9.4 DETERMINING QUANTITATIVE PRECIPITATION FORECAST DURATION TO OPTIMIZE RIVER FORECAST SERVICES

A. Juliann Meyer* and Lisa D. Holts
National Weather Service Missouri Basin River Forecast Center, Pleasant Hill, Missouri

Michael M. DeWeese and R. Holly Reckel
National Weather Service North Central River Forecast Center, Chanhassen, Minnesota

Noreen O. Schwein
National Weather Service, Lenexa, KS (retired)

1. INTRODUCTION

Flooding in the United States causes huge economic losses, personal property damage, and human fatalities. The 30-year average annual losses due to flooding from 1983 through 2012 were $8.17 billion in damages and 89 deaths per year (NOAA 2013). The National Weather Service (NWS) is charged with forecasting river levels as part of the NWS mission to protect life and property and enhance the nation's economy. River Forecast Center (RFC) hydrologists use hydrologic models to produce routine, daily forecasts and event-driven flood and water resources forecasts for the nation.

These models use a variety of meteorological parameters such as precipitation and air temperature, which can have a strong influence on the accuracy of river forecasts. To further the NWS's mission, there is a continuing need to improve the accuracy and increase the lead time provided by these forecasts by determining the best input to incorporate into the hydrologic model. One of these variables, Quantitative Precipitation Forecasts (QPF), can increase the lead time of a river forecast and provide earlier warning of a flood however, the amount, timing and location can be extremely challenging for NWS meteorologists to forecast, particularly as the forecast time period increases.

The purpose of this study was to investigate the errors in QPF as well as the impact of QPF on river forecast accuracy in order to determine the optimum number of hours of QPF to use in those forecasts. The study area encompassed a portion of the Midwest in the NWS Missouri Basin River Forecast Center (MBRFC) and North Central River Forecast Center (NCRFC) areas of responsibility.

Past studies have looked at the effect of QPF on river forecasts (Schwein 1996), the quality of the distribution and quantity of QPF (Cokely and Meyer 1991), and the impact of QPF on river forecasts and flood warnings (Meyer 2003).

* *Corresponding author address:* A. Juliann Meyer NWS/Missouri Basin RFC, Pleasant Hill, MO 64080; e-mail: julie.meyer@noaa.gov

In 2009, Schwein, Lander and Jones published "*Optimization of QPF Durations Used in River Forecasts,*" which recommended the number of hours of QPF to use in river forecasts was 24 from September through May (fall, winter, and spring) and 18 from June through August (summer). NWS Central Region management decision was to use 24 hours of QPF year-round as there was concern that changing the QPF duration in summer may be confusing for NWS partners and the general public. With that decision, the RFCs were given the option they could re-evaluate the decision at a later date. The Central Region RFCs revisited the issue with this study and two main goals:

- Evaluate the impact of the various QPF durations on river forecasts (the first study focused solely on precipitation errors but did suggest looking at river forecast errors as a follow-on study)
- Evaluate benefit of longer lead times using newly acquired data with durations out to 72 hours (the first study was limited to 24 hours)

The study was designed to address the above goals by meeting the following objectives:

a) Analyze QPF errors in 6-hour time increments similar to the 2009 study except on a hydrologic forecast basin scale instead of a 4x4 km Hydrologic Rainfall Analysis Project (HRAP) grid, using statistical methods similar to the 2009 study.
b) Analyze QPF errors through the analysis of 6-hour incremental QPF errors, as well as 6-to72-hour cumulative QPF errors
c) Identify the optimal number of QPF time periods based upon river forecast verification metrics that balance river forecast lead time with river forecast accuracy.

These three objectives were approached though a 3-part methodology. The first part was done similar to the 2009 study looking at positive QPF errors greater than 2.54 mm (0.10 inch) for each 6-hour time step.

The second part expanded on that, using positive and negative errors greater than 0.254 mm (0.01 inch) for both individual 6-hour increments as well as accumulated errors through 72 hours. Part 3 focused on river forecast errors using forecasts that included QPF time periods from 0 to 72 hours. Conclusions were drawn based on the use of statistical significance testing, quantification of greatest relative change in errors (e.g., increasing error with increasing time periods) for Mean Absolute Error (MAE) and Mean Error (ME), and comparisons of Probability of Detection (POD) with Hydrologic False Alarm Rate (HFAR) and Mean Error.

2. BACKGROUND

There are two RFCs located in the NWS Central Region. The North Central River Forecast Center provides forecasts for United States portions of the Hudson Bay drainage, Souris, miscellaneous streams and rivers into the Great Lakes, and the upper Mississippi River basin (NCRFC), while the Missouri Basin River Forecast Center (MBRFC) forecasts its namesake and the St. Mary Basin in Montana. Both RFCs are in the Midwestern U.S. and have many hydrometeorological and geological similarities. As a result of the 2009 study, both RFCs began using 24 hours of QPF operationally year-round starting June 1, 2009. Prior to this, the two RFCs had differing local policy with regard to the use of QPF in river forecasts: NCRFC used 24 hours of QPF for the entire year, while MBRFC used 12 hours of QPF from April through September and 24 hours of QPF October through March.

In May 2009, Central Region Headquarters (CRH) and the two RFCs outlined the scope of this follow-up study. The original design of this study used QPF durations of 0, 6, 12, 18, 24, and 48 hours. Early in the data collection process, based on the recommendation for a QPF durations study in the NWS RFC Verification Team's Final Report (Demargne, 2009), the duration of 72 hours of QPF was added.

3. CLIMATE AND HYDROLOGIC SUMMARY

The study spanned 13 months from June 1, 2009 through June 30, 2010. In the June 2009-2010 time period, precipitation averaged slightly above normal across the Midwest (as depicted in Figure 1), except for the northern portions of Minnesota, Wisconsin and Michigan.

The study period started out with a mix of above and below normal precipitation across both the MBRFC and NCRFC. October of 2009 was the wettest on record for much of the region, with two to four times the normal rainfall in most of the study area as shown in Figure 2. September and November 2009 were much drier than normal. Many areas also had the snowiest winters on record from December 2009 to February 2010.

Precipitation was more variable across the study area during the first half of 2010 with several wet and dry areas distributed throughout. June of 2010 was another much wetter period for most of the region with 150-300% of normal as shown in Figure 3, breaking several monthly and daily rainfall records across Nebraska through Iowa and Illinois. Additional maps are available at http://water.weather.gov/precip.

FIG. 1. The Midwest as defined by the U.S. Census Bureau

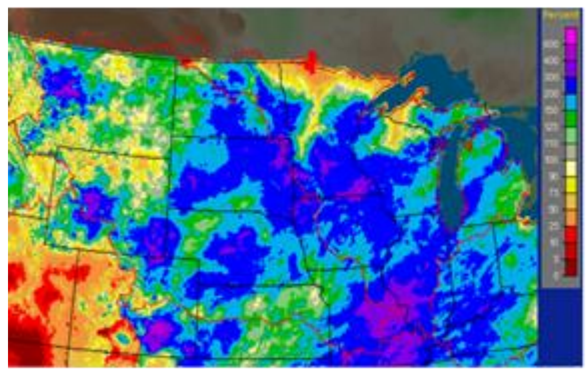

FIG. 2. Percent of Normal Precipitation for October 2009 was much wetter than normal across the study area

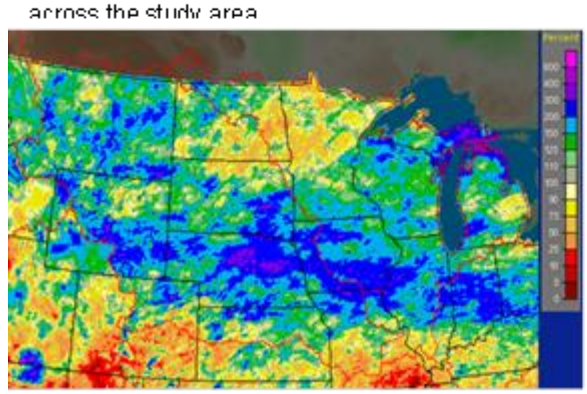

FIG. 3. Percent of Normal Precipitation for June 2010 was much wetter than normal across much of the study area.

The United States Geological Survey (USGS) historical stream flows for Water Year 2009 (Oct 2008 - Sep 2009) show normal to much higher than normal values for most of the Missouri and Upper Mississippi River basins. For Water Year 2010 (Oct 2009 - Sep 2010), USGS stream flows were also significantly above normal for the majority of Missouri and Upper Mississippi River basins. See Figures 4 and 5 below

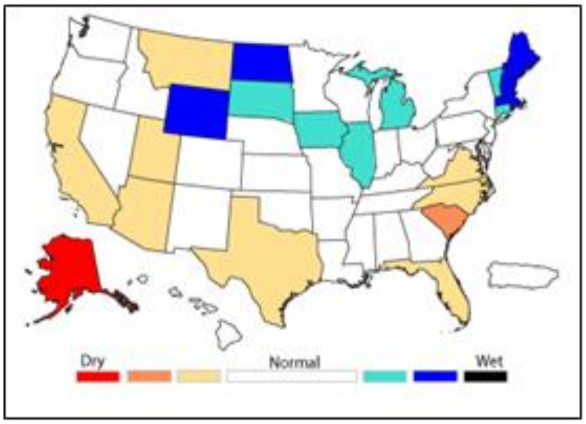

FIG. 4. USGS 2009 water year streamflow indicates near normal to above normal streamflow.

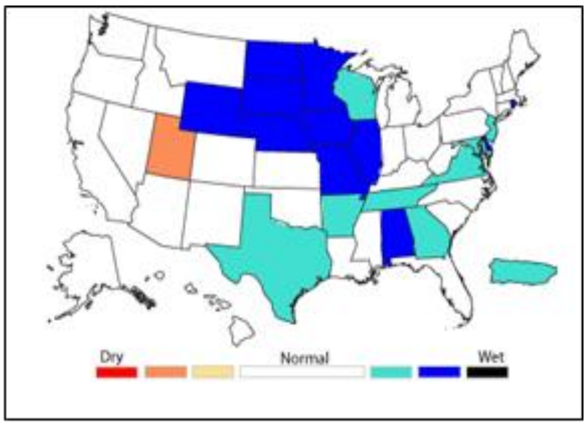

FIG. 5. USGS 2010 water year streamflow indicates much above normal streamflow.

4. DATA COLLECTION AND QUALITY CONTROL

The study was divided into three parts. The first two parts looked only at QPF errors using individual 6-hour time step precipitation data, accumulated precipitation data from six hours up to 72 hours, and 24-hour daily precipitation totals out to three days in the analyses. In Part 3, both precipitation and river forecast errors were analyzed. River forecast basins were selected from a wide geographic area that is representative of the forecast conditions that exist across both RFCs' areas of responsibility. For MBRFC, basins in Colorado, Wyoming, and Montana were excluded due to the mountainous terrain being quite dissimilar from NCRFC's area. To ensure an adequate sample size for Part 3, river forecast points were selected based on the frequency that observations reached Forecast Issuance Stage (FIS) or higher. It should be noted with this sampling constraint, the selected points were not evenly distributed but were concentrated in more hydrologically active geographic regions.

The precipitation data encompassed 1,164 basins in NCRFC's area and 906 basins in MBRFC's area. The morning RFC QPF data for twelve 6-hour time periods were collected for all three parts of this study: UTC 1200-1800, 1800-0000, 0000-0600 and 0600-1200 for Forecast Day 1-3. The QPF data collected were the Hydrometeorological Analysis and Support (HAS) QPF, which is the operational QPF used as input to river forecasts. The observed data collected was a combination of Mean Areal Precipitation (MAP) and multisensory (radar and rain gage) information (MAPX). Observed precipitation estimates used as ground truth in forecast operations vary throughout the year based mainly on weather conditions, such as icing, that can negatively impact radar precipitation estimates. MAP is calculated based on 24-hour precipitation reports and 6-hour synoptic station observations calculated over the basin areas using the Thiessen method (Chow 1964). MAPX is based on 1-hour radar quantitative precipitation estimate (QPE) over an HRAP grid spacing (Fulton 1998) and point-based hourly precipitation observations calculated over the basin areas. MAP and MAPX data were collected in time steps matching the QPF data. The QPF and MAP/MAPX gridded data values were then averaged to the river forecast basin areas with sizes averaging around 300 square miles.

At the time of this study, the CR RFCs did not use more than 24 hours of QPF in their routine forecasts. However, they did create model-driven contingency forecasts that incorporated 72 hours of QPF and produced an ensemble forecasts with a range of forecast stage levels in 6-hour time steps over a 5-day period. Besides having a greater number of forecasts to analyze, using these ensemble forecasts in this study had the additional advantage of producing more objective results in that no forecaster modifications to the hydrologic model were incorporated. Rather, errors resulted from changes in model inputs, such as QPF, thereby making QPF-based conclusions easier to cite.

Monitoring scripts were put in place to ensure the precipitation and river stage data were generated and posted to the RFC archive database. As the data had already been reviewed in operations, the main quality control for the precipitation data was ensuring the data successfully posted to the database. For the forecast stage data collected for Part 3, a more detailed review of the data was necessary since it had not been previously reviewed and occasionally, the raw model output would contain erroneous data. This raw model output contained all the forecasts used in this study. When erroneous data was found, it was eliminated from the dataset.

5. PART 1: PRECIPITATION ERROR ANALYSIS FOR INDIVIDUAL 6-HOUR TIME PERIODS

5.1 *Methods and Analysis*

Similar to the previous QPF optimization study by Schwein, Lander and Jones (2009), Part 1 analyzed QPF error datasets for each 6-hour period for forecast minus observed (F-O) errors where the forecast was at least 2.54 mm (0.10 inch). Statistical means for each time period of QPF errors were tested for significant differences in hopes of finding an obvious break where QPF errors increased significantly from one time period to the next. Due to the extremely large sample sizes, they were reduced to a random five percent sample of each dataset (Hamburg 1977) and then analyzed in a similar manner as the earlier study, by precipitation categories and seasons. While the 2009 study considered geographic regions within the RFCs, the regions in this study were defined by each RFC area as a whole, and then the two areas combined into a single dataset. The seasons were defined as fall (September 1 through November 30), winter (December 1 through the end of February), spring (March 1 through May 31) and summer (June 1 through August 31). Also similar to the earlier study, Part 1 of this study focused on the impact of over-forecasting; under-forecasting (F-O < 0) was not considered. Student's t-test for equal means was performed at the alpha=0.05 level for each 6-hour dataset to draw conclusions regarding significant differences between two QPF time periods.

MS Excel 2010 with the Data Analysis toolkit was used to compute the descriptive statistics and Student's t-test. Local computer applications were written to compile the datasets used in the MS Excel spreadsheet.

5.2 *Results and Conclusions*

Two-tailed Student's t-tests for equal means were conducted on the QPF errors for each forecast period and season on each RFC's dataset separately, followed by the two RFCs datasets combined. Unlike the 2009 study, significant differences were common within the first 12 or 18 hours. Beyond that, if summer data was extracted from the datasets, results were similar in that they indicated more QPF could be used (no significant differences until later periods). The t-test results for Part 1 did not indicate a clear break. Therefore, no conclusion could be drawn as to the optimal QPF duration to use.

6. PART 2: PRECIPITATION ERROR ANALYSIS ACCUMULATED THROUGH 72 HOURS

6.1 *Methods and Analysis*

This analysis also focused on the precipitation data only, but looked at all the data where the forecast and/or observation was >= 0.254 mm (0.01 inch). Datasets were created to accumulate precipitation totals for six to 72 hours. In addition to the individual time periods, accumulated precipitation was analyzed in two different ways: a) three 24-hour totals for each forecast day, and b) accumulated values from 6 hour total all the way to 72 hour totals.

The two RFCs examined both under- and over-forecasting, with forecast precipitation errors <= -0.254 mm (-0.01 inch) and errors >= 0.254 mm (0.01 inch) in the analysis, (i.e. eliminated pairs where the forecast minus observed (F-O) value was zero). The statistics utilized for the analyses were Mean Error (ME), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Pearson's correlation (CORR), and the categorical statistics Probability of Detection (POD) and Traditional False Alarm Rate (TFAR). The forecast precipitation error statistics were conditioned both by forecast and observed data in the following classes: MIN, 0.254 mm, 2.54 mm, 6.35 mm, 12.70 mm, 25.40 mm, and MAX (MIN, 0.01 inch, 0.10 inch, 0.25 inch, 0.50 inch, 1.00 inch, and MAX). The accumulated dataset (6 to 72 hours of QPF) with errors of <= 2.54 mm (-0.10 inch) and errors >= 2.54 mm (0.10 inch) was used in the t-tests. All three datasets were analyzed by season.

The RFCs used the NWS Interactive Verification Program (IVP) to pair the observed and forecast data values and analyze the precipitation with a variety of statistical metrics. MS Excel 2010 with the Data Analysis toolkit was again used for the Student's t-test and to generate charts.

Local computer applications were written to compile the pairs used by IVP for 1-, 2-, and 3-day precipitation totals, and 6- to 72-hour precipitation totals. The same local applications that were written for Part 1 were used to extract a random five percent dataset for the Student's t-test analysis of the accumulated (6- to 72-hour totals) precipitation dataset.

6.2 *Results and Conclusions*

While IVP provided a wealth of descriptive statistics (ME, MAE, POD, FAR) in this second part of the study, these statistics did not provide objective conclusive results in determining the optimal QPF duration that should be used routinely in the RFC's river models. Overall, the error statistics qualitatively showed little change moving forward in time. No clear break was seen where errors became much greater.

The Student's t-test was again used, but this time with the accumulated precipitation errors from 0 to 72 hours of QPF. The test was conducted for each accumulated total and season, on each RFC's dataset separately and then the two RFC datasets combined. Again, due to the large size of the datasets, a random five percent of each dataset was used in the analysis. For most combinations through the accumulating time steps, except for the summer data alone, the p-values were greater than the selected 0.05 alpha level, indicating no significant difference. Similar to Part 1 results, significant differences were shown in the first 12 hours. T-test results for the summer data showed more significant differences with an indicated break in the 36-48 hour time frame. Given the results thus far, no conclusion could be drawn as to the optimal QPF duration to use in river forecasts at NCRFC and MBRFC.

7. PART 3: PRECIPITATION AND STAGE ERROR ANALYSIS

7.1 *Methods and Analysis*

Unlike the previous two parts of the study, which looked at the precipitation data only, Part 3 examined the QPF along with the impact the various QPF durations had on the river stage forecasts. Data for all basins and forecast points, both precipitation and stage, were collected for the entire 13-month study period. QPF was analyzed and tested in a similar manner as to Parts 1 and 2. River stage locations used in the analysis were selected based on river forecast point response times (Fast, Medium and Slow), along with the number of events that occurred in the 13-month study period. For this study a minimum of 12 Fast (generally time to crest < 24 hours), eight Medium (time to crest >= 24 hours and < 60 hours), and four Slow (time to crest >= 60 hours) stations were selected from each RFC. Stations were selected based on greatest flood activity, or locations with the highest number of observed flood events. The selections were based on the number of actual observed events. Figure 6 shows the geographic distribution of stations.

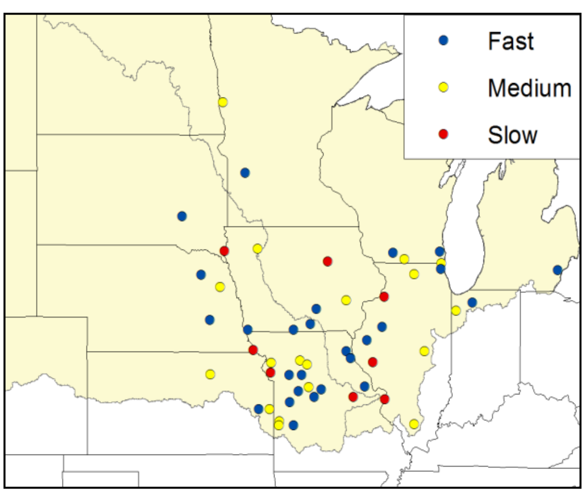

FIG. 6. Geographical Distribution of forecast points selected for this study.

As the forecast stage and flow variables are not completely independent of each other, the serial correlation nature of the data had to be taken into account for the t-test on river forecasts. To lessen the effects of serial correlation in assessing performance of river forecasts and recognizing that lack of correlation does not imply independence of events, the travel time from headwaters to forecast points was used as a lag between forecast/observation pairs. The longest travel time for Fast, Medium and Slow responding rivers was used for all data points in each category. For the analysis, only forecast observation pairs were used that were separated by at least the following lag times: Fast - four days, Medium - seven days, and Slow - 22 days. While it is recognized that not all correlation can be removed due to the day to day dependency of the numerous parameters input to river forecasts, this method was seen as the best way to remove the majority of it in order to conduct parametric statistical analyses on the river forecast/observation pairs.

7.1.1 *Student's t-Test on Precipitation Data*

Unlike the extremely large datasets of Parts 1 and 2, the precipitation dataset for Part 3 focused on data for the basins that feed the Fast and Medium response time locations used in the river stage forecast analysis. Basins for Slow response time locations were not considered as results would be very similar to the analyses in parts 1 and 2. For these t-tests the data were analyzed by season, and whether the forecast minus observed values were less than zero (F-0 < 0), indicating under-forecasting, or greater than zero, (F-O > 0 indicating over-forecasting. The results indicate that the optimal QPF time duration for forecasting river stages likely lies in the 24- to 48-hour range.

7.1.2 *Student's t-Test on River Forecasts*

While looking at the t-test on river forecasts, the Fast and Medium response time locations were analyzed separately for each RFC. The filtered dataset for the Slow response time rivers was quite limited. Therefore, data for the two RFCs were combined to one dataset. For each response time dataset, the river forecast errors were separated into two groups based on 1) the condition that the observation was greater than or equal to flood Stage (FS), and 2) the condition that the forecast was greater than or equal to FS. These groups were further separated by QPF time periods (0, 6, 12, 18, 24, 48 and 72 hours). T-tests for equal means were performed on the river forecast error datasets in those QPF time periods. As with the precipitation analyses in Parts 1 and 2, the results were inconclusive and other methods of analysis were used to determine the optimal QPF duration to incorporate into NCRFC and MBRFC's river models.

7.1.3 *Error and Categorical Statistics*

The main analysis developed by the RFCs was to use IVP statistical output to determine the optimal QPF. In addition, the RFCs also computed the percent difference, or relative change, for each statistic. All statistics calculated were combined as a whole, as well as divided by season. One general characteristic common in both RFCs that was also considered in the final conclusions was that the overall river forecast bias for each RFC was low. The scatter plots in Figures 7 and 8 show the river forecasts vs. the corresponding observations and the associated low forecast bias for each RFC. Under-forecasting is indicated below the diagonal, "perfect forecast", line while over-forecasting is plotted above the diagonal. The closer the forecast/observed pair plots to the diagonal, the better the forecast. Since precipitation inputs are one of the strongest drivers of the river model (Linsley, Kohler, and Paulhus, 1975), one likely reason for the low bias is the limited amount of QPF used (i.e., 24 hours) in the 1-5 day forecasts.

The study compared various IVP output for a particular statistic (e.g., MAE, ME), looking at the different QPF durations and computing the relative change of those error statistics from one QPF duration to the next. Plots of HFAR vs. POD by forecast lead time (1-5 days) and by season were also produced. The analysis used QPF durations of 6, 12, 18, 24, 48, and 72 hours and then 0, 24, 48 and 72 hours in side by side comparisons. The second analysis with equal time period intervals was performed due to concern that unequal time period intervals would adversely skew the conclusions. Examples of the various plots used in this analysis are shown in Figures 9 through 12. In general, when conditioned on the observation, error statistics decreased with increasing QPF while the opposite was true when conditioned on the forecast (Figures 9 & 11). In the relative change graphs, the annotated number over the bar was the "decision" QPF for that dataset. For results conditioned on the forecast (forecast category), the lesser number in the range of QPF hours (e.g., 24-48 hours of QPF) was used since errors generally increase in time when conditioned on the forecast. The higher number of hours was selected for the ranges in the observed category when errors typically decrease with increasing hours of QPF. Figure 10 shows the greatest change for spring season errors was from 24 to 48 hours, thus, 24 hours was selected. The other seasons suggest using no QPF. However, Figures 11 and 12 show spring and summer to be 24 hours and fall and winter, 48-72 hours.

The study also examined POD vs. HFAR. HFAR where the best value of HFAR is 0 and the best value of POD is 1. HFAR had a tendency to worsen (increase towards 1) with increasing QPF while POD improved (increase towards 1). In an attempt to determine a balance between the two, plots of HFAR vs. POD were created. Results for MBRFC (Figure 13) showed greater increase in POD compared to HFAR in the 0 to 24 and 0 to 48 hour range. For example, in Figure 14, notice Summer 2009 shows a more significant rise in HFAR compared to little improvement in POD going from 24 to 48 hour QPF, while Summer 2010 shows a similar trend but with fewer hours of QPF (12 to 18 hours). Fall shows more rise in HFAR going from 24 to 48 hour QPF compared to improvements in POD with 0 to 24 hour QPF.

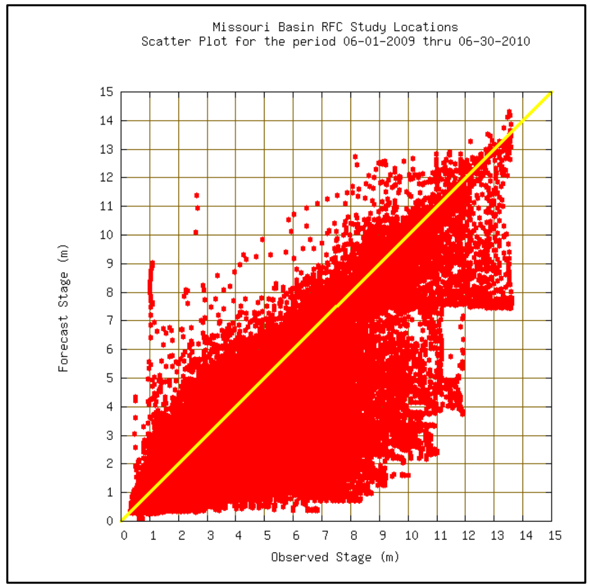

FIG 7. MBRFC scatterplot Forecast vs. Observed pairs indicates an under forecasting bias.

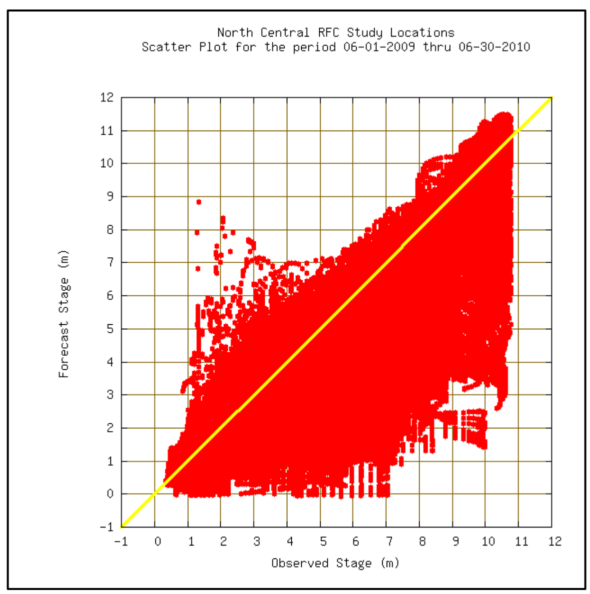

FIG. 8. NCRFC Scatterplot Forecast vs. Observed pairs indicated an under forecasting bias.

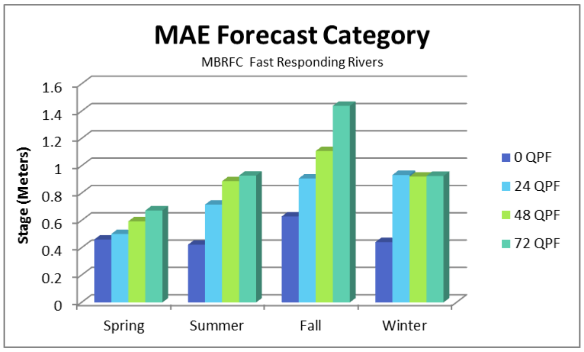

FIG. 9. MBRFC Mean Absolute Errors by Forecast Category indicates increasing error with increasing QPF duration.

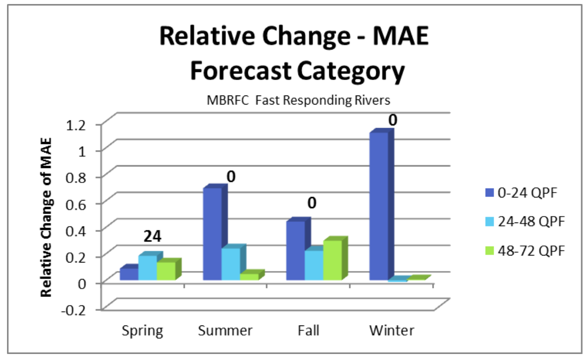

FIG. 10. MBRFC MAE Relative Change by Forecast Category indicates less QPF is better in river forecasts.

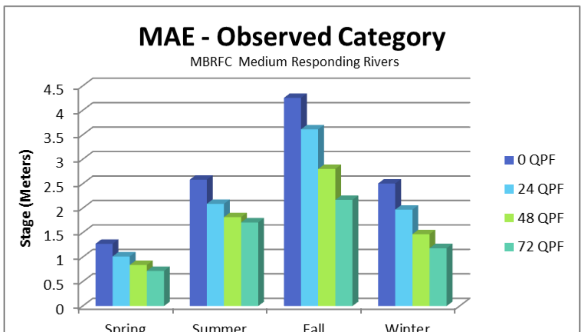

FIG. 11. MBRFC Mean Absolute Errors by Observed Category indicates decreasing error with increasing QPF duration.

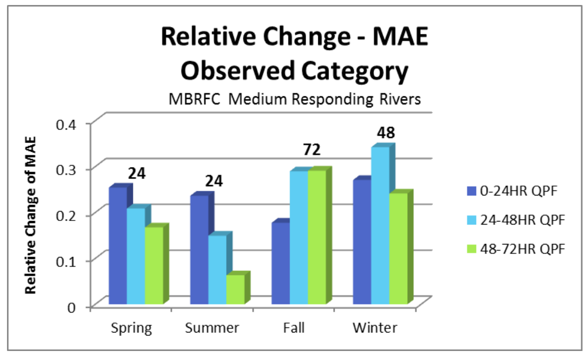

FIG. 12. MBRFC MAE Relative Change by Observed Category indicates more QPF is better in river forecasts.

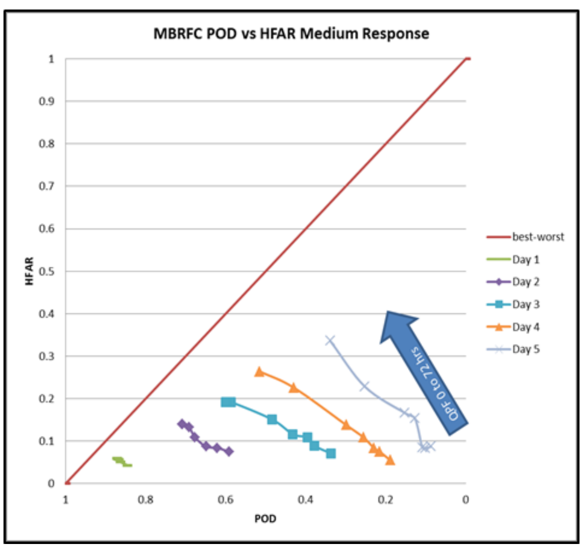

FIG 13. MBRFC - POD vs. HFAR by leadtime day for Medium response time locations shows a greater improvement in POD than hinderance in HFAR from 0-24 hrs QPF.

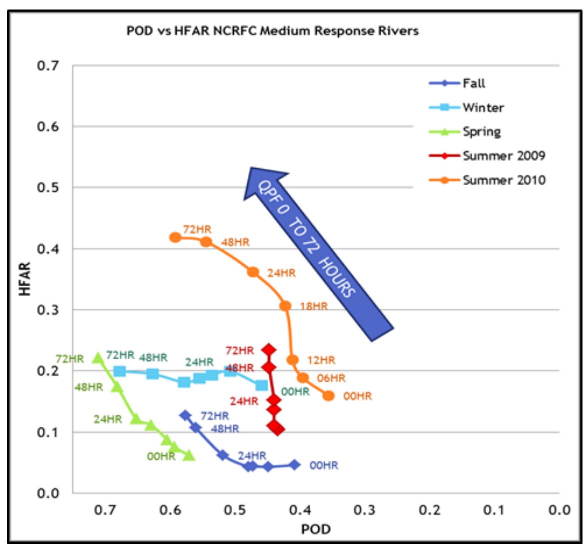

FIG. 14. NCRFC - POD vs. HFAR by seasons for Medium response time locations.

7.2 *MBRFC Results*

The conclusions in Table 1 are based on the review of all the data and graphs for part 3. For the MBRFC study area, it appeared the optimal QPF duration was in the range of 24 to 48 hours. During analysis, results were focused on the Fast and Medium response time locations, because Slow response time locations were not greatly affected by QPF proximity. That is, as long as it rains somewhere upstream, the water will contribute to the streamflow at the basin outlet.

For QPF durations 0, 6, 12, 18, 24, 48, and 72 hours, the relative change in the error statistics, ME and MAE, were considered. Results tended to indicate using 18-24 hours of QPF, when analyzing by forecast category (conditioned on the forecast). When analyzing by observed category, the data indicated 48-72-hour QPF was best to use. The categorical statistics, POD, again indicated that 48-hour QPF should be used, while HFAR showed 24-hour QPF was the best option.

When evenly distributing the QPF durations to 0, 24, 48, and 72 hours, the relative change from 0-24 hours QPF implied that using some QPF in the river forecasts overwhelmingly confirms the current philosophy on the operational use of QPF in river forecasts. The POD and error statistics by observed category both indicate 24-48 hour QPF being optimal. Conversely, the HFAR and error statistics by observed category show 0-24 hour QPF is optimal.

The study also looked at the POD vs. HFAR plots for Fast and Medium response time locations, to find a balance between under and over forecasting. Fast response time locations indicated using 24-hour QPF, while Medium response time locations favored 48-hour QPF.

Combining all the methods of analysis, the data indicated there was enough evidence to support using 48-hour QPF in the fall and winter, and 24-hour QPF during the spring and summer. It should be noted there was some evidence to support using 48-hour QPF duration in the spring. However, MBRFC concluded that the optimal QPF duration that should be used routinely in the river model was to continue using 24-hour QPF duration in the spring and summer, and to use a longer QPF duration of 48-hours in the fall and winter.

Considering…	MBRFC Summary of Conclusions
Considered QPF durations: 0, 6, 12, 18. 24, 48, & 72 hours • MAE by Forecast Category • ME by Forecast Category • HFAR • POD • MAE by Observed Category • ME by Observed Category	▪ Focused on forecasts above Flood Stage for Fast and Medium response time rivers ▪ Considered the raw statistics as well as relative change calculations ▪ Conclusion: optimal QPF duration is 24- to 48-hours QPF
Considered QPF durations: 0, 24, 48, & 72 hours • MAE by Forecast Category • ME by Forecast Category • HFAR • POD • MAE by Observed Category • ME by Observed Category	▪ Focused on forecasts above Flood Stage during study period for Fast, Medium, and Slow response time rivers ▪ Considered the raw statistics as well as relative change calculations ▪ Conclusion: optimal QPF duration is 24- to 48-hours QPF
Considered QPF durations: 0, 6, 12, 18. 24, 48, & 72 hours • POD vs. HFAR for Lead Time • POD vs. HFAR for Seasons	▪ Focused on Forecasts Above Flood Stage for Fast and Medium response time rivers ▪ Conclusion: optimal QPF duration is 24- to 48-hours QPF

TABLE 1. MBRFC summary of conclusions indicate that the optimal QPF duration for the MBRFC forecast area would be 24-48-hours.

7.3 *NCRFC Results*

Similar to MBRFC, NCRFC error analysis was conducted by conditioning both for observed and forecast categories to determine forecast discrimination and forecast reliability, respectively. Discrimination answers the question of when the observation is above flood stage, or when flooding is occurring, was it forecast? Reliability answers the question if the forecast was above flood stage, was the observation also above flood stage, or did flooding actually occur?

When conditioned on observations, both ME (bias) and MAE analysis showed a decreasing trend in bias and error increasing QPF durations as shown in Figure 15. Thus, the results clearly indicated an improvement in forecast discrimination when using longer periods of QPF.

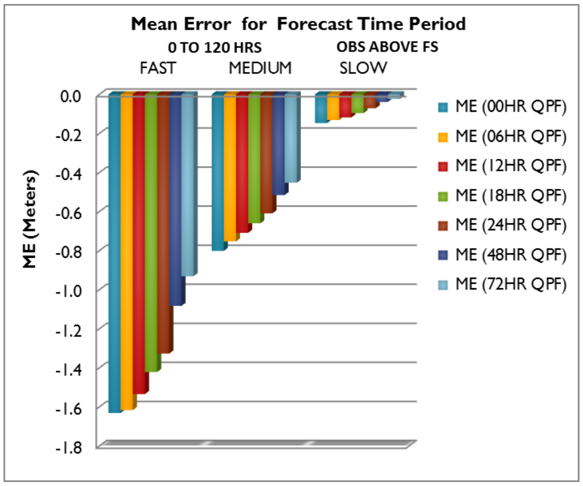

FIG. 15. NCRFC ME for observations above flood stage indicated improvement in forecast discrimination with more QPF.

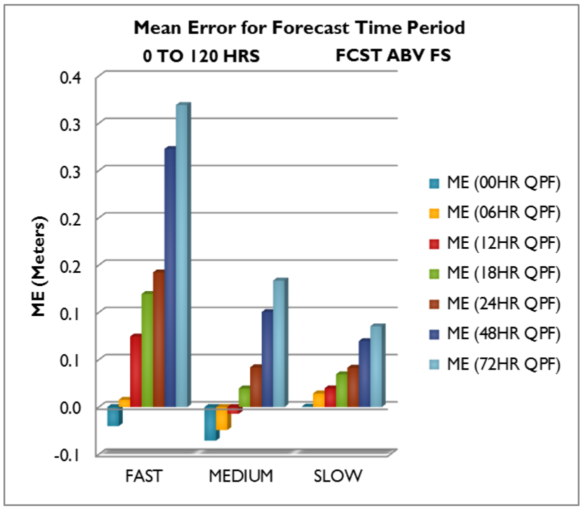

FIG. 16. NCRFC ME for forecasts above flood stage indicates increasing QPF duration reduces forecast reliability.

When conditioned on the forecasts, however, incorporating additional periods of QPF actually increased the error and bias as shown in Figure 16. So adding additional QPF actually decreased forecast reliability.

When conditioned by observations, the results show that there is a consistent improvement in bias with additional periods of QPF across all seasons (Figure 17). When looking at the results conditioned by forecast, the seasonal analysis clearly showed that summer exhibits much more bias than the remainder of the year (Figure 18). This seems logical given the climatology of the summer season, which is characterized by highly variable convective rainfall. Conversely, the fall and winter seasons when precipitation is predominantly stratiform, one can see relatively minor changes in bias with longer duration QPF.

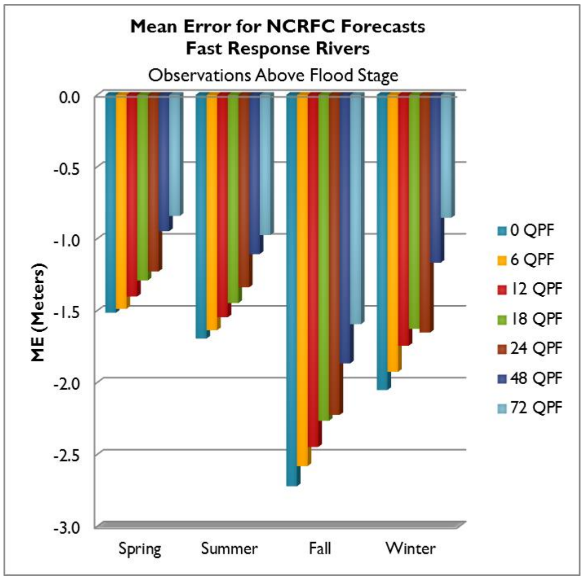

FIG. 17. NCRFC Seasonal ME for observations above flood stage indicates decreased bias with increasing QPF duration.

POD (Figure 19) and HFAR (Figure 20) both increased with longer QPF durations, yielding no obvious optimum QPF duration on initial review. In addition, analysis of the relative change in verification metrics showed no clear breakpoints either. Results of a comparison between POD and HFAR on a seasonal basis, however, indicated a higher summer rate of HFAR vs. POD using additional QPF beyond 24 hours. All comparisons of HFAR and POD during winter showed a greater increase in POD vs. HFAR using 48 and 72 hours of QPF.

Comparisons of POD vs. HFAR and Mean Error vs. POD over Day 1 through 5 lead times for fast, medium, and slow response time locations. Figure 21 demonstrates the results of POD compared to ME on a seasonal basis, with all QPF durations labeled as data points. Trends were examined to determine inflection points indicating where gains could be made without increasing negative effects. For example, the red line representing the summer of 2010 shows an improvement in POD between 18 and 24 hours of QPF. But the trend shows a large increase in ME with no improvement in POD between 24 and 48 hours of QPF.

Based on the combined results, the NCRFC analysis indicated that there were no consistent results on an annual basis. The subsequent seasonal analysis provided evidence for two recommendations, 24 hours of QPF in spring and summer, and 48-72 hours of QPF in fall and winter. Conclusions for NCRFC are summarized in Table 2.

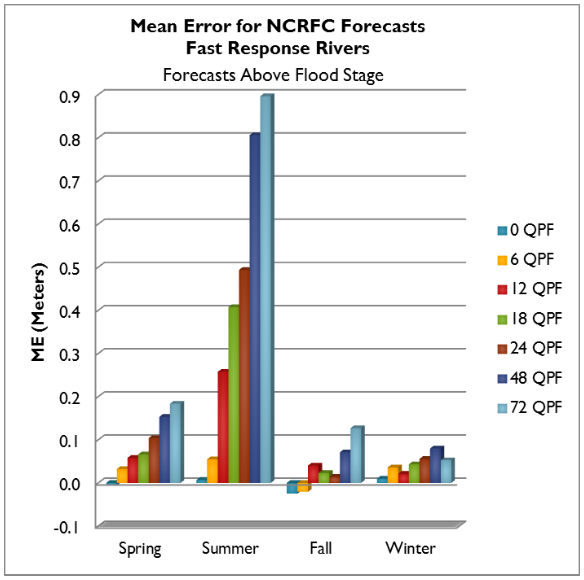

FIG. 18. NCRFC Seasonal ME for forecasts above flood stage shows much more bias in Summer forecasts.

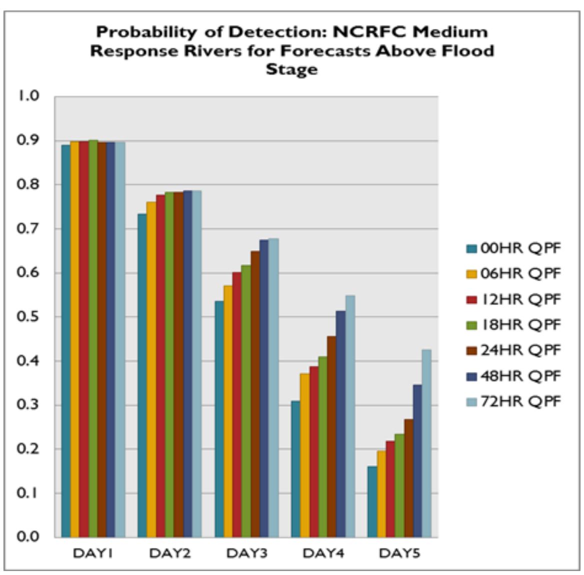

FIG. 19. POD for lead times Day 1 to 5 increases with increasing QPF duration for each lead time day.

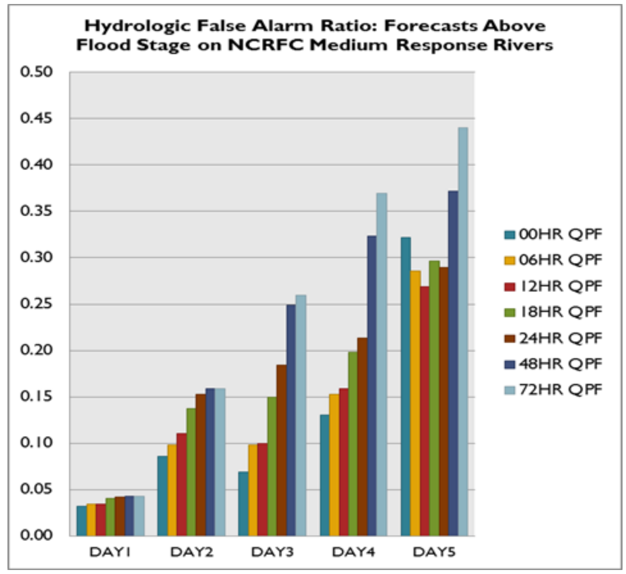

FIG. 20. HFAR for lead times Day 1 to 5 shows increasing error with increasing QPF duration.

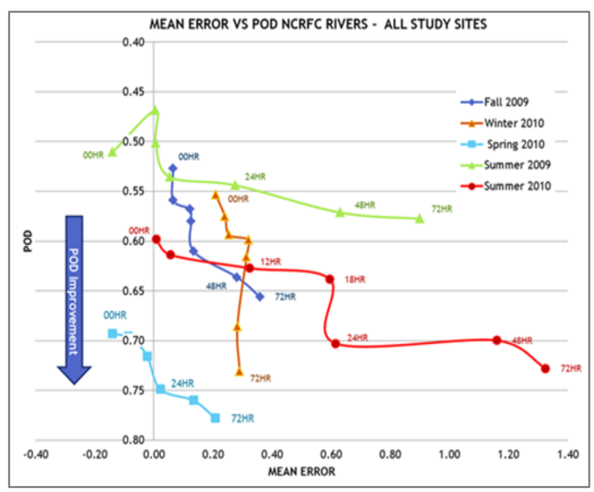

FIG. 21. Seasonal comparison of ME vs. POD for all NCRFC study sites shows improvement in forecasts during summer 2010 between 18 and 24 hours QPF.

Considering…	NCRFC Recommendation for best QPF Time Horizon…
POD vs HFAR POD vs ME HFAR vs ME Seasonality	• Focus on Forecasts Above Flood Stage for Fast and Medium Response Rivers… • Conclusion: 24 hours QPF Spring and Summer, 48hr-72 hours QPF Fall and Winter
ME vs POD over all lead times MAE vs. POD over all lead times	• Focus on Forecasts Above Flood Stage for Fast, Medium, and Slow Response Rivers and errors over lead times Day 1 to 5… • Conclusions: 24 hours QPF for Fast and Medium Response and 24 to 48 hour QPF for Slow Response.

TABLE 2. NCRFC summary of conclusions states that 24 hours QPF for Fast and Medium responding rivers and during Spring and Summer is optimal while more QPF can be handled during Fall and Winter as well as for Slow responding rivers.

8. SUMMARY AND RECOMMENDATIONS

Each River Forecast Center analyzed river stage forecasts for selected Fast, Medium, and Slow response time locations for 24-28 forecast points in its area of responsibility. This analysis involved computing several statistics including, but not limited to: Mean Error and Mean Absolute Error by both observed and forecast conditioning, POD, and HFAR. The two RFCs worked together to arrive at an objective approach to analyze the various plots. The two RFCs found:

- Overall low bias (under-forecasting) for points actively flooding
- For statistics conditioned on forecasts above FS, errors increased with increasing QPF duration (degrading forecast quality)
- For statistics conditioned on observations above FS, errors decreased with increasing QPF duration (improving forecast quality)
- POD increased with increasing QPF time duration (improving forecast quality)
- HFAR increased with increasing QPF time duration (degrading forecast quality)
- Significance testing (Student's t-test) was inconclusive

Analysis of the impact of the various QPF durations on the river stage forecasts indicated there was enough supporting evidence to recommend the following number of hours of QPF be used routinely in the river forecasts at Missouri Basin and North Central River Forecast Centers:

- 24 hours of QPF Spring and Summer (March 1 through August 31)
- 48 hours of QPF Fall and Winter (September 1 through the end of February)

It was further suggested the above dates not be hardwired but somewhat flexible depending on the conditions at the time. It would also be acceptable to exceed the routine hours of QPF when circumstances warrant, as currently practiced.

9. ACKNOWLEDGEMENTS

The authors wish to thank Kris Lander (CRH), Jennifer Bushur (MBRFC), and Pedro J. Restrepo (NCRFC) for their significant input and contributions to this manuscript. Julie DeMargne and James Brown provided consultation in water resources verification and statistics, and Hank Herr provided IVP technical expertise. We also appreciate the additional insights from Kevin Werner (CBRFC), Scott Dummer (MBRFC), Scott Lindsey (APRFC),Tom Adams (formerly of OHRFC), Tony Anderson (ABRFC), and David Welch (LMRFC) concerning QPF, river forecasts and statistics during this study.

REFERENCES

Chow, V., 1964: *Handbook of Applied Hydrology: A Compendium of Water-resources Technology,* McGraw-Hill Book Company, 1468 pp.

Cokely, D. A. and A. J. Meyer, December 1991: *A Comparison of Quantitative Precipitation Forecasts to Observed Mean Areal Precipitation* CR Tech. Attachment 91-26, 13 pp.

Demargne, J., September 2009: *Towards Standard Verification Strategies For Operational Hydrologic Forecasting - Report of the NWS Hydrologic Forecast Verification Team,* National Weather Service available at http://www.nws.noaa.gov/oh/rfcdev/docs/NWS-Hydrologic-Forecast-Verification-Team_Final-report_Sep09.pdf

Fulton, R., August 1998:_ *WSR-88D Polar-to-HRAP Mapping,* National Weather Service, available at http://amazon.nws.noaa.gov/articles/HRL_Pubs_PDF_May12_2009/HRL_PUBS_381-443/393_WSR-88D_POLAR_TO_HRAP_MAPPING.pdf

Hamburg, M., 1977: *Statistical Analysis for Decision Making 2nd Edition*, Harcourt Brace Jovanovich, Inc., 801 pp.

Herr, H., August 2007, NWS OHD, *Interactive Verification Program User's Manual*, 94 pp.

Linsley, R. K., M. A. Kohler, and J. L. H. Paulhus, 1975: *Hydrology for Engineers 2nd Edition,* McGraw-Hill Book Company, 482 pp.

Meyer, A. J. , June 2003: *Good QPF, Bad QPF: It's a Matter of Perspective*, USDA CMETO-362 course paper – unpublished, 10 pp.

NOAA National Weather Service Hydrologic Information Center: *Flood Loss Data*, available at http://www.nws.noaa.gov/oh/hic/

Schwein, N. O., K. S. Lander, and A. C. Jones, January 2009: *Optimization of Quantitative Precipitation Forecast time Durations used in River Forecasts,* Proceedings, 23rd Conference on Hydrology, AZ, Amer. Meteor. Soc., available at https://ams.confex.com/ams/pdfpapers/149707.pdf

Schwein, N. O., January 1996: *The Effect of Quantitative Precipitation Forecasts on River Forecasts,* NOAA CR Tech. Memorandum NWS CR-110, 39 pp.