VICTORIA UNIVERSITY
MELBOURNE AUSTRALIA

Calibration, validation and uncertainty analysis of a SWAT water quality model

ORIGINAL ARTICLE

Calibration, validation and uncertainty analysis of a SWAT water quality model

Sushil K. Das[1] · Amimul Ahsan[2,3] · Md. Habibur Rahman Bejoy Khan[2] · Abdullah Gokhan Yilmaz[4] · Shakil Ahmed[5,6] · Monzur Imteaz[3] · Muhammad Atiq Ur Rehman Tariq[7,8] · Md. Shafiquzzaman[9] · Anne W. M. Ng[8,10] · Nadhir Al-Ansari[11]

Abstract
Sediment and nutrient pollution in water bodies is threatening human health and the ecosystem, due to rapid land use changes and improper agricultural practices. The impact of the nonpoint source pollution needs to be evaluated for the sustainable use of water resources. An ideal tool like the soil and water assessment tool (SWAT) can assess the impact of pollutant loads on the drainage area, which could be beneficial for developing a water quality management model. This study aims to evaluate the SWAT model's multi-objective and multivariable calibration, validation, and uncertainty analysis at three different sites of the Yarra River drainage area in Victoria, Australia. The drainage area is split into 51 subdrainage areas in the SWAT model. The model is calibrated and validated for streamflow from 1990 to 2008 and sediment and nutrients from 1998 to 2008. The results show that most of the monthly and annual calibration and validation for streamflow, nutrients, and sediment at the three selected sites are found with Nash–Sutcliffe efficiency values greater than 0.50. Furthermore, the uncertainty analysis of the model shows satisfactory results where the p-factor value is reliable by considering 95% prediction uncertainty and the d-factor value is close to zero. The model's results indicate that the model performs well in the river's watershed, which helps construct a water quality management model. Finally, the model application in the cost-effective management of water quality might reduce pollution in water bodies due to land use and agricultural activities, which would be beneficial to water management managers.

Keywords Land use · Sediment · Phosphorus · Nitrogen · SWAT · Yarra River

✉ Amimul Ahsan
aa@iut-dhaka.edu

✉ Nadhir Al-Ansari
nadhir.alansari@ltu.se

1 College of Engineering and Science, Victoria University, Melbourne, VIC, Australia

2 Department of Civil and Environmental Engineering, Islamic University of Technology, Gazipur, Bangladesh

3 Department of Civil and Construction Engineering, Swinburne University of Technology, Melbourne, Australia

4 La Trobe University, Melbourne, VIC 3086, Australia

5 Department of Civil Engineering, East West University, Dhaka, Bangladesh

6 Global Institute for Water Security, University of Saskatchewan, Saskatchewan, Canada

7 Centre of Excellence in Water Resources Engineering, University of Engineering and Technology, Lahore 54890, Pakistan

8 Faculty of Science and Technology, Charles Darwin University, Ellengowan, Dr, Brinkin, NT 0810, Australia

9 Department of Civil Engineering, College of Engineering, Qassim University, Buraidah 51452, Kingdom of Saudi Arabia

10 Energy and Resources Institute, Charles Darwin University, Darwin, Ellengowan, Dr, Brinkin, NT 0810, Australia

11 Department of Civil, Environmental and Natural Resources Engineering, Lulea University of Technology, 971 87 Lulea, Sweden

Introduction

The excessive land use changes and various agricultural practices are causing the accumulation of sediments in water bodies and a shortage of potable water, which threatens human health and ecosystems. The impact of the alteration of tropical rainforest to agricultural land decreases the porosity of the topsoil resulting in leaching of nutrients, incremental runoff, eutrophication of water and erosion. The complexity of agricultural land management is immense as several parameters, such as chemical transport, hydrological factors, topographical features, land use patterns, and soil characteristics, are required to determine the transport mechanism of non-point source (NPS) pollution. The management of drainage areas with mathematical computer models is becoming a cost-efficient evaluation means in abridging the difficult challenges (water pollution) from agricultural practices and land use activities (Guo et al. 2002; Ahsan et al. 2023).

The unique hydrological location and limited water quality data availability have greatly influenced the application of water quality models for the drainage areas in Australia (Grayson et al. 1999; Croke and Jakeman 2001; Ahsan et al. 2023). The limited information on soil types, erosion or geographical ecosystem and land use data also complicated the advancement of water quality models (Kragt and Newham 2009). A few investigators (Borah and Bera 2003; Ahsan et al. 2023) recommended the implementation of physics-based models in agricultural NPS contamination modeling as the nature of pollution is scattered and compulsive. Thus, creating an efficient water quality management strategy, becomes a great challenge for the drainage area in the poor data situations of Australia. U.S. EPA (2002) suggested the use of the technically sound, vigorous, and justifiable model for generating model outcomes for projects ranging from regulation to investigation. For this, the ability of a model for assessing the simulation and forecast of streamflow and component of water quality is done by the calibration, validation, sensitivity, and uncertainty analysis (Rafik et al. 2023). Furthermore, it was found that the calibration and validation processes became more complicated due to the increased model parameters, multiple locations, and number of subdrainage areas responded to the forecasting (like streamflow, nitrogen, sediment, and phosphorus). In such a case, van Griensven et al. (2002) suggested that the number of factor adjustments during calibration can be reduced by a sensitivity analysis for the solution of such types of models. Multiple objectives are used in the calibration of a drainage area model to assess the pact between the observed and simulated values. The optimization of the objectives is performed using several goodness-of-fit

predictors (for example, coefficient of determination, and Nash–Sutcliffe efficiency), several variables (sediment, water, nutrients, and energy), and several sites (Niraula et al. 2012; Nasirzadehdizaji and Akyuz 2022).

The multiple factors in the water quality model complicate the calibration process by assessing the relationships between numerous estimated outcomes and one factor. This can happen when the alteration of one factor causes one estimated variable to more directly overlap with measured values and another estimated variable to less directly overlap with calculated values (White and Chaubey 2005). Thus, a stepwise calibration process in rational order is performed frequently, because of the overlapping between factors and estimated outcomes and measurement ambiguity (Madsen 2003). The SWAT model simplifies the calibration process problem by performing an orderly optimization of the objective function: (1) Surface runoff, (2) Total flow and base flow, (3) Nitrogen, (4) Phosphorus, and (5) Sediment (Santhi et al. 2001; Kirsch et al. 2002; Xue et al. 2022). A flowchart for the calibration of streamflow, nutrients, and sediment recommended by Arnold et al. (2012) (modified from Santhi et al. 2001) is applied in this research, as shown in Fig. 10 (Appendix). The hydrologic outcomes (base flow, total streamflow, and surface runoff) are initially calibrated as they cause an impact on the other outcome factors (nutrients and sediments). Furthermore, after hydrologic calibration, the sediment is calibrated as sediment can affect the transport of phosphorus in the catchment, and after phosphorus calibration, nitrogen calibration is performed due to the huge uncertainty in phosphorus estimation by the model from multiple sources (Campbell and Edwards 2001; Oduor et al. 2023). The validation of the model is performed similarly to the calibration process with the estimated and measured values in assessing, whether the objective function is fulfilled with a different dataset other than the calibration process (Aibaidula et al. 2023). The validation helps to assess the model whether calibrated properly if not, and then revision of the calibration process needs to be done again.

The uncertainty of the model seeks to quantitatively evaluate the reliability of the model's outcome. The various sources of modeling uncertainties should be analyzed, and the outcome of the model can be characterized with an acceptable limit (Gupta et al. 1998; Vrugt et al. 2003; Aibaidula et al. 2023). The extent of the uncertainties in the model was calculated using the p-factor and d-factor values (Abbaspour et al. 2004). The ideal value of the p-factor was considered, where the uncertainty was 95% in the prediction of the observed data. However, the d-factor ideal value is close to zero, which means that the uncertainty in the estimated outcome is reduced.

The study aims to evaluate the calibration, validation, and uncertainty analysis of a water quality model at different sites of the middle Yarra River drainage area in

Victoria, Australia with SWAT. The calibration of the streamflow at three sites is done from 1990 to 2002 and validation from 2003 to 2008. Furthermore, the calibration of the sediment and nutrients is done from 1998 to 2004 and validation from 2005 to 2008. This is the first time at three different sites calibration, validation, and uncertainty analysis are done for managing the quality of water.

Methodology

Location

In the Victoria state of Australia, the Yarra River covers a drainage area of 4000 km^2, which supplies potable water to the mass people and many agricultural industries. It covers most of the agricultural land cover other than natural vegetation and urban land cover. The drainage area is segregated into lower, middle, and upper portions because of land use activities.

The middle yarra drainage area (MYDA) is chosen for this study (Fig. 1), which covers an area of 1511 km^2 and is one of the largest pollutant flow sources in the Port Phillip Bay territory (EPA Victoria 1999; Melbourne Water 2010).

Model data insertion

The SWAT2005 model's interface, named ArcSWAT, is utilized in this study, which is created by the USDA-ARS as a hydrological tool for drainage areas (Arnold et al. 1998; Winchell et al. 2009). The SWAT model has the potential to simulate the effect of sediment, agricultural chemicals, and water loads on complicated drainage areas with differing land use, soils, and management situations over a long period. Moreover, the model also has a robust uncertainty, sensitivity, and autocalibration analysis tool. Due to its multiple facilities, the model has been widely recognized by many researchers to do long-duration continuous simulations in the agricultural drainage area (Gassman et al. 2015; Ahsan et al. 2023). Table 1 depicts the essential data along with their sources for the model's calibration and operation. The monitoring locations and digital input maps for streamflow and climate data are shown in Fig. 2.

The SWAT model applies ASTER 30 m GDEM in this study, as shown in Fig. 2a. The soil labels, as shown in Fig. 2b, follow the prevalent principal profile form illustrated by the Factual Key in brackets of the Australian Soil Classification System (Northcote 1979; Isbell 2002). In the drainage area, two prevailing soil varieties are 35% Dermosol and 54% Sodosol, and their several soil characteristics are presented. Furthermore, about 32% of the total
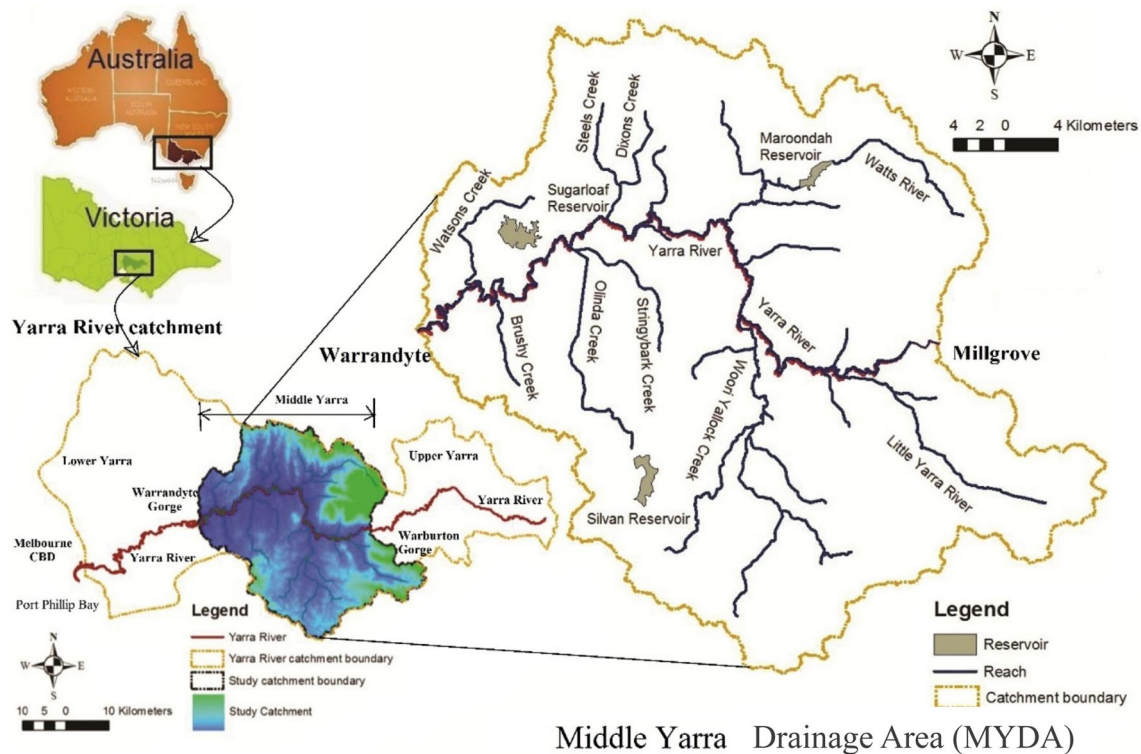

Fig. 1 The middle yarra drainage area (MYDA) (Das et al. 2022)

Table 1 SWAT model data reference (Das et al. 2022)

Data	Reference
Digital elevation model (DEM)	ASTER 30 m GDEM, jointly developed by The Ministry of Economy, Trade, and Industry (METI) of Japan and the United States National Aeronautics and Space Administration (NASA) (http://asterweb.jpl.nasa.gov/gdem-wist.asp)
Soil	Atlas of Australian Soils from Australian Soil Resource Information System (ASRIS) developed by CSIRO and Department of Agriculture, Fisheries and Forestry (DAFF) (http://www.asris.csiro.au)
Land use	50 m grid raster data for the period of 1997 to May 2006 collected from Australian Bureau of Agricultural and Resource Economics and Sciences (ABARES) (http://adl.brs.gov.au/landuse)
Climate	16 rainfall stations, and 4 temperature (max and min), solar radiation, wind speed and relative humidity stations data from SILO climate database (http://www.longpaddock.qld.gov.au/silo) and Bureau of Meteorology (BoM) (http://www.bom.gov.au/climate/data/)
Streamflow and water quality	Daily streamflow and water quality monthly grab sample data for the monitoring stations from Melbourne Water (http://www.melbournewater.com.au/)
Crop management practices	Manure, fertilizer type and application rate, tillage practices, cropping seasons, and irrigation rate from Australian Bureau of Statistics (http://www.abs.gov.au), Department of Environment and Primary Industries (http://www.depi.vic.gov.au/) and Melbourne Water (http://www.melbournewater.com.au/)
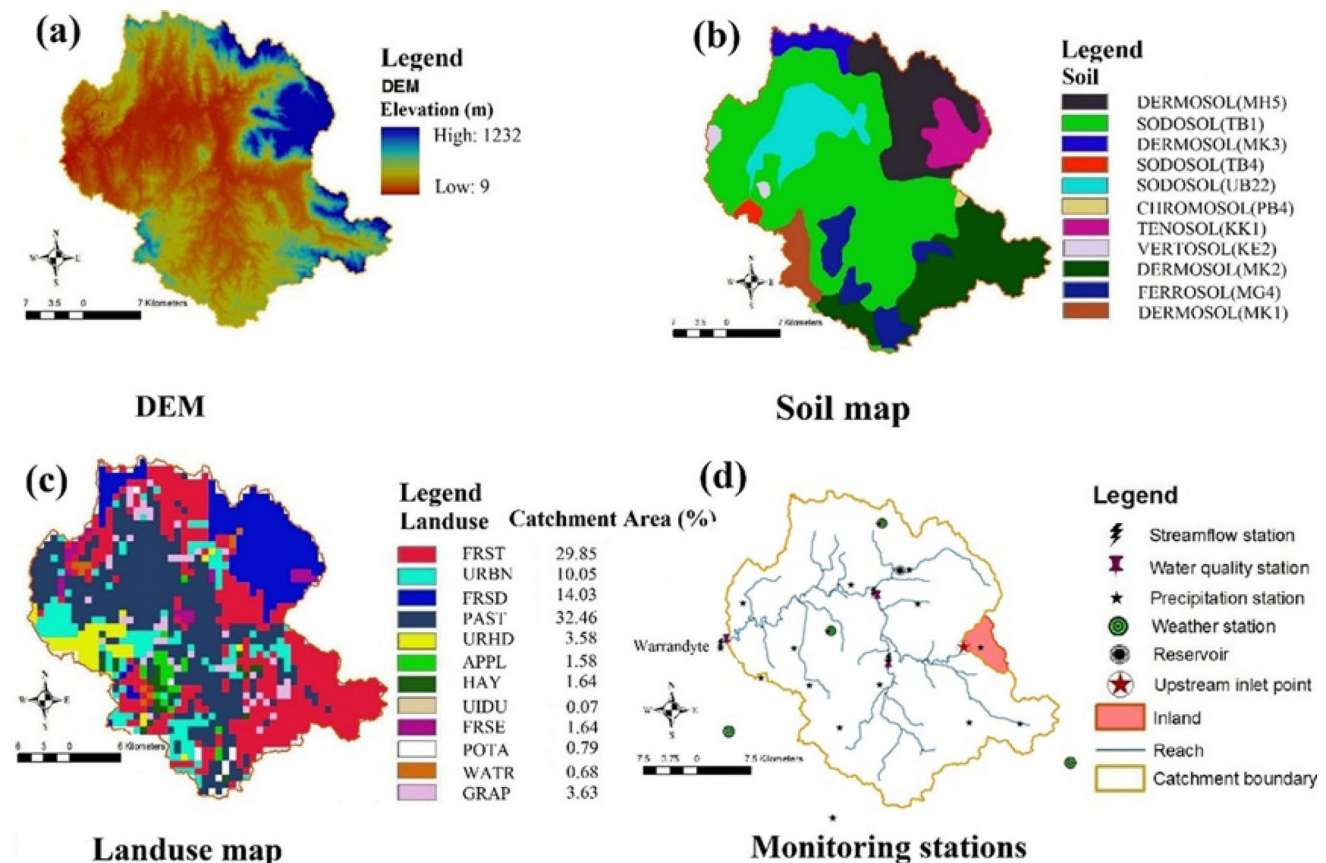
Fig. 2 Soil varieties, DEM, land use, and monitoring locations map for the SWAT model at the MYDA (Das et al. 2022)

region is covered by pastures, as shown in Fig. 2c of the MYDA. As per the SWAT model guideline, the land use category is re-categorized and created in accordance with the conditions of the SWAT model for the MYDA, as the model has a pre-defined land use category that generates a link with the land use map (Winchell et al. 2009).

The climate data are compiled for the 1980–2008 phase. The MYDA average monthly temperature and

precipitation are shown in Fig. 11 (Appendix). The highest rainfall appears in September, and the lowest rainfall appears in February. Thus, the highest temperature alters from 11.4 °C (July) to 25.3 °C (February), and the lowest temperature alters from 4.4 °C (July) to 12.3 °C (February). An immediate decrease in yearly mean precipitation (from 1140 to 922 mm) from 1997 forward is shown in Fig. 12 (Appendix), implying terrible drought events in the MYDA, recognized as a millennium drought.

The calibration of base flow, total streamflow, and surface runoff can accurately represent the subsurface and surface hydrological processes. In this study, an automatic digital filter-based software named "Base flow Filter Program" (USDA-ARS 1999) is utilized to segregate base flow (Arnold et al. 1995). In the MYDA, the base flow split revealed that it contributes mostly to the total streamflow (about 75%). The addition of streamflow and water quality from the Upper Yarra into the MYDA is done by the SWAT model "upstream inlet point" function, and the streamflow and water quality data are compiled from the Millgrove station (Fig. 2d).

SWAT model: system, calibration, validation, sensitivity, and uncertainty analysis

In accordance with the instructions of the SWAT 2005 version interface named ArcSWAT, all the required datasets and input database files (Winchell et al. 2009) are compiled and arranged for this study model. The MYDA is partitioned into 51 subdrainage areas and 431 hydrological response divisions (HRDs). Curve Number (C.N.), Penman–Monteith, and Muskingum techniques to assess runoff, PET, and channel routing with NPK transformation are used for hydrological method modeling.

The inbuilt sensitivity, auto-calibration, and uncertainty tool of the SWAT performs the uncertainty, calibration, and sensitivity analysis of the model in the MYDA three sites. In the SWAT model, 26 streamflow, 6 sediments, and 9 nutrient factors are present, where each factor has a primary value and default limits. During the calibration operation, the model outcome variables are altered by evaluating the preliminary values of the factors that are sensitive, while other factors remain the same. In order to assess, the streamflow factors in the SWAT model, as per Van Griensven et al. (2006), the LH-OAT (Latin-Hypercube and One-Factor-At-a-Time) sensitivity analysis technique is utilized. Furthermore, the auto-calibration of ParaSol (SCE-UA) is applied to the streamflow factors by assessing the sensitivity outcomes. Moreover, ParaSol (SCE-UA) also delivers options for determining uncertainty analysis (Van Liew and Veith 2010; Green and Van Griensven 2008).

Selected sites of the MYDA for calibration, validation, and uncertainty analysis

The assessment of the model in supplement to the graphical approaches mentioned by Moriasi et al. (2007) is done by Nash–Sutcliffe efficiency (NSE), percent bias (PBIAS), and the ratio of the root mean square error to the standard deviation of calculated data (RSR). The optimum value of PBIAS and RSR is 0, where negative and positive signs of PBIAS imply overprediction and underprediction, respectively. As suggested by Moriasi et al. (2007), a simulated model can be evaluated as satisfactory if RSR ≤ 0.70 and, NSE > 0.50 and if PBIAS < ± 25% for Streamflow and NSE > 0.50 and RSR ≤ 0.70, and if PBIAS ≤ ± 55% for TSS and PBIAS ≤ ± 70% for TN and TP for a monthly time step. The uncertainty analysis is assessed as "good" simulations if a 95% probability range of the p-factor and d-factor estimation for each simulated factor is achieved, otherwise termed as "not good" simulation (Van Liew and Veith 2010; Green and Van Griensven 2008). Finally, the coefficient of correlation (R^2) is considered for the assessment of the model outcome.

Melbourne Water Corporation's monitoring network is the main streamflow and water quality monitoring system, which is placed in the Yarra River drainage area. The Yarra River drainage area comprises 70 streamflow gauging stations and 33 water quality monitoring stations for measuring the varying quality of water. By assessing the stations, three sites are selected based on the multisite, multi-objective, and multivariable calibration objectives of MYDA water quality management, as shown in Fig. 3. They are mentioned as Site-1, Site-2, and Site-3. The locations of the sites on the basis of streamflow and water quality monitoring stations are shown in Tables 6 and 7 (Appendix). Furthermore, the three sites are spatially distributed over the entire MYDA, and site 3 is the drainage area outlet.

Results and discussion

Calibration and validation of SWAT model at different sites of MYDA

Streamflow

In accordance with the sensitivity results, 15 factors are categorized as extremely significant and significant in accordance with the classification of ratings by Van Griensven et al. (2006). Details of the factors, durations of calibration and validation, ParaSol (SCE-UA) auto-calibration (Reckhow 1994; Aibaidula et al. 2023), and manual tuning of the SWAT are found in Das et al. (2022) and Ahsan et al. (2023).

Fig. 3 Selected sites in the MYDA for calibration, uncertainty, and validation analysis
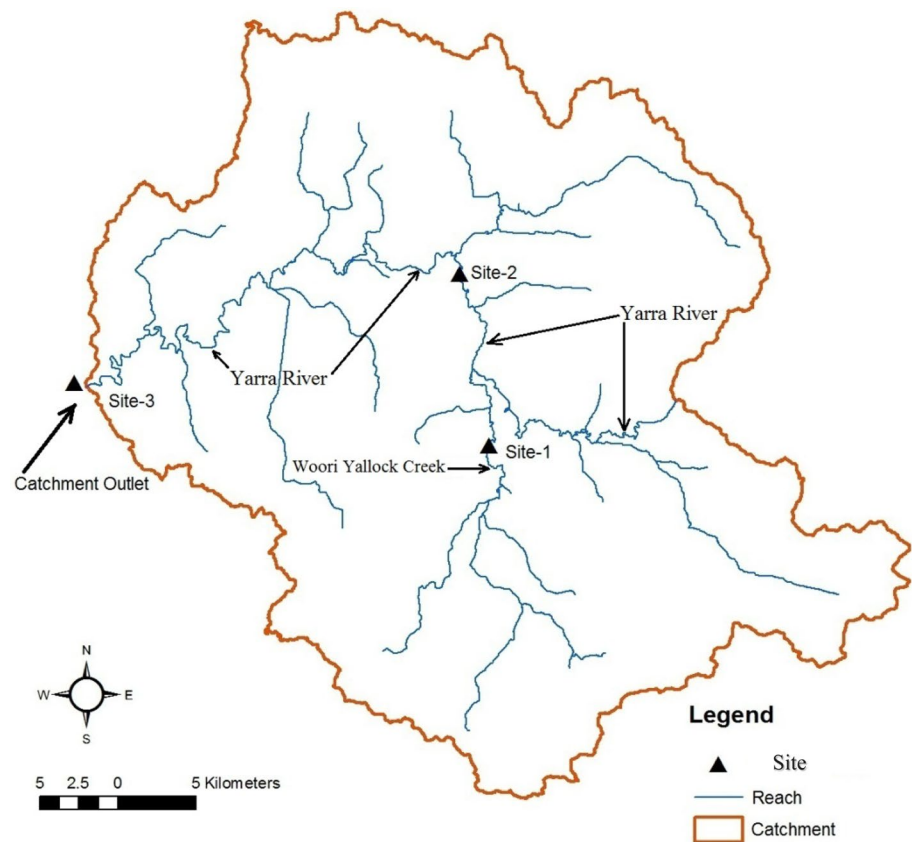
During the calibration period for total streamflow and base flow at all three sites (monthly and annual), the SWAT model operates very well ($R^2 > 0.75$, NSE > 0.75, RSR < 0.60, and PBIAS 10%, as shown in Table 2). Likewise, runoff calibrations (monthly and annually) at all 3 sites are also acceptable, as per the suggestion of Moriasi et al. (2007). However, the site 2 calibration outcome (monthly and annual) (NSE > 0.78, $R^2 > 0.90$, RSR < 0.50, and PBIAS $< 22\%$) of total streamflow, base flow, and runoff imply that it is more satisfactory than the other 2 sites. Through the calibration, the model, in general, underpredicts the flows in monsoon (1990–1996) and overpredicts the flows in summer (1997–2002) based on monthly data. The overprediction in the dry period may be due to the underprediction of the usage of water in irrigation, as the duration of the usage of water is challenging to evaluate due to the storage. Overall, during calibration, the model underpredicts the runoff, base flow, and total streamflow (monthly and annual), whereas the PBIAS outcome is positive, representing underestimation. Furthermore, the runoff underprediction is much higher than base flow, where PBIAS values of runoff are greater.

In validation, total streamflow performance ratings are satisfactory based on annual and monthly data at 3 sites (NSE > 0.75, RSR < 0.50, $R^2 > 0.75$, and PBIAS 10%, as shown in Table 2), but the annual value of site 1 is

unsatisfactory (NSE < 0.50 and RSR > 0.7). However, the base flow and runoff results of site 1 are unsatisfactory, with NSE < 0.70 and RSR > 0.70, even though there is some easing on the monthly and annual step guideline. Furthermore, the validation of sites 2 and 3 is satisfactory in terms of runoff and base flow.

The SWAT model simulation findings from this study are found to be reliable with other SWAT research such as Kirsch et al. (2002), Green and Van Griensven (2008), and Gasirabo et al. (2023). In NSW, the Mooki catchment also overestimates some lower flows and small peaks, while underestimating the peak runoff in the SWAT model simulation (Vervoort 2007). Additionally, Watson et al. (2003) found that the Woady Yaloak River watershed overpredicts the low flows in the SWAT model application in Victoria.

Sediment and nutrients

In accordance with the sensitivity analysis findings, 13 parameters are rated as extremely significant, as per the rating classification by Van Griensven et al. (2006). These are CH_COV, CH_EROD, NPERCO, PHOSKD, PPERCO, RCHRG_DP, SOL_LABP, SOL_NO3, SOL_ORGN, SOL_ORGP, SPCON, SPEXP, and ULSE_P from higher to lower levels, respectively. ParaSol (SCE-UA) auto-calibration is done at the MYDA's outlet on the 13 most important

Table 2 Calibration (1990–2002) and validation (2003–2008) of streamflow at 3 sites of the MYDA

			Monthly				Annual			
			R^2	NSE	PBIAS	RSR	R^2	NSE	PBIAS	RSR
Total streamflow	Calibration	Site-1	0.81	0.79	10	0.45	0.85	0.79	10	0.46
		Site-2	0.94	0.93	6	0.27	0.96	0.92	6	0.28
		Site-3	0.93	0.89	10	0.34	0.96	0.87	10	0.36
	Validation	Site-1	0.80	0.76	22	0.49	0.77	0.49	20	0.72
		Site-2	0.88	0.88	1	0.35	0.94	0.85	1	0.38
		Site-3	0.82	0.82	-3	0.43	0.87	0.81	− 3	0.43
Base flow	Calibration	Site-1	0.83	0.60	6	0.63	0.95	0.94	2	0.25
		Site-2	0.93	0.93	2	0.26	0.95	0.94	2	0.25
		Site-3	0.93	0.89	6	0.33	0.95	0.88	6	0.35
	Validation	Site-1	0.71	0.56	26	0.66	0.67	0.22	26	0.89
		Site-2	0.83	0.83	-6	0.41	0.90	0.77	− 6	0.48
		Site-3	0.81	0.79	-11	0.46	0.84	0.71	− 11	0.54
Runoff	Calibration	Site-1	0.66	0.59	20	0.64	0.84	0.63	20	0.60
		Site-2	0.91	0.84	21	0.40	0.98	0.79	21	0.46
		Site-3	0.84	0.80	23	0.45	0.97	0.76	23	0.49
	Validation	Site-1	0.85	0.77	11	0.47	0.82	0.65	11	0.59
		Site-2	0.87	0.85	19	0.39	0.98	0.77	19	0.47
		Site-3	0.82	0.79	19	0.46	0.87	0.70	19	0.55

Positive and negative PBIAS values mean underprediction and overprediction, respectively in percent. Monthly simulations are satisfactory if NSE > 0.50 and RSR ≤ 0.70, and if PBIAS ± 25% for streamflow as per Moriasi et al. (2007)

parameters (as stated above) for sediment (TSS) and nutrients (TP and TN). The simulated values for sediment (TSS) and nutrients (TP and TN) are calibrated with the data from 1998 to 2004, and then confirmed with the data from 2005 to 2008. This time period captures all changes that could happen in the patterns of sediment and nutrients. In addition, manual tuning of the SWAT model in-stream and nutrient cycling parameters is applied for in-stream and nutrient cycling process calibration. The different sites of the MYDA's calibration and validation are shown in Table 3.

Warrandyte (Site-3) In this site, the calibration outcomes for the monthly are satisfactory, but the annual are very good for sediment (TSS monthly: NSE > 0.50, RSR < 0.70, R^2 > 0.75, and PBIAS < 15%, and TSS annual: NSE > 0.75, RSR < 0.50, R^2 > 0.90, and PBIAS < 15%, as shown in Table 3), as per Moriasi et al. (2007). Likewise, for sites- 1 and 2, the calibration results of the TN monthly and annual are also very good. However, the TP monthly outcomes are satisfactory (NSE > 0.50, RSR ≤ 0.70, R^2 > 0.50, and PBIAS < 25%), and the annual outcomes are very good (NSE > 0.65, RSR < 0.70, R^2 > 0.85, and PBIAS < 25%), as shown in Table 3. Likewise, for sites- 1 and 2, the model underpredicts the monthly peak loads for TP and TSS loads (Fig. 4), but overpredicts the TN loads.

In the validation operation, the model outcomes were satisfactory for monthly and annual TSS loads (NSE > 0.65, R^2 > 0.75, RSR < 0.60, and PBIAS < ± 30%, as shown

in Table 3), as per Moriasi et al. (2007). The TN loads (monthly) are very good (NSE > 0.75, R^2 > 0.80, RSR < 0.50, and PBIAS < ± 25%), as shown in Table 3. Furthermore, the validation outcome of TP monthly loads is satisfactory (NSE > 0.65, R^2 > 0.95, RSR < 0.60, but PBIAS ≤ ± 70), but annual loads, are unsatisfactory (RSR > 0.70 and NSE < 0.50), as shown in Table 3. The model overpredicts the monthly loads as shown in Fig. 5, as the PBIAS values are negatively shown in Table 3. The TSS and TN overpredictions, are much less than the TP.

Yarra Grange (Site-2) In this region, the calibration outcomes for the monthly and annual sediment are very good (NSE > 0.75, RSR < 0.50, R^2 > 0.90, and PBIAS < 15%, as shown in Table 3), as per Moriasi et al. (2007). Likewise, the results of the TN loads of monthly and annuals are also very good (NSE > 0.75, RSR < 0.50, R^2 > 0.90, and PBIAS < ± 25% shown in Table 3). However, the TP outcomes are unsatisfactory, like at Site-1 (RSR > 0.70 and NSE ≤ 0.50), as shown in Table 3. The model underpredicts the monthly peak loads for TP and TSS loads (Fig. 6) but overpredicts the TN loads for similar reasons at Site-1.

In the validation operation, the model outcomes are satisfactory for monthly and annual TSS loads (monthly: NSE > 0.65, R^2 > 0.85, PBIAS < ± 30% but RSR < 0.70, and annual: E_{NS}^2 > 0.65, R^2 > 0.85, RSR < 0.60, and

Table 3 Calibration (1998–2004) and validation (2005–2008) of TSS, TN, TP at 3 sites of the MYDA

			Monthly				Annual			
			R^2	NSE	PBIAS	RSR	R^2	NSE	PBIAS	RSR
Sediment (TSS)	Calibration	Site-1	0.67	0.67	8	0.58	0.84	0.74	8	0.51
		Site-2	0.92	0.89	14	0.32	0.91	0.80	14	0.45
		Site-3	0.77	0.62	2	0.61	0.91	0.78	2	0.47
	Validation	Site-1	0.97	0.63	− 29	0.61	0.99	0.77	− 29	0.48
		Site-2	0.86	0.66	− 17	0.63	0.85	0.72	− 17	0.53
		Site-3	0.78	0.68	− 17	0.56	0.85	0.69	− 17	0.55
Total nitrogen (TN)	Calibration	Site-1	0.80	0.77	− 8	0.48	0.93	0.76	− 8	0.49
		Site-2	0.92	0.92	− 1	0.28	0.98	0.98	− 1	0.15
		Site-3	0.86	0.79	15	0.46	0.96	0.79	15	0.46
	Validation	Site-1	0.67	0.55	− 21	0.67	0.72	0.43	− 21	0.75
		Site-2	0.84	0.83	− 8	0.42	0.99	0.89	− 8	0.32
		Site-3	0.83	0.78	− 5	0.47	0.98	0.95	− 5	0.22
Total phosphorus (TP)	Calibration	Site-1	0.47	0.46	13	0.74	0.58	0.49	13	0.72
		Site-2	0.48	0.45	18	0.74	0.74	0.50	18	0.71
		Site-3	0.53	0.51	19	0.70	0.89	0.69	19	0.56
	Validation	Site-1	0.87	0.61	− 81	0.62	0.95	0.05	− 81	0.97
		Site-2	0.95	0.62	− 53	0.61	0.94	0.01	− 53	1
		Site-3	0.97	0.71	− 56	0.54	0.97	0.35	− 56	0.81

Positive and negative PBIAS values mean under prediction and over prediction, respectively in percent. Monthly simulations are satisfactory if NSE > 0.50 and RSR ≤ 0.70, and if PBIAS ≤ ± 55% for TSS and PBIAS ≤ ± 70% for TN and TP as per Moriasi et al. (2007)
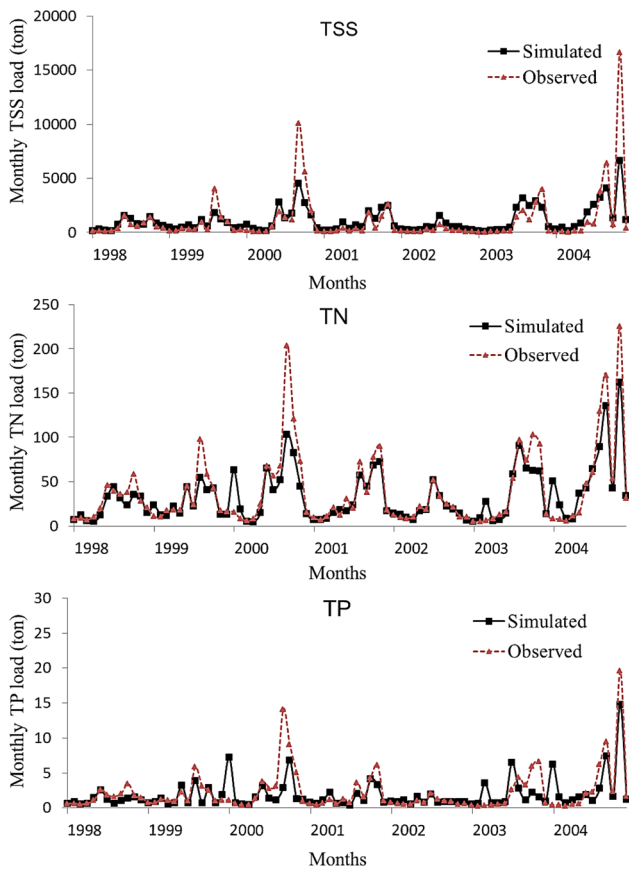

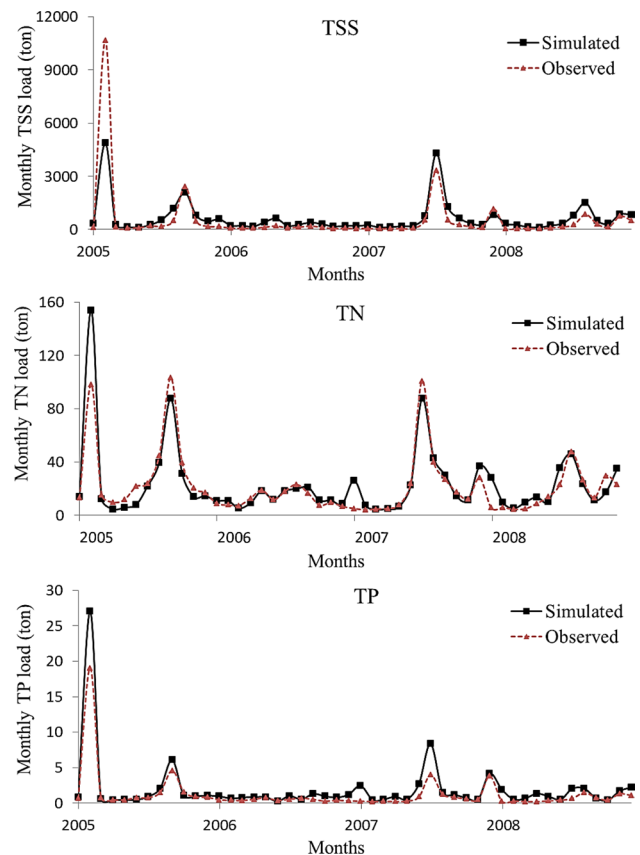**Fig. 4** Warrandyte (Site-3) monthly calibration of TP, TN and TSS (Ahsan et al. 2023)

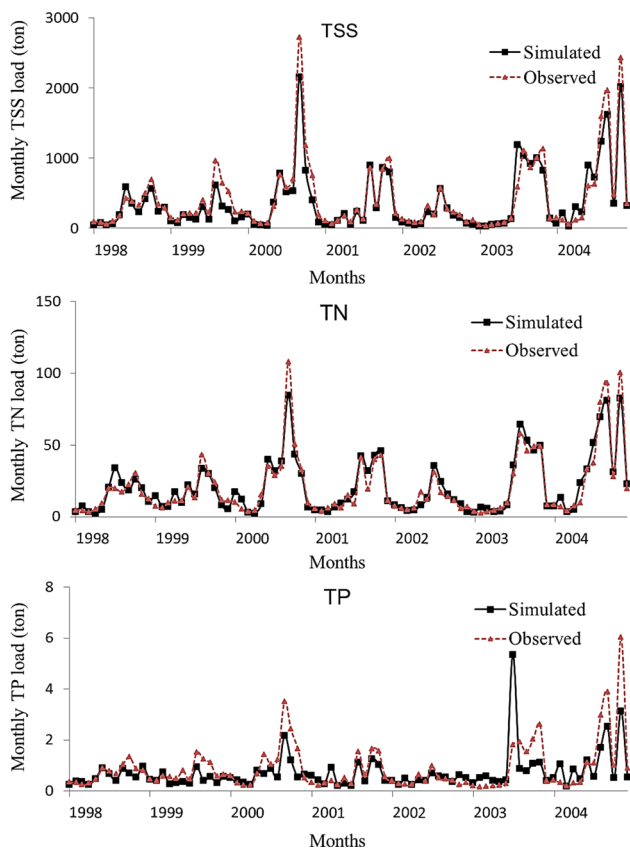**Fig. 5** Warrandyte (Site-3) monthly validation of TSS, TN and TP (Ahsan et al. 2023)

Fig. 6 Yarra Grange (Site-2) monthly calibration of TSS, TN and TP

Fig. 7 Yarra Grange (Site-2) monthly validation of TSS, TN and TP

PBIAS < ±30% as shown in Table 3), as per Moriasi et al. (2007). The annual and monthly TN loads are good (NSE > 0.75, RSR < 0.50, R² > 0.80, and PBIAS < ±25%), as shown in Table 3. Furthermore, the validation outcomes of TP loads for monthly are satisfactory (NSE > 0.50, RSR < 0.70, R² > 0.95, and PBIAS ≤ ±70), but the annual outcomes are unsatisfactory (RSR > 0.70 and NSE < 0.50, as shown in Table 3). The model overpredicts the monthly peak loads, as shown in Fig. 7, as the PBIAS values are negative, as shown in Table 3. The TSS and TN overprediction are much less than the TP.

Woori Yallock (Site-1) In this region, the calibration outcomes for the monthly and annual sediment are good (TSS monthly: NSE > 0.50, R² > 0.75, RSR < 0.70, and PBIAS < 15%, and TSS annual: NSE > 0.75, R² > 0.90, RSR < 0.50, and PBIAS < 15%, as shown in Table 3), according to Moriasi et al. (2007). The calibration results of the TN monthly and annuals are also very good (NSE > 0.75, R² ≥ 0.80, RSR < 0.50, and PBIAS ≤ ±15%). However, the TP outcomes are unsatisfactory (NSE ≤ 0.50, R² < 0.60, RSR > 0.70), as shown in Table 3. The model underpredicts the monthly peak loads for TP and TSS loads (Fig. 8) as the
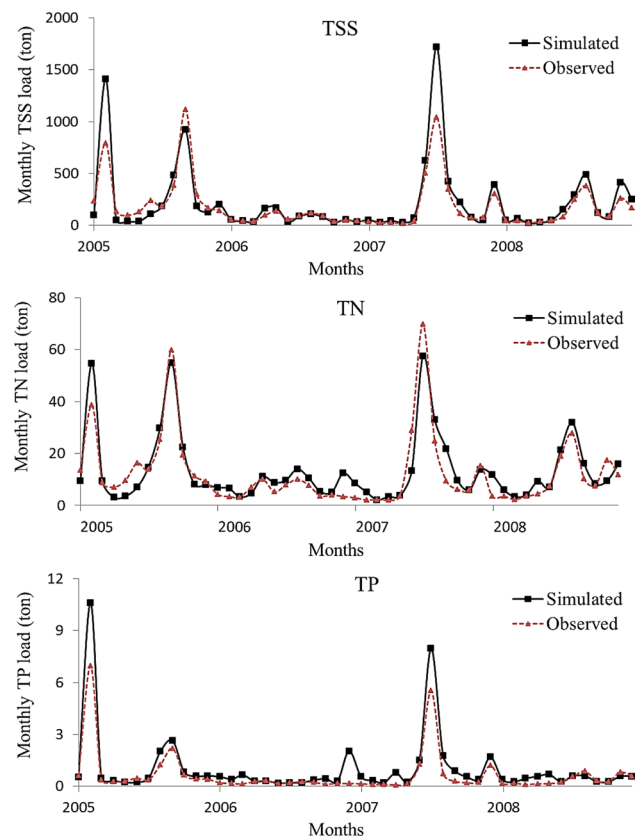
PBIAS values are positive, illustrating underprediction, but overpredicts the TN loads as the PBIAS values are negative.

In the validation operation, the model outcomes were satisfactory for monthly and annual TSS loads (monthly: NSE > 0.50, R² > 0.95, RSR < 0.70, and PBIAS < ±55%, and annual: NSE > 0.75, R² > 0.95, RSR < 0.50, and PBIAS < ±30%, as shown in Table 3), as per Moriasi et al. (2007). The monthly TN loads are satisfactory (NSE > 0.50, R² > 0.65, RSR < 0.70, and PBIAS < ±70%), but the annual loads are unsatisfactory (RSR > 0. 70 and NSE < 0.50), as shown in Table 3. Furthermore, the validation outcomes of TP loads for annual and monthly are both unsatisfactory (NSE < 0.50, RSR > 0.70, and PBIAS ≥ ±70, as shown in Table 3). The model overpredicts the monthly peak loads, as shown in Fig. 9, as the PBIAS values are negative, as shown in Table 3. The TSS and TN overpredictions are much less than the TP.

Calibration and validation outcomes

The simulation results of the MYDA SWAT model are consistent with those of other SWAT studies. The SWAT model overpredicts the load of nutrients in the validation period because of the minor variation in the outcome of manure
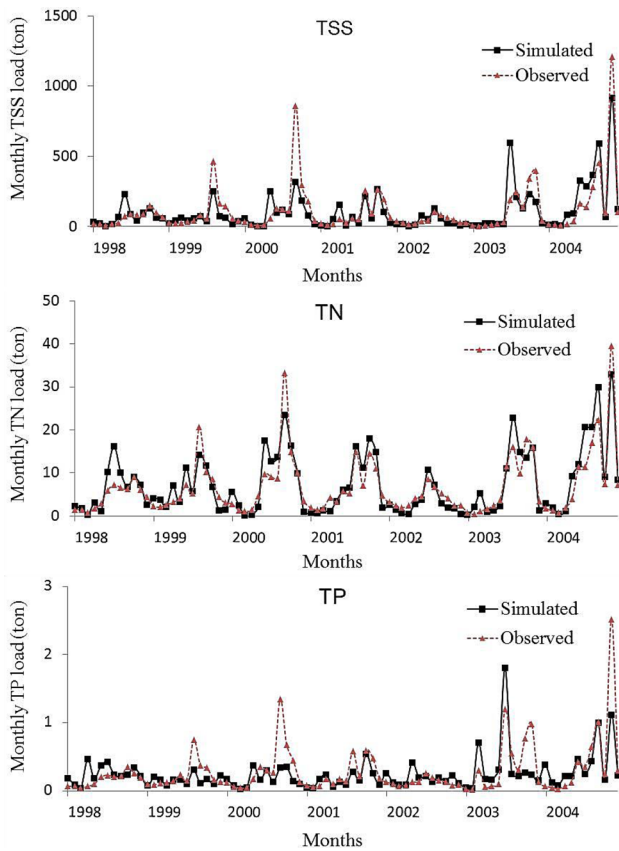
Fig. 8 Woori Yallock (Site-1) monthly calibration of TP, TN and TSS

Fig. 9 Woori Yallock (Site-1) monthly validation of TP, TN and TSS

or fertilizer that occurred during the calibration period, according to Green and Van Griensven (2008). Furthermore, they also showed that overprediction might occur because of precipitation events that occurred soon after manure or fertilizer was used. Likewise, Neitsch et al. (2002) found that for correct modelling, a precise report on the quantity and date of pesticide, manure or fertilizer handling is vital, and it is not often present. Moreover, load overprediction may occur with randomly defined usage dates, which overlap with a rainy day. However, farmers do not need to use manure or fertilizer on rainy days. As a result, Holvoet et al. (2005) recommended tackling this dilemma with a correct calibration using inverse modelling approaches and which case is used in this study (adjusting the usage and dates of fertilizers and manure during the calibration process). Furthermore, in the Nil catchment in Belgium, Holvoet et al. (2005) performed a sensitivity examination with SWAT on pesticides and hydrology and found that the errors that may occur in precipitation or usage rates are much less significant than the date of pesticide usage.

In general, the SWAT model performed poorly (in terms of NSE values) in calibration at site-1 which includes the subdrainage areas of the Woori Yallock Creek as shown in Tables 2 and 3. The Woori Yallock Creek is a contributing
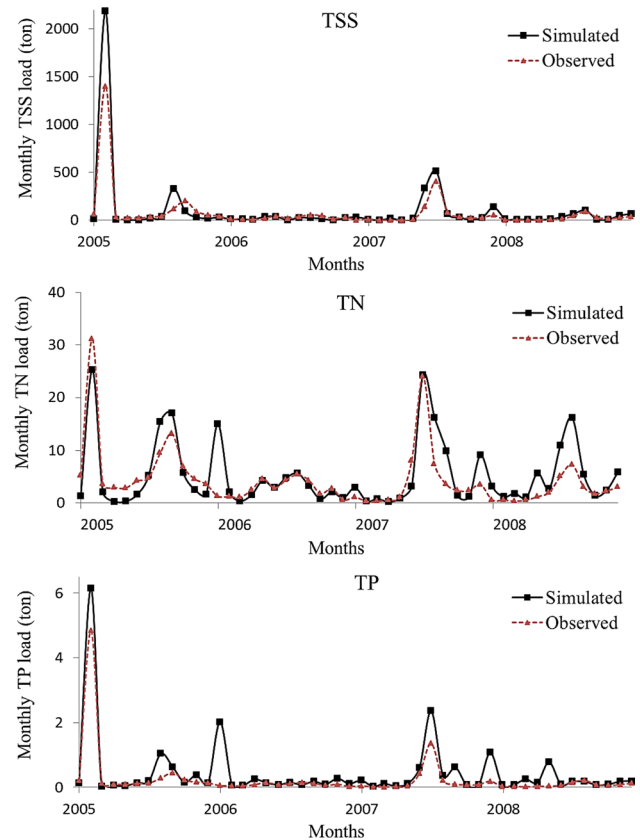
branch to the Yarra River, where the flow rate is very low and ceases to flow during dry periods. In addition, fewer water quality grab sample data were available for this site (site-1). Moreover, the input data for developing the model, such as crop and land management data, were coarse (i.e., sparse) at site-1 compared to the other two sites (site-2 and 3). This is also reflected in the uncertainty analysis with low p-factor values as shown in Tables 4 and 5. The streamflow statistics at the data sites are shown in Table 8 (Appendix) and the constituent concentration statistics at the data sites are shown in Table 9 (Appendix).

Uncertainty analysis of nutrients, sediment, and streamflow

The outcome of the uncertainty analysis of nutrients, sediment, and streamflow is shown in Table 4. The outcome shows that the model's prediction is fairly reliable in the sense that the uncertainty limit is narrow, which means the d-factor values are very little. However, the p-factor values are very small too. For the hydrology of four tributaries in the Lake Tana drainage area in Ethiopia, Setegn et al. (2008) discovered similar results. Their results show that the d-factor varies from 0.02 to 0.01 and the p-factor varies

Table 4 MYDA's water quality uncertainty results

	Streamflow			TSS			TN			TP		
	Site-1	Site-2	Site-3	Site-1	Site-2	Site-3	Site-1	Site-2	Site-3	Site-1	Site-2	Site-3
p-factor (%)	14	31	19	8	13	5	17	25	17	1	13	10
d-factor	0.16	0.12	0.10	0.12	0.09	0.02	0.19	0.14	0.09	0.04	0.15	0.12

Table 5 MYDA's water quality uncertainty results (considering uncertainty in observed data)

	Streamflow			TSS			TN			TP		
	Site-1	Site-2	Site-3	Site-1	Site-2	Site-3	Site-1	Site-2	Site-3	Site-1	Site-2	Site-3
p-factor (%)	27	56	53	25	37	13	31	50	39	15	29	25
d-factor	0.16	0.12	0.10	0.12	0.09	0.02	0.19	0.14	0.09	0.04	0.15	0.12

from 15 to 21%. The TSS and TP predictions show more uncertainties from the uncertainty outcomes.

There may be two causes; the p-factor values in the MYDA's water quality are very small. Firstly, the model uses a very low tolerance value of the objective function for determining 'good' and 'not good' simulations. For this, the uncertainty limits are narrow and bracket small numbers for the observed data. The uncertainty analysis done by Yang et al. (2008) in the Chaohe Basin in China by the same model shows a p-factor value of 18% and a d-factor value of 0.08 in the calibration duration.

Abbaspour et al. (2004) did the uncertainty analysis by the SUFI-2 method, where they considered $\pm 10\%$ uncertainty in the calibration of observed streamflow data. If it is considered in the similar way, i.e., $\pm 10\%$ uncertainty in the streamflow, nutrients, and sediment observed data in the model, then the p-factor values were increased significantly as shown in Table 5 compared to the p-factor values in Table 4.

Conclusions

The development of a reliable drainage area water quality model and its validation on the basis of the real-world drainage area is a great challenge. Such a model could save money and time. Therefore, the MYDA water quality management model is developed as per the guidelines stated. The SWAT-inbuilt LH-OAT and Parasol (SCE-UA) are used in this study for assessing multisite, multifactors, and multi-objective uncertainty, and auto-calibration analysis. The calibration and validation durations of streamflow, sediment and nutrients are different. Firstly, streamflow calibration is done at 3 sites followed by sediment, and nutrients are done. The sensitivity analysis results show that hydrologic factors dominated the highest factors globally for streamflow, sediment, and nutrients. Furthermore,

the identification of streamflow parameters in the SWAT model can be potentially made by the water quality factors (TN and TP), where a single factor is correlated to multiple factors.

Streamflow calibration outcomes illustrate good agreement between simulated and observed flows (runoff, base flow, and total streamflow). The calibration and validation outcomes of TN and TSS are acceptable with minor exclusions; however, TP outcomes are unsatisfactory. The MYDA water quality management system overestimates flow in dry years and underestimates flow in wet years. Moreover, the SWAT model underpredicts TN, TSS, and TP monthly peak loads in calibration and overpredicts in validation. It is noticed that the streamflow estimation produces a higher percentage of runoff from MYDA water quality management in monsoon periods. The outcome of the uncertainty analysis for the streamflow, nutrients, and sediment illustrates that the model's estimation is reliable as the limit of uncertainty is narrow with very small values of d-factor and p-factor. The TSS and TP estimations illustrate more uncertainties. The SWAT model significantly aids in the management of the MYDA's water quality. The outcomes of the calibration and validation of the drainage area water quality show great performance not only on the catchment outlet but also through the entire MYDA, which simplifies the complex task and labor in the calibration method. This study shows the implication of SWAT by assessing calibration, validation, and uncertainty outcomes in the MYDA water quality management. The simulated results of SWAT in MYDA are within acceptable limits with minor exclusions, which imply it can efficiently predict the sediment, streamflow, and nutrient loads.

Appendix

See Figs. 10, 11 and 12.

Fig. 10 Flowchart for calibration process of flow, sediment, and nutrients in the drainage area models (from Arnold et al. 2012; modified from Santhi et al. 2001)
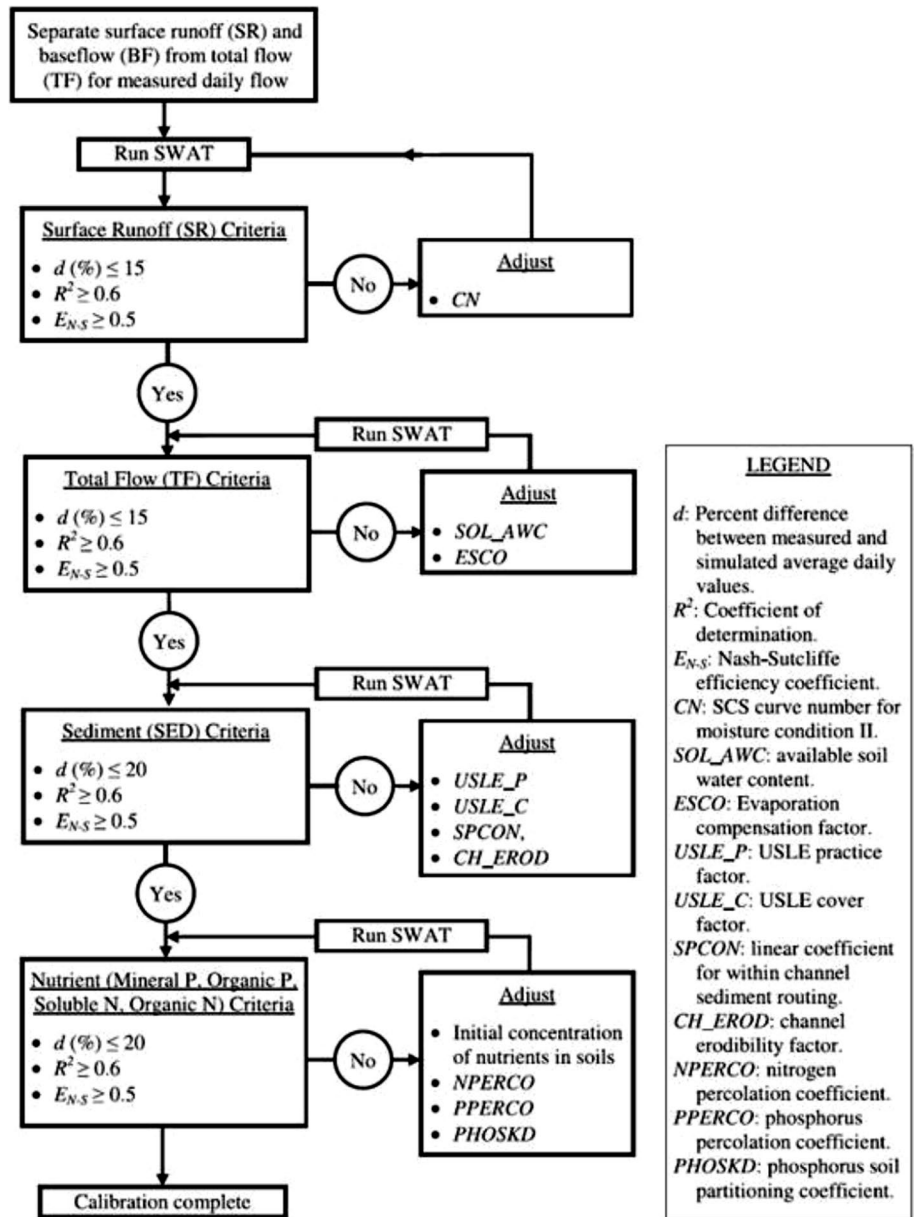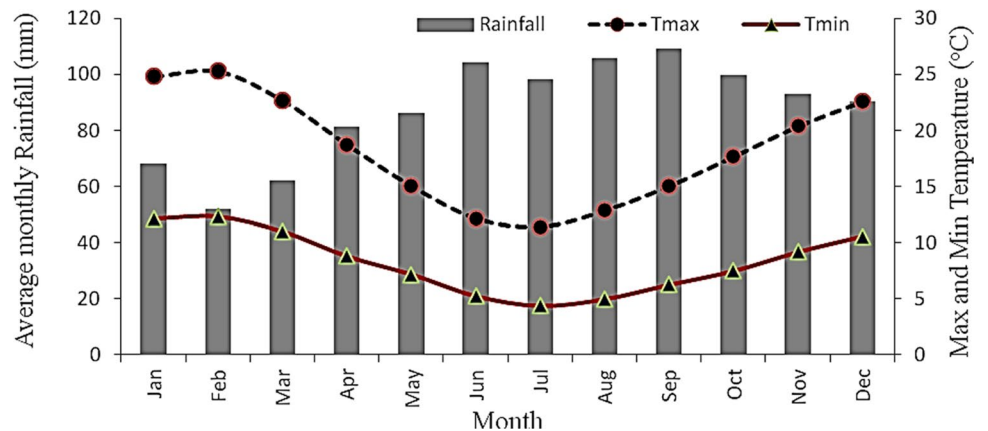

Fig. 11 Variation of average monthly rainfall and temperature (min and max) in the MYDA (Ahsan et al. 2023)
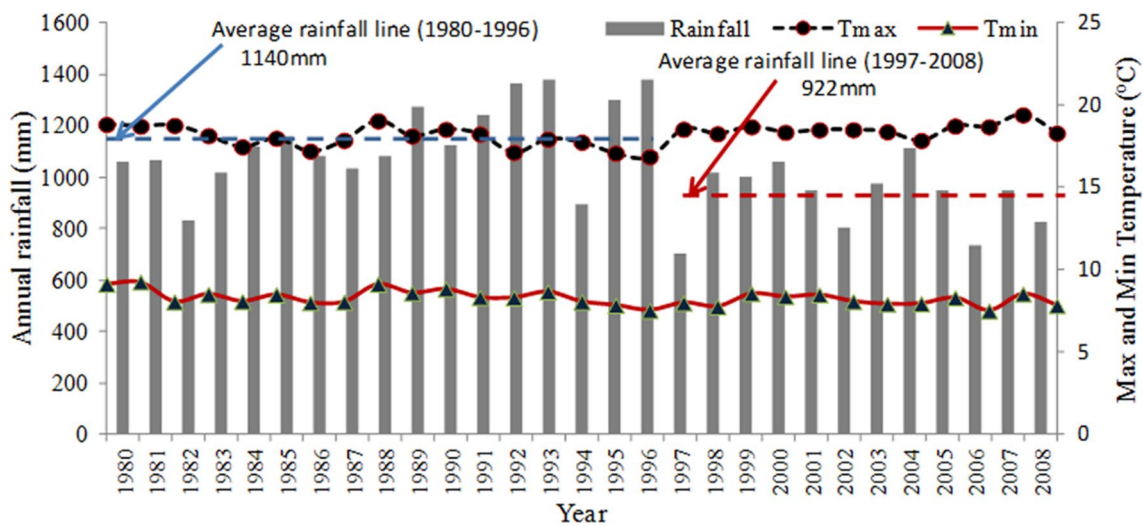
Fig. 12 Variation of annual rainfall and temperature (min and max) in the MYDA (Ahsan et al. 2023)

Table 6 Streamflow Observing Stations in the MYDA

Sites	Location	Latitude (^{o}S)	Longitude (^{o}E)
Site-1	Woori Yallock Creek at Woori Yallock	− 37.765	145.512
Site-2	Yarra River at Yarra Grange	− 37.667	145.476
Site-3	Yarra River at Warrandyte	− 37.740	145.212

Table 7 Water Quality Observing Locations in the MYDA

Sites	Location	Latitude (^{o}S)	Longitude (^{o}E)
Site-1	at Warburton Highway, Woori Yallock	− 37.777	145.508
Site-2	at Maroondah Highway, Healesville	− 37.678	145.491
Site-3	at Kangaroo Ground-Warrandyte Road, Warrandyte	− 37.738	145.219

Table 8 Streamflow statistics at the data sites

Data site	Record period of data available	Mean daily streamflow (m^3/s)									Record period
		Total period (1990–2008)			Calibration period (1990–2002)			Validation period (2003–2008)			
		Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Mean
Site-1	1975–2008	0.09	2.37	82.3	0.31	2.79	63.7	0.09	1.47	82.3	2.53
Site-2	1980–2008	0.96	10.20	194.1	1.78	11.89	194.1	0.96	6.53	103.7	10.08
Site-3	1970–2008	1.55	13.77	209.7	2.16	16.21	209.7	1.55	8.46	102.6	13.70

Table 9 Constituent concentrations statistics at the data sites

| Data site | Constituent | Constituent concentrations (mg/l) | | | | | | | | | | | |
| | | Total period (1998–2008) | | | | Calibration period (1998–2004) | | | | Validation period (2005–2008) | | | |
		Min	Mean	Max	Obs	Min	Mean	Max	Obs	Min	Mean	Max	Obs
Site-1	TN	0.58	1.19	2.71	129	0.72	1.25	2.71	82	0.58	1.09	1.51	47
	TP	0.01	0.04	0.15		0.01	0.04	0.15		0.01	0.03	0.08	
	TSS	1	12	75		2	14	75		1	8	32	
Site-2	TN	0.33	0.79	2.29	131	0.37	0.83	2.29	83	0.33	0.72	1.32	48
	TP	0.01	0.03	0.12		0.01	0.03	0.12		0.01	0.03	0.07	
	TSS	1	11	54		4	13	54		1	8	48	
Site-3	TN	0.42	1.13	2.93	133	0.42	1.17	2.52	85	0.53	1.07	2.93	48
	TP	0.01	0.06	0.35		0.03	0.06	0.16		0.01	0.05	0.35	
	TSS	1	15	310		1	16	120		1	14	310	

Obs Observations (number of available grab samples)

Data Availability The necessary data set is given in Appendix. The additional data that support the findings of this study are available on request from the corresponding authors.

Declarations

Conflict of interests The authors declare that they have no conflict of interest.

Ethical approval and consent to participate Not applicable.

Consent for publication Not applicable.

References

Abbaspour KC, Johnson CA, Van Genuchten MT (2004) Estimating uncertain flow and transport parameters using a sequential uncertainty fitting procedure. Vadose Zone J 3(4):1340–1352

Ahsan A, Das SK, Khan MHRB, Ng AW, Al-Ansari N, Ahmed S, Shafiquzzaman M (2023) Modeling the impacts of best management practices (BMPs) on pollution reduction in the Yarra River catchment, Australia. Appl Water Sci 13(4):98

Aibaidula D, Ates N, Dadaser-Celik F (2023) Uncertainty analysis for streamflow modeling using multiple optimization algorithms at a data-scarce semi-arid region: Altinapa reservoir watershed, Turkey. Stoch Env Res Risk Assess 37(5):1997–2011

Arnold JG, Allen PM, Muttiah R, Bernhardt G (1995) Automated base flow separation and recession analysis techniques. Ground Water 33(6):1010–1018

Arnold JG, Moriasi DN, Gassman PW, Abbaspour KC, White MJ, Srinivasan R, Santhi C, Harmel RD, Van Griensven A, Van Liew MW (2012) Swat: model use, calibration, and validation. Trans ASABE 55(4):1491–1508

Arnold JG, Srinivasan R, Muttiah RS, Williams JR (1998) Large area hydrologic modeling and assessment part I: model development. J Am Water Resour Assoc 34(1):73–89

Borah DK, Bera M (2003) Watershed-scale hydrologic and nonpoint-source pollution models: review of mathematical bases. Trans ASAE 46(6):1553–1566

Cambell KL, Edwards DR (2001) Phosphorus and water quality impacts. In: Ritter WF, Shirmohammadi A (eds) Agricultural nonpoint source pollution: watershed management and hydrology. Lewis Publishers, Boca Raton, Florida, pp 91–109

Croke BFW, Jakeman AJ (2001) Predictions in catchment hydrology: an Australian perspective. Mar Freshw Res 52(1):65–79

Das SK, Ahsan A, Khan MHRB, Tariq MAUR, Muttil N, Ng AWM (2022) Impacts of climate alteration on the Hydrology of the Yarra River Catchment, Australia using GCMs and SWAT model. Water 14:445

EPA Victoria (1999), Protecting the environmental health of yarra catchment waterways policy impact assessment, Report No. 654, Melbourne, Australia

Gasirabo A, Xi C, Kurban A, Liu T, Baligira HR, Umuhoza J, DufatanyeEdovia U (2023) SWAT model calibration for hydrological modeling using concurrent methods, a case of the Nile Nyabarongo River basin in Rwanda. Front Water 5:1268593

Gassman P, Jha M, Wolter C, Schilling K (2015) Evaluation of alternative cropping and nutrient management systems with soil and water assessment tool for the raccoon river watershed master plan. Am J Environ Sci 11(4):227

Grayson RB, Argent R, Western A (1999b) Scoping study for the implementation of water quality management frameworks: Final Report, CEAH report 2/99, University of Melbourne, Victoria, Australia

Green CH, van Griensven A (2008) Autocalibration in hydrologic modeling: using Swat 2005 in small-scale watersheds. Environ Model Softw 23(4):422–434

Guo Y, Markus M, Demmissie M (2002) Uncertainty of Nitrate-N load computations for agricultural watersheds. Water Resour Res 38(10):1–12

Gupta HV, Sorooshian S, Yapo PO (1998) Toward improved calibration of hydrologic models: multiple and noncommensurable measures of information. Water Resour Res 34(4):751–763

Holvoet K, van Griensven A, Seuntjens P, Vanrolleghem PA (2005) Sensitivity analysis for hydrology and pesticide supply towards the river in Swat. Phys Chem Earth, Parts a/b/c 30(8):518–526

Isbell R (2002) The Australian soil classification, Revised. CSIRO, Melbourne, Australia

Kirsch K, Kirsch A, Arnold JG (2002) Predicting sediment and phosphorus loads in the rock river basin using swat. Trans ASAE 45(6):1757–1769

Kragt ME, Newham LTH (2009) Developing a water-quality model for the george catchment, tasmania, landscape logic technical report No. 16, Tasmania

Madsen H (2003) Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. Adv Water Resour 26(2):205–216

Melbourne Water (2010), Rural land program—water sensitive farm design, Melbourne Water, Victoria

Moriasi DN, Arnold JG, Van Liew MW, Bingner RL, Harmel RD, Veith TL (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans ASABE 50(3):885–900

Nasirzadehdizaji R, Akyuz DE (2022) Predicting the potential impact of forest fires on runoff and sediment loads using a distributed hydrological modeling approach. Ecol Model 468:109959

Neitsch SL, Arnold JG, Kiniry JR, Williams JR (2005) Soil and water assessment tool theoretical documentation version 2005, Blackland Research Center, Temple, TX, USA

Neitsch SL, Arnold JG, Kiniry JR, Srinivasan R, Williams JR (2004) Soil and water assessment tool input/output file documentation version 2005, Blackland Research Center, Temple, TX, USA

Neitsch SL, Arnold JG, Srinivasan R, Grassland S (2002) Pesticides fate and transport predicted by the soil and water assessment tool (Swat): Atrazine, Metolachlor and Trifluralin in the Sugar Creek Watershed. Brc Publication # 2002–03, pp 96

Niraula R, Norman LM, Meixner T, Callegary JB (2012) Multi-gauge calibration for modeling the semi-arid santa cruz watershed in Arizona-Mexico border area using swat. Air, Soil Water Res 5:41–57

Northcote KH (1979) A Factual Key for the Recognition of Australian Soils, 4th edn. Rellim Technical Publishers, Glenside, SA

Oduor BO, Campo-Bescós MÁ, Lana-Renault N, Kyllmar K, Mårtensson K, Casalí J (2023) Quantification of agricultural best management practices impacts on sediment and phosphorous export in a small catchment in southeastern Sweden. Agric Water Manag 290:108595

Rafik A, Brahim YA, Amazirh A, Ouarani M, Bargam B, Ouatiki H, Bouslihim Y, Bouchaou L, Chehbouni A (2023) Groundwater level forecasting in a data-scarce region through remote sensing data downscaling, hydrological modeling, and machine learning: a case study from Morocco. J Hydrol Regional Stud 50:101569

Reckhow KH (1994) Water quality simulation modeling and uncertainty analysis for risk assessment and decision making. Ecol Model 72(1–2):1–20

Santhi C, Arnold JG, Williams JR, Dugas WA, Srinivasan R, Hauck LM (2001) Validation of the swat model on a large river basin with point and nonpoint sources. J Am Water Resourc Assoc 37(5):1169–1188

Setegn SG, Srinivasan R, Dargahi B (2008) Hydrological modelling in the Lake Tana Basin, Ethiopia using swat model. Open Hydrol J 2(1):49–62. https://doi.org/10.2174/1874378100802010049

Thorsen M, Refsgaard JC, Hansen S, Pebesma E, Jensen JB, Kleeschulte S (2001) Assessment of uncertainty in simulation of nitrate leaching to aquifers at catchment scale. J Hydrol 242(3):210–227

U.S. EPA (2002), Guidance for quality assurance project plans for modeling, EPA QA/G-5M. Report EPA/240/R-02/007, Washington, D.C.: U.S. EPA, Office of Environmental Information

USDA-ARS (1999), U.S. Department of agriculture-agricultural research service, soil and water assessment tool, Swat: Baseflow Filter Program, viewed 15 May 2010, <http://swatmodel.tamu.edu/software/baseflow-filter-program>

van Griensven A, Francos A, Bauwens W (2002) Sensitivity analysis and auto-calibration of an integral dynamic model for river water quality. Water Sci Technol 45(9):325–332

van Griensven A, Meixner T, Grunwald S, Bishop T, Diluzio M, Srinivasan R (2006) A global sensitivity analysis tool for the parameters of multi-variable catchment models. J Hydrol 324(1–4):10–23. https://doi.org/10.1016/j.jhydrol.2005.09.008

Van Liew MW, Veith TL (2010) Guidelines for using the sensitivity analysis and auto-calibration tools for multi-gage or multi-step calibration in Swat

Vervoort RW (2007) Uncertainties in calibrating swat for a semi-arid catchment in Nsw (Australia). In: Proceedings of the IVth SWAT conference, Delft (The Netherlands), pp 2–5

Vrugt JA, Gupta HV, Bouten W, Sorooshian S (2003) A shuffled complex evolution metropolis algorithm for estimating posterior distribution of watershed model parameters. In: Sorooshian S, Gupta HV, Rousseau AN, Turcotte R (eds) Q Duan. Calibration of Watershed Models, AGU Washington

Watson BM, Selvalingam S, Ghafouri M (2003) Evaluation of swat for modelling the water balance of the Woady Yaloak River Catchment, Victoria. In: Post and David (Eds.), MODSIM 2003 : International Congress on Modelling and Simulation, Jupiters Hotel and Casino, 14–17 July 2003 : integrative modelling of biophysical, social and economic systems for resource management solutions : proceedings, Modelling and Simulation Society of Australia and New Zealand Inc, Canberra. ACT, pp 1–6

White KL, Chaubey I (2005) Sensitivity analysis, calibration, and validations for a multisite and multivariable swat model. J Am Water Resour Assoc 41(5):1077–1089

Winchell M, Srinivasan R, Di Luzio M, Arnold JG (2009) *Arcswat 2.3.4 Interface for Swat2005, User's Guide*, Grassland, Soil and Water Research Laboratory, Temple, TX (USA)

Xue J, Wang Q, Zhang M (2022) A review of non-point source water pollution modeling for the urban–rural transitional areas of China: Research status and prospect. Sci Total Environ 826:154146

Yang J, Reichert P, Abbaspour KC, Xia J, Yang H (2008) Comparing uncertainty analysis techniques for a swat application to the Chaohe basin in China. J Hydrol 358(1):1–23