Ethiopian Journal of Water Science and Technology (EJWST)

DOI: https://doi.org/10.59122/EJWST657

Vol. 8, 2025, pages: 95~118 ISSN (Print): 2220-7643

Comparative Evaluation of Model Sensitivity, Calibration, and Parameter Uncertainty in Streamflow Simulation Using SWAT and HBV Light in the Geba Watershed, Ethiopia

Abebe Temesgen Ayalew[1], Kinfe Bereda Mirani[2], Yohannes Mehari Andiye[1],

[1]Faculty of Hydraulic and Water Resources Engineering, Arba Minch Water Technology Institute, Arba Minch University, P.O. Box 21, Arba Minch, Ethiopia
[2] Faculty of Water Supply and Environmental Engineering, Arba Minch Water Technology Institute, Arba Minch University, P.O. Box 21, Arba Minch, Ethiopia

Corresponding Author: Yohannes Mehari Andiye; yohannes.mehari@amu.edu.et orcid: 0009-0007-7153-5474

Abstract

Evaluating and simulating streamflow is extremely useful for managing water resources in almost all regions, particularly in arid ones. This study focuses on the performance of two hydrological models, HBV Light and SWAT, in streamflow simulation in northern Ethiopia. The models were evaluated using an ensemble modeling approach, which integrated Monte Carlo simulations for HBV Light and SWAT, while SWAT-CUP was used for calibration and validation. During the sensitivity analysis, key parameters controlling the model outputs were identified. For HBV Light, the parameters, K2, MAXBAS, and BETA, reflect subsurface processes, whereas SWAT, CN2, GWQMN, and SOL_AWC were used to control surface runoff. During calibration and validation, SWAT demonstrated statistically superior performance in modeling streamflow (R^2=0.73, NSE=0.81) and (R^2 =0.72, NSE=0.72) respectively. While HBV Light recorded a performance of (R^2=0.71, NSE=0.70) during calibration and (R^2= 0.71, NSE= 0.71) during validation,which was closer to the observed streamflow. This indicates during the validation phase, SWAT still performed better but HBV Light demonstrated narrower predictive uncertainties at 95% along with more identifiability of the parameters that reduced the problem of equifinality. The bottom line of this case was that SWAT was statistically better, while HBV Light was more transparent and reliable with uncertainties. All things considered, both models could simulate streamflow, but their differences suggested that context-based choice would be optimal. While predictive consistency and uncertainty portrayal were most important, HBV Light was advantageous, and SWAT was better suited for use in cases that rank calibration accuracy above all else. The hydrology of the catchment could be better understood if streamline decision-making for the effective and sustainable management of water resources in the Geba Catchment and similar semi-arid areas were combined, or if multi-model ensembles were utilized.

Keywords: Streamflow simulation, Hydrological modeling, HBV Light model, SWAT model, Uncertainty analysis, Parameter sensitivity

Received: 11 September, 2025; Accepted: 10 November, 2025 Published: 22 December, 2025

1. INTRODUCTION

Hydrological modeling plays a fundamental role in understanding the dynamics of catchment processes and supporting water resource management across diverse climatic regions. Globally, rainfall–runoff models have evolved from empirical formulations to sophisticated process-based tools that integrate climatic, hydrological, and land surface interactions. These models are instrumental in flood prediction, water allocation, and impact assessment under changing climate conditions (Ayele & Ayalew, 2024; Jehanzaib et al., 2022; S. Mekonnen et al., 2023; Mosbahi et al., 2023; Onyutha, 2022, 2024). In Ethiopia, their application has expanded to evaluate runoff behavior in data-scarce basins where hydro-climatic variability and land-use changes are pronounced (Mekonnen & Manderso, 2023; Mekoya & Workneh, 2024; Nonki et al., 2023; Vogeti et al., 2023). However, model reliability remains challenged by limited data availability, structural complexity, and parameter uncertainty issues that this study addresses through a comparative evaluation of HBV Light and SWAT models within the Geba Catchment (Asnake et al., 2021; Haleem et al., 2022; Leta et al., 2021; Lv et al., 2022; Worku et al., 2022). The rainfall runoff model is helpful to study the non-linear behavior of the catchment.

In Ethiopia, there are a large number of hydrological modeling works in different basins under different climates, land use, soil, geology, and topographical settings (Emiru et al., 2022; Lv et al., 2022; Magidi et al., 2021). The applied models have different structures to simulate discharge. Their ability to simulate discharge in a comparative approach has not yet been explored. Furthermore, identifying better model conceptualization, the parsimony issue, identifiability of parameters, and predictive ability are the key points in understanding catchment behavior at various spatio-temporal scales. Comparative modeling studies thus would enable us to gain an insight into the weaknesses and strengths of the models in capturing the various response mode of the hydrograph by evaluating the credibility of the model through better process representation, parsimony, identifiability, and uncertainty assessment.

Moreover, in the study area there are no further researches were aimed to evaluate the performance of the model under realistic simulations. Thus, Geba Catchment is selected because of the availability of data to conduct comparative catchment analysis using HBV light versus SWAT models for better understanding of rainfall - runoff relationship in the sub-basin. Numerous hydrological models have been used to simulate the rainfall-runoff process (Dash

et al., 2021; Emiru et al., 2022; Lv et al., 2022; Magidi et al., 2021; Santos et al., 2022). As catchments vary, the applications of the models show variations. In this regard, comparative modeling studies would help to find the appropriate model for better representation of the catchment situation. Knowledge of catchment behavior and effect assessment are essential in data-poor regions such as the Tekeze Basin.

The structural formulation of the models in semi-arid basins like the Geba Watershed is a crucial determinant of the reliability of the simulation. As the conceptual model of lump is linked with the idea of hydrology, the HBV Light Model uses the components of storage by highlighting the processes of subsurface flow and recession. This implies its applicability in data-scarce regions with high groundwater-surface water interactions. On the other hand, the SWAT Model has a semi-distributed structure that splits the basin into sub-watersheds and Hydrological Response Units (HRUs) thereby representing spatial variability in land use, soil, and topography explicitly. This structure comparison affects sensitivity and uncertainty of model performance, and HBV Light has better performance to contain temporal flow range due to few data. SWAT is capable of representing the complexity of surface processes by explicitly accounting for their spatial variability in a detailed and distributed manner. Therefore, comparison shows the impact of model architecture on the representation of processes in semi-arid hydrological conditions.

2. MATERIALS AND METHODS

2.1 Study Area Description

Geba Catchment is situated in Tigray Regional State and drains the northeastern portion of the Tekeze River Basin. The upper portion of the catchment, which encompasses around 2437.52 km^2, is the main focus of this study. The geographical location of the study area is between 13°20′49″ and 14°10′13″ latitudes and 39°21′00″ and 39°46′30″ longitudes East (Figure 1). The basin exhibits a significant elevation gradient, ranging from a minimum of 1740 m to a maximum of 3300 m above sea level. (Ahmad et al., 2023). Geological formations and erosion factors greatly influence the topography of the basin. The main rivers have steep hills and rugged escarpments.

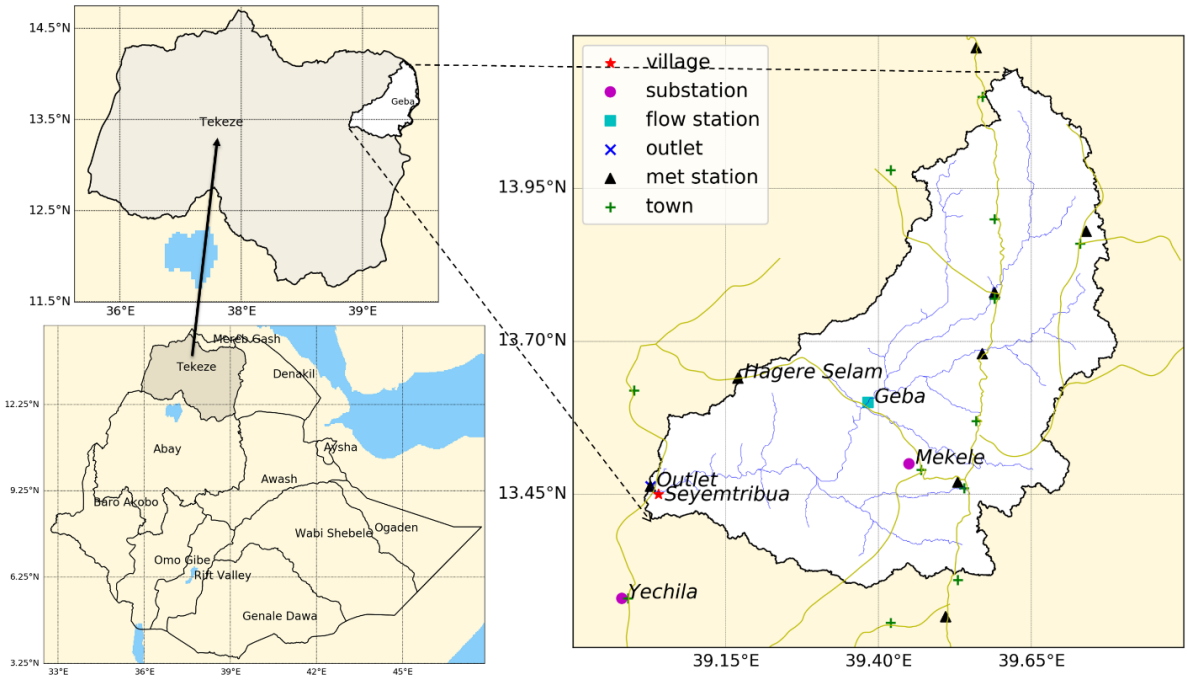


Figure 1. Location map of the Geba catchment and meteorological stations

Note: (met station and sub-station in the legend stand for meteorological stations and secondary observation points, respectively.

2.2 Data Collection

The hydrological and meteorological datasets utilized in this study were obtained from the Ministry of Water and Energy and the Ethiopian Meteorological Institute. The dataset comprises streamflow records from 2000 to 2023 and meteorological data spanning 1992 to 2023. In addition, the Ethiopian Mapping Agency supplied the Digital Elevation Model (DEM), Land Use/Land Cover (LULC), and soil maps which served as essential spatial inputs for the analysis. The LULC map used in this study represents the year 2020 and includes six primary land cover classes: cultivated land, grassland, shrub land, forest, urban areas, and barren land. The DEM (30 m) and soil maps (FAO classification) were integrated with the LULC data as essential spatial inputs for both models.

2.3 Data Quality and Resolution

The input datasets used for both models were rigorously screened and standardized for spatial and temporal consistency. Meteorological data (rainfall, temperature, and potential evapotranspiration) were available at a daily time step for the period 1992–2023. Spatial inputs, Digital Elevation Model (DEM, 30 m resolution), soil map (FAO classification), and Land Use/Land Cover (LULC, 2020 classification), were resampled to a uniform grid resolution

compatible with the SWAT model. Streamflow data from 2000–2023 were used for calibration and validation. Data reliability was verified using double mass curve analysis and statistical outlier detection. These steps ensured that both temporal and spatial data quality should meet the standards required for robust hydrological modeling in data-limited contexts.

2.4 Rainfall Data Screening

First, a visual review of daily rainfall data was used to screen meteorological stations in the research area for rough rainfall data. This initial step focused on identifying obvious data entry errors or non-physical rainfall values (e.g., negative rainfall). Next, a statistical outlier analysis was performed on the daily and monthly rainfall records. Specifically, the Inter-quartile Range (IQR) method was employed to flag extreme values (outliers) within each station. These flagged outliers were then cross-checked with records from adjacent stations to determine if they represented true, extreme events or if they were errors. Data filtering was carried out from 1992 to 2023. This was due to extensive gaps in rainfall records at some stations and the absence of long, overlapping periods of observation. To screen for inconsistencies and non-homogeneity among stations, the Double Mass Curve analysis was utilized. This involved plotting the cumulative rainfall of each station against the concurrent cumulative rainfall of a strong, consistent base station or the average of a group of reliable stations. Any significant change in the slope of the curve indicated a non-homogeneity or inconsistency in the data of the station. These inconsistencies may arise from factors such as station relocation, instrument changes, or long-term systematic errors. These issues were corrected as described in the data analysis section.Time series plotting of the daily rainfall was used to create a graphic comparison of the rainfall data, further highlighting temporal consistency. Figure 2 illustrates the long-term yearly rainfall computed over the period 1992-2023. It clearly shows strong seasonal variability in rainfall distribution. The highest rainfall occurred from June to September while smaller rainfall records were observed from March to May. Very low rainfall is observed during the dry months between October and February. This seasonal pattern reflects the dominant summer monsoon system that controls rainfall in the Geba Catchment.

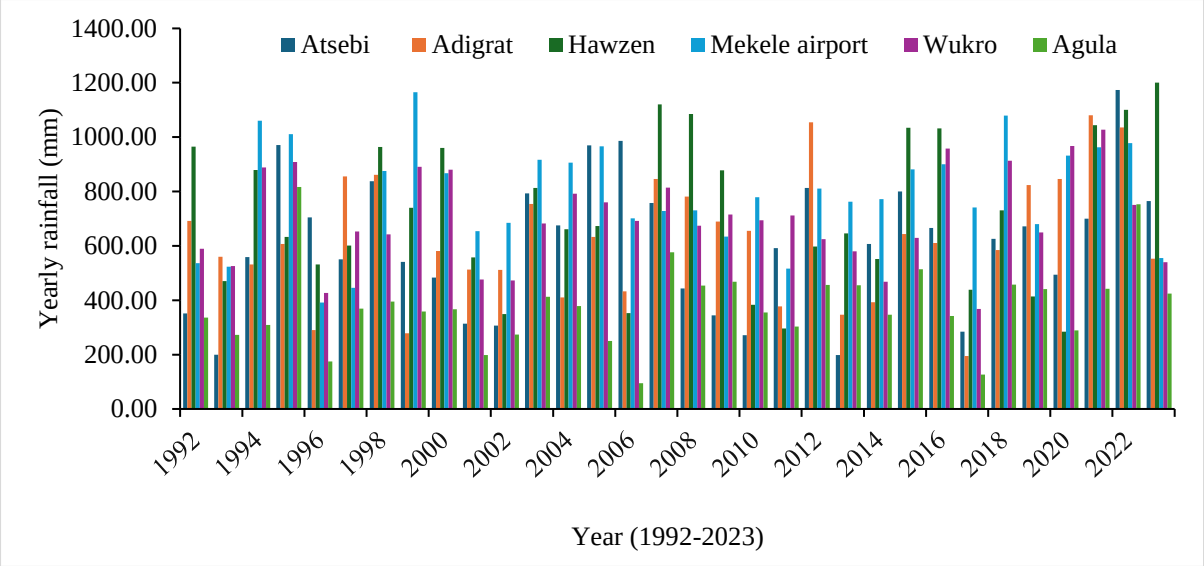


Figure 2. Yearly rainfall at selected rainfall stations

2.5 Data Analysis

The normal ratio approach and station average were utilized to fill the gaps in all stations. The homogeneity and consistency of rainfall were examined using a double mass curve which was also utilized to correct conflicting data. For reference, evapotranspiration (ETo) was estimated using the Penman–Monteith method, which is regarded as the standard approach when conventional meteorological variables such as solar radiation, wind speed, temperature, and humidity are available (Daniel & Abate, 2022). Nevertheless, not every station in this research region had access to such data. The Hargreaves technique was used to assess potential evapotranspiration since most stations contain both maximum and minimum temperature. The Thiessen Polygon Method was employed to convert the point rainfall measurements from the selected meteorological stations into areal rainfall inputs. This approach ensures that the rainfall input for each sub-basin accurately reflects the weighted average of the contributing stations.

2.6 Model Description and Mathematical Formulation

HBV Light Model:

The HBV Light Model is a conceptual rainfall–runoff model that represents the hydrological cycle using linked storage components for snow, soil moisture, groundwater, and routing. The model simulates runoff based on the following core relationships:

$$P = P_{eff} + \Delta S \quad (1)$$

$$Q = K_0 S_1 + K_1 S_2 + K_2 S_3 \quad (2)$$

where P_{eff} is the effective precipitation, ΔS is the change in storage, Q is total runoff, and K_0, K_1, K_2 and K_3 are recession coefficients for upper and lower zones. The terms S_1, S_2 and S_3 represent storage components of the system at different time levels. The nonlinear soil moisture relationship is controlled by the BETA parameter which determines the partitioning between infiltration and quick runoff.

SWAT Model:

The SWAT model is a semi-distributed, process-based hydrological model that divides a basin into sub-basins and Hydrological Response Units (HRUs). The water balance in each HRU is expressed as:

$$SW_t = SW_o + \sum_{i-1}^{t}(R_{day} - Q_{surf} + E_a - W_{seep} - Q_{gw}) \quad (3)$$

where SW_t is final soil water content, SW_0 is initial soil water content, t is time in days, R_{day} is the amount of precipitation on day i, Qsurf is the amount of surface runoff on day i, Ea is the amount of evapotranspiration on day i, Wseep is amount of water entering the vadose zone and Qgw is the amount of return flow on day i, all are in mm of water.

2.7 Model Sensitivity Analysis

The HBV-Light model sensitivity analysis was conducted using a Monte Carlo simulation approach, whereas sensitivity analysis for the SWAT model was performed automatically using the SUFI-2 algorithm within SWAT-CUP. The impact of each model parameter was assessed by systematically exploring its predefined range through continuous variation. Parameters were typically perturbed within ±20% of their initial best estimates, or across their full permissible ranges, to evaluate their influence on the objective function. This approach allows robust assessment of model sensitivity by quantifying the response of simulated outputs to parameter uncertainty.

2.8 Time Period Specified: Calibration and Validation

The model calibration for both models is conducted from 2000 to 2023. When the agreement between observed and simulated flow is good, model parameter and model performance evaluation factor are fixed.

2.9 Model Performance

For the calibration and validation periods of this work, the model performance was assessed using different model evaluation criteria (Al-Taei et al., 2023). The integrated assessment of the study involved three complementary stages: sensitivity analysis, calibration and validation, and uncertainty quantification which is performed independently for both models and then comparatively analyzed. Performance measures used in calibration and validation include theCoefficient of Determination (R^2), Nash–Sutcliffe Efficiency (NSE), PBIAS, and calculated as follows (Nash & Sutcliffe, 1970 , Santhi *et al*., 2001):

$$R^2 = \frac{\left[\sum_{i=1}^{n}(O_i - O_{ave})(S_i - S_{ave})\right]}{\sum_{i=1}^{n}(O_i - O_{avg})^2 \sum_{i=1}^{n}(S_i - S_{ave})^2} \quad (4)$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(O_i - S_i)^2}{\sum_{i=1}^{n}(O_i - O_{ave})^2} \quad (5)$$

$$PBIAS = 100\left[\frac{\sum_{i=1}^{n}(O_i - S_i)}{\sum_{i=1}^{n}O_i}\right] \quad (6)$$

where O_i is observed stream flow, O_{avg} is average observed Stream flow, S_i is simulated Stream flow, S_{ave} is average Simulated Stream flow and n is total number of observations

This explicit formulation provides clarity on the physical and mathematical bases of both models and their integration through sensitivity, calibration, and uncertainty evaluations.

2.10 Uncertainty Analysis for Both Models

Model simulations are inherently subject to uncertainty arising from parameter selection, model structure, performance, and calibration procedures. To quantify these uncertainties, uncertainty analyses were conducted using Monte Carlo simulation for the HBV-Light model and the SUFI-2 algorithm implemented in SWAT-CUP for the SWAT model (Angelakis et al., 2023; Azizi et al., 2021). Model uncertainty was evaluated using the p-factor, defined as the percentage of observed data bracketed by the 95% prediction uncertainty, and the r-factor,

representing the average thickness of the uncertainty band. An optimal model performance is characterized by a higher p-factor combined with a lower r-factor, reflecting a balance between predictive reliability and precision; these metrics were therefore used as key performance indicators in this study.

For the HBV-Light model, parameter and uncertainty analyses were performed using a Monte Carlo framework with 150,000 parameter sets to ensure comprehensive exploration of the parameter space and minimize the likelihood of overlooking optimal parameter combinations. This number of simulations was considered sufficient given the model's computational efficiency and is consistent with recommendations from previous hydrological studies for robust uncertainty characterization (Pianosi et al., 2016; Seibert, 2000). The resulting ensemble of simulations was used to derive upper and lower uncertainty bounds of the simulated streamflow, thereby capturing variability associated with different parameter choices (Ouallali et al., 2024; Sahu et al., 2023).

For the SWAT model, uncertainty analysis was conducted using the SUFI-2 algorithm within the SWAT-CUP framework, where parameter uncertainty was propagated through multiple calibration iterations; the number of iterations employed is reported in the results section. While parameter uncertainty plays a critical role in simulated streamflow, previous studies suggest that rainfall–runoff modeling may yield acceptable performance across a wide range of parameter sets, allowing the selection of an optimal solution when multiple parameter combinations provide reasonable fits to observed data (Mustafa et al., 2023; Ouallali et al., 2024; Sahu et al., 2023).

3. RESULTS AND DISSCUSSION

3.1 HBV Light Model

3.1.1 HBV Model Sensitivity

Throughout the simulation, the other model parameters were either insensitive or less sensitive than MAXBAS, BETA and K2 which were the most sensitive at Geba Catchment. Figure 3, Figure 4 and Figure 5 shows the subsurface or ground water dominance. This is the main process for the HBV-Light model because its K2 (recession or storage coefficient at box 2) is sensitive to all objective functions, unlike other processes.

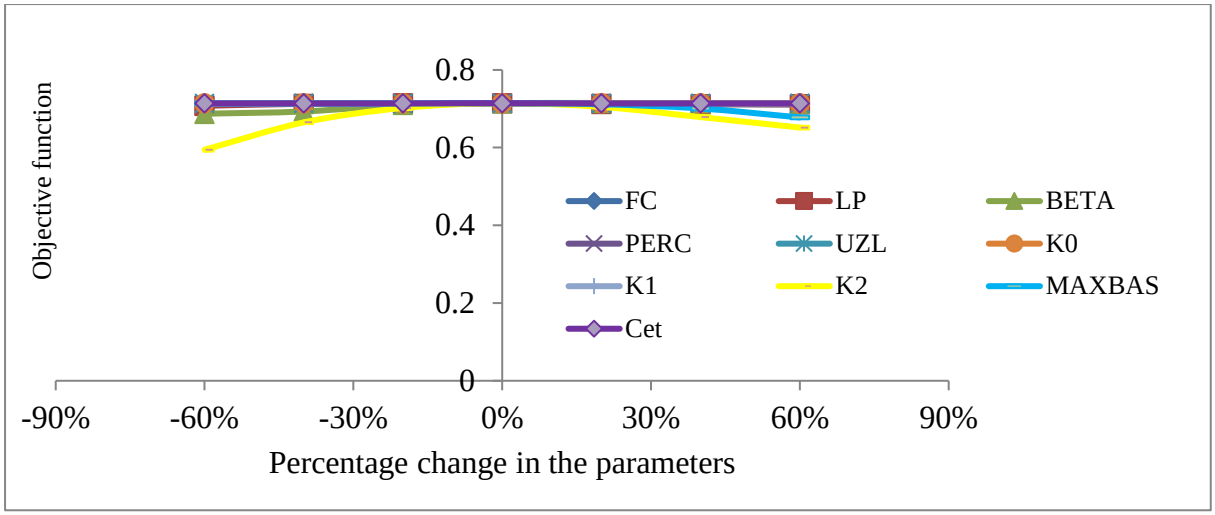


Figure 3. Percentage change in temperature versus R^2

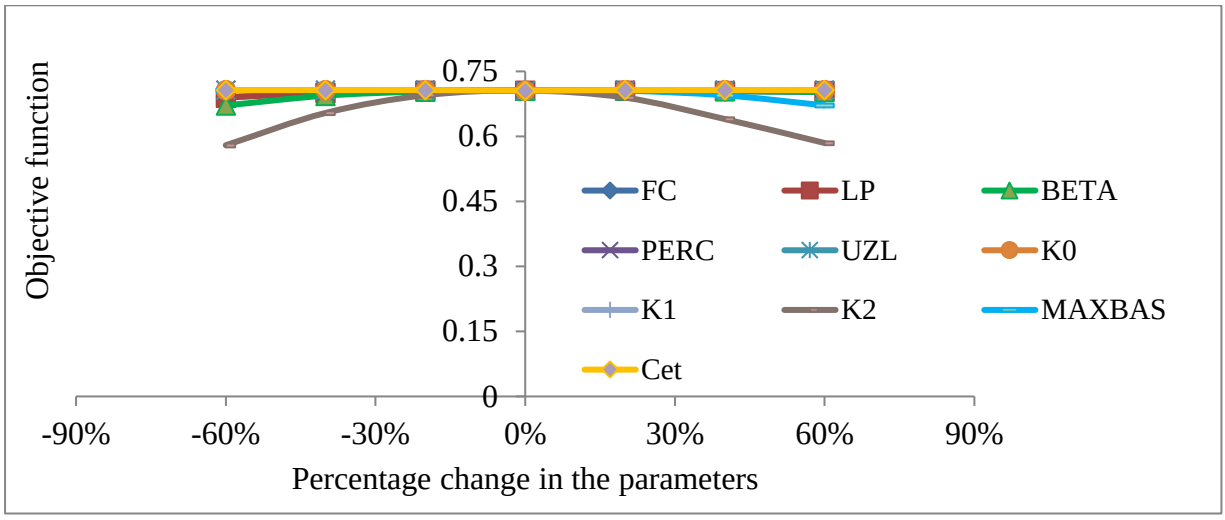


Figure 4. Percentage change in temperature versus NSE

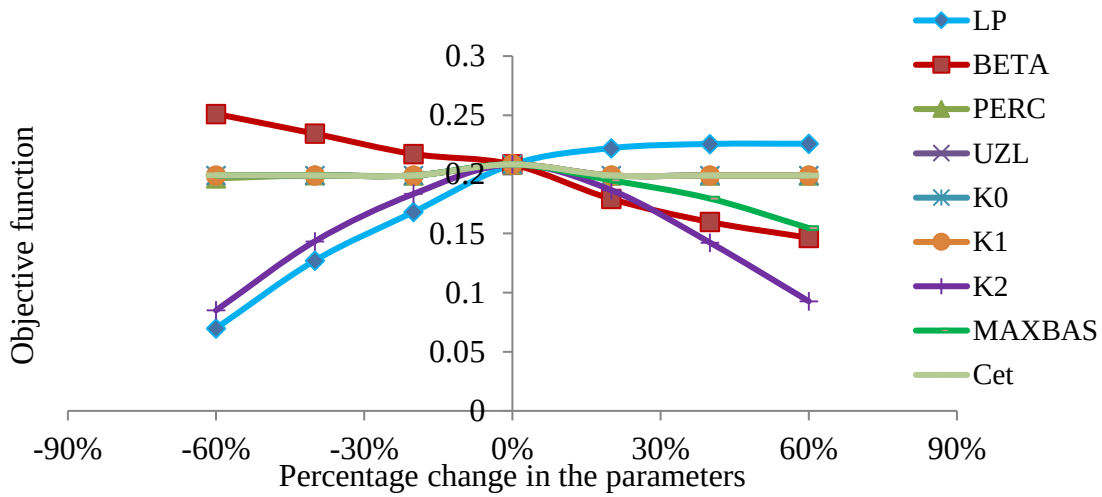


Figure 5. Percentage change in temperature versus relative volume error

3.1.2 HBV Model Simulation

The Geba Watershed model performance by HBV light model had been satisfactory for eighteen years (from 2000 to 2018) and from 2019 to 2023, with R^2 = 0.71, NSE = 0.707, and R^2 = 0.74, NSE = 0.73 for the calibration and validation period as shown in Figure 6 and Figure 7 respectively. The consistent underestimation of low flows during the validation period was likely related to the simplified groundwater representation in the HBV model. The model uses fixed recession coefficients that may not fully reflect the slow groundwater release observed in the Geba Catchment. In addition, limited low-flow observations increase uncertainty in calibrating parameters that control base flow and recession behavior. The optimized model parameter value for HBV Light model is presented in Table 1.

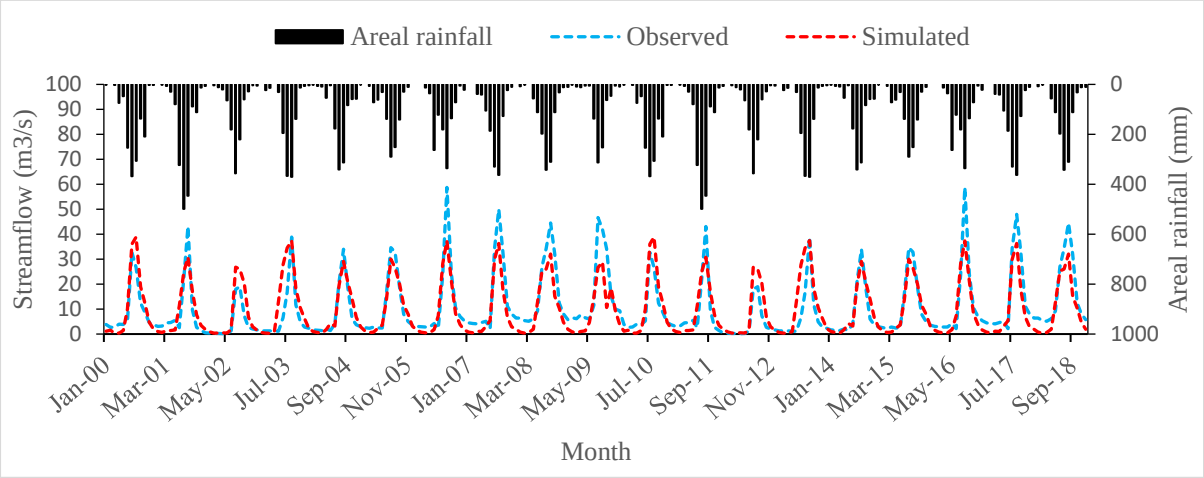


Figure 6. Flow simulation for calibration in HBV Light model

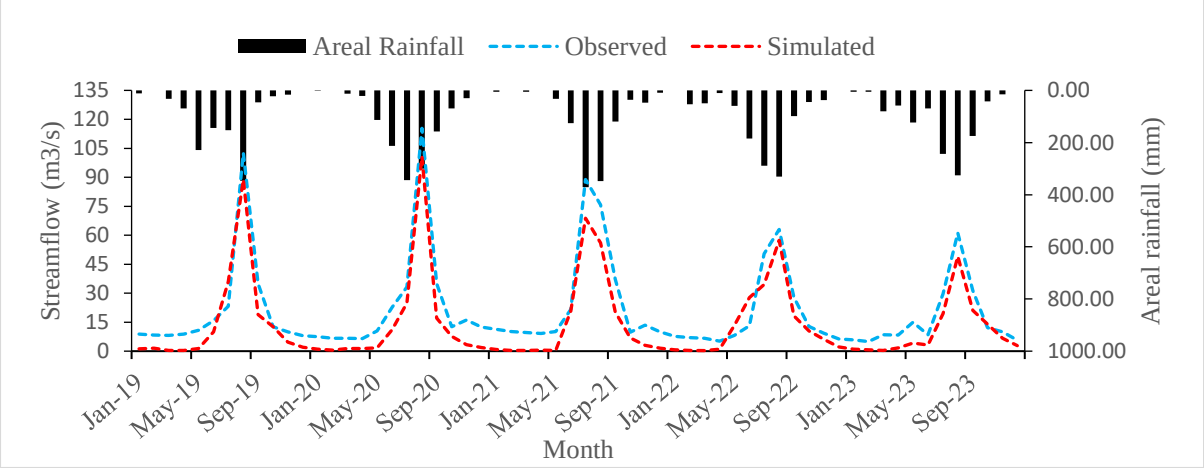


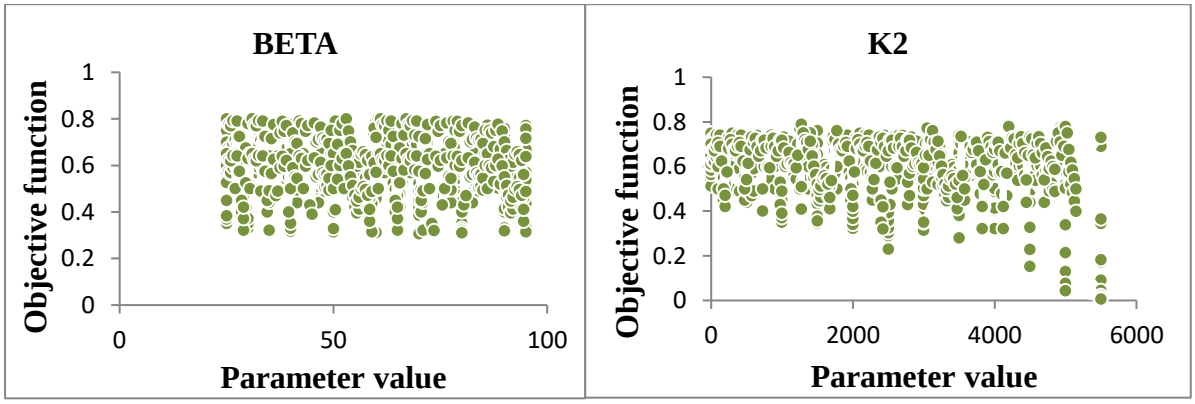
Figure 7. Flow simulation for validation in HBV Light model

Table 1. Summary of optimized model parameter values for HBV Light Model

Parameter	unit	Valid range	Optimized parameter value for calibration
FC	mm	(0,inf)	850
LP	_	[0,1]	0.8
BETA	_	(0,inf)	0.85
PERC	mm/Δt	[0,inf)	60
UZL	mm	[0,inf)	50
K0	1/Δt	[0,1)	0.85
K1	1/Δt	[0,1)	0.55
K2	1/Δt	[0,1)	0.65
MAXBAS	Δt	[1,100]	1
Cet	1/OC	[0,1]	0.01

3.1.3 HBV Model Parameter and Uncertainty Analysis

The calibration results of the HBV-Light model indicated the presence of parameter equifinality, whereby multiple parameter combinations produced comparable performance metrics, complicating the identification of a unique parameter set that best represents actual catchment processes. Despite this, the HBV-Light calibration exhibited relatively narrow uncertainty bounds and consistent parameter identifiability, suggesting reduced parameter ambiguity and a better-constrained equifinal behavior within the HBV framework (Figure 8). The concept of equifinality in hydrological modeling has been widely documented, with several studies reporting substantial equifinal and poorly constrained parameter sets during model calibration (Beven & Binley, 1992, Bammou et al., 2024; Ben Khélifa & Mosbahi, 2022; Bogale, 2021).



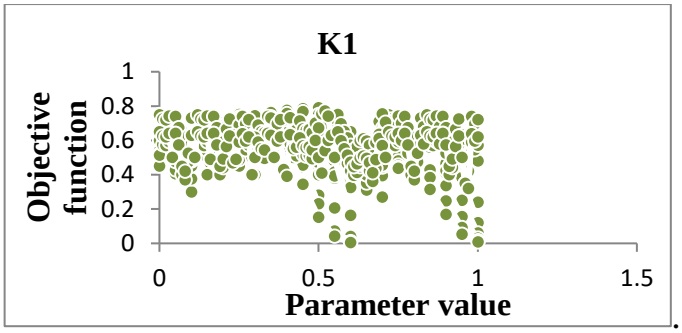


Figure 8. Dotty plot representing identifiability of model parameters in HBV Light Model

As shown in Figure 9 and 10, the majority of the simulated flow fell between the interval or uncertainty range. Only parameter uncertainty was taken into account in this investigation. As a result, the simulated outcome of the flow was trustworthy. Consistent with past studies addressing model uncertainty in Ethiopian basins (Bogale, 2021; Nonki et al., 2023; Ouallali et al., 2024), the majority of simulated flows in this study fell within the 95% predictive uncertainty range, confirming the reliability of the Monte Carlo-based uncertainty characterization. It was found that the simulation result was beyond the uncertainty range for the Muger Catchment Abay Basin in Ethiopia.

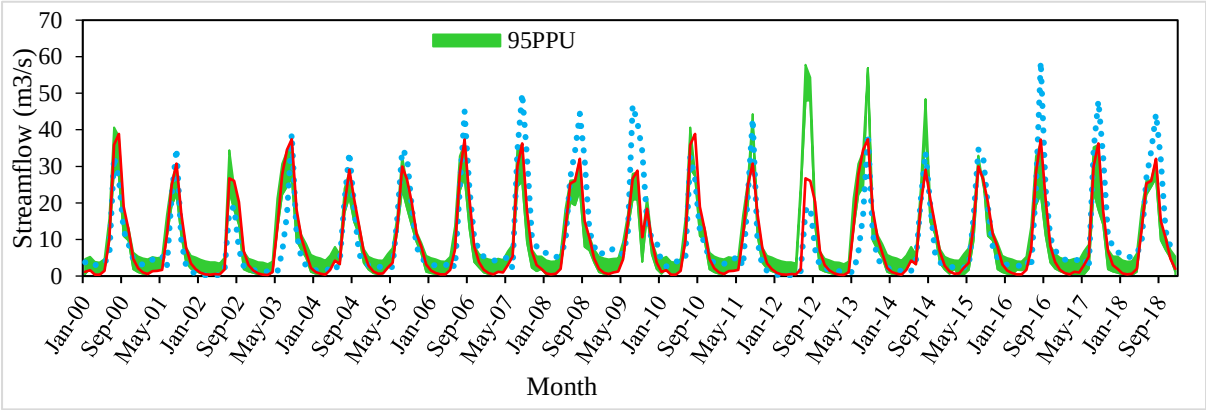


Figure 9. Uncertainty analysis for calibration period

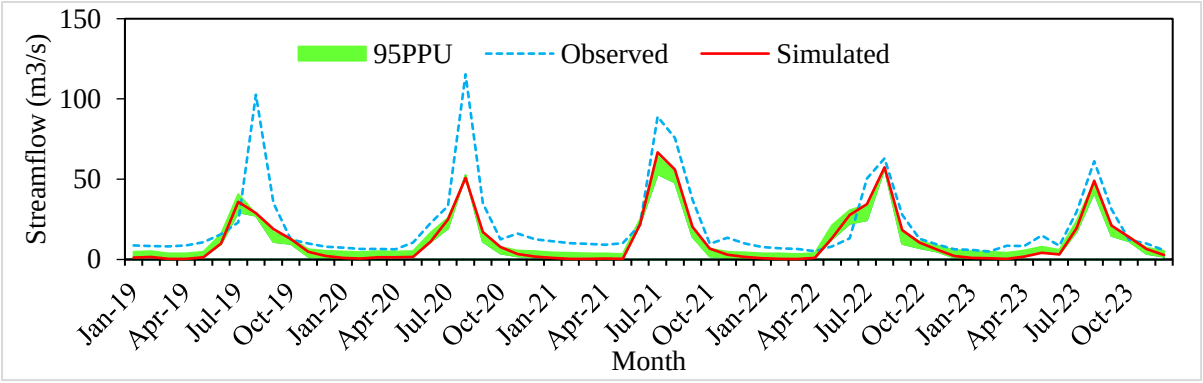


Figure 10. Uncertainty analysis for validation period

3.2 SWAT Model

3.2.1 SWAT Model Sensitivity

It is important to note that direct comparison of parameter sensitivity between SWAT and HBV Light is not based on numerical values as each model employs distinct structural and conceptual formulations. Instead, the comparison focuses on the relative influence of dominant parameters and the hydrological processes they represent. HBV-Light focuses on subsurface flow control, whereas SWAT emphasizes surface runoff generation. The most sensitive factors were land use and antecedent soil water conditions (CN2), followed by the soil characteristics of the watershed (SOL_AWC) and ground water determinant parameters for flow (GWQMN) (Figure 11). The delay time for aquifer recharging (GW_DELAY) and the depth of the soil layer from the top to the bottom (SOL_Z) was a soil parameter mainly controlling surface runoff generation whereas the remaining parameters were not significantly sensitive to runoff simulation.

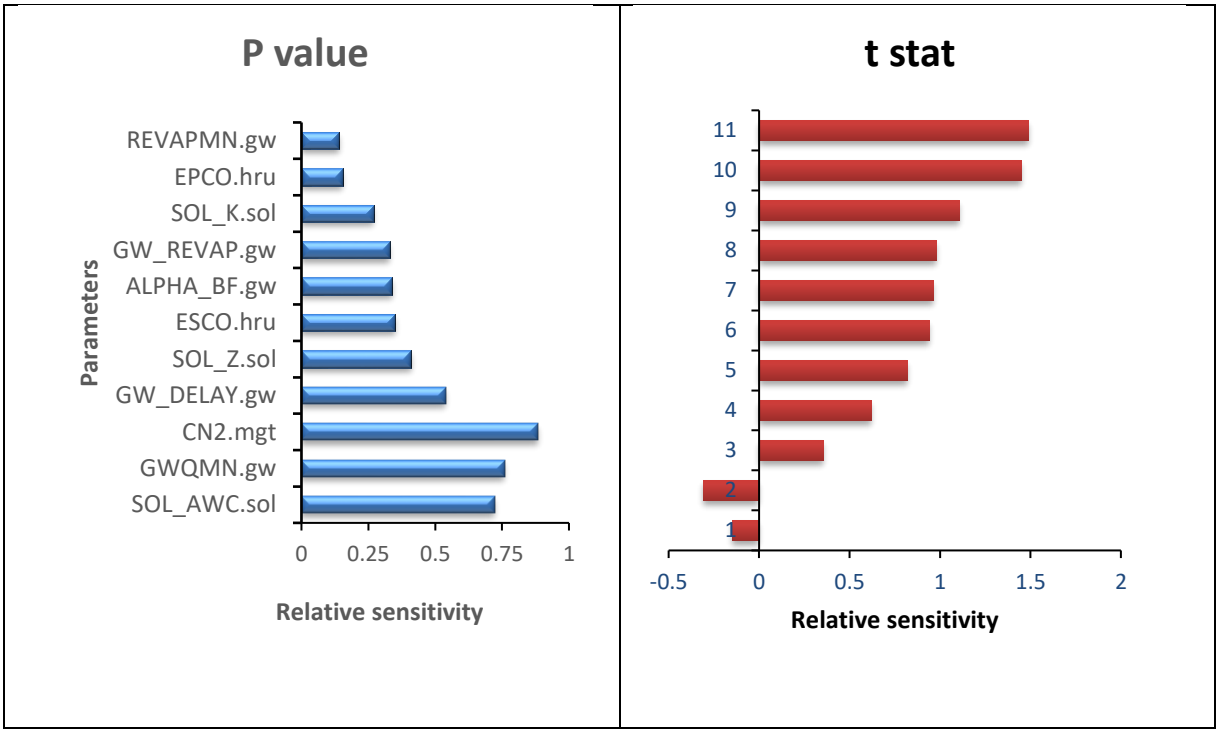


Figure 11. Model parameter sensitivity ranking

In Figure 11, the p-values denote the significance level associated with each t-statistic of the parameter in the sensitivity ranking. Parameters with p-values approaching zero (typically ≤ 0.05) are considered statistically significant. The initial figure presented rounded values; this was corrected to show precise significance levels, indicating that parameters CN2, SOL_AWC,

and GWQMN exhibited the highest sensitivity at 5% significance level. Surface dominance should be the parameter identified in the SWAT Model as land use and antecedent soil water conditions (CN2) were the most sensitive of the model parameters.

Note: RS- Relative sensitivity values of model parameters have a value; the t Stat gives a measure of sensitivity (larger absolute values are more sensitive); the p value establishes the significance of the sensitivity (a value near zero has greater significance); and "R_" and "V_" denote relative change and replacement to the initial parameter values, respectively. $0 \leq RS < 0.05$, $0.05 \leq RS < 0.2$, and $0.02 \leq RS < 1.0$ are considered small to negligible. $RS \geq 1.0$ is quite high.

3.2.2. SWAT Model Calibration and Validation

The monthly time-based statistical values for R^2, NSE, and PBIAS over eight years (January 1, 2000 to December 31, 2018) and the validation period (January 1, 2019 to December 31, 2023) were 0.81, 0.73, -11% and 0.72, 0.72, -11% for calibration and validation, respectively (Figure 12 and Figure 13). These results were comparable to those of the study by (Costea et al., 2022; Duc & Sawada, 2023; Esmaeilpour, 2024). By repeatedly altering the parameters 1500 times, the model was automatically calibrated. The value that fits is presented in Table 2.Following adjustments, the model test indicated that R^2, NSE, and PBIAS were 89.60%, 86.36%, and -8.16%,respectively. Consequently, the goal functions were met.

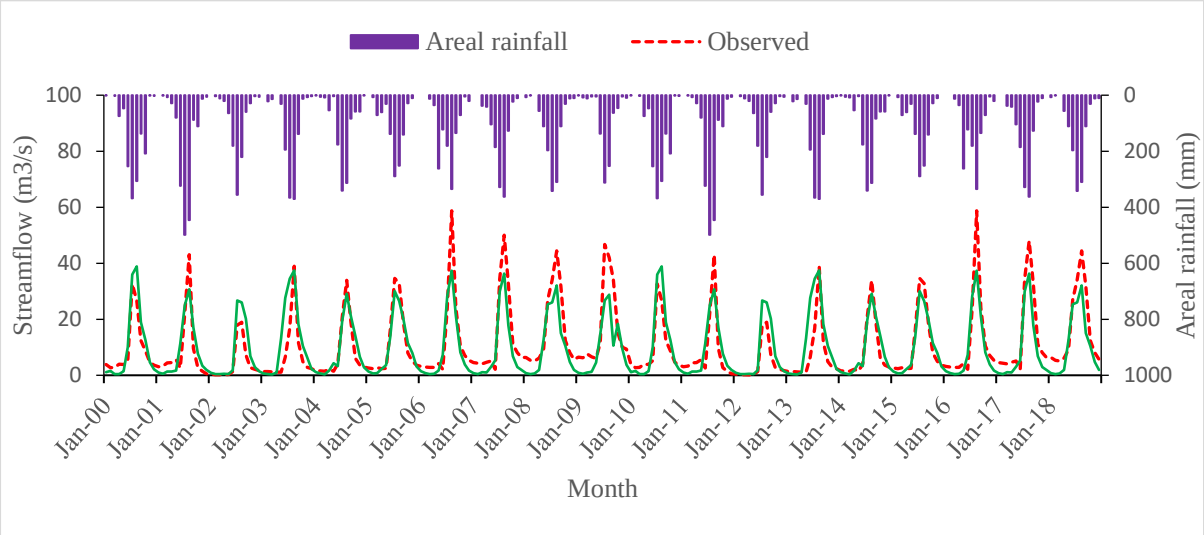


Figure 12. Observed and simulated flow hydrographs during calibration period

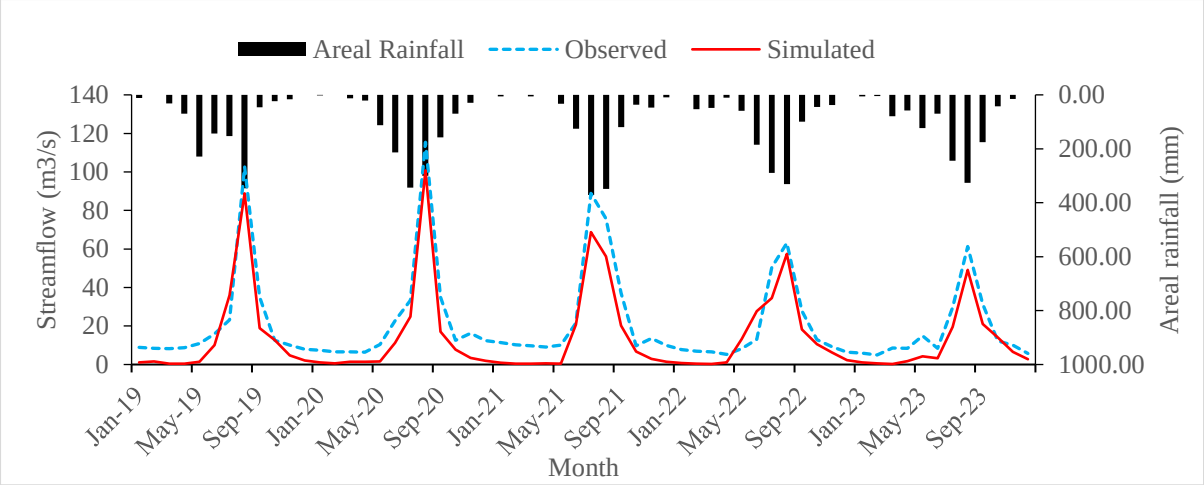


Figure 13. Observed and simulated flow hydrographs during validation period

Table 2. Summary of optimized model parameter values for SWAT Model

Parameter	Valid range	Optimized parameter value for calibration
ALPHA_BF	0-1	0.72
CN2	35-98	91
GWQMN	0-5000	700
ESCO	0-1	0.64
SOL_AWC	0-1	0.93
CANMAX	0-10	7.2
REVAPMN	0-500	445
GWREVAP	0.02-0.2	0.06
SOL_Z	0-3500	2315
SOL_K	0-2000	1645
GW_DELAY	0-500	265

3.2.3 SWAT Model Parameter and Uncertainty Analysis

The relationship between parameter values and the goal function is shown in Figure 14. This graph serves to illustrate the distribution of sample points and provide a sense of parameter sensitivity for the more sensitive parameters chosen during calibration. As in Figure 14, the sample points were dispersed throughout the objective function, leading to the best identifiability of parameters that fall within this range. The results were comparable to those of (Mathevet et al., 2023; Mustafa et al., 2023).

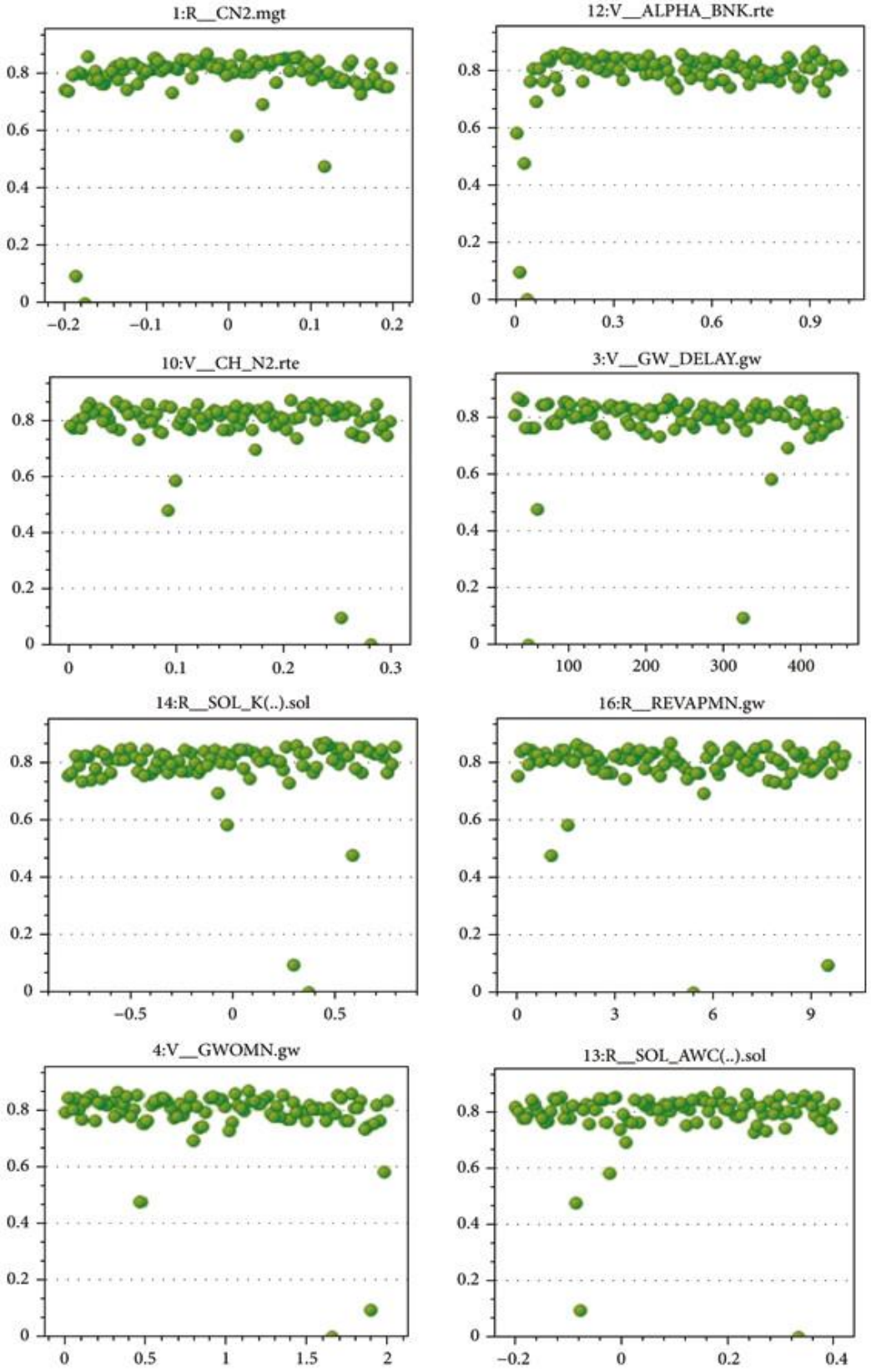


Figure 14 Sample dotty plot for selective sensitive parameters

N.B: The x-axis indicates parameter range and the y axis for objective function

The distribution of the sample points showed that the majority of the points were spread farther from the goal function as compared to the HBV model. This may imply that the HBV light model was less dependable in terms of parameter identifiability.

SUFI-2, 95PPUs, the p-factor, and the r-factor served as indicators of the uncertainty. The p-factor and r-factor of the calibration for monthly streamflow were 0.69 and 0.64, respectively. This showed that the 95% prediction interval (PPU) could bracket around 69% (out of a perfect 100%) of the recorded monthly stream flow with an extremely narrow 95% PPU band of 0.64 (near a perfect 0) during calibration (Ouallali et al., 2024). A well-balanced relationship between the p-factor and r-factor was maintained during the SUFI-2 calibration process. The obtained p-factor of 0.69 and r-factor of 0.64 reflect an optimal equilibrium, ensuring that approximately 70% of the observed flow values were bracketed within a narrow predictive uncertainty band, thereby confirming both precision and reliability of the calibrated model (Figure 15 and 16).

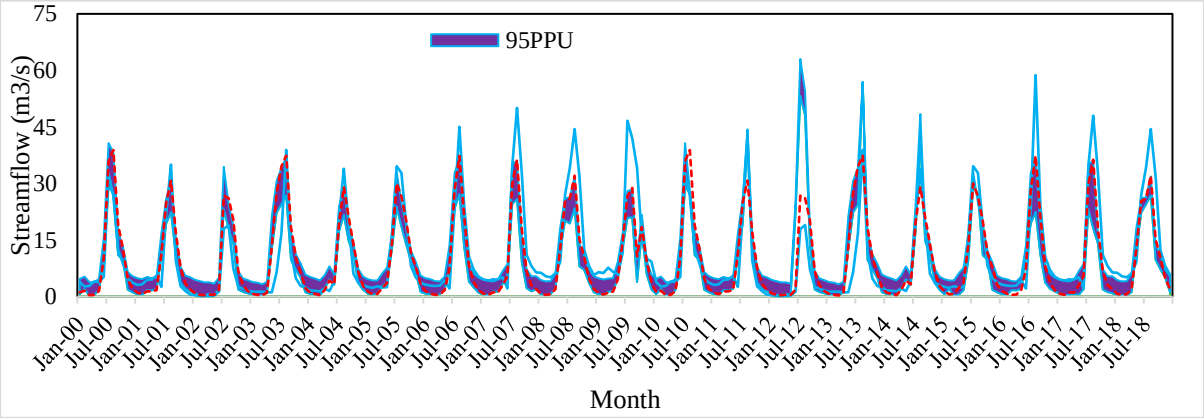


Figure 15 Uncertainty during the calibration period

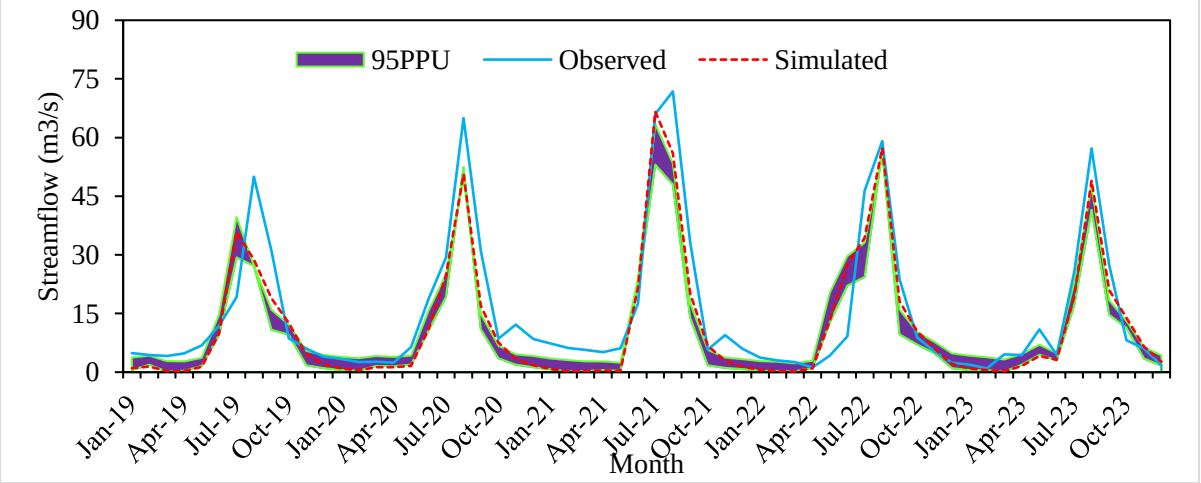


Figure 16 Uncertainty during the validation period

Table 3. Evaluation of the model performance of HBV Light and the SWAT Model

Streamflow Simulation	R^2	NSE	PBIAS (%)
HBV light model calibration	0.71	0.70	-4
HBV light model validation	0.71	0.71	-3
SWAT model calibration	0.81	0.73	-11
SWAT model validation	0.72	0.72	-11

The study period (2000–2023) was characterized by notable climate variability and gradual land use modification within the Geba Catchment. Although the models were calibrated to observed conditions, these inter-annual fluctuations influenced runoff generation patterns and model responses. Future studies should explicitly integrate climate and land use dynamics to evaluate their long-term influence on model robustness and parameter stability. Such extensions would enhance model applicability under changing environmental conditions.

Generally, the comparative evaluation showed the contrasting hydrological representations of the two models. While SWAT demonstrated stronger statistical efficiency and replicated observed hydrographs, it exhibited broader uncertainty ranges owing to greater parameter interdependence. Conversely, HBV Light showed narrower predictive intervals and enhanced identifiability, thereby reducing parameter equifinality. These findings highlighted that model choice should be context-driven: HBV Light offered superior reliability in data-limited basins emphasizing uncertainty quantification whereas SWAT was preferable when calibration precision and spatial process representation were prioritized.

4 CONCLUSION

A detailed comparative analysis of the HBV Light and SWAT model was done in the modelling of streamflow in the Geba Catchment. The evaluation was conducted using the best calibration methods and uncertainty tests to assess statistical effectiveness and predictive accuracy.. Sensitivity analysis showed parameter dominance that were opposite to those in HBV Light with the outputs being influenced mainly by subsurface processes (K2, MAXBAS, BETA) In SWAT, surface hydrological processes are represented by CN2, GWQMN, and SOL_AWC. The SWAT model shows good statistical performance with ENS values of 0.81 and 0.75 in calibration and validation, respectively. However, HBV Light, having ENS values of 0.70 and 0.73, indicated that SWAT fitted better to observed overall flows. This made SWAT especially appropriate in situations where an accurate reproduction of the observed streamflow was of

interest. Conversely, HBV Light had 95% predictive uncertainty margins and more reliable parameter identifiability. This largely neutralized equifinality problems which were eminent in SWAT stochastic calibration. These features highlighted a major trade-off. Although SWAT was doing much better in achieving more similar statistical approximation of the observations, HBV Light was much better in the predictability of the results. In other words, HBV Light representation of uncertainty was more transparent. The models differed from each other by demonstrating different behavior: HBV Light was characterized by the use of the groundwater processes, but SWAT was characterized by the surface runoff. Such difference demonstrated the nature of the complex hydrology of the Geba Catchment. Seasonal variations in land use and the presence of clay-based soils characterize the process of runoff generation. Although SWAT was statistically more accurate, its nature of underestimating low flows as well as overestimating peak flows along with the wider range of uncertainty of its prediction, could be restrictive to the application of this model in the long-term water resource planning. On the other hand, HBV Light had a significant strength in its ability to restrict the behavior of the parameters, and limited the parameters to slimmer predictive spaces, which made it especially useful in data-sparse contexts, and integrative watershed management. HBV-Light, however, has some limitations. Its performance can vary depending on catchment characteristics, and certain surface hydrological processes are not represented as accurately as in other models.In general, the study demonstrated that the process of model selection could be seen as an optimization between statistical performance, predictive reliability, and uncertainty transparency. It is better to use a more careful, context-based strategy: HBV Light is most appropriate when it is needed to make strong predictions in case of uncertainty, whereas SWAT can be more suitable in cases where the main goal is high calibration accuracy. Integrating the understanding of the two models or using multi-model ensembles may offer a better picture of the catchment hydrology, leading to the development of improved decisions in the sustainability of water resources and water management in the semi-arid areas. Implications of the findings also include the need to have strong data collection, careful model analysis, and adaptability of catchment hydrological dynamics in the application of conceptual hydrological models.

Acknowledgement: Not available

Author contributions: Abebe Temesgen Ayalew was involved in conceptualization, design, and material preparation, and wrote the main manuscript text. Kinfe Bereda Mirani and Yohannes

Mehari Andiye were involved in data collection, supervision, and revision of the first draft and revised draft. All authors reviewed the manuscript.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability: All data generated or analyzed during this study are included in this published article. Additional data is available on request.

Declarations: All authors have read, understood, and have complied with the statement "Ethical Responsibilities of Authors" as in the instructions for authors.

Competing interests: The authors declare no competing interests.

REFERENCES

Ahmad, M. N., Shao, Z., & Javed, A. (2023). Modelling land use/land cover (LULC) change dynamics, future prospects, and its environmental impacts based on geospatial data models and remote sensing data. *Environmental Science and Pollution Research*, *30*(12), 32985–33001. https://doi.org/10.1007/s11356-022-24442-2

Al-Taei, A. I., Alesheikh, A. A., & Darvishi Boloorani, A. (2023). Land Use/Land Cover Change Analysis Using Multi-Temporal Remote Sensing Data: A Case Study of Tigris and Euphrates Rivers Basin. *Land*, *12*(5). https://doi.org/10.3390/land12051101

Angelakis, A. N., Capodaglio, A. G., Valipour, M., Krasilnikoff, J., Ahmed, A. T., Mandi, L., Tzanakakis, V. A., Baba, A., Kumar, R., Zheng, X., Min, Z., Han, M., Turay, B., Bilgiç, E., & Dercas, N. (2023). Evolution of Floods: From Ancient Times to the Present Times (ca 7600 BC to the Present) and the Future. *Land*, *12*(6). https://doi.org/10.3390/land12061211

Asnake, K., Worku, H., & Argaw, M. (2021). Assessing the impact of watershed land use on Kebena river water quality in Addis Ababa, Ethiopia. *Environmental Systems Research*, *10*(1). https://doi.org/10.1186/s40068-020-00208-y

Ayele, M. A., & Ayalew, A. T. (2024). Modeling of watershed intervention techniques to rehabilitate sediment yield hotspot areas in Hare watershed, rift valley basin, Ethiopia. *Modeling Earth Systems and Environment*, *10*(1), 425–441. https://doi.org/10.1007/s40808-023-01808-0

Azizi, S., Ilderomi, A. R., & Noori, H. (2021). Investigating the effects of land use change on

flood hydrograph using HEC-HMS hydrologic model (case study: Ekbatan Dam). *Natural Hazards*, *109*(1), 145–160. https://doi.org/10.1007/s11069-021-04830-6

Bammou, Y., Benzougagh, B., Igmoullan, B., Ouallali, A., Kader, S., Spalevic, V., Sestras, P., Billi, P., & Marković, S. B. (2024). Optimizing flood susceptibility assessment in semi-arid regions using ensemble algorithms: a case study of Moroccan High Atlas. *Natural Hazards*, *120*(8), 7787–7816. https://doi.org/10.1007/s11069-024-06550-z

Ben Khélifa, W., & Mosbahi, M. (2022). Modeling of rainfall-runoff process using HEC-HMS model for an urban ungauged watershed in Tunisia. *Modeling Earth Systems and Environment*, *8*(2), 1749–1758. https://doi.org/10.1007/s40808-021-01177-6

Bogale, A. (2021). Morphometric analysis of a drainage basin using geographical information system in Gilgel Abay watershed, Lake Tana Basin, upper Blue Nile Basin, Ethiopia. *Applied Water Science*, *11*(7), 1–7. https://doi.org/10.1007/s13201-021-01447-9

Costea, A., Bilasco, S., Irimus, I. A., Rosca, S., Vescan, I., Fodorean, I., & Sestras, P. (2022). Evaluation of the Risk Induced by Soil Erosion on Land Use. Case Study: Guruslău Depression. *Sustainability (Switzerland)*, *14*(2), 1–19. https://doi.org/10.3390/su14020652

Daniel, H., & Abate, B. (2022). Effect of climate change on streamflow in the Gelana watershed, Rift valley basin, Ethiopia. *Journal of Water and Climate Change*, *13*(5), 2205–2232. https://doi.org/10.2166/wcc.2022.059

Dash, S. S., Sena, D. R., Mandal, U., Kumar, A., Kumar, G., Mishra, P. K., & Rawat, M. (2021). A hydrological modelling-based approach for vulnerable area identification under changing climate scenarios. *Journal of Water and Climate Change*, *12*(2), 433–452. https://doi.org/10.2166/wcc.2020.202

Duc, L., & Sawada, Y. (2023). A signal-processing-based interpretation of the Nash-Sutcliffe efficiency. *Hydrology and Earth System Sciences*, *27*(9), 1827–1839. https://doi.org/10.5194/hess-27-1827-2023

Emiru, N. C., Recha, J. W., Thompson, J. R., Belay, A., Aynekulu, E., Manyevere, A., Demissie, T. D., Osano, P. M., Hussein, J., Molla, M. B., Mengistu, G. M., & Solomon, D. (2022). Impact of Climate Change on the Hydrology of the Upper Awash River Basin, Ethiopia. *Hydrology*, *9*(1). https://doi.org/10.3390/hydrology9010003

Esmaeilpour, M. (2024). *Flood risk assessment using fuzzy logic and HEC-HMS model in Ojan Chay watershed , Northwest of Iran*. *13*(2), 65–91. https://doi.org/10.22067/geoeh.2023.82202.1359

Haleem, K., Khan, A. U., Ahmad, S., Khan, M., Khan, F. A., Khan, W., & Khan, J. (2022). Hydrological impacts of climate and land-use change on flow regime variations in upper Indus basin. *Journal of Water and Climate Change*, *13*(2), 758–770. https://doi.org/10.2166/wcc.2021.238

Jehanzaib, M., Ajmal, M., Achite, M., & Kim, T. W. (2022). Comprehensive Review: Advancements in Rainfall-Runoff Modelling for Flood Mitigation. *Climate*, *10*(10), 1–17. https://doi.org/10.3390/cli10100147

Leta, M. K., Demissie, T. A., & Tränckner, J. (2021). Modeling and prediction of land use land cover change dynamics based on land change modeler (Lcm) in nashe watershed, upper blue nile basin, Ethiopia. *Sustainability (Switzerland)*, *13*(7). https://doi.org/10.3390/su13073740

Lv, Y., Yang, D., Yuan, R., Yang, K., & Zhong, H. (2022). A novel multivariate signal processing-based fault diagnosis approach of rotating machinery under various operating conditions. *Measurement Science and Technology*, *33*(7). https://doi.org/10.1088/1361-6501/ac60d5

Magidi, J., Nhamo, L., Mpandeli, S., & Mabhaudhi, T. (2021). Application of the random forest classifier to map irrigated areas using google earth engine. *Remote Sensing*, *13*(5), 1–15. https://doi.org/10.3390/rs13050876

Mathevet, T., Le Moine, N., Andréassian, V., Gupta, H., & Oudin, L. (2023). Multi-objective assessment of hydrological model performances using Nash–Sutcliffe and Kling–Gupta efficiencies on a worldwide large sample of watersheds. *Comptes Rendus - Geoscience*, *355*(S1), 117–141. https://doi.org/10.5802/crgeos.189

Mekonnen, S., Dessie, M., Tadesse, A., Nega, H., & Zewdu, A. (2023). Predicting the daily flow in ungauged catchments of the eastern part of the upper Blue Nile basin, Ethiopia. *Sustainable Water Resources Management*, *9*(3), 1–16. https://doi.org/10.1007/s40899-023-00856-9

Mekonnen, Y. A., & Manderso, T. M. (2023). Land use/land cover change impact on streamflow using Arc-SWAT model, in case of Fetam watershed, Abbay Basin, Ethiopia. *Applied Water Science*, *13*(5), 1–19. https://doi.org/10.1007/s13201-023-01914-5

Mekoya, A., & Workneh, M. (2024). Comparison of potential evapotranspiration methods in Ethiopia. *Discover Atmosphere*, *2*(1). https://doi.org/10.1007/s44292-024-00003-9

Mosbahi, M., Kassouk, Z., Benabdallah, S., Aouissi, J., Arbi, R., Mrad, M., Blake, R., Norouzi, H., & Béjaoui, B. (2023). Modeling Hydrological Responses to Land Use Change in

Sejnane Watershed, Northern Tunisia. *Water (Switzerland)*, *15*(9). https://doi.org/10.3390/w15091737

Mustafa, A., Szydłowski, M., Veysipanah, M., & Hameed, H. M. (2023). GIS-based hydrodynamic modeling for urban flood mitigation in fast-growing regions: a case study of Erbil, Kurdistan Region of Iraq. *Scientific Reports*, *13*(1), 1–18. https://doi.org/10.1038/s41598-023-36138-9

Nonki, R. M., Amoussou, E., Lenouo, A., Tshimanga, R. M., & Houndenou, C. (2023). Sensitivity and identifiability analysis of a conceptual-lumped model in the headwaters of the Benue River Basin, Cameroon: implications for uncertainty quantification and parameter optimization. *Hydrology Research*, *54*(9), 1036–1054. https://doi.org/10.2166/nh.2023.243

Onyutha, C. (2022). A hydrological model skill score and revised R-squared. *Hydrology Research*, *53*(1), 51–64. https://doi.org/10.2166/NH.2021.071

Onyutha, C. (2024). Pros and cons of various efficiency criteria for hydrological model performance evaluation. *Proceedings of the International Association of Hydrological Sciences*, *385*, 181–187. https://doi.org/10.5194/piahs-385-181-2024

Ouallali, A., Kader, S., Bammou, Y., Aqnouy, M., Courba, S., Beroho, M., Briak, H., Spalevic, V., Kuriqi, A., & Hysa, A. (2024). Assessment of the Erosion and Outflow Intensity in the Rif Region under Different Land Use and Land Cover Scenarios. *Land*, *13*(2), 1–20. https://doi.org/10.3390/land13020141

Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016). Environmental Modelling & Software Sensitivity analysis of environmental models : A systematic review with practical work fl ow. *Environmental Modelling and Software*, *79*, 214–232. https://doi.org/10.1016/j.envsoft.2016.02.008

Sahu, M. K., Shwetha, H. R., & Dwarakish, G. S. (2023). State-of-the-art hydrological models and application of the HEC-HMS model: a review. *Modeling Earth Systems and Environment*, *9*(3), 3029–3051. https://doi.org/10.1007/s40808-023-01704-7

Santos, D., Cardoso-Fernandes, J., Lima, A., Müller, A., Brönner, M., & Teodoro, A. C. (2022). Spectral Analysis to Improve Inputs to Random Forest and Other Boosted Ensemble Tree-Based Algorithms for Detecting NYF Pegmatites in Tysfjord, Norway. *Remote Sensing*, *14*(15). https://doi.org/10.3390/rs14153532

Seibert, J. (2000). Multi-criteria calibration of a conceptual runoff model using a genetic algorithm. *Hydrology and Earth System Sciences*, *4*(2), 215–224.

Vogeti, R. K., Srinivasa Raju, K., Nagesh Kumar, D., Rajesh, A. M., Somanath Kumar, S. V., & Jha, Y. S. K. (2023). Application of hydrological models in climate change framework for a river basin in India. *Journal of Water and Climate Change*, *14*(9), 3150–3165. https://doi.org/10.2166/wcc.2023.188

Worku, M. A., Feyisa, G. L., & Beketie, K. T. (2022). Climate trend analysis for a semi-arid Borana zone in southern Ethiopia during 1981–2018. *Environmental Systems Research, 11*(1). https://doi.org/10.1186/s40068-022-00247-7