



## **Urban sanitary sewer modelling in cold climate**

Test case using automatic calibration and precipitation forecast in Finland

PEDRO PAULO ALMEIDA SILVA

August 15<sup>th</sup>, 2019



BRANDENBURG UNIVERSITY OF TECHNOLOGY  
COTTBUS-SENFTENBERG  
Erasmus Mundus Joint Master Programme

**This dissertation was accepted as the final thesis report of the degree of Joint Master of Science in Hydroinformatics and Water Management.**

**Academic Supervisor:** Frank Molkenthin, PhD  
Head of EuroAquaee Programme at  
Brandenburgische Technische Universität , Cottbus, Germany

**Institutional Supervisor:** Markus Sunela, PhD  
Chief Technology Officer, Fluidit Ltd.,  
Tampere, Finland

**Thesis Presentation:** August 21<sup>st</sup>, 2019

**Funding:**

Euroaquaee is a joint Master Programme funded by Erasmus+. This dissertation also received support from Fluidit Ltd. company where the student performed a professional practice with the title of Master Thesis Worker.



Funded by the  
Erasmus+ Programme  
of the European Union



**Declaration:**

*I hereby declare that this work was written independently without any unauthorized assistance or sources not given credit within the work. All words, phrases, passages, and data taken from other sources have been properly cited. No parts of this work in the same form or a similar form have ever been previously handed in to fulfill an examination.*

---

Pedro Paulo Almeida Silva



# Contents

<b>Contents</b>	<b>ii</b>
<b>List of Tables</b>	<b>iv</b>
<b>List of Figures</b>	<b>v</b>
<b>Abstract</b>	<b>vii</b>
<b>Dedication</b>	<b>vii</b>
<b>Acknowledgements</b>	<b>viii</b>
<b>Acronyms</b>	<b>ix</b>
<b>1 Introduction</b>	<b>12</b>
1.1 Sewer networks . . . . .	12
1.2 Problem description and objective of the thesis . . . . .	14
1.3 Thesis structure . . . . .	17
<b>2 Methodology</b>	<b>19</b>
2.1 Introduction . . . . .	19
2.2 Offline Model . . . . .	20
2.3 Online Model . . . . .	21
2.4 Parameter Optimization Algorithm . . . . .	23
<b>3 Sewer model</b>	<b>26</b>
3.1 Alternative methods to model rainfall dependent infiltration and inflow . . . . .	27
3.2 Physics-based hydrological model . . . . .	29
3.2.1 Rainfall-Runoff . . . . .	29
3.2.2 Snowpack and snowmelt . . . . .	31
3.2.3 Infiltration . . . . .	34
3.2.4 Aquifer and groundwater flow . . . . .	36
3.3 Unit hydrograph hydrological model . . . . .	37
3.4 Hydraulic model . . . . .	39
<b>4 Case study and model set up</b>	<b>42</b>
4.1 Jokela Town . . . . .	42
4.2 Jokela's Hydraulic Model . . . . .	43
4.3 Data and Parameter Estimation . . . . .	45
4.3.1 Sanitary Sewer Flow Data . . . . .	45
4.3.2 Decomposition of flow Components . . . . .	46
4.3.3 Meteorological data and parameters . . . . .	48

4.3.4	Precipitation data . . . . .	49
4.3.5	Temperature data . . . . .	50
4.3.6	Wind and evaporation data . . . . .	50
4.3.7	Snow depth data and snowpack parameters . . . . .	51
4.3.8	Weather Forecast . . . . .	53
4.3.9	Topographic data . . . . .	55
4.3.10	Terrain data and runoff parameters . . . . .	55
4.3.11	Soil Superficial deposits and Infiltration Parameters . . . . .	59
4.3.12	Water table and groundwater flow parameters . . . . .	62
4.3.13	synthetic unit hydrograph . . . . .	67
<b>5</b>	<b>Results and discussion</b>	<b>71</b>
5.1	Calibration . . . . .	71
5.1.1	Physics-based model calibration . . . . .	72
5.1.2	Unit hydrograph model calibration . . . . .	74
5.2	Model validation with historical data . . . . .	78
5.3	Seasonal water balance . . . . .	84
5.4	Simulations with forecast data . . . . .	85
5.5	Discussions for a continuous simulation application . . . . .	88
<b>6</b>	<b>Conclusions and recommendations</b>	<b>92</b>
6.1	Conclusions . . . . .	92
6.2	Recommendations for future research . . . . .	93
<b>Bibliography</b>		<b>95</b>
<b>A Appendix</b>		<b>100</b>

## List of Tables

3.1	Snowpack and snowmelt parameters range [38] . . . . .	33
3.2	Modified Horton infiltration parameters range[38] . . . . .	36
4.1	Estimated values of sanitary sewer network (SSN) flow components in 2018 for Jokela Catchment in cubic metres . . . . .	49
4.2	Snowpack and snowmelt estimated parameters . . . . .	53
4.3	SWMM runoff parameter estimation . . . . .	59
4.4	Modified Horton infiltration parameter estimation . . . . .	62
4.5	Groundwater flow parameter estimation . . . . .	66
4.6	Range of T and K parameters by Vallabhaneni and Burgess [50] . . . . .	70
5.1	Events used for calibration and validation. . . . .	72
5.2	Goodness-of-fit tests. Adapted from [25] . . . . .	73
5.3	Calibrated parameters for physics-based model . . . . .	73
5.4	Evaluation of events used for calibration of physics-based model parameters . . . . .	74
5.5	Range of unit hydrograph method parameters for optimization algorithm . . . . .	77
5.6	Evaluation of events used for calibration of unit hydrograph model parameters . . . . .	77
5.7	Event 5 and 6 simulation results evaluation for the number of iterations of the DDS optimization algorithm . . . . .	78
5.8	Season-based validation results of proposed hydrological models . . . . .	82
5.9	Event-based validation results of proposed hydrological models . . . . .	83
5.10	Wastewater balance of validation simulations versus data-based estimated . . . . .	85

## List of Figures

1.1	Percentage of population served by different types of sanitation system. [48] . . . . .	12
1.2	Wet-weather Infiltration . . . . .	16
1.3	Defects on the sewer network [13] . . . . .	17
1.4	Sewer overflows [4] [27] . . . . .	17
2.1	Hydrological model development flowchart . . . . .	21
2.2	System diagram an online forecasting model . . . . .	22
3.1	Wet-weather flow components. Modified from [50] . . . . .	27
3.2	Precipitation Losses relative to a Sanitary Sewer Network . . . . .	29
3.3	Non-linear reservoir model [38] . . . . .	30
3.4	Extra $D_s$ and $\%_{routed}$ for Non-linear reservoir model. Modified from Rossman and Huber [38] . . . . .	31
3.5	Seasonal variation of melt coefficients [38] . . . . .	32
3.6	Horton infiltration capacity decay and recovery curves. Modified from [38] . . . . .	35
3.7	Representation of two-zone groundwater model in SWMM [38] . . . . .	36
3.8	RTK method components for short, medium, long-term and resulting hydrographs [50]	38
3.9	Seasonal response to rainfall relationship [50] . . . . .	39
4.1	Location of Jokela in Finland and Vantaanjoki river basin . . . . .	42
4.2	Jokela's sanitary sewer network . . . . .	44
4.3	Dry-weather flow simulated versus measured in Jokela pumping station . . . . .	44
4.4	Ourliers removal . . . . .	45
4.5	RDII x DWF estimate for 2018 of total wastewater load . . . . .	46
4.6	Estimated dry-weather hydrograph obtained from EPA SSOAP tool . . . . .	47
4.7	RDII and DWF estimate for 2018 of total wastewater load of Jokela SSN . . . . .	48
4.8	Comparison of the recorded precipitation amount of the years 2018 and 2019. . . . .	50
4.9	Comparison of rain gauge and radar-based total precipitation amount measurements for Jokela's catchment . . . . .	51
4.10	Snow depth and Temperature Measurements FMI [19] . . . . .	52
4.11	Results of parameter estimation based on literature review and manual calibration of the snowpack and snowmelt parameters . . . . .	52
4.12	Numerical weather forecast models coverage. [20] . . . . .	54

4.13	Topographic data used for each process of the physics-based model . . . . .	55
4.14	Jokela subcatchment division . . . . .	57
4.15	Comparison of sewershed delineation using Filled and Burned DEM . . . . .	58
4.16	Delineation methods. D1 (left) and D3 (right) . . . . .	59
4.17	Pervious and Impervious division and land cover. . . . .	60
4.18	Jokela catchment soil superficial deposits . . . . .	61
4.19	Locations of observation wells around Jokela's catchment . . . . .	63
4.20	Groundwater table measurements. . . . .	64
4.21	Dupuit-Forcheimer lateral seepage to adjacent channel. Modified from [10] . . . . .	64
4.22	Position of h2 for each subcatchment . . . . .	65
4.23	Groundwater table first simulation from January to May of 2018 . . . . .	67
4.24	RTK unit hydrograph model scheme . . . . .	69
4.25	Representation of Jokela's Hydrological Model using RTK UH method . . . . .	70
5.1	Results of calibration and validation of the physics-based model for the dormant season	79
5.2	Results of calibration and validation of the unit hydrograph model for the dormant season . . . . .	80
5.3	Validation of the growth season . . . . .	81
5.4	Hypothetical shift on simulated flows. Results of physics-based model manually advanced in five days . . . . .	82
5.5	Observed snow depth versus simulated depth of water equivalent of 2019 . . . . .	83
5.6	Different simulated flow components of 2019 dormant season using the physics-based model . . . . .	84
5.7	Preliminary analysis of model performance with precipitation forecasted by the HARMONIE model . . . . .	87
5.8	Comparison of simulation results with different initial conditions for the physics-based model . . . . .	89

## ABSTRACT

Part of stormwater and snowmelt in cold climate regions infiltrate the urban sanitary sewer networks (SSNs) increasing the load, and the cost, for wastewater treatment plants (WWTPs). This may cause overflows when the system capacity is exceeded. This study investigated the use of an empirical and a physics-based hydrological model, available in stormwater management model (SWMM). The aim was to simulate the rainfall-derived infiltration and inflow (RDII) of a sewer network as part of an operational system able to forecast wastewater flows based on future weather conditions. The empirical model used was the RTK unit hydrograph. The physics-based model included rainfall-runoff, snowpack, snowmelt, infiltration, aquifer, and groundwater flow processes in SWMM. The hydrological models were coupled to a hydraulic model and applied to a test case in Jokela, a town located in the south of Finland. A methodology for parameter estimation of both models was proposed in this study. The results of the study suggest that both models can be used for the identification of sanitary sewer overflows (SSOs), scenario assessment, and as part of an operational forecasting system with certain level of accuracy. Better agreement with measured flow data was obtained with the physics-based model for specific conditions of cold climates with less than 1 % error on volume estimates during long-term simulations. Wastewater flow decomposition, data treatment, analysis of initial conditions, and simulations using forecasted precipitation were demonstrated and discussed. Sensitive parameters for calibration were defined and a method to limit their possible range was proposed. This seems to facilitate the implementation of an automatic calibration routine to assist the operation and management of the network on a near to real-time basis.

**Keywords:** continuous simulation, modeling, urban sanitary sewer network, sewer model, automatic calibration, SWMM5, forecast.

## **DEDICATION**

In dedication of my family that provided me full support for this journey. Especial dedication to mother, Maria Almeida da Silva who had to be far from her son.

## **ACKNOWLEDGEMENTS**

The author acknowledges EuroAqua+ Joint Master Programme and the European Union for the organisation and funding which made this study possible for me and provided the same conditions to several other students.

Professors, students, assistants, and staff of EuroAqua+ programme from all five participant universities played an important role on teaching and organising this educational programme and assist to development of this master thesis. Special gratitude to Dr. Frank Molkenthin from Brandenburg Technical University for his tutoring during this thesis.

Thanks to Fluidit Ltd. company in Finland for providing all necessary support. My sincere recognition for its owners and employees who contributed to different parts of this study. Acknowledgments to Dr. Markus Sunela for his patience and valuable guidance. Kalervo Kylätie and Timo Ranta-Pere for their proactive help and availability during the project. Janne Väyrynen, Mika Kuronen, Panu Kuitunen, Johan Ånäs, and Lauri Rantala who helped through explanations of the hydraulic model, the development of necessary auxiliary scripts, and valuable discussions and opinions.

The author Tuusula Water Utility (Tuusulan Vesihuolto) and Hanna Riihinen for providing the data used and access to their sewer network model.

## ACRONYMS

<b>ARMAX</b>	autoregressive-moving-average model
<b>API</b>	application programming interface
<b>ATI</b>	antecedent temperature index
<b>BWF</b>	base wastewater flow
<b>CSN</b>	combined sewer network
<b>CSO</b>	combined sewer overflow
<b>DDS</b>	dynamically dimensioned search
<b>DEM</b>	digital elevation model
<b>DWF</b>	dry-weather flow
<b>ECMWF</b>	European Centre of Medium Range Forecasts
<b>EPA</b>	U.S. Environmental Protection Agency
<b>FMI</b>	Finnish Meteorological Institute
<b>FWC</b>	free water capacity
<b>GTK</b>	Geological Survey of Finland
<b>GWI</b>	groundwater infiltration
<b>HD-DDS</b>	Hybrid Discrete Dynamically Dimensioned Search
<b>IA</b>	initial abstraction
<b>I/I</b>	infiltration and inflow
<b>LAM</b>	limited area model
<b>NLS</b>	Finnish National Land Survey
<b>NRW</b>	non-revenue water
<b>NWS</b>	U.S. National Weather Service
<b>RDII</b>	rainfall-derived infiltration and inflow
<b>RNM</b>	ratio of negative melt
<b>SCADA</b>	supervisory control and data access
<b>SCF</b>	snow catch factor
<b>SSN</b>	sanitary sewer network
<b>SWSN</b>	stormwater sewer network

<b>SSOAP</b>	sanitary sewer overflow analysis and planning
<b>SSO</b>	sanitary sewer overflow
<b>SYKE</b>	Finnish Environment Institute
<b>SWMM</b>	stormwater management model
<b>TMFG</b>	Traffic Management Finland Group
<b>UH</b>	unit hydrograph
<b>WWF</b>	wet-weather flow
<b>WWTP</b>	wastewater treatment plant
<b>XML</b>	extensible markup language

# 1. INTRODUCTION

## 1.1. Sewer networks

Sanitary sewer networks (SSNs) are urban infrastructure systems used to collect and transport wastewater from human developments, such as households, factories, hospitals, schools, commercial and business facilities to a disposal point, such as the wastewater treatment plant (WWTP). The collection sewer network is usually owned and operated by municipalities, water utilities, and authorities. [50]

Collection and treatment of wastewater is an important infrastructure of urban conglomerates because of its impacts for human health and environmental protection. Although the creation of the first sanitary sewer systems date back to ca. 4000 BC [18], as for the year 2019, an estimate of 26 % of the world's human population - 2,0 billion people - still do not have access to basic sanitation facilities. An estimate of 432.000 annual diarrhoeal deaths are caused by poor sanitation. The transmission of several other diseases, such as hepatitis A, typhoid and polio can also be associated with precarious sanitary conditions [52]. The amount of human population still defecating on the open environment, and using different types of sanitation system varies in different parts of the planet as depicted on in Figure 1.1.

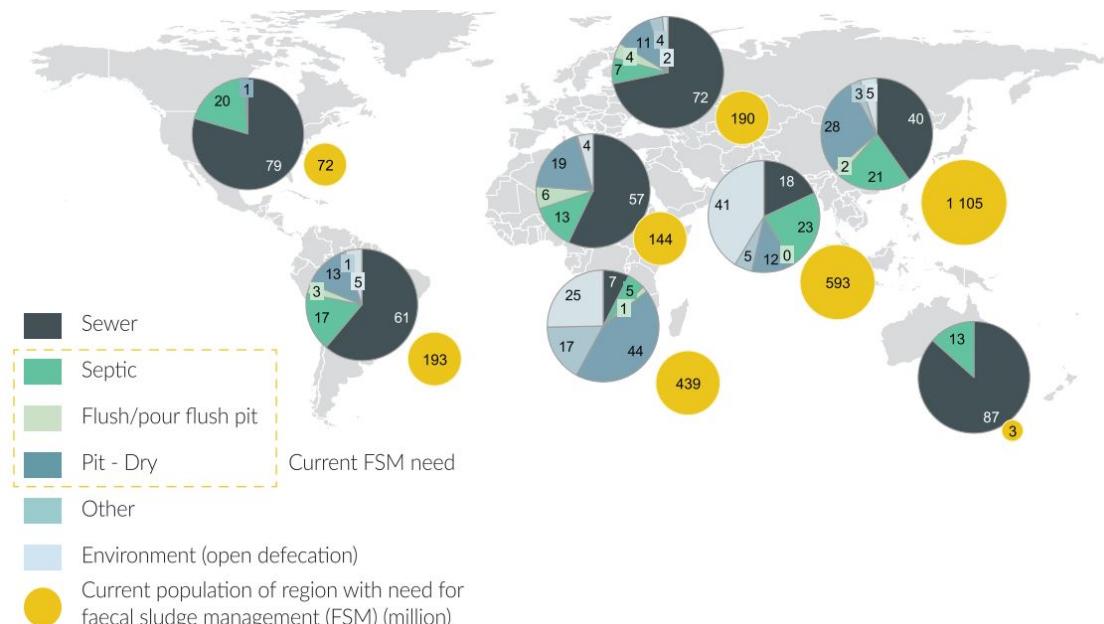


Figure 1.1.: Percentage of population served by different types of sanitation system. [48]

There is, therefore, the need to develop more SSN, specially in undeveloped and developing regions

to reduce the amount of contamination, deaths, and environmental degradation caused by open defecation. Creation of new sewer systems are, however, costly.

Managing the existent sewer system can be also a challenging task. These systems are, in some cases, combined diverting the domestic sewage and the excessive stormwater in urban areas. This systems are named as combined sewer network (CSN) and important assets for sanitation and urban drainage. The capacity of existing sewer systems is often the reason for costly rehabilitation and expansion projects. Predictions of increase on human migration from rural areas to urban areas also pose a challenge for the systems which were designed and built, in some cases, more than 100 years ago.

The planet's urban population rose from 750 million to 4,2 billion in the past 70 years. The world's population growth prediction in combination with the urbanization estimates that an additional 2,5 billion people will live in urban areas by 2050. Examples of large human agglomeration are mega cities, such as Tokyo, New Delhi, Shanghai, Mexico city, and São Paulo, all with more than 20 million inhabitants [47]. The rapid increase of urban population is sometimes translated proportionally to a necessity of a fast increase of the sewer network capacity, which requires from the network owners and operators a more frequent assessment of their system.

Sanitary sewer systems built several years ago were often designed to serve a much lower number of users. The aging of the infrastructure can also compromise its performance, because defects, such as pipe cracks, appear in the network over time. In SSN these defects increase the amount of inflow and infiltration from rainfall and snowmelt compromising its operation. combined sewer networks (CSNs) and stormwater sewer networks (SWSNs) are also impacted by urbanization as more impervious areas produce an increase on surface runoff quantity. Other factors, such as climate change can contribute to a capacity exceedance if the prediction of more frequent and more intense events, such as rainfall and snowmelt, becomes reality [26].

Thus, the list of the mentioned challenges of the SSN management, as for the date of this study are:

- Population and urbanization growth.
- Aging pipes and network components.
- Climate change.

Larger urban conglomerates requires an also large sewer network. This systems can extent to several kilometers under the soil surface with a large number of components. These components can be defined as static, such as pipes, joints, manholes, and dynamic components, such as pumping stations and control vales.

The challenges listed above represent changes and adaptations for the sewer systems. To understand the impacts caused by this changes, the assessment of different scenarios can be used. It is, however, complex engineering problem due to the size of the network usually with several static, and some, dynamic components influencing the flow characteristics.

The design of new sewer system, modification of existing networks, and assessment of different scenarios can be achieved by engineers and planners by using a dynamic model as a tool. A dynamic model can simulate the water quantity and quality of sewer networks by representing the physical components of the networks and their properties influencing the dynamics of the fluid. The water transport is simulated by governing equations derived from mass and momentum conservation laws.

The sewer model can be a powerful tool for network owners. It can be used to, relatively quickly, simulate the behavior of their network under different scenarios. This can assist the design of new parts of the network or act as a predictive tool targeting the reduction of problems, such as overflows caused by capacity exceedance. During the snowmelt, the temperature of the wastewater in the SSN may drop. This can impact the wastewater treatment process. Therefore, this process may be helped by knowing, in advance, when the snowmelt will occur.

## **1.2. Problem description and objective of the thesis**

This study proposed to address, at least part, of the solution for the following problem:

1. Inflow and infiltration of water from rainfall and snowmelt to sanitary sewer network (SSN) increases the wastewater flow rate, the cost of wastewater treatment, and the risk of SSOs. How can the RDII be modeled accounting for snowmelt and rainfall to improve the operations of the network with a tool for analysis and prediction?

This thesis had the following objectives:

1. Compare and analyze the performance of two different hydrological modeling approaches present in stormwater management model (SWMM) to simulate the rainfall-derived infiltration and inflows (RDIIIs).
2. Couple the hydrological models to the hydraulic model to simulate SSN flows on a real case study of a town in Finland.
3. Propose a methodology for parameter estimation and calibration of the hydrological models.
4. Use the built sewer model with forecasted precipitation data and identify key aspects of using the proposed model as part of an operational forecast system able to predict the SSN flows during all seasons.

Modelling the sanitary sewer network flows can help cities to understand and solve issues that impact their society, environment and economy. Factors influencing the behavior of the system have to be investigated before, and during, the development of a sanitary sewer model. As described by Vallabhaneni and Burgess [50], the flows in the network usually have two different behaviors that can be classified as:

1. Dry-weather flow (DWF).

## 2. Wet-weather flow (WWF).

Therefore, a successful forecast of the total flow in the sanitary sewer system is only achieved if the model is capable of simulating DWF and WWF.

Typically, DWF pattern describing its variation over time can be estimated by analyzing historical data from flow measurements available along the network or at the downstream end during dry periods [9]. Intuitively, more water is discharged into the sewer during day time than during nighttime. More complexity is added when trying to estimate the WWF in the sanitary sewer network. Flow increases in the network due to inflow and infiltration caused rainfall or snowmelt. This increase of inflow, which finds its way into the sanitary sewer network during and after a wet weather period, is classified as the rainfall-derived infiltration and inflow (RDII). Figure 1.2 depicts the representation of the infiltration of both, stormwater and snowmelt in the SSN during different seasons.

A sanitary sewer model aiming to simulate both DWF and WWF, can be further divided in two different models: hydrological and hydraulic model. The hydrological model in this context is used to transform and divert water coming from rainfall or snowmelt simulating the WWF. A hydrological model approximates physical process occurring in the nature, such as snow accumulation, infiltration, losses for evapotranspiration, etc. to divert the incoming water with a spatial and temporal distribution. The hydraulic model is the representation of the physical components of the sewer network which connects the domestic, commercial, industrial, etc, sewage to the disposal point.

The challenge on RDII estimations lies upon the different ways stormwater enters a sanitary sewer network (SSN) [31]. Stormwater can flow into the SSN directly through foundation and roof drain connections, leaky manhole covers, or stormwater drains. After reaching the soil surface and percolating, the stormwater infiltrates through defects in the network components, such as: damaged pipes, joints, manholes, etc, [38]. Defects of the pipe network are shown in Figure 1.3

Quantity of inflow and infiltration (I/I) increases proportionally with the intensity of the rainfall and snowmelt. Urban areas located in cold climate can have a significant amount of RDII because of snowmelt. Therefore, the snow accumulation and snowmelt process were considered during this study.

As already mentioned, rainfall-derived infiltration and inflow (RDII) is responsible for increasing the flow rate in sanitary sewer systems. This increase may cause sanitary sewer overflow (SSO) due to exceedance of the system capacity [38]. In some cities, sanitary sewer is combined with stormwater sewer network. The capacity of this coupled systems may also be exceeded during an event causing combined sewer overflow (CSO) [50]. When the capacity is exceeded, untreated water rejected by the wastewater treatment plants (WWTPs) is released into surface waters [14], such as rivers and streams. Upstream capacity-related issues may cause wastewater to find its way into basements or streets [38], such as depicted in Figure 1.4. Untreated water released on the surface water bodies or urban area increases the risk of human contamination by infectious diseases.

Although these are well known problems, they are still present in urban centers around the globe. Frequency of SSOs and CSOs are seasonal dependent. Their frequency may also increase, if the current prediction of more frequent intensive rainfall events is realized [14], enlarging the damage to the

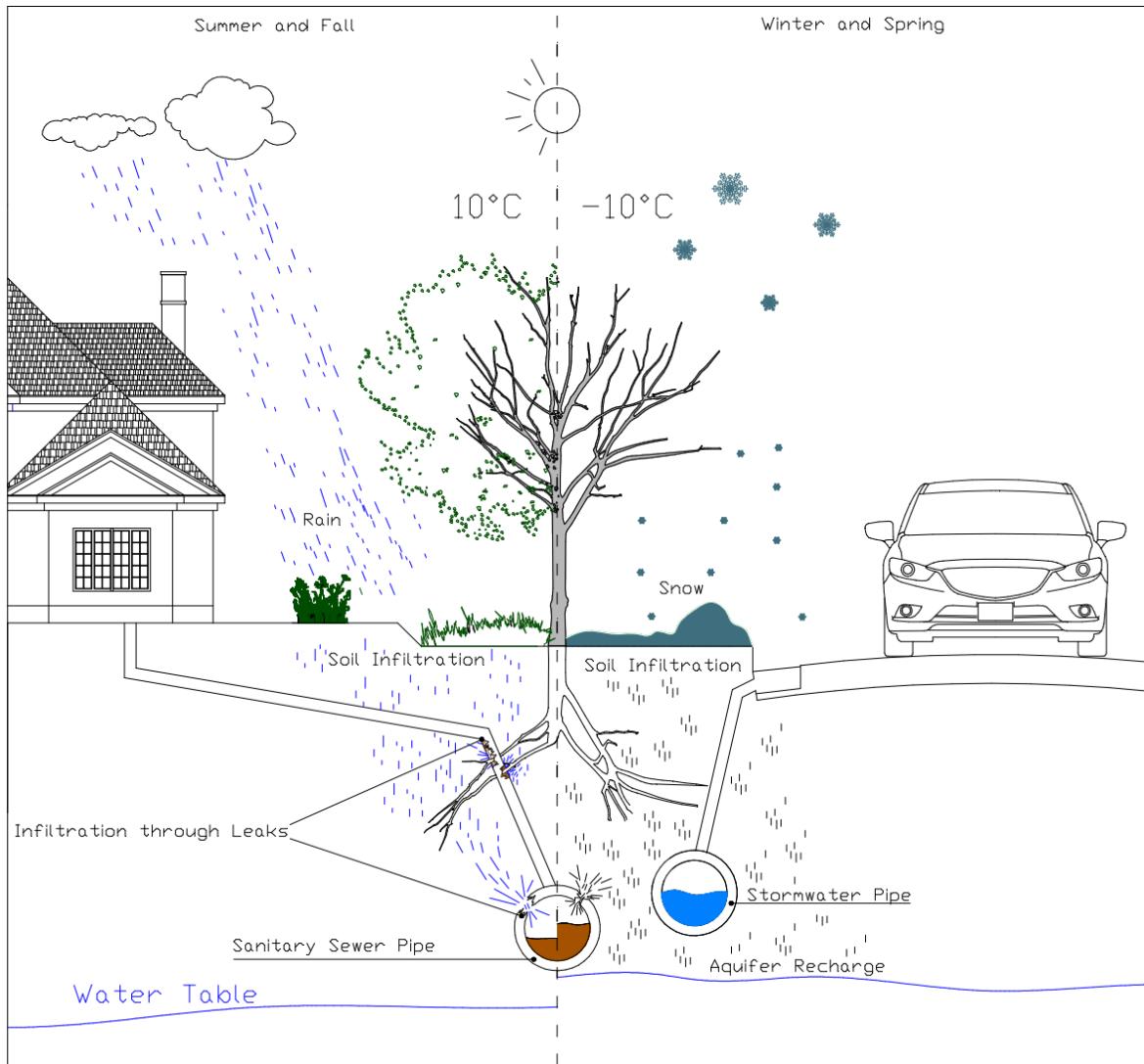


Figure 1.2.: Wet-weather Infiltration

human and environment health caused by the overflows. Moreover, wastewater overflows can cause conflicts in the society when streams or rivers are used both as an option to dispose wastewater and recreational area as described by Heikkinen et al. [24]. Giving the SSN owners, such as cities, municipalities and water utilities the ability to predict SSOs and CSOs was one of the prior motivations of this study as previously described. However, there are other benefits of modelling the complete flow in the SSNs, such as:

- Reproduction of past events for analysis.
- Assessment of the network behavior inclusion of new users, also under different scenarios.
- Estimate parts of the network more likely to be defective due to higher RDII flows.

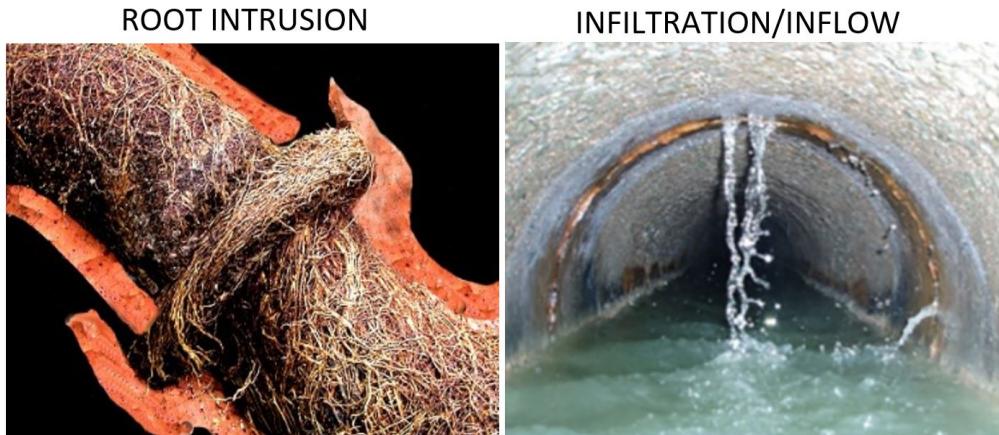


Figure 1.3.: Defects on the sewer network [13]



Figure 1.4.: Sewer overflows [4] [27]

RDII can increase the inflow of wastewater to WWTPs for weeks due to possible long response times [31]. Intuitively, the operational cost of the plant also raises since more wastewater needs to be treated. Results from simulations over the year might be able to identify increases over time on the flow pattern in specific pipes or sub-divisions of the network. This would be an indicator to carry further inspections and evaluate whether the infrastructure is damaged allowing infiltration.

### 1.3. Thesis structure

The study is divided in six chapters.

1. Chapter 1 presents an introduction to the topic of the thesis, a brief problem description and the thesis objectives.
2. Chapter 2 discuss the methodology proposed, more detailed questions and answers that helped guide the work presented in this study.

3. Chapter 3 provides an overview of alternative methods to be used for achieving the proposed objective and a literature review of the hydrological models chosen.
4. Chapter 4 presents the case study, overall characteristics of the catchment, and detailed description of the data used and model set up. Preliminary calibration of parts separated parts of the model are also available in this chapter as part of initial parameter sensitive analysis.
5. Chapter 5 contains the calibration and validation of the model as results of the work and discussion of the findings.
6. Chapter 6 presents the conclusions of the work and recommendations for future works.

## **2. METHODOLOGY**

### **2.1. Introduction**

This study investigates the main aspects behind the development of two hydrological models and proposes a methodology for the model set up. Even though this study does not propose a complete description of a real implementation of an integrated, operational and early warning system, it focuses the development and data acquisition for the hydrological models aiming a near to real-time operation. This purpose influenced decisions of methods to be used and data fetching routines. Thus, key aspects which governed the methodology presented can be classified as:

- Fetching of real-time forecast data;
- Automatic calibration and validation;
- The supervisory control and data access (SCADA) flows data treatment;
- The SSN flows decomposition;
- Forecast possible overflows, capacities.

There is an existent hydraulic model of the study site. Therefore, details about the existing model are briefly presented later on chapter 4. Focus is given here to the hydrological model set up. The methodology was divided, conceptually, in two parts:

1. Offline model set up.
2. Online model set up.

During the offline model set up the input data necessities are assessed. The parameter estimation is carried based on spatial data of the study site or literature review. The first simulations with previously estimated parameters are also part of the offline model set up. Therefore Online model part aims to define a methodology to assess the input data requirements. As an example, input data can be the meteorological forecast data. This part also aims to investigate the characteristics of the proposed hydrological models regarding the necessity of continuous calibration and parameter updates for near to real-time applications.

A simple distinction of the two steps, offline and online models can be, perhaps, better understood by the different use of the sewer model after the conclusion of each phase. In short, the offline sewer model can be used for design and analysis purpose whereas the online model can be used for a forecast and early warning system.

## 2.2. Offline Model

The sanitary sewer model can be defined by the combination of the description of the hydrological and hydraulic states of the sewershed. The methodology presented here focus on the creation of the hydrological model with the assumption that the hydraulic model was already existent. The decision of which type of hydrological model, that in this case is used to simulate the RDII, is defined as the first step for the offline model set up. Some alternative models for simulating RDII are presented later on chapter 3. Another assumptions taken when proposing the following methodology is that the historical recorded of the discharge data is available for the point of interest in the SSN.

Figure 2.1 depicts a proposed flow chart of the necessary steps taken in order to develop the offline model. The second step, as shown in the chart, is the evaluation of the available data, which in some cases, motivates a reconsideration of the hydrological model chosen on the first step. In case there is more than one type of a specific data input to support the creation of the chosen hydrological model, a decision for the most suitable must be made. As an example, sometimes there are two types of precipitation measurements, such as radar-based and rain gauge. Often, the data used in hydrological models require some amount of pretreatment, which converts the raw information given in the data to friendly formats for the chosen hydrological model. Sometimes an interactive process is required in case of problems when processing the selected input data and a different choice must be made.

In theory, the processing of the input data can be the only necessary step to estimate the model parameters. However, uncertainties related to the input data measurements, such as errors associated with the instrument, and approximations of hydrological process affect the accuracy of the data. Therefore, a range of possible values instead of a unique absolute value. A sensitivity analysis is proposed in this methodology to identify, and rank, the estimated parameters by their impact on the final simulated result through an evaluation criteria defined in agreement with the purpose of the model. The most sensitive parameters can then be chosen for the calibration process, which compares the simulation results with observed data and adjust the selected parameters, within the proposed range, to minimize the differences between simulated and observed values. After the calibration of the parameters, the model performance can be assessed by simulating another rainfall or snowmelt events.

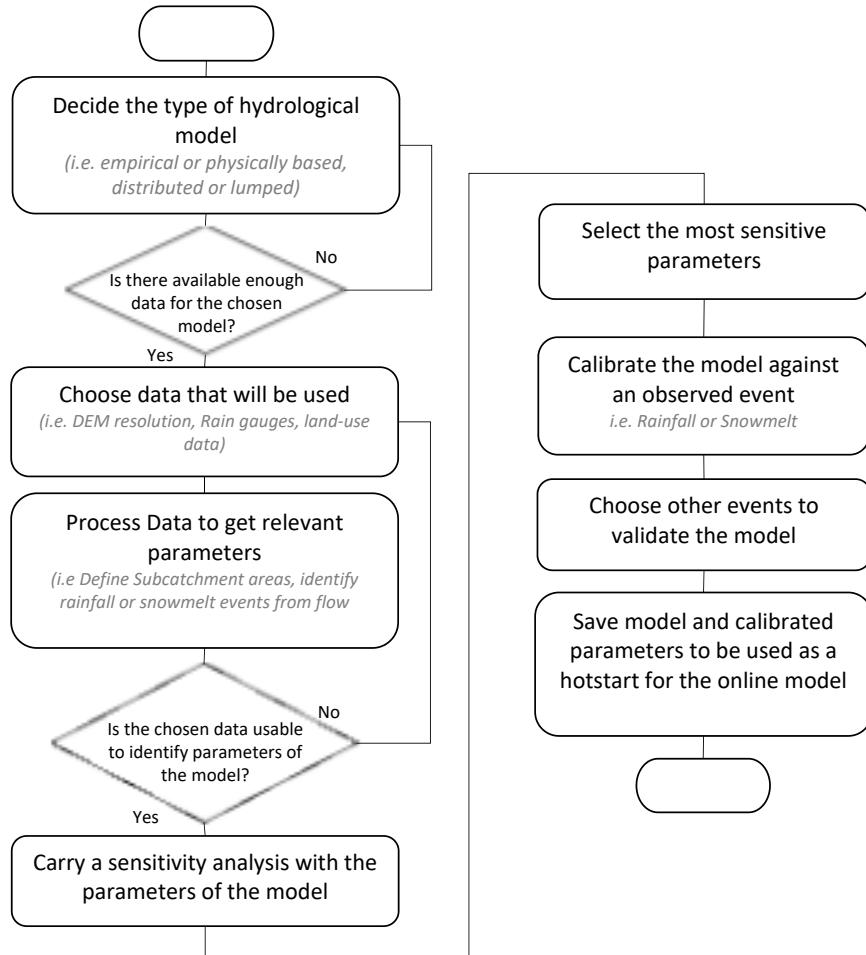


Figure 2.1.: Hydrological model development flowchart

### 2.3. Online Model

The online model can be characterized as an autonomous operational system able to communicate in near to real-time with different data providers, with latency measured in minutes. The framework focus on the collection and processing all necessary data for the sanitary sewer model simulations. The proposed methodology also considers a continuous calibration of the model parameters using an optimization algorithm.

Although several parts of the offline model set up can be automatized after choosing hydrological model, its reproducibility for other catchments is dependent on the available data used for parameter calibrations. Therefore, the offline model set up is posed as a prerequisite for the online model. Differently than the offline model, the online model can be considered as a closed loop system as depicted in the Figure 2.2.

Choi and Ball [11] defined two types of parameters for a hydrological process:

- Measured Parameters: parameters directly measured, such length of channels/pipes, catchment land-use, or recorded rainfall depth.
- Inferred Parameters: parameters not directly measured, and coefficients used by empirical models that approximate complex physical processes. Approximate characteristics of the system (i.e. imperviousness) and processes (i.e. flow coefficients, such as hydraulic conductivity or manning's).

The characterization of parameters is relevant to the calibration of the model, which is later discussed in this section.

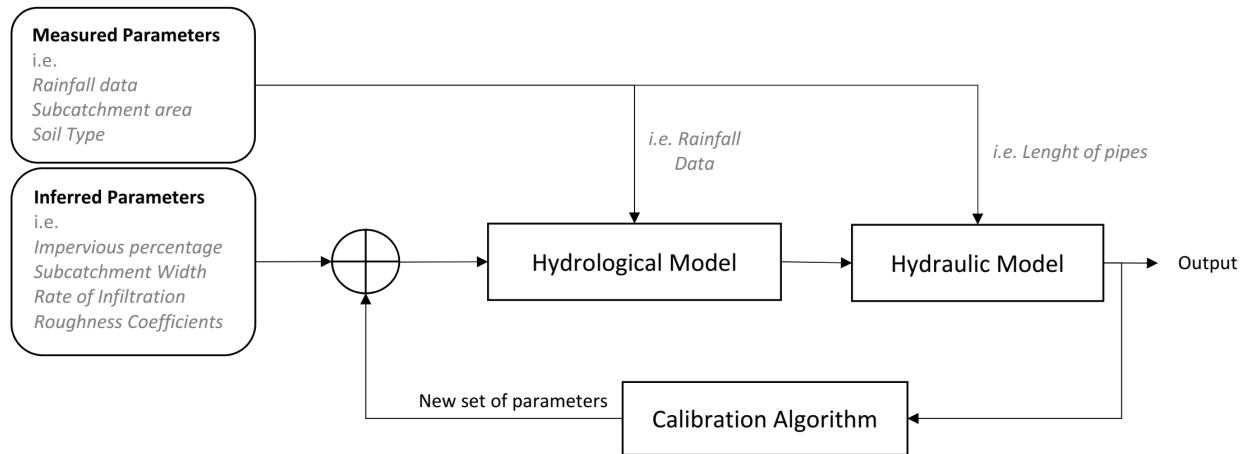


Figure 2.2.: System diagram an online forecasting model

Simple applications in python language [33] were developed during this study. They were used to automatize the SSN historical flow data treatment process and weather forecast data acquisition from the suppliers' application programming interfaces (APIs). These applications are freely available in the A. The flow data treatment application can be used for other case studies elsewhere. The weather forecast application can be used with three different weather models for Finland, with two different models for the north of Europe, and one of the models can be used to retrieve weather forecast data of the places within the European continent. More information about the weather forecast data is available on chapter 4.

Important points of the online simulation process are discussed on the following items:

- Frequency of simulations:

Dependant on many factors, such as the type of hydrological and hydraulic model chosen, calibration process, determined data acquisition routine, hardware used to process the information, etc.

- Relevant information obtained from the simulation results:

The online mode set up aims for an operational and early warning system. Therefore, information of the hydraulic states in the network before, during, and after the occurrence of a rainfall or snowmelt are relevant. Prior to the beginning of an event the forecast of the flow discharges can be considered important. This information concerns the SSN operations, such as the control of pumping stations. The peak discharge is also a critical information concerning the warning system for SSOs, or combined sewer overflows (CSOs) in case of a combined sewer. The period after the discharge peak when the flows in the SSN reduces towards customary levels remain relevant for operational purposes, for example, to schedule a repair in the network.

- Choice of parameters for calibration:

Inferred parameters can be considered the least known values since no measurements were carried. Therefore, they are the first candidates to be chosen for the calibration process. The sensitivity analysis carried during the offline model set up is used to identify which of the inferred parameters have greater impact to the results. The number of parameters necessary to simulate a sanitary sewer model and the range of possible values for each parameter can be a very large number. This increases the search space for the calibration algorithm. This may require many iterations to identify an optimal set of parameters. More than one calibration run is often necessary since calibration algorithms not always converge to a solution expected.

- Frequency of calibrations:

The number calibration simulation are dependant on the chosen model. It is first necessary to evaluate if the previously select parameter values are no longer valid. Seasonal weather conditions in cold climates modify the hydrological behavior of a sewershed. This can affect some parameters, such as parameters used for the infiltration module. Therefore, in such cases the calibration process can be used for adjustments on the parameters to better represent the actual state of the sewershed. A calibration run might happen in scale of hours, days, weeks, months, etc.

- Evaluation function:

Deciding if some parameter should be updated by the new value proposed process can be considered as an important part of the calibration process. The purpose of the model and the expected information should be defined and expressed in the evaluation function. This can be static or dynamic since the target may change on an event basis. As an example, It can be more relevant for the network operator to predict total volumes of events happening during winter season whereas peak discharge is prioritized during summer storms.

## 2.4. Parameter Optimization Algorithm

Models that approximate the hydrological behavior of catchments usually have a large number of parameters. These parameters are used for different processes of the model. The also large number

of combinations for the total set of parameters makes the calibration of catchment models a complex task. Optimization techniques are used to identify the optimum value of the objective function, often through a heuristic approach. However, there are often different sets of values for the parameters leading to similar optimum results [22]. Because of the mentioned complexity, the manual calibration of such models can be a laborious and tedious task.

Optimization algorithms mimics the process of manual calibration. According to Dent et al. [15] a calibration algorithm workflow can be defined as:

1. Change parameters of the model.
2. Run the model simulation.
3. Measure error of simulated vs observed.
4. Repeat previous steps and choose the most fit set of parameters.

For hydrologic models with multiple subcatchments and parameters, the number of possible combinations of parameters are of several orders of magnitude. This complexity motivates the adoption of optimization algorithms for automatic calibration.

To understand how rapidly the search space increases one can imagine a given hydrologic model with only one catchment and three different parameters as the area, slope, and roughness. Assuming that area is constant, but each of the other two can assume five other possible values. The possible combination of parameters to represent the catchment would be 165. Considering that the catchments is divided in two subcatchments. In this case, the possible number of combinations would raise to 74613.

*dynamically dimensioned search* (DDS) is a stochastic single-solution based heuristic global search algorithm that was created for the automatic calibration process of watershed simulation models. It was developed to find good global solution for a set of parameters faster than previously available search algorithms. DDS was proposed by Tolson and Shoemaker in 2007 [46].

The name *dynamically dimensioned* is related to the ability of the algorithm to scale the search based on user-specified maximum number of iterations. Global search approach is used for the first iterations. DDS switches to a local search approach by selecting and reducing the search space when the number of iterations nears the maximum allowed. The algorithm reduces the search space by strategic reduction of the number of parameters to be calibrated when it approximates to the end of the search. It also respects the constraints of each parameter given by the user. Therefore, it does not choose values for a parameter out of the specified range.

Other relevant aspects of an automatic calibration and DDS algorithm are:

- Automatic calibration can avoid modelers subjective judgments on parameter estimation difficult to be reproduced or explained. Sometimes decisions are made based on previous calibration experiences. It can accelerate the process of calibration as a multiple objectives evaluation can be included in the process, such as comparison of peak flows, hydrograph shape, or total volume [15];

- DDS was created for computationally expensive calibrations [7]. Therefore, it is suitable for infiltration and inflow models where a possible large number of parameters should be simultaneously optimized;
- DDS typically converges finding a good solution for a set of parameters and successfully avoid poor local solutions [46];
- For models with subdivisions of the space domain, such as SWMM [28], comparisons available in the literature have proven that DDS is one of the fastest to converge and the best finding good solutions. In other words, DDS does outperform other algorithms for complex models[46, 51, 7];

Refer to Sunela [40] for real case application using the DDS algorithm.

### **3. SEWER MODEL**

The wastewater flow in the urban sanitary sewer networks (SSNs) can be decomposed as dry-weather flow (DWF) and wet-weather flow (WWF). The DWF can be described as the customary observed flow containing wastewater from human developments when no rainfall or snowmelt occurs. The DWF typically follows a cyclic pattern of usage similar to the freshwater consumption. The wet-weather flow (WWF) represents the SSN flows during and after a rainfall or snowmelt event occurs. dry-weather flow (DWF) can be further divided in two components:

1. Base Waste Flow (BWF): inflow of waste water coming from households, commercial and industrial sites.
2. Groundwater Infiltration (GWI): Water from aquifers that infiltrates into the network thought defects, such as pipe cracks and leaky joints [50].

The choice of the hydrological model in this study aims the representation of RDII, which is the incremental flow into the sanitary sewer system caused by precipitation (rainfall or snowmelt). Figure 3.1 shows the typical characteristics of different components of sanitary sewer flow. RDII needs to be first separated from DWF when processing raw data coming from flow meters. More about the methods to separate the components are discussed in chapter 4.

As mentioned in chapter 1. There are different ways stormwater or snowmelt finds its way into the sanitary sewer pipes, which ideally would have only wastewater from urban developments, such as households, commercial centers, factories, etc. The flow increase in the SSN can be triggered by events, such as storm, snowmelt or increase of soil moisture content. From rainfall or snowmelt water flows over the soil surface and into the sanitary sewer through manhole leaky covers or directly from roof-drain and foundation connections. This extra amount added to the wastewater is generally observed few hours after the beginning of the storm up to days after a period with intense snowmelt. [38]. The flow above the customary values observed as a long-term effect (days or weeks) can be explained by the features of subsurface flow. Once the water infiltrates, it moves through the pores in the soil with a much slower velocity before entering the network system through its defects.

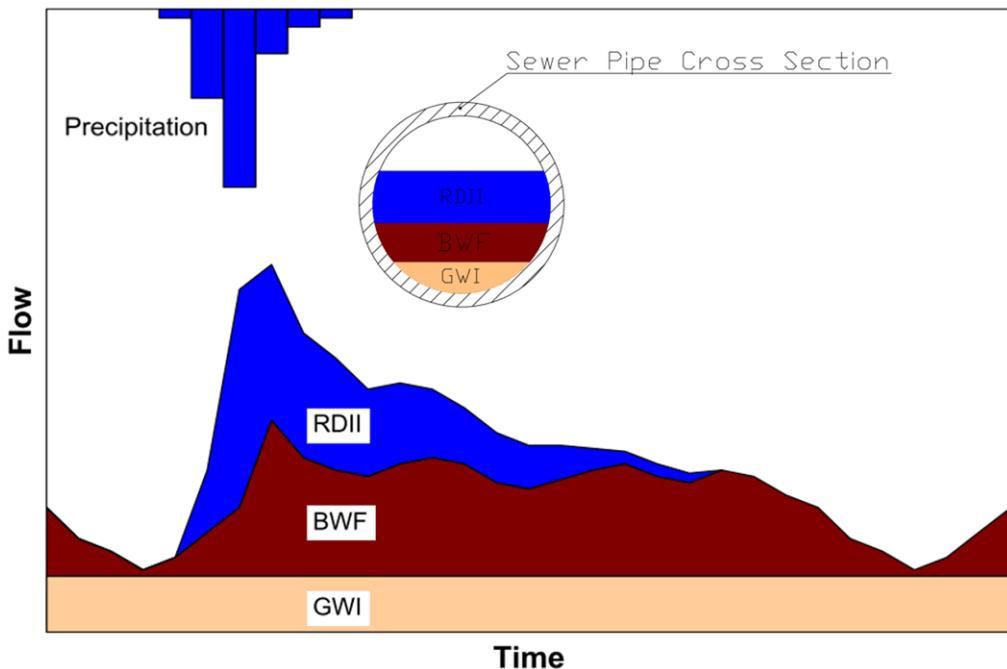


Figure 3.1.: Wet-weather flow components. Modified from [50]

### 3.1. Alternative methods to model rainfall dependent infiltration and inflow

Rainfall dependent infiltration and inflow have been modeled with different methods. Bennett et al. [9][9] carried a literature review and case study in 1999 of around 10 different methods for quantifying RDII. The study concluded that only the regression and unit hydrograph methods are suitable when applying continuous simulation for long-term modelling. However, a physics-based method was not assessed. The unit hydrograph (UH) method also provided the best consistent match to storm peaks among the benchmarked methods [50].

Vallabhaneni and Burgess [50] and [16] also considered sewer network rehabilitation capabilities as a factor for evaluation of the methods, and suggested that regression should only be used when at least more than 2 years of recorded flow and rainfall data is available. When no flow measurements are available, the constant unit rate method [9] seems to be useful since it accounts for spatial characteristics (topographical data) of the sewershed, information of pipe characteristics and population. Moreover, the study concluded that the unit hydrograph RTK method can be useful to estimate the amount of the wet-weather flow is relative to the direct inflow and the amount relative to infiltration. Knowing whether RDII is more impacted by inflow or infiltration is relevant when evaluating the sanitary sewer network for rehabilitation.

It is important to mention that the studies also concluded that there is no RDII quantification method that can be universally applied, since their use depend on available data and characteristics of the catchment. The goal of Vallabhaneni and Burgess [50] and EPA [16] reviews were to choose

the most suitable method to be first implemented in a toolbox named as Sanitary Sewer Overflow Analysis and Planning (SSOAP) that is later discussed.

Physics-based models have also been used to model sanitary sewer flows. Robinson [36] modelled groundwater infiltration into SSN using SWMM5 and two aquifer approach to investigate rehabilitation method of urban catchment in Seattle, USA. Tavakol-Davani et al. [42] used SWMM5 rainfall-runoff module to study the impacts of climate change to combined sewer overflows (CSOs) in Toledo, Ohio, USA. Choi [12] PhD dissertation compared the use of three physics-based models: A roof downspout; Sump pump, and leaky lateral models with RTK unit hydrograph method. Even though there are studies available in the literature using physics-based models applied to sanitary sewer flows, it seems to be easier to find cases where the process are analyzed separately with few including also snowmelt models to the processes.

Therefore, a comparison of a physics-based model and empirical model as hydrological inputs to the sanitary sewer hydraulic model was proposed in this study. SWMM 5 was chosen and the two available routines (SWMM physics-based modules and RTK unit hydrographs) are used to simulate the SSN flows in the study area.

SWMM 5 was chosen for this study for three main reasons:

1. It includes snowpack and snowmelt, infiltration, runoff and aquifer models, as well as the RTK unit hydrograph.
2. Free open source.
3. Well documented user manuals and several published case studies.
4. Hydraulic model of the study area was already built in SWMM.

The approach chosen focuses on the flows in the sanitary sewer network (SSN) without further information of the stormwater sewer network (SWSN) or any stormwater harvesting system. These artificial units are treated as losses as depicted in Figure 3.2.

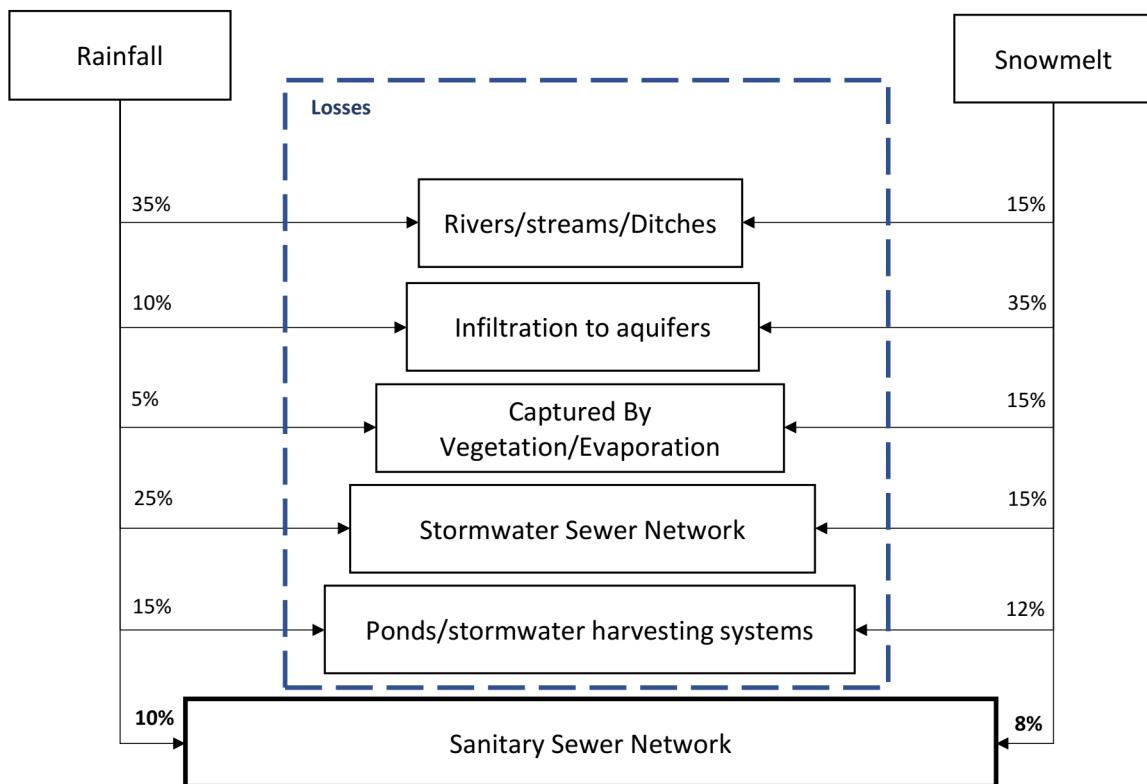


Figure 3.2.: Precipitation Losses relative to a Sanitary Sewer Network

## 3.2. Physics-based hydrological model

The use of stormwater management model (SWMM) in this study aimed to model four processes happening simultaneously in the watershed to simulate fast, medium and long-term response observed in SSN wet-weather flows. The four processes/SWMM modules are described here as: Runoff; Snowpack and Snowmelt; Infiltration; and Aquifer/Groundwater. A summary of the four modules is presented on the following sections based on Rossman and Huber [38].

### 3.2.1. Rainfall-Runoff

The area in SWMM is discretized by subcatchments. The size of each subcatchment is dependent on the purpose of the model. More on subcatchment delineation is discussed in Chapter 4. The rainfall-runoff is computed in SWMM for each one of the subcatchments using a non-linear reservoir model as depicted in Figure 3.3 and Equation 3.1 [38].

$$\frac{\partial d}{\partial t} = i - e - f - q \quad (3.1)$$

Where:

$i$  = precipitation rate [m/s]

$e$  = surface evaporation rate [m/s]

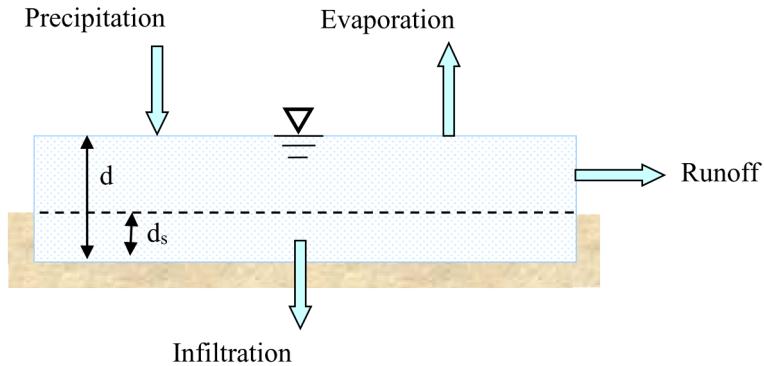


Figure 3.3.: Non-linear reservoir model [38]

$f$  = infiltration rate [m/s]

$q$  = runoff rate [m/s]

Runoff happens when water exceeds the depression storage ( $D_s$ ) and the overland flow is assumed as uniform in a rectangular channel expressed by Gauckler–Manning–Strickler formula (3.2). Each subcatchment in SWMM can be divided in three portions: 1. Pervious; 2. Impervious; 3. Impervious without  $D_s$ .

$$q = \frac{1.49 \cdot W \cdot S^{1/2}}{A \cdot n} \cdot (d - d_s)^{5/3} \quad (3.2)$$

Where:

$A$  = area [ $\text{m}^2$ ]

$W$  = Flow width [m]

$S$  = Slope [1]

$n$  = Manning's roughness coefficient [ $\text{s}/\text{m}^{1/3}$ ]

Runoff can be divided and routed to three different areas: 1. Outlet (node within the pipe network); 2. Pervious or impervious portion of the subcatchment; and 3. Other subcatchment. The modeller can input a percent of runoff routed ( $\%_{routed}$ ) as a parameter for SWMM model.

No information of stormwater sewer network (SWSN) was assessed in this study. Therefore, the amount of stormwater that finds its way into the SWSN is treated as a loss (see Figure 3.2). It was also assumed that the fast response on sanitary sewer wet-weather hydrograph follows the same pattern as surface runoff. However, with a reduction in its volume. One can imagine that the hydrograph coming from a subcatchment entering the SWSN will have the same shape as the "short-term hydrograph" entering the sanitary sewer network (SSN), but with greater volume.

The volume of water from precipitation lost to the SWSN can be represented by two existent parameters in SWMM model: 1. depression storage ( $D_s$ ); and/or 2. percent routed ( $\%_{routed}$ ). Therefore, the values chosen for these two parameters in this study may differ greatly from other SWMM models

focused more on modeling SWSN. Figure 3.4 represents the conceptual difference considered in this study in comparison with the original non-linear reservoir model.

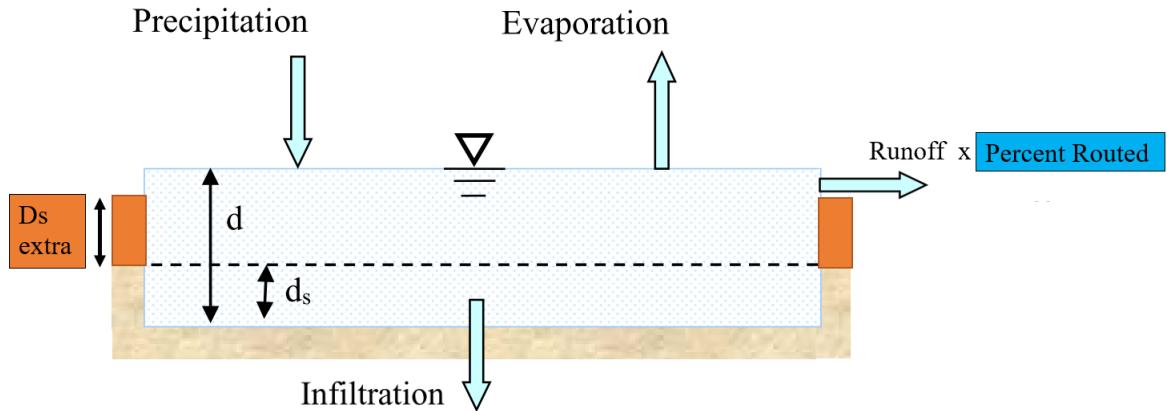


Figure 3.4.: Extra  $D_s$  and  $\%_{routed}$  for Non-linear reservoir model. Modified from Rossman and Huber [38]

$D_s$  accounts for the amount of water "absorbed" by the watershed from precipitation before runoff occurs. Wetting and ponding of the surface, and interception are usually the losses modeled by this parameter. The  $D_s$  extra and  $\%_{routed}$  represented in Figure 3.4 are the increment value for  $D_s$  and  $\%_{routed}$  to represent losses to SWSN system. The parameter estimation is discussed later in Chapter 4.

### 3.2.2. Snowpack and snowmelt

Snowpack and snowmelt module was used to simulate the variations of flows in the sanitary sewer network (SSN) occurring during winter conditions since a considerable quantity of infiltration occurs during snowmelt periods.

Snowpack and snowmelt calculation routines available in SWMM were based on models developed by U.S. National Weather Service (NWS) [6, 5]. SWMM models the depth of water equivalent as the snowpack. The depth is increased during snow accumulation periods and decreased, when snowmelt occurs. The amount of water released from the snowpack during snowmelt is transformed in precipitation rate [mm/h] and summed to the rainfall as "net precipitation" that is used as input to compute surface runoff. Therefore, snowmelt calculations are part of runoff module [38].

Three of the key parameters for the snowmelt routine are:

1.  $T_a$ : air temperature of the current time step [ $^{\circ}\text{C}$ ].
2.  $T_{base}$ : The base temperature of which snowmelt starts to occur [ $^{\circ}\text{C}$ ].
3.  $DHM$ : melt coefficient [mm/h- $^{\circ}\text{C}$ ].

These three parameters are used in the linear type Equation 3.3 to compute the snowmelt [mm/h] during dry periods. Calculations of snowmelt during wet periods (rainfall greater than 0.51 mm/h) take also in consideration the wind speed and local atmospheric pressure. Refer to Anderson [6, 5] or Rossman and Huber [38] for detailed description of snowmelt calculations during rainfall.

$$SMELT = DHM \cdot (T_a - T_{base}) \quad (3.3)$$

Melt coefficient ( $DHM$ ) varies seasonally and is calculated based on a sinusoidal equation and two user-supplied constants: Minimum melt coefficient ( $DHM_{min}$ ) which occurs on December 21th; and Maximum melt ( $DHM_{max}$ ) occurring on June 21th as depicted in Figure 3.5.

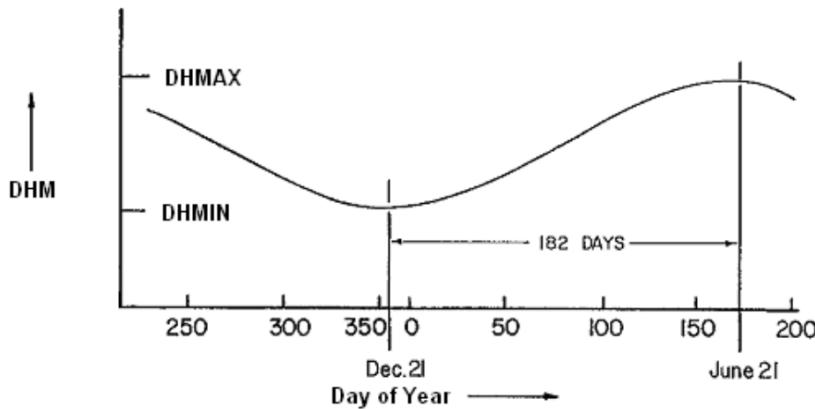


Figure 3.5.: Seasonal variation of melt coefficients [38]

Before snowmelt occurs, the snowpack status has to be assessed. For this, there are two conditions:

1. The snowpack has to be heated with air temperatures higher than  $T_{base}$ .
2. Snowmelt has to fill the voids within the snowpack. Meaning that there is a quantity of water contained in the snowpack and it is considered to be a fraction of the "depth of water equivalent" and named as fraction of free water capacity (FWC).

The heat content in the pack is calculated and FWC a user-supplied value. Therefore, liquid melt will only become a component of "net precipitation" after the two conditions, mentioned above are satisfied.

The difference between heat content of the snowpack and  $T_{base}$  is named as "cold content" ( $COLDC$ ). This variable is used to compute how much heat is necessary to be transferred to the snowpack before snowmelt can occur, as the first condition mentioned above. The  $COLDC$  value is updated every time step based on the heat transfer between the pack and the atmosphere. The variation of the cold content ( $\Delta CC$ ) is calculated every time step assuming a negative value during melting periods. The following two user-supplied constant fractions are necessary to compute  $\Delta CC$ : the ratio of negative melt (RNM); and the antecedent temperature index (ATI) weight ratio ( $TIPM$ ).

The rate of which heat transfer occurs is calculated based on SWMM's internal parameter of antecedent temperature index (*ATI*) which is function of  $T_a$  and *TIPM*. Values of *TIPM* towards tending to zero represent a thicker pack which warms and cools slowly as a greater weight is given to more antecedent temperatures. Equation 3.4 is used when  $T_a < T_{base}$  and equation 3.5 when  $T_a > T_{base}$  [38].

$$\Delta CC = RNM \cdot DHM \cdot (ATI - T_a) \cdot TimeStep \quad (3.4)$$

$$\Delta CC = -SNOWMELT \cdot RNM \cdot TimeStep \quad (3.5)$$

The ratio of negative melt (RNM) in equations 3.4 and 3.5 is used to account for a reduced heat transfer during periods without "actual liquid melt". Snow plowing and areal depletion were not used in this study for lack of data and simplicity. Other four parameters used were:

- $U$ : Monthly average wind speed [m/s]
- $Z_{el}$ : elevation above mean sea level [m]
- $T_{div}$ : Dividing temperature between snowfall and rainfall [ $^{\circ}\text{C}$ ]
- $SCF$ : Snow catch factor [ratio]

Where parameters  $U$  and  $Z_{el}$  were used to compute the influence of wind speed on the melting of snow during rainfall periods and parameters  $T_{div}$  and  $SCF$  used to define the amount of snowfall from raw precipitation input data.

Table 3.1 depict all parameters used for the snowpack and snowmelt module in this study and their proposed range based on other study cases available in the literature.

Table 3.1.: Snowpack and snowmelt parameters range [38]

Parameter	Proposed Range	Units
$T_{div}$	0–2	[ $^{\circ}\text{C}$ ]
$SCF$	1–2	[1]
$T_{base}$	-4–0	[ $^{\circ}\text{C}$ ]
$DHM_{min - max}$	0.019–0.11	[mm/h- $^{\circ}\text{C}$ ]
$RNM$	0–1	[1]
$FWFRAC$	0.01–0.25	[1]
$TIPM$	0–1	[1]
$T_a; Z_{el}; U$	Location Based (see Chapter 4)	

RNM and TIPM bare the full possible range. However, suggestions available in the literature were used as initial values in this study. All other ranges of parameters in Table 3.1, except by

$DHM_{min-max}$ , were proposed as suggested by Rossman and Huber [38] in the SWMM hydrology reference manual.

Tikkanen [43] suggested values for the degree-hour melt coefficients ( $DHM_{min-max}$ ) when modeling a catchment in Finland based on values calibrated by Valeo and Ho [49] for a catchment in Calgary, Canada. Tikkanen used reduced values of the melt coefficients, in comparison to Valeo and Ho. The purpose of the reduction was to account for fewer solar radiation due to difference in latitude. Valeo and Ho calibrated different values of  $DHM_{min-max}$  for snow covered pervious and impervious areas, varying from 0,02 for  $DHM_{min}$  to 0,15  $DHM_{max}$ . Therefore, the proposed range in this study was based on Tikkanen and Valeo and Ho findings.

### 3.2.3. Infiltration

An infiltration model was used in this study to assess long-term simulations (up to 6 months) of the winter periods using snowmelt routine. As groundwater infiltration (GWI) is one of the components of sanitary sewer network (SSN) flows, an aquifer and groundwater inflow models were included. The gradient of groundwater infiltration to the SSN is dependent on the water table elevation. Therefore, the infiltration routine was included as a way to recharge the modeled aquifer by varying the saturated zone elevation (water table) providing a connection between effective precipitation and the GWI component.

SWMM version 5.1 offers the modeller five different infiltration models.

1. Horton infiltration.
2. Modified Horton.
3. Green-Ampt.
4. Modified Green-Ampt.
5. Curve number.

The Modified Horton method [2, 3] was chosen among the options for three main reasons:

1. It is simply one of the default methods available in SWMM.
2. It has the same parameters as the well known Horton method which parameter estimates are suggested in SWMM hydrology manual. [38].
3. Appears to be more accurate for low intensity rainfall events than the original Horton method [38].

The two governing equations of the method describes the infiltration capacity decay during wet periods: the Equation 3.6 and its recovery curve during dry periods, Equation 3.8. Figure 3.6 depicts an example of these two curves and how the infiltration capacity would change over time.

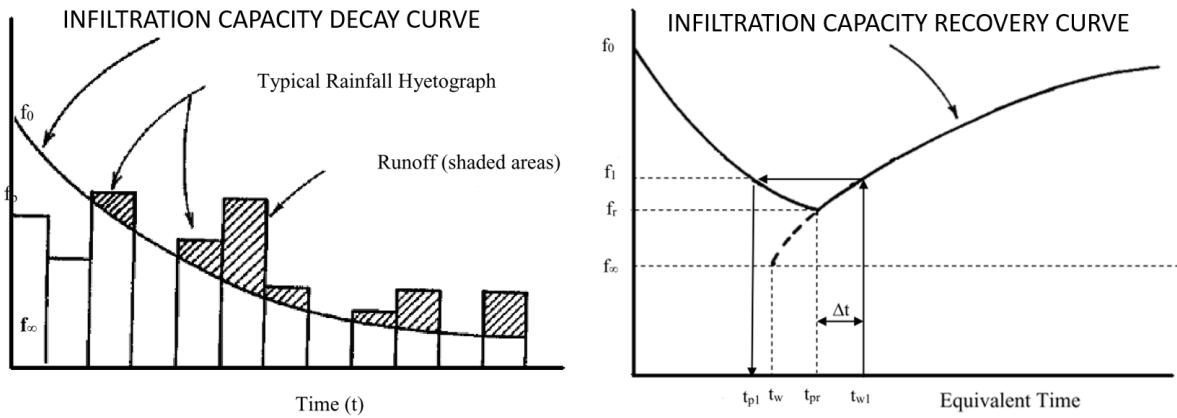


Figure 3.6.: Horton infiltration capacity decay and recovery curves. Modified from [38]

$$F = f_{\infty} \cdot t + \frac{(f_0 - f_{\infty})}{k_d} \cdot (1 - e^{-k_d \cdot t}) \quad (3.6)$$

Where:

$F$  = cumulative infiltration capacity [ft]

$f_{\infty}$  = minimum or equilibrium value of infiltration capacity at  $t = \infty$  [ft/sec]

$f_0$  = maximum or initial value of infiltration capacity at  $t = 0$  [ft/sec]

$t$  = equivalent time [sec]

$k_d$  = decay coefficient [ $\text{sec}^{-1}$ ]

It is important to mention that Equation 3.6 is an integrated form of Horton's original equation. SWMM uses integrated form to consider the intensity of the rainfall event also as a function of the infiltration capacity reduction [38].

$$\frac{df_r}{dt} = kr \cdot (f_0 - f_r) \quad (3.7)$$

Where:

$f_r$  = infiltration capacity during recovery [ft]

$f_{r0}$  = maximum or initial value of infiltration capacity at  $t = 0$  [ft/sec]

$k_r$  = regeneration coefficient [1/sec]

$t$  = time [sec]

The infiltration capacity at time  $t$  after integrating 3.7 when infiltration capacity is  $f_{r0}$  is:

$$f_r = f_0 - (f_0 - f_{r0}) \cdot e^{-k_d \cdot t} \quad (3.8)$$

SWMM computational scheme first checks for wet-period (rainfall/snowmelt) or dry period to ap-

ply either of the equations 3.6 or 3.8, and compute the current infiltration capacity and the amount of water infiltrating the soil. More details of the equations and computational scheme are available in [38].

Table 3.2 presents a rough estimate of the range of four input parameters for Horton infiltration model. The range was extracted from EPA SWMM user help.

Table 3.2.: Modified Horton infiltration parameters range[38]

Parameter	Typical Range	Units
Maximum infiltration rate	8,50– 254	mm/h
Minimum infiltration rate	0,254–120	mm/h
Decay coefficient	2–7	h <sup>-1</sup>
Drying Time	2–14	days

### 3.2.4. Aquifer and groundwater flow

There are medium and long-term wet-weather infiltration observed in sanitary sewer network (SSN) measured flow data. This infiltration raises the flow above its average for days or even weeks as discussed in the previous sections. It is challenging or not possible to model the short, medium and long-term hydrographs using only SWMM's runoff module since its parameters, such as roughness and slope would be distorted as an attempt to reproduce the delayed flows. Moreover, a short-term response can occur at the same period as the medium and long-term responses. One can imagine a high intensity rainfall happening right after a snowmelt period. The delayed portion of snowmelt infiltrates and slowly recharges the aquifer and discharges into the SSN, while rainfall causes runoff being fully discharged in minutes or hours. The Aquifer and Groundwater Flow module of SWMM was implemented together with the aforementioned modules as an attempt to represent the most important hydrological processes happening in the sewershed. Aquifer in SWMM is represented as a two-zones model containing an unsaturated and a saturated zone as depicted in Figure 3.7.

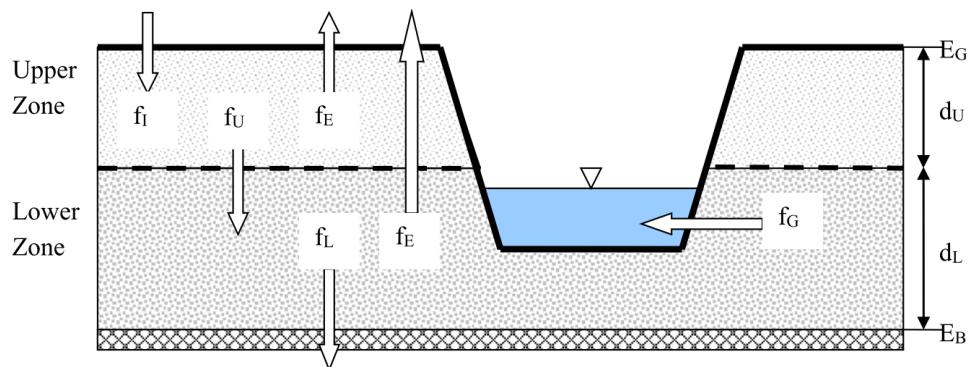


Figure 3.7.: Representation of two-zone groundwater model in SWMM [38]

The first zone is intermediary located between the soil surface and the groundwater table. Six fluxes among soil surface and the two aquifer zones can be computed every time step varying the elevation of the groundwater table and, therefore, changing the size of the aquifer zones. The variation of the saturated zone elevation ( $d_L$ ) affects the flow from the aquifer to the receiving node in the pipe network system ( $f_G$ ).

Modellers can use a customized equation to describe  $f_G$  flux using available parameters from the two-zone model. SWMM's hydrology reference manual [38] mentions options, such as Linear Reservoir, Dupuit-Forcheimer Lateral Seepage, and Hooghoudt's Tile Drainage. However, it is important to remember that groundwater table elevation is constant along the subcatchment limiting the representation of the pressure gradient between the saturated zone and the receiving node to the difference between  $d_L$  and the elevation of water surface in the receiving node. Note that these two elevations can vary every time step and govern changes in  $f_G$ .

### 3.3. Unit hydrograph hydrological model

Flows are higher than average into the sanitary or combined sewer network during and after a storm. This extra quantity of stormwater flows into the network through the drainage system configuration, such as roof drain connections, foundation connections, and sump pump or through defects existent in the network, such as leaky manhole covers, leaky joints, and pipe cracks [38]. Different pipe configurations and defects in the network causes also distinct volumes and rate of infiltration entering the sanitary collection system. As an example, stormwater flows into the collection system shortly after falling over the roof of a residence in case a roof drain is directly connected to the sanitary sewer network. The impervious and steep characteristics of a roof surface diverts the stormwater before infiltration or any other loss occur. Less intuitively, flows and depths into the sanitary sewer network remain higher than the average for hours, and sometimes, even weeks after the rainfall [31]. This long-term response is usually caused when stormwater first percolating the permeable soil before entering the sanitary network system through the defects aforementioned [38]. In summary, wet-weather flows into sanitary sewer system is impacted by different phenomena which affect the time and volume of stormwater infiltration.

To simulate the inflow and infiltration caused by a rainfall event, SWMM incorporates a synthetic unit hydrograph method, also called as RTK unit hydrograph because of its main parameters:

- $R$  = fraction of precipitation that enters the network (area below the hydrograph) [fraction]
- $T$  = time for the hydrograph peak [h]
- $K$  = parameter to define recession time [h]

The method was created specifically to simulate the portion of the wastewater flow affected by rainfall, namely the rainfall-derived infiltration and inflow (RDII). The RTK unit hydrograph method

uses a combination of three unit triangular hydrographs to account for the short, medium and long-term effects of RDII. The three unit hydrographs are then summed using a convolution process to create a final resultant RDII hydrograph that replicates the overall response of the system to a pulse of the rainfall event as shown in Figure 3.8. The resultant simulated RDII is then compared to the RDII hydrograph derived from real measured flows from a point in the sanitary sewer system. The best fit to the measured RDII is achieved through parameter calibration. Each of three unit hydrograph has three required parameters ( $R$ ,  $T$ , and  $K$ ) and may have other three optional parameters to represent the initial abstraction [50]. The number of parameters for each unit hydrograph ranges from 3 to 6. The range is multiplied by 3 for the resulting RDII hydrograph when all three unit hydrographs are combined. The total number of required parameters is 9 and increased to 18 when optional parameters are considered.

The optional initial abstraction parameters represent the rainfall amount necessary for RDII to occur. Therefore, these parameters account for initial rainfall loss and represent, for instance, precipitation amount retained by rainfall harvesting systems and vegetation as well as the soil antecedent moisture condition [38]. The three parameters that govern the capacity of initial abstraction are:

- $D_{max}$  = maximum depth of precipitation amount available [mm]
- $D_{rec}$  = recovery rate [mm/day]
- $D_0$  = amount of used at the start of the simulation [mm]

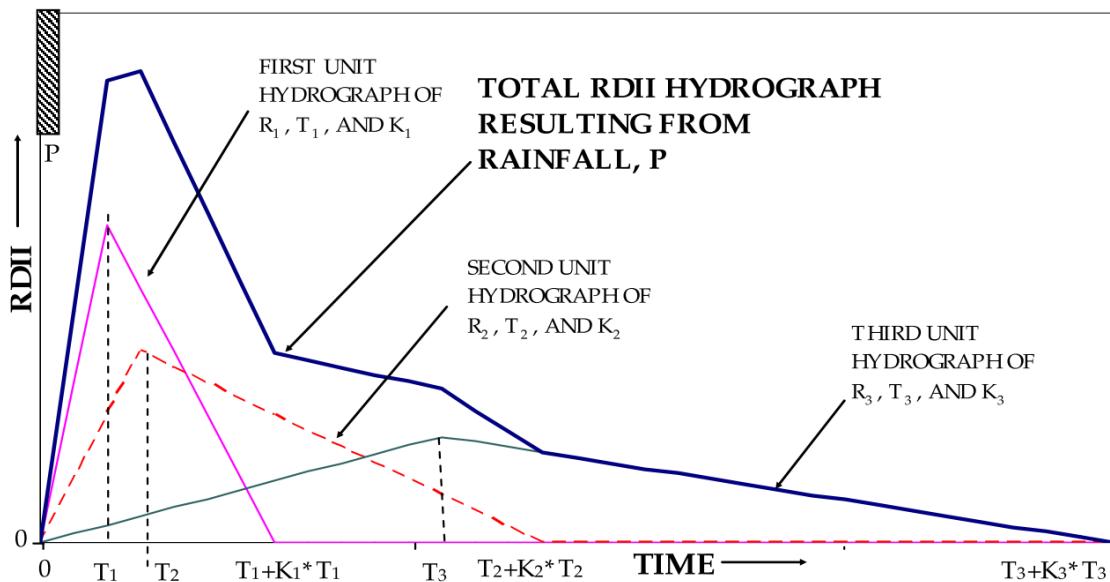


Figure 3.8.: RTK method components for short, medium, long-term and resulting hydrographs [50]

One can imagine the initial abstraction as a linear reservoir or bucket model standing between the precipitation input data and the unit hydrograph model. Precipitation data is fills the bucket and

reaches the unit hydrograph only the the bucket is full. The height of the bucket is analog to  $D_{max}$ , the amount of water available in the bucket before the precipitation starts is analog to  $D_{rec}$  whereas  $D_0$  can be imagined as the rate of seepage of a hole in the bottom of the bucket reducing the amount of water in the bucket and recovering its capacity.

As the RTK required and optional parameters are used to simplify hydrological process of the sewershed area, they are also prone to seasonal variations that affect these processes. For instance, the volume of infiltration into the pipe network can be greater during winter and snowmelt periods than summer and fall. Vallabhaneni and Burgess [50] described a case study that used RTK unit hydrograph model for sanitary sewer overflow (SSO) planning and analysis of a sanitary sewer network in Ann Arbor city located in southeast Michigan, USA. The location of Ann Arbor is also relevant for this study due to the snow accumulation of winter seasons (roughly 45 cm). It was observed that R-values and initial abstraction parameters presented a seasonal variation as depicted in Figure 3.9. On one hand, the dormant season (November to May) had a greater volume of rainfall infiltrating the sewer. On the other hand,  $D_{max}$  ( $V_o$  in the Figure 3.9) was greater during the growth season (April to October) when vegetation grows and reduces the amount of rainfall available to infiltrate the pipe network.

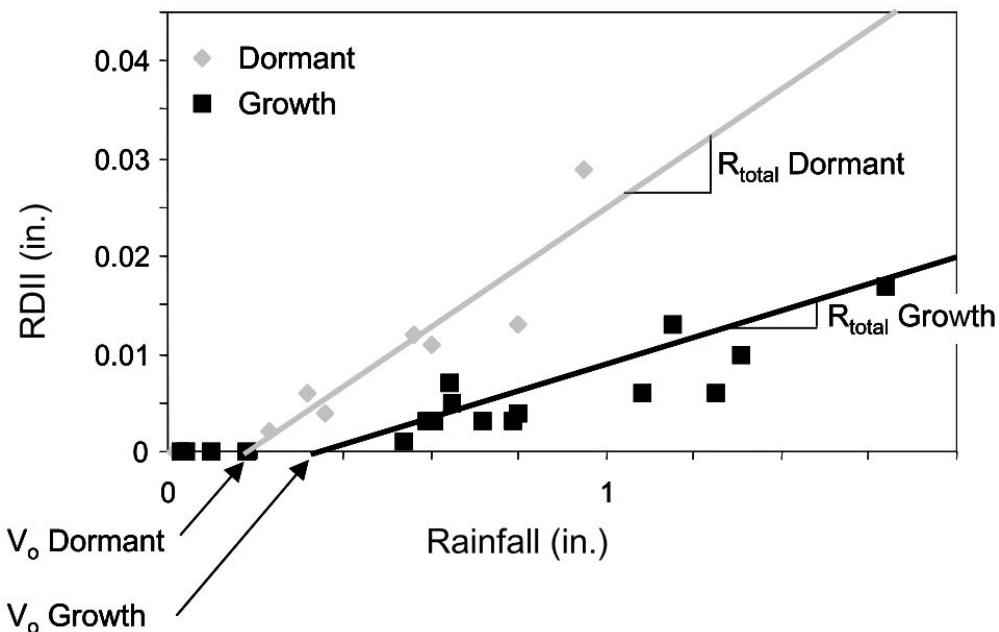


Figure 3.9.: Seasonal response to rainfall relationship [50]

### 3.4. Hydraulic model

SWMM computes the hydraulic state of the network by solving a continuity and momentum equations (3.9, 3.10) with a numerical solution procedure. These equations describe one-dimensional, gradually varied, unsteady free surface flow. Flow depth and discharge rate ( $Q$ ) for each conduit and water level in the nodes are computed each time step [37].

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = 0 \quad (3.9)$$

$$\frac{\partial Q}{\partial t} + \frac{\partial(\frac{Q^2}{A})}{\partial x} + g \cdot A \cdot \frac{\partial H}{\partial x} + g \cdot A \cdot S_f = 0 \quad (3.10)$$

Where:

$A$  = flow cross-sectional area [m<sup>2</sup>]

$Q$  = flow rate [m<sup>3</sup>/s]

$t$  = time [s]

$x$  = distance [m]

$H$  = hydraulic head of water in the conduit ( $Z + Y$ ) [m]

$Z$  = conduit invert elevation [m]

$Y$  = conduit water depth [m]

$S_f$  = friction slope (head loss per unit length)

$g$  = acceleration of gravity [m/s<sup>2</sup>]

Equations 3.9 and 3.10 are also known as Saint Venant equations and are used to compute the discharge ( $Q$ ) in the conduits. The head values ( $H$ ) at the nodes are computed with a nodal continuity equation 3.11 that assumes conservation of flow where the change in the volume is equal to the difference of flow entering the node and outflows [37].

$$\frac{\partial H}{\partial t} = \frac{\sum Q}{A_{sn} + \sum A_{sl}} \quad (3.11)$$

Where:

$\sum Q$  = net flow into the node-link (inflow - outflow) [m<sup>3</sup>]

$A_{sn}$  = node's storage surface area [m<sup>2</sup>]

$\sum A_{sl}$  = Sum of the surface area contributed by the links connected to the node [m<sup>2</sup>]

The dynamic wave routing method was chosen for this study's simulations with inertial damping. The dynamic wave solves the full form of Saint Venant equations. More simplified routing methods are also available in SWMM engine, such as kinematic wave, but were not assessed in this study. SWMM's inertial damping routine was also utilized in this study. The damping reduces the inertial component of the momentum equation 3.10 when sub-critical to supercritical flow changes occur. This procedure is used to assure more stability to the solution. The space is discretized in SWMM with a node-link approach where each node is connected to half of length of each conduit connected to it. The time domain was discretized using a variable time step length controlled by courant stability criteria. Finite difference approximations replace derivatives of the governing equations. An implicit backwards Euler method scheme is then used to approximate equations 3.9, 3.10 and 3.11 and compute ( $Q$ ) and head values ( $H$ ) at each node-link of the next time step [37].

The friction component of equation 3.10 ( $S_f$ ) is calculated using a manning's formula for free surface flow for and Darcy-Weisbach or Hazen-Williams formula for pressurized flow which are not described here for simplicity. The Darcy-Weisbach equation was used in this study and no assessment or comparison with Hazen-Williams formula was done [37].

Dynamic wave routing method requires the information of the hydraulic head at each outfall and inflows entering each node as boundary conditions. The inflows entering each node, considered as the upstream boundary condition, varies with time and are of particular interest in this study. The two external inflows entering some nodes of the network are dry-weather flow (DWF) and rainfall-derived infiltration and inflow (RDII). DWF can be determined by the wastewater load estimated of each building connected to the node varying with an average consumption pattern for the area. RDII inflows are given by the proposed hydrological model.

## 4. CASE STUDY AND MODEL SET UP

The town of Jokela in Finland was chosen as the study area. Hydraulic model and flow measurements were provided by Tuusula Water Utility (Tuusulan Vesihuolto). The following sections provide a summary of characteristics of the catchment, data providers and data treatment, and parameter estimation for physics based and unit hydrograph methods.

### 4.1. Jokela Town

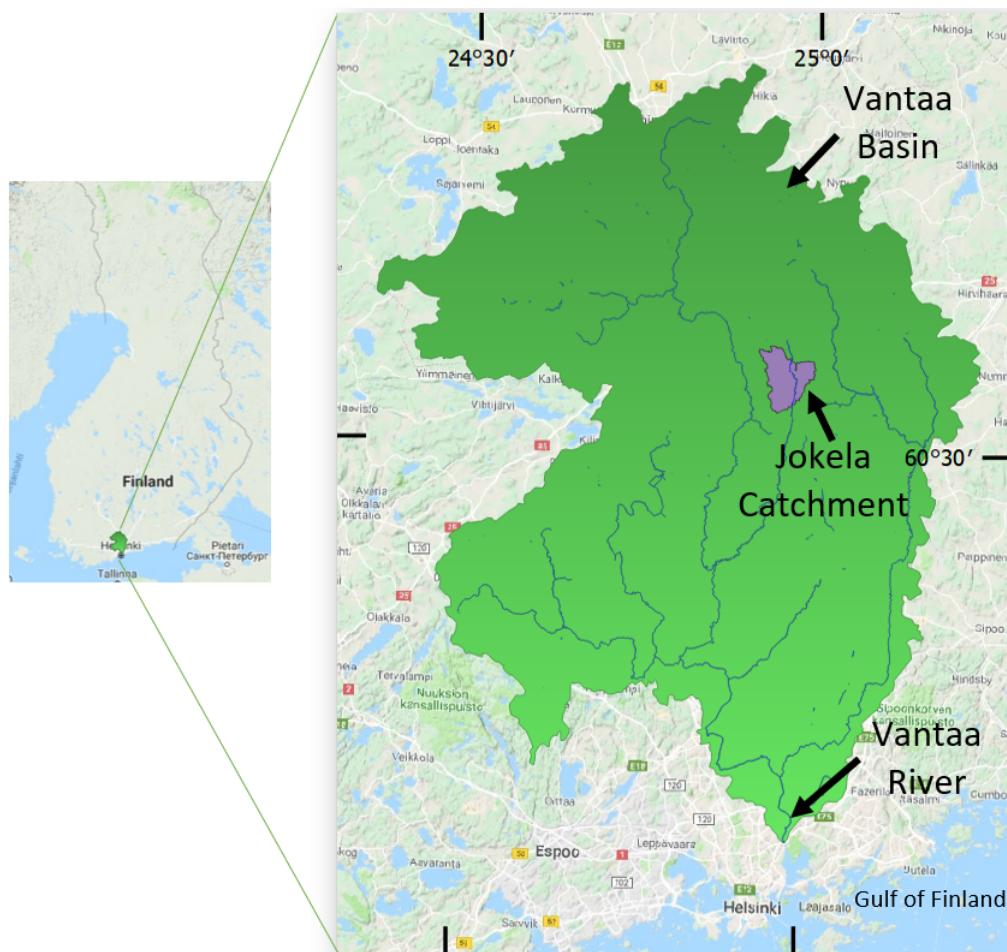


Figure 4.1.: Location of Jokela in Finland and Vantaanjoki river basin

Jokela is a town located in south region of Finland with about 40 kilometers from the capital city of Helsinki. It is part of Tuusula municipality. Jokela's urban area is mostly residential with an area

estimated of 315 hectares. There are approximately 6.500 residents and 2.975 buildings (households, commercial, and factories). The area is rather flat with an average slope of 3 %. Its land use can be considered semi-urban and roughly divided as c.a. 43 % of its catchment area with forests and semi-natural areas, 32 % with urban artificial surfaces, 13 % of transitional woodland or shrub and 12 % or pastures. Its soil superficial deposit is mostly clay (66 %) and sand (20 %). Its main drainage stream has an average width of roughly 6,5 meters width and cuts the catchment flowing from north to south draining to Vantaanjoki (Vantaa River) and traveling approximately 50 kilometers before reaching the Gulf of Finland.

## 4.2. Jokela's Hydraulic Model

Time-varying boundary condition was assigned for upstream as dry-weather flow (DWF) entering the nodes based on user consumption and daily pattern usage data for Jokela town. This was estimated from billed water distribution data for each house hold when the hydraulic model was built. Normal flow depth of the connecting conduit was assigned to the outfalls as downstream boundaries. Initial head and discharge values were set to zero. The pumping stations are represented as a storage unit node component followed by two pumps link component.

The seven pumping stations depicted in Figure 4.2 were used to divide areas for the hydrological model and receive RDII inflows with the last one downstream considered as the system outlet (Jokela pumping station). From Jokela pumping station the wastewater is pumped towards southeast through remaining parts, not considered in this study, until the outlet representing the waste water treatment plant of Tuusula urban drainage network. There are 1.550 DWF inflows representing the demand of buildings entering the Jokela's network system through several nodes. The network length is approximately 46,7 km, average slope of 5,6 %, average diameters of 0,22 m and daily dry-weather inflow of approximately 766 m<sup>3</sup> for the upstream parts of all seven pumping stations.

No calibration was carried for the hydraulic model. A simple comparison between simulated and measured dry-weather flow (DWF) hydrographs of the network's Jokela outlet was carried. Relatively dry days of raw inflow data in October 2018 were chosen for the comparison depicted in Figure 4.3 with the first two days being Thursday and Friday and the last two the weekend.

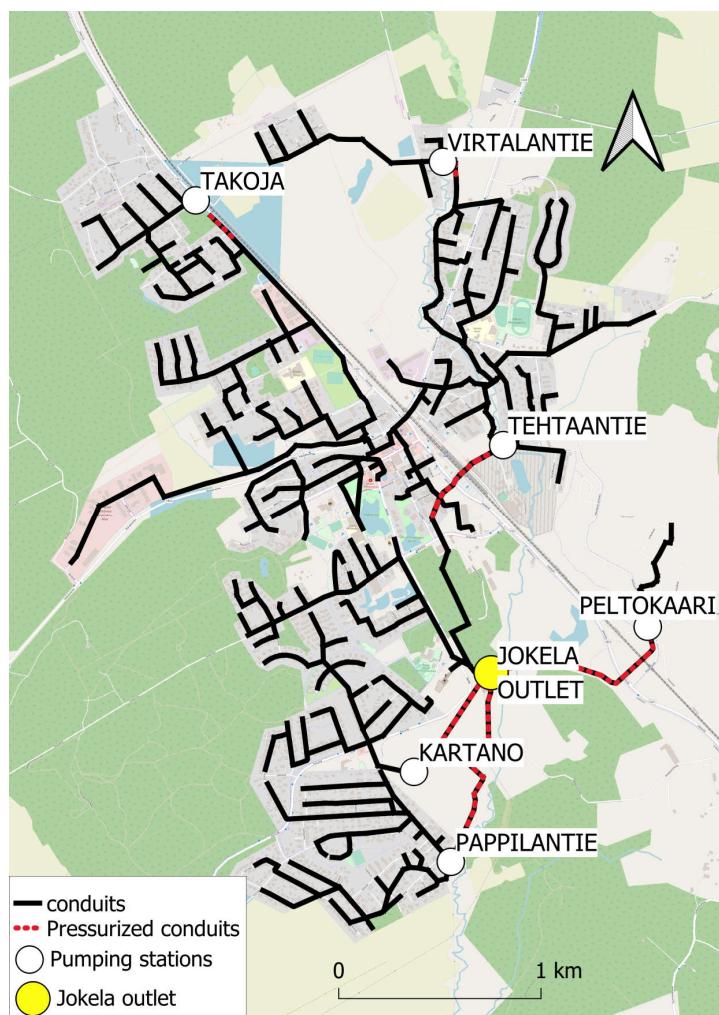


Figure 4.2.: Jokela's sanitary sewer network

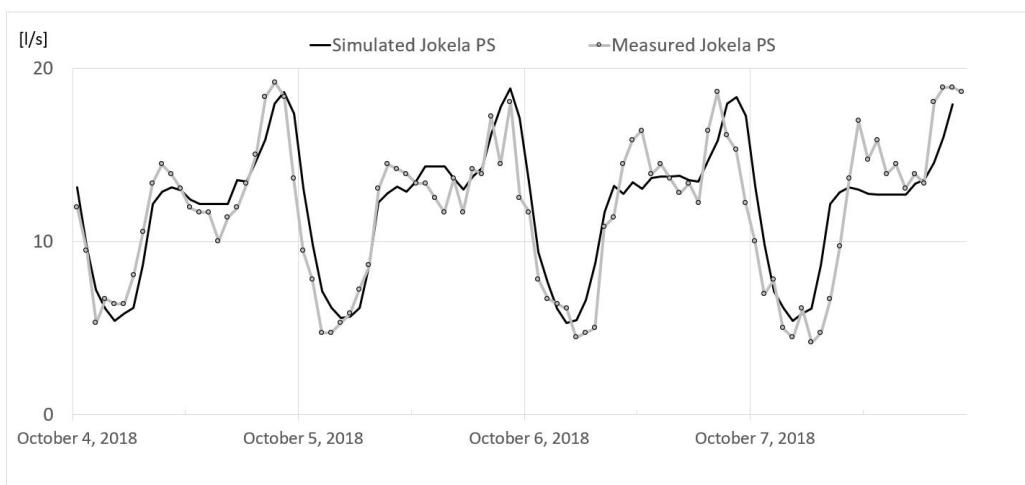


Figure 4.3.: Dry-weather flow simulated versus measured in Jokela pumping station

## 4.3. Data and Parameter Estimation

This section presents the details of parameter estimation for the physics-based model (SWMM modules) and RTK-Unit Hydrographs and the data sets used for model construction, calibration and validation.

### 4.3.1. Sanitary Sewer Flow Data

Flow meter data was provided by Tuusula Water Utility which is responsible for urban drainage system management of Jokela town and all Tuusula municipality. Data was measured at SSN pumping stations, two of which are used for calibration and validation purposes in this study. One of the meters is located at the last pumping station (Jokela Pumping Station - subcatchment D4S4) of the delineated area and it is considered the system outlet. Hourly measurements were recorded during almost whole period of year 2018. The first six months of 2019 were also recorded and used in this study, but following descriptions of the procedures were concentrated in 2018 data.

The flow data was preprocessed to eliminate missing data, outliers and normalize it to reduce noisy measurements as an attempt to capture the overall evolution of the flow. Figure 4.4 depicts the distribution of measured data before and after the elimination of outliers and missing data. The outliers, here defined as flows greater than 200 L/s, appeared as spikes measured only at one data interval unit (1 h). This suggests that these measurements are in fact unreal values and were filtered out. Missing measurements were filled in using simple linear interpolation. Zero flow measurements were treated as missing data too. When periods longer than 12 hours of missing data were observed no interpolation was carried. These values were left out of the future analysis.

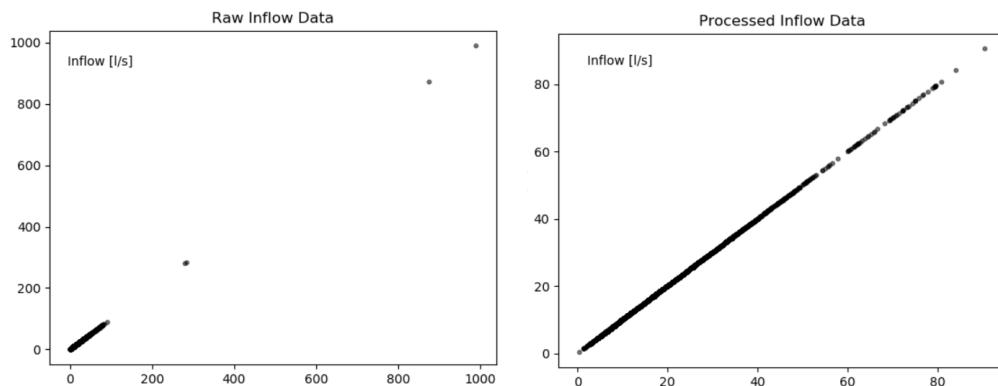


Figure 4.4.: Outliers removal

Outliers presenting extreme flow were relatively easy to identify and filter out using a threshold value. However, unreal measurements presenting customary flow rates were harder to filter using the threshold method. An example of such measurement is depicted in Figure 4.5 on February 22nd at 20:00. The flow data was then smoothed using Savitzky-Golay filter [39] with window length of 11 and 5th order polynomial. It is important to mention that care should be taken when smoothing

the data, since it can erroneously reduce the peak flow measured. Periods with flows higher than average and peak flows are most important when simulating wet-weather flow conditions in SSN. Higher order polynomials better reproduce the peak flows, but also keep unrealistically low flows. The pretreatment procedure is mostly used for auto-regressive models that use historical flow data not only for comparing the results and calibrating, but also to build the model itself. Li et al. [29] performed similar pretreatment on a sanitary sewer flow data derived from pump ON/OFF operation to transform non-stationary data into stationary before applying autoregressive-moving-average model (ARMAX).

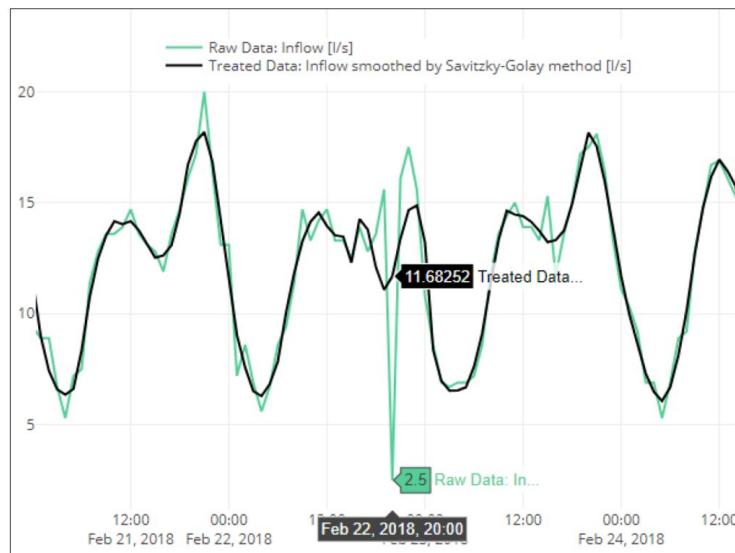


Figure 4.5.: RDII x DWF estimate for 2018 of total wastewater load

Data set of 2019 was also pretreated as described above. An external script written in python [33] was developed to perform the flow data treatment (see in the A). Other relevant tool used was the EPA SSOAP toolbox [50]. It helped to identify RDII flows from the raw flow data, as described in the following section.

#### 4.3.2. Decomposition of flow Components

It was relevant to identify the different components of the sanitary sewer network (SSN) to estimate the amount of RDII flows and compare measured with modelled when analyzing the water balance for short or long-term simulations. The estimated "measured" RDII flows were identified using EPA SSOAP toolbox.

EPA SSOAP toolbox has many functionalities. As its core, it allows the users to visually build RTK unit hydrographs through an interactive process by adjusting RTK parameters and comparing the output hydrograph with the hydrograph measured [50]. However, the SSOAP toolbox was mostly used in this study to identify precipitation events and identify the flow components in Jokela's SSN. RTK parameters were estimated in this study using an optimization algorithm as described in the Chapter 2.

Two input data were used in SSOAP: the sanitary sewer flow data of Jokela pumping Station; and the precipitation data. SSOAP utilizes precipitation data to identify wet-periods and define precipitation events. Days without wet-weather flow (WWF) are then filtered to define a mean dry-weather flow (DWF) hydrograph for weekdays and weekends. Current day and previous two days with any precipitation record were filtered out of the data set for weekdays and weekends. Past three to seven days with recorded total precipitation higher than 10 millimetres and 15 holidays in Finland were also filtered out of the set. A total of 66 week days and 32 weekend days remained from DWF identification process. The output DWF hydrographs are depicted in Figure 4.6. Weekdays have average DWF of 11,4 L/s with the standard deviation of 1,75, and 12,5 L/s and the standard deviation of 2,35 for weekends. Refer to Vallabhaneni and Burgess [50] for detailed information about the SSOAP toolbox.

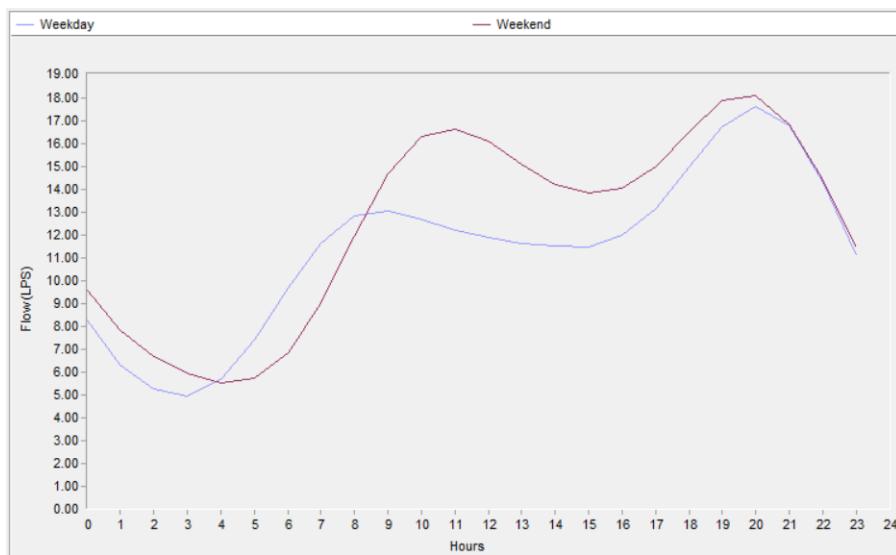


Figure 4.6.: Estimated dry-weather hydrograph obtained from EPA SSOAP tool

The next step was subtracting the DWF estimated from metered flow to quantify RDII flow ( $RDII = MeteredFlow - DWF$ ). First, a constant groundwater infiltration rate value was defined. This is important to further divide the DWF into two parts: basewaste flow (BWF) and GWI. The GWI is estimated as a percentage of the minimum nighttime flow. SSOAP help file suggest that in residential areas typically 90 % of the minimum nighttime flow represents GWI. However, a different approach was chosen based on available information from the water utility. It was known that an estimate of 766 m<sup>3</sup>/day is pumped to the Jokela town area from the water distribution system. It was assumed that almost all water supplied eventually finds its way to the sanitary sewer network, meaning that the daily average BWF is roughly 766 m<sup>3</sup>/day. To achieve this value for BWF with 2018 recorded flow an estimate of 55 % of daily minimum nighttime flow was attributed to GWI with the rest (45 %) representing wastewater flow. Table 4.1 depicts the estimated daily average flow components of Jokela's SSN with approximately 47 % of the total wastewater amount transported to the WWTP being non-revenue water (NRW) from RDII and the constant GWI. Figure 4.7 shows the total monthly RDII and DWF. Approximately 88 % of total annual RDII enters the network between January and May of 2018.

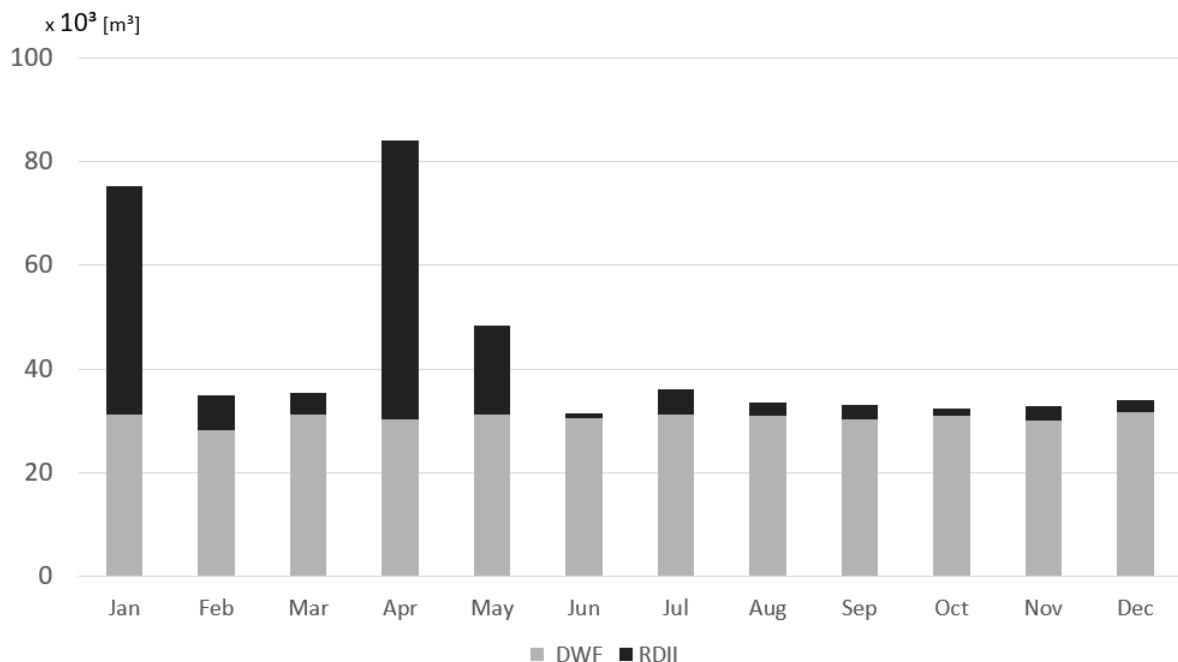


Figure 4.7.: RDII and DWF estimate for 2018 of total wastewater load of Jokela SSN

Flow data set of the first six months of 2019 were also analyzed using EPA SSOAP toolbox assuming the dry-weather estimation made with the flow data set of the year 2018.

It is important to remember that the RDII flow can also be further divided (i.e. short-term hydrograph from runoff and long-term hydrograph from sub-surface infiltration), and the varying groundwater infiltration happening after rainfall or snowmelt is considered here different than the constant GWI used in SSOAP toolbox.

#### 4.3.3. Meteorological data and parameters

Historical meteorological measurements and weather forecasts were used in this study. Firstly, as input for hydraulic models during calibration and validation process. Secondly, to simulate and preliminary analyze the performance using model forecast. Precipitation data is used for both physics-based and RTK unit hydrograph method. Snow depth measurements were used as input for the snowpack and snowmelt SWMM module process and to calibrate its parameters. Precipitation, temperature and snowdepth historical measurements were fetched from FMI's weather station 101130 - hyvinkää hyvinkäänkylä, located about 10 km northwest from Jokela Catchment. Wind speed and direction data were fetched from FMI's weather station 103794 - Mäntsälä Hirvihaara located about 15 km northeast from Jokela Catchment. These were the closest stations from Jokela Catchment. All meteorological data sets had 1 hour recording interval.

Table 4.1.: Estimated values of sanitary sewer network (SSN) flow components in 2018 for Jokela Catchment in cubic metres

<b>Month</b>	<b>Obs Flow</b>	<b>DWF</b>	<b>BWF</b>	<b>RDII</b>	<b>RDII + GWI</b>	<b>% RDII</b>	<b>% RDII + GWI</b>
Jan	2.429	1.007	765	1.423	1.665	58,6	68,5
Feb	1.241	1.006	764	237	480	19,1	38,6
Mar	1.136	1.010	767	136	379	11,9	33,4
Apr	2.801	1.011	768	1.790	2.033	63,9	72,6
May	1.556	1.010	767	547	789	35,1	50,7
Jun	929	1.015	770	32	277	3,4	29,8
Jul	1.041	1.007	765	159	401	15,3	38,6
Aug	932	1.003	762	82	324	8,8	34,7
Sep	1.028	1.011	768	94	337	9,1	32,8
Oct	986	1.003	762	39	280	4,0	28,4
Nov	1.066	1.004	762	89	330	8,3	31,0
Dec	1.025	1.022	776	77	323	7,5	31,6
<b>Total</b>	<b>16.169</b>	<b>12.110</b>	<b>9.195</b>	<b>4.704</b>	<b>7.618</b>	<b>29,1</b>	<b>47,1</b>

#### 4.3.4. Precipitation data

Precipitation is one of the most important inputs for hydrological models. The same precipitation data set was included for both of the methods: physics-based and RTK Unit Hydrograph.

Precipitation data of entire 2018 and part of 2019 recorded by 101130 - hyvinkää hyvinkäänkylä rain gauge was assessed. The total amount of precipitation recorded in 2018 was 493 mm with 199 mm recorded in the first six months of the year. The first half of 2019 presented a sum of 285 millimetres. approximately 43 % more than the previous year for the same period. Figure 4.8 depicts the monthly accumulated precipitation. It is possible to observe also considerable monthly variation when comparing the two years. No pretreatment was applied to the precipitation data recorded in the hyvinkää hyvinkäänkylä station.

A simple comparison of rain gauge and radar recordings of a rainfall event that happened on 16th July, 2019 was carried for the study area to assess the possible variations of the two methods. The radar data grid size used was about 1 km, with one hour interval precipitation amount recordings. This data set is also available in the Finnish Meteorological Institute (FMI) open data platform ([19]). Values of grids falling within one subcatchment were summed to represent the rainfall records of each subcatchment. Therefore, using radar information allows the creation of one SWMM rain gauge component for each subcatchment providing finer resolution for the precipitation spatial distribution. There are considerable differences between radar and gauge records for the evaluated event in the study area, as depicted in Figure 4.9. Each subcatchment has its own hyetograph with different shape and peak values. Total rain gauge measurements captured only 37 % of the rainfall amount in comparison with radar during 13 hour event (4:00 to 16:00) for the entire catchment area. Rain gauge also recorded less than half of the hours recorded by radar presenting a different hyetograph shape.

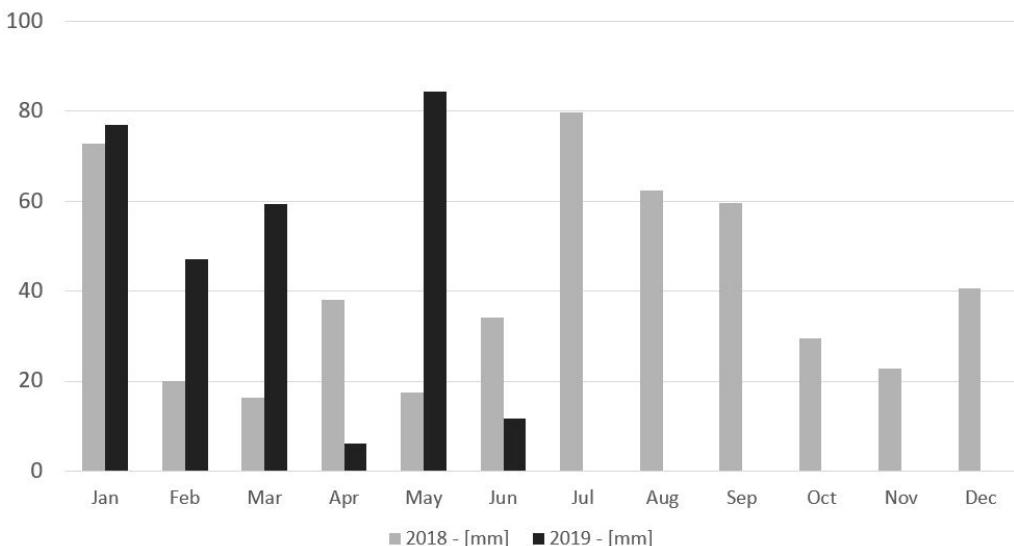


Figure 4.8.: Comparison of the recorded precipitation amount of the years 2018 and 2019.

Therefore, rain gauge measurements used seem to underestimate rainfall over the study area.

#### 4.3.5. Temperature data

Temperature data was used for the snowpack and snowmelt routine in SWMM for the physics-based model. The data set provided data for the entire year of 2018 with hourly recording intervals. It is important to mention that, according to the Finnish Meteorological Institute (FMI), 2018 was among the ten hottest years ever recorded [21]. The hottest temperature of the year was 30,7°C, recorded on 1st of August, and the coldest was -25,5°C, measured on 22nd of February. Figure 4.10 depicts the mean daily air temperature of the data set used. Only one missing data point was found and filled using linear interpolation between the neighbouring measured values.

#### 4.3.6. Wind and evaporation data

SWMM uses the wind data to calculate for the snowpack routine. The monthly wind average speed was calculated from FMI's weather station id 103794 - Mäntsälä Hirvihaara, and added as input to SWMM physics-based model. The RTK Unit hydrograph method did not use wind speed information. The monthly wind average speed ranged between 7,6 km/h, recorded on May to 11,7 km/h, recorded on February.

Mean annual pan evaporation estimates of 343 mm/year for Finland [30] was used as basis for the monthly evaporation rate input to the physics-based model. Evaporation was set to occur only during dry-weather periods. Most of the total annual evaporation was set to occur during the beginning of the growth season, with the highest evaporation rate of 93 mm/month occurring on July. A total of 300 millimetres evaporation was assumed to occur from May to September. The evaporation input is used

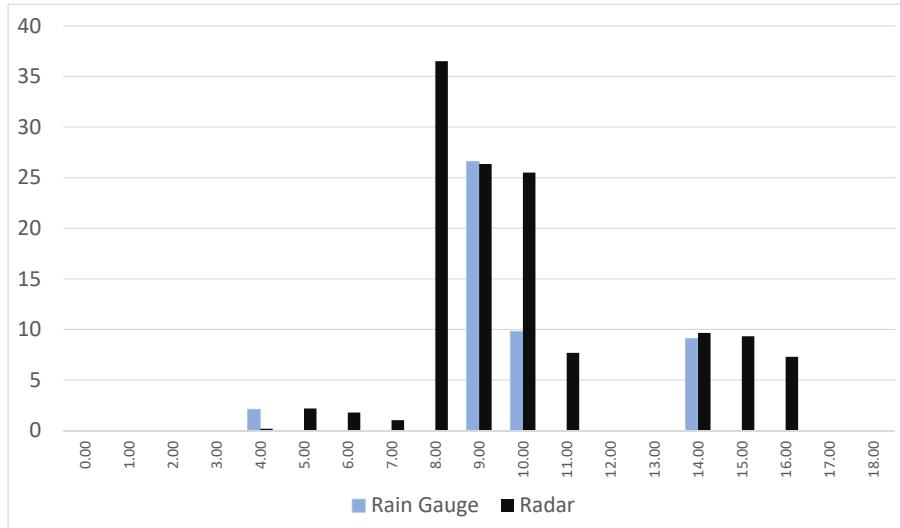


Figure 4.9.: Comparison of rain gauge and radar-based total precipitation amount measurements for Jokela's catchment

for the computations of the water balance in the physics-based model. It is used for the calculations in the subcatchments, upper-zone of the aquifer, and in the snowpack.

#### 4.3.7. Snow depth data and snowpack parameters

Snow depth measurements were taken in the same location as precipitation and temperature. The maximum snow depth in 2018 was 35 cm recorded on April, ten days before the complete melt of the pack. The average snow depth from January to April was 21,4 cm. Data of the year 2019 was also used in this study, mostly for validation, with a maximum of 69 cm recorded on February and 37,5 cm average values from January to April. It is important to remember that SWMM computes depth of water equivalent instead of thickness of the snow pack as described in Section 3.2.2. Therefore, thickness of snow pack is used here to calibrate the depth of water equivalent mostly to evaluate periods of snow accumulation and snow melt. The amount of water equivalent present in the snow pack is expected to be lower at the beginning of the winter as increase as snow melts and freezes again increasing the water content in the pack.

Estimation of parameters was carried based on the ranges proposed in Chapter 3. Sensitivity analysis was carried by manually varying individual parameters while others were kept constant. It was found that the degree-hour melt coefficients ( $DHM_{min}$ ,  $DHM_{max}$ ), dividing temperature ( $T_{div}$ ), base temperature ( $T_{base}$ ) and snow catch factor ( $SCF$ ).

Base temperature and dividing temperature were initially estimated by observing relations among hourly temperature and snow depth. Whenever the snow depth measurement varied, values of temperature where collected. Temperatures during the increase of the snow depth were stored to estimate the dividing temperature( $T_{div}$ ) whereas temperatures during decrease of snow depth were used to es-

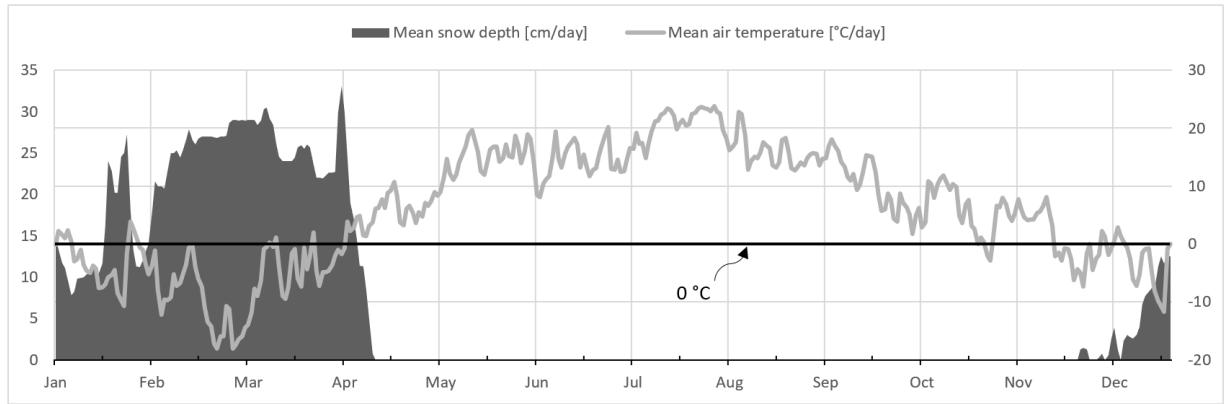


Figure 4.10.: Snow depth and Temperature Measurements FMI [19]

timate the base temperature( $T_{base}$ ). Periods, when the temperatures were out of the selected range were discarded. (refer to Table 3.1 for the selected range). The mean of the stored temperature was then used as the final estimation of  $T_{div}$  and  $T_{base}$ , and were respectively 0,1 °C, and -1,9 °C. It is important to emphasize that this was a rough estimate used only as the initial guess for the parameter calibration.  $DHM_{min}$  and  $DHM_{max}$  were set initially as the limits of Table 3.1.  $TIPM$  and  $RNM$  were set as used by Rossman and Huber [38] and Anderson [6]. There was no information assessed about the rain gauge deficiency to record snowfall and fraction of free water capacity. Therefore,  $SCF$  and  $FWFRAC$  values were set initially to their middle value of their respective range. Two parameters for initial condition of the simulation are also required: the Initial depth of water equivalent ( $SD_0$ ); and the initial free water ( $FW_0$ ). The first was estimated as 10 % of the first measured value of snow depth assuming, a ratio of 10:1 for the snow pack depth and water equivalent depth, as a rule of thumb suggested by Rossman and Huber [38].

Figure 4.11 shows the results of the simulation using the initially estimated parameters and the results of a simulation with parameters manually calibrated. Only the first four months of 2018 are reported, between 1st of January and 15th of April, when all snow had already melted.

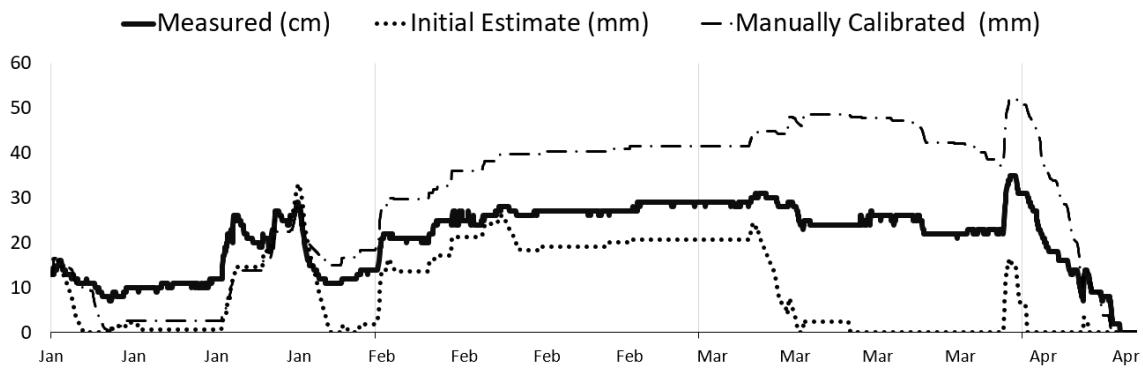


Figure 4.11.: Results of parameter estimation based on literature review and manual calibration of the snowpack and snowmelt parameters

By a visual analysis of the results of the first simulation and comparison to the measured snow depth values, it is possible to identify a higher rate of melt and snow accumulation than the measured values. This was corrected after a manual calibration of parameters. A lower rate of snow accumulation was achieved by assuming that the rain gauge was able better measure the snowfall. Therefore,  $SCF$  was reduced to 1,2. To reduce melting rate, the temperature which snow melt starts ( $T_{base}$ ) was increased to its upper limit of 0 °C. Although the results were better, higher melt rate caused the snow pack to completely melt three weeks too early. Therefore, the minimum melt coefficient ( $DHM_{min}$ ) was lowered further than its lower range limit from 0,019 mm/h-°C to 0,001 mm/h-°C.

Table 4.2 depicts the initially estimated parameters, parameter values after manual calibration against measured data, and changes in the previously proposed range based on literature review (see Table 3.1) after results of the manual calibration. The only changed was a reduction of  $DHM_{min}$  parameter by approximately 47 %.

Table 4.2.: Snowpack and snowmelt estimated parameters

parameter	initially estimated	manually calibrated	units
$T_{div}$	0,1	-	[°C]
SCF	1,5	1,2	[1]
Tb	- 1,9	0	[°C]
DHM - DHM	0,019–0,10	0,009–0,03	[mm/h-°C]
RNM	0,6	-	[1]
FWFRAC	0,13	-	[1]
TIPM	0,5	-	[1]
SD0	13	-	[mm]
FW0	0	-	[mm]

#### 4.3.8. Weather Forecast

Precipitation forecast output from the numerical weather forecast model HARMONIE was used in this study. The Finnish Meteorological Institute (FMI) provides the model results through its free open data platform [19]. FMI also provides data from HIRLAM forecast model. Both are limited area models (LAMs) where HARMONIE covers scandinavia and HIRLAM a larger area of the north hemisphere (see Figure 4.12). HARMONIE model was chosen for this study due to its finer horizontal grid resolution of 2,5 km. The model is updated every 6 hours (00, 06, 12 and 18 UTC) with forecasts for approximately the next 55 hours [19].

It is important to emphasize that the HARMONIE is only one of the weather models used by FMI. The institute uses different weather forecast models. Models from the European Centre of Medium Range Forecasts (ECMWF) combined with the meteorologists experience are used to update the published weather forecast, which is not included to their open data platform. Therefore, the results of

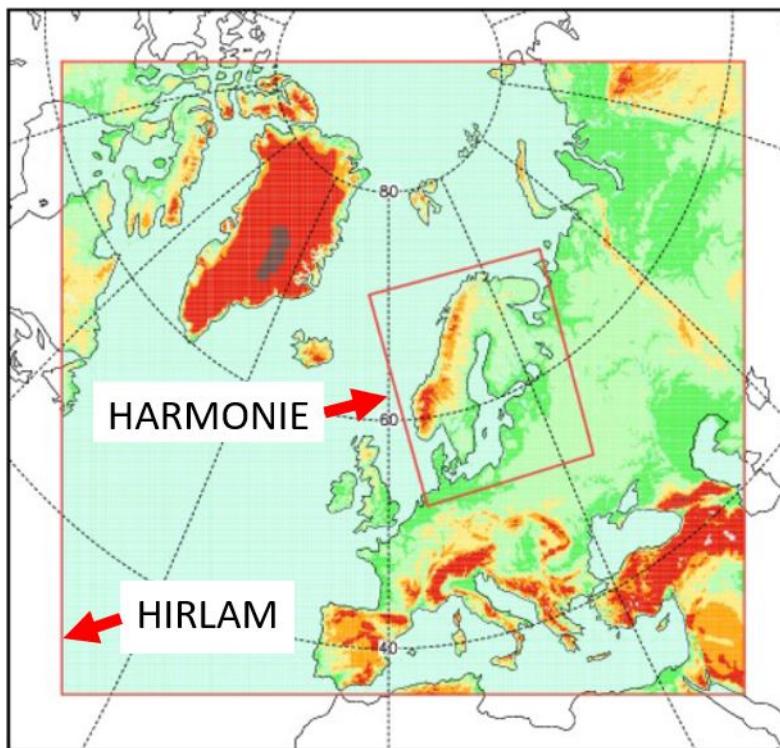


Figure 4.12.: Numerical weather forecast models coverage. [20]

the hydrological models proposed in this study are prone also to the limitations of using only one numerical weather forecast model as input for the weather forecast. The Traffic Management Finland Group (TMFG) provides also weather forecast conditions for Finnish roads. This can be used as an alternative, or together with HARMONIE model, for locations in Finland. Forecasts of weather variables such as precipitation and temperature can be fetched based of the road weather forecast. the id of the roads crossing, or nearby, the catchment area can be used to filter precipitation forecast for the region. TMFG's provides the road weather forecast through its open data portal (Digitraffic, [44]) for the next 12 h with 2 h interval. Precipitation forecast is given in seven different ranges: one for dry weather, three for rainfall intensity and three for snowfall intensity [45]. The road weather forecast was not used for the analysis in Chapter 5, since the service was temporarily interrupted during the summer period of 2019.

HARMONIE data used is provided as point value pair time series in extensible markup language (XML). Queries to FMI's WFS request forecasted values for specific spatial location chosen by latitude and longitude coordinates. Jokela's pipe network centroid location was calculated and used as latitude = 60.55922 and longitude = 24.97789 coordinates. TMFG's weather road forecast in javaScript Object Notation (JSON). Queries using specific road codes are possible - such as segments of the road 45 located on the east side of Jokela Town. An external python script [33] was written to collect and parse data into time series in friendly format for SWMM model application. All the scripts developed in this study are available on the appendix.

#### 4.3.9. Topographic data

Topographic data was assessed to estimate parameters used to model different flows occurring in Jokela's catchment. The chart presented in Figure 4.13 shows the relation among topographic data sets and the process being modeled. Data description and initial parameter estimates are provided in the following sections.

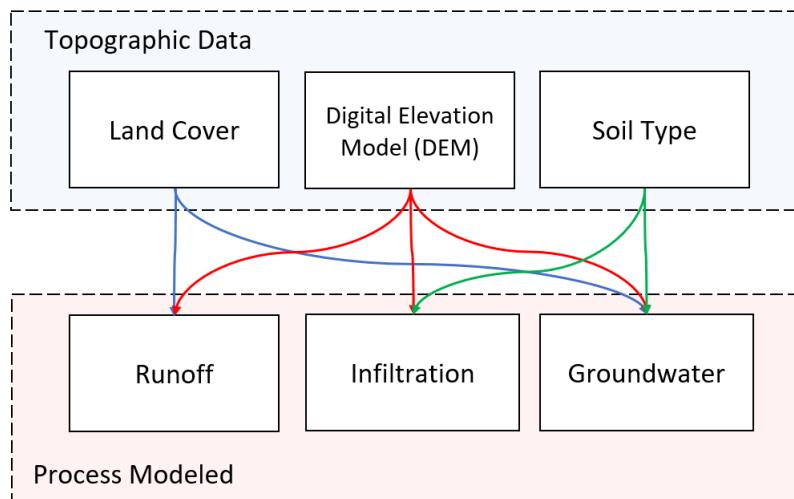


Figure 4.13.: Topographic data used for each process of the physics-based model

Description of the data and initial estimates for parameters are provided in the following sections.

#### 4.3.10. Terrain data and runoff parameters

Digital elevation model (DEM) was used in many parts of this study. It was used in the runoff model to delineate the subcatchments that provided surface area information for all the SWMM modules used. The DEM used in this study was provided by the National Land Survey of Finland (NLS) through its open data platform. The DEM use had a 2 x 2 m resolution. The data was collected in the summer of 2015. The vertical accuracy varies from 0,3 to 1 meter [32]. The DEM provided had already bridges cut off and a "filling" operation was carried using QGIS 3.6.1 [34] to exclude possible holes existent in the data set.

The definition of the surface area influencing the RDII is rather conceptually challenging. As mentioned previously, surface and groundwater flow influence the SSN flows and their delimited area respective area of influence are different. The watershed is usually divided by its elevation which determines the direction of the surface runoff. The groundwater flow depends on the characteristics of the aquifer, which can have a very different delimited area in comparison with the surface watershed area. This can be imagined with the classical example of a rain drop falling over a specific point on the soil surface. The drop can flow over the surface towards the direction of the steeper slope or infiltrate, and then flow through the porous in the soil to very different direction than it would flow, if it remained on the surface.

Boundaries of a catch basin can be defined using the highest elevations, such as topographic crest and road center line, or lowest points of the area such as rivers and streams [28]. Cadastral parcels, which are property developed in urban areas, are also often used to delineate the catchment basin. However, parcels are rather divided for administrative reasons than soil or subsoil characteristics. Thus, it can be challenging to estimate some topographical parameters, such as soil surface slope and roughness using administrative areas. As a hypothetical exercise, one can imagine the possible issue when using parcels when modelling an urban area surrounded by mountains. In case the delineated area is limited only to the properties, no information relative to the mountains' is included in the parameter estimation, and simulated, hydrographs can mismatch the observed volume and shape.

The sewershed delineation is, therefore, a discretization of the space domain. In this study, delineation is based on the area division that supplies water to a specific point within the SSN. The DEM is usually the data used for delineation of the area of influence when modeling natural rivers. This method, however, is not conceptually valid for SSN, since the network's slopes not always follow the soil surface slopes estimated with the DEM.

After leaving the buildings, the wastewater is transported by the SSN conveyed with the gravity force, when possible. When the wastewater reaches points in the system where it needs to be transported to higher elevations, extra energy may be needed. Pumping stations are then installed to supply energy for the fluid to overcome the height change and flow forward towards the disposal point. In other words, when the wastewater leaves the buildings, it should flow in the SSN towards a pumping station, where it receives energy to overcome the elevation increase, or directly to the disposal point. Therefore, the pumping stations in the SSN are collection points where wastewater coming from the upstream pipe network flows to. This is one of the reason why, in this study, the space domain was discretized based on the pumping stations. When stormwater infiltrates the SSN it is diverted to the next pumping station downstream or directly to the disposal outlet. The second reason is that flow meter observations were recorded in two of the pumping stations within the catchment, and the remaining stations are probably the best candidates to receive flow meter devices in the future.

DEM, pumping station locations and pipe network were the data used for the subcatchment delineation. QGIS application was chosen to perform GIS operations and delineate sewersheds, because it is freely available and open source. It also has automatic delineation tools, such as GRASS - r.watershed [1]. The proposed subcatchment partition for Jokela town is depicted in Figure 4.14

Pumping stations are usually placed in locations with lower surface elevation, when compared to their upstream pipe network, which diverts the wastewater utilizing gravitational force reducing energy consumption. However, differences between soil surface and network still exists in some parts of the pumping station upstream service area. This happens because no information about soil type and infiltration rates is used when delineating the areas. Therefore, subcatchment of pumping station one (PS 1) may overlap with the pipe network from the upstream pumping station (PS 2). Only automatic sewershed delineation procedure and DEM are not enough to avoid overlapping without further data processing. To overcome this, a pipe network buffer area can be delineated and summed

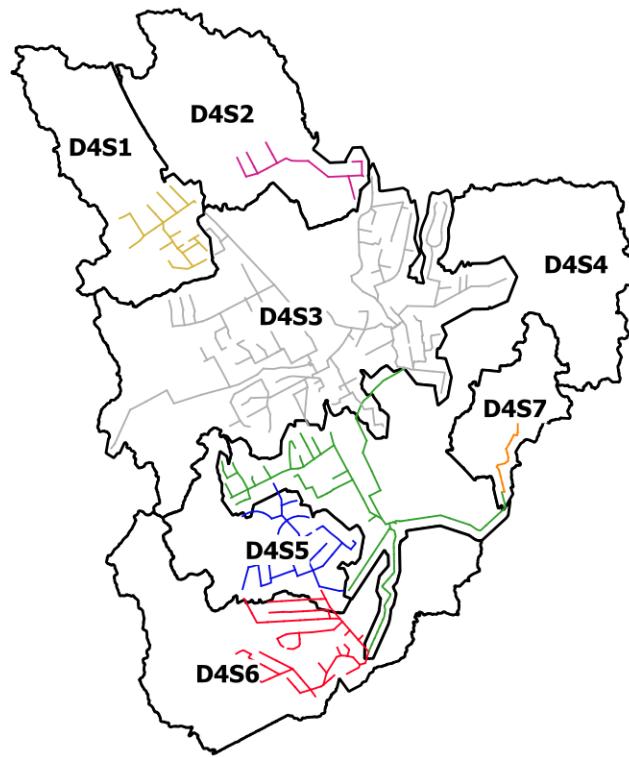


Figure 4.14.: Jokela subcatchment division

with its subcatchment or the pipe network vector data can be "burned" into the DEM raster data. Figure 4.15 depicts the differences on watershed delineation using *GRASS – r.watershed* and a "Filled DEM" and "Burned DEM".

The pipe network included in the DEM acts as artificial streams, successfully limiting the subcatchment borders when pipes upstream from different pumping station do not overlap. Some level of manual adjustment was still required when pipe crossings exist. Another way used and proposed in this study to delineate the area influencing RDII flows is based on the pipes dimensions. It was proposed that the area of influence is proportional to the size of the network components (i.e. pipe length and diameter). This assumes that the amount of defect (i.e. pipe cracks) is proportional to the pipe size. In this study, a combination of the buffered area, proportional to the pipe length and diameter, and the soil surface topographic crest was used to delineate the subcatchments, as depicted in Figure 4.14. Further adjustments could still be done considering roads, railways, artificial barriers and streams present in the study area. Figure 4.16 depicts the two different delineation methods (D1 = buffer over pipe size, D3 = topographic crest delineation) used for the final delineation (D4 = D1 + D3).

For organizational purposes, subcatchments were named after delineation method (i.e. D4) and subcatchment number (i.e. S2). The letters were dropped and the subcatchments are referred are referred simply using their numbers (i.e. 42) from now on.

To define other parameters used in SWMM's runoff module a Land Cover data set was assessed.

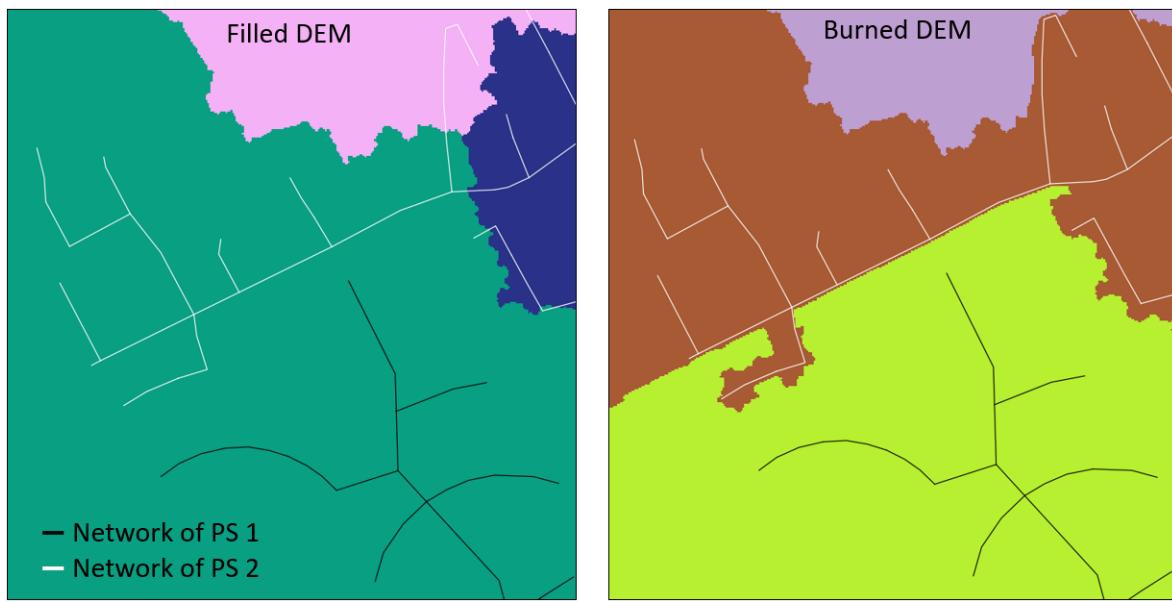


Figure 4.15.: Comparison of sewershed delineation using Filled and Burned DEM

The Finnish Corine Land Cover 2018 (CLC2018) distributed by SYKE's open data platform [41] was selected for this purpose. This data is an ensemble of different data sets, such as topographic database, digital road database of Finland, buildings, land parcels and also data interpreted from satellite images. The source data are from the 2016–2017 period and the final raster data has 20 m resolution. This data set was produced by SYKE as part of EU Copernicus Land monitoring project [17] and follows its standard nomenclature for land cover class. The raster data set has four hierarchy levels of land cover class and several sub-classes [41]. If all four hierarchy levels and their sub-classes are used, a better spatial representation of land cover is obtained. However, the choice of which level to use depends on the size of the delineated sewershed. For this study three levels of hierarchy were assumed to suffice. The vectorized and clipped data set for the study area is depicted in Figure 4.17.

Land cover of artificial surfaces were assumed as impervious areas whereas agricultural, forest, semi-natural and wetlands were assumed as pervious areas. This division was then used to calculate the percentage of impervious area, an input parameter for the SWMM runoff module. The roughness coefficient (the  $n$  parameter in Manning's equation) was also estimated based on the land cover data. For this, a relation between standard Copernicus land cover class description and Manning's roughness coefficient table available in the literature [38] was created. An area weighted average calculation was carried for the different types of land cover and, therefore, roughness coefficients are present within a subcatchment. In other words, roughness coefficients of land cover with higher percentage over the total subcatchment's area have higher influence in the final value assigned to the subcatchment. This was done for both impervious and pervious area roughness estimate.

The depression storage parameter were roughly estimated based on recommendations from Denver Urban Drainage and Flood Control District (UDFCD, 2007). The percentage of impervious land covers

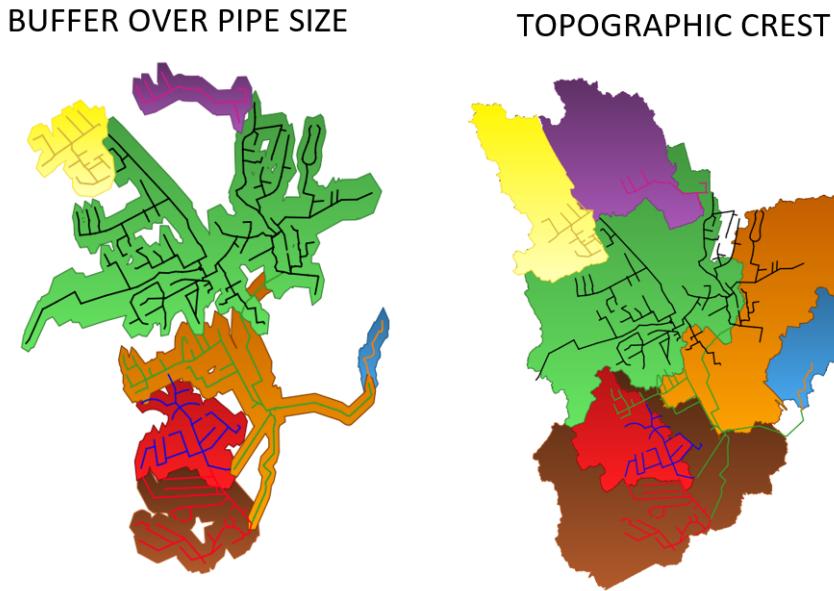


Figure 4.16.: Delineation methods. D1 (left) and D3 (right)

areas within a subcatchment were multiplied by 2,54 millimetres and pervious land cover areas by 10,16 millimetres before the area weighted average was taken. Table 4.3 depicts the estimated values for runoff parameters for each subcatchment as well as some relations to their respective SSN.

Table 4.3.: SWMM runoff parameter estimation

<i>SC</i>	<i>area</i> [ha]	<i>area / area<sub>total</sub></i> [%]	<i>Mean Slope</i> [%]	<i>n<sub>imp</sub></i> [s/m <sup>1/3</sup> ]	<i>n<sub>perv</sub></i> [s/m <sup>1/3</sup> ]	<i>area<sub>imp</sub></i> [%]	<i>Ds<sub>imp</sub></i> [mm]	<i>Ds<sub>perv</sub></i> [mm]	<i>Area/ net</i> [m]	<i>Mean elev</i> [m]
41	134	9,70	6,33	0,0321	0,0810	34,6	0,880	6,64	488	87,4
42	157	11,6	6,05	0,0323	0,075	10,6	0,270	9,08	889	80,2
43	383	28,3	5,90	0,0317	0,082	52,5	1,33	4,83	179	77,1
44	314	23,2	4,89	0,0310	0,0770	25,1	0,640	7,61	315	72,0
45	97,7	7,2	6,05	0,0310	0,081	49,4	1,25	5,14	240	77,8
46	212	15,7	6,01	0,0309	0,0760	29,1	0,740	7,20	363	73,0
47	58,1	4,3	2,61	0,0350	0,0920	1,30	0,0300	10,0	930	71,1
sum* or mean	1.355*	100*	5,41	0,0323	0,0806	28,9	0,740	7,22	486	76,9

*SC* = subcatchment

#### 4.3.11. Soil Superficial deposits and Infiltration Parameters

The soil type coverage data was fetched from Geological Survey of Finland (GTK) [23] through its open data online service, and was used to estimate the three parameters of the Horton infiltration (see Section 3.2.3). The available information was obtained as vector data containing superficial deposits of Finland with material produced between 1972–2007. GIS operations were carried using QGIS

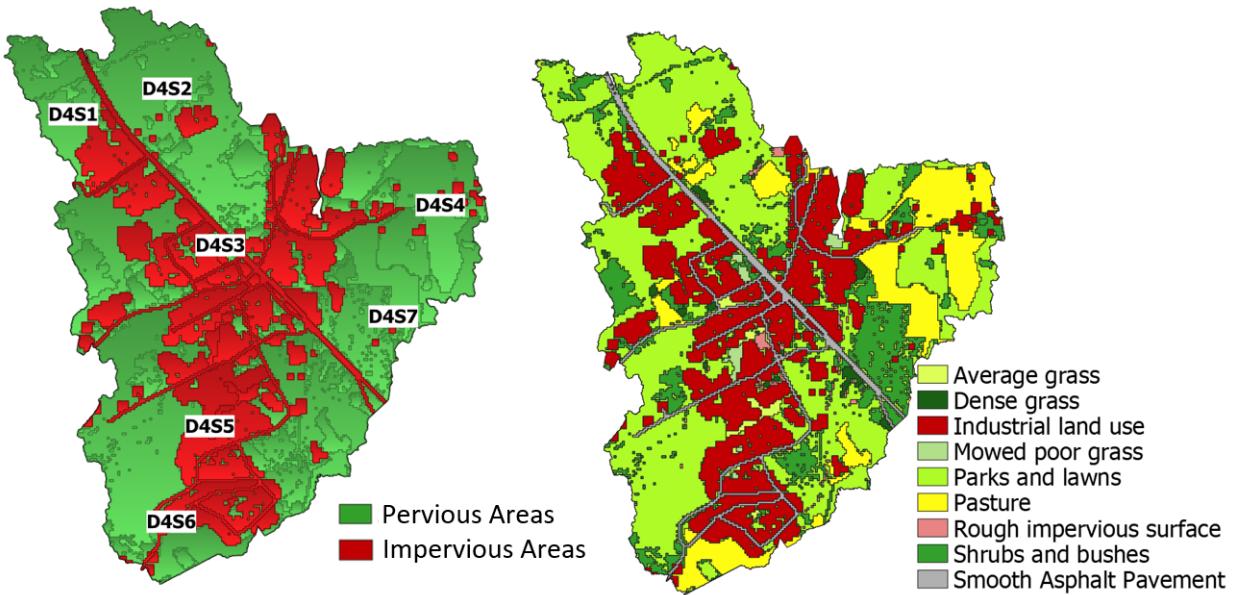


Figure 4.17.: Pervious and Impervious division and land cover.  
Impervious and Pervious areas (left) and level 3 from SYKE's corine land cover for Jokela catchment (right) [41]

to delimit the area concerned Jokela's catchment. Coverage of superficial deposit is depicted in Figure 4.18. Mixed soil types were simplified to facilitate model's parameter estimations.

As explained in Chapter 3, there are four parameters necessary to satisfy the Modified Horton infiltration model: initial infiltration capacity ( $f_0$ ); minimum infiltration capacity ( $f_\infty$ ), decay coefficient ( $k_d$ ), and recovery coefficient ( $k_r$ ) that is calculated based on drying time. For simplicity, the parameters were averaged by the whole area of Jokela catchment without considering subcatchment's divisions per pumping station. Therefore, it is assumed that all subcatchments have the same parameters for the infiltration model. Estimation of parameters was carried as follows:

1. Initial infiltration capacity ( $f_0$ ): values based on the soil type for DRY soils multiplied by 1.2 to account for vegetation present in the catchment [38].
2. Minimum infiltration capacity ( $f_\infty$ ): considered equal to saturated hydraulic conductivity with values estimated by Rawls et al. [35] and available in SWMM user help.
3. Decay coefficient ( $k_d$ ): set to 4 [ $\text{h}^{-1}$ ] [38].
4. Recovery coefficient ( $k_r$ ) in days: Equation 4.1 as function of minimum infiltration capacity in inches/hour.

$$k_r = \frac{3.125}{\sqrt{f_\infty}} \quad (4.1)$$

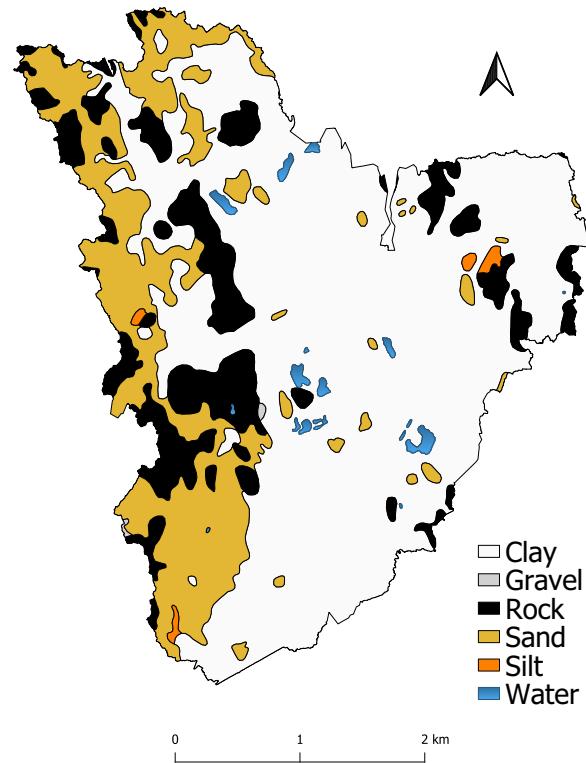


Figure 4.18.: Jokela catchment soil superficial deposits

The estimation of each parameter for each soil type is depicted in Table 4.4. Area and percentage of coverage for each soil type was also calculated to spatially weight  $f_0$ ,  $f_\infty$ , and  $k_r$  using Equation 4.2 to obtain a unique value of each parameter to represent the entire catchment.

$$\text{Parameter} = \sum_n^m (\text{Parameter}_n \cdot \text{Coverage}_n \cdot 0, 1) \quad (4.2)$$

where:

$\text{parameter} = f_0, f_\infty \text{ or } k_r$

$n = \text{soil type (excluding water)}$

Equation 4.2 was applied to the estimated parameters of Table 4.4. The result of  $k_r$  was set to its upper limit range since it was greater than 14 days because of the predominance of soil types with relatively low saturated hydraulic conductivity (*Clay and Rock*). The following unique parameters were obtained and included to SWMM model:

$$f_0 = 52,2 \text{ mm/h}, f_\infty = 24,6 \text{ mm/h}, k_d = 4 \text{ h}^{-1}, k_r = 14 \text{ days}$$

Table 4.4.: Modified Horton infiltration parameter estimation

soil type	area [ha]	coverage [%]	$f_0$	$f_\infty$	$k_r$
Clay	888	66,0	30,5	0,3	29
Sand	268	19,9	152	118	1,50
Rock	182	13,5	3,0	0,03	90,9
Silt	5,90	0,4	91,4	6,5	6,20
Gravel	1,20	0,1	1.524	1.180	0,50
Water	11,5	0,9	-	-	-
Total	1.356	100	-	-	-

#### 4.3.12. Water table and groundwater flow parameters

As mentioned in the literature Bennett et al. [9], Vallabhaneni and Burgess [50], Barden et al. [8], infiltration into the sewer pipes can be caused by seasonal variations in the elevation of the groundwater table or other condition that increased soil moisture content causing a temporary saturated zone. The elevation of the groundwater table around Jokela town was assessed in this section as an attempt to identify a possible correlation with seasonal variations of the water table and the flow increase measured in the SSN.

Information of water table levels was collected from the Finnish Environmental Institute (SYKE) through its open data service [41] and provided by Tuusula Water Utility. Data of three observation wells from SYKE were available surrounding and Jokela town and one from Tuusula Water Utility database within the delineated catchment as depicted in Figure 4.19. The recording period and routines among the four stations vary considerably: from one record per month to one record per year.

Data of measurements from 2004 to 2016 of station 0118651 -located southeast from Jokela- were collected. The years 2007, 2008, and 2017 were left out of the analysis because there were less than six months recorded. All the eleven years records showed an increased groundwater table from March to May.

Only yearly measurements were available for the station 0154356 located around 5 km west from Jokela. The records are from different months, mostly during spring and summer. Therefore, assessment of monthly variation for the same year was not possible. However, the available data suggests slightly higher water table levels on average from January to June for the period of 1999–2017.

Only Station 0118651 and HP 12 had measurements from 2018. This year was relevant to compare with flow measurements available in the sewer network. According to the measurements of these two stations, the water table elevation was higher in the first five months of the year (Jan–May) with an increase during the period from February to May and a decrease, from May to December 2018. Monthly mean of measured groundwater table from all SYKE's station as well as monthly measurements of 2018 from stations 0118651 and HP 12 were plotted in Figure 4.20.

By comparing the two plots in Figure ?? and Figure 4.20, it is possible to observe that a correlation

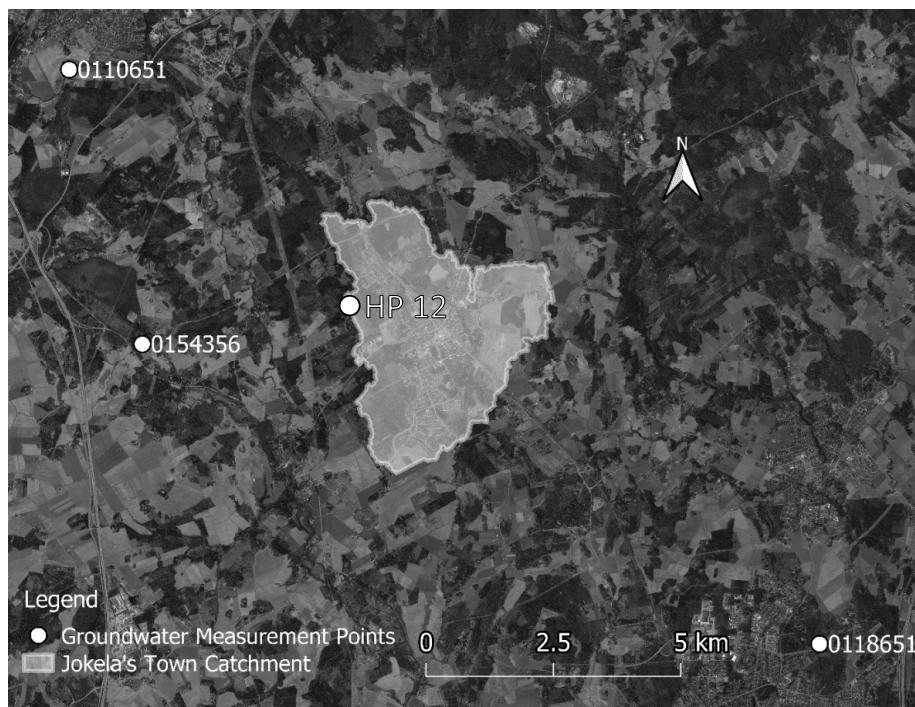


Figure 4.19.: Locations of observation wells around Jokela's catchment

between higher groundwater table and infiltration amount for the first five months into the sanitary sewer of 2018 exists. The estimated amount of RDII peaks in April as the groundwater table elevation measured by the closest observation well to the Jokela's SSN (H 12).

As with all the other data assessed in this study, care should be taken when extrapolating the behavior of the measurements to different years. Groundwater elevation also varies on year by year basis and the same behavior is not always observed. However, the plots in Figure 4.20 presenting the mean values for two different observation wells suggests that an increase of the groundwater table occurs from February to May on average from 2004–2018.

For simplicity, it was assumed that all subcatchments modeled in SWMM share an aquifer with almost all the same parameters. Differences are in the parameters used to describe the groundwater infiltration and the aquifer's bottom elevation. Therefore, the estimation of parameters is divided here in two parts: the aquifer parameters and the groundwater flow parameters. The choice for different groundwater infiltration parameters for each of the subcatchments was made as an attempt to investigate differences in the groundwater infiltration quantity during the calibration. In other words, a higher rate of groundwater infiltration ( $f_G$ ) in one of the subcatchments could suggest that, there are more defects (i.e. pipe cracks) to its network and, therefore, could be candidate for further rehabilitation analysis.

Dupuit-Forcheimer lateral seepage Equation (4.3) was chosen to calculate  $f_G$ . There was no indication that this choice was the best for the study area. The reasons why Dupuit-Forcheimer was chosen

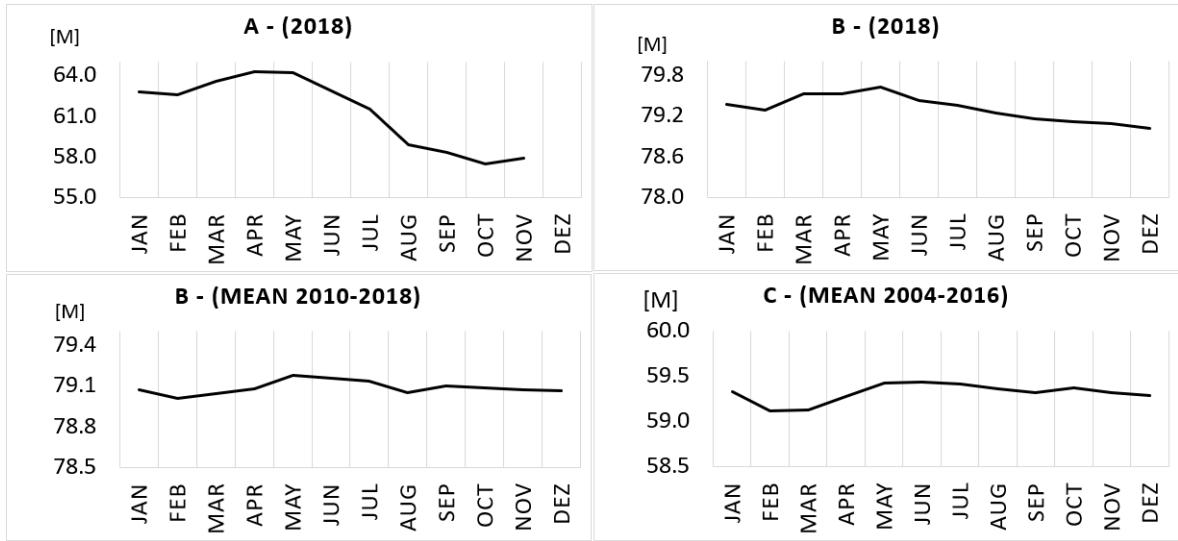


Figure 4.20.: Groundwater table measurements.

Two upper plots are measurements from 2018 only whereas the other two are monthly mean values of the period informed. (A = H12, B = 0110651, C = 0118651). Data from Tuusulan Vesihuolto and SYKE [41]

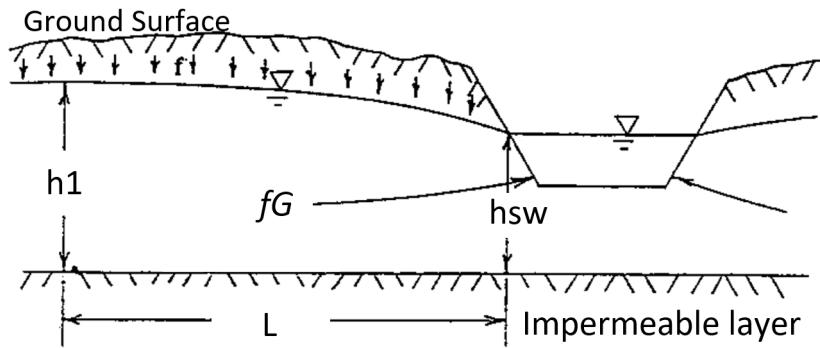


Figure 4.21.: Dupuit-Forcheimer lateral seepage to adjacent channel. Modified from [10]

over the three examples described in in SWMM's hydrology reference manual [38] were: 1. Unlike the linear reservoir method, all parameters of the equation could be estimated based on topographical data before calibration process, 2. Simpler than Hooghoudt's method since the groundwater infiltration flow is evaluated only at one node. Dupuit-Forcheimer seepage scheme is depicted in Figure 4.21.

$$f_G = A_1 \cdot d_L^2 - A_3 \cdot d_L \cdot h_{sw} \quad (4.3)$$

where:

$f_G$  = Flux from saturated zone to receiving node [m/s]

$$A_1 = -A_3 = 2K_s \cdot L^{-2} [\text{s m}]^{-1}$$

$K_s = f_\infty$  = Saturated hydraulic conductivity [m/s]

$L$  = Distance between  $h_1$  and  $h_{sw}$  [m]

$d_L$  = Saturated zone elevation [m]

$h_{sw}$  = Elevation of water inside the receiving node with same reference as the other elevations [m]

It was assumed that the water table elevation distribution is similar to the surface elevation spatial distribution to estimate  $L$  parameter. This is a rough estimate, since there is no indication that the water table elevation follows the surface elevation in the study area. Because of this, the location of highest surface elevation point for each subcatchment was found using available DEM data and QGIS raster calculator. This point was assumed to be  $h_1$  and its distribution is showed in Figure 4.22. The chosen equation is valid only for flows entering the receiving node. Thus, flow occurs only when  $d_L > h_{sw}$ .



Figure 4.22.: Position of  $h_2$  for each subcatchment

The distance between these two points was measured and assigned to  $L$ , for each subcatchment's groundwater component. The saturated hydraulic conductivity ( $K_s$ ) was assumed to be constant for all subcatchments. The  $K_s$  value is the same as the estimated for the modified Horton infiltration method (24,6 mm/h). The  $d_L$  parameter is updated every time step of the simulation as part of the computations. However, its initial value,  $d_{L0}$ , is required at the beginning of the simulation. The  $d_{L0}$  parameter was assume as the the same value of the bottom elevation of each receiving node ( $h_B$ ), as depicted in Table 4.5. The aquifer bottom elevation used as reference for all the other elevations was also set as  $h_B$ , since no information of the impermeable zone was assessed.  $h_{sw}$  value is obtained from hydraulic model flow routing as the wastewater depth plus  $h_B$ .

The aquifer parameters were estimated from soil properties data tables available in the literature ([35],[38]). These tables were used for the infiltration parameter estimation. Equation 4.2 was used to average the values from each subcatchment to the entire catchment area. The parameters and respective estimated values are as follows:

- porosity = 0,41

- field capacity = 0,27
- wilting point = 0,18
- tension slope = 28
- hydraulic conductivity slope (HCO) = 48,6
- fraction of evaporation = 0,35
- lower Evap. Depth = 5 m
- lower GW Loss Rate (seepage) = 1 m/s
- initial unsat. zone moisture = 0,27

As mentioned in Section 4.1, there are streams crossing the delineated catchment area not included in the model. This can be observed when checking the two-aquifer model water budget since interaction between aquifer and surface water (rivers and streams) is not modeled omitting a probably considerable baseflow loss. This loss is represented, in this study, by the seepage to deeper aquifers ( $f_L$ ). The constant rate of seepage is a user-supplied parameter and was used as one of the calibration parameters to simulate long-term groundwater table elevation.

Table 4.5.: Groundwater flow parameter estimation

subcatchment	$h_B$ [m]	L [m]	$A1$ [ $s \cdot m$ ] $^{-1}$
41	75,3	1.108	4,3E-05
42	66,2	2.026	1,3E-05
43	60,0	2.040	1,3E-05
44	62,2	1.397	2,7E-05
45	58,1	1.324	3,0E-05
46	52,2	1.797	1,6E-05
47	64,2	1.319	3,0E-05

Calibration of  $L$  parameter may apply as the  $h_1$  points were located at the borders of the subcatchments using the rough estimation described above. No information of the spatial distribution of the groundwater head was obtained within the subcatchments. Conceptually, the  $L$  distance is closer to the upper limit of the possible range. The calibration would, most likely, move towards decreasing the value of  $L$ . This decrease would result on an increase for  $A1$  values, which increases the groundwater infiltration rate. A flow rate not greater than the observed is expected when using the initial values as suggested in Table 4.5, assuming a constant saturated hydraulic conductivity ( $K_s$ ). Therefore,  $f_G$  is likely closer to its lower limit range and calibration efforts should focus on increasing  $f_G$  by decreasing  $L$ .

First simulation was carried with the values presented in Table 4.5. The result of the first simulation of the groundwater table ( $d_L$ ) from subcatchment 41 is plotted in Figure 4.23. The variation of the simulated groundwater table did not follow completely the pattern observed from well HP 12, even though both simulated and measured data are from the same delineated area (subcatchment 41). There is an overall increase of the water table observed from the simulated values from Jan to May. However, a decrease on  $d_L$  from February to March was simulated, but not existent according to the observed data.

It is also important to note that the average elevation of the bottom of the junctions in Jokela's SSN is higher than the groundwater table measured (Figure 4.20, suggesting that the groundwater infiltration into the SSN is not caused by the water table elevation per se, but by the increase in the soil moisture during aquifer recharge periods, such as snowmelt periods. This is the reason the initial  $d_L$  elevations were set as the bottom elevation of the network's receiving node instead of the measured values. The groundwater elevation results from the model simulations is rather a representation of the increase in soil moisture caused by infiltration than the actual groundwater table levels of the aquifer in the sewershed.

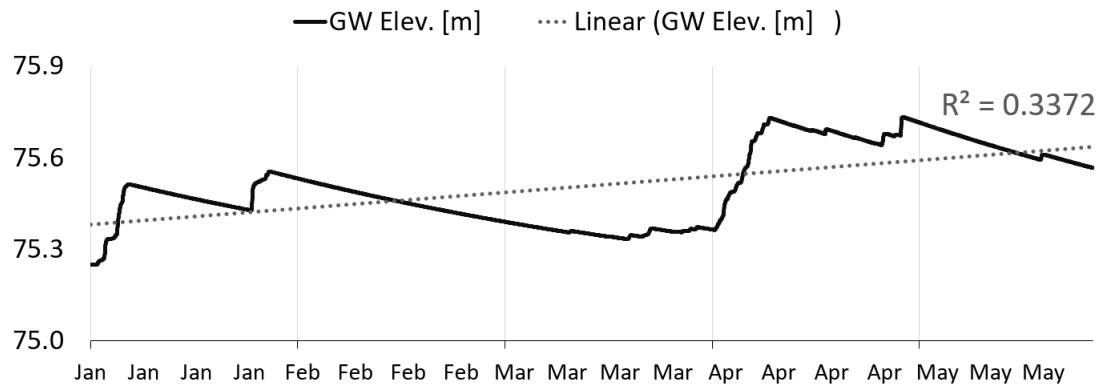


Figure 4.23.: Groundwater table first simulation from January to May of 2018

#### 4.3.13. synthetic unit hydrograph

Three synthetic unit hydrographs of the RTK method were added to each of the seven pumping stations as a representation of the hydrological response of each of the delineated subcatchments as illustrated in Figure 4.25. Other than the required parameters ( $R$ ,  $T$ , and  $K$ ), two out of the three optional parameters ( $D_{max}$  and  $D_{rec}$ ) of the initial abstraction were used in this study. No information of the initial soil moisture for each simulation condition was assessed and, therefore,  $D_0$  was set to zero. As seven subcatchments were modeled, the number of parameters to be estimated per subcatchment was 15:

- $R_{short}, R_{medium}, R_{long}$
- $T_{short}, T_{medium}, T_{long}$

- $K_{short}, K_{medium}, K_{long}$
- $Dmax_{short}, Dmax_{medium}, Dmax_{long}$
- $Drec_{short}, Drec_{medium}, Drec_{long}$

A literature review was carried to find a suitable range for the estimation of parameters. The range, however, may vary considerably according to the type of watershed being studied. As an example, one may expect a different range of parameters between highly urbanized and semi-urban or rural areas.

As stated in [50] R-values are proportional to the sewershed delineated area. As a hypothetical exercise, one can image the differences in the area of the two delineation methods depicted in Figure 4.16. D3 has larger area and, consequently, receives a greater rainfall volume than D1. Therefore, the rainfall volume is proportional to the delineated subcatchment's area since rainfall data from one point in space (rain gauge) is extrapolated for the entire catchment. The RDII amount, however, is not affected by the delineation of sewersheds remaining unchanged regardless the size of the delineated area. Thus, when relating precipitation and RDII amount for an event, different R-values are obtained with different delineated areas - even when the same event is analyzed. For simplicity, the same delineated area used by the physics-based model was considered for the unit hydrograph method (see Figure 4.14). Different range of  $R$  parameters are expected in this study - mostly smaller R-values - when compared to other studies that used the same method, but distinct sewershed delineation often leading to smaller areas of cadastral parcels. The study area 1.355[ha] must be considered - if comparison is desired.

Precipitation input data also influences the estimation of R-values. Therefore, R-values estimation is prone to the limitations of the precipitation measurement method. When precipitation measurements - used for calibration of R-values - fails to capture the realistic precipitation volume, estimated R-values also represent erroneous fractions of RDII. Since the unit hydrograph method can be classified as empirical curve fitting, rain gauge data that systemically underestimate precipitation can also succeed on predicting wet-weather flows with relatively good volume estimate in the sanitary sewer network, since R-values are calibrated. However, rain gauge measurements are also affected by wind speed and characteristics of the site and instrument and can fail to record a rainfall event with same intensity under different wind conditions. Wind speed or soil characteristics are not assessed by the unit hydrograph method in this study.

In summary, R-values are influenced by sewershed area, precipitation input, amount and type of network defects, and catchment's soil characteristics. Although time parameters ( $T$  and  $K$ ) are also influenced by the sewershed's features, they can be less impacted by the volume errors caused by precipitation measurement since they are more affected by the rainfall time distribution, the hyetograph shape. In some cases, different precipitation measurement methods - such as rain gauges and radar - do capture similar rainfall distribution as concluded by (Wride et al. [53]). Thus,  $T$  and  $K$  can be more influenced by the soil, land use, and network characteristics, described by parameters, such as saturated hydraulic conductivity, impervious percentage, and the amount of roof and foundation

drains, that can be assessed to estimate how fast precipitation inflows and infiltrates the sewer when loss methods are included to the model.

Moreover, the RTK unit hydrograph method per se does not account for a snowmelt routine - which is a relevant source of inflow and infiltration in cold climates. This process, if necessary, can be externally simulated and converted to equivalent water amount and combined to the precipitation input. The measured precipitation data used in this study does not explicitly separate snowfall and rainfall. Therefore, its use before any modification could lead to simulations of nonexistent flows in the network considering that snowfall rather accumulates on the soil surface when temperatures are smaller than 0 °C. To overcome this issue during snow season, two extra data sets were used in addition to precipitation. Hourly temperature records and an estimated threshold temperature were used to filter out snowfall from precipitation data classifying all precipitation occurring below 0,5 °C as snowfall that accumulates in the snow pack. Historical measured snow depth data was then used to roughly estimate a snowmelt time series. The reduction of snow depth was converted to equivalent water amount assuming a 1:10 ratio. For each 1 centimeter reduction recorded of the snowpack depth, an 1 millimeter of water amount was generated and added to the filtered precipitation data creating an "effective precipitation" to be used by the unit hydrographs. Figure 4.24 depicts the proposed scheme.

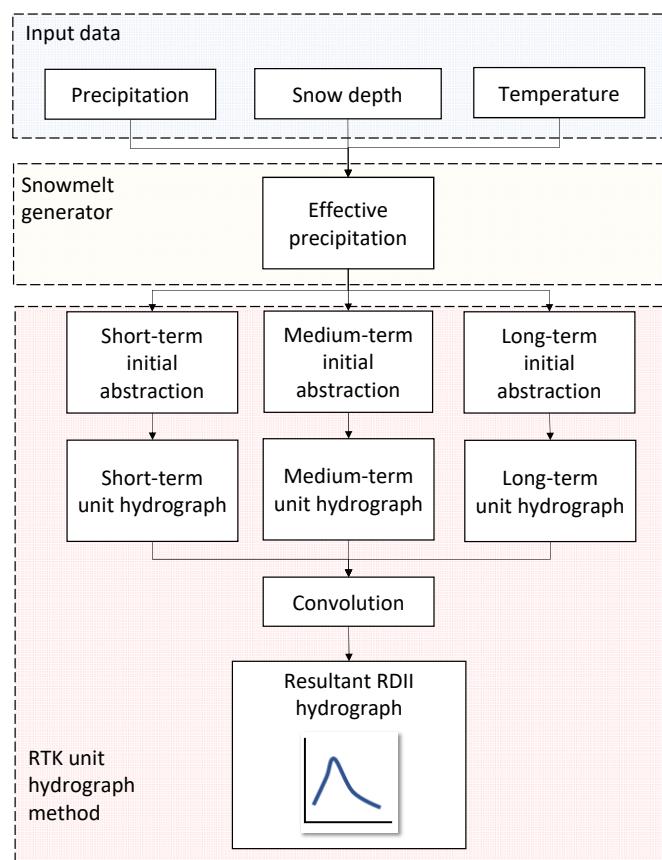


Figure 4.24.: RTK unit hydrograph model scheme

Barden et al. [8] applied RTK unit hydrograph method for continuous simulation of a catchment with residential dominant land use and obtained a range of 4 to 26 % for R-values for 14 months recorded period. The study obtained the best results when applying seasonal RTK-values with monthly initial abstraction (IA) parameters and a well performing model for moderate and large storms with seasonal RTK and IA parameters. These findings highlight the impact of seasonal variations - often divided as dormant and growth season - expected for this study and other sewersheds under cold climate elsewhere. As previously explained, considerable differences on R-values are expected for this study when compared to the literature. Therefore, a method based of available data was used to define a range of R-values and is described later on the following Section. Vallabhaneni and Burgess [50] proposes a range for  $T$  and  $K$  parameters in urban sewersheds depicted in Table 4.6. This range was used by this study for the growth season (from June to November) and as bases for dormant season (from December to May) parameters.

Table 4.6.: Range of T and K parameters by Vallabhaneni and Burgess [50]

Curve	T [h]	K
Short-Term	0.5–2	1–2
Medium-Term	3–5	2–3
Long-Term	5–10	3–7

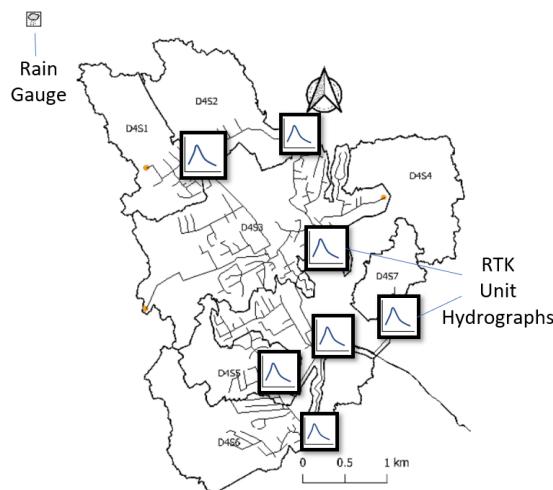


Figure 4.25.: Representation of Jokela's Hydrological Model using RTK UH method

## 5. RESULTS AND DISCUSSION

### 5.1. Calibration

Snowmelt or rainfall events were filtered and chosen for calibration and validation purposes based on divided periods analyzing precipitation and flows in the network system. The year of 2018 was divided into two periods: dormant and growth season. The dormant season was characterized, in this study, by the existence of snow cover, which usually occurs - for the study area - from the beginning of the winter to the beginning of spring. In 2018 - year with available data for calibration - the dormant season was classified from December to May. Winter in Finland varies from 100 days, in the southwest, to 200 days, in the north (Lapland) [21]. Thus, the dormant season accounts for at least 27 % of the year in Finland. Parts of spring, summer and autumn are characterized as growth season which ranged, in 2018, from June to November for the study area. Therefore, dormant and growth season had each 6 months for the study area.

The division of periods was necessary due to the differences encountered in the SSN flows. RDII hydrographs of dormant season occur during longer periods (weeks) than the growth season (days). Higher flows during dormant season were observed for weeks and even months. In 2018, the longest RDII hydrograph started from early April and ended around 25th of May having approximately 22 days (530 hours) of flows higher than 18 L/s - dry-weather average maximum flow. Much shorter RDII hydrographs occur during the growth season, impacting SSN's flows only for hours or days. By comparing of the two data sets, it was possible to observe that RDII winter hydrographs correlates with snowmelt periods, with the latter happening first, which was assumed as the reduction of the snow pack depth. Correlation was also observed for some events of the growth season. The growth season had RDII hydrographs peaking before precipitation was even recorded by the rain gauge and some RDII hydrographs happening without apparent occurrence of precipitation. As briefly described in Section 4.3.4, time shifted rainfall is one of limitations when extrapolating rain gauge point data of the weather station located 10 km away from the catchment that affects the calibration process.

As one of the purposes of the study was to explore hydrological models for an operational system able to provide flow forecasts, relatively short events with low intensity where wet-weather flows were not observed were not considered for calibration purposes. Events were characterized by the start of the rainfall and the end of the RDII hydrograph relative to the previous rainfall. As an example, a 12-hour period RDII hydrograph was observed, starting after 2 h after the beginning of a 4h-rainfall. Thus, the event duration was considered to be 14 h (12 h + 2 h). Events with volume lower than 15 millimetres and duration inferior to 6 h were not considered. The EPA SSOAP toolbox automatic routine for event determination was used to filter the events of interest.

Table 5.1 depicts the nine events of 2018 where the first three events regard the dormant season and the remaining six events occurred during the growth season. Four events of 2019 were used for validation of the dormant season. Events 4, 5, and 6 of 2018 were used for calibration of growth season and 7, 8 and 9 for validation. Furthermore, validation of the calibrated models was also carried with weather forecast data from the HARMONIE model.

Table 5.1.: Events used for calibration and validation.

Event	Type	Start and End Time [m/d - h]	Event Duration [h]	Precip. [h]	Precip. [mm]	Precip. Peak(s) [mm/h]	RDII [mm]	Peak Flow(s) [L/s]
2018								
1	C	01/1 - 00	24/1 - 23	576	92	58,3 ; 86,1*	***	2,46
2	C	25/1 - 00	18/2 - 22	599	49	33,5 ; 58,2*	***	1,22
3	C	04/4 - 20	26/5 - 03	1.232	43	42,6 ; 77,3*	***	3,92**
4	C	21/6 - 20	22/6 - 10	15	4	17,4	7	0,04
5	C	02/7 - 23	04/7 - 17	42	10	18,1	7 ; 3	0,07
6	C	04/7 - 20	07/7 - 10	62	6	38,2	21	0,21
7	V	20/8 - 17	21/8 - 13	20	5	30,3	9,6	0,04
8	V	11/9 - 17	14/9 - 00	55	8	18,2	5,5	0,08
9	V	15/9 - 08	19/9 - 23	111	11	21,8	5,4 ; 3,3	0,09
2019								
10	V	10/2 - 06	14/2 - 10	100	20	8,5*	***	0,38
11	V	15/2 - 02	18/2 - 21	91	-	-*	***	0,14
12	V	18/2 - 22	22/2 - 21	95	6	3,2*	***	0,25
13	V	10/3 - 13	24/4 - 11	1.078	54	49,6*	***	4,13

C = Calibration, V = Validation. Peak flow(s) of total recorded SSN flow (DWF + RDII).

\* Effective precipitation used for the unit hydrograph method.

\*\* Underestimated due to 6 days of missing data.

\*\*\* Several events of low intensity recorded.

Simulation results were compared with measured flows and evaluated through goodness-of-fit measures adopted from the literature [25] and described in Table 5.2. No evaluation of the goodness-of-fit tests was performed in this study.

### 5.1.1. Physics-based model calibration

Calibrations of the physics-based model were also divided in two periods: dormant and growth seasons. Long-term simulations were done for the dormant season - when snow cover was present - due to long events (from 575 to 1.240 hours) and time influenced snowmelt coefficient parameters ( $DHM_{max}$ ,  $DHM_{min}$ ). The period simulated was from January to May. The long-term simulation was used to investigate the model performance when parameters are constant. Three major RDII hydrographs were observed during the chosen period, events 1, 2 and 3 from Table 5.1. Each of them corresponding to major snowmelt events of the year 2018 with the latter associated to the complete

Table 5.2.: Goodness-of-fit tests. Adapted from [25]

goodness-of-fit tests	equation	interpretation
NSE	$1 - \frac{\sum_{i=1}^n (Q_{obs,i} - Q_{sim,i})^2}{\sum_{i=1}^n (Q_{obs,i} - Q_{m,obs})^2}$	values range from $-\infty$ to 1. Value 1 represents a perfect model. The value 0 means that the model performance is as accurate as observed mean over the observed data set. Values less means that observed mean is better than the simulated model.
absolute volume error	$\left[ 1 - \frac{\sum_{i=1}^n Q_{sim,i}}{\sum_{i=1}^n Q_{obs,i}} \right] \cdot 100$	Values range from 0 to $+\infty$ . The best fit receives the value 0. Plus and minus signs are added in front of the value and represent, respectively, overestimation and underestimation
peak flow rate absolute error	$\left[ 1 - \frac{Q_{sim,max}}{Q_{obs,max}} \right] \cdot 100$	

where  $n$  is the total number of data points;  $Q_{obs,i}$  and  $Q_{sim,i}$  are the  $i$ th flow value;  $Q_{m,obs}$  and  $Q_{m,sim}$  are the mean of the observed and simulated flow;  $Q_{obs,max}$  and  $Q_{obs,min}$  are the maximum and minimum value in the data set;

melt of the snowpack. There were six days with missing flow records from the start of event 3 and were, therefore, not considered for simulations evaluations.

Events of the growth season were calibrated with shorter simulations due to also shorter event duration. Events 5 and 6 were considered to be the same as an attempt to calibrate inflow parameters for two distinct events in one simulation. Table 5.3 depicts the parameters chosen for calibration of each module and their calibrated values for each calibration method and event.

Manual calibration was carried for both periods by modifying sensitive parameters of each module, such as snowmelt coefficients ( $DHM_{max}$  and  $DHM_{min}$ ), aquifer's saturated hydraulic conductivity ( $K_s$ ), seepage rate, conductivity slope ( $HCO$ ), and groundwater flow coefficient  $A1$ . The percent routed ( $\%routed$ ), and flow width were the calibrated parameters of the runoff routine.

Table 5.3.: Calibrated parameters for physics-based model

season	runoff		aquifer			A1	snowpack		NSE	volume error [%]
	%routed	width	Seepage	$K_s$	$HCO$		Dmax	Dmin		
dormant	94	10 %*	10	24,6	37	122 %*	0,030	0,015	0,59	-12,4
									0,87	+4,70
growth	98,35	10 %*	450	450	26	122 %*	-	-	0,68	+2,86
									0,56	+21,7
									0,89	-0,25

\* Percent reduced or increased from previously estimated value (i.e. subcatchment 45 had the width reduced from 240 to 24 and A1 increased from 0,000030 to 0,00065). The change was applied to all subcatchments. Previous values available in Chapter 4.

\*\* Multiple peaks. Period of complete snow melt.

The runoff block parameter  $\%routed$  was calibrated as described in Section 3.2.1 to simulate losses caused by streams and stormwater sewer network not included in the model. The percentage routed ( $\%routed$ ) diverts the precipitation volume to the pervious subareas which are on average 71 % of

the catchment's area. The precipitation and snowmelt amount diverted to the pervious subarea is then available for infiltration process which, consequently, varies the groundwater table.  $\%routed$  can be related to R-values of the unit hydrograph method as the ratio between precipitation and RDII amount.

The groundwater inflow was, firstly, abruptly interrupted by SWMM routine which sets the groundwater inflow to zero, since the water table drops below the wastewater level in the receiving node (Jokela Storage Unit). SWMM uses this routine whenever the groundwater flow parameter  $A3$  value is set to different than zero. To avoid sudden drops on the groundwater inflow component,  $A3$  parameter was set to zero. This simplifies the firstly proposed Dupuit-Forcheimer equation 4.3 by excluding the second component resulting in equation 5.1.

$$f_G = A_1 \cdot d_L^2 \quad (5.1)$$

Results of the events used for parameter calibration of the physics-based model are depicted in Table 5.4. Volume, flow peak amount and timing were also evaluated.

Table 5.4.: Evaluation of events used for calibration of physics-based model parameters

<b>season</b>	<b>event(s)</b>	<b>NSE</b>	<b>volume error [%]</b>	<b>peak error [%]</b>	<b>peak shifted [h]</b>
dormant	1	0,63	-16,5	-14,0	0
	2	0,89	-1,38	-6,96	0
	3	0,68	-16,84	**	**
growth	4	0,56	+21,70	+0,18	-1
	5; 6	0,89	-0,25	-0,11	+1

\*\* Multiple peaks. Period of complete snow melt.

### 5.1.2. Unit hydrograph model calibration

Unit hydrograph method was calibrated using the dynamically dimensioned search (DDS) algorithm. One unit hydrograph was assigned to each of the subcatchments (seven in total). The division of periods with snow cover and without was considered. Each unit hydrograph had 15 parameters to be calibrated, R, T, K, Dmax, Drec for each of the three terms. There were 105 different parameters in total to be calibrated. Therefore, a range of parameters had to be defined to constrain the optimization algorithm search space. The ranges of R-values were estimated using available data and RDII amount estimated. The time parameters (T-values and K-values) were estimated by analyzing RDII hydrographs time length.

R-values were estimated by comparing RDII amount of each event and its precipitation amount. This ratio was assumed to be an estimate of the R-total ( $R_{short} + R_{medium} + R_{long}$ ) and represent the fraction of the precipitation that infiltrated the sanitary sewer network (SSN). An average of the ratio

was taken with all events for each period. During snow cover (dormant season) an estimated average of 3,34 % of effective precipitation infiltrated the network and the growth season ratio estimation was of 0,36 %, one order of magnitude smaller than the estimates for the dormant season. Therefore, R-total of each of the seven resultant RDII hydrographs was constrained to a limited offset from the estimated ratio for each period.

The range of possible T-values for the optimization algorithm were estimated based on the event's total duration time. The shortest and longest event's duration was defined as the T-total lower and upper limits. Therefore, the lower limit for the dormant season was set to 576 h, same duration as the first event, and the upper limit for the same period was set as 1.232 h, same as event 3. The same assumption was used for the growth season. K-values range for the growth season were assumed to be the same as suggested by Vallabhaneni and Burgess [50] and doubled values for the dormant season.

No soil or land cover data was assessed for the initial abstraction parameters (IA). For simplicity,  $D_{rec}$  and  $D_{max}$  were assumed to have the same range varying from 0,5 mm to 20 mm and gradually distributed for all unit hydrographs. R-total, T-total estimates were then divided among the short, medium and long-term hydrographs with the following assumptions:

dormant season:

- The upper and lower limits of  $R_{total}$  were calculated as function of the average  $R_{total}$ , 0,334, estimated from the data set. It was assumed that the upper limit can be three times greater than the average and the lower bound three times smaller.
- To further divide the total response ratio to all the three unit hydrographs the  $R_{medium}$  and  $R_{long}$  were assumed to account for 90 % of  $R_{total}$  for the dormant season with 45 % each.
- The upper and lower limits of  $T_{total}$  were assumed as the same as the longest and shortest event duration.
- To further divide the  $T_{total}$  range as bounds for the three unit hydrographs the following was assumed: short-term hydrograph impacts up to 2 % of the total event's duration, medium-term hydrograph 30 % and long-term hydrograph can occur for 150 % of the entire event duration.

growth season:

- The upper and lower limits of  $R_{total}$  were calculated as function of the average  $R_{total}$ , 0,036, estimated from the data set. It was assumed that the upper limit can be six times greater than the average and the lower bound six times smaller.
- To further divide the total response ratio to all the three unit hydrographs the  $R_{short}$  and  $R_{medium}$  were assumed to account for 90 % of  $R_{total}$  for the growth season with 45 % each.
- The upper and lower limits of  $T_{total}$  were assumed as the same as the longest and shortest event duration.

- To further divide the  $T_{total}$  range as bounds for the three unit hydrographs the following was assumed: Short-term hydrograph impacts up to 6 % of the total event's duration, medium-term hydrograph 20 % and long-term hydrograph can occur for 150 % of the entire event duration.

Equation 5.2 was used to calculate T-values based on K-values and event's duration. For instance, the upper limit of the  $T_{short}$  value of the dormant period was estimated based on event 3 - longest event duration of the season with 1.232 hours - and the assumption that a short response hydrograph accounts for only 2 % of the total duration time if the event, which is about 24,6 hours. The time estimated (24,6 h) is then added to equation 5.2 to calculate the upper bound of  $T_{short}$ . In other words, when the upper bound of  $T_{short}$  and  $K_{short}$  is chosen, the maximum short-term unit hydrograph duration equals 24,6 hours.

$$T = \frac{T_E}{1 + K} \quad (5.2)$$

Where:

$T_E$  = estimated time [h]

The lower limit of all  $T_{medium}$  was set as the upper bound of  $T_{short}$  range and  $T_{long}$  lower bounds were, similarly, constrained by  $T_{medium}$  range. This assures  $T_{short} < T_{medium} < T_{long}$  for consistency as suggested by [50]. The EPA SSOAP help file also suggest constant T and K-values for the two seasons. However, preliminary analysis of RDII hydrographs showed considerable differences on the ranges T and K-values for the dormant and growth season and were, therefore, separated by season.

Table 5.5 depicts the proposed range of parameters limiting the search space for the dynamically dimensioned search (DDS) algorithm. In summary, ranges for  $R_{total}$  and  $T_{total}$  were estimated based on effective precipitation data and RDII hydrographs estimated from flow meter data. Further division for short, medium and long-term hydrographs were estimated on assumed ratios used to distribute  $R_{total}$  and  $T_{total}$  among the R and T-values of all three unit hydrographs. Ranges for K-values were estimated based on values suggested in the literature. IA parameter's chosen assumes a relatively wider range for losses for interception and/or surface wetting in comparison to the physics-based model parameters ( $Ds_{imp}$ , and  $Ds_{perv}$ ).

It is important to highlight that the method used to estimate the range of parameters is empirical and with a set of assumptions as described previously. There is no evidence that the chosen method is suitable for other catchments elsewhere. The main reason to define a range based on available data and preset factors is to automatize the calibration routine with ranges capable of self-adjustment with time.

Results of the events used for parameter calibration of the unit hydrograph method are depicted in Table 5.6. Shift on time to peak was also evaluated.

Table 5.7 depicts the results evaluated for each number of iterations chosen for the DDS optimization algorithm. Flows were evaluated on the downstream most pumping station (PS1) located in subcatchment 44. Flows of the pumping station 2 (PS2), located in subcatchment 43, were evaluated

Table 5.5.: Range of unit hydrograph method parameters for optimization algorithm

<b>unit hydrograph</b>	<b>parameter</b>	<b>dormant season</b>		<b>growth season</b>	
		lower	upper	lower	upper
short	R	0,0011	0,0100	0,0003	0,0097
	T	3,84	4,928	0,45	2,22
	K	2	4	1	2
	Drec	0,5	5	0,5	5
	Dmax	0,5	5	0,5	5
medium	R	0,0050	0,0451	0,0003	0,0097
	T	4,928	52,8	2,22	5,55
	K	4	6	2	3
	Drec	1	10	1	10
	Dmax	1	10	1	10
long	R	0,0050	0,0451	0,00006	0,0022
	T	52,8	123,2	5,55	20,81
	K	6	14	3	7
	Drec	2	20	2	20
	Dmax	2	20	2	20
		Rtotal	0,0006	0,0808	0,0111
					0,1000

Table 5.6.: Evaluation of events used for calibration of unit hydrograph model parameters

<b>season</b>	<b>event(s)</b>	<b>R-total</b>	<b>NSE</b>	<b>volume error [%]</b>	<b>peak error [%]</b>	<b>peak shifted [h]</b>
dormant	2	0,0500*	0,85	+7,37	+1,42	0
	3		0,65	-9,89	**	**
growth	4	0,00585*	0,84	+11,2	+3,59	-1
	5, 6		0,83	+3,02	+0,81	0

\* Sum of the average of R-values from all subcatchments. Individual values of all parameters available in the appendix.

\*\* Multiple peaks. Period of complete snow melt.

as for reference, but not included in the calibration process. All events were calibrated using, at least, three runs of 600 iterations. The calibrated parameters with the best fit considering the absolute volume error, peak flow rate absolute error and NSE among the three runs were chosen. For the event 5, 6, the accuracy of the simulated model, evaluated by the NSE, proportionally rose with the number of iterations up to 1.000. No significant number of iterations and optimization runs were carried to evaluate the performance of the DDS algorithm. However, it was observed that errors on volume and peak estimates should be included as a multi-objective function for the evaluation routine of the DDS algorithm since calibrations with better NSE values did not represent more accurate volume or peak estimation.

The observed flow from the pumping station PS2 was not included in the calibration process. Therefore, the values of NSE for PS2 presented in Table 5.7 were paired with the simulation results of PS1. The values are expressed to illustrate that it was possible to obtain similar accuracy for the upstream points even without considering them for the calibration. This approach can be used, for example, to estimate the RDII for upstream, ungauged points of the network.

Table 5.7.: Event 5 and 6 simulation results evaluation for the number of iterations of the DDS optimization algorithm

Number of Iterations	NSE	
	PS1	PS2
100	0,82	0,81
300	0,85	0,85
600	0,86	0,86
1000	0,88	0,79

## 5.2. Model validation with historical data

Historical precipitation data recorded by the rain gauge to validate both models and events forecasted by the HARMONIE model were used to preliminary assess the use of the models to predict flows in the sanitary sewer system. Seven events were used for the validation with historical precipitation data where three events (7, 8 and 9) of the growth season of 2018, and four events (10, 11, 12 and 13) of the dormant season of 2019. Three low intensity rainfall events during the transition from dormant to growth season were used to assess the model response with precipitation predicted by the HARMONIE model.

The events chosen based on historical precipitation and flow data are depicted in Table 5.1. Events used to evaluate the model performance using the forecasted precipitation were defined based only on observed flows. The limited amount of forecasted data used in this study allowed only a preliminary analysis of the model performance.

Figure 5.1 shows the calibration and validation result of the long-term simulation of the dormant

season using the physics-based model. Events 10, 11, 12 and 13 were simulated using the estimated and calibrated parameters described in the previous sections. The results of calibration and validation for the same period using the unit hydrograph method are depicted in Figure 5.2.

The calibrations of the growth season were carried for each event. Therefore, for simplicity, only the validation is presented as a graphic result for both models in Figure 5.3. Table 5.8 depicts the evaluation of the long-term validation of both models and Table 5.9 presents an event-based evaluation of the validations.

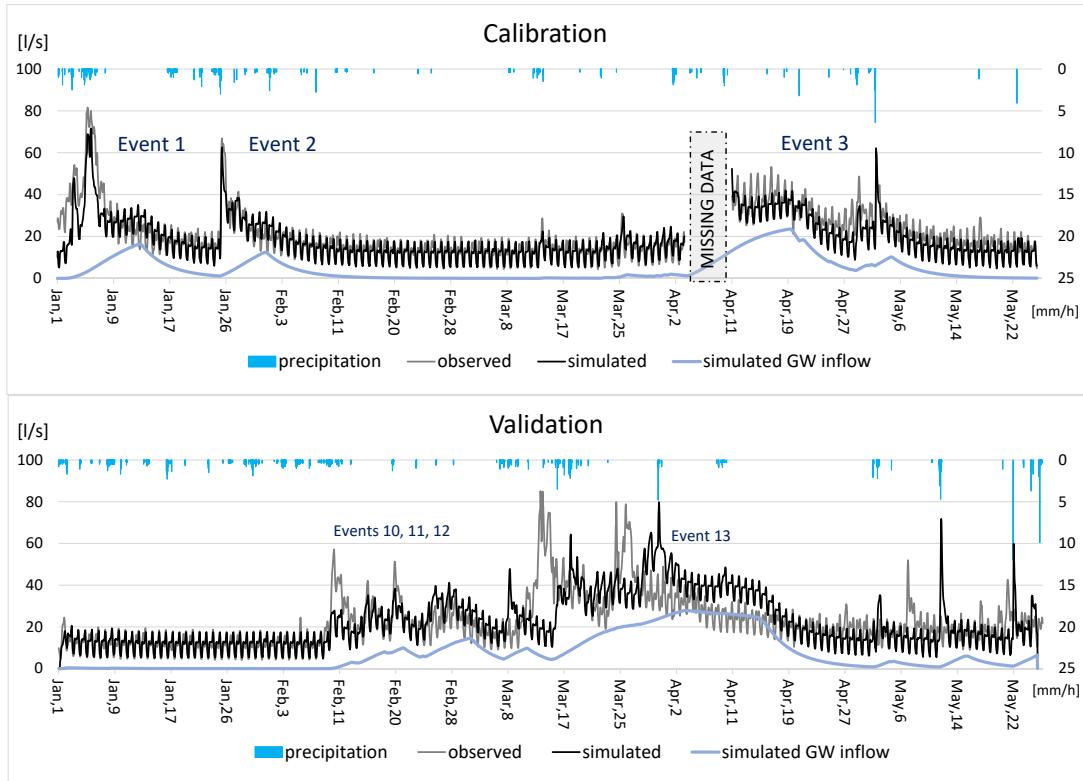


Figure 5.1.: Results of calibration and validation of the physics-based model for the dormant season

For the dormant season, the physics-based model outperformed the unit hydrograph model according to the chosen evaluation criteria. Both models received negative NSE values when simulating the flows of the complete snowmelt period (Event 13). The peak flows of event 13 were delayed in about five days when using the physics-based model. The unit hydrograph model failed to capture the magnitude of the peak flows during event 13 and, therefore, underestimated the volume by 8,4 %. It is important to highlight that the input "effective precipitation" (also called "net precipitation") containing both snowmelt and rainfall are different between the two models. The underestimation of the volume can be caused by the simplistic method used to quantify the snowmelt based on the snow depth measurements used as input to the unit hydrograph method. Therefore, changes of the snow pack specific gravity are neglected. The snowpack and snowmelt routines of the physics-based model

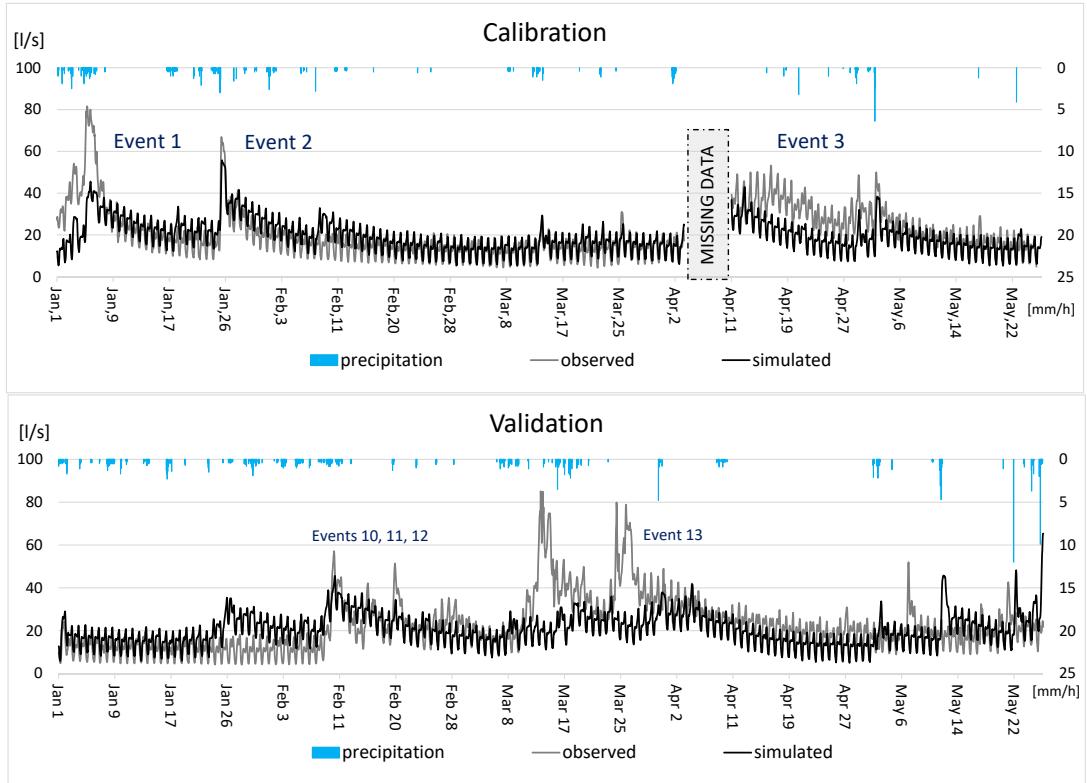


Figure 5.2.: Results of calibration and validation of the unit hydrograph model for the dormant season

computes the amount of water depth equivalent to avoid estimating the specific gravity of the snow pack which changes throughout the winter.

Figure 5.4 depicts the simulated flows during and after the complete melt of the snow pack should a shift of five days on the snowmelt have occurred. This curve cannot be considered as a valid result since it was manually shifted, but it suggests that a more accurate (NSE 0,88) flow estimation could be obtained using the physics-based model. Snowmelt parameters and the temperature input data are the two hypothesis explored here as the causes for the delayed flows observed during the complete melt of the snow pack in 2019:

1. The first hypothesis suggests that there are seasonal variations of the snowmelt parameters. These parameters could be dynamically adjusted as function of predicted weather conditions. This could overcome the time shift error. The maximum snowmelt rate ( $DHM_{max}$ ) and/or the threshold temperature for snowmelt to occur ( $T_{base}$ ) are some of the parameters that could be adjusted to shift the snowmelt in time. Solar radiation, which was not explicitly estimated in this study, is an example of a possible weather condition that can be considered to update the snowmelt parameters or the cold content, which is used in SWMM's heat transfer calculations. It is possible to identify that the snowmelt occurs in accordance with the observed flow data,

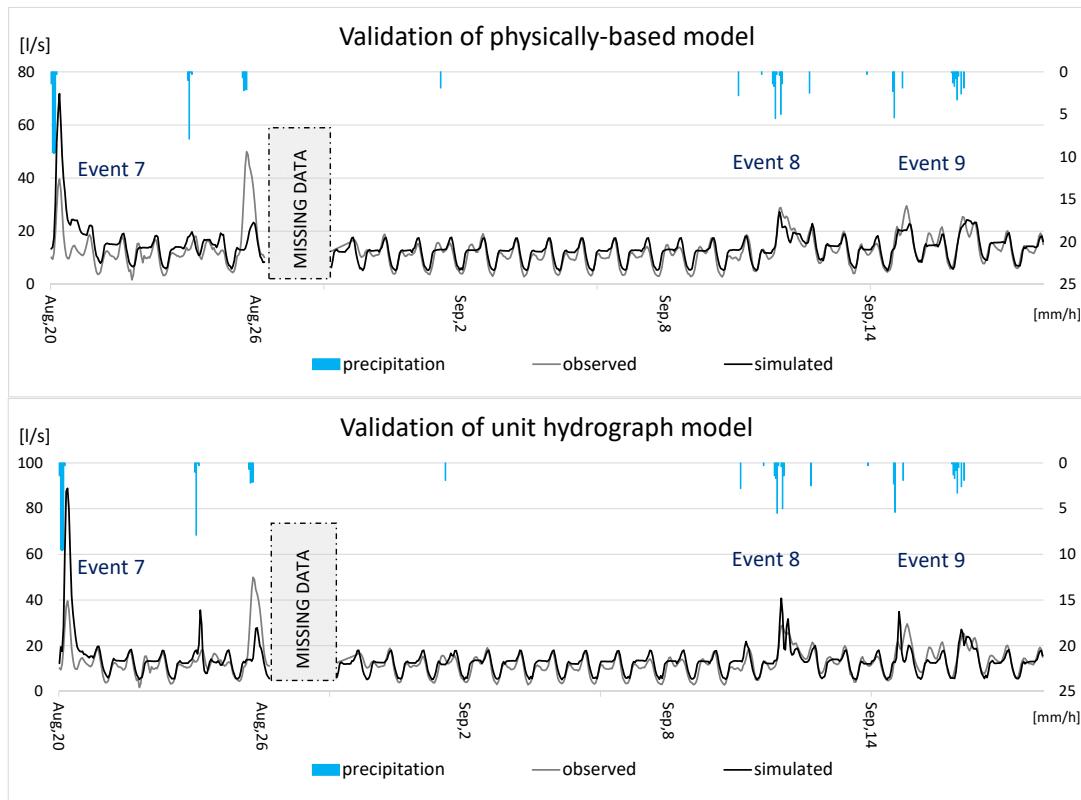


Figure 5.3.: Validation of the growth season

after manually shifting the simulated flows (see Figure 5.4) The groundwater infiltration rate fluctuates in apparent accordance to the total observed flow. Therefore, it suggests that no adjustment is required for the infiltration, aquifer, and groundwater flow parameters. However, even after manually shifting the snowmelt, errors on the estimated flows occurred, such as the underestimation of the first peak flow observed during event 13 by an absolute peak flow error (APE1) of -32,3 %.

2. The temperature input data can be the other factor influencing the time when the snowmelt occurs. The temperature measurements and snow pack depth, used for calibration and validation process, were collected from the weather station located about 10 km northwest from the catchment on a fairly open garden [19]. Therefore, the measurements were not collected in Jokela's urban area and may be also influencing the delay of the snowmelt observed from the simulation results. The snowmelt time of the simulation is in accordance with the measured reduction of the snow pack depth as depicted in Figure 5.5. This suggests that the snowmelt parameters, previously calibrated, succeeded on estimating, with relative accuracy, the time of snowmelt for the location where the measurements were taken. If true, no adjustments on the snow melt parameters, as suggested by the first hypothesis, would be necessary. However, this cannot be confirmed as no measurements of the snow pack were available for Jokela's town.

Table 5.8.: Season-based validation results of proposed hydrological models

model	season	event	NSE	volume error [%]
physics-based unit hydrograph	dormant	10, 11, 12; 13	0,20	+1,53
	growth	7, 8, 9	0,40	+8,94
	dormant	10, 11, 12, 13	0,12	-8,40
	growth	7, 8, 9	0,18	+6,51

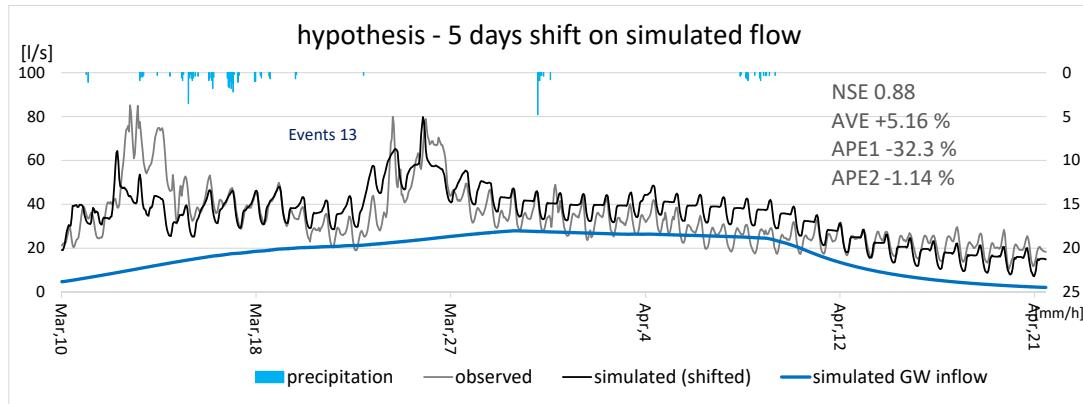


Figure 5.4.: Hypothetical shift on simulated flows. Results of physics-based model manually advanced in five days

Both models performed similarly for the growth season in relation to the total volume error and better results were obtained with the physics-based model regarding the NSE evaluation as depicted in Table 5.8. Both models overestimated the flows of event 7 by more than 45 % of volume and peak discharge. Event 8 was better estimated by the physics-based model with less than 1 % volume error with 6 % error of the peak discharge, and exact match of peak timing. The first peak observed of event 9 was underestimated by the physics-based model and advanced by the unit hydrograph simulation. Similar performance was observed with both models during the second peak of event 9.

Table 5.9.: Event-based validation results of proposed hydrological models

<b>model</b>	<b>season</b>	<b>event</b>	<b>NSE</b>	<b>Volume Error [%]</b>	<b>Peak Error[%]</b>	<b>Peak shifted [h]</b>
growth		7	-1,79	+45,3	+46,0	0
		8	0,64	+0,688	-5,93	0
		9	0,55	-4,02	-40,92 ; -8,436	+3 ; +3
- based	dormant	10	-0,84	-65,3	-105	+3
		11	0,64	-0,72	-19,7	+2
		12	0,58	-9,87	-34,0	+3
	growth	13	-0,72	+5,90	-32,3 ; +1,14	+112 ; +122
		7	-4,24	+48,2	+56,2	0
		8	0,44	-4,74	+28,5	-1
unit hydrograph	dormant	9	0,50	-10,6	+15,2 ; +6,04	-7 ; -2
		10	0,55	-3,77	-32,0	+1
		11	0,51	-3,52	-30,2	+2
		12	0,14	-9,87	-53,2	+2
		13	-0,58	-52,3	*	*

\* Simulation did not reproduce the two flow peaks observed

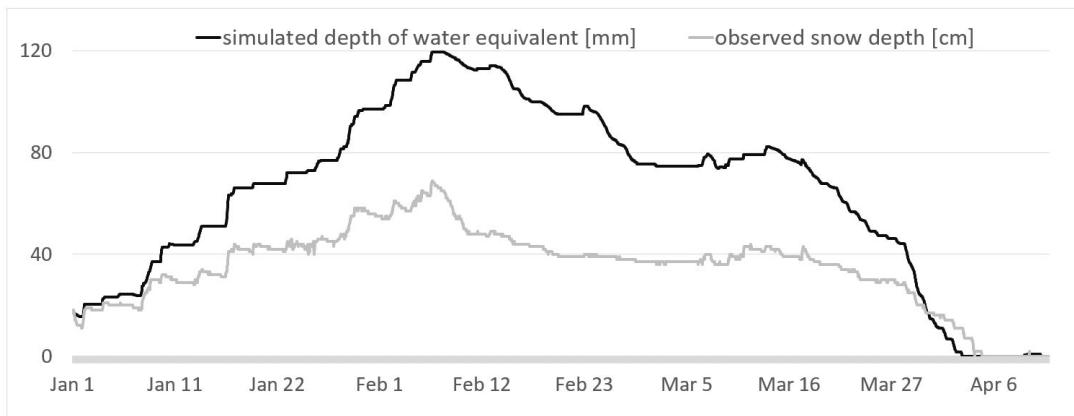


Figure 5.5.: Observed snow depth versus simulated depth of water equivalent of 2019

### 5.3. Seasonal water balance

The volume produced by simulations of the dormant period of 2019 and the growth period of 2018 were compared with the volume estimated based on the available data using the SSOAP toolbox. Table 5.10 depicts the wastewater balance separated by rainfall-derived infiltration and inflow (RDII) and dry-weather flow (DWF) with a percent of RDII in relation to the precipitation amount of the period. The dormant period received 270 mm of precipitation, being this real number somewhat larger due to the unknown snow catch deficiency of the rain gauge. A total of 288 mm precipitation was recorded for the growth period. The RDII results of the unit hydrograph and physics-based model simulations were separated from the total flow by subtracting the simulated DWF. The RDII amount of the unit hydrograph was defined by the R-values. The RDII regarding the physics-based model can be further divided in the portion relative to the groundwater infiltration and the runoff amount as depicted in Figure 5.6.

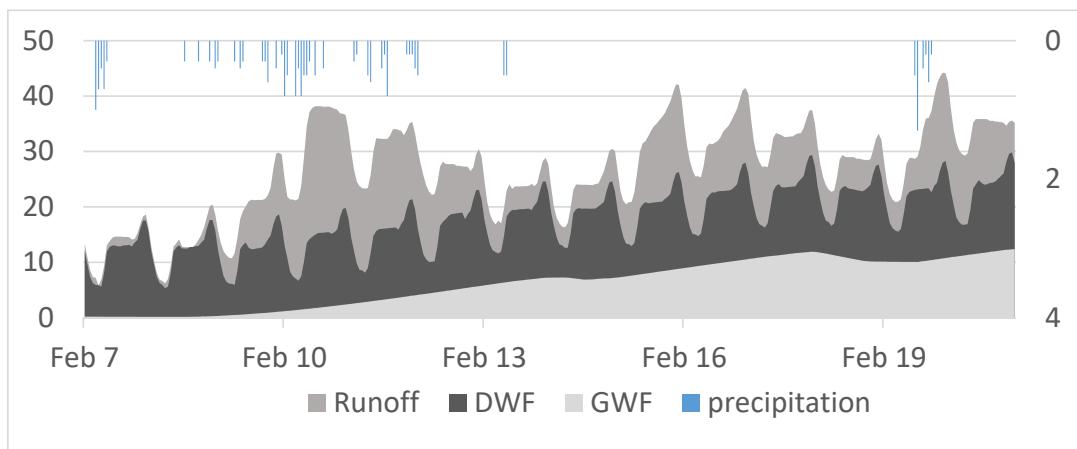


Figure 5.6.: Different simulated flow components of 2019 dormant season using the physics-based model

Table 5.10.: Wastewater balance of validation simulations versus data-based estimated

season	flow component	simulated [m <sup>3</sup> ]		estimated [m <sup>3</sup> ]	volume error [%]	
		PB	UH		PB	UH
dormant	DWF	151.003	149.165	152.344	-0,9	-2
	RDII	135.746	108.874	135.530	+0,2	-20
	total flow	286.749	258.040	287.874	-0,4	-10
growth	DWF	214.996	218.323	216.010	-0,5	+1
	RDII	19.390	13.930	17.499	+11	-20
	total	233.386	232.253	214.160	+9	+8

DWF = base waste flow + constant groundwater infiltration

PB = physics based model ; UH = unit hydrograph model

dormant season from 1st of January to 31st of May 2019.

growth season from 1st of June to 31st of November 2018.

#### 5.4. Simulations with forecast data

The three events defined (EF1, EF2, EF3) for a preliminary analysis of the model performance using forecasted precipitation as input are depicted in Figure 5.7. The precipitation forecast was updated 12 h before the observed peak flow of event EF1. No precipitation was predicted and, therefore, no response was observed from the hydrological models. The rain gauge, located around 10 km from the study site, also did not capture any rainfall. After the peak flow a low intensity rainfall was predicted with a total of 3,9 mm for event's EF1 period, not sufficient to overcome the initial loss of the models and simulate wet-weather flows.

The precipitation forecast was updated 9 h before the greatest peak flow rate of Event EF2 and 3 h after. Zero rainfall was forecasted. The rain gauge recorded a 12 mm/h intensity rainfall 12 h after the peak discharge. This was probably the rainfall which caused the peak flow, but since no rain was predicted, no wet-weather response was observed from the simulations. A forecast update of 24th of May predicted a rainfall with its peak happening 23 h later. Although certain underestimation occurred, the forecast, made 23 h in advance, matched the exact hour of the rainfall peak provoking a response of the hydrological models.

The rain gauge records are used as a comparison only for the rainfall timing. They have different measurement units. The rain gauge was measured with rainfall intensity [mm/h] and the forecast data as incremental volume [mm] for the forecast.

The models used the parameters calibrated for the growth season since, according to the measurements, no snow cover was present after 6th of May, 2019. However, events EF1 and EF2, evaluated with forecasted precipitation, occurred during the transition time from the dormant to the growth season and both underestimate the observed flows in the sanitary sewer. According to the data of 2018 and 2019, the groundwater infiltration rate can remain higher for more than two months after

## *5. Results and discussion*

the complete snow melt. This suggest that the transition between dormant and growth season can be a rather long period. The approach used in this study, for both models, underestimate the flows during the transition period.

In case no hot start file or long enough simulations to preserve the previous hydrological states of the sewershed, the long recession curve of the groundwater infiltration of the transition period could be simulated by adjusting the initial water table level and the seepage rate of the aquifer module of the physics-based model. For the unit hydrograph model, this can be simulated as an external inflow that slowly diminishes until the beginning of growth season. No further simulations of the transition period was carried in this study, but simulations of the dormant period with different initial water table elevations of the physics-based model were carried for a preliminary assessment for continuous simulation.

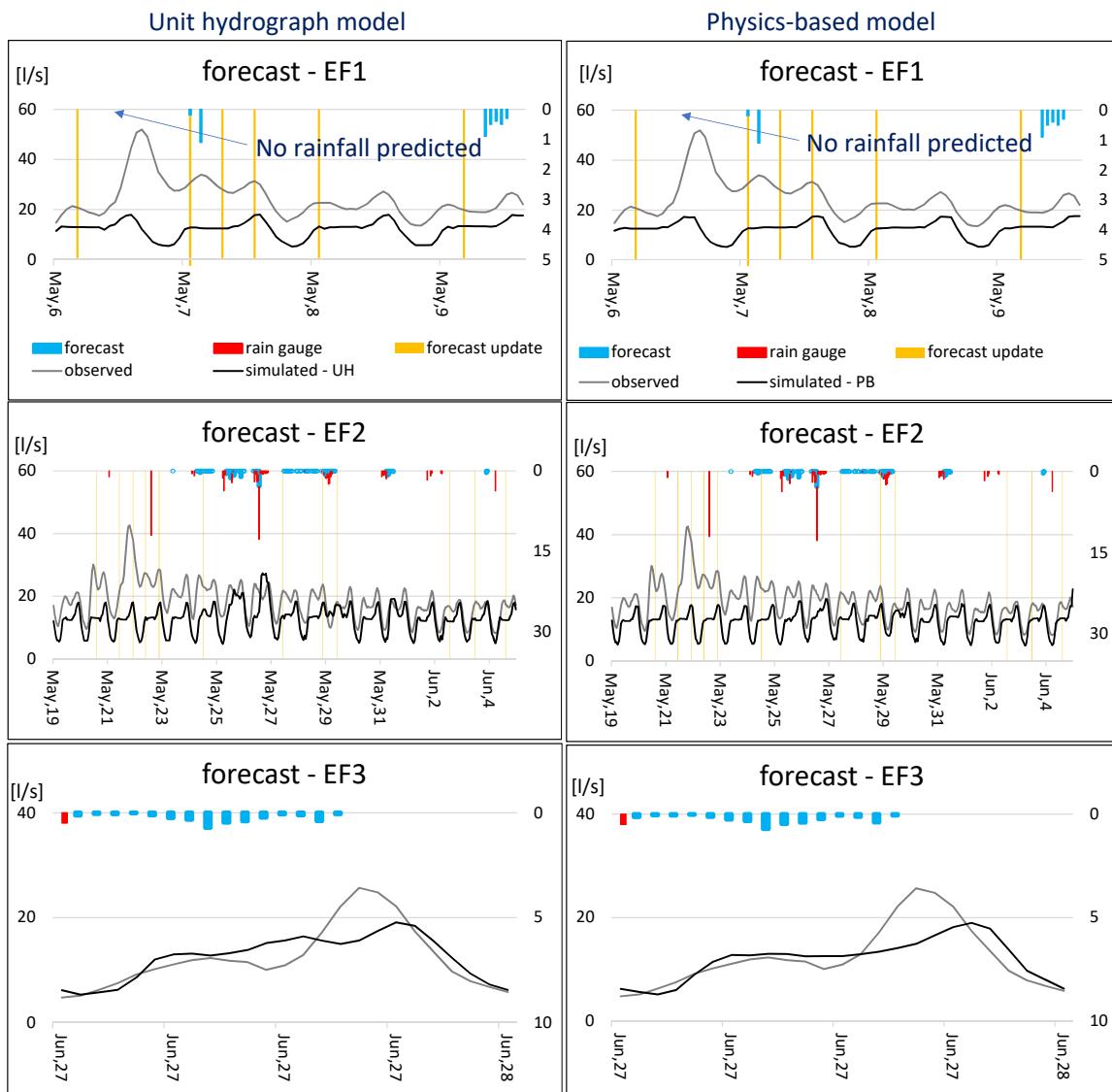


Figure 5.7.: Preliminary analysis of model performance with precipitation forecasted by the HARMONIE model

## 5.5. Discussions for a continuous simulation application

A continuous application is defined here as an operational system able to forecast the total discharge on chosen points of the system continuously with preset update intervals. The sewer model proposed in this study can be used as part of such application to estimate the inflow caused by external infiltration during wet periods. The accuracy is, however, limited to several uncertainties, such as input data accuracy, parameter estimation, physical processes simplifications, and initial conditions.

The antecedent condition has a considerable impact on the flow for the following hours after the simulation, as depicted in Figure 5.8. The simulated event occurred during the dormant season of 2019. Short-term simulations of event 11 were carried using the calibrated physics-based model for the same season. Event 10 occurred shortly before rose the flow rate in the system. Distinct initial conditions were assigned to each of the six simulations. The varied initial conditions were: the simulation start time (START); the initial water depth equivalent in the snow pack (SWE); fraction of free water capacity (snow storage) and initial groundwater table elevation ( $GWT_{initial}$ ). SWE was updated based on a 1:10 ratio of the measured snow depth for the time of the start of each simulation. A short description of the chosen initial conditions and results obtained for each simulation is provided below:

1. simulation started 14 days before the first reported flow. The initial water table elevation was set to zero and the previous event 10 was also included in the simulation increasing the groundwater table levels and the groundwater inflow (GWF) prior to the beginning of event 11. The peak flow was underestimated by 3,39 % and the volume was overestimated by an approximate of 16 %.
2. simulation 2 had the start time updated to seven days prior the event and SWE to 57,1 mm. A similar result compared to simulation 1 regarding volume and peak errors with the groundwater inflow as the main difference. The shorter period simulated in advance (7 days less than simulation 1) missed part of the snowmelt amount of the previous event reducing the groundwater table elevation and the groundwater inflow.
3. Simulation 3 started three days prior the event period.  $GWT_{initial}$  remained zero. The simulated flows at the beginning of the event were, no longer, influence by groundwater inflow component as the simulation start time skipped the snowmelt of the previous period that recharged the aquifer.
4. Simulation 4 behaved similarly to simulation 3 as  $GWT_{initial}$  remained zero.
5.  $GWT_{initial}$  was updated to 10 cm above the bottom elevation of all receiving nodes for simulation 5. The updated  $GWT_{initial}$  impacted the groundwater inflow resulting in a similar rate as the simulations with much earlier simulation start. However, the peak flow was still underestimated.

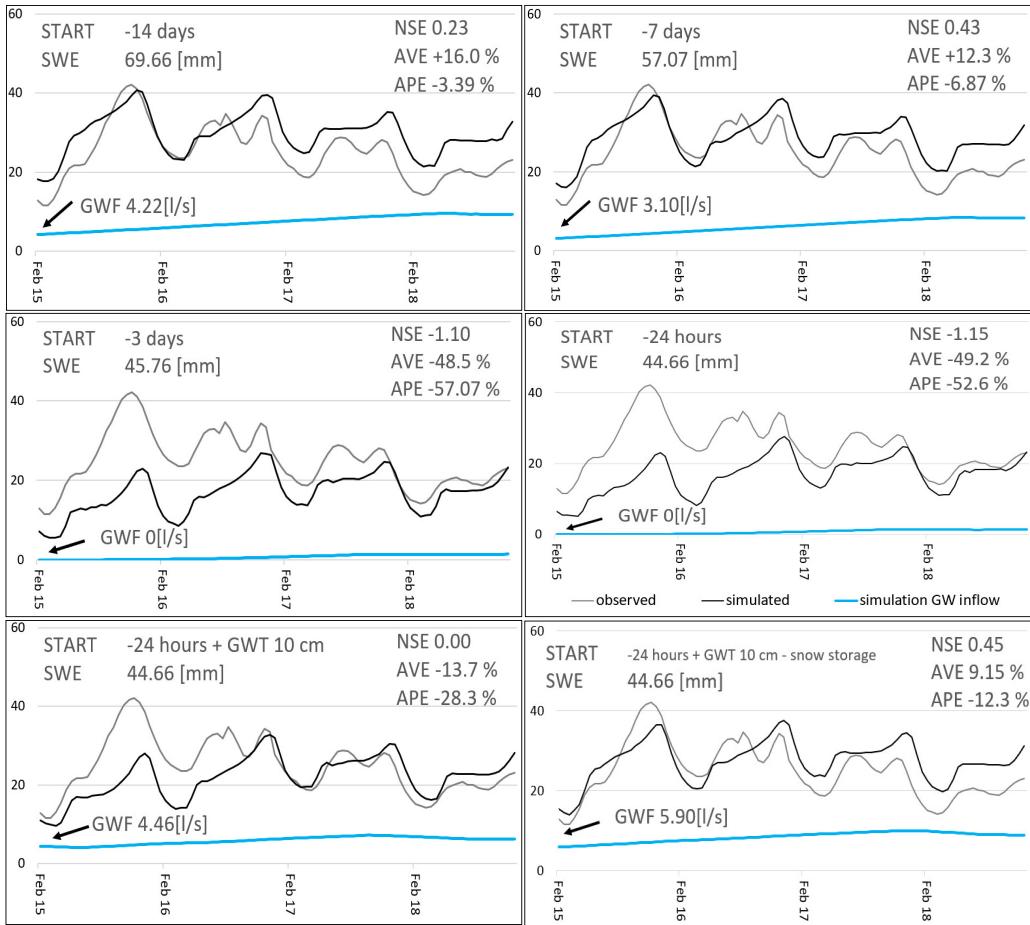


Figure 5.8.: Comparison of simulation results with different initial conditions for the physics-based model

6. Simulation 6 had the same initial parameters as simulation 5, except by the snow storage fraction which was set to zero. The missing volume of the first day observed from simulation 5 was then returned since the snowmelt did not need to fill the porous space of the snowpack before running off. The results of simulation 6 measured by the goodness-of-fit tests were similar, if not better, than simulations with much earlier start time, such as simulation 1.

The groundwater inflow increased as the snowmelt recharged the aquifer and rose the groundwater table elevation throughout the event period. This was observed in all the six simulations. The average observed flow, however, decreased during the recession period. A change on the aquifer parameters would be required, such as the increase of the seepage rate, to better estimate the recession curve of this specific event. However, this would affect the reproduction of other events of the dormant season, since the aquifer parameters were calibrated based on three events of the long-term simulation. Therefore, the errors of the recession curve are limited by the multi-event calibration approach.

The different results, obtained by varying the initial conditions, shows the need to preserve the previous hydrological states, influenced by the past events, for a continuous simulation application. The initial conditions can be obtained from a hot start file which contains the hydraulic and hydrological states from a specific time of the last simulation. A new hot start file can be generated at the end of the current simulation. This can mimic a long-term simulation, but with short intervals, preserving the hydrological states of the system for the next simulation.

Moreover, some parameters can be updated based on measured data to avoid propagation of errors from the previous simulations. These parameters can be, for example, the parameters assessed by the six simulations depicted in Figure 5.8. The initial depth of water equivalent of snow pack can be updated based on snow depth measured data and estimated water content of the snow pack. The initial water table  $GWT_{initial}$ , which limits the groundwater inflow, could be estimated by subtracting the dry-weather flow component from the flow metering data of the previous day.

An example of a simplified computational scheme for a continuous simulation application is suggested below:

1. Connect to the data provider's API and the water utility SCADA system and retrieve the weather forecast and the last flow recorded flow data.
2. Assimilate the weather forecast data converting, if not already, the data in time series input format used by the hydrological models. Check the quality of the flow metering data with a predefined routine that, for example, remove outliers and estimate values of the missing data.
3. Estimate and update the model's initial parameters based on weather forecast and flow metering data. For the physics-based model the updated parameters can be the initial depth of water equivalent of snow pack, Initial water table elevation, and update snowmelt parameters, such as snowmelt threshold ( $T_b$ ) temperature and maximum melt rate ( $DHM_{max}$ ). For the unit hydrograph model external inflow representing the groundwater flow infiltration based on previous flow records can be updated.
4. Run the simulations, in parallel, with different data sources, if existent, and the two hydraulic models using the hot start file with the previously estimate hydrological and hydraulic states of the system. As an example, four simulation run simultaneously if two forecast model predictions are available.
5. Repeat the process when the next updated forecast is available and start the calibration routine if the current time matches the previously time set for the calibration.

A continuous model calibration can be also be introduced as part of a real-time application. As time passes, new events can be used for the calibration of the model parameters. For this, the system should be able to autonomously perform the event detection on a similar fashion as the procedure done in this study using the EPA SSOAP toolbox. The frequency of calibration varies according to the model being used and season to be simulated. Events of the dormant season can have long duration, such

as two months. Therefore, less frequent calibrations, such as one per season, could be applied. Since only two seasons were analyzed in this study, there is no evidence that more than one calibration is required for the dormant season.

Calibrations of the growth season can be more frequently than the dormant season with one per month or even one every two weeks due to the shorter duration of the event. The amount of iterations of the optimization algorithm depends on the available computational processing power and number of computational nodes to be calculated by SWMM. NSE higher than 0,75 was obtained with 600 iterations of the DDS optimization algorithm. A calibration run of an event duration of a week lasted for about six hours, with preemption, and 1 min simulation time of the model using SWMM 5.1 and the unit hydrograph model. Moreover, the possibility of including new input data, such as historical radar-based rainfall, for new calibrations is relevant to assure flexibility for the process as more accurate input data becomes available.

## **6. CONCLUSIONS AND RECOMMENDATIONS**

### **6.1. Conclusions**

The primary goal of this study was to create two hydrological models, available in SWMM, to simulate the rainfall-derived infiltration and inflow (RDII) component of a sanitary sewer network (SSN) flow.

The output hydrograph of the hydrological models were added to specific points of an existing hydraulic model containing all pipes and nodes of the network. The dry-weather components of the SSN flow were included as part of the hydraulic model, and were estimated based on data of the water distribution and the usage pattern. The coupling between the hydrological and hydraulic model was carried to assess the relevant aspects of using such models on a continuous simulation application.

The wastewater flow of the sanitary sewer of a town located on the south of Finland was calibrated and simulated for an entire year period. Two sets of parameters were used for each model due to different hydrological behavior of winter season typical of regions with a cold climate.

Both models successfully simulated the RDII flows in response of a rainfall or snowmelt event occurring within the sewershed. The level of accuracy varied among the events with slightly better results obtained during winter and early spring by the physics-based model due to its snowpack and snowmelt routine.

GIS tools available on Qgis software allowed the automatic delineation and division of the 1.355 ha catchment in seven subcatchments, for each pumping stations, based on the digital elevation model. GIS operations also were used to estimate several parameters required for the physics-based model with available topographic data. The EPA SSOAP toolbox was used and found to be a helpful aid to identify the different components of SSN flow.

When topographic data, such as the digital elevation model (DEM), the land use, and soil type coverage are available, several parameters of the physics-based model can be estimated with equal or similar framework as described in this study. Parameters estimated with the aforementioned topographic data are necessary for the runoff, groundwater flow, and aquifer processes of SWMM physics-based model. After the estimation of most of the parameter of the model based on the available data, relatively accurate calibration results were obtained (NSE 0,54–0,89) by varying only the following eight sensible parameters: the percentage of runoff routed to the pervious subarea and flow width of each subcatchment of the runoff module; seepage rate, saturated hydraulic conductivity, conductivity slope, and flow coefficient ( $A_1$ ) of the aquifer and groundwater flow module; and the minimum and maximum melt coefficient of the snowpack module.

This study also proposed a method to define the possible range of the RTK unit hydrograph parameters based on precipitation and flow metering data. The range was used to limit the search-space of

the optimization algorithm, which successfully found values for the 15 parameters of each subcatchment within the proposed range. A range of parameters for the physics-based model were proposed based on a literature review and calibration of each individual process.

It was found that, after conceptual adaptations and calibrations, the SWMM physics-based model successfully reproduced the flows in the SSN and the snow pack variations throughout the winter and spring of 2018. Calibration of the percentage of runoff routed to the pervious subareas of the subcatchments allowed the simulation of losses not included in the model, such as losses for the stormwater sewer network, surface ponds, and rain harvesting systems. The groundwater infiltration simulation was achieved by setting the aquifer's bottom elevation to the same elevation of the receiving node and the - probably higher than usual - seepage rate assigned to the aquifer's saturated zone to the aquitard allowed a more sensible fluctuation than customary groundwater table variations. This approach was crucial to simulate the long recession infiltration observed in the SSN after a snow melt period.

The comparison of rain gauge and radar-based precipitation measurements, simulations carried with different initial conditions, and preliminary analysis of the model's performance using precipitation forecast were useful to emphasize the importance of the spatial distribution of the recorded variables used in the models, especially during a continuous simulation for a near to real-time application to predict future flows in the SSN. The frequency of calibrations for a continuous application is suggested as one per season for the parameters used to simulate the winter and early spring (dormant season) and twice a month, if necessary, for the rest of the seasons when snow cover is not present.

## **6.2. Recommendations for future research**

The performance of both models can be compared when simulating less frequent, and worst-case scenario, events, such as high intensity rainfalls during snow melt periods to assess the design and current risk of sanitary sewer overflow (SSO) and combined sewer overflow (CSO). The assessment of the unit hydrograph performance when using the same "effective rainfall" (snowmelt + rainfall) of the physics-based model, not carried in this study, can be a topic for a future research.

Remediation strategies for the sewer network can be further studied also using simulation of historical events and both models proposed in this study. A more extent evaluation of the accuracy of the simulated flow upstream, when only a downstream recorded flow is used for the calibration of the models, could be an useful research to identify, usually ungauged, parts of the network candidates for rehabilitation or replacement due to high RDII.

Much can be improved, and further studied, for continuous applications regarding the calibration process used in this study. The use of an optimization algorithm, such as the dynamically dimensioned search (DDS), can be implemented also to calibrate the physics-based model with the sensitive parameters used in this study for a manual calibration.

The use of a multi-objective function, able to weight the importance of volume, peak flow, and overall performance when evaluating the simulation results, is necessary to correctly identify the best fit in agreement with the objective of the model. Identifying a framework with a unique penalty

## *6. Conclusions and recommendations*

function specifically for sanitary sewer networks for all seasons could be an interesting research and useful for automatized continuous calibration of the sewer model.

A meta-optimization framework which searches for the optimum range parameters to be calibrated, constrained by physics meaningful values, and the use of discrete design variables to reduce the available search space can be also further investigated for the calibrated parameters proposed in this study for both hydrological models.

A framework to adjust and match the current measured flow and the next forecasted flow to avoid discontinuity on the results would be helpful for a forecast system. The assessment of more than one, if available, forecast model and historical rainfall measurements can be defined as precedent study for continuous simulations of a flow forecast system. Also for a continuous and near real-time application, questions, such as how often should the initial parameters, influencing SWMM's aquifer's soil moisture and snow pack cold content, be updated during a continuous simulation application to reduce the possible error propagation.

## BIBLIOGRAPHY

- [1] GRASS GIS. URL <https://grass.osgeo.org/>.
- [2] A. Osman Akan. Horton Infiltration Equation Revised. *Journal of Irrigation and Drainage Engineering*, 118(5):828–830, 1992.
- [3] R.J. Akan, A.O. and Houghtalen. *Urban Hydrology, Hydraulics, and Stormwater Quality: Engineering Applications and Computer Modeling*. Wiley, 2003. ISBN 9780471431589.
- [4] Alan Cressler. Combined Stormwater/Sewage Overflow, Roswell, GA, Sept. 2009, 2009. URL <https://www.usgs.gov/media/images/combined-stormwatersewage-overflow-roswell-ga-sept-2009>.
- [5] Eric Anderson. Snow accumulation and ablation model-SNOW-17, NWSRFS user manual. Technical Report January, 2006. URL [http://www.nws.noaa.gov/oh/hr1/nwsrfs/users{\\_}manual/part2/{\\_}pdf/22snow17.pdf](http://www.nws.noaa.gov/oh/hr1/nwsrfs/users{_}manual/part2/{_}pdf/22snow17.pdf).
- [6] Eric A . Anderson. National Weather Service River Forecast System- Snow Accumulation And Ablation Model. Technical report, Washinton, DC, 1973. URL <https://repository.library.noaa.gov/view/noaa/13507>.
- [7] Richard Arsenault, Annie Poulin, Pascal Côté, and François Brissette. Comparison of Stochastic Optimization Algorithms in Hydrological Model Calibration. *Journal of Hydrologic Engineering*, 2013. ISSN 1084-0699. doi: 10.1061/(asce)he.1943-5584.0000938.
- [8] Greg Barden, Tim Fallara, Hunter Kelly, Fang Cheng, Benjamin J. Sherman, and Edward Burgess. Comparison of RDII Unit Hydrograph Approaches for Continuous Simulation using SWMM 5. *Journal of Water Management Modeling*, 6062:209–222, 2011. doi: 10.14796/jwmm.r241-12.
- [9] David Bennett, Reggie Rowe, Marie Strum, David Wood, Nancy Schultz, Kelly Roach, Mike Spence, and Virgil Adderley. Using flow prediction technologies to control sanitary sewer overflows. Technical report, 1999.
- [10] H Bouwer. *Groundwater Hydrology*. McGraw-Hill, New York, 1978.
- [11] Kyung Sook Choi and James E. Ball. Parameter estimation for urban runoff modelling. *Urban Water*, 2002. ISSN 14620758. doi: 10.1016/S1462-0758(01)00072-3.

- [12] Nam Jeong Choi. *Understanding sewer infiltration and inflow using impulse response functions derived from physics-based models*. PhD thesis, University of Illinois at Urbana-Champaign. URL <https://www.ideals.illinois.edu/handle/2142/90540>.
- [13] City of San Luis Obispo - CA. Sewer Lateral Pipe Material. URL <https://www.slocity.org/government/department-directory/utilities-department/wastewater/wastewater-collections/sewer-lateral-pipe-material/>.
- [14] Ana Maria de Roda Husman, Ton de Nijs, Ankie Sterk, Jack F. Schijven, and Heleen de Man. Climate change impact on infection risks during bathing downstream of sewage emissions from CSOs or WWTPs. *Water Research*, 105:11–21, 2016. ISSN 00431354. doi: 10.1016/j.watres.2016.08.053.
- [15] S Dent, R Blair Hanna, and L T Wright. Automated Calibration using Optimization Techniques with SWMM RUNOFF. *Journal of Water Management Modeling*, pages 220–238, 2004. ISSN 2292-6062. doi: 10.14796/JWMM.R220-18. URL [www.chijournal.org](http://www.chijournal.org).
- [16] EPA. Review of Sewer Design Criteria and RDII Prediction Methods January 2008. Technical report. URL <https://nepis.epa.gov/Adobe/PDF/P1008BP3.pdf>.
- [17] European Environment Agency (EEA). © European Union, Copernicus Land Monitoring Service 2017, 2017.
- [18] Giovanni De Feo, George Antoniou, Hilal Franz Fardin, Fatma El-gohary, Xiao Yun Zheng, Ieva Reklaityte, David Butler, Stavros Yannopoulos, and Andreas N Angelakis. The Historical Development of Sewers Worldwide. (June):3936–3974, 2014. doi: 10.3390/su6063936.
- [19] FMI. Open data - Finnish Meteorological Institute, . URL <https://en.ilmatieteenlaitos.fi/open-data>.
- [20] FMI. Numerical weather prediction, . URL <https://en.ilmatieteenlaitos.fi/numerical-weather-prediction>.
- [21] FMI. Year 2018 was warm and dry - Finnish Meteorological Institute, 2019. URL <https://public.wmo.int/en/media/news-from-members/year-2018-was-warm-and-dry-finnish-meteorological-institute>.
- [22] Monomoy Goswami and Kieran Michael O'Connor. Comparative assessment of six automatic optimization techniques for calibration of a conceptual rainfall-runoff model. *Hydrological Sciences Journal*, 2007. ISSN 02626667. doi: 10.1623/hysj.52.3.432.
- [23] GTK. Geological Survey of Finland (GTK) Spatial Data Products. URL <https://hakku.gtk.fi/en/locations/search>.

- [24] Milja Heikkinen, Paula Schönach, and Ilmo Massa. Politicization of wastewater overflows: The case of the Vantaa River, Finland. *Water Policy*, 18:1454–1472, 2016. doi: 10.2166/wp.2016.011. URL [https://helda.helsinki.fi//bitstream/handle/10138/187724/Wastewater{\\_\]overflow{\\_\]pre{\\_\]copy{\\_\]edit{\\_\]final.pdf?sequence=1](https://helda.helsinki.fi//bitstream/handle/10138/187724/Wastewater{_]overflow{_]pre{_]copy{_]edit{_]final.pdf?sequence=1).
- [25] Sharif Hossain, Guna Alankarage Hewa, and Subhashini Wella-Hewage. A comparison of continuous and event-based rainfall-runoff (RR) modelling using EPA-SWMM. *Water (Switzerland)*, 11(3), 2019. ISSN 20734441. doi: 10.3390/w11030611.
- [26] IPCC. *Climate Change 2014: Synthesis Report, Contribution of working Groups I, II and III to the Fifth Assess*, volume 218. 2014. ISBN 9789291691432. doi: 10.1016/S0022-0248(00)00575-3.
- [27] Kimberly Davis. Solving Sewer Overflows Green Or Gray WERF Shows The Way, 2016. URL <https://www.watersonline.com/doc/solving-sewer-overflows-green-or-gray-werf-shows-the-way-0001>.
- [28] Joong Gwang Lee, Christopher T. Nietch, and Srinivas Panguluri. SWMM Modeling Methods for Simulating Green Infrastructure at a Suburban Headwatershed : User's Guide. (October), 2017.
- [29] Jiuling Li, Keshab Sharma, Yiqi Liu, Guangming Jiang, and Zhiguo Yuan. Real-time prediction of rain-impacted sewage flow for on-line control of chemical dosing in sewers. *Water Research*, 149:311–321, 2019. ISSN 18792448. doi: 10.1016/j.watres.2018.11.021. URL <https://doi.org/10.1016/j.watres.2018.11.021>.
- [30] Toshitsugu Moroizumi, Naoya Ito, Jari Koskiaho, and Sirkka Tattari. Long-term trends of pan evaporation and an analysis of its causes in Finland. Technical report. URL [http://ousar.lib.okayama-u.ac.jp/files/public/5/52684/20160528114009612690/B03-PanFinland{\\_\]Moroizumi.pdf](http://ousar.lib.okayama-u.ac.jp/files/public/5/52684/20160528114009612690/B03-PanFinland{_]Moroizumi.pdf).
- [31] Charles Mosley, Shawn Dent, Paul Kadota, Leonard T. Wright, and Yassine Djebbar. Comparing Rainfall Dependent Inflow and Infiltration Simulation Methods. *Journal of Water Management Modeling*, (January), 2001. doi: 10.14796/jwmm.r207-16.
- [32] NSL. National Land Survey of Finland. URL <https://tiedostopalvelu.maanmittauslaitos.fi/tp/kartta?lang=en>.
- [33] Python Software Foundation. Python 3.7. URL <https://www.python.org/>.
- [34] Qgis.org. QGIS. URL <https://qgis.org/en/site/index.html>.
- [35] Walter J. Rawls, Donald L. Brakensiek, and Norman Miller. Green-ampt Infiltration Parameters from Soils Data. *Journal of Hydraulic Engineering*, 109(1):62–70, 1983. ISSN 0733-9429. doi: 10.1061/(asce)0733-9429(1983)109:1(62).

- [36] Brent Robinson. Modeling Sanitary Sewer Groundwater Inflow Rehabilitation Effectiveness in SWMM5 Using a Two Aquifer Approach. *Journal of Water Management Modeling*, (Sanipor 2014):1–12, 2015. doi: 10.14796/jwmm.c385.
- [37] Larry A Roesner. Storm Water Management Model Manual. *Usepa*, I(July), 2017.
- [38] Lewis A Rossman and Wayne C Huber. Storm Water Management Model Reference Manual. Volume I - Hydrology (Revised). EPA/600/R-15/162A. I:231, 2016. URL <https://www.epa.gov/water-research/storm-water-management-model-swmm>.
- [39] Abraham SAVITZKY and Marcel J.E. GOLAY. Soothing and differentiation of data by simplified least squares procedures. *Anal. Chem.*, 36:1627–1639, 1964, 1964.
- [40] Markus Sunela. *Real-Time Control Optimization of Water Distribution System with Storage*. 2017. ISBN 9789949831654. doi: 10.3917/lett.051.0109.
- [41] SYKE. Finnish Environment Institute (SYKE) Open Environmental Information Systems. URL [https://www.syke.fi/fi-FI/Avoin{\\_}tieto/Ymparistotietojarjestelmat](https://www.syke.fi/fi-FI/Avoin{_}tieto/Ymparistotietojarjestelmat).
- [42] Hessam Tavakol-Davani, Erfan Goharian, Carly H. Hansen, Hassan Tavakol-Davani, Defne Apul, and Steven J. Burian. How does climate change affect combined sewer overflow in a system benefiting from rainwater harvesting systems? *Sustainable Cities and Society*, 27:430–438, 2016. ISSN 22106707. doi: 10.1016/j.scs.2016.07.003.
- [43] Henri Tikkanen. *Hydrological modeling of a large urban catchment using a stormwater management model (SWMM) (MSc thesis)*. 2013.
- [44] TMFG. Road traffic | Digitraffic, . URL <https://www.digitraffic.fi/en/road-traffic/>.
- [45] TMFG. Open Data, . URL <https://www.tmfq.fi/fi/tmfq/avoindata>.
- [46] Bryan A. Tolson and Christine A. Shoemaker. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research*, 2007. ISSN 00431397. doi: 10.1029/2005WR004723.
- [47] UN. 2018 Revision of World Urbanization Prospects | Multimedia Library - United Nations Department of Economic and Social Affairs, 2018. URL <https://www.un.org/development/desa/publications/2018-revision-of-world-urbanization-prospects.html>.
- [48] UN-World Water. *World Water Development Report 2017*. 2017. ISBN 9789231002014. URL [www.unwater.org](http://www.unwater.org).

- [49] C. Valeo and C.L.I. Ho. Modelling urban snowmelt runoff. *Journal of Hydrology*, 299(3-4): 237–251, 2004. ISSN 0022-1694. doi: 10.1016/J.JHYDROL.2004.08.007. URL <https://www.sciencedirect.com/science/article/pii/S0022169404003683>.
- [50] Srinivas Vallabhaneni and Edward H Burgess. Computer Tools for Sanitary Sewer System Capacity Analysis and Computer Tools for Sanitary Sewer System. *Environmental Protection*, (October):1–104, 2007.
- [51] M. Wallner, U. Haberlandt, and J. Dietrich. Evaluation of different calibration strategies for large scale continuous hydrological modelling. *Advances in Geosciences*, 2012. ISSN 16807340. doi: 10.5194/adgeo-31-67-2012.
- [52] WHO. World Health Organization: Sanitation, 2019. URL <https://www.who.int/news-room/fact-sheets/detail/sanitation>.
- [53] Derek Wride, Mi Chen, D Ph, and Ralph Johnstone. Characterizing the Spatial Variability of Rainfall Across a Large Metropolitan Area. 2:1–10, 2004.

## A. APPENDIX

Digital appendix available on <https://github.com/pedrohydroweb/Thesis-hydroinformatics>.

Content:

- Tables with calibrated parameters for the RTK unit hydrograph method. Values for each sub-catchment and season.
- Python script for measured wastewater flow treatment.
- Python scripts to collect and parse weather forecast information from the Finnish Meteorological Institute (FMI) (HIRLAM and HARMONIE models) and the weather road forecast from the Traffic Management Finland Group (TMFG).