T.R.
GEBZE TECHNICAL UNIVERSITY
INSTITUTE of EARTH and MARINE SCIENCES

EVALUATING THE EFFECTS of SOIL DATA QUALITY on THE SWAT MODEL RUNOFF PREDICTION PERFORMANCE in A HIGHLY INDUSTRIALIZED CATCHMENT; CASE STUDY in SAZ-ÇAYIROVA STREAM, TURKEY

HALİL NURULLAH ORUÇ
THESIS SUBMITTED for THE DEGREE of MASTER of SCIENCE
DEPARTMENT of EARTH and MARINE SCIENCES

GEBZE
2021

T.R.
GEBZE TECHNICAL UNIVERSITY
INSTITUTE of EARTH and MARINE SCIENCES

EVALUATING THE EFFECTS of SOIL DATA QUALITY on THE SWAT MODEL RUNOFF PREDICTION PERFORMANCE in A HIGHLY INDUSTRIALIZED CATCHMENT; CASE STUDY in SAZ-ÇAYIROVA STREAM, TURKEY

HALİL NURULLAH ORUÇ
THESIS SUBMITTED for THE DEGREE of
MASTER of SCIENCE
DEPARTMENT of EARTH and MARINE SCIENCES

THESIS SUPERVISOR
Asst. Prof. Dr. MELTEM ÇELEN

GEBZE
2021

T.C.
GEBZE TEKNİK ÜNİVERSİTESİ
YER ve DENİZ BİLİMLERİ ENSTİTÜSÜ

ENDÜSTRİYELLEŞTİRİLMİŞ BİR HAVZADA TOPRAK VERİ KALİTESİNİN SWAT MODELİNİN AKIŞ TAHMİN PERFORMANSI ÜZERİNDEKİ ETKİLERİNİN DEĞERLENDİRİLMESİ; SAZ-ÇAYIROVA DERESİ ÖRNEĞİ, TÜRKİYE

HALİL NURULLAH ORUÇ
YÜKSEK LİSANS TEZİ
YER ve DENİZ BİLİMLERİ ANABİLİM DALI

DANIŞMANI
Dr. Öğr. Üyesi MELTEM ÇELEN

GEBZE
2021

GEBZE TEKNİK ÜNİVERSİTESİ	**YÜKSEK LİSANS JÜRİ ONAY FORMU**

GTÜYer ve Deniz Bilimleri Enstitüsü Yönetim Kurulu’nun …..…/…..…/..…..… tarih ve ..…../…… sayılı kararıyla oluşturulan jüri tarafından 25/05/2021 tarihinde tez savunma sınavı yapılan Halil Nurullah ORUÇ’un tez çalışması Yer Bilimleri Anabilim Dalında YÜKSEK LİSANS tezi olarak kabul edilmiştir.

JÜRİ

ÜYE
(TEZ DANIŞMANI) :

ÜYE :

ÜYE :

ONAY

Gebze Teknik Üniversitesi Yer ve Deniz Bilimleri Enstitüsü Yönetim Kurulu’nun
…..…/…..…/..…..… tarih ve ………/……... sayılı kararı.

İMZA/MÜHÜR

SUMMARY

In this thesis study, the flow rate prediction performance of the model established by using the Specified Soil Database (SSM) formed by laboratory and field studies and the international Food and Agricultural Organization (FAO) soil databases as soil input layer for the Soil and Water Assessment Tool (SWAT) model was evaluated.

Twenty-nine soil samples were collected representing the entire Saz-Çayırova. 7 soil physicochemical properties were analyzed out of 13 parameters. These parameters were configured in the GIS environment and the SSM database consisting of 15 different soil combinations was created. Differences between FAO and SSM are depth, hydraulic conductivity and texture values can be explained by the fact that the FAO database accepts average values for the whole of the Marmara Basin.

Calibration and validation processes were applied for the the catchment outlet. Nash Sutcliffe Efficiency (NSE) value calculated between the predicted dataset produced using SSM and observation runoff was in acceptable range (0.44) even before calibration, while this value was below the threshold value (0.11) for FAO dataset. NSE were found as 0.66 and 0.35 for the models established with SSM and FAO, respectively. In this context, predictive performance of model established using SSM were seen to be acceptable for the Saz-Çayırova catchment.

As a result of this thesis study, it has been concluded that even though the preparation of model input parameters and calibration processes required high labor intensity, the soil databases should be specified with laboratory and field studies for watersheds dominated by industrial or residential areas.

Key Words: SWAT, FAO, Surface Runoff

ÖZET

Bu tez çalışmasında, Toprak ve Su Değerlendirme Aracı (SWAT) modeli için girdi olarak laboratuvar ve arazi çalışmaları ile oluşturulan spesifiye edilmiş toprak veritabanı (SSM) ve uluslararası Food and Agricultural Organization (FAO) toprak veritabanları kullanılarak kurulan SWAT'ın yüzeysel akışı tahminleme performansı değerlendirilmiştir. Bu doğrultuda, çalışma alanı olarak endüstriyel alanlar ile domine edilmiş Saz-Çayırova deresi seçilmiştir.

SSM veritabanının hazırlanması için havzanın tamamını temsil eden 29 toprak örneği alınmış ve SWAT için gerekli 13 toprak parametresinden 7 tanesi analiz edilmiştir. Bu parametreler GIS ortamında yapılandırılarak 15 farklı toprak kombinasyonundan oluşan SSM veritabanı oluşturulmuştur. SWAT modeli çalıştırılmadan önce SSM ile FAO veritabanı sonuçları karşılaştırılmıştır. Toprak tekstürünü oluşturan parametreler FAO veritabanında yaklaşık eşit yüzdelikteyken, SSM veritabanında toplam kum ve silt içeriğinin %78 olması iki toprak veritabanı arasındaki en dikkat çekici farktır.

SSM veritabanı sonuçları istatistik performans kriterleri göz önüne alınarak kalibrasyondan önce bile kabul edilebilir (NSE=0.44) seviyede bulunmuş iken, FAO veritabanı ile kurulan modelde akış tahmin performansı kabul edilebilir seviyenin altında bulunmuştur (NSE=0.11). Kalibrasyon işlemlerinden sonra, havza mansabında SSM veritabanı kullanılarak kurulan modelde FAO ile kurulan modele nazaran daha iyi sonuçlar bulunmuştur. SSM ve FAO ile kurulan model akış tahminleri için NSE değerleri sırasıyla 0.66 ile 0.35 bulunmuştur. Bu durum, noktasal kaynakların tanımlanmadığı ve endüstriyel alanlar ile çevrili Saz-Çayırova havzası için yeterli aralıkta bulunmuştur.

Bu tez çalışması sonucunda, endüstriyel alanlar ile domine edilmiş havzalarda modelin kurulumu ve kalibrasyon işlemleri sırasında yüksek emek yoğunluğu gerektirmesine rağmen toprak veritabanlarının laboratuvar ve saha çalışmaları ile spesifiye edilmesi gerektiği bulunmuştur.

Anahtar Kelimeler: SWAT, FAO, Yüzeysel akış

ACKNOWLEDGEMENTS

I would like to thank TUBITAK for financially support with 119Y032 TUBITAK Project and Gebze Technical University Scientific Research Projects Directorate (No: 2017 A105-54)

I would like to express my special appreciation to my advisor Asst. Prof. Meltem Çelen for her guidance on both research as well as on my career have been invaluable.

I would like to deeply thank Professor Dr. Salim Öncel for his help and suggestions in field and laboratory studies. I'd also like to thank thesis monitoring committee members Associated Professor Dr. Fatih Gülgen for their advice.

I appreciate all the support I have received from my colleagues in Institute of Earth and Marine Sciences and Gebze Technical University. I'd like to special thank Env. Engineer Sinem Vural for her help in field and laboratory studies. I would also like to thank Muhsin Mahmoody Vanolya and Dr. Micheal Yu for their contribution in specification of model input parameters and visualization.

I'd like to thank my family, my mother Zerrin Oruç, my father Selami Oruç, and my sister Zübeyde Zoroğlu for their support and believing in my ability. I'd like to deeply thank my wife Elif Oruç for her support and patience. I am forever grateful.

TABLE of CONTENTS

LIST of ABBREVIATIONS and ACRONYMS

Abbreviation and Acronyms		Explanations
CORINE	:	Coordination of Information on the Environment
DEM	:	Digital Elevation Model
FAO	:	Food and Agricultural Organization
GIS	:	Geographical Information System
NSE	:	Nash–Sutcliffe efficiency
SSM	:	Specified Soil Map produced by laboratory and field studies
SUFI-2	:	Sequential Uncertainty Fitting
SWAT	:	Soil and Water Assessment Tool
TNSD	:	Turkey Ministry of Agriculture and Forest
UNEP	:	United Nations Environment Program

LIST of FIGURES

LIST of TABLES

1. INTRODUCTION

In this thesis, the hydrological cycle of the Kocaeli Saz Stream Catchment is modeled using the Soil and Water Assessment Tool (SWAT) model. The soil database which is obtained from field and laboratory studies and soil database of The Food and Agriculture Organization (FAO) is used in order to compare the effect of both soil databases on hydrological processes. In this chapter, the theoretical background of the thesis study is presented. In this context, water resources over the World and Turkey, the hydrological modeling and the aim of the thesis were included in Chapter 1.1, in Chapter 1.2 and in Chapter 1.3 respectively.

1.1. Water Resources in World and Turkey

Water is a natural resource with no alternative for life. Water is one of the most principal parts of the ecosystem where living things interact with each other, and consequently the sustainability of the ecosystem depends on water. In addition to meeting fundamental needs, water is the source of social and economic developments such as agriculture, energy production, industry and transportation [1]. For this reason, the first settlements were chosen close to the water. Urbanization, developments in industry and agriculture that are growing with the population increase the demand for water. With these developments, environmental pollution problems have also arisen. Therefore, the whole world has accepted that the quantity of water as well as its quality should be preserved [1], [2].

One in three people in the world lives in regions with insufficient water resources. Approximately 20% of the world's population cannot reach clean drinking water and 50% lack sufficient quality water resources [2]. The United Nations Environment Program (UNEP) states that there is 1.4 billion km^3 of water in the world, but a very small proportion of this amount is available [2]. More than 3 billion people are expected to suffer from water shortage in 2025, according to UNEP. By 2050, it is estimated that 40% of the world's population, which is expected to be over 9 billion, will suffer from water shortage [1]. Only 2.4% of the world's water is freshwater. In these fresh water sources, 68.6% are in glaciers, 30.1% are groundwater and the remaining 1.3% are in rivers, lakes and atmosphere [1], [2].

There are many criteria in determining and classifying the water presence of countries in terms of usable water per capita per year: poorest countries in terms of water resources (<1000 m^3), water shortage (<2000 m^3) and water rich countries (8000 - 10000 m^3), respectively [1]. The water resources on earth and the distribution of the population by continents have no obviously linear relationship (Figure 1.1). Asian continent (highest population) constitutes 60% of the total world population, while it has only 36% of the total fresh water resources. In contrast, North and South America have 14% of the world's population, while 41% of freshwater resources. While the European and African continents have approximately 14% of the world's population, they have 8% and 11% of the world's water resources, respectively. The ratio of Australia and islands to the world population is 1% and its share in water resources is 5%. Countries with high water resources and low population density are water rich.

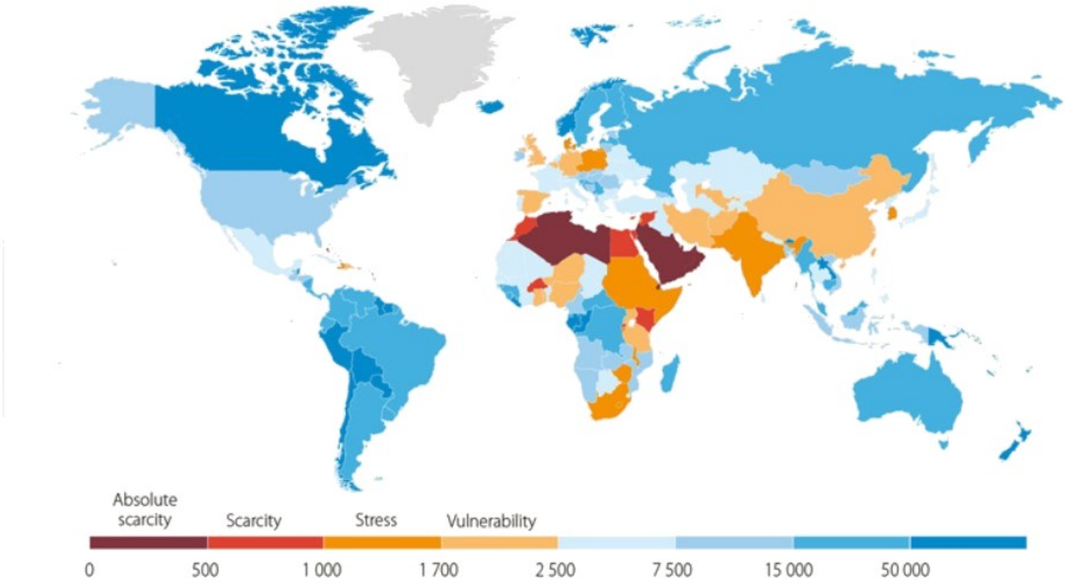


Figure 1.1: Total renewable water resources (m^3 per capita per year).

The average rainfall of Turkey is 643 mm per year, which corresponds to an average of 501 km^3 of water per year [3]. Approximately 274 km^3 of this water returns to the atmosphere through evaporation, 69 km^3 of average rainfall feeds groundwater, rest of the rainfall (158 km^2) flows into the seas and lakes in basins [4]. With 41 km^3 of water feeding aquifers and 7 km^3 of water coming from transboundary streams, Turkey's total water potential is 234 km^3. Only 44 km^3 out of 112 km^3 of annual average ground water potential can be used [4], [5].

Turkey is not a rich country in terms of available water, it is one of the countries that may experience water shortages in the future. Amount of water available per capita in Turkey, which was over 3000 m^3 about twenty years ago, has fallen to around 1400 m^3 in recent years [5]. The General Directorate of State Hydraulic Works (DSİ) indicates that Turkey, whose population will reach 85 million by 2030, will become a water-deprived country with 1,100 m^3 of usable water per capita (DSİ, 2015). Given these data, between 2050 and 2100, it is inevitable that Turkey will struggle with a very serious water crisis. Therefore, identifying, protecting and managing restricted water resources in a sustainable manner has become highly important [5].

1.2. Hydrological Modeling

Although the general structure of the hydrological cycle is easy to understand and interpret, it is just as difficult to measure and analyze the processes in the system. How these processes respond to land use/land cover change and the characteristic soil structure of the region is very important in terms of land and water resources management. Topographic conditions, land use / land cover, climatic conditions and anthropogenic activities etc. are among the main factors affecting the hydrological process. The temporal and spatial structure of the hydrological process can be better understood by modeling studies in which these factors are considered together.

Hydrological models can be used for many purposes. For example, evaluation of river quality, pollutant transport and hydrodynamics, hydrology and quality of lakes, wetlands and dam lakes, groundwater hydrogeology; management of flood, drought and watershed; runoff-volume simulation along the channel network, and the estimation of the flow at the watershed output [6].

Watershed-based hydrological modeling has two main objectives; the first is to understand the functioning of the system and the effects of changes in the system, the second is to produce time series for prediction and designs. On the other hand, the objectives of hydrological modeling can be detailed as follows in the conclusion of the studies about water management [7].

- Forecasting the data which needed for the planning, operation and development of streams or dams made to increase the usability of water resources.

- Generation of hydrological data in watershed where measurement data is insufficient,
- To reveal the effect of land use change and climate change in catchments,
- Calculation, managing and estimating of water budgets to ensure water allocation on a sectoral basis,
- Determination and estimation of runoff to be used in mass balance research for the assessment of water quality in surface waters,
- Hydrological modeling studies are carried out for estimation of hydrological events such as drought and flood, which are considered as natural disasters.

Many Geographical Information System (GIS) technologies and many physically based models have been developed to serve the purposes listed above. SWAT [7], WEAP (Water Evaluation and Planning System) [8], HEC-HMS (Hydrologic Engineering Center, Hydrologic Modeling System) [9], MIKE SHE [10], HSPF (Hydrologic Simulation Program – Fortran) [11], AGNPS (Agricultural Non Point Source Model) [12], HEC-RAS (Hydrologic Engineering Centers River Analysis System) [13] can be given as examples among the most preferred models worldwide. In general, models can be classified as physical / mathematical, stochastic / deterministic, static / dynamic, optimization / simulation, and rounded parameters / scattered parameters [2]. The SWAT model which is used in this thesis is a deterministic, dynamic, semi-scattered parameter simulation model. SWAT was detailed in Chapter 3.2.

1.3. Aim of Thesis

For the hydrological modeling, some soil parameters such as the number of soil layers, profile depths, organic matter content, hydraulic conductivity, available water holding capacity were generally calculated or accepted according to the default values in the literature. However, these values were generally based on foreign studies and cannot reflect the unique characteristics of catchment. For this reason, conducting field and laboratory analyzes and literature review to reveal unique soil characteristics will enable a more accurate decision about the soil condition of the studied catchment. In the light of the aforementioned information, comparing high labor-intensive soil analyzes to international soil databases such as FAO, and identifying the similarities or differences will help researchers better understand the specific hydrological characteristics of the studied catchment.

The main purpose of this study is testing the preferability of the soil database of FAO generally used as input parameters in hydrological modeling studies in Turkey on small-scale and industrialized catchments. For this purpose, the goals determined within the scope of the thesis study are listed as follows;

- Hydrological Model establishment for Saz-Çayırova Stream, located in the north-western part of Turkey and under industrial pressure, using SWAT.
- Determining the soil properties in the catchment by laboratory and field studies,
- Evaluating the effect of FAO and soil database produced from this study on modeled streamflow,
- Revealing the potential of two different soil databases to estimate the surface flow using the SWAT-CUP semi-automatic calibration technique.

The novelty of this study is interpreting the response of hydrological models to elaborating soil database for small and industrialized catchments. It is believed that the effect of elaborating the soil parameter on hydrological model performance in such catchments will make an important contribution to the literature.

2. LITERATURE REVIEW

2.1. Studies in Saz-Çayırova Stream Catchment

The discharge gauging station of the General Directorate for Hydraulic Works (GDHV) on the Saz Stream is currently closed, where flow rate measurements were made between 1982 and 1986. In addition, some stream rehabilitation works were carried out by the Kocaeli Water and Sewerage Administration (ISU) for the stream.

A few water quality and land use studies were carried out in Saz Stream catchment. Atasayan et al., (2013) investigated the change in the natural landscape by detecting the changes in the natural vegetation of the Saz Stream protection band for the period between 2003 and 2013 through satellite images [14]. Various water quality studies were carried out in order to analyze discharge of industrials to the Saz Stream. Heavy metal concentrations were determined at 15 points by Hız (2000) and it was stated in the study that the relationship between the examined parameters is complex and these relationships cannot be explained by binary correlations [15]. The study conducted by Kanburoğlu (2002) was examined that copper, iron, cadmium metals affect the Vida faba L. (pods) chromosomes [16].

Six master's thesis were generated on the website of the Turkey Higher Education Institution for Saz Stream. These studies generally examined the chemical and biological structure of the water in the stream. There are no studies for analyzing and modeling parameters affecting hydrological processes of the stream.

2.2. Hydrological Modeling Studies Using SWAT

Many studies have been conducted with the SWAT model to test the model performance in large and small-scale basins in Africa, America, Europe and the Far East. For instance, Rossi et al., (2009) conducted a hydrological modeling study in the Lower Mekong River watershed, which has 630,000 km^2 area. The model was calibrated and validated for the years between 1985 and 1992, and between 1993 and 2000, respectively. Nash-Sutcliffe value calculated for predicted and observed surface runoff was found as 0.8 and 1. This study draws attention to the need for powerful computer hardware [17].

Güngör and Göncü (2013) carried out a hydrological modeling study to determine the best water management strategies in the Lower Porsuk Stream basin. The modeling period was between 1978-2009. Acceptable results were obtained in the calibration and validation processes and irrigation scenarios were applied in the watershed. In this scenario, it was assumed that the gutter irrigation system was switched to the drip irrigation system, and the change in the stream flow was determined for 1 year. The result of this scenario has revealed that the amount of surface water will increase by approximately 87% during this year [18].

The study carried out by Pott and Fohrer (2017) investigated the hydrological processes using SWAT in a small-scale rural basin located north part of Germany. The model results were found to be good for both in the calibration and validation steps of the study. The results show that 52.8% of the water budget was base flow, 7.7% was subsurface flow, 34.3% was drainage flow and the remaining was surface flow [19].

Di Luzio et al., (2005) conducted a comprehensive study to determine the effects of different GIS database on runoff and sediment. A small-scale basin (21.3 km^2) has been selected to determine these effects more clearly. Two different DEMs, three different land use maps and two different soil maps were used and 12 different database combinations were used in this study. Firstly, they found different DEMs cause small differences in the sub-basin boundaries. Secondly, soil database and land use maps have a significant effect on runoff and sediment efficiency. It was stated that new researches should be conducted in different watersheds worldwide in order to elaborate these effects [20].

In a study conducted by Güzel (2010), hydrological processes of Köyceğiz Dalyan Basin were examined using SWAT. The model was operated between 1976 and 2008, and monthly precipitation, runoff, subsurface flow, groundwater flow values were modelled. One of the most critical outputs in this study is that irrigation water used in agricultural activities feeds the groundwater flow [21].

Göncü and Güngör (2013) modeled the Porsuk Basin using SWAT. The model was operated between 1978 and 2004, calibrated between 1998 and 2004 and validated for all years. It was aimed to decide on the most appropriate water management strategy by applying different irrigation scenarios to the model [18].

Akiner and Akkoyunlu (2012) produced missing and future precipitation data for the Melen Basin using Artificial Neural Networks technique. This data set is integrated into the SWAT model, and future runoff was modeled. The results showed that water transfer from Melen Basin to Istanbul is a suitable water management practice [22].

Studies examining the soil physical and chemical properties in various basins in Turkey are existing. For instance, Dengiz and Başkan (2007) examined conventional geological materials in their study to determine the relationship between soil profile development and land shape in Ankara Soğulca Basin [23].

Aydınalp and Arslan (2003) aimed to classify main soil groups in the Western Black Sea Basin according to international soil classification systems. The physical and chemical properties of the soil were determined by examining 10 profile depths of 7 main soil groups [24].

Dengiz et al., (2007) analyzed the current land use, topographic, geological and geomorphological maps of the Büyükçay Basin in the GIS environment and conducted field observations in order to reveal the soil characteristics to provide information regarding the planning of the basin. In this study, 10 soil profile depths were examined [23].

In literature studies, different soil databases such as State Soil Geographic database (STATSGO), Soil Survey Geographic database (SSURGO), soil map of Food and Agricultural Organization (FAO) were examined in order to evaluate the effect and prediction performance of models over the surface runoff [25], [26], [27].

Levick et al., (2004) compared effects of FAO, STATSGO and SSURGO soil databases in the Walnut Gulch Watershed in Arizona. Higher surface runoff values were estimated using the STATSGO database compared to SSURGO. In addition, about two times more streamflow was estimated using FAO compared to SSURGO [25]. The difference in streamflow is explained by the difference in data resolution. SWAT was mostly used in order to examine the effects of these mentioned soil databases on hydrology [28], [27].

Peschel et al., (2003) investigated the effect of soil database resolution over surface runoff in the Leon basin, which is located in Texas [29]. The total surface runoff obtained using SSURGO soil database was found to be higher than the STATSGO database, unlike the study conducted by Levick et al., (2004). High runoff values were attributed to higher resolution; even though the study did not draw any firm conclusion.

Lenhart et al., (2002) compared the two soil databases having different resolutions, by artificial simulation method, in a small-scale basin and found that the hydrological model results are relatively sensitive to the physical properties of the soil such as bulk density, available water capacity, soil texture and saturated hydraulic conductivity [30].

Setegn et al., (2010) generated the soil structure of the Upper Mississippi River Basin by field and laboratory studies and created its hydrological budget using the Soil and Water Assessment (SWAT) model. The critical point in this study is that no calibration and validation processes have been performed for the model. It was emphasized that similar results were obtained when compared with the results of three different past studies performed for the same basin, and that the model yielded acceptable results for hydrological components without calibration [26].

Sajikumar et al., (2015) studied the impact of land use change on runoff of two basins in Kerala India, using the SWAT model. The effect of land use on runoff has been clearly observed by changing the forest land suitable for agriculture. In this study, it was emphasized that storing rainfall for agriculture, and growing plants with high evaporation potential are important factors affecting runoff [31].

A study conducted by Li et al., (2012) in the upper and middle parts of the Taoerhe River basin, which is located in the northeast part of China, was aimed to determine the effects of land use on monthly and annual runoff. Study was carried out under humid, dry and moderate climatic conditions using the SWAT model. As a result of the application, the land use change has a secondary effect on surface runoff estimation, where rainfall has a very important role and effect on climatic conditions. In addition, it has been determined that the effect of changes in climate conditions and land use on monthly runoff is greater during rainy periods [32].

One of the most remarkable studies using the SWAT was carried out by Iskender and Sajikumar (2015). In this study, the surface runoff prediction performance of the SWAT model in the Manali Basin located in the south of India lacking observation data was evaluated. According to the statistical performance criteria of the model, realistic results could not be obtained in the SWAT. However, it was emphasized that more realistic results can be achieved by modifying the topographic and climate data in the long term period [31]. It is also stated that this situation can be solved by using the Green & Ampt method and integrating hourly rainfall and flow data into the model [33].

3. MATERIAL and METHOD

3.1. Study Area

Saz - Çayırova river system is located in the Marmara Basin, northwest parts of Turkey (Figure 3.1). Saz-Çayırova Stream which is heavily affected by environmental factors and runs along the Kocaeli-Istanbul border, located in one of the most important industrial regions of Turkey, has been selected as the study area. The length of the Saz stream is about 10 km and its width varies between 2-20 m. The catchment area is approximately 50 km^2 and is classified as a small-scale catchment [14]. Stream network was obtained from the on-going Scientific and Technological Research Council of Turkey (TUBITAK) Project (No:119Y032). The stream network has been digitized using aerial photographs and complied to the GIS environment in this project. In addition, the networks are taken as reference while determining sub-basins in the watershed delineation step in the SWAT model. The study area consists of two main sub-basins. These are Saz stream sub-basin (20 km^2) and Çayıtova creek sub-basin (30 km^2). According to the results of water quality monitoring studies, pollution load from Çayırova Stream is higher than the pollution load from Saz Stream.

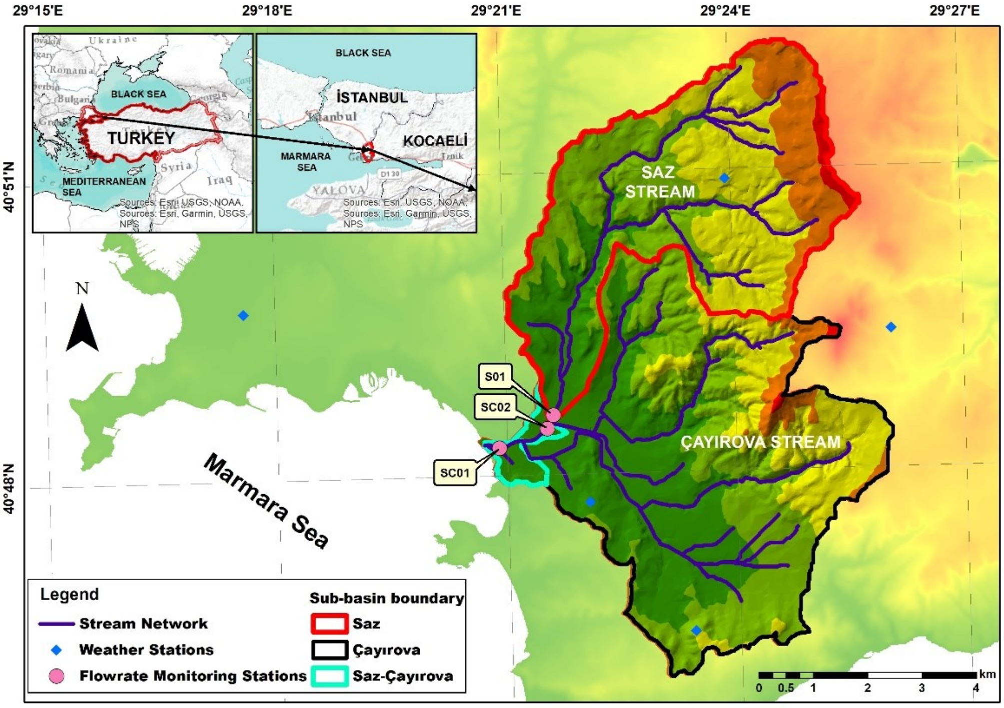


Figure 3.1: Location map of Saz - Çayırova river system.

The Saz- Çayırova Stream network is mostly included in the industrial and residential areas and it has become difficult to detect these networks visually without using Geographic Analysis Tools. Gebze Technical University, two different dense residential areas (Şekerpınar, Güzeltepe, Gebze) and continuous/discontinuous industrial areas are located within the catchment.

3.2. The SWAT Model

The SWAT (Soil and Water Assessment Tool) model is an efficient method of evaluating water sources and nonpoint source pollution with its many applications at different scales and environmental conditions, [11], [12], [34]. The SWAT is a hydrology and water quality modeling tool that efficiently reveals the effects of water, sediment and pollution sources in watershed management. The SWAT is an advanced model and compatible with GIS (Geographic Information Systems) software.

The SWAT was developed by USDA (United States Department of Agriculture) and ARS (Agricultural Research Service) which are funded from EPA (Environmental Protection Agency), and incorporated into EPA BASINS 3.0 water quality modeling system (Figure 3.2). Background of the SWAT model was firstly established to be used in combination with CREAMS (Chemicals, Runoff, and Erosion) to examine the impact of land management on water, sediment, plant nutrients and pesticides, which are directly separated from a piece of land in rural basins [7]. Subsequently, the Groundwater Loading Effects on Agricultural Management Systems (GLEAMS) [35] model was developed to determine the effects of groundwater chemicals on basins dynamics [36]. In the next step, EPIC (Erosion Productivity Impact Calculator) [37] was developed to simulate the effects of erosion on plant productivity. The latest SWAT version, established for modeling agricultural lands, has a similar structure to Water Resources Simulator (SWRRB) in rural basins [7]. But, the SWRRB model only allows simulation in 10 sub-basins. However, the SWAT was created to evaluate flow kinetics and allow more detailed simulations of large-scale basins [7].

The SWAT model can model hydrological processes such as surface runoff, infiltration, percolation, evapotranspiration, lake and reservoir storage, groundwater flow discharge, and plant nutrients and pesticide loads. Daily or long term simulations can be made for the separate multiple sub-basins. The input data are meteorological, topographic, soil, vegetation, and land use data of the studied watershed. The modeled watershed is divided into unique hydrological units called hydrological response units (HRUs), which differ according to drainage areas and input parameters. The process is conducted for each HRU [38].

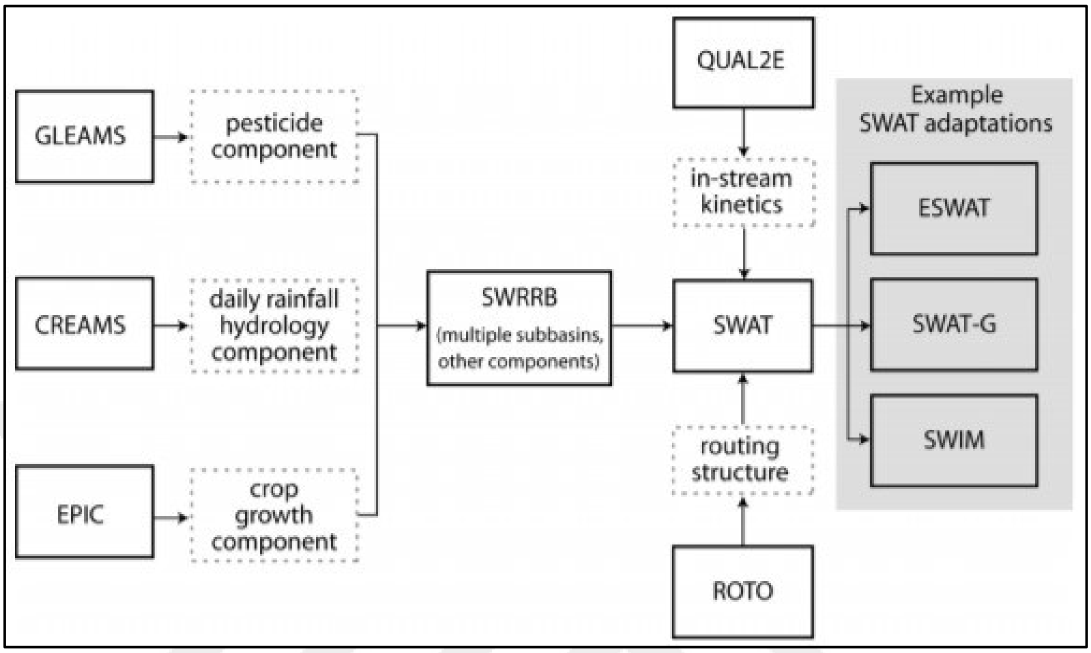


Figure 3.2: Historical development of SWAT.

The background of the SWAT model is mainly based on water balance. Water balance is a complex combination and interaction between hydrological runoff such as precipitation, evapotranspiration, surface and groundwater flow discharge. The details of the SWAT model hydrological processes were given below (Figure 3.3). Properly modeling of these hydrological components allows researchers to examine processes such as erosion, plant growth, and nutrient cycling. In the SWAT model, the surface and groundwater flow reaching the river and the sea can be modeled separately. SWAT can model shallow aquifer systems for groundwater. While the SWAT model contributes directly to the surface (shallow) runoff, it accepts that the effect of the free surface aquifer (deep) system is outside the basin. In this respect, the surface flow calculated for each sub-basin in the SWAT model consists of three sections. 1) Surface flow, 2) Sub-surface flow, 3) Groundwater flow. However, the amount of infiltration that can reach the deep aquifer is neglected.

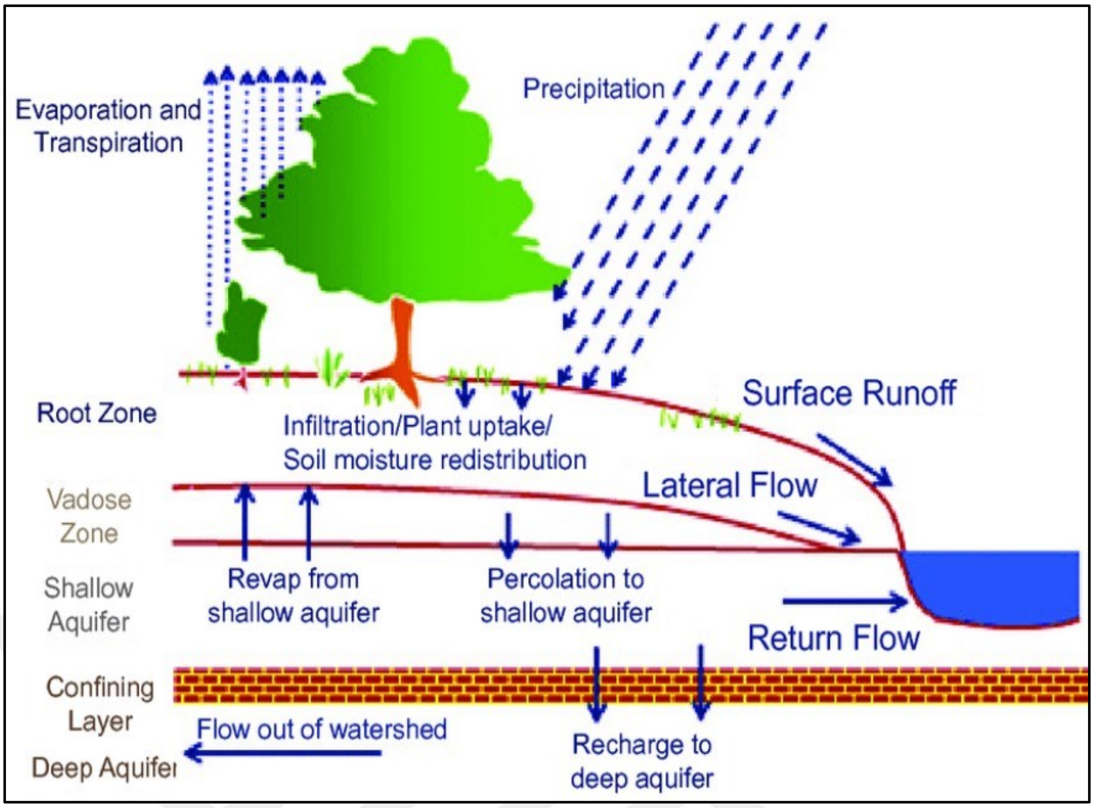


Figure 3.3: Schematic representation of the hydrologic cycle.

In the SWAT model, calculations take place in two steps. The first step is the precipitation-flow distribution depending on the water mass balance for each sub-basin [39].The second is the transportation section that connects all sub-basins to determine a positionally open outlet of watershed. The following water balance equation (Equation 3.1) is used in SWAT [34].

$$SW_t = SW_0 + \sum_{i=1}^{t}\left(R_{day} - Q_{surf} - E_a - W_{seep} - Q_{gw}\right) \tag{3.1}$$

Where SW_t is the final soil water content (mm H_2O), SW_0 is the initial soil water content on day i (mm H_2O), t is the time (day), R_{day} is the amount of precipitation on day i (mm H_2O), $Qsurf$ is the amount of surface runoff on day i (mm H_2O), E_a is the amount of evapotranspiration on day i (mm H_2O), $Wseep$ is the amount of water entering the vadose zone from the soil profile on day i (mm H_2O), and Q_{gw} is the amount of return flow on day i (mm H_2O).

The inputs for the model are; Digital Elevation Model (DEM), Land Use/Land Cover, Soil Properties, Meteorological, Management Data.

3.2.1. Digital Elevation Model (DEM)

A digital elevation model is required to define the watershed. DEM is also used to calculate sub-basin parameters such as slope, slope length and characterize stream network properties, i.e. channel slope, length and width [40]. The Grid resolution unit must be in meters, kilometers, feet, yards, miles or in decimal degrees, whereas the altitude unit is meters, centimeters, yards, feet or inches.

3.2.2. Land Use and Land Cover

Land Use/Land Cover data is one of the most important GIS data used in the model. Land covers are important for calculating different flow, nutrient loads and erosion rates [40]. Land use / land cover change also indirectly affects economically and socio-economically, as well as affecting biodiversity, climate change and global warming. Remote sensing and Geographic Information Systems (GIS) techniques are frequently used for analyzing land use / land cover changes in terms of present and past [41]. In present, more detailed and accurate data are obtained in terms of temporal and spatial dimensions by handling advanced GIS techniques, radar and aerial photographs, high resolution satellite images and image processing techniques in an integrated structure.

For the SWAT model, the land use / land cover map must be ESRI GRID, in shape file or in feature class formats. The categories selected in the land use and land cover map should be reclassified into the SWAT database of land cover / plant species. The user has three options to reclassify the land use and cover categories:

- Creating land use / land cover lookup table in ArcSWAT interface (interface includes USGS LULC and NLCD 1992 lookup tables).
- Writing SWAT land use / land cover codes for each category.
- Creating a user lookup table defining SWAT codes for different categories.

CORINE (Coordination of Information on the Environment) is the land use / land cover data determined by the European Environment Agency. This dataset was produced by computer-aided visual interpretation method on satellite images [42]. The CORINE Project has been executed by 13 countries. The dataset reflecting 1990, 2000, 2006, 2012, 2018 years have been created for the European Union member countries and Turkey [43], [44]. In Turkey, the CORINE Project was started by Turkey

Statistical Institute (TSI) in 2001. This project has been carried out by the Republic of Turkey Ministry of Agricultural and Forestry since 2005. In this study, the CORINE-2018 (Coordination of Information on the Environment) land use/land cover classification was used.

3.2.3. Soil Data

One of the most important input data for hydrological modelling is the digitized soil maps of the studied area. Soil maps that are produced as a result of soil etude and mapping studies and related reports presented constitute the soil database. The accuracy of the reports and detailed information provide acceptable results for future studies [5].

The terms Genetic or Taxonomic classification and Soil Taxonomy refers to the natural classification of soil types. These systems examine soils according to their natural properties and compare them with each other. The systems reveal the relationships of the most important natural properties of soils without any practical and applied purpose. The main natural classification systems are Russian, German, French, Australian and the old American classification systems. In addition to these, The New American Soil Taxonomy and FAO/UNESCO soil classification system are used in scientific research all over the world.

The digital soil map of Turkey is classified according to the Old American Soil Classification System in Turkey since 1958. This map is scaled of 1/25 000 and compatible with GIS. Users can access data about; Main soil groups (23 types), soil characteristics, other soil characteristics (salinity, alkalinity, stony, rocky etc.), erosion level sensibility (1-non or very low, 2- medium, 3- high, 4-very high), area types (bare rocks, rubble areas, ricer flooded areas, sandy shores, reed bed on muddy areas, permanent snowed areas), groups of land capability classes (lands suitable or not suitable for ploughing, lands not suitable for agriculture etc.) and sub groups of land capability classes (non-capability due to slope, poor drainage, flooding or wetness, climate restriction etc.) in soil map of Turkey.

The soil map must be in the ESRI GRID, shapefile or feature class format and should cover at least 95% of the simulated area. The categories specified in the soil map had to be linked to the SWAT soil database. The users can add soil types and properties to the SWAT soil database using the ArcSWAT editing database tool or by importing SWAT soil files (.sol). The users have four options for linking the soil map and the soil database [34]:

- Using a STATSGO polygon (MUID) number: The SWAT soil database contains information about all soil physicochemical properties for all polygons in the USA. This option (Stmuid) uses data for the dominant soil type for the map category.
- Using the Stmuid + Seqn option: In this option, the user can choose another soil type from the dominant soil types in the MUID.
- Using the name option: The model allows the users to specify the user soil type name specified in the soil map. In this case, the user must import the SWAT soil file (.sol) or write the soil data to the soil database.
- Using the S5id option: If the S5id option is selected, data from the specified soil series is used to represent the map unit. To use this option, a US soil database must be installed.

More Detailed information on the integration of these soil data into the SWAT model is described in the Chapter 5.

3.2.4. Meteorological Data

Obtaining the right meteorological dataset is a vital step for the simulations. Rainfall and temperature data are the main meteorological parameters for the SWAT model. In addition, solar radiation, wind speed, relative humidity parameters can be produced by the model using precipitation and temperature data or the user can provide these values into the model. Gage stations should be selected to fully represent the studied watershed. Latitude, longitude and altitude values of the stations should be introduced to the model. Unlike other hydrology models, the SWAT model can interpolate missing data.

The SWAT model needs 6 different meteorological data in daily resolution. These are; precipitation, minimum temperature, maximum temperature, solar radiation, relative humidity and wind speed. Users can introduce or obtain this data to the model by providing it on a daily basis, or they can also generate data with the help of the WGEN [45] model which is embedded in the SWAT using a few years of average monthly data.

The meteorological parameters required for SWAT can be obtained from Turkish State Meteorological Service (TSMS) in Turkey. Measurement data in hourly and daily resolution available from TSMS can be used for the user's data input section.

3.2.5. Management Data

The main purpose of watershed-based hydrological models is to observe the effect of anthropogenic activities on a system. Land and water management activities are important as the basis of this assessment. The SWAT management option is available in an HRU. Operation data such as planting, harvesting, irrigation applications, nutrient applications, pesticide applications, and tillage can be obtained by researchers from the local farmers.

The management file is divided into two parts. First of all, the initial conditions or management practices that never change during simulation are summarized. The second part contains a list of management processes that take place at specific times.

General management variables, including initial conditions, are listed below:

- Initial plant growth parameters
- General management parameters
- Urban management parameters
- Irrigation management parameters
- Tile drain management parameters
- Management operations

Scheduled management operations are mainly; Planting/beginning of growing season, irrigation operation, fertilizer application, pesticide application, harvest/kill operation, tillage operation, grazing operation, auto irrigation and fertilizer initialization, street sweeping operation, release/impound operation, continuous fertilizer operation, end of year operation. All the management options listed above are described in detail in the SWAT input / output file document [34].

3.2.6. Hydrological Response Units (HRUs)

The paths of the water in the study catchment are determined by simulating the SWAT model [39] owing to the hydrological response units. The Hydrological Response Units (HRUs) are produced by overlay analysis in the GIS environment, considering the land use / land cover, soil and topographic characteristics for each sub-basin. The obtained HRUs represent units with similar hydrological characteristics.

HRUs, which act as the unit of modeling, simplify the modeling process and improve numerical efficiency. The representation of HRUs produced for this study using FAO and SSM databases is detailed in Chapter 5.6.

3.2.7. The SWAT-CUP Calibration Technique

Sensitivity parameters are determined by the users and their range is modified during semi-automated model calibration. The model is run and the necessary outputs are obtained from the SWAT output files (.TxtInOut). Using the SWAT_CUP interface, calibration, uncertainty or sensitivity analysis of SWAT model outputs can be easily performed. The main function of the interface is to provide a link between the input and output of the calibration program. There are five different optimization methods defined for calibration and uncertainty analysis in SWAT-CUP. These are Particle Swarm Optimization (PSO), Sequential Uncertainty Fiting (SUFI-2), Markov Chain Monte Carlo (MCMC), Parameter Solution (PARASOL) and Generalized Likelihood Uncertainty Estimation (GLUE) methods. SUFI_2 algorithms is the most preferred optimization method used for the SWAT model calibration. This method is obviously based on the sensitivity analysis between the observed and the simulated values. During the sensitivity analysis, the most effective parameters required in the calibration and validation process are achieved [46], [47]. In this study, the SUFI-2 optimization algorithm, which has the ability to analyze the sensitivity of the parameters, was preferred for the calibration.

Various SWAT parameters related to runoff can be estimated using the SUFI-2 algorithm. Uncertainty in SUFI-2 is defined as the inconsistency between the measured variables and the simulation variables. To explain this uncertainty, the measured data must be protected, except for outlier values. Thus, SUFI-2 combines calibration with uncertainty analysis to find uncertainty parameters. In SUFI-2, the model output uncertainty is measured at 95% estimation uncertainty (95PPU), and the input parameter uncertainty is expressed as a uniform distribution [48].

Model performance statistics are used to test the consistency of simulation values with observed values. Different statistical methods such as p-factor, r-factor, R^2, NSE, adj-R^2, RMSE, SSQR, PBIAS were defined in order to test the performance of the hydrological model. In this study, Determination coefficient (R^2), Nash-Sutcliffe coefficient of efficiency (NSE) and Percent Bias (PBIAS) values were used to evaluate the quality of model.

3.2.8. Summary of Data Acquisition

The summary of the input data is given on Table 3.1. The DEM having a resolution of 5 m was generated using aerial photographs which were taken from The Turkish General Command of Mapping. Files containing contour data were combined using the mosaic command of ArcGIS 10.4.1. Land use/land cover map were obtained from the CORINE-2018 Project. The soil map of FAO and Specified Soil Maps created as a result of laboratory and field studies (SSM) were used separately. Also, slope classes, stream network structures, sub-basin boundaries, catchment boundaries and height bands were determined by using DEM.

Table 3.1: Summary of data collection methods.

Parameter	Summarized Obtained Method
Digital Elevation Model	Using aerial photographs which were taken from Turkish General Command of Mapping
Land use and Land Cover	Coordination of Information on The Environment (CORINE) - 2018
Soil Data	Specified with Laboratory and Field Studies Food and Agricultural Organization (FAO) Soil Map [49]
Meteorological Data	Turkish General Directorate of Meteorology (TMMS) & National Environmental Prediction Center (NCEP)

3.3. Soil Data Preparation

Soil structure is a critical factor for understanding processes such as surface runoff, infiltration, crop growth and soil water. According to the soil map prepared by Turkey Ministry of Agriculture and Forest (TNSD), Saz-Çayırova Stream Watershed includes 3 main soil groups. These are Rendzinas (R), Brown forest soil without lime (N) and brown soils without lime (U). However, approximately 35% of the watershed is described as residential area and soil data is not defined for those regions. In addition, some of the parameters required for SWAT are not included in the soil database prepared by TNSD. Therefore, field and laboratory studies were carried out to obtain the required parameters for the SWAT. The specified soil database (SSM) was digitized in the GIS environment.

Since the catchment is dominated by industrial and residential areas, attention has been paid to take soil samples from points where the natural soil structure is not disturbed. Accordingly, 29 soil sampling points were chosen to represent the entire catchment (Figure 3.4). To determine the locations of soil sampling points, main soil groups of the catchment, slope / depth values obtained from the TNSD database and land use/cover status were considered. The degraded soil samples were taken approximately 1 kg from a depth of 0-30 cm.

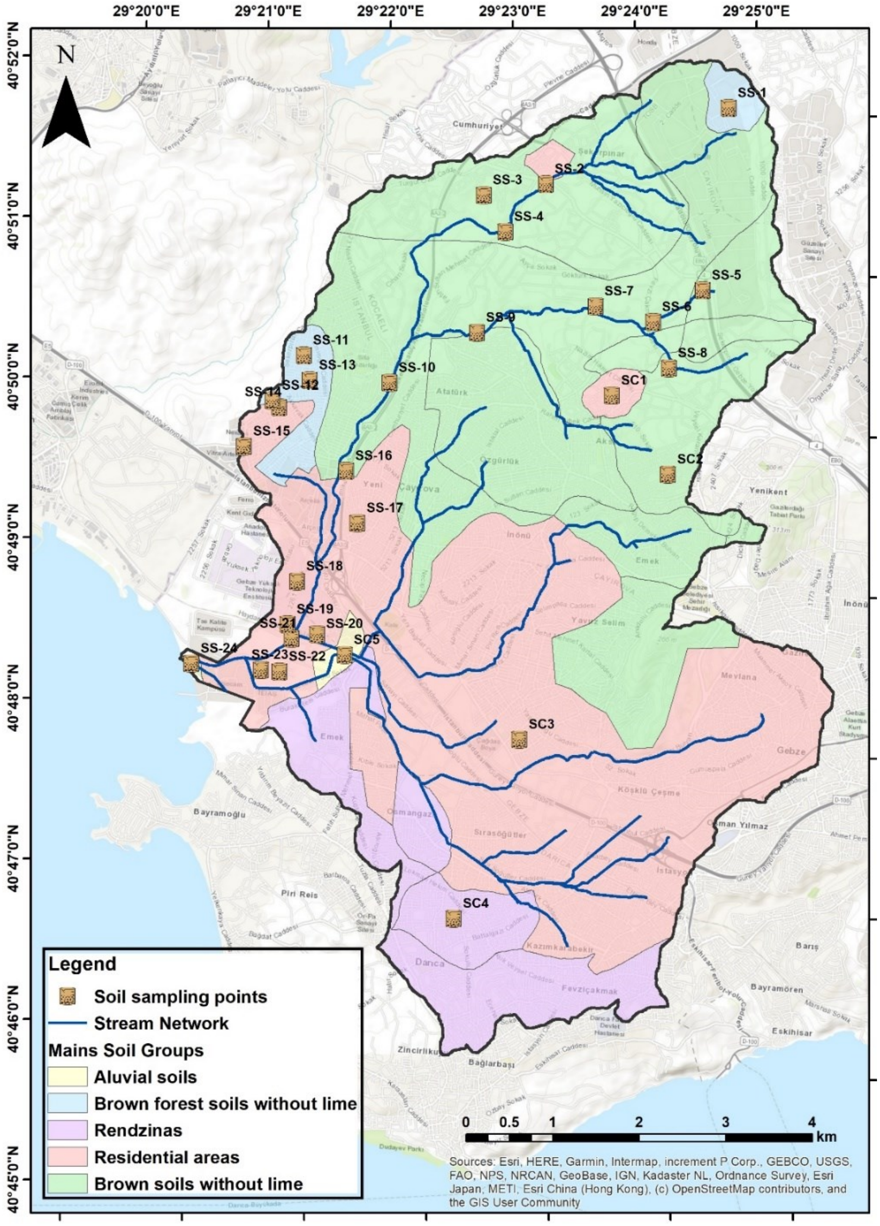


Figure 3.4: Soil sampling points and main soil groups provided by the Ministry of Agriculture and Forest.

The photos of soil sampling points are shown in Figure 3.5. During the field studies, equal amounts of soil samples were taken at 10 meters' intervals of the selected point and mixed in order to analyze the soil samples more homogeneously.

Figure 3.5: Selected soil sampling points.

Numerous physical and chemical soil parameters are used as input variables in SWAT. These parameters are hydrologic group, maximum rooting depth, maximum crack volume, texture of the specified soil, bulk density, available water content, saturated hydraulic conductivity, organic carbon content, clay-silt-sand-rock/gravel content, albedo, soil erodibility factor. In this study, some soil data parameters were adapted from the SWAT Manual [39]. The description of these parameters and determination methods for this study are shown in Table 3.2. While bulk density, organic carbon content (%), Maximum rooting depth, Depth from soil surface to bottom of layer, clay - silt - sand - gravel content (%), USLE soil erodibility factor were analyzed and calculated by field and laboratory studies, the rest of these parameters are obtained from the literature studies [50], [39], [51].

Table 3.2: Soil parameters required for SWAT.

Description	Determination Method
Hydrologic group	Turkey General Directorate of Rural Services (TNSD)
Depth from soil surface to bottom of layer (mm)	Field studies
Maximum rooting depth (mm)	Field studies
Fraction of porosity from which anions are excluded	Neitsch et al., [39]
Maximum crack volume (m^3/m^3)	Neitsch et al., [39]
Bulk density (g/cm^3)	Laboratory Analysis
Available water content (mm/mm)	Neitsch et al., [39]
Saturated hydraulic conductivity (mm/h)	Laboratory Analysis
Organic Carbon content (%)	Laboratory Analysis
Clay, silt, sand, rock/gravel content (%)	Laboratory Analysis
Albedo	Katharina [52]
USLE soil erodibility factor	Calculate according to Williams et al., [37]

Soil data preparation chapter is presented in detail in two subtitles named “soil input parameters obtained from the literature” and “soil input parameters measured by field and laboratory studies”.

3.3.1. Soil Input Parameters Obtained from The Literature

Many soil physicochemical properties that interact directly or indirectly with each other are used as input variables in SWAT. For instance, the erodibility factor of the studied region is a fraction of soil texture (clay, silt, sand and rock content) and organic carbon content [50]. The soil parameters obtained from literature studies are detailed as follows;

- Soil hydrologic groups are one of the important soil parameters defined by the US Natural Resources Conservation Service (NRSC) and determined by considering the infiltration characteristics. Four main hydrologic groups have been defined worldwide. These groups include high infiltration rate (A, about range between 7.5 - 10 mm h^{-1}), moderate infiltration rate (B, about range between 3 - 7.5 mm h^{-1}), slow infiltration rate (C, about range between 0.8 - 7.5 mm h^{-1}) and very slow infiltration rate (D, smaller than 0.8 mm h^{-1}). In this study, hydrologic group values were determined by combining the main soil groups with soil properties in the TNSD database [53]. The dominant hydrologic group of the Saz-Çayırova catchment is found as C.

- Most soil minerals are negatively charged in normal pH. Therefore, the interaction of these particle surfaces with anions such as nitrate is repulsive. This repulsion is called fraction of porosity from which anions are excluded [54]. It is an optional parameter for the model and 0.5 value was used in this study.
- When soil samples are dried in an ash oven, the soil water content can range from zero to the saturation level of the soil (maximum-level). This available water content is defined in two intermediate stages of transfer between plant and soil. These are field capacity and permanent wilting point, respectively. Field capacity is the water content found when a well-moistened soil is discharged for about two days. Permanent wilting point is the water content found when plants begin to fade. The Available water content parameter is defined as the difference of these two parameters. Field capacity and permanent wilting point parameters were obtained from the SWAT documentation [34].
- Moist soil albedo refers to the fraction of solar radiation reflected by the surface soil body to the amount incident upon it. The albedo values determined according to the study of Fricke K. [52] considering the texture classes.
- Maximum crack volume (m^3 / m^3) parameters, 0.5 was chosen for this study [34], [39].

3.3.2. Soil Input Parameters Measured by Field and Laboratory Studies

Soil texture analysis, which directly affects other variables and is critical for its determination, was performed first. Soil albedo, available water content, saturated hydraulic conductivity values was determined according to the soil texture. In order to reveal the soil texture, gravel/rock (particle diameter > 2mm) and sand (particle diameter ranges between 2 and 0.5 mm) content was determined by sieve analysis. The amount of smaller diameter grains such as silt and clay was analyzed using the Mastersizer-2000 Malvern device in the Materials Science and Engineering

Laboratory of Gebze Technical University. Photographs showing the order of texture analysis are given in Figure 3.6.

Figure 3.6: Laboratory analysis steps for soil texture analysis, a) weighing soil samples, b) sieve analysis for gravel and sand, c) sieve analysis for silt and clay, d) fine particle size collection for Mastersizer analysis.

In order to define the soil texture type triangle plot was used considering the measured rock / gravel, sand, silt and clay percentages (Figure 3.7). For the study area, soil groups were determined as sandy loam (82.6%), silt loam (13.6%) and loamy sand (3.8%), respectively. Combining the data on soil textures and main soil groups in the GIS environment, soil sub-polygons of the SSM were created.

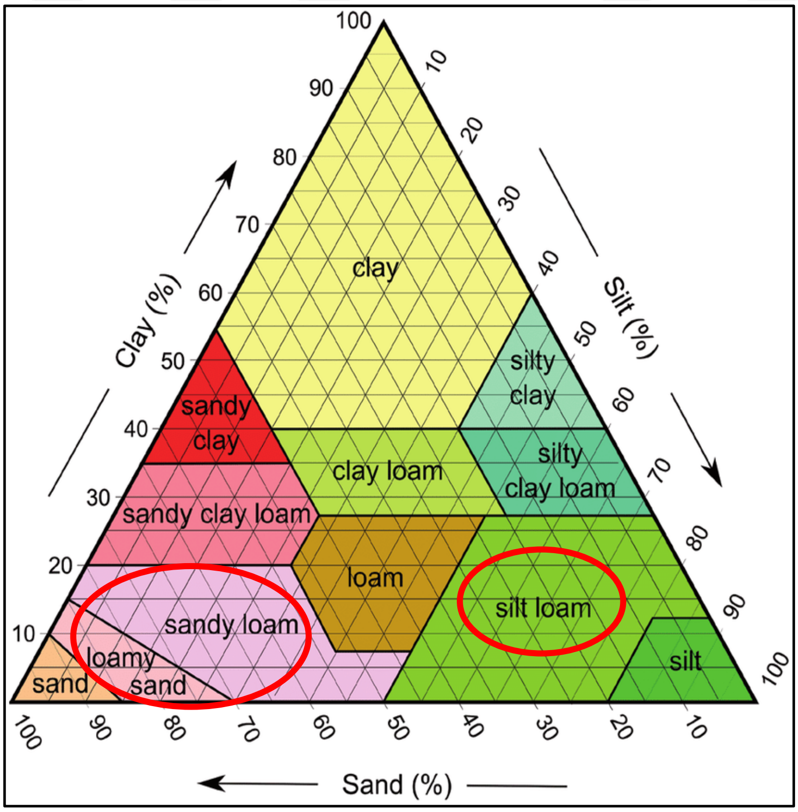


Figure 3.7: Definition of the soil texture using the soil triangle plot.

Organic carbon content was determined based on the organic matter content of the soil samples using the equation given below (Equation 3.2).

$$\text{Organic matter content (\% soil weight)} = 1.72\ \times \text{Organic carbon content} \tag{3.2}$$

Loss-on-Ignition method was used for analyzing the organic matter content of soil samples [55]. After drying 10 g soil samples in the oven, samples were burned in the muffle furnace at 550 °C for 4 hours. In the second step, the soil sample was reweighed and the difference was accepted as organic matter. In order to validate the obtained results the organic carbon content in a selected soil sample was determined by Walkley-Black method as well [56].

Soil bulk density expresses the ratio of mass of solid particles to the volume of the soil sample. In this study, bulk density was determined by ratio of oven dried soil samples to total volume of soil sample.

Saturated hydraulic conductivity is an important physical property that controls the permeability of the soils and therefore the transportation and washing of salt and agricultural chemicals in the soils. It is also an important parameter for evaluation of drainage problems in basins [57]. In this study, hydraulic conductivity was measured as follows; Soil samples were placed in a tin box with holes in the bottom. Secondly, the box was placed very slowly into a water-filled container. Water entering through the holes in the bottom of the box passes through the sample and reaches a certain level above the surface. Then, the box was removed from the water and the time until the water on the soil disappeared was measured using a stopwatch. As the last step, hydraulic conductivity is calculated with the formula below (Equation 3.3).

$$\mathrm{K} = \frac{\mathrm{L}}{\mathrm{t}} * \log_{\mathrm{e}}\left(1 + \frac{\mathrm{h}}{\mathrm{L}}\right) \tag{3.3}$$

Where K is the saturated hydraulic conductivity (mm/h), L is thickness of soil sample (mm), h is height of water over the soil sample (cm), t is time of water removed from the top surface of the sample (cm).

The Universal Soil Loss Equation (USLE) estimates long-term average annual erosion rate based on the field slope, soil type, precipitation pattern, topography of the field, crop activity and management practices. USLE Equation is given as follows;

$$A = R \text{ x } K_{USLE} \text{ x } LS \text{ x } C \text{ x } P \tag{3.4}$$

A is the potential long-term average annual soil loss in tonnes per hectare per year. R is the rainfall and runoff factor by geographic location. LS is the slope length-gradient factor. The LS factor represents the rate of soil loss under given conditions to the ratio in a field with a "standard" slope steepness of 9% and slope length of 22.13 m (72.6 ft). C is the crop/vegetation and management factor. C is used to determine the relative effectiveness of soil and crop management systems in terms of preventing soil loss. The C factor is a ratio comparing the soil loss from land under a specific crop and management system to the corresponding loss from continuously fallow and tilled land. P is the support practice factor. It reflects the effects of applications that will reduce the amount and rate of water flow and thus reduce the amount of erosion. The P factor represents the ratio of soil loss through a support application to straight farming down and up the slope [37]. These factors are usually used in the calibration step in SWAT application.

USLE soil erodibility factor is a measure of the sensitivity of soil particles to separation and transport by precipitation and discharge. Soil texture is an important parameter affecting this variable as well as structure of soil, organic matter content and permeability. The following equations (Equation 3.5- 3.9) were used in this study to calculate K_{USLE} which represent the soil erodibility factor.

$$K_{USLE} = f_{csand} \times f_{cl-si} \times f_{orgc} \times f_{hisand} \tag{3.5}$$

Where; f_{csand} is a factor that gives low soil erodibility factors for high-coarse sand soils and high values for low-sand soils, $f_{(cl-si)}$ is the low erodibility factor in soils with high clay-silt rate, f_{orgc} is a factor that reduces soil erodibility for soils with high organic carbon content and f_{hisand} is a factor that reduces soil erodibility in soils with extremely high sand content.

$$f_{csand} = \left(0.2 + 0.3 \times \exp\left[-0.256 \times m_s \times \left(1 - \frac{m_{silt}}{100}\right)\right]\right) \tag{3.6}$$

$$f_{cl-si} = \left(\frac{m_{silt}}{m_c + m_{silt}}\right)^{0.3} \tag{3.7}$$

$$f_{orgc} = \left(1 - \frac{0.25 \times orgC}{orgC + \exp[3.72 - 2.95 \times orgC]}\right) \tag{3.8}$$

$$\mathrm{f_{hisand}} = \left[1 - \frac{0.7 \times \left(1 - \frac{\mathrm{m_s}}{100}\right)}{\left(1 - \frac{\mathrm{m_s}}{100}\right) + \exp\left[-5.51 + 22.9 \times \left(1 - \frac{\mathrm{m_s}}{100}\right)\right]}\right] \tag{3.9}$$

Where; *ms* is the sand content (0.05-2.00 mm diameter) (%), m_{silt} is the silt content (0.002- 0.05 mm diameter) (%), m_c is the percent clay content (< 0.002 mm diameter), and *org C* is the percent organic carbon content of the layer (%).

Depth from soil surface to bottom of layer refers to the layer that prevents the plant roots from developing downward. Plants supply water and nutrients from the upper soil, and a depth of up to 150 cm significantly affects plant growth. In this study, maximum rooting depth is assumed as soil profile depth. The depth of the soil profile is determined by field studies and compared to soil depths of Turkish National Soil Database for verification.

3.4. Flow Rate Monitoring

Flow rate is the most critical parameter for water yield and regime in hydrology studies. Hence, in this study, locations of flow rate monitoring stations were determined to represent all river networks of the Saz Stream (Figure 3.8). Flow rate values were determined by the multiplication of water velocity (m s^{-1}) and cross sectional area (m^2) of the stream. Water velocity was measured using the "Hydrometer 2 JDC" current meter. In order to define the cross-sectional area, firstly, the depth of the water column was measured with 20 cm intervals. Secondly, a reference depth point has been set. Taking into account this reference depth point, the cross-sectional area was calculated for each daily and monthly flow rate measurements (Equation 3.10).

$$\mathrm{Q = A \times V} \tag{3.10}$$

Where Q is the flow rate (m^3 s^{-1}), A is the cross-sectional area (Calculated in R ver.4.0.3 Programming Language – m^2) and V is measured velocity (m s^{-1}).

Each flow rate monitoring point was also defined as the outlet of 8 sub-basins designed in the SWAT, as present in Figure 4.1 in Chapter 4. The daily flow rate measurements were carried out at the flow rate monitoring stations namely SD-0-1, SCD-0-2 and SCD-0-1. The station namely SCD-0-2 represents the total flow coming from Saz stream and Çayırova stream. The SD-0-1 represents flow coming only from the Saz stream. These two monitoring stations were used to evaluate the model performance after calibration. The station SCD-0-1 represents the outlet of the Saz-Çayırova stream. At this station, flow rate was measured by Kocaeli Metropolitan Municipality Water and Sewerage Administration (ISU). These values were used to compare the observed and predicted runoff before and after calibration. Measurements were carried out at the other 5 flow monitoring stations, shown in Figure 3.8, in monthly periods within the scope of TUBİTAK project (No: 119Y032). These measurements were used to evaluate the general characteristics of the catchment, especially the flow rates from Saz Stream.

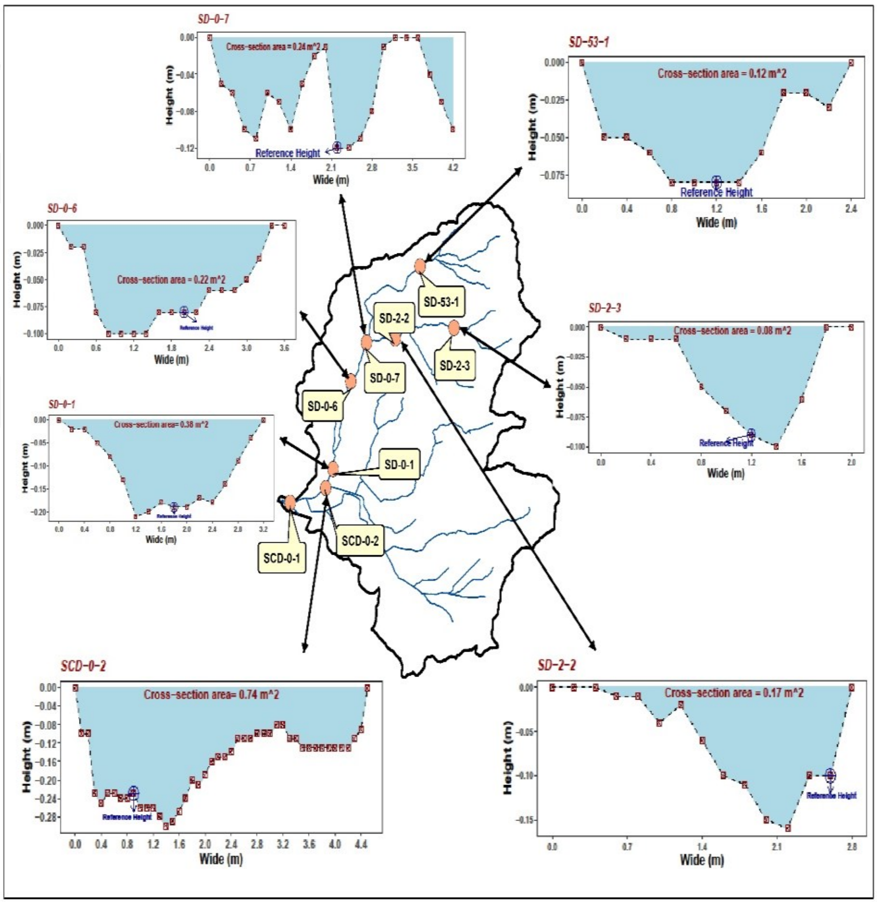


Figure 3.8: Location and cross-sectional area of flow rate monitoring points.

Locations, measurement periods and measurement frequencies of 8 flow rates monitoring points were shown in Table 3.3. It should be noted that flow rate measurements could not be carried out at the SCD-0-2 station between April 3rd and June 1st 2020 due to the COVID-19 precautions.

Table 3.3: Location of surface runoff monitoring points and measurement periods and frequencies.

ID	Outlet Points	Measurement Period	Measurement Periodicity
1	SD-53-1	June - December 2020	Monthly
2	SD-2-3	June - December 2020	Monthly
3	SD-2-2	June - December 2020	Monthly
4	SD-0-7	June - December 2020	Monthly
5	SD-0-6	June - December 2020	Monthly
6	SD-0-1	July - December 2020	Daily
7	SCD-0-2	February - December 2020	Daily
8	SCD-0-1	1 January 2016 – 23 October 2020	Daily

The photos taken at the flow rate monitoring stations were illustrated in Figure 3.9. Where the station namely SCD-0-1, SCD-0-2 and SD-0-1 are located in Gebze Technical University Campus, the remainings are located outside the campus.

Figure 3.9: The flow rate monitoring stations throughout the Saz Stream catchment.

4. RESULTS of LABORATORY and FIELD STUDIES

4.1. Temporal and Spatial Variation of Flow Rate Measurements

Sub-basins were formed in the watershed delineator module of the SWAT model in order to interpret the variation of the flow rate across the catchment easily (Figure 4.1). The flow rate monitoring points are designed as the outlet of each sub-basin.

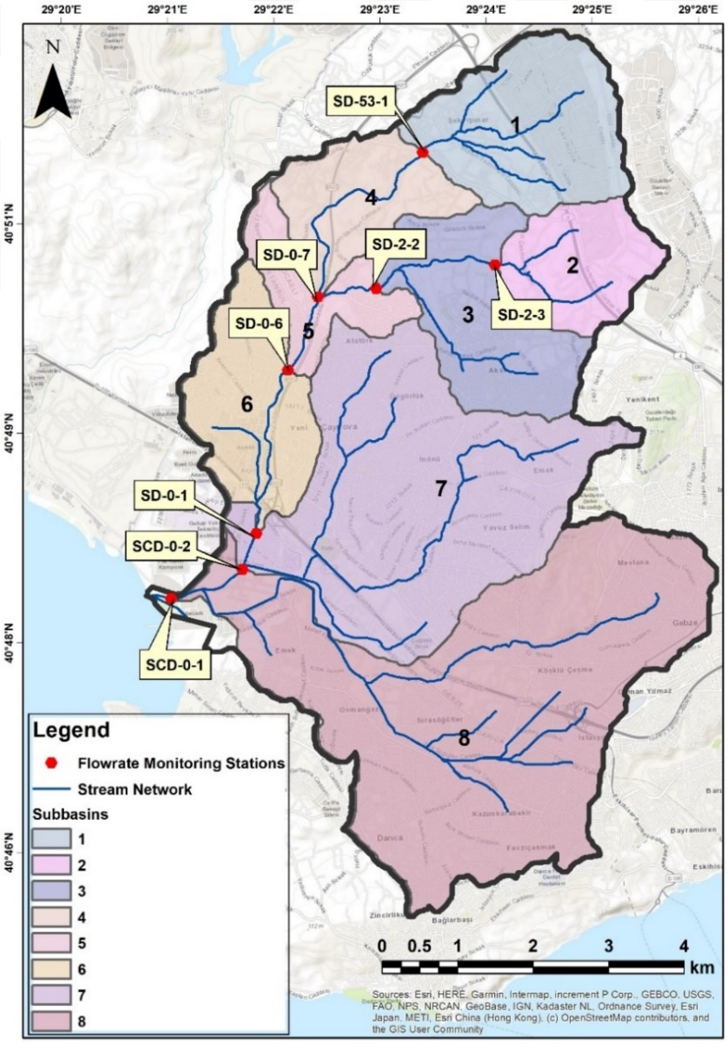


Figure 4.1: Sub-basins of Saz-Çayırova catchment area.

Surface-flow measurement points have been chosen to represent the entire Saz Stream, within the scope of TÜBİTAK Project (No: 119Y032).

The surface-flow variability of the Saz Stream was illustrated in Figure 4.2. The monthly measurement points (seen in Table 3.3) were used to assess the runoff condition of the catchment before model run. Considering the spatial distribution of the measurement points, flow rates increase linearly at the stations from the upstream to downstream of the watershed, as expected. The lowest value was observed in SD-0-7 point. This station is representing the rural area and it was one of the most affected points by flood events on rainy days. The highest surface-flow values were observed at SCD-0-2, which is the closest station to the downstream of the catchment. The most important reason for this condition is that the load from the Çayırova Stream Network is also included at SCD-0-2. Although the area of Subbasin-1 (about 5 km^2) is larger than Subbasin-2 (2.8 km^2) and Subbasin-3 (4.4 km^2) lower surface flow rate values were observed in SD-53-1 representing Subbasin-1 (Figure 4.1), comparing to SD-2-3 and SD-2-2 (representing Subbasin-2 and Subbasin-3). This indicates that watershed is more fed by waters coming from the Subbasin-2 and Subbasin-3 compared to Subbasin-1.

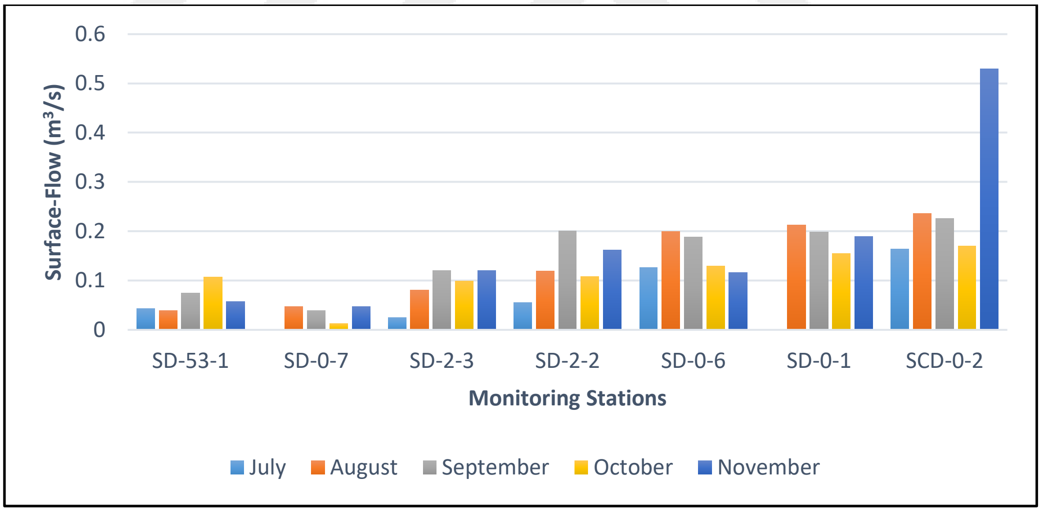


Figure 4.2: Temporal and spatial variation of flow rate values.

Considering the temporal change, the highest average flow rate values was observed in November, while the lowest flow rate values were observed in July. This situation is directly proportional to the meteorological events in the catchment. In addition, the relatively high flow rate values observed in August may be attributed to residential and industrial discharges. In October, a relative decrease was observed in the flow rate values compared to September and August, for all stations except for the station SD-53-1.

Measurement period of SCD-0-2 representing loads from Saz and Çayırova stream networks, were between February 24th and December 1st 2020. But, flow rate measurements could not be carried out between April 3rd and June 1st 2020 due to Covid-19 pandemic measures (Figure 4.3). The measurement period at the SD-0-1 representing only the load from the Saz stream network was between July 17th and December 1st 2020. Daily oscillation of both observed monitoring stations was similar, except for certain days. To support this situation, the statistical determination coefficient (R^2) of the flow rate values calculated for two measurement stations was found to be 0.78. The most prominent episode in Figure 4.3 took place on August 22nd, 2020. At this episode, 8 mm daily average precipitation was observed and flow rates were measured as 1.3 m^3/s and 5.8 m^3/s at SD-0-1 and SCD-0-2, respectively. The relatively high difference between them shows the severity of the load coming from Çayırova Stream at rainfall terms. Additionally, there was a lower episode on October 20th 2020 compared to the episode on August 22nd 2020. On the specified date, the average rainfall was measured as 0.25 mm, while the flow rates at SD-0-1 and at SCD-0-2 were measured as 1.13 m^3/s and 1.56 m^3/s, respectively. This episode may indicate a source discharging into the Saz stream.

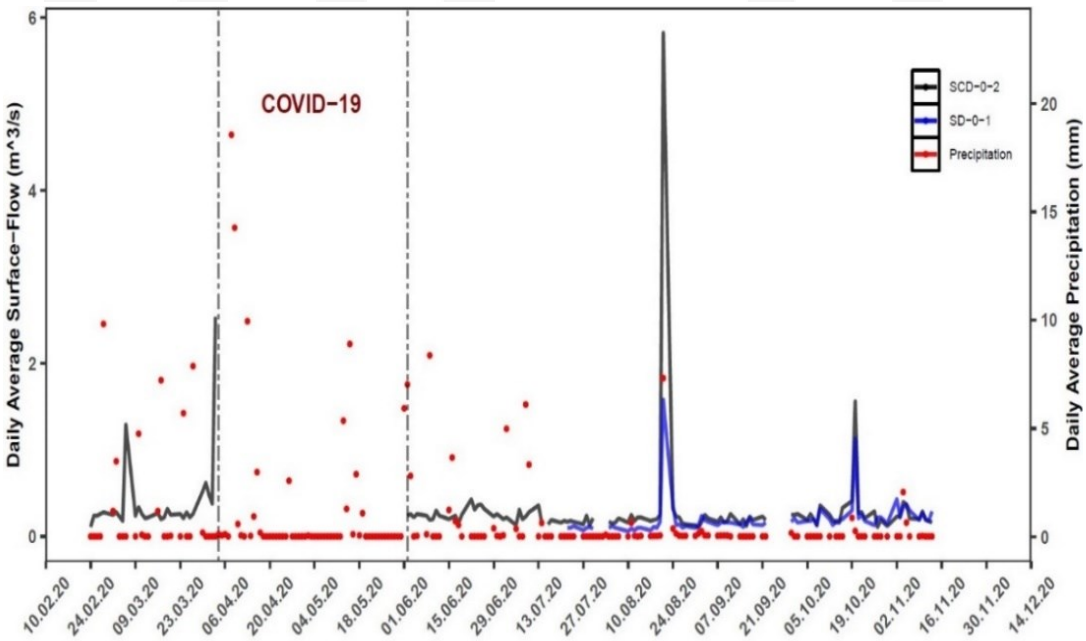


Figure 4.3: Temporal variation of precipitation and daily flow rates at the station SCD-0-2 and station SD-0-1.

Station SCD-0-1, which represents the catchment outlet, was used in the calibration and validation processes in the SWAT model. Thus, the temporal variation of the flow rates between January 1st 2017 and December 1st 2020 at the SCD-0-1, obtained from the Kocaeli Metropolitan Municipality Water and Sewerage Administration (ISU) were also examined. This station represents the load from Subbasin-8 in addition to Subbasin-7 represented by the SCD-0-2.

The rainfall measurements are weighted according to the Thiessen Polygon method [58] with the data obtained from Çayırova, Darıca, Tuzla and Gebze Meteorology Stations, which are located within the catchment (See Figure 5.4 in Chapter 5). Then, the variation of rainfall measurements compared to the flow rate measurements at the Station SCD-0-1 (Figure 4.4).

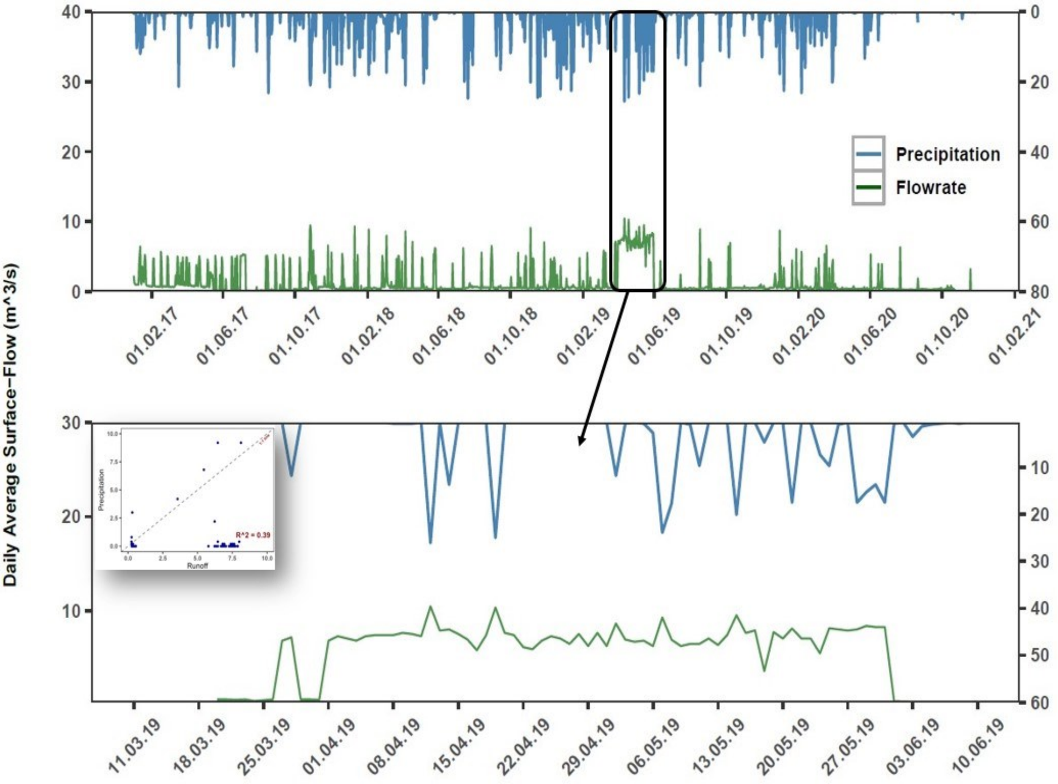


Figure 4.4: Variation of the daily flow rate of SCD-0-1 and precipitation in the catchment.

Low determination coefficient value ($R^2 = 0.09$, significance level $<2e^{-16}$) was observed between the daily flow rate and rainfall values. Several episodes were seen in the studied periods. The episodes seen on 31 October 2017 (Q = 8.98 m^3/s, Precipitation = 15.9 mm), on March 13th 2018 (Q = 6.6 m^3/s, Precipitation = 14.23 mm) and on December 29th 2019 (Q = 8.741 m^3/s, Precipitation = 13.45 mm) could be explained by meteorological processes. Contrary to expectations, the episode seen between March and June 2019 did not appear to be related to meteorological events.

While the flow rate variation between these dates ranged from 7.7 m^3 /s to 10.3 m^3/s, the maximum precipitation was recorded as 16 mm. In this episode, significant correlation ($R^2 = 0.39$ seen bottom of Figure 4.4) was not found between flow rate and precipitation. In this context, this episode may have been associated with anthropogenic activities such as wastewater treatment discharges, residential discharges throughout the catchment.

Temporal variation of flow rates of SCD-0-1 and SCD-0-2, representing another Çayırova network, was examined in order to understand whether the variation is significant or not. The determination coefficient (R^2) was found to be 0.52 (significance level$<2e^{-16}$) for the runoff values measured at these two stations. On the other hand, the average of the flow rate value (4.18 m^3/s) at the SCD-0-1 was about 15 times those of the SCD-0-2 (0.31 m^3/s). This relationship shows that the Çayırova Stream has the greatest contribution to the flow rate at the SCD-0-1.

4.2. Preparation of Specified Soil Map (SSM)

Soil sampling studies were carried out at 29 points representing the entire catchment in June 2020 (See Figure 4.4). The points are chosen considering the main soil group given in Turkish National Soil Maps (TNSD) and land use/land cover classification obtained from CORINE-2018.

SSM database was formed as follows; Firstly, the spatial distribution of soil profile depth (mm), erodibility factor, hydraulic conductivity (mm/h), organic carbon (%), clay-sand-silt-rock content (%) values measured was interpolated using the Inverse Distance Weighted (IDW) interpolation technique. Secondly, the interpolated parameters were merged according to soil texture and main soil groups. Thirdly, the merged map was reclassified and 15 different soil regions were created within the catchment that can be integrated into the SWAT model. Furthermore, each soil region was classified according to the main soil groups taken from the TNSD database and the texture class that determined with soil texture triangle plot. Three main soil groups have been defined for the catchment, namely rendzinas (R – 16.54%), brown forest soils without lime (N - 10.96%), brown soils without lime (U – 37.5%).

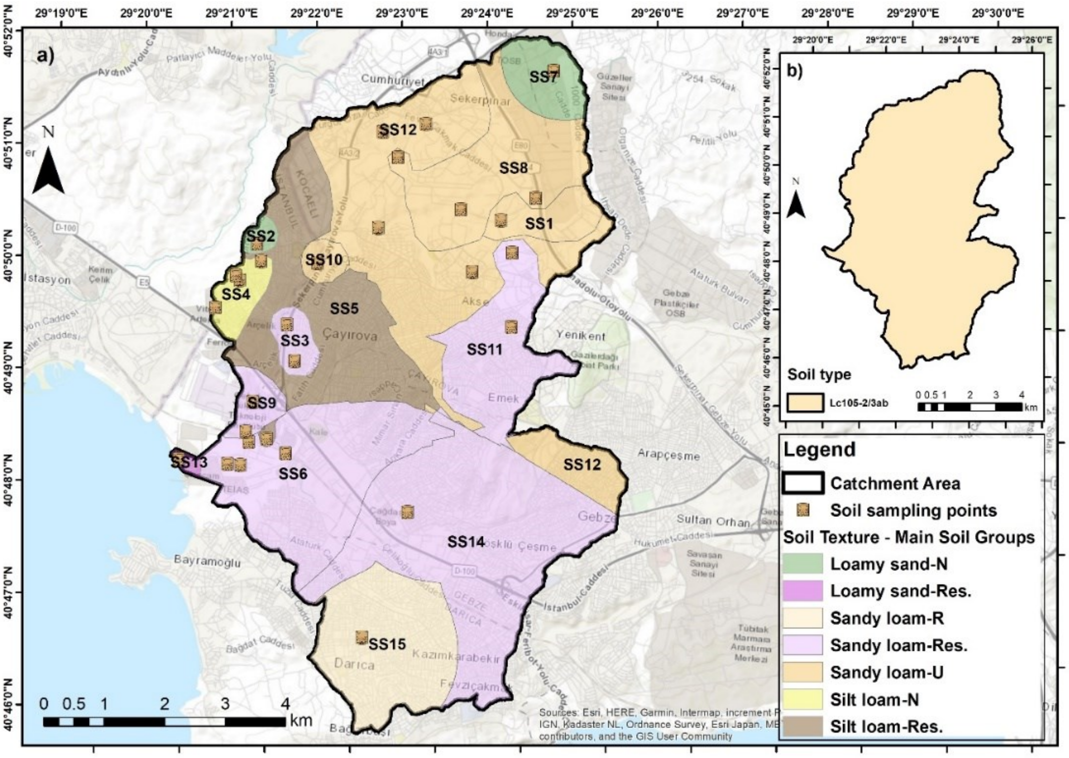


Figure 4.5: Specified Soil Database (SSM) (a), and FAO soil database (b).

The remaining areas were classified as Residential areas (Res. – 35.1%) in the TNSD. As a result of the soil analysis, 3 different soil textures were determined; sandy-loam (82.6%), silt-loam (13.8 %), loamy-sand (3.6%). After the combination of main soil groups and texture classification, all physicochemical properties have been added to the "soil look-up table" in the appropriate format to run the SWAT model.

The Global Food and Agricultural Organization (FAO) soil database, which is more homogeneous than the SSM database, was used to examine the effect of different soil databases on surface runoff. In this soil database, the dominant soil type for the entire Marmara Basin, also including the study area, is defined as Chromic Luvisols with medium to fine texture (sandy-loam), undulating to rolling, and reddish to brown color.

4.2.1. Spatial Distribution of Measured Soil Parameters

Spatial distributions of clay, silt, sand, rock / gravel parameters (%) that form the soil texture analyzed by laboratory studies along the Saz-Çayırova catchment were given in Figure 4.6. Clay is the most functional fraction in soil due to its negative charge source. Therefore, clay affects not only surface runoff, but also many physical, chemical and biological soil properties such as retention of plant nutrients, soil aeration, soil temperature and moisture [59]. Relatively lower clay content (2.68 – 4.17%) was observed at higher elevations, while the southeast part of the catchment (in the GTU campus) has a relatively higher clay content (5.71 – 7.22%). The lowest clay content was observed in the central parts of the Çayırova catchment and outlet of the watershed. While the silt content was relatively high in the western part of the catchment (39 - 49%), which also represents the residential area, it was found to be the lowest in the southern parts of the Çayırova region (10 - 19%). The sand content (left-bottom in Figure 4.6) ranges from 18 to 76% throughout the catchment. In addition, the soil texture at the outlet of the catchment mostly consists of sand (approximately 75%). As a result of the field studies, generally high sand content was observed in all soil samples taken from the region. The rock/gravel content of the region was found to be high (36 - 50%) in the southern parts of the Çayırova catchment, in contrast to the silt distribution. In summary, the soil texture within the catchment area is found to be rich in sand and silt content.

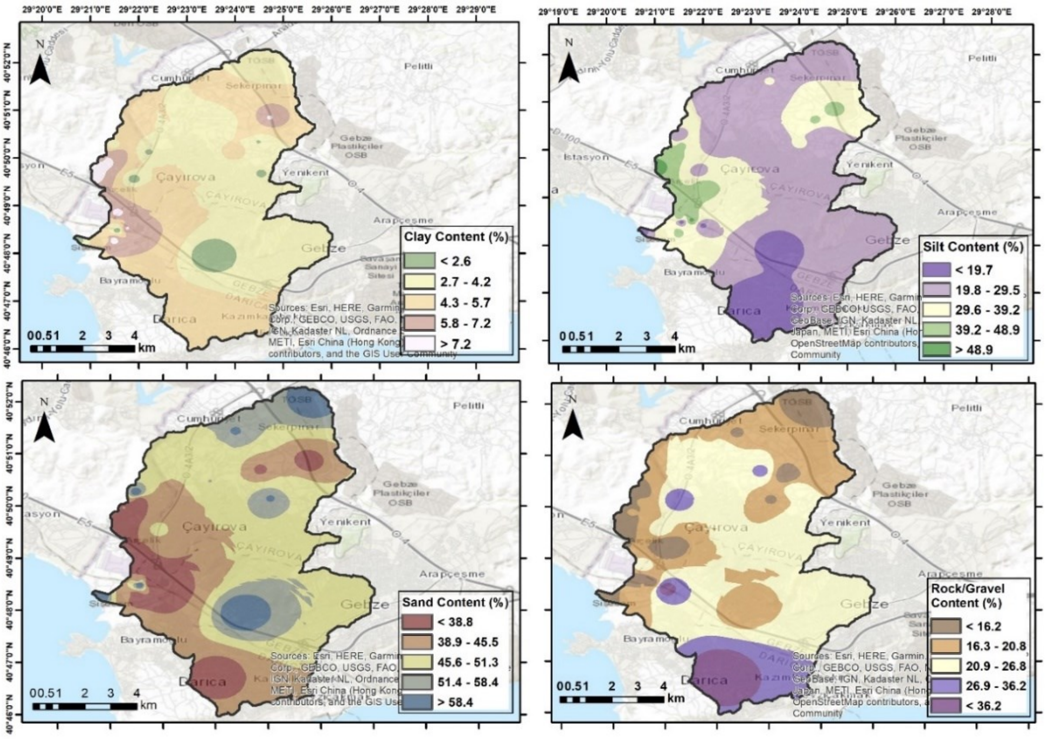


Figure 4.6: Spatial distribution of soil texture properties in the Saz-Çayırova catchment.

Spatial distributions of soil profile depth (mm), bulk density (g cm^{-3}) organic carbon content (%) and erodibility factor parameters (%) along the Saz-Çayırova catchment were given in Figure 4.7. The relationship of soil profile depth to surface runoff is critical for interpreting infiltration characteristics of soil, and water bodies emerging from shallow aquifers. In addition, the relationship is important for modeling the lateral transport of water in the soil profile [60]. The soil profile depth varies between 250 and 900 mm across the studied catchment, and the average depth was determined as 498 mm. The maximum soil profile depth (900 mm) was observed in the northeastern parts of the catchment, which also represents the industrial zone, while the minimum soil profile depth (250 mm) was analyzed near residential areas and at the downstream sites of the watershed. Soil bulk density is a key variable in estimating hydrological properties of soil, defining rainfall-infiltration- base or surface runoff-erosion relationships and soil quality [61]. Soil bulk density values measured in the laboratory ranges from 1.12 to 1.63 g/cm^3 and the average value is 1.4 g/cm^3. While high bulk density values were observed at the regions with higher elevations in north-east parts of the catchment, low values (1.12-1.22 g/cm^3) were observed in the southern parts of the catchment.

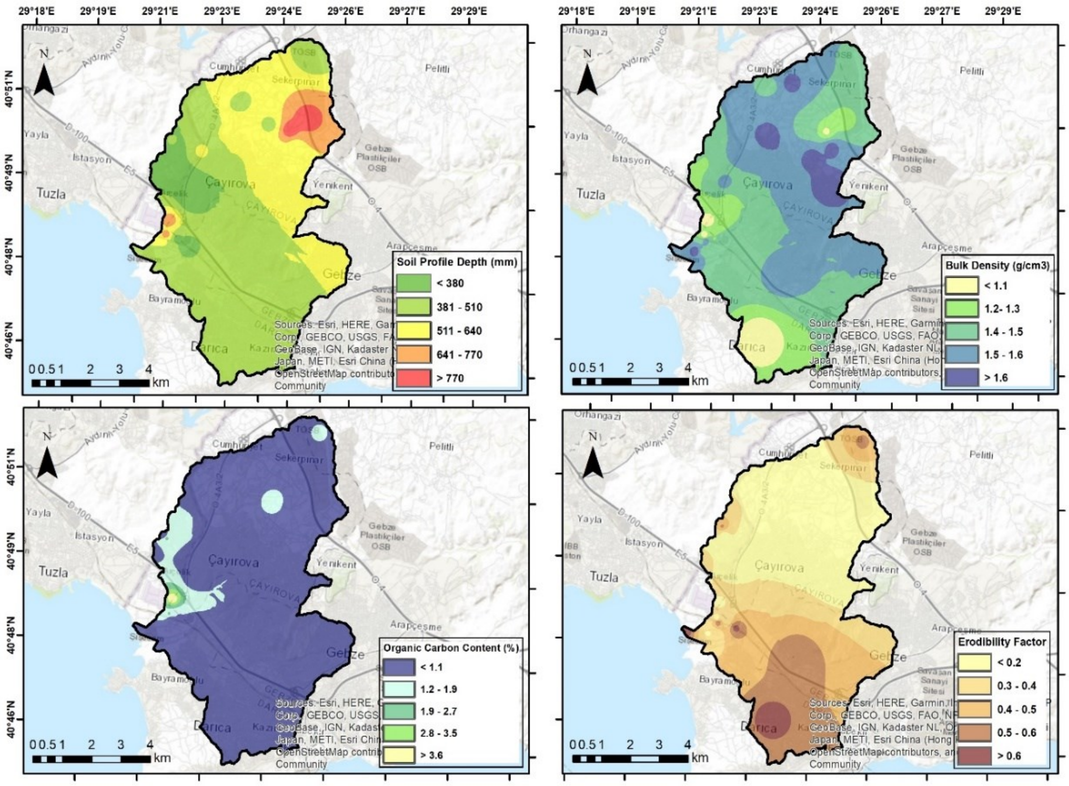


Figure 4.7: Spatial distribution of soil physicochemical parameters.

The soil parameters required for the SWAT model interact with each other directly or indirectly, as mentioned previously. One of the most considerable examples is the organic carbon content (%). In many studies, high correlations have been observed between organic carbon content and area-averaged available water content, bulk density and soil texture [62], [63], [64], [65]. In this study, determination coefficient (R^2) value was found as 0.78 between organic carbon content and bulk density. In addition, the R^2 value between the organic carbon content and the clay content was found as 0.82. It can be said that this results supports the previous studies in literature. The organic carbon content of the catchment is relatively low. While high carbon content appears in the south parts of the Saz sub-basin (1.91 - 4.26%), organic carbon contents of the remaining areas are generally varied between 0.3 - 1.1%.

The soil erodibility factor is a key factor expressing the resistance of the soil to the impact exerted by the forces of water erosion such as precipitation and surface runoff. In another perspective, the erodibility factor represents not only the susceptibility of the soil to erosion but also the runoff. The erodibility factor varies between 0.23 and 0.53 throughout the watershed. Higher erodibility factor values were calculated in the Çayırova sub-basin compared to Saz sub-basin.

Hydraulic conductivity is one of the most important hydro-geological factors because of direct interaction with other soil physico-chemical parameters, managing the flow of fluids, and pollutant transport in soils or aquifers [66]. Saturated hydraulic conductivity values range between 4.36 and 61.18 mm/h across the study area (Figure 4.8). High values were observed in the down-stream sites of the Saz stream and within the Çayırova sub-basin. The lowest hydraulic conductivity values were observed in the regions close to the residential areas (western parts of the catchment). The average conductivity value is calculated as 31.91 mm/h for the whole catchment.

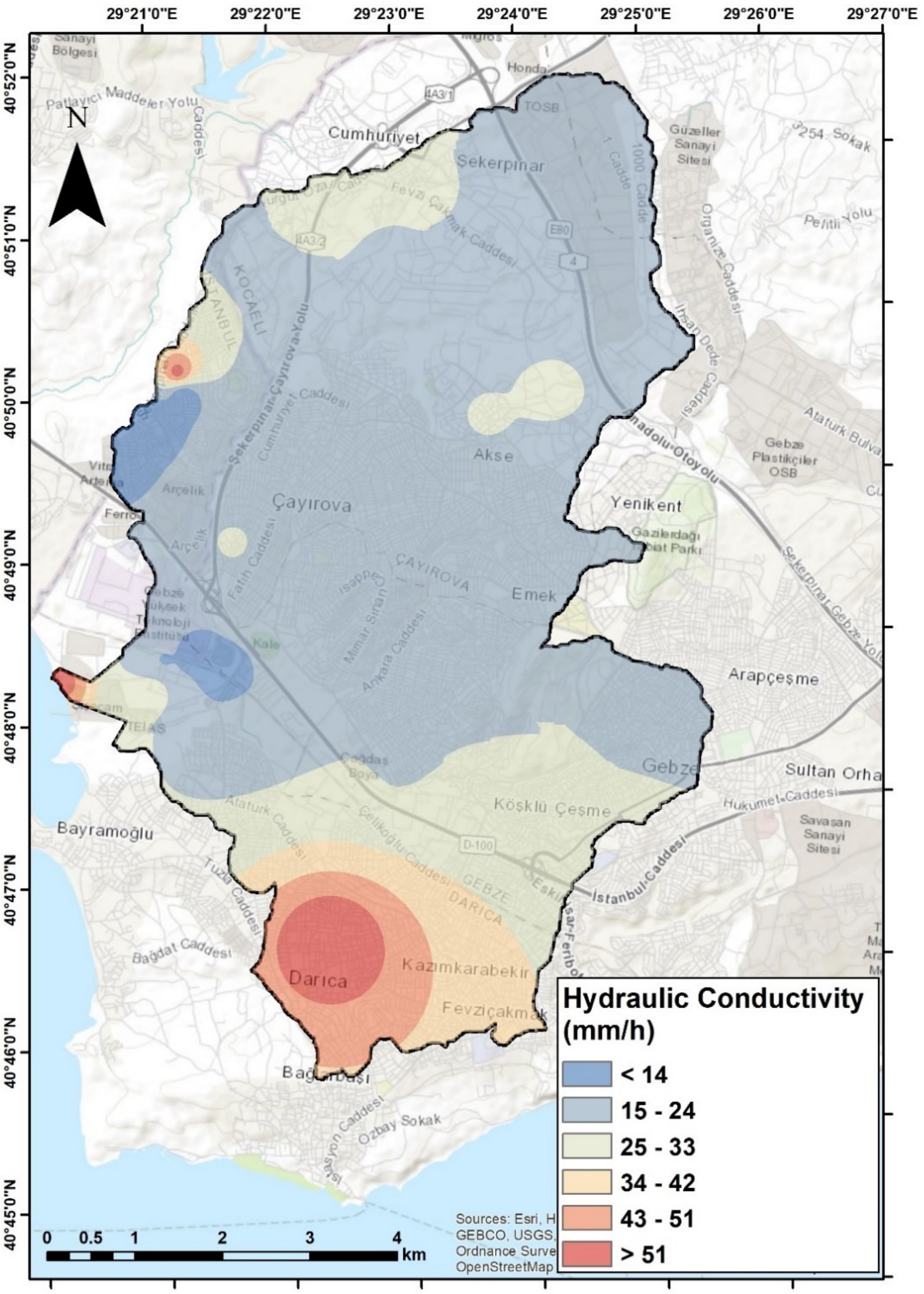


Figure 4.8: Spatial distribution of hydraulic conductivity within the Saz-Çayırova catchment.

4.2.2. Comparison of SSM and FAO Soil Database

The specified soil parameters by laboratory and field studies is summarised in Table 4.1. In the SSM database, soil parameters were examined in one layer, while those of the FAO database were examined in two layers. Soil profile depth and maximum rooting depth (mm) were selected equal in SSM. The depth values in the specified soil database were found at almost half of the values given in the FAO database. While soil texture (sand, silt and clay) values are almost evenly distributed in the FAO database, sand and silt content are formed about 78% of the SSM database. Additionally, the clay content values in SSM are 5 and 8 times lower than the layer-1 and layer-2 in FAO database, respectively. Another parameter showing a high difference in the two soil databases is saturated hydraulic conductivity (mm/h). The hydraulic conductivity was defined as 7.16 and 3.19 mm / h for layer-1 and layer-2 in the FAO soil database, respectively. Also, the average saturated hydraulic conductivity value was measured as 31.67 mm/h for SSM. These differences observed in depth, hydraulic conductivity and texture values can be explained by the fact that the FAO database accepts average values for the whole of the Marmara Basin. Another difference is in the soil hydrological group; while the dominant hydrological group was embedded as D (Infiltration rate ranges between 0 - 0.8 mm/h) in the FAO database, this parameter was defined as C (Infiltration rate ranges between 0.8 - 3 mm/h) in SSM.

Table 4.1: Summarized soil parameter for SSM and FAO databases used in SWAT modeling.

Sampling ID	Specified Soil Database (SSM)		FAO Database	
	Average	Min. - Max.	Layer-1	Layer-2
Profile Depth - Maximum rooting depth (mm)	493	250-900	840	840
Bulk density (g/cm^3)	1.4	1.12 - 1.63	1.3	1.5
Organic Carbon (%)	1.1	0.32 - 4.28	1.2	0.5
Rock/ Gravel (%)	17.6	5.55 - 34.9	10	0
Sand (%)	43.3	17.9 - 75.5	31	33
Silt (%)	34.1	12.7-58.7	32	25
Clay (%)	5.1	1.1-8.7	27	42
USLE soil erodibility factor	0.29	0.23-0.53	0.25	3.16
Saturated hydraulic conductivity (mm/h)	31.91	13.27-61.18	7.16	3.19

5. APPLICATION of THE SWAT MODEL

The version ArcSWAT 2012 was used as the model interface. Thereafter, the necessary data was pre-processed to be used by the model. The input layers, calibration and validation steps of the SWAT model are shown in Figure 5.1. In order to reveal hydrological processes of the Saz-Çayırova Stream. Firstly, the model input parameters (DEM, land use / land cover, soil database, slope definition, meteorological parameters) were modified to fit the input files of the SWAT model. In the second step, the relationship between predicted and observed runoff values generated from FAO and SSM soil databases was evaluated without applying any model calibration. Lastly, the relationship between predicted and observed runoff values was examined after applying semi-automatically calibration technique.

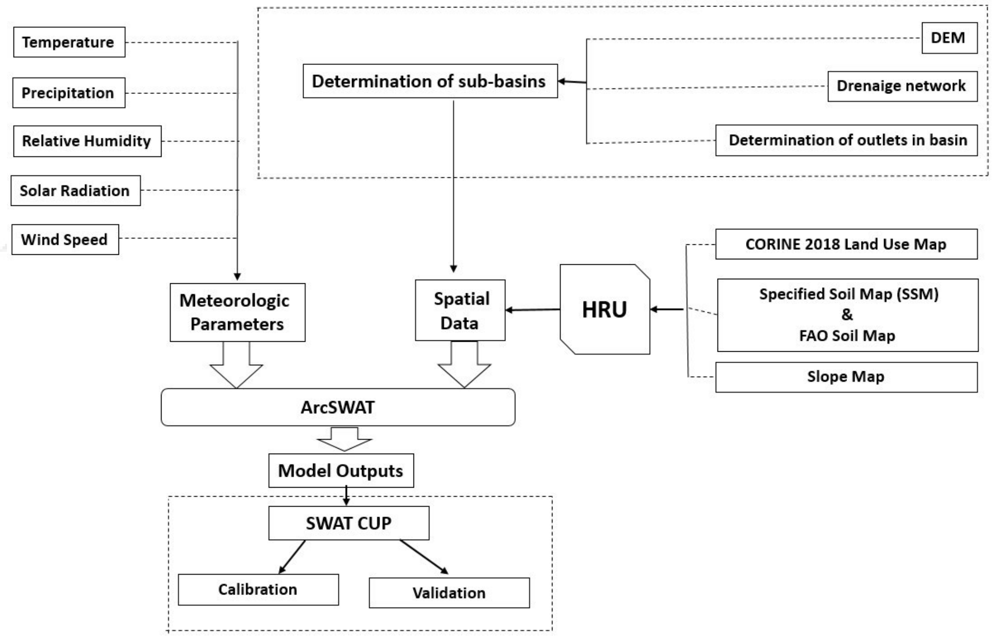
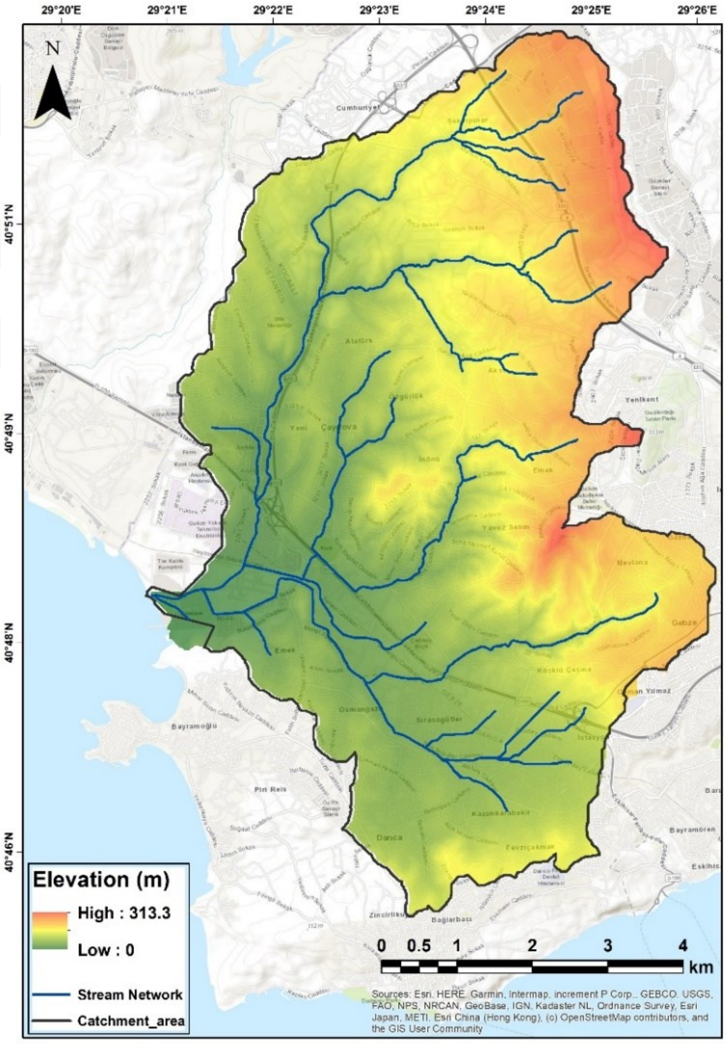


Figure 5.1: Flow chart of SWAT model.

5.1. Preparation of Model Input Layers

5.1.1. Watershed Delineator

During the initial setup phase of the model, the digital elevation model (DEM), which is the most basic base file on which many operations can be performed, was uploaded to the system (Figure 5.2.a). After this process is completed, the minimum, maximum and threshold values on the DEM for the sub-basins that can be created based on it are shown in the model. Based on this information, minimum and maximum sub-basin boundaries were calculated as 26 and 5246 ha, respectively. The recommended threshold value was set manually as 60 ha and the drainage network was determined according to this value. This threshold value was seen as the value that gives the closest one to the current drainage network produced by using aerial photographs. The height of the catchment was calculated as minimum sea level, maximum 331 m and average 100.7 m. It can be said that the variation in elevation is relatively high. This especially affects the amount of water to surface runoff.



Sources: Esri, HERE, Garmin, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), (c) OpenStreetMap contributors, and the GIS User Community

Figure 5.2: The Digital Elevation Model used in the watershed delineation steps of the SWAT.

The Sub-basin outlet points and the observation points were proposed by the system, after the drainage network was determined. These points were manually replaced with outlet points of 8 sub-basins, taking into account the previously defined flow rate measurement points. Since the branches of Çayırova stream are in the form of a channel, it has been accepted as a single sub-basin.

5.1.2. Land use and Land Cover Data

Since the properties of the CORINE land use data used to create hydrological units do not match the land use database of the SWAT model, the land use code equivalent of the CORINE-2018 data should be determined. Any standard method is not defined for the matching. In this study, the equivalent of the CORINE codes were determined by considering the CORINE Level-I (Anderson Classification) and Level-III codes. The considered CORINE codes and their SWAT code equivalents are given in Table 5.1.

Table 5.1: Land use codes of CORINE and SWAT database and percentages.

CLC Code	SWAT Code	Anderson Classes level I	Type	Percentage (%)
111	URHD	Urban or built-up land	Continuous urban fabric	19.35
112	URMD		Discontinuous urban fabric	35.35
121	UIDU		Industrial or commercial units	29.01
122	UTRN		Road and rail networks and associated land	0.99
133	URCS		Construction sites	0.51
141	URLD		Green urban areas	0.54
231	PAST	Agricultural land	Pastures	1.34
242	CRGR		Complex cultivation patterns	6.16
243	CRDY		Land principally occupied by agriculture	0.50
321	GRAS	Forest	Natural grasslands	1.48
324	TUWO		Transitional woodland-shrub	4.74
523	WATR	Water	Sea and ocean	0.03

The CORINE method is based on remote sensing techniques [42]. The CORINE Land Cover classification is based on a three-level hierarchical basis, and each level corresponds to an individual scale. Level-III mapped on the scale of 1/100,000 and has information in 44 categories. Level-II corresponds to 1/500,000 and Level-I mapped on the scale of 1/1,000,000. According to CORINE Level-I, 73 % of the catchment consists of built-up lands, 15 % agricultural lands, %10 rangeland and %2 road and rail networks. According to Level-I classification, the study area is dominated by industrial and residential areas.

Considering the CORINE Level-III 11 land cover classification scheme applied over the study area was illustrated in Figure 5.3.

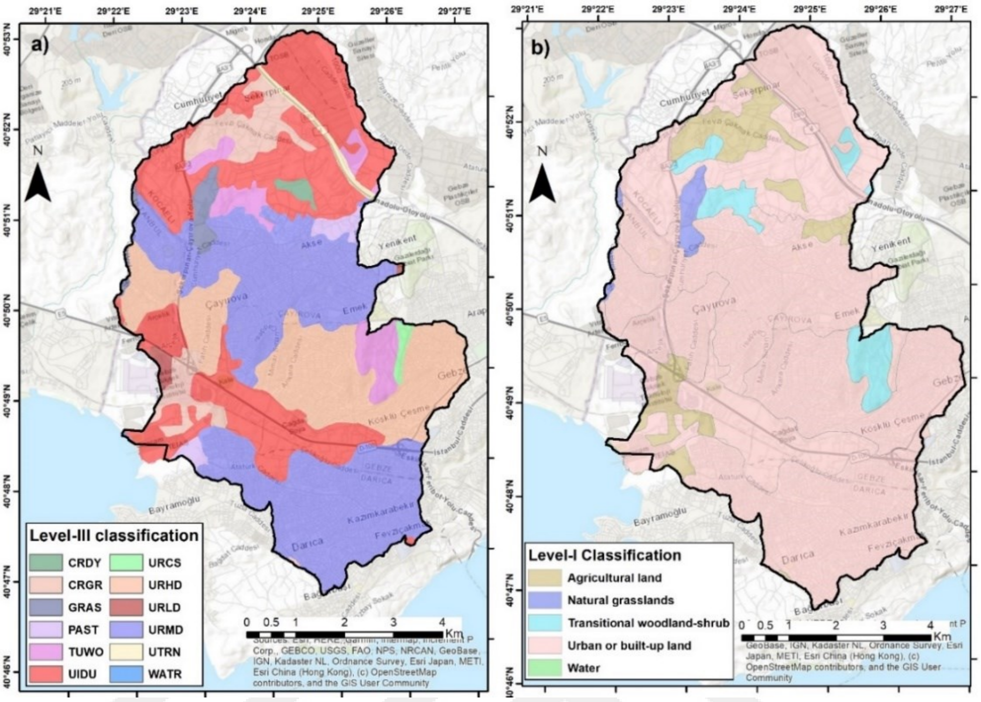


Figure 5.3: The spatial distribution of the land use classes of the region; a) CORINE Level-III classification, b) CORINE Level-I classification.

5.1.3. Soil Data

The Global Food and Agricultural Organization (FAO, 2007) soil database and Specified Soil Database (SSM) which produced laboratory and field studies were separately integrated into the SWAT in order to test the usability of global soil databases in small-scale catchment.

Defined physicochemical soil properties for each soil region must be imported to the MS-Access file embedded in the SWAT. In the same time, each soil region's names must be integrated into the attribute table of the digitized soil map. In this way, defined soil physicochemical properties corresponding to each soil region can be evaluated by the model for hydrological processes (Table 5.2).

Soil variables were fitted for the profile depths of 300 and 1000 mm, respectively [67]. All the physicochemical properties of the soil required for the SWAT model are available in the FAO soil database in appropriate format. The FAO soil database integrated into the SWAT model is given in Table 5.3.

Table 5.2: The SSM database (user_soil table) integrated into the SWAT model for the Saz-Çayırova catchment.

SNAM	HYDGRP	TEXTURE	SOL_Z1	SOL_BD1	SOL_AWC1	SOL_K1	SOL_CBN1	CLAY1	SILT1	SAND1	ROCK1	SOL_ALB1	USLE_K1
SS1	C	Sandy loam-U	900.00	1.19	0.11	21.80	0.73	5.63	41.47	45.30	7.60	0.14	0.28
SS2	C	Loamy sand-N	400.00	1.41	0.07	55.80	1.14	3.07	20.42	65.40	11.11	0.15	0.42
SS3	D	Sandy loam-Res.	250.00	1.38	0.11	21.80	0.54	3.63	33.55	46.90	15.92	0.14	0.27
SS4	D	Silt loam-N	287.50	1.33	0.13	7.20	1.17	7.77	47.33	31.13	13.77	0.13	0.28
SS5	C	Silt loam-Res.	300.00	1.32	0.13	7.20	1.01	7.52	52.58	27.50	12.40	0.13	0.29
SS6	D	Sandy loam-Res.	437.50	1.55	0.14	13.27	0.64	7.44	37.08	41.08	14.40	0.14	0.32
SS7	C	Loamy sand-N	400.00	1.41	0.07	55.80	1.14	3.07	20.42	65.40	11.11	0.15	0.42
SS8	B	Sandy loam-U	600.00	1.37	0.12	16.93	1.09	4.82	31.22	39.73	24.22	0.14	0.26
SS9	C	Sandy loam-Res.	750.00	1.14	0.12	14.50	3.15	5.89	46.75	30.40	16.96	0.14	0.26
SS10	C	Sandy loam-U	550.00	1.50	0.11	21.80	1.02	2.47	20.53	42.10	34.90	0.14	0.24
SS11	C	Sandy loam-Res.	550.00	1.59	0.11	21.80	0.67	2.56	21.61	50.20	25.63	0.14	0.24
SS12	C	Sandy loam-U	561.67	1.52	0.11	21.80	0.60	4.23	24.86	50.78	20.12	0.14	0.25
SS13	D	Loamy sand-Res.	500.00	1.52	0.07	55.80	0.47	2.16	16.31	75.53	6.00	0.15	0.53
SS14	C	Sandy loam-Res.	500.00	1.52	0.11	21.80	0.54	1.12	12.65	69.73	16.50	0.14	0.47
SS15	B	Sandy loam-R	500.00	1.80	0.10	1.20	0.50	5.00	10.00	35.00	50.00	0.15	0.55

Table 5.3: The FAO soil database (user_soil table) integrated into the SWAT model for the Saz-Çayırova catchment.

Layer	SNAM	HYDGRP	TEXTURE	SOL_Z	SOL_BD	SOL_AWC	SOL_K	SOL_CBN	CLAY	SILT	SAND	ROCK	SOL_ALB	USLE_K
1	Lc105-2/3-ab	D	Sandy loam	300	1.3	0.146	7.6	1.2	31	28	40	0	0.0484	0.2512
2				1000	1.5	0.146	3.19	0.5	43	25	33	0	0.1867	0.2512

[1] HYDGRP: Hydrological Group, ANION_EXCL; Fraction of porosity from which anions are excluded, SOL_CRK; Maximum crack volume, SOL_Z; Depth from soil surface to bottom of layer (mm), SOL_BD; Bulk density (g/cm3), SOL_AWC; Available water content (mm/mm), SOL_K; Saturated hydraulic conductivity (mm/h), SOL_CBN; Organic Carbon content (%), SOL_ALB; Albedo, USLE_K; USLE soil erodibility factor.

5.1.4. Slope Definition

One of the data required by the model is the definition of slope classes. DEM, previously provided in watershed delineator step, was used as input parameter for this classification. The model allows for up to 5 slope classes. These classes have a great impact in HRU generation. For this study, the slope classes were divided into 3 classes as <5%, 5-12% and >12% (Figure 5.4). These classes constitute 28.67%, 38% and 33.33% of the catchment, respectively.

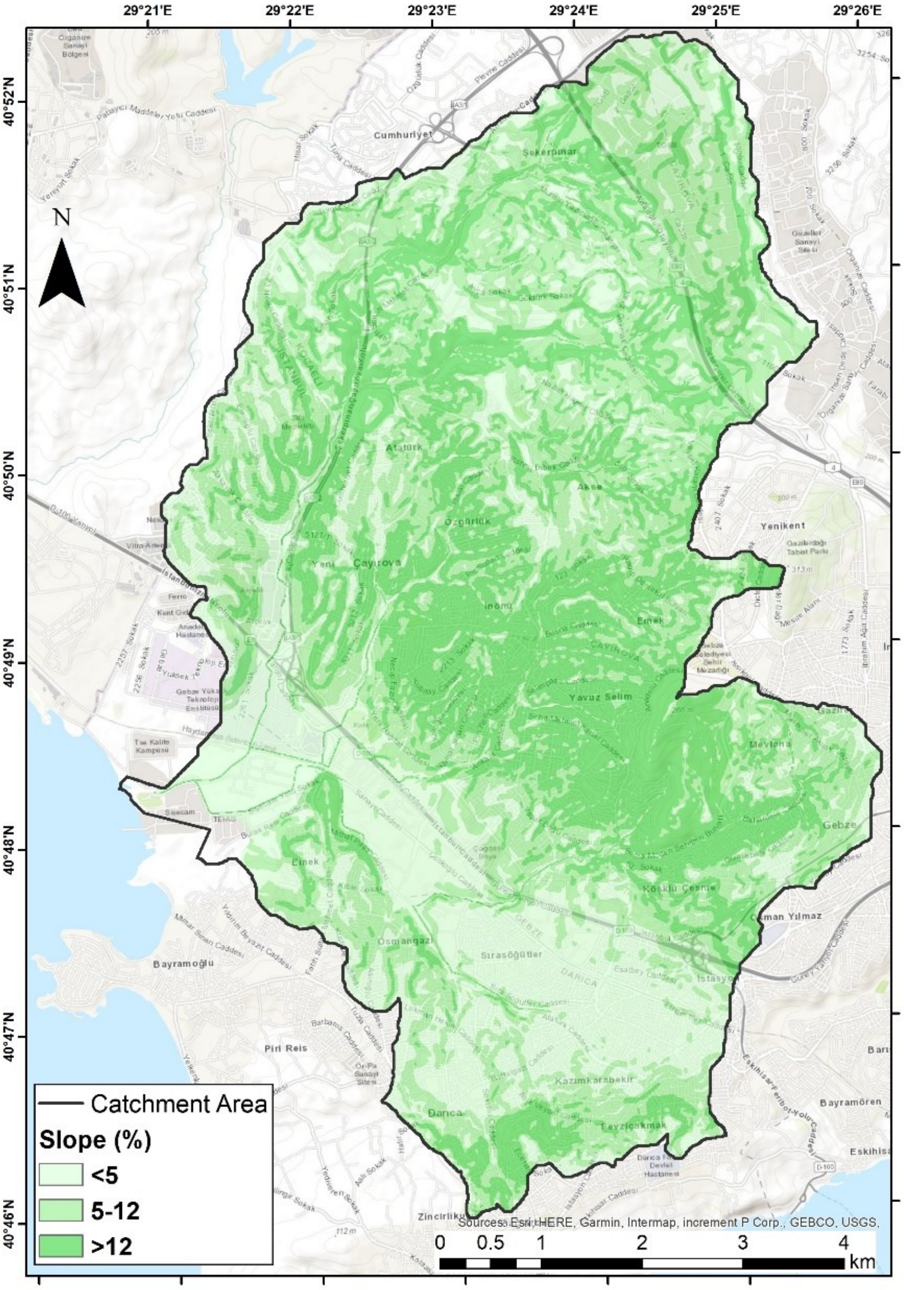
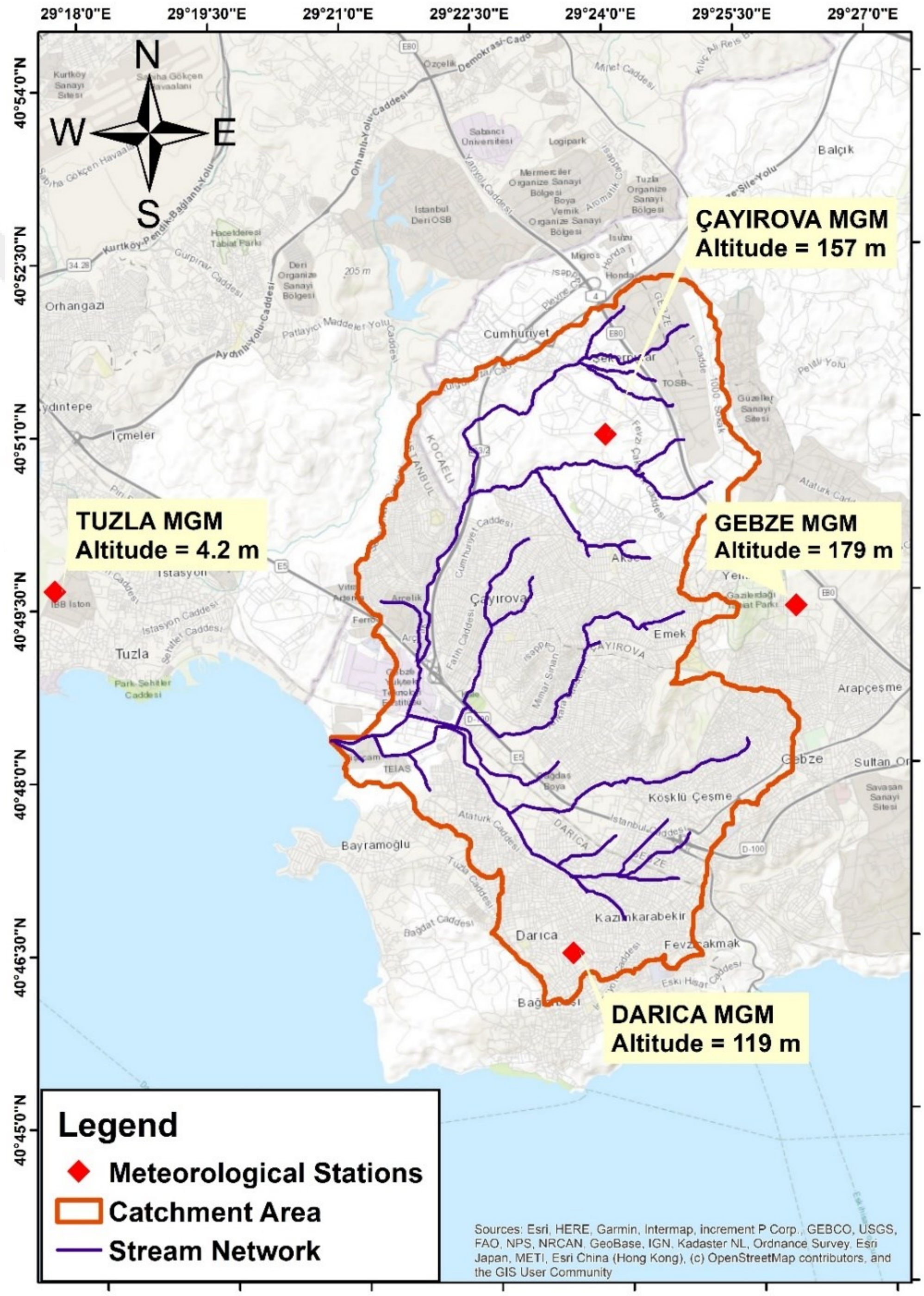


Figure 5.4: Slope map of the catchment.

5.1.5. Meteorological Data

The Meteorological data derived from the stations located at Çayırova (40.848452, 29.398757), Gebze (40.822998°, 29.434191°), Tuzla (40.827800°, 29.293090°) and Darıca (40.773599°, 29.389993°) were used to represent the whole catchment (Figure 5.5).



Sources: Esri, HERE, Garmin, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), (c) OpenStreetMap contributors, and the GIS User Community

Figure 5.5: Location of meteorological stations in Saz-Çayırova catchment.

The Thiessen Polygon method was applied to interpret the variability of observed and predicted runoff with precipitation at SCD-0-1 point (catchment outlet). As a result of this method, the meteorology station that best represents the catchment is Çayırova station of the Turkish General Directorate of Meteorology (MGM).

Wind speed and direction values of 4 meteorological stations representing all directions of the catchment were taken into consideration. Wind rose was composed, using R programming language, in order to determine the dominant wind direction of the Saz-Çayırova catchment (Figure 5.6). Dominant wind direction of the watershed is northeastern (about 58%). Besides, southern (13%) and south-eastern (17%) wind direction was introduced. The wind speed affecting the catchment is relatively low. Considering the dominant wind direction, it is seen that the wind speed of 3-4 m/s is dominant over the catchment.

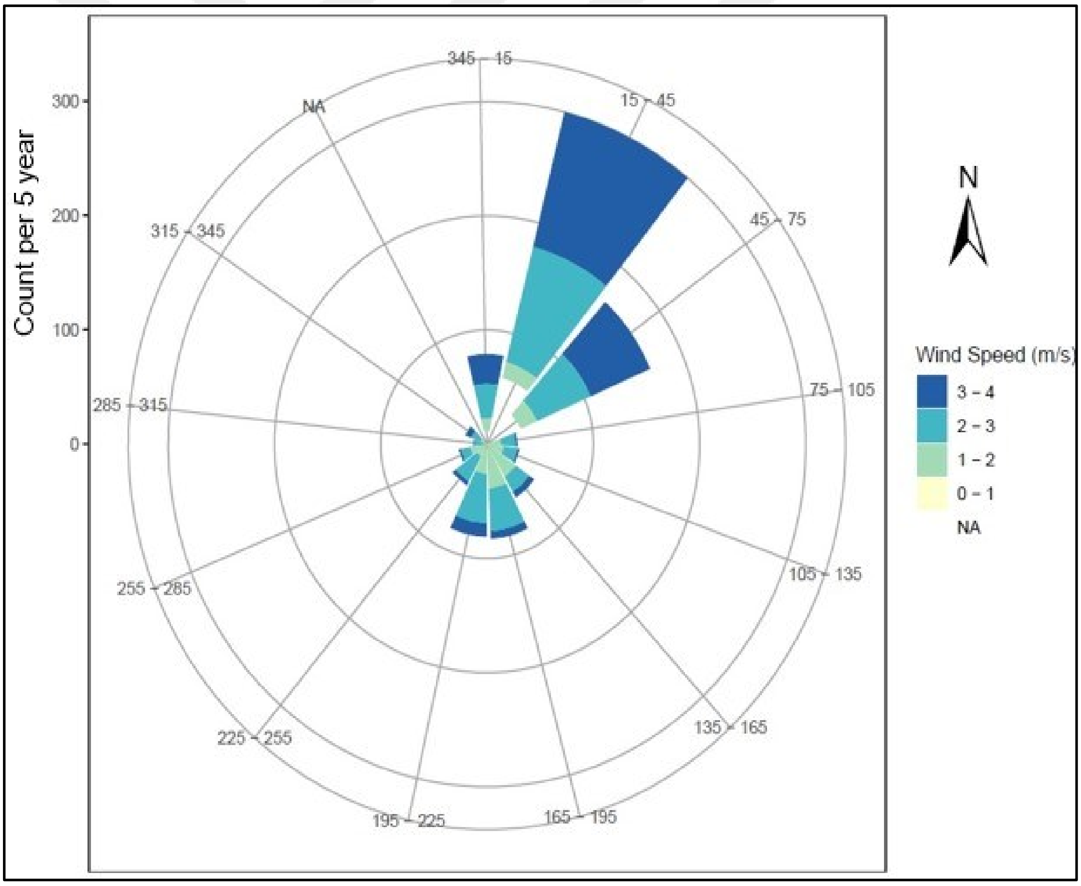


Figure 5.6: Wind speed (m/s) and direction (°) for Saz-Çayırova catchment.

Meteorological data that play an important role for hydrological processes were obtained from meteorological stations of Turkish General Directorate of Meteorology (MGM), located in the catchment area. In addition, the meteorological data produced by the National Environmental Prediction Center (NCEP), which operates in connection with the National Oceanic and Atmospheric Office (NOAA) established to investigate the weather and sea events in the world, was used. The meteorological data obtained from MGM was hourly resolution. This data was converted to daily resolution, using R programming language, and integrated into the SWAT. The provided parameters are precipitation, maximum temperature, minimum temperature, solar radiation, wind speed and relative humidity. The dew point temperature, the number of dry days, the probability of wet days required for the modeling produced using the obtained from MGM. In addition, NCEP re-projection data used for the days that no measurements exist for MGM stations.

The two most important parameters that illustrate the climate dynamics of any region are precipitation and temperature. For this reason, the temporal variation of the obtained temperature, dew point temperature and precipitation parameters for the working period in the 4 MGM stations were examined in monthly periods (Figure 5.7). Variation of temperature and dew-point temperature have similar seasonal trends for 4 stations. While small differences were observed between the two meteorological parameters during the winter season, high differences were observed in the summer season. For instance, while the average temperature value of 4 stations was 5 °C in the winter season, the average dew point temperature value was recorded as 4 °C. In summer, the average temperature value was recorded as 25 °C, while the average dew point temperature value was 17 °C. In this context, the lower difference between the two parameters in winter is related to the amount of moisture in the atmosphere. In terms of monthly average precipitation, the variation of values at the 4 stations has a similar trend to temperature values. For example, similar peaks occurred at 4 MGM stations for episodes seen on October 13rd 2016 and on November 24th 2018. On May 25th 2019, the precipitation at Darıca station was approximately 5 mm, while precipitation was recorded approximately 1 mm at the other stations. The measured monthly maximum rainfall was 7 mm in a 5-year period. In this context, the average monthly rainfall is relatively low (approximately 2 mm) in the region.

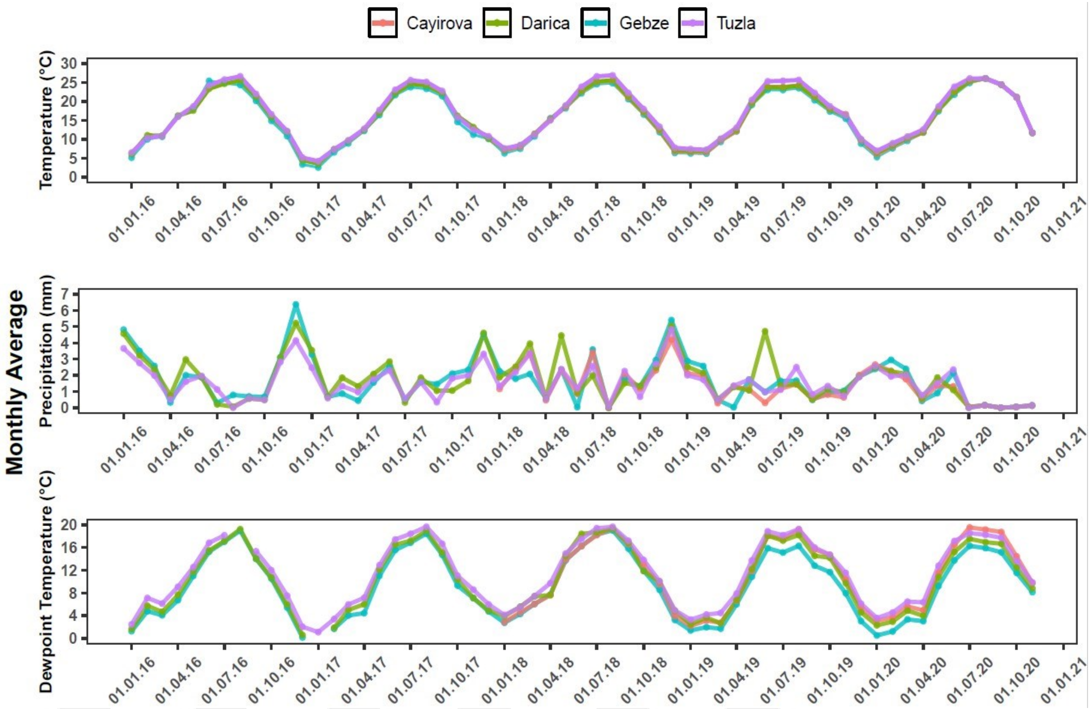


Figure 5.7: Variation of precipitation, temperature and dewpoint temperature measured at the four meteorological stations.

Examining the variation between parameters obtained from selected weather stations is important to understand the dominant climatic conditions prevailing in the studied catchment. Thus, pearson correlation coefficients were examined among the 4 MGM stations for all meteorological parameters. Among the selected weather stations, pearson correlation coefficients for temperature values vary between 0.9 and 0.96. There is a similar situation for dew point temperature values (0.81 - 0.92). For precipitation, the highest correlation was observed between Çayırova and Gebze stations (0.84). The lowest correlation was calculated between Çayırova and Darıca stations (0.57). However, Pearson correlation coefficients are generally affected by outlier values. In this context, the temporal variation of all examined meteorological parameters showed similar trends in 4 stations. This can be explained by the low altitude differences among the location of meteorological stations.

The obtained meteorological data consists of a 5-year data set covering the years 2016-2021. The values of each parameter were prepared as a .txt file in a single column in the month and year order. Afterwards, the statistical values of the meteorological data were calculated and integrated into the "WGEN file".

Calibration and validation processes must be determined for simulation of the model, after all data of the SWAT model has been successfully inserted. Modeling period is determined according to the time interval of meteorological input data. For this reason, the modeling period was set between January 1st 2016 and December 1st 2020. The first year, 2016, was determined as the warm-up year for the modelling and a 5 year of simulations was carried out until December 1st 2020.

5.2. Examination of Model Outputs

5.2.1. Hydrological Response Units (HRUs) Definition

Hydrological response units (HRUs) must be defined, after determining the catchment boundary, sub-basins, the land use, soil layer and slope classes belonging to the study area, as listed above (Figure 5.8). Hydrological response units are the smallest area in a homogeneous structure created on the base of land use, slope and elevation properties in catchment. Neitsch et al., (2011) stated that the number of sub-basins defined by the model is insufficient to examine the unique features produced in the dataset [34]. Thus, HRUs definition is critical in order to examine the unique characteristics of the watershed.

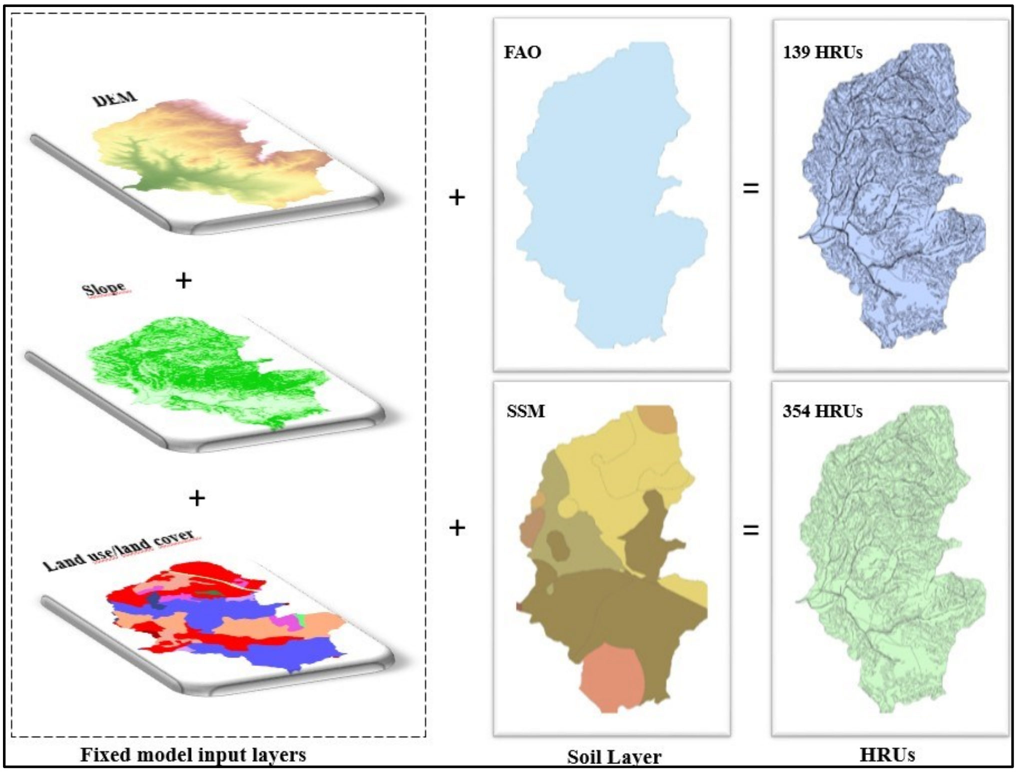


Figure 5.8: Hydrological response units (HRUs) definition for Saz-Çayırova catchment.

In this study, the default threshold value of 200 ha was given in the watershed delineation section of the model and 80 and 156 HRUs were defined for FAO and SSM soil databases, respectively. However, as a result of field observations and GIS analysis, this was found to be insufficient and the threshold value was reduced to 60 ha. Thus, 139, 354 HRUs were defined for the FAO and SSM soil databases, respectively. The model gives the opportunity to define a threshold percentage for land use, soil, and slope that could be included in the model during HRU definition. In the model established for SSM and FAO soil databases, threshold values were set as 15% for each layer.

5.2.2. Evaluation of the Model Results Without Calibration

The model results obtained using both soil databases were compared before applying calibration processes. The examination of the uncalibrated model outputs is significant because of the fact that calibration processes mask differences of the model results of the both soil data sets. According to the model results, surface runoff was better estimated by using SSM. (Figure 5.9).

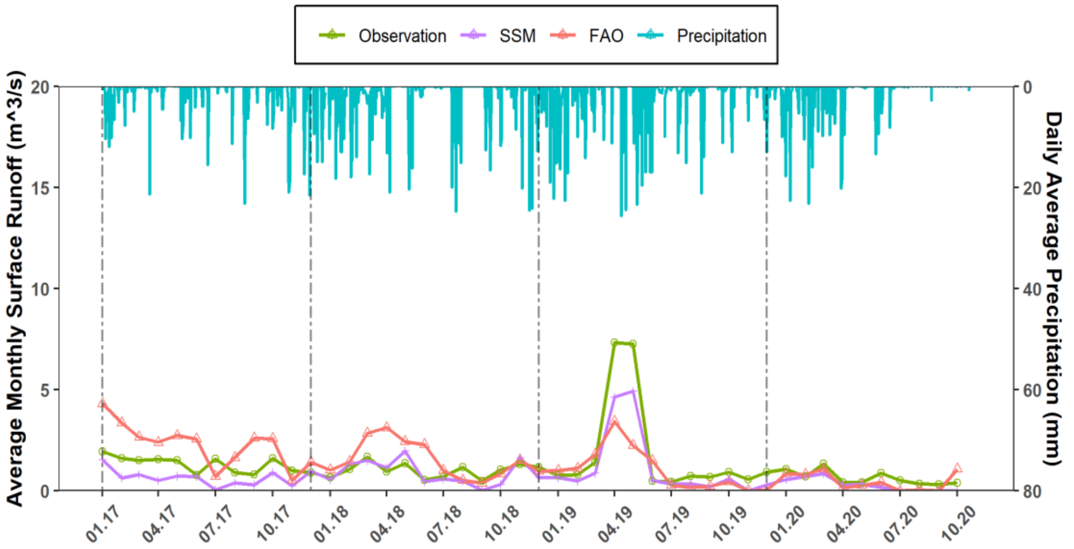


Figure 5.9: Average monthly time series of observed and predicted flow rate values that produced using FAO and SSM before calibration.

Monthly average runoff values were underestimated by the model established using the SSM, while they were overestimated by the model established using FAO database. On the other hand, predicted runoff values for the period between April and

June 2019 were underestimated. This episode is thought to be associated with the anthropogenic activities such as heavy industrial, residential discharges, as stated in Chapter 4.1. Two obvious peaks appear in the observed flow rate data in this episode. In this period, two peaks having low values were also observed in the results of the model with SSM database, while there is one peak in the results of the model with FAO database.

The determination coefficient (R^2) between the surface runoff values predicted using the SSM and the FAO database was calculated as 0.15. A possible explanation for this result can be the high difference in physicochemical soil parameters such as hydraulic conductivity, erodibility factor, clay content, dominant hydrologic group, which directly affects the surface runoff.

In the seasonal approach, the higher average monthly surface runoff was observed in the winter season (December, January, February) and the lower values were observed in the summer season (especially July), as expected. Periodically, apparent episodes were observed in the spring in 2018 (average runoff = 2.56 m^3/s) and in March and in April 2019 (average runoff = 7.43 m^3/h). The episode in 2018 can be explained by the average monthly precipitation data obtained from the 4 meteorological stations. While the first episode in March 2019 correlated with meteorological events, the second episode could be associated with local natural processes or anthropogenic activities of the catchment. Besides, the temporal variability of surface runoff in winter and autumn is more consistent than those of summer and spring for the both of two the soil database. Considering the variation of surface runoff generated from SSM and FAO relative to observation values, average monthly runoff oscillation is more similar in 2020 compared to other examined years. This interpretation was supported by the R^2 values between observed and predicted data of each year. The R^2 value calculated for 2020 is 0.46 and 0.68 for the FAO and SSM databases, respectively, while it varies between 0.2 and 0.35 for other years.

The relationship of observed and predicted flow rates with precipitation has high variation. This relationship is independent from each other especially in the autumn and spring season, while they have similar trends during the winter months. While the R^2 value between precipitation and observed runoff values is 0.26, it is approximately 0.11 for the precipitation and predicted runoff values. The main reason for the low determination coefficient values can be explained by the fact that the studied catchment is dominated by industry and residential areas (about 73%).

The both simulation results were compared with the observed runoff values without applying any calibration technique in order to introduce the results, straightforwardly (Figure 5.10). The proximity of the scattered points to the 1:1 (y=x) line shows the predictive performance of the model. The first noticeable case in the plot is that the model predictive performance is insufficient in daily low runoff conditions. While NSE and R^2 values for SSM were calculated as 0.21 and 0.17, the same values for FAO were calculated as 0.11 and 0.08, respectively. This shows that predictive performance of the model established with SSM was relatively better than the model established using FAO. Additionally, daily surface runoff examination of the SWAT model was not recommended in most research [46], [68], [69]. The results of this study obtained before the calibration support this explanation. The second case is that relatively high runoff conditions observed in 2019 was generally underestimated in both simulations.

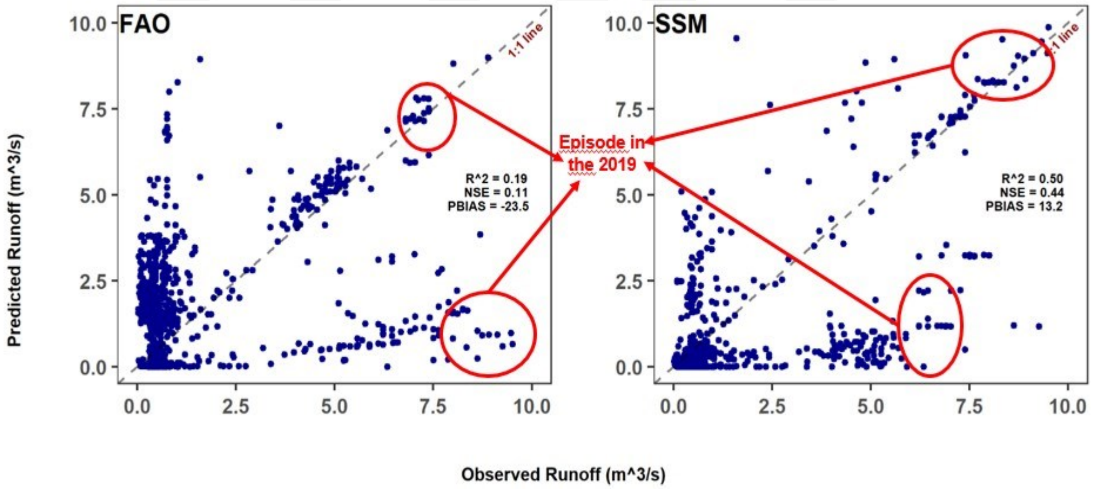


Figure 5.10: Scatter plot of observation and uncalibrated prediction runoff values produced from SSM and FAO (red circle represent the episode seen between May and July, 2019).

Determination coefficient (R^2), Nash-Sutcliffe coefficient of efficiency and The Percentage error statistics (PBIAS) values were used to evaluate the quality of model results without applying any calibration techniques. Motovilov et al., (1999) stated that if the NSE value produced after calibration was between 1 and 0.75, the flow estimation was excellent and if the NSE value was between 0.75 and 0.36, the flow estimation was acceptable [69]. For this study, the NSE value calculated as 0.19 and

0.50 for FAO and SSM, respectively. In this respect, the SSM database was making acceptable estimates even before calibration. The NSE values between the produced runoff values using the FAO database and the observation data remained below the acceptable threshold value (0.36). The Percentage error statistics (PBIAS) indicate the closeness of the negative or positive distance of model simulation values to observed values. PBIAS values are -23.5 and 0.2, for FAO and SSM respectively. Additionally, it has been observed that the runoff values estimated from both soil databases have a significant relationship with the observation data, but the outlier values affect the determination coefficient value as they directly affect the model performance.

5.2.3. Evaluation Model Results After Calibration

The SWAT-CUP (SWAT Calibration and Uncertainty Procedures) program was used for the calibration of the model. The SWAT-CUP program is the most preferred software for automatic calibration of SWAT model outputs. Calibration, uncertainty or sensitivity analysis of SWAT model outputs can be done easily with the SWAT-CUP program. In the program, SUFI-2 (Sequential Uncertainty Fitting) algorithm was used among 5 different optimization methods for calibration steps [70].

In the calibration process, the most critical variable is known to be runoff. After runoff calibration, other hydrological parameters such as evaporation, sediment transport or water quality variables are calibrated [68], [71]. Calibration processes were applied assuming that each subbasin is independent from each other. Calibration processes can be performed with daily, monthly and annual average surface runoff data in SWAT-CUP. In this study, calibration and validation processes were carried out using the daily average surface flow values. The SWAT-CUP program was run 5 times (500 iterations in each run) for each subbasin and a total of 2500 simulations were carried out. The measurement periods of SCD-0-2 and SD-0-1 are February 24th – December 1st and July 17th –December 1st, respectively. Considering this short period, only calibration procedures were carried out at these points. After the calibration, the model output was validated. In literature studies, it was recommended to separate the calibration and validation processes as 70% and 30% of the studied period in the dataset, respectively [34], [72]. Validation and calibration processes are performed at the same measurement point. Validation is generally preferred apart from the calibration period. In this study, the validation process was carried out only at the

SCD-0-1 point, representing the outlet of the catchment, considering the studied period between January 1st, 2020 and October 23rd, 2020.

Detailed explanation of all sources of uncertainty, observation data error, structure of the model, setup procedures in SUFI-2 algorithm has been detailed by Abbaspour et al., (2007). The ranges of the selected parameters, their final values, and 16 flow parameters were used in the calibration processes in order to obtain the best fit between the simulated and observed daily runoff were given in Table 5.4.

Table 5.4: Inventory of SWAT-CUP sensitivity parameters, descriptions, calibration ranges and best fitted parameter values of surface runoff for FAO and SSM database.

Streamflow Parameters	Definition	Calibration Range		Best fitted values	
		Min	Max	SSM	FAO
r__CN2.mgt	USDA Soil Conservation Service (SCS) curve number for moisture condition II	-0.5	0.5	-0.35	-0.23
r__ALPHA_BF.gw	Base flow alpha factor (days)	-0.5	0.5	-0.125	-0.103
r__GW_DELAY.gw	Ground water delay (days)	-0.5	0.5	-0.25	-0.12
r__GWQMN.gw	Threshold depth of water in the shallow aquifer required for return flow to occur (mm)	-0.5	0.5	0.35	0.28
r__SOL_K().sol	Soil saturated hydraulic conductivity (mm h^{-1})	-0.5	0.5	-0.15	-0.13
r__USLE_K().sol	Soil erodibility factor in USLE	-0.5	0.5	-0.35	0.31
r__SOL_AWC().sol	Available water capacity of the soil layer (mm mm^{-1})	-0.5	0.5	0.35	0.32
a__OV_N.hru	Overland Manning roughness	0	0.8	0.2	0.6
r__HRU_SLP.hru	Average slope steepness (m m^{-1})	-0.2	0.2	0.12	-0.14
a__ESCO.hru	Soil evaporation compensation factor	0	1	0.45	0.21
r__SURLAG.bsn	Surface runoff lag coefficient	-0.5	0.5	0.35	0.13
r__SOL_BD().sol	Soil bulk density (g/cm^3)	-0.5	0.5	-0.15	-0.12
r__RCHRG_DP.gw	Deep aquifer percolation fraction	0	1	0.25	0.34
r__SOL_Z().sol	Depth from soil surface to the bottom of layer (mm)	-0.5	0.5	-0.05	-0.12
a__CH_K2.rte	Effective channel hydraulic conductivity (mm h^{-1})	0	150	122.5	103.6
a__CH_N2.rte	Manning coefficient for main channel	0.01	0.2	0.198	0.093

[1] Type of change: r_ = relative change, a_=absolute change

While selecting the calibration parameters, both soil databases have been modified so that the relative error value is below 5%. The SCS curve number for moisture condition II (CN2), soil available water capacity of the soil layer (SOL-AWC) and soil evaporation compensation factor (ESCO) are the default water balance parameters to calibrate the surface flow and suggested in the SWAT manual. The determination coefficient value (R^2 = 0.50), before calibration, was in the acceptable range with using the SSM, while the value was low using the FAO soil database (R^2 = 0.19). For this reason, it has been tried to select parameters that can affect the results obtained from both databases together. Calibration range of CN2 was adjusted between -0.5 and 0.5. After that, the best fitted value for the SSM and FAO was -0.35 and -0.23, respectively. The ESCO was assigned as an absolute value in all HRUs. The SOL-AWC was analyzed with variations of 50% to find the best fitting value. All other parameters were adjusted considering literature studies.

5.2.3.1. Applying Calibration Process at Station SCD-0-2 and SD-0-1

The calibration of daily flow measurements was performed using the flow rate values measured at the stations SCD-0-2 and SD-0-1 (Figure 5.11). Calibration processes for the SCD-0-2 point were carried out between February 24th 2020 and December 1st 2020. The two models established results were found to underestimate the observed runoff values. In an episode seen in June, the runoff produced from the FAO database also appears as a noticeable peak. In the episode observed on August 21st, the model results were seen consistent with the observed runoff. The main reason for this can be explained by the recorded high precipitation event (8 mm) on August 21th. Another point that stands out in the model results is that the values obtained for autumn period are quite low. While the average of the observation values was about 0.49 m^3/s during the Autumn, it was estimated as 0.08 and 0.07 m^3/s by the models established using FAO and SSM, respectively. The main driving reason for this underestimation can be associated with relatively low total mean precipitation (4.59 mm) value during of this 3-month.

Calibration processes at station SD-0-1 were carried out between July 17th and December 1st 2020. The most noticeable difference between the models established using FAO and SSM was seen on August 17th. While the observed flow value was measured as 0.18 m^3/s, the produced flow values by the models established using FAO and SSM were found to be 3.33 m^3/s and 0.13 m^3/s, respectively. The observed flow values have a better fit with the model established using SSM compared to the model established using FAO. SD-0-1 represents the Subbasin-6 output (Seen in Figure 4.1). One of the distinguishing characteristics of this subbasin is that about 72% of the soil texture is sand. Considering that this information is in the SSM database and that the only difference between the two models is the soil layer, it was not surprising that the model established using SSM better explain the episode seen on August 17th in subbasin-6.

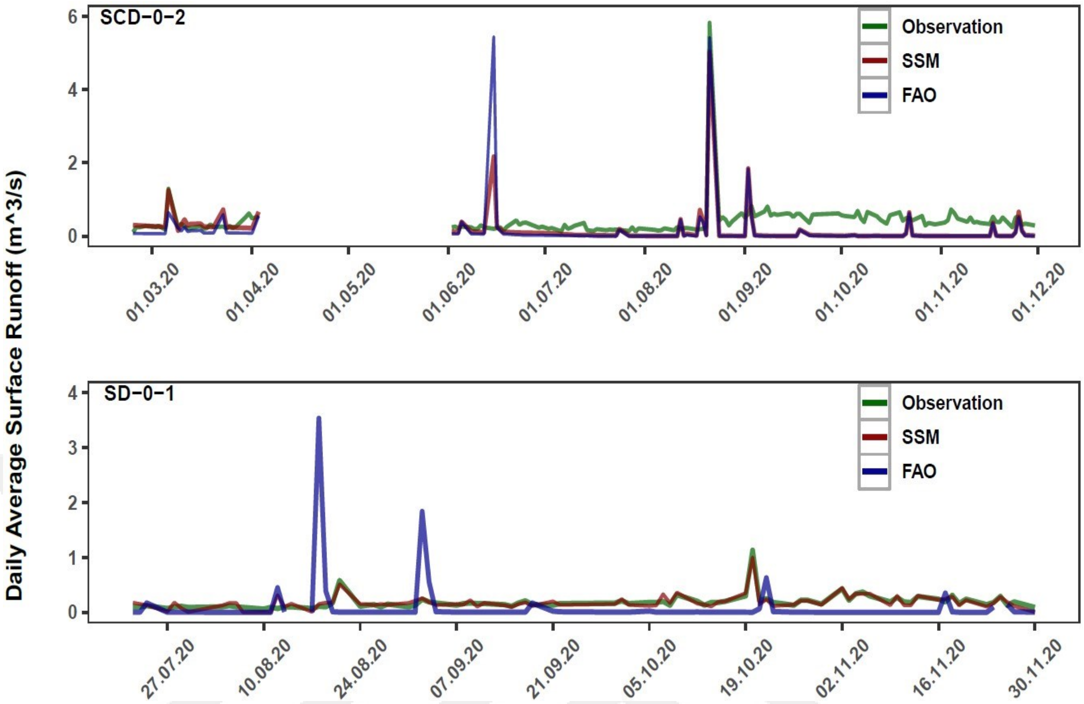


Figure 5.11: Comparison of daily observation and prediction surface runoff at the Subbasin-7 (SCD-0-2) and Subbasin-6 (SD-0-1) outlet.

Comparison of calibrated model results and observation runoff at SCD-0-2 and SD-0-1 monitoring stations are detailed in Figure 5.12. In the SCD-0-2, more compatible results were detected in the relatively low runoff values predicted using SSM. In the established model using the FAO database, the outlier value seen at low runoff values has a direct effect on the predictive performance of the model. Furthermore, the results of the both model setups using the two different soil databases were considerably compatible with high values. This situation indicates that predictive performance of the model is better at comparatively high values, such as at the catchment outlet (SCD-0-1).

Considering the model results obtained for station SD-0-1, it is seen that the performance of both models are better than the results obtained at the station SCD-0-2. In the model established using FAO, runoff values were more deviated from the 1:1 line compared to the model established using SSM. In the established model using the SSM, excellent fitting was analyzed for extreme runoff values. High runoff effects the surface flow conditions and low runoff effects the base flows because of increasing infiltration period. In the model established using SSM, the observed relatively good results for low runoff conditions indicates that the model prediction performance at the base flow better than the high runoff conditions.

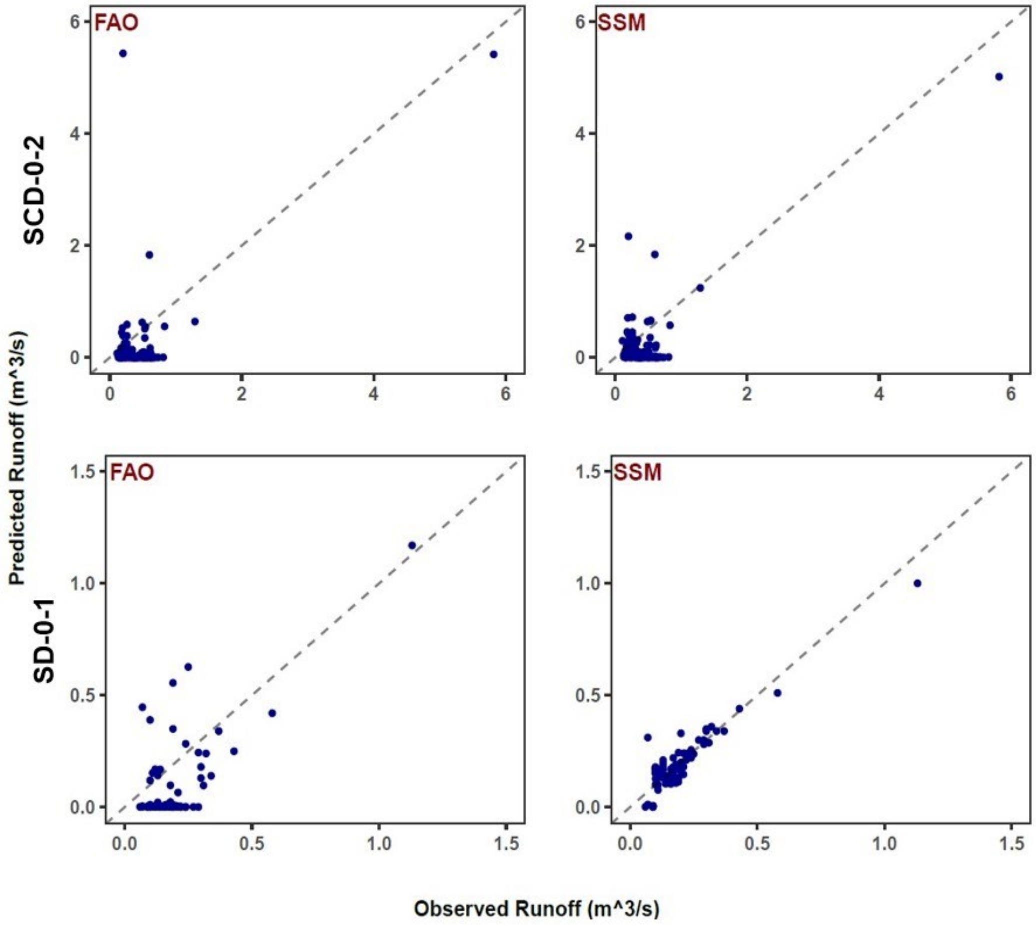


Figure 5.12: Evaluating the calibration of runoff at the SCD-0-2 and SD-0-1.

5.2.3.2. Applying Calibration and Validation Processes at SCD-0-1

The monthly temporal variation of calibrated and validated runoff values produced by using FAO and SSM databases with observation data was illustrated in Figure 5.13. Firstly, peak runoff values which formed after precipitation were consistent with model results, depending on the increasing amount of precipitation. Furthermore, predicted baseflow which formed after the rainfall were observed compatible with the decrease of the observed runoff. Relative deviation was observed during the decrease of flow. The main reason for this is that the SURLAG parameter, which is the surface flow delay coefficient, cannot be determined separately for each sub-basin, but a single parameter value can be entered into the mode for the whole basin [72].

Two relatively extreme runoff values (peak) were observed in 2019, as mentioned in the Chapter 5.2.2. The model results established with SSM are compatible with the second peak. Other hand, two peaks were seen in the model results established with FAO. However, these values are lower than the observation. Therefore, although the results of the model established with SSM has a single peak, a closer prediction was made. In the seasonal approach, the results of the model established with SSM better predicted the runoff values in summer and winter compared to the model results established with FAO. For instance, the model results using SSM and the high flow rates measured on June 15th, 2017 were quite compatible.

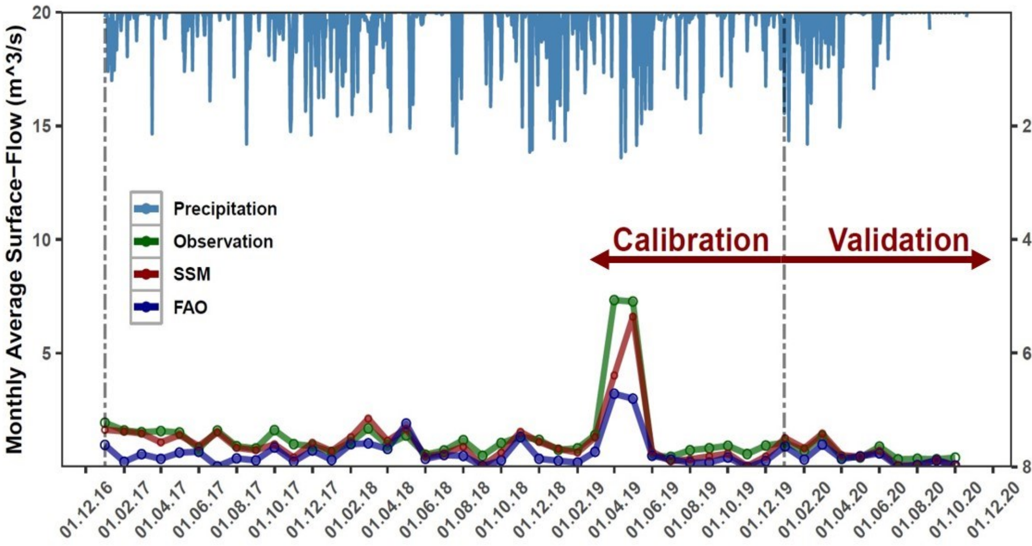


Figure 5.13: Comparison of monthly observation and prediction surface runoff in the catchment outlet and their variation with daily average precipitation.

The scatter plots of the calibrated model results produced using the FAO and SSM database comparing observation values measured at SCD-0-1 are presented in Figure 5.14. Considering the modeled runoff results using FAO database, the prediction scale of runoff values after calibration was degraded to a lower range. While the observed flow values change between 0.1 m^3/s and 1.5 m^3/s, the calibrated flow values change between 0.1 m^3/s and 5 m^3/s. Corresponding to the range of these observed flow values, uncalibrated flow values changes between 0.1 m^3/s and 9.2 m^3/s. Besides, better model performance was observed at high runoff conditions (7.5 m^3/s - 10 m^3/s) compared to low runoff conditions (< 2.5 m^3/s) in model established using FAO.

Considering the modeled runoff results using SSM database, no significant difference was found for the lower values before and after calibration. While the observed flow values change between 0.1 m^3/s and 2.5 m^3/s, uncalibrated and calibrated flow values change between approximately 0.1 m^3/s and 5 m^3/s. In addition, the high runoff values observed in 2019 were relatively closer to the 1:1 line, after calibration. In general terms, it can be said that calibration parameters selected were effective at low runoff values (0.1 m^3/s - 2.5 m^3/s) for the model established using FAO database, while they were effective at medium (2.5 m^3/s - 5 m^3/s) and high (7.5 m^3/s - 10 m^3/s) flow values for the model established using SSM.

The R^2 value between the uncalibrated and calibrated flow values produced from the two soil databases was found to be 0.42 and 0.52, respectively. These relatively high R^2 values can be explained by the close values of soil parameters such as organic carbon, silt content in both databases.

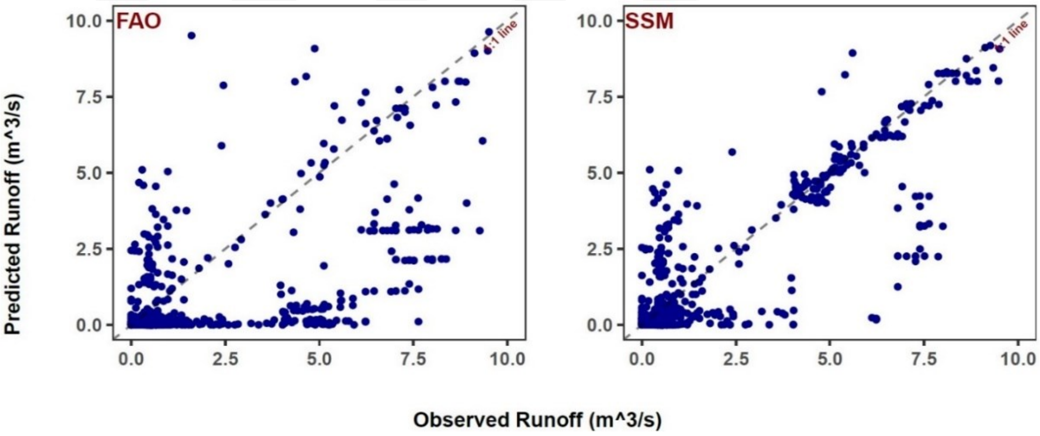


Figure 5.14: Comparison of daily observation and calibrated prediction values produced from SSM and FAO soil database.

Following the calibration procedure, the predicted values were validated using the observed values at the station SCD-0-1 between January 1st and October 23rd 2020 (Figure 5.15). Model results established with FAO show that the observation values are better predicted at high flow conditions than at low flow, considering 1:1 line. The prediction performance of the model using the SSM soil database, was seen excellent at high flow rates. Contrarily, the estimation performance found to be low for runoff values between 0.1 and 1 m^3/s.

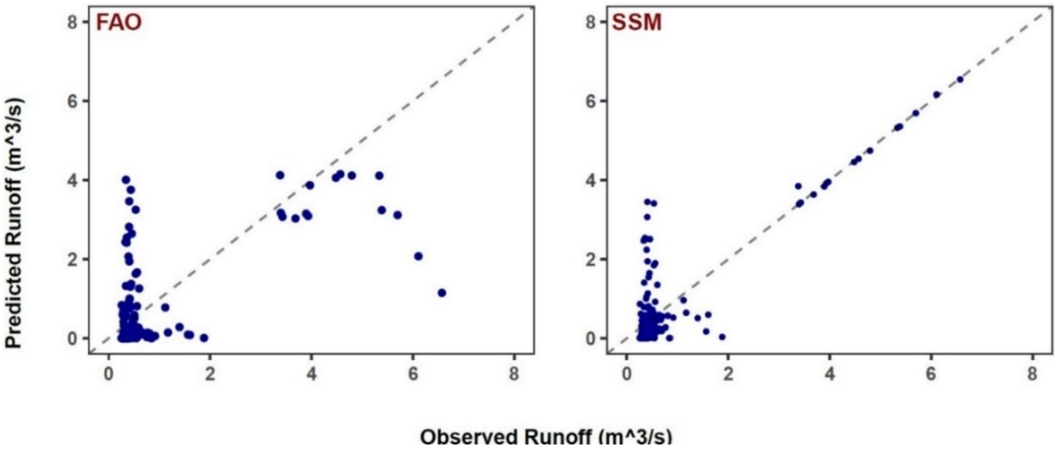


Figure 5.15: Evaluating the validation results of runoff values produced using FAO and SSM database.

The performance statistics between the runoff values produced by using SSM and FAO databases and the flow values measured at SCD-0-1, SCD-0-2 and SD-0-1 stations are shown in Table 5.5. Considering the established model with the SSM, the R^2 value, which indicates the dimension between simulation and observation values during the calibration phase, was found as 0.83 and 0.78 at the SD-0-1 and SCD-0-1, respectively. At the SCD-0-2, the R^2 value is lower than the other stations (0.59). For the SCD-0-1, NSE value was in the acceptable range (0.36 - 0.75) even before calibration. After calibration, the obtained value was relatively high (0.66) although still in the same range. The driving reason for this situation is the direct effect of daily extreme oscillation on the performance of the model. NSE value was found in the excellent value range (> 0.75) at station SD-0-1. PBIAS values were found to be -14.3, -22.3 and -4.32 for SCD-0-1, SCD-0-2 and SD-0-1, respectively. PBIAS was found in the "very good" range for the SD-0-1, while it was classified as "acceptable" for other stations. Moreover, the validation results for the SCD-0-1 were consistent with the calibration results.

Considering the model results using the FAO database for SCD-0-1, SCD-0-2 and SD-0-1, R^2 values were found to be 0.43, 0.38 and 0.49, respectively. NSE values for the same stations calculated as 0.35, 0.22 and 0.37, respectively. The calculated PBIAS values were 22.31, 37.2 and 25.2, orderly. According to these results, the SD-0-1 point was found in the "acceptable" range. NSE values, below the acceptable threshold before calibration in the established model outputs with FAO, fell into acceptable range after calibration. This situation is directly related to the reduction of the estimation range at low flow values before calibration with the sensitivity parameters determined in the calibration step. Validation results for the SCD-0-1 were also classified as “acceptable” and appeared to be consistent with the calibration results.

Table 5.5: Statistical evaluation of the SWAT model outputs in the calibration and validation process.

Station	SSM						FAO					
	Calibration			Validation			Calibration			Validation		
	R^2	NSE	PBIAS	R^2	NSE	PBIAS	R^2	NSE	PBIAS	R^2	NSE	PBIAS
SCD-0-1	0.78	0.66	-14.3	0.76	0.63	-12.1	0.43	0.35	22.31	0.41	0.26	-15.1
SCD-0-2	0.59	0.43	-22.3		-		0.35	0.28	-37.2		-	
SD-0-1	0.83	0.77	-4.32				0.37	0.32	29.2			

The outputs of the model established using SSM appear higher prediction performance than the outputs of the model established using FAO at all monitoring stations. Among all observation stations and both established models, the best runoff prediction has been analyzed at the SD-0-1 monitoring station, which represents the Subbasin-7 outlet (Figure 4.1). This situation can be explained by the fact that only the Saz stream is represented at this station and Çayırova stream, which is affected more by industrial loads, joins the Saz stream after this point.

In summary, the SWAT model results were evaluated before and after calibration. Firstly, the monitoring station, namely SCD-0-1, which represents the outlet of the catchment, was used to evaluate the model results without calibration. NSE value (0.44) calculated between the predicted runoff established model using SSM and observation runoff was in acceptable range, while this value (0.36) was below the threshold value for the established model using FAO. Second, the model results were re-evaluated by applying calibration. After calibration, the model results showed that the performed better at SCD-0-1 monitoring station than FAO. Furthermore, the episode seen in 2019 after model calibration was found to be relatively well predicted in the model established using the SSM database. This prediction is a good example of elaborating the soil database to explain episodes over the studied catchment. Calibration procedures were also applied for SCD-0-2 and SD-0-1 monitoring stations, which represent Subbasin-7 and Subbasin-6 outlets. Statistically, the highest model performance was observed with SSM at the SD-0-1 monitoring station, representing the runoff coming from the Saz stream (NSE: 0.77, R^2: 0.83, PBIAS: -4.32).

6. CONCLUSION and SUGGESTIONS

In this study, Specified Soil Maps created by laboratory and field studies (SSM) and the international FAO soil database were used in order to evaluate the runoff prediction performance of SWAT in Saz-Çayırova stream, which is dominated by intensive industry and residential areas. For the SSM, twenty-nine soil samples representing the entire catchment were collected. Seven parameters were measured by field and laboratory studies. After that, fifteen different soil sub-categories were formed by configuring these parameters in the GIS environment. It is found that sand and silt content is dominant in the soil texture of the catchment. Relatively high clay content was found downstream (approximately 6%). This situation can be associated with the fact that stream energy decreases, thereby increasing the tendency of suspended solids to precipitate. Additionally, Erodibility factor values (USLE_K) were relatively high in the downstream of catchment and Çayırova stream networks (USLE_K> 0.5). Jan et al., (2009) stated that the sensitivity to water erosion is high in regions having erodibility factor values greater than 0.5 [73]. The most noticeable differences in the two databases are soil texture and saturated hydraulic conductivity (mm h^{-1}). While soil texture (sand, silt and clay) values are almost evenly distributed in the FAO database, the sum of the sand and silt content are formed about 78% of the SSM. Main reason for the difference is that the entire Marmara basin was lumped in only one soil type (Lc-105-2/3ab) in the FAO database. About 4 times difference of saturated hydraulic conductivity values in SSM and FAO is clearly related to different the sand contents of the two databases.

Studies evaluating the runoff prediction performance of the SWAT model using different global soil databases are available [25], [74], [75]. Mengistu et al., (2007) integrated STATSGO and SSURGO soil databases into the SWAT to evaluate the flow predictive performance in the Turkey Creek Watershed outlet (126 km^2). After calibration, NSE values for SSURGO and STATSGO were 0.72 and 0.61, respectively [75]. Similarly, Bhandri et al., (2019) tested these soil databases for Lower Cumberland watershed, Tennessee (242 km^2). After calibration, NSE values were calculated as 0.69 and 0.58 for SSURGO and STATSGO, respectively [74]. In our study, after calibration NSE values at the catchment outlet (SCD-0-1) were calculated as 0.66 and 0.35 for SSM and FAO, respectively. The first driving reason for this is that soil databases of the mentioned studies are highly elaborating databases embedded in the SWAT. Another possible reason for this situation is that the SWAT model was run for catchments dominated by agricultural areas in the mentioned studies, contrary to our study. Besides, both studies stated that elaborating soil databases better reflects the infiltration potential in the specific regions in the catchments. In our study, the infiltration potential around the SD-0-1 point was found to be high (having about 72% sand content of soil texture). Likewise, the better prediction performance of the model established with SSM for this region (Figure 5.14) supports this explanation.

A few studies on comparing the soil database specified by field and laboratory applications with the FAO database are also existing in the literature. Busico et al., (2019) integrated FAO (SL1) and two soil databases formed by field studies (SL2 and SL3) separately into the SWAT model [76]. The study area was classified as small-scale (155 km^2) and highly urbanized (approximately 43%) basin. SL1 was chosen as the best one in describing watershed streamflow, despite it being characterized by the more homogenous soil database. On the contrary, it has been recorded that detailed soil databases give better results in specific studies such as runoff forecasting and examination of flood events at sub-basin level. The second explanation supports our study results. Study conducted by Yassine et al., (2020) formed a specified soil database (TAMEDSOIL) for Tamedroust catchment, covering an area of 642 km^2. The SWAT model was established using the TAMEDSOIL and FAO database in order to evaluate the runoff prediction performance for the years between 1993 and 2009. Five soil parameters were analyzed by laboratory studies. After calibration the NSE value was 0.65 for both soil databases. Validation results were 0.74 and 0.76 for FAO and TAMEDSOIL, respectively [77]. In our study, better results were found after calibration in the SWAT model established with the SSM database. The most important reason for this is that the number of soil parameters analyzed in our study is higher. But in contrast, the model results established with FAO in this study are better than results obtained in our study. This situation can be attributed to the two different reasons. The first is that the Tamedroust basin is dominated by bare soil and agricultural areas. The second is that the running period of the SWAT in Tamedroust basin is 16 years which is longer than our study.

Preferable results were found in SWAT simulations using SSM database before and after calibration compared to established models using FAO. Notwithstanding, the soil map produced by laboratory and field studies requires high labor intensity.

Detailed soil databases containing the soil physico-chemical parameters needed by the hydrological models have not been created throughout Turkey. Since the available soil databases are prepared according to main soil groups, these databases reduce the reliability of hydrological model outputs. The origination of national soil database inventory and determining the soil physico-chemical properties throughout Turkey is critical because of the unique characteristics of the basins in Turkey. By this means, more accurate projections can be developed by researchers for the protection of our natural resources.

On the other hand, as a result of field studies, many point discharge sources have been identified along the Saz-Çayırova catchment. These sources could not be integrated into the model due to lack of data and their effect on the results could not be observed. This situation can be correlated to the relatively low predictions seen at the catchment outlet and SCD-0-2 monitoring station. Accordingly, open source data of point discharge sources is critical to reveal more clearly the physical and chemical processes in the studied watershed. Considering all these reasons, the performance of the SWAT model has been found to be good for the Saz-Çayırova catchment dominated by industrial areas.

In this thesis study, the runoff predictive performance of the model was tested using two different soil databases. Subsequently, sub-dynamics of the model were established to make simulation or forecasting. These sub-dynamics are important in revealing the possible effects of human-induced interventions such as population increase/decrease, reclamation studies, land use change on water quality or soil erosion/sediment yield in each sub-basin. In the future studies, it is concluded that better results can be obtained by integrating into the model discharges of point sources and longer term runoff values.

REFERENCES

[1] WWAD, (2019), "The United Nations World Water Development Report", UNESCO.

[2] USGS (2016), "Global water distribution", United States Geological Survey.

[3] Web-1, (2020), https://mgm.gov.tr/veridegerlendirme/yillik-toplam-yagis-verileri.aspx, (Acces Date: 26/10/2020).

[4] DSI, (2018), "General Directorate of State Hydraulic Works 2018 Annual Report, Ankara"

[5] KB, (2018), "10th Development Plan (2019-2023) Water Resources Management and Security", Turkey Ministry of Development, Ankara.

[6] Cüceloğlu G., (2013), "Model supported hydrological analysis of Darlık", İstanbul Technical University.

[7] Arnold, J.G, Srinivasan R., Muttiah R. S., Williams J. R., (1990), "Large Area Hydrologic Modeling and Assesment Part I: Model Development," Journal of the American Water Resources Association.

[8] SEI, (2011), "WEAP User Manual. Stockholm Environment Institute," USA.

[9] HEC, (2001), "Hydrologic Modeling System HES-HMS, User's Manuel," US Army Corps of Engineers Hydrologic Engineering Center, USA.

[10] DHI, (2007), "Mike She User Guide", DHI Software, Denmark.

[11] Bicknell B. R., Imhoff J., Kittle J., Jobes T., Donigian A.S., (2000), "Hydrological Simulation Program – Fortran User's Manual. Release 12", US EPA.

[12] Young R.A., Onstad A.S., Bosch D.D., Anderson W.P., (1989), "AGNPS - a Nonpoint-Source Pollution Model for Evaluating Agricultural Watersheds", Journal of Soil and Water Conservation.

[13] HEC, (2016), "HEC-RAS River Analysis System", US Army Corps of Engineers Hydrologic Engineering Center, USA.

[14] Atasayan Ö., Küçükyağcı P.Ö., Çalışkan Ö., (2013), “Çayırova (Saz) Deresi Çevresinin Doğal Peyzaj Dönüşümünün Değerlendirilmesi”, Landscape Architecture 5th Congress, Adana, Turkey.

[15] Hız H. (2000), "The pollution of Çayırova stream and its impact on the Marmara Sea", Gebze Technical University, Kocaeli, Turkey.

[16] Kanburoğlu E. (2002), "Çayırova Deresi'nde tespit edilmiş olan ağır metal iyonlarının Vicia faba L. kromozomları üzerine etkisi", Gebze Technical University, Kocaeli.

[17] Rossi C. G., Srinivasan R., Jiraynot K., Duc T. L, Sonvannabouth P., Binh N., Gassman P. W., (2009), "Hydrologic evaluation of the lower Mekong River basin with the Soil and Water Assessment Tool model", International Agricultural Engineering Journal.

[18] Güngör Ö., Göncü S, (2013), "Application of the soil and water assessment tool model on the Lower Porsuk Stream Watershed", Hydrological Processes.

[19] Pott C. A., Fohrer N, (2017), "Hydrological modeling in a rural catchment in Germany", Applied Research and Agrotechnology.

[20] Di Luzio M., Arnold J. G., Srinivasan R., (2005), "Effect of GIS data quality on small watershed stream flow and sediment simulations", Hydrological Processes.

[21] Güzel Ç., (2010), "Application of SWAT Model in a Watershed in Turkey", MSc. Thesis, İstanbul Technical University, İstanbul.

[22] Akıner M. E., Akkoyunlu A., (2012), "Modeling and forecasting river flow rate from the Melen Watershed, Turkey", Journal of Hydrology.

[23] Dengiz O., Göl C., Başkan O., (2007), "Soil Properties and Soil Mapping of Büyükçay Basin, Çankırı, Turkey", Artvin Coruh University Journal of Forestry Faculty.

[24] Aydınalp C., Arslan Y., (2003), "Antalya Havzasındaki Büyük Toprak Gruplarının FAO/UNESCO (1990), FitzPatrick (1988) ve Toprak Taksonomisi (USDA Soil Taxonomy, 1998) Sistemlerine Göre Sınıflandırılması", Anadolu J. of AARI.

[25] Levick L. R., Semmens D. J., Guertin D. P., Burns I. S., Scott S. N., Unkrich C. L., Goodrich D. C., (2004), "Adding global soils data to the automated geospatial watershed assessment tool (AGWA)", Second International Symposium on Transboundary Waters Management, Tucson.

[26] Setegn S. G., Srinivasan R., Melesse A. M. and Dargahi B., (2010), "SWAT model application and prediction uncertainty analysis in the Lake Tana Basin, Ethiopia" Hydrological Processes.

[27] Kroner S. M., Cozzie D. A., (1999), "Data Collection for The Hazardous Waste Identification Rule. Section 7.0" Soil Data US Environmental Protection Agency Office of Solid Waste.

[28] Bren L., (2015), "Forest Hydrology and Catchment Management" Springer Netherlands.

[29] Peschel J. M., Haan P. K., Lacey R. E., (2003), "A SSURGO Pre-Processing Extension 798 for the Arc View Soil and Water Assessment Tool" ASAE Paper no. 032123, St.799 Joseph, Michigan.

[30] Lenhart T., Eckhardt K., Fohrer N., Frede H. G., (2002), "Comparison of two different approaches of sensitivity analysis", Physics and Chemistry of the Earth, Parts A/B/C.

[31] Sajikumar N., Remya R. S., (2015), "Impact of Land Cover and Land Use Change on Runoff Characteristics", Journal of Environmental Management.

[32] Li L., Jiang D., Hou X., and Li J., (2012), "Simulated Runoff Responses to Land Use in the Middle and Upstream Reaches of Taoerhe River Basin, Northeast China, in Wet, Average and Dry Years", Hydrological Processes.

[33] Ravindra V. K., Bbabagrahi S., (2011), "Green-Ampt Infiltration Models for Varied Field Conditions: A Revisit", Water Resources Management.

[34] Neitsch, S. L., Arnold, J. G., Kiniry, J. R., Srinivasan, R., Williams, J. R. (2011), “Soil and water assessment tool: Input/output file documentation”, Soil and Water Research Laboratory.

[35] Leonard R., Knisel W., Still D., (1987), "Groundwater loading effects of agricultural management systems", Transactions of the ASAE.

[36] Knisel W. G., (1980), "CREAMS: a field scale model for Chemicals, Runoff, and Erosion from Agricultural Management Systems USA", Department of Agriculture and Conservation research report (USA).

[37] Williams J., Jones C., Dyke P. T., (1984), "A modeling approach to determining the relationship between erosion and soil productivity", Transactions of the ASAE.

[38] Wincell M., Srinivasan R., Di Luzio J., Arnold J., (2010), "ArcSWAT Interface for SWAT2009 User's Guide", Texas, ABD.

[39] Neitsch S. L, Arnold J. G., Kiniry J. R. and Williams J. R., (2009), “Soil And Water Assessment Tool Theoretical Documentation, Version 2009”, Texas, U.S.A: Agricultural Research Service and Blackland Research Center.

[40] Busteed P. R., Storm D. E., White M. J., Stoodley S. H., (2009), "Using SWAT to Target Critical Source Sediment and Phosphorus Areas in the Wister Lake", American Journal of Environmental Science.

[41] Lillesand T., R. Kiefer (2004), “Remote Sensing and Image Interpretation”, New York: John Wiley & Sons, Inc.

[42] CORINE (1997), "The CORINE Project Methodology", European Environmental Agency.

[43] Ferance J., Jaffrain G., Soukup T., Hazeu G. (2010), "Determining changes and flows in European landscapes 1990-2000 using CORINE land cover data", Agriculture, Ecosystems and Environment.

[44] Karagüllü E., Aydoğdu M., Dedeoğlu F., Peşkircioğlu M., Yıldız H., Yerdelen A. (2002), "Project of Determination of Agricultural Areas in Gaziantep Province Using Satellite Images", Tarla Bitkileri Merkez Araştırma Enstitüsü.

[45] Sharpley A. N., Williams J. R. (1990), "EPIC-Erosion Productivity Impact Calculator", U.S. Department of Agriculture, Agricultural Research Service, Technical Bulletin, USA.

[46] Abasspour K. C., Yang J., Maximov I., Siber R., Bogner K., J. Mieleitner, J. Zobrist, R. Srinavasan (2007), "Modeling Hydrology and Water Quality in The Pre-Alpine/Alpine Thur Watershed Using SWAT", Journal of Hydrology.

[47] Jajarmizadeh M., Harun S, R. Abdullah, M. Salarpour (2012), "Using Soil and Water Assessment Tool for Flow Simulation and Assessment of Sensitive Parameters Applying SUFI-2 Algorithm", Caspian Journal of Applied Sciences Research.

[48] Rostaminan R., Jaleh A., Afyuni M., Mousavi S. F., M. Heidarpour, A. Jalalian, K. C. Abbaspour (2010), "Application of a SWAT Model for Estimating Runoff and Sediment in Two Mountainous Basins in Central Iran", Journal of Hydro-Environment Research.

[49] FAO, (2007), "Digital Soil Map of the World," 28.02.2007. http://www.fao.org/geonetwork/srv/en/metadata.show%3Fid=14116.

[50] Williams R. H., (1995), "Effect of sheep, deer and cattle dung on herbage production and soil nutrient content," Grass and Forage Science.

[51] Özer Z., (1990), “Hydrological and hydraulic principles in the design of water structures,” Ankara: Ministry of Agriculture, Forestry and Rural Affairs, General Directorate of Rural Services Basin Improvement and Ponds Department.

[52] Katharina F., (2014), “Analysis and Modelling of Water Supply and Deman Under Climate Change, Land Use Transformation and Socio-Economic Development”, Texas.

[53] Öztürk D., Batuk F., (2011), "SCS Yüzey Akış Eğri Numarasının Uzaktan Algılama ve Coğrafi Bilgi Sistemleri ile Belirlenmesi", TUFUAB VI. Technical Symposium, Turkey.

[54] Thomas G., McMahon M., (1972) "The relation between soil characteristics,water movement and nitrate concentration of ground water", Kentucky Water Resources Institute Research Report No. 52, Lexington.

[55] Oliver H., Andre L., Gerry L., (2001) "Loss on Ignition as a Method for Estimating Organic and Carbonate Content in Sediments: Reproducibility and Comparability of Results", Journal of Paleolimnology.

[56] Filho C. C., Laurenço A., Fonseca I. B. (2002), "Aggregate stability under different soil management systems in a red latosol in the state of Parana, Brazil", Soil and Tillage Research.

[57] Nielsen D. R., Kutilek M. (1994), "Soil Hydrology", Catena Verlag, Cremlingen, Germany.

[58] Thiessen A. H. (1911), "Precipitation for large areas, monthly", Weather Review.

[59] Sun W., Xu G., Gong P., Liang S, (2006), "Fractal analysis of remotely sensed images: A review of methods and applications", International Journal of Remote Sensing.

[60] Gonzalo S. A., Pablo G., Saman R., Howard S. W., (2018), "On the appropriate definition of soil profile configuration and initial conditions for land surface–hydrology models in cold regions", Hydrology and Earth System Sciences.

[61] Assouline S., (2011), "Bulk Density of Soils and Impact on Their Hydraulic Properties", in Encyclopedia of Agrophysics.

[62] Manns H. R., Aaron A. B., (2014), "Importance of soil organic carbon on surface soil water content variability among agricultural fields", Journal of Hydrology.

[63] Saxton, Rawls (2006), "Soil water characteristic estimates by texture and organic matter for hydrologic solutions", Soil Science Society of America Journal.

[64] Famiglietti Ryu, (2006), "Multi-scale spatial correlation and scaling behavior of surface soil moisture", Geophysical Research Letters.

[65] Stevens A., Wasemael B. V., Bartholomeus D., Rosillon B., Tychon B., Ben-Dor E., (2008), "Laboratory, field and airborne spectroscopy for monitoring organic carbon content in agricultural soils", Geoderma.

[66] Jun L., (2015), "Identification of Forensic Information from Existing Conventional Site-Investigation Data", Introduction to Environmental Forensics (Third Edition).

[67] FAO, (1998), "FAO-Unesco soil map of the world, revised legend, with corrections and updates", World Soil Resources Report 60, FAO, Rome; reprinted with updates as Technical Paper 20 by ISRIC, Wageningen.

[68] Larose M., Heathman G. C., Norton L. D., Engel B., (2007), "Hydrologic and Atrazine Simulation of the Cedar Creek Watershed Using the SWAT Model", Journal of Environmental Quality.

[69] Motovilov Y. G., Gottschalk L., Engeland K., A. Rodhe (1999), "Validation of a distributed hydrological model against spatial observations", Agricultural and Forest Meteorology.

[70] Meaurio M., Zabaleta A., Uriarte J. A., Srinavasan R., I. Antigüedad (2015), "Evaluation of SWAT models performance to simulate streamflow spatial origin. The case of a small forested watershed", Journal of Hydrology.

[71] Santhi C., Allen P. M., Arnold J. G., Tuppad P., (2008), "Regional estimation of base flow for the conterminous United States by hydrologic landscape regions", Journal of Hydrology.

[72] Arnold J., Daniel M., Gassman P., (2012), "SWAT: Model use, calibration, and validation".

[73] Jan V., Janecek M., Tıpp M. l (2007), "Revised Soil Erodibility K-factor for Soils in the Czech Republic", Soil and Water Research..

[74] Bhandari K., Arnaud K., Thakali R., Kaira A., (2018), “Effects of Soil Data Resolution on the Simulated Stream Flow and Water Quality: Application of Watershed-Based SWAT Model”, World Environmental and Water Resources Congress, Minneapolis, Minnesota.

[75] Mengistu G., John E. M. (2007), "Effects of soil data resolution on SWAT model stream flow and water quality predictions", Journal of Environmental Management.

[76] 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.

[77] Yassine B., Aicha R., Namira P. E. A., Lorena L., (2019), "Understanding the effects of soil data quality on SWAT model performance and hydrological processes in Tamedroust watershed (Morocco)", Journal of African Earth Sciences.

BIOGRAPHY

Halil Nurullah ORUÇ works as a research assistant at Gebze Technical University Earth and Marine Science Institute. ORUÇ received his BSc in Environmental Engineering in 2016 from Kocaeli University, Kocaeli, Turkey, and he is currently a MSc student on Earth Science at Gebze Technical University Earth and Marine Science Institute, Turkey.

APPENDICES

Appendices A: Studies

Oruc, H., Çelen, M., Gülgen, F., Öncel, M. S., Vural, S., & Kılıç, B. (2020). Sensitivity of the SWAT Model to Soil Data Parameterization; Case Study in Saz - Çayırova Stream, Turkey. 5th EurAsia Waste Management Symposium, 26/10/2020, İstanbul.