Journal of
Water & Climate Change

 Journal of Water and Climate Change Vol 15 No 2, 703 doi: 10.2166/wcc.2023.548

Modeling surface water potential using the SWAT model combined with principal component analysis in the ungauged Gelana watershed, Ethiopia

Habtamu Daniel

College of Engineering, Hydraulic and Water Resources Engineering, Wolaita Sodo University, Wolaita Sodo, Ethiopia
E-mail: habtedanny12@gmail.com

HD, 0000-0002-6591-2364

ABSTRACT

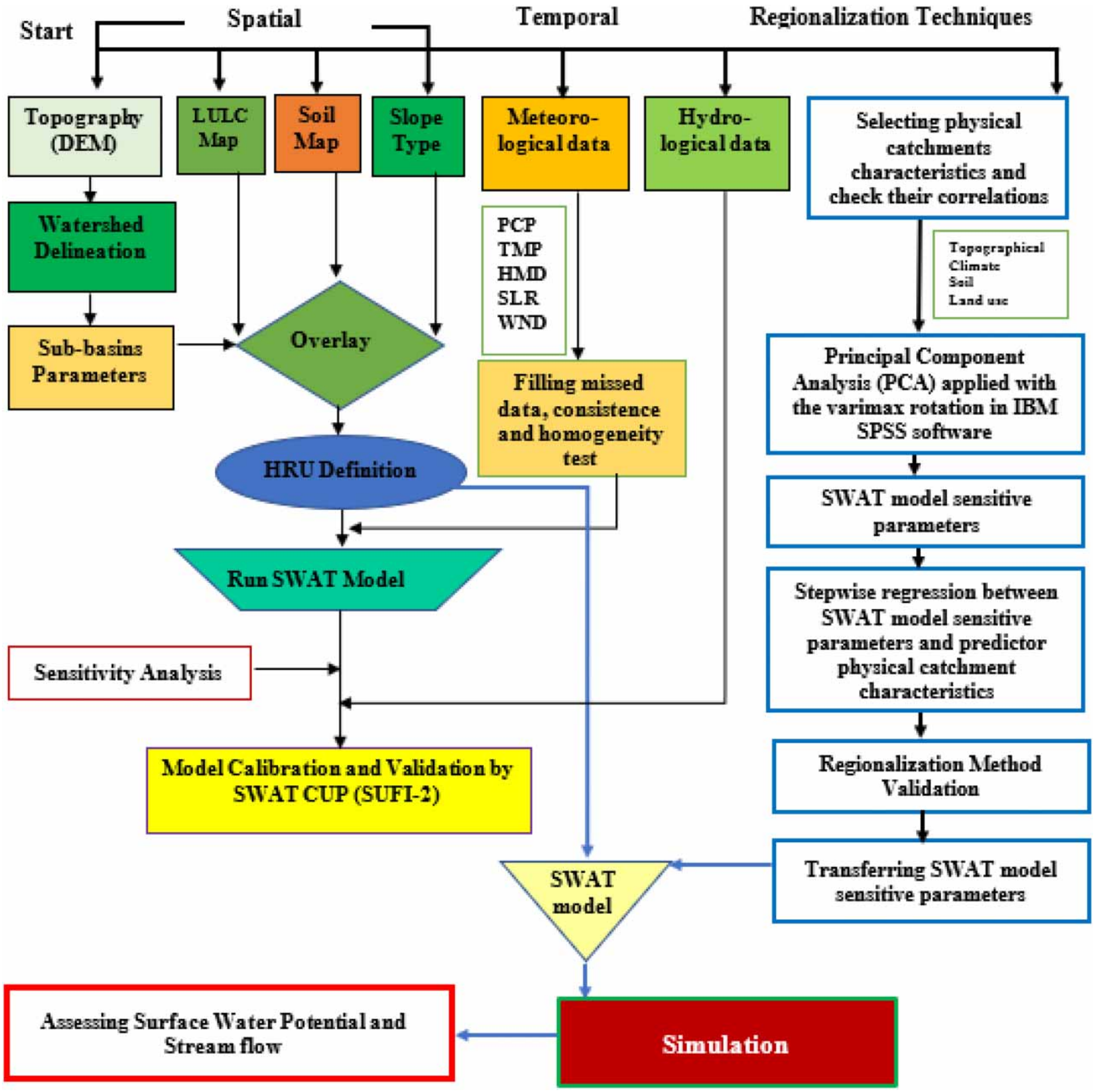
This study aims to model and assess surface water potential in an ungauged watershed using the Soil and Water Assessment Tool (SWAT), principal component analysis (PCA), and regression-based regionalization techniques in the Gelana River, Ethiopia. The SWAT model was calibrated (and validated) for the 1989–2007 (2008–2015) period, and it showed a very good performance to model the river flow. The 18 physical catchment characteristics that affect the production of streamflow were selected for correlation, and to make an equation using 19 optimized SWAT model parameters. These characteristics were categorized as two climate descriptors, three soil descriptors, seven land use land cover descriptors, and six topographical descriptors. The regression equations for each SWAT parameter with a function of the physical characteristics and principal components were developed. Then, SWAT model-sensitive parameters were transferred and validated. Results reveal that the watershed has a greatest surface runoff volume of 162.14 MCM in May and a least runoff of 9.15 MCM in January. The contribution of the water balance during the spring season is highest with a total surface runoff of 103.72 mm. From the whole Gelana watershed area of 336,460 ha, 778.4 MCM of yearly surface runoff was produced. The maximum mean monthly river flow is 15.7 m^3/s in May.

Key words: principal component analysis, regionalization, SPSS, stepwise multiple linear regression, SWAT, ungauged watershed

HIGHLIGHTS

- The 18 physical catchments characteristics were selected for correlation, and to make an equation with optimized SWAT parameters.
- From the entire Gelana watershed area of 336,460 ha, 778.4 MCM of annual surface runoff was produced.
- Streamflow data play a significant role in existing and upcoming engineering design and water resources management in and around the watershed.

GRAPHICAL ABSTRACT



1. INTRODUCTION

Accurate information on stream flows is the basis for the management, planning, and design of water resources projects like hydropower, irrigation project development, water supply, flood forecasting and control, and a sustainable aquatic life and ecosystem. Likewise, it is necessary for studies in the watersheds and basins. However, researchers and project designers have a crucial problem to address in order to achieve success on hydrological modeling in the watersheds. The streamflow gauging stations in river basins are commonly installed on main rivers due to the challenge of estimating daily streamflow with its stochastic and complicated structure (Burgan 2022). The main problem is the unavailability of gauged streamflow series data for calibration and validation of the hydrological models. Streamflow estimation in ungauged and scarcely gauged watersheds are a key study area in surface water hydrology, since streamflow data are essential for management and development of surface water resources. Moreover, it is necessary for water and land use managers, administrators, planners, builders, engineers, recreationists, and for all sectors. Besides, the daily, monthly, seasonal, and annual streamflow data are very useful for characterizing streamflow variability. Unfortunately, in many cases, several watersheds are ungauged at the river outlets. Thus, it is possible to model the ungauged watershed using regionalization methods and hydrological models (Guo *et al.* 2021; Daniel & Abate 2022).

Besides, industrialization and urbanization have a substantial effect on the hydrological cycle. The land use land cover and climate change are the foremost issues that can change the hydrological process of the catchment (Shahid *et al.* 2018, 2021). Moreover, the variation of spatial patterns of landscape disturbs water balance components in the watersheds (Nannawo *et al.* 2021). Assessing the surface water potential of a river is crucial for providing the strategic evidence needed for long-term planning of water resources, and managing the water projects. Thus, it is vital to study a river's potential for effective water resources planning and management. The surface water is a renewable resource, of which the quality and quantity are space–time dependent, that can possibly make economic returns and is the source of rainfed agriculture. Water resources development, water resources utilization, and integrated water resources management are the best programs and are known as a tool for sustainable economic growth, water related conflict management, and poverty reduction in developing countries (Ashine 2021). Rivers and lakes have served as the key sources of water during human history, and found less than 0.3% of nearly 3% of the Earth's freshwater (Beza *et al.* 2023). However, the rural people who depend on agriculture and livestock production for their livelihoods require an effective management of surface water. Moreover, irrigation development is necessary for sustainable and reliable agricultural development in Ethiopia by satisfying the demands of food security and poverty reduction (Ashine 2021). These sustainable developments are ensured by assessment of the potential of available surface water resources and designing best utilization mechanisms on the watershed level.

The regionalization method is transforming hydrological information from gauged watersheds to ungauged watersheds. The aim of parameter regionalization is providing effective hydrological information to watersheds (Guo *et al.* 2021; Daniel & Abate 2022). The regionalization methods can be used effectively to estimate streamflow in ungauged watersheds (Arsenault *et al.* 2019). Numerous types of regionalization approaches have been proposed in the last few decades for predictions in ungauged watersheds (Tegegne & Kim 2018). In general, there are four major regionalization methods: similarity-based (spatial proximity and physical similarity) method, regression-based method, hydrological signatures-based method, and catchment runoff-response similarity approach (Tegegne & Kim 2018; Guo *et al.* 2021).

The spatial proximity refers to finding one or more donor watersheds (gauged) that are adjacent to the target watersheds (ungauged) in space, and regionalizing its parameters to the target watersheds by interpolation or averaging. It is proposed based on watershed spatial similarity that adjacent spaces have autocorrelation characteristics (Beck *et al.* 2016; Tegegne & Kim 2018; Guo *et al.* 2021). Hydrological signature refers to static and dynamic indicators that can reflect the hydrological characteristics of watersheds on different time scales (Zhang *et al.* 2018; Guo *et al.* 2021). The watersheds that have similar runoff responses and rainfall characteristics are considered hydrologically similar. The runoff-response approach linked with the Soil and Water Assessment Tool (SWAT) model calibrated parameters is the integrated output of the model and expresses all of the interactions related to the hydrological phenomenon within a watershed (Tegegne & Kim 2018).

The regression-based regionalization does not need to define the similarity measure, but it establishes the relationship between hydrological parameters and watershed characteristics (Guo *et al.* 2021). Regression analysis is the most recommended approach used to set up a regional model for the assessment of model parameters in the ungauged watersheds. In addition, the stepwise multiple linear regression techniques can be applied in modeling hydrological responses such as surface runoff from the watersheds (Sharma *et al.* 2015; Daniel & Abate 2022).

Principal components analysis (PCA) is the statistical method for reducing a large number of interrelated variables into a smaller number of dominant variables, and has been used in many areas of scientific research (Wuttichaikitcharoen & Babel 2014). Large datasets are increasingly common and every so often problematic to interpret. PCA is an adaptive data analysis technique for reducing the dimensionality of such datasets, increasing interpretability, and minimizing the loss of information. It is developed by forming new uncorrelated variables that continually maximize variance (Jolliffe *et al.* 2016; Daniel & Abate 2022). PCA is a data-analytic technique that obtains linear transformations of a group of correlated variables such that certain optimal conditions are achieved. However, the transformed variables are uncorrelated (Daniel & Abate 2022). The spatially distributed rainfall information and the large dataset do not always lead to higher model performance. In this case, performing the PCA is important for circumstances where reliable, and long-recorded hydrological data are not available for modeling (Hu *et al.* 2007). The relevance of PCA is mainly due to three reasons: it can represent the variance of a scalar field with comparatively fewer independent coefficients, it can remove redundant variables in a multivariate dataset, and it can represent physically independent processes (Syed *et al.* 2004).

PCA is the best method to get a better correlation and group the optimized parameters into physically significant components. According a study in Kanhiya Nala watershed, India, by Sharma *et al.* (2015), the PCA in grouping geomorphic parameters of a watershed for hydrological modeling clearly reveals that some of the parameters are strongly correlated with the components, and it is a good method for screening out the insignificant parameters from the analysis. Moreover, Gajbhiye & Sharma (2015), who studied the application of the PCA method for the interpretation and grouping of water quality parameters, concluded that the results of PCA reflect a good outlook on the water quality monitoring and interpretation of the surface water. Furthermore, evaluating the variability of physical variables contributing to the hydrological cycle, the principal components (PCs) showed proper cyclicity at seasonal and annual timescales (Syed *et al.* 2004).

SWAT is the semi-distributed hydrological model that is computationally effective for modeling and assessing the watershed hydrological process. It is physically based, continuous in time and is capable of simulating long periods for computing the effects of management changes (Daniel & Abate 2022; Daniel 2023).

Lack of streamflow data for most of the watersheds makes it difficult for many sectors in the Rift valley basin watersheds in Ethiopia to assess water resources availability, and any related study in the watersheds and sub-basins. Assessment of water resources in a watersheds and sub-basins is very significant for proper management, planning, decision making, domestic sectors, industrial sectors, disaster management, agricultural activity, design of bridges, and dams. According to Daniel & Abate (2022), there are several occurrences of flooding in the Gelana watershed area; irrigated agriculture field and water supply are increasing over time. So, for proper management of these phenomena, the historical streamflow data are necessary. Nonetheless, the Gelana River is ungauged in the watershed outlet. Therefore, the objective of this study is ungauged watershed modeling and assessing surface water potential using SWAT, PCA, and regression-based regionalization techniques in the Gelana River, since streamflow estimation plays a significant role in existing and forthcoming engineering design and water resources management in and around Gelana watershed. Subsequently, hydrological data are significant for planning and management of surface water resources in the watershed. In addition, the streamflow data are essential for characterizing streamflow changeability in the Gelana River.

2. MATERIALS AND METHODS

2.1. Study area

The Gelana River is situated in the Lake Abaya sub-basin, Rift valley basin in the southeastern part of Ethiopia. The Gelana watershed is located in between latitude 5°25′15.9″N to 6°17′32.7″N and longitude 37°49′54.1″E to 38°21′22.7″E (Figure 1). The Gelana River originates from the Gedeo zone, Yirga Chefe highland area that drains into and meets the Lake Abaya with a watershed area of about 3,364.6 km^2. It is a perennial river. This paper also includes the Gidabo river, Hare river, and Kulfo rivers located in the Rift valley basin, lake Abay-Chamo sub-basin.

The topographic feature of the Gelana watershed has diverse altitudinal difference which ranges from 1,171 to 3,167 m above mean sea level. In addition, the classification of the Gelana watershed ranges from humid in the highlands to semi-arid in the lowlands of the watershed. Rainfall patterns in the Gelana watershed have a bimodal profile with an absolute peak in May and relative peak in October, with the main rain occurring from March to May (Belg), and from August to November (end of Kiramet to Tseday). Moreover, relatively intensive rainfall was received in April, May, September, and October, with the maximum mean monthly rainfall received in May at the Yirga Chefe station. The minimum mean monthly rainfall was recorded at the Hagere Mariam station in January; also, in all other stations, the lowest rainfall occurred from December to February but started to increase in March. Additionally, the mean monthly maximum and minimum temperatures vary from 20.87 °C in July at Fiseha Genet to 33.60 °C in March at Arba Minch, and 8.39 °C in January at Yirga Chefe to 18.54 °C in March at Arba Minch, respectively. The highest maximum and minimum temperatures were recorded in the Arba Minch station in March.

Furthermore, according to the land use/ land cover classification, nine major land uses/ land cover types were identified in the Gelana watershed. However, the major part of the watershed was covered by agricultural land which covered about 1,336.23 km^2 (39.70%) of watershed area, and the lowest part of the watershed was covered by water body which accounts for about 0.19 km^2 (0.01%) of the watershed from the whole study area (Table 1). Also, based on the dominant characteristics, the soil of the study area was classified into four major groups: Humic Nitisols, Chromic Luvisols, Eutric Fluvisols, and Eutric Vertisols. But Humic Nitisols is the dominant soil type covering an area of 1,514.90 km^2 (45.02%), whereas Eutric Vertisols soil type covering the lowest area of 363.16 km^2 (10.82%) of the watershed from the total area (Table 2).

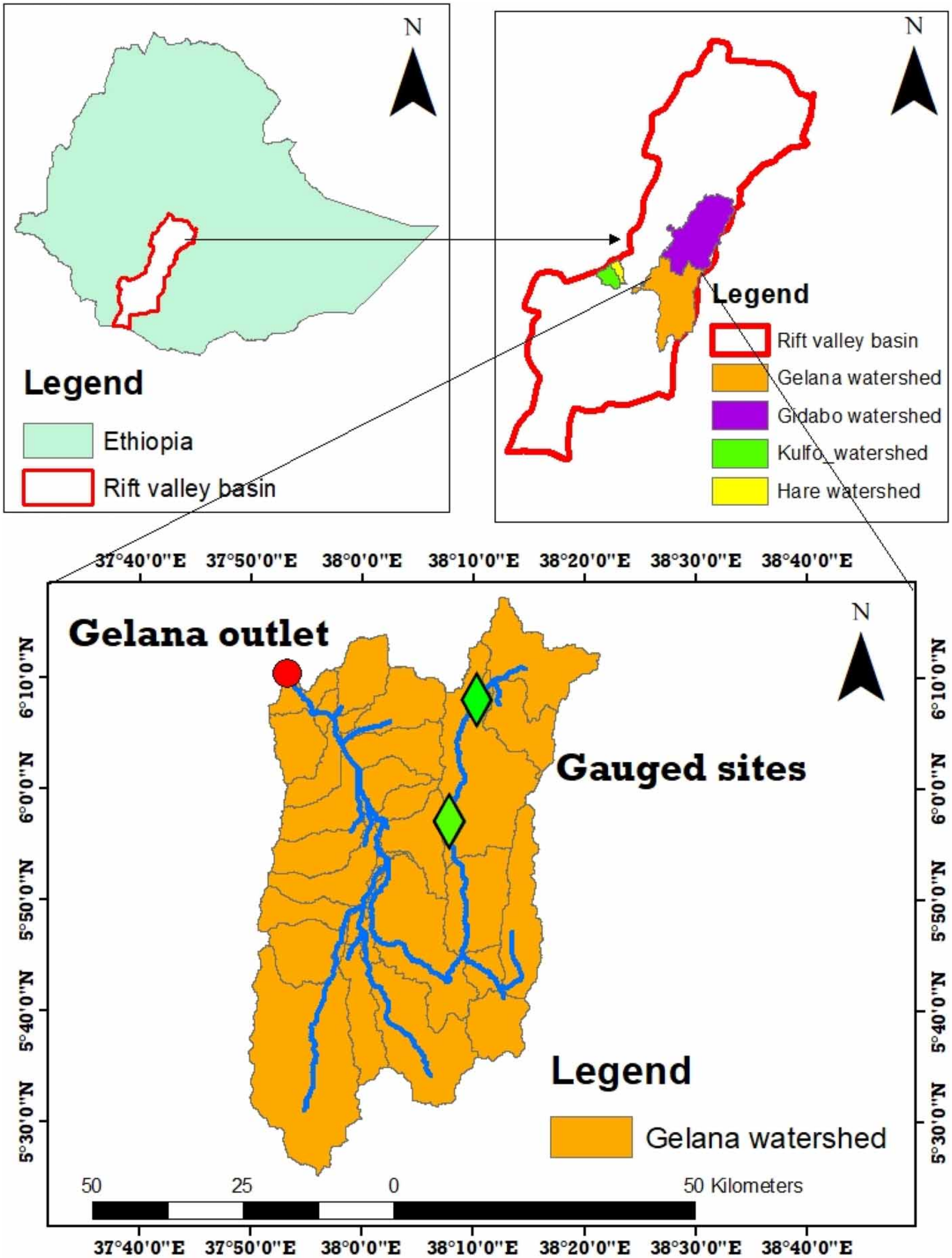


Figure 1 | Location map of the study area.

Table 1 | Land use land cover of the study area and their areal coverage

No	Types of land use/cover	SWAT code	Area (km^2)	Percentage covered (%)
1	Agricultural	AGRL	1,336.23	39.70
2	Forest-Evergreen	FRSE	983.84	29.25
3	Shrub land	RNGB	693.07	20.61
4	Bare land	BARR	132.19	3.93
5	Forest-Mixed	FRST	112.37	3.34
6	Settlement	URBN	68.95	2.05
7	Grass land	PAST	28.87	0.86
8	Wetland	WETL	8.88	0.26
9	Water body	WATR	0.19	0.01
	Total		3,364.6	100

Table 2 | Major soil types of the study area with areal coverage

No	Types of soil	SWAT code	Area (km^2)	Percentage covered (%)
1	Humic Nitisols	NTu	1,514.90	45.02
2	Chromic Luvisols	LVx	865.27	25.64
3	Eutric Fluvisols	FLe	621.27	18.52
4	Eutric Vertisols	VRe	363.16	10.82
	Total		3,364.6	100

2.2. Data sources

2.2.1. Materials and software used

2.2.1.1. Meteorological data. Climate data used as input for the SWAT model were collected from the National Meteorological Agency of Ethiopia. The main data, sources and their descriptions are summarized in Table 3. These include rainfall, maximum and minimum temperature, solar radiation, wind speed, and relative humidity with varied data lengths. There are seven meteorological stations in and around the Gelana watershed. However, only two stations were synoptic stations, which have all climate variables data. These are Arba Minch and Hagere Mariam stations (Figure 2 and Table 4). Moreover, climate data and the locations of the neighbor watersheds as shown in Table 5.

2.3. Hydrological (streamflow) data

Hydrological data (streamflow) of Gelana, Gidabo, Kulfo, and Hare river streamflow data were obtained from the Ministry of Water, Irrigation, and Energy (MoWIE) from 1980 to 2015. The Gelana River streamflow was gauged in two nearly stations

Table 3 | Major data and their sources

Data	Sources of data	Descriptions
Terrain	From Alaska satellite facility (https://asf.alaska.edu/)	DEM (12.5 m $\times$12.5 m)
Observed climate data	National Meteorological Agency (NMA)	Rainfall, Maximum and minimum temperature, Wind speed, Solar radiation and Humidity
Hydrological data (streamflow)	From Ministry of Water, Irrigation, and Energy (MoWIE)	Gelana streamflow gauged at Tore and Yirga Chefe. Also, the Gidabo, Kulfo, and Hare river streamflow data.
Land use land cover and Soil map	From Ministry of Water, Irrigation, and Energy (MoWIE)	Land use and land cover map, and Soil map

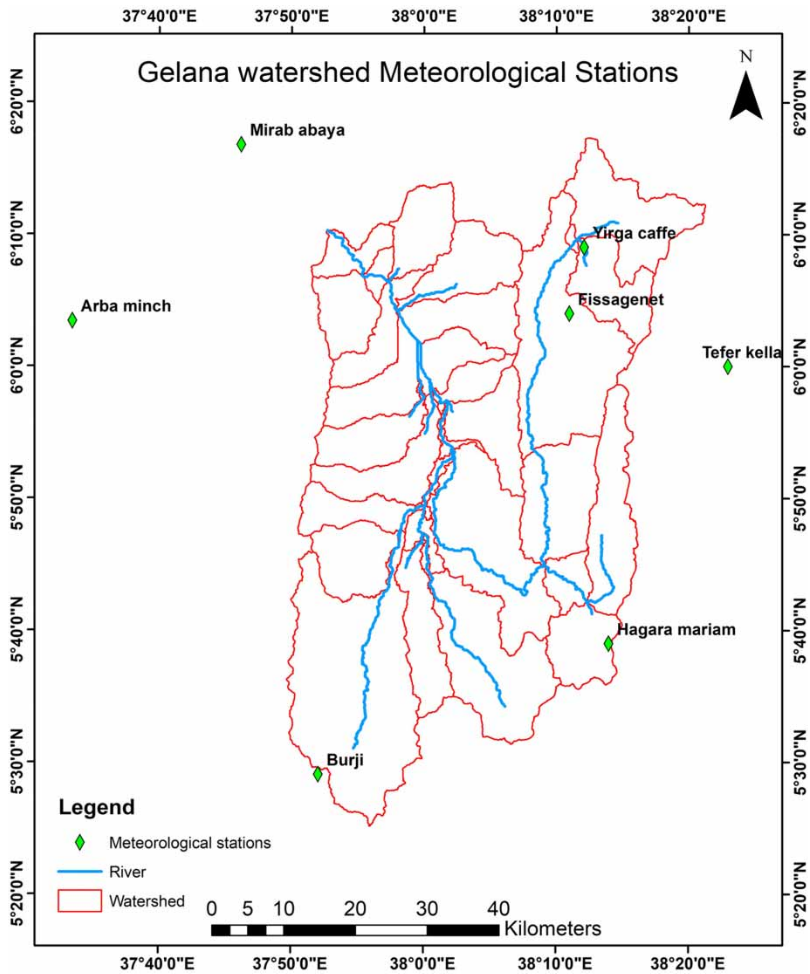


Figure 2 | Meteorological stations of the watershed.

Table 4 | Climate data and the locations in the Gelana watershed

Stations name	**Latitude**	**Longitude**	**Elevation**	**Data length**
Mirab Abaya	6.28	37.77	1,221	1987–2019
Arba Minch	6.06	37.56	1,207	1987–2019
Burji	5.48	37.87	1,815	1987–2018
Fiseha Genet	6.07	38.18	2,240	1987–2019
Hagere Mariam	5.65	38.23	1,861	1987–2019
Tefere Kella	6.00	38.38	1,870	1987–2019
Yirga Chefe	6.15	38.2	1,856	1987–2019

(Figure 1), the upper location is the Yirga Chefe station and the downstream is the Tore station. The maximum mean monthly streamflow was recorded in May and October, and the minimum flow occurred from December to March in both gauging stations. The highest mean monthly streamflow was around 9.98 m^3/s at the Tore station, and 9.97 m^3/s at the Yiga Chefe station in October. The major materials and software used and their purposes are tabulated and summarized in Table 6. In addition, the locations and availability of the streamflow gauging stations are mentioned in Table 7.

Table 5 | Climate data and the locations of the neighbor watersheds

Stations name	Latitude	Longitude	Elevation	Watersheds
Aleta Wendo	6.60	38.42	1,947	
Aposto	6.74	38.37	1,762	
Hagere Selam	6.47	38.52	2,809	Gidabo watershed
Yirga Chefe	6.15	38.20	1,856	
Dilla	6.34	38.30	1,579	
Mirab Abaya	6.28	37.77	1,221	
Arba Minch	6.06	37.56	1,207	Hare watershed and
Chencha	6.23	37.58	2,632	Kulfo watershed
Zenga	6.35	36.95	1,229	

Table 6 | Major materials and software used in this study

No	Material name	Purposes
1	Arc GIS 10.4.1	Used to obtain the physical parameters and spatial information of the watersheds.
2	Arc SWAT 2012 model	Used to delineate the watersheds, and simulate the streamflow for gauged and ungauged watersheds.
3	SWAT-CUP 2012	For sensitivity analysis, calibration, validation, and uncertainty analysis of the SWAT model.
4	IBM SPSS statistics software	Used to check the correlations and covariances of the physical characteristics of the catchments, to develop the principal component analysis, and stepwise regression equation.

Table 7 | Available streamflow gauging stations and their locations

No	River name/gauging station	Latitude	Longitude
1	Gelana at Tore	5.89	38.14
2	Gelana at Yirga Chefe	6.15	38.18
3	Gidabo at Aposto	6.75	38.38
4	Hare at Arba minch	6.07	37.60
5	Kulfo at Sikela	6.03	37.53

2.4. General framework of the study

The general framework of the study is shown in Figure 3.

2.5. Hydrologic modeling using SWAT

2.5.1. Weather database

The SWAT model requires the daily rainfall, maximum and minimum temperature, solar radiation, wind speed and relative humidity. Not all meteorological stations were synoptic stations, however, Arba Minch and Hager Mariam were synoptic stations (having all types of climatic data) used for generating the remaining weather database for the other five meteorological stations which have no full climate data. After calculating the WXGEN parameters, the corresponding location table was prepared according to the SWAT format, and then loaded into the model. SWAT takes data of each climatic variable from the nearest weather station measured from the meteorological stations.

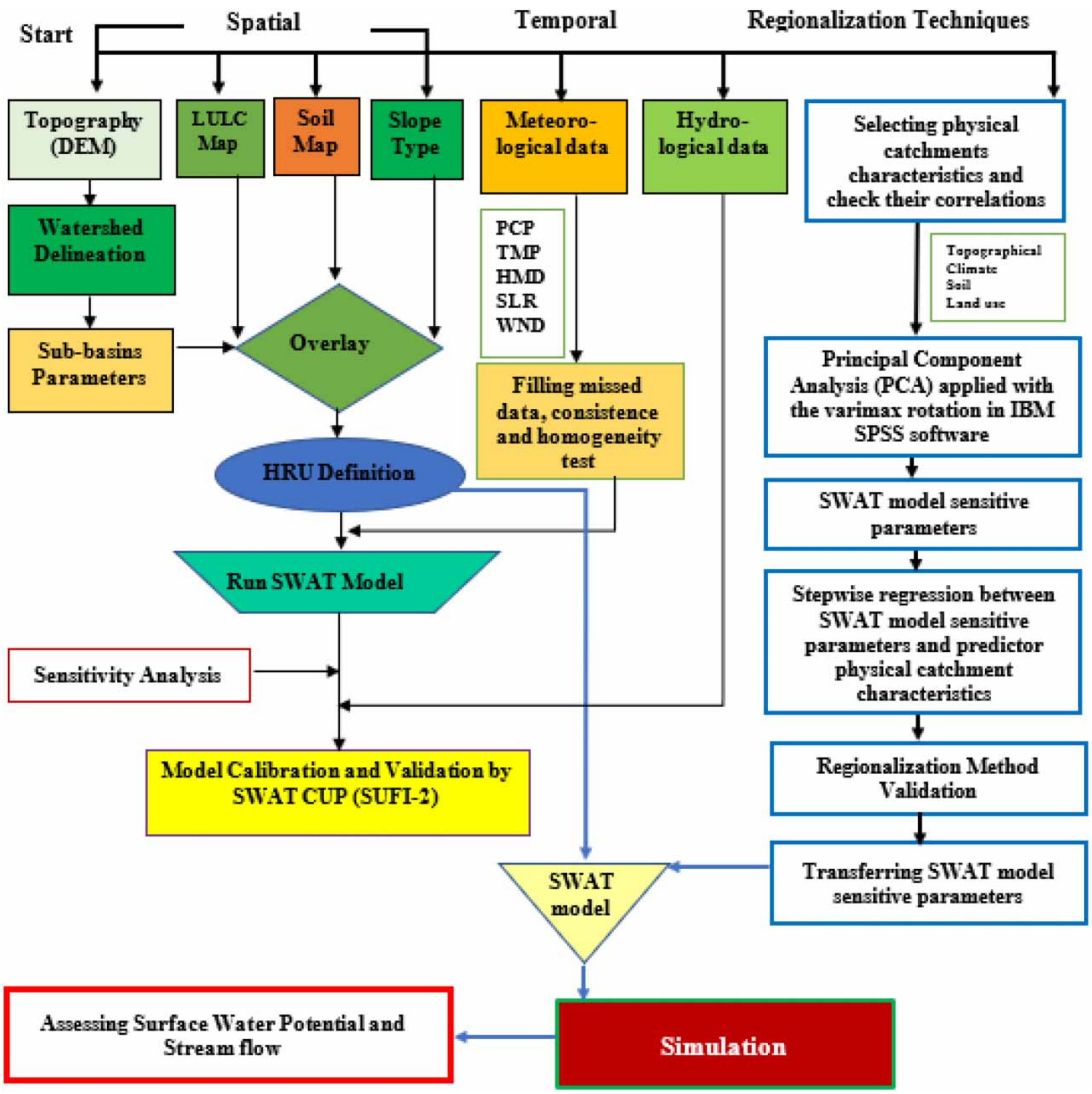


Figure 3 | The general framework of the study. PCP indicates precipitation, TMP indicates temperature, HMD indicates relative humidity, SLR indicates solar radiation, and WND indicates wind speed.

2.5.2. Watershed delineation

This was done by loading the DEM (12.5 m × 12.5 m) in the Arc SWAT 2012 model. Then, the stream network was generated by using a threshold area that defines the origin of a stream. For watershed delineation, the locations of the streamflow gaugings were added manually as sub-basin outlets. This ensures that the model calibration was done at the exact location. Based on this, the total area of the watershed was 3,364.6 km^2.

2.5.3. Hydrological response unit

The prepared land use and soil map were loaded in the delineated watershed by Arc SWAT, and then the slope was classified. The multiple slope option was selected, i.e., five classes (0–5, 5–10, 10–15, 15–20, and above 20). The land use land cover, soil map, and slope classes were reclassified corresponding with the parameters in the SWAT database. Then, all these physical properties were made to be overlaid for HRU definition. The HRU was defined by considering the threshold levels. In multiple HRU definitions, a threshold level was used to eliminate minor land uses, soils, and slope classes in each sub-basin. Subdividing the sub-watershed into areas having unique land use, soil and slope combinations makes it possible to study the differences in streamflow and evapotranspiration for different land cover, soil, and slope. For multiple HRU definitions, 5% land use land cover, 10% soil, and 15% slope threshold were used. Finally, the Gelana watershed was categorized into 441 HRUs with 29 sub-basins.

2.5.4. SWAT simulation

The SWAT input files were organized and the model was set to run. At the end, it simulates the streamflow in the Gelana watershed. The daily weather data such as rainfall, temperatures, solar radiation, wind speed, and relative humidity were loaded in the SWAT model. A total of 29 years including 2-year warm-up periods of the seven climatic stations from 1 January, 1987 to 31 December, 2015 were used for SWAT simulation depending on data availability.

2.5.5. Hydrologic water balance

Water balance is the driving force of the all processes that occur in the watershed hydrologic cycle (Daniel 2023). The hydrologic cycle simulated by SWAT is based on the following water balance equation:

$$\mathrm{SW_t} = \mathrm{SW_o} + \sum_{t=1}^{t} (R_{\mathrm{day}} - Q_{\mathrm{surf}} - E_a - W_{\mathrm{deep}} - Q_{\mathrm{gw}}) \tag{1}$$

where $\mathrm{SW_t}$ is the final soil water content in mm, $\mathrm{SW_o}$ is the initial soil water content in a day in mm, t is the time in days, R_{day} is the amount of precipitation in a day (in mm) i, Q_{surf} is the amount of surface runoff in a day in mm, E_a is the amount of evapotranspiration in a day in mm, W_{deep} is the amount of water entering the vadose from the soil profile in a day (mm), Q_{gw} is the amount of the return flow in a day (mm).

2.5.6. Estimation of surface runoff

The Soil Conservation Service (SCS) curve number was utilized to appraise the surface runoff due to its capability to utilize everyday input information. The SCS runoff equation is an empirical model that came into common use within the 1950s. The model was created to supply a steady premise for evaluating the amounts of runoff under varying land use and soil types (Daniel 2023).

The SCS curve number equation is:

$$Q_{\mathrm{surf}} = \frac{(R_{\mathrm{day}} - I_a)^2}{R_{\mathrm{day}} - I_a + S} \tag{2}$$

where Q_{surf} is the accumulated runoff or rainfall excess (mm H_2O), R_{day} is the rainfall depth for the day (mm H_2O), I_a is the initial abstractions which include surface storage, interception, and infiltration prior to runoff (mm H_2O), and S is the retention parameter (mm H_2O).

Hence, a surface runoff will occur when R_{day} is greater than I_a. The curve number is mainly dependent on the types of soil and land uses/land cover of the watershed (Daniel 2023).

Furthermore, the retention parameter varies spatially due to changes in soils, land use, management, slope, and temporally due to changes in soil water content (Daniel 2023). The retention parameter is defined as:

$$S = 25.4 \left(\frac{100}{\mathrm{CN}} - 10 \right) \tag{3}$$

where CN is the curve number for the day.

The initial abstraction, I_a is commonly approximated as 0.2S. Finally, the SCS curve number equation becomes

$$Q_{\mathrm{surf}} = \frac{(R_{\mathrm{day}} - 0.2S)^2}{R_{\mathrm{day}} + 0.8S} \tag{4}$$

2.6. Sensitivity analysis, calibration, and validation

SUFI-2 (Sequential Uncertainty fitting version 2) embedded in the SWAT-CUP is used for the sensitivity analysis, calibration, and validation of the hydrological model (SWAT model).

2.6.1. Sensitivity analysis

Owing to a large number of flow parameters in the SWAT model, identifying the most sensitive parameters is necessary to improve the calibration of the hydrological model. Sensitivity analysis is identifying the most sensitive parameters that strongly influence the flow process. The parameter ranking is taken from the last iterations of SUFI-2 based on *t*-Stat and *p*-value. The larger the absolute value of *t*-Stat and the smaller the *p*-value the more sensitive the parameter (Dibaba *et al.* 2020; Daniel & Abate 2022). Sensitivity analysis is the method of determining the rate of change in model output concerning changes in model inputs (parameters), and it is essential to identify key parameters and the parameter precision required for calibration (Arnold *et al.* 2012). It is also a measure of the effect of the change of one parameter on another. It minimizes the number of parameters to be used in the calibration step by making use of the most sensitive parameters largely controlling the behavior of the simulated process. It reduces the time required for the calibration and validation process. Moreover, it increases the accuracy of calibration by reducing uncertainty (Gassman *et al.* 2014).

The SUFI-2 was given several iterations to reach acceptable results. Each iteration provided the suggested values for the new parameters to be used in the next iteration. Finally, it achieved an acceptable result with the values of the Nash–Sutcliffe, coefficient of determination, percent of bias, and other uncertainty analysis statistical parameters.

The SWAT-CUP was run monthly from the period 01 January, 1987 to 31 December, 2015. The first 2 years from 01 January, 1987 to 31 December, 1988 were used for model warm-up.

The initial 21 parameters set were selected based on studies undertaken in the Rift valley basin watersheds located near the Gelana watershed (Shanka 2017; Demmissie *et al.* 2018). In addition, they were selected by considering the scenario (default) results from the Txt-In-Out folder after the SWAT model simulation.

Generally, the streamflow simulation considered numbers of hydrological input parameters of groundwater (.gw), management (.mgt), soil (.sol), hydrologic response units (.hru), routine (.rte) and sub-basin (.bsn) categories as shown in Table 8.

2.6.2. Calibration

Calibration of the hydrological model is the process of estimating model parameters by comparing the model prediction with the observed data for the same condition (Dibaba *et al.* 2020). The SWAT model was constructed with state-of-the-art components in an effort to simulate the processes physically and realistically. Inputs of the model are physically based (i.e., based on readily accessible information). This indicates SWAT is not a 'parametric model' with a formal optimization procedure (as part of the calibration process) to fitting any data (Santhi *et al.* 2001).

Calibration is the determination to better parameterize a model to a given set of local conditions, thus reducing the prediction uncertainty. It is performed by wisely selecting values for model input parameters within their respective uncertainty ranges (Arnold *et al.* 2012). This involves comparing the model results generated with the use of historic meteorological data to recorded streamflow. Generally, calibration succeeded in identifying the sensitive parameters by comparing model-simulated streamflow with observed streamflow data for the period of 01 January, 1989–31 December, 2007.

2.6.3. Validation

Validation is used to test the calibrated model without further parameter adjustments with an independent dataset and the results are compared to the remaining observational data to evaluate the model prediction (Dibaba *et al.* 2020). It is the process of representing that a given site-specific model is capable of making sufficiently accurate simulations (Santhi *et al.* 2001).

Validation was the last stage of the modeling, verifying the performance of the SWAT model for simulated flows in the periods 01 January, 2008–31 December, 2015.

2.7. SWAT model performance evaluation

Performance of the model simulation with the observed streamflow is expressed by statistics techniques such as coefficient of determination, Nash–Sutcliffe efficiency (NSE), and percent bias (Abbaspour *et al.* 2015; Daniel & Abate 2022). Furthermore, see Table 9.

Coefficient of determination (R^2) designates the proportion of the variance in measured data explained by the model, and is widely used for model evaluation. It is oversensitive to high extreme values (outliers) and insensitive to additive and

Table 8 | SWAT model parameters selected for sensitivity analysis

No	Parameters	Description	Default range	Category
1	CN2	SCS runoff curve number (Initial SCS CN II value)	35–98	.mgt
2	ALPHA_BF	Baseflow alpha factor [days]	0–1	.gw
3	GW_DELAY	Groundwater delay [days]	0–500	.gw
4	GWQMN	Threshold depth of water in the shallow aquifer required for return flow to occur [mm]	0–5,000	.gw
5	ALPHA_BNK	Baseflow alpha factor for bank storage [days]	0–1	.rte
6	CH_N2	Manning's 'n' value for the main channel	−0.01 to 0.3	.rte
7	GW_REVAP	Groundwater 'revap' coefficient	0.02–0.2	.gw
8	ESCO	Soil evaporation compensation factor	0–1	.hru
9	SOL_Z	Depth from the soil surface to bottom of layer	0–3,500	.sol
10	SOL_K	Saturated hydraulic conductivity	0–2,000	.sol
11	HRU_SLP	Average slope steepness [m/m]	0–1	.hru
12	SURLAG	Surface runoff lag time	0.05–24	.bsn
13	RCHRG_DP	Deep aquifer percolation fraction	0–1	.gw
14	EPCO	Plant uptake compensation factor	0–1	.hru
15	SOL_AWC	Available water capacity of the soil layer	0–1	.sol
16	SLSUBBSN	Average slope length [m]	10–150	.hru
17	REVAPMN	Threshold depth of water in the shallow aquifer for 'revap' to occur [mm]	0–500	.gw
18	OV_N	Manning's 'n' value for overland flow	0.01–1	.hru
19	BIOMIX	Biological mixing efficiency	0–1	.mgt
20	CH_K2	Effective hydraulic conductivity in main channel alluvium [mm/hr]	−0.01 to 500	.rte
21	CANMX	Maximum canopy storage	0–100	.hru

Table 9 | SWAT model performance evaluation formulas

Formula	Value	Rating
$R^2 = \dfrac{\left[\sum_{i=1}^{n}(O_i - O_{avg})(S_i - S_{avg})\right]^2}{\sqrt{\sum_{i=1}^{n}(O_i - O_{avg})^2 \sum_{i=1}^{n}(S_i - S_{avg})^2}}$	$0.75 < R^2 \leq 1$ $0.65 < R^2 \leq 0.75$ $0.5 \leq R^2 \leq 0.65$ $R^2 < 0.5$	Very good Good Satisfactory Unsatisfactory
$\mathrm{NSE} = 1 - \dfrac{\sum_{i=1}^{n}(O_i - S_i)^2}{\sum_{i=1}^{n}(O_i - O_{\mathrm{avg}})^2}$	$0.75 < \mathrm{NSE} \leq 1$ $0.65 < \mathrm{NSE} \leq 0.75$ $0.5 \leq \mathrm{NSE} \leq 0.65$ $\mathrm{NSE} < 0.5$	Very good Good Satisfactory Unsatisfactory
$\mathrm{PBIAS} = \dfrac{\sum_{i=1}^{n} O_i - \sum_{i=1}^{n} S_i}{\sum_{i=1}^{n} O_i} \times 100$	$\mathrm{PBIAS} < \pm 10\%$ $\pm 10\% \leq \mathrm{PBIAS} \leq 15\%$ $\pm 15\% < \mathrm{PBIAS} \leq 25\%$ $\mathrm{PBIAS} > \pm 25\%$	Very good Good Satisfactory Unsatisfactory

Here, R^2 is the coefficient of determination, NSE is the Nash–Sutcliffe Efficiency, and PBIAS is Percent Bias, O_i is the ith observed data, O_{avg} is the mean of the observed data, S_i is the ith simulated data, S_{avg} is the mean of model-simulated data, and n is the total number of events.

proportional differences between model predictions and measured data (Santhi *et al.* 2001; Moriasi *et al.* 2007). It shows how the simulated data correlates to the observed data. It ranges from 0 to 1, with higher values indicating less error variance (Moriasi *et al.* 2007; Daniel & Abate 2022).

NSE determines the relative magnitude of the residual variance compared to the measured data variance. NSE is very commonly used, which provides extensive information on reported values. It is also the best objective function for reflecting the overall fit of a hydrograph. It is less sensitive to high extreme values due to the squared differences (Moriasi *et al.* 2007; Daniel & Abate 2022). It measures how well trends in the observed data are reproduced by the simulated results over a specified period and for a specified time step. NSE ranges between $-\infty$ and 1. Values between 0 and 1 are acceptable levels of performance, whereas values NSE ≤ 0 indicate unacceptable performance (Moriasi *et al.* 2007; Daniel & Abate 2022).

Percent Bias (PBIAS) measures the mean tendency of the simulated data to be larger or smaller than their observed counterparts. Its values for streamflow tend to vary more among different autocalibration methods during dry years than wet years. This fact should be considered when attempting to do a split sample evaluation, one for calibration and one for validation (Santhi *et al.* 2001; Moriasi *et al.* 2007). It characterizes the percent mean deviation between observed and simulated flows. The best value of PBIAS is 0, with low values indicating accurate model simulation (Daniel & Abate 2022).

2.8. Uncertainty analysis of the SWAT model

SWAT model parameters account for uncertainty from driving variables, conceptual models, parameters, and measured data. Uncertainty analysis measures the goodness of fit and the 95% prediction uncertainty between simulated and observed streamflow. It is performed after sensitivity analysis by using SWAT-CUP software (Abbaspour *et al.* 2015).

The propagation of uncertainties in model outputs in SUFI-2 is expressed as the 95% probability distribution. The *p*-factor and *r*-factor statistics are used to quantify the fit between the result expressed as 95PPU and observation. 95PPU is calculated by the 2.5 and 97.5% levels of the cumulative distribution of the output variables. The degree of uncertainties is measured as the *p*-factor, and the measure quantifies the strength of uncertainty analysis by the *r*-factor. The percentage of observations covered by the 95PPU varies from 0 to 1 with the ideal value of 1, while for the *r*-factor, i.e., the thickness of the 95PPU optimal value is around 1 (Dibaba *et al.* 2020; Daniel & Abate 2022).

$$\bar{r} = \frac{1}{n}\sum_{t_i}^{n}(y_{t_i}^{M}, 97.5\% - y_{t_i}^{M},\ 2.5\%) \tag{5}$$

$$r\text{–factor} = \frac{p\text{–factor}}{\sigma_{\text{obs}}} \tag{6}$$

where $y_{t_i}^{M}$, 97.5% and $y_{t_i}^{M}$, 2.5% represents the upper and lower boundaries of the 95PPU, and σ_{obs} is the standard deviation of the measured data.

2.9. Regionalization method

This method transposes hydrological parameters information from gauged watersheds to ungauged watershed, and it is efficient for estimating streamflow in ungauged watershed outlets. The Gelana River streamflow was not gauged at the outlet of the watershed, due to the fact that the SWAT model optimized parameters were transferred from gauged donor watersheds to the ungauged watershed outlet.

The regionalization approach was done by the following steps. First, the SWAT model was calibrated for gauged watersheds against observed streamflow to establish good-performing sensitive parameter sets. Next, relationships were evaluated between SWAT model-sensitive parameters and physical catchment characteristics to develop the regionalization model. Following this, the model-sensitive parameters were defined based on the physical catchment characteristics and PCs from the ungauged watershed. We then substituted the ungauged watershed physical catchment characteristics and PC values in the developed regionalization equation, and got the sensitive parameters values. Finally, the regionalization model was validated by using the new sensitive parameters in the neighbor watershed. Then, the transferred SWAT model-sensitive parameters values were used to simulate the streamflow in the ungauged watershed.

The rivers used for regionalization were Gelana River gauged at Tore, Gelana River gauged at Yirga Chafe, Gidabo River gauged at Aposto, Hare River gauged at Arba Minch, and Kulfo River gauged at Sekala. Among them, the Gelana River gauged at Tore and Yirga Chefe, Gidabo River, and Hare River were selected as sensitive parameters donors, whereas the Kulfo river is the most neighboring river to the ungauged part of the Gelana River and was used for validation of regionalized parameters. The targeted location was the Gelana watershed outlet near Lake Abaya.

To develop a regionalization model, the optimized SWAT model parameters were assessed during calibration at the gauged watersheds by SWAT-CUP (SUFI-2). The neighboring (or nearby) watersheds must share physical attributes, such as land use land cover, soil, elevation, slope, and climate data; and they were becoming the donor watersheds. The optimized hydrological model parameters were transferred by using the regression method to the ungauged site, i.e., Gelana watershed outlet.

The 18 physical catchments characteristics were selected for correlation, and to make an equation with optimized SWAT parameters. The selected physical catchment characteristics should directly or indirectly affect the production of streamflow in the watershed. These characteristics were categorized as: two climate descriptors, three soil descriptors, seven land use land cover descriptors, and six topographical descriptors. They were prepared for developing the regionalization method as presented in Table 10.

The physical catchment characteristics that relate to the topography of the watersheds were extracted from the DEM. The others related to the land use land cover and soil were obtained from the land use land cover map and soil map, respectively. Similarly, the physical catchment characteristics under climate descriptors were obtained from the meteorological data as made available by NMA.

2.10. PCA

PCA is a data-analytic technique that obtains linear transformations of a group of correlated variables such that certain optimal conditions are achieved. However, the transformed variables are uncorrelated (Daniel & Abate 2022).

The spatially distributed rainfall information and the large dataset do not always lead to higher model performance. In this case, performing the PCA is important for circumstances where reliable, and long-recorded hydrological data are not available for modeling (Hu *et al.* 2007). PCA is used to identify a small number of derived variables from a larger number of original variables to simplify the subsequent analysis of the data (Wuttichaikitcharoen & Babel 2014). The relevance of PCA is mainly due to three reasons: it can represent the variance of a scalar field with comparatively fewer independent

Table 10 | Selected physical catchments characteristics

No	Physical catchment characteristics	Symbol
	Topographical descriptors	
1	Drainage area (km^2)	Area
2	Mean elevation (m)	ME
3	Length of the longest flow path (km)	LLP
4	Topographic wetness index	TWI
5	Aspect	Aspect
6	Flow accumulation	FA
	Climate descriptors	
1	Mean annual rainfall (mm)	MAR
2	Mean annual potential evapotranspiration (mm)	PET
	Soil descriptors	
1	Saturated hydraulic conductivity (mm/hr)	Ksat
2	Available water capacity of the soil layer	Swc
3	Bulk Density Moist (g/cc)	BDM
	Land use land cover descriptors	
1	Forest-Evergreen (%)	% FRSE
2	Agricultural (%)	% AGRL
3	Shrubland (%)	% RNGB
4	Forest-Mixed (%)	% FRST
5	Bare land (%)	% BARR
6	Grassland (%)	% PAST
7	Settlement (%)	% URBN

coefficients, it can remove redundant variables in a multivariate dataset, and it can represent physically independent processes (Syed *et al.* 2004).

PCA was applied to identify the factors influencing the Gelana watershed streamflow. The first step adopted in the PCA was the selection of a set of catchment characteristics indicators for the study area. The initial set consisted of the 18 physical catchment characteristics under four categories as shown in Table 10. The next step was the assessment of the suitability of data for the PCA using the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity. KMO is a test conducted to examine the strength of the partial correlation between the variables. It tests the ratio of item correlations to partial item correlations. If the partials are similar to the raw correlations, it means that the items do not share much variance. The range of KMO is from 0.0 to 1.0; however, the score of 0.50 is suggested as the minimum value for a good PCA (Wuttichaikitcharoen & Babel 2014). Bartlett's test of sphericity checks for the hypothesis that the correlation matrix is an identity matrix. The null hypothesis of the test is that the variables are orthogonal, i.e., not correlated. The alternative hypothesis is that the variables are not orthogonal, i.e., they are correlated enough to where the correlation matrix diverges significantly from the identity matrix. The significance value for the analysis led us to reject the null hypothesis, and conclude that there are correlations in the dataset that are appropriate for the PCA. The score from Bartlett's test of sphericity with significance at 95% ($p < 0.05$) was considered appropriate for the PCA (Wuttichaikitcharoen & Babel 2014). The last step was the determination of dominant factors. The PCA with Varimax rotation was performed to identify the PCs or subsets from a larger dataset. The extraction method of Varimax with Kaiser normalization was employed for the selection of the dominant factors. Kaiser's criterion (eigen values) rule states that, only components with eigen values of 1.0 or more are retained for further investigation and indicate highly correlated factor loadings in the PCs (Wuttichaikitcharoen & Babel 2014).

2.11. Regression analysis

Regression-based regionalization can be traced back to 1960 when Nash (1960) tried to find the correlation between the unit hydrograph and the catchment attributes, and simultaneously, Dalrymple (1960) had proposed the idea for flood frequency curve transposition. The main core of regression-based regionalization is to establish a certain regression relationship between model parameters and catchment descriptors. This relationship is used to predict the model parameters of target catchments, and then to achieve the purpose of a simulation or prediction runoff in ungauged catchments (Guo *et al.* 2021).

Stepwise regression analysis is one of the common approaches used for streamflow regionalization in both hydrologic model dependent and independent approaches. It is done based on the hydro-climatological and morphological attributes of the catchments (Swain & Patra 2015). The most commonly used procedure for selecting the best regression equation is stepwise linear regression analysis using the probability of 5% for selecting a factor, which is performed by using SPSS (Wuttichaikitcharoen & Babel 2014).

A stepwise multiple linear regression was used in this study to predict SWAT model-sensitive parameters from several independent physical catchment characteristics. It was assumed that better relations can be established using multiple physical catchment characteristics than when only one physical catchment characteristic is used (Rientjes *et al.* 2011). Therefore, relations between physical catchment characteristics and model parameters were assessed through stepwise multiple linear regression analysis.

The significance of the multiple linear regression equations was tested by evaluating the significance of individual coefficients and by a test of overall significance. Its statistical significance was tested through the coefficient of determination (R^2), t-stat, and p-value (at the significance level of 5%) of the regression statistics in the regression summary output; and check hydrologically relevant. This is a guarantee that the regression equations could be used for developing each parameter. It was performed for each model-sensitive parameter.

The multiple linear regression model is computed as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_n X_n \tag{7}$$

where Y is the predicted SWAT model parameter, X_1, X_2, X_3, X_n are the sets of predictors catchment characteristics, β_0: the intercept of the regression line, β_1, β_2, β_3, β_n: the coefficients of catchment features.

2.12. Regionalization model validation

Before its use of a regionalization model output, it must be tested where predicted and observed streamflow was compared for a neighbor gauged watershed that was not used when establishing the regionalization model. Therefore, the Kulfo watershed was used for validation only, and not to derive the regionalization equation for the ungauged watershed. Moreover, it is the nearest watershed to the ungauged part of the Gelana watershed outlet. In addition, it is located in a similar climate condition (semi-arid zone) with an ungauged part of the study watershed. Thus, the objective functions NSE and R^2 defined during the calibration and validation period were used to check the ability of the regionalization model (Daniel & Abate 2022).

3. RESULTS AND DISCUSSION

3.1. Hydrological modeling

3.1.1. Sensitivity analysis

The sensitive parameters which affect the hydrological (SWAT) model output with their ranks were evaluated by using SWAT-CUP (SUFI-2). The small *p*-value and large absolute value of *t*-Stat indicate the most sensitive parameter in the watershed. The parameters with medium, high, and very highly sensitive values that affect the model output significantly were used to calibrate and validate the hydrological model (Daniel & Abate 2022; Daniel 2023). The sensitive parameters in descending order are shown in Table 11. Hence, the eight topmost sensitive hydrological parameters in the Gelana watershed are ALPHA_BF.gw (Baseflow alpha factor in days), RCHRG_DP.gw (Deep aquifer percolation fraction), CH_K2.rte (Effective hydraulic conductivity in main channel alluvium), CN2.mgt (SCS runoff curve number), GWQMN.gw (Threshold depth of water in the shallow aquifer required for return flow to occur in mm), SOL_K(..).sol (Saturated hydraulic conductivity), SLSUBBSN.hru (Average slope length), and HRU_SLP.hru (Average slope steepness). The result indicates that the hydrological process of the watershed depends mainly on the action of these parameters. The ALPHA_BF.gw (Baseflow alpha factor in days) was found to be a very highly sensitive parameter.

Table 11 | Sensitivity parameters for the SWAT model

Rank	Parameters Name	*t*-Stat	*p*-Value	Fitted value
1	V__ALPHA_BF.gw	−31.87	0.00	0.000016
2	V__RCHRG_DP.gw	−16.52	0.00	0.31
3	V__CH_K2.rte	13.51	0.00	9.94
4	R__CN2.mgt	−5.45	0.00	−0.30
5	V__GWQMN.gw	4.45	0.00	942.68
6	R__SOL_K(..).sol	−3.05	0.00	−0.25
7	R__SLSUBBSN.hru	2.77	0.01	1.56
8	V__HRU_SLP.hru	−2.58	0.01	0.58
9	R__SOL_Z(..).sol	−1.26	0.21	−0.08
10	V__ESCO.hru	−1.14	0.25	0.62
11	V__SURLAG.hru	0.93	0.35	11.48
12	R__SOL_AWC(..).sol	−0.93	0.35	−0.09
13	R__ALPHA_BNK.rte	0.89	0.38	0.48
14	V__OV_N.hru	0.78	0.44	0.22
15	V__GW_DELAY.gw	0.67	0.50	32.35
16	V__GW_REVAP.gw	−0.60	0.55	0.12
17	V__EPCO.hru	−0.37	0.71	0.20
18	R__CH_N2.rte	0.35	0.72	−0.03
19	V__REVAPMN.gw	0.34	0.73	363.83

Here, *R* means an existing parameter value is multiplied by (1 + a fitted value), *V* means an existing parameter value is to be replaced by a fitted value.

3.1.2. SWAT model calibration and validation

The selected sensitive parameters were used for calibration and validation of the SWAT model with SWAT-CUP (SUFI-2) on monthly time steps from 1989 to 2015. From the total flow gauging periods about two-thirds of the time steps, i.e., from 1989 to 2007, were nominated for the calibration period, and the remaining around one-third periods i.e., from 2008 to 2015, were used for validation of the SWAT model gauged at the Tore station in the Gelana watershed (Figures 4 and 5).

The summary of SWAT model calibration and validation results are presented in Table 12. The calibration outcomes on average monthly streamflow illustrate that the SWAT model can capture the observed streamflow with R^2, NSE, and PBIAS of 0.74, 0.74, and 4.8, respectively. Correspondingly, the mean monthly streamflow validation shows R^2, NSE, and PBIAS of 0.71, 0.67, and 7.5, respectively. Additionally, the percentages of observations bracketed by the 95PPU were 78 and 66% of the observation, and *r*-factor equals 0.89 and 0.76 during calibration and validation periods, respectively. Overall, according to Santhi *et al.* (2001) and Moriasi *et al.* (2007), the SWAT model attained a good fit between observation and simulation in the Tore station of the Gelana watershed.

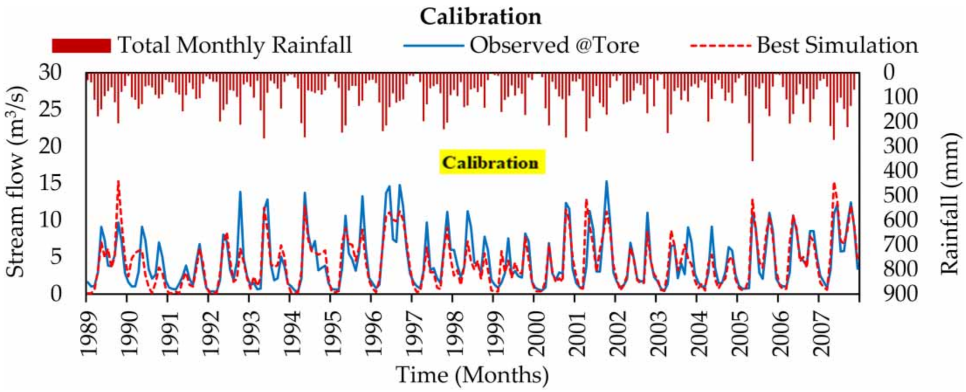


Figure 4 | SWAT model calibration in the Tore gauging station of the Gelana River.

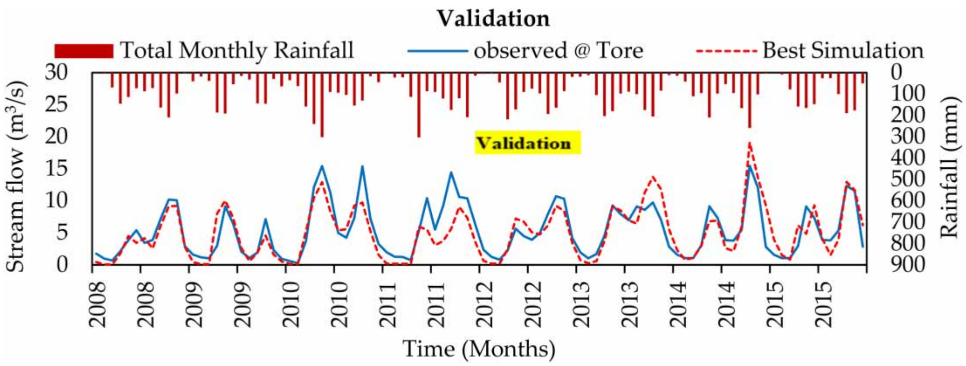


Figure 5 | SWAT model validation in the Tore gauging station of the Gelana River.

Table 12 | Summary of SWAT model performance in the Tore gauging station

	Evaluation parameters					Mean flow (m³/s)	
	R^2	NSE	PBIAS	*p*-factor	*r*-factor	Observed	Simulated
Calibration	0.74	0.74	4.8	0.78	0.89	4.44	4.23
Validation	0.71	0.67	7.5	0.66	0.76	5.34	4.94

3.2. Regionalization method

3.2.1. SWAT model performance

The SWAT model-sensitive parameters were transposed from gauged neighbor rivers to an ungauged Gelana River outlet. For donors' rivers and validation river, the calibration and validation of hydrologic model (SWAT) simulation were done to check the efficiency of the model performance as shown in Table 13 and Figure 6. The outcomes show that the donors and the validation rivers demonstrate a good agreement between the observed and model simulated flow (Table 13). Besides, the SWAT model sensitive parameters values at neighbor gauged catchments are shown in Table 14.

3.2.2. Physical characteristics of the catchments

The selected 18 physical catchment characteristics as shown in Table 15 should have a significant effect on the production of streamflow in the watershed, and relation with SWAT model-sensitive parameters. The inter-correlations of the selected 18 physical catchment characteristics were evaluated. The bold values indicate the inter-correlation coefficients above ± 0.60 as shown in Table 16. In addition, Kaiser–Meyer-Olkin's measure of sampling adequacy is 0.731, Bartlett's test of sphericity significance is 0.000. Therefore, the data was suitable for the PCA.

Generally, depending on the inter-correlation values, all selected physical catchment characteristics were in the range of moderate to strong correlation. The correlation matrix of physical catchment characteristics reveals strong correlation (correlation coefficient greater than ± 0.9), good correlation (correlation coefficient greater than ± 0.75), and moderately correlated (correlation coefficient greater than ± 0.60) (Daniel & Abate 2022). Hence, it is complex to cluster the parameters into components and to assess any physical significance at this step. The inter-correlation values are therefore directed to the PCA.

3.2.3. PCA

PCA applied with the varimax rotation in IBM SPSS statistics software shows three PCs with an eigen value greater than 1.00 as shown in Tables 17 and 18, and Figures 7 and 8 accounted for the total cumulative variance of 95.4% as per their eigen values. The first component has described about 55.0% of the variance in the catchment characteristics, the second component described 33.4%, and the third component described around 7.0%. Thus, 95.4% of the variance was explained by only the three components (Daniel & Abate 2022).

The extraction method of PCA by a varimax rotated component matrix was determined as shown in Table 19. The first component (PC1) is strongly correlated (more than ± 0.90) with Swc, BDM, and PET, and moderately correlated (more than ± 0.60) with FRST, Aspect, Ksat, AGRL, FRSE, and MAR, based on higher loading factors which may be termed as soil descriptors component. The second component (PC2) is strongly correlated with Area, LLP, and FA, and moderately correlated with BARR, RNGB, and ME, which may be termed as the topographical descriptors component. The third component (PC3) is well correlated (more than ± 0.75) with URBN and PAST, and moderately correlated with AGRL, FRSE, and MAR, which may be termed as land use land cover descriptors component (Daniel & Abate 2022).

The prominent variables for following regression analyses were selected based on loading factors (Wuttichaikitcharoen & Babel 2014). The upper most three variables with the highest loading factor and more than ± 0.60 were nominated as representative variables of each of the PCs. Nine of the parameters were very highly correlated. However, other parameters screen-out due to their significance, followed by regrouping the remaining variables into the physically significant factors (Sharma *et al.* 2015).

Table 13 | SWAT model performance in the gauging stations of the neighbor rivers

Gauging stations	Calibration (1989–2007)			Validation (2008–2015)		
	R^2	NSE	PBIAS	R^2	NSE	PBIAS
Gelana at Yirga Chefe	0.65	0.64	−3.5	0.67	0.56	−0.1
Kulfo at Sekala	0.65	0.63	7.5	0.63	0.55	5.2
Gidabo at Aposto	0.64	0.63	−3.3			
Hare at Arbaminch	0.61	0.61	−1.3			

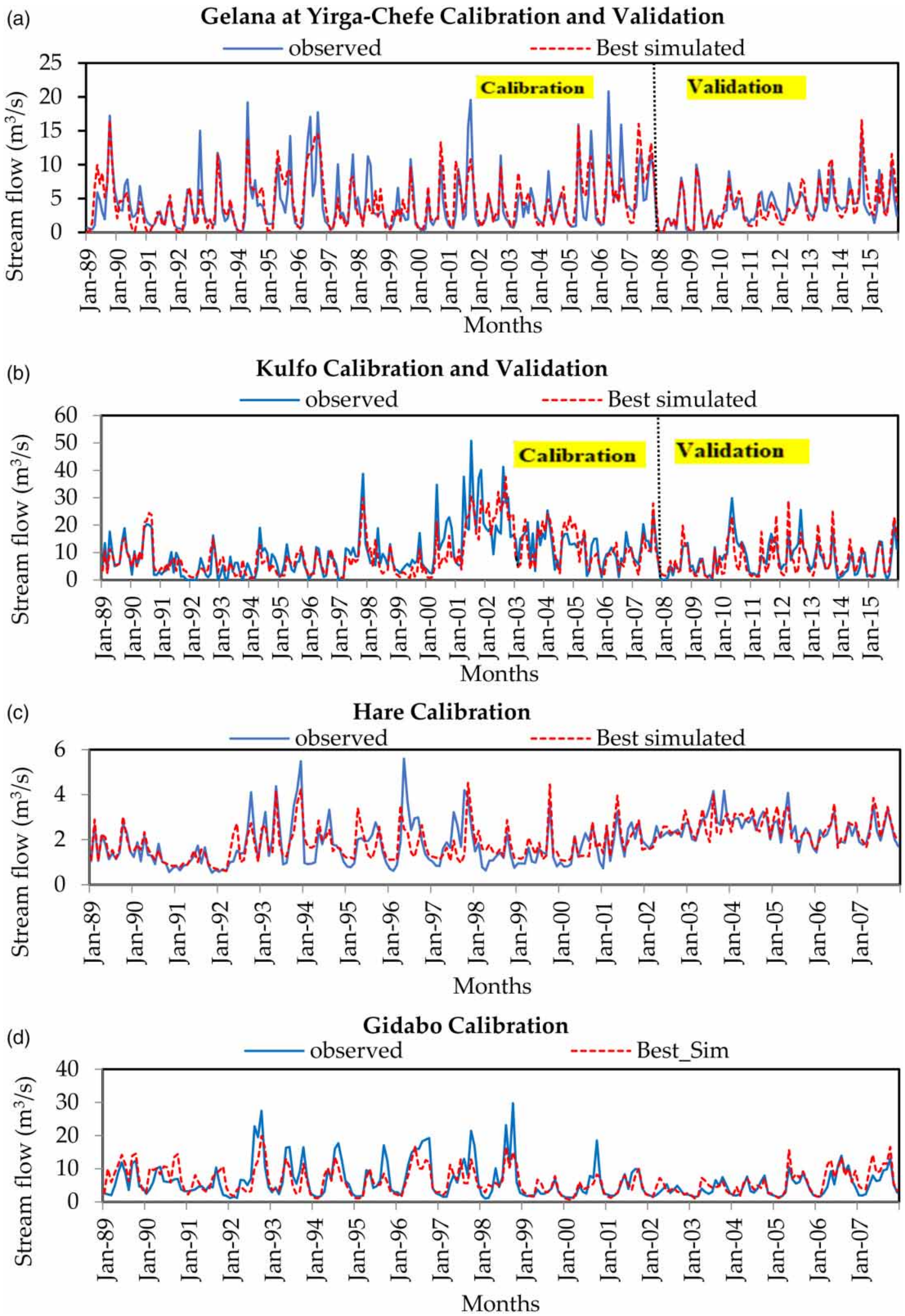


Figure 6 | The SWAT model calibration and validation in the nearby watersheds: (a) Gelana River gauged at Yirga cheffe, (b) Kulfo river at Sikala, (c) Hare river gauged at Arba minch, and (d) Gidabo river gauged at Aposto.

The screen-out parameters from the analysis having less significance in describing the component variance in the watershed were FRST, Aspect, Ksat, FRSE, MAR, BARR, RNGB, and ME. Generally, the correlation matrix and PC matrix were developed from these nine parameters. All nine parameters were supposed to be the forcing factors of streamflow either positive or negative effects, which can be used subsequently as predictor variables in regression analysis. These predictor variables are; from PC1: Swc, BDM, and PET; from PC2: Area, LLP, and FA; and from PC3: URBN, PAST, and AGRL were selected.

Table 14 | SWAT model optimized parameters values at neighbor gauged catchments

No	Parameters name	Gelana at Tore	Gelana at Yirga- Chafee	Gidabo at Aposto	Hare at Arba Minch	Kulfo at Sikala
1	V__ALPHA_BF.gw	0.000016	0.000065	0.00002	0.00007	0.61
2	V__RCHRG_DP.gw	0.31	0.94	0.12	0.39	0.37
3	V__CH_K2.rte	9.94	64.09	4.45	0.04	15.27
4	R__CN2.mgt	−0.30	−0.09	−0.27	−0.34	−0.09
5	V__GWQMN.gw	942.68	2,487.67	1,140.69	882.61	68.36
6	R__SOL_K(..).sol	−0.25	−0.03	−0.18	−0.15	−0.12
7	R__SLSUBBSN.hru	1.56	0.59	1.33	1.08	0.06
8	V__HRU_SLP.hru	0.58	0.41	0.70	0.10	0.11
9	R__SOL_Z(..).sol	−0.08	−0.07	1.12	2.31	1.97
10	V__ESCO.hru	0.62	0.13	0.25	0.11	0.98
11	V__SURLAG.hru	11.48	12.41	11.42	2.02	4.14
12	R__SOL_AWC(..).sol	−0.09	0.15	0.12	−0.05	−0.07
13	R__ALPHA_BNK.rte	0.48	0.92	0.67	0.26	0.20
14	V__OV_N.hru	0.22	0.74	0.43	0.17	0.35
15	V__GW_DELAY.gw	32.35	15.50	20.52	216.54	30.09
16	V__GW_REVAP.gw	0.12	0.15	0.04	0.09	0.09
17	V__EPCO.hru	0.20	0.91	0.11	0.37	0.30
18	R__CH_N2.rte	−0.03	0.20	0.01	−0.03	−0.02
19	V__REVAPMN.gw	363.83	209.43	95.90	349.21	138.14

Communalities estimate the variance in each variable accounted for by the components. All communality values in Table 19 are very high, which indicates that the extracted components represent the variables well.

3.2.4. Regression equation

One parameter from the significant components may form a set of independent parameters at a time of modeling the hydrologic responses (Sharma *et al.* 2015; Daniel & Abate 2022). After the stepwise regression between SWAT model-sensitive parameters and predictor physical catchment characteristics, the correlation coefficients were identified and the physical relevancy of each index to each parameter was checked. The catchment indices indicate better correlation and were hydrologically relevant to each SWAT model parameter concerning catchment response. Then, they were nominated and regressed over each SWAT model parameter.

Finally, the regression equation for each SWAT parameter with a function of the physical characteristics and PCs were developed after checking the hydrological, and statistical significance through R^2, *t*-test, and *p*-value of the regression statistics (Supplementary material, Appendix A).

The Supplementary material (Appendix A) values were analyzed by the multiple linear regression equation as:

$$\text{SWAT model sensetive parameters} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_n X_n \tag{8}$$

where X_1, X_2, X_3, X_n are the sets of predictors catchment characteristics, β_0: the intercept of the regression line, β_1, β_2, β_3, β_n: the coefficients of catchment features.

Then, substitute the physical characteristics and PC values of the Gelana watershed outlet. Thus, the SWAT model-sensitive parameters were transferred to the ungauged outlet of the Gelana watershed as shown in Table 20.

3.2.5. Regionalization model validation

The performance of the regionalization model was tasted by calibration and validation on the nearby watershed, before using the transferred sensitive parameters in an ungauged watershed (Gelana outlet). Thus, the Kulfo watershed was used to

Table 15 | The 18 physical catchment characteristics

Catchments	Area	ME	LLP	TWI	Aspect	FA	MAR	PET	% FRSE	% AGRL	% RNGB	% FRST	% BARR	% PAST	% URBN	Ksat	Swc	BDM
Gelana at Tore	622.6	2,064.6	61.2	6.4	196.5	2,360.0	1,361.4	351.2	83.9	10.6	0.9	0.1	0.0	0.1	4.4	21.8	0.1	1.2
Gelana at Yirga Chefe	282.3	2,190.9	26.1	6.2	194.8	970.0	1,370.4	347.9	85.1	9.0	0.7	0.1	0.0	0.0	4.8	20.2	0.2	1.4
Gidabo at Aposto	630.0	2,091.4	47.7	6.5	195.2	1,465.3	1,202.7	298.1	61.8	25.8	0.2	0.0	0.0	11.5	0.7	20.2	0.2	1.4
Hare at Arba minch	186.3	2,497.1	35.4	5.9	150.0	1,453.7	948.6	512.7	0.0	73.0	0.8	14.6	0.0	11.4	0.1	36.4	0.1	1.0
Kulfo at Sikela	370.0	2,296.2	39.2	5.7	159.0	1,500.9	1,100.1	511.5	0.0	73.4	2.1	10.4	0.0	14.0	0.0	24.1	0.1	1.1
Gelana at outlet	3,364.6	1,718.8	170.3	6.7	178.9	5,740.6	1,079.8	384.4	29.3	39.7	20.6	3.3	3.9	0.9	2.1	21.8	0.1	1.2

Table 16 | The inter-correlation of 18 physical catchment characteristics (PCCs)

PCCs	Area	ME	LLP	TWI	Aspect	FA	MAR	PET	% FRSE	% AGRL	% RNGB	% FRST	% BARR	% PAST	% URBN	Ksat	Swc	BDM
Area	1.00																	
ME	**−0.87**	1.00																
LLP	**0.99**	**−0.85**	1.00															
TWI	**0.67**	**−0.84**	**0.66**	1.00														
Aspect	0.10	−0.56	0.07	**0.70**	1.00													
FA	**0.98**	**−0.83**	**1.00**	**0.62**	0.04	1.00												
MAR	−0.21	−0.27	−0.23	0.30	**0.88**	−0.24	1.00											
PET	−0.19	**0.60**	−0.16	**−0.81**	**−0.96**	−0.11	**−0.72**	1.00										
% FRSE	−0.10	−0.35	−0.11	0.55	**0.95**	−0.13	**0.94**	**−0.88**	1.00									
% AGRL	−0.06	0.49	−0.04	**−0.66**	**−0.97**	−0.02	**−0.90**	**0.91**	**−0.99**	1.00								
% RNGB	**0.98**	**−0.78**	**0.97**	0.53	−0.05	**0.96**	−0.30	−0.03	−0.22	0.07	1.00							
% FRST	−0.21	**0.65**	−0.18	**−0.73**	**−0.99**	−0.14	**−0.84**	**0.96**	**−0.91**	**0.95**	−0.07	1.00						
% BARR	**0.99**	**−0.79**	**0.98**	0.59	−0.01	**0.97**	−0.29	−0.09	−0.18	0.02	**1.00**	−0.11	1.00					
% PAST	−0.42	**0.60**	−0.42	−0.59	**−0.63**	−0.44	**−0.62**	0.54	**−0.69**	**0.76**	−0.38	**0.65**	**−0.60**	1.00				
% URBN	0.03	−0.35	0.03	0.39	**0.74**	0.04	**0.86**	−0.59	**0.86**	**−0.86**	−0.02	**−0.72**	0.00	**−0.91**	1.00			
Ksat	−0.27	**0.68**	−0.23	−0.53	**−0.83**	−0.18	**−0.75**	**0.76**	**−0.68**	**0.72**	−0.17	**0.88**	−0.17	0.45	**−0.62**	1.00		
Swc	0.04	−0.50	−0.01	0.45	**0.87**	−0.06	**0.82**	**−0.81**	**0.76**	**−0.77**	−0.07	**−0.90**	−0.05	−0.37	0.53	**−0.96**	1.00	
BDM	−0.01	−0.38	−0.09	0.57	**0.86**	−0.15	**0.71**	**−02.89**	**0.79**	**−0.80**	−0.12	**−0.86**	−0.08	−0.37	0.51	**−0.77**	**0.88**	1.00

The bolded values indicate the high inter-correlated parameters.

Table 17 | Total variance explained by IBM SPSS software

	Initial eigenvalues			Extraction sums squared loadings			Rotation sums of squared loadings		
Component	**Total**	**% of variance**	**Cumulative %**	**Total**	**% of variance**	**Cumulative %**	**Total**	**% of variance**	**Cumulative %**
1	9.895	54.971	54.971	9.898	54.971	54.971	7.598	42.210	42.210
2	6.015	33.417	88.388	6.015	33.417	88.388	6.375	35.416	77.626
3	1.256	6.975	95.364	1.256	6.975	95.364	3.193	17.738	95.364
4	0.625	3.475	98.838						
5	0.209	1.162	100.000						
6	0	0	100.000						
7	0	0	100.000						
8	0	0	100.000						
9	0	0	100.000						
10	0	0	100.000						
11	0	0	100.000						
12	0	0	100.000						
13	0	0	100.000						
14	0	0	100.000						
15	0	0	100.000						
16	0	0	100.000						
17	0	0	100.000						
18	0	0	100.000						

Table 18 | Summary of PCs for catchment characteristics

PCs	Eigenvalues	Percent of variance (%)	Cumulative variance (%)
1	9.895	54.971	54.971
2	6.015	33.417	88.388
3	1.256	6.975	95.364

validate the regionalization model (Figure 9). According to Santhi *et al.* (2001), and Moriasi *et al.* (2007), the general results indicate that the $R^2 = 0.62$ and NSE $= 0.51$ during the calibration, and $R^2 = 0.66$ and NSE $= 0.52$ during validation. These outcomes reveal a satisfactory agreement of the observed and simulated streamflow.

Therefore, the PCA coupled with the stepwise regression method regionalization model has an adequate performance in the nearby watershed. So, the transferred SWAT model-sensitive parameters can govern the Gelana watershed outlet (ungauged part) streamflow simulation.

3.3. Water balance and surface water potential

SWAT model simulation in Tables 21 and 22 reveals that the rainfall, surface runoff, lateral flow, total water yield, and river flow are directly related, and inversely related to temperature, evapotranspiration and potential evapotranspiration in the watershed. Average monthly water balance in Table 21 reveals that the most noteworthy mean monthly rainfall (190.41 mm) and surface runoff (48.19 mm) are recorded in May. Similarly, the maximum lateral flow contribution to river flow is evaluated in May (27.14 mm). Their subsequent result is the highest water yield in May (128.64 mm). Additionally, the watershed has the greatest runoff volume of 162.14 MCM and the least runoff of 9.15 MCM in May and January, respectively.

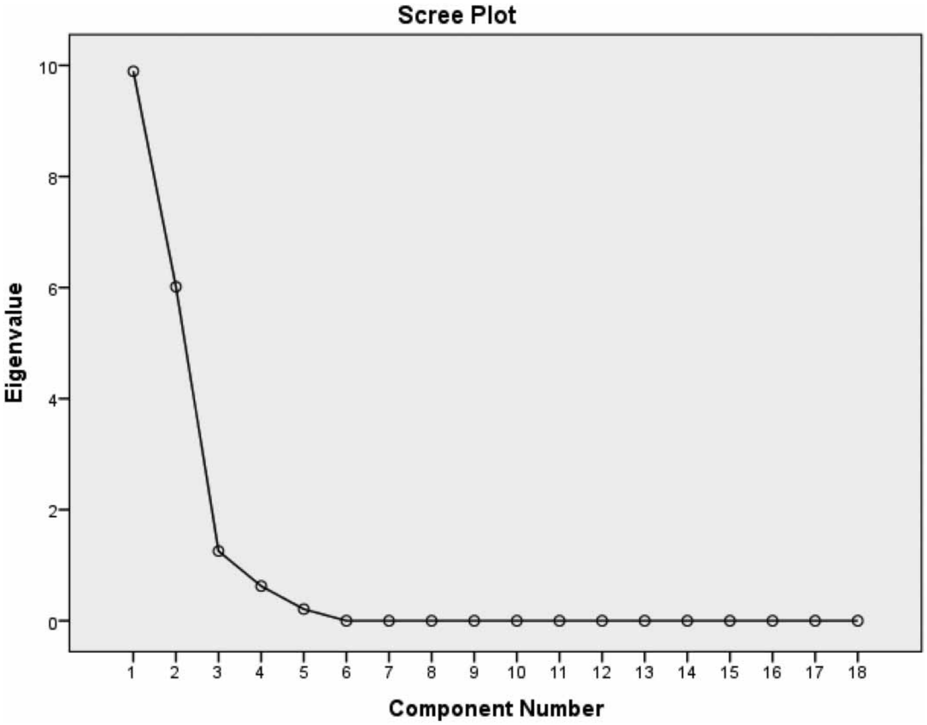


Figure 7 | Scree plot by IBM SPSS software.

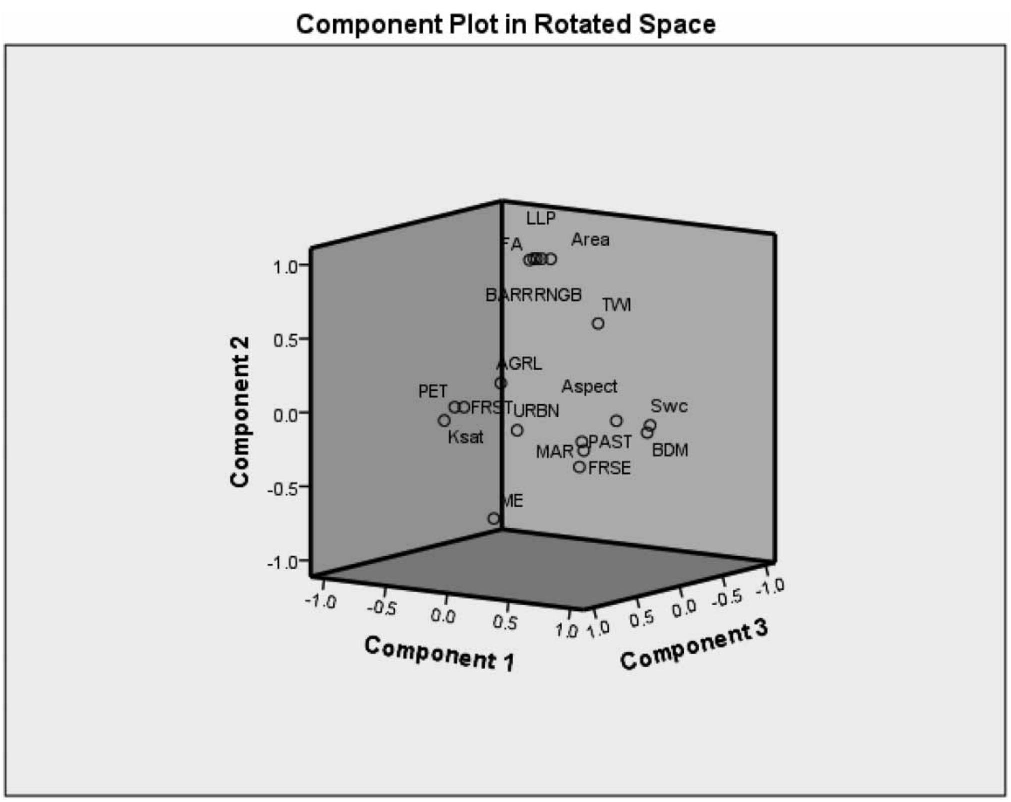


Figure 8 | Component plot in rotated space.

Table 19 | Results of PCA (Varimax rotated component matrix)

Factor	Eigenvectors			Communalities
	PC1	PC2	PC3	
Swc	**0.950**	−0.023	0.110	0.914
BDM	**0.931**	−0.073	0.118	0.885
PET	**− 0.909**	−0.143	−0.273	0.920
FRST	−0.709	−0.159	−0.381	0.996
Aspect	0.696	0.048	0.425	0.985
Ksat	−0.688	−0.210	−0.122	0.847
AGRL	−0.676	−0.017	**− 0.619**	0.985
FRSE	0.665	−0.139	0.608	0.985
MAR	0.619	−0.257	0.602	0.945
Area	0.062	**0.997**	−0.001	0.998
LLP	0.009	**0.997**	0.027	0.994
FA	−0.056	**0.989**	0.076	0.988
BARR	−0.047	0.689	0.005	0.980
RNGB	−0.081	0.683	−0.010	0.973
ME	−0.504	−0.639	−0.150	0.980
TWI	0.598	0.545	0.212	0.819
URBN	0.418	0.003	**0.897**	0.980
PAST	−0.288	−0.400	**− 0.864**	0.990

The bold values indicates the extraction method by PCA, rotation method; Varimax with Kaiser normalization indicate highly correlated variables and factor loadings in the PCs.

Table 20 | SWAT model-sensitive parameters transferred to the ungauged Gelana watershed outlet

No	Parameters name	Transposed value
1	V__ALPHA_BF.gw	0.0001
2	V__RCHRG_DP.gw	0.25
3	V__CH_K2.rte	15.42
4	R__CN2.mgt	−0.18
5	V__GWQMN.gw	1,221.33
6	R__SOL_K(..).sol	−0.21
7	R__SLSUBBSN.hru	1.30
8	V__HRU_SLP.hru	0.43
9	R__SOL_Z(..).sol	0.15
10	V__ESCO.hru	0.42
11	V__SURLAG.bsn	8.24
12	R__SOL_AWC(..).sol	−0.06
13	R__ALPHA_BNK.rte	0.49
14	V__OV_N.hru	0.32
15	V__GW_DELAY.gw	33.60
16	V__GW_REVAP.gw	0.09
17	V__EPCO.hru	0.38
18	R__CH_N2.rte	−0.05
19	V__REVAPMN.gw	272.70

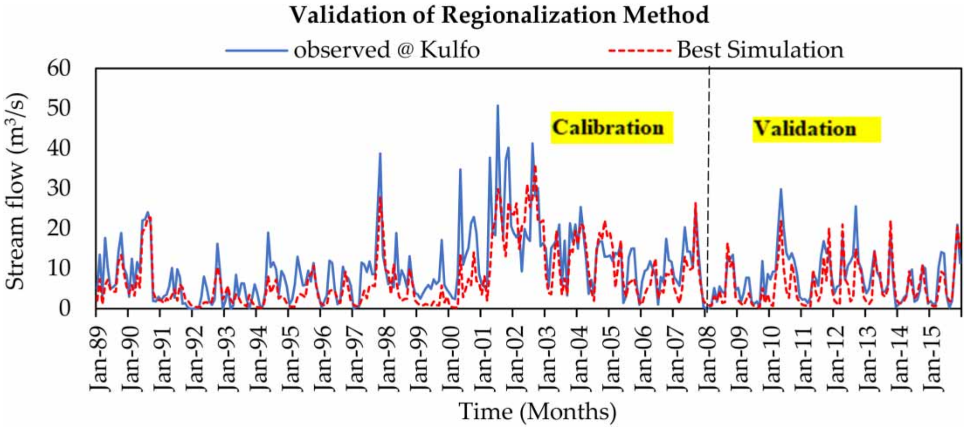


Figure 9 | Regionalization model validation.

Table 21 | Average monthly water balance values of the watershed

Month	Rainfall (mm)	SUR_Q (mm)	LAT_Q (mm)	WYLD (mm)	ET (mm)	PET (mm)	Av. temp
January	23.72	2.72	2.14	32.09	15.01	46.16	20.54
February	24.95	3.44	2.32	18.97	15.73	41.67	21.51
March	78.64	14.75	7.59	32.81	28.46	42.90	21.69
April	181.57	40.78	21.82	83.75	27.48	31.49	20.63
May	190.41	48.19	27.14	128.64	25.42	27.93	19.82
June	83.14	14.30	11.53	89.05	16.15	19.02	19.21
July	64.85	8.12	7.64	67.36	11.89	14.52	18.89
August	86.25	12.34	10.63	62.46	15.66	18.95	19.22
September	128.53	21.14	14.51	73.47	21.07	25.71	19.58
October	169.64	40.66	25.41	117.27	23.97	28.81	19.57
November	84.93	20.16	13.33	95.09	24.32	36.25	19.58
December	34.81	4.74	3.47	58.04	19.59	43.05	19.78

SURF_Q is the surface runoff, LAT_Q is the later flow in the watershed, WYLD is Water yield, ET is evapotranspiration, PET is potential evapotranspiration, and Av. temp is the average temperature.

Table 22 | Average annual water balance values of the watershed

No	Water balance	Depth (mm)
1	Rainfall	1,151.5
2	Surface runoff	231.34
3	Lateral flow	147.54
4	Total water yield	859.00
5	Evapotranspiration	244.75
6	Potential evapotranspiration	376.45
7	Return flow	133.51
8	Recharge to deep aquifer	144.41
9	Percolation to shallow aquifer	577.62
10	Re-evaporation from shallow aquifer	33.51

There are four seasons in Ethiopia: Winter (Bega) includes December, January, and February; Spring (Belg) includes March, April, and May; Summer (Kiremt) includes June, July, and August; and Autumn (Tseday) includes September, October, and November (Daniel & Abate 2022). The contribution of the water balance during the spring season is high with a total rainfall of 450.62 mm depth, total surface runoff of 103.72 mm, and lateral soil flow of 56.55 mm. However, the water balance is low during the winter season with a total rainfall of 83.48 mm, surface runoff depth of 10.9 mm, and lateral soil flow of 7.93 mm. In addition, the contribution of the water balance during the autumn season is relatively higher than the summer season. A corresponding study was done in the Somodo Watershed, Ethiopia by Ashine (2021).

The result in Table 22 shows that the mean annual rainfall of the watershed is 1,151.5 mm, the surface water runoff is 231.34 mm, and potential evapotranspiration is 376.45 mm. From the whole Gelana watershed area of 3,364.6 km^2, 778.4 million m^3 yearly surface runoff was produced.

Generally, the rainfall pattern of the Gelana watershed is the bimodal profile with an absolute peak in May and relative peak in October, with the maximum rain occurring (wet season) from March to May, and from August to November. Furthermore, the evapotranspiration is dependent on the crop growth, air temperature and soil water content. Consequently, the average temperature rise results in increasing potential evapotranspiration and actual evaporation, which could be a critical factor for the reduction of total water yield. As the forms of water are exposed to losses, owing to the rises in temperature, the evaporation is also a factor for the reduction of the surface runoff and river flow in the Gelana watershed.

Results of a comparative study in the Jewuha watershed, Awash basin, Ethiopia, by Beza *et al.* (2023) show that the SWAT model performs very well to demonstrate the surface water compared to the ground water within the wet season.

3.4. Estimated streamflow

The Gelana River at the outlet was estimated by using PCA coupled with the stepwise regression regionalization method. The SWAT model-sensitive parameters were transferred to the ungauged outlet of the Gelana watershed, then simulate the streamflow using the transposed sensitive parameters at the outlet of the watershed. The estimated hydrological data (streamflow) was for the period 1989–2015, since the available climate data are in these periods, and the hydrological model (SWAT) simulates from available weather input data.

The estimated mean monthly streamflow of the Gelana watershed at outlet around 3,364.6 km^2 from 1989 to 2015 were shown in Figure 10. The peak streamflow was estimated in May and October, and the low streamflow occurred from December to March. The maximum mean monthly streamflow was estimated about 15.7 and 10.25 m^3/s at the outlet in May and October, respectively.

Furthermore, the highest mean annual streamflow was estimated around 9.3 m^3/s in 2010, and the minimum flow was seen around 1.07 m^3/s in 2002 in the Gelana watershed outlet (Figure 11). Generally, the monthly and yearly streamflow show similar trends with the gauged Gelana River at upstream in Tore and Yirgacheffe stations.

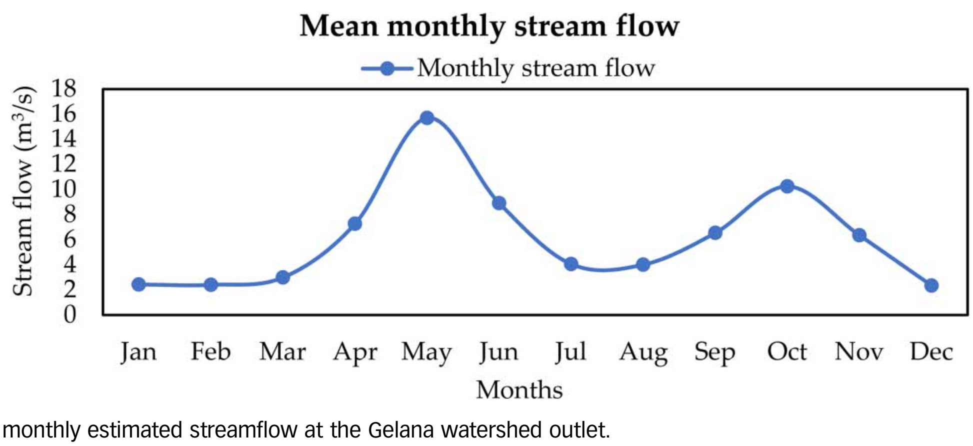


Figure 10 | Mean monthly estimated streamflow at the Gelana watershed outlet.

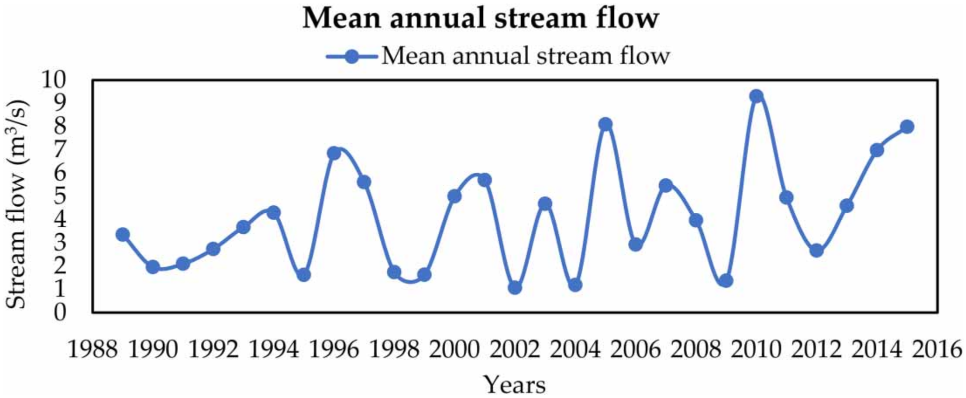


Figure 11 | Mean annual estimated streamflow at the Gelana watershed outlet.

Therefore, the surface water potential of the Gelana watershed was 778.4 MCM annually from a total area of 336,460 ha. Thus, surface water potential within the ungauged watershed is adequate to create any pressure and gravity driven structures for the purpose of obtaining agrarian efficiency and for water supply within the watershed zone.

4. CONCLUSION

This study aims to model and assess surface water potential in an ungauged watershed using the SWAT, PCA, and regression-based regionalization techniques in the Gelana River, Ethiopia. The five rivers, namely Gelana River gauged at Tore, Gelana River gauged at Yirgacheffe, Gidabo river, Hare river, and Kulfo river were used for regionalization. The SWAT model was calibrated (and validated) for the 1989–2007 (2008–2015) period, on the five rivers. Thus, the 19 sensitive parameters were selected by using SWAT-CUP (SUFI-2). Next, the 18 physical catchments characteristics were selected for correlation, and to make an equation with 19 optimized SWAT model parameters. These characteristics were categorized as: two climate descriptors, three soil descriptors, seven land use land cover descriptors, and six topographical descriptors. The regression equation for each SWAT parameter with a function of the physical characteristics and PCs were developed. The performance of the regionalization model was validated on the Kulfo watershed, and reveals a satisfactory agreement. So, the transferred SWAT model-sensitive parameters can govern the Gelana watershed outlet (ungauged part) stream flow simulation. Results reveal that the watershed has the greatest runoff volume of 162.14 MCM and a least runoff of 9.15 MCM in May and January, respectively. The contribution of the water balance during the spring season is high with a total rainfall of 450.62 mm depth, total surface runoff of 103.72 mm, and lateral soil flow of 56.55 mm. However, the water balance is low during the winter season with a total rainfall of 83.48 mm, surface runoff depth of 10.9 mm, and lateral soil flow of 7.93 mm. From the whole Gelana watershed area of 3,364.6 km^2, 778.4 million m^3 yearly surface runoff was produced. Consequently, the average temperature rise results in increasing potential evapotranspiration and actual evaporation, which could be a critical factor for the reduction of total water yield. And, the maximum mean monthly streamflow was estimated about 15.7 and 10.25 m^3/s at the outlet in May and October, respectively. In general, surface water potential within the ungauged watershed is adequate to create any pressure and gravity driven structures for the purpose of obtaining agrarian efficiency and for water supply. Therefore, hydrological data are essential for management and development of surface water resources. Likewise, it is necessary for water and land use managers, administrators, planners, builders, engineers, recreationists, and for all sectors. Besides, the daily, monthly, seasonal, and annual streamflow data are very useful for characterizing streamflow variability.

4.1. Limitations and directions for future research

This study considered only stepwise multiple linear regression coupled with PCA regionalization techniques. Thus, this study should be extended by comparing different regionalization methods followed by discussion and conclusion based on the results of different methods. Moreover, the results were based on limited SWAT model-sensitive parameters and physical catchment characteristics for the production of streamflow. However, if the physical catchment characteristics were increased the results become more perfect. Thus, the next study should consider several descriptors from different categories. Besides, only five river records were used for regionalization in this study. Consequently, future studies should consider several rivers gauging from different climatic zones and river basins for regionalization of parameters, and for validation of the transposed parameters. Furthermore, the stepwise linear regression analysis was done using the 5% probability for selecting a factor, which is performed using SPSS. Hence, this study should be extended by using other models/software or MCDM approaches instead of SPSS software.

ACKNOWLEDGEMENTS

The author would like to acknowledge the National Meteorological Agency; and Ministry of Water, Irrigation, and Energy (MoWIE) of Ethiopia for providing the Meteorological, Hydrological, Soil, and Land use landcover data.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Abbaspour, K. C., Rouholahnejad, E., Vaghefi, S., Srinivasan, R., Yang, H. & Kløve, B. 2015 A continental-scale hydrology and water quality model for Europe: Calibration and uncertainty of a high-resolution large-scale SWAT model. *Journal of Hydrology* **524**, 733–752. https://doi.org/10.1016/j.jhydrol.2015.03.027.

Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., Santhi, C., Harmel, R. D., Van Griensven, A., Van Liew, M. W., Kannan, N. & Jha, M. K. 2012 SWAT: Model use, calibration, and validation. *Transactions of the Asabe* **55** (4), 1491–1508.

Arsenault, R., Breton-Dufour, M., Poulin, A., Dallaire, G. & Romero-Lopez, R. 2019 Streamflow prediction in ungauged basins: analysis of regionalization methods in a hydrologically heterogeneous region of Mexico. *Hydrological Sciences Journal* **64** (11), 1297–1311. https://doi.org/10.1080/02626667.2019.1639716.

Ashine, E. 2021 Modeling surface water potential of somodo watershed. *Irrigat Drainage Sys Eng* **10**, 290.

Beck, H. E., van Dijk, A. I. J. M., de Roo, A., Miralles, D. G., McVicar, T. R., Schellekens, J. & Bruijnzeel, L. A. 2016 Global-scale regionalization of hydrologic model parameters. *Water Resources Research* **52** (5), 3599–3622. https://doi.org/10.1002/2015WR018247.

Beza, M., Hailu, H. & Teferi, G. 2023 Modeling and Assessing Surface Water Potential Using Combined SWAT Model and Spatial Proximity Regionalization Technique for Ungauged Subwatershed of Jewuha Watershed, Awash Basin, Ethiopia. 2023.

Burgan, H. I. 2022 Comparison of different ANN (FFBP, GRNN, RBF) algorithms and multiple linear regression for daily streamflow prediction in Kocasu River, Turkey. *Fresenius Environment bulletin* **31**, 4699–4708.

Daniel, H. 2023 Hydrological response to climate change in the Deme watershed, Omo-Gibe Basin, Ethiopia. *Journal of Water and Climate Change* **14** (4), 1112–1131. https://doi.org/10.2166/wcc.2023.248.

Daniel, H. & Abate, B. 2022 Effect of climate change on stream flow in the gelana watershed, rift valley basin, Ethiopia. *Journal of Water and Climate Change* **13** (5), 1–28. https://doi.org/10.2166/wcc.2022.059.

Demmissie, N. G., Demissie, T. A. & Tufa, F. G. 2018 Predicting the impact of climate change on Kulfo River flow. *Hydrology* **6** (3), 78. https://doi.org/10.11648/j.hyd.20180603.11.

Dibaba, W. T., Demissie, T. A. & Miegel, K. 2020 Watershed hydrological response to combined land use/land cover and climate change in highland Ethiopia: Finchaa catchment. *Water (Switzerland)* **12** (6). https://doi.org/10.3390/w12061801

Gajbhiye, S. & Sharma, S. K. 2015 Application of Principal Components Analysis for Interpretation and Grouping of Water Quality Parameters. April. https://doi.org/10.14257/ijhit.2015.8.4.11

Gassman, P. W., Sadeghi, A. M. & Srinivasan, R. 2014 Applications of the SWAT model special section: Overview and insights. *Journal of Environmental Quality* **43** (1), 1–8. https://doi.org/10.2134/jeq2013.11.0466.

Guo, Y., Zhang, Y., Zhang, L. & Wang, Z. 2021 Regionalization of hydrological modeling for predicting streamflow in ungauged catchments: A comprehensive review. *Wiley Interdisciplinary Reviews: Water* **8** (1), 1–32.

Hu, T., Wu, F. & Zhang, X. 2007 Rainfall-runoff modeling using principal component analysis and neural network. 235–248. https://doi.org/10.2166/nh.2007.010

Jolliffe, I. T., Cadima, J. & Cadima, J. 2016 *Principal Component Analysis: A Review and Recent Developments*.

Moriasi, D. N., Arnold, J. G., Liew, M. W. V., Bingner, R. L., Harmel, R. D. & Veith, T. L. 2007 Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE* **50** (3), 885–900.

Nannawo, A. S., Lohani, T. K. & Eshete, A. A. 2021 Exemplifying the effects using WetSpass model depicting the landscape modifications on long-term surface and subsurface hydrological water balance in Bilate Basin, Ethiopia. *Advances in Civil Engineering* **2021**, 1–20.

Rientjes, T. H. M., Perera, B. U. J., Haile, A. T., Reggiani, P. & Muthuwatta, L. P. 2011 Regionalization for lake level simulation – The case of Lake Tana in the Upper Blue Nile, Ethiopia. *Hydrology and Earth System Sciences* **15** (4), 1167–1183. https://doi.org/10.5194/hess-15-1167-2011

Santhi, C., Arnold, J. G., Williams, J. R., Dugas, W. A., Srinivasan, R. & Hauck, L. M. 2001 Validation of the SWAT model on a large river basin with point and non-point sources. *Journal of the American Water Resources Association* **37** (5), 1169–1188.

Shahid, M., Cong, Z. & Zhang, D. 2018 Understanding the impacts of climate change and human activities on streamflow: A case study of the Soan River basin, Pakistan. *Theoretical and Applied Climatology* **134**, 205–219.

Shahid, M., Rahman, K. U., Haider, S., Gabriel, H. F., Khan, A. J., Pham, Q. B. & Anh, D. T. 2021 Quantitative assessment of regional land use and climate change impact on runoff across Gilgit watershed. *Environmental Earth Sciences* **80**, 1–18.

Shanka, A. S. 2017 Evaluation of climate change impacts on run-off in the Gidabo River Basin: Southern Ethiopia. *Environment Pollution and Climate Change* **01** (03), 1–6. https://doi.org/10.4172/2573-458x.1000129.

Sharma, S. K., Gajbhiye, S. & Tignath, S. 2015 Application of principal component analysis in grouping geomorphic parameters of a watershed for hydrological modeling. *Applied Water Science* **5** (1), 89–96. https://doi.org/10.1007/s13201-014-0170-1.

Swain, J. B. & Patra, K. C. 2015 Streamflow Estimation in Ungauged Catchments Using Regional Flow Duration Curve: Comparative Study. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001509.

Syed, T. H., Lakshmi, V., Paleologos, E., Lohmann, D., Mitchell, K. & Famiglietti, J. S. 2004 Analysis of process controls in land surface hydrological cycle over the continental United States. **109**, 1–11. https://doi.org/10.1029/2004JD004640.

Tegegne, G. & Kim, Y. O. 2018 Modelling ungauged catchments using the catchment runoff response similarity. *Journal of Hydrology* **564** (July), 452–466. https://doi.org/10.1016/j.jhydrol.2018.07.042.

Wuttichaikitcharoen, P. & Babel, M. S. 2014 Principal component and multiple regression analyses for the estimation of suspended sediment yield in ungauged Basins of Northern Thailand. 2412–2435. https://doi.org/10.3390/w6082412.

Zhang, Y., Chiew, F. H. S., Li, M. & Post, D. 2018 Predicting runoff signatures using regression and hydrological modeling approaches. *Water Resources Research* **54** (10), 7859–7878. https://doi.org/10.1029/2018WR023325.

First received 6 September 2023; accepted in revised form 13 December 2023. Available online 26 December 2023