International Journal of Energy and Water Resources
https://doi.org/10.1007/s42108-020-00075-4

ORIGINAL ARTICLE

Suitability of averaged outputs from multiple rainfall-runoff models for hydrological extremes: a case of River Kafu catchment in East Africa

C. Onyutha[1] · C. J. Amollo[1] · J. Nyende[1] · A. Nakagiri[1]

Received: 2 March 2020 / Accepted: 5 May 2020

Abstract

In this study, seven rainfall-runoff models were applied to model daily River Kafu flows from 1952 to 1981. Among others, models from the rainfall-runoff library of the eWater toolkit were applied. Optimal parameters of each model were obtained based on an automatic calibration strategy. Averaging in terms of simple arithmetic mean, hereinafter taken as the multi-model ensemble (MME), was performed to independently and identically distributed events separately extracted from the outputs of the individual models. How well the MME captured variation and frequency of observed hydrological extremes was assessed. Models performed better for high flows than low flows. Absolute model average biases on quantiles with return periods from 1 to 30 years were over the ranges 5.5–83.6% and 11.6–57.7% for high flows and low flows, respectively. It is envisaged that making model structures flexible and performing calibration with objective functions constrained to extreme events can enhance simultaneous capturing of high flows and low flows. The amount of variance in annual maxima series that could be explained by the multi-model ensemble was 73.4% and ranged from 35.1 to 82.5% for the individual models. This made the multi-model ensemble better than outputs from six of the seven models. For the annual minima flows, the multi-model ensemble yielded the smallest root mean squared error but the third largest coefficient of determination. Notably, the suitability of the multi-model ensemble in capturing the hydrological extremes depends on the selected goodness-of-fit measure, approach for combination of model outputs, number of models considered and length of data used.

Keywords East Africa · High flows · Hydrological extremes · Low flows · Multi-model ensemble · Rainfall-runoff model · River Kafu catchment

Introduction

Due to the changing climate, many locations of the world, especially those in arid environments, are faced with the challenge of water insecurity. This affects various aspects of water resources management, something which requires careful planning. River Kafu catchment in Uganda is primarily a rangeland supporting vast herds of livestock. In addition, River Kafu provides water for other domestic purposes and major economic activities like fishing and agriculture. The catchment has been affected by the potential impacts of climate change and variability. Furthermore, the catchment also experiences alteration due to the natural or anthropogenic activities and is continuously being faced with water scarcity, deteriorating water quality, floods and droughts, and these issues are all negatively impacting on Uganda's quest for economic and social development.

Within the framework of the integrated water resources management, modeling can be used to understand the hydrological functioning of a catchment with possible analyses of scenarios related to planning and operation of water resources management applications. Some of these applications include reservoir operation, irrigation, hydropower generation and ecological management practices. For hydrological modeling, both physically or process-based and conceptual lumped models exist. Examples of physically based hydrological models include the Soil and Water Assessment Tool SWAT (Arnold et al. 1998) and MIKE Système Hydrologique Européen (SHE) (Abbott et al. 1986a,

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s42108-020-00075-4) contains supplementary material, which is available to authorized users.

✉ C. Onyutha
conyutha@kyu.ac.ug

[1] Department of Civil and Building Engineering, Kyambogo University, P.O. Box 1, Kyambogo, Kampala, Uganda

Published online: 26 May 2020

b), Waterloo Flood Forecasting Model WATFLOOD (Kouwen 1988, 2000; Kouwen and Mousavi 2002), The Coupled Routing and Excess Storage CREST (Wang et al. 2011), HydroGeoSphere (Therrien et al. 2005), Institute of Hydrology Distributed Model IHDM (Beven et al. 1987; Calver and Wood 1995), and the Modified Interaction Soil Biosphere Atmosphere MISBA (Kerkhoven and Gan 2006). In a fully distributed model, the catchment is treated as independent cells taking into account the spatial heterogeneities of landuse, soil, elevation and climatic variables (such as precipitation and evapotranspiration) in the generation of the various sub-flow components. Physical models consider the use of several parameters (stemming from numerous assumptions), thereby complicating calibration and analyses. On the other hand, a conceptual model has simple structure with few parameters and does not require many model inputs. In modeling using a lumped conceptual model, the catchment is treated as a single unit to simulate flow at the catchment outlet. Examples of lumped conceptual rainfall-runoff models include Hydrological Model focusing on sub flows variation (HMSV) (Onyutha 2019), Australian Water Balance Model (AWBM) (Boughton 2004), TANK model (Sugawara and Fuyuki 1956), SACRAMENTO model (Burnash 1995), Identification of Unit Hydrographs and Component Flows from Rainfall, Evaporation and Stream (IHACRES) flow data (Jakeman et al. 1990; Jakeman and Hornberger 1993), SIMHYD (Porter and McMahon 1972), Hydrologiska Byråns Vattenavdelning (HBV) (Bergström 1976), Soil Moisture Accounting and Routing (SMAR) model (O'Connell et al. 1970) and Nedbør-Afstrømnings Model (NAM) (Nielsen and Hansen 1973; Danish Hydraulic Institute DHI 2007).

In hydrological modeling, lumped conceptual models tend to be preferred to the fully distributed models. Considering only the East African region (where the study area is located), lumped conceptual models were applied in several studies (Legesse et al. 2003, 2004; Easton et al. 2010; Betrie et al. 2011; Bitew and Gebremichael 2011; Ebrahim et al. 2013). Each of these studies applied a single hydrological model. The choice and use of a particular models among others can lead to large bias in reproducing observed extreme hydrological events (Onyutha 2016). To even out such a possible bias in hydrological analysis, the use of several models is required to support decision related to management of water resources applications. Many lumped conceptual models are not readily available (in terms of their codes and/or installable software) for researchers and this limits their applications for hydrological modeling. The Rainfall Runoff Library (RRL) (Podger 2004) comprises several lumped conceptual models (including AWBM, TANK model, SACRAMENTO model, SMAR and SIMHYD) which are readily available to be acquired upon registration by researchers via the link https://toolkit.ewater.org.au/Tools/RRL (accessed 16 April 2020). From the same link, also IHACRES can be obtained as a standalone (not contained within the RRL).

The RRL models have been applied in several studies such as Haque et al. (2014), Ankit and Tiwari (2015), Kumar et al. (2015), Yu and Zhu (2015), Zhang et al. (2016), Kunnath-Poovakka and Eldho (2019), Yang et al. (2019), Onyutha (2019), Jaiswal et al. (2020) and Esmali et al. (2020). Jaiswal et al. (2020) found TANK to perform better than the AWBM when applied to simulate rainfall runoff over the Tandula basin of Chhattisgarh (India). Esmali et al. (2020) concluded that IHACRES performed better in simulating hydrological conditions for semi-arid and semi-humid than arid watersheds in Iran. Haque et al. (2014) applied AWBM to assess uncertainty in hydrological modeling of ungauged catchments in Australia. Ankit and Tiwari (2015) applied the AWBM to model monthly river flow of Bina basin in India. Using the AWBM, Zhang et al. (2016) modeled monthly streamflow in the Poyang Lake basin. Kunnath-Poovakka and Eldho (2019) applied AWBM and SACRAMENTO to simulate monthly flows of a number of catchments including Adhala, Bhandardara, Gargaon, Kadwa and Mula all in the upper Godavari river basin, India. Daily flow of the Kesinga catchment of Mahanadi River Basin, located in Odisha, India was modeled by Kumar et al. (2015) using the AWBM, SIMHYD, SACRAMENTO, SMAR and TANK. The river flow of the Fernow catchment in north-central West Virginia, USA was modeled using the AWBM and SIMHYD (Yu and Zhu 2015). Onyutha (2016) applied the AWBM, IHACRES, SACRAMENTO, SIMHYD and TANK to simulate the river flow in the upper Blue Nile basin in Ethiopia. Another model readily available to be obtained by modelers is the recently introduced HMSV (Onyutha 2019) which was tested also on the upper Blue Nile. The HMSV has a framework which allows optimization of the model with respect to hydrological extremes.

To address the issue of uncertainties from models, modeling results can be combined to obtain a single set of predictions. Combination of results from models of varying structural complexity yields multi-model ensemble; however, if several realizations from one deterministic model are combined, we obtain single model ensemble (Baker and Ellison 2008). Single model ensemble also comprises combination of results from various models of the same structural type. The concept of ensemble hydrological modeling goes back to the 1970s (Twedt et al. 1977). However, most of (if not all) the studies on multi-model ensemble in hydrology focused on mean hydrological conditions. Examples of such studies include Rojas et al. (2008), Diks et al. (2010), Arsenault et al. (2015) and Kumar et al. (2015). Hydrological modeling with focus on frequency and trends in floods and droughts is vital for planning of risk-based water resources management.

By the time of writing this paper, there were no studies that applied hydrological models to model the variation in the flows of the River Kafu catchment moreover with the focus on hydrological extremes. The purpose of this study conducted between January 2019 and February 2020 at Kyambogo University, Kampala, Uganda was to compare the performance of several lumped conceptual models in the simulation of daily River Kafu flows. This was done while investigating how well an ensemble from the multiple models could capture the variation and frequency of the observed hydrological extremes.

Materials and methods

Study area

The catchment of River Kafu (Fig. 1) is located in the western part of Uganda and falls within the Albertine region. River Kafu has a catchment area of 15,983 km^2 which extends longitudinally from 31°10′ E to 32°41′ E and between 0°11′ N and 1°56′ N in the north-south direction. The catchment of River Kafu is in the western part of Uganda and comprises areas from a number of districts including Hoima, Kibale, Kyankwanzi, Luwero, Masindi, Nakasongola and Nakaseke. River Nkusi passes via the swamp from which River Kafu emanates. River Kafu has two tributaries including River Mayanja and River Lugogo. River Kafu empties into the Victoria Nile at the location with longitude = 32.095 and latitude = 1.647. River Kafu catchment is surrounded by a number of lakes including Kyoga, Victoria, Albert, Wamala, Edward and George. There are also a number of mountains in between which River Kafu catchment is located including Rwenzori, Elgon and Moroto. Figure 1 comprises the Digital Elevation Model (DEM) as the background map. The hole-filled DEM derived from the USGS/NASA (Jarvis et al. 2008) and processed by the International Centre for Tropical Agriculture (CIAT-CSI-SRTM) using interpolation methods described by Reuter et al. (2007) was downloaded online via the link https://srtm.csi.cgiar.org (accessed 03 December 2019). It is noticeable that the altitude is higher in the southern than northern part of the catchment.

River flow data

Daily time series of River Kafu flows observed at the main hydrological station (with station ID 83213) along Gulu–Kampala highway (32°02′28″ N, 01°32′44″ E) was obtained from the Department of Water Resources Management under the Ministry of Water and Environment, Uganda. The obtained river flow data were recorded from March 1952 to October 2018.

Rainfall data

Daily rainfall data from a total of 34 gauge stations were obtained from the Uganda National Meteorological Authority in Uganda. However, only 15 stations (see Table 1) within and around the catchment were considered for rainfall-runoff

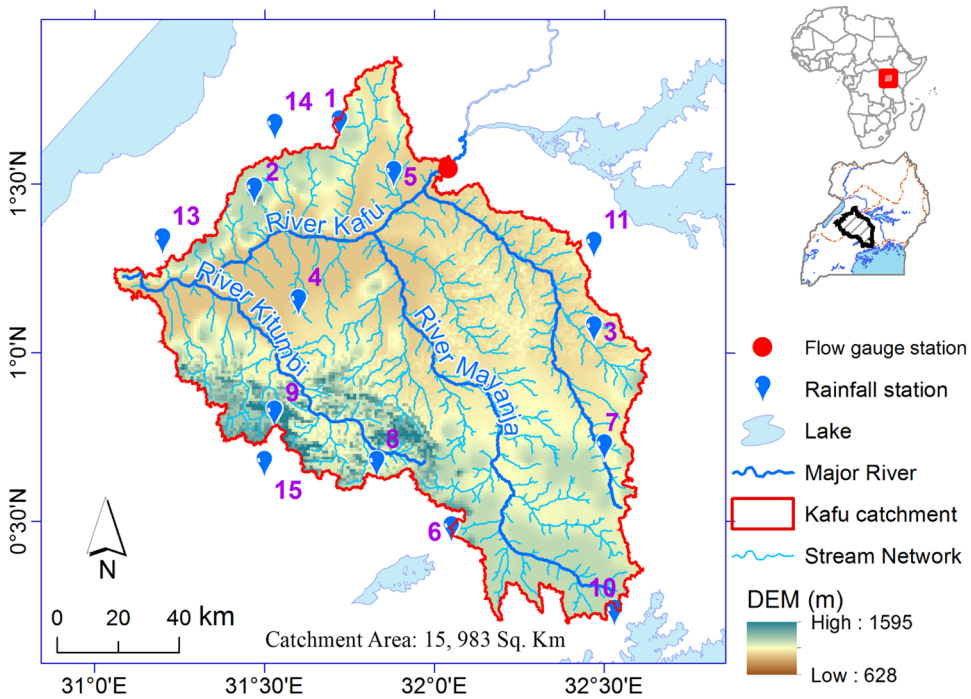

Fig. 1 River Kafu catchment and selected hydro-meteorological stations

Table 1 Rainfall stations considered used in this study

S. no	Latitude (°)	Longitude (°)	Station ID	Description	Missing record (%)
1	1.68	31.72	88,310,030	Masindi Met	5.2
2	1.48	31.47	88,310,060	Bulindi	2.7
3	1.07	32.47	88,310,180	Kakoge Gombola	2.4
4	1.15	31.6	88,310,190	Butembe	16.4
5	1.53	31.88	88,310,280	Namyekudo	21.6
6	0.477	32.05	89,320,610	Bakijulula	4.1
7	0.72	32.5	89,320,010	Bukalasa Agric	13.2
8	0.67	31.83	89,310,100	Bukuya Gombolola	15.8
9	0.82	31.53	89,310,210	Buta – Mubende	8.9
10	0.23	32.53	89,320,760	Kajjansi Fish Farm	10.8
11	1.32	32.47	88,310,020	Nakasongola	17.5
12	1.23	30.82	88,310,020	Kyangwali	20.8
13	1.33	31.2	88,310,170	Kiziranfumbi	22.8
14	1.67	31.53	88,310,240	Nyabyeya	1.6
15	0.67	31.5	89,310,200	Namungo	14.1

modeling. The data from these selected rainfall stations covered the period 1952–1981. The 15 stations with their locations shown in Fig. 1 were selected because their percentages of missing records were less than 25%. In several areas of the sub-Saharan Africa, it is difficult to obtain rainfall data with the missing records less than 10%. The missing data records were in-filled using the inverse distance weighted interpolation technique.

Lumped conceptual models

A total of seven lumped conceptual models were used in this study. Six of the seven models are internationally well known including AWMB, SACRAMENTO, SIMHYD, TANK, SMAR and IHACRES. The details of these models can be obtained from RRL (Podger 2004). The detail of the seventh model, HMSV can be obtained from Onyutha (2019). All the seven models were selected because they are freely available to hydrological modelers, easy to learn and apply, applicable mostly in modeling runoff in a lumped conceptual way which is directly linked to the purpose of this study. The structural details of the selected models can be obtained from the Supplementary Material S1 (see Figure S1.1–S1.7).

Temperature data

To compute potential evapotranspiration (PET) required for hydrological modeling, the minimum (T_{min}) and maximum (T_{max}) temperature were required. Due to unavailability of observed temperature data over the selected catchment, daily series of T_{min} and T_{max} of the Princeton global forcing (PGF) (Sheffield et al. 2006) covering the period 1948–2008 were downloaded via the link https://hydrology.princeton.edu/data/pgf/0.5deg (accessed 12 May 2019). The computation of PET was based on the T_{min} and T_{max} over the period 1951–1981 as that for the river flow series.

The daily PET was computed based on the Hargreaves formula using (Hargreaves and Samani 1982; Hargreaves and Samani 1985)

$$ET_o = 0.0023R_a\left(T_{mean} + 17.8\right)\left(T_{max} - T_{min}\right)^{0.5}, \quad (1)$$

where ET_o (mm/day) is the daily PET, T_{mean} (°C) denotes the mean of T_{min} and T_{max} and R_a stands for the incoming solar radiation (W/m^2).

Rainfall-runoff modeling

For each of the seven selected models, daily model inputs were obtained as catchment-wide averaged rainfall and PET. The model inputs were prepared separately in the format required by each model. In the next step, the initial conditions and ranges of the parameters for each model were set.

Calibration can be done manually or automatically. In this study, to eliminate subjectively due to manual calibration, the parameters of each of the models were changed using automatic calibration strategy. For all the models taken from the e-Water toolkit (AWBM, SACRAMENTO, TANK, SIMHYD and SMAR), the calibration was based on the shuffled complex evolution (SCE) (Duan 1994) algorithm incorporated in the RRL. The calibration of the HMSV was based on the generalized likelihood uncertainty estimation (Beven and Binley 1992). The calibration of IHACRES followed the procedure detailed by Croke et al. (2005). Eventually, the automation of IHACRES calibration was twofold. In

the first step, the linear module calibration was controlled by fixed transfer function. To do so, cross-correlation was used to calculate the delay between rainfall and stream flow. Second, using non-linear module calibration, a grid search through the parameter space was performed to obtain the best parameter set. It is important to note that the calibration of the HMSV was also twofold, hereinafter represented as HMSV(1) and HMSV(2). First, in the HMSV(1), the conventional approach was adopted just like other traditional models. Conventionally, the calibration was done based on the overall water balance without focus to extreme conditions. In the second approach, the HMSV(2) comprised a specific calibration framework in which the mismatch between the high flows and low flows was reduced iteratively as detailed by Onyutha (2019).

Calibration of each model was performed using daily rainfall and PET data from 1952 to 1961. The Nash–Sutcliffe efficiency (NSE) (Nash and Sutcliffe 1970) was used as an optimization objective function for the calibration. To determine the validity of the calibrated model, daily rainfall and PET data from1962 to 1981 were used. During validation, the optimal parameters obtained during calibration were kept constant and used to perform model simulation. The "goodness-of-fit" for both calibration and validation periods were obtained in terms of the NSE

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}\left(Q_{m,i} - X_{o,i}\right)^2}{\sum_{i=1}^{n}\left(X_{o,i} - \overline{X}_o\right)^2}, \tag{2}$$

where $Q_{\mathrm{m},i}$ = modeled Flow, X_o = observed Flow, $\overline{X}_o$ = the mean of the X_o's and n = the sample size. The best (worst) performance of a model in terms of the NSE is 1 (∞).

Comparison of model performance under extreme hydrological conditions

To obtain hydrological extreme values from both observed and simulated flows, the procedure of extracting nearly independent peak over threshold (POT) events developed by Onyutha (2017) was used. This procedure is incorporated in a tool called FAN-Stat (Onyutha 2017) which stands for frequency analyses considering non-stationarity. Before the extraction of independent low flow events, the original flow was transformed through inversion. This type of flow transformation enables the frequency analysis of low flow quantiles to be based on a similar approach applicable when considering high flows. The extracted nearly independent high flow and low flow events were used to conduct frequency analyses. Frequency of an extreme flow event was in terms of the return period (in years).

Empirically, to get the return period of an event, the extracted events were ranked from the highest to the lowest. In the ranking, the largest and smallest events were assigned ranks of 1 and m, respectively, where m was taken as the sample size. Let n denote total number of data record length (in years) and consider r_i as the rank of the ith event. The empirical return period T for the ith event was computed as $T = n/r_i$ in years. Graphically, the quantiles were plotted against the log-transformed values of T.

The performance of each model in reproducing extreme hydrological conditions was evaluated in terms of root mean squared error (RMSE) (Eq. 3), and model average bias (MAB) (Eq. 4),

$$\mathrm{RMSE}\left(\mathrm{m^3/s}\right) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Q_{m,i} - X_{o,i}\right)^2}, \tag{3}$$

$$\mathrm{MAB}\ (\%) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Q_{\mathrm{m},i} - X_{o,i}}{X_{o,i}} \times 100\right), \tag{4}$$

and for an ideal model (or with the best performance), the value MAB or RMSE is zero.

Apart from the POT events, river flow series comprising the maximum events in each year was extracted from observed flow as well as modeled outputs from each model. Similarly, annual minima series was obtained from each modeled outputs as well as observed flow. Co-variation in observed and annual minima series or annual maxima flows was assessed in terms of the coefficient of determination (R^2) or R-squared using

$$R^2 = \frac{\left(\sum_{i=1}^{n}\left(Q_{\mathrm{m},i} - Q^m\right)\left(X_{o,i} - \overline{X}_o\right)\right)^2}{\sum_{i=1}^{n}\left(Q_{\mathrm{m},i} - Q^m\right)^2 \sum_{i=1}^{n}\left(X_{o,i} - \overline{X}_o\right)^2}, \tag{5}$$

where Q^m denotes the mean value of modeled quantiles, and $Q_{\mathrm{m},i}$,$X_{o,i}$ and $\overline{X}_o$ are as defined for previous equations. The best and worst model can be given by R^2 equal to 1 and 0, respectively.

Multiple model ensemble

Outputs from models can be combined in a number of ways including weighted average method, constrained multiple linear regression, multi-model super ensemble and simple model average. For simplicity, the results of the various models from this study were combined through simple averaging. Since the focus was on hydrological extremes, POT, annual maxima flows and annual minima events were first extracted. In the next step, the POTs from each model were sorted from the highest to the lowest. The quantiles of corresponding empirical return periods were averaged to obtain the multiple model ensemble (MME).

The annual maxima flows for corresponding year were also combined to yield another set of the MME. The final set of MME was obtained by averaging the annual minima series for the corresponding year. Observed and modeled hydrological extremes (annual maxima or minima series) were assessed in terms of R-squared (Eq. 5). It is vital to note that in obtaining MME, some models which do not perform well tend to be eliminated from the multi-model pool for combination of models' outputs. Such a consideration would be necessary where there are several models of different structural complexities applied in a study focusing on mean hydrological conditions. Therefore, in this study which focused on hydrological extremes, all the models were used to obtain the MME.

Results and discussion

Performance of models under moderate hydrological condition

Figure 2 shows observed discharge and simulated flows from the various models. The parameters of each model used to obtain results presented in Fig. 2 can be obtained

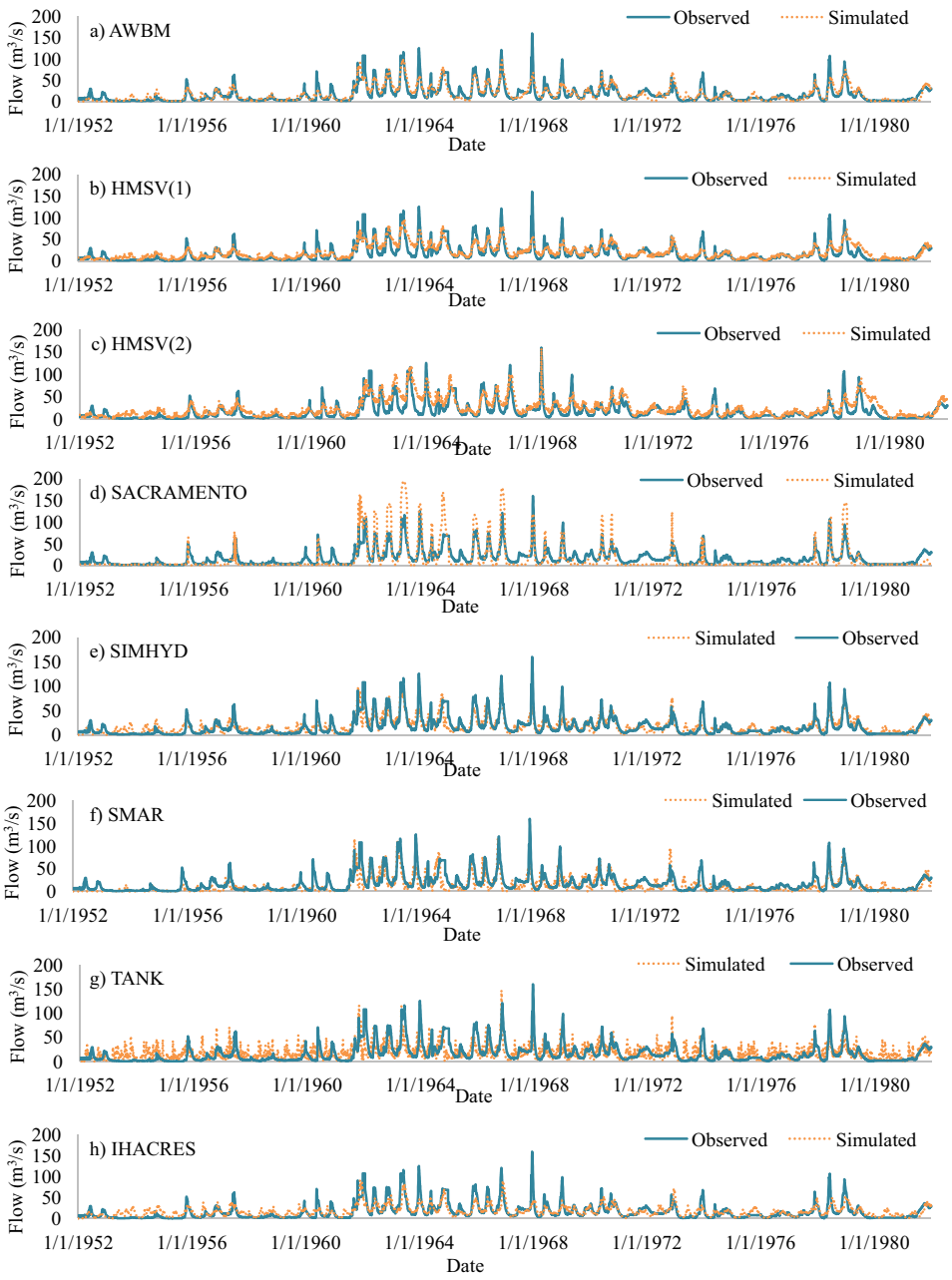

Fig. 2 Observed discharge and simulated flow based on **a** AWBM, **b** HMSV(1), **c** HMSV(2), **d** SACRAMENTO, **e** SIMHYD, **f** SMAR, **g** TANK and **h** IHACRES

in Supplementary Material S1 (see Tables T1.1–T1.7). Graphically, it is noticed that all the models captured the overall variation in the flow with time (Fig. 2a–f). However, the peak high flows were under- or over-estimated in some cases. For instance, most of the peak high flows were over-estimated by SACRAMENTO (Fig. 2c). Generally, the models exhibited satisfactory performances in reproducing mean hydrological condition.

The statistical "goodness-of-fits" for the calibration and validation can be seen in Fig. 3. Statistically, the best performance of a hydrological model would be indicated by the NSE of 1. However, in this study, the calibration results considering the overall water balance showed that the highest NSE values of 0.54 and 0.52 were obtained by the HMSV(1) and AWBM, respectively. The NSE values of the remaining models were below 0.5. In several past studies (Kunnath-Poovakka and Eldho 2019; Onyutha 2019; Yang et al. 2019; Zhang et al. 2016; Ankit and Tiwari 2015; Kumar et al. 2015; Yu and Zhu 2015), the performance of these RRL models (AWBM, TANK, SAC, SIMHYD and SMAR) applied in this research was noted to differ. For monthly flow of the Bina basin in India, AWBM yielded NSE of 0.82 and 0.62 for calibration and validation, respectively (Ankit and Tiwari 2015). Simulation of monthly streamflow in the Poyang Lake basin using the AWBM yielded NSE greater than 0.9 (Zhang et al. 2016). AWBM and SACRAMENTO when applied by Kunnath-Poovakka and Eldho (2019) to simulate monthly flows of a number of catchments in the upper Godavari river basin yielded NSE values below 0.5 for Adhala, Gargaon and Kadwa catchments. For the monthly flow of Bhandardara catchment, the NSE of AWBM and SACRAMENTO was slightly above and below 0.5 (Kunnath-Poovakka and Eldho 2019). However, for Mula catchment, Kunnath-Poovakka and Eldho (2019), obtained NSE of about 0.6 using both the AWBM and SACRAMENTO. All the above studies applied the models using monthly data. However, in this study, the hydrological models were applied using daily hydro-meteorological data. Even when the modeling was done for various catchments using daily series, the NSE values still differed among models. For instance, in modeling daily flow of the Kesinga catchment in Odisha (India), Kumar et al. (2015) applied the AWBM, SIMHYD, SACRAMENTO, SMAR and TANK and the NSE of each model was above 0.5. In simulation of hydrological extreme conditions in the Blue Nile basin, Onyutha (2016) applied the AWBM, IHACRES, SACRAMENTO, SIMHYD and TANK and got calibration NSE values for the models to varying extents between 0.65 and 0.85. In studies by Kumar et al. (2015) and Onyutha (2016), manual and automatic calibrations of the RRL models were adopted, respectively. Yu and Zhu (2015) found NSE of 0.49 and 0.46 for the AWBM and SIMHYD, respectively, when applied to Fernow in north-central West Virginia, USA. There are several reasons why results of hydrological models differ across catchments. Models are of different structures and this makes their compatibility with available data to be of varying extents. Results of the same model applied to various catchments also differ because catchments are not the spatially similar with respect to size, topography, soils, geology and hydrological conditions (Onyutha 2016) as well as space–time variability of rainfall and PET. The extent to which hydrological conditions may also be impacted upon by human factors and climate variability (Pirnia et al. 2019; Rutkowska et al. 2017) varies from one catchment to another and this can lead to differences in how models respond to the hydro-meteorological inputs from various catchments.

Considering the full time series, the NSE values for AWBM, HMSV(1), SACRAMENTO, SIMHYD, SMAR, TANK, HMSV(2) and IHACRES were 0.70, 0.67, 0.35, 0.56, 0.41, 0.14, 0.53 and 0.48, respectively. In other words, the NSE values were higher for validation than calibration periods. In hydrological modeling, due to the difficulty in transferability of the model performance with respect to the variation in the catchment hydro-climatic conditions over time, the NSE of calibration tends to be higher than that for validation period. However, as seen from Fig. 3, the NSE values were higher for the validation than calibration period. To explain this, it is vital to first note that around 1961, there occurred a step jump in the flow mean (Fig. 2). In other words, the flow values were higher for the period after than before 1961. On a cursory look at Eq. (2), it can be realized that using the squares of the difference between the

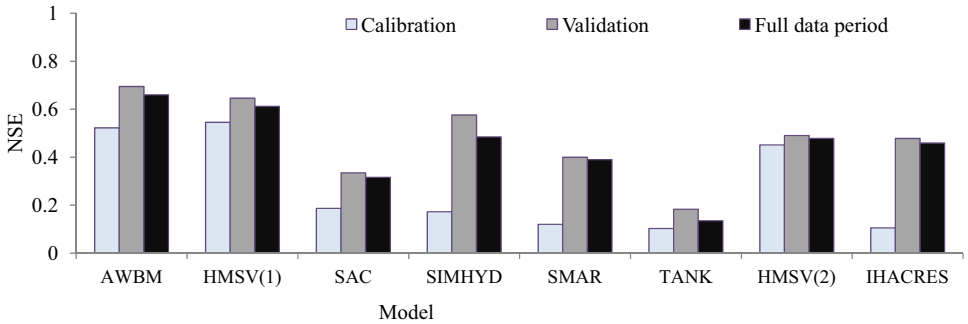

Fig. 3 Statistical measure (NSE) of the mismatch between observed and simulated flow

simulated and the observed variable at each time step, the errors are larger for high flows than low flows. Eventually, the process of minimizing errors during calibration premised on NSE as the objective function enhances the performance of the model for the period during which the flow values are large. This could explain the higher NSE values obtained for validation than those yielded during calibration.

Performance of models considering hydrological extremes

Figure 4 shows graphical assessment of the performance of the various models in reproducing nearly independent hydrological extreme events. Considering high flows (Fig. 4a), it is noticeable that for return periods from 1 to 3 years, most models (except SACRAMENTO which showed systematic over-estimation for all the peak high flow events), exhibited good performance. However, for return periods greater than 3 years, most models tended to under-estimate the high flow quantiles. For low flows (Fig. 4b), most models underestimated the low flow quantiles. The relationship between the extreme flow events and the reduced variate (here in this study, considered in terms of log-transformed return period) is expected to be linear for a normal tailed flow. In this way, the slope of such a line characterizes the variation in the extreme flow conditions something that should be adequately reproduced by a model. In this line, it is noticeable that the scatter points for SMAR, SIMHYD and SACRAMENTO appeared almost like horizontal lines. In other words, the slope of return period curves were captured fairly by HMSV(1) and AWBM. The best performance, through visual judgment, was obtained using the HMSV(2). For high flows, the MME fell closer to the observed quantiles than outputs from a number of models (Fig. 4a). Similarly for low flows (Fig. 4b), MME of low flows was better than the quantiles from a number of individual models. Nevertheless, the MME performed well for return periods from 1 to 3 years. Between 3 and 30 years, the deviation between MME and observed quantiles increased systematically. This was due to the large deviations of the outputs from models especially for high return periods.

It can be seen that the slope of the linear relationship between observed quantiles and return periods was reproduced better by the high flows than low flows. This is because the design of most hydrological models are structurally suited to capture high flows than low flows (Staudinger et al. 2011; Onyutha 2016). Moreover, the use of objective functions constrained to overall water balance during calibration gives more weights to high flows than low flows (Onyutha 2016). This requires the hydrological models to have their structures flexible so that they could adaptively capture both high flows and low flows.

Figure 5 shows statistical performance of models in reproducing extreme high flow and low flow events considering return periods from 1 to 30 years. The smaller the values of MAB and RMSE, the better is the model. It can be noticed that for both high flows (Fig. 5a, b) and low flows (Fig. 5c, d), SACRAMENTO (denoted as SAC) exhibited the largest values of MAB and RMSE. As already seen from Fig. 4, the largest mismatch between the observed and simulated quantiles was obtained for model outputs from SAC. The performance of the AWBM and HMSV(1) was comparable. However, HMSV(2) had the lowest values of MAB and RMSE for both high flows and low flows; this was because the calibration scheme adopted for the HMSV(2) is, as highlighted before in section of methodology, tailored toward reproducing such independent extreme hydrological events as shown in Fig. 4. This suggests that an improved performance of traditional models can be obtained if their calibrations are based on specialized frameworks focusing on high flows and low flows (Onyutha 2019). Perhaps, instead of the NSE, another objective function which gives equal weights to both high flows and low flows could be adopted. This is because, the use of the NSE during calibration makes traditional models to perform better for high flows than low flows. For the annual maxima series (Fig. 5a, b), the MME yielded the smallest MAB (1.05%) and second smallest RMSE (10.84%). However, for the annual minima series

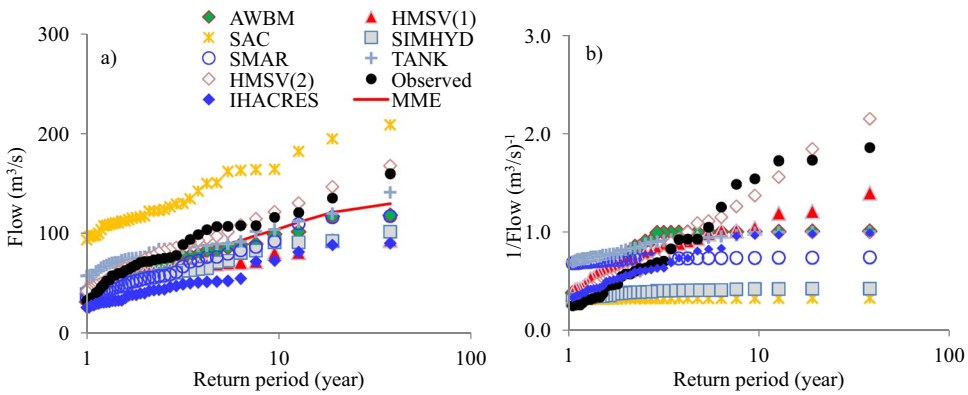

Fig. 4 Frequency of observed and simulated POT **a** high flows and **b** low flow quantiles. The legend entries of chart (**b**) are as those for (**a**). SAC stands for SACRAMENTO

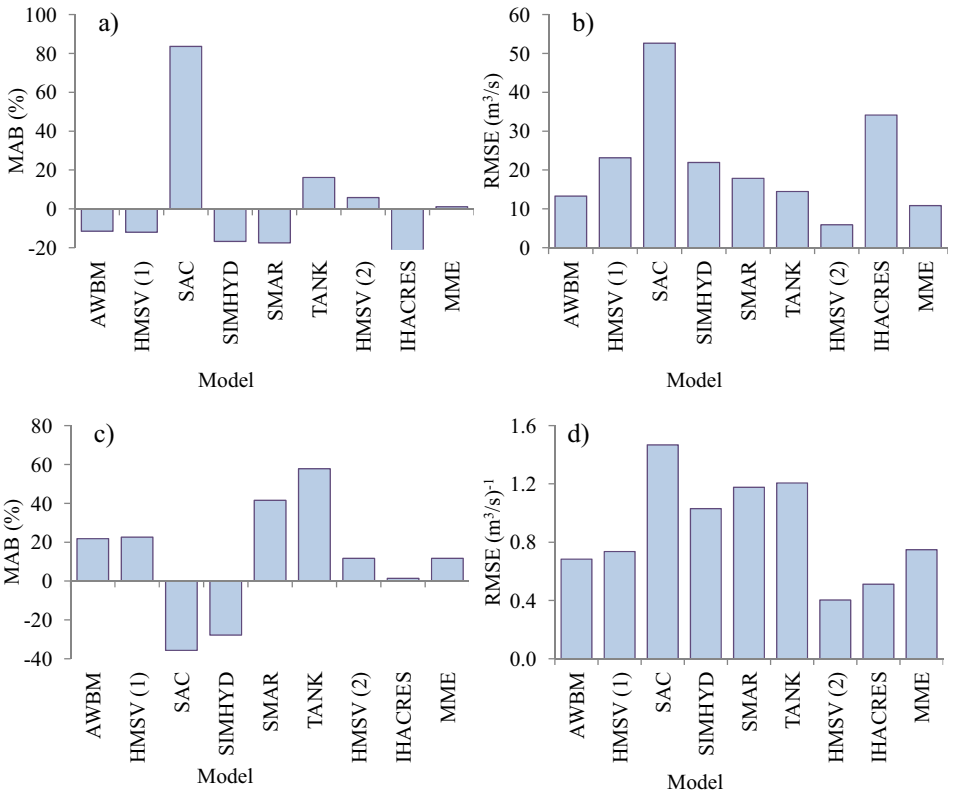

Fig. 5 Statistical measure (**a**, **c**) MAB and (**b**, **d**) RMSE for (**a**, **b**) high flows and (**c**, **d**) low flows. In the label of the horizontal axis, SAC and stands for SACRAMENTO

(Fig. 5c, d), the MME yielded the second smallest MAB (11.64%) and the third lowest RMSE (0.75%).

An important question to answer would be why the difficulty of hydrological models in capturing hydrological extremes. Poor performance of models could be indicative of the issue of data limitation and/or data quality issues. Using the AWBM, Haque et al. (2014) showed that the accuracy with which a hydrological model tended to reproduce the variation in observed river flow depended on the length of the data. Lack of data that exist due to short historical and spatially insufficient observations fully affects application of hydrological models (Grayson et al. 2002). This is because questionable quality of model inputs reduces the compatibility of the model structure with the available structure and makes calibration of the model a difficult task to achieve in modeling.

Figure 6 shows comparison of observed and simulated annual maxima flows and annual minima series. For the period 1952–1981 (when rainfall data was available for rainfall-runoff modeling), the various models captured well the variation in the annual maxima series (Fig. 6a). However, observed annual minima flow events (Fig. 6b) were under-estimated by a number of models including SACRAMENTO, SMAR and TANK. For both the annual maxima flows and annual minima series, the MME fell closer to the observed hydrological extremes than the outputs from a number of models (Fig. 6a, b). This highlights the importance of the MME in capturing the temporal variation in hydrological extremes.

To understand the recent variability, annual maxima and annual minima flow events over the period 1982–2017 were also plotted in Fig. 6. It is noticeable from Fig. 6a that the annual maxima series exhibited an increasing trend ($p = 0.0037$) for which the null hypothesis H_0 (no trend) can be rejected ($p < 0.05$). Increasing high flows indicate a shift towards wet conditions with possible occurrences of flooding events. River Kafu catchment experiences bimodal rainfall pattern with March–April–May and September–October–November (SON) rainy seasons. Rainfall events of large intensities tend to occur during the SON season. Due to the increasing rainfall, there are a number of bridges on River Kafu which have recently (between 2010 and 2020) been often submerged by the flooding water thereby hampering movement of vehicles. Several homes and agricultural fields have also been inundated, thereby leading to loss of property and displacements of large number local population in the River Kafu catchment.

Despite the increasing annual maxima events (Fig. 6a), it can be seen that the annual minima series exhibited a decreasing trend (Fig. 6b). The significance of the non-zero slope of a linear variation of the annual minima flows with time yielded $p = 0.0620$ meaning that the H_0 (no trend) could

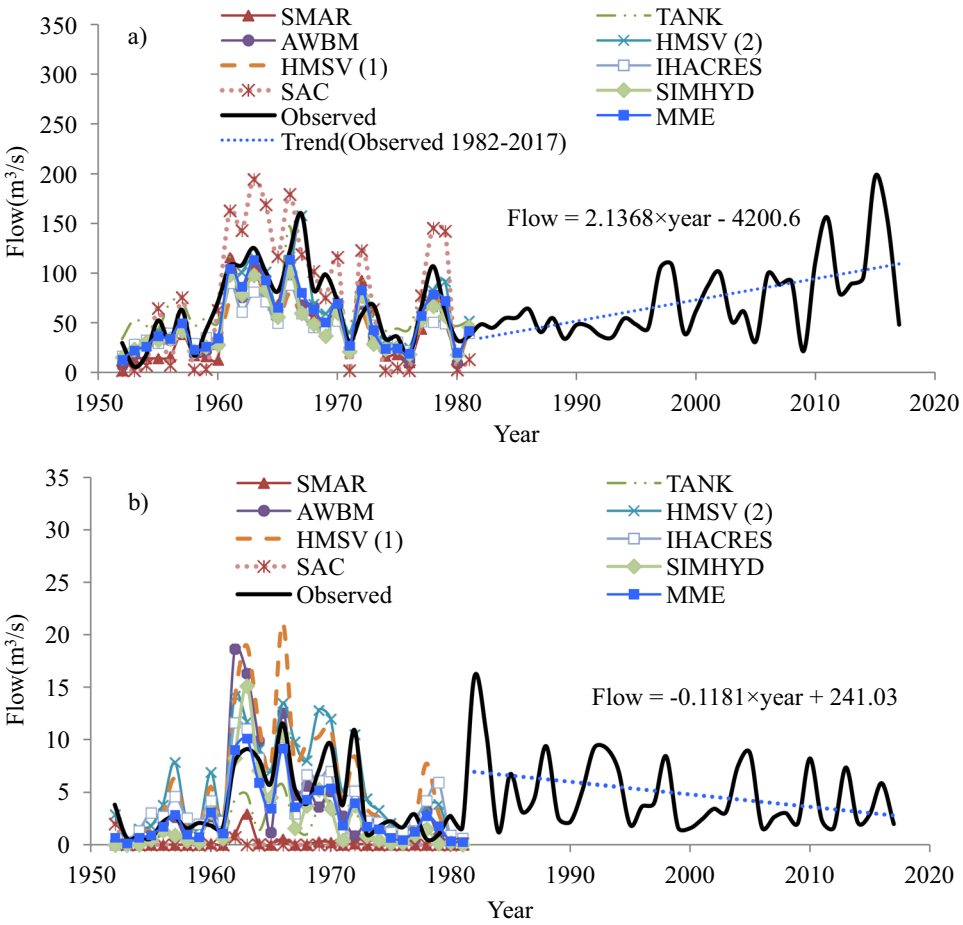

Fig. 6 Time series plots of observed and simulated **a** annual maxima series and **b** annual minima series

not be rejected ($p > 0.05$). Nevertheless, negative trend in low flows indicate a shift towards dry conditions with increased possibility of prolonged droughts. If the negative trend continues into the future due to decreasing rainfall during dry seasons, there will be prolonged droughts. There are several ways in which prolonged drought can affect the local population. Low rainfall total, high evaporation rates and depleted soil moisture coupled with high temperature (as common phenomena in a number of areas in East Africa (Onyutha 2020)) can cause drying of crops thereby leading to food shortage. Drought-related phenomena such as drying up of grasses from pasture or rangelands, and subsequent lack of water for animals to drink can lead to death of livestock. Given that cattle keeping is a major occupation across the River Kafu catchment, prolonged droughts can prompt the pastoralists to move far and wide in search of pasture and water for livestock. Migration from one place to another by pastoral communities (or nomadism) is a way of adapting to drought. However, occupants of an area and the immigrants are normally forced to compete for resources (such as water, land and pasture) and this is a potential source of conflicts. For the farming communities, drought-tolerant crop varieties can be planted during dry seasons. To circumvent the effects of prolonged droughts, income-generating activities outside farming should be promoted.

The coincidence of positive trend in high flows and decreasing low flows indicate an increase in flow intermittency or the existence of longer dry and wet spells or stronger wet-dry variations. This suggests a changing behaviour of River Kafu catchment as a hydrological system in response to rainfall and PET as the inputs. Changes in catchment behaviour or water resources can be driven by climate variability or human factors (Rutkowska et al. 2017; Pirnia et al. 2019). Some of the human activities which can exacerbate flooding across the River Kafu catchment include the massive deforestation, over-grazing and bush burning. In such cases, it is vital that measures to prevent rampart deforestation (especially for charcoal and expansion of agricultural areas) while promoting tree planting programmes should be considered in the catchment management plan of the River Kafu watershed. Since rainfall data were limited to period before 1982, future research should focus on determining whether hydrological models can replicate the variability in observed flow after 1981 till recent. Amidst data scarcity, the use of reanalyses or satellite precipitation and evaporation products such as Climate Forecast System

Reanalysis (CFSR), Gravity Recovery and Climate Experiment (GRACE) and Parameter elevation Regression on Independent Slopes Model (PRISM) could be explored. Fully distributed model could be used to understand the possible impacts human factors on water resources. Furthermore, it is envisaged that separation of the impacts of climate variability and human factors in express terms when undertaken in a future research will support planning of landuse management plan across the study area.

Figure 7 shows performance models in reproducing variation in hydrological extremes. Unlike for Fig. 5, the metrics RMSE and R-squared plotted in Fig. 7 were obtained with the annual maxima and minima series not sorted. The RMSE values were larger for peak high flows (Fig. 7a) than low flows (Fig. 7b). This was because high flows and low flows are of large and small magnitudes, respectively. Therefore, the differences between observed and modeled series were larger for high flows than low flows.

With respect to the RMSE from the annual maxima series (Fig. 7a), the MME was better than the outputs from all the models except HMSV(2). However, for the annual minima flow*s* (Fig. 7b), the MME performed better than the outputs of all the models. The amount of variance in observed hydrological extremes that could be explained by the MME was 73.4% and 68.9% for the annual maxima flow*s* and annual minima series, respectively. This indicated that the MME comprised the second and third best results in reproducing variation in observed annual maxima and annual minima series, respectively.

The best performance (smallest RMSE) for annual maxima series (Fig. 7a) and annual minima series (Fig. 7b) was by HMSV(2) and IHACRES, respectively. The largest RMSE values for annual maxima flow*s* and annual minima series were 38 m^3/s and 5 m^3/s, respectively, and all were obtained by SACRAMENTO. The lowest R-squared values for annual maxima flow*s* and annual minima series were 35.1% and 0.5% obtained by TANK (Fig. 7c) and SACRAMENTO (Fig. 7d), respectively. Considering the annual maxima flow*s* (Fig. 7c), the R-squared for each model was above 50%.

The amount of variance in the extreme flows that could be explained by the seven hydrological models, on average, was 64.4% and 49.2% for the annual maxima flows and annual minima series, respectively. This shows the difficulty for the hydrological models to capture variability in flows is larger for low flows than high flows. Not all the variance in extreme hydrological conditions was explained by the models. This means that there are other factors which control variation in river flows but were not captured by the hydrological models applied in this study. Such factors among others include climate variability, climate change, human factors, and data quality.

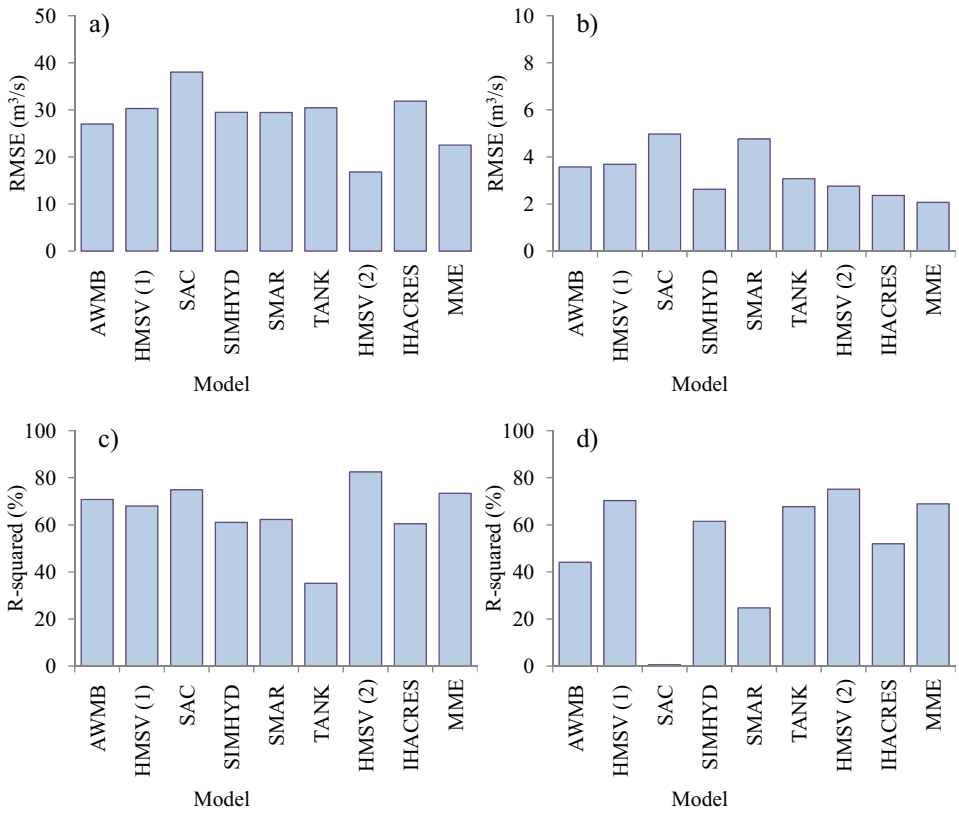

Fig. 7 Statistical measure **a**, **b** RMSE and **c**, **d** R-squared for (**a**, **c**) annual maxima series and (**c**, **d**) annual minima series. SAC stands for SACRAMENTO

Conclusions

In this study, seven lumped conceptual models were compared in terms of their performance to reproduce observed flow of River Kafu catchment. Apart from the HMSV and IHACRES, five models from the RRL (including AWMB, SACRAMENTO, TANK, SIMHYD and SMAR) were applied to simulate River Kafu flow using meteorological data from 1952–1981. Further analysis was conducted using river flow data from 1982 to 2018.

The hydrological models performed better for high flows than low flows. This followed the design of a typical traditional model structure tailored towards capturing high flows than low flows. The lowest R-squared values in reproducing the variation in observe annual maxima flows and annual minima series by the models were 35.1% and 0.5% obtained by TANK and SACRAMENTO, respectively. Considering the annual maxima flows, the R-squared for each model was above 50%. The largest root mean squared errors for annual maxima flows and annual minima series were 38 m^3/s and 5 m^3/s, respectively, and all were obtained by SACRAMENTO. The amount of variance in the annual maxima flows and annual minima series that could be explained by MME was 73.4% and 68.9%, respectively.

Using the conventional calibration strategy based on the objective function constrained to the overall water balance, the performance of the selected models was comparable. The hydrological extremes reproduced by the HMSV using the specific HMSV-based calibration framework were less biased than those based on the conventional schemes of traditional models. This indicated that the calibration of models could be done in a case-specific way. For instance, regarding climate change impact investigation which requires careful analyses of hydrological extremes, the calibration of the models could be premised on either low flows or high flows, whichever is central to the study under consideration.

Further analysis showed that the river flow series comprising the maximum and minimum event in each year over the period 1981–2018 was found to exhibit increasing ($p = 0.0037$) and decreasing ($p = 0.0620$) trends, respectively. Thus, the magnitudes of flooding events and severity of hydrological droughts were increasing. This could be due to the impacts of climate variability and human factors on water resources. Therefore, a future study is recommended to assess landuse change and climate variability impacts on the water resources in River Kafu catchment.

This study showed that for a particular catchment, the performance of models can differ. It is vital to compare results from various models if a particular model is to be chosen to support decision regarding applications related to floods and droughts. Importantly, to take into account the limitations of individual models, MME can be used to support planning and decision related to catchment management. It is recommended that in obtaining the MME for high flows or low flows, selection of suitable models should be based on some criteria with respect to stipulated values of a number of metrics such as NSE, MAB and RMSE. Furthermore, the suitability of the MME based on various combination approaches should be investigated for each catchment or basin under consideration.

Acknowledgements The authors are grateful to the Cooperative Research Centre for Catchment Hydrology in Australia for granting access to download the Rainfall Runoff Library. Acknowledgment is given to the Princeton Global Forcing, and the Ministry of Water and Environment, Uganda for data used in this study. The authors acknowledge that this paper was partly based on a study by Amollo (2020) under the supervision of Charles Onyutha (Ph.D.) and Jacob Nyende (Ph.D.).

Author contributions All authors were involved in the conception of the study. CO and CJA were further involved in data acquisition, modeling, analysis, interpretation, drafting manuscript and making necessary revisions. JN and AN reviewed and made comments on the draft manuscripts. All authors reviewed and approved the manuscript for submission.

Funding This research received no external funding.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Abbott, M. B., Bathrust, J. C., Cunge, J. A., O'Connell, P. E., & Rasmussen, J. (1986a). An introduction to European hydrological system–Système Hydrologique Européen (SHE) Part 1. History and philosophy of physically based distributed modeling system. *Journal of Hydrology, 87*, 45–59.

Abbott, M. B., Bathrust, J. C., Cunge, J. A., O'Connell, P. E., & Rasmussen, J. (1986b). An introduction to European hydrological system: Système Hydrologique Européen (SHE) Part 2. Structure of a physically based distributed modeling system. *Journal of Hydrology, 87*, 61–77.

Amollo, C. J. (2020). *Comparing the performance of different lumped conceptual hydrological models: a case study of River Kafu Catchment*. MSc. Dissertation, Department of Civil and Building Engineering, Kyambogo University, Kampala, Uganda.

Ankit, B., & Tiwari, H. L. (2015). Rainfall runoff estimation using RRL toolkit. *International Journal of Engineering Research and Technology, 4*(5), 595–599.

Arnold, J. G., Srinivasan, R., Muttiah, R. S., & Williams, J. R. (1998). Large area hydrologicmodeling and assessment-Part

I: model development. *Journal of American Water Resources Association, 34*(1), 73–89.
Arsenault, R., Gatien, P., Renaud, B., Brissette, F., & Martel, J.-L. (2015). A comparative analysis of 9 multi-model averaging approaches in hydrological continuous streamflow simulation. *Journal of Hydrology, 529*, 754–767.
Baker, L., & Ellison, D. (2008). The wisdom of crowds—ensembles and modules in environmental modeling. *Geoderma, 147*, 1–7.
Bergström, S. (1976). *Development and application of a conceptual runoff model for Scandinavian catchments, SMHI RHO 7*. Norrköping: SMHI.
Betrie, G. D., Mohamed, Y. A., Van Griensven, A., & Srinivasan, R. (2011). Sediment management modelling in the Blue Nile Basin using SWAT model. *Hydrology and Earth System Sciences, 15*(3), 807–818.
Beven, K., & Binley, A. M. (1992). The future role of distributed models: Model calibration and predictive uncertainty. *Hydrological Processes, 6*, 279–298.
Beven, K. J., Calver, A., & Morris, E. M. (1987). *The Institute of hydrology distributed model*. Wallingford: Institute of Hydrology. **(Report 98)**.
Bitew, M. M., & Gebremichael, M. (2011). Are satellite-gauge rainfall products better than satellite-only products for Nile hydrology? In A. M. Melesse (Ed.), *Nile River Basin* (pp. 129–141). Amsterdam: Springer.
Boughton, W. (2004). The Australian water balance model. *Environ Model Softw, 19*(10), 943–956.
Burnash, R. J. C. (1995). The NWS River forecast system-catchment modeling. In V. P. Singh (Ed.), *Computer Models of Watershed Hydrology* (pp. 311–366). Colorado: Water Resources Publications.
Calver, A., & Wood, W. L. (1995). The institute of hydrology distributed model. In V. P. Singh (Ed.), *Computer Models of Watershed Hydrology* (pp. 595–626). Colorado: Water Resources Publications.
Croke, B. F. W., Andrews, F., Spate, J., & Cuddy, S. M. (2005). *IHACRES User Guide, Technical Report 2005/19, iCAM*. Canberra: The Australian National University.
DHI. (2007). *Reference manual of MIKE11—a Modeling System for Rivers and Channels*. Hørsholm: DHI Water & Environment.
Diks, C. G. H., & Vrugt, J. A. (2010). Comparison of point forecast accuracy of model averaging methods in hydrologic applications. *Stochastic Environmental Research and Risk Assessment, 24*, 809–820.
Duan, Q. (1994). The SCE-UA method. *Journal of Hydrology, 158*, 265–284.
Easton, Z. M., Fuka, D. R., White, E. D., Collick, A. S., Ashagre, B. B., McCartney, M., et al. (2010). A multi basin SWAT model analysis of runoff and sedimentation in the Blue Nile, Ethiopia. *Hydrology and Earth System Sciences, 14*(10), 1827–1841.
Ebrahim, G. D., Jonoski, G. Y., van Griensven, A., & Di Baldassarre, G. (2013). Downscaling technique uncertainty in assessing hydrological impact of climate change in the Upper Beles River Basin, Ethiopia. *Hydrology Research, 44*(2), 377–398.
Esmali, A., Golshan, M., & Kavian, A. (2020). Investigating the performance of SWAT and IHACRES in simulation streamflow under different climatic regions in Iran. *Atmósfera*. https://doi.org/10.20937/ATM.52740.
Grayson, R., Blöschl, G., Western, A. W., & McMahon, T. A. (2002). Advances in the use of observed spatial patterns of catchment hydrological response. *Advances in Water Resources, 25*(8), 1313–1334.
Haque, M., Rahman, A., Hagare, D., & Kibria, G. (2014). Parameter uncertainty of the AWBM model when applied to an ungauged catchment. *Hydrological Processes, 29*(6), 1493–1504.
Hargreaves, G. H., & Samani, Z. A. (1982). Estimation of potential evapotranspiration. *Proceedings of the American Society of Civil Engineers, 108*, 223–230.
Hargreaves, G. H., & Samani, Z. A. (1985). Reference crop evapotranspiration from temperature. *Transactions of the American Society of Agricultural Engineers, 1*, 96–99.
Jaiswal, R. K., Ali, S., & Bharti, B. (2020). Comparative evaluation of conceptual and physical rainfall–runoff models. *Applied Water Science, 10*(48), 1–24. https://doi.org/10.1007/s13201-019-1122-6.
Jakeman, A. J., & Hornberger, G. M. (1993). How much complexity is warranted in a rainfall-runoff model? *Water Resources Research, 29*(8), 2637–2649.
Jakeman, A. J., Littlewood, I. G., & Whitehead, P. G. (1990). Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. *Journal of Hydrology, 117*, 275–300.
Jarvis, A., Reuter, H. I., Nelson, A., & Guevara, E. (2008). *Hole-filled seamless SRTM data V4, International Centre for Tropical Agriculture (CIAT)*. https://srtm.csi.cgiar.org. Accessed 20 Dec 2019.
Kerkhoven, E., & Gan, T. Y. (2006). A modified ISBA surface scheme for modelling the hydrology of Athabasca River Basin with GCM-scale data. *Advanced in Water Resources, 29*, 808–826.
Kouwen, N. (1988). WATFLOOD. A micro-computer based flood forecasting system based on real-time weather radar. *Canadian Water Resources Journal, 13*, 62–77.
Kouwen, N. (2000). *WATFLOOD/SPL9: Hydrological model and flood forecasting system*. Waterloo: Department of Civil Engineering, University of Waterloo.
Kouwen, N., & Mousavi, S. F. (2002). WATFLOOD/SPL9 hydrological model & flood forecasting system. In: Singh, V. P. and Frevert, D.K., (Ed.), Mathematical models of large watershed hydrology, pp. 649–685.
Kumar, A., Singh, R., Jena, P. P., Chatterjee, C., & Mishra, A. (2015). Identification of the best multi-model combination for simulating river discharge. *Journal of Hydrology, 525*, 313–325.
Kunnath-Poovakka, A., & Eldho, T. I. (2019). A comparative study of conceptual rainfall-runoff models GR4J, AWBM and Sacramento at catchments in the upper Godavari river basin, India. *Journal of Earth System Sciences*, *128*(33), 1–15. https://doi.org/10.1007/s12040-018-1055-8.
Legesse, D., Vallet-Coulomb, C., & Gasse, F. (2003). Hydrological response of a catchment to climate and land use changes in Tropical Africa: case study south central Ethiopia. *Journal of Hydrology, 275*(1–2), 67–85.
Legesse, D., Vallet-Coulomb, C., & Gasse, F. (2004). 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, 18*(3), 487–504.
Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology, 10*(3), 282–290.
Nielsen, S. A., & Hansen, E. (1973). Numerical simulation of the rainfall-runoff process on a daily basis. *Nordic Hydrology*, *4*(3), 171–190.
O'Connell, P. E., Nash, J. E., & Farrel, J. P. (1970). River flow forecasting through conceptual models. Part 2. The Brosna catchment at Ferbane. *Journal of Hydrology, 10*, 317–329.
Onyutha, C. (2016). Influence of hydrological model selection on simulation of moderate and extreme flow events: A case study of the Blue Nile basin. *Advances in Meteorology*, *2016*, 7148326. https://doi.org/10.1155/2016/7148326.
Onyutha, C. (2017). On rigorous drought assessment using daily time scale: Non-stationary frequency analyses, revisited concepts, and a new method to yield non-parametric indices. *Hydrology, 4*(4), 48. https://doi.org/10.3390/hydrology4040048.

Onyutha, C. (2019). Hydrological model supported by a step-wise calibration against sub-flows and validation of extreme flow events. *Water*, *11*(2), 244. https://doi.org/10.3390/w11020244.

Onyutha, C. (2020). Analyses of rainfall extremes in East Africa based on observations from rain gauges and climate change simulations by CORDEX RCMs. *Climate Dynamics*. https://doi.org/10.1007/s00382-020-05264-9.

Pirnia, P., Darabi, H., Choubin, B., Omidvar, E., Onyutha, C., & Haghighi, A. T. (2019). Contribution of climatic variability and human activities to stream flow changes in the Haraz River basin, northern Iran. *Journal of Hydro-environment Research, 25*, 12–24.

Podger, G. (2004). *User Guide—Rainfall Runoff Library*. Australia: Cooperative Research Centre for Catchment Hydrology.

Porter, J. W., & McMahon, T. A. (1972). A model for the simulation of streamflow data from climatic records. *Journal of Hydrology, 13*, 297–324.

Reuter, H. I., Nelson, A., & Jarvis, A. (2007). An evaluation of void filling interpolation methods for SRTM data. *International Journal of Geographical Information Sciences, 21*, 983–1008.

Rojas, R., Feyen, L., & Dassargues, A. (2008). Conceptual model uncertainty in groundwater modeling: combining generalized likelihood uncertainty estimation and Bayesian model averaging. *Water Resources Research, 44*, 1–16.

Rutkowska, A., Willems, P., Onyutha, C., & Mlocek, W. (2017). Temporal and spatial variability of extreme river flow quantiles in the Upper Vistula River basin Poland. *Hydrological Processes, 31*(7), 1510–1526.

Sheffield, J., Goteti, G., & Wood, E. F. (2006). Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *Journal of Climate, 19*(13), 3088–3111.

Staudinger, M., Stahl, K., Seibert, J., Clark, M. P., & Tallaksen, L. M. (2011). Comparison of hydrological model structures based on recession and low flow simulations. *Hydrology and Earth System Sciences, 15*, 3447–3459.

Sugawara, M., & Fuyuki, F. A. (1956). *Method of revision of River discharge by means of a Rainfall model*. Saitama: The Geosphere Research Institute of Saitama University.

Therrien, R., McLaren, R. G., Sudicky, E. A., & Panday, S. M. (2005). *HydroGeoSphere: A three-dimensional numerical model describing fully-integrated subsurface and surface flow and solute transport* (p. 322). Waterloo: Groundwater Simulations Group, University of Waterloo.

Twedt, T. M., Schaake, J. C., & Peck, E. L. (1977). National weather service extended streamflow prediction. In: Proc. 45th Western Snow Conference, Albuquerque, New Mexico, pp. 52–57. https://westernsnowconference.org/sites/westernsnowconference.org/PDFs/1977Twedt.pdf. Accessed 19 Apr 2020.

Wang, J., Yang, H., Li, L., Gourley, J. J., Sadiq, I. K., Yilmaz, K. K., et al. (2011). The coupled routing and excess storage (CREST) distributed hydrological model. *Hydrological Sciences Journal, 56*(1), 84–98.

Yang, Q., Zhang, H., Wang, G., Luo, S., Chen, D., Peng, W., et al. (2019). Dynamic runoff simulation in a changing environment: A data stream approach. *Environmental Modelling and Software, 112*, 157–165.

Yu, B., & Zhu, Z. (2015). A comparative assessment of AWBM and SimHyd for forested watersheds. *Hydrological Sciences Journal, 60*(7–8), 1200–1212.

Zhang, Q., Liu, J., Singh, V. P., Gu, X., & Chen, X. (2016). Evaluation of impacts of climate change and human activities on streamflow in the Poyang Lake basin, China. *Hydrological Processes, 30*(14), 2562–2576.