10.4 OPTIMIZATION OF QUANTITATIVE PRECIPITATION FORECAST TIME HORIZONS USED IN RIVER FOREASTS

Noreen O. Schwein*
National Weather Service Central Region Headquarters, Kansas City, Missouri
Kristopher S. Lander
National Weather Service Central Region Headquarters, Kansas City, Missouri
Adam C. Jones
University of Oklahoma, Norman, Oklahoma

1. INTRODUCTION

In the early to mid 1990s, the National Weather Service (NWS) began issuing river forecasts that included forecasts for precipitation amounts (Quantitative Precipitation Forecasts (QPF)) as part of the nationwide routine (e.g., daily) and event-driven (flood) forecasts. This was mainly due to recommendations from post flood assessments, particularly from report on the Great Flood of 1993 (NOAA, 1994). Prior to 1993, use of QPF was sporadic across the U.S. with most of the use being event-driven as part of an internal, contingency forecast. Numerous recommendations from The Great Flood of 1993 Natural Disaster Survey Report cited the need to incorporate QPF into river forecasts on a routine basis to ensure an increase in the lead time to the event, as communicated to the public via statements and warnings. The report also discussed the pros and cons of using QPF stating “Even with large river systems, unless the timing, magnitude, and location of the predicted rainfall can be accurately delineated, errors in the timing and magnitude of downstream crest forecasts can be substantial.” The need for increased lead time had to be weighed against minimizing the issuance of false alarms due to over-forecasting the height of a river. Confidence in the accuracy of QPFs varied across the country and resulted in no policy being established as to the future time period (forecast time horizon) that should be used. NWS River Forecast Centers (RFCs) established local policy according to various local studies. While these studies provided some guidance as to the forecast time horizon to use, the studies were somewhat limited both in the amount of data analyzed and geographic scope.

*Corresponding author Address: Noreen O. Schwein, NWS Central Region Headquarters, 7220 NW 101st Terrace, Kansas City, MO, 64153-2371; e-mail: noreen.schwein@noaa.gov.

There are two RFCs in the NWS Central Region (CR): one provides forecasts for the upper Mississippi River Basin and the other for the Missouri River Basin. While both are located in the Midwestern U.S. with many hydrometeorological and geological similarities, the two RFCs have differing local policy with regard to the use of QPF in river forecasts. These differences can be problematic when translated to consistent service provided to the public. For example, forecasts for the lower Mississippi River use hydrologic inputs from four other RFCs: North Central RFC (NCRFC), Ohio River RFC (OHRFC), Arkansas-Red Basin RFC (ABRFC) and the Missouri Basin RFC (MBRFC). Figure 1 depicts the five RFC areas. Only two of these RFCs (NCRFC and OHRFC) use a common QPF forecast time horizon (24 hours) on a routine basis, year round. ABRFC uses 12 hours of QPF year round and MBRFC uses 12 hours in the spring and summer seasons, 24 hours in the fall and winter. The downstream RFC responsible for the lower Mississippi River (LMRFC) routinely uses 12 hours of QPF. Since future precipitation can cause a significant impact (positive or negative) on the resultant river forecast, varying precipitation inputs into the hydrologic models can also cause a significant impact. This paper will attempt to determine a common optimal QPF forecast

Figure 1 RFC areas contributing to Mississippi River forecasts

time horizon to be used by both CR RFCs; results will be shared with other NWS regional management for their possible use, as well.

2. DATA

QPF error data (forecast minus observation (F-O)) was provided by the NWS National Precipitation Verification Unit in gridded form (4x4 km^2) and spanned a 39-month time period from October 2004 to January 2008. The data was taken from the Hydrometeorological Prediction Center (HPC) daily 1200 UTC forecasts (HQPF) in four 6-hour time steps: 1200-1800, 1800-0000, 0000-0600 and 0600-1200 UTC. HPC forecasts are used by all RFCs as initial guidance and were therefore considered consistent QPFs to analyze. Since the impact of less than .254 cm (0.10 inch) of precipitation was deemed negligible on river forecasts, this study concentrated only on

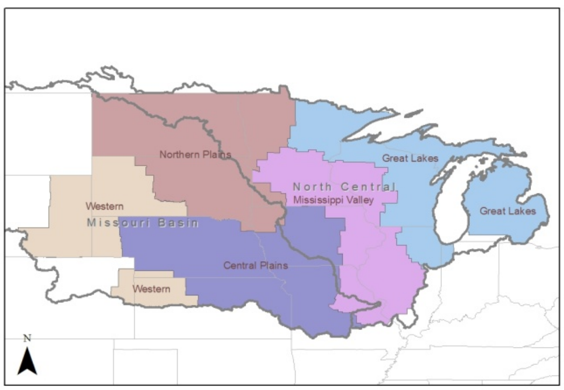


Figure 2 NWS Climate Regions in Central Region

Precip Depth Category	HQPF Value Range (cm)	HQPF Value Range (in)
A	0.25 ≤ HQPF < 0.64	0.10 ≤ HQPF < 0.25
B	0.64 ≤ HQPF < 1.27	0.25 ≤ HQPF < 0.50
C	1.27 ≤ HQPF < 2.54	0.50 ≤ HQPF < 1.00
D	2.54 ≤ HQPF < 5.08	1.00 ≤ HQPF < 2.00
E	5.08 ≤ HQPF < 7.62	2.00 ≤ HQPF < 3.00
F	HQPF ≥ 7.62	HQPF ≥ 3.00

Table 1 Forecast Precipitation Categories

days when at least .254 cm of precipitation was forecast. The data was further stratified by CR management determined geographic regions (Figure 2): Central Plains, Great Lakes, Mississippi Valley, Northern Plains and Western; precipitation categories (Table 1) and seasons: fall (September-November), winter (December-February), spring (March-May) and summer (June-August). The regions were based on a combination of climatology and NWS field office boundaries.

Microsoft Excel spreadsheets were organized by season, then subdivided by regions and precipitation categories for each of the four 6-hour time periods. Time periods are referred herein as first (1200 – 1800 UTC), second (1800 – 0000 UTC), third (0000 – 0600 UTC) and fourth (0600 – 1200 UTC) periods. Each record represented one 4x4 km^2 grid that fell within the above stated season/climate region/depth hierarchy.

3. DATA ANALYSIS

To find an optimum forecast time horizon, a good approach seemed to be comparing QPF forecast errors from one time period to another. Absolute values of the forecast errors were first analyzed for significant differences. Table 2 shows the breakdown of the number of grids analyzed, over 20 million. The amount of data was somewhat unwieldy, but the stratification helped to make it more manageable. Table 3 shows an example of this stratification. A combination of Microsoft Excel and SAS JMP IN software were used for the statistical analysis. Two-tailed Student's t-Tests for unequal variance were conducted for each forecast period, precipitation category, climate region and season. These tests were conducted on the entire range of data for the absolute value of the forecast – observed records. Sample results from one test are presented in Table 3. For most combinations, the p values were much less than the selected 0.05 alpha level, and were deemed significantly differently. The individual records for each forecast error grid showed very large negative errors where heavy rain occurred and was not forecast. This likely contributed to the significant differences, especially since such a large volume of data would be sensitive to small differences (Sall and Lehman, 1996). Positive error values, indicative of over-forecasting, were not nearly as extreme; both are shown in Table 4.

	Precipitation Forecast Absolute Errors Grid Count				
	1st Pd Forecast	2nd Pd Forecast	3rd Pd Forecast	4th Pd Forecast	Total
Fall	1,284,686	1,223,394	1,047,218	1,130,043	4,685,341
Winter	698,947	767,068	773,904	781,052	3,020,971
Spring	1,556,215	1,287,747	1,510,227	1,596,944	5,951,133
Summer	1,714,479	1,432,830	1,769,810	1,900,085	6,817,204
Total	5,254,327	4,711,039	5,101,159	5,408,124	20,474,649

Table 2 QPF Error grid count by season and forecast time period for all regions and precipitation categories

Central Plains Category A	**6-hour Forecast Begin time**			
	1200 UTC	**1800 UTC**	**0000 UTC**	**0600 UTC**
1200 UTC	--	--	--	--
1800 UTC	0.00	--	--	--
0000 UTC	0.000014	0.00	--	--
0600 UTC	0.00	0.0000013	0.0042	--

Table 3 T-test results for the fall season, Central Plains, category A QPF errors

Precipitation Forecast - All Errors (cm)								
	1st Pd Forecast		2nd Pd Forecast		3rd Pd Forecast		4th Pd Forecast	
Precipitation Category	Max	Min	Max	Min	Max	Min	Max	Min
A	0.63	-22.73	0.63	-12.19	0.63	-18.48	0.63	-28.75
B	1.26	-29.44	1.26	-14.77	1.26	-15.55	1.26	-22.80
C	2.53	-18.34	2.53	-7.65	2.53	-11.83	2.53	-27.25
D	4.68	-9.01	3.57	-1.27	4.60	-4.84	5.03	-8.96
E	5.72	-1.47					7.56	-5.05
F							8.10	7.82

Table 4 Maximum and Minimum QPF errors (F-O)

As stated previously, RFCs used local studies to determine the QPF forecast horizon to use. These studies were based on occasions when QPF was over-forecast. Since one of the main reasons for limiting the forecast time horizon for QPFs is to minimize over-forecasting and false alarms, analyzing errors greater than zero rather than the absolute value of errors would certainly be valuable for optimizing the number of forecast time periods to use. For this reason we next looked at positive errors, (F-O) > 0. As with the absolute errors, positive errors were stratified in a similar manner. In order to make the analysis less unwieldy, the 17+ million grids were decreased to 306 using means and testing by groups and pairs. With more than 30 observations in each group, parametric statistics could be used, however, the distribution indicated the data were not normally distributed thus nonparametric methods were also used.

The data were first separated by time periods. Statistical results from JMP IN shown in Figure 3 and detailed in Appendix 1 indicate a significant difference in the analysis of variance with a p value of 0.0408; the t-test indicates a significant difference with data paired with the 0600 UTC forecast period. The Tukey-Kramer and Wilcoxon tests point toward no significant difference for any of the four time periods. The diamonds in Figure 3 form the 95% confidence intervals of the hypothesized mean (that the difference of the means equals zero). Diamonds that do not overlap indicate a significant difference (Sall and Lehman, 1996). The first three forecast time periods (1200, 1800 and 0000 UTC) overlap which may indicate they are not significantly different. The diamond for the 4th period forecast (0600 UTC) shows the greatest difference. For this reason, the next step was to omit the 0600 UTC data from the comparison. Figure 4 shows the graphical results of comparing only the first three forecast periods.

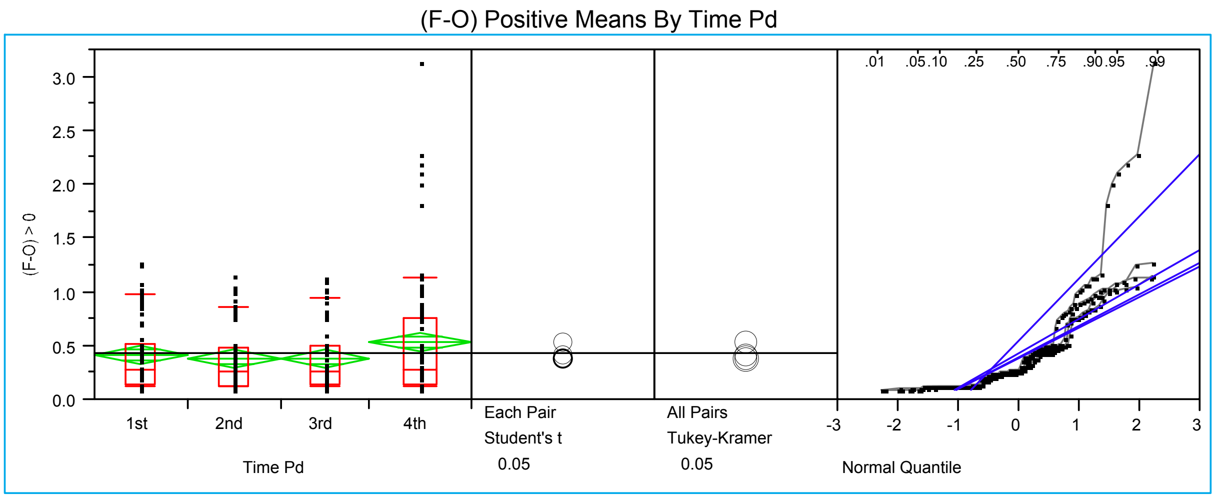


Figure 3 One-way Analysis of Variance (ANOVA) output for all forecast periods with normal quantile plot

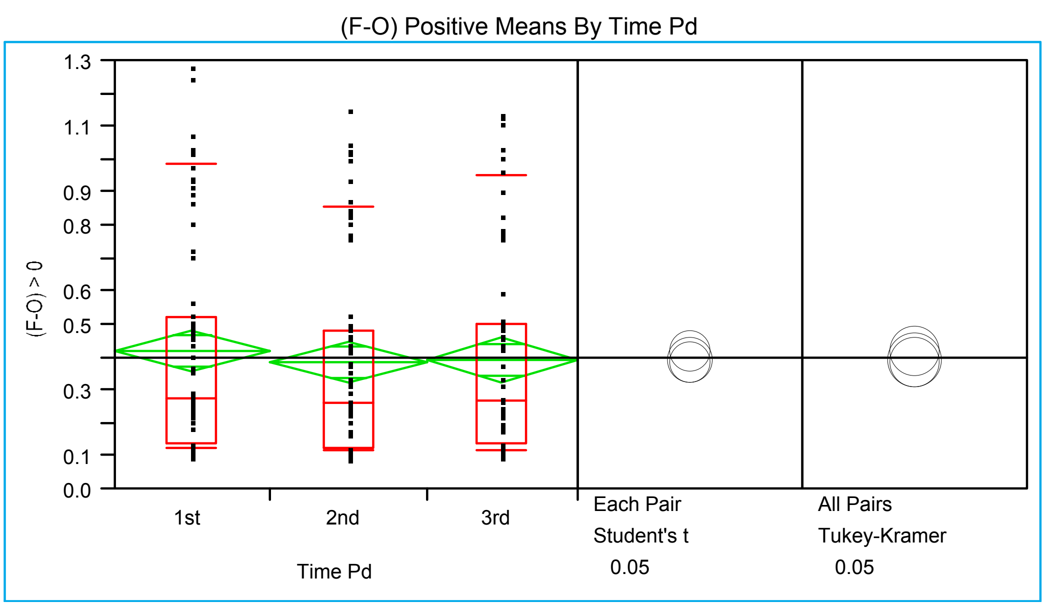


Figure 4 One-way Analysis of Variance (ANOVA) output for first three forecast periods

The overlapping diamonds suggest no significant difference. The p value from the ANOVA was 0.69 (compared to the previous 0.04), indicating no significant difference; the other tests also showed no significant difference.

As stated previously, the two CR RFCs vary their QPF time horizons by season. Figure 5 shows the results looking at all time periods for all seasons. Here, the summer season stood out, and the p values from both parametric and nonparametric supported a significant difference, as well. The winter season also appeared to be different although this was not indicated in the tests. Review of the summer errors showed a greater than 7.62 cm (3 in) positive error. As a matter of fact, this was the only occasion in all the 17+ million grids where a forecast in precipitation category F occurred (there were a total of 17 grids for this 4th period forecast on 20 July 2006). The

F-O Positive Means by Season

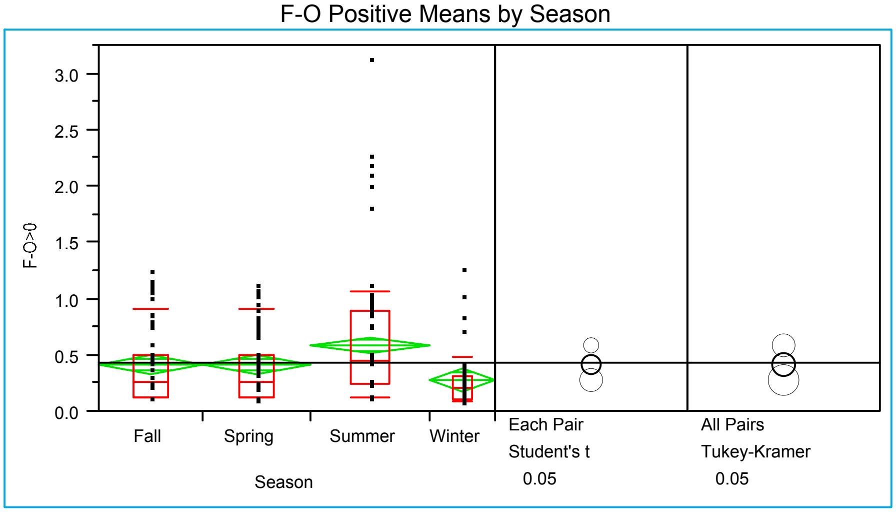


Figure 5 One-way Analysis of Variance (ANOVA) output by season for all seasons

record was omitted as an outlier and the analysis rerun. Figure 6 and the resultant tests show the summer season to continue to be significantly different. Statistical tests on the winter months were mixed. The seasonal test presented in Figure 5 was modified to remove the fourth QPF period from the summer season, while the fall, spring and winter seasons were unmodified and included all four periods of QPF. The parametric results presented in Figure 7 demonstrate that the 18-hour QPF period in summer is not statistically different from 24-hour periods of QPF used during the fall and spring periods. This infers that the 18-hour summer QPFs produced verification results similar to spring and

(F-O) Positive Means By Season

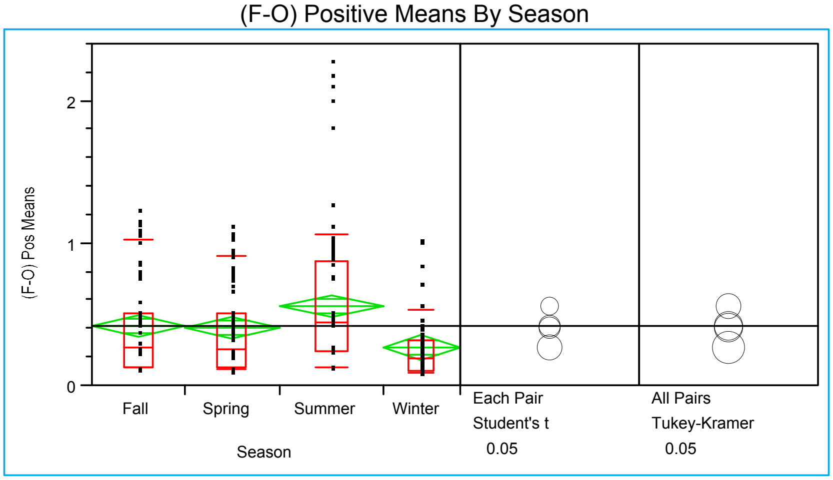


Figure 6 One-way Analysis of Variance (ANOVA) by season for all seasons omitting outlier

fall seasons. The results presented in Figure 7 also show a winter period being significantly different with a mean below the overall seasonal mean. This result might suggest that the winter season could likely accommodate additional periods of QPF beyond the 24-hour QPF period. However, generalized trends in the data suggest that additional periods of QPF will result in an increased error therefore, further study for QPFs beyond the 4th period (24-hour) forecast are needed.. Otherwise, it should be noted that the lower mean for the winter season points to less of a negative impact on the river forecast (i.e., lower false alarm rate).

Removing all of the summer data continued to show (Figure 8) no significant difference between the fall and spring seasons; significant with winter.. Running the data by time period and omitting the summer season showed no significant difference between any season (Figure 9).

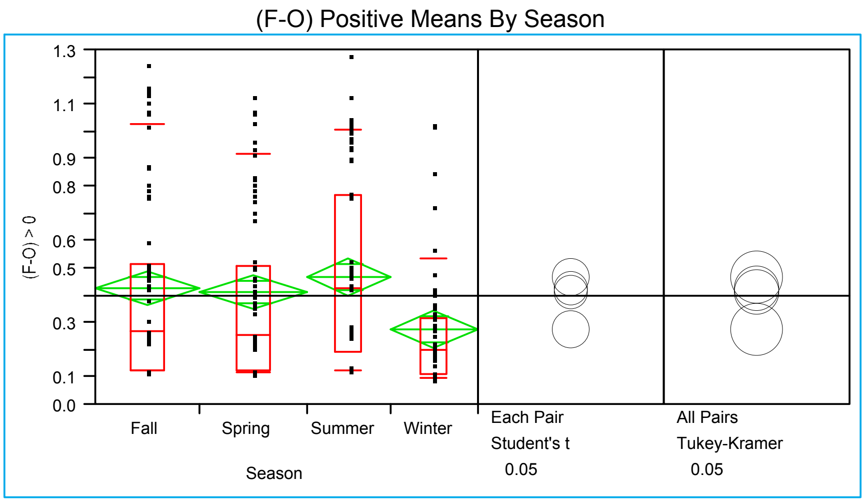
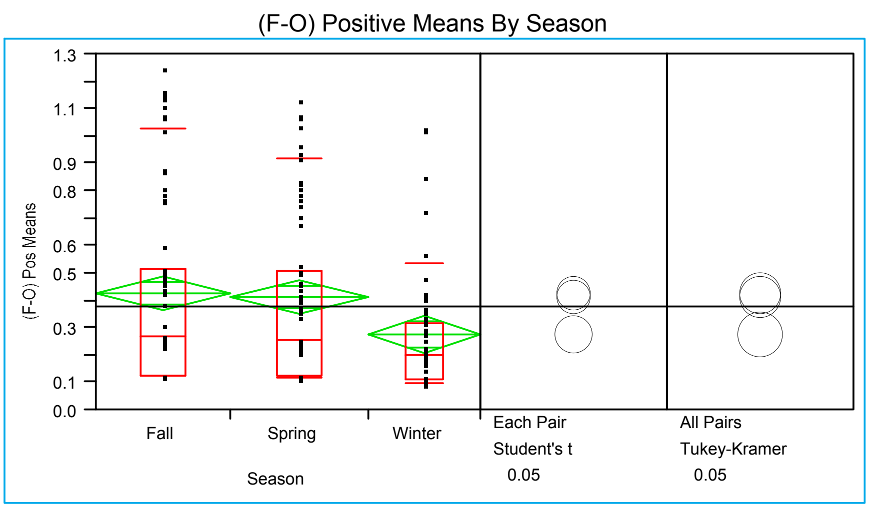


Figure 7 One-way Analysis of Variance (ANOVA) by season for all season using 18-hours of QPF for summer and 24-hours of QPF for all other seasons



Means for Oneway ANOVA omitting summer

Level	Number	Mean	Std Error
Fall	77	0.425714	0.03250
Spring	79	0.410759	0.03208
Winter	63	0.274921	0.03593

Figure 8 One-way Analysis of Variance (ANOVA) by seasons omitting the summer season entirely

(F-O) Positive Means By Time Pd

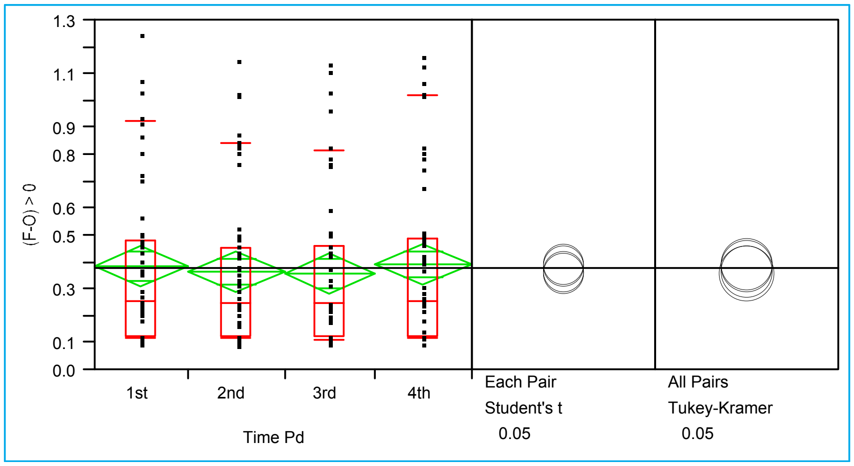


Figure 9 One-way Analysis of Variance (ANOVA) output omitting the summer season

Using this data to optimize the QPF time horizon, we were leaning toward separating out the summer season and using a shorter time period (18 hours). Although the winter could be extracted out per this methodology, there seemed no reason to bounce back and forth between the fall, winter and spring seasons when the mean positive error (over-forecasting) was less in the winter. This thought was also supported by HPC verification statistics which showed threat scores much lower in the summer months (example in Figure 10).

Finally, we used the same methodology as above, this time by the different climate regions in CR, and found no significant differences (Figure 11). Although the Mississippi Valley was a bit suspect, the statistical output did not support it (p value for the F-test was 0.58).

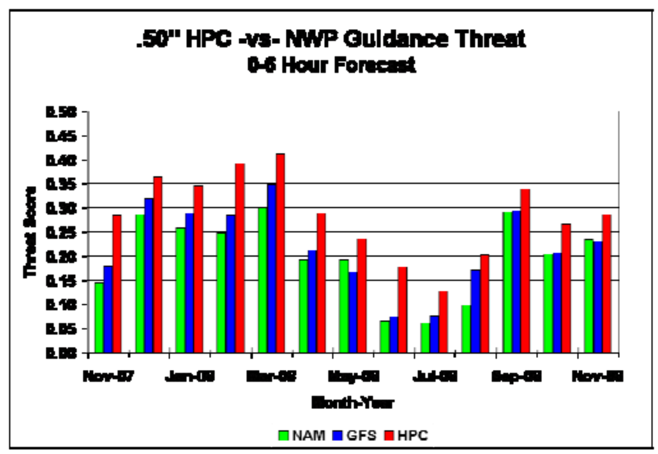


Figure 10 Example Threat Score plot from the NWS Hydrometeorological Prediction Center

(F-O) Positive Means By Region

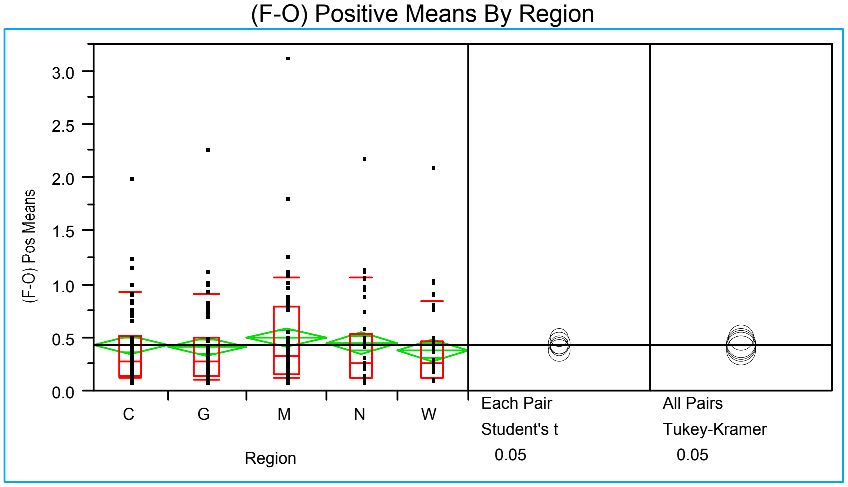


Figure 11 One-way Analysis of Variance (ANOVA) by region, all times and all seasons

4. FURTHER REVIEW OF THE DATA

In order to make a final proposal regarding the optimization of QPF time horizons, we looked at the data to get an idea of the magnitude of the errors and the possible impact on river forecasts. While specific studies could not be cited, experienced hydrologic forecasters have stated that less than 0.254 cm (0.10 inch) of rainfall in a 6-hour period has little if any impact on the height of a river. Higher amounts can cause a rise, however, this varies by season and antecedent conditions of the soil. For example, a 2.54 cm rainfall in six hours may cause significant flooding in the early spring when the ground is still frozen, but may cause no effect in the summer when the ground is drier and a canopy of vegetation absorbs a portion of the rain. It was because of these variations of effects on the river that we stratified the observations as we did by season and ranges of precipitation.

Separating the precipitation amounts into categories provided some interesting results. Table 5 shows the count for positive errors by forecast time period and precipitation categories. Dividing by similar totals for all errors, an areal coverage of positive error was calculated (Table 6). The coverage of over-estimated errors is greater than 80% for categories A-C, and 90-100% for D-F. However, 92% of those positive errors are in the A-B category, less than 1.27 cm (0.50 in). While there are probably times when less than 1.27 cm of rain over a large area would result in false alarm warnings, it may not occur that often. Further study should be done to determine this impact.

	Precipitation Forecast Positive Errors Grid Count				
Precipitation Category	1st Pd Forecast	2nd Pd Forecast	3rd Pd Forecast	4th Pd Forecast	Total
A	2,747,533	2,662,570	3,087,297	3,053,247	11,550,647
B	1,092,229	969,270	1,057,516	1,127,859	4,246,874
C	403,751	267,944	253,369	378,889	1,303,953
D	46,037	19,691	14,857	38,522	119,107
E	842	0	0	2,744	3,586
F	0	0	0	17	17
Total	4,290,392	3,919,475	4,413,039	4,601,278	17,224,184

Table 5 QPF positive error grid count by precipitation categories and forecast time horizons

Precipitation Forecast Positive Errors Grid Coverage				
Precipitation Category	1st Pd Forecast	2nd Pd Forecast	3rd Pd Forecast	4th Pd Forecast
A	0.82	0.83	0.87	0.85
B	0.80	0.84	0.84	0.86
C	0.85	0.88	0.88	0.81
D	0.91	0.99	0.97	0.92
E	0.96			0.97
F				1.00

Table 6 QPF positive areal error by precipitation categories and forecast time horizons

5. CONCLUSIONS

Precipitation is one of the most influential parameters in the river forecast; largely driving hydrologic models. However, forecasting future precipitation is a tremendous challenge. Accurate precipitation forecasts could provide great benefits with longer lead time to protect life and property and enhance the Nation's economy. This paper questioned one aspect of this difficult challenge: how far into the future should we extend precipitation forecasts that are incorporated into river forecasts?

This QPF forecast time horizon question was analyzed by forecast time periods, season, climatological regions and ranges of precipitation amounts. Parametric and nonparametric statistical tests were conducted. When analyzing for all time periods, significant differences were shown in QPFs issued in the 4th period (0600-1200 UTC). When analyzing for all seasons, significant differences were shown for the summer season. Further analysis showed no significant difference between spring and fall seasons for all four time periods when omitting summer. There was a difference during the winter season, but due to a smaller mean error, it was decided to group winter with the fall and spring seasons. HPC threat scores corroborated the finding of lower QPF skill during the summer months.

The authors submit the optimal decision for Central Region is for routine river forecasts to use an 18-hour QPF time horizon in the summer, and 24 hours during the remaining time of the year. However, during times of higher confidence events, river forecasts should be flexible to include longer QPF time horizons. The outlier forecast on 20 July 2006 shows an attempt to forecast for heavy rain events in the Midwest (events that often occur in that region after midnight (NOAA 1994)), but also points to the difficulty in accurately forecasting heavy precipitation events further into the future. Until meteorological models, forecasters and technology can improve to better pinpoint the location and timing of these heavy rain events, particularly with summer convection, NWS RFC forecasters should attempt to get the best lead time on these systems as possible, using the tools available.

6. ACKNOWLEDGMENTS

The authors wish to acknowledge Letitia Soulliard of the Hydrometeorological Prediction Center for her time and effort in gathering and transmitting the data to NWS Central Region Headquarters. Ms. Soulliard was extremely helpful and responsive to our requests and we are extremely grateful.

8. REFERENCES

Microsoft Office Excel 2007 (2006): Microsoft Corporation.

National Oceanic and Atmospheric Administration (NOAA) , 1994: Natural Disaster Survey Report, The Great Flood of 1993, 244 pp.

Sall, J., and A. Lehman, SAS Institute Inc., 1996: JMP Start Statistics. Duxbury Press, 520 pp.

7. APPENDIX – Example of Analysis of Variance output for the four forecast time periods (ref. Figure 3)

Oneway Anova

Summary of Fit

RSquare	0.02696
RSquare Adj	0.017294
Root Mean Square Error	0.396507
Mean of Response	0.43817
Observations (or Sum Wgts)	306

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	1.315532	0.438511	2.7892
Error	302	47.479844	0.157218	Prob>F
C Total	305	48.795375	0.159985	0.0408

Means for Oneway Anova

Level	Number	Mean	Std Error
1st	76	0.423684	0.04548
2nd	76	0.383553	0.04548
3rd	72	0.390972	0.04673
4th	82	0.543659	0.04379

Std Error uses a pooled estimate of error variance

Means Comparisons

Dif=Mean[i]-Mean[j]	4th	1st	3rd	2nd
4th	0.000000	0.119974	0.152686	0.160106
1st	-0.11997	0.000000	0.032712	0.040132
3rd	-0.15269	-0.03271	0.000000	0.007420
2nd	-0.16011	-0.04013	-0.00742	0.000000

Alpha= 0.05

Comparisons for each pair using Student's t

t

1.96788

Abs(Dif)-LSD	4th	1st	3rd	2nd
4th	-0.12186	-0.00427	0.026667	0.035865
1st	-0.00427	-0.12658	-0.09561	-0.08645
3rd	0.026667	-0.09561	-0.13005	-0.1209
2nd	0.035865	-0.08645	-0.1209	-0.12658

Positive values show pairs of means that are significantly different.

Comparisons for all pairs using Tukey-Kramer HSD

q*

2.58342

Abs(Dif)-LSD	4th	1st	3rd	2nd
4th	-0.15998	-0.04313	-0.01275	-0.003
1st	-0.04313	-0.16617	-0.13575	-0.12604
3rd	-0.01275	-0.13575	-0.17072	-0.16104
2nd	-0.003	-0.12604	-0.16104	-0.16617

Positive values show pairs of means that are significantly different.

Wilcoxon / Kruskal-Wallis Tests (Rank Sums)

Level	Count	Score Sum	Score Mean	(Mean-Mean0)/Std0
1st	76	11962	157.395	0.442
2nd	76	10899	143.408	-1.147
3rd	72	10645	147.847	-0.620
4th	82	13465	164.207	1.281

1-way Test, Chi-Square Approximation

ChiSquare	DF	Prob>ChiSq
2.6362	3	0.4512