*EASTERN REGION TECHNICAL ATTACHMENT*
*NO. 91-5A*
*JUNE, 1991*

# VERIFICATION OF RIVER STAGE AND QUANTITATIVE PRECIPITATION FORECASTS

*Richard E. Arkell*
*Robert E. LaPlante*
*National Weather Service Forecast Office*
*Charleston, West Virginia*

## 1. INTRODUCTION

River stage forecasts have been routinely issued for years by National Weather Service Forecast Offices (WSFO's). More recently, some WSFO's have begun to produce quantitative precipitation forecasts (QPF). To date, relatively little automated verification of river forecasts has been conducted. In addition, a local automated QPF verification program only recently became available. Consequently, it has been difficult to objectively evaluate river forecasts, and to a lesser extent QPF, and provide useful feedback to the forecaster.

With the development of the River Stage Forecast Verification (RSFV) Program (LaPlante 1991), and the QPF Verification Program (Palko 1991), WSFO's now have the ability to evaluate both the river stage forecasts and QPF. The purpose of this study is to examine the performance of river forecasts, and QPF, in the WSFO Charleston forecast area, and suggest how these forecasts may be improved.

## 2. QPF VERIFICATION

WSFO Charleston has been issuing 24-hour QPF twice daily for 4 years. The Ohio River Forecast Center (OHRFC) in Cincinnati, OH, and the Lower Mississippi River forecast Center (LMRFC) in Slidell, LA, have been using QPF in their stage forecasts at all forecast points in the WSFO Charleston forecast area except those on the mainstem Ohio River. The OHRFC currently omits QPF generated flow from the Ohio because QPF is not currently forecast daily for the entire Upper Ohio basin.

The 1200 UTC QPF issued by WSFO Charleston were analyzed using the local QPF verification software. The program consists of two parts: a module that builds a data base automatically in AFOS on a daily basis; and a module to generate verification statistics that is run on demand. The first module extracts observed 6- and 24-hour rainfall amounts, and 6- and 24-hour rainfall estimates from WSFO Charleston's daily QPF. The verification statistics are categorical, with observed versus forecast frequencies for nine categories.

## 3. RIVER STAGE FORECAST VERIFICATION

Daily public river stage forecasts issued each morning by WSFO Charleston for 1, 2, and 3 days, were evaluated by using the RSFV program. Note, a national initiative is underway to verify flood forecasts categorically (i.e., no flooding; minor, moderate, major, near record, and record flooding (Morris 1988). This categorical approach, while well suited for flood crests, does not have the resolution to quantify errors and biases in the daily forecasts.

Like the QPF verification program, the RSFV software consists of two parts: a module that builds an AFOS data base automatically on a daily basis; and a

module to generate verification statistics that is run on demand. The first module extracts river stage observations and forecasts from an AFOS product in the format used by the OHRFC and LMRFC; non-daily (high water) forecast points are not verified. The system can store up to a year's worth of data. In addition, a SHEF (Standard Hydrometeorological Exchange Format) decoder was developed at Charleston (Young 1991) that decodes 3-day SHEF forecasts into the plain language format required by the RSFV program, thus making the verification software transportable to any office that issues SHEF forecasts.

Verification statistics are generated on demand by running the verification module. Statistics can be computed for a single station, a basin, or the entire forecast area. The scores for single stations include mean river stage, maximum observed and forecast stage, minimum observed and forecast stage, mean absolute error (MAE), and bias. MAE and bias are presented both in feet, and as a percentage of the mean observed stage during the verification period. Scores are calculated for all days, and also for those days when the observed stage equals or exceeds a preset "significant" stage that is usually defined as 1/2 or 3/4 bankfull. Consequently, statistics for "high water" events can be separated from overall scores.

## 4. QPF SCORES

The QPF verification program was run for the 1200 UTC, 24-hour QPF during the period July 2, 1990, to January 17, 1991. This is the same time interval used for the river stage verification. The data set includes a total of 2299 24-hour basin-average QPF forecasts. The results are shown in Tables 1 and 2.

Rainfall events with a quarter of an inch, or less, were forecast too infrequently; events of .26 to 1.00 inch were forecast with approximately the right frequency, while events with greater amounts were overforecast. Overall, 61% of the forecasts were in the correct category, and 90% of the forecasts were within one category. When the zero forecast QPF category is excluded, these values are 41% and 79%, respectively.

QPF skill can be measured by use of the Heidke Skill Score. Values of this score can range from -1 to 1, with 1 being perfect skill, while a score of 0 indicates no improvement over random chance. The WSFO Charleston forecasts were modestly skillful as indicated by a Heidke Skill Score of 0.386.

## 5. RIVER FORECAST SCORES

We will discuss overall river forecasting performance, and then see how forecasts vary between different size basins. For point of reference, average stages for individual stations during the study period ranged from 17 to 29 ft on the Ohio River, and from 1.25 to 16 ft for the headwaters. Maximum observed 7 am stages on the Ohio River ranged from 38 to 52 ft, or from 3 ft below flood stage to 8 ft above. Maximum observed 7 am stages on the headwater rivers ranged from 6 to 29 ft, or from 13 ft below flood stage to flood stage.

Charleston verifies 27 stations in its forecast area. The data set analyzed in this report consists of 156 days of forecasts, and includes a total of 12,082 individual stage forecasts. The results are shown in Table 3 in both percentage values (percentage of the mean observed stage) and absolute values (feet). Percentage values are not strictly proportional to absolute values due to the averaging method.

### 5.1. Overall River Forecasting

For all days, and all forecast points combined, the MAE increased from day 1 to day 3 by a factor of two-thirds in percent and doubled in feet.

The biases for all days combined was slightly positive (overforecasting the stage height) for day 1, and became increasingly negative (underforecasting the stage height) for days 2 and 3.

The number of "high water" cases amounted to 7% of the total data set, or 849 individual forecasts. Generally, the MAE and bias values for high water cases were three to six times larger than for all days combined.

### 5.2 Mainstem, Midpoint, and Headwater Categories

The overall data set was broken into three subsets to investigate differences in MAE and bias based on basin size. Mainstem river forecast points (for which QPF is not incorporated into the RFC river models), were defined as the 8 points on the Ohio River which have drainages of 26,000 to 56,000 sq mi. Midpoint forecast points consisted of the four locations with 2,500 to 10,500 sq mi of drainage. For the purposes of this paper, the headwater points comprised the 15 locations with 220 to 1,500 sq mi of drainage. Since the verification statistics for the midpoint category were basically between the mainstem and headwater categories, we will limit our discussion to the two extremes.

Referring to Table 3 for headwater and mainstem points, we see that the MAE was considerably higher in percentages, for headwater points. However, the mainstem points had higher MAE's (in feet) for all days combined, while the values were similar during "high water" days. MAE's were highest at both mainstem and headwater points (in both percent and feet) on day 3 during "high water."

The different signal inherent in the verification statistics for percent and feet illustrates the need to normalize these data. A method was devised in which the MAE and bias in units of feet were converted to units of cubic feet per second (cfs) per square mile, thus making the verification statistics for different basin sizes somewhat comparable. This was done by using the rating tables (which relates stage heights to flow rates), to determine the change in cfs, per foot change in stage, at each forecast point, for both the mean daily stage and for the mean "high water" stage during the study period. Based on this information, along with the corresponding drainage area, normalized MAE's and biases were derived. The technique is illustrated in Table 4.

Area-normalized verification statistics were calculated for all cases (Table 5). The results are similar to the percentage statistics, but are even more pronounced. The area-normalized MAE for headwater points was considerably higher than for mainstem points, especially during "high water" events, when it was approximately 10 times larger.

The biases generally followed the same pattern of increasing with projection (Table 3) as for the MAE's. One exception was the +4% bias for headwater points on day 1. This may be a reflection of the positive bias in 24-hour QPF. Overall, the percentages emphasized larger biases for headwater points than the mainstem, especially during "high water." However, the signal was again different when the statistics were examined in feet. For all days combined, the mainstem bias is twice as large as the headwater bias. The headwater bias in feet is still greater for high water events, but not as much. This reaffirms the need to normalize the data.

The area-normalized bias statistics (Table 5) support the percentage statistics. The largest negative biases were at headwater points during high water. On day 3, the bias for headwater points during high water was -11.6 cfs/per sq mi, over 12 times the value of -0.9 cfs/sq mi for mainstem locations.

### 5.3 Individual Forecast Points

Statistics were also generated for all 27 individual forecast points. The results showed that the largest drainage areas in the mainstem sample exhibited the smallest forecast errors. For example, Huntington, WV, with the largest drainage area (56,000 sq mi), had MAE's and biases of about two-thirds those for all the mainstem forecast points combined.

On the other hand, the smallest drainage areas had the largest deviations. Camden-on-Gauley, with a drainage area of 236 sq

mi, had MAE's and biases which were generally 1 1/2 times larger than the combined headwater points.

## 6. OPTIONS FOR THE FUTURE

Continuing verification efforts are needed. Based on the results presented here, there are some options we can consider to address the negative bias in river stages. We can keep producing only a first day QPF, and make manual adjustments of stages at the WSFO for days 2 and 3. We can drop the second and third day forecasts for the smaller headwater points where MAE's and biases are high. Another possibility is to extend the QPF out to 48 or 72 hours. Of course, the impact of this longer range QPF on the stage forecasts would depend on the skill of the QPF, as well as other hydrologic considerations. Extending the QPF is possible because the OHRFC and LMRFC can currently handle QPF out to 72 hours in their models. In addition, the hydrologic models could be improved to better handle base flow and recession characteristics for the forecast hydrograph. Of course, these are just some possible options, and further study is needed.

## 7. CONCLUSIONS

The RSFV and QPF verification programs have provided an opportunity to quantitatively evaluate river stage forecasts and QPF, and to examine the relationship between them. Use of this software also has demonstrated the importance of statistical verification.

The river forecast verification results for this study revealed a substantial negative bias in the stage forecasts for days 2 and 3. Headwater points had the largest bias. This negative bias showed up not only in high water forecasts, where one would expect some "low side" bias due to the nature of extreme event forecasting, but also, to a lesser extent, in the overall forecasts. The 24-hour QPF indicated some skill; however, the QPF amounts greater than an inch were forecast too frequently.

Lastly, the results show the need to quantify forecast errors from different sources. We will not know the relative importance of QPF, versus limitations in the hydrologic models, until parallel model runs are made over an extended period of time with and without QPF. Such runs would be especially important given current efforts to develop QPF forecasting techniques, and to verify both QPF and river stage forecasts.

### References


LaPlante, R., 1991: Local river stage forecast verification. NOAA Eastern Region Computer Programs and Problems NWS ERCP, NWS Eastern Region Headquarters, Bohemia, NY (in press).

Morris, D., 1988: A categorical, event oriented, flood forecast verification system for National Weather Service hydrology. NOAA Technical Memorandum NWS HYDRO 43, NWS Office of Hydrology, Silver Spring, MD, 74 pp.

Palko, J., 1991: QPS Verification Software. NOAA Eastern Region Computer Programs and Problems NWS ERCP No. 47, NWS Eastern Region Headquarters, Bohemia, NY (in press).

Young, R., 1991: River forecast program (multiple SHEF files to public format). NOAA Eastern Region Computer Programs and Problems NWS ERCP, NWS Eastern Region Headquarters, Bohemia, NY (in press).

| | FORECAST CATEGORY (inches) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **OBSERVED CATEGORY (inches)** | 0 | .01-.25 | .26-.50 | .51-.75 | .76-1.00 | 1.01-1.25 | 1.26-1.50 | 1.51-2.00 | >2.00 | TOTAL |
| 0 | **1025** | 57 | 13 | 2 | 0 | 0 | 1 | 0 | 0 | 1098 |
| .01- .25 | 345 | **260** | 113 | 22 | 16 | 15 | 3 | 4 | 0 | 778 |
| .26- .50 | 7 | 72 | **69** | 41 | 20 | 6 | 4 | 14 | 3 | 236 |
| .51- .75 | 9 | 13 | 24 | **29** | 16 | 1 | 4 | 3 | 2 | 101 |
| .76-1.00 | 3 | 6 | 9 | 8 | **7** | 3 | 2 | 5 | 6 | 49 |
| 1.01-1.25 | 0 | 0 | 5 | 3 | 1 | **3** | 1 | 4 | 0 | 17 |
| 1.26-1.50 | 0 | 0 | 0 | 3 | 0 | 1 | **2** | 0 | 0 | 6 |
| 1.51-2.00 | 0 | 0 | 3 | 1 | 0 | 0 | 1 | **2** | 4 | 11 |
| > 2.00 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | **1** | 3 |
| TOTAL | 1389 | 408 | 236 | 109 | 61 | 29 | 19 | 32 | 16 | 2299 |

**Table 1.** QPF contingency table for WSFO Charleston for July 2, 1990 to January 17, 1991.

| Category (inches) | # Of Fcsts | # Of Obsvd | Bias (Fcst/Obsvd) | % Correct | % Within One Category |
|---|---|---|---|---|---|
| 0 | 1389 | 1098 | 1.27 | 74 | 97 |
| .01- .25 | 408 | 778 | .52 | 64 | 95 |
| .26- .50 | 236 | 236 | 1.00 | 29 | 87 |
| .51- .75 | 109 | 101 | 1.08 | 27 | 72 |
| .76-1.00 | 61 | 49 | 1.24 | 11 | 39 |
| 1.01-1.25 | 29 | 17 | 1.71 | 10 | 24 |
| 1.26-1.50 | 19 | 6 | 3.17 | 11 | 21 |
| 1.51-2.00 | 32 | 11 | 2.91 | 6 | 6 |
| > 2.00 | 16 | 3 | 5.33 | 6 | 31 |
| TOTAL | 2299 | 2299 | | 61 | 90 |

**Table 2.** QPF bias and percent correct scores for WSFO Charleston, July 2, 1990 to January 17, 1991.

**ALL FORECAST POINTS COMBINED**

| | | 24 HR | | 48 HR | | 72 HR | |
|---|---|---|---|---|---|---|---|
| | | (%) | (ft) | (%) | (ft) | (%) | (ft) |
| *All days* | Number of cases..... | 4119 | | 4011 | | 3952 | |
| | Mean absolute error.. | 9.3 | .51 | 12.4 | .78 | 15.2 | 1.04 |
| | Bias ..... | 2.3 | .02 | -3.0 | -.28 | -9.7 | -.72 |
| *High water* | Number of cases..... | 285 | | 283 | | 281 | |
| | Mean absolute error.. | 7.8 | 1.56 | 14.9 | 3.18 | 22.2 | 4.86 |
| | Bias..... | -2.5 | -.38 | -9.8 | -1.74 | -18.6 | -3.71 |

**MAINSTEM FORECAST POINTS**

| | | 24 HR | | 48 HR | | 72 HR | |
|---|---|---|---|---|---|---|---|
| | | (%) | (ft) | (%) | (ft) | (%) | (ft) |
| *All days* | Number of cases..... | 1223 | | 1191 | | 1172 | |
| | Mean absolute error.. | 3.3 | .67 | 6.0 | 1.21 | 9.1 | 1.83 |
| | Bias..... | -0.7 | -.14 | -2.7 | -.54 | -6.5 | -1.31 |
| *High water* | Number of cases..... | 193 | | 193 | | 193 | |
| | Mean absolute error.. | 4.7 | 1.48 | 9.6 | 3.07 | 15.3 | 4.94 |
| | Bias..... | -0.4 | -.08 | -3.3 | -1.05 | -10.2 | -3.29 |

**HEADWATER FORECAST POINTS**

| | | 24 HR | | 48 HR | | 72 HR | |
|---|---|---|---|---|---|---|---|
| | | (%) | (ft) | (%) | (ft) | (%) | (ft) |
| *All days* | Number of cases..... | 2294 | | 2234 | | 2202 | |
| | Mean absolute error.. | 12.6 | .46 | 15.9 | .63 | 18.9 | .74 |
| | Bias..... | 4.0 | .10 | -3.3 | -.18 | -12.0 | -.49 |
| *High water* | Number of cases..... | 67 | | 68 | | 66 | |
| | Mean absolute error.. | 16.5 | 1.91 | 29.9 | 3.81 | 42.4 | 5.17 |
| | Bias..... | -8.3 | -1.17 | -26.8 | -3.57 | -42.4 | -5.17 |

(Note: High water is defined as stage greater than or equal to the "significant stage.")

**Table 3.** River forecast verification for WSFO Charleston, July 2, 1990 to January 17, 1991.

Examples based on Mean Absolute Error (MAE) for day 1 during high water events:

| MAE (ft) | | Avg. change in cfs per 1 ft change in stage | | MAE (cfs) | | Avg. drainage area (sq mi) | | MAE (cfs/sq mi) |
|---|---|---|---|---|---|---|---|---|
| | | | | **MAINSTEM FORECAST POINTS** | | | | |
| 1.48 | x | 11,800 | = | 17,464 | / | 42,000 | = | 0.42 |
| | | | | **HEADWATER FORECAST POINTS** | | | | |
| 1.91 | x | 1,800 | = | 3,438 | / | 800 | = | 4.30 |

**Table 4.** Basin normalization method for mean absolute error and bias.

| | | **MAINSTEM FORECAST POINTS** | | |
|---|---|---|---|---|
| | | **24 HR** | **48 HR** | **72 HR** |
| | | | (cfs/sq mi) | |
| *All days* | Mean absolute error..... | .19 | .34 | .51 |
| | Bias..... | -.04 | -.15 | -.36 |
| *High water* | Mean absolute error..... | .42 | .86 | 1.39 |
| | Bias..... | -.02 | -.30 | -.92 |
| | | **HEADWATER FORECAST POINTS** | | |
| | | **24 HR** | **48 HR** | **72 HR** |
| | | | (cfs/sq mi) | |
| *All days* | Mean absolute error..... | .52 | .71 | .83 |
| | Bias..... | .11 | -.20 | -.55 |
| *High water* | Mean absolute error..... | 4.30 | 8.57 | 11.63 |
| | Bias..... | -2.63 | -8.03 | -11.63 |

(Note: High water is defined as stages greater than or equal to the "significant stage.")

**Table 5.** Area normalized mean absolute error and bias for river stage forecasts for WSFO Charleston, July 2, 1990 to January 17, 1991.