Hydrological model optimization using multi-gauge calibration (MGC) in a mountainous region

Sead Ahmed Swalih and Ercan Kahya

ABSTRACT

It is a challenge for hydrological models to capture complex processes in a basin with limited data when estimating model parameters. This study aims to contribute in this field by assessing the impact of incorporating spatial dimension on the improvement of model calibration. Hence, the main objective of this study was to evaluate the impact of multi-gauge calibration in hydrological model calibration for Ikizdere basin, Black Sea Region in Turkey. In addition, we have incorporated the climate change impact assessment for the study area. Four scenarios were tested for performance assessment of calibration: (1) using downstream flow data (DC), (2) using upstream data (UC), (3) using upstream and downstream data (Multi-Gauge Calibration – MGC), and (4) using upstream and then downstream data (UCDC). The results have shown that using individual gauges for calibration (1 and 2) improve the local predictive capacity of the model. MGC calibration significantly improved the model performance for the whole basin unlike 1 and 2. However, the local gauge calibrations statistical performance, compared to MGC outputs, was better for local areas. The UCDC yields the best model performance and much improved predictive capacity. Regarding the climate change, we did not observe an agreement amongst the future climate projections for the basin towards the end of the century.

Key words | climate change, hydrological modelling, multi-gauge calibration, SWAT

Sead Ahmed Swalih (corresponding author)
Ercan Kahya
Civil Engineering Department,
Istanbul Technical University,
Istanbul,
Turkey
E-mail: *seahhu@yahoo.com*

HIGHLIGHTS

- Four calibration techniques were investigated: (1) downstream calibration (DC), (2) upstream calibration (UC), (3) calibration with both upstream and downstream data (MGC), and (4) calibration with first from upstream data, then downstream data sequentially (UCDC).
- UC and DC improve the predictive capacity of the model only for the region where calibration data is used. They gave a better statistical performance for the particular region compared with MGC and UCDC.
- MGC calibration significantly improved the model performance for the whole basin unlike the local gauge calibrations.
- UCDC technique gave best model performance, where the model performance improved for the whole study area far better than the other calibration techniques.
- No agreement observed amongst future climate projections for the study area.

doi: 10.2166/hydro.2020.034

INTRODUCTION

Hydrologic models have been widely used in hydrological research and water resource management studies. Hydrological models are increasingly used to simulate changes in the basin management, not only for investigating the impacts of external influences, but also for discovering the impacts of future data series (observing climate change, drought and flood studies) (Zhang *et al.* 2016). The physical hydrological model, namely Soil and Water Assessment Tool (SWAT), is a river basin scale model requiring specific information about weather, soil properties, topography, vegetation, and land management practices occurring in a basin (Arnold *et al.* 1998; Neitsch *et al.* 2011; Winchell *et al.* 2013). It has gained international acceptance as a robust interdisciplinary basin modeling tool. However, calibration has been one of the major challenges in developing hydrological models in such a way that simulations will be in good agreement with observed flows. The SWAT model calibration has been improving due to continuing advancements in the calibration techniques (Van Griensven *et al.* 2006; Arnold *et al.* 2012a, 2012b; Ercan *et al.* 2014).

In past decades, many calibration methods have been developed for SWAT, including manual calibration and automated procedures using the shuffled complex evolution (SCE) method and others. Arnold *et al.* (2012a, 2012b) developed the SWAT-CUP program that facilitates the calibration and validation of SWAT model using the semi-automated approach (SUFI2) (Abbaspour 2015). The sensitivity analysis of SWAT-CUP helps check model parameters that have a high impact on the simulations and make the necessary changes accordingly and model parameters must only be calibrated in a reasonable range that could give meaning on the ground (Van Griensven *et al.* 2006). For calibration studies, it is important to spatially account for hydrological processes in order to improve the efficiency of model run time and incorporate the impact of uncertainty in the model (Klemes 1986). This task is one of the major challenges that enhances SWAT model calibration. Therefore, the aim of this study is to investigate the impact of incorporating hydrological processes spatially in order to improve model calibration. In addition, the impact of the model calibration techniques on the hydrological modeling of the basin and climate change impact assessment will be studied.

LITERATURE REVIEW

A number of attempts have been conducted by hydrologists to incorporate model parameters spatially to improve model accuracy. In 2005, a multi-variable and multi-site approach to the calibration and validation of the SWAT model for the Motueka catchment, New Zealand, was conducted and found that the method improved the model calibration and validation (Nash–Sutcliffe coefficient of 0.78 and 0.72 respectively) (Cao *et al.* 2005). In another study, spatially distributed calibration at sub-basin level and temporal validation at the stream gauges outlet points was conducted to incorporate the spatial and temporal hydrological patterns for two river basins in the Ohio, USA, region and found out that spatially distributed calibration and validation of the basins improve the model predictive capacity (Santhi *et al.* 2008). Similar studies conducted by Lu *et al.* (2015), Bai *et al.* (2017) and Zhang *et al.* (2017) in the Yingluoxia, Miyun and Baihe basins, China, comparing single- and multi-site calibration and validation of the SWAT model, showed that the parameters determined from multi-site calibration and validation were far better than those from the single site calibrations. Haas *et al.* (2016) studied a joint multi-metric calibration of flow simulated by SWAT in Treene basin, Germany. The results showed that the approach used has improved the calibrated predictions, where adequate model runs with good performance for different hydrological conditions for the flow were detected. Contrary to the above, Shresthaa *et al.* (2016) indicated that multi-site calibration did not improve simulations of flow and sediments compared to single-site calibration.

Changes have occurred in several aspects of the atmosphere and surface that alter the global energy budget of the Earth causing the climate to change (Solomon *et al.* 2007). The Intergovernmental Panel on Climate Change (IPCC), set up by the World Meteorological Organization (WMO) to provide information on climate change, has confirmed with high confidence that the gradual increase in the average temperature of the Earth's surface is due to the growing concentration of greenhouse gases (GHGs) in the atmosphere, attributed to burning of fossil fuels and changes in land use and land cover (IPCC 2001). Today, Regional Circulation Models (RCMs) are applied to determine the likely effects of these changes in the GHGs. It is to be mentioned here that a GCM/RCM is a Global/Regional Circulation Model that could depict different pictures for the future weather over a region, and consequently affect the hydrologic regime of basins (Elshamy *et al.* 2009). Increases have occurred in the number of heavy precipitation events. The availability of observational data restricts the types of extremes that can be analyzed. Confidence in these estimates is higher for some climate variables (e.g. temperature) than for others (e.g. precipitation).

Some studies have been conducted to investigate the effects of multiple stream gauging stations' data on basin-wide calibration and validation in Turkey. The impact of size of sub-basins on the hydrologic parameters and their spatial variability in the estimation of the hydrologic parameters in the western Black Sea region were studied by Kocyigit *et al.* (2017) and Akay *et al.* (2018). The studies indicated that multi-site calibration and validation resulted in satisfactory outcomes for the direct flow hydrograph but not the peak flow prediction. These studies have only focused on the application of hydrological models, and very limited attempts were made in the model parametrization and calibration technique studies for basins in Turkey. Therefore, there is an obvious need to assess the impact of spacial dimension and calibration and validation techniques in hydrological modeling.

Consequently, the main objective of this study is to evaluate the impact of multi-site calibration on the performance of hydrological model calibration and validation. To achieve this, we used four different calibration techniques: (1) calibration using gauge data only from the downstream area (DC); (2) calibration using gauge data only from the upstream area (UC); (3) calibration using gauge data from both the upstream and downstream areas of the basin (Multi Gauge Calibration – MGC); and (4) calibration using gauge data first from the upstream area only, then using the data from the downstream area sequentially (UCDC). We analyzed the impact of these techniques on the hydrologic regime of the basin. Hydrological model simulations and statistical analysis, as well as climate change impact assessments, have been used in the study.

STUDY AREA AND DATA

The Rize province is located in the northeastern Anatolian mainland of Turkey, so-called the eastern Black Sea region, with an area of 3,920 km^2. The average annual precipitation in the region is nearly 2,250 mm and total annual runoff is about 2,745 million m^3 based on the reference period 1960–1996 (Sen & Kahya 2017). The study area is a mountainous region with elevations reaching higher than 3,000 m (Figure 1) and a very wet climate (Kahya *et al.* 2008a). In addition, Rize province falls into the homogeneous streamflow region (Kahya *et al.* 2008b).

Digital elevation model (DEM), soil and land use maps

The Digital Elevation Model (DEM), the soil map, and the land use map of the basin were obtained from Istanbul Technical University (ITÜ) – Hydraulics and Water Resource Engineering Laboratory. The digital elevation model (DEM) with 28-meter spatial resolution was used which is quite sufficient for hydrological modelling studies. A total of 22 sub-basins were identified after delimiting the basin boundary in ArcSWAT (Figure 1).

DATA AND METHODOLOGY

Weather and hydrological data

The observed precipitation data was accessed from the Turkish Meteorological Directorate (MGM). We used the maximum and minimum temperature data from the Climate Forecast System Reanalysis (CFSR) which was proven to have good accuracy with the observed temperature (Duana *et al.* 2019). The precipitation data in the vicinity of the Ikizdere basin was used for the SWAT simulations. In addition, the daily flow data of three flow gauging stations found in Ikizdere basin were received from the Turkish General

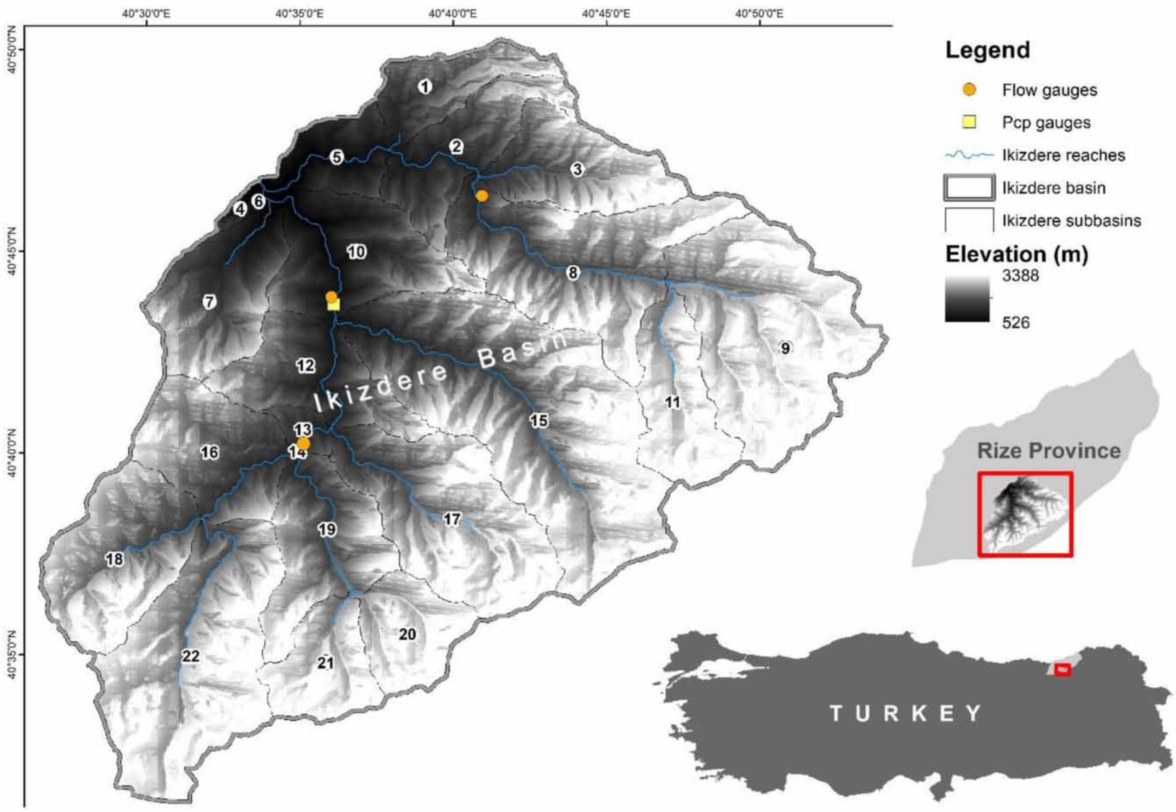

Figure 1 | The Rize Province, delineation of the study area, and sub-basins of Ikizdere basin.

Directorate of State Hydraulic Works (DSI) (for the years 1971–1996) to carry out the model calibration and validation. The flow gauging stations of Camlıkderesi or Ikizdere (station no. E22A015) and Tözköyderesi (station no. E22A033) were used to assess the impacts of spatial streamflow data on the overall model optimization performance. The third gauge of Cimil-Koknar (station no. D22A096) in Ikizdere basin was insufficiently small to be used for the study. After failing to simulate the Kaptanpaşa basin of Rize province within acceptable accuracy, due to a shortage of observed climate and streamflow data, the Ikizdere basin was selected for our study as sufficient data is available for hydrological modelling, which is also recommended by Kahya *et al.* (2015). Moreover, the mean monthly values of temperature, wind speed, solar radiation and humidity were generated from the Climate Forecast System Reanalysis (CFSR) data, which can be accessed from the website of US National Centers for Environmental Prediction (NCEP-CFSR). The climate change data for the three climate models and two scenarios were downloaded from the official IPCC data repository (www.ipcc-data.org/sim/gcm_clim/SRES_AR4/index.html)

Model calibration and sensitivity analysis

SWAT is a river basin scale model developed to predict the impact of basin management on water, sediment, nutrients, and agricultural and chemical yields (Arnold *et al.* 1998). In this study, the ArcSWAT program was used in setting up the model for the study area. The observed precipitation data from the Turkish Meteorological agency stations was employed in the SWAT model calibration. The model uses only one climate station in each sub-basin when a single station is available within the sub-basin; otherwise, it demands the mean data of nearby stations. Once the delineation of the basin is completed, Hydrological Response Units (HRU) were categorized with respect to the soil, land use and slope maps overlay. Categorizing all the similar parcels of land with similar hydrology, SWAT employs a lumped calculation for parameters instead of a distributed calculation for the sake of saving time and resources. Once the HRU and sub-basin analysis is completed, the ArcSWAT interface then inputs climate data (such as precipitation and temperature) to finalize the SWAT model setup. The two streamflow series in Camlıkdere station located at the downstream and Tözköydere station located at the upstream of the study area were used for the model calibration and validation.

The SWAT model setup based on the study area was simulated using precipitation data for the period 1971–1996. The observed precipitation was assumed the only dominant climate variable that has direct and significant impacts on streamflow downstream. The simulations were run on a monthly time step and a warm-up period of three years for the model to adjust itself according to the data and basin conditions. The output of the model in the warm-up period was not used for further analysis. Following successful setting procedures for SWAT, it was calibrated and validated using the remaining historical streamflow data. More specifically, the first three years (1971–1973) were used in the model warm-up, and the next 12 years (1974–1990) in the model calibration, and finally the remaining six years (1991–1996) for the model validation. The model was calibrated using the following four methods: the first one using Ikizdere streamflow data (Downstream Calibration – DC), the second using Tözköydere streamflow data (Upstream Calibration – UC), the third using both Ikizdere and Tözköydere streamflow data (Multi-Gauge Calibration – MGC), and finally the fourth calibration using Tözköydere streamflow before the calibration using Ikizdere streamflow data (Upstream Calibration before Downstream Calibration – UCDC). From the review of relevant literatures, the technique employed in this research is quite different from what has been done in the past (at least for the Black Sea region). Thus, it will have a significant contribution in highlighting the significance of incorporating spatial dimension in hydrological model parametrization studies.

In the phase of model calibration and sensitivity analysis, we adopted the SWAT-CUP auto-calibration tool (Abbaspour 2015). The Latin Hypercube One-factor-At-a-Time design (LH-OAT) method proposed by Morris (1991) was implemented to carry out the sensitivity analysis method which uses a stratified sampling approach that better covers the sampling hypercube with fewer samples (White & Chaubey 2005; Van Griensven *et al.* 2006). The results are the relative sensitivities based on the linear approximations, providing only partial information about the sensitivity of objective function to the model parameters. In general, the larger the absolute value of *t*-stat and the smaller the *p*-value, the more sensitive is the parameter. The most sensitive model parameters are essential in model calibration as these parameters will affect the model more than less sensitive parameters. The four model calibration methods were tested for performance. In each method, a set of five calibration steps were conducted until the simulated streamflow at a desirable performance is attained as compared to the observed streamflow. It is important to note that each calibration step consists of 500 iteration runs. The final step was validation to evaluate the performance of the calibrated model with a separate precipitation forcing.

Performance analysis

To evaluate the performance of the model calibration techniques, the following five statistical methods were used: (i) Nash–Sutcliff Efficiency (NS), which measures the closeness of the simulated value to the observed value (Equation (1)); (ii) Coefficient of Determination (R^2) (Equation (2)); (iii) Root Mean Square Error (RMSR) (Equation (3)); (iv) Percentage Bias (PBIAS) (Equation (4)), and finally (v) *p*-factor (the percentage of observed time series covered by the 95% prediction uncertainty, or in short 95PPU), which is a measure of uncertainty in the model (input data, parameter, model, etc.). The NS, a major objective function criterion, was used to assess the degree of fitness exhibited by the simulation with that observed. These methods are discussed by Moriasi *et al.* (2007) in detail. In addition, Molina-Navarro *et al.* (2017) advised the use of NS as an objective function when addressing a multi-site and multi-variable calibration studies:

$$\mathrm{NS} = 1 - \left[\frac{\sum_{i=1}^{n}\left(Y_i^{\mathrm{obs}} - Y_i^{n}\right)^2}{\sum_{i=1}^{n}\left(Y_i^{\mathrm{obs}} - Y_i^{\mathrm{mean}}\right)^2}\right] \tag{1}$$

$$R^2 = \left[\frac{n\sum(G_iS_i) - (\sum G_i)(\sum S_i)}{\sqrt{\left(n(\sum G_i^2) - (\sum G_i)^2\right)\left(n(\sum S_i^2) - (\sum S_i)^2\right)}}\right]^2 \tag{2}$$

$$\mathrm{RSR} = \frac{\mathrm{RMSE}}{\mathrm{STDEV_{obs}}} = \frac{\sqrt{\sum_{i=1}^{n}(Y_i^{\mathrm{obs}} - Y_i^{n})^2}}{\sqrt{\sum_{i=1}^{n}(Y_i^{\mathrm{obs}} - Y^{\mathrm{mean}})^2}} \tag{3}$$

$$\mathrm{PBIAS} = 100 * \left[\frac{\sum_{i=1}^{n}(Y_i^{\mathrm{obs}} - Y_i^{n})}{\sum_{i=1}^{n}(Y_i^{\mathrm{obs}})}\right] \tag{4}$$

where Y_i^{obs} is the ith observation; Y_i^n is the ith simulated value; Y_i^{mean} is the mean of observed rainfall data; G_i is ground precipitation measurements; S_i is satellite precipitation estimates, and n is the total number of data pairs or observations.

Climate change assessment

In this study, three regionally downscaled GCM climate models (MIMR, INCM3 and HADCM3) and two scenarios (A2 and A1B) were used. It is evident that the three climate models are not consistent in their estimation of future precipitation change. Except for the A1B scenario of INCM3, all the other climate change scenarios predicted the same or decreased precipitation across the seasons. The summary of IPCC climate change data used in our study is presented in Figure 2. The values indicate the projected precipitation in mm at the end of the century (2080–2099) with respect to the baseline period (1961–1990). We implemented the climate change analysis by directly substituting the percentage change in precipitation forecast by the climate model for each month of the year in the SWAT input/output files (TxtInOut folder). We copied the 'swat.exe' executable file to run the SWAT model from the TxtInOut folder after making the necessary change in all the sub-basin (.sub) files.

RESULTS AND DISCUSSION

In this study, we used the SWAT model using the SWAT-CUP model optimization tool to assess our four model optimization techniques. In addition, we assessed the impact of climate change for the study area.

Model calibration and sensitivity analysis

The SWAT model of the study area was set-up and calibrated using the observed flow data. The output of the model calibration, fitted parameter values and ranks of the sensitivity analysis are illustrated in Supplementary Material, Annex 1. Some of the parameters are significantly affected by the streamflow stations used in the model calibration. CH_N2 and CN parameters have shown to be sensitive, which indicated the direct impact of river channel characteristics on streamflow. In addition, SFTMP and SNO50COV were found to be sensitive which indicates the effect temperature on the flow regime of the basin. The majority of model parameters are less sensitive when it comes to the location of gauge data used for calibration.

Irrespective of the four model calibration techniques used, the best parameter values fitted for each calibration

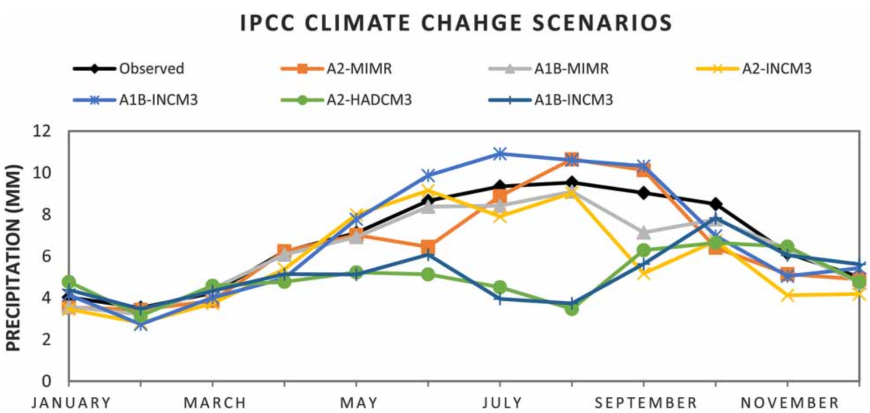

Figure 2 | The three climate change models and two scenarios of IPCC used in the study.

technique are quite similar along the calibration techniques used. The only exceptions are: CN2, CH_K2, LAT_TTIME, ALPHA_BF, GWQMN, PLAPS, SLSOIL and SOL_K. Most of these parameters are characterized by their effect on the river channel and soil characteristics.

Model performance

Five statistical parameters: NS, R2, RSR, PBIAS and *p*-factor, were calculated to evaluate the performance of the model and the calibration techniques (Table 1). According to Moriasi *et al.* (2007), model simulation can be judged as satisfactory if NS > 0.50 and RSR < 0.70, and if PBIAS < ±25% for streamflow analysis. The SWAT model was calibrated using four methods (DC, UC, MGC, and UCDC). The calibration with Tözköydere flow data (UC) gave a good NS parameter (0.8 and 0.7 for calibration and validation respectively). That is due to low flow intensity upstream of the basin caused by a smaller contribution area for the river outlet. This makes calibration easier for the objective functions to achieve a large number in a short period of time. The calibration with only Ikizdere flow (DC) gave the NS to be 0.7 both for calibration and validation. The DC calibration technique resulted in a smaller NS value due to a larger contribution area for flow causing high flow at the outlet. Maximizing the objective function (NS) for a larger basin is more difficult than a basin of small area. The third technique (MGC) used both Ikizdere and Tözköydere flow for the model calibration. The objective function for both gauging stations were lower than that achieved in the previous steps (0.7 and 0.6 for Ikizdere calibration and validation respectively, 0.8 and 0.6 for Tözköydere calibration and validation respectively). The fourth method calibrated the model using Tözköydere flow before calibrating with Ikizdere flow values. The final model parameters best fits calibrated using the Tözköydere flows were not changed when calibrating using Ikizdere flows. That way, the interdependence of the upstream and downstream flows was maintained. Although the objective function was lower for the third technique (MGC), the mean, standard deviation and PBIAS values were found to perform better than the first two calibration techniques. The best objective function performance was observed for the fourth technique (UCDC), which gave far better NS

Table 1 | The statistical evaluation of model parametrization using flow data for Ikizdere Basin

	DC[a]				UC[b]				MGC[c]				UCDC[d]			
	Ikizdere		Tözköydere		Ikizdere		Tözköydere		Ikizdere		Tözköydere		Ikizdere		Tözköydere	
Statistic parameter	Calib.	Valid.	Calib.	Valid.	Calib.	Valid.	Calib.	Valid.	Calib.	Valid.	Calib.	Valid.	Calib.	Valid.	Calib.	Valid.
NS	0.72	0.69	–	0.69	–	0.66	0.81	0.74	0.71	0.60	0.75	0.64	0.74	0.75	0.70	0.84
R^2	0.75	0.75	–	0.69	–	0.70	0.82	0.76	0.74	0.64	0.75	0.65	0.76	0.79	0.72	0.84
RSR	0.52	0.56	–	0.55	–	0.58	0.43	0.51	0.54	0.60	0.50	0.60	0.51	0.50	0.55	0.40
PBIAS	10.10	22.70	–	4.7	–	17.1	8.40	13.90	13.20	11.70	3.90	11.70	10.40	18.40	−0.10	6.20
p-factor	0.86	0.72	–	0.57	–	0.60	0.76	0.58	0.82	0.50	0.81	0.55	0.53	0.48	0.66	0.42
Mean_Sim (Mean_Obs)	11.8 (13.1)	10.7 (13.8)	–	6.19 (6.50)	–	11.4 (13.8)	5.9 (6.5)	5.6 (6.5)	11.4 (13.1)	10.6 (13.8)	6.2 (6.4)	5.7 (6.5)	11.7 (13.1)	11.3 (13.8)	6.4 (6.4)	6.1 (6.5)
StDev_Sim (StDev_Obs)	11.1 (11.4)	11.4 (12.9)	–	5.93 (7.17)	–	11.6 (12.9)	6.0 (6.9)	6.5 (6.9)	10.6 (11.4)	10.9 (12.9)	5.9 (6.7)	6.0 (7.2)	10.3 (11.4)	10.9 (12.9)	6.7 (6.7)	6.3 (7.2)

[a]The parameters were calculated for the flows of Ikizdere.
[b]The parameters were calculated for the flows of Tözköydere.
[c]The calibration was conducted using both Ikizdere and Tözköyere flows.
[d]The calibration was conducted using Tözköyere flow (upstream) before using the Ikizdere flows (downstream).

values both for the upstream and downstream regions of the basin. The other observation was calibrating the model using the upstream gauge data (UC) which gave more accurate model performance when compared to the calibration using the downstream flow data (DC). The calibration of the model with the flow data of Ikizdere resulted in improved NS values at the Tözköydere location too. However, the reverse did not show a similar outcome (Table 1).

Temporal performance

In the temporal analysis, the monthly average simulated flow values are plotted against time. The calibration of the model using the upstream flow (UC) (NS of 0.81 for calibration) compared to the calibration using the downstream flow (DC) (NS of 0.72 for calibration) demonstrated good accuracy (Figures 3 and 4). The main reason is that the upstream flow intensity is lower, making it less difficult for the algorithm to capture the distribution including the peak flow values. When the flow value increases downstream, it becomes difficult to estimate the flows, especially the peak flows resulting in lower calibration performance. When both the upstream and downstream flow data are used for model calibration simultaneously (MGC), the algorithm adjusts all model parameters providing the best optimization value (NS in our case) (Figure 5).

In addition, the temporal analysis proved the superiority of the fourth technique (UCDC) as compared to the first three techniques, since the hydrological model was optimized using two flow gauging stations one after the other (Figure 6). This allows the model to be optimized sequentially, starting from the upstream region down to the downstream region. The other techniques optimize the model with local gauge data contrary to the holistic approach implemented in UCDC. This results in loss of information about the characteristics of the basin which could otherwise be useful to improve the model performance. In the UCDS technique, we made sure not to change the parameters that were set using the upstream flow when calibrating using the downstream flow in order to maintain the inter-dependence of upstream–downstream flow.

Climate change assessment

The monthly average flow values for the simulated flow, as well as six climate change scenarios compared with the observed flow at Ikizdere sub-basin outlet, is plotted in Figure 7. There is no consistent trend in the forecasted flow for the various climate model precipitations. For the

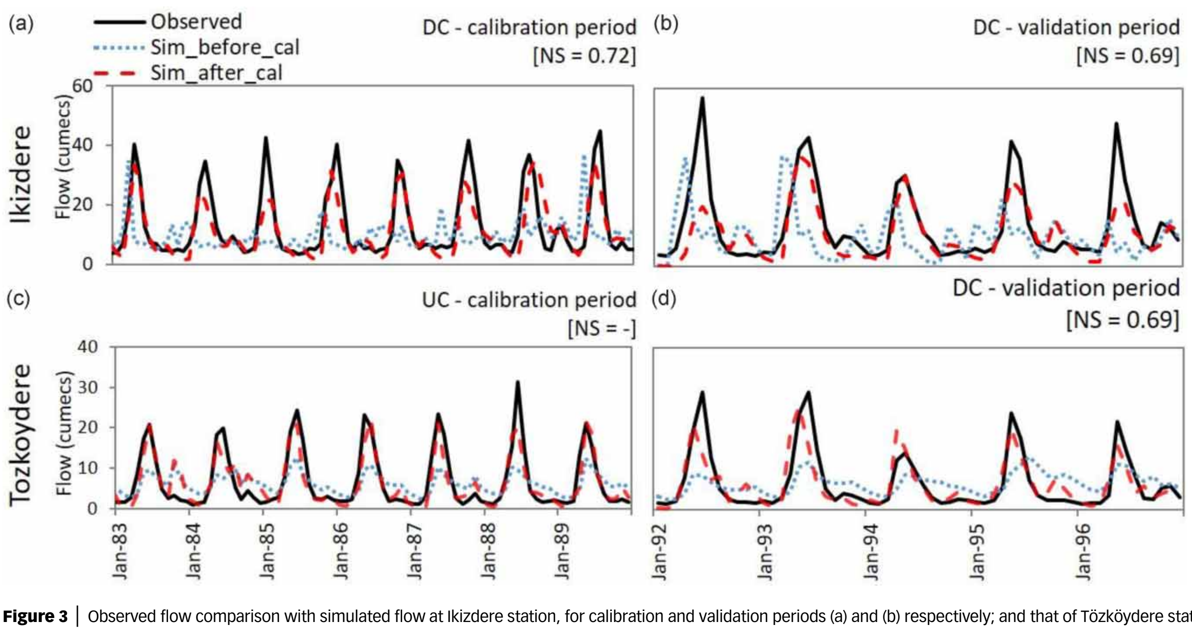

Figure 3 | Observed flow comparison with simulated flow at Ikizdere station, for calibration and validation periods (a) and (b) respectively; and that of Tözköydere station validation (c) and (d) (for DC – only Ikizdere flow used for model optimization).

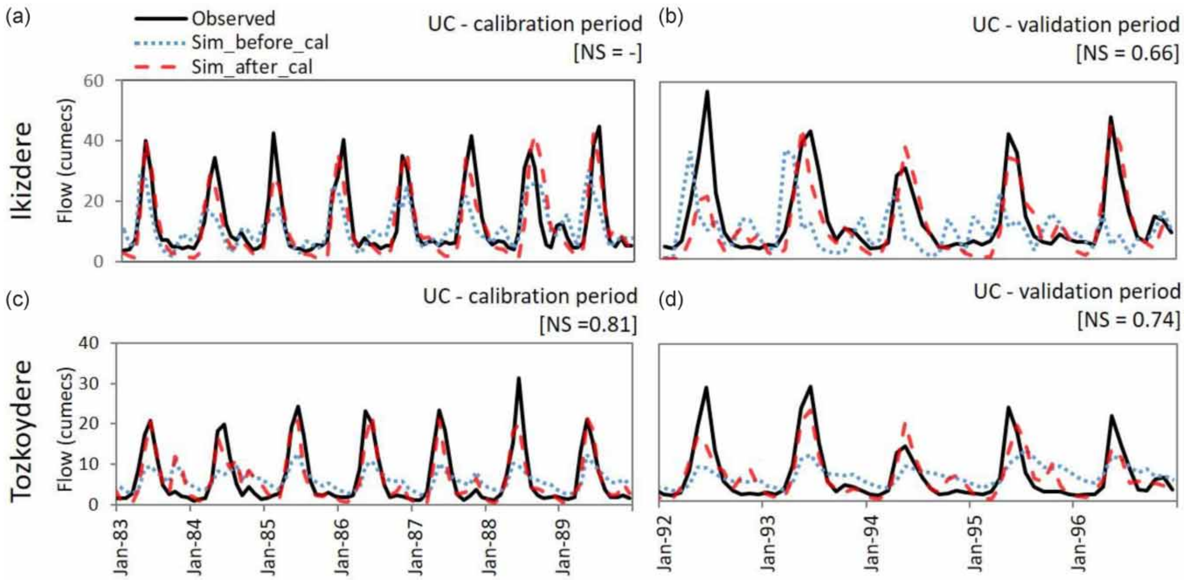

Figure 4 | Observed flow comparison with simulated flow at Ikizdere station, for calibration and validation periods (a) and (b) respectively; and that of Tozkoydere station validation (c) and (d) respectively (for UC – only Tözköydere flow used for model optimization).

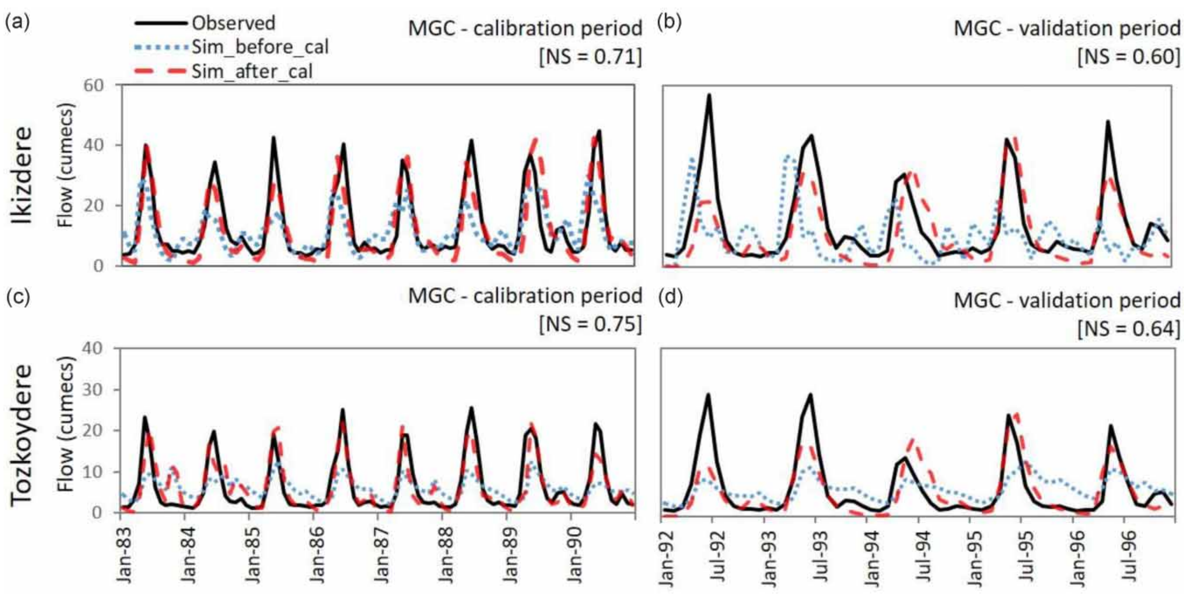

Figure 5 | Observed flow comparison with simulated flow after calibration and validation (a) and (b) respectively for Ikizdere station; (c) and (d) for Tözköydere station (for MGC – both Ikizdere and Tözköydere flow used for model optimization).

winter, all the climate models forecast a decrease in flow for the river. The other seasons forecasted a slight increase in flow. All in all, we did not observe an agreement amongst the climate projections forecast for the Rize basin.

CONCLUSIONS

It is quite challenging for a hydrological model to capture the complex processes in a basin, especially for mountainous regions with limited input data. Estimating the

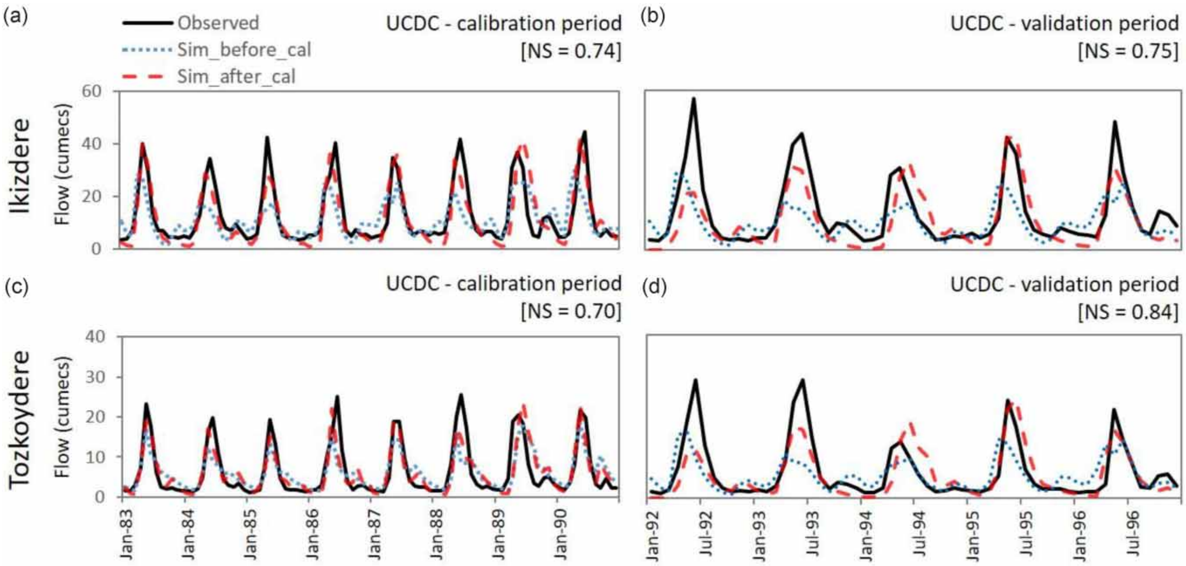

Figure 6 | Observed flow comparison with simulated flow after calibration and validation (a) and (b) respectively for Ikizdere station; (c) and (d) for Tözköydere station (for UCDC – Tözköydere flow used before using Ikizdere for model optimization).

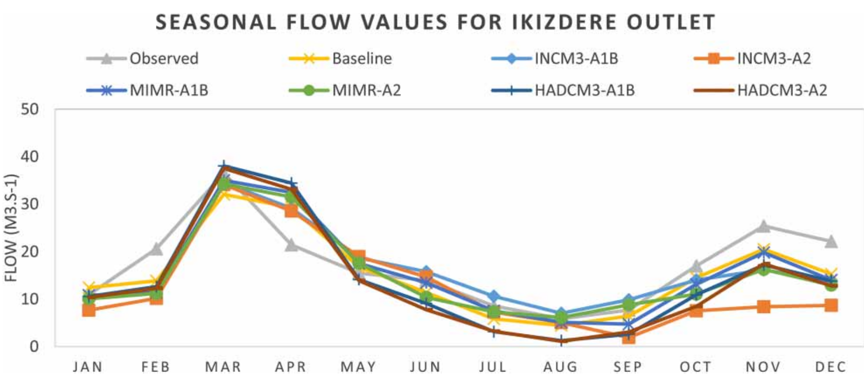

Figure 7 | Seasonal average flow of the observed, baseline simulation and simulations with climate change forcing at Ikizdere outlet.

parameters that represent the various hydrological processes is one of the main issues hydrologists need to solve since it is impossible to measure all of the hydrological parameters. In order to assess the impact of calibration techniques on the model performance, four methods were tested, calibration using flow data: (1) only from upstream area (US), (2) only from downstream area (DC), (3) using both upstream and downstream areas of the basin–multi-gauge calibration (MGC), and (4) first the upstream then downstream flow data (UCDC). The study was aimed at assessing the performance of multi-gauge calibration over single gauge calibration techniques.

The results of our analysis have shown that using individual gauges for calibration improve the predictive ability of the model for the particular region. The DC calibration technique resulted in a lower performance compared to the UC, due to a larger drainage area for the outlet causing high flow. In calibration, maximizing the objective function (NS) for a large basin area is more difficult than for a basin of small area. Model calibration using the MGC technique significantly improved model performance as a whole, unlike the single gauge calibrations. This is due to the fact that the model obtains much wider information on the basin characteristics to adjust its parameters when more than one gauging station is used simultaneously for calibration. The UCDC calibration technique gave the best model performance since the model performance was improved step by step starting from the upstream area going to the downstream area of the basin. This technique not only improved the model for each region, but also the interdependence of the two regions was maintained since the model parameters calibrated by the upstream flow were not changed when calibrating the model with the downstream flow. Our study findings are in agreement with what has been reported by Cao *et al.* (2005), Santhi *et al.* (2008), Haas *et al.* (2016), Lu *et al.* (2015), Zhang *et al.* (2017), Bai *et al.* (2017) and Akay *et al.* (2018). They all reported improvements in model calibration performance with multi-site (multi-gauge) calibration compared with single-gauge calibrations for their studies. Evaluating the results of previous studies and our study, we can confidently say multi-gauge calibration gives much better model calibration performance than single-gauge calibration techniques for the Ikizdere basin, Black Sea region. However, we still recommend more studies to be carried out on other basins to reach a comprehensive conclusion about multi-gauge calibration techniques for the Black Sea region. Regarding the climate change assessment, we have not observed a consistent trend in the forecasted flow by the various climate model predictions. The IPCC climate projections forecast did not show a visible shift in the hydrologic regime for the Rize basin.

ACKNOWLEDGEMENTS

We would like to acknowledge TUBİTAK (the Turkish Scientific and Technological Research Council) for providing the funds to carry out this study. We are also indebted to the Turkish State Meteorological Service (MGM) and General Directorate of Hydraulic Works (DSI) for the observed weather and flow data access.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

REFERENCES

Abbaspour, K. C. 2015 *SWAT-CUP: SWAT Calibration and Uncertainty Programs – A User Manual*. Eawag – Swiss Federal Institute of Aquatic Science and Technology, Switzerland.

Akay, H., Koçyiğit, M. B. & Yanmaz, A. M. 2018 Effect of using multiple stream gauging stations on calibration of hydrologic parameters and estimation of hydrograph of ungauged neighboring basin. *Arabian Journal of Geosciences* **11**, 282–293.

Arnold, J. G., Srinivasan, R., Muttiah, R. S. & Williams, J. R. 1998 Large area hydrologic modeling and assessment part I: model development. *Journal of the American Water Resources Association* **34** (1), 73–89.

Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., Santhi, C., Harmel, R. D., van Griensven, A., Van Liew, M. W., Kannan, N. & Jha, M. K. 2012a SWAT: model use, calibration, and validation. *American Society of Agricultural and Biological Engineers (ASABE)* **55** (4), 1491–1508.

Arnold, J. G., Kiniry, J. R., Srinivasan, R., Williams, J. R., Haney, E. B. & Neitsch, S. L. 2012b *Soil & Water Assessment Tool – Input/Output Documentation, Version 2012*. Texas Water Institute, Texas, USA.

Bai, J., Shen, Z. & Yan, T. 2017 Comparison of single- and multi-site calibration and validation: a case study of SWAT in the Miyun Reservoir watershed, China. *Journal of Frontiers of Earth Science* **11** (3), 592–600.

Cao, W., Bowden, W. B., Tim, D. T. & Fenemor, A. 2005 Multi-variable and multi-site calibration and validation of SWAT in a large mountainous catchment with high spatial variability. *Journal of Hydrological Processes* **20**, 1057–1073.

Duana, Z., Tuoa, Y., Liu, J., Gaod, H., Song, X., Zhang, Z., Yangb, L. & Mekonnen, D. F. 2019 Hydrological evaluation of open-access precipitation and air temperature datasets using SWAT in a poorly gauged basin in Ethiopia. *Journal of Hydrology* **569**, 612–626.

Elshamy, M. E., Seierstad, I. A. & Sorteberg, A. 2009 Impacts of climate change on Blue Nile flows using bias-corrected GCM scenarios. *Hydrology and Earth System Sciences* **13**, 551–565.

Ercan, M. B., Goodall, J. L., Castronova, A. M., Humphrey, M. & Beekwilder, N. 2014 Calibration of SWAT models using the cloud. *Journal of Environmental Modelling & Software* **62**, 188–196.

Haas, M. B., Guse, B., Pfannerstill, M. & Fohrer, N. 2016 A joined multi-metric calibration of river discharge and nitrate loads with different performance measures. *Journal of Hydrology* **536**, 534–545.

IPCC 2001 *Technical Summary: Climate Change Impacts, Adaptation and Vulnerability: A Report of Working Group II*. Intergovernmental Panel on Climate Change (IPCC), Geneva, Switzerland.

Kahya, E., Kalaycı, S. & Piechota, T. C. 2008a Streamflow regionalization: case study of Turkey. *Journal of Hydrologic Engineering* **13** (4), 205–214.

Kahya, E., Demirel, M. C. & Bég, A. O. 2008b Hydrologic homogeneous regions using monthly streamflow in Turkey. *Earth Sciences Research Journal* **12** (2), 181–193.

Kahya, E., Özger, M., Şeker, D. Z., Karaca, M., Can, I., Kömüşcü, A. Ü., Bozkurt, D., Şen, O., Mehr, A. D., Erdem, H. & Bagheri, F. 2015 *Determination of Flood Risk on Basin Boundaries of the Rize Province: Current and Future Situation by Climate and Hydrological Models (Turkish)*. Scientific and Technological Research Council of Turkey (TÜBİTAK), Istanbul.

Klemes, V. 1986 Dilettantism in hydrology: transition or destiny? *Journal of Water Resources Research* **22** (9), 177S–188S.

Kocyigit, M. B., Akayl, H. & Yanmaz, A. M. 2017 Effect of watershed partitioning on hydrologic parameters and estimation of hydrograph of an ungauged basin: a case study in Gokirmak and Kocanaz, Turkey. *Arabian Journal of Geosciences* **10**, 331–344.

Lu, Z., Zou, S., Xiao, H., Zheng, C., Yin, Z. & Wang, W. 2015 Comprehensive hydrologic calibration of SWAT and water balance analysis in mountainous watersheds in northwest China. *Journal of Physics and Chemistry of the Earth* **79–82**, 76–85.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D. & Veith, T. L. 2007 Model evaluation guidelines for systematic quantification of accuracy in basin simulations. *American Society of Agricultural and Biological Engineers (ASABE)* **50** (3), 885–900.

Molina-Navarro, E., Andersen, H. E., Nielsen, A., Thodsen, H. & Trolle, D. 2017 The impact of the objective function in multi-site and multi-variable calibration of the SWAT model. *Journal of Environmental Modelling & Software* **93**, 255–267.

Morris, M. D. 1991 Factorial sampling plans for preliminary computational experiments. *Journal of Technometrics* **33** (2), 161–174.

Neitsch, S. A., Arnold, A. D., Kiniry, A. S. & Williams, D. A. 2011 *Soil and Water Assessment Tool – Theoretical Documentation Version 2009*. Texas Water Resources Institute, Technical Report No. 406. Texas.

Santhi, C., Kannan, N., Arnold, J. G. & Di Luzio, M. 2008 Spatial calibration and temporal validation of flow for regional scale hydrologic modeling. *Journal of the American Water Resources Association (JAWRA)* **44** (4), 829–846.

Sen, O. & Kahya, E. 2017 Determination of flood risk: a case study in the rainiest city of Turkey. *Journal of Environmental Modelling & Software* **93**, 296–309. doi:10.1016/j.envsoft.2017.03.030.

Shresthaa, M. K., Recknagela, F., Frizenschaf, J. & Meyer, M. 2016 Assessing SWAT models based on single and multi-site calibration for the simulation of flow and nutrient loads in the semi-arid Onkaparinga catchment in South Australia. *Journal of Agricultural Water Management* **175**, 61–71.

Solomon, S., Qin, D., Manning, M., Alley, R. B., Berntsen, T., Bindoff, N. L., Chen, Z., Chidthaisong, A., Gregory, J. M., Hegerl, G. C., Heimann, M., Hewitson, B., Hoskins, B. J., Joos, F., Jouzel, J., Kattsov, V., Lohmann, U., Matsuno, T., Molina, M., Nicholls, N., Overpeck, J., Raga, G., Ramaswamy, V., Ren, J., Rusticucci, M., Somerville, R., Stocker, T. F., Whetton, P., Wood, R. A. & Wratt, D. 2007 Technical summary. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor & H. L. Miller, eds). Cambridge University Press, Cambridge, UK and New York, NY, USA.

Van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M. & Srinivasan, R. 2006 A global sensitivity analysis tool for the parameters of multi-variable catchment models. *Journal of Hydrology* **324**, 10–23.

White, K. L. & Chaubey, I. 2005 Sensitivity analysis, calibration, and validations for a multisite and multivariable SWAT model. *Journal of the American Water Resources Association (JAWRA)* **41** (5), 1077–1089.

Winchell, M., Srinivasan, R., Di Luzio, M. & Arnold, J. 2013 *ArcSWAT Interface for SWAT2012 – User's Guide*. Blackland Research and Extension Centre & Grassland, Soil and Water Research Laboratory, USDA Agricultural Research Service, Texas, USA.

Zhang, L., Jin, X., He, C., Zhang, B., Zhang, X., Li, J., Zhao, C., Tian, J. & DeMarchi, C. 2016 Comparison of SWAT and DLBRM for hydrological modeling of a mountainous watershed in arid northwest China. *Journal of Hydrologic Engineering (ASCE)* **21** (5), 04016007 (1–11).

Zhang, Y., Shao, Q., Zhang, S., Zhai, X. & She, D. 2017 Comparison of single- and multi-site calibration and validation: a case study of SWAT in the Miyun Reservoir watershed, China. *Journal of Frontiers of Earth Science* **301**, 54–61.

First received 6 February 2020; accepted in revised form 19 October 2020. Available online 15 December 2020