

Comparative Verification of Recent Quantitative Precipitation Forecasts in the National Weather Service: A Simple Approach for Scoring Forecast Accuracy

JEROME P. CHARBA

Meteorological Development Laboratory, Office of Science and Technology, National Weather Service, NOAA, Silver Spring, Maryland

DAVID W. REYNOLDS*

Hydrometeorological Prediction Center, National Centers for Environmental Prediction, National Weather Service, NOAA, Camp Springs, Maryland

BRETT E. McDONALD⁺

Office of Meteorology, National Weather Service, NOAA, Silver Spring, Maryland

GARY M. CARTER[#]

Scientific Services, Eastern Region Headquarters, National Weather Service, NOAA, Bohemia, New York

(Manuscript received 13 June 2002, in final form 29 October 2002)

ABSTRACT

Comparative verification of operational 6-h quantitative precipitation forecast (QPF) products used for streamflow models run at National Weather Service (NWS) River Forecast Centers (RFCs) is presented. The QPF products include 1) national guidance produced by operational numerical weather prediction (NWP) models run at the National Centers for Environmental Prediction (NCEP), 2) guidance produced by forecasters at the Hydrometeorological Prediction Center (HPC) of NCEP for the conterminous United States, 3) local forecasts produced by forecasters at NWS Weather Forecast Offices (WFOs), and 4) the final QPF product for multi-WFO areas prepared by forecasters at RFCs. A major component of the study was development of a simple scoring methodology to indicate the relative accuracy of the various QPF products for NWS managers and possibly hydrologic users. The method is based on mean absolute error (MAE) and bias scores for continuous precipitation amounts grouped into mutually exclusive intervals. The grouping (stratification) was conducted on the basis of observed precipitation, which is customary, and also forecast precipitation. For ranking overall accuracy of each QPF product, the MAE for the two stratifications was objectively combined. The combined MAE could be particularly useful when the accuracy rankings for the individual stratifications are not consistent. MAE and bias scores from the comparative verification of 6-h QPF products during the 1998/99 cool season in the eastern United States for day 1 (0–24-h period) indicated that the HPC guidance performed slightly better than corresponding products issued by WFOs and RFCs. Nevertheless, the HPC product was only marginally better than the best-performing NCEP NWP model for QPF in the eastern United States, the Aviation (AVN) Model. In the western United States during the 1999/2000 cool season, the WFOs improved on the HPC guidance for day 1 but not for day 2 or day 3 (24–48- and 48–72-h periods, respectively). Also, both of these human QPF products improved on the AVN Model on day 1, but by day 3 neither did. These findings contributed to changes in the NWS QPF process for hydrologic model input.

1. Introduction

By the mid to late 1990s, quantitative precipitation forecasts (QPFs) were being ingested into streamflow

models at all 13 National Weather Service (NWS) River Forecast Centers (RFCs) across the conterminous United States and Alaska (Fig. 1a) (National Weather Service 1999).¹ The NWS process by which the ingested QPF product was prepared (henceforth termed the QPF process) was complex. It involved a series of QPF products, including output from numerical weather prediction (NWP) models run centrally at the National Centers for Environmental Prediction (NCEP), products issued for the conterminous United States (CONUS) by forecasters at NCEP's Hydrometeorological Prediction Cen-

* Current affiliation: WFO Monterey, National Weather Service, NOAA, Monterey, California.

⁺ Current affiliation: WFO Riverton, National Weather Service, NOAA, Riverton, Wyoming.

[#] Current affiliation: Office of Hydrologic Development, National Weather Service, NOAA, Silver Spring, Maryland.

Corresponding author address: Dr. Jerome P. Charba, W/OST21, Meteorological Development Laboratory, 1325 East West Hwy., Rm. 10410, Silver Spring, MD 20910.
E-mail: Jerome.Charba@noaa.gov

¹ In the western United States, many local NWS forecast offices have been issuing QPFs for RFC hydrologic applications for over 30 years.

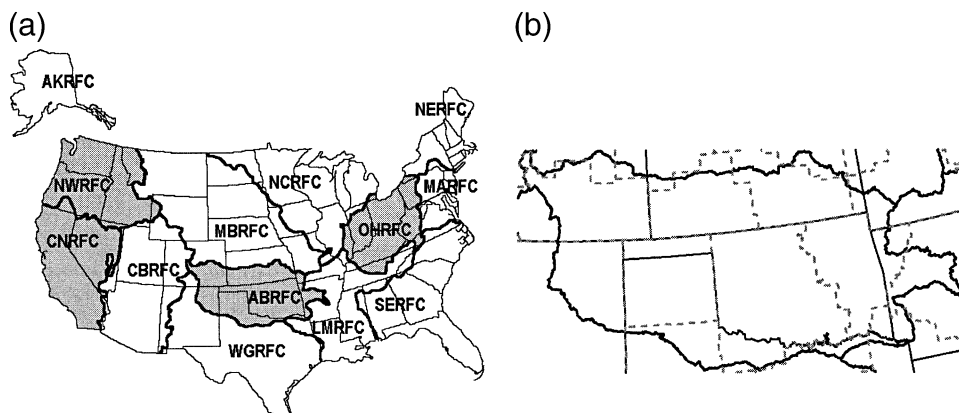


FIG. 1. (a) RFC service areas over CONUS and Alaska. In the RFC call letters, AK stands for Alaska, NW for Northwest, CN for California–Nevada, CB for Colorado Basin, MB for Missouri Basin, NC for North Central, AB for Arkansas–Red Basin, WG for West Gulf, LM for Lower Mississippi, SE for Southeast, OH for Ohio, MA for Mid-Atlantic, and NE for Northeast. The study was conducted for RFCs highlighted with shading. (b) WFO HSAs (light dashed lines) within the ABRFC area (heavy solid line).

ter (HPC), a product issued by forecasters at Weather Forecast Offices (WFOs) for the hydrologic service area (HSA; Fig. 1b) of each, and a final modified WFO QPF prepared by forecasters/hydrologists at RFCs (Fig. 1a). (Although each QPF product was intended to serve as guidance for the subsequent one, it will be shown later that they often did not agree with one another.)

In March of 1999, a team of NWS meteorologists and hydrologists (see acknowledgments for team composition) was commissioned to study the existing QPF process to assess its overall effectiveness. A glaring weakness in the QPF process was that the inherent QPF products were not being verified, at least not in a comparative sense. Thus, the team undertook a study to formulate and conduct an objective comparative verification—the subject of this article. One facet of this pioneering effort involved wrestling with a number of challenging “data problems” that included short historical samples and incomplete geographical coverage of some QPF products, QPF and verification data of diverse types and purposes, and disparate data archiving formats. Because of limitations on the verification study imposed by these data problems, the formulation of conclusions and recommendations regarding the QPF process had to rely in part on supplemental information obtained from a questionnaire survey of the WFOs and RFCs (noted in section 8).

An even greater challenge involved formulating a verification methodology to measure the potential benefit of each QPF product for RFC streamflow models. These models currently incorporate precipitation (both observed and forecast) in standard 6-h periods to model drainage-basin runoff over those periods (Burnash 1995). Because runoff usually increases rapidly with increasing 6-h precipitation, a key requirement for the design of the verification methodology was for it to be especially sensitive to forecaster skill in predicting heavy 6-h amounts.

In most previous QPF verification studies, forecast skill was measured by first converting precipitation expressed as continuous amounts into “exceedance” categories (yes–no statements indicating whether precipitation equals or exceeds selected threshold amounts) and then computing performance measures for each threshold (e.g., Bosart 1980; Charba and Klein 1980; Gyakum and Samuels 1987; Olson et al. 1995; Mesinger 1996). Such a scoring approach is most useful when threshold amounts that would result in a hazardous event (such as flooding) are predetermined. In current RFC operations, hydrologic models compute the volume of precipitation runoff over drainage basins such that the stage (height) of flow at specific points along the stream (river) is determined (Fread et al. 1995). Because these models ingest QPFs as continuous precipitation amounts, the forecasts should be scored accordingly.

Continuous QPF scoring was conducted in recent studies by Colle et al. (1999, 2000), wherein exceedance intervals were used to stratify performance scores by precipitation amount. However, this kind of stratification might not capture the richness of the score-versus-precipitation-amount relationship, because most precipitation intervals are broad. Also, the basis of the stratification was observed precipitation. Because forecast precipitation is also used in streamflow models, an additional stratification based on forecast precipitation should provide beneficial information to hydrologic users. A complicating factor occurs when scores for the two stratifications are inconsistent—a challenge that had to be addressed in this study.

This QPF verification study includes several extensions to similar studies in the formal literature. One is that the comparative scoring included all operational QPF products in the NWS (national as well as local; sections 2, 3, and 4), which is unprecedented. Because these QPF products and the verification data had diverse forms and formats, extensive postprocessing was re-

quired to achieve needed standardization (section 5). The verification methodology (section 6), whose purpose was to rank the accuracy of the various products, incorporates a few techniques not previously applied for QPFs, and it shares a few components of the generalized verification methods of Murphy and Winkler (1987) and Murphy et al. (1989). The results obtained from application of the methodology to two recent historical samples are given in section 7, section 8 contains the findings and operational ramifications in the NWS, and a summary is provided in section 9.

2. Operational QPF products

The 6-h QPF products that compose the QPF process begin with national guidance generated by operational numerical and statistical models run at NCEP. NCEP NWP model QPFs generated by the Nested Grid Model (NGM; Hoke et al. 1989), the Aviation (AVN) Model run of the Global Spectral Model (Kanamitsu et al. 1991), and the "early" Eta Model (Black 1994) were all included in this study (at least in the early stages). [See Mesinger (1996) for a concise description of this suite of NCEP NWP models.] Also, included initially were QPF products generated by an NGM model output statistics (MOS) model (Antolik 2000), which was developed at the NWS Meteorological Development Laboratory (MDL).²

Three manually produced products composed the balance of the QPF process at the time of this study. One of these is the QPF graphic issued for the CONUS at HPC (Olson et al. 1995). A second product was a composite of local QPFs from 119 WFOs over the CONUS, where a local forecast applies to the WFO's HSA (Fig. 1b). The form of the WFO QPFs was graphical in the eastern United States and alphanumeric station/point data in the western United States. The third product was a graphic generated by hydrometeorological analysis and support (HAS) forecasters (Fread et al. 1995) at RFCs. In the western United States, the role of HAS forecasters was only to monitor the shortest-range WFO QPFs for consistency with the most recent precipitation observations and to request WFO updates when inconsistencies appeared. Thus, a separate RFC product was not involved in the comparative verification in the West.

The three-tier human component of the QPF process was designed to capitalize on the expertise and experience of forecasters within each layer. The design assumed that 1) forecasters at HPC have special expertise in the use of numerical and statistical QPF guidance produced by collocated computer models, 2) WFO line forecasters have in-depth knowledge of local climatic

effects that should lead to finescale improvement of the national HPC guidance, and 3) HAS forecasters have a working understanding of the sensitivity of streamflow models to precipitation, which should translate to appropriate adjustment of the QPF product from WFOs.

3. Verification data

Two kinds of precipitation data were used for verifying the QPF products. In the eastern United States, we used the Stage III (precipitation) Analysis (Fread et al. 1995; Fulton et al. 1998), which is now being used extensively at RFCs as antecedent precipitation input into streamflow models (National Weather Service 1999). This hourly precipitation analysis involves automated processing of radar-estimated precipitation from the modern Weather Surveillance Radar-1988 Doppler (WSR-88D) network (Fread et al. 1995) together with supplemental gauge measurements and interactive quality control by RFC HAS forecasters. A positive attribute of this data type is its high spatial resolution: the Hydrologic Rainfall Analysis Project (HRAP) grid has a mesh length of about 4 km at mid-latitudes. West of the U.S. Continental Divide, radar-estimated precipitation is not reliable because of extensive radar-beam blockage by the mountainous terrain (Westrick et al. 1999). Therefore, conventional gauge measurements were used in this region.

4. Verification domain

The study was limited to the cool season of the year for two reasons. First, because of the great difficulty in accurately forecasting precipitation amount during any time of the year, it was believed that verification results would be more meaningful during the cool season when such forecasting is somewhat less difficult (e.g., Charba and Klein 1980; Olson et al. 1995). Second, the vast majority of the normal annual precipitation in the western United States occurs during the cool season (Charba et al. 1998; Groisman et al. 2001). The incidence of frozen precipitation (especially snow) in winter results in error in both the Stage III Analysis and in the gauge measurements [see Fulton et al. (1998); Colle et al. (2000), and the references therein], but the adverse impact on the verification should be the same for all QPF products and thus of little importance in this study.

Several aspects of the verification domain, including the historical sampling periods, the RFC service areas, and forecast-validity periods, are summarized in Table 1. Two RFC areas were selected to represent the eastern United States for the period-I (1 October 1998–1 March 1999) verification. They were the Arkansas–Red Basin RFC (ABRFC) and the Ohio RFC (OHRFC) (Fig. 1a). ABRFC and OHRFC were selected largely because of the availability of archives of the Stage III Analysis and WFO and RFC QPF products. In addition, the precipitation regime for these two geographical regions is be-

² An additional QPF product from an MDL statistically based model, initially included in the study, was the Local Advanced Weather Interactive Processing System (AWIPS) MOS Program (LAMP) QPF product (Charba 1998). This gridded product was being produced in an experimental mode at the time of the study.

TABLE 1. Domains for the two sampling periods. The RFC areas are defined in the text. Four 6-h QPFs that span a 24-h period compose day 1 for period I; these 6-h QPFs were extended to compose day 2 and day 3 during period II. The day-1, day-2, and day-3 forecast ranges are 12–36, 36–60, and 60–84 h, respectively, after the 0000 UTC model cycle time (see text).

	RFC areas	Historical period	6-h QPF temporal coverage
Period I	ABRFC, OHRFC	1 Oct 1998–31 Mar 1999	Day 1
Period II	CNRFC, NWRFC	1 Nov 1999–31 Mar 2000	Days 1–3

lieved to be representative of much of the eastern United States during the cool season, because both areas are located near a principal midwinter cyclone storm track (Reitan 1974). A second verification period (period II), which spanned 1 November 1999–31 March 2000, involved the western United States (Table 1) because it was found that the available WFO QPF archives for period I in this region contained updated forecasts. (This circumstance would have resulted in an unfair advantage for the WFOs because the updates were issued within the period of validity of the forecasts.) To represent the western United States, we selected two of the three RFCs located west of the Continental Divide (Fig. 1a), which were California–Nevada RFC (CNRFC) and Northwest RFC (NWRFC). These centers were chosen because of the availability of archives of the WFO QPF products and precipitation gauge data from a relatively dense network (illustrated for CNRFC in Fig. 2). Further, these areas receive most of their high annual precipitation during the cool season, as noted above.

All 6-h QPF products included in the study were based on the 0000 UTC (NWP model) cycle for both period I and period II. For period I, the QPFs, which span four consecutive 6-h valid periods and project 12–36 h from 0000 UTC, are denoted “day 1” in Table 1 and Fig. 3. Note from Fig. 3 that the WFO and RFC

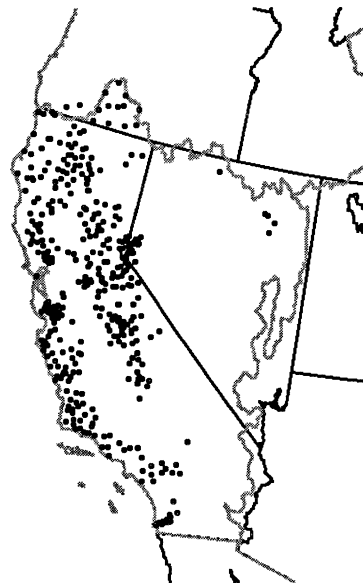


FIG. 2. Precipitation gauge network (dots) for the CNRFC service area (light boundary).

QPFs could have benefited from later observations, because the issuance time was later than that for the NWP models and HPC. For period II, the temporal coverage of the 6-h QPF products was extended to include two subsequent 24-h periods (denoted “day 2” and “day 3” in Table 1). It is noted that, although the forecast lead times increase for the individual 6-h periods composing each “day” and thus the corresponding verification scores should degrade (especially for day 1), for brevity and because of the shortness of the sampling periods, verification data within each day were combined during scoring. Also, an additional HPC QPF product valid for a 24-h period at the day-1 range (Olson et al. 1995) was also involved in this study. This product was not included in the comparative evaluation of concern, but it was used for formulating the verification methodology (section 6).

5. Preprocessing the QPF products and verification data

Because each of the operational QPF products and verification data types was unique with regard to form (graphic, grid, or points) and spatial resolution, extensive processing was required to achieve necessary standardization for the verification. Also, it is noted that, because this study commenced in March of 1999, some of the QPF products have changed. The following brief discussion of preprocessing is applicable to the operational products in use at that time.

a. Eastern United States: Period-I verification

For the two eastern RFC areas, the form of the QPF products and verification data exhibited high uniformity.

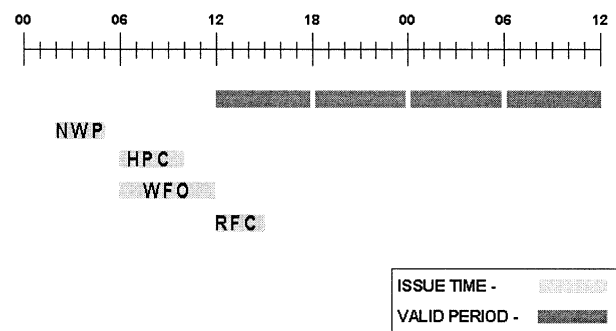


FIG. 3. QPF product issuers (acronyms defined in text) and issue times (light shading) relative to four 6-h valid periods (dark shading) composing the day-1 forecast period. Timescale is UTC.

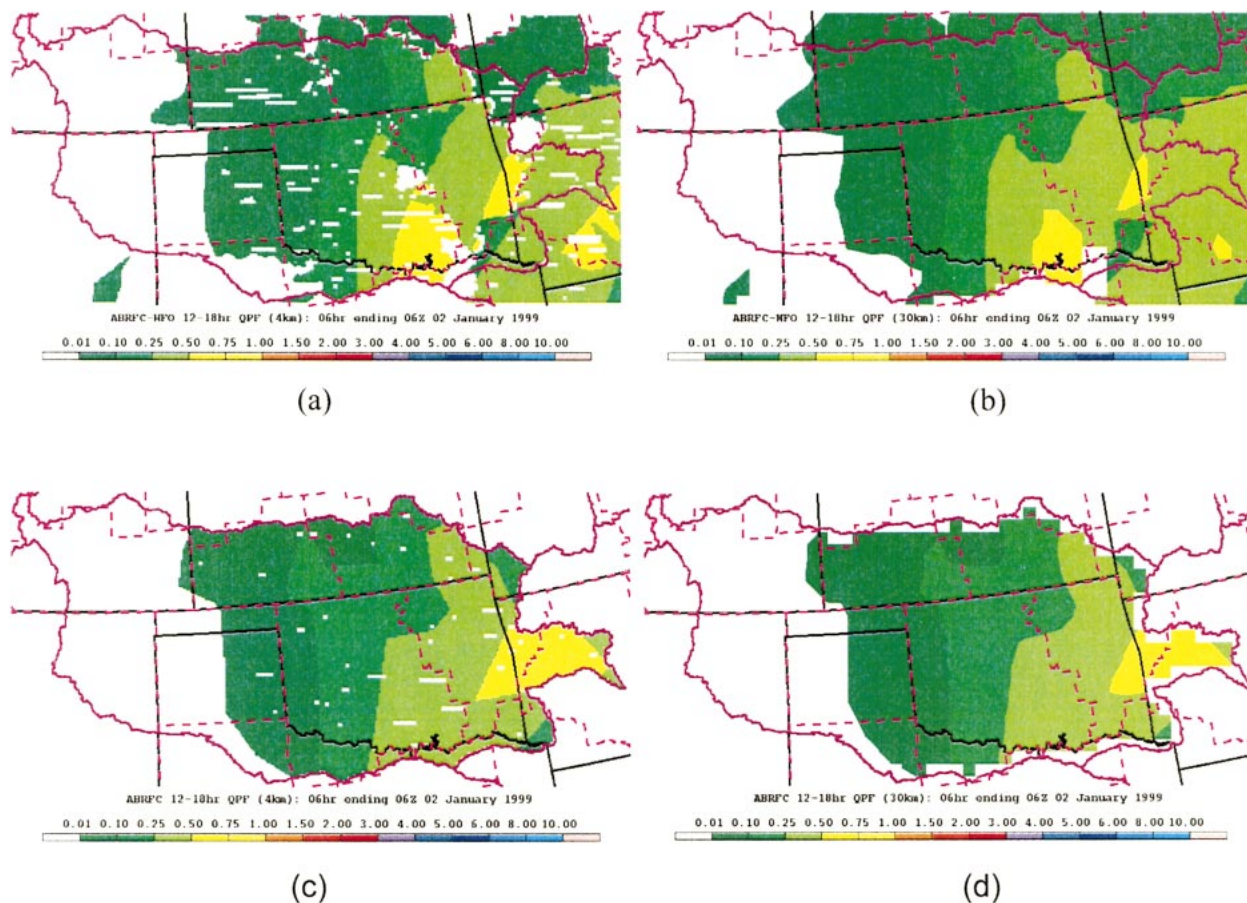


FIG. 4. WFO QPF product (in.) for ABRFC area (solid purple boundary) and the 6-h period ending 0600 UTC 2 Jan 1999 on (a) 4- and (b) 30-km grid. Corresponding RFC QPF product on (c) 4- and (d) 30-km grid. WFO HSA boundaries are superimposed (dashed lines). Clear streaks in (a) and (c) reflect missing gridpoint values.

Thus, the required preprocessing involved only some problems in the gridding of the graphical HPC QPF products and diverse grid meshes in which the various products were available. To be specific, each of the QPF products (except the NGM MOS QPF) and the verification data were represented on a grid. [These gridded QPF data were treated as representing spatially averaged precipitation in this study, because the human-generated QPF products are defined as such (National Weather Service 1999) and most modelers agree that NWS operational model-generated precipitation should be treated as areal rather than point values.] The choice of conducting the comparative verification on a polar stereographic grid with a standard longitude of 105°W was obvious, because all QPF products and the verification data were archived with a grid of this type. The choice of the grid mesh of 31.75 km at 60°N (about 30 km at midlatitudes) was driven largely by the facts that this grid had been in use for verification at NCEP since 1984 (Olson et al. 1995) and that the HPC and NWP model products were archived on it. Further, although the WFO and RFC QPF products (as well as the stage-III verification data) were represented on the 4-km HRAP grid,

visual inspection of many cases revealed no evidence that the finest-scale features in the QPF maps approached this spatial resolution (see Fig. 4 for an example case). Thus, it seemed safe to assume that rendering these fields on the 30-km grid would not significantly degrade the finest spatial scales.

Some properties of the graphical HPC 6- and 24-h QPF products presented limitations in rendering this product on the 30-km verification grid. In particular, the manually drawn isohyets composing these charts (at the time of this study) began with 0.25 in. (Fig. 5); that is, precipitation under 0.25 in. was not forecast. Thus, in the automated interpolation from the isohyetal field to the 30-km grid [see Ruth (1992) for the method], precipitation amounts under 0.25 in. were “bogused” by adding (prior to the interpolation) a fictitious 0.00-in. contour just outside the CONUS border. Thus, it appeared imperative that the bogused data values in the HPC QPF product should be excluded from the verification (or at least their number minimized; see section 6c). Another gridding problem involved forecaster annotations of the “maximum point amount” within the heaviest isohyets (see Fig. 5). Because an appropriate

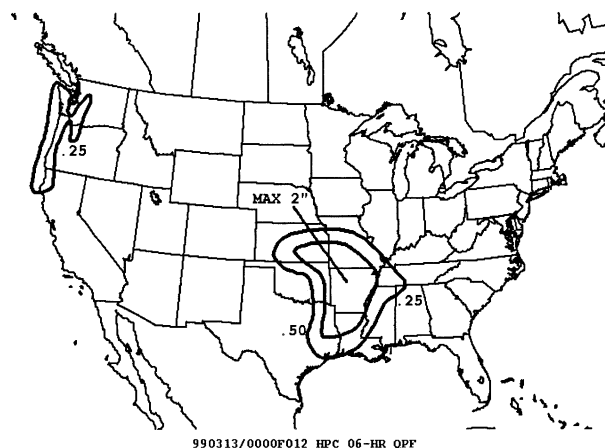


FIG. 5. Example HPC manual 6-h QPF graphic applicable to the period-I verification.

method for incorporating these localized peak amounts in the graph-to-grid interpolation was not apparent, they were excluded. Evidence of an adverse impact of this interpolation limitation on the HPC scores for the period-I verification is noted in the next section. HPC QPF graphics were upgraded for the period-II verification: forecasters were directed to add the zero-precipitation isohyet and to incorporate local peak amounts into the isohyetal pattern. Thus, both of the HPC scoring limitations affecting the period-I verification in the eastern United States were averted in the period-II verification in the West.³

The operational inauguration of the WFO and RFC QPF products in support of the NWS QPF process in the eastern United States occurred in the mid to late 1990s (National Weather Service 1999). The first step in the WFO product preparation consisted of the forecaster drawing isohyets of 6-h mean areal precipitation for the WFO HSA (Fig. 1b). Next, this localized contour map was rendered on a subset of the 4-km HRAP grid using automated graph-to-grid interpolation (Fenbers 1995). Such WFO subgrids, transmitted to the associated RFC and composited with other WFO subgrids, formed the WFO QPF product used in this study (Fig. 4a). Note the spatial discontinuities in the composite QPF field, which occur at neighboring WFO HSA boundaries. The corresponding RFC QPF product was also generated through application of interactive software (Fenbers 1993), whereby the RFC HAS forecaster drew the QPF isohyetal field with the WFO composite as an underlay field. The HAS forecaster typically modified the composite by correcting for perceived QPF error and removing the characteristic spatial disconti-

³ Some readers may question whether the HPC QPF product should have been included in the period-I comparative verification given the scoring limitations involved. Although this doubt is valid, the product was retained because of its status as an integral part of the NWS QPF process. Besides, it was found to score well (in a relative sense).

nities at WFO boundaries. The process was completed as the new RFC map was also automatically rendered on the 4-km grid. The RFC product, corresponding to the WFO product in Fig. 4a, is shown in Fig. 4c. Note that spatial discontinuities seen in the WFO product are absent in the latter.⁴

The verification preprocessing required for the WFO and RFC QPF products (as well as the stage-III verification data) involved transposing from the 4-km grid to the 30-km grid.⁵ This transposition was done by averaging the fine-mesh gridpoint values within 30-km grid boxes centered on the coarse mesh grid points. Figures 4b and 4d show that this rerendering of the respective WFO and RFC products did not significantly degrade their spatial resolution. Figures 6a and 6b show that the corresponding transposition of the stage-III data makes the spatial resolution of the verification data more consistent with the resolution of these (and the other) QPF products.

b. Western United States: Period-II verification

A mix of point and gridded QPF product and verification data resulted in added complexity to the period-II comparative verification for the two western RFC areas (Table 1). As for the East, the NWP model and HPC products, which were on the 30-km grid, are assumed to represent spatially averaged precipitation. The verification data and WFO QPFs, on the other hand, pertain to specific points (Figs. 2 and 7, respectively), and, moreover, the WFO forecast points are a subset of the verification points. Especially considering the complex mountainous terrain composing the CNRFC and NWRFC areas, the inconsistency in scales represented in the spatially averaged QPF and point verification data will have an adverse impact on scores for verifying NWP model and HPC QPF products (Cherubini et al. 2002).

Another factor relevant to the western verification is that different preprocessing procedures between the CNRFC and NWRFC were in operation for the point (WFO) QPF and precipitation gauge data prior to ingestion into locally run streamflow models. For CNRFC, composited point QPFs from 10 WFOs within the service area (Fig. 7) were objectively “distributed” onto the 4-km HRAP grid. The precipitation distribution

⁴ Although this process could still result in QPF inconsistencies at neighboring RFC boundaries, such inconsistencies should have a lesser negative hydrologic impact than those between neighboring WFO HSAs. The reason is the boundaries separating RFC service areas were drawn (long ago) such that splitting of major watersheds is minimized.

⁵ Although the QPF and observed precipitation inputs to RFC streamflow models in the eastern United States are based on the 4-km grids, the specific inputs consist of mean areal precipitation (MAP) for predefined stream subbasins. The MAP products, which are obtained by averaging the 4-km grid values within the subbasins, were not included in this verification study.

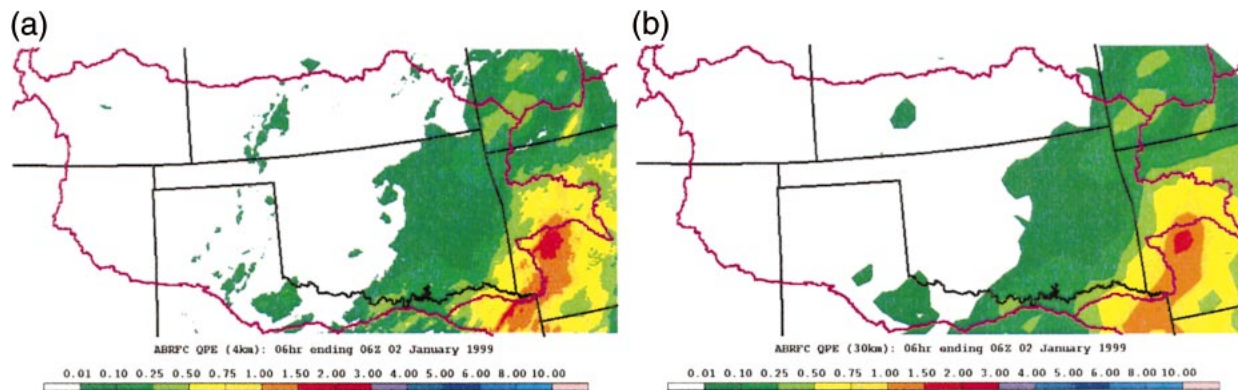


FIG. 6. (a) Stage III (hourly precipitation) Analysis (4-km grid; in.) for ABRFC area accumulated over the 6-h period ending 0600 UTC 2 Jan 1999, and (b) the same data field rendered on the 30-km grid.

model, called Mountain Mapper (MM; Henkel and Peterson 1996), preserves a point QPF value at the grid point closest to it and applies modeled climatic precipitation data, called the Precipitation–Elevation Regression on Independent Slopes Model (PRISM; Daly et al. 1994), in the QPF assignment to all other grid points. *Because an identical gridding procedure was applied to the CNRFC gauge data used in the verification (Fig. 2), an inherent correlation with the gridded WFO QPF product arises.*⁶ In operations at NWRFC, on the other

⁶ As for eastern RFCs, both the QPF and observed precipitation inputs to streamflow models at CNRFC consisted of corresponding MAP values derived from the corresponding 4-km grids, neither of which was used in this study.

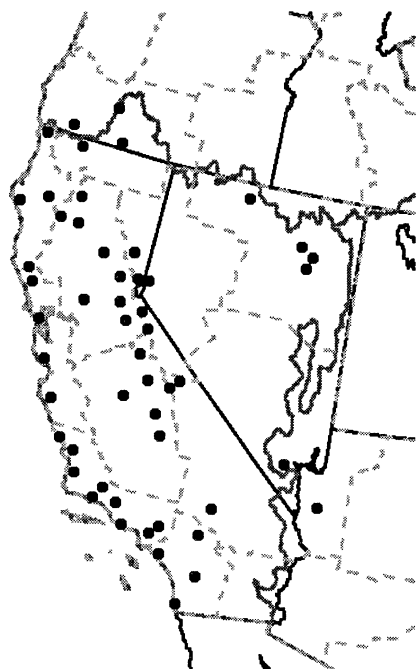


FIG. 7. WFO forecast points (dots) and WFO HSA boundaries (dashed lines) within the CNRFC area (light solid boundary). The WFO forecast points are a subset of the verification points in Fig. 2.

hand, the MM gridding procedure was not used for preparing the QPF (and observed precipitation data) for streamflow model input. Instead, the point WFO QPFs and observed precipitation data are ingested directly into the streamflow model.

To address scoring inequity concerns between the point WFO QPFs and the gridded NWP model and HPC QPF products in the CNRFC verification, three scoring approaches (summarized in Table 2) were applied. The most direct approach (approach 1) was to verify all QPF products at the WFO forecast points (Fig. 7) using the corresponding point (gauge) data for verification. In this approach, the gridded NWP model and HPC QPFs were interpolated to the WFO points. In approach 2, all QPF products and the verifying data were in gridded form. The gridding of the point WFO QPF and verifying data resulted from application of the MM model, whereas the NWP model and HPC QPFs were in their original (“raw”) gridded forms. Approach 3 for the WFO and verifying data was identical to approach 2, whereas for the NWP model and HPC products new “MM grids” were derived. These new MM grids were obtained by first interpolating from the original 30-km grids to the WFO points and subsequently applying the MM model to obtain 4-km grids. Note that in approaches 2 and 3, the 4-km grids were coarsened to obtain 30-km grids. As indicated in Table 2, only approach 1 was applied in the NWRFC verification because the MM gridding technique was not used in operations at this RFC when this study was conducted.

Several aspects of the three verification approaches are noteworthy. The approach-1 method results in a negative impact on the NWP model and HPC scores because the forms of the spatially averaged QPFs and (point) verification data are inconsistent. In approach 2, the QPF form inconsistency is ostensibly removed, because all products and verifying data represent areal precipitation. In actual fact, however, the QPF product inconsistency inherent to approach 1 is not entirely eliminated because the point-specific WFO QPF and verification data are conserved in the 4-km MM grids. Also, the inherent

TABLE 2. Forms (station points or gridded) of the QPF products and verification data embodied in three verification approaches used for two western RFCs. The preprocessing (where needed) associated with these data forms is indicated in parentheses; “interpolated” denotes bilinear interpolation from the 30-km grid and “MM” stands for application of the Mountain-Mapper model. For the WFO QPF and verification data, approach 3 is identical to approach 2, and this approach differs somewhat from approach 3 for the NWP model and HPC QPFs (see text).

RFC	Approach	WFO QPF	NWP model and HPC QPF	Verification data
CNRFC	1	Station points	Station points (interpolated)	Station points
	2	Gridded (MM)	Gridded (raw)	Gridded (MM)
	3	Gridded (MM)	Gridded (MM)	Gridded (MM)
NWRFC	1	Station points	Station points (interpolated)	Station points

correlation involving the PRISM climatological values (“climatology factor”) that artificially benefits the WFO scoring is not involved in the scoring of the model and HPC products. In approach 3, in which all products are based on MM grids, the climatology factor benefits all products equally, and so this approach should be the most equitable among the three. However, the point specificity in the verification data remains as a problem for the model and HPC products.

6. Formulation of a scoring methodology

Scoring techniques for QPFs, which address some requirements for the RFC user, are discussed in this section. In streamflow models, the volume of precipitation over drainage basins is used to determine runoff (Burnash 1995). Because runoff can increase rapidly with increasing precipitation amount, an important scoring requirement is that performance measures be stratified (grouped) by precipitation amount. Such scoring stratification is best examined on the basis of a historical sample that contains a broad range of QPF and observed precipitation amounts. Because the ranges of QPF and observed precipitation amounts are much greater for 24-h periods than for the 6-h periods for which this comparative verification is directed, the former period was used in formulating the scoring methodology. Another factor in this choice was that the length of the available sample of the 6-h QPF product for HPC was short for the period-I sample (covering only 1 January–31 March 1999), whereas the corresponding HPC 24-h product was available for the full 1 October 1998–31 March 1999 period (Table 1). Corresponding 24-h QPF products for the NWP models, WFOs, and RFCs were based on summations of the 6-h products.

a. Scoring measures

Scoring measures that describe the accuracy of (or error in) spatially averaged QPFs in continuous form essentially quantify error in the volume of predicted precipitation. Analogous accuracy measures for continuous QPFs at points measure precipitation depth error at those points. Appropriate scores are the mean absolute error (MAE) and the root-mean-square error (rmse; Wilks 1995), which are defined as

$$\text{MAE} = \frac{\sum_{i=1}^N |(F_i - O_i)|}{N} \quad \text{and} \quad (1)$$

$$\text{rmse} = \left[\frac{\sum_{i=1}^N (F_i - O_i)^2}{N} \right]^{1/2}, \quad (2)$$

respectively. In these equations, F_i and O_i refer to forecast and observed precipitation amount, respectively, for point i (alternatively the center point of a verification grid box), and the summation is over the number of cases N in the sample. Inspection of (1) and (2) reveals that both scores have values of 0 when the QPF absolute error $|F_i - O_i|$ is 0 for all i , and they have positive values otherwise. Note also that each score (especially the rmse because the error is squared) is sensitive to large errors in a sample—even a few of them.

Because RFC streamflow models conserve water mass (in space and time) over hydrologic basins (Burnash 1995), mass balance between forecast and observed precipitation is important. The degree of mass balance is measured by the bias (Wilks 1995), which is defined as

$$\text{bias} = \frac{\sum_{i=1}^N F_i}{\sum_{i=1}^N O_i}, \quad (3)$$

where the notation is the same as before. Note that the bias for a sample is always positive (or 0) and that forecasts with no bias have a value of 1.0 (hereinafter called perfect bias). Note further that QPFs with little bias can have considerable utility in streamflow models even though their accuracy might not be high. Note from (1), (2), and (3), however, that QPFs with low (high) error are more likely to have a good (poor) bias than for the reverse situation.

b. Score stratification according to precipitation amount

In some previous literature articles (e.g., Murphy and Winkler 1987; Murphy et al. 1989; Brooks and Doswell 1996), forecast performance scores, such as those defined in the previous section, have been called summary measures. This terminology reflects the common appli-

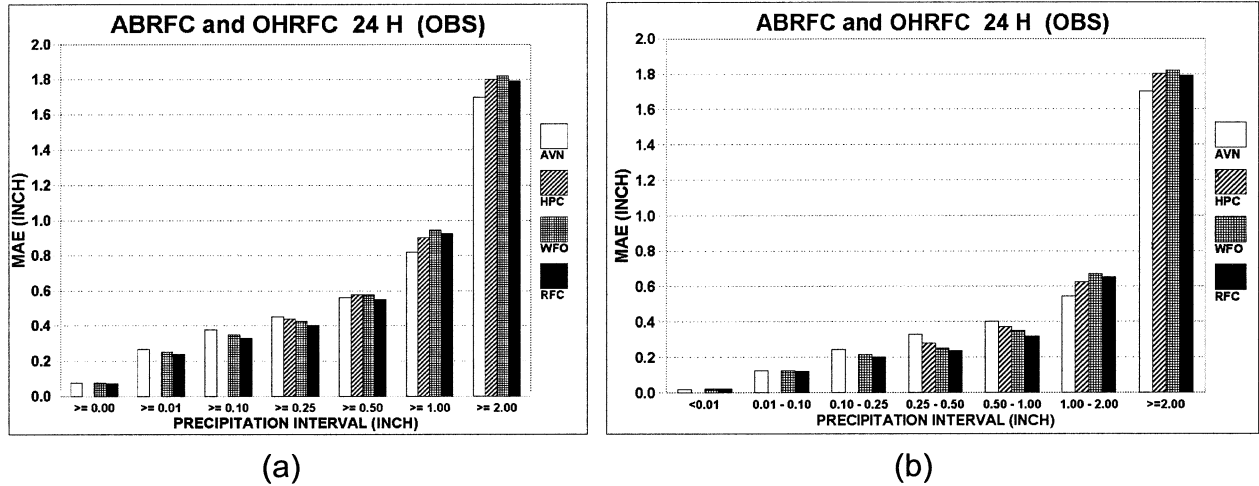
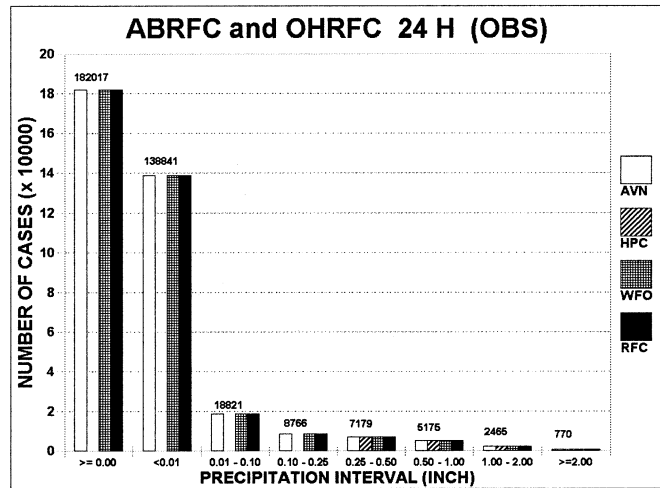


FIG. 8. MAE for 24-h AVN, HPC, WFO, and RFC QPF products vs observed precipitation amount (denoted OBS in title) on the basis of (a) exceedance intervals and (b) mutually exclusive (ME) intervals. (c) The subsample sizes corresponding to (b). The underlying sample is from 1 Oct 1998 to 31 Mar 1999 (period I) and the ABRFC and OHRFC areas combined. HPC is not shown for intervals under 0.25 in. because HPC did not predict precipitation amounts in this range during period I (see text).

(c)



cation of these scores to the full available sample of the forecasts and observations, and, thus, a single score value summarizes forecast accuracy. Such scores were applied in this study, but their stratification by precipitation amount provides considerable additional information. For example, Fig. 8a is a plot of the MAE for 24-h QPFs from the AVN Model, HPC, WFOs, and RFCs based on the two eastern RFCs (Fig. 1a) during period I (Table 1). In this plot, the MAEs are stratified by observed precipitation amount on the basis of exceedance intervals, following the stratification technique used by Colle et al. (1999, 2000) among others. As expected, the MAEs rise with increasing precipitation amount—gradually for most threshold amounts and abruptly for the highest threshold (≥ 2.00 in.). Of course, with exceedance intervals, the subsample over which each MAE value is computed overlaps to a degree with the subsample for every other MAE value. The individual MAE values are consequently partially redundant. To address this score overlap ambiguity, we ap-

plied mutually exclusive precipitation intervals (henceforth called ME intervals).

Figure 8b is identical to Fig. 8a except that ME intervals were used. Note that the MAE rise with increasing precipitation amount is now more gradual for all but the heaviest precipitation interval; of course, the MAE for the highest interval is identical in both plots. Thus, because of the slower initial rise in MAE, Fig. 8b also exhibits a sharper increase in QPF error with the heaviest precipitation interval. In essence, the difference in shapes of the plots between Figs. 8a and 8b occurs because the small ME intervals in the latter sharpen the specificity of the QPF error as a function of precipitation amount. Because this method of specifying error as a function of precipitation amount should provide additional information to both forecasters and hydrologic users, ME intervals were adopted for score stratification in this study. The only known previous use of this method of QPF error stratification was in a conference preprint article (Schultz and Snook 1996) and

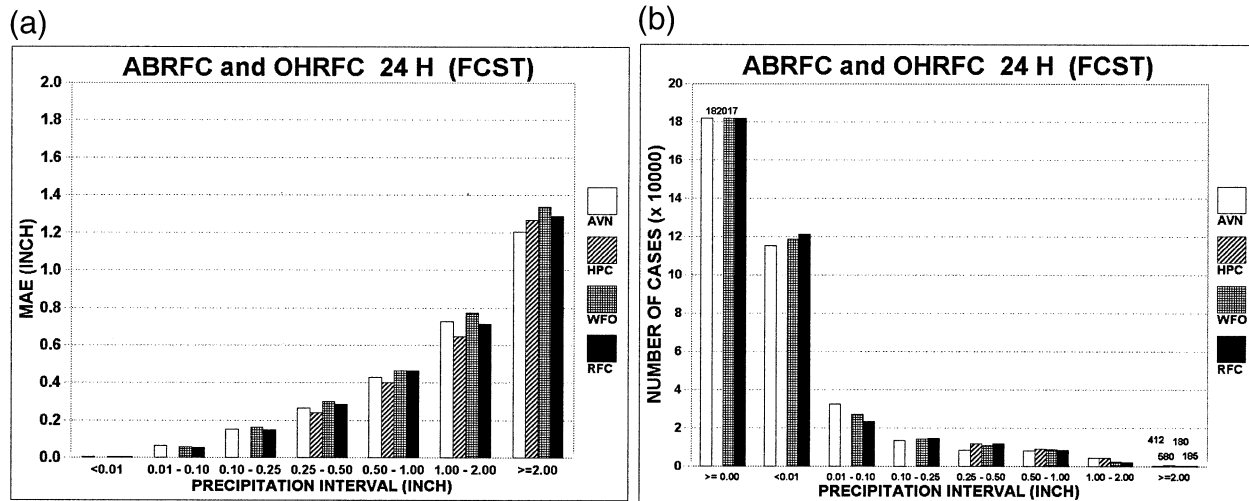


FIG. 9. (a) MAE vs forecast precipitation (denoted FCST in the title) in ME intervals and (b) the subsample sizes corresponding to (a). In (b) the number of cases is plotted above the bars for the first and last ME precipitation intervals. Otherwise as in Fig. 8.

in doctoral work by McDonald (1998), one of the co-authors of this study. It is noteworthy, however, that Murphy and Winkler (1987), Murphy et al. (1989), and Brooks and Doswell (1996) applied ME intervals for “diagnostic” verification of NWS temperature forecasts in a “distributions oriented” approach. In a subsequent section, we discuss how the stratification of the “summary measures” scores by using ME intervals in this study represents a significant step toward a distributions-oriented verification approach.

c. Score stratification according to observed and forecast precipitation

Stratification of verification scores on the basis of *observed* precipitation, as in Fig. 8b, has been applied in all formal QPF verification articles known to the authors [recent studies include Junker et al. (1992), Olson et al. (1995), Schultz and Snook (1996), Colle et al. (1999, 2000), and Mao et al. (2000)]. A positive attribute of this approach is that the subsample for each precipitation interval is identical for each QPF product being comparatively scored (Fig. 8c), and thus a comparison of the scores is strictly appropriate. On the other hand, forecast precipitation is an important input variable for streamflow models, and so performance scores stratified by QPF amount should be relevant to RFC users.

A plot of MAE (for the same sample as used in Fig. 8) in which the stratification now is according to intervals of forecast precipitation is shown in Fig. 9a. Note that in this figure the rise in MAE with increasing forecast precipitation is more uniform than for observed precipitation (Fig. 8b). Also, the peak MAEs, which are for ≥ 2.00 -in. forecasts, are substantially less than the corresponding MAEs stratified by observed precipitation. This result implies that for those days in which

very heavy precipitation is forecast the accuracy of those forecasts is better than when such precipitation occurs irrespective of the forecast. However, part of the increased error in the heaviest interval of observed precipitation may arise from very large error associated with rare cases of extremely large observed precipitation amounts.

Score stratification based on forecast precipitation involves a drawback in comparative verification applications, however, in that the subsamples corresponding to the precipitation intervals vary among the different forecast products (Fig. 9b). As a consequence, in a strict sense it is not appropriate to compare scores among the various QPF products with this stratification approach. However, if the subsamples among the QPF products do not vary greatly, a rough score comparison is possible. Indeed, Fig. 9b shows that the subsamples among the products are generally similar, especially for precipitation amounts under 2.00 in.

[A noteworthy point concerns the bogused HPC QPFs below 0.25 in. (see section 5a) for the period-I verification in the eastern United States. As noted in the captions of the scoring charts presented thus far, HPC scores are not shown for intervals under 0.25 in. because of the presence of the bogused QPFs. Although exclusion of these HPC scores removes contamination from the bogused QPFs in the case of the forecast-conditioned stratification (Fig. 9), in the corresponding observed precipitation stratification bogused values could appear for cases of observed precipitation of ≥ 0.25 in. Thus, an experiment was conducted to see what impact the bogused QPFs had on the HPC MAEs in this figure. We found that while the fractions of bogused QPFs for intervals of observed precipitation above 0.25 in. were not minor (they ranged from 24.9% for the 0.25–0.50-in. interval to 2.6% for ≥ 2.00 in.), the removal of these cases from the HPC sample resulted in a negligible

change in the corresponding MAE values (not shown). Thus, the adverse impact of the small fractions of bogus HPC QPFs was essentially lost when mixed with the large error in valid QPFs. Therefore, the presence of the bogus HPC QPFs for precipitation intervals above the 0.25-in. threshold was ignored.]

As noted before, the stratification of summary measures, such as MAE (or rmse), by observed and forecast precipitation amount employed in this study shares a component of the distributions-oriented verification approach advanced by Murphy and Winkler (1987). These authors show that the joint probability distribution of forecasts and observations, on which this verification approach is based, contains all non-time-dependent information about the relationship between these variables. Moreover, they demonstrate how the joint distribution is more easily understood when factored into conditional and marginal distributions, for which the baseline variable constitutes the forecasts in one factorization and the observations in the other (see the above-cited article for elaboration). In addition, Murphy et al. (1989) identify several classes of diagnostic verification methods based on the approach, one of which consists of summary measures of the joint, conditional, and marginal probability distributions. Of relevance is that the technique of stratifying the MAE on observed precipitation on the one hand (Fig. 8b) and on forecast precipitation on the other (Fig. 9a) is an example of a summary measure of the conditional probability distributions of concern. Also, the frequency distributions for observed precipitation (Fig. 8c) and forecast precipitation (Fig. 9b) constitute the marginal distributions in the factorizations. Thus, we find that the verification techniques adopted for this study involve a significant step toward a distributions-oriented approach. Still, it is important to mention that development of the joint and conditional probability distributions of the QPFs and verifying precipitation data (had we pursued the distributions-oriented verification scheme) would have been difficult, considering the high “complexity” and “dimensionality” of the problem and the small verification samples involved (Murphy 1991).

d. Combining accuracy scores for observed and forecast precipitation stratifications

In this study, an important objective was to rank the accuracy of the QPF products on the basis of the forecast error scores given by (1) or (2). In the case of MAE, low values are associated with high rank, and vice versa for high MAEs. On careful comparison of the MAE charts for the observed and forecast precipitation stratifications (Figs. 8b and 9a, respectively), we find that the accuracy rankings among the QPF products for some precipitation intervals are inconsistent. Clear inconsistencies are seen in the three ME intervals spanning 0.25–2.00 in., and a slight inconsistency is seen in the ≥ 2.00 -in. interval. These rank inconsistencies should not be

surprising because the distributions of observed and forecast precipitation amount (Figs. 8c and 9b, respectively) are not the same. Nevertheless they present a dilemma for objectively ranking overall QPF product accuracy, wherein both score stratifications are considered to be important.

A technique that addresses the ranking dilemma involves combining MAEs based on the two stratifications into single scores. This merging is accomplished computationally by looping through the verification sample twice, wherein stratification based on observed precipitation is used in one pass and forecast precipitation is used in the other. In essence, the combined MAE, MAE_C , within an ME interval is defined as

$$MAE_C = \frac{N_F MAE_F + N_O MAE_O}{N_F + N_O}, \quad (4)$$

where MAE_F and MAE_O are the corresponding MAEs stratified by forecast precipitation and observed precipitation, respectively, and N_F and N_O are the corresponding subsamples sizes. When viewed from the standpoint of a single map for a QPF product and the verifying precipitation field, MAE_C is the mean of the absolute forecast error within the envelope of the areas where forecast precipitation and observed precipitation lie within the interval. Note that where the forecast and observed areas coincide (overlap), a point corresponding to the forecast and the observation is counted twice.

The chart for MAE_C and the distribution of the combined subsamples corresponding to Figs. 8 and 9 is shown in Figs. 10a and 10b, respectively. A careful comparison of the QPF product rankings based on MAE_C (Fig. 10a) with corresponding rankings based on MAE_O and MAE_F (Figs. 8b and 9a, respectively) reveals that MAE_C blends the latter two scores as intended. Note also, from Fig. 10b, that the subsamples corresponding to MAE_C are sums of the subsamples corresponding to MAE_O and MAE_F (Figs. 8c and 9b, respectively). Thus, where the rankings based on MAE_O and MAE_F diverge, as for the AVN and HPC in the 1.00–2.00-in. interval in Figs. 8b and 9a for example, MAE_C yields a ranking that appropriately weights the two “disparate” MAE values. Of course, for precipitation intervals within which a QPF product has either the best or worst ranking for *both* MAE_O or MAE_F , that ranking is retained in MAE_C . Further, even when the product rankings are not changed appreciably in MAE_C , the *relative* MAE values among the products can change substantially, as seen in the ≥ 2.00 -in. interval in Figs. 8b, 9a, and 10a. In the latter example, the MAE_C values for HPC and AVN fell relative to the WFO and RFC values because the MAE_F values among all QPF products are lower than the corresponding MAE_O values, and HPC and the AVN had many more cases corresponding to MAE_F . Thus, MAE_C can serve as a useful tool for ranking the QPF products when an account of both types of score stratifications is needed.

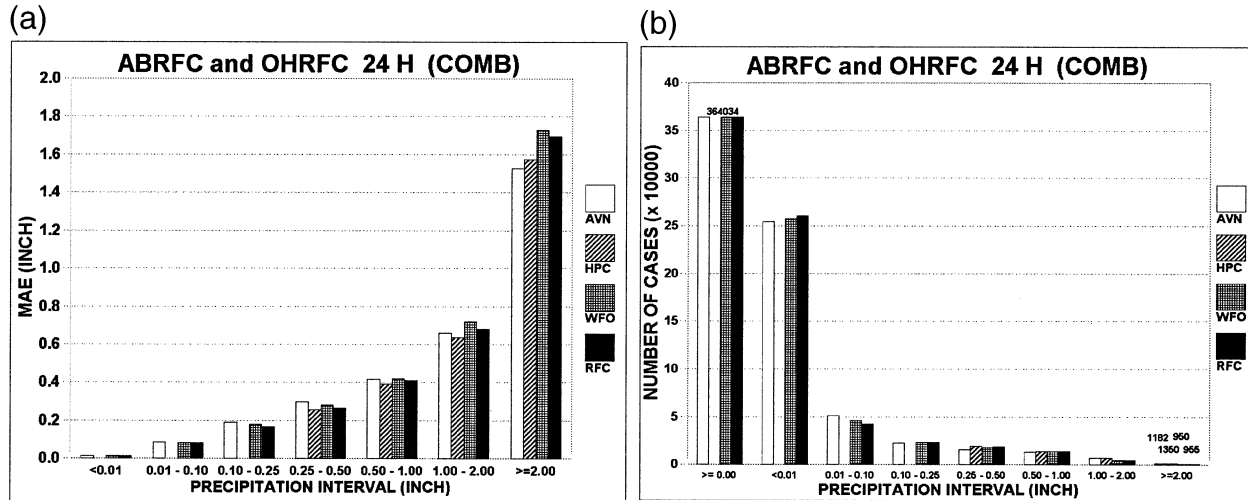


FIG. 10. (a) MAE vs observed and forecast precipitation combined (denoted COMB in the title) in ME intervals, and (b) the subsample sizes corresponding to (a). In (b) the number of cases is plotted above the bars for the first and last precipitation intervals. Otherwise as in Fig. 8.

It is important to point out that some users of the accuracy scores developed for this study might prefer the scores individually stratified by observed and forecast precipitation rather than the combined scores. These users would likely include operational forecasters who wish to understand the error characteristics of the forecasts when amounts in various ranges are forecast or observed. Moreover, on the basis of the interpretations of the conditional probability distributions provided in Murphy and Winkler (1987) and Murphy et al. (1989), it could be argued that absolute accuracy scores stratified on the basis of the forecasts contain different information (about the forecasts) than corresponding scores stratified on the basis of the observations. In specific terms, MAE stratified (conditioned) on forecast precipitation amount could be interpreted as forecast error that *in part* arises from the degree to which the QPFs are not calibrated (are conditionally biased). The corresponding MAE stratified on observed precipitation indicates the degree to which a forecasted amount “discriminates” among possible values of the observed amount [also see Wilks (1995) and Brooks and Doswell (1996) for discussions of this topic]. Thus, these two stratifications appear to provide different views of the QPF error, which some users might prefer. Therefore, at the time of writing, *both* the combined and separate absolute accuracy scores are provided online at a Web site (<http://www.hpc.ncep.noaa.gov/npvu/>) that contains verification data from the recently implemented national QPF verification program (the implemented verification methods are from this study). Nevertheless, because the scope of the current study does not extend beyond providing a simple overall accuracy ranking of the various QPF products, only the combined scores are considered henceforth.

e. Scores used for assessing QPF performance

In section 6a, we noted that rmse provides a measure of absolute forecast accuracy that could complement that provided by MAE. To see if rmse could benefit this study, this score was applied to the same sample of 24-h QPFs as was used for the MAE applications in the previous sections. It was found that, although the magnitudes of the rmse were higher than corresponding values of the MAEs, the rankings of the QPF products based on the two scores were essentially identical (not shown). This finding was also noted in samples involving the 6-h QPF products. Thus, for the purpose of ranking forecast accuracy in this study, we concluded that the two scores provided essentially redundant information and that only one of them was needed. Because of its greater simplicity, MAE was chosen.

MAE was used as the primary tool to judge QPF performance, but the forecast bias also has relevance for the hydrologic user. As indicated previously, its role in this study was to describe overforecasting/underforecasting properties of the various QPF products. Scoring tests for the sample at hand with the bias conditioned on observed and forecast precipitation (as for MAE) resulted in scores with little utility. In particular, the bias conditioned on observed (forecast) precipitation indicated extreme overforecasting (underforecasting) of very light precipitation and severe underforecasting (overforecasting) of very heavy precipitation (not shown). Such extreme conditional bias for very small and very large precipitation amounts probably arises as a consequence of the large error inherent in QPFs together with the characteristic spatial patterns of precipitation fields. For example, for the case in which the bias is conditioned on very light observed precipitation amounts, one can readily envision frequent situations in

which just a small portion of the relevant area (or a few of those points) is paired with forecast amounts that could be greater by as much as one or two orders of magnitude. The bias given by (3) would be very large in this case. In converse, when the condition is heavy observed precipitation amounts, a very small bias could arise because a small portion of those areas could have matching forecast values that are up to two orders of magnitude smaller. An inverse conditional bias scenario arises when forecast precipitation is the conditioning variable. In essence, because the forecast and observed values are paired and the condition is only on one of them, the conditional bias takes into account the accuracy of the forecasts. Thus, it is virtually unimaginable that perfect conditional bias could be achieved unless the forecasts were perfect.

An unconditional bias (for a particular ME interval) was specified by simply summing all forecast and observed precipitation amounts [numerator and denominator, respectively, in (3)] in a sample that falls within the interval. The essential distinction between this bias specification and that for conditional bias is that here the forecast and observed precipitation data are not paired; rather, the forecast and observed precipitation data are summed independent of one another. Thus, perfect bias could be obtained for an ME interval even when paired forecast and observed amounts *never* fall in the interval; that is, the forecasts are grossly inaccurate. This unconditional bias specification is appropriate for hydrologic applications, because unbiased forecasts over a large watershed can have substantial utility despite exhibiting poor accuracy. Thus, only the unconditional bias was used in this study.

For the spatially averaged ABRFC and OHRFC precipitation data used in this section, the (unconditional) bias for the various QPF products is shown in Fig. 11. The QPF products generally exhibit moderate overforecasting for light and moderate precipitation amounts and strong underforecasting for the heaviest amounts. The overall relationship of bias to precipitation amount is similar to that documented in previous QPF verification studies (e.g., Junker et al. 1992; Olson et al. 1995; Mesinger 1996; Colle et al. 2000).

f. Ranking performance among the QPF products

For this study, MAE_C provides an objective basis on which to rank forecast accuracy among the QPF products,⁷ and the bias provides a diagnosis of the degree of overforecasting or underforecasting. Because of the relevance to river flood forecasting, MAE_C values for the heaviest precipitation intervals warranted dominant consideration in the ranking process. Because the heaviest

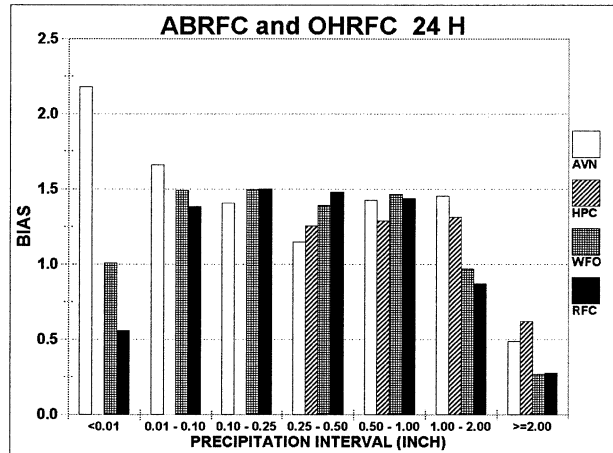


FIG. 11. Bias vs precipitation amount in ME intervals. Otherwise as in Fig. 10.

est events are also the most rare, their sample size must be carefully considered.

In the frequency distribution of 24-h observed precipitation for the combined areas of ABRFC and OHRFC (Fig. 8c), we see that the subsample for ≥ 2.00 in. (770 events) is clearly the smallest among all precipitation intervals. To allow one to gain an appreciation of how this rare event total was formed, Table 3a contains the number of these events (when at least one occurred) in individual 24-h periods for the separate RFC areas. Note that the 770 events were reasonably well dispersed over the sampling period and the two RFC areas: there were a total of 22 periods in which an event occurred, over the 6-month sample. On the other hand, the individual 24-h counts exhibit a wide variation, and the large number of events that appear for a small number of periods in the individual RFC areas (82 or more events occurred in five periods) would make statistical hypothesis tests for the MAE scores difficult to apply [because of likely high auto- and serial correlations in the data (Wilks 1995)]. Further, even if statistical testing indicated the MAE differences were significant, this would not ensure significance in a practical sense, that is, from the standpoint of, say, economic value to a user. Thus, statistical testing was not employed in this study. Instead, for all verification samples used in this study for which the distribution of observed precipitation events was reasonably robust, we chose, as a rough criterion, a difference in MAE_C among the QPF products of about 10% or more to indicate a potentially meaningful difference in forecast accuracy. Application of this criterion to the 24-h QPFs in Fig. 10a shows that the improvement (in MAE_C) of HPC and the AVN on the WFOs is meaningful in the highest two precipitation intervals. This approach for assessing meaningful improvements in QPF accuracy was also applied for the 6-h QPF products.

⁷ Note, however, that no single scoring measure (such as MAE) can completely describe forecast quality for all users (Murphy and Ehrendorfer 1987).

TABLE 3. Number of nonzero observed precipitation events of (a) ≥ 2.00 in. in a 24-h period and (b) ≥ 1.00 in. in a 6-h period for the RFC area and sampling period indicated. Also listed is the total number of events in parentheses and the number of (24 or 6 h) periods with one or more events. The sampling periods begin on the first day of the month indicated and end on the last day of the month indicated. For ABRFC and OHRFC, an event pertains to the mean precipitation amount over a 30-km grid box; the number of grid boxes is 743 for ABRFC area and 580 for the OHRFC area. For CNRFC and NWRFC, an event pertains to a point (gauge) precipitation amount; over the sampling period, the average number of gauge observations was 36 in the CNRFC area and 41 in the NWRFC area.

(a)				
RFC	Sampling period	No. ≥ 2.00 -in. events in 24-h period	(Total)	No. of 24-h periods
ABRFC	Oct 1998–Mar 1999	100, 21, 90, 108, 8, 118, 10, 1, 9, 3, 10, 8, 42, 2	(530)	14
OHRFC	Oct 1998–Mar 1999	35, 1, 33, 31, 38, 18, 82, 2	(240)	8
(b)				
RFC	Sampling period	No. of ≥ 1.00 -in. events in 6-h period	(Total)	No. of 6-h periods
ABRFC	Oct–Dec 1998	16, 21, 37, 81, 1, 10, 7, 3, 1, 1, 8, 25, 61, 69, 55, 51, 58, 17, 2, 12, 1, 18, 38, 14, 6, 3, 17, 75, 51, 10, 47, 31, 21, 12, 1, 3, 5	(889)	37
ABRFC	Jan–Mar 1999	3, 2, 8, 2, 5, 6, 2, 32, 9, 1, 4, 3, 31, 54, 21, 6, 3	(192)	17
OHRFC	Oct–Dec 1998	6, 2, 1, 3, 3, 17, 51, 23, 2, 13, 1, 9, 1, 1, 31, 8, 2, 12, 14	(200)	19
OHRFC	Jan–Mar 1999	6, 2, 46, 14, 5, 1, 29, 16, 2, 34, 60, 28, 1, 1, 1, 18, 1, 2, 10, 1, 3	(281)	21
CNRFC	Nov 1999–Mar 2000	1, 1, 2, 1, 1, 4, 3, 3, 1, 1, 1, 1, 2, 2, 1, 1, 1, 1, 1, 1, 1, 6, 2, 3, 3, 4, 5, 4, 3, 2, 3, 2, 1, 3, 2, 3, 3, 2, 3, 4, 10, 3, 2, 1, 1, 5, 1, 1, 3, 2, 1, 1, 1, 1, 1, 4, 3, 1, 1, 1, 1	(137)	62
NWRFC	Nov 1999–Mar 2000	1, 1, 1, 1, 4, 1, 1, 1, 1, 1, 1, 5, 4, 5, 6, 4, 5, 1, 1, 1, 1, 2, 1, 6, 9, 1, 1, 2, 1, 1, 1, 4, 3, 2, 5, 3, 1, 1, 1	(92)	39

7. Results from the comparative verification of 6-h QPF products

The comparative verification of 6-h QPF products embodied those from three NCEP operational NWP models, two MDL statistical models, and three layers of human-generated products (see section 2), but for conciseness results are presented only for what was found to be the best NWP model for QPF (AVN) and the three manual (HPC, WFO, and RFC) products.⁸ Also, because the primary purpose of the comparative scoring was to rank the overall accuracy of the various QPF products for the RFC user, the MAE scores for only the combined scoring stratification are presented.

a. Period I: Eastern United States

As for the 24-h QPF scores in the previous section, the 6-h QPF scores for ABRFC and OHRFC were combined in the period-I verification (Table 1) because they

⁸ Some readers may be surprised that the NCEP global/spectral AVN Model produced better QPF scores than the regional Eta Model (e.g., Mesinger 1996). Verification statistics (computed independent of this study and available at the time of writing online at <http://www.hpc.ncep.noaa.gov/html/hpcoverif.shtml>) show a year-to-year rotation in the superiority of QPFs from the two models. Annual verification statistics (available at this Web site) that span the two cool seasons included in this study show a slight superiority for AVN, which is consistent with the findings from this study. Further, Mesinger (1998) noted that Eta is less competitive during winter than during summer.

Scores from the MDL statistical models are excluded from presentation because uniqueness in the product forms from these models precluded strict comparison with other products.

were similar. Figure 12 shows MAE and bias for the AVN, HPC, WFO, and RFC QPF products, wherein statistics for the four 6-h periods composing day1 are combined. The historical sample used for this figure was limited to the second half of period I (1 January–31 March 1999), because the gridded HPC 6-h QPF product was not available for the first half. Note that the heaviest precipitation interval for which the scores are shown is ≥ 1.00 in. As shown in Table 3b, there were only 38 (17 + 21) 6-h periods in which the ≥ 1.00 -in. event occurred in the ABRFC and OHRFC areas within the January–March 1999 record. The corresponding number of 30-km grid boxes with this event, totaling 473 (192 + 281 in Table 3b), is considered to be adequate to yield stable relative scores, especially because the associated numbers of forecast grid boxes among the various QPF products exceeded this number; they were in the range 533–885 (not shown).

The MAE and bias scores for the QPF products in Fig. 12 exhibit several noteworthy features. MAEs for HPC are slightly better than those for the WFOs and RFCs (over the three precipitation intervals for which a comparison is possible; Fig. 12a), and it is surprising that the AVN scored about as well as HPC. On the basis of the 10% MAE difference criterion noted earlier, either the HPC or AVN QPF accuracy in the heaviest two precipitation intervals is better than the corresponding accuracy of either the WFO or RFC QPFs. Another feature in Fig. 12a is that the magnitude of the QPF error for all products is very large, especially for ≥ 1.00 in. for which the MAE is only slightly less than the lower bound of the interval. The corresponding bias

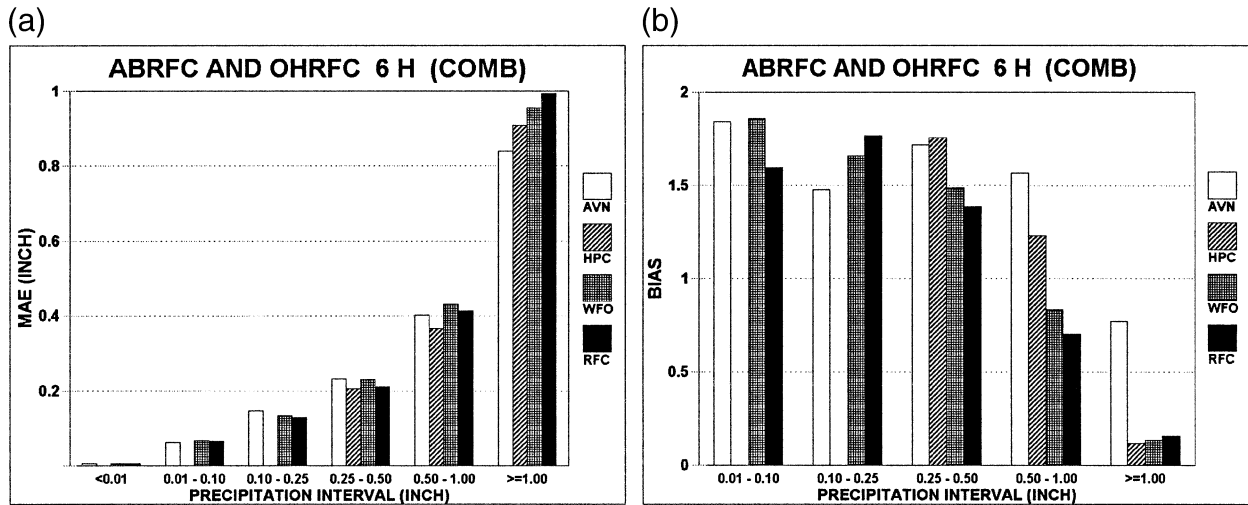


FIG. 12. (a) MAE and (b) bias for the 6-h AVN, HPC, WFO, and RFC QPF products for day 1 and the ABRFC and OHRFC areas combined. “COMB” in the titles denotes the combined forecast and observed precipitation stratification in the scoring. The HPC product is not shown for intervals of less than 0.25 in. because HPC did not forecast such amounts during period I. In (b), the bias is not shown for <0.01 in. because the AVN value was well beyond the upper bound of the scale. The sample (shortened) spans the period 1 Jan–31 Mar 1999.

chart (Fig. 12b) shows that each of the three manual QPF products overforecast precipitation below 0.50 in. and severely underforecast precipitation above 1.00 in. The corresponding bias for the AVN over the three comparative intervals shows a slight overall improvement on the three manual products, because it does not exhibit the severe underforecasting in the highest interval. Also, note HPC’s extreme bias dropoff from the 0.50–1.00-in. interval to ≥1.00 in. This result is probably a reflection of the graph-to-grid interpolation limitation for the HPC QPF graphic noted in section 5a, wherein annotations of the maximum point amount within the heaviest isohyet were ignored. Of course, this scoring deficiency should also have had an adverse impact on the corresponding HPC MAE for ≥1.00 in. in Fig. 12a.

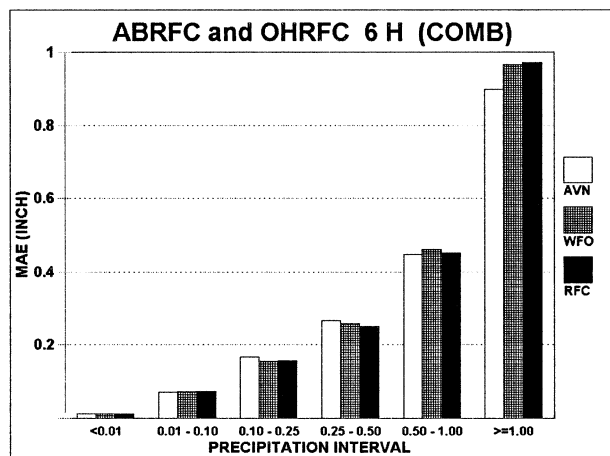


FIG. 13. As in Fig. 12a but with the HPC QPF product excluded and for the full period-I sample (1 Oct 1998–31 Mar 1999).

Figure 13 shows the comparative MAEs for the QPF products over the full 6-month period-I sample (1 October 1998–31 March 1999), as made possible by excluding the HPC product. The figure shows that for the ≥1.00-in. interval the improvement in MAE for the AVN over the WFOs and RFCs almost meets the 10% criterion. If the distribution of heavy precipitation storms during the second 3 months of the full period (second 3 months was used for Fig. 12) is similar to that for the first 3 months, then it should be appropriate to compare the MAE scores in Fig. 13 with those in Fig. 12a. Table 3b shows that, although the sizes of the heavy precipitation storms (as indicated by the number of ≥1.00-in. events per 6-h period over the ABRFC and OHRFC areas) were much larger during the first half of the 6-month period than for the second half, the number of storms for the two periods was not greatly different [total of 56 (37 + 19) 6-h periods had one or more ≥1.00-in. events during first half vs 38 (17 + 21) for the second half]. Thus, the comparison of results in the two figures should be permissible. Comparison of results reveals consistency in the relative performances of the QPF products common to both figures. This consistency adds credence to the scores based on the shorter sample in Fig. 12a. It is also noteworthy that the corresponding bias plot for the full sample (not shown) also exhibited no appreciable change from that obtained with the shorter sample.

b. Period II: Western United States

Verification scores for CNRFC and NWRFC in the western United States from period II (1 November 1999–31 March 2000) are provided for the three veri-

fication approaches discussed in section 5b and summarized in Table 2. In regard to robustness of the verification samples, Table 3b shows that the frequency of the number of 6-h periods for which the rarest precipitation event (≥ 1.00 in.) occurred at CNRFC and NWRFC is somewhat higher than for the period-I sample at ABRFC and OHRFC; the smaller number of reported events at the western RFCs arises because the average number of gauge observations was far less than the number of grid boxes for the eastern RFCs (36 gauge observations for CNRFC and 41 for NWRFC as compared with 743 grid boxes for ABRFC and 580 for OHRFC). Thus, as for the eastern RFCs, the western sample for the rarest event is considered to be adequate to yield approximately stable relative scores. Also, in contrast to scoring results for the eastern United States, the HPC scores in the West are shown for all precipitation intervals (bogus QPFs were not involved), and the scores for all QPF products span days 1–3.

The MAE and bias scores for the “station points” scoring approach (approach 1 in Table 2) are shown in Fig. 14 for the combined CNRFC and NWRFC areas (the scores were similar for the two areas) and separately for days 1, 2, and 3. An important feature in the MAE charts is that the WFOs scored better than HPC and AVN on day 1, though they did not improve on HPC on days 2 or 3. The superior performance of the WFOs on day 1 is also evident in the corresponding bias chart, whereas for days 2 and 3 HPC had a better bias in the heaviest precipitation interval. These results suggest that the WFOs focused their effort on day 1 whereas HPC applied model guidance more uniformly over the three forecast days. AVN is seen to underperform the WFOs and HPC in terms of both MAE and bias, especially on day 1. The typical bias trend of overforecasting of light precipitation and underforecasting of heavy amounts is especially striking for AVN.

As noted previously, HPC and AVN are inappropriately scored in the station points approach, because these QPF products represent spatially averaged precipitation whereas the verifying precipitation amounts apply to specific stations. Thus, the gridded scoring approaches (approaches 2 and 3 in Table 2) should be more equitable. (Recall that the gridded approaches were applied only for CNRFC because the MM gridding method was not implemented at NWRFC at the time of this study.) Note also that the MM (gridded) scoring approach is appropriate for the WFOs because the WFO (station) QPFs and verifying gauge observations are conserved in the MM gridding scheme.

Figure 15 shows comparative scores for the station points and gridded scoring approaches for the WFOs and HPC (AVN is not included for brevity, and day-2 scores are not shown because they were similar to those for day 3, as in Fig. 14). [Recall that only one gridded approach applies to the point WFO QPFs and verification data—gridded approaches 2 and 3 are identical—whereas for HPC (and AVN) approach 2 involved the

raw grids while approach 3 involved the MM grids.] Figure 15 shows that, in terms of the station-points approach, the WFOs scored better than HPC at CNRFC, not only on day 1 but even on day 3, though to a lesser degree for the latter. With the gridded scoring approaches, though, the improvement in HPC’s scores (over the corresponding station-points scores) was greater than that for the WFOs, such that by day 3 HPC’s MAE and bias were better in most precipitation intervals. Two points are noteworthy. One is that the general improvement in the scores from the station-points to the gridded approaches is expected, in part because the gridding procedures involve spatial averaging and thus increased spatial coherence in both the forecast and verifying fields (Charba and Klein 1980; Bosart 1980; Gyakum and Samuels 1987; Mesinger 1996; Schultz and Snook 1996). Another likely factor in the score improvement is the artificial correlation introduced by the MM gridding of both the forecasts and observations. The greater and more consistent improvement in MAE and bias scores for HPC likely reflects the inappropriateness of verifying the spatially averaged HPC QPFs on the basis of point precipitation data. This finding is consistent with a similar result presented by Cherubini et al. (2002).

Based on the above findings, the most meaningful comparison of performance for the AVN, HPC, and WFO QPF products in the western United States involves the gridded scoring with the MM approach. The scores based on this method for CNRFC are shown for all three QPF products for day 1 and day 3 in Fig. 16. The MAE charts show that the WFOs maintain their accuracy superiority over HPC and the AVN for the heaviest precipitation interval on day 1. By day 3, HPC achieved a similar improvement in accuracy on the WFOs in the ≥ 1.00 -in. interval [the corresponding scores for day 2 were similar (not shown)]. The AVN had the poorest MAE on day 1, but by day 3 it scored slightly better than the WFOs and almost equal to HPC. Despite the superior MAE scored by the WFOs on day 1, their bias was not better than HPC’s bias. Also, on day 3 HPC’s bias for the heaviest precipitation interval is clearly better than that for the WFOs and AVN.

A comparison of the gridded scores for the western United States with those for the East reveals interesting findings. The day-3 MAEs for CNRFC (Fig. 16) are clearly better than the day-1 MAEs for ABRFC and OHRFC (Figs. 12a and 13). This result may be surprising to some because of the forecasting impediment in the western United States imposed by the upstream data-void Pacific Ocean. The QPF performance strength in the West is believed to be reflective of the topographic focusing of precipitation by major mountain chains, which makes positioning of precipitation areas less difficult than in the East. (It might also explain why RFCs in the eastern United States are not using the newly available day-2 and day-3 QPFs in hydrologic models,

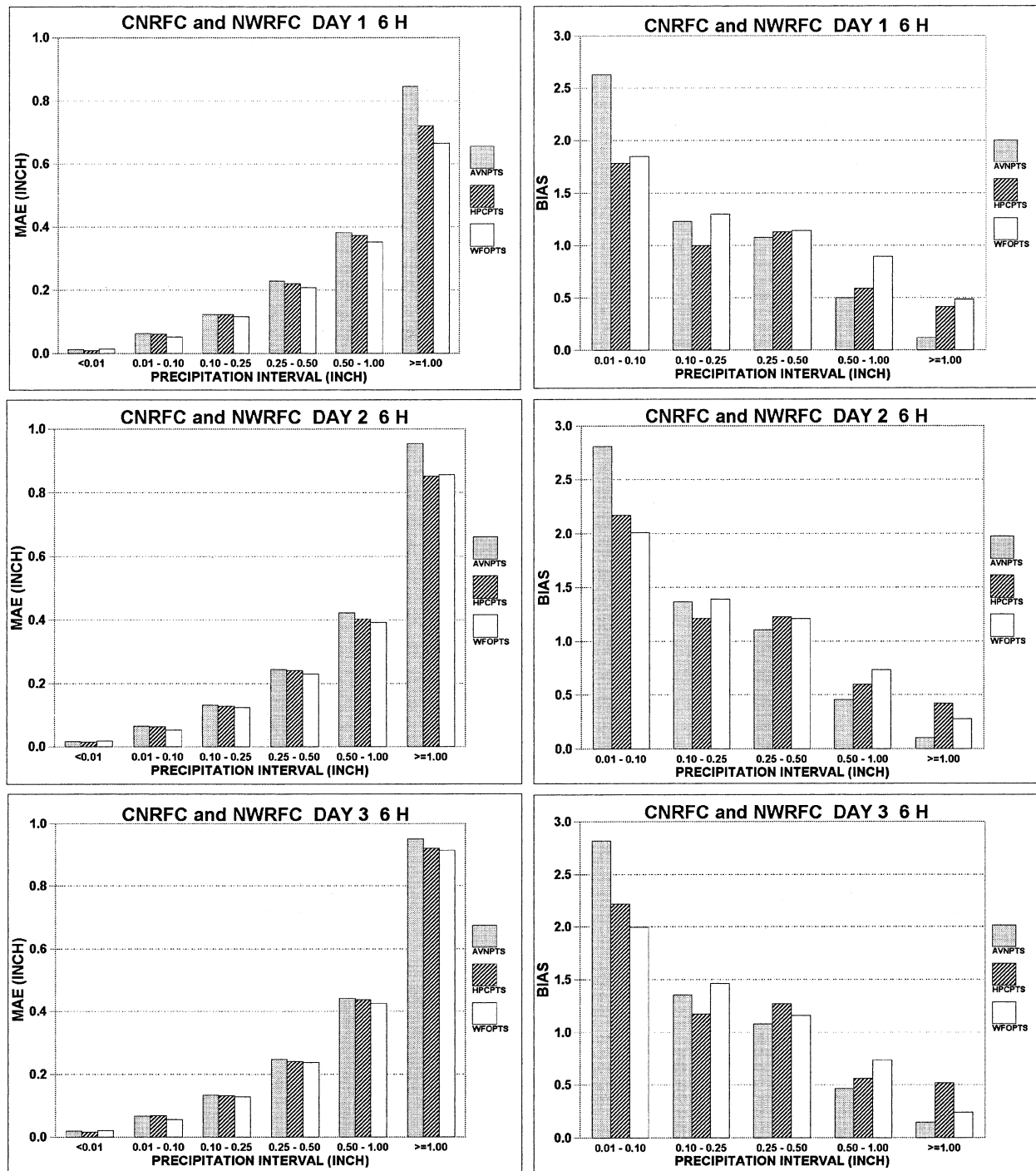


FIG. 14. (left) MAE and (right) bias during period II (1 Nov 1999–31 Mar 2000) based on station points scoring for (top) day 1, (center) day 2, (bottom) day 3, and the CNRFC and NWRFC areas combined. The <0.01-in. interval is not shown in the bias charts because the bias is not defined for 0 and nonmeasurable gauge precipitation.

except in special meteorological situations.) Another finding gleaned from the regional comparison is that HPC did not exhibit the extreme underforecasting bias in the heaviest precipitation interval in the West (Fig.

16), which marred its bias performance in the East (Fig. 12b). This result supports the conclusion drawn earlier that the scoring limitation for HPC during period I had an adverse impact on HPC's scores in the East.

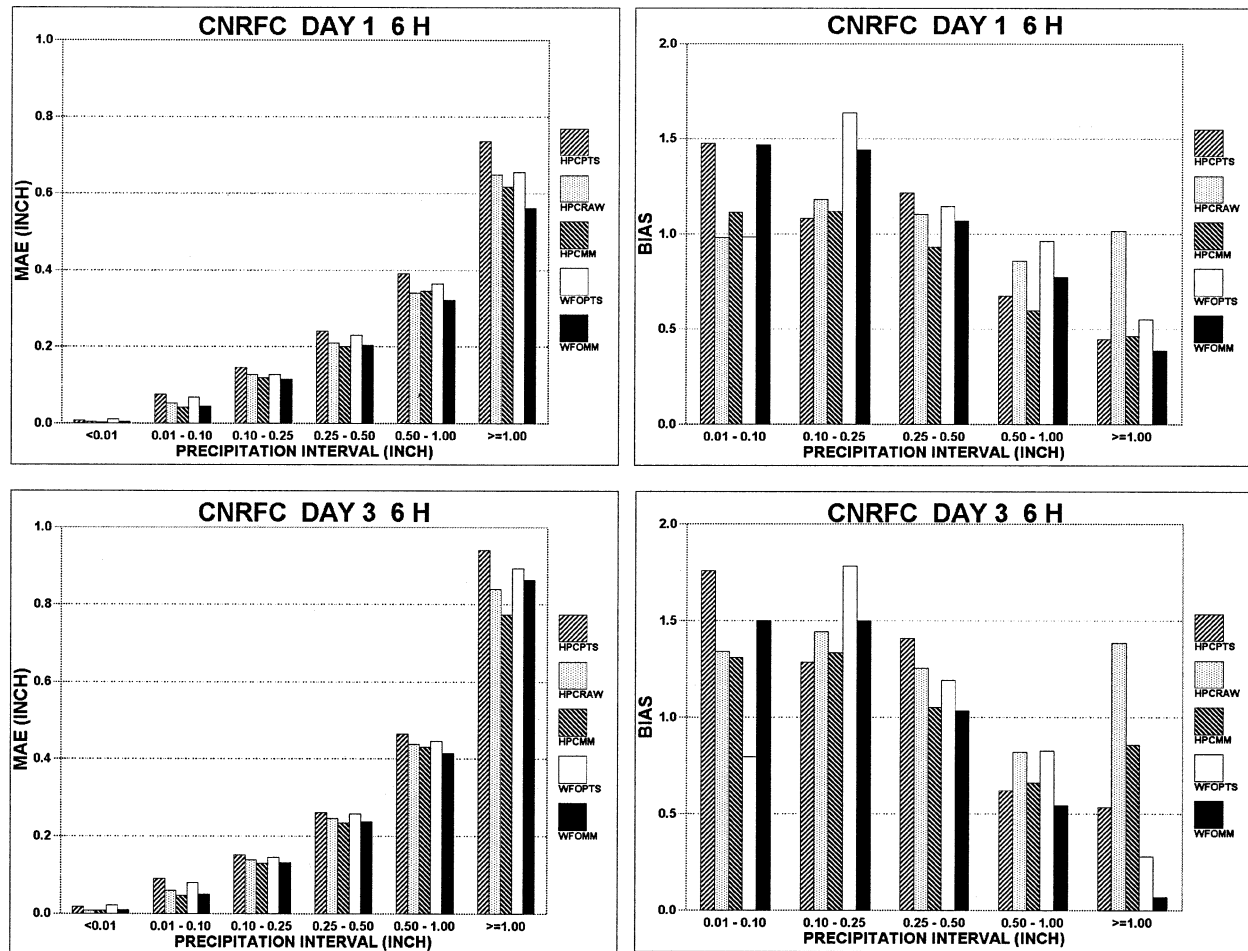


FIG. 15. (left) MAE and (right) bias for (top) day 1 and (bottom) day 3 based on the station points and two gridded scoring approaches over the CNRFC area. In the legend, “PTS” denotes station points scoring, “RAW” denotes the raw grids scoring, and “MM” denotes MM gridded scoring. Otherwise as in Fig. 14.

8. Discussion

a. Contributions of human forecasters to QPF performance

A principal aim of this study was to assess whether each of the three stages of human intervention (effectively two stages in the western United States) in the NWS QPF process contributed additional accuracy to the final product input to RFC streamflow models. The verification statistics presented in the previous section did not indicate an accuracy contribution by WFOs and RFCs to the available HPC product in the eastern United States. In the western United States the WFOs improved on the HPC product for the day-1 period but not for days 2 and 3.

To gain added insight into the human role in the QPF process, a series of case studies for heavy rain events were performed for both the eastern and western United States. This effort, which involved in-

spection of graphical presentations of all QPF products for each case as well as associated verification statistics, yielded several findings. One was that the adjustments human forecasters at the various NWS offices made to the NWP guidance usually involved changes in the spatial distribution and timing of the model precipitation. However, as we have seen in the previous section, such adjustments apparently did not consistently result in a more accurate forecast, at least for the eastern United States where verification statistics for day 1 indicated the best human QPF product (HPC) made at best a marginal improvement on the AVN Model (and the WFOs and RFCs performed slightly worse than this model). In the western United States, a clear improvement on the AVN was achieved by the WFOs and HPC in the same forecast range. For a single heavy precipitation event, Fig. 17 illustrates how WFO and HPC forecasters correctly redistributed the AVN-Model QPF to reflect the topo-

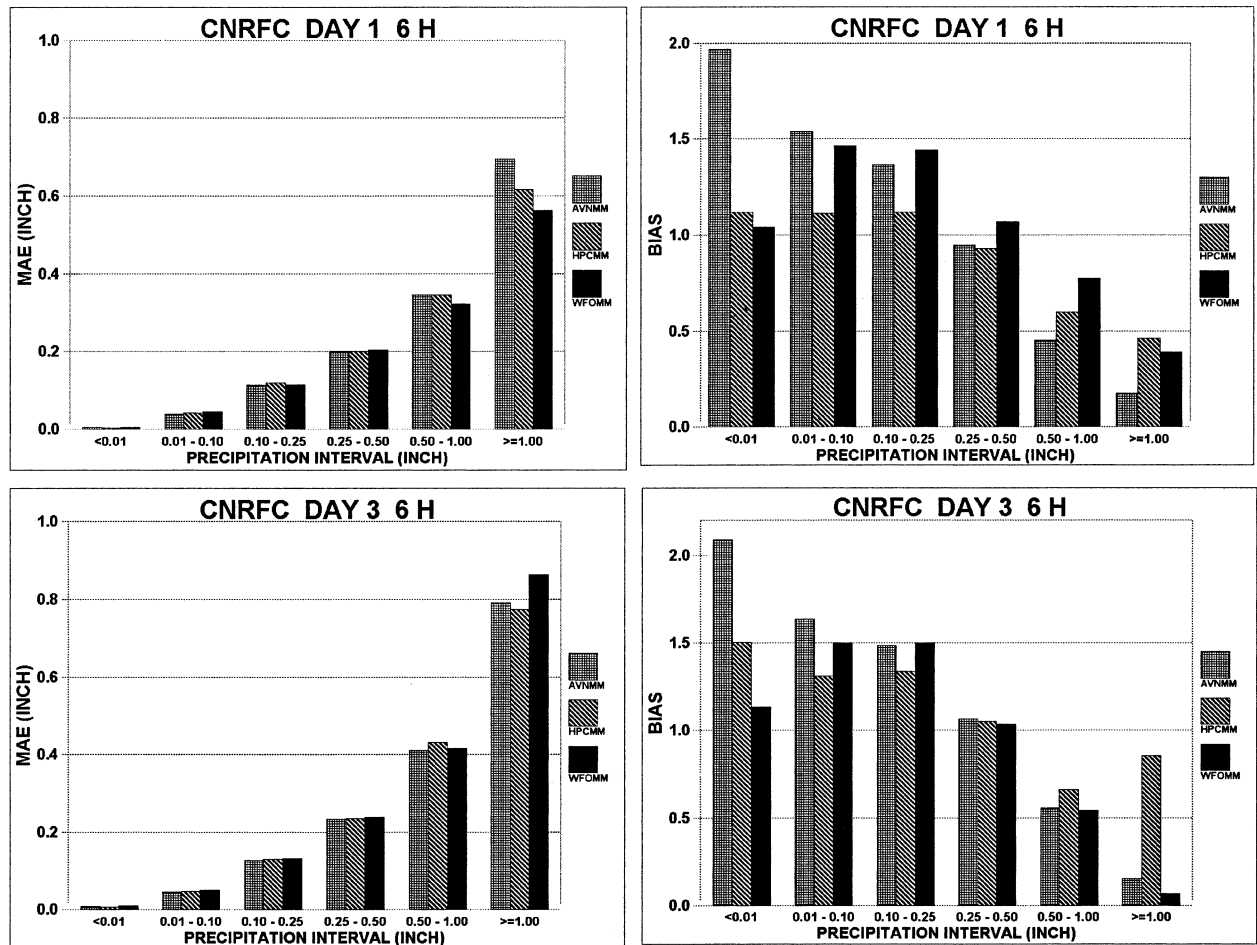


FIG. 16. (left) MAE and (right) bias for (top) day 1 and (bottom) day 3 over the CNRFC area based on the MM (gridded) scoring approach. Otherwise as in Fig. 15.

graphic focusing of precipitation by the Coastal Mountain and Sierra Nevada ranges.⁹ It also shows that WFO forecasters improved on the HPC guidance for this day-1 6-h period, in conformity with findings from the verification statistics for the western United States.

Another finding was the high degree of disparity among the manual QPF products issued by the various NWS offices. Especially in the eastern United States for period I, the HPC and WFO QPF products many times diverged, sometimes even in major map features. An example is seen by comparing the HPC QPF product in Fig. 18 with the corresponding WFO and RFC QPF products in Fig. 4. Note that the HPC QPF maximum in Fig. 18 is located in western Arkansas, the WFO

maximum is in southeast Oklahoma (Fig. 4b), and the RFC maximum lies between these locations (Fig. 4d).

An additional problem was observed in the WFO QPF composite for an RFC area. The problem consisted of spatial discontinuities in the QPF pattern along the HSA boundaries of adjacent WFOs, an example of which is shown in Fig. 4. Such a product inconsistency, as well as that noted in the previous paragraph, is believed to have arisen from largely independent interpretations of the various model guidance by forecasters at HPC and at the individual WFOs. This assertion is supported by findings gleaned from written responses to a comprehensive questionnaire sent to all WFOs and RFCs in conjunction with this study.¹⁰ WFO forecasters indicated

⁹ The relatively coarse spatial resolution of the AVN Model limits its capability to reflect orographic forcing even by major mountain ranges. The NCEP Eta Model contains higher grid and topographical resolutions, but it did not score better than AVN for the cool-season samples used in this study, even in the western United States.

¹⁰ Detailed results from the questionnaire survey of WFOs and RFCs are available in the "Final Report of the Quantitative Precipitation Forecast Process Assessment," which was available online (<http://www.nws.noaa.gov/er/hq/QPF/>). It is noted that the responses to the survey represent a good cross section of these NWS field offices, because 101 of the 117 WFOs and all 13 of the RFCs provided complete information.

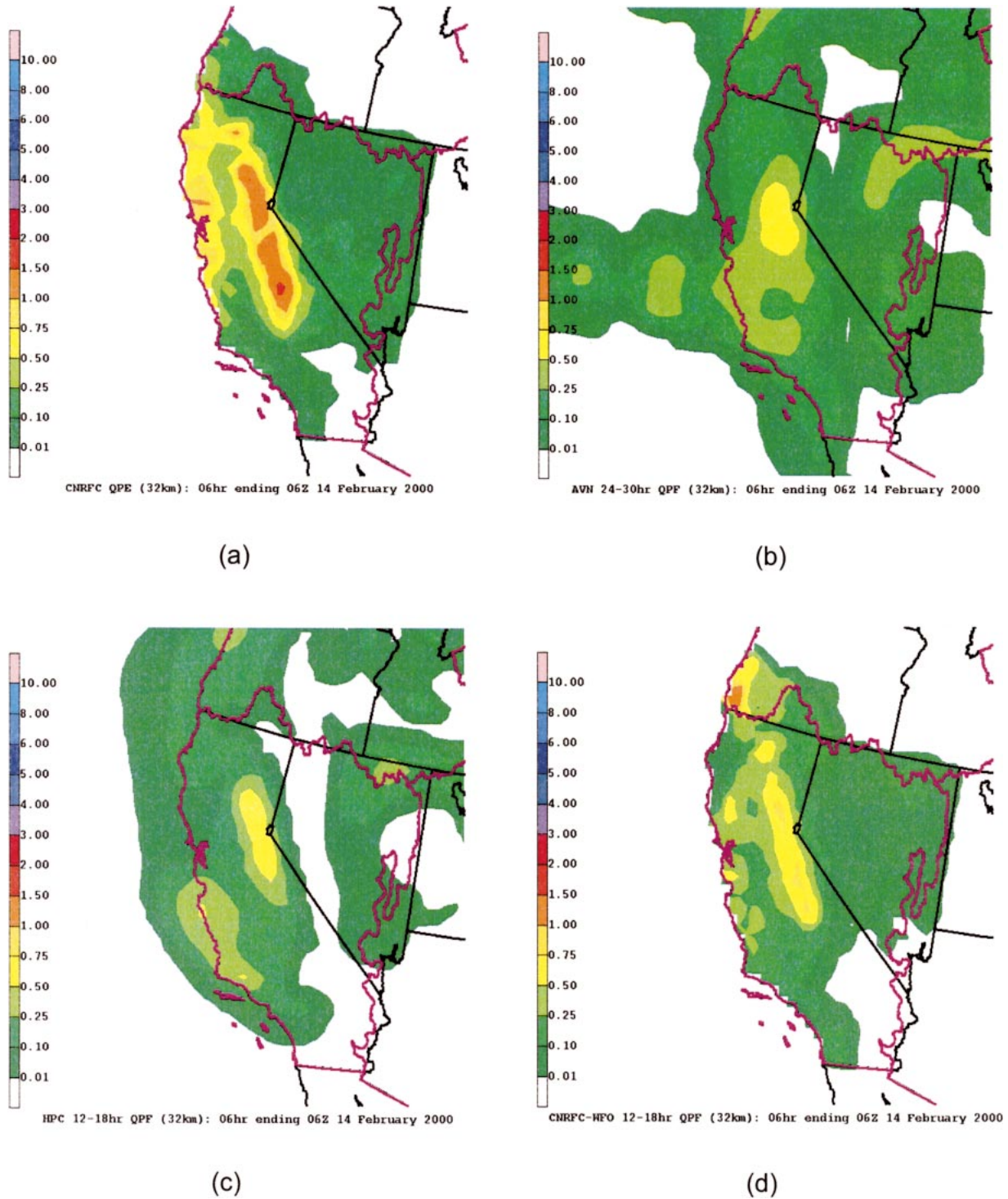


FIG. 17. (a) Observed (MM) precipitation analysis, (b) AVN QPF (raw), (c) HPC QPF (raw), and (d) WFO QPF (MM) (in.) on 30-km grid for 6-h period ending 0600 UTC 14 Feb 2000. The notations raw and MM are defined in section 7b.

that they were more likely to use the direct NWP model guidance than the HPC QPF guidance product. Further, the propensity for independent production of the QPF products evidently applies to neighboring WFOs, judg-

ing from the frequent spatial inconsistencies in the WFO QPF composite. In fact, an additional finding from the case studies was that a common function of HAS forecasters in the preparation of the RFC QPF product in

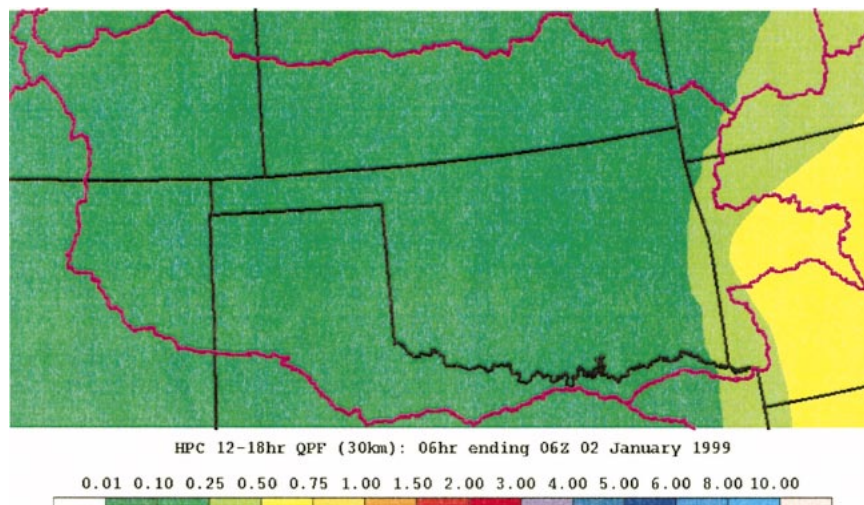


FIG. 18. HPC QPF product (in.) for ABRFC area (solid boundary) and the 6-h period ending 0600 UTC 2 Jan 1999 on the 30-km grid. Precipitation amounts under 0.25 in. are bogused.

the eastern United States was to “smooth out” spatial incoherence in the WFO QPF composite (see Fig. 4).

b. Ramifications of the study for NWS operations

Changes to the NWS QPF process have been implemented recently as a result of findings from the QPF verification presented in this article and a survey of all WFOs and RFCs (noted earlier), and from managerial considerations. The management considerations involved the necessity of optimizing the use of fixed human resources in the NWS in the face of increasing demand for products and services. The principal change in the QPF process is that HPC’s role has been greatly expanded. The WFO’s reduced role consists of monitoring the HPC 6-h QPF product and coordinating needed modifications with the affiliated RFC or HPC. WFOs retain responsibility for issuing official NWS hydrologic forecast products, such as flood outlooks, watches, and warnings.

Among its increased duties, HPC now routinely produces the 6-h QPF product four times daily instead of twice daily, and, to meet the requirements for the western United States, these products have been extended from day 1 to include days 2 and 3 during the cool season of the year. The RFC role in the new process involves the issuance of nonscheduled updates to the HPC product (rather than to the former WFO product) for one or more of the early 6-h periods of day 1. Such updates, which are coordinated with HPC, are issued as rapidly changing weather conditions warrant. In essence, the new QPF process emphasizes an enhanced partnership between HPC and the RFCs.

A more indirect result of this study is that a national precipitation verification program has been inaugurated within the NWS. Since October of 2000, verification statistics for NWS QPF products for the preceding

month have been made available near the beginning of the following month. This rapid feedback to forecasters and various users of these products allows for constant performance monitoring and adjustments to the products as needed. The verification procedures and methodology developed in this study have been adopted in this program. Complete information concerning the verification program and continuously updated verification statistics were available online (<http://www.hpc.ncep.noaa.gov/npvu/>) at the time of writing. The verification statistics show that the new NWS QPF process is functioning in the manner envisioned, and the QPF product scores are generally improving with each forward step of the process from beginning to end. A follow-up article to demonstrate this encouraging result is pending.

9. Summary and conclusions

Comparative verification of operational 6-h QPF products that form the NWS QPF process was presented. These products included national QPF guidance from NCEP NWP model output and from human forecasters at HPC. It also included QPF products issued for local geographical areas by human forecasters at WFOs and RFCs. Extensive postprocessing of these diverse QPF products and of the verifying precipitation observations was required to achieve consistent scoring. In the eastern United States, the verification was conducted on a grid with a 30-km mesh, and in the western United States the verification was conducted on both this grid and at irregularly spaced points.

A significant component of the study was development of scoring techniques for ranking the accuracy of the various QPF products for NWS RFC streamflow models. Because the volume of precipitation runoff within drainage basins is modeled in this application, QPFs expressed as continuous amounts were scored ac-

cordingly. Also, because heavy precipitation amounts usually result in greater runoff than light amounts, judicious stratification of the scoring measures according to precipitation amount was applied. For the score stratification, it was found that mutually exclusive precipitation intervals provided a clearer picture of QPF error versus precipitation amount than did exceedance intervals. Also, in addition to the usual score stratification based on observed precipitation, corresponding stratification based on forecast precipitation was also applied, because the latter parameter is a major input to stream-flow models. For the purpose of ranking the overall accuracy of the QPF products, MAE for the two stratifications was objectively combined. This technique was especially helpful in ranking forecast accuracy when indications based on the separate stratifications were not consistent.

MAE and bias scores from the comparative verification of 6-h QPF products in the eastern United States for day 1 (0–24-h period) during the 1998/99 cool season showed that the HPC (manual) QPF guidance product performed slightly better than corresponding products issued for local areas by WFO and RFC forecasters. Still, the HPC QPF was only marginally better than the best-performing NCEP NWP model for QPF, which was the AVN Model. In the western United States for the 1999/2000 cool season, the WFOs achieved an improvement on HPC for day 1, but they failed to do so for days 2 and 3 (24–48- and 48–72-h periods, respectively). Also, both of these human QPF products showed an improvement on the AVN Model on day 1, but by day 3 neither of them improved on it. Comparison of the eastern and western U.S. scores revealed that the day-3 QPF scores in the West were better than day 1 scores in the East. This result indicates the reduction in QPF difficulty stemming from the topographic focusing of heavy precipitation by the major mountain ranges in the western United States overcompensates for a presumed increase in QPF difficulty arising from the upstream data-void Pacific Ocean.

The study findings and managerial considerations have led to two significant changes at NWS. One is that the human involvement in the process by which QPFs are produced for input into RFC hydrologic models was streamlined. In particular, the role of HPC has been increased, the partnership between HPC and the RFCs has been enhanced, and the WFO role has been reduced to monitoring and coordinating HPC and RFC QPF products. (WFOs nevertheless retain all responsibility associated with hydrologic forecast products.) A second, more indirect result of the study was the institution of a near-real time national QPF verification program at the NWS. The verification procedures developed herein were implemented in this program.

Acknowledgments. This study benefited from contributions from many persons in the NWS and beyond. Among them are members of the team selected to eval-

uate the NWS QPF process, who (in addition to the authors) were Thomas Graziano and Elizabeth Page of the NWS Office of Meteorology, Roger Pierce of the NWS Office of Hydrology, Rusty Pfof of NWS Southern Region, John Halquist of NWS Central Region, and Gregg Rishel of NWS Western Region. Also, Bill Lawrence of ABRFC; Mark Fenbers of OHRFC; Owen Rhea, Rob Doornbos, and Michael Ekern of CNRFC; Donald Laurine and Harold Opitz of NWRFC; Chris Hill of WFO Seattle; and Steve Chiswell of Unidata made various contributions that together include providing archives of the WFO and RFC QPF data, stage-III data, rain gauge data, assisting in accessing data archives, and engaging in useful discussions. Brett McDonald (one of the coauthors), who was a key worker in the objective verification aspect of this study, was supported in part by an appointment to the COMET Postdoctoral Fellowship Program and to the UCAR Visiting Scientist Program, which is sponsored by the National Weather Service and administered by the University Corporation for Atmospheric Research pursuant to National Oceanic and Atmospheric Administration Award NA97WD0082. Also, we thank NWS Director Jack Kelly, who initiated the study and commissioned the team; Greg Mandt, director of the NWS Office of Services who selected the team members; and NWS Meteorological Development Laboratory Director Harry Glahn for giving the lead author time away from normal duties to draft and revise the manuscript. The manuscript was strengthened by the comments of three anonymous reviewers; those from one reviewer were especially helpful for relating how aspects of the verification methodology were linked to previous studies.

REFERENCES

- Antolik, M. S., 2000: An overview of the National Weather Service's centralized statistical quantitative precipitation forecasts. *J. Hydrol.*, **239**, 306–337.
- Black, T., 1994: The new NMC mesoscale model: Description and forecast experiments. *Wea. Forecasting*, **9**, 265–278.
- Bosart, L. F., 1980: Evaluation of LFM-2 quantitative precipitation forecasts. *Mon. Wea. Rev.*, **108**, 1087–1099.
- Brooks, H. E., and C. A. Doswell, 1996: A comparison of measures-oriented and distributions-oriented approaches to forecast verification. *Wea. Forecasting*, **11**, 288–303.
- Burnash, R. J. C., 1995: The National Weather Service river forecast system-catchment model. *Computer Models of Watershed Hydrology*, V. P. Singh, Ed., Water Resources Publications, 311–366.
- Charba, J. P., 1998: The LAMP QPF products. Part I: Model development. *Wea. Forecasting*, **13**, 934–965.
- , and W. H. Klein, 1980: Skill in precipitation forecasting in the National Weather Service. *Bull. Amer. Meteor. Soc.*, **61**, 1546–1555.
- , Y. Liu, M. H. Hollar, B. Exley, and A. Belayachi, 1998: Gridded climatic monthly frequencies of precipitation amount for 1-, 3-, and 6-h periods over the conterminous United States. *Wea. Forecasting*, **13**, 25–57.
- Cherubini, T., A. Ghelli, and F. Lalaurette, 2002: Verification of precipitation forecasts over the Alpine region using a high-density observing network. *Wea. Forecasting*, **17**, 238–249.

- 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.
- , C. F. Mass, and K. J. Westrick, 2000: MM5 precipitation verification over the Pacific Northwest during the 1997–99 cool seasons. *Wea. Forecasting*, **15**, 730–744.
- 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.
- Fenbers, M. J., 1993: HASQPF user's guide. National Weather Service Eastern Region, NOAA, U.S. Department of Commerce, 57 pp. [Available from National Weather Service Eastern Region, 630 Johnson Ave., Bohemia, NY 11716-2626.]
- , 1995: WINQPF user's guide. NOAA Eastern Region Computer Programs NWS ERC-25MC, National Weather Service Eastern Region, NOAA, U.S. Department of Commerce, 57 pp. [Available from National Weather Service Eastern Region, 630 Johnson Ave., Bohemia, NY 11716-2626.]
- Fread, D. L., and Coauthors, 1995: Modernization in the National Weather Service river and flood program. *Wea. Forecasting*, **10**, 477–484.
- Fulton, R. A., J. P. Breidenbach, D.-J. Seo, D. A. Miller, and T. O'Bannon, 1998: The WSR-88D rainfall algorithm. *Wea. Forecasting*, **13**, 377–395.
- 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.
- Gyakum, J. R., and K. J. Samuels, 1987: An evaluation of quantitative and probability-of-precipitation forecasts during the 1984–85 warm and cool seasons. *Wea. Forecasting*, **2**, 158–168.
- Henkel, A., and C. Peterson, 1996: Can deterministic quantitative precipitation forecasts in mountainous regions be specified in a rapid, climatologically-consistent manner with Mountain Mapper functioning as the tool for mechanical specification, quality control, and verification? *Abstracts, Fifth National Heavy Precipitation Workshop*, State College, PA, NWS/NOAA, 31 pp. [Available from Office of Climate, Water, and Weather Services, W/OS, 1325 East West Hwy., Silver Spring, MD 20910.]
- Hoke, J. E., N. A. Phillips, G. J. DiMego, J. J. Tuccillo, and J. G. Sela, 1989: The Regional Analysis and Forecast System of the National Meteorological Center. *Wea. Forecasting*, **4**, 323–334.
- Junker, N. W., J. E. Hoke, B. E. Sullivan, K. F. Brill, and F. J. Hughes, 1992: Seasonal and geographic variations in quantitative precipitation prediction by NMC's Nested-Grid Model and Medium-Range Forecast Model. *Wea. Forecasting*, **7**, 410–429.
- Kanamitsu, M., and Coauthors, 1991: Recent changes implemented into the global spectral system at NMC. *Wea. Forecasting*, **6**, 425–435.
- Mao, Q., S. F. Mueller, and H. M. H. Juang, 2000: Quantitative precipitation forecasting for the Tennessee and Cumberland River watersheds using the NCEP Regional Spectral Model. *Wea. Forecasting*, **15**, 29–45.
- McDonald, B. E., 1998: Sensitivity of precipitation forecast skill to horizontal resolution. Ph.D. dissertation, University of Utah, 135 pp.
- Mesinger, F., 1996: Improvements in quantitative precipitation forecasts with the Eta regional model at the National Centers for Environmental Prediction. *Bull. Amer. Meteor. Soc.*, **77**, 2637–2649.
- , 1998: Comparison of quantitative precipitation forecasts by the 48- and by the 29-km Eta Model: An update and possible implications. Preprints, *16th Conf. on Weather Forecasting and Analysis*, Phoenix, AZ, Amer. Meteor. Soc., J106–J107.
- Murphy, A. H., 1991: Forecast verification: Its complexity and dimensionality. *Mon. Wea. Rev.*, **119**, 1990–1601.
- , and M. Ehrendorfer, 1987: On the relationship between accuracy and value in the cost-loss situation. *Wea. Forecasting*, **2**, 243–251.
- , and R. L. Winkler, 1987: A general framework for forecast verification. *Mon. Wea. Rev.*, **115**, 1330–1338.
- , B. G. Brown, and Y.-S. Chen, 1989: Diagnostic verification of temperature forecasts. *Wea. Forecasting*, **4**, 485–501.
- National Weather Service, 1999: The modernized end-to-end forecast process for quantitative precipitation information: Hydrometeorological requirements, scientific issues, and service concepts. National Weather Service, NOAA, U.S. Department of Commerce, 187 pp. [Available from Office of Climate, Water, and Weather Services, W/OS, 1325 East West Hwy., Silver Spring, MD 20910.]
- Olson, D. A., N. W. Junker, and B. Korty, 1995: Evaluation of 33 years of quantitative precipitation forecasting. *Wea. Forecasting*, **10**, 498–511.
- Reitan, C. H., 1974: Frequencies of cyclones and cyclogenesis for North America, 1951–1970. *Mon. Wea. Rev.*, **102**, 861–868.
- Ruth, D. P., 1992: The systematic interpolative radial search (SIRS)—a method to compute gridpoint values from contours. TDL Office Note 92-14, National Weather Service, NOAA, U.S. Department of Commerce, 16 pp. [Available from Meteorological Development Laboratory, W/OST2, 1325 East West Hwy., Silver Spring, MD 20910.]
- Schultz, P., and J. S. Snook, 1996: Objective verification of mesoscale model precipitation forecasts. Preprints, *15th Conf. on Weather Forecasting and Analysis*. Norfolk, VA, Amer. Meteor. Soc., J124–J126.
- Westrick, K. J., C. F. Mass, and B. A. Colle, 1999: The limitations of the WSR-88D radar network for quantitative precipitation measurement over the coastal western United States. *Bull. Amer. Meteor. Soc.*, **80**, 2289–2298.
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.