University of Tartu
Faculty of Science and Technology
Institute of Ecology and Earth Sciences
Department of Geography

Master Thesis in Geoinformatics for Urbanized Society (30 ECTS)

Assessing the effect of spatial input data quality on the SWAT model's sensitivity in the Porijõgi catchment

Mahdiyeh Sepehrar

Supervisors:
Alexander Kmoch, Evelyn Uuemaa

Tartu 2021

SWAT mudeli tundlikkus ruumilistele sisendandmetele Porijõe valgla näitel

Lühikokkuvõte

Hüdroloogilised mudelid on oluliseks abivahendiks hüdroloogiliste süsteemide keerukuse mõistmisel ning veeressursside arendamisel, haldamisel ja planeerimisel. Üheks levinumaks neist on SWAT, mida on paljudes riikides kasutatud heitveekoguste ja veekvaliteedi modelleerimisel. Hüdroloogiliste mudelite sisendiks on uurimisala ruumiandmed, näiteks kõrgusinfo, maakasutus, mullastik ja geoloogia. Sisendandmete kvaliteet, resolutsioon ja selle teisendamine võivad oluliselt mõjutada mudeli määramatust. Kuigi üldjuhul eelistatakse kõrge resolutsiooniga piirkondlikke sisendandmeid, ei pruugi mudel sellegipoolest olla täpsem kui madalama resolutsiooniga globaalsete andmete puhul. Käesoleva uuringu eesmärgiks oli selgitada välja sisendandmete eraldusvõime mõju SWAT mudeli parameetrite tundlikkusele ja määramatusele. Uurimisalaks oli Porijõe valgla Tartu linna lähedal. Töös võrreldi globaalse ja piirkondliku täpsusega sisendandmetega SWAT mudeleid. Võrdluse aluseks olid Porijõe valglas perioodil 2000–2013 simuleeritud igakuiste vooluhulkade statistilised näitajad. Tundlikkuse analüüsiks ja mudeli kalibreerimiseks kasutati SUFI-2 algoritmi. Uuringu tulemusena selgus, et modelleerimise täpsust mõjutas oluliselt mullastikuandmete valik, kusjuures kõrgema resolutsiooniga piirkondlikud andmed suurendasid mudeli määramatust.

Märksõnad: hüdroloogiline modelleerimine, SWAT, määramatus, tundlikkus, kalibreerimine

CERCS kood: P510 Füüsiline geograafia, geomorfoloogia, mullateadus, kartograafia, klimatoloogia

Abstract

Hydrological models are an essential tool to understand the complexity of the hydrological systems and are crucial for water resource development, management, and planning. One of the hydrological models which has been extensively used in many countries for modeling discharge as well as water quality is the SWAT model. Spatial information input data such as elevation, land use, and soil or geology is essential for every semi-distributed or distributed hydrological model. The quality, resampling methods, and spatial resolution of these spatial input datasets introduce a great deal of uncertainty into the model (Sharma&Tiari, 2014). Fine resolution regional data is preferred for modeling purposes, but it is not always reliable that these data can lead to better model performances. The overall objective of this study is to examine the impact of input data on SWAT model parameter sensitivity and uncertainty. The study area for this research is the Porijõgi Catchment near Tartu city. Several models were built where and global and reginal data with the different spatial resolution was used, including the new Estonian high-resolution EstSoil-EH soil dataset. The research uses statistical criteria to evaluate SWAT model performance for monthly simulated stream flows between 2000-2013. The Sufi-2 algorithm was used for sensitivity analysis and model calibration. The result unexpectedly indicated lower performance for models with high-resolution regional soil compared to models with global soil data.

Keywords: hydrology, model, SWAT, uncertainty, sensitivity, calibration.

CERCS Code: P510 Physical geography, geomorphology, pedology, cartography, climatology

Table of Contents

Abbreviation 1
Introduction 2
1.Theoretical overview 4
1.1. Hydrological system 4
1.2. Hydrological models 4
1.3. Soil Water Assessment Tool (SWAT) model 5
1.4. SWAT model sensitivity analysis, calibration, and validation 7
1.5. The effect of input data on SWAT model's uncertainty 8
1.5.1. Land cover 9
1.5.2. Soil 10
2. Data and methodology 11
2.1. Study area 11
2.2. Input Data 12
2.2.1. Elevation data 12
2.2.2. Soil data 15
2.2.3. Observed climate and flow data 19
2.3. Model setup 22
2.4. Calibration Method 22
2.4.1 Nash–Sutcliffe Efficiency 23
2.4.2 Mann Whitney U test 24
3.Results 25
3.1. Model global sensitivity results 25
3.2. Model Evaluation Performance 27
3.3. Parameter uncertainty of the top 5% simulations 30
3.3.1. Parameter uncertainty of ALPHA_BF 30
3.3.2. Parameter uncertainty of CN2 32
3.3.3. Parameter uncertainty of SOL_K 33
3.3.4 Parameter uncertainty of SOL_BD 35
3.3.5. Parameter uncertainty of EPCO 36
3.3.6. Parameter uncertainty of GW_DELAY 38
3.4 Validation 39
3.5. Parameter uncertainty in Validation period of the top 5% simulations 42

3.5.1. Parameter uncertainty of SOL_BD in validation period 42
4.Discussion... 44
5. Conclusion 47
Kokkuvõte.. 48
References.. 50

Abbreviation

Abbreviation	Details
ARM	Agricultural Runoff Management
CREAMS	Chemicals, Runoff, and Erosion from Agricultural Management Systems
CN2	Initial Soil Conservation Service (SCS) runoff curve number for moisture condition II)
DEM	Digital Elevation Model
ETAK	Estonian Topographic Database
GIS	Geographic Information Systems
HEC-HMS	Hydrologic Engineering Center-Hydrologic Modeling System
HRU	Hydrologic response unit
HSPF	Hydrological Simulation Program - FORTRAN
IIASA	International Institute for Applied Systems Analysis
LULC	land use/landcover
NRCS	Natural Resources Conservation Service
SMTMP	Base temperature of snowmelt
SOL_AWC	Available water capacity of the soil layer
SOL_K	Saturated hydraulic conductivity
SWAT_CUP	SWAT calibration uncertainties program
SWM	Watershed Model
USDA	United States Department of Agriculture
95PPU	95 Percent Prediction Uncertainty

Introduction

Hydrological models were built to overcome the incomprehensibility and complexity of hydrological systems (Xu, 2002; Woessner, 2012). Hydrological models are the simplified simulation of the real world. Researchers and decision-makers currently use a wide range of hydrological models, yet their application depends on the purpose of the use. Some models are used for research to increase the knowledge about hydrological processes. In contrast, others are used for prediction and simulation purposes by decision-makers to inform the most effective decision for operation and planning (Sorooshian et al., 2008).

Computer-based hydrologic models are essential tools for water resource development, planning, and management since they enable long-term simulations of the effects of watershed processes and management activities (Singh & Woolhiser, 2002). Hydrological models have also made it possible to evaluate the best management practices in watersheds (Arabi et al., 2006; Douglas-Mankin et al., 2010). Additionally, these models facilitate the simulations of different conservation programs, and they can be used to design policies to mitigate water and soil quality degradation by determining suitable conservation programs for watershed settings (Moriasi et al., 2007).

There are several hydrological models suitable for estimating current water availability, such as the Chemicals, Runoff, and Erosion from Agricultural Management Systems (CREAMS), the Hydrologic Simulation Package-FORTAIN (HSPF), the Hydrologic Engineering Center-Hydrologic Modeling System (HEC-HMS), and (SWAT) (Arnold et al., 1998; Montecelos-Zamora et al., 2018).

For choosing the appropriate model, the investigation of the model's data requirements and functionality is essential. (Viessman et al., 1989).

Thus, a good hydrologic model must correctly reflect agricultural management and land use changes and their effects on streamflow (Arnold et al., 1998). Six essential characteristics of a good hydrologic model are high spatial detail, computational efficiency, readily available Inputs, continuous-time representation, the ability to simulate land management scenarios, and the ability to provide reasonable results. However, the implementation of these models often requires integrating geographic information systems (GIS), remote sensing, and multiple databases for the development of the model input parameters and the analysis and visualization of the simulation results (He, 2003).

Even though there are several hydrological models available to explore data change effects on water flow amount (Krysanova et al., 2017; Bormann et al., 2009; Exbrayat et al., 2014), SWAT is the most frequently used water quantity-quality model (Fu et al., 2019; Ray, 2018). The model was developed to face present and future challenges in hydrological modeling. SWAT model has become one of the most used models in hydrological modeling over the past 20 years (Arnold et al., 1998) and is widely used in northern European watersheds to better assess the hydrological changes (Piniewski et al., 2018). The model uses a large range of temporal and spatial scales, environmental conditions, and management practices. It can consider climate change and land use scenarios to simulate the stream flows and pollutant transfer (Bieger et al., 2017).

The latest focus of hydrological modeling applications is on the effect of spatial input data on model output (Chaplot, 2014). These uncertainties are usually associated with the data source, resample techniques, and resolution. Spatial input data such as elevation, land use, soil, or geology are an essential requirement of every semi-distributed or distributed model. These data are available in different resolutions and can be collected from various sources, but the latter can provide different information regardless of having the exact resolution (Sharma&Tiari, 2014).

A trustworthy input dataset is one of the essential requirements for producing a reliable model response by reducing the uncertainty (Wang&Wu, 2015; Beven, 2016).

The overall objective of this study is to examine the impact of input data quality on SWAT model's sensitivity in the Porijõgi catchment, Estonia. The research tries to answer the following questions:

Does high-resolution reginal soil, and land use/cover input improve SWAT model's predictive reliability of flow?

What is the uncertainty of the model with reginal data input compared to global data?

The hypothesis is that higher resolution reginal soil and land use data will increase model performance in the case study area compared to using coarser resolution global soil and land use datasets.

The research is using statistical criteria to evaluate SWAT model performance for monthly simulated stream flows. In addition to graphical representation, statistical metrics such as the Nash-Sutcliffe Efficiency (NSE) will be used to evaluate model performance. In addition, Mann–Whitney U test will be done for comparison of parameter uncertainty in 95 percentiles between models.

1.Theoretical overview

1.1. Hydrological system

Hydrology deals with the appearance, movement, and storage of water on earth. Water appears in liquid, solid, and vapor phases, and in distinct ways, is transported through surface, subsurface, and atmosphere; it is stored in vegetation, soil, flood. Therefore, understanding the underlying physics and processes involved in it and estimating the quantity and quality of water in the various phases and stores is crucial (Salas et al., 2014).

Hydrological system is defined as a set of chemicals, physical and or biological processes that act upon input variables and result in output variables. The variable here is the system’s characteristic, which can be measured and can assume different numbers at different times. Parameters define these characteristics of hydrological systems that may remain constant in time or change (Dooge, 1973).

As most hydrological systems are complex and cannot be easily understood and monitored in detail, managing these water resources is difficult. The problems associated with water resources management involve complex processes from surface and subsurface level to their interface level. (Sophocleous, 2002; Srivastava et al., 2013). Abstraction for understanding and simplifying these systems is necessary. This can be done through hydrological system analysis or modeling. The main goal of hydrological system modeling is to study system operation and predict the output by using the hydrological model(Nyeko, 2010).

1.2. Hydrological models

The power of computers has been remarkably recognized in hydrology after the digital revolution since the 1960s. Numerical and statistical simulation is often used in many studies as a way of computing. Watershed Model (SWM), developed by Crawford and Linsley in 1966, was the first try to model hydrological cycle in watersheds; many advancements were made on this model (Jajarmizadeh et al., 2012).

Later, many models were made for hydrological modeling; some examples of these models include: The Hydrologic Modeling System (HEC-HMS) was designed to simulate the complete hydrologic processes of dendritic watershed systems. Many traditional hydrologic analysis procedures such as event infiltration, unit hydrographs, and hydrologic routing were included in this model (Neitsch et al., 2011).

Research scientists developed (CREAMS) model from the United States Department of Agriculture (USDA_ Agricultural Research Service). The main objective of the model was to aid the Natural Resources Conservation Service (NRCS) specialists to estimate nonpoint-source pollution from agricultural fields and study the impacts of different management practices (Neitsch et al., 2011).

Hydrological Simulation Program - FORTRAN (HSPF) is an extensive program for simulating the hydrology and water quality for regular and toxic organic pollutants; the program was released in March 1997. It is the combination of watershed-scale Agricultural Runoff Management (ARM) and non-point source pollution models (NPS) models in basin-scale also, it is the only model that allows the integrated simulation of soil and land contaminant runoff process (Neitsch et al., 2011).

Researchers developed another semi-distributed model as a compromise between lumped and fully distributed models to defeat the difficulties of distributed models. The algorithms in these models are simple but physically based (Arnold et al., 1993). As one of the semi-distributed hydrological models, the SWAT model was originally developed to predict discharge from ungauged basins (Arnold et al., 1998).

1.3. Soil Water Assessment Tool (SWAT) model

SWAT is a river basin watershed scaled model which was developed for USDA Agricultural Research Service. SWAT consists of features of several sub-models and is a direct development of Simulator for Water Resources in Rural Basins (SWRRB) (Williams et al., 1985; Arnold et al.,1990). The primary goal of the model is to predict the impact of land management practices on water, sediment, and agricultural chemical yields in large complex watersheds with different soils, land uses, and management practices over a long period of time (Neitsch et al., 2011). Therefore, the SWAT model has been successfully used to simulate water flow, sediment, and nutrient loadings (Rosenthal & Hoffman, 1999). It has been extensively used in different countries for discharge prediction also soil and water conservation (Patel & Srivastava, 2013; Spruill et al., 2000; Zhang et al., 2010).

The SWAT model is a continuous, long-term, distributed-parameter model that has the ability to simulate surface flow, soil erosion, subsurface flow, sediment transfer, and movement of nutrients through watersheds (Arnold et al., 1998). The model allows simulation of various physical processes such as water runoff, deposition, sediment generation, and nutrition transport. The hydrological cycle simulated in the model is based on the water balance equation:

$$SW_t = SW_0 + i=1t\ (R_{day} - Q_{surf} - E_a - W_{seep} - Q_{lat} - Q_{gw})$$

where SW_t is the final soil water content (mm water), SW_0 is the initial soil water content on day i (mm water), t is the time (days), R_{day} is the amount of precipitation on day i (mm water), Q_{surf} is the amount of surface runoff on day i, (mm H_2O), E_a is the amount evapotranspiration on day i (mm H2O),W_{seep} is the amount of water entering the vadose zone from the soil profile on day i (mmwater),Q_{lat} is the lateral flow from soil to channel and Q_{gw} is the amount of return flow on day i (mm water) (Neitsch et al., 2005) .

The model interface takes the data inputs and subdivides the basin based on the direction of drainage overland flow into sub-basins. After that, it delineates the HRU, a unique combination of soil type and land cover for each sub-basin (Bieger et al., 2017).

Soil water content, surface runoff, nutrient cycling, and management practices are simulated for each HRU, and the results are aggregated for the sub-basin by weighted average. The input data are incorporated into the model using the tables containing information about the hydrological characteristics of relevant soil. The model also requires daily climate information about solar radiation, precipitation, wind speed, humidity, and maximum, minimum temperatures. Finally, the location of existing streamflow and rainfall gauging stations should be included inside the model. Based on these data, the model estimate evapotranspiration establishes the water balance of each HRU (Bieger et al., 2017). In this research, Arc SWAT version 2012 was used for modeling the watershed; moreover, SWAT cup version 2007 was used for calibration and validation of models.

1.4. SWAT model sensitivity analysis, calibration, and validation

The process-based nature of the SWAT model parameters makes them dependent on realistic parameter ranges. Sensitivity analysis and determination of the sensitive parameters are the initial steps in the calibration and validation process of the SWAT. The determination of the parameters range is usually based on expert judgment or sensitivity analysis. Sensitivity analysis is determining the rate of change in model output based on changes applied to model input. Therefore, it is necessary to identify key parameters and their precision for model calibration (Ma et al., 2000).

Guo and Su (2019) investigated the streamflow in the Shiyang basin based on SWAT. They found out that the Runoff curve number (CN2) and the Base temperature of the snowmelt (SMTMP) are found to be the most sensitive parameters, which implies that the generations of surface runoff and snowmelt were extremely crucial for streamflow in this basin. Moreover, the uncertainty analysis of streamflow prediction indicated that simulation performance could be further improved by parameter optimization. In addition, sensitivities of 21 input parameters have been analyzed using the SUFI-2 algorithm in SWAT_CUP for the Langat River Basin by Khalid et al. (2016). In this basin, five SWAT input parameters showed the most sensitivity for reginal and global sensitivity procedures, including CN2.mgt, Groundwater delay (GW_DELAY.gw), SLOPE.hru, available water capacity of the soil layer (SOL_AWC.sol), and Saturated hydraulic conductivity (SOL_K.sol).

According to the new research done on sources of uncertainty from various types of input datasets and model parameters by SWAT application for the Minjiang River watershed, 20 parameters were used, and 15 sensitive parameters for this basin were chosen. The analysis revealed that uncertainties related to the stream network precision and specific SWAT parameters are the most critical factors (Koo et al., 2020).

SWAT sensitivity analysis can be performed locally or globally. Local sensitivity analysis is changing parameter values one at a time while global is changing all parameter values. As the sensitivity of the parameters often depends on the values of other parameters, therefore the problem with one at time analysis is the ignorance of the correct values of the other fixed parameters. Also, the disadvantage of the global sensitivity analysis is the high number of simulations needed and being expensive. At the same time, the strength of this method is the more robust depiction of model uncertainty by comprehensively accounting for parameter interactions. Both procedures are a necessary step in the calibration process (Arnold et al., 2012).

The next step is the calibration of the model, which is better parameterizing the model to a given set of local conditions by reducing the prediction uncertainty. For calibration process input parameters, values should be carefully selected based on a comparison of model prediction(output)for given assumed conditions for observed data for the same condition. The final step is model validation, demonstrating that the model can make accurate simulations based on the project goal. This process can be done by running the model by parameters determined during calibration and comparing the results to the observed data used during the calibration process. Calibration can be done manually or using autocalibration tools in SWAT-CUP (Arnold et al., 2012).

1.5. The effect of input data on SWAT model's uncertainty

Using a model to simulate the reality has proven to be challenging due to the many possible errors such as forcing data, model structure, input data parameter estimation, and use of goodness-of-fit criteria. Also, watershed response prediction can involve two types of errors: 1) the systematic model error can be accrued regardless of input 2) errors due to inaccuracies of the input data (Troutman, 1983). SWAT as a semi-distributed hydrological model requires spatial information as input data, such as topography (elevation), LULC (land use/landcover), soils, or hydro- geology. Wagenet and Hutson (1996) reported that even though the use of GIS significantly enhanced the capability of models to simulate watersheds, but the scale that GIS or spatial data such as land use/cover, soil, and elevation should be collected and used is a major concerned which needs to be studied. In addition, the authors mentioned that modeling results could be sensitive to the quality and nature of input variables, and interpretation of model output is limited to the resolution and quality of input environment data.

Shirmommadi et al. (2006) mentioned that uncertainty in estimates of input parameters derived from different land use and soil datasets could result in a significant portion of uncertainty in streamflow modeling. Di Lazio et al. (2005) indicated the interaction and aggregation of various soil and land use layers at the hydrologic response unit (HRU) level play a crucial role in describing the hydrological response in a realistic manner. Another source of uncertainty is caused by different sources of available data on model predictive uncertainty. The potential impact caused by various sources of available data toward model prediction (flow) is still rarely being investigated (Heathman et al., 2009).

The number of studies done on the effect of spatial input data resolution on uncertainty in surface flow and water quality response predictions is limited. Most of these studies either have focused on the rainfall-runoff process only or investigated the effects of Digital Elevation Model (DEM) resolution and cell aggregation on topography and uncertainty in modeled flow (Ma, 1993; Wang et al., 2000).

Choubey et al. (2005) assessed the effect of DEM, LULC, and soil map resolution on SWAT output error. In this study, the maps have been resampled to different resolutions, and the best resolution map was used for calibration of the model, and the optimum parameter was transferred to other model setups. The authors imply that not using observed data to calculate relative error and transferring parameter values would evade uncertainty of observed data and parameter computation. Moreover, studies showed that measured soil data and meteorological data had been shown to affect model output accuracy (e.g., Wagenet & Hutson, 1996; Wilson et al., 1996). Heathman et al. (2009) evaluated the use of different combinations of soil and land use maps from various sources on modeled streamflow in SWAT. The authors found identified the influence of interaction, pre-processing, and aggregation of unique combinations of GIS input layers on simulated streamflow. The results indicated that land use has more effect on streamflow estimates than soil map. Hoang et al. (2018) found out that the resolution and complexity of the spatial input data do not improve the model’s performance nor reduce parameter and output uncertainty. However, spatial and temporal observations can be used in finding suitable parameter sets and reducing uncertainty and prediction. Although its wide usage, the SWAT model has rarely been applied in the Baltics, especially in Estonia. Only a few studies have been using SWAT to assess streamflow or water quality. None of the studies have analyzed the impact of spatial data on sensitivity and uncertainty in-depth on model performance with reginal data sources (Wielgat et al., 2021; Čerkasova et al., 2019).

1.5.1. Land cover

Predictive reliability of the SWAT model based on land use data spatial resolution and image classification was investigated by Luzio et al. (2017). The author showed that the low-resolution model had just slightly better predictive reliability. However, the impact of classification was unclear in this research.

Pai et al. (2003) showed and highlighted the effect of published land use data categorical errors on SWAT model uncertainty. Yen et al. (2015) used different land use maps from multiple sources to investigate the model's uncertainty and the potential impact of cross transferring optimal calibration parameters between models. Their result showed that use of varying data source may not only alter prediction and associated uncertainty also has a direct impact on transferability of model parameters.

1.5.2. Soil

A study done on error transmission of use of different soil data to model predictions in mountainous watershed indicated that model outputs are not sensitive to soil resolution. Still, soil data choice can significantly affect the application of watershed models in terms of the goodness-of-fit indicator, predicted data, and related uncertainty (Kmoch et al., 2019). Moriasi and Starks (2010a) investigate the effect of soils and precipitation dataset on SWAT streamflow simulation performance, and calibration parameters they found out that there were no significant differences in the model monthly performance statistics between the higher resolution soil and the lower resolution (Kumar & Merwade, 2009).

2. Data and methodology

2.1. Study area

The Porijõgi drainage basin with total area of 258 km^2 is one of the Emajõgi river sub-catchments. The total area for the catchment delineated against measurement Reola station is 240 km^2.The catchment is located on the borders of two southern Estonian till plain and Otepää heights (Varep 1964) with a landscape representative of the whole of south Estonia. The central and northern part of the catchment is in the southern Estonian moraine plain 5-10 km south of Tartu city (58°23'N; 26°44'E). The elevation of the plateau in this area varies between 30-60 above mean sea level, and slopes are 5-6%; also, reliefs are undulated. Primeval valleys split the landscape, 3-5 km wide and up to 40 m deep, formed by streams (Varep, 1964). The southern part of drainage lies on the northern slope of Otepää heights. The elevation of this region is up to 120 m; the relative heights reach 30-35 m. The depth to the water table varies depending on relief and geomorphologic conditions (0.5-20 m). The whole catchment area's bedrock is formed by Devonian sandstone, which is covered by loamy sandy till. The northern part of the catchment is covered by patches of fields, grasslands, and forests, while the southern part a very mosaic landscape (Mander et al.,1994). Coniferous and mixed types of forests cover about 32% of the investigated catchments, the rest being evergreen and deciduous forest (5%) and agricultural land that formed roughly 50% of the drainage area (Iital et al., 2010).

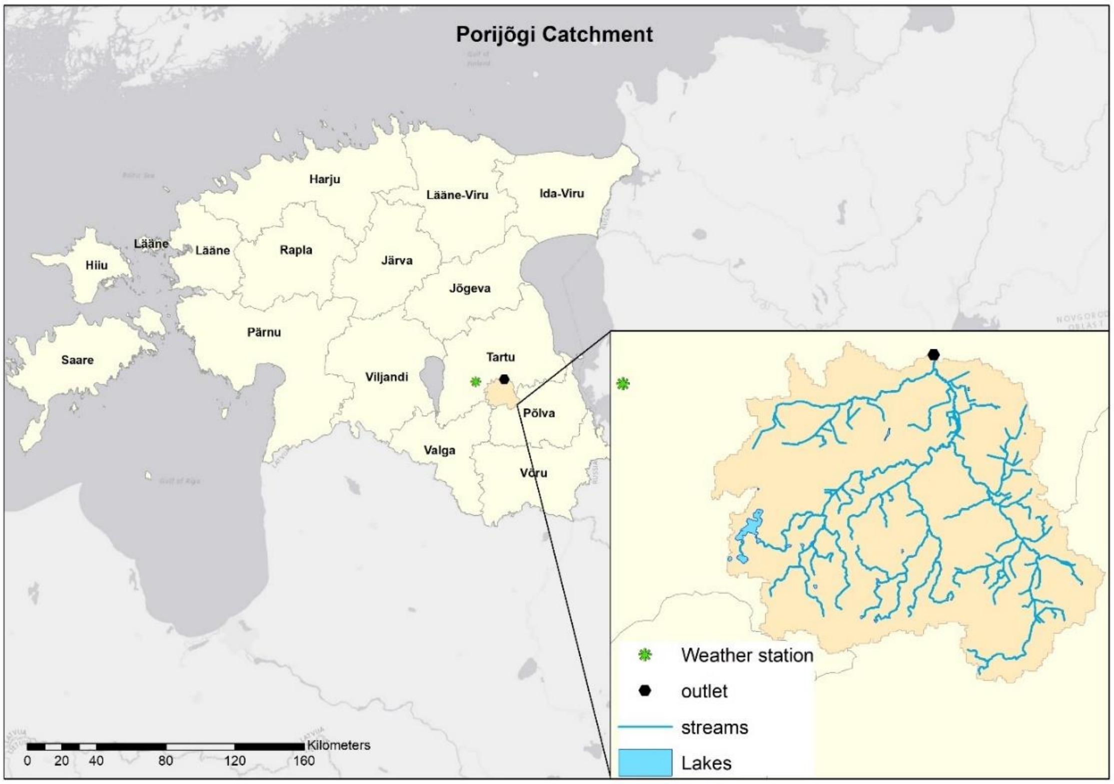

Figure 1:Study area map.

2.2. Input Data

2.2.1. Elevation data

Merit HydroDEM with spatial resolution of 75 m was used as digital elevation model (Figure 2) (MERIT Hydro, 2019). The data is prepared as 5-degree × 5-degree tiles in WGS84 spatial reference system. For all simulations, the same DEM data was used. The drainage threshold for all setups was 250 meters. The number of subbasins for all models was 53, and the slopes were classified into two classes: 1-5% and more than 5%. All data was projected into the Estonian national coordinate reference system (EPSG:3301).

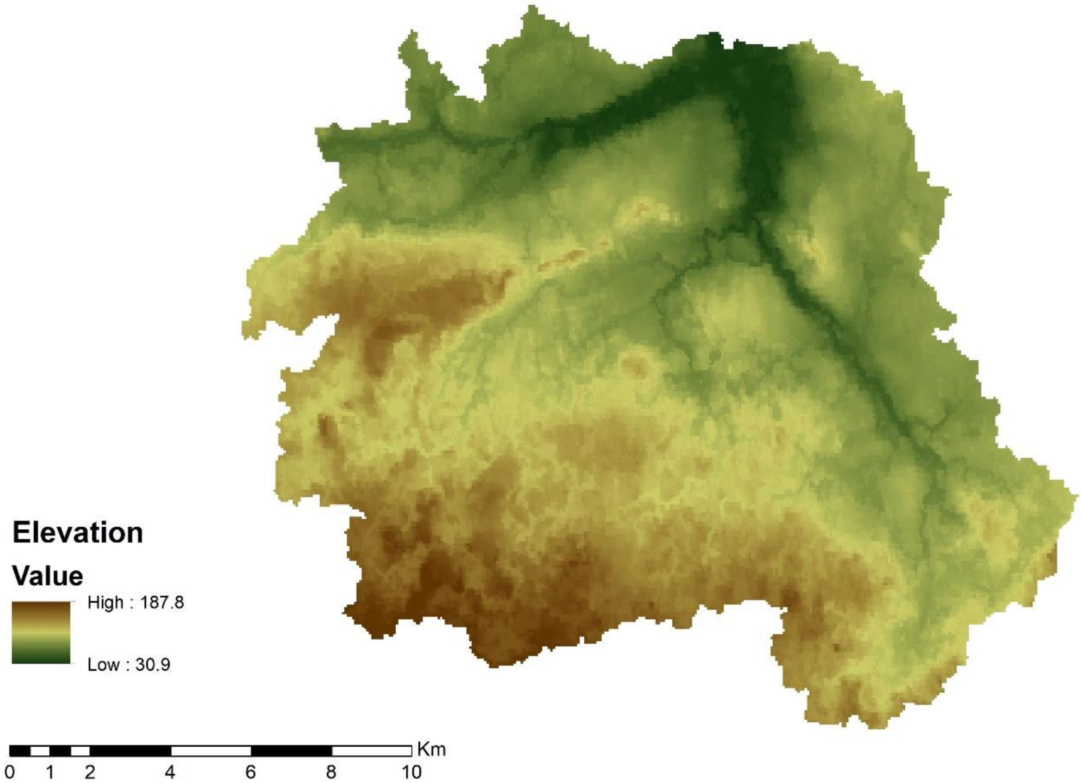

Figure 2: Elevation map of the Porijõgi catchment.

For the global land cover data, CORINE land cover data for 2012 were used. The Porijõgi catchment has only 9 CORINE land cover classes out of 44. CORINE data has a spatial resolution of 100 meters and an accuracy of 85 %. This land cover is made by the cooperation of different countries as satellite images. (Copernicus Land Monitoring Service,2021). Based on CORINE land cover majority of the land use in this area belongs to the generic-agricultural lands and mixed forest with 49.9%. and 32.3% (Table1). Urban areas and water bodies with less than 1% have a minor share of the catchment land use (Figure 3).

Table 1. CORINE land cover types and percentage

Land use code	Definition	Percentage
URML	Urban Medium Density	0.4
AGRL	Generic-Agriculture	49.8
PAST	Pasture/Hay	4.9
RNGE	Grasslands/Herbaceous	0.7
RNGB	Range Shrubland	6.1
FRSD	Deciduous Forest	0.4
FRSE	Evergreen Forest	4.9
FRST	Mixed Forest	32.3
WATR	Water	0.4

For the reginal land use/cover data, the Estonian Topographic Database (ETAK) was obtained from the Estonian Land Board (ETAK,2020). It consists of 13 classes (Table 2). In ETAK, the mixed forest percentage is 45.8% which is significantly more than in the CORINE land cover dataset with 32%. Also, the land class generic or fields in ETAK is 38%, around 11% less than CORINE land cover classes. ETAK is more detailed than CORINE; especially this can be seen in urban land use and wetlands distribution (Figure 3).

Table 2. ETAK land use types and percentage

Land use code	Definition	Percentage
UTRN	Urban Transportation	0.5
UIDU	Urban Industrial	0.3
UTBN	Residential / public building	0.3
URLD	Private yard	2.2
AGRL	Generic-Agriculture	38
PAST	Pasture/Hay	8.7
RNGB	Range Shrubland	0.9
RNGE	Grasslands/Herbaceous	0.01
FRST	Mixed Forest	45.8
WETL	Wetland	<0.01
WETN	Emergent/Herbaceous Wetlands	0.1
WETF	Woody Wetlands	2.2
WATR	Water	0.9

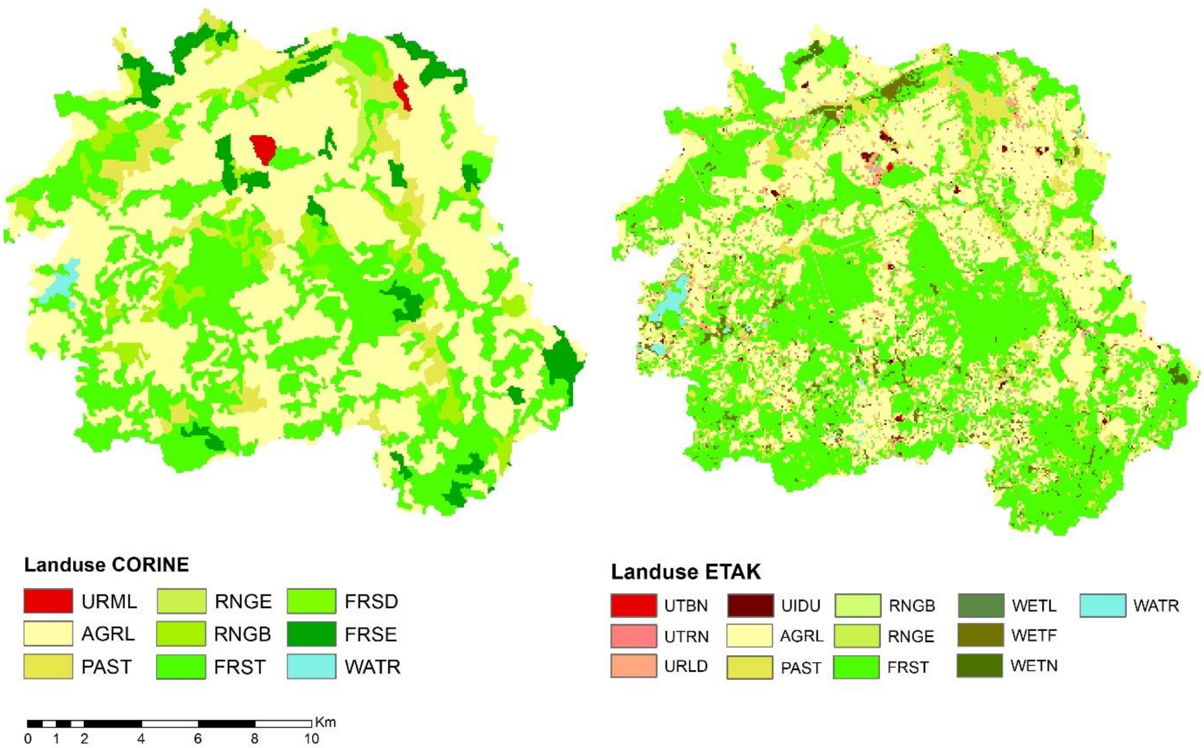

Figure 3:Land use map of the Porijõgi catchment.

2.2.2. Soil data

The Harmonized World Soil Database (HWSD) was used as global level data for soil. The database is a 1 km spatial resolution raster with over 15 000 different soil mapping units. The resulting raster database consists of 21 600 rows and 43 200 columns linked to harmonized soil property data. Different parameters of soil (organic carbon, pH, water storage capacity, soil depth, cation exchange capacity of the soil and the clay fraction, total exchangeable nutrients, lime and gypsum contents, sodium exchange percentage, salinity, textural class, and granulometry) are gathered in this data. This data was produced with the cooperation of the International Institute for Applied Systems Analysis (IIASA) and the Food and Agriculture Organization of the United Nations (FAO). The HWSD soil data is widely used in SWAT modeling. The spatial resolution of soil data in this database is lower; the catchment in this database has three soil classes. The majority of soil texture is clay loam-sand, with 67.3 % in the catchment (Table3). The soil map is shown in Figure4 (Fischer et al., 2008).

Table 3. HWSD soil texture and groups percentage

Soil Group	Soil Texture	Percentage
Mixture (Gleyic Luvisols, Haplic Podzols)	Clay loam-Sand	67.3
Mixture (Gleyic Luvisols, Eutric Plansols, Eutric Gleysols)	Sandy clay-Loam	29.8
Terric Histosols	Clay loam	2.8

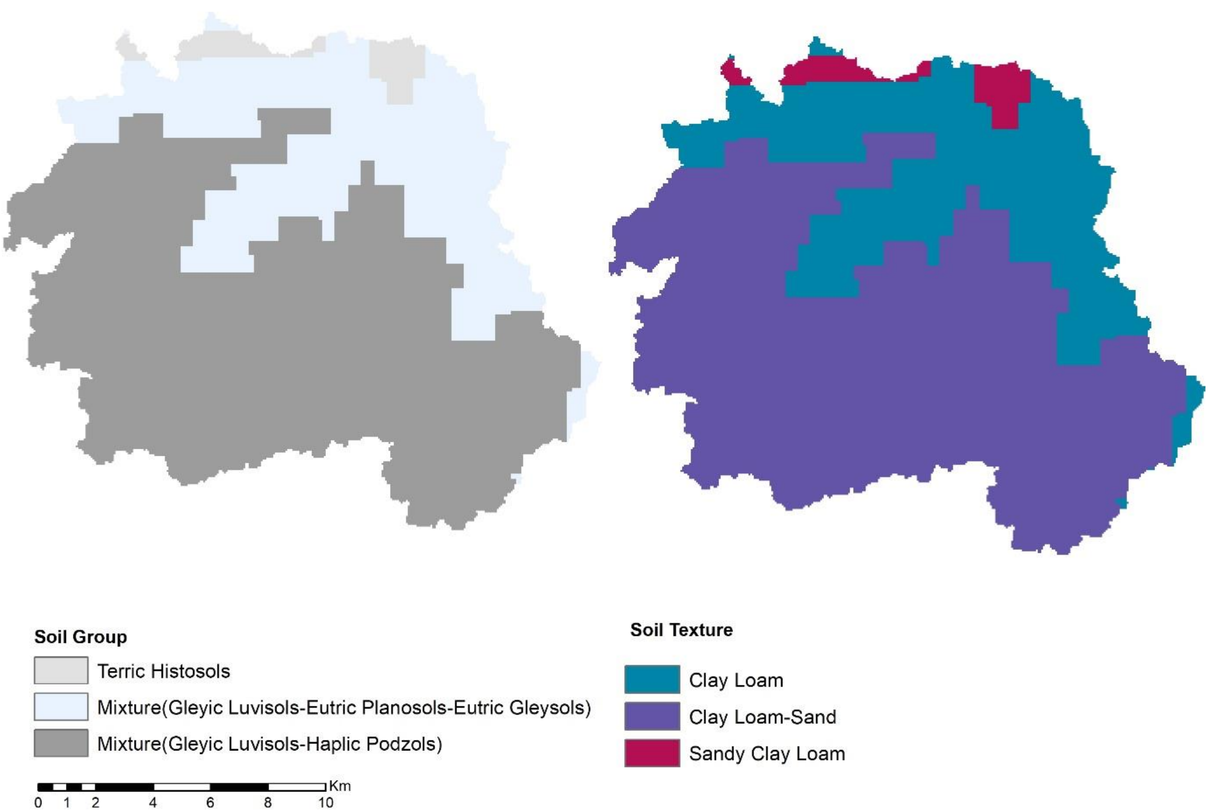

Figure 4:HWSD soil map of the Porijõgi catchment.

For reginal level soil data, EstSoil-EH (Kmoch et al., 2021) was used. The soil map is a synthesized version of 20 extended eco-hydrological variables for Estonia. It is the most detailed and information-rich dataset for soils in Estonia; it mapped more than 750 000 soil units throughout Estonia at the scale of 1:10 000, which include variables such as soil profiles (e.g., layers, depths), texture (clay, silt, and sand components), rockiness, and physical variables related to water and carbon (bulk density, hydraulic conductivity, organic carbon content) (Kmoch et al., 2019). Computation of this map in model input is hard because of the extensive number of soil classes; therefore, the raster soil map was reclassified by sieve tool in GIS which removes raster polygons smaller than a provided threshold size (in pixels) and replaces them with the pixel value of the largest neighbor polygon either four or eight-pixel connection should be chosen for the determination. In this study, the soil map was generalized with sieve tool with the threshold of 50 focal and connection to 8 neighboring pixels. The number of classes decreased to 731 soil classes after generalization. The main soil groups are shown in Table 3. Only soil groups in top layers were chosen for the visualization. The majority of the catchment soil group consists of Umbisols with 52.56 % (Table4). After generalization, this soil group decreased in the catchment and only consist of 40 % of the catchment the smallest percentage is for Retisols with 0.07 in EstSoil-EH map, but after generalization, this number increased to 12% also soil group Cambisol is not included in soil map after generalization (Figure 5). (Figure 6) and (Annex 1) show the percentage of the textures are in EstSoil-EH.

Table 4. EstSoil-EH Main Soil Group

Soil Group	Percentage EstSoil-EH Soil Groups	Percentage Generalized Soil Groups
Umbisols	52.6	39.7
Fluvisols	2.2	28.8
Regosols	0.1	1.6
Gleysols	0.9	5.1
Histosols	5.2	5.3
Podzols	0.2	0.2
Retisols	0.1	12.8
Luvisols	17.6	8.2
Cambisol	2.2	-

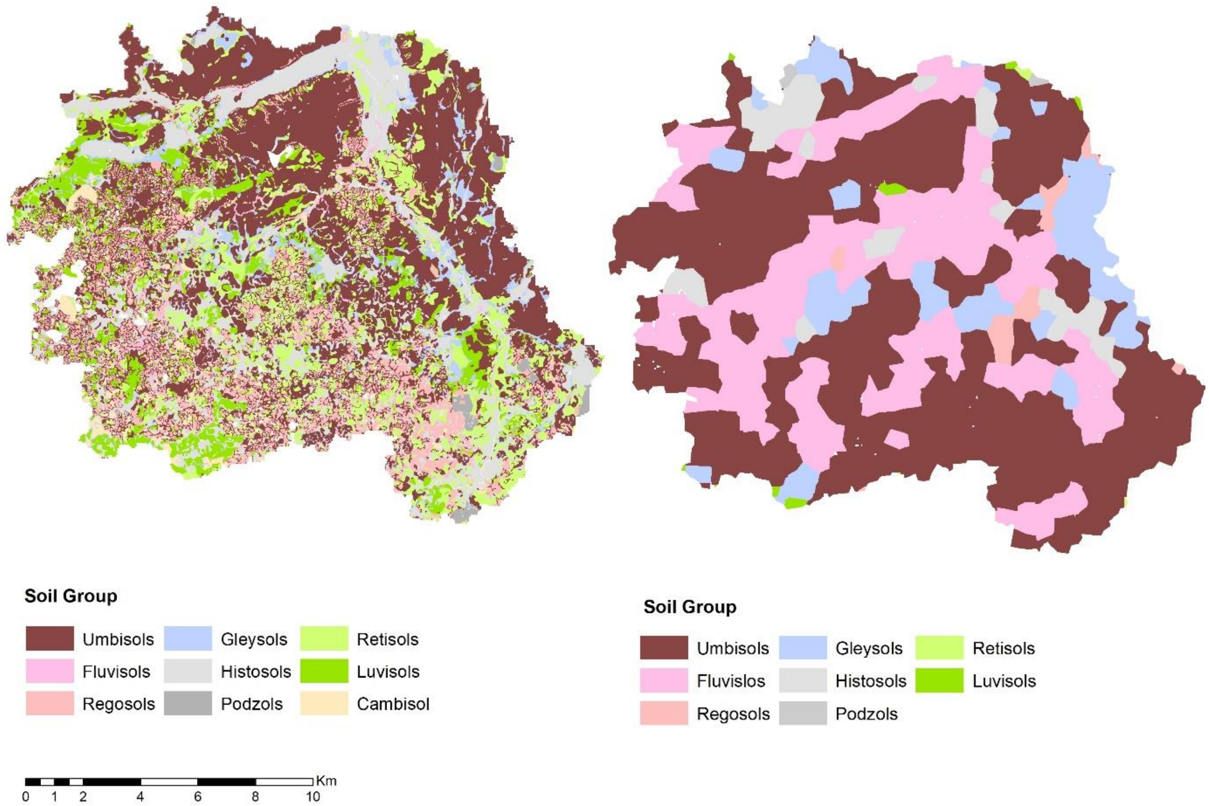

Figure 5: Soil groups in EstSoil-EH map of the Porijõgi catchment.

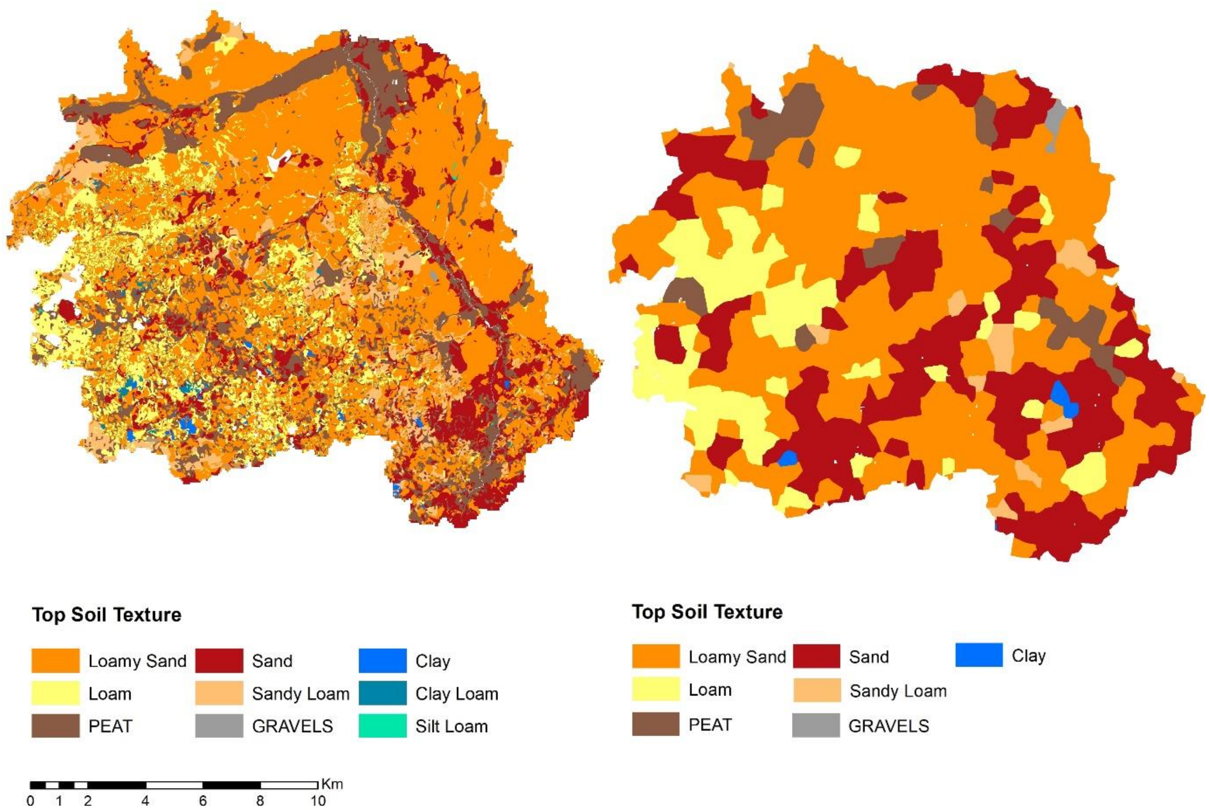

Figure 6: Soil Textures in EstSoil-EH map of the Porijõgi catchment.

2.2.3. Observed climate and flow data

In the SWAT model, climate variables are either generated weather data or input data from the stations. Daily precipitation, solar radiation, wind speed minimum-maximum temperature, and relative humidity were collected from the records of Tõravere station (N 58.264, E 26.461) and was used as the model's climate input. Data consist of daily records from 2000-2013. In average, 2008 with 60.78mm was the wettest year, and 2011 with 35.4 was driest year. The warmest year is 2008 and the coldest year is 2012 also the average water flow was in 2010, and the lowest one is in the year 2006 (Table5).

Table 5. Yearly average temperature, streamflow and precipitation

Year	Average Temperature (°C)	Average Flow ($m^3.s^{-1}$)	Average Precipitation(mm)
2000	8.9	1.4	42.5
2001	7.0	1.4	61.5
2002	6.8	1.4	39.2
2003	7.4	1.8	55.9
2004	7.9	2.0	50.8
2005	6.7	1.5	42.0
2006	8.2	1.1	38.9
2007	8.2	1.4	48.4
2008	9.1	2.0	60.8
2009	6.9	2.3	56.0
2010	6.4	2.4	58.4
2011	8.9	1.9	35.4
2012	6.1	1.8	53.6
2013	8.3	1.5	37.6

The observed flow data for Porijogi are daily observed flow from the Reola hydrometry station, which is located in N 58. 1624, E 26. 4431. Data consist of daily flow records from 2000-2013 for the catchment. The temperature and streamflow follow a similar trend; as temperature increases in the catchment, the streamflow increases (Figure 7).

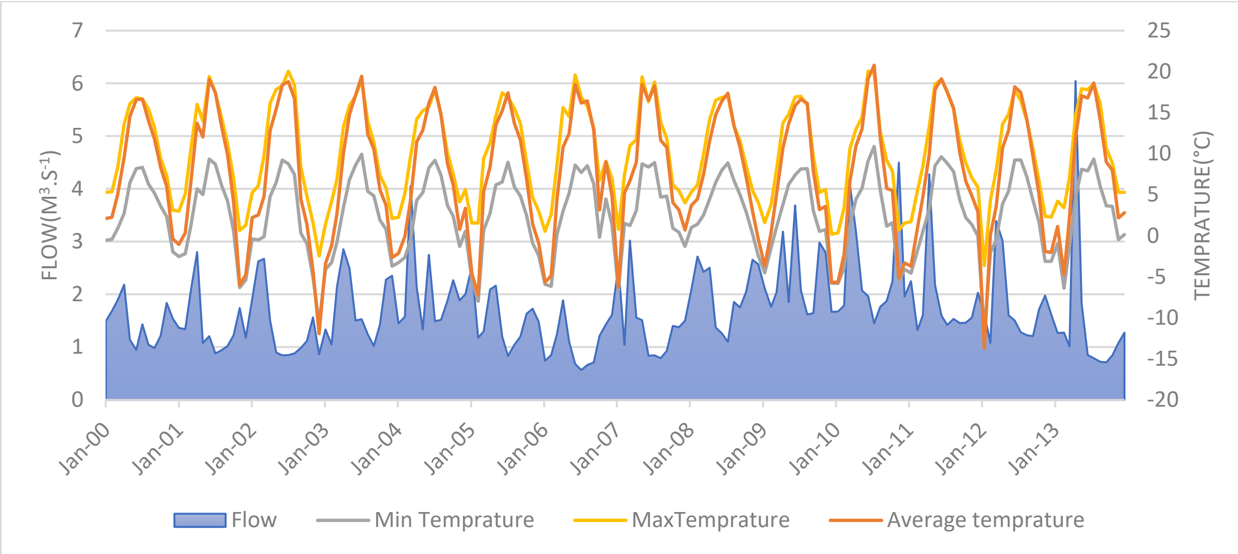

Figure 7: Monthly streamflow and temperature for the year 2000-2013 in the Porijõgi catchment.

The flow and precipitation graph (Figure8) indicates that increase in precipitation played an essential role in the increase in streamflow in catchment except for month April, in this month because of temperature increase and melting of the snow and frozen soil even though the precipitation declines the flow in the catchment significantly increased.

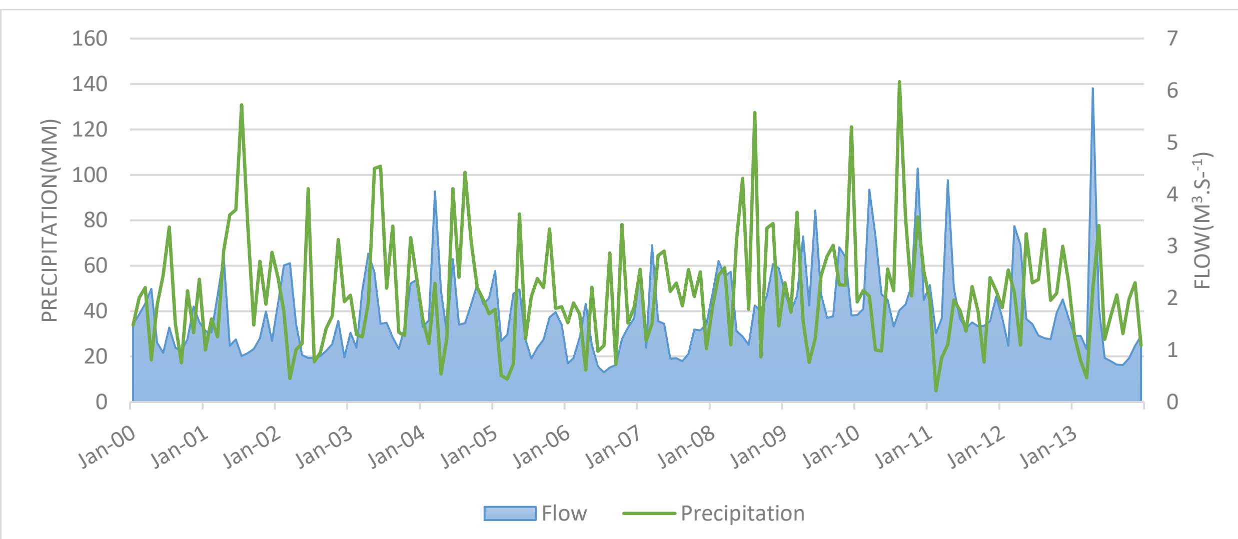

Figure 8: Monthly streamflow and precipitation for the year 2000-2013 in the Porijõgi catchment.

2.3. Model setup

Four model setups for the catchment were made for the time period of 2000-2013 (Table 6). The period of 2000-2003 was chosen as the warm-up period for the models. The following model setups were created:

Table 6. SWAT model setups description

Model	Input data description
m1	EstSoil-EH / ETAK land use
m2	EstSoil-EH /CORINE
m3	HWSD soil/ ETAK land use
m4	HWSD soil/CORINE

The number of subbasins was kept unchanged in all models. The number of HRU-s in the models depends on the distribution of land use, soil, and slopes. The first model, setup m1, has 602 HRU, m2 has 661 HRU, and m3 has 282 HRU final model m4 has 289 HRU.

2.4. Calibration Method

Due to the general uncertainty and spatial variability in the input data, the model is always calibrated and validated based on observed flow data. In this research, for minimizing variation between observed and predicted flow, calibration and uncertainty analyses were done by using the SUFI-2 algorithm available in the soil calibration and uncertainty programs (SWAT-CUP 2007) software (Abbaspour, 2009). SUFI-2 uses parameter values to represent uncertainty in parameters. Then it uses Latin hypercube sampling to draw independent parameter sets for each calibration run. Latin hypercube sampling is a stratified random sampling algorithm that guarantees to cover the parameter space evenly, contrary to, for example, Monte Carlo-based random sampling algorithms (Abbaspour et al., 2007b). In combination with SUFI2, this reduces the number of simulations necessary to achieve good results. SUFI2 propagates parameter uncertainty sets for each calibration runs by calculating model output uncertainty variables (cumulative distribution of output variables) at 2.5% and 97.5%; This is referred to as 95% prediction uncertainty (95PPU). The goal of SUFI-2 is to generate 95PPU results of simulated flow, which envelop most observed flow (Abbaspour et al., 2007b). In this research, twenty-one parameters for calibration were used; the parameters and their range were chosen based on swat model calibration literature (Guo and Su,2019; Koo et al., 2020) (Annex 2).

SUFI-2 generated the best parameters with the smallest uncertainties. These parameters were chosen for the final calibration of the models.

The first iteration was done for baseline check without parameter changes, and then twelve initial iterations were made as standard protocol for all the four models, one-at-a-time or small groups of 2 or 4 for each of the models for initial calibration to achieve a reasonable base performing models (Annex 3).

The selection and order are also informed by how often the base parameters for calibration were chosen in the literature, with the exception of the two parameters related to soil which were chosen based on the focus of the study. Snow parameters should be calibrated in snow-dominated catchments, like Porijõgi, as they have a strong interaction with all other parameters on the performance because they strongly relate to water availability in the system.

In this study, Nash Sutcliffe Efficiency coefficient (NSE) was chosen as a metric for evaluation of the model performance.

2.4.1 Nash–Sutcliffe Efficiency

Nash–Sutcliffe Efficiency (NSE) is used to quantitatively describe the accuracy of model outputs, especially in hydrological modeling.

It helps in the assessment between the predicted and observed flow. NSE is a normalized statistic that determines the relative magnitude of the residual variance(“noise”) compared with the measured data variance (“information”) (Nash & Sutcliffe, 1970). NSE is computed as following equation.

$$NSE = 1 - \frac{(\sum_{i=1}^{n} Q_i - P_i)^2}{(\sum_{i=1}^{n} Q_i - O)^2}$$

NSE is the prediction efficiency, O_i is the observed condition at the time i, O is the mean of the observed values overall times, and P_i is the predicted value at the time I (Nash & Sutcliffe, 1970).

2.4.2 Mann Whitney U test

The Mann-Whitney U test null hypothesis stipulates that the two groups come from the same population. In other terms, it stipulates that the two independent groups are homogeneous and have the same distribution. The two variables corresponding to the two groups, represented by two continuous cumulative distributions, are then called stochastically equal.

The Mann-Whitney U test initially implies the calculation of a U statistic for each group. These statistics have a known distribution under the null hypothesis identified by Mann and Whitney (1947). Mathematically, the Mann-Whitney U statistics are defined by the following equation, for each group:

$U_x = n_x n_y + ((n_x(n_x + 1))/ 2) - R_x$ (1)

$U_y = n_x n_y + ((n_y (n_y + 1))/ 2) - R_y$ (2)

where n_x is the number of observations or participants in the first group, n_y is the number of observations or participants in the second group, R_x is the sum of the ranks assigned to the first group and R_y is the sum of the ranks assigned to the second group (Nachar, 2008).

3.Results

3.1. Model global sensitivity results

Based on the previous publications related to hydrological simulation using SWAT as well as our own experience, we selected 21 key parameters for model calibration (Table 7). The twelve initial iterations were made as standard protocol for all the four models, one-at-a-time or small groups of 2 or 4 for each of the models for initial calibration to achieve a reasonable base performing models. The global sensitivity analysis provides t-stat for impact and p-value for statistical significance for each parameter based on their variations during the many simulations. Parameters with a p-value less than 0.05 were chosen as sensitive parameters for each model. Thirteen parameters ALPHA_BF, SOL_K, GW_DELAY, and SMFMN proved to be sensitive in m1, according to the global sensitivity analysis (Table1). Model m2 had similar results to m1; in this model, thirteen parameters ALPHA_BF, GW_DELAY, SOL_K, and SMFMN proved to be sensitive. Sensitivity analysis showed that eleven parameters were sensitive in m3 including ALPHA_BF, CN2, SMFMX, and GW_DELAY. Model m4 with twelve sensitive parameters in global sensitivity analysis showed similar sensitivity results to m3 in this model ALPHA_BF, CN2, SMFMX, and SMTMP were sensitive.

Table 7. Global uncertainty result for 4 model setups (gray color show the sensitive parameters in each model)

Parameter Name	m1		m2		m3		m4	
	t-Stat	P-Value	t-Stat	P-Value	t-Stat	P-Value	t-Stat	P-Value
R_CH_N2.rte	0.04	0.96	0.04	0.97	0.04	0.97	0.05	0.96
V_ALPHA_BNK.rte	-0.37	0.71	-0.41	0.69	-0.40	0.69	-0.39	0.70
V_LAT_TTIME.hru	-0.63	0.53	-0.57	0.57	0.49	0.63	0.56	0.58
V_TIMP.bsn	0.67	0.50	0.64	0.52	-0.76	0.44	-0.87	0.38
V_GW_REVAP.gw	0.76	0.44	0.76	0.45	0.67	0.51	0.58	0.56
V_GWQMN.gw	0.84	0.40	0.95	0.34	-1.46	0.14	2.25	0.02
V_SNO50COV.bsn	-1.90	0.06	-1.94	0.05	-1.78	0.08	-2.69	0.01
V_ESCO.hru	1.94	0.05	1.87	0.06	-1.75	0.08	-2.69	0.01
V_EPCO.hru	-2.06	0.04	-2.04	0.04	-1.94	0.05	-1.83	0.07
R_SOL_BD(..).sol	2.28	0.02	2.34	0.02	4.77	0.00	4.82	0.00
A_RCHRG_DP.gw	2.71	0.01	2.77	0.01	2.60	0.01	2.46	0.01
V_SMTMP.bsn	2.92	0.00	2.95	0.00	-9.03	0.00	9.93	0.00
R_CN2.mgt	2.93	0.00	2.73	0.01	-11.59	0.00	-13.99	0.00
V_REVAPMN.gw	3.38	0.00	3.06	0.00	-1.46	0.14	-1.76	0.08
V_SNOCOVMX.bsn	3.42	0.00	3.39	0.00	3.30	0.00	3.27	0.00
V_SFTMP.bsn	-3.96	0.00	-4.17	0.00	-9.03	0.00	-9.52	0.00
V_SMFMX.bsn	-4.04	0.00	-4.16	0.00	-11.25	0.00	-11.84	0.00
V_SMFMN.bsn	-4.15	0.00	-4.18	0.00	-6.48	0.00	-6.55	0.00
V_GW_DELAY.gw	4.34	0.00	4.87	0.00	9.84	0.00	9.68	0.00
R_SOL_K(..).sol	4.46	0.00	4.22	0.00	1.61	0.11	1.53	0.13
V_ALPHA_BF.gw	52.31	0.00	51.67	0.00	28.85	0.00	27.01	0.00

For final uncertainty analysis, the results of sensitive parameters were joined in all four-model (Table 8). By choosing the sensitive parameters which are same in all the four models and combing the ranges by taking the minimum and maximum range of the parameter in all models. The models are sensitive to surface runoff, soil bulk, basin parameters, groundwater delay, snowfall, and snowmelt parameters. Ten sensitive parameters were used for the uncertainty analysis of the model without use of the parameters related to snow.

Table 8. Sensitive parameters in the four models (gray color shows parameters chosen for uncertainty analysis)

Parameter Name	Low_bound	Up_bound	New_low_bound	New_up_bound
a__RCHRG_DP.gw	-0.03	0.07	-0.06	0.07
r__CN2.mgt	-0.16	0.00	-0.23	0.05
r__SOL_BD.sol	-0.15	0.43	-0.44	0.43
r__SOL_K.sol	0.85	11.40	-0.50	16.68
v__ALPHA_BF.gw	-0.10	0.39	-0.18	0.39
v__EPCO.hru	0.13	1.04	0.13	1.49
v__ESCO.hru	0.49	0.95	0.49	0.95
v__GW_DELAY.gw	-6.73	63.96	-6.73	87.64
v__GWQMN.gw	1.69	5.08	1.69	5.08
v__REVAPMN.gw	2.47	10.43	-1.51	10.43
v__SFTMP.bsn	-7.31	-0.34	-7.31	1.39
v__SMFMN.bsn	-22.66	8.08	-22.66	10.38
v__SMFMX.bsn	0.24	16.77	-5.85	16.77
v__SMTMP.bsn	0.07	4.38	-1.70	4.38
v__SNOCOVMX.bsn	47.58	78.63	32.18	78.83

3.2. Model Evaluation Performance

The model performance of models in the 95 percentile results showed satisfactory results for all model setups. The performance of model m3 and m4 with HWSD soil and range of 0.47-0.61 and 0.48- 0.61 respectively were similar to each other and better than m1 and m2 with range of 0.38-0.52 and 0.42- 0.54 and input soil of EstSoil-EH (Figure 9).

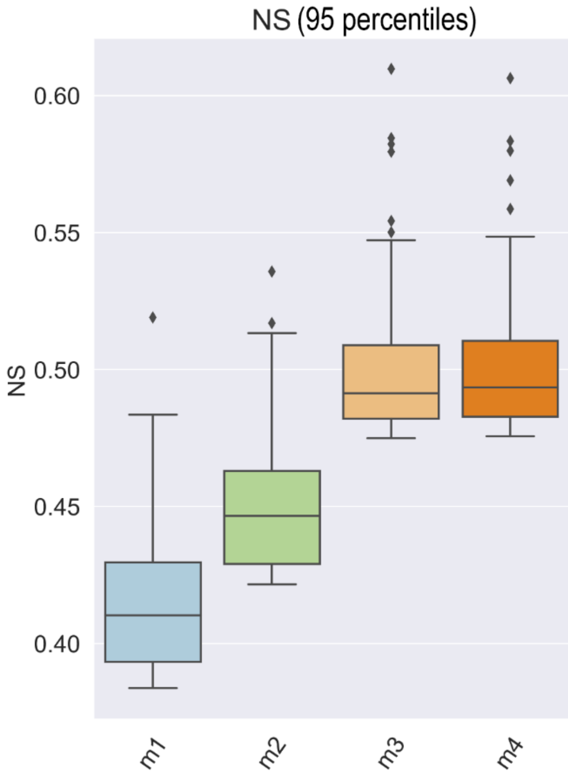

Figure 9: Models range in 95 percentiles (5% best simulations) for four model setups

The (95 Percent Prediction Uncertainty) 95PPU plot for model m1 and m2 indicates that 95PPU plot is not completely covering the observed flow, especially in low flows (Figure 10 and Figure 11).

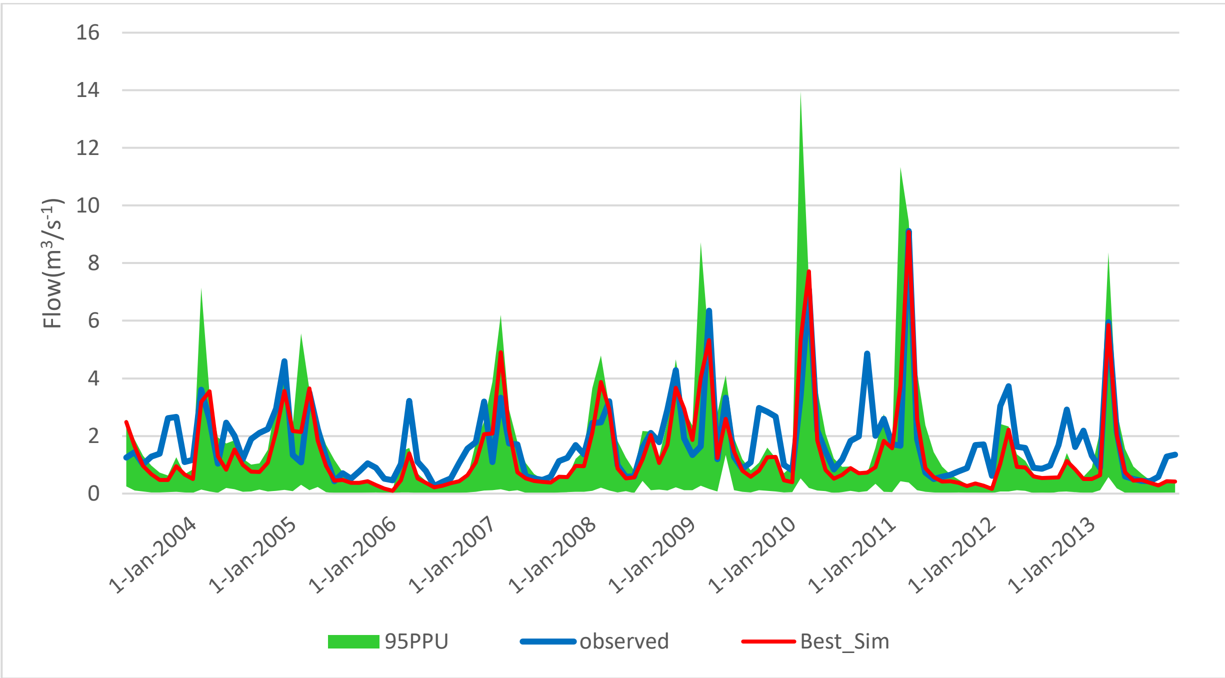

Figure10: Simulated streamflow for model m1 in 95PPU.

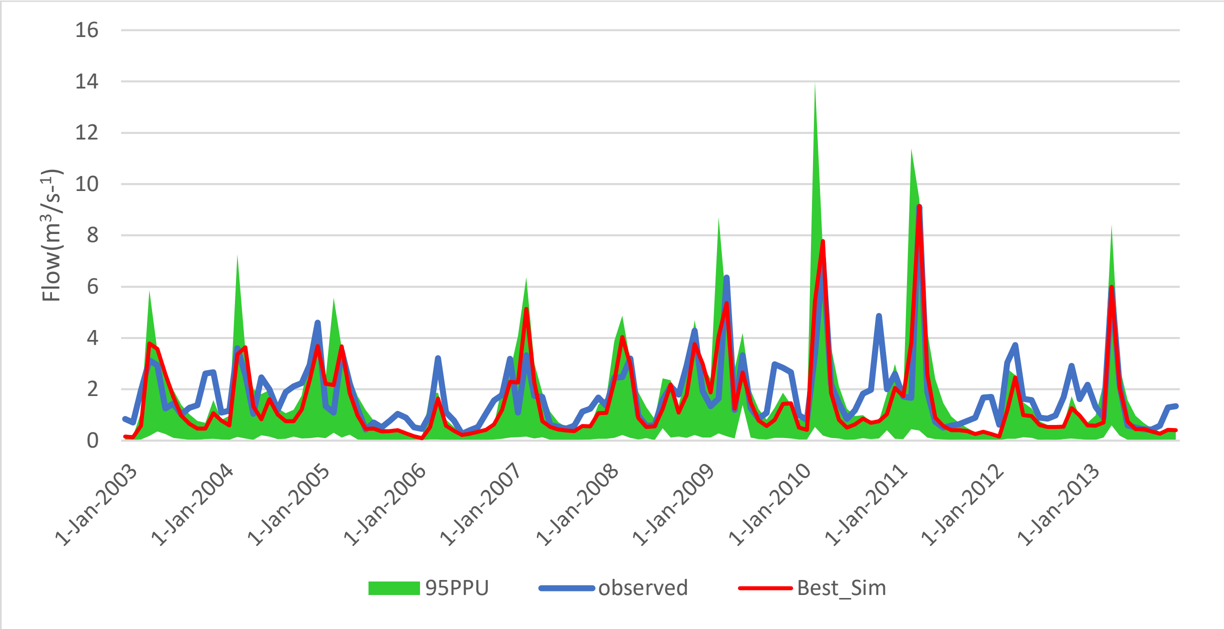

Figure 11: Simulated streamflow for model m2 in 95PPU.

The 95PPU plot for model m3 and m4 with global soil shows that these models performed better and most of the low flows were covered by model simulation (Figure 12 and Figure 13).

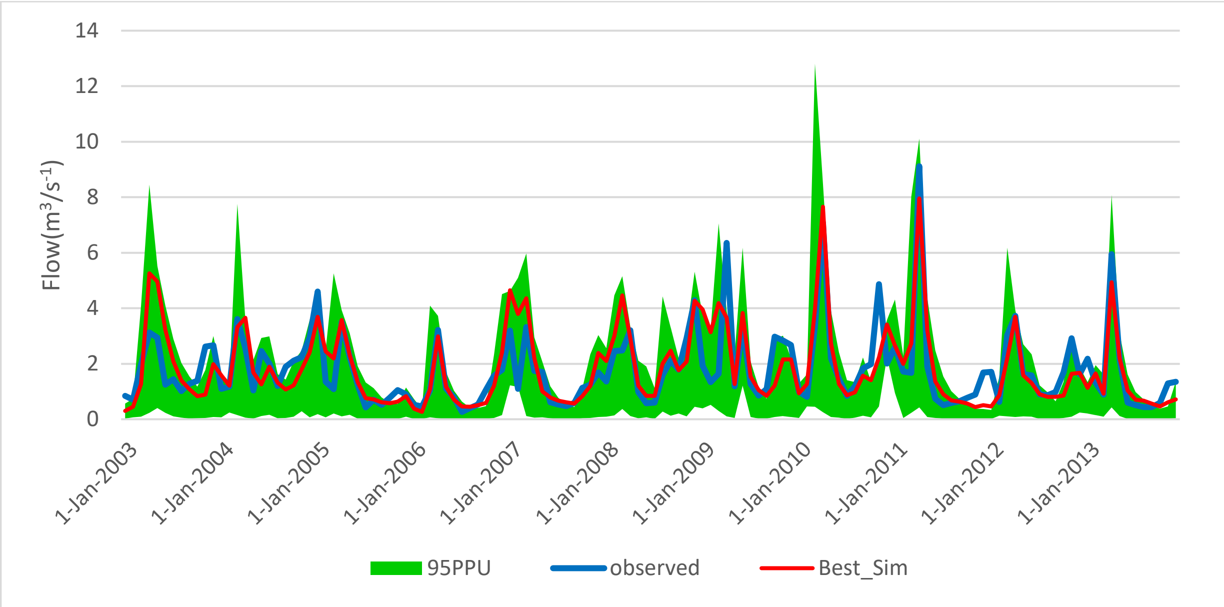

Figure 12: Simulated streamflow for model m3 in 95PPU.

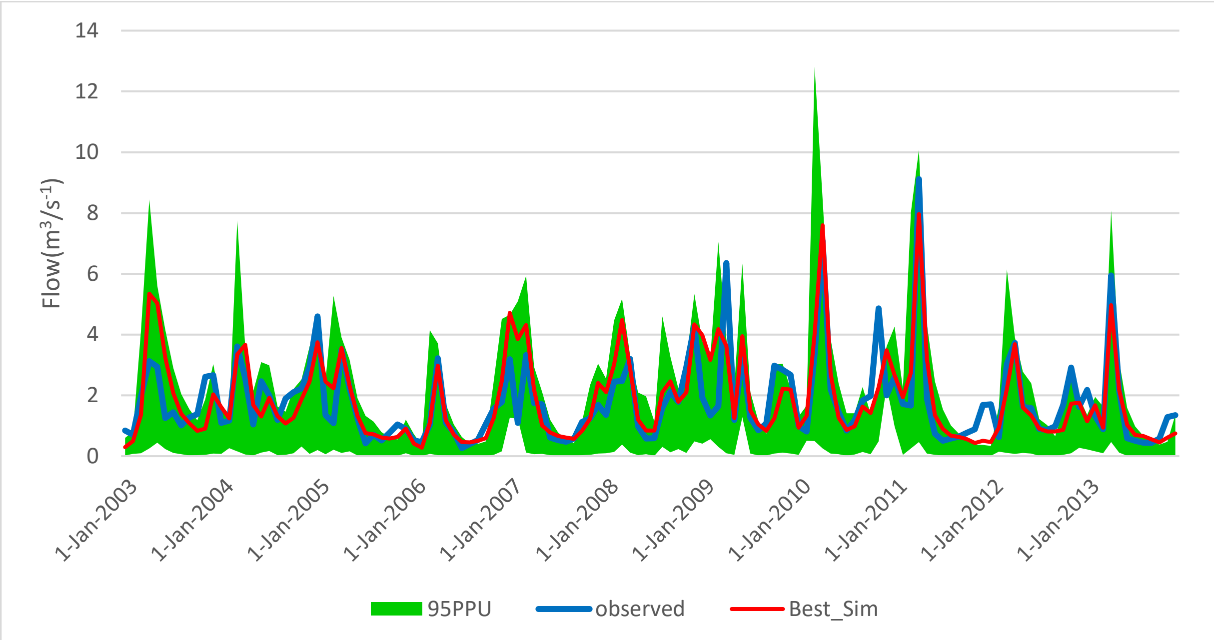

Figure 13: Simulated streamflow for model m4 in 95PPU.

3.3. Parameter uncertainty of the top 5% simulations

Uncertainty analyses were done based on ten sensitive parameters, RCHRG_DP, REVAPMN, GWQMN, ESCO, EPCO, GW_DELAY, ALPHA_BF, CN2, SOL_K, SOL_BD, related to generic and input dataset-specific parameters. The uncertainty result for the parameter based on the objective function of NSE shows that parameters range in models with the same soil data is almost similar to each other. The shorter the range of the model's parameter, the less the uncertainty associated with it. The parameter range of ALPHA_BF, CN2, SOL_K, and SOL_BD have great difference between 2 model pairs, m1-m2 and m3-m4, while ESCO and GW_DELAY had moderate difference other parameters did not show any significant difference between models with different input data.

3.3.1. Parameter uncertainty of ALPHA_BF

The baseflow alpha-factor (ALPHA_BF) is a direct index of groundwater flow response to changes in recharge. The parameter range varies from less than 0.05 to around 0. 4 (Figure 14). Models m1 and m2 have slight positive skewness and similar variation compared to m3, and m4. Models m3 and m4 have a slightly wider distribution than m1 and m2 (Figure15).

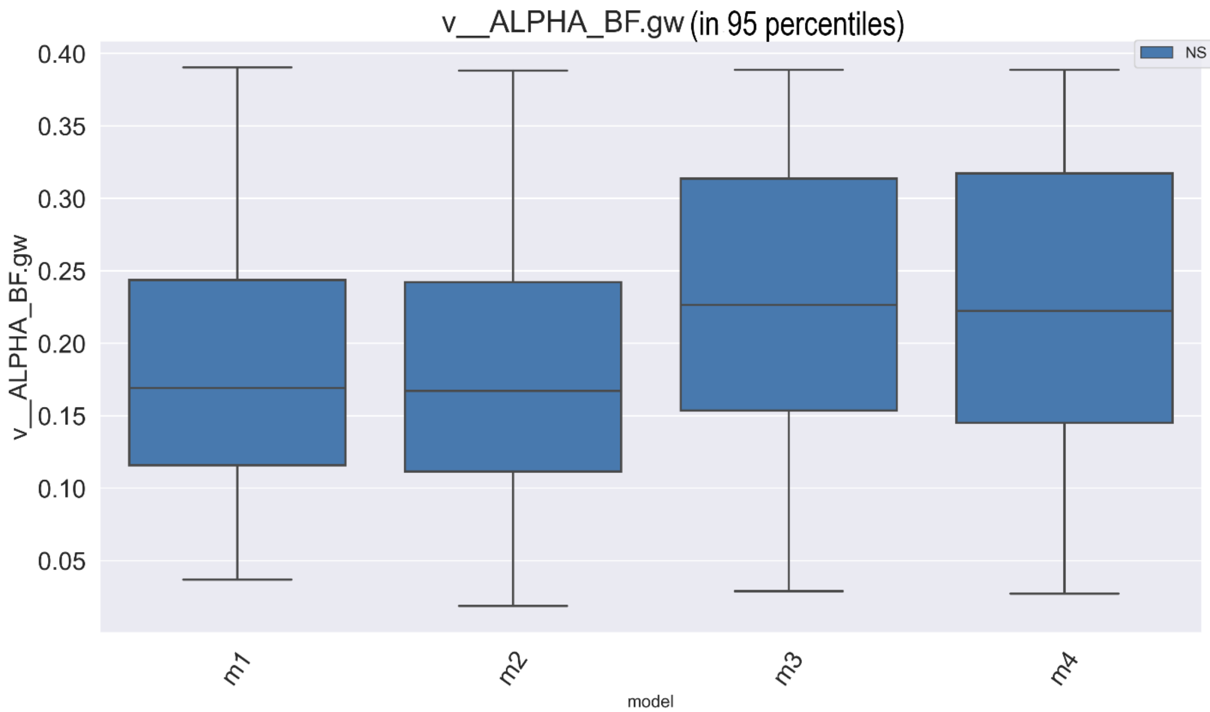

Figure 14: Range of the parameter ALPHA_BF in the four models.

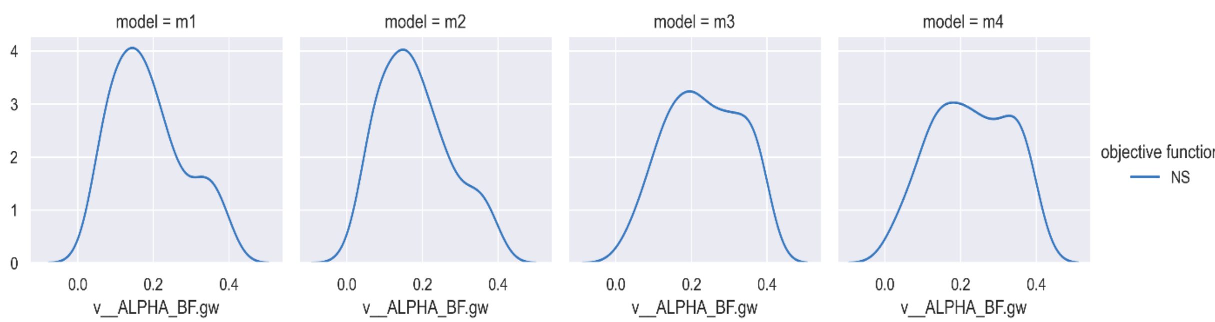

Figure 15: Distribution plot of ALPHA_BF for the simulation in 95 percentiles.

In Mann-Whitney U test, the behavior of the parameter in models is compared to each other. The highest value, 1, indicates the highest similarity between models, and 0 demonstrates the significant difference between models. Mann-Whitney U test results indicated (Table 9) that m1 and m2 with 0.74 had similar results and were utterly different from m3 and m4.

Table 9. Mann-Whitney U test result for parameter ALPHA_BF

	m1	m2	m3	m4
m1	1.00	0.74	0.00	0.00
m2	0.74	1.00	0.00	0.00
m3	0.00	0.00	1.00	0.74
m4	0.00	0.00	0.74	1.00

3.3.2. Parameter uncertainty of CN2

The parameter CN2(Initial Soil Conservation Service (SCS) runoff curve number for moisture condition II) is the function of the soil's permeability, land use, and antecedent soil water condition. The range of this parameter in m1 and m2 is similar to each other and ranges between -0.16 to 0, while m3 and m4, the parameter range is slightly less and less dispersed (Figure 16). In addition, m3 and m4 have slight positive skewness (Figure 17). The parameter range in model m4 has the least dispersion, uncertainty and ranges between -0.16 to -0.02.

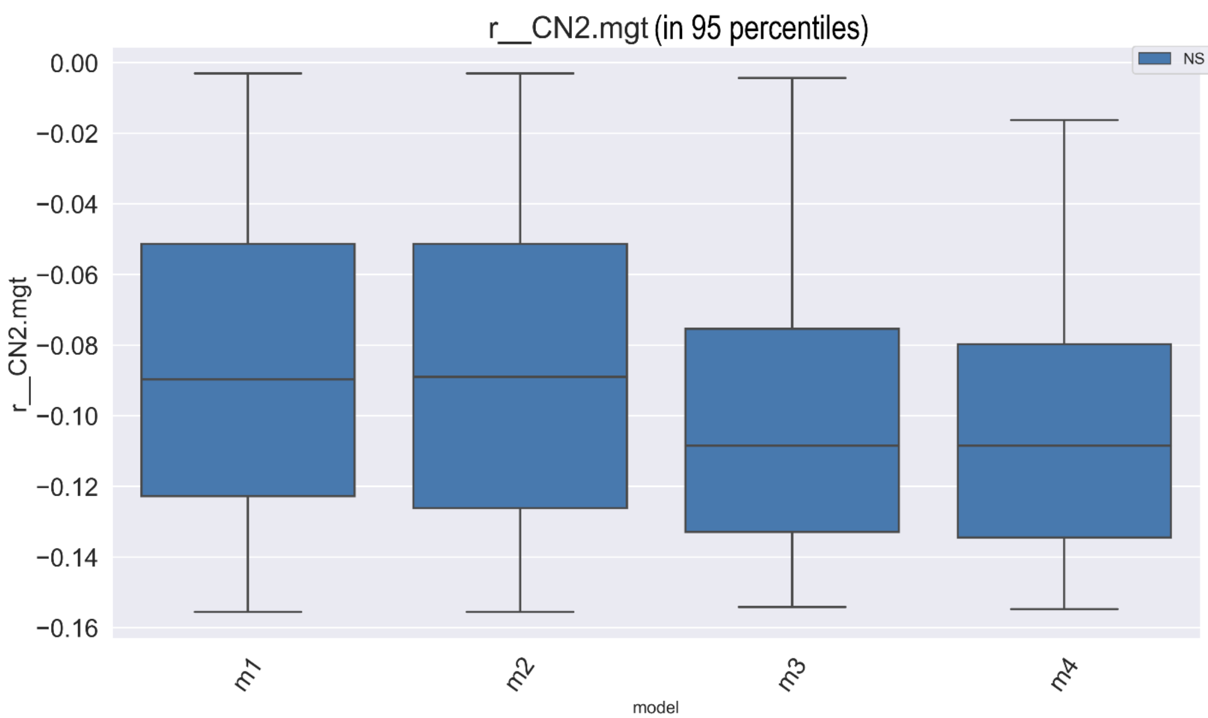

Figure 16: Range of the parameter CN2 in the four models.

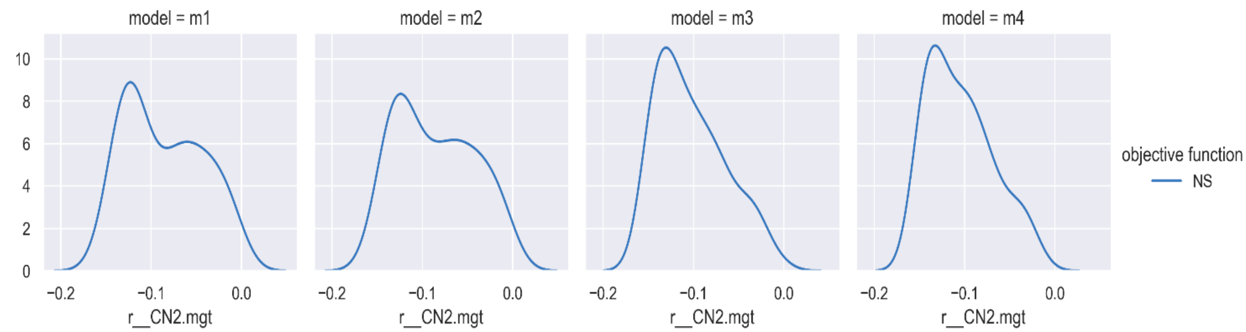

Figure 17: Distribution plot of CN2 for the simulation in 95 percentiles.

The Mann-Whitney U test result for this parameter showed that the parameter results in model m1 and m2 with value 1 are strongly similar. Model m3 and m4 with a test result of 0.72 are homogeneous. In contrast, model m3 and m1, and m2 with the result of 0.01 are slightly identical but completely disparate from model m4 (Table 10).

Table10. Mann-Whitney U test result for parameter CN2

	m1	m2	m3	m4
m1	1.00	1.00	0.01	0.00
m2	1.00	1.00	0.01	0.00
m3	0.01	0.01	1.00	0.72
m4	0.00	0.00	0.72	1.00

3.3.3. Parameter uncertainty of SOL_K

Parameter saturated hydraulic connectivity SOL_K relates soil water flow rate (flux density) to the hydraulic gradient and measures the ease of water movement through the soil. The range of the parameter in models m1 and m2 is between 2 to 15 (Figure18), has negative skewness and is less scattered than model m2 and m3 with range of around 1.5 to 15 (Figure 19).

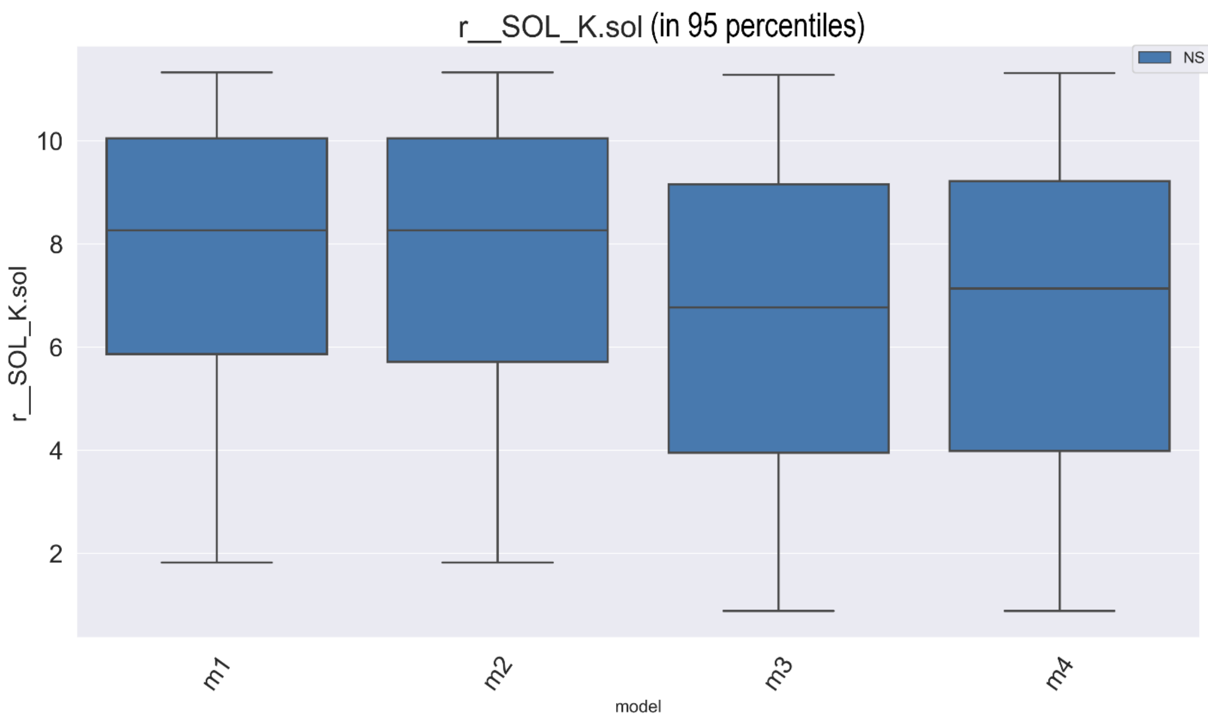

Figure 18: Range of the parameter SOL_K in the four models.

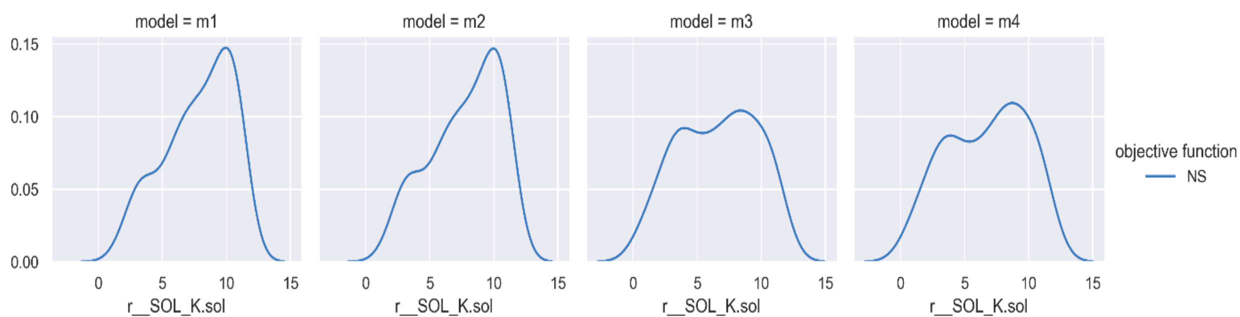

Figure 19: Distribution plot of SOL_K for the simulation in 95 percentiles.

Mann-Whitney U test for this parameter for model m1 and m2 is 0.97 high similarity between models while for model m3 and m4 is 0.77. Models m1, m2, and m3 with 0.01 had a low similarity, and m1, m2 and m4 ,0.02 were different (Table 11).

Table 11. Mann-Whitney U test result for parameter SOl_K

	m1	m2	m3	m4
m1	1.00	0.97	0.01	0.02
m2	0.97	1.00	0.01	0.02
m3	0.01	0.01	1.00	0.77
m4	0.02	0.02	0.77	1.00

3.3.4 Parameter uncertainty of SOL_BD

Moist bulk density SOL_BD expresses the ratio of the mass of the solid particles to the total volume of the soil. The range of this parameter is -0.01 to 0.4 for the four models (Figure 20). The result for m1 and m2 had negative skewness (Figure 21).

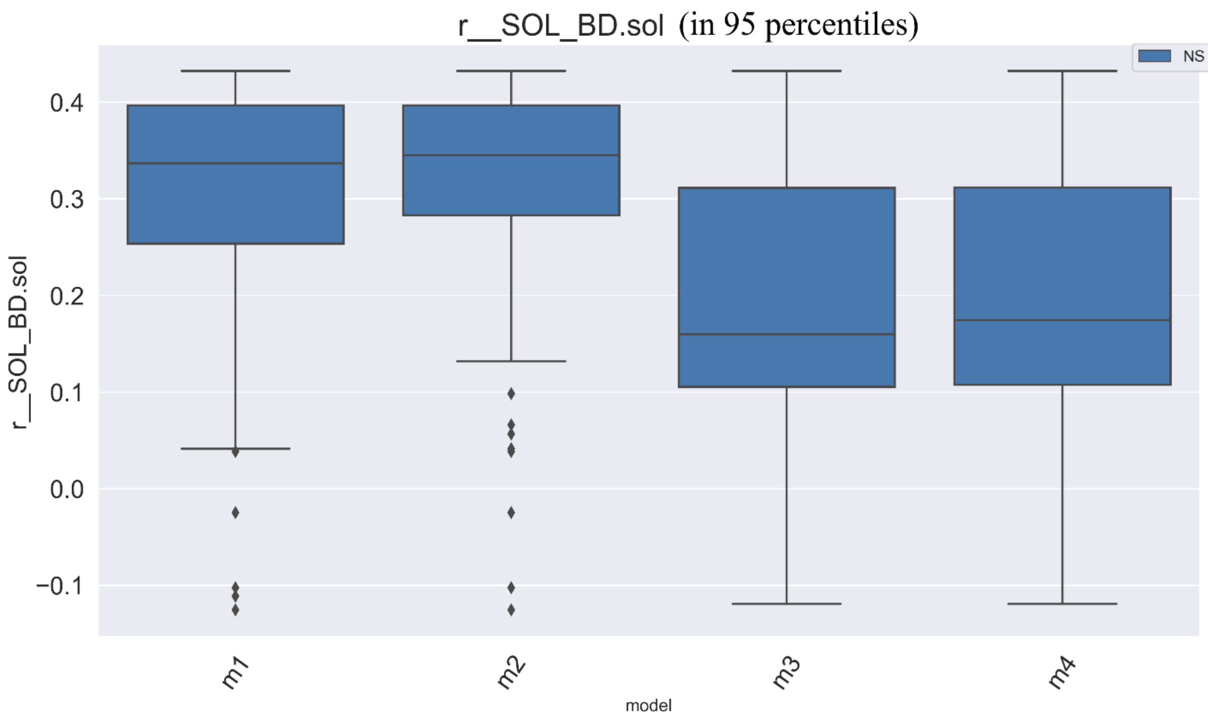

Figure 20: Range of the parameter SOL_BD in the four models.

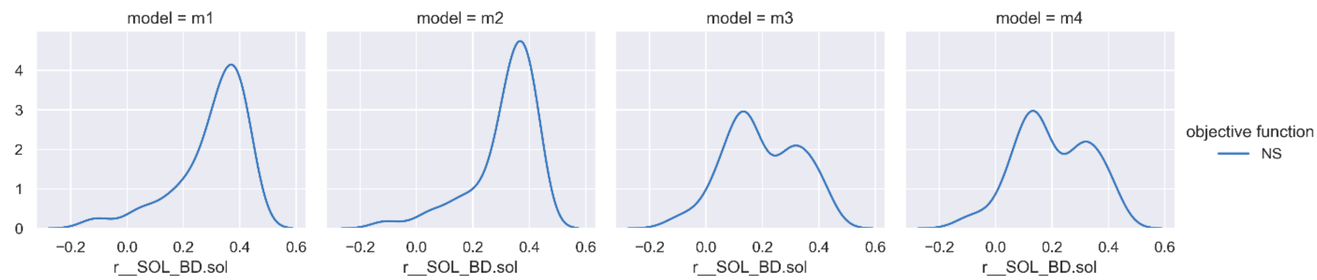

Figure 21: Distribution plot of SOL_BD for the simulation in 95 percentiles.

The results of Mann-Whitney U also indicated that m1 and m2 with 0.56 test result are similar to each other while they are strongly different from m3 and m4 (Table 12).

Table 12. Mann-Whitney test result for parameter SOL_BD

	m1	m2	m3	m4
m1	1.00	0.56	0.00	0.00
m2	0.56	1.00	0.00	0.00
m3	0.00	0.00	1.00	0.84
m4	0.00	0.00	0.84	1.00

3.3.5. Parameter uncertainty of EPCO

Parameter plant uptake compensation factor EPCO range of this parameter is slightly different between models m1- m2 and m3-m4 (Figure 22). There is a slight positive skewness in the four models (Figure 23).

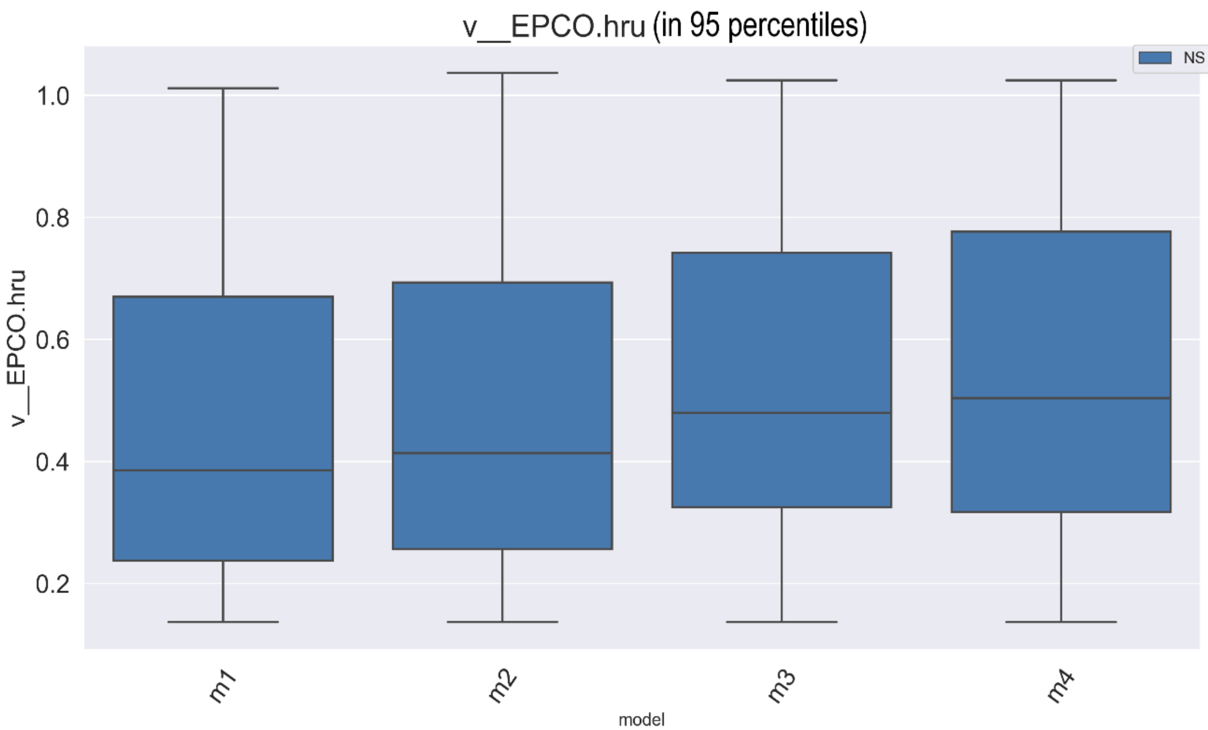

Figure 22: Range of the parameter EPCO in the four models.

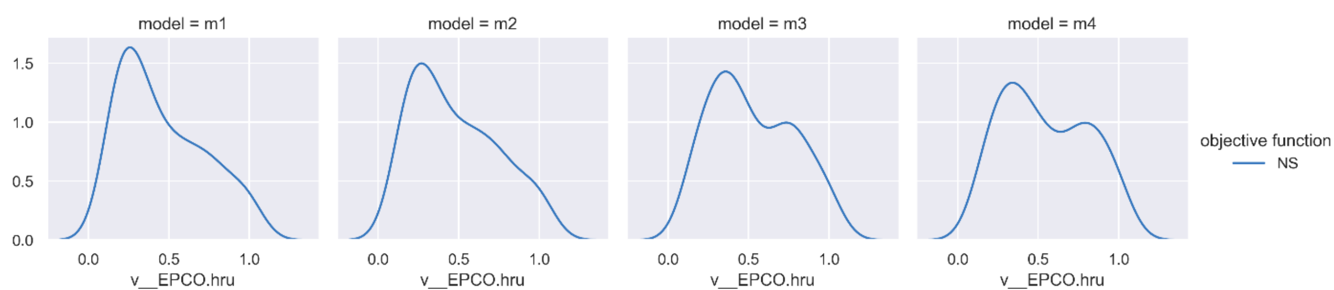

Figure 23: Distribution plot of EPCO for the simulation in 95 percentiles.

Mann-Whitney U tests showed that m1 and m4 with 0.84 test results are more similar than m1-m3 with 0.74 and m1-m2 with 0.54 test results. Model m2 and m3 with 0.23 are different from each other (Table 13).

Table 13. Mann-Whitney U test result for parameter EPCO

	m1	m2	m3	m4
m1	1.00	0.54	0.74	0.84
m2	0.54	1.00	0.24	0.52
m3	0.74	0.24	1.00	0.43
m4	0.84	0.52	0.43	1.00

3.3.6. Parameter uncertainty of GW_DELAY

Parameter range for groundwater delay time (GW_DELAY) is slightly different between model pairs m1-m2 and m3-m4 (Figure 24). Models m3 and m4 have slight negative skewness (Figure 25).

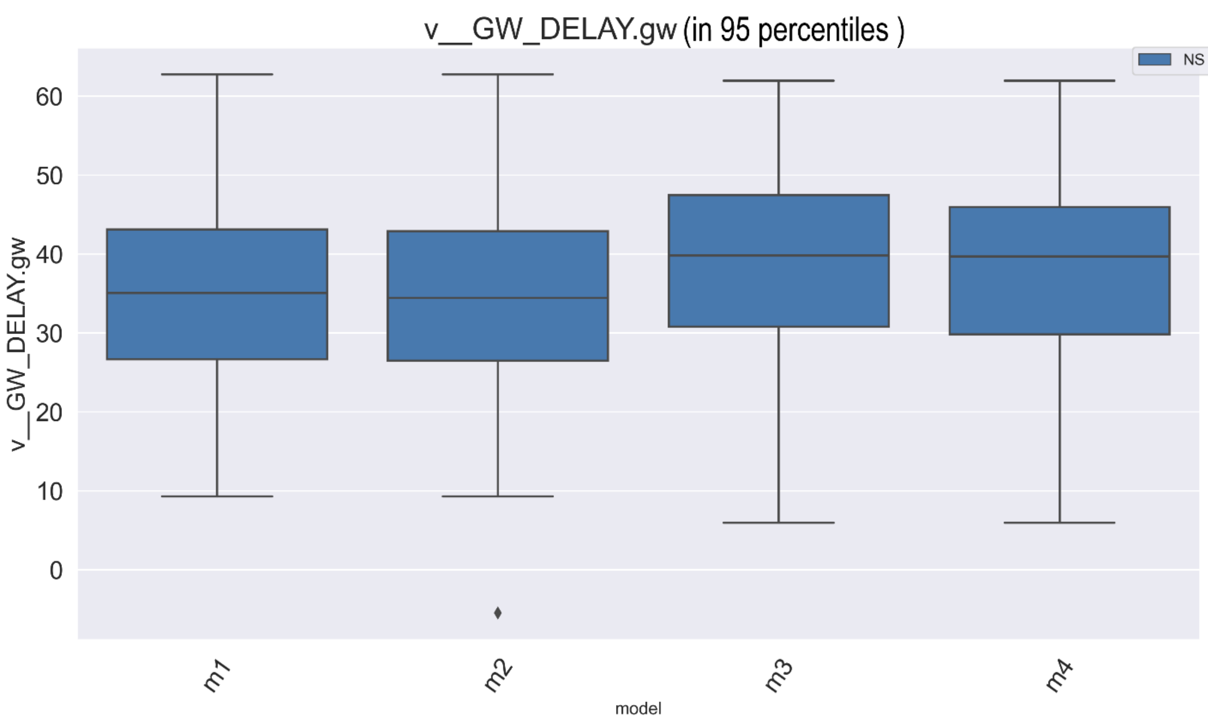

Figure 24: Range of the parameter GW_DELAY in the four models.

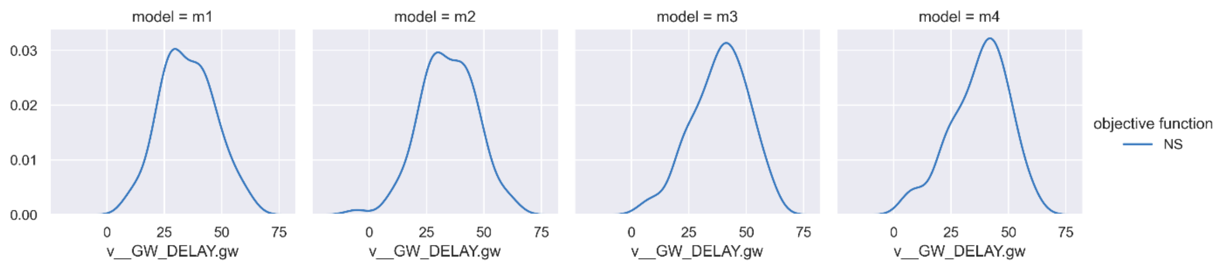

Figure 25: Distribution plot of GW_DELAY for the simulation in 95 percentiles.

Mann-Whitney U test result indicates that m1 and m2 are similar to each other with 0.78 test results, while m3 and m4 with 0.89 have the highest similarity (Table 14).

Table 14. Mann-Whitney U test result for parameter GW_DELAY

	m1	m2	m3	m4
m1	1.00	0.78	0.64	0.54
m2	0.78	1.00	0.86	0.75
m3	0.64	0.86	1.00	0.89
m4	0.54	0.75	0.89	1.00

3.4 Validation

Validation of model was done during the period 2014–2018, the performance of the models after validation is unsatisfactory. Maximum NSE values for model m3 and m4 with 0.38 and 0.39 respectively is better than models m1 and m2 with 0.26 and 0.29 values.

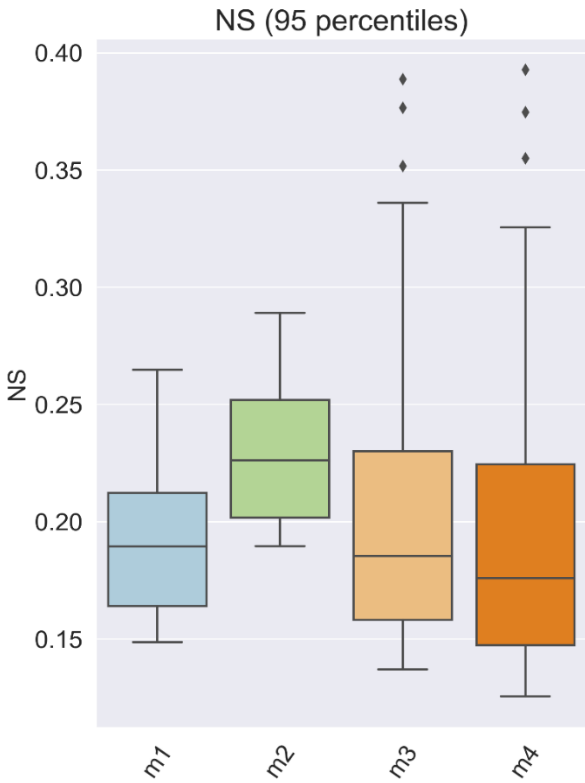

Figure 26: Models range in Validation period (5% best simulations) for four model setups

The (95 Percent Prediction Uncertainty) 95PPU plot for validation period shows that model simulation in models m1 and m2 with EstSoil-EH is not covering most of the flow peaks (Figure 27 and Figure 28).

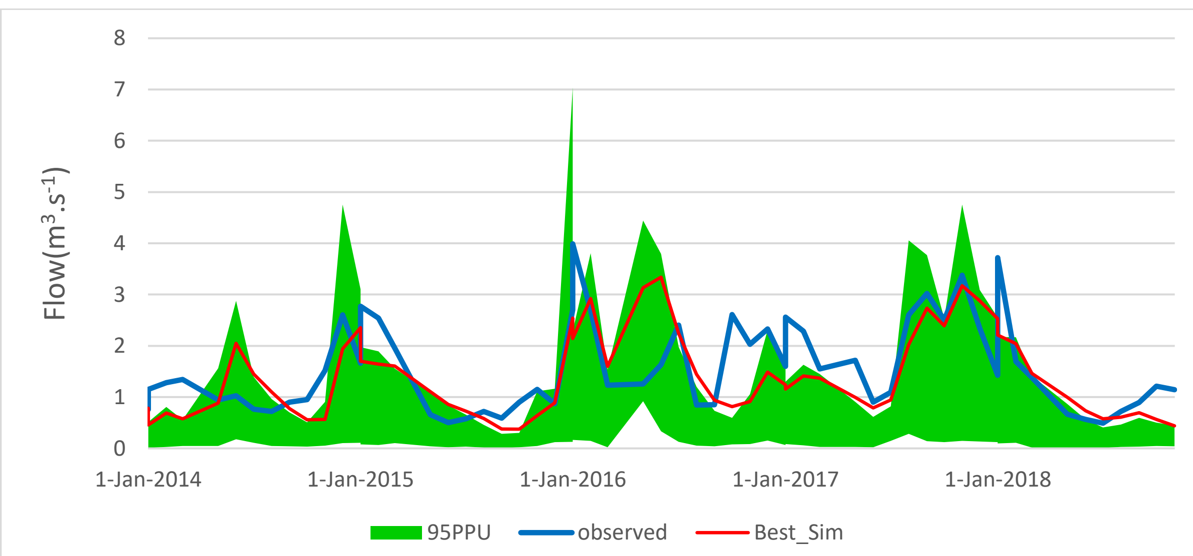

Figure 27: Simulated streamflow for model m1 in 95PPU in validation period.

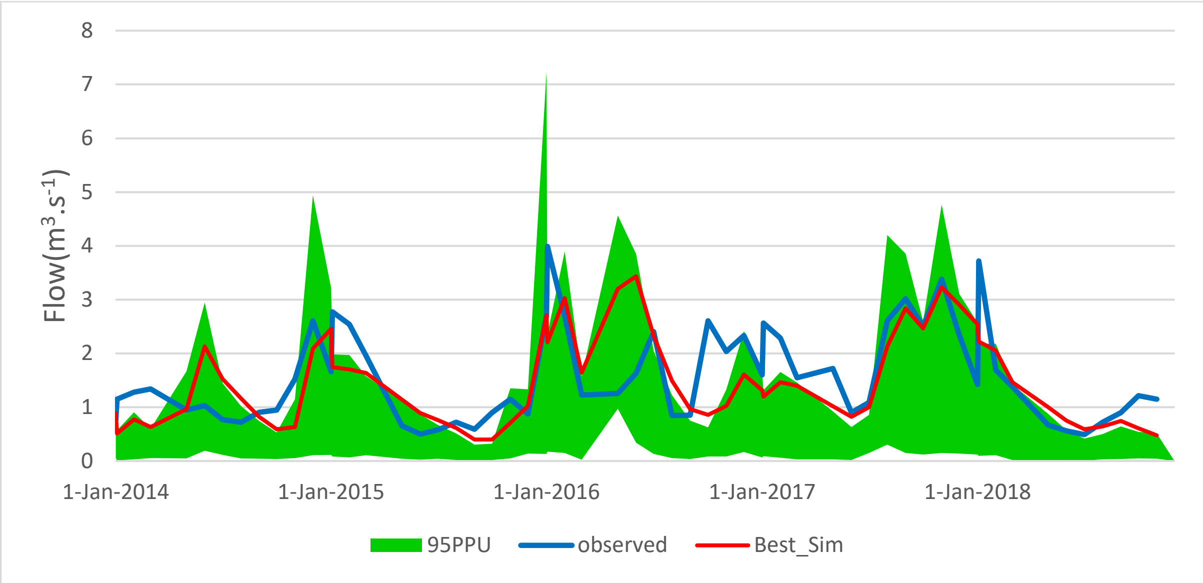

Figure 28: Simulated streamflow for model m2 in 95PPU in validation period.

The 95PPU plot for model m3 and m4 with global soil shows some flows were not covered by model simulation (Figure 29 and Figure 30).

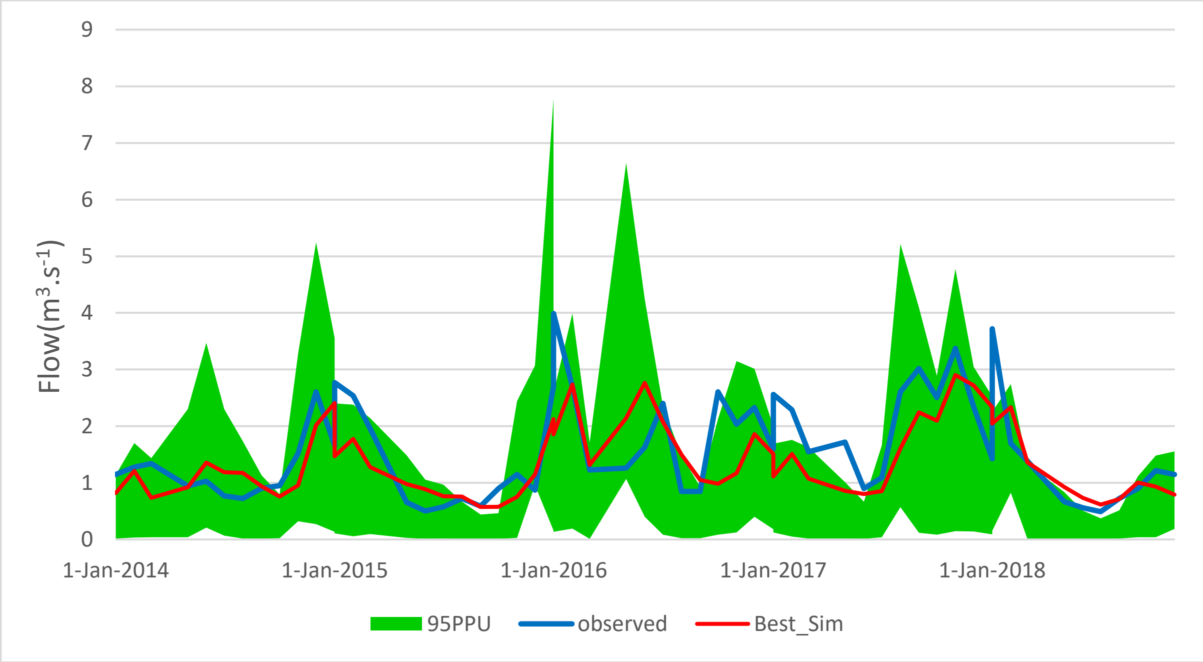

Figure 29: Simulated streamflow for model m3 in 95PPU in validation period.

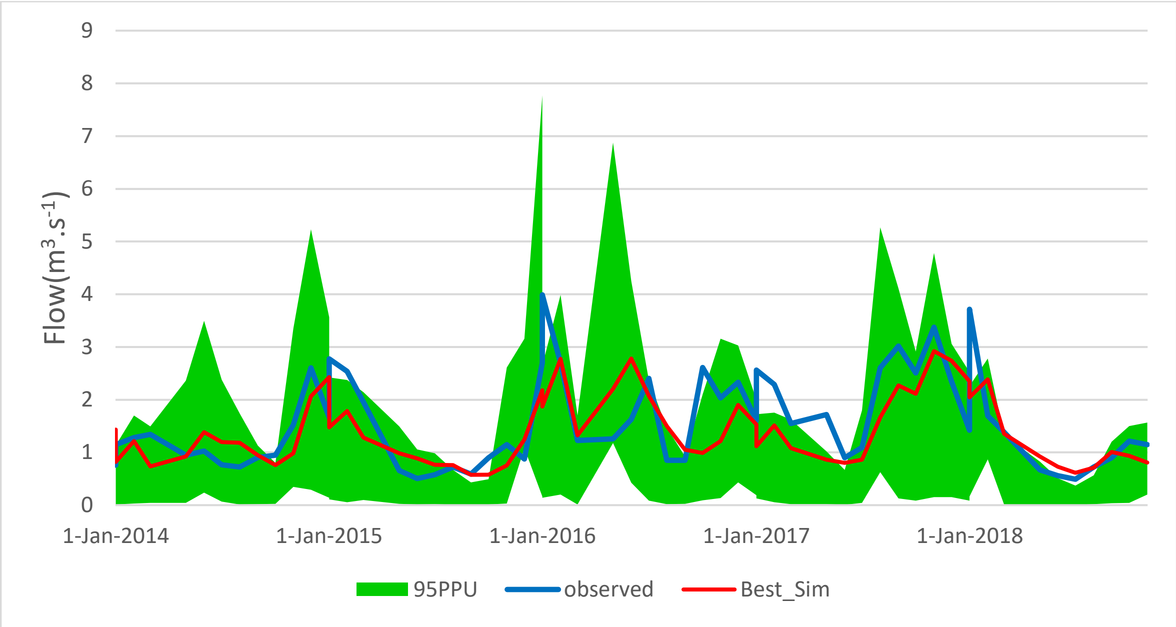

Figure 30: Simulated streamflow for model m4 in 95PPU in validation period.

3.5. Parameter uncertainty in Validation period of the top 5% simulations

The uncertainty result for the parameter based on the objective function of NSE shows that parameters behavior in models in calibration and validation period is similar. The parameter range of ALPHA_BF, GW_DELAY, SOL_K did not change significantly, while ESCO and EPCO had slightly more difference in calibration and validation period. Parameter CN2 had moderate change but SOL_BD had obvious change after the validation period.

3.5.1. Parameter uncertainty of SOL_BD in validation period

Similar to calibration period the range of this parameter is -0.01 to 0.4 for models m1 and m2(Figure 31). The result for m1 and m2 had negative skewness (Figure 32). After validation, the range of the parameter is 0-0.4 there for the uncertainty of the models m1 and m2 decreased.

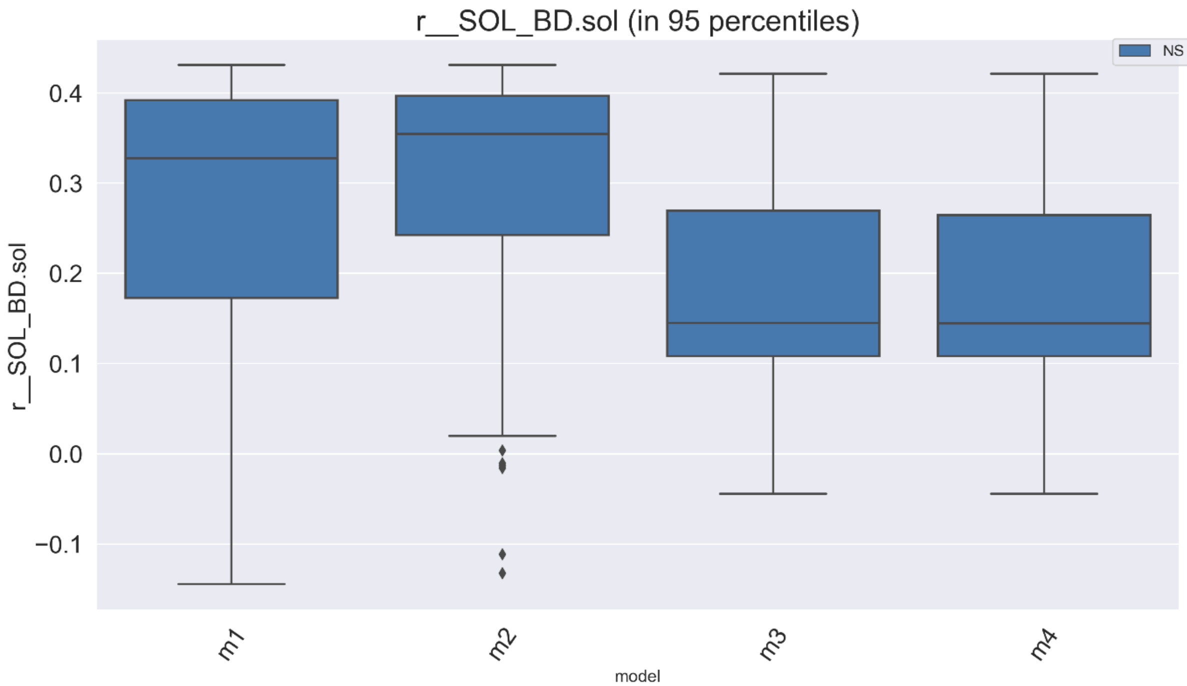

Figure 31: Range of the parameter CN2 in the four models after validation.

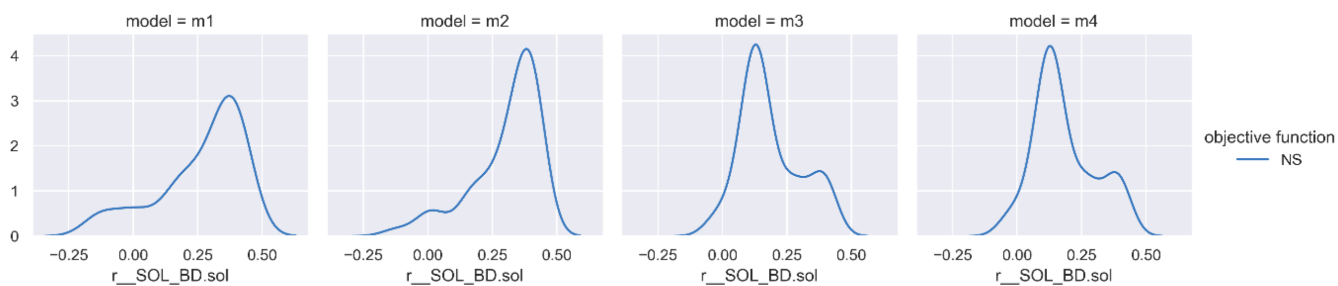

Figure 25: Distribution plot of CN2 for the simulation in 95 percentiles in validation period.

4.Discussion

Using a model to simulate reality has proven to be challenging due to the many possible errors (Troutman, 1983). SWAT as a semi-distributed hydrological model requires spatial information as input data, such as topography (elevation), LULC (land use/landcover), soils, or (hydro-) geology. The modeling results could be sensitive to the quality and nature of input variables, and interpretation of model output is limited to the resolution and quality of input environment data (Wagenet & Hutson,1996).

The goal of the research was to examine the impact of input data quality on the SWAT model's sensitivity. For achieving the runoff simulation and better representation of the catchment area, the time-intensive and careful effort of calibration of SWAT model parameters was made. Twenty-one parameters were applied in model calibration following previous studies (Guo & Su,2019; Koo et al., 2020) that suggested them as potentially sensitive. After careful parameter adjustment, fifteen parameters proved to be statistically significantly sensitive in this study. The models are highly sensitive to surface runoff, soil bulk, basin parameters, groundwater delay, snowfall, and snowmelt parameters.

The uncertainty analysis was done based on the ten sensitive parameters for the top 5% (95 percentiles) simulations of each model. The result of uncertainty analysis showed that the NSE values for models' range are satisfactory results for four models with NSE>0.5. The models are highly sensitive to snowmelt, temperature, and snowfall due to the persistent snowpack in the catchment; these parameters were not used in uncertainty analysis for better illustration of the effect of land use and soil data on the model.

High-quality reginal soil and land use/cover input did not improve SWAT model's predictive reliability, especially in the simulation of low flows. The more detailed soil map generated a lower range, while the models with global soil maps showed a higher range and better representation of high flows. Similar to previous studies by Camargos et al. (2018) high flows were predicted more efficiently in the four model setups compared to lower flows suggesting that swat model is slightly affected by special input data for representation of high flows. The variance between model's performance and better performance of the models with global soil and land use/landcover shows that low flows are sensitive to the different various input data; according to Wagenet & Hutson (1986) modeling results could be sensitive to the quality and nature of input variables.

For finding the uncertainty of the model with global input compared to reginal data, the uncertainty of parameters in 95 percentiles were analyzed. Models with the same soil map performed similarly to each other. In addition, models with CORINE land cover as input slightly performed slightly better than ETAK land use. This result is in line with Luzio et al. (2017) study that showed, the low-resolution land use model had just slightly better predictive reliability. The parameter uncertainty results indicated a strong difference for four parameters, AlPHA_BF, CN2, SOL_K, and SOL_BD, between models with different soil maps and moderate difference for parameters EPCO and GW_DELAY. After validation, the performance of all models was unsatisfactory. Behavior of the parameters did not change significantly after the validation the highest change was in SOL_BD. The overall lower model performances in validation period and missed peaks in flow simulations are understandable duo to the shorter validation period.

The amount of baseflow contribution from groundwater to the stream is influenced by many basin characteristics such as climate, soil, topography, and landcover. In this study, as climate and topography input data were consistent in model input, there was no significant difference between models with different land use; therefore, inconsistency in soil data can affect the baseflow in models. Studies show that the lower infiltration, high rates of evapotranspiration, and runoff reduce baseflow in the catchment (Brutsaert, 2005, Gardner et al., 2010, Price, 2011). The parameter uncertainty result for baseflow contribution from groundwater (AlPHA_BF) shows just a slight difference between models with HWSD and EstSoil-EH data inputs. Overall, the positive range of the parameter and lower performance of the models in low flow simulation, especially in EstSoil-EH, indicates a model difference in groundwater parameter distribution and its attempt to increase the baseflow.

The runoff curve number (CN2) parameter is one of the land use parameters in SWAT, showing surface runoff. One of the critical properties that affect runoff is infiltration, and if infiltration decreases, runoff increases, and groundwater storage decrease(McCulloch & Robinson, 1993, Nie et al., 2011). The result for this parameter shows that model m3 and m4 have less variability and need minor adjustment than model m1, m2. As the parameter range for CN2 is negative, the model attempts to decrease this parameter value, especially in models with EstSoil-EH, and therefore increase groundwater.

The parameter soil bulk density (SOL_BD) indicates the density of the soil. Higher bulk density makes the soil less porous and reduces the infiltration of the water. The parameter SOL_BD showed a similar result with a strong tendency toward higher values in models with EstSoil-EH input. The estimation of Sol_BD in Est Soil-EH was directly extracted from soil organic carbons (SOC). The pedotransfer model used for estimation of the SOC and BD shows an anti-proportional relationship. Therefore, higher SOC values cause lower BD values (Kmoch et al., 2019). The strong tendency for the higher BD values in the parameter range indicates lower values of the BD in EstSoil-EH. These results are in line with the recent feedback of EstSoil-EH, which shows that higher SOC values, especially for Agricultural fields.

Saturated hydraulic conductivity (Sol_K) shows the velocity of the water's infiltration through the soil. The higher density soil has less infiltration and less hydraulic connectivity. Parameter uncertainty results (Figure 18) show that Sol_K in models with EstSoil-EH data input was significantly skewed towards the higher values. Models' strong attempt to increase the range of this parameter shows the effect of the lower bulk density and higher infiltration results on this parameter.

The parameter (EPCO) is the plant evaporation compensation. A decrease of the ESCO value results in a reduction of streamflow, as more water is extracted from lower levels of the soil layer to meet the evaporative demands. In addition (GW_REVAP) implies water movement from the shallow aquifer into the overlying unsaturated soil layers. As the evaporation rate increases up to the potential evapotranspiration rate, it reduces, thereby the baseflow. Results showed a moderate difference between models for these parameters, and the model attempted to increase the range of these parameters to increase streamflow slightly.

5. Conclusion

Spatial input data such as elevation, land use, and soil or geology is essential for all semi-distributed or distributed hydrological models. The quality, resampling methods, and spatial resolution of these spatial input datasets introduce a great deal of uncertainty into the model. Only a few studies have been using SWAT to assess streamflow or water quality; none have analyzed the impact of spatial data on sensitivity and uncertainty in-depth on model performance with reginal data sources (Wielgat et al., 2021; Čerkasova et al., 2019).

This study provides detailed insights into the impact of different input data quality on the SWAT model's sensitivity. Spatial data with different resolutions can impact the model performance. Four model setups were analyzed by using global and reginal land use and soil data. Performance of the all-model setups was satisfactory and models with same soil data performed similar.

Studies done by Camargos et al. (2018), shows that high flows were predicted more efficiently in the four model setups than lower flows, suggesting that the swat model is slightly affected by special input data for representation of high flows. There is a notable difference among the setups when predicting low flows. Setups with HWSD soil were more efficient in the representation of the low flows compared to setups with EstSoil-EH. Models with the same soil map performed similarly to each other.

Parameter uncertainty in 95 percentiles showed the less distribution of parameters for models with HWSD soil map. The higher range of the parameters and strong tendency toward higher values indicates a possible inconsistency in EstSoil-EH data due to the higher values of soil organic carbons and lower soil bulk density estimation in this map. Validation result showed lower performance models, but this result is comprehensible duo to the shorter validation time and the goal of research to confirm parameters behavior and not to simulate the best model.

In conclusion, better quality data did not increase the performance of the swat model, and global input data represented better model performance. It should be concluded that the improvement of the EstSoil-EH soil map can result in better model performance and decrease the uncertainty of the input data parameters in this catchment as this input is directly affecting the hydrological modeling.

Kokkuvõte

SWAT mudeli tundlikkus ruumilistele sisendandmetele Porijõe valgla näitel

Hüdroloogia uurib Maal oleva vee liikumist, jaotumist ja kvaliteeti. Vett esineb nii vedelal, tahkel kui aurustunud kujul. Vee oleku faasid on omavahel seotud veeringe tsüklite kaudu, mis kirjeldavad vee liikumist maapinna ja atmosfääri vahel ning selle talletumist taimestikus, mullastikus ja veekogudes. Veehulga ja -kvaliteedi hindamise oluliseks osaks on hüdrosfääri protsesside mõistmine (Salas et al., 2014). Veeressursside efektiivset majandamist raskendab hüdroloogiliste süsteemide, sh pinna-, põhja- ja tarbevee vaheliste vastastikuste mõjude ja protsesside keerukus (Sophocleous, 2002; Srivastava et al., 2013). Antud protsesside uurimiseks on vajalik nende lihtsustamine ja modelleerimine hüdroloogilise modellerimise abil (Nyeko, 2010).

Vooluhulga modelleerimiseks on kasutatud mitmeid mudeleid (Krysanova et al., 2017; Bormann et al., 2009; Exbrayat et al., 2014), millest kõige levinum on *Soil and Water Assessment Tool* ehk SWAT (Fu et al., 2019; Ray, 2018). Seejuures on seni vähe uuritud SWATi sisendina kasutatavate ruumiandmete eraldusvõime mõju vooluhulga või veekvaliteedi modelleerimise tulemustele (Wielgat et al., 2021; Čerkasova et al., 2019). Hüdroloogiliste mudelite sisendiks on uurimisala ruumiandmed, näiteks kõrgusinfo, maakasutus, mullastik ja geoloogia. Sisendandmete kvaliteet, resolutsioon ja selle teisendamine võivad oluliselt mõjutada mudeli määramatust. Käesoleva uuringu eesmärgiks oli selgitada välja sisendandmete eraldusvõime mõju SWAT mudeli parameetrite tundlikkusele ja määramatusele Porijõe valgla näitel, võrreldes selleks globaalseid sisendandmeid kohalikega. Hüpoteesi kohaselt eeldati, et kõrgema resolutsiooniga kohalike andmete kasutamine parandab mudeli toimimist uurimispiirkonnas.

ArcSWAT keskkonnas loodi neli erineva sisendandmete kombinatsiooniga mudelit. Neist esimese (M1) puhul kasutati Eesti mullastikuandmestikku EstSoil-EH ja Euroopa maakatte andmestikku CORINE. Teise (M2) puhul kasutati mõlemal juhul kohalikke andmeid, kolmanda (M3) puhul globaalset mullaandmestikku HWSD ja kohalikku maakasutust ning neljanda (M4) sisendandmed olid vastavalt HWSD ja CORINE. Mudeli kalibreerimiseks valiti kirjanduse põhjal 21 parameetrit. Vaadeldud ja modelleeritud vooluhulga vahelise varieeruvuse vähendamiseks viidi läbi mudeli kalibreerimine ja määramatuse analüüs tarkvarapaketi SWAT-CUP algoritmi SUFI-2 abil (Abbaspour, 2009). Kalibreerimise käigus teostati 13 mudeli iteratsiooni ja tundlikkuse analüüsi jaoks tehti kokku 2000 simulatsiooni. Algoritm tuvastas 15 tundlikku parameetrit, mida kasutati seejärel tundlikkuse analüüsil.

Mudelite prognoosivõimet hinnati Nash–Sutcliffe'i koefitsiendiga (NSE). NSE jäi M1 puhul vahemikku 0.38–0.52, M2 puhul 0.42–0.54, M3 0.47–0.61 ning M4 puhul vahemikku 0.48–0.61. Seega oli prognoosivõime parem globaalset mullaandmestikku kasutanud mudelite M3 ja M4 korral. Määramatuse analüüs näitas, et mulla lasuvustiheduse ja küllastunud veejuhtivuse väärtusi iseloomustas negatiivne asümmeetria. Mann–Whitney test tuvastas, et kohalikku mullandmestikku rakendanud mudelite M1 ja M2 parameetrid erinesid statistiliselt oluliselt M3 ja M4 omadest. Eesti mullakaardi EstSoil-EH jaoks tuletati lasuvustihedus orgaanilise mullasüsiniku (SOC) põhjal ning saadud lasuvustiheduse ruumiline jaotus ei ole täielikus kooskõlas varasemate samalaadsete uuringutega. Seega võib oletada, et mudel võis SOC väärtusi teatud maakasutuse klasside (nt põllumaa) puhul üle hinnata.

References

-Abbaspour, K.C.,(2009). SWAT-CUP2. *SWAT Calibration and Uncertainty Programs*, Version 2.

-Abbaspour, K.C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J., Zobrist, J., Srinivasan, R., (2007b) . Modelling hydrology and water quality in the pre-alpine/ alpine Thur watershed using SWAT. *Journal of Hydrology* 333 (2), 413e430.

-Arabi, M., Govindaraju, R.S., Hantush, M.M. (2006). Role of watershed subdivision on evaluation of long-term impact of best management practices on water quality. J. Am. *Water Resour*. Assoc. 42, 513–528.

-Arheimer, B.; et al.(2017) , Intercomparison of regional-scale hydrological models and climate change impacts projected for 12 large river basins worldwide_A synthesis. *Environmental Research Letters*, 12(10), 105002. https://doi.org/10.1088/1748-9326/aa8359.

- Arnold, J. G., Srinivasan, R., Muttiah, R. S., & Williams, J. R. (1998). Large area hydrologic modeling and assessment part I: Model Development. Journal of the American Water Resources Association, 34(1), 73–89. https://doi.org/10.1111/j.1752-1688.1998.tb05961.x.

-Arnold, J.G., Srinivasan, R.S. Muttiah, Williams J.R. (1998). Large Area Hydrologic Modeling and Assessment Part I: Model Development. *Journal of the American Water Resources Association* 34(1):73-89. DOI: 10.1111/j.1752-1688. 1998.tb05961. x.

- Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., Santhi, C., Harmel, R. D., Griensven, A. van, Van Liew, M. W., Kannan, N., M. K. Jha, M. K. (2012). SWAT: model use, calibration and validation, *Transactions of the ASABE*. 55(4): 1491-1508.

-Beven, K. (2016). Facets of uncertainty: Epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication. *Hydrol. Sci*. J., 61, 1652–1665.

-Bieger, K., Arnold, J.G., Rathjens, H. J. White, M.D., Bosch, D. M., Allen, P., Volk, M., Srinivasan, R. (2017). Introduction to SWAT+, a completely restructured version of soil and water assessment tool. USA: *journal of American water resources.* vol. 53, No. 1.

-Brutsaert, W. (2005). Hydrology: An Introduction. Cambridge University Press, New York.

-Bormann, H.; Breuer, L.; Gräff, T.; Huisman, J.A.; Croke, B. (2009) Assessing the impact of land use change on hydrology by ensemble modelling (LUCHEM) IV: Model sensitivity to data aggregation and spatial (re-)distribution. Adv. *Water Resour*, 32, 171–192.

-Breuer, L.; Vaché, K.B.; Julich, S.; Frede, H.-G. (2008). Current concepts in nitrogen dynamics for mesoscale catchments. *Hydrol. Sci*. J., 53, 1059–1074.

-Camargos, C., Julich, S., Houska, T., Bach, M., & Breuer, L. (2018). Effects of Input Data Content on the Uncertainty of Simulating Water Resources. *Water*, *10*(5), 621. https://doi.org/10.3390/w10050621

-Chaplot, V. (2014). Impact of spatial input data resolution on hydrological and erosion modeling: Recommendations from a global assessment. *Physics and Chemistry of the Earth, Parts A/B/C*, *67–69*, 23–35. https://doi.org/10.1016/j.pce.2013.09.020.

- Copernicus Land Monitoring Service. (2021). CORINE Land Cover — European Environment Agency (EEA) -Copernicus Programme. Retrieved January 7, 2021, from https://land.copernicus.eu/pan-european/corine-land-cover.

-Čerkasova, N., Umgiesser, G., & Ertürk, A. (2019). Assessing Climate Change Impacts on Streamflow, Sediment and Nutrient Loadings of the Minija River (Lithuania): A Hillslope Watershed Discretization Application with High-Resolution Spatial Inputs. *Water*, *11*(4), 676. https://doi.org/10.3390/w11040676

-Di Luzio, M., Arnold, J.C., Srinivasan. R. (2005). Effect of GIS data quality on watershed stream flow and sediment simulations. *I Hydrological processes* 19.629-6311.

-Do, N. C., Razavi, S. (2020). Correlation Effects? A Major but Often Neglected Component in Sensitivity and Uncertainty Analysis. *Water Resources Research*, *56*(3). https://doi.org/10.1029/2019wr025436.

-Dooge, J. C. (1972). Mathematical models of hydrologic systems. Proceedings of the International Symposium on Modeling Techniques in Water Resources Systems, Ottawa, Canada, 1:171-189.

-Douglas-Mankin, K.R., Srinivasan, R., Arnold, J. (2010). Soil and water assessment tool (SWAT) model: current developments and applications. *Trans. ASABE* 53. 1423–1431.

-ETAK, Estonian Land Board, landuse types, retrieved in 2020 or 03.01.2021 https://geoportaal.maaamet.ee/eng/Maps-and-Data/Estonian-Topographic-Database/Download-Topographic-Data-p618.html.

-Exbrayat, J.-F.; Buytaert, W.; Timbe, E.; Windhorst, D.; Breuer, L. (2014). Addressing sources of uncertainty in runoff projections for a data scarce catchment in the Ecuadorian Andes. Clim. Chang, 125, 221–235.

-Fischer, G., F. Nachtergaele, S. Prieler, H.T. van Velthuizen, L. Verelst, D. Wiberg. (2008). *Global Agro-ecological Zones Assessment for Agriculture (GAEZ 2008)*. IIASA, Laxenburg, Austria and FAO, Rome, Italy.

-Fu, B., Merritt, W.S., Croke, B.F.W., Weber, T.R., Jakeman, A.J. (2019). A review of catchment-scale water quality and erosion models and a synthesis of future DISCover from 1 km AVHRR data. Int. J. Rem. Sens. 21, 1303–1330. https://doi.org/10.1080/014311600210191.

-Gardner, W.P., Susong, D.D., Solomon, D.K., Heasler, H. (2010). Snowmelt hydrograph interpretation: revealing watershed scalehydrologic characteristics of the Yellowstone volcanic plateau. J. Hydrol. 383 (3–4), 209–222.

-Guo, J., & Su, X. (2019). Parameter sensitivity analysis of SWAT model for streamflow simulation with multisource precipitation datasets. *Hydrology Research*, *50*(3), 861–877. https://doi.org/10.2166/nh.2019.083.

-He, C. (2003). Integration of geographic information systems and simulation model for watershed management. Environ. Model. *Softw*. 18, 809–813.

-Heathman, G.C. & Larose, M. & Ascough, James. (2009). Soil and Water Assessment Tool evaluation of soil and land use geographic information system data sets on simulated stream flow. *Journal of Soil and Water Conservation*. 64. 10.2489/jswc.64.1.17.

-Hoang, L., Mukundan, R., Moore, K. E. B., Owens, E. M., and Steenhuis, T. S. (2018). "The effect of input data resolution and complexity on the uncertainty of hydrological predictions in a humid vegetated watershed", *Hydrology and Earth System Sciences*, vol. 22, no. 11, pp. 5947–5965, doi:10.5194/hess-22-5947-2018.

-Jajarmizadeh, M., Harun, S., & Salarpour, M. (2012). A review on theoretical consideration and types of models in hydrology. *Journal of Environmental Science and Technology*, *5*(5), 249–261. https://doi.org/10.3923/jest.2012.249.261.

-Khalid, K., Ali, M. F., Rahman, N. F. A., Mispan, M. R., Haron, S. H., Othman, Z., & Bachok, M. F. (2016). Sensitivity Analysis in Watershed Model Using SUFI-2 Algorithm. *Procedia Engineering*, *162*, 441–447. https://doi.org/10.1016/j.proeng.2016.11.086.

-Kmoch, A., Kanal, A., Astover, A., Kull, A., Virro, H., Helm, A., Ostonen, I., & Uuemaa, E. (2019). *derived from the Soil Map of Estonia*. *October* 1–29. https://doi.org/10.5281/ZENODO.3473290.

Koo, H., Chen, M., Jakeman, A. J., & Zhang, F. (2020). A global sensitivity analysis approach for identifying critical sources of uncertainty in non-identifiable, spatially distributed environmental models: A holistic analysis applied to SWAT for input datasets and model parameters. *Environmental Modelling & Software*, *127*, 104676. https://doi.org/10.1016/j.envsoft.2020.104676.

-Kumar, S.; Merwade, V. (2009). Impact of Watershed Subdivision and Soil Data Resolution on SWAT Model Calibration and Parameter Uncertainty. *J. Am. Water Resour. Assoc*, 45, 1179–1196.

-Luzio, Mauro & Leh, Mansoor & Cothren, Jackson & Brahana, J. & Asante, Kwasi. (2017). Effects of land use land-cover data resolution and classification methods on SWAT model flow predictive reliability. International Journal of Hydrology Science and Technology. 7. 39. 10.1504/IJHST.2017.10001829.

Ma, L., J. C. Ascough II, L. R. Ahuja, M. J. Shaffer, J. D. Hanson, and K. W. Rojas. (2000). Root Zone Water Quality Model sensitivity analysis using Monte Carlo simulation. *Trans. ASAE* 43(4): 883-895.

-Ma, S. (1993). Integrating GIS and Remote Sensing Approaches with NEXRAD to Investigate the Impact of Spatial Scaling of Hydrologic Parameters on Storm-Runoff Modeling. Ph.D. Dissertation, University of Oklahoma, Norman, Oklahoma.

-McCulloch, J. S., & Robinson, M. (1993). History of forest hydrology. Journal of Hydrology, 150(2–4), 189–216. https://doi.org/10.1016/0022-1694(93)90111-l.

-Mander, Ü., Kuusemets, V. and Treier, K. (1994). Variation in groundwater quality in South Estonian rural areas. IAHS Proc. of GQM ’93: Groundwater Quality Management, Tallnin, Estonia, 6-9 September 1993. *IAHS Publ*. No. 220, pp. 81-93.

-MERIT Hydro: global hydrography datasets. (2019). MERIT Hydro: Global Hydrography Datasets. http://hydro.iis.u-tokyo.ac.jp/%7Eyamadai/MERIT_Hydro/.

-Montecelos-Zamora, Y., Cavazos, T., Kretzschmar, T., Vivoni, E.R., Corzo, G., Molina-Navarro, E. (2018). Hydrological Modeling of Climate Change Impacts in a Tropical River Basin: A Case Study of the Cauto River, Cuba. *Water*, *10*, 1135.

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

-Moriasi, D., Arnold, J., Van Liew, M., Bingner, R., Harmel, R., Veith, T., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE* 50 (3), 885e900.

-Moriasi, D., & Starks, P. (2010). Effects of the resolution of soil dataset and precipitation dataset on SWAT2005 streamflow calibration parameters and simulation accuracy. *Journal of Soil and Water Conservation*, 65(2), 63–78. https://doi.org/10.2489/jswc.65.2.63.

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L.(2007). Model evaluation guidelines for systematic quantification of accuracy in watershed

simulations. *Am. Soc. Agric. Biol*. Eng. 50, 885–900.
-Nachar, N. (2008). The Mann-Whitney U: A Test for Assessing Whether Two Independent Samples Come from the Same Distribution. Tutorials in Quantitative Methods for Psychology, 4(1), 13–20. https://doi.org/10.20982/tqmp.04.1.p013
-Nash, J., Sutcliffe, J. (1970). River flow forecasting through conceptual models' part IdA discussion of principles. *Journal of Hydrology* 10 (3), 282e290.
-Neitsch, S., Arnold, J., Kiniry, J., Williams, J., King, K. (2005). Soil and Water Assessment Tool: Theoretical Documentation, Version 2005. Texas, USA.

Neitsch, S. L., Arnold, J. G., Kiniry, J. R., & Williams, J. R. (2011). Soil & Water Assessment Tool Theoretical Documentation Version 2009. *Texas Water Resources Institute*, 1–647. https://doi.org/10.1016/j.scitotenv.2015.11.063.

-Nyeko, M. (2010). Land use change in ASWA basin-northern Uganda: opportunities and constrains to water resources management, University of Naples Federico *II*, faculty of agriculture.

-Pai, Naresh & Saraswat, Dharmendra. (2013). Impact of land use and land cover categorical uncertainty on SWAT hydrologic modeling. *Transactions of the ASABE*. 56. 1387-1397. 10.13031/trans.56.10062.

-Patel DP, Gajjar CA, Srivastava PK. (2013). Prioritization of Malesari mini watersheds through morphometric analysis: a remote sensing and GIS perspective. *Environ Earth Sci* 69:2643–2656.

-Piniewski, M., Szcześniak, M., Huang, S. and Kundzewicz, Z. W. (2018). Projections of runoff in the Vistula and the Odra river basins with the help of the SWAT model, *Hydrol. Res*., 49(2), 303–317, doi:10.2166/nh.2017.280.

-Price, K. (2011). Effects of watershed topography, soils, land use, and climate on baseflow hydrology in humid regions: a review.Prog. *Phys. Geogr*. 35 (4), 465–492.
-Ray, L.K., (2018). Limitation of automatic watershed delineation tools in coastal region.

Ann. GIS 24, 261–274. https://doi.org/10.1080/19475683.2018.1526212.

-Rosenthal, W.D., and D.W. Hoffman. (1999). Hydrologic modeling/GIS as an aid in locating monitoring sites. *Trans. ASAE* 42(6): 1591-1598.

- Salas, J. Govindaraju, R., Anderson, M. A., Arabi, M., Francés, F., Suarez, W., Lavado, W., Green, T. (2014). Introduction to Hydrology.10.1007/978-1-62703-595-8_1.

-Sharma, A.; Tiwari, K.N. (2014). A comparative appraisal of hydrological behavior of SRTM DEM at catchment level. J. *Hydrol*, 519, 1394–1404.

-Shirmohammadi, A., Chaubey, I., Harmel, R. D., Bosch, D. D., Muñoz-Carpena, R., Dharmasri, C., Sexton, A., Arabi, M., Wolfe, M. L., Frankenberger, J., Graff, C., & Sohrabi, T. M. (2006). UNCERTAINTY IN TMDL MODELS. *Transactions of the ASABE*, *49*(4), 1033–1049. https://doi.org/10.13031/2013.21741.

-Singh, V.P., Woolhiser, D.A. (2002). Mathematical modeling of watershed hydrology. J. *Hydrol*. Eng. 7, 270–292.

-Singh, J., Knapp, H.V., Arnold, J., Demissie, M. (2005). Hydrological modeling of the iroquois river watershed using HSPF and SWAT1. *JAWRA Journal of the American Water Resources Association* 41 (2), 343e360.

-Spruill, C.A., S.R. Workman, and J.L. Taraba. (2000). Simulation of daily and monthly stream discharge from small watersheds using the SWAT model. *Trans. ASAE* 43(6): 1431-1439.

-Sophocleous M (2002) Interactions between groundwater and surface water: the state of the science. *Hydrogeol J* 10:52–67.

-Sorooshian, S., Hsu, K.-, Coppola, E., Tomassetti, B., Verdecchia, M., & Visconti, G. (2008). Hydrological Modelling and the Water Cycle: Coupling the Atmospheric and Hydrological Models (Water Science and Technology Library, 63) (1st Corrected ed. 2008, Corr. 2nd printing 2008 ed.). *Springer*.

-Srivastava PK, Han D, Ramirez MR, Islam T (2013b) Machine learning techniques for downscaling SMOS satellite soil moisture using MODIS land surface temperature for hydrological application. *Water Resour Manag* 27:3127–3144.

-Varep, E. (1964). The landscape regions of Estonia. Publications on Geography IV. Acta et Commerstationes Universitatis Tartuensis, 156: 3-28.

-Viessman, W.J., Lewis, G.L., Knapp, J.W. (1989). Hydrologic modelling. In: Viessman, W.J., Lewis, G.L., Knapp, J.W. (Eds.), Introduction to Hydrology, third ed. Harper and Row Publishers, *New York*, pp. 491–670.

-Wagenet, R. J. and J. L. Hutson. (1996). Scale-Dependency of Solute Transport Modeling/GIS Applications. *J. Env. Qual*. 25:499-510.

-Wang, M., A. T. Hjelmfelt, and J. Garbrecht. (2000). DEM Aggregation for Watershed Modeling. *J. Amer. Water Resour*. Assoc.36(3):529-584.

-Wang, H.; Wu, Z.; Hu, C. (2015). A Comprehensive Study of the Effect of Input Data on Hydrology and non-point Source Pollution Modeling. *Water Resour. Manag*, 29, 1505–1521.

-Wielgat, P., Kalinowska, D., Szymkiewicz, A., Zima, P., Jaworska-Szulc, B., Wojciechowska, E., Nawrot, N., Matej-Lukowicz, K., & Dzierzbicka-Glowacka, L. A. (2021). Towards a multi-basin SWAT model for the migration of nutrients and pesticides to Puck Bay (Southern Baltic Sea). *PeerJ*, *9*, e10938. https://doi.org/10.7717/peerj.10938

-Troutman, B. M. (1983). Runoff Prediction Errors and Bias in Parameter Estimation Induced by Spatial Variability of Precipitation. *Water Resour. Res*. 19(3):791-810.

-Williams, J.R., A.D. Nicks, and J.G. Arnold. 1985. Simulator for water resources in rural basins. *J. Hydrol. Engr*. 111(6): 970-986.

-Wilson, J. P., W. P. Inskeep, J. M. Wraith, and R. D. Snyder. (1996). GIS-Based Solute Transport Modeling Applications: Scale Effects of Soil and Climate Data Input. *J. Environ. Qual*. 25:445-453.

-Woessner, W.W. (2012). Formulating, applying and constraining hydrological models: modeling 101. In: Water Center and School of Natural Resources Seminar. University of Nebraska, USA, pp. 1–65.

-Xu, C. (2002). Textbook of hydrological models. In: Xu, C. (Ed.), Uppsala (Sweden). Uppsala University Department of Earth Sciences Hydrology.
-Zhang X, Srinivasan R, Liew MV. (2010). On the use of multi-algorithm, genetically adaptive multi-objective method for multi-site calibration of the SWAT model. *Hydrol Process* 24:955–969.

Annexes

Annex 1. EstSoil-EH Main Soil Texture

Soil Texture	Percentage EstSoil-EH map	Percentage Generalized Map
Clay	2.6	0.4
Gravels	0.2	0.3
Loam	11.5	12.9
Loamy sand	49.9	52

Peat	17.6	5.4
Sand	11.5	28.9
Silt Loam	0.0	-
Clay loam	0.3	-
Sandy loam	6.4	-

Annex 1. Swat model input parameters

number	Parameter factor	Parameter range	Parameters unite	description
1	RCHRG_DP.gw	-0.062505 - 0.043449	-	Deep aquifer percolation fraction
2	CH_N2.rte	-0.342738 - 0.1354	-	Manning’s “n” value for the main channel
3	CN2.mgt	-0.225268 - 0.05	-	Initial SCS runoff curve number for moisture condition
4	SOL_BD.sol	-0.438979 - 0.147948	Mg/m3 or g/cm3	Moist bulk density

5	SOL_K.sol	-0.5 - 16.677691	Mm/hr	Saturated hydraulic conductivity (water movement through soil)
6	ALPHA_BF.gw	-0.176526 - 0.38496	-	Baseflow alpha factor (groundwater flow response to changes in recharge) per day
7	ALPHA_BNK.rte	-0.340375 - 0.621719	-	Baseflow alpha factor for bank storage(days).
8	EPCO.hru	0.519428 - 1.492366	-	Plant uptake compensation factor
9	ESCO.hru	0.712238 - 1.187601	-	Soil evaporation compensation factor
10	GWQMN.qw	-0.002619 - 5.818706	mmH2O	Threshold depth of water in shallow aquifer required for return flow to occur
I1	GW_DELAY.gw1	17.469163 -87.643723		The delay time
12	GW_REVAP.gw	0.117698 - 8	-	Ground water revap coefficient
13	LAT_TTIME.hru	-7.333207 -113.157242		Lateral flow travel time(days)
I14	REVAPMN.gw	-1.510072-10.387309	mmH2O	Threshold depth of water in the shallow aquifer for 'revap' or percolation to the deep aquifer to occur
15	SFTMP.bsn	-4.703057-1.389648	°C	Snowfall temperature
16	SMFMN.bsn	-15.167006-10.379701	-	Melt factor for snow on December 21
17	SMFMX.bsn	-5.847014-13.204706	-	Melt factor for snow on june21
18	SMTMP.bsn	-1.701399-2.715629	-	Snowmelt base temperature
19	SNO50COV.bsn	0.117494-0.63984	-	Fraction of snow volume represented by snocovmx that corresponds to 50% snow cover
20	SNOCOVMX.bsn	32.175999-78.831711	-	Minimum snow water content that corresponds to 100% snow cover
21	TIMP.bsn	-0.380653-1.137321	-	Snowpack temperature lag factor

Iteration number	**Calibrated parameter**
Iteration 2	v__SFTMP.bsn
Iteration 3	v__SMFMN.bsn
Iteration 4	v__SMFMX.bsn

Iteration5	v__SMTMP.bsn
Iteration6	v__SNO50COV.bsn
Iteration7	v__TIMP.bsn
Iteration8	r__CN2.mgt
Iteration9	r__CH_N2.rte
Iteration10	v__ALPHA_BF.gw- a__RCHRG_DP.gw
Iteration11	v__ALPHA_BNK.rte-v__EPCO.hru-v__ESCO.hru- v__LAT_TTIME.hru.
Iteration12	v__GWQMN.gw- v__GW_DELAY.gw- v__GW_REVAP.gw- v__REVAPMN.gw
Iteration13	r__SOL_BD.sol- r__SOL_K..sol

Annex 2. Parameters calibrated in each iteration

Non-exclusive licence to reproduce thesis and make thesis public

I, Mahdiyeh Sepehrar

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to

reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright, “Assessing the effect of spatial input data quality on the SWAT model’s sensitivity in the Porijõgi catchment” supervised by Dr. Alexander Kmoch, Evelyn Uuemaa

Mahdiyeh Sepehrar
Tartu, 23.05.2021