

Article

Comparing the performances of multiple rainfall-runoff models of a karst watershed

Md Moudud Hasan¹, Md. Shariot-Ullah^{2*}, Ajoy Kumar Saha³ and M. G. Mostofa Amin²

¹Department of Agricultural and Industrial Engineering, Hajee Mohammad Danesh Science and Technology University, Dinajpur-5200, Bangladesh

²Department of Irrigation and Water Management, Bangladesh Agricultural University, Mymensingh-2202, Bangladesh

³Department of Irrigation and Water Management, Sylhet Agricultural University, Sylhet-2202, Bangladesh

*Corresponding author: Md. Shariot-Ullah, Department of Irrigation and Water Management, Bangladesh Agricultural University, Mymensingh-2202, Bangladesh. E-mail: msu_iwm@bau.edu.bd

Received: 31 March 2021/Accepted: 26 April 2021/ Published: 30 April 2021

Abstract: Different modeling concepts, a simple (black-box) to a fully distributed modeling (white-box), were used to develop a rainfall-runoff model based on the watershed characteristics to estimate runoff at the watershed outlet. A conceptual (grey-box) model is usually a balance between the black-box and white-box model. In this study, three grey-box models were developed by varying model structures for a karst watershed. The performance of the grey-box models was evaluated and compared with a semi-distributed type (white-box) model that was developed using the Soil and Water Assessment Tool in a previous study. The evaluation was carried out using goodness-of-fit statistics and extreme flow analysis using WETSPRO (Water Engineering Time Series Processing tool). Nash-Sutcliffe efficiencies (NSE) of the grey-box models were from 0.39 to 0.77 in the calibration period and from 0.30 to 0.61 in the validation period. However, the white-box model performed better in terms of NSE but has a higher bias. The best grey-box model performed better in simulating extreme flow, whereas the white-box (SWAT) model adequately simulated daily flows.

Keyword: extreme flow; WETSPRO; conceptual model; SWAT

1. Introduction

Rainfall-runoff models are often used to predict the behavior of a natural hydrological system (Abdollahi *et al.*, 2017; Amin *et al.*, 2017; Cislighi *et al.*, 2020). In reality, the natural system is very complex and consists of many components and processes. A model is a simplified representation of the actual system. The complexity of a model varies from a very simple to a more elaborate representation of the main processes of the system. Based on the degree of complexity, models are classified as (i) detailed physically-based models (white-box), (ii) conceptual models (grey-box), and (iii) empirical models (black-box) (Beck *et al.*, 1990; Willems, 2015). White-box models are considered the most complex, whereas black-box models are the simplest ones.

In white-box models, physical processes are represented using many equations to establish a relation between the input and output of the system. The prediction of a white-box model is relatively accurate and reliable, but it requires a powerful computing system and much time to simulate the results. It also requires a large amount of information and field data about the system to represent the processes accurately (Biftu & Gan, 2001; Chen *et al.*, 2016; Praskievicz & Chang, 2009; Wang *et al.*, 2010; Yang *et al.*, 2014). On the other hand, black-box models do not represent physical processes. In black-box models, a relation is established between the input and output of the model using observed input and output data. As black-box models do not include any physical process, they often provide inaccurate results especially when used for extrapolation or used for a dynamic system (Vaze *et al.*, 2011; Willems, 2015).

A grey-box model is a balance between the detailed physically-based model and the empirical model. In grey-box models, the main physical processes are represented in a lumped way by a few processes. Performances of a grey-box model can be improved by calibration once the main processes are included in the model (Jaiswal *et al.*, 2020; Kumar *et al.*, 2015; Willems, 2015). In this modeling, a systematic approach is applied where the whole system is analyzed as a unit rather than looking deeply into its components individually.

The complexity of a model depends on the purpose of its use, such as designing a water distribution project, forecasting the flood of a watershed, and prediction of frequency and amount of low flow (Eheart & Tornil, 1999; Smakhtin, 2001). Modeling for a flood management project mainly deals with the peak flow of a stream (Hsu & Wei, 2007; Khan & Valeo, 2016; Sudheer *et al.*, 2003). In contrast, a detailed simulation of the hydrological processes in a watershed is needed for predicting the transport of different pollutants (Amin *et al.*, 2017; Collick *et al.*, 2015), sediment (James *et al.*, 2010; Kizhisseri *et al.*, 2006), and for evaluating a farming system (Amin *et al.*, 2018). Therefore, a model's performance should be evaluated using multi-criteria analysis to understand its applicability. The Water Engineering Time Series Processing tool (WETSPRO) (Willems, 2004a, 2009) is used to evaluate a rainfall-runoff model. It can conduct flow filtering (peak and low flow) using a numeric digital filter. Nearly independent peaks and low flow values can be extracted using independent variables by this tool. Hence, it provides information about the model's performance to simulate extreme flows.

Developing a rainfall-runoff model for a karst watershed is a difficult process due to the complicated natural features of such watershed (Amin *et al.*, 2017). The presence of underground fractures in the bedrock and preferential flow paths enhance groundwater recharge in karst watersheds (Hartmann *et al.*, 2015). Thus, the infiltration capacity is high, which limits surface runoff and reduces actual evapotranspiration during wet conditions (Malard *et al.*, 2016). In addition, the network of sub-surface channels within karst aquifers increases the speed of groundwater flow (Fulton *et al.*, 2005) compared to water movement through non-karst aquifers. Moreover, the area outside the topo-graphic watershed boundary can discharge inside the basin and vice versa due to underground fractures. All these factors make it difficult to develop an efficient model of such a watershed. Therefore, a fully distributed or semi-distributed type model such as the Soil and Water Assessment Tool 'SWAT' (Arnold *et al.*, 1998) is often used for detailed rainfall-runoff simulation in a karst watershed (Collick *et al.*, 2015; Amin *et al.*, 2017). However, the performance of the grey-box model for a karst watershed should be evaluated to understand its suitability and limitation. In addition, a comparison of the performance of a grey-box with a white-box model can provide information about its suitability in a complicated watershed.

In this context, the performance of a grey-box model was evaluated based on multi-criteria analysis to understand its applicability better. In this study, a rainfall-runoff grey-box model was developed using the system approach concept for the karst watershed, and its performance was compared with the white box model (semi-distributed SWAT model) in terms of statistical-goodness of fit and extreme flow analysis.

2. Materials and Methods

2.1. Watershed overview

Spring Creek watershed in Center County, Pennsylvania, USA, was selected for this study (Figure 1). Geographically, this watershed covers 370 km² area situated between 40°40'–40°59'N and 77°38'–78°00'W in the Appalachian Ridge and Valley physiographic province of the upper Chesapeake Bay watershed. The fourth order stream Spring Creek discharges into Bald Eagle Creek. There are three gauging stations of the United States Geological Survey (USGS) in the watershed. The average elevation of the watershed is approximately 370 m above mean sea level (amsl). The study area has a temperate climate with hot, humid summers and cold winters. The mean annual temperature in the study area is 10.1°C, the average annual precipitation is 1060 mm, and annual actual evapotranspiration ranged 268–768 mm.

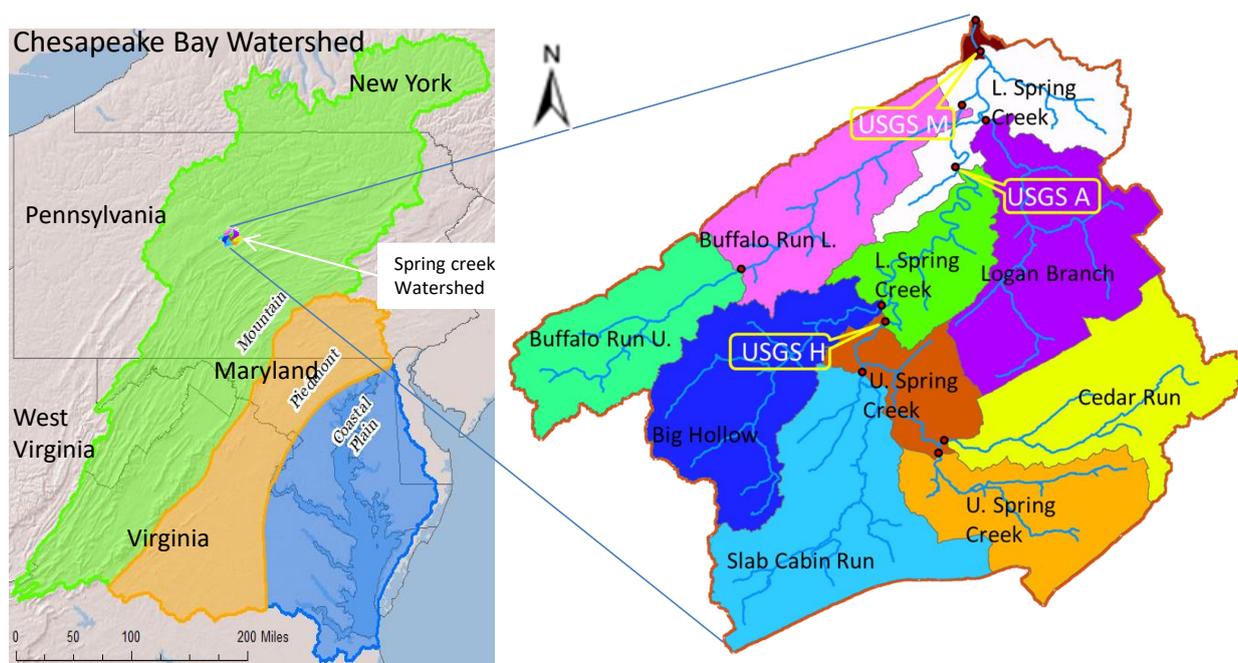


Figure 1. The location of Spring Creek watershed (Source: United States Geological Survey and <http://www.chesapeakebay.net>).

The Spring Creek watershed has Karst geologic formations. Taylor (1997) estimated based on the observed data from 1968 to 1994 that 33–60% of precipitation was lost through actual evapotranspiration, and 33–58% joined the stream, 9–16% of stream flow originated from direct runoff, whereas base-flow contribution was 87–91%. In the Spring Creek watershed, 34% of the land was used for agriculture, whereas 23% and 43% were developed and forests, respectively. The impervious area in the watershed increased from 3.1% in 1938 to 13.3 in 2006 (Brooks *et al.*, 2011). Watershed management practices have been applied in highly impervious areas to minimize runoff quantity (Amin *et al.*, 2017).

2.2. Data collection and evaluation

2.2.1. Data sources

The daily discharge, precipitation, maximum daily temperature, minimum daily temperature, and relative humidity of the watershed for 12 years (2002 to 2013) were used in this study to develop the grey-box model. The same data set was used to develop the white box model using SWAT in a previous study (Amin *et al.*, 2017). The SWAT-simulated daily discharge was used to compare with the grey-box model output. Observed daily discharge data were collected from the USGS gauging station for the Spring Creek watershed (<http://waterdata.usgs.gov/pa/nwis/rt>). Weather data was collected from the Chesapeake Community Modeling Program (<http://ches.communitymodeling.org/>) and the Pennsylvania State Climatologist (<http://climate.psu.edu/>).

2.2.2. Daily discharge data

The discharge data at the outlet of the watershed was used in the current study. The runoff and precipitation data were shown in Figure 2. It was observed that there were frequent peak flows and some long low flow periods as well.

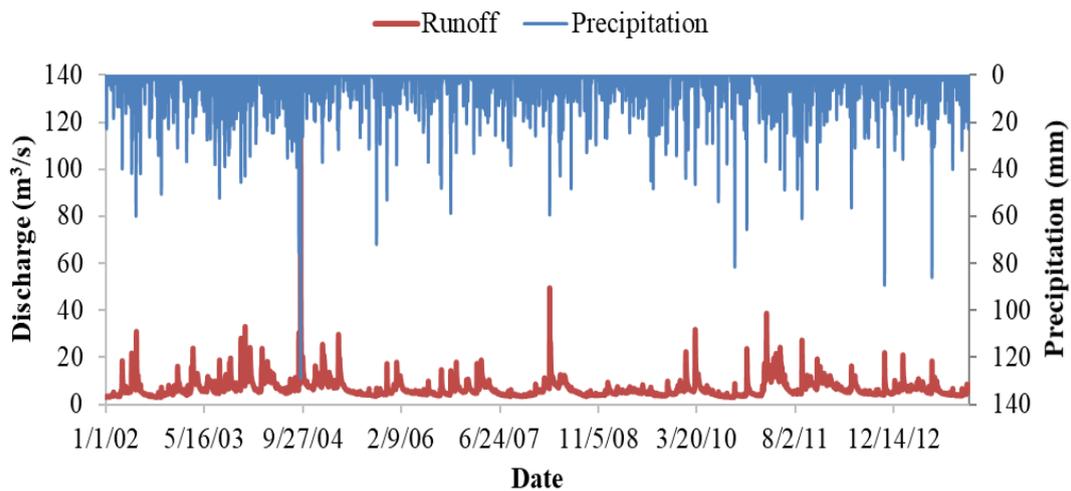


Figure 2. Plotting of time series of observed runoff (discharge) and rainfall (precipitation).

2.2.3. Daily precipitation data

The collected precipitation data was recorded in the station named "State College PA US" which is situated inside the watershed. It was assumed that the collected precipitation data is the average precipitation that was uniformly distributed over the watershed. The study area is in the temperate region, so the precipitation data contain rainfall and snowfall. Figure 2 illustrates that precipitation occurred around the year, but there was high precipitation variability between July and September.

2.2.4. Daily temperature data

The lowest temperature observed in January–February starts increasing from March until it reaches a peak in July. Then, the temperature starts decreasing for the rest of the months (Figure 3). The daily minimum temperature mostly remains below freezing between December and March, and snowfall occurs mainly during this below-freezing temperature period.

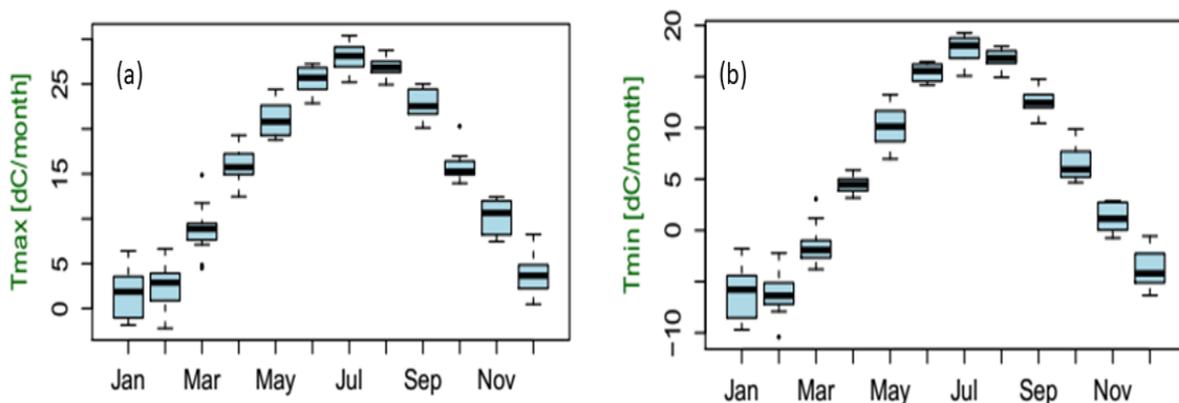


Figure 3. Monthly maximum (Tmax) (a) and minimum (Tmin) (b) temperature variation over the year in the study area (dC=degree Celsius).

2.2.5. Water balance of the watershed

The water balance of the watershed was checked using observed precipitation, observed runoff (stream flow), and the SWAT simulated evapotranspiration data. The cumulative volume of precipitation, runoff, and evapotranspiration was calculated for 12 years and plotted in Figure 4. The water balance of the watershed can be expressed using Eq. 1.

$$\text{Precipitation} = \text{Runoff} + \text{Evapotranspiration} + \text{Goundwater outflow} - \text{Groundwater inflow} \quad (1)$$

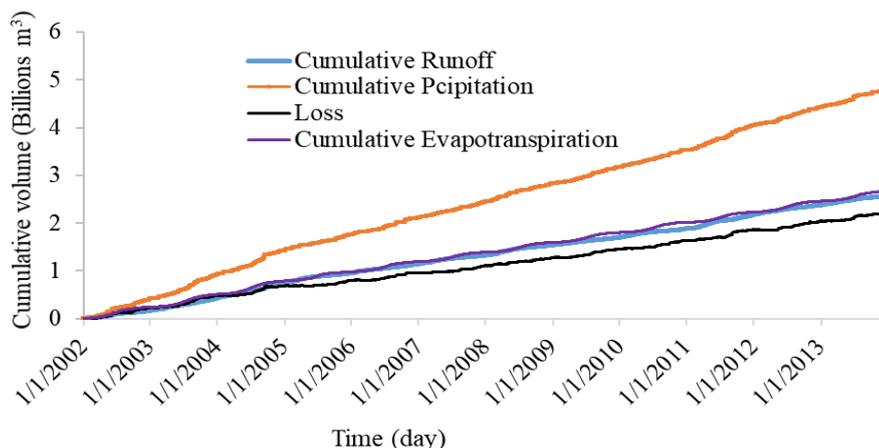


Figure 4. Variation of cumulative runoff, precipitation, evapotranspiration, and loss (equation (2)) in the water balance of Spring Creek watershed for the study period.

For long-term analysis, it is usually assumed that groundwater inflow is equal to groundwater outflow. As discussed in the study area section, the groundwater basin is 22% larger than the surface watershed, so there might be extra groundwater inflow in the water balance. To check groundwater contribution in water balance, the loss volume was calculated by subtracting runoff from precipitation.

$$\text{Loss} = \text{Precipitation} - \text{Runoff} \quad (2)$$

The calculated loss volume was then plotted together with cumulative precipitation, runoff, and evapotranspiration in Figure 4. Figure 4 shows that evapotranspiration was higher than the loss volume, which indicates that there was groundwater inflow contribution from outside the watershed. So, the water balance equation was presented as follows:

$$\text{Precipitation} = \text{Runoff} + \text{Loss} \quad (3)$$

$$\text{Loss} = \text{Evapotranspiration} + \text{Goundwater outflow} - \text{Groundwater inflow} \quad (4)$$

It was also observed that runoff was equal to 54% of the precipitation that is the average runoff coefficient (0.54).

2.3. Grey-box model construction

The main objective of this study was to build a simple but acceptable model, so a grey-box model was constructed by including different processes in a lumped way. At first, the water balance equation was developed, and the relationships between different hydrological processes were analyzed. Then, a simple conceptual model was constructed, and the model's structure was changed until the model provides an acceptable result. The different steps followed to build the grey-box models are described in the following subsections.

2.3.1. Model-1

At first, a simple model structure was used. The precipitation-runoff model was considered as a rainfall-runoff model; snow melting or snowpack was not used. The daily net rainfall was calculated by multiplying the loss factor with the daily precipitation.

$$\text{Daily net rainfall } (P_{\text{net}}) = \text{Daily precipitation} * \text{Loss factor} \quad (5)$$

The loss factor (L_f) was calibrated by comparing the cumulative volume of observed discharge and net rainfall. The loss factor value was selected for which the cumulative volume of the observed discharge almost equal to the cumulative volume of net rainfall. Once the daily net rainfall (P_{net}) was obtained, it was divided into three runoff components: overland flow, interflow, and baseflow. Initially, the values of overland flow portion (W_{OF}), interflow portion (W_{IF}), and baseflow portion (W_{BF}) were derived by analyzing the observed discharge in

WETSPRO (Willems, 2004a). Hence, the input for overland flow, interflow, and baseflow was calculated using the following equation:

$$q_{inOF} = P_{net} \times W_{OF} \quad (6)$$

$$q_{inIF} = P_{net} \times W_{IF} \quad (7)$$

$$q_{inBF} = P_{net} \times W_{BF} \quad (8)$$

Here, q_{inOF} , q_{inIF} and q_{inBF} are the input for overland flow, interflow, and baseflow, respectively.

From the analysis of observed data in WETSPRO, it was found that the portion of base flow (W_{BF}) is around 0.65 of total flow and the portion of quick flow (interflow + overland flow) is 0.35 of total flow. It was also observed that the portion of interflow (0.25) is higher than the overland flow (0.10). In case of heavy rainfall, the portion of overland flow should increase; hence, the flow portion between interflow and overland flow was switched based on critical net precipitation $P_{c,net}$. The value of $P_{c,net}$ was adjusted during calibration.

The input-output relation of the water system can be described using a linear reservoir model (Willems, 2004b). The linear reservoir model was used to represent the input-output relation of three runoff components. It was assumed that three reservoirs are in the parallel position; that is, the runoff component contributes to runoff separately. Equation (9) was used to calculate the outflow of each runoff component, i.e., the overland outflow (q_{outOF}), interflow outflow (q_{outIF}), and baseflow outflow (q_{outBF}) from the input of overland flow, interflow, and baseflow, respectively.

$$q_{out}(t) = \exp\left(-\frac{1}{k}\right) q_{out}(t-1) + \left(1 - \exp\left(-\frac{1}{k}\right)\right) q_{in}(t) \quad (9)$$

Here, $q_{out}(t)$ is the outflow at time t (m^3/s); $q_{out}(t-1)$ is the outflow at time $t-1$ (m^3/s); $q_{in}(t)$ is the inflow at time t (m^3/s); k is the flow recession constant. The initial value of the recession constant of overland flow (k_{OF}), interflow (k_{IF}), and baseflow (k_{BF}) was determined by analyzing the observed flow in WETSPRO.

Finally, the model outflow was obtained by adding the outflow of overland flow, interflow, and baseflow.

$$Q_{out}(t) = q_{outOF} + q_{outIF} + q_{outBF} \quad (10)$$

where $Q_{out}(t)$ is the model outflow which is runoff at time t .

2.3.2. Model-2

In Model-1, the loss was the same throughout the year. However, the idea of calculating net rainfall as a function of daily average temperature can be used in this study. So, a new loss equation (Eq. 11) was introduced in the second approach.

$$L = \begin{cases} sT_{avg} & \text{if } T_{avg} > 0 \\ 0 & \text{if } T_{avg} \leq 0 \end{cases} \quad (11)$$

Here, L is the loss (mm); s is the linear regression slope (constant).

Therefore, the net rainfall equation was also changed to Eq. 12.

$$P_{net_Tavg} = \text{Daily precipitation} - L \quad (12)$$

The rest of the model structure was the same as the first approach for Model-1. However, all the parameters were calibrated and optimized for this new model.

2.3.3. Model-3

The inclusion of the snow melting process could improve the model performance. Therefore, the snowpack and snow melting process were included in the third approach to improve the model performance. At first, it was decided whether precipitation is rainfall or snowfall based on a critical temperature ($T_{c,snow}$). The critical temperature is the average temperature of the day for which the precipitation will be considered as snowfall. This parameter was adjusted during calibration. However, in reality, the form of precipitation depends on many meteorological factors. Snow melting rate was calculated using an equation adapted from Moussav *et al.* (1989). The snow melting equation is as follows:

$$M = \begin{cases} C(T_{max}^* - T_{base}) & \text{if } T_{max}^* > T_{base} \\ 0 & \text{if } T_{max}^* < T_{base} \end{cases} \quad (13)$$

Here, M is the melting rate (mm day^{-1}); T_{max}^* is the daily maximum temperature over freezing point ($^{\circ}\text{C}$); T_{base} is the base temperature ($^{\circ}\text{C}$); C is the melting factor ($\text{mm day}^{-1} \text{ } ^{\circ}\text{C}^{-1}$). The optimum value of melting factor, C and base temperature, T_{base} was selected during calibration.

The snowpack process was added in this model to accumulate snowfall in winter. The daily snowpack amount was calculated using the following equation.

$$S_{\text{pack}_t} = S_{\text{pack}_{t-1}} - M_{a_{t-1}} + P_{\text{snow}_t} \tag{14}$$

Here, S_{pack_t} and $S_{\text{pack}_{t-1}}$ is the snowpack amount in mm in day t and $t-1$, respectively, P_{snow_t} is the snow fall in mm in day t , and $M_{a_{t-1}}$ is the actual melting of snow in mm in day $t-1$. The actual melting was calculated using the following equation:

$$M_{a_t} = \begin{cases} M_t & \text{if } S_{\text{pack}_t} > M_t \\ S_{\text{pack}_t} & \text{if } S_{\text{pack}_t} < M_t \\ 0 & \text{if } S_{\text{pack}_t} \leq 0 \end{cases} \tag{15}$$

where M_{a_t} is the actual melting (mm) in day t , M_t is the snow melting potential of the day t and S_{pack_t} is the amount of accumulated snow up to t^{th} day.

The rest of the model structure was the same as in the previous approaches. However, all the parameters were calibrated again for this new approach. A diagram of the model is shown in Figure 5.

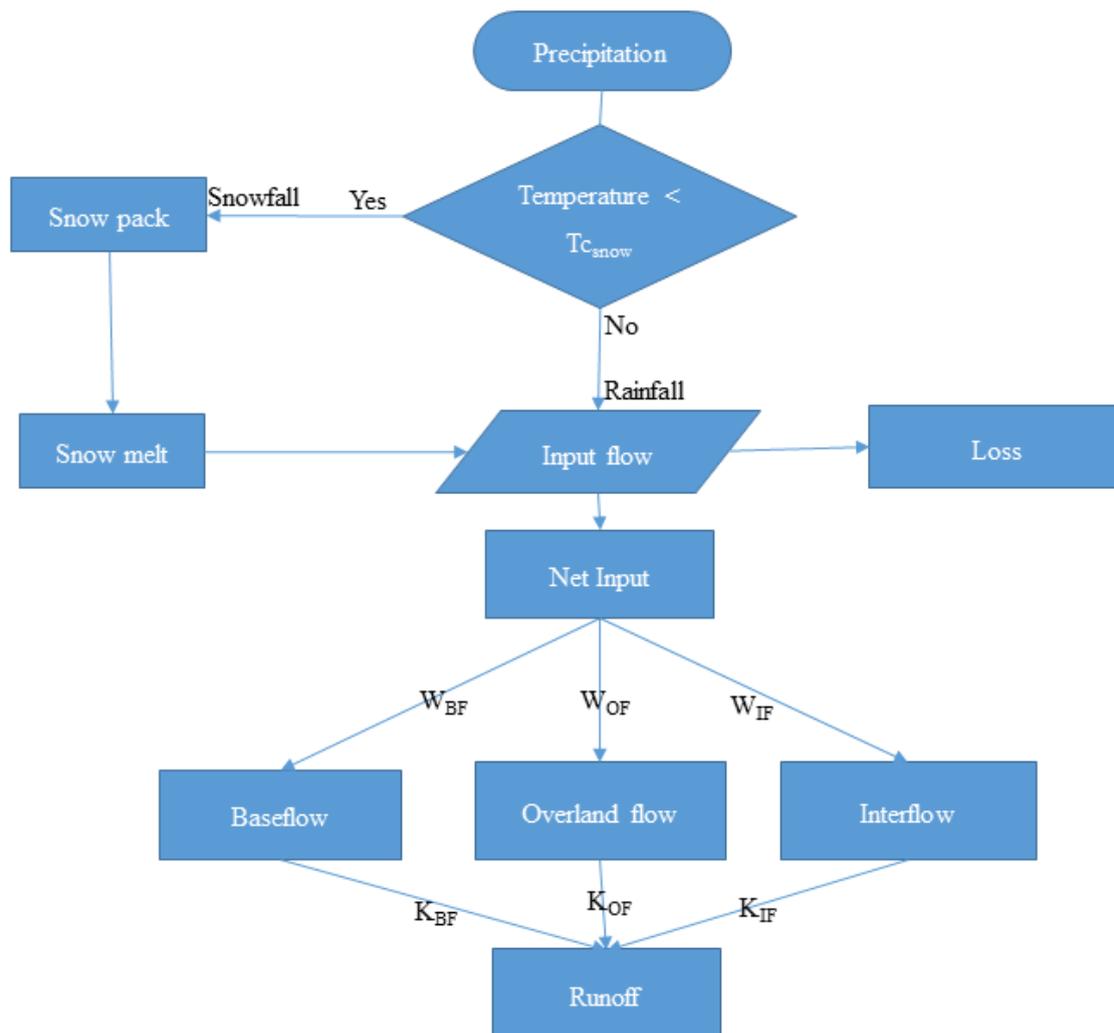


Figure 5. Grey-box model diagram of Model-3.

2.3.4. Grey-box model calibration

The model was calibrated for each parameter from six years of daily data (2002 to 2007), and the loss factor was calibrated based on water balance. The water balance was checked by comparing the cumulative net rainfall volume with the cumulative observed outflow. The other parameters were calibrated by trial and error while the performance of the model was observed simultaneously. The excel solver was also used to optimize the parameters where the objective function was to maximize the Nash–Sutcliffe efficiency (NSE) (Eq. 16). After automatic optimization by the solver, fine-tuning was done manually to minimize the overfitting.

2.4. White-box model

A semi-distributed model of this watershed was developed using Topo-SWAT in a previous study (Amin *et al.*, 2017). The simulated runoff of this model was taken as an output of a white-box model. This model was devolved using different types of weather, soil attributes, and terrain data. Weather and streamflow data of several stations were used to calibrate and validate the model. Ten-meter resolution of elevation data was used to define the terrain of the watershed. In addition, a spatial combination of the FAO-UNESCO Digital Soil Map of the World (FAO, 2007) and topographically derived wetness classes were used in this model. Hence, this semi-distributed model was developed using more data (particularly spatially distributed data) than the grey box model.

2.5. Model performance evaluation

The performance of the grey-box models was evaluated by both visual inspection and statistical goodness-of-fit. Then, the best grey-box model was compared with the white-box model (SWAT model). The models were calibrated using the observed data for the period of 2002 to 2007 and validated for the period of 2008 to 2013. The statistical goodness-of-fit methods were carried out by using the Nash–Sutcliffe efficiency (NSE) (Eq. 16) and percent bias (PBIAS) (Eq. 17) (Moriassi *et al.*, 2007). In addition, the coefficient of determination (R^2) was used to assess the precision of the simulated values with respect to the observed.

$$NSE = 1 - \frac{\sum_{i=1}^n (Q_{o_i} - Q_{s_i})^2}{\sum_{i=1}^n (Q_{o_i} - \bar{Q}_o)^2} \quad (16)$$

$$PBIAS = \frac{\sum_{i=1}^n (Q_{o_i} - Q_{s_i})}{\sum_{i=1}^n (Q_{o_i})} \times 100 \quad (17)$$

where n is the number of observations, and Q_{o_i} and Q_{s_i} are the i th observed and simulated runoff, respectively.

2.6. Extreme flow analysis

The Water Engineering Time Series Processing (WETSPRO) (Willems, 2004a) tool was used for extreme flow analysis. The observed flow, grey-box model output, and white-box model output were plugged into the WETSPRO tool to evaluate the model performance in terms of extreme flow simulation. The flow was filtered using a numeric digital filter of WETSPRO to extract independent peak-over-threshold (POT) values. Three parameters were needed for POT analysis. Those parameters are (i) maximum fraction of base flow in the peak flow (f), (ii) independence period (k_p) between two peaks, and (iii) minimum peak height (q_{lim}). The k_p can be taken equal to the recession constant of quick flow or higher. The f can be taken as the upper limit of the base flow fraction in the peak flow. The upper limit of small noise peaks, which needs to be avoided to be selected as POTs, can be used as q_{lim} . The parameters used in the POT section are shown in Table 1. Those parameter values were estimated by analyzing observed flow. At first, the observed time series were filtered into three components of baseflow, interflow, and overland flow. Then, independent peak flow and low flow were calculated through POT selection. The same procedure was also applied to the output of the models.

Table 1. Parameters used in POT selection using the method independent of base flow.

Parameters	Peak flow period	Low flow period
Maximum fraction of base flow in the peak flow (f) [-]:	0.4	0.3
Independence period (k_p) [day]:	30	130
Minimum peak height (q_{lim}) [m^3/s]:	1	1

High flow values receive higher weights in statistical goodness of fit calculation. The Box-Cox (BC) transformation of values can minimize the effect of such consequences. Therefore, POT values were transformed using Eq. 18.

$$BC(Q) = \frac{Q^\lambda - 1}{\lambda} \quad (18)$$

The value of parameter λ ranges from 0 to 1, and it should be adjusted to reach homoscedasticity in the residuals. In this study, homoscedasticity was obtained for $\lambda = 0.25$ for both peak and low flows.

2.7. Statistical test

The statistical hypothesis test of the difference between simulated and observed extreme flows was conducted using the Wilcoxon signed-rank test (Rey & Neuhäuser, 2011), a non-parametric type test. The non-parametric test was performed because the extreme flows were not normally distributed. Pair-wise Wilcoxon signed-rank test was performed using the function "wilcox.test" of R statistical analysis software (R Core Team, 2021).

3. Results and Discussion

3.1. Calibrated parameter of grey-box model

The calibrated parameters of different models are shown in Table 2. Model-1 and Model-2 had the same number of parameters, while Model-3 had the maximum number of parameters. The value of the parameters related to flow portion (e.g., overland flow, base flow, inter-flow) were the same in both Model-2 and Model-3. The parameters related to net-input calculation were changed due to the changes in the model structures.

Table 2. List of the parameters used in the different models with calibrated values.

Parameter name	Unit	Model-1	Model-2	Model-3
Loss factor, L_f	-	0.55		
Slope parameter of loss, S	-		0.6	0.56
The critical temperature for snow fall, $T_{c\text{ snow}}$	°C			1
Base temperature, T_{base}	°C			0
Melting factor, C	mm day ⁻¹ °C ⁻¹			0.6
Critical net precipitation, $P_{c\text{ net}}$	m ³ /s	360	360	360
Overland flow portion, W_{BF}	-	0.08	0.08	0.08
Overland flow recession constant, k_{OF}	days	1	1	1
Interflow Portion, W_{BF}	-	0.22	0.35	0.35
Interflow recession constant, k_{IF}	days	20	30	30
Base flow portion, W_{BF}	-	0.7	0.57	0.57
Base flow recession constant, k_{BF}	days	170	170	170
Total number of parameters		8	8	11

3.2. Model performance evaluation

3.2.1. Simulated and observed outflow

The observed outflow and simulated outflow of different models are shown in Figure 6 for both calibration and validation periods. It was observed that Model-1 was unable to simulate peak flow and low flow. A fixed loss factor was used in Model-1; hence, the simulated outflow of model-1 did not have seasonal variation. The simulated outflow values of Model-2 and Model-3 were almost similar except in winter (2/9/2008) when Model-2 overestimated the flow. This is because Model-2 did not have the snowfall accounting process. Model-3 successfully simulated small peaks and low flows. However, it underestimated some peak flows. On the other hand, the white-box (SWAT) model was able to simulate high and medium peaks. However, it overestimated some of the peak values. Seasonal variation was strongly visible in the white box model than the grey box model. This means that the white-box model was suitable for simulating the daily runoff process.

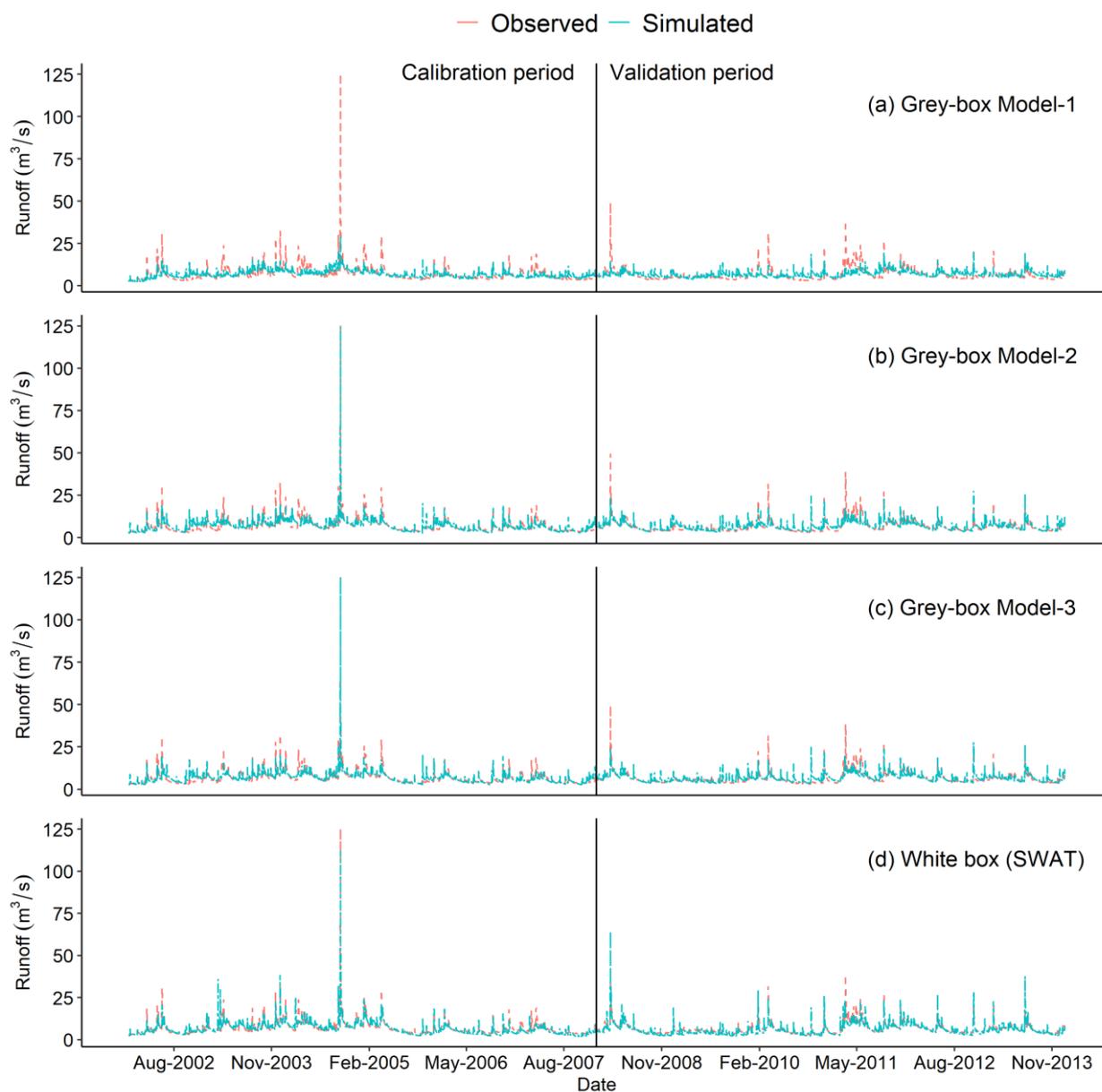


Figure 6. Observed and simulated runoff of different models in calibration and validation periods.

3.2.2. Statistical goodness-of-fit

The performance indicators of the models based on statistical goodness of fit are shown in Table 3. Model-3 had the highest NSE and R^2 in both calibration and validation periods compare to other grey-box models (Table 3). Model-3 also had the lowest bias in the validation period. Model-2 had similar goodness-of-fit statistics to Model-3. However, Model-1 performed poorly with NSE below 0.5. Hence, Model-3 was chosen as a grey-box model for further analysis in the study.

The white-box model performed better than the grey-box models regarding NSE and R^2 in both calibration and validation periods. However, PBIAS was higher in the white-box model than the grey-box models. It was also observed that the PBIAS of the white-box model was higher in both the validation and calibration periods than the grey-box models. The positive values of PBIAS means that it overestimated the runoff. Although the white-box model had higher NSE than the grey box model, the NSE values of both models were within the acceptable range (>0.50) (Moriassi *et al.*, 2007).

Table 3. Model performance evaluation based on statistical goodness-of-fit methods.

Model	Calibration (2002-2007)			Validation (2008-2013)		
	NSE	R ²	PBIAS	NSE	R ²	PBIAS
Model-1 (grey-box)	0.39	0.41	1.71	0.30	0.30	-3.66
Model-2 (grey-box)	0.70	0.70	1.36	0.55	0.56	-2.12
Model-3 (grey-box)*	0.77	0.77	2.66	0.61	0.61	-1.15
SWAT (white-box)	0.78	0.80	5.62	0.69	0.76	8.38

*Selected grey-box model

3.2.3. Peak flow analysis

The model performance concerning separated peak flow is shown in Figure 7. It was observed that both models were able to simulate high flow up to 30 m³/s. However, for the flow higher than 30 m³/s or a return period of more than two years, the grey-box model (Model-3) slightly underestimated the flow. However, a very high flow was well simulated by the grey-box model. Overall, there was no significant difference between the grey-box model predictions and the observed peak values (p-value = 0.11). On the other hand, the white-box model's peak values were significantly different from the observed peaks (p-value= 9.12×10⁻⁰⁸) and the peaks of the grey-box model (p-value = 0.0027). In the scatter plot of peak flows after the Box-Cox transformation (Figure 7b), the grey-box model underestimated the peak flows as the mean deviation was lower than the bisector. The plotting points are more scattered for the white-box model than the grey-box model. This means that the prediction uncertainty for the white-box model was higher.

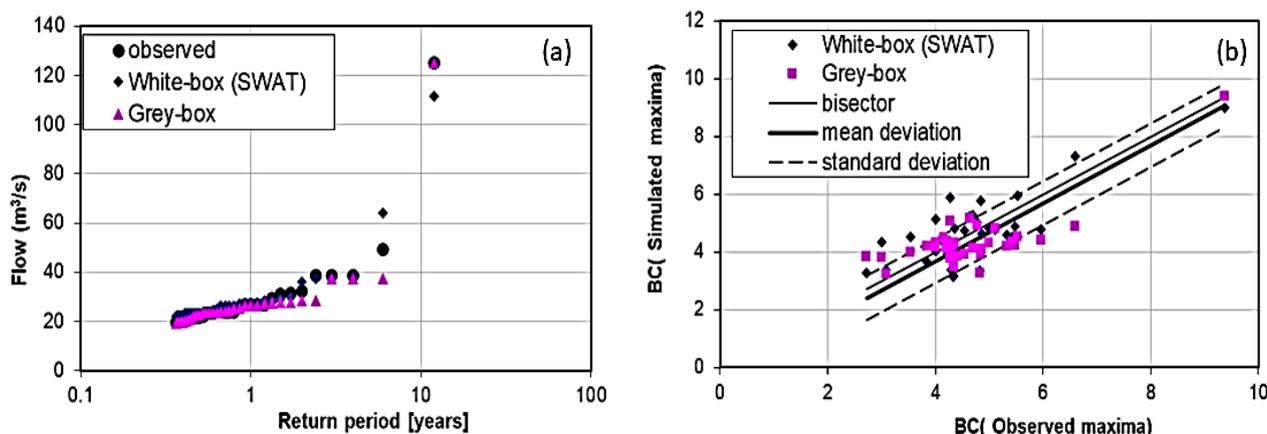


Figure 7. Graphical comparison of observed and simulated peak flows (a) and scatter plot of peak flows after the Box-Cox transformation (b).

3.2.4. Low flow analysis

Both models underestimated the low flows (Figure 8). The observed values of low flow were significantly different from the predicted values for both the grey-box (p-value = 0.00031) and white-box models (p-value = 3.05×10⁻⁰⁵). The mean deviation was lower than the bisector (Figure 8b), which indicates that the model underestimated the low flow. The degree of under-estimation was higher for the white-box model as most of the scatter points were below the bisector line and out of the standard deviation line. The performance of the grey-box model was better in low flow simulation.

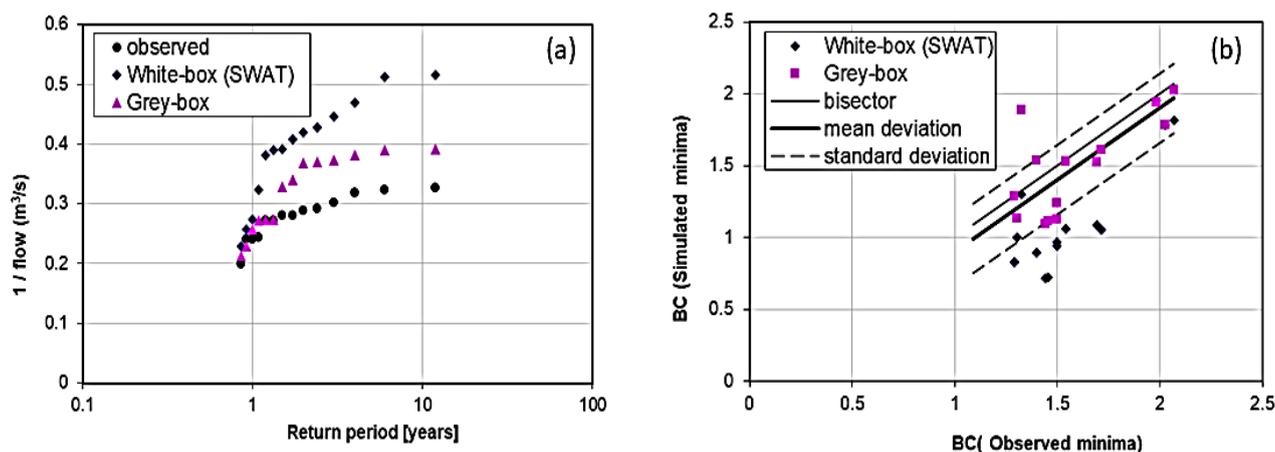


Figure 8. Graphical comparison of observed and simulated low flows (a) and scatter plot of low flows after the Box-Cox transformation (b).

4. Conclusions

Multiple rainfall-runoff conceptual (grey-box) models were developed using different model structures for a karst watershed. The performance of the selected grey-box model was compared with a semi-distributed SWAT (white-box) model. The performance of the models was evaluated in terms of goodness-of-fit statistics and extreme flow analysis. According to overall goodness-of-fit, the white-box model performed better with the highest NSE and R^2 ; however, the best grey-box model was also acceptable as NSE was higher than 0.60 in both calibration and validation periods. The white-box model can simulate seasonal variation better, whereas the grey-box model simulated extreme flow (peak and low flow) more efficiently than its counterpart did. Therefore, we can conclude that a conceptual (grey-box) model can be used for extreme flow analysis, whereas a semi-distributed model is suitable for daily runoff simulation.

Conflict of interest

None to declare.

Authors' contribution

Md Moudud Hasan: Manuscript editing, model development, data analysis.

Md. Shariot-Ullah: Manuscript editing, resource.

Ajoy Kumar Saha: Data investigation, review & editing.

M. G. Mostofa Amin: Data collection, model development, review and editing, supervision.

References

- Abdollahi K, I Bashir, B Verbeiren, MR Harouna, A Van Griensven, M Huysmans and O Batelaan, 2017. A distributed monthly water balance model: formulation and application on Black Volta Basin. *Environmental Earth Sciences*, 76: 198.
- Amin MGM, TL Veith, AS Collick, HD Karsten and AR Buda, 2017. Simulating hydrological and nonpoint source pollution processes in a karst watershed: A variable source area hydrology model evaluation. *Agricultural Water Management*, 180: 212–223.
- Amin MGM, HD Karsten, TL Veith, DB Beegle and PJA Kleinman, 2018. Conservation dairy farming impact on water quality in a karst watershed in northeastern US. *Agricultural Systems*, 165: 187–196.
- Arnold JG, R Srinivasan, RS Muttiah and JR Williams, 1998. Large area hydrologic modeling and assessment. Part I: model development. *J. Am. Water Resour. Assoc.*, 34: 73–89.
- Beck MB, FM Kleissen and HS Wheater, 1990. Identifying flow paths in models of surface water acidification. In *Reviews of Geophysics (Vol. 28, Issue 2, pp. 207–229)*.
- Biftu GF and TY Gan, 2001. Semi-distributed, physically based, hydrologic modeling of the paddle river basin, Alberta, using remotely sensed data. *Journal of Hydrology*, 244: 137–156.
- Brooks RP, RF Carline, JS Shortie, SE Yetter, D Weller and K Boomer, 2011. Analysis of BMP implementation performance and maintenance in spring creek, an agriculturally-influenced watershed in Pennsylvania. *USDA*, 1–19.

- Chen Y, J Li, and H Xu, 2016. Improving flood forecasting capability of physically based distributed hydrological models by parameter optimization. *Hydrology and Earth System Sciences*, 20: 375–392.
- Cislaghi A, D Masseroni, C Massari, S Camici, and L Brocca, 2020. Combining a rainfall–runoff model and a regionalization approach for flood and water resource assessment in the western Po Valley, Italy. *Hydrological Sciences Journal*, 65: 348–370.
- Collick AS, DR Fuka, PJA Kleinman, AR Buda, JL Weld, MJ White, TL Veith, RB Bryant, CH Bolster and ZM Easton, 2015. Predicting phosphorus dynamics in complex terrains using a variable source area hydrology model. *Hydrological Processes*, 29: 588–601.
- Eheart JW and DW Tornil, 1999. Low-flow frequency exacerbation by irrigation withdrawals in the agricultural midwest under various climate change scenarios. *Water Resources Research*, 35: 2237–2246.
- Fulton BJW, EH Koerkle, SD Mcauley, SA Hoffman and LF Zarr, 2005. Hydrogeologic setting and conceptual hydrologic model of the Spring Creek basin, Centre County, Pennsylvania. U.S. Geological Survey Scientific Investigations Report 2005-5091, June, 83. <http://www.csa.com/partners/viewrecord.php?requester=gs&collection=ENV&recid=15961414>
- Hartmann A, T Gleeson, R Rosolem, F Pianosi, Y Wada and T Wagener, 2015. A large-scale simulation model to assess karstic groundwater recharge over Europe and the Mediterranean. *Geoscientific Model Development*, 8: 1729–1746.
- Hsu NS and CC Wei, 2007. A multipurpose reservoir real-time operation model for flood control during typhoon invasion. *Journal of Hydrology*, 336: 282–293.
- Jaiswal RK, S Ali and B Bharti, 2020. Comparative evaluation of conceptual and physical rainfall–runoff models. *Applied Water Science*, 10(1).
- James SC, CA Jones, MD Grace and JD Roberts, 2010. Advances in sediment transport modelling. *Journal of Hydraulic Research*, 48: 754–763.
- Khan UT and C Valeo, 2016. Short-term peak flow rate prediction and flood risk assessment using fuzzy linear regression. *Journal of Environmental Informatics*, 28: 71–89.
- Kizhisseri AS, D Simmonds, Y Rafiq and M Borthwick, 2006. An evolutionary computation approach to sediment transport modelling. *Coastal Dynamics 2005 - Proceedings of the Fifth Coastal Dynamics International Conference*. [https://doi.org/10.1061/40855\(214\)81](https://doi.org/10.1061/40855(214)81)
- Kumar A, R Singh, PP Jena, C Chatterjee and A Mishra, 2015. Identification of the best multi-model combination for simulating river discharge. *Journal of Hydrology*, 525: 313–325.
- Malard A, M Sinreich and PY Jeannin, 2016. A novel approach for estimating karst groundwater recharge in mountainous regions and its application in Switzerland. *Hydrological Processes*, 30: 2153–2166.
- Moriasi DN, JG Arnold, MW Van Liew, RL Binger, RD Harmel and TL Veith, 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50: 885–900.
- Moussav M, G Wyseure and J Feyen, 1989. Estimation of melt rate in seasonally snow-covered mountainous areas. *Hydrological Sciences Journal*, 34(December 2014), 249–263.
- Praskievicz S and H Chang, 2009. A review of hydrological modelling of basin-scale climate change and urban development impacts. *Progress in Physical Geography*, 33: 650–671.
- R Core Team, 2021. R: A Language and Environment for Statistical Computing. <https://www.r-project.org/>
- Rey D and M Neuhäuser, 2011. Wilcoxon-Signed-Rank Test. In *International Encyclopedia of Statistical Science* (pp. 1658–1659). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-04898-2_616
- Smakhtin VU, 2001. Low flow hydrology: a review. *Journal of Hydrology*, 240: 147–186.
- Sudheer KP, PC Nayak and KS Ramasastri, 2003. Improving peak flow estimates in artificial neural network river flow models. *Hydrological Processes*, 17: 677–686.
- Taylor LE, 1997. *Water Budget for the Spring Creek Basin*. 184.
- Vaze J, P Jordan, R Beecham, A Frost and G Summerell, 2011. Guidelines for rainfall-runoff modelling Towards best practice model application. In *eWater Cooperative Research Center* (Issue December). Water Research Cooperation.
- Wang XL, VR Swail and A Cox, 2010. Dynamical versus statistical downscaling methods for ocean wave heights. *International Journal of Climatology*, 30: 317–332.
- Willems P, 2004a. WETSPRO : Water Engineering Time Series PROcessing tool. Manual, 22.
- Willems P, 2009. A time series tool to support the multi-criteria performance evaluation of rainfall-runoff models. *Environmental Modelling and Software*, 24: 311–321.
- Willems P, 2015. *Systems Approach to Water Management*. 61–63.
- Willems P, 2004b. Parsimonious model for combined sewer overflow pollution. 10 p.

Yang H, G Wang, Y Yang, B Xue and B Wu, 2014. Assessment of the impacts of land use changes on nonpoint source pollution inputs upstream of the three gorges reservoir. *The Scientific World Journal*, 2014. <https://doi.org/10.1155/2014/526240>