

PROBABILISTIC FLOOD FORECAST

USING BAYESIAN METHODS

PROBABILISTIC FLOOD FORECAST USING BAYESIAN METHODS

By SHASHA HAN, M.Sc., B.Sc.

Faculty of Engineering

Civil Engineering

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

DOCTOR OF PHILOSOPHY

McMaster University © Copyright by Shasha Han, June 2019

DOCTOR OF PHILOSOPHY (2019)

McMaster University

Civil Engineering

Hamilton, Canada

TITLE:

Probabilistic Flood Forecast Using

Bayesian Methods

AUTHOR:

Shasha Han

M.Sc. (Beijing Normal University)

B.Sc. (Qingdao University of Science and Technology)

SUPERVISOR:

Dr. Paulin Coulibaly

NUMBER OF PAGES:

xxiii, 209

Lay Abstract

Flood is one of the top weather related hazards and causes serious property damage and loss of lives every year worldwide. If the timing and magnitude of the flood event could be accurately predicted in advance, it will allow time to get well prepared, and thus reduce its negative impacts. This research focuses on improving flood forecasts through advanced Bayesian techniques. The main objectives are: (1) enhancing reliability and accuracy of flood forecasting system; and (2) improving the assessment of predictive uncertainty associated with the flood forecasts. The key contributions include: (1) application of Bayesian forecasting methods in a semi-urban watershed to advance the predictive uncertainty quantification; and (2) investigation of the Bayesian forecasting methods with different inputs and models and combining bias correction technique to further improve the forecast performance. It is expected that the findings from this research will benefit flood impact mitigation, watershed management and water resources planning.

Abstract

The number of flood events and the estimated costs of floods have increased dramatically over the past few decades. To reduce the negative impacts of flooding, reliable flood forecasting is essential for early warning and decision making. Although various flood forecasting models and techniques have been developed, the assessment and reduction of uncertainties associated with the forecast remain a challenging task. Therefore, this thesis focuses on the investigation of Bayesian methods for producing probabilistic flood forecasts to accurately quantify predictive uncertainty and enhance the forecast performance and reliability.

In the thesis, hydrologic uncertainty was quantified by a Bayesian post-processor - Hydrologic Uncertainty Processor (HUP), and the predictability of HUP with different hydrologic models under different flow conditions were investigated. Followed by an extension of HUP into an ensemble prediction framework, which constitutes the Bayesian Ensemble Uncertainty Processor (BEUP). Then the BEUP with bias-corrected ensemble weather inputs was tested to improve predictive performance. In addition, the effects of input and model type on BEUP were investigated through different combinations of BEUP with deterministic/ensemble weather predictions and lumped/semi-distributed hydrologic models.

Results indicate that Bayesian method is robust for probabilistic flood forecasting with uncertainty assessment. HUP is able to improve the deterministic forecast from the hydrologic model and produces more accurate probabilistic forecast. Under high flow

condition, a better performing hydrologic model yields better probabilistic forecast after applying HUP. BEUP can significantly improve the accuracy and reliability of short-range flood forecasts, but the improvement becomes less obvious as lead time increases. The best results for short-range forecasts are obtained by applying both bias correction and BEUP. Results also show that bias correcting each ensemble member of weather inputs generates better flood forecast than only bias correcting the ensemble mean. The improvement on BEUP brought by the hydrologic model type is more significant than the input data type. BEUP with semi-distributed model is recommended for short-range flood forecasts.

Acknowledgements

I would like to express my heartfelt gratitude to my supervisor Prof. Paulin Coulibaly, for providing me all the support, for taking great effort to guide my research in each step, and for giving me lots of encouragement that helped to overcome the difficulties in my study. I greatly appreciate his patience, caring, contribution and supervision throughout my Ph.D. study. Without his help, it would not have been possible to achieve this milestone in my life.

I would also like to express my sincere gratitude to the other members of my supervisory committee, Prof. Altaf Arain and Prof. Sarah Dickson, for their helpful advices and insightful comments that have improved my research. I am also extremely thankful to Prof. Daniela Biondi at University of Calabria for hosting me as a visiting student and providing valuable technical support.

I thank all my colleagues at McMaster Water Resources and Hydrologic Modeling Lab (WRHML) for sharing time and knowledge, for helping me with any question, and for making my graduate life fun and exciting. I would also like to thank all my friends for their friendship and endless support, I enjoyed every moment with them.

I acknowledge Toronto and Region Conservation Authority (TRCA), Water Survey of Canada (WSC), Environment and Climate Change Canada (ECCC) and Canadian Surface Prediction Archive (CaSPAr) for the study data. I also acknowledge the financial support from the China Scholarship Council (CSC), the Natural Science and Engineering

Research Council (NSERC) of Canada, the NSERC Canadian FloodNet, the School of Graduate Studies, the Department of Civil Engineering at McMaster University and the McMaster Centre for Climate Change (MCCC).

Last and foremost, I forever grateful for all my family members, my lovely daughter Heju (Hailey), my husband Yuze (Sam), my parents and my parents-in-law. My very special thank you goes to my husband for his love, encouragement and accompany. My deepest thank you goes to my parents for their support, dedication and unconditional love over my whole life.

Table of Contents

Chapter 1. Introduction	1
1.1 Flood Risk and Flood Forecasting Systems.....	1
1.2 Probability Theory and Predictive Uncertainty	4
1.3 Research Objectives and Thesis Outline.....	6
1.4 References.....	7
Chapter 2. Bayesian Flood Forecasting Methods: A Review.....	9
2.1 Abstract.....	10
2.2 Introduction.....	11
2.3 Overview of Advances in Bayesian Forecasting System (BFS).....	13
2.3.1 Fundamentals of BFS.....	13
2.3.2 Developments in BFS from 1999-2015	16
2.4 Application of Bayesian Flood Forecasting Methods.....	23
2.4.1 BFS Application in River Stage Forecasting.....	23
2.4.2 BFS Application in Real-time Flood Forecasting.....	29
2.4.3 Other Bayesian Method Application in Flood Estimation.....	34
2.4.4 Review and Discussion of the Predictive Uncertainty Assessment Approaches in Flood Forecasting	36
2.5 Conclusions and Future Work	45
2.6 Acknowledgments.....	46
2.7 References.....	47
Chapter 3. Assessing Hydrologic Uncertainty Processor Performance for Flood Forecasting in a Semiurban Watershed.....	58
3.1 Abstract.....	59
3.2 Introduction.....	60
3.3 Methodology.....	64
3.3.1 Hydrologic Uncertainty Processor (HUP)	64
3.3.1.1 Background of the Bayesian Processor.....	64

3.3.1.2 Normal Quantile Transform (NQT).....	67
3.3.1.3 Dependence Parameters	68
3.3.1.4 Prior Density and Prior Distribution	69
3.3.1.5 Posterior Density and Posterior Distribution	70
3.3.2 Rainfall-Runoff Models	71
3.3.2.1 HYMOD	71
3.3.2.2 GR4H	73
3.3.2.3 Models Setup	74
3.4 Study Area and Data	76
3.5 Results and Discussion	80
3.5.1 Rainfall-Runoff Model Calibration.....	80
3.5.2 Application of the Precipitation-dependent HUP	83
3.5.2.1 Estimation of Prior Distribution.....	83
3.5.2.2 Estimation of Likelihood Function	88
3.5.2.3 Estimation of Posterior Distribution	91
3.5.3 Discussion of Results and Performance Assessment.....	93
3.6 Conclusions.....	102
3.7 Acknowledgements.....	104
3.8 References.....	105
Chapter 4. Probabilistic Flood Forecasting Using Hydrologic Uncertainty	
Processor with Ensemble Weather Forecasts.....	112
4.1 Abstract	113
4.2 Introduction.....	113
4.3 Methodology	117
4.3.1 Multivariate Bias Correction Algorithm.....	118
4.3.2 Principle of Bayesian Ensemble Uncertainty Processor	121
4.4 Study Area and Data	125
4.5 Application and Discussion	127
4.5.1 Hydrologic Model	127

4.5.2 Calibration of Bayesian Processor	132
4.5.3 Bias Correction of Ensemble Weather Forecasts.....	136
4.5.4 Experiments and Comparisons	138
4.5.4.1 Scenario Results Comparisons.....	139
4.5.4.2 Comparison of Reliability.....	149
4.5.4.3 Forecast Hydrographs	151
4.6 Conclusions.....	154
4.7 Acknowledgments.....	156
4.8 References.....	157
Chapter 5. Assessing the Effects of Input and Model Type on Bayesian Ensemble Uncertainty Processor.....	163
5.1 Abstract.....	164
5.2 Introduction.....	165
5.3 Study Watershed and Data.....	168
5.4 Methodology	170
5.4.1 Methodology Overview	170
5.4.2 Hydrologic Model and Model Setup.....	171
5.4.3 Total Uncertainty Assessment	174
5.5 Results and Analysis	177
5.5.1 Comparison between Weather Predictions and Observations	177
5.5.2 SAC-SMA Calibration.....	180
5.5.3 Calibration of Hydrologic Uncertainty Processor.....	185
5.5.4 Scenario Results Comparisons.....	187
5.5.5 Example Forecast Hydrographs	193
5.6 Conclusions and Discussion	195
5.7 Acknowledgments.....	196
5.8 References.....	198
Chapter 6. Conclusions and Recommendations.....	203
6.1 Conclusions.....	203

6.1.1 Bayesian Flood Forecasting Methods	203
6.1.2 Hydrologic Uncertainty Processor with Different Models	204
6.1.3 Bayesian Ensemble Uncertainty Processor with Ensemble Weather Forecasts	205
6.1.4 Sensitivity of Bayesian Ensemble Uncertainty Processor	205
6.1.5 General Conclusions	206
6.2 Recommendations for Future Research	206
6.3 References	209

List of Figures

Figure 1-1 Number of floods in Canada (Data source: Canadian Disaster Database; the final bar only covers data for 2010-2016).....	3
Figure 1-2 Total estimated flood costs in Canada (Data source: Canadian Disaster Database).....	3
Figure 2-1 Structure of basic Bayesian Forecasting System	15
Figure 2-2 Developments of BFS method	17
Figure 3-1 Humber River Watershed in Southern Ontario, Canada.....	77
Figure 3-2 Model parameter estimation and uncertainty: (a) HYMOD optNVE; (b) HYMOD optKGE; (c) GR4H optNVE; and (d) GR4H optKGE. Box plots show the spread of optimized parameters for 20 calibration events.	81
Figure 3-3 Model performance evaluation. Event 4, 7, 14 and 23 are validation events, the remaining are calibration events.	83
Figure 3-4 Marginal distributions Γ_{nv} of observed discharges H_n for (a) $n = 1$; and (b) $n = 6$ conditional on precipitation indicator $v = 0, 1$	85
Figure 3-5 Dependence structure of the prior densities under $v = 1$ in the (a) transformed space; and (b) original space for lead time equal to 1, 3, 5 hours. h_n = actual river discharge in original space for every lead time n ; and w_n = actual river discharge in transformed normal space for every lead time n	87
Figure 3-6 Marginal distributions Λ_{nv} of simulated discharges S_n for (a) $n = 1$; and (b) $n = 6$ conditional on precipitation indicator $v = 0, 1$	89
Figure 3-7 Dependence structure of the likelihood function under $v = 1$ in the (a) transformed space; and (b) original space for lead time equal to 2, 4, 6 hours. s_n = model river discharge in original space for every lead time n ; and x_n = model river discharge in the transformed normal space for every lead time n	90
Figure 3-8 Hydrologic uncertainty bound generated by HUP for event 18: (a) 1 hour ahead forecast by HUP-HYMOD; (b) 6 hours ahead forecast by HUP-HYMOD; (c) 1 hour ahead forecast by HUP-GR4H; and (d) 6 hours ahead forecast by HUP-GR4H.	95

Figure 3-9 Hydrologic uncertainty bound generated by HUP for event 7: (a) 1 hour ahead forecast by HUP-HYMOD; (b) 6 hours ahead forecast by HUP-HYMOD; (c) 1 hour ahead forecast by HUP-GR4H; and (d) 6 hours ahead forecast by HUP-GR4H.....	96
Figure 3-10 Comparison of CRPS between HUP-HYMOD and HUP-GR4H for different lead times, and CRPS value for (a) HUP-HYMOD validation events; (b) HUP-GR4H validation events; (c) HUP-HYMOD calibration events; and (d) HUP-GR4H calibration events	99
Figure 3-11 Comparison of reliability plots between HUP-HYMOD and HUP-GR4H for different lead times, and reliability plots for (a) HUP-HYMOD calibration events; (b) HUP-GR4H calibration events; (c) HUP-HYMOD validation events; and (d) HUP-GR4H validation events.	102
Figure 4-1 Flowchart of methodology	118
Figure 4-2 Study area: Humber River Watershed.....	126
Figure 4-3 Calibration and validation for MAC-HBV	131
Figure 4-4 Marginal distribution of observed discharge for h18, h36, h54 and h72 conditional on precipitation indicator	133
Figure 4-5 Marginal distribution of simulated discharge for s18, s36, s54 and s72 conditional on precipitation indicator	134
Figure 4-6 Comparison of energy distance score	137
Figure 4-7 Scatter plot comparison for short-range forecasts between using GEPS data only and combining GEPS with HUP.....	140
Figure 4-8 Scatter plot comparison for medium-range forecasts between using GEPS data only and combining GEPS with HUP.....	141
Figure 4-9 Scatter plot comparison for short-range forecasts between using GEPS mean with HUP	142
Figure 4-10 Comparison of NSE for different scenarios: (a) short-range; (b) medium-range.....	145
Figure 4-11 Comparison of CRPS for different scenarios: (a) short-range; (b) medium-range.....	148

Figure 4-12 Comparison of reliability plot for different scenarios: (a) short-range; (b) medium-range	151
Figure 4-13 Ensemble forecasts and probabilistic forecasts: (a) using GEPS-raw as input; (b) using GEPS-BC as input (take June 27, 2015 for example)	153
Figure 5-1 Study watershed and model setup	169
Figure 5-2 Flowchart of methodology	171
Figure 5-3 Comparison between RDPS and observation for (a) precipitation and (b) temperature	179
Figure 5-4 Comparison between REPS mean and observation for (a) precipitation and (b) temperature	179
Figure 5-5 Observed vs. simulated hydrographs at basin outlet for lumped and semi-distributed SAC-SMA: (a) calibration; (b) validation	183
Figure 5-6 Estimated HUP parameters for (a) lumped SAC-SMA and (b) semi-distributed SAC-SMA.....	186
Figure 5-7 Nash-Sutcliffe coefficient (NSE) and Kling-Gupta efficiency (KGE) for different scenarios as a function of forecast lead time.....	188
Figure 5-8 Continuous ranked probability score (CRPS) and its reliability component for different scenarios for various forecast lead times: (a) all flows; (b) peak flows.....	191
Figure 5-9 Percent of peak flows lie within (a) the 50% uncertainty bound (25-75% quantile) and (b) the 90% uncertainty bound (5-95% quantile)	192
Figure 5-10 Forecast hydrograph for a high flow event issued on April 15, 2018 using Dressed RDPS+Semi-dist	194
Figure 5-11 Forecast hydrograph for a high flow event issued on April 16, 2018 using REPS+Semi-dist	194

List of Tables

Table 2-1 Summary of studies on BFS approaches applied for river stage forecasting (1999-2015).....	19
Table 2-2 Summary of studies on BFS approaches applied for real-time flood forecasting (1999-2015).....	21
Table 2-3 Comparison of the predictive uncertainty assessment approaches in flood forecasting (in chronological order).....	38
Table 3-1 Model parameters of HYMOD and GR4H	74
Table 3-2 Main characteristics of the selected flood events	79
Table 3-3 Sample statistics and marginal prior distributions of observed discharge	86
Table 3-4 Sample statistics and prior distributions of simulated discharge.....	89
Table 3-5 Calibrated dependence parameters of posterior distribution for HUP-HYMOD	92
Table 3-6 Calibrated dependence parameters of posterior distribution for HUP-GR4H..	92
Table 3-7 Comparison of NSE between HYMOD/GR4H and HUP-HYMOD/GR4H (median value)	97
Table 3-8 Comparison of mean CRPS under different conditions	100
Table 4-1 Optimized parameters of MAC-HBV and SNOW-17.....	129
Table 4-2 Dependence parameters of Hydrologic Uncertainty Processor.....	135
Table 4-3 Brief descriptions of the seven application scenarios.....	138
Table 5-1 Summary of SAC-SMA and SNOW-17 parameters	181
Table 5-2 Performances of SAC-SMA models	184

List of Abbreviations

BC	Bias Correction
BEF	Bayesian Ensemble Forecast
BEUP	Bayesian Ensemble Uncertainty Processor
BFS	Bayesian Forecasting System
BJP	Bayesian Joint Probability
BMA	Bayesian Model Averaging
BNN	Bayesian Neural Network
BNs	Bayesian Networks
BPEM	Bayesian Processor Of Ensemble Members
BPO	Bayesian Processor Of Output
CaSPAr	Canadian Surface Prediction Archive
CDD	Canadian Disaster Database
CDF	Cumulative Distribution Function
CHPS	Community Hydrologic Prediction System
COR	Correlation Coefficient
CRPS	Continuous Ranked Probability Score
DA	Data Assimilation
DBM	Data-Based Mechanistic
Delft-FEWS	Delft Flood Early Warning System
EBFS	Ensemble Bayesian Forecasting System

EBFSR	Ensemble Bayesian Forecasting System With Randomization
EFF	Ensemble Flood Forecasting
EFFS	European Flood Forecasting System
EMOS	Ensemble Model Output Statistics
EnKF	Ensemble Kalman Filter
ESP	Ensemble Streamflow Prediction
GA	Genetic Algorithms
GDPS	Global Deterministic Prediction System
GEM	Global Environmental Multiscale Model
GEPS	Global Ensemble Prediction System
GEV	Generalized Extreme Value
GLMPP	General Linear Model Post-Processor
GLUE	Generalised Likelihood Uncertainty Estimation
HBM	Hierarchical Bayesian Model
HBN	Hierarchical Bayesian Network
HMOS	Hydrologic Model Output Statistics
HUP	Hydrologic Uncertainty Processor
IBUNE	Integrated Bayesian Uncertainty Estimator
ICK	Indicator Co-Kriging
IEF	Input Ensemble Forecaster
INT	Integrator

IQR	Interquartile Range
IUP	Input Uncertainty Processor
KGE	Kling-Gupta Efficiency
LN	Log-Normal
MBC	Multivariate Bias Correction
MCP	Model Conditional Processor
MSW	Modified Shapiro-Wilk Test
NFFS	National Flood Forecasting System
N-pdf _t	N-Dimensional Probability Density Function Transform
NQT	Normal Quantile Transform
NSE	Nash-Sutcliffe Efficiency
NVE	Nash Volume Error
NWP	Numerical Weather Prediction
NWSRFS	National Weather Service River Forecast System
PD-HUP	Precipitation-Dependent Hydrologic Uncertainty Processor
PET	Potential Evapotranspiration
PF	Particle Filter
PFC	Peak Flow Criterion
PFF	Probabilistic Flood Forecast
PF-MCMC	Particle Filter-Markov Chain Monte Carlo
PF-SIR	Particle Filter-Sequential Importance Resampling

PI-HUP	Precipitation-Independent Hydrologic Uncertainty Processor
PQPF	Probabilistic Quantitative Precipitation Forecast
PRDF	Probabilistic River Discharge Forecast
PRSF	Probabilistic River Stage Forecast
PRVF	Probabilistic Runoff Volume Forecast
PSO	Particle Swarm Optimization
PSTF	Probabilistic Stage Transition Forecast
PUP	Precipitation Uncertainty Processor
QDM	Quantile Delta Mapping
QR	Quantile Regression
RDPS	Regional Deterministic Prediction System
REPS	Regional Ensemble Prediction System
RMSE	Root Mean Square Error
SB	Sub-basin
SD	Standard Deviation
SMA	Soil Moisture Accounting
SWE	Snow Water Equivalent
UH	Unit Hydrograph
UNEEC	Uncertainty Estimation Based On Local Errors And Clustering
UTC	Coordinated Universal Time
VE	Volume Error

List of Symbols

$A_{nv}, B_{nv}, D_{nv}, T_{nv}$	Parameters of the posterior distribution
$a_{nv}, b_{nv}, d_{nv}, \sigma_{nv}$	Parameters of the likelihood function
c_{nv}	Parameter of transition density
C_{nv}, t_{nv}	Parameters of the prior distribution
f_{nv}	Likelihood function
g_{nv}	Prior density of actual river discharge
G_{nv}	Prior distribution of actual river discharge
h_n	Actual river discharge (a realization)
H_n	Actual river discharge (random variable)
j	Ensemble member
l	Distribution type
n	Lead time
p	Conditional quantile
q	Standard normal density
Q	Standard normal distribution
Q^{-1}	Inverse of normal standard distribution
s_n	Model river discharge (a realization)
S_n	Model river discharge (random variable)
t	Time scale

v	Precipitation occurrence probability
V	Precipitation indicator
w_n	Actual river discharge in normal space (a realization)
W_n	Actual river discharge in normal space (random variable)
x_n	Model river discharge in normal space (a realization)
X_n	Model river discharge in normal space (random variable)
α	Scale parameter of the distribution
β	Shape parameter of the distribution
γ	Shift parameter of the distribution
γ_{nv}	Marginal density of actual river discharge
κ	Expected density of model river discharge
$\bar{\Lambda}_{nv}$	Marginal distribution of model river discharge
$\bar{\lambda}_{nv}$	Marginal density of model river discharge
ϕ_{nv}	Posterior density of actual river discharge
Ψ_{nv}	Predictive distribution of actual river discharge
ψ_{nv}	Predictive density of actual river discharge
Γ_{nv}	Marginal prior distribution of actual river discharge
Π_{nv}	Conditional output distribution of model river discharge
Φ_{nv}	Posterior distribution of actual river discharge

Declaration of Academic Achievement

This thesis was prepared in a sandwich style in accordance with the regulations provided by the School of Graduate Studies at McMaster University. It includes the published and submitted papers listed below:

Chapter 2: Bayesian Flood Forecasting Methods: A Review by S. Han and P. Coulibaly, Journal of Hydrology, 551, 340-351, doi: 10.1016/j.jhydrol.2017.06.004, 2017. (With permission from publisher)

Chapter 3: Assessing Hydrologic Uncertainty Processor Performance for Flood Forecasting in a Semiurban Watershed by S. Han, P. Coulibaly and D. Biondi, Journal of Hydrologic Engineering, 24(9), 05019025, doi: 10.1061/(ASCE)HE.1943-5584.0001828, 2019. (With permission from publisher)

Chapter 4: Probabilistic Flood Forecasting Using Hydrologic Uncertainty Processor with Ensemble Weather Forecasts by S. Han and P. Coulibaly, Journal of Hydrometeorology, 20(7), 1379-1398, doi: 10.1175/JHM-D-18-0251.1, 2019. (With permission from publisher)

Chapter 5: Assessing the Effects of Input and Model Type on Bayesian Ensemble Uncertainty Processor by S. Han and P. Coulibaly, Advances in Water Resources, under review, manuscript number ADWR_2019_465.

For Chapter 2, S. Han conducted the literature review on Bayesian flood forecasting methods under the supervision of Dr. P. Coulibaly. S. Han wrote the manuscript and Dr. P. Coulibaly reviewed and edited it, the paper was published in Journal of Hydrology in 2017. For Chapter 3, S. Han conducted the modelling and computational work with the supervision of Dr. P. Coulibaly and consultation of Dr. D. Biondi. Dr. D. Biondi provided guidance on programming for Hydrologic Uncertainty Processor. S. Han wrote the manuscript, Dr. P. Coulibaly and Dr. D. Biondi reviewed and edited it, and the paper was published in ASCE Journal of Hydrologic Engineering in 2019. For Chapter 4 and 5, S. Han conducted the modelling and computational work with the guidance and supervision of Dr. P. Coulibaly. S. Han wrote the manuscript and Dr. P. Coulibaly reviewed and edited it. Chapter 4 was published in Journal of Hydrometeorology in 2019, and chapter 5 was submitted to Advances in Water Resources. The work reported here was undertaken from September 2014 to June 2019.

Chapter 1. Introduction

1.1 Flood Risk and Flood Forecasting Systems

Over the past few decades, climate change has led to raised air temperature and intense precipitation events, and the more frequent precipitation events have in turn resulted in increased risk of flooding (Allen et al. 2014). In Canada, according to the Canadian Disaster Database (CDD), the number of floods has dramatically grown from under 10 in the 1930s to over 60 in the 2000s (Figure 1-1). There is also a rapid growth of the estimated flood costs since 1960s, and the total cost during 2010s reaches 7.18 billion dollars (Figure 1-2). Flood is among the top weather-related killers (Chin 2006), to reduce and alleviate the negative impacts of flooding, flood forecasting provides an essential tool to allow for mitigation action. Structural measures such as construction of flood control facilities usually require huge capital investment (Mays 2010), in many regions flood forecasting as a nonstructural measure is the only effective and affordable way for flood protection (WMO, 2006).

Various flood forecasting systems were developed worldwide. For example, the European Flood Forecasting System (EFFS) was developed for European countries in 2003 (De Roo et al. 2003). The National Flood Forecasting System (NFFS) has been used in England since 2005 and Wales since 2006 (Werner et al. 2009). The Scotland Flood Early Warning System (FEWS) was expanded into a national system in 2007 (Cranston and Tavendale 2012). The National Weather Service River Forecast System (NWSRFS) has been employed in United States for over 30 years, and was replaced by the

Community Hydrologic Prediction System (CHPS) in recent years (Roe et al. 2010). The NFFS and FEWS adopted the same open shell called Delft Flood Early Warning System (Delft-FEWS). Delft-FEWS can incorporate any forecasting model through a model adapter and combine a wide range of data (Werner et al. 2013).

Hydrologic models (or rainfall-runoff models) are embedded in the forecasting system; most of them are mathematical representation of the hydrologic cycle in the watershed, including hydrologic processes such as precipitation, infiltration, interception, evapotranspiration and runoff. Each model has its own unique structure and characteristic, which is defined by a set of model parameters. The main input to the hydrologic model is precipitation; others may include air temperature, soil moisture, topography, physical parameters and so on, and the model output is streamflow or discharge (Devia et al. 2015). Prior to the forecast, the model parameters should be calibrated using historical data. In the forecast mode, the model is fed by forecasted weather data and produces estimated streamflow or discharge. The model outputs for the future are used to make decisions on warnings of floods and watershed management (Jain et al. 2018).

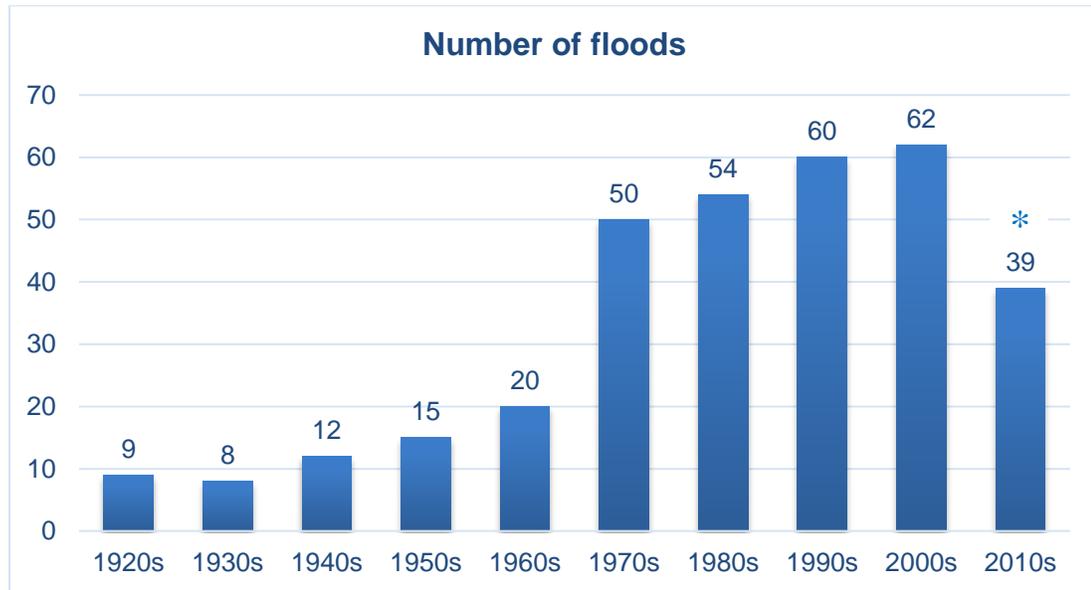


Figure 1-1 Number of floods in Canada (Data source: Canadian Disaster Database; the final bar only covers data for 2010-2016)

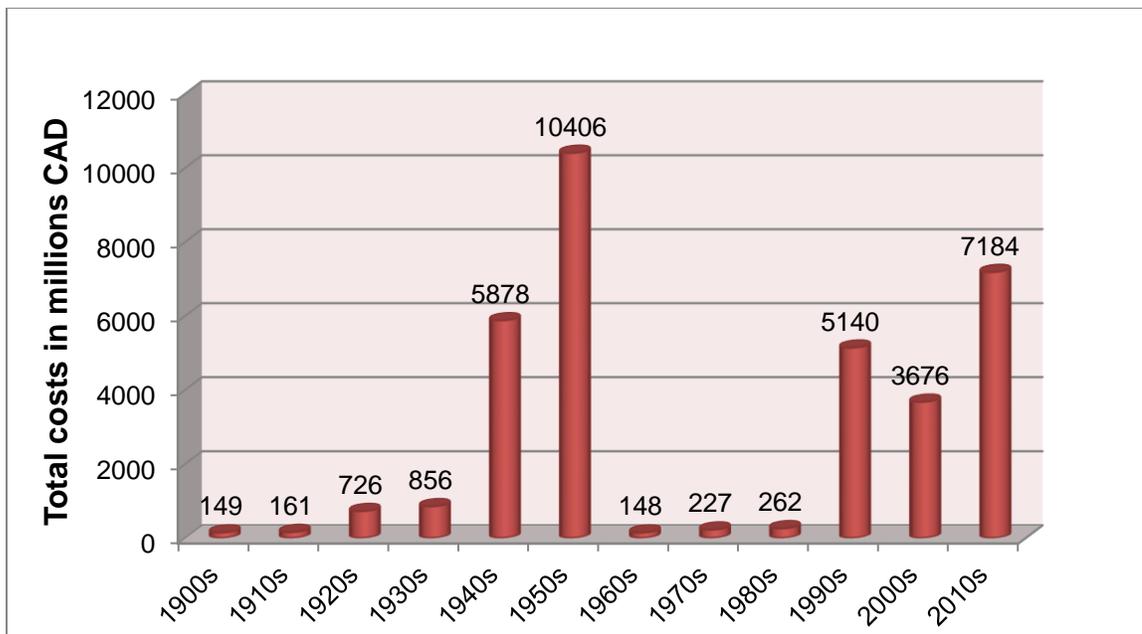


Figure 1-2 Total estimated flood costs in Canada (Data source: Canadian Disaster Database)

1.2 Probability Theory and Predictive Uncertainty

Hydrologic model could be used in a deterministic or probabilistic manner. Deterministic model attempts to provide an exact representation of the physical system and gives a single estimate of model response (Farmer and Vogel 2016). It ignores the uncertainty and has certain limitations. First, the deterministic representation is a simplified version of the real system and relies on limited knowledge, as the statistician George Box said, “All models are wrong, some are useful” (Box et al. 2005). Second, the reality for a particular point in time cannot be exactly or perfectly represented by the measurements due to the measurement errors. Third, most mathematical models cannot be solved exactly and only have an approximate solution (Reich and Cotter 2015). As such, deterministic forecast is giving way to probabilistic forecast, which considers uncertainty from various sources and assigns a probability to each of the different model outcomes.

Based on the probability theory, there are three different ways to derive probability (Leonard and Hsu 1999; Reich and Cotter 2015):

- (1) Replicate a large number of experiments under identical conditions and record the frequency of an event to occur, the probability of this event is the frequency of occurrence over the total number of outcomes.
- (2) Identify possible alternatives that are equally likely to occur, and assign each of the alternatives with equal probability.
- (3) Estimate the probability based on current knowledge of the system or previous experiment of a similar system, and then revise the prior probability to posterior probability when new information is available.

In this thesis, the Bayesian methods were used to develop probabilistic flood forecast, which belongs to the last category of probability. The prior distribution is first estimated based on all the received information at the onset of the forecast. After running the model, the new information from the model forecasts is used to update the prior distribution into posterior distribution, and everything we know about the future condition and associated uncertainty is summarized in this Bayesian posterior distribution or predictive distribution. Predictive uncertainty is defined as “the expression of a subjective assessment of the probability of occurrence of a future (real) event conditional upon all the knowledge available up to the present (the prior knowledge) and the information that can be acquired through a learning inferential process” (Rougier 2007). Although a variety of techniques about predictive uncertainty assessment has been developed, however, great challenges are still remaining. First, the uncertainties in flood forecast comes from various sources, it is difficult to accurately quantify all of them or even just the major uncertainties. Some proposed methods only allow to measure limited sources such as model parameter uncertainty or input data uncertainty. Second, predictive uncertainty is different from simulation uncertainty or forecast sensitivity. It requires a probability distribution for the future true value conditional on the model predictions, and cannot be fully represented by a distribution of model predictions based on observations or the sensitivity of the model predictions (Todini 2009). Third, there is an increasing need for the flexibility of the framework; it should be able to work with various hydrologic models and different types of data. Thus, further research is necessary to address these challenges.

1.3 Research Objectives and Thesis Outline

The research presented in this thesis focuses on the uncertainty assessment in flood forecasting through Bayesian methods, and the goal of this work is to enhance the accuracy and reliability of flood forecasting system and improve the assessment of predictive uncertainty associated with flood forecasts. To achieve the overall objective, four journal papers have been completed and are presented in chapter 2-5 of the thesis.

This thesis includes six chapters. Chapter 1 is an overview of the research background and context. Chapter 2 presents a comprehensive review of the Bayesian flood forecasting methods (1999-2016) and a summary of alternative predictive uncertainty assessment methods. Chapter 3 presents an application of the Hydrologic Uncertainty Processor (HUP) for probabilistic flood forecasting to quantify the hydrologic uncertainty, and assesses the performance of HUP combined with different hydrologic models. Chapter 4 extends the HUP into the Bayesian Ensemble Uncertainty Processor (BEUP) to quantify the dominant uncertainties, and integrates bias correction of ensemble weather forecasts with BEUP to enhance the predictive performance. Chapter 5 investigates the effects of different input data types and different hydrologic model types on the performance of BEUP. Chapter 6 is a summary of major conclusions and recommendations for future research.

1.4 References

- Allen, M.R., Barros, V.R., Broome, J., Cramer, W., Christ, R., Church, J.A., Clarke, L., Dahe, Q., Dasgupta, P., Dubash, N.K., Edenhofer, O., 2014. IPCC fifth assessment synthesis report-climate change 2014 synthesis report.
- Box, G.E., Hunter, J.S., Hunter, W.G., 2005. Statistics for Experimenters. In Wiley Series in Probability and Statistics. Wiley Hoboken, NJ, USA.
- Chin, D.A., 2006. Water Resources Engineering. Pearson Prentice Hall.
- Cranston, M.D., Tavendale, A.C.W., 2012. Advances in operational flood forecasting in Scotland. Proceedings of the Institution of Civil Engineers - Water Management 165(2), 79–87. doi:10.1680/wama.2012.165.2.79
- De Roo, A.P.J., Gouweleeuw, B., Thielen, J., Bartholmes, J., Bongioannini-Cerlini, P., Todini, E., Bates, P.D., Horritt, M., Hunter, N., Beven, K., Pappenberger, F., Heise, E., Rivin, G., Hils, M., Hollingsworth, A., Holst, B., Kwadijk, J., Reggiani, P., Dijk, M. Van, Sattler, K., Sprokkereef, E., 2003. Development of a european flood forecasting system. International Journal of River Basin Management 1(1), 49–59. doi:10.1080/15715124.2003.9635192
- Devia, G.K., Ganasri, B.P., Dwarakish, G.S., 2015. A Review on Hydrological Models. Aquatic Procedia 4, 1001–1007. doi:10.1016/j.aqpro.2015.02.126
- Farmer, W.H., Vogel, R.M., 2016. On the deterministic and stochastic use of hydrologic models. Water Resources Research 52(7), 5619–5633. doi:10.1002/2016WR019129
- Jain, S.K., Mani, P., Jain, S.K., Prakash, P., Singh, V.P., Tullos, D., Kumar, S., Agarwal, S.P., Dimri, A.P., 2018. A Brief review of flood forecasting techniques and their applications. International Journal of River Basin Management 16(3), 329–344. doi:10.1080/15715124.2017.1411920
- Leonard, T., Hsu, J.S.J., 1999. Bayesian Methods: An Analysis for Statisticians and Interdisciplinary Researchers. Cambridge University Press.

- Mays, L.W., 2010. *Water Resources Engineering*. John Wiley & Sons.
- Reich, S., Cotter, C., 2015. *Probabilistic Forecasting and Bayesian Data Assimilation*. Cambridge University Press.
- Roe, J., Dietz, C., Restrepo, P., Halquist, J., Hartman, R., Horwood, R., Olsen, B., Opitz, H., Shedd, R., Welles., E., 2010. NOAA’s Community Hydrologic Prediction System. Proceedings from the 4th Federal Interagency Hydrologic Modeling Conference.
- Rougier, J., 2007. Probabilistic inference for future climate using an ensemble of climate model evaluations. *Climatic Change* 81(3-4), 247–264. doi:10.1007/s10584-006-9156-9
- Todini, E., 2009. Uncertainties in Environmental Modelling and Consequences for Policy Making, chap. Predictive Uncertainty Assessment in Real Time Flood Forecasting. doi:10.1007/978-90-481-2636-1
- Werner, M., Cranston, M., Harrison, T., Whitfiel, D., Schellekens, J., 2009. Recent developments in operational flood forecasting in England, Wales and Scotland. *Meteorological Applications* 16(1), 13–22.
- Werner, M., Schellekens, J., Gijsbers, P., van Dijk, M., van den Akker, O., Heynert, K., 2013. The Delft-FEWS flow forecasting system. *Environmental Modelling and Software* 40, 65–77. doi:10.1016/j.envsoft.2012.07.010
- World Meteorological Organization, 2006. *WMO flood Forecasting Initiative: Final Report of the Synthesis Conference on Improved Meteorological and Hydrological Forecasting*.

Chapter 2. Bayesian Flood Forecasting Methods: A Review

Summary of Paper 1: Han, S. and Coulibaly, P. (2017). Bayesian Flood Forecasting Methods: A Review. *Journal of Hydrology*, 551, 340-351.

This research is an extensive literature review on Bayesian forecasting methods used in flood forecasting, and the focus is on research works from 1999 until 2016. The main topics include:

- Overview of fundamentals of Bayesian forecasting system (BFS)
- Recent advances in BFS
- Literature review on BFS application
- Advantages and limitations of Bayesian forecasting methods
- Pros and cons of alternative predictive uncertainty assessment methods
- Future research direction in Bayesian flood forecasting

Key results of this research include:

- Bayesian method can provide an effective and advanced approach for probabilistic flood forecasting;
- Bayesian forecasting system is able to consider all major sources of uncertainty and produce more accurate and reliable flood forecasts;
- Some emerging Bayesian forecasting methods could overcome certain limitations and reduce predictive uncertainty.

2.1 Abstract

Over the past few decades, floods have been seen as one of the most common and largely distributed natural disasters in the world. If floods could be accurately forecasted in advance, then their negative impacts could be greatly minimized. It is widely recognized that quantification and reduction of uncertainty associated with the hydrologic forecast is of great importance for flood estimation and rational decision making. Bayesian forecasting system (BFS) offers an ideal theoretic framework for uncertainty quantification that can be developed for probabilistic flood forecasting via any deterministic hydrologic model. It provides suitable theoretical structure, empirically validated models and reasonable analytic-numerical computation method, and can be developed into various Bayesian forecasting approaches. This paper presents a comprehensive review on Bayesian forecasting approaches applied in flood forecasting from 1999 till now. The review starts with an overview of fundamentals of BFS and recent advances in BFS, followed with BFS application in river stage forecasting and real-time flood forecasting, then move to a critical analysis by evaluating advantages and limitations of Bayesian forecasting methods and other predictive uncertainty assessment approaches in flood forecasting, and finally discusses the future research direction in Bayesian flood forecasting.

Results show that the Bayesian flood forecasting approach is an effective and advanced way for flood estimation, it considers all sources of uncertainties and produces a predictive distribution of the river stage, river discharge or runoff, thus gives more accurate and reliable flood forecasts. Some emerging Bayesian forecasting methods (e.g.

ensemble Bayesian forecasting system, Bayesian multi-model combination) were shown to overcome limitations of single model or fixed model weight and effectively reduce predictive uncertainty. In recent years, various Bayesian flood forecasting approaches have been developed and widely applied, but there is still room for improvements. Future research in the context of Bayesian flood forecasting should be on assimilation of various sources of newly available information and improvement of predictive performance assessment methods.

Key words: Probabilistic flood forecast; Bayesian forecasting system; Uncertainty quantification; Predictive distribution; Predictive density function; Probability

2.2 Introduction

According to the fifth IPCC (Intergovernmental Panel on Climate Change) climate assessment report, extreme weather events were increased during the 21st century due to climate change (IPCC, 2014). Accelerated hydrological cycle leads to increased frequency of intense precipitation events and enhanced fluctuation in streamflow to some extent, which in turn results in more frequent floods and droughts (Reggiani & Weerts, 2008). Floods were seen as one of the most common and largely distributed natural disasters in the world, and caused significant damage to life and property over the past few decades (Balica et al., 2013). So there is an increasing need for flood control measures, both structural and non-structural. Among them, flood forecasting and estimation is an effective method that allows time for mitigating action. If floods could be predicted accurately in advance, then their negative impacts could be minimized.

Hydrologic models used for forecasting river stage, river discharge or runoff volumes are usually deterministic, and forecast results are normally exhibited as time series of estimates. However, their estimates are not free of error and contain limited amount of information though operationally simple. From the viewpoint of a decision maker who must make a rational flood mitigation decision based on the information provided by a hydrologic forecaster, a point estimate of the predictand may be insufficient (Krzysztofowicz, 1999; Krzysztofowicz, 2001b). In order to provide more valuable information, the uncertainty associated with the predictand needs to be quantified in terms of probability distribution and degree of certitude, decisions should be made according to this probability distribution instead of just a single value of estimate (Krzysztofowicz, 1983). The growing demand for forecast products and the increasing capability to quantify predictive uncertainty give an impetus for research into probabilistic forecasting of hydrologic variates.

It is widely recognized that proper uncertainty quantification associated with a hydrologic forecast is of great importance for both operational application and scientific research (Biondi et al., 2010). In recent years many approaches have been developed for uncertainty quantification and reduction, but there are still challenges as uncertainties could arise from a variety of sources (Biondi & De Luca, 2012). Among the methodologies well suited for flood forecasting process, Bayesian forecasting system (BFS) provides an ideal theoretic framework that can be developed for different purposes using probabilistic forecast of inputs via any deterministic hydrologic model. It considers

and quantifies all sources of uncertainties which gives more reliable estimation (Krzysztofowicz, 1999).

This paper provides a comprehensive review on Bayesian flood forecasting approaches and discusses the research direction within this field. BFS can be developed for diversified probabilistic forecasting systems suitable for various purposes. Here the paper only focuses on the review of BFS approaches used for flood forecasting from the year of 1999 until now.

2.3 Overview of Advances in Bayesian Forecasting System (BFS)

2.3.1 Fundamentals of BFS

Bayesian forecasting system is a robust theoretical framework that can be used for probabilistic forecast through deterministic hydrologic model of any complexity (Krzysztofowicz, 1999). In the domain of flood forecasting, BFS could be developed to produce probabilistic river stage forecast (PRSF), probabilistic river discharge forecast (PRDF) or probabilistic runoff volume forecast (PRVF) at any time step.

In the BFS, the total uncertainty associated with the hydrologic forecast is broken down into two sources: precipitation uncertainty and hydrologic uncertainty. Precipitation uncertainty is related to the future average precipitation amount. Hydrologic uncertainty is the aggregate of all other uncertainties. These sources include: imperfections of the hydrologic model (e.g. model structure, model parameters), measurement errors of physical variables (e.g. temperature, streamflow, and precipitation), incorrect temporal and spatial downscaling of the total precipitation (e.g. deterministic forecast of spatial

disaggregation of total precipitation amount into subbasins, deterministic forecast of subperiods' precipitation amount from temporal disaggregation of total amount) and so on. In the first place, precipitation uncertainty and hydrologic uncertainty are quantified respectively, and then integrated together to produce a probabilistic forecast (Krzysztofowicz, 1999; Krzysztofowicz & Kelly, 2000a; Krzysztofowicz & Herr, 2001; Krzysztofowicz, 2002). It is technically impractical and perhaps unnecessary to specifically quantify every source of uncertainty. Usually only a few sources dominate the contribution to the total uncertainty, therefore a compromise between the exactness and practicality can be reached by limiting the decomposition into the dominant uncertainties and all other uncertainties in the aggregate (Krzysztofowicz, 1999; Krzysztofowicz & Kelly, 2000a).

The decomposition method of uncertainties leads to the fundamental structure of BFS shown in Figure 2-1. There are two processors attach to the hydrologic model. One processor propagates the precipitation uncertainty into the output uncertainty under the assumption of nonexistence of hydrologic uncertainty. Another processor maps the hydrologic uncertainty into the output uncertainty based on the assumption that no precipitation uncertainty exists within this process. The two uncertainties are then incorporated together to generate a probabilistic forecast and this incorporation is nonmonotonic and nonadditive. Therefore, the BFS consists of three interrelated structural components: (1) Input uncertainty processor (IUP), the dominant source of input uncertainty is future precipitation, thus this processor is also called precipitation uncertainty processor (PUP), (2) Hydrologic uncertainty processor (HUP), (3) Integrator

(INT). If the hydrologic predictand is river stage, then for PUP, the distribution of precipitation amount and response function induces the distribution of model river stage. For HUP, given the marginal prior distribution of actual river stage, prior dependence parameters, likelihood dependence parameters and marginal initial distribution of model river stage, the posterior distribution and posterior density can be derived by Bayesian revision process. Based on the output of PUP and HUP, the task of INT is to produce the predictive distribution and predictive density (Krzysztofowicz, 1999; Krzysztofowicz, 2002).

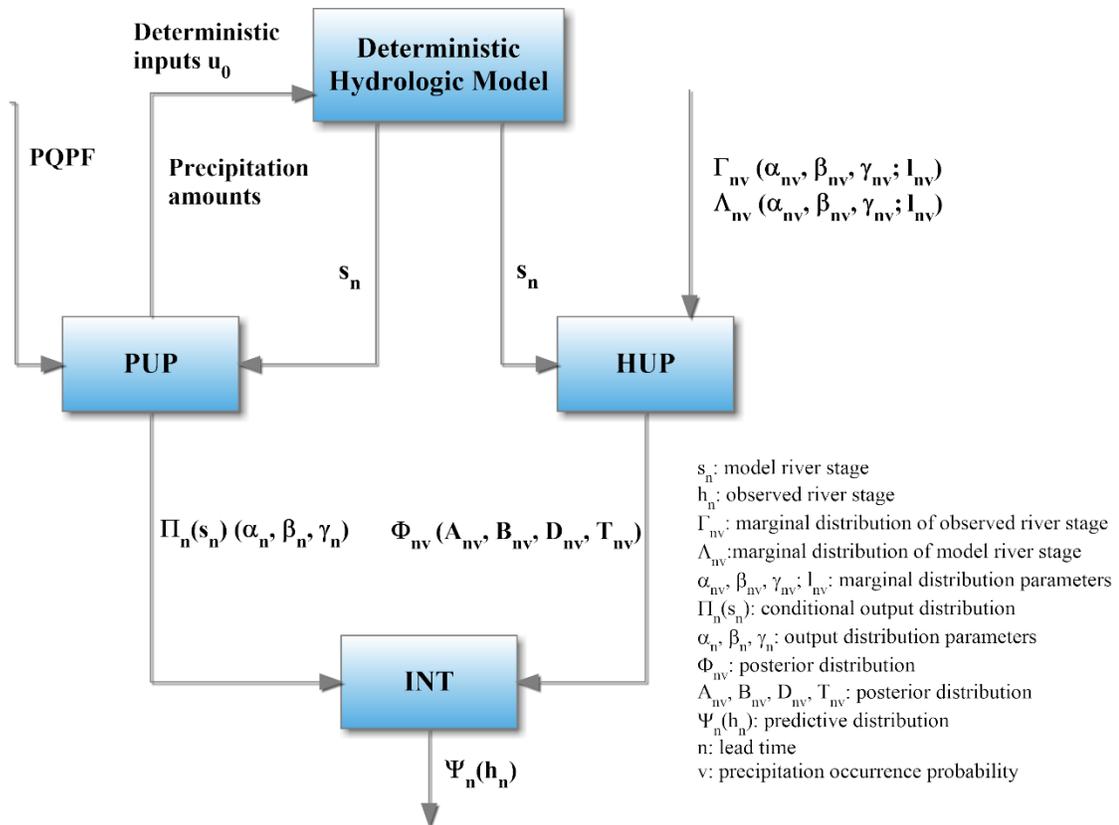


Figure 2-1 Structure of basic Bayesian Forecasting System

2.3.2 Developments in BFS from 1999-2015

Since Krzysztofowicz introduced BFS in 1999, it has been gaining in popularity worldwide. Then two types of BFS were formed, one is to obtain a probabilistic river stage forecast (PRSF) on the basis of probabilistic quantitative precipitation forecast (PQPF), another one is to generate probabilistic stage transition forecast (PSTF) in accordance with PQPF. These two types of BFS rest on the same theoretic structure, but the second BFS provide more information such as river stage process evolution besides each river stage. There are two kinds of HUP within the BFS: precipitation-independent hydrologic uncertainty processor (PI-HUP) and precipitation-dependent hydrologic uncertainty processor (PD-HUP). PD-HUP is composed of two branches, one is under the condition of precipitation occurrence, and another one is under nonoccurrence of precipitation. Hence provides more reliable information for prior distribution and likelihood function. Later on, with the introduction of ensemble weather prediction, Bayesian ensemble forecast (BEF) for flood was developed. The ensemble flood forecast can be generated by Monte Carlo simulation and is called ensemble Bayesian forecasting system (EBFS). Based on EBFS, a modified method named ensemble Bayesian forecasting system with randomization (EBFSR) was proposed. EBFSR is a more operationally feasible and computational efficient approach (Herr & Krzysztofowicz, 2015a).

Based on the PRSF or the PSTF, probabilistic flood forecast (PFF) can be obtained either approximately or exactly. In real-time flood forecasting, BEF could also be generated by Bayesian ensemble uncertainty processor (BEUP). In order to address model structure

uncertainty, an integrated Bayesian uncertainty estimator (IBUNE) was proposed then to combine multi-model prediction. Recently a sequential Bayesian multi-model combination method was applied and showed its ability to overcome the limitations that exist in data assimilation (DA) and Bayesian model averaging (BMA) methods. A graphical presentation of the developments in BFS method can be found in Figure 2-2. These Bayesian approaches are efficient tools for flood forecasting with uncertainty estimate.

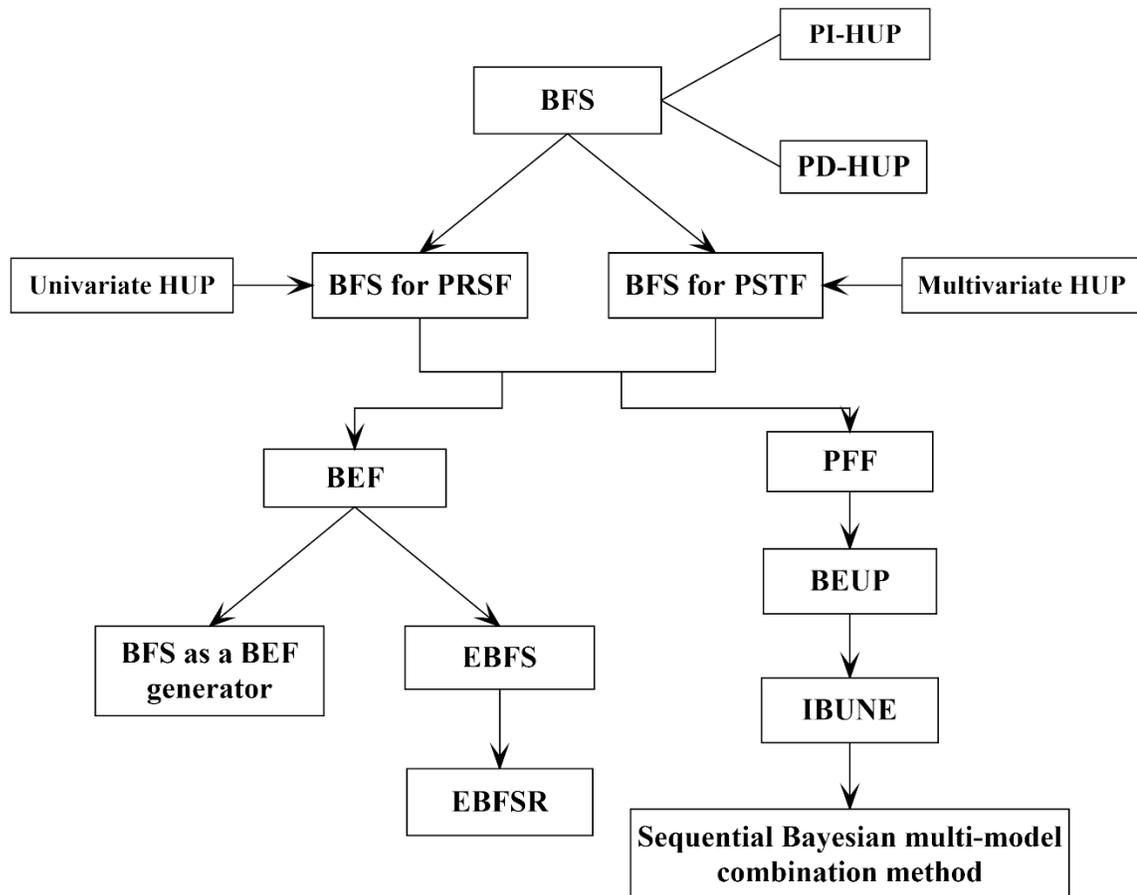


Figure 2-2 Developments of BFS method

There are over 30 published papers in the application and development of Bayesian flood forecasting from 1999 until now. A summary of the BFS applied in river stage forecasting is shown in Table 2-1, and the BFS applications in real-time flood forecasting are summarized in Table 2-2. In the following section, the characteristics of these Bayesian flood forecasting methods and the major findings will be further discussed.

Table 2-1 Summary of studies on BFS approaches applied for river stage forecasting (1999-2015)

References	Key Objectives	Methods	Major Findings/ Contributions
Krzysztofowicz, R. (1999)	Introduce BFS theory	Interpreted the principles of Bayesian predictive inference	5 general properties of BFS: <ul style="list-style-type: none"> ▪ Decomposition of total uncertainty ▪ Predictive result: predictive distribution ▪ Bayesian inference ▪ Self-calibration ▪ Coherence property
Kelly, K. S., & Krzysztofowicz, R. (2000)	Describe theory of PUP	Implemented PUP under different timing pattern, probability and initial condition	<ul style="list-style-type: none"> ▪ Response function: two-piece power function ▪ Conditional distribution: two-piece Weibull distribution ▪ PQPF: deterministic equivalence principle
Krzysztofowicz, R., & Kelly, K. S. (2000)	Describe theory of HUP	Parametric HUP based on Bayesian formulation	<ul style="list-style-type: none"> ▪ Significance of hydrologic uncertainty ▪ Nonlinear and heteroscedastic dependence structure ▪ Properties of Meta-Gaussian model
Krzysztofowicz, R., & Herr, H. D. (2001)	Quantify the hydrologic uncertainty by PD-HUP	PD-HUP based on Meta-Gaussian model	<ul style="list-style-type: none"> ▪ Informativeness of PD-HUP ▪ Nonlinear and heteroscedastic dependence structure ▪ Non-stationarity of prior density and likelihood function
Krzysztofowicz, R. (2001)	Describe theory of INT	INT with PD-HUP	<ul style="list-style-type: none"> ▪ Predictive density: bimodal and asymmetric ▪ Superiority of the PD-HUP over PI-HUP, Monte Carlo simulation and ensemble techniques
Krzysztofowicz, R. (2002)	Produce a short-term PRSF based on PQPF	Presented the synthesis of BFS for PRSF	<ul style="list-style-type: none"> ▪ Theoretically derived structure ▪ Empirically validated models ▪ Parsimonious analytic-numerical method

Maranzano, C. J., & Krzysztofowicz, R. (2004)	Seek the simplest likelihood and prior dependence structures	BFS for PSTF based on multivariate HUP	<ul style="list-style-type: none"> ▪ HUP: two-branch structure ▪ Prior distribution: first-order Markov dependence structure ▪ Likelihood function: second-order conditional dependence structure
Krzysztofowicz, R., & Maranzano, C. J. (2004a)	Implement the multivariate HUP for PSTF	Multivariate HUP based on Meta-Gaussian model	<ul style="list-style-type: none"> ▪ Non-stationarity of prior distribution and likelihood function ▪ Any form of marginal distributions ▪ Nonlinear and heteroscedastic dependence structure
Krzysztofowicz, R., & Maranzano, C. J. (2004b)	Produce a short-term PSTF based on PQPF	Presented the synthesis of BFS for PSTF	<ul style="list-style-type: none"> ▪ Comparison between BFS for PRSF and BFS for PSTF ▪ The PSTF is computationally simple but difficult to use and display compared with the Markovian PSTF
Herr, H. D., & Krzysztofowicz, R. (2010)	Determine the smallest ensemble size for BEF	Used BFS as a generator of BEF	<ul style="list-style-type: none"> ▪ PRSF and PFF: several hundreds ▪ PSTF: several thousands
Herr, H. D., & Krzysztofowicz, R. (2015)	Develop EBFSR based on EBFS	<ul style="list-style-type: none"> ▪ EBFS using Monte Carlo simulation ▪ EBFSR ▪ Compared EBFS and EBFSR 	EBFSR was more computationally efficient and operationally feasible than EBFS

Table 2-2 Summary of studies on BFS approaches applied for real-time flood forecasting (1999-2015)

References	Key Objectives	Methods	Major Findings/ Contributions
Krzysztofowicz, R. (2002)	Obtain bounds on and approximations to the PFF from PRSF	<ul style="list-style-type: none"> ▪ Bounds on PFF: Fréchet bounds and tighter bounds ▪ Approximations to PFF: DLI and RLI 	<ul style="list-style-type: none"> ▪ Construction methods are simple to obtain bounds on and estimators of the PFF ▪ Parameter is invariant with forecast point, hydrologic season and time step
Ajami, N. K., et al. (2007)	Develop IBUNE to confront uncertainty in input errors, model parameters and structure	IBUNE with SAC-SMA, HYMOD and SWB	<ul style="list-style-type: none"> ▪ Useful and applicable technique ▪ Improved model prediction uncertainty bounds
Reggiani, P., & Weerts, A. H. (2008a)	Implement HUP for the operational flood forecasting system	<ul style="list-style-type: none"> ▪ HBV for hydrologic response ▪ Embedded HUP in Delft-FEWS 	<ul style="list-style-type: none"> ▪ Inclusion of water level observations improved accuracy ▪ Inadequacy of linear regression
Reggiani, P., & Weerts, A. H. (2008b)	Produce PQPF by a modified BPO	<ul style="list-style-type: none"> ▪ HIRLAM for rainfall forecasting ▪ BPO for weather model output 	<ul style="list-style-type: none"> ▪ BPO has larger skill in predicting event occurrence or nonoccurrence rather than event depth ▪ Uncertainty due to spatial and temporal variability of precipitation cannot be ignored
Todini, E. (2008)	Introduce MCP to assess predictive uncertainty	<ul style="list-style-type: none"> ▪ MCP: simple lag-1 Markov model, phase-space approach ▪ Compared MCP with HUP and BMA 	<ul style="list-style-type: none"> ▪ Reconciled physically based with data driven models ▪ Reduced predictive uncertainty
Reggiani, P., et al. (2009)	Implement HUP for an ensemble streamflow forecast	<ul style="list-style-type: none"> ▪ HBV model ▪ BEUP for uncertainty assessment 	<ul style="list-style-type: none"> ▪ BEUP was an effective uncertainty assessment tool ▪ Establishment of a cost function to express the economic value
Biondi, D., et al. (2009)	Quantify the predictive uncertainty on river discharge	<ul style="list-style-type: none"> ▪ Used RISE model to simulate hydrologic response ▪ Applied PD-HUP for BFS 	<ul style="list-style-type: none"> ▪ Hydrologic uncertainty grew with increased discharge, high precipitation and increased lead time

Biondi, D., et al. (2009)	Assess the predictive uncertainty on water discharge	<ul style="list-style-type: none"> ▪ RISE model ▪ BFS based on PD-HUP 	<ul style="list-style-type: none"> ▪ Hydrologic uncertainty grew with increased discharge and lead time ▪ Inadequacy of liner model
Biondi, D., et al. (2010)	Assess uncertainty through PD-HUP	<ul style="list-style-type: none"> ▪ RISE model ▪ PD-HUP for real-time forecasting 	<ul style="list-style-type: none"> ▪ Hydrologic uncertainty grew with increased discharge, high precipitation and increased lead time
Biondi, D., & De Luca, D. L. (2012)	Apply BFS for real-time forecasting	<ul style="list-style-type: none"> ▪ PRAISE model for rainfall forecasting ▪ RISE model to simulate hydrological response ▪ BFS based on PD-HUP 	<ul style="list-style-type: none"> ▪ Crucial role of HUP ▪ Importance to develop predictive distribution ▪ Necessity for PD-BFS
Biondi, D., & De Luca, D. L. (2013)	Assess the performance of BFS	<ul style="list-style-type: none"> ▪ PRAISE model ▪ RISE model ▪ Assessed the BFS performance under UD, PD, HD and TD 	<ul style="list-style-type: none"> ▪ TD presented the best predictive results ▪ Importance of multifaceted view of assessment
Krzysztofowicz, R. (2014)	Derive the exact PFF from PSTF	<ul style="list-style-type: none"> ▪ Exceedance functions ▪ Isoprobability time series ▪ Distribution of time to flooding 	<ul style="list-style-type: none"> ▪ RLI was accurate and simple for real-time flood forecasting
DeChant, C. M., & Moradkhani, H. (2014)	Address model and initial condition uncertainty	<ul style="list-style-type: none"> ▪ Combine ESP with ensemble data assimilation ▪ Combine ESP with Sequential Bayesian Combination 	<ul style="list-style-type: none"> ▪ ESP combined with DA and SBC increased the reliability

2.4 Application of Bayesian Flood Forecasting Methods

2.4.1 BFS Application in River Stage Forecasting

Over the past few decades, flood forecasting has received considerable attention. In order to fully explore flood forecast uncertainty and improve forecast accuracy, Krzysztofowicz introduced a Bayesian Forecasting System (Krzysztofowicz, 1999). He interpreted the basic principles of Bayesian predictive inference and constructed numerical examples to show the quantification and integration of the uncertainties in a BFS. He indicated that BFS has five general properties: (1) The BFS decomposes the total uncertainty into input uncertainty and hydrologic uncertainty. Each of them is quantified respectively and then integrated into the probabilistic forecast. (2) The predictive result is shown as predictive density function. (3) The predictive density is obtained by revising prior density based on available information. (4) It owns a property of self-calibration. Suppose the probabilistic forecast of inputs is well calibrated, so does the BFS. (5) It guarantees a coherence property which guards against poor forecast. If the informativeness of the hydrologic model production is lower than the prior density, the predictive density will automatically converge to the state of prior density.

The first type of the BFS was to produce a short-term PRSF, PQPF was input into a deterministic hydrologic model that simulate the hydrologic response of watershed to precipitation (Kelly & Krzysztofowicz, 2000; Krzysztofowicz & Herr, 2001; Krzysztofowicz & Kelly, 2000a; Krzysztofowicz, 2001a, 2002). Subsequent papers detailed the principle and the formulation of each component separately. The study area throughout the series of case studies was Eldred, Pennsylvania, situated in the upstream

of the Allegheny River which covers a drainage area of 1430 km². The data used was a single input of average basin precipitation with 6-hour time interval. As the input to PUP, PQPF consists of two parts: (1) the probabilistic forecast of the total precipitation quantity, (2) the deterministic forecast of the temporal disaggregation. Forecasting the disaggregation deterministically can get almost the same model river stage distribution as forecasting probabilistically, this finding is called deterministic equivalence principle. For PUP, the output distribution can be represented by a five-parameter two-piece Weibull distribution and the response function can be represented by a two-piece power function. The PQPF scheme performed well for PRSF (Kelly & Krzysztofowicz, 2000). For HUP, it was a parametric model developed through Bayesian formulation under a family of meta-Gaussian distribution. The hydrologic uncertainty was verified as significant to the total uncertainty and cannot be ignored. The prior density and likelihood function was found nonstationary and a 1st order Markov model of river stage process seemed to be suitable for modeling the prior density. It also revealed that the dependence structure among the actual river stage was nonlinear and heteroscedastic, and the dependence structure among the model river stage and actual river stage was nonlinear and heteroscedastic as well (Krzysztofowicz & Kelly, 2000). As for the INT, it output a numerical representation of predictive distribution and predictive density of actual river stage, and the predictive density could be bimodal and asymmetric (Krzysztofowicz, 2001b, 2014). The overall BFS offered suitable theoretical structure, empirically validated models and reasonable analytic-numerical computation method. For a blueprint

of BFS operational implementation, the reader is referred to Krzysztofowicz (2002a, 2014a).

Suppose the information contained in PQPF was inadequate to predict the river stage on a special occasion, then the predictive distribution is the same as the prior distribution. If the PQPF was perfect, the posterior distribution could serve as a representative of the predictive distribution. Similarly, if the hydrologic model didn't have enough predictive capability, then the posterior distribution converges to the prior distribution. If the hydrologic model was perfect, then the output distribution represents the predictive distribution (Krzysztofowicz, 2001a).

There are two versions of HUP: PI-HUP and PD-HUP. PD-HUP has two branches that are conditional on the precipitation occurrence. Thus besides the two branches of posterior distribution from HUP, the INT also needs the conditional probabilities that combine them. The PD-HUP proved to provide more information for both prior distribution and likelihood function than the PI-HUP. Because large hydrologic uncertainty will appear when precipitation occurs, the posterior distribution under precipitation occurrence or nonoccurrence are distinct. The BFS with PD-HUP was shown superior over the BFS with PI-HUP, Monte Carlo simulation and ensemble techniques. Monte Carlo simulation and ensemble forecasting are only techniques for executing PUP to estimate output distribution, without HUP and INT, they don't yield probabilistic forecasts (Krzysztofowicz & Herr, 2001; Krzysztofowicz, 2001a).

The second type of the analytic-numerical BFS was proposed to produce a short-term PSTF with PQPF input into a deterministic hydrologic model that simulates the

hydrologic response of watershed to precipitation (Krzysztofowicz & Maranzano, 2004a, 2004b; Maranzano & Krzysztofowicz, 2004). Unlike PRSF that demonstrates sequential predictive n-step transition density functions, PSTF demonstrates a finite sequence of infinite families of predictive one-step transition density functions. The PSTF quantifies the total uncertainty about the river stage process evolution in time besides the uncertainty of each river stage. To produce a PRSF, the first BFS requires a univariate HUP that generates a family of posterior distributions of actual river stage in order to quantify hydrologic uncertainty about the n-step transition. To produce a PSTF, the second BFS requires a multivariate HUP that produces a family of posterior joint distribution of actual river stage process conditional on the model river stage process to quantify hydrologic uncertainty associated with the one-step transition. The posterior joint distribution could be decomposed into posterior one-step transition distributions, each of them can be obtained from a prior distribution and a likelihood function through Bayes theorem. The first BFS and the second BFS are based on the same PQPF input and theoretic structure, meet same design requirement and share many assumptions, but the BFS for PSTF requires more numerical calculation and a more general HUP.

Some studies were performed to seek the simplest mathematical structure which is sufficient to capture the empirical dependence structures of prior density functions and likelihood functions and to implement the general multivariate HUP for short-term PSTF. The hydrologic models selected for application in the four forecast points in USA were lumped antecedent precipitation index (API)-based hydrologic model, lumped Sacramento catchment model, and semi-distributed Sacramento catchment model,

respectively (Krzysztofowicz & Maranzano, 2004b; Maranzano & Krzysztofowicz, 2004). The main conclusions are shown below. (1) For both the BFS-PSTF and BFS-PRSF system, the HUP must consist of two branches, each conditional on the occurrence and non-occurrence of precipitation. The prior distributions and the likelihood functions in these two branches can be non-stationary. (2) In the HUP-PSTF system, for the prior one-step transition distributions, Markov dependence structure of order one seemed adequate. While for the likelihood functions, conditional dependence structure of order two can meet the requirements. (3) With respect to the HUP-PRSF system, for the prior n-step transition distributions, Markov dependence structure of order one was adequate. And for the likelihood functions, conditional dependence structure of order one was adequate. (4) The multivariate HUP tolerates all forms of marginal distribution, and it allows for an asymmetric and bimodal conditional density function and a heteroscedastic and nonlinear dependence structure. (5) The likelihood parameters have no relation between the level of hydrologic uncertainty and the study area, the basin size, the spatial disaggregation of precipitation, the hydrologic model type and the storm type.

With respect to forecast output, comparison between the PSTF and the Markovian PSTF leads to the conclusion that the PSTF is computationally simple but difficult to use and display. Because it specifies dependence of order n to the predictive one-step transition density functions. Conversely, the Markovian PSTF is computationally complex but simple to use and display because it specifies dependence of order one to the predictive one-step Markov transition density functions (Krzysztofowicz & Maranzano, 2004a).

The introduction of ensemble weather prediction in recent years has served as a reference in the domain of ensemble flood forecasting (EFF). EFF starts from a central analysis which is also called control forecast, and generates other members by perturbing the initial conditions and the parameter values. Thus ensemble forecasts can be obtained which was found can provide a numerical quantification and representation of predictive uncertainty and give more information for flood forecasting and early warning (Cloke & Pappenberger, 2009; DeChant & Moradkhani, 2014).

For Bayesian ensemble forecast, two approaches were tested to generate BEF: (1) attach an ensemble generator to the analytic-numerical BFS (Herr & Krzysztofowicz, 2010a, 2014); (2) implement the BFS exactly and entirely through Monte Carlo simulation, which is called EBFS (Herr & Krzysztofowicz, 2015). EBFS has three structural components: input ensemble forecaster (IEF), deterministic hydrologic model, hydrologic uncertainty processor (HUP). The function of IEF is to produce an ensemble precipitation time series that inputs to the hydrologic model, and other two components work the same as in traditional BFS. An ensemble of inputs produced by IEF is transformed deterministically by the hydrologic model to an ensemble of outputs, the model outputs next transformed stochastically into an ensemble of predictands by the HUP. Besides the general properties possessed by BFS, one additional property is required for EBFS: that is it should generate an ensemble size large enough to meet the accuracy requirement. As indicated in Herr & Krzysztofowicz (2010b), in order to avoid significant errors, the smallest ensemble size required for the PRSF and the PFF was on the order of several hundreds, and for PSTF the smallest ensemble size required was on the order of several

thousands. However, the computing time increased linearly with the increase of ensemble size, which can be a hurdle for operational use. This gave an impetus to the emergence of EBFSR. EBFSR is a refined method based on EBFS and it takes advantage of auxiliary randomization to reduce the model runs to generate large Bayesian ensemble size. Experiments have shown that EBFSR is more computational efficient and operationally feasible than EBFS (Herr and Krzysztofowicz, 2015a). Since the ensemble size was determined by the randomization factor and the number of hydrologic models runs, methods to find the best values of these two factors were discussed in (Herr and Krzysztofowicz, 2015b).

2.4.2 BFS Application in Real-time Flood Forecasting

Based on the PRSF or the PSTF produced by BFS, PFF can be constructed either approximately or exactly, which specifies a sequence of maximum river stages exceedance functions within a time step. The notion of time interval is important because it is required to assess the total risks of flooding and it is needed to obtain the temporal distribution of flooding. For the method on how to establish bounds on and approximations to the PFF via a short term PRSF, see (Krzysztofowicz, 2002b). As to how to derive the exact PFF via a short term PSTF, reader should refer to (Krzysztofowicz, 2014b). The PFF may be displayed in three formats: (1) the sequence of maximum river stage exceedance functions, (2) the isoprobability of the maximum river stage quantiles, and (3) the temporal distribution of flooding.

Several studies have adapted the BFS to evaluate the total uncertainty associated with water level or river discharge for real-time flood forecasting. For precipitation forecasting,

the High-Resolution Limited-Area Model (HIRLAM) (Reggiani & Weerts, 2008b) or a stochastic model called Prediction of Rainfall Amount Inside Storm Events (PRAISE) (Biondi & De Luca, 2012, 2013) was employed. A modified Bayesian processor of output (BPO) was implemented in one study to process the quantitative precipitation forecasts output from the numerical weather prediction model, the processor represented a larger skill in predicting the occurrence or nonoccurrence of the event rather than in forecasting the event depth. For large basins, the uncertainty due to spatial-temporal variability of precipitation cannot be ignored and should be addressed (Reggiani & Weerts, 2008b). As for hydrologic response, it was simulated by Hydrologiska Byrans Vattenbalansavdelning model (HBV) (Reggiani et al., 2009; Reggiani & Weerts, 2008a) or by the rainfall-runoff model namely Runoff by Infiltration and Saturation Excess (RISE). RISE is a deterministic process-oriented model especially suitable for catchments with small and medium size (Biondi et al., 2009; Biondi & De Luca, 2012, 2013; Biondi et al., 2010). In real-time flood forecasting systems, sometimes hydrological and hydraulic models were interlinked and embedded in a data management environment. Delft-FEWS (Flood Early Warning System) was chosen as the platform in Reggiani & Weerts (2008a) to integrate these models due to its flexibility in model integration and facilitation for data assimilation.

Some applications were carried out for the operational flood forecasting of the river Rhine (160,000 km²) (Reggiani et al., 2009; Reggiani & Weerts, 2008a). Instead of assuming a Markov chain process for the river system, Reggiani & Weerts (2008a) proposed a linear regression for prior density taking upstream observing stations into

consideration. The integration of water level observations from these stations significantly improved the accuracy of prior distribution and revised posterior distribution. However, the linear relation seemed not adequate. Reggiani et al (2009) extended the focus of the Bayesian processor from a deterministic forecast to an ensemble streamflow forecasts named BEUP, the processor translated the prior distribution into an ensemble of posterior distributions and then averaged into a single posterior meta distribution to represent the ensemble forecast. BEUP was shown as an effective uncertainty assessment tool when compared with non-Bayesian uncertainty assessment on the basis of criteria of ranked probability skill score.

Other applications were conducted in Turbolo Creek catchment (29 km²), a semi-arid region situated in Southern part of Italy and the data used within these studies included rainfall amount, temperature and discharge values which were sampled at a 20 minutes time interval (Biondi et al., 2009; Biondi & De Luca, 2012, 2013; Biondi, et al., 2010). For the HUP applied, it rests on three assumptions: (1) precipitation dependent structure conditioned on the occurrence or nonoccurrence of precipitation, (2) nonstationarity of both model river discharge and actual river discharge, (3) meta-Gaussian formulation for all the conditional distributions. It was found that the hydrologic uncertainty grew with increased forecasted discharge and increased lead time, and it would be higher if precipitation occurred (Biondi, Sirangelo, et al., 2009; Biondi, Versace, et al., 2009; Biondi et al., 2010). The real-time simulation of storm event indicated that the linear regression for the dependence structure between observed and simulated discharge was inadequate for high values (Biondi et al., 2009). Results also revealed the crucial role of

HUP to produce predictive distribution of river discharge, highlighted the importance of this complete predictive distribution, and emphasized the superiority of precipitation dependent BFS in real-time flood forecasting (Biondi & De Luca, 2012).

The applications of BFS for real-time flood forecasting have been described in many papers, but its performance assessment is still rare. In order to perform comprehensive analysis of the predictive ability of BFS, graphical tools and scalar metrics learned from meteorology and well suited for probabilistic forecast were used (Biondi & De Luca, 2013). The verification tools included calibration which concerns the statistical consistency between forecasts and observations, sharpness that characterizes the marginal distribution of observations and forecasts, accuracy that considers the joint distribution of the forecasts and observations, and together with continuous ranked probability score (CRPS). Based on these assessment criteria, the interaction of different sources of uncertainty and its impact on the prediction performance were discussed under four different hypotheses: (1) predictive distribution (UD) which relates to the implementation of a perfect hydrologic model; (2) prior distribution (PD) which corresponds to a non-informative PQPF; (3) posterior distribution (HD) that relates to a perfect input; and (4) total predictive distribution (TD) which considers both sources of uncertainty. TD showed best calibration, PD was slightly underestimated, HD showed slight overestimation, while UD lead to marked overestimation. TD provided the widest forecast intervals, UD represented the narrowest bands and relatively higher sharpness, while PD and HD showed intermediate forecast interquartile range (IQR) widths. In terms of Root Mean Square Error (RMSE), TD and HD indicated a better predictive skill

than UD and PD under lead time of 1 h. In summary, TD showed the best predictive capability for river discharge prediction. It also emphasized the importance to perform the multifaceted analysis of prediction attributes (Biondi and De Luca, 2013).

As an alternative to HUP, a new Bayesian processor, model conditional processor (MCP) which combines the observations with the model forecasts in a multi-Normal space was shown as an effective approach to reconcile physically based models with data driven models and to reduce the predictive uncertainty in some extent (Coccia et al, 2010; Todini, 2008, 2012, 2013). In addition, a flexible and hybrid Bayesian multi-model combination framework IBUNE was applied by combining multi-model predictions so that the prediction error of one model can be counteracted by other models. It takes input error, model parameters error and model structural deficiency into account. The results suggested that the input error and model structural uncertainty cannot be ignored, IBUNE was proved to be a useful and applicable technique and able to improve model prediction uncertainty bounds (Ajami et al., 2007). DA and BMA also gained attention recently as they can reduce forecast uncertainty. However, DA methods are limited to one single model, and BMA methods are limited to Gaussian likelihood assumption of predictive distribution and fixed value for model weights. Thus a sequential Bayesian multi-model combination method was proposed which can overcome these limitations (Moradkhani et al., 2009). In ensemble streamflow prediction (ESP), it was combined with ensemble data assimilation and sequential Bayesian combination to address the model and initial condition uncertainty, and it did increase the accuracy of probabilistic forecasts (DeChant & Moradkhani, 2014).

2.4.3 Other Bayesian Method Application in Flood Estimation

Flood frequency analysis is also an important aspect in flood control and water resources management, and it aims at estimating future flood behavior based on historic peak flows. As an alternative to the traditional analysis, a rainfall-runoff model placed in Bayesian inference was used for flood frequency estimation, based on the limited available information in ungauged or poorly gauged basin, the posterior parameters distribution was assessed (Biondi & De Luca, 2015). Both event-based and continuous simulation methods were considered and compared with purely statistical approach. The use of regional hydrological signatures reduced the uncertainty bounds on simulated peak discharge, and the continuous simulation method matched better with the statistical flood frequency analysis (Biondi & De Luca, 2015). It was also demonstrated that the combination of extra information reduced the estimation uncertainty (Viglione et al., 2013). In addition, a Bayesian hierarchical model that consists of several layers, was proposed and better presented the spatial and temporal variability (Yan & Moradkhani, 2014). Multi-model ensemble approach based on BMA can be developed for flood frequency analysis as well. The BMA method provided more robust prediction than single model and the major uncertainty in flood frequency analysis lied in model structure (Yan & Moradkhani, 2015).

Other Bayesian methods applied in the research field of flood forecasting include Bayesian neural network (BNN) for rainfall-runoff modelling (Khan & Coulibaly, 2006), Bayesian networks (BNs) for drought forecasting (Madadgar & Moradkhani, 2014b),

Bayesian dynamic modelling for time series analysis (Mike, 2013) and Bayesian Multi-modeling for streamflow forecast (Madadgar & Moradkhani, 2014a).

The main inputs to a flood forecast model are precipitation and temperature data, advances in their forecast can improve flood forecast accuracy. BMA is a method designed to average the different competing models so that uncertainty accompanied by model selection could be quantified (Hoeting et al., 1999). It can be adopted in streamflow (Najafi & Moradkhani, 2016), precipitation (Sloughter et al., 2007) and temperature estimation (Raftery et al., 2005). Bayesian processor of output (BPO) is a method designed to process model output, and fuse it with observed data to quantify the uncertainty about the predictand (Krzysztofowicz, 2004). (Marty et al., 2015) tried to combine BMA and BPO in a new framework to form a new methodology called Bayesian processor of ensemble members (BPEM) aiming at post processing an ensemble of the numerical weather prediction (NWP) model output, primarily temperature. The forecast skill of BPEM was found to slightly outperform BMA, BPO and climatological forecast according to the assessment of continuous ranked probability score (CRPS) and reliability component. Other methods conducted include Bayesian weighting approach for precipitation forecasting based on observation likelihood to determine a set of weights for each member (Raynaud et al., 2015), hierarchical Bayesian model (HBM) for temperature forecasting (Di Narzo and Cocchi, 2010), hierarchical Bayesian network (HBN) for soil moisture data assimilation (Qin et al., 2013).

2.4.4 Review and Discussion of the Predictive Uncertainty Assessment Approaches in Flood Forecasting

Besides the Bayesian forecasting methods discussed before, there are other various approaches developed to assess predictive uncertainty in flood forecasting application. From their theoretical base, all the predictive uncertainty assessment methods used in flood forecasting can be classified into 5 types: (1) model error analytical methods, which are based on the statistical analysis of model errors, e.g. Quantile regression (QR) (Weerts et al., 2011) and meta-Gaussian approach (Montanari & Grossi, 2008); (2) ensemble based techniques, relying on numerous simulations or multi-models to generate the output probability distribution, e.g. BMA (Duan et al., 2007), hydrologic model output statistics (HMOS) (Regonda et al., 2013), general linear model post-processor (GLMPP) (Zhao et al., 2011), and non-Gaussian copulas approach (Madadgar et al., 2014); (3) Bayesian methods, in which Bayesian theory is used to estimate the probability distribution of predictand conditional on available information, e.g. BFS (Krzysztofowicz, 2002a), model conditional processor (MCP) (Todini, 2008, 2012), BEF (Herr & Krzysztofowicz, 2010, 2015; Reggiani et al., 2009), IBUNE (Ajami et al., 2007) and Sequential Bayesian Multi-model Combination (DeChant & Moradkhani, 2014); (4) data assimilation (DA) methods, e.g. ensemble Kalman filter (EnKF) (Dechant & Moradkhani, 2012) and particle filter (PF) (Salamon & Feyen, 2009); (5) machine learning techniques, e.g. uncertainty estimation based on local errors and clustering (UNEEC) (Dogulu et al., 2014; Solomatine & Shrestha, 2009), grey neural networks (Alvisi & Franchini, 2012), grey number theory based approach (Alvisi et al., 2013), and Bayesian neural network

(BNN) (Khan & Coulibaly, 2006). No method is perfect, each of them has advantages and limitations. The choice of a method highly depends on the available data and the research purpose. The detailed discussion of Prons and Cons of each approach can be found below and also summarized in Table 2-3.

Table 2-3 Comparison of the predictive uncertainty assessment approaches in flood forecasting (in chronological order)

References	Key Objectives	Methods	Pros	Cons
Krzysztofowicz, R. (1999); Krzysztofowicz, R. (2002); Krzysztofowicz, R., & Maranzano, C. J. (2004); Biondi, D., & De Luca, D. L. (2012)	Assess the total uncertainty associated with PRSF, PSTF or real-time flood forecasting	Bayesian Forecasting System (BFS)	<ul style="list-style-type: none"> • Flexible and robust theoretical structure • Able to quantify all sources of uncertainties 	Rest on structural and distributional assumptions
Montanari, A., & Brath, A. (2004); Montanari, A., & Grossi, G. (2008)	Assess the uncertainty associated with rainfall-runoff simulations	Meta-Gaussian Approach	<ul style="list-style-type: none"> • Low computational requirement • Quite straightforward to apply 	<ul style="list-style-type: none"> • Require model error assumption • Consider limited sources of uncertainty
Khan, M. S., & Coulibaly, P. (2006); Zhang, X., et al. (2009); Zhang, X., et al. (2011)	Apply BNN for uncertainty estimation of streamflow simulation	Bayesian Neural Network (BNN)	Can address the problem of overfit and underfit	Identification of interaction between different uncertainties need to be improved
Duan, Q., et al. (2007); Vrugt, J. A., & Robinson, B. A. (2007); Liang, Z., et al. (2013)	Develop reliable probabilistic streamflow forecast	Bayesian Model Averaging (BMA)	Use multi-model to reduce the model structure uncertainty	Might cause large uncertainty in real-time predictions
Vrugt, J. A., & Robinson, B. A. (2007); Salamon, P., & Feyen, L. (2009); DeChant, C. M., & Moradkhani, H. (2012); Moradkhani, H., et al. (2012)	Quantify uncertainty in hydrologic forecasting	Data Assimilation (DA)	<ul style="list-style-type: none"> • Assimilate various sources of data • Update states automatically 	<ul style="list-style-type: none"> • EnKF: assumption of Gaussian error distribution • PF: computational demand increase in some cases
Ajami, N. K., et al. (2007)	Build IBUNE to confront input uncertainty, model parameter uncertainty and model structure uncertainty	Integrated Bayesian Uncertainty Estimator (IBUNE)	<ul style="list-style-type: none"> • Bayesian multi-model combination framework • Produce improved model prediction uncertainty bounds 	Expensive to be used in real-time operational application

Scholzel, C. & Friederichs, P. (2008); Madadgar, S., et al. (2014)	Introduce the non-Gaussian copulas approach for ensemble flow forecast	Non-Gaussian Copulas Approach	<ul style="list-style-type: none"> • Able to catch covariance structure • Account for connection between forecasts and observations 	Don't solve dimensionality problem
Todini, E. (2008); Todini, E. (2013)	Propose the MCP as an alternative to HUP to assess predictive uncertainty	Model Conditional Processor (MCP)	<ul style="list-style-type: none"> • More capable for multi-model and multi-location application • MCP algorithm and its derivation are more simple than HUP 	Little consideration for precipitation input uncertainty
Wang, Q. J., et al. (2009); Wang, Q. J. & Robertson, D. E. (2011); Zhao, T., et al. (2015)	Quantify predictive uncertainty by post process deterministic streamflow forecast	Bayesian Joint Probability (BJP)	<ul style="list-style-type: none"> • Low requirement for input data • Capable to apply in real-time forecasts 	<ul style="list-style-type: none"> • Might cause bound-related issue • Sensitive to initial catchment condition
Solomatine, D. P. & Shrestha, D. L. (2009); Dogulu, N., et al. (2014)	Estimate model uncertainty by UNEEC and compare with GLUE, meta-Gaussian and QR	Uncertainty Estimation based on Local Errors and Clustering (UNEEC)	<ul style="list-style-type: none"> • Can express the whole model error distribution • No requirement for residual pdf assumption 	<ul style="list-style-type: none"> • Issue of extrapolation • Accuracy highly depend on regression model
Reggiani, P., et al. (2009); Herr, H. D., & Krzysztofowicz, R. (2010); Herr, H. D., & Krzysztofowicz, R. (2015)	Produce Bayesian ensemble forecast that provide numerical characterization of predictive uncertainty	Bayesian Ensemble Forecast (BEF)	Combine BFS with ensemble technique	Computing time increases with the increase of ensemble size
Moradkhani, H., et al. (2009); DeChant, C. M., & Moradkhani, H. (2014)	Assess initial condition uncertainty, future climate uncertainty and model errors with streamflow forecasting	Sequential Bayesian Multi-model Combination method	<ul style="list-style-type: none"> • Sequential updating to blend multiple models • Not limited to fixed model weight as in BMA 	Require to combine with other technique for some cases, e.g. DA

Brown, J. D., & Seo, D. J. (2010)	Quantify and reduce bias in ensemble hydrologic forecasts	Non-Parametric Postprocessor	Good for variables lack of parametric distribution	<ul style="list-style-type: none"> • Don't quantify sampling uncertainty with extreme events • May not good for dynamic modeling
Weerts, A. H., et al. (2011)	Estimate hydrological uncertainty associated with water level forecast	Quantile Regression (QR)	<ul style="list-style-type: none"> • Relatively simple to apply • Require very few assumptions • Easy to understand the theory 	<ul style="list-style-type: none"> • Require long time period of data • Don't include non-linear model
Zhao, L., et al. (2011); Ye, A., et al. (2014); Ye, A., et al. (2015)	Adjust and reduce bias in ensemble streamflow predictions	General Linear Model Post-Processor (GLMPP)	<ul style="list-style-type: none"> • Easily applied to ensemble forecasts • Remove mean biases effectively 	Sensitive to different settings
Steenbergen, N. V., et al. (2012)	Use a non-parametric data-based approach for probabilistic flood forecasting	Non-Parametric Data-based Approach	Generate more realistic confidence interval	Unable to assess the uncertainty for value out of the historical range
Alvisi, S. & Franchini, M. (2012)	Quantify the uncertainty with river stage forecasting	Grey Neural Networks	Capture non-linear relationship between variables without physical process simulation	Require excessive computing time
Alvisi, S., et al. (2013)	Quantify the uncertainty with rr model and compare with GLUE	Grey Number Theory based Approach	Significantly reduce the computation time	Require good quality of data
Regonda, S. K., et al. (2013)	Generate ensemble streamflow forecasts from single-valued forecasts	Hydrologic Model Output Statistics (HMOS)	<ul style="list-style-type: none"> • Preserve temporal correlation over successive lead times • Less expensive to develop and apply 	<ul style="list-style-type: none"> • Limited sample size with operational forecasts • Consider limited sources of uncertainty

Quantile regression (QR) was introduced by Koenker (2005), and applied by Weerts et al (2011) to estimate hydrologic uncertainty associated with water level forecast. QR provides a predictive uncertainty assessment method that is easy to understand and relatively simple to apply. Its main limitation is that it requires long time period of data in order to generate ideal results, and unlike UNEEC, QR relies on linear regression. The meta-Gaussian approach, which is based on the regression of model errors over the model forecast to assess rainfall-runoff modeling uncertainty, was found quite straightforward to apply and has low computational requirement, but the model errors need to be assumed as Gaussian and homoscedastic (Montanari & Grossi, 2008; Montanari & Brath, 2004). The predictive uncertainty can also be estimated by analyzing the residuals in a non-parametric way, as the non-parametric data-based approach presented by Steenbergen et al (2012).

Another probabilistic streamflow forecast, BMA proposed by Raftery (1993), was used in several studies (Duan et al., 2007; Liang et al., 2013; Vrugt & Robinson, 2007). It uses multi-model to reduce the uncertainty due to model selection, and the model with better performance gets higher BMA weight. Due to these advantages, BMA was also widely used in meteorology. Some studies found that it will lead to less biased predictive distribution and smaller uncertainty when integrating other methods with BMA, for example, copula-embedded BMA (Cop-BMA) (Madadgarand & Moradkhani, 2014) and combination of Genetic Algorithms (GA) and BMA (Zhang et al., 2009). One disadvantage is that it may cause large uncertainty in real-time predictions.

Hydrologic model output statistics (HMOS) introduced by Regonda et al (2013) as an uncertainty post processor, is able to generate ensemble streamflow forecasts from single-valued forecasts. The ensemble members can preserve the temporal correlation over successive lead times. For further improvement, the issue of limited sample size obtained from the operational forecasts should be considered. In meteorological applications, model output statistics (MOS) (Glahn & Lowry, 1972) and ensemble model output statistics (EMOS) (Gneiting et al., 2005) were widely adopted, which are based on the same theory as HMOS. The general linear model post-processor (GLMPP) developed by Zhao et al (2011), was found capable to remove mean biases effectively in ensemble streamflow predictions (Ye et al., 2014, 2015). In operational application, one may pay attention to the parameter setting as the results are sensitive to different settings. The non-Gaussian copulas approach have been recently introduced in climatological applications (Scholzel & Friederichs, 2008), and its variant, the multivariate post-processor was used to set up uncertainty post processors for forecasted flow (Madadgar et al., 2014). This multivariate copula-based method is able to catch covariance structure and account for the joint connection between forecasts and observations. The limitation of the copula approach is that it doesn't deal with the problem of dimensionality, obtaining the parametric distribution for high dimensional variables still remains complicated.

BFS provides a flexible and robust theoretical structure and is able to quantify all sources of uncertainties, but it rests on certain structural and distributional assumptions. As an alternative to HUP, which is one essential component of BFS, MCP is more capable for multi-model and multi-location application and its derivation is more simple than HUP

(Todini, 2008, 2013). By combining BFS with ensemble technique, BEF has the ability to generate Bayesian ensemble forecasts to numerically characterize the predictive uncertainty. Based on a Bayesian multi-model combination framework, IBUNE can provide improved model prediction uncertainty bounds. However, BEF and IBUNE show potential weakness in terms of operational application, for example, the computing time of BEF increases with the increase of ensemble size (Herr & Krzysztofowicz, 2010, 2015), while IBUNE is computationally expensive to be used in real-time operational application (Ajami et al., 2007). Not limited to the fixed model weights as in BMA method, Sequential Bayesian Multi-model Combination has the flexibility to assign higher weights to better performed models for a certain time period, and the adjustable parameters are sequentially updated to blend multiple models, sometimes it is required to combine with other techniques such as DA to obtain adequate outcome.

Due to the ability to assimilate various sources of data, merge observations and simulations optimally and update the state continuously, DA method has gained popularity in quantifying uncertainty in hydrologic forecasting, such as EnKF (Dechant & Moradkhani, 2012; Vrugt & Robinson, 2007), PF (Salamon & Feyen, 2009). Several variants of PF like particle filter-sequential importance resampling (PF-SIR) (Dechant & Moradkhani, 2012), particle filter-markov chain monte carlo (PF-MCMC) (Moradkhani et al., 2012) have been proposed. EnKF is computationally efficient, and could achieve better performance compared with BMA (Vrugt & Robinson, 2007), whereas the theory depends on the assumption of gaussian error distribution. It is noteworthy that in prediction mode, the Kalman filter cannot be directly used for assessing predictive

uncertainty since future observations are not available, thus the model forecasts are now required to work as “pseudo measurements” of the future predicted quantities (Todini, 2006, 2016). PF is more capable of preserving the spatial distribution pattern than EnKF (Dechant & Moradkhani, 2012), but in filtering algorithms, the problem of weight collapse might happen due to particle weight disparity, if trying to handle this issue by adding a resampling step, the computational cost will be high.

Machine learning techniques have also been used in order to assess the predictive uncertainty, for example, uncertainty estimation based on local errors and clustering (UNEEC) (Dogulu et al., 2014; Solomatine & Shrestha, 2009), grey neural networks (Alvisi & Franchini, 2012), grey number theory based approach (Alvisi et al., 2013), Bayesian neural network (BNN) (Khan & Coulibaly, 2006; Zhang et al., 2009, 2011). The advantage of these machine learning approaches is that they can capture the non-linear relationship between variables without explicitly modeling the physical process, and improved model uncertainty estimation can be generated compared with other methods, such as meta-Gaussian and QR (Alvisi et al., 2013; Solomatine & Shrestha, 2009). The main limitation is their inability to extrapolate beyond the boundary of the training dataset.

Under the condition that the variables don't have a parametric distribution form, non-parametric postprocessor applied by Brown & Seo (2010) is a good way to quantify and reduce bias in ensemble hydrologic forecasts, it is based on Bayesian optimal linear estimation and is similar to indicator co-kriging (ICK). The shortcoming of this method is the lack of ability for dynamic modeling as it doesn't model joint distribution of

observations across different lead time. If the quality of data record is not ideal, Bayesian joint probability (BJP) will be a possible method to consider for probabilistic streamflow forecasting. BJP has low requirement for the input data, which means it can contain some missing record, and it should be able to work in real-time forecasts (Zhao et al., 2015). On condition that the variable has a physical range, bound-related issue might happens (Wang et al., 2009). The extended BJP is applicable to streams with zero flows, however, it is sensitive to the initial catchment condition (Wang & Robertson, 2011).

2.5 Conclusions and Future Work

A comprehensive review of Bayesian methods applied to flood forecasting over the last two decades is provided. Bayesian flood forecasting is an advanced and effective way for probabilistic flood forecast with uncertainty estimate. It offers an ideal theoretical structure and quantifies all sources of uncertainties, thus can reduce predictive uncertainty to some extent and lead to a more reliable and accurate forecast. As shown in this paper, the Bayesian flood forecasting approaches have been developed rapidly and widely applied since 1999, which brings great confidence in the research field of accurate and reliable flood forecast and estimation. However, there is still room for improvement and some challenges to overcome. First, only limited types of river basin were studied to develop and test the Bayesian forecasting approaches, it remains unknown whether these Bayesian flood forecasting approaches are suitable for different watersheds with different sizes and different physical and climatic characteristics. Second, more forecast products could be developed to express the predictive uncertainty. Third, most of the previous studies used single average precipitation amount as input to the Bayesian forecasting

methods. However, many other sources of data may have added value (e.g., radar data, numerical weather prediction data, and remotely sensed data), thus it is expected that the Bayesian approach could be more flexible to assimilate various sources of newly available information. This does require further research. Finally, as performance assessment of Bayesian flood forecasting is rare, it might be a potential research direction. Overall, significant work is needed in the future to address all the above issues in the context of flood forecasting with uncertainty estimate.

2.6 Acknowledgments

This work was supported by the Natural Science and Engineering Research Council (NSERC) Canadian FloodNet (grant number: NETGP 451456) and the China Scholarship Council (CSC). The authors would like to thank the four anonymous reviewers for their suggestions and comments that helped to improve the quality of this paper.

2.7 References

- Ajami, N. K., Duan, Q., & Sorooshian, S. (2007). An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction. *Water Resources Research*, 43(1). <http://doi.org/10.1029/2005WR004745>
- Alvisi, S., Bernini, A., & Franchini, M. (2013). A conceptual grey rainfall-runoff model for simulation with uncertainty. *Journal of Hydroinformatics*, 15(1), 1–20. <http://doi.org/10.2166/hydro.2012.069>
- Alvisi, S., & Franchini, M. (2012). Grey neural networks for river stage forecasting with uncertainty. *Physics and Chemistry of the Earth*, 42-44, 108–118. <http://doi.org/10.1016/j.pce.2011.04.002>
- Balica, S. F., Popescu, I., Beevers, L., & Wright, N. G. (2013). Parametric and physically based modelling techniques for flood risk and vulnerability assessment: A comparison. *Environmental Modelling & Software*, 41, 84–92. <http://doi.org/10.1016/j.envsoft.2012.11.002>
- Biondi, D., & De Luca, D. L. (2015, April). Model parameters conditioning on regional hydrologic signatures for process-based design flood estimation in ungauged basins. In *EGU General Assembly Conference Abstracts* (Vol. 17, p. 10146).
- Biondi, D., & De Luca, D. L. (2015, August). Process based design flood estimation under parameter uncertainty: a case study in southern Italy. In *IUGG General Assembly Conference Abstracts*.
- Biondi, D., Sirangelo, B., & Versace, P. (2009, September). Application of a precipitation dependent HUP to a small watershed in Southern Italy. In *11th Plinius Conference on Mediterranean Storms* (Vol. 11, p. 11661).
- Biondi, D., Versace, P., & Sirangelo, B. (2009, April). Uncertainty assessment through a precipitation dependent HUP: an application to a small Southern Italy catchment. In *EGU General Assembly Conference Abstracts* (Vol. 11, p. 9489).

- Biondi, D., & De Luca, D. L. (2012). A Bayesian approach for real-time flood forecasting. *Physics and Chemistry of the Earth, Parts A/B/C*, 42-44, 91–97.
<http://doi.org/10.1016/j.pce.2011.04.004>
- Biondi, D., & De Luca, D. L. (2013). Performance assessment of a Bayesian Forecasting System (BFS) for real-time flood forecasting. *Journal of Hydrology*, 479, 51–63.
<http://doi.org/10.1016/j.jhydrol.2012.11.019>
- Biondi, D., & De Luca, D. L. (2015). Process-based design flood estimation in ungauged basins by conditioning model parameters on regional hydrological signatures. *Natural Hazards*, 79(2), 1015–1038. <http://doi.org/10.1007/s11069-015-1889-1>
- Biondi, D., Versace, P., & Sirangelo, B. (2010). Uncertainty assessment through a precipitation dependent hydrologic uncertainty processor: An application to a small catchment in southern Italy. *Journal of Hydrology*, 386(1-4), 38–54.
<http://doi.org/10.1016/j.jhydrol.2010.03.004>
- Brown, J. D., & Seo, D. J. (2010). A nonparametric postprocessor for bias correction of hydrometeorological and hydrologic ensemble forecasts. *Journal of Hydrometeorology*, 11(3), 642–665. <http://doi.org/10.1175/2009JHM1188.1>
- Cloke, H. L., & Pappenberger, F. (2009). Ensemble flood forecasting: A review. *Journal of Hydrology*, 375(3-4), 613–626. <http://doi.org/10.1016/j.jhydrol.2009.06.005>
- Coccia, G., Francés, F., Camilo Múnera, J., & Todini, E. (2010). Application of a Bayesian processor for predictive uncertainty assessment in real time flood forecasting. In *EGU General Assembly Conference Abstracts* (Vol. 12, p. 13006).
- Dechant, C. M., & Moradkhani, H. (2012). Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting. *Water Resources Research*, 48(4).
<http://doi.org/10.1029/2011WR011011>
- DeChant, C. M., & Moradkhani, H. (2014). Toward a reliable prediction of seasonal forecast uncertainty: Addressing model and initial condition uncertainty with

- ensemble data assimilation and Sequential Bayesian Combination. *Journal of Hydrology*, 519, 2967–2977. <http://doi.org/10.1016/j.jhydrol.2014.05.045>
- Di Narzo, A. F., & Cocchi, D. (2010). A Bayesian hierarchical approach to ensemble weather forecasting. *Journal of the Royal Statistical Society Series C-Applied Statistics*, 59(3), 405–422. <http://doi.org/10.1111/j.1467-9876.2009.00700.x>
- Dogulu, N., López López, P., Solomatine, D. P., Weerts, A. H., & Shrestha, D. L. (2014). Estimation of predictive hydrologic uncertainty using the quantile regression and UNEEC methods and their comparison on contrasting catchments. *Hydrology and Earth System Sciences*, 19, 3181–3201. <http://doi.org/10.5194/hess-19-3181-2015>
- Duan, Q., Ajami, N. K., Gao, X., & Sorooshian, S. (2007). Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Advances in Water Resources*, 30(5), 1371–1386. <http://doi.org/10.1016/j.advwatres.2006.11.014>
- Glahn, H. R., & Lowry, D. A. (1972). The use of model output statistics (MOS) in objective weather forecasting. *Journal of Applied Meteorology*, 11(8), 1203-1211. [http://doi.org/10.1175/1520-0450\(1972\)011<1203:TUOMOS>2.0.CO;2](http://doi.org/10.1175/1520-0450(1972)011<1203:TUOMOS>2.0.CO;2)
- Gneiting, T., Raftery, A. E., Westveld, A. H., & Goldman, T. (2005). Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Monthly Weather Review*, 133(5), 1098–1118. <http://doi.org/10.1175/MWR2904.1>
- Herr, H. D., & Krzysztofowicz, R. (2010a). Bayesian ensemble forecast of river stages and ensemble size requirements. *Journal of Hydrology*, 387(3-4), 151–164. <http://doi.org/10.1016/j.jhydrol.2010.02.024>
- Herr, H. D., & Krzysztofowicz, R. (2010b). Bayesian ensemble forecast of river stages and ensemble size requirements. *Journal of Hydrology*, 387(3-4), 151–164. <http://doi.org/10.1016/j.jhydrol.2010.02.024>

- Herr, H. D., & Krzysztofowicz, R. (2014). Addendum to Bayesian ensemble forecast of river stages and ensemble size requirements. *Journal of Hydrology*, 515, 304–306.
<http://doi.org/10.1016/j.jhydrol.2014.04.065>
- Herr, H. D., & Krzysztofowicz, R. (2015). Ensemble Bayesian forecasting system Part I: Theory and algorithms. *Journal of Hydrology*, 524, 789–802.
<http://doi.org/10.1016/j.jhydrol.2014.11.072>
- Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, 14(4), 382–401.
<http://doi.org/10.2307/2676803>
- Kelly, K. S., & Krzysztofowicz, R. (2000). Precipitation uncertainty processor for probabilistic river stage forecasting. *Water Resources Research*, 36(9), 2643–2653.
<http://doi.org/10.1029/2000WR900061>
- Khan, M. S., & Coulibaly, P. (2006). Bayesian neural network for rainfall-runoff modeling. *Water Resources Research*, 42(7). <http://doi.org/10.1029/2005WR003971>
- Koenker, R. (2005). *Quantile Regression* (No.38). Cambridge university press.
- Krzysztofowicz, R. (1983). A Bayesian Markov Model of the flood forecast process. *Water Resources Research*, 19(6), 1455–1465.
<http://doi.org/10.1029/WR019i006p01455>
- Krzysztofowicz, R. (1999). Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resources Research*, 35(9), 2739–2750.
<http://doi.org/10.1029/1999WR900099>
- Krzysztofowicz, R. (2001a). Integrator of uncertainties for probabilistic river stage forecasting: Precipitation-dependent model. *Journal of Hydrology*, 249(1-4), 69–85.
[http://doi.org/10.1016/S0022-1694\(01\)00413-9](http://doi.org/10.1016/S0022-1694(01)00413-9)
- Krzysztofowicz, R. (2001b). Integrator of uncertainties for probabilistic river stage forecasting: precipitation-dependent model. *Journal of Hydrology*, 249(1-4), 69–85.
[http://doi.org/10.1016/S0022-1694\(01\)00413-9](http://doi.org/10.1016/S0022-1694(01)00413-9)

- Krzysztofowicz, R. (2002a). Bayesian system for probabilistic river stage forecasting. *Journal of Hydrology*, 268(1-4), 16–40. [http://doi.org/10.1016/S0022-1694\(02\)00106-3](http://doi.org/10.1016/S0022-1694(02)00106-3)
- Krzysztofowicz, R. (2002b). Probabilistic flood forecast: bounds and approximations. *Journal of Hydrology*, 268(1-4), 41–55. [http://doi.org/10.1016/S0022-1694\(02\)00107-5](http://doi.org/10.1016/S0022-1694(02)00107-5)
- Krzysztofowicz, R. (2004). Bayesian processor of output: A new technique for probabilistic weather forecasting. *Bulletin of the American Meteorological Society*, (iii), 3803–3807.
- Krzysztofowicz, R. (2014a). Corrigendum to “Bayesian system for probabilistic river stage forecasting” [J. Hydrol. 268 (1–4) (2002) 16–40]. *Journal of Hydrology*, 515, 346. <http://doi.org/10.1016/j.jhydrol.2014.04.051>
- Krzysztofowicz, R. (2014b). Probabilistic flood forecast: Exact and approximate predictive distributions. *Journal of Hydrology*, 517, 643–651. <http://doi.org/10.1016/j.jhydrol.2014.04.050>
- Krzysztofowicz, R., & Herr, H. D. (2001). Hydrologic uncertainty processor for probabilistic river stage forecasting: precipitation-dependent model. *Journal of Hydrology*, 249(1-4), 46–68. [http://doi.org/10.1016/S0022-1694\(01\)00412-7](http://doi.org/10.1016/S0022-1694(01)00412-7)
- Krzysztofowicz, R., & Kelly, K. S. (2000). Hydrologic uncertainty processor for probabilistic river stage forecasting. *Water Resources Research*, 36(11), 3265–3277. <http://doi.org/10.1029/2000WR900108>
- Krzysztofowicz, R., & Maranzano, C. J. (2004a). Bayesian system for probabilistic stage transition forecasting. *Journal of Hydrology*, 299(1-2), 15–44. <http://doi.org/10.1016/j.jhydrol.2004.02.013>
- Krzysztofowicz, R., & Maranzano, C. J. (2004b). Hydrologic uncertainty processor for probabilistic stage transition forecasting. *Journal of Hydrology*, 293(1-4), 57–73. <http://doi.org/10.1016/j.jhydrol.2004.01.003>

- Liang, Z., Wang, D., Guo, Y., Zhang, Y., & Dai, R. (2011). Application of bayesian model averaging approach to multimodel ensemble hydrologic forecasting. *Journal of Hydrologic Engineering*, 18(11), 1426–1436.
[http://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000493](http://doi.org/10.1061/(ASCE)HE.1943-5584.0000493)
- Madadgar, S., & Moradkhani, H. (2014a). Improved Bayesian multimodeling: Integration of copulas and Bayesian model averaging. *Water Resources Research*, 50(12), 9586–9603. <http://doi.org/10.1002/2014WR015965>
- Madadgar, S., & Moradkhani, H. (2014b). Spatio-temporal drought forecasting within Bayesian networks. *Journal of Hydrology*, 512, 134–146.
<http://doi.org/10.1016/j.jhydrol.2014.02.039>
- Madadgar, S., & Moradkhani, H. (2014c). Improved Bayesian multimodeling: Integration of copulas and Bayesian model averaging. *Water Resources Research*, 50(12), 9586–9603. <http://doi.org/10.1002/2014WR015965>
- Madadgar, S., Moradkhani, H., & Garen, D. (2014). Towards improved post-processing of hydrologic forecast ensembles. *Hydrological Processes*, 28(1), 104–122.
<http://doi.org/10.1002/hyp.9562>
- Maranzano, C. J., & Krzysztofowicz, R. (2004). Identification of likelihood and prior dependence structures for hydrologic uncertainty processor. *Journal of Hydrology*, 290(1-2), 1–21. <http://doi.org/10.1016/j.jhydrol.2003.11.021>
- Marty, R., Fortin, V., Kuswanto, H., Favre, A.C., & Parent, E. (2015). Combining the Bayesian processor of output with Bayesian model averaging for reliable ensemble forecasting. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64(1), 75–92. <http://doi.org/10.1111/rssc.12062>
- Montanari, A., & Grossi, G. (2008). Estimating the uncertainty of hydrological forecasts: A statistical approach. *Water Resources Research*, 44(12).
<http://doi.org/10.1029/2008WR006897>

- Montanari, A., & Brath, A. (2004). A stochastic approach for assessing the uncertainty of rainfall-runoff simulations. *Water Resources Research*, 40(1).
<http://doi.org/10.1029/2003WR002540>
- Moradkhani, H., Hsu, K., & Sorooshian, S. (2009, April). Reducing Streamflow Forecasting Uncertainty: A Sequential Bayesian Approach Inspired by Data Assimilation for Multi-model Combination and Forecasting. In *EGU General Assembly Conference Abstracts* (Vol.11, p. 1067).
- Moradkhani, H., Dechant, C. M., & Sorooshian, S. (2012). Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-Markov chain Monte Carlo method. *Water Resources Research*, 48(12).
<http://doi.org/10.1029/2012WR012144>
- Najafi, M. R., & Moradkhani, H. (2015). Ensemble combination of seasonal streamflow forecasts. *Journal of Hydrologic Engineering*, 21(1), 04015043.
[http://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001250](http://doi.org/10.1061/(ASCE)HE.1943-5584.0001250)
- Pachauri, R. K., et al. (2014). Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II, III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.
- Qin, S., Ma, J., & Wang, X. (2013). Development of a hierarchical Bayesian network algorithm for land surface data assimilation. *International Journal of Remote Sensing*, 34(6), 1905–1927. <http://doi.org/10.1080/01431161.2012.727495>
- Raftery, A. E. (1993). Bayesian model selection in structural equation models. *Sage Focus Editions*, 154, 163–163.
- Raftery, A. E., Gneiting, T., Balabdaoui, F., & Polakowski, M. (2005). Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, 133(5), 1155–1174. <http://doi.org/10.1175/MWR2906.1>
- Raynaud, L., Pannekoucke, O., Arbogast, P., & Bouttier, F. (2015). Application of a Bayesian weighting for short-range lagged ensemble forecasting at the convective

- scale. *Quarterly Journal of the Royal Meteorological Society*, 141(687), 459–468.
<http://doi.org/10.1002/qj.2366>
- Reggiani, P., Renner, M., Weerts, A. H., & Van Gelder, P. A H. J. M. (2009).
Uncertainty assessment via Bayesian revision of ensemble streamflow predictions in
the operational river Rhine forecasting system. *Water Resources Research*, 45(2), 1–
14. <http://doi.org/10.1029/2007WR006758>
- Reggiani, P., & Weerts, A. H. (2008a). A Bayesian approach to decision-making under
uncertainty: An application to real-time forecasting in the river Rhine. *Journal of
Hydrology*, 356(1-2), 56–69. <http://doi.org/10.1016/j.jhydrol.2008.03.027>
- Reggiani, P., & Weerts, A. H. (2008b). Probabilistic quantitative precipitation forecast
for flood prediction: An application. *Journal of Hydrometeorology*, 9(1), 76–95.
<http://doi.org/10.1175/2007JHM858.1>
- Regonda, S. K., Seo, D. J., Lawrence, B., Brown, J. D., & Demargne, J. (2013). Short-
term ensemble streamflow forecasting using operationally-produced single-valued
streamflow forecasts - A Hydrologic Model Output Statistics (HMOS) approach.
Journal of Hydrology, 497, 80–96. <http://doi.org/10.1016/j.jhydrol.2013.05.028>
- Salamon, P., & Feyen, L. (2009). Assessing parameter, precipitation, and predictive
uncertainty in a distributed hydrological model using sequential data assimilation
with the particle filter. *Journal of Hydrology*, 376(3-4), 428–442.
<http://doi.org/10.1016/j.jhydrol.2009.07.051>
- Schoelzel, C., & Friederichs, P. (2008). Multivariate non-normally distributed random
variables in climate research-introduction to the copula approach. *Nonlin. Processes
Geophys.*, 15(5), 761-772.
- Sloughter, J. M. L., Raftery, A. E., Gneiting, T., & Fraley, C. (2007). Probabilistic
quantitative precipitation forecasting using Bayesian model averaging. *Monthly
Weather Review*, 135(9), 3209–3220. <http://doi.org/10.1175/MWR3441.1>

- Solomatine, D. P., & Shrestha, D. L. (2009). A novel method to estimate model uncertainty using machine learning techniques. *Water Resources Research*, 45(12). <http://doi.org/10.1029/2008WR006839>
- Todini, E. (2012, April). A Comparison of Uncertainty Processors (HUP, BMA and MCP) in the Normal Space. In *EGU General Assembly Conference Abstracts* (vol. 14, p. 13892).
- Todini, E. (2006). Rainfall-runoff models for real-time forecasting. In *Encyclopedia of Hydrological Sciences* (Part 11, Paper 123). Chichester, UK: John Wiley & Sons.
- Todini, E. (2008). A model conditional processor to assess predictive uncertainty in flood forecasting. *International Journal of River Basin Management*, 6(2), 123–137. <http://doi.org/10.1080/15715124.2008.9635342>
- Todini, E. (2013). From HUP to MCP: Analogies and extended performances. *Journal of Hydrology*, 477, 33–42. <http://doi.org/10.1016/j.jhydrol.2012.10.037>
- Todini, E. (2016). Predictive uncertainty assessment and decision making. In V. P. Singh (Ed), *Handbook of Applied Hydrology, Second Edition* (Chapter 26, pp.1-15). New York, US: McGraw-Hill Education.
- Van Steenbergen, N., Ronsyn, J., & Willems, P. (2012). A non-parametric data-based approach for probabilistic flood forecasting in support of uncertainty communication. *Environmental Modelling and Software*, 33, 92–105. <http://doi.org/10.1016/j.envsoft.2012.01.013>
- Viglione, A., Merz, R., Salinas, J. L., & Blöschl, G. (2013). Flood frequency hydrology: 3. A Bayesian analysis. *Water Resources Research*, 49(2), 675–692. <http://doi.org/10.1029/2011WR010782>
- Vrugt, J. A., & Robinson, B. A. (2007). Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging. *Water Resources Research*, 43(1). <http://doi.org/10.1029/2005WR004838>

- Wang, Q. J., & Robertson, D. E. (2011). Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences. *Water Resources Research*, 47(2). <http://doi.org/10.1029/2010WR009333>
- Wang, Q. J., Robertson, D. E., & Chiew, F. H. S. (2009). A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites. *Water Resources Research*, 45(5). <http://doi.org/10.1029/2008WR007355>
- Weerts, A. H., Winsemius, H. C., & Verkade, J. S. (2011). Estimation of predictive hydrological uncertainty using quantile regression: examples from the National Flood Forecasting System (England and Wales). *Hydrology and Earth System Sciences*, 15,(1).
- Yan, H., & Moradkhani, H. (2014). A regional Bayesian hierarchical model for flood frequency analysis. *Stochastic Environmental Research and Risk Assessment*, 29(3), 1019–1036. <http://doi.org/10.1007/s00477-014-0975-3>
- Yan, H., & Moradkhani, H. (2016). Toward more robust extreme flood prediction by Bayesian hierarchical and multimodeling. *Natural Hazards*, 81(1), 203-225. <http://doi.org/10.1007/s11069-015-2070-6>
- Ye, A., Duan, Q., Schaake, J., Xu, J., Deng, X., Di, Z.,...& Gong, W. (2015). Post-processing of ensemble forecasts in low-flow period. *Hydrological Processes*, 29(10), 2438–2453. <http://doi.org/10.1002/hyp.10374>
- Ye, A., Duan, Q., Yuan, X., Wood, E. F., & Schaake, J. (2014). Hydrologic post-processing of MOPEX streamflow simulations. *Journal of Hydrology*, 508, 147–156. <http://doi.org/10.1016/j.jhydrol.2013.10.055>
- Zhang, X., Liang, F., Srinivasan, R., & Van Liew, M. (2009). Estimating uncertainty of streamflow simulation using Bayesian neural networks. *Water Resources Research*, 45(2). <http://doi.org/10.1029/2008WR007030>
- Zhang, X., Liang, F., Yu, B., & Zong, Z. (2011). Explicitly integrating parameter, input, and structure uncertainties into Bayesian Neural Networks for probabilistic

hydrologic forecasting. *Journal of Hydrology*, 409(3), 696–709.

<http://doi.org/10.1016/j.jhydrol.2011.09.002>

Zhang, X., Srinivasan, R., & Bosch, D. (2009). Calibration and uncertainty analysis of the SWAT model using Genetic Algorithms and Bayesian Model Averaging.

Journal of Hydrology, 374(3), 307–317.

<http://doi.org/10.1016/j.jhydrol.2009.06.023>

Zhao, L., Duan, Q., Schaake, J., Ye, A., & Xia, J. (2011). A hydrologic post-processor for ensemble streamflow predictions. *Advances in Geosciences*, 29, 51–59.

<http://doi.org/10.5194/adgeo-29-51-2011>

Zhao, T., Wang, Q. J., Bennett, J. C., Robertson, D. E., Shao, Q., & Zhao, J. (2015).

Quantifying predictive uncertainty of streamflow forecasts based on a Bayesian joint probability model. *Journal of Hydrology*, 528, 329–340.

<http://doi.org/10.1016/j.jhydrol.2015.06.043>

Chapter 3. Assessing Hydrologic Uncertainty Processor Performance for Flood Forecasting in a Semiurban Watershed

Summary of Paper 2: Han, S., Coulibaly, P. and Biondi, D. (2019). Assessing Hydrologic Uncertainty Processor Performance for Flood Forecasting in a Semiurban Watershed. *Journal of Hydrologic Engineering*, 24(9), 05019025.

This research work applied the precipitation-dependent Hydrologic Uncertainty Processor (HUP) in a semiurban watershed to quantify the hydrologic uncertainty with flood forecast, and compared the performance of HUP with different hydrologic models under different flow conditions.

Key findings of this research include:

- HUP is able to correct the deterministic forecast from the hydrologic model, and produces more accurate probabilistic forecast with quantification of hydrologic uncertainty.
- For low peak flow events, HUP combining with different hydrologic models show comparable performance.
- For high peak flow events, the better deterministic forecast is yielded from the hydrologic model, the better probabilistic forecast is produced by applying HUP with that hydrologic model.

3.1 Abstract

A key challenge in enhancing flood forecast relies in the difficulty of reducing predictive uncertainty. The Precipitation-Dependent Hydrologic Uncertainty Processor (HUP) is a flexible model independent Bayesian processor that can be used with any hydrologic model to provide probabilistic forecast. This study investigates the use of HUP with different hydrologic models for hydrologic uncertainty quantification in a flood forecasting scheme for a semi-urban watershed of southern Ontario (Canada). The purpose is to better understand predictive uncertainty and enhance flood forecasting system reliability in semi-urban conditions. HUP is based on Bayes' theorem, it updates the prior distribution given available information at the forecast time to obtain the posterior distribution that is close to future unknown actual value. In this study, HYMOD and GR4H were selected to work with HUP, and the Bayesian processor was calibrated using a number of selected flood events from 2005 to 2014. The performance of the processor was assessed by graphical tools and performance metrics, like reliability plots, Nash Sutcliffe efficiency (NSE), and continuous ranked probability score (CRPS). Results show that HUP provides a robust framework and a reliable analytic-numerical method for the quantification of hydrologic uncertainty, the actual values are well captured by the uncertainty bounds, the CRPS values are relatively small, and reliability curves lie close to the bisector. The comparison between the NSE calculated from the output of the sole deterministic hydrologic model (HYMOD/GR4H) and from the median of the predictive distribution produced by HUP-HYMOD/HUP-GR4H, demonstrates that HUP has the ability to improve the deterministic forecast. For low peak flow events, HUP

combining with different hydrologic models presents similar predictive performance, while for high peak flow events, a well performed deterministic model is required in HUP to produce an accurate probabilistic forecast.

Author keywords: Hydrologic Uncertainty Processor; uncertainty quantification; posterior distribution; Bayes theorem; flood forecasting

3.2 Introduction

Over the past few decades in Canada, the frequency of floods has been increasing, and the estimated total costs of large flood events could exceed 5 billion dollars over a ten year period (The City of Windsor 2012). Therefore, in the research field of non-structural flood management, it is extremely important to address flood forecasting challenges and enhance flood forecasting systems. Accurate quantification of uncertainties associated with deterministic forecast is one of these challenges. As deterministic forecast that gives a point estimate of the predictand through model simulation can have limited value to decision-makers (Reggiani and Weerts 2008), the scientific hydrologic community shows an increasing interest in probabilistic forecast, which not only predicts the point forecast but also expresses the associated degree of confidence under uncertainty.

Uncertainties within flood forecasting are manifold, they could arise from various sources, including model uncertainty (uncertainty of model parameter and model structure), measurement uncertainty, uncertainty associated with initial condition, insufficient and incomplete data input, uncertainty of future precipitation and temperature, etc. Although uncertainty quantification (UQ) associated with the forecast is an essential support to

make an effective decision, sometimes uncertainties are difficult to interpret probabilistically and are not always fully and accurately quantified.

Research in recent years has been significantly focused on the quantification and reduction of uncertainty and many methodologies have been introduced. Approaches that are suitable for predictive uncertainty quantification include Data-Based Mechanistic (DBM) with Kalman Filter (Young 2002; Romanowicz et al. 2006), Bayesian Joint Probability (BJP) (Wang et al. 2009; Wang and Robertson 2011; Zhao et al. 2015), Bayesian Forecasting System (BFS) proposed by Krzysztofowicz (1999), Model Conditional Processor (MCP) by Todini (2008, 2013), Data Assimilation (DA) (Dechant and Moradkhani 2012; Moradkhani et al. 2012; Salamon and Feyen 2009; Vrugt and Robinson 2007), Quantile Regression (QR) (Weerts et al. 2011), Non-Gaussian Copulas Approach (Madadgar and Moradkhani 2014; Schoelzel and Friederichs 2008), Non-Parametric Databased Approach (Van Steenbergen et al. 2012), Grey Neural Networks (Alvisi and Franchini 2012), Bayesian Neural Network (BNN) (Khan and Coulibaly 2006; Zhang et al. 2009, 2011), Hydrologic Model Output Statistics (HMOS) (Regonda et al. 2013), and etc. The predictive uncertainty quantification techniques have been largely improved, however, there are still some aspects that need further investigation. Most of the approaches quantify the predictive uncertainty in a lumped way, the contribution of each individual source of error to the total uncertainty is not clear yet, and the understanding of the dominant individual uncertainties needs to be improved (Zhang et al. 2011).

Among the predictive uncertainty quantification approaches, BFS provides a robust and flexible theoretical framework for probabilistic forecasting, considers all the major sources of uncertainties and can work with any deterministic hydrologic model. BFS breaks down the total uncertainty into precipitation uncertainty and hydrologic uncertainty, the hydrologic uncertainty processor (HUP) is a component of BFS that specifically quantifies hydrologic uncertainty and provides probabilistic forecast under the assumption of a perfect forecast of precipitation amount (Liu et al. 2018). A series of papers have described the theory of each BFS component (Kelly and Krzysztofowicz 2000; Krzysztofowicz 1999, 2001; Krzysztofowicz and Herr 2001; Krzysztofowicz and Kelly 2000), and several studies have presented the application of BFS for river stage forecasting (Herr and Krzysztofowicz 2010, 2015; Krzysztofowicz 2002) as well as for real-time flood forecasting (Reggiani and Weerts 2008; Reggiani et al. 2009; Biondi et al. 2010; Biondi and De Luca 2013, 2012).

BFS has been applied for limited types of watersheds at daily or hourly time scale and its performance on other types of watersheds needs to be further investigated. This study focuses on application of the HUP component to a semi-urban watershed encompassing hundreds of square kilometers at hourly time step. Semi-urban watersheds include both rural areas which are open and sparsely populated and urban areas which are heavily built-up and densely populated. Humber River watershed, the watershed selected for the present study, is comprised of half urban area in downstream and half rural area located in upstream. During high intensity precipitation events, water level in urban area rises rapidly due to the impervious surface, and then the increased flow in urban downstream

is fed by the water stored in rural area upstream, often resulting in extended flood duration. The complexity and variability of the rainfall-runoff response in semi-urban watersheds, due to mixture of urban and rural lands, is likely to lead to multiple flow peaks, and increase prediction uncertainty (Fletcher et al. 2013). In addition, enhancing the reliability of flood forecasts through better uncertainty estimate is of particular need in populated semi-urban watersheds.

Previous studies have focused on the application of HUP to assess hydrologic uncertainty (Krzysztofowicz and Herr 2001; Krzysztofowicz and Kelly 2000; Biondi et al. 2010; Reggiani and Weerts 2008; Reggiani et al. 2009; Liu et al. 2016; Liu et al. 2018), but very few studies have compared the performance of HUP working with different rainfall-runoff models. There are many post-processing approaches that aim to improve flood predictions considering the information content from the outputs of different hydrological models. The Bayesian Model Averaging (BMA), introduced by Raftery (1993), has been frequently applied in hydrology to combine forecasts from different hydrological models and estimate predictive uncertainty as a weighted mean of the predictive distributions of individual models. Other approaches that make use of the model-based ensemble information have been proposed to estimate the predictive uncertainty distribution, e.g. the already mentioned QR and MCP as well as the EMOS (Ensemble Model Output Statistics) by Gneiting et al. (2005) and the non parametric approach by Brown and Seo (2010). A recent review paper by Han and Coulibaly (2017) considers most of the proposed Bayesian approaches, while a comprehensive review of the commonly used

statistical post-processing methods for uncertainty estimation of meteorological (e.g. precipitation) and hydrological (e.g. streamflow) forecasts, is provided in Li et al. (2017).

This contribution aims at investigating the use of HUP with different hydrologic models in a semi-urban watershed, and provides a comprehensive analysis of HUP performance under different flow conditions. The objectives of this study are: (i) to accurately quantify hydrologic uncertainty associated with flood forecast through a precipitation-dependent HUP given the information available at the forecast time and assumption of perfect precipitation input; (ii) to compare the predictability of HUP combined with different deterministic hydrologic models under different lead times; (iii) to assess the predictive performance of HUP for different magnitudes of peak flow events (low peak flow event, medium peak flow event, high peak flow event).

3.3 Methodology

3.3.1 Hydrologic Uncertainty Processor (HUP)

3.3.1.1 Background of the Bayesian Processor

BFS consists of three components: precipitation uncertainty processor (PUP), hydrologic uncertainty processor (HUP) and integrator (INT). Precipitation uncertainty (related to the forecast of the total average precipitation amount) is quantified in PUP under the hypothesis of nonexistence of hydrologic uncertainty; the purpose of the HUP, a Bayesian processor, is to quantify the hydrologic uncertainty (related to the aggregate of all sources of error other than precipitation amount error, including spatial and temporal downscaling of precipitation forecast, and measurement and estimation errors of other

model inputs) associated with flood forecast under the assumption that precipitation uncertainty is zero; then these two uncertainties are combined in INT (Krzysztofowicz and Herr 2001). Based on Bayes' theorem, i.e. combining a prior distribution with a likelihood function, HUP outputs a posterior distribution conditional on initial states and deterministic forecasts, that is able to provide a complete characterization of uncertainty (Liu et al. 2018).

Two versions of HUP can be distinguished: a precipitation-independent HUP and a precipitation-dependent HUP. Precipitation-independent HUP is a one branch processor where all the hydrologic processes are analyzed together, while precipitation-dependent HUP is a processor with two branches, each conditional on the precipitation occurrence: if precipitation occurs, the hydrological process will be allocated to and analyzed in the first branch; if there is no precipitation, the hydrological process will be assigned to the second branch. Some studies found that hydrologic uncertainty grew under occurrence of precipitation, as more components in hydrologic models are active when precipitation occurs and the precipitation-dependent HUP turned out to be more efficient and informative than precipitation-independent HUP (Krzysztofowicz and Herr 2001), hence the precipitation-dependent HUP was employed in this study.

Define V as an indicator of precipitation occurrence or non-occurrence over the forecast period, it is binary with $V = 0$ indicating basin precipitation amount is equal to 0 and $V = 1$ indicating precipitation amount is larger than 0. Define n ($n = 1, 2, \dots, N$) as the lead time in units of hours, and time step $t_n - t_{n-1}$ is fixed at a value of 1 hour. Let h_n denote the actual river discharge at the outlet of the basin for every lead time $n \in \{1, \dots, N\}$; at time

t_0 , h_0 is the observed river discharge, the actual river discharge at time t_n is uncertain and thus treated as a random variable H_n . Similarly, let s_n denote the model river discharge for every lead time n , produced from the hydrologic model to estimate H_n and stands for a realization of variate S_n . Supposing there is no hydrologic uncertainty, s_n would be expected to equal h_n for every n , the impact of hydrologic uncertainty is to give rise to a probability distribution of the actual river discharge H_n .

For each precipitation indicator v ($v = 0, 1$) and every lead time n ($n = 1, \dots, N$), the HUP gives a family of posterior distributions Φ_{nv} and posterior densities ϕ_{nv} for H_n . The prior density g of H_n exists before the forecast, and is characterized in terms of marginal density and a family of transition densities. The posterior density of actual river discharge H_n can be given by the Bayes theorem via the revision of the prior density on the basis of all the available information. Conditional on the model river discharge s_n and observed river discharge h_0 at time t_0 , the posterior density ϕ_{nv} takes the form (Krzysztofowicz 1999):

$$\phi_{nv}(h_n|s_n, h_0) = \frac{f_{nv}(s_n|h_n, h_0)g_{nv}(h_n|h_0)}{\kappa(s_n|h_0)} \quad (3-1)$$

Where f is the likelihood function derived by looking at the relationship between predictands and observations, and the expected density κ of model river discharge S_n , conditional on the observed river discharge h_0 , can be obtained from Eq. (3-2) below by combining g_{nv} and f_{nv} into a Bayesian revision process (Krzysztofowicz 1999):

$$\kappa(s_n|h_0) = \int_{-\infty}^{\infty} f_{nv}(s_n|h_n, h_0)g_{nv}(h_n|h_0)dh_n \quad (3-2)$$

More detailed information about the formula derivation can be found in (Krzysztofowicz and Kelly 2000; Krzysztofowicz 2002; Krzysztofowicz and Herr 2001). As for the steps about how to implement the precipitation-dependent HUP, they will be discussed in detail below.

3.3.1.2 Normal Quantile Transform (NQT)

A meta-Gaussian approach (Krzysztofowicz and Kelly 2000; Krzysztofowicz and Herr 2001) was adopted in this study to apply the precipitation-dependent HUP. Within this meta-Gaussian model, each variate, H_n and S_n , was transformed into normally distributed variate, W_n and X_n , respectively. In the transformed space, the variates are assumed to follow Gaussian distribution and the stochastic dependence structure is modelled in terms of normal linear equations, which makes it easy for regression and parameter estimation. Then the estimation results are transformed back into the original space, resulting in meta-Gaussian distributions.

Based on the matching between the historical discharge record and the discharge output from the hydrologic prediction, a joint sample $\{(v; s_1, \dots, s_N; h_0, h_1, \dots, h_N)\}$ can be formed. For every $v \in \{0, 1\}$ and every $n \in \{0, 1, \dots, N\}$, define Γ_{nv} as marginal prior distribution of H_n , and $\bar{\Lambda}_{nv}$ as marginal initial distribution of S_n . The corresponding densities are represented by γ_{nv} and $\bar{\lambda}_{nv}$. Once the appropriate marginal distribution type is determined, each original variate is transformed into a normal variate through the standard normal inverse Q^{-1} of the marginal distribution, the process is called normal quantile transform (NQT) and presented by these equations (Krzysztofowicz and Kelly 2000):

$$W_n = Q^{-1}(\Gamma_{nv}(H_n)), n = 0, 1, \dots, N \quad (3-3)$$

$$X_n = Q^{-1}(\bar{\Lambda}_{nv}(S_n)), n = 0, 1, \dots, N \quad (3-4)$$

3.3.1.3 Dependence Parameters

In the transformed space, the stochastic dependence structures of transition density and likelihood function are characterized by two linear regressions. By looking at the relationship between actual river discharge at time t_n and t_{n-1} in normal space, the dependence parameters of the transition densities c_{nv} ($v = 0, 1; n = 1, \dots, N$) can be defined as (Krzysztofowicz and Herr 2001):

$$E(W_n | W_{n-1} = w_{n-1}, V = v) = c_{nv} w_{n-1} \quad (3-5)$$

$$Var(W_n | W_{n-1} = w_{n-1}, V = v) = 1 - c_{nv}^2 \quad (3-6)$$

The parameters c_{nv} are regression constants, and the residual is statistically independent of W_{n-1} and normally distributed with zero mean and variance $1 - c_{nv}^2$.

By looking at the distribution of the model river discharge at time t_n conditional on actual river discharge at t_n and actual river discharge at the forecast time t_0 in the normal space, the dependence parameters of the likelihood function a_{nv} , b_{nv} , d_{nv} and σ_{nv} ($v = 0, 1; n = 1, \dots, N$) can be defined as (Krzysztofowicz and Herr 2001):

$$E(X_n | W_n = w_n, W_0 = w_0, V = v) = a_{nv} w_n + d_{nv} w_0 + b_{nv} \quad (3-7)$$

$$Var(X_n | W_n = w_n, W_0 = w_0, V = v) = \sigma_{nv}^2 \quad (3-8)$$

The parameters a_{nv} , b_{nv} and d_{nv} are regression constants, and the residual is statistically independent of other variables and normally distributed with zero mean and variance σ_{nv}^2 .

3.3.1.4 Prior Density and Prior Distribution

Given the prior dependence parameters c_{nv} defined in Eq. (3-5), the parameters that characterize meta-Gaussian prior density and prior distribution can be calculated (Krzysztofowicz and Herr 2001):

$$C_{nv} = \prod_{i=1}^n c_{iv} \quad (3-9)$$

$$t_{nv}^2 = 1 - C_{nv}^2 \quad (3-10)$$

The meta-Gaussian prior distribution of actual river discharge H_n on the n th hour, conditional on precipitation event $V = v$, and given the observed river discharge $H_0 = h_0$ at the forecast time, can be derived (Krzysztofowicz and Herr 2001):

$$G_{nv}(h_n|h_0) = Q\left(\frac{Q^{-1}(\Gamma_{nv}(h_n)) - C_{nv}Q^{-1}(\Gamma_{0v}(h_0))}{t_{nv}}\right) \quad (3-11)$$

And the corresponding meta-Gaussian prior density takes the form (Krzysztofowicz and Herr 2001):

$$g_{nv}(h_n|h_0) = \frac{\gamma_{nv}(h_n)q(Q^{-1}(G_{nv}(h_n|h_0)))}{t_{nv}q(Q^{-1}(\Gamma_{nv}(h_n)))} \quad (3-12)$$

Where, Q is the standard normal distribution, Q^{-1} is its inverse and q stands for its corresponding density.

3.3.1.5 Posterior Density and Posterior Distribution

Given the parameters defined in Eqs. (3-7) - (3-10), the dependence parameters used in posterior density and posterior distribution A_{nv} , B_{nv} , D_{nv} and T_{nv} can be calculated as follows (Krzysztofowicz and Herr 2001):

$$A_{nv} = \frac{a_{nv} t_{nv}^2}{a_{nv}^2 t_{nv}^2 + \sigma_{nv}^2} \quad (3-13)$$

$$B_{nv} = \frac{-a_{nv} b_{nv} t_{nv}^2}{a_{nv}^2 t_{nv}^2 + \sigma_{nv}^2} \quad (3-14)$$

$$D_{nv} = \frac{C_{nv} \sigma_{nv}^2 - a_{nv} d_{nv} t_{nv}^2}{a_{nv}^2 t_{nv}^2 + \sigma_{nv}^2} \quad (3-15)$$

$$T_{nv}^2 = \frac{t_{nv}^2 \sigma_{nv}^2}{a_{nv}^2 t_{nv}^2 + \sigma_{nv}^2} \quad (3-16)$$

The meta-Gaussian posterior distribution of actual river discharge H_n on the nth hour, conditional on precipitation event $V = v$ and the model river discharge $S_n = s_n$ which output from the hydrologic model based on a perfect precipitation input, and given the observed river discharge $H_0 = h_0$ at the forecast time, can be derived (Krzysztofowicz and Herr 2001):

$$\begin{aligned} & \Phi_{nv}(h_n | s_n, h_0) \\ &= Q \left(\frac{Q^{-1}(\Gamma_{nv}(h_n)) - A_{nv} Q^{-1}(\bar{\Lambda}_{nv}(s_n)) - D_{nv} Q^{-1}(\Gamma_{0v}(h_0)) - B_{nv}}{T_{nv}} \right) \end{aligned} \quad (3-17)$$

And the corresponding meta-Gaussian posterior density takes the form (Krzysztofowicz and Herr 2001):

$$\phi_{nv}(h_n|s_n, h_0) = \frac{\gamma_{nv}(h_n)q\left(Q^{-1}(\Phi_{nv}(h_n|s_n, h_0))\right)}{T_{nv}q\left(Q^{-1}(\Gamma_{nv}(h_n))\right)} \quad (3-18)$$

HUP can be used independently with any operational flood forecasting system; it offers an analytic expression for probabilistic forecast distribution considering hydrologic uncertainty. The calibration of HUP can be done off-line, and the estimated parameters are ready to be used in real-time forecasting (Krzysztofowicz and Kelly 2000; Krzysztofowicz and Herr 2001).

3.3.2 Rainfall-Runoff Models

The HUP processor is a post-processor of a deterministic hydrologic model that statistically analyzes the model output and observations based on Bayes' rule in order to produce probabilistic forecast. In this study, the rainfall-runoff models selected to cooperate with HUP are the HYdrological MODel (HYMOD) and the modèle du Génie Rural à 4 paramètres Horaire (GR4H).

3.3.2.1 HYMOD

The HYMOD model was introduced by Boyle in (2001) for hydrological model uncertainty analysis. It is a lumped conceptual rainfall-runoff model, widely used in the literature, that despite its simplicity has proven to provide very good results (Quan et al. 2015; Sun et al. 2015). The model considers a runoff generation process based on a simple rainfall excess model as in the probability-distributed theory proposed by Moore (1985). The watershed is regarded as composed of infinite independent units and the

distribution function of the spatial variability of the water storage capacity within the basin is defined as (Boyle 2001):

$$F(C) = 1 - \left(1 - \frac{C}{C_{max}}\right)^{beta} \quad (3-19)$$

Where, F is the cumulative probability of a given storage capacity; C is the water storage capacity (mm) while C_{max} is the maximum water storage capacity (mm), and $beta$ is the degree of spatial variability within this basin. The amount of the water storage depends on the rate of precipitation and evapotranspiration over a certain time period; when the water storage is over the maximum water storage capacity, the excess of water is treated as runoff. For each time step, the actual evapotranspiration is equal to potential evapotranspiration if enough water is stored, otherwise it equals the available stored water. The runoff volume is divided into overland flow and subsurface flow according to the partition coefficient α . The overland flow is considered as quick flow and routed through three identical linear reservoirs to the outlet, the flow rate in this routing is described by the recession coefficient K_{quick} of each reservoir. While the subsurface flow is considered as slow flow, and goes through one parallel reservoir, with recession coefficient K_{slow} , to the outlet of the basin. Eventually, the quick flow and slow flow arriving at the outlet at that time step are combined to estimate the total streamflow of the watershed outlet. The description and range of model parameters are presented in Table 3-1, and more details about the model structure can be found in Sun et al. (2015).

3.3.2.2 GR4H

GR4H is an hourly lumped rainfall-runoff model, first introduced by Mathevet in (2005) for flood estimation in headwater basins. It is derived from the modèle du Génie Rural à 4 paramètres Journalier (GR4J), which is a daily lumped continuous rainfall-runoff model (Perrin et al. 2003). GR4H and GR4J share similar model structure. GR4H includes two major modules and four parameters to calibrate: $x1$ and $x2$ for production module; $x3$ and $x4$ for routing module. The description of model parameter is presented in Table 3-1.

In the production module, rainfall amount and potential evapotranspiration are subtracted to determine the net rainfall P_n and net evapotranspiration E_n , the fraction P_s of P_n that goes to the production store S is determined by a soil moisture accounting store (SMA), and the remaining amount $P_n - P_s$ is the effective rainfall. Percolation leakage $Perc$ coming from the production store plus the effective rainfall define the water amount that finally reaches the routing process $P_r = Perc + (P_n - P_s)$. The production store level is then updated to serve as the initial production store value for the next time step. In the routing module, P_r is divided into two flow components: 90% indirect flow being routed by a unit hydrograph UH and a nonlinear routing store R, and 10% direct flow being routed via a single unit hydrograph. A water loss or gain function F is applied to both components to find out the groundwater exchange along the way. The total streamflow Q is finally given by the sum of the flow coming from the two branches: output of indirect flow Q_r and output of direct flow Q_d . The differences between GR4H and GR4J are the percolation rate in the production function and the number of UH in the routing function

(Andreassian et al. 2006). More details about the model structure can be found in Mathevet (2005).

Table 3-1 Model parameters of HYMOD and GR4H

Parameter	Description	Unit	Range	Calibrated parameters	
				optNVE	optKGE
HYMOD					
<i>C_{max}</i>	Maximum storage capacity	mm	10-700	261.41	273.25
<i>Beta</i>	Degree of spatial variability of the soil moisture capacity	--	0.1-2	1.58	1.69
<i>Alpha</i>	Factor distributing the flow between slow and quick release reservoirs	--	0.2-0.99	0.35	0.21
<i>K_{slow}</i>	Residence time of the slow release reservoir	day	0.001-0.1	0.01	0.02
<i>K_{quick}</i>	Residence time of the quick release reservoirs	day	0.01-0.99	0.17	0.28
GR4H					
<i>x₁</i>	Maximal capacity of production store	mm	1-2000	1387.77	1247.38
<i>x₂</i>	Water exchange coefficient	mm	-10-5	-0.66	-0.59
<i>x₃</i>	Maximal capacity of the routing store	mm	1-500	2.38	2.33
<i>x₄</i>	Time parameter of unit hydrograph	hour	0.5-96	21.63	20.33

3.3.2.3 Models Setup

HYMOD and GR4H were set up using event-based approach, the selection of events are presented in the following section. The two rainfall-runoff models were calibrated for each calibration event independently to obtain the optimized parameter sets, the average values of the parameter sets were calculated as the final calibrated parameters which used for validation events. Particle swarm optimization (PSO) was employed as the

optimization algorithm based on a previous study (Razavi and Coulibaly 2016). Two objective functions were used: NVE (combined Nash Sutcliffe efficiency and volume error) (Samuel et al. 2011) and KGE (Kling-Gupta efficiency) (Gupta et al. 2009). The closer NVE or KGE is to 1, the better the model is performing.

NVE considers both low and high peak flow (Samuel et al. 2011):

$$NVE = 0.5NSE - 0.1VE + 0.25NSE_{log} + 0.25NSE_{sqr} \quad (3-20)$$

Where NSE is Nash Sutcliffe efficiency, and VE is volume error. NSE_{log} is NSE based on the logarithm of discharge, NSE_{sqr} is NSE based on square of discharge, and they can be calculated as below (Samuel et al. 2011):

$$NSE_{log} = 1 - \left[\frac{\sum_{i=1}^N (\log Q_{sim} - \log Q_{obs})^2}{\sum_{i=1}^N (\log Q_{obs} - \overline{\log Q_{obs}})^2} \right] \quad (3-21)$$

$$NSE_{sqr} = 1 - \left[\frac{\sum_{i=1}^N (Q_{sim}^2 - Q_{obs}^2)^2}{\sum_{i=1}^N (Q_{obs}^2 - \overline{Q_{obs}^2})^2} \right] \quad (3-22)$$

Q_{sim} and Q_{obs} are simulated and observed discharge, respectively. $\overline{Q_{obs}}$ is the mean value of observed discharges, and N is the number of discharge values.

NSE and mean squared error (MSE) can be decomposed into three components, representing the correlation, the bias and a measure of variability. KGE is formulated through calculating the Euclidian distance of the three components from the ideal point (Gupta et al. 2009):

$$KGE = 1 - \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (3-23)$$

Where r is the linear correlation coefficient between observed and simulated discharge, α is calculated as the ratio between standard deviation in simulated values and standard deviation in observed values, β is the ratio between mean simulated and mean observed discharge.

3.4 Study Area and Data

The Humber River Watershed, located in Southern Ontario, Canada (Figure 3-1), was selected as a case study to apply the HUP for flood forecasting because it is identified as a flood-prone area and the risk of flooding is a key issue in some regions of this watershed. The watershed covers 911 km², flowing from the headwaters on the Niagara Escarpment and Oak Ridges Moraine down through flat clay, till plains and marshes to the Lake Ontario. Elevation in the watershed varies from 75 to 488 m a.s.l., the northwestern part is characterized by hummocky terrain and steep slopes, while the central part is relatively flat. As of 2015, land use within the Humber River Watershed was approximately 54% rural, 33% urban and remaining 13% urbanizing, with 32% under natural cover (17% forest, 9% meadow, 3% successional and 2% wetland) (TRCA 2008).

A continental climate moderated by Great Lakes is found in the Humber Watershed, the climate characteristic is affected by cold dry air masses from the north and warm moist masses from the south. Based on the historical climate records, the average annual precipitation ranges from 798 to 933 mm, and the mean annual temperature ranges between 6.5 and 8 °C. The fluctuations of precipitation and temperature depend on the

elevation and the distance to Lake Ontario. As for the mean annual evapotranspiration, it ranges from 469 mm along the shoreline of the lake, to 517 mm in higher elevation areas (TRCA 2008). Heavy precipitation events in this region occurred more often in summer time from June to August.

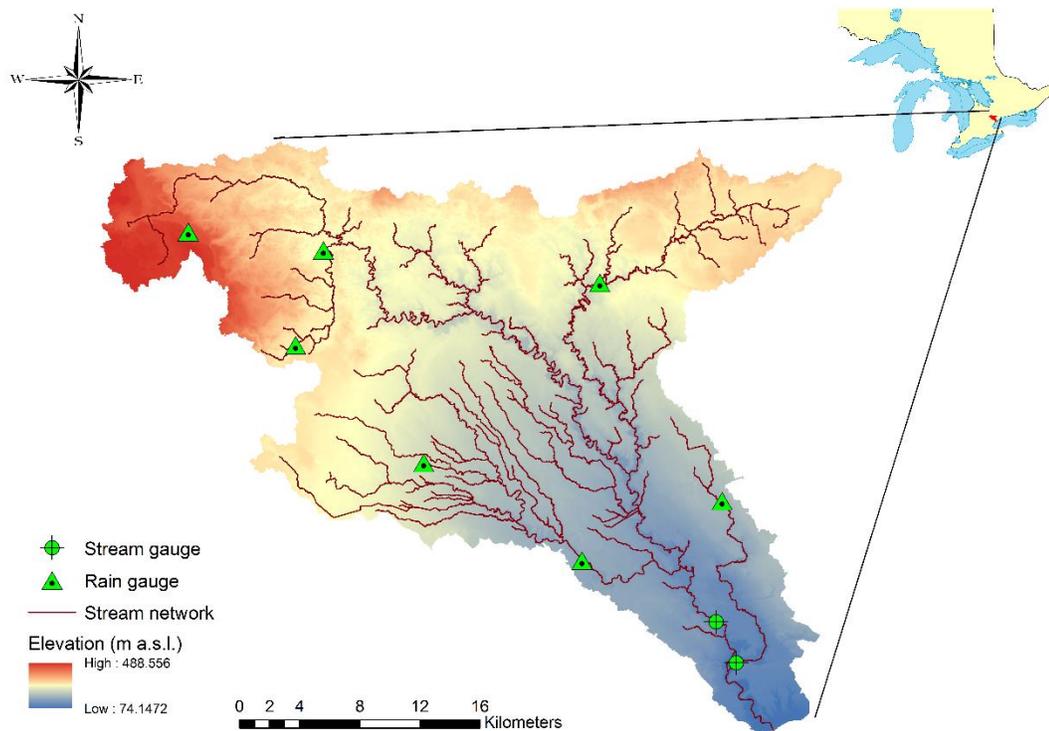


Figure 3-1 Humber River Watershed in Southern Ontario, Canada

The two hydrologic models employed with the HUP, namely HYMOD and GR4H, were calibrated against hourly observed discharge using hourly precipitation and evapotranspiration as inputs. Evapotranspiration was calculated based on temperature through an adjusted potential evapotranspiration model proposed by Oudin et al. (2005a,

2005b). Precipitation and temperature data were provided by Toronto and Region Conservation Authority (TRCA). The discharge data were from Water Survey of Canada. The available period of the dataset was from 2005 until 2014. Thiessen polygon method was used to calculate mean areal precipitation; the flows from the two stream gauges near the outlet, Humber River at Weston and Black Creek near Weston, were considered to estimate the total outflow of the Humber Watershed.

Based on the available period of historic record, precipitation events that have been noted to have produced relatively high flows (total rainfall amount larger than 20 mm) were selected as candidate flood events to calibrate the model and assess the predictive performance of the processor (TRCA and AMEC 2012). Main characteristics of the 24 selected flood events are presented in Table 3-2, including initial discharge, peak discharge, runoff volume, and rainfall depth. Runoff volume is defined as the cumulative flow volume during the flood event (Reddy and Ganguli 2012), and was estimated using the formula from Yue (2001). The average rainfall depth of these events is 54.20 mm, and the peak discharge ranges from 27.01 to 355.23 m³/s. For some events, the flow reaches the peak in a few hours, leading to potential flooding in this region. Based on the 25% (50.59 m³/s) and 75% (90.14 m³/s) quantile of the empirical distribution of the peak discharges for the available sample, the events were classified as high peak flow event ($Q_{peak} \geq 90.14 \text{ m}^3/\text{s}$), medium peak flow event ($50.59 \text{ m}^3/\text{s} < Q_{peak} < 90.14 \text{ m}^3/\text{s}$) and low peak flow event ($Q_{peak} \leq 50.59 \text{ m}^3/\text{s}$). With the aim of having sets of independent events, according to the dates of the events (at least one event per year for calibration, and the validation events are from different year) and their hydrograph characteristics

including the magnitude of peak flow and shape of hydrography (validation events were selected to cover different types of events), 20 events were chosen for model calibration and 4 events (event 4, event 7, event 14 and event 23) were chosen for validation.

Table 3-2 Main characteristics of the selected flood events

Event ID	Date (mm/dd/yyyy)	Q_{peak} (m ³ /s)	Q_0 (m ³ /s)	Runoff volume (10 ⁶ m ³)	Rainfall depth (mm)
1	8/19/2005	282.43	3.90	16.81	53.30
2	8/31/2005	49.86	4.32	3.01	24.01
3	9/26/2005	49.56	3.75	4.10	34.90
<u>4</u>	11/15/2005	50.59	4.56	11.53	48.21
5	4/23/2006	70.00	4.79	14.16	36.56
6	7/10/2006	81.05	3.04	14.29	66.74
<u>7</u>	10/17/2006	56.25	3.64	15.23	64.23
8	5/15/2007	81.05	3.85	13.20	47.12
9	7/23/2008	90.14	4.18	14.81	83.37
10	8/9/2008	80.17	4.04	10.40	55.21
11	8/9/2009	46.87	4.56	9.88	56.66
12	8/20/2009	62.72	2.60	6.21	19.94
13	5/7/2010	83.12	4.52	7.73	37.64
<u>14</u>	7/23/2010	104.11	4.49	10.44	43.93
15	9/28/2010	50.30	3.03	7.50	41.43
16	5/14/2011	90.10	4.98	20.32	64.21
17	10/20/2011	90.61	4.91	17.43	75.63
18	11/29/2011	147.24	4.98	26.36	75.22
19	10/27/2012	53.83	4.76	20.99	77.98
20	5/21/2013	27.01	4.90	5.17	34.18
21	5/29/2013	200.21	4.88	17.28	64.53
22	7/8/2013	355.23	4.74	36.14	81.90
<u>23</u>	7/27/2014	67.31	2.56	8.87	29.83
24	9/10/2014	76.75	3.79	16.33	84.12

Validation events are underlined

3.5 Results and Discussion

3.5.1 Rainfall-Runoff Model Calibration

The optimized parameter sets obtained independently for each calibration event by optimizing NVE (optNVE) and KGE (optKGE) are presented in Figure 3-2 on a logarithmic scale. The final calibrated model parameters, chosen as the average parameter values of the calibration events, are listed in Table 3-1. In Figure 3-2, the red line in the middle of the boxplot stands for the median value, the central box represents the interquartile range of the data, the whiskers above and below the box show the location of the maximum and minimum, and the red crosses are outliers. Comparison between optNVE and optKGE indicates that for some parameters the box plots have similar distribution but slightly different median, while for other they have similar median but different distribution. For HYMOD in both cases, the box plots for C_{max} and K_{slow} are relatively tall, indicating a wider distribution range, and the box plots for $Beta$ are comparatively short, indicating a high level of consistency for different events. In the same way for GR4H, the box plots for x_2 are relatively short and the box plots for the other three parameters are tall. Optimal parameter values for different events using different objective functions could be very different, selecting the best model parameter set for future flood event forecast becomes tricky, therefore error in model parameter estimation is a large contributor to hydrologic uncertainty.

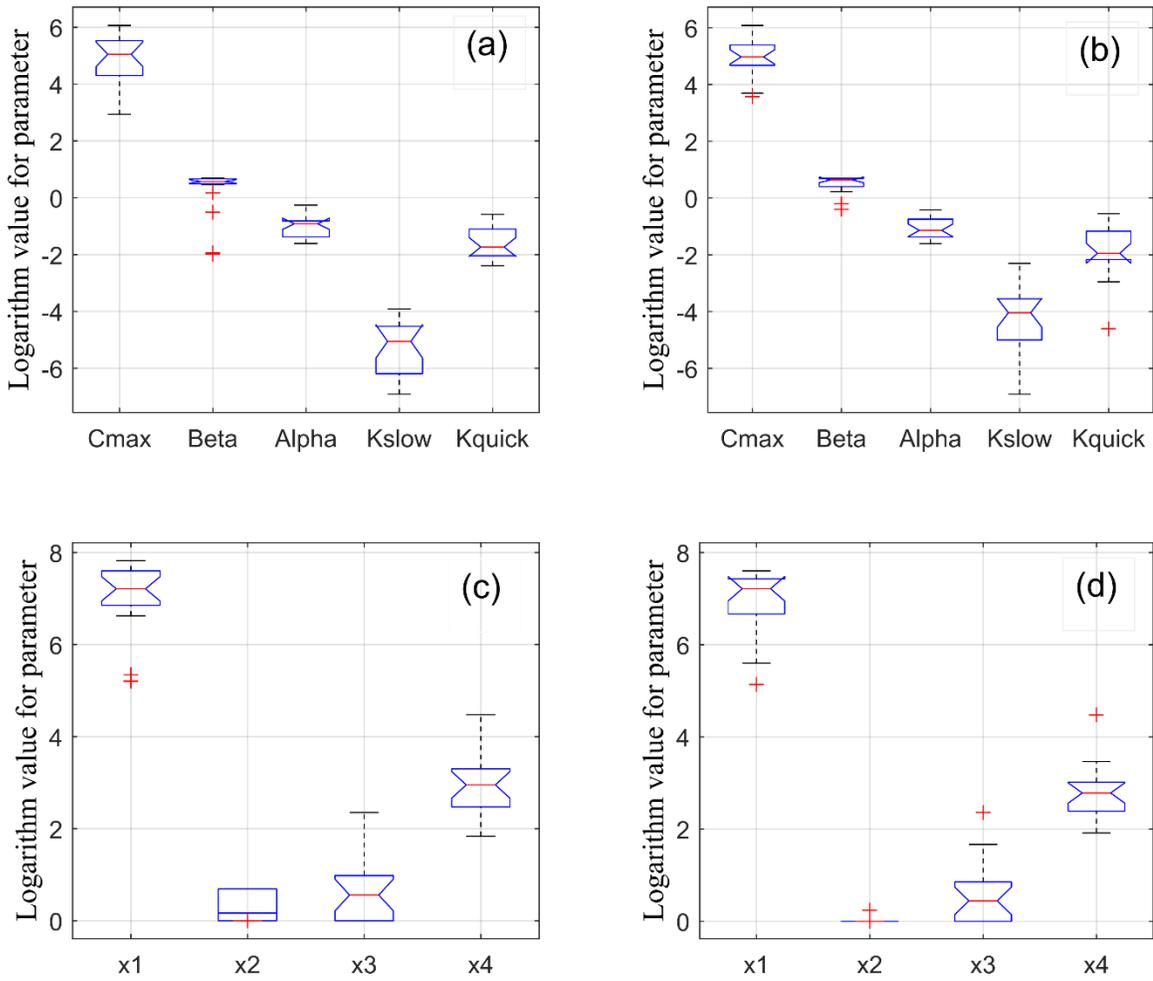


Figure 3-2 Model parameter estimation and uncertainty: (a) HYMOD optNVE; (b) HYMOD optKGE; (c) GR4H optNVE; and (d) GR4H optKGE. Box plots show the spread of optimized parameters for 20 calibration events.

To assess performances for these 4 scenarios, different models calibrated using different objective functions, were evaluated in terms of NSE, modified peak flow criterion (MPFC), percentage error for peak flow and runoff volume. The comparisons of the scatterplots can be found in Figure 3-3. The MPFC was modified from PFC (Coulibaly et al. 2001) and computed by $MPFC=1-PFC$. MPFC statistics are specifically for assessing

accuracy of peak flow forecast and a value equal to 1 indicates a perfect fit. The NSE values for HYMOD range from 0.57 to 0.95 while for GR4H range from 0.33 to 0.85. The peak flow differences between the observed value and simulated value using HYMOD are within -30% and 10%, while the peak flow differences using GR4H are within -60% and 20%. The average MPFC values for the 4 scenarios (HYMOD optNVE, HYMOD optKGE, GR4H optNVE, GR4H optKGE) are 0.76, 0.78, 0.73 and 0.74, respectively. The runoff volume differences for HYMOD varies between -20% and 20%, and for GR4H varies between -30% and 10%. Overall, the calibration and validation results show that HYMOD performed much better than GR4H. The use of objective function KGE gave slightly better performance than objective function NVE for event-based simulation. The two calibrated models using objective function KGE were then passed to HUP to analyze the hydrologic uncertainty.

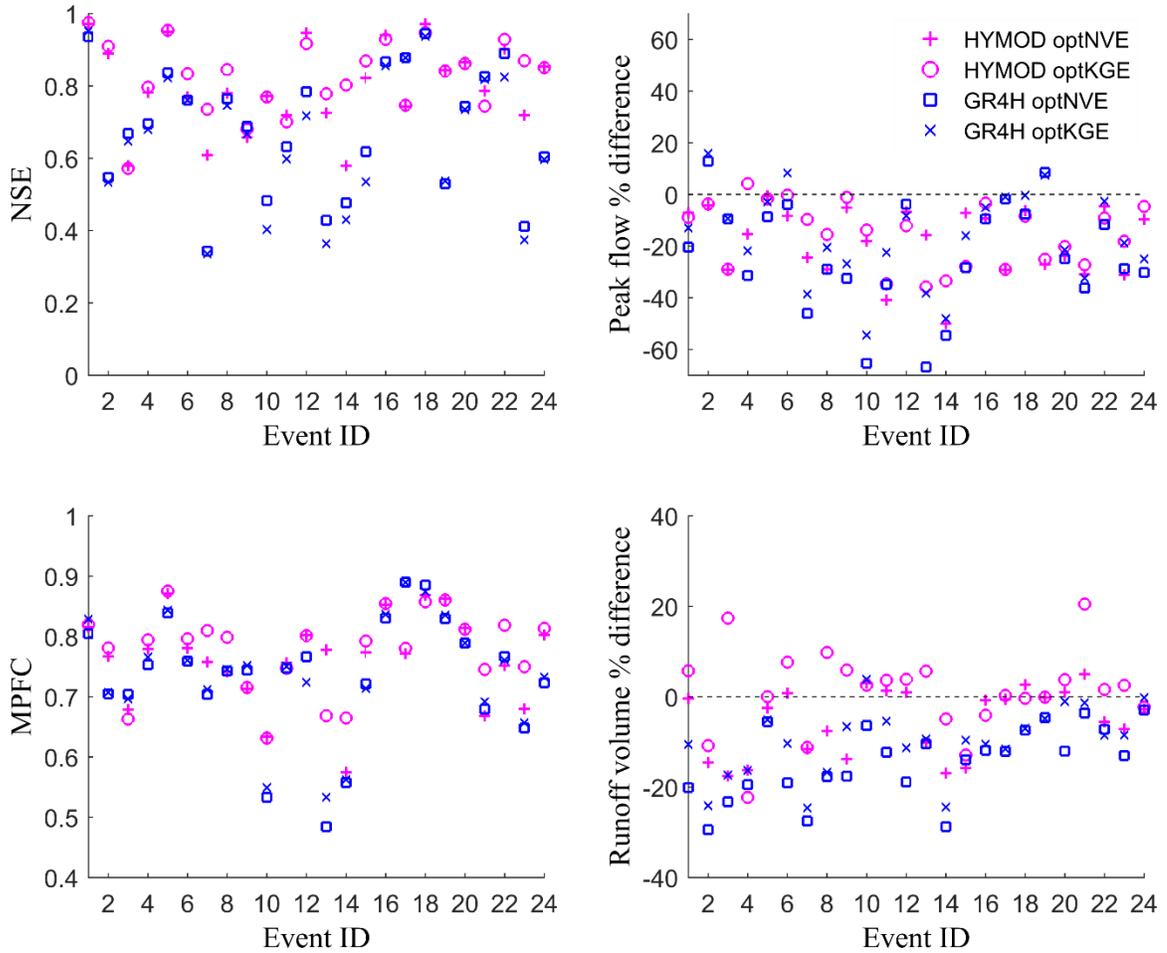


Figure 3-3 Model performance evaluation. Event 4, 7, 14 and 23 are validation events, the remaining are calibration events.

3.5.2 Application of the Precipitation-dependent HUP

3.5.2.1 Estimation of Prior Distribution

For the application of the precipitation-dependent HUP, all the data samples from the 20 calibration events were put together to estimate the HUP parameters. Recall that v is precipitation indicator and h_n is actual river discharge for every lead time n . In this study, lead times up to 6 hours were tested as it is the duration of precipitation forecast assumed

to know at each forecast occasions, and it is also assumed that beyond 6 hours the precipitation forecast is null. From the observed discharge data, joint samples $\{(v; h_0, h_1, h_2, \dots, h_N)\}$ were extracted and were classified according to $v = 0$ or $v = 1$. Then for each v and each n , the corresponding subsample $\{h_n\}$ was used to estimate the marginal prior distribution Γ_{nv} of H_n .

The modified Shapiro-Wilk test (MSW) developed by Ashkar and Aucoin (2012) is a powerful method to determine the goodness of fit for non-normal distribution, a MSW value of 1 indicates perfect fit. Among all the 15 distributions tested by MSW, Generalized Extreme Value (GEV) generated the largest MSW value 0.99 for $v = 0$, thus GEV is the best distribution function for $v = 0$. While Log-normal (LN), GEV and Inverse Gaussian produced the best MSW value 0.98 for $v = 1$, and LN was selected as the most suitable distribution for $v = 1$ because it fits better for the tails. Figure 3-4 illustrates the goodness of fit for the selected distributions.

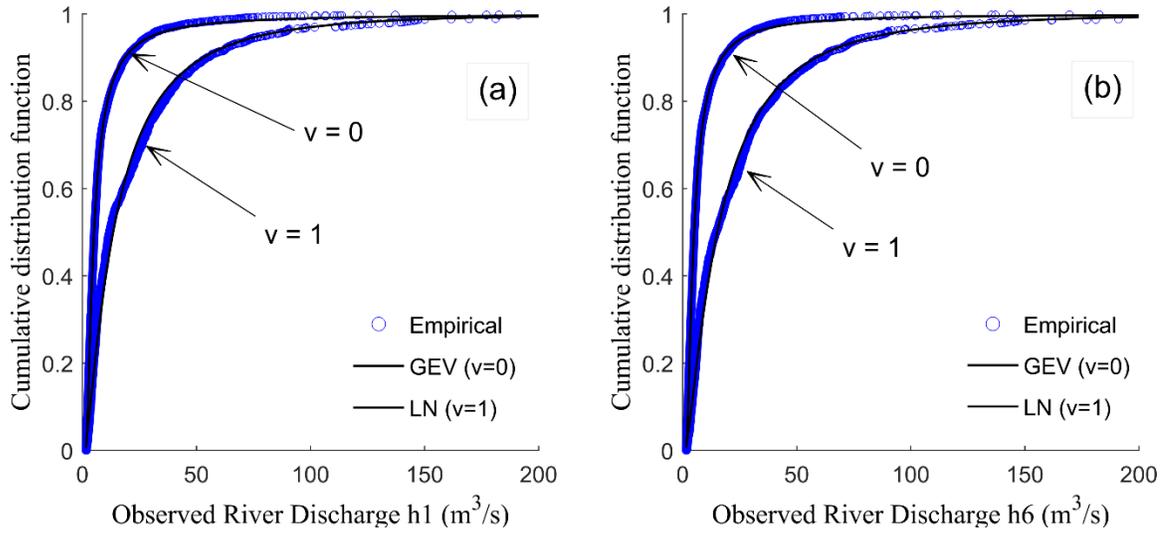


Figure 3-4 Marginal distributions Γ_{nv} of observed discharges H_n for (a) $n = 1$; and (b) $n = 6$ conditional on precipitation indicator $v = 0, 1$.

Table 3-3 describes the statistics and the parameter values for each subsample. In the branch of $v = 0$, the standard deviation of H_n demonstrates a decreasing trend with a falling mean observed discharge, while for $v = 1$, the standard deviation (S.D.) indicates an increasing trend as the mean value rises. As expected, the uncertainty measured by standard deviation is uniformly higher under precipitation occurrence ($v = 1$) than under nonoccurrence of precipitation ($v = 0$). Conditional on precipitation indicator v , the prior dependence structure can be obtained. The Pearson's correlation coefficient c_{nv} was estimated by looking at the linear relationship between joint transformed subsample $\{(w_{n-1}, w_n)\}$, under the hypothesis of a first order markov chain process, as defined by Eq. (3-5). The prior distribution parameter C_{nv} and t_{nv} shown in Table 3-5 were computed by Eqs. (3-9) and (3-10). C_{nv} characterizes the dependence structure between H_n and H_0 , the

decreasing trend in C_{nv} for both branches indicates that the informativeness of H_0 decreases as the lead time grows.

Table 3-3 Sample statistics and marginal prior distributions of observed discharge

Precipitation event	Variate	Sample		Distribution type	Parameters of Γ_{nv}			Correlation coefficient c_{nv}
		Mean	S.D.		α	β	γ	
$\nu = 0$	$h0$	9.80	17.29	GEV	0.73	2.63	4.04	--
	$h1$	9.54	16.71	GEV	0.72	2.58	4.01	0.999
	$h2$	9.30	16.24	GEV	0.71	2.53	3.99	0.998
	$h3$	9.14	15.76	GEV	0.70	2.50	3.97	0.997
	$h4$	9.04	15.50	GEV	0.69	2.49	3.96	0.995
	$h5$	8.96	15.50	GEV	0.69	2.48	3.96	0.995
	$h6$	8.93	15.40	GEV	0.68	2.48	3.97	0.994
$\nu = 1$	$h0$	23.83	35.69	LN	2.55	1.07	--	--
	$h1$	24.75	36.25	LN	2.60	1.07	--	0.975
	$h2$	25.57	36.63	LN	2.65	1.06	--	0.975
	$h3$	26.18	37.07	LN	2.68	1.05	--	0.979
	$h4$	26.55	37.29	LN	2.70	1.05	--	0.983
	$h5$	26.85	37.14	LN	2.71	1.05	--	0.983
	$h6$	27.00	37.22	LN	2.72	1.05	--	0.987

Sample size: 5473 for $\nu = 0$, 1607 for $\nu = 1$

Figure 3-5 shows the dependence structure conditional on $\nu = 1$ under certain lead time n , as we are more interested in high precipitation conditions in flood forecasting. It presents the scatterplot of (w_n, w_{n-1}) and (h_n, h_{n-1}) along with the conditional median of W_n and H_n ($p = 0.5$), and the conditional quantiles for $p = 0.1$ and $p = 0.9$, obtained from Eqs. (3-24) and (3-25), defining the 80% central credible interval around the median. ρ_{nv} assesses the

relationship between h_n and h_{n-1} in the original space is Spearman's rank correlation coefficient. The upper three scatterplots indicate a linear and homoscedastic dependence structure in the transformed space between W_n and W_{n-1} , while in the lower three scatterplots, which are derived in the original space, the relationship between H_n and H_{n-1} is less evident.

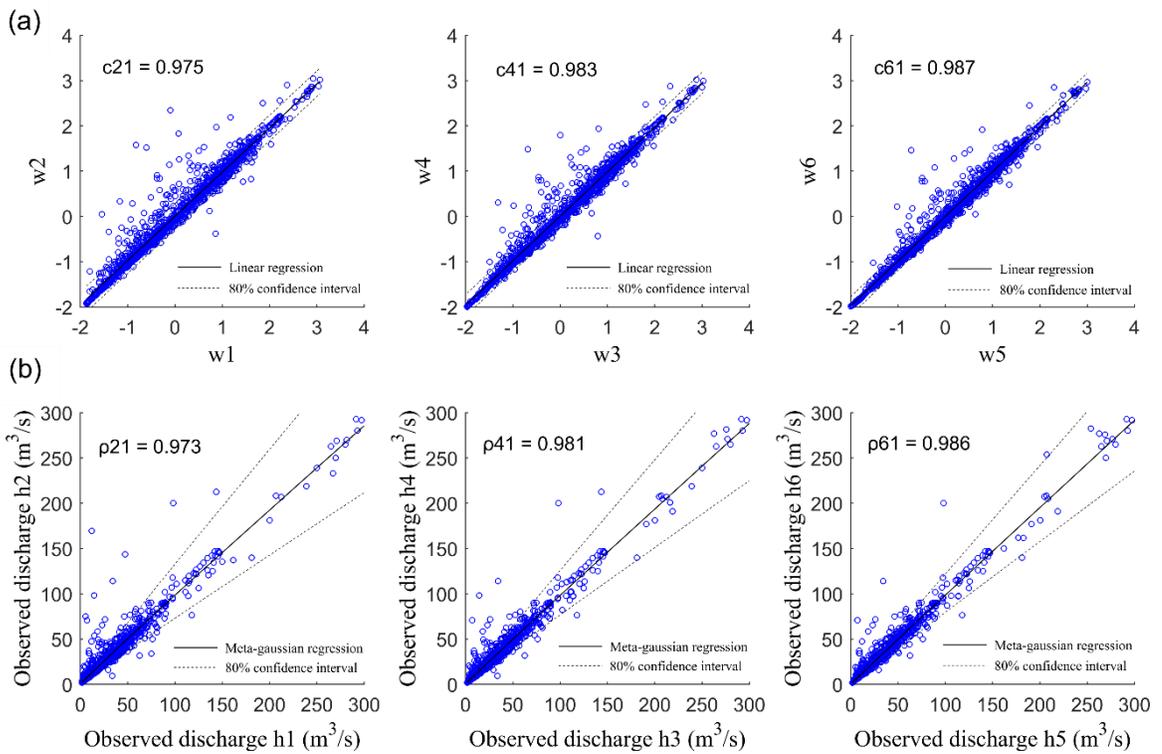


Figure 3-5 Dependence structure of the prior densities under $v = 1$ in the (a) transformed space; and (b) original space for lead time equal to 1, 3, 5 hours. h_n = actual river discharge in original space for every lead time n ; and w_n = actual river discharge in transformed normal space for every lead time n .

$$w_n(p|w_{n-1}, v) = c_{nv}w_{n-1} + (1 - c_{nv}^2)^{\frac{1}{2}}Q^{-1}(p) \quad (3-24)$$

$$h_n(p|h_{n-1}, v) = \Gamma_{nv}^{-1}(Q(c_{nv}Q^{-1}(\Gamma_{n-1,v}(h_{n-1}))) + (1 - c_{nv}^2)^{\frac{1}{2}}Q^{-1}(p) \quad (3-25)$$

3.5.2.2 Estimation of Likelihood Function

From the simulated discharge produced by HYMOD, $\{(v; s_1, s_2, \dots, s_N)\}$ were extracted and matched with $\{(v; h_0, h_1, h_2, \dots, h_N)\}$, providing joint samples $\{(v; s_1, s_2, \dots, s_N; h_0, h_1, h_2, \dots, h_N)\}$ conditional on precipitation indicator. The sample statistics are expressed in Table 3-4. Similar to the estimation of prior distribution, for each v and each n , the corresponding subsample $\{s_n\}$ was used to construct the marginal distribution \bar{L}_{nv} of S_n . Again, for $v = 0$ GEV is the best one according to MSW test, and LN, GEV and Inverse Gaussian are the most suitable one for $v = 1$. GEV gave the largest MSW value 0.99 for $v = 0$, while for $v = 1$, LN, GEV and Inverse Gaussian gave the best MSW value 0.98: LN was selected in coherence with the marginal distribution of observed discharges. The parameters $\bar{\alpha}$, $\bar{\beta}$ and $\bar{\gamma}$ of the distributions for \bar{L}_{nv} are reported in Table 3-4. Figure 3-6 exemplifies the goodness of fit for the selected distributions.

Each subsample $\{(v; s_1, s_2, \dots, s_N)\}$ was processed through NQT by Eq. (3-4) to obtain $\{(v; x_1, x_2, \dots, x_N)\}$ in the normal space. The joint sample $\{(v; x_1, x_2, \dots, x_N; w_0, w_1, w_2, \dots, w_N)\}$ was used to estimate the likelihood function parameters a_{nv} , b_{nv} , d_{nv} and σ_{nv} . These parameters were defined by the linear regression between x_n , w_n and w_0 based on Eqs. (3-7) and (3-8) (Table 3-5). For every v and every n , $a_{nv} \neq 1$ and $\sigma_{nv} \neq 0$, implying that hydrologic uncertainty exists, otherwise S_n should equal to H_n under the perfect forecast. In all cases $d_{nv} \neq 0$, indicating S_n is stochastically dependent on H_0 to some degree, the negligibly small b_{nv} value conveys that S_n is dominated by H_n and H_0 .

Table 3-4 Sample statistics and prior distributions of simulated discharge

Precipitation event	Variate	Sample		Distribution type	Parameters of $\bar{\Lambda}_{nv}$		
		Mean	S.D.		$\bar{\alpha}$	$\bar{\beta}$	$\bar{\gamma}$
$\nu = 0$	$s1$	9.80	17.29	GEV	0.73	2.63	4.05
	$s2$	9.58	16.83	GEV	0.73	2.58	3.96
	$s3$	9.39	16.41	GEV	0.73	2.54	3.89
	$s4$	9.23	16.06	GEV	0.72	2.51	3.82
	$s5$	9.09	15.82	GEV	0.72	2.49	3.75
	$s6$	8.97	15.62	GEV	0.71	2.47	3.69
$\nu = 1$	$s1$	23.94	35.70	LN	2.56	1.07	--
	$s2$	23.63	34.71	LN	2.55	1.07	--
	$s3$	23.38	33.83	LN	2.55	1.07	--
	$s4$	23.19	33.06	LN	2.54	1.07	--
	$s5$	22.99	32.33	LN	2.54	1.07	--
	$s6$	22.86	31.70	LN	2.54	1.07	--

Sample size: 5473 for $\nu = 0$, 1607 for $\nu = 1$

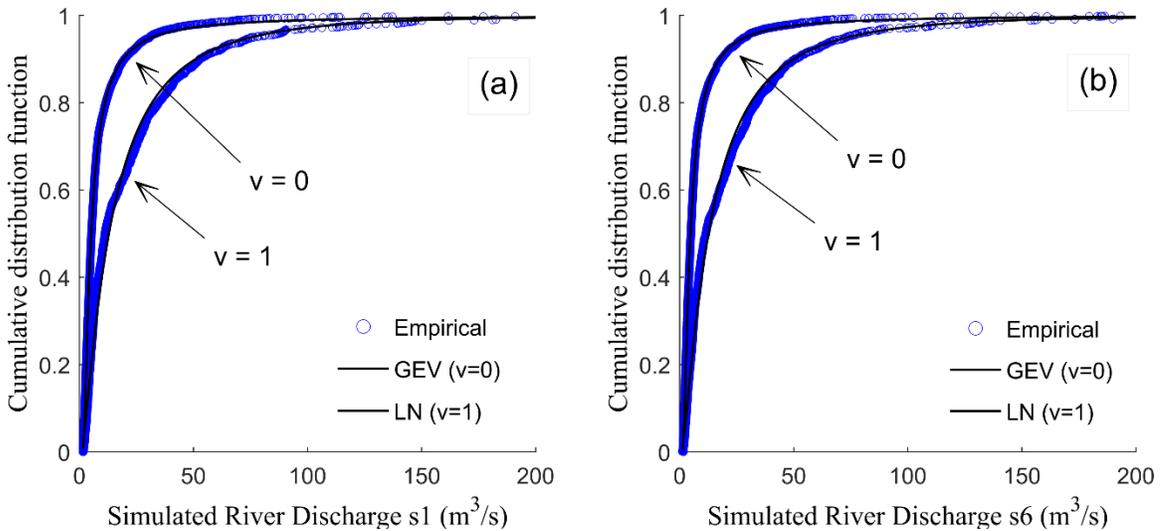


Figure 3-6 Marginal distributions Λ_{nv} of simulated discharges S_n for (a) $n = 1$; and (b) $n = 6$ conditional on precipitation indicator $\nu = 0, 1$

The likelihood dependence structure between S_n and H_n conditional on $\nu = 1$ is revealed in Figure 3-7, it shows how the linear regression of X_n on W_n (the upper three panels) is mapped into the original space of S_n versus H_n (the lower three panels). The 80% central credible interval around the median is also indicated in the normal and the real space according to Eqs.(3-26) and (3-27), respectively. The graphical evidence supports that the nonlinear and heteroscedastic dependence structure between S_n and H_n is better captured by the likelihood function in the normal space.

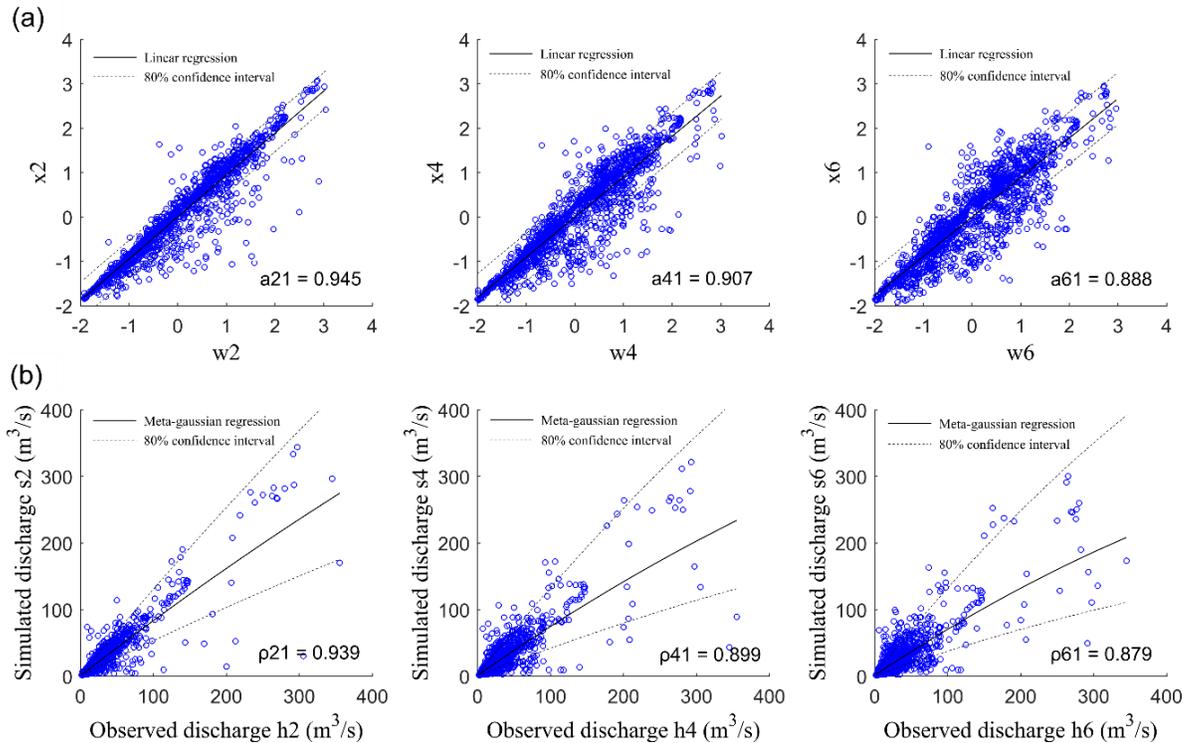


Figure 3-7 Dependence structure of the likelihood function under $\nu = 1$ in the (a) transformed space; and (b) original space for lead time equal to 2, 4, 6 hours. s_n = model river discharge in original space for every lead time n ; and x_n = model river discharge in the transformed normal space for every lead time n .

$$x_n(p|w_n, w_0, v) = a_{nv}w_n + d_{nv}w_0 + b_{nv} + \sigma_{nv}Q^{-1}(p) \quad (3-26)$$

$$\begin{aligned} s_n(p|h_n, h_0, v) \\ = \bar{A}_{nv}^{-1}(Q(a_{nv}Q^{-1}(\Gamma_{nv}(h_n)) + d_{nv}Q^{-1}(\Gamma_{n0}(h_0)) + b_{nv} \\ + \sigma_{nv}Q^{-1}(p))) \end{aligned} \quad (3-27)$$

3.5.2.3 Estimation of Posterior Distribution

According to the Bayes' equation, the posterior knowledge is equal to the prior knowledge updated using likelihood function. The parameters of posterior distribution A_{nv} , B_{nv} , D_{nv} and T_{nv} , obtained for HYMOD and GR4H are shown in Table 3-5 and Table 3-6, respectively. To summarize, the precipitation dependent HUP processor is specified by estimating 185 (30N+5) parameters:

- (i) Parameters of the marginal prior distribution Γ_{nv} of actual river discharge H_n , $\{(\alpha_{nv}, \beta_{nv}, \gamma_{nv}): v = 0, n = 0, 1, \dots, N; (\alpha_{nv}, \beta_{nv}): v = 1, n = 0, 1, \dots, N\}$
- (ii) Parameters of the marginal distribution \bar{A}_{nv} of model river discharge S_n , $\{(\bar{\alpha}_{nv}, \bar{\beta}_{nv}, \bar{\gamma}_{nv}): v = 0, n = 1, \dots, N; (\bar{\alpha}_{nv}, \bar{\beta}_{nv}): v = 1, n = 1, \dots, N\}$
- (iii) Parameters of prior distributions, $\{(C_{nv}, t_{nv}): v = 0, 1; n = 1, \dots, N\}$
- (iv) Dependence parameters of likelihood function, $\{(a_{nv}, b_{nv}, d_{nv}, \sigma_{nv}): v = 0, 1; n = 1, \dots, N\}$
- (v) Dependence parameters of posterior distribution, $\{(A_{nv}, B_{nv}, D_{nv}, T_{nv}): v = 0, 1; n = 1, \dots, N\}$

These parameters were used during validation period to quantify hydrologic uncertainty and assess the performance of the processor.

Table 3-5 Calibrated dependence parameters of posterior distribution for HUP-HYMOD

Precipitation event	Lead time n	Likelihood function				Prior distribution		Posterior distribution			
		a_{nv}	b_{nv}	d_{nv}	σ_{nv}^2	C_{nv}	t_{nv}^2	A_{nv}	B_{nv}	D_{nv}	T_{nv}
$\nu = 0$	1	0.04	0.00	0.96	0.00	1.00	0.00	0.89	0.00	0.11	0.04
	2	0.16	0.00	0.84	0.00	1.00	0.01	0.31	0.00	0.69	0.07
	3	0.29	0.00	0.71	0.01	0.99	0.01	0.35	0.00	0.64	0.10
	4	0.36	0.00	0.64	0.02	0.99	0.02	0.38	0.00	0.62	0.13
	5	0.39	0.00	0.60	0.03	0.98	0.03	0.37	0.00	0.62	0.16
	6	0.42	0.00	0.57	0.04	0.98	0.04	0.37	0.00	0.62	0.19
$\nu = 1$	1	0.07	0.00	0.93	0.00	0.98	0.05	1.20	0.00	-0.22	0.21
	2	0.15	0.00	0.85	0.01	0.95	0.09	1.12	0.00	-0.16	0.28
	3	0.22	0.00	0.79	0.02	0.93	0.13	1.17	0.00	-0.23	0.31
	4	0.27	0.00	0.75	0.03	0.92	0.16	1.08	0.00	-0.16	0.34
	5	0.31	0.00	0.71	0.04	0.90	0.19	1.08	0.00	-0.17	0.35
	6	0.35	0.00	0.68	0.05	0.89	0.21	1.00	0.00	-0.09	0.37

Table 3-6 Calibrated dependence parameters of posterior distribution for HUP-GR4H

Precipitation event	Lead time n	Likelihood function				Prior distribution		Posterior distribution			
		a_{nv}	b_{nv}	d_{nv}	σ_{nv}^2	C_{nv}	t_{nv}^2	A_{nv}	B_{nv}	D_{nv}	T_{nv}
$\nu = 0$	1	1.42	0.00	-0.72	0.52	1.00	0.00	0.00	0.00	1.00	0.04
	2	0.92	0.00	-0.22	0.53	1.00	0.01	0.01	0.00	0.99	0.07
	3	0.72	0.00	-0.02	0.53	0.99	0.01	0.02	0.00	0.98	0.11
	4	0.69	0.00	-0.01	0.54	0.99	0.02	0.03	0.00	0.97	0.14
	5	0.70	0.00	-0.01	0.54	0.98	0.03	0.04	0.00	0.96	0.17
	6	0.72	0.00	-0.04	0.55	0.98	0.04	0.05	0.00	0.94	0.20
$\nu = 1$	1	0.44	0.00	0.42	0.27	0.98	0.05	0.08	0.00	0.91	0.22
	2	0.52	0.00	0.35	0.27	0.95	0.09	0.17	0.00	0.81	0.29
	3	0.59	0.00	0.29	0.26	0.93	0.13	0.26	0.00	0.71	0.33
	4	0.63	0.00	0.26	0.25	0.92	0.16	0.32	0.00	0.64	0.36
	5	0.68	0.00	0.22	0.25	0.90	0.19	0.39	0.00	0.58	0.37
	6	0.71	0.00	0.19	0.25	0.89	0.21	0.42	0.00	0.54	0.38

3.5.3 Discussion of Results and Performance Assessment

As an example, Figure 3-8 and Figure 3-9 show the HUP outputs for event 18 (selected from calibration set) and event 7 (selected from validation set). Peak flow in event 18 is 147.24 m³/s which is considered as a high peak event, while event 7 is a medium peak event with the peak flow of 56.25 m³/s. In each figure, the observed discharge, the deterministic forecast derived from the rainfall-runoff model and the uncertainty bound generated by the probabilistic forecast via HUP are compared under 4 different conditions: (i) 1 hour ahead forecast by HUP-HYMOD; (ii) 6 hour ahead forecast by HUP-HYMOD; (iii) 1 hour ahead forecast by HUP-GR4H; (iv) 6 hour ahead forecast by HUP-GR4H. From lead time $n = 1$ to $n = 6$, the discharge at time t_n is forecasted n hours ahead, at each forecast time, the model updates all the available information (e.g. initial discharge), and feed by precipitation for the next n hours (observed precipitation is assumed as the perfect precipitation forecasts). For lead time equals to 1 hour, whether high peak flow or medium peak flow event, basically all the observed discharges fall within the uncertainty bound expressed by 25% to 75% quantile of the predictive distribution. As the lead time increases, the uncertainty bound becomes wider, especially for high discharge values. When lead time equals 6 hours, as expected, a deterioration in the probabilistic forecast can be seen compared with 1 hour ahead forecast, since some observed discharges lie outside the uncertainty bound. In terms of different rainfall-runoff models combined with HUP, results are similar in HUP-HYMOD and HUP-GR4H for shorter lead time (e.g. 1 hour); for larger lead time (e.g. 6 hours), HUP-HYMOD

performs better, particularly for event 7 as the observed values are slightly better captured by the uncertainty bound.

The behavior of the predictive power with time largely depends on the ability of rainfall-runoff model to accurately reproduce the streamflow and on the forecast horizon of the precipitation forecasts. The performance of HUP is expected to be higher at times smaller than the concentration time of the catchment as there is still effect of measured precipitation and known initial condition, after the duration of the precipitation forecast plus the time of concentration, it is expected that the effect of assuming a null precipitation mainly influences the deterioration of discharge forecasts. Additionally, the rainfall-runoff transformation would be improved by including more information about the spatial distribution of precipitation and by using a more complex semi-distributed or distributed hydrologic model, which may reduce the width of the uncertainty bound. A different pattern of GR4H is revealed in Figure 3-9 as compared to observations or HYMOD. This may be because the UH used in GR4H has m ordinates which determined by the time parameter $x4$, and the water is staggered into m UH inputs, thus the effective rainfall is spread over several successive time steps. Nevertheless, the effect of the HUP significantly improved the outcomes of the GR4H model.

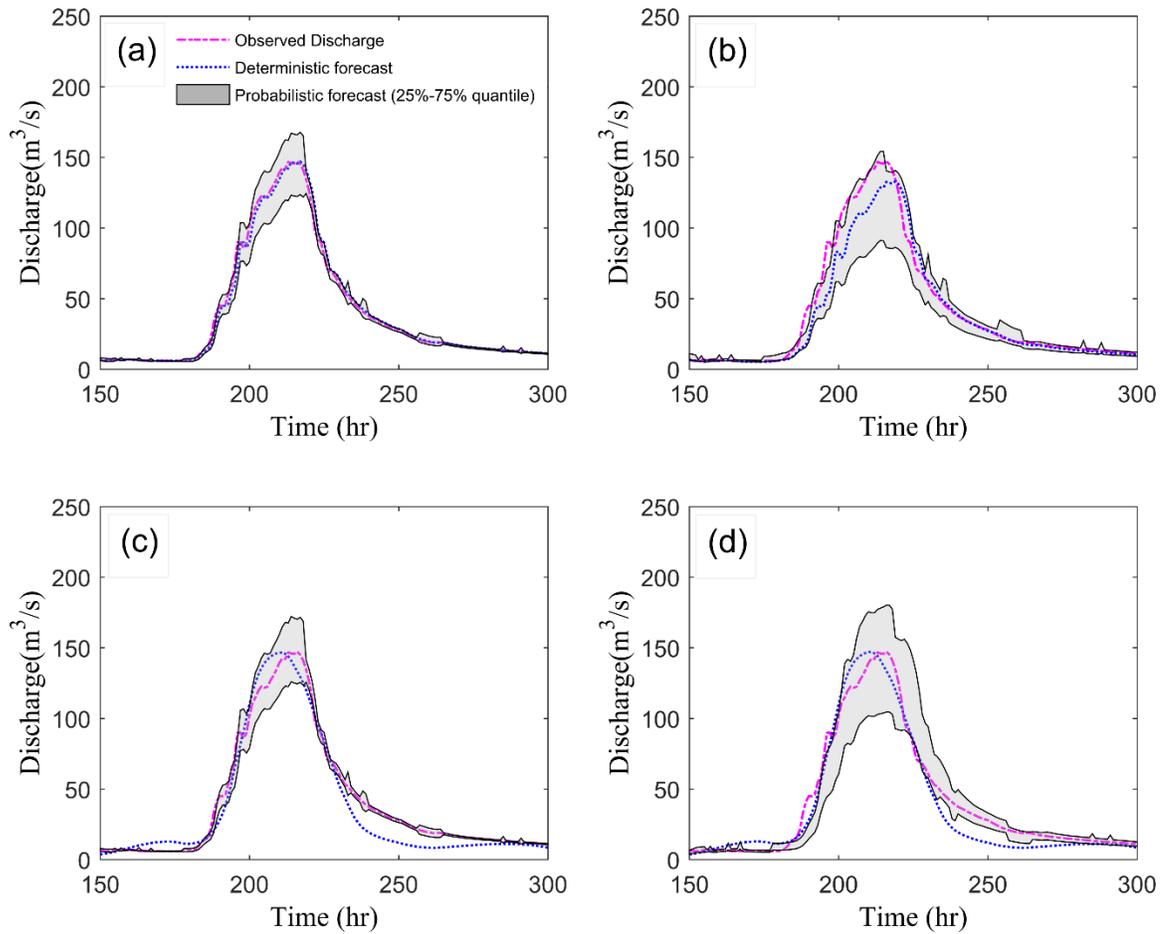


Figure 3-8 Hydrologic uncertainty bound generated by HUP for event 18: (a) 1 hour ahead forecast by HUP-HYMOD; (b) 6 hours ahead forecast by HUP-HYMOD; (c) 1 hour ahead forecast by HUP-GR4H; and (d) 6 hours ahead forecast by HUP-GR4H.

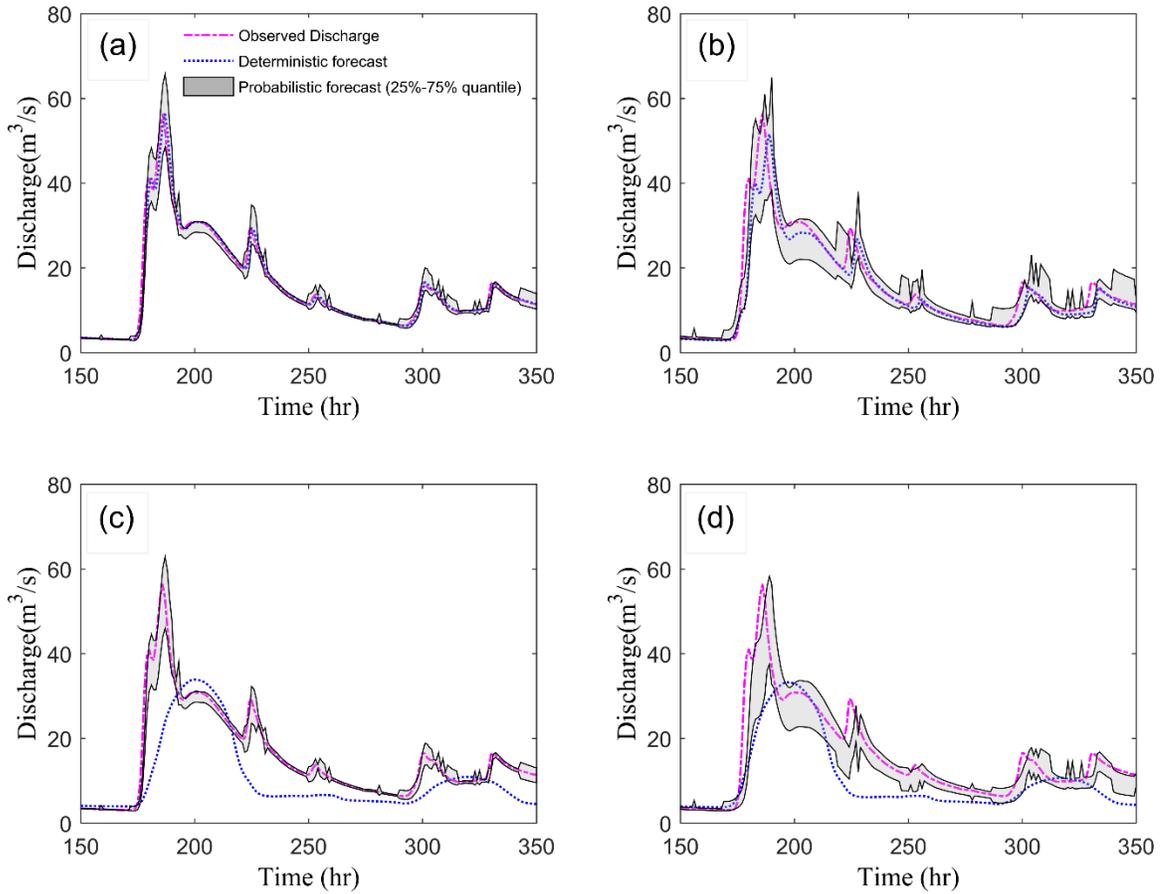


Figure 3-9 Hydrologic uncertainty bound generated by HUP for event 7: (a) 1 hour ahead forecast by HUP-HYMOD; (b) 6 hours ahead forecast by HUP-HYMOD; (c) 1 hour ahead forecast by HUP-GR4H; and (d) 6 hours ahead forecast by HUP-GR4H.

The performance in terms of accuracy of the deterministic forecast and probabilistic forecast, considering the median of the predictive distribution, were compared in terms of NSE value. As presented in Table 3-7, HYMOD and HUP-HYMOD show similar NSE values for lead time $n = 1$, but for lead time $n = 2$ up to 6, HUP-HYMOD presents larger NSE than HYMOD, indicating improved probabilistic forecasts. Accordingly, HUP-GR4H has greater NSE values than GR4H for all the lead times.

Table 3-7 Comparison of NSE between HYMOD/GR4H and HUP-HYMOD/GR4H
(median value)

Mean NSE		$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$
Calibration events	HYMOD	0.94	0.85	0.78	0.74	0.70	0.67
	HUP-HYMOD	0.94	0.86	0.81	0.77	0.74	0.70
	GR4H	0.71	0.69	0.56	0.56	0.56	0.55
	HUP-GR4H	0.94	0.85	0.79	0.73	0.69	0.66
Validation events	HYMOD	0.95	0.85	0.77	0.72	0.67	0.64
	HUP-HYMOD	0.95	0.87	0.81	0.77	0.74	0.71
	GR4H	0.50	0.50	0.50	0.50	0.49	0.49
	HUP-GR4H	0.94	0.84	0.74	0.66	0.60	0.55

Continuous ranked probability score (CRPS) was used to assess the predictive ability of HUP. CRPS is one of the most widely used metrics for probabilistic and ensemble forecasts verification (Hersbach 2000). CRPS measures how well the predictive distributions match the observed values by considering both the location and spread of the distribution. It can be expressed as in Eq. (3-28) (Hersbach 2000) where $\Phi_n(\cdot)$ is the predictive distribution, $h_{a,n}$ is the observed value and $H(\cdot)$ is the Heaviside function, its perfect score of 0 is only achieved when $\Phi_n(\cdot) = H(\cdot)$.

$$CRPS = \int_0^{+\infty} [\Phi_n(h_n) - H(h_n - h_{a,n})]^2 dh_n \quad (3-28)$$

The comparison of CRPS values between HUP-HYMOD and HUP-GR4H for different lead times is shown in Figure 3-10, and Table 3-8 illustrates the comparison of mean CRPS under different types of events. According to Figure 3-10, it is obvious that the

CRPS for all the events goes up with the increase of lead time. The CRPS values are relatively low, with the majority under 4.00, revealing good performance for the HUP processor in most of the cases. Among all the 24 events, the distributions of CRPS between HUP-HYMOD and HUP-GR4H are different for event 7, 14, 21 and 22, and very similar for other events. Event 14, 21 and 22 are flood events with high peak flow: 104.11 m³/s, 200.21 m³/s, and 355.23 m³/s, respectively. Event 7 is a multi-peak event. As indicated in Table 3-8, for low peak flow events, the mean CRPS value of HUP-HYMOD and HUP-GR4H is quite similar for every lead times, but for high peak flow events, smaller CRPS values are obtained for HUP-HYMOD compared to HUP-GR4H, indicating the better performance of the former vs. the latter. The hydrologic model calibration results indicate that HYMOD performed better than GR4H for almost all the type of the events, however, using HUP as a post-processor of GR4H presents comparable performance to HUP-HYMOD for low peak flow events. While for high peak flow events, the better deterministic forecast is generated from the hydrologic model, the better probabilistic forecast is produced by applying HUP to that hydrologic model.

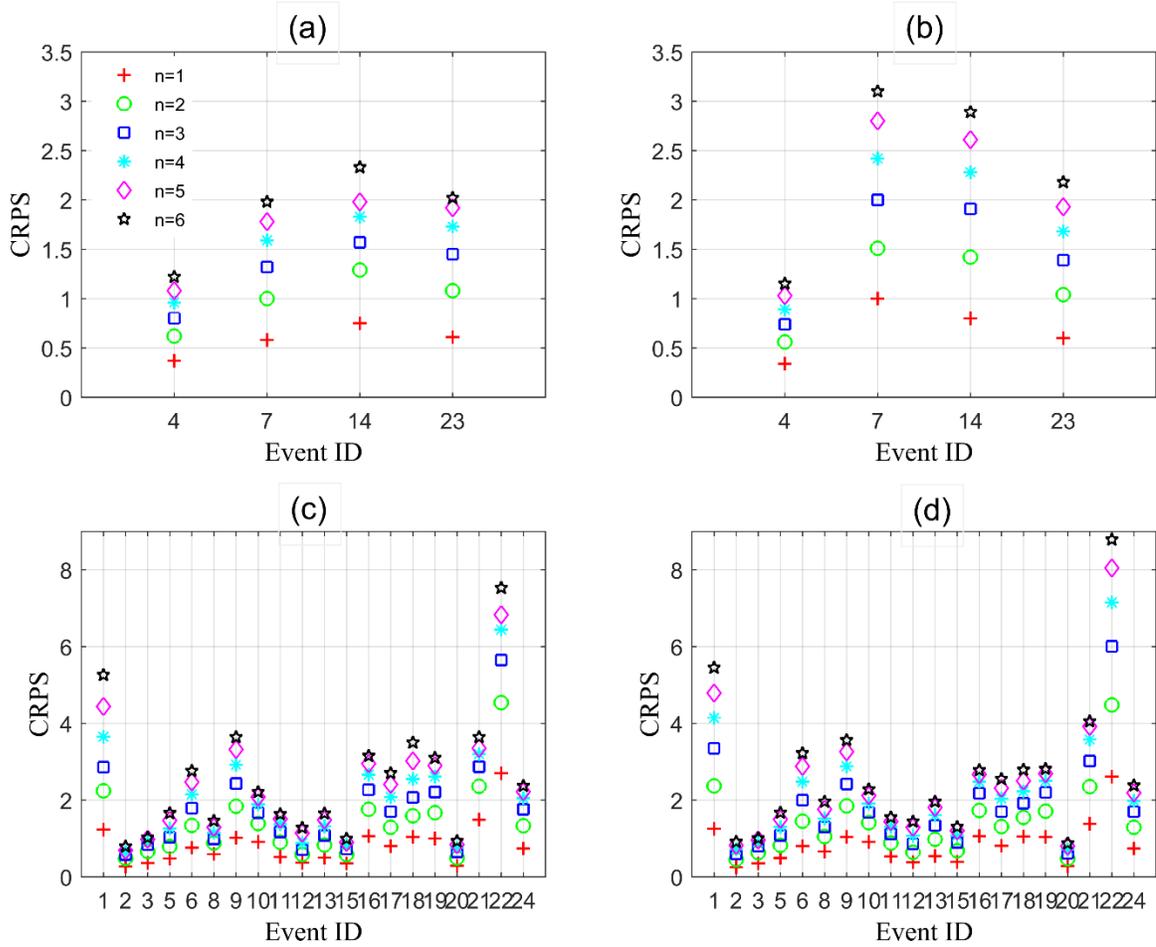


Figure 3-10 Comparison of CRPS between HUP-HYMOD and HUP-GR4H for different lead times, and CRPS value for (a) HUP-HYMOD validation events; (b) HUP-GR4H validation events; (c) HUP-HYMOD calibration events; and (d) HUP-GR4H calibration events

Based on the numerical results in Table 3-8, hydrologic uncertainty increases with increasing lead time, which leads to deterioration of the probabilistic forecast, although a decreasing trend is indicated in the deterioration with increasing lead time. For example, considering the HUP-HYMOD validation events, the mean CRPS value for lead time 1 hour is 0.58, while the mean CRPS for lead time 6 hour is 1.89 which increased by 226%,

while the relative differences of CRPS between lead time n and $n-1$ ($n = 2$ to 6) are about 72%, 28%, 18%, 10%, 11%, respectively, revealing a decreasing trend. HUP performance also decreases with high discharge value. For example, as shown in Table 3-8, the mean CRPS values by HUP-HYMOD for low, medium and high peak flow events are 1.10, 2.30 and 4.16, respectively ($n = 6$). The trends indicate that increasing lead time and high discharge decrease the structural robustness of HUP in capturing actual model predictive uncertainty.

Table 3-8 Comparison of mean CRPS under different conditions

Mean CRPS	Lead time	Calibration events	Validation events	Low peak flow events	Medium peak flow events	High peak flow events
HUP-HYMOD	$n = 1$	0.82	0.58	0.36	0.73	1.33
	$n = 2$	1.38	1.00	0.62	1.23	2.22
	$n = 3$	1.75	1.29	0.79	1.58	2.79
	$n = 4$	2.07	1.53	0.92	1.87	3.29
	$n = 5$	2.31	1.69	1.00	2.11	3.67
	$n = 6$	2.56	1.89	1.10	2.30	4.16
HUP-GR4H	$n = 1$	0.83	0.68	0.36	0.75	1.32
	$n = 2$	1.41	1.13	0.61	1.27	2.25
	$n = 3$	1.84	1.51	0.79	1.65	2.98
	$n = 4$	2.18	1.82	0.94	1.95	3.57
	$n = 5$	2.45	2.09	1.04	2.19	4.03
	$n = 6$	2.67	2.33	1.13	2.39	4.42

In addition, reliability of the forecasts was analyzed using a reliability plot suggested by Laio and Tamea (2007). The reliability plot represents the Z_i values against their empirical cumulative function R_i/n , in which Z_i is the cumulative density function

corresponding to the observed discharge x_i , and R_i/n is the associated rank in the ordered Z_i vector divided by the sample size n . Kolmogorov confidence bands, parallel to the bisector with a distance of $q(\alpha)/\sqrt{n}$ ($q(\alpha = 0.05) = 1.358$), were also included in the graph to test the uniformity. As indicated in the evaluation criterion shown in Laio and Tamea (2007), the forecast is deemed reliable under the condition that the plotted $(Z_i, R_i/n)$ pairs lie close to the bisector and distributed within the confidence bands, otherwise, problem regarding the prediction bias or the spread of the probabilistic distribution is revealed.

Figure 3-11 presents the comparison of reliability plots between HUP-HYMOD and HUP-GR4H for different lead times. The Kolmogorov 5% confidence bands, displayed as the dashed lines, are larger in validation events than in calibration events, this is because the sample size in validation set is smaller than in calibration set, causing wider acceptability limit. The results show that most of the reliability curves are distributed around the bisector and lie within significance bands, indicating relatively reliable forecasts. For both HUP-HYMOD and HUP-GR4H calibration events, the reliability curves closely follow the bisector for small lead times, while suggest a tendency to away from the bisector as the lead time increases. For HUP-HYMOD and HUP-GR4H validation events, parts of the reliability curves lie below the bisector, revealing underestimation of the predictions. However, the degree of the under-prediction using HUP-HYMOD is less than using HUP-GR4H.

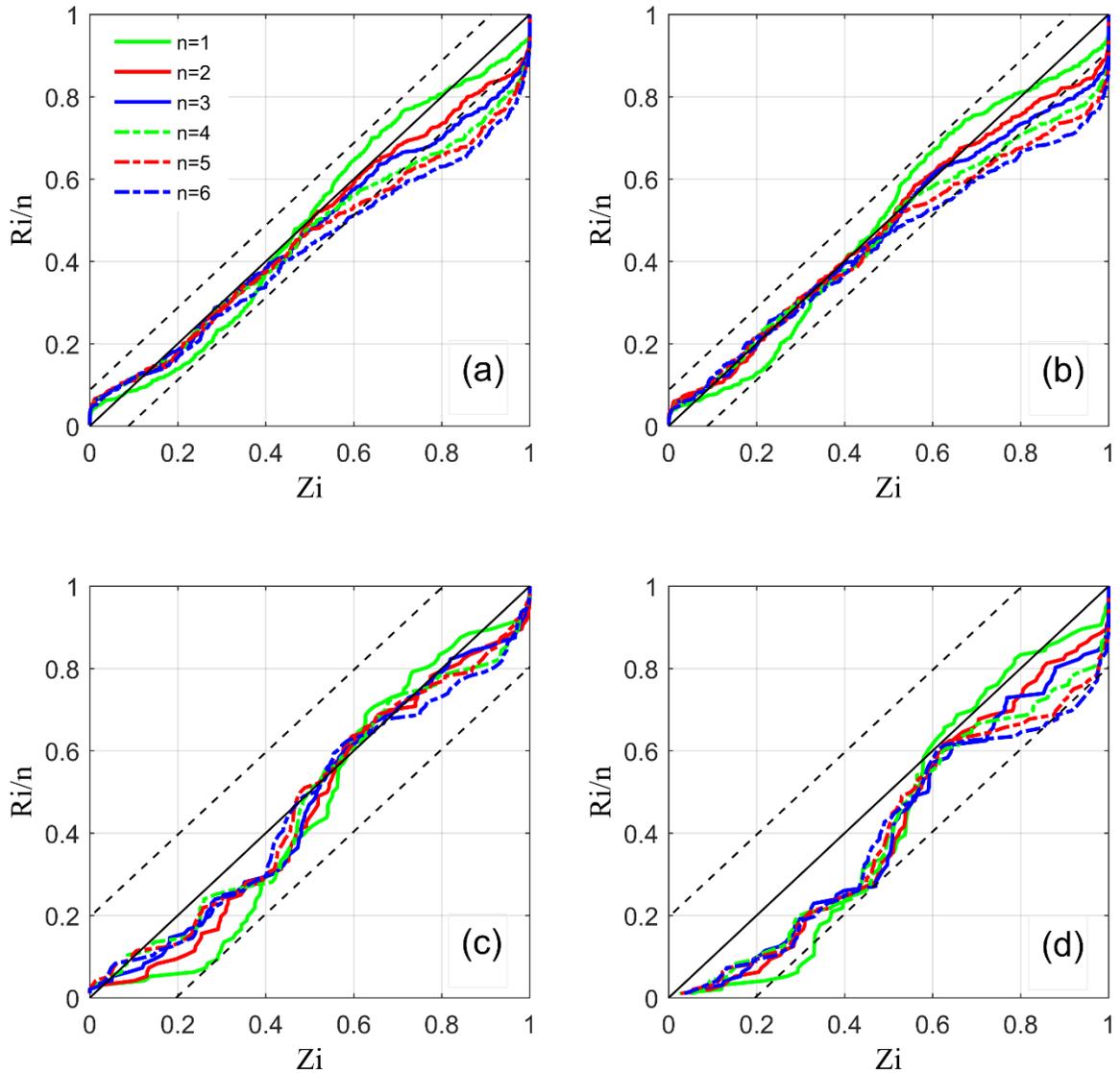


Figure 3-11 Comparison of reliability plots between HUP-HYMOD and HUP-GR4H for different lead times, and reliability plots for (a) HUP-HYMOD calibration events; (b) HUP-GR4H calibration events; (c) HUP-HYMOD validation events; and (d) HUP-GR4H validation events.

3.6 Conclusions

The paper highlights the feasibility and benefits of using precipitation-dependent HUP in a semi-urban watershed to quantify the hydrologic uncertainty associated with flood

forecast under the assumption of perfect precipitation input. Based on all available information at the time of the forecast, particularly the prediction of deterministic hydrologic model, the Bayesian processor is able to revise a prior distribution into a posterior distribution of the future actual value of the predictand. The procedure for HUP parameter estimation demonstrates the ability of the meta-Gaussian approach to capture the nonlinear and heteroscedastic dependence structures of the variables in the normal space. The predictive strength of the HUP was explored by analyzing its predictive distributions, CRPS values and reliability plots. This analysis revealed that hydrologic uncertainty has direct impact on the forecast result and cannot be overlooked. HUP is proved as a robust method for hydrologic uncertainty quantification, the actual discharges are well captured by the uncertainty bound produced from the processor, the CRPS values for most of the cases are relatively low, and the most of the reliability curves in reliability plots lie near the bisector and within the significance bands.

The performance of HUP combined with two different hydrologic models HYMOD and GR4H was tested. Results suggest that HUP has the ability to correct the deterministic forecast from HYMOD and GR4H, and produces a reliable predictive distribution which contains more valuable information. For low peak flow condition, HYMOD shows better performance than GR4H, and combining HUP with HYMOD doesn't show significant gain in performance compared with HUP-GR4H. However, under high peak flow condition, HYMOD outperforms GR4H and the combined HUP-HYMOD also significantly outperforms HUP-GR4H, indicating that through the post process of HUP, a better performing deterministic model produces better probabilistic forecast. As expected,

hydrologic uncertainty increases as the predicted discharge increases, and grows with increasing lead time. The increased hydrologic uncertainty leads to deterioration of the processor performance, but the degree of deterioration decreases with the increase of lead time.

The method can be easily used in any operational flood forecasting system to obtain probabilistic forecasts that account for hydrologic uncertainty. It has the ability to correct the deterministic forecast (e.g. from HYMOD and GR4H), and produce a more reliable predictive distribution which contains more valuable information.

3.7 Acknowledgements

This research was funded by the Natural Science and Engineering Research Council (NSERC) Canadian FloodNet (Grant number NETGP 451456), and by the China Scholarship Council (CSC). The authors would like to thank the Toronto and Region Conservation Authority (TRCA) and Water Survey of Canada for providing the study data. The authors acknowledge four anonymous reviewers for their comments that helped to improve the manuscript.

3.8 References

- Alvisi, S., and Franchini, M. (2012). “Grey neural networks for river stage forecasting with uncertainty.” *Physics and Chemistry of the Earth*, 42–44, 108–118.
- Andreassian, V., Bergstrom, S., Chahinian, N., Duan, Q., Gusev, Y. M., Littlewood, I., ... & Moussa, R. (2006). “Catalogue of the models used in MOPEX 2004 / 2005.” *IAHS publication*, 307, 41.
- Ashkar, F., and Aucoin, F. (2012). “Choice between competitive pairs of frequency models for use in hydrology: a review and some new results.” *Hydrological Sciences Journal*, 57(6), 1092–1106.
- Biondi, D., and De Luca, D. L. (2012). “A Bayesian approach for real-time flood forecasting.” *Physics and Chemistry of the Earth, Parts A/B/C*, 42–44, 91–97.
- Biondi, D., and De Luca, D. L. (2013). “Performance assessment of a Bayesian Forecasting System (BFS) for real-time flood forecasting.” *Journal of Hydrology*, 479, 51–63.
- Biondi, D., Versace, P., and Sirangelo, B. (2010). “Uncertainty assessment through a precipitation dependent hydrologic uncertainty processor: An application to a small catchment in southern Italy.” *Journal of Hydrology*, 386(1–4), 38–54.
- Boyle, D. P. (2001). “Multicriteria calibration of hydrologic models.” Ph.D. Thesis, Dep. of Hydrol. and Water Resour., Univ. of Ariz., Tucson.
- Brown, J. D., and Seo, D. J. (2010). “A Nonparametric Postprocessor for Bias Correction of Hydrometeorological and Hydrologic Ensemble Forecasts.” *Journal of hydrometeorology*, 11, 642–665.
- Coulibaly, P., Anctil, F., and Bobée, B. (2001). “Multivariate Reservoir Inflow Forecasting Using Temporal Neural Networks.” *Journal of Hydrologic Engineering*, 6(5), 367–376.
- Dechant, C. M., and Moradkhani, H. (2012). “Examining the effectiveness and

- robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting.” *Water Resources Research*, 48(4), 1–15.
- Gneiting, T., Raftery, A. E., Westveld, A. H., & Goldman, T. (2005). “Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation.” *Monthly Weather Review*, 133, 1098-1118.
- Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F. (2009). “Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling.” *Journal of Hydrology*, 377(1–2), 80–91.
- Han, S., and Coulibaly, P. (2017). “Bayesian flood forecasting methods: A review.” *Journal of Hydrology*, 551, 340-351.
- Herr, H. D., and Krzysztofowicz, R. (2010). “Bayesian ensemble forecast of river stages and ensemble size requirements.” *Journal of Hydrology*, 387(3–4), 151–164.
- Herr, H. D., and Krzysztofowicz, R. (2015). “Ensemble Bayesian forecasting system Part I: Theory and algorithms.” *Journal of Hydrology*, 524, 789–802.
- Hersbach, H. (2000). “Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems.” *Weather and Forecasting*, 15(5), 559–570.
- Kelly, K. S., and Krzysztofowicz, R. (2000). “Precipitation uncertainty processor for probabilistic river stage forecasting.” *Water Resources Research*, 36(9), 2643–2653.
- Khan, M. S., and Coulibaly, P. (2006). “Bayesian neural network for rainfall-runoff modeling.” *Water Resources Research*, 42(7), n/a-n/a.
- Krzysztofowicz, R. (1999). “Bayesian theory of probabilistic forecasting via deterministic hydrologic model.” *Water Resources Research*, 35(9), 2739–2750.
- Krzysztofowicz, R. (2001). “Integrator of uncertainties for probabilistic river stage forecasting: precipitation-dependent model.” *Journal of Hydrology*, 249(1–4), 69–85.
- Krzysztofowicz, R. (2002). “Bayesian system for probabilistic river stage forecasting.” *Journal of Hydrology*, 268(1–4), 16–40.

- Krzysztofowicz, R., and Herr, H. D. (2001). “Hydrologic uncertainty processor for probabilistic river stage forecasting: precipitation-dependent model.” *Journal of Hydrologic*, 249(1–4), 46–68.
- Krzysztofowicz, R., and Kelly, K. S. (2000). “Hydrologic uncertainty processor for probabilistic river stage forecasting.” *Water Resources Research*, 36(11), 3265–3277.
- Krzysztofowicz, R., and Maranzano, C. J. (2004). “Bayesian system for probabilistic stage transition forecasting.” *Journal of Hydrology*, 299(1–2), 15–44.
- Laio, F., and Tamea, S. (2007). “Verification tools for probabilistic forecasts of continuous hydrological variables.” *Hydrology and Earth System Sciences*, 11, 1267-1277.
- Li, W., Duan, Q., Miao, C., Ye, A., Gong, W., and Di, Z. (2017). “A review on statistical postprocessing methods for hydrometeorological ensemble forecasting. ” *Wiley Interdisciplinary Reviews: Water*, 4(6), e1246.
- Liu, Z., Guo, S., Xiong, L., and Xu, C. Y. (2018). “Hydrologic uncertainty processor based on a copula function.” *Hydrological Sciences Journal*, 63(1), 74-86.
- Liu, Z., Guo, S., Zhang, H., Liu, D., and Yang, G. (2016). “Comparative study of three updating procedures for real-time flood forecasting. ” *Water Resources Management*, 30(7), 2111-2126.
- Maranzano, C. J., and Krzysztofowicz, R. (2004). “Identification of likelihood and prior dependence structures for hydrologic uncertainty processor.” *Journal of Hydrology*, 290(1–2), 1–21.
- Madadgar, S., and Moradkhani, H. (2014). “Improved Bayesian multimodeling: Integration of copulas and Bayesian model averaging.” *Water Resources Research*, 50(12), 9586–9603.
- Mathevet, T. (2005). “Which rainfall–runoff model at the hourly time-step? Empirical development and intercomparison of rainfall–runoff models on a large sample of

- watersheds.” Ph.D. thesis (in french), ENGREF Univ., Paris, France.
- Moore, R. J. (1985). “The probability-distributed principle and runoff production at point and basin scales.” *Hydrological Sciences Journal*, 30(2), 273-297.
- Moradkhani, H., Dechant, C. M., and Sorooshian, S. (2012). “Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-Markov chain Monte Carlo method.” *Water Resources Research*, 48(12).
- Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., and Loumagne, C. (2005a). “Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 2 - Towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling.” *Journal of Hydrology*, 303(1-4), 290-306.
- Oudin, L., Michel, C., and Anctil, F. (2005b). “Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 1 - Can rainfall-runoff models effectively handle detailed potential evapotranspiration inputs?” *Journal of Hydrology*, 303(1-4), 275-289.
- Perrin, C., Michel, C., and Andréassian, V. (2003). “Improvement of a parsimonious model for streamflow simulation.” *Journal of Hydrology*, 279(1-4), 275-289.
- Quan, Z., Teng, J., Sun, W., Cheng, T., and Zhang, J. (2015). “Evaluation of the HYMOD model for rainfall-runoff simulation using the GLUE method.” *IAHS-AISH Proceedings and Reports*, 368(August 2014), 180-185.
- Raftery, A. E. (1993). “Bayesian model selection in structural equation models.” in K.A. Bollen and J.S. Long (Eds.), *Testing Structural Equation Models*, pp. 163-180. Newbury Park, Calif. Sage.
- Razavi, T., and Coulibaly, P. (2016). “An evaluation of regionalization and watershed classification schemes for continuous daily streamflow prediction in ungauged watersheds.” *Canadian Water Resources Journal / Revue canadienne des ressources hydriques*, 1-19.
- Reddy, M. J., and Ganguli, P. (2012). “Bivariate Flood Frequency Analysis of Upper

- Godavari River Flows Using Archimedean Copulas.” *Water Resources Management*, 26(14), 3995–4018.
- Reggiani, P., Renner, M., Weerts, A. H., and Van Gelder, P. A. H. J. M. (2009). “Uncertainty assessment via Bayesian revision of ensemble streamflow predictions in the operational river Rhine forecasting system.” *Water Resources Research*, 45(2), 1–14.
- Reggiani, P., and Weerts, A. H. (2008). “A Bayesian approach to decision-making under uncertainty: An application to real-time forecasting in the river Rhine.” *Journal of Hydrology*, 356(1–2), 56–69.
- Regonda, S. K., Seo, D. J., Lawrence, B., Brown, J. D., and Demargne, J. (2013). “Short-term ensemble streamflow forecasting using operationally-produced single-valued streamflow forecasts - A Hydrologic Model Output Statistics (HMOS) approach.” *Journal of Hydrology*, 497, 80–96.
- Romanowicz, R. J., Young, P. C., and Beven, K. J. (2006). “Data assimilation and adaptive forecasting of water levels in the river Severn catchment, United Kingdom.” *Water Resources Research*, 42(6).
- Salamon, P., and Feyen, L. (2009). “Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter.” *Journal of Hydrology*, 376(3–4), 428–442.
- Samuel, J., Coulibaly, P., and Metcalfe, R. A. (2011). “Estimation of continuous streamflow in Ontario ungauged basins: Comparison of regionalization methods.” *Journal of Hydrologic Engineering*, 16(5), 447–459.
- Schoelzel, C., and Friederichs, P. (2008). “Multivariate non-normally distributed random variables in climate research – introduction to the copula approach.” *Nonlin. Processes Geophys*, 15, 761–772.
- Sun, W., Ishidaira, H., Bastola, S., and Yu, J. (2015). “Estimating daily time series of streamflow using hydrological model calibrated based on satellite observations of

- river water surface width: Toward real world applications.” *Environmental Research*, 139, 36–45.
- The City of Windsor. (2012). “City of Windsor Climate Change Adaptation Plan.” <<http://www.citywindsor.ca/residents/environment/Environmental-Master-Plan/Documents/Windsor%20Climate%20Change%20Adaptation%20Plan.pdf>> (Mar. 07, 2017)
- Todini, E. (2008). “A model conditional processor to assess predictive uncertainty in flood forecasting.” *International Journal of River Basin Management*, 6(2), 123–137.
- Todini, E. (2013). “From HUP to MCP: Analogies and extended performances.” *Journal of Hydrology*, 477, 33–42.
- Toronto and Region Conservation Authority. (2008). “Humber River Watershed Plan.”
- TRCA and AMEC. (2012). “Hydrologic impact of future development on flood flows and mitigation requirements in the Humber River Watershed-Report submitted by AMEC Environment & Infrastructure”. Toronto, ON, Canada.
- Van Steenbergen, N., Ronsyn, J., and Willems, P. (2012). “A non-parametric data-based approach for probabilistic flood forecasting in support of uncertainty communication.” *Environmental Modelling and Software*, 33, 92–105.
- Vrugt, J. A., and Robinson, B. A. (2007). “Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging.” *Water Resources Research*, 43(1), 1–15.
- Wang, Q. J., and Robertson, D. E. (2011). “Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences.” *Water Resources Research*, 47(2), 1–19.
- Wang, Q. J., Robertson, D. E., and Chiew, F. H. S. (2009). “A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites.” *Water Resources Research*, 45(5), 1–18.
- Weerts, A. H., Winsemius, H. C., and Verkade, J. S. (2011). “Estimation of predictive

- hydrological uncertainty using quantile regression: Examples from the National Flood Forecasting System (England and Wales).” *Hydrology and Earth System Sciences*, 15(1), 255–265.
- Young, P. C. (2002). “Advances in real-time flood forecasting. ” *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 360(1796), 1433-1450.
- Yue, S. (2001). “A bivariate gamma distribution for use in multivariate flood frequency analysis.” *Hydrological Processes*, 15(6), 1033–1045.
- Zhang, X., Liang, F., Srinivasan, R., and Van Liew, M. (2009). “Estimating uncertainty of streamflow simulation using Bayesian neural networks.” *Water Resources Research*, 45(2), 1–16.
- Zhang, X., Liang, F., Yu, B., and Zong, Z. (2011). “Explicitly integrating parameter, input, and structure uncertainties into Bayesian Neural Networks for probabilistic hydrologic forecasting.” *Journal of Hydrology*, 409(3–4), 696–709.
- Zhao, T., Wang, Q. J., Bennett, J. C., Robertson, D. E., Shao, Q., and Zhao, J. (2015). “Quantifying predictive uncertainty of streamflow forecasts based on a Bayesian joint probability model.” *Journal of Hydrology*, 528, 329–340.

Chapter 4. Probabilistic Flood Forecasting Using Hydrologic Uncertainty Processor with Ensemble Weather Forecasts

Summary of Paper 3: Han, S. and Coulibaly, P. (2019). Probabilistic Flood Forecasting Using Hydrologic Uncertainty Processor with Ensemble Weather Forecasts. *Journal of Hydrometeorology*, 20(7), 1379-1398.

This research work extended the Hydrologic Uncertainty Processor (HUP) into an ensemble prediction framework to assess both meteorological uncertainty and hydrologic uncertainty, which constitutes the Bayesian Ensemble Uncertainty Processor (BEUP), and integrated the BEUP with bias-corrected ensemble weather forecasts to enhance the forecast performance.

Key findings of this research include:

- The performances of BEUP are promising for short-range forecasts (3h – 24h).
- HUP can improve the performance for both short-range and medium-range forecasts, and the improvement is significant for short lead times and becomes less evident as lead time grows.
- The best scenario for short-range forecast is applying bias correction to each ensemble plus applying HUP.
- Bias correcting each ensemble of weather forecasts produces better predictive performance than just bias correcting the ensemble mean.

4.1 Abstract

Recent advances in the field of flood forecasting have shown increased interests in probabilistic forecast as it provides not only the point forecast but also the assessment of associated uncertainty. Here, an investigation of hydrologic uncertainty processor (HUP) as a post-processor of ensemble forecasts to generate probabilistic flood forecasts, is presented. The main purpose is to quantify dominant uncertainties and enhance flood forecast reliability. HUP is based on Bayes' theorem and designed to capture hydrologic uncertainty. Ensemble forecasts are forced by ensemble weather forecasts from Global Ensemble Prediction System (GEPS) which are inherently uncertain, the input uncertainty propagates through the model chain and integrates with hydrologic uncertainty in HUP. The bias of GEPS was removed using multivariate bias correction, and several scenarios were developed by different combinations of GEPS with HUP. The performance of different forecast horizons for these scenarios was compared using multifaceted evaluation metrics. Results show that HUP is able to improve the performance for both short-range and medium-range forecasts; the improvement is significant for short lead times and becomes less obvious with increasing lead time. Overall, the performances for short-range forecasts when using HUP are promising, and the most satisfactory result for short-range is obtained by applying bias correction to each ensemble member plus applying HUP post processor.

4.2 Introduction

As the fifth IPCC (Intergovernmental Panel on Climate Change) climate assessment report concludes, acceleration of hydrological cycle due to climate change has led to

more frequent floods over the last few decades (IPCC 2015). Large floods during recent years, such as the 2013 Southern Ontario flood (Canada), 2017 Texas floods (United States), and 2007 United Kingdom floods, raised more demand of reliable flood forecasting system and methods (Reggiani et al. 2009). If flood can be forecasted accurately in advance, up to 35 percent of flood damage can be reduced by mitigation actions (United Nations 2004). Various flood forecasting models and techniques have been developed; however, adequate assessment of uncertainties associated with the forecast remains a challenging task.

A variety of uncertainties affect the forecast performance, including uncertainty related to model structure and parameter, uncertainty of weather forecasts, measurement error in observations, etc. No matter where the uncertainty comes from, the total predictive uncertainty has to be addressed (Reggiani and Weerts 2008a). Therefore, probabilistic forecast accompanied by uncertainty evaluation is gaining more interest to supplement the traditional deterministic forecast. Many predictive uncertainty assessment methods have been introduced and applied in flood forecasting experiments [see a recent review by (Han and Coulibaly 2017)]. This includes methods such as Model Conditional Processor (MCP) (Todini 2008), Data Assimilation (DA) (Vrugt et al. 2005), Quantile Regression (QR) (Koenker 2005), Hydrologic Model Output Statistics (HMOS) (Regonda et al. 2013), Ensembles Model Output Statistics (EMOS) (Gneiting et al. 2005), Bayesian Model Averaging (BMA) (Raftery et al. 2005) and so on. In this study, Bayesian forecasting system (BFS) introduced by Krzysztofowicz (1999) is selected due to its several salient properties: (i) it can produce probabilistic forecast through any

deterministic hydrologic model, (ii) the system is targeted to quantify all sources of uncertainties, (iii) it is able to update the prior distribution to posterior distribution based on Bayes' theorem by assimilating all the available information at the forecast time.

BFS consists of three parts: (i) input uncertainty processor (IUP), (ii) hydrologic uncertainty processor (HUP), and (iii) integrator (INT). As the name suggests, IUP is designed to quantify input uncertainty from the basin average precipitation amount during the forecast period; HUP aims to quantify hydrologic uncertainty which is the aggregate of all other uncertainties, including measurement and estimation error of model inputs, model structural and parametric uncertainty, model initial condition uncertainty and so on; and INT combines them together. Detailed descriptions of each component are shown in a sequence of paper (Krzysztofowicz and Kelly 2000; Krzysztofowicz and Maranzano 2004; Krzysztofowicz 2002, 2001; Krzysztofowicz and Herr 2001; Kelly and Krzysztofowicz 2000). Nowadays, weather forecasts are mostly outputs obtained from running different numerical weather prediction models or applying different perturbations (Schefzik 2016). Given the recent advances and popularity of ensemble weather products, instead of using the original probabilistic quantitative precipitation forecast (PQPF) in IUP, ensemble forecasts are served as the IUP component of the BFS in this study. In the case, the ensemble weather forecasts are an auxiliary randomization of future meteorological conditions (Reggiani and Weerts 2008a). Thus the inherent input uncertainty propagates through the model chain, and integrates with hydrologic uncertainty addressed by HUP to estimate total predictive uncertainty (Reggiani et al. 2009). In this context, HUP is performed as a hydrologic post-processor of ensemble

forecasts forced by ensemble weather forecasts, and this approach is referred to as Bayesian ensemble uncertainty processor (BEUP) in Reggiani et al. (2009).

Due to the variable atmospheric condition, imperfect orography in the model, unavoidable simplifications of the physics and thermodynamic processes, uncertainty in model parameterization and limited spatial resolution, weather forecasts generated from global or regional weather prediction models inherently exhibit systematic biases relative to observations (Eden et al. 2012). Thus, weather forecasts should be bias corrected or post-processed before practical application (Maraun 2016). Here, ensemble weather forecasts produced by the Global Ensemble Prediction System (GEPS) are used, their bias is removed in different ways, resulting in different weather forecast datasets to drive the Bayesian ensemble uncertainty processor (BEUP). Since many flood forecasting centers across Canada use deterministic weather forecasts (e.g. Global Deterministic Prediction System GDPS, Regional Deterministic Prediction System RDPS) instead of ensemble weather forecasts (e.g. Global Ensemble Prediction System GEPS, Regional Ensemble Prediction System REPS) to force their hydrologic models, therefore, besides ensemble weather forecast datasets, the ensemble mean is also tested as a potential substitute for the deterministic weather forecast.

The contribution of this work is to integrate for the first time (to our best knowledge) the bias correction (meteorological post-processing) and the HUP (hydrologic post-processing) in flood forecasting, and provide a comprehensive assessment of the predictive performance of HUP using different combinations of weather forecast inputs.

The main objectives of this research include: (i) showing the applicability of BEUP for

enhanced probabilistic flood forecasts using GEPS ensemble forecasts, as an alternative to the use of deterministic weather forecasts as currently practiced by Canadian hydrologic forecast centers and in other countries; (ii) assessing the predictive performance of HUP with bias-corrected ensemble weather forecasts; (iii) investigating the forecast performance of using different weather forecast datasets, including raw GEPS, bias-corrected GEPS, and their ensemble mean.

4.3 Methodology

An overview of the methodology used for the probabilistic flood forecast with total uncertainty assessment is presented in a flowchart (see Figure 4-1). Based on historical observations, the hydrologic model is calibrated prior to the forecast time. The calibrated hydrologic model is passed to HUP to analyze the model uncertainty. In HUP, the hydrologic model imitates the forecasts using meteorological observations given available information at the forecast time, and the forecasted discharges are statistically analyzed in comparison with observed discharges for different lead times. On the basis of Bayesian theory, HUP updates the prior distribution into the posterior distribution conditional on model forecast and initial condition. The HUP parameters which characterize the uncertainty expressed in the posterior distribution are estimated beforehand. In the forecast mode, the ensemble weather forecasts are used to run the hydrologic model after bias correction, and the model outputs are then passed to the calibrated HUP. Finally the ensemble posterior distributions generated by HUP are lumped into one representative predictive distribution. Overall, the methodology includes four major parts: (i) calibration of hydrologic model, (ii) calibration of Hydrologic

Uncertainty Processor, (iii) bias correction of ensemble weather forecasts, and (iv) application of ensemble weather forecasts with Hydrologic Uncertainty Processor. Detailed explanation about the bias correction method and the Bayesian ensemble processor are shown below.

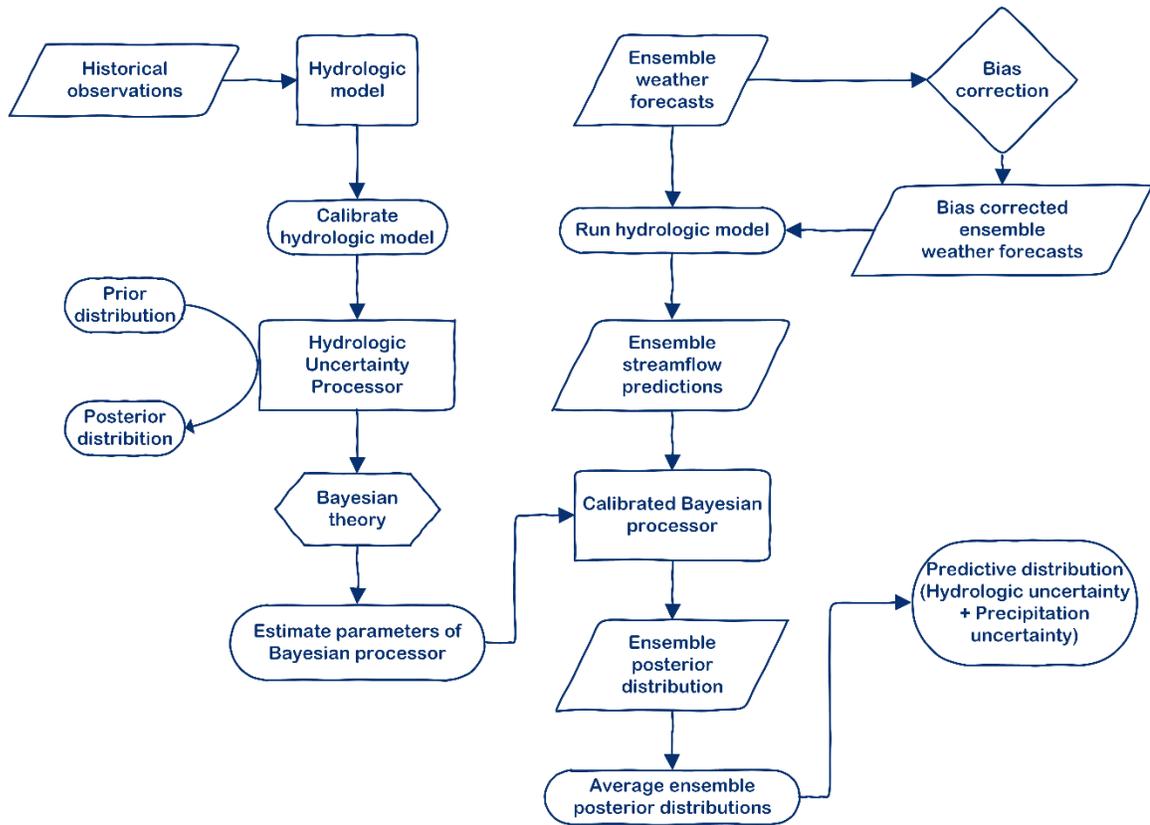


Figure 4-1 Flowchart of methodology

4.3.1 Multivariate Bias Correction Algorithm

Some of the popular bias correction methods include quantile mapping (Maurer and Hidalgo 2008), equidistant quantile mapping (Li et al. 2010) and equiratio quantile mapping (Wang and Chen 2014); however, these methods only apply to bias correcting

individual variable and do not consider the correlation between variables. As an alternative to univariate bias correction, multivariate bias correction has been developed in order to also correct the dependence structure. Three multivariate bias correction (MBC) algorithms were proposed: MBC Pearson correlation (MBC_p) (Cannon 2016), MBC rank correlation (MBC_r) (Cannon 2016), and MBC N-dimensional probability density function transform (MBC_n) (Cannon 2017), their performances were compared, and MBC_n showed best results (Cannon 2017). Therefore, MBC_n was applied in this study to bias correct the ensemble weather forecasts.

MBC_n algorithm matches multivariate distribution using quantile delta mapping (QDM) and the N-dimensional probability density function transform (N-pdf_t). Three datasets should be prepared: the matrix of observations during the calibration period X_T , the matrix of climate model output during the calibration period X_S , and the matrix of climate model output during the projected period X_P . To use MBC_n, we assume that (i) observations during calibration period are the perfect representation of the true historical weather; (ii) the statistical characteristics of the biases between the model outputs and observations during the calibration period also apply to the projected period (Cannon et al. 2015; Cannon 2017). Three steps are involved in MBC_n (Cannon 2017): (i) apply the orthogonal rotation to the source, projection, and target datasets, where superscript [j] stands for the jth iteration of the N-pdf_t, and R is the uniformly distributed random orthogonal rotation matrix;

$$\tilde{X}_S^{[j]} = X_S^{[j]} R^{[j]}$$

$$\tilde{X}_P^{[j]} = X_P^{[j]} R^{[j]} \quad (4-1)$$

$$\tilde{X}_T^{[j]} = X_T^{[j]} R^{[j]}$$

(ii) correct each variable in rotated source and projection data by QDM based on the corresponding variable in rotated target data

$$\begin{aligned} \hat{x}_S^{(n)[j]}(i) &= \tilde{x}_S^{(n)[j]}(i) + \tilde{F}_T^{(n)[j]-1} \left(\tilde{F}_S^{(n)[j]} \left(\tilde{x}_S^{(n)[j]}(i) \right) \right) \\ &\quad - \tilde{F}_P^{(n)[j]-1} \left(\tilde{F}_S^{(n)[j]} \left(\tilde{x}_S^{(n)[j]}(i) \right) \right) \end{aligned} \quad (4-2)$$

$$\begin{aligned} \hat{x}_P^{(n)[j]}(i) &= \tilde{x}_P^{(n)[j]}(i) + \tilde{F}_T^{(n)[j]-1} \left(\tilde{F}_P^{(n)[j]} \left(\tilde{x}_P^{(n)[j]}(i) \right) \right) \\ &\quad - \tilde{F}_S^{(n)[j]-1} \left(\tilde{F}_P^{(n)[j]} \left(\tilde{x}_P^{(n)[j]}(i) \right) \right) \end{aligned}$$

if the variable, such as precipitation, contains absolute zero value, the addition/subtraction operators should be replaced by multiplication/division operators as below

$$\begin{aligned} \hat{x}_S^{(n)[j]}(i) &= \tilde{x}_S^{(n)[j]}(i) \\ &\quad \times \tilde{F}_T^{(n)[j]-1} \left(\tilde{F}_S^{(n)[j]} \left(\tilde{x}_S^{(n)[j]}(i) \right) \right) / \tilde{F}_P^{(n)[j]-1} \left(\tilde{F}_S^{(n)[j]} \left(\tilde{x}_S^{(n)[j]}(i) \right) \right) \end{aligned} \quad (4-3)$$

$$\begin{aligned} \hat{x}_P^{(n)[j]}(i) &= \tilde{x}_P^{(n)[j]}(i) \\ &\quad \times \tilde{F}_T^{(n)[j]-1} \left(\tilde{F}_P^{(n)[j]} \left(\tilde{x}_P^{(n)[j]}(i) \right) \right) / \tilde{F}_S^{(n)[j]-1} \left(\tilde{F}_P^{(n)[j]} \left(\tilde{x}_P^{(n)[j]}(i) \right) \right) \end{aligned}$$

suppose the dataset is $I \times N$ matrix, n is the n th variable in the matrix and i is the i th value of this variable, \tilde{F} means the empirical cumulative distribution function (cdf) of the rotated dataset, and \tilde{F}^{-1} denotes its inverse; (iii) apply the inverse rotation to the resulting data

$$X_S^{[j+1]} = \hat{X}_S^{[j]} R^{[j]-1}$$

$$X_P^{[j+1]} = \hat{X}_P^{[j]} R^{[j]-1} \quad (4-4)$$

$$X_T^{[j+1]} = X_T^{[j]}$$

Repeat step (i) - (iii) until the multivariate distribution of corrected source data matches target data. A corresponding R package is available for download from

<https://CRAN.R-project.org/package=MBC>.

4.3.2 Principle of Bayesian Ensemble Uncertainty Processor

The Hydrologic Uncertainty Processor (HUP) is formulated as a Bayesian processor which post-process the model outputs through Bayesian revision. It revises the prior density which is based on past evidence, through the likelihood function which brings in various hydrologic uncertainty sources to the process, and yields posterior density that expresses the aggregation of these uncertainties. In this study, the total predictive uncertainty associated with flood forecast is assessed using HUP with ensemble weather forecasts following Reggiani et al. (2009) approach. This involves applying Bayesian revision for each ensemble streamflow forecast and lump the ensemble posterior distributions into a single posterior meta-distribution as a representative function.

However, instead of using linear regression to parameterize the prior density, first-order Markov chain proposed by Krzysztofowicz and Kelly (2000) is employed. The reason for this is that the first-order Markov chain was applied to a watershed of 1450 km², and the basin size herein is similar to that study area size. While the linear regression was applied to super large watershed with an area of 160,000 km², which is unlikely to behave like first-order Markov (Reggiani and Weerts 2008b). And also, instead of using one-branch HUP processor (Krzysztofowicz and Kelly 2000), a two-branch HUP processor (Krzysztofowicz and Maranzano 2004; Krzysztofowicz 2002, 2001; Krzysztofowicz and Herr 2001) that is conditional on precipitation occurrence is adopted, since the two-branch processor was found to be more efficient and informative (Krzysztofowicz and Herr 2001). The algebraic manipulations of this Bayesian ensemble processor are summarized below, more details about the formula derivation are described in (Reggiani and Weerts 2008b; Reggiani et al. 2009), and more details about the HUP can be found in (Krzysztofowicz and Kelly 2000; Krzysztofowicz and Herr 2001; Krzysztofowicz 2002).

Following the notation in Krzysztofowicz's papers, let define n ($n = 1, \dots, N$) as forecast lead time, and v as precipitation indicator, with $v = 1$ indicates precipitation occurrence, while $v = 0$ means no precipitation. Let H_n denotes discharge observation at time t_n , and the observed discharge at forecast time t_0 is H_0 . Let S_n denotes the modeled discharge resulting from historical precipitation observation, and $S_{n,j}$ denotes the modeled discharge resulting from ensemble weather forecast with ensemble member $j = 1, \dots, J$. The corresponding lowercase letters h_n , h_0 , s_n and $s_{n,j}$ stand for realizations of variates H_n , H_0 , S_n and $S_{n,j}$, respectively.

In practice, through a process called normal quantile transform (NQT), H_n and S_n are transformed into variate W_n and X_n , respectively. For every $v \in \{0, 1\}$ and every $n \in \{0, 1, \dots, N\}$, the NQT steps include: first match H_n with marginal prior distribution $\Gamma(\cdot)$ (corresponding density is γ) and match S_n with marginal initial distribution $\bar{\Lambda}(\cdot)$ (corresponding density is $\bar{\lambda}$), and then perform the standard normal inverse $Q^{-1}(\cdot)$ for both. NQT makes it easy to fit regression and estimate parameters in normal space, in the end the expression results are transformed back into the original space.

In the normal space, the dependence parameter of the transition density c_{nv} is defined by the linear regression below

$$W_n = c_{nv}W_{n-1} + \varepsilon_n \quad (4-5)$$

Where residual ε_n is stochastically independent of W_{n-1} , and normally distributed with zero mean and variance $1 - c_{nv}^2$. The dependence parameters of the likelihood function a_{nv} , b_{nv} , d_{nv} and σ_{nv} can be defined by the following linear regression

$$X_n = a_{nv}W_n + d_{nv}W_0 + b_{nv} + \theta_n \quad (4-6)$$

where residual θ_n is stochastically independent of (W_n, W_0) , and normally distributed with zero mean and variance σ_{nv}^2 (Krzysztofowicz and Kelly 2000).

Given the dependence parameters estimated by Eq. (4-5) the parameters of prior distribution C_{nv} and t_{nv} ($v = 0, 1; n = 1, \dots, N$) are defined in Table 4-2. The parametric expression of prior distribution, conditional on discharge observation at time t_0 , is as follow

$$G_{nv}(h_n|h_0) = Q\left(\frac{Q^{-1}(\Gamma_{nv}(h_n)) - C_{nv}Q^{-1}(\Gamma_{0v}(h_0))}{t_{nv}}\right) \quad (4-7)$$

The corresponding prior density is

$$g_{nv}(h_n|h_0) = \frac{\gamma_{nv}(h_n)q(Q^{-1}(G_{nv}(h_n|h_0)))}{t_{nv}q(Q^{-1}(\Gamma_{nv}(h_n)))} \quad (4-8)$$

Given the dependence parameters derived from Eq. (4-6), the parameters of posterior distribution A_{nv} , B_{nv} , D_{nv} and T_{nv} ($v = 0, 1$; $n = 1, \dots, N$) are defined in Table 4-2. The posterior distribution for each ensemble member, conditional on each ensemble forecast and discharge observation at time t_0 , is expressed as

$$\begin{aligned} & \Phi_{nv,j}(h_n|s_{n,j}, h_0) \\ &= Q\left(\frac{Q^{-1}(\Gamma_{nv}(h_n)) - A_{nv}Q^{-1}(\bar{\Lambda}_{nv}(s_{n,j})) - D_{nv}Q^{-1}(\Gamma_{0v}(h_0)) - B_{nv}}{T_{nv}}\right) \end{aligned} \quad (4-9)$$

And the corresponding posterior density is

$$\phi_{nv,j}(h_n|s_{n,j}, h_0) = \frac{\gamma_{nv}(h_n)q\left(Q^{-1}\left(\Phi_{nv,j}(h_n|s_{n,j}, h_0)\right)\right)}{T_{nv}q(Q^{-1}(\Gamma_{nv}(h_n)))} \quad (4-10)$$

Next the ensemble posterior densities are integrated into one representative predictive density by averaging Eq. (4-10) over the ensemble $s_{n,j}$

$$\bar{\phi}_{nv,j}(h_n|\bar{s}_n, h_0) = \frac{1}{J} \sum_{j=1}^J \phi_{nv,j}(h_n|s_{n,j}, h_0) \quad (4-11)$$

The corresponding cumulative probability distribution is stated as follow

$$\bar{\Phi}_{nv,j}(h_n | \bar{s}_n, h_0, EF) \quad (4-12)$$

with EF stands for ensemble forecast (Reggiani et al. 2009).

4.4 Study Area and Data

The Humber River Watershed was chosen as study area to apply the BEUP with ensemble weather forecasts used as input for uncertainty assessment in flood forecasting. The watershed is located in Southern Ontario, Canada with a total drainage area of 911 km². A detailed basin description can be found in Han et al. (2018). This region is “flood-vulnerable”, and recent extreme hydrometeorological events (e.g. 2013 Southern Ontario Flash Flood) further emphasize the requirement for enhancing flood forecasting system in this populated region of Ontario.

Two types of data were used in this study: observed precipitation, temperature, and discharge from gauging stations (from January 2011 to December 2015); gridded precipitation and temperature forecasts from Global Ensemble Prediction System (GEPS) (from June 2015 to December 2015). The hourly gauged precipitation and temperature time series were provided by Toronto and Region Conservation Authority (TRCA), the hourly discharge time series were provided by Water Survey of Canada, and the GEPS data were from Environment and Climate Change Canada (ECCC). As shown in Figure 4-2, the 15 rain gauges and 5 temperature gauges were used to calculate mean areal precipitation and temperature, and the 2 stream gauges near the outlet were added up to estimate the total outflow. There are some days during winter that none of the 15 rain gauges has data, in order to obtain continuous time series, the missing mean areal

precipitation was filled first by nearby EC (Environment Canada) stations and then using linear interpolation.

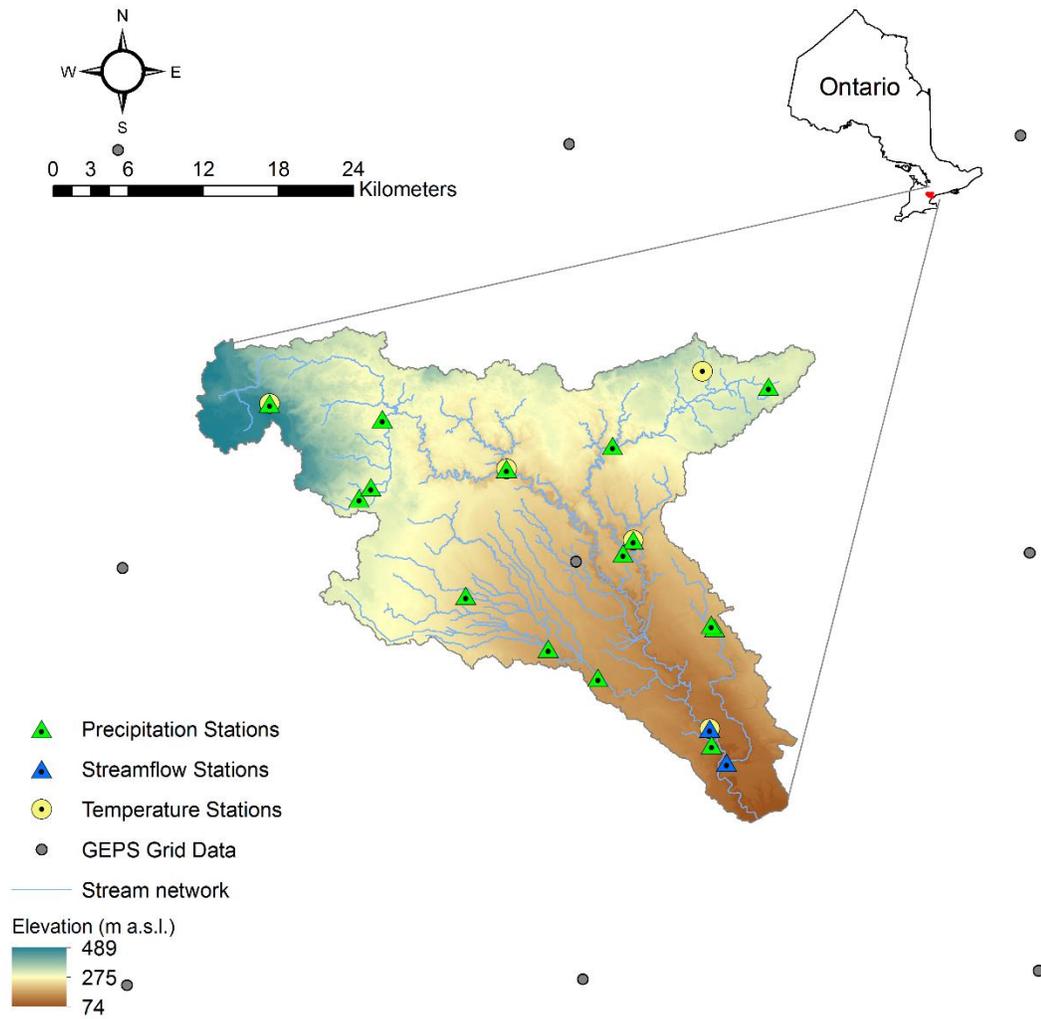


Figure 4-2 Study area: Humber River Watershed

Two modes were used to set up the system: historical mode and forecast mode. In historical mode, the system was forced by observed precipitation and temperature, with

January 2011 to December 2014 as calibration and January 2015 to May 2015 as validation. In forecast mode that starts from June 2015 to December 2015, the system was forced by ensemble weather forecasts from GEPS which consists of 1 control and 20 perturbed members, each forecast of GEPS data is initialized at 00Z daily and produces a forecast output every 3 hours for 384 hours. The system spatial resolution is 0.5 degree.

4.5 Application and Discussion

4.5.1 Hydrologic Model

A lumped conceptual rainfall-runoff model namely MAC-HBV (McMaster University Hydrologiska Byråns Vattenbalansavdelning), following the model structure of HBV (Bergström 1976), was used to simulate the hydrologic response. MAC-HBV was introduced by Samuel et al. (2011) and applied in many hydrological studies (Razavi and Coulibaly 2016, 2017), it adopts a similar concept of the HBV model from Merz and Blöschl (2004) and uses a modified routing routine following Siebert (1999). Instead of using the simplified Thornwaite formula to account for potential evapotranspiration (PE), an adjusted PE calculation approach proposed by Qudin et al. (2005a, 2005b) was adopted as it is more efficient in dealing with hourly PE.

MAC-HBV consists of a snow routine, a soil moisture routine, a response routine, and a routing routine. For the snow routine, the simple degree-day concept is replaced with SNOW-17 snow accumulation and ablation model (Anderson 2006) which is more capable of dealing with snow (Houle et al. 2017; He et al. 2011a,b), it requires only temperature and precipitation as inputs, while model output includes snow water

equivalent (SWE), and rain plus snowmelt which is passed to soil moisture routine after adjustment by rainfall correction factor PXADJ. The soil moisture routine represents the changes in soil moisture storage and the contribution to runoff entering into response routine. The soil moisture storage is controlled by rainfall, snowmelt and actual evapotranspiration. The runoff amount depends on the soil box water content, and its maximum value f_c , and a non-linear runoff generation controlling parameter β . The response routine comprises two reservoirs: an upper soil reservoir and a lower soil reservoir, it represents the water storage in upper zone and lower zone and estimates the total outflow of these two reservoirs. Recharge from soil moisture routine flow into the upper soil reservoir, and part of the water permeate into the lower soil reservoir based on the percolation rate parameter c_{perc} , thus the total outflow includes three parts: (i) outflow from the upper zone which is controlled by a flow recession coefficient k_0 if the water storage exceeds the threshold value l_{suz} , (ii) outflow from upper zone which is determined by flow recession coefficient k_1 if l_{suz} is not exceeded, and (iii) a slow outflow from lower zone affected by flow recession coefficient k_2 . In the routing routine, a triangular weighting function determined by parameter $maxbas$ is used to estimate the final runoff. All the parameter descriptions are presented in Table 4-1, and a more detailed description of each routine and corresponding equations can be found in Samuel et al. (2011).

Table 4-1 Optimized parameters of MAC-HBV and SNOW-17

Model Parameters	Descriptions	Ranges	Units	Optimized Parameters
MAC-HBV				
fc	Maximum soil box water content	60 – 600	mm	374.60
lp/fc	Limit for potential evaporation/Maximum soil box water content	0.1 – 0.9	mm/mm	0.90
beta	A non-linear parameter controlling runoff generation	0.1 – 10	–	0.76
k0	Flow recession coefficient in an upper soil reservoir (for soil moisture exceeds a threshold lsuz value)	1 – 30	days	11.06
lsuz	A threshold value used to control response routing on an upper soil reservoir	1 – 100	mm	5.77
k1	Flow recession coefficient in an upper soil reservoir	15 – 100	days	31.45
cperc	A constant percolation rate parameter	0.01 – 3	mm/days	0.08
k2	Flow recession coefficient in a lower soil reservoir	100 – 500	days	261.06
maxbas	A triangular weighting function for modeling a channel routing routine	1 – 3	days	2.00
alpha1	An exponent in relation between outflow and storage representing non-linearity of storage-discharge relationship of lower reservoir	0.5 – 1.25	–	1.11
PXADJ	Rainfall correction factor	0.1 – 1	–	0.52
SNOW-17				
scf	Snow fall correction factor	0.4 – 1.6	–	0.76
uadj	The average wind function during rain-on-snow periods	0.01 – 0.22	mm/mb/°C	0.09
mbase	Base temperature for nonrain melt factor	0 – 1	°C	0.43

mfmax	Maximum melt factor	0.2 – 2	mm/6h/°C	1.62
mfmin	Minimum melt factor	0.02 – 0.7	mm/6h/°C	0.04
tipm	Antecedent snow temperature index parameter	0.01 – 0.99	–	0.71
nmf	Maximum negative melt factor	0.05 – 0.5	mm/6h/°C	0.18
plwhc	Percent of liquid water capacity	0 – 0.4	–	0.00
pxtemp1	Lower Limit Temperature dividing transition from snow	-2 – 2	°C	-1.90
pxtemp2	Upper Limit Temperature dividing rain from transition	1 – 3	°C	1.19

In order to match the GEPS format in forecast mode, MAC-HBV model was calibrated in 3 hourly time step using particle swarm optimization algorithm (PSO) approach (Razavi and Coulibaly 2017) with modified NVE (combined Nash Sutcliffe efficiency and volume error) as objective function (Samuel et al. 2011; Razavi and Coulibaly 2017). As shown in Eq. (4-13), MNVE (modified NVE) gives more weight to NSE_{sqr} , which is better at reflecting the performance for capturing high flows, since we are more interested in high flow rather than low flow in the context of flood forecasting.

$$MNVE = 0.5NSE - 0.1VE + 0.5NSE_{sqr} \quad (4-13)$$

Calibration and validation results are shown in Figure 4-3. For calibration period, MPFC [modified peak flow criteria, $MPFC=1-PFC$, equation for PFC can be found in (Coulibaly et al. 2001)], which provides more accurate performance evaluation for flood period with 1 indicating a perfect forecast, could reach 0.81, NSE (Nash Sutcliffe efficiency) can reach 0.62 and VE (volume error) is 0.23. For validation period, MPFC is near 0.80, NSE

value is 0.60 and VE is 0.02. The MPFC values for both calibration and validation are fairly good, indicating that the model performs well in simulating peak flows. Although the NSE values are in the acceptable model range, they may appear relatively low as the model was optimized with emphasis on high flows rather than both low and high flows. The emphasis here been on enhancing flood forecasting, the model performance on the higher flows is satisfactory. Given that the model is in 3 hourly time step, the NSE values are well acceptable, and after post-processing of HUP, the performance can be further improved. The optimized parameters of MAC-HBV and SNOW-17 used in the forecast mode are presented in Table 4-1.

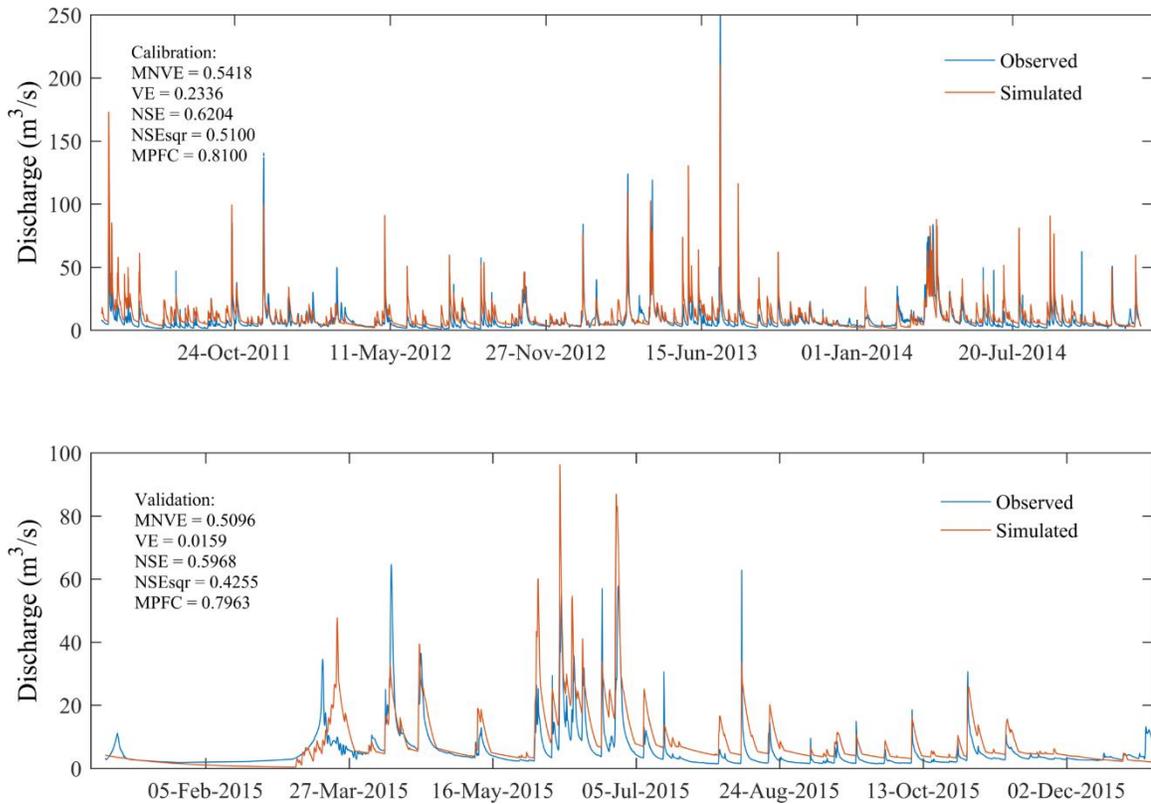


Figure 4-3 Calibration and validation for MAC-HBV

4.5.2 Calibration of Bayesian Processor

To specify the Bayesian processor, parametric expressions for the family of prior density, likelihood function and posterior density should be obtained (Reggiani and Weerts 2008b). The same time period as the hydrologic model calibration was used to estimate these parameters, and lead time up to 72 hours was considered. For every lead time n and each precipitation indicator v , the corresponding discharge observation sub-sample h_n was extracted to estimate marginal prior distribution. According to the modified Shapiro-Wilk test (MSW) proposed by Ashkar and Aucoin (2012), which is a useful approach to determine the goodness of fit for non-normal distribution, kernel was tested to be the most suitable distribution function. The goodness of fit for kernel distribution are presented in Figure 4-4 for selected lead times. In the normal space after NQT process, following Eq. (4-5) and coefficient definition for C_{nv} and t_{nv} in Table 4-2, parameters for prior distribution were calculated and shown in Table 4-2. As lead time grows, C_{nv} shows a decreasing trend and t_{nv} shows an increasing trend for both branches, indicating dependence structure between H_n and H_0 is weakened with the increase of lead time.

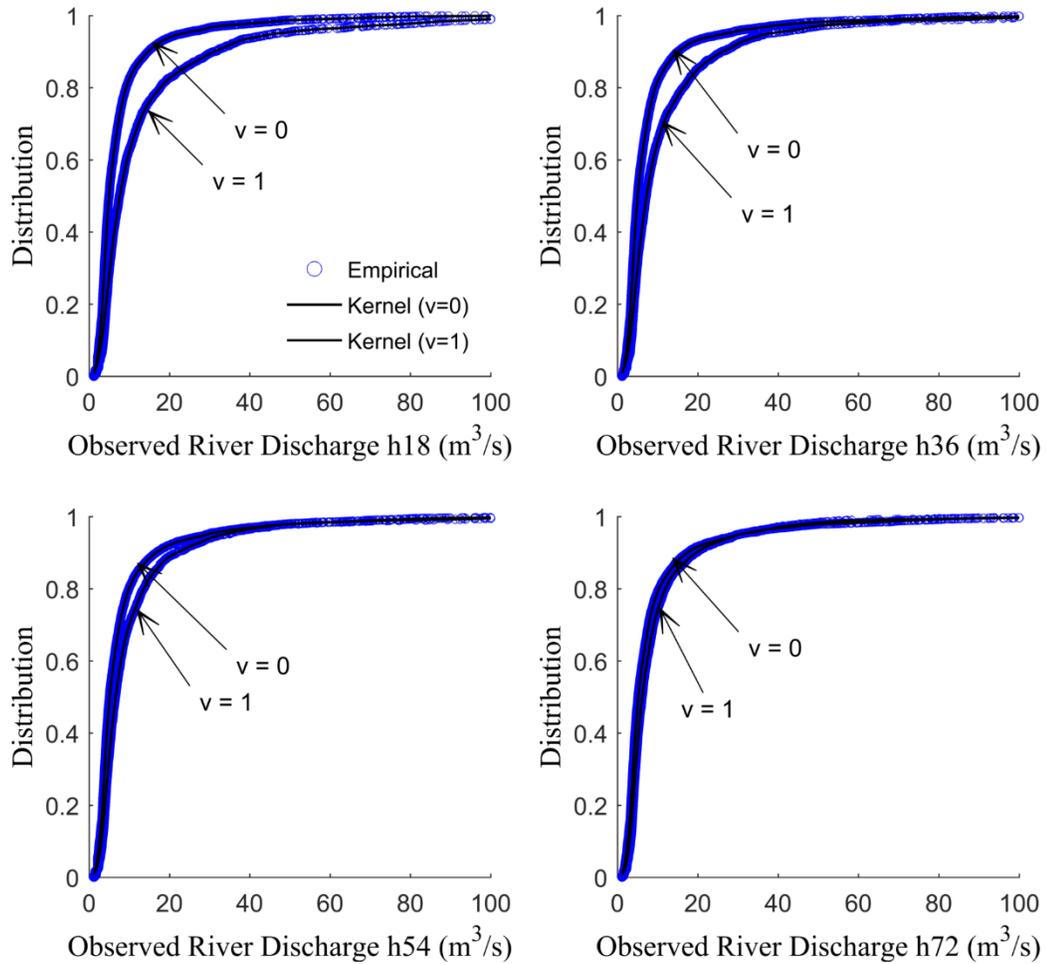


Figure 4-4 Marginal distribution of observed discharge for h18, h36, h54 and h72 conditional on precipitation indicator

Similarly, for every n and each v , the corresponding simulated discharge sub-sample s_n that matches h_n was derived to estimate marginal distribution. Again, kernel was determined as the best function based on MSW test, the goodness of fit of kernel distribution for selected lead times are displayed in Figure 4-5. Then in the transformed space, following Eq. (4-6) and coefficient definition for A_{nv} , B_{nv} , D_{nv} , and T_{nv} in Table 4-2, parameters for posterior distribution were computed and presented in Table 4-2. It is

noted that B_{nv} values are very small for all the cases and thus approximated to zero. As lead time increases, A_{nv} increases and D_{nv} decreases for both branches, suggesting the forecast is less affected by H_0 and more influenced by S_n with increasing lead time. These HUP parameters characterize the prior distribution and posterior distribution based on Eqs. (4-7) to (4-12) they are calibrated offline beforehand and will be used in forecast mode for probabilistic forecasting.

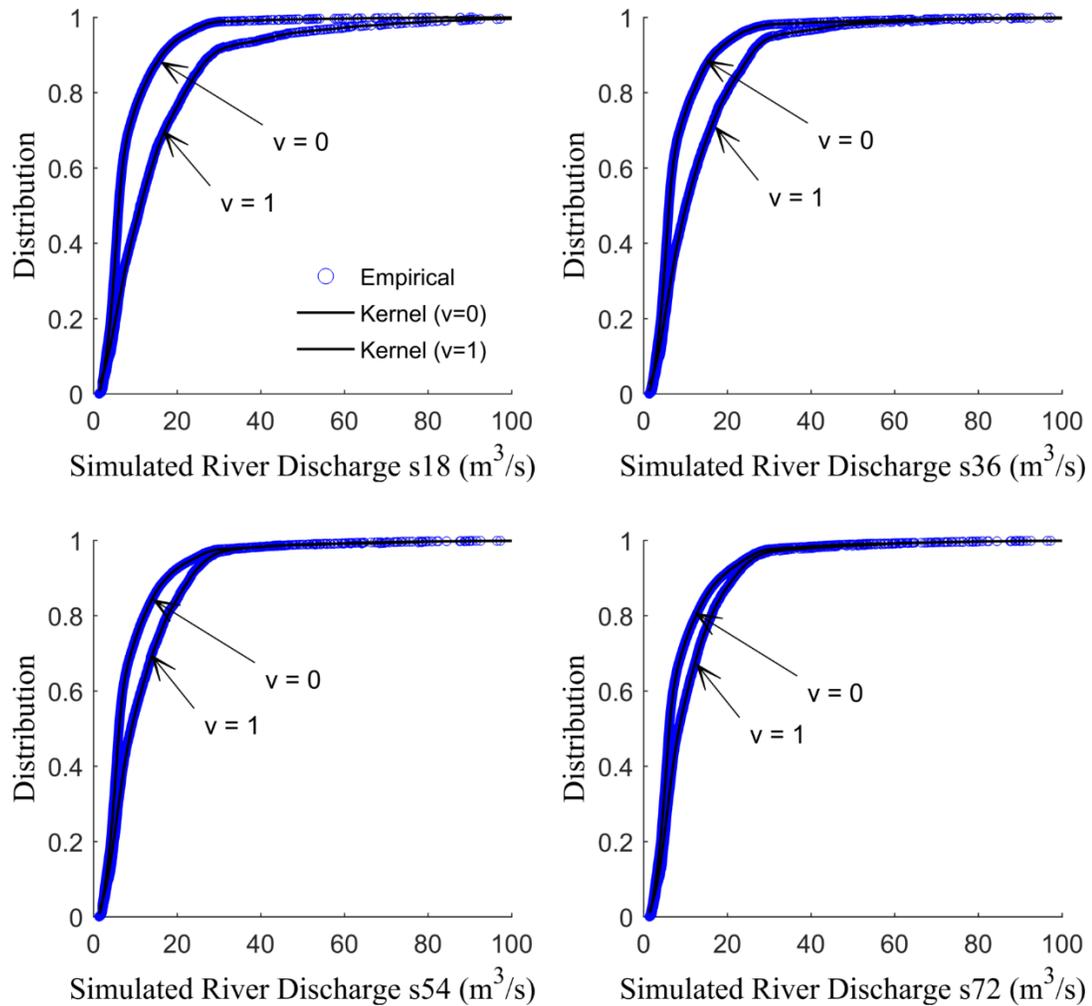


Figure 4-5 Marginal distribution of simulated discharge for s18, s36, s54 and s72 conditional on precipitation indicator

Table 4-2 Dependence parameters of Hydrologic Uncertainty Processor

Coefficient definition	Prior distribution				Posterior distribution							
	$C_{nv} = \prod_{i=1}^n c_{iv}$		$t_{nv}^2 = 1 - C_{nv}^2$		$A_{nv} = \frac{a_{nv}t_{nv}^2}{a_{nv}^2t_{nv}^2 + \sigma_{nv}^2}$		$B_{nv} = \frac{-a_{nv}b_{nv}t_{nv}^2}{a_{nv}^2t_{nv}^2 + \sigma_{nv}^2}$		$D_{nv} = \frac{C_{nv}\sigma_{nv}^2 - a_{nv}d_{nv}t}{a_{nv}^2t_{nv}^2 + \sigma_{nv}^2}$		$T_{nv}^2 = \frac{t_{nv}^2\sigma_{nv}^2}{a_{nv}^2t_{nv}^2 + \sigma_{nv}^2}$	
Lead time n	v = 0	v = 1	v = 0	v = 1	v = 0	v = 1	v = 0	v = 1	v = 0	v = 1	v = 0	v = 1
3	1.00	0.92	0.01	0.14	-0.01	0.22	0.00	0.00	1.01	0.75	0.08	0.36
6	0.98	0.87	0.03	0.24	0.00	0.38	0.00	0.00	0.98	0.58	0.18	0.42
9	0.96	0.85	0.07	0.28	0.03	0.47	0.00	0.00	0.95	0.50	0.26	0.44
12	0.94	0.83	0.11	0.32	0.07	0.51	0.00	0.00	0.90	0.46	0.33	0.45
15	0.92	0.81	0.15	0.34	0.11	0.53	0.00	0.00	0.86	0.44	0.38	0.45
18	0.90	0.79	0.19	0.37	0.14	0.55	0.00	0.00	0.82	0.42	0.42	0.46
21	0.88	0.78	0.23	0.40	0.18	0.57	0.00	0.00	0.79	0.40	0.45	0.48
24	0.86	0.76	0.26	0.42	0.21	0.58	0.00	0.00	0.75	0.39	0.48	0.49
27	0.84	0.74	0.29	0.45	0.24	0.59	0.00	0.00	0.72	0.37	0.50	0.50
30	0.82	0.73	0.32	0.47	0.27	0.60	0.00	0.00	0.70	0.35	0.52	0.52
33	0.80	0.71	0.35	0.50	0.29	0.61	0.00	0.00	0.67	0.34	0.54	0.53
36	0.79	0.69	0.38	0.52	0.31	0.61	0.00	0.00	0.65	0.33	0.55	0.54
39	0.77	0.68	0.41	0.53	0.33	0.62	0.00	0.00	0.63	0.33	0.57	0.55
42	0.75	0.67	0.44	0.55	0.36	0.62	0.00	0.00	0.61	0.32	0.58	0.56
45	0.73	0.66	0.46	0.57	0.37	0.62	0.00	0.00	0.59	0.31	0.59	0.57
48	0.72	0.65	0.49	0.58	0.39	0.61	0.00	0.00	0.57	0.31	0.60	0.58
51	0.70	0.63	0.51	0.60	0.41	0.61	0.00	0.00	0.56	0.30	0.61	0.59
54	0.69	0.62	0.53	0.61	0.42	0.62	0.00	0.00	0.54	0.30	0.62	0.60
57	0.67	0.61	0.55	0.63	0.44	0.62	0.00	0.00	0.53	0.29	0.62	0.61
60	0.66	0.60	0.57	0.64	0.45	0.61	0.00	0.00	0.51	0.28	0.63	0.62
63	0.64	0.59	0.59	0.65	0.46	0.62	0.00	0.00	0.50	0.28	0.64	0.62
66	0.63	0.58	0.61	0.66	0.47	0.62	0.00	0.00	0.49	0.28	0.64	0.62
69	0.61	0.57	0.62	0.68	0.48	0.62	0.00	0.00	0.48	0.27	0.65	0.63
72	0.60	0.55	0.64	0.69	0.49	0.62	0.00	0.00	0.47	0.27	0.65	0.64

4.5.3 Bias Correction of Ensemble Weather Forecasts

In this study, GEPS ensemble weather forecasts were bias corrected using MBCn in three different ways: (i) bias correct each ensemble separately to get bias corrected ensembles, (ii) bias correct each ensemble first and then average them to get bias corrected ensemble mean, (iii) calculate the ensemble mean first and then bias correct the mean. As a result, five types of forecast dataset were obtained: (i) GEPS-raw which means raw GEPS data, (ii) GEPS-BC which is bias-corrected GEPS ensembles, (iii) GEPS-raw-mean that represents ensemble mean of GEPS-raw, (iv) GEPS-BC-mean which is the ensemble mean of bias-corrected GEPS data, and (v) GEPS-mean-BC which stands for bias-corrected GEPS-raw-mean. After bias correction, the raw GEPS and the bias-corrected GEPS data were compared in terms of energy distance score.

Energy distance score measures the statistical discrepancy between sample x and y from two multivariate distributions; energy distance equals zero if and only if distributions of x and y are identical, and it increases with the increase of the discrepancy of distributions (Cannon, 2017; Székely and Rizzo, 2013). The squared energy distance is defined as (Rizzo and Székely, 2016):

$$D^2(F, G) = 2E\|x - y\| - E\|x - x'\| - E\|y - y'\| \geq 0 \quad (4-14)$$

Where F and G denote the cdfs of sample x and y , respectively. $\|\cdot\|$ is the Euclidean norm and E is the expected value, x' denotes an independent and identically distributed copy of x ; similarly, y' is an independent and identically distributed copy of y . The comparison of energy distance score between observations and the five datasets are presented in Figure

4-6. Energy distances between GEPS-raw and observations range from 11.19 to 21.53, while energy distances between GEPS-BC and observations are all within 2 which indicates a large improvement when compared with GEPS-raw. Energy distance between GEPS-raw-mean and observations is 22.75, after bias correcting GEPS-raw-mean, the energy distance drops down to 2.28, and the energy distance between GEPS-BC-mean and observations also reduced to 4.49. The results suggest that MBCn bias correction can significantly decrease the statistical discrepancy between meteorological observations and ensemble weather forecasts.

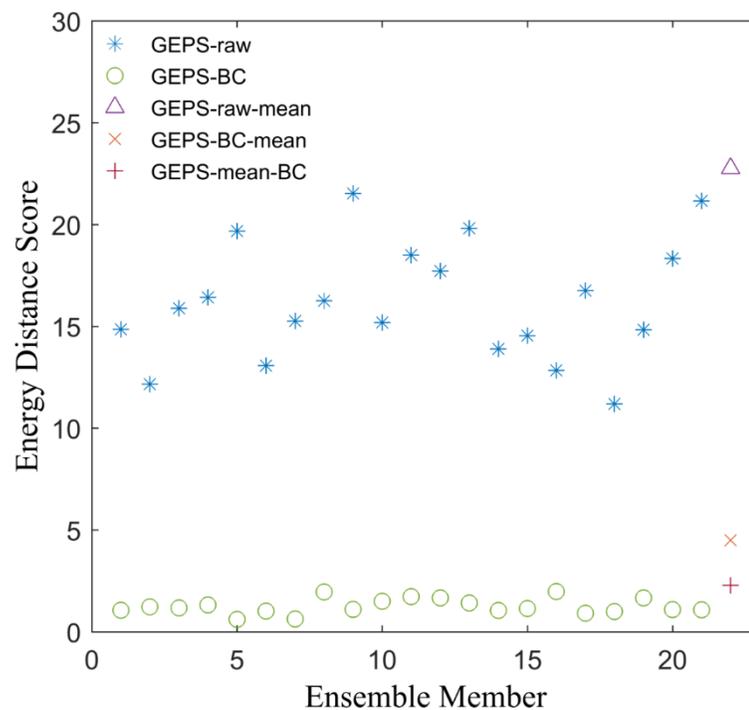


Figure 4-6 Comparison of energy distance score

4.5.4 Experiments and Comparisons

The five types of GEPS dataset generated from bias correction were used as inputs to the calibrated HUP, resulting in seven application scenarios summarized in Table 4-3. The forecast performance of these seven different scenarios were evaluated and compared using multiple verification metrics and visual graphical tools, such as scatter plots, correlation coefficient r , root mean squared error (RMSE), NSE, continuous ranked probability score (CRPS), reliability plots and forecast hydrographs (Jha et al. 2018; Verkade et al. 2017). The forecast horizons considered vary from short-range forecasts (3h - 24 h herein) and medium-range forecasts (24h - 72h herein).

Table 4-3 Brief descriptions of the seven application scenarios

Scenario	Brief description
GEPS-raw	use GEPS-raw data only to run the model and get ensemble forecast
GEPS-BC	use GEPS-BC only to derive ensemble forecast
GEPS-raw+HUP	use HUP post-process the ensemble forecast from GEPS-raw to get probabilistic forecast
GEPS-BC+HUP	use HUP post-process the ensemble forecast from GEPS-BC to obtain probabilistic forecast
GEPS-raw-mean+HUP	apply HUP with GEPS-raw-mean to generate probabilistic forecast
GEPS-BC-mean+HUP	use GEPS-BC-mean with HUP to generate probabilistic forecast
GEPS-mean-BC+HUP	use GEPS-mean-BC with HUP to generate probabilistic forecast

4.5.4.1 Scenario Results Comparisons

In order to get single-valued forecast derivative for performance evaluation, the mean value is used for ensemble forecasts, and the median is used for probabilistic forecasts. For short-range forecasts, scatter plots of single-valued forecasts versus observations for all the seven scenarios (from Table 4-3) are presented in Figure 4-7 and Figure 4-9 (the first four scenarios are in Figure 4-7, and the other three are in Figure 4-9). For medium-range forecasts, scatter plots comparison between using GEPS only and combining GEPS with HUP (the first four scenarios) are presented in Figure 4-8. For both short-range and medium-range, results are shown for selected forecast lead times only instead of all the lead times. In all the plots, the forecast-observation pairs are marked by blue point, the 1:1 diagonals are emphasized by solid black lines, and the x-axes and y-axes are identical. Vertically, the scatter plot panel in each column are from the same scenario which is indicated at the top of the graph. Horizontally, the scatter plot panel in each row stands for the same lead time as indicated at the right edge of the graph. Metrics including r , NSE and RMSE are calculated and presented for every scatter plot.

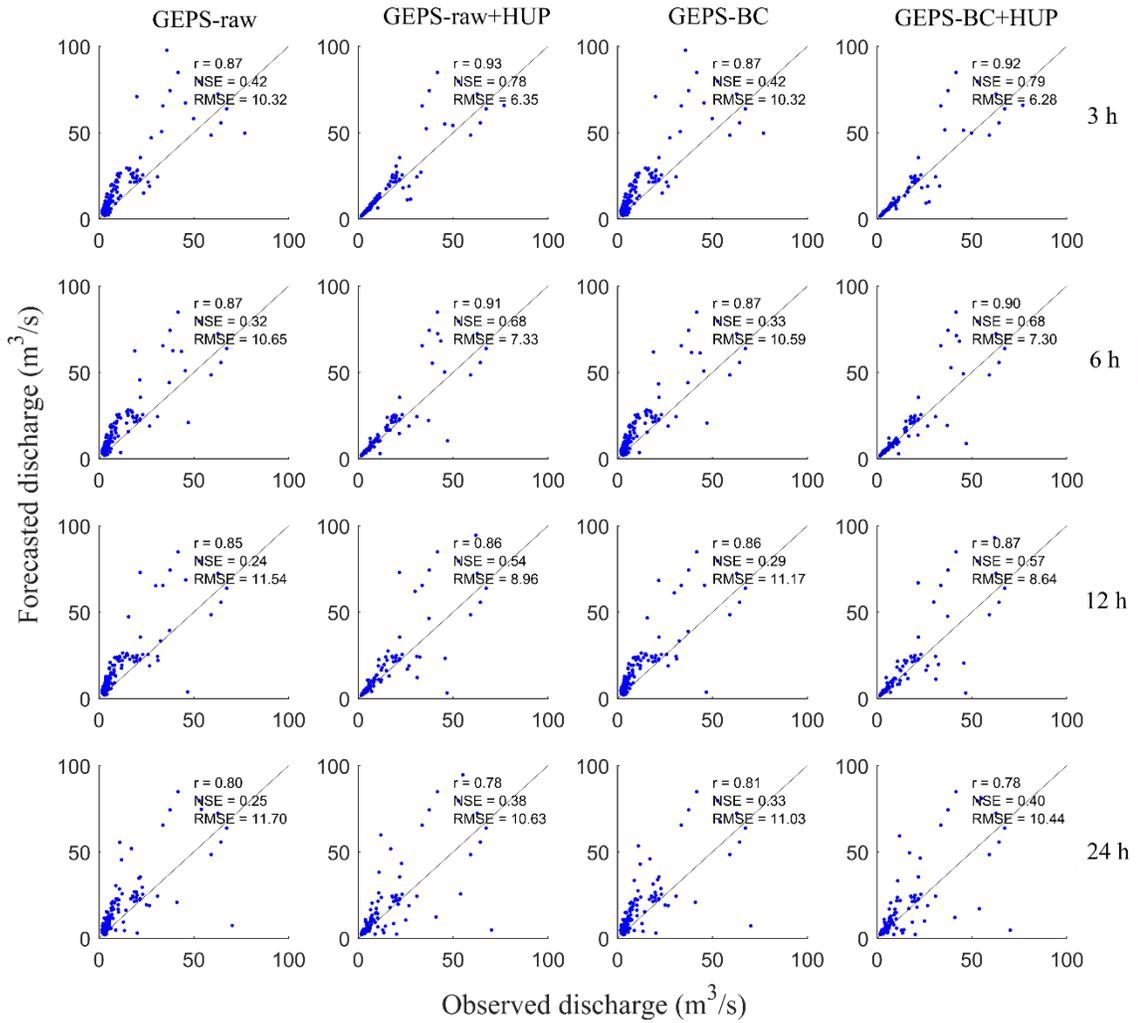


Figure 4-7 Scatter plot comparison for short-range forecasts between using GEPS data only and combining GEPS with HUP

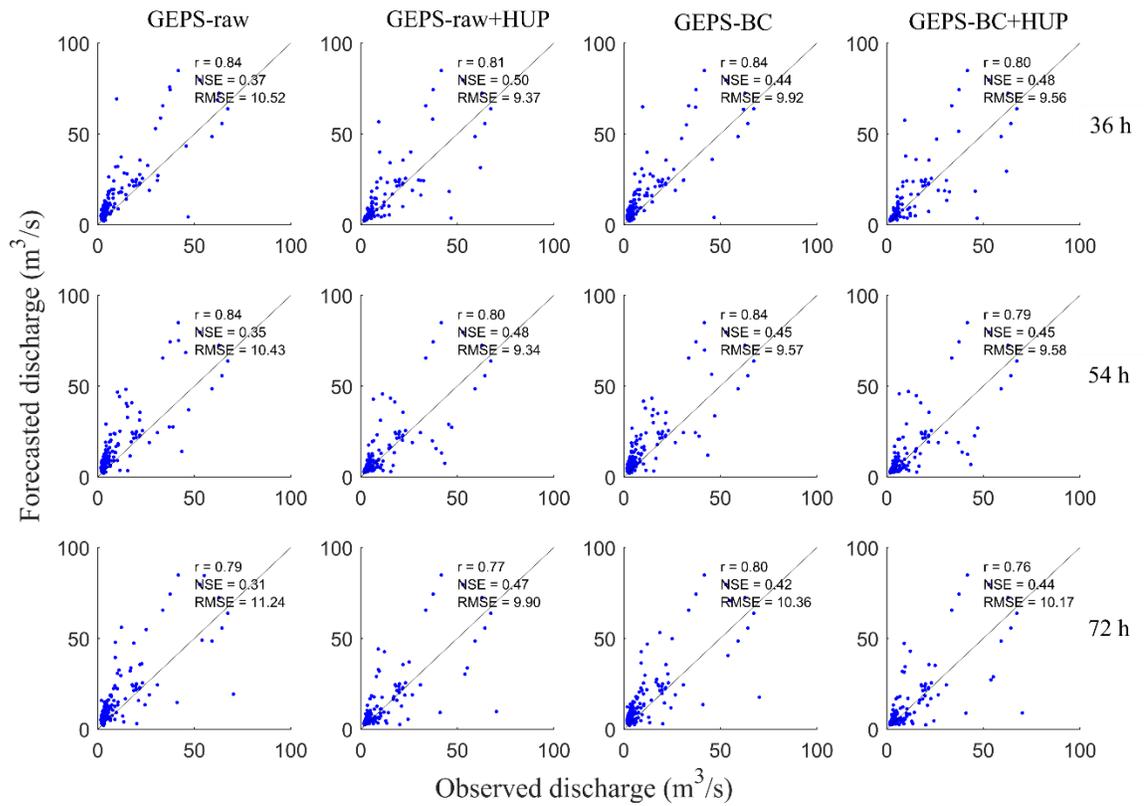


Figure 4-8 Scatter plot comparison for medium-range forecasts between using GEPS data only and combining GEPS with HUP

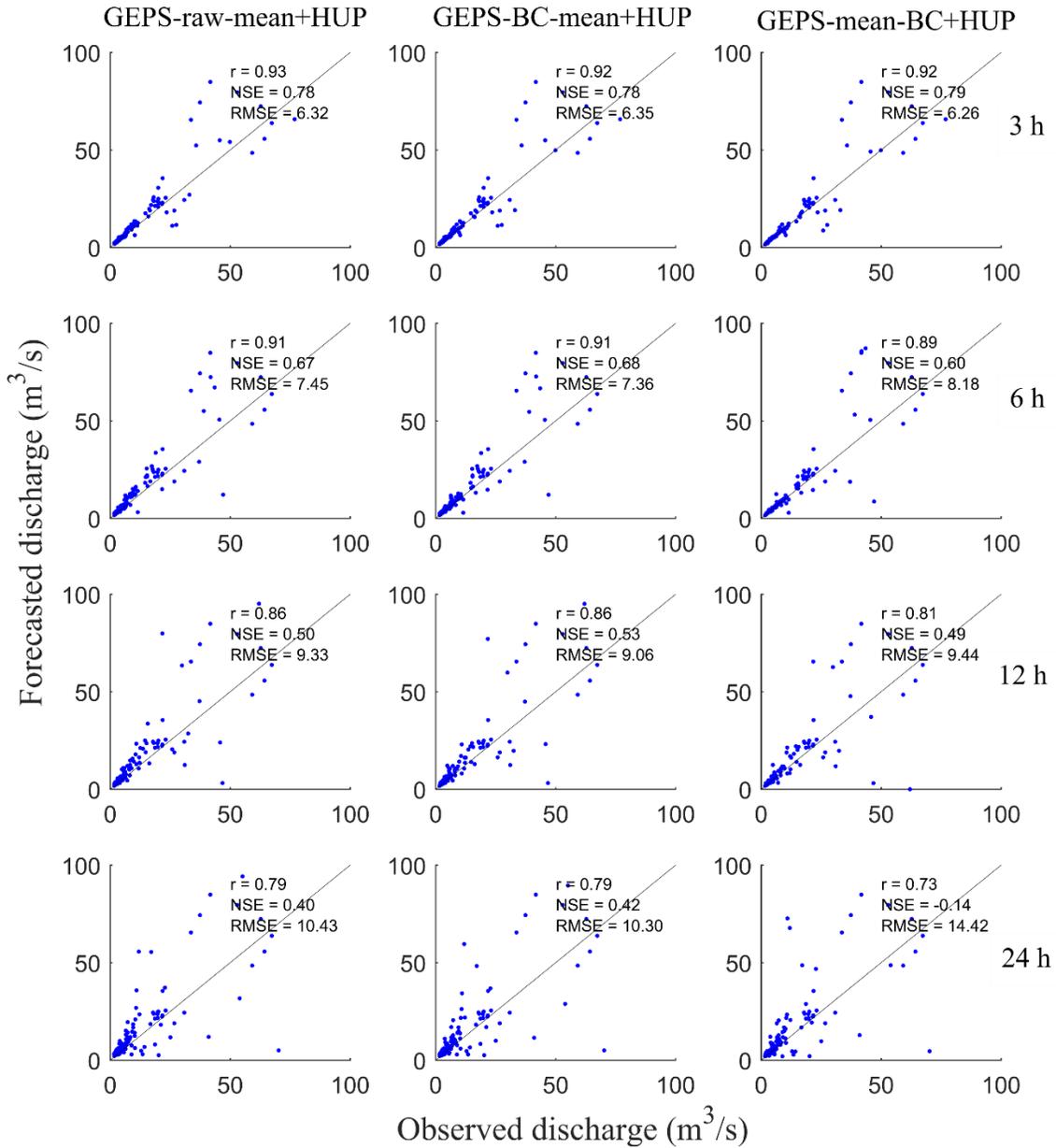


Figure 4-9 Scatter plot comparison for short-range forecasts between using GEPS mean with HUP

As shown in Figure 4-7 and Figure 4-9, Results indicate that for all the scenarios of short-range forecasts, the pairs are more spread with increasing lead time, and larger flow

values have higher spread. As lead time grows, correlation coefficient r decreases, NSE and RMSE follow similar pattern and suggest a worsening trend. Comparison between Figure 4-7 and Figure 4-8 indicates that the performance of combining HUP is promising for short-range forecasts, and worsens for medium-range forecasts. For same forecast lead times, comparisons between using GEPS data only and combining GEPS with HUP (GEPS-raw versus GEPS-raw+HUP and GEPS-BC versus GEPS-BC+HUP) reveal that HUP is able to improve the forecast performance, as lower RMSE and higher NSE are obtained when HUP is used, and the improvement is significant for short-range forecasts. Comparisons between using bias-corrected GEPS and raw GEPS (GEPS-raw versus GEPS-BC and GEPS-raw+HUP versus GEPS-BC+HUP) indicate that bias correcting ensemble weather forecasts could also improve the performance for most of the lead times. However, the improvement is less obvious for medium range forecasts. Scatter plot comparison for short-range forecasts between using GEPS mean with HUP is presented in Figure 4-9, there is no notable difference between GEPS-raw-mean+HUP and GEPS-BC-mean+HUP. For GEPS-mean-BC+HUP, the results are promising for small lead times, while for higher lead times, it shows the most unsatisfactory performance among the seven scenarios. For example, the NSE value for lead time 24 hours is negative, even worse than just using the raw data. This indicates that bias correcting each ensemble member outperforms only bias correcting the ensemble mean.

The NSE values for all the forecast lead times is further analyzed in Figure 4-10, with short-range showing in the upper graph and medium-range showing in the lower graph. The seven different scenarios are divided into three groups, and each group is plotted in

the same color but different line style. For both short-range and medium-range, except GEPS-mean-BC+HUP, NSE values are quite similar across the other four Bayesian scenarios. NSE for these four Bayesian scenarios are generally higher than non-Bayesian scenarios, and the differences become less evident for medium-range forecasts. As for GEPS-mean-BC+HUP which represented by orange dotted line, NSE values are acceptable for short lead times, but deteriorate as lead time exceeds 21 hours, revealing the performance is not stable when GEPS-mean-BC is used as input. The GEPS ensemble forecasts are generated by Global Environmental Multiscale Model (GEM) with different physics parameterizations, data assimilation cycles, and sets of perturbed observations. Therefore, due to the different ensemble configurations, each ensemble may have its unique biases that need to be dealt with independently, simply bias correcting the ensemble mean could possibly result in poor performance (Cui et al. 2012). Comparison between different forecast horizons demonstrates that the performances of short-range forecast are better than medium-range for the Bayesian scenarios. For both graphs, the green solid line for GEPS-BC is located above the green dashed line for GEPS-raw, this further proves the improved performance after bias correction of input data. In terms of NSE results, the best scenario for short-range is GEPS-BC+HUP, followed by GEPS-BC-mean+HUP which presents comparable result over lead time 15 hours. The best scenario for medium range is GEPS-raw-mean+HUP, followed by GEPS-BC-mean+HUP which shows comparable result from lead time 24 hours to 45 hours.

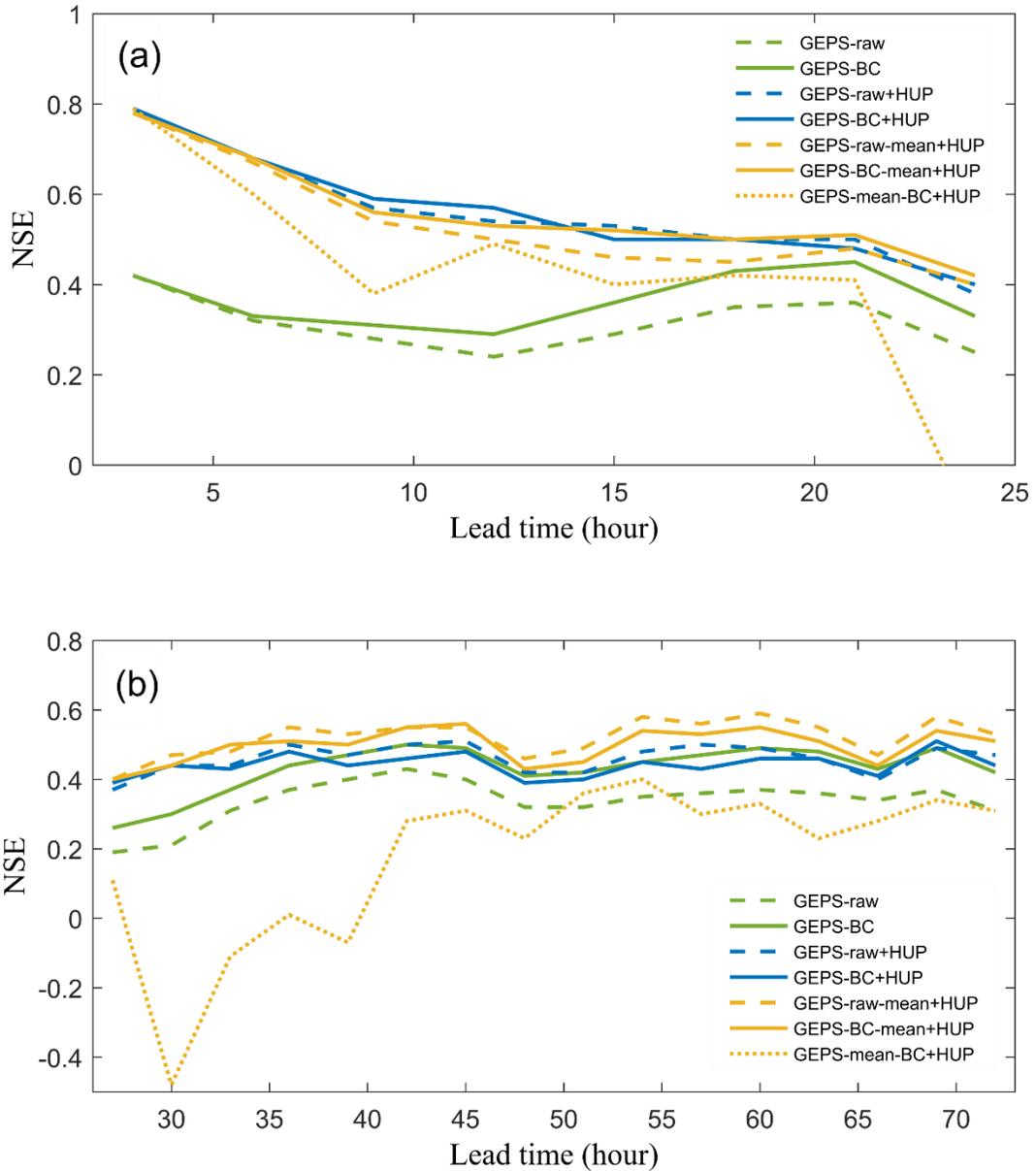


Figure 4-10 Comparison of NSE for different scenarios: (a) short-range; (b) medium-range

In terms of skill of full ensemble forecast and probabilistic forecast, the overall accuracy is assessed by CRPS. CRPS is a commonly used metric for ensemble or probabilistic

verification through measuring the spread of ensembles or forecast distributions against observations:

$$CRPS = \int_{-\infty}^{\infty} (F_{zf}(t) - F_{zo}(t))^2 dt \quad (4-15)$$

Where F_{zf} denotes cdf of ensemble or probabilistic forecast, and F_{zo} denotes cdf of observation. F_{zo} is a Heaviside function, and it equals 1 when values are greater than observed value and otherwise equal to 0. The forecast is considered more accurate if CRPS gets more close to zero.

Figure 4-11 presents the comparison of mean CRPS for different scenarios for both short-range (upper graph) and medium-range (lower graph). The line style and color for each scenario (Figure 4-11) are consistent with Figure 4-10. According to visual inspection, the results are in accordance with the single-valued verification results. For the Bayesian scenarios, the CRPS values are very low for short-range (below 3.00) and gradually rises with increasing lead time, the CRPS values for medium-range are generally higher than short-range as can be expected. The CRPS values for non-Bayesian scenarios such as GEPS-raw and GEPS-BC are consistently higher than the Bayesian scenarios for both short-range and medium-range, indicating HUP which worked as post-processor of ensemble forecast is able to improve the forecast performance across all lead times. However, the improvement becomes less pronounced as lead time grows. The deviation between GEPS-raw+HUP and GEPS-BC+HUP is subtle, and this also applies to GEPS-raw-mean+HUP and GEPS-BC-mean+HUP. For short-range, the CRPS values for GEPS-mean-BC+HUP are similar to the other four Bayesian scenarios. While for

medium-range, the CRPS result for GEPS-mean-BC+HUP is worse than the other four Bayesian scenarios, and converges with GEPS-BC at lead time of 63 hours. This reveals that bias correcting the ensemble mean only rather than bias correcting each ensemble may result in unstable performance. For both graphs, the CRPS values for GEPS-BC are lower than GEPS-raw, and larger differences are shown as lead time increases. The comparison demonstrates that MBCn bias correction is able to improve the ensemble forecast for all the forecast lead times, and the improvement increases with the increase of lead time. In general, in terms of CRPS results, the best scenario for short-range forecast is GEPS-BC+HUP. GEPS-raw-mean+HUP shows the best performance for medium-range, and GEPS-BC-mean+HUP takes second place.

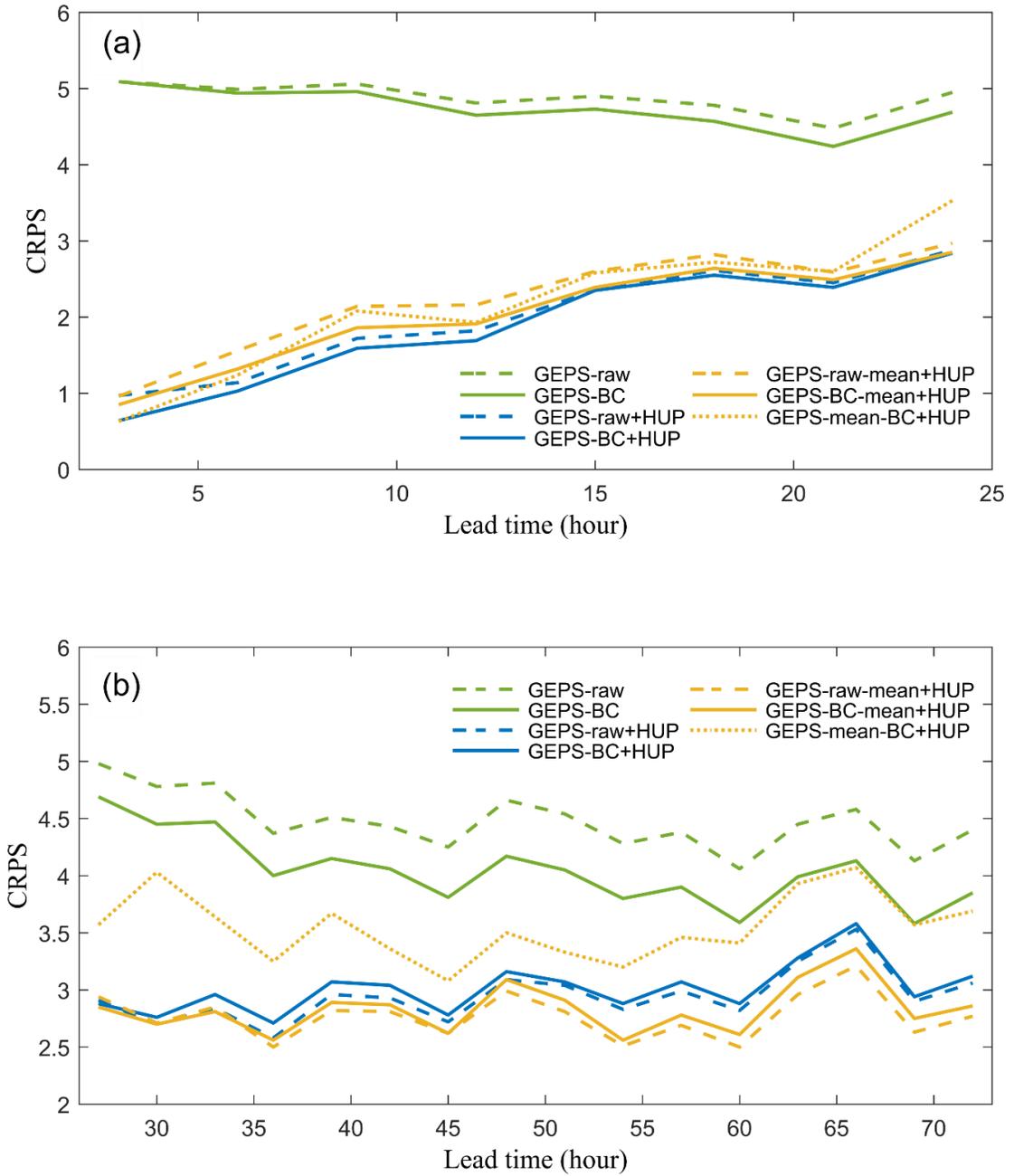


Figure 4-11 Comparison of CRPS for different scenarios: (a) short-range; (b) medium-range

4.5.4.2 Comparison of Reliability

In addition, the reliabilities of the best three scenarios (GEPS-BC+HUP for short-range, GEPS-raw-mean+HUP and GEPS-BC-mean+HUP for medium-range) identified are assessed using reliability plot introduced by Laio and Tamea (2007). Reliability plot is used to evaluate the degree to which the probabilistic forecasts are reliable. Provided that X_i is the observed flow at time t_i , then Z_i is the cdf value derived from the probabilistic forecast that corresponding to X_i , and mathematically expressed by $Z_i = P_i(X_i)$. And R_i denotes the corresponding rank of Z_i when Z_i values are sorted in increasing order, their empirical cumulative distribution function is calculated using R_i divided by sample size n . Consequently, reliability plot is a plot of Z_i values versus R_i/n , the shape of the reliability curves are used to judge the reliability of the forecast. Besides, Kolmogorov confidence bands are shown along with the reliability curve in the same plot. They are two lines parallel to the bisector at the same distance, one is located above, and another is situated below. The distance between the confidence band and the bisector line depends on the significance level α and computed via $q(\alpha)/\sqrt{n}$, here α of 0.05 is used and $q(\alpha=0.05)=1.36$. The forecast is deemed reliable under the condition that the $(Z_i, R_i/n)$ pairs are distributed close to bisector and remain inside the confidence bands; otherwise, forecast issues are detected.

Figure 4-12 shows the reliability plots of the three best probabilistic forecast scenarios for both short-range (upper graph) and medium-range (lower graph) for a selection of lead times, the $(Z_i, R_i/n)$ points for different lead time are presented by different marker type and color. The bisector line is emphasized by black solid line, and the Kolmogorov

confidence bands are plotted in black dashed line. The evaluation criterion of the reliability curve can be found in Laio and Tamea (2007). It presents several possible outcomes, curve below the bisector line indicates under prediction, while curve above the line means over prediction. The S-shaped curve reveals problem of spread of the distribution, either narrow forecast or large forecast. For the reliability plots obtained, in most cases the forecast can be considered relatively reliable, since most of the points are distributed within the significance band and near the bisector. For GEPS-raw-mean+HUP, few points for lead time 3 hours and 6 hours are located outside the confidence bands, and few points for lead time 12 hours and 24 hours reach the confidence line. While for GEPS-BC-mean+HUP, after bias correction process of GEPS forecasts, these points move more close to the bisector, indicating that for short-range forecasts, using bias-corrected GEPS as input appears to be more reliable than using raw GEPS as input. As such, even though GEPS-raw based scenarios show comparable performance with GEPS-BC based scenarios in terms of forecast skill, bias correction of ensemble weather inputs is still recommended, and the improvement brought by bias correction would be further enhanced if a longer training dataset is available (Cui et al. 2012). We should also note that it is possible that a prediction can pass the test but has no operational value, it is recommended to use this verification method together with some other method to make a multifaceted assessment (Laio and Tamea 2007). Overall, in terms of probabilistic and ensemble verification measures along with single-valued verification measures, it turns out that the performances of short-range probabilistic forecasts are good, and GEPS-BC+HUP performs best for short-range.

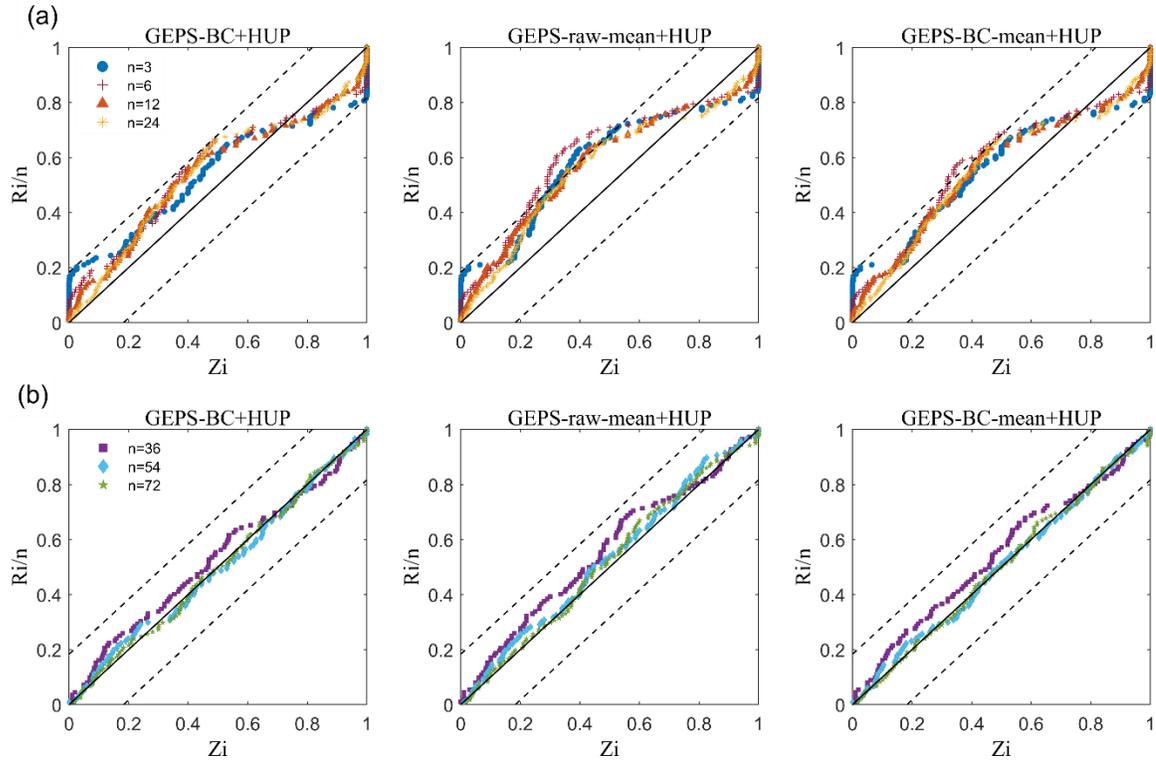


Figure 4-12 Comparison of reliability plot for different scenarios: (a) short-range; (b) medium-range

4.5.4.3 Forecast Hydrographs

A sample of forecast hydrographs at the watershed outlet on 27 June 2015 00Z is shown in Figure 4-13. The upper figure is the probabilistic forecast using GEPS-raw as input data, and the lower one presents the forecast results using GEPS-BC as input. The solid red line is the observed discharge, and the shaded area demonstrates 30% probability interval of probabilistic forecast using GEPS with HUP. The mean of ensemble forecasts is represented by the black dash-dot line, and the median of the predictive distribution from probabilistic forecast is exhibited by the blue dashed line. For both figures, the

ensemble mean lies above the observed line, indicating a flow overestimation. However, after post-processing by HUP, the predictive median move closer to the observed, and the predictive distribution or so-called uncertainty bound could capture most of the observations. These results further indicate the Bayesian revising effect of the HUP processor. Because of the large uncertainties of ensemble formation and the coarse spatial resolution, the quality of the raw GEPS is very limited because of the time lag and large bias which is very challenging to correct. Although some biases could be reduced by the MBCn as shown in FIG. 13, there is still some remaining bias – suggesting that there is still room for improvement. This may require a longer training dataset, observation and weather forecasts with higher spatial resolution, or alternative bias correction methods. It should be noted that Figure 4-13 is only one example on a particular forecast time, and not necessarily the general behavior of all the forecast hydrographs. The behavior of the probabilistic forecast, demonstrated by the uncertainty bound, is conditional on the initial condition at the particular forecast time, the ensemble discharge forecasts forced by corresponding ensemble weather data, and the precipitation indicator (that defines which branch it should be assigned to in the two-branch HUP) which determined by the precipitation forecast.

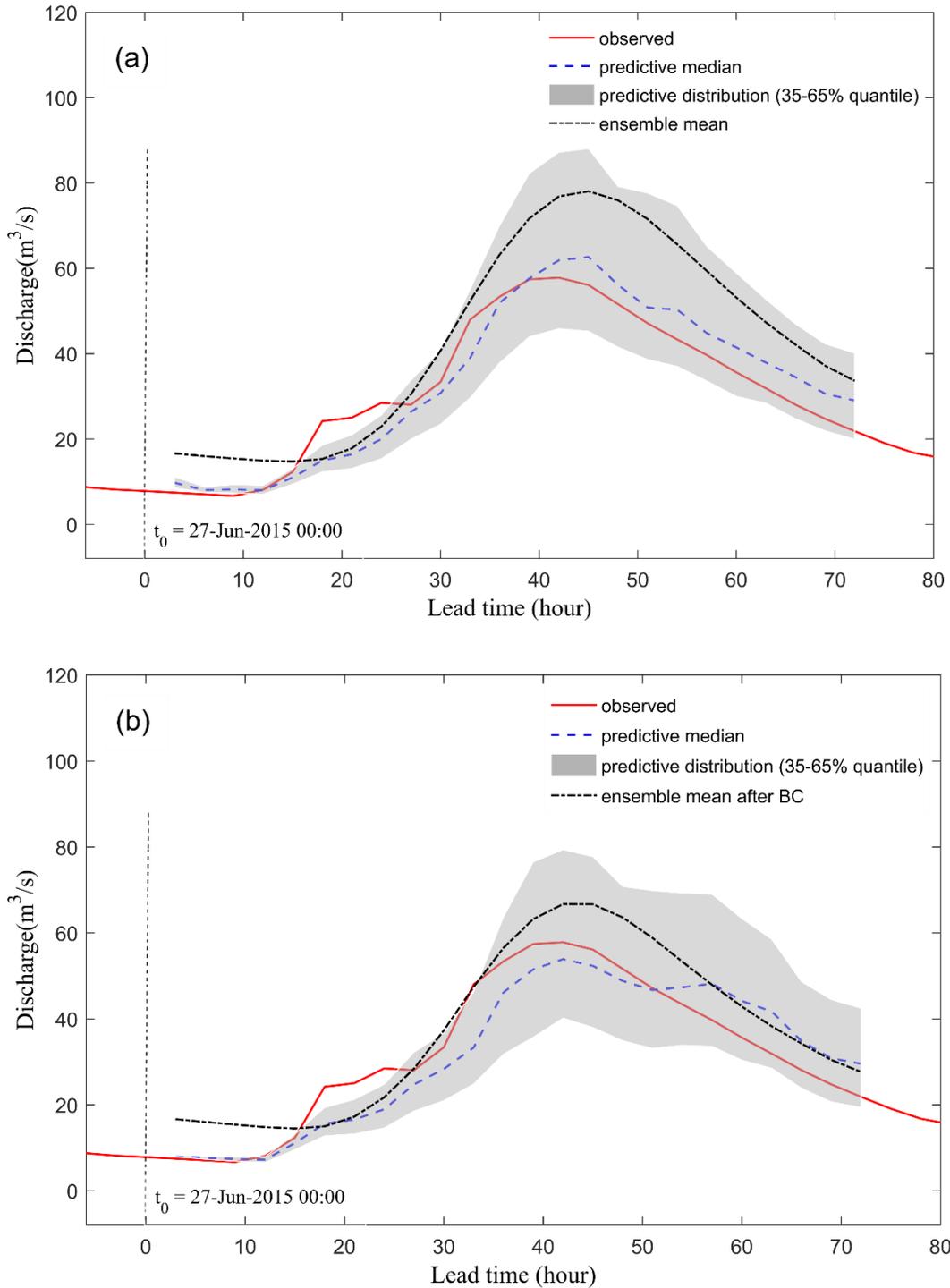


Figure 4-13 Ensemble forecasts and probabilistic forecasts: (a) using GEPS-raw as input; (b) using GEPS-BC as input (take June 27, 2015 for example)

4.6 Conclusions

This paper has presented an application of Hydrologic Uncertainty Processor for post-processing ensemble streamflow forecasts which were forced by ensemble weather forecasts. Conditional on the ensemble forecasts and initial condition, the Bayesian processor updates the prior density derived from historical observations to ensemble posterior densities via likelihood functions. The ensemble of revised posterior densities are subsequently lumped into a representative one to assess the uncertainties. To remove the bias of the ensemble weather forecasts, GEPS was bias corrected through MBCn approach in three different ways, resulting in five sets of forecast data. Consequently, seven different forecast scenarios were developed by using raw GEPS and bias-free GEPS independently, as well as using them together with HUP. The prediction skills of different forecast horizons for different scenarios were assessed and compared through various verification metrics. Based on detailed analysis of the results, the following conclusions can be drawn.

- (i) On the whole, the performances of the Bayesian scenarios are promising for short-range (3h - 24h) forecasts, but showed little to no improvement for medium-range (24h - 72h) forecasts. The best scenario for short-range forecast is GEPS-BC+HUP, which is applying bias correction to each ensemble plus applying hydrologic uncertainty processor.
- (ii) HUP, a hydrologic post-processor for resulting ensemble forecasts from raw GEPS or bias-corrected GEPS, is able to improve the performance for both short-range and medium-range forecasts. This is indicated by lower RMSE and CRPS, and higher r and NSE when HUP is used (except for GEPS-mean-BC scenario). The improvement is

significant for short lead times and becomes less evident as forecast lead time grows. Most of the forecasts from the selected Bayesian scenarios appeared reliable as the points in the reliability plot are located within the bands and close to the bisector line.

(iii) MBCn, works like a meteorological post-processor of weather forecasts, can greatly reduce the statistical discrepancy between GEPS and weather observations. It could also yield improved short-range flood forecasts, which is indicated by improved NSE, CRPS and reliability plot. However, the improvement is less obvious compared with HUP.

(iv) The performance of GEPS-mean-BC+HUP is not stable and deteriorates at certain point, indicating that each ensemble member should be bias corrected instead of just bias correcting the ensemble mean, under the condition that bias-free ensemble weather forecasts are preferred.

(v) For both short-range and medium-range, GEPS-BC outperforms GEPS-raw; however, results are quite similar between GEPS-raw+HUP and GEPS-BC+HUP, as well as GEPS-raw-mean+HUP and GEPS-BC-mean+HUP. It reveals that the performance difference between using raw and bias-corrected weather forecasts becomes less noticeable after Bayesian revision process. However, bias correction does enhance the forecast reliability.

Future work will involve testing alternative ensemble weather forecasts with longer archive and higher spatial resolution, alternative meteorological post-processing methods along with hydrologic post-processing technique, to further assess the potential of HUP for operational flood forecasting.

4.7 Acknowledgments

This work was supported by the Natural Science and Engineering Research Council (NSERC) Canadian FloodNet (Grant # NETGP-451456) and the China Scholarship Council (CSC). The data were obtained from Toronto and Region Conservation Authority, Water Survey of Canada and Environment and Climate Change Canada. The authors would like to thank Dr. Daniela Biondi at University of Calabria for her technical support. The authors are grateful to Dr. Alex J. Cannon for making the bias correction code available. The authors acknowledge two anonymous reviewers for their comments that helped to improve the manuscript.

4.8 References

- Anderson, E., 2006: Snow accumulation and ablation model–SNOW-17. US National Weather Service, Silver Spring, MD.
- Ashkar, F., and F. Aucoin, 2012: Choice between competitive pairs of frequency models for use in hydrology: a review and some new results. *Hydrological Sciences Journal*, **57**, 1092–1106, doi:10.1080/02626667.2012.701746.
- Bergström, S., 1976: Development and application of a conceptual runoff model for Scandinavian catchments.
- Cannon, A. J., S. R. Sobie, and T. Q. Murdock, 2015: Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes?. *Journal of Climate*, **28**, 6938–6959, doi:10.1175/JCLI-D-14-00754.1.
- Cannon, A. J., 2016: Multivariate bias correction of climate model output: Matching marginal distributions and intervariable dependence structure. *Journal of Climate*, **29**, 7045–7064, doi:10.1175/JCLI-D-15-0679.1.
- Cannon, A. J., 2017: Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables. *Climate Dynamics*, **50**, 31–49, doi:10.1007/s00382-017-3580-6.
- Coulibaly, P., F. Anctil, and B. Bobée, 2001: Multivariate reservoir inflow forecasting using temporal neural networks. *Journal of Hydrologic Engineering*, **6**, 367–376, doi:10.1061/(ASCE)1084-0699(2001)6:5(367).
- Cui, B., Z. Toth, Y. Zhu, and D. Hou, 2012: Bias correction for global ensemble forecast. *Weather and Forecasting*, **27**, 396–410, doi:10.1175/WAF-D-11-00011.1.
- Eden, J. M., M. Widmann, D. Grawe, and S. Rast, 2012: Skill, correction, and downscaling of GCM-simulated precipitation. *Journal of Climate*, **25**, 3970–3984, doi:10.1175/JCLI-D-11-00254.1.

- Gneiting, T., A. E. Raftery, A. H. Westveld III, and T. Goldman, 2005: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Monthly Weather Review*, **133**, 1098-1118, doi: org/10.1175/MWR2904.1
- Han, S., and P. Coulibaly, 2017: Bayesian flood forecasting methods: A review. *Journal of Hydrology*, **551**, 340–351, doi:10.1016/j.jhydrol.2017.06.004.
- Han, S., P. Coulibaly, and D. Biondi, 2018: Assessing hydrologic uncertainty processor performance for flood forecasting: A case study in a semi-urban watershed. Submitted to *Journal of Hydrologic Engineering*, HEENG-4291.
- He, M., T. S. Hogue, K. J. Franz, S. A. Margulis, and J. A. Vrugt, 2011a: Characterizing parameter sensitivity and uncertainty for a snow model across hydroclimatic regimes. *Advances in Water Resources*, **34**, 114–127, doi:10.1016/j.advwatres.2010.10.002.
- He, M., T. S. Hogue, K. J. Franz, S. A. Margulis, and J. A. Vrugt, 2011b: Corruption of parameter behavior and regionalization by model and forcing data errors: A Bayesian example using the SNOW17 model. *Water Resources Research*, **47**, 1–17, doi:10.1029/2010WR009753.
- Houle, E. S., B. Livneh, and J. R. Kasprzyk, 2017: Exploring snow model parameter sensitivity using Sobol' variance decomposition. *Environmental Modelling and Software*, **89**, 144–158, doi:10.1016/j.envsoft.2016.11.024.
- Intergovernmental Panel on Climate Change, 2015: Climate change 2014: Mitigation of climate change. Cambridge University Press.
- Jha, S. K., D. L. Shrestha, T. A. Stadnyk, and P. Coulibaly, 2018: Evaluation of ensemble precipitation forecasts generated through post-processing in a Canadian catchment. *Hydrology and Earth System Sciences*, **22**, 1957–1969, doi:10.5194/hess-22-1957-2018.
- Kelly, K. S., and R. Krzysztofowicz, 2000: Precipitation uncertainty processor for probabilistic river stage forecasting. *Water Resources Research*, **36**, 2643-2653, doi:10.1029/2000WR900061.

- Koenker, R., 2005: Quantile Regression (No.38). Cambridge University Press.
- Krzysztofowicz, R., 1999: Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resources Research*, **35**, 2739–2750, doi:10.1029/1999WR900099.
- Krzysztofowicz, R., 2001: Integrator of uncertainties for probabilistic river stage forecasting: precipitation-dependent model. *Journal of Hydrology*, **249**, 69–85, doi:10.1016/S0022-1694(01)00413-9.
- Krzysztofowicz, R., 2002: Bayesian system for probabilistic river stage forecasting. *Journal of Hydrology*, **268**, 16–40, doi:10.1016/S0022-1694(02)00106-3.
- Krzysztofowicz, R., and K. S. Kelly, 2000: Hydrologic uncertainty processor for probabilistic river stage forecasting. *Water Resources Research*, **36**, 3265–3277, doi:10.1029/2000WR900108.
- Krzysztofowicz, R., and H. D. Herr, 2001: Hydrologic uncertainty processor for probabilistic river stage forecasting: precipitation-dependent model. *Journal of Hydrology*, **249**, 46–68, doi:10.1016/S0022-1694(01)00412-7.
- Krzysztofowicz, R., and C. J. Maranzano, 2004: Bayesian system for probabilistic stage transition forecasting. *Journal of Hydrology*, **299**, 15–44, doi:10.1016/J.JHYDROL.2004.02.013.
- Laio, F., and S. Tamea, 2007: Verification tools for probabilistic forecast of continuous hydrological variables. *Hydrology and Earth System Sciences*, **11**, 1267–1277.
- Li, H., J. Sheffield, and E. F. Wood, 2010: Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *Journal of Geophysical Research Atmospheres*, **115**, doi:10.1029/2009JD012882.
- Maraun, D., 2016: Bias Correcting Climate Change Simulations - a Critical Review. *Current Climate Change Reports*, **2**, 211–220, doi:10.1007/s40641-016-0050-x.
- Maurer, E. P., and H. G. Hidalgo, 2008: Utility of daily vs. monthly large-scale climate

- data: An intercomparison of two statistical downscaling methods. *Hydrology and Earth System Sciences*, **12**, 551–563, doi:10.5194/hess-12-551-2008.
- Merz, R., and G. Blöschl, 2004: Regionalisation of catchment model parameters. *Journal of Hydrology*, **287**, 95–123, doi:10.1016/j.jhydrol.2003.09.028.
- Oudin, L., F. Hervieu, C. Michel, C. Perrin, V. Andréassian, F. Anctil, and C. Loumagne, 2005a: Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 2 - Towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling. *Journal of Hydrology*, **303**, 290–306, doi:10.1016/j.jhydrol.2004.08.026.
- Oudin, L., C. Michel, and F. Anctil, 2005b: Which potential evapotranspiration input for a lumped rainfall-runoff model?: Part 1 - Can rainfall-runoff models effectively handle detailed potential evapotranspiration inputs? *Journal of Hydrology*, **303**, 275–289, doi:10.1016/j.jhydrol.2004.08.025.
- Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, **133**, 1155-1174.
- Razavi, T., and P. Coulibaly, 2016: Improving streamflow estimation in ungauged basins using a multi-modelling approach. *Hydrological Sciences Journal*, **61**, 2668–2679, doi:10.1080/02626667.2016.1154558.
- Razavi, T., and P. Coulibaly, 2017: An evaluation of regionalization and watershed classification schemes for continuous daily streamflow prediction in ungauged watersheds. *Canadian Water Resources Journal*, **42**, 2–20, doi:10.1080/07011784.2016.1184590.
- Reggiani, P., and A. H. Weerts, 2008a: Probabilistic quantitative precipitation forecast for flood prediction: An application. *Journal of Hydrometeorology*, **9**, 76–95, doi:10.1175/2007JHM858.1.
- Reggiani, P., and A. H. Weerts, 2008b: A Bayesian approach to decision-making under

- uncertainty: An application to real-time forecasting in the river Rhine. *Journal of Hydrology*, **356**, 56–69, doi:10.1016/j.jhydrol.2008.03.027.
- Reggiani, P., M. Renner, A. H. Weerts, and P. A. H. J. M. Van Gelder, 2009: Uncertainty assessment via Bayesian revision of ensemble streamflow predictions in the operational river Rhine forecasting system. *Water Resources Research*, **45**, 1–14, doi:10.1029/2007WR006758.
- Regonda, S. K., S., D.-J. Seo, B. Lawrence, J. D. Brown, and J. Demargne, 2013: Short-term ensemble streamflow forecasting using operationally-produced single-valued streamflow forecasts - A Hydrologic Model Output Statistics (HMOS) approach. *Journal of Hydrology*, **497**, 80–96, doi:10.1016/j.jhydrol.2013.05.028.
- Rizzo, M. L., and G. J. Székely, 2016: Energy distance. *Wiley Interdisciplinary Reviews: Computational Statistics*, **8**, 27–38, doi:10.1002/wics.1375.
- Samuel, J., P. Coulibaly, and R. A. Metcalfe, 2011: Estimation of continuous streamflow in Ontario ungauged basins: comparison of regionalization methods. *Journal of Hydrologic Engineering*, **16**, 447–459, doi:10.1061/(ASCE)HE.1943-5584.0000338.
- Schefzik, R., 2016: Combining parametric low-dimensional ensemble postprocessing with reordering methods. *Quarterly Journal of the Royal Meteorological Society*, **142**, 2463–2477, doi:10.1002/qj.2839.
- Seibert, J., 1999: Regionalisation of parameters for a conceptual rainfall-runoff model. *Agricultural and Forest Meteorology*, **98–99**, 279–293, doi:10.1016/S0168-1923(99)00105-7.
- Székely, G. J., and M. L. Rizzo, 2013: Energy statistics: A class of statistics based on distances. *Journal of Statistical Planning and Inference*, **143**, 1249–1272, doi:10.1016/j.jspi.2013.03.018.
- Todini, E., 2008: A model conditional processor to assess predictive uncertainty in flood forecasting. *Intl. J. River Basin Management*, **6**, 123–137, doi:10.1080/15715124.2008.9635342.

United Nations, 2004: Guidelines for reducing flood losses. U.N. Int. Strategy Disaster Reduct., Geneva, Switzerland.

Verkade, J. S., J. D. Brown, F. Davids, P. Reggiani, and A. H. Weerts, 2017: Estimating predictive hydrological uncertainty by dressing deterministic and ensemble forecasts; a comparison, with application to Meuse and Rhine. *Journal of Hydrology*, **555**, 257–277, doi:10.1016/j.jhydrol.2017.10.024.

Vrugt, J. A., C. G. H. Diks, H. V. Gupta, W. Bouten, and J. M. Verstraten, 2005: Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. *Water Resources Research*, **41**, 1–17, doi:10.1029/2004WR003059.

Wang, L., and W. Chen, 2014: A CMIP5 multimodel projection of future temperature, precipitation, and climatological drought in China. *International Journal of Climatology*, **34**, 2059–2078, doi:10.1002/joc.3822.

Chapter 5. Assessing the Effects of Input and Model Type on Bayesian Ensemble Uncertainty Processor

Summary of Paper 3: Han, S. and Coulibaly, P. (2019). Assessing the Effects of Input and Model Type on Bayesian Ensemble Uncertainty Processor. *Advances in Water Resources*, under review.

The goal of this research was to investigate the effects of weather input type and hydrologic model type on the performance of Bayesian Ensemble Uncertainty Processor (BEUP). Thus, BEUP with deterministic/ensemble weather predictions and lumped/semi-distributed hydrologic models were tested and compared.

Key findings of this research include:

- BEUP performs well in capturing peak flows.
- BEUP with the semi-distributed hydrologic model outperforms the lumped model in terms of accuracy and reliability.
- The improvement brought by the hydrologic model type is more significant than the input data type.
- Using BEUP with semi-distributed model is recommended for short-term flood forecast (1 day ahead) with uncertainty estimation.

5.1 Abstract

Meteorological uncertainty and hydrologic uncertainty are two major sources of uncertainty in flood forecast, to adequately account for both of them, the Bayesian Ensemble Uncertainty Processor (BEUP) was applied to post-process the ensemble forecasts. The meteorological uncertainty, which is represented by ensembles of weather prediction, propagates through the hydrologic model from the weather input to streamflow output. After the post-processing of BEUP, hydrologic uncertainty is quantified and is then added as another layer of uncertainty. The integration of these two uncertainties provides an estimation of the predictive total uncertainty. To investigate the factors that have impact on the performance of BEUP, two types of forcing data were used: deterministic weather predictions with ensemble dressing and ensemble weather predictions, and the hydrologic model combined with BEUP was set up in both lumped and semi-distributed scheme. The different combinations of input data type and hydrologic model type lead to four different probabilistic forecast scenarios, and their performances were compared by various evaluation metrics. Comparisons among the different scenarios indicate that using BEUP with semi-distributed model yields more accurate and reliable flood forecasts than with lumped model. It was also shown that using dressed deterministic weather predictions as input outperforms the use of ensemble weather predictions. Results show that the uncertainty bound generated from BEUP capture well the peak flows. It was also found that the improvement brought by model type is more significant than data type. BEUP combined with semi-distributed model is recommended for short-term flood forecast with uncertainty estimate.

Keywords: Flood forecast; Uncertainty assessment; Ensemble forecast; Post-processing; Bayesian theory.

5.2 Introduction

Flood forecasting based on numerical model provides an effective nonstructural measure in flood management (Munoz and Constantinescu, 2018; Schendel and Thongwichian, 2017). Traditionally, the flood forecasts are produced in a deterministic way, which give only a single value for a particular time, but the forecast uncertainty is not considered (Jones and Kay, 2007; Linde et al., 2017). Over the last two decades, ensemble forecasts forced by ensembles of Numerical Weather Predictions (NWP) have gained popularity (Schaake et al., 2007). However, the uncertainty considered for these forecasts is limited to the uncertainty in NWP, and the ensemble output may not be sufficient to represent the full distribution of predictive probability (Biondi and Todini, 2018). Therefore, whether the forecasts are deterministic or ensemble, they have to be treated using post-processing approach to adequately account for the total uncertainty (Li et al., 2017).

Reviews of the post-processing methods for hydrologic forecasts can be found in Li et al. (2017) and Han and Coulibaly (2017). Some commonly used methods include the Hydrologic Uncertainty Processor (HUP) which updates prior distribution into posterior distribution by assimilating new information (Han et al., 2019; Krzysztofowicz and Herr, 2001), Model Conditional Processor (MCP) which is able to deal with multi-model, multi-site and multi-lead time problem (Coccia and Todini, 2011; Todini, 2013), Bayesian Model Averaging (BMA) which is designed to assess the posterior mean and variance of predictand conditional on several model forecasts (Fragoso et al., 2018;

Masadgar and Moradkhani, 2014), Quantile Regression (QR) which is based on statistical analysis of model error (Weerts et al., 2011), Ensemble Model Output Statistics (EMOS) which post-processes ensemble forecasts using multiple linear regression (Hemri et al., 2015), and etc. There are many post-processors available in the literature, and the Bayesian Ensemble Uncertainty Processor (BEUP) investigated herein is among the most robust approaches.

BEUP is an extension of HUP which was developed to quantify hydrologic uncertainty based on Bayes' theorem, and BEUP has adapted HUP to an ensemble prediction framework to treat multiple responses derived from meteorological ensembles (Reggiani et al., 2009). In hydrologic prediction, meteorological uncertainty (uncertainties in weather predictions) and hydrologic uncertainty (aggregate of uncertainties related to modelling process) are two dominant sources of uncertainty (Seo et al., 2006), and BEUP is able to address both of them. Within the framework of BEUP, the meteorological uncertainty is estimated by ensembles of weather prediction and propagates through the hydrologic model. The hydrologic uncertainty which is quantified using HUP is then added on, and the final output from BEUP which is in the form of predictive distribution gives an estimation of the total uncertainty (Verkade et al., 2017). Previous works have proved that BEUP could improve the forecast performance and provides more adequate quantification of predictive uncertainty (Han and Coulibaly, 2019; Reggiani et al., 2009). However, it still remains unknown which factors could affect the performance of BEUP. For example, the effect of the type of NWP (deterministic vs. ensemble) used as inputs

is unknown. Similarly, the effect of the type of hydrologic model (semi/distributed vs. lumped) used with BEUP is not documented.

There are different types of weather prediction product that could be used to force the BEUP system. In Canada, with the launch of Canadian Surface Prediction Archive (CaSPAr) in 2017, long archives of weather predictions are becoming easily accessible, including deterministic weather predictions such as Regional Deterministic Prediction System (RDPS) and ensemble weather predictions such as Regional Ensemble Prediction System (REPS). Also, the hydrologic model used with BEUP could be set up either lumped or semi-distributed. The lumped model treats the entire basin as one homogenous unit, while the semi-distributed model divides the basin into smaller sub-basins and each sub-basin has a different parameter set (Lobligeois et al., 2014). The contribution of this work is to investigate the impact on the performance of BEUP brought by using different weather prediction inputs and using different types of hydrologic model. The ultimate goal being to identify the best combination (input and model type) for use with BEUP in probabilistic flood forecasting.

To achieve this, RDPS with an ensemble dressing and REPS were used as input data to force both lumped and semi-distributed hydrological models, generating ensembles of streamflow forecasts. These ensemble forecasts were then post-processed via BEUP, producing probabilistic forecasts with uncertainty estimation. The major objectives of this research are: (1) to compare the performance of BEUP driven by deterministic and ensemble weather predictions for different forecast horizons; (2) to evaluate the predictive performance of BEUP based on lumped and semi-distributed model across

different forecast horizons; (3) to identify the best combination of input-model-BEUP to adequately account for the total predictive uncertainty and enhance flood forecasts through the BEUP post-processor.

5.3 Study Watershed and Data

The study area selected to conduct the research is Humber River watershed; it is a flood-prone area which is managed by Toronto Region Conservation Authority (TRCA). Humber River watershed is located in the south-central of the Greater Toronto Area (the most populous metropolitan region in Canada), it flows from Niagara Escarpment and Oak Ridges Moraine and drains to Lake Ontario with approximately 911 km² total drainage area (Figure 5-1). The watershed is characterized as a semi-urban watershed which covers a mixture of agricultural, urban and rapidly urbanizing land uses (TRCA, 2008). Please refer to Han et al. (2019) for a more detailed watershed description.

Hourly gauge precipitation and temperature observations were provided by TRCA, and hourly gauge discharge observations were provided by Water Survey of Canada (WSC). Kriging method was used to infill the missing precipitation and temperature data (Bostan et al., 2012; Meng et al., 2013). After data processing, the available time period for all the observation datasets is from January 2008 to December 2018. As shown in Figure 5-1, the eleven precipitation stations and the eleven temperature stations used are spatially distributed over the watershed, and the five flow stations selected are either located on the main river or the main tributary (Awol et al., 2018).

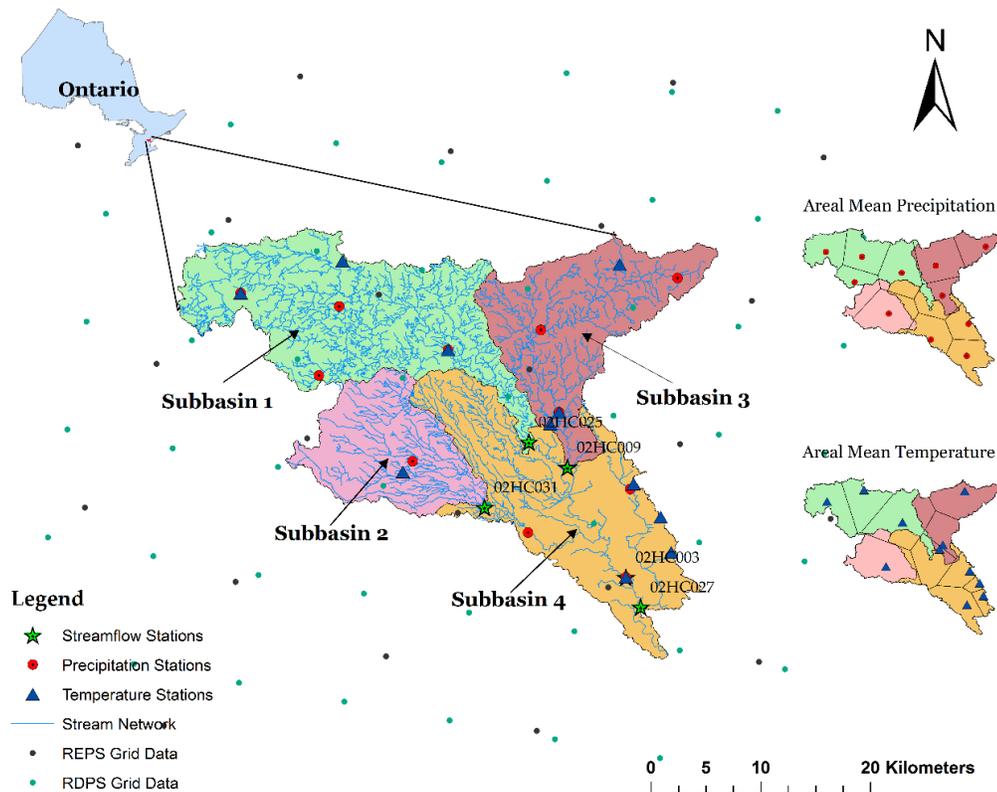


Figure 5-1 Study watershed and model setup

The precipitation and temperature predictions were obtained from Canadian Surface Prediction Archive (CaSPAR: <https://caspar-data.ca/>). Two different prediction products were collected: deterministic weather predictions from Regional Deterministic Prediction System (RDPS) and ensemble weather predictions from Regional Ensemble Prediction System (REPS). They are both gridded hourly forecasts. RDPS issued 4 times per day at UTC [00, 06, 12, 18] with each forecast up to 84 hours, the available archive is from January 2015 until now, and the spatial resolution is 10 km. REPS issued 2 times per day

at UTC [00, 12], and each forecast is up to 72 hours. REPS data are ensemble predictions which include 1 control member and 20 perturbed members. The available archive is from May 2017 until today with 15 km spatial resolution.

5.4 Methodology

5.4.1 Methodology Overview

The flowchart of the methodology is shown in Figure 5-2. The regional deterministic weather predictions (herein RDPS) are first dressed by a spread of meteorological uncertainty, then the dressed RDPS and regional ensemble weather predictions (herein REPS) are used as inputs to run the lumped and semi-distributed hydrologic models. Following this, the produced ensemble discharge forecasts are post-processed by BEUP processor. Prior to the forecast, the hydrologic model and the Hydrologic Uncertainty Processor (HUP) in BEUP have been calibrated in advance. The inherent meteorological uncertainty is presented by the ensemble spread, through the model propagation, it integrates with the hydrologic uncertainty which is quantified by HUP. Finally, a representative full distribution is obtained to estimate the predictive total uncertainty. Based on the different weather prediction inputs and the different spatial structures in the hydrologic model, four forecast approaches are developed: (1) lumped model with BEUP forced by dressed RDPS (Dressed RDPS + Lump), (2) semi-distributed model with BEUP forced by dressed RDPS (Dressed RDPS + Semi-dist), (3) lumped model with BEUP forced by REPS (REPS + Lump), and (4) semi-distributed model with BEUP forced by REPS (REPS + Semi-dist). Details about the model setup and uncertainty assessment are described in the following section.

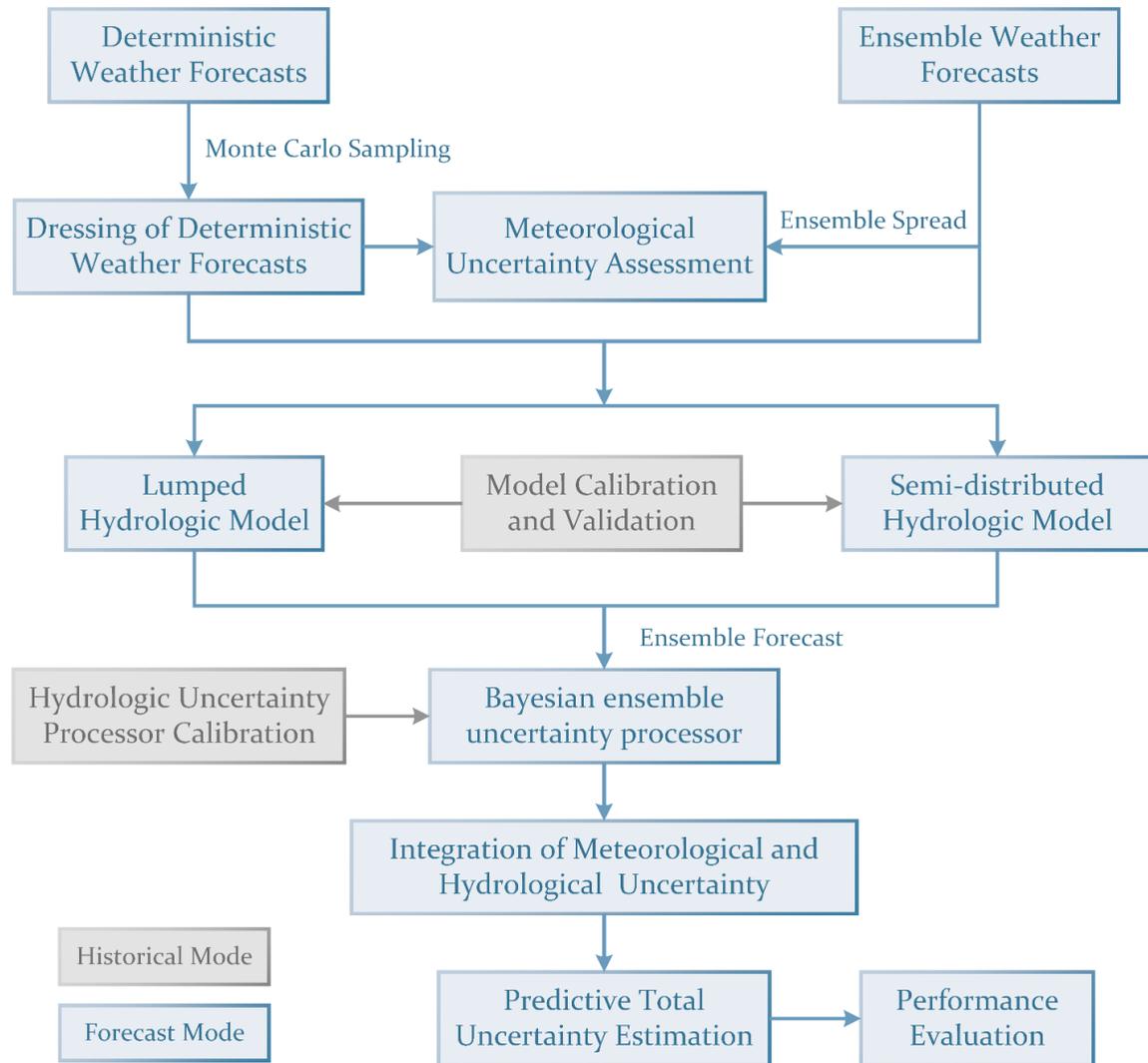


Figure 5-2 Flowchart of methodology

5.4.2 Hydrologic Model and Model Setup

Sacramento soil moisture accounting (SAC-SMA) model (Burnash et al., 1973) was applied to simulate the runoff. It is a conceptual hydrologic model used by National Weather Service River Forecast System (NWSRFS) for river streamflow forecasting across the United States. Due to the increasing operational requirements, Community

Hydrologic Prediction System (CHPS), an enhanced open forecasting system based on SAC-SMA is now replacing NWSRFS (Roe et al., 2010). Within CHPS, snowpack input for SAC-SMA is from SNOW-17 snow accumulation and ablation model (Anderson, 2006). Thus, in our version of SAC-SMA, the original snow routine was replaced by the SNOW-17 (Bennett et al., 2018). The inputs to the SAC-SMA include precipitation, temperature and potential evapotranspiration (PET), where PET was estimated using an adjusted PET calculation method based on temperature (Oudin et al., 2005), and the model output is simulated streamflow.

We refer to Razavi and Coulibaly (2017) for the structure of the SAC-SMA model used herein. In the model, the basin is divided into upper zone and lower zone, and water is stored in the form of tension water and free water. Tension water is firmly bound to the soil particles and only can be removed by evaporation or evapotranspiration, while free water is not bound to the soil particles and can be moved by gravitational forces. There are six state variable reservoirs in this model: additional impervious area content (ADIMC), upper zone tension water storage content (UZTWC), upper zone free water storage content (UZFWC), lower zone tension water storage content (LZTWC), lower zone free primary water storage content (LZFPC), and lower zone free secondary water storage content (LZFSC) (Razavi and Coulibaly, 2017). When precipitation, including rainfall and snowmelt, occurs over the basin, the portion that falls on the impermeable area becomes direct runoff, and the other portion that falls on the permeable area enters UZTWC. Once the UZTWC is filled, the excess water which beyond the capacity is accumulated in the UZFWC, and the amount of water in UZFWC is available to produce

interflow and percolate to the lower zone. When the UZFWC is totally filled, the excess precipitation over the capacity will generate surface runoff. The percolated water goes to LZTWC first and is then divided between LZFPC and LZFSC after LZTWC meets its maximum capacity. LZFPC supplies primary base flow and LZFSC supplies supplemental base flow. Therefore, the total runoff comes from five sources: (1) direct runoff from ADIMC, (2) interflow from UZFWC, (3) surface runoff from UZFWC, (4) primary base flow from LZFPC, and (5) supplemental base flow from LZFSC.

Our version of SAC-SMA has 26 parameters, 16 for SAC-SMA and 10 for SNOW-17, descriptions of these parameters and their ranges are presented in Table 5-1. The model was calibrated in hourly time step against discharge measurements, with 2008-2013 used as calibration period and 2014-2017 used as validation period (the first year was treated as warm-up). Particle swarm optimization (PSO) was employed as optimization algorithm (Razavi and Coulibaly, 2017), and modified Nash-volume error (MNVE) which is expressed in Eq. (5-1) was used as objective function (Samuel et al., 2012). Where NSE_{sqr} is NSE calculated using the logarithm of discharge, it reflects the accuracy for simulating high flows, and VE is volume error.

$$MNVE = NSE_{sqr} - 0.1VE \quad (5-1)$$

In this study, the SAC-SMA model was set up in two different ways: lumped SAC-SMA and semi-distributed SAC-SMA. As indicated in Figure 5-1, following the stream network, the basin was divided into four sub-basins in the semi-distributed SAC-SMA model: sub-basin 1 (SB1), sub-basin 2 (SB2), sub-basin 3 (SB3) and sub-basin 4 (SB4).

Station 02HC0025, 02HC0031, and 02HC009 were used as the pour point of SB1, SB2 and SB3 respectively, station 02HC003 and 02HC027 were summed up to calculate the basin outlet flow. Mean precipitation and temperature were calculated for each sub-basin using Thiessen Polygon method (Figure 5-1) and were used as inputs to simulate runoff at the pour point for each sub-basin. Muskingum routing method (Mohammad, 2014; Oyekanmi and Oladepo, 2017) was applied to route the flow from upstream to downstream, thus the total basin outflow was the summation of routed flow from SB1, SB2, SB3 and simulated flow from SB4. For the lumped SAC-SMA, the same precipitation and temperature information were used, but the areal mean was calculated for the entire basin. Accordingly, a single runoff output was simulated at the basin outlet.

5.4.3 Total Uncertainty Assessment

The total predictive uncertainty breaks down into two major sources: meteorological uncertainty and hydrologic uncertainty. For REPS, the ensemble members are obtained by perturbing through initial state, boundary condition and physical tendency of weather prediction model (Environment and Climate Change Canada, n.d.), thus the meteorological uncertainty is characterized by the spread of the ensembles. While for RDPS, which are single deterministic predictions of weather variables, the meteorological uncertainty was dressed through a sampling of the meteorological ensembles. Based on the analysis of the historical residuals between RDPS and observations (from 2015 to 2017), it was found that there is approximately 40% error for precipitation forecasts and on average 2 degrees of error for temperature forecasts. We assume that the statistical characteristics for the historical period will not change in the

very near future. Therefore, to represent meteorological uncertainty for the forecast period (the year 2018), other ensemble members were derived equiprobably via Monte Carlo sampling from 40% error for precipitation and ± 2 degrees for temperature. These ensembles were combined with the raw RDPS to construct the dressed RDPS dataset which has the same ensemble size as REPS. The meteorological uncertainty represented by the ensembles of REPS and dressed RDPS propagates from precipitation and temperature inputs to simulated discharge.

The hydrologic uncertainty was quantified by a precipitation-dependent Hydrologic Uncertainty Processor (HUP) under the assumption that the model inputs are perfect (Krzysztofowicz and Herr, 2001). Following Bayes' theorem, HUP revises prior distribution into posterior distribution by incorporating likelihood function. Here, "prior" means a prior belief based on the available information prior to the forecast, likelihood function carries new information from the model output and brings in uncertainty, and "posterior" means the updated belief after integrating all the new information through the Bayesian revision process. As a result, the hydrologic uncertainty is quantified by the updated posterior distribution. The procedure for HUP estimation is summarized in the following, and we refer to Krzysztofowicz (2002), Krzysztofowicz and Herr (2001) and Han et al. (2019) for a detailed description of the HUP including the mathematical background, Bayesian formulation, and all related equations.

In HUP, the forecast lead time is expressed using n with $n = 1, \dots, N$, and the indicator of precipitation occurrence is denoted by v ($v = 1$: precipitation occurs; $v = 0$: no precipitation). Suppose the forecast time is t_0 , the observed discharge at time t_n is denoted

by H_n and the modeled discharge at t_n is S_n , and realizations of H_n and S_n are expressed by h_n and s_n , respectively. Before actual forecasting, HUP was calibrated beforehand according to the following steps.

Step 1: Through matching between simulated discharge forecasts and corresponding discharge observations for different lead times, obtain joint sample $(h_0, h_1, h_2, \dots, h_N; s_1, s_2, \dots, s_N)$ for both branch $v = 0$ and $v = 1$.

Step 2: For both v and every n , estimate marginal prior distribution Γ_{nv} for H_n (corresponding density is γ_{nv}) and estimate marginal initial distribution $\bar{\Lambda}_{nv}$ for S_n .

Step 3: Following the estimated marginal distribution functions, transform each variate h_n and s_n into normally distributed w_n and x_n via standard normal inverse Q^{-1} .

Step 4: In the normal space, estimate the parameters of the transition densities from the linear regression between w_n and w_{n-1} , based on these dependence parameters, calculate the parameters that defining the prior distributions.

Step 5: In the normal space, based on the linear regression between x_n , w_n and w_0 and residual analysis between the expected value and true value of x_n , estimate the parameters of the likelihood functions.

Step 6: Given the parameters of prior distributions and likelihood functions, calculate the parameters A_{nv} , B_{nv} , D_{nv} , and T_{nv} that defining the updated posterior distributions.

These calibrated HUP parameters and estimated marginal distributions are ready to use in the forecast period. Conditional on forecasted discharge and initial condition, the

posterior distribution Φ_{nv} of the actual river discharge which takes hydrologic uncertainty into account can be defined by the equation (Krzysztofowicz and Herr, 2001):

$$\begin{aligned} & \Phi_{nv}(h_n|s_n, h_0) \\ &= Q \left(\frac{Q^{-1}(\Gamma_{nv}(h_n)) - A_{nv}Q^{-1}(\bar{\Lambda}_{nv}(s_n)) - D_{nv}Q^{-1}(\Gamma_{0v}(h_0)) - B_{nv}}{T_{nv}} \right) \end{aligned} \quad (5-2)$$

The corresponding posterior density ϕ_{nv} can be derived by (Krzysztofowicz and Herr, 2001):

$$\phi_{nv}(h_n|s_n, h_0) = \frac{\gamma_{nv}(h_n)q \left(Q^{-1}(\Phi_{nv}(h_n|s_n, h_0)) \right)}{T_{nv}q(Q^{-1}\Gamma_{nv}(h_n))} \quad (5-3)$$

When it comes to ensemble forecasts which are forced by dressed RDPS and REPS, the application of HUP to these ensembles constitutes BEUP. Where HUP is used to post-process each ensemble, producing ensemble posterior distributions, and then these ensemble posterior distributions are averaged into one representative predictive distribution. Details about the BEUP can be found in Han and Coulibaly (2019) and Reggiani et al. (2009). After post-processing of BEUP, the meteorological uncertainty and hydrologic uncertainty are lumped together to provide an estimation of the total uncertainty.

5.5 Results and Analysis

5.5.1 Comparison between Weather Predictions and Observations

The quality of raw RDPS and REPS, including precipitation and temperature predictions, were assessed by comparing with corresponding observations across various lead times.

Figure 5-3 shows the correlation coefficient between RDPS and observation for the entire basin and each sub-basin, and Figure 5-4 shows the correlation coefficient between REPS mean and observation for the entire basin and each sub-basin. The correlation coefficient (COR) measures the strength of the linear relationship between two variables. For both precipitation and temperature, the COR presents a decreasing trend as lead time grows. The repeating pattern displayed for temperature is due to the forecast cycle, RDPS predictions are performed every 6 hours and REPS predictions are performed every 12 hours. COR for RDPS precipitation ranges between 0.1 and 0.7, COR for REPS precipitation ranges between 0.2 and 0.7, and COR for both RDPS and REPS temperature are above 0.97. This indicates that the quality of temperature predictions is very good, while the quality for precipitation predictions is much more variable and declines quickly with forecast lead time. Comparison between RDPS and REPS suggests that REPS have slightly better COR than RDPS for higher lead times.

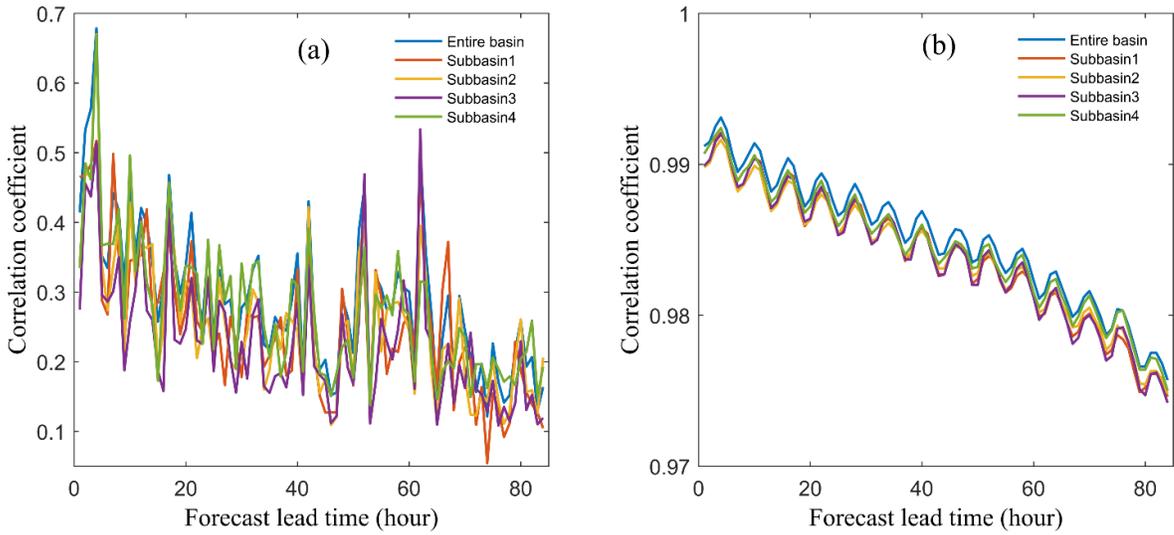


Figure 5-3 Comparison between RDPS and observation for (a) precipitation and (b) temperature

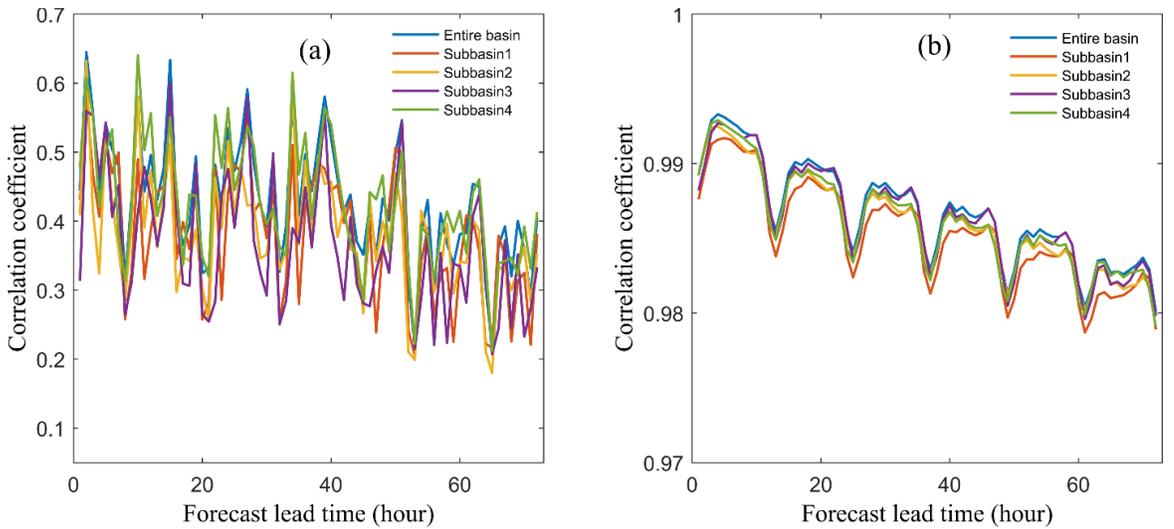


Figure 5-4 Comparison between REPS mean and observation for (a) precipitation and (b) temperature

5.5.2 SAC-SMA Calibration

The optimized parameters for the lumped and semi-distributed SAC-SMA are presented in Table 5-1, lumped SAC-SMA yields one parameter set for the entire basin, while semi-distributed SAC-SMA offers different parameter sets for each sub-basin. The hydrographs of observed and simulated discharges for the basin outlet are shown in Figure 5-5, with calibration period presented in the upper half and validation period in the lower half. Observed discharges are expressed by the black dot, simulated discharges from the lumped SAC-SMA are expressed by the red line, and simulated discharges from the semi-distributed SAC-SMA are expressed by the blue dashed line. Results from the hydrographs reveal that the semi-distributed model performs better than the lumped model in matching the high flows.

Table 5-1 Summary of SAC-SMA and SNOW-17 parameters

Parameter	Description	Unit	Range	Lumped	Semi-distributed			
					SB1	SB2	SB3	SB4
<i>SAC-SMA</i>								
UZTWM	Upper zone tension water capacity	mm	1–150	35.46	59.35	106.94	23.37	103.16
UZFWM	Upper zone free water capacity	mm	1–150	3.07	58.18	74.44	120.27	32.85
UZK	Upper zone free water drainage rate	d ⁻¹	0.1–0.5	0.24	0.48	0.49	0.49	0.11
PCTIM	Fraction of the minimum impervious area	–	0–0.1	0.10	0.09	0.09	0.01	0.09
ADIMP	Fraction of the additional impervious area	–	0–0.4	0.30	0.14	0.32	0.16	0.03
ZPERC	Maximum percolation rate	–	1–250	153.56	174.19	108.29	131.95	81.87
REXP	Exponent for the percolation demand equation	–	1–5	3.90	3.39	1.76	3.13	2.55
LZTWM	Lower zone tension water capacity	mm	1–500	228.38	154.86	197.10	333.38	353.29
LZFSM	Lower zone supplemental free water capacity	mm	1–1000	466.79	717.82	729.54	394.44	271.58
LZFPM	Lower zone primary free water capacity	mm	1–1000	919.55	804.14	400.76	641.52	438.04
LZSK	Lower zone supplemental free water drainage rate	d ⁻¹	0.01–0.25	0.04	0.01	0.02	0.01	0.06
LZPK	Lower zone primary free water drainage rate	d ⁻¹	0.0001–0.025	0.0001	0.0007	0.0001	0.02	0.02
PFREE	Fraction of percolated water goes directly into lower zone free water storages	–	0–0.6	0.59	0.59	0.53	0.57	0.59

PEADJ	ET-demand adjustment factor	–	0.5–1.4	0.76	1.19	1.28	1.39	1.05
PXADJ	Precipitation adjustment factor	–	0.5–1.4	1.39	0.97	1.39	1.22	1.38
Rq	Residence time parameter of quick-flow	d	0–0.99	0.00	0.10	0.15	0.01	0.39
SNOW-17								
scf	Snow fall correction factor	–	0.4–1.6	1.29	0.79	0.67	1.01	0.44
uadj	The average wind function during rain-on-snow periods	mm/mb/°C	0.01–0.22	0.08	0.05	0.09	0.16	0.10
mbase	Base temperature for non-rain melt factor	°C	0–1	0.12	0.90	0.42	0.02	0.83
mfmax	Maximum melt factor	mm/6h/°C	0–2	1.04	1.61	1.18	1.79	0.82
mfmin	Minimum melt factor	mm/6h/°C	0–0.7	0.13	0.17	0.01	0.002	0.01
tipm	Antecedent snow temperature index parameter	–	0.01–0.99	0.29	0.10	0.34	0.65	0.05
nmf	Maximum negative melt factor	mm/6h/°C	0.01–0.5	0.34	0.48	0.50	0.17	0.46
plwhc	Percent of liquid water capacity	–	0–0.4	0.001	0.02	0.01	0.003	0.11
pxtemp1	Lower Limit Temperature dividing transition from snow	°C	-2–2	-0.39	-1.92	-1.99	0.04	1.60
pxtemp2	Upper Limit Temperature dividing rain from transition	°C	1–3	2.18	2.43	2.22	2.89	1.58

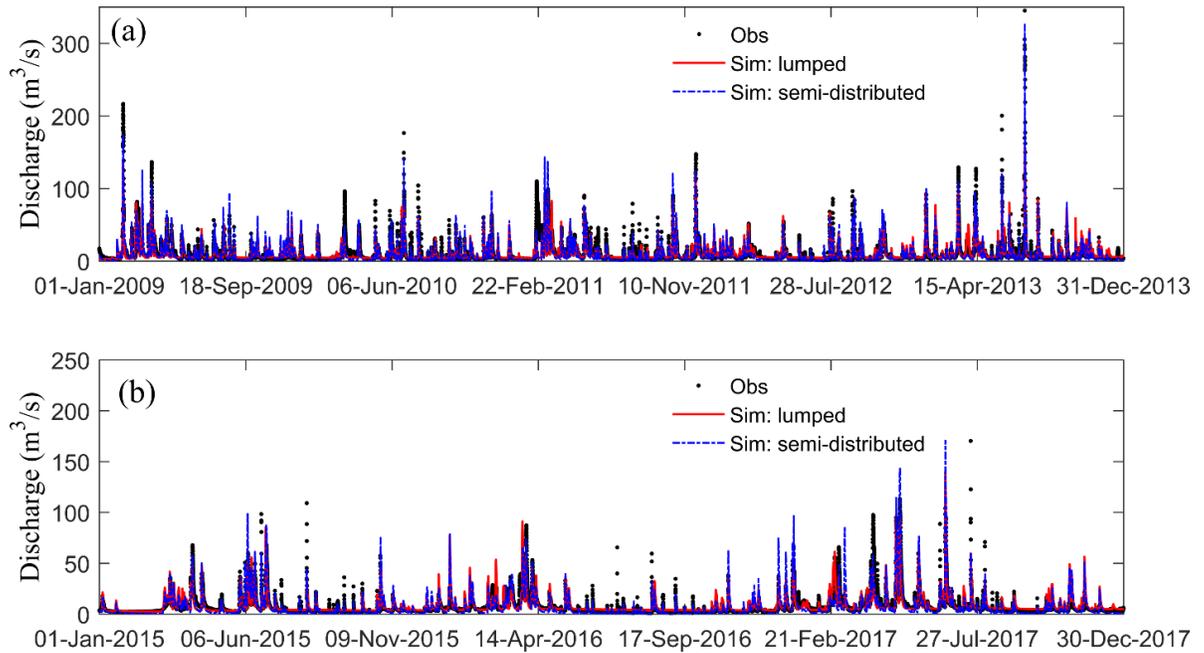


Figure 5-5 Observed vs. simulated hydrographs at basin outlet for lumped and semi-distributed SAC-SMA: (a) calibration; (b) validation

The performances of SAC-SMA models were evaluated using KGE (Kling-Gupta efficiency), NSE (Nash-Sutcliffe coefficient), previously mentioned NSE_{sqr} , VE (volume error) and PFC (peak flow criteria), and the results are summarized in Table 5-2. KGE measures the model efficiency by computing the Euclidian distance from its ideal point (Gupta et al., 2009), NSE assesses the model efficiency for both low flows and high flows by calculating residual variance over measured data variance (Nash and Sutcliffe, 1970), and NSE_{sqr} determines the model efficiency with more tendency to high flows. For KGE, NSE, and NSE_{sqr} , the perfect value is 1, the closer the value to 1, the more efficient the model is. VE measures the difference between the summation of simulations and the summation of observations relative to the latter, PFC provides a model performance

measure for peak flow only (Coulibaly et al., 2000). For VE and PFC, a value of 0 indicates a perfect fit.

Table 5-2 Performances of SAC-SMA models

	All flows					Peak flows				
	Lumped*	Semi distributed				Lumped*	Semi distributed			
		SB1	SB2	SB3	SB4*		SB1	SB2	SB3	SB4*
Calibration										
KGE	0.68	0.71	0.65	0.71	0.80	0.61	0.69	0.69	0.65	0.73
NSE	0.60	0.57	0.55	0.39	0.62	0.47	0.40	0.50	0.33	0.49
NSE _{sqr}	0.42	0.59	0.50	0.44	0.76	0.40	0.55	0.49	0.44	0.76
VE	0.12	0.21	0.27	0.05	0.03	0.05	0.13	0.15	0.05	0.12
PFC	--	--	--	--	--	0.15	0.10	0.17	0.16	0.11
Validation										
KGE	0.69	0.59	0.45	0.65	0.73	0.68	0.61	0.60	0.59	0.68
NSE	0.53	0.49	0.51	0.26	0.52	0.39	0.31	0.48	0.16	0.35
NSE _{sqr}	0.41	0.70	0.61	0.38	0.38	0.35	0.66	0.63	0.39	0.31
VE	0.21	0.36	0.46	0.03	0.04	0.03	0.29	0.31	0.11	0.07
PFC	--	--	--	--	--	0.14	0.10	0.11	0.14	0.13

Note: basin outlet marked with asterisk*

The performances for all flows are presented on the left in Table 5-2. For the basin outlet calibration period, the semi-distributed model has higher KGE, NSE, NSE_{sqr} and lower VE compared to the lumped model. For the validation period, the semi-distributed model has higher KGE, similar NSE, slightly lower NSE_{sqr}, and better VE. The performances for peak flows which only consider the flows over 75% quantile are presented on the

right. Compared to the lumped model, the semi-distributed model presents better KGE, NSE, NSE_{sqr} , and PFC statistics for peak flows during calibration, and same KGE and better PFC during validation. For the sub-basins in the semi-distributed model, the NSE values seem relatively low, but the KGE, NSE_{sqr} , and PFC values are good. These results suggest that, in terms of simulating both low flows and high flows as indicated by NSE, lumped and semi-distributed models show comparable performance. However, in terms of simulating high flows as reflected by PFC and NSE_{sqr} , the performance of the semi-distributed model is more satisfactory.

5.5.3 Calibration of Hydrologic Uncertainty Processor

The calibrated lumped and semi-distributed SAC-SMA are passed into HUP to analyze the hydrologic uncertainty, time period of 2008 to 2017 was used for estimation of HUP, and lead time was tested up to 84 hours to make it consistent with the maximum forecast lead time of RDPS and REPS. Recall that the predictive distribution produced from HUP is defined by Eq. (5-2) depending on marginal distributions for H_n and S_n and parameters A_{nv} , B_{nv} , D_{nv} and T_{nv} . Based on the modified Shapiro-Wilk (MSW) test (Ashkar and Aucoin, 2012), kernel distribution was determined as the best distribution type as it shows the best fit for both H_n and S_n . The estimated parameters A_{nv} , B_{nv} , D_{nv} and T_{nv} for both branches across different lead times are presented in Figure 5-6. For both lumped and semi-distributed model, B_{nv} values are near 0. T_{nv} shows an increasing trend, T_{n1} values are higher than T_{n0} for small lead times and the two lines converge for higher lead times. This indicates that the variance of forecast distribution increases with the increase of lead time and the forecast is more uncertain when precipitation occurs. A_{nv} is

increasing while D_{nv} is decreasing as lead time grows, revealing that as lead time increases the predictive distribution is more dominated by the new information which is from model forecast and less dominated by the available information prior to the forecast. A_{n1} values are greater than A_{n0} while D_{n1} values are lower than D_{n0} , this means that under the condition of precipitation occurrence the forecast is less affected by the prior belief.

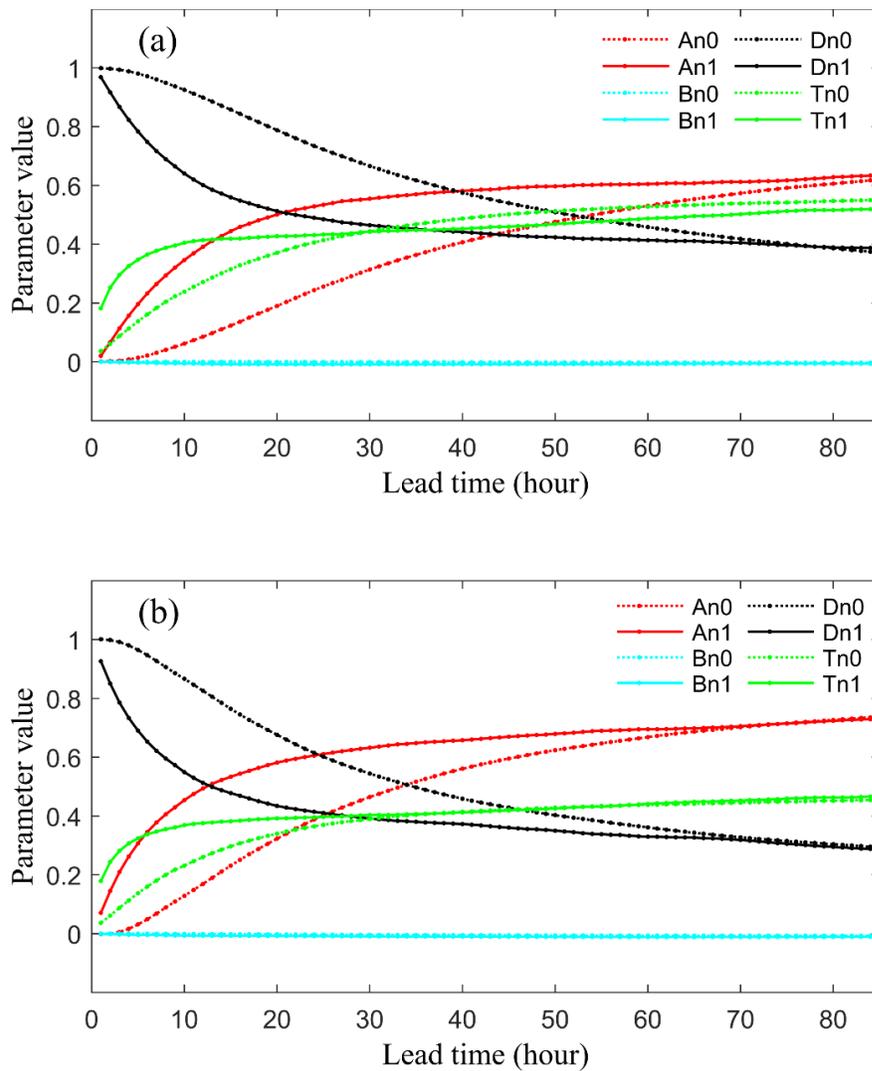


Figure 5-6 Estimated HUP parameters for (a) lumped SAC-SMA and (b) semi-distributed SAC-SMA

5.5.4 Scenario Results Comparisons

During the forecast period, dressed RDPS and REPS were used as forcing data for the calibrated lumped and semi-distributed SAC-SMA, and the ensemble forecasts generated from the model were post-processed by the calibrated BEUP, resulting in four different probabilistic forecast scenarios which are all considering the total predictive uncertainty. The performances of the four scenarios were compared using various evaluation metrics, including the accuracy of single-valued forecasts and the probabilistic characteristics of the forecast distributions.

Predictive mean from the probabilistic forecasts was used as the single-valued forecast, and its accuracy was assessed by NSE and KGE. The comparisons of the deterministic metrics for different scenarios as a function of forecast lead time are presented in Figure 5-7. Where Dressed RDPS+Lump is represented by the magenta dashed line, Dressed RDPS+Semi-dist is represented by the magenta solid line, REPS+Lump is represented by the blue dashed line, and REPS+Semi-dist is represented by the blue solid line. Scenarios with the same input data use the same line color, and scenarios with the same model type use the same line style. For all the scenarios, NSE and KGE statistics are getting worse as forecast lead time increases. Dressed RDPS+Semi-dist and REPS+Semi-dist show similar NSE and KGE within lead time of 24 hours, but they differ for longer lead times with Dressed RDPS+Semi-dist showing higher value. In general, NSE and KGE follow similar pattern with $\text{Dressed RDPS+Semi-dist} > \text{REPS+Semi-dist} > \text{Dressed RDPS+Lump} > \text{REPS+Lump}$. NSE and KGE values of the best two scenarios are above 0.5 under the lead time of 24 hours, indicating that the 1 day forecasts using these two

approaches are acceptable in terms of accuracy. Overall, in terms of model efficiency and accuracy, using the semi-distributed model with BEUP significantly outperforms using lumped model with BEUP. BEUP with dressed RDPS as input outperforms BEUP with REPS as input. The best two scenarios are Dressed RDPS+Semi-dist and REPS+Semi-dist with the former being the best combination for use with BEUP.

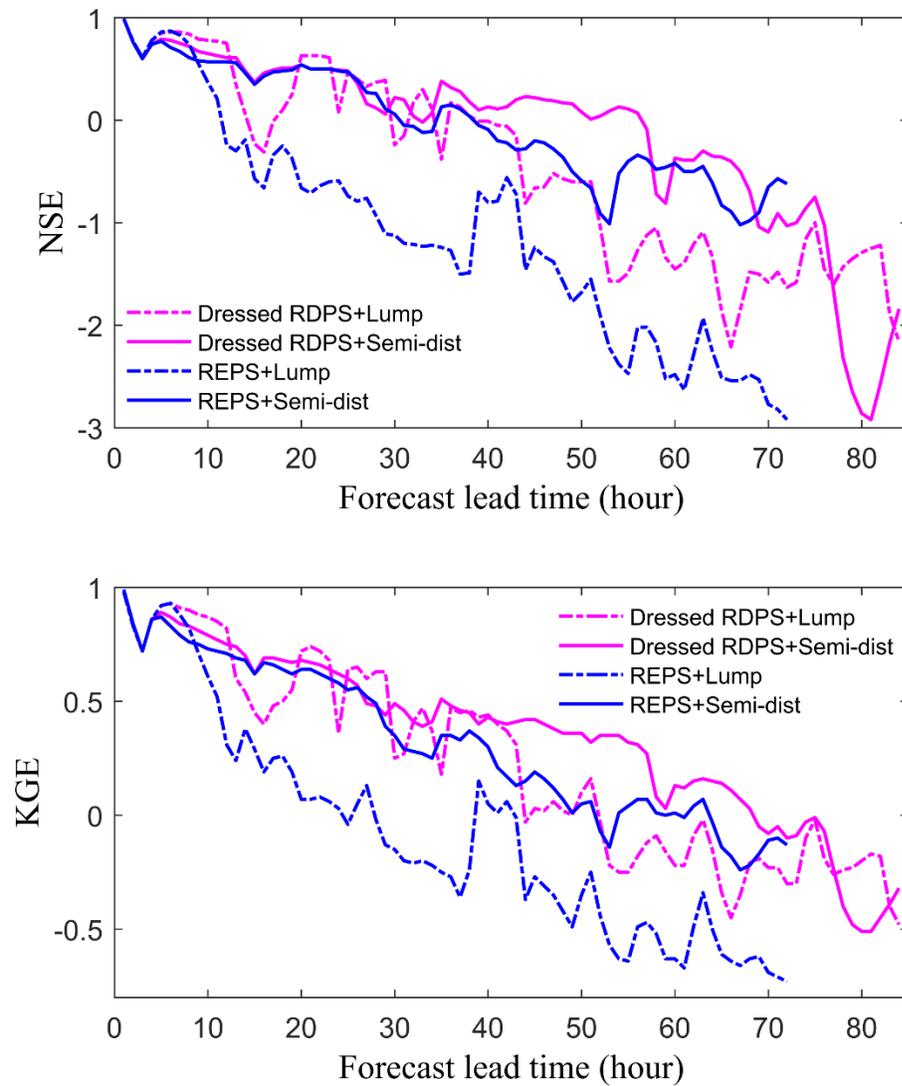


Figure 5-7 Nash-Sutcliffe coefficient (NSE) and Kling-Gupta efficiency (KGE) for different scenarios as a function of forecast lead time

The probabilistic characteristics of the forecast distributions were assessed using continuous ranked probability score (CRPS) and its decomposition. CRPS is a measure of how well the forecast distribution match the observation; if the forecast is perfect, the CRPS will be 0. To provide a detailed picture of the forecast distribution, mean CRPS can be decomposed into three parts as expressed in Eq. (5-4): reliability, resolution, and uncertainty (Hersbach, 2000).

$$\overline{CRPS} = reliability - resolution + uncertainty \quad (5-4)$$

The reliability tests whether the frequency of observations falls within each probability bin is nearly equal; it should be 0 with a perfectly reliable forecast. The resolution measures how much improvement gained from the forecast system compared to the forecast based on climatology, the higher the resolution, the more improvement it could get. The uncertainty is equal to the best achievable CRPS value when only climatological information is available. As such, CRPS and its reliability component are most frequently used.

The comparison of CRPS and its reliability component for different scenarios are presented in Figure 5-8, the same scenario is expressed by the same line style and color as in Figure 5-7. Results for all flows are shown on the left and results for peak flows (flows over 75% quantile) are shown on the right. The CRPS and reliability for all flows and peak flows demonstrate similar pattern, all worsen with increasing lead time. When considering all flows, the CRPS values do not show significant difference among the scenarios, they gradually grow from near 0.3 to around 6. The reliability values of

Dressed RDPS+Lump, Dressed RDPS+Semi-dist and REPS+Semi-dist are similar, REPS+Lump appears less reliable than the other three scenarios. CRPS and reliability are low for short lead times, indicating a relatively skillful and reliable short-term forecast.

For peak flows, the differences of CRPS and reliability between the scenarios are more obvious. The CRPS worsens with higher flow as higher CRPS is shown when considering peak flows, it gradually increases from near 0.3 to around 9. Dressed RDPS+Lump, Dressed RDPS+Semi-dist and REPS+Semi-dist show comparable CRPS, while CRPS for REPS+Lump is a little bit worse for higher lead times. For peak flow reliability, the inferiority of REPS+Lump is more evident, and Dressed RDPS+Semi-dist is more reliable than REPS+Semi-dist and Dressed RDPS+Lump beyond lead time of 40 hours. In short, Dressed RDPS+Lump, Dressed RDPS+Semi-dist and REPS+Semi-dist indicate comparable forecast skill as reflected by the CRPS, with REPS+Lump showing a slightly worse performance. The forecast reliability can be sorted as: Dressed RDPS+Semi-dist > REPS+Semi-dist \approx Dressed RDPS+Lump > REPS+Lump.

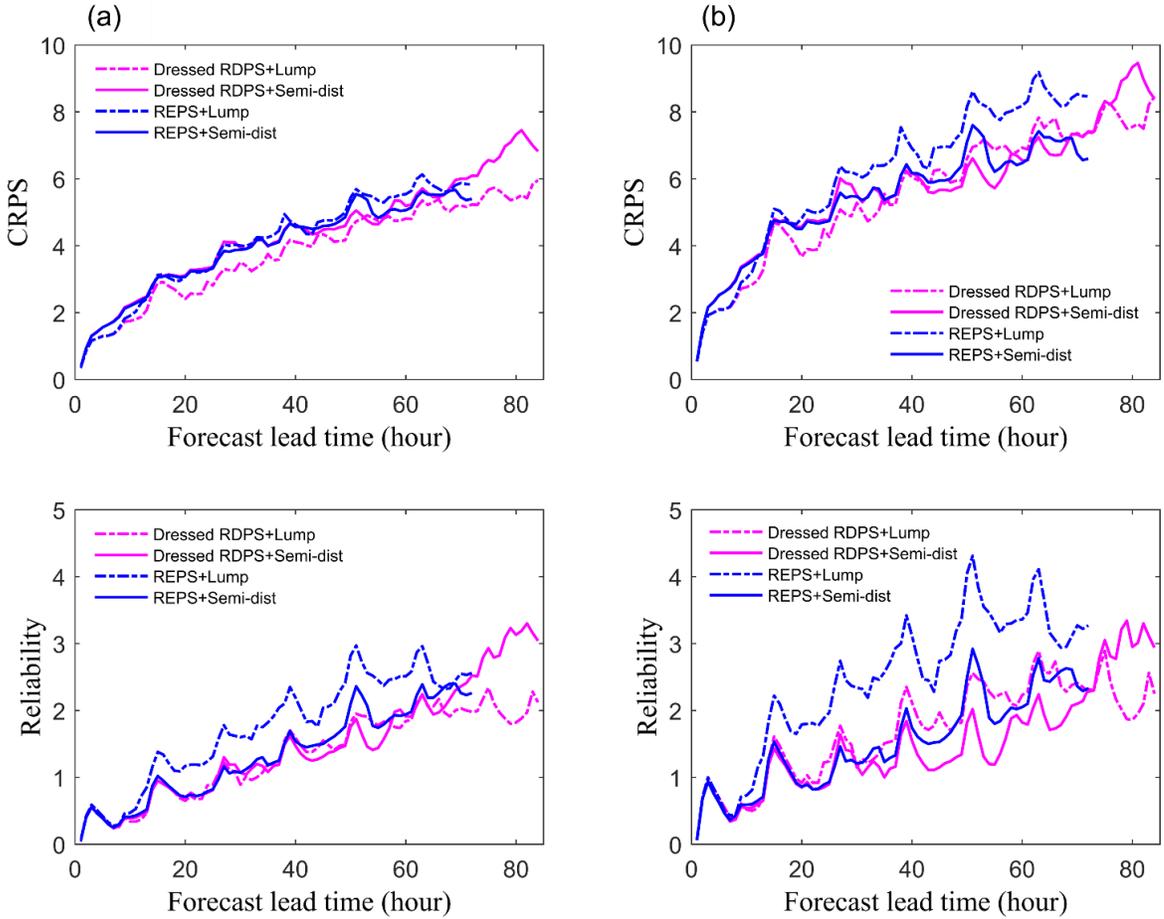


Figure 5-8 Continuous ranked probability score (CRPS) and its reliability component for different scenarios for various forecast lead times: (a) all flows; (b) peak flows

In addition, hit rate was used to provide a direct measure of the forecast quality, here it calculates how much percent of peak flows is captured by the defined uncertainty bounds from the probabilistic forecasts (Ramos et al., 2007). Figure 5-9 presents the hit rate for both 50% and 90% uncertainty bound, 50% uncertainty bound is defined by 25-75% quantile of the forecast distribution, and 90% uncertainty bound is defined by 5-95% quantile. Hit rates for 50% uncertainty bound are all above 30%, and are higher than 40%

within a lead time of 36 hours. Hit rates for 90% bound are all beyond 70% and most of the cases are greater than 80%, expressing the ability in capturing peak flows is satisfactory. With respect to the comparison between scenarios, none of the scenarios performs consistently better than another when forecast lead time increases. However, REPS+Semi-dist appears the best performing in term of hit rate for longer lead times.

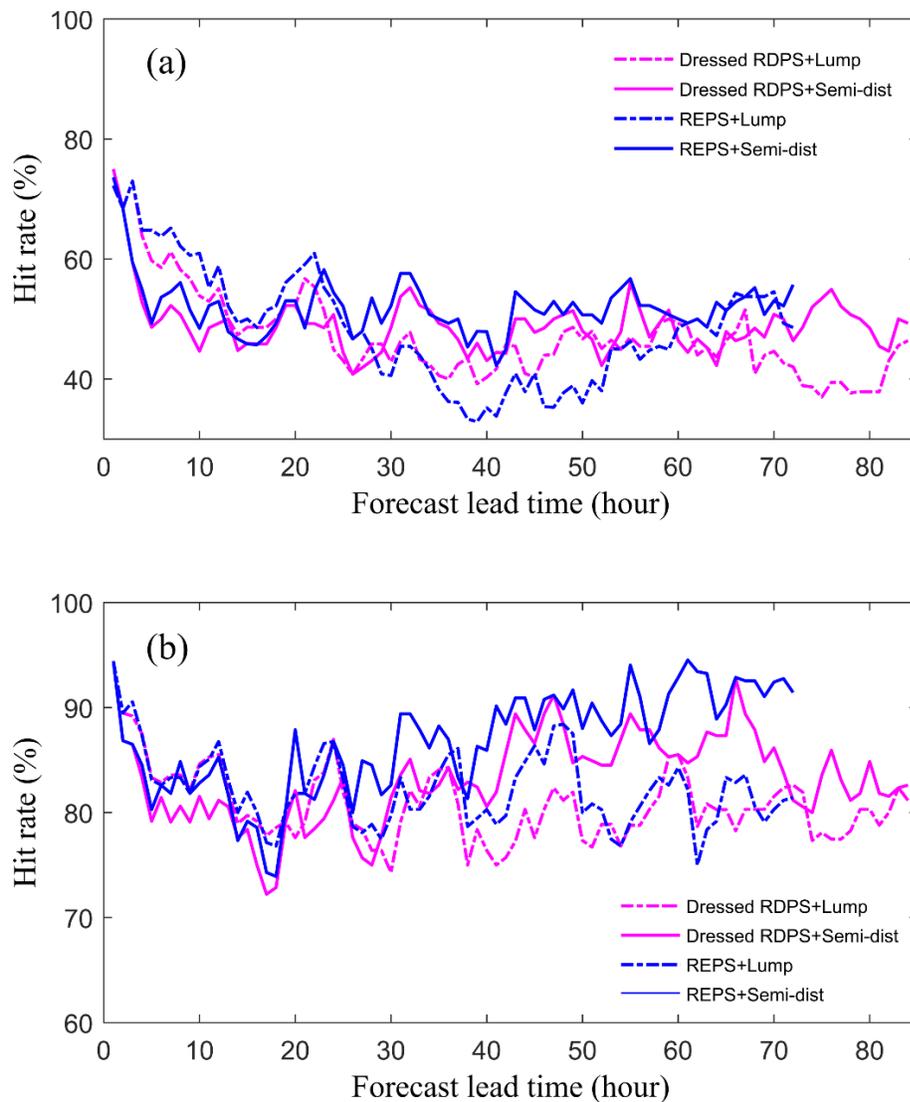


Figure 5-9 Percent of peak flows lie within (a) the 50% uncertainty bound (25-75% quantile) and (b) the 90% uncertainty bound (5-95% quantile)

5.5.5 Example Forecast Hydrographs

Two examples of the forecast hydrograph using the best two scenarios identified are demonstrated. Figure 5-10 is a forecast hydrograph for a high flow event issued on April 15, 2018 using Dressed RDPS+Semi-dist, and Figure 5-11 is a forecast hydrograph for the same event issued 12 hours later using REPS+Semi-dist. For each hydrograph, to demonstrate how to communicate the uncertainty, both the 50% (25-75% quantile) uncertainty bound and a narrower 30% (35-65% quantile) uncertainty bound derived from the predictive distribution are presented. Observed discharge is represented by the red line, predictive mean is expressed by the blue line, and predictive median is expressed by the black line. For both forecasts, peaks are well captured by the uncertainty bounds. Predictive mean tends to overpredict the peak flow as it lies above the observation, this is because the forecast distributions are right-skewed and have a tail on the right side of the distribution; thus the predictive mean is closer to the tail which gives a relatively larger flow. However, the forecast is acceptable considering the uncertainty bound performs well in capturing high flows and predictive median shows a good match with observation. Further improvement may be achieved by applying bias correction or post-processing to the weather predictions (Khajehei and Moradkhani, 2017; Khajehei, et al., 2018).

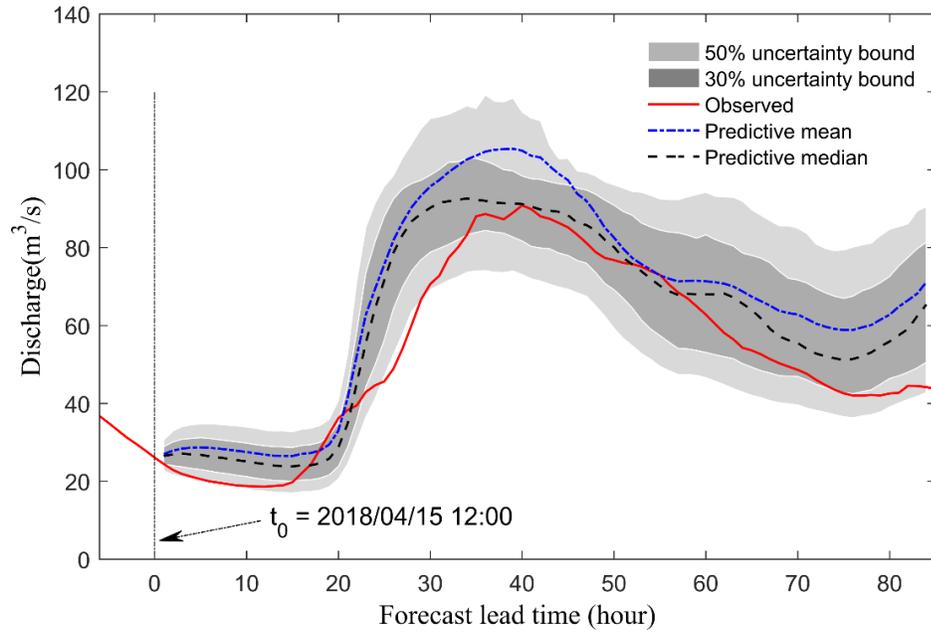


Figure 5-10 Forecast hydrograph for a high flow event issued on April 15, 2018 using Dressed RDPS+Semi-dist

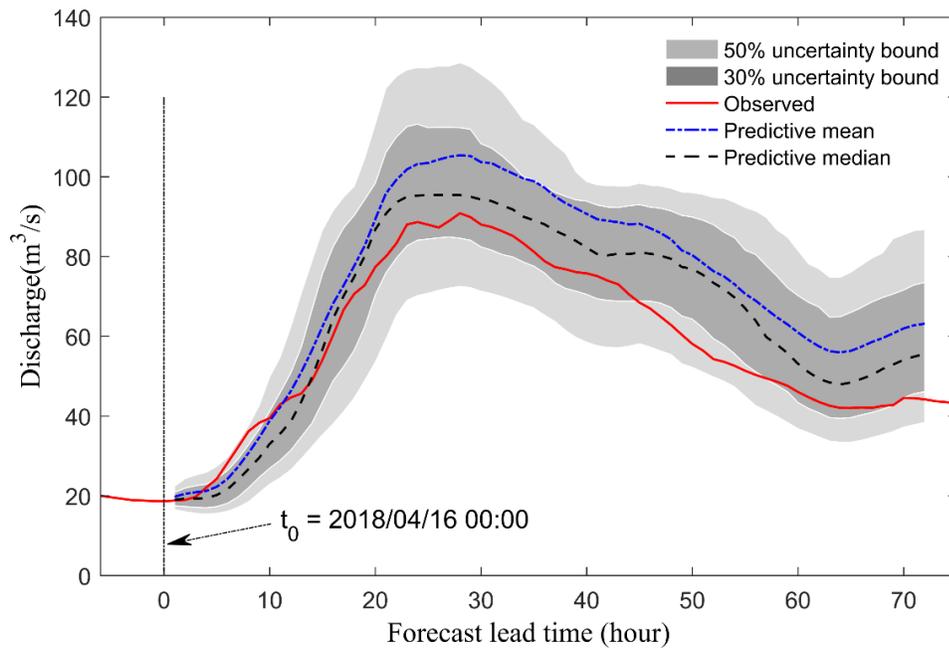


Figure 5-11 Forecast hydrograph for a high flow event issued on April 16, 2018 using REPS+Semi-dist

5.6 Conclusions and Discussion

This paper has presented an application of BEUP as a post-processor of ensemble forecasts to assess the total predictive uncertainty associated with flood forecast. In BEUP, each ensemble was post-processed using HUP, and the ensemble posterior distributions produced were lumped together into one representative predictive distribution. The total uncertainty represented by the predictive distribution is the integration of two major uncertainty sources: meteorological uncertainty which is estimated by the ensembles of weather prediction and hydrologic uncertainty which can be quantified through the post-processing of HUP. To identify the factors that affect the performance of BEUP in flood forecasting context, BEUP with different weather inputs and different hydrologic model types were tested. The different weather inputs include REPS and dressed RDPS which is the RDPS with an ensemble dressing, the different model types include a lumped SAC-SMA and a semi-distributed SAC-SMA which are both calibrated with emphasis on high flows. Therefore, the different combinations led to four forecast scenarios all with predictive uncertainty estimation using BEUP, and the performance of these different scenarios were assessed and compared using multiple evaluation metrics.

Results indicate that BEUP provides a robust post-processing method for total uncertainty estimation, the uncertainty bounds produced perform well in capturing peak flows as indicated by the high peak flow hit rate. In terms of accuracy, efficiency and reliability, BEUP with the semi-distributed model is more accurate and reliable than with the lumped model, and using dressed RDPS as forcing data outperforms using REPS as

forcing data. In terms of forecast skill, no significant difference is revealed among the four different scenarios, indicating they have similar forecast skill. In general, the best two scenarios are Dressed RDPS+Semi-dist and REPS+Semi-dist. Dressed RDPS+Semi-dist has better NSE, KGE and reliability, while REPS+Semi-dist has higher peak flow hit rate. For 1-day ahead forecasts, the two best scenarios are promising approaches as suggested by their high NSE and KGE values along with low CRPS and reliability score values. Comparisons among the scenarios reveal that the improvement brought by the hydrologic model type is more significant than the input data type. The semi-distributed model, which considers spatial variability within the watershed compared to the lumped model, is recommended for use with BEUP for flood forecasting.

There are some aspects that may further enhance the forecast performance and could be tested in future work. For example, there are usually systematic biases exhibited in the weather predictions, application of bias correction or post-processing to the weather prediction inputs before running the model might lead to an improved performance. Additionally, dividing the basin into more sub-basins in the semi-distributed model will further reflect the spatial variability, and thus may produce better forecast results. Also, the ensemble dressing of RDPS could be conducted in an alternative way, or could be changed into an alternative uncertainty dressing technique. However, these complementary work would not change the key conclusions drawn from this paper.

5.7 Acknowledgments

This research was supported by the Natural Science and Engineering Research Council (NSERC) Canadian FloodNet with grant number NETGP-451456 and the China

Scholarship Council (CSC). Special thanks to Dr. Daniela Biondi from University of Calabria for the technical support. The authors are grateful to the Toronto Region Conservation Authority (TRCA) for providing precipitation and temperature data, the Water Survey of Canada (WSC) for providing discharge data, and the Canadian Surface Prediction Archive (CaSPAr) from where we downloaded the weather prediction products.

5.8 References

- Anderson, E., 2006. Snow accumulation and ablation model–SNOW-17.
- Ashkar, F., Aucoin, F., 2012. Choice between competitive pairs of frequency models for use in hydrology: a review and some new results. *Hydrological Sciences Journal* 57(6), 1092–1106. doi:10.1080/02626667.2012.701746
- Awol, F.S., Coulibaly, P., Tolson, B.A., 2018. Event-based model calibration approaches for selecting representative distributed parameters in semi-urban watersheds. *Advances in Water Resources* 118, 12–27. doi:10.1016/j.advwatres.2018.05.013
- Bennett, K.E., Cherry, J.E., Balk, B., Lindsey, S., 2018. Using MODIS estimates of fractional snow cover extent to improve streamflow forecasts in Interior Alaska. *Hydrology and Earth System Sciences Discussions*. doi:10.5194/hess-2018-96
- Biondi, D., Todini, E., 2018. Comparing Hydrological Postprocessors Including Ensemble Predictions Into Full Predictive Probability Distribution of Streamflow. *Water Resources Research* 54(12), 9860–9882. doi:10.1029/2017WR022432
- Bostan, P.A., Heuvelink, G.B., Akyurek, S.Z., 2012. Comparison of regression and kriging techniques for mapping the average annual precipitation of Turkey. *International Journal of Applied Earth Observations and Geoinformation* 19, 115–126. doi:10.1016/j.jag.2012.04.010
- Burnash, R.J., Ferral, R.L., McGuire, R.A., 1973. A generalized streamflow simulation system : conceptual modeling for digital computers.
- Coccia, G., Todini, E., 2011. Recent developments in predictive uncertainty assessment based on the model conditional processor approach. *Hydrology and Earth System Sciences* 15(10), 3253–3274. doi:10.5194/hess-15-3253-2011
- Coulibaly, P., Anctil, F., Bobée, B., 2000. Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. *Journal of Hydrology* 230(3-4), 244–257. doi:10.1016/S0022-1694(00)00214-6

- Fragoso, T.M., Bertoli, W., Louzada, F., 2018. Bayesian Model Averaging: A Systematic Review and Conceptual Classification. *International Statistical Review* 86(1), 1–28. doi:10.1111/insr.12243
- Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology* 377(1-2), 80–91. doi:10.1016/j.jhydrol.2009.08.003
- Han, S., Coulibaly, P., Biondi, D., 2019. Assessing Hydrologic Uncertainty Processor performance for flood forecasting in a semiurban watershed. *Journal of Hydrologic Engineering*, in press.
- Han, S., Coulibaly, P., 2017. Bayesian flood forecasting methods: A review. *Journal of Hydrology* 551, 340–351. doi:10.1016/j.jhydrol.2017.06.004
- Han, S., Coulibaly, P., 2019. Probabilistic flood forecasting using Hydrologic Uncertainty Processor with ensemble weather forecasts. *Journal of Hydrometeorology*, in press.
- Hemri, S., Lisniak, D., Klein, B., 2015. Multivariate postprocessing techniques for probabilistic hydrological forecasting. *Water Resources Research* 51(9), 7436–7451. doi:10.1111/j.1752-1688.1969.tb04897.x
- Hersbach, H., 2000. Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. *Weather and Forecasting* 15(5), 559–570. doi:10.1175/1520-0434(2000)015<0559:dotcrp>2.0.co;2
- Jones, D.A., Kay, A.L., 2007. Uncertainty analysis for estimating flood frequencies for ungauged catchments using rainfall-runoff models. *Advances in Water Resources* 30(5), 1190–1204. doi:10.1016/j.advwatres.2006.10.009
- Khajehei, S., Ahmadalipour, A., Moradkhani, H., 2018. An effective post-processing of the North American multi-model ensemble (NMME) precipitation forecasts over the continental US. *Climate dynamics* 51(1-2), 457-472. doi:10.1007/s00382-017-3934-

0.

- Khajehi, S., Moradkhani, H., 2017. Towards an improved ensemble precipitation forecast: A probabilistic post-processing approach. *Journal of Hydrology* 546, 476-489. doi:10.1016/j.jhydrol.2017.01.026.
- Krzysztofowicz, R., 2002. Bayesian system for probabilistic river stage forecasting. *Journal of Hydrology* 268(1-4), 16–40. doi:10.1016/S0022-1694(02)00106-3
- Krzysztofowicz, R., Herr, H.D., 2001. Hydrologic uncertainty processor for probabilistic river stage forecasting: precipitation-dependent model. *Journal of Hydrology* 249(1-4), 46-48. doi:10.1016/S0022-1694(01)00412-7
- Li, W., Duan, Q., Miao, C., Ye, A., Gong, W., Di, Z., 2017. A review on statistical postprocessing methods for hydrometeorological ensemble forecasting. *Wiley Interdisciplinary Reviews: Water* 4(6), e1246. doi:10.1002/wat2.1246
- Linde, N., Ginsbourger, D., Irving, J., Nobile, F., Doucet, A., 2017. On uncertainty quantification in hydrogeology and hydrogeophysics. *Advances in Water Resources* 110, 166–181. doi:10.1016/j.advwatres.2017.10.014
- Lobligeois, F., Andréassian, V., Perrin, C., Tabary, P., Loumagne, C., 2014. When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation using 3620 flood events. *Hydrology and Earth System Sciences* 18(2), 575–594. doi:10.5194/hess-18-575-2014
- Masadgar, S., Moradkhani, H., 2014. Improved Bayesian multimodeling: Integration of copulas and Bayesian model averaging. *Water Resources Research* 50(12), 9586–9603. doi:10.1002/2015WR017273. Received
- Meng, Q., Liu, Z., Borders, B.E., 2013. Assessment of regression kriging for spatial interpolation – comparisons of seven GIS interpolation methods. *Cartography and Geographic Information Science* 40(1), 28–39. doi:10.1080/15230406.2013.762138
- Mohammad, Q.H., 2014. Calculating the Coefficients of Muskingum and Muskingum-Cunge Methods for a reach from Shatt- Al-Hilla river. *Journal of University of*

Babylon 22(4), 749–762.

Munoz, D.H., Constantinescu, G., 2018. A fully 3-D numerical model to predict flood wave propagation and assess efficiency of flood protection measures. *Advances in Water Resources* 122, 148–165. doi:10.1016/j.advwatres.2018.10.014

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I - A discussion of principles. *Journal of Hydrology* 10(3), 282–290.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., Loumagne, C., 2005. Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 2 - Towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling. *Journal of Hydrology* 303(1-4), 290–306. doi:10.1016/j.jhydrol.2004.08.026

Oyekanmi, M.O., Oladepo, K.T., 2017. Application of Basic Excel Programming to Linear Muskingum Model for Open Channel Routing. *Civil and Environmental Research* 9, 39–54.

Ramos, M.H., Bartholmes, J., Pozo, J.T., 2007. Development of decision support products based on ensemble forecasts in the European flood alert system. *Atmospheric Science Letters* 8(4), 113–119. doi:10.1002/asl

Razavi, T., Coulibaly, P., 2017. An evaluation of regionalization and watershed classification schemes for continuous daily streamflow prediction in ungauged watersheds. *Canadian Water Resources Journal* 42(1), 2–20. doi:10.1080/07011784.2016.1184590

Reggiani, P., Renner, M., Weerts, A.H., Van Gelder, P.A.H.J.M., 2009. Uncertainty assessment via Bayesian revision of ensemble streamflow predictions in the operational river Rhine forecasting system. *Water Resources Research* 45(2), 1–14. doi:10.1029/2007WR006758

Regional Ensemble Prediction System, URL:

<http://data.ec.gc.ca/data/weather/products/regional-ensemble-prediction-system/>

(accessed 4.26.19).

- Roe, J., Dietz, C., Restrepo, P., Halquist, J., Hartman, R., Horwood, R., Olsen, B., Opitz, H., Shedd, R., Welles., E., 2010. NOAA's community hydrologic prediction system. Proceedings from the 4th Federal Interagency Hydrologic Modeling Conference.
- Samuel, J., Coulibaly, P., Metcalfe, R.A., 2012. Identification of rainfall-runoff model for improved baseflow estimation in ungauged basins. *Hydrological Processes* 26(3), 356–366. doi:10.1002/hyp.8133
- Schaake, J.C., Hamill, T.M., Buizza, R., Clark, M., 2007. HEPEX: The hydrological ensemble prediction experiment. *Bulletin of the American Meteorological Society* 88(10), 1541–1548. doi:10.1175/BAMS-88-10-1541
- Schendel, T., Thongwichian, R., 2017. Considering historical flood events in flood frequency analysis: Is it worth the effort? *Advances in Water Resources* 105, 144–153. doi:10.1016/j.advwatres.2017.05.002
- Seo, D.J., Herr, H.D., Schaake, J.C., 2006. A statistical post-processor for accounting of hydrologic uncertainty in short-range ensemble streamflow prediction. *Hydrology and Earth System Sciences Discussions* 3(4), 1987–2035. doi:10.5194/hessd-3-1987-2006
- Todini, E., 2013. From HUP to MCP: Analogies and extended performances. *Journal of Hydrology* 477, 33–42. doi:10.1016/j.jhydrol.2012.10.037
- Verkade, J.S., Brown, J.D., Davids, F., Reggiani, P., Weerts, A.H., 2017. Estimating predictive hydrological uncertainty by dressing deterministic and ensemble forecasts; a comparison, with application to Meuse and Rhine. *Journal of Hydrology* 555, 257–277. doi:10.1016/j.jhydrol.2017.10.024
- Weerts, A.H., Winsemius, H.C., Verkade, J.S., 2011. Estimation of predictive hydrological uncertainty using quantile regression: Examples from the National Flood Forecasting System (England and Wales). *Hydrology and Earth System Sciences* 15(1), 255–265. doi:10.5194/hess-15-255-2011

Chapter 6. Conclusions and Recommendations

6.1 Conclusions

This thesis focuses on probabilistic flood forecasting using Bayesian methods to adequately account for predictive uncertainty and improve the forecast performance and reliability. First, a Bayesian processor, Hydrologic Uncertainty Processor (HUP), was employed with different hydrologic models to quantify the hydrologic uncertainty in flood forecast. Next, a Bayesian Ensemble Uncertainty Processor (BEUP), which is an extension of HUP, was integrated with bias-corrected ensemble weather forecasts to assess the major uncertainties. Last, the sensitivity of the BEUP post-processor was explored by combining with different weather input types and different hydrologic model types. The main conclusions of the thesis are summarized as below.

6.1.1 Bayesian Flood Forecasting Methods

- Bayesian method provides a robust approach for probabilistic flood forecasting, the associated uncertainty is expressed in the posterior distribution that can be derived by combining the prior distribution with the likelihood function.
- Bayesian forecasting system is able to quantify all major sources of uncertainty and provides more reliable and accurate flood forecasts, all the information about the total uncertainty are summarized in the final predictive distribution.
- The framework of Bayesian forecasting system is quite flexible which allow it to be adapted to various purposes and work with any deterministic hydrologic model.

- Some advanced Bayesian forecasting methods (e.g. ensemble Bayesian forecasting system and Bayesian multi-model combination) are capable of overcoming certain limitations (e.g. single model or fixed model weight) and reducing the predictive uncertainty.

6.1.2 Hydrologic Uncertainty Processor with Different Models

- Hydrologic uncertainty is one large source of uncertainty in flood forecasting and cannot be overlooked. HUP provides a reliable and accurate analytic-numerical method for hydrologic uncertainty quantification based on the results of NSE, CRPS and reliability, and the uncertainty bound generated from HUP can well capture the observations.
- HUP, works as a hydrological post-processor, is able to improve the deterministic forecast from the hydrologic model, and yields more accurate probabilistic forecast with quantification of hydrologic uncertainty.
- As expected, hydrologic uncertainty increases as forecast lead time grows and also increases with increasing flow volume, the increased hydrologic uncertainty leads to deterioration of the HUP performance.
- For low peak flow events, after post-processing of HUP, a poorly performing hydrologic model could produce comparable probabilistic forecast as the better performing hydrologic model. While for high peak flow events, a better performing hydrologic model produces better probabilistic forecast after running HUP.

6.1.3 Bayesian Ensemble Uncertainty Processor with Ensemble Weather Forecasts

- BEUP is an extension of HUP for post-processing ensemble forecasts, and it considers meteorological uncertainty and hydrologic uncertainty within the framework. The performances of BEUP are promising for short-range forecasts (3h – 24h) and have little improvement for medium-range forecasts (24h – 72h).
- Bias correction of ensemble weather forecasts could greatly reduce its statistical discrepancy with observation, and bias correcting each ensemble member of weather forecasts produces better flood forecasts than just bias correcting the ensemble mean.
- As indicated by NSE, r , RMSE and CRPS, HUP is capable of enhancing the predictive performance for both short-range and medium-range forecasts. The improvement is significant for short lead times, and tends to be less obvious as lead time increases.
- For short-range forecast, the best results are obtained by applying both meteorological post-processing and hydrologic post-processing. Meteorological post-processing means applying bias correction to each ensemble member of weather inputs, and hydrologic post-processing means applying HUP to each ensemble of streamflow forecasts.

6.1.4 Sensitivity of Bayesian Ensemble Uncertainty Processor

- BEUP is a robust post-processing method for total major uncertainty estimation. As indicated by the high percentage of peak flow hit rate, the uncertainty bound produced from BEUP performs well in capturing the peak flows.

- BEUP combined with semi-distributed hydrologic model yields more accurate and reliable flood forecasts than combined with lumped hydrologic model. Using deterministic weather prediction with an ensemble dressing as input to BEUP outperforms using ensemble weather prediction as input.
- The improvement caused by the hydrologic model type (e.g. lumped or semi-distributed) is more significant than the improvement brought by weather input data type (e.g. deterministic or ensemble weather predictions).
- BEUP with semi-distributed hydrologic model is recommend for short-term (1 day ahead) flood forecasting. Using dressed deterministic weather prediction as forcing data to the system show better NSE, KGE and reliability, and using ensemble weather prediction as forcing data show better peak flow hit rate.

6.1.5 General Conclusions

The general conclusions from the thesis are listed as follows:

- Both HUP and BEUP are robust methods for probabilistic flood forecast with uncertainty quantification.
- Short-term flood forecasts using HUP and BEUP are accurate and reliable.
- Bias corrected meteorological inputs and well calibrated hydrologic models are recommended to use with Bayesian methods (HUP, BEUP) for flood forecasting.

6.2 Recommendations for Future Research

One topic that needs further research is assimilation of various sources of information within the Bayesian method. The main input to a flood forecast model is precipitation,

and the unknown future precipitation forms the primary source of uncertainty. There are various ways to collect precipitation data: rain gauge measured data, radar-derived rainfall data and satellite-based data. Radar could provide useful information in terms of spatial variability of precipitation, but they are often accompanied by biases and uncertainties. Techniques for adjusting radar rainfall to rain gauge measurements were developed (Sinclair and Pegram, 2005; Mazzetti and Todini, 2009). In the meantime, with the development of remotely sensed techniques which are able to improve the spatio-temporal resolution and reduce latency, advanced methods for retrieving precipitation from satellite-based microwave and infrared measurement have been developed (Joyce et al., 2004; Kubota et al., 2007). In addition, many other types of data may also add value, such as soil moisture. Most of the previous studies used limited sources of data as inputs, if multiple sources of data could be assimilated and blended together properly, it is expected the forecast accuracy and skill could be further improved.

Another interesting and challenging topic is the integration of Bayesian ensemble flood forecasting with multi-model for advanced flood forecasting. Previous studies have been conducted separately on Bayesian ensemble forecasting or on Bayesian multi-model combination, further research is required to integrate them to take advantage of both systems. If these two components could be combined appropriately in an adaptive and flexible framework, the model uncertainty and input data uncertainty could be greatly reduced and better quantified, and thus lead to significantly improved flood forecast with extended lead time.

Lastly, there have been great advances in probabilistic flood forecast with predictive uncertainty quantification, such as the previously mentioned methods in Chapter 2 and some recently proposed methods including HEAVEN (Hybrid Ensemble and Variational Data Assimilation framework for Environment systems) (Abbaszadeh et al., 2019), SDMU (state-dependent model uncertainty estimation method) (Pathiraja et al., 2018), and EPFM (Evolutionary Particle Filter with Markov Chain Monte Carlo) (Abbaszadeh et al., 2018). However, large efforts are still needed about how to communicate these uncertainties. For example, it would be more beneficial for the hydrology community and the public at large if the advanced technology on flood forecasting could be converted into software or smartphone app. The users just click and provide input; the system will run and give an answer with associated probability (Singh, 2019).

6.3 References

- Abbaszadeh, P., Moradkhani, H., Daescu, D.N., 2019. The Quest for Model Uncertainty Quantification: A Hybrid Ensemble and Variational Data Assimilation Framework. *Water Resources Research* 55(3), 2407-2431. doi: 10.1029/2018WR023629
- Abbaszadeh, P., Moradkhani, H., Yan, H., 2018. Enhancing hydrologic data assimilation by evolutionary Particle Filter and Markov Chain Monte Carlo. *Advances in Water Resources* 111, 192-204. doi: 10.1016/j.advwatres.2017.11.011
- Joyce, R.J., Janowiak, J.E., Arkin, P.A., Xie, P., 2004. CMORPH: A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution. *Journal of Hydrometeorology* 5(3), 487–503. doi:10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2
- Kubota, T., Shige, S., Hashizume, H., Aonashi, K., Takahashi, N., Seto, S., Hirose, M., Takayabu, Y.N., Ushio, T., Nakagawa, K. and Iwanami, K., 2007. Global Precipitation Map Using Satellite-Borne Microwave Radiometers by the GSMaP Project: Production and Validation. *IEEE Transactions on Geoscience and Remote Sensing* 45(7), 2259–2275.
- Mazzetti, C., Todini, E., 2009. Combining weather radar and raingauge data for hydrologic applications, pp. 230.
- Pathiraja, S., Moradkhani, H., Marshall, L., Sharma, A., Geenens, G., 2018. Data-driven model uncertainty estimation in hydrologic data assimilation. *Water Resources Research* 54(2), 1252-1280. doi: 10.1002/2018WR022627
- Sinclair, S., Pegram, G., 2005. Combining Radar and Rain Gauge Rainfall Estimates Using Conditional Merging. *Atmospheric Science Letters* 6(1), 19–22. doi:wiley.com/10.1002/asl.85
- Singh, V.P., 2019. Talk in 2019 FloodNet Annual General Meeting, Hamilton, Ontario, June 20, 2019.