Check for updates

Water Science & Technology

IWA PUBLISHING

© 2023 The Authors

Water Science & Technology Vol 88 No 7, 1621 doi: 10.2166/wst.2023.297

Examining the open-source datasets for water quantity and quality using the soil and water assessment tool (SWAT)

İsmail Bilal Peker and Sezar Gülbaz*

Department of Civil Engineering, Istanbul University-Cerrahpaşa, 34320 Avcılar, Istanbul, Türkiye
*Corresponding author. E-mail: sezarg@iuc.edu.tr

İBP, 0000-0001-9133-6797; SG, 0000-0002-2274-6896

ABSTRACT

Water quality modeling is very important for the management of water resources. In this study, the upper part of the Porsuk Basin in Türkiye is analyzed using SWAT. In the analysis, in addition to the data provided by the General Directorate of State Hydraulic Works (DSI), the freely available flow and water quality data from the GEMStat data portal were used. This study presents a discussion of the practicality of GEMStat data for a water quality model. For this purpose, firstly, the SWAT model was constructed with freely available global data sources on elevation, land use/land cover, and soil type. Then, the model flow outputs were calibrated and validated for both DSI and GEMStat data in three different time periods. As a result, the flow calibration and validation success in the daily time step is 0.64 and 0.44 according to the Nash–Sutcliffe efficiency (NSE). The model was also validated using GEMStat flow data and calibrated using GEMStat water quality data such as nitrate (NO_3), total suspended solids (TSS), and dissolved oxygen (DO) with a reasonable value. Hence, the results showed that GEMStat flow and water quality data can be used as auxiliary open-source data in the modeling process.

Key words: GEMStat data portal, SWAT, Upper Porsuk Basin, water quality model

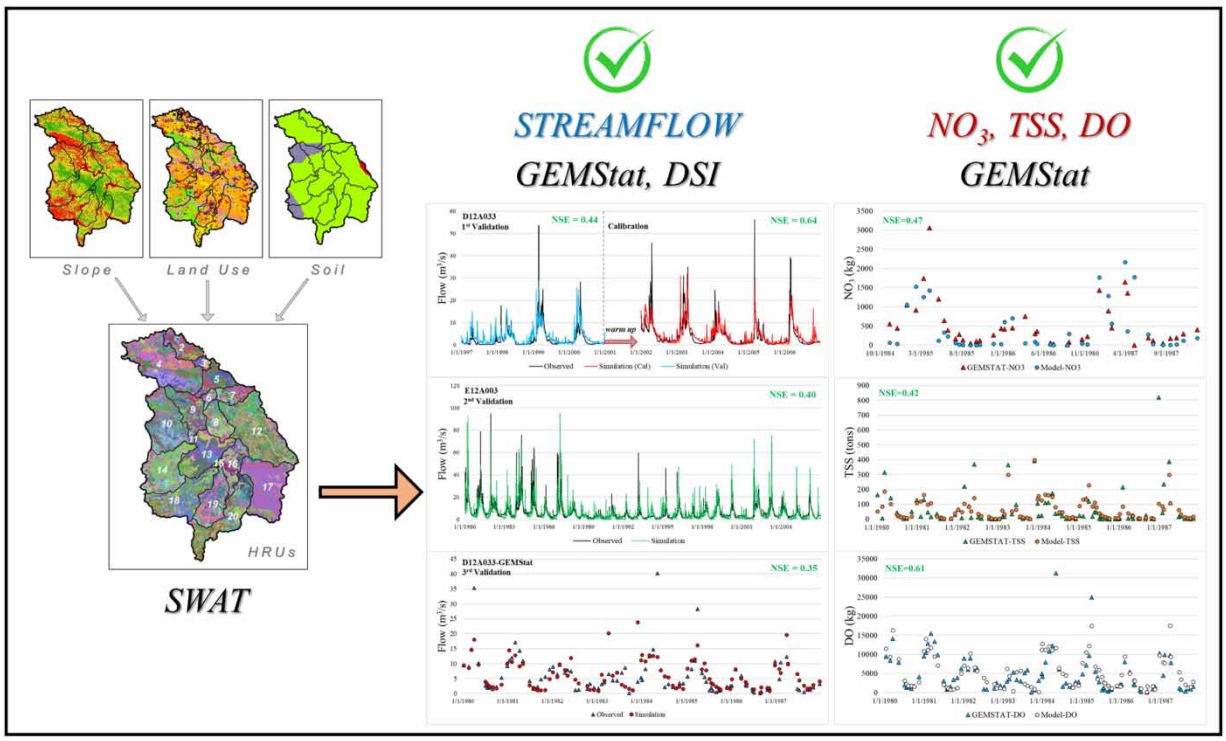
HIGHLIGHTS

- The importance of data monitoring and the benefit of the freely available data were emphasized.
- The usability of the data obtained from the global data provider portal GEMStat was tested using SWAT.
- Model-generated flow and water quality data showed reasonable fit (NSE values between 0.35 and 0.64) with the real data in the daily time step.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (http://creativecommons.org/licenses/by/4.0/).

Downloaded from http://iwaponline.com/wst/article-pdf/88/7/1621/1307524/wst088071621.pdf
by guest

GRAPHICAL ABSTRACT



1. INTRODUCTION

The protection of aquatic ecosystems against pollution is an important global priority (United Nations General Assembly 2015). The studies on this issue aimed at conserving clean water resources and increasing the water quality, which depends on the measurement of the parameters that cause pollution. It is obvious that freshwater resources are becoming increasingly polluted due to agriculture and urban activities (Carpenter *et al.* 1998; Bhateria & Jain 2016). In recent years, measuring and monitoring water quality parameters in rivers, lakes, wetlands, or coastal waters has gained importance (USEPA 2000; Kirschke *et al.* 2020). Monitoring water quantity and quality is necessary for informed decision-making and minimizing uncertainties regarding water management (WMO 2009; Stewart 2015) even if these procedures are laborious and costly (Barcelona *et al.* 1985; Chapman 1996; Zessner 2021). These measurements offer multiple benefits such as flood forecasts and warnings, protecting lives and property, flood plain mapping, determining environmental or ecological flows, and regulating pollutant discharges (USGS 2006). Conceptually, such measurements are also essential for watershed modeling studies performed for the management and regulation of water quality. Moreover, calibration and validation of the watershed models require long-term and accurate observations of water quantity and quality parameters, which are always limited (Sivapalan 2003).

Although watershed models are robust tools for predicting water quality in a region, modeling procedures are challenging due to the data scarcity issue, especially in low- and middle-income countries (Wagner *et al.* 2009). Therefore, a simple but scientifically proven hydrological model that does not require too much data would be a useful tool for water quantity and quality estimations (Kwakye & Bárdossy 2020). Choosing an appropriate model and obtaining the necessary data is critical to construct a reliable model. Knowledge about the water quantity and quality of the basin can be obtained from modeling studies. There are numerous modeling tools for examining water quality parameters (Tsakiris & Alexakis 2012; Wang *et al.* 2013; Gao & Li 2014; Gülbaz 2019; Ejigu 2021; Bai *et al.* 2022). HSPF (Hydrological Simulation Program-FORTRAN) (Barnwell & Johanson 1981), WASP (Water Quality Analysis Simulation Program) (Di Toro *et al.* 1983), QUAL2E (Brown & Barnwell 1987), SWAT (Soil and Water Assessment Tool) (Arnold *et al.* 1998), and AQUATOX (Park *et al.* 2008) are some commonly used water quality models. Among them, SWAT is an effective tool for both water quantity and quality assessment,

Downloaded from http://iwaponline.com/wst/article-pdf/88/7/1621/1307524/wst088071621.pdf
by guest

especially in agricultural watersheds. Costa *et al.* (2021) showed that SWAT is the most frequently used water quality model in the world.

Besides, obtaining the data required for modeling is as important as the selection of an appropriate modeling tool. Although many physical-based (such as elevation, land use/land cover, and soil type) and meteorological (such as precipitation and temperature) global data sources are available, obtaining previously measured water quality data is relatively difficult. Calibration or output testing in water quality models is also quite challenging. Studies in this field need great effort and data archive. Required data can be measured by researchers, obtained from local organizations, or both. Also, it is reported that open and free data sources were for hydrology studies (Newcomer *et al.* 2022). Researchers prefer using different data sources if they did not conduct field measurements in the region. There are a few continental and global datasets such as CESI (Canadian Environmental Sustainability Indicators program), GEMStat (Global Freshwater Quality Database), GLORICH (GLObal RIver CHemistry), and Waterbase and WQP (Water Quality Portal) (Virro *et al.* 2021). Only GEMStat provides datasets about the water quality of some rivers in Türkiye. Among these rivers, water quality measurement data for Porsuk Stream is available in the GEMStat data portal. Furthermore, some researchers collected their own field measurements in the Porsuk Basin (Orak 2006; Yüce *et al.* 2006; Solak 2009; Gürel 2011; Çelen *et al.* 2014; Köse *et al.* 2018; Şahin 2018) and some gathered data from local organizations (Yerel 2010; Güngör 2011). In the present study, the GEMStat data portal was chosen as the exclusive global source for water quality measurement data pertaining to Turkish rivers among all available options. The primary reason for selecting GEMStat data was its unique coverage of Turkish rivers. GEMStat is highly favored due to its user-friendly interface and convenient access to data through its visual database, which includes mapped representations. Furthermore, it is important to highlight that GEMStat offers this valuable data at no cost. Both water quantity (flow) and water quality (nitrate (NO_3), total suspended solids (TSS), and dissolved oxygen (DO)) data were used. The aim of this study is to investigate the availability of global water quality data, rather than to present a comprehensive water quality model. Accordingly, flow and water quality data obtained from the model were calibrated and validated in three different time periods. Therefore, the usability of GEMStat data in such a study is also examined. Our results are believed to provide more insight into the usability of GEMStat data in future studies.

2. MATERIALS AND METHODS

2.1. Study area and data

The upper part of the Porsuk Basin, which is an important branch of the Sakarya River, was selected as the study area. Porsuk Stream feeds the Sakarya River, which is one of the 25 main basins in Türkiye. It is an important water source in Türkiye, especially the upper part of the Porsuk Basin (a part up to Eskişehir province) is used as a drinking and irrigation water basin (Köse *et al.* 2016). Therefore, examining the water quality in this region is of great importance for the health of ecosystems.

The basin lies between 29°35′–30°30′ east longitudes and 38°43′–39°38′ north latitudes. The location of the basin is shown in Figure 1. The Upper Porsuk Basin covers an area of about 3,947 km^2. The mean elevation of the basin is 1,160 m, ranging between 904 and 2,184 m. Agricultural lands are dominant in the basin (about 47%). Besides, bare land (about 14%) and forests (13%) are the other most common land use classes. The soil types in the basin are mostly (about 90%) Bk45-2bc coded, C group hydrological soil class.

The spatial data consists of a digital elevation model (DEM), land use/land cover, and soil type maps. SRTM Shuttle Radar Topography Mission (SRTM) with 30 m cell size was utilized for watershed delineation and stream network process. The European Environment Agency – Coordination of Information on the Environment (EEA-CORINE) with 100 m cell size and the Digital Soil Map of the World (FAO-DSMW) scale 1:5,000,000 by the Food and Agriculture Organization of the United Nations Educational, Scientific and Cultural Organization (FAO-UNESCO) data were employed for generating land use/land cover and soil type, respectively. In this study, the preference for CORINE and FAO-DSMW data is widespread on a global scale, particularly among users of the SWAT model (ShangGuan *et al.* 2014; Abbaspour *et al.* 2019; Busico *et al.* 2020; López-Ballesteros *et al.* 2023). Despite the coarse resolution and relatively outdated nature of the FAO-DSMW data, one of the reasons it is favored in this study is its capability to directly align with the SWAT model codes. The availability of this matching system contributes significantly to the preference for this data in the context of this study. The meteorological data were gathered from global weather data derived from Climate Forecast System Reanalysis (CFSR) (Fuka *et al.* 2014). The available date range for this data is 1979–2013. Also, there are two flow stations on the upper part of the Porsuk River: The

Downloaded from http://iwaponline.com/wst/article-pdf/88/7/1621/1307524/wst088071621.pdf
by guest

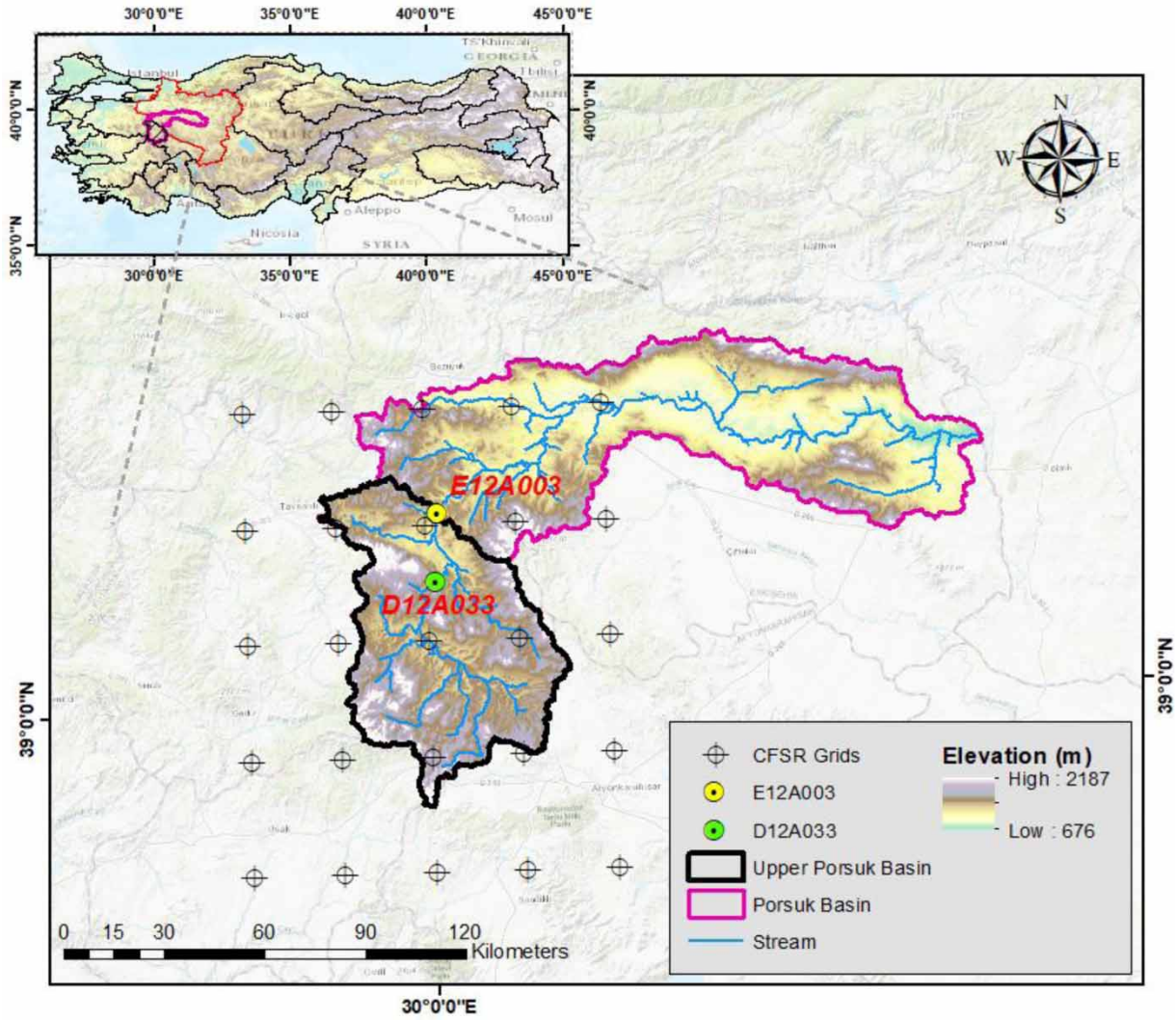


Figure 1 | Location of the stations in the study area.

first one (Station Code: E12A003) is located at the outlet of the whole basin and has continuous discharge data over the 1980–2006 period. The second station (Station Code: D12A033) is located in the middle of the basin. The continuous discharge data of this station can be obtained between the years 1997 and 2006. The second station has also non-continuous data for both instantaneous flows, NO_3, TSS, and DO. These parameters were gathered from the United Nations Environment Programme Global Environment Monitoring System (UNEP GEMS, UNEP-GEMS/Water Programme 2006). The NO_3 data covered the years 1984–1987, while discharge, TSS, and DO data covered 1980–1987. Therefore, the period 1979–2006 was selected as the modeling period. Table 1 shows the details of the data.

2.2. Model setup

The SWAT (Soil and Water Assessment) model is a deterministic, semi-distributed, process-based watershed model developed by the USDA (United States Department of Agriculture) (Arnold *et al.* 1998; Neitsch *et al.* 2011). Successful applications of the SWAT model have been demonstrated by different disciplines in various areas from the field scale to the basin scale, in different climatic zones with different geographical conditions (Gassman *et al.* 2007; Ahn & Kim 2019).

For the current study, the ArcSWAT extension was used in the model setup for the Upper Porsuk Basin. The model inputs can be easily incorporated into the model in the GIS environment. In the first step, the Upper Porsuk Basin was subdivided into 20 sub-basins considering the DEM. Then, 586 HRUs were automatically created by combining the spatial model inputs

Downloaded from http://iwaponline.com/wst/article-pdf/88/7/1621/1307524/wst088071621.pdf
by guest

Table 1 | Data description, source, and scale/resolution used in the modeling

Data type	Data source	Scale/Resolution
HRU definition data		
Digital elevation model (DEM)	SRTM	Grid cell 30 × 30 m
Land use/land cover	CORINE (year 1990)	Grid cell 100 × 100 m
Soil	FAO-UNESCO DSMW	Scale 1:5,000,000
Meteorological data		
Precipitation	CFSR	Grid cell ~38 km
Max./Min. temperature		
Relative humidity		
Solar radiation		
Wind speed		
Calibration/Validation data		
Flow	DSI and GEMStat	Ground station
Water quality (NO_3, TSS, and DO)	GEMStat	Ground station

(elevation, land use/land cover, and soil). Ten elevation bands were used. The meteorological data prepared in an appropriate format covering the whole modeling period (1979–2013) were entered into the model. Depending on the meteorological data, the Penman-Monteith method (Monteith 1965) was preferred for evapotranspiration calculation.

3. RESULTS

3.1. Flow calibration and validation

In the selected modeling period (1979–2006), the simulation flow outputs were calibrated and validated using the data from two stations (D12A033 and E12A003). Among these stations, the relatively more reliable one, the data obtained from DSI (D12A033), was used for the calibration. The period 2002–2006, which contains more up-to-date and continuous data compared to other years, was selected as the calibration period. The other continuous period for the same station, between 1997 and 2000, was used in the first validation period. In addition, the continuous data of the other station coded E12A003 were run with calibrated parameters, and the model was verified again. Finally, a third verification procedure was performed using the instantaneous flow data for the same station gathered from the GEMStat data portal (D12A033-GEMStat) in the period 1980–1987. In the simulations, a one-year warm-up period was used to stabilize the model using initial conditions. In Figure 2, the selected calibration and validation periods are demonstrated.

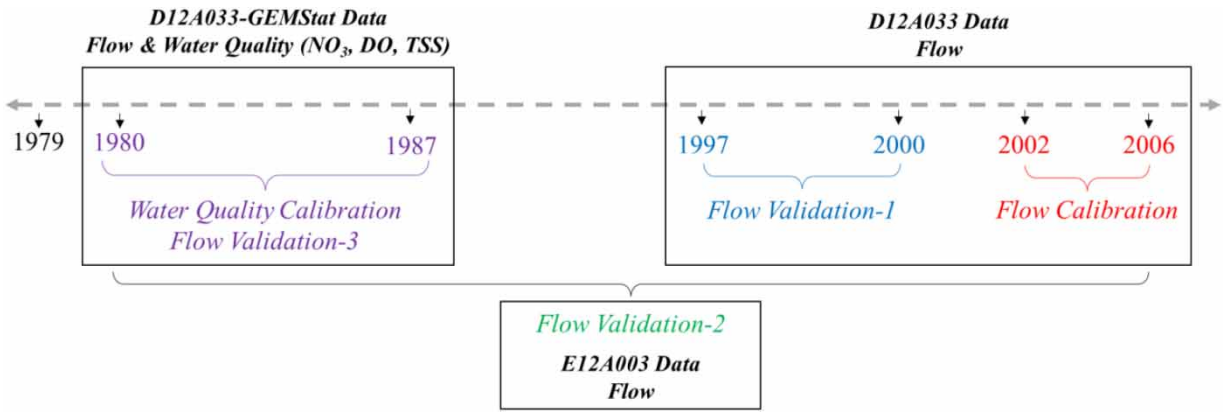


Figure 2 | Calibration and validation periods.

Downloaded from http://iwaponline.com/wst/article-pdf/88/7/1621/1307524/wst088071621.pdf
by guest

The SWAT-CUP (SWAT-Calibration and Uncertainty Program) automatic calibration program (Abbaspour 2012) was used to obtain optimized parameters. The SUFI-2 (Sequential Uncertainty and Fitting-version2) algorithm (McKay *et al.* 1979) in SWAT-CUP was employed for calibration. With the global sensitivity analysis performed in SWAT-CUP before the calibration, sensitive parameters were determined. The result of the sensitivity analysis points to parameter CN2 as the most sensitive parameter supported by a significant absolute *t*-stat. Consistent with previous research results using the SWAT model (White & Chaubey 2005; Van Griensven *et al.* 2006; Arnold *et al.* 2012a, 2012b; Abbaspour *et al.* 2015; Khalid *et al.* 2016), this was not surprising, emphasizing the expected importance of the CN2 parameter. As shown in Table 2, the parameters with the highest absolute *t*-stat and *p*-value < 0.05 are the most sensitive to flow outputs. In the initial analysis, calibration utilizing seven parameters that passed the sensitivity threshold (p-value < 0.05) yielded results falling short of the acceptable performance. Subsequently, the count of sensitive parameters was incremented to 20, accompanied by a revision of the performance criteria to attain more reasonable levels. The calibration was made at the daily time step and over four iterations with 500 simulations. The sensitive parameters, descriptions, change methods, and the final calibrated values used for the calibration procedure are presented in Table 3.

Table 2 | Global sensitivity analysis results

Number	Parameter name	*p*-value	*t*-stat	Absolute *t*-stat
1	CN2	0.0000	−39.9703	39.9703
2	SMFMN	0.0000	−5.1006	5.1006
3	SOL_Z	0.0000	4.8760	4.8760
4	ESCO	0.0003	−3.6491	3.6491
5	SOL_AWC	0.0060	2.7683	2.7683
6	SFTMP	0.0073	−2.7033	2.7033
7	GW_DELAY	0.0102	2.5870	2.5870
8	CH_K1	0.0548	1.9285	1.9285
9	SMFMX	0.0660	−1.8461	1.8461
10	TLAPS	0.0856	−1.7253	1.7253
11	CH_N1	0.1796	1.3454	1.3454
12	GWQMN	0.1917	−1.3089	1.3089
13	REVAPMN	0.1967	−1.2942	1.2942
14	SMTMP	0.2056	−1.2689	1.2689
15	RCHRG_DP	0.2087	−1.2601	1.2601
16	TIMP	0.6528	−1.2087	1.2087
17	ALPHA_BF	0.2572	1.1355	1.1355
18	SOL_K	0.3118	−1.0133	1.0133
19	SLSUBBSN	0.6863	−0.7521	0.7521
20	GW_REVAP	0.5602	−0.7352	0.7352
21	PLAPS	0.5046	−0.6681	0.6681
22	EPCO	0.4628	0.5833	0.5833
23	LAT_TTIME	0.2278	0.4504	0.4504
24	SOL_BD	0.4526	0.4044	0.4044
25	SURLAG	0.7608	−0.3048	0.3048
26	FFCB	0.7648	−0.2995	0.2995
27	OV_N	0.8171	−0.2316	0.2316
28	CANMX	0.8906	0.1377	0.1377
29	CH_N2	0.9318	0.0856	0.0856
30	CH_K2	0.9795	0.0257	0.0257

Downloaded from http://iwaponline.com/wst/article-pdf/88/7/1621/1307524/wst088071621.pdf
by guest

Table 3 | Calibrated values for sensitive parameters with descriptions, change methods, and initial ranges

Parameters	Descriptions	Initial ranges		Change method*	Calibrated value
TLAPS	Temperature lapse rate (°C/km)	−6	−3	v_	−5.60
SFTMP	Snowfall temperature (°C)	−3	3	v_	−1.01
SMTMP	Snowmelt base temperature (°C)	−3	3	v_	1.18
SMFMX	Melt factor for snow on June 21 (mm H_2O/°C-day)	3	6	v_	4.30
SMFMN	Melt factor for snow on December 21 (mm H_2O/°C-day)	0	3	v_	1.63
TIMP	Snowpack temperature lag factor	0	1	v_	0.36
ESCO	Soil evaporation compensation factor	0.7	1	v_	0.82
CN2	Initial SCS runoff curve number for moisture condition II	−0.3	0.3	r_	−0.30
ALPHA_BF	Baseflow alpha factor (1/days)	0	1	v_	0.58
GWQMN	Threshold depth of water in the shallow aquifer required for return flow to occur (mm H_2O)	100	3,000	v_	1,402.56
GW_DELAY	Groundwater delay time (days)	1	50	v_	1.02
GW_REVAP	Groundwater revap coefficient	0.02	0.2	v_	0.14
REVAPMN	Threshold depth of water in the shallow aquifer for revap or percolation to the deep aquifer to occur (mm H_2O)	100	300	v_	187.95
RCHRG_DP	Deep aquifer percolation fraction	0	0.5	v_	0.31
CH_K1	Effective hydraulic conductivity in tributary channel alluvium (mm/h)	−0.3	0.3	r_	−0.03
CH_N1	Manning's '*n*' value for the tributary channels	−0.3	0.3	r_	−0.19
SOL_Z	Depth from soil surface to bottom of layer (mm)	−0.3	0.3	r_	0.22
SOL_AWC	Available water capacity of the soil layer (mm H_2O/mm soil)	−0.3	0.3	r_	−0.14
SOL_K	Saturated hydraulic conductivity (mm/h)	−0.3	0.3	r_	0.30
SLSUBBSN	Average slope length (m)	−0.3	0.3	r_	−0.10

v_ indicates that the parameter value has been changed directly. r_ indicates that the parameter value has been changed relatively.

The success of the model was determined by the NSE (Nash–Sutcliffe efficiency) as an objective function. The coefficient of determination (R^2) and percentage bias (PBIAS) were also considered as performance criteria. Accordingly, the success of calibration was 0.64 and 0.44 of NSE values in the calibration and first validation periods, respectively (Figure 3(a)). Furthermore, the analysis revealed that the coefficient of determination (R^2) values were determined to be 0.65 and 0.50, parallel to the NSE values. The calibrated model was also tested at a different station with the code E12A003 for a long and continuous period (1980–2006). This 26-year complete period is important since it covers the entire time interval used in the modeling. The success of the daily simulation was 0.40 for NSE (Figure 3(b)) and 0.52 for R^2. Finally, the non-continuous data from station D12A033-GEMStat in different time ranges were also tested to verify if the data were suitable for calibration. The reason for using this data (in the period 1980–1987) for modeling is because water quality data are also available in the same period. To discuss the water quality outputs which are presented in the following parts of the study, the flow data in this period should also be verified. Accordingly, the performance of the model in simulating the flow data with 96 measuring points in this period was found to be 0.35 for NSE and 0.40 for R^2. Accordingly, it was determined that the model simulates the flow data acceptably. It highlights that the main point to be clarified is the satisfactory nature of simulations when the results of both model calibration and validation processes are examined. Specifically, during the calibration period, NSE of 0.64 and R^2 of 0.65 are achieved. In the validations, the obtained results include an R^2 value of 0.50, 0.52, and 0.40, as well as NSE values of 0.44, 0.40, and 0.35. All of these values are considered to be at an acceptable level, further reinforcing the notion of obtaining reliable simulations. The pictorial representation of the scattered points of the instantaneous flow data is given in Figure 3(c). A coherent trend between measured and simulated flow data was observed. Besides, the validation metric employed is obtained from GEMStat stream flow values recorded at the time of water quality measurements. These stream flow values are utilized to validate the simulations performed by the SWAT model. It is noteworthy that the low

Downloaded from http://iwaponline.com/wst/article-pdf/88/7/1621/1307524/wst088071621.pdf
by guest

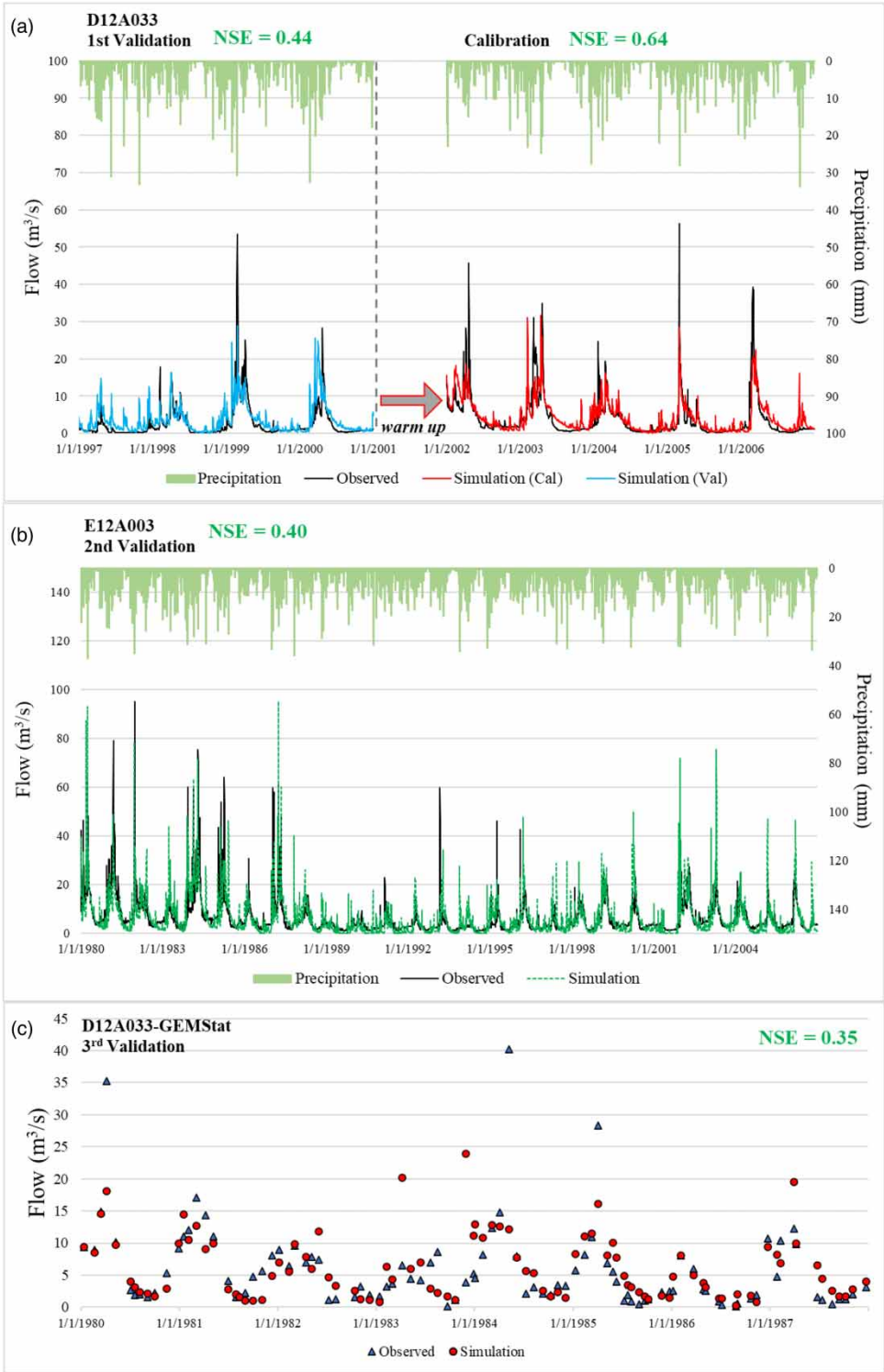


Figure 3 | Precipitation and flow relation for (a) the calibration and the first validation, (b) the second validation, and flow values for (c) the third validation periods.

NSE value (0.35) is mainly due to comparing only 96 discrete measurements of stream flow against the continuous simulation results generated by the model, as opposed to a continuous comparison. However, despite these limitations, the fact that the NSE value remains positive and close to 0.50 (Moriasi *et al.* 2007) is indeed encouraging.

Downloaded from http://iwaponline.com/wst/article-pdf/88/7/1621/1307524/wst088071621.pdf
by guest

3.2. Water quality: NO_3, TSS, and DO

After the flow calibration and validation procedure, simulations of the water quality data were examined in terms of the obtained variables (NO_3, TSS, and DO). The water quality variables provided by the GEMStat data portal were used to calibrate the SWAT model outputs. Calibration was done manually using values in the literature (Arnold *et al.* 2012a, 2012b). Nitrate-related and TSS-related parameters were tested with multiple re-runs and their sensitivities were determined accordingly. Also, the most effective parameters were adjusted as recommended by Arabi *et al.* (2008), Kannan (2012), Qiu *et al.* (2012), and Abbaspour *et al.* (2015) to improve the performance. As a result, the selected eight parameters were adjusted manually, and the values showing the highest performance were fit. The calibrated values of the determined parameters and the descriptions of these parameters are given in Table 4. Also, results were adjusted in kg for NO_3 and in tons for TSS. Furthermore, DO concentration in the stream obtained from GEMStat was converted from mg/L to kg to calibrate with the data obtained from the SWAT model. In the transformation of water quality variables such as load from concentration (mg/L), flow data provided from the GEMStat data portal, which is previously validated in Section 3.1, were used (third validation period).

Figure 4 shows that the NO_3, TSS, and DO outputs of the model – run with the parameters fitted after manual calibration – are coherent with the GEMStat data. In addition, as seen in Figure 4, these are non-continuous data for certain days (96 measurements). Although a perfect fit was not achieved, the model outputs fit the data almost accurately. The metrics of this pictorial trend are given in Table 5. NSE was taken into account in the calibration, as well as R^2 and PBIAS values were presented as other performance criteria. In addition, the calibration periods and the number of samples for the water quality variables analyzed are given in Table 5. The results indicate that the GEMStat data showed reasonable agreement with the model outputs. When NSE values were selected as the objective function, 0.47, 0.42, and 0.61 values were obtained for NO_3, TSS, and DO, respectively. Furthermore, the R^2 values of 0.57, 0.42, and 0.62 align with acceptable levels, similar to the NSE performance. Although PBIAS values of 31.42 are poor performance for NO_3, the absolute PBIAS values of 0.24 and 8.15 are very good performance for TSS and DO. These results indicate that the developed model generates acceptable results.

In addition to the model performance metrics listed in Table 5, the measurement values for the same study area in the literature were examined. Thus, it was aimed to avoid one-sidedness basing the reliability of GEMStat data on SWAT model simulations alone. For this purpose, the river water quality observation data obtained by different researchers in the Upper Porsuk Basin are listed in Table 6. It provides the opportunity to discuss the GEMStat NO_3, TSS, and DO data in terms of order with the available observation data values in the literature. Based on the findings, it is evident that the GEMStat data align with the observations from other data, exhibiting a similar range. Although there may be a few exceptional cases in the GEMStat data regarding TSS, these instances are minimal and inconsequential in terms of the overall situation.

4. CONCLUSIONS

Determining the hydrological behavior of the Upper Porsuk Basin is an important step to assess water quality parameters. For this purpose, as a first stage, the flow outputs were examined with the basin model developed on SWAT. The basin model was

Table 4 | Calibrated values for water quality parameters with their descriptions

Parameter	Descriptions	Calibrated value
RCN	Concentration of nitrogen in rainfall (mg N/L)	1
NPERCO	Nitrate percolation coefficient	0.2
CMN	Rate factor for humus mineralization of active organic nutrients (N and P)	0.0003
CH_COV1	Channel erodibility factor	0.1
CH_COV2	Channel cover factor	0.1
SPCON	Linear parameter for calculating the maximum amount of sediment that can be re-entrained during channel sediment routing	0.00015
SPEXP	Exponent parameter for calculating sediment re-entrained in channel sediment routing	1.05
PRF	Peak rate adjustment factor for sediment routing in the main channel	1.7

Downloaded from http://iwaponline.com/wst/article-pdf/88/7/1621/1307524/wst088071621.pdf
by guest

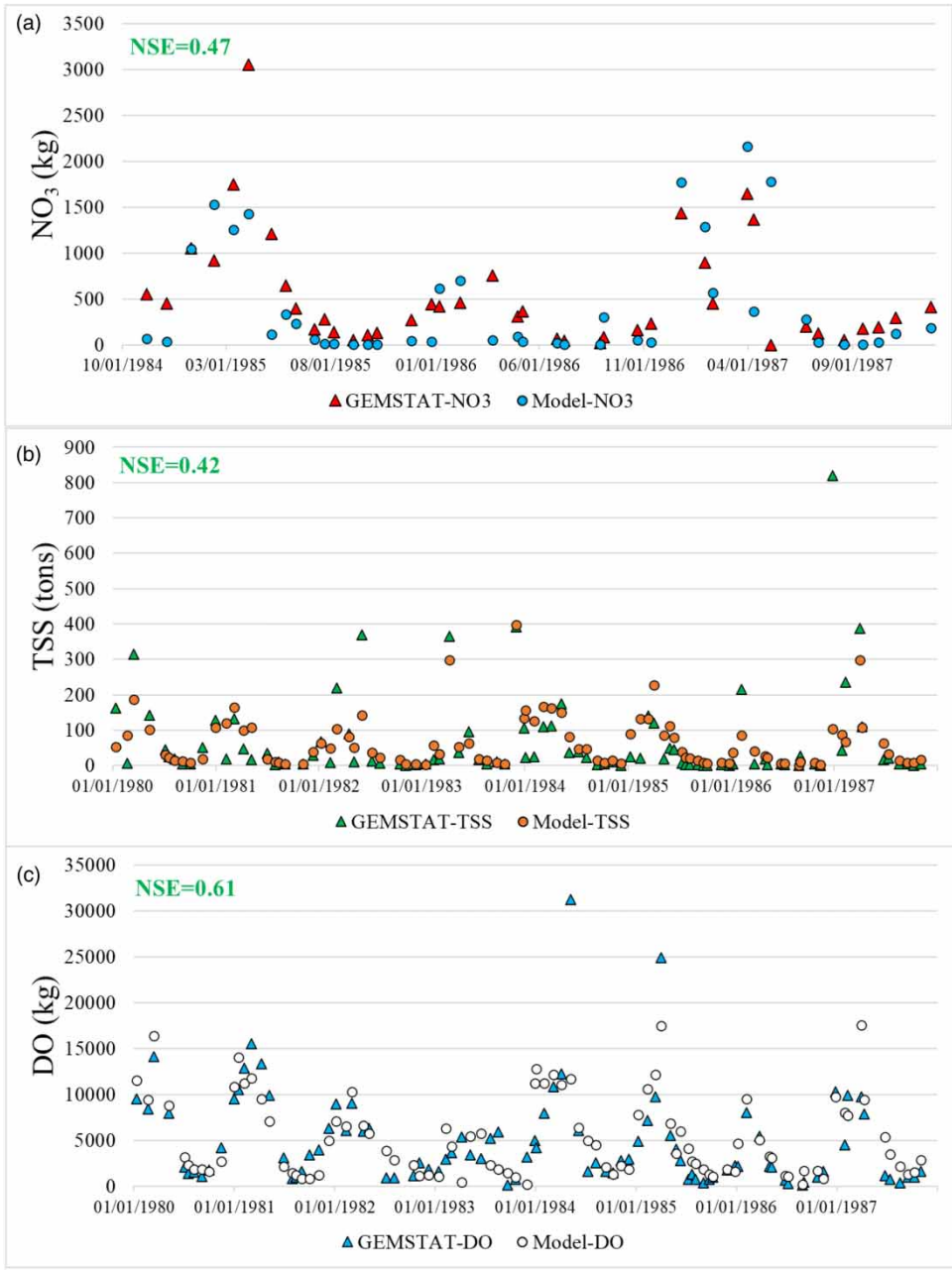


Figure 4 | Calibrated outputs for (a) NO_3, (b) TSS, and (c) DO in the sampling periods with the GEMStat measurement points.

tested by analyzing the flow outputs in the calibration and validation periods. Thus, a reasonable and concrete flow model for the Upper Porsuk Basin was achieved. In the first stage, a sufficient amount of information was obtained to proceed with the development of a water quality model. The NSE values varied between 0.35 and 0.44 for three validation periods. Although these values were below the NSE value of 0.64 obtained in the calibration, they were within the acceptable range. In the second stage of the study, the water quality outputs were calibrated. For this purpose, NO_3, TSS, and DO values provided by the GEMStat data portal were compared with the SWAT model outputs, and the performance of the model was determined. Accordingly, NSE values were found to be 0.47, 0.42, and 0.61 for NO_3, TSS, and DO, respectively.

SWAT is a reliable and widely used modeling tool in terms of water quantity and quality outputs. GEMStat data portal is a very important data source to control outputs of water quality models like SWAT. Only a few measured datasets that can be

Downloaded from http://iwaponline.com/wst/article-pdf/88/7/1621/1307524/wst088071621.pdf
by guest

Table 5 | Model performance results for flow and water quality variables with the periods and number of samples

Variable	Period	Source	Number of samples	Performance criteria		
				NSE	R^2	PBIAS
Flow-Calibration	1.1.2002–12.31.2006	DSI	Continuous-5 year daily	0.64	0.65	− 11.36
Flow-Validation1	1.1.1997–12.31.2000	DSI	Continuous-4 year daily	0.44	0.50	− 35.19
Flow-Validation2	1.1.1980–12.31.2006	DSI	Continuous-26 year daily	0.40	0.52	+ 1.77
Flow-Validation3	1.1.1980–12.31.1987	GEMStat	96 measurements	0.35	0.40	− 7.36
NO_3	1.1.1984–12.31.1987	GEMStat	40 measurements	0.47	0.57	+ 31.42
TSS	1.1.1980–12.31.1987	GEMStat	96 measurements	0.42	0.42	− 0.24
DO	1.1.1980–12.31.1987	GEMStat	96 measurements	0.61	0.62	− 8.15

Table 6 | NO_3, TSS, and DO data available in the literature in the Upper Porsuk Basin with GEMStat

Data source	Ranges			Observation date
	NO_3-N (mg/L)	TSS* (mg/L)	DO (mg/L)	
GEMStat	0.5–2.15	3–500	6.2–12.4	1984–1987
Köse *et al.* (2016)	1.01	NA	7.56	2015
Orak (2006)	NA	NA	8.24–9.11	2001–2002
Solak (2009)	NA	280.2–497.3	4.22–10.13	2006
Güngör (2011)	NA	90	NA	2003–2005
Gürel (2011)	0.15–2.15	NA	1.9–13	2009
Şahin (2018)	0.5–6	NA	9–11	2016

*Ranges and number of samplings for TSS

Range	Number of sampling
3–500	91
500–1,000	2
>1,000	3

used to calibrate water quality models are available and obviously, taking samples and analyzing them would be costly and require a lot of time. Global or continental-scale data portals such as GEMStat help researchers overcome this difficulty. Collection of data from rivers is very important for the calibration and validation of water quality models. In this study, the practicability of the GEMStat global data was tested with a SWAT water quality model for a basin in Türkiye. Furthermore, we conducted a comparison between GEMStat data and readily available measurements of NO_3, TSS, and DO obtained from previous field studies conducted in the Upper Porsuk River section. The similarity in range between the observed values reported in the literature and the GEMStat data serves as an additional validation, supporting the reliability of GEMStat from a different perspective. Therefore, simulated NO_3, TSS, and DO data can be validated using this freely available dataset. The results of this study showed that by sharing previous measurements over the internet, time and labor can be saved since difficult field measurements to calibrate and test water quality models will no longer be required. However, it should be noted that the data in the study is dated before the 2000s. For this reason, we strongly recommend the development of up-to-date and open-source free data portals. As the use of open-source and easily accessible data portals such as GEMStat becomes widespread by researchers or institutions for control of outputs, modeling studies would be more accurate and effective. In conclusion, data scarcity, which is the most important issue in strategy planning practices for water pollution and management, can be overcome.

DATA AVAILABILITY STATEMENT

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

Downloaded from http://iwaponline.com/wst/article-pdf/88/7/1621/1307524/wst088071621.pdf
by guest

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Abbaspour, K. C. 2012 *SWAT-CUP-2012. SWAT Calibration and Uncertainty Program – A User Manual*. Swiss Federal Institute of Aquatic Science and Technology, Dübendorf.

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

Abbaspour, K. C., Vaghefi, S. A., Yang, H. J. & Srinivasan, R. 2019 Global soil, landuse, evapotranspiration, historical and future weather databases for SWAT applications. *Scientific Data* **6** (1). https://doi.org/10.1038/s41597-019-0282-4.

Ahn, S. R. & Kim, S. J. 2019 Assessment of watershed health, vulnerability and resilience for determining protection and restoration priorities. *Environmental Modelling & Software* **122**, 103926.

Arabi, M., Frankenberger, J. R., Engel, B. A. & Arnold, J. G. 2008 Representation of agricultural conservation practices with SWAT. *Hydrological Processes* **22** (16), 3042–3055. https://doi.org/10.1002/hyp.6890.

Arnold, J. G., Srinivasan, R., Muttiah, R. S. & Williams, J. R. 1998 Large area hydrologic modeling and assessment Part I: Model development. *Journal of American Water Resources Association* **34** (1), 73–89.

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. 2012a SWAT: Model use, calibration, and validation. *Transactions of the ASABE* **55** (4), 1491–1508.

Arnold, J. G., Kiniry, J. R., Srinivasan, R., Williams, J. R., Haney, E. B. & Neitsch, S. L. 2012b *Soil and Water Assessment Tool Input/Output Documentation Version 2012*. Texas Water Resources Institute, Texas, p. 7.

Bai, J., Zhao, J., Zhang, Z. & Tian, Z. 2022 Assessment and a review of research on surface water quality modeling. *Ecological Modelling* **466**, 109888.

Barcelona, M., Gibb, J. P., Helfrich, J. A. & Garske, E. E. 1985 *Practical Guide for Groundwater Sampling*. Illinois State Water Survey ISWS Contract Report, Champaign, p. 374.

Barnwell, T. O. & Johanson, R. 1981 HSPF: A comprehensive package for simulation of watershed hydrology and water quality. In: *Nonpoint Pollution Control: Tools and Techniques for the Future*. Interstate Commission on the Potomac River Basin, 1055 First Street, Rockville, MD 20850.

Bhateria, R. & Jain, D. 2016 Water quality assessment of lake water: A review. *Sustainable Water Resources Management* **2**, 161–173.

Brown, L. C. & Barnwell, T. O. 1987 *The Enhanced Stream Water Quality Models QUAL2E and QUAL2E-UNCAS (EPA/600/3-87-007)*. US Environmental Protection Agency, Athens.

Busico, G., Colombani, N., Fronzi, D., Pellegrini, M., Tazioli, A. & Mastrocicco, M. 2020 Evaluating SWAT model performance, considering different soils data input, to quantify actual and future runoff susceptibility in a highly urbanized basin. *Journal of Environmental Management* **266**, 110625. https://doi.org/10.1016/j.jenvman.2020.110625.

Carpenter, S. R., Caraco, N. F., Correll, D. L., Howarth, R. W., Sharpley, A. N. & Smith, V. H. 1998 Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecological Applications* **8** (3), 559–568.

Çelen, M., Karpuzcu, M., Onkal Engin, G., Tetzlaff, B. & Wendland, F. 2014 Modelling total phosphorus input pathways in the Porsuk reservoir catchment in Turkey. *Environmental Earth Sciences* **72** (12), 5019–5034.

Chapman, D. V., World Health Organization, UNESCO & United Nations Environment Programme 1996 *Water Quality Assessments: A Guide to the Use of Biota, Sediments and Water in Environmental Monitoring*, 2nd edn. (Chapman, D., ed.). E & FN Spon, London.

Costa, C. M. d. B., Leite, I. R., Almeida, A. K. & Almeida, I. K. d. 2021 Choosing an appropriate water quality model – a review. *Environmental Monitoring and Assessment* **193** (1), 38.

Di Toro, D. M., Fitzpatrick, J. J. & Thomann, R. V. 1983 *Water Quality Analysis Simulation Program (WASP) and Model Verification Program (MVP) – Documentation*. Hydroscience, Inc., Westwood, NY, for U.S. EPA, Duluth, MN, Contract No. 68-01-3872.

EEA-CORINE Land Cover 2000 *Copernicus Land Monitoring Service*. CORINE Land Cover, Copenhagen, Denmark. Available from: https://land.copernicus.eu/pan-european/corine-land-cover (accessed 20 November 2022).

Ejigu, M. T. 2021 Overview of water quality modeling. *Cogent Engineering* **8** (1), 1891711.

FAO-DSMW Food and Agriculture Organization of the United Nations *FAO Digital Soil Map of the World (DSMW)*. Available from: http://www.fao.org/geonetwork/srv/ (accessed 20 November 2022).

Fuka, D. R., Walter, M. T., MacAlister, C., DeGaetano, A. T., Steenhuis, T. S. & Easton, Z. M. 2014 Using the climate forecast system reanalysis as weather input data for watershed models. *Hydrological Processes* **28** (22), 5613–5623.

Gao, L. & Li, D. 2014 A review of hydrological/water-quality models. *Frontiers of Agricultural Science and Engineering* **1** (4), 267–276.

Gassman, P. W., Reyes, M. R., Green, C. H. & Arnold, J. G. 2007 The soil and water assessment tool: Historical development, applications, and future research directions. *Transactions of the ASABE* **50** (4), 1211–1250.

Gülbaz, S. 2019 Water quality model for non-point source pollutants incorporating bioretention with EPA SWMM. *Desalination and Water Treatment* **164**, 111–120.

Downloaded from http://iwaponline.com/wst/article-pdf/88/7/1621/1307524/wst088071621.pdf
by guest

Güngör, Ö. 2011 *Determination and Modeling of Suspended Solids Transport in the Lower Porsuk Stream Watershed*. Master's Dissertation, Anadolu University. CoHE Thesis Center (Thesis no: 283728).

Gürel, E. 2011 *Determination of Water Quality of Porsuk Stream*. Master's Dissertation, Eskişehir Osmangazi University. CoHE Thesis Center (Thesis no: 299820).

Kannan, N. 2012 *SWAT Modeling of the Arroyo Colorado Watershed; TR-426*. Texas Water Resources Institute, College Station, TX, USA.

Khalid, K., Ali, M. S. a. M., Rahman, N. F. A., Mispan, M. R., Haron, S. H., Othman, Z. & Bachok, M. F. 2016 Sensitivity analysis in watershed model using SUFI-2 algorithm. *Procedia Engineering* **162**, 441–447.

Kirschke, S., Avellán, T., Bärlund, I., Bogardi, J. J., Carvalho, L., Chapman, D., Dickens, C. W. S., Irvine, K., Lee, S. B., Mehner, T. & Warner, S. 2020 Capacity challenges in water quality monitoring: Understanding the role of human development. *Environmental Monitoring and Assessment* **192**, 298.

Köse, E., Çiçek, A., Uysal, K., Tokatlı, C., Arslan, N. & Emiroğlu, Ö. 2016 Evaluation of surface water quality in Porsuk Stream. *Anadolu University Journal of Science and Technology C – Life Sciences and Biotechnology* **4** (2), 81–93.

Köse, E., Emiroğlu, Ö., Çiçek, A., Tokatli, C., Başkurt, S. & Aksu, S. 2018 Sediment quality assessment in Porsuk Stream Basin (Turkey) from a multi-statistical perspective. *Polish Journal of Environmental Studies* **27** (2), 747–752.

Kwakye, S. O. & Bárdossy, A. 2020 Hydrological modelling in data-scarce catchments: Black Volta basin in West Africa. *SN Applied Sciences* **2**, 628.

López-Ballesteros, A., Nielsen, A., Castellanos-Osorio, G., Trolle, D. & Senent-Aparicio, J. 2023 DSOLMap, a novel high-resolution global digital soil property map for the SWAT+ model: Development and hydrological evaluation. *Catena* **231**, 107339. https://doi.org/10.1016/j.catena.2023.107339.

McKay, M. D., Beckman, R. J. & Conover, W. J. 1979 Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21** (2), 239–245.

Monteith, J. L. 1965 Evaporation and the environment. In: *19th Symposia of the Society for Experimental Biology*. Vol. 19, pp. 205–234.

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

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, Forney, TX, USA, pp. 1–618.

Newcomer, M., Dogulu, N., Iravani, H., Dembélé, M., Uysal, G., Roy, T. & Fischer, S. 2022 Open and free datasets for hydrology research: Insights, challenges and opportunities. In: *IAHS-AISH Scientific Assembly 2022*, 29 May–3 Jun 2022, Montpellier, France. IAHS2022-310.

Orak, E. 2006 *Water Quality Modelling With Fuzzy Logic İn Porsuk Creek*. Master's Dissertation, Gebze Institute of Technology. CoHE Thesis Center (Thesis no: 182578).

Park, R. A., Clough, J. S. & Wellman, M. C. 2008 AQUATOX: Modeling environmental fate and ecological effects in aquatic ecosystems. *Ecological Modelling* **213** (1), 1–15.

Qiu, L., Zheng, F. & Yin, R. S. 2012 SWAT-based runoff and sediment simulation in a small watershed, the loessial hilly-gullied region of China: Capabilities and challenges. *International Journal of Sediment Research* **27** (2), 226–234.

Şahin, M. 2018 *Assessment of Porsuk Stream According to Water Quality Indexes and Determination of Dam Lake Trophic Level*. Doctoral Dissertation, Anadolu University. CoHE Thesis Center (Thesis no: 509764).

ShangGuan, W., Dai, Y., Duan, Q., Liu, B. & Yuan, H. 2014 A global soil data set for earth system modeling. *Journal of Advances in Modeling Earth Systems* **6** (1), 249–263. https://doi.org/10.1002/2013ms000293.

Sivapalan, M. 2003 Prediction in ungauged basins: A grand challenge for theoretical hydrology. *Hydrological Processes* **17** (15), 3163–3170.

Solak, C. N. 2009 *The Determination of Pollution in the Felent Creek (Porsuk-Kütahya) by Using Some Aquatic Organisms*. Doctoral Dissertation, Dumlupınar University. CoHE Thesis Center (Thesis no: 237946).

SRTM Shuttle Radar Topography Mission 1 Arc-Second Global. doi:10.5066/F7PR7TFT. Available from: https://www.usgs.gov/ (accessed 20 November 2022).

Stewart, B. 2015 Measuring what we manage – the importance of hydrological data to water resources management. *Proceedings of IAHS* **366**, 80–85.

Tsakiris, G. & Alexakis, D. 2012 Water quality models: An overview. *European Water* **37**, 33–46.

UNEP-GEMS/Water Programme 2006 *UNEP Global Environment Monitoring System*. Water Programme. Available from: www.gemstat.org (accessed 20 November 2022).

United Nations General Assembly 2015 *Transforming Our World: The 2030 Agenda for Sustainable Development*. Available from: https://sdgs.un.org/sites/default/files/publications/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf (accessed 20 November 2022).

USEPA 2000 *National Water Quality Inventory*. EPA 841-R-02-001, Office of Water, United States Environmental Protection Agency (EPA), Washington, DC, USA, p. 460.

USGS 2006 *Benefits of the USGS Stream Gauging Program – Users and Uses of US Streamflow Data*. Available from: http://water.usgs.gov/osw/pubs/nhwc_report.pdf

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. *Journal of Hydrology* **324** (1–4), 10–23.

Downloaded from http://iwaponline.com/wst/article-pdf/88/7/1621/1307524/wst088071621.pdf
by guest

Virro, H., Amatulli, G., Kmoch, A., Shen, L. & Uuemaa, E. 2021 GRQA: Global river water quality archive. *Earth System Science Data* **13** (12), 5483–5507.

Wagner, S., Kunstmann, H., Bárdossy, A., Conrad, C. & Colditz, R. R. 2009 Water balance estimation of a poorly gauged catchment in West Africa using dynamically downscaled meteorological fields and remote sensing information. *Physics and Chemistry of the Earth, Parts A/B/C* **34** (4–5), 225–235.

Wang, Q., Li, S., Jia, P., Qi, C. & Ding, F. 2013 A review of surface water quality models. *The Scientific World Journal* **2013**, 231768.

White, K. L. & Chaubey, I. 2005 Sensitivity analysis, calibration, and validations for a multisite and multivariable SWAT model. *Journal of the American Water Resources Association* **41** (5), 1077–1089.

WMO 2009 *Guide to Hydrological Practices Volume II, Management of Water Resources and Application of Hydrological Practices*, 6th edn. WMO-No. 168, Geneva.

Yerel, S. 2010 Water quality assessment of Porsuk River, Turkey. *E-Journal of Chemistry* **7** (2), 593–599.

Yüce, G., Pinarbasi, A., Ozcelik, S. & Ugurluoglu, D. 2006 Soil and water pollution derived from anthropogenic activities in the Porsuk River Basin, Turkey. *Environmental Geology* **49** (3), 359–375.

Zessner, M. 2021 Monitoring, modeling and management of water quality. *Water* **13**, 1523.

First received 26 November 2022; accepted in revised form 8 September 2023. Available online 23 September 2023

Downloaded from http://iwaponline.com/wst/article-pdf/88/7/1621/1307524/wst088071621.pdf
by guest