Research Article

Influence of Hydrological Model Selection on Simulation of Moderate and Extreme Flow Events: A Case Study of the Blue Nile Basin

Charles Onyutha

Faculty of Technoscience, Muni University, P.O. Box 725, Arua, Uganda

Correspondence should be addressed to Charles Onyutha; conyutha@gmail.com

Five hydrological models were applied based on data from the Blue Nile Basin. Optimal parameters of each model were obtained by automatic calibration. Model performance was tested under both moderate and extreme flow conditions. Extreme events for the model performance evaluation were extracted based on seven criteria. Apart from graphical techniques, there were nine statistical "goodness-of-fit" metrics used to judge the model performance. It was found that whereas the influence of model selection may be minimal in the simulation of normal flow events, it can lead to large under- and/or overestimations of extreme events. Besides, the selection of the best model for extreme events may be influenced by the choice of the statistical "goodness-of-fit" measures as well as the criteria for extraction of high and low flows. It was noted that the use of overall water-balance-based objective function not only is suitable for moderate flow conditions but also influences the models to perform better for high flows than low flows. Thus, the choice of a particular model is recommended to be made on a case by case basis with respect to the objectives of the modeling as well as the results from evaluation of the intermodel differences.

1. Introduction

According to the Intergovernmental Panel on Climate Change, IPCC [1], the continued increase in the greenhouse gas emissions will lead to the warming and changes in the various components of the climate system. With respect to the hydrological system, these changes are reflected in the increased frequency and severity of risk-based water disasters such as drought and floods which inflict severe damage on public life and property in many parts of the world on a yearly basis. The Nile Basin in which the study area is located is not exceptional to such hydrological disasters. Although the Ethiopian highlands experience rainfall in excess of 1000 mm annually, extreme low rainfall conditions also tend to occur in the Blue Nile Basin (BNB). Thus, both floods and drought occur in the study area. Some typical examples of the occurrences of hydrometeorological extremes in the BNB include the daily rainfall of 88 mm (which was the highest since late 1890s) in Khartoum on the 31st July 1920 [2], the severe floods in the months of August and September in 1988 in Sudan [3],

the devastating drought of the early 1980s in the Ethiopian highlands [3], and drought conditions over the 1970s and 1980s which led to the decline in annual rainfall by 30% [4].

In support of the investigations of impacts of climate variability and change on hydrological extremes, hydrological models may be applied. The ability of such models to capture the extreme high and low flow events is important for planning and management of risk-based water resources applications. Because the study area is dominated by cultivated area based on rain-fed agriculture, normal hydrological events (e.g., annual mean flows) which may be indicative of the mean net rainfall totals are vital for the management of agricultural practices. Distributed models (from which simulated runoff can be obtained at any point of interest within the catchment) require more detailed spatial inputs to reflect the spatiotemporal variability in the runoff. Conceptual models (which are simpler in their structures than the fully distributed process-based models) are designed mostly to simulate lumped runoff at the catchment outlet (as aimed at in this study) based on catchment-wide averaged inputs.

Some examples of rainfall-runoff models include Nedbør-Afstrømnings Model (NAM) (Danish Hydraulic Institute DHI [5–7]), Hydrologiska Byråns Vattenavdelning (HBV) [8], Probability Distribution Model (PDM) [9], Precipitation Runoff Modeling System (PRMS) [10], SIMHYD, that is, the simplified version of HYDROLOG [11], TANK model [12], Sacramento (SAC) model [13], Australian Water Balance Model (AWBM) [14], and Identification of Unit Hydrographs and Component Flows from Rainfall, Evaporation and Stream (IHACRES) flow data model [15–17]. For the study area, several rainfall-runoff models have been applied to model the hydrological regimes of the Blue Nile. Examples of the models applied include Soil and Water Assessment Tool (SWAT) by [18–21], HBV by [22, 23], Hydrologic Engineering Centre-Hydrological Modeling System (HEC-HMS) by [24], and PRMS by [25, 26]. Most of the above studies applied only one hydrological or rainfall-runoff model. Because of the tendency of the models to differ among themselves in terms of their structural complexity and set of parameters for calibration, the use of only one model in hydrological investigations leads to lack of insight about the influence of the model selection on the modeled results [27]. According to [27], the influence of the selection of a hydrologic modeling approach which is seldom investigated is critical for impact assessment, for example, of climate variability and change on the water resources of the BNB. The use of modeled results for data scarce region like the BNB to aid water resources planning and management decisions requires clear assessment of the possible uncertainties and their communication. Water management decisions taken amidst uncertainty of the scientific supporting information may lead to unnecessarily lavish expenditure of the limited economic resources, for example, for risk-based applications. To judge the confidence in the results of modeling amidst data limitation and quality problem, evaluation of a number of models in particular study is of paramount importance to the scientific community especially those involved in impact investigations, for example, of climate variability and change on hydrology.

This study is therefore aimed at exploring the influence of model selection on the simulation of (1) daily moderate and extreme flow events, (2) temporal changes in the flows from (1), and (3) hydrological extreme quantiles as a simultaneous function of aggregation levels and return periods.

2. Study Area, Data, and Selected Models

2.1. Study Area and Data. The BNB (Figure 1) with the flow outlet at Khartoum has a drainage area of about 325,000 km^2 which extends in both Ethiopia and Sudan in Africa. The Blue Nile with a total length of about 1,460 km flows into and out of Lake Tana and emanates from the Ethiopian highlands based on the two main tributaries including the Dinder and Rahad rivers. From Lake Tana to El Diem which is at the Ethiopian-Sudanese border, the length of the Blue Nile is about 940 km. The basin can receive annual rainfall up to about 2000 mm, though in some years it can reduce to less than 1000 mm. The climate of the basin is characterized by seasonal migration of the intertropical convergence zone.
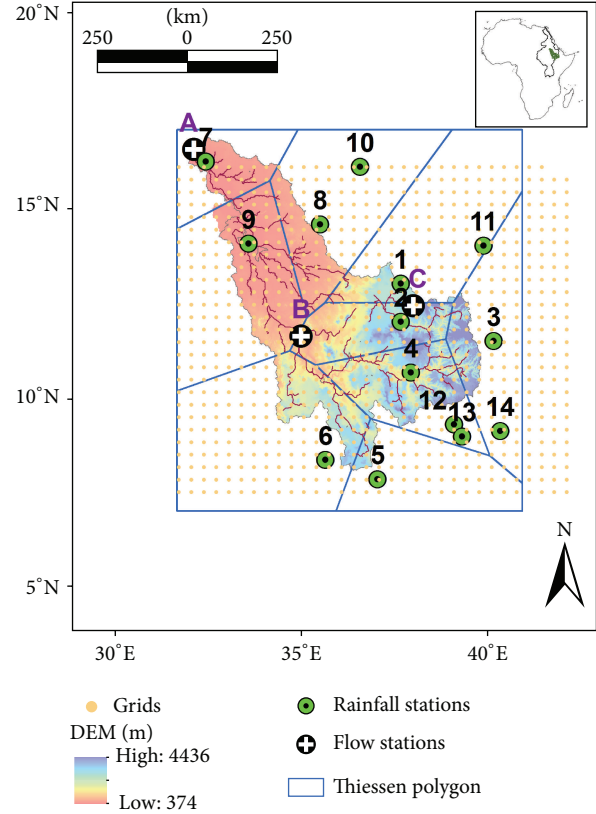

Figure 1: Location of the BNB. The labels of the flow and rainfall stations are consistent with those in Table 1. The Digital Elevation Model (DEM) used as the background map was obtained online from the International Centre for Tropical Agriculture, CIAT-CSI SRTM website, http://srtm.csi.cgiar.org/ (accessed: 20th October, 2010).

Daily rainfall data at 14 locations (Figure 1) in and around the BNB were obtained online from the Global Historical Climatology Network (GHCN) [28, 29] via the link http://www.ncdc.noaa.gov/oa/climate/ghcn-daily/ (accessed: 11th June, 2014). Gridded (0.3° × 0.3°) Climate Forecast System Reanalysis (CFSR) rainfall data from the National Centers for Environmental Prediction (NCEP) were also obtained via the web link http://cfs.ncep.noaa.gov/cfsr/ (accessed: 3rd February, 2016). The rainfall data obtained from the GHCN were mainly from 1965 to 1990. The percentage of the missing rainfall records from the GHCN was minimal (Table 1). From 1991 to 2000, the rainfall data from the GHCN were augmented by the daily CFSR data. The missing rainfall records were in-filled using the inverse distance weighted interpolation technique as applied before for the rainfall of the Nile Basin by [30]. Consider Ω_H as the missing rainfall intensity at the meteorological station H, let k be the distance between H and another station in the neighborhood being used for interpolation, and take η as the power parameter. Using τ other neighboring rainfall stations, Ω_H for a particular period was estimated using

$$\Omega_H = \frac{\sum_{i=1}^{\tau} \Omega_i \times k_i^{-\eta}}{\sum_{i=1}^{\tau} k_j^{-\eta}}. \tag{1}$$

TABLE 1: Meteorological and hydrological stations.

S. number	GHCN_ID	Station	Long.	Lat.	PM (%)
		Rainfall			
1	ET000063331	Gondar	37.42	12.55	0.0
2	ET000063332	Bahar Dar	37.42	11.60	0.0
3	ET000063333	Combolcha	39.73	11.12	0.3
4	ET000063334	Debre Marcos	37.67	10.33	0.0
5	ET000063402	Jimma	36.83	7.67	0.1
6	ET000063403	Gore	35.53	8.15	0.6
7	SU000062721	Khartoum	32.55	15.60	0.3
8	SU000062752	Gedaref	35.40	14.03	0.0
9	SU000062762	Sennar	33.62	13.55	0.8
10	SU000062730	Kassala	36.40	15.47	0.9
11	ET000063330	Makale	39.48	13.50	1.1
12	ET000063450	Addis Ababa-Bole	38.75	9.03	0.5
13	ET000063451	Harar Meda	38.95	8.73	0.6
14	ET000063453	Metehara	39.90	8.87	0.7
		Flow			
A	***	Khartoum	32.51	15.64	0.0
B	***	El Diem	34.93	11.24	0.0
C	***	Ribb	37.73	12	0.0

***: missing station ID; PM: percentage of missing records; long.: longitude [°]; lat.: latitude [°].

For a realistic value of the interpolated record, η was set to 2. This was to obtain optimal weights to the points both far and near the station with the missing record.

Daily flow series from 1965 to 2000 observed at Khartoum was adopted from a study by [23]. Daily flow data at El Diem and Ribb observed from 1980 to 2000 were obtained from personal sources. Minimum and maximum temperatures were also downloaded from the data sources similar to those of the rainfall series.

2.2. Rainfall-Runoff Models. Five internationally well-known models including IHACRES, AWBM, TANK, SAC, and SIMHYD were obtained from the "eWater Toolkit" of the Cooperative Research Centre for Catchment Hydrology in Australia via the link http://www.toolkit.net.au/ (accessed: 25th August, 2015). These models were selected because (1) they are freely available online, (2) by modeling the runoff in rather lumped than distributed way, they conform with the aim of the study, and (3) their robustness for hydrological modeling has been demonstrated for several climatic zones by several studies, a few examples of which include [31–37].

For the details on the structures and rainfall-runoff generation processes which these models can simulate, the reader is referred to [15–17, 38] for IHACRES and to [39] for AWBM, TANK, SAC, and SIMHYD. However, some brief description on each model as well as the general structural

details can be obtained from Appendices A.1 to A.5. The model parameters are also presented in Appendices B.1 to B.2.

3. Methodology

3.1. Rainfall-Runoff Modeling. All the selected five models use catchment-averaged rainfall and potential evapotranspiration (PET) as inputs. Based on the PET and the soil water storages, the actual evapotranspiration is calculated by the models. The PET was computed using the FAO Penman-Monteith method [40] considering minimum and maximum temperature.

First and foremost, before applying the models to the entire BNB to simulate the long-term runoff at the Khartoum flow outlet, it was deemed important to investigate the influence of spatial extent of the catchment on the capacity of the models to simulate the runoff. Daily river observed from 1980 to 2000 at Khartoum, El Diem, and Ribb with the catchment areas of 325,000 km^2, 176,000 km^2, and 1,592 km^2, respectively, were used. Based on the rainfall stations upstream of each flow station, Thiessen polygons were constructed. The average of the observed daily rainfall in each polygon was merged with the mean of the gridded CFSR rainfall data (within the same polygon) so as to cover the period similar to that of the flows. The blended rainfall data from all the constructed polygons were jointly used to compute the catchment-wide average rainfall for the model inputs.

To limit the influence of subjectivity in the manual changing of model parameters, automatic calibration was done for all the models. A number of optimizers exist for automatic calibration including the uniform random search, Rosenbrock multistart search, Rosenbrock single start search, multistart pattern search, single start pattern search, generic algorithm [41, 42], and shuffled complex evolution (SCE) [43–45]. The SCE was adopted in this study because of two reasons. Firstly, it has been demonstrated to be efficient and effective in the calibration of a number of rainfall-runoff models (see, e.g., [44, 46–48]). Secondly, the SCE has also been widely applied for automatic rainfall-runoff calibrations (see, e.g., [49–52]). According to [6], the SCE method combines four optimization strategies to simultaneously evolve a number of potential solutions towards the region of the global optimum of the objective function. These SCE strategies include the simplex method, competitive evolution, controlled random search, and complex shuffling. Note should be taken that the adoption of the SCE was possible for the models incorporated within the Rainfall Runoff Library (RRL) of the "eWater Toolkit," that is, AWBM, TANK, SAC, and SIMHYD. For IHACRES which was obtained as a stand-alone model, the calibration process is automated in two steps. Firstly, the linear module calibration which is controlled by fixed transfer function is done using cross-correlation to calculate the delay between rainfall and stream flow. Secondly, the nonlinear module calibration is used to perform a grid search through the parameter space to obtain the best parameter set. Further detail on this calibration procedure for the IHACRES can be found in [38]. For the calibration of all the models, the

optimization objective function was in terms of the well-known Nash-Sutcliffe Efficiency (NSE) [53].

After assessing the influence of spatial variation on the adequacy of the modeling results, the models were rebuilt using the long-term model inputs considering the entire BNB (i.e., with the flow outlet at Khartoum). To obtain catchment-wide rainfall average, Thiessen polygon was constructed using all the 14 rainfall stations. For each model, the automatic calibration was done based on daily series from 01/01/1965 to 31/12/1990. The validation of the models was done using daily data from 01/01/1991 to 31/12/2000. The "goodness-of-fit" between the observed and modeled flows was assessed using the NSE for both calibration and validation periods.

3.2. Comparison of Observed and Modeled Flows

3.2.1. "Goodness-of-Fit" of the Entire Time Series.
Firstly, the "goodness-of-fit" between the observed and modeled flows considering the entire or full-time series (covering both the calibration and validation periods) was, again, assessed using the NSE. Secondly, to reduce/increase the sensitivity of the NSE to high/low flows, the logarithmic values of observed and modeled flows were to be used. However, in this study, because some modeled flows were zero in value for all the models (except TANK), the use of logarithmic values to compute NSE was not feasible. Eventually, two other forms of statistical "goodness-of-fit" metrics which consider the absolute relative deviations between the observed and modeled flows as presented by [54] were used. Consider Q_m as the modeled flow, Q_o the observed flow, $\overline{Q}_o$ the mean of observed flow, and n the sample size. The relative efficiency (E_f) and index of agreement (I_d) were computed using

$$E_f \, [-] = 1 - \frac{\sum_{i=1}^{n} |Q_{m,i} - Q_{o,i}|^{P_q}}{\sum_{i=1}^{n} |Q_{o,i} - \overline{Q}_o|^{P_q}},$$

$$I_d \, [-] = 1 - \frac{\sum_{i=1}^{n} |Q_{o,i} - Q_{m,i}|^{P_q}}{\sum_{i=1}^{n} (|Q_{m,i} - \overline{Q}_o| + |Q_{o,i} - \overline{Q}_o|)^{P_q}}. \tag{2}$$

If the assessment of the models is with respect to high flows, values of the power P_q greater than 1 can be used. However, in this study, $P_q = 1$ was used to obtain a balance between high flows and low flows given that the full-time series was being considered.

Graphically, comparison of the simulated and observed cumulative flows was also made to investigate the water balance closing capacity of each model. Using the annual mean flow, scatter plot points of observed versus modeled series was also made. This was after Box-Cox (BC) transformation procedure (as will be described in Section 3.2.2).

3.2.2. "Goodness-of-Fit" of Flow Extremes.
Using the entire or full-time series (covering both the calibration and validation periods), flow extremes were extracted based on a number of criteria for the model performance evaluation. Firstly, the forecast accuracy metrics H_{FC} and L_{FC} relevant for high

flows and low flows, respectively, as proposed by [55] were computed using

$$H_{FC} \, [-] = \frac{\left(\sum_{j=1}^{N_H} \left(Q_{o,j} - Q_{m,j} \right)^2 Q_{m,j}^2 \right)^{0.25}}{\left(\sum_{j=1}^{N_H} Q_{o,j}^2 \right)^{0.5}},$$

$$L_{FC} \, [-] = \frac{\left(\sum_{j=1}^{N_L} \left(Q_{o,j} - Q_{m,j} \right)^2 Q_{m,j}^2 \right)^{0.25}}{\left(\sum_{j=1}^{N_L} Q_{o,j}^2 \right)^{0.5}}, \tag{3}$$

where N_L is number of low flow events lower than one-third of the mean low flow observed; N_H is number of high flow events greater than one-third of the mean peak flow observed. In this study, the mean peak high flow was computed as the average of the events above the median flow. Similarly, the mean low flow was taken as the average of the hydrological events less than the median flow. According to [55], the best or perfect model performance is given by H_{FC} or L_{FC} equal to zero. However, it is noticeable that, even for the worst models with the all values of Q_m equal to zero, the values of H_{FC} or L_{FC} also become zero. Eventually, caution must be taken to use the metrics H_{FC} and L_{FC} alongside other statistical "goodness-of-fit" measures. In this study, the model average bias (M_{AB}) and root mean squared error (R_{MSE}) were selected to supplement the evaluations using H_{FC} and L_{FC}. M_{AB} and R_{MSE} were computed using (4) by setting N equal to N_H or N_L. For an unbiased model, M_{AB} is ideally equal to zero. Consider

$$M_{AB} \, (\%) = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{Q_{m,i} - Q_{o,i}}{Q_{o,i}} \times 100 \right),$$

$$R_{MSE} \, (\text{m}^3 \, \text{s}^{-1}) = \left(\frac{1}{N} \sum_{i=1}^{N} (Q_{m,i} - Q_{o,i})^2 \right)^{0.5}. \tag{4}$$

To investigate the effect of adjustment of extreme flow extraction threshold on the model performance, H_{FC}, L_{FC}, M_{AB}, and R_{MSE} were also computed by setting N_H (or N_L) from (3) and N of (4) in terms of (1) the number of events higher (or lower) than the 95th (5th) percentile of observed flows and (2) the number of annual maxima and minima flows.

Graphically, plots of the Box-Cox (BC) transformed observed versus modeled series were made for the minimum as well as maximum flows in each year. To give similar weights to the maximum and minimum flows (q) so as to obtain homoscedastic model residuals, the parameter (λ) of the BC [56] transformation (see (5)) was set to 0.25. Consider

$$BC \, (q) = \frac{q^{\lambda} - 1}{\lambda}. \tag{5}$$

3.2.3. Analyses of Temporal Changes.
The differences in the changes between observed and modeled flows were investigated using the maximum, mean, and minimum flow in each year. The considered changes in flow were in terms of the variability and subtrends. Subtrends are short-durational

changes in trend direction over unknown periods within the series [56]. The identification and assessment of the temporal subtrends and investigation of variability are vital to ascertain the possibility of any intervention of, for example, climate fluctuations on the hydrology.

The subtrend analysis was done using the cumulative rank difference (CRD) technique [56]. To graphically reveal the hidden short-durational changes (e.g., jumps in the mean, subtrends, etc.) within the annual mean flows and the minimum as well as maximum flow in each year, CRD plots were used. To construct the CRD plots, the following steps were taken:

(i) Considering v as the number of times a data point is exceeded, w as the number of times a data point appears within the given sample and n as the sample size, the difference (d) between the exceedance and nonexceedance counts of the data points is computed using

$$d_i = 2v_i - (n - w_i) \quad \text{for } 1 \le i \le n. \tag{6}$$

To determine v or w, each data point is counted as if it was not considered before [57]. Considering the imaginary series (3, 2, 7, 4, 3, 4), $n = 6$ and for $i = 1, 2, \ldots, n$, $v = (3, 5, 0, 1, 3, 1)$, $w = (2, 1, 1, 2, 2, 2)$, and $d = (2, 5, -5, -2, 2, -2)$.

(ii) The cumulative sum (c) of d from (6) is given by

$$c_i = \sum_{j=1}^{i} d_j \quad \text{for } 1 \le i \le n. \tag{7}$$

For the series in Step (i), $c_i = (2, 7, 2, 0, 2, 0)$. It can be checked that, at $i = n$, c_i is always zero.

(iii) The CRD plot was made in terms of the graph of c_i against the time unit of the series.

The short-durational changes from the CRD plots can be identified based on the graphical guidelines illustrated in Figure 2 which was generated based on synthetic series P of $n = 200$. In the CRD plot, the $c = 0$ line is taken as the reference (i.e., the case when the data points within the series are of the same value). The values above or below this reference are considered to characterize subtrends in the series. If the series has no trend, the scatter points cross the reference several times (see case (a) and (1)). When the series has a positive/negative trend, most of (if not all) the scatter points take the form of a curve above/below the reference (see case (b) and (2)). It is possible that different changes can occur within the same series. For instance, when the series has its first/second half characterized by positive/negative trend, two curves are formed such that the first one is above the reference and the other is below the $c = 0$ line (see case (c) and (3)). For a step upward/downward jump in the mean of the series (assuming there is no trend in both parts of the subseries before and after the step jump), the scatter points form lines which meet at a point (call it the vertex) above/below the reference (see case

(d) and (4)). For an upward/downward jump, the slope of the first line is positive/negative, while that for the second one is negative/positive. For further details on the use of the CRD plots to identify changes in the series and how the significance of the visualized overall trend can be tested, the reader is referred to [57].

Variability was investigated using the nonparametric anomaly indicator method (NAIM) [58]. NAIM relies on nonparametric rescaling of the data followed by temporal convolution or aggregation of the series. According to World Meteorological Organization (WMO) [59], aggregation of series allows study of the general or summarized and representative behavior of the data. To implement the NAIM to compute anomalies in the series, the following was done:

(i) A block length (L_B) based on a sensitivity analysis was selected. This was done while ensuring that the periods of high or low anomalies were as nearly independent as possible. To capture the decadal oscillations (if any), sensitivity analysis was done using L_B in the range from 2 to 10 years. Finally $L_B = 10$ years was adopted because it gave a more representative smoothing of the series than for other values such as $L_B = 2, 5,$ and 7 years.

(ii) Each value of d from (6) was transformed using $X_i = -1 \times d_i$ for $1 \le i \le n$; this was to make the temporal variation of the rescaled series match that of the given data while circumventing the influence of possible outliers (if any) on the variability analysis.

(iii) Using L_B and rescaled series X from Steps (i) and (ii), respectively, temporal aggregation was done in an overlapping way using (8) to obtain the mean (M) of the subseries in each time slice j; consider

$$M_j = \frac{1}{L_B} \sum_{i=j}^{z} X_i, \tag{8}$$

where $z = j + (L_B - 1)$ and $1 \le j \le (n + 1 - L_B)$.

(iv) The anomaly (%) in each slice j was calculated as a ratio of $100M_j$ to $(n - 1)$. Actually, $(n - 1)$ is equal to the possible maximum absolute value of X from Step (ii). Thus, the possible maximum anomaly is 100%.

(v) Finally, the anomaly (%) was plotted against the midpoints of the time slices. From the plot, the long-term mean of the series is represented by the zero percent anomaly. The fluctuation of the anomalies above or below the reference was considered to characterize the variability in the series.

If the L_B of 10 years selected in Step (i) is moved in an overlapping way as described by (8), the resultant time slices based on the data considering Khartoum flow outlet, that is, from 1965 to 2000, would be 1965–1974, 1966–1975, ..., 1991–2000. The mid-points of the time slices (as required in Step (v)) against which the anomalies from Step (iv) are plotted would be 1970, 1971, ..., 1996.

For validity of the NAIM results, the significance of the anomalies can be tested to verify the null hypothesis (H_0)
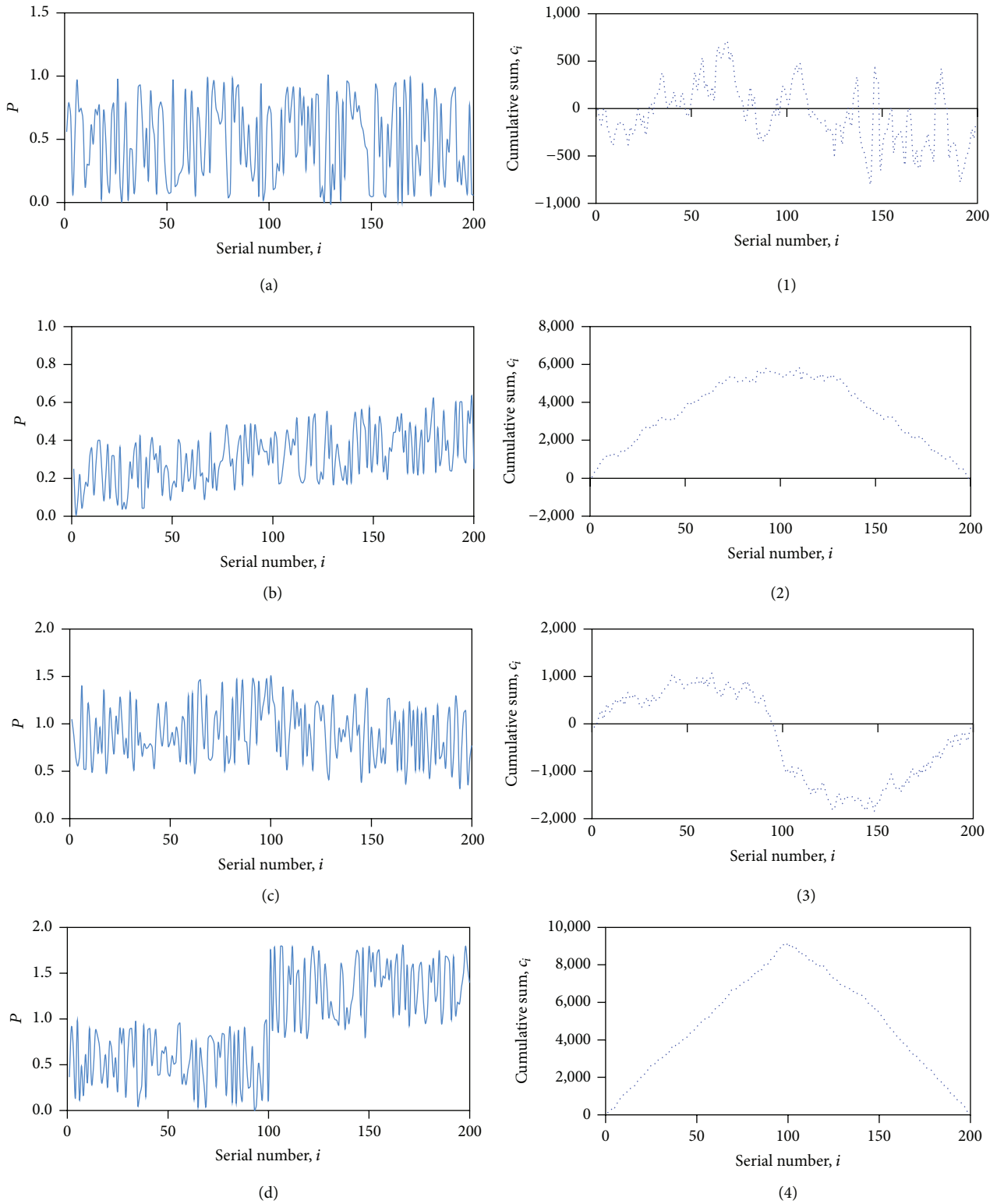
(a)

(1)

(b)

(2)

(c)

(3)

(d)

(4)

FIGURE 2: The plots for (a)–(d) synthetic series, P of $n = 200$, and (1)–(4) the cumulative effects of the temporal variations for the corresponding series.

that the variability in the series is caused by only natural randomness; that is, there is no persistence in the temporal variation. Nonparametric bootstrapping [60] by Monte Carlo simulations is used to test H_0. Consider N_{MC} as the number of Monte Carlo simulations; the variability bounds in the form of $(100 - \alpha)\%$ Confidence Interval (CI) were constructed by (1) applying the NAIM using L_B to the original series, (2) randomly shuffling the original full-time series, (3) applying the NAIM using L_B to the shuffled series, (4) repeating Steps (2)-(3) N_{MC} times to obtain N_{MC} sets of anomaly values, (5) ranking, for each set, the anomaly values from the highest to the lowest, and (6) taking the upper and lower limits of the $(100 - \alpha)\%$ CI as the $[0.005 \times \alpha\% \times N_{MC}]$th and $[\{1 - (0.005 \times \alpha\%)\} \times N_{MC}]$th anomaly values, respectively. H_0 is accepted if the anomaly values from Step (1) fall within the CI. If the upper/lower CI limit is upcrossed/downcrossed, H_0 is rejected. In this study, $\alpha\%$ and N_{MC} were set to 5% and 1000, respectively.

To evaluate the performance of the models in the simulation of the observed temporal variability, the following steps of the nonparametric skill score (SC, %) were taken:

(i) The zero percent anomaly was taken as the reference for natural randomness in the data. The symbols "+" or "−" were assigned to a particular time slice if its anomaly was greater or less than zero, respectively. If the anomaly was zero, "∗" was assigned.

(ii) Again, the symbol "+" was assigned to a particular time slice if its anomaly was greater than or equal to the upper limit of the 95% CI; otherwise, the symbol "−" was assigned.

(iii) Furthermore, the symbol "+" was assigned to a particular time slice if its anomaly was less than or equal to the lower limit of the 95% CI; otherwise, "−" was assigned.

(iv) Rewards and penalties were awarded for the scores assigned to the anomalies from the models. For a given model, if its modeled flow correctly obtained the anomaly sign (i.e., + or −) similar to that of the observed flow in a particular time slice, a score of +1 was given to the model; otherwise a penalty of −1 was assigned. This procedure was repeated separately for each of the criteria from Steps (i) to (iii).

(v) The number of criteria under which the scores were assigned was determined and denoted as u; in this case $u = 3$ because there were three aspects of accuracy tested, that is, in Steps (i) and (iii). Consider that $h_1, h_2, \ldots,$ and h_u denote the sum of scores for the first, second, . . ., and the uth criterion, and if $k_1, k_2, \ldots,$ and k_u are the sample sizes of events for which scores were assigned using the first, second, . . ., and the uth criterion, generally the balance (b) between the total of the rewards and penalties would be given by $b = (h_1/k_1) + (h_2/k_2) + \cdots + (h_u/k_u)$.

(vi) The SC (%) for a particular model was calculated using

$$SC\,(\%) = \begin{cases} b \times 100 \times u^{-1} & \text{if } b > 0 \\ 0 & \text{if } b \leq 0. \end{cases} \tag{9}$$

The best model performance is given by SC of 100%. The worst model is that with SC of zero.

3.2.4. Amplitude-Duration-Frequency Relationships. Amplitude-Duration-Frequency (ADF) relationships combine Extreme Value Distribution (EVD) over a range of aggregation levels. ADF relationships can be considered as one of the most important tools for risk-based water engineering and management. ADF relationships are used for design, operation, and/or management of water supply projects (e.g., dikes, dams, and irrigation systems) [61] or urban drainage facilities such as sewer conduits [62]. ADF relationships constructed for flow and rainfall are called QDF and IDF, respectively. In this study, QDF relationships were constructed for both observed and simulated flows.

The first step in the construction of QDF relationships is the selection of aggregation levels, that is, durational intervals over which flows are averaged. To cover the relevant water resources management or water engineering applications as agriculture, irrigation, hydropower, domestic water supply, pollution control, and so forth, aggregation levels may be selected in the range of 1–90 days for high flows and the range of 1–365 days for low flows [63].

Equation (8) with the term L_B taken as the aggregation level and X as the original (i.e., unrescaled) flow is used in averaging of the series. For analyses of low flows from nonephemeral rivers (i.e., when $X > 0$), the aggregation of X is done after transformation of X by $(1/X)$. For low flow analyses, mostly Weibull or Fréchet distributions are used. However, the transformation of the series using $(1/X)$ makes the low flows to follow the Generalized Pareto Distribution GPD [64] or exponential instead of Weibull or Fréchet distribution. This transformation allows both the low flows and high flows to be analyzed in a similar way. Consider $G(x)$ as the cumulative distribution function of the GPD, and assume that $(1/X)$ low flows exhibit normal tail in the exponential quantile plot (i.e., $-\ln\{1 - G(x)\}$ in abscissa and x in ordinate). As shown by [65], to calibrate exponential distribution (with scale (α) and location or threshold (x_t) parameters) to the series ($S = 1/X$) above a specified threshold s_t,

$$G\,(s) = P\{S \leq s \mid S \geq s_t\} = 1 - \exp\left\{-\frac{(s - s_t)}{\alpha}\right\}. \tag{10}$$

This equation can be transferred towards a distribution for X (using $x_t = 1/s_t$) as follows:

$$\begin{aligned} G\,(s) &= P\{X \leq x \mid X \leq x_t\} = P\{S \geq s \mid S \geq s_t\} \\ &= 1 - P\{S \leq s \mid S \geq s_t\}, \end{aligned} \tag{11}$$

$$G(s) = \exp\left\{-\frac{(s - s_t)}{\alpha}\right\} = \exp\left\{-\frac{(x^{-1} - x_t^{-1})}{\alpha}\right\}, \quad (12)$$

$$G(x) = P\{X \le x \mid X \le x_t\} = \exp\left\{-\frac{(x^{-1} - x_t^{-1})}{\alpha}\right\}$$
$$= \exp\left(-\frac{x^{-1}}{\alpha}\right)\left\{\exp\left(-\frac{x_t^{-1}}{\alpha}\right)\right\}^{-1}. \quad (13)$$

For values lower than x_t, it can be noted that (13) matches the Fréchet distribution $G(x) = \exp(-x^{-\eta}/\alpha)$ with $\eta = 1$.

The next step after the temporal averaging of the series is the selection of independent hydrological extremes for each aggregation level. In line with frequency analysis, the requirement that the data should be independent and identically distributed can be achieved through extraction, from the full series, of either Annual Maxima Series (AMS) or the Partial Duration Series (PDS)/Peak Over Threshold (POT). The main advantage of the AMS in which the time slice is chosen to be one hydrological year is that it produces extremes with stronger independence compared to those of the POT method. However, it has the disadvantages that the sample is rather small, and because the second highest event in each year is not considered, some useful information for the definition of the extreme value region is lost. In the POT method, all events above a certain truncation (threshold) level are extracted, thereby leading to a more reasonable sample size for extreme value analysis than that of the AMS. Thus, the POT method intuitively provides a more consistent definition of the extreme value domain than the AMS approach [66]. Eventually, the POT method of extracting independent hydrological extremes was adopted in this study. The POT events were selected using the independence criteria based on the flow threshold and the time between successive flow extremes as presented by [67].

The step that follows the POT extraction is fitting of the EVD to the independent extreme events. It is known that the POT events as used in this study follow the GPD (see (14) and (15)) which can be valid for values of x above the threshold x_t. Consider

$$G(x) = 1 - \left\{1 + \gamma\frac{(x - x_t)}{\alpha}\right\}^{-1/\gamma} \quad \text{for } \gamma \ne 0, \quad (14)$$

$$G(x) = 1 - \exp\left\{-\frac{(x - x_t)}{\alpha}\right\} \quad \text{for } \gamma = 0. \quad (15)$$

Three classes of the GPD can be identified based on the shape parameter γ, that is, normal ($\gamma = 0$), heavy ($\gamma > 0$), and light ($\gamma < 0$) tails. When $\gamma \ge 0$, the upper GPD tail goes up to infinite values and for $\gamma < 0$, the GPD has a final right-end point. The GPD for $\gamma = 0$ equals the exponential distribution. The parameters of the EVD were estimated using the weighted linear regression (WLR) technique based on the quantile incremental properties of the GPDs [68]. The WLR technique was adopted because it was recently demonstrated by [66] to be ostensibly more robust to capture EVD tails than other well-known parameter estimation methods such as the method of moments, maximum likelihood, and L-moment approach. To compute the parameters of the GPD (see (14)), Pareto quantile plot (i.e., $-\ln\{1 - G(x)\}$ in abscissa and $\ln(x)$ in ordinate) is used. In this plot, the GPD (see (14)) appears as a line and the slope of this line especially in the tail approximates to γ. Eventually, for the GPD (see (14)), α is computed using (16), and γ can be estimated by least square weighted linear regression based on the Hill weights [69] using (17). To compute the parameters of the exponential distribution (see (15)), exponential quantile plot (i.e., $-\ln\{1 - G(x)\}$ in abscissa, and x in ordinate) is used. In this plot, the exponential distribution takes the form of a line whose slope is equal to α which can be estimated again based on the Hill weights [69] using (18). Consider t as the number of POT events above the threshold x_t and

$$\alpha_t = \gamma_t \times x_t, \quad (16)$$

$$\gamma_t = \frac{1}{t - 1}\left\{\sum_{i=1}^{t-1}\ln(x_i)\right\} - \ln(x_t), \quad (17)$$

$$\alpha_t = \frac{1}{t - 1}\left\{\sum_{i=1}^{t-1}x_i\right\} - x_t. \quad (18)$$

The optimal x_t is obtained as the value of x for which the mean squared error (MSE) is minimal. For the GPD (see (14)) and exponential distribution (see (15)), the MSE of the WLR in the Pareto and exponential quantile plot can be given by (19) and (20), respectively. Consider

$$\text{MSE}_t = \frac{1}{t - 1}\left\{\sum_{i=1}^{t-1}\left(-\ln\left(\frac{i}{t}\right)\right)^{-1} \cdot \left(\ln\left(\frac{x_i}{x_t}\right) - \gamma_t \ln\left(\frac{t}{i}\right)\right)^2\right\}, \quad (19)$$

$$\text{MSE}_t = \frac{1}{t - 1}\left\{\sum_{i=1}^{t-1}\left(-\ln\left(\frac{i}{t}\right)\right)^{-1} \cdot \left(x_i - x_t - \alpha_t \ln\left(\frac{t}{i}\right)\right)^2\right\}. \quad (20)$$

The second last step of QDF construction comprises the estimation of the flow quantiles. Considering r as the data record length in years and i as the rank of the POT events ($i = 1$ for the highest), the relationship between the T-year flow Q_T and the return period T based on calibrated GPD and empirical data can be calculated using (21) and (22), respectively. Consider

$$T\text{(years)} = \left(\frac{r}{t}\right)\frac{1}{1 - G(Q_T)}, \quad (21)$$

$$T\text{(years)} = \left(\frac{r}{i}\right). \quad (22)$$

Empirically, Q_T is obtained as the flow value corresponding to T computed using (22). Based on (22), it is noticeable that the
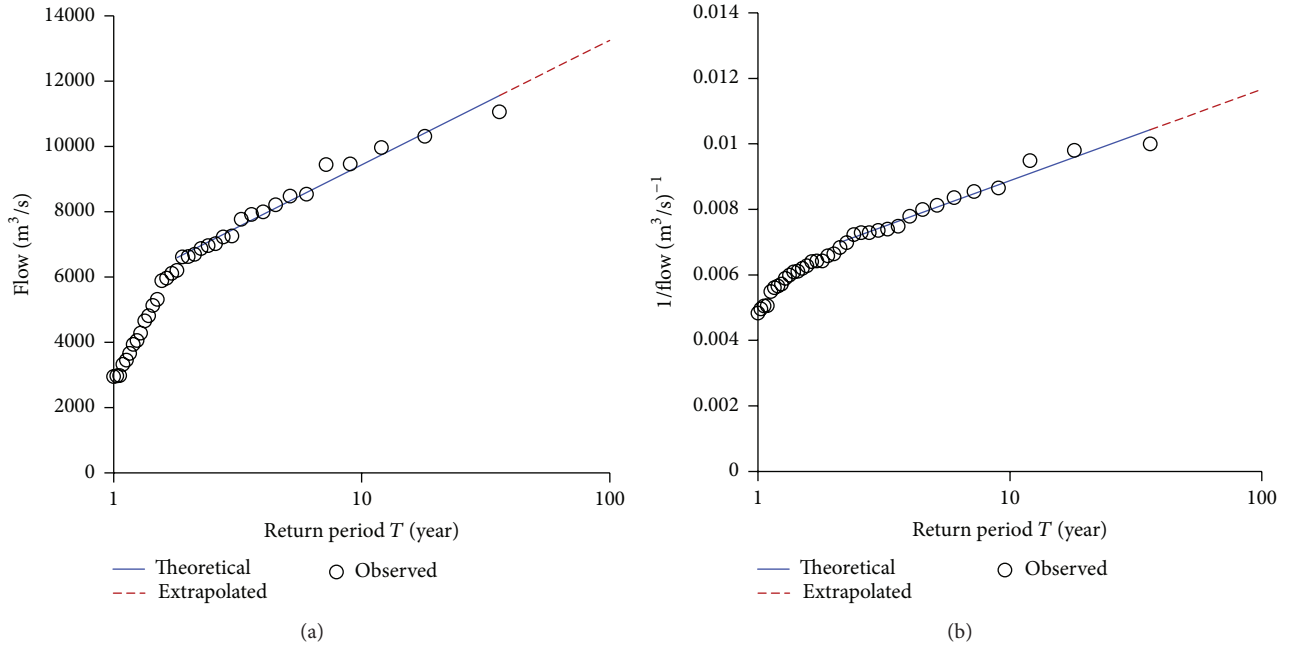
FIGURE 3: POT events in exponential quantile plots for (a) high flows and (b) low flows of 1-day aggregation level. The regression line is the fitted EVD.

empirical Q_T can only be estimated for T values not exceeding r. For T greater than r, the theoretical quantiles can be estimated for the GPD (see (14)) and exponential distribution (see (15)) using (23) and (24), respectively. Extrapolation requires the choice of T values to be made. In this study, to minimize the possible uncertainty boost in the extreme value analyses due to finite sample size, T was ensured not to exceed 100 years (i.e., less than three times the data record length). Besides, the range of T from 5 to 100 can be used in planning and designing of multipurpose risk-based water engineering applications such as hydraulic structures along river systems (bridges, culverts, etc.). T around 100 years may be used for medium-sized flood protection systems, flood plain development, and so forth [62]. In the construction of the QDF relationships, for each aggregation level the theoretical quantiles Q_T are computed for all the selected return periods. Consider

$$Q_T = \exp\left(\ln(x_t) + \gamma\left\{\ln(T) - \ln\left(\frac{r}{t}\right)\right\}\right) \tag{23}$$
$$\text{for } \gamma \neq 0,$$

$$Q_T = x_t + \alpha\left\{\ln(T) - \ln\left(\frac{r}{t}\right)\right\} \quad \text{for } \gamma = 0. \tag{24}$$

Figure 3 shows examples of calibrated exponential distribution or normal tailed GPD as regression lines in exponential quantile plots for both high flows and low flows. It is noticeable that the linear behavior of the quantiles is obtained towards the tail of the distribution of events. The regression lines can be seen to be extrapolated up to T of 100 years for both high flows (Figure 3(a)) and low flows (Figure 3(b)). For low flows, back transformation is applied to the $(1/X)$ transformed T-year events to obtain the actual quantiles.

Finally, the QDF relationships comprise the compiled values of Q_T for all the aggregation levels as well as the different selected return periods. This makes possible the estimation of cumulative volumes of water during drought or flood periods at various aggregation levels or return periods.

4. Results and Discussions

4.1. Rainfall-Runoff Modeling

4.1.1. Effect of Catchment Size on Simulation of Runoff. Figure 4 shows the performance of the models in simulating runoff for catchments of different drainage areas using hydrometeorological inputs from 1980 to 2000. It is noticeable that the flows of Ribb (Figures 4(k)–4(o)) are less smooth than those of El Diem (Figures 4(f)–4(j)) and Khartoum (Figures 4(a)–4(e)). Normally, the time taken by the water to reach the catchment outlet when the area or size is large tends to be long. The increasing delay in the catchment response as the catchment size increases leads to low runoff volume from rainfall following the large losses of rain water by infiltration, evaporation, and so forth. Given that the flow is cumulative in volume from upstream to downstream, the changes in flow due to the fluctuations in rainfall-runoff tend to be more smoothened as the catchment size increases. However, regardless of the catchment size, it is noticeable that the observed hydrographs are well reproduced by those of the modeled flows. The statistical evidence of the agreement between the observed and modeled flows for all the catchments is provided in Table 2. The NSE for each model fell in the range of 0.70–0.81 regardless of the catchment size. Similarly, the correlation between the observed and modeled flows was between 0.8 and 0.9. This

(a)

(b)

(c)
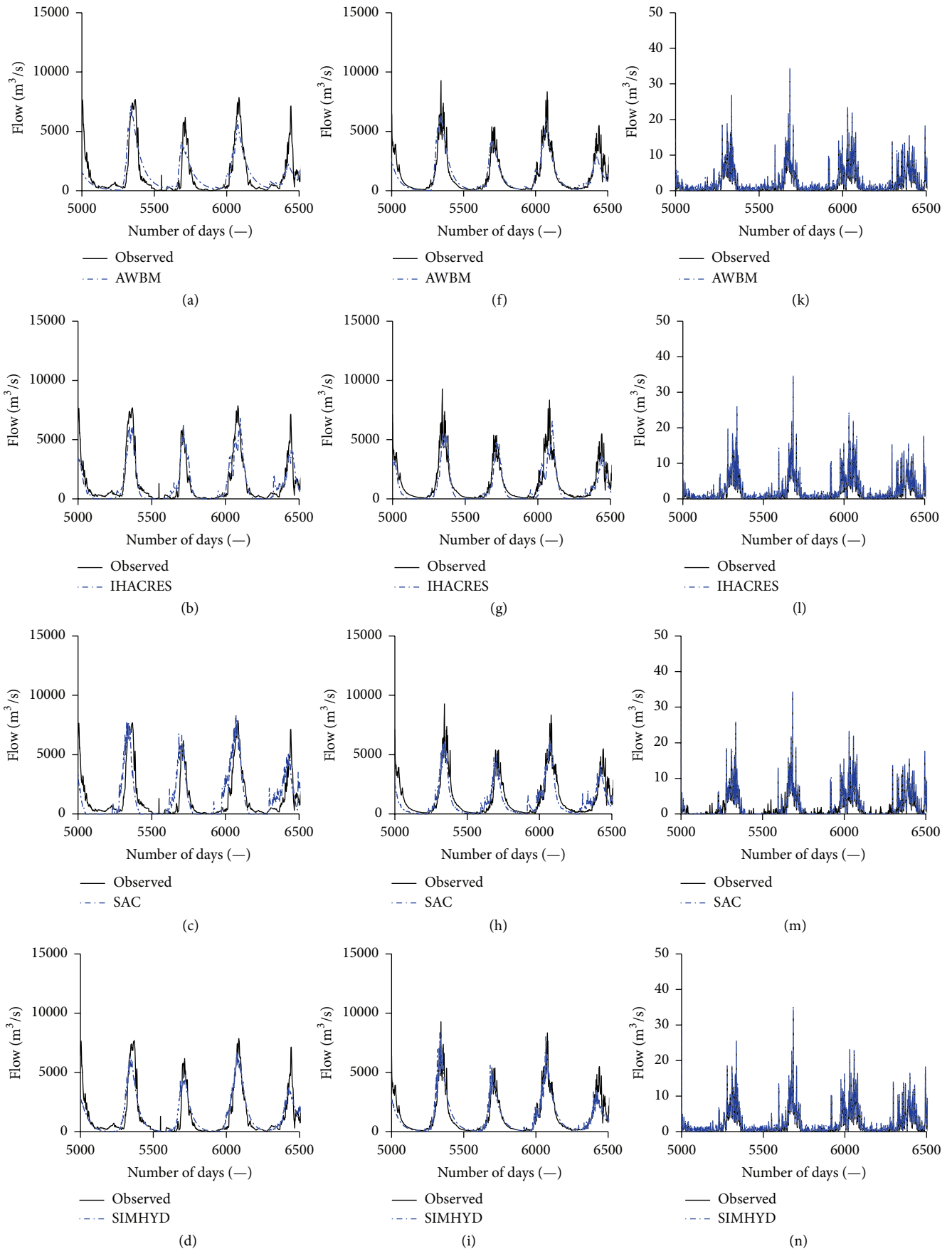
(d)

(f)

(g)

(h)

(i)

(k)
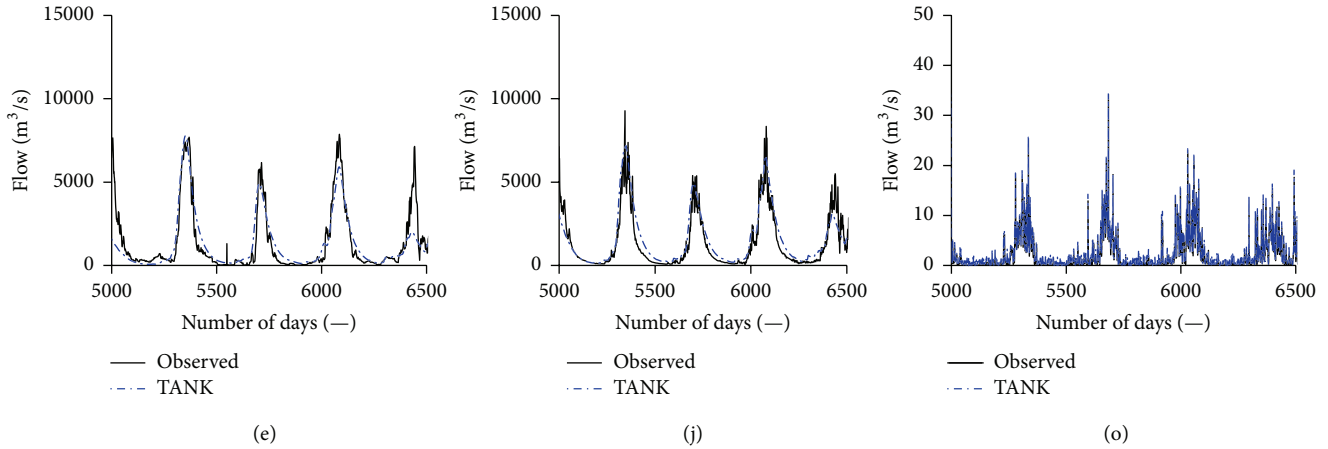
(l)

(m)

(n)

Figure 4: Continued.

FIGURE 4: Observed and modeled daily flows at (a)–(e) Khartoum, (f)–(j) El Diem, and (k)–(o) Ribb based on (a, f, k) AWBM, (b, g, l) IHACRES, (c, h, m) SAC, (d, i, n) SIMHYD, and (e, j, o) TANK. The number of days from 5000 to 6500 in the horizontal axis covers the period of flows from 9/8/1993 to 17/10/1997.

TABLE 2: Statistical evaluation of the model performance.

Model	Khartoum		El Diem		Ribb	
	NSE	Corr.	NSE	Corr.	NSE	Corr.
AWBM	0.71	0.81	0.80	0.86	0.71	0.82
IHACRES	0.76	0.78	0.77	0.81	0.70	0.81
SAC	0.72	0.82	0.78	0.84	0.72	0.80
SIMHYD	0.78	0.85	0.81	0.86	0.74	0.83
TANK	0.70	0.79	0.79	0.85	0.76	0.84

Corr.: correlation between observed flow and simulated flow.

TABLE 3: Statistical measures of the agreement between the observed and simulated flows.

Model	AWBM	IHACRES	SAC	SIMHYD	TANK
Calibration	0.86	0.73	0.74	0.75	0.79
Validation	0.68	0.65	0.70	0.61	0.71

indicated the adequacy of the models to simulate long-term runoff from the BNB. Eventually, to evaluate the performance of the models in reproducing the variability and quantiles of the flows considering the entire BNB, all the models were applied to simulate runoff at Khartoum and the modeling results are presented in Section 4.1.2.

4.1.2. Model Results for the Runoff at Khartoum. Figure 5 shows the observed and simulated daily flows from 1965 to 2000. The calibration and validation periods are shown by the number of days in the ranges of 1–9496 and 9497–13149, respectively. It is visually noticeable that the performance of each model (Figures 5(a)–5(e)) is acceptable. However, the tendency of the SIMHYD (Figure 5(d)) to overestimate the high peak flows is also evident. The model parameters following the SCE technique of calibration are listed in Appendices B.1 and B.2.

Table 3 shows the statistical "goodness-of-fit" from the different models considering calibration and validation periods separately. The NSE values for calibration were higher than those for validation. With respect to the calibration results, the validation NSE for AWBM, IHACRES, SAC, SIMHYD, and TANK dropped by 20.9, 11.0, 5.4, 20.0, and 10.1%, respectively. The average drop of NSE by 13.5% (considering all the models) could be indicative of the difficulty of the models to capture the runoff generation dynamics due

to the possible change in catchment behavior. As stated in Section 3.1, the model validation was from 1991 to 2000. The land management policy in Ethiopia (where the upper part of the BNB is located) from 1991 to 2000 was different from that of the 1975–1990 period. This difference in policy was due to the changes in the government or political regimes in 1975 and 1991 [70]. For instance, from 1975 onwards, the landlord-tenant relationship was abolished by the "land-to-the-tillers" policy; however, by 1991, the government decreed state ownership of land [70]. The combination of such governmental policy changes and other factors such as widespread poverty associated with frenzied population growth promotes anthropogenic influences on the hydrology through overgrazing, deforestation, significant expansion of urbanized areas, and so forth. Such anthropogenic factors affect hydrology by altering the amount of infiltration into the soil, velocity of the overland runoff, the rate and amount of evaporation, and so forth [23], thereby modifying the catchment response to the rainfall input. The capacity of the models to capture such changes in catchment behavior tends to differ as seen in the differences in the NSE drop. However, because the NSE values for all the models were higher than 0.6 for the validation, the simulated flows were considered adequate for the comparison of the model performance in capturing the variability of the flows.

4.2. Comparison of the Observed and Modeled Flows

4.2.1. "Goodness-of-Fit" of the Full-Time Series. Figure 6 shows the results of graphical comparison of observed and
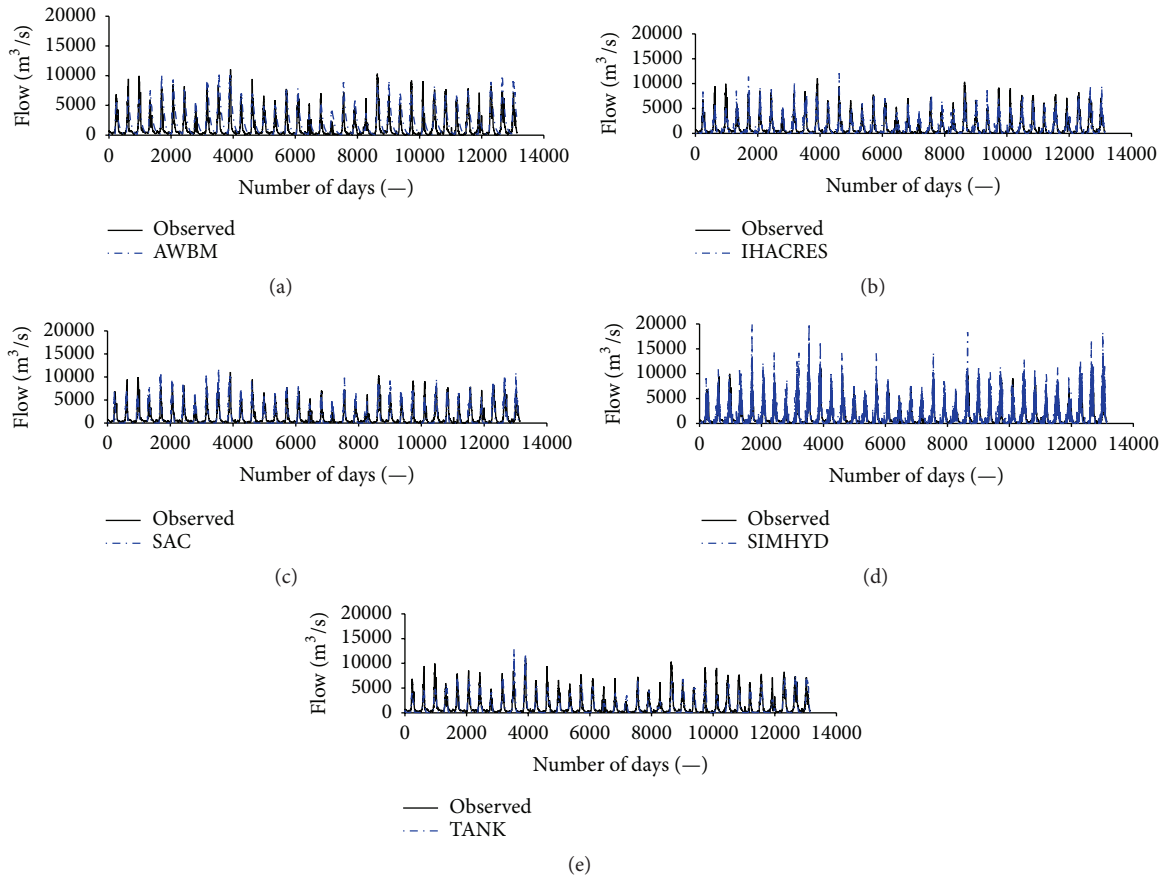
FIGURE 5: Observed and modeled daily flows from (a) AWBM, (b) IHACRES, (c) SAC, (d) SIMHYD, and (e) TANK. In the horizontal axis, the calibration period (from 1/1/1965 to 31/12/1990) covers the number of days in the range of 1–9496. Similarly, the validation period (from 1/1/1991 to 31/12/2000) is from 9497–13149 days.
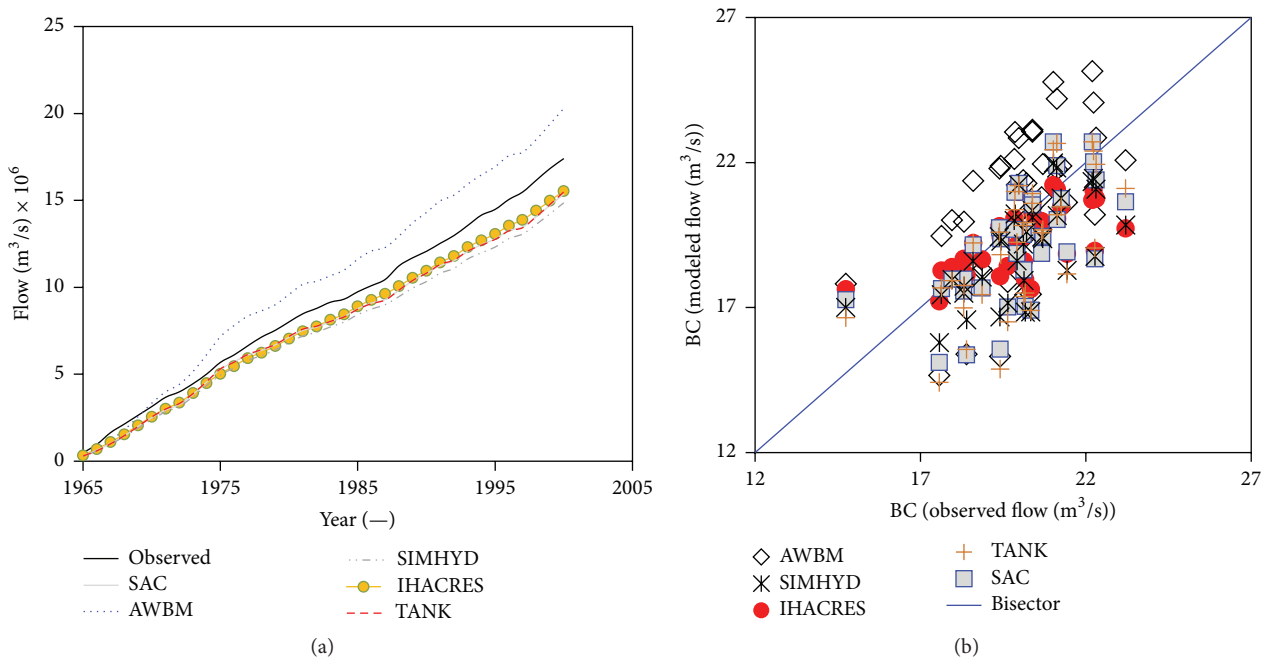

FIGURE 6: Performance of the models in terms of (a) the cumulative flow and (b) observed versus simulated annual mean flow. For the labels of the axes in chart (b), "BC" appearing before "()" shows that the flow (m^3/s) values plotted were obtained after applying Box-Cox (BC) transformation using (5).

simulated flows combined over both calibration and validation periods. It is seen that whereas the AWBM overestimated the observed cumulative flows, underestimations were obtained for the rest of the other models (Figure 6(a)). Furthermore, the deviations between the observed and simulated flows from the various models can be confirmed by most of the scatter points of the AWBM (and the rest of the models) falling above (below) the bisector (Figure 6(b)). The performances of IHACRES, SAC, SIMHYD, and TANK are shown to be comparable (Figures 6(a) and 6(b)). The average biases in reproducing observed cumulative flows for AWBM, IHACRES, SAC, SIMHYD, and TANK were 17.68, −6.86, −10.18, −12.56, and −10.37%, respectively. One probable cause of these model biases indicating the deficiency in capturing the observed water balance closure could be the overall structural differences among the models in capturing the dynamics of the runoff generation in the study area. Another reason for the model biases would be the low quality of the meteorological model inputs. However, as already seen in Table 1, the quality of the data at each station was satisfactory for rainfall-runoff modeling. The performance of the rainfall from these stations, say, to explain the temporal variation in the runoff would be good if they are to be used in a case-specific way, for example, each station for a particular subbasin where it is located. However, in the implementation of Thiessen polygons for the computation of catchment-wide average rainfall, all the stations are considered so long as they fall within the polygons irrespective of whether they are located over the catchment or not. Although all the 14 rainfall stations from Table 1 were used to construct Thiessen polygons, it is noticeable from Figure 1 that stations 3, 5-6, and 10–14 are outside the BNB boundary. With respect to the physical reality, rainwater from these stations, 3, 5-6, and 10–14, cannot contribute directly to the overland runoff of the BNB. The consideration of data from stations located outside the catchment as done in this study (following the limitation of stations with data) lowers the accuracy of the temporal variation of the lumped catchment-wide rainfall to capture the variation in the observed overland flow; thus it becomes a possible reason for the mismatch between measured and modeled flows.

Figure 7 shows the statistical measures of agreement between observed and simulated flows. It is noticeable that all the metrics including E_f, I_d, and NSE for each model were above 0.5. Furthermore, considering each statistical metric, the performances of the models are evidently comparable. This shows that, without giving particular focus to either flooding or drought conditions, the influence of model selection on the simulation of the runoff of the study area may be minimal. Nonetheless, based on which model obtained the highest statistical "goodness-of-fit," the best performance with respect to E_f and I_d was realized from TANK, followed by IHACRES. The lowest values of E_f and I_d were obtained by AWBM and SIMHYD, respectively. However, with respect to NSE, the best performing model was AWBM, followed by TANK. The lowest NSE was obtained by SIMHYD. It can be seen that the value of E_f was lower than those of I_d and NSE for each model. For the AWBM, the NSE was greater than both E_f and I_d. On the contrary, the rest of the models
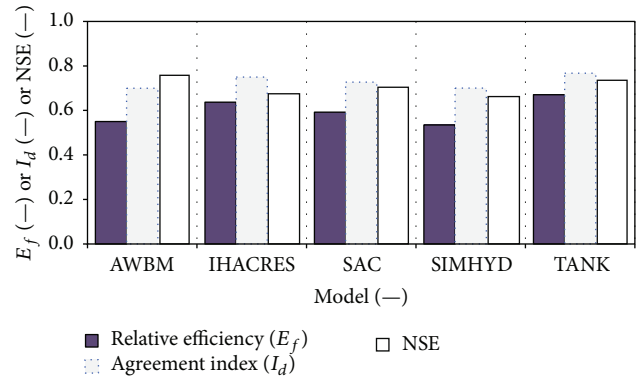

Figure 7: Performance of the models in terms of the forecast accuracy and NSE.

exhibited higher value of I_d than for the NSE. This shows that the judgment of the model performance depends on the selected statistical "goodness-of-fit" measure. It is therefore recommended that a number of statistical metrics be used in evaluating model performance in rainfall-runoff modeling.

4.2.2. "Goodness-of-Fit" of Flow Extremes. Figure 8 shows the capacity of the models to reproduce extreme events. The performances of the models are comparable for capturing high flow extremes (Figures 8(a), 8(c), and 8(e)), though SIMHYD tended to overestimate the high flows. Generally, the values of L_{FC} (Figures 8(b), 8(d), and 8(f)) were greater than those of H_{FC}. Except for the annual minima flow (Figure 8(f)), the values of L_{FC} from the various models are also fairly comparable (Figures 8(b) and 8(d)). For the annual minima flow, L_{FC} was zero for SAC because all the modeled minimum flow in each year was zero for the entire calibration and validation periods. To supplement the evaluation using H_{FC} and L_{FC}, results based on M_{AB} and R_{MSE} are presented in Table 4. The overall best model (i.e., with the lowest M_{AB}) to capture both high and low flow conditions was AWBM. It is noticeable that the magnitude of a particular statistical measure for evaluating the model performance tended to differ with the variation in the criteria for extracting the flow extremes from the full-time series. This shows that, for the intercomparison of the performance of the models with respect to extreme conditions, high and low flow events should be extracted based on a number of criteria. In this study, seven criteria were considered. The results of the use of six criteria are presented in Figure 8. The last (i.e., the seventh) criterion which entailed extracting events which are nearly independent and identically distributed was applicable to both high and low flows and is very vital for applications of extreme value analyses as seen in the construction of QDF relationships (see Sections 3.2.4 and 4.2.3).

Figure 9 shows the graphical model performance in terms of the Box-Cox transformed observed and modeled flows. For an ideal model (which is said to be unbiased), the scatter points in Figure 9 would all fall along the bisector (i.e., 45° line in the Cartesian plane). Consistent with the graphical observation made on Figure 5(d) and Figures 8(a), 8(c), and 8(e), it is noticeable from Figure 9(a) that the SIMHYD
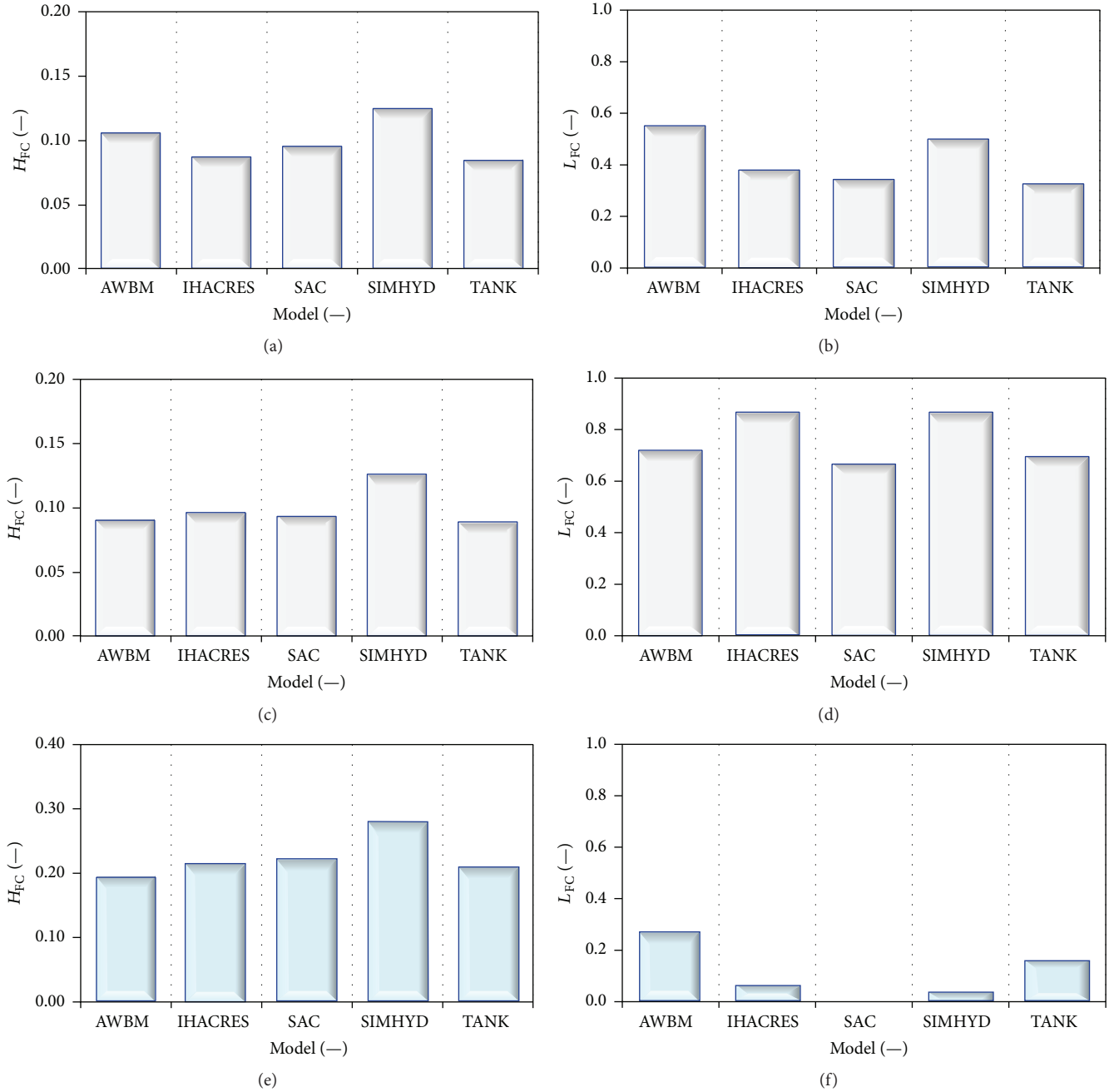
FIGURE 8: Model performance considering (a) flow events greater than one-third of the observed peak high flow mean, (b) flows less than one-third of the observed low flow mean, (c) flows greater than the 95th percentile of the observed flow, (d) flows less than the 5th percentile of the observed flow, (e) maximum flow in each year, and (f) minimum flow in each year.

overestimated the maximum flow in each year. As seen from Table 4, this overestimation by SIMHYD equaled M_{AB} of 62.57%. It is shown that the TANK model underestimated most of the annual maximum flow events. Figure 9(b) shows that the only realistic estimations of the minimum flow in each year were obtained by the AWBM. All the other models including IHACRES, SAC, SIMHYD, and TANK performed poorly in capturing the observed annual minimum flow. The worst performance for the annual minima flow was by SAC. This was for the reason that SAC produced zero flows during

every dry period in each year; thus, the scatter points of its annual minima flows appear as a horizontal line (Figure 9(b)) with the constant value of −4 after applying the Box-Cox transformation using (5) with the parameter λ set to 0.25 as mentioned before.

4.2.3. Analyses of Temporal Changes. Figure 10 shows the temporal anomalies in the observed and simulated flows. The details of the temporal evolution of the anomalies characterizing the variability in the annual maxima and
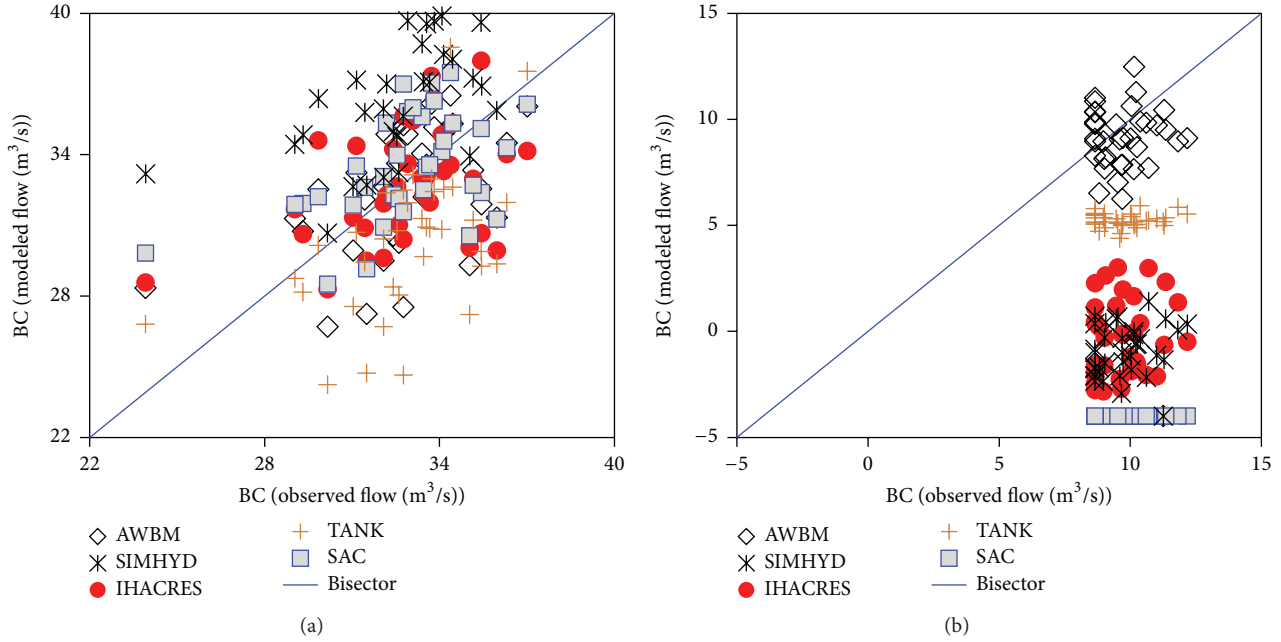
FIGURE 9: Model performance evaluation for (a) maximum flow in a year and (b) minimum flow in a year. For the labels of the axes, "BC" appearing before "()" shows that the flow (m³/s) values plotted were obtained after applying Box-Cox (BC) transformation using (5).

TABLE 4: Statistical measures of model performance for flow extremes.

Metric	AWBM	IHACRES	SAC	SIMHYD	TANK
		Maximum flow in each year			
R_{MSE} (m³/s)	1882.10	2022.53	1943.30	5371.12	2613.39
M_{AB} (%)	0.44	2.80	10.17	62.57	−21.81
		Minimum flow in each year			
R_{MSE} (m³/s)	75.35	149.41	161.23	150.55	124.06
M_{AB} (%)	−9.18	−98.63	−100.00	−99.54	−78.61
	Extreme events greater than one-third of the observed peak high flow mean				
R_{MSE} (m³/s)	2133.59	1972.62	1978.95	2705.44	1808.53
M_{AB} (%)	−10.89	5.49	−5.59	7.66	−5.81
	Extreme events less than one-third of the observed low flow mean				
R_{MSE} (m³/s)	300.82	413.04	780.03	504.00	311.20
M_{AB} (%)	−13.57	17.73	94.54	28.22	67.02
	Extreme events greater than the 95th percentile of the observed flow				
R_{MSE} (m³/s)	1638.19	2926.47	2237.11	3629.58	2326.94
M_{AB} (%)	−0.19	31.00	12.12	29.74	24.19
	Extreme events less than the 5th percentile of the observed flow				
R_{MSE} (m³/s)	208.70	251.36	535.34	306.96	108.75
M_{AB} (%)	−15.27	−27.78	96.99	37.68	57.13

annual minima are presented in Table 5. It is shown that the observed maximum flow in each year was generally

above the reference from 1970 to 1978 and again from 1991 to mid 1990s (Figure 10(a)). The annual maximum flow is shown to be below the reference from the late 1970s to early 1990s. This decrease in the flow was significant at the level of 5%. Visually, it is evident that all the models captured quite well the temporal variability in the observed annual maxima. The temporal variability of the annual maxima is comparable with that of the mean annual flows (Figure 10(b)). For the annual minima series, it is shown in Figure 10(c) and Table 5 that the flow events of 1970–1983 (1984–1996) were above (below) the reference. The anomalies in annual minima flow before and after 1984 were positive and negative, respectively. Nonetheless, the observed annual minima flow exhibited a steadily significant decrease from 1970 to mid 1990s. Statistically, the skill scores of the models in capturing the anomalies of variability in the annual maxima were 66.7% (AWBM), 63.0% (IHACRES), 63.0% (SAC), 70.4% (SIMHYD), and 66.7% (TANK). For annual mean flow, skill scores of 55.6% (AWBM), 59.1% (IHACRES), 55.6% (SAC), 62.9% (SIMHYD), and 70.4% (TANK) were obtained. Generally, the models performed poorly in capturing the variability in the annual minima flow with the skill scores of 18.5% (AWBM), 14.8% (IHACRES), 33.3% (SIMHYD), and 0% (TANK and SAC). According to [71], low flows are often poorly reproduced by most rainfall-runoff models which are tailored to capture flooding conditions. Besides, the overall water-balance-based objective function (as applied for calibration in this study) seemingly influences the capacity of the models to perform better in capturing the temporal changes in observed high flows compared to low flows. One option (which can still remain food for thought) would be for the model developers to revise their model structures to capture both low flows and high flows in an acceptably
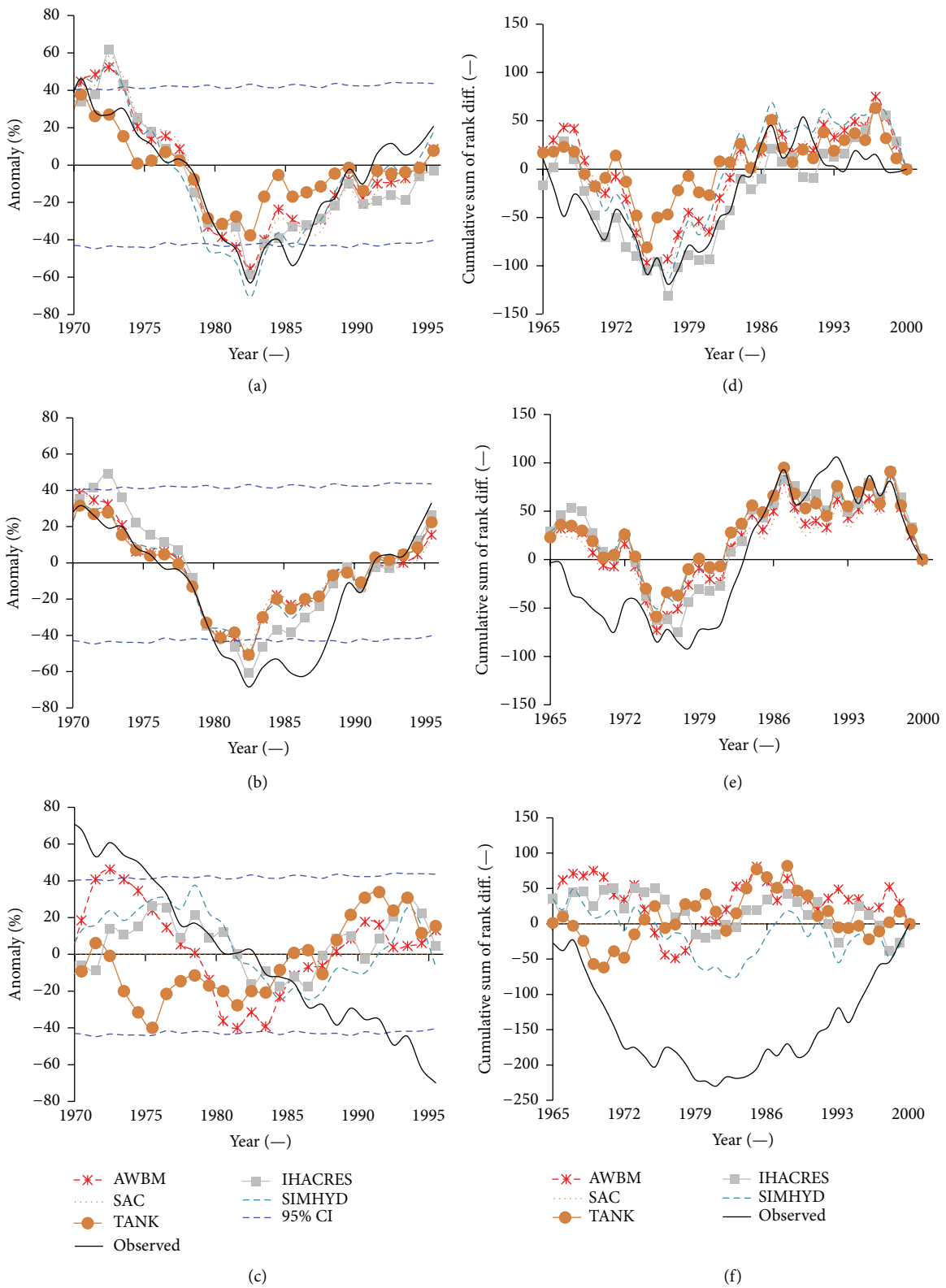
FIGURE 10: Temporal variation in terms of (a)–(c) NAIM results and (d)–(f) CRD results for (a, d) maximum flow in each year, (b, e) mean annual flow, and (c, f) minimum flow in each year. The charts (a)–(c) and (d)–(f) share the same legend at the bottom of (c) and (f), respectively.

TABLE 5: Temporal anomalies in the annual maxima and annual minima flows.

Year	Obs.	AWBM	IHACRES	SAC	SIMHYD	TANK	Obs.	AWBM	IHACRES	SAC	SIMHYD	TANK
	Maximum flow in each year						Minimum flow in each year					
1970	+	+	+	+	+	+	+#	−	−	*	−	−
1971	+#	+#	+	+	+#	+	+#	+	−	*	+	−
1972	+	+#	+	+#	+#	+	+#	+#	−	*	+	+
1973	+	+#	+#	+#	+#	+	+#	+#	+	*	+	−
1974	+	+#	+#	+#	+	+	+#	+	+	*	+	−
1975	+	+	+	+	+	+	+#	+	+	*	+	−
1976	+	+	+	+	+	+	+#	+	+	*	+	−
1977	+	+	+	+	+	+	+	+	+	*	+	−
1978	+	+	+	+	−	+	+	+	+	*	+	−
1979	−	−	−	−	−	−	+	+	+	*	+	−
1980	−	−	−	−	−#	−	+	−	+	*	+	−
1981	−	−	−	−	−#	−	+	−	+	*	+	−
1982	−#	−#	−	−	−#	−	+	−	−	*	−	−
1983	−#	−#	−#	−#	−#	−	+	−	−	*	−	−
1984	−#	−	−#	−#	−#	−	−	−	−	*	−	−
1985	−	−	−	−	−	−	−	−	−	*	−	−
1986	−#	−	−	−	−#	−	−	−	−	*	−	+
1987	−	−	−	−	−	−	−	−	−	*	−	+
1988	−	−	−	−	−	−	−	−	−	*	−	−
1989	−	−	−	−	−	−	−	+	+	*	−	+
1990	−	−	−	−	−	−	−	+	+	*	−	+
1991	−	−	−	−	−	−	−	+	−	*	−	+
1992	+	−	−	−	−	−	−	+	+	*	+	+
1993	+	−	−	−	−	−	−#	+	+	*	+	+
1994	+	−	−	−	−	−	−#	+	+	*	+	+
1995	+	−	−	−	+	−	−#	+	+	*	+	+
1996	+	+	−	+	+	+	−#	+	+	*	−	+

The symbols "+" and "−" denote anomaly greater and less than zero, respectively.
The symbol "∗" indicates anomaly of zero percent. "Obs." denotes observed change.
The cells with superscript symbol "#" are for anomalies significant at the level of 5%.
Each anomaly was placed at the center of the 10-year time slice, for example, 1965–1974, 1966–1975, . . ., 1991–2000, obtained based on the NAIM procedure as described in Section 3.2.3.

simultaneous way. Secondly, for optimization of the numerical performance schemes during calibration, the use of water-balance-based objective function should be combined with other criteria which seek to strike a balance in the model performance for both high flows and low flows jointly. Some of the criteria which can be combined together include the adequacy of the models to capture the overall shape of the hydrograph, peak high flows, low flows, and temporal changes (variability and subtrends) in high flows and low flows.

For short-durational changes in trend direction, the annual maxima and mean flow (Figures 10(d) and 10(e)) exhibited both negative and positive subtrends. This result is consistent with that of [23] that found decreasing and increasing short-durational trends over the periods of 1965–1983 and 1984–2000, respectively, in the Blue Nile flow at Khartoum. Like for the NAIM results, all the models

again exhibited close agreement in capturing the temporal cumulative effect of the variations or subtrends in the annual maxima and mean flows. The correlation coefficients between the CRD curves for observed and simulated annual maxima flows were 0.86 (AWBM), 0.85 (IHACRES), 0.86 (SAC), 0.93 (SIMHYD), and 0.82 (TANK). For the annual mean flows, correlation values of 0.87 (AWBM), 0.86 (IHACRES), 0.86 (SAC), 0.89 (SIMHYD), and 0.88 (TANK) were obtained. For the annual mean flow, there was more discrepancy between the observed and modeled flows with respect to decreasing than increasing subtrends. Nevertheless, some slight discrepancy between the temporal subtrends of observed and modeled flows from all the models was exhibited from 1991 to 1999. This discrepancy could be an indication of the slight change in catchment behavior with respect to the rainfall-runoff generation dynamics. The arguable evidence of this change in the catchment behavior was already presented in
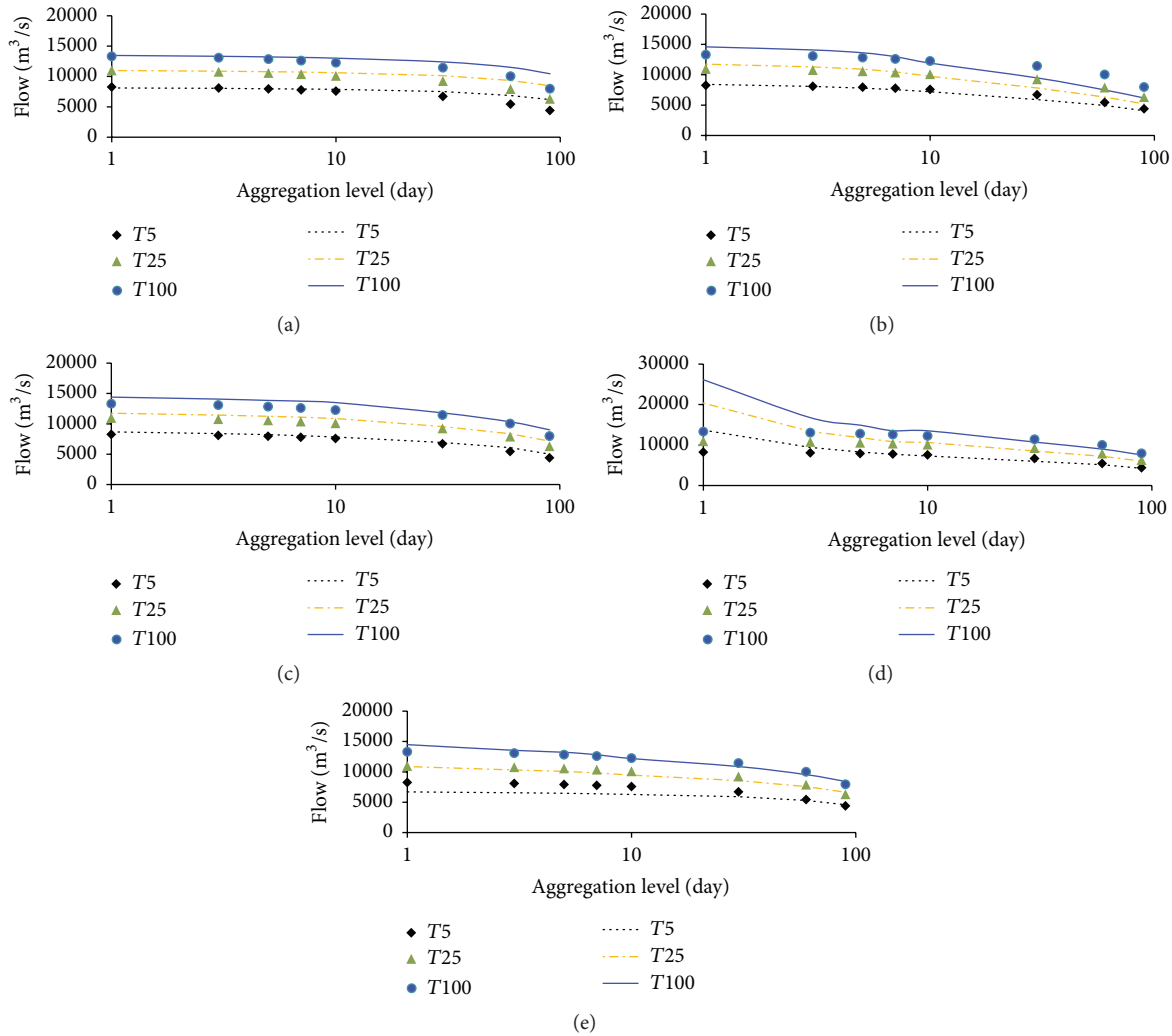
FIGURE 11: QDF relationships for observed peak high flows and modeled flows from (a) AWBM, (b) IHACRES, (c) SAC, (d) SIMHYD, and (e) TANK. Markers and lines are for observed and modeled flows, respectively. In the legend, for example, $T5$ denotes T-year curve for $T = 5$ years.

Section 4.2.1. Consistent with the NAIM result in Figure 10(c), it is noticeable from Figure 10(f) that the entire CRD curve fell below the reference, thereby indicating a dominantly decreasing trend in the observed annual minimum flows [23]. The models again performed poorly in capturing the subtrends in the annual minima flows. The correlation coefficients between the CRD curves for observed and simulated annual minima flows were 0.28 (AWBM), −0.06 (IHACRES), 0.00 (SAC), 0.49 (SIMHYD), and −0.41 (TANK). Based on the obtained results, the best and consistent model performance to reproduce the observed cumulative temporal variation for annual maxima and mean flows was SIMHYD.

4.2.4. Amplitude-Duration-Frequency Relationships. Figures 11 and 12 show the QDF relationships for extreme high flows and low flows, respectively, compiled from the quantiles estimated based on exponential distribution for return period T of 5, 25, and 100 years and aggregation levels of 1, 3, 5, 7, 10, 30, 60, and 90 days (for high flows) and 1, 10, 30, 90,

150, 180, 240, and 365 days (for low flows). The slope of each T-year curve on the QDF relationships is positive (for high flows) and negative (for low flows). It is shown that SIMHYD (Figure 11(d)) overestimated the high flow quantiles for aggregation level of one day. However, for higher aggregation levels, the biases in the estimation of quantiles reduced considerably. For aggregation levels higher than 10 days, IHACRES (Figure 11(b)) underestimated the quantiles for $T =$ 100 years. This might be due to the inadequacy of the model to reproduce higher quantiles in the tail of the distribution of extreme event, thereby leading to uncertainty in the EVD calibration for extrapolation. For AWBM (Figure 11(a)), SAC (Figure 11(c)), and TANK (Figure 11(e)), the quantiles from the observed and modeled peak high flows are noticeably comparable for all the aggregation levels. For low flows, poor performance of all the selected models (Figures 12(b)–12(e)) except AWBM (Figure 12(a)) is evident.

The model biases in reproducing the daily aggregation observed high flow quantiles for T between 1 and
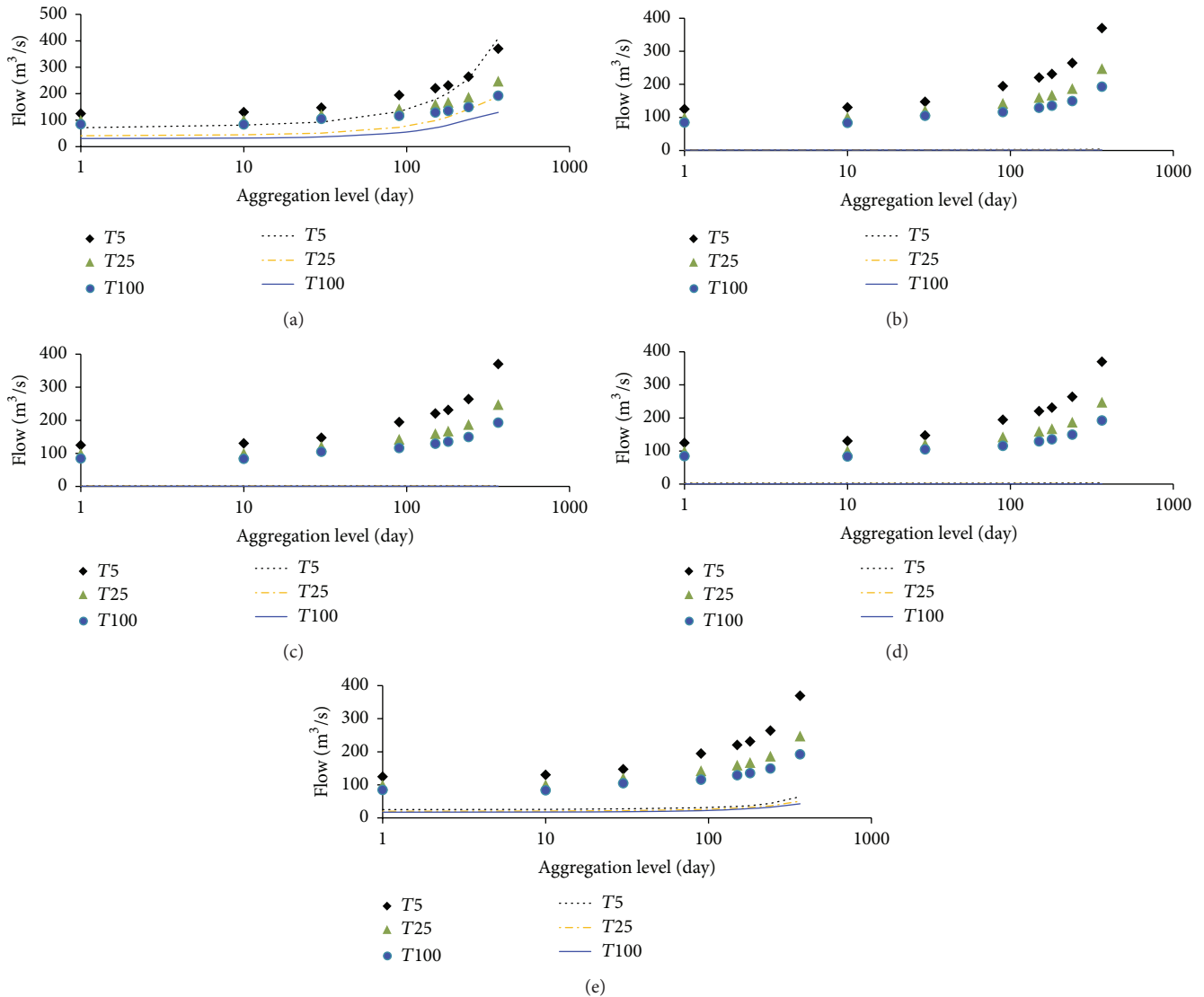
FIGURE 12: QDF relationships for observed extreme low flows and modeled flows from (a) AWBM, (b) IHACRES, (c) SAC, (d) SIMHYD, and (e) TANK. Markers and lines are for observed and modeled flows, respectively. In the legend, for example, $T25$ denotes T-year curve for $T = 25$ years.

40 years were −0.55% (AWBM), 5.30% (IHACRES), 6.35% (SAC), 79.56% (SIMHYD), and −6.28% (TANK). For low flows, the corresponding biases of −53.51% (AWBM), −99.98% (IHACRES), −100% (SAC), −99.99% (SIMHYD), and −79.46% (TANK) were obtained. One reason for the biases in the models could be due to the observation errors in the flows. For instance, in the flooding conditions when the flow exceeds the river banks, observation errors become large as a result of the bias from the rating curve extrapolation. On the other hand, the difference in the biases indicates the influence of model selection on the simulation of extreme events. If the results from models are to be used to support decision for risk-based management, the model with the least bias would be selected.

5. Conclusions

In this study, the influence of hydrological model selection on the rainfall-runoff simulation was demonstrated using five well-known models including AWBM, IHACRES, Sacramento (SAC), SIMHYD, and TANK applied based on hydrometeorological data from the BNB. Optimal parameters of all the selected models were obtained through automatic calibration using Nash-Sutcliffe Efficiency as the objective function. Each model was tested by evaluating its capacity in capturing moderate and extreme hydrological events and the temporal changes in the flows. The temporal changes in the flows were assessed in terms of variability and subtrends (short-durational changes in trends within the series). Variability was computed using the nonparametric anomaly

indicator method and sub-trends were analyzed based on the cumulative rank difference technique. The performance of the models were evaluated using nine statistical "goodness-of-fit" measures including Nash-Sutcliffe Efficiency, relative deviation efficiency, low flow forecast accuracy, index of agreement, high flow forecast accuracy, model bias, root mean squared error, correlation, and skill scores. The model performance was evaluated based on high flows, low flows, and the overall shape of the hydrograph. To evaluate the performance of the models in reproducing quantiles as a simultaneous function of different return periods and aggregation levels, Amplitude-Duration-Frequency (ADF) relationships constructed using observed and modeled hydrological extremes were compared.

Without giving particular focus to either flooding or drought conditions, the performance of all the models in simulating the runoff of the study area was found acceptable. With respect to considering the overall shape of the hydrograph, the best performance in terms of relative efficiency and index of agreement was realized from TANK, followed by IHACRES. However, using the Nash-Sutcliffe Efficiency, AWBM was the best performing model. The deficiency of the models in capturing the observed water balance closure evaluated using bias ranged from -6.86% (IHACRES) to 17.68% (AWBM). Thus, the best model to reproduce observed cumulative flow was IHACRES. On an important note, the choice of the "goodness-of-fit" measures in performance evaluation was found to influence the selection of a particular model. Therefore, caution should be taken to use a number of measures in intercomparison of hydrological models.

The skill scores in reproducing temporal variability in observed annual maximum flows varied from 63.0% (IHACRES) to 70.4% (SIMHYD). Correspondingly, for the minimum flows in each year, the best and worst performances were with skill score of 14.8% (IHACRES) and 0% (SAC), respectively. For the annual mean flow, the lowest and the highest skill scores were 55.6% (AWBM and SAC) and 70.4% (TANK), respectively. Whereas the changes in observed annual minimum flows were dominated by a negative trend, the simulated flows from all the models exhibited no trend. Eventually, the correlation between the temporal subtrends from observed and simulated annual minima flows ranged from 0.0 (SAC) to 0.49 (SIMHYD). The observed annual maxima as well as annual mean flow exhibited both negative and positive subtrends over the periods of 1965–1984 and 1985–2000, respectively. These observed positive and negative subtrends were adequately reproduced by the simulated runoffs from all the models. Coefficients of the correlation between observed and modeled annual mean flows were above 0.85 for all the models. For maximum flows in each year, correlation varied from 0.82 (TANK) to 0.93 (SIMHYD). Therefore, the best performing model to reproduce the observed temporal variability at decadal time scale was SIMHYD (for annual maxima and minima flow) and TANK (for annual mean flow). For short-durational changes in trend directions, the best and consistent model performance to reproduce the observed cumulative temporal variation for annual maxima and mean and minima flows was SIMHYD.

For the ADF relationships, the model bias in reproducing the daily aggregation observed high flow quantiles for return periods between 1 and 40 years ranged from -0.55% (AWBM) to 79.56% (SIMHYD). Correspondingly, the biases for low flows were from -53.51% (AWBM) to -100% (SAC). Generally, for the simulation of flooding and drought conditions, it was found that the use of the overall water-balance-based objective function influences the capacity of the models to perform better in capturing the changes as well as quantiles of observed high flows compared to low flows. This might not be surprising because low flows are often poorly reproduced by most rainfall-runoff models which are tailored to capture flooding conditions [71]. Whereas the need to revise concepts on model structures to simultaneously capture both low flows and high flows acceptably may still remain food for thought, it is recommended that the use of water-balance-based objective function be combined with other criteria for optimization of the numerical performance schemes. Such other criteria may include the overall shape of the hydrograph, peak high flows, low flows, variability, and subtrends in low and high flows. Furthermore, it was generally found that the choice of the criteria for extracting extreme events from the full series for the purpose of intermodel comparison influences the model performance. Thus, caution must be taken to test model performance with respect to extreme events extracted based on a number of criteria. In this study, a total of seven criteria were considered including the extraction of the flow extremes based on (1) events greater than one-third of the observed peak high flow mean, (2) events less than one-third of the observed low flow mean, (3) high flow events greater than the 95th percentile of the observed flow, (4) low flows less than the 5th percentile of the observed flow, (5) maximum flow in each year, (6) minimum flow in each year, and (7) events which are nearly independent and identically distributed (e.g., those used for construction of the ADF relationships). Considering all the above criteria in a combined way, the overall best model (i.e., with the best "goodness-of-fits") to satisfactorily capture both high and low flow conditions jointly was the AWBM. Because the performance of a particular model may differ from one criterion to another, an attempt to interrelate the performance indicators across all various criteria should be augmented by the expert judgment of the modeler.

It can be remarked that prudence must be exercised in the choice of a particular model which can be made on a case by case basis in line with the objectives of the hydrological modeling study. Evaluation of the intermodel differences is vital in the choice of which hydrological models to apply for impact investigations, for example, of climate variability and change on water resources. Importantly, whereas the influence of model selection may be minimal in the simulation of normal flow events, it must be considered carefully for high flows or low flows to minimize under- and/or overestimation of hydrological extremes. The selection of a particular model for simulating extreme flow events is, in turn, influenced by the choice of (1) the "goodness-of-fit" measures for evaluating the model performance and (2) the criteria for extracting extreme events for the model performance evaluation or intended application. Furthermore, note should be taken that
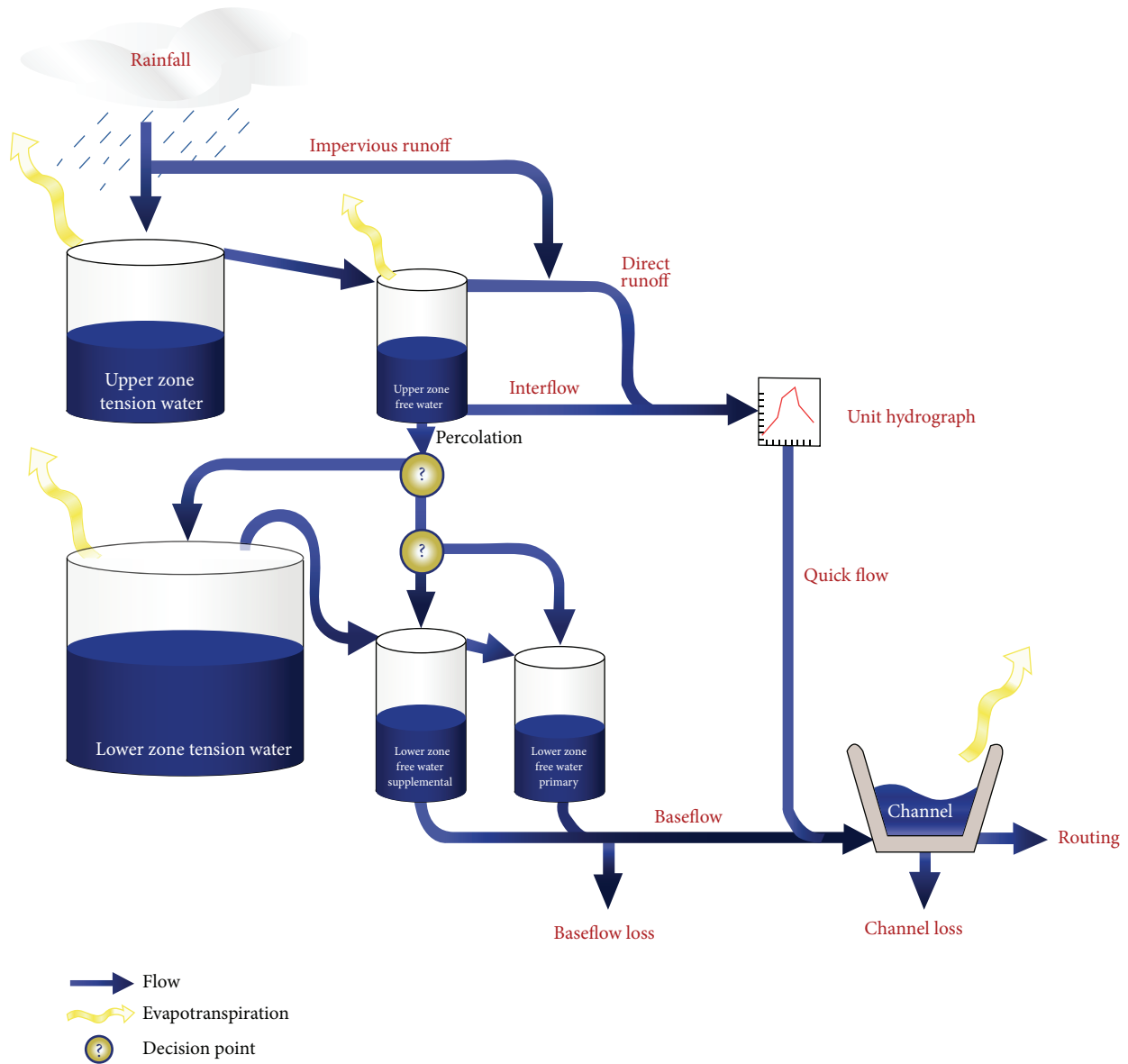
FIGURE 13: The SAC model structure (Source: RRL [39]).

the performance of a hydrological model may be influenced by (1) the quality of the data which can be translated into the ease or difficulty of the model calibration, (2) the method of model calibration, (3) incompatibility of the given model structure with (a) the data availability, (b) the uniqueness of catchments with respect to their sizes, topography, soils, vegetation, rocks, hydrological conditions, and anthropogenic modifications, (c) spatiotemporal scales of the model inputs, and so forth.

Finally, whereas the model selection has been demonstrated to influence the adequacy of the simulated runoff, the author does not intend that the overall message from this paper be taken in a negative way. The results from this study are rather to underscore acumen about how hydrological models can be selected to support decisions related to applications regarding risk-based water-related analyses,

agricultural practices, environmental planning and management, and so forth.

For readers interested in implementing the CRD trend test and/or performing variability analyses using the NAIM as done in this study, the author has made freely available to the public an Excel-based VBA-coded tool named "CRD-NAIM" which can be downloaded online from https://sites.google.com/site/conyutha/tools-to-download.

Appendix

A. Structures of Selected Rainfall-Runoff Models

A.1. The Sacramento (SAC) Model Structure (Adopted from the RRL [39]). In the SAC model, the soil moisture storage can
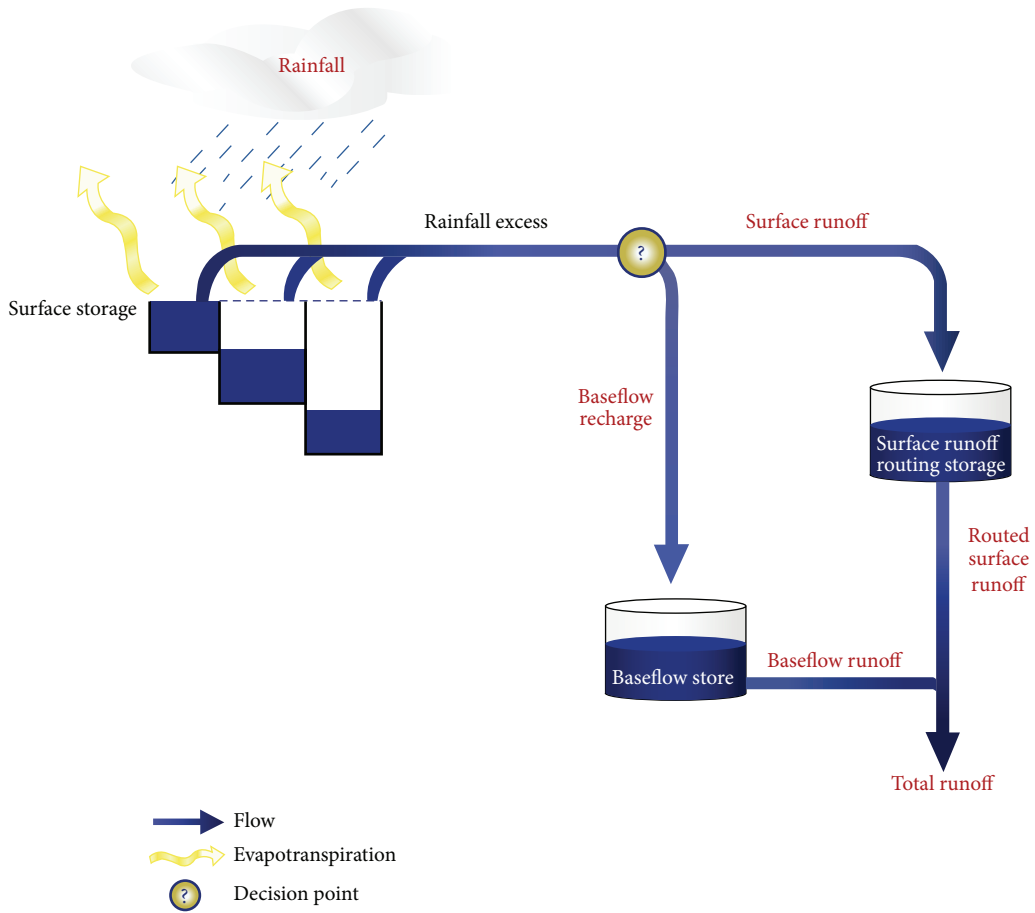
FIGURE 14: The AWBM structure (Source: RRL [39]).
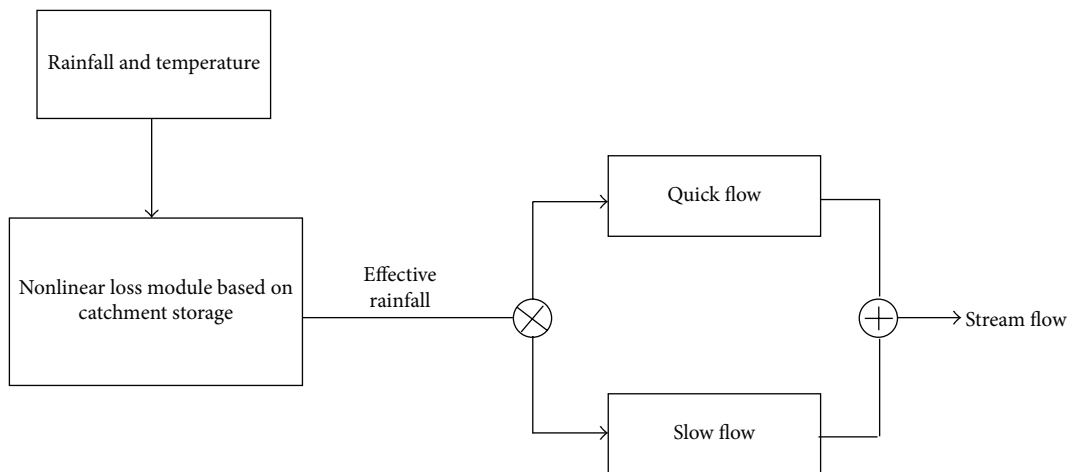
FIGURE 15: The IHACRES model structure.

be reduced/increased by evaporation/rainfall. The depth at which the absorbed rainfall fraction will infiltrate depends on the size and relative wetness of the storages. Runoff is generated by the fraction of the rainfall in excess of that absorbed. The runoff is transformed into surface flow through unit hydrograph. Lateral flows which also occur from the soil moisture stores are combined with the surface flow to yield the total stream flow. The rainfall-runoff generation processes

FIGURE 16: The SIMHYD model structure (Source: RRL [39]).

in the SAC model are controlled by mainly 16 parameters of which 5 define the size of soil moisture stores, 3 calculate the rate of lateral outflows, 3 estimate the percolation water from the upper to the lower soil moisture stores, 2 compute direct runoff, and 3 determine losses in the system (see Figure 13).

A.2. The AWBM Structure (Adopted from the RRL [39]). The AWBM for the catchment water balance takes rainfall to three surface water stores based on their moisture contents. Each surface water store is considered independently of the others. Evaporation is subtracted from each of the stores.

The moisture in excess of the storage capacity becomes either the surface runoff or recharge into the groundwater. The baseflow and the surface runoff are routed separately and later combined into the total flow at the outlet of the catchment. In total, there are generally 8 parameters which control the rainfall-runoff generation by the AWBM (see Figure 14).

A.3. The IHACRES Model Structure. The IHACRES model is characterized by the catchment storage from which rainfall is lost through evaporation, and the remaining fraction contributes to the effective rainfall. Quick flow and slow flow are
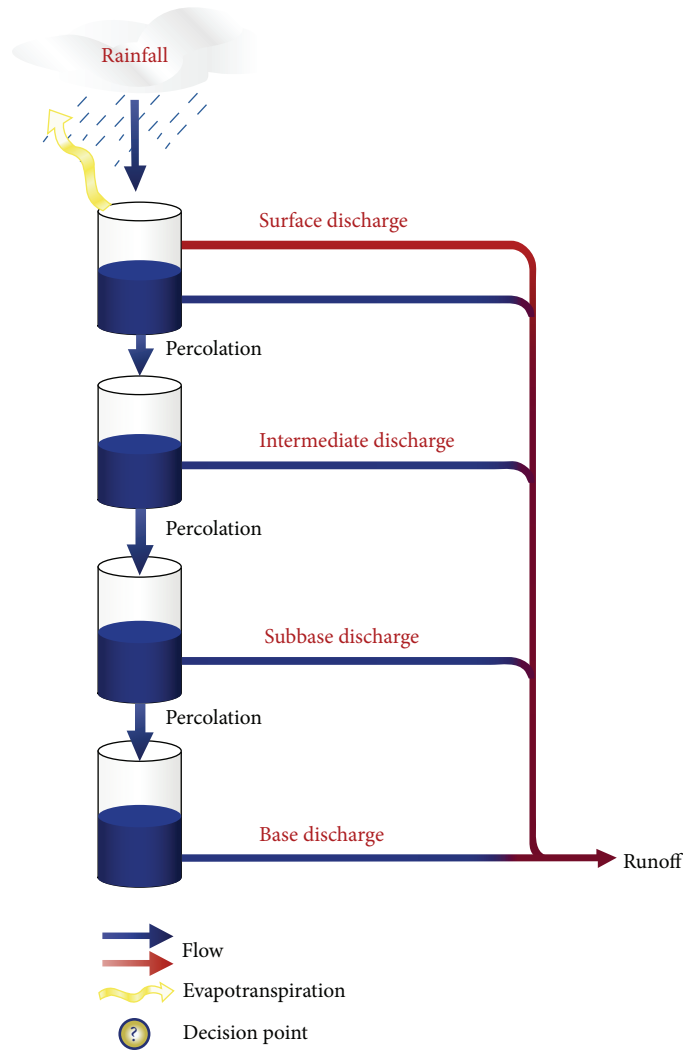
Figure 17: The TANK model structure (Source: RRL [39]).

generated from the effective rainfall. Using unit hydrograph concept, the catchment is taken as linear reservoirs in series and/or parallel. This concept is premised on the assumed linear relationship between effective rainfall and stream flow. The total flow at the catchment outlet comprises the baseflow and quick flow in a combined way. The rainfall-runoff generation processes in the TANK model are controlled by 11 parameters (see Figure 15).

A.4. The SIMHYD Structure (Adopted from the RRL [39]). In the SIMHYD, rainfall is lost by evaporation and infiltration. The fraction that infiltrates is transformed into interflow, groundwater store, and soil moisture store. Interflow and baseflow are each assumed to follow the soil wetness in a linear way. The infiltration excess becomes the surface runoff. The surface runoff and interflow form the quick flow which can be combined with the baseflow to yield the stream flow. The SIMHYD model has 9 model parameters for calibration (see Figure 16).

A.5. The TANK Model Structure (Adopted from the RRL [14]). The TANK model has four tanks arranged vertically in series. The rainfall is fed into the top tank. The rainfall is lost by evaporation from the top tank downwards in a sequential way. The outlets at the sides of the tanks generate different components of the total runoff, that is, surface runoff, intermediate runoff, subbase runoff, and baseflow from the first through to the fourth tank, respectively. The TANK model has 18 parameters for calibration (see Figure 17).

B. Parameters of Selected Rainfall-Runoff Models

B.1. List of Model Parameters after Calibration of SAC, IHACRES, and AWBM. See Table 6.

B.2. List of Model Parameters after Calibration of TANK and SIMHYD. See Table 7.

TABLE 6

S. number	Parameter	Unit	Value
	SACRAMENTO		
1	Additional fraction of pervious area (Adimp)	(—)	0.0023
2	Lower Zone Free Water Primary Maximum (Lzfpm)	(mm)	12.987
3	Lower Zone Free Water Supplemental Maximum (Lzfsm)	(m)	24.998
4	Ratio of water in LZFPM (Lzpk)	(mm)	0.061
5	Ratio of water in LZFSM (Lzsk)	(mm)	0.035
6	Lower Zone Tension Water Maximum (Lztwm)	(mm)	241.28
7	Impervious fraction of the basin (Pctim)	(—)	0.003
8	Minimum proportion of percolation (P_{free})	(—)	0.658
9	Exponential percolation rate (R_{exp})	(—)	0.661
10	Fraction of water unavailable for transpiration (Rserv)	(—)	0.310
11	Catchment portion that loses water by evaporation (Sarva)	(—)	0.010
12	Fraction of base flow which is groundwater flow (Side)	(—)	0.330
13	Flow volume through porous material (Ssout)	$m^3/s/km^2$	0.001
14	Upper Zone Free Water Maximum (Uzfwm)	1/day	79.728
15	Ratio of water in UZFWM (Uzk)	1/day	0.0103
16	Upper Zone Tension Water Maximum (Uztwm)	1/day	99.659
17	Factor applied to PBASE (Zperc)	(—)	11.479
	IHACRES		
1	Delay	(day)	28.00
2	Recession rate 1 ($\alpha^{(s)}$)	(1/day)	−0.888
3	Peak response 1 ($\beta^{(s)}$)	(—)	0.112
4	Time constant 1 ($\tau^{(s)}$)	(day)	8.424
5	Volume proportion 1 ($\nu^{(s)}$)	(°C)	1.000
6	Mass balance term (c)	(—)	0.005
7	Drying rate at reference temperature (t_w)	(°C/day)	2.000
8	Temperature dependence of drying rate (f)	(—)	0.000
9	Reference temperature (t_{ref})	(°C)	20.00
10	Moisture threshold for producing flow (l)	(mm)	0.000
11	Power on soil moisture (p)	(—)	1.000
	AWBM		
1	Fraction of catchment area for the first store ($A1$)	(—)	0.634
2	Fraction of catchment area for the second store ($A2$)	(—)	0.366
3	Base flow index (BFI)	(—)	0.629
4	Storage capacity of first store ($C1$)	(mm)	24.961
5	Storage capacity of second store ($C2$)	(mm)	140.793
6	Storage capacity of third store ($C3$)	(mm)	301.082
7	Base flow recession constant (K_{base})	(day)	0.999
8	Surface flow recession constant (K_{surf})	(day)	0.972

TABLE 7

S. number	Parameter	Unit	Value
	TANK		
1	Depth below the top outlet of the first tank ($H11$)	(mm)	407.26
2	Overland runoff from the top outlet of first tank ($a11$)	(m^3/s)	0.0584
3	Overland runoff from the lower outlet of first tank ($a12$)	(m^3/s)	0.0029
4	Intermediate runoff ($a21$)	(m^3/s)	0.1308
5	Subbase runoff ($a31$)	(m^3/s)	0.9775
6	Base flow ($a41$)	(m^3/s)	0.9517
7	Alpha	(—)	0.5339
8	Outflow from the bottom of the first tank ($b1$)	(m^3/s)	0.0031
9	Outflow from the bottom of the second tank ($b2$)	(m^3/s)	0.9276
10	Outflow from the bottom of the third tank ($b3$)	(m^3/s)	0.0106
11	Water depth in the first tank ($C1$)	(mm)	90.607
12	Water depth in the second tank ($C2$)	(mm)	78.564
13	Water depth in the third tank ($C3$)	(mm)	48.240
14	Water depth in the fourth tank ($C4$)	(mm)	6.489
15	Depth below the lower outlet of the first tank ($H12$)	(mm)	146.888
16	Depth below the outlet of the second tank ($H21$)	(mm)	73.086
17	Depth below the outlet of the third tank ($H31$)	(mm)	2.4874
18	Depth below the outlet of the fourth tank ($H41$)	(mm)	5.8069
	SIMHYD		
1	Baseflow coefficient	(—)	0.0107
2	Impervious threshold	(—)	1.5412
3	Infiltration coefficient	(—)	308.16
4	Infiltration shape	(—)	2.8373
5	Interflow coefficient	(—)	0.1534
6	Pervious fraction	(—)	0.9922
7	Rainfall interception store capacity	(mm)	4.9412
8	Recharge coefficient	(—)	0.2806
9	Soil moisture store capacity	(mm)	440.91

Competing Interests

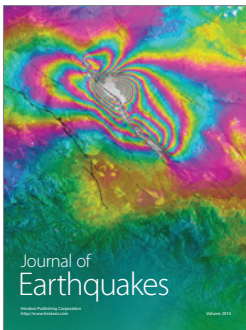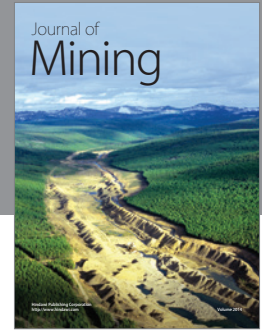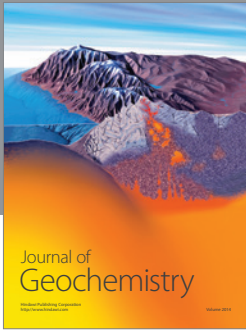The author declares that there are no competing interests regarding the publication of this paper.

Acknowledgments

.noaa.gov/oa/climate/ghcn-daily/ (accessed: 11th June, 2014) of the Global Historical Climatology Network (GHCN) data [28, 29] and http://cfs.ncep.noaa.gov/cfsr/ (accessed: 3rd February, 2016) of the Climate Forecast System Reanalysis (CFSR).

References

[1] IPCC, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the IPCC*, Cambridge University Press, New York, NY, USA, 2013.

[2] M. Hulme and A. Trilsbach, "The August 1988 storm over Khartoum: its climatology and impact," *Weather*, vol. 44, no. 2, pp. 82–90, 1989.

[3] J. V. Sutcliffe, G. Dugdale, and J. R. Milford, "The Sudan floods of 1988," *Hydrological Sciences Journal*, vol. 34, no. 3, pp. 355–364, 1989.

[4] M. Hulme, "Rainfall changes in Africa: 1931–1960 to 1961–1990," *International Journal of Climatology*, vol. 12, no. 7, pp. 685–699, 1992.

[5] DHI, "Reference manual, DHI water & environment," in *MIKE11—a Modeling System for Rivers and Channels*, pp. 278–325, DHI Water & Environment, Hørsholm, Denmark, 2007.

[6] H. Madsen, "Automatic calibration of a conceptual rainfall-runoff model using multiple objectives," *Journal of Hydrology*, vol. 235, no. 3-4, pp. 276–288, 2000.

[7] S. A. Nielsen and E. Hansen, "Numerical simulation of the rainfall-runoff process on a daily basis," *Nordic Hydrology*, vol. 4, no. 3, pp. 171–190, 1973.

[8] S. Bergström, *Development and Application of a Conceptual Runoff Model for Scandinavian Catchments*, SMHI RHO 7, SMHI, Norrköping, Sweden, 1976.

[9] R. J. Moore, "The PDM rainfall-runoff model," *Hydrology and Earth System Sciences*, vol. 11, no. 1, pp. 483–499, 2007.

[10] S. L. Markstrom, R. S. Regan, L. E. Hay et al., *PRMS-IV, the Precipitation-Runoff Modeling System, Version 4*, Techniques and Methods, book 6, chapter B7, U.S. Geological Survey Techniques and Methods, 2015.

[11] J. W. Porter and T. A. McMahon, "A model for the simulation of streamflow data from climatic records," *Journal of Hydrology*, vol. 13, pp. 297–324, 1971.

[12] M. Sugawara, "Tank model," in *Computer Models of Watershed Hydrology*, V. P. Singh, Ed., pp. 165–214, Water Resources Publications, Littleton, Colo, USA, 1995.

[13] R. J. C. Burnash, "The NWS River forecast system-catchment modeling," in *Computer Models of Watershed Hydrology*, V. P. Singh, Ed., pp. 311–366, Water Resources Publications, Littleton, Colo, USA, 1995.

[14] W. Boughton, "The Australian water balance model," *Environmental Modelling & Software*, vol. 19, no. 10, pp. 943–956, 2004.

[15] A. J. Jakeman and G. M. Hornberger, "How much complexity is warranted in a rainfall-runoff model?" *Water Resources Research*, vol. 29, no. 8, pp. 2637–2649, 1993.

[16] A. J. Jakeman, I. G. Littlewood, and P. G. Whitehead, "Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments," *Journal of Hydrology*, vol. 117, no. 1–4, pp. 275–300, 1990.

[17] I. G. Littlewood and A. J. Jakeman, "A new method of rainfall-runoff modelling and its applications in catchment hydrology," in *Environmental Modelling*, P. Zannetti, Ed., vol. 2, pp. 143–171, Computational Mechanics, Southampton, UK, 1994.

[18] G. D. Betrie, Y. A. Mohamed, A. Van Griensven, and R. Srinivasan, "Sediment management modelling in the Blue Nile Basin using SWAT model," *Hydrology and Earth System Sciences*, vol. 15, no. 3, pp. 807–818, 2011.

[19] M. M. Bitew and M. Gebremichael, "Are satellite-gauge rainfall products better than satellite-only products for nile hydrology?" in *Nile River Basin*, A. M. Melesse, Ed., chapter 6, pp. 129–141, Springer, Amsterdam, Netherlands, 2011.

[20] Z. M. Easton, D. R. Fuka, E. D. White et al., "A multi basin SWAT model analysis of runoff and sedimentation in the Blue Nile, Ethiopia," *Hydrology and Earth System Sciences*, vol. 14, no. 10, pp. 1827–1841, 2010.

[21] S. G. Setegn, R. Srinivasan, and B. Dargahi, "Hydrological modelling in the Lake Tana Basin, Ethiopia using SWAT model," *Open Hydrology Journal*, vol. 2, no. 1, pp. 49–62, 2008.

[22] S. G. Gebrehiwot, J. Seibert, A. I. Gärdenäs, P.-E. Mellander, and K. Bishop, "Hydrological change detection using modeling: half a century of runoff from four rivers in the Blue Nile Basin," *Water Resources Research*, vol. 49, no. 6, pp. 3842–3851, 2013.

[23] C. Onyutha and P. Willems, "Identification of the main attribute of river flow temporal variations in the Nile Basin," *Hydrology and Earth System Sciences Discussions*, vol. 12, no. 11, pp. 12167–12214, 2015.

[24] G. Y. Ebrahim, A. Jonoski, A. van Griensven, and G. Di Baldassarre, "Downscaling technique uncertainty in assessing hydrological impact of climate change in the Upper Beles River Basin, Ethiopia," *Hydrology Research*, vol. 44, no. 2, pp. 377–398, 2013.

[25] D. Legesse, C. Vallet-Coulomb, and F. Gasse, "Hydrological response of a catchment to climate and land use changes in Tropical Africa: case study south central Ethiopia," *Journal of Hydrology*, vol. 275, no. 1-2, pp. 67–85, 2003.

[26] D. Legesse, C. Vallet-Coulomb, and F. Gasse, "Analysis of the hydrological response of a tropical terminal lake, Lake Abiyata (main Ethiopian rift valley) to changes in climate and human activities," *Hydrological Processes*, vol. 18, no. 3, pp. 487–504, 2004.

[27] M. T. Taye, P. Willems, and P. Block, "Implications of climate change on hydrological extremes in the Blue Nile basin: a review," *Journal of Hydrology: Regional Studies*, vol. 4, pp. 280–293, 2015.

[28] J. M. Menne, I. Durre, B. Korzeniewski et al., *Global Historical Climatology Network-Daily (GHCN-Daily), Version 3*, NOAA National Climatic Data Center, Asheville, NC, USA, 2012.

[29] M. J. Menne, I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston, "An overview of the global historical climatology network-daily database," *Journal of Atmospheric and Oceanic Technology*, vol. 29, no. 7, pp. 897–910, 2012.

[30] C. Onyutha and P. Willems, "Spatial and temporal variability of rainfall in the Nile Basin," *Hydrology and Earth System Sciences*, vol. 19, no. 5, pp. 2227–2246, 2015.

[31] E. Abushandi and B. Merkel, "Modelling rainfall runoff relations using HEC-HMS and IHACRES for a single rain event in an arid region of Jordan," *Water Resources Management*, vol. 27, no. 7, pp. 2391–2409, 2013.

[32] E. C. Carcano, P. Bartolini, M. Muselli, and L. Piroddi, "Jordan recurrent neural network versus IHACRES in modelling daily streamflows," *Journal of Hydrology*, vol. 362, no. 3-4, pp. 291–307, 2008.

[33] A. Hope, J. Decker, and P. Jankowski, "Utility of gridded rainfall for IHACRES daily river flow predictions in southern California watersheds," *Journal of the American Water Resources Association*, vol. 44, no. 2, pp. 428–435, 2008.

[34] R. N. Jones, F. H. S. Chiew, W. C. Boughton, and L. Zhang, "Estimating the sensitivity of mean annual runoff to climate change using selected hydrological models," *Advances in Water Resources*, vol. 29, no. 10, pp. 1419–1429, 2006.

[35] J. Luo, E. Wang, S. Shen, H. Zheng, and Y. Zhang, "Effects of conditional parameterization on performance of rainfall-runoff model regarding hydrologic non-stationarity," *Hydrological Processes*, vol. 26, no. 26, pp. 3953–3961, 2012.

[36] B. Yu and Z. Zhu, "A comparative assessment of AWBM and SimHyd for forested watersheds," *Hydrological Sciences Journal*, vol. 60, no. 7-8, pp. 1200–1212, 2015.

[37] Y. Zhang and F. H. S. Chiew, "Relative merits of different methods for runoff predictions in ungauged catchments," *Water Resources Research*, vol. 45, no. 7, Article ID W07412, 2009.

[38] B. F. W. Croke, F. Andrews, J. Spate, and S. M. Cuddy, *IHACRES User Guide*, Technical Report 2005/19, iCAM, School of Resources, Environment and Society; The Australian National University, Canberra, Australia, 2nd edition, 2005.

[39] G. Podger, *User Guide—Rainfall Runoff Library*, Catchment Modelling Toolkit, CRC for Catchment Hydrology, 2004.

[40] R. G. Allen, L. S. Pereira, D. Raes, and M. Smith, "Crop evapotranspiration-guidelines for computing crop water requirements," FAO Irrigation and Drainage Paper 56, FAO-Food and Agriculture Organization of the United Nations, Rome, Italy, 1998.

[41] J. H. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, Mich, USA, 1975.

[42] Q. J. Wang, "The genetic algorithm and its application to calibrating conceptual rainfall-runoff models," *Water Resources Research*, vol. 27, no. 9, pp. 2467–2471, 1991.

[43] Q. Duan, "Optimal use of the SCE-UA global optimization method for calibrating watershed models," *Journal of Hydrology*, vol. 158, no. 3-4, pp. 265–284, 1994.

[44] Q. Duan, S. Sorooshian, and V. Gupta, "Effective and efficient global optimization for conceptual rainfall-runoff models," *Water Resources Research*, vol. 28, no. 4, pp. 1015–1031, 1992.

[45] Q. Y. Duan, V. K. Gupta, and S. Sorooshian, "Shuffled complex evolution approach for effective and efficient global minimization," *Journal of Optimization Theory and Applications*, vol. 76, no. 3, pp. 501–521, 1993.

[46] Q. Duan, S. Sorooshian, and V. K. Gupta, "Optimal use of the SCE-UA global optimization method for calibrating watershed models," *Journal of Hydrology*, vol. 158, no. 3-4, pp. 265–284, 1994.

[47] T. Y. Gan and G. F. Biftu, "Automatic calibration of conceptual rainfall-runoff models: optimization algorithms, catchment conditions, and model structure," *Water Resources Research*, vol. 32, no. 12, pp. 3513–3524, 1996.

[48] M. Thyer, G. Kuczera, and B. C. Bates, "Probabilistic optimization for conceptual rainfall-runoff models: a comparison of the shuffled complex evolution and simulated annealing algorithms," *Water Resources Research*, vol. 35, no. 3, pp. 767–773, 1999.

[49] V. A. Cooper, V. T. V. Nguyen, and J. A. Nicell, "Evaluation of global optimization methods for conceptual rainfall-runoff model calibration," *Water Science and Technology*, vol. 36, no. 5, pp. 53–60, 1997.

[50] T. Y. Gan, E. M. Dlamini, and G. F. Biftu, "Effects of model complexity and structure, data quality, and objective functions on hydrologic modeling," *Journal of Hydrology*, vol. 192, no. 1–4, pp. 81–103, 1997.

[51] S. Sorooshian, Q. Duan, and V. K. Gupta, "Calibration of rainfall-runoff models: application of global optimization to the Sacramento soil moisture accounting model," *Water Resources Research*, vol. 29, no. 4, pp. 1185–1194, 1993.

[52] P. O. Yapo, H. V. Gupta, and S. Sorooshian, "Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data," *Journal of Hydrology*, vol. 181, no. 1–4, pp. 23–48, 1996.

[53] J. E. Nash and J. V. Sutcliffe, "River flow forecasting through conceptual models part I—A discussion of principles," *Journal of Hydrology*, vol. 10, no. 3, pp. 282–290, 1970.

[54] P. Krause, D. P. Boyle, and F. Bäse, "Comparison of different efficiency criteria for hydrological model assessment," *Advances in Geosciences*, vol. 5, pp. 89–97, 2005.

[55] P. Coulibaly, F. Anctil, and B. Bobée, "Multivariate reservoir inflow forecasting using temporal neural networks," *Journal of Hydrologic Engineering*, vol. 6, no. 5, pp. 367–376, 2001.

[56] G. E. Box and D. R. Cox, "An analysis of transformations," *Journal of the Royal Statistical Society, Series B: Methodological*, vol. 26, pp. 211–252, 1964.

[57] C. Onyutha, "Identification of sub-trends from hydro-meteorological series," *Stochastic Environmental Research and Risk Assessment*, vol. 30, no. 1, pp. 189–205, 2016.

[58] C. Onyutha, "Variability of seasonal and annual rainfall in the River Nile riparian countries and possible linkages to ocean–atmosphere interactions," *Hydrology Research*, vol. 47, no. 2, pp. 171–184, 2015.

[59] WMO, "Hydrological data," in *Manual on Low-Flow Estimation and Prediction*, Operational Hydrology Report No. 50, WMO-No.1029, p. 138, WMO, Geneva, Switzerland, 2008.

[60] A. C. Davison and D. V. Hinkley, *Bootstrap Methods and Their Application*, Cambridge University Press, Cambridge, UK, 1997.

[61] L. M. Nhat, Y. Tachikawa, and K. Takara, "Establishment of intensity-duration-frequency curves for precipitation in the monsoon area of Vietnam," *Annals of Disaster Prevention Research Institute Kyoto University*, vol. 49, pp. 93–103, 2006.

[62] C. Onyutha and P. Willems, "Empirical statistical characterization and regionalization of amplitude–duration–frequency curves for extreme peak flows in the Lake Victoria Basin, East Africa," *Hydrological Sciences Journal*, vol. 60, no. 6, pp. 997–1012, 2015.

[63] C. Onyutha and P. Willems, "Uncertainties in flow-Duration-Frequency relationships of high and low flow extremes in Lake Victoria Basin," *Water*, vol. 5, no. 4, pp. 1561–1579, 2013.

[64] I. Pickands, "Statistical inference using extreme order statistics," *The Annals of Statistics*, vol. 3, no. 1, pp. 119–131, 1975.

[65] C. Onyutha, "Statistical modelling of FDC and return periods to characterise QDF and design threshold of hydrological extremes," *Journal of Urban and Environmental Engineering*, vol. 6, no. 2, pp. 132–148, 2012.

[66] C. Onyutha and P. Willems, "Uncertainty in calibrating generalised Pareto distribution to rainfall extremes in Lake Victoria basin," *Hydrology Research*, vol. 46, no. 3, pp. 356–376, 2015.

[67] M. Lang, T. B. M. J. Ouarda, and B. Bobée, "Towards operational guidelines for over-threshold modeling," *Journal of Hydrology*, vol. 225, no. 3-4, pp. 103–117, 1999.

[68] J. Beirlant, J. L. Teugels, and P. Vynckier, *Practical Analysis of Extreme Values*, Leuven University Press, Leuven, Belgium, 1996.

[69] B. M. Hill, "A simple general approach to inference about the tail of a distribution," *The Annals of Statistics*, vol. 3, no. 5, pp. 1163–1174, 1975.

[70] S. G. Gebrehiwot, A. Taye, and K. Bishop, "Forest cover and stream flow in a headwater of the blue nile: complementing observational data analysis with community perception," *Ambio*, vol. 39, no. 4, pp. 284–294, 2010.

[71] M. Staudinger, K. Stahl, J. Seibert, M. P. Clark, and L. M. Tallaksen, "Comparison of hydrological model structures based on recession and low flow simulations," *Hydrology and Earth System Sciences*, vol. 15, no. 11, pp. 3447–3459, 2011.