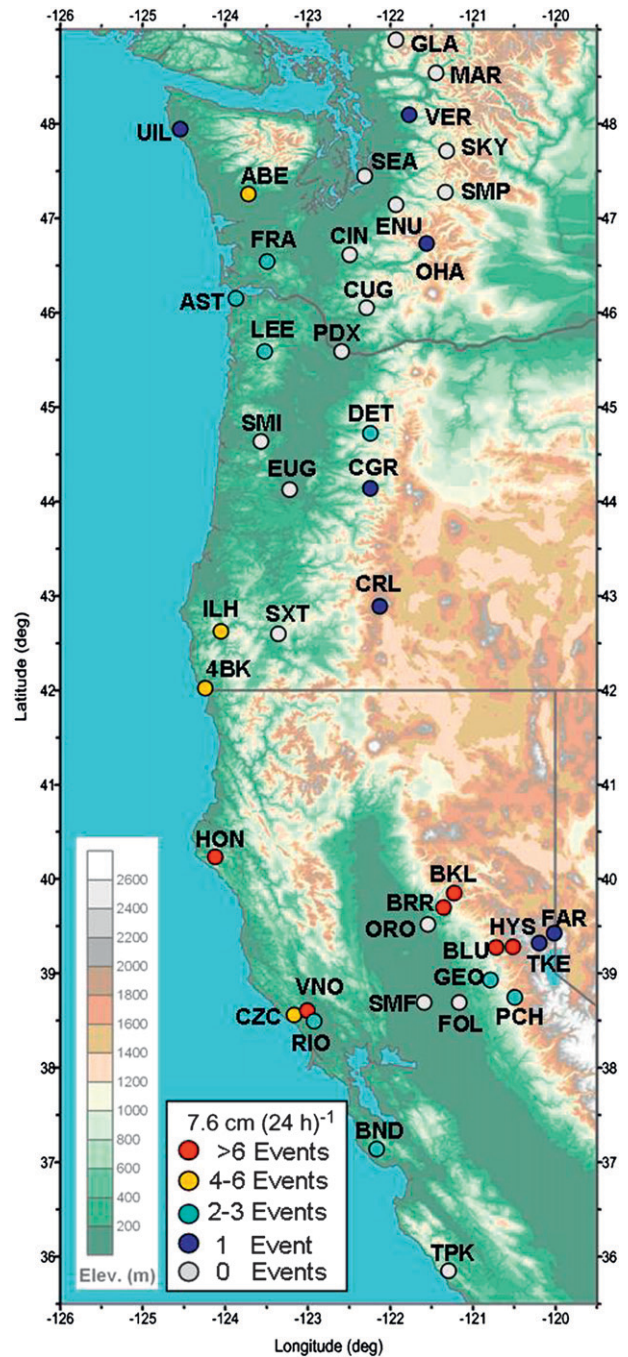


FIG. 10. Map of the 41 verification sites indicating the number of observed days when >7.6 cm (24 h)⁻¹ [3 in. (24 h)⁻¹] of precipitation was observed.

values are lower than the biases calculated previously in section 3b for all events, with precipitation exceeding 0.25 cm (24 h)⁻¹ [0.1 in. (24 h)⁻¹] in the CNRFC (see section 4b; 1.01 for day 1, 0.95 for day 2, and 0.79 for day 3). For the NWRFC, the extreme event biases are higher than the biases seen when all events are considered

TABLE 2. Summary of observed, correctly predicted (day 1), and incorrectly predicted (day 1) precipitation events $>7.6 \text{ cm (24 h)}^{-1}$ [$3.0 \text{ in. (24 h)}^{-1}$] for the CNRFC (CN) and NWRFC (NW) regions. Those dates for which there was only one site either observed or predicted are excluded. Indications of AR potential based on Special Sensor Microwave Imager (SSM/I) data using the methods from Neiman et al. (2008b) are also noted.

Date	RFC region	No. observed sites	No. correctly predicted sites (hits)	No. incorrectly predicted sites (false alarms)	AR
6 Nov 2005	NW	3	2	0	x
2 Dec 2005	CN	7	4	0	x
19 Dec 2005	CN	7	0	0	x
22 Dec 2005	CN	2	1	0	x
23 Dec 2005	CN-NW	4	1	0	x
25 Dec 2005	NW	3	3	0	
26 Dec 2005	CN	2	0	0	x
28 Dec 2005	CN-NW	6	4	0	x
29 Dec 2005	CN	1	1	1	
31 Dec 2005	CN-NW	14	11	0	x
1 Jan 2006	CN-NW	1	1	4	x
6 Jan 2006	NW	2	0	0	x
10 Jan 2006	NW	7	1	0	x
11 Jan 2006	NW	3	0	0	x
17 Jan 2006	NW	1	0	5	x
30 Jan 2006	NW	4	3	0	x
27 Feb 2006	CN	2	0	0	x
28 Feb 2006	CN	6	6	5	x
6 Mar 2006	CN	4	3	0	x
12 Apr 2006	CN	3	1	0	x

(see section 3b; 1.16 for day 1, 1.14 for day 2, and 1.10 for day 3). This comparison highlights the importance of treating QPF verification for extreme events with special attention.

An additional benefit of focusing on extreme events is that it allows for the identification of key atmospheric conditions for which better forecasts are needed. In the case of the HMT-2006 winter season, there were 20 dates for which $>7.6 \text{ cm (24 h)}^{-1}$ [$3.0 \text{ in. (24 h)}^{-1}$] was observed or predicted within the CNRFC and/or NWRFC study area. Those dates for which there was only one site either observed or predicted are excluded. Several recent studies have concluded that ARs play a major role in extreme precipitation events along the West Coast. Ralph et al. (2006) showed that all seven of the flood events on the Russian River during the 8-yr period where the necessary experimental observations were available were associated with well-defined AR conditions striking the coast. Similarly, Neiman et al. (2008a) concluded that a strong and slow-moving AR struck the Pacific Northwest in November 2006 and was a major contributor to the record-breaking precipitation in the region. The key paper for this study, however, is Neiman et al. (2008b), which provided a catalogue of all AR events developed using methods founded on techniques developed by Zhu and Newell (1998) and Ralph et al. (2004, 2005a). The date for each event was compared with the catalogue of AR events documented

independently by Neiman et al. (2008b), which reveals that 18 of 20 dates of extreme precipitation in the CNRFC and/or NWRFC were associated with land-falling AR conditions. Of the sites that observed $>7.6 \text{ cm (24 h)}^{-1}$ [$3.0 \text{ in. (24 h)}^{-1}$] during these particular dates, approximately 50% were correctly predicted (POD of 0.50) and the FAR was 0.26. This analysis points to the importance of evaluating QPF effects in the context of AR conditions, which are ideal for generating the orographic precipitation (Ralph et al. 2005a) that is primarily the cause of the extreme precipitation in this region. Jankov et al. (2009) has taken a step in this direction by evaluating microphysical parameterizations and low-level winds in a state-of-the-art mesoscale model.

4. Defining extreme precipitation events nationally

Although the analysis presented above focuses on QPF verification for the West Coast, the scope of NOAA's QPF services is national, and thus QPF verification needs to occur on a national level. To assess QPF performance throughout each region of the United States, suitable 24-h accumulated precipitation thresholds are derived using a method developed here to objectively determine appropriate extreme precipitation thresholds for different regions. The approach takes advantage of the decades-long data available from the COOP network (National

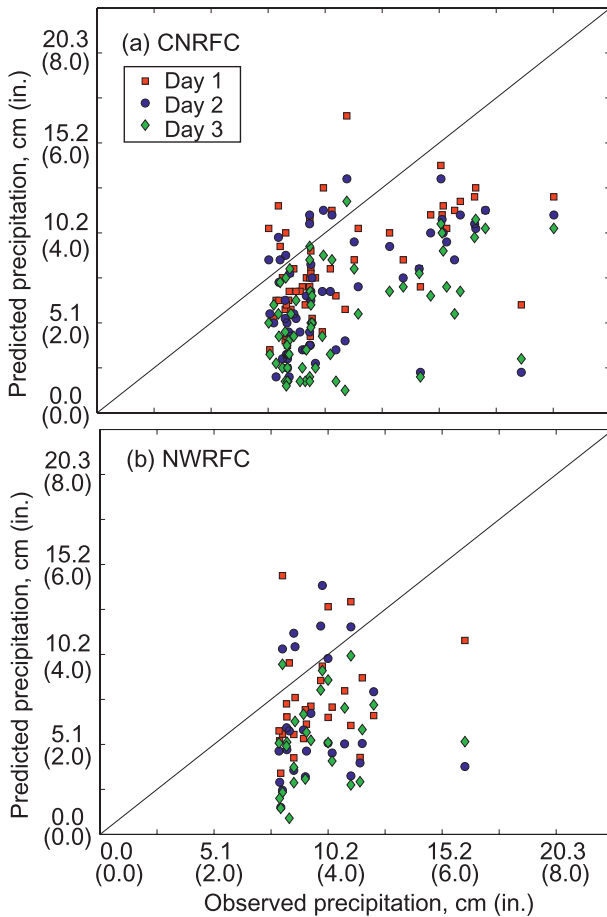


FIG. 11. Scatterplot of observed accumulated precipitation vs predicted accumulated precipitation for forecast intervals of day 1 (red squares), day 2 (blue circles), and day 3 (green diamonds) for observed events with precipitation exceeding $7.6 \text{ cm (24 h)}^{-1}$ [$3 \text{ in. (24 h)}^{-1}$] in the (a) CNRFC and (b) NWRFC regions.

Weather Service 1989), which includes 24-h accumulated precipitation totals. Although the COOP data have a variety of limitations (e.g., the samples are not gathered at the same time across all sites), the availability of roughly 6000 sites nationally and their long-term nature make them invaluable for assessing the frequency of occurrence of extreme precipitation amounts. Additionally, the COOP sites represent point measurements; however, the QPE used elsewhere in this paper are obtained from $4 \text{ km} \times 4 \text{ km}$ gridded datasets. For this reason, the COOP data have not been used directly for verification of the gridded forecasts.

Because of the geographic variability of precipitation behavior in different regions of the contiguous United States (e.g., land-falling tropical storms in the Southeast, atmospheric rivers on the West Coast, mesoscale convective systems in the central plains, and monsoons in the Southwest), variations in the nature of the extreme

events could be significant. Given that NOAA has already organized its river forecast system into 13 RFCs (Fig. 12) and that river forecasting is a major requirement driver for accurate extreme QPF, the approach taken here is to use the RFC regions as the verification domains.

To quantify the frequency of occurrence of extreme precipitation within subregions of the United States, the COOP data were segregated into 13 areas, each one representing a RFC region. Table 3 lists the number of COOP sites in each region, which varies from 134 to 948 sites. The next step was to identify only those COOP-site days (24-h periods) when measurable precipitation was recorded at a site in the region. The total number of these “wet” days for each RFC region is also shown in Table 3 and ranges from $\sim 1.2 \times 10^6$ site days in the Colorado Basin RFC (CBRFC) region to $\sim 4.1 \times 10^6$ in the North Central RFC (NCRFC). These differences result not only from geographic variations in the frequency of wet days but also from the length of the COOP site records, which varies greatly from region to region and from site to site (no minimum record length was required, but the average record length was 33 complete years per site). In total, 5886 COOP sites were used, which represented almost 30 million wet days nationally. While there are clearly variations in sampling distributions with altitude and other factors, the sheer size of this dataset lends some statistical significance to the results.

The frequency of occurrence of daily rainfall events was calculated for a wide range of thresholds of observed daily rainfall. Only those days that recorded any rainfall were considered, and thus the results correspond to the fraction of wet days exceeding the various daily precipitation thresholds. These occurrences were calculated separately for each RFC so as to reveal geographical variations in the frequency of extreme rainfall across the nation. The results for each RFC are shown in Fig. 13, where there are two distinct clusters of behavior and one outlier, especially in the frequency range of interest from roughly 1/100 to 1/1000. Interestingly, the $7.6\text{-cm (24 h)}^{-1}$ [$3.0 \text{ in. (24 h)}^{-1}$] threshold used in the analyses of the CNRFC and NWRFC regions presented earlier corresponds almost exactly to a frequency of 1/100, while the $12.7\text{-cm (24 h)}^{-1}$ [$5.0 \text{ in. (24 h)}^{-1}$] threshold corresponds nearly to 1/1000 for the RFCs that most frequently experience extreme rainfall. On the basis of the knowledge gained in HMT and on the utility of the $7.6\text{-cm (24 h)}^{-1}$ [$3.0 \text{ in. (24 h)}^{-1}$] and $12.7\text{-cm (24 h)}^{-1}$ [$5.0 \text{ in. (24 h)}^{-1}$] thresholds developed independently from this COOP analysis, the thresholds of 1/100 and 1/1000 are adopted in the following analysis. Table 3 shows the exact values for each RFC corresponding to

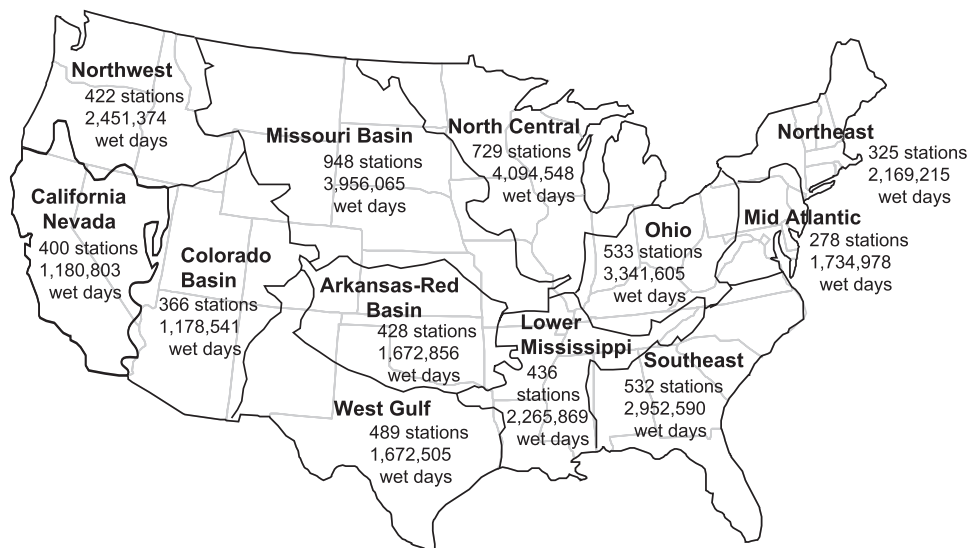


FIG. 12. Map showing the contiguous United States (CONUS) NWS RFC regions. Each region is labeled with the number of COOP sites within that region that were used for the analysis of frequency of occurrence of precipitation amounts and the number of COOP-site days with measurable precipitation during the period of record examined.

these frequencies. Remarkably, they fall into three nearly distinct groups, corresponding roughly to $3.8 \text{ cm (24 h)}^{-1}$ [$1.5 \text{ in. (24 h)}^{-1}$] in the CBRFC, $5.1 \text{ cm (24 h)}^{-1}$ [$2.0 \text{ in. (24 h)}^{-1}$] for the northern tier of RFCs, and $7.6 \text{ cm (24 h)}^{-1}$ [$3.0 \text{ in. (24 h)}^{-1}$] in the southern tier of RFCs plus the CNRFC. These analyses represent a perspective on extreme events that complements the “return period” approach emphasized in the NOAA Atlas 2 (e.g., Miller et al. 1973).

The differences seen between the CNRFC and NWRFC in the data from the one winter studied in detail (i.e., the 2005/06 cool season) are consistent with this

climatological result where $5.1 \text{ cm (24 h)}^{-1}$ [$2.0 \text{ in. (24 h)}^{-1}$] events have the same frequency in the NWRFC as do the $7.6 \text{ cm (24 h)}^{-1}$ [$3.0 \text{ in. (24 h)}^{-1}$] events in the CNRFC. In short, the CNRFC is more likely to receive more extreme daily precipitation amounts.

On the basis of the results for each RFC, thresholds for extreme 24-h precipitation for each RFC are proposed (Fig. 14). [Note, although the Alaska–Pacific RFC (APRFC) is not shown in Fig. 14, the extreme precipitation thresholds for APRFC are shown in Fig. 13 and Table 3.] By slightly rounding the thresholds, a simple range of thresholds can be used. It is useful to

TABLE 3. Analysis of COOP daily precipitation amounts as a function of NWS RFC domains.

RFC	No. COOP sites	No. wet site days	Threshold of 0.01 frequency [cm (in.)]	Threshold of 0.001 frequency [cm (in.)]
Lower Mississippi	436	2 265 869	7.9 (3.1)	14.2 (5.6)
Alaska–Pacific	134	998 375	7.9 (3.1)	18.3 (7.2)
California–Nevada	400	1 180 803	7.6 (3.0)	14.2 (5.6)
Southeast	532	2 952 590	7.6 (3.0)	13.7 (5.4)
West Gulf	489	1 672 505	7.6 (3.0)	14.5 (5.7)
Arkansas–Red Basin	428	1 672 856	7.1 (2.8)	13.0 (5.1)
Middle Atlantic	278	1 734 978	5.6 (2.2)	10.4 (4.1)
Ohio	533	3 341 605	5.3 (2.1)	9.4 (3.7)
Missouri Basin	948	3 956 065	5.3 (2.1)	9.7 (3.8)
North Central	729	4 094 548	5.1 (2.0)	9.4 (3.7)
Northeast	325	2 169 215	5.1 (2.0)	9.4 (3.7)
Northwest	422	2 451 374	5.1 (2.0)	9.4 (3.7)
Colorado Basin	366	1 178 541	3.8 (1.5)	7.1 (2.8)
National sums (contiguous)	6,020	29 669 324		

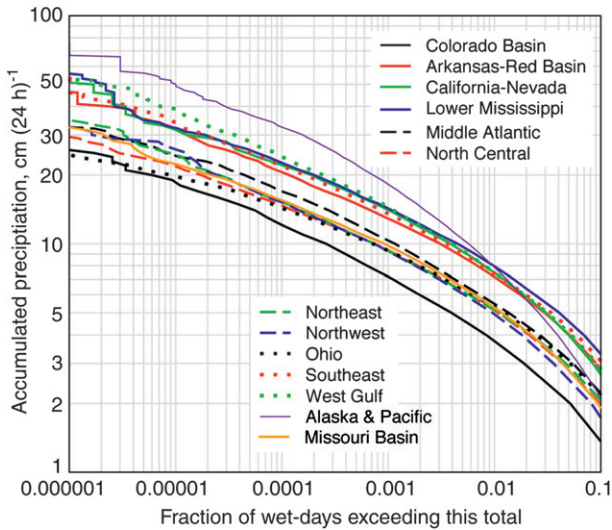


FIG. 13. Frequency of occurrence of different daily precipitation amounts observed by the COOP network as a function of the 13 NWS RFC domains.

note that the values found here that define extreme precipitation are consistent with the range used by Groisman et al. (2001) to explore long-term climate trends in heavy precipitation, also based on COOP data and a frequency-of-occurrence analysis.

5. Conclusions and future work

The prediction of extreme precipitation events is a critical challenge with many important applications. This paper uses data from the NWS’s CNRFC and NWRFC to document the performance of QPFs during HMT-2006, which took place in the winter of 2005/06 on the West Coast, where the landfall of ARs led to extreme precipitation. The challenge of predicting the 62 events in the CNRFC domain where >7.6 cm (3.0 in.) of precipitation was observed in 24 h, including the 16 events that exceeded 12.7 cm $(24\text{ h})^{-1}$ [5.0 in. $(24\text{ h})^{-1}$], was illustrated using a combination of POD, FAR, CSI, MAE, and bias. The method was then applied to the NWRFC as well, which showed that the approach developed initially for the CNRFC area is also applicable to the Washington–Oregon area. Table 4 shows the performance for the CNRFC for events >7.6 cm $(24\text{ h})^{-1}$ [3.0 in. $(24\text{ h})^{-1}$] and for those >12.7 cm $(24\text{ h})^{-1}$ [5.0 in. $(24\text{ h})^{-1}$] at forecast lead times of days 1–3. The CNRFC POD for events exceeding 7.6 cm $(24\text{ h})^{-1}$ [3.0 in. $(24\text{ h})^{-1}$] was 0.5, while the POD for events exceeding 12.7 cm $(24\text{ h})^{-1}$ [5.0 in. $(24\text{ h})^{-1}$] was 0.06. Also, the MAE for the extreme events was found to be about 50% of the observed precipitation for all lead times. The bias for these events worsened with greater lead time, decreasing from 0.72 at day 1 lead time to 0.49 for day 3 lead

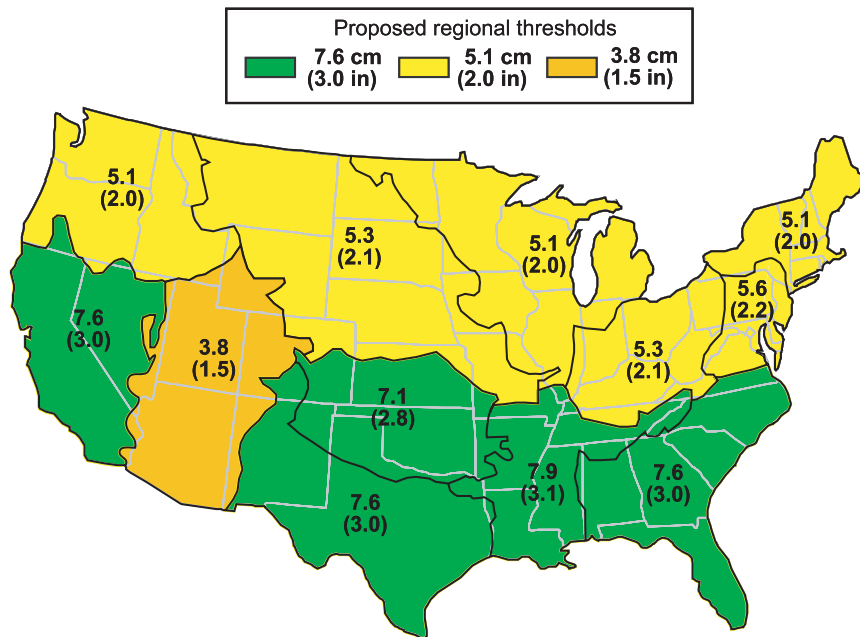


FIG. 14. CONUS NWS RFC regional thresholds for daily precipitation amounts recommended for future use by the NWS in tracking QPF performance in high-impact extreme precipitation events using the POD, FAR, CSI, bias, and MAE statistical methods. In each RFC region, the upper number is the threshold of the top 1% of precipitation events over the period from 1950 to 2007 in centimeters and the bottom number is in inches.

TABLE 4. Summary of skill score values for events with accumulated precipitation >7.6 cm (24 h) $^{-1}$ [3.0 in. (24 h) $^{-1}$] and >12.7 cm (24 h) $^{-1}$ [5.0 in. (24 h) $^{-1}$] in the CNRFC and NWRFC regions.

CNRFC metric	>7.6 cm (24 h) $^{-1}$ [3 in. (24 h) $^{-1}$]			>12.7 cm (24 h) $^{-1}$ [5.0 in. (24 h) $^{-1}$]		
	Day 1	Day 2	Day 3	Day 1	Day 2	Day 3
CSI	0.42	0.32	0.21	0.06	0.06	0.00
POD	0.50	0.40	0.24	0.06	0.06	0.00
FAR	0.26	0.38	0.35	0.50	0.50	N/A
MAE {cm (24 h) $^{-1}$ [in. (24 h) $^{-1}$]}	3.1 (1.2)	4.0 (1.6)	5.0 (2.0)	5.3 (2.1)	6.4 (2.5)	8.1 (3.2)
RMSE {cm (24 h) $^{-1}$ [in. (24 h) $^{-1}$]}	3.7 (1.5)	4.9 (1.9)	5.9 (2.3)	5.8 (2.3)	7.3 (2.9)	8.6 (3.4)
Bias	0.88	0.82	0.60	0.72	0.63	0.49

NWRFC metric	>7.6 cm (24 h) $^{-1}$ [3 in. (24 h) $^{-1}$]			>12.7 cm (24 h) $^{-1}$ [5.0 in. (24 h) $^{-1}$]		
	Day 1	Day 2	Day 3	Day 1	Day 2	Day 3
CSI	0.29	0.29	0.16	0.00	0.00	0.00
POD	0.50	0.44	0.22	0.00	0.00	0.00
FAR	0.59	0.56	0.63	1.00	1.00	N/A
MAE {cm (24 h) $^{-1}$ [in. (24 h) $^{-1}$]}	3.3 (1.3)	4.4 (1.7)	4.6 (1.8)	5.5 (2.2)	8.3 (3.3)	11.0 (4.3)
RMSE {cm (24 h) $^{-1}$ [in. (24 h) $^{-1}$]}	4.0 (1.6)	5.0 (2.0)	5.3 (2.1)	5.9 (2.3)	9.3 (3.6)	11.0 (4.3)
Bias	1.47	1.58	1.22	1.39	0.83	0.32

time. Comparing these biases to the biases for all events (i.e., including those with only light precipitation), which were 1.01 (day 1) and 0.79 (day 3), shows that biases for QPFs for the extreme events were worse (i.e., greater underprediction) than for weaker events. This highlights the special challenges of QPFs in the high-impact events.

Similarities and differences between the CNRFC and NWRFC were evident in the verification scores for extreme events, particularly the seasonal bias, MAE, and FAR. Interestingly, the POD values were nearly the same for both RFCs for events >7.6 cm (24 h) $^{-1}$ [3.0 in. (24 h) $^{-1}$]. Another difference between the two RFCs is in the number of events >7.6 cm (24 h) $^{-1}$ [3.0 in. (24 h) $^{-1}$]. The total number in the NWRFC (28 events), for the same period, is less than half the 62 events observed in the CNRFC. Many factors may contribute to the differences observed between the RFCs. For instance, the climatology of extreme events varies by RFC (Figs. 13 and 14), and details of the forecast procedures and tools differ from one RFC to another. These differences likely contribute to differing QPF performance; although, at the higher precipitation thresholds, the relatively smaller sample size makes it difficult to interpret. The design of this study did not allow for the documentation of the effects of variations in the predictability of key hydrometeorological phenomena from one region to another or of the differing forecast procedures and tools.

In addition to variations of hydrometeorological and forecasting procedures and tools from one RFC to another, the topography of each RFC varies. As mentioned in section 3b, the orographic nature of the extreme precipitation on the West Coast strongly influences the QPF errors by providing a non-time-varying forcing, which allows QPFs in mountainous regions to have

greater skill than in other regions without such strong and fixed forcing. However, the presence of complex terrain can also act to compound errors in QPFs under some conditions. For example, errors in lower-tropospheric wind direction cause QPF errors related to terrain orientation and rain shadowing (Brady and Waldstreicher 2001; Ralph et al. 2003), errors which become more significant as the terrain becomes more complex. Another source of errors is due to the existence of threshold phenomena, such as blocking [where blocking is a sensitive function of upwind stratification, winds, and mountain height (Smith 1979)], which can displace orographic precipitation upwind of mountains (e.g., Neiman et al. 2002). Problems with model microphysics are also important and can be amplified by mountains; for example, if the microphysics packages cause too much rain out in one place, then the downwind mountains will have less moisture or vice versa (Colle et al. 1999).

Partly to determine if the difference between the number of extreme events was an outlier (i.e., would the Pacific Northwest states normally have more extreme precipitation events), COOP daily precipitation totals were examined going back decades at roughly 400 sites in the CNRFC domain and 422 sites in the NWRFC domain. The COOP diagnostics show that the threshold defining the top 1% of 24-h precipitation events is 5.1 cm (24 h) $^{-1}$ [2.0 in. (24 h) $^{-1}$] in the NWRFC domain and 7.6 cm (24 h) $^{-1}$ [3.0 in. (24 h) $^{-1}$] in the CNRFC. This finding, and the method used to derive it, was then applied to each of the 12 RFCs in the continental United States (Fig. 14). The thresholds of extreme accumulated precipitation range from 3.8 cm (24 h) $^{-1}$ [1.5 in. (24 h) $^{-1}$] (CBRFC) to 7.6 cm (24 h) $^{-1}$ [3.0 in. (24 h) $^{-1}$] (CNRFC, WGRFC, and SERFC). Using the

same approach, thresholds were also derived for the top 0.1% of events. {It should be noted that the correspondence between the $7.6 \text{ cm (24 h)}^{-1}$ [$3.0 \text{ in. (24 h)}^{-1}$] thresholds used initially in this paper to develop an extreme event verification method and the 1% threshold based on COOP data is coincidental.} Because the thresholds derived from COOP data are observationally based, they represent a reliable baseline against which to evaluate changes in QPF methods and tools. However, comparisons against gridded data need to factor in the differences between gridded and point data, which are grid-resolution dependent. Also, since COOP observations are daily data, they are not useful in evaluating QPFs on finer time resolutions, such as every 6 h.

The methodology developed in this paper, both in terms of objective determination of regionally relevant thresholds defining extreme precipitation events and measuring QPF performance, could be readily applied to the existing forecast system (QPE and QPF). It is suggested here that NOAA begin measuring NWS extreme QPF performance by this method. Doing so would establish a measure that is more relevant to many of NOAA's customers' needs, and it would help stimulate focused research and prototyping (e.g., Neiman et al. 2009), leading to new forecast methods tailored to this key forecast challenge and NOAA service.

Future work will involve applying the method to the QPE and QPF data from all RFC forecast points, both as retrospectively as possible and in the future, as well as to gridded forecast products. QPF from the National Centers for Environmental Prediction (NCEP)'s Hydrometeorological Prediction Center (HPC) could be evaluated in this manner, as could model-based QPFs. Also, QPFs with 6-h time resolution could be explored (requiring better time resolution than COOP data can provide), including the quantification of the timing of the most extreme precipitation within the 24-h accumulation period. In addition, analysis of longer accumulation times (e.g., 72 h) will provide a broader view helpful for water management applications. Together, this work, overall, would establish a baseline against which to measure future QPF performance that is critical to many applications.

Establishing this performance measure will also help guide future research to improve extreme QPF, such as by improving microphysics in numerical models (e.g., Colle et al. 1999), by developing probabilistic QPF methods (e.g., Jankov et al. 2009; Yuan et al. 2009), and by better observing critical parameters (e.g., water vapor; Neiman et al. 2008a) and physical processes, such as ARs and their focused water vapor transport (Neiman et al. 2009; Jankov et al. 2009). Efforts already under way at HMT and the Developmental Testbed Center

(DTC; Bernardet et al. 2008) are taking the next steps in developing these measures and using them to assess state-of-the-art mesoscale numerical model forecasts.

Acknowledgments. Thanks to Wes Junker and Mike Ekern of the NWS for providing key data and useful guidance. Without the efforts of the overall HMT team, especially David Kingsmill and Tim Schneider, this work would not have been possible. The authors also wish to thank Lynn Johnson and Ed Tollerud for providing feedback and helpful suggestions.

REFERENCES

- Anthes, R. A., 1983: Regional models of the atmosphere in middle latitudes. *Mon. Wea. Rev.*, **111**, 1306–1335.
- Antolik, M. S., 2000: An overview of the National Weather Service's centralized statistical quantitative precipitation forecasts. *J. Hydrol.*, **239**, 306–337.
- Bernardet, L., and Coauthors, 2008: The Developmental Testbed Center and its Winter Forecasting Experiment. *Bull. Amer. Meteor. Soc.*, **89**, 611–627.
- Bosart, L. F., 1980: Evaluation of LFM-2 quantitative precipitation forecasts. *Mon. Wea. Rev.*, **108**, 1087–1099.
- Brady, R. H., and J. S. Waldstreicher, 2001: Observations of mountain wave–induced precipitation shadows over northeast Pennsylvania. *Wea. Forecasting*, **16**, 281–300.
- Charba, J. P., D. W. Reynolds, B. E. McDonald, and G. M. Carter, 2003: Comparative verification of recent quantitative precipitation forecasts in the National Weather Service: A simple approach for scoring forecast accuracy. *Wea. Forecasting*, **18**, 161–183.
- Colle, B. A., K. J. Westrick, and C. F. Mass, 1999: Evaluation of MM-5 and Eta-10 precipitation forecasts over the Pacific Northwest during the cool season. *Wea. Forecasting*, **14**, 137–154.
- Daly, C., R. P. Neilson, and D. L. Phillips, 1994: A statistical–topographic model for mapping climatological precipitation over mountainous terrain. *J. Appl. Meteor.*, **33**, 140–158.
- Dettinger, M., K. Redmond, and D. Cayan, 2004: Winter orographic precipitation ratios in the Sierra Nevada—Large-scale atmospheric circulations and hydrologic consequences. *J. Hydrometeorol.*, **5**, 1102–1116.
- Fread, D., and Coauthors, 1995: Modernization in the National Weather Service River and Flood Program. *Wea. Forecasting*, **10**, 477–484.
- Groisman, P. Y., R. W. Knight, and T. R. Karl, 2001: Heavy precipitation and high streamflow in the contiguous United States. *Bull. Amer. Meteor. Soc.*, **82**, 219–246.
- Hoaglin, D. C., F. Mosteller, and J. W. Tukey, 1983: *Understanding Robust and Exploratory Data Analysis*. Wiley, 447 pp.
- Jankov, I., J.-W. Bao, P. J. Neiman, P. J. Schultz, H. Yuan, and A. B. White, 2009: Evaluation and comparison of microphysical algorithms in WRF-ARW model simulations of atmospheric river events affecting the California Coast. *J. Hydrometeorol.*, **10**, 847–870.
- Lazo, J. K., R. E. Morss, and J. L. Demuth, 2009: 300 billion served: Sources, perceptions, uses, and values of weather forecasts. *Bull. Amer. Meteor. Soc.*, **90**, 785–798.
- Miller, J. F., R. H. Frederick, and R. J. Tracey, 1973: *California*. Vol. 11, *Precipitation-Frequency Atlas of the Western United*

- States, NOAA Atlas 2, National Weather Service, 50 pp. [Available from National Technical Information Service, 5285 Port Royal Rd., Springfield, VA 22161.]
- Morss, R. E., and F. M. Ralph, 2007: Use of information by National Weather Service forecasters and emergency managers during the CALJET and PACJET-2001. *Wea. Forecasting*, **22**, 539–555.
- National Research Council, 1995: *Flood Risk Management and the American River Basin: An Evaluation*. National Academies Press, 256 pp.
- , 1999: *Improving American River Flood Frequency Analysis*. National Academies Press, 132 pp.
- National Weather Service, 1989: National Weather Service Observing Handbook No. 2: Cooperative station observations, Office of Systems Operations, Observing Systems Branch, 94 pp. [Available online at <http://weather.gov/om/coop/Publications/coophandbook2.pdf>.]
- , 1999: The modernized end-to-end forecast process for quantitative precipitation information: Hydrometeorological requirements, scientific issues, and service concepts. National Weather Service, 187 pp. [Available from the Office of Climate, Water, and Weather Services, W/OS, 1325 East West Hwy., Silver Spring, MD 20910.]
- Neiman, P. J., F. M. Ralph, A. B. White, D. E. Kingsmill, and P. O. G. Persson, 2002: The statistical relationship between upslope flow and rainfall in California's coastal mountains: Observations during CALJET. *Mon. Wea. Rev.*, **130**, 1468–1492.
- , —, G. A. Wick, Y.-H. Kuo, T.-K. Wee, Z. Ma, G. H. Taylor, and M. D. Dettinger, 2008a: 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.
- , —, —, J. D. Lundquist, and M. D. Dettinger, 2008b: 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.
- , A. B. White, F. M. Ralph, D. J. Gottas, and S. I. Gutman, 2009: A water vapor flux tool for precipitation forecasting. *J. Water Manage.*, **162**, 83–94.
- Olson, D. A., N. W. Junker, and B. Korty, 1995: Evaluation of 33 years of quantitative precipitation forecasting at the NMC. *Wea. Forecasting*, **10**, 498–511.
- Pandey, G. R., D. R. Cayan, and K. P. Georgakakos, 1999: Precipitation structure in the Sierra Nevada of California during winter. *J. Geophys. Res.*, **104**, 12 019–12 030.
- Ralph, F. M., P. J. Neiman, D. E. Kingsmill, P. O. G. Persson, A. B. White, E. T. Strem, E. D. Andrews, and R. C. Antweiler, 2003: The impact of a prominent rain shadow on flooding in California's Santa Cruz Mountains: A CALJET case study and sensitivity to the ENSO cycle. *J. Hydrometeorol.*, **4**, 1243–1264.
- , —, and G. A. Wick, 2004: Satellite and CALJET aircraft observations of atmospheric rivers over the eastern North Pacific Ocean during the El Niño winter of 1997/98. *Mon. Wea. Rev.*, **132**, 1721–1745.
- , —, and R. Rotunno, 2005a: Dropsonde observations in low-level jets over the Northeastern Pacific Ocean from CALJET-1998 and PACJET-2001: Mean vertical-profile and atmospheric-river characteristics. *Mon. Wea. Rev.*, **133**, 889–910.
- , and Coauthors, 2005b: Improving short-term (0–48 h) cool-season quantitative precipitation forecasting: Recommendations from a USWRP Workshop. *Bull. Amer. Meteor. Soc.*, **86**, 1619–1632.
- , P. J. Neiman, G. A. Wick, 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, doi:10.1029/2006GL026689.
- Reynolds, D., 2003: Value-added quantitative precipitation forecasts: How valuable is the forecaster? *Bull. Amer. Meteor. Soc.*, **84**, 876–878.
- Smith, R. B., 1979: The influence of mountains on the atmosphere. *Advances in Geophysics*, Vol. 21, Academic Press, 87–230.
- Yuan, H., C. Lu, J. A. McGinley, P. J. Schultz, B. D. Jamison, L. Wharton, and C. J. Anderson, 2009: Evaluation of short-range quantitative precipitation forecasts from a time-lagged multimodel ensemble. *Wea. Forecasting*, **24**, 18–38.
- Zhu, Y., and R. E. Newell, 1998: A proposed algorithm for moisture fluxes from atmospheric rivers. *Mon. Wea. Rev.*, **126**, 725–735.