

**2016 Forecast Verifications at the Lower Mississippi River Forecast
Center Using Varying QPF**

W. Scott Lincoln and David Schlotzhauer
Lower Mississippi River Forecast Center

July 2016

Table of Contents

Introduction	2
1.0 Verification of Long Range Mississippi River Forecasts with Two Durations of QPF	3
1.1 Overview	3
1.2 Methodology.....	4
1.3 Results.....	4
1.4 Operational Applications & Future Work	11
2.0 Verification of Headwater Basin Forecasts with Varying QPF	15
2.1 Overview	15
2.2 Past Work.....	15
2.3 Methodology.....	18
2.4 Results.....	20
2.5 Discussion.....	24
2.6 Operational Considerations and Future Work.....	24
3.0 References	25

Introduction

An important way to evaluate river forecast skill is through verification. During a verification analysis, forecasted values are compared to subsequent observations for that same point in time. Forecasts are typically compared to observations with a wide variety of lead times to better understand how forecast skill changes as an event gets closer or further away in time. Quantitative precipitation forecasts (QPF) that are used to create river forecasts can also be evaluated by utilizing numerous durations of QPF and comparing which river forecast provides the best results.

From summer 2015 through summer 2016, two forecast verification analyses were conducted by the National Weather Service (NWS) Lower Mississippi River Forecast Center (LMRFC). The first verification was for long range forecasts on the Mississippi River where forecasts utilizing two different QPF durations were compared. The second verification was for several smaller watersheds across the LMRFC area comparing a wide range of QPF durations. This report is broken into two separate sections which present the findings of each of these verification efforts.

1.0 Verification of Long Range Mississippi River Forecasts with Two Durations of QPF

1.1 Overview

For the Mississippi River, missed forecasts, both too high and too low, can have substantial costs to the economy, particularly the shipping industry. The most frequent requests from the shipping industry often are Mississippi River forecasts with additional lead time (forecasts that go farther into the future) and additional QPF. Forecasts with additional lead time are important because it can mitigate the costs related to low flow and high flow events. Forecasts with additional QPF are also important because operational forecasts which use 1 day of QPF (as of fall 2015*) drop off quickly after the first few days due to “dry weather” in the model. For a watershed as large as the Mississippi River, continuous dry weather for days rarely occurs except during rare drought conditions, making longer-range forecasts often verify much lower than observations.

In August 2013, the LMRFC began issuing experimental forecasts using 16 days of QPF. This QPF with 16 days of future rainfall came from the ensemble mean of the North American Ensemble Forecast System (NAEFS). These forecasts were issued once per week, and, contrary to normal operational forecasts, received only minimal quality control. In October 2015, a verification of these NAEFS-based forecasts was conducted as well as a verification of the corresponding operational forecasts which utilized only 1 day of QPF. Both forecasts go 28 days into the future. Hereafter, the 28-day forecast utilizing 1 day of QPF will be referred to as the “Operational” forecast and the 28-day forecast utilizing 16-days of QPF will be referred to as the “NAEFS” forecast.

**Operational forecasts for the Mississippi River began utilizing 2 days of QPF in spring 2016.*

1.2 Methodology

Verification software at the LMRFC creates pairs of forecast values and observation values for each forecast in the database, and then computes error statistics on those pairs. Verification pairs were created for the August 1st, 2013, through October 14th, 2015, period. Root mean squared error (RMSE), maximum error value (MAXERR), mean error (ME), and mean absolute error (MAE) were computed for lead times of 1, 2, 3, 4, 5, 7, 14, 21, and 28 days for 8 forecast locations on the Mississippi River.

1.3 Results

At every location analyzed, for most forecast lead times, the NAEFS forecast showed less error (RMSE) than the operational forecast (Table 1). Error statistics were often similar when looking at lead times of 1-5 days (Figure 1). By 7 days of lead time, NAEFS forecasts began performing better for most locations, and this improved skill continued for lead times up to at least 21 days. The reduction in error was less pronounced for most locations at 28 days lead time.

Another important statistic analyzed was the ME. ME provides a way of looking at forecast bias; where ME is near 0.0, forecasts have little bias (even if they significant error both above and below the observations). The ME statistic showed improvement with the NAEFS forecasts over the operational forecasts for all locations, especially for the 7-21 day lead time. MAE provides a way of looking at forecast error regardless of whether or not many forecasts overall had little to no bias. Daily forecasts alternating between 5ft below observations and 5ft above observations would yield 0.0 for ME but 5.0 for MAE; this makes MAE an important context for using ME in summarizing verification statistics. The patterns found in MAE were very similar to that of RMSE. ME and MAE statistics are summarized by Figure 2, Figure 3, Figure 4, Figure 5, Figure 6, Figure 7, Figure 8, and Figure 9.

Table 1. RMSE statistics for each of the 8 locations analyzed.

Lead Time	Cairo, IL		Memphis, TN		Arkansas City, AR		Vicksburg, MS		Natchez, MS		Red River Landing, LA		Baton Rouge, LA		New Orleans, LA	
	Oper.	NAEFS	Oper.	NAEFS	Oper.	NAEFS	Oper.	NAEFS	Oper.	NAEFS	Oper.	NAEFS	Oper.	NAEFS	Oper.	NAEFS
All	9.55	6.50	7.53	5.27	6.02	4.11	5.75	4.16	5.29	3.88	4.76	3.66	4.16	3.11	1.66	1.42
1	0.58	0.67	0.29	0.30	0.38	0.41	0.32	0.32	0.22	0.24	0.23	0.33	0.20	0.28	0.28	0.28
2	1.01	1.18	0.56	0.59	0.63	0.74	0.50	0.54	0.40	0.44	0.41	0.59	0.39	0.51	0.38	0.35
3	1.22	1.61	0.77	0.82	0.83	1.03	0.74	0.86	0.66	0.71	0.71	0.88	0.60	0.76	0.47	0.47
4	1.88	2.05	1.09	1.20	0.93	1.25	0.97	1.20	0.90	1.02	0.94	1.11	0.80	0.93	0.53	0.59
5	2.90	2.48	1.39	1.69	1.09	1.44	1.19	1.46	1.16	1.32	1.15	1.35	1.01	1.14	0.65	0.65
7	4.83	3.25	2.57	2.46	1.53	1.96	1.68	2.01	1.67	1.87	1.69	1.79	1.43	1.53	0.84	0.83
14	9.88	5.78	7.67	4.50	5.66	3.64	5.15	3.74	4.73	3.59	4.11	3.26	3.39	2.78	1.42	1.33
21	12.69	7.76	10.45	6.60	8.58	5.27	8.12	5.26	7.66	4.95	6.86	4.64	5.84	3.85	2.21	1.77
28	10.05	10.18	8.58	8.47	7.27	6.57	7.35	6.47	7.48	6.08	6.93	5.81	5.90	4.90	1.99	2.23

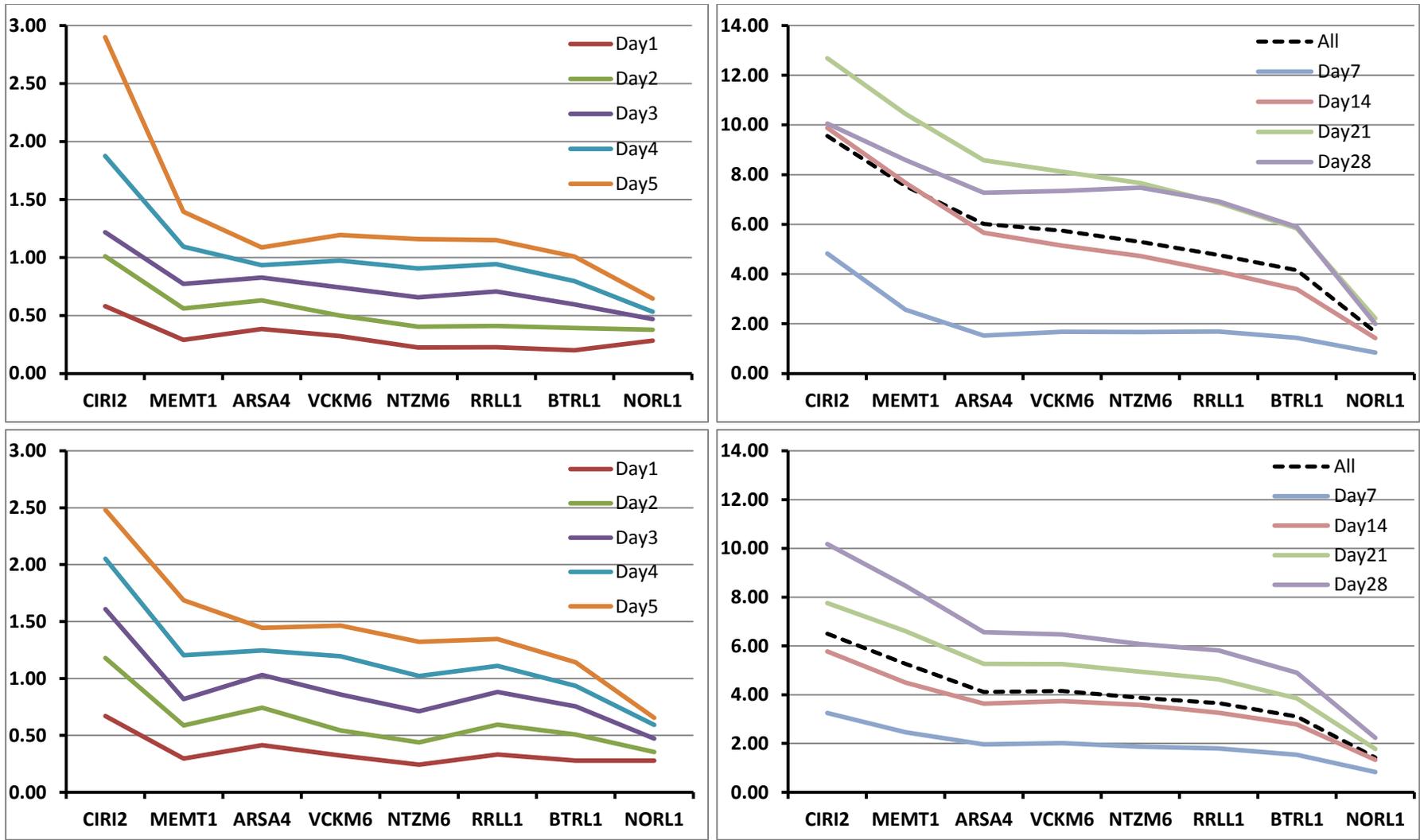


Figure 1. Verification results for the "Operational" (top) and "NAEFS" (bottom) forecasts for locations on the Mississippi River. RMSE statistics for both short lead times (left) and long lead times (right) were better with the NAEFS forecasts compared to the Operational forecasts. The dashed black line (right) indicates the RMSE statistics for all lead times combined.

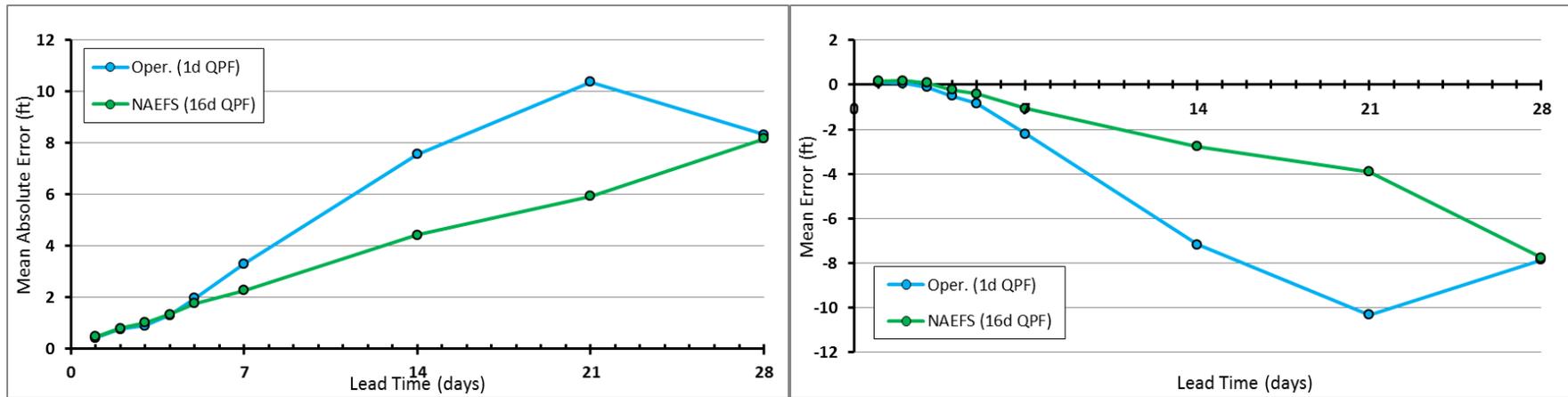


Figure 2. MAE (left) and ME (right) statistics for the Mississippi River at Cairo, IL, covering 1-28 days of lead time.

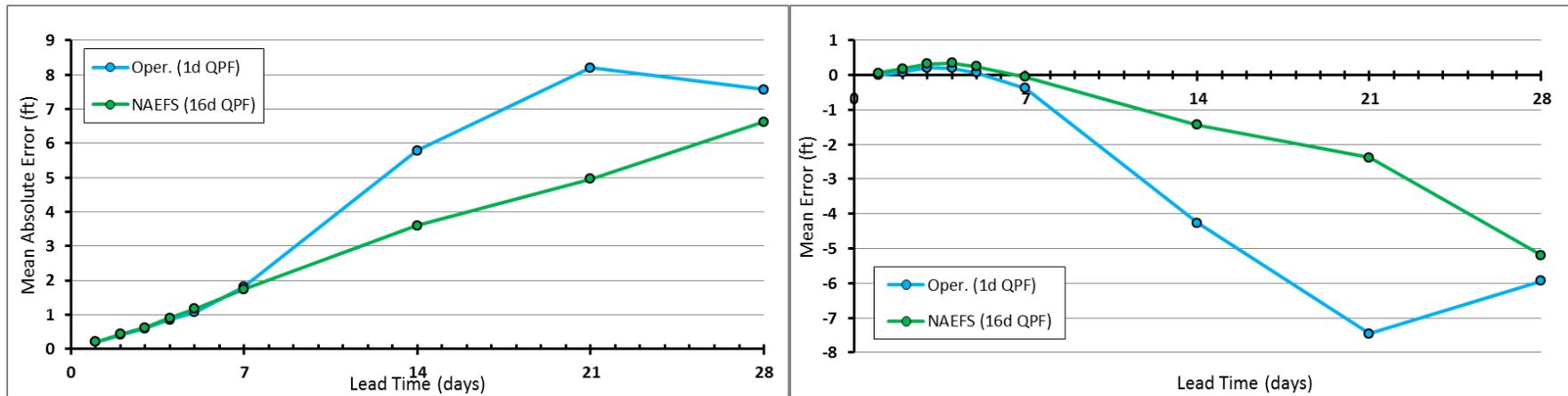


Figure 3. MAE (left) and ME (right) statistics for the Mississippi River at Memphis, TN, covering 1-28 days of lead time.

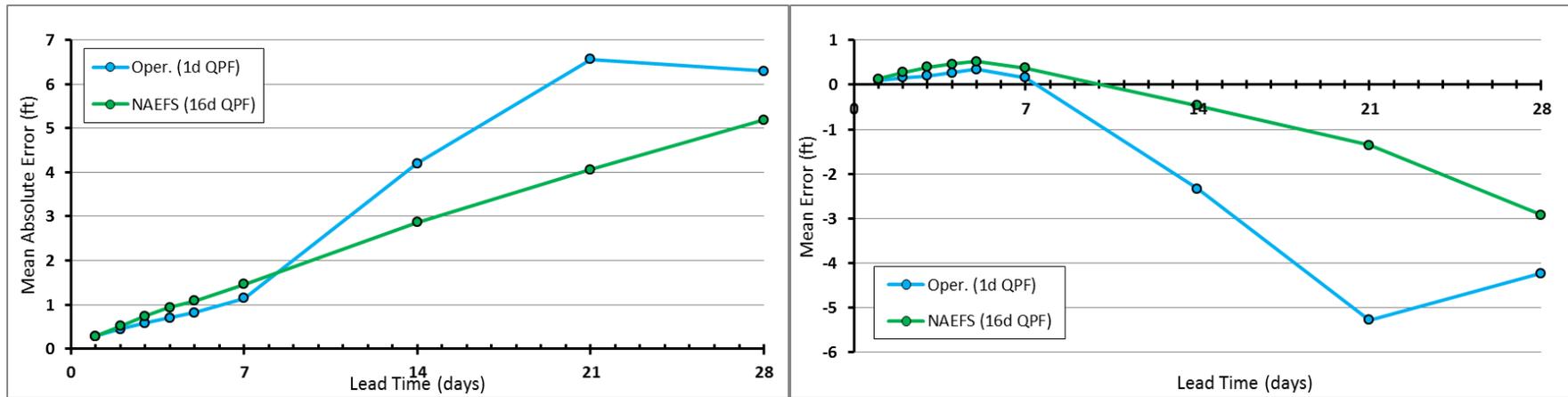


Figure 4. MAE (left) and ME (right) statistics for the Mississippi River at Arkansas City, AR, covering 1-28 days of lead time.

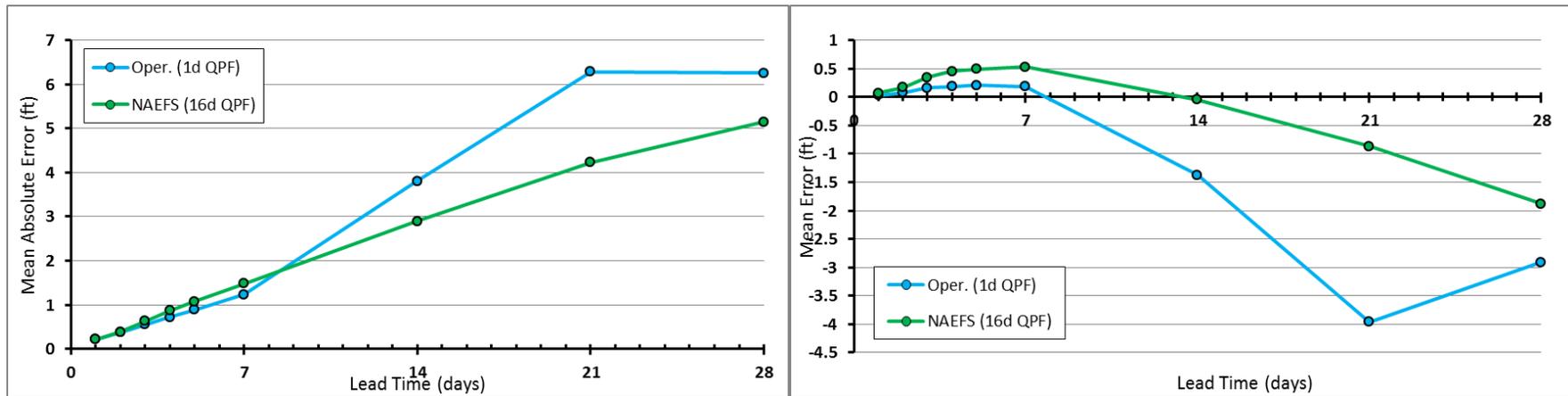


Figure 5. MAE (left) and ME (right) statistics for the Mississippi River at Vicksburg, MS, covering 1-28 days of lead time.

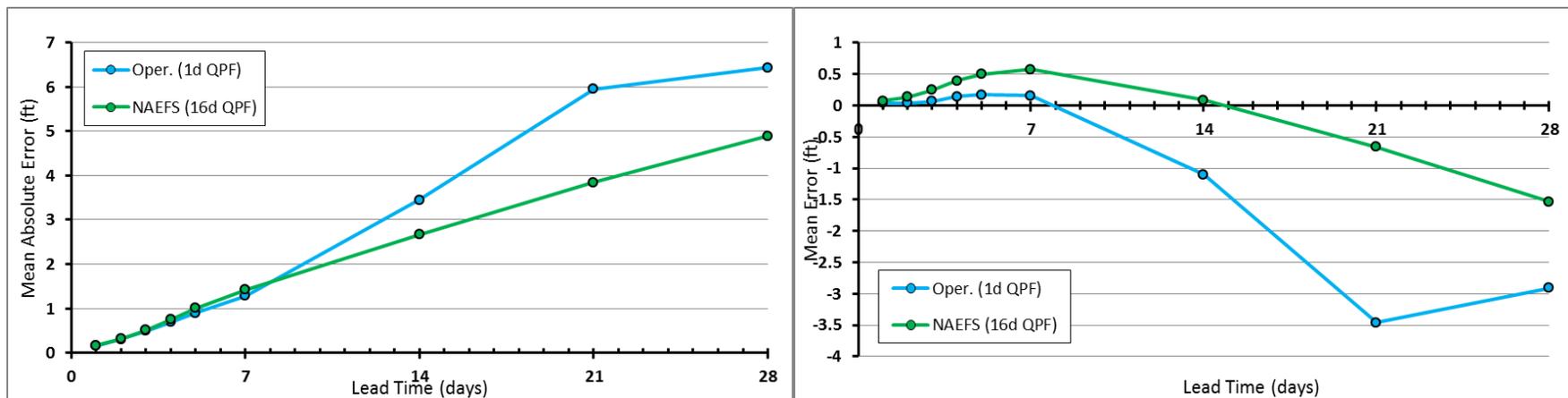


Figure 6. MAE (left) and ME (right) statistics for the Mississippi River at Natchez, MS, covering 1-28 days of lead time.

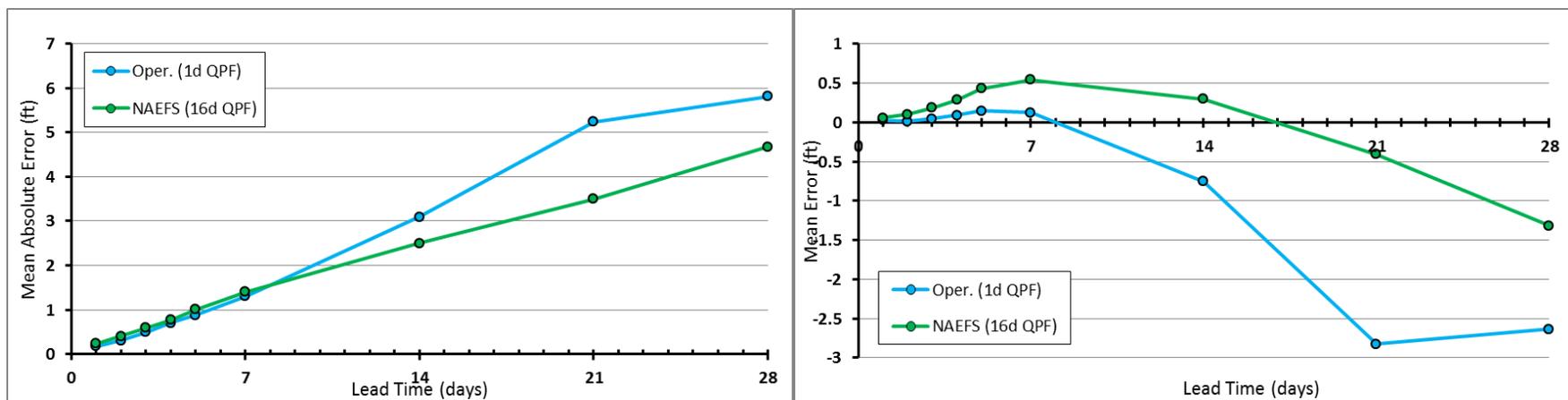


Figure 7. MAE (left) and ME (right) statistics for the Mississippi River at Red River Landing, LA, covering 1-28 days of lead time.

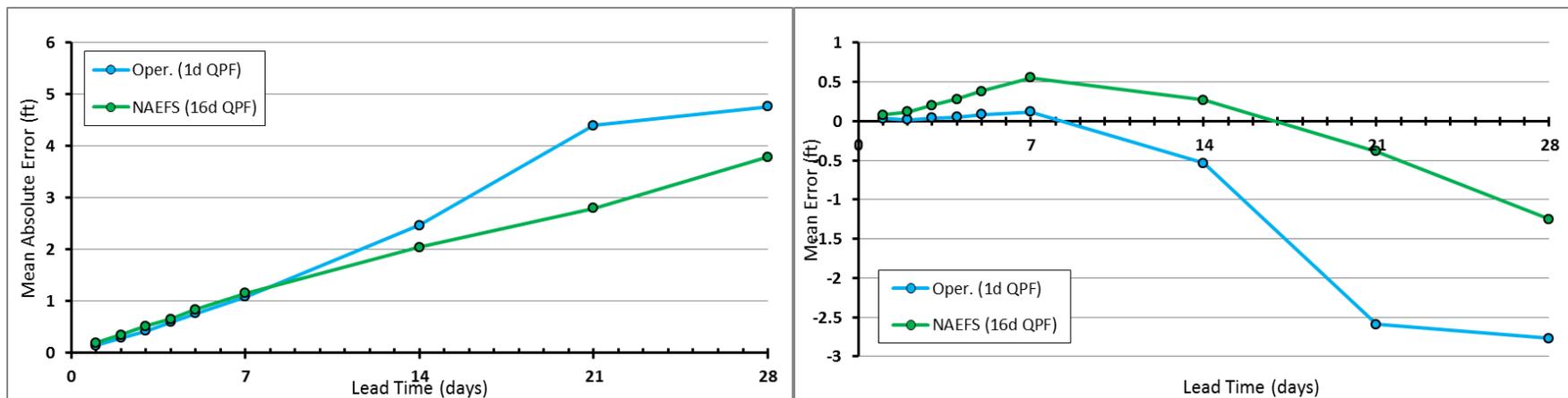


Figure 8. MAE (left) and ME (right) statistics for the Mississippi River at Baton Rouge, LA, covering 1-28 days of lead time.

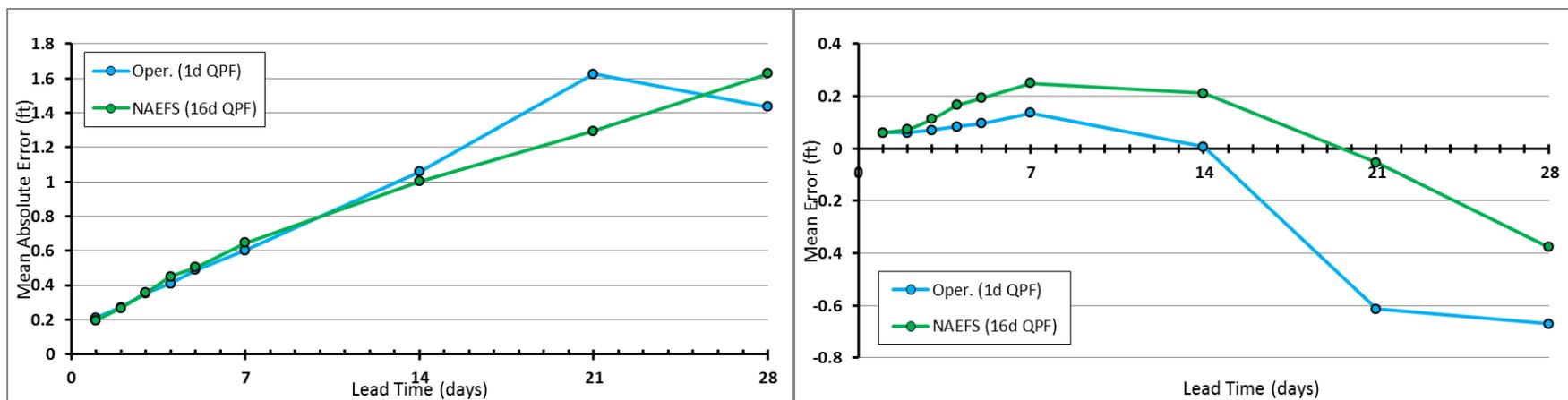


Figure 9. MAE (left) and ME (right) statistics for the Mississippi River at New Orleans, LA, covering 1-28 days of lead time.

1.4 Operational Applications & Future Work

One potential use of the verification statistics for the Mississippi River would be the ability to provide additional forecast uncertainty guidance to NWS partners. The deterministic forecast, either from the operational QPF or extended QPF from the NAEFS, could be provided along with a shaded area of uncertainty based upon the verification of past forecasts. A similar forecast visualization is provided by the National Hurricane Center (NHC) when they issue track forecasts for tropical cyclones. The shaded area, or “cone,” around the official forecast tracks are determined based upon past forecast errors.

An example application of this NHC forecast uncertainty method to Mississippi River forecasts is illustrated by both Figure 10 and Figure 11. Figure 10 shows forecasts with an uncertainty range, both above and below, determined by the magnitude of ME. Figure 11 shows forecasts with an uncertainty range determined by the magnitude of ME, but the uncertainty range is centered on the bias-corrected forecast based upon MAE.

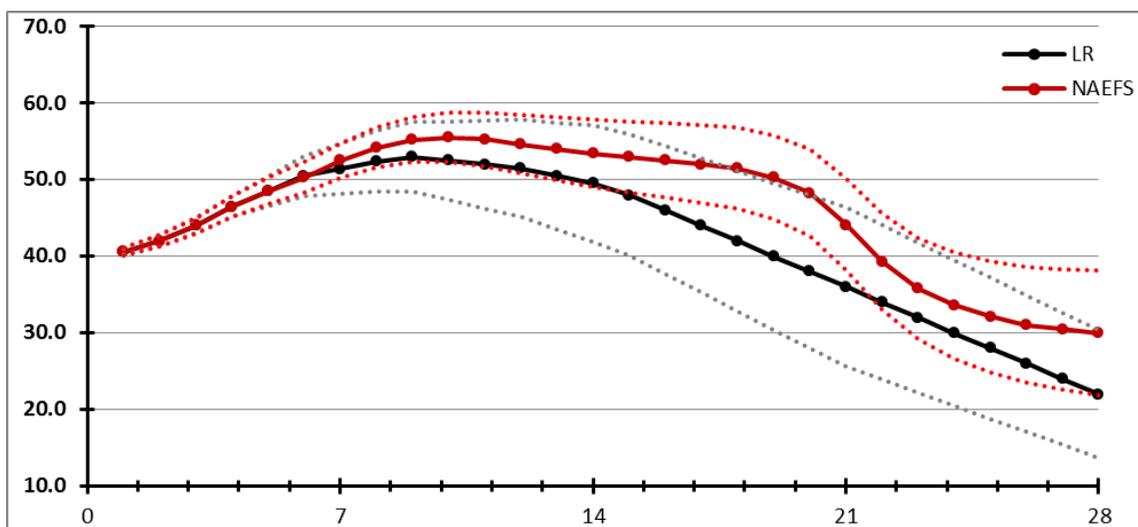


Figure 10. Hypothetical forecasts for the Mississippi River at Cairo, IL, from the 1 day QPF operational forecast (labeled "LR", black) and the 16 day QPF NAEFS forecast (red) based upon the 12/25/2015 operational forecast. Uncertainty is illustrated by the shaded lines on either side of the forecast, created by both adding and subtracting the ME from the forecast ordinate.

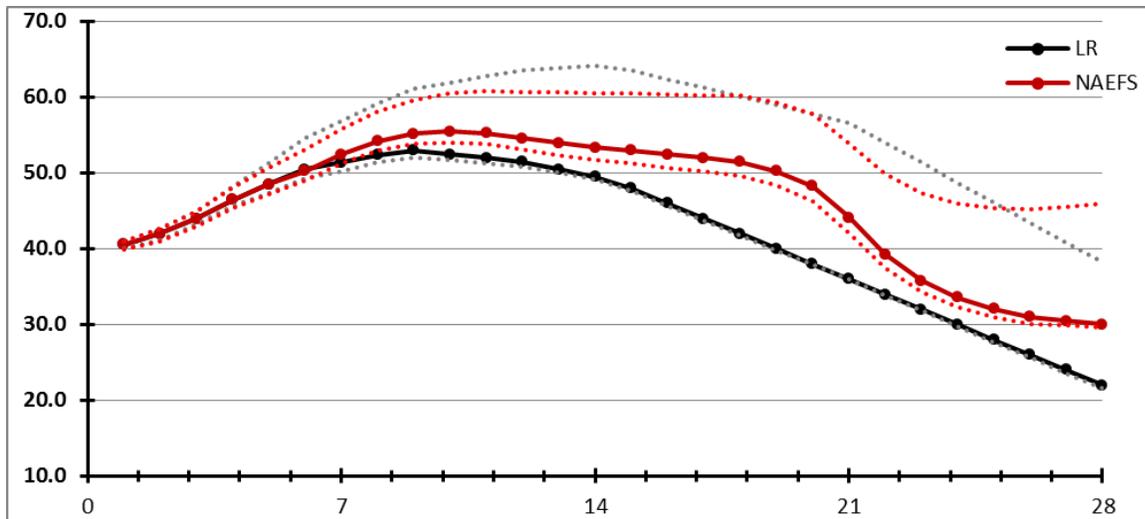


Figure 11. Hypothetical forecasts for the Mississippi River at Cairo, IL, from the 1 day QPF operational forecast (labeled "LR", black) and the 16 day QPF NAEFS forecast (red) based upon the 12/25/2015 operational forecast. Uncertainty is illustrated as in Figure 10 but the ME is added and subtracted to the forecast ordinates after they are bias-corrected using the MAE values for each forecast lead time.

This technique was also used to brief close partners to the LMRFC on the potential for significant flooding on the Mississippi River in late December 2015 through early January 2016 (Figure 12). Based upon the verification analysis results, LMRFC forecasters were able to provide a forecast with additional QPF, and also provide some context to that forecast (on average, with 7 days lead time, this forecast verifies within +/- x.x feet). Because statistics show significant reduction in forecast error using additional QPF from the NAEFS (even with minimal human QC of resulting forecasts), these types of internal forecasts to partner agencies should continue to be provided. Feedback from our partners should be evaluated to determine usefulness.

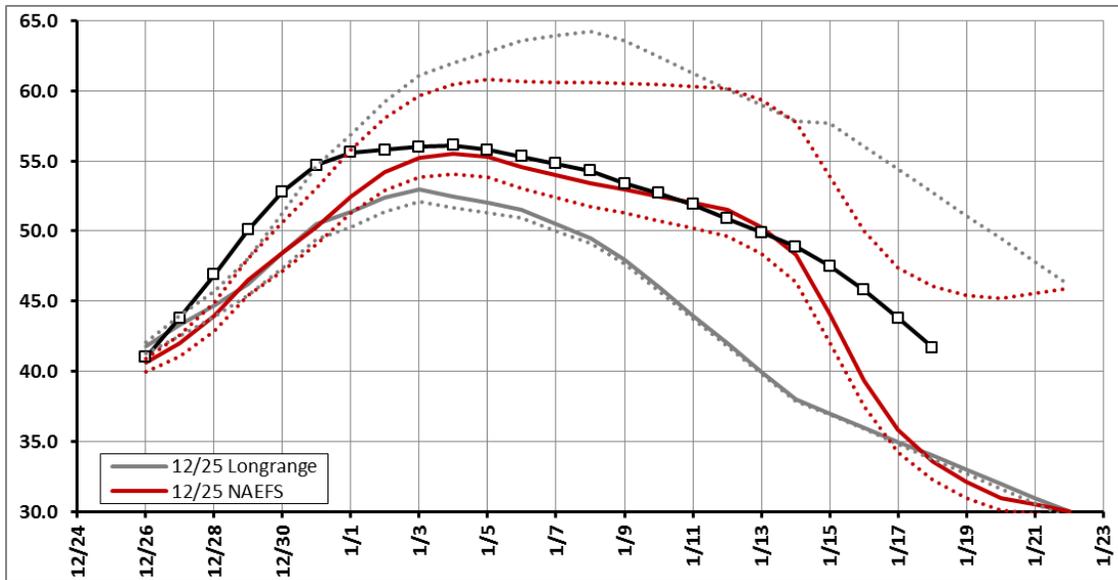


Figure 12. Uncertainty ranges for both the operational 1 day QPF (labeled "Longrange") and the 16 day QPF NAEFS forecasts issued on 12/25/2016 compared to observations (black line). The beginning of the rise was outside of the forecast uncertainty ranges (due to human regulation to mitigate flood impacts) but the crest was near the center.

2.0 Verification of Headwater Basin Forecasts with Varying QPF

2.1 Overview

In the previous section, a verification analysis was presented that dealt specifically with the Mississippi River comparing only two (2) different QPF durations. Another verification analysis was recently conducted which was meant to cover the smaller river systems covered by LMRFC. Optimal QPF duration is also important for the remainder of the LMRFC forecast area, and a wide range of QPF durations are available to forecasters. LMRFC policy is for forecasters to use best judgement, and collaboration with local NWS WFOs, to determine the duration of QPF to use operationally. Until spring 2016, the default QPF duration (unless changed manually by LMRFC forecasters) was 12hrs for all areas other than the Appalachians. Past work by a Southern Region (SR) NWS QPF team attempted to standardize QPF methodology within the region. This section will provide an overview of this past work on determining the optimal QPF duration and will also present recent efforts to build upon these past analyses.

2.2 Past Work

In late 2010, a team was assembled with members from multiple SR NWS offices to discuss the issue of how to best utilize QPF for river forecasts within the region. The Southern Region River Forecast Center QPF Strategy Team (hereafter referred to as “QPF Team”) set out to “apply a uniform and scientific analysis to [determining the most effective QPF duration]” (Waller, et al., 2013). Due to the unique areas covered by each SR RFC, it was recognized that no “one size fits all” QPF duration could be applied to the entire region. Instead, each river forecast center conducted an analysis of forecasts using multiple QPF durations to determine what the optimal QPF duration would be for that RFC. A final report by this QPF Team was completed in July 2013 (Waller, et al., 2013).

The methodology used by the QPF Team focused solely on “hydrologically significant” forecasts, so any event with less than 0.1 inches of rainfall observed was removed, cutting the sample size roughly in half. Sites with significant impact from human regulation were also supposed to be removed. The sites selected for analysis by the LMRFC representatives of the QPF Team are shown by Figure 13. The time range of analyzed forecasts spanned from July 2008 and December 2012. The QPF Team evaluated forecasts based upon mean absolute error (MAE), bias (B), false alarm rate (FAR), and critical success index (CSI). The statistics were calculated and plotted for each of the study locations (Figure 14) and evaluated by the team to determine the best QPF duration. Finally, the optimal QPF duration determined for each forecast location was averaged for the entire RFC. Results were also summarized by “warm season” (March 15 to October 15) and “cold season” (October 15 to March 15) times of the year.

Once the analysis was completed for each RFC, the individual RFCs completed and submitted reports summarizing their local results. LMRFC's report was completed in July 2013. The LMRFC report indicated that the average optimal QPF duration for the entire year was about 48 hours (Figure 15), with individual basins ranging from 12 hours to 72 hours (Smith & Roth, 2013). For the warm season this average was slightly higher, near 54 hours, and for the cool season this average was slightly lower, near 43 hours.

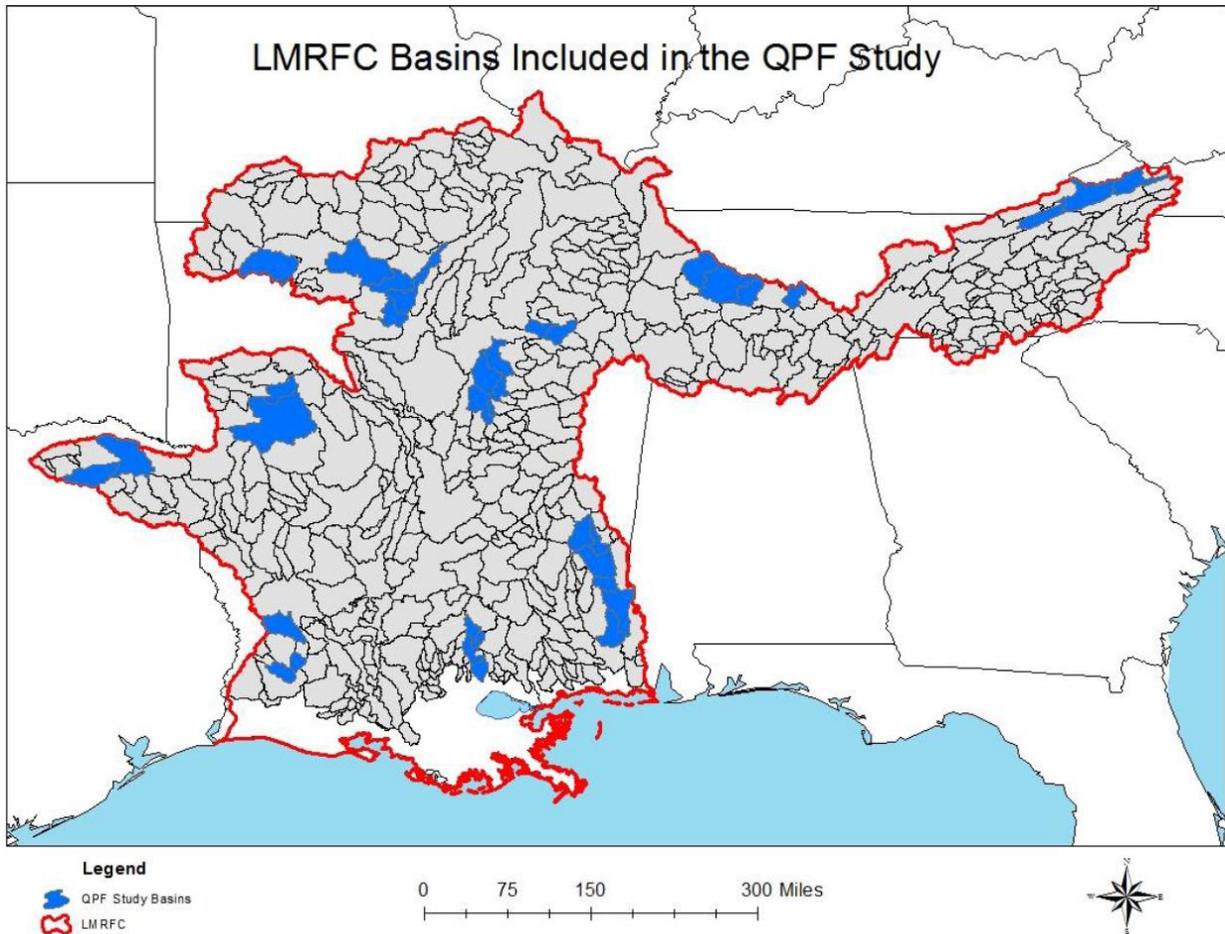


Figure 13. Basins studied in the LMRFC report to the SR QPF Team. From Smith & Roth (2013).

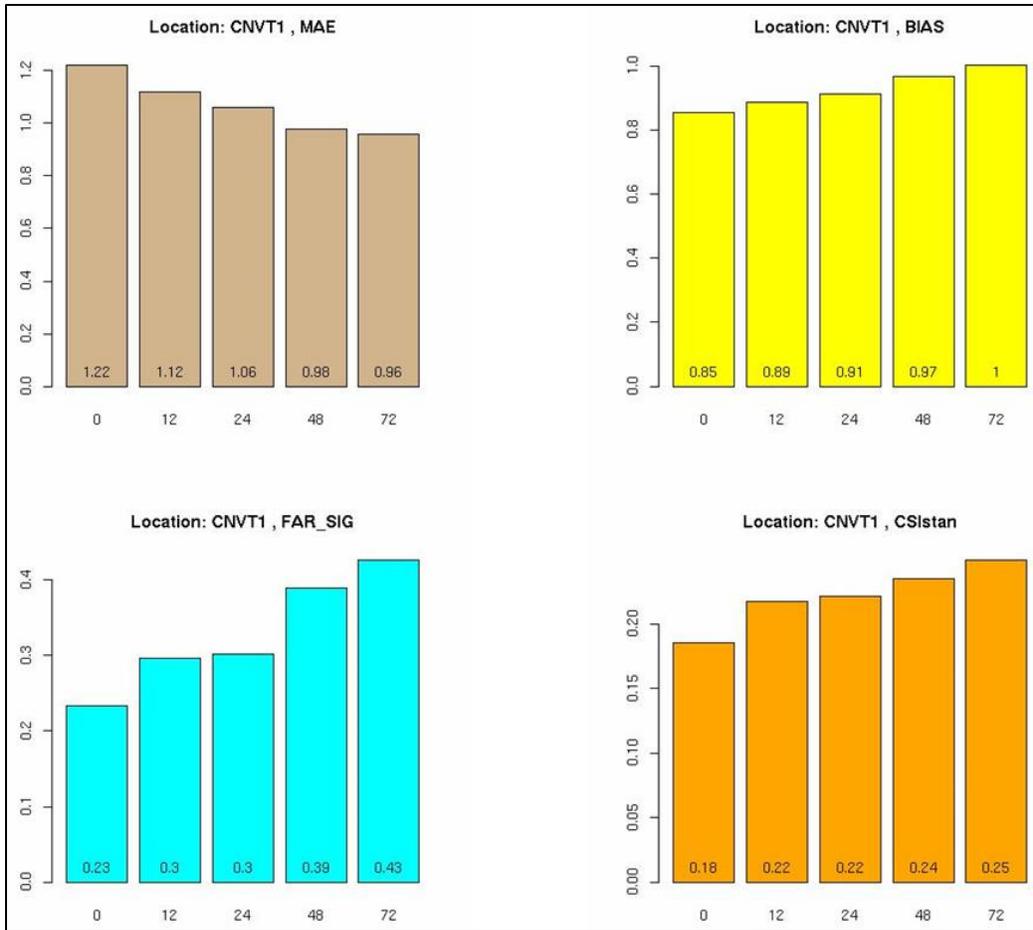


Figure 14. Example output statistics from the LMRFC report to the SR QPF Team. From Smith & Roth (2013). For this particular site, Duck River at Centerville, TN (CNVT1), all verification statistics except for FAR improved as the QPF duration increased.

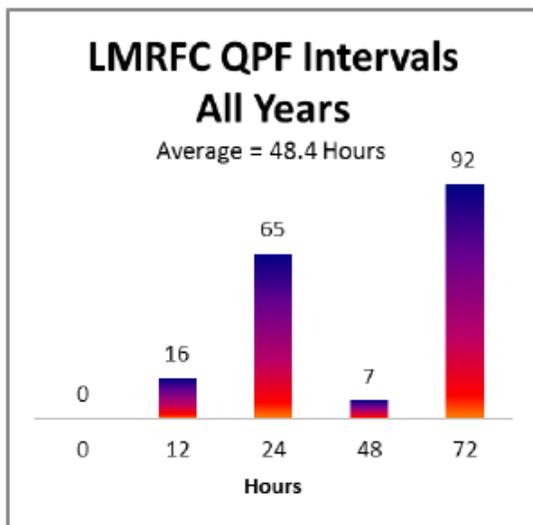


Figure 15. Summarized results of the LMRFC report to the SR QPF Team. From Smith & Roth (2013).

The LMRFC report also concluded that although many sites indicated 72 hours of QPF duration as optimal, there often was an increase in the FAR with additional QPF (see Figure 14 for an example). A review of the reported data showed that of the 36 basins studied by LMRFC for the QPF Team, 33 had a FAR less than 50% and 21 had a FAR less than 30%. In the LMRFC report, a FAR of 0.3 or higher (both cool and warm seasons) was considered unacceptable. In cases where the FAR was deemed higher than acceptable, a shorter QPF duration was deemed “optimal.” Besides FAR, virtually every single studied location showed a pattern of continuously improving error statistics with increased QPF. Based upon this information, the recommendation by LMRFC was for 24 hours of QPF to be used in all seasons. The results from the LMRFC report were summarized along with the results of the other SR RFCs in the “Final Report of the Southern Region River Forecast Center Quantitative Precipitation Forecast Strategy Team” (Waller, et al., 2013).

2.3 Methodology

Although the LMRFC participated in the QPF Team study and developed recommendations for the optimal QPF duration, some significant caveats remained. Of the 36 study basins, only five (5) were headwater basins (where streamflow has no upstream basin component). Of the study basins, 22 had some form of human regulation which could alter natural streamflow, and of those 7 were classified as being impacted by “major regulation” even though one of the listed specifications in the QPF Team summary report was that “forecast points immediately downstream from a major reservoir [were removed]” (Waller, et al., 2013). The usage of non-headwater basins can be problematic because the differences in simulated hydrographs include not only differences in QPF for that local basin, but also include differences in QPF for all upstream basins as well as any difference in lag time over the range of flows sampled. The usage of basins with human regulation is also problematic because the simulated hydrographs would be dependent upon reservoir releases that are assumed based upon past conditions, and the reservoir releases can change greatly depending on the flow. By using headwater basins only and removing basins that have significant human regulation, the number of variables is reduced, and the signal from QPF duration differences may be stronger.

After renewed interest in the optimal QPF duration issue in 2015, the data for LMRFC’s area was investigated once more. For this instance, an altered methodology was used. This verification used only headwater basins. Basins with major regulation were excluded. A total of 51 basins were analyzed, comparing the 12 hour, 24 hour, 48 hour, and 72 hour QPF durations for batch forecasts from January 1st, 2010, through January 1st, 2013. In contrast with the earlier study, this analysis did not remove events with less than 0.1 inches of QPF. Verification

statistics maximum error (MAXERR), root mean squared error (RMSE), mean error (ME), and mean absolute error (MAE) were calculated. A series of python scripts were developed to automate several tasks: the organization of output data by basin, the summarizing of errors for each basin, and the creation of plots to visualize the error statistics.

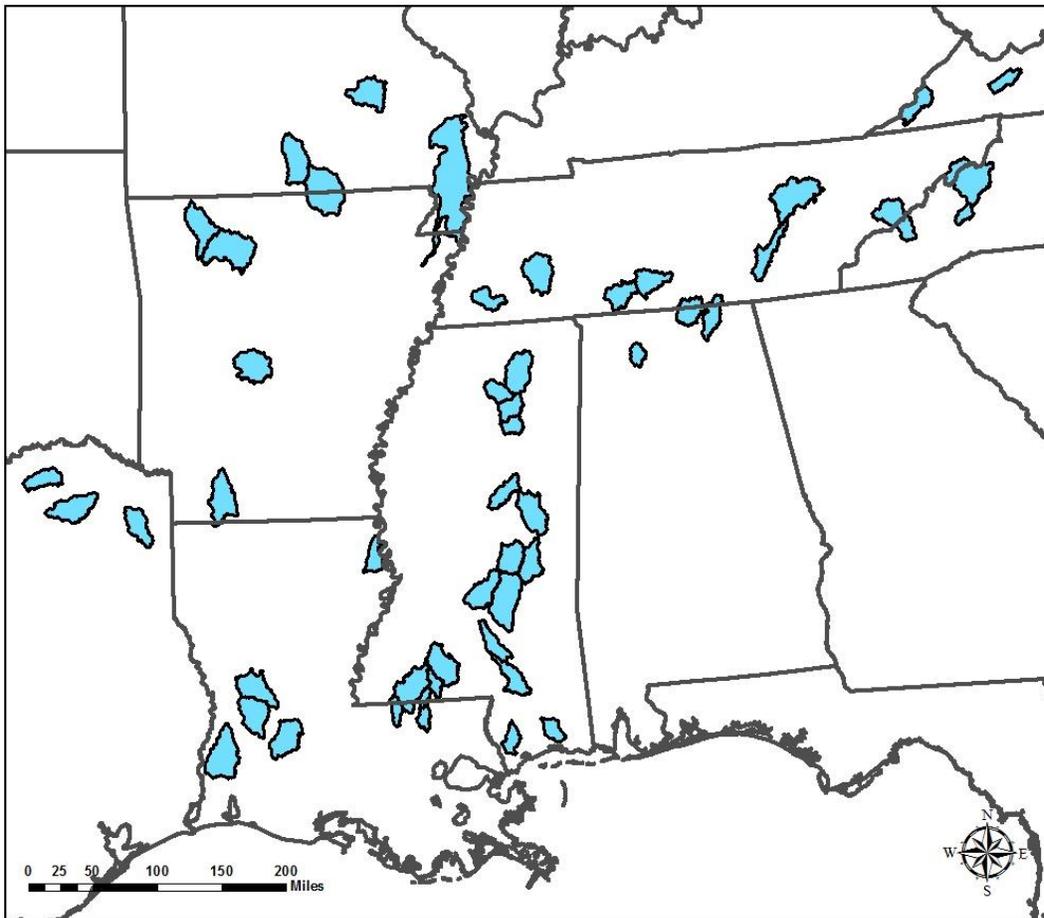


Figure 16. Headwater basins where forecasts were verified by differing QPF durations.

2.4 Results

Plots were created for each of the 51 basins studied showing ME and MAE. MAE was visualized as a bounded area on either side of ME (Figure 17). One plot was created for each QPF duration, and plots separated forecasts based upon the event being above or below flood stage. A trend toward improvement in forecast error statistics would be shown by the ME line moving closer to 0.0 and the width area between MAE lines decreasing. Similar to the results of the 2013 QPF Team report, virtually every study basin showed an improvement in MAE and ME (bias) statistics as the QPF duration increased.

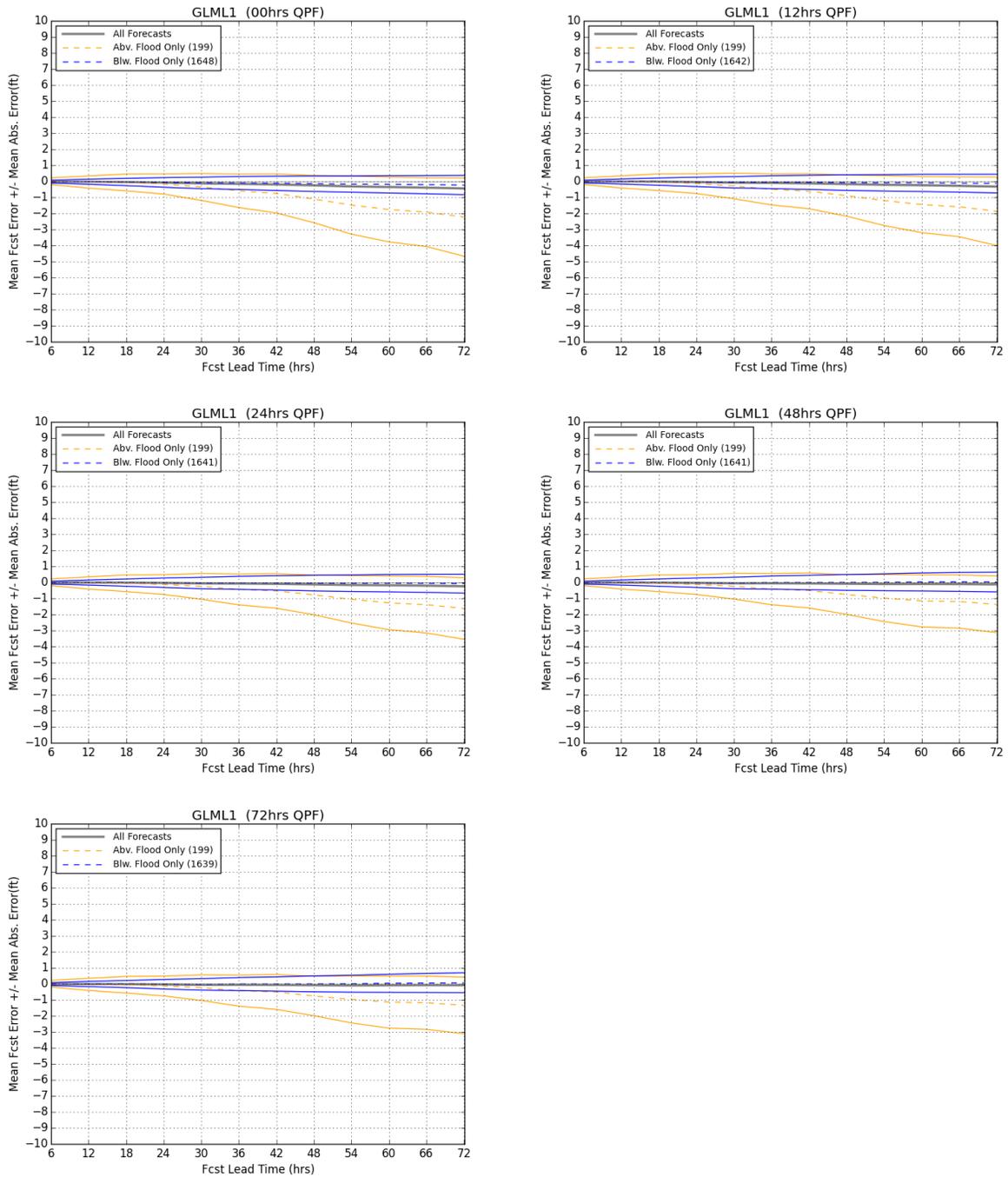


Figure 17. Example script output for Calcasieu River at Glenmora, LA (GLML1). A plot for each QPF duration was created, and each plot separates forecasts by above flood events (yellow) and below flood events (blue).

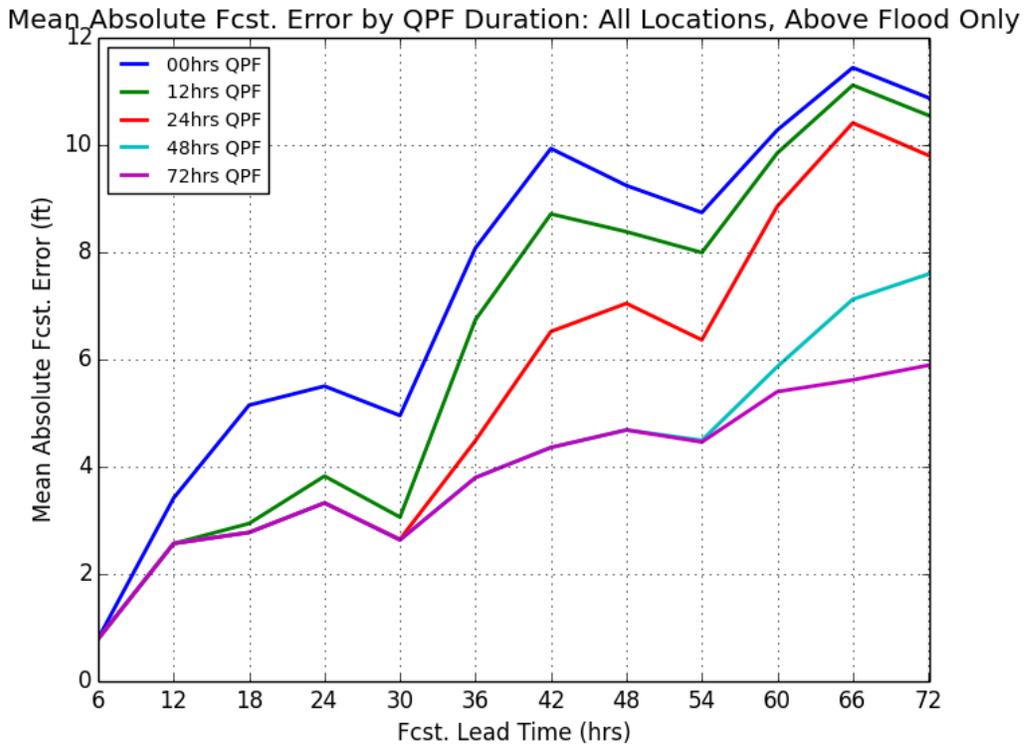
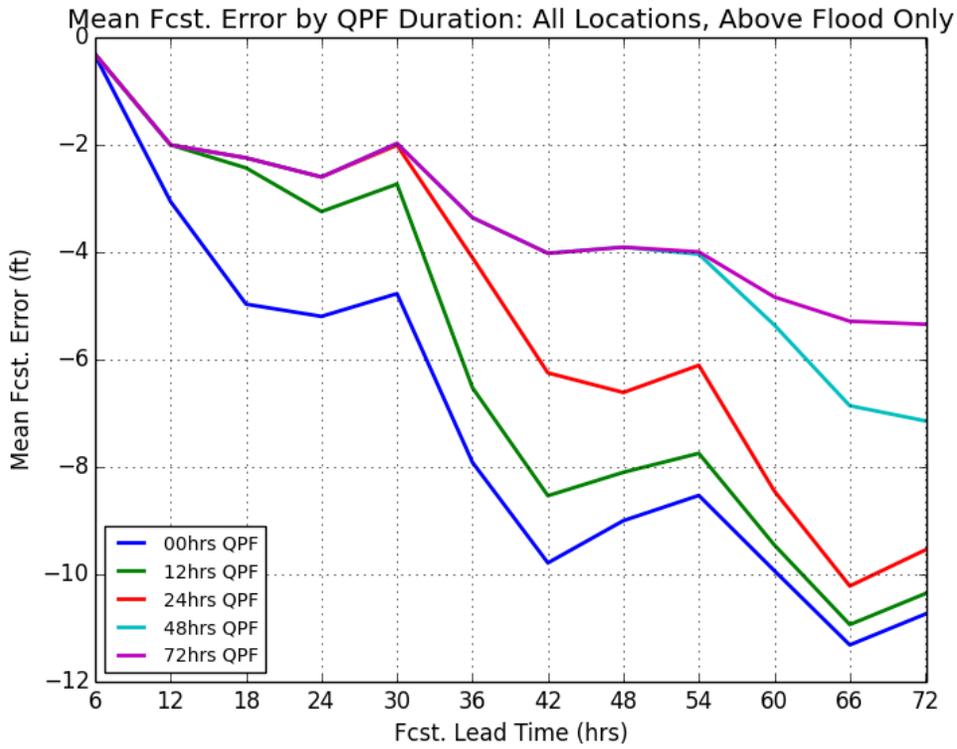


Figure 18. Summary of ME (above) and MAE (below) statistics for all study basins, for above flood stage events only. Both ME and MAE improved as QPF duration increased, with the largest gains leveling off by 48 hours of QPF.

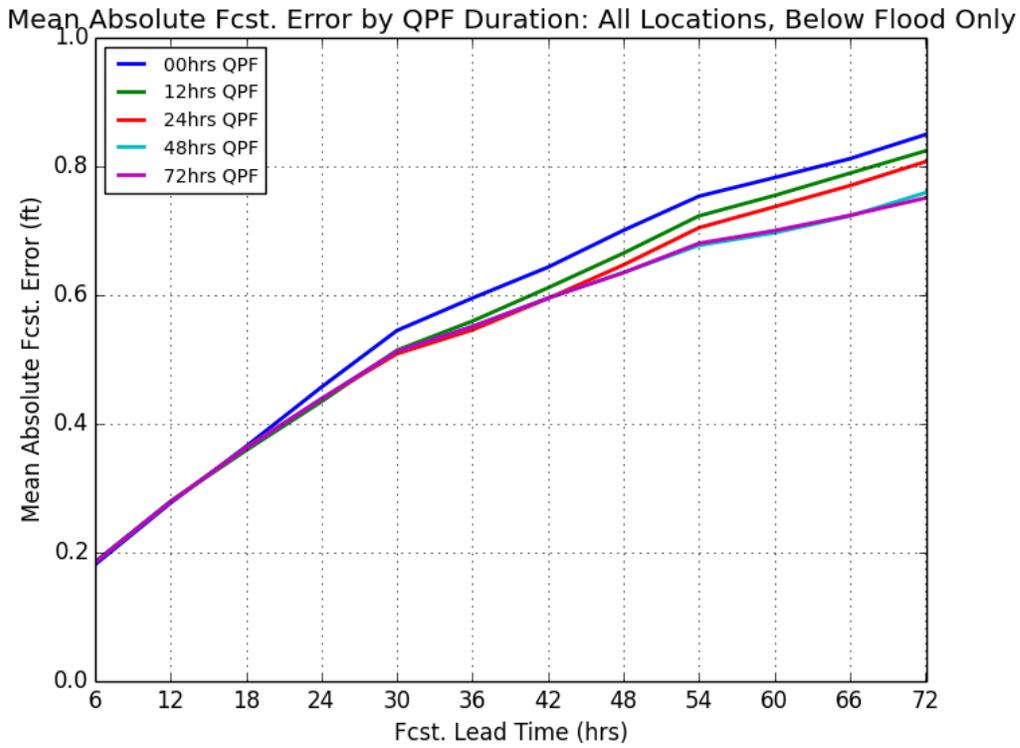
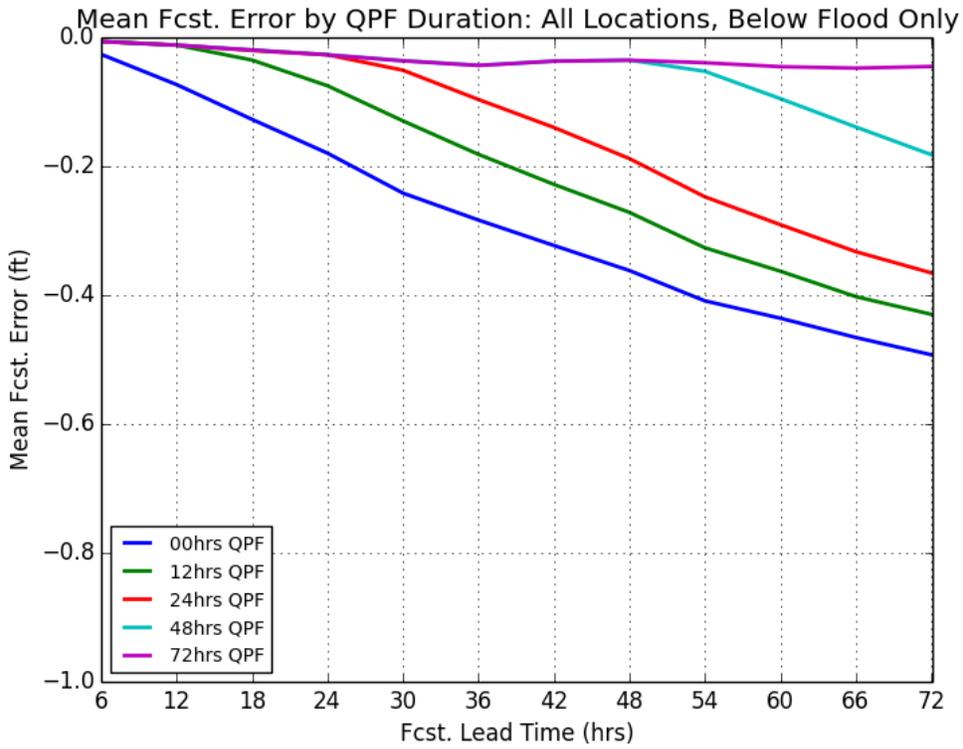


Figure 19. Summary of ME (above) and MAE (below) statistics for all study basins, for below flood stage events only. Both ME and MAE improved as QPF duration increased, with the largest gains leveling off by 48 hours of QPF.

2.5 Discussion

The verification statistics made available for this new study did not include FAR or CSI, so they cannot be compared to the 2013 QPF Team analysis. ME can be directly compared, and MAE can be indirectly compared (to bias), however. Even considering the issues noted with the chosen basins for the 2013 study, the results of this study appear to be very consistent with the previous results. There is a very clear, consistent tendency for reduced error statistics as QPF duration increases. Only a few study locations have ME values go above 0.0, similar to the 2013 analysis where only two (2) study locations had bias values go above 1.0 with longer QPF durations. Forecast error statistics for 72 hours of QPF show the greatest improvement compared to 12 hours of QPF when looking at only above flood events. The 3 day river forecast has almost four (4) times the error of 1 day forecast when only 12 hours of QPF is used; with 72hrs of QPF, the error drops to less than a multiple of two (2). It should be noted, however, that almost half of the study basins had no above flood events during the three (3) year study period.

In the 2013 study, FAR was the one error statistic that rarely improved with additional QPF. This conceptually makes sense, as using less QPF will always lower a forecast crest, moving forecasts away from flood stage. Choosing a FAR value that is “acceptable” is not easy, however, and limited guidance exists. Statistics for tornado warnings issued from 1994 through 2016 indicate FAR values ranging from 0.70 to 0.80, with an indicated goal FAR of 0.72 or less (National Weather Service, 2016). The 0.30 FAR threshold used by the 2013 QPF Team study was much lower, but a “better” value is elusive. “Acceptable” FAR values likely vary by weather phenomena and may even vary by end user of NWS products.

2.6 Operational Considerations and Future Work

Suggestions for future work on the QPF duration topic include the addition of error statistics for the human-reviewed river forecasts onto plots using the automatically-generated forecasts. Analysis of additional forecasts (beyond the 3 years used here) might also make conclusions stronger. It might also be helpful to remove hydrologically insignificant (<0.1 inches QPF) events for a more direct comparison to past verification studies.

3.0 References

National Weather Service. (2016). *NWS GPRA Metrics*. Retrieved July 2016, from https://verification.nws.noaa.gov/services/gpra/NWS_GPRA_Metrics.pdf

Smith, J., & Roth, K. (2013). *LMRFC Report to the NWS Southern Region River Forecast Center QPF Strategy Team*. National Weather Service. Retrieved July 2016

Waller, G., Anderson, T., Lowe, L., Paul, J., Crowley, L., Smith, J., . . . Wright, J. (2013). *Results, Conclusions, and Recommendations for Including Quantitative Precipitation Forecasts for Use in Operational Hydrologic Forecasts in the Southern Region River Forecast Centers*. National Weather Service. Retrieved July 2016