ENSEMBLE STREAMFLOW FORECASTING FOR THE UPPER TRINITY RIVER BASIN IN TEXAS

by

MANABENDRA SAHARIA

Presented to the Faculty of the Graduate School of
The University of Texas at Arlington in Partial Fulfillment
of the Requirements
for the Degree of

MASTER OF SCIENCE IN CIVIL ENGINEERING

THE UNIVERSITY OF TEXAS AT ARLINGTON

August 2013

ACKNOWLEDGEMENTS

It is difficult to overstate my gratitude to my advisor, Dr. Dong-Jun Seo. Without his active supervision and constant support, this journey wouldn't have been possible.

My sincere appreciation also goes out to Dr. Kevin He of National Weather Service who provided critical technical and scientific help throughout the course of this work. I wish to thank Mr. Robert Corby, Dr. Hank Herr, Dr. Satish Regonda, Dr. Limin Wu, Mr. Mark Fresch, Mr. Gautam Sood, Ms. Xuning Tan and Dr. James Brown, for help at various stages of this work. I also thank Dr. Kevin He, Dr. Hyeok Choi and Dr. Xinbao Yu for serving on my thesis committee. I am also very thankful to all my colleagues in the Hydrology and Water Resources Laboratory (HWRL) for providing a stimulating research environment.

My sincere thanks to Dr. Rajib Bhattacharjya (IIT Guwahati), Prof. Sharad K. Jain (IIT Roorkee), Dr. Parthasarathi Choudhury (NIT Silchar) and Dr. Parthajit Roy (NIT Silchar) for inspiring me to tread this path.

This thesis would not have been possible without the love and support of my family. My deepest gratitude to my parents for teaching me the importance of learning and all their sacrifices for me. Also my brother, Dhiraj, of whom I am very proud of and Bandita, for supporting me in every step.

July 15, 2013

ABSTRACT

ENSEMBLE STREAMFLOW FORECASTING FOR THE

UPPER TRINITY RIVER BASIN IN TEXAS

MANABENDRA SAHARIA, M.S.

The University of Texas at Arlington, 2013

Supervising Professor: Dr. Dong-Jun Seo

By allowing for routine use of longer-lead quantitative precipitation forecast (QPF) in hydrologic prediction, ensemble forecasting offers hope for extending the lead time for short-range streamflow forecasting. In this work, this potential is assessed by comparatively evaluating ensemble streamflow hindcasts forced by Day 1-3 QPF with those forced by Day 1 QPF for five headwater basins in the Upper Trinity River Basin in North Texas. The hindcasts are generated for a 7-year period of 2004 to 2010 using the Hydrologic Ensemble Forecast Service (HEFS), which operates on the Community Hydrologic Prediction System (CHPS) of the National Weather Service (NWS). HEFS includes the Meteorological Ensemble Forecast Processor (MEFP), Ensemble Postprocessor (EnsPost) and Ensemble Verification System (EVS). In this study, MEFP is used to generate ensemble QPFs from the West Gulf River Forecast Center (WGRFC)-produced single-valued QPFs, EnsPost is used to post-process the streamflow hindcasts in terms of correcting hydrologic bias involved and EVS is used to verify the precipitation and streamflow ensemble hindcasts. The results show that: (1) The ensemble QPFs produced from single-valued QPFs using MEFP are generally skillful and reliable, (2) Using Day 1-3 single-valued QPF via HEFS significantly increases the skill in short-range Ensemble Streamflow Prediction (ESP) forecasts; and (3) Post-processing of ESP ensembles via EnsPost improves discrimination and reliability of the raw ESP ensembles. Finally, the issues and challenges are identified and future research recommendations are provided.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Figure Page

LIST OF TABLES

Table Page

CHAPTER 1

INTRODUCTION

Predicting river flows from observations and predictions of precipitation and other meteorological variables is an important topic in applied hydrology. Accurate river flow forecasting is not only important for flood forecasting, but also for a range of applications related to design, operation and management of water resources structures. River forecasts with sufficient lead time are important for issuing warnings and enacting preventive measures, including evacuation. Increasing lead time also provides significant economic benefits as it facilitates cost-effective water management and provides protection from water shortages.

Currently, the National Weather Service (NWS) West Gulf River Forecast Center (WGRFC) produces single-valued river stage forecasts for the Texas region. For future precipitation, the operational practice at WGRFC is to input single-valued quantitative precipitation forecast (QPF) out to only 12-24 hours with no precipitation assumed thereafter. In a single-valued mode, getting longer-range QPF right is extremely difficult due to lack of predictability. The objective of this study is to produce skillful forecasts of streamflow at longer lead times using longer-lead QPFs and post-processing of simulated streamflow. The premise of this work is that, via ensemble forecasting, one may utilize the skill in longer-lead QPF and hence extend the lead time of streamflow forecasts.

Single-valued forecasting is currently the dominant paradigm in streamflow forecasting. Streamflow forecasts are, however, inherently uncertain due to uncertainties in inputs, initial conditions (IC), parameters and structures of hydrologic models. To allow risk-based decision making, streamflow forecasts should convey uncertainty information. Quantifying predictive uncertainty in streamflow forecasts is an active area of research and is recognized as one of the most pressing needs in operational hydrology today.

Owing to the disparate sources of uncertainty involved, hydrologic prediction with reliable uncertainty estimation is a challenge (e.g. [1], [2], [3]). In the past several years, a number of quantitative and qualitative approaches have been used to assess uncertainty and utilize that

information in decision making (for reviews; see [4], [5], [6]). A large number of methods have emerged for quantification of uncertainty in operational hydrologic forecasting. They are based on one of the two modeling approaches - the lumped approach, which directly models the total uncertainty and the source-specific approach which models specific sources of uncertainty separately and combines them to get the total uncertainty. A distinction is made between input and hydrologic uncertainty, the latter of which includes uncertainties in the ICs, parameters and structures, which are usually lumped together and modeled as hydrologic uncertainty ([19], [20], [21], [22], [2], [4], [23]). An example of the lumped approach is Model Output Statistics (MOS; [24], [25], [26]), in which a statistical relationship is estimated between historical forecasts and observations. Based on the relationship, real-time forecasts are subjected to an adjustment and the uncertainty is estimated from the historical residuals surrounding the regression relationship. As source-specific approaches model key sources of uncertainty, they can incorporate human forecaster judgment and prior information ([27], [28]) to make targeted improvements in modeling ([29], [30], [31]).

The first hydrologic ensemble prediction operation was developed by the National Weather Service (NWS; [35]) in the 1970's. Ensemble Streamflow Prediction (ESP) was added to the NWS River Forecast System (NWSRFS) to generate streamflow ensembles up to 90 days ahead based on historical precipitation and temperature observations. The system was further improved by using seasonal "climate outlooks" produced by the Climate Prediction Center (CPC) of the National Centers for Environmental Prediction (NCEP) and short-range QPF produced by the NWS River Forecast Centers (RFC). Recently, an application called the Hydrologic Ensemble Forecast Service (HEFS) has been developed by the NWS for ensemble forecasting of precipitation, temperature and streamflow from one hour to one year into the future ([9], [10]). HEFS adopts the source-specific method by modeling meteorological and hydrologic uncertainties separately via the Meteorological Ensemble Forecast Processor (MEFP) and Ensemble Postprocessor (EnsPost). Ensemble forecasts of precipitation are produced by MEFP from single-valued QPFs based on the bivariate probability distribution between the observed precipitation and the single-valued QPF [12]. Streamflow prediction takes place in two steps. Firstly, "raw" streamflow forecasts are generated from the future ensemble forcings produced by MEFP. The resulting

streamflow ensembles may contain biases in the mean and in the higher-order moments. These "raw" streamflow forecasts are post-processed with EnsPost, which accounts for hydrologic uncertainty and reduces biases [2]. EnsPost models the joint probability distribution of the observed and simulated streamflows in normal quantile space (NQT; [1], [2]). Currently, Version 1 of HEFS is scheduled for operational use in all RFCs by 2014.

In this work, HEFS is used to assess the value of utilizing Day 2-3 QPF in short-term ESP via ensemble hindcasting and verification. The organization of this thesis is as follows. Following this introductory chapter is Chapter 2, which presents details of HEFS and its various components. Chapter 3 describes the approach and methodology adopted for this study, as well as a description of the experiments. The results from the various experiments and their analysis is provided in Chapter 4. Chapter 5 summarizes the thesis, including the conclusions and future research recommendations drawn from this research. Concluding this thesis is the bibliography.

CHAPTER 2

HYDROLOGIC ENSEMBLE FORECAST SERVICE (HEFS)

HEFS refers to a suite of modules in the Community Hydrologic Prediction System (CHPS) that can be used for a wide spectrum of applications such as flood forecasting, water supply management and water quality forecasting. The ability to issue reliable and skillful probabilistic predictions is dependent on accurately accounting for the various sources of uncertainty. HEFS explicitly accounts for the major sources of uncertainty by modeling meteorological and hydrologic uncertainties separately. Reliable and skillful ensemble forecasts of precipitation and temperature are necessary to produce reliable and skillful hydrologic ensemble forecasts ([2], [7], [8]). While ensemble precipitation forecasts are widely available from various numerical weather prediction (NWP) models, they are generally not very reliable due to the various sources of error in NWP model predictions. The purpose of MEFP is to produce reliable ensemble forecasts of future forcings that capture the skill contained in the source forecasts (single-valued or ensemble).

HEFS propagates the forcing ensembles generated by MEFP through a suite of hydrologic models to get raw streamflow forecasts. For hydrologic uncertainty, HEFS does not model individual sources of uncertainty, such as ICs, parametric and structural uncertainties. Instead, it models the collective uncertainty as a single stochastic term using EnsPost [2]. A hindcast is a simulated forecast generated by going back in time and treading the forecast process as if one is trying to predict the future without the benefit of any verifying truth. All precipitation and streamflow hindcasts (raw and post-processed) are verified using Ensemble Verification System (EVS; [5]). Figure 2.1 shows the schematic of various components within HEFS.

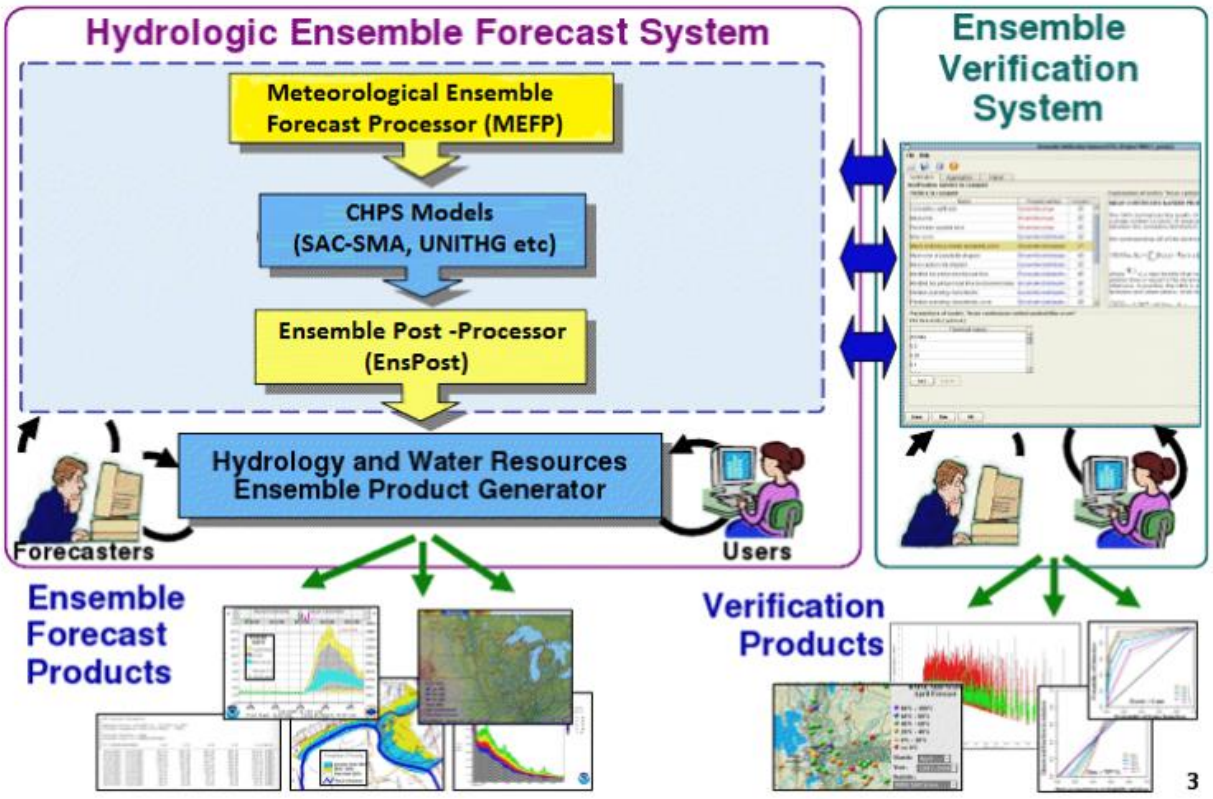

Figure 2.1. Schematic of Hydrologic Ensemble Forecast Service (HEFS) [43]

2.1 Meteorological Ensemble Forecast Processor (MEFP)

MEFP models the uncertainty in the future forcings of precipitation and temperature and has two components: the MEFP Parameter Estimator (MEFP PE) and the MEFP Ensemble Generator (MEFP EG). It is an integral component of HEFS and models the uncertainty associated with future mean areal precipitation (FMAP) and future mean areal temperature (FMAT) that are inputs to the hydrologic model. MEFP is capable of utilizing forecast information from multiple sources of weather and climate prediction models such as the NCEP's Global Ensemble Forecast System (GEFS) forecasts and single-valued forecasts produced by the RFC ([7], [11], [12]). Additionally, climatological ensembles up to one year can also be generated from historical observations or from statistically smoothed climatology from historical observations [32].

MEFP is based on the statistical technique of Schaake et al. (2007) [11] to generate ensemble forecasts of precipitation and temperature from single-valued QPF and quantitative

temperature forecasts (QTF), respectively. It is based on the modeled bivariate probability distributions between the observed precipitation and temperature and the single-valued QPF and QTF, respectively. MEFP also includes an improved precipitation ensemble generation technique based on a mixed-type bivariate meta-Gaussian distribution model of Wu et al. (2011) [12]. For precipitation or temperature, MEFP PE estimates the joint probability distribution between observations and single-valued forecasts. MEFP EG then samples precipitation or temperature ensembles from the conditional distribution given the single-valued QPF or QTF, respectively. Finally, Schaake Shuffle [15] is used to rank the ensemble members and construct the time series according to the ranks of historical observations.

2.2 Ensemble Postprocessor (EnsPost)

The forcing ensembles generated using MEFP are ingested into the suite of hydrologic models of CHPS to generate raw streamflow ensembles that reflect input uncertainty. These raw ESP ensembles do not account for hydrologic uncertainty and may be biased in their mean, spread and higher-order moments. In order to correct for biases and to model the hydrologic uncertainty, HEFS uses the Ensemble Postprocessor, or EnsPost, developed by Seo et al. [2]. EnsPost attempts to model all sources of hydrologic uncertainty in a lumped fashion to generate post-processed streamflow ensembles that reflect both input and hydrologic uncertainty.

The importance of accounting for hydrologic uncertainty in ensemble forecasting, especially in short-term forecasting, has been documented in several studies [16]. EnsPost employs a combination of probability matching and recursive linear regression [2]. It stratifies the regression according to season and flow magnitude in order to reflect seasonality and flow regime-dependent temporal correlation structures in streamflow. Historical time series of observed and simulated flow are converted into normal variates via Normal Quantile Transformation (NQT; [36], [37]). A regression is obtained for each combination of season and flow regime and optimal parameters are obtained by minimizing a combination of error statistics. Two kinds of error models are currently available in EnsPost, the Stochastic Error Model (ERS) and the Deterministic Error Model (ERD). In the real-time mode, the 6-hourly streamflow forecasts are aggregated to daily forecasts, the daily forecasts are post-processed, and the post-processed daily streamflow ensembles are disaggregated for sub-daily periods. The efficacy of

EnsPost is subject to availability of long-term streamflow data and stationarity of streamflow climatology.

2.3 Ensemble Verification System (EVS)

Most of the techniques used to verify ensemble forecasts were pioneered in meteorology [13]. EVS developed by Brown et al. (2010) [5] is a flexible and modular open-source component of HEFS. EVS is programmed in Java and designed to verify forcing and hydrologic ensembles. Using this tool, forecasts can be evaluated for a range of forecast horizon, space-time scale, seasonality and magnitude of event. Forecasts can also be aggregated in time, such as daily precipitation values from 6-hourly forecasts or aggregated spatially to verify ensembles across several discrete locations. Graphical and numerical outputs can be generated in several different formats. The quality of a forecast can be evaluated by measuring key attributes such as reliability or unbiasedness of the forecast probabilities, discrimination (ability to distinguish between events and non-events) and skill relative to a baseline forecasting system ([17], [13]). Brown et al. (2010) [5] also presents a diagnostic verification strategy for hydrologic ensembles based on selection of summary verification metrics and analysis of the relative contributions of different sources of error. EVS is currently used operationally at several NWS RFCs and supports scientific research and development ([7], [12], [5]).

2.4 Community Hydrologic Prediction Service (CHPS)

Since the 1970's, the NWS has been using NWSRFS for river forecasting which is now being replaced by CHPS. CHPS is a customized version of Delft-Flood Early Warning System (FEWS) developed by Deltares of Netherlands. Being an open source tool adopted by a number of forecasting agencies around the world, CHPS/FEWS provides a common infrastructure for hydrologic modeling for cost-effective infusion of advances in science and technology ([40], [41], [42]).

CHAPTER 3

APPROACH AND METHODOLOGY

This chapter describes the approach and methodology used in this study, the hindcasting experiments conducted, the study area and the data used.

3.1 Hindcasting Experiments

To assess the value of Day 2-3 QPF to short-range ESP, a set of ensemble hindcasting experiments and ensemble verification were designed and carried out. The hindcasting process follows the schematic in Figure 3.1. This section describes the process in some detail.

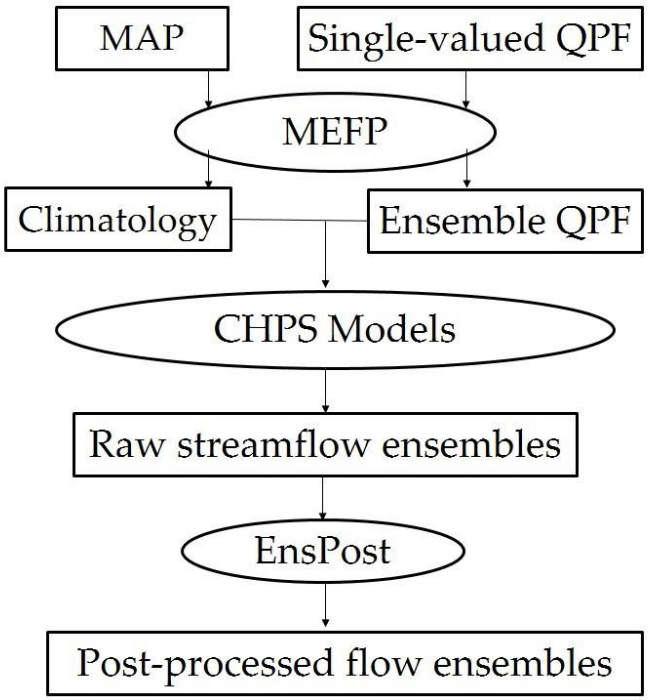

Figure 3.1 Schematic view of data flow of information in HEFS

Experiment 1. Generate Day 1 ensemble QPF (EQPF) from Day 1 single-valued QPF and Day 2-14 ensemble precipitation hindcasts from climatology, input the Day 1-14 ensemble precipitation hindcasts into ESP to generate raw ESP ensembles. Here EQPF and ensemble precipitation hindcast are used interchangeably,

Experiment 2. Apply EnsPost to the raw ESP ensembles in Experiment 1,

Experiment 3. Generate Day 1-3 EQPF from Day 1-3 single-valued QPF and Day 4-14 ensemble precipitation hindcasts from climatology, input Day 1-14 ensemble precipitation hindcasts into ESP to generate raw ESP ensembles, and

Experiment 4. Apply EnsPost to the raw ESP ensembles in Experiment 3.

By comparing the verification results between the ensembles forced by longer-lead and short-lead QPFs, one may quantify the contribution of longer-lead QPF to skill in ESP and increase in lead time. From the verification results of raw and post-processed streamflow forecasts, the improvement in skill due to streamflow ensemble post-processing may be assessed. To quantify the skill from each operation, verification was performed on precipitation, raw streamflow and post-processed streamflow hindcasts using EVS.

3.2 Study Area and Dataset

The above hindcasting experiments were conducted for five headwater basins of the Upper Trinity River: Big Sandy Creek near Bridgeport (BRPT2), Denton Creek near Justin (DCJT2), Elm Fork of the Trinity River near Gainesville (GLLT2), West Fork of the Trinity River near Jacksboro (JAKT2) and Clear Creek near Sanger (SGET2). Figure 3.2 and Table 3.1 show the locations and characteristics of the basins, respectively. Improved streamflow prediction for headwater basins improves prediction downstream. As such, skillful streamflow prediction for these basins is very important for the Dallas-Fort Worth metropolitan area (DFW) for flood warning, water supply, reservoir operations and water quality management. It is noted here that DFW is one of the fastest growing urban areas in the U.S. and almost all of its water needs are met by surface water sources [39].

Table 3.1. Basin ID

Characteristics	BRPT2	DCJT2	GLLT2	JAKT2	SGET2
Latitude	33.23	33.12	33.62	33.29	33.34
Longitude	-97.69	-97.29	-97.15	-98.08	-97.18
Area (km^2)	862.47	1036.00	450.66	1769.00	764.05
Mean Elev. (m)	229	197	227	279	193
Action Stage (m)	3.35	2.74	6.10	5.49	7.01
Flood Stage (m)	3.65	3.05	6.71	6.10	7.62
Time to Peak (hours)	24	12	12	24	12

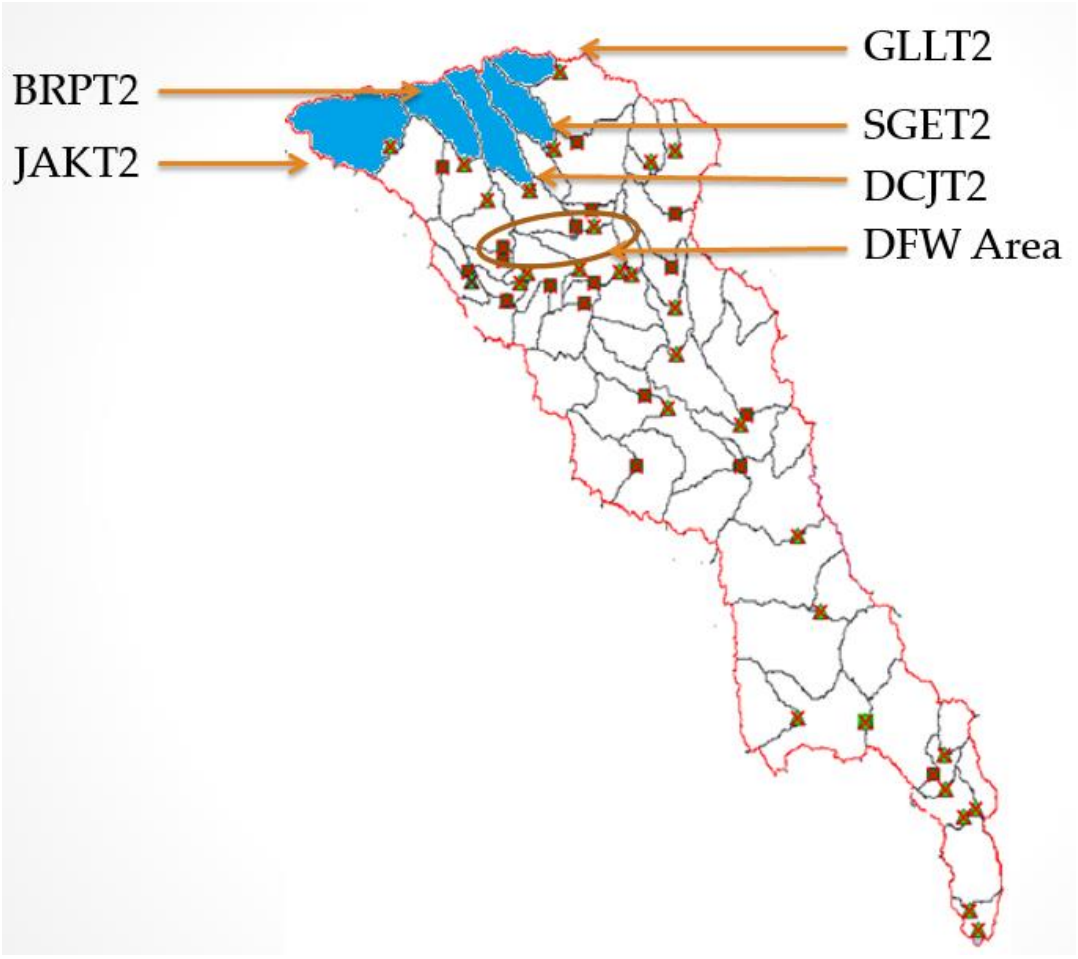

Figure 3.2. Map of Upper Trinity River Basin

For our hindcasting study, the following data was used (the period of record is shown in parentheses):

- Historical single-valued 6-hour QPF for 2004-2010 for 1 to 3 days into the future operationally produced by the WGRFC,
- MAP for 1960 to 2010 produced by the WGRFC and
- Observed mean daily streamflow for 1960-2012 produced by the U.S. Geological Survey (USGS) and provided by WGRFC.

Observed precipitation for each basin consists of the 6-hour observed MAP produced by WGRFC using rain-gage observations in the earlier years and the Multisensor Precipitation Estimator (MPE) estimates since mid-1990s. Precipitation observations are available at 0Z, 6Z, 12Z and 18Z from 1960 to 2010.

Based on the availability of historical single-valued QPF, precipitation hindcasts were generated for the 7-year period of 2004-2010. It is possible that the period may not be representative of the hydroclimatology of the study area over a longer period. Investigation of long-term changes in hydroclimatology, however, was beyond the scope of this work. Each ensemble hindcast consisted of 46 members with lead time of up to 14 days (336 hours in 6 hour increments). Raw streamflow hindcasts were generated with CHPS using the precipitation hindcasts from MEFP. As Texas does not have a significant snow season, SNOW-17 [38] and hence temperature data was not used. The Sacramento Soil Moisture Accounting model (SAC-SMA, [34]) was used for rainfall-runoff modeling and unit hydrograph was used for hydrologic routing.

CHAPTER 4

RESULTS

This section presents the results from the end-to-end multi-year hindcasting using HEFS. Ensemble precipitation and streamflow hindcasts generated for an approximately 7-year period (2004-2010) are verified for all observed-forecast pairs.

Verification results of ensemble precipitation hindcasts generated from 1 and 1-3 Day RFC QPFs via MEFP are compared. The ensemble hindcasts generated using MEFP are verified against the observed MAP in daily amounts and the results are presented by forecast lead time and magnitude of the precipitation amount.

4.1 Correlation Coefficient

In order to first compare the results in a single-valued forecast sense, the correlation coefficient is used. Correlation Coefficients are calculated between the ensemble mean and observation for both precipitation and streamflow. It is a measure of strength in linear association between the forecasts and observations.

Figure 4.1 shows the correlation coefficient of Day 1 ensemble mean QPF and observed precipitation as a function of lead time for BRPT2. As expected, the Day 1 EQPF shows significant correlation for the first day but drops off to negligible levels after that. Each line represents a different threshold of precipitation amount expressed as percentiles of observed precipitation in the 7-yr period (50^{th}, 75^{th}, 90^{th} and 95^{th} percentiles). Figure 4.2 shows the correlation coefficient of Day 1-3 ensemble mean QPF and observed precipitation plotted against lead time. Note the increased correlation in Days 2 and 3. As noted, the objective of this study is to utilize this skill in Days 2 and 3 to increase the quality and lead time of streamflow forecasts. For higher thresholds, the sample size is rather small and, as such, significant sampling uncertainties exist in the verification statistics. For the 90^{th} and 95^{th} percentiles, the precipitation values are only 1.3843 and 3.6347 (mm) for BRPT2, 1.4986 and 3.6677 (mm) for DCJT2, 1.3157 and 4.1249 (mm) for GLLT2, 1.3106 and 3.4366 (mm) for JAKT2 and 1.3817 and 3.7109 (mm) for SGET2, respectively.

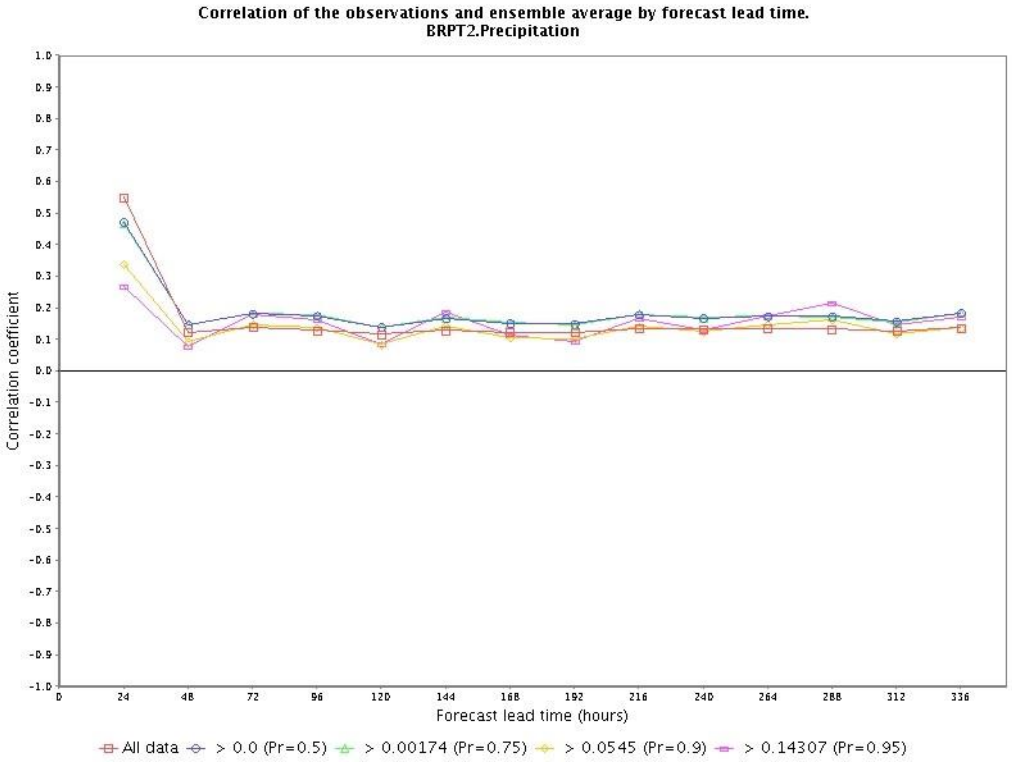

Figure 4.1. Correlation of mean of Day 1 EQPF and observed precipitation for BRPT2

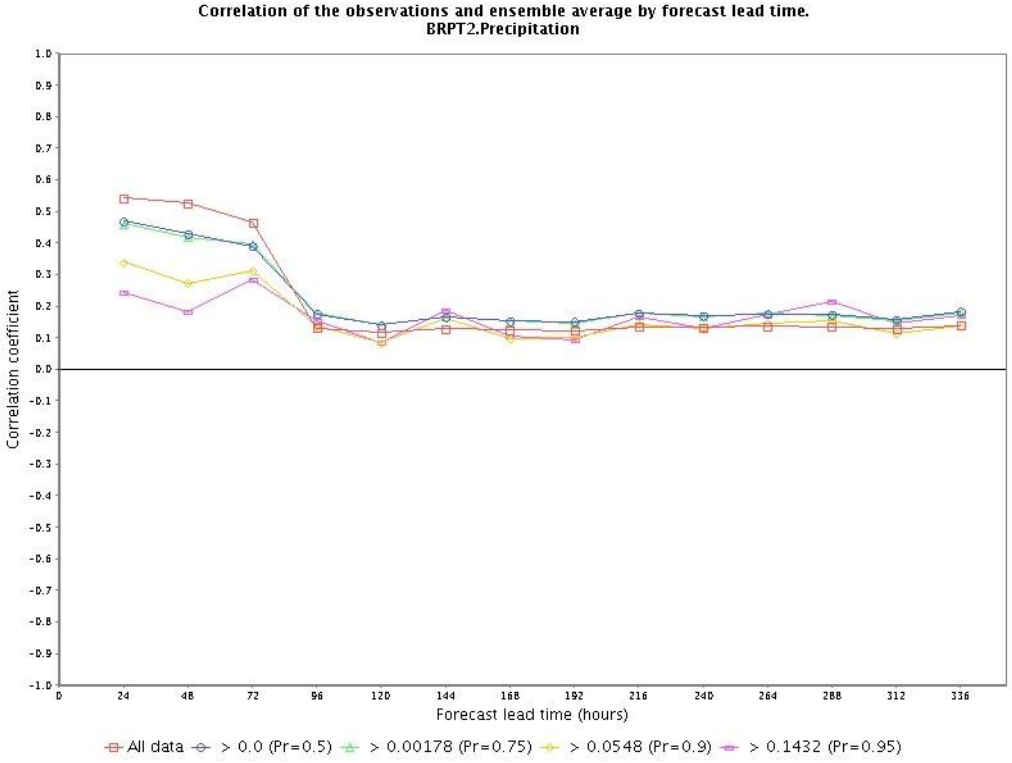

Figure 4.2. Correlation of mean of Day 1-3 EQPF and observed precipitation for BRPT2

For DCJT2, Figures 4.3 and 4.4 show the correlation coefficient of Day 1 as well as Day 1-3 ensemble mean QPF and observed precipitation respectively. The skill for the 95th percentile is negligible for Day 1 but increases for longer lead times. This is a sampling artifact; the sample size is very small (~100) and hence there exists large sampling uncertainty.

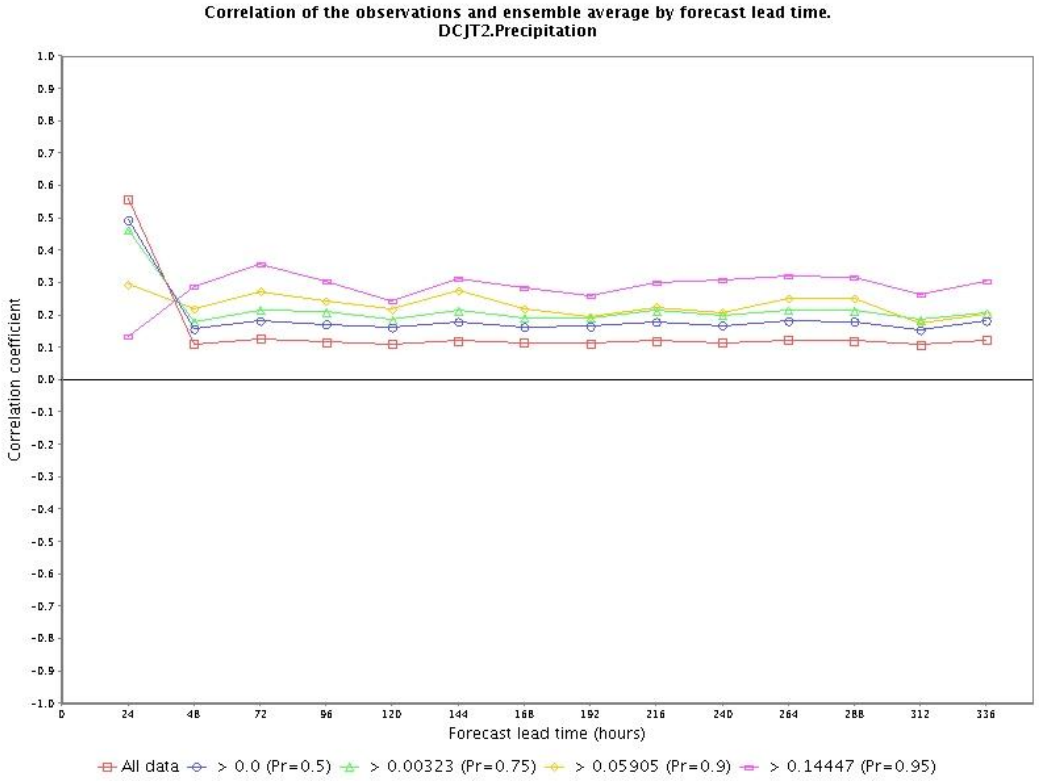

Figure 4.3. Correlation of mean of Day 1 EQPF and observed precipitation for DCJT2

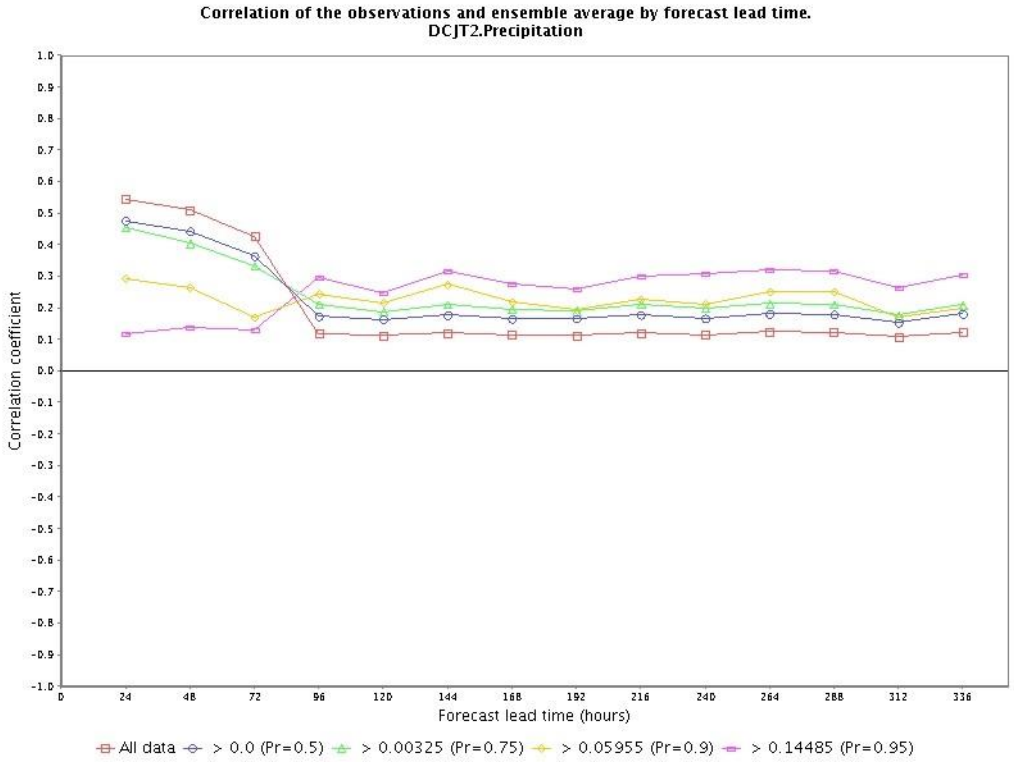

Figure 4.4. Correlation of mean of Day 1-3 EQPF and observed precipitation for DCJT2

Figures 4.5, 4.7 and 4.9 show the correlation coefficients of Day 1 ensemble mean QPF and observed precipitation against lead time for GLLT2, JAKT2 and SGET2 respectively. Also, Figures 4.6, 4.8 and 4.10 show the correlation coefficients of Day 1-3 ensemble mean QPF and observed precipitation for GLLT2, JAKT2 and SGET2 respectively. The correlation coefficients for these basins follow a pattern similar to DCJT2's. The skill for 95th percentile for GLLT2, JAKT2 and SGET2 is negligible for Day 1 and increases for longer lead times, which can be attributed to existence of sampling uncertainty.

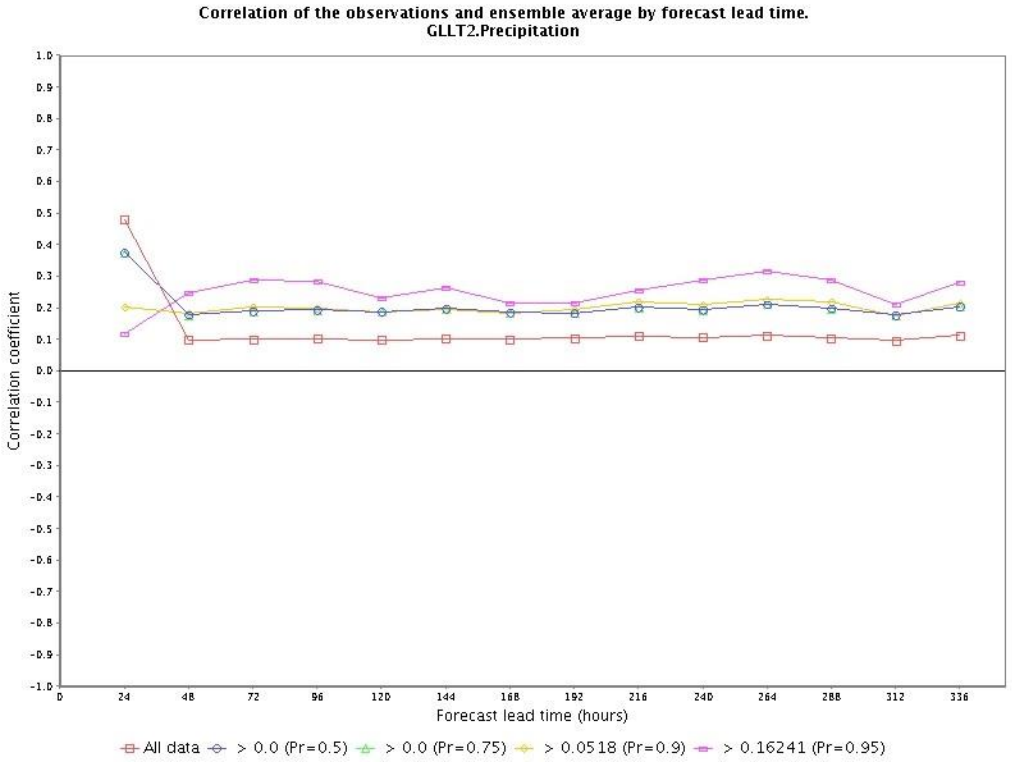

Figure 4.5. Correlation of mean of Day 1 EQPF and observed precipitation for GLLT2

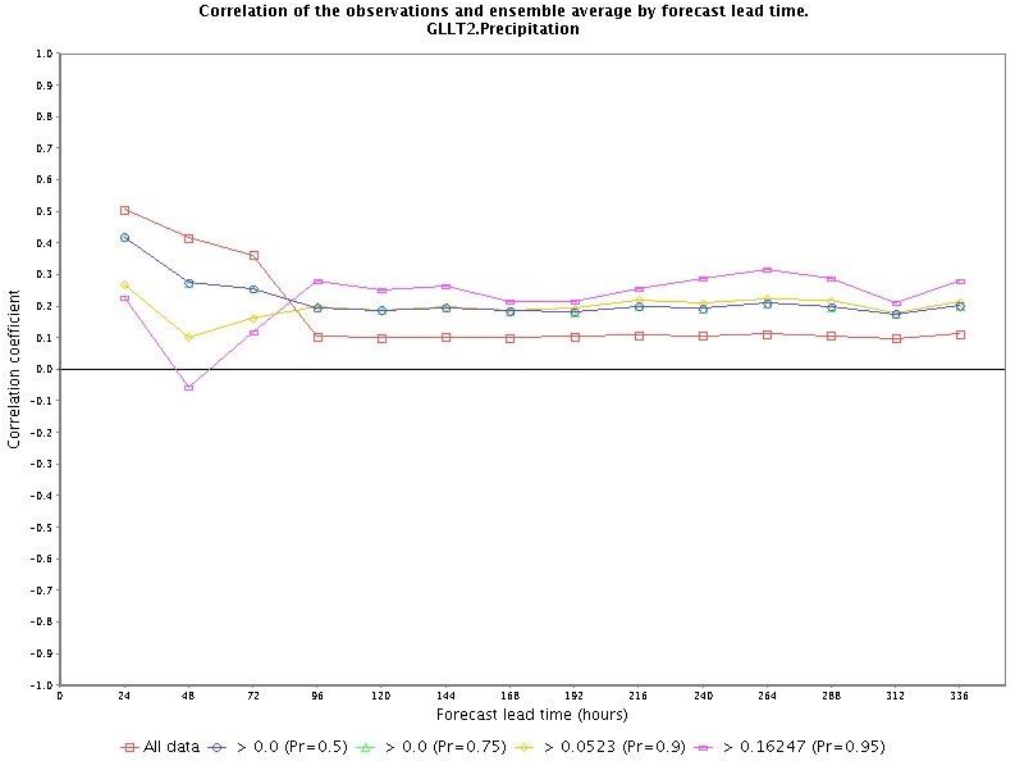

Figure 4.6. Correlation of mean of Day 1-3 EQPF and observed precipitation for GLLT2

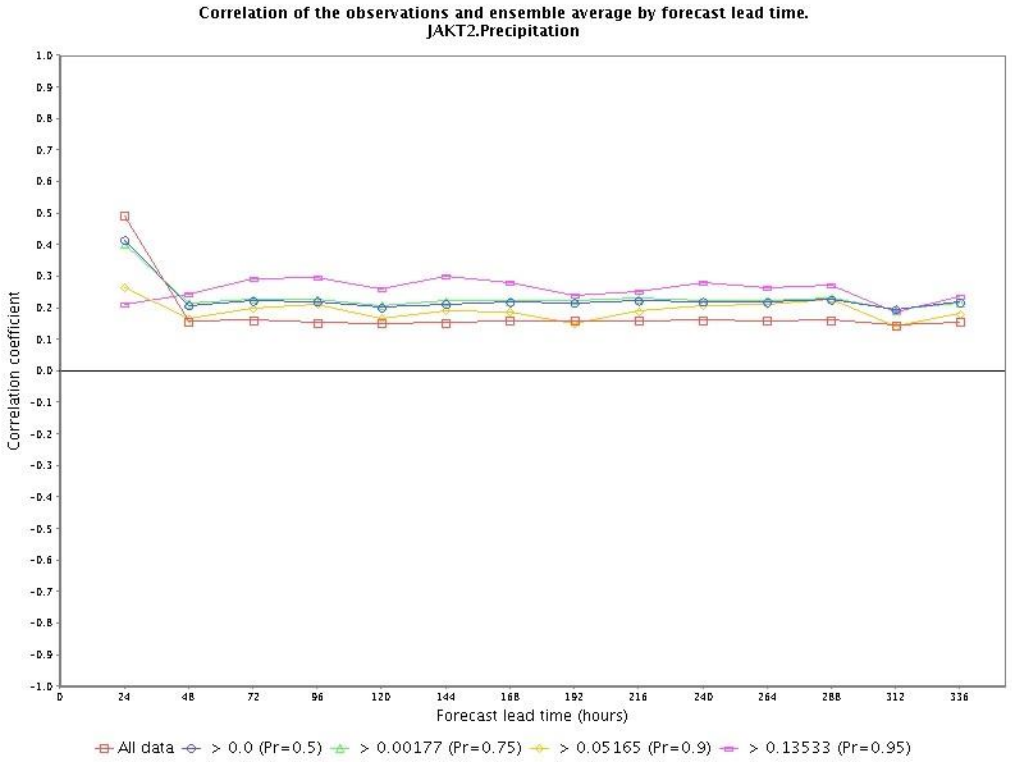

Figure 4.7. Correlation of mean of Day 1 EQPF and observed precipitation for JAKT2

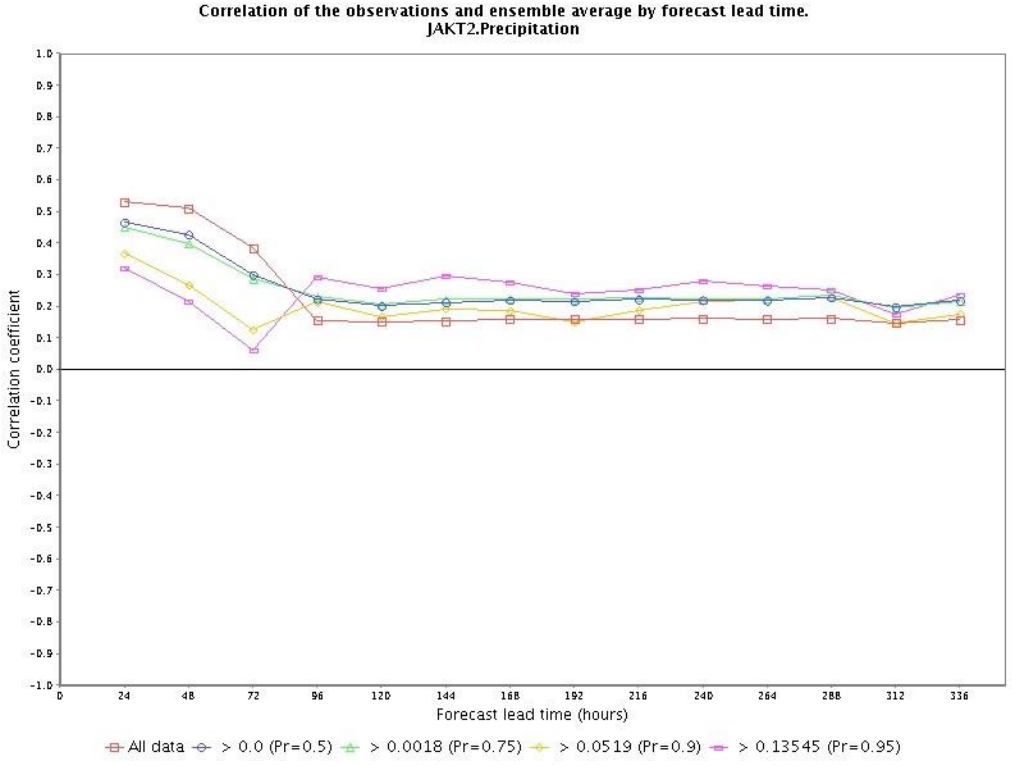

Figure 4.8. Correlation of mean of Day 1-3 EQPF and observed precipitation for JAKT2

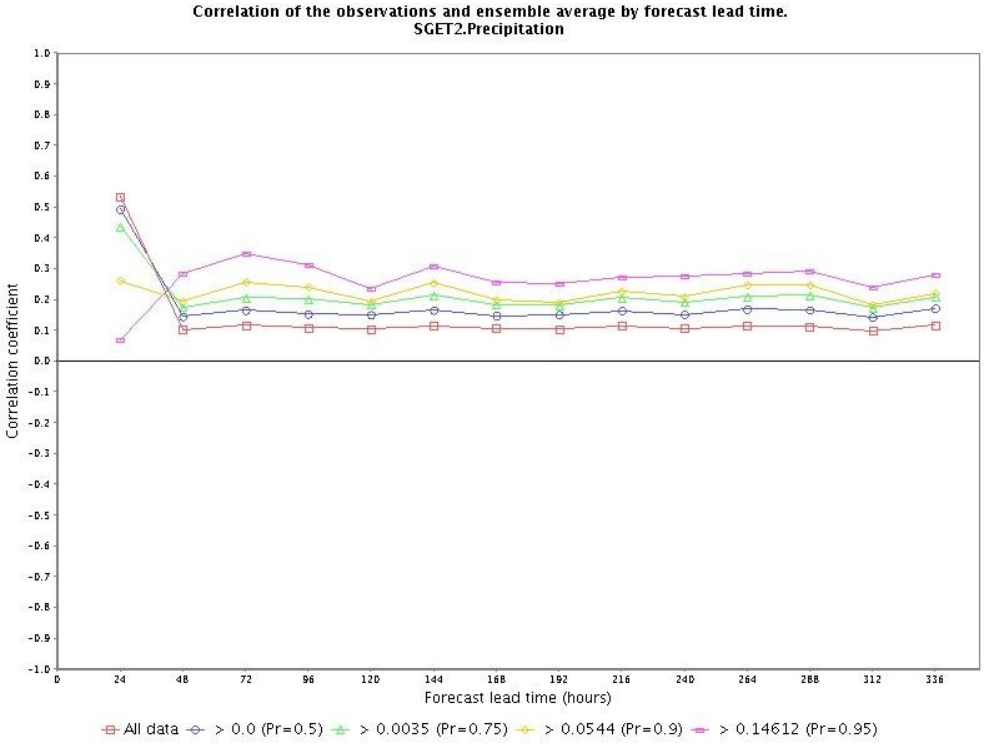

Figure 4.9. Correlation of mean of Day 1 EQPF and observed precipitation for SGET2

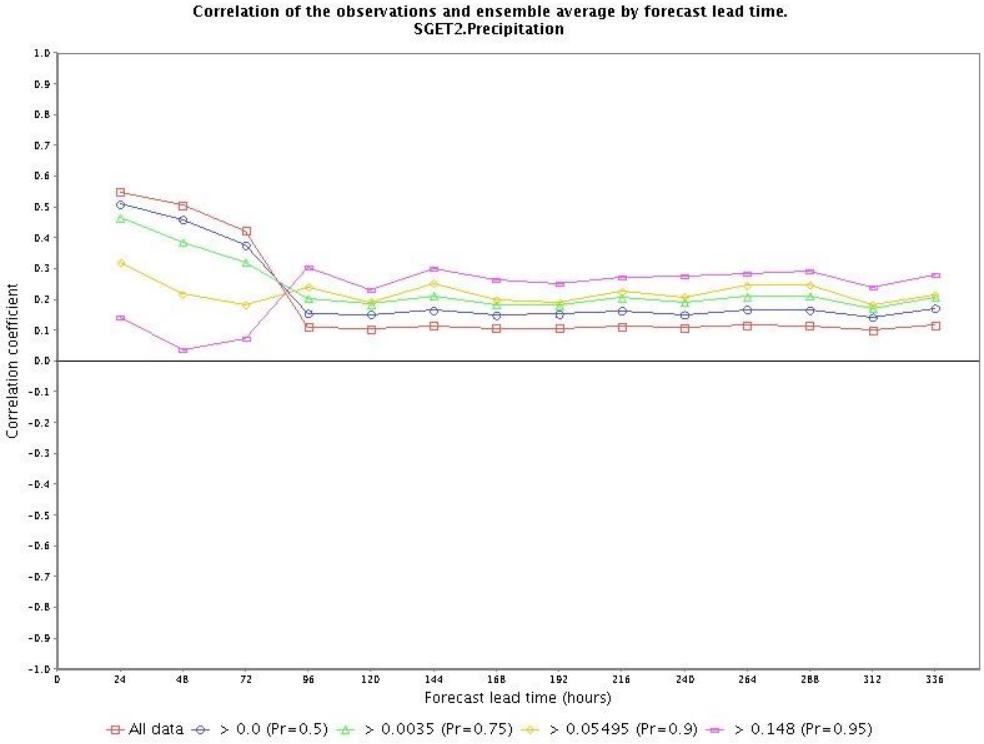

Figure 4.10. Correlation of mean of Day 1-3 EQPF and observed precipitation for SGET2

Figure 4.11 shows the correlation between the mean of raw ESP ensembles forced by Day 1 EQPF and observed streamflow for BRPT2. The correlation decreases gradually after the first day, as expected. Figure 4.12 shows the correlation between the mean of raw ESP ensembles forced by Day 1-3 EQPF and observed streamflow. The large increase in correlation for Days 2 and 3 represents the marginal value in the ensemble mean sense of using Day 2-3 QPF in streamflow forecasting. Figure 4.13 is the same as Figure 4.12, but for the post-processed streamflow ensembles forced by Day 1 EQPF. Note that Figure 4.13 does not show any marked increase in correlation over Figure 4.11. This may be due to the fact that correlation measures strength of association only in the mean of the ensemble instead of the whole ensemble. Note that the purpose of EnsPost is to improve reliability (i.e. probabilistic unbiasedness) of the ensembles, rather than to increase correlation in the ensemble mean sense. Below, the ensemble verification results are presented which assess the improvement in skill due to the use of post-processing by examining reliability and discrimination. Figure 4.14 is the same as Figure 4.13 but for the post-processed streamflow ensembles forced by Day 1-3 EQPF. Compared to Figure 4.13, Figure 4.14 shows an increase in correlation, which represents the value in the ensemble mean sense afforded by the use of longer-lead QPF.

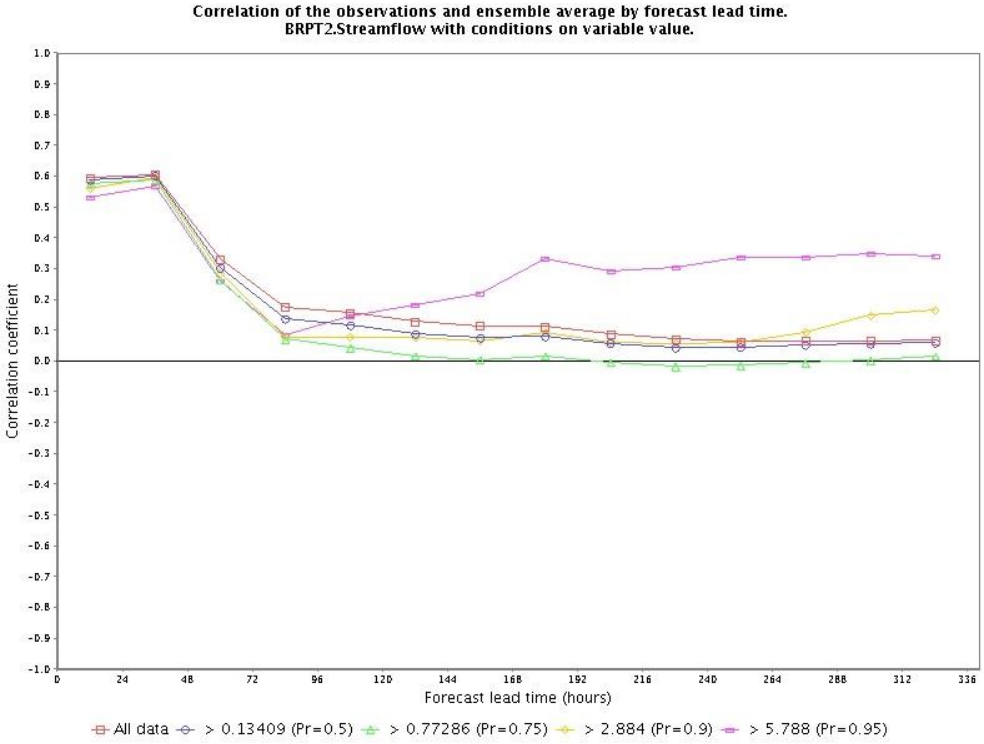

Figure 4.11. Corr. of mean of Day 1 EQPF-forced ESP ensembles and obs. flow (BRPT2)

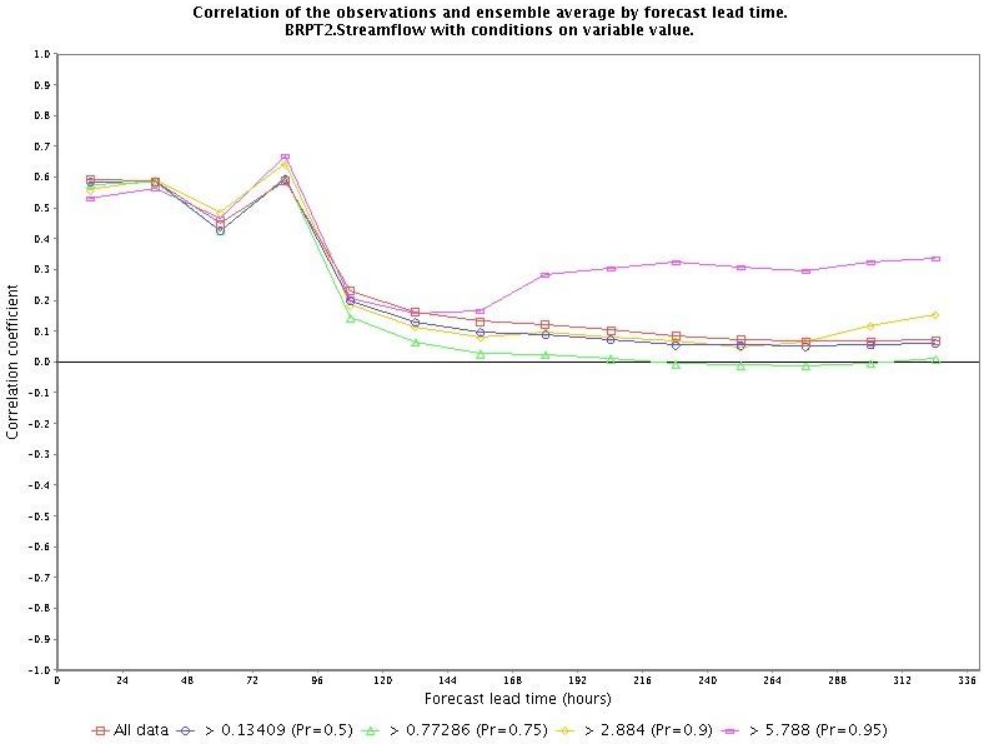

Figure 4.12. Corr. of mean of Day 1-3 EQPF-forced ESP ensembles and obs. flow (BRPT2)

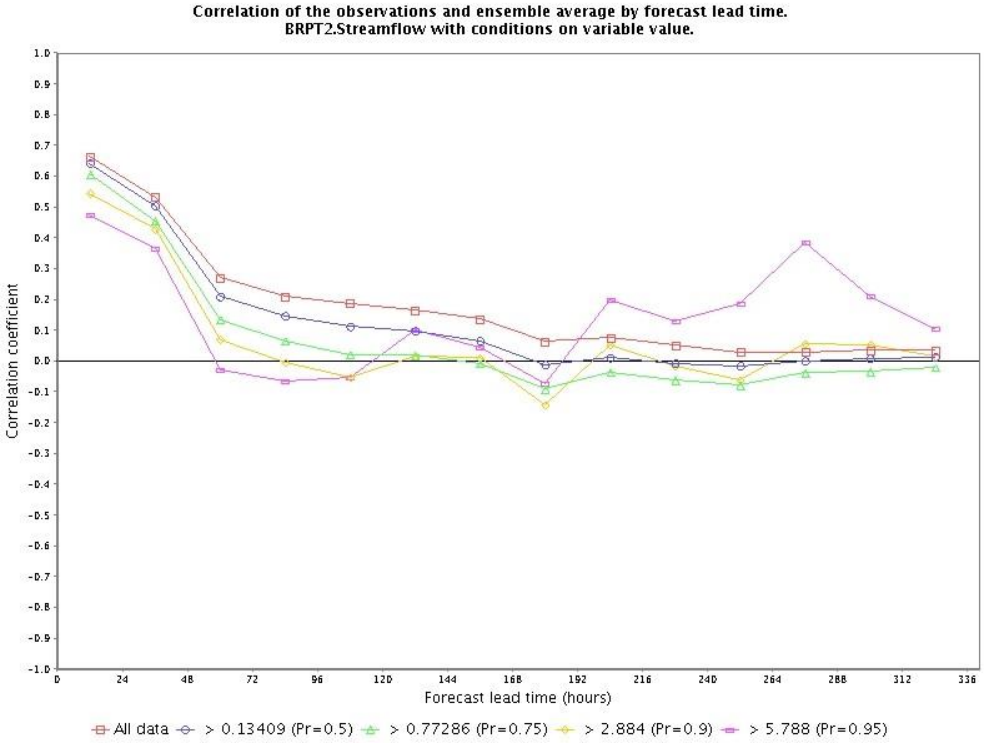

Figure 4.13. Corr. of mean of Day 1 EQPF-forced PP'ed ensembles and obs. Flow (BRPT2)

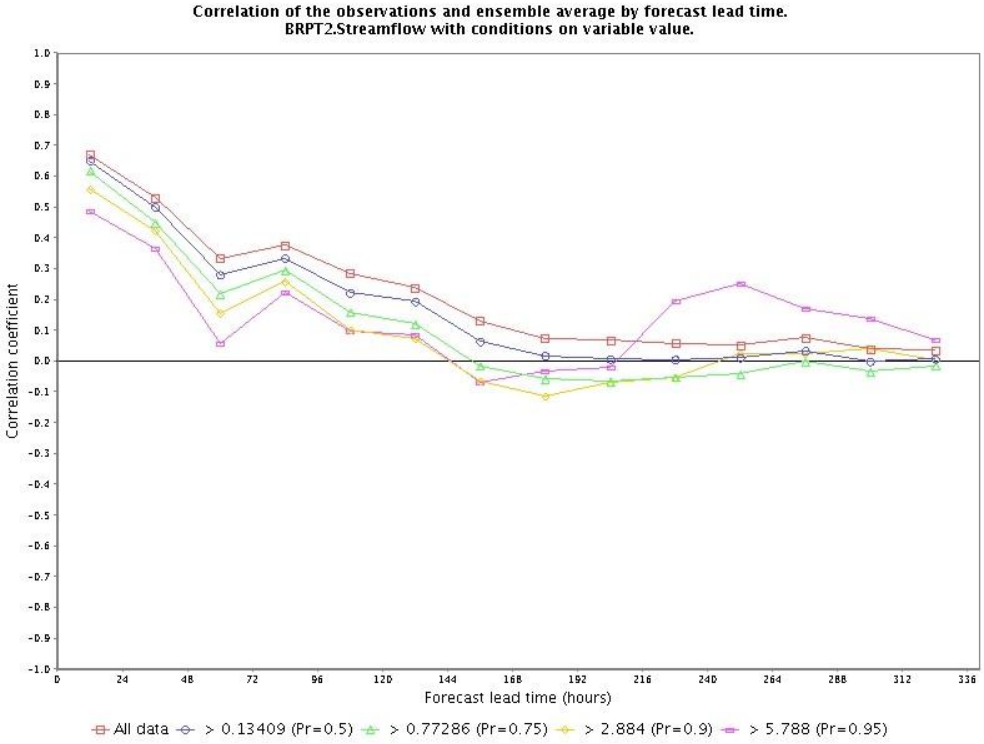

Figure 4.14. Corr. of mean of Day 1-3 EQPF-forced PP'ed ensembles and obs. Flow (BRPT2)

DCJT2 follows a pattern similar to BRPT2, where the ESP ensembles forced by Day 1-3 EQPF show higher correlation (Figure 4.16) for Days 2 and 3 compared to those forced by Day 1 EQPF (Figure 4.15). The post-processed ensembles forced by Day 1 EQPF (Figure 4.17) do not show higher correlation compared to the corresponding raw ESP ensembles (Figure 4.15) (see the explanation above for BRPT2). But comparison of the post-processed ensembles forced by Day 1 EQPF (Figure 4.17) and Day 1-3 EQPF (Figure 4.18) shows increase in correlation in the latter due to using longer-lead QPF.

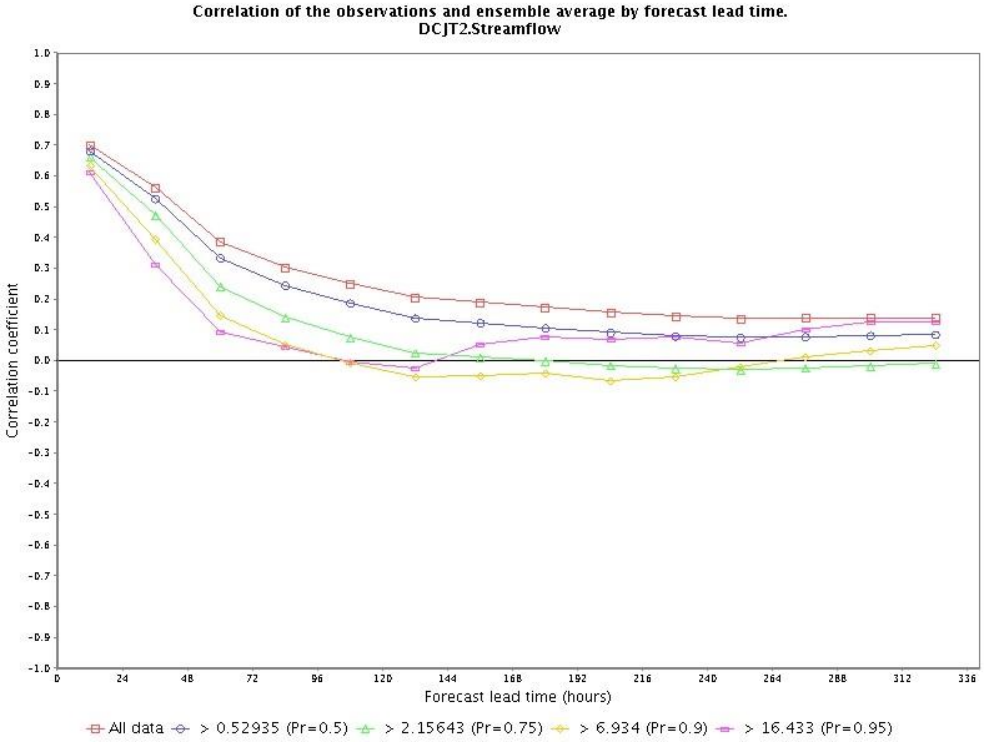

Figure 4.15. Corr. of mean of Day 1 EQPF-forced ESP ensembles and obs. Flow (DCJT2)

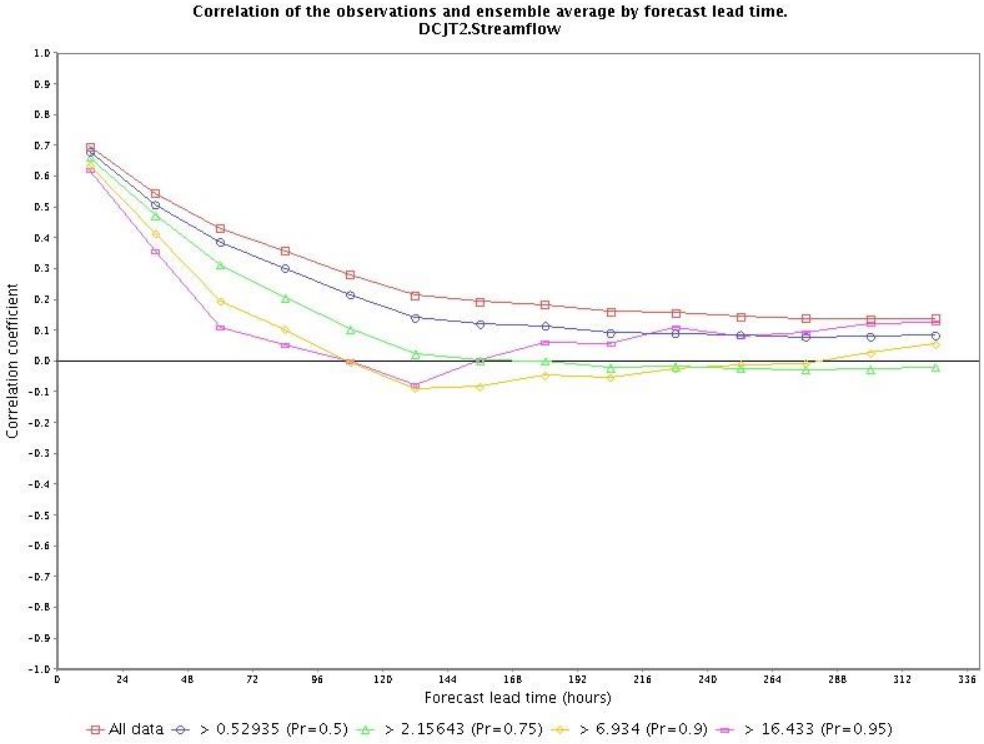

Figure 4.16. Corr. of mean of Day 1-3 EQPF-forced ESP ensembles and obs. Flow (DCJT2)

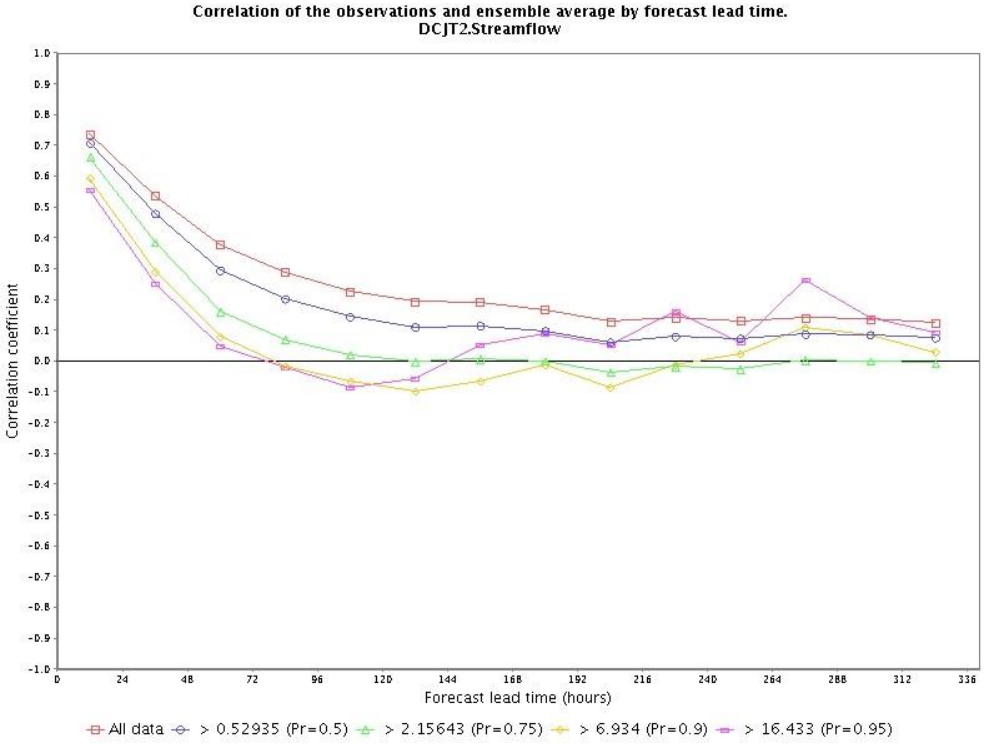

Figure 4.17. Corr. of mean of Day 1 EQPF-forced PP'ed ensembles and obs. Flow (DCJT2)

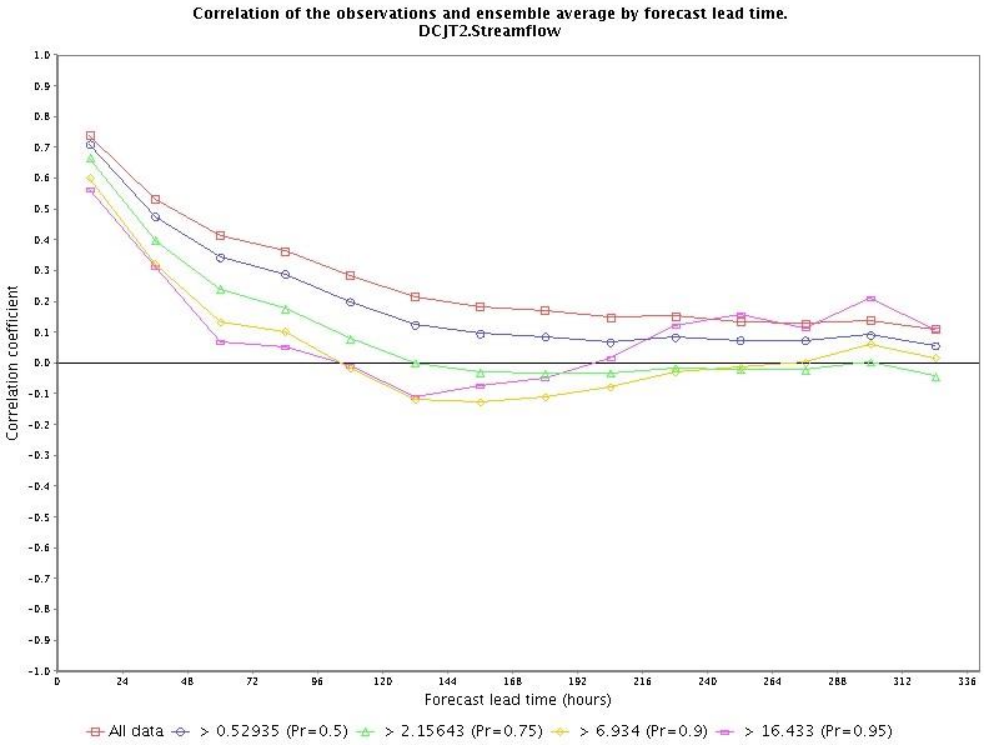

Figure 4.18. Corr. of mean of Day 1-3 EQPF-forced PP'ed ensembles and obs. Flow (DCJT2)

GLLT2 results are similar to BRPT2 and DCJT2 as well. The Day 1-3 EQPF-forced ESP ensembles (Figure 4.20) show higher correlation than the Day 1 EQPF-forced ESP ensembles (Figure 4.19). Comparing the post-processed ensembles forced by the Day 1 QPF (Figure 4.21) and those by the Day 1-3 QPF (Figure 4.22), one may see the increase in correlation due to the use of longer-lead QPF in streamflow forecasting. Overall, the correlation in ensemble mean of forecast streamflow for GLLT2 is slightly smaller than BRPT2 and DCJT2 in the first 1-3 days. This reduced correlation for GLLT2 may be due to the smaller size of the basin; one may expect smaller skill in QPF and shorter basin memory both of which contribute to reduced predictability in streamflow.

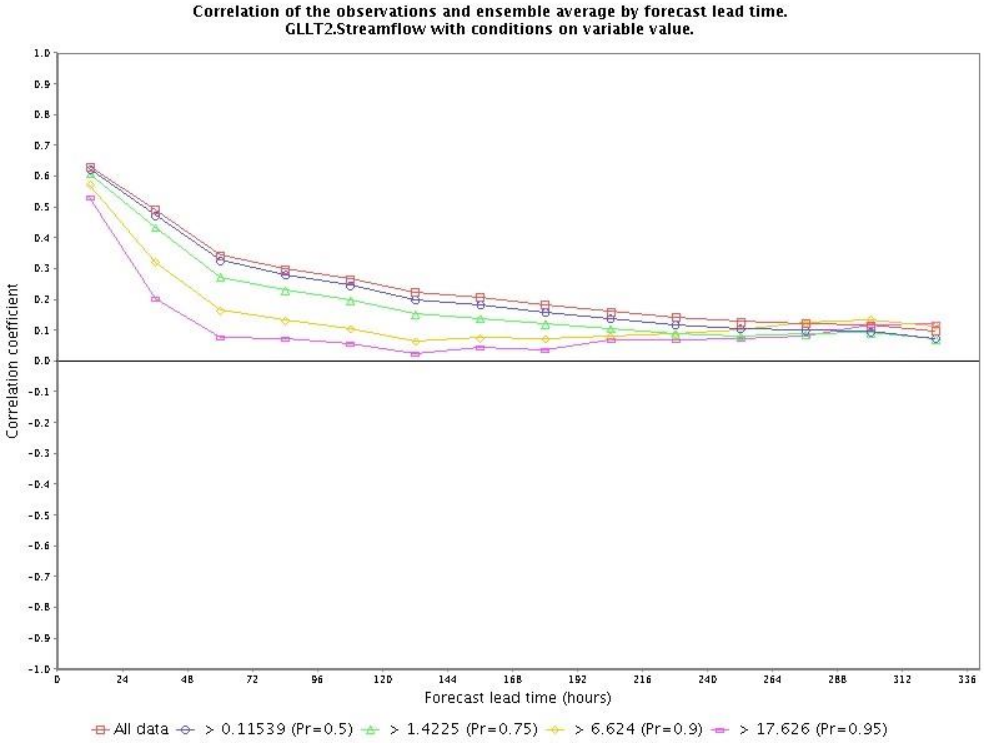

Figure 4.19. Corr. of mean of Day 1 EQPF-forced ESP ensembles and obs. Flow (GLLT2)

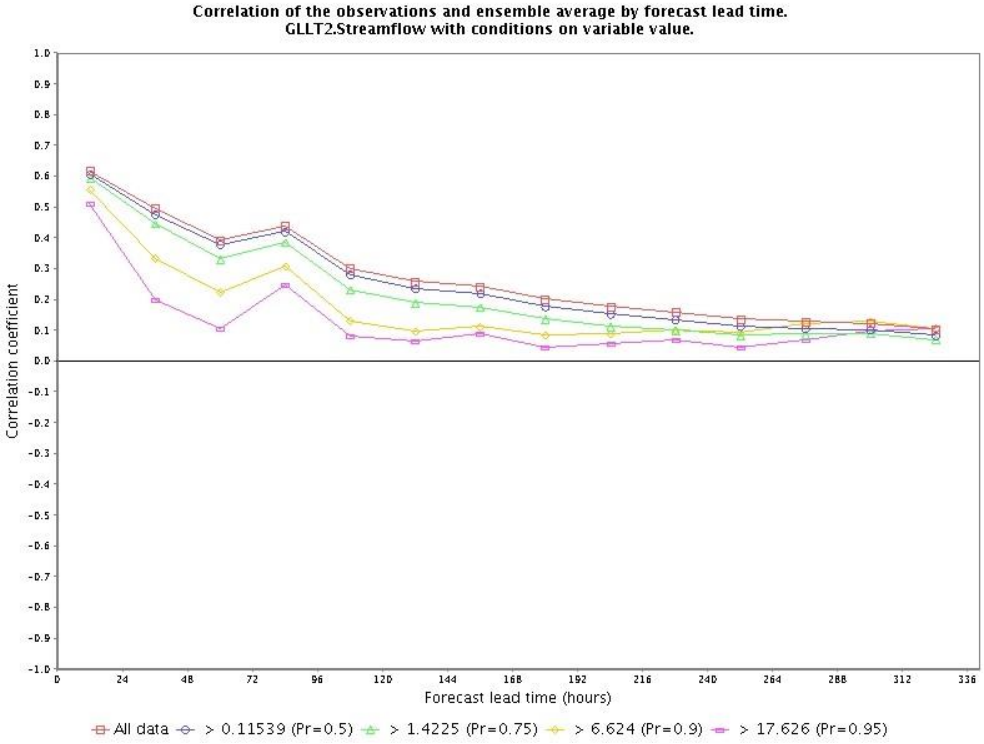

Figure 4.20. Corr. of mean of Day 1-3 EQPF-forced ESP ensembles and obs. Flow (GLLT2)

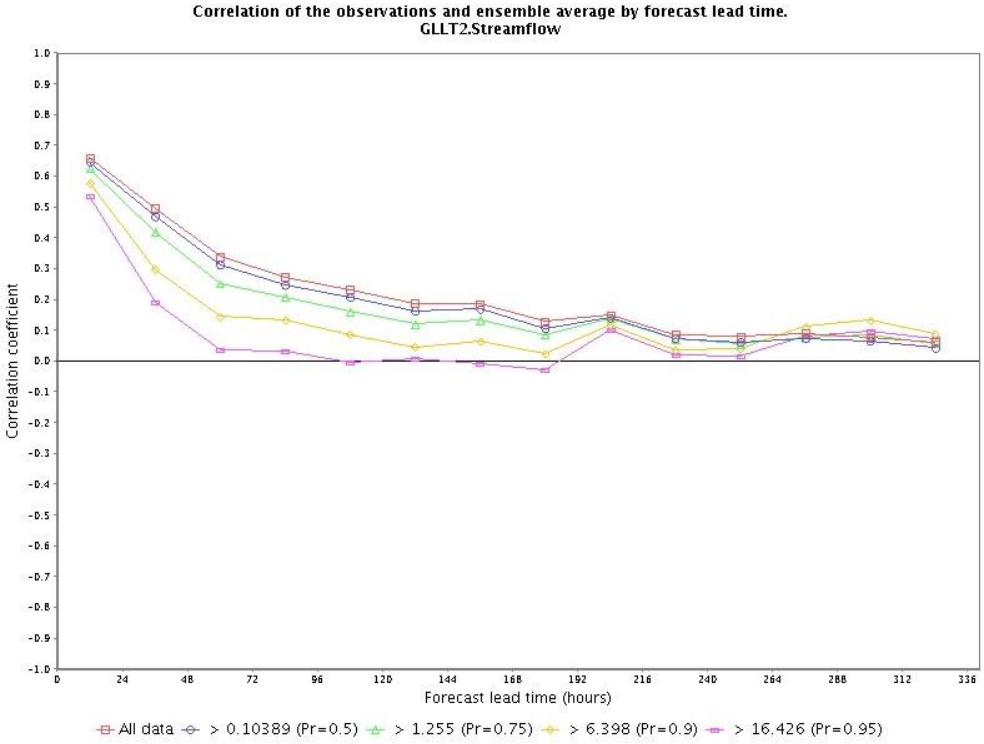

Figure 4.21. Corr. of mean of Day 1 EQPF-forced PP'ed ensembles and obs. Flow (GLLT2)

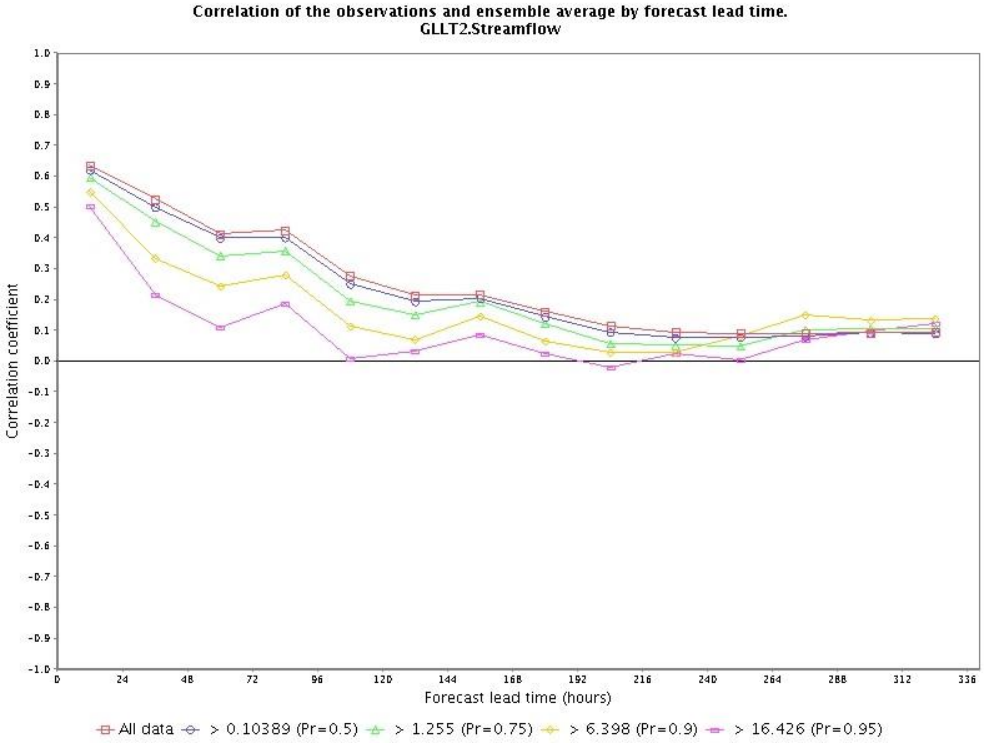

Figure 4.22. Corr. of mean of Day 1-3 EQPF-forced PP'ed ensembles and obs. Flow (GLLT2)

JAKT2 and SGET2 show decreased correlation compared to the other three basins while the pattern remains the same. For JAKT2, Figures 4.23 and 4.24 show the correlation between the observed flow and the ESP ensemble mean hindcasts forced by Day 1 and Day 1-3 EQPF respectively. Figure 4.25 and 4.26 show the correlation coefficients of the mean of corresponding post-processed ensembles and observed flow. Similarly for SGET2, Figures 4.27 and 4.28 show the correlation coefficients of the mean of ESP ensembles forced by Day 1 and Day 1-3 EQPF respectively. And Figures 4.29 and 4.30 show the correlation coefficients of the mean of the corresponding post-processes ensembles. In all the basins, one may see increased correlation for Days 2 and 3 due to the use of longer-lead QPF and EnsPost.

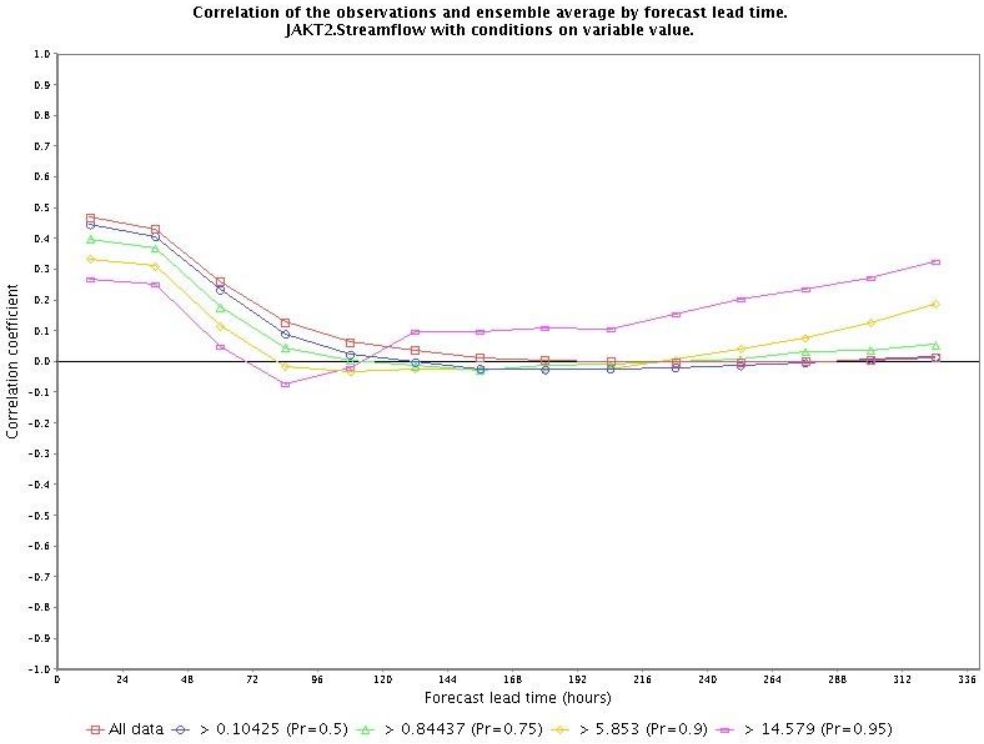

Figure 4.23. Corr. of mean of Day 1 EQPF-forced ESP ensembles and obs. Flow (JAKT2)

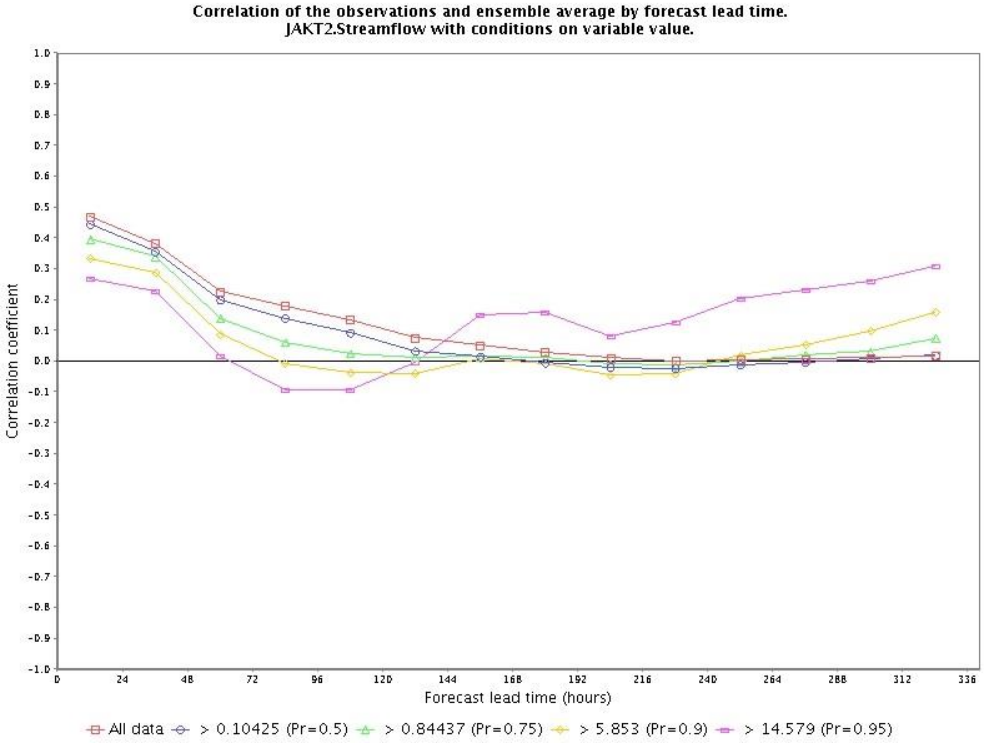

Figure 4.24. Corr. of mean of Day 1-3 EQPF-forced ESP ensembles and obs. Flow (JAKT2)

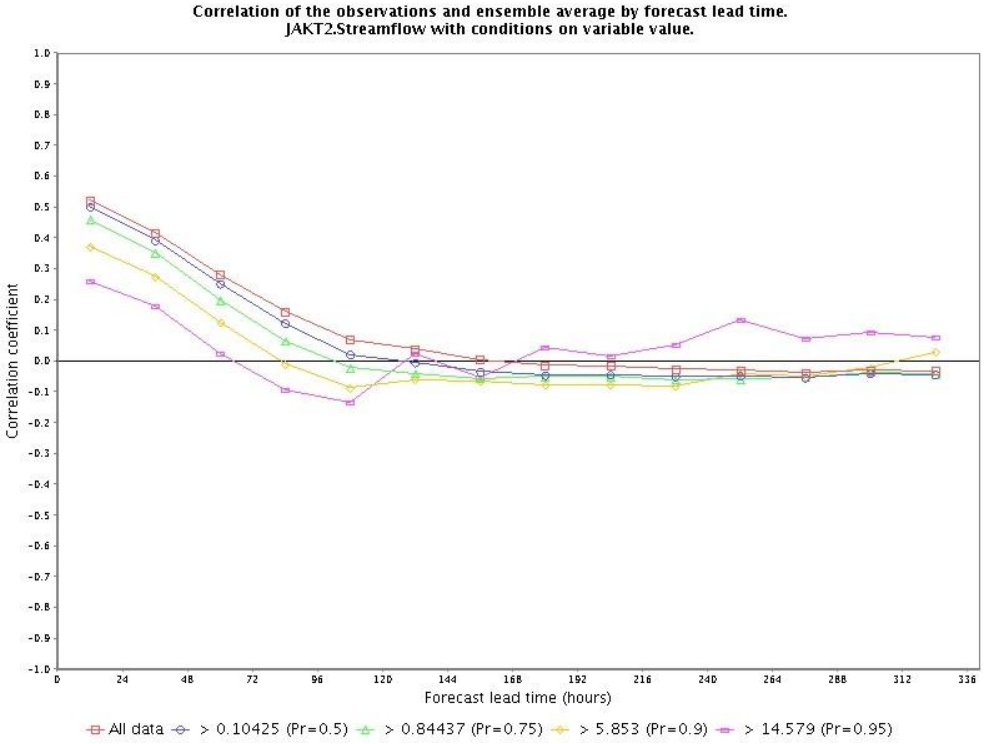

Figure 4.25. Corr. of mean of Day 1 EQPF-forced PP'ed ensembles and obs. Flow (JAKT2)

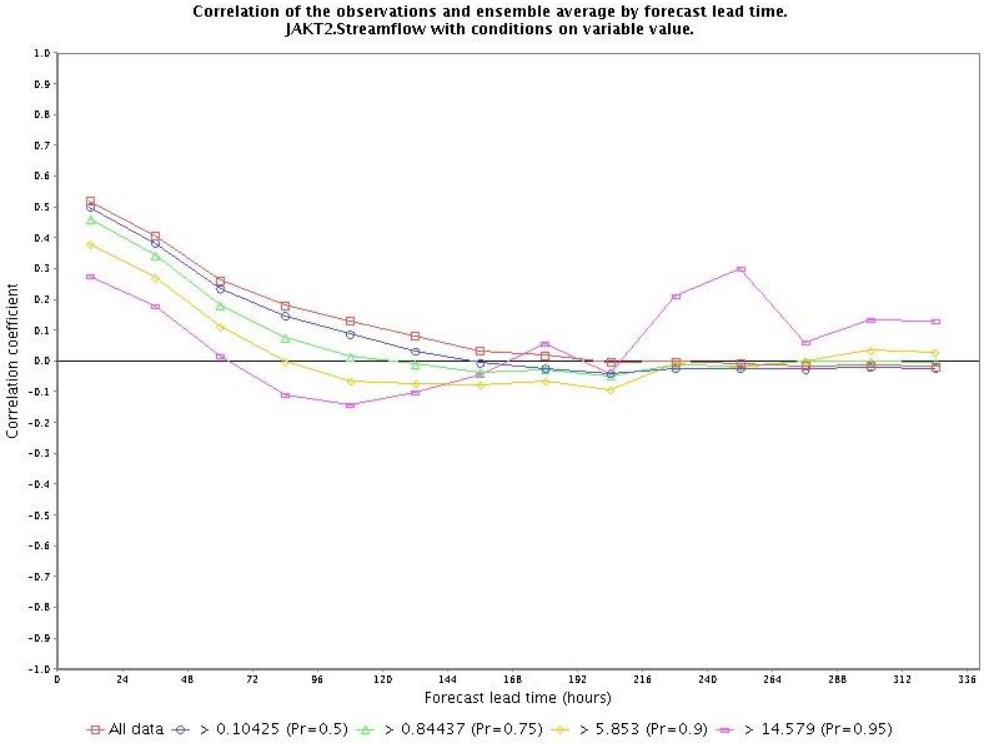

Figure 4.26. Corr. of mean of Day 1-3 EQPF-forced PP'ed ensembles and obs. Flow (JAKT2)

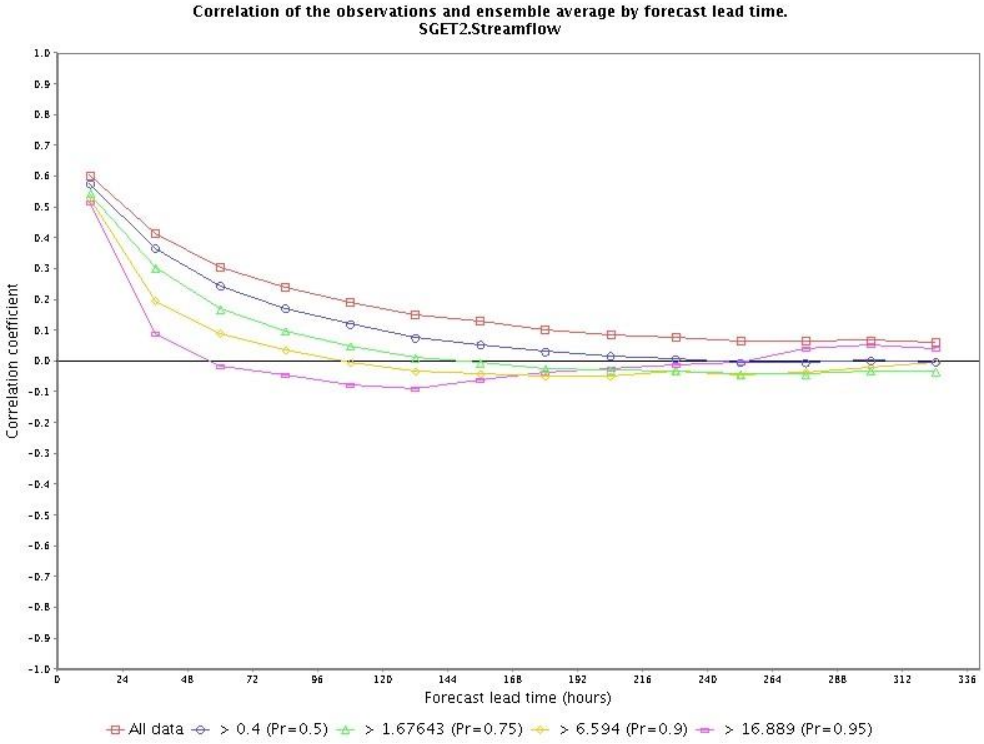

Figure 4.27. Corr. of mean of Day 1 EQPF-forced ESP ensembles and obs. Flow (SGET2)

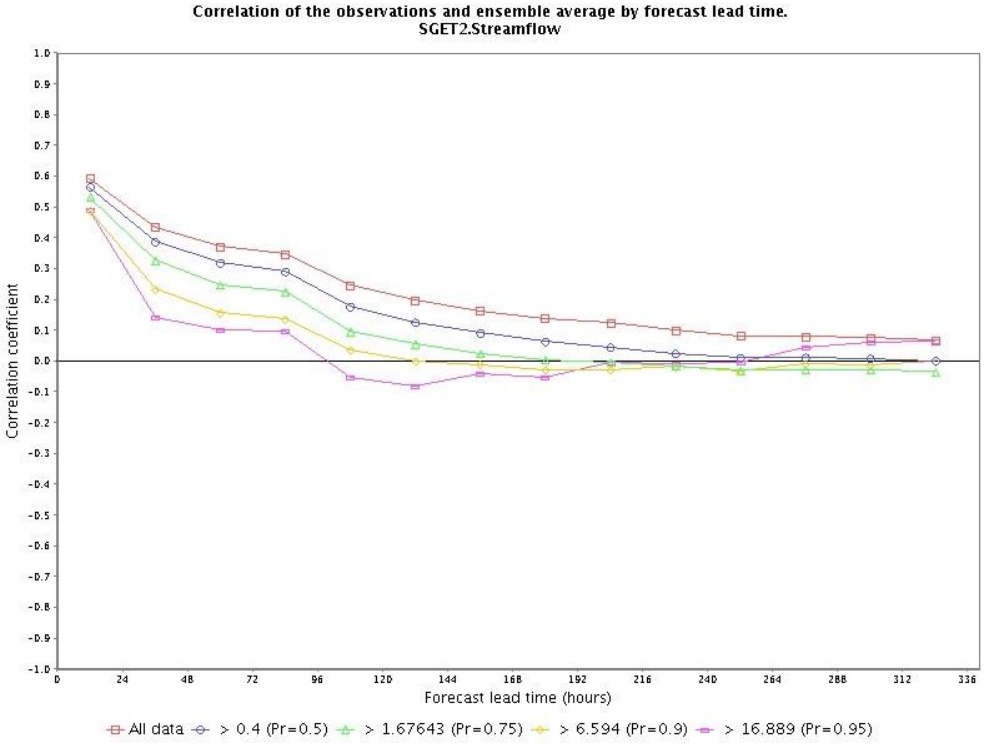

Figure 4.28. Corr. of mean of Day 1-3 EQPF-forced ESP ensembles and obs. Flow (SGET2)

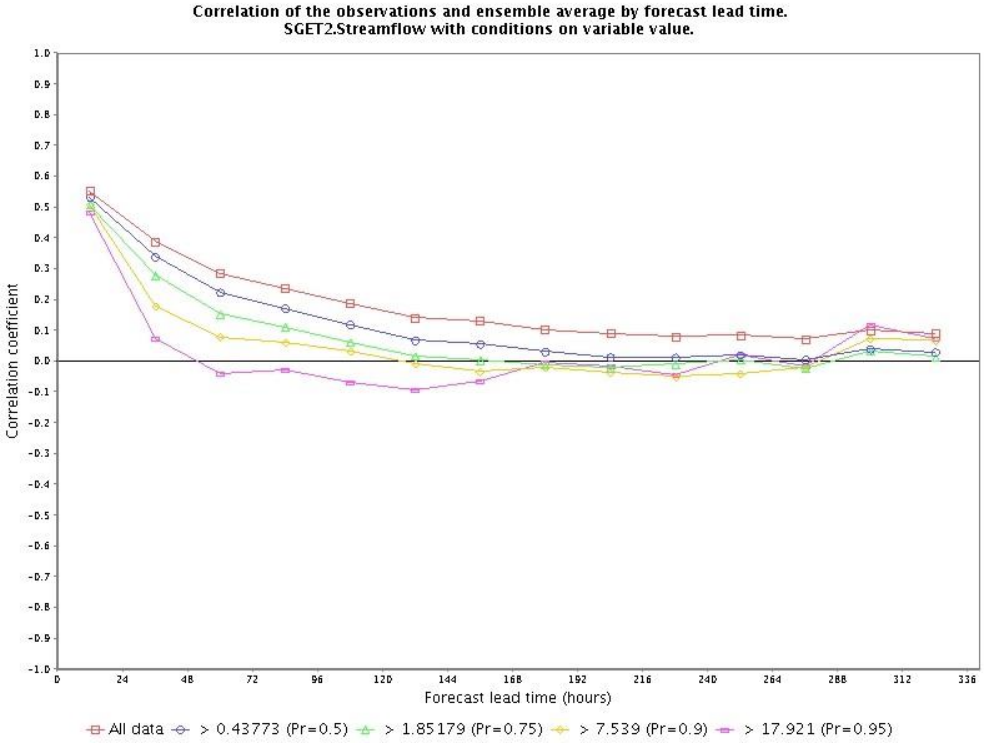

Figure 4.29. Corr. of mean of Day 1 EQPF-forced PP'ed ensembles and obs. Flow (SGET2)

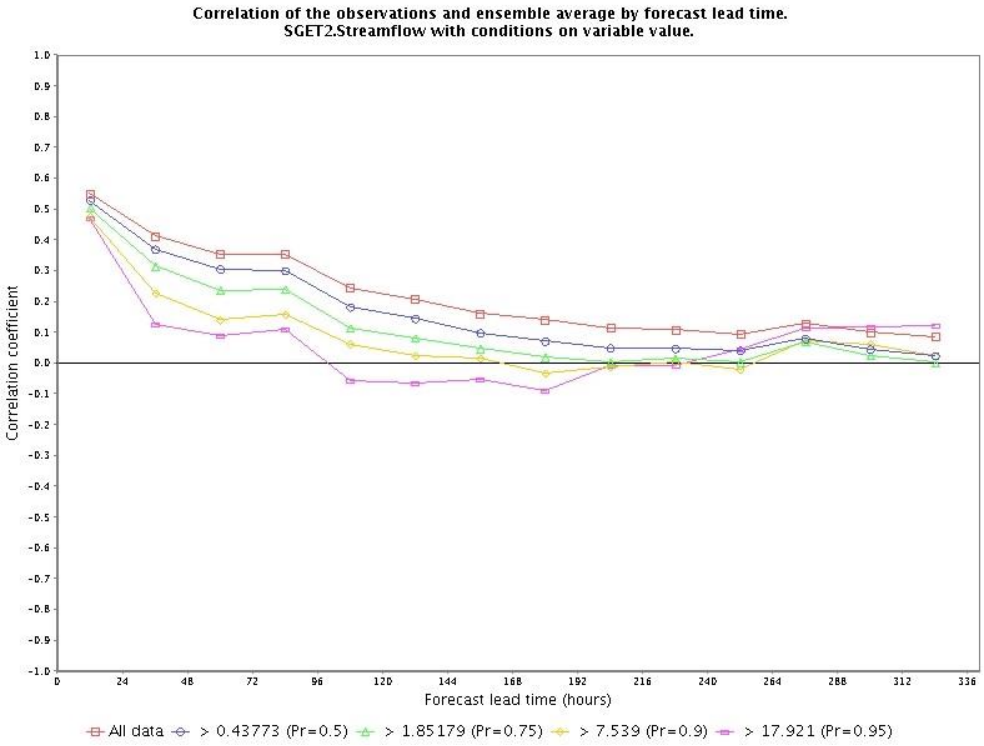

Figure 4.30. Corr. of mean of Day 1-3 EQPF-forced PP'ed ensembles and obs. Flow (SGET2)

Single-value verification metrics such as correlation coefficient measures skill in the single-valued sense only. To assess the skill in the ensemble sense, ensemble verification is necessary ([5], [8]). Different ensemble verification measures evaluate different attributes such as reliability, resolution, discrimination and sharpness. The verification tool used in this work, EVS, includes many verification measures ([5], [33]). Of them, only the reliability and relative operating characteristic (ROC) results are presented below as they are complementary measures that assess disparate attributes.

4.2. Reliability Diagram

Probabilistic forecasts that are unbiased against the frequency of the verifying observations are termed reliable. This probabilistic unbiasedness of forecasts is a requisite for risk-based decision making. Reliability diagram, which plots forecast probability against the frequency of verifying observations, can detect conditional biases in probability in ensemble forecasts based on the deviation of the reliability curves from the diagonal, which represents perfect reliability. The reliability curve is a representation of the bivariate distribution between the forecast and the

observed conditioned on the forecast whereas ROC is that conditioned on the observation. As such, reliability diagram and ROC are independent attributes and insensitive to each other.

Figures 4.31 and 4.32 show the reliability diagrams for Day 1 and Day 1-3 EQPF for BRPT2. Figures 4.32, 4.33 and 4.34 show the reliability diagrams for Days 1 through 3, respectively. Figures 4.35, 4.36, 4.37 and 4.38 show the reliability diagrams of ESP ensembles at increasing lead times forced by Day 1 and Day 1-3 EQPF. Note in these figures the lack of reliability in raw ESP ensembles. That the forcing precipitation ensembles are reasonably reliable, but the ESP ensembles are not, is an indication that a large hydrologic uncertainty exists. Figures 4.39, 4.40, 4.41 and 4.42 show the corresponding post-processed ensembles forced by Day 1 and Day 1-3 EQPF. Note that post-processing greatly improves reliability of these forecasts, an indication that EnsPost is largely successful in modeling hydrologic uncertainty for this basin. For comprehensive results, we present below only those for BRPT2. Generally speaking, the results for the other basins are qualitatively similar. There are, however, notable differences, in particular, with JAKT2 and SGET2 due to reduced quality in hydrologic simulation which are described below as well.

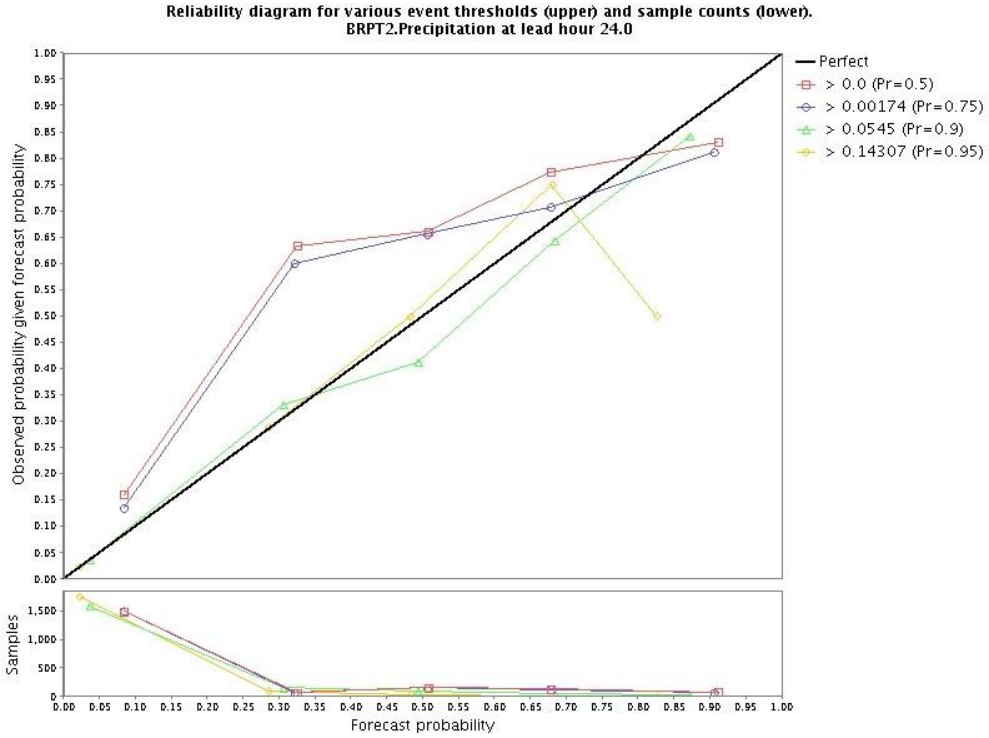

Figure 4.31. Reliability Diagram of Day 1 EQPF at Day 1 for BRPT2

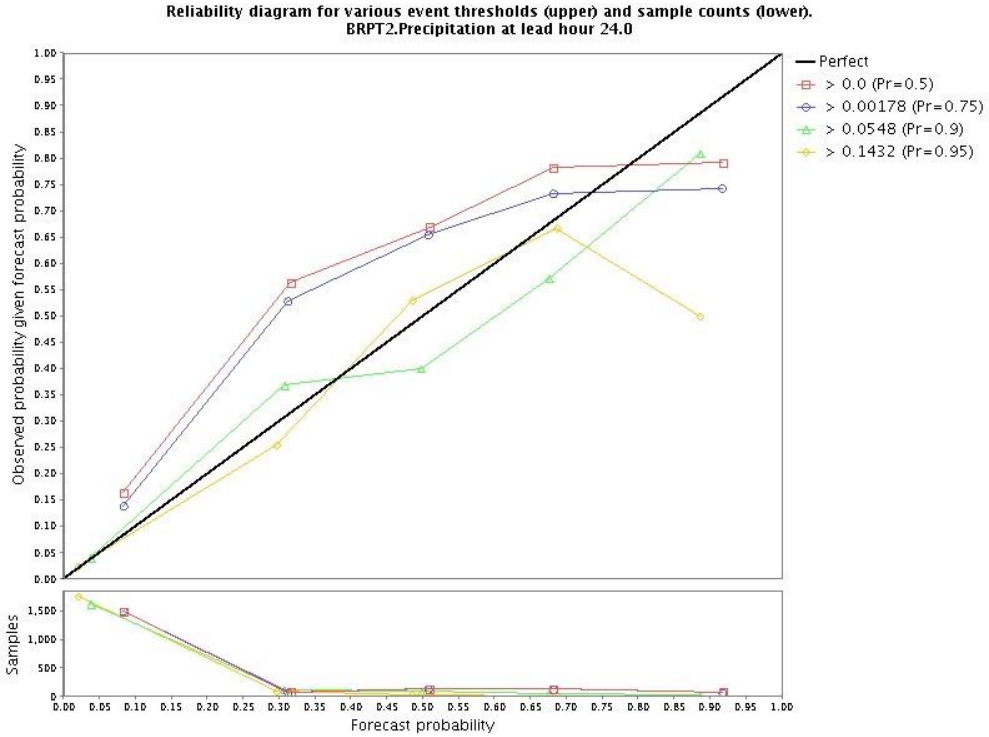

Figure 4.32. Reliability Diagram of Day 1-3 EQPF at Day 1 for BRPT2

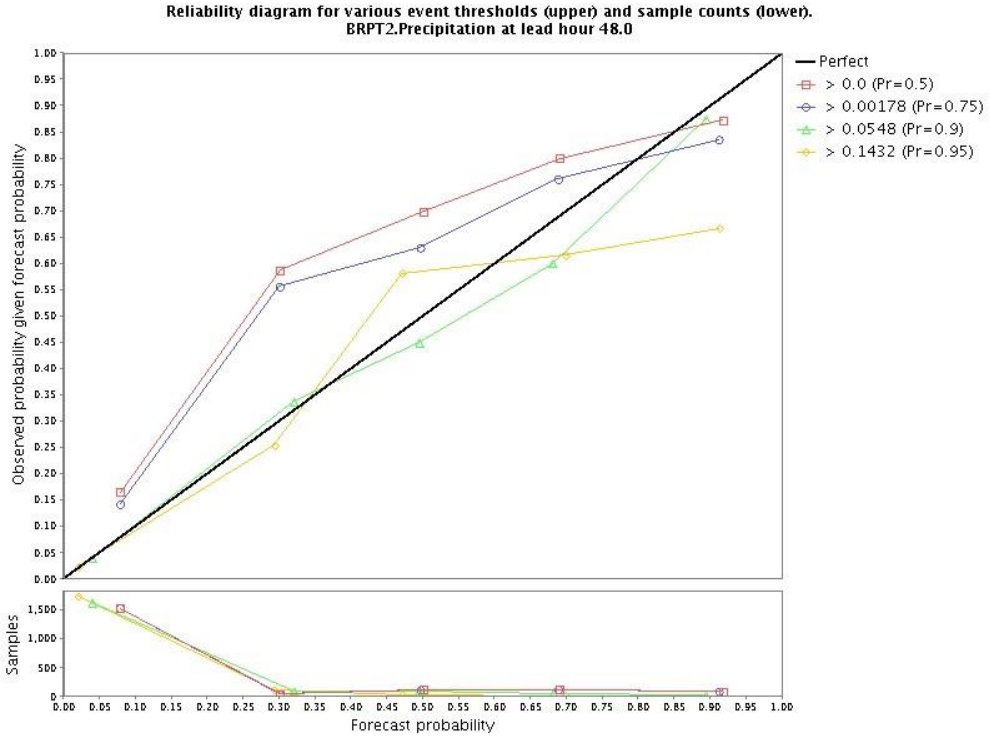

Figure 4.33. Reliability Diagram of Day 1-3 EQPF at Day 2 for BRPT2

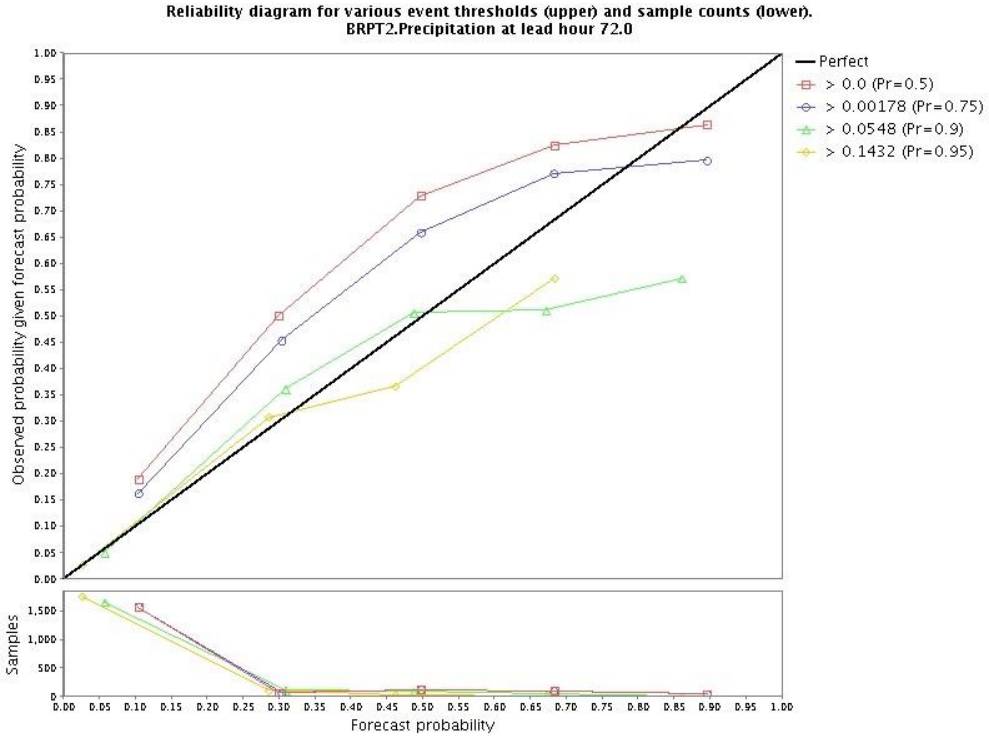

Figure 4.34. Reliability Diagram of Day 1-3 EQPF at Day 3 for BRPT2

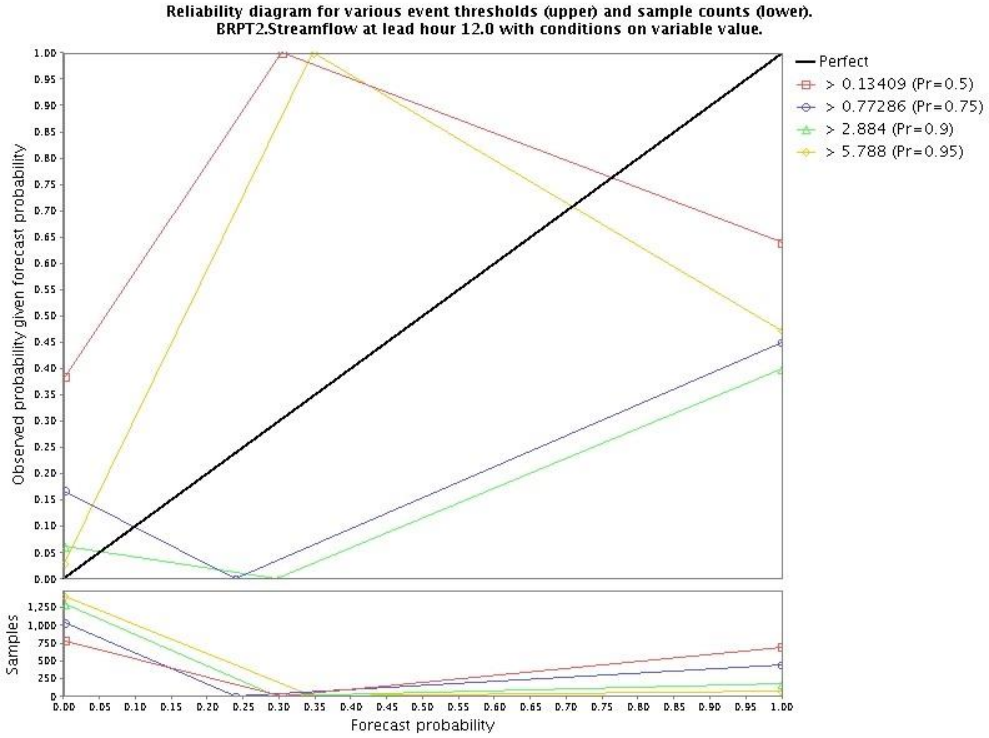

Figure 4.35. Reliability Diagram of Day 1 EQPF-forced ESP ensembles at Day 1 for BRPT2

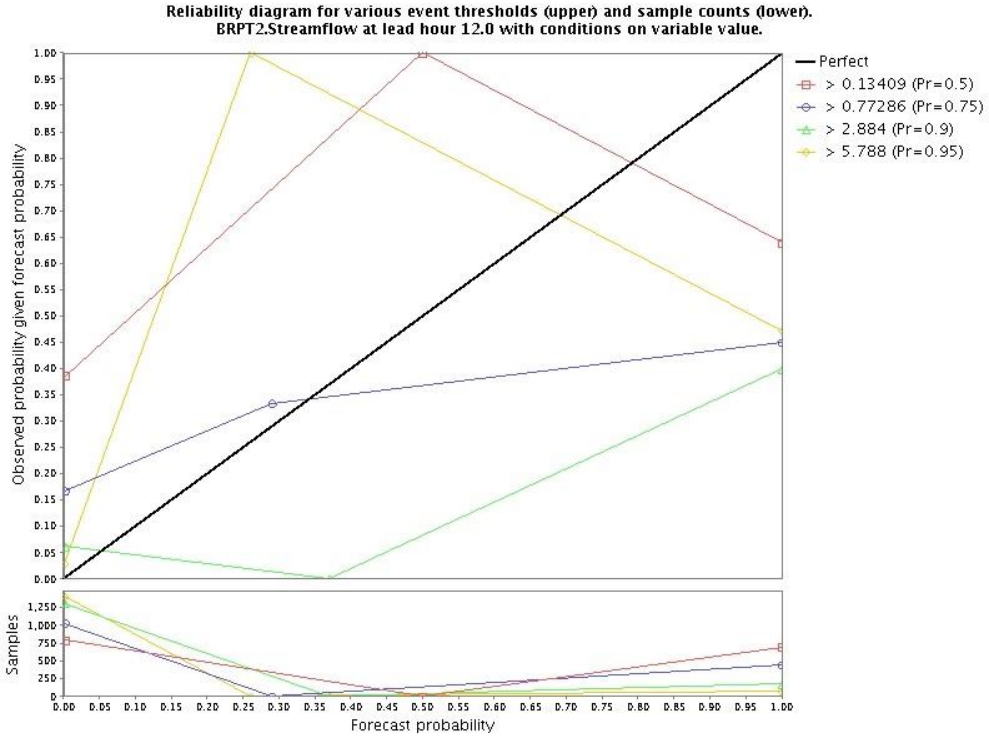

Figure 4.36. Reliability Diagram of Day 1-3 EQPF-forced ESP ensembles at Day 1 for BRPT2

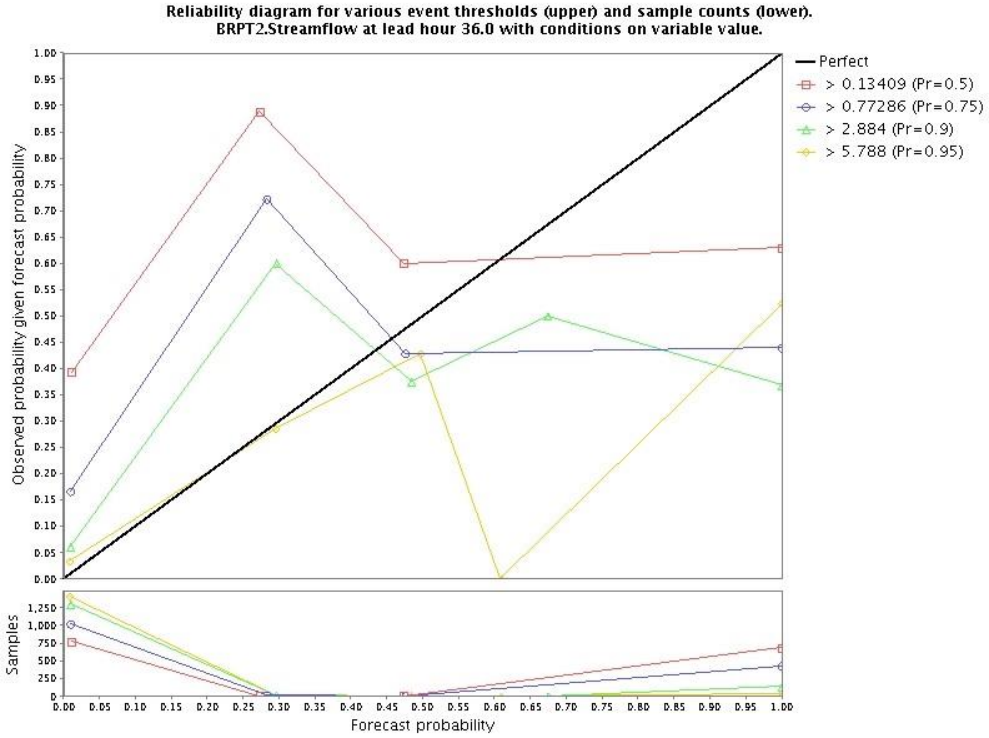

Figure 4.37. Reliability Diagram of Day 1-3 EQPF-forced ESP ensembles at Day 2 for BRPT2

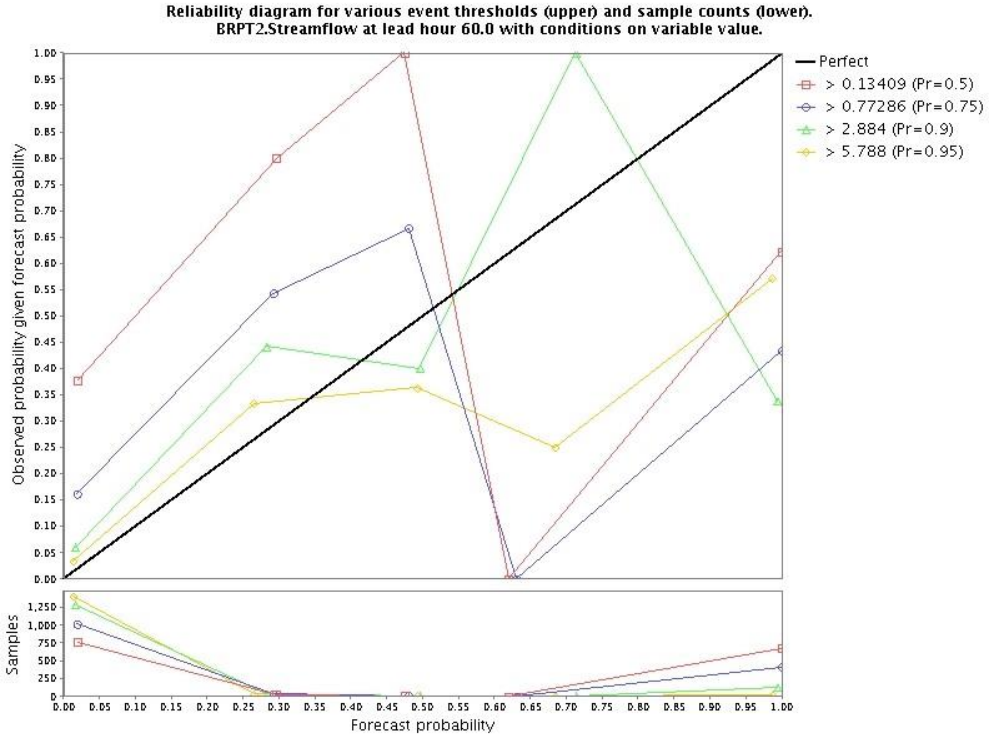

Figure 4.38. Reliability Diagram of Day 1-3 EQPF-forced ESP ensembles at Day 3 for BRPT2

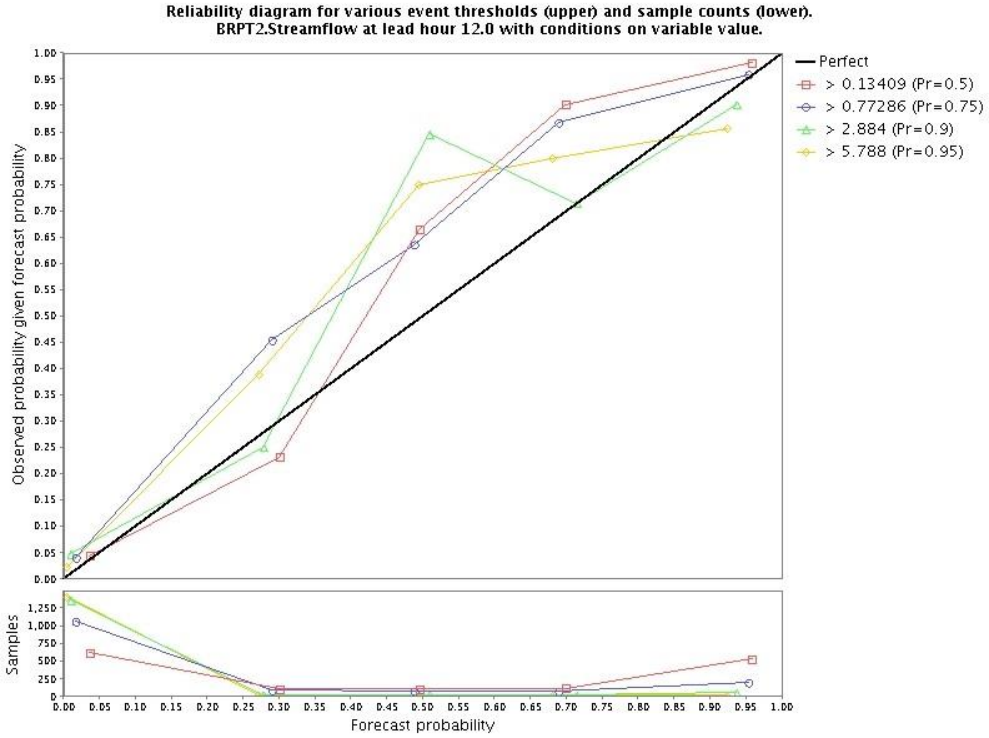

Figure 4.39. Reliability Diagram of Day 1 EQPF-forced PP'ed ensembles at Day 1 for BRPT2

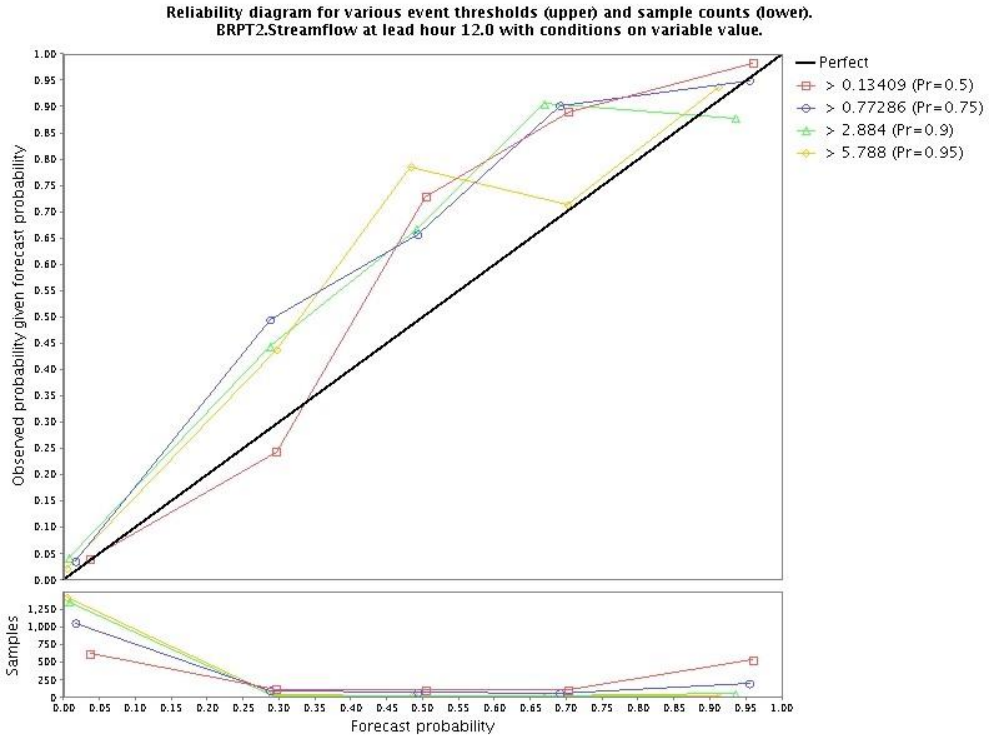

Figure 4.40. Reliability Diagram of Day 1-3 EQPF-forced PP'ed ensembles at Day 1 for BRPT2

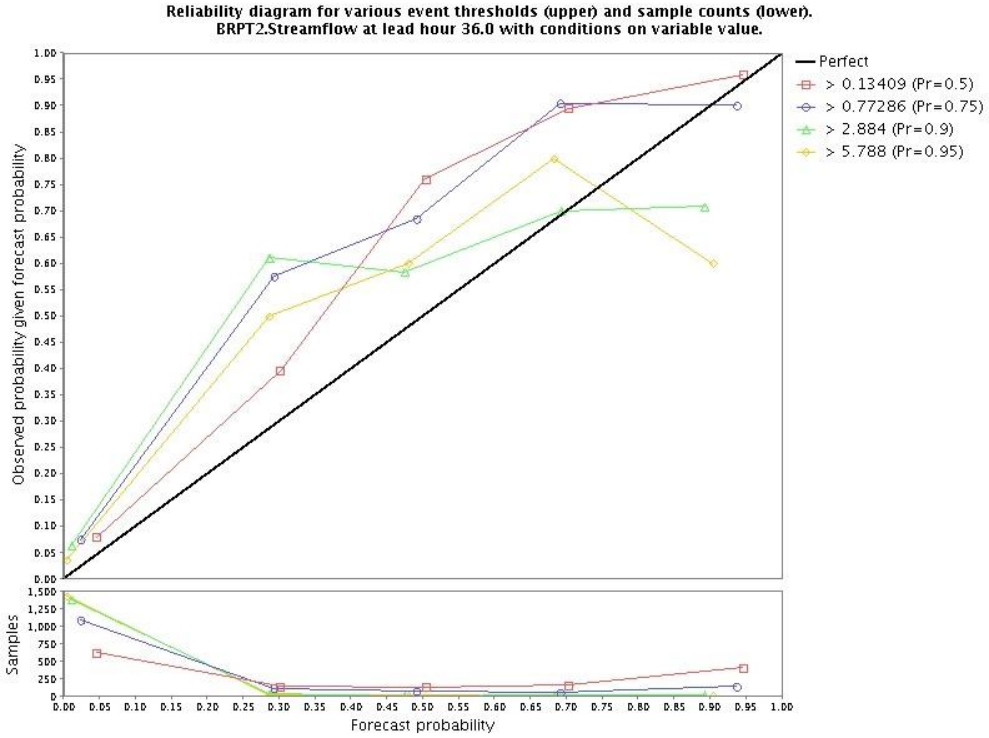

Figure 4.41. Reliability Diagram of Day 1-3 EQPF-forced PP'ed ensembles at Day 2 for BRPT2

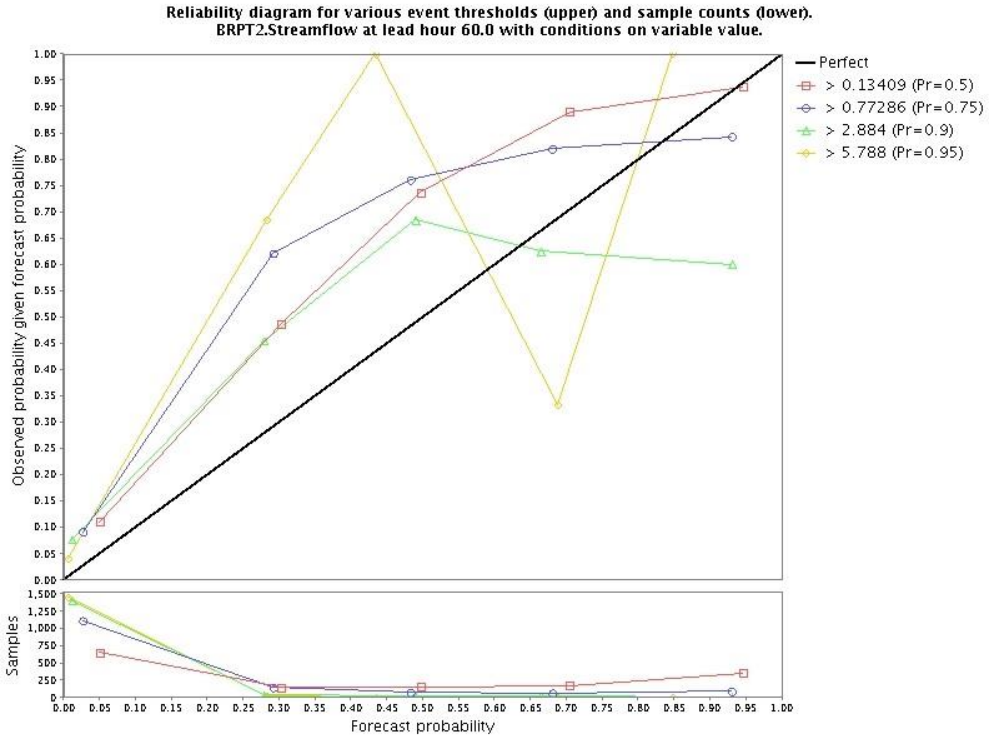

Figure 4.42. Reliability Diagram of Day 1-3 EQPF-forced PP'ed ensembles at Day 3 for BRPT2

4.3 Relative Operating Characteristics (ROC) and ROC Score

ROC measures the ability of a forecast to distinguish between an event and a non-event where an event is defined by the user. The ROC curve plots the Probability of Detection (POD) against the false alarm rate (FAR) at different levels of exceedance probability [17]. The area between the ROC curve and the diagonal line is referred to as the ROC area. The diagonal line (e.g. ROC of climatological forecasts) represents no skill. A perfect forecast would produce an ROC area made of the upper triangle. ROC score is defined as the ROC area multiplied by two. Note that forecasts with perfect and no discriminatory skill would yield a ROC score of unity and zero, respectively.

Figures 4.43 and 4.44 show the ROC scores of Day 1 and Day 1-3 EQPFs for BRPT2, respectively. In Figure 4.43, the ROC score drops to a negligible level after the first day as expected. ROC scores of Day 1-3 EQPF are significant for Days 2 and 3. As noted above, the objective of this work is to utilize this skill in Days 2 and 3 ensemble precipitation forecasts to produce more skillful streamflow forecasts. Figures 4.45 and 4.46 represent the ROC score of the ESP ensembles forced by the Day 1 and Day 1-3 EQPF, respectively. In both the cases, the ROC scores for the highest thresholds (90^{th} and 95^{th} percentiles) are smaller than those for the 50^{th} and 75^{th} percentiles in the first few days. This is due to reduced predictability in higher flows. Comparing the ESP ensembles forced by Day 1-3 EQPF to those by Day 1 EQPF, one may see an increased ROC score in Days 2 and 3 for all thresholds. This shows the marginal improvement in skill in Days 2 and 3 streamflow ensembles due to the use of longer-lead QPF. Figures 4.47 and 4.48 show the ROC score for post-processed ESP ensembles forced by Day 1 and Day 1-3 EQPF, respectively. Note that, for low thresholds, post-processing greatly increases the ROC score. For higher thresholds, however, the gain in ROC score is significantly smaller due to the limited predictability in high flow conditions. Comparing Figure 4.48 to Figure 4.47, one may see the improvement in ROC score due to longer-lead EQPF.

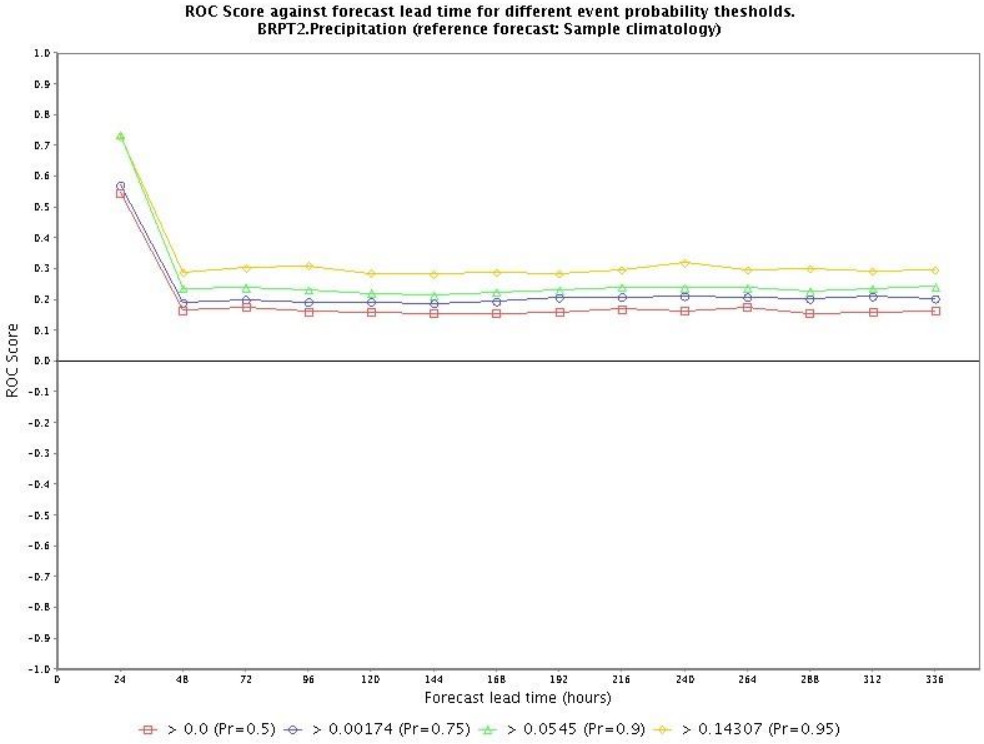

Figure 4.43. ROC Score of Day 1 EQPF for BRPT2

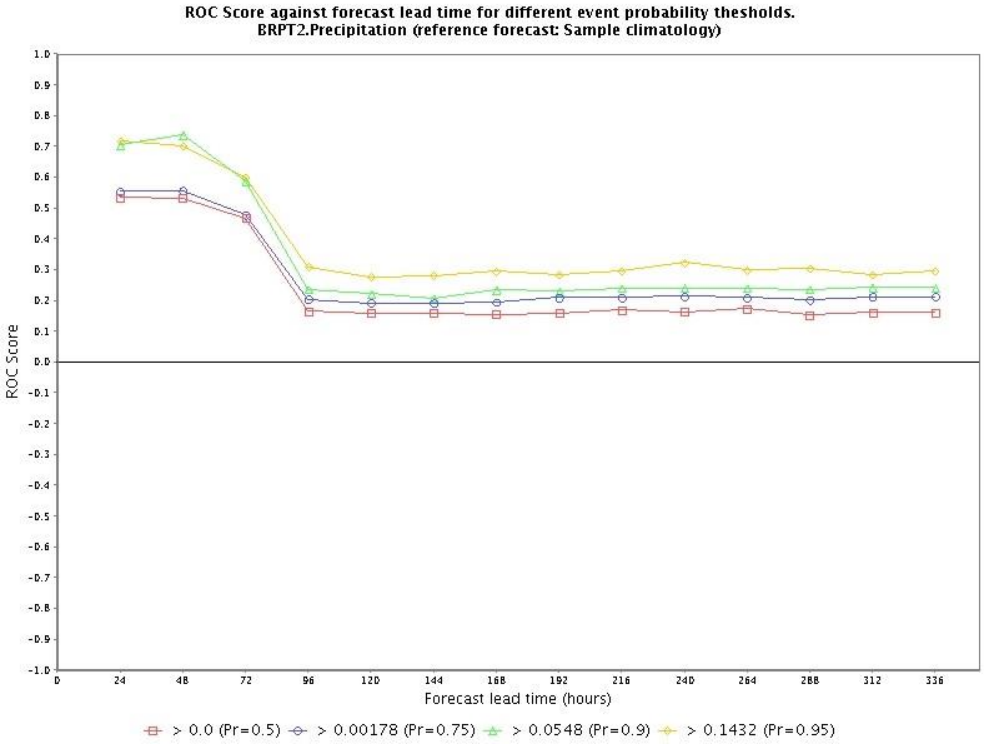

Figure 4.44. ROC Score of Day 1-3 EQPF for BRPT2

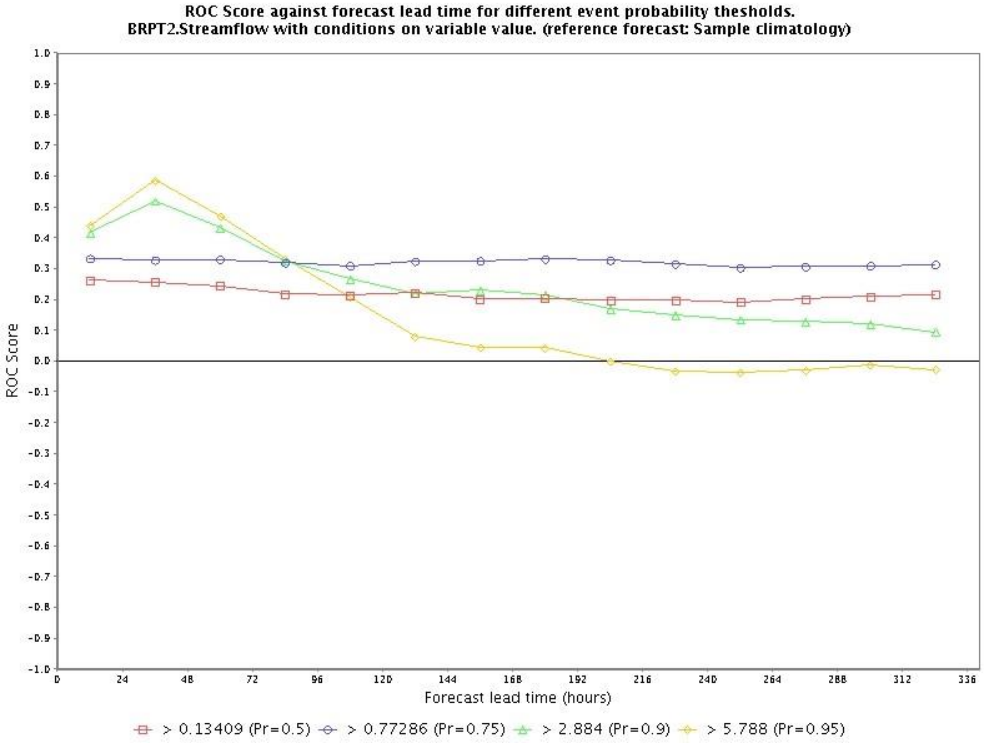

Figure 4.45. ROC Score of Day 1 EQPF-forced ESP ensembles for BRPT2

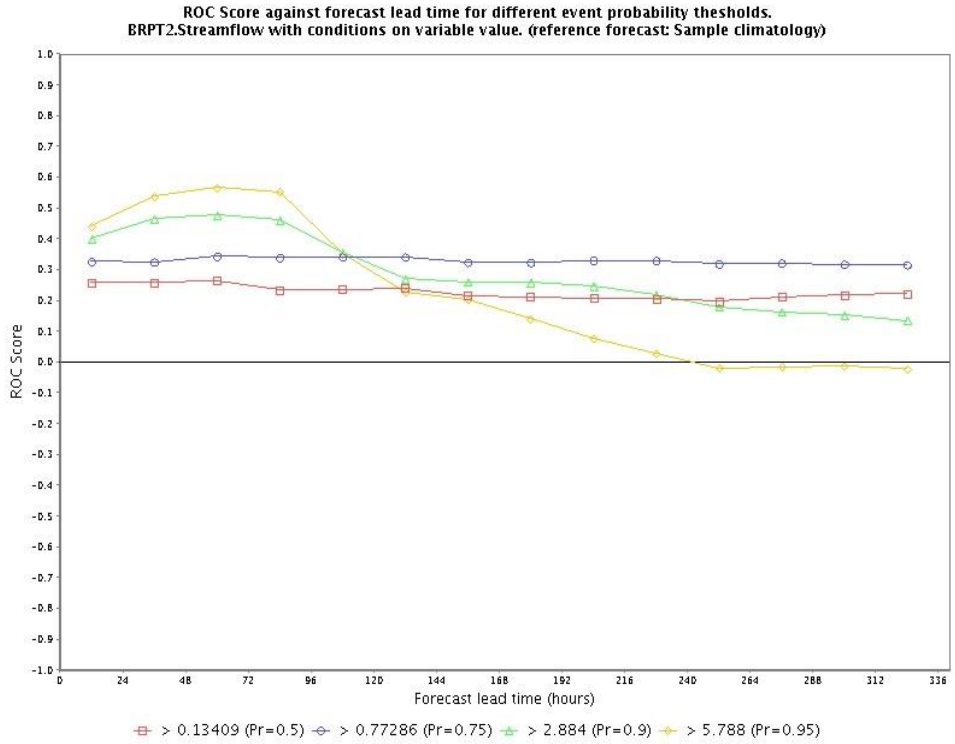

Figure 4.46. ROC Score of Day 1-3 EQPF-forced ESP ensembles for BRPT2

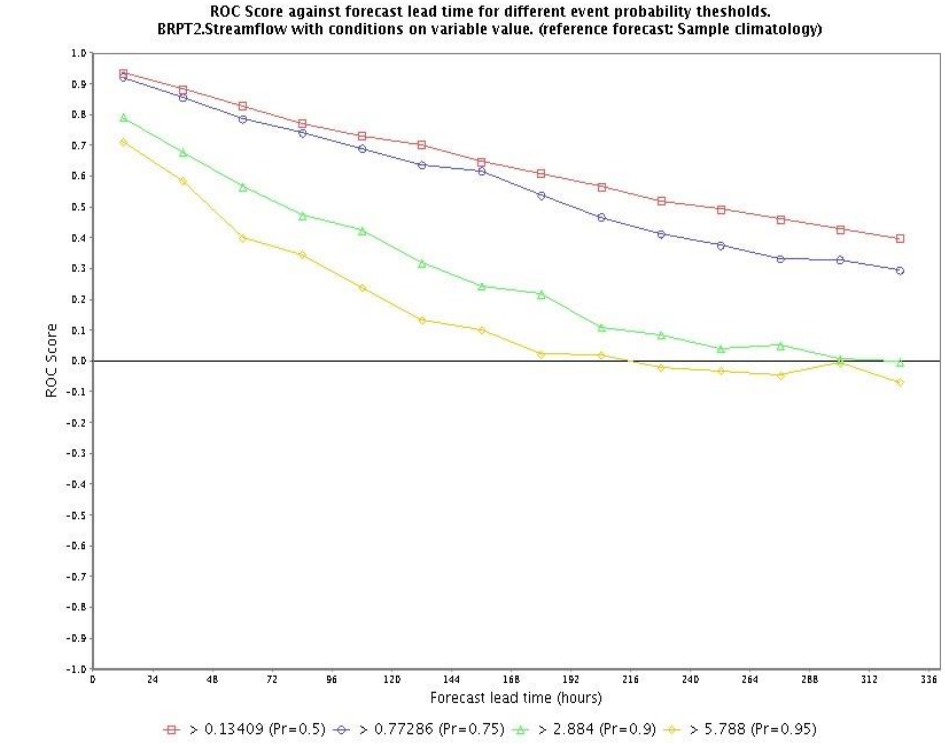

Figure 4.47. ROC Score of Day 1 EQPF-forced PP'ed ensembles for BRPT2

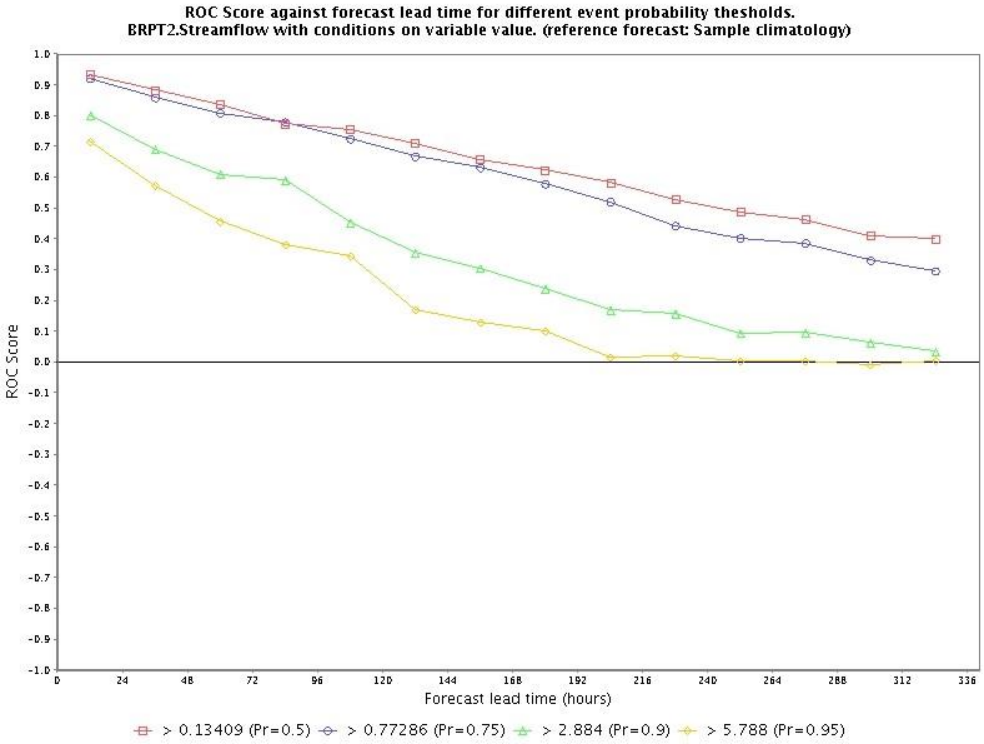

Figure 4.48. ROC Score of Day 1-3 EQPF-forced PP'ed ensembles for BRPT2

In order to compare the improvement (or lack thereof) in the ROC score due to longer-lead QPF and ensemble post-processing more easily, the ROC scores are plotted corresponding to the Hindcasting Experiments 1 through 4 , 1) Day 1 QPF-forced ESP ensembles, 2) Day 1 QPF-forced post-processed ensembles, 3) Day 1-3 QPF-forced ESP ensembles and 4) Day 1-3 QPF-forced post-processed ensembles, for each of the four thresholds (50^{th}, 75^{th}, 90^{th}, and 95^{th} percentiles). Figures 4.49-4.52, 4.53-4.56, 4.57-4.60, 4.61-4.64 and 4.65-4.68 show the ROC scores at 50^{th}, 75^{th}, 90^{th} and 95^{th} percentile thresholds for BRPT2, DCJT2, GLLT2, JAKT2 and SGET2, respectively. For BRPT2, DCJT2 and GLLT2, the improvement in ROC score due to longer-lead QPF and ensemble post-processing is significant. Note that the ROC scores of post-processed ensembles (forced by either Day 1 or Day 1-3 EQPF) decrease as the threshold increases. This is due to reduced predictability in high flow conditions. The increase in ROC scores due to post-processing is most significant at the 50% threshold and the increase becomes less significant as the threshold increases, a reflection of decreasing predictability as the magnitude of flow increases. From these scores, one may see the generally positive effects of post-processing in ensemble streamflow forecasting. Compared to ESP ensembles forced by Day 1 EQPF, one may see the increased skill in ESP ensembles forced by Day 1-3 EQPF due to longer-lead QPF. The improvement in skill due to post-processing is nominal for JAKT2 and SGET2, whether forced by Day 1 or Day 1-3 QPF. It is suspected that the reduced performance by EnsPost is due to the fact that, currently, EnsPost cannot effectively deal with ephemeral basins for which no streamflow may occur at certain times of the year, and that the quality of hydrologic simulation is generally low. Additional investigation is necessary, however, for quantitative assessment of them. The fraction of the hindcast period for which there was no flow is about 36% for BRPT2, 26% for DCJT2, 13% for GLLT2, 45% for JAKT2 and 23% for SGET2.

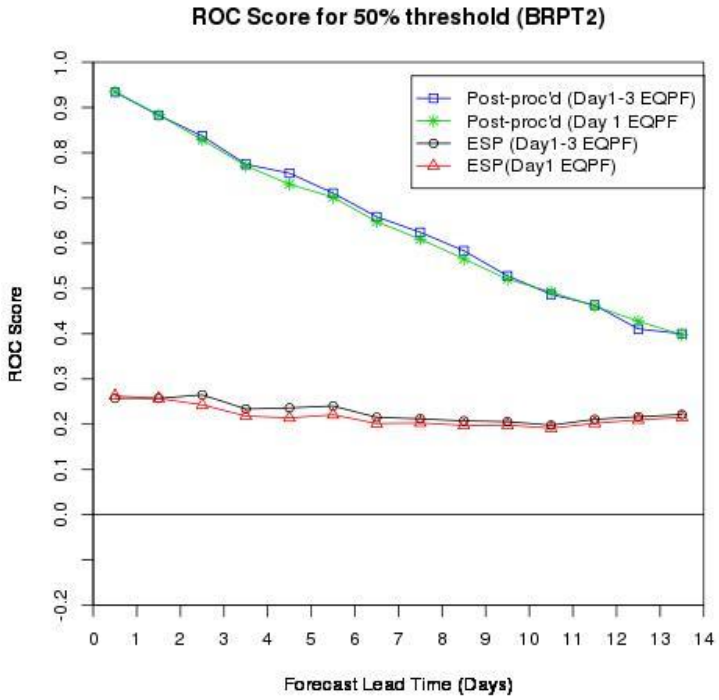

Figure 4.49. ROC Scores at 50% Threshold (BRPT2)

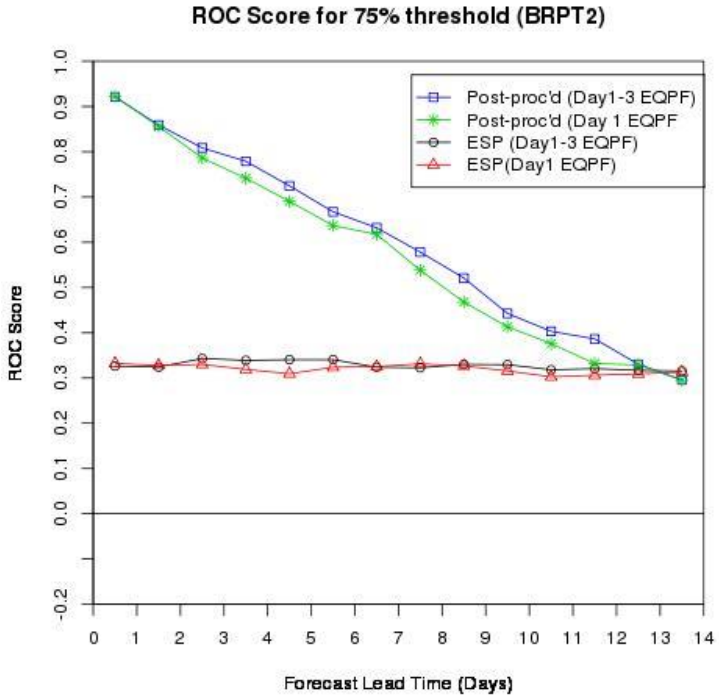

Figure 4.50. ROC Scores at 75% Threshold (BRPT2)

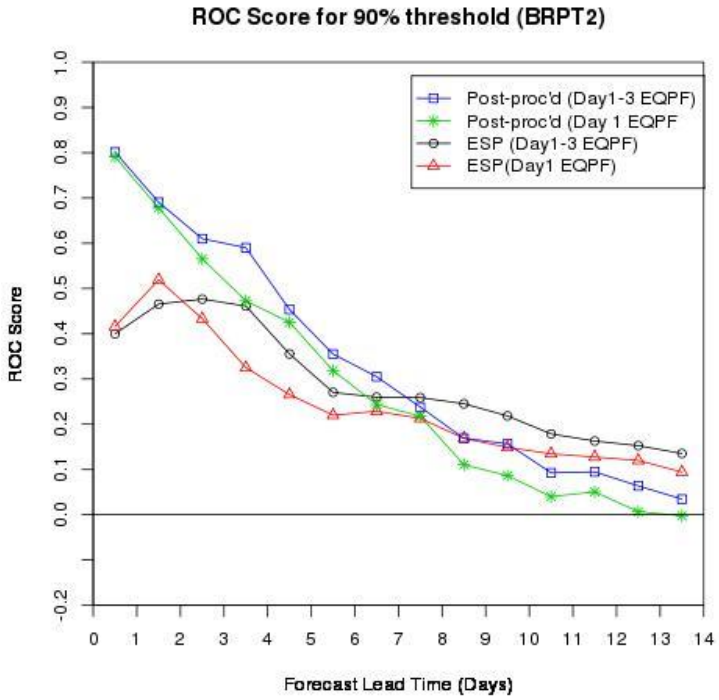

Figure 4.51. ROC Scores at 90% Threshold (BRPT2)

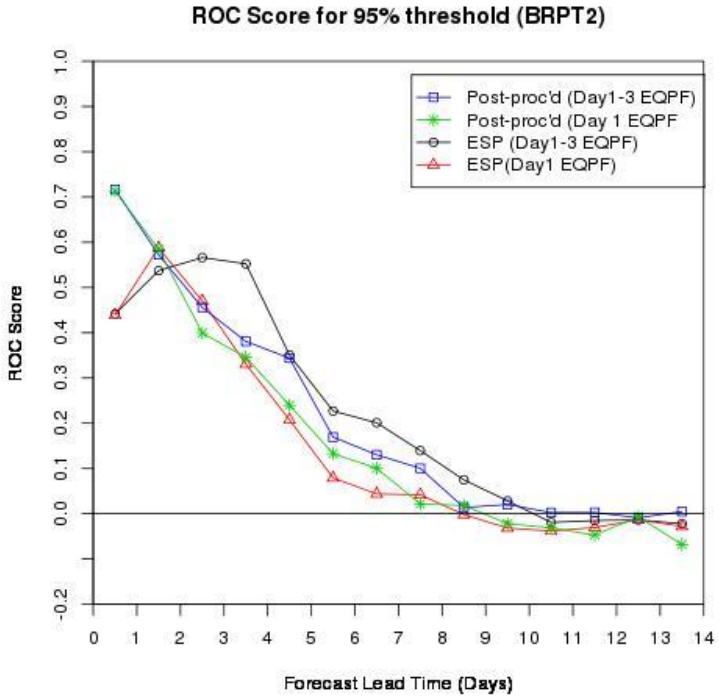

Figure 4.52. ROC Scores at 95% Threshold (BRPT2)

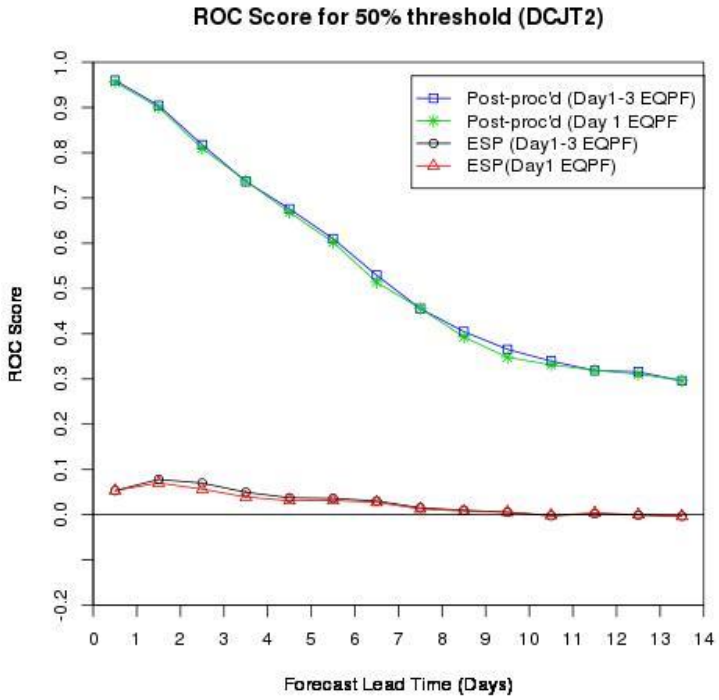

Figure 4.53. ROC Scores at 50% Threshold (DCJT2)

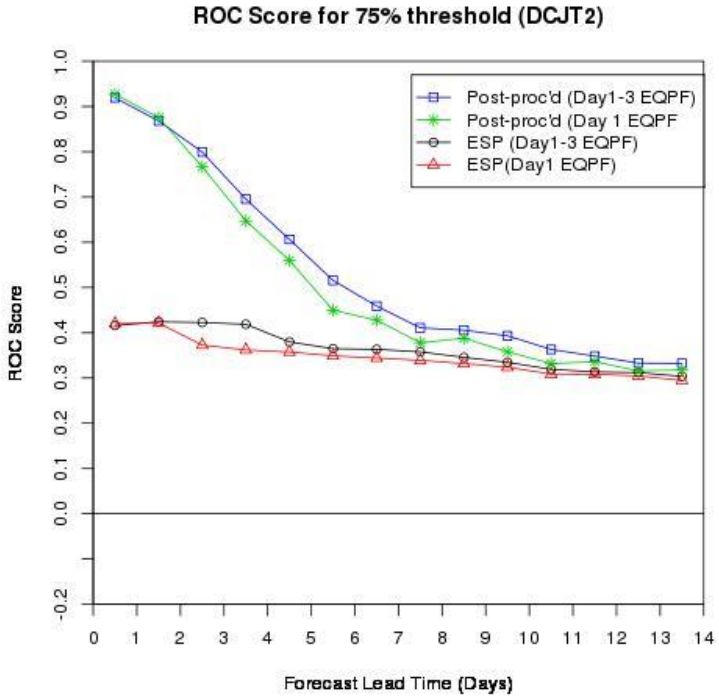

Figure 4.54. ROC Scores at 75% Threshold (DCJT2)

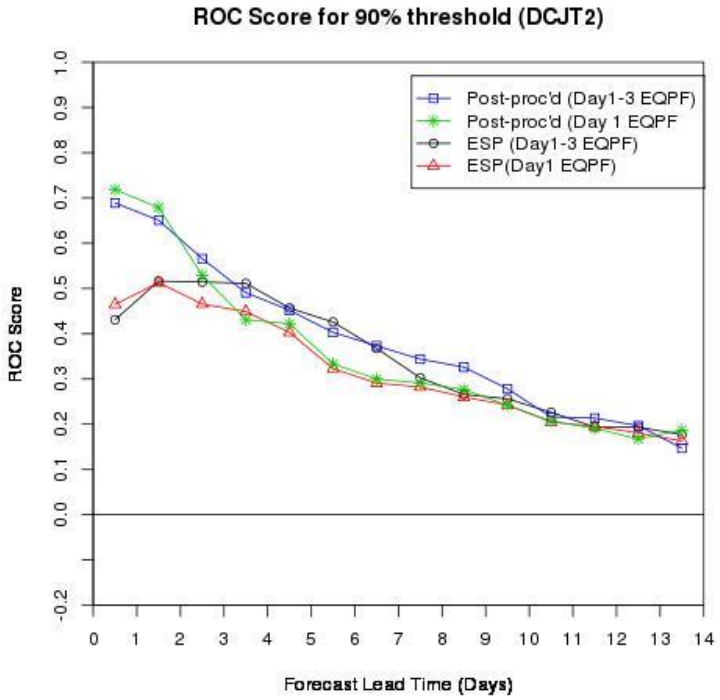

Figure 4.55. ROC Scores at 90% Threshold (DCJT2)

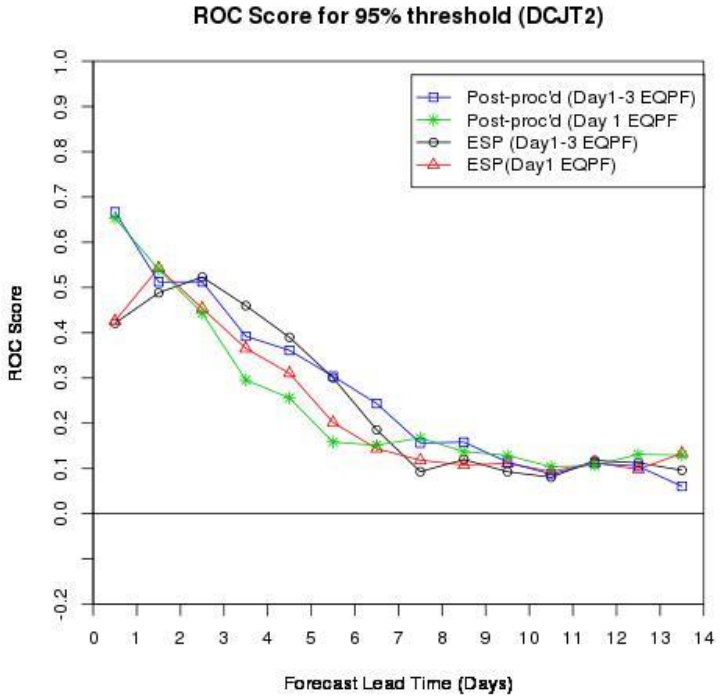

Figure 4.56. ROC Scores at 95% Threshold (DCJT2)

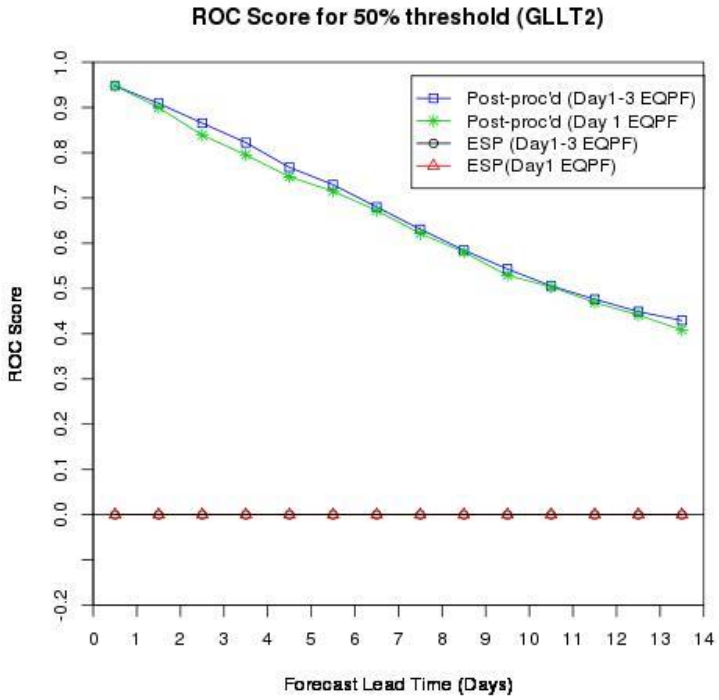

Figure 4.57. ROC Scores at 50% Threshold (GLLT2)

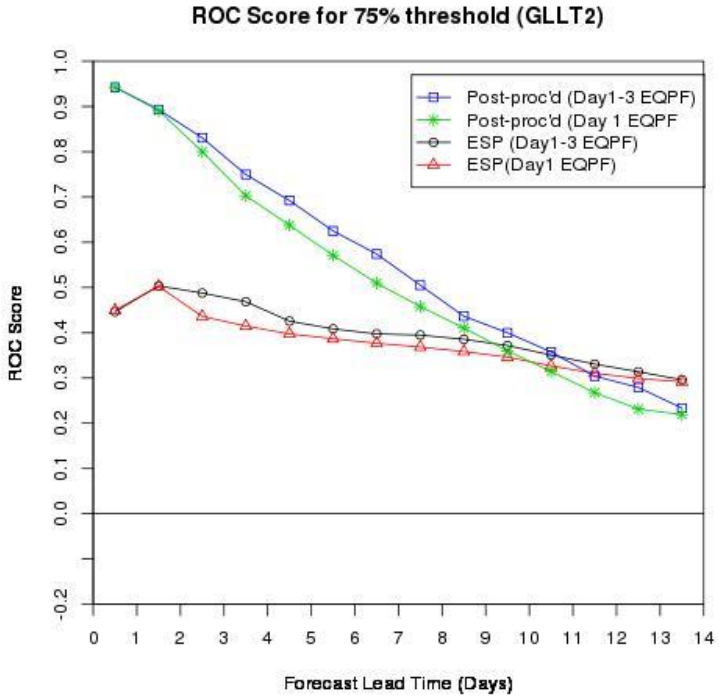

Figure 4.58. ROC Scores at 75% Threshold (GLLT2)

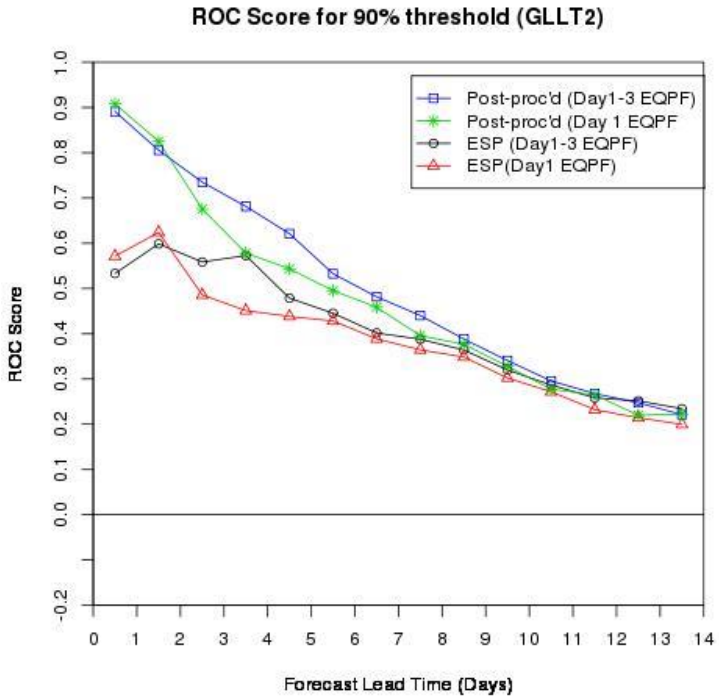

Figure 4.59. ROC Scores at 90% Threshold (GLLT2)

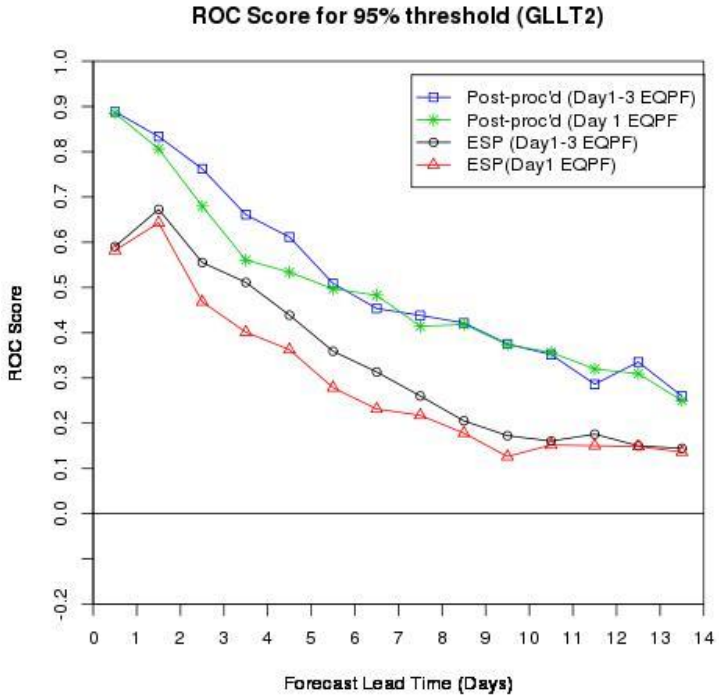

Figure 4.60. ROC Scores at 95% Threshold (GLLT2)

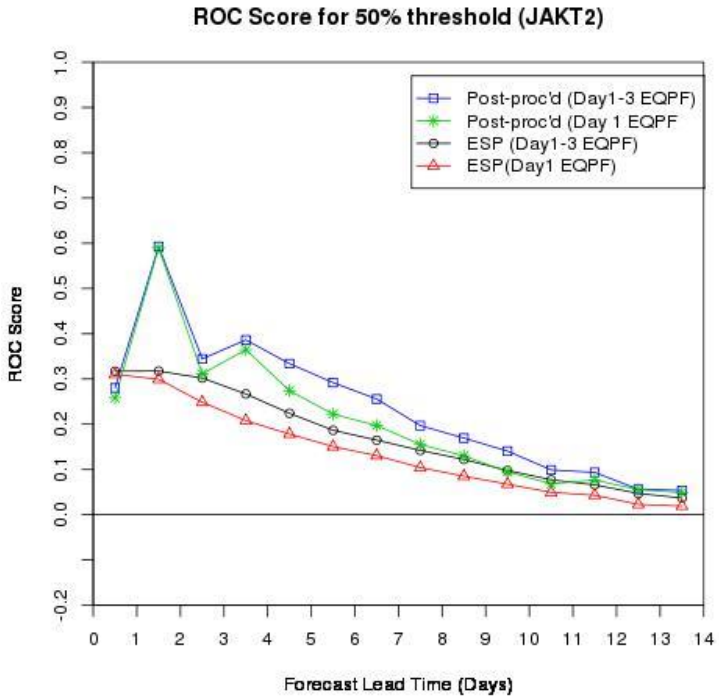

Figure 4.61. ROC Scores at 50% Threshold (JAKT2)

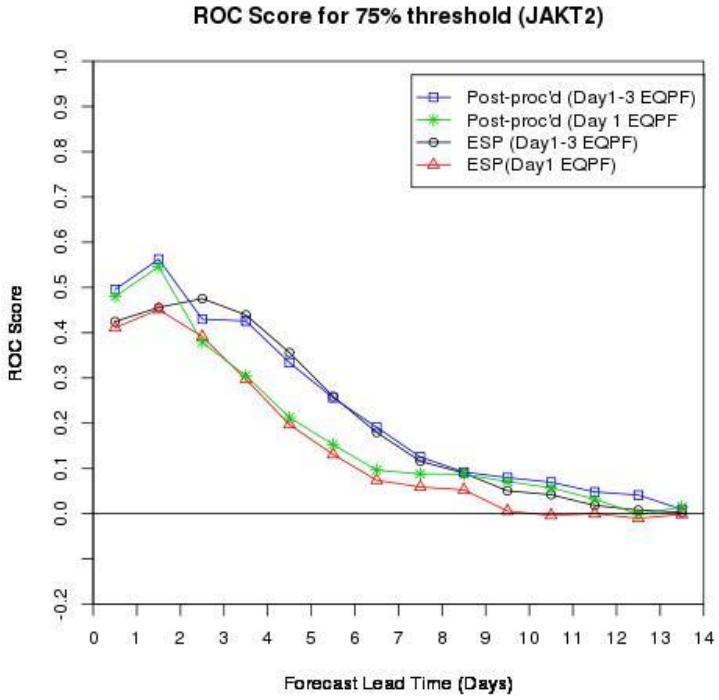

Figure 4.62. ROC Scores at 75% Threshold (JAKT2)

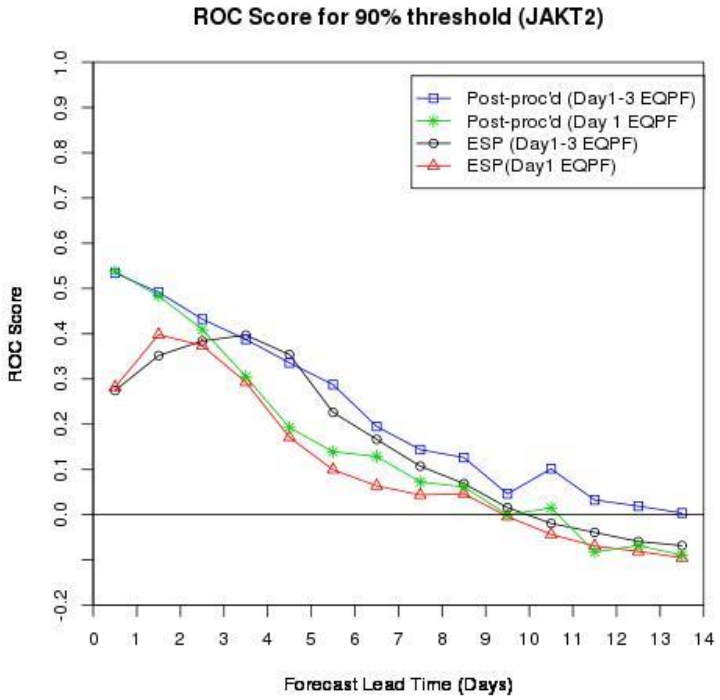

Figure 4.63. ROC Scores at 75% Threshold (JAKT2)

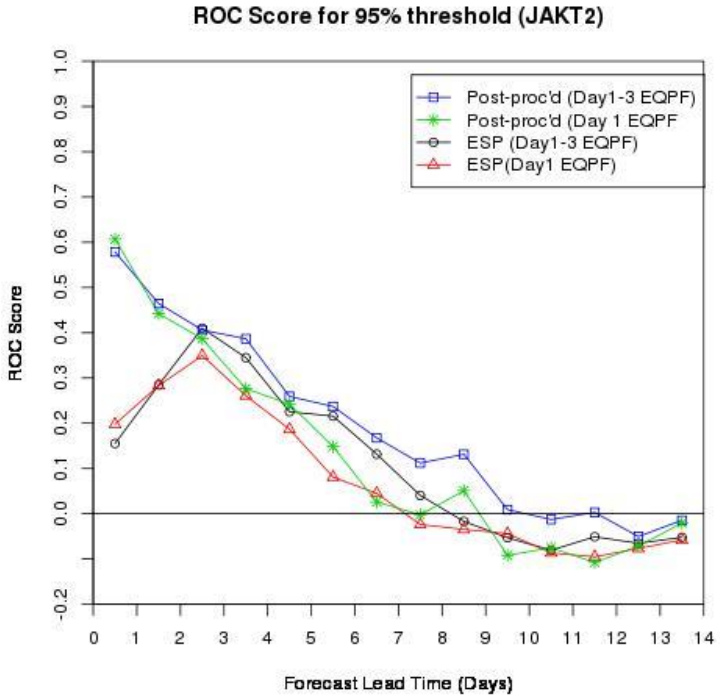

Figure 4.64. ROC Scores at 95% Threshold (JAKT2)

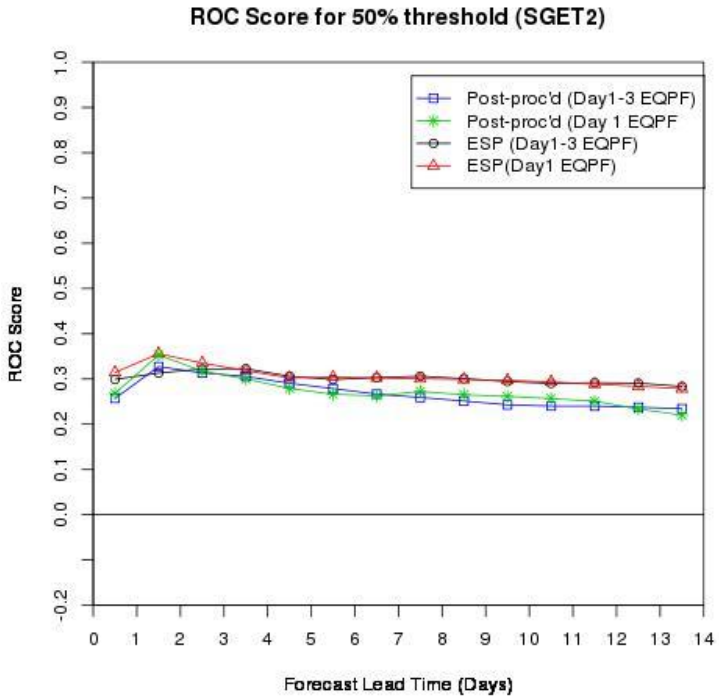

Figure 4.65. ROC Scores at 50% Threshold (SGET2)

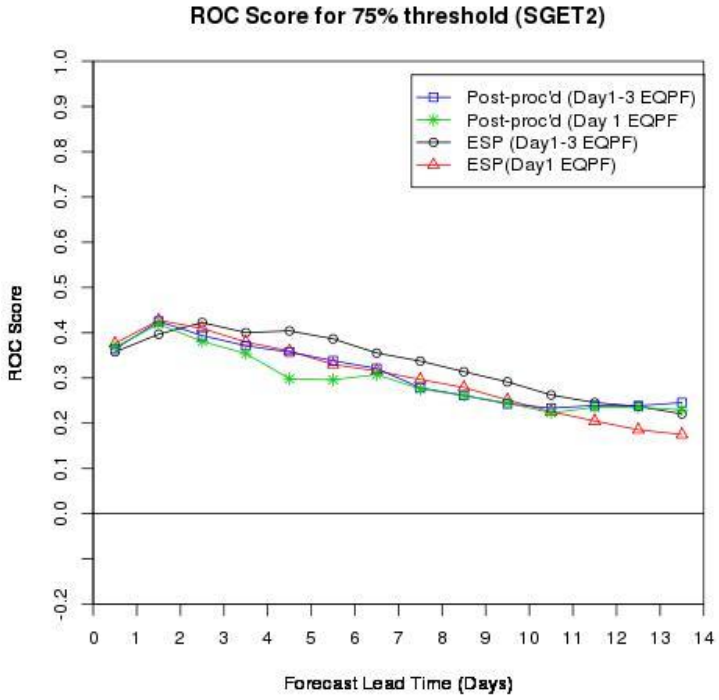

Figure 4.66. ROC Scores at 75% Threshold (SGET2)

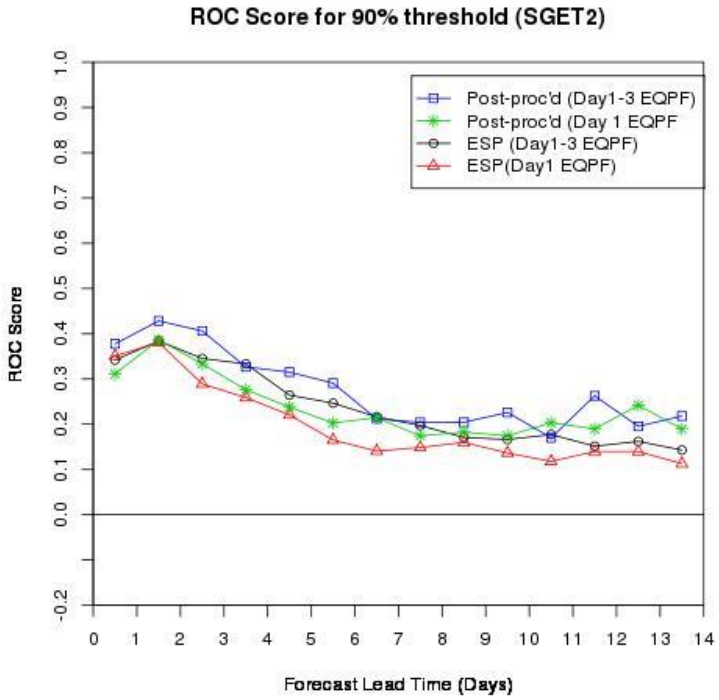

Figure 4.67. ROC Scores at 90% Threshold (SGET2)

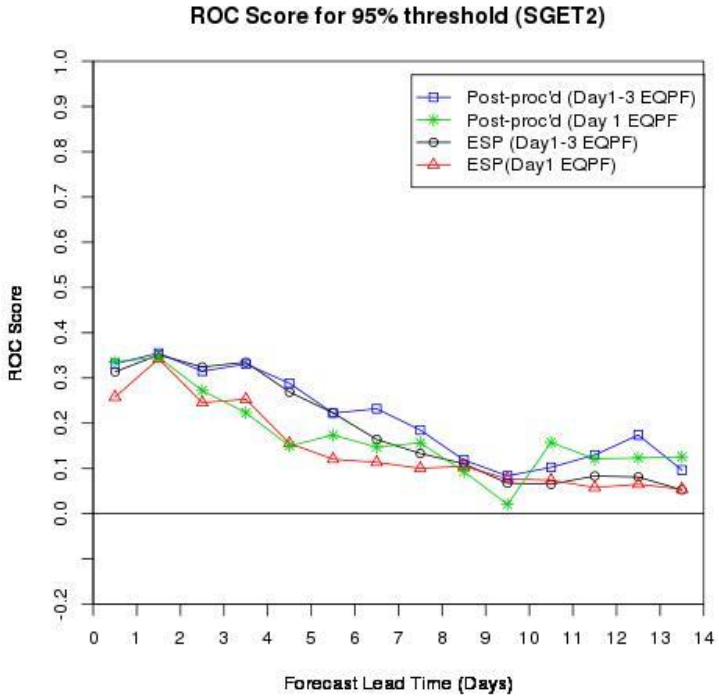

Figure 4.68. ROC Scores at 95% Threshold (SGET2)

4.4 Analysis of ROC according to acceptable FAR

To relate increase in ROC scores to improvement in forecast quality in a single-valued sense, the ROC results shown above are expressed as changes in POD according to specific levels of FAR. Some decision makers may prefer a lower FAR at the expense of a lower POD. Others may prefer a higher POD even if it may increase FAR. An FAR of 1 (%) denotes an extreme aversion to false alarms while an FAR of 50 (%) denotes an extreme tolerance to false alarms. A positive bar denotes an increase in POD by the addition of Day 2-3 QPF and/or EnsPost. Figures 4.69 through 4.73 show the increase in POD corresponding to the FAR levels of 1, 5, 10, 20 and 50 (%), respectively. They show the increase in POD in Day 1-3 EQPF-forced ESP over Day 1 EQPF-forced ESP for all five basins. The addition of Day 2-3 QPF does not lead to an increase in POD over Day 1 EQPF-forced ESP ensembles, for Day 1 for all basins. The greatest gains in POD are seen for Days 3-5. These results are subject to sampling uncertainties and may change as sample size increases, but it can be clearly seen that there is an increase in POD in ESP forecasts due to usage of Day 2-3 QPF.

Now, the acceptable FAR is fixed at 5 (%) and increase in POD due to usage of EnsPost is examined. Figure 4.74 shows the increase in POD due to the addition of Day 2-3 QPF and EnsPost. Compared with Figure 4.70, which shows the increase in POD due to the addition of Day 2-3 QPF, a substantial increase in POD due to usage of Day 2-3 QPF and EnsPost can be seen. Figures 4.75 and 4.76 as well as Figures 4.77 and 4.78 are the same figures as Figures 4.70 and 4.74, but for the 90^{th} and 75^{th} threshold flow, respectively. Note that as the threshold flow decreases, the impact of Day 2-3 QPF decreases and that of EnsPost increases; this is because hydrologic uncertainty is dominant over input uncertainty in low flow conditions and is much more easily corrected than in high flow conditions owing to persistence

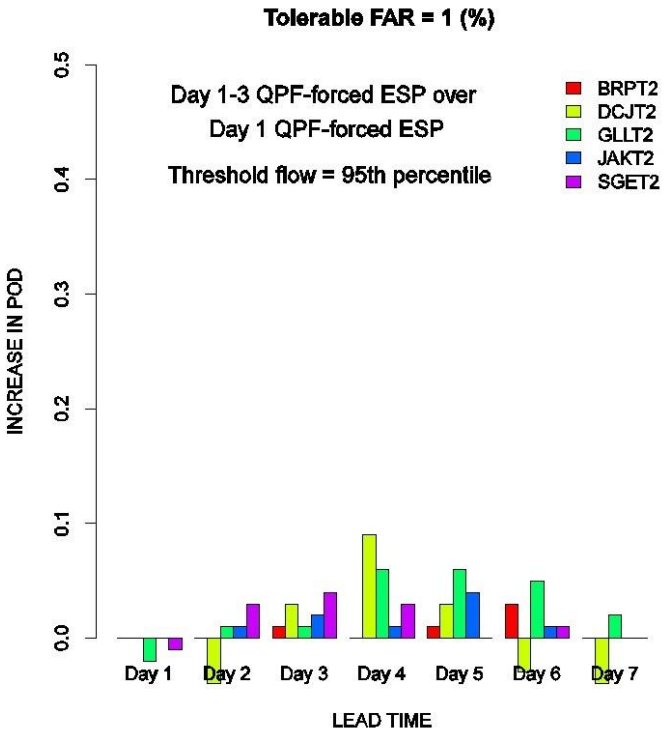

Figure 4.69. Increase in POD in ESP forecast due to Day 2-3 QPF at FAR 1%

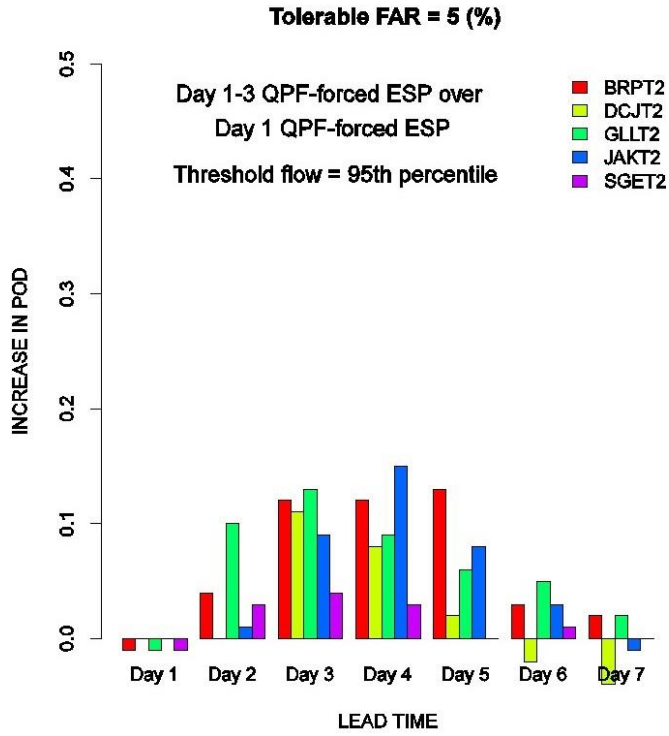

Figure 4.70. Increase in POD in ESP forecast due to Day 2-3 QPF at FAR 5%

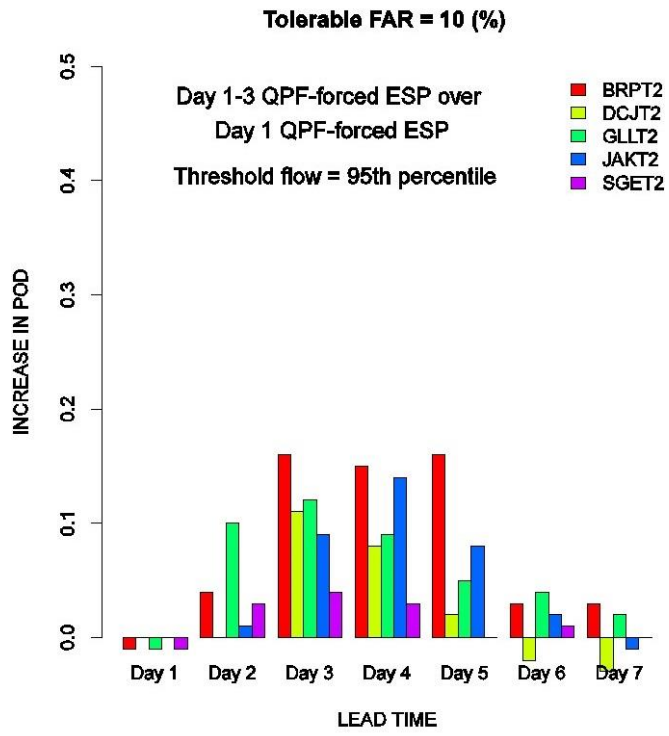

Figure 4.71. Increase in POD in ESP forecast due to Day 2-3 QPF at FAR 10%

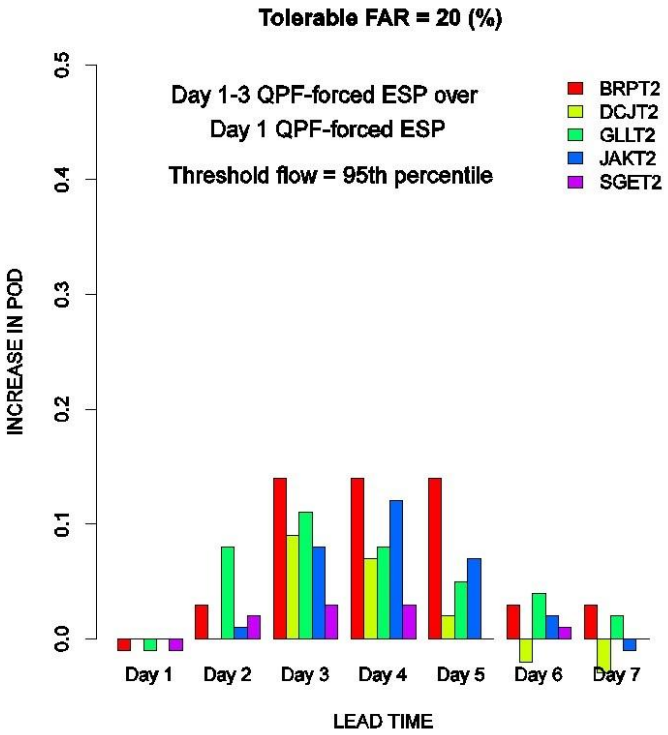

Figure 4.72. Increase in POD in ESP forecast due to Day 2-3 QPF at FAR 20%

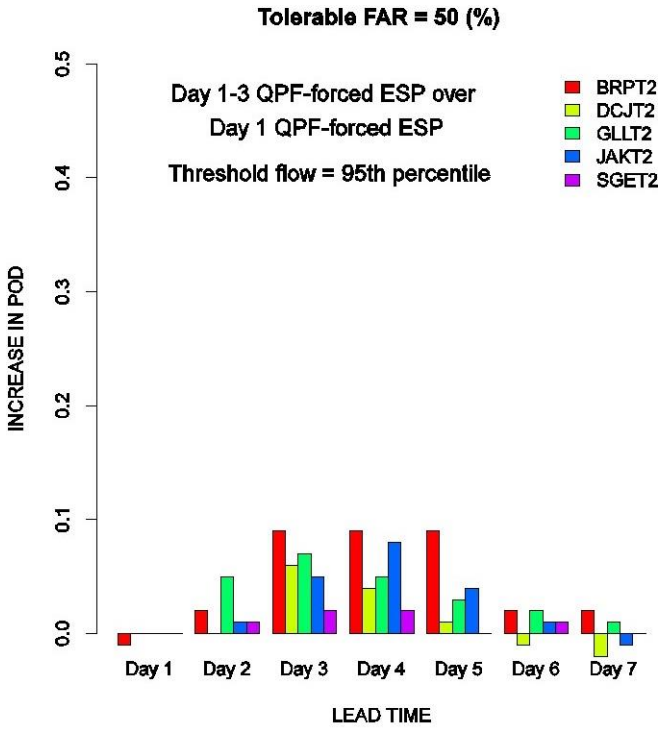

Figure 4.73. Increase in POD in ESP forecast due to Day 2-3 QPF at FAR 50%

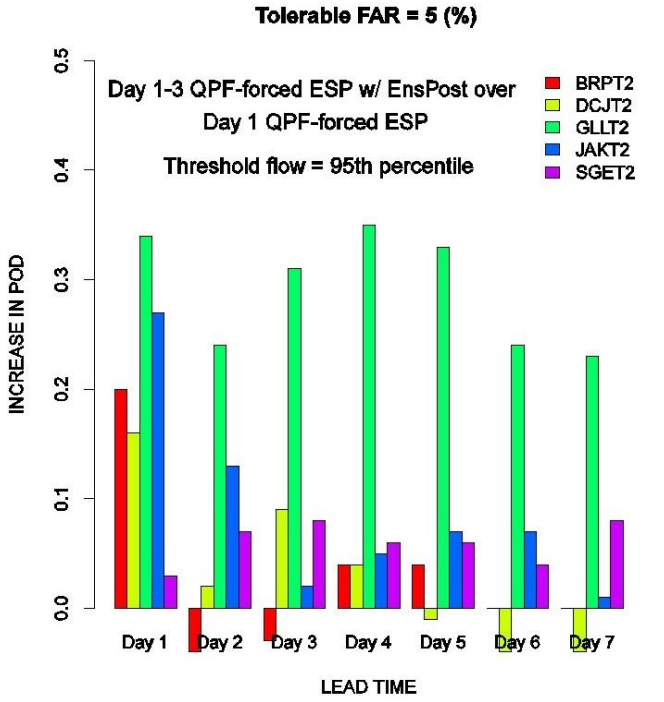

Figure 4.74. Increase in POD due to Day 2-3 QPF and EnsPost at FAR 5% and 95% threshold

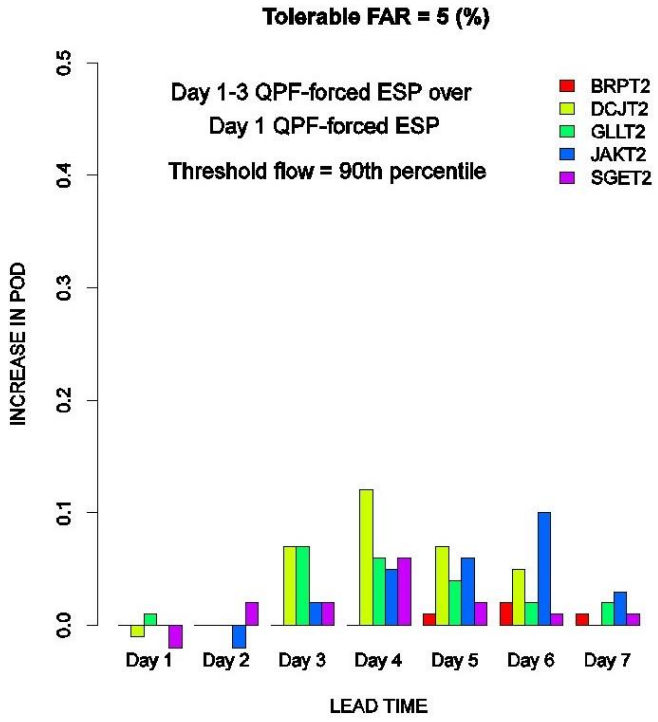

Figure 4.75. Increase in POD due to Day 2-3 QPF at FAR 5% and 90% threshold

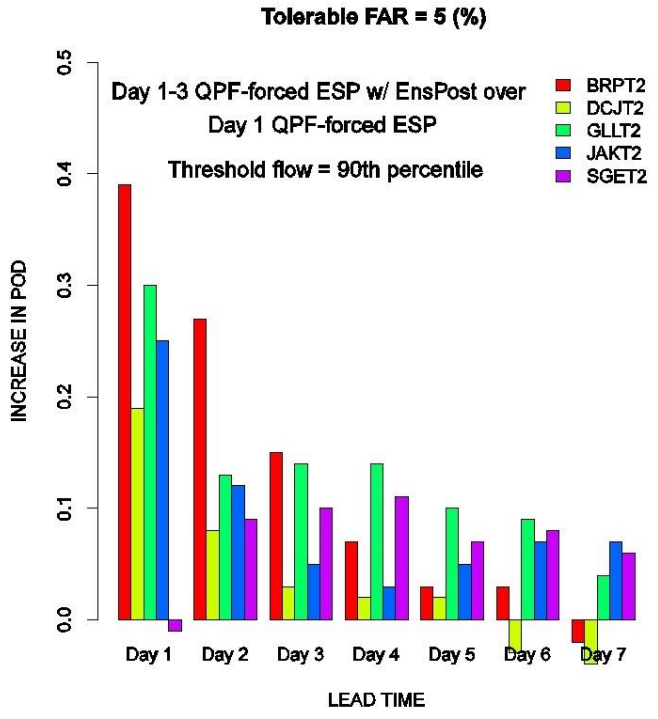

Figure 4.76. Increase in POD due to Day 2-3 QPF and EnsPost at FAR 5% and 90% threshold

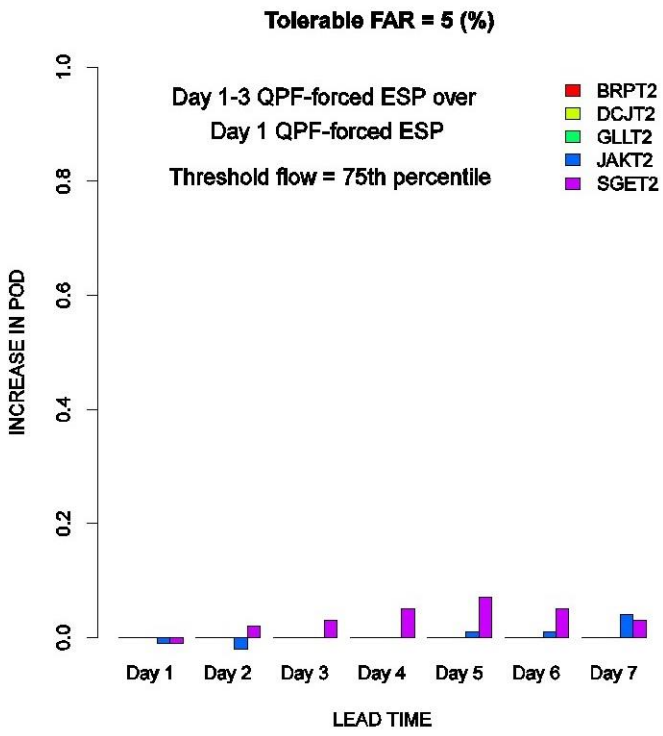

Figure 4.77. Increase in POD due to Day 2-3 QPF at FAR 5% and 75% threshold

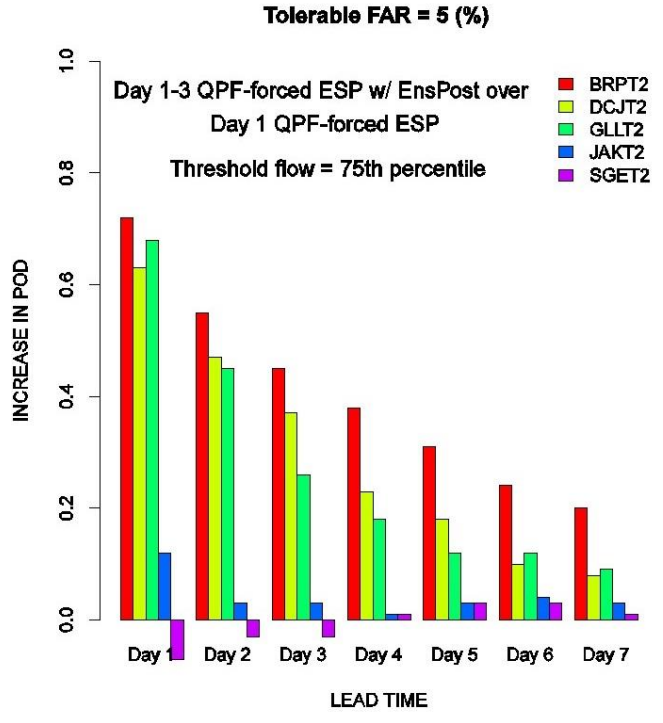

Figure 4.78. Increase in POD due to Day 2-3 QPF and EnsPost at FAR 5% and 75%Threshold

CHAPTER 5

CONCLUSIONS AND FUTURE RESEARCH RECOMMENDATIONS

The objective of this study was to assess the improvement in skill in short-range ensemble streamflow forecasting due to the use of longer-lead quantitative precipitation forecasts (QPF) and ensemble post-processing. For the study, a series of hindcasting experiments were conducted using the Hydrologic Ensemble Forecast Service (HEFS) following Demargne et al. (2010, 2013) ([14], [10]). Ensemble forecasts of precipitation and streamflow were generated using HEFS for a 7 year period between 2004 and 2010. The precipitation, raw and post-processed streamflow ensembles were verified using the Ensemble Verification System (EVS) to assess the quality of hindcasts. The study area consisted of five headwater basins in the Upper Trinity River Basin in Texas. These study basins offer a tough test for HEFS, both hydrometeorologically and hydrologically. Precipitation is dominated by convection which has very limited predictability. The basins are flashy with fast-rising streamflow when they respond to rainfall but also with periods of no streamflow.

The Meteorological Ensemble Forecast Processor (MEFP) was used to generate ensemble precipitation hindcasts from the single-valued QPF produced by WGRFC for 1 to 3 days of lead time and climatology for 4 to 14 days of lead time. The precipitation hindcasts are ingested into a suite of hydrologic models to produce raw streamflow forecasts. For all basins, the streamflow hindcasts were generated via the Community Hydrologic Prediction System (CHPS) using the Sacramento Soil Moisture Accounting model (SAC-SMA) and unit hydrograph (UNIT HG). The raw streamflow ensembles produced from CHPS accounts only for input uncertainty and may be biased in their mean and spread. The raw streamflow ensembles are post-processed by the Ensemble Postprocessor (EnsPost) to produce streamflow ensembles that account for bias and uncertainty in the hydrologic models. The precipitation and streamflow ensembles were verified using the Ensemble Verification System (EVS).

The main findings from this work are as follows.

1. The ensemble QPFs from single-valued QPFs generated by the Meteorological Ensemble Forecast Processor (MEFP) are generally skillful and reliable as have been found in previous studies [12].
2. Compared to using Day 1 QPF only, using Day 1-3 single-valued QPF via HEFS significantly increases the skill in short-range ESP forecasts. The Relative Operating Characteristic (ROC) results indicate that the addition of Day 2-3 QPF increase the probability of detection (POD) of 95th percentile flow by about 10 percent for Day 3-4 streamflow prediction and extends the lead time by about a day.
3. Post-processing of ESP ensembles via EnsPost generally improves discrimination and reliability of the raw ESP ensembles. For high flow thresholds, the addition of Day 2-3 QPF is more important than post-processing as the input uncertainty is more dominant. For low flow thresholds, however, the opposite holds true as the hydrologic uncertainty is dominant.
4. To the best of the author's knowledge, this is the first time that HEFS has been successfully used to produce ensemble hindcasts outside of NWS, thus demonstrating portability of the software.

The main recommendations for future research are as follows:

1. Extend the study to a large number of basins for large-sample verification, especially for large events.
2. Carry out similar hindcasting and verification experiments using the Global Ensemble Forecasting System (GEFS) reforecast dataset. This is expected to further extend the lead time of ensemble streamflow forecasts.
3. Improve EnsPost to handle ephemeral basins.
4. Develop and implement the parametric uncertainty processor and the ensemble data assimilator (DA). Due to the factors contributing to nonstationarity in the streamflow processes such as urbanization and possibly climate change, purely statistical techniques for modeling and reducing hydrologic uncertainty are of limited applicability. Modeling and reducing parametric and initial condition uncertainties are necessary to whiten the residual uncertainty as much as possible and hence to allow uncertainty modeling under nonstationarity.

REFERENCES

[1] R. Krzysztofowicz, "Bayesian theory of probabilistic forecasting via deterministic hydrologic model", Water Resources Research, vol. 35.9, p. 27392750, 1999.

[2] D.-J. Seo, H. Herr, and J. Schaake, "A statistical post-processor for accounting of hydrologic uncertainty in short-range ensemble streamflow prediction", Hydrology and Earth System Sciences, vol. 3, pp. 1987-2035, 2006.

[3] F. Pappenberger and K. Beven, "Ignorance is bliss: Or seven reasons not to use uncertainty analysis", Water Resources Research, vol. 42, p. W05302, 2006.

[4] L. Matott, J. Babendreier, and S. Parucker, "Evaluating uncertainty in integrated environmental models: A review of concepts and tools", Water Resources Research, vol. 45, p. WO6421, 2009.

[5] J. D. Brown, J. Demargne, D.-J. Seo, and Y. Liu, "The ensemble verification system (EVS): a software tool for verifying ensemble forecasts of hydrometeorological and hydrologic variables at discrete locations", Environmental Modelling & Software, vol. 25, pp. 854-872, 2010b.

[6] M. H. Ramos, S. J. van Andel, and F. Pappenberger, "Do probabilistic forecasts lead to better decisions?", Hydrology and Earth System Sciences, vol. 9, pp. 13 569-13 607, 2012.

[7] J. Demargne, L. Wu, D.-J. Seo, J. Schaake, E. Boegh, H. Kunstmann, T. Wagener, A. Hall, L. Bastidas, S. Franks, et al., "Experimental hydrometeorological and hydrological ensemble forecasts and their verification in the us national weather service", in Proceedings of the Symposium Quantification and Reduction of Predictive Uncertainty for Sustainable Water Resources Management, vol. 4.2. Perugia, Italy: IAHS Press, 2007, pp. 655-717.

[8] J. Demargne, J. D. Brown, Y. Liu, D.-J. Seo, L. Wu, Z. Toth, , and Y. Zhu, "Diagnostic verification of hydrometeorological and hydrologic ensembles", Atmospheric Science Letters, vol. 11(2), pp. 114-122, 2010.

[9] D.-J. Seo, J. Demargne, L. Wu, Y. Liu, J. D. Brown, S. Regonda, , and H. Lee, "Hydrologic ensemble prediction for risk-based water resources management and hazard mitigation", in 4th Federal Interagency Hydrologic Modeling Conference, Las Vegas, NV., June 27-July 1, 2010.

[10] J. Demargne, L. Wu, S. Regonda, J. Brown, H. Lee, M. He, D.-J. Seo, R. Hartman, M. Fresch, , and Y. Zhu, "The science of NOAA's operational hydrologic ensemble forecast service", Bulletin of the American Meteorological Society, 2013.

[11] J. Schaake, J. Demargne, R. Hartman, M. Mullusky, E. Welles, L. Wu, H. Herr, X. Fan, and D. J. Seo., "Precipitation and temperature ensemble forecasts from single-value forecasts", Hydrology and Earth System Sciences, vol. 2, pp. 655-717, 2007.

[12] L. Wu, D.-J. Seo, J. Demargne, J. Brown, S. Cong, and J. Schaake, "Generation of ensemble precipitation forecast from single-valued quantitative precipitation forecast via meta-gaussian distribution models", Journal of Hydrology, vol. 399(3-4), pp. 281-298, 2011.

[13] D. Wilks, Statistical Methods in the Atmospheric Sciences, 2nd ed., Ed. San Diego: Elsevier, 2006.

[14] J. Demargne, Y. Liu, J. Brown, D.J. Seo, L. Wu , A. Weerts and M. Werner, "Hydrologic Ensemble Hindcasting & Verification in the U.S. National Weather Service", EGU, Vienna, 2010

[15] M. Clark, S. Gangopadhyay, L. Hay, B. Rajagopalan, and R. Wilby, "The Schaake shuffle: A method for reconstructing space-time variability in forecasted precipitation and temperature fields", Journal of Hydrometeorology 5.1: 243-262, 2004.

[16] A. Velázquez, T. Petit, A. Lavoie, M.-A. Boucher, R. Turcotte, V. Fortin, and F. Anctil, "An evaluation of the Canadian global meteorological ensemble prediction system for short-term hydrological forecasting", Hydrol. Earth Syst. Sci., 13, 2221-2231, 2009.

[17] I.T. Jolliffe, and D.B. Stephenson, "Forecast Verification. A Practitioner's Guide in Atmospheric Science", John Wiley and Sons, Chichester, 240 pp, 2012.

[18] D-J. Seo, "Conditional bias-penalized kriging", Stochastic Environmental Research and Risk Assessment. DOI 10.1007/s00477-012-0567-z, 2012.

[19] G.B.M. Heuvelink, James D. Brown, and E. E. Van Loon, "A probabilistic framework for representing and simulating uncertain environmental variables", International Journal of Geographical Information Science 21.5: 497-513, 2007.

[20] D. Kavetski, G. Kuczera, and S. W. Franks, "Calibration of conceptual hydrological models revisited: 1. Overcoming numerical artefacts", Journal of Hydrology 320.1 : 173-186, 2006.

[21] D. Kavetski, G. Kuczera, and S. W. Franks, "Bayesian analysis of input uncertainty in hydrological modeling: 2. Application", Water Resources Research 42.3: W03408, 2006.

[22] J. Schaake, K. Franz, A. Bradley, and Buizza, R., "The Hydrologic Ensemble Prediction EXperiment (HEPEX)", Hydrology and Earth Systems Sciences Discussion 3: 3321–3332, 2006.

[23] Y. Liu, and H. V. Gupta, "Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework", Water Resources Research 43, W07401, doi:10.1029/2006WR005756, 2007.

[24] H. Glahn, and D. Lowry, "The Use of Model Output Statistics (MOS) in Objective Weather Forecasting", Journal of Applied Meteorology 11(8), 1203-1211., 1972.

[25] T. Gneiting, and A.E. Raftery, "Weather forecasting with ensemble methods", Science 310(5746), 248-249., 2005.

[26] S. Regonda, D-J. Seo, B. Lawrence, J.D. Brown, and J. Demargne, "Short-term ensemble streamflow forecasting using operationally produced single-valued streamflow forecasts – A Hydrologic Model Output Statistics (HMOS) approach", Journal of Hydrology, 2013.

[27] J.B. Kadane, and L.J. Wolfson, "Experiences in elicitation", The Statistician 47, 3–19, 1998.

[28] R. Rojas, L. Feyen, and A. Dassargues, "Sensitivity analysis of prior model probabilities and the value of prior knowledge in the assessment of conceptual model uncertainty in groundwater modelling", Hydrological Processes 23(8), 1131-114, 2009.

[29] G.B.M. Heuvelink, "Error Propagation in Environmental Modelling with GIS", Taylor and Francis, London, 1998.

[30] J. Oakley, and A. O'Hagan, “Probabilistic sensitivity analysis of complex models: a Bayesian approach”, Journal of the Royal Statistical Society, B66, 751–769, 2004.

[31] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola, “Global sensitivity analysis”, The Primer, John Wiley and Sons, Chichester, 2008

[32] Meteorological Ensemble Forecast Processor (MEFP) User’s Manual, National Weather Service, Office of Hydrologic Development.

[33] J. D. Brown, “Ensemble Verification System (EVS) User’s Manual”, Hydrologic Ensemble Prediction group, Office of Hydrologic Development, National Weather Service

[34] J.C.R. Burnash, “The NWS river forecast system-catchment modelling”, V.P. Sigh (Ed.), Computer Models of Watershed Hydrology, Water Resources Publications, Colorado, pp. 311–366, 1995.

[35] G. N. Day, “Extended streamflow forecasting using NWSRFS”, Journal of Water Resources Planning and Management, 111(2), 157–170, 1985.

[36] K. S. Kelly, and R. Krzysztofowicz, "A bivariate meta-Gaussian density for use in hydrology", Stochastic Hydrology and Hydraulics 11.1, 17-31, 1997.

[37] R. Krzysztofowicz, "Transformation and normalization of variates with specified distributions", Journal of Hydrology 197.1-4: 286-292, 1997.

[38] E. A. Anderson, “National Weather Service River Forecast System – Snow accumulation and ablation model”, NOAA Technical Memorandum NWS HYDRO-17. US Dept of Commerce, Silver Spring, Maryland, USA, 1973.

[39] Dallas Water Utilities (DWU), “2012 Drinking Water Quality Report”, 2012.

[40] M. Werner, J. Schellekens, P. Gijsbers, M. van Dijk, O. van den Akker, K. Heynert, “The Delft-FEWS flow forecasting system”, Environmental Modelling & Software, Volume 40, February Pages 65-77, ISSN 1364-8152, http://dx.doi.org/10.1016/j.envsoft.2012.07.010, 2013.

[41] M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli, & G. Guariso, “Delft FEWS: A proven infrastructure to bring data, sensors and models together." International Congress on Environmental Modelling and Software, iEMS, 2008.

[42] E. de Rooij and M. Werner. "Delft FEWS: an open interface that connects models and data streams for operational forecasting systems.", EGU General Assembly Conference Abstracts, Vol. 12, 2010.

BIOGRAPHICAL STATEMENT

Manabendra Saharia was born in Guwahati, Assam, India, in 1987. He received his B.Tech degree in Civil Engineering from National Institute of Technology, Silchar, India, in 2011. His current research interest is hydrologic forecasting.