

Bias-correction schemes for calibrated flow in a conceptual hydrological model

Kue Bum Kim, Hyun-Han Kwon and Dawei Han

ABSTRACT

We explore post-processing methods that can reduce biases in simulated flow in a hydrological model (HYMOD). Here, three bias-correction methods are compared using a set of calibrated parameters as a baseline (Cases 1 and 5). The proposed bias-correction methods are based on a flow duration curve (Case 2), an autoregressive model based on residuals obtained from simulated flows (Case 3), and a rating curve (Case 4). A clear seasonality representing a more substantial variability in winter than summer was evident in all cases. The extended range of residuals was usually observed in winter, indicating that the HYMOD may not reproduce high flows appropriately. This study confirmed that bias-corrected flows are more effective than the baseline model in terms of correcting a systematic error in the simulated flow. Moreover, a comparison of root mean square error over different flow regimes demonstrates that Case 3 is the most effective at correcting systematic biases over the entire flow regime. Finally, monthly water balances for all cases are evaluated and compared during both calibration and validation periods. The water balance in Case 3 is also closer to the observed values. The effects of different post-processing approaches on the performance of bias-correction are examined and discussed.

Key words | bias correction, hydrological model, residual and autocorrelation

Kue Bum Kim

Water Resources Policy Division, Ministry of Environment,
Sejong-si,
South Korea

Hyun-Han Kwon (corresponding author)

Department of Civil and Environmental Engineering,
Sejong University,
Seoul,
South Korea
E-mail: hkwon@sejong.ac.kr

Dawei Han

Water and Environmental Management Research Centre, Department of Civil Engineering,
University of Bristol,
Bristol,
UK

HIGHLIGHTS

- We explore three bias-correction methods that can reduce biases in simulated flow.
- This is the first attempt to combine calibration and bias correction in the runoff simulation.
- This study confirmed that bias correction approaches are more effective in terms of correcting a systematic error.
- The monthly water balance estimated from the proposed approach was close to that of the observed value.

INTRODUCTION

Hydrological models are essential tools (Madsen 2000; Wagener *et al.* 2003) for understanding water cycles that provide a basis for decision-making related to water resource

management, land planning, infrastructure design, and water supply. They are also imperfect representations of the real world due to model uncertainties, measurement errors, sampling errors stemming from insufficient data, climate variability, and inadequate descriptions of initial and boundary conditions. Understanding the hydrological behavior of a catchment is a complicated task (De Vos *et al.*

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

doi: 10.2166/nh.2021.043

2010b). Typical hydrological modeling processes include: (1) parameterization based on observed data; (2) sensitivity analysis to diagnose which parameter is the most essential during calibration; and (3) calibration, either manually or automatically (Eckhardt & Arnold 2001; Eckhardt *et al.* 2005; Muleta & Nicklow 2005; Abbaspour *et al.* 2007; Bekele & Nicklow 2007; Van Griensven & Meixner 2007; Green & Van Griensven 2008; Zhang *et al.* 2009).

Commonly used performance metrics in the form of objective functions in rainfall–runoff model calibration are Nash–Sutcliffe efficiency and root mean square error (RMSE), which can be obtained by comparing observed and simulated streamflows. However, the limited ability of these metrics to cover different aspects of hydrologic behavior can lead to systematic bias after calibration. Outputs from a calibrated hydrological model also have systematic biases, although model parameters can be obtained by minimizing the difference between the observed and simulated flows within an optimization framework. Model efficiency is influenced more by high flows than low flows (Pushpalatha *et al.* 2012) without explicitly treating them, and information on temporal variability of flow is poorly preserved. Attempts to explore these issues and improve model uncertainties have been made. For example, a combination of several performance measures was applied to explore different aspects of the fit between observed and simulated flow (Gupta *et al.* 1998; Wagener *et al.* 2001a; Freer *et al.* 2004).

Apart from model calibration, statistical post-processing approaches for ensemble streamflow predictions have been introduced to improve model prediction skills by reducing systematic biases (Bogner & Kalas 2008; Weerts *et al.* 2011; Zhao *et al.* 2011). The main objective of post-processing is to reduce biases in the mean and variance of raw ensemble streamflow predictions. A number of post-processing approaches are available to correct simulated streamflow in a hydrologic model. One such approach is a quantile mapping of the cumulative distributions of a prediction onto those of the observed distributions (Shi *et al.* 2008; Madadgar *et al.* 2014). Another correction approach is based on a Bayesian model (Krzysztofowicz & Maranzano 2004; Brown & Seo 2010, 2013). A simple bias-correction method and generalized linear regression (Zhao *et al.* 2011) have been applied to reduce uncertainties in model predictions. Ensemble

streamflow forecasts along with statistical post-processing produce superior performance compared with raw ensemble hydrological forecasts (Krzysztofowicz & Kelly 2000; Krzysztofowicz & Maranzano 2004; Montanari & Brath 2004; Seo *et al.* 2006; Wood & Schaake 2008; Weerts *et al.* 2011). These studies suggest that post-processing can improve a conventional calibration approach in an efficient way (Shi *et al.* 2008; Yuan & Wood 2012; Ye *et al.* 2015).

Previous studies have focused separately on the hydrological model calibration process and the statistical post-processing approach. This study took an integrated approach that combined post-processing techniques (i.e., bias-correction methods) with the calibration of a hydrological model. Specifically, three bias-correction models (Cases 2, 3, and 4) were proposed to reduce systematic biases in streamflow simulation. The proposed bias-correction approaches were evaluated along with a conceptual hydrological model. To the best of our knowledge, few studies have combined hydrological model calibration and bias-correction approaches in a single model. Model performance was also compared with simulated flow obtained from an existing calibration approach (Case 1) and a sub-annual calibration approach (Case 5). A brief overview of the post-processing procedures considered in this study follows:

- Case 1 was based on the use of an existing calibration approach without bias correction.
- Case 2 was based on flow duration curves (FDCs) of observed and calibrated flows.
- Case 3 was based on an autoregressive model constructed from time series of the residual, which were calculated from observed and simulated flows.
- Case 4 was also based on an autoregressive model constructed from time series of the residual. However, the residual was calculated from observed and simulated river levels, which were converted using a rating curve.
- Case 5 was based on a sub-annual calibration approach without bias correction. Two different parameter sets are obtained by calibrating the dry season (spring and summer) and the wet season (autumn and winter) separately.

We compared the performance of the simulated flow from a well-calibrated model as a baseline (Cases 1 and 5)

with the three bias-correction models (Cases 2, 3, and 4) to address the following questions:

- (1) Can a seasonal pattern and systematic trend in simulated flow from a rainfall–runoff model be identified?
- (2) Can a bias-correction approach reduce systematic biases in the residual for both calibration and validation periods? What is the most effective method of correcting biases of an entire flow regime?
- (3) Can simulated flow with bias correction effectively reproduce observed water balance?

The balance of this study is organized as follows. In the section ‘Case Study Area and the Hydrological Model’, we



Figure 1 | Map showing the location of the Thorverton catchment in the UK.

describe the study area, data, and hydrological model. The three different post-processing methods introduced in this study are summarized in the section ‘Bias-Correction Scheme’. The results of different bias-correction methods are presented in the ‘Results and Discussion’. The main conclusions and discussions of this study are provided in the section ‘Discussion and Conclusion’.

CASE STUDY AREA AND THE HYDROLOGICAL MODEL

Study area and data

The Thorverton catchment has an area of 606 km² and is a sub-catchment of the Exe catchment. The Exe catchment is in the southwest of England and has an area of 1,530 km² and an average annual rainfall of 1,088 mm. [Figure 1](#) provides an overview of the Thorverton catchment. A daily time series of the observed precipitation, potential evapotranspiration, and flow data (1961–1990) from the Thorverton catchment was obtained from the UK Met Office. Rating curves were acquired from the National River Flow Archive. Thirty-year (1961–1990) mean rainfall and temperature data for this catchment are presented in [Figure 2](#). Winter is the wettest season and summer is the warmest.

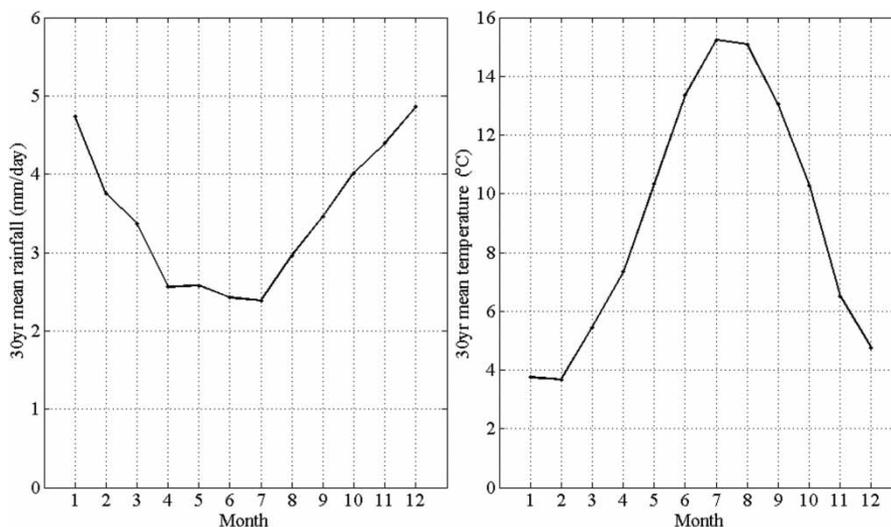


Figure 2 | The 30-year mean monthly rainfall and temperature record for the Thorverton catchment.

Hydrological model and calibration

The conceptual rainfall–runoff model used in this study, HYMOD, has five parameters. The model consists of a simple excess rainfall model based on the probability distributed principle (Moore 1985) and was applied in several recent studies (Boyle et al. 2003; Wagener et al. 2001b; Vrugt et al. 2003; De Vos et al. 2010a). Model parameters are described in Table 1, and the model structure is illustrated in Figure 3. The cumulative distribution function of the water storage capacity, C , has the following form:

$$F(C) = 1 - \left(1 - \frac{C(t)}{C_{\max}}\right)^{b_{\text{exp}}}, \quad 0 \leq C(t) \leq C_{\max} \quad (1)$$

where C_{\max} is the maximum soil moisture storage capacity in the catchment, and b_{exp} controls the degree of spatial variability of the soil moisture capacity. The excess rainfall can be transformed into the runoff, which is divided into quick and slow flows based on partitioning factor α . The runoffs are routed through three identical quick-flow tanks and a parallel slow-flow tank. The flow rates are determined

Table 1 | Model parameters considered in the HYMOD

Parameter	Unit	Range	Description
C_{\max}	mm	1–500	Maximum soil moisture storage capacity
b_{exp}	–	0.01–1.99	Spatial variability of soil moisture capacity
α	–	0.01–0.99	Quick-/slow-flow distribution factor
R_s	Day	0.01–0.99	Recession coefficient for slow-flow tank
R_q	Day	0.01–0.99	Recession coefficient for quick-flow tank

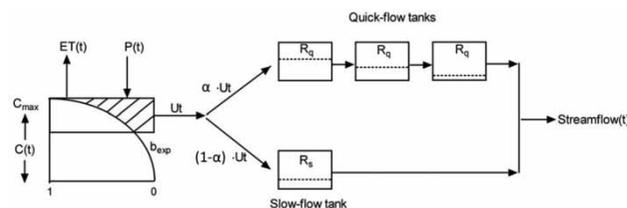


Figure 3 | Structure of the HYMOD (adopted from Vrugt et al. (2002)).

by the recession coefficient for the quick-flow tank (R_q) and slow-flow tank (R_s).

For the optimization algorithm, we used dynamically dimensioned search (DDS; Tolson & Shoemaker 2007), which is a simple, single-objective, and heuristic global search algorithm. DDS searches a set of parameters globally and incrementally by localizing the parameter space as the number of iterations reaches a predefined maximum number of simulations. The procedure from the global to the local scale is carried out by probabilistically decreasing the number of model parameters in the neighborhood. The DDS algorithm easily converges to global solutions by perturbing the current solution values in randomly selected dimensions with magnitudes of the perturbations randomly sampled from a normal distribution with a mean of zero. More details can be found in Tolson & Shoemaker (2007).

Calibration was conducted to obtain an optimal set of parameters with an objective function based on the RMSE by minimizing the difference between the observed and simulated flows, and the optimized parameters were tested in the validation period by:

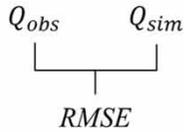
$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (Q_{\text{obs},i} - Q_{\text{sim},i})^2}{N}} \quad (2)$$

where Q_{sim} and Q_{obs} are the simulated and observed runoff, respectively. i is the i th day, and N is the number of days in the calibration period. The effect of different objective functions and their combinations in the calibration of rainfall–runoff models was not explored.

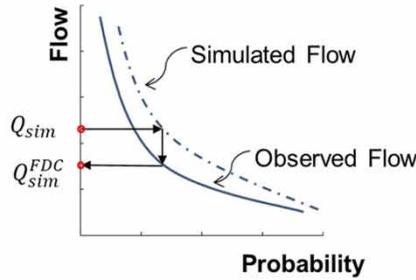
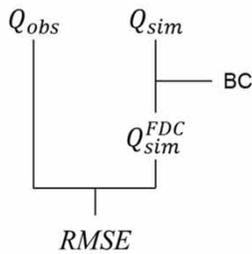
BIAS-CORRECTION SCHEME

First, hydrological data, including precipitation and temperature from 1961 to 1990, were compiled and used to calibrate the HYMOD. Second, different transfer functions (i.e., FDC, autoregressive relations, and rating curve) were constructed to map the simulated flow onto the observed flow in the process of bias correction. Third, remaining hydrological data from 1991 to 2008 were used for validation purposes. Figure 4 illustrates a schematic diagram of the bias-correction approaches and the improved

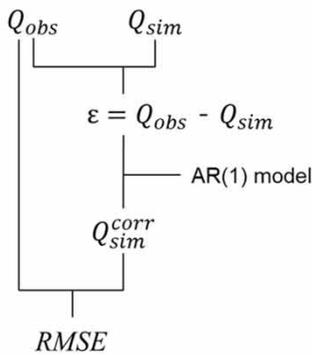
(Case 1) Conventional method



(Case 2) Bias correction based on FDC

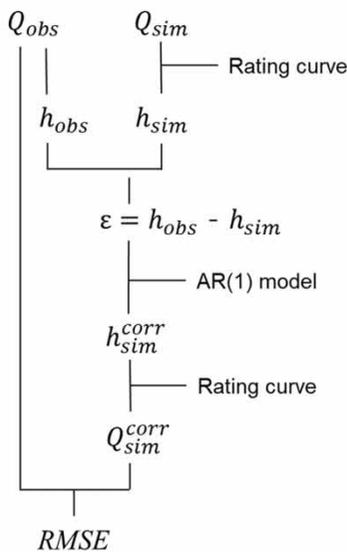


(Case 3) Bias correction based on AR model



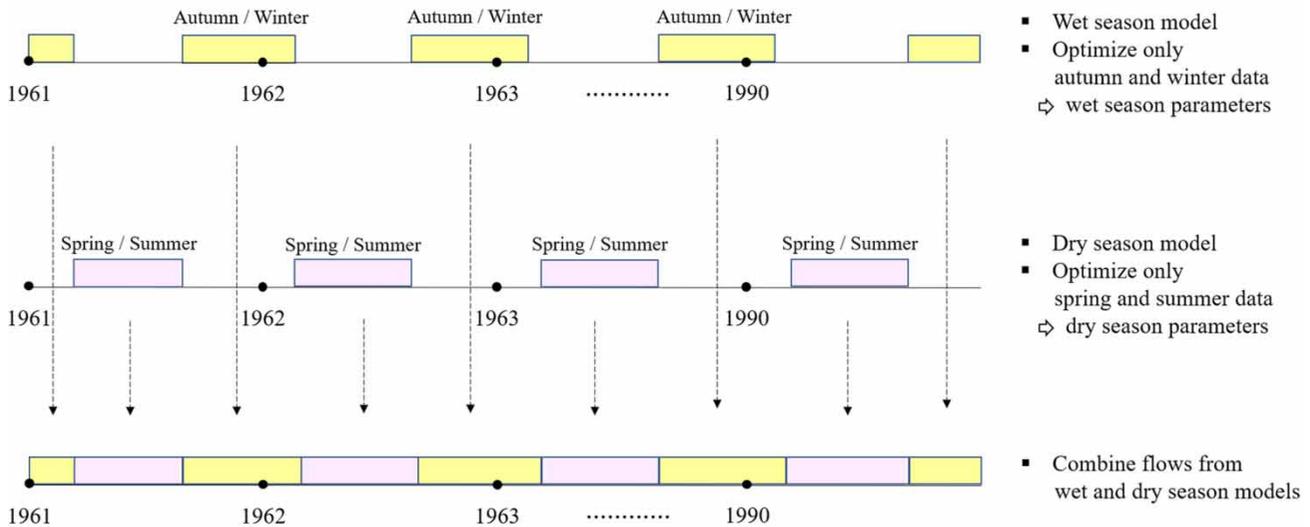
- $\varepsilon = Q_{obs} - Q_{sim}$
- $\varepsilon_{t+1} = a \times \varepsilon_t + b$
- $Q_{sim}^{corr} = Q_{sim} + \varepsilon$

(Case 4) Bias correction based on Rating curve



- $\varepsilon = h_{obs} - h_{sim}$
- $\varepsilon_{t+1} = a \times \varepsilon_t + b$
- $h_{sim}^{corr} = h_{sim} + \varepsilon$
- $h_{sim}^{corr} \rightarrow Q_{sim}^{corr}$

Figure 4 | Schematic concept of the conventional calibration method (Case 1), three different post-processing methods for correcting bias of simulated flows (Cases 2, 3, and 4), and sub-annual calibration scheme (Case 5). (Continued.)

(Case 5) Sub-annual calibration scheme**Figure 4** | Continued.

sub-annual calibration scheme proposed in this study. This study compared five different schemes, including a conventional calibration method (Case 1), three post-processing methods that were applied to the simulated flow in the context of bias correction (Cases 2, 3, and 4), and sub-annual calibration based on wet and dry seasons separately (Case 5). The parameters obtained from the calibration process (Case 1) were subsequently used to simulate streamflow sequences as a baseline for the other three cases. In other words, three post-processing methods were aimed at correcting the simulated flow from a conventional calibration process (Case 1). In Case 2 (the FDC model), the cumulative probability of simulated flow at a given time was obtained from the FDC of the simulated flow, and the corresponding observed probability was then taken from that of the observed value (FDC from observation) to adjust the calibrated flow further. The processes were as follows:

- First, the FDC was built with data from the entire calibration period for both observed and simulated flows.
- Second, the probability of the simulated flow ($FDC_{sim}(Q_{sim})$) was computed from the simulated FDC.
- Third, the bias-corrected flow (Q_{sim}^{corr}) was computed by mapping the probability of simulated flow obtained from the simulated FDC onto the observed FDC

($FDC_{obs}(Q_{obs})$). The equation can be expressed as follows:

$$Q_{sim}^{corr} = FDC_{obs}^{-1} [FDC_{sim}(Q_{sim})] \quad (3)$$

where FDC_{sim} is constructed from the simulated flow, FDC_{obs}^{-1} is an inverse function built with the observed flow, and Q_{sim}^{corr} is the bias-corrected flow. The FDC played the role of a transfer function for bias correction of the simulated flow by matching the FDC of observed flow for the same period. The transfer functions were estimated for 30 years (1961–1990), and these transfer functions were then applied to the remaining period (1991–2008) for validation.

The following bias-correction process is based on a first-order autoregressive model of the residual $AR_Q(1)$, named Case 3:

- First, the residual series was obtained by subtracting the simulated flow from that of the observed (Equation (4)).
- Second, the residual was assumed to follow the first-order autoregressive model, $AR_Q(1)$, as written in Equation (5).
- Third, the bias-corrected flow was obtained by adding the modeled residuals from $AR_Q(1)$ to the simulated flow

(Equation (6)) as follows:

$$\varepsilon_t = Q_{\text{obs},t} - Q_{\text{sim},t} \quad t = 1, 2, 3, \dots, T \quad (4)$$

$$\varepsilon_{t+1} = a \times \varepsilon_t + b \quad (5)$$

$$Q_{\text{sim},t+1}^{\text{corr}} = Q_{\text{sim},t+1} + \varepsilon_{t+1}, \quad (6)$$

where t is time index (days) and T is the number of records. ε_{t+1} is the modeled residual at time $t + 1$, while a and b are coefficients of the AR_Q(1) model.

In the proposed bias-correction method (Case 4), AR_H(1) was applied in the rating curve domain. In other words, the simulated flow was mapped to the rating curve to obtain the water stage instead of flow. Moreover, the residual was assumed to follow the first-order autoregressive model as adopted by Case 3. The key aspects of the bias-correction methods in Case 4 can be summarized as follows:

- First, the simulated flow was converted to the water stage through a rating curve.
- Second, the residual sequences were obtained by subtracting the simulated water stage from that of the observed value (Equation (7)).
- Third, the AR_H(1) autoregressive model was applied to the residual series (Equation (8)).
- Fourth, the bias-corrected water stage was synthesized by adding the modeled residuals from AR_H(1) to the simulated water stage (Equation (9)).
- Fifth, the bias-corrected flow was obtained by mapping the water stage onto the rating curve (Equation (10)).

$$\varepsilon_t = h_{\text{obs},t} - h_{\text{sim},t} \quad t = 1, 2, 3, \dots, T \quad (7)$$

$$\varepsilon_{t+1} = a \times \varepsilon_t + b \quad (8)$$

$$h_{\text{sim},t+1}^{\text{corr}} = h_{\text{sim},t+1} + \varepsilon_{t+1} \quad (9)$$

$$h_{\text{sim},t+1}^{\text{corr}} \rightarrow Q_{\text{sim},t+1}^{\text{corr}} \quad (10)$$

Kim & Han (2017) explored sub-annual calibration schemes. They found that the model calibrated on the sub-annual period schemes generally performs better than the model calibrated on the entire datasets without its separation. Therefore, the sub-annual calibration method is

additionally considered as a reference (Case 5) for comparison with those of the bias-correction schemes (Cases 2, 3, and 4). The purpose of this experiment was to explore whether an improved calibration framework can be more effective than a post-processing approach to the simulated flow. More specifically, the model (HYMOD) was calibrated for two distinct seasons (i.e., dry and wet seasons), and the obtained parameter sets were subsequently used in the validation phase. During calibration, the model was kept running continuously until the optimization was done only for the chosen seasons. The calibrated flow data from sub-annual calibration schemes are then combined together to examine the model performance over the entire time period.

RESULTS AND DISCUSSION

Comparison of results during the calibration period

Figure 5 compares residual distributions over the year for five different cases. Case 1, which is based on a conventional calibration process without bias correction, is used as a baseline for the other three, while Case 2 corrects for bias in the FDC domain. A temporal pattern in the residuals that may be influenced by seasonality in hydrologic variables (i.e., precipitation and temperature) can be seen in Figure 5. No clear trend emerged with respect to the residual because the distribution was nearly symmetric. The relative range (i.e., the range/the average flow) in summer was greater than that in winter for five different cases. The ratio of relative range in summer to that in winter ranges from 1.49 to 1.83, indicating that the HYMOD may not adequately capture low flows during summer. The degree of reduction in the range of the residual in Case 1 was primarily the extent to which bias correction based on FDC (Case 2) was effective. The median of the residual continues to converge toward zero in all months of the year, and the reduction ratio in the range of residual in the Case 2 with respect to Case 1 was significantly larger in summer (21.3–21.8%) than in winter (6.3–8.4%) as summarized in Table 2. This suggests that the FDC model (Case 2) reduced bias in the low flow more effectively than in the high flow. For Case 3, a clear seasonality with a large variability in winter was evident in line with Case 2. The degree of

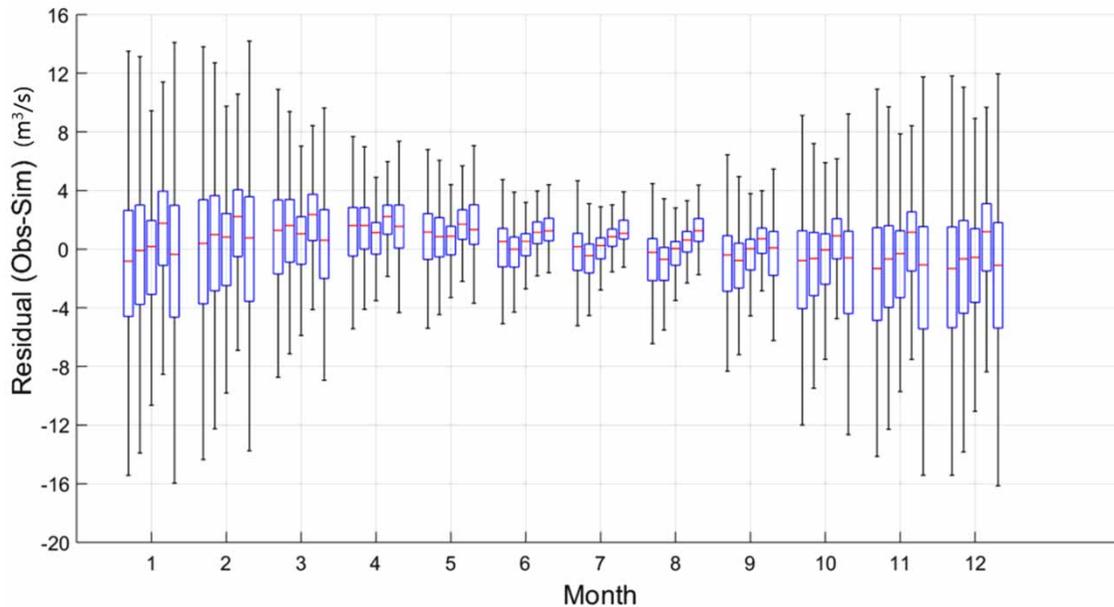


Figure 5 | Comparison of monthly residuals for each case during the calibration period. The boxplots of each month are the result of Cases 1, 2, 3, 4, and 5 (from left to right).

reduction in the range of the residual was more pronounced in Case 3, as shown in Table 2. The reduction ratio in the winter (26.9%–30.7%) was substantially larger than that of the Case 2 (6.3–8.4%), confirming the efficacy of the AR_F model compared with the FDC model in reducing bias in the simulated flow. More importantly, the median of the residual remained close to zero, particularly for those treated with the first-order autoregressive model on residuals, $AR_Q(1)$, compared with Case 3. The monthly distribution of the residual in Case 4 was largely comparable to that of Case 3 in terms of seasonality and variability. Furthermore, the reduction ratio expanded slightly over the year, relative to Case 3. However, the median of the residual tended toward positive values that led in turn to an increase in bias associated with underestimation of the flow, indicating that the efficacy of treating biases in the simulated flow may be limited in its application to bias correction. For Case 5, a clear seasonality with large variability in winter was evident as in other cases. The reduction ratio in the winter (–5.2 to –0.5%) showed negative values, which implies that the performance of the sub-annual calibration scheme model has not been improved compared with that of the conventional calibration method (Case 1). However, the reduction ratio in summer (41.4–48.9%) was largely comparable to that of

Cases 3 and 4. This indicates that the sub-annual calibration model performs better in the low flows than in the high flows.

The relationship between the observed flow and the residual was further explored in the context of residual distribution by categorizing the residual into six classes from –60 to $60 \text{ m}^3 \text{ s}^{-1}$, a range that includes most (99.8%) of the residuals. The labels of categorized residuals in Figure 6 indicate the upper boundary of each class. Here, our focus was to explore the effects of bias correction in adjusting residual distribution with respect to the magnitude of the flow. Clear tendencies for both negative and positive residuals were evident. For Case 1, residuals showed an increasing trend as the flow magnitude increased. The results indicate that the model had a systematic error in simulating high flows, and this may have been due in part to a limited ability to represent the fast-runoff component. In Case 1, negative residuals were associated with relatively low flows of between approximately 5 and $80 \text{ m}^3 \text{ s}^{-1}$, while positive residuals were associated with relatively high flows of 20 – $170 \text{ m}^3 \text{ s}^{-1}$. In particular, after bias correction, the positive residual associated with the high flow showed a decreasing tendency, while the negative residual associated with the high flow appeared to increase. However, this does not necessarily mean that it had no important effects

Table 2 | Range of monthly residual for each case

Month	Range	Range of residuals										
		Case 1			Case 2		Case 3		Case 4		Case 5	
		Value	Value	RR (%)	Value	RR (%)	Value	RR (%)	Value	RR (%)		
1	Max–Min	29.0	27.1	6.3	20.2	30.3	20.2	30.2	30.5	– 5.2		
	Interquartile	7.2	6.8		5.0		5.1		7.6			
2	Max–Min	28.4	26.0	8.4	19.7	30.7	18.2	35.9	28.5	– 0.5		
	Interquartile	7.1	6.5		4.9		4.6		7.1			
3	Max–Min	20.2	17.1	15.2	13.0	35.6	12.6	37.5	18.7	7.3		
	Interquartile	5.0	4.3		3.2		3.2		4.7			
4	Max–Min	13.3	11.3	14.9	8.6	35.0	7.9	40.4	11.8	11.6		
	Interquartile	3.3	2.8		2.2		2.0		2.9			
5	Max–Min	12.5	10.7	14.4	7.8	37.6	8.1	35.6	10.8	13.6		
	Interquartile	3.1	2.7		2.0		2.0		2.7			
6	Max–Min	10.5	8.2	21.8	6.0	42.6	5.9	43.7	6.2	41.4		
	Interquartile	2.6	2.1		1.5		1.5		1.5			
7	Max–Min	10.1	7.9	21.8	5.7	43.4	4.7	53.5	5.2	48.9		
	Interquartile	2.5	2.0		1.4		1.2		1.3			
8	Max–Min	11.5	9.0	21.3	6.4	43.7	5.6	50.8	6.2	45.8		
	Interquartile	2.9	2.3		1.6		1.4		1.6			
9	Max–Min	15.1	12.3	19.1	8.4	44.4	6.9	54.4	12.0	20.8		
	Interquartile	3.8	3.1		2.1		1.7		3.0			
10	Max–Min	21.3	17.2	19.0	13.7	35.4	11.0	48.4	22.5	– 5.5		
	Interquartile	5.3	4.3		3.4		2.7		5.6			
11	Max–Min	25.3	22.3	11.9	18.3	27.8	16.1	36.2	27.9	– 10.2		
	Interquartile	6.3	5.6		4.6		4.0		7.0			
12	Max–Min	27.5	25.3	7.9	20.1	26.9	18.4	33.1	28.8	– 4.6		
	Interquartile	6.9	6.3		5.0		4.6		7.2			

'Max–Min' represents a range between the maximum value and the minimum value. 'interquartile' represents lower and upper quartiles (25th and 75th percentiles). 'RR' represents a reduction ratio of residuals for Cases 2, 3, 4 and 5 against Case 1.

on the correction of negative bias (i.e., overestimation of simulated flow).

The increase in the negative bias can be partially attributed to a bias that cannot be fully explained by the proposed bias-correction schemes. More specifically, the positive bias relating to the high flow may have migrated to the negative bias with the effect of bias correction being most pronounced at high flows, resulting in a decrease in the positive residual. A more detailed exploration of this migration pattern is provided in the following section. For Case 5, as is the case with Case 1, negative residuals were associated with relatively low flows and positive residuals were associated with relatively high flows except for the residual range from 0 to $20 \text{ m}^3 \text{ s}^{-1}$.

Figure 7 shows the distribution of residuals along with the flow to compare Case 1 and the other three cases; conditional heteroscedastic variances on the magnitude of flow are evident in all cases. As expected, the bias-corrected flows for the three cases appear to be more effective than the baseline (Case 1) in correcting systematic bias. Among the three bias-correction schemes, the performance of Cases 3 and 4 was superior in terms of overall efficiency compared with Case 2. The shape of the marginal distribution of residual converged to zero with reduced variance through the use of bias-correction approaches. The reduced variability of the residual is represented graphically in Figure 7. Overall, the marginal distribution of the residual in Case 2 was largely similar to that of Case 1 and a significant decrease in

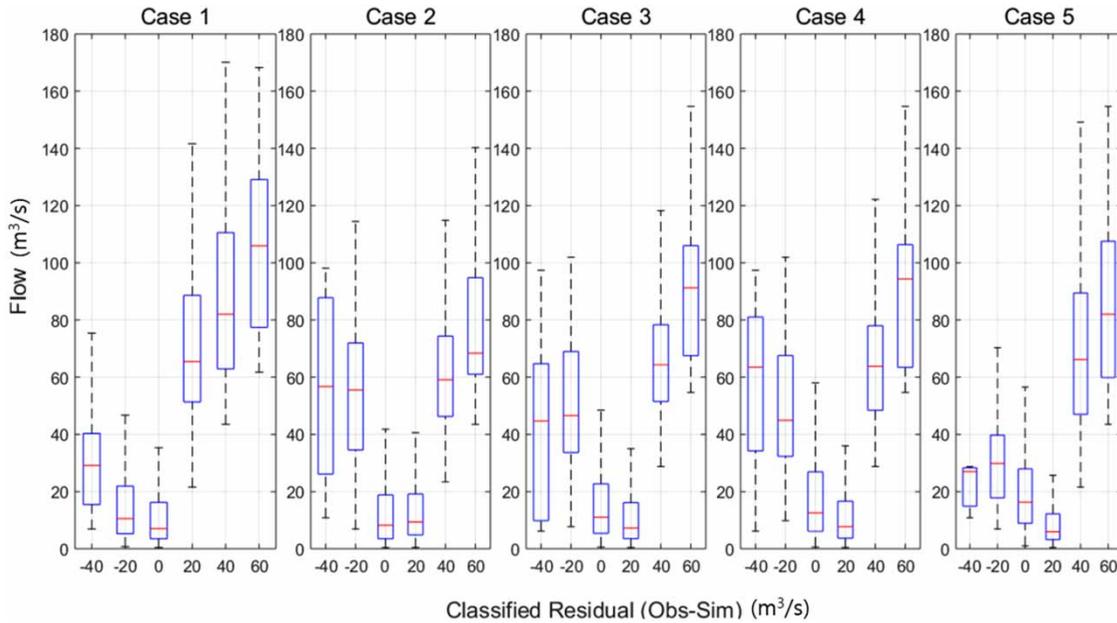


Figure 6 | Boxplots for observed flow with respect to the magnitude of residual for five different cases during the calibration period. Labels on the x-axis indicate the upper boundary of the residual range from -60 to $60 \text{ m}^3 \text{ s}^{-1}$.

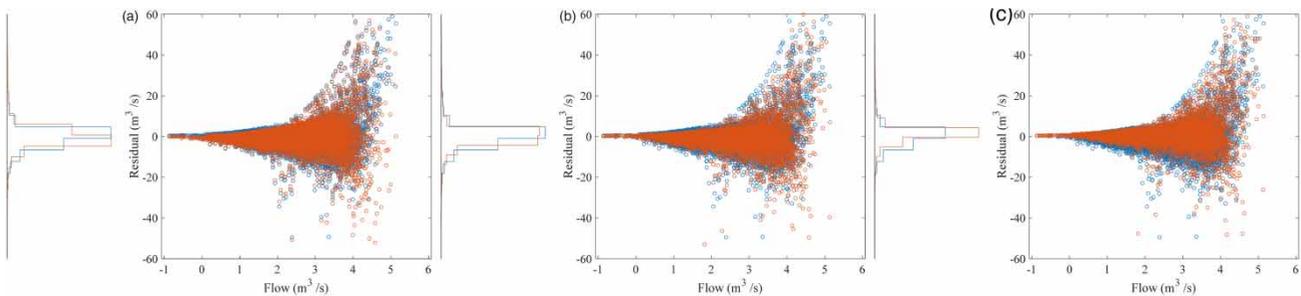


Figure 7 | Distribution of residuals along with the flow for a comparison between Case 1 and the other three cases. Here, the x-axis is the log-transformed flow.

negative bias was evident in Cases 2, 3, and 4. However, negative biases ranging from -40 to -60 were slightly increased by the use of the bias-correction approach, which is depicted in Figure 6. The slight increase in negative residual in a certain range can be attributed to the shift from a positive bias after bias correction, but the shift toward the negative reported here was not significant.

Table 3 provides a comparison of the RMSE in different cases during the calibration period. The RMSE for Cases 1 and 2 is comparable to each other, while an apparent reduction in the RMSE is seen in Cases 3 and 4, which can be attributed to the efficacy of bias-correction approaches proposed in this study. The residuals in the RMSE were squared before averaging them to ensure a

Table 3 | Comparison of RMSE over Cases 1, 2, 3, 4, and 5 during the calibration period ($\text{m}^3 \text{ s}^{-1}$). Values in bold correspond to the lowest RMSE value

Flow range	Bias-correction methods				
	Case 1	Case 2	Case 3	Case 4	Case 5
Entire flow	7.219	7.277	6.429	6.638	7.483
High flow (75th–100th)	13.05	13.355	12.069	12.511	13.313
Intermediate flow (25th–75th)	4.189	3.943	3.123	3.101	4.744
Low flow (0–25th)	1.727	1.521	1.115	1.286	1.295

relatively high penalty on a large residual was assigned (Kim *et al.* 2018); an increase in RMSE could therefore be expected, particularly for the significant difference in the

simulation. In general, the results imply that the FDC model (Case 2) played a limited role in correcting large residuals, whereas the $AR_Q(1)$ (Case 3) and $AR_H(1)$ (Case 4) bias-correction approaches appeared to be much more effective in terms of removing a large residual. To further investigate the efficacy of bias-correction methods, a comparison of RMSE over different flow regimes (i.e., low, intermediate, and high flow) is provided in Table 3. Low, intermediate, and high flows are classified by quartiles based on the 25th and 75th percentiles. Cases 3 and 4 appeared to be most effective in correcting systematic biases over the entire flow regime. Compared with Case 1, Case 2 performed slightly better for intermediate- and low-flow regimes, while a slight increase in RMSE was observed for high flow. The sub-annual calibration model (Case 5) were mainly effective in terms of reproducing low-flow regimes compared with Case 1.

Table 4 presents the monthly water balances of the simulated flows for all four cases during the calibration period. The water balance of Case 3 was closer to that of the observed value, in line with the results presented in Figure 7 and Table 3. The water balance for Case 2 on an annual basis was equal to the observed value because bias correction was based on an FDC. However, monthly differences in water balance were substantially increased. Apart from Case 2, Case 3 was the most similar to the observed values.

Results during the validation period

As illustrated in Figure 5 during the calibration period, Figure 8 compares residual distributions over the year for the five different cases during the validation period. Besides, Figure 9 illustrates the relationship between the observed flow and the residual in the context of residual distribution. Both the results are comparable to that of the calibration period. The median of residual distributions for Case 3 becomes closer to zero, while other cases are relatively distant from zero. Thus, a comparison of validation experiment and simulation again confirms that the systematic bias of the simulated flow can be more effectively reduced by Case 3, as already seen in the calibration period.

A comparison of RMSE during the validation period for each case is summarized in Table 5. The efficiency of Case 2 was not as good as the conventional calibration scheme (Case 1), but the other three cases (3, 4, and 5) were superior to Case 1. This indicates that the FDC model (Case 2) played a limited role in reducing bias, especially for correcting unseen data during the validation period. This may have been due to non-stationary flow, which in turn led to a change in the FDC, as illustrated in Supplementary Figure S1. However, the autoregressive model-based bias-correction approach (Case 3) was most effective for both

Table 4 | Monthly water balance of simulated flows of Cases 1, 2, 3, 4, and 5 during the calibration period

Month	Water balance					
	Obs.	Case 1	Case 2	Case 3	Case 4	Case 5
Jan	25,939	25,358	25,149	25,438	23,464	25,347
Feb	23,612	22,955	22,857	23,202	21,446	22,945
Mar	17,259	16,130	15,709	16,521	15,025	16,550
Apr	11,691	10,447	10,172	10,947	9,680	10,360
May	8,071	7,225	7,185	7,550	6,512	6,613
Jun	4,977	5,027	5,229	4,833	4,015	3,976
Jul	4,270	4,571	4,875	4,389	3,653	3,358
Aug	5,094	5,845	5,956	5,469	4,621	4,471
Sep	6,834	7,508	7,558	7,269	6,264	7,283
Oct	13,664	14,879	14,702	14,327	12,971	15,478
Nov	20,028	21,332	21,108	20,826	19,188	21,742
Dec	25,575	26,343	26,513	26,027	24,162	26,522
Total	167,014	167,620	167,014	166,799	151,003	164,643

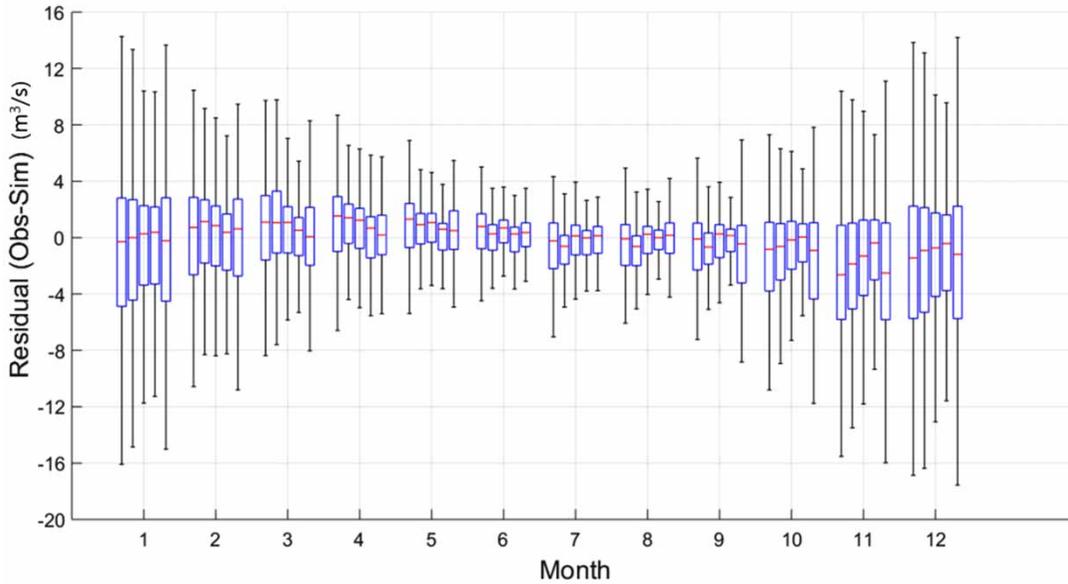


Figure 8 | Comparison of monthly residuals for each case during the validation period. The boxplots of each month are the result of Cases 1, 2, 3, 4, and 5 (from left to right).

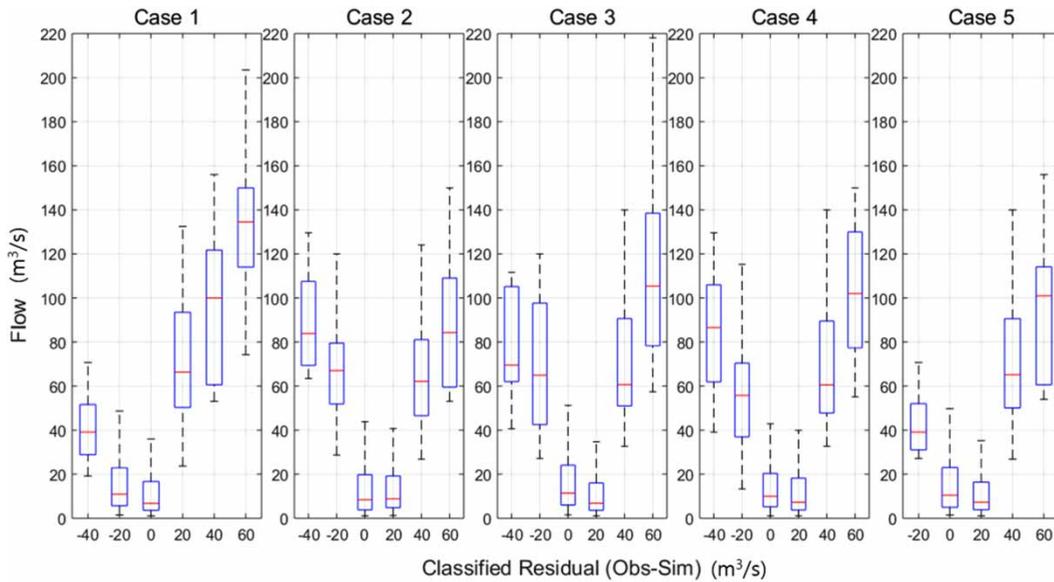


Figure 9 | Boxplots for observed flow with respect to the magnitude of residual for five different cases during the validation period. Labels on the x-axis indicate the upper boundary of the residual range from -60 to $60 \text{ m}^3 \text{ s}^{-1}$.

Table 5 | Comparison of RMSE for Cases 1, 2, 3, 4, and 5 during the validation period ($\text{m}^3 \text{ s}^{-1}$)

Period	RMSE				
	Case 1	Case 2	Case 3	Case 4	Case 5
Validation (1991–2008)	8.256	8.339	7.684	8.108	8.083

calibration and validation periods. Table 6 presents the monthly water balance during the validation period for each case. Unlike the results reported in Table 4, the water balance of Case 2 represented a large difference in the residual compared with the other four cases on an annual basis. This may be because the temporal pattern of the bias-

Table 6 | Monthly water balance of simulated flows of Cases 1, 2, 3, 4, and 5 during the validation period

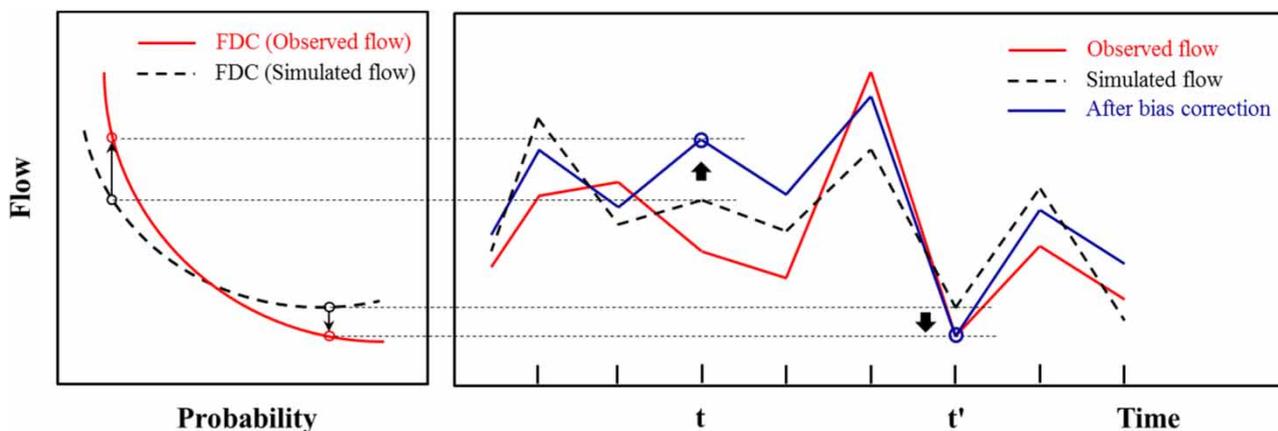
Month	Water balance					
	Obs.	Case 1	Case 2	Case 3	Case 4	Case 5
Jan	17,150	16,156	15,043	16,455	15,422	16,214
Feb	12,317	12,008	11,017	12,045	15,534	12,040
Mar	10,396	9,697	8,884	9,873	13,957	10,031
Apr	7,402	6,682	6,209	6,864	12,883	7,011
May	4,801	4,256	4,065	4,375	7,977	4,371
Jun	3,847	3,692	3,597	3,603	6,575	3,602
Jul	3,462	3,761	3,665	3,571	4,263	3,460
Aug	3,312	3,632	3,553	3,475	3,675	3,384
Sep	4,168	4,471	4,332	4,278	3,517	4,726
Oct	9,030	9,275	8,670	9,063	4,853	9,360
Nov	14,374	15,632	14,488	15,070	7,358	15,481
Dec	17,714	17,509	16,603	17,540	12,718	17,456
Sum	107,974	106,770	100,126	106,212	108,732	107,134

corrected flow from the FDC was relaxed in the sense that the FDC was constructed in a probability space rather than a temporal domain. Moreover, monthly differences in water balance were substantially improved by Case 3, among the other cases.

DISCUSSION AND CONCLUSIONS

The FDC model effectively reduced biases during the calibration period but was ineffective with respect to both RMSE and water balance during the validation period. As

illustrated in Figure 10, the empirical FDCs were applied for adjusting biases in the simulated flow against the observed flow, and in such cases, it may not be appropriate to correct point-wise biases over time, leading to degradation of model efficiency. For example, at time t , the flow after bias correction moved further away from the observation, which led to a larger error. However, at the time t' , the bias-corrected flow approached the observed value. In this study, FDCs were constructed for the entire period of observed and simulated flow. Building FDCs in several groups associated with temporal scales can minimize the impact of this issue. More specifically, flow data for the

**Figure 10** | Schematic of the bias-correction approach using the FDC_model (Case 2) and two different examples for underestimation and overestimating in the process of bias correction.

construction of the FDC curve can be divided into different temporal scales, such as biannual, seasonal, or monthly periods, to build FDCs. By comparing (or optimizing), the results of different periods of FDC models, the optimal period that could reduce biases for both the calibration and validation periods can be identified, which is another research topic.

In this study, an AR(1) first-order autoregressive process was applied and the obtained results confirmed the usefulness of reducing the bias of simulated flow by examining the hypothesis that the residuals were autocorrelated, but further investigation is needed. For example, autoregressive processes can be expanded to a general form and an autoregressive integrated moving-average model, and the most appropriate form of the AR model can be explored by comparing Akaike's information criterion or Bayesian information criterion.

Hydrological models are imperfect representations of real hydrological circulation due to incomplete model structure, insufficient data, measurement errors, and inadequate initial and boundary conditions. The main objective of this study was to assess whether post-processing methods can reduce the biases in hydrological model outputs. Results of three bias-correction methods were compared with the conventionally calibrated flow and sub-annually calibrated flow. The main conclusions can be summarized as follows:

1. Monthly residuals were compared across four different cases. Generally, there was a clear seasonality representing a more substantial variability in winter than summer. The range of the residual usually expanded in winter, demonstrating that the HYMOD may not reproduce high flows adequately. For a systematic comparison, the reduction ratio in the range of the residual with respect to Case 1 was evaluated. An overall reduction in the residuals of bias-corrected flows (Cases 2, 3, and 4) was clear. A substantial decrease in the range of residual was identified in Cases 3 and 4. However, the residual from Case 4 showed an increasing tendency toward positive values, and the efficacy of treating biases may be limited.
2. This study examined the effects of bias correction on adjusting residual distribution with respect to the magnitude of the flow. There was a clear increasing tendency for the residual as the magnitude of the flow increased, as seen in the monthly distribution of residuals. The positive residual related to high flow showed a decreasing trend, whereas the negative residual associated with high flow appeared to increase after the bias correction. There is evidence of the shift from a positive bias toward negative bias and *vice versa*, with the effect of bias correction being most pronounced at high flows, resulting in a decrease in the positive residual (or an increase in the negative residual).
3. This study confirmed that the bias-corrected flows with three cases were more effective than the baseline model (Case 1), in terms of correcting a systematic error in simulated flow. The residual distribution converged to zero with reduced variability, indicating the efficacy of bias correction. A significant decrease in negative residual with high flow appeared in Cases 2, 3, and 4. However, negative residuals in a certain range were slightly higher due to the shift from a positive residual after bias correction.
4. A more detailed comparison with the RMSE during the calibration and validation periods revealed that the RMSE in Cases 1 and 2 were similar to each other, while a clear reduction in the RMSE was seen in Cases 3 and 4 during the calibration phase. On the other hand, Case 3 was more effective than the other three cases, which were comparable with each other in terms of the RMSE during the validation phase. A comparison of the RMSE over different flow regimes demonstrated that Case 3 appeared the most effective in correcting systematic biases in the entire flow regime. Finally, monthly water balances for all the cases were compared during both calibration and validation periods. The monthly water balance of Case 3 was close to that of the observed value, whereas Case 2 showed considerable differences in the residual compared with the other three cases.
5. An AR model (Cases 3 and 4) showed relatively good efficiency than the conventional calibration method (Case 1) and the FDC model (Case 2) at correcting systematic biases both in calibration and validation periods. Here, the AR model is only applicable when the lagged residual is obtained from observations at the previous time step. For example, the AR model cannot be applied to correct the simulated flow in the future period (e.g., 2050–2100).

To alleviate this issue, the non-stationary nature of the FDC needs to be explored for better representation of time-varying behavior in the residual, including the identification of the optimal temporal scales for the construction of an FDC curve within an optimization framework. Moreover, future research should include a rigorous comparison of the bias-correction approach under different conditions (e.g., different hydrological models, catchment size, climatic conditions, land use, and terrain) to provide a general guideline.

6. A sub-annual calibration scheme (Case 5) was introduced to explore whether an improved calibration framework can be more effective than the post-processing approach to the simulated flow. The results showed that this model performance did not show much improvement compared with three post-processing approaches.

ACKNOWLEDGEMENTS

This work was funded by the Korea Meteorological Administration Research and Development Program under Grant KMI (KMI 2018-07010).

DATA AVAILABILITY STATEMENT

All relevant data are available from an online repository or repositories. The data used in this study are available from the UK Met Office (<https://www.metoffice.gov.uk/research/climate/maps-and-data/about/archives>) and the National River Flow Archive (<https://nrfa.ceh.ac.uk/data/station/info/45001>).

REFERENCES

- Abbaspour, K. C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J., Zobrist, J. & Srinivasan, R. 2007 Modelling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. *Journal of Hydrology* **333** (2–4), 413–430.
- Bekele, E. G. & Nicklow, J. W. 2007 Multi-objective automatic calibration of SWAT using NSGA-II. *Journal of Hydrology* **341** (3–4), 165–176.
- 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** (2), 95–102.
- Boyle, D. P., Gupta, H. V. & Sorooshian, S. 2003 Multicriteria calibration of hydrologic models. In: *Calibration of Watershed Models* (Q. Duan, H. Gupta, S. Sorooshian, A. Rousseau, & R. Turcotte, eds). American Geophysical Union, AGU, Washington DC, pp. 185–196.
- Brown, J. D. & Seo, D.-J. 2010 A nonparametric postprocessor for bias correction of hydrometeorological and hydrologic ensemble forecasts. *Journal of Hydrometeorology* **11** (3), 642–665.
- Brown, J. D. & Seo, D. J. 2013 Evaluation of a nonparametric postprocessor for bias correction and uncertainty estimation of hydrologic predictions. *Hydrological Processes* **27** (1), 83–105.
- De Vos, N., Rientjes, T. & Gupta, H. 2010a Diagnostic evaluation of conceptual rainfall–runoff models using temporal clustering. *Hydrological Processes* **24** (20), 2840–2850.
- De Vos, N., Rientjes, T. & Gupta, H. V. 2010b Diagnostic evaluation of conceptual rainfall–runoff models using temporal clustering. *Hydrological Processes* **24** (20), 2840–2850.
- Eckhardt, K. & Arnold, J. 2001 Automatic calibration of a distributed catchment model. *Journal of Hydrology* **251** (1–2), 103–109.
- Eckhardt, K., Fohrer, N. & Frede, H. G. 2005 Automatic model calibration. *Hydrological Processes: An International Journal* **19** (3), 651–658.
- Freer, J. E., McMillan, H., McDonnell, J. & Beven, K. 2004 Constraining dynamic TOPMODEL responses for imprecise water table information using fuzzy rule based performance measures. *Journal of Hydrology* **291** (3–4), 254–277.
- Green, C. & Van Griensven, A. 2008 Autocalibration in hydrologic modeling: using SWAT2005 in small-scale watersheds. *Environmental Modelling and Software* **23** (4), 422–434.
- Gupta, H. V., Sorooshian, S. & Yapo, P. O. 1998 Toward improved calibration of hydrologic models: multiple and noncommensurable measures of information. *Water Resources Research* **34** (4), 751–763.
- Kim, K. B. & Han, D. 2017 Exploration of sub-annual calibration schemes of hydrological models. *Hydrology Research* **48** (4), 1014–1031.
- Kim, T.-J., Kwon, H.-H. & Lima, C. 2018 A Bayesian partial pooling approach to mean field bias correction of weather radar rainfall estimates: application to Osungsan weather radar in South Korea. *Journal of Hydrology* **565**, 14–26.
- Krzysztofowicz, R. & Kelly, K. S. 2000 Hydrologic uncertainty processor for probabilistic river stage forecasting. *Water Resources Research* **36** (11), 3265–3277.
- Krzysztofowicz, R. & Maranzano, C. J. 2004 Hydrologic uncertainty processor for probabilistic stage transition forecasting. *Journal of Hydrology* **293** (1–4), 57–73.
- Madadgar, S., Moradkhani, H. & Garen, D. 2014 Towards improved post-processing of hydrologic forecast ensembles. *Hydrological Processes* **28** (1), 104–122.

- Madsen, H. 2000 Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. *Journal of Hydrology* **235** (3–4), 276–288.
- Montanari, A. & Brath, A. 2004 A stochastic approach for assessing the uncertainty of rainfall-runoff simulations. *Water Resources Research* **40**, 1.
- Moore, R. 1985 The probability-distributed principle and runoff production at point and basin scales. *Hydrological Sciences Journal* **30** (2), 273–297.
- Muleta, M. K. & Nicklow, J. W. 2005 Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. *Journal of Hydrology* **306** (1–4), 127–145.
- Pushpalatha, R., Perrin, C., Le Moine, N. & Andréassian, V. 2012 A review of efficiency criteria suitable for evaluating low-flow simulations. *Journal of Hydrology* **420**, 171–182.
- Seo, D.-J., Herr, H. & Schaake, J. 2006 A statistical post-processor for accounting of hydrologic uncertainty in short-range ensemble streamflow prediction. *Hydrology and Earth System Sciences Discussions* **3** (4), 1987–2035.
- Shi, X., Wood, A. W. & Lettenmaier, D. P. 2008 How essential is hydrologic model calibration to seasonal streamflow forecasting? *Journal of Hydrometeorology* **9** (6), 1350–1363.
- Tolson, B. A. & Shoemaker, C. A. 2007 Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research* **43**, 1.
- Van Griensven, A. & Meixner, T. 2007 A global and efficient multi-objective auto-calibration and uncertainty estimation method for water quality catchment models. *Journal of Hydroinformatics* **9** (4), 277–291.
- Vrugt, J. A., Bouten, W., Gupta, H. V. & Sorooshian, S. 2002 Toward improved identifiability of hydrologic model parameters: The information content of experimental data. *Water Resources Research* **38** (12), 48-1–48-13.
- Vrugt, J. A., Gupta, H. V., Bouten, W. & Sorooshian, S. 2003 A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research* **39**, 8.
- Wagener, T., Boyle, D. P., Lees, M. J., Wheater, H. S., Gupta, H. V. & Sorooshian, S. 2001a A framework for development and application of hydrological models. *Hydrology and Earth System Sciences* **5** (1), 13–26.
- Wagener, T., Boyle, D. P., Lees, M. J., Wheater, H. S., Gupta, H. V., Sorooshian, S. *et al.* 2001b A framework for development and application of hydrological models. *Hydrology and Earth System Sciences Discussions* **5** (1), 13–26.
- Wagener, T., McIntyre, N., Lees, M., Wheater, H. & Gupta, H. V. 2003 Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis. *Hydrological Processes* **17** (2), 455–476.
- Weerts, A., Winsemius, H. & Verkade, J. 2011 Estimation of predictive hydrological uncertainty using quantile regression: examples from the National Flood Forecasting System (England and Wales). *Hydrology and Earth System Sciences* **15** (1), 255–265.
- Wood, A. W. & Schaake, J. C. 2008 Correcting errors in streamflow forecast ensemble mean and spread. *Journal of Hydrometeorology* **9** (1), 132–148.
- Ye, A., Duan, Q., Schaake, J., Xu, J., Deng, X., Di, Z., Miao, C. & Gong, W. 2015 Post-processing of ensemble forecasts in low-flow period. *Hydrological Processes* **29** (10), 2438–2453.
- 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**, 12.
- Zhang, X., Srinivasan, R. & Bosch, D. 2009 Calibration and uncertainty analysis of the SWAT model using Genetic Algorithms and Bayesian model averaging. *Journal of Hydrology* **374** (3–4), 307–317.
- 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.

First received 3 March 2020; accepted in revised form 10 November 2020. Available online 4 January 2021