Environmental Engineering and Management Journal

June 2024, Vol. 23, No. 6, 1151-1162
http://www.eemj.icpm.tuiasi.ro/; http://www.eemj.eu
http://doi.org/10.30638/eemj.2024.094

"Gheorghe Asachi" Technical University of Iasi, Romania

EFFECTS OF VARIOUS LAND USE LAND COVER (LULC) DATA ON HYDROLOGICAL MODEL PERFORMANCES

İsmail Bilal Peker[1], Gökhan Cüceloğlu[2], Sezar Gülbaz[1]*, Yusuf Serengil[3]

[1]Civil Engineering Department, Istanbul University-Cerrahpaşa, Avcılar, Istanbul, Turkey
[2]Institute of Earth and Marine Sciences, Gebze Technical University, Gebze, Kocaeli, Turkey
[3]Forest Engineering Department, Istanbul University-Cerrahpaşa, Bahçeköy, Istanbul, Turkey

Abstract

Land cover is a significant input to hydrological models, and its features may affect model performances. To evaluate its impact on evapotranspiration, surface runoff, and water yield, we tested six open-source LULC data products (GLCC, GLC 2000, GlobCover 2005, GLCNMO V1, CLC 1990, and PELCOM) in the Emet-Orhaneli Basin located in western Anatolia. The Soil and Water Assessment Tool (SWAT) was employed to assess hydrological responses. Following the model calibration with observed streamflow data, the changes in outputs over the 1980-2012 period were compared temporally and spatially. The results revealed that temporal and spatial changes in evapotranspiration (up to 2%) and water yield were slight (up to 7%), whereas surface runoff varied more significantly in monthly and interannual intervals. The surface runoff values varied up to 70% for different LULC data in the basin scale and more distinct variations at the subbasin scale. The surface runoff values were highest (80.92 mm) in the case of using GLCC and lowest (48.13 mm) in PELCOM case. We concluded that the LULC data is crucial for estimating surface runoff and peak flow, while it is less effective in estimating evapotranspiration and total water yield. Our results may guide hydrologic modellers in selecting LULC data for specific conditions and purposes.

Key words: forest dominated basin, hydrological modeling, land cover, land use, SWAT

Received: September, 2023; Revised final: January, 2024; Accepted: February, 2024; Published in final edited form: June, 2024

1. Introduction

Land cover refers to the physical and biological material that covers the Earth's surface while land use defines the type and intensity of human activities occurring in a particular location (LaGro, 2005). Land use land cover (LULC) data can be collected using specific technologies such as satellite images, aerial photographs, sensors, and other tools. These data are processed and analyzed through geographic information systems (GIS) software (Bey et al., 2016). LULC data is an essential input for simulations of processes such as water quantity and quality in hydrological systems (Dwarakish and Ganasri, 2015). Especially physical-based watershed models, it is critical to accurately define the land use characteristics of the basin in the model for the reliability of the results. Field measurements can be conducted to acquire data on soil or land use (Georgakakos and Baumer, 1996; Oruç et al., 2022; Yan et al., 2015). However, these measurements are laborious and costly (Gibbs et al., 2007). For this reason, data sets produced on a global or regional-scale using remote sensing methods are available on various platforms (Grekousis et al., 2015). The reliability of these data sets and their uncertainty is an important concern in modeling studies. Therefore, examining the potential effect of these datasets, which should be accurately defined in watershed models, on the hydrological results in the local regions is quite important.

Publicly available LULC data is provided by many institutions such as the United States Geological

* Author to whom all correspondence should be addressed: e-mail: sezarg@iuc.edu.tr; Phone: +90212 473 70 70; Fax: +90 212 473 71 80
Pre-published manuscript in Research Square: https://www.researchsquare.com/article/rs-3197488/latest

Survey (USGS), Food and Agriculture Organization (FAO), European Environment Agency (EEA), and International Union for Conservation of Nature (IUCN). These datasets are obtained using satellite images with different spatial and temporal resolutions. Grekousis et al. (2015) presented a large-scale survey of LULC data generated at 21 global and 43 regional-scales. Accordingly, Global Land Cover Characterization (GLCC) (Loveland et al., 2000), Global Land Cover 2000 (GLC 2000) (Bartholomé and Belward, 2005), GlobCover (Bicheron et al., 2008) and the Global Land Cover by National Mapping Organizations (GLCNMO) (Tateishi et al, 2011), as well as Coordination of Information on the Environment (CORINE) (Buttner et al., 1998) and the Pan-European Land Cover Monitoring Project (PELCOM) (Mücher et al., 2000) are some of the frequently used public global- and European-scale datasets. These six datasets are classified by different classification systems according to their purpose and are stored at different spatial resolutions. Even if different global- and European-scale LULC input data for hydrological modeling are successfully used for different purposes, the suitability or applicability of these global products to simulate hydrology cannot be guaranteed. Geological and environmental differences between regions require detailed investigation and accuracy assessment when using remote sensing data (Alawi and Ozkul, 2023). For this reason, these products should be tested and evaluated in order to understand their effect on hydrological outputs.

In recent years, numerous GIS-based watershed modeling tools, employing LULC data through mathematical equations to describe hydrological processes, have been introduced. Among them, SWAT (Arnold et al., 1998) is a semi-distributed process model used worldwide (Aloui et al., 2023; Gassman et al., 2007). Global-scale land use data has been successfully used in many SWAT applications (Abbaspour et al., 2019; Ali et al., 2023; Peker and Gulbaz, 2023). Several studies have examined the effects on model outputs using different LULC data with a hydrological model (Busari et al., 2021; Chirachawala et al., 2020; Huang et al., 2013; M'Barek et al., 2023; Romanowicz et al., 2005). In a study conducted by El-Sadek and Irvem (2014), three different LULC data (GLCC, GlobCover, and CORINE) were used in the SWAT model. Their results indicated that the SWAT model was sensitive to the LULC input data. In another study, two different LULC data produced by CORINE and LANDSAT 7 ETM were examined with SWAT (Cuceloglu et al, 2021). Although the authors observed that both data yielded similar results in streamflow simulations, spatial differences were reported at the subbasin-scale.

The present study's main objective is to examine the effects and applicability of six different LULC datasets on the hydrological responses. The Emet and Orhaneli Basin in Türkiye, characterized by a significant presence of forests, was selected as the study area for this research endeavor. Recent studies have relied on watershed modeling to investigate the substantial influence of forested areas on surface runoff (Ding et al., 2022; Li et al., 2020; Luo et al., 2020; Zhang et al., 2020). These studies consistently reveal that a reduction in forest cover leads to increased runoff, whereas afforestation has the opposite effect. The current study focuses on assessing the impact of global Land Use and Land Cover (LULC) data, utilized as input for a hydrological model, on surface runoff within a watershed predominantly covered by forests. First, separate models were set up using SWAT with six various LULC datasets (GLCC, GLC 2000, GlobCover 2005, GLCNMO V1, CLC 1990 and PELCOM). Each individual model was calibrated for streamflow. Second, the hydrological responses (evapotranspiration, surface runoff, water yield) were evaluated temporarily and spatially for each model. Within this context, monthly and annual changes in mean evapotranspiration (ET), surface runoff (SURQ), and water yield (WYLD) values obtained during the simulation period of 1980-2012 were investigated. In addition, variations in the average values at the subbasin-scale were obtained, and the spatial variations in model results were determined. As a result, the hydrological responses of six LULC input data on model simulations were compared and presented.

2. Materials and methods

2.1. Study area and data

Emet-Orhaneli is a subbasin of the Susurluk Basin, one of Türkiye's 25 major basins (Fig. 1). The Emet and Orhaneli streams, which form the basin unite, merge to form Mustafa Kemal Pasa River, and drain into Uluabat Lake. The drainage area of this basin is 9537 km^2. The mean elevation is approximately 1042 m and ranges from 20 m (at the outlet part) to 2072 m (in the high parts of Murat Mountain where the Orhaneli Stream originates). The average slope is approximately 18%. Although the LULC classes of the basin vary in the data from various sources, it is mostly covered with agricultural lands, forests, and range lands. C-type alluvial soil is dominant in the basin.

Geospatial (elevation, LULC, and soil type) and meteorological (precipitation, max/min temperature, relative humidity, solar radiation, wind speed) input data are required to create a hydrological model base. The 90 m resolution Shuttle Radar Topography Mission (SRTM) (Jarvis et al., 2008) provided by the National Aeronautics and Space Administration (NASA) and 1:5000000 scale Digital Soil Map of World (DSMW) (http://www.fao.org/geonetwork/srv/) data provided by FAO were used for elevation and soil type data, respectively.

LULC data obtained from six various sources were used in the current study. These datasets were GLCC, GLC 2000, GlobCover 2005, GLCNMO V1, CLC 1990, and PELCOM. The characteristics of the

land use data are summarized in Table 1. For meteorological data, Climate Forecast System Reanalysis (CFSR) (Fuka et al., 2014), which is widely used in SWAT studies (Tan et al., 2021) was used. This data includes daily precipitation, maximum/minimum temperature, relative humidity, solar radiation, and wind speed.

2.2. Analysis and classification of the LULC data

The LULC products from GlobCover, GLCNMO, and CLC datasets capture different periods. GlobCover provides separate data versions for 2005 and 2009, GLCNMO for 2003 (V1), 2008 (V2), and 2013 (V3), and CORINE offers distinct data versions for CLC (CORINE Land Cover) in 1990, 2000, 2006, 2012, and 2018. In the current study, versions of GlobCover 2005, GLCNMO V1, and CLC 1990 were used to represent the model period of the datasets. The properties of the selected LULC data are given in Table 1. Six various LULC maps used for comparison are shown in Fig. 2. The distribution of LULC classes obtained for these data is shown in Fig. 3 as areal percentages. In all datasets, three predominant classes were observed in the Emet-Orhaneli Basin: agricultural areas, forests, and ranges. These three classes spatially varied among different data. The GLCC data identified agricultural areas in approximately 87% of the basin, whereas PELCOM indicated a brushes coverage of 70%. In contrast, the GlobCover 2005 data revealed that approximately 71% of the total basin area was covered by forests. In the GLC 2000, GLCNMO V1 and CLC 1990 data, none of the categories-agriculture, forest, and range-demonstrated dominance in terms of coverage or representation. The study results compare the effects of six different LULC data on the model results in their classification formats. Six various LULC data are classified with diverse types of details, reflecting their unique nature and intended applications. The representation of these classification formats in the hydrological model for the basin in this study exhibits variations. In this context, six different models were created and calibrated for six different data.

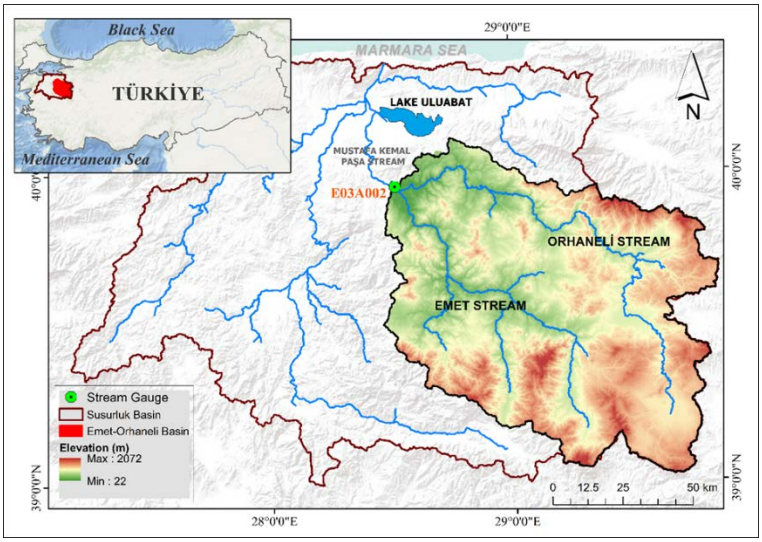

Fig. 1. Location and boundary of Emet-Orhaneli Basin

Table 1. Properties of LULC data

Scale	*LULC data*	*Spatial resolution*	*Time span*	*Satellite*	*Classification system*	*Producer*
GLOBAL	GLCC	1 km	1992-1993	AVHHR[1]	IGBP[2] 17 classes	USGS, UNL[3]
	GLC 2000	1 km	1999-2000	SPOT[4] 4	FAO LCCS[5] 22 classes	JRC[6]
	GlobCover 2005	300 m	2004-2006	MERIS FR[7]	FAO LCCS 22 classes	ESA[8]
	GLCNMO V1	1 km	2003	MODIS[9]	FAO LCCS 22 classes	ISCGM[10]
EUROPEAN	CLC 1990	100 m	1986-1999	Landsat 4,5 TM	5 main classes	EEA
	PELCOM	1 km	1995-1999	AVHRR	16 classes	JRC

[1]AVHRR: Advanced Very High-Resolution Radiometer, [2]IGBP: The International Geosphere–Biosphere Programme, [3]UNL: University of Nebraska-Lincoln, [4]SPOT: Satellite Pour l'Observation de la Terre, [5]LCCS: Land Cover Classification System, [6]JRC: Joint Research Center, [7]MERIS FR: The Medium Resolution Imaging Spectrometer Full Resolution, [8]ESA: European Space Agency, [9]MODIS: Moderate Resolution Imaging Spectroradiometer, [10]ISCGM: International Steering Committee for Global Mapping

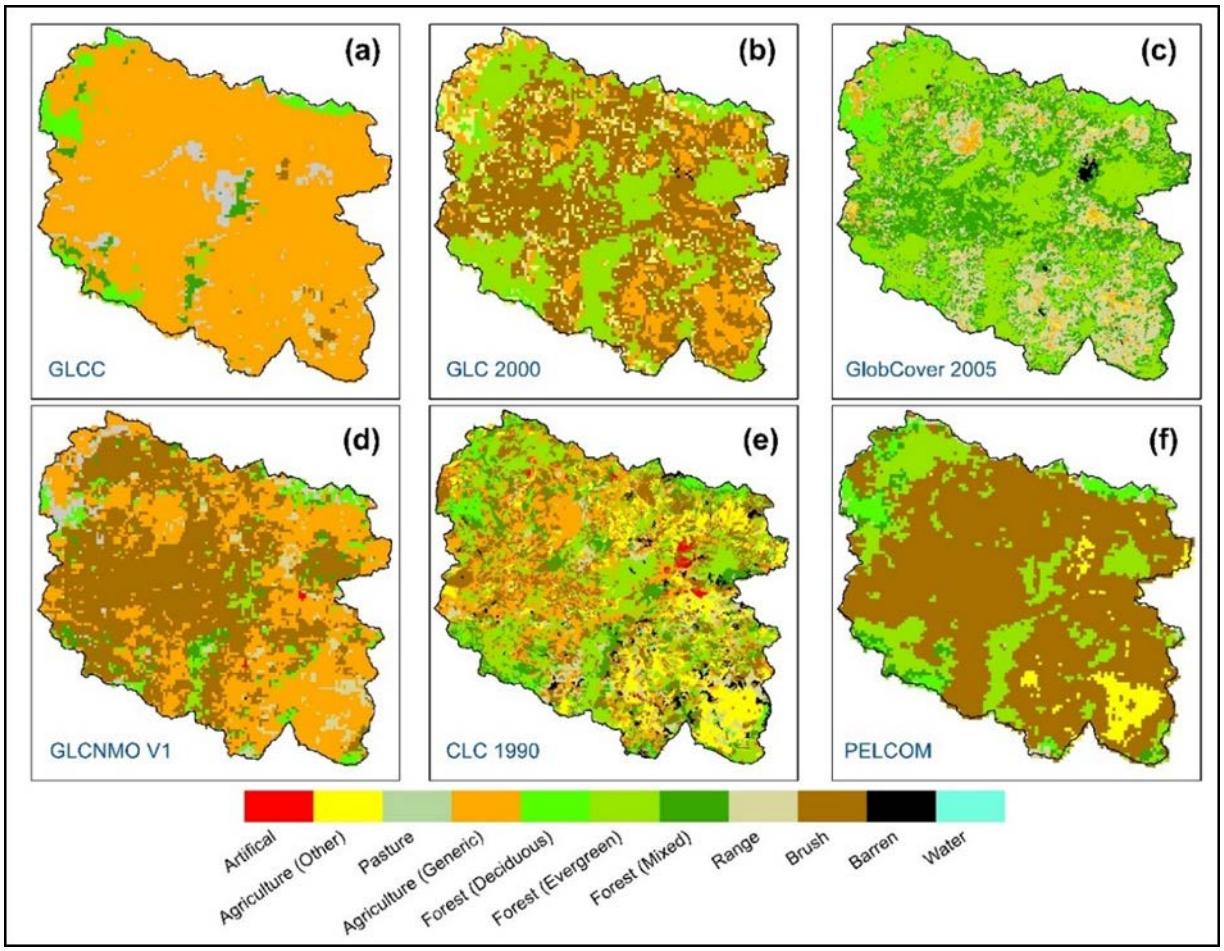

Fig. 2. General view of (a) GLCC, (b) GLC 2000, (c) GlobCover 2005, (d) GLCNMO V1, (e) CLC 1990, (f) PELCOM data

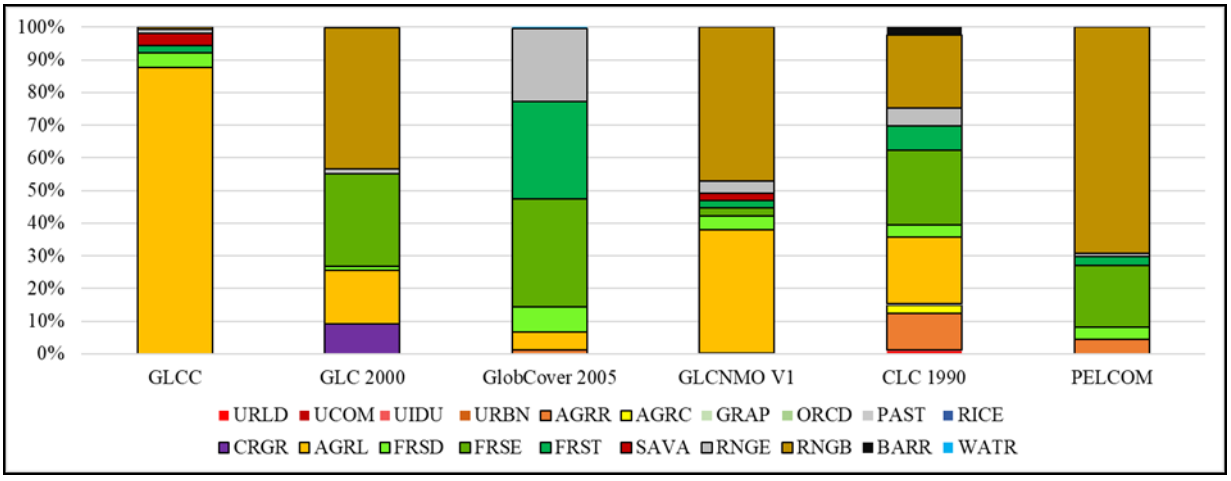

Fig. 3. LULC data classification with SWAT codes

2.3. Hydrological modeling

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; Bieger et al. 2016; Neitsch et al. 2011). The model has a wide range of applications, including flow and nonpoint source pollution modeling in changing environments. On the other hand, the simplification of flow routing through a modified rational method and determination of peak runoff rate with an empirical-type rational formula result in the incapability of detailed flood routing for single events. The hydrological simulation of a watershed involves two major divisions in SWAT. The first part is the land phase of the hydrologic cycle (Fig. 4), which regulates the quantities of water, sediments, nutrients, and pesticides transported to the main channel in each subbasin. The second part is the water or routing phase of the hydrologic cycle, addressing the movement of water, sediments, nutrients, pesticides, etc., through

the channel network of the watershed to the outlet (Neitsch et al., 2011) (Eq. 1).

$$SW_t = SW_0 + \sum_{i=1}^{t}(R_{day} - Q_{surf} - E_{\alpha} - W_{seep} - Q_{gw}) \quad (1)$$

where: SW_t is the final soil water content (mm), SW_0 is the initial soil water content on day i (*mm*), t is the time (days), R_{day} is the amount of precipitation on day i (mm), Q_{surf} is the amount of surface runoff on day i (mm), E_{α} is the amount of evapotranspiration on day i (mm), w_{seep} is the amount of water entering the vadose zone from the soil profile day i (mm), and Q_{gw} is the amount of return flow on day i (mm).

In SWAT, surface runoff was estimated by the SCS-CN method, which depends on land use and soil type characteristics (Eqs. 2-3) (Neitsch et al., 2011).

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

$$S = 25.4(\frac{1000}{CN} - 10) \quad (3)$$

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

In the model, the basin area is first divided into subbasins based on topography, and each subbasin is then divided into hydrological response units (HRU) according to slope, LULC, and soil type characteristics (Neitsch et al., 2011). The basin in all models was separated into 43 subbasins using the automatic basin delineation option in SWAT based on the digital elevation model. The lowest subbasin threshold area was selected as 100 km^2. Each subbasin was further divided into five elevation bands to increase the representation of elevation-related variations. The slope was divided into three classes as <5%, 5-15%, and >15%. To create HRUs, no threshold values were defined for slope, land use, and soil type variables. The models were run between 1979 and 2012, depending on the data periods of the CFSR and the streamflow observation station. Also, Penman-Monteith (Monteith, 1965) method was used for ET calculation in the model. The latest version of ArcSWAT 2012.10.26 version, which can run on ArcGIS/ArcMap 10.8, was used as the modeling tool to perform these calculations. In the models created, only the LULC data (GLCC, GLC 2000, GlobCover 2005, GLCNMO V1, CLC 1990, and PELCOM) were modified, and all other model conditions were kept constant.

LULC data with different classifications were used in modeling. Model-specific codes must be defined to identify LULC data in SWAT. For this reason, the classifications of these data were matched in accordance with SWAT-specific LULC codes and inputted the model. The nearest features for the relevant LULC class were considered for these classifications. The number of classes in these categorizations performed with the same procedure is variable for each data.

Considering the CFSR data covering 1979-2012, the model calibration period was determined as 1980-2000 (21 years), and the validation period as 2001-2012 (12 years). The data obtained by E03A002 streamflow observation station located at the basin outlet were used for the observed flow values in these periods. These data are open source and can be accessed from the annals of Turkish State Hydraulic Works (DSI). In all models, 1979 (1 year) was run as the warming year. Sequential Uncertainty Fitting - Version 2 (SUFI-2) Latin Hypercube Sampling (LHS) algorithm (McKay et al., 1979), which is included in the public version of the SWAT-Calibration and Uncertainty Program (SWAT-CUP) (Abbaspour et al., 2004, 2007) was used for the calibration process. Nash-Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970). metric was considered the algorithm's objective function. Each model established with different LULC data was examined by running 300 simulations with the same parameters and the initial value ranges of these parameters. The parameters, descriptions, units and initial value ranges used in the calibration are shown in Table 2. The calibrated parameters obtained were run for the validation period, and the models were validated for different periods.

In Fig. 5, the flowchart of study is depicted, offering a concise representation of the connections between inputs, the hydrological model framework, calibration, and outputs.

3. Results

In the analysis, six different LULC data were compared for two aspects: first, the streamflow calibration and validation based on the gauge data at the basin outlet were compared according to the success metrics. At this step, the trendy conditions and visual harmony in the hydrographs were also examined. Secondly, monthly and annual average values of three important hydrological cycle components (ET, surface runoff, and water yield) of the model results were compared for the given period. In the second part of the study, comparisons were evaluated as temporal values for the study period and spatially at the subbasin-scale on average.

3.1. Streamflow calibration and validation results

Streamflow results obtained by the models built with six different LULC data with their different classifications were calibrated based on the gauge data. NSE was considered the objective function in the calibrations. In addition, the coefficient of determination (R2) and Kling-Gupta Efficiency (KGE) (Gupta et al., 2009) metrics were also compared. Depending on the characteristics of the parameters, the suitable set is obtained by modifying

them through either relative or direct replacement. The fitting values of parameters with obtained by the calibrations are given in Table 2. Among these parameters, CN2 and SOL_AWC show relative spatial variability in HRUs (denoted by r_). Others are direct change of parameter values (denoted by v_). For example, for each model of LULC data, there is a -0.09% difference between the CN2 initial value and its calibrated value. Thus, at the basin average, the initial and calibrated values of CN2 are 80.90 and 73.62 for GLCC, 79.19 and 72.06 for GLC 2000, 79.61 and 72.44 for GlobCover 2005, 78.71 and 71.63 for GLCNMO V1, 79.79 and 72.61 for CLC 1990, and 76.69 and 69.79 for PELCOM. While the calibrated CN2 value is the same relative decrease in all models, physical values vary depending on LULC and soil groups. Additionally, six interrelated groundwater process parameters were identified as sensitive parameters in the calibration process. This result is similar to studies conducted in the literature (Abbaspour et al., 2015; Akbaş et al., 2020). The complexity and uncertainty in groundwater processes, which are difficult to observe, may make these parameters more effective in the model. Consequently, the results of the calibration and validation of six different models are given in Table 3.

Fig. 6(a) exhibits the simulated and observed streamflow values throughout the simulation period of 1980-2012. Hence, simulations conducted during both the calibration and validation periods may demonstrate similarity and have a closely related trend. While the timing of the peaks in streamflow is consistent for the entire model period, there are differences in the peak flow amounts according to the observation values. In this case, peak flows in SWAT simulations are generally higher than observation values. Fig. 6(b, c) show hydrographs in the randomly selected water year 1981 and 2002 which was chosen for detailed examination, respectively. The dashed black line depicts the observation flow. The simulation lines are tightly grouped together, with the red lines having relatively higher values indicating the GLCC data and the green lines at the lower values representing the GlobCover 2005 data. In both figures, the values for ascent and descent are nearly coincident, with only minor disparities in peak values.

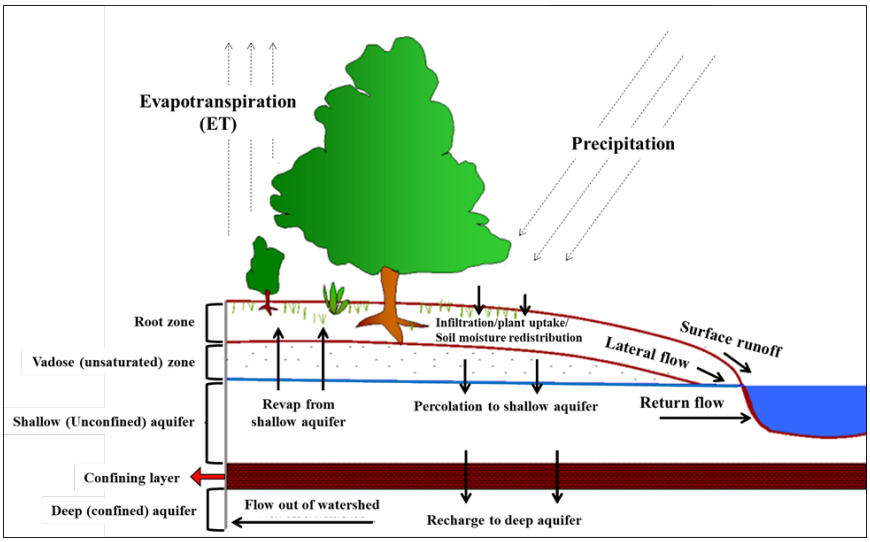

Fig. 4. Schematic representation of the hydrologic cycle in SWAT model (modified from Neitsch et al., 2011)

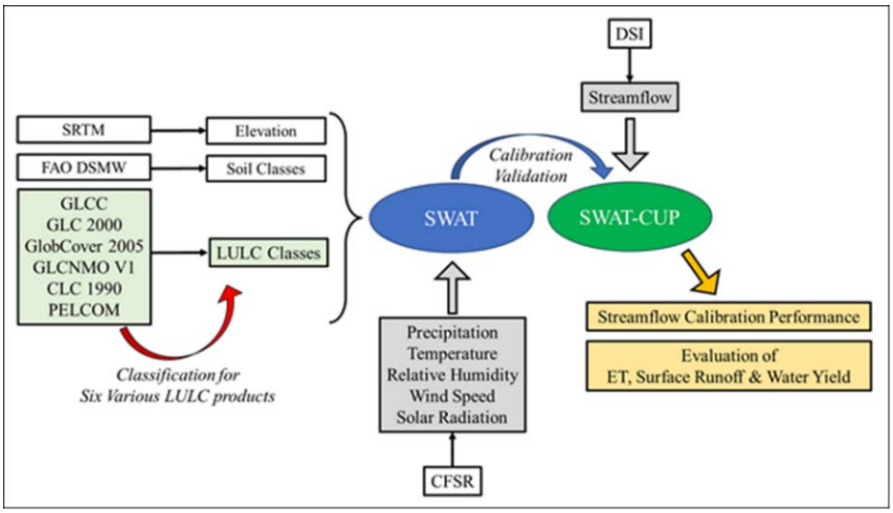

Fig. 5. Flowchart of the study

Table 2. Descriptions, ranges, and fitted values of the sensitive parameters

Parameters	***Description***	***Unit***	***Range***		***Fitted Value***
			Min	***Max***	
v_TLAPS	Temperature lapse rate	ºC	-8	-4	-4.18
v_SFTMP	Snowfall temperature	ºC	-3	3	0.17
v_SMTMP	Snowmelt base temperature	ºC	-3	3	-0.51
v_SMFMX	Melt factor for snow on June 21	mm H_2O/ºC-day	3	6	3.21
v_SMFMN	Melt factor for snow on December 21	mm H_2O/ºC-day	0	3	1.69
v_TIMP	Snowpack temperature lag factor	unitless	0	1	0.08
r_CN2	SCS runoff curve number	unitless	-0.2	0.2	-0.09
r_SOL_AWC	Available water capacity of the soil layer	mm H_2O/mm	-0.2	0.2	-0.18
v_ESCO	Soil evaporation compensation factor	unitless	0.7	1	0.73
v_GW_DELAY	Groundwater delay time	days	0	100	12.5
v_ALPHA_BF	Baseflow alpha factor	days	0	1	0.46
v_RCHRG_DP	Deep aquifer percolation fraction	unitless	0	0.2	0.18
v_GW_REVAP	Groundwater "revap" coefficient	unitless	0	0.2	0.17
v_REVAPMN	Threshold depth of water in the shallow aquifer for "revap" to occur	mm H_2O	0	500	89.17
v_GWQMN	Threshold depth of water in the shallow aquifer for return flow to occur	mm H_2O	0	5000	1641.67

** The term "r_" is used for the relative adjustment of a parameter within a given range and the term "v_" means directly replacing the parameter value with the assigned value.*

Table 3. Performance metrics of SWAT for flow calibration and validation (NSE, R2, KGE)

Land use data	***Calibration (1980-2000)***			***Validation (2001-2012)***		
	NSE	***R2***	***KGE***	***NSE***	***R2***	***KGE***
GLCC	0.61	0.74	0.73	0.28	0.61	0.55
GLC 2000	0.65	0.75	0.77	0.58	0.69	0.75
GlobCover 2005	0.66	0.75	0.78	0.61	0.70	0.78
GLCNMO V1	0.65	0.75	0.77	0.50	0.67	0.69
CLC 1990	0.65	0.75	0.77	0.50	0.66	0.69
PELCOM	0.64	0.74	0.77	0.61	0.70	0.78

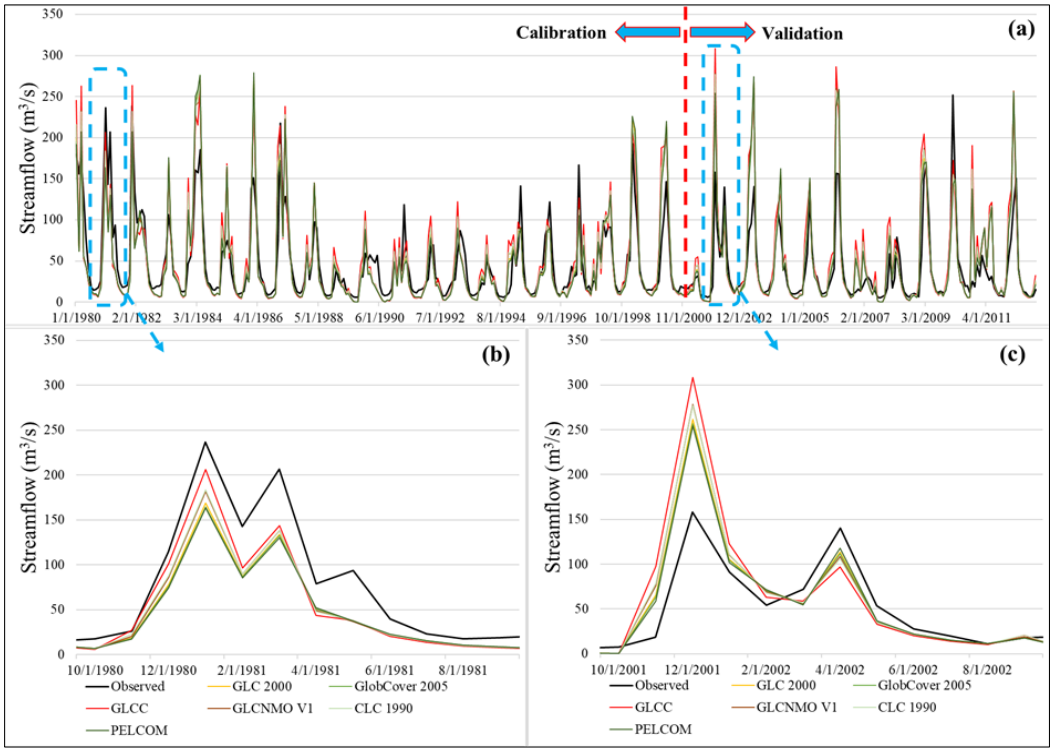

Fig. 6. Hydrographs (a) in the modelling period between 1980-2012, (b) for 1981, (c) 2002 water year

3.2. Monthly and annual change of ET, surface runoff, and water yield

Figure 7 shows the changes in the mean monthly ET, surface runoff, and water yield values in the given period. Accordingly, various LULC data sources were compared with different their classification details. In addition, the observational streamflow data measured at the basin outlet for water yields were compared with the model results.

Table 4 gives the mean annual values of three hydrological cycle components as well as the precipitation. The evapotranspiration/precipitation (ET/PREC), surface runoff/precipitation (SURQ/PREC), and water yield/precipitation (WYLD/PREC) ratios for different LULC data in the entire period are presented.

According to Table 4, the variations in the SURQ/PREC ratio are more noticeable compared to others. In this case, it would be meaningful to elaborate the spatial changes in surface runoffs. Fig. 8 presents the spatial variations in the mean annual surface runoff values, as well as the mean precipitation, at the subbasin scale. Moreover, Fig. 9 was designed with the intention of providing a clearer representation of the variations in surface runoffs at the subbasin scale. In Fig. 9(a), the average values of each subbasin are depicted for the six LULC datasets. Additionally, Fig. 9(b-g) illustrate the deviations of surface runoffs for each LULC dataset compared to the average values. This visual approach enables a more distinct observation of the disparities in surface runoffs among the different LULC data.

Figure 10 exhibits box and whisker plots for each LULC data, providing a statistical visualization of mean surface runoff (SURQ) values in subbasins. By utilizing this graphical representation, it is possible to examine the minimum, maximum, median, and interquartile range (25th and 75th percentiles) values of the SURQ results for each LULC data across 43 subbasins. This allows for a comprehensive statistical analysis of the data distribution and variability. From a statistical perspective, Fig. 10 provides an overview of the surface runoffs in 43 sub-basins by displaying key statistical measures such as the minimum, lower quartile (25th percentile), median (50th percentile), upper quartile (75th percentile), and maximum values. These measures allow for a comprehensive analysis of the data distribution and can be compared alongside the average values to gain further insights into the variability of surface runoffs within the subbasins

4. Discussion

The Emet-Orhaneli Basin consists of agricultural lands, forests, and ranges. However, there were differences in ratios of LULC classes among data sources (Figs. 2-3) to affect hydrological responses of the calibrated models. Accordingly, three different discussions are made based on the results: first, the performance metrics (NSE, R2, and KGE) were compared according to the observed streamflow data for the whole basin outlet. Second, mean monthly and annual ET, surface runoff, and water yield distributions were evaluated over the entire study period (1980-2012). Finally, spatial changes in surface runoff values were observed at the subbasin-scale.

In agreement with previous studies (Chirachawala et al., 2020; Cuceloglu et al, 2021; El-Sadek and Irvem, 2014; M'Barek et al., 2023), the calibrated stream flow (water yield) results in this study showed minimal variations when employing different LULC data. This result was demonstrated in this study for six various LULC data, two of which had not been tested before (GLCNMO V1 and PELCOM). All datasets showed above acceptable calibration performance (Moriasi et al., 2007). Results were evaluated with the NSE metric, which is a widely used metric in determining model performance. Accordingly, a slightly better calibration result according to NSE was obtained with GlobCover 2005 (0.66) and the worst performance was obtained with GLCC (0.61) data. Calibration results of other data varied between 0.64 and 0.65.

In the validation period, however, the metric values decreased, and the GLCC data had an NSE value below 0.50. The poor performance of the GLCC data in terms of NSE value may be attributed to factors such as the aggregation of LULC classes and the low resolution of the dataset (El-Sadek and Irvem, 2014; Romanowicz et al., 2005). However, the relatively better performance of other coarse resolution data suggests that resolution alone may not be the primary reason for the GLCC performance. In addition to the evaluation of performance metrics, hydrograph trends were found to be very close to each other. As an example, slight differences can be observed in the peak conditions of the 1981 and 2002 water years (Fig. 6 b-c). The GLCC had a higher peak value than the other datasets, however, no particular trend differences were found throughout the entire study period. However, the use of low-resolution LULC data could be more desirable since it offers the advantage of minimizing processing and calibration efforts (M'Barek et al., 2023).

For annual and monthly means in the simulation period, ET and water yield showed almost similar trends for all LULC data, while the surface runoff values were relatively different. Moreover, the water yield values were above the observations (overestimation) in all datasets (Fig. 7c). In addition, the ET/PREC, SURQ/PREC, and WYLD/PREC ratios were in the order of 0.5, 0.1, and 0.2, respectively (Table 4). The SURQ/PREC ratio was more sensitive to LULC data compared to other ratios (ET/PREC and WYLD/PREC). This result indicates that different LULC data affect surface runoff more than ET and water yield. Fig. 7(b) illustrates that GLCC (87% agricultural) had the highest surface runoff values, while PELCOM data (70% brushes) had the lowest values. The surface runoff values of GLCC data are extreme compared to the others. This

data stands out from the others due to their underrepresentation of forested areas within the basin.

While the monthly and annual average values of the simulations allowed us to interpret the overall results for the entire basin, visualizations were made to examine the changes in the subbasin-scale. Similar to Cuceloglu et al. (2021), spatial differences in surface runoff values were more striking compared to water yield values. For this purpose, in Fig. 8, surface runoff variations were presented at the subbasin-scale. Fig. 8 revealed that the GLCC data exhibits a higher surface runoff variation than the other datasets. This difference in the GLCC data can be caused by the distribution of the highly sensitive CN2 parameter in the basin, which varies spatially on the basis of HRU. When GLCC data was used in the SWAT model, high CN values and therefore surface runoff values were observed because agriculture dominant HRUs were produced.

On the other hand, low values were observed in PELCOM data because HRUs with high brush class were produced. In Fig. 9, the differences of the data from the mean values at the subbasin scale indicate that GLCC produces higher surface runoff values and PELCOM produces lower values. Hence, it was unveiled in which sub-basins the surface runoffs deviated from the averages, indicating either higher or lower values. The GLCC model subbasins have higher than average values (differences of more than 30 mm), while the PELCOM has low (differences of less than 20 mm) extreme values. In the subbasins of GlobCover 2005, there are no variations exceeding 20 mm from the average values, however, there are values within the range of 10 to 20 mm that fall below the average.

The subbasins of the other three data (GLC 2000, GLCNMO V1 and CLC 1990) differ in the range of ±10 mm and are closer to the mean values. Also, in Fig. 10, GLCC appears to have larger variability than the other five data. Besides the difference in mean value (red dots) and median, box and whisker length indicate a wide range in 43 subbasin surface runoff values for GLCC.

5. Conclusion

The LULC data is a fundamental input of a hydrological model. Using global data sources simplifies the accessibility and application of this input. However, there have been concerns about model performances while using these sources due to their resolution and compatibility issues. Our results revealed that the resolution and classification of the LULC layer had only a slight effect on streamflow in the basin size we studied. Specifically, the streamflow calibration performances for the whole basin outlet were comparable and met the acceptable level for all datasets (NSE>0.50). GlobCover 2005 data yielded slightly better results than the other data in terms of streamflow calibration at the whole basin outlet.

While the temporal and spatial differences in the mean monthly and annual ET and water yield values were quite small, surface runoff values showed relatively larger variability. In particular, the GLCC data showed high surface runoff values from the other datasets. ET and streamflow (water yield) performance of SWAT was not affected by the LULC data source; however, surface runoff performance changed significantly. The effects at the subbasin scale were more apparent.

We concluded that the choice of LULC data in hydrologic modelling with SWAT does not have a significant impact on water yield evaluations in the basin size we studied. However, it can significantly influence hydrologic components such as surface runoff to affect sediment and water quality analyses.

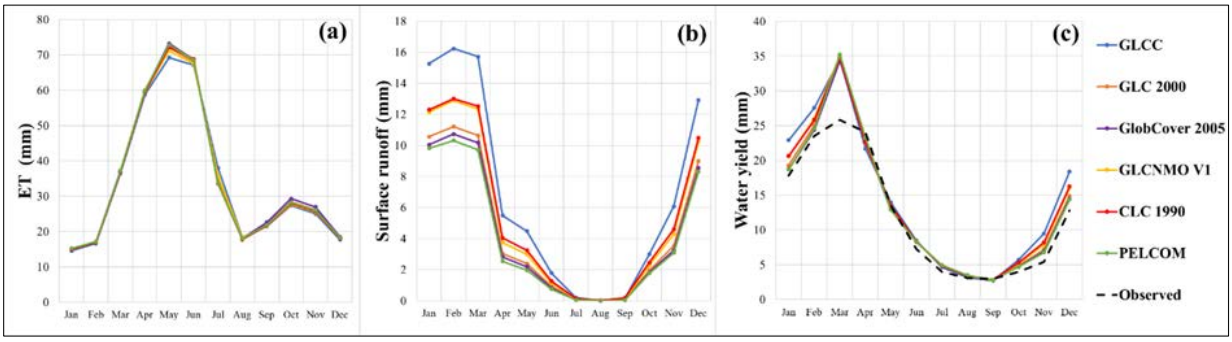

Fig. 7. Mean monthly temporal changes in 1980-2012 for (a) ET, (b) surface runoff, (c) water yield

Table 4. Mean annual results of surface runoff, water yield and ET in the entire modelling period

Data source	*PREC (mm)*	*ET (mm)*	*SURQ (mm)*	*WYLD (mm)*	*ET/PREC*	*SURQ/PREC*	*WYLD/PREC*
GLCC	777.2	410.6	80.92	172.61	0.53	0.10	0.22
GLC 2000		416.2	53.08	161.36	0.54	0.07	0.21
GlobCover 2005		419.4	50.26	158.11	0.54	0.06	0.20
GLCNMO V1		414	61.97	165.17	0.53	0.08	0.21
CLC 1990		413.9	63.96	165.84	0.53	0.08	0.21
PELCOM		416.6	48.13	159.45	0.54	0.06	0.21

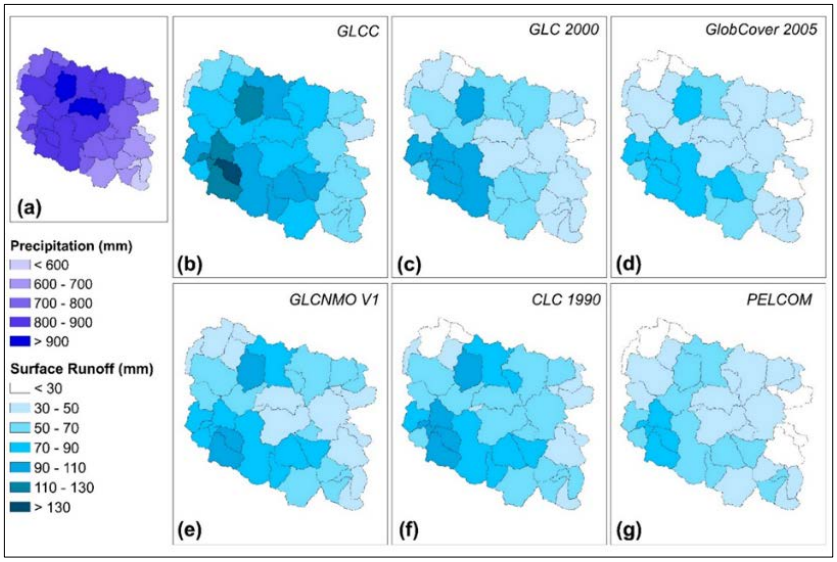

Fig. 8. Mean annual spatial changes in (a) precipitation and surface runoff distributions of (b) GLCC, (c) GLC 2000, (d) GlobCover 2005, (e) GLCNMO V1, (f) CLC 1990, (g) PELCOM at the subbasin-scale

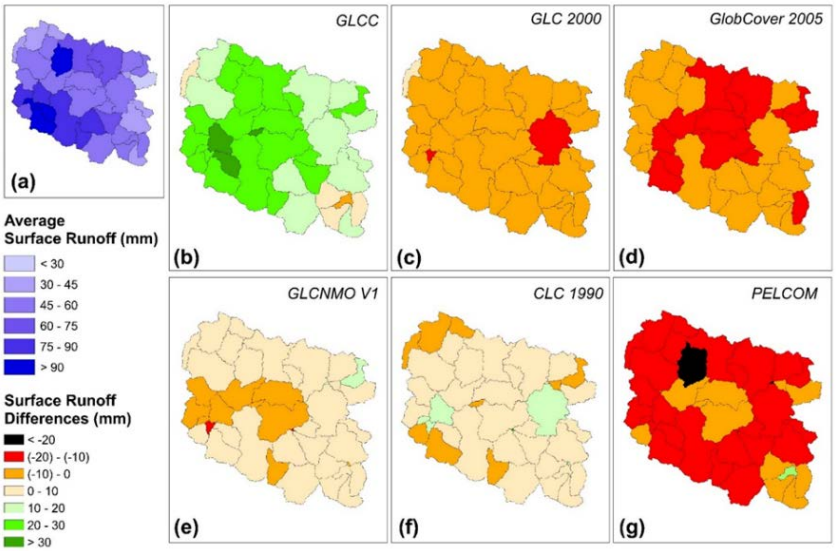

Fig. 9. Surface runoff differences from (a) mean values of all LULC data for (b) GLCC, (c) GLC 2000, (d) GlobCover 2005, (e) GLCNMO V1, (f) CLC 1990, (g) PELCOM in subbasin scale

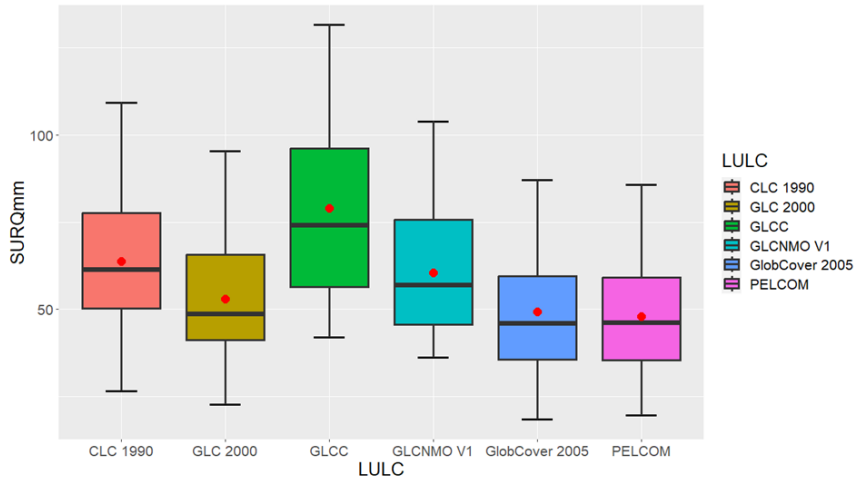

Fig. 10. Box-Whisker plots of surface runoff (SURQ) values for each LULC data in subbasin means

Acknowledgements
The authors would also like to thank the General Directorate of State Hydraulic Works (DSI) for their data support in undertaking this work.

References

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

Abbaspour K.C., Johnson C.L., Van Genuchten M.T., (2004), Estimating uncertain flow and transport parameters using a sequential uncertainty fitting procedure, *Vadose Zone Journal*, **3**, 1340-1352.

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**, 263, https://doi.org/10.1038/s41597-019-0282-4.

Abbaspour K.C., Vejdani M., Haghighat S., Yang J., (2007), *SWAT-CUP Calibration and Uncertainty Programs for SWAT*, The 4th International SWAT Conference, Delft, Netherlands, 1596-1602.

Akbaş A., Freer J., Özdemir H., Bates P., Turp M.T., (2020), What about reservoirs? Questioning anthropogenic and climatic interferences on water availability, *Hydrological Processes*, **34**, 5441-5455.

Alawi S.A., Ozkul S.D., (2023), Evaluation of land use/land cover datasets in hydrological modelling using the SWAT model, *H2Open Journal*, **6**, 63-74.

Ali M., Popescu I., Jonoski A., Solomatine D., (2023), Remote sensed and/or global datasets for distributed hydrological modelling: a review, *Remote Sensing*, **15**, 1642, https://doi.org/10.3390/rs15061642.

Aloui S., Mazzoni A., Elomri A., Aouissi J., Boufekane A., Zghibi A., (2023), A review of Soil and Water Assessment Tool (SWAT) studies of Mediterranean catchments: Applications, feasibility, and future directions, *Journal of Environmental Management*, **326**, 116799, https://doi.org/10.1016/j.jenvman.2022.116799.

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

Bartholomé E., Belward A., (2005), GLC2000: a new approach to global land cover mapping from Earth observation data, *International Journal of Remote Sensing*, **26**, 1959-1977.

Bey A., Díaz A., Maniatis D., Marchi G., Mollicone D., Ricci S., Bastin J., Moore R., Federici S., Rezende M., Patriarca C., Turia R., Gamoga G., Abe H., Kaidong E., Miceli G., (2016), Collect earth: land use and land cover assessment through augmented visual interpretation, *Remote Sensing*, **8**, 807, https://doi.org/10.3390/rs8100807.

Bicheron P., Defourny P., Brockmann C., Schouten L., Vancutsem C., Huc M., Bontemps S., Leroy M., Achard F., Herold M., Ranera F., Arino O., (2008), GlobCover-Products Description and Validation Report. Toulouse (France), MEDIAS-France, JRC49240, On line at: https://publications.jrc.ec.europa.eu/repository/handle/JRC49240

Bieger K., Arnold J.G., Rathjens H., White M.J., Bosch D.D., Allen P.M., Volk M., Srinivasan R., (2016), Introduction to SWAT+, a Completely Restructured Version of the Soil and Water Assessment Tool, *Journal of the American Water Resources Association (JAWRA)*, **53**, 115-130.

Busari I.O., Demirel M., Newton A., (2021), Effect of Using Multi-Year Land Use Land Cover and Monthly LAI Inputs on the Calibration of a Distributed Hydrologic Model, *Water*, **13**, 1538, https://doi.org/10.3390/w13111538.

Buttner G., Feranec J., Jaffrain G., Mari L., Maucha G., Soukup T., (1998), *The European CORINE Land Cover Database*, Proceedings of ISPRS Commission VII Symposium, Budapest, September 1-4, 633-638.

Chirachawala C., Shrestha S., Babel M.S., Virdis S.G., Wichakul S., (2020), Evaluation of global land use/land cover products for hydrologic simulation in the Upper Yom River Basin, Thailand, *Science of the Total Environment*, **708**, 135148, https://doi.org/10.1016/j.scitotenv.2019.135148.

Cuceloglu G., Seker D.Z., Tanik A., Ozturk I., (2021), Analyzing Effects of Two Different Land Use Datasets on Hydrological Simulations by Using SWAT Model, *International Journal of Environment and Geoinformatics*, **8**, 172-185.

Ding B., Zhang Y., Yu X., Jia G., Wang Y., Wang Y., Zheng P., Li Z., (2022), Effects of forest cover type and ratio changes on runoff and its components, *International Soil and Water Conservation Research*, **10**, 445-456.

Dwarakish G., Ganasri B., (2015), Impact of land use change on hydrological systems: A review of current modeling approaches, *Cogent Geoscience*, **1**, 1115691, https://doi.org/10.1080/23312041.2015.1115691.

El-Sadek A., Irvem A., (2014), Evaluating the impact of land use uncertainty on the simulated streamflow and sediment yield of the Seyhan River basin using the SWAT model, *Turkish Journal of Agriculture and Forestry*, **38**, 515-530.

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**, 5613-5623.

Gassman P.W., Steenhuis T.S., Green C., Arnold J.G., (2007), The Soil and Water Assessment Tool: Historical Development, Applications, and Future Research Directions, *Transactions of the ASABE*, **50**, 1211-1250.

Georgakakos K.P., Baumer O.W., (1996), Measurement and utilization of on-site soil moisture data, *Journal of Hydrology*, **184**, 131-152.

Gibbs H.K., Brown S.A., Niles J.L., Foley J.A., (2007), Monitoring and estimating tropical forest carbon stocks: making REDD a reality, *Environmental Research Letters*, **2**, 045023, https://10.1088/1748-9326/2/4/045023.

Grekousis G., Mountrakis G., Kavouras M., (2015), An overview of 21 global and 43 regional land-cover mapping products, *International Journal of Remote Sensing*, **36**, 5309-5335.

Gupta H.V., Kling H., Yilmaz K.K., Martinez G., (2009), Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *Journal of Hydrology*, **377**, 80-91.

Huang J., Zhou P., Zhou Z., Huang Y., (2013), Assessing the influence of land use and land cover datasets with different points in time and levels of detail on watershed modeling in the north river watershed, China, *International Journal of Environmental Research and Public Health*, **10**, 144-157.

Jarvis A., Reuter H.I., Nelson A., Guevara E., (2008), Hole-filled SRTM for the globe Version 4, CGIAR Consortium for Spatial Information (CGIAR-CSI), On line at: http://srtm.csi.cgiar.org/

LaGro J.A., (2005), *Land-use Classification*, In: *Encyclopedia of Soils in the Environment*, Elsevier Academic Press, 321-328.

Li Z., Zhou P., Shi X., Li Y., (2020), Forest effects on runoff under climate change in the Upper Dongjiang River Basin: insights from annual to intra-annual scales, *Environmental Research Letters*, **16**, 014032, https://10.1088/1748-9326/abd066.

Loveland T.R., Reed B.C., Brown J.F., Ohlen D.O., Zhu Z., Yang L., Merchant J.A., (2000), Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data, *International Journal*, **21**, 1303-1330.

Luo J., Zhou X., Rubinato M., Li G., Tian Y., Zhou J., (2020), Impact of multiple vegetation covers on surface runoff and sediment yield in the small basin of Nverzhai, Hunan Province, *China, Forests*, **11**, 329, https://doi.org/10.3390/f11030329.

M'barek S.A., Bouslihim Y., Rochdi A., Miftah A., (2023), Effect of LULC data resolution on hydrological and erosion modeling using SWAT model, *Modeling Earth Systems and Environment*, **9**, 831-846.

McKay M.D., Beckman R.J., Conover W.J., (1979), A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics*, **21**, 239, https://doi.org/10.2307/1268522.

Monteith J.L., (1965), Evaporation and environment, *Symposia of the Society for Experimental Biology*, **19**, 205-234.

Moriasi D.N., Arnold J.R., 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**, 885-900.

Mücher C.A., Steinnocher K., Champeaux J.L., Griguolo S., Wester K., Heunks C., Vn Katwijk V., (2000), *Establishment of a 1-km Pan-European Land Cover Database for Environmental Monitoring*, Geoinformation for all, XIXth congress of the International Society for Photogrammetry and Remote Sensing (ISPRS), Int. Arch. Photogramm. remote Sens., vol. 33, 702-709.

Nash J.E., Sutcliffe J.V., (1970), River flow forecasting through conceptual models part I - A discussion of principles, *Journal of Hydrology*, **10**, 282-290.

Neitsch S.L., Arnold J.G., Kiniry J.R., Williams J.R., (2011), Soil and Water Assessment Tool Theoretical Documentation. Temple, Texas 76502: USDA-153 ARS Grassland Soil and Water Research Laboratory, and Texas A&M University, Blackland Research and Extension Center, On line at: https://www.ars.usda.gov/plains-area/temple-tx/grassland-soil-and-water-research-laboratory/

Oruç H.N., Çelen M., Gülgen F., Öncel S., Vural S., Kılıç B., (2022), Evaluating the effects of soil data quality on the SWAT runoff prediction Performance; A case study of Saz-Cayirova catchment, Turkey, *Urban Water Journal*, **20**, 1592-1607.

Peker İ.B., Gülbaz S., (2023), Examining the open-source datasets for water quantity and quality using the soil and water assessment tool (SWAT), *Water Science and Technology*, **88**, 1621-1634.

Romanowicz A., Vanclooster M., Rounsevell M., La Junesse I., (2005), Sensitivity of the SWAT model to the soil and land use data parametrisation: a case study in the Thyle catchment, Belgium, *Ecological Modelling*, **187**, 27-39.

Tan M.L. Gassman P.W., Liang J., Haywood J., (2021), A review of alternative climate products for SWAT modelling: Sources, assessment and future directions, *Science of the Total Environment*, **795**, 148915, https://doi.org/10.1016/j.scitotenv.2021.148915.

Tateishi R., Uriyangqai B., Al-Bilbisi H., Ghar M.A., Tsend-Ayush J., Kobayashi T., Kasimu A., Hoan N.T., Shalaby A., Alsaaideh B., Enkhzaya T.G., Sato H., (2011), Production of global land cover data - GLCNMO, *International Journal of Digital Earth*, **4**, 22-49.

Yan W.Y., Shaker A., El-Ashmawy N., (2015), Urban land cover classification using airborne LiDAR data: A review, *Remote Sensing of Environment*, **158**, 295-310.

Zhang H., Wang B., Liu D., Zhang M., Leslie L.M., Yu Q., (2020), Using an improved SWAT model to simulate hydrological responses to land use change: A case study of a catchment in tropical Australia, *Journal of Hydrology*, **585**, 124822, https://doi.org/10.1016/j.jhydrol.2020.124822.