Journal of Hydrology 508 (2014) 147–156

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Hydrologic post-processing of MOPEX streamflow simulations

Aizhong Ye [a,*], Qingyun Duan [a], Xing Yuan [b], Eric F. Wood [b], John Schaake [c]

[a] College of Global Change and Earth System Science, Beijing Normal University, Beijing, China
[b] Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, USA
[c] Consultant, Annapolis, MD, USA

ARTICLE INFO

Article history:
Received 21 November 2012
Received in revised form 23 October 2013
Accepted 30 October 2013
Available online 7 November 2013
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Ezio Todini, Associate Editor

Keywords:
Post-processor
Calibration
Hydrologic model
MOPEX

SUMMARY

There are many approaches to improve hydrologic model predictions, including pre-processing to deal with input uncertainty, data assimilation to treat initial and boundary condition uncertainty, model calibration to reduce parametric uncertainty. Hydrologic post-processing is an approach for treating uncertainties from hydrologic model outputs propagated from all upstream sources. It works by relating model outputs (e.g., streamflow) to corresponding observations through a statistical model. This paper compares the effect of post-processing and model calibration in improving hydrologic forecasts under different hydroclimatic conditions and across different models.

Observed and simulated daily streamflow data from the Second Workshop on Model Parameter Estimation Experiment (MOPEX) were used for the comparisons described above. The results from 7 hydrologic models showed that post-processing alone was better than the results from hydrologic model calibrations for 12 basins in the eastern United States. The predictive QQ plot indicates that the predictive distributions of post-processed ensemble streamflow simulations are reliable. Post-processed results were similar for different hydrologic models, but were quite different for different basins. In terms of ensemble prediction, post-processing results tended to be over-confident. In general, post-processing can improve hydrological forecasts and reduce uncertainty in wet basins, but caution should be taken when applying post-processing to dry basins where there are many zeros values in the data.

1. Introduction

There are many approaches to improve hydrologic model predictions by reducing uncertainties from various sources including model inputs, initial and boundary conditions, model structure and model parameters. Model inputs, especially precipitation inputs, are most critical to good hydrologic predictions. There are two basic ways to improve precipitation inputs: the improvement of quantitative precipitation estimates (QPE) and the improvement of quantitative precipitation forecast (QPF). For accurate and reliable QPE, it is necessary to establish a good observational network to measure precipitation amount and a good methodology to mesh measurements from different sources (Xie et al., 2007; Cherubini et al., 2002; Zhou et al., 2008). To improve QPF generated by numerical weather prediction (NWP) models, statistical post-processing techniques (referred to pre-processing in hydrologic modelling community) are often used (Clark and Slater, 2006; Glahn et al., 2008; Schaake et al., 2007). Data assimilation techniques are usually used to improve the estimates of initial and boundary conditions (Andreadis and Lettenmaier, 2006; Slater and Clark, 2006). Model calibration is the common approach to reduce uncertainties due to incorrect specification of model parameters (Duan et al., 1992). Hydrologic post-processing works directly on hydrologic model outputs by using a statistical model to represent the relationship between model outputs (e.g., streamflow) and corresponding observations. It serves the purpose of removing model biases from all upstream uncertainty sources and is the final step before the issue of actual hydrologic forecasts.

As hydrologic models become more complicated, data assimilation and model calibration are becoming more difficult and require more computational resources. In the case of high resolution distributed modelling over a large basin, the heterogeneity of model states and model parameters over a large number of hydrological modelling units make it difficult to apply data assimilation techniques such as Ensemble Kalman Filter (Andreadis and Lettenmaier, 2006) or traditional optimization methods such as SCE-UA (Shuffled Complex Evolution method developed at the University of Arizona) (Duan et al., 1994,1992). Moreover, human interventions, such as reservoir management, which are not considered in most hydrologic models, can complicate data assimilation or model calibration of the affected basins. In contrast, since hydrologic post-processing only considers the model output-observation relationship, it can overcome several limitations complicating model calibration and data assimilation. First, only

* Corresponding author. Tel./fax: +86 10 58804191.
E-mail address: azye@bnu.edu.cn (A. Ye).

0022-1694/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jhydrol.2013.10.055

limited computational resources are required for post-processing. Second, the post-processing approach may be applicable in management-affected basins.

Post-processing quantifies and reduce uncertainties in model predictions, renders the predicted probability unbiased in model-predicted flow (Krzysztofowicz and Kelly, 2000; Seo et al., 2006). The main idea of post-processing is to calculate the conditional probability (Seo et al., 2000) of observed flow given forecast flow. Krzysztofowicz and his associates proposed a Bayesian Processor of Output (BPO) to post-process streamflow forecasts (Krzysztofowicz and Kelly, 2000; Krzysztofowicz and Maranzano, 2004). Based on this Bayesian framework, Todini proposed a Model Conditional Processor (MCP) to compute the posterior probability distribution of streamflow prediction error based on all information available, including observations and predictions from one or more models (Todini, 2008; Coccia and Todini, 2011). Montanari and Brath (2004) proposed a meta-Gaussian approach in order to estimate the probability distribution of the rainfall-runoff prediction error using multivariate regression which relates streamflow errors to several predictor variables. Weerts et al. (2011) presented a quantile regression approach to quantify the streamflow predictive uncertainty, in which the streamflow observations and predictions are divided into different quantiles in order to avoid any prescriptive assumption of the variables in the regression. Seo et al. (2006) presented a simple and inexpensive statistical post-processor for ensemble streamflow prediction (ESP). Another approach, based on the state-space model and wavelet transformation, was used to correct errors of simulated (forecasted) discharge (Bogner and Kalas, 2008). Wood and Schaake (2008) used the correlation between forecast ensemble means and observations to generate a conditional forecast, with means and spreads lying between climatological means and spreads (when the forecast has no skill) and the raw forecast mean, with zero spread (when the forecast is perfect). In many of the approaches described above, they make use of the Normal Quantile Transform (NQT) to project the observations and the predictions into the Normal space.

These approaches can remove or reduce systematic biases and provide post-processed ensemble streamflow forecasts that are significantly better than raw forecasts. Does that mean that the post-processor can replace model calibration in streamflow forecasts? Shi et al. (2008) addressed this question with regard seasonal hydrologic forecasting for several western U.S. basins and found that post-processing alone is almost as effective as hydrologic model calibration. Yuan and Wood (2012) found that post-processing streamflow forecasts from a coupled ocean-land-atmosphere seasonal forecast model, where its land surface model is uncalibrated, has comparable performance to a well-calibrated hydrologic model driven by bias-corrected meteorological forcing. This finding indicates that a global climate forecast model with an uncalibrated land surface component can provide a useful streamflow forecast after post-processing. However, the above studies are based on one or two specific hydrologic models. Whether the post-processor can replace model calibration in different hydroclimatic conditions for different hydrologic models is still uncertain.

This paper systematically compares results from post-processing and calibration of 7 models in 12 basins in the Southeastern U.S. The organization of the paper is as follows: section 2 describes the General Linear Model Post-Processor (GLMPP); section 3 introduces the data and study domain; section 4 presents results and discussion; and section 5 provides conclusions.

2. Model description

We selected the General Linear Model Post-Processor (GLMPP), which was recently developed (Zhao et al., 2011), to compare the post-processing and calibration results. GLMPP has the following properties: (1) it removes mean bias; (2) it produces an ensemble of members representing, in an "equally-likely" sense, the observed hydrograph being predicted; and (3) it preserves temporal scale dependency relationships in streamflow hydrographs and uncertainty in the predictions.

Fig. 1 is the flow chart illustrating the procedure of GLMPP. Given observed and simulated streamflow data, first select the forecast date and forecast window. The forecast window consists of an analysis period prior to the forecast date and a forecast period from the forecast date. The lengths of analysis and forecast periods are N_a and N_f, respectively. The Generalized Linear Model Post-Processor (GLMPP) for the forecast date can be expressed as:

$$Z_{1.2} = A \cdot Z_2 + B \cdot E \tag{1}$$

where $Z_{1,2}$ is the predict and given the predictor vector, Z_2, i.e., $Z_{1,2} = Z_1|Z_2$. A and B are parameter matrices. Z_1 is the observed streamflow for the forecast period (i.e., $Z_1 = \{Q_{o,f}(t), t = 1,\ldots,N_f\}$), Z_2 is the predictor vector, which is made of simulated and observed streamflow for the analysis period, $\{Q_{s,\boldsymbol{a}}\}$ and $\{Q_{o,\boldsymbol{a}}\}$, and simulated streamflow for the forecast period $\{\boldsymbol{Q_{o,f}}\}$ (i.e., $Z_2 = \{Q_{s,f}(t), t = 1,\ldots,Nf, Q_{o,a}(t), t = 1,\ldots,N_a; Q_{s,a}(t), t = 1,\ldots,N_a\}$), A and B are coefficient matrices of the regression equation, $\boldsymbol{E}$ is the random error term. If $\boldsymbol{E}$ is identical independently distributed (i.d.d.) Gaussian, then Eq. (1) can be easily solved by multivariate linear regression.

To make E i.d.d., all observed and simulated streamflow values undergo a normal quantile transform (NQT). To make the parameter estimates more robust, a "buffer" period with a length of N_b (in days) is introduced to include data (i.e., Z_1 and Z_2) for $N_b/2$ days prior to the analysis period and $N_b/2$ days after the forecast period to enlarge the data sample size for calibrating GLMPP.

Before solving A and B, we first compute covariance matrices: $\Sigma = \begin{bmatrix} \Sigma_{11}\Sigma_{12} \\ \Sigma_{21}\Sigma_{22} \end{bmatrix}$, where $\Sigma_{11}, \Sigma_{12}, \Sigma_{21}$ and Σ_{22} are the covariance matrices of Z_1, Z_1 and Z_2, Z_2 and Z_1 and Z_2, respectively. Based on linear regression, we obtain $A = \Sigma_{12} \cdot \Sigma_{22}^{-1}$ and $BB^T = \Sigma_{11} - \Sigma_{12} \cdot \Sigma_{22}^{-1} \cdot \Sigma_{21}$.

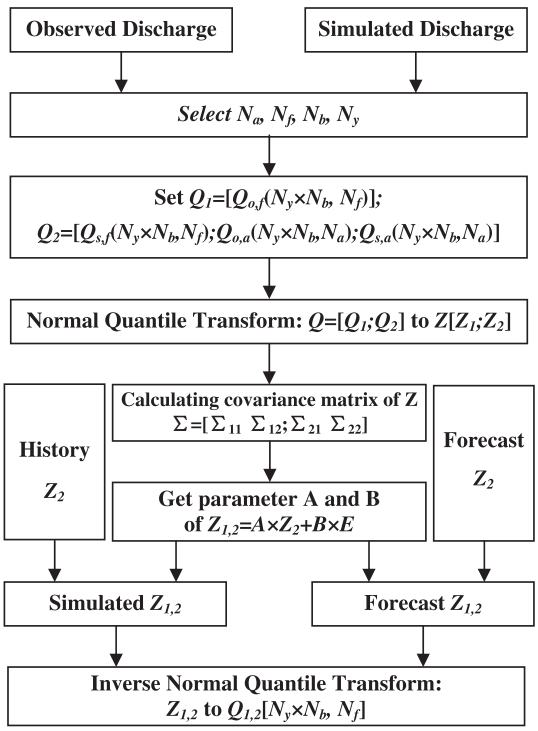

Fig. 1. Flow chart for the post-processor.

There is a set of A and B for each forecast date. For each day in a calendar year, we generated 365 sets of A and B. After A and B are solved, we can use Monte Carlo method to generated streamflow ensembles.

GLMPP is different from the MCP approach proposed by Todini (2008) because the former is a linear regression of the NQT transformed input–output pairs, as opposed to the Bayesian framework employed by the latter. This method also differs from that of Montanari and Brath (2004) in that GLMPP divides the data into analysis period and forecast period and includes a buffer period. The use of an analysis period plays a role similar to data assimilation, while the buffer period augments the sample sizes for more robust model calibration.

Model performance measures include the Nash-Sutcliffe efficiency value (NSE), correlation coefficient (R), water balance bias, and Root Mean Square Error (RMSE). They are computed as follows:

$$NSE = \left[1 - \frac{\sum(Q_C - Q_O)^2}{\sum(Q_O - \overline{Q_O})^2}\right] \tag{2}$$

$$R = \frac{\sum(Q_c - \overline{Q}_c)(Q_o - \overline{Q}_o)}{\sqrt{\sum(Q_c - \overline{Q}_c)^2 \sum(Q_o - \overline{Q}_o)^2}} \tag{3}$$

$$Bias = \frac{SD}{OD} - 1 \tag{4}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Q_{ci} - Q_{oi})^2} \tag{5}$$

where $Q_O, Q_C, \overline{Q_O}, \overline{Q_c}$ are observed, simulated, average observed and average simulated discharges, SD is the sum of simulated discharges, and OD is the sum of observed discharges. A NSE score of 1 indicates perfect forecast. The long term mean as a forecast would result a NSE of 0. A negative NSE value suggests that the forecast is worse than the long term average. A perfect $RMSE$ score is 0. The squared difference is used in the calculation, so $RMSE$ gives more weight to high than low values.

In order to evaluate the reliability of predictive distributions, Laio and Tamea (2007) suggested the use of the predictive QQPLOT (Thyer et al., 2009; Biondi and De Luca, 2013), which does not require a subjective binning of the data. Let us denote z_i as the correct probabilistic forecast of realization x_i ($i = 1,\ldots,n$) and plot it against their empirical cumulative distribution R_i/n, where R_i is the rank (position) of the ith value in the ordered vector of z_i values. The degree of the departure from the bisector (the 1:1 line) represents

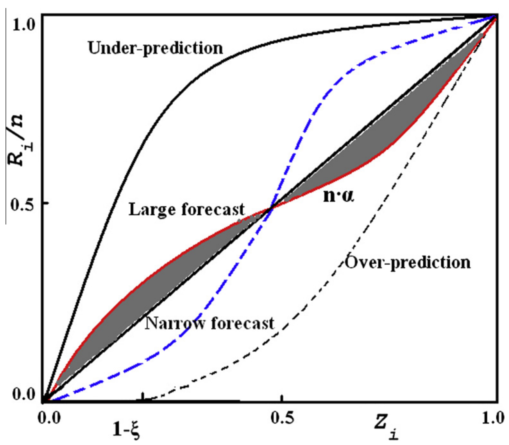

Fig. 2. Schematic of the predictive QQ plot and derived indexes (Laio and Tamea, 2007; Renard et al., 2010).

interpretable deficiencies (see Fig. 2). There are two indices, the area between the QQ curve and the bisector (α) and the complement of the fraction of z_i values equal to 0 or 1 that corresponds to the proportion of observed discharge values outside the range of the predictive distribution (ξ), that quantify the reliability (Renard et al., 2010). α and ξ are computed as follows:

$$\alpha = \frac{1}{n}\sum_{i=1}^{n}\left|z_i - \frac{Ri}{n}\right| \tag{6}$$

$$\xi = 1 - \frac{1}{n}\sum_{i=1}^{n}\beta \text{ with } \beta = \begin{cases} 1 & \text{if } z_i = 0 \text{ or } z_i = 1 \\ 0 & \text{otherwise} \end{cases}$$

α varies between 0.5 (worst reliability, with all realizations outside their predictive range) and 0 (perfect reliability), while ξ varies between 0 (all realizations outside their predictive range) and 1 (no incompatible realizations).

3. Data and study domain

Data used in this study are the simulated and observed daily streamflow data from the Second Workshop on Model Parameter Estimation Experiment (MOPEX) database (Duan et al., 2006). There are 7 different models in the MOPEX database (Table 1), which were run using uncalibrated and calibrated parameters for 12 Southeastern U.S. basins (Table 2, Fig. 3). The uncalibrated parameters are the so called "default parameters" defined based on geomorphologic properties such as soil and vegetation characteristics, topographical features (e.g., slope and elevation), or experiments. Different models may determine their default parameters differently. Calibrated parameters were obtained through optimization which minimizes the aggregated differences between simulated and observed streamflow values. Streamflow data used in this study covers the period from 1962 to 1997. The ratio of annual runoff to precipitation is from 0.15 to 0.63 (Table 2), so hydroclimatic conditions are different among the 12 basins.

4. Results and discussion

4.1. GLMPP parameters

We used a set of algorithmic parameters suggested by Zhao et al. (2011) for the GLMPP, who have done a comprehensive study of the effects of the length of analysis period, the length of forecast period and the length of the buffer period. According to Zhao et al. (2011), the analysis period preceding the forecast are used to compensate for unwanted effects of imperfect estimates of initial conditions on forecasts. The length of the analysis period N_a should be greater than 5. So we used a value of 10 for N_a in this study. The length of the forecast period N_f was chosen based on the user needs. We chose a value of 30 for N_f for this study. A "buffer" period N_b is used to increase the sample size for the GLMPP. A reasonable value for N_b can increase robustness of the GLMPP results. For

Table 1
Hydrological models from MOPEX (Duan et al., 2006).

ID	IDModel	Note
1	grj4	GR4J model
2	isba	ISBA model
3	noa	NOAH model
4	sac	Sacramento model
5	swap	SWAP model
6	swb	Simple Water Balance model
7	vic	VIC model
8	mean	Multi-models mean

Table 2
River basins and complementary information[a].

Basin ID	USGS ID	Long.	Lat.	Area (km^2)	Flood flow (mm/day)	Annual prec.(mm)	Annual runoff (mm)	Runoff/prec.	Station name
B1	07378500	−90.9903	30.4639	3315	17.29	1560	612	0.39	Amite River Near Denham Springs, LA
B2	03451500	−82.5786	35.6092	2448	20.21	1378	795	0.58	French Broad River At Asheville, NC
B3	03054500	−80.0403	39.15	2372	26.41	1164	736	0.63	Tygart Valley River At Philippi, WV
B4	07186000	−94.5661	37.2456	3015	13.65	1075	300	0.28	Spring River Near Waco, MO
B5	01608500	−78.6544	39.4469	3810	12.98	1043	339	0.33	S Branch Potomac River Nr Springfield, WV
B6	01643000	−77.3661	39.3869	2116	15.62	1042	421	0.4	Monocacy R At Jug Bridge Nr Frederick, MD
B7	01668000	−77.5181	38.3222	4134	41.16	1028	375	0.36	Rappahannock River Nr Fredericksburg, VA
B8	03179000	−81.0106	37.5439	1020	13.37	1017	419	0.41	Bluestone River Nr Pipestem, WV
B9	03364000	−85.9256	39.2	4421	8.78	1014	377	0.37	East Fork White River At Columbus, IN
B10	05455500	−91.7156	41.4664	1484	6.7	881	261	0.3	English River At Kalona, IA
B11	08172000	−97.6506	29.6661	2170	6.07	819	170	0.21	San Marcos River At Luling, TX
B12	08167500	−98.3833	29.8603	3406	29.12	761	116	0.15	Guadalupe River Nr Spring Branch, TX

[a] Basin ID numbers in accordance with annual precipitation.

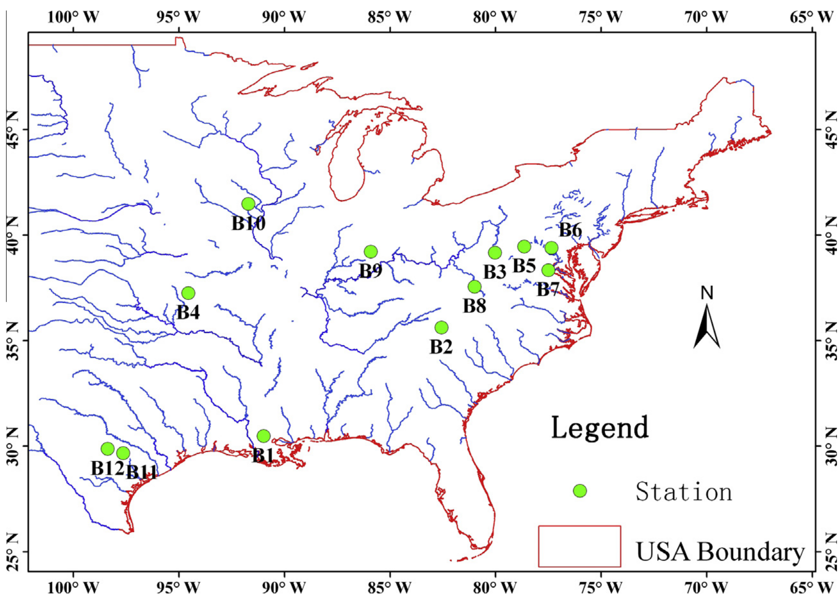

Fig. 3. Location of the basins.

this study, a value of 14 is chosen for N_b. The same values for N_a, N_f, and N_b are used in all basins for all models.

We used a split data method to evaluate the effectiveness of GLMPP. Data from 1962 to 1980 were used to calibrate the post-processor parameters, and data from 1981 to 1997 were used for verification.

4.2. Comparison of raw streamflow simulations and post-processed streamflow simulations

There are two sets of raw streamflow simulations for each model and in each basin, one using uncalibrated model parameters and another using calibrated parameters. For those raw streamflow simulations, we applied GLMPP to post-process them. We calculated four performance indices as described in Eqs. (2)–(5) for both the raw and post-processed streamflow simulations, which are denoted as "uncal", "cal", "postuncal" and "postcal", respectively. The orders of models and basins in the following figures are the same as shown in Tables 1 and 2.

We selected different periods to calibrate and verify GLMPP. The calibration period is 1962–1980, and the verification period is 1981–1997. The performance indices for the four sets of streamflow simulations are given in Figs. 4 and 5. Note that in these figures, horizontal axis denotes different basins, while vertical axis denotes different models. For the first, second and fourth columns, "red[1]" color indicates good (preferred) scores and "blue" color implies bad scores. For the third column (i.e., bias), the preferred score is zero, which corresponds to "light green" color. The results suggest that streamflow simulations using calibrated model parameters are better than that of the uncalibrated model parameters in both the calibration and verification period. Streamflow simulations after post-processing are generally much better than that of the calibrated

[1] For interpretation of color in Figs. 4, 5 and 8, the reader is referred to the web version of this article.

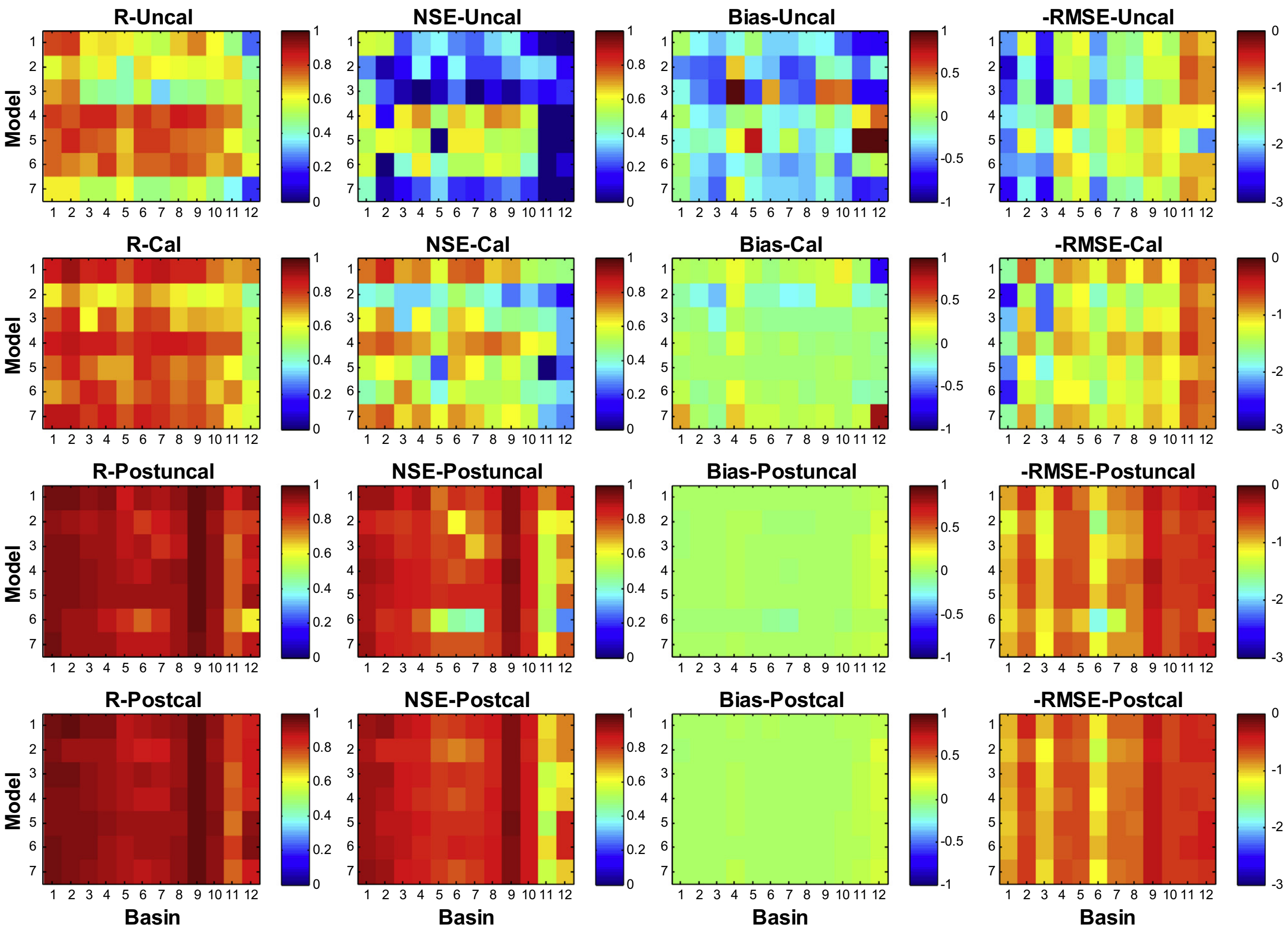

Fig. 4. Performance indices calculated for raw and post-processed streamflow simulations during the calibration period (1962–1980). Rows 1 to 4 denote cases for raw uncalibrated (Uncal), raw calibrated (Cal), post-processed of uncalibrated streamflow, and post-processed of calibrated streamflow, respectively. Columns 1 to 4 denote correlation coefficient, NS efficiency, bias and RMSE, respectively.

model parameters. Post-processed streamflow simulations are slightly worse than those using calibrated model parameters for the verification period in two dry basins. The reason post-processing generally leads to better performance indices than model calibration alone is that post-processing works directly to correct the errors in the model outputs. It can even implicitly deal with alteration to streamflow values due to human activities, if these activities are not totally spurious. Model calibration is generally done to ensure that the aggregate errors over the calibration period are minimized. But it cannot effectively deal with the model structural errors that lead the model to have different predictive errors in different seasons and in different places.

To compare the performance of different models in different basins, we calculated the mean performance indices for all models (Fig. 6) and for all basins (Fig. 7). The mean indices are calculated as follows: first, we obtain model performance indices for model and for each basin (see Fig. 4 and Fig. 5); second, we averaged all indices over each model or over each basin (see Figs. 6 and 7). Figs. 6 and 7 clearly show that the performance indices for raw streamflow simulations (i.e., the first two bars for each basin) are much worse than that for post-processed streamflow simulations (i.e., the third and fourth bars for each basin) for the calibrated period. The performance indices for post-processed streamflow simulations are mostly better than that for raw streamflow simulations in the verification period except for two dry basins (i.e., B11 and B12), in which the indices for post-processed streamflow are actually worse than that of the raw calibrated streamflow simulations. This result is due to the fact that streamflow data for the dry basins contain too many zeros, which lead to truncated Normal distributions after observed and simulated streamflow values are transformed to the Normal space via NQT. The truncated Normal distribution would result in incorrect solution of Eq. (1) because the Gaussian assumption for GLMPP is violated. This problem is actually not limited to GLMPP, but also some other post-processing models that use NQT to achieve the Normality requirement for the data. Certain ways were attempted to alleviate this problem, including the use of quantile regression, Bayesian approach that does not depend on Gaussian assumption (Krzysztofowicz and Herr, 2001; Weerts et al., 2011; Todini, 2008) or considering two distributions for positive and negative errors (Montanari and Grossi, 2008).

Since there are significant seasonal variations in climatic conditions over the different basins, the performance indices may also display seasonal patterns. To investigate seasonal variation in performance indices, we analyzed the long-term average monthly streamflow for the verification period. Fig. 8 displays the long-term monthly average raw and post-processed streamflow values for the Gr4j model (The results for other models are basically the same and are, therefore, not shown). We note that the uncalibrated raw streamflow monthly values (i.e., the green solid line) are quite different from the observations (i.e., the solid black line) in most basins. Calibrated raw streamflow values (i.e., blue dash line) are generally closer to observations than uncalibrated streamflow values, but significant errors still exist for most basins in calibrated streamflows. All post-processed streamflow values (i.e., red line) are very close to observations (we only plotted post-processed streamflows for the uncalibrated case because the post-processed streamflows for

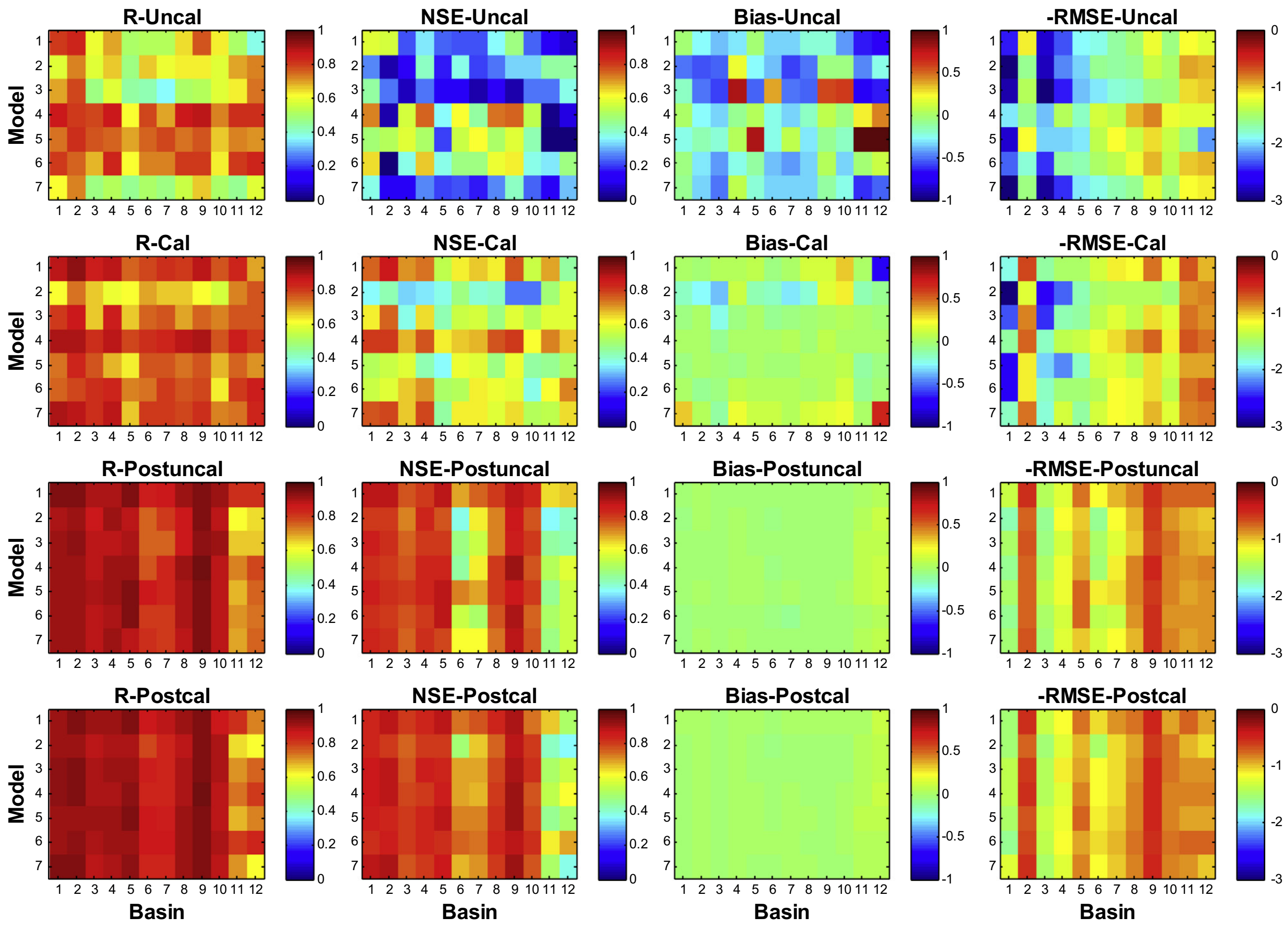

Fig. 5. Performance indices calculated for raw and post-processed streamflow simulations during the verification period (1981–1997). Rows 1 to 4 denote cases for raw uncalibrated (Uncal), raw calibrated (Cal), post-processed of uncalibrated streamflow, and post-processed of calibrated streamflow, respectively. Columns 1 to 4 denote correlation coefficient, NS efficiency, bias and RMSE, respectively.

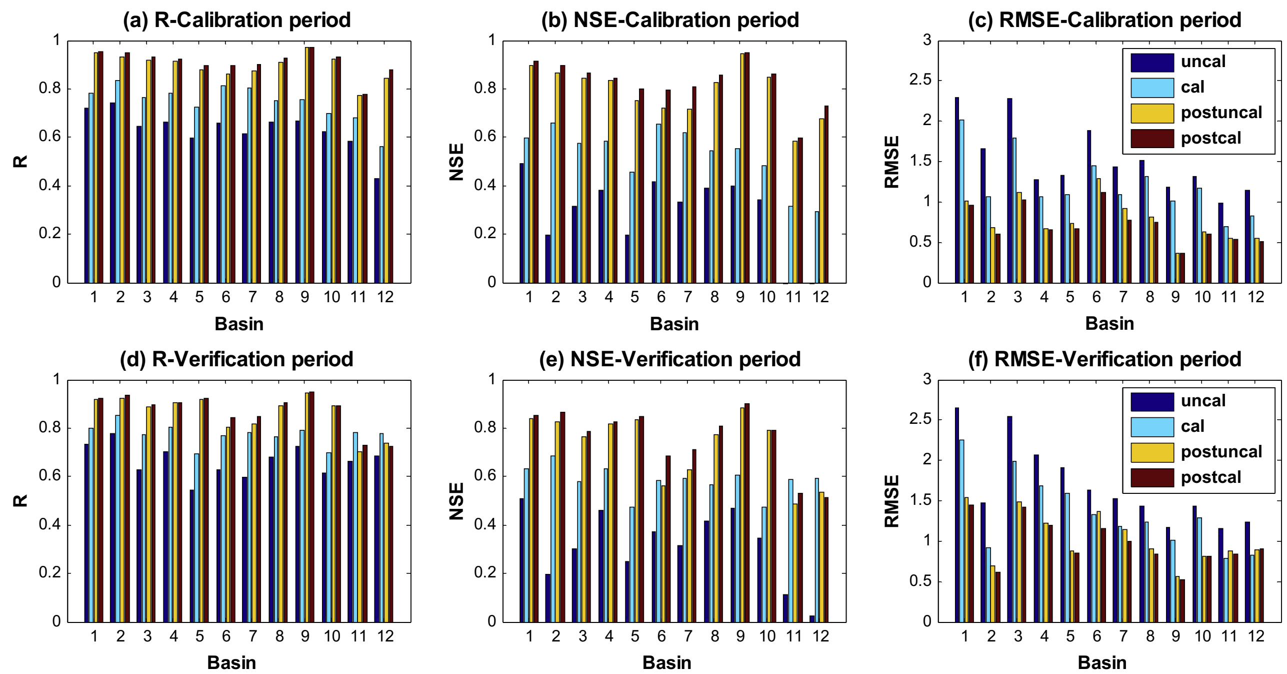

Fig. 6. Performance indices of raw and post-processed streamflow simulations averaged over 7 models. Top row denotes indices for calibration period, while lower row denotes indices for verification period. Columns 1 to 3 denote correlation coefficient, NS efficiency and RMSE, respectively.

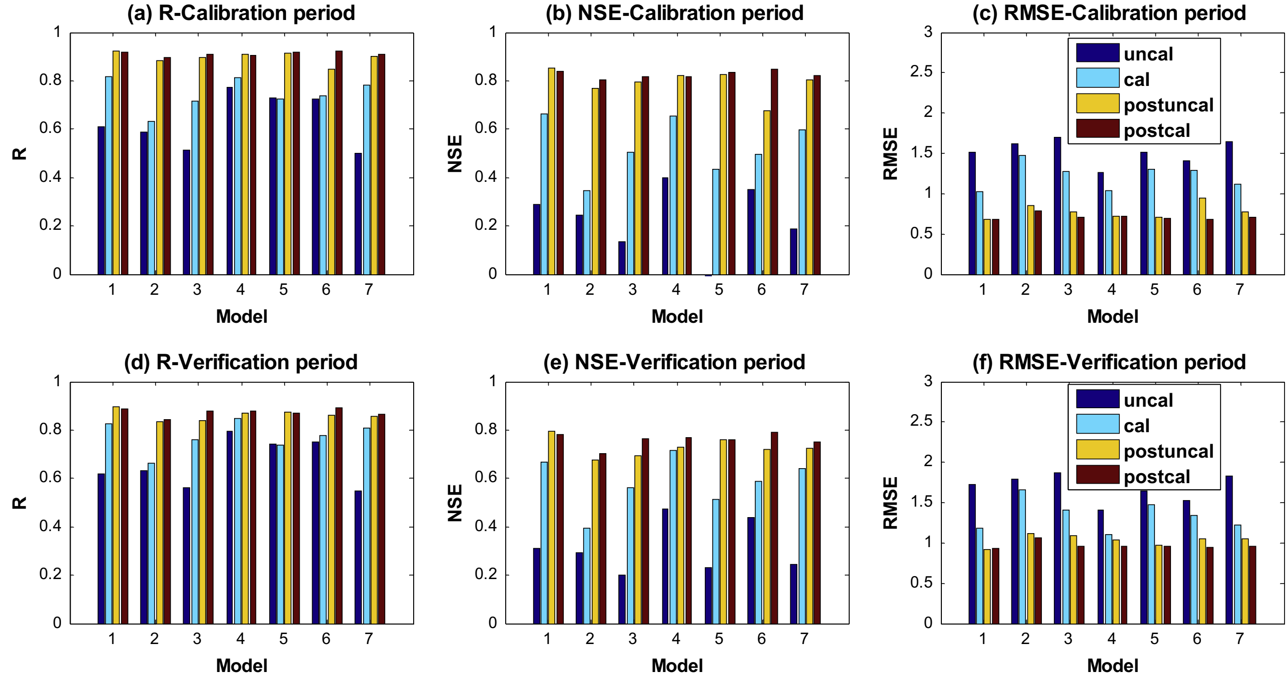

Fig. 7. Performance indices of raw and post-processed streamflow simulations averaged over 12 basins. Top row denotes indices for calibration period, while lower row denotes indices for verification period. Columns 1 to 3 denote correlation coefficient, NS efficiency and RMSE, respectively.

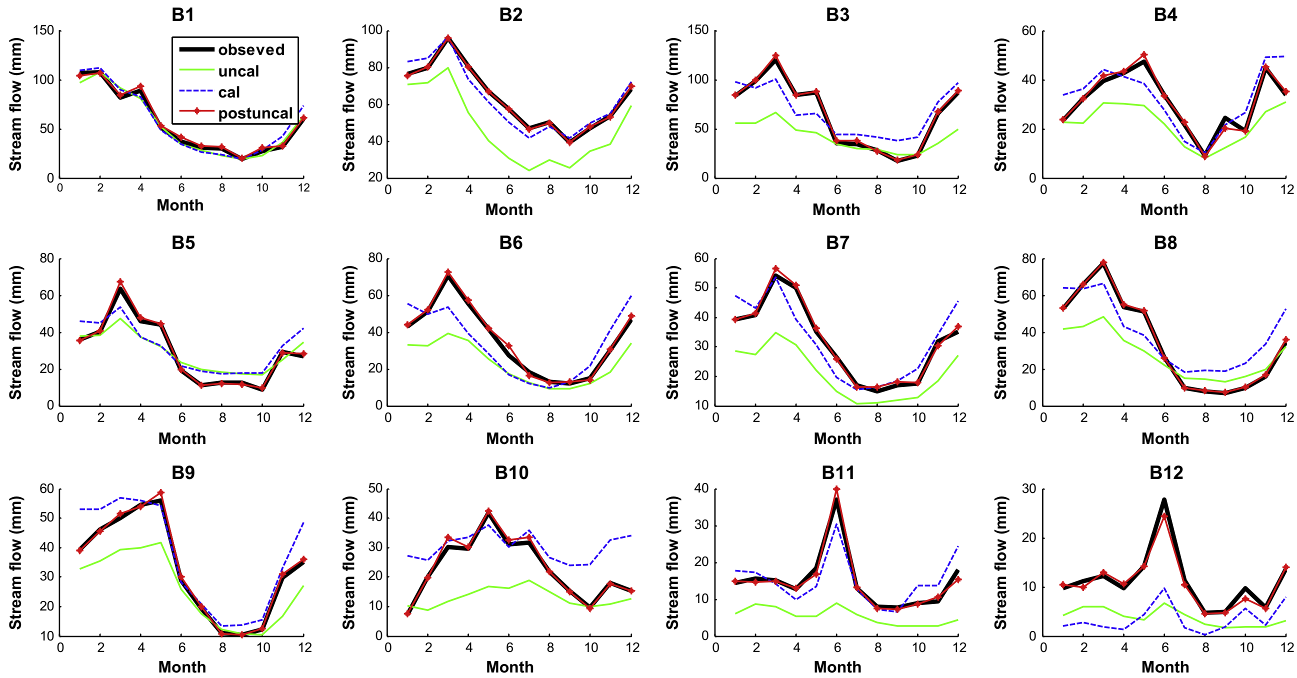

Fig. 8. Long-term average monthly hydrographs for the verification period for the 12 river basins by the Gr4j model.

the calibrated case are almost the same). The results in Fig. 8 indicate that the post-processing is quite effective in removing systematic biases for all basins and for all different seasons.

4.3. Evaluation of ensemble spread of the post-processed streamflow values

Once GLMPP is calibrated, we can generate post-processed ensemble streamflow values given the raw streamflow simulations using a Monte Carlo approach. For each model and each basin, 50 ensembles are produced. The rank histogram is used to evaluate whether these ensembles include the observations being predicted as equi-probable members. A perfect histogram of ranks will be uniform (Wilks, 2011; Yuan and Wood, 2012). We plotted the rank histograms of post-processed ensemble streamflow simulations for each basin for Gr4j model in Fig. 9 (other models exhibit similar results and are not shown here). There are 51 ranks ordered from small to large. The rank value is the frequency of observed values falling within the rank interval. We found that there are obvious differences in the ranked histograms for different basins. There

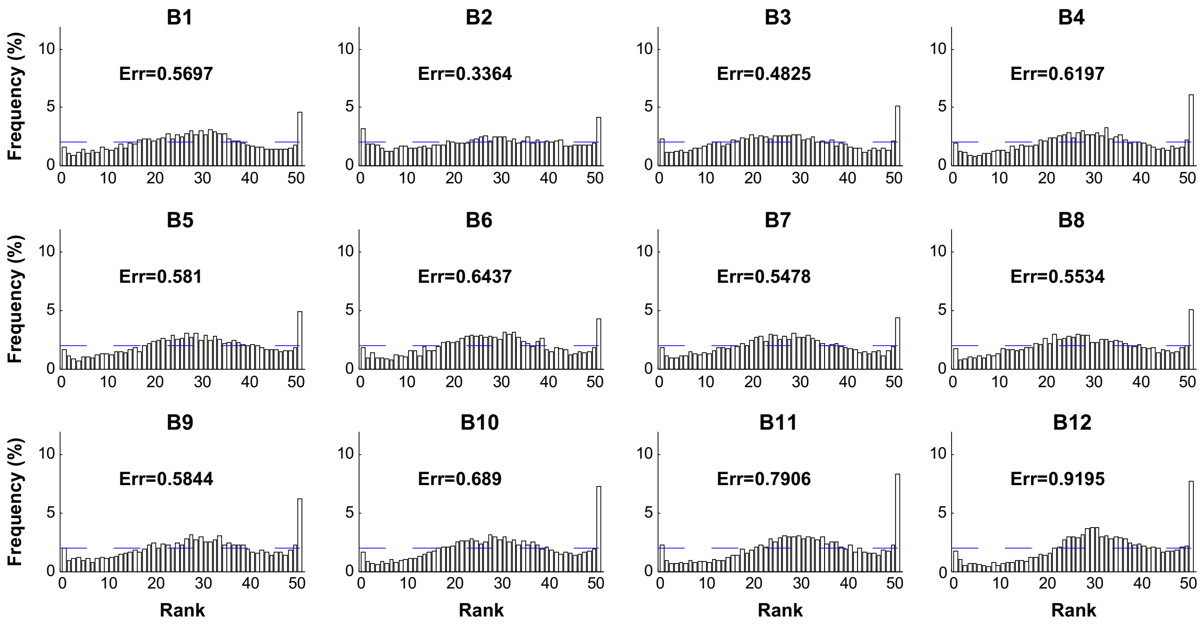

Fig. 9. Rank histograms for post-processed ensemble streamflow simulations in day 1 for the verification period by the Gr4j model. The horizontal dashed lines indicate perfect uniformity.

are three major features: (1) the result is reasonable and the ranks are uniform (i.e., Basin 2); (2) dome-shaped rank histograms with the peaks located in the centers seem to suggest certain under-confidence in the ensemble simulations (all basins except Basin 2); and (3) the right-most ranks are high (all basins except Basin 2), indicating that a significant number of ensemble members exceed the ensemble upper uncertainty range. These results suggest that the spread of ensemble might be further improved.

The predictive QQ plots of the post-processed ensemble simulations are shown in Fig. 10. It is clear from this figure that the curves closely follow the bisector lines for all basins. This indicates that the predictive distributions of post-processed ensemble streamflow simulations are reliable. The reliability indices α and ξ are shown in Table 3. We note that the values of α are close to zero and values of ξ are close to one, further demonstrating that the predictive distributions of the post-processed ensemble streamflow simulations are reliable. The results above are based on an ensemble size of 50. We conducted additional test on the effect of different ensemble sizes and found that 50 is adequate to capture the uncertainty range. This because is that GLMPP generates ensemble members by sampling randomly from the Normal distribution.

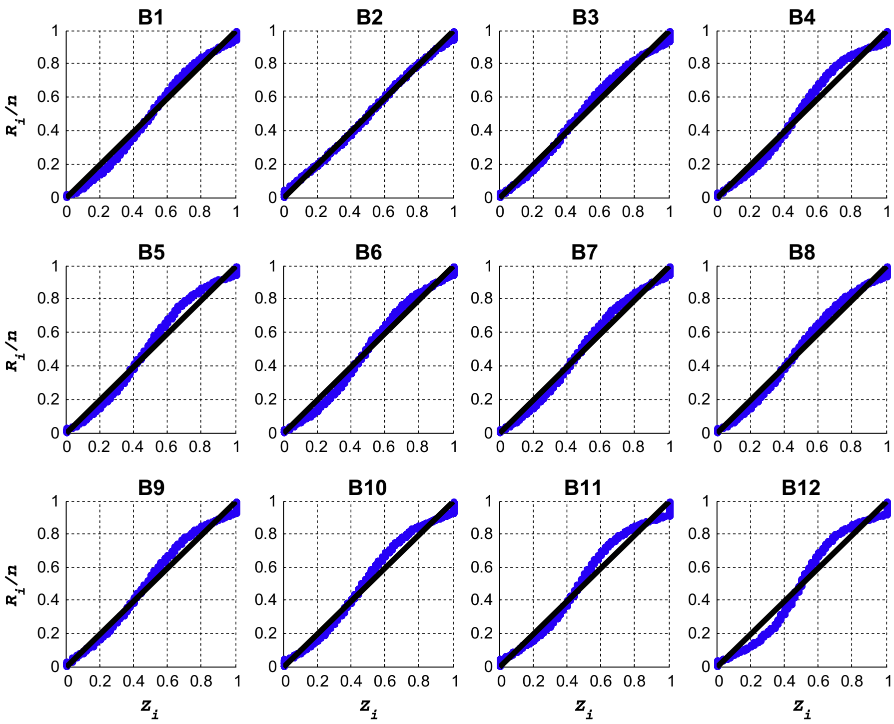

Fig. 10. Predictive QQ plot for the verification period (1981–1997) for the 12 river basins by the Gr4j model.

Table 3
The values of α and ξ for predictive distribution.

Model	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12
α												
grj4	0.03	0.01	0.03	0.04	0.04	0.04	0.03	0.03	0.03	0.04	0.04	0.06
isba	0.06	0.04	0.03	0.02	0.04	0.04	0.05	0.04	0.02	0.02	0.04	0.06
noa	0.03	0.02	0.03	0.05	0.03	0.04	0.04	0.03	0.04	0.04	0.03	0.04
sac	0.03	0.02	0.02	0.04	0.04	0.03	0.03	0.02	0.03	0.03	0.04	0.05
swap	0.03	0.01	0.02	0.04	0.03	0.04	0.04	0.03	0.03	0.03	0.04	0.05
swb	0.03	0.02	0.03	0.03	0.04	0.03	0.03	0.05	0.03	0.03	0.03	0.04
vic	0.03	0.01	0.02	0.03	0.03	0.02	0.02	0.03	0.03	0.03	0.02	0.05
ξ												
grj4	0.92	0.90	0.90	0.90	0.92	0.92	0.92	0.92	0.90	0.88	0.87	0.89
isba	0.89	0.85	0.88	0.90	0.91	0.88	0.88	0.88	0.88	0.86	0.84	0.87
noa	0.90	0.87	0.89	0.91	0.90	0.92	0.90	0.88	0.90	0.88	0.85	0.88
sac	0.91	0.89	0.90	0.90	0.92	0.92	0.91	0.90	0.90	0.88	0.89	0.90
swap	0.92	0.90	0.90	0.90	0.92	0.91	0.92	0.91	0.89	0.87	0.88	0.91
swb	0.90	0.89	0.88	0.88	0.88	0.89	0.88	0.81	0.88	0.85	0.83	0.85
vic	0.90	0.88	0.89	0.89	0.91	0.91	0.90	0.89	0.89	0.86	0.85	0.89

5. Conclusions

We conducted a comprehensive evaluation of GLMPP on seven different hydrologic models using streamflow simulation data from twelve basins from the Second Workshop on MOPEX. We found that GLMPP can effectively reduce the mean biases in streamflow simulations in both the calibration and verification periods. GLMPP post-processed streamflow simulations are generally better than that obtained through model calibration alone. The performance indices for post-processed streamflow simulations are similar for different hydrologic models in the same basin but are different among basins. The performance indices for post-processed streamflow simulations were best for the wet basins. In dry (semi-humid) basins, the performance indices for post-processed streamflow simulations in the calibration period are better than those in the verification period. This suggests a deficiency in the approaches that make use of NQT to achieve the Normality for the data for arid or semi-arid regions and alternative statistical techniques may be required to better understand the predictive uncertainty conditioned on model predictions.

GLMPP is used in this study to post-processing streamflow simulations that are generated using observed meteorological forcing data. It can be modified to post-processing hydrologic predictions generated using meteorological predictions, assuming these predictions are well calibrated to get rid of biases in both the predictive means and spreads. GLMPP can be generalized to multi-model ensemble predictions by extending the GLM approach to a multivariate setting.

Acknowledgments

This study was supported by National Science and Technology support plan Program (No. 2013BAB05B04) and the CRSRI Open Research Program (Program: CKWV2012324/KY). We further thank two anonymous reviewers, Andras Bardossy, Editor-in-Chief, Ezio Todini, Associate Editor for their valuable comments which improved the article significantly.

References

Andreadis, K.M., Lettenmaier, D.P., 2006. Assimilating remotely sensed snow observations into a macroscale hydrology model. Advances in Water Resources 29, 872–886.
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.
Bogner, K., Kalas, M., 2008. Error-correction methods and evaluation of an ensemble based hydrological forecasting system for the Upper Danube catchment. Atmospheric Science Letters 9, 95–102.
Cherubini, T., Ghelli, A., Lalaurette, F., 2002. Verification of precipitation forecasts over the alpine region using a high-density observing network. Weather and Forecasting 17, 238–249.
Clark, M.P., Slater, A.G., 2006. Probabilistic quantitative precipitation estimation in complex terrain. Journal of Hydrometeorology 7, 3–22.
Coccia, G., Todini, E., 2011. Recent developments in predictive uncertainty assessment based on the model conditional processor approach. Hydrology and Earth System Sciences 15, 3253–3274.
Duan, Q., Sorooshian, S., Gupta, V., 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resources Research 28, 1015–1031.
Duan, Q., Sorooshian, S., Gupta, V.K., 1994. Optimal use of the SCE-UA global optimization method for calibrating watershed models. Journal of Hydrology 158, 265–284.
Duan, Q., Schaake, J., Andréassian, V., et al., 2006. Model Parameter Estimation Experiment (MOPEX): an overview of science strategy and major results from the second and third workshops. Journal of Hydrology 320, 3–17.
Glahn, B., Peroutka, M., Wiedenfeld, J., et al., 2008. MOS uncertainty estimates in an ensemble framework. Monthly Weather Review 137, 246–268.
Krzysztofowicz, R., Herr, H.D., 2001. Hydrologic uncertainty processor for probabilistic river stage forecasting: precipitation-dependent model. Journal of Hydrology 249, 46–68.
Krzysztofowicz, R., Kelly, K.S., 2000. Hydrologic uncertainty processor for probabilistic river stage forecasting. Water Resources Research 36, 3265–3277.
Krzysztofowicz, R., Maranzano, C.J., 2004. Hydrologic uncertainty processor for probabilistic stage transition forecasting. Journal of Hydrology 293, 57–73.
Laio, F., Tamea, S., 2007. Verification tools for probabilistic forecasts of continuous hydrological variables. Hydrology and Earth System Sciences 11, 1267–1277.
Montanari, A., Brath, A., 2004. A stochastic approach for assessing the uncertainty of rainfall-runoff simulations. Water Resources Research 40, W01106.
Montanari, A., Grossi, G., 2008. Estimating the uncertainty of hydrological forecasts: a statistical approach. Water Resources Research 44, W00B8.
Renard, B., Kavetski, D., Kuczera, G., Thyer, M., Franks, S.W., 2010. Understanding predictive uncertainty in hydrologic modeling: the challenge of identifying input and structural errors. Water Resources Research 46, W05521.
Schaake, J., Demargne, J., Hartman, R., et al., 2007. Precipitation and temperature ensemble forecasts from single-value forecasts. Hydrology and Earth System Sciences Discussions Discussions 4, 655–717.
Seo, D.J., Perica, S., Welles, E., Schaake, J.C., 2000. Simulation of precipitation fields from probabilistic quantitative precipitation forecast. Journal of Hydrology 239, 203–229.
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 Discussions 3, 1987–2035.
Shi, X., Wood, A.W., Lettenmaier, D.P., 2008. How essential is hydrologic model calibration to seasonal streamflow forecasting? Journal of Hydrometeorology 9, 1350–1363.
Slater, A.G., Clark, M.P., 2006. Snow data assimilation via an ensemble kalman filter. Journal of Hydrometeorology 7, 478–493.
Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, SW., Srikanthan, S., 2009. Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: a case study using Bayesian total error analysis. Water Resources Research 45, W00B14, http://dx.doi.org/10.1029/2008wr006825.
Todini, E., 2008. A model conditional processor to assess predictive uncertainty in flood forecasting. International Journal of River Basin Management 6, 123–137.
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, 255–265.
Wilks, D.S., 2011. Statistical Methods in the Atmospheric Sciences, third ed. Academic Press.

Wood, A.W., Schaake, J.C., 2008. Correcting errors in streamflow forecast ensemble mean and spread. Journal of Hydrometeorology 9, 132–148.
Xie, P., Chen, M., Yang, S., et al., 2007. A gauge-based analysis of daily precipitation over East Asia. Journal of Hydrometeorology 8, 607–626.
Yuan, X., Wood, E.F., 2012. Downscaling precipitation or bias-correcting streamflow? Some implications for coupled general circulation model (CGCM)-based ensemble seasonal hydrologic forecast. Water Resources Research 48, W12519. http://dx.doi.org/10.1029/2012 WR012256.
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.
Zhou, T., Yu, R., Chen, H., Dai, A., Pan, Y., 2008. Summer precipitation frequency, intensity, and diurnal cycle over China: a comparison of satellite data with rain gauge observations. Journal of Climate 21, 3997–4010.