Hydrology Research

Hydrology Research Vol 55 No 3, 412 doi: 10.2166/nh.2024.188

Multi-gauge calibration comparison for simulating streamflow across the Major River Basins in Madagascar: SWAT + Toolbox, R-SWAT, and SWAT + Editor Hard calibration

Rakotoarimanana Zy Harifidy[a,*], Ishidaira Hiroshi[b], Souma Kazuyoshi[b], Magome Jun[b], Rakotoarimanana Zy Misa Harivelo[c] and Carlos Antonio Fernández-Palomino[d,e]

[a] Integrated Graduate School of Medicine, Engineering and Agricultural Sciences, University of Yamanashi, Kofu 400-8511, Japan
[b] Interdisciplinary Centre for River Basin Environment, University of Yamanashi, Kofu 400-8511, Japan
[c] College of Civil Engineering, Tongji University, No. 1239, Si Ping Rd., Shanghai 200092, China
[d] Research Department II – Climate Resilience, Potsdam Institute for Climate Impact Research, Potsdam, Germany
[e] Institute of Environmental Science and Geography, University of Potsdam, Potsdam, Germany
*Corresponding author. E-mail: rakzyharifidy@yahoo.fr; g22dtk04@yamanashi.ac.jp

RZH, 0000-0002-5748-2990

ABSTRACT

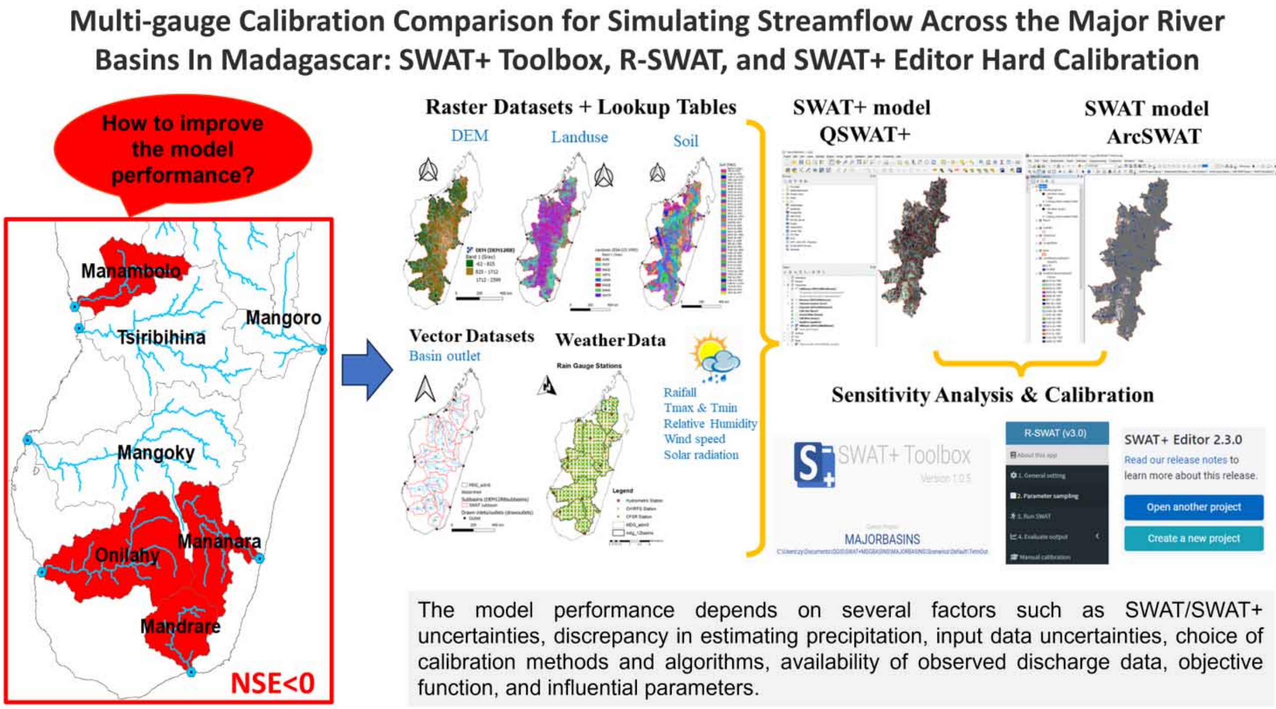
This paper aims to improve the Soil and Water Assessment Tool (SWAT) model performance across the Major River Basins in Madagascar (MRBM), specifically for SWAT simulation in the Manambolo, Onilahy, Mananara, and Mandrare basins. A multi-gauge calibration was carried out to compare the performance of SWAT+ Toolbox, and R-SWAT, SWAT+ Editor Hard calibration on a monthly time step for the periods 1982–1999. We found that the SWAT+ model generated greater surface runoff, while the SWAT model resulted in higher groundwater flow in both CSFR and CHIRPS datasets. It has been demonstrated that the SWAT+ Toolbox had more potential in calibrating runoff across the MRBM compared to R-SWAT. Calibration in both methods led to a reduction in surface runoff, percolation, water yield, and curve number but increased the lateral flow, evapotranspiration (ET), and groundwater flow. The results showed that the multi-gauge calibrations did not significantly enhance simulation performance in the MRBM compared to single-site calibration. The performance of the SWAT+ model for runoff simulation within the SWAT+ Toolbox and R-SWAT was unsatisfactory for most basins (NSE < 0) except for Betsiboka, Mahavavy, Tsiribihina, Mangoro, and Mangoky basins (NSE $= 0.40$–0.70; $R^2 = 0.45$–0.80, PBIAS$\leq \pm 25$), whether considering the CHIRPS or CSFR datasets. Further study is still required to address this issue.

Key words: multi-gauge calibration, Major River Basins in Madagascar (MRBM), R-SWAT, SWAT + Editor, SWAT + Toolbox

HIGHLIGHTS

- The SWAT+ model generated greater surface runoff, while the SWAT 2012 model resulted in higher groundwater flow.
- CSFR data exhibited high precipitation compared to the CHIRPS data.
- The SWAT + Toolbox showed greater potential in calibrating runoff across the MRBM compared to R-SWAT.
- Multi-gauge calibrations did not significantly enhance simulation performance in the MRBM compared to single-site calibration.

GRAPHICAL ABSTRACT



INTRODUCTION

The Soil and Water Assessment Tool (SWAT) model is a semi-distributed river basin model that operates on a daily time step (Arnold *et al.* 2012). SWAT has undergone continuous model refinement and enhancement. SWAT 2012 determines the hydrological response units (HRUs) in every sub-basin based on soil, land use, and slope. SWAT+ offers more flexibility in defining management schedules, routing constituents, and connecting managed flow systems to the natural stream network compared to SWAT 2012 (Bieger *et al.* 2017). One of the modifications in SWAT+ involves incorporating landscape units and managing the flow and movement of pollutants across the landscape. Several studies have tested the application of the SWAT 2012 and SWAT+ models globally such as in the US (Wu *et al.* 2020), in Africa (Chawanda *et al.* 2024), and in Europe (Wagner *et al.* 2022). These studies demonstrate that the models performed efficiently in simulating streamflow, sediment, snow, soil nutrient loss, impact of land use, and climate change.

In recent years, significant advancements have been observed in the application of process-based hydrological models, marking notable progress in addressing the complex challenges of water resources systems (Mohammadi *et al.* 2024). One of the complexities in utilizing the SWAT model involves conducting parameter calibration, sensitivity assessment, and uncertainty analysis. Calibration and validation are essential processes to reduce uncertainties and ensure an effective application of hydrological models. Calibration involves adjusting the parameters that affect SWAT model predictions using observed data (Wagener *et al.* 2004). Validation, on the other hand, involves comparing SWAT results with observed data without making any changes to the influencing factors. The parameter adjustment process can be manual or automatic. Manual calibration requires the expertise of the modeler in understanding the hydrological behavior of the watersheds of the study area and is time-consuming (Kannan *et al.* 2008). Automatic calibration is based on objective functions with optimization algorithms to search for optimal values of parameters (Sorooshian & Gupta 1995). An objective function is to measure the numerical difference between model output and observations (Pechlivanidis *et al.* 2011). To date, various methods have been developed and introduced to calibrate and validate the SWAT model such as SWAT-CUP (Soil and Water Assessment Tool Calibration and Uncertainty Program) (Abbaspour *et al.* 2007), IPEAT+ (Integrated Parameter Estimation and Uncertainty Analysis Tool Plus) (Yen *et al.* 2019), R-SWAT (R environment for SWAT) (Nguyen *et al.* 2022), SWAT + Editor (Desktop interface to SWAT+) (Bieger *et al.* 2017), SWAT + Toolbox (C#-coded software compatible with the SWAT+ model) (Chawanda 2021).

Even though single-gauge calibration has been the standard practice for many years, multi-gauge calibration has gained prominence in recent hydrological research due to its potential to improve model accuracy and robustness. Researchers continue to advance methodologies for effectively implementing multi-gauge calibrations despite the challenges associated with data quality, computational complexity, and parameter transferability. Several studies have explored various methodologies for multi-gauge calibrations such as Bayesian approaches, machine learning techniques, and advanced optimization algorithms to handle the increased complexity of calibrating models. Those studies conducted a comparison of the single- and multi-gauge-based calibrations for hydrological modeling. As reported by Gong *et al.* (2012), and Singh & Saravanan (2022), values obtained through a single-gauge calibration hold greater significance compared to those from a multi-gauge calibration. However, previous studies have demonstrated the importance of the multi-gauge approach to calibrate the observed data for large catchments (Wang *et al.* 2012; Song *et al.* 2021). Many researchers have achieved excellent results while using the multi-calibration methods for discharge, sediment, nutrients, and evaporation (Shrestha *et al.* 2016; Odusanya *et al.* 2019).

Rainfall input data plays a crucial role in simulating streamflow in hydrological modeling. Previous studies evaluated the accuracy and reliability of rainfall data from the Climate Forecast System Re-analysis/CFSR (Saha *et al.* 2014) and Climate Hazards Group InfraRed Precipitation with Station data/CHIRPS (Funk *et al.* 2015) for simulating the streamflow in Africa using the SWAT model. They found that CHIRPS had good performance compared to the CSFR data (Le & Pricope 2017; Duan *et al.* 2019; Akoko *et al.* 2021). Other studies also found that the combination of CFSR with CHIRPS precipitation data provides highly accurate results (Dile *et al.* 2016; Bayissa *et al.* 2017; Duan *et al.* 2019). Regarding the case of Madagascar, previous studies assessed the reliability of several gridded precipitation datasets at the daily, monthly, seasonal, and annual time steps. Their analysis demonstrated CHIRPS data to better represent the country's rainfall with low biases compared to other gridded rainfall data (Kling–Gupta Efficiency (KGE) = 0.88, correlation coefficient (CC) = 0.91, root mean square error (RMSE) = 51.1, and bias = –0.6) (Randriatsara *et al.* 2022; Ramahaimandimby *et al.* 2022; Ollivier *et al.* 2023). An accurate simulation of hydrological processes is therefore important for water resource management across the MRBM.

Our prior research has involved the application of SWAT 2012, CSFR data, and single-gauge calibration methods using the SUFI-2 (Sequential Uncertainty Fitting) algorithm in SWAT-CUP to the MRBM. However, the SWAT model performed poorly in the Southern basins. Another challenge in conducting long-term shifts in river flow on a national level is the lack of observed data as well as the differences in climatic and hydrological characteristics among regions. Considering these research gaps, the current study explores several approaches to improve the SWAT model's performance across the MRBM. First, we will assess the effectiveness of updating the SWAT model with the newly restructured version, SWAT+, and compare its performance in simulating streamflow. Additionally, we will examine the impacts of various precipitation data sources, such as CSFR (Re-analysis) and CHIRPS (Satellite) on the streamflow. Furthermore, our investigation extends to evaluating the influence of different calibration methods and algorithms including the application of SWAT+Toolbox (Dynamically Dimension Search/DDS) and R-SWAT (Uniform Latin Hypercube Sampling/LHS). These approaches will provide a deeper understanding of the factors affecting the SWAT model's performance and valuable baseline information for the MRBM.

The main aim of this study is therefore to improve the SWAT model performance across the MRBM, specifically for SWAT simulation in the Manambolo, Onilahy, Mananara, and Mandrare basins. A multi-gauge calibration for catchment discharge is carried out to test the approaches mentioned above. The results of this research could provide reasonable recommendations for streamflow calibration in data-scarce large river basins. The findings will be beneficial for future water resources planning, reducing ecological risks, and addressing potential social crises arising from water scarcity in Madagascar.

METHODS

Study area

The area for which modeling was performed in the present study is the Major River Basins in Madagascar (MRBM). The MRBM covers an area of over 10,000 km^2, each with a total area of about 320,373.20 km². The following basins are presented in Figure 1*:* Sofia, Mahajamba, Mahavavy, Betsiboka, Maningory, Manambolo, Tsiribihina, Mangoro, Mangoky, Onilahy,

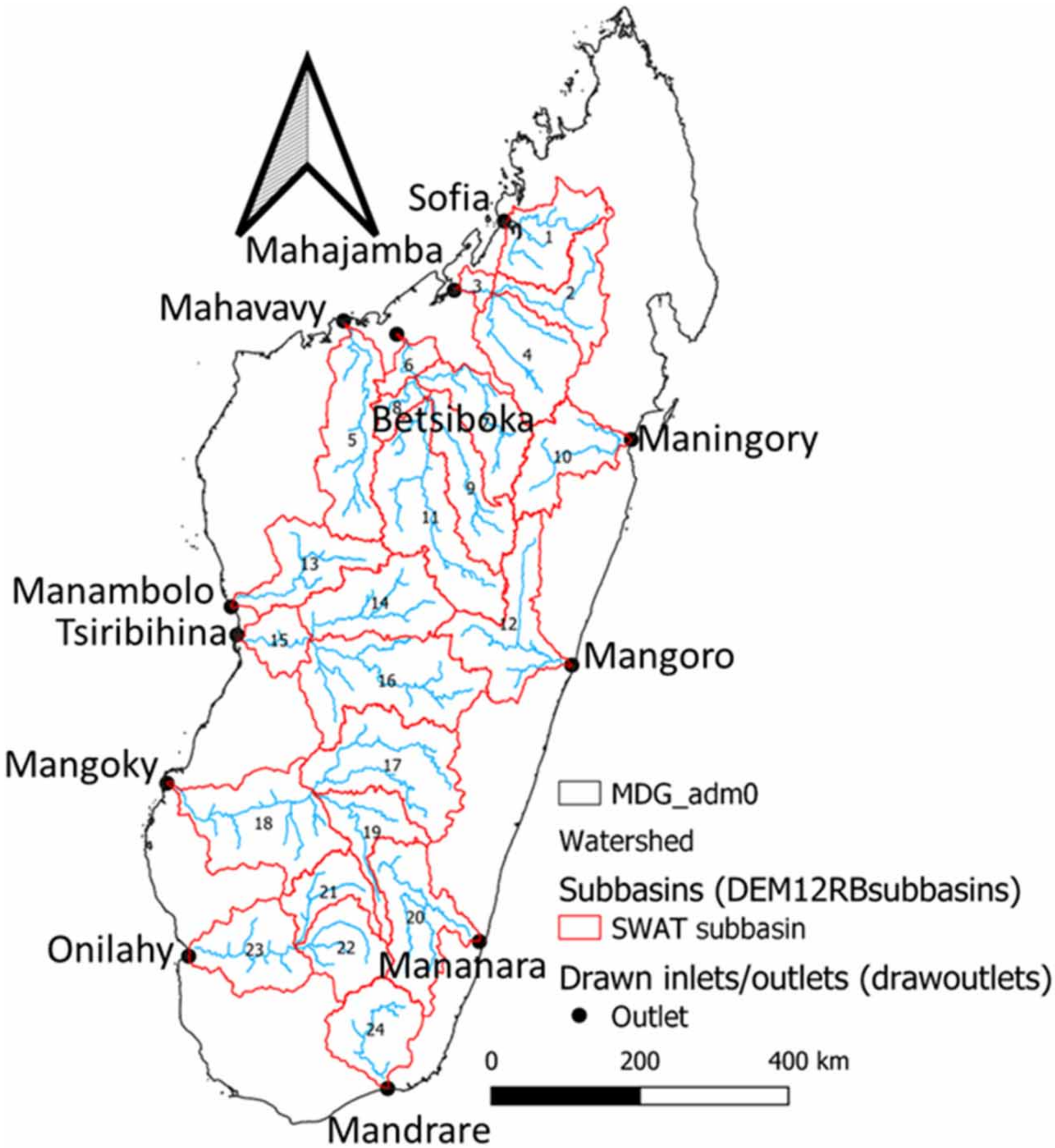


Figure 1 | Map showing the MRBM.

Mananara, and Mandrare. These basins face many challenges such as water scarcity and conflict of use (Rakotoarimanana & Ishidaira 2022), the impact of climate change (Raholijao *et al.* 2019), population growth (INSTAT 2019), and significant changes in land use patterns (Tiandraza *et al.* 2023). In addition, there has been limited research conducted on river streamflow within these sub-basins due to the lack of comprehensive data which is a serious challenge for Madagascar (Rabezanahary *et al.* 2021). Consequently, predicting the behavior of hydrological systems at the MRBM is a challenging task, and requires a proper model to gain meaningful insights.

SWAT model setup

In this study, SWAT 2012 and SWAT+ were set up for the MRBM, and their performances in simulating streamflow were assessed through a comparative analysis. ArcGIS 10.6.1 and QGIS 3.28 software were used including their extensions

ArcSWAT and QSWAT for this analysis. We used the Penman–Monteith equation, SCS Curve Number, and Muskingum Cunge methods to estimate potential evapotranspiration (PET), water partitioning in surface runoff and infiltration, and routing.

For further details, the SWAT 2012 and SWAT + user manuals are available at https://swat.tamu.edu/media/69296/SWAT-IO-Documentation-2012.pdf and https://swatplus.gitbook.io/io-docs/. Figure 2 describes the method used in this study to effectively address the research objectives and ensure the reliability of our findings.

Four model scenarios were set up using SWAT 2012 and SWAT+ models and two types of rainfall data (CHIRPS and CSFR). Precipitation is a critical input data when simulating streamflow. Both models require daily data for precipitation, maximum and minimum air temperature, solar radiation, relative humidity, and wind speed. The models can read observed weather data or generate values using the weather generator. A weather generator module (WGEN) is capable of filling in the missing weather data and simulating weather parameters when observed data are unavailable (Richardson 1981). Climate data will be generated in two instances: when the user specifies that simulated weather data will be used or when there are missing values in the observed weather data.

SWAT employs a nearest-neighbor approach, assigning each sub-basin the precipitation value from the closest weather station (Neitsch *et al.* 2011). This approach could potentially influence the calculation of average precipitation on a watershed scale. While the input file must contain data for the entire period of simulation, the record does not have to begin with the first day of simulation. SWAT can search for the beginning date in the file, saving editing time on the user's part. Once SWAT identifies the record for the initial simulation day, it ceases processing the year and date, skipping subsequent dates. Consequently, the data for the remaining simulation days must be sequentially listed (*source:* https://swat.tamu.edu/media/69317/ch06_input_pcp.pdf).

In SWAT + , each spatial object will be associated with weather stations closest to its centroid for precipitation, temperature, solar radiation, relative humidity, and wind speed. As these stations may be located at different positions, multiple combinations of weather stations may be necessary for a SWAT+ setup. These combinations are documented as a record in weather-sta.cli, each assigned a unique name referenced by connect files for various spatial objects. Furthermore, the closest weather generator station is specified by name, pointing to weather-wgn.cli. Users also have the option to designate the

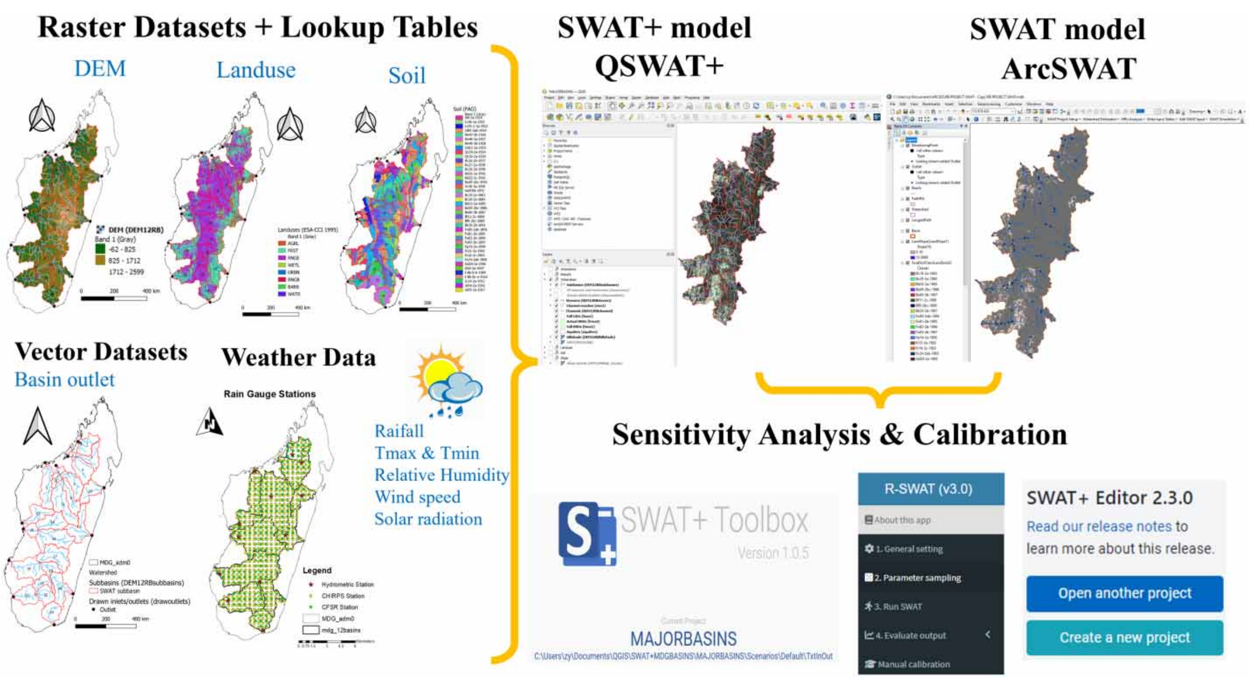


Figure 2 | Schematics showing the input data, the SWAT + / SWAT model setup, and calibration methods (SWAT + Toolbox, R-SWAT, SWAT + Editor).

name of an atmospheric deposition record, pointing to atmo.cli (source: https://swatplus.gitbook.io/io-docs/introduction/climate).

The simulations in SWAT 2012 and SWAT+ models started with watershed delineation. We used identical input data for the DEM, land use, and soil type for both models. The next step is to define the HRUs which are subdivided into LSUs in the SWAT+ model. We set an HRU threshold of 5% for land use, 5% for soil type, and 5% for slope. After generating the HRUs, we input two types of weather data. One scenario was set up with CSFR data (precipitation, min. and max air temperature, relative humidity, wind speed, and solar radiation). The other scenario used the combination of CHIRPS rainfall data and CSFR data (min. and max air temperature, relative humidity, wind speed, solar radiation). Finally, we ran the models with a 3-year warm-up period. The simulation length was 36 years (1979–2014) for CSFR data and 34 years (1981–2014) for CHIRPS data. Data for this study were collected using global satellite observations data and can be applied to the MRBM.

Calibration and validation

In this study, we carried out a multi-gauge calibration for catchment discharge using only the results from the SWAT+ model and compared the performance of SWAT + Toolbox, R-SWAT, and SWAT + Editor Hard Calibration on a monthly time step for the period 1982–1999. A comprehensive overview of the data used, experimental design principles, and calibration methods used in this study are provided in the Supplementary material, Appendix A. The Nash–Sutcliffe coefficients (NSE), the coefficient of determination (R^2), and the percentage bias (PBIAS) were used to compare the performance of the SWAT+ model for runoff simulation. According to Moriasi *et al.* (2007), NSE > 0.5, $R^2 > 0.5$ and PBIAS $\leq \pm 25$ were used to assess the satisfactory performance of the model for simulating discharge.

- Calibration method in the SWAT + Toolbox

The SWAT + Toolbox enables users to undertake sensitivity assessments, both manual and automatic calibration, among other functions. The calibration was conducted utilizing a semi-automated calibration routine provided within the SWAT + Toolbox (Chawanda 2021). Parameter optimization was achieved by employing the DDS algorithm integrated into the SWAT + Toolbox. Previous research has demonstrated that DDS is well-suited to dealing with complex calibration of distributed watershed models. DDS quickly reaches effective calibration solutions and can effortlessly steer clear of unfavorable local optima (Tolson & Shoemaker 2007). A step-by-step guide for the user manual is available at https://celray.github.io/docs/swatplus-toolbox/v1.0/index.html.

- Calibration method in the R-SWAT

The R-based SWAT (R-SWAT) was initially developed by Wu & Liu (2012) by transforming the Fortran-based SWAT model into an R function. R-SWAT is an interactive graphical user interface tool in the R environment for SWAT parameter calibration, sensitivity and uncertainty analyses, and visualization (Nguyen *et al.* 2022). Parameter optimization was achieved by using the Uniform LHS approach integrated into the R-SWAT. This approach has been used in uncertainty analysis by SWAT-CUP 'SUIF2' algorithm (Abbaspour *et al.* 2007). A previous study argues that LHS can produce a more efficient representation of samples compared to the Monte Carlo method (Yu *et al.* 2001). A detailed user manual that provides a systematic and sequential guide is available at https://github.com/tamnva/R-SWAT/wiki/R-SWAT-User-Manual.

- Calibration method in the SWAT + Editor

The SWAT + Editor is a desktop interface to SWAT+ which provides the capability to modify the input files, run simulations, and conduct sensitivity analysis and calibration. It operates as an independent application without GIS, which enables ArcSWAT and QSWAT users to easily collaborate with individuals lacking GIS expertise or access. Two types of calibrations are available in SWAT + Editor. Soft calibration focuses on mass balance whereas hard calibration focuses on getting simulated and observed hydrographs as closely as possible. The results of the fitted values from the calibration in SWAT + Toolbox and R-SWAT were put back in SWAT + Editor to conduct the hard calibration.

RESULTS AND DISCUSSION

Comparison results of the models before calibration

SWAT 2012 divided the MRBM into 144 sub-basins while the SWAT+ divided the MRBM into 24 sub-basins and 222 landscape units (including both floodplain and upslope). After applying a 5% filter of thresholds, the SWAT+ model setups

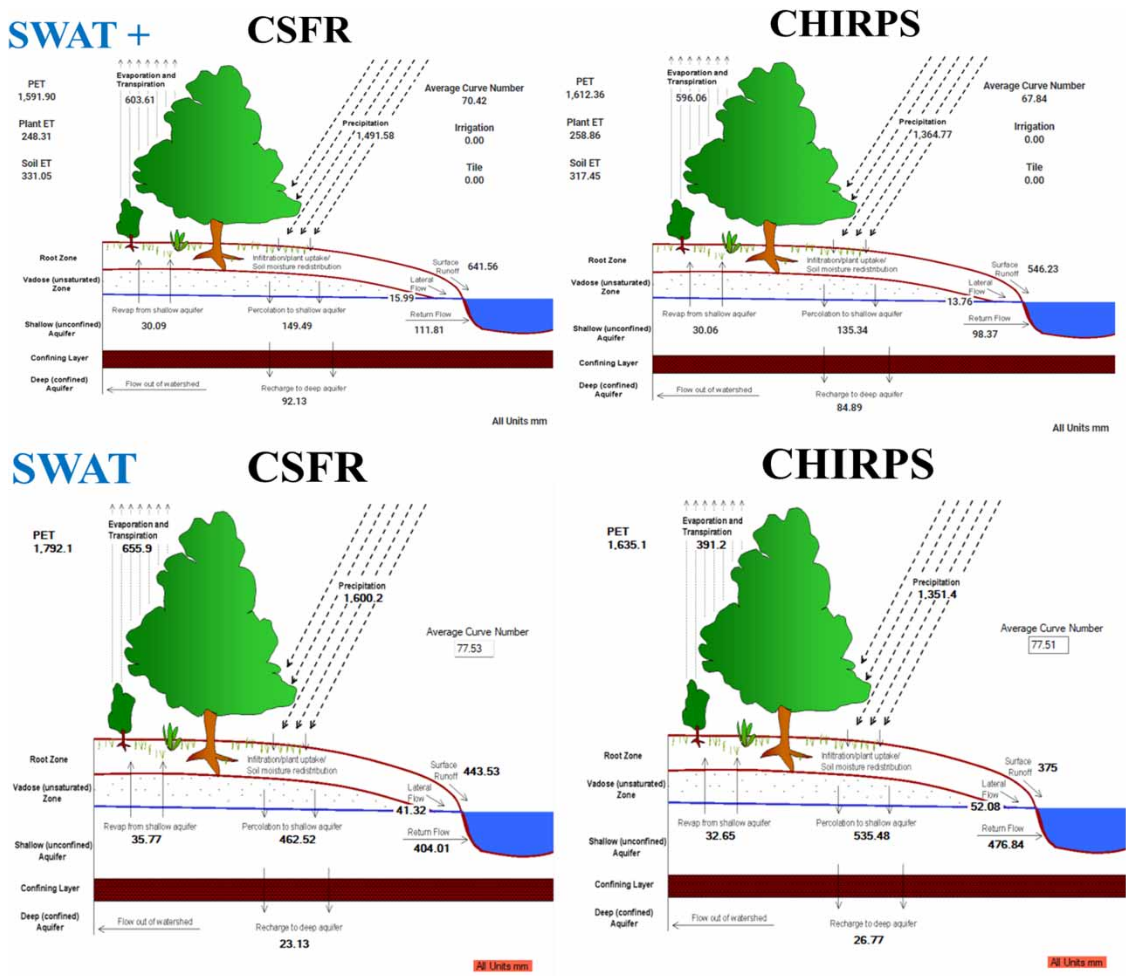


Figure 3 | SWAT + /SWAT model water balance ratios before calibration.

produced approximately 2,429 HRUs, which is larger compared to the SWAT 2012 about 1,774 HRUs. Figure 3 illustrates the difference in the water balance ratios between CSFR and CHIRPS rainfall data modeling in SWAT and SWAT+ models before calibration.

Figure 3 shows a large difference in the SWAT 2012/SWAT+ model water balance ratios between the two rainfall datasets. We found that SWAT+ produced more surface runoff compared to the SWAT 2012 model for both CSFR (641.56/443.53 mm) and CHIRPS (546.23/ 375.00 mm) data. However, the groundwater flow generated by the SWAT model was higher compared to the SWAT+ model for both CSFR (404.01/111.81 mm) and CHIRPS (476.84/ 98.37 mm) data. Wagner *et al.* (2022) also found similar results when they compared the performance of two SWAT+ models with the SWAT model version in the Kielstau catchment in Northern Germany. These findings may be explained in terms of the implementation of landscape units in the SWAT+ model (Bieger *et al.* 2017) which promotes the interaction of hydrological processes amongst HRUs in the MRBM. We also found that the SWAT+ model was easy to manipulate and had several features compared to the SWAT 2012. The most striking result to emerge from Figure 3 is the difference in precipitation values between SWAT+ and SWAT 2012 for the CSFR dataset (1,491.58 and 1,600.2 mm). However, the simulated precipitation using the CHIRPS dataset shows minimal variation in the SWAT+ (1,364.77 mm) and SWAT models (1,351.4 mm). This finding can be attributed to the models' internal processes, how they handle precipitation inputs, and the discretization of sub-basins, as SWAT employs a significantly larger number of sub-basins compared to SWAT+. In SWAT, a watershed is divided into several sub-watersheds, which are subsequently divided into hydrologic response units

(HRUs) (Arnold *et al.* 2012). SWAT+ adopts the landscape unit (LSU) approach to enhance the distribution and allocation of management operations across landscapes (Bieger *et al.* 2017). This modification provides greater flexibility in watershed discretization and configuration to better capture the intricacies of a watershed (Rathjens *et al.* 2015). The CHIRPS precipitation dataset has a high resolution (0.25°) and is consistent in capturing the daily rainfall pattern compared to the CSFR dataset (Musie *et al.* 2019; Upadhyay *et al.* 2022). Higher resolution data provides more detailed information but demands more computational resources (Singh & Saravanan 2022). It should be emphasized that we did not interpolate weather station data as SWAT will automatically distribute the weather data to the sub-basins by using data from only the gauge station that is nearest to the centroid of each sub-basin (Tuo *et al.* 2016). In addition, both models use different interpolation methods for distributing precipitation data within the watershed. SWAT+ employs a more detailed method for representing precipitation distribution, while SWAT 2012 uses a simpler approach (Arnold *et al.* 2012; Bieger *et al.* 2017). Both models require daily precipitation as input and do not directly change the amount of precipitation. The models simulate how the precipitation flows through the watershed by considering various factors such as evaporation, infiltration, interception, storage, and release. These processes essentially determine how much of the input precipitation eventually becomes runoff. While the total precipitation amount remains the same, SWAT and SWAT+ can differ in their representation and calculation of these intermediate steps, leading to apparent discrepancies in the simulated precipitation values. Four model scenarios show that the PET contributes more to the water loss in the MRBM compared to other water balance components. Moreover, the results showed that the SWAT model generates a high curve number compared to the SWAT+ model. These results indicate that a significant portion of rainfall is likely to become surface runoff rather than infiltrating into the soil. Table 1 presents the performance of both models in simulating the streamflow. The Nash–Sutcliffe Efficiency (NSE), the Coefficient of Determination (R^2), and the Percent Bias (PBIAS) were used to evaluate the performance of the uncalibrated SWAT+ and SWAT models. Table 1 shows a great difference between the uncalibrated SWAT+ and SWAT in both CHIRPS and CSFR datasets.

As shown in Table 1, uncalibrated SWAT 2012 and SWAT+ models simulated the streamflow across the MRBM with unsatisfactory performance metrics (NSE < 0.5, $R^2 < 0.5$ and PBIAS $\geq \pm 25$). Both models and rainfall input datasets indicated a negative value of the NSE (NSE < 0) for most basins. The low NSE values in both models were due to systematic underestimation of the streamflow (negative value of PBIAS). Comparing the four results, it can be seen that the SWAT+ model setups perform better than the SWAT 2012 model based on NSE (–1.92 to 0.65 for SWAT + ; −40.8 to 0.63 for SWAT). However, the findings suggest that the R^2 is greater for the SWAT 2012 compared to the SWAT+ model. The CSRF-driven SWAT

Table 1 | Model performance before calibration

	SWAT + model						SWAT model					
	CHIRPS data			CSRF data			CHIRPS data			CSRF data		
Basin name	R^2	NSE	PBIAS	R^2	NSE	PBIAS	R^2	NSE	PBIAS	R^2	NSE	PBIAS
Sofia	0	−0.4	−135.3	0	−0.47	−115.7	0.03	−2.3	−128.6	0.41	−2.19	−195
Mahajamba	0	−0.63	−126.6	0	−0.99	−54.8	0.21	−3.55	−140.8	0.75	−1.93	−67.2
Mahavavy	0.13	−0.74	−32.7	0.07	−1.23	−58.4	0.16	−2.31	27.2	0.65	0.19	−28
Betsiboka	0.13	−0.68	−70.2	0.09	−1.44	−93.6	0.08	0.22	58.8	0.53	−0.09	82.3
Maningory	0.01	−1.07	−61.6	0.02	−0.7	−58.6	0.02	−3.91	−46	0.88	0.27	−50.7
Manambolo	0.02	−0.94	−48.5	0.02	−1.92	−71.9	0.14	−2.51	−18.4	0.35	−0.89	−35.9
Tsiribihina	0.79	0.65	4.7	0.26	−0.43	88.1	0.45	0.11	65.4	0.73	0.63	−11.3
Mangoro	0.01	0.18	−217.3	0.01	−0.43	−94.4	0.05	−4.92	−202.3	0.68	−1.3	−62.3
Mangoky	0.21	0.01	53.5	0.06	−0.71	63.5	0.45	0.16	68.3	0.6	0.25	−49.6
Onilahy	0.01	−1.44	−126.5	0	−1.37	−125	0.1	−40.8	−73.4	0.33	−37.18	−555.5
Mananara	0.03	−0.71	−102.3	0.01	−0.7	−101.2	0.09	−21.1	−329.6	0.29	−14.67	−269.1
Mandrare	0.05	−0.6	−108.4	0.04	−0.32	−106.5	0.2	−10.07	−211.5	0.14	−21.17	−448.2

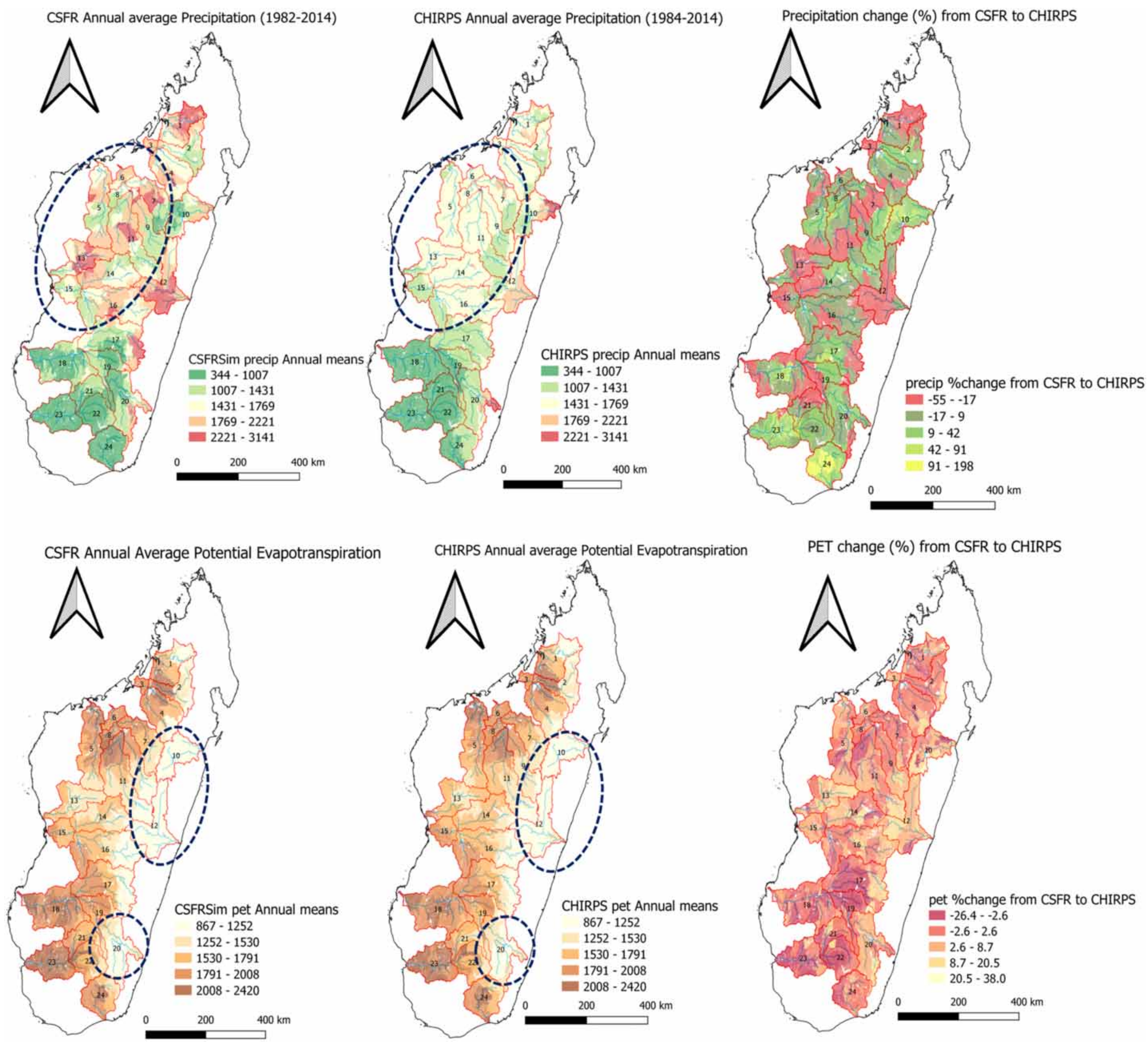


Figure 4 | Spatial visualization of annual average precipitation and PET using CSFR and CHIRPS data in SWAT+ before calibration.

model seems to be a favorable option (R^2 range of 0.14–0.88). Overall, these results demonstrate the need for calibration to further improve the model performance.

Experimental evidence on the impact of precipitation data selection on water balance components using the new version SWAT+ model was studied. The maps in Figure 4 illustrate the spatial distribution of the CSFR and CHIRPS rainfall and evapotranspiration (ET) data across the MRBM.

These maps show great differences in the spatial distribution of the precipitation and ET across the MRBM in SWAT+ before the calibration. The mean annual precipitation for the CFSR and CHIRPS ranged from 344 to 1,431 mm/year in the Southern basins (Mangoky, Onilahy, and Mandrare). CSFR data showed a high value of precipitation (ranging from 1,007 to 3,141 mm/year) compared to the CHIRPS data (ranging from 1,007 to 1,769 mm/year) in the Western basins (Mahavavy, Betsiboka, Manambolo, and Tsiribihina). The difference between the two products (change from CSFR to CHIRPS) ranged from –55 to −9%. However, both datasets indicate that the Northern basins (Sofia and Mahajamba basins) and Eastern basins (Maningory, Mangoro, and Mananara basins) received abundant precipitation compared to the Western and Southern basins. From Figure 4, both CHIRPS and CSFR data demonstrated high ET in the downstream areas of all basins (ranging from 1,530 to 2,420 mm/year) except for the Maningory, Mangoro, and Mananara basins (ranging from 867 to 1,252 mm/year). CSFR data generated larger ET compared to the CHIRPS data for the Betsiboka, Mahajamba Tsiribihina, and Mangoky basins (change from CSFR to CHIRPS ranged from –26.4 to −2.6%).

Result of sensitivity analysis and calibration

Multi-gauge calibration and validation were performed for streamflow and conducted at a monthly time step for the periods of 1982 to 1999. We used the same parameters and the same number of simulations (500 iterations) for both calibrations in the SWAT + Toolbox and R-SWAT. The sensitivity analysis revealed that the same parameters governing the streamflow were found to be sensitive for both calibrations in the SWAT + Toolbox and R-SWAT. Therefore, these six most sensitive parameters were used for calibration. It is important to note that these parameters had been found the most sensitive based on our previous single-site calibration using the SUFI-2 algorithm in SWAT-CUP. The following six most sensitive parameters were used: Soil Conservation Service runoff curve number for moisture condition II (CN2), Available Water Capacity of the soil layer (SOL_AWC), Manning's n value for overland flow (OV_N), Soil evaporation compensation factor (ESCO), Threshold depth from the surface to water table for revap to occur (REVAP_MIN), and Groundwater revap coefficient (REVAP_CO). Table 2 compares the initial ranges and final fitted values during the calibration in the SWAT + Toolbox and R-SWAT using CHIRPS and CSFR data.

As Table 2 shows, there is a significant difference between the best-fitted value of OV_N and REVAP-MIN parameters in both SWAT + Toolbox and R-SWAT. We found that CHIRPS data generated a very high fitted value of OV_N and REVAP-MIN in R-SWAT compared to the SWAT + Toolbox. On the contrary, CSFR data produced a high fitted value of OV_N and REVAP-MIN in the SWAT + Toolbox compared to R-SWAT. According to a previous study, a high OV_N value indicates a significant roughness on the bottom of the channel that reduces the speed of water flow (Li & DeLiberty 2021). A high value of REVAP-MIN indicates a relatively large depth from the surface to the water table in the soil before evaporation or recharge ET processes are considered in the model. The infiltration rate was determined by the CN2 and SOL_AWC. The initial range of CN2 in SWAT + Toolbox and R-SWAT was equal to minus or plus 20%. Results show a decrease of 20% in the CN2 value for both calibrations in SWAT + Toolbox and R-SWAT which indicates a decrease in runoff and an increase in baseflow across the MRBM. High values of the best-fitted AWC were found in the case of R-SWAT, while the optimal value of ESCO was significant in the case of SWAT + Toolbox. These results demonstrate the capability of the soil to retain a significant amount of water, which can be beneficial for plants and agriculture, especially in a basin with insufficient rainfall (Odusanya *et al.* 2019). Best-fitted REVAP_CO was close to 0 for both calibrations in SWAT + Toolbox and R-SWAT. This result indicates that the movement of water from the aquifer to the root zone is restricted.

SWAT+ performance during calibration and validation

The SWAT+ model was set up for the MRBM with two model scenarios comparing CSFR and CHIRPS rainfall data. The two model scenarios were calibrated and validated using multi-gauge calibration in SWAT + Toolbox and R-SWAT for the periods

Table 2 | Parameters initial ranges for calibration and their fitted values

				SWAT + Toolbox		R-SWAT	
Cal_parm	**chg_type**	**Min**	**Max**	**CHIRPS**	**CSFR**	**CHIRPS**	**CSFR**
CN2	Abschg	−0.2	0.2	−0.20	−0.19	−0.10	−0.09
AWC	Abschg	0.01	1	0.24	0.08	0.83	0.79
OV_N	Abschg	0.01	30	13.94	11.11	19.91	7.52
ESCO	Absval	0	1	1.00	1.00	0.54	0.07
REVAP_CO	Absval	0.02	0.2	0.11	0.14	0.13	0.04
REVAP_MIN	Absval	0	50	1.50	2.69	39.43	2.66

Note: Relative Change (abschg): absolute change increases/decreases the current value by a specified value. Replace (Absval): the specified value takes the place of the old parameter value.

1982–1999. It is important to note that calibration and validation periods vary for each basin. Several statistical metrics were used to evaluate the effectiveness and accuracy of the SWAT model. A value of 1 of NSE represents a perfect match between simulated and observed flow. A high value of the Coefficient of Determination (R^2) indicates the proportion of the variance in the observed and simulated flow data. R^2 ranges from 0 to 1, with 1 indicating a perfect fit. The Percent Bias (PBIAS) measures the average tendency of the model to overpredict or underpredict the observed flow. A value close to zero indicates a good performance of the SWAT model.

We applied the same calibration parameters, objective functions (NSE), and 500 simulations for both calibration methods. Table 3 presents the statistical results of the calibrated model parameters across the MRBM.

Table 3 demonstrates that the performance of the SWAT+ model for runoff simulation within the SWAT + Toolbox and R-SWAT was unsatisfactory for most basins, whether considering the CHIRPS or CSFR dataset. During calibration and validation, the NSE values were negative for the Sofia, Mahajamba, Maningory, Manambolo, Onilahy, Mananara, and Mandrare basins. In contrast, the NSE values fell within the range of 0.40 to 0.70 for the Betsiboka, Mahavavy, Tsiribihina, Mangoro, and Mangoky basins.

In general, the results revealed that the CHIRPS dataset performed well, with five out of the 12 basins achieving an NSE > 0.45. In contrast, the CSFR dataset fared less favorably, as only three out of the 12 basins exhibited an NSE > 0.45 during both the calibration and validation periods, whether using the SWAT + Toolbox or R-SWAT. When we utilized the SWAT + Toolbox for calibration, we observed an NSE range of −3.27 to 0.77, an R^2 range of 0.09–0.88, and a PBIAS range of −71.26 to 66.87. During the validation process, we obtained an NSE range of −32.57 to 0.69, with an R^2 range of 0.002–0.92 and a PBIAS range of −363.23 to 65.81 for all basins. On the other hand, when we employed R-SWAT for calibration, we found that the NSE values fell within the range of −2.27 to 0.70, while the R^2 and PBIAS values ranged from 0.03 to 0.81 and −83.86 to 74.02, respectively. During validation, the NSE values were in the range of −19.63 to 0.73, the R^2 values in the range of 0.05–0.95, and the PBIAS values in the range of −296.07 to 72.42 across all basins. The results of percent bias PBIAS provide insight into whether the low NSE values in this scenario are due to systematic overestimation or underestimation issues. An implication of these findings is the potential of the SWAT + Toolbox compared to R-SWAT in calibrating runoff across the MRBM. However, a couple of case studies demonstrated the efficiency of R-SWAT for modeling monthly streamflow and nitrate nitrogen (Wu & Liu 2014) and as a decision-support tool for integrated watershed management (Udías *et al.* 2016). It is important to note that the SWAT + Toolbox uses the channel morphology output files (channel_sdmorph_mon) for calibration while R-SWAT uses the channel output files (channel_sd_mon). A channel is defined as a flowing water body transporting water from one point to another whereas channel morphology provides a detailed breakdown of how sediments are being transported and deposited within the channels. SWAT+ does not just simulate how the water flows but also how it shapes the landscape and moves sediments around. In addition, both calibration tools use different interfaces (QGIS/R scripts), data analysis capabilities, and algorithms for parameter sensitivity analysis and calibration (DDS/ Uniform LHS). The difference in these results can be explained by the calibration tool's structure uncertainties, data uncertainties, and model parameter uncertainties (Pechlivanidis *et al.* 2011). The main concern of the paper was to improve the SWAT model performance in the Manambolo, Onilahy, Mananara, and Mandrare basins. However, the results proved to be inefficient for the two model runs to simulate the flow at Manambolo (NSE = –0.24, −1.23), Onilahy (NSE = –0.11, −0.03,), Mananara (NSE = 0.16, −32.57), and Mandrare (NSE = 0.02, –0.76) basins during the calibration and validation periods, respectively. These findings indicate that the model performance depends not only on the choice of calibration method but, more significantly, on factors such as limited observed discharge data, the uncertainty in rainfall data, the influential parameters, and the objective function used for the calibration approach. McIntyre *et al.* (2002) confirmed that data insufficiency in calibration results in uncertainty in parameter estimates. Therefore, a longer calibration period and effective adjustment of the parameter sets are required for better model calibration in these basins.

Figure 5 displays the hydrographs of monthly observed and best-simulated streamflow in SWAT + Toolbox and R-SWAT across the MRBM.

Overall, as shown in Figure 5, the two SWAT+ model scenarios using CSFR and CHIRPS data did not accurately replicate the overall pattern and timing of both low and high flow events in the MRBM during both calibrations in SWAT + Toolbox and R-SWAT. However, the observed and simulated flow matched well in Betsiboka, Mahavavy, Tsiribihina, Mangoro, and Mangoky basins. The graphs demonstrate that both CSFR and CHIRPS precipitation data resulted in significant peak flow, with CHIRPS data exhibiting a capability to forecast peak flow that closely matched the observed flow for both calibrations in

Table 3 | Model performance statistics of the two multi-gauge calibration methods used in this study

	CHIRPS-SWAT + Toolbox						CSFR-SWAT + Toolbox						CHIRPS R-SWAT						CSFR R-SWAT					
	Calibration			Validation			Calibration			Validation			Calibration			Validation			Calibration			Validation		
Basin name	R^2	NSE	PBIAS	R^2	NSE	PBIAS	R^2	NSE	PBIAS	R^2	NSE	PBIAS	R^2	NSE	PBIAS	R^2	NSE	PBIAS	R^2	NSE	PBIAS	R^2	NSE	PBIAS
Sofia	0.60	0.46	−13.84	0.5	−1.23	−74.72	0.42	−0.52	−71.26	0.51	−3.39	−95.54	0.57	−0.32	−56.95	0.57	−4.65	−156.46	0.46	−1.08	−83.86	0.59	−4.85	−150.56
Mahajamba	0.74	−1.02	9.01	0.69	−3.54	−4	0.7	−0.82	0.79	0.68	−9.8	−34.8	0.68	−2.27	− 21.96	0.68	−2.02	−0.12	0.74	−0.57	−19.20	0.7	−4.28	−16.49
Mahavavy	0.66	0.37	23.88	0.72	0.61	18.69	0.67	0.35	17.09	0.56	0.41	7.3	0.72	0.31	19.08	0.74	0.67	13.30	0.69	0.5	14.99	0.59	0.55	14.15
Betsiboka	0.79	0.57	17.07	0.77	0.65	18.67	0.69	−0.09	−12.83	0.66	0.4	−14.76	0.79	0.55	19.41	0.78	0.73	23	0.67	0.58	22.41	0.64	0.59	25.59
Maningory	0.45	−3.27	17.84	0.32	−0.13	30.55	0.25	−1.91	52.87	0.26	−0.06	41.18	0.52	−1.69	17.43	0.32	0.01	31.87	0.27	−1.51	46.39	0.27	0.02	43.53
Manambolo	0.37	−0.24	27.51	0.23	−1.23	3.13	0.42	−1.7	−42.38	0.45	−7.48	−116.84	0.36	−0.55	4.82	0.31	−1.85	−36.73	0.43	−0.76	−28.74	0.46	−4.38	−93.43
Tsiribihina	0.75	0.68	14.09	0.88	0.63	15.81	0.8	0.77	12.48	0.6	0.35	−21.95	0.81	0.7	30.93	0.78	0.45	43.22	0.81	0.66	35.15	0.62	0.61	10.99
Mangoro	0.73	0.52	22.22	0.64	0.48	23.05	0.44	−0.91	−39.19	0.45	−3.13	−92.49	0.77	0.59	24.50	0.66	0.49	23.40	0.47	0.26	2.33	0.55	−0.27	−42.64
Mangoky	0.81	0.41	23.58	0.90	0.69	22.99	0.41	0.35	16.2	0.54	0.48	21.24	0.54	0.03	23.06	0.76	0.48	24.75	0.59	0.14	21.11	0.41	0.29	23.59
Onilahy	0.18	−0.43	27.71	0.74	−1.4	−74.84	0.13	−0.11	11.31	0.51	−0.03	−65.34	0.05	−0.92	59.41	0.55	−2.09	− 51.43	0.25	−0.13	52.12	0.36	−0.82	72.42
Mananara	0.88	0.16	−39.51	0.96	−32.57	−363.23	0.42	0.04	−14.03	0.39	−0.61	−95.49	0.29	0.01	44.92	0.95	−19.63	− 296.07	0.07	−0.82	28.07	0.29	−0.15	−3.89
Mandrare	0.41	0.02	−6.8	0.002	−0.73	51.75	0.09	−0.07	66.87	0.04	−0.86	65.81	0.33	−0.43	− 1.29	0.05	−1.15	61.09	0.03	−0.18	74.02	0.06	−0.79	69.63

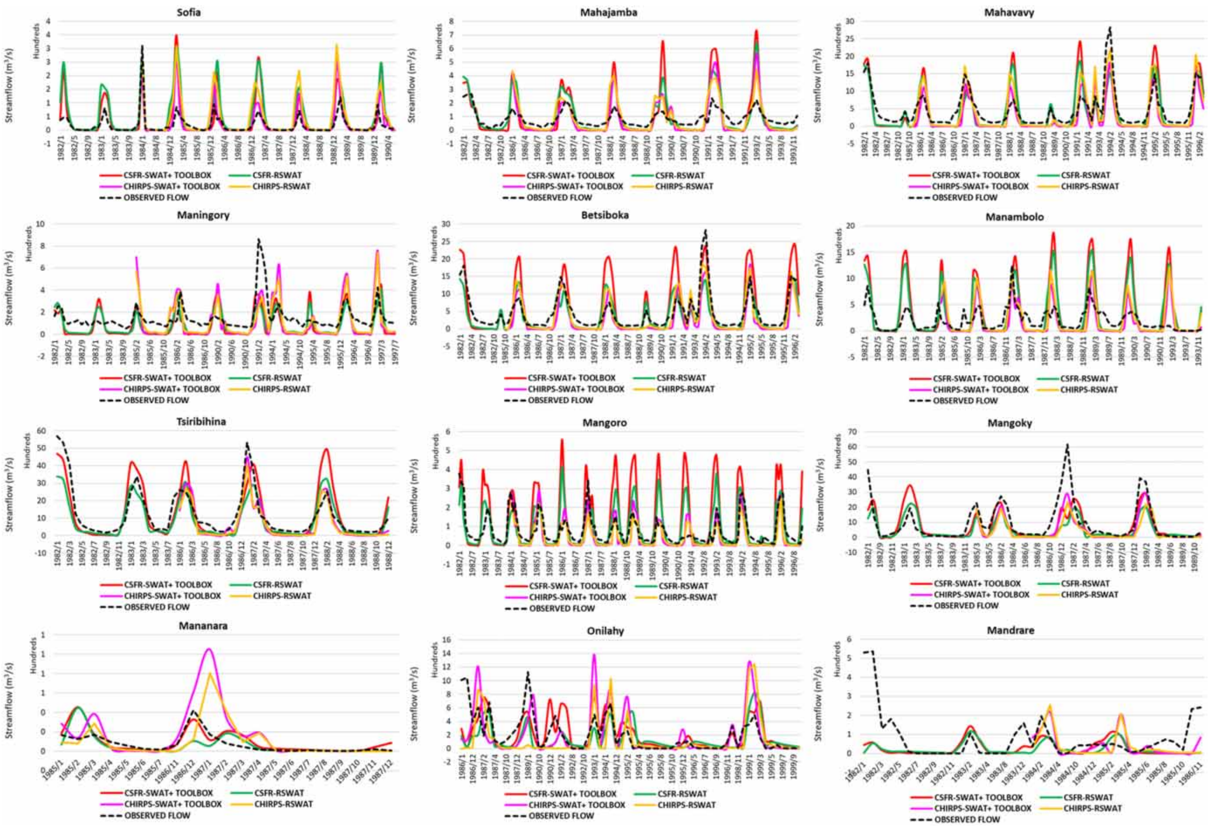


Figure 5 | Comparison of calibrated streamflow across the MRBM for the periods 1981–1999: Observed flow (Dotted black line), CHIRPS-SWAT + Toolbox (Pink), CHIRPS-R-SWAT (Orange), CSFR-SWAT + Toolbox (Red), and CSFR-R-SWAT (Green).

SWAT + Toolbox and R-SWAT. Interestingly, there is a large difference between the observed and predicted low flows in Mangoro, Mahajamba, and Manambolo basins for both scenarios and calibration methods. The model completely failed to predict the high flows in Mananara, Onilahy, and Mandrare basins in which CHIRPS data generated high peak flow compared to the CSFR data.

Taken together, these results imply that multi-gauge calibrations using SWAT + Toolbox and R-SWAT did not significantly improve simulation performance in the Manambolo, Onilahy, Mananara, and Mandrare basins compared to the results of our previous study using a single-site calibration in SWAT-CUP. We found that NSE ranged from 0.4–0.9 for the Mahajamba, Betsiboka, Maningory, Tsiribihina, and Mangoky basins; 0.23–0.68 for the Sofia, Mahavavy, and Mangoro basins; $NSE = 0.15$ for Manambolo basin and $NSE \leq 0$ for Onilahy, Mananara, and Mandrare basins during calibration and validation. However, the previous study compared R-SWAT and SWAT-CUP in calibrating streamflow and found satisfactory and minor differences in the results (Nguyen *et al.* 2022). It is interesting to note that calibration in R-SWAT required less time compared to the SWAT + Toolbox and SWAT-CUP. According to a previous study, R-SWAT rewrites a file with all updated parameter values at once (Nguyen *et al.* 2022). Moreover, R-SWAT provides a high level of customization, data analysis, and visualization tools of the model outputs with other R packages. On the other hand, the SWAT + Toolbox provides options for manual and automatic calibration as well as a good visualization of the results within QGIS which can make it more accessible to users (Chawanda *et al.* 2020). Overall, SWAT + Toolbox and R-SWAT provide flexible platforms for testing new parameter sensitivity and optimization packages with complex hydrological models. A community of users and support resources are also available at https://groups.google.com/g/R-SWAT and https://groups.google.com/g/swatplus.

Before proceeding to the water balance analysis, it is important to examine the impact of precipitation data selection (CSFR and CHIRPS) on the model after calibration.

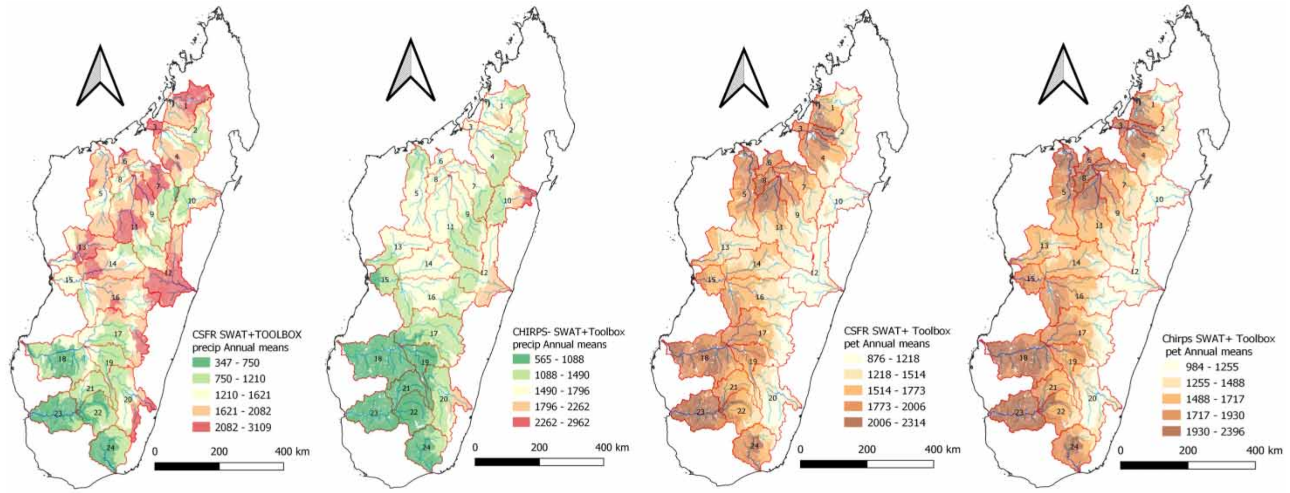


Figure 6 | Comparison of the annual average precipitation and evapotranspiration precipitation after calibration in SWAT + Toolbox.

The fitted values of the most sensitive parameters from the calibration SWAT + Toolbox and R-SWAT were put back in the SWAT + Editor to conduct the hard calibration. The spatial distribution of the water balance components after calibration is displayed in Figure 6.

These maps showed a large difference between precipitation and ET across the MRBM after the calibration in the SWAT + Toolbox. CSFR data produced high precipitation in the Western basins compared to the CHIRPS data. The ET was lower in the Eastern basins for both datasets. In general, these trends show that there has been minimal change in the spatial distribution of the precipitation and ET compared to the results before calibration.

Results of water balance

The differences in the annual average water balance components between CHIRPS and CSFR data calibrated in SWAT + Toolbox and R-SWAT are shown in Figure 7. The annual average precipitation, surface runoff, lateral flow, percolation, ET, water yield, PET, groundwater flow, and curve number provide a deeper understanding of the basin's hydrological processes and are crucial for effectively managing the water resources at the MRBM. The mean annual PET was calculated using the Penman–Monteith equation, the default method in SWAT + .

Figure 7 shows that among the water balance components, the PET was the most dominant (about 1,600 mm/year on average) while the lateral flow had the least influence (about 23 mm/year on average). The average annual amount of precipitation (1,360 mm/year for CHIRPS; 1,490 mm/year for CSFR) and ET (1,610 mm/year for CHIRPS; 1,590 mm/year for CSFR) were similar before and after model calibration for SWAT + Toolbox and R-SWAT. Calibration in both methods led to a reduction in surface runoff, percolation, water yield, and curve number but increased the lateral flow, ET, and groundwater flow. Surprisingly, SWAT + Toolbox produced a great value of ET in both CHIRPS (636 mm/year) and CSFR (664 mm/year) datasets compared to the calibration by R-SWAT (577 mm/year for CHIRPS; 525 mm/year for CSFR). In general, the CSFR dataset produced substantial quantities of precipitation, surface runoff, percolation, water yield, and curve number, whereas the CHIRPS dataset resulted in elevated ET and PET.

CONCLUSION

Two versions of the SWAT model, specifically SWAT 2012 and SWAT + , were set up for the Major River Basins in Madagascar (MRBM) and their performances in simulating streamflow were compared. The main aim of this study is to improve the SWAT model performance across the MRBM, specifically for SWAT simulation in the Manambolo, Onilahy, Mananara, and Mandrare basins. In this context, we also tried to investigate how the uncertainty of the CHIRPS and CSFR rainfall input data affects the streamflow across the MRBM. A multi-gauge calibration for catchment discharge was carried out to compare the performance of SWAT + Toolbox, and R-SWAT, SWAT + Editor hard calibration on a

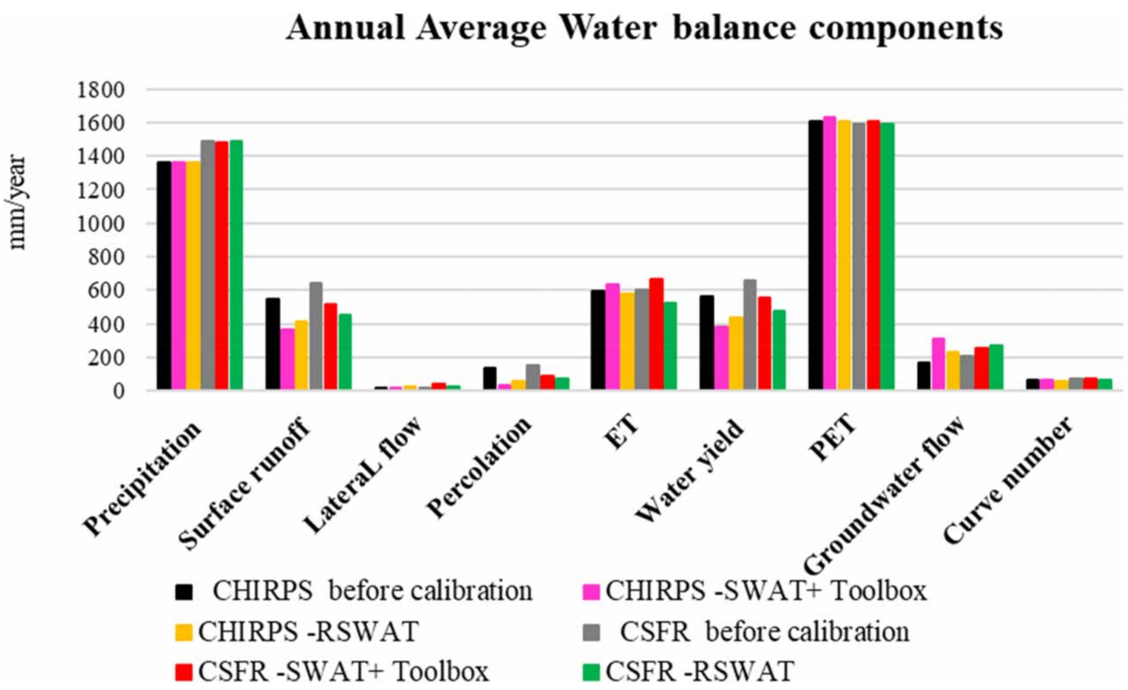


Figure 7 | Annual average water balance components simulated by SWAT+ model: CHIRPS before calibration (black), CHIRPS-SWAT + Toolbox (Pink), CHIRPS-R-SWAT (Orange), CSFR before calibration (Grey), CSFR-SWAT + Toolbox (Red), CSFR-R-SWAT (Green).

monthly time step for the periods 1982–1999. The proposed calibration approaches provided convenient and adaptable platforms for experimenting with advanced hydrological models' parameter sensitivity and optimization tools at large basins scale.

This study has shown that the SWAT+ model setups produced approximately 2,429 HRUs, which is larger compared to the SWAT 2012 about 1,774 HRUs. SWAT+ generated greater surface runoff than the SWAT 2012 model did for both the CSFR and CHIRPS data sets. On the other hand, the SWAT model produced higher groundwater flow compared to the SWAT+ model for both CSFR and CHIRPS data. It was also shown that the CSFR data showed a high value of precipitation compared to the CHIRPS data, especially in Mahavavy, Betsiboka, Manambolo, and Tsiribihina basins. The ET was very low for the Maningory, Mangoro, and Mananara basins for both rainfall datasets. Based on the results, it can be concluded that the performance of the SWAT+ model for runoff simulation within the SWAT + Toolbox and R-SWAT was unsatisfactory for most basins, whether considering the CHIRPS or CSFR dataset. It has been demonstrated that the multi-gauge calibration was able to sufficiently simulate the streamflow for Betsiboka, Mahavavy, Tsiribihina, Mangoro, and Mangoky basins, with NSE ranging from 0.40 to 0.70 during calibration and validation in both SWAT + Toolbox and R-SWAT. Our study also found that multi-gauge calibrations using SWAT + Toolbox and R-SWAT did not significantly improve simulation performance in the Manambolo, Onilahy, Mananara, and Mandrare basins compared to the single-site calibration using SWAT-CUP software. Previous studies simulated the streamflow in Africa and found a negative value of the NSE ($NSE < 0$) in Manambolo, Onilahy, Mananara, and Mandrare basins (Schuol *et al.* 2008; Xie *et al.* 2012). An implication of these findings is the potential of the SWAT + Toolbox compared to R-SWAT in calibrating runoff across the MRBM. Calibration in both methods led to a reduction in surface runoff, percolation, water yield, and curve number but increased the lateral flow, ET, and groundwater flow. The results indicated that CSFR data produced substantial quantities of precipitation, surface runoff, percolation, water yield, and curve number, whereas the CHIRPS dataset resulted in elevated ET and PET. Despite the simulation flexibility offered by SWAT +, we conclude that CSRF-driven SWAT 2012 in combination with the SWAT-CUP calibration method is more appropriate for runoff simulation in the MRBM. The SWAT 2012 is a stable, well-established, and validated tool compared to the SWAT+ which is still under active development. The SWAT 2012 has a long history of applications and extensive validation studies, making it a trusted tool for hydrological simulations. We acknowledge that the optimization algorithm's influence is just one aspect of model calibration. The uncertainties associated with precipitation estimates play a substantial role in influencing the estimation of hydrological model parameters and water balance components (Fernandez-Palomino *et al.* 2022; Wang *et al.* 2023; Ye *et al.* 2012). This broader perspective underscores the complexity of calibration processes and the need for comprehensive evaluations.

The most important limitation lies in the fact that the model performance depends not only on the choice of calibration method but, more significantly, on factors such as limited observed discharge data, the uncertainty in rainfall data, the influential parameters, and the objective function used for the calibration approach. Moreover, multi-objective calibration in large watersheds may have limitations due to the heterogeneity of factors influencing hydrological response, such as climate, land use, and soil. Calibration is an ongoing process offering constant chances for improvement. It is crucial to acknowledge that the discrepancy in estimating precipitation is a constraint in our study. Future research could address this issue by allocating precipitation through the estimation of areal averages for each sub-basin. Estimating areal averages for each sub-basin based on the gridded precipitation product could potentially alleviate variations in precipitation values arising from watershed discretization. Therefore, further study of the issue is still required for the MRBM. This study offers some important insights into multi-gauge calibration using different calibration methods and rainfall input data on a large basin scale.

ACKNOWLEDGEMENTS

The authors sincerely thank the Madagascar National Meteorological and Hydrological Service for providing the observed discharge data, the Climate Hazards Center (CHC) at the University of California, Santa Barbara (UCSB) and the National Oceanic and Atmospheric Administration (NOAA) for providing the CHIRPS and CSFR climate data. We are also grateful to the SWAT + and R-SWAT online communities for their support.

FUNDING

This work was supported by JST, the establishment of university fellowships towards the creation of science and technology innovation, Grant Number JPMJFS2117, University of Yamanashi.

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., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J., Zobrist, J. & Srinivasan, R. 2007 Modelling hydrology and water quality in the pre-alpine/alpine thur watershed using SWAT. *Journal of Hydrology* **333** (2–4), 413–430. https://doi.org/10.1016/j.jhydrol.2006.09.014.

Akoko, G., Le, T. H., Gomi, T. & Kato, T. 2021 A review of SWAT model application in Africa. *Water* **13** (9), 1313. https://doi.org/10.3390/w13091313.

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. doi:10.13031/ 2013.42256.

Bayissa, Y., Tadesse, T., Demisse, G. & Shiferaw, A. 2017 Evaluation of satellite-based rainfall estimates and application to monitor meteorological drought for the upper blue Nile basin, Ethiopia. *Remote Sensing* **9** (7), 669. https://doi.org/10.3390/rs9070669.

Bieger, K., Arnold, J. G., Rathjens, H., White, M. J., Bosch, D. D., Allen, P. M., Volk, M. & Srinivasan, R. 2017 Introduction to SWAT +, a completely restructured version of the soil and water assessment tool. *Journal of the American Water Resources Association* **53**, 115–130. https://doi.org/10.1111/1752-1688.12482.

Chawanda, C. J. 2021 SWAT+ Toolbox v0.7. swatplustoolboxinstallerv076.zip. https://doi.org/10.5281/zenodo.6331716 (Accessed 05 September 2023).

Chawanda, C. J., Arnold, J., Thiery, W. & van Griensven, A. 2020 Mass balance calibration and reservoir representations for large-scale hydrological impact studies using SWAT +. *Climatic Change* **163**, 1307–1327. https://doi.org/10.1007/s10584-020-02924-x.

Chawanda, C. J., Nkwasa, A., Thiery, W. & van Griensven, A. 2024 Combined impacts of climate and land-use change on future water resources in Africa. *Hydrology and Earth System Sciences* **28** (1), 117–138. https://doi.org/10.5194/hess-28-117-2024, 2024.

Dile, Y. T., Karlberg, L., Daggupati, P., Srinivasan, R., Wiberg, D. & Rockström, J. 2016 Assessing the implications of water harvesting intensification on upstream–downstream ecosystem services: a case study in the lake tana basin. *Science of the Total Environment* **542**, 22–35. https://doi.org/10.1016/j.scitotenv.2015.10.065.

Duan, Z., Tuo, Y., Liu, J., Gao, H., Song, X., Zhang, Z., Yang, L. & Mekonnen, D. F. 2019 Hydrological evaluation of open-access precipitation and air temperature datasets using SWAT in a poorly gauged basin in Ethiopia. *Journal of Hydrology* **569**, 612–626. https://doi.org/10.1016/j.jhydrol.2018.12.02.

Fernandez-Palomino, C. A., Hattermann, F. F., Krysanova, V., Lobanova, A., Vega-Jácome, F., Lavado, W., Santini, W., Aybar, C. & Bronstert, A. 2022 A novel high-Resolution gridded precipitation dataset for Peruvian and Ecuadorian watersheds: development and hydrological evaluation. *J Hydrometeorol* **23**, 309–336. https://doi.org/10.1175/JHM-D-20-0285.1.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A. & Michaelsen, J. 2015 The climate hazards infrared precipitation with stations – a new environmental record for monitoring extremes. *Scientific Data* **2** (1), 1–21. https://doi.org/10.1038/sdata.2015.66.

Gong, Y., Shen, Z., Liu, R., Hong, Q. & Wu, X. 2012 A comparison of single-and multi-gauge based calibrations for hydrological modeling of the upper daning river watershed in China's three gorges reservoir region. *Hydrology Research* **43** (6), 822–832. https://doi.org/10.2166/nh.2012.021.

INSTAT 2019 Résultats Globaux Du Recensement Général De La Population Et De L'habitation De 2018 De Madagascar (RGPH-3) Tome 1, INSTAT-CCER, 2019. www.instat.mg/p/resultats-definitifs-du-rgph-3.

Kannan, N., Santhi, C. & Arnold, J. G. 2008 Development of an automated procedure for estimation of the spatial variation of runoff in large river basins. *Journal of Hydrology* **359** (1–2), 1–15. https://doi.org/10.1016/j.jhydrol.2008.06.00.

Le, A. M. & Pricope, N. G. 2017 Increasing the accuracy of runoff and streamflow simulation in the nzoia basin, western Kenya, through the incorporation of satellite-derived CHIRPS data. *Water* **9** (2), 114. https://doi.org/10.3390/w9020114.

Li, Y. & DeLiberty, T. 2021 Evaluating hourly SWAT streamflow simulations for urbanized and forest watersheds across northwestern Delaware, US. *Stochastic Environmental Research and Risk Assessment* **35**, 1145–1159. https://doi.org/10.1007/s00477-020-01904-y.

McIntyre, N., Wheater, H. & Lees, M. 2002 Estimation and propagation of parametric uncertainty in environmental models. *Journal of Hydroinformatics* **4** (3), 177–198. https://doi.org/10.2166/hydro.2002.0018.

Mohammadi, B., Vazifehkhah, S. & Duan, Z. 2024 A conceptual metaheuristic-based framework for improving runoff time series simulation in glacierized catchments. *Engineering Applications of Artificial Intelligence* **127**, 107302. https://doi.org/10.1016/j.engappai.2023.107302.

Moriasi, D., Arnold, J., van Liew, M., Bingner, R., Harmel, D. & Veith, T. 2007 Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE* **50**, 885–900. https://doi.org/10.13031/2013.23153.

Musie, M., Sen, S. & Srivastava, P. 2019 Comparison and evaluation of gridded precipitation datasets for streamflow simulation in data scarce watersheds of Ethiopia. *Journal of Hydrology* **579**, 124168. https://doi.org/10.1016/j.jhydrol.2019.124168.

Neitsch, S. L., Arnold, J. G., Kiniry, J. R. & Williams, J. R. 2011 *Soil and Water Assessment Tool Theoretical Documentation Version 2009*. Texas Water Resources Institute, College Station, TX. https://swat.tamu.edu/media/99192/swat2009-theory.pdf

Nguyen, T. V., Dietrich, J., Dang, D. T., Tran, D. A., Doan, B. V., Sarrazin, F. J., Abbaspour, K. & Srinivasan, R. 2022 An interactive graphical interface tool for parameter calibration, sensitivity analysis, uncertainty analysis, and visualization for the soil and water assessment tool. *Environmental Modelling & Software* **156**, 105497. https://doi.org/10.1016/j.envsoft.2022.105497.

Odusanya, A. E., Mehdi, B., Schürz, C., Oke, A. O., Awokola, O. S., Awomeso, J. A., Adejuwon, J. O. & Schulz, K. 2019 Multi-site calibration and validation of SWAT with satellite-based evapotranspiration in a data-sparse catchment in southwestern Nigeria. *Hydrology and Earth System Sciences* **23** (2), 1113–1144. https://doi.org/10.5194/hess-23-1113-2019.

Ollivier, C. C., Carrière, S. D., Heath, T., Olioso, A., Rabefitia, Z., Rakoto, H. & Satgé, F. 2023 Ensemble precipitation estimates based on an assessment of 21 gridded precipitation datasets to improve precipitation estimations across Madagascar. *Journal of Hydrology: Regional Studies* **47**, 101400. doi:10.1016/j.ejrh.2023.101400.

Pechlivanidis, I. G., Jackson, B. M., Mcintyre, N. R. & Wheater, H. S. 2011 Catchment scale hydrological modeling: a review of model types, calibration approaches, and uncertainty analysis methods in the context of recent developments in technology and applications. *Global NEST Journal* **13** (3), 193–214.

Rabezanahary, T. M. F., Rahaman, M. H. & Zhai, J. 2021 Assessment of the future impact of climate change on the hydrology of the mangoky river, Madagascar using ANN and SWAT. *Water* **13** (9), 1239. https://doi.org/10.3390/w13091239.

Raholijao, N., Arivelo, T. A., Rakotomavo, Z. A. P. H., Voahangin-dRakotoson, D., Srinivasan, G., Shanmugasundaram, J., Dash, I. & Qiu, J. 2019 *Les Tendances Climatiques et les Futurs Changements Climatiques A Madagascar-2019*. Government of Madagascar, Antananarivo, Madagascar. https://open-library.cirad.fr/files/6/2390__Les_tendances_climatiques_et_les_futures_changements_climatiques_a_Madagascar_2019.pdf.

Rakotoarimanana, Z. H. & Ishidaira, H. 2022 Analysis of river basin management in Madagascar and lessons learned from Japan. *Water* **14** (3), 449. https://doi.org/10.3390/w14030449.

Ramahaimandimby, Z., Randriamaherisoa, A., Jonard, F., Vanclooster, M. & Bielders, C. L. 2022 Reliability of gridded precipitation products for water management studies: the case of the ankavia river basin in Madagascar. *Remote Sensing.* **14** (16), 3940. https://doi.org/10.3390/rs14163940.

Randriatsara, H. H. R. H., Hu, Z., Xu, X., Ayugi, B., Sian, K. T. C. L. K., Mumo, R. & Ongoma, V. 2022 Evaluation of gridded precipitation datasets over Madagascar. *International Journal of Climatology* **42** (13), 7028–7046. https://doi.org/10.1002/joc.7628.

Rathjens, H., Oppelt, N., Bosch, D. D., Arnold, J. G. & Volk, M. 2015 Development of a grid-based version of the SWAT landscape model. *Hydrological Processes* **29** (6), 900–914. https://doi.org/10.1002/hyp.10197.

Richardson, C. W. 1981 Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research* **17** (1), 182–190. https://doi.org/10.1029/WR017i001p00182.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y. T., Chuang, H. Y., Iredell, M. & Ek, M. 2014 The NCEP climate forecast system version 2. *Journal of Climate* **27** (6), 2185–2208. doi: http://dx.doi.org/10.1175/JCLI-D-12-00823.1.

Schuol, J., Abbaspour, K. C., Yang, H., Srinivasan, R. & Zehnder, A. J. 2008 Modeling blue and green water availability in Africa. *Water Resources Research* **44** (7), 212–221. https://doi.org/10.1029/2007WR006609.

Shrestha, M. K., Recknagel, F., Frizenschaf, J. & Meyer, W. 2016 Assessing SWAT models based on single and multi-site calibration for the simulation of flow and nutrient loads in the semi-arid onkaparinga catchment in South Australia. *Agricultural Water Management* **175**, 61–71. https://doi.org/10.1016/j.agwat.2016.02.009.

Singh, L. & Saravanan, S. 2022 Assessing streamflow modeling using single and multi-site calibration approach on bharathpuzha catchment, India: a case study. *Model. Earth Syst. Environ.* **8**, 4135–4148. https://doi.org/10.1007/s40808-022-01353-2.

Song, Y., Zhang, J. & Lai, Y. 2021 Influence of multisite calibration on streamflow estimation based on the hydrological model with CMADS inputs. *Journal of Water and Climate Change* **12** (7), 3264–3281. https://doi.org/10.2166/wcc.2021.115.

Sorooshian, S. & Gupta, V. K. 1995 Model calibration. Computer models of watershed hydrology edited by Singh, V.P. Water Resources Publication, Highlands Ranch, Co.

Tiandraza, F., Qu, S., Hu, S., Mkono, C. N., Tikhomirova, A. & Randrialahamady, S. N. 2023 Response of ecosystem services to land Use change in Madagascar island, Africa: a Mul-ti-Scale perspective. *International Journal of Environmental Research and Public Health* **20** (4), 3060. https://doi.org/10.3390/ijerph20043060.

Tolson, B. A. & Shoemaker, C. A. 2007 Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research* **43** (1), W01413. https://doi.org/10.1029/2005WR004723.

Tuo, Y., Duan, Z., Disse, M. & Chiogna, G. 2016 Evaluation of precipitation input for SWAT modeling in alpine catchment: a case study in the adige river basin (Italy). *Science of the Total Environment* **573**, 66–82. https://doi.org/10.1016/j.scitotenv.2016.08.034.

Udías, A., Malago, A., Pastori, M., Vigiak, O., Reynaud, A., Elorza Tenreiro, F. J. & Bouraoui, F. 2016 A Decision Support Tool (R-SWAT-DS) for Integrated Watershed Management. In *8th International Congress on Environmental Modelling and Software; International Environmental Modelling and Software Society (Organiser). Toulouse, (France): Proceedings of the 8th International Congress on Environmental Modelling and Software*. pp. 944–951. JRC102736. https://core.ac.uk/reader/148688510.

Upadhyay, S., Silwal, P., Prajapati, R., Talchabhadel, R., Shrestha, S., Duwal, S. & Lakhe, H. 2022 Evaluating magnitude agreement and occurrence consistency of CHIRPS product with ground-based observations over medium-sized river basins in Nepal. *Hydrology* **9** (8), 146. https://doi.org/10.3390/hydrology9080146.

Wagener, T., Wheater, H. S. & Gupta, H. V. 2004 *Rainfall-Runoff Modelling in Gauged and Ungauged Catchments*. Imperial College Press, London, UK, pp. 1–306. doi:10.1142/p335.

Wagner, P. D., Bieger, K., Arnold, J. G. & Fohrer, N. 2022 Representation of hydrological processes in a rural lowland catchment in northern Germany using SWAT and SWAT + . *Hydrological Processes* **36** (5), e14589. https://doi.org/10.1002/hyp.14589.

Wang, S., Zhang, Z., Sun, G., Strauss, P., Guo, J., Tang, Y. & Yao, A. 2012 Multi-site calibration, validation, and sensitivity analysis of the MIKE SHE model for a large watershed in northern China. *Hydrology and Earth System Sciences* **16** (12), 4621–4632. https://doi.org/10.5194/hess-16-4621-2012.

Wang, J., Zhuo, L., Han, D., Liu, Y. & Rico-Ramirez, M. A. 2023 Hydrological model adaptability to rainfall inputs of varied quality. *Water Resour Res* **59**, e2022WR032484. https://doi.org/10.1029/2022WR032484.

Wu, Y. & Liu, S. 2012 Automating calibration, sensitivity, and uncertainty analysis of complex models using the R package flexible modeling environment (FME): sWAT as an example. *Environmental Modelling & Software* **31**, 99–109. https://doi.org/10.1016/j.envsoft.2011.11.013.

Wu, Y. & Liu, S. 2014 Improvement of the R-SWAT-FME framework to support multiple variables and multi-objective functions. *Science of the Total Environment* **466**, 455–466. https://doi.org/10.1016/j.scitotenv.2013.07.048.

Wu, J., Yen, H., Arnold, J. G., Yang, Y. E., Cai, X., White, M. J., Santhi, C., Miao, C. & Srinivasan, R. 2020 Development of reservoir operation functions in SWAT+ for national environmental assessments. *Journal of Hydrology* **583**, 124556. https://doi.org/10.1016/j.jhydrol.2020.124556.

Xie, H., Longuevergne, L., Ringler, C. & Scanlon, B. R. 2012 Calibration and evaluation of a semi-distributed watershed model of Sub-Saharan Africa using GRACE data. *Hydrology and Earth System Sciences* **16** (9), 3083–3099. https://doi.org/10.5194/hess-16-3083-2012.

Ye, B., Yang, D. & Ma, L. 2012 Effect of precipitation bias correction on water budget calculation in upper Yellow River, China. *Environmental Research Letters* **7**, 025201. https://doi.org/10.1088/1748-9326/7/2/025201.

Yen, H., Park, S., Arnold, J. G., Srinivasan, R., Chawanda, C. J., Wang, R., Feng, Q., Wu, J., Miao, C., Bieger, K. & Daggupati, P. 2019 IPEAT + : A built-in optimization and automatic calibration tool of SWAT + . *Water* **11** (8), 1681. https://doi.org/10.3390/w11081681.
Yu, P. S., Yang, T. C. & Chen, S. J. 2001 Comparison of uncertainty analysis methods for a distributed rainfall–runoff model. *Journal of Hydrology* **244** (1–2), 43–59. https://doi.org/10.1016/S0022-1694(01)00328-6.

First received 30 November 2023; accepted in revised form 8 February 2024. Available online 27 February 2024