U.S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
OFFICE OF SCIENCE AND TECHNOLOGY
METEOROLOGICAL DEVELOPMENT LABORATORY

MDL OFFICE NOTE 11-1

# DERIVING IMPROVED 6-H PROBABILISTIC QPFs (PQPFs) BY BLENDING TWO MODEL-PRODUCED PQPFs: PRELIMINARY RESULTS

Jerome P. Charba
Frederick G. Samplatsky

October 2011

# DERIVING IMPROVED 6-H PROBABILISTIC QPFs (PQPFs) BY BLENDING TWO MODEL-PRODUCED PQPFs: PRELIMINARY RESULTS

Jerome P. Charba
Frederick G. Samplatsky

## 1. INTRODUCTION

Over the past several years, the Meteorological Development Laboratory (MDL) has been producing experimental 6-h probabilistic quantitative precipitation forecasts (PQPFs) with a “high-resolution” MOS (model output statistics) approach (HR; Charba and Samplatsky 2011b, henceforth referenced as CS). Since 1 February 2010, the Hydrometeorological Prediction Center (HPC) of National Centers for Environmental Prediction has also been producing experimental 6-h PQPFs with a multi-model QPF distributions method, where the HPC deterministic QPF is the mode of the distribution (Novak et al. 2011). Subsequently, MDL has been conducting ongoing comparative verification of the HR and HPC PQPFs on the 4-km national Hydrologic Rainfall Analysis Project (HRAP) grid (CS), which is native to HR.

A significant benefit of conducting the comparative verification on the HRAP grid is that the verifying “Stage IV” precipitation data are also native to this grid. While this requires interpolating the HPC PQPFs from a 32-km grid, standard bi-quadratic interpolation well preserves the original grids. Note that prior to the verification the Stage IV precipitation data are subjected to supplemental quality control at MDL to remove sporadic residual errors (CS).

Comparative Brier skill scores (BSS) for the contiguous United States (CONUS) have consistently shown that HR and HPC PQPFs have similar skill considering all precipitation thresholds in the day 1 to day 3 forecast range (12 – 30 h and 60 – 78 h forecast projections, respectively). This is shown in Fig. 1 for recent cool and warm season samples; close inspection of this figure reveals HR with slightly better skill for light precipitation thresholds, HR and HPC have about equal skill for moderate thresholds, and HPC has slightly better skill for heavy thresholds[1].

[1] Comparative scoring with a different methodology at HPC (Novak et al. 2011) has yielded a similar HR versus HPC skill ranking for light precipitation thresholds, but for heavy precipitation thresholds the skill for HR is much lower than that seen here (not shown). The latter result is likely due primarily to spatial averaging of the Stage IV precipitation in the HPC scoring, as the HR PQPFs apply to full-resolution Stage IV data which is not mentioned in Novak et al.

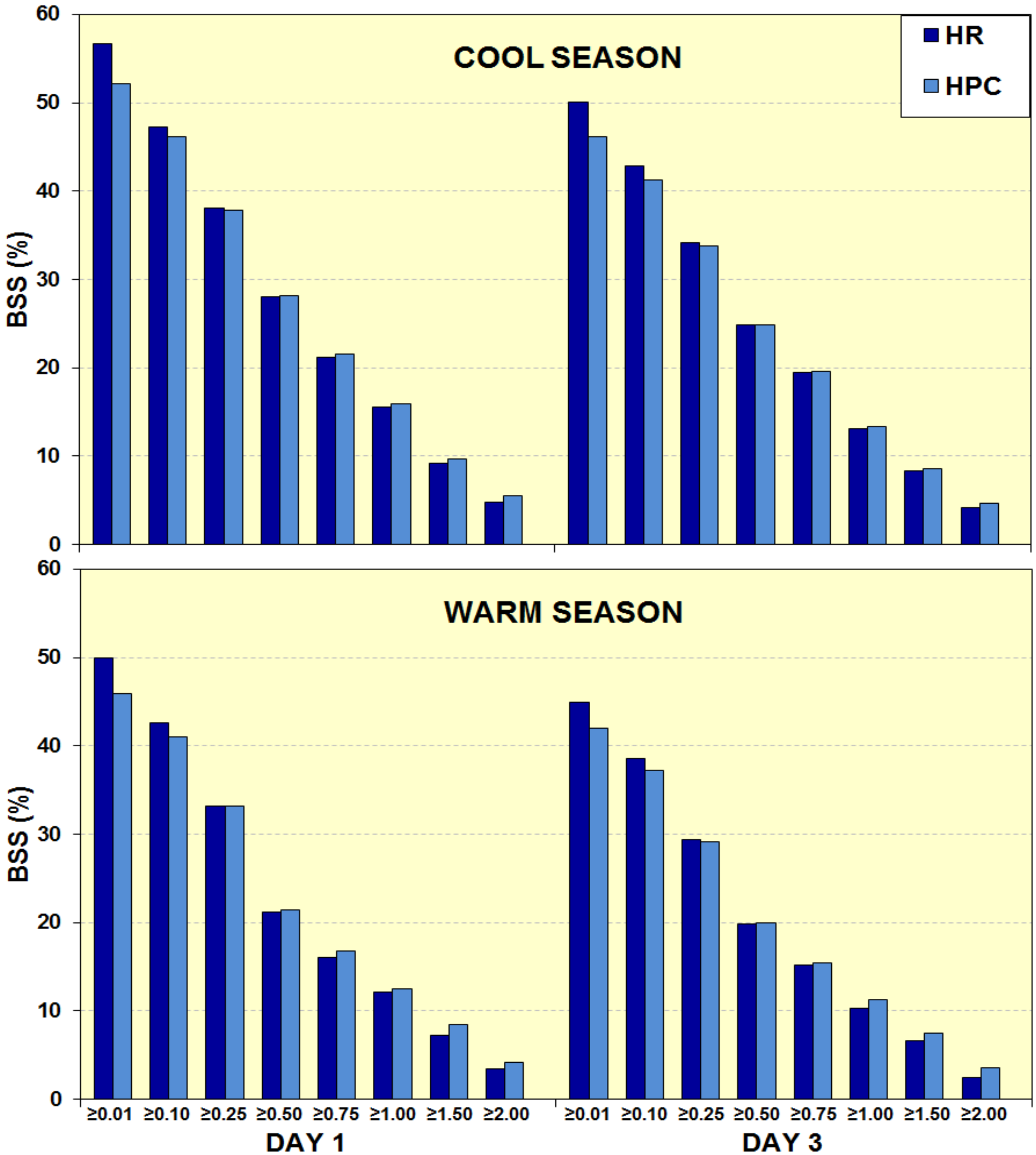


Figure 1. Seasonal Brier skill score (BSS) for the HR and HPC PQPFs on the HRAP grid, the precipitation (in.) thresholds shown, the 1200 UTC cycle, and the CONUS domain. The verification samples consisted of all days during 1 October 2010 – 31 March 2011 (cool season) and 1 April – September 30, 2010 (warm season) for the day 1 and day 3 forecast periods.

On the other hand, we have found that the reliability and sharpness of the HR and HPC PQPFs have substantial contrasts, where HR has somewhat better reliability while HPC has better sharpness. This finding is illustrated in Fig. 2 for day 2 (36 – 54 h forecast projections) for the same verification samples as for Fig. 1 [the reliability and sharpness properties for day 1 and day 3 (not shown) are quite similar to those for day 2)]. These findings suggest the two PQPF products may complement one another, and an improved PQPF product could result by blending them, as also noted by Charba and Samplatsky (2011c). In this technical note, we discuss two blending techniques and PQPF performance evaluations to investigate this hypothesis.

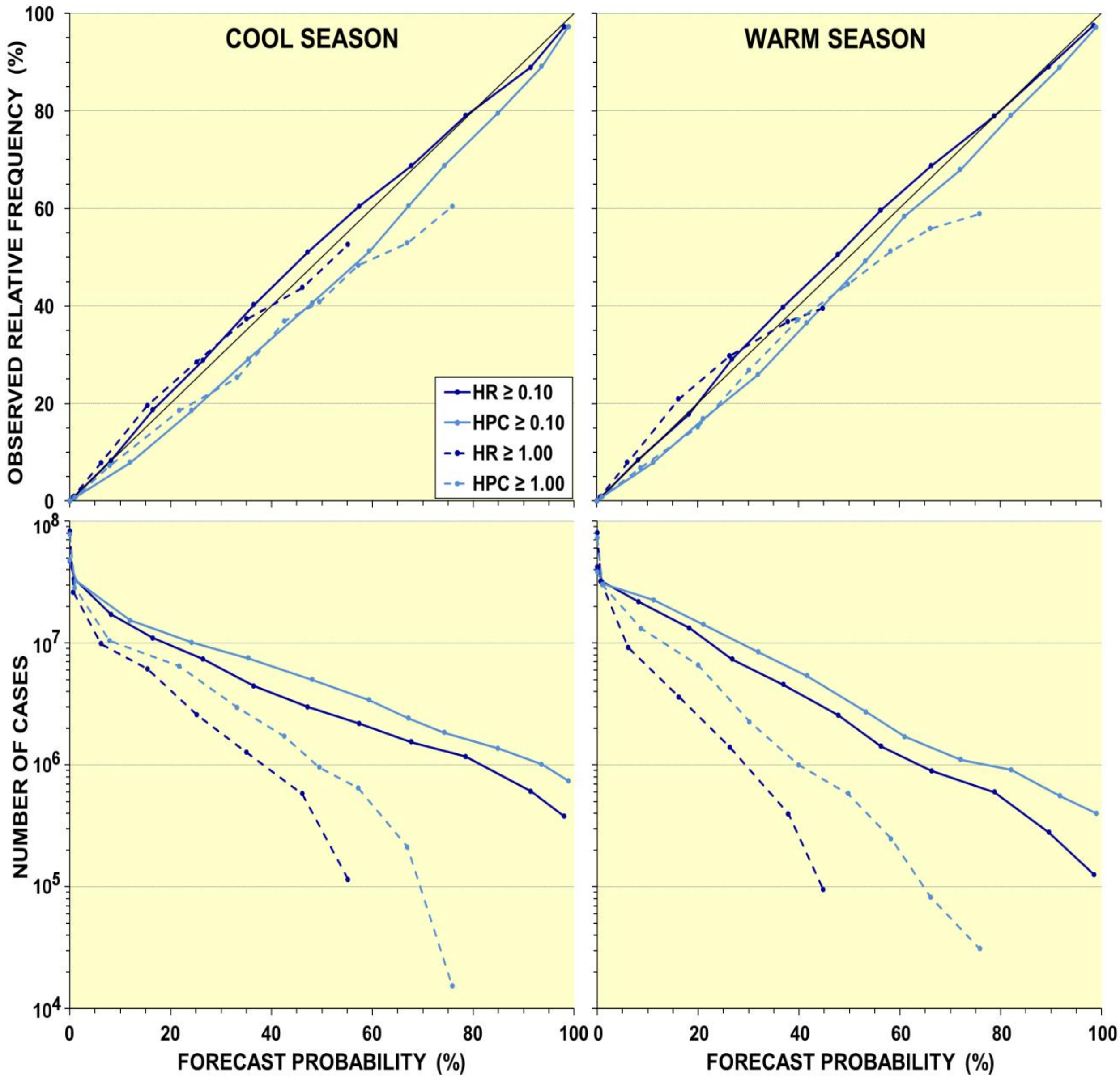


Figure 2. Day 2 HR and HPC PQPF reliability (upper) corresponding to Fig. 1 for ≥ 0.10 and ≥ 1.00 in., where better reliability for HR is indicated by a closer curve alignment to the perfect reliability (straight diagonal) lines. Corresponding probability distributions are shown in lower panels, where better sharpness for HPC is indicated by more cases of upper probabilities and, for ≥ 1.00 in., higher peak probabilities. For all charts the probability is plotted at the mean value for an interval.

## 2. PQPF BLENDING METHODS

One blending technique consisted of a simple average of the HR and HPC PQPFs, which we refer to as AVG. The other method, denoted RGR, used screening linear regression to develop and apply PQPF prediction equations, where the HR and HPC PQPFs in multiple forms were used as candidate predictors. The predictor forms included the HR and HPC PQPFs (raw and smoothed versions for HPC), grid binary forms, and products of the two PQPFs (raw and smoothed versions).

For the RGR development, the binary predictand consisted of yes/no occurrences (1/0 values) of 6-h precipitation thresholds, consisting of ≥ 0.01, ≥ 0.10, ≥ 0.25, ≥ 0.50, ≥ 0.75, ≥ 1.00, ≥ 1.50, and ≥ 2.00 in. Thus, the regression equation for ≥ 0.01 in. yields the probability of precipitation (PoP). For all remaining thresholds the predictand was conditional on the occurrence of ≥ 0.01 in. Since the resulting PQPFs are conditional (CPQPF), unconditional PQPFs were obtained by multiplying CPQPF by PoP. This conditional predictand approach as well as the grid binary and product forms of the HR and HPC PQPF predictors can account for non-linear predictor-predictand relationships despite the use of linear prediction equations (CS).

Since the HPC PQPFs have a relatively short history, the RGR developmental sample was short considering that occurrences of the heavy 6-h precipitation thresholds are rare events. To augment the samples, the data were combined over broad areas of the CONUS in either of two ways. In one case the data were combined over the entire CONUS, which is referred to as “national” (RGR_NAT). In the other case the data were combined within each of 14 *overlapping* “regional” geographical areas (RGR_REG; see Charba and Samplatsky 2011a for a description of the overlapping-regions method).

Since sample shortness was a concern, especially for RGR_REG, the data for the RGR development were also combined for the 0000 and 1200 UTC cycles. To test the performance of the RGR PQPFs every sixth day of each cycle was withheld from the developmental sample to form the independent test sample. Finally, for both RGR_REG and RGR_NAT the samples/regression equations were stratified by six-month cool (October – March) and warm (April – September) seasons to account for contrasting seasonal heavy precipitation mechanisms. Thus, both the development and test samples for the cool season were taken from 1 February - 31 March 2010 and 1 October 2010 - 31 March 2011; for the warm season the samples were from 1 April – 30 September 2010 and 1 April – 30 June 2011.

## 3. COMPARATIVE SCORING OF THE VARIOUS PQPFS

In this section, we examine the comparative skill, reliability, and sharpness of the HR, HPC, AVG, RGR_NAT, and RGR_REG PQPFs for the cool season and warm season test samples noted above. Fig. 3, which contains seasonal BSSs for day 1 and day 3, shows that skill for RGR_NAT and RGR_REG is clearly higher than that for HR, HPC and AVG, especially for moderate and heavy precipitation thresholds and the cool season. More minor features in Fig. 3 are: (1) the skill ranking of HR versus HPC is essentially the same as in Fig. 1 even though the associated verification samples are quite different; (2) the skill for AVG lies between that for HR and HPC for the light precipitation thresholds, but it is slightly higher than either for all remaining thresholds; (3) the skill for RGR_REG is only marginally higher than for RGR_NAT, which is not surprising considering the shortness of the developmental samples[2].

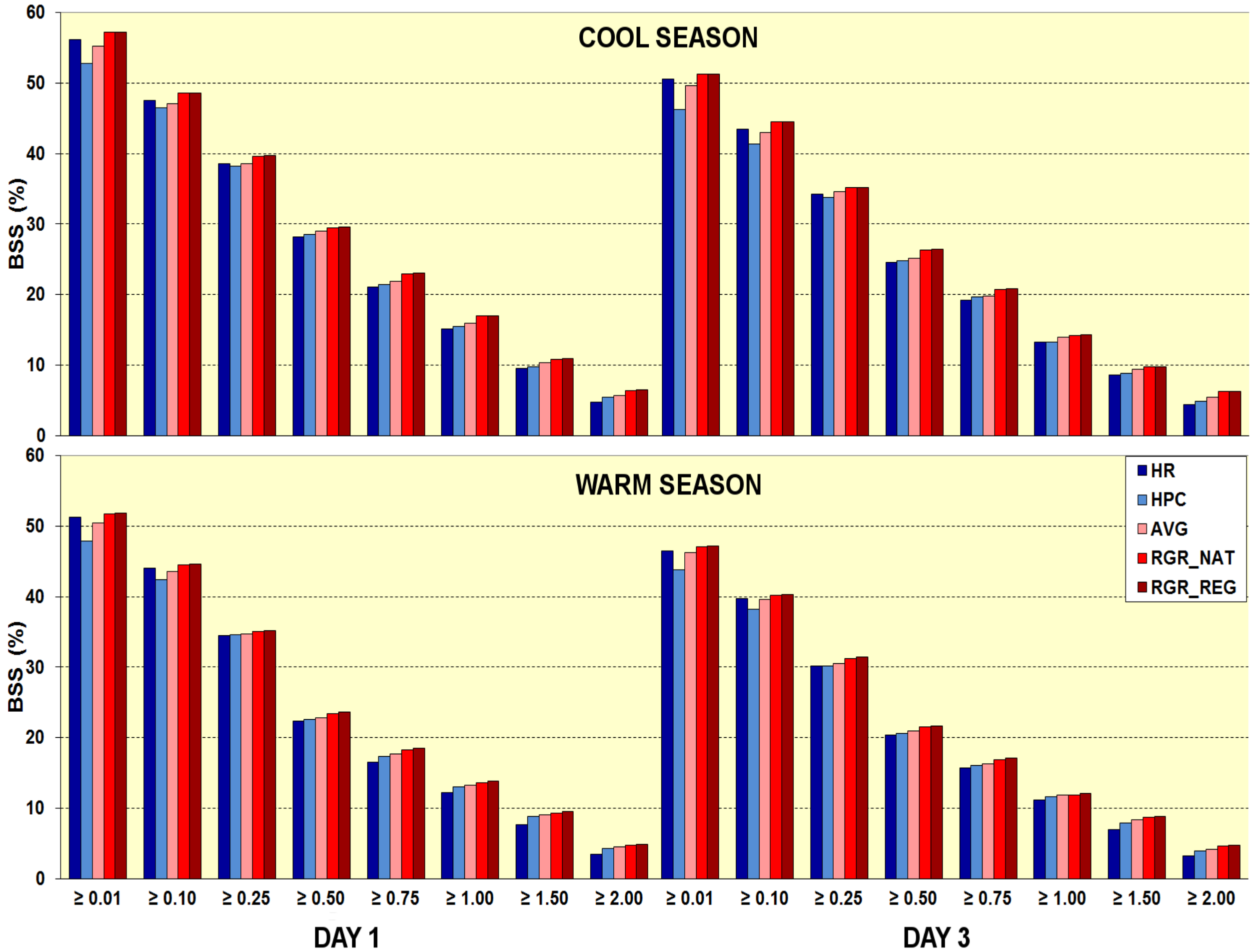


Figure 3. BSS for HR, HPC, AVG, RGR_NAT, and RGR_REG PQPFs, the precipitation thresholds (in.) shown, the cool season (top), the warm season (bottom), day 1 (left), and day3 (right). See text for the legend notation and verification samples.

[2] In similar experiments involving regional versus national PQPF regression equations based on much longer developmental samples, Charba and Samplatsky (2011b) found that regionalization yielded a larger skill enhancement.

Reliability and probability distribution diagrams, corresponding to Fig. 3 for a mid-range precipitation threshold (≥ 0.50 in.) and day 2, are shown in Fig. 4. Here we find that the reliability for RGR_REG and RGR_NAT is quite good, and it is better than that for HR, HPC, and AVG. The sharpness for RGR_REG is also good, although HPC's sharpness is somewhat better especially for the warm season. More minor features are: (1) the sharpness for RGR_REG is somewhat better than that for RGR_NAT for upper probabilities; (2) the reliability and sharpness curves for AVG generally lie between those for HR and HPC; (3) the reliability and sharpness properties for HR and HPC are qualitatively similar to those in Fig. 2 despite the different precipitation thresholds and verification samples involved.

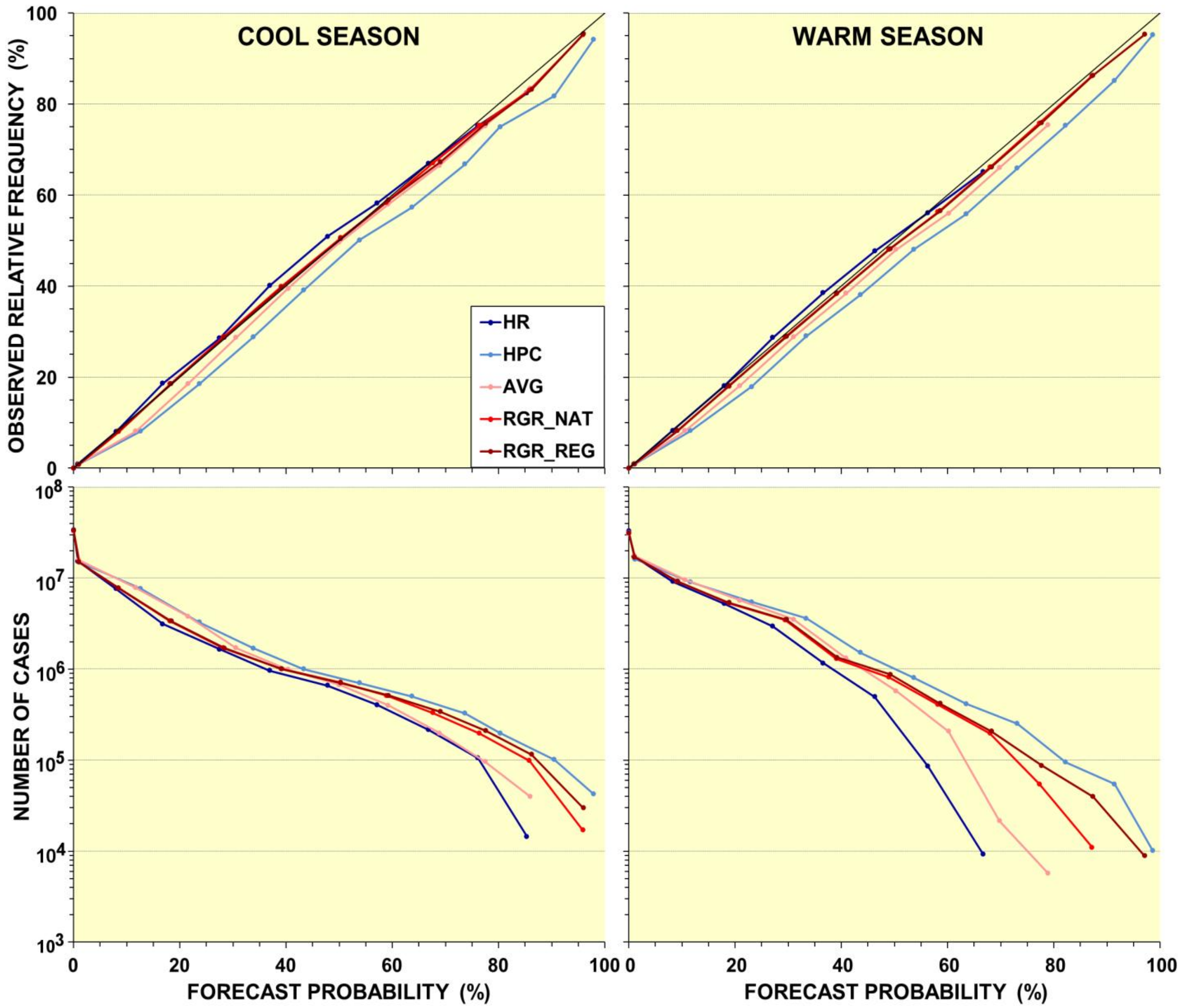


Figure. 4. Seasonal reliability and probability distribution diagrams corresponding to Fig. 3 for ≥ 0.50 in. and day 2 (corresponding diagrams for day 1 and day 3 are similar to these).

Corresponding reliability and probability distributions for light (≥ 0.10 in.) and heavy (≥ 1.00 in.) precipitation thresholds for HR, HPC, and RGR_REG are shown in Fig. 5

(AVG and RGR_NAT are not shown for simplicity and because their PQPF performance lags that for RGR_REG.) Again, we find the reliability for RGR_REG is better than that for both HR and HPC, especially for ≥ 0.10 in. Also, the sharpness for RGR_REG is almost as good as that for HPC. Comparing Fig. 5 with Fig. 4, we see general consistency of the reliability and sharpness properties for HR, HPC, and RGR_REG between the two figures.

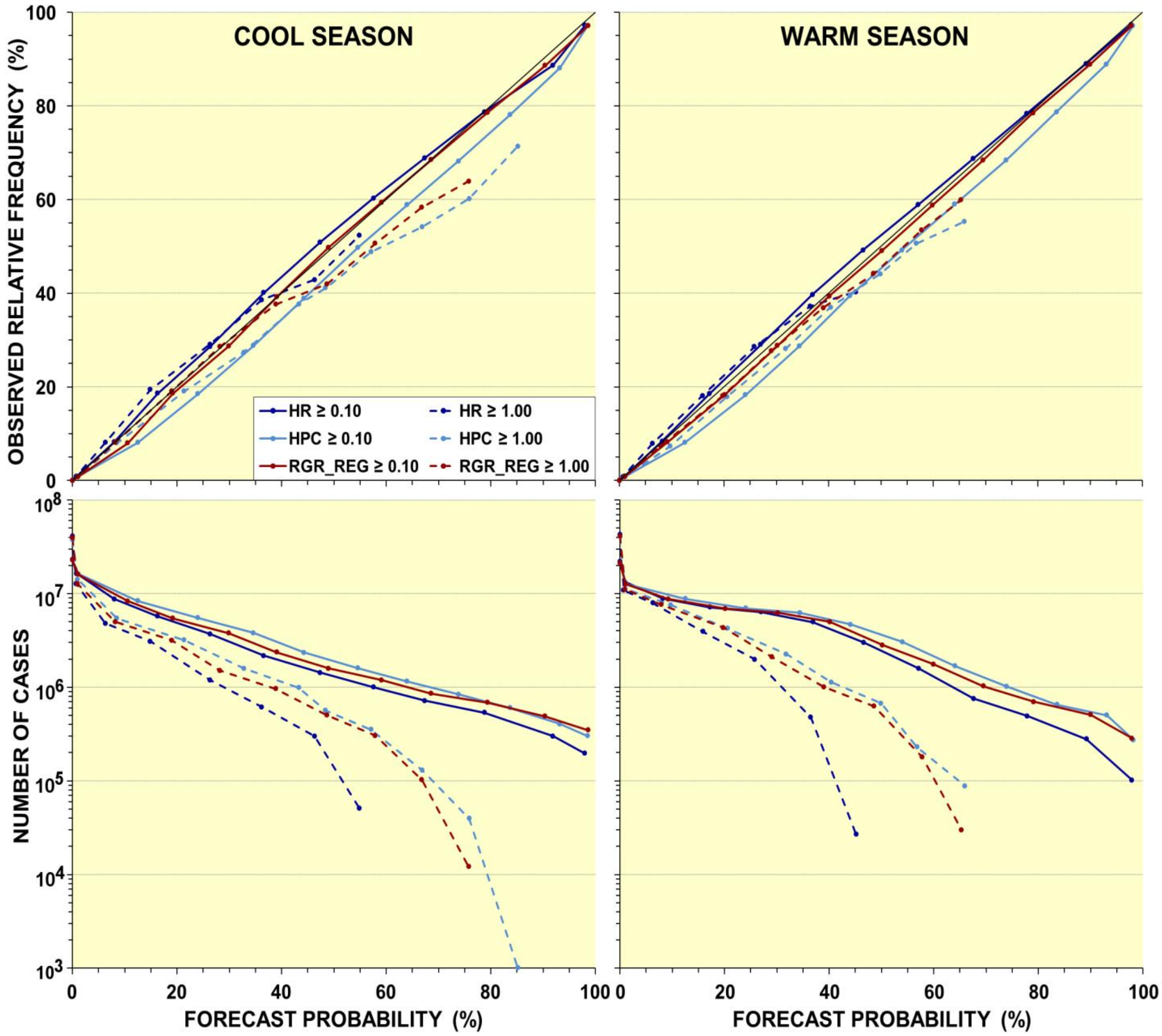


Figure. 5. HR, HPC, and REG PQPF reliability and probability distributions for the same samples as for Fig. 4 for the day 2 forecast period (corresponding charts for day 1 and day 3 are similar to these).

## 4. PQPF MAPS FOR SELECTED HEAVY PRECIPITATION CASES

Stage IV precipitation for 0600 – 1200 UTC, 01 January 2011 (cool season case) along with the associated 72-h probability forecast of ≥ 0.50 in. for each of the PQPF models is shown in Fig. 6 (RGR_NAT and RGR_REG are henceforth denoted NAT and REG, respectively). Note that HPC probabilities are much more focused geographically and peak values are far higher than those for HR. As expected, the AVG probabilities lie about mid-way between those for HR and HPC, while maps for NAT and REG appear more similar to that for HPC than HR. Also, peak probabilities for REG are almost as high as those for HPC, while peaks for NAT are slightly lower[3]. Considering the three day (72 hours) lead time, each PQPF map is reasonably well matched with the observed precipitation field despite the southward positioning error.

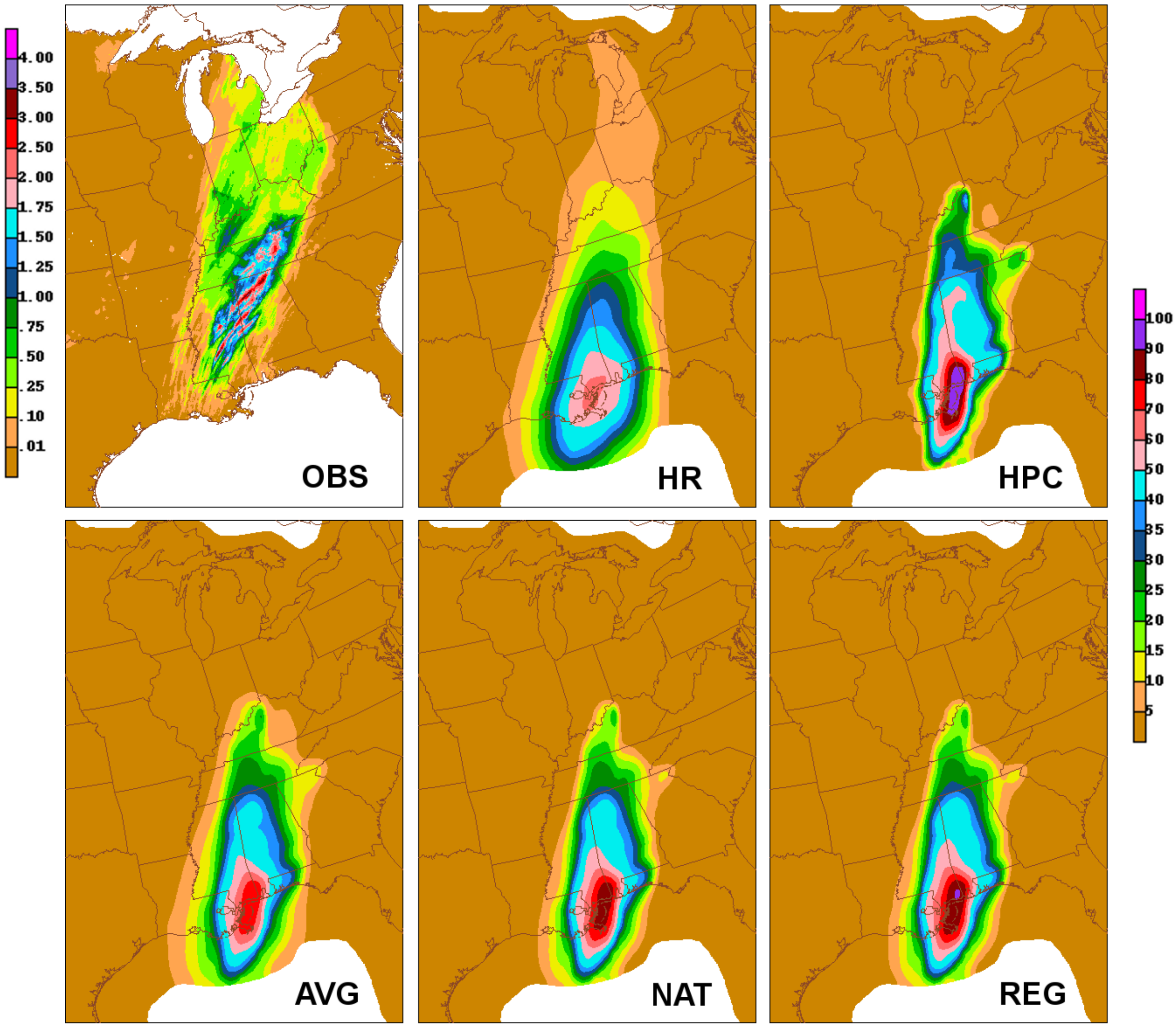


Figure 6. Stage IV precipitation (OBS; in.) and HR, HPC, AVG, RGR_NAT, and RGR_REG 72-h forecast probability (%) of ≥ 0.50 in. for 0600 – 1200 UTC, 01 January 2011 (RGR_NAT and RGR_REG are denoted NAT and REG, respectively).

[3] Very light smoothing was applied the blended probabilities, which enhances spatial coherency in contour maps and adds marginal skill improvement (not shown).

The observed precipitation field for 1800 – 0000 UTC, 10 - 20 December 2010 along with associated HR, HPC, and REG 12-h probabilities of ≥ 0.10 in. are shown in Fig. 7 (AVG and NAT are omitted as for Fig. 5). For this cool season event, the HPC map shows probabilities over 90 % scattered over a broad area of the western U.S., while such probabilities for HR are limited to a small part of California. Given the strong contrasts in probability magnitudes, map patterns, and spatial gradients between HPC and HR, the REG map seems more heavily weighted towards HR, which is consistent with HR's skill improvement over HPC for the light precipitations thresholds (Fig. 3).

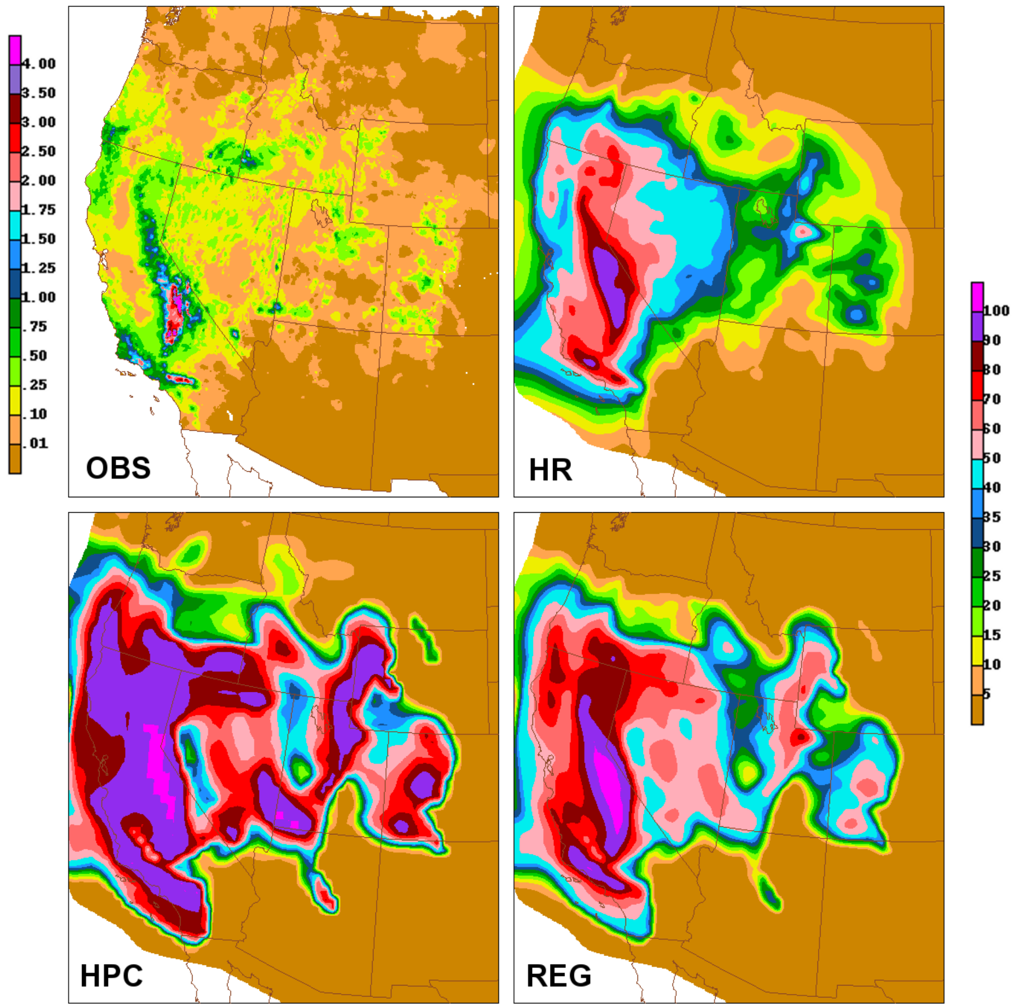


Figure 7. As in Fig. 6, except for 1800 – 0000 UTC, 19 – 20 December 2010, a 12-h lead time, ≥ 0.10 in., and HR, HPC, and REG only.

The last case is a warm season event, involving 12-h probabilities of ≥ 1.00 in. valid 1800 UTC – 0000 UTC, 27 – 28 August 2011-- Hurricane Irene (Fig. 8). Here, the upper probabilities for HPC are far higher than those for HR, and yet the coverage of probabilities above 80 % is similar for REG and HPC. Thus, the HPC probabilities clearly have a greater impact on the REG hybrid, which is consistent with HPC's superior (warm season) skill for ≥ 1.00 in. (Fig. 3).

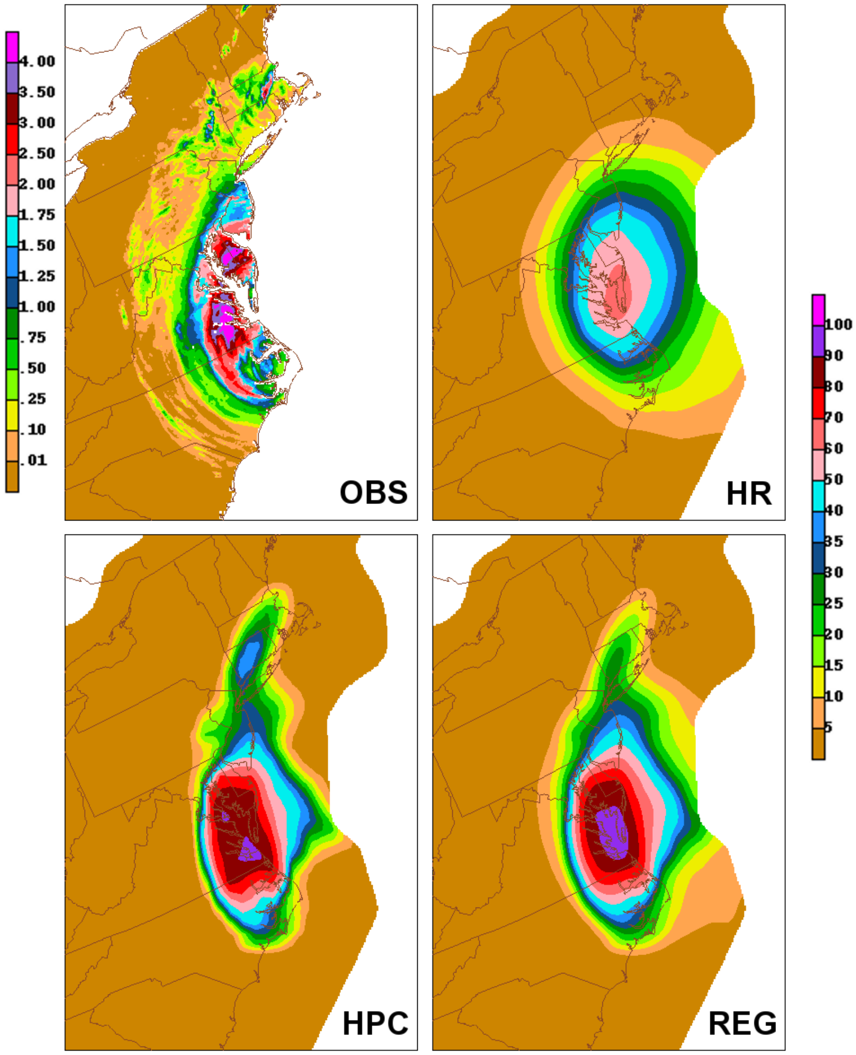


Figure 8. As in Fig. 7, except for 1800 - 0000 UTC, 27 - 28 August 2011, a 12-h lead time, and ≥ 1.00 in.

It is worthy to note that for each of the three cases considered here, the HR PQPFs have better spatial coherency/continuity than the HPC PQPFs. Specifically, the HPC maps commonly exhibit very steep probability gradients in some locations and gradual gradients in others. Very steep PQPF gradients may not be realistic, especially for heavy precipitation thresholds and long forecast projections. The smoother HR map patterns seem more realistic, and this smoothness tends to be incorporated into the REG hybrid PQPF maps. At the same time, we have seen that the hybrid PQPF maps mostly incorporate the peak probabilities and, to some degree, the sharp spatial gradients that characterize the HPC PQPF maps. Thus, the hybrid PQPFs largely incorporate attractive map properties of both the HR and HPC PQPFs.

## 5. DISCUSSION

Figures. 3 - 8 indicate the linear regression blending of the HR and HPC PQPFs resulted in more skillful and more reliable PQPFs than exhibited by either of the component PQPFs. However, it is important to recall that the verification samples underlying these findings were taken from the same historical periods as the samples used to develop the regression hybrid (section 1). That is, the verification samples were formed by withholding every sixth day of the 0000- and 1200-UTC development samples for the regression equations. This could result in an internal correlation between the development and test samples for the regression-hybrid PQPFs, which would give them an unfair advantage in the comparative scoring with HR, HPC and AVG. Thus, the present finding of enhanced PQPF performance with regression blending should be regarded as preliminary.

Unfortunately, the authors are not aware of a robust test of the potential development - test sample inter-dependence given the short historical samples presently available. Thus, a more definitive test of the regression-hybrid PQPF performance awaits the accumulation of a longer historical record of the HPC PQPFs (a longer historical archive is presently available only for HR). It is expected that at least two full cool and warm seasons of data are required (the samples herein consist of about 1.5 seasons), which will become available in the near future.

## 6. PRELIMINARY FINDINGS AND COMMENTS

Preliminary findings from this study suggest that linear regression blending of the HR and HPC PQPFs result in improved PQPF skill and reliability compared that for the input HR and HPC PQPFs and a simple average blend of these inputs. Also, the sharpness of the regression-hybrid PQPFs approached the sharpness level of HPC, where the sharpness was superior to all PQPF models considered. At the same time, the high spatial coherency property of the HR PQPF maps was evident in the regression-hybrid PQPF maps.

On the other hand, the formulation of more robust conclusions regarding PQPF enhancement with regression blending of the HR and HPC PQPFs awaits longer historical samples than those presently available. Also, longer development samples will support further stratification of the regression PQPF blending, which could further enhance the forecast performance. In particular, enhancements should result from extending the seasonal stratification from two seasons to three or more seasons (CS) and developing separate regression equations for the 0000- and 1200-UTC cycles.

With expected operational implementation of the HR PQPFs in the near future, the regression-hybrid PQPFs could be implemented as a small add-on to the HR PQPF program. The latter PQPFs could then replace the HR PQPFs and additional QPF products could be derived from the former through post-processing.

## 7. REFERENCES


Charba, J. P., and F. G. Samplatsky, 2011a: Regionalization in fine grid GFS MOS 6-h quantitative precipitation forecasts. *Mon. Wea. Rev.*, **139**, 24-38.

___, and ___, 2011b: High resolution GFS-based MOS quantitative precipitation forecasts on a 4-km grid. *Mon. Wea. Rev.*, **139**, 39-68.

___, and ___, 2011c: Comparative verification of experimental HR QPFs with model and forecaster-prepared QPFs in the NWS. *Extended Abstracts, 24th Conference Weather Analysis and Forecasting/20th Conf. on Numerical Weather Prediction*, Amer. Meteor. Soc., Seattle, WA, **10.B2**.

Novak, D. R., K. Brill, M. Eckert, R. Oravec, B. Sullivan, R. Bann, F. Barthold, and M. Bodner, 2011: Quantifying extreme rainfall threats at the Hydrometeorological Prediction Center. *Extended Abstracts, 24th Conference Weather Analysis and Forecasting/20th Conf. on Numerical Weather Prediction*, Amer. Meteor. Soc., Seattle, WA, **14A.4**.