

Post-processing of ensemble forecasts in low-flow period

Aizhong Ye,^{1,2*} Qingyun Duan,^{1,2} John Schaake,³ Jing Xu,^{1,2} Xiaoxue Deng,^{1,2} Zhenhua Di,^{1,2}
Chiyuan Miao^{1,2} and Wei Gong^{1,2}

¹ State Key Laboratory of Earth Surface Processes and Resource Ecology, College of Global Change and Earth System Science, Beijing Normal University, Beijing, 100875, China

² Joint Center for Global Change Studies, Beijing, 100875, China

³ Consultant, Annapolis, MD, USA

Abstract:

For water supply, navigational, ecological protection or water quality control purposes, there is a great need in knowing the likelihood of the river level falling below a certain threshold. Ensemble streamflow prediction (ESP) based on simulations of deterministic hydrologic models is widely used to assess this likelihood. Raw ESP results can be biased in both the ensemble means and the spreads. In this study, we applied a modified general linear model post-processor (GLMPP) to correct these biases. The modified GLMPP is built on the basis of regression of simulated and observed streamflow calculated on the basis of canonical events, instead of the daily values as is carried out in the original GLMPP. We conducted the probabilistic analysis of post-processed ESP results falling below pre-specified low-flow levels at seasonal time scale. Raw ESP forecasts from the 1980 to 2006 periods by four different land surface models (LSMs) in eight large river basins in the continental USA are included in the analysis. The four LSMs are Noah, Mosaic, variable infiltration capacity and Sacramento models. The major results from this study are as follows: (1) a modified GLMPP was proposed on the basis of canonical events; (2) post-processing can improve the accuracy and reduce the uncertainty of hydrologic forecasts; (3) post-processing can help deal with the effect of human activity; and (4) raw simulation results from different models vary greatly in different basins. However, post-processing can always remove model biases under different conditions. Copyright © 2014 John Wiley & Sons, Ltd.

KEY WORDS low flow; ensemble forecast; post-processor; NLDAS data

Received 9 June 2014; Accepted 28 September 2014

INTRODUCTION

Low flow is the minimum flow in a river during the dry periods of the year (Smakhtin, 2001). The international glossary of hydrology (WMO, 1974) defines low flow as ‘flow of water in a stream during prolonged dry weather’. Low flows are normally derived from groundwater discharge or surface discharge from lakes, marshes or melting glaciers (Smakhtin, 2001). For water supply, navigational, ecological protection or water quality control purposes, there is a great need in knowing the likelihood of the river level falling below certain low-flow thresholds. Ensemble streamflow prediction (ESP) is a commonly used approach by hydrologic forecasters to assess this likelihood of future streamflow magnitude (Franz *et al.*, 2003). ESP creates a probabilistic outlook of future streamflow levels for lead times ranging from a few days to a few seasons, based on streamflow simulations from a deterministic hydrologic modelling system that is

run with multiple meteorological forcing inputs and current basin conditions (Day, 1985).

Because of uncertainties in meteorological forcing, initial conditions and hydrologic model structure and parameters, raw ESP results can be biased in both the ensemble means and the spreads. This is particularly true for low flows as most hydrologic models are developed with limited interest in replicating low flows well (Shaw and Riha, 2012). Streamflow forecast performance measures, such as the Nash–Sutcliffe efficiency (NSE) value or the Root-Mean-Square-Error (RMSE), are heavily weighted towards predictions of high flows than low flows, i.e. if the model simulates flood peaks well but low flows poorly, the NSE or RMSE values would still be good (Pushpalatha *et al.*, 2012). Low flow predictions can also be complicated by human activities that alter natural streamflow via reservoir regulation or water diversion for irrigation (Döll and Siebert, 2002; Sperna Weiland *et al.*, 2010; Wisser *et al.*, 2010; Falloon *et al.*, 2011). The effects of the reservoir manifest in simulated hydrographs that will have a tendency to over-predict high flows and under-predict low flows, making ESP streamflow simulations incomparable to observations.

*Correspondence to: Aizhong Ye, State Key Laboratory of Earth Surface Processes and Resource Ecology, College of Global Change and Earth System Science, Beijing Normal University, Beijing, 100875, China.
E-mail: azye@bnu.edu.cn

For ESP streamflow predictions to be meaningful, it is necessary to remove their inherent biases. There are a number of approaches to improve raw ESP streamflow predictions. One approach is to improve meteorological observations and forecasts that are used to drive the hydrologic models (Schaake *et al.*, 2006). Past research has also indicated that better initial conditions account for much of the seasonal streamflow predictability in ESP, especially for snow-dominated basins (Shukla *et al.*, 2013). Another approach is, therefore, to improve the initial conditions in the hydrologic model through better assimilation of observations (Slater and Clark, 2006). The third approach is to improve hydrologic model structure and to develop better parameter estimation methods (Duan *et al.*, 2006; Singh and Frevert, 2006). Multimodel ensemble strategy has also been promoted recently to account for model structural uncertainty (Ajami *et al.*, 2007; Duan *et al.*, 2007). The fourth approach is to perform hydrologic post-processing of the hydrologic model outputs (Brown and Seo, 2010; Brown and Seo, 2013; Madadgar *et al.*, 2014). Hydrologic post-processing addresses uncertainties of hydrologic model outputs by using a statistical model to represent the relationship between model outputs and corresponding observations (Krzysztofowicz and Kelly, 2000; Seo *et al.*, 2006). It can remove or reduce model systematic biases from all upstream uncertainty sources and can be regarded as the final step before the issue of actual hydrologic forecasts (Shi *et al.*, 2008; Yuan and Wood, 2012).

The methods of post-processing can be divided into three categories. One approach is a simple bias correction method, such as the degree of mass balance (DMB) (McCullor and Stull, 2008) and the linearly weighted DMB (Bourdin and Stull, 2013). Another approach is using probability calibration methods based on Bayesian framework (Todini, 2008; Brown and Seo, 2010; Coccia and Todini, 2011; Brown and Seo, 2013). The quantile mapping approach is another technique for post-processing, which maps the probability distribution of ensemble predictions with the observed frequency (Shi *et al.*, 2008; Madadgar *et al.*, 2014).

Zhao *et al.* (2011) developed a generalized linear model-based approach (general linear model post-processor, GLMPP) and compared the post-processing results with calibration results of hydrological models. The idea of GLMPP is similar to the quantile mapping approach, but in detail, it is different. GLMPP is a linear regression analysis of the NQT transformed input–output pairs. They demonstrated the effectiveness of GLMPP using the streamflow simulation results from one basin in the Model Parameter Estimation Experiment (MOPEX) database, where MOPEX stands for Model Parameter Estimation Experiment (Duan *et al.*, 2006), and reported very promising results (Zhao *et al.*, 2011). Ye *et al.* (2014) performed a comprehensive evaluation of the effectiveness of GLMPP in

improving the MOPEX streamflow simulations issued by seven different models and for 12 MOPEX basins and found that GLMPP can improve streamflow simulations significantly, especially in wet and temperate basins. However, they also found that performance of GLMPP in dry basins might not be good. This is probably because in dry basins, streamflow can have a strong seasonal signature and there may be many days with zero streamflow values. GLMPP may fail in these cases because the underlying probability distributions of the considered variables are truncated because of the presence of many zero values or the monthly streamflow in the forecast period. Another issue with the original GLMPP is that the general linear model is built on regression of simulated and observed streamflow daily values no matter how long the forecast lead time is. In seasonal streamflow predictions, forecasters are more concerned about the predictability of average monthly or seasonal values than the exact values on specific days during the season. For this reason, this paper proposes a modified GLMPP that is built on the basis of regression of simulated and observed streamflow values calculated on the basis of canonical events (CE), which can be daily, monthly or seasonal events. CEs correspond to streamflow events with specific lead times and durations. For example, a CE can be the average daily streamflow from day 6 to day 10 or monthly streamflow in the forecast period.

This paper systematically compares the performance of the post-processed and raw-simulated streamflow simulations in an ESP setup for four models and eight different large river basins in the USA. The organization of the paper is as follows: we have the Method Section, which describes the GLMPP based on canonical events (GLMPP-CE); we then introduce the Data and Study Domain Section, present the Post-processing Results and Discussion Section and finally, provide the Conclusions Section.

METHOD

General linear model post-processor based on canonical events

The GLMPP is a hydrological post-processor used to predict future streamflow observations based on raw streamflow predictions (Zhao *et al.*, 2011; Ye *et al.*, 2014). GLMPP possesses the following properties: (1) GLMPP removes bias (i.e. $|\overline{Q_s}/\overline{Q_o} - 1|$) in the streamflow simulations, $\overline{Q_s}$ and $\overline{Q_o}$ being the simulation and observation over the evaluation period, respectively; (2) GLMPP produces an ensemble of streamflow members representing, in an ‘equally-likely’ sense, the hydrograph being predicted; and (3) GLMPP preserves space–time variability to be consistent with observation.

The GLMPP is basically a multivariate linear regression model relating forecasted streamflow values to the

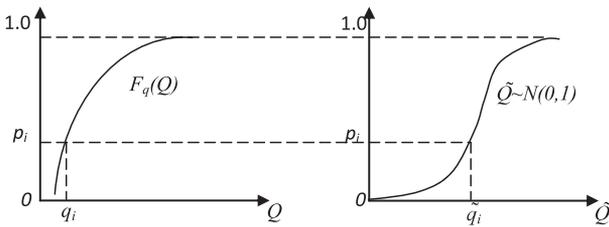


Figure 1. Schematic diagram for Normal Quantile Transform

observed values. The forecasted and observed streamflow values first go through the Normal Quantile Transform (NQT) (Krzysztofowicz, 1997; Bogner *et al.*, 2011) to ensure these variables are normally distributed.

The NQT is a procedure to ensure that considered variables (observed and simulated streamflow) are normally distributed. The NQT involves the following steps: (1) compute the cumulative frequency curve of the data set, $Q = \{q_1, q_1, \dots, q_M\}$ (Figure 1); (2) each data point q_i corresponds to a unique frequency value, p_i ; (3) replace q_i in Q with the corresponding value \tilde{q}_i , which is a Normal variate from a standard Normal distribution $N(0,1)$.

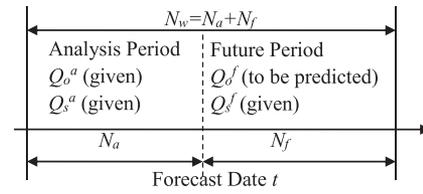
Given the forecast date N , the analysis period before this date and the forecast period after this date, the mathematical expression of GLMPP is expressed as follows:

$$Y = A \cdot X + B \cdot E \tag{1}$$

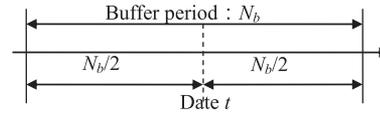
where $Y = [\tilde{Q}_o^f]$ is the predictand (i.e. the streamflow observations in the forecast period) and $X = [\tilde{Q}_s^f \ \tilde{Q}_o^a \ \tilde{Q}_s^a]^T$ is the predictor (i.e. streamflow forecasts in the forecast period and streamflow observations and simulations in the analysis period); $X = [\tilde{Q}_s^f \ \tilde{Q}_o^a \ \tilde{Q}_s^a]^T$ and $Y = [\tilde{Q}_o^f]$ are the NQT transformed values of $Q_2 = [Q_s^f, Q_o^a, Q_s^a]$, and $Q_1 = [Q_o^f]$, respectively. A and B are the undetermined coefficient matrices in the regression equation. $E \sim N(0,1)$ is a normally distributed random variable with a mean of zero and a standard deviation of 1. See Figure 2a for the schematic illustration of the forecast date, the analysis period and the forecast period.

To solve for the coefficient matrices A and B , we need to fill in the predictand and predictor vectors using N_y years of historical hindcast data and the corresponding observations:

$$\begin{aligned} Q_o^a &= [Q_{o,1}^a \ Q_{o,2}^a \ \dots \ Q_{o,i}^a \ \dots \ Q_{o,N_y}^a], \\ Q_s^a &= [Q_{s,1}^a \ Q_{s,2}^a \ \dots \ Q_{s,i}^a \ \dots \ Q_{s,N_y}^a], \\ Q_o^f &= [Q_{o,1}^f \ Q_{o,2}^f \ \dots \ Q_{o,i}^f \ \dots \ Q_{o,N_y}^f], \\ Q_s^f &= [Q_{s,1}^f \ Q_{s,2}^f \ \dots \ Q_{s,i}^f \ \dots \ Q_{s,N_y}^f], \end{aligned}$$



(a) A schematic of the forecast date, the analysis period and forecast period



(b) The schematic of the buffer period used to enlarge the sample size

Figure 2. The data window for the general linear model post-processor based on canonical events

where $Q_{o,i}^a$ and $Q_{s,i}^a$ are the observed and simulated streamflow in the analysis period; $Q_{o,i}^f$ and $Q_{s,i}^f$ are the observed and forecast streamflow in the forecast period; i (ranging from 1 to N_y) is the year indices; a means analysis period; f means forecast period; o means observation; and s means simulation. To enlarge the sample size to solve GLMPP, if we assume that the coefficient matrices A and B for the dates within a time window (called the buffer period, see Figure 2b) around the forecast date are similar (i.e. the data samples come from the same distribution), then, the full data matrices are shown as follows:

$$\begin{aligned} Q_{o,i}^a &= \begin{bmatrix} q_{i,N-\frac{N_b}{2}-N_a}^o & \dots & q_{i,N-N_a}^o & \dots & q_{i,N+\frac{N_b}{2}-N_a}^o \\ q_{i,N-\frac{N_b}{2}-N_a+1}^o & \dots & q_{i,N-N_a+1}^o & \dots & q_{i,N+\frac{N_b}{2}-N_a+1}^o \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ q_{i,N-\frac{N_b}{2}-1}^o & \dots & q_{i,N-1}^o & \dots & q_{i,N+\frac{N_b}{2}-1}^o \end{bmatrix}, \\ Q_{s,i}^a &= \begin{bmatrix} q_{i,N-\frac{N_b}{2}-N_a}^s & \dots & q_{i,N-N_a}^s & \dots & q_{i,N+\frac{N_b}{2}-N_a}^s \\ q_{i,N-\frac{N_b}{2}-N_a+1}^s & \dots & q_{i,N-N_a+1}^s & \dots & q_{i,N+\frac{N_b}{2}-N_a+1}^s \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ q_{i,N-\frac{N_b}{2}-1}^s & \dots & q_{i,N-1}^s & \dots & q_{i,N+\frac{N_b}{2}-1}^s \end{bmatrix}, \\ Q_{o,i}^f &= \begin{bmatrix} q_{i,N-\frac{N_b}{2}}^o & \dots & q_{i,N}^o & \dots & q_{i,N+\frac{N_b}{2}}^o \\ q_{i,N-\frac{N_b}{2}+1}^o & \dots & q_{i,N-N_a+1}^o & \dots & q_{i,N+\frac{N_b}{2}+1}^o \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ q_{i,N-\frac{N_b}{2}+N_f-1}^o & \dots & q_{i,N+N_f-1}^o & \dots & q_{i,N+\frac{N_b}{2}+N_f-1}^o \end{bmatrix}, \end{aligned}$$

$$Q_{s,i}^f = \begin{bmatrix} q_{i,N-\frac{N_b}{2}}^s & \cdots & q_{i,N}^s & \cdots & q_{i,N+\frac{N_b}{2}}^s \\ q_{i,N-\frac{N_b}{2}+1}^s & \cdots & q_{i,N-N_a+1}^s & \cdots & q_{i,N+\frac{N_b}{2}+1}^s \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ q_{i,N-\frac{N_b}{2}+N_f-1}^s & \cdots & q_{i,N+N_f-1}^s & \cdots & q_{i,N+\frac{N_b}{2}+N_f-1}^s \end{bmatrix},$$

where $q_{i,k}^o$ and $q_{i,k}^s$ are the observed and simulated daily streamflow (m^3/s) on day k in year i ; N is the present day from 1 to 365 in a year; N_a is the number of data points in the analysis period (i.e. the period prior to the day the forecast is made); and N_f is the number of data points in the forecast period (Figure 2). With the Gaussian assumption made for all variables in Equation (1), it can be easily solved by multivariate linear regression. To make the parameter estimates more robust, a ‘buffer’ period with a length of N_b (in days) is introduced to include more data pairs (i.e. Y and X pairs corresponding to days immediately before and after the forecast date considered) to enlarge the data sample size. The data pairs for $N_b/2$ days prior to the forecast date and $N_b/2$ days after the forecast date are included in the regression equation when solving GLMPP. A and B are coefficient matrices of the regression equation in the N day. If we denote Σ_{11} as the covariance matrix of Y , Σ_{22} as the covariance matrix of X , and Σ_{12} (or Σ_{21}) as the covariance matrices between Y and X (or X and Y), then we obtain $A = \Sigma_{12} \cdot \Sigma_{22}^{-1}$ and $BB^T = \Sigma_{11} - \Sigma_{12} \cdot \Sigma_{22}^{-1} \cdot \Sigma_{21}$. When we obtain A and B , then set $Q_{o,i}^a = [q_{i,N-N_a}^o \dots q_{i,N-1}^o]^T$, $Q_{s,i}^a = [q_{i,N-N_a}^s \dots q_{i,N-1}^s]^T$, $Q_{s,i}^f = [q_{i,N}^s \dots q_{i,N+N_f-1}^s]^T$, Q_o^a , Q_s^a , and Q_s^f , we can obtain the streamflow observations in the forecast period with Equation (1).

In the original GLMPP, \tilde{Q}_o^f , \tilde{Q}_s^f , \tilde{Q}_o^a , and \tilde{Q}_s^a are the daily streamflow values. In this study, we modify the original GLMPP and rename it as GLMPP-CE. In GLMPP-CE, the values of \tilde{Q}_o^f , \tilde{Q}_s^f , \tilde{Q}_o^a , and \tilde{Q}_s^a correspond to the values of CEs, which are defined as the average values of streamflow over a certain specific period. This modification is made for seasonal streamflow predictions, which have lead times of several months into the future. At this time scale, we are more interested in the streamflow values averaged or totalled over a period (e.g. 30 or 90 days) instead of the exact value on a particular date several months away from the forecast date. For instance, we are interested in the total seasonal streamflow volume flowing into a reservoir 6 months from now, but we have little interest in knowing the exact lowest flow on the 180th day in the future because we doubt the future lowest flow value.

For each day in a calendar year, we obtain A 's and B 's by solving Equation (1). Thus, 365 sets of A 's and B 's are obtained. Using the A 's and B 's and raw streamflow

simulations, we can generate streamflow ensembles by using the Monte Carlo method.

Figure 3 is the flow chart illustrating the procedure of GLMPP-CE. Given observed and simulated streamflow data, we first set the forecast date, choose the time window and define the CEs. The time window consists of an analysis period prior to the forecast date and a forecast period from the forecast date. Each CE has a start time (CE_b) and an end time (CE_e). The streamflow value for each CE is the average of observed or simulated streamflow over the period from CE_b and CE_e .

Determination of the low-flow period and low-flow probability forecast

Low-flow period (LFP) refers to the dry season. We use low-flow threshold (LFT) to determine LFP. LFP is the period when daily streamflow is less than LFT. LFT (m^3/s) is the streamflow when accumulated frequency is 0.2. LFT is determined as follows: (1) obtain the observed annual daily streamflow; (2) calculate the accumulated frequency observed of the daily streamflow; and (3) find the observed

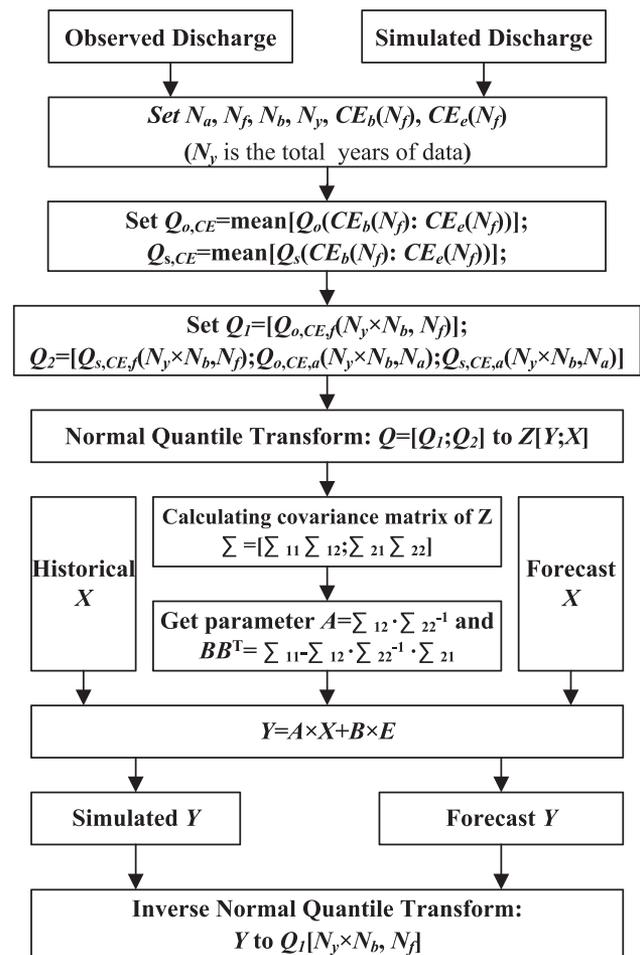


Figure 3. Flow chart for the post-processor on canonical events

daily streamflow, which corresponds to the accumulated frequency of 0.2.

We defined low-flow probability as the probability when the streamflow is less than LFT. GLMPP-CE can afford ensemble streamflow forecast, so we can obtain the post-process low-flow probability from ensemble forecast in each day. The deterministic simulated streamflow and observed streamflow are single value, the probability is 1 if the streamflow is less than LFT; otherwise, the probability is 0.

Model performance measures

To evaluate GLMPP-CE, we consider the following model performance measures: the *NSE* value calculated on inverse flows (Pushpalatha *et al.*, 2012), correlation coefficient (*R*), water balance bias (*Bias*) and RMSE. They are computed as follows:

$$NSE = \left[1 - \frac{\sum_{i=1}^N \left(\frac{1}{Q_{o,i} + \varepsilon} - \frac{1}{Q_{s,i} + \varepsilon} \right)^2}{\sum_{i=1}^N \left(\frac{1}{Q_{o,i} + \varepsilon} - \frac{1}{Q_{o,i} + \varepsilon} \right)^2} \right] \quad (2)$$

$$R = \frac{\sum_{i=1}^N (Q_{s,i} - \bar{Q}_{s,i})(Q_{o,i} - \bar{Q}_{o,i})}{\sqrt{\sum_{i=1}^N (Q_{s,i} - \bar{Q}_{s,i})^2 \sum_{i=1}^N (Q_{o,i} - \bar{Q}_{o,i})^2}} \quad (3)$$

$$Bias = \frac{\bar{Q}_s}{\bar{Q}_o} - 1 \quad (4)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Q_{s,i} - Q_{o,i})^2} \quad (5)$$

where $Q_o, Q_s, \bar{Q}_o, \bar{Q}_s$ are observed, simulated, average observed and average simulated streamflow discharges (m^3/s), ε is a small constant and N is the total number of time steps. A *NSE* score of 1 indicates perfect forecast of flow observations. Long-term mean of inverse flow

means *NSE* of 0. A negative *NSE* value suggests that the forecast is worse than the long-term average. A *RMSE* value of 0 indicates that forecasts are equal to flow observations. The squared difference is used in the calculation, so *RMSE* gives more weight to high than low values of errors.

In order to evaluate the reliability of the predictive distributions, the rank histogram is used here to diagnose whether the spread of the ensembles is reasonable for both the raw and the post-processed streamflow simulations. A perfect rank histogram would show observations evenly spread across equi-probability bins (Wilks, 2011; Yuan and Wood, 2012). The perfect rank value is 1 divided by ensemble members.

DATA AND STUDY DOMAIN

We used simulated daily streamflow data for eight large river basins (Table I) in the USA (see Figure 4 for location of the river basins), which are available at (<ftp://nomad6.ncep.noaa.gov/pub/raid2/wd20yx/nldas/Streamflow/>). These streamflow simulations are generated by four different land surface hydrological models in the North American Land Data Assimilation System (NLDAS) (Lohmann *et al.*, 2004; Mitchell *et al.*, 2004). The four models are Mosaic model, Noah model, Sacramento model (Sac) and variable infiltration capacity (VIC) model. The Noah model is the land surface model of the National Centers for Environmental Prediction (Chen *et al.*, 1996). The Mosaic land surface model is a surface-vegetation-atmosphere transfer scheme with a mosaic approach (Koster and Suarez, 1996). The VIC model is a macroscale hydrologic model that solves full water and energy balances, originally developed by Liang *et al.* (1996). The Sacramento model (SAC) is a conceptual rainfall-runoff model (Burnash *et al.*, 1973; Lohmann *et al.*, 2004). The simulated streamflow data covers the period from 1980 to 2006. Corresponding streamflow observations are available from the U.S. Geological Survey website (<http://waterdata.usgs.gov/usa/nwis/sw>). The eight river basins have different hydro-climatic conditions, and seven river basins of them are

Table I. River basin information

Basin	ID	Area (km ²)	Annual prec. (mm)	Long.	Lat.	Site name
B1	14105700	613827	477	-121.17	45.61	Columbia River at The Dalles, OR
B2	06934500	1353269	539	-91.44	38.71	Missouri River at Hermann, MO
B3	03611500	525768	1092	-88.74	37.15	Ohio River at Metropolis, IL
B4	07263450	409296	742	-92.36	34.79	Arkansas River at Murray Dam Near Little Rock, AR
B5	05474500	308209	798	-91.37	40.39	Mississippi River at Keokuk, IA
B6	01646502	29940	1006	-77.13	38.95	Potomac River (Adjusted) Near Washington, DC
B7	09421500	444701	351	-114.74	36.02	Colorado River Below Hoover Dam, AZ, NV
B8	02428400	55615	1359	-87.55	31.62	Alabama River at Claiborne L&D Near Monroeville, PA

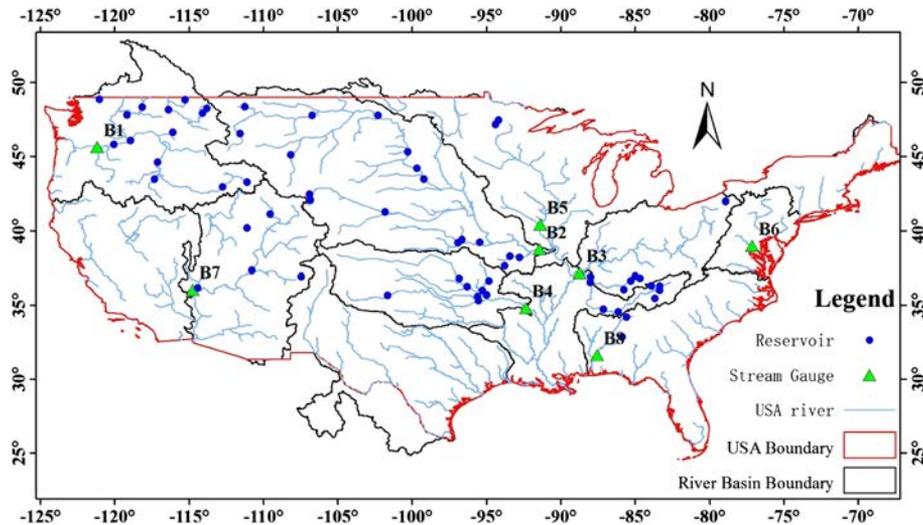


Figure 4. Spatial location of the streamflow gauges and reservoirs for the eight river basins

Table II. Large reservoirs information in eight basins. Capacity is referring to the total capacity of all large reservoirs in the basin

Basin	Number of reservoirs	Capacity (km ³)	Annual streamflow (m ³ /s)	Annual water (km ³)	Ratio (%)
B1	14	45.66	5042	159	29
B2	17	120.63	2578	81	148
B3	13	35.93	8325	263	14
B4	8	16.93	1350	43	40
B5	2	2.52	2278	72	4
B6	0	0.00	362	11	0
B7	5	76.18	429	14	564
B8	2	3.77	872	28	14

Annual streamflow is mean annual streamflow of the river. And annual water is the total annual streamflow volume of the river. Ratio = Capacity/Annual water. Ratio can show human activity impact. The natural streamflow is likely to alter if the Ratio is big.

regulated by large reservoirs (as marked by blue dots in Figure 4). The annual average precipitation is from 351 mm (B7) to 1359 mm (B8) (Cosgrove *et al.*, 2003). The annual average runoff is from 30 mm (B7) to 499 mm (B3). The land covers of eight basins are different. The land cover of B7 is desert and meadow, and the land cover of B3 is forest. The reservoir information for the eight basins is listed in Table II.

Determination of the low-flow period

The LFP can be determined from the observed daily streamflow during 1980 to 2006 using the method in the Method Section.

Table III lists the streamflow values corresponding to the accumulated frequency of 0.2. Figure 5(a) displays

Table III. Low-flow threshold (m³/s)

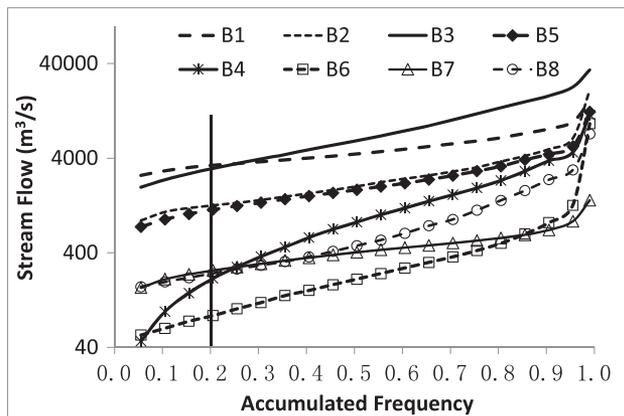
	B1	B2	B3	B4	B5	B6	B7	B8
All data	3370	1266	3115	214	1144	87	261	242
Low-flow period	2860	1104	2280	123	1025	61	214	191

observed daily streamflow frequency curves for the eight basins during 1980 to 2006. Figure 5(b) displays observed daily streamflow frequency curves for the eight basins in LFP during 1980 to 2006. Figure 6 shows the observed long-term average monthly streamflow for the eight basins. We note that the streamflow values during the period from 1 July to 1 October are almost always less than the LFTs for all basins except B7 (B7 on 1 October to 1 January). Accordingly, we set the dry season to the period from 1 July to 1 October for all basins except B7 (B7 on 1 October to 1 January). The LFTs from all data are used to determine whether it is dry season, and then the LFTs from only the LFP data are used to determine whether it is a low-flow event occurs in the dry season.

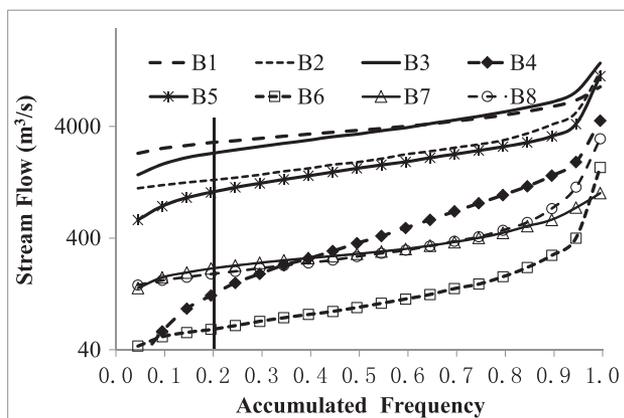
POST-PROCESSING RESULTS AND DISCUSSION

GLMPP-CE parameters

We used a set of GLMPP parameters suggested by Zhao *et al.* (2011). The length of the forecast period is set to 93 days. For the first two weeks, each CE corresponds to the daily streamflow value during that period. For the period from day 15 to day 31, day 32 to day 62 and day 63 to day 93, the average streamflow values for these periods are also treated as three separate CEs. The length of the analysis period, N_{as} , is set to 10 days, and the length



(a)



(b)

Figure 5. Observed daily streamflow frequency curves for the eight basins. (a) Whole year, (b) low-flow period (1 July to 1 October from 1980 to 2006, B7 on 1 October to 1 January)

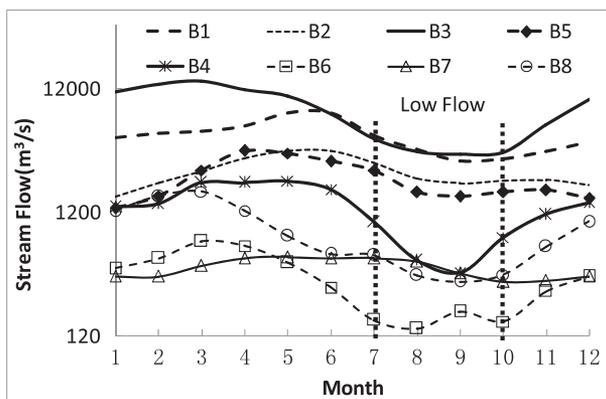


Figure 6. Long-term observed monthly streamflow averages for the eight basins

of the buffer period, N_b , is set to 15 days. There are seven days in the past, one day in the present and seven days in the future.

Comparing the long-term average streamflow simulations to the observed values

We analysed long-term average raw daily streamflow simulations in order to know the accuracy of the raw streamflow simulations. Table IV exhibited the long-term annual average observed streamflow discharges and the relative biases of the long-term annual average streamflow simulations from the four models in the eight river basins, respectively. From Table IV, it is clear that there are significant biases in the simulated streamflow for all models in all basins, with biases as large as 292%. The annual average streamflow over the eight basins shows that the Mosaic and Sac model are under-predicted and the Noah and VIC model are over-predicted (Table IV). Figure 7 shows the long-term observed (black dash line) and bias-corrected simulated daily streamflow from the four models for the eight basins. The bias-corrected method is to make average simulated streamflow equal to average observed streamflow. We note that there are still significant seasonal differences between streamflow simulations and observations for all models and all basins after systematic biases are removed. Particularly in the basins with large reservoir capacities (i.e. B2 and B7), the streamflow simulations from all models are generally larger than the observed values in the wet season, but the reverse is likely true in the dry season. This is probably due to the existence of reservoirs in these basins, which may decrease the streamflow out of the reservoirs during the wet season and increase the streamflow out of the reservoirs during the dry season. Because none of the models in NLDAS have explicit reservoir considerations, the difference between simulated and observed streamflow values are significant. In this study, we will investigate if post-processing can help to reduce the biases in the streamflow simulations in all basins and also implicitly consider the reservoir effect in basins such as B2 and B7.

Evaluating the accuracy of the raw and post-processed streamflow simulations

The post-processed ensemble forecast means are first compared with raw simulations of daily streamflow over different lead times, using observed streamflow. Figure 8 shows the correlation coefficients between the raw-simulated streamflow values and the observed streamflow values corresponding to different CEs (1–17) and different forecast dates in a calendar year for the four models and eight basins. The differences in correlation coefficients are compared between different models across all basins and seasons. The correlation coefficients for the Colorado River below Hoover Dam (basin 7) are especially low, probably due to the presence of Hoover Dam above the stream gauge. Figure 9 displays the

Table IV. Long-term annual average observed and simulated streamflow (m³/s) and Relative biases (%) from 1980 to 2006

ID	Observed streamflow	Simulated streamflow (m ³ /s)				Relative biases (%)			
		Mosaic	Noah	Sac	VIC	Mosaic	Noah	Sac	VIC
B1	5042	2844	7100	3548	5418	-44	41	-30	7
B2	2578	1192	5310	2334	4980	-54	106	-9	93
B3	8325	3814	8536	4804	9985	-54	3	-42	20
B4	1350	726	2281	1198	2883	-46	69	-11	113
B5	2278	972	4048	1126	3585	-57	78	-51	57
B6	362	125	429	195	519	-65	18	-46	43
B7	429	500	1680	713	902	17	292	66	110
B8	872	612	1066	705	1666	-30	22	-19	91
Average	2655	1348	3806	1828	3742	-42	79	-18	67

Sac, Sacramento model; VIC, variable infiltration capacity.

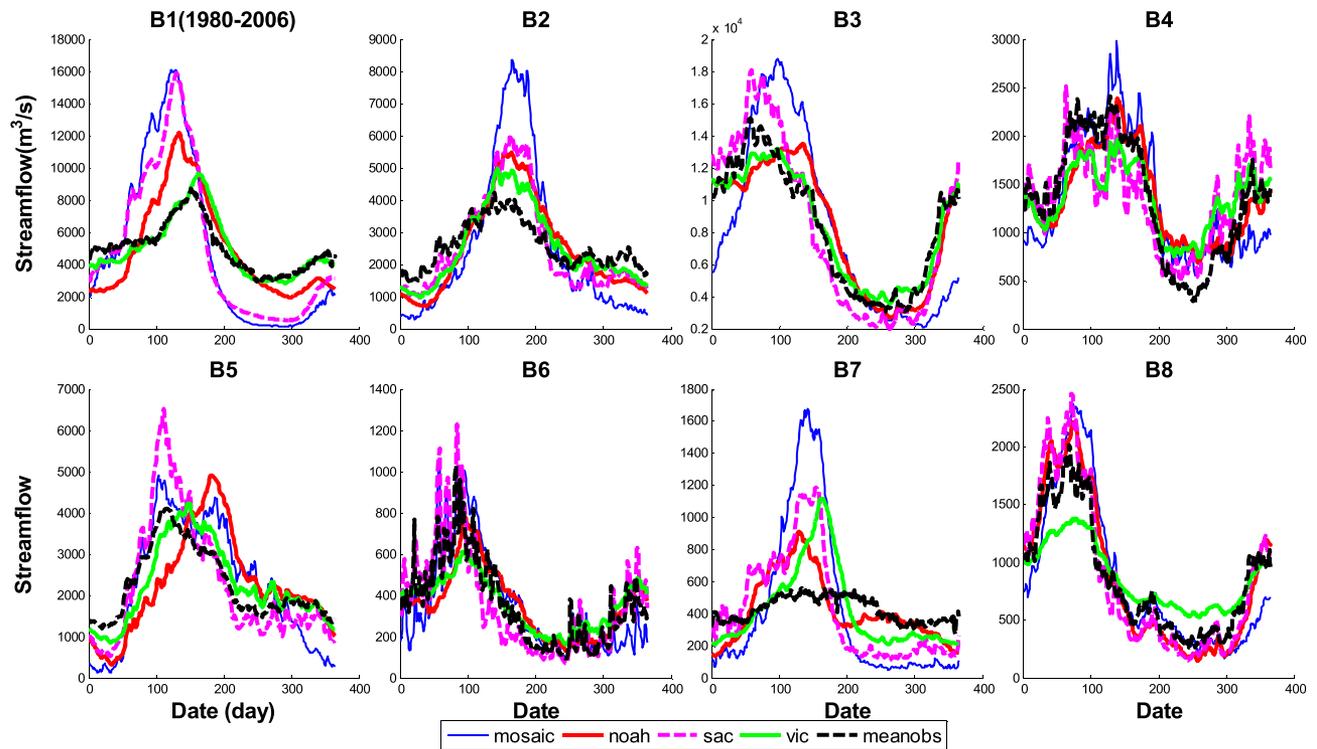


Figure 7. Observed and bias-corrected simulated daily streamflow climatology for the eight basins from 1980 to 2006

correlation coefficients between the post-processed simulated streamflow values and the observed streamflow values for the four models and the eight basins. Comparing Figures 8 and 9, it is clear that the correlation coefficients of the post-processed streamflow values are much better than those of the raw streamflow simulation values for all models and in all basins. The correlation coefficients are high for the events 15–17 in Figure 8 (exception of B7). In Figure 9, the coefficients are generally higher, so the post-processing has useful skills especially for lead times smaller than 2 weeks. Of particular note, the events 15–17 have lead times of

2 weeks or more. This suggests that the post-processed streamflow has useful skills for lead times up to 2 weeks in the future.

The results in Figure 9 were generated using GLMPP-CE with an analysis period of 10 days, i.e. $N_a=10$. To verify the importance of the analysis period, we also experimented with GLMPP-CE using no analysis period, i.e. $N_a=0$ days, and the results are shown in Figure 10. Figure 10 shows lower correlation coefficients when no analysis period is used, as compared with higher correlation coefficients displayed in Figure 9. This is particularly true for the streamflow events of the first

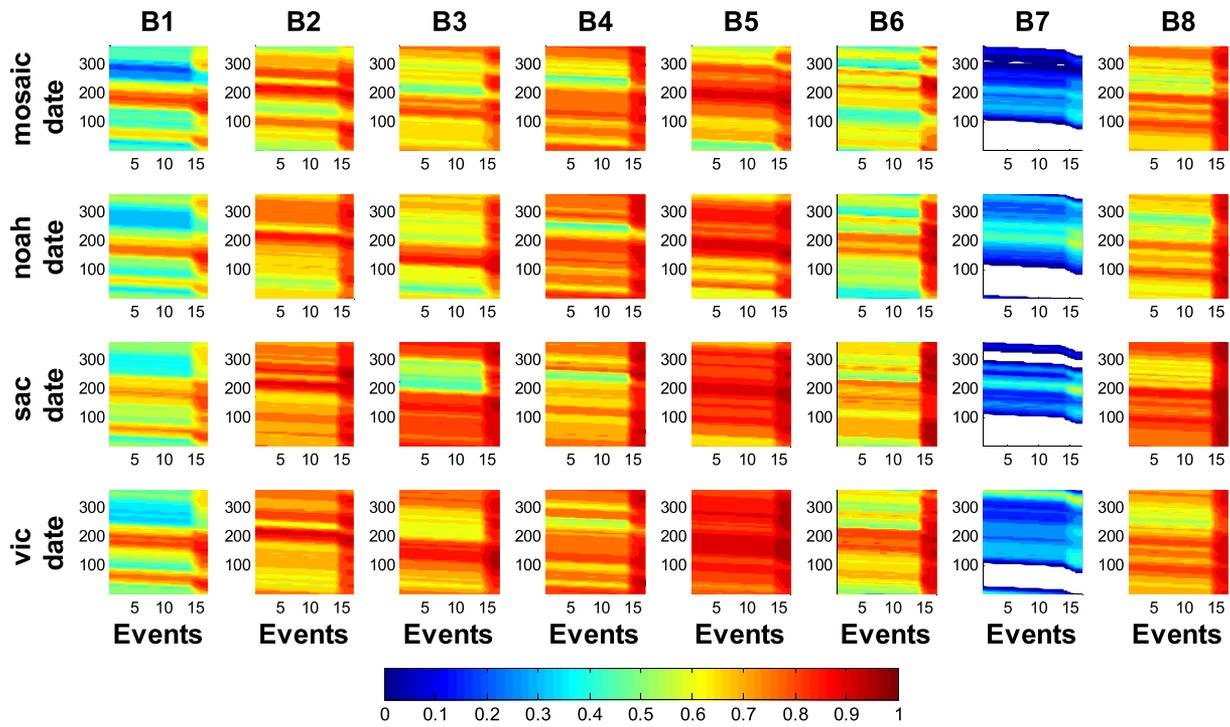


Figure 8. Correlation coefficients between the raw-simulated streamflow values and the observed streamflow values for the four models and the eight basins from 1980 to 2006. Horizontal axis denotes different canonical events (1–17), and vertical axis denotes the different dates in an annual calendar. The plot is empty if the correlation coefficients are less than zero

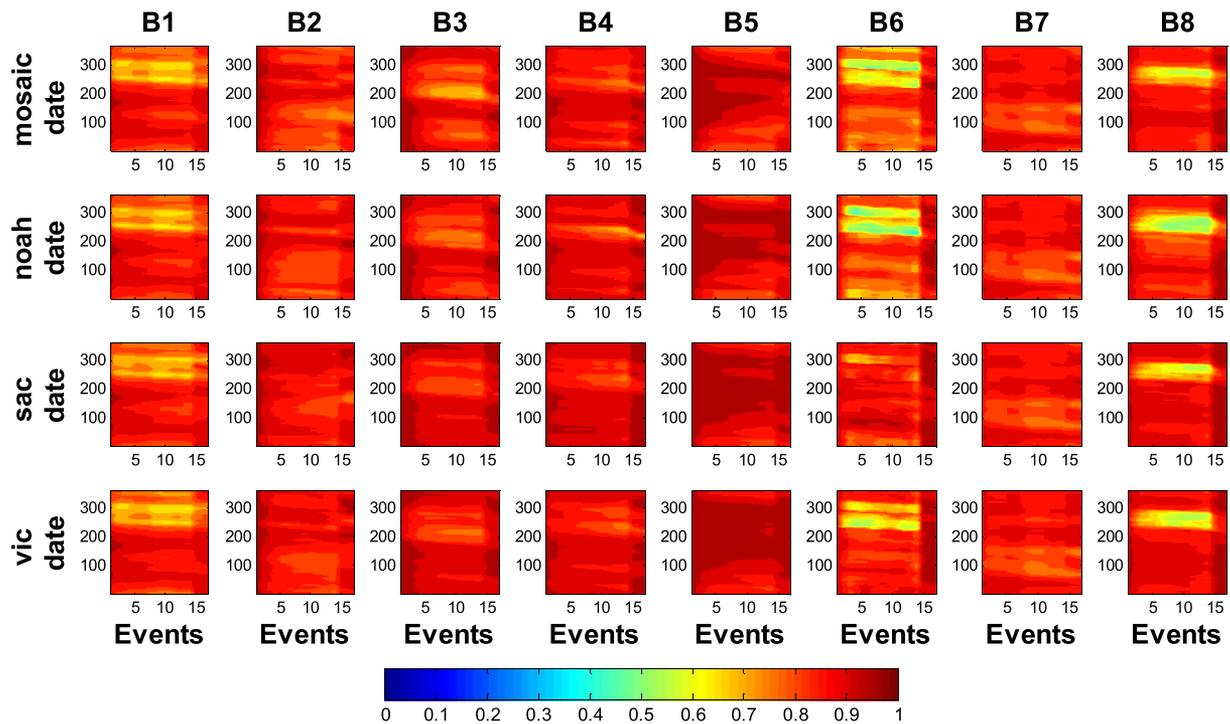


Figure 9. The correlation coefficients between the post-processed simulated streamflow values with an analysis period ($N_a = 10$ days) and the observed streamflow values for the four models and the eight basins from 1980 to 2006. Horizontal axis denotes different canonical events (1–17), and vertical axis denotes the different dates in an annual calendar

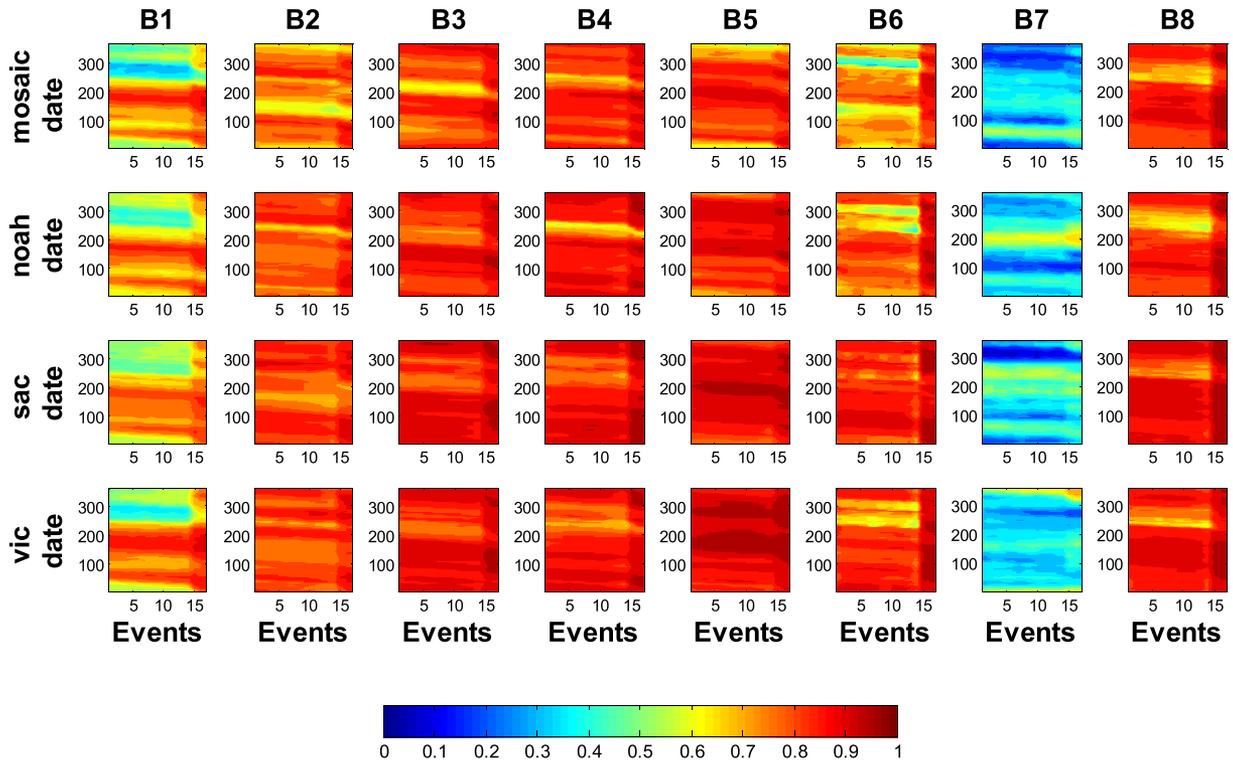


Figure 10. The correlation coefficients between the post-processed simulated streamflow values with no analysis period ($N_a=0$) and the observed streamflow values for the four models and the eight basins from 1980 to 2006. Horizontal axis denotes different canonical events (1–17) and vertical axis denotes the different dates in an annual calendar

14 days in the period, when initial conditions have big influence on streamflow predictions. For events with lead times longer than 2 weeks, the effects of initial conditions diminish and, therefore, the correlation coefficients for events 15–17 are not affected.

Evaluation of the performance of GLMPP-CE for predicting streamflow values in the dry season

In this section, we focus our attention to the performance of GLMPP-CE in the dry season, i.e. the forecast period from 1 July to 1 October. Figure 11 displays the means of the streamflow simulations made from 1 July to 1 October from 1980 to 2006 by the four models and for the eight basins (B7 on 1 October to 1 January). The lines include the means of the raw streamflow simulations (blue-dotted lines), the post-processed streamflow simulations (the red dash lines) and the observed streamflow simulations (black solid lines) in Figure 11. The raw results (NLDAS data) are always shown to deviate from the observed results, although the post-processed results always match the observed results well. The raw results from different models vary greatly in different basins, with Mosaic raw results generally smaller than the observed and Vic and Noah raw results generally bigger than the observed. The results from this experiment suggest that GLMPP-CE can effectively

remove systematic biases in the raw streamflow simulations in all cases.

If we examine the streamflow forecast for a particular year (say year 2000), we notice that a similar improvement of the post-processed streamflow simulations over the raw streamflow simulations is apparent (Figure 12).

To quantitatively evaluate the performance of the post-processed streamflow simulations by GLMPP-CE and the raw streamflow simulations, we calculated the four performance indices as described in Equations (2)–(5) for all models and all basins. Figure 13 only shows the Noah model results because the other models' results are similar to the results of Noah model. The performance indices from the raw simulations are denoted as 'raw' and from post-processed as 'post'. The basins in the following figures are indexed in the same order as shown in Table I. 'Red' indicates good performance and 'blue' implies bad performance for columns 1 (correlation coefficient), 2 (NSE) and 4 (RMSE). For column 3, 'green' is desired, i.e. bias equal to zero. Figure 13 clearly indicates that the performance indices for the post-processed streamflow simulations are superior to those of the raw streamflow simulations. The correlation coefficients of the post-processed streamflow simulations for all models in all basins are close to 1, better than the corresponding values for the raw streamflow simulations.

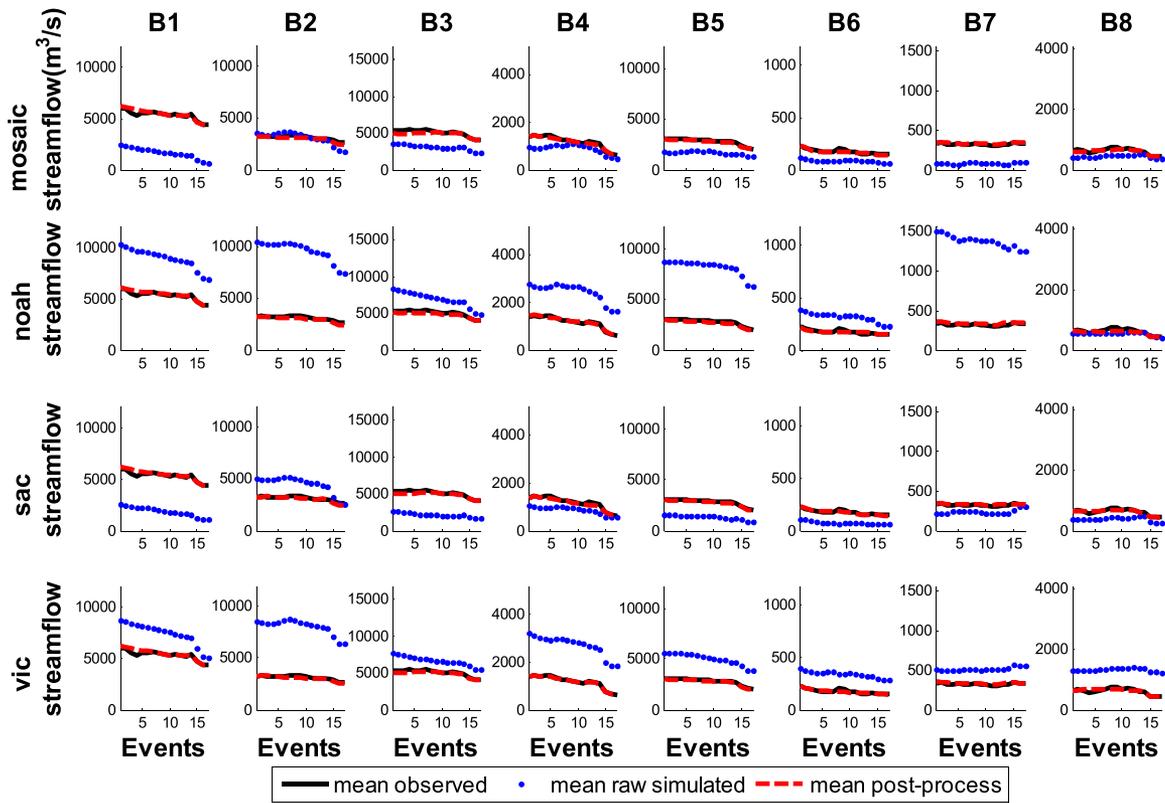


Figure 11. Long-term average daily streamflow in the seven river basins (B1–B8 except B7) on 1 July to 1 October and B7 on 1 October to 1 January from 1980 to 2006

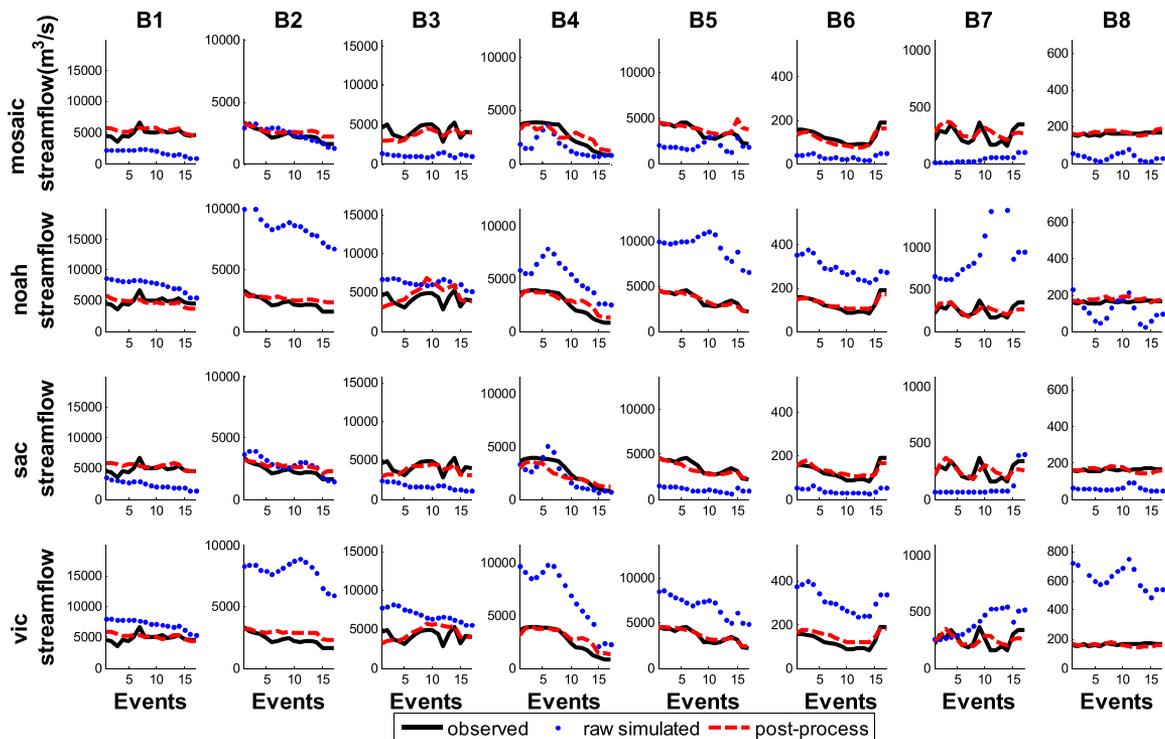


Figure 12. Daily hydrographs in the eight river basins on 1 July to 1 October 2000 (B7 on 1 October to 1 January)

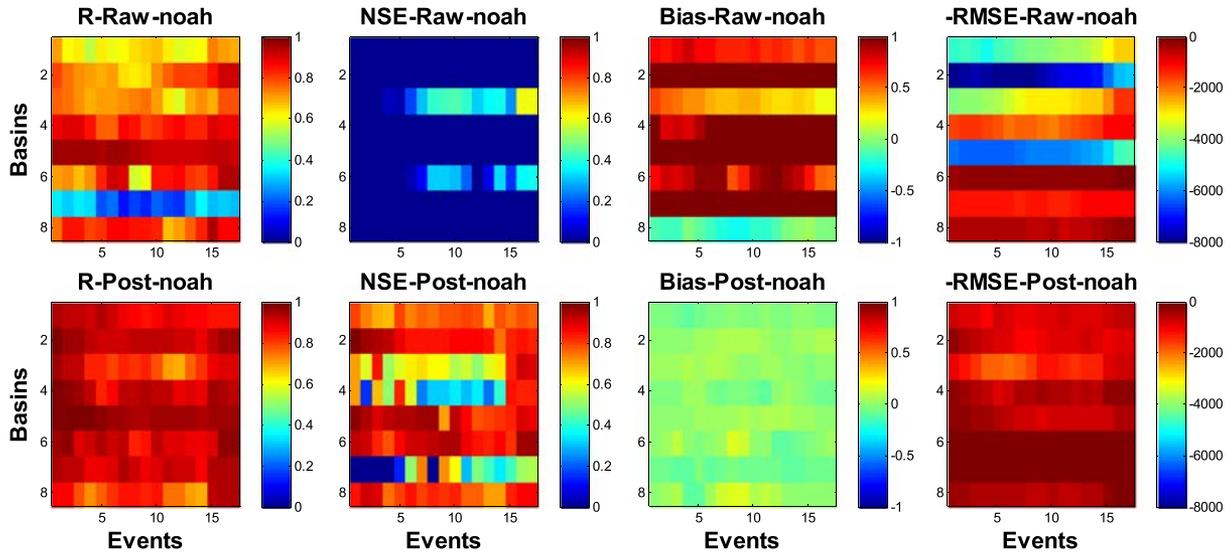


Figure 13. Performance indices of the raw and post-processed streamflow simulations for the dry period by Noah model from 1980 to 2006. The top row corresponds to the results for raw streamflow simulations, whereas the bottom row corresponds to the results for the post-processed streamflow simulations. RMSE unit is m^3/s

Interms of the NSE index, we notice that for most of the raw streamflow simulations, NSE indexes are less than 0.2 (blue), indicating that the raw simulations poorly matched with the observations. On the other hand, the post-processed NSE indexes are greater than 0.7 (red), suggesting their superior performance of post-processed streamflow simulation over the raw streamflow simulations. As for biases,

the values for the raw streamflow simulations can be either over-predicted (red) or close to zero (green, B8). For post-processed results, the biases are always close to zero (green). Dramatic improvement of the RMSE indices is also noticeable. Most of the raw streamflow simulations RMSE are greater than $1000 m^3/s$, whereas most of the post-processed RMSE are less than $1000 m^3/s$.

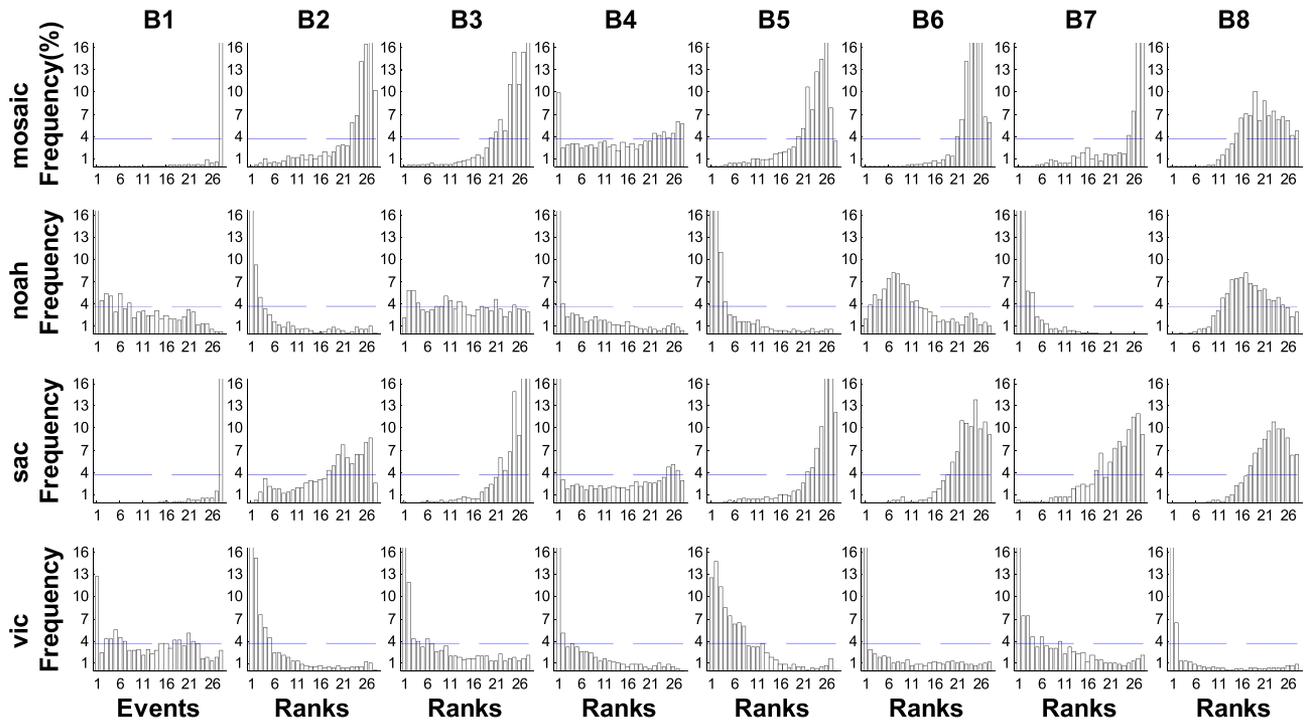


Figure 14. Rank histograms for raw ensemble streamflow simulated from 1980 to 2006. Perfect rank histograms value is 3.57% (the perfect rank value is 1 divided by ensemble members)

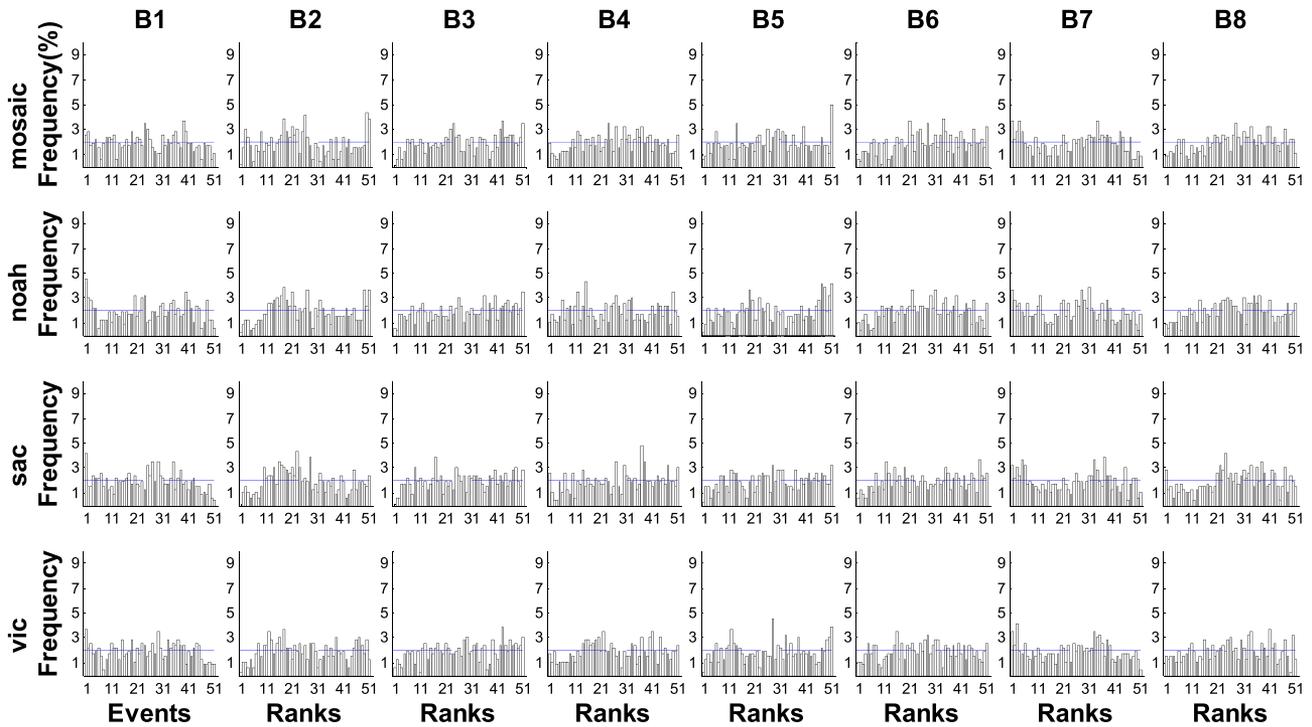


Figure 15. Rank histograms for ensemble streamflow post-processing from 1980 to 2006. Perfect rank value is 1.96%

Evaluation of the ensemble spread of the raw and post-processed streamflow simulations

Once the GLMPP-CE is set, we can use the Monte Carlo method to generate ESPs based on raw streamflow simulations. For the raw streamflow simulations (NLDAS simulated streamflow data), we have 27 ensemble members, each of which corresponds to the streamflow simulation from each year from 1980 to 2006. For post-processed streamflow simulations, we can generate any number of ensembles. The raw streamflow ensembles have 28 ranks (including rank 0) and the post-processed streamflow ensembles have 51 ranks (50 ensemble members). The rank histograms for the raw and post-processed streamflow ensembles of the four models are shown in Figures 14 and 15. The ranks of the raw ensembles for all models are highly variable for all eight basins (Figure 14), whereas the rank of the post-processed ensembles are relatively uniform (Figure 15), indicating that post-processing significantly improves the ensemble spread. The rank histograms of the Mosaic and Sac models show that the ranks to the right side are generally high, suggesting that these models tend to under-predict the streamflow. On the other hand, the rank histograms of the Noah and VIC models indicate that the ranks to the left side are high for most basins, implying that these models over-predict the streamflow. These findings are consistent with the daily streamflow shown in Figures 11 and 12.

Evaluation of the low-flow probability forecasts

The GLMPP-CE can provide low-flow probability forecasts based on ensemble streamflow forecasts. Figure 16 shows the probabilities of observed (black line), raw-simulated (blue dash line) and post-processed (red dash line) daily streamflow values falling below the LFT during the evaluation period. The probability based on post-processed simulation reasonably matches the frequency of observations. The same cannot be said about probability based on raw simulation.

Figure 17 compares the average values of the low-flow probability based on raw (blue-dotted lines) and post-

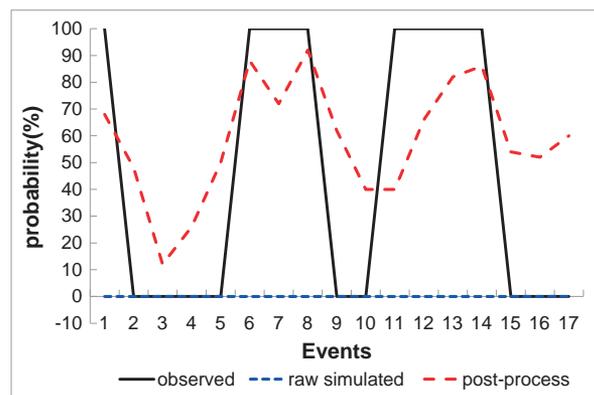


Figure 16. Low-flow probability of variable infiltration capacity model in the B7 basin on 1 October to 1 January 2000

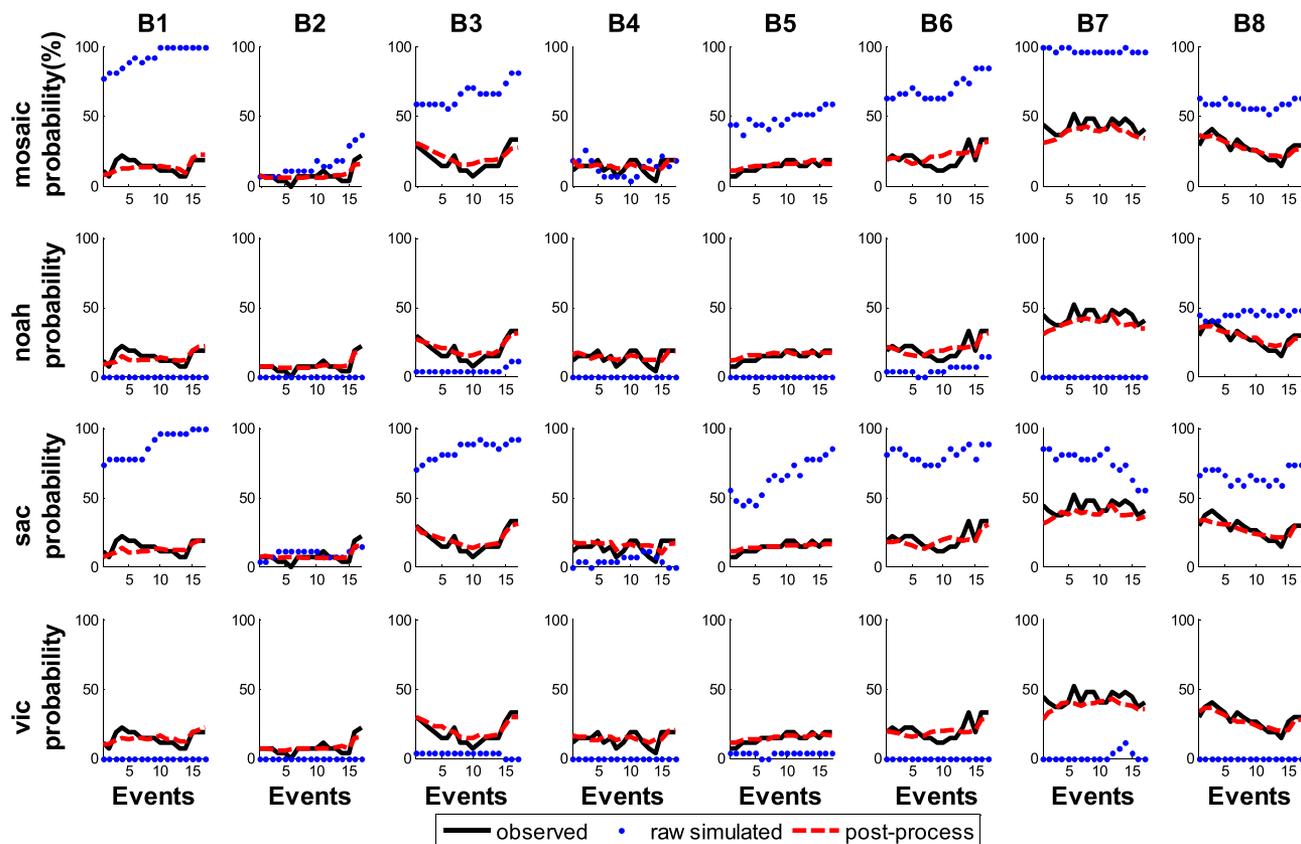


Figure 17. Long-term mean low-flow probability in the seven river basins (B1–B8 except B7) on 1 July to 1 October from 1980 to 2006 (B7 on 1 October to 1 January). The observed is the low-flow frequency. The raw-simulated is the low-flow frequency. The post-process is the long-term mean flow probability

processed (the red dash lines) streamflow against observed probability (black solid lines) for the evaluation period by the four models and for the seven basins. The raw results are shown to deviate consistently from the observed results, although the post-processed results always match the observed results well.

CONCLUSIONS

In this study, we presented a version of GLMPP based on CEs, GLMPP-CE. We used the GLMPP-CE to post-process the streamflow simulations for eight river basins in the USA generated by the four land surface hydrology models in NLDAS. We focused our attention to the performance of raw and post-processed streamflow simulations in the predictions of low-flow events.

We found that GLMPP-CE can effectively reduce the mean biases in the streamflow simulations. Another interesting finding from the study is that GLMPP-CE can deal implicitly with the effect of reservoir regulations, which alter the natural streamflow, as several basins are heavily regulated by reservoirs (e.g. basins 2 and 7). Because none of the models in NLDAS have explicit

reservoir considerations, we can see the significant difference between raw-simulated streamflow of hydrological models and observed streamflow. However, these hydrological models have the skill to simulate natural rainfall-runoff hydrographs; reservoir operation is always in accordance with some rules, such as reservoirs may decrease the streamflow out of the reservoirs during the wet season and increase the streamflow out of the reservoirs during the dry season. So the biases are systematic biases in the streamflow simulations. GLMPP-CE is a statistical model, which can find the biases and remove them. Of course, it is difficult to remove random error. Furthermore, post-processing by GLMPP-CE can help improve the spread of the streamflow ensemble predictions and afford low-flow probability forecast. This suggests that using GLMPP-CE within the ESP framework can help improve the accuracy as well as the reliability of the ESP. Raw simulations from different models vary greatly in different basins. But post-processing can always improve the performance of streamflow simulations by all models under different conditions. We also found that using CEs as the factors in regression helps us achieve significant skill in events with long lead times.

ACKNOWLEDGEMENTS

This study was supported by the Natural Science Foundation of China (No. 41475093), National Science and Technology Support Plan Program (NO. 2013BAB05B04) and the Fundamental Research Funds for the Central Universities (NO. 2013YB32). Special thanks Dr Youlong Xia at Environmental Modeling Center NOAA/NWS/NCEP for providing the NLDAS flow data.

REFERENCES

- Ajami NK, Duan Q, Sorooshian S. 2007. An integrated hydrologic Bayesian multimodel combination framework: confronting input, parameter, and model structural uncertainty in hydrologic prediction. *Water Resources Research* **43**: W01403.
- Bogner K, Pappenberger F, Cloke HL. 2011. Technical note: the normal quantile transformation and its application in a flood forecasting system. *Hydrology and Earth System Sciences Discussions* **8**: 9275–9297.
- Bourdin DR, Stull RB. 2013. Bias-corrected short-range member-to-member ensemble forecasts of reservoir inflow. *Journal of Hydrology* **502**: 77–88.
- Brown JD, Seo D-J. 2010. A nonparametric postprocessor for bias correction of hydrometeorological and hydrologic ensemble forecasts. *Journal of Hydrometeorology* **11**: 642–665.
- Brown JD, Seo D-J. 2013. Evaluation of a nonparametric post-processor for bias correction and uncertainty estimation of hydrologic predictions. *Hydrological Processes* **27**: 83–105.
- Burnash RJC, Ferral RL, McGuire RA. 1973. A generalized streamflow simulation system: conceptual modeling for digital computers. Technical Report, Joint Federal and State River Forecast Center, U.S. National Weather Service and California State Department of Water Resources, Sacramento, California, 204 pp.
- Chen F, Mitchell K, Schaake J, Xue Y, Pan H, Koren V, Duan Q, Betts A. 1996. Modeling of land-surface evaporation by four schemes and comparison with FIFE observations. *Journal of Geophysical Research* **101**: 7251–7268.
- 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.
- Cosgrove BA, Lohmann D, Mitchell KE, Houser PR, Wood EF, Schaake JC, Robock A, Marshall C, Sheffield J, Duan Q, Luo L, Higgins RW, Pinker RT, Tarpley JD, Meng J. 2003. Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *Journal of Geophysical Research, [Atmospheres]* **108**: 8842.
- Day G. 1985. Extended streamflow forecasting using NWSRFS. *Journal of Water Resources Planning and Management* **111**: 157–170.
- Döll P, Siebert S. 2002. Global modeling of irrigation water requirements. *Water Resources Research* **38**: 1037–1046.
- Duan Q, Schaake J, Andréassian V, Franks S, Goteti G, Gupta HV, Gusev YM, Habets F, Hall A, Hay L, Hogue T, Huang M, Leavesley G, Liang X, Nasonova ON, Noilhan J, Oudin L, Sorooshian S, Wagener T, Wood EF. 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.
- Duan Q, Ajami NK, Gao X, Sorooshian S. 2007. Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Advances in Water Resources* **30**: 1371–1386.
- Falloon P, Betts R, Wiltshire A, Dankers R, Mathison C, McNeill D, Bates P, Trigg M. 2011. Validation of river flows in HadGEM1 and HadCM3 with the TRIP river flow model. *Journal of Hydrometeorology* **12**: 1157–1180.
- Franz KJ, Hartmann HC, Sorooshian S, Bales R. 2003. Verification of national weather service ensemble streamflow predictions for water supply forecasting in the Colorado river basin. *Journal of Hydrometeorology* **4**: 1105–1118.
- Koster RD, Suarez MJ. 1996. Energy and water balance calculations in the Mosaic LSM. NASA Tech. Memo., 104606, vol. 9, 60 pp.
- Krzysztofowicz R. 1997. Transformation and normalization of variates with specified distributions. *Journal of Hydrology* **197**(1–4): 286–292.
- Krzysztofowicz R, Kelly KS. 2000. Hydrologic uncertainty processor for probabilistic river stage forecasting. *Water Resources Research* **36**: 3265–3277.
- Liang X, Wood E, Lettenmaier D. 1996. Surface and soil moisture parameterization of the VIC-2L model: evaluation and modifications. *Global and Planetary Change* **13**: 195–206.
- Lohmann D, Mitchell KE, Houser PR, Wood EF, Schaake JC, Robock A, Cosgrove BA, Sheffield J, Duan Q, Luo L, Higgins RW, Pinker RT, Tarpley JD. 2004. Streamflow and water balance intercomparisons of four land surface models in the North American Land Data Assimilation System project. *Journal of Geophysical Research, [Atmospheres]* **109**: D07S91.
- Madadgar S, Moradkhani H, Garen D. 2014. Towards improved post-processing of hydrologic forecast ensembles. *Hydrological Processes* **28**: 104–122.
- Mccollor D, Stull R. 2008. Hydrometeorological Accuracy Enhancement via Postprocessing of Numerical Weather Forecasts in Complex Terrain. *Weather and Forecasting* **23**: 131–144.
- Mitchell KE, Lohmann D, Houser PR, Wood EF, Schaake JC, Robock A, Cosgrove BA, Sheffield J, Duan Q, Luo L, Higgins RW, Pinker RT, Tarpley JD, Lettenmaier DP, Marshall CH, Entin JK, Pan M, Shi W, Koren V, Meng J, Ramsay BH, Bailey AA. 2004. The multi-institution North American Land Data Assimilation System (NLDAS): utilizing multiple GCIIP products and partners in a continental distributed hydrological modeling system. *Journal of Geophysical Research, [Atmospheres]* **109**: D07S90.
- Pushpalatha R, Perrin C, Moine NL, Andréassian V. 2012. A review of efficiency criteria suitable for evaluating low-flow simulations. *Journal of Hydrology* **420–421**: 171–182.
- Schaake J, Franz K, Bradley A, Buizza R. 2006. The Hydrologic Ensemble Prediction Experiment (HEPEX). *Hydrology and Earth System Sciences Discussions* **3**: 3321–3332.
- Seo D-J, Herr HD, Schaake JC. 2006. A statistical post-processor for accounting of hydrologic uncertainty in short-range ensemble streamflow prediction. *Hydrology and Earth System Sciences Discussions* **3**: 1987–2035.
- Shaw SB, Riha SJ. 2012. Examining individual recession events instead of a data cloud: using a modified interpretation of $dQ/dt-Q$ streamflow recession in glaciated watersheds to better inform models of low flow. *Journal of Hydrology* **434–435**: 46–54.
- Shi X, Wood AW, Lettenmaier DP. 2008. How essential is hydrologic model calibration to seasonal streamflow forecasting? *Journal of Hydrology* **9**: 1350–1363.
- Shukla S, Sheffield J, Wood EF, Lettenmaier DP. 2013. On the sources of global land surface hydrologic predictability. *Hydrology and Earth System Sciences Discussions* **10**: 1987–2013.
- Singh VP, Frevert DK. 2006. *Watershed Models*. CRC Taylor and Francis: Boca Raton, FL.
- Slater AG, Clark MP. 2006. Snow data assimilation via an ensemble Kalman filter. *Journal of Hydrometeorology* **7**: 478–493.
- Smakhtin VU. 2001. Low flow hydrology: a review. *Journal of Hydrology* **240**: 147–186.
- Sperna Weiland FC, Van Beek LPH, Kwadijk JJC, Bierkens MFP. 2010. The ability of a GCM-forced hydrological model to reproduce global discharge variability. *Hydrology and Earth System Sciences Discussions* **7**: 687–724.
- Todini E. 2008. A model conditional processor to assess predictive uncertainty in flood forecasting. *International Journal of River Basin Management* **6**: 123–137.
- Wilks DS. 2011. *Statistical Methods in the Atmospheric Sciences*, 3rd edn. Academic Press: New York.
- Wisser D, Fekete BM, Vörösmarty CJ, Schumann AH. 2010. Reconstructing 20th century global hydrography: a contribution to the Global Terrestrial Network-Hydrology (GTN-H). *Hydrology and Earth System Sciences* **14**: 1–24.
- WMO. 1974. *International Glossary of Hydrology*. World Meteorological Organization: Geneva.

- Ye A, Duan Q, Yuan X, Wood EF, Schaake J. 2014. Hydrologic post-processing of MOPEX streamflow simulations. *Journal of Hydrology* **508**: 147–156.
- Yuan X, Wood EF. 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.
- 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.