# Physically-Based Streamflow Predictions in Ungauged Basin with Semi-Arid Climate

**Dmitry Kozlov and Anghesom Ghebrehiwot**

**Abstract** Rainfall-runoff models that utilize reanalysis datasets as driving variables have been widely applied for generating hydrological responses in data-sparse regions. Apparently, there are various requirements that affect the choice of a particular method of hydrologic investigation. In the present study, Soil and Water Assessment Tool (SWAT) and precipitation-runoff (MIKE 11-NAM) models were selected to simulate streamflows from a small watershed with semi-arid climate, using Climate Forecast System Reanalysis (CFSR) as driving inputs. As such, models that provide reliable streamflow predictions from regions with similar climate settings, whose errors and uncertainties are within acceptable ranges, can be identified. The main characteristics of performance criteria indicate that the SWAT model relatively outperform the MIKE 11-NAM model. However, while most of the statistical evaluations prove the acceptable performance of the SWAT model, broad range of prediction uncertainties during calibration and validation were also reflected. Among the possible sources of errors, errors due to forcing data are most likely to be accounted for the unsatisfactory portions of both models. Therefore, to minimize model uncertainty and thereupon improve its performance, in-situ data collection need to be incontestably boosted up. The study also highlights the need for further investigation on the possible mechanisms of proper application of CFSR that avoid erroneous streamflow predictions from similar regions.

**Keywords** Eritrea · MIKE 11-NAM · Rainfall-runoff · Reanalysis datasets · Sensitivity analysis · Simulation · SWAT model

## 1 Introduction

Physically-based mathematical models are used to analyse and predict hydrological and biogeochemical processes within river catchments, including the flow of water, sediment, chemicals, nutrients, and microbial organisms, within watersheds, as well as quantify the impact of human activities on these processes. Such models

D. Kozlov · A. Ghebrehiwot (✉)
Moscow State University of Civil Engineering, Yaroslavskoe shosse, 26, Moscow, Russia 129337

have been useful tools in underpinning our understanding about the dynamic interactions between climate and land surface hydrology [1–3] and providing the missing information as a basis for decision-making. As such, a broad spectrum of critical environmental and water resources problems have been addressed with the support of physically-based mathematical models. Nevertheless, a bulk of evidence demonstrates that there are limitations in existing models due to lack of full representation of the complex hydrologic systems and spatio-temporal variability of the hydrological and meteorological components. As a result, recent studies [3–5] emphasized the need for the development of watershed models that make use of the widest possible information available and underpin the current practices of sustainable management of river basins, as well as new techniques that integrate economic, social and environmental perspectives.

Rainfall-runoff models are one of the extensively applied predictive tools for generating hydrological responses. Whichever rainfall-runoff model we select for whatsoever purpose it may be, it remains to be only an approximate representation of the real processes. Despite the efforts put in to overcome the fundamental problem of extensive difference in the spatial and time scales of hydrological models through the application of downscaling of model outputs, selection of an appropriate hydrologic model is yet one of the critical issues. Conspicuously, the effectiveness of a model largely depends on availability and reliability of historical ground information; the efficacy of a model is lower and more uncertain in ungauged regions and vice-versa. At global scale, river basins in many parts of the world are not only ungauged but also experience a significant reduction in the ground hydrometric networks [3, 6–9]. Regional studies in East Africa[1] [10] have shown that the political and socio-economic situation in this part of the African continent has not been conducive to conventional hydrological data collection. Moreover, these problems are enormously exacerbated by the consequences of anthropogenic and climatic changes. To overcome the gap in shortage of data from conventional hydrometric networks, numerous researchers in the field [11–14] have investigated the application of remote sensing-based information.

The use of global climate reanalysis datasets for modelling streamflow has shown that the effectiveness of the model depends on the source and resolution of the input datasets and climate of the region of interest. For example, the CFSR of the National Centres for Environmental Prediction (USA) and ERA-Interim were used to model daily and monthly streamflows using the SWAT model of a river basin, located in the Sudan-Sahel region [15]. They found that the ERA-Interim datasets generated better results compared to the former. Similarly, the use of SWAT and MIKE 11-NAM in the conditions of South Africa [16] and Eritrea [11] showed low statistical representativeness between precipitation data from the CFSR and field rainfall measurements, as well as overall water imbalance. An assessment of the applicability of CSFR for modelling hydrological processes within the boundaries of five river basins with different hydrological and climatic conditions in Ethiopia

---

[1] WMO and GWP, *Integrated Drought Management Programme Handbook of Drought Indicators and Indices*, no. 1173. 2016.

and the United States was carried out [13]. They found that the use of input variables from the CFSR provides modelling of streamflow as good as the outputs from the use of model inputs from ground-based weather stations. Thus, despite the fact that conventional in-situ hydrometric data remain the most accurate and reliable sources of input information, the use of reanalysis datasets as alternative source for modelling runoff in ungauged or poorly gauged river basins has been proposed. One of the most sophisticated and widely used models that make use of reanalysis datasets is SWAT model. It is "a conceptual, continuous-time model developed to assist water resource managers in assessing water supplies and non-point source pollution on watersheds and large river basins" [17] and operates at a daily time step. The SWAT model has got worldwide recognition. For example, SWAT and global climate models were used to study the formation of streamflow in Russia [18], the United States [19], the hydrological situation of Africa [20], including the impact of climate change on the availability of fresh water on the African continent [21]. But, certain shortcomings of SWAT model were noted [22], especially in terms of comparing the simulation results with long-term in-situ data on daily runoff and/or discharge of pollutants.

The SWAT model is complex with semi-distributed parameters, so its use requires a large amount of input data, which makes it difficult to parameterize and calibrate. For the SWAT model, special computational algorithms were created, which are based on the method of multidimensional mathematical optimization, including the SWAT-CUP software module [23, 24]. SWAT-CUP is designed for the purpose of auto-calibration and uncertainty analysis for the SWAT model and combines five different optimization algorithms: sequential analysis of all possible sources of uncertainty SUFI-2, the genetic algorithm of swarm intelligence (PSO), as well as methods of general probabilistic uncertainty estimation (GLUE), parametric solution (ParaSol) and Monte Carlo with Markov chains (MCMC), which allows us to use various objective functions and criteria. The advantage of SWAT-CUP is that it combines several calibration and uncertainty analysis procedures into a single interface, making the model calibration procedure more understandable and faster.[2] Despite the fact that the SUFI-2 algorithm was quite effective for large-scale models, the equifinality problem is still one of the most acute for the calibration of the parameters of the hydrological model [25].

River basins in Eritrea are characterised by the spatial and temporal variability of climate and geophysical characteristics, landuse and climate changes. The majority of river basins do not have regular observation network, or characterized by a lack of high-quality field data. Under such circumstances, the development of models and schemes for water management planning remains to be a complex task [26]. A recent survey shows that the management, collection and processing of hydrometric networks at national level are declining. In the contrary, there are a lot of ongoing nation-wide water resources related development projects [26, 27], including the construction of reservoirs, diversion structures, expansion of agriculture and settlements. Thus, given the lack of high-quality field observational data, on the one hand,

---

[2] SWAT-CUP 2012: SWAT Calibration and Uncertainty Programs - A User Manual," *Sci. Technol.*, 2014.

and the ongoing intensive water management activities, on the other, the applicability of satellite-based climate data becomes a timely, important and urgent task in the region. As such, we recently evaluated the applicability of the conceptual model "precipitation-runoff" (MIKE 11-NAM) for streamflow simulations from the Mereb-Gash river basin using CFSR [11]. Nevertheless, the results failed to satisfy the model acceptance criterion and accordingly three tasks were suggested as future courses of action: the transformation of CFSR information into a more realistic one; the evaluation of reanalysis data from other sources at different time scales and resolutions; and the study of the effectiveness of other software systems. Therefore, the objectives of the study were as follows: (i) to use the SWAT to establish a hydrological model of the Debarwa subbasin in the upper reaches of the Mereb-Gash River basin with a monthly estimated time interval; and (ii) to evaluate the effectiveness of the SWAT in comparison with MIKE 11-NAM for the purposes of modelling streamflow in the conditions of the specified subbasin.

## 2 Materials and Methods

The major work of the study required physically-based SWAT and MIKE 11-NAM models establishment in order to evaluate the various water balance components at subbasin level and monthly time intervals. The latter model's setup, details of the procedures and working principles can be referred to the authors' recent publication [11]. Thus, the discussion of this section mainly focused on the SWAT model. These models were predominantly established using freely available data. Other than the topographic, soil and landuse data, SWAT requires climatic data at daily or sub-daily time steps. Major input data for SWAT include digital elevation model (DEM), landuse, soil properties, and daily weather data. These data were complemented by additional sources, provided by the Ministry of Land, Water and Environment, Department of Water Resources, Eritrea. Finally, the findings of both models were evaluated and intercompared using different statistical evaluation techniques, which is presented in the ensuing section. Generally, the modelling procedures include model setup, calibration, uncertainty analysis, sensitivity analysis, validation, and analyses such as climate change, best management practices, risk analysis, etc.

Debarwa is a small watershed in the upper reaches of Mereb-Gash basin. The location, landuse and other hydrologic features of the subbasin are depicted in Fig. 1. The total area of the catchment is approximately 200 km$^2$, with an altitudinal range from 1905 to 2550 m above msl. It is a mountainous (50% of area has a slope greater than 10%) covered by sparse shrubs and agriculture. The soil type in the area is dominated by Eutric Nitosols of clay soils, followed by Humic Cambisols of clay-loam. Both soil categories fall under the third hydrologic group (C); that is, the soils have a slow infiltration and water transmission rates when thoroughly wetted, as well as a layer that impedes downward movement of water or have moderately fine to fine texture. The Debarwa watershed lies in moist highlands zone where temperature varies from 0 °C to 32 °C and an average annual rainfall of 547 mm. Climate in
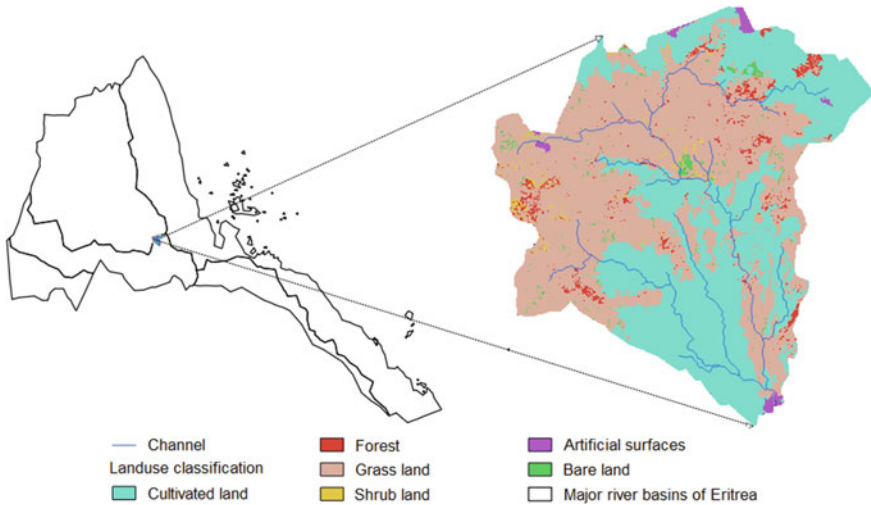
**Fig. 1** Location and landuse maps of the study area

the catchment can be characterized as moderate with December-January being the coldest and March–April the hottest. Maximum precipitation occurs in the summer season, specifically in the months of July and August with a monthly mean rainfall of 185 mm and 175 mm, respectively. The watershed has one global weather station and one flow measuring station at the outlet.

In the SWAT program, a watershed is divided into multiple subbasins, which are then further subdivided into hydrologic response units (HRUs). HRUs are defined as "lands with similar spatial datasets, namely topography, landuse, and soil types and all the components of the soil water balance could be determined on an HRU basis, with the assumption that similar HRUs would have similar hydrologic characteristics" [24]. Simulation of watershed hydrology is done in the land phase, which controls the amount of water, sediment, nutrient, and pesticide loadings to the main channel in each subbasin, and in the routing phase, which is the movement of water, sediments, etc., through the streams of the subbasins to the outlets. Besides, the soil processes include lateral flow from the soil, return flow from shallow aquifers, and tile drainage, which transfers water to the river; shallow aquifer recharge, and capillary rise from shallow aquifer into the root zone, and finally deep aquifer recharge, which removes water from the system. The climate-driven hydrological cycle provides moisture and energy inputs. In this regard, global CFSR data from the National Centre for Atmospheric Research (USA) was utilized. The QSWAT 2012 interface was used to set up and parameterize the model. On the basis of DEM and the stream network, a threshold drainage area of 3 $km^2$ was chosen to discretize the watershed into 13 subbasins, which were further subdivided into 61 HRUs based on soil, landuse, and slope. Each HRU is normally thought to be a uniform unit where water balance calculations are made. A schematic representation of the model setup is shown in Fig. 2. The climate
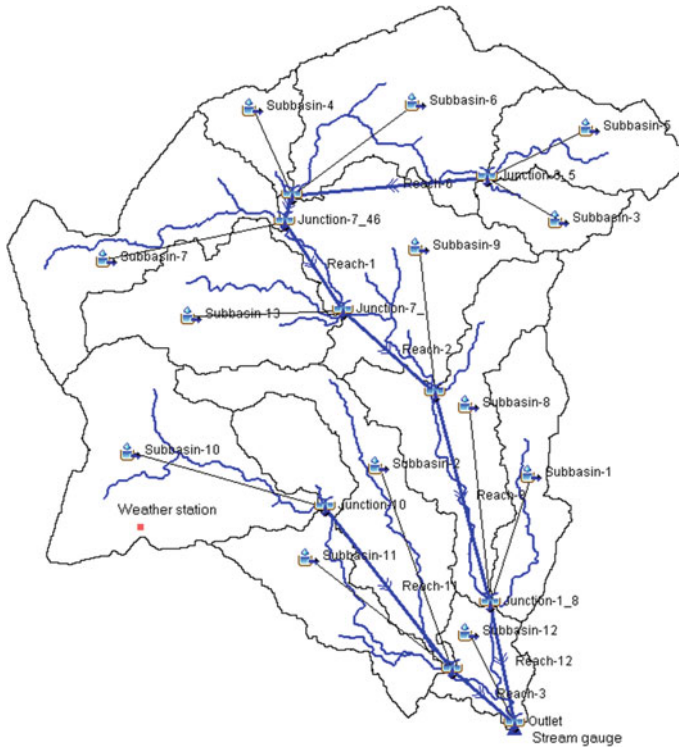
**Fig. 2** Schematic representation of the Debarwa subbasin

variables based on the CFSR as well as data on daily and monthly water consumption in the catchment area were entered into the model as input information. For the estimated time period of the simulation, the interval from 1994 to 2010 was considered. Approximately two–thirds of the data was used for calibration and the rest for validation. The initial and final runs were performed using SUFI-2. The calculations didn't consider point and distributed sources of pollution, bottom sediments, nitrogen and phosphorus loads, reservoir regulations, and the spatial variability of some other parameters.

The initial selection of parameters depends on the behaviour of the initial model result before any calibration. The SWAT-CUP program has the provision of ten independently evaluated objective functions and an additional multi-objective function—a combination of two or more objective functions. As has been clearly articulated in various literatures [23, 25], the outputs corresponding to each objective function are normally unique, leading to the conditionality of objective functions. As such, multi-objective function has been suggested to overcome the problem of conditionality. On the other hand, model uncertainty could be minimized if and only if we clearly identify the sources of uncertainty. Possible sources of uncertainty in hydrologic modelling [11, 23, 24, 29] can be categorized as follows: (i) model input data;

(ii) model assumptions and simplifications; (iii) the science underlying the model; (iv) stochastic uncertainty also known as variability; and (v) code uncertainty, such as numerical approximations and undetected software errors. It would be unrealistic to expect a perfect model performance at the end because of the aforementioned sources of errors as well as many activities that occur in the watershed.

Successful application of hydrological models largely depends on the calibration and sensitivity analysis of the parameters [23, 28]. Calibration and validation procedures are effectively used only with field observations. The information about the measured daily or monthly streamflow data is important for these procedures. SUFI-2 in the SWAT-CUP module [30] was employed for calibration and validation procedures. The SUFI-2 algorithm covers a wide range of parameter uncertainties at the beginning of calculations, as a result of which the observational data initially falls into the 95% uncertainty forecast (95PPU-confidence probability). 95PPU is the interval between 2.5% and 97.5% of the total distribution of the output simulated variable (water flow) obtained using an efficient Latin hypercube sampling algorithm, excluding 5% of the worst simulations [30]. Then, with each iterative step, the uncertainty interval narrows, and simultaneously two indices are checked that determine the degree of agreement and uncertainty of the model: the *P*-factor (the percentage of measurement results that fall into the 95PPU), ranging from 0 to 1, and the *R*-factor (the ratio of the average width of the 95PPU interval to the standard deviation of the corresponding measured value). In an ideal situation, when the simulation results are exactly (100%) consistent with the observational data, the *P*-factor is 1. A *P*-factor value of 0.70 or higher is considered sufficient for the results of streamflow modelling. The *P*-factor and *R*-factor of 1 are iterations that exactly match the measurement results. The desired value of the *R*-factor, determined by Eq. (1), is considered as acceptable if its value is less than 1.50 [30].

$$R - factor_j = \frac{\frac{1}{n_j} \sum_{t_i=1}^{n_j} \left( x_s^{t_i,97.5\%} - x_s^{t_i,2.5\%} \right)}{\sigma_{oj}} \tag{1}$$

where $x_s^{t_i,97.5\%}$ and $x_s^{t_i,2.5\%}$ are the upper and lower boundary of the 95PPU at time-step $t$ and simulation $i$, $n_j$—the number of data points, and $\sigma_{oj}$—the standard deviation of the $j$th observed variable.

As has been discussed, the SUFI-2 optimization algorithm allows the use of various objective functions, out of which the Nash and Sutcliffe efficiency (*NS*) was used (*NS* = 1.0 being optimal value and $0.75 < NS \leq 1$ being acceptable). In addition, the coefficient of determination ($0.70 < R^2 < 1.0$), modified coefficient of determination ($bR^2$), per cent bias ($PBIAS < \pm 25$), and ratio of the root mean squared error to the standard deviation of measured data ($RSR \leq 0.6$), whose corresponding equations are represented by Eqs. (2–6), were also additional criteria for statistical model evaluations.

$$R^2 = \frac{\left[\sum_i \left(Q_{m,i} - \overline{Q}_m\right)\left(Q_{s,i} - \overline{Q}_s\right)\right]^2}{\sum_i \left(Q_{m,i} - \overline{Q}_m\right)^2 \sum_i \left(Q_{s,i} - \overline{Q}_s\right)^2} \tag{2}$$

$$bR^2 = \left\{ \begin{array}{l} |b|R^2 \ if \ |b| \leq 1 \\ |b|^{-1}R^2 \ if \ |b| > 1 \end{array} \right\} \tag{3}$$

$$NS = 1 - \frac{\sum_i \left(Q_m - Q_s\right)_i^2}{\sum_i \left(Q_{m,i} - \overline{Q}_m\right)^2} \tag{4}$$

$$PBIAS = 100 \times \frac{\sum_i^n \left(Q_m - Q_s\right)_i}{\sum_i^n Q_{m,i}} \tag{5}$$

$$RSR = \sqrt{\sum_{i=1}^n \left(Q_m - Q_s\right)_i^2} \Big/ \sqrt{\sum_{i=1}^n \left(Q_{m,i} - \overline{Q}_m\right)^2} \tag{6}$$

where $Q$—a variable (e.g., discharge); $m$ and $s$—stand for observed and simulated variables; $b$—slope of the regression line between the observed and simulated variables; and $i$—the $i$th observed or simulated data.

## 3 Results and Discussion

Rainfall and corresponding simulated daily streamflow from SWAT model prior to calibration in SWAT-CUP program, as well as observed streamflow at the outlet of the watershed was analysed and evaluated. As such, absolute overlapping in the seasonality of rainfall and corresponding simulated and observed streamflows were noticed; a large amount of rainfall produced high flows and vice versa. However, a considerable quantitative mismatch between the simulated and observed streamflows ($R^2 = 0.10$) was realized at this stage. This disparity was in fact a signal that our calibration may not yield a perfect fit by all means possible.

During parameterization process, SWAT-CUP provides two different methods of sensitivity analysis: one-at-a-time and global. In this study, the latter method was applied, where all selected parameters change at a time and uses multi-regression computation. The SUFI-2 program permits up to 1000 iterations for one complete iterative run. The global sensitivity uses the $P$-value and $t$-stats for analysing the sensitivity of selected parameters to prioritize them; large $t$-stat and lower $P$-value indicate higher parameter sensitivity and vice versa. The study area is a watershed characterized as ungauged or poorly gauged with a limited in-situ hydrometric data. Besides, SWAT contains a large number of variable parameters involved in the calibration process. In such conditions, calibration of all parameters causes great difficulties. Therefore, first we need to select the most significant parameters, which are thought to represent the hydrological processes, for the calibration procedure. To this end, the sensitivity analysis of randomly selected 15 parameters was carried out within

the SUFI-2 procedure (Table 1), out of which those that have the greatest influence on the formation of streamflow in the study area are identified. After a series of tests in SWAT-CUP, it was found that the top most sensitive parameters include CN, SHALLST, and RCHRG_DP.

Considering the dynamics and radical uncertainty of daily flows, calibration was limited to monthly flows. Accordingly, the performance of the best parameter sets chosen during the sensitivity analysis was evaluated by two statistical evaluations:

**Table 1** Parameter sensitivity analysis and calibrated monthly streamflow values

| Parameter | Description | Sensitivity | | Simulated values | | |
|---|---|---|---|---|---|---|
| | | $t$-stat | $P$-value | Fitted | Min | Max |
| CN2* | Soil conservation service (SCS) curve number for moisture condition II | −32.53 | 0.00 | −0.29 | −0.30 | 0.10 |
| SHALLST | Initial depth of water in the shallow aquifer (mm of $H_2O$) | −7.69 | 0.00 | 3308 | 1000 | 5000 |
| RCHRG_DP | Deep aquifer percolation fraction | −5.62 | 0.00 | 0.06 | 0.00 | 0.80 |
| ALPHA_BF | Baseflow alpha factor (1/days) | −3.44 | 0.00 | 0.30 | 0.00 | 0.50 |
| EPCO | Plant uptake compensation factor | −1.83 | 0.07 | 0.90 | 0.30 | 1.00 |
| CH_N2 | Manning's "$n$" value for the main channel | −1.30 | 0.20 | 0.24 | −0.01 | 0.30 |
| SURLAG | Surface runoff lag coefficient | −0.21 | 0.84 | 21.46 | 6.00 | 24.00 |
| REVAPMN | Threshold depth of water in the shallow aquifer for "*revap*" or percolation to the deep aquifer to occur (mm of $H_2O$) | 0.69 | 0.49 | 355.6 | 0.00 | 400 |
| OV_N.hru | Manning's "$n$" value for overland flow | 0.82 | 0.41 | 0.54 | 0.10 | 1.00 |
| GW_DELAY | Groundwater delay time (days) | 1.49 | 0.14 | 468.85 | 150 | 500 |
| CH_K2 | Effective hydraulic conductivity in main channel alluvium (mm/hour) | 1.55 | 0.12 | 190.30 | 100 | 400 |
| FFCB | Initial soil water storage expressed as a fraction of field capacity water content | 2.22 | 0.03 | 0.38 | 0.20 | 1.00 |

(continued)

**Table 1** (continued)

| Parameter | Description | Sensitivity | | Simulated values | | |
|---|---|---|---|---|---|---|
| | | *t*-stat | *P*-value | Fitted | Min | Max |
| ESCO | Soil evaporation compensation factor | 2.40 | 0.02 | 0.25 | 0.20 | 1.00 |
| GW_REVAP | Groundwater "*revap*" coefficient | 4.73 | 0.00 | 0.09 | 0.02 | 0.20 |
| GWQMN | Threshold depth of water in the shallow aquifer required for return flow to occur (mm of $H_2O$) | 6.58 | 0.00 | 4982.50 | 1500 | 5000 |

*The change is relative whereas the change in all other parameters is replacement with other value

(i) model prediction uncertainty and (ii) model performance evaluation. Uncertainty analysis refers to the propagation of all model input uncertainties to model outputs, which stem from the lack of knowledge of physical model inputs to model parameters and model structure. Identification of all acceptable model solutions in the face of all input uncertainties can provide us with model uncertainty in SWAT-CUP as 95PPU. Once the model is parameterized and the ranges are assigned, the model is normally run some 300–1000 times [23]. After all simulations are completed, the provision of post-processing option in SWAT-CUP calculates the objective function and the 95PPU for all observed variables in the objective function. The prediction uncertainty, which is represented by the shaded regions for the calibration (Fig. 3) and validation (Fig. 4) processes, is expressed by the 95PPU in SUFI-2. As a result, *P*-factor values were estimated to be 0.34 and 0.43 for calibration and validation, respectively (Table 2). In other words, only 34% and 43% of the observed streamflows are bounded by the 95PPU during calibration (1997–2001 and 2007–2010) and validation periods (2002–2006), respectively. On the other hand, the *R*-factor values are also equal to 2.56 and 3.48 for calibration and validation periods, respectively (Table 2). The
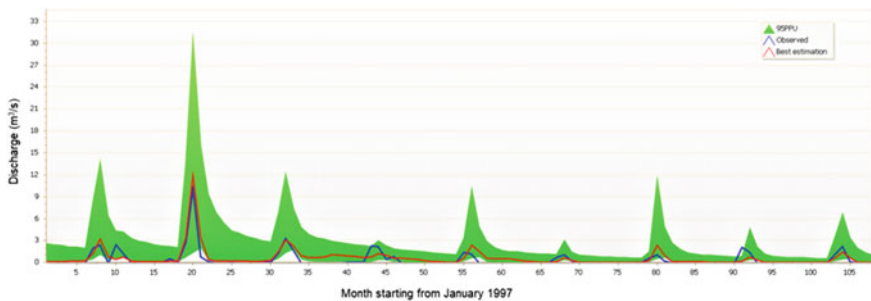


**Fig. 3** Comparison of observed and simulated monthly streamflows during calibration period (1997–2001 and 2007–2010)
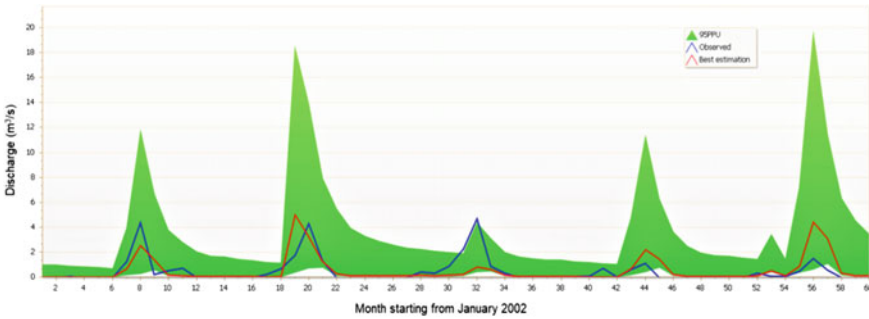
**Fig. 4** Comparison of observed and simulated monthly streamflows during validation period (2002–2006)

**Table 2** Summary of statistics for calibration and validation procedures

| Process | Uncertainty prediction | | Objective function | | | | |
|---|---|---|---|---|---|---|---|
| | *P*-factor | *R*-factor | *R2* | *bR2* | *NS* | *PBIAS* | *RSR* |
| Calibration | 0.34 | 2.56 | 0.80 | 0.79 | 0.73 | −42.0 | 0.52 |
| Validation | 0.43 | 3.48 | 0.32 | 0.18 | 0.12 | −9.8 | 0.94 |

calibrated and validated values of *P*-factor and *R*-factor are clearly outside of the recommended ranges [30], i.e., *P*-factor > 0.70 and *R*-factor < 1.50.

Five model performance indicators were employed, out of which *NS* was used as the major objective function as has been described above. The other four performance indices include $R^2$, b$R^2$, *PBIAS*, and *RSR*. Results as tabulated in Table 2 clearly show that all the performance indicators for the calibration period ($R^2$, b$R^2$, and *NS* > 0.70, and *RSR* < 0.60) are in fairly acceptable ranges. In other words, the statistical indices indicate that there is a good agreement between the observed and simulated streamflows. On the contrary, the corresponding model performance indicators for validation ($R^2$ and b$R^2$ < 0.40, *NS* < 0.50, and *RSR* > 0.70) are evaluated as unsatisfactory. *PBIAS* measures the average tendency of the simulated data to be larger or smaller than their observed counterparts. Positive values represent model underestimation bias and negative values indicate model overestimation bias [31]. So, *PBIAS* values-based model performance during calibration could be evaluated as unsatisfactory (*PBIAS* > ± 25), whereas that of validation is evaluated as acceptable (*PBIAS* < ± 10). *PBIAS*-values show model overestimation by 42% and 9.8% during calibration and validation, respectively.

To understand the issue of conditionality, an investigation on the effect of objective function choice on the model performance was explored by running SUFI-2 post-processing alone. This procedure does not require the running of the SWAT model again. Accordingly, three objective functions were tested, namely *NS*, *PBIAS* and $R^2$ against other indicators. The graphical visualization (Fig. 5) and model performance indicators (Table 3) clearly illustrate how the choice of objective function affects the
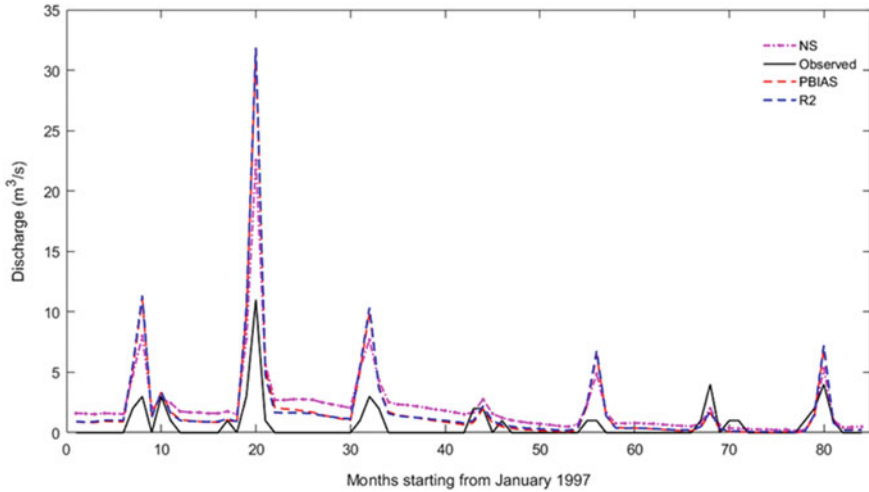
**Fig. 5** Effect of objective function selection on calibration solutions

**Table 3** Effect of choice of objective function on calibration solutions

| Type of function | Uncertainty prediction | | Objective function | | | | |
|---|---|---|---|---|---|---|---|
| | $P$-factor | $R$-factor | $R^2$ | $NS$ | $bR^2$ | $PBIAS$ | $RSR$ |
| $NS$ | 0.21 | 2.03 | 0.71 | -1.41 | 0.44 | -225.4 | 1.55 |
| $PBIAS$ | 0.05 | 0.48 | 0.76 | -3.14 | 0.33 | -185.39 | -0.58 |
| $R2$ | 0.05 | 0.35 | 0.76 | -3.31 | 0.32 | -193.37 | -0.61 |

calibration solution. While each objective function produced unique solutions, which was also reported by many researchers [23, 25], overestimation of simulated flows, especially peak flow and baseflow, could be clearly detected in all of the outputs in this particular case.

In the preceding section, we realized that overall performance of the SWAT model, verified by the use of statistical evaluations, was unsatisfactory. Unsatisfactory performance of the SWAT model was specifically magnified during the analysis of model prediction uncertainty in calibration and validation processes (Table 2). At this stage, it was necessary to think of possible sources of errors and uncertainties. Accordingly, we arrived at the conclusion that errors due to input climate data (e.g., precipitation) had considerable influence on the unacceptable model outputs. Because, considerable overestimation of the CFSR-based precipitation as compared to field observations had been reported in the authors' recent works [10, 11]. This situation directed us to compare the outputs from physically process-based distributed SWAT with a semi-distributed MIKE 11-NAM so as to come up with a model with relatively better performance. While the former is discussed in the preceding sections, the latter's analyses are briefly discussed in the ensuing paragraph.

MIKE 11-NAM model has less number (9) of basic parameters than that of SWAT model. The list of these parameters, their descriptions, lower and upper limits and fitted values during calibration are presented in Table 4. The fitted values are the optimal values that were obtained through iterative process and manual and automatic calibrations. Having seen these values, we were able to realize that some of them are far beyond our realistic expectations (e.g., runoff coefficient, baseflow, etc.). Because, Debarwa catchment is characterised by mountainous, low infiltration rate as a result of poor soil conditions and vegetation cover. Besides, it remains dry for much of the year due to its ephemeral nature. During rainy days, the watershed experiences flash floods [26] with short durations of flows (time to peak, time base, time lag) and lower or almost zero baseflows. Thus, a runoff coefficient of 0.10 and high values of baseflows, in some cases, are deemed to be quite irrelevant. At this point in time, it is very difficult to verify the other fitted values owing to the absence of field data.

The intercomparison between simulated monthly streamflows of SWAT and MIKE 11-NAM models, as well as observed flows for calibration (Fig. 6 and Fig. 7) and validation (Fig. 8), respectively, were analysed. Moreover, these outputs were evaluated using various objective functions whose values for calibration and validation are summarized in Tables 5 and 6, respectively. All of the performance indicators discernibly show that MIKE 11-NAM is far less satisfactory; the statistical indictors

**Table 4** MIKE 11-NAM model basic parameters for calibration and validation procedures

| Parameter | Unit | Description | Fitted | Lower bound | Upper bound |
|---|---|---|---|---|---|
| $U_{max}$ | mm | Upper limit of the amount of water in the surface storage, representing interception, depression, and surface storages | 20 | 10 | 20 |
| $L_{max}$ | mm | Maximum water content in the lower zone storage, representing the soil moisture below the surface from which plants draw water for transpiration | 300 | 100 | 300 |
| $CQOF$ | – | Overland flow runoff coefficient that determines the distribution of excess rainfall into overland flow and infiltration | 0.10 | 0.10 | 1.00 |
| $CK_{IF}$ | hour | Time constant for interflow from the surface storage | 967.46 | 500 | 1000 |
| $CK_{12}$ | hour | Time constant for overland flow and interflow routing, routed through two linear reservoirs in series | 44.13 | 10 | 50 |
| $TOF$ | – | Threshold values for overland flow | 0 | 0 | 0.99 |
| $TIF$ | – | Threshold values for interflow | 0 | 0 | 0.99 |
| $TG$ | – | Threshold values for groundwater recharge | 0.98 | 0 | 0.99 |
| $CK_{BF}$ | hour | Time constant for routing baseflow | 4000 | 1000 | 4000 |
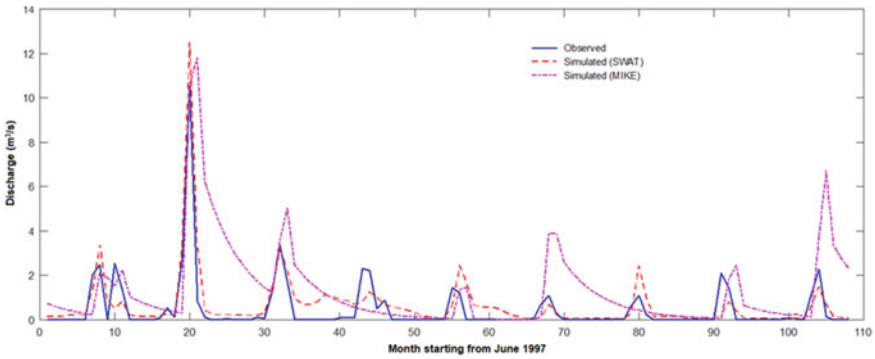
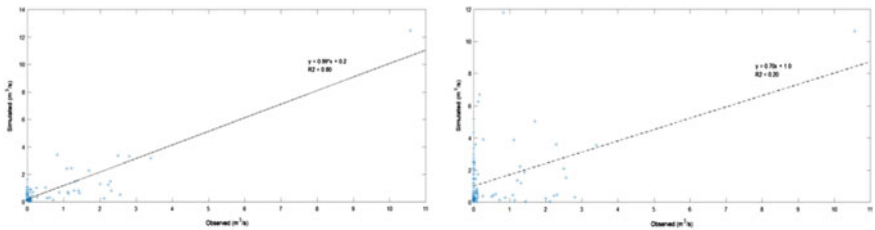**Fig. 6** Comparison of observed and simulated monthly streamflows during calibration



**Fig. 7** Correlation between observed and simulated monthly streamflows during calibration: SWAT (left) and MIKE 11-NAM (right)
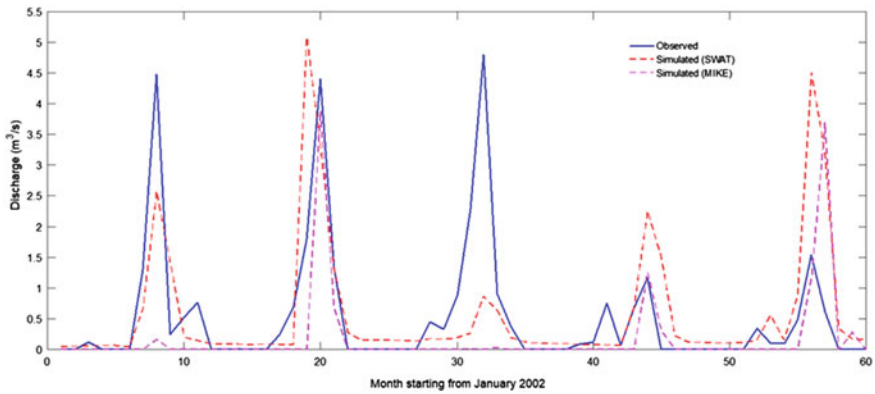


**Fig. 8** Comparison of observed and simulated monthly streamflows during validation

**Table 5** Performance evaluation of selected models during calibration period

| Type of model | Objective function | | | | |
|---|---|---|---|---|---|
| | $R^2$ | $bR^2$ | NS | RSR | PBIAS |
| SWAT | 0.80 | 0.77 | 0.73 | 0.52 | -42 |
| MIKE 11-NAM | 0.20 | 0.19 | -1.62 | 1.61 | 188 |

**Table 6** Performance evaluation of selected models during validation

| Type of model | Objective function | | | | |
|---|---|---|---|---|---|
| | $R^2$ | $bR^2$ | NS | RSR | PBIAS |
| SWAT | 0.32 | 0.22 | 0.12 | 0.93 | 9.80 |
| MIKE 11-NAM | 0.17 | 0.05 | -0.01 | 1.00 | -64.10 |

are less than the allowable ranges and the visual graphical comparisons of observed and simulated do not fairly coincide. In addition, Fig. 7 shows a better correlation between observed and simulated streamflows in SWAT ($R^2 = 0.80$) than MIKE 11-NAM ($R^2 = 0.20$). Therefore, based on the statistical evaluations and visual graphical comparisons, it is fair to say that the SWAT model, without forgetting the issue of uncertainty as has been described above, strikingly outperformed MIKE 11-NAM during calibration and validation procedures.

Physically-based models play an important role in obtaining hydrological and biogeochemical information in catchments that are not sufficiently studied from a hydrological point of view in arid and semi-arid regions. While some models are complex others are fairly simple. The former types of models normally require significant amounts of reference information and have a large number of parameters, whereas the latter require less reference information and have fewer parameters. The effectiveness and suitability of physically-based models for hydrological predictions in ungauged and/or poorly gauged river basins depends on numerous factors such as data availability and computational facility, knowledge and experience of the user, the type of the problem, and economics. It is understandable that a given approach will seldom satisfy all of these requirements, and consequently one approach will seldom be uniformly better than the other under all circumstances. Each model, regardless of its complexity, has its own strengths and weaknesses. A choice among approaches depends on their systematic evaluations, which, in turn, entails construction of an objective function, use of goodness-of-fit criterion, sensitivity analysis, error analysis, and comparison and ranking.

In view of the above facts, physically-based models with semi-distributed and lumped parameters, namely SWAT and MIKE 11-NAM, which are widely used for hydrological response predictions in arid and semi-arid regions, were studied. As noted earlier, to overcome the limitation of reference information, the technology of using satellite climate reanalysis datasets (e.g., CFSR) has drawn the attention of researchers in the field. However, these applications are mainly constrained by lack of in-situ data for calibration and validation procedures and significant amounts of model uncertainty. Thus, cautious application of reanalysis datasets has been suggested. SWAT model, which uses reanalysis datasets as well as other databases, which are available in the public domain as driving inputs without any modifications, was employed. To ascertain the model efficiency and identify models with acceptable uncertainty, it was necessary to intercompare with other models, out of which MIKE 11-NAM was selected. In this respect, based on the performance evaluations of both models, promising results have been achieved. However, the current approach

requires additional endeavours and verifications that ensure the required level of certainty is attained. In this regard, some possible insights have been proposed.

Sensitivity analysis shows the portion of parameters in the model output uncertainties. More sensitive parameters have a higher share of model uncertainties than less sensitive ones in the model output if that parameter is left uncalibrated. Therefore, sensitivity analysis is the first step that should be taken into consideration in model calibration. However, not all sensitive parameters may be calibrated in ungauged catchments. In this study, there were no measured parameters and hence, it is recommended that further efforts should be made to use all available data sources of the catchment under study. This helps to exclude less sensitive parameters from calibration and avoid unnecessary and arbitrary adjustments of parameters. Generally, the SWAT model uncertainty, represented by *P*-factor and *R*-factor, were found to be outside of the acceptable limits for calibration and validation periods. Thus, other approaches that intend to make CFSR and other reanalysis datasets suitable for hydrologic and environmental investigations in the region need to be investigated.

## 4   Conclusion

The choice of a particular method of hydrologic investigation depends on various requirements such as data availability, approach, economic, and the like. It is conceivable that a given approach will seldom satisfy all of these requirements. This fact was substantiated by our recent work [25] on simulation of streamflows from Mereb-Gash river basin in Eritrea using MIKE 11-NAM model and reanalysis datasets as model inputs. As such, a comparable research was necessary in order to identify the most effective and suitable models for the conditions of the region under consideration. SWAT and MIKE 11-NAM physically-based were applied to simulate streamflows from a Debarwa subbasin within the Mereb-Gash river basin. The reason for focusing at subbasin level was to reduce the accumulated errors that we had experienced when the larger river basin was considered. Findings indicated that SWAT relatively outperformed MIKE 11-NAM in terms of overall model efficiency. Nevertheless, while most of the objective functions proved the acceptable efficiency of the former model, it also reflected a lot of uncertainties during calibration and validation procedures. Yet the uncertainties due to the use of SWAT model were greatly decimated as compared to that of MIKE 11-NAM. Among the different sources of model errors, we believe, errors due to forcing data are highly likely to be accounted for lower performances. However, this does not mean that the plausible scenario that the model's performance could be influenced by other sources of errors is totally out of consideration.

Even though reanalysis datasets have apparently great advantage over in-situ observations in terms of their simplicity, the findings from this study underscored the need for critical re-examination of the former. In this respect, we would like to

suggest the following approaches. Firstly, to minimize model uncertainty and there-upon improve its performance, ground data collection systems need to be strengthened as much as possible. Secondly, further investigation on the applicability of CFSR datasets to simulate streamflows shall be carried out in the near future; for example, downscaling or upscaling of the forcing datasets, depending on the overall situation of projects, would be a possible option in this direction. This could be done with the help of local hydrometric information, for example, long-term annual rainfall. Otherwise, using the CFSR datasets without any modifications are likely to end up in erroneous predictions in semi-arid regions. Finally, an intercomparison of the currently addressed models and other models, irrespective of their complexity, are suggested as future course of work.

# References

1. Ghebrehiwot A, Kozlov D (2019) Vestn MGSU 14:8
2. Hrachowitz M et al (2013) Hydrol Sci J 58:6
3. Sivapalan M et al (2003) Hydrol Sci J 48:6
4. Montanari A et al (2013) Hydrol Sci J 58:6
5. McMillan H et al (2016) Hydrol Sci J
6. Fekete B, Vörösmarty C (2007) The current status of global river discharge monitoring and potential new technologies complementing traditional discharge measurements in Predictions in Ungauged Basins. In Proceedings of the PUB Kick-off meeting 20–22 November 2002, Brasilia, Brazil IAHS publication
7. Vörösmarty C et al (2001) Eos 82:5
8. Shiklomanov A, Lammers R (2009) Environ Res Lett 4:9
9. Shiklomanov A, Lammers R, Vörösmarty C (2002) Eos, Washington. DC 83**:**2
10. Ghebrehiwot A, Kozlov D (2020) Vestn MGSU 15:1
11. Ghebrehiwot A, Kozlov D (2020) Vestn MGSU 15:7
12. Dile Y, Srinivasan R (2014) J Am Water Resour Assoc 50:5
13. Fuka D, Walter M, Macalister C, Degaetano A, Steenhuis T, Easton Z (2014) Hydrol Process 28:22
14. Auerbach D, Easton Z, Walter M, Flecker A, Fuka D (2016) Hydrol Process 30:19
15. Nkiaka E, Nawaz N, Lovett J (2017) Hydrology 4:1
16. Mararakanye N, Le Roux J, Franke A (2020) Phys Chem Earth 117
17. Arnold J, Srinivasan R, Muttiah R, Williams J (1998) J Am Water Resour Assoc 34:1
18. Bugaec A, Garcman B, Tereshkina A (2018) Meteorologija i gidrologija 5 Russian)
19. Arnold J, Srinivasan R, Muttiah R, Allen P (1999) J Am Water Resour Assoc 35:5
20. Schuol J, Abbaspour K, Yang H, Srinivasan R, Zehnder A (2008) Water Resour Res 44:7
21. Schuol J, Abbaspour K, Srinivasan R, Yang H (2008) J Hydrol. 352(1–2)
22. Gassman P, Reyes M, Green C, Arnold J (2007) Trans ASABE 50:4
23. Abbaspour K, Rouholahnejad E, Vaghefi S, Srinivasan R, Yang H, Kløve B (2015) J Hydrol 524
24. Arnold J et al (2012) Trans ASABE 55:4
25. Yang D, Musiake K (2003) Hydrol Process 17:14
26. Kozlov D, Ghebrehiwot A (2019) Mag Civ Eng 3:87
27. A. Gehbrehiwot, D. Kozlov, *E3S Web Conf.* **97** (2019)
28. Abbaspour K, Vaghefi S, Yang H, Srinivasan R (2019) Sci Data 6:1
29. Abbaspour K, Johnson C, van Genuchten M (2004) Vadose Zo J 3:4
30. Gupta H, Sorooshian S, Yapo P (1999) J Hydrol Eng 4:2