Probabilistic Flood Forecasting Using Hydrologic Uncertainty Processor with Ensemble Weather Forecasts

SHASHA HAN

Department of Civil Engineering, McMaster University, Hamilton, Ontario, Canada

PAULIN COULIBALY

Department of Civil Engineering, and School of Geography and Earth Sciences, McMaster University, Hamilton, Ontario, Canada

(Manuscript received 29 November 2018, in final form 16 April 2019)

ABSTRACT

Recent advances in the field of flood forecasting have shown increased interests in probabilistic forecasting as it provides not only the point forecast but also the assessment of associated uncertainty. Here, an investigation of a hydrologic uncertainty processor (HUP) as a postprocessor 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's theorem and designed to capture hydrologic uncertainty. Ensemble forecasts are forced by ensemble weather forecasts from the Global Ensemble Prediction System (GEPS) that are inherently uncertain, and 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- 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 the short range is obtained by applying bias correction to each ensemble member plus applying the HUP postprocessor.

1. Introduction

As the fifth Intergovernmental Panel on Climate Change (IPCC) climate assessment report concludes, acceleration of the hydrological cycle due to climate change has led to more frequent floods over the last few decades (IPCC 2014). Large floods during recent years, such as the 2013 Southern Ontario flood (Canada), the 2017 Texas floods (United States), and the 2007 United Kingdom floods, increased demand for reliable flood forecasting systems and methods (Reggiani et al. 2009). If floods can be forecasted accurately in advance, up to 35% 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 parameters, 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 forecasting 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), ensemble model output statistics (EMOS; Gneiting et al. 2005), Bayesian model averaging (BMA; Raftery et al. 2005), and so on. In this study, the Bayesian forecasting

Corresponding author: Shasha Han, hans6@mcmaster.ca

system (BFS) introduced by Krzysztofowicz (1999) is selected due to its several salient properties: (i) it can produce probabilistic forecasts through any deterministic hydrologic model, (ii) the system is targeted to quantify all sources of uncertainties, and (iii) it is able to update the prior distribution to a posterior distribution based on Bayes's theorem by assimilating all the available information at the forecast time.

BFS consists of three parts: (i) an input uncertainty processor (IUP), (ii) a hydrologic uncertainty processor (HUP), and (iii) an 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 (Scheffzik 2016). Given the recent advances and popularity of ensemble weather products, instead of using the original probabilistic quantitative precipitation forecast (QPPF) in IUP, ensemble forecasts serve as the IUP component of the BFS in this study. In this 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 a Bayesian ensemble uncertainty processor (BEUP) in Reggiani et al. (2009).

Due to variable atmospheric conditions, 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 postprocessed before practical application (Maraun 2016). Here, ensemble weather forecasts produced by

the Global Ensemble Prediction System (GEPS) are used, and their bias is removed in different ways, resulting in different weather forecast datasets to drive the 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., GEPS, Regional Ensemble Prediction System (REPS)] to force their hydrologic models, therefore, besides the ensemble weather forecast datasets, the ensemble mean is also tested as a potential substitute for deterministic weather forecasting.

The contribution of this work is to integrate for the first time (to our best knowledge) the bias correction (meteorological postprocessing) and the HUP (hydrologic postprocessing) in flood forecasting and to 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; and (iii) investigating the forecast performance of using different weather forecast datasets, including raw GEPS, bias-corrected GEPS, and their ensemble mean.

The remainder of the paper is organized as follows. Section 2 gives a more detailed description of the applied methodology, including the bias correction method and Bayesian ensemble uncertainty processor, followed by a presentation of the study area and data in section 3. Section 4 presents the application results and compares the performance of different scenarios. Section 5 draws the conclusions.

2. Methodology

An overview of the methodology used for the probabilistic flood forecast with total uncertainty assessment is presented in a flowchart (see Fig. 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,

FIG. 1. Flowchart of the methodology.

HUP updates the prior distribution into the posterior distribution conditional on the model forecast and initial condition. The HUP parameters that 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 the hydrologic model, (ii) calibration of the hydrologic uncertainty processor, (iii) bias correction of ensemble weather forecasts, and (iv) application of ensemble weather forecasts with the hydrologic uncertainty processor. Detailed explanations about the bias correction method and the Bayesian ensemble processor are shown below.

a. 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

an 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 (MBCp; Cannon 2016), MBC rank correlation (MBCr; Cannon 2016), and MBC N -dimensional probability density function transform (MBCn; Cannon 2018); their performances were compared, and MBCn showed best results (Cannon 2018). Therefore, MBCn was applied in this study to bias correct the ensemble weather forecasts.

The MBCn algorithm matches the multivariate distribution using quantile delta mapping (QDM) and the N -dimensional probability density function transform (N -pdf). Three datasets should be prepared: the matrix of observations during the calibration period $\mathbf{X}_T$, the matrix of climate model output during the calibration period $\mathbf{X}_S$, and the matrix of climate model output during the projected period $\mathbf{X}_P$. To use MBCn, we assume that (i) observations during calibration period are the perfect representation of the true historical weather and (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 2018). Three steps are involved in MBCn (Cannon 2018): (i) apply the orthogonal rotation to the source, projection, and target datasets,

$$\begin{aligned}\tilde{\mathbf{X}}_S^{[j]} &= \mathbf{X}_S^{[j]} \mathbf{R}^{[j]}, \\ \tilde{\mathbf{X}}_P^{[j]} &= \mathbf{X}_P^{[j]} \mathbf{R}^{[j]}, \text{ and} \\ \tilde{\mathbf{X}}_T^{[j]} &= \mathbf{X}_T^{[j]} \mathbf{R}^{[j]},\end{aligned}\quad (1)$$

where superscript $[j]$ stands for the j th iteration of the N -pdf, and $\mathbf{R}$ is the uniformly distributed random orthogonal rotation matrix; (ii) correct each variable in rotated source and projection data by QDM based on the corresponding variable in rotated target data

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

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

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

suppose the dataset is an $I \times N$ 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

$$\begin{aligned}\mathbf{X}_S^{[j+1]} &= \tilde{\mathbf{X}}_S^{[j]} \mathbf{R}^{[j]-1}, \\ \mathbf{X}_P^{[j+1]} &= \tilde{\mathbf{X}}_P^{[j]} \mathbf{R}^{[j]-1}, \text{ and} \\ \mathbf{X}_T^{[j+1]} &= \tilde{\mathbf{X}}_T^{[j]}.\end{aligned}\quad (4)$$

Repeat steps i–iii until the multivariate distribution of the corrected source data matches the target data. A corresponding R package is available for download from <https://CRAN.R-project.org/package=MBC>.

b. Principle of a Bayesian ensemble uncertainty processor

The HUP is formulated as a Bayesian processor that postprocesses the model outputs through Bayesian revision. It revises the prior density, which is based on past

evidence, through the likelihood function that brings in various hydrologic uncertainty sources to the process, and yields a 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.'s (2009) approach. This involves applying Bayesian revision for each ensemble streamflow forecast and lumps the ensemble posterior distributions into a single posterior metadistribution as a representative function. However, instead of using linear regression to parameterize the prior density, a 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 linear regression was applied to a super-large watershed with an area of 160 000 km², which is unlikely to behave like a first-order Markov chain (Reggiani and Weerts 2008b). Also, instead of using a 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) and Reggiani et al. (2009), and more details about the HUP can be found in Krzysztofowicz and Kelly (2000), Krzysztofowicz and Herr (2001), and Krzysztofowicz (2002).

Following the notation in Krzysztofowicz's papers, let define n ($n = 1, \dots, N$) as forecast lead time, and v as a precipitation indicator, with $v = 1$ indicating precipitation occurrence, while $v = 0$ means no precipitation. Let H_n denote the discharge observation at time t_n , and the observed discharge at forecast time t_0 is H_0 . Let S_n denote the modeled discharge resulting from historical precipitation observation and $S_{n,j}$ denote 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 to 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, and 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:

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

where residual Ξ_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:

$$X_n = a_{nv} W_n + d_{nv} W_0 + b_{nv} + \Theta_n, \quad (6)$$

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. (5), the parameters of prior distribution C_{nv} and

t_{nv} ($v = 0, 1; n = 1, \dots, N$) are defined in Table 1. The parametric expression of prior distribution, conditional on discharge observation at time t_0 , is as follows:

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

The corresponding prior density is

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

Given the dependence parameters derived from Eq. (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 1. The posterior distribution for each ensemble member, conditional on each ensemble forecast and discharge observation at time t_0 , is expressed as

$$\Phi_{nvj}(h_n|s_{nj}, h_0) = Q\left\{\frac{Q^{-1}[\Gamma_{nv}(h_n)] - A_{nv}Q^{-1}[\bar{\Lambda}_{nv}(s_{nj})] - D_{nv}Q^{-1}[\Gamma_{0v}(h_0)] - B_{nv}}{T_{nv}}\right\}, \quad (9)$$

and the corresponding posterior density is

$$\phi_{nvj}(h_n|s_{nj}, h_0) = \frac{\gamma_{nv}(h_n)q\{Q^{-1}[\Phi_{nvj}(h_n|s_{nj}, h_0)]\}}{T_{nv}\{Q^{-1}[\Gamma_{nv}(h_n)]\}}. \quad (10)$$

Next, the ensemble posterior densities are integrated into one representative predictive density by averaging Eq. (10) over the ensemble s_{nj} :

$$\bar{\phi}_{nvj}(h_n|\bar{s}_n, h_0) = \frac{1}{J} \sum_{j=1}^J \phi_{nvj}(h_n|s_{nj}, h_0). \quad (11)$$

The corresponding cumulative probability distribution is stated as follows:

$$\bar{\Phi}_{nvj}(h_n|\bar{s}_n, h_0, \text{EF}), \quad (12)$$

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

3. Study area and data

The Humber River watershed was chosen as study area to apply the BEUP, with ensemble weather forecasts used as the input for the 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. (2019). 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) and gridded precipitation and temperature forecasts from GEPS (from June 2015 to December 2015). The hourly gauged precipitation and temperature time series were provided by the Toronto and Region Conservation Authority (TRCA), the hourly discharge time series were provided by the Water Survey of Canada, and the GEPS data were from Environment and Climate Change Canada (ECCC). As shown in Fig. 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, so in order to obtain

FIG. 2. Study area: Humber River watershed.

continuous time series, the missing mean areal precipitation was filled first by nearby Environment Canada (EC) stations and then using linear interpolation.

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–December 2014 as calibration and January 2015–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 0000 UTC daily and produces a forecast output every 3 h for 384 h. The system spatial resolution is 0.5° .

4. Application and discussion

a. Hydrologic model

A lumped conceptual rainfall–runoff model, McMaster University Hydrologiska Byråns Vattenbalansavdelning (MAC-HBV), 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 has been applied in many hydrological studies (Razavi and Coulibaly 2016, 2017); it adopts a similar concept to the HBV model from Merz and Blöschl (2004) and uses a modified routing routine following Seibert (1999). Instead of using the simplified Thornwaite formula to account for potential evapotranspiration (PE), an adjusted PE calculation approach proposed by Oudin et al. (2005a,b) 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 the 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) and requires only temperature and precipitation as inputs, while model output includes snow water equivalent (SWE) and rain plus snowmelt, which is passed to the soil moisture routine after adjustment by the 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

TABLE 2. 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 ⁻¹	0.90
Beta	A nonlinear 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 day ⁻¹	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 nonlinearity 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	Snowfall correction factor	0.4–1.6	—	0.76
Uadj	The average wind function during rain-on-snow periods	0.01–0.22	mm hPa ⁻¹ °C ⁻¹	0.09
Mbase	Base temperature for non-rain melt factor	0–1	°C	0.43
Mfmax	Maximum melt factor	0.2–2	mm (6 h) ⁻¹ °C ⁻¹	1.62
Mfmin	Minimum melt factor	0.02–0.7	mm (6 h) ⁻¹ °C ⁻¹	0.04
Tipm	Antecedent snow temperature index parameter	0.01–0.99	—	0.71
Nmf	Maximum negative melt factor	0.05–0.5	mm (6 h) ⁻¹ °C ⁻¹	0.18
plwhc	Percent of liquid water capacity	0–0.4	—	0.00
pxtemp1	Lower limit temperature dividing transition from snow	–2 to 2	°C	–1.90
pxtemp2	Upper limit temperature dividing rain from transition	1–3	°C	1.19

controlled by rainfall, snowmelt, and actual evapotranspiration. The runoff amount depends on the soil box water content, its maximum value fc, and a nonlinear runoff generation controlling parameter beta. The response routine comprises two reservoirs: an upper soil reservoir and a lower soil reservoir that represent the water storage in the upper zone and lower zone and estimate the total outflow of these two reservoirs. Recharge from soil moisture routine flows into the upper soil reservoir, and part of the water permeates into the lower soil reservoir based on the percolation rate parameter cperc. Thus, the total outflow includes three parts: (i) outflow from the upper zone which is controlled by a flow recession coefficient k0 if the water storage exceeds the threshold value lsuz, (ii) outflow from upper zone which is determined by flow recession coefficient k1 if lsuz is not exceeded, and (iii) a slow outflow from the lower zone affected by flow recession coefficient k2. 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 2, and a more detailed description of each routine and corresponding equations can be found in Samuel et al. (2011).

To match the GEPS format in forecast mode, the MAC-HBV model was calibrated in a 3-hourly time step using the particle swarm optimization algorithm (PSO) approach (Razavi and Coulibaly 2017) with the modified combined Nash–Sutcliffe efficiency (NSE) and volume error (VE) (NVE) as the objective function (Samuel et al. 2011; Razavi and Coulibaly 2017). As shown in Eq. (13), modified NVE (MNVE) 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:

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

Calibration and validation results are shown in Fig. 3. For calibration period, modified peak flow criteria [MPFC; MPFC = 1 – PFC; the equation for PFC can

FIG. 3. Calibration and validation for MAC-HBV.

be found in [Coulibaly et al. \(2001\)](#)], which provides more accurate performance evaluation for a flood period with 1 indicating a perfect forecast, could reach 0.81, NSE can reach 0.62, and VE is 0.23. For validation period, MPFC is near 0.80, the 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 has been on enhancing flood forecasting, so the model performance on the higher flows is satisfactory. Given that the model is in a 3-hourly time step, the NSE values are acceptable, and after postprocessing 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 2](#).

b. 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 h was considered. For every lead time n and each precipitation indicator v , the corresponding discharge observation

subsample 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 nonnormal distribution, a kernel was tested to be the most suitable distribution function. The goodness of fit for the kernel distribution is presented in [Fig. 4](#) for selected lead times. In the normal space after an NQT process, following [Eq. \(5\)](#) and the coefficient definitions for C_{nv} and t_{nv} in [Table 1](#), parameters for the prior distribution were calculated and shown in [Table 1](#). 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.

Similarly, for every n and each v , the corresponding simulated discharge subsample s_n that matches h_n was derived to estimate the marginal distribution. Again, a kernel was determined as the best function based on MSW test, the goodness of fit of the kernel distribution for selected lead times is displayed in [Fig. 5](#). Then in the transformed space, following [Eq. \(6\)](#) and the coefficient definitions for A_{nv} , B_{nv} , D_{nv} , and T_{nv} in [Table 1](#), parameters for the posterior distribution were computed and presented in [Table 1](#). It is noted that B_{nv} values are very small for all the cases and are thus approximated to zero. As lead time increases, A_{nv} increases and D_{nv} decreases for both branches, suggesting that the forecast is less affected by H_0 and more influenced by S_n with

FIG. 4. Marginal distribution of observed discharge for h18, h36, h54, and h72 conditional on the precipitation indicator.

increasing lead time. These HUP parameters characterize the prior distribution and posterior distribution based on Eqs. (7)–(12), and they are calibrated offline beforehand and will be used in forecast mode for probabilistic forecasting.

c. 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 the bias-corrected ensemble mean, and (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 the bias-corrected GEPS ensembles; (iii) GEPS-raw-mean, which represents the ensemble mean of GEPS-raw; (iv) GEPS-BC-mean, which is the ensemble mean of the 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.

The energy distance score measures the statistical discrepancy between sample x and y from two multivariate distributions; the 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 2018; Székely and Rizzo 2013). The squared energy distance is defined as (Rizzo and Székely 2016)

$$D^2(F, G) = 2Ex - y - Ex - x' - Ey - y' \geq 0, \quad (14)$$

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

FIG. 5. Marginal distribution of simulated discharge for s18, s36, s54, and s72 conditional on the precipitation indicator.

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.

d. Experiments and comparisons

The five types of GEPS datasets generated from bias correction were used as inputs to the calibrated HUP, resulting in seven application scenarios summarized in Table 3. The forecast performance of these seven different scenarios were evaluated and compared using multiple verification metrics and visual graphical tools, such as scatterplots, correlation coefficient r , root-mean-square 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 (3–24 h herein) and medium-range forecasts (24–72 h herein).

1) SCENARIO RESULTS COMPARISONS

To get a 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, scatterplots of single-valued forecasts versus observations for all seven scenarios (from Table 3) are presented in Fig. 7 and

FIG. 6. Comparison of energy distance scores.

TABLE 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 to postprocess the ensemble forecast from GEPS-raw to get probabilistic forecast
GEPS-BC+HUP	Use HUP to postprocess 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

Fig. 9 (the first four scenarios are in Fig. 7, and the other three are in Fig. 9). For medium-range forecasts, scatterplot comparisons between using GEPS only and combining GEPS with HUP (the first four scenarios) are presented in Fig. 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 a blue point, the 1:1 diagonals are emphasized by solid black lines, and the x axes and y axes are identical. Vertically, the scatterplot panels in each column are from the same scenario indicated at the top of the graph. Horizontally, the scatterplot panels in each row stand 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 scatterplot.

As shown in Figs. 7 and 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, and NSE and RMSE follow similar pattern and suggest a worsening trend. Comparison between Fig. 7 and Fig. 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. Scatterplot comparison for short-range forecasts between using GEPS mean with HUP is presented in Fig. 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, they show the most unsatisfactory performance among the seven scenarios. For example, the NSE value for a lead time of 24 h 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 are further analyzed in Fig. 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 with a different line style. For both the short and medium range, except GEPS-mean-BC+HUP, NSE values are quite similar across the other four Bayesian scenarios. NSEs 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 is represented by the orange dotted line, NSE values are acceptable for short lead times, but deteriorate as lead times exceed 21 h, 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 results over lead time 15 h. The best scenario for the medium range is GEPS-raw-mean+HUP, followed by GEPS-BC-mean+HUP, which shows comparable results from lead times 24 to 45 h.

FIG. 7. Scatterplot comparison for short-range forecasts between using GEPS data only and combining GEPS with HUP.

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:

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

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

Figure 11 presents the comparison of mean CRPS for different scenarios for both the short range (top) and medium range (bottom). The line style and color for each scenario (Fig. 11) are consistent with Fig. 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 the short range (below 3.00) and gradually rise with increasing lead time, and the CRPS values for the medium range are generally higher than the 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 the short and medium range, indicating HUP, which worked as a postprocessor of the ensemble forecast, is able to improve the forecast

FIG. 8. Scatterplot comparison for medium-range forecasts between using GEPS data only and combining GEPS with HUP.

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 the short range, the CRPS values for GEPS-mean-BC+HUP are similar to the other four Bayesian scenarios. Meanwhile, for the 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 h. 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 the short-range forecast is GEPS-BC+HUP. GEPS-raw-mean+HUP shows the best performance for the medium range, and GEPS-BC-mean+HUP takes second place.

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 a reliability plot introduced by [Laio and Tamea \(2007\)](#). A 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 corresponds to X_i and is mathematically expressed by $Z_i = P_i(X_i)$. The term 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, the reliability plot is a plot of Z_i values versus R_i/n , and 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

FIG. 9. Scatterplot comparison for short-range forecasts between using GEPS mean with HUP.

band and the bisector line depends on the significance level α and is computed via $q(\alpha)/\sqrt{n}$, here an 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 12 shows the reliability plots of the three best probabilistic forecast scenarios for both the short range (top) and medium range (bottom) for a selection of lead times; the $(Z_i, R_i/n)$ points for different lead times are presented by different marker types and colors. The bisector line is emphasized by a black solid line, and the Kolmogorov confidence

bands are plotted in a black dashed line. The evaluation criterion of the reliability curve can be found in [Laio and Tamea \(2007\)](#). It presents several possible outcomes: a curve below the bisector line indicates underprediction, while a curve above the line means overprediction. The S-shaped curve reveals a 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 times 3 and 6 h are located outside the confidence bands, and few points for lead times 12 and 24 h reach

FIG. 10. Comparison of NSE for different scenarios: (a) short range and (b) medium range.

the confidence line. While for GEPS-BC-mean+HUP, after bias correction process of GEPS forecasts, these points move closer 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 the short range.

3) FORECAST HYDROGRAPHS

A sample of forecast hydrographs at the watershed outlet at 0000 UTC 27 June 2015 is shown in Fig. 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

FIG. 11. Comparison of CRPS for different scenarios: (a) short range and (b) medium range.

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-dotted 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 postprocessing by HUP, the predictive median moved 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 Fig. 13 is only one example at 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

FIG. 12. Comparison of reliability plots for different scenarios: (a) short range and (b) medium range.

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 is determined by the precipitation forecast.

5. Conclusions

This paper has presented an application of HUP for postprocessing ensemble streamflow forecasts that 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 ensembles 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 the 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 a 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 (3–24 h) forecasts, but showed little to no improvement for medium-range (24–72 h) forecasts. The best scenario for short-range forecasts is GEPS-BC+HUP, which applies bias correction to each ensemble plus applies HUP.
- (ii) HUP, a hydrologic postprocessor 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.

FIG. 13. Ensemble forecasts and probabilistic forecasts: (a) using GEPS-raw as input and (b) using GEPS-BC as input (taking 27 Jun 2015 as the example).

- (iii) MBCn, which works like a meteorological postprocessor 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 a 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 the short range and medium range, GEPS-BC outperforms GEPS-raw; however, the results are quite similar between GEPS-raw+HUP and GEPS-BC+HUP, as well as between GEPS-raw-mean+HUP and GEPS-BC-mean+HUP. This reveals that the performance difference between using raw and bias-corrected weather forecasts becomes less noticeable after the 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 and alternative meteorological postprocessing

methods, along with hydrologic postprocessing technique, to further assess the potential of HUP for operational flood forecasting.

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 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.

REFERENCES

- Anderson, E., 2006: Snow accumulation and ablation model—SNOW-17. User's manual, NWS, 61 pp., http://www.nws.noaa.gov/oh/hrl/nwsrfs/users_manual/part2/_pdf/22snow17.pdf.
- Ashkar, F., and F. Aucoin, 2012: Choice between competitive pairs of frequency models for use in hydrology: A review and some new results. *Hydrol. Sci. J.*, **57**, 1092–1106, <https://doi.org/10.1080/02626667.2012.701746>.
- Bergström, S., 1976: Development and application of a conceptual runoff model for Scandinavian catchments. University of Lund Department of Water Resources Engineering Bulletin Series A, Vol. 52, SMHI Rep. 7, 134 pp.
- Cannon, A. J., 2016: Multivariate bias correction of climate model output: Matching marginal distributions and inter-variable dependence structure. *J. Climate*, **29**, 7045–7064, <https://doi.org/10.1175/JCLI-D-15-0679.1>.
- , 2018: Multivariate quantile mapping bias correction: An N-dimensional probability density function transform for climate model simulations of multiple variables. *Climate Dyn.*, **50**, 31–49, <https://doi.org/10.1007/s00382-017-3580-6>.
- , 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? *J. Climate*, **28**, 6938–6959, <https://doi.org/10.1175/JCLI-D-14-00754.1>.
- Coulbaly, P., F. Anctil, and B. Bobée, 2001: Multivariate reservoir inflow forecasting using temporal neural networks. *J. Hydrol. Eng.*, **6**, 367–376, [https://doi.org/10.1061/\(ASCE\)1084-0699\(2001\)6:5\(367\)](https://doi.org/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. *Wea. Forecasting*, **27**, 396–410, <https://doi.org/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. *J. Climate*, **25**, 3970–3984, <https://doi.org/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. *Mon. Wea. Rev.*, **133**, 1098–1118, <https://doi.org/10.1175/MWR2904.1>.
- Han, S., and P. Coulbaly, 2017: Bayesian flood forecasting methods: A review. *J. Hydrol.*, **551**, 340–351, <https://doi.org/10.1016/j.jhydrol.2017.06.004>.
- , —, and D. Biondi, 2019: Assessing hydrologic uncertainty processor performance for flood forecasting in a semiurban watershed. *J. Hydrol. Eng.*, [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001828](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001828), in press.
- 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. *Adv. Water Resour.*, **34**, 114–127, <https://doi.org/10.1016/j.advwatres.2010.10.002>.
- , —, —, —, and —, 2011b: Corruption of parameter behavior and regionalization by model and forcing data errors: A Bayesian example using the SNOW17 model. *Water Resour. Res.*, **47**, W07546, <https://doi.org/10.1029/2010WR009753>.
- Houle, E. S., B. Livneh, and J. R. Kasprzyk, 2017: Exploring snow model parameter sensitivity using Sobol' variance decomposition. *Environ. Modell. Software*, **89**, 144–158, <https://doi.org/10.1016/j.envsoft.2016.11.024>.
- IPCC, 2014: *Climate Change 2014: Mitigation of Climate Change*. Cambridge University Press, 1465 pp., <https://doi.org/10.1017/CBO9781107415416>.
- Jha, S. K., D. L. Shrestha, T. A. Stadnyk, and P. Coulbaly, 2018: Evaluation of ensemble precipitation forecasts generated through post-processing in a Canadian catchment. *Hydrol. Earth Syst. Sci.*, **22**, 1957–1969, <https://doi.org/10.5194/hess-22-1957-2018>.
- Kelly, K. S., and R. Krzysztofowicz, 2000: Precipitation uncertainty processor for probabilistic river stage forecasting. *Water Resour. Res.*, **36**, 2643–2653, <https://doi.org/10.1029/2000WR900061>.
- Koenker, R., 2005: *Quantile Regression*. Econometric Society Monographs, No. 38, Cambridge University Press, 349 pp., <https://doi.org/10.1017/CBO9780511754098>.
- Krzysztofowicz, R., 1999: Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resour. Res.*, **35**, 2739–2750, <https://doi.org/10.1029/1999WR900099>.
- , 2001: Integrator of uncertainties for probabilistic river stage forecasting: precipitation-dependent model. *J. Hydrol.*, **249**, 69–85, [https://doi.org/10.1016/S0022-1694\(01\)00413-9](https://doi.org/10.1016/S0022-1694(01)00413-9).
- , 2002: Bayesian system for probabilistic river stage forecasting. *J. Hydrol.*, **268**, 16–40, [https://doi.org/10.1016/S0022-1694\(02\)00106-3](https://doi.org/10.1016/S0022-1694(02)00106-3).
- , and K. S. Kelly, 2000: Hydrologic uncertainty processor for probabilistic river stage forecasting. *Water Resour. Res.*, **36**, 3265–3277, <https://doi.org/10.1029/2000WR900108>.
- , and H. D. Herr, 2001: Hydrologic uncertainty processor for probabilistic river stage forecasting: Precipitation-dependent model. *J. Hydrol.*, **249**, 46–68, [https://doi.org/10.1016/S0022-1694\(01\)00412-7](https://doi.org/10.1016/S0022-1694(01)00412-7).
- , and C. J. Maranzano, 2004: Bayesian system for probabilistic stage transition forecasting. *J. Hydrol.*, **299**, 15–44, <https://doi.org/10.1016/j.jhydrol.2004.02.013>.
- Laio, F., and S. Tamea, 2007: Verification tools for probabilistic forecast of continuous hydrological variables. *Hydrol. Earth Syst. Sci.*, **11**, 1267–1277, <https://doi.org/10.5194/hess-11-1267-2007>.
- Li, H., J. Sheffield, and E. F. Wood, 2010: Bias correction of monthly precipitation and temperature fields from Inter-governmental Panel on Climate Change AR4 models using equidistant quantile matching. *J. Geophys. Res.*, **115**, D10101, <https://doi.org/10.1029/2009JD012882>.

- Maraun, D., 2016: Bias correcting climate change simulations - A critical review. *Curr. Climate Change Rep.*, **2**, 211–220, <https://doi.org/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. *Hydrol. Earth Syst. Sci.*, **12**, 551–563, <https://doi.org/10.5194/hess-12-551-2008>.
- Merz, R., and G. Blöschl, 2004: Regionalisation of catchment model parameters. *J. Hydrol.*, **287**, 95–123, <https://doi.org/10.1016/j.jhydrol.2003.09.028>.
- Oudin, L., C. Michel, and F. Anctil, 2005a: Which potential evapotranspiration input for a lumped rainfall-runoff model?: Part 1—Can rainfall-runoff models effectively handle detailed potential evapotranspiration inputs? *J. Hydrol.*, **303**, 275–289, <https://doi.org/10.1016/j.jhydrol.2004.08.025>.
- , F. Hervieu, C. Michel, C. Perrin, V. Andréassian, F. Anctil, and C. Loumagne, 2005b: Which potential evapotranspiration input for a lumped rainfall-runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling. *J. Hydrol.*, **303**, 290–306, <https://doi.org/10.1016/j.jhydrol.2004.08.026>.
- Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.*, **133**, 1155–1174, <https://doi.org/10.1175/MWR2906.1>.
- Razavi, T., and P. Coulibaly, 2016: Improving streamflow estimation in ungauged basins using a multi-modelling approach. *Hydrol. Sci. J.*, **61**, 2668–2679, <https://doi.org/10.1080/02626667.2016.1154558>.
- , and —, 2017: An evaluation of regionalization and watershed classification schemes for continuous daily streamflow prediction in ungauged watersheds. *Can. Water Resour. J.*, **42**, 2–20, <https://doi.org/10.1080/07011784.2016.1184590>.
- Reggiani, P., and A. H. Weerts, 2008a: Probabilistic quantitative precipitation forecast for flood prediction: An application. *J. Hydrometeor.*, **9**, 76–95, <https://doi.org/10.1175/2007JHM858.1>.
- , and —, 2008b: A Bayesian approach to decision-making under uncertainty: An application to real-time forecasting in the river Rhine. *J. Hydrol.*, **356**, 56–69, <https://doi.org/10.1016/j.jhydrol.2008.03.027>.
- , 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 Resour. Res.*, **45**, W02428, <https://doi.org/10.1029/2007WR006758>.
- Regonda, S. K., 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. *J. Hydrol.*, **497**, 80–96, <https://doi.org/10.1016/j.jhydrol.2013.05.028>.
- Rizzo, M. L., and G. J. Székely, 2016: Energy distance. *Wiley Interdiscip. Rev. Comput. Stat.*, **8**, 27–38, <https://doi.org/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. *J. Hydrol. Eng.*, **16**, 447–459, [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000338](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000338).
- Schefzik, R., 2016: Combining parametric low-dimensional ensemble postprocessing with reordering methods. *Quart. J. Roy. Meteor. Soc.*, **142**, 2463–2477, <https://doi.org/10.1002/qj.2839>.
- Seibert, J., 1999: Regionalisation of parameters for a conceptual rainfall-runoff model. *Agric. For. Meteorol.*, **98–99**, 279–293, [https://doi.org/10.1016/S0168-1923\(99\)00105-7](https://doi.org/10.1016/S0168-1923(99)00105-7).
- Székely, G. J., and M. L. Rizzo, 2013: Energy statistics: A class of statistics based on distances. *J. Stat. Plan. Inference*, **143**, 1249–1272, <https://doi.org/10.1016/j.jspi.2013.03.018>.
- Todini, E., 2008: A model conditional processor to assess predictive uncertainty in flood forecasting. *Int. J. River Basin Manage.*, **6**, 123–137, <https://doi.org/10.1080/15715124.2008.9635342>.
- United Nations, 2004: Guidelines for reducing flood losses. U.N. International Strategy for Disaster Reduction, 79 pp., https://sustainabledevelopment.un.org/content/documents/flood_guidelines.pdf.
- 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. *J. Hydrol.*, **555**, 257–277, <https://doi.org/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 Resour. Res.*, **41**, W01017, <https://doi.org/10.1029/2004WR003059>.
- Wang, L., and W. Chen, 2014: A CMIP5 multimodel projection of future temperature, precipitation, and climatological drought in China. *Int. J. Climatol.*, **34**, 2059–2078, <https://doi.org/10.1002/joc.3822>.