Aalto University
School of Chemical
Technology

School of Chemical Technology
Degree Programme of Chemical Technology

Saara Lehtinen

SIMULATION OF STORMWATER QUALITY IN AN URBAN CATCHMENT USING THE STORMWATER MANAGEMENT MODEL (SWMM)

Master's thesis for the degree of Master of Science in Technology submitted for inspection, Espoo, 10 November, 2014.

Supervisor Professor Harri Koivusalo

Instructor M.Sc. Gerald Krebs

Aalto University
School of Chemical
Technology

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of master's thesis

Author Saara Lehtinen	
Title of thesis Simulation of stormwater quality in an urban catchment using the Stormwater Management Model (SWMM)	
Department Department of Civil and Environmental Engineering	
Professorship Water Resources Engineering	**Code of professorship** Yhd-12
Thesis supervisor Professor Harri Koivusalo	
Thesis advisor M.Sc. (Tech.) Gerald Krebs	
Date 10.11.2014 **Number of pages** 87 + 5	**Language** English

Abstract

In the face of climate change, population growth and urbanization an understanding of stormwater quality processes and their prediction in urban areas are essential to make good use of stormwater and to minimize its detrimental impacts on the population and the environment.

In this study a stormwater quality model calibration was conducted using the Stormwater Management Model (SWMM) for an urban catchment in Lahti, Finland by utilizing rainfall, runoff and turbidity data from the catchment outlet. The continuously observed turbidity data was converted to TSS concentrations, which represented water quality. With the aid of a genetic algorithm the calibration was conducted for five model parameters: the maximum build-up, build-up exponent, wash-off coefficient and wash-off exponent, which appear in the exponential build-up and wash-off functions of the model and the initial build-up of pollutants on the catchment surfaces. Three different sets of parameter boundaries and different calibration sequences of one, two or three events were used. A sensitivity analysis was conducted before and after the calibration to investigate the interrelations of the parameters and their effect on the model output.

The simulation results showed that the exponential functions are adaptable within the requirement that fluctuations in water quality are related to fluctuations in runoff. A good performance was obtained for the calibration sequence, when three events were used. However, even three events are too small a sample of the vast variety of rainfall-runoff events and only a few of the validation events were able to be predicted with the calibrated parameters. Additionally, the importance of accurate data is well demonstrated.

The results indicate that including the initial build-up as a calibration parameter improves the model performance. The parameters show complex interrelations, but some clear tendencies and regularities were identified. The same set of parameters was assigned for all impervious surfaces with the exception of one optimization, where the parameters for roof surfaces were optimized independently. Interestingly, this optimization produced the best performing parameters for the calibration sequence, which encourages for more detailed stormwater quality modelling.

This study provides the obtained ranges and behaviour of the quality parameters as an approximation or a comparison for future stormwater quality modelling applications in similar urban catchments.

Keywords Urban stormwater quality, modelling, SWMM, quality parameterization

Aalto-yliopisto
Kemian tekniikan korkeakoulu

Aalto-yliopisto, PL 11000, 00076 AALTO
www.aalto.fi

Diplomityön tiivistelmä

Tekijä Saara Lehtinen

Työn nimi Kaupunkivaluma-alueen huleveden laadun simulointi SWMM-mallilla

Laitos Yhdyskunta- ja ympäristötekniikan laitos

Professuuri Tekninen vesitalous **Professuurikoodi** Yhd-12

Työn valvoja Professori Harri Koivusalo

Työn ohjaaja DI Gerald Krebs

Päivämäärä 10.11.2014 **Sivumäärä** 87 + 5 **Kieli** Englanti

Tiivistelmä

Ilmastonmuutoksen, väestönkasvun ja kaupungistumisen myötä hulevesien rooli kaupunkialueilla tulee yhä tärkeämmäksi. Hulevesien laatuprosesseja tulisi ymmärtää paremmin ja niitä pitäisi pystyä mallintamaan, jotta hulevettä voitaisiin hyödyntää ja toisaalta sen haitallisia vaikutuksia ihmisiin ja ympäristöön voitaisiin vähentää.

Tässä työssä hulevesien määrää ja laatua kuvaava SWMM-malli kalibroitiin kaupunkivaluma-alueelle Lahdessa hyödyntäen sadanta-, valunta- ja sameusaineistoa valuma-alueen purkupisteestä. Sameus muutettiin kokonaiskiintoaineen konsentraatioksi, joka edusti huleveden laatua. Geneettisen algoritmin avulla suoritettiin kalibrointi viidelle parametrille: saasteiden lähtökertymälle valuma-alueen pinnoilla, sekä maksimikertymälle, kertymiseksponentille, huuhtoutumiskertoimelle ja huuhtoutumiseksponentille, jotka esiintyvät mallin eksponentiaalisissa kertymis- ja huuhtoutumisfunktioissa. Kalibroinnissa käytettiin kolmea erilaista parametrien vaihteluvälien joukkoa ja erilaisia yhden, kahden tai kolmen tapahtuman sarjoja. Herkkyysanalyysi suoritettiin ennen ja jälkeen kalibroinnin, jotta parametrien keskinäisiä suhteita ja vaikutuksia mallin tuloksiin voitaisiin tutkia.

Työ osoittaa, että eksponentiaaliset funktiot ovat mukautuvia sillä ehdolla, että veden laadun vaihtelut seurailevat valunnan vaihteluita. Parhaat tulokset kalibrointijaksolle saatiin laajoilla vaihteluväleillä ja käyttämällä kolmea kalibrointitapahtumaa yhtä aikaa. Silti kolmekin tapahtumaa on hyvin pieni otos edustamaan kaikkia erilaisia sadanta-valuntatapahtumia, minkä vuoksi vain muutama validointitapahtuma pystyttiin ennustamaan kalibroiduilla parametreilla. Lisäksi tarkan ja luotettavan aineiston tärkeys tulee selvästi näkyviin.

Työn tulokset viittaavat siihen, että lähtökertymän lisääminen kalibrointiparametriksi parantaa mallin toimivuutta. Laatuparametrit ovat monimutkaisesti sidoksissa toisiinsa, mutta joitakin selviä taipumuksia ja säännönmukaisuuksia voidaan tunnistaa. Optimoinneissa läpäisemättömille pinnoille annettiin samat laatuparametrit, paitsi yhdessä optimoinnissa kattopintojen parametrien annettiin vaihdella itsenäisesti. Tämä optimointi tuotti parhaan tuloksen kalibrointijaksolle, mikä kannustaa tutkimaan myös yksityiskohtaisempaa huleveden laadun mallinnusta.

Tämä työ tarjoaa tietoa SWMM-mallin laatuparametrien käyttäytymisestä. Parametreille saatuja vaihteluvälejä voidaan käyttää suuntaa-antavina arvoina ja vertailukohtana tulevissa hulevesien laadun mallinnussovelluksissa samankaltaisilla kaupunkivaluma-alueilla.

Avainsanat Hulevesien laatu, mallintaminen, SWMM, laatuparametrisointi

Acknowledgements

This study was carried out at the Aalto University School of Engineering as a part of the research project '*Urban laboratory for sustainable built environment'* (Urban Laboratory project for short). It is a co-operative research project between Aalto University, Helsinki University and Lahti University of Applied Sciences and is mainly funded by the European Regional Development Fund. The main core of the project is building a strong foundation for research in the field of sustainable urbanization in the Lahti area with a focus on controlling the water cycle in the urban environment. I am happy to have been introduced to such an interesting and current topic.

First of all I want to thank Professor Harri Koivusalo, who supervised this thesis, and my advisor M.Sc. Gerald Krebs for the guidance and feedback he gave me throughout the project. I felt he was always ready to help. It is also important for me to acknowledge Juhani Järveläinen, Tuukka Ryynänen, Marjo Valtanen and Nora Sillanpää, who helped me in the beginning of my work, and all the others who were part of the Urban Laboratory project. In addition I want to thank my co-workers for the time I spent with the Water and Environmental Engineering research group. Maa- ja vesitekniikan tuki ry. provided additional funding for this thesis, for which it is thankfully acknowledged.

I appreciate the ever so delightful and uplifting company of my wonderful friends and thank them for being such an inspiration. My close friend Tea Miller deserves a special mention for the peer support that was most welcome both in and outside of our studies.

Finally, I wish to praise my dear parents and siblings. I have always been able to trust them to have my back. I owe gratitude to my parents for their ability to make me focus on the essentials in life and for their overwhelming encouragement and support.

Helsinki 4.11.2014

Saara Lehtinen

Table of contents

Abbreviations

AP	Ainonpolku (the study catchment)
BMP	Best Management Practices
CAL(nr)	Calibration event (number)
CSO	Combined sewer overflow
CSS	Combined sewer system
EIA	Effective impervious area
EMC	Event mean concentration
FMI	The Finnish Meteorological Institute
FTU	Formazin turbidity units
HSF	Hot start file
LID	Low Impact Development
LSB	Lahti Science and Business Centre
MnC	Manning's roughness coefficient (*n*) for conduit flow
MnO	Manning's roughness coefficient (*n*) for overland flow
NTU	Nephelometric turbidity units
NSGA-II	Nondominated sorting genetic algorithm II
OR1, O1-O15	Optimizations
RS1, RS2, RS3	Range set 1, 2 and 3, respectively
Stor	Depression storage
SWMM	Stormwater Management Model
TIA	Total impervious area
TSS	Total Suspended Solids
VAL(nr)	Validation event (number)

GOODNESS-OF-FIT CRITERIA:

CORR	Linear correlation coefficient
E	Nash-Sutcliffe efficiency
PFE	Peak flow error (%)
SSE	Sum of squared error
TLE	Total load error (%)
VE	Volume error (%)

CALIBRATION PARAMETERS:

BE	Build-up exponent
IB	Initial build-up
MB	Maximum build-up
WC	Wash-off coefficient
WE	Wash-off exponent
RBE	Build-up exponent for roofs
RIB	Initial build-up for roofs
RMB	Maximum build-up for roofs
RWC	Wash-off coefficient for roofs
RWE	Wash-off exponent for roofs

List of figures

List of tables

1 Introduction

The world is facing three major global trends, which will affect the human societies for centuries to come: climate change, population growth and urbanization. Climate change is expected to increase precipitation in some areas of the world and decrease it in others, which poses challenges to both drainage of excess water and acquiring water for use. In high-latitude regions like Finland, precipitation is likely to increase (US EPA 2014a). Population growth will intensify urbanization and due to increasing imperviousness, the risks relating to stormwater flooding will increase respectively. Therefore the role of urban stormwater management becomes increasingly important. Urban stormwater runoff, by definition, is the runoff that results from rainfall over areas that are changed from their natural condition by human activities (House et al. 1993). Increasing the area of paved surfaces alters the hydrological cycle by reducing infiltration and at the same time accelerating surface runoff. When, additionally, the amount of precipitation increases, there is a large volume of stormwater that needs to be managed. Drainage systems are needed in urban environments to ensure the functionality of the urban infrastructure and public health (Barbosa et al. 2012). Runoff caused by rainfall needs to be safely either retained in suitable places inside the urban catchment or conveyed away from the catchment.

Stormwater releases and transports pollutants from urban surfaces to nearby natural receiving waters. Pollutants can have various detrimental implications for the ecology and quality of the waters by for instance altering the natural flow regimes, causing erosion and increasing eutrophication. Achieving and maintaining a good quality of water is increasingly important because of the objective of sustainability and the growing demand of potable water all over the world.

Stormwater quality (and quantity) modelling serves as an important tool for stormwater management. Extensive stormwater monitoring requires equipment and budget resources, but modelling allows for the use of less data and can be used to simulate time intervals beyond the monitoring period (Vezzaro & Mikkelsen 2012). This further allows for estimations and predictions over long periods and helps managers in planning stormwater management systems for selected urban areas. A wide variety of models addressing stormwater quantity and quality have been developed (Niemczynowicz 1999). The stormwater quantity models are well developed and widely adopted, whereas the quality models are less well established. The pollutant accumulation and wash-off processes are complex and encompass high temporal and spatial variations, which makes accurate and reliable stormwater quality modelling a challenging task (Dotto et al. 2010).

More research has been prompted to be carried out (e.g. Hossain et al. 2010) in order to better understand the dynamics of build-up and wash-off of pollutants in different environments and to derive parameters for the build-up and wash-off functions that could be used in urban stormwater modelling and planning. The motivation for this study is to develop knowledge on stormwater quality parameters, which is expected to set a starting point to stormwater quality modelling and more effective stormwater management in Finnish conditions.

1.1 Objectives and scope of the study

The main goal of this thesis was the parameterization of the water quality module of the US EPA Stormwater Management model (SWMM) for a Finnish urban catchment. The more profound idea behind this study was to provide more knowledge about the quality model, how it can be applied and what the prerequisites are for it to function well in modelling stormwater quality.

The following specific objectives were necessary to reach the final goal:

1. Analysis of rainfall, runoff and turbidity data from the study catchment to identify suitable events to be used for model calibration and validation.
2. Defining a correlation between turbidity and total suspended solids (TSS) concentration that can be used for converting the continuously measured turbidity into continuous TSS concentrations.
3. Investigation of the interrelations of the quality parameters and their effect on the model output by sensitivity analyses before and after calibration.
4. Calibration and validation of the quality model parameterization with different parameter boundaries for different sets of events.

The simulations were conducted by using solely the exponential build-up and wash-off functions of the SWMM quality model. This study focused on the simulation of stormwater quality processes over impervious areas in the present conditions of the study catchment. The performance statistics of the sensitivity analyses and the optimizations are provided in order for the results to be openly analysed, discussed and interpreted. The results can be used in future studies as guidance, comparison or approximate parameter values for similar urban catchments.

2 Theoretical framework

2.1 Urban hydrology

Stormwater runoff is formed over surfaces during rainfall or snowmelt events, when water over land surfaces is not infiltrated or otherwise percolated into the ground (US EPA 2014b). Stormwater from urban areas is sometimes called urban runoff as a distinction from runoff from rural or forested areas. Herein the term stormwater is used and refers to stormwater from urban (constructed) areas unless otherwise specified.

Urbanization has dramatic impacts on the hydrology of a catchment by altering the way water flows, infiltrates and evaporates. The impacts are various but result mostly from the elevated fraction of impervious areas in urban catchments compared to natural catchments. Impervious surfaces are mostly artificial surfaces such as roofs and different pavements of roads, driveways, parking lots, playfields, sidewalks, etc. In urban areas many natural surfaces are also compacted very tight, reducing their infiltration capacity. Rainfall cannot infiltrate from impervious surfaces, which causes stormwater to flow over the urban surfaces until it evaporates, meets a pervious surface where it can infiltrate or becomes collected into an underground pipe network.

When infiltration is reduced, the runoff response to rainfall becomes faster and leads to increases in total and direct runoff volumes, higher peak flows, shorter times of recession and shorter times of concentration, which is defined as the time needed for water to flow from the hydraulically outermost point of the basin to the outlet of the basin (Fletcher et al. 2013, Sillanpää 2013, Shuster et al. 2005, Barbosa et al. 2012, Haan et al. 1994, p. 75). Most impervious surfaces also transport water more effectively than natural pervious surfaces. For example, the movement of water is hindered more over a grass field than it is over smooth asphalt. Impervious smooth surfaces cause the stormwater to flow more rapidly to the inlets, pits and low-lying areas of the catchment and the stormwater infrastructure concentrates the distributed stormwater flows very rapidly so that infiltration is inhibited (Shuster et al. 2005). Faster collection of stormwater can result in floods at the outlet of the catchment or earlier along the pipe network if the water does not have enough time to discharge and the capacity is exceeded. As a consequence of flooding and fast movement of water, areas downstream of urban catchments may also experience severe stream bank erosion problems (Sillanpää 2013).

Effective impervious area (EIA) that is the impervious area with a direct hydraulic connection to the drainage system is the main contributor to surface runoff during most rainfall events (Shuster et al. 2008, Shuster et al. 2005). In urban areas where EIA is large, small- and medium-sized rainfall events are of main concern, because they occur frequently and even though the total event runoff volume is not large, they can cause considerable harm. Booth et al. (2002) found that the runoff generation from a storm of 2-year recurrence interval in a catchment of 10% level of EIA yielded the same discharge volume as a storm of 10-year recurrence interval in the same catchment before urbanization. For larger rainfall events the importance of total impervious area (TIA) increases as the disconnected impervious area also starts contributing to runoff (Sillanpää 2013).

Urbanization causes reduced recharge of groundwater through inhibition of infiltration. Furthermore the quality of groundwater may be impacted, but that depends on the quality of stormwater (Shuster et al. 2008, Shuster et al. 2005). The removal of vegetation, reduction of infiltration and the urban microclimate cause changes in evapotranspiration. The field of evapotranspiration in urban contexts remains quite poorly explored, but has recently gained more interest with the motivation of knowing how the urban water balance can be influenced by stormwater management (Fletcher et al. 2013).

2.2 Stormwater quality

2.2.1 Stormwater pollutants and their impact on receiving waters

Anthropogenic activity generates pollutants on urban surfaces that are transported to receiving natural water bodies by stormwater. Urbanization increases both the absolute amount of pollutants on surfaces and surface runoff, resulting in increased mobilization of pollutants (Barbosa et al. 2012). The pollutants in stormwater include nutrients, heavy metals, bacteria and suspended solids. In Helsinki, for instance, it was recently reported that the threshold values suggested by the province of Stockholm for stormwater were exceeded for copper, zinc and oil hydrocarbons (Airola et al. 2014).

Stormwater pollution originates from various sources; traffic, pesticides, cut grass and leaves, spillages, illegal wastewater connections, litter, wet and dry atmospheric deposition, faeces, de-icing chemicals and sand, to mention a few. Nutrients in urban stormwater originate for example from fossil fuel combustion as dry and wet atmospheric deposition, plant debris and fertilizers. The main nutrients are nitrogen and phosphorus in their different forms (e.g. ammonium, nitrate, phosphate and organic forms). Excessive nitrogen and phosphorus inputs cause eutrophication in the receiving waters by accelerating the growth of phytoplankton and macrophytes. Accumulation of organic material in the sediments causes hypoxic or even anoxic conditions and makes the sediments a poor habitat for most species (House et al. 1993). The increase of nutrients in water causes changes in the species diversity and ecosystem structure and can also lead to toxic algal blooms (Burian et al. 2001, Valtanen et al. 2014a).

Stormwater is recognized as the most important source of heavy metals (Barbosa et al. 2012). The most common metals in stormwater are copper (Cu), zinc (Zn) and lead (Pb) (Sillanpää 2013) of which Pb and Zn originate mostly from traffic. Heavy metals also originate from corrosion of buildings, atmospheric deposition, industrial activities and spills (Herngren et al. 2005). Heavy metals are readily attached to suspended solids and especially to the finer particles, which are easily transported by stormwater runoff.

In addition to suspended solids, nutrients and heavy metals, urban stormwater includes microbiological pollutants and bacteria that originate from degrading vegetation, soils and animal faeces (for example from geese, pigeons and dogs). Bacteria, such as *Shigella sonnei* and *Giardia lamblia*; viruses, such as rotavirus; and parasites, such as *Cryptosporidium*, might cause illness and be dangerous to humans (James & Joyce 2004). The presence of microbial pathogens in water may therefore restrict their recreational use (e.g. for swimming) (Mallin et al. 2008).

Pollutants can cause acute, chronic or cumulative impacts in the receiving waters. Acute effects are short-term, but can still be serious. One example of an acute impact is a large chemical spill, which can cause serious fish kills. Gradual build-up of pollutants, such as nutrients and metals, causes cumulative contamination in the receiving waterways. The way the pollutants impact is dependent also on the size of the receiving water. A small pond responds faster than a lake or a sea. Pollutants are present in stormwater in different forms. Organic matter is typically in colloidal form, heavy metals are typically dissolved and many pollutants have fractions of them attached to suspended particles (Barbosa et al. 2012). Therefore it is relevant to plan the management practices so that all forms are considered.

The discharge of stormwater into receiving water bodies has diverse impacts on the geomorphology of the receiving waters (House et al. 1993). The high peaks and runoff volumes can cause erosion and channel migration in the receiving streams. Altered bankside and riparian vegetation also cause changes in the temperature regime of the water body, which can further affect the stream organisms. Urbanization decreases the amount of coarse sediments and increases the amount of fine particles, which leads to habitat simplification (Barbosa et al. 2012, Fletcher et al. 2013).

2.2.2 Turbidity and total suspended solids

Many pollutants, like metals, pesticides and nutrients, are bound to suspended solids, which is why the total suspended solids (TSS) concentration is often used as a primary indicator pollutant e.g. in modelling (Egodawatta et al. 2007, Akan & Houghtalen 2003). Miguntanna et al. (2013) note, though, that suspended solids do not always represent other pollutants, such as nitrogen, well, because the physico-chemical characteristics are quite different. However, as total suspended solids are an important pollutant in the receiving waters also as such, the TSS concentration is used in the present study as the representative measure for water quality.

In addition to the release of contaminants transported by suspended solids, the chemical and physical alterations in water bodies caused by increased concentration of suspended solids include the depletion of dissolved oxygen in the water, temperature changes, reduction in penetration of light, which further restricts the photosynthesis of aquatic plants, and infilling of channels and reservoirs through deposition of suspended solids (Bilotta & Brazier 2008). These alterations have further unwanted effects such as higher costs in water treatment, decreased longevity of hydraulic structures such as dams and decrease in the recreational quality of the water body.

Turbidity is the measure of light scattering and, consequently, reduction in the transparency of water due to colloidal and suspended matter. It is measured using in-situ equipment, which record either the loss in intensity (i.e. attenuance turbidimeters) or scattering (i.e. nephelometric turbidimeters) of a beam of light in the water (Bilotta & Brazier 2008). Depending on the used method and equipment, turbidity is most commonly measured in nephelometric turbidity units (NTU) or in Formazin turbidity units (FTU) (Wilde & Gibs 1998). These two units are considered comparable in value.

Ideally the loads of suspended sediments coming out of a catchment would be computed by combining the flow and the concentration that are both measured continuously. However, it is not possible to continuously measure the concentration of suspended solids in water, as it requires sampling and laboratory analyses (Linjama et al. 2009). Automatic water samplers are available, but they also suffer from cost and reliability problems (Grayson et al. 1996). Measuring turbidity is not costly, the measurements are done on-site and the data logger can continuously store and send the data to the user. Turbidity is therefore the most popular surrogate measure of suspended solids.

The correlation of turbidity and suspended solids concentration in water is superficially straightforward as the more there are solids suspended in the water, the more turbid the water is. Nonetheless, there are factors that make the correlation more complicated and actually a high turbidity may not always correspond to a high concentration of suspended sediments. In addition to the amount of the particulate matter, turbidity is affected by the size and shape of the particles and the presence of phytoplankton, dissolved minerals and humic substances (Christensen et al. 2002, Bilotta & Brazier 2008). The turbidity of water from forested areas for example does not compare well with the turbidity of water from urban areas as the forests release much more humic substances in the waters than urban areas. Also the size of the suspended solids plays a key role as the large particles may not stay “suspended” in the water and the smaller particles may not be possible to be examined with the filters used in measuring suspended solids in the water.

Literature suggests that there are seasonal differences in the correlation of turbidity and suspended solids even at the same location (Grayson et al. 1996, Jones 2008). In Finland the four seasons affect the composition of sediments in stormwater. For example the snow cover and the sanding of streets in winter result in high concentrations of sand in the sewers in early spring when snow is melting. Another example would be the pollen of different trees and other plants or falling leaves that affect the characteristics of the stormwater at different times of the year.

Even two similar urban locations are not necessarily comparable regarding the turbidity-TSS correlation if the prevailing weather conditions are not the same. Yet oftentimes the most important defining factor regarding the type of particles in stormwater is considered to be the predominant land use of the catchment. The assumption behind this is that the matter that is released by stormwater is more or less the same from for example any urban area, commercial area, industrial area and forested area. In the current study the seasonal and soil type differences went unheeded as it was assumed for simplicity that land use is the key factor in determining the correlation between turbidity and suspended solids. This simplification was used in validating the turbidity-TSS correlation obtained for this study with literature values.

2.2.3 Pollutant build-up and wash-off characteristics

Various factors, from the geographical location and climatic conditions to the percentage of imperviousness and road layout, influence the generation, build-up and wash-off of pollutants in a catchment (Liu et al. 2012). The spatial distribution of urban features and how they are hydraulically connected to the drainage system are important in defining how the catchment responds to rainfall (Sillanpää 2013, Liu et al. 2012). The response of a catchment to rainfall further affects the water quality from the catchment. A large share of urban surfaces is typically covered by transport-related functions. Schueler (1994) for example reported a transport-related imperviousness of 63 to 70% at 11 residential, multifamily and commercial areas. Egodawatta et al. (2009) and Vaze and Chiew (2002) considered that road surfaces are the most important areas that cause urban stormwater quality issues. Roads are subjected to all kinds of leakages from vehicles and the road surfaces get eroded because of abrasion resulting in particles on the road surfaces that are washed off by runoff. Roads are also typically inclined in order for them to drain efficiently, which makes the stormwater to flow faster and release pollutants more easily from the surface.

The rate of pollutant build-up and the total amount of accumulated pollutants are significantly higher for roads than they are for example for roofs (Egodawatta & Goonetilleke 2008a). However, due to a coarser surface structure, pollutants from road surfaces do not wash off as easily as they do from roof surfaces (Egodawatta et al. 2009). Due to their smooth surface structure pollutants can be easily washed off from roof surfaces even during small rainfall events and therefore pollutants from roofs can be a substantial contributor to the water quality issues in urban streams. Roofs cover a relatively large surface area in many urban catchments. Egodawatta and Goonetilleke (2008a) for example reported roof surface coverages of 19, 23 and 38% for three urban locations, which they considered typical for most residential catchment. Pollutants that build up on roofs originate mostly from atmospheric deposition and are therefore finer than the particles on road surfaces (Egodawatta et al. 2009). Another important source of pollutants on roofs is the degradation of roof materials (Van Metre & Mahler 2003), which is affected by the acidity of rainfall and the cladding material.

Build-up characteristics over any surface are most affected by the surrounding land use activities, traffic and climatic conditions that dictate the quality and quantity of pollutants over and around the surfaces. The rate of build-up is higher right after a storm and reduces then gradually as the dry period continues. The majority, 80%, of total amount of pollutants build up during the first 7 dry days (Egodawatta et al. 2009).

Most studies cited in this Section are from an Australian research group, who have used simulated rainfall over roads and roofs to study the wash-off and build-up characteristics of different surfaces (e.g. Egodawatta et al. 2009). They have found that wash-off can vary with both particle size (Miguntanna et al. 2013) and particle density (Egodawatta et al. 2007). The importance of size and density of the particles is dependent on the surface characteristics. According to Egodawatta and Goonetilleke (2008b) the wash-off of pollutants is not dependent on the initial pollutant availability, but depends more on the type, structure and e.g. the slope of the catchment.

The first flush is a specific characteristic of stormwater wash-off. It is a phenomenon, where "the initial period of a runoff event produces a higher pollutant concentration peak often preceding the runoff peak" (Egodawatta & Goonetilleke 2008b). This means that a small amount of water is able to mobilize a relatively large amount of pollutants from the surfaces resulting in a high concentration in the beginning of an event, which afterwards dilutes as less mobile pollutants and at the same time more water are available. According to Van Metre and Mahler (2003) the first 2.6 mm of artificial rainfall (applied over a rooftop) mobilize most of the pollutants. When harvesting rainwater for domestic use the initial volume is commonly bypassed to avoid the pollutants resulting from the first flush. Egodawatta et al. (2009) remark, though, that because the initial stage of a rainfall usually has a smaller intensity, the particles that remain after the first flush may be mobilized later if the intensity increases. Therefore only discarding the first flush is not a sufficient treatment method.

2.2.4 Limitations of stormwater quality measurements

Uncertainty in environmental analysis arises from physical phenomena, errors in observations and errors in modelling. Every measurement has some degree of uncertainty, because measurements are only a representative of reality (Gandin 1988). Estimating uncertainty and recognizing the possible errors is important to guarantee safety and reliability in urban stormwater management. It is a difficult task, because natural processes are typically complex, non-linear, dynamic and unpredictable (Bertrand-Krajewski & Muste 2007a). The automatic samplers and other equipment used in stormwater quality measurements suffer from ice, temperature and pressure changes, gases, litter and other foreign matter and contamination, which can cause breaks in the monitoring (Linjama et al. 2009, Bertrand-Krajewski & Muste 2007b). Errors in data can be classified into two categories: random and systematic. Random errors are evident and caused by a multitude of independent factors. Random errors do not depend on the measured value as opposed to systematic errors, which are mainly caused by a scale shift of the instrument or some persistent factor which is not accounted for. Gross errors are caused by malfunctioning of the equipment or mistakes during data processing, transmission and reception, but they are usually also easily detectable (Gandin 1988).

Measurement errors affecting stormwater quality modelling can be originated from the rainfall, runoff or water quality measurements. Rainfall, runoff and, in this case, turbidity measurements are often conducted at the outlet of the study catchment. This means that the recordings at the outlet are expected to represent the processes over the entire catchment, which undoubtedly causes error. Point rainfall measurements cannot represent the real rainfall phenomenon temporally or spatially entirely correctly (Niemczynowicz 1999), but are often considered to provide sufficient estimation. Possible errors for rainfall measurements, especially tipping bucket rain gauges, include “catching problems” caused by wind (drops do not fall vertically), splashing (drops hit the measurement equipment but splash away from it) and wetting (drops adhere to the bucket and do not fall). Additionally the equipment might be subject to vandalism. Urban stormwater runoff can be erroneously measured if the amount of water in the sewers is increased or reduced by some other source than rainfall (e.g. someone washing their car outside or irrigation of a garden).

Turbidity measurements are subject to a vast array of sources of error. Stormwater transports litter that can cause blockage to the measurement equipment or cover the sensor so that it gives very high turbidity values. Dirt in stormwater can also agglomerate over the sensor gradually so that it gives increasingly higher turbidity values. Occasional high values can result from a larger object transported by the flow blinding the sensor. Many new turbidity sensors are equipped with automatic cleaning systems, but they still do not guarantee a correct measurement and maintenance at the measurement location and data inspection and verification are still needed (Bertrand-Krajewski & Muste 2007b).

2.3 Stormwater management

2.3.1 Centralized systems

Nowadays, as polluted water is a common concern, stormwater and sanitary sewerage is demanded and regulated by specific laws in all developed countries, even though there are especially rural areas that still lack the sewage systems (De Feo et al. 2014). The objective is that at least all wastewater is purified before being released into the receiving waters, and strict controls on point source polluters have been implemented. In most developed areas wastewater treatment efficiency has greatly advanced during recent decades and pollution from wastewater treatment plants has greatly diminished. For example in Europe the EU environmental legislation requires the member states to upgrade their water and wastewater treatment systems so that they can fulfil the strict requirements (De Feo et al. 2014). Urban stormwater runoff is mainly non-point pollution and thus not easily regulated. Stormwater is therefore recognized as the main source of pollution for urban receiving water bodies in developed countries (Kostarelos et al. 2011, Valtanen et al. 2014a).

In most parts of the world stormwater is collected and transported from urban areas to natural waters via an underground pipe network. Where stormwater is collected, it is transported either via a combined sewer system (CSS), where wastewater and stormwater flow together in a single pipe, or a separate sewer system, where the waters flow in separate pipes. Both of the mentioned sewage systems have their advantages and disadvantages. CSS is the older system of the two and is common for example in old European city centres. The most important advantage of CSS is that both wastewater and stormwater are transported to a treatment facility where they are purified before being released into the environment. The biggest disadvantage of CSS is the occurrence of combined sewer overflows (CSOs), which arise when the capacity of the sewers or the treatment facility is exceeded during heavy rainfall or snowmelt. CSOs can be frequent, occurring for example once every three days (American City & County, 1996). As the high volume of stormwater flows together and mixes with wastewater from households and industries in CSS, CSOs contain, in addition to stormwater, human faeces and industrial wastes that end up directly polluting the receiving waters (US EPA 1999).

Mainly because of CSOs the separate sewer system began advancing in the late 1870’s. In Finland, the first separate sewer system was introduced in 1938 in the capital, Helsinki, and from then onwards the separate system was implemented in most new areas around the country

(Sillanpää 2013). In Lahti the most part of the sewage system consists of separate sewers, only the older part of the city centre having a combined sewer network (Lahti Aqua 2014). In the separate sewer system most commonly only wastewater is treated and the stormwater is led directly to a nearby natural water body. Thus, compared to the combined sewer system, the amount of water to be purified in the treatment facilities is smaller, which makes the wastewater treatment more feasible. Collecting stormwater separately from wastewater also reduces the risks of CSOs, which is one of the main reasons to favour the separate sewer system and in some areas the already existing CSSs are therefore being replaced by separate sewer systems (US EPA 1999). However, transforming a CSS to a separate sewer system is neither simple nor cheap. Furthermore, as already previously stated, untreated stormwater from urban areas carry various pollutants that, if not removed, end up degrading the receiving waters (Barbosa et al. 2012, US EPA 1999). The pollutants from urban surfaces and the pollutants in wastewater are different and have different impacts, which is why stormwater should be dissociated from CSOs when talking about pollution originating from different urban waters. This study concentrates specifically on the simulation of surface pollution transported by stormwater.

2.3.2 Decentralized systems

Careful planning and reasoned management of urban stormwater are needed to avoid the adverse hydrological and ecological changes they cause in the urban area and in the receiving waters; conveying peak flows into the sewage systems and “out of sight” is not enough. During recent decades the management of urban runoff already at the site by decentralized solutions (as opposed to the traditional centralized systems where stormwater is collected to a centralized collection system and managed off-site) has become a growing trend (Barbosa et al. 2012, Sillanpää 2013). These integrated approaches in stormwater management take into consideration the quantity, quality and amenity management as three equally important aspects. The aim is to mitigate changes and minimize all detrimental effects caused by land use development as well as sustain evapotranspiration, groundwater recharge and re-use of stormwater at the site (Sillanpää 2013, Aaltonen et al. 2008). Fletcher et al. (2013) recognized that a more natural water balance in urban environments benefits also the liveability of cities and stormwater is increasingly considered as a resource instead of a pure nuisance.

The new stormwater management systems have different denominations according to their focus and the country they were developed in. Low Impact Development (LID) and Best Management Practice (BMP) are the most common. Herein these management methods are called Best Management Practises or BMPs. The BMPs can be any type of management practices, structural or non-structural, such as political decisions, pollutant prevention strategies, street cleaning or rainwater retention (Barbosa et al. 2012). Examples of structural BMPs are porous pavements, storage basins, green roofs, rain gardens and vegetated swales (US EPA 2014c).

Stormwater quality can be mainly managed by affecting either the urban hydrology and the stormwater quantity or by affecting directly the pollutants and how the stormwater can transport them. Street sweeping is a common practice and is efficient in removing litter and large particles from the surfaces which are then not transported to the outlet of the catchment. However, as street sweeping removes only the larger particles from the street surfaces, the finer particles that for example carry the majority of heavy metals are left on the roads to be readily removed by stormwater. (Herngren et al. 2005) Street sweeping on its own is therefore not sufficient for stormwater quality management.

When water is retained in the catchment before it enters the drainage network, the volume of water reaching the outlet is smaller and the stormwater flows are temporally more evenly distributed so that the risk of flooding is reduced. Storage basins usually only delay runoff from the catchment, but depending on the design (that is whether the basin has an impervious or pervious bottom), the quantity of stormwater to be managed can also be reduced. Porous pavements, rain gardens and vegetated swales both delay and reduce the runoff peak through infiltration and evaporation. In addition to the storage capacity, these structures filter and clean water that percolates into groundwater. Growing vegetation uses water and nutrients and has an important role in reducing water quantity and improving water quality.

Building green roofs is a trend at the moment, but their peak flow reduction is less than for the other BMPs (Lee et al. 2010). Yet they have effect and they also contribute to the amenity of the urban area by diversifying the landscape. Some BMP structures can be made visible and designed to serve as decorations or monuments, such as fountains or other more constructed water elements. The best result is gained when the management controls are distributed throughout the catchment and combined strategically.

2.4 Stormwater quality modelling

Gaining more knowledge of stormwater quality and developing water quality computation tools is important now as urbanization is accelerating rapidly. Flood control requires the stormwater to be directed away from cities while rainwater harvesting has become an interesting solution alternative for the growing need for potable water. Therefore the quality of both stormwater directed away from cities and stormwater collected for the use of cities, need to be modelled. Roofs in urban areas would be the primary source for rainwater harvesting and therefore modelling the water quality also from roof surfaces is important (Egodawatta et al. 2009). Climate change poses a further challenge for future stormwater quality management (Fletcher et al. 2013) as the behaviour of stormwater and the stormwater quality strongly depend on the climate conditions.

Effective stormwater management and treatment designs require stormwater quality estimation prior to implementation of the management systems (Akan & Houghtalen 2003) and modelling is the most convenient estimation tool (Egodawatta & Goonetilleke 2008b). A stormwater quality model combines mathematical procedures in describing the changes in stormwater quality as a response to a rainfall-runoff event in a catchment. In stormwater quality modelling the pollutant build-up, wash-off and transport processes are mathematically represented by a set of equations. Many mathematical formulations are available to represent the processes with varying levels of accuracy.

The quality of stormwater is a sum of a multitude of factors, which makes stormwater quality modelling a complex task and requires simplifications in the construction of models. Of the stormwater quality processes wash-off is the least investigated (Egodawatta et al. 2007), but modelling build-up is also a challenge, because the processes depend on many varying factors. Pollutant build-up and wash-off processes vary widely for different kinds of surfaces (e.g. roads, roofs, parking lots and playgrounds) in different environments. If the processes need to be described in detail, a separation should also be made in some cases between different materials of the same type of surface (e.g. roofs). However, often the required data is limited and therefore all these things cannot be taken into consideration and even crude generalizations are needed. Yet to advance in the field of stormwater management the physical, chemical and biological processes should be understood better (Obropta & Kardos 2007).

Pollutants in urban stormwater are usually regarded to generate on surfaces that are aggregated as a single land use category, such as ‘industrial’, ‘commercial’ or ‘residential’ areas. This division is crude as there is a great variety in the existence and portions of different specific surface types within these land uses in different locations. The pollutant processes are not consistent within the same land use as site-specific characteristics play an important role in how the pollutant build-up and wash-off are formulated (Liu et al. 2012). Pollutant generation in stormwater from a commercial area in Beijing is not necessarily even close to that from a commercial area in Helsinki. Research on the generation of pollutants in stormwater from specific types of surfaces such as roads, roofs, parking lots or different kinds of pavements has not been widely conducted, so these kinds of generalizations and aggregations have still been inevitable (Brodie & Dunn 2010). This is probably the reason why stormwater modelling is limited to a very general level and considering aggregated land uses like commercial and residential areas is still the predominant way of studying stormwater quality. No examples of stormwater quality modelling using a high spatial resolution were found in the literature.

Model calibration has typically been done in large catchments with mixed land uses (e.g. Cho & Seo 2007, Temprano et al. 2006) or for very specific small plots (e.g. Egodawatta & Goonetilleke 2006, 2008a), but not much research has been conducted for catchments that are large enough to represent more than one surface type, yet specific enough to provide results that are transferable to other locations with the same kind of land use. In this study it was considered that the selected catchment is representative of an urban area in the northern countries and the results could therefore offer a benchmark for future modelling purposes.

Studies have been conducted on the relationship of event mean concentrations (EMC) of pollutants, most often total suspended solids (TSS), and different kinds of rainfall-runoff characteristics (runoff volume, runoff intensity, rainfall intensity, rainfall depth etc.) over some specific surface types, such as roads, highways and roofs (Gilbert & Clausen 2006). Using only EMCs, however, does not give much information about the dynamics of the stormwater quality response to runoff. The first flush phenomenon for example cannot be captured if constant EMC is used. In this study the dynamics of fluctuations in the water quality were studied with the use of an exponential function that relates the pollutant concentration to runoff nonlinearly.

3 Material and methods

The methodology followed in this thesis is presented in Figure 1 and explained in this chapter.

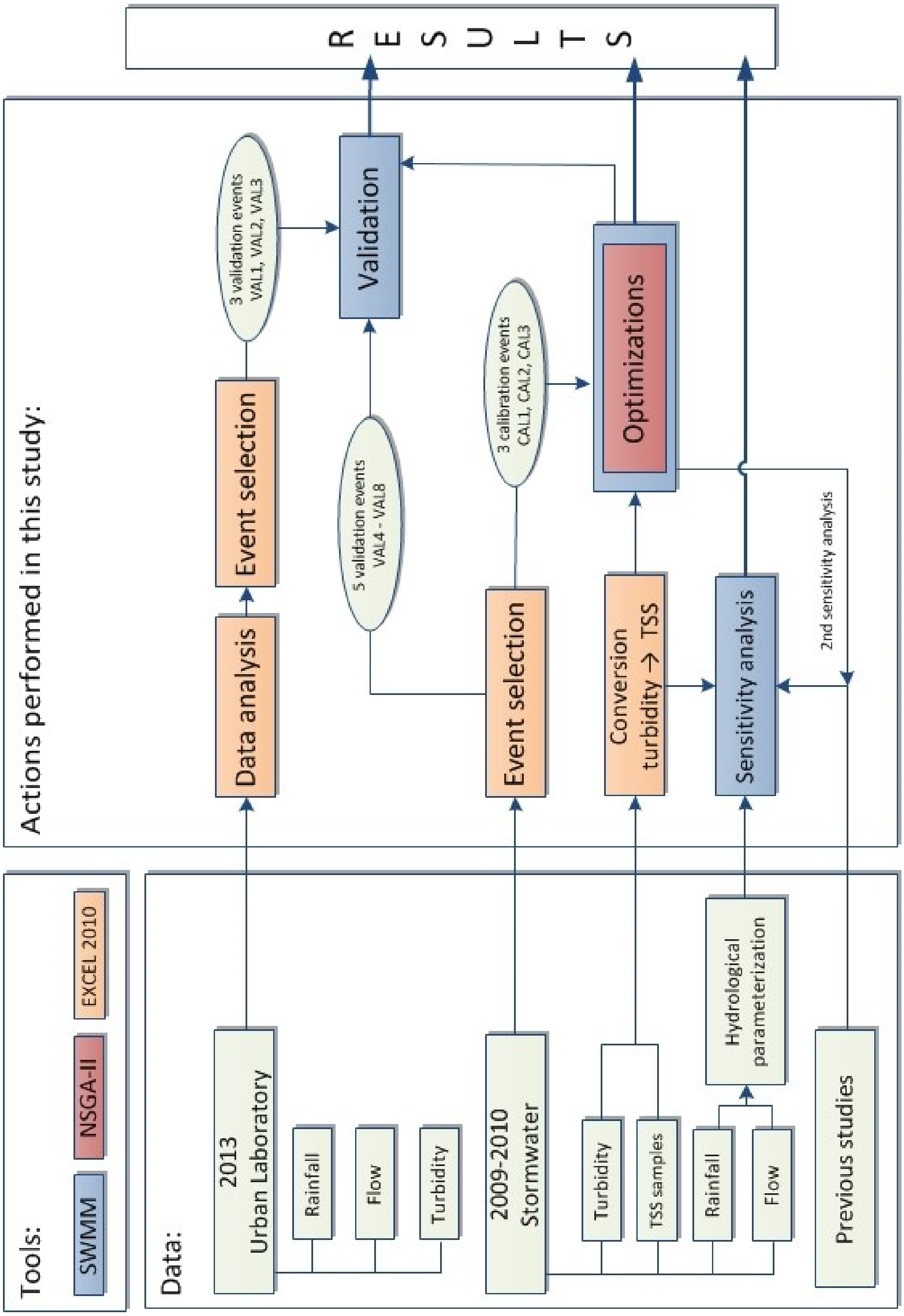


Figure 1. *The tools, data and methodology. Data is presented in light green and blue, red and orange are used to connect the action and the tool it was performed with.*

3.1 Study site

The study site is located in the city of Lahti in southern Finland, 100 km northeast from the Finnish capital, Helsinki. Lahti inhabits 102,000 people and belongs to the boreal climate zone. The annual means of precipitation and air temperature are 633 mm and 4.1 °C, respectively. The study catchment called Ainonpolku (AP) is situated close to the city centre and is urbanized with apartment blocks, a few office buildings and detached houses. It is 6.63 ha in size and has an imperviousness of 54%. The catchment area is drained through separate sewer system that is comprised of circular pipes with a total length of 3.23 km (Krebs et al. 2014). The study catchment is one of several catchments that have been studied in the Lahti area for the purposes of urban stormwater research. The earlier data collection (2008-2010) was conducted within the framework of the Stormwater research programme (Valtanen et al. 2009) and the later studies (e.g. Pajari 2014, Tikkanen 2013) are part of the Urban Laboratory project that started in 2012. More information about the study catchment can be found in (Valtanen et al. 2014a).

Before this study was initiated, the hydrological part of the SWMM model had already been calibrated for the AP catchment by Krebs et al. (2014). The calibration was conducted with high spatial resolution (the catchment was divided into 784 subcatchments that are illustrated in Figure 2) and had yielded successful results. The calibrated model was used as the hydrological parameterization and formed the basis for the water quality model in this study. The details of the hydrological model calibration can be found in Krebs et al. (2014).

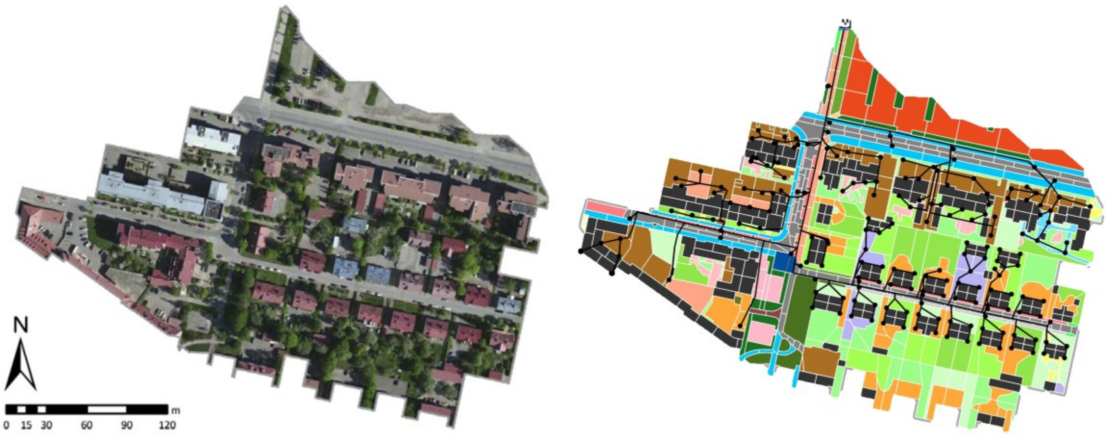


Figure 2. *An aerial image of the Ainonpolku study catchment (left) and the division of the catchment to subcatchments for modelling purposes (right). Pictures provided by Gerald Krebs.*

3.2 Data

3.2.1 Stormwater programme data 2009–2010

The data used in this study from the Stormwater programme (Valtanen et al. 2009) covered the period from 2009 to 2010. It included rainfall data at one-minute interval at the outlet of the Ainonpolku catchment, 10-minute rainfall data from Lahti Science and Business Centre (LSB) situated 2.7 km from the study catchment, and hourly and daily recordings from Lahti-Laune, a station operated by the Finnish Meteorological Institute (FMI) situated 3.3 km from the study catchment. Locations of the weather stations in Lahti in relation to AP are presented in Figure 3. Turbidity and flow were measured at the outlet of the AP catchment at one-minute recording interval. Further details of the measurements at the catchment can be found in Valtanen et al. (2014b).

3.2.2 Urban Laboratory project data 2013

The data that was collected as part of the Urban Laboratory project consisted of discharge and turbidity measurements at the outlet of the AP catchment. A tipping bucket precipitation recording has taken place at Ainonpolku as well, but the measurements started only at the end of August 2013 and the measuring interval was three hours, which was a too coarse time resolution to support the purposes of this study.

The discharge was measured with a Nivus OMC Pro CF flow meter at one-minute recording interval. The measurement error is less than +/- 2 mm for the water level and <1% for the flow velocity. The flow meter uses ultrasound to determine the water level and delay line method with cross-correlation to determine the flow profile (Kestävän ympäristön kaupunkilaboratorio 2014). The automatic turbidity measurements were conducted with a Thermo Scientific RT 114 gauge. The data sets from Ainonpolku were stored using an online database for the project.

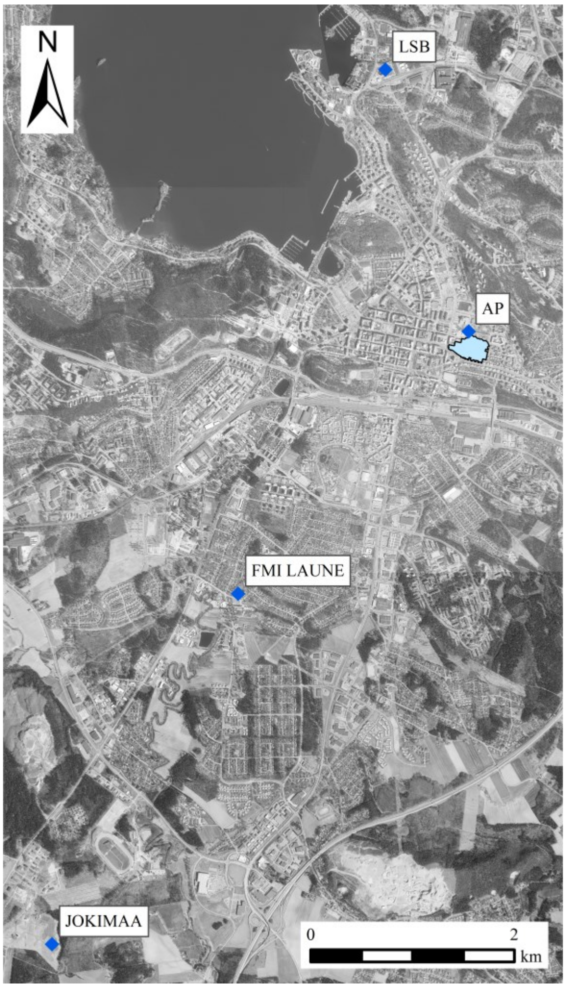


Figure 3. *The rain stations in Lahti in relation to the Ainonpolku catchment (area marked).*

3.2.3 Additional data

To replace the missing rainfall at Ainonpolku from 2013, close-by rainfall measurements were compiled. The available additional data were hourly and daily rainfall data from Lahti-Laune, one-minute data from LSB, and ten-minute rainfall data from Jokimaa, a weather station situated 7.21 km from the catchment. The rainfall intensities at Jokimaa and LSB were measured by a Vaisala WXT520 weather transmitter that detects the impact of raindrops hitting the sensor cover and converts the signal to volume (Vaisala Oyj 2012). The Laune hourly rainfall data was provided by the Finnish Meteorological Institute. Rainfall data at ten-minute interval from Jokimaa was received from the University of Helsinki.

3.3 Data analysis

The hydrological data for 2009-2010 was checked and processed by Krebs et al. (2014) and was used as such in this study. In the selection of suitable calibration events from the 2009-2010 data, the representativeness of the turbidity measurements was determined by visual inspection and by estimation of the logic of the turbidity response to runoff. For example, if turbidity did not change when there was significant runoff, it was concluded that the turbidity measurements were not successful and the event was not selected. Some potentially faulty events were however included in the validation along with healthy looking events, because comparing the model output against questionable data can give new insights about how the model works. These events also fell between the successfully measured events and were therefore easy to simulate.

The discharge and turbidity data collected in 2013 was scrutinized to look into its quality and reliability, as it was raw data with no previous data quality assurance. First the data sets were separately inspected for missing or unrealistic values. Microsoft Excel (version Office 2010) was used in the processing of the data. The rainfall data from different locations in Lahti had been examined earlier, and were now further examined to see whether the dynamics of rainfall and runoff also match the dynamics of turbidity. Scatter plots of rainfall against runoff were examined first over the whole monitoring period and the data was separated into events. Only events that had a reasonable rainfall-runoff correlation were chosen for a more detailed analysis of consistency between rainfall and runoff and between runoff and turbidity.

3.4 Correlating turbidity with total suspended solids

The concentration of total suspended solids can be used as a representative for stormwater quality. Because continuous measurement of concentrations in stormwater is infeasible, continuous turbidity measurements were used in the Urban Laboratory project to indicate stormwater quality. Suspended particles scatter light that goes into water. The more there are particles in the water the more light is scattered and the more turbid the water is. Hence there is a correlation between the turbidity and the total amount of suspended solids in the water, which is widely acknowledged in the literature (e.g. Bertrand-Krajewski 2004, Al-Yaseri et al. 2013). The correlation between turbidity and TSS concentration is not the same everywhere, though, because the turbidity of water is dependent on for example the material and shape of the particles which are for their part dependent on the soil type and many other characteristics of the catchment.

According to the Finnish Environment Institute (2014) it can be broadly estimated that 1 FNU is equivalent to 1 mg/l of suspended solids. This is a very broad generalization and a more precise correlation was sought for. This was obtained with the help of samples that had been taken intermittently during the runoff events of interest. The TSS concentrations that were analysed from the samples were plotted as a function of their corresponding turbidity measurements and linear correlation estimation was generated. The obtained correlation was backed up with correlations from previous studies from similar study areas.

3.5 US EPA Stormwater Management Model 5.0 (SWMM 5.0)

3.5.1 Hydrology

The EPA Stormwater Management Model (SWMM), developed by the United States Environmental Protection Agency (US EPA), is one of the most widely used urban water quality models (Borris et al. 2013, Chow et al. 2012, Obropta & Kardos 2007). It is used for both single event and continuous simulation of quantity and quality processes of stormwater from primarily urban catchments.

The model is divided into Atmosphere, Land Surface, Groundwater and Transport compartments. The Atmosphere compartment defines the precipitation and pollutants from air that are included as inputs to the system via Rain Gage objects. The Subcatchment objects represent the Land Surface compartment that receives data from the Atmosphere compartment and conveys infiltration data to the Groundwater compartment and surface runoff (including pollutant loadings) to the Transport compartment. The Transport compartment contains all kinds of conveyance elements like pipes, channels and pumps that transport water to the outlet or to treatment facilities. Predefined inputs can be used so that every compartment does not need to be in use. (Rossman 2010)

The catchments in the model are considered as nonlinear reservoirs, which receive water from rainfall and from nearby and adjacent catchments (Krebs et al. 2013). The water infiltrates, evaporates and turns into runoff in the catchment according to the individual catchment properties. Infiltration occurs only from pervious surfaces and runoff is only produced when the water depth in the catchment exceeds the storage depth of the catchment that is the sum of surface ponding, surface wetting and interception capacity.

3.5.2 Water quality

The water quantity and quality parts of the SWMM model are independent from each other so that calibration and validation of the hydrology can and is recommended to be done separately from the calibration of the water quality model. Stormwater quality is modelled by simulating the build-up and wash-off processes of pollutants inside the catchment. In the model the pollutants accumulate on the catchment surfaces during dry weather according to a build-up function, and a wash-off function dictates how the pollutants wash off during wet weather. The build-up and wash-off functions include coefficients, which have to be provided as input parameters.

The user can decide whether the build-up of pollutants during dry weather is governed by a power function, an exponential function or a saturation function. The build-up power function describes a mass of pollutant build-up accumulation (B)

$$B = Min(C_1, C_2 t^{C_3}) \tag{1}$$

where C_1 is the maximum build-up (mass per unit area or curb length), C_2 is a build-up rate constant, t is the elapsed time and C_3 is a time exponent. The larger the time exponent C_3 is, the faster pollutants build up. Also the larger the rate constant C_2 is, the faster the maximum is reached.

The saturation function describes a build-up that declines until a saturation value is reached,

$$B = \frac{C_1 t}{S + t} \quad (2)$$

where S is the half-saturation constant (days to reach half of the maximum build-up).

The exponential function describes a build-up that approaches the maximum limit asymptotically,

$$B = C_1(1 - e^{-C_2 t}) \quad (3)$$

where C_1 is the maximum build-up (herein abbreviated as MB) in mass per area or curb lenght, C_2 is the build-up rate constant (1/d) and t is the elapsed time (d). The curves of the three different build-up functions are graphed in Figure 4. In this study the exponential function was chosen for build-up. The build-up rate constant is called herein the build-up exponent (BE).

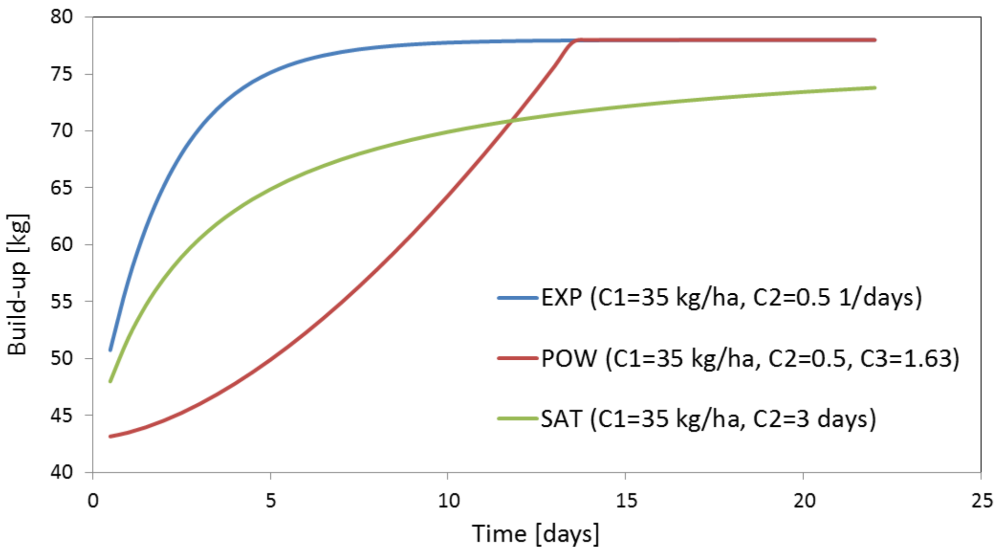


Figure 4. *The build-up curves with a power function (POW), exponential function (EXP) and saturation function (SAT).*

The properties associated with the wash-off of pollutants during rainfall-runoff events are described by an exponential function, a rating curve or event mean concentration (EMC). Rating curve wash-off describes a wash-off load (W) (in kg/s when using SI units) that is proportional to the runoff raised to a power,

$$W = C_4 Q^{C_5} \quad (4)$$

where C_4 is the wash-off coefficient, C_5 is the wash-off exponent and Q is the runoff rate in flow units. The EMC wash-off is the rating curve wash-off function where the exponent is 1, and the coefficient C_4 is the wash-off concentration in kg/l when using SI units.

The exponential wash-off function describes a wash-off load (W) (in kg/h when using SI units)

$$W = C_4 q^{C_5} B \quad (5)$$

where C_4 is the wash-off coefficient (herein abbreviated WC), C_5 is the wash-off exponent (herein abbreviated WE), q is the runoff rate per unit area (mm/h) and B is the total mass of built-up pollutant in mass units, not in mass per area as in the build-up functions (Equations 1-3). (Rossman 2010)

The exponential function (Equation 5) was selected for this study, because it is the only function that is dependent on the accumulated build-up and it can produce a nonlinear wash-off curve. The rating curve can also produce a nonlinear wash-off curve, but it is independent from build-up. The EMC wash-off curve is linear and independent from build-up, which means that it only shows a predefined concentration multiplied by the simulated flow and therefore basically gives no additional information. A fictional discharge curve and associated pollutant wash-off curves calculated with the described wash-off equations are presented in Figure 5.

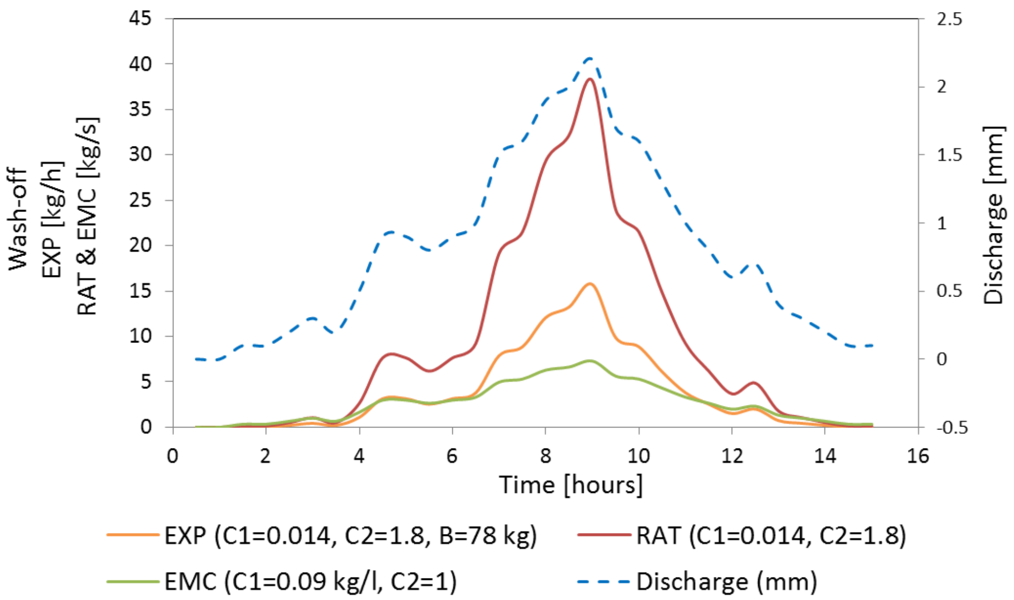


Figure 5. *The wash-off curves expressed by the exponential function (EXP), the rating curve function (RAT) and the event mean concentration (EMC).*

The amount of pollutants available in the beginning of the simulation is either given directly as input to the model or calculated by the model according to the user-defined number of dry days prior to the beginning of the simulation. If the number of dry days in the latter case is set to zero, the build-up will start to accumulate from zero at the start of the analysis.

Each subcatchment in the model is assigned the land uses it contains along with the percentages of land use cover inside the subcatchment. The user can define as many different land uses as needed. The resolution of the model can therefore be altered according to the needs of the user and the available data. Land use in this context can be for example industrial, residential or commercial, when the resolution is coarse, or street, pavement, roof, parking lot, green etc., as in the case of this study, or even more precise, when for example roofs of different material could be named as different land uses. In this study the division of the whole catchment into subcatchments was detailed and all the subcatchments consisted of only one land use category.

The land uses can be assigned the same or different build-up and wash-off functions and the parameters of the functions for each land use can be different. If the functions and the parameters are the same, pollutants build up and wash off similarly in every subcatchment. In the present study all impervious subcatchments were given the same build-up and wash-off functions (both exponential) and also the same set of associated parameters. The pervious subcatchments were not given any build-up or wash-off functions to ensure that pervious areas would not contribute to the loads. The difference in the contributions to concentrations was thus mainly defined by the hydrology and not by the build-up and wash-off characteristics of different subcatchments.

There is an option to impose street cleaning information and pollutant removal efficiencies associated with best management practices (BMP) on the land uses. The user can define whether there is street sweeping and how large a fraction of pollutants on the surface is removed. As there was no information on street sweeping or BMPs at Ainonpolku, they were neglected in this study.

All pollutants that are created in the model are given a name, a concentration unit and the concentrations of the pollutant in rain water, ground water and in any infiltration or inflow. This means that in addition to the build-up on the surfaces the pollutants in the outflow of a catchment could originate from ground water or rain water or any infiltration that could be for example a single polluter somewhere inside the catchment. The build-up and wash-off qualities are land use dependent, but the concentrations in rain water, ground water and infiltration/inflow water are the same over the whole catchment. If there is a pollutant whose concentration is a fixed fraction of another pollutant concentration, this relationship is also possible to be given in the model. In this study TSS was defined as the only pollutant type. Only build-up on the surfaces was considered and all the other abovementioned means of pollution accumulation were set as zero.

3.6 Goodness-of-fit criteria

The performance criteria to evaluate the model performance during the sensitivity analysis, the calibration and the validation were adopted from the hydrological parameter optimization of Krebs et al. (2014). The model performance criteria were the Nash-Sutcliffe efficiency *E* (Equation 6), the linear correlation coefficient *CORR* (Equation 7) and the per cent total load error *TLE* (Equation 8). The sum of squared errors *SSE* (Equation 9) was used as objective function in the calibration along with the linear correlation coefficient. The per cent volume error VE (Equation 10) is used to represent the model performance when the hydrological part of SWMM is addressed.

$$E = 1 - \frac{\sum_{i=1}^{n}\left(C_{o,i} - C_{m,i}\right)^2}{\sum_{i=1}^{n}\left(C_{o,i} - \overline{C_o}\right)^2} \tag{6}$$

$$CORR = \frac{\sum_{i=1}^{n}\left(C_{o,i} - \overline{C_o}\right)\left(C_{m,i} - \overline{C_m}\right)}{\sqrt{\sum_{i=1}^{n}\left(C_{o,i} - \overline{C_o}\right)^2 \sum_{i=1}^{n}\left(C_{m,i} - \overline{C_m}\right)^2}} \tag{7}$$

$$TLE = \frac{L_o - L_m}{L_o} 100 \tag{8}$$

$$SSE = \sum_{i=1}^{n}\left(C_{o,i} - C_{m,i}\right)^2 \tag{9}$$

$$VE = \frac{V_o - V_m}{V_o}100 \tag{10}$$

where $C_{o,i}$ and $C_{m,i}$ are the observed and modelled TSS concentrations in [mg/l], respectively, $\overline{C_o}$ and $\overline{C_m}$ [mg/l] are the observed and modelled mean concentration values, respectively, L_o and L_m [g] are the observed and modelled total load, respectively, V_o and V_m [mm] are the observed and modelled total discharge volume, respectively, and n is the number of observations. Wash-off loads for each minute are obtained by multiplication of the concentration by the flow value of the minute. The simulated concentration and flow are used in calculating the simulated load, and the observed concentration and flow are used for the measured load. The total loads are calculated as a sum over the duration of one event or the whole calibration period.

The use of the Nash-Sutcliffe efficiency and linear correlation coefficient has been suggested by Moriasi et al. (2007). *TLE* for water quality simulation corresponds with the deviation of runoff volumes for hydrological modelling that was suggested by the ASCE Task Committee (1993).

3.7 Sensitivity analyses

Sensitivity analyses are conducted to find out how sensitive the model is towards the parameters of interest. Often a sensitivity analysis is conducted before calibration to better understand the structural reliability of the model (Mustonen 1986) and to be able to decipher the appropriate parameters for the calibration. If the model is very insensitive towards some parameters, they should be left outside of the calibration. On the other hand, the sensitive parameters are better to be included in the calibration.

Often the sensitivity analysis is conducted again with the calibrated parameters, because the parameters may have changed notably from the initially guessed values. The sensitivity of a model towards a parameter might be dependent on the parameter value in the way that the closer the parameter value is to the real value the more sensitive the model is toward it. If the sensitivity is tested far from the real value, alterations in the parameter value do not necessarily affect the simulation result relatively as much. According to Mustonen (1986) the purpose of an *a posteriori* sensitivity analysis is to study what kinds of outputs are possible with certain parameter distributions. In the present study the sensitivity analyses served as tools to increase the understanding of the interrelationships of the parameters.

In this study the influence of five water quality parameters and three hydrological parameters on the simulation results (the TSS model output) was investigated with sensitivity analyses before and after the calibration. The sensitivity analysis for the hydrological parameters was conducted previously by Krebs et al. (2014) who noted that the model performance regarding hydrology is dominated most by the depression storage and the Manning's roughness coefficient for closed conduit flow. These two parameters and the Manning's n for overland flow were added in the first sensitivity analysis of this study to investigate their effect on the water quality results. The actual quality parameters investigated in both sensitivity analyses were the initial build-up (IB) of pollutants on the surfaces and four parameters (the maximum build-up (MB), the build-up exponent (BE), the wash-off coefficient (WC) and the wash-off exponent (WE)) that appear in the exponential build-up and wash-off functions (Equations 3 and 5).

Three consecutive events in July 2009 were selected for the sensitivity analyses and the model calibration. The first event started on 7 July and the last one ended on 12 July. The event lead times were 10, 17 and 20 hours, the rainfall depths were 3.5, 22.0 and 23.7 mm and the peak runoff flows were 13.3, 196.0 and 252.0 l/s for the first, the second and the third event, respectively. The dry time between the first and the second event was 27 h and the dry time between the second and the third event was 43 h. The basic information about the calibration and validation events is presented in Table 1.

Table 1. *The rainfall events selected for calibration (CAL) and validation (VAL). VAL1, VAL2 and VAL3 are from the Urban Laboratory project and the other events are from the Stormwater programme data.*

Name	Date	Duration [h]	Peak flow [l/s]	Rainfall depth [mm]	Observed discharge [mm]	Antecedent dry time* [days]	Total load [kg]	EMC [mg/l]
CAL1	7. - 8.7.2009	10	13.3	3.5	1.1	6.5	7.8	105.6
CAL2	9. - 10.7.2009	17	195.9	22.0	11.6	1	133.0	172.3
CAL3	11. - 12.7.2009	20	252.0	23.7	11.9	2	88.2	112.3
VAL1	13. - 14.6.2013	26	45.0	10.0	4.8	2.5	192.2	605.3
VAL2	13. - 15.8.2013	33	128.4	75.8	26.4	2	1658.2	947.4
VAL3	17. - 18.10.2013	29	56.6	13.8	9.7	7	75.1	117.1
VAL4	13. - 14.7.2009	24	72.0	17.5	8.3	1	57.1	103.3
VAL5	24. - 25.7.2009	9	41.8	7.0	3.3	3	16.6	75.4
VAL6	31.7. - 1.8.2009	18	217.7	10.8	5.2	6.5	62.2	181.4
VAL7	29. - 30.8.2009	23	143.6	17.9	7.1	2.5	53.4	113.5
VAL8	13.6.2010	12	31.1	8.8	3.5	0.5	40.1	175.5

** dry time defined as 12 h during which no more than 1 mm of rain has occurred*

For the purposes of the first sensitivity analysis, the analysis was set to start on 27 June when the previous rainfall event had taken place before the calibration events. The model was run first with initial guesses for the parameters (for fast build-up and intermediate wash-off), which were obtained from an earlier study by Borris et al. (2013). The values of the quality parameters were then perturbed independently by ±20% and ±50% keeping the rest of the parameters fixed and a model run was performed after every perturbation. The Manning's roughness coefficient for conduits was perturbed by ±30%, the Manning's *n* for overland flow was perturbed by ±40% and the depression storage by ±50% keeping the other parameters unchanged.

In the second sensitivity analysis one set of calibrated parameters were taken as the initial conditions and each parameter was independently perturbed by ±50% after which the model was run and the results compared with the calibration result. Only the quality parameters were included in the second sensitivity analysis and the hydrological parameters were left out.

The linear correlation coefficient, the Nash-Sutcliffe efficiency and the total load error were selected as the goodness-of-fit criteria for the sensitivity analyses and calculated for every model run. The goodness-of-fit criteria were calculated for all events together and for every event separately, and compared with the corresponding values obtained with the initial parameter set.

3.8 Optimization of parameters

The sensitivity analysis, calibration and validation of the stormwater quality model in this study were conducted using the hydrological parameterization of Krebs et al. (2014) for the same study catchment.

Generally the calibration of a stormwater quality model is a process where the deviation between the observed and simulated pollutant load in stormwater is minimized by adjusting build-up and wash-off parameters. SWMM does not provide loads as a time series output, but a time series of pollutant concentration can be examined. In this study the simulated TSS concentration with each tested parameter set was thus compared to the TSS concentration that was converted from the observed stormwater turbidity at the outlet of the study catchment.

The parameters in the exponential build-up and wash-off functions presented in Section 3.5.2 were chosen for calibration. The initial build-up, herein defined as the amount of pollutants on the surfaces of the catchment at the beginning of the simulation, was decided to be included in the calibration in addition to the maximum build-up, the build-up exponent, the wash-off coefficient and the wash-off exponent.

In SWMM there are two ways to assign an initial build-up; either by giving it directly to each subcatchment or by letting the model calculate the build-up from zero according to the number of dry days over which the pollutants build up on the surfaces before the beginning of the simulation. The initial build-up from the study catchment was unknown and it is very difficult to estimate as there were no measurements from the catchment surfaces. The initial build-up could not be given to the model as input, but neither was it desired to start accumulating pollutant build-up from zero, because practically the catchment surfaces are never completely clear of pollutants (Wang et al. 2011). Assigning a certain build-up to the subcatchments would be as much speculation as letting the pollutant build-up start from zero. Therefore the initial build-up was considered as a calibration parameter. Another way of getting past the issue of initial build-up would be to start the analysis so much before the first calibration event that the initial conditions would not have effect. This would, however, only raise another dilemma, as it is not known how long the initial build-up affects the simulation results. Neither is it feasible to start the simulation too much before the first calibration event as it makes the calibration procedure longer.

The calibration procedure used the genetic multi-objective optimization algorithm NSGA-II (nondominated sorting genetic algorithm II) (Deb et al. 2002). The NSGA-II and SWMM were linked through an R-script originally developed by Peter Steinberg/Herrera Environmental Consultants and modified by Krebs et al. (2014) to suit the purposes of this study. The objective functions for the optimizations were chosen to be the linear correlation coefficient (or Pearson's correlation coefficient) (Equation 7) multiplied by -1 and the sum of squared error (Equation 9). The objective was to minimize both of these functions, so that a perfect fit for the observed and simulated concentration curve would get a value of 0 for the *SSE* and a value of -1 for the -*CORR*.

In the optimization each parameter could fluctuate between predefined minimum and maximum values. These parameter boundaries are difficult to estimate as the maximum build-up and initial build-up are the only values that are actually physically measurable and the rest are mathematical coefficients. Measuring values for MB and IB is possible, but laborious and approximate, and finding values for the other parameters needs to be done by trial runs with field data or based on values obtained for other similar catchments, which is also very approximate as the parameter values vary a lot for different catchments.

In this study, three different sets of the abovementioned parameter boundaries were applied in the optimization of the parameters. The first set (called Range Set 1, RS1) was determined according to values obtained in earlier studies from similar urban catchments. Range Set 2 (RS2) was applied after RS1. The extremes of every parameter range in RS2 were meant to be set so far from the expected parameter value that the best fit could be selected freely. Range Set 3 (RS3) was the last one to be applied. For RS3 the parameters whose values altered from the allowable minimum to the maximum in the previous optimizations were given an even larger range and the parameters whose values did not alter that much, were given a smaller range. The third range set needed to be used, because unexpectedly many of the optimizations got stuck in the extremes of the previously determined ranges. To avoid unrealistic optimization results ensuing from mostly overlapping ranges of the initial and maximum build-up, a punishment function was added in the R-script, so that the initial build-up could never end up larger than the maximum build-up. Otherwise all parameters were adjusted independently inside the defined ranges.

The three events that were used in the sensitivity analysis were also selected for the optimization. To see the impact of the number and type of events on performance statistics and the parameter values, the optimizations were conducted with different combinations of the three calibration events. Each of the events independently, every two-event combination and the three events together were therefore optimized using Range Set 1. To see the effect of the parameter ranges on the performance statistics and the parameters the first event, the first and the second event and all three events were optimized with the three different parameter range sets.

In addition to the continuous calibration sequence, the simulation time was further reduced by using the Hot Start File (HSF) possibility in SWMM. A Hot Start File is a binary file, which contains the hydraulic and water quality information for the drainage system at the end of a run (Rossman 2010). HSF used in this study was made by running a simulation starting on 27 June, when the previous significant rainfall event took place, until 7 July just before the start of the first calibration event and saving the simulation result as HSF. As no water quality variables were assigned in creating HSF, it only contained the hydrological information. Before using the HSF in the actual optimizations it was verified that the model produces the same hydrological outcome for the calibration events with HSF and without it when the simulation is started much earlier. Using HSF saved the time the model needs for running the simulation from 27 June until 7 July and because the genetic algorithm requires thousands of simulations, the time saved is substantial. The same HSF was used as the initial conditions for all the optimizations (except for O14) so that the calibration results are comparable.

Runoff and turbidity data were only available from the catchment outlet and therefore the processes inside the catchment could not be studied closer. One of the assumptions used in this study was that impervious surfaces are the predominant source of suspended solids in stormwater runoff and the pervious areas do not contribute to water quality at the outlet. The pervious area of the study catchment, which was 31.5% of the total area, only contributed 1.3% of the total runoff to the catchment outlet. This supports the exclusion of pervious areas. Only the surface types that are mainly impervious were thus assigned a land use in SWMM where build-up and wash-off of pollutants could occur. The subcatchments whose surfaces were covered with vegetation were not assigned any land use and did not contribute to load accumulation or wash off. In contrast to loads, water can still flow to and from them as defined by the hydrological part of the model. The other subcatchments that had different types of surface (roof, pavement, street, parking lot) were assigned 100% coverage of a land use that in this study was called "Not Green". This "Not Green" land use was given the build-up and wash-off properties dictated by the Equations 3 and 5 and their relevant parameters. This means that the accumulation and wash-off characteristics of all the other subcatchments than the green areas were the same. The differences between the subcatchments and the individual contribution of each of them to the total wash-off of the study catchment are thus dictated by the hydrological characteristics of the subcatchments.

3.9 Validation of calibration results

Five events from the Stormwater programme data from 2009-2010 and three events from the Urban Laboratory data from 2013 were simulated for the purpose of validation of the calibrated model. These events included periods that did not look very promising regarding the cohesion of rainfall, runoff and turbidity, but they were included because of the additional information they could offer. The simulation of 2009 and 2010 events was done as a continuous simulation starting from the first calibration event using the same Hot Start File as in the calibration (see Section 3.8). The 2013 validation period was simulated from 30 May 2013 onwards without any initial build-up as there was no data from 2011 and 2012.

The validation period of 2009-2010 was simulated with four parameter sets (optimizations OR1, O7, O10 and O13). The 2013 validation period was only simulated with two parameter sets (optimizations OR1 and O13). The calibration and validation results were evaluated by the Nash-Sutcliffe efficiency, the linear correlation coefficient and the total load error. The goodness-of-fit criteria were separately calculated for every validation event.

4 Results and discussion

4.1 Data analysis

Data analysis for the Stormwater data from 2009-2010 can be found from earlier publications (Krebs et al. 2013, Krebs et al. 2014, Valtanen 2009). The data analysis for the Urban Laboratory data is explained in the following chapters.

4.1.1 Discharge and turbidity data from Ainonpolku 2013

To find suitable runoff events to be used in the calibration and validation of the SWMM quality model, and to check the data for possible measurement errors or other problems, the raw data that had been collected for the Urban Laboratory project in 2013 was examined.

In the analysis of the data from 2013, the discharge and turbidity data were first separately reviewed. It was observed that in both time series approximately once every three hours a time stamp with one minute accuracy was represented twice. The minute was mainly, but not always, the same in the discharge series and in the turbidity series, so the two series needed to be gone through separately. This error was found to extend over the whole year of data in both the discharge and the turbidity data.

When the time stamps were viewed with one second accuracy, it was noted that the "double" time stamps actually fell on different seconds of the minute, which means that the two values represented the same minute, but both of the data points were real. It was assumed that these extra time stamps resulted from an error in the recording of the measurement device (e.g. incorrect sensor programming) and the extra data point, which is always the latter point of the two on the same minute, were manually removed from both the discharge and the turbidity data.

In addition to the extra time stamps, the discharge data was missing some time stamps, but much more sparsely than the extra stamps. These isolated missing values could be due to inappropriate sensor operation or for example maintenance or repair (Bertrand-Krajewski & Muste 2007b). If the missing time stamps were isolated, a value was interpolated between the value of the previous minute and the following minute. If there were longer periods of time missing in the data, these were naturally not surveyed in selecting suitable events for this study. There were inexplicable peculiarities in the discharge data, such as a 24 hour skip from the time stamp of 5.6.2013 7:12 to the time stamp of 6.6.2013 7:13. All these irregularities were identified from the data and the event selection was made considering only the quality-assured parts of the data.

The events for the discharge quantity analysis were first chosen based on how consistent the Jokimaa rainfall data was with the discharge. Closer inspection of the selected events revealed that there were occasional discharge drops to zero in the middle of a rainfall event. There was thus clearly some data missing also when the time stamps were correct. Figure 6 presents one of these occasions, when suddenly at 4:47 the flow drops to zero in the middle of an event.

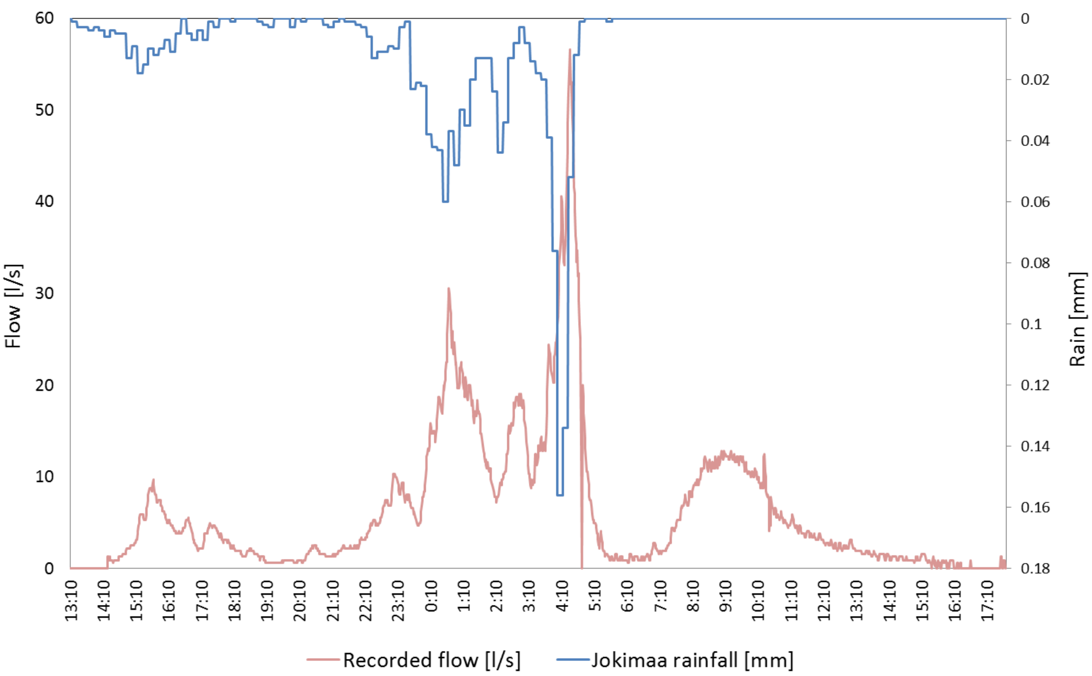


Figure 6. *An example of a missing data point (at 4:47) in the middle of an event.*

In addition to the double time stamps and the occasional missing data, there were the following oddities in the turbidity data. The turbidity values in AP in 2013 alternated widely, from less than 100 NTU to nearly 4000 NTU. The very high stable turbidity values (3900 NTU lasting over a day as an example) could be the consequence of some larger particle, for example a leaf, stuck on the sensor. However, when the turbidity stayed abnormally high, but was still fluctuating, the high values were not caused by any fixed blockage. In this case dirt covering the turbidimeter could explain the high turbidity values. However, regular cleansing of the measurement equipment took place during the whole monitoring period, which should have prevented the dirt impact on the readings.

4.1.2 Rainfall data from Lahti 2013

As there was no rainfall data available from the AP catchment, other close-by collected rainfall measurements were tested. The observed discharge data from Ainonpolku was compared to rainfall data from Lahti-Laune, Lahti Science and Business Centre and Jokimaa measuring stations. Rainfall data from LSB had to be completely discarded because of unsystematic errors and a gross underestimation of the rain volume throughout the whole monitoring period. Also the rainfall data from the FMI station Lahti-Laune was not used as the hourly recording interval resulted too coarse for modelling the water quality. The Jokimaa rainfall data was the most usable for the purposes of this study. Both timing and volume of the rain in Jokimaa fit well to the AP discharge during most events and the resolution of approximately 10 minutes was considered to capture the event characteristics.

The raw rainfall data was examined similarly as the discharge and turbidity data. The recording interval remained mainly in 10 minutes apart from some irregularities most likely caused by maintenance breaks. All the data was converted to one-minute data by dividing the rainfall volume of each time stamp evenly over the previous ten minutes. If the recording interval changed from ten minutes to some other amount of minutes (in this case the recording interval altered between 8 and 25 minutes), the volume of that time stamp was divided accordingly.

The conversion to one-minute rainfall data ensured that SWMM input requirements are met. SWMM interprets that every rainfall value is constant over the recording interval that is specified for the rain gage in question (Rossman 2010). SWMM required that the recording interval remains unchanged during the simulation. The only errors are made when the data from several minutes is distributed evenly although the rain might have fallen more sharply, but this could only be avoided with more precise measurements.

During the monitoring period, May – October 2013, both rain and discharge data were missing at some points. Typical data breaks lasted from a couple of hours to several days.

4.1.3 Event selection 2013

A rainfall-runoff event was defined by Sillanpää (2013) to start from the beginning of precipitation and to last until the flow rate has returned to the pre-event baseflow rate. This definition was applied in this study with the distinction that as there was no baseflow at the studied catchment in 2013, an event was considered to end when the discharge returned to zero.

Because rainfall data from a relatively distant location (7.21 km from the study catchment) had to be used in the simulations, it was necessary to check first whether the rainfall produced the same kind of discharge as had been recorded. If the simulated stormwater runoff would not be within acceptable error, the stormwater quality could not be simulated successfully. Some events that looked consistent at first when inspected in a large scale had to be discarded in the end owing to data breaks in either the discharge data or the rainfall data. For many events it could be easily stated that the Jokimaa rain and AP discharge did not fit together and they had to be left out of the study. When there were no obvious mismatches in the recorded rainfall and discharge data, the SWMM hydrological parameterization of Krebs et al. (2014) was used to produce the discharge. The simulated discharge response to rainfall was well accurate for the Stormwater programme data from 2009-2010 so it was concluded that the earlier hydrological model parameterization was reliable. If, then, the discharge simulated using the rainfall from Jokimaa in 2013 did not fit to the recorded discharge in Ainonpolku, it was concluded that the recorded discharge was in reality produced by a different rainfall than the one recorded in Jokimaa.

Most of the events from 2013 had to be discarded due to incoherence already after inspecting the rainfall and the discharge. An example is presented in Figure 7, where it is observed that the pattern of the simulated stormwater flow is nearly the same as the recorded (the correlation coefficient is high, $CORR$ = 0.89), but the scale is different, which shows as an unacceptable Nash-Sutcliffe efficiency E = 0.04 and a large volume error VE = 81.8%. The measured runoff coefficient for this event would be 0.24, which is lower than the runoff coefficient (0.44 with a standard deviation of 0.17) defined by Valtanen et al. (2014b) for the study catchment.

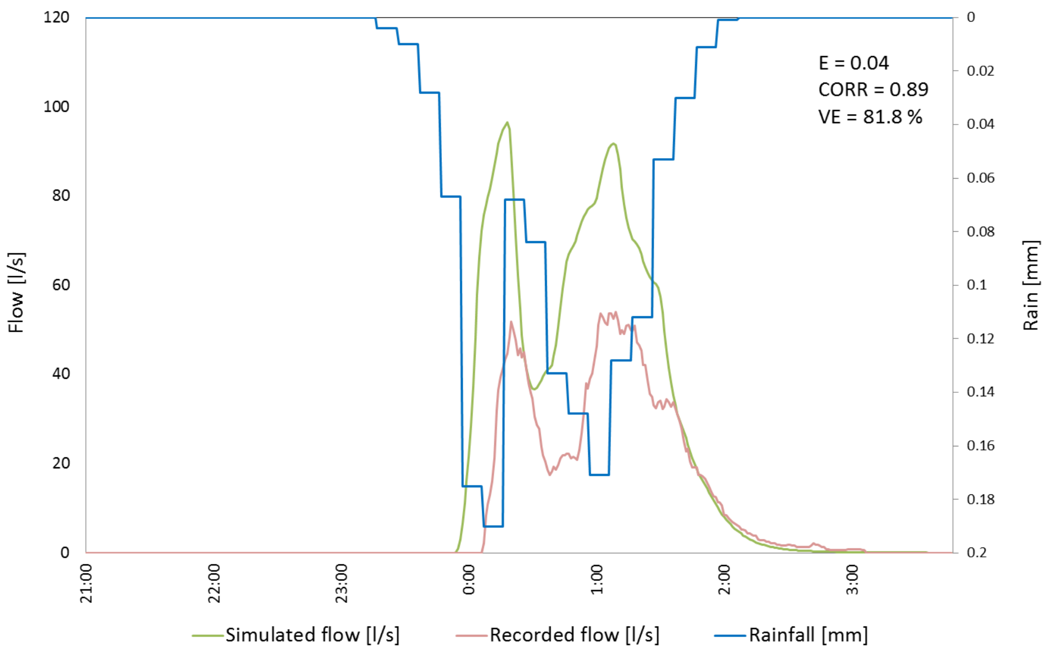


Figure 7. *An event from 2013 (8 Aug 2013 21:00 – 9 Aug 2013 3:50), which was discarded from further inspection due to a scale difference in the recorded and simulated stormwater runoff.*

Three events from 2013 showed a reasonably good match between rainfall and runoff, which is the prerequisite for a reasoned stormwater quality simulation (Figure 8, Table 1). The recorded rain of these events in Jokimaa and discharge in AP fit by visual inspection well together, although the simulated flow does not produce exactly same kind of response. The best fit according to the Nash-Sutcliffe efficiency and the correlation coefficient is gained for the event VAL1. The good *E* value (0.78) is due to the good fit of the high peaks, even though some of the low peaks in the middle of the event are not produced in the simulation. The fit in VAL2 looks visually good, although the *E* is only 0.28. The low *E* value is most likely due to the poor fit of the high peaks, contrary to VAL1. The VAL3 has an overall good fit with an *E* of 0.65 and a *CORR* of 0.84. The volume error becomes fairly large (39.7%) because of the latter part of the event, where there is observed flow, but no rainfall.

The problem with using Jokimaa rain data with Ainonpolku discharge data is foremost the considerable distance between them. Even though the measurements in both locations were accurate, they did not necessarily fit together because of the distance. The larger the distance between two measuring points, the more there are changes in the distribution and water content of the nimbi and accordingly in the rainfall pattern. Between Ainonpolku and Jokimaa there are differences in altitude that affect the formation of clouds more than flat ground. Even the rainfall dynamics were the same over Jokimaa and over Ainonpolku, it cannot be known for sure whether the rain fell over both locations at the same time or if there was a lag, and where the rain fell first. The selected events VAL1, VAL2 and VAL3 are likely due to a large uniform rain front, when the timing and volume of the rainfall are roughly the same in Jokimaa and in Ainonpolku.

Because of the drawbacks of the 2013 data, it was used only in the validation of the quality model calibration. After the rainfall and discharge fit was examined, the turbidity data was taken into examination. The turbidity data for the validation events VAL1, VAL2 and VAL3 is presented and examined in Section 4.6 about validation.

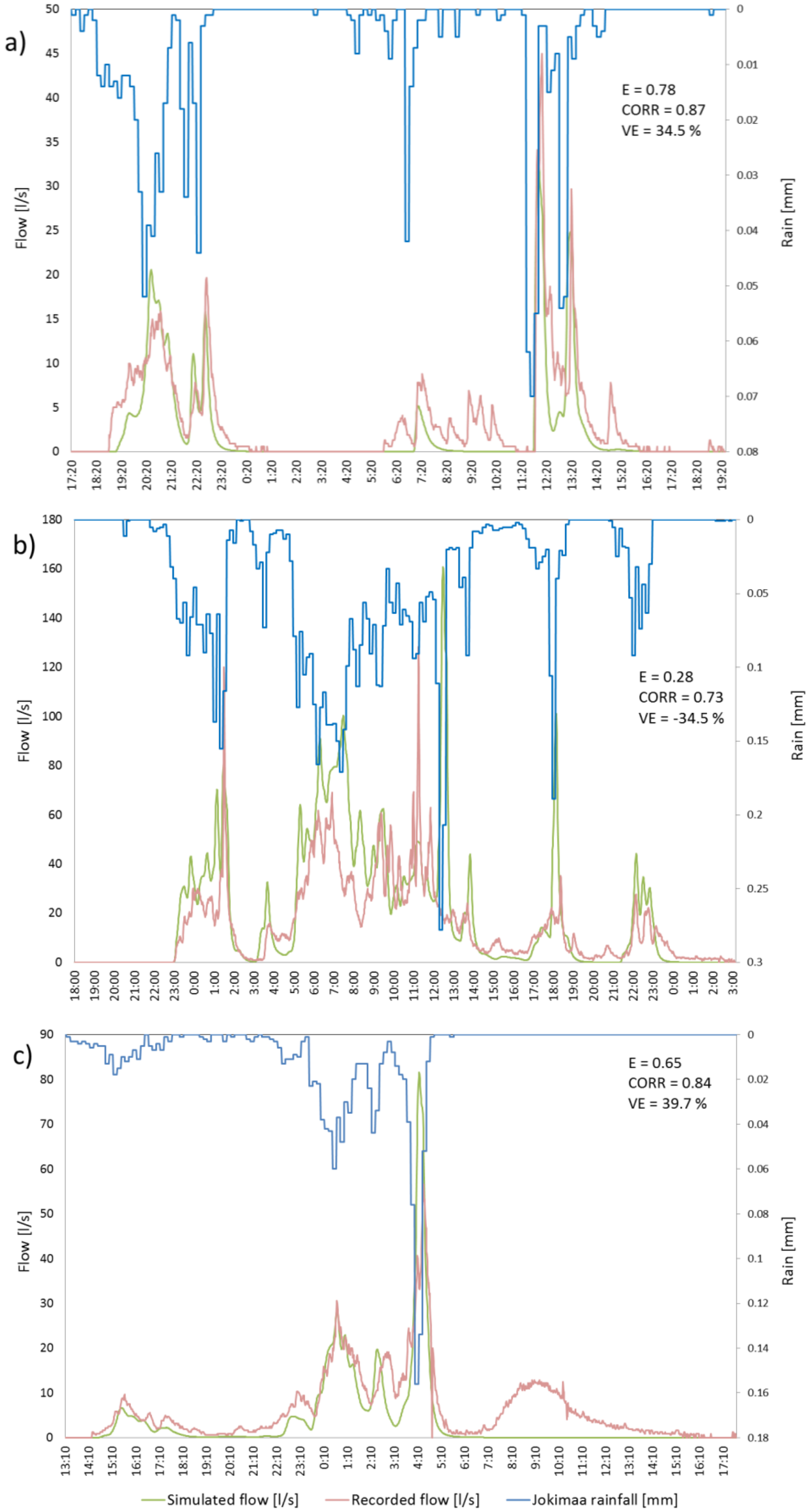


Figure 8. *Rainfall recorded at the Jokimaa station, flow recorded at the Ainonpolku catchment outlet and flow simulated with the Jokimaa rainfall for the validation events VAL1 (a), VAL2 (b) and VAL3 (c).*

4.2 Turbidity and TSS correlation

Line et al. (2013) reported that total suspended solids estimates from turbidity measurements would not correlate well with reality for low TSS values (or for low turbidity values). Due to presence of dissolved material in water a small value of turbidity may correspond to zero concentration of suspended solids.

Altogether 54 TSS assays had been made during rainfall-runoff events in 2009-2010 of which 29 were used with their corresponding turbidity values to create a scatter plot from which the correlation between turbidity and TSS concentration could be obtained. Not all of the assays were used, because by visual inspection of some events it could be noted that there was no correlation between the two variables. It was assumed that no correlation was an indication of an error with either the turbidity measurements or the TSS measurements.

The intercept of the linear regression equation was set to zero, because it was assumed that when there are no suspended solids in the water it cannot be turbid either. The graph that shows the correlation of turbidity and the TSS concentration is shown in Figure 9.

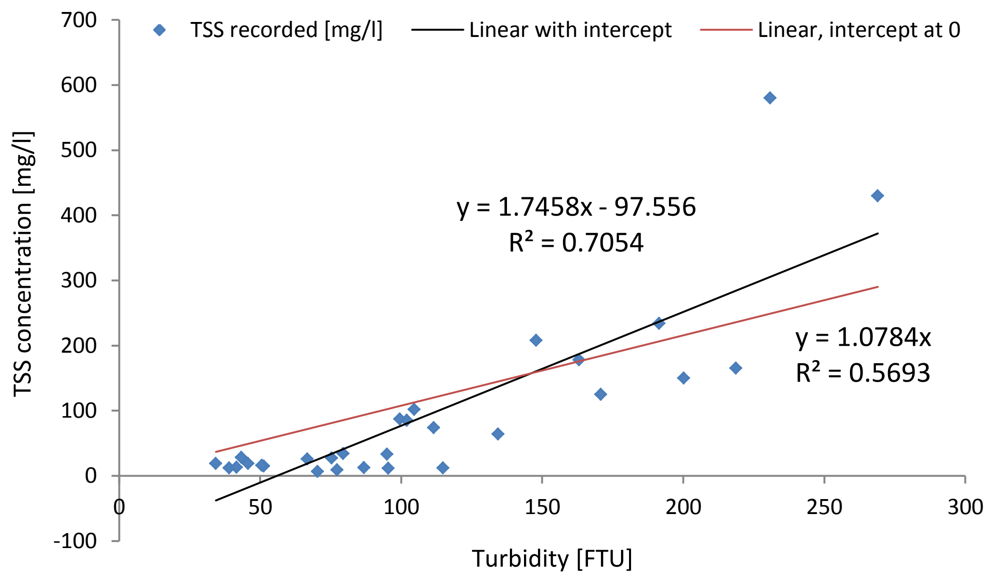


Figure 9. *Turbidity-TSS correlation when intercept is settled freely and when intercept is forced to 0.*

The correlation equation for turbidity and TSS thus became

$$C_{TSS} = 1.0784T \tag{11}$$

where T is the turbidity in the given unit. The correlation was acceptable with an R^2 value of 0.57 and a p-value less than 10^{-8}, within the criteria for significance. Fixing the intercept to zero in this case reduces the influence of the small TSS values on the regression slope, which was in accordance with the suggestions of Line et al. (2013). Equation 11 was selected to be used in the conversion of the measured turbidity to total suspended solids.

The relationship in Equation 11 is straightforward and simplified, and it corresponds well with previous studies. Settle et al. (2007) obtained a linear relationship between suspended solids and turbidity with a multiplier (for turbidity) of 1.06, and Métadier and Bertrand-Krajewski (2012) developed a polynomial function that gives a similar correlation as Equation 11. TSS for the first calibration event calculated with the three different above-mentioned functions are presented in Figure 10, where it can be observed that the outcome does not differ substantially. A direct correlation for turbidity and total suspended solids has been widely used, although the values of the coefficients differ.

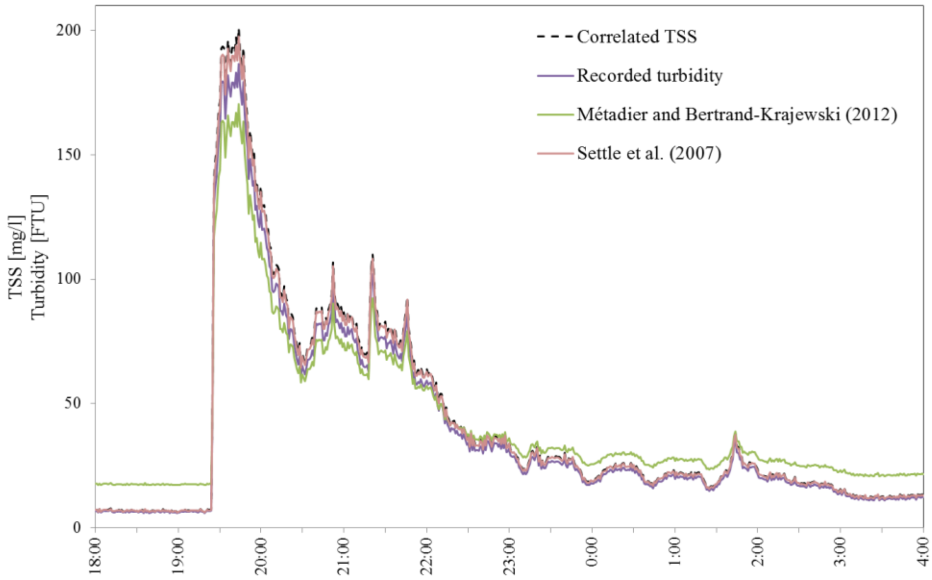


Figure 10. *The recorded turbidity and the TSS correlated according to two correlations from literature and according to the correlation determined in this study for the first calibration event.*

The correlation for turbidity and total suspended solids concentration (Equation 11) was determined with the data from 2009, when the turbidity values were lower than the turbidity values in 2013. The extrapolation of Equation 11 outside of its calibration data (Figure 9) may affect the results in the validation made with the 2013 data.

4.3 Calibration events

The three calibration events CAL1, CAL2 and CAL3 are presented in Figure 11. The hydrological parameterization of Krebs et al. (2014) had yielded good results as can be observed in the graphs where the Nash-Sutcliffe efficiency, linear correlation and volume error of the simulated flow are presented. It can be observed that the correlated TSS concentration does not differ much from the recorded turbidity.

The simulated flow nicely follows the observed flow, with Nash-Sutcliffe efficiencies over the satisfactory limit of >0.50, linear correlations over 0.70 and volume errors less than 20%. These events were selected for the sensitivity analyses and the calibration not only because of good model performance but also because a continuous sequence of events enabled the completion of multi-event simulations in feasible runtimes. A sequence of events also allows for build-up calibration, which is essential in the stormwater quality model.

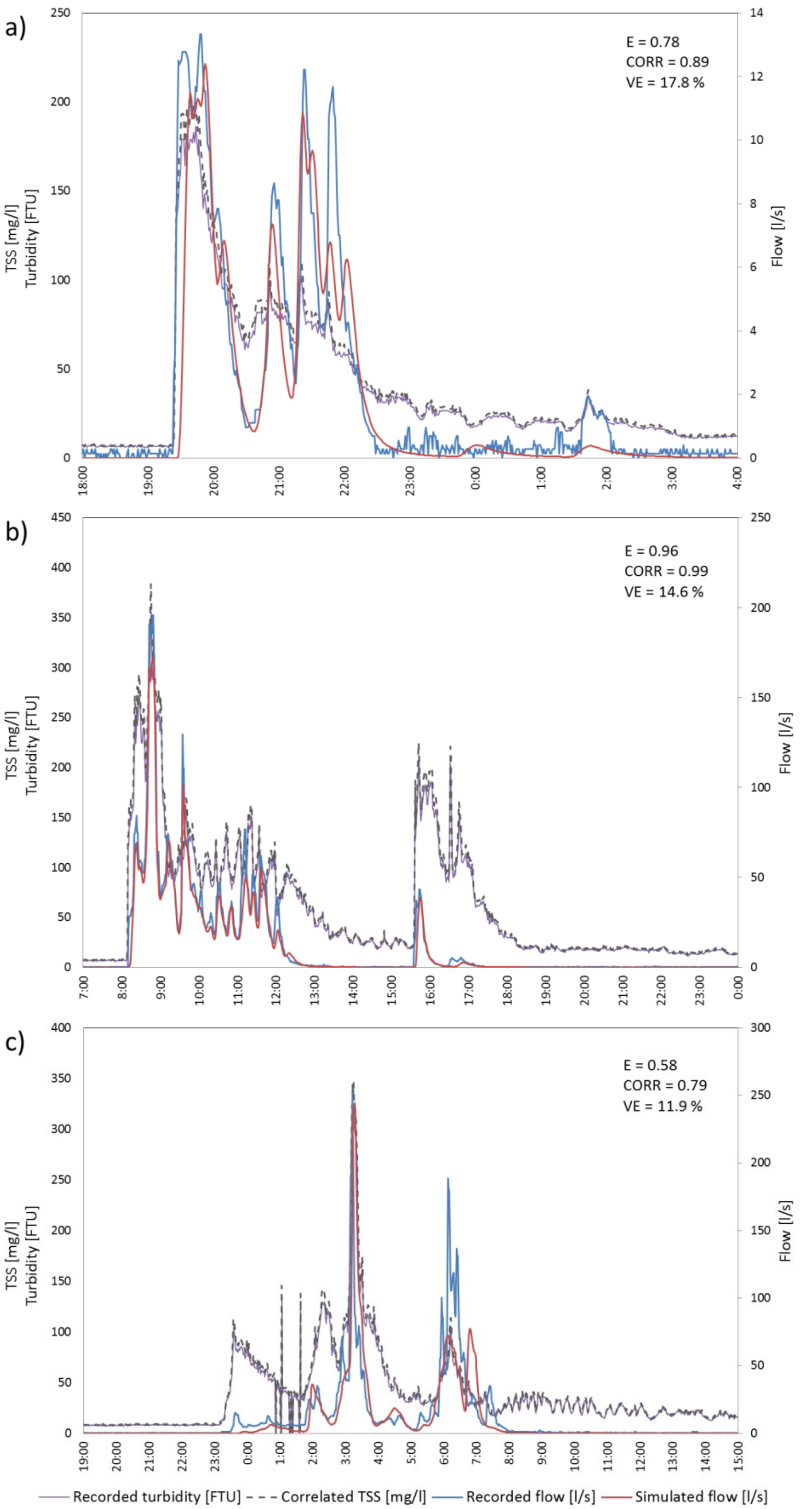


Figure 11. *The recorded turbidity, correlated TSS concentration, recorded flow and simulated flow of the calibration events CAL1 (a), CAL2 (b) and CAL3 (c) with the Nash-Sutcliffe efficiency, linear correlation and volume error (VE) of the simulated flow.*

4.4 Sensitivity analyses

4.4.1 Initial values

In this study two sensitivity analyses were conducted; one before calibration and another after calibration. *Before calibration* the model sensitivity was investigated for five quality parameters and three hydrological parameters, which were the Manning's roughness coefficient for conduit flow (MnC) and for overland flow (MnO) and depression storage (Stor). These three hydrological parameters were included, because according to Krebs et al. (2014) they were the hydrological parameters towards which SWMM was the most sensitive (affected the runoff the most) in the study catchment. The simulated pollutant concentration is directly dependent on runoff to a power in the exponential function that was selected to simulate pollutant wash-off (Equation 5). Changes in flow may thus have a considerable effect on the wash-off of pollutants and therefore it was considered important to investigate the sensitivity of the most important hydrological parameters as well. The initial values for the hydrological parameters were obtained from the hydrological parameterization of Krebs et al. (2014) for the study catchment.

The sensitivity analysis of quality parameters before calibration was conducted with values obtained from literature, because no information existed on the parameter values from the study catchment. Furthermore, the quality parameters are difficult to estimate without measurements and modelling as only the maximum and initial build-up are physically bound and the rest of the parameters are purely mathematical. Borris et al. (2013) studied a typical residential catchment in Northern Sweden with a separate sewer system and an imperviousness of 35%. The catchment resembled the current study catchment enough so that the set of quality parameters they used for the quality parameters could be used as initial values in this study. Fast build-up rate and intermediate wash-off rate were chosen, so that the build-up exponent in Equation 3 was 0.9 and the wash-off coefficient of Equation 5 was 0.04. Borris et al. (2013) did not take the initial build-up into consideration.

In the first sensitivity analysis of this study the simulations, which were used to investigate the sensitivity the other parameters than the initial build-up, started 27 June with a zero initial build-up. The build-up at the beginning of the first calibration event on 7 July (that is the actual calibration parameter initial build-up, IB) was calculated by running the simulation from 27 June until the beginning of the first calibration event and by dividing the remaining build-up with the total impervious area and by assigning the result of the division on all subcatchments that contribute to runoff. In testing the sensitivity of the model towards the initial build-up this was the parameter that was then perturbed and the Hot Start File was used so that the simulation of water quality could be initiated from the beginning of the first calibration event directly.

The sensitivity analysis *after calibration* was conducted with the optimized parameter values of the optimization O10. This optimization was selected for the sensitivity analysis, because it was conducted for the whole calibration sequence, it had yielded good results and the parameter values were suitable. For example the build-up exponent had been optimized to a value of 5, whereas in the optimizations O13 and O7 it was optimized to a value of 8 and 0.08, respectively. It was regarded convenient to conduct the sensitivity analysis with a value falling between the minimum and maximum of optimized values. The hydrological parameters were not included in the second sensitivity analysis as the first sensitivity analysis revealed them insensitive.

Table 2 presents the parameter values used in the sensitivity analysis before calibration and the optimization O10 used in the sensitivity analysis after calibration, and Figure 12 presents the performance against the observed concentration. It can be noted that the simulated concentration curve produced with the initial values is too high for the first event, but the pattern fits nicely. The simulation also fails to reach the high peak of the third event, but the total load error for the third event is relatively small. The good fit in the second event implied that the parameters are suitable for the first sensitivity analysis as no calibration had yet been conducted. The optimization O10 gives good results for all the calibration events. A comparative inspection of the two sensitivity analyses for each parameter follows in the next Sections. The performance statistics of the sensitivity analyses are presented in Appendix 1 and Appendix 2.

Table 2. *The initial quality parameter values in the first (Initial guess) and the second (O10) sensitivity analysis.*

	IB	BE	MB	WC	WE	MnO	MnC	Stor
Initial guess	34.5	0.9	35	0.04	1.15	0.001-0.667	0.011-0.015	0.1-4.82
O10	262.9	5	270.4	0.001	1.456	-	-	-

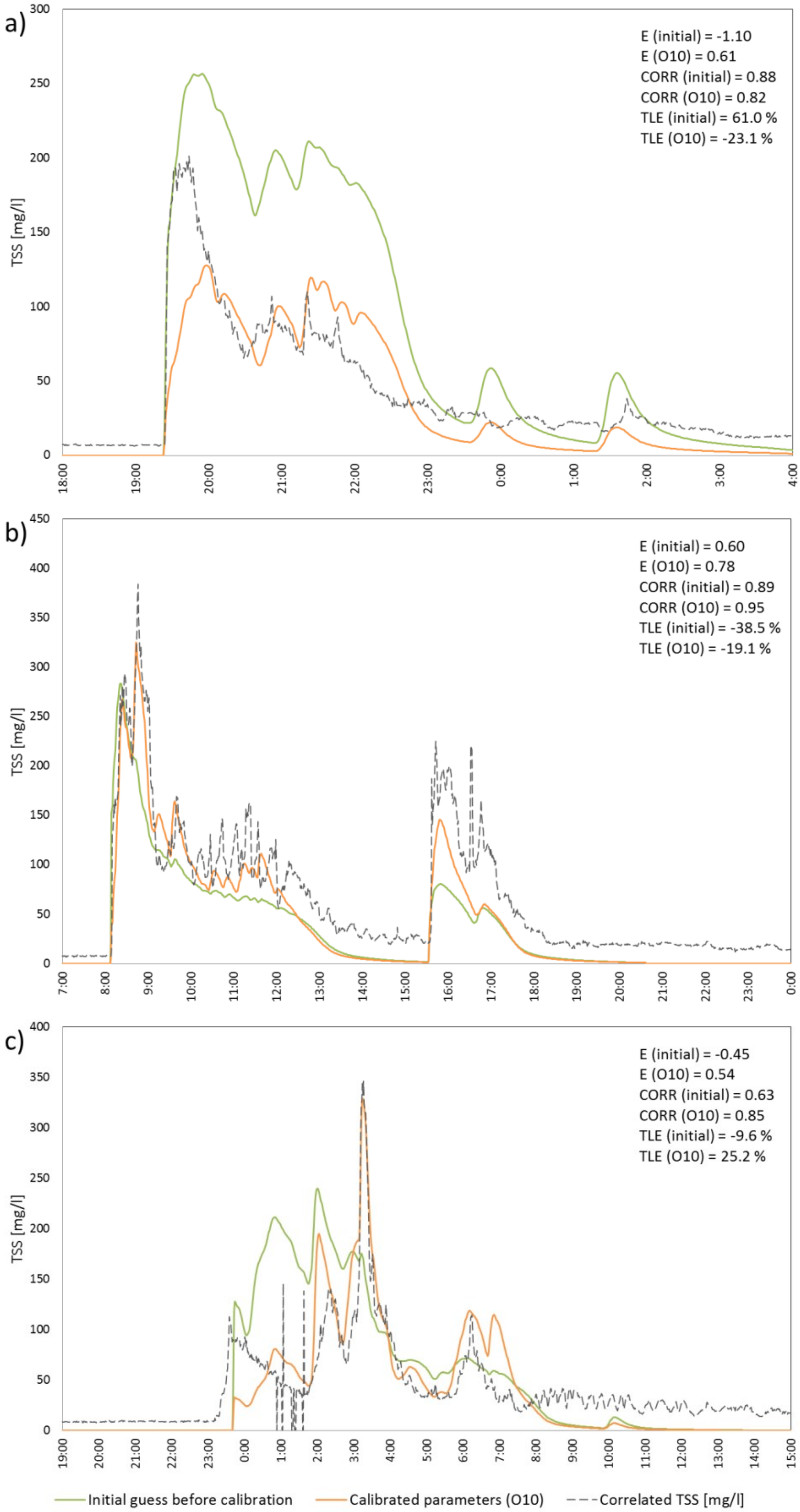


Figure 12. *The simulated TSS concentration with the initial values for the parameters before the calibration and with the optimized parameters obtained in optimization O10 with their corresponding Nash-Sutcliffe efficiencies, linear correlations and total load errors for CAL1 (a), CAL2 (b) and CAL3 (c).*

4.4.2 Maximum build-up

Figure 13 presents the effect of perturbations in the maximum build-up in the first sensitivity analysis. The maximum build-up had the largest effect on the total load of suspended solids of all the parameters. A deviation in MB changed the total loads and peak loads as much as the change in the parameter. That is a 20% change in the parameter value caused a 20% change to the same direction in the concentrations and thus also in the total load. This means that the fraction of pollutants that is removed by stormwater runoff remains the same, but the total amount of pollutants can be scaled up by altering the maximum build-up. The same interpretation has been reported in previous studies from e.g. Alley and Smith (1981) and Borris et al. (2013). Perturbations in maximum build-up did not change the dynamics of any event at all in the first sensitivity analysis. The change in the correlation coefficient was 0% for the events individually and for all the events jointly.

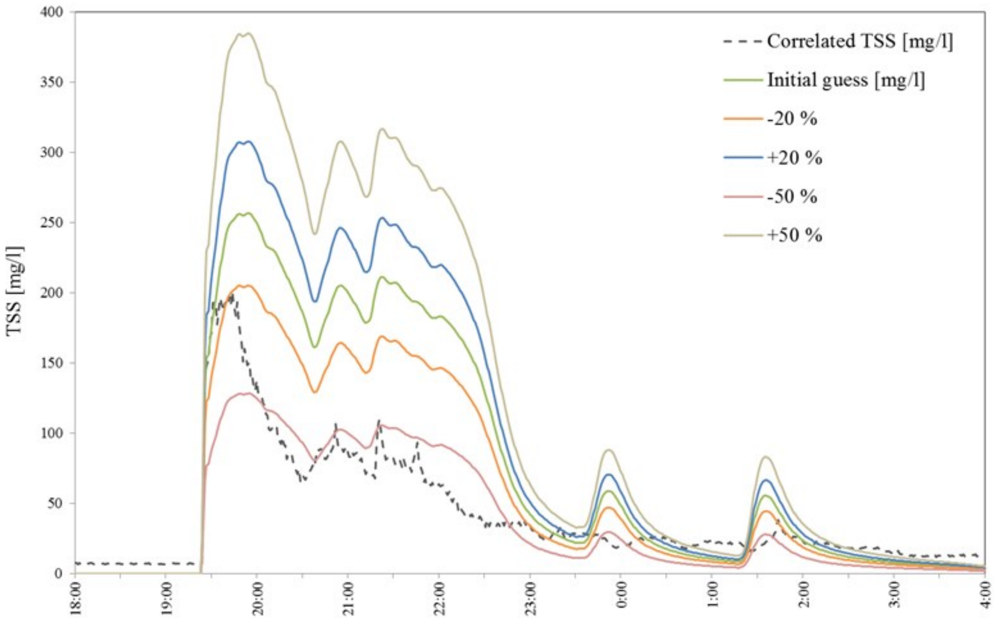


Figure 13. *The effect of perturbations in the maximum build-up for the first calibration event in the first sensitivity analysis.*

The first sensitivity analysis thus suggests that maximum build-up is purely a coefficient and affects the simulated concentrations by a common multiplier throughout the whole simulation. The second sensitivity analysis, however, gave a different response. When the initial build-up was also calibrated, the sensitivity of the model toward the maximum build-up was reduced. Figure 14 demonstrates the effect of a ±50% change in the maximum build-up for the first and the last calibration event when there was an initial build-up of 262.9 kg/ha on the catchment surfaces. It can be observed that for the first event a reduction in the maximum build-up does not have an effect on the model output (the concentration curves of O10 and -50% are on top of each other) and a 50% increase in the maximum build-up affects the simulation result only a little when compared to how it affected in the first sensitivity analysis. For the third event (and the second event that is not included in the picture) the alterations in both directions affect the simulation result. This suggests that including initial build-up in the simulation reduces the effect of maximum build-up throughout the simulation, but the most in the beginning. The importance of including initial build-up in the simulation is therefore emphasized when the calibration is conducted using only a few events.

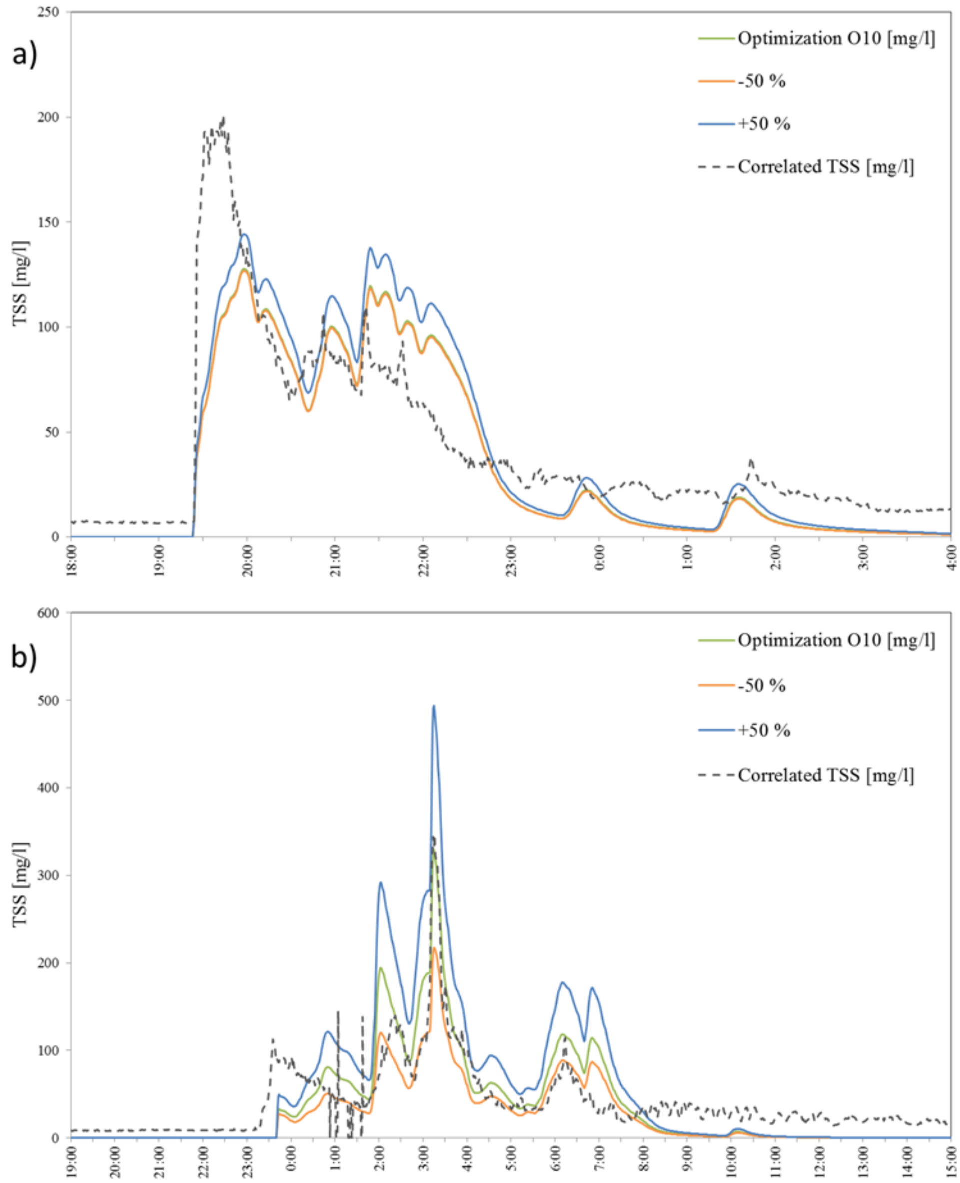


Figure 14. *The effect of a ±50% change in the maximum build-up for the first (a) and the last calibration event (b) in the sensitivity analysis after calibration.*

In the first sensitivity analysis the maximum build-up was the most sensitive parameter for all the three events. When the initial build-up was included in the calibration, the importance of the MB was significantly reduced. The effect of a -50% change in MB in the second sensitivity analysis only changed the *TLE* for all events by -19.5% in comparison to -50% when the initial build-up was not taken into account in the first sensitivity analysis. The effect of a +50% change in MB on the other hand caused almost the same change in *TLE* as in the earlier sensitivity analysis, +49%. The model was not very sensitive to the maximum build-up in the first event anymore; a -50% change only caused a -0.9% change in the TLE, and a +50% change caused a 14.2% increase in the TLE. In the second and third events a +50% change caused a 50% increase in the TLE, but a -50% did not affect as much (-7.1% and -30.5% change for the second and third event, respectively). This indicates, that the smaller the maximum build-up is, the less it affects when the initial build-up is present.

Borris et al. (2013) neglected the initial build-up and therefore draw the conclusion that the maximum build-up is a coefficient whose absolute value would not be important as it could be scaled down or up. When the initial build-up is included, the maximum build-up still affects as a coefficient so that the wash-off pattern remains the same independent of the maximum build-up, only the scale changes. The correlation coefficient does not change in the second sensitivity analysis either. Still, the effect of initial build-up is evident when the total loads are inspected; a 50% reduction in the maximum build-up reduced the total load of the third event with 30.5% compared to 50% in the first sensitivity analysis. It also seems that the larger the maximum build-up is the less important the initial build-up is, because a 50% increase in the maximum build-up in the second and third events caused a 50% increase in the total load like in the first sensitivity analysis. When the entire calibration sequence and the total load error are regarded, the maximum build-up is the second most important parameter (after wash-off exponent, discussed later in Section 4.4.6).

The maximum build-up has a physical basis. It is the maximum amount of pollutants that can build up on the surfaces of a catchment and could be determined with measurements of debris on surfaces. The amount of pollutants over the same surface type in different locations could be measured over the course of time, when the build-up dynamics could be looked into and an average maximum could be found for the specific surface type and used also as an approximate value in other similar places.

The maximum build-up alternates even within the same surface type according to the surroundings and activities of the selected area. In some time the constituents over surfaces stop accumulating as they are resuspended by wind and for example vehicles or pedestrians moving over the surface (Haiping & Yamada 1996). Any approximations for the MB would still be welcome for simulation purposes as at the moment only few studies (none to the author's knowledge in urban areas in Finland) report about measurement-based quality parameter values.

4.4.3 Initial build-up

In the first sensitivity analysis, perturbations in initial build-up had large impacts on the simulation result over the first event, but a minor impact on the last event, which can be observed in the goodness-of-fit values in Appendix 1. The impact on the first event was similar to the impact of the maximum build-up so that the correlation coefficient did not change much, but the per cent change in the total load was nearly as large as the change in the parameter.

In the second sensitivity analysis the effect of the initial build-up is similar to that of the maximum build-up in the second sensitivity analysis, but to the opposite direction (Appendix 2). The model is most sensitive towards perturbations in the initial build-up in the *beginning* of the simulation and the effect *reduces* with time. Also it seems that increasing the initial build-up affects the simulation output more than decreasing it. Decreasing the IB by 50% reduced the total load of the first event by 35.8% but had no effect on the simulations of the second and third event. Increasing the IB by 50%, on the other hand, increased the total load of the first, second and third event by 48.7%, 40.2% and 15.1%, respectively.

The main result from the sensitivity analyses for maximum and initial build-ups are that both parameters affect mostly as coefficients and do not change the pattern of the concentration curve, but change the scale. When both parameters are calibrated, the initial build-up is more important in the beginning of the simulation and the maximum build-up becomes more important when the simulation continues. Also these two parameters are interrelated so that the larger the initial build-up is, the less sensitive the model is toward the maximum build-up and vice versa; the larger the maximum build-up is, the less important the initial build-up is.

The initial build-up of pollutants has been both included and ignored in earlier studies, which was commented also by Wang et al. (2011). Temprano et al. (2006) tested the sensitivity of SWMM for the number of antecedent dry days. The calculation of antecedent dry days started from clean bed, but as the model uses the number of antecedent dry days to calculate the build-up before the beginning of the simulation, this method could also be considered as a way to take the initial build-up into account. Even still, the assumption remains that the previous event before the first simulation event has cleaned the surfaces of the catchment. Avellaneda et al. (2009) also assumed that every event would have enough energy to remove all pollutants from the surfaces. Wang et al. (2011) on the other hand write that it is practically impossible to wash off pollutants completely.

Temprano et al. (2006) found a low level of sensitivity of the model for the number of dry days, which is not very well in agreement with the findings of this study if the number of dry days and initial build-up are considered to represent the same effect. Temprano et al. (2006) mention, though, that the low sensitivity they found could have been caused by the poor sedimentation in the sewer system due to a steep slope. This exemplifies the uniqueness of every catchment and how the catchment characteristics are in close relationship with the quality parameters.

In the model development study of Wang et al. (2011) they used null residual build-up after previous event for one model and a residual build-up equal to the maximum build-up in another. The model that included the initial build-up produced better simulation results than the model that is based on the assumption of null initial pollutant. However, assuming an initial build-up to be the maximum build-up is also just an assumption and might have essentially no more truth-value than assuming null initial build-up or calibrating it. The problem with including the initial build-up in the calibration is that it represents a true physical value, which is the result of processes in the catchment before the simulation period. Yet in the calibration the accumulation and wash-off in the catchment *before* the start of the analysis have no effect in the calibrated initial build-up, as it is optimized to any value that combines well and gives a good outcome with the rest of the parameters *during* the calibration sequence. It is the calibration sequence that dictates where the initial build-up settles, not the processes before the calibration sequence, although in reality it is vice versa. Even still, because the assumption of a predefined build-up (of zero or equal to the maximum build-up) is erroneous to begin with, it is recommended to include the initial build-up in the calibration.

Because the MB and IB are bound together and are so sensitive, it would be advisable to measure also the initial build-up. The measurement strategy described for the maximum build-up would work for the initial build-up too. When measuring initial build-up, though, probably the most interesting time to measure is after a rainfall event. Or better still, before and after an event when the pollution removal potential of rainfall events of different sizes could be studied. If the initial build-up after an event would be known, even approximately, it would be quite convenient to start the SWMM simulation always after an event and start building up the pollutants from there until the maximum. The initial and maximum build-ups could still be calibrated, but the ranges in which their values can alter in the optimization could be smaller which would make the calibration more reliable.

4.4.4 Build-up exponent

The build-up exponent determines how fast the build-up maximum is reached. The build-up curve rises more steeply and reaches the level of maximum build-up faster when BE gets larger (Figure 15). The level of maximum build-up is not reached even in three months when BE is around 0.05. When BE increases to 0.5, the level of MB is already reached in about ten days and when BE is 1, MB is reached in about 5 days. When BE is given a value of 3, the maximum level is reached in two days and when BE is 8, the maximum is reached in one day.

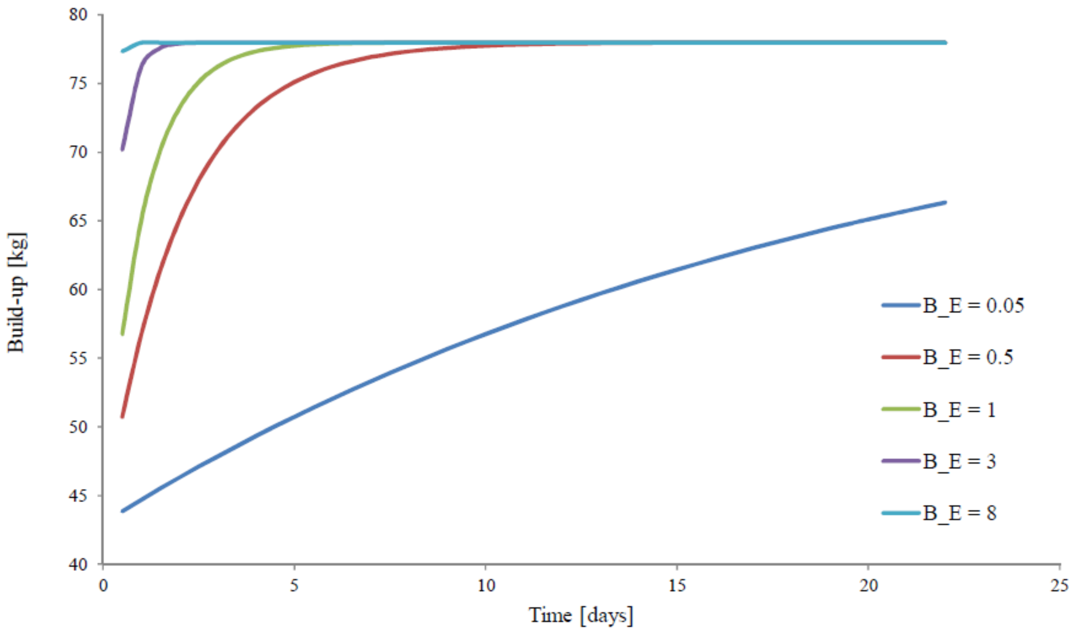


Figure 15. *The exponential build-up curve plotted against time with different build-up exponents.*

The build-up exponent is the quality parameter to which the model has the least sensitivity. In the graphs of Appendix 3 and Appendix 4 this is clearly demonstrated as the BE curve is nearly horizontal in all the graphs for both sensitivity analyses. The exceptions are the graphs for the third event, where the part of the curve on the negative side (smaller BE) has a larger slope than the part on the negative side (larger BE). This demonstrates the aforementioned matter that the smaller the BE is, the more sensitive the model is to changes in it.

The influence of the build-up exponent on the simulation output during the calibration period is observed the easiest in the second peak of the second event (Figure 16). With the conditions set by the other quality parameters, this peak is not reached when BE is small. The larger the BE is, the better the simulated total suspended solids concentration fits the observed concentration curve. This is due to a very fast accumulation of solids which is represented by the high BE. If the accumulation is really fast, the time between the two peaks is enough to accumulate so much particulate matter on the surfaces that it can be washed off by a large intensity rainfall even at the end of the event. If the build-up exponent is smaller, the build-up rate is not fast enough to accumulate particulate matter and compensate for the wash-off.

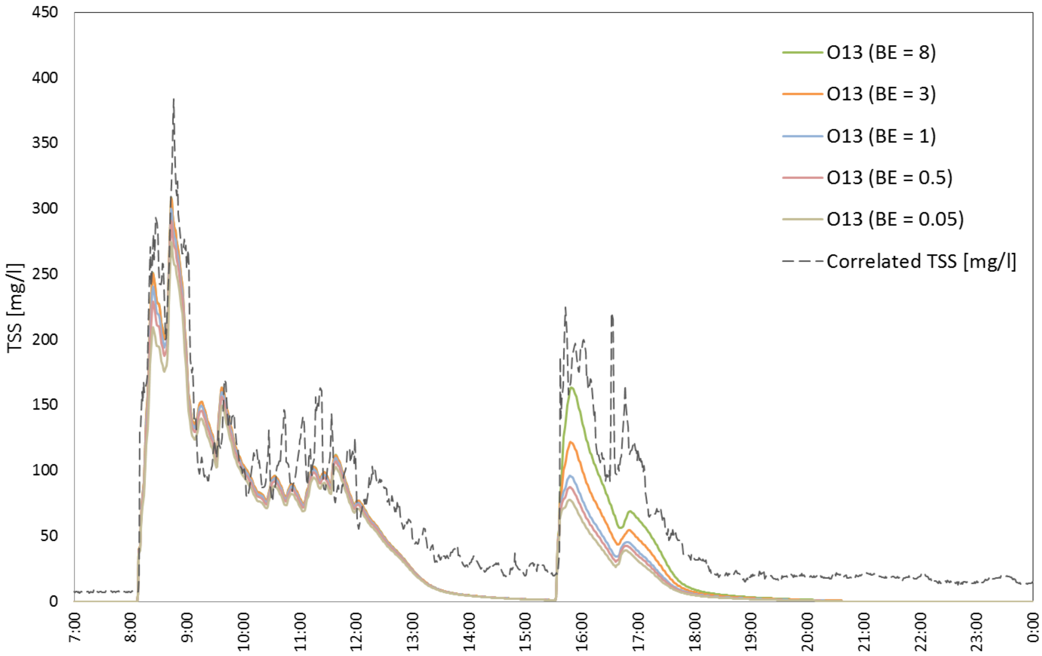


Figure 16. *The effect of BE on the modelled concentration. These simulations have been run with the optimized parameters (O13) only altering the build-up exponent value.*

The model is more sensitive to BE during the last event than it is during the first event. When the simulation starts, there is yet enough matter on the surfaces dictated in the calculations mostly by the initial build-up and the maximum build-up. When the simulation progresses the building up of pollutants has to commence so that there would be something for the upcoming events to wash off. The further the simulation continues, the more important the build-up exponent gets. However, as mentioned, out of the quality parameters the model is the least sensitive towards the build-up exponent.

The simulation results in the two sensitivity analyses for the third event are presented in Figure 17. The differences in the two sensitivity analyses are presumably due to the different scale of the parameter values and not due to the presence or absence of initial build-up. The BE value ranges between 0.45 and 1.35 in the first sensitivity analysis and between 2.5 and 7.5 in the second sensitivity analysis. It is clear that when the exponent is smaller, its perturbations cause a change in the simulation result, but when the exponent is large, the changes are negligible.

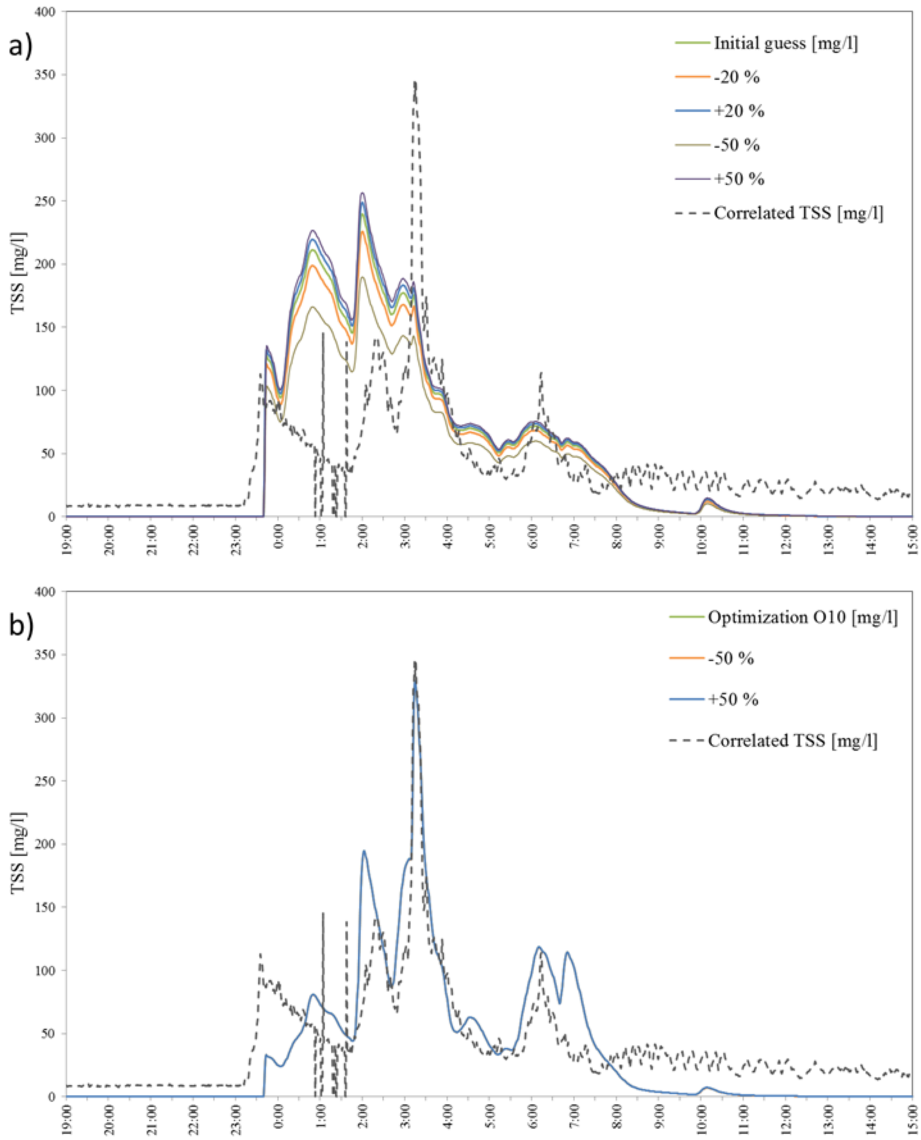


Figure 17. *The effect of perturbations in build-up exponent in the sensitivity analysis before (a) and after calibration (b) for the third event.*

4.4.5 Wash-off coefficient

The correlation coefficient changes a little when WC is perturbed in the first sensitivity analysis (Appendix 1). The maximum change is observed in the third event when WC is altered by 50% when the correlation coefficient changes from 0.63 to 0.58. For the whole three event period the change in the correlation coefficient alters between -7.4% and 8.9% when the WC value is increased by half and reduced to half of the initial value, respectively. For the second and third event when there was a decrease in the parameter value the correlation became stronger and for an increase in the parameter value, the correlation degraded, but for the first event the effect was the opposite.

In both of the sensitivity analyses there is slightly less effect on the total loads from an increase than from a decrease in the parameter value. Also, the first event is affected more than the second and third events, which are affected almost equally by the parameter deviations. The model is not as sensitive to the wash-off coefficient as it is to the maximum build-up, but otherwise the effect of WC on the simulated concentration curve is similar to that of the build-up maximum. When WC or MB grows, the concentration grows. MB increases the *amount* of pollutants on the surfaces, when there are more pollutants available for wash-off. Increasing WC increases the *efficiency* of wash-off, which means that runoff can release more pollutants. In both cases the pollutant concentration increases in the runoff.

The sensitivity analyses before and after calibration gave similar results for WC even though the values differ by two orders of magnitude so that WC values altered in the ranges of 0.02-0.06 before the calibration and 0.0005-0.0015 after the calibration. It is therefore assumed that the absolute value of WC is not much affecting the sensitivity of the model for WC, but the *proportion* of WC to the build-up parameters is. WC needs to be smaller in the second than first sensitivity analysis, because all the build-up parameters have notably higher values than in the first sensitivity analysis. The absolute value of WC changes according to the build-up parameter values, but as the proportion of the parameters remains, the sensitivity of the model towards WC remains the same.

4.4.6 Wash-off exponent

The wash-off pattern during an event is most affected by the wash-off exponent. In Equation 5 the stormwater flow is the base of the wash-off exponent. The concentration of washed-off pollutants is therefore always dependent on the flow, but depending on WE, the concentration response to the flow can vary a lot. The amount of pollutants in stormwater can increase or decrease as a function of flow.

In the first sensitivity analysis the values tested for WE were 1.15, 0.92, 1.38, 0.575 and 1.725 (Figure 18). The concentration curves obtained with values larger than one were following the flow curve so that the concentration curve went up and down with the flow curve. This is normally observed in the catchments during rainfall-runoff events. A larger amount of water flowing on the surfaces releases more pollutants from the surfaces and thus the concentration grows higher. SWMM is also able to produce the opposite kind of response, where the pollutant concentration is even reduced when more water flows on the surfaces. This can be observed in Figure 18 when WE is given a value less than 1 (-20% and -50%). When the WE is 0.92 the concentration curve does not rise a lot when the flow curve rises. For the smallest value, 0.575, the response is even almost the reverse so that when flow increases, the pollutant concentration decreases.

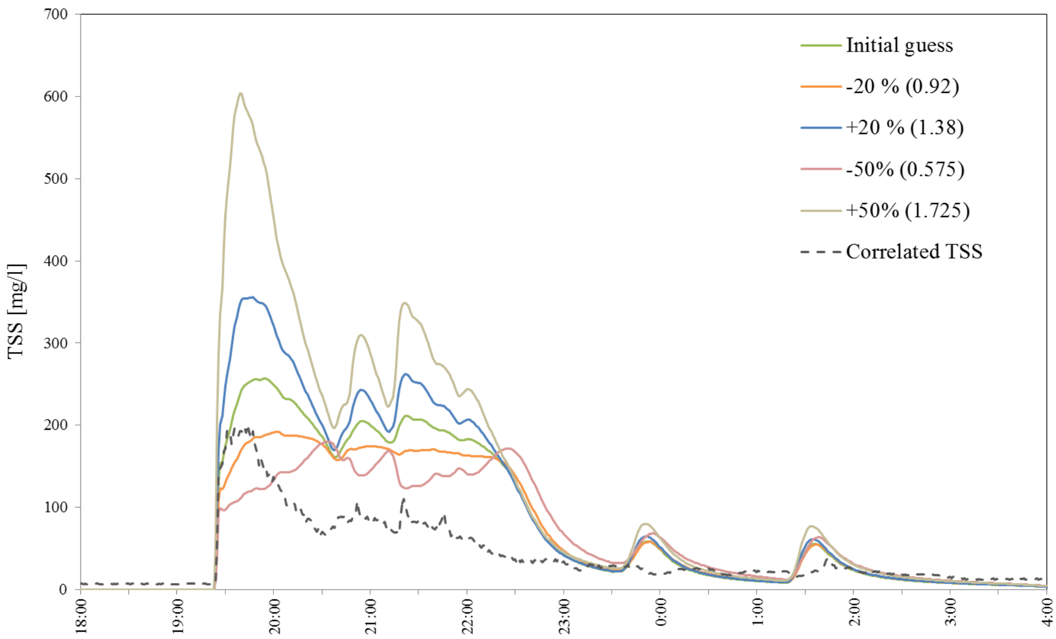


Figure 18. *The concentration response to alterations in the wash-off exponent during the first calibration event in the first sensitivity analysis.*

The wash-off exponent affects the dynamics of the TSS output the most. The correlation coefficient between the observed and simulated concentration curve changed in the first sensitivity analysis by 14% with a -20% change in the parameter value and 45% with a -50% change in the parameter value for the last event. In the first sensitivity analysis alterations in the parameter value have slightly less impact on the total load than the maximum build-up. In the sensitivity analysis after calibration, on the other hand, the total loads are affected much more. In the second sensitivity analysis a 20% decrease in the parameter value causes, for example, a 49.7% decrease in the total load of the third event and a 50% increase in the parameter value increases the total load by 228.5%.

The large impact of changes in the wash-off exponent in the second sensitivity analysis might be more due to the absolute value of WE than due to the presence of initial build-up. In the first sensitivity analysis WE ranged between 0.92 and 1.725, where 1.725 is in the range of the optimized values for WE, and in the later sensitivity analysis it ranged more widely between 0.728 and 2.184.

WE has been kept constant in many earlier studies. Egodawatta and Goonetilleke (2010) and Borris et al. (2013) assumed that WE is a constant irrespective of land use or surface type, and only changed the wash-off coefficient with surface type. Baffaut and Delleur (1990) kept both the wash-off exponent and coefficient constant for high and low intensity rainfall events. In this study, too, WE did not alter very much between the different optimizations, but as the model is sensitive to the parameter, the results were clearly improved by the calibration of WE. When all the events and sensitivity analyses are considered together with graphs of Appendix 3 and Appendix 4, the main result is that the SWMM model is the most sensitive to the wash-off exponent. The wash-off exponent is the most important parameter to define the shape of the wash-off curve and is therefore crucial in modelling the stormwater quality. WE is also a parameter whose value can only be tested by simulations so it is suggested to be included in the calibration at least within a narrow range. If the number of calibration parameters needs to be reduced, WE could be kept constant after it has been tested for a few optimizations, where it tends to settle. The more nonlinear the water quality response is to runoff, the more important it is to include the wash-off exponent in the calibration to improve the model performance.

4.4.7 Depression storage

The depression storage was altered in the first sensitivity analysis by ±50%. The initial depression storage ranged from 0.1 to 4.82 mm (the weighted average for the entire catchment being 1.86 mm (Krebs et al. 2014)) and the new values were between 0.05-2.41 for the reduced and between 0.15-7.23 for the increased depression storage (Appendix 1). In the hydrological parameterization conducted by Krebs et al. (2014) the volume error between the observed and the simulated flows was 3.1% initially. Reducing the depression storage gave a volume error of -3.7% and increasing the depression storage increased the VE to 9.3% so that the change was approximately 6% to both directions.

Perturbations in the depression storage affected the results of the first event the most (Appendix 1). The first event was also the only event where increasing the depression storage reduced the total load. For the second and third event increasing the depression storage increased the total load. For the third event the correlation coefficient did not significantly change along with perturbations. All in all decreasing the depression storage had a larger effect on the quality simulation results than increasing the depression storage when the performance for the entire calibration sequence was considered.

When depression storage is decreased (less water is retained in the subcatchments), the total load of the first event increases, but the total load of the second and third events decreases. This is probably due to larger wash-off along with larger flow in the beginning of the calibration period, when less pollutant is left to be washed off during the last two events. When the depression storage is increased (more water is retained in the subcatchments), the opposite happens; not that much is washed-off in the beginning and thus more is left to be washed off during the later events.

The wash-off curve in Figure 19 shows that the smaller the depression storage, the faster the catchment responds to rainfall, resulting in higher peaks at the beginning of an event. The less there are pits and other depressions where the water can be retained, the more water is available to flow on the surfaces and the more solid matter can be extracted from the surfaces. It can be seen from the graphs in Appendix 3 and Appendix 4 that the sensitivity of the quality model towards changes in the depression storage is still very low compared to the actual quality parameters (Sections 4.4.2-4.4.6).

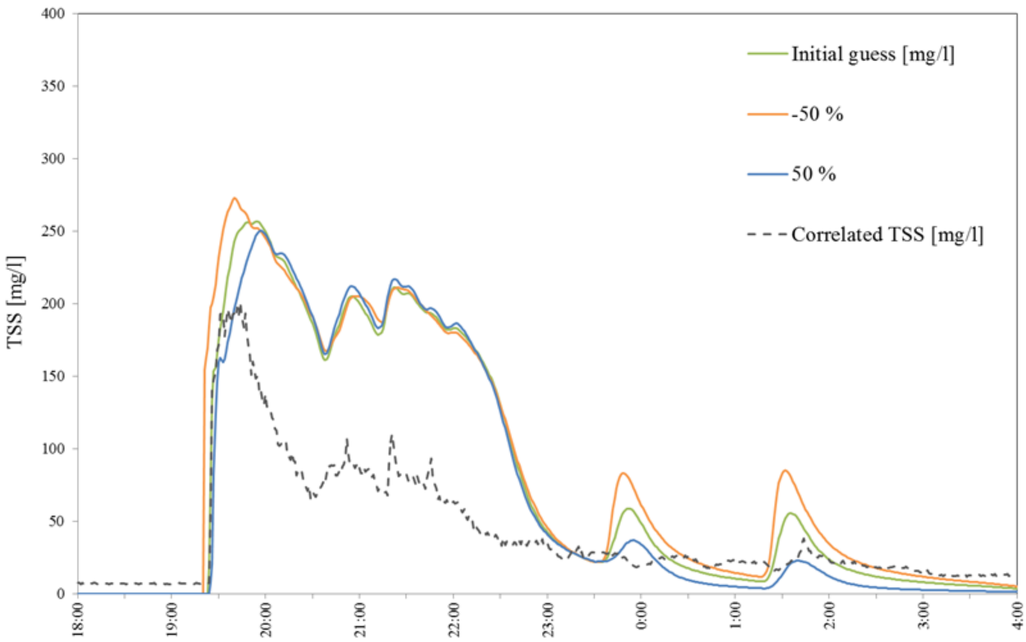


Figure 19. *Effects of perturbations in the depression storage during the first event in the sensitivity analysis before calibration.*

4.4.8 Manning's roughness coefficients

The effect on water quality of perturbations in Manning's roughness coefficient for conduits and overland flow was examined (Appendix 1). The largest effects could be observed during the first event, but all in all the effects were small; changes in both Manning's *n* for conduits and Manning's *n* for overland flow changed the *TLE* by ±1% at the most and the correlation coefficient changed at the most by only 2.2%. The model was a bit more sensitive towards changes in the Manning's *n* for conduit flow as its value was altered ±30% and the value of Manning's *n* for overland flow was altered ±40% and yet the changes in the goodness-of-fit values were of the same magnitude. In this catchment the conduit flow affects the water quality more than overland flow.

Krebs et al. (2014) reported that for this specific study catchment the hydrological model was more sensitive for the Manning's *n* for conduit flow. Altering the Manning's *n* values did not affect the flow volume much. In the hydrological model the initial volume error for the flow was 3.1% and when the Manning's *n* values were reduced, the VE became 2.8% and 3.2% for the overland and the conduit flow, respectively. When the Manning's *n* values were increased, the VE became 3.3% and 3.1% for the overland and conduit flow, respectively. The peak flows on the other hand were affected more by the perturbations. For both of the Manning's *n* values the initial peak flow error (*PFE*) was -35%. The decrease in the Manning's *n* for overland flow changed the *PFE* to -39.3% and the increase changed it to -29.1%. The *PFE* for the decreased Manning's *n* for conduit flow was -38.3% and for the increased value it was -16.3% (Krebs et al. 2014). The model being more sensitive for the Manning's *n* for conduit flow therefore shows in the peak flows. The effects of perturbations in the hydrological parameters depend on the catchment characteristics and the hydrology affects the quality through affecting flow.

The more a surface delays the movement of flowing water, the larger the value of Manning's *n* is for the surface. When the movement of water is delayed it shows as lower peak flows and accelerating the movement of water shows as higher peaks. The pollutant concentration follows the flow pattern and thus when the Manning's *n* is increased the simulated TSS concentration has a tendency to increase with flow. On the other hand when the Manning's *n* is reduced, the flow and the concentration of TSS are observed earlier and the curve rises and declines more steeply. This is because water flows faster and can therefore remove more particulate matter from the surfaces. Also, the water reaches the outlet of the catchment (where the quality and flow are measured) faster. Even though these alterations can be observed, it is clear from the graphs in Appendix 3 and Appendix 4 that the stormwater quality submodel is the least sensitive towards changes in the Manning's *n* values.

4.4.9 Summary of the sensitivity analyses

The most sensitive quality parameter is the wash-off exponent followed by the maximum build-up, the initial build-up, and the wash-off coefficient. The build-up exponent is the least sensitive quality parameter.

The hydrological parameters affect the flow, and through that have some effect on the wash-off, but they do not affect the build-up characteristics. Although the effect of perturbations in the hydrological parameters could have been studied also in the second sensitivity analysis, studying the hydrological parameters more profoundly was considered unnecessary as the hydrological model was already calibrated. All in all the hydrological parameters only affect the water quality to a small degree in this kind of simulation and can therefore be left out of inspection. The primary prerequisite for a sound water quality calibration is a working hydrological model and only after the hydrological model is validated and working, the water quality model can be calibrated. The hydrological parameters should not be mixed in the quality calibration, but it is better to calibrate the water quality model with the wash-off and build-up parameters alone.

There are various interrelations between the five quality parameters. The most evident example is the interrelation between the maximum and the initial build-up. The initial build-up affects most in the beginning of the simulation and the maximum build-up starts affecting more after the influence of IB starts fading. The initial build-up is herein recommended to be included in the study when stormwater quality is simulated. A careful sensitivity analysis helps to determine suitable ranges for the calibration parameters and is the more important the more parameters are included in the calibration. Parameter ranges from similar catchments and climates can be used as help.

4.5 Optimization results

4.5.1 Optimization setup and criteria

The quality parameters were calibrated using the multi-objective optimization algorithm NSGA-II (Deb et al. 2002). The optimizations were stopped when the parameter values did not change substantially within the best 30 optimization results when the optimizations were ranked according to the smallest *SSE* and according to the largest linear correlation coefficient. The best optimization runs with respect to *SSE* and linear correlation did not need to yield identical parameter sets. The best optimization result was considered to be the one with the smallest *SSE* even though the linear correlation would not be the best. Several previous studies have also found the *SSE* to produce the best fitting parameters (e.g. Avellaneda et al. 2009, Haiping & Yamada 1996). The performance statistics are very similar for the concentration curve that gives the smallest *SSE* and the one that gives the best linear correlation. An example of this difference is presented in Figure 20 where the best results according to *SSE* and *CORR* for optimization O14 are illustrated. The difference in the Nash-Sutcliffe efficiency is in the third decimal, the difference in *CORR* is 0.008 and the difference in *TLE* is 0.9%.

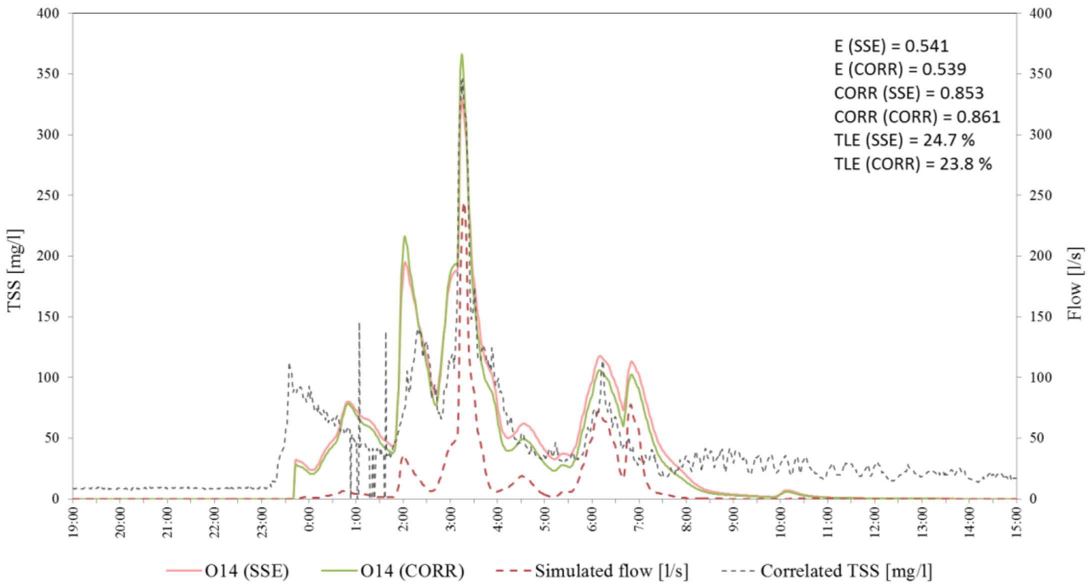


Figure 20. *The best optimization result as per SSE and correlation coefficient (CORR) for optimization O14 (third calibration event). The difference can be noticed but is not very big.*

Two optimizations with identical setup were run simultaneously to check if the genetic algorithm produces similar results. Some optimizations (O8-O12) were run only once, because of time limitations and because according to the first several optimizations it seemed that the algorithm ends up with essentially the same parameter values for similar runs.

Reaching the optimum parameter values with the NSGA-II requires thousands of simulation runs and is time-consuming. To reduce the computation time the optimizations were run with a Hot Start File starting directly from the beginning of the first calibration event and the initial build-up was included as a calibration parameter.

4.5.2 Parameter boundaries

The parameter range sets introduced in Section 3.8 were used in the calibration procedure. The allowable ranges in which the parameters were let to fluctuate and settle in the optimizations were defined by RS1 (Range Set 1) for optimizations OR1 and O1-O7, by RS2 (Range Set 2) for optimizations O8-O10 and by RS3 (Range Set 3) for optimizations O11-O14 (Table 3). The entire calibration sequence was simulated using the optimized parameter set of every optimization (O1-O15 and OR1) and the resulting simulated concentration of TSS was compared with the observed (correlated) TSS concentration. Table 4 presents the optimized parameter values for every optimization. The optimization OR1 and its results are discussed in Section 4.5.6.

Table 3. *The parameter ranges (RS), initial values in the optimizations and the range of calibrated values for the initial build-up (BI), maximum build-up (MB), build-up exponent (BE), wash-off coefficient (WC) and wash-off exponent (WE).*

Parameter	Minimum			Maximum			Initial	Calibrated range (all)	Calibrated range (best)
Range set	RS1	RS2	RS3	RS1	RS2	RS3			
IB	20	0	0	50	1000	450	35	3 - 265	44.9 - 265
MB	20	1	30	250	1000	450	35	20.3 - 286	161.6 - 286
BE	0.08	0.01	0.001	1.11	5	8	0.09	0.049 - 8	0.08 - 8
WC	0.013	0.001	0.001	1.4	30	1.5	0.25	0.001 - 0.36	0.001 - 0.013
WE	0.16	0.01	0.9	1.8	10	2.2	1.15	1.11 - 1.8	1.349 - 1.470

Table 4. *The parameter values optimized for each combination of events and each parameter range set. The background colour indicates the range set so that OR1 and O1-O7 used RS1, O8-O10 used RS2 and O11-O14 used RS3. The parameter values for roofs are indicated as RMB, RBE, RWC, RWE and RIB.*

Opti-mization	Optimized for event(s)	Calibration parameters (two parallel optimization results)									
		MB		**BE**		**WC**		**WE**		**IB**	
OR1	1, 2, 3 (roofs separate)	43.8	36.4	0.08	0.08	0.013	0.014	0.906	0.718	42.6	35.8
		(RMB)		**(RBE)**		**(RWC)**		**(RWE)**		**(RIB)**	
		(114.1)	(119.0)	(1.11)	(1.11)	(0.013)	(0.013)	(1.683)	(1.695)	(40.9)	(47.0)
O1	1	20.3	23.2	0.17	0.08	0.015	0.015	1.8	1.8	20	20
O2	2	109.5	110.3	1.11	1.11	0.013	0.013	1.111	1.107	20	20
O3	3	123.2	122.4	0.08	0.08	0.013	0.013	1.520	1.518	50	50
O4	1, 2	96.2	99.5	1.11	1.11	0.015	0.014	1.115	1.125	34.5	36.5
O5	2, 3	140.8	135.5	0.08	0.08	0.014	0.015	1.369	1.362	50	50
O6	1, 3	132.0	140.3	0.08	0.08	0.013	0.013	1.523	1.524	30.4	29.9
O7	1, 2, 3	164.2	161.6	0.08	0.08	0.013	0.013	1.349	1.355	45.6	44.9
O8	1	23.2	-	0.69	-	0.240	-	1.438	-	3.6	-
O9	1, 2	148.2	-	5	-	0.004	-	1.328	-	79.4	-
O10	1, 2, 3	270.4	-	5	-	0.001	-	1.456	-	262.9	-
O11	1	263	-	0.049	-	0.360	-	1.287	-	3	-
O12	1, 2	157	-	7.546	-	0.004	-	1.291	-	65	-
O13	1, 2, 3	286	262	7.992	7.992	0.001	0.001	1.422	1.470	265	253
O14	1, 2, 3 (1.5.2009)	266	267	8	8	0.001	0.001	1.462	1.458	-	-
O15	1, 2, 3 (7.7.2009)	180	181	8	8	0.002	0.002	1.418	1.414	-	-

The parameter boundaries of RS1 (Table 3) were defined according to parameter values calibrated in earlier studies and kept relatively narrow to support a faster convergence. Cambez et al. (2008) calibrated a build-up rate constant (build-up exponent) of 0.08 1/day for an 80 ha urban area in Portugal. This was the smallest BE found from the literature and was therefore set as the minimum for the first range set in this study. The upper boundary for the build-up rate constant in the exponential build-up function was set to 1.11 which was the maximum of the range Cho and Seo (2007) determined for the build-up exponent in a Korean watershed of varied land uses. Very different values for the build-up maximum were encountered in the literature ranging from 17.5 kg/ha (Temprano et al. 2006) to 560 kg/ha (Alley & Smith 1981) the most being less than 300 kg/ha. The upper boundary for MB was first set at 250 kg/ha. Earlier calibration results of the initial build-up were not found from the studied literature. In most of the studies the issue was not addressed at all, which was considered to indicate that the initial amount of build-up on the surfaces was set to zero. Without prior knowledge the first range for IB was set to 20-50 kg.

The smallest wash-off coefficient, 0.013, was used by Borris et al. (2013) to express a low wash-off rate of pollutants (Table 3). The largest possible wash-off coefficient in RS1 was set at 1.4 although a value as large as 46 was used by Temprano et al. (2006). The majority of wash-off coefficients used in earlier studies ranged between 0.013-1.2. The wash-off exponent in previous studies ranged between 0.16 and 9.83 (Cho & Seo 2007). The minimum boundary was taken from Cho and Seo (2007), but the maximum boundary was set to a lower value, 1.8, because in the other examined studies the value for WE stayed at a little over 1.

It can be noted from Table 4 that RS1 was for many parts too narrow. BE reached either the minimum or the maximum boundary in every optimization except for O1 (the first simulation) and the wash-off coefficient got stuck in the minimum boundary in most optimizations. The initial build-up went to the minimum in RS1 for the first two optimizations to reduce the scale of the concentration curve and to the maximum for the optimizations O3 and O5 for the opposite reason; to increase the scale of the concentration curve. The wash-off exponent went to the maximum in the optimization O1 most likely because both IB and MB were so small that the wash-off parameters needed to be large to increase the concentrations in the output. Only the maximum build-up kept well inside the boundaries of RS1.

After the optimizations O1-O7 (Table 4) were finalized, RS2 was defined (Table 3). In RS2 the ranges for all the parameters were meant to be sufficiently wide for parameters to fluctuate freely without having optimized values at the boundaries. Still the wash-off coefficient went to the minimum in optimization O10 and the build-up exponent went to the maximum in O9 and O10. Thereafter, the third and the last range set, RS3, was created (Table 3). The range for BE was extended more to see whether it would increase beyond RS2 or whether it could settle to a smaller value than the minimum in the RS1. The minimum boundary for WC was not changed as the precision of 0.001 was thought to be already quite small and more precise values had not been met in earlier studies. The ranges for MB and WE were constricted as they had not fluctuated extensively in the previous optimizations to restrict the number of parameter combinations that produce the same result.

When all optimizations are examined (Table 4), it is noted that BE tends to go to the maximum in the optimizations where the third event is present and also in the optimizations jointly for the first and second event, except in RS1, where BE goes to the minimum for all the optimizations where the third event is included. In the second event there is a relatively large peak after the first one, which can only be reached if the build-up in the beginning of the second event is so large that all the pollutants cannot be washed off by the first peak or if the build-up of pollutants is very fast. The same applies for the third event; the later occurring peaks cannot be reached if the build-up is not large. Increasing BE allows for a faster build-up of pollutants, when there is more pollutants on the surfaces and accordingly the peaks can be better reached. BE goes to the minimum in RS1, because the minimum boundary for WC is too large. WC increases the wash-off too much even when it is in its minimum and the only way to reduce it in this case is to reduce the *amount* of pollutants, which is done by decreasing BE. Also IB and MB have an effect on reducing the amount of pollutants.

In the optimizations for the first event BE is not so high and on the other hand WC tends to be higher than in the other optimizations (Table 4). The build-up does not need to be large or fast for the first event, which is small and a sufficient amount of pollutants is provided by the initial build-up. Some of the accumulated build-up still needs to be washed off, so rather the wash-off coefficient is larger. For the second and third event it is the other way round. The wash-off coefficient tends to be smaller so that the pollutants would not wash-off too fast and on the other hand BE is higher to guarantee a fast enough pollutant build-up so that there is something to wash-off even at the end of the calibration sequence. BE and WC are in this way closely related.

RS3 produced the best calibration results. The optimization for the first event separately produced a poor *E* for the entire calibration sequence, but the rest of the optimizations with RS3 produced an *E* ranging from 0.61 to 0.72, a *CORR* ranging from 0.86 to 0.89 and a *TLE* ranging from -2.5% to -7.3% for the entire calibration sequence. The good results obtained with RS3 are explained by the build-up characteristics being able to come closer to the observed behaviour of the catchment with the more loose parameter boundaries for BE. However, even if the results get better with the larger BE, increasing the parameter further might not represent reality well as the pollutants might not in reality build up that fast. Earlier studies have suggested that the maximum build-up is reached for example in three weeks (Egodawatta & Goonetilleke 2010), although in the study of Borris et al. (2013) 80% of the maximum build-up was reached with a fast rate in less than two days, but the maximum build-up they used was fairly small, 35 kg/ha.

Table 5 presents the goodness-of-fit values for the simulations whose parameter values were calibrated using all the three calibration events together. The performance statistics of all the simulations are presented Appendix 5. Figure 21 illustrates a comparison of the TSS concentration curves of the best optimization results from every range set (O7, O10 and O13). It is noted that O10 and O13 produce a very similar curve and they reach higher in the latter peaks of the second and third events, but stay lower during the entire first event. The initial and maximum build-up values of O7 are much lower than those of O10 and O13, which seems a bit surprising as it is exactly the first event where the optimization O7 still produces a higher concentration. In the sensitivity analyses it was concluded that IB has most effect at the beginning of the simulation when it would be logical that a smaller IB would leave the concentration curve lower than in the simulations with high IB. The reason for the higher concentration curve of O7 does not lie in the build-up characteristics, but in the wash-off. The wash-off coefficient is 13 times higher in O7 than in O10 and O13, which results in a much faster wash-off. Fast wash-off can also be observed in the third calibration event, where O7 curve rises the highest in the beginning, but does not rise as high as O10 and O13 in the end of the event. The effect of a high BE could also be the reason to the high peaks of O10 and O13 compared to those of O7.

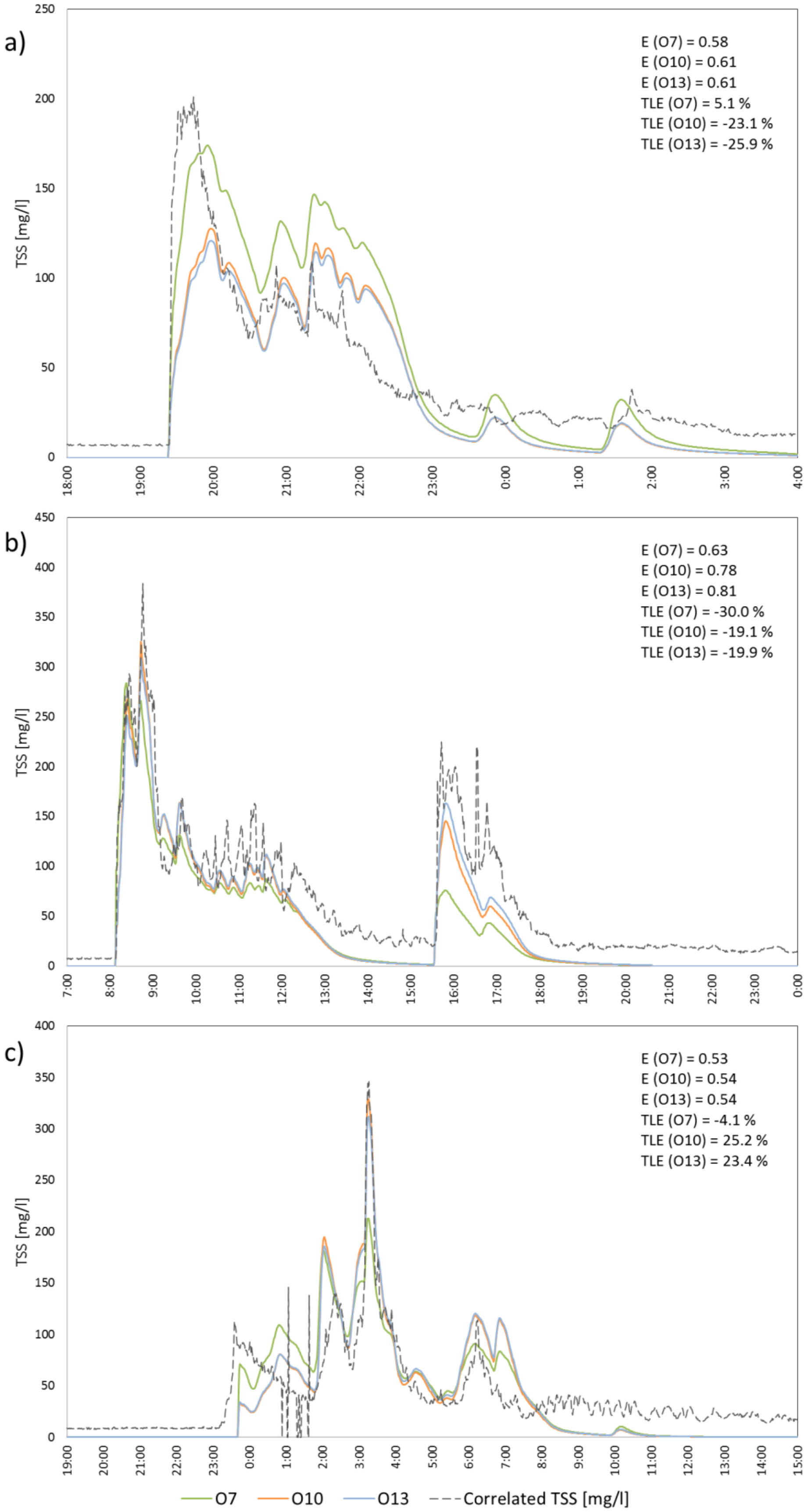


Figure 21. *The results of calibration for the three events jointly with RS1 (O7), RS2 (O10) and RS3 (O13) with their corresponding Nash-Sutcliffe efficiencies and total load errors for CAL1 (a), CAL2 (b) and CAL3 (c).*

Table 5. *The goodness-of-fit values for the simulations whose parameter values were calibrated using all the three calibration events together. These optimizations gave the best calibration results. The model continuity errors are presented for four optimizations.*

Optimization	All events			1st event			2nd event			3rd event			Runoff quality continuity error	Quality routing continuity error
	E	CORR	TLE	E	CORR	TLE	E	CORR	TLE	E	CORR	TLE		
OR1	0.71	0.87	-9.2%	0.67	0.85	-13.2%	0.74	0.92	-23.1%	0.65	0.87	12.0%	-9.2%	6.7%
O7	0.61	0.83	-18.8%	0.58	0.88	5.1%	0.63	0.90	-30.0%	0.53	0.80	-4.1%	-9.4%	2.2%
O10	0.70	0.88	-2.2%	0.62	0.82	-23.1%	0.78	0.95	-19.1%	0.54	0.85	25.2%	-2.3%	1.0%
O13	0.72	0.88	-3.4%	0.61	0.82	-25.9%	0.81	0.95	-19.9%	0.54	0.85	23.4%	-2.2%	0.8%
O14	0.72	0.89	-2.5%	0.62	0.82	-21.6%	0.81	0.95	-19.4%	0.54	0.84	24.7%	-	-
O15	0.66	0.86	-4.4%	0.19	0.78	-62.9%	0.84	0.96	-19.1%	0.48	0.84	23.0%	-	-

4.5.3 Model continuity errors

Model continuity errors represent the mass balance errors in the computation according to Equation 12.

$$E_m = S_{init} + L_{in} - (S_{final} + L_{out}) \tag{12}$$

where E_m is the model continuity error, S_{init} and S_{final} are the initial and final storage, respectively, and L_{in} and L_{out} are the total inflow and outflow, for the drainage system, respectively. Rossman (2010) suggests that if the continuity errors exceed 10%, the validity of the results should be questioned. The quality continuity errors of the model in this study were over 10% for some of the optimizations and were elevated, over 9%, for the optimizations OR1 and O7 (Table 5). The smallest quality continuity errors in the simulations over the whole calibration period were produced with O10 and O13. The continuity errors are in the same order as the total load errors for the entire calibration sequence. It was considered that the errors in OR1, O7, O10 and O13 were acceptable as they did not exceed the limit of 10% and these optimizations were selected later for validation.

In Rossman (2010) the most common reason for high continuity errors is too long computational time steps. Shorter time steps were tested in the study, but no change was observed in the continuity errors. Using HSF was not the reason for the elevated error levels, as simulations without it produced errors of the same magnitude. The effect of alterations in the parameters was tested by changing the parameter values independently. When the build-up parameters were altered, the errors remained at the initial level, which indicated that the quality continuity errors are not produced by the build-up calculations. The wash-off coefficient affected the continuity errors the most. The higher the wash-off rate from the catchment was the higher the continuity errors were. WC is 13 times higher in OR1 and O7 than in O10 and O13, which could be the reason for the difference in the continuity errors for these optimizations.

The infiltration loss of pollutants was found to be high when WC was high. Slower wash-off, indicated by a smaller WC, diminishes the infiltration loss. Fast flowing water washes off pollutants more effectively than slowly flowing water and thus has a higher pollution concentration in the beginning of an event. The high concentration water fills the depression storages and infiltrates first carrying the washed-off pollutants with it. The amount of infiltrating water remains the same when the wash-off is slower, but as the pollutant concentration is smaller, the amount of pollutants infiltrating is also lower.

The high spatial resolution of the study catchment includes several small subcatchments over which only a small amount of pollutants is produced. Rounding of pollutant concentrations can therefore exaggerate the amount of pollutants and be the source of continuity errors. The overall quality continuity errors are negative (Table 5), which means that the amount of pollutants coming out of the catchment is higher than what there is available. The complex routing, where stormwater can flow through different impervious and pervious surfaces before entering the network might also produce errors. Differences in the modelling time step and the reporting time step might produce differences in the calculated values and the output values.

Even though the continuity errors were large, minimizing them in SWMM should not be solely used as the calibration criteria for the model. This approach does not suit these build-up and wash-off functions, because even senseless quality parameter combinations can produce a good outcome with no continuity errors. For example a parameter set where the maximum build-up is smaller than the initial build-up can give a good outcome for the simulation with no quality continuity errors, but still the basis is illogical. The problem of illogical parameter combinations is present with any goodness-of-fit criteria unless the parameters have been given a sensible range or sensible correlations with each other.

4.5.4 Events

Acceptable calibration results (with a Nash-Sutcliffe efficiency E ranging from 0.52 to 0.72) for the entire calibration sequence were obtained with most of the optimizations. The performance statistics of every optimized parameter set can be observed in Appendix 5. In this study the performance of the optimizations is evaluated considering all of the calibration events jointly *and* separately, when the best performance statistics were achieved with parameter sets that were optimized for the whole calibration period of three events (Table 5).

The best performance for every individual event is obtained with parameters that were optimized for that specific event. This can be expected, but what needs to be remarked here is that the parameters yielding very good results for one event, yielded poor results for the other two events. Also, parameter optimization for events 1 and 2 produced good results for events 1 and 2, but very bad results for the third event. Accordingly, optimizations against any combination of two events produced good results for the calibration events but poor results for the third event. The fewer events are included in the calibration, the better the result is for the calibration event(s), but at the same time the model performance degrades for the other events.

Temprano et al. (2006) obtained good results in their study by using only one event for calibration of the SWMM quality model and two events for validation. The small number of used events does not, however, make the results very trustful. In this study optimizations O1, O2, O4, O8 and O11, which were calibrated for only the first event or the first and the second event, produced poor results with an efficiency E ranging from -15.02 (O11) to 0.11 (O1) for the entire calibration sequence. Optimization O11 was the worst example producing the lowest goodness-of-fit values for the other two events (Appendix 5).

The other reason behind the poor performance for the entire calibration sequence by calibrating only the first event is that the first event represents the other too events poorly as it is much smaller and has different dynamics. Krebs et al. (2014) and Warwick and Tadepalli (1991) noted that it is generally more difficult to calibrate small events. Calibration for only the third calibration event produces better results, but in general the results are better the more events are included. The less there is variation in the dynamics and scale of the calibration events, the easier it is to find parameter combinations that can reproduce a similar output. In reality the rainfall-runoff events are variable and therefore any simple event or combination of events will not be able to represent reality. There is reason to think that more than two events are needed for a sound calibration and the characteristics of the selected events play an important role.

The quality model performance statistics for CAL3 were lower than for CAL1 and CAL2, when the individual event calibrations with the range set RS1 were compared (Appendix 5). The efficiency *E* for event 3 (O3) was 0.60 with a linear correlation of 0.84 and a *TLE* of -11%. The same fitness values were approximately 0.87 and 0.83, 0.95 and 0.91 and -11.6% and -14.3% for events 1 (O1) and 2 (O2), respectively. The third event has the lowest hydrological performance (see Section 4.3) and the performance of flow simulation is directly reflected to the water quality simulation. The third event also has the most difference between the observed flow dynamics and the observed fluctuations in TSS concentration. These are the likely reasons for a worse fit between the simulated and observed concentrations as the concentration is related to the flow pattern. Also there are some peaks in the beginning of the third event that are most likely erroneous measurements and cannot therefore be reproduced by the model.

The results of calibrations in RS1 for one event (Figure 22), for two events (Figure 23) and their comparison to optimization O7 illustrate how optimizing only for the small event leaves the concentrations of the larger events too low and optimizing for large events overestimates the concentrations of the small event. Including another event in the calibration equalizes the results of different optimizations, but still the optimization O5 for CAL2 and CAL3 for example overestimates the concentrations of CAL1. Interestingly the optimizations O2 (optimized for CAL2) and O4 (for CAL1 and CAL2) produce a very similar result for the second and third event although the result for the first event is better in O4. WC is a little larger in O4 which allows for a larger wash-off also during the first event, but MB is smaller so that the build-up does not grow very high for the third event.

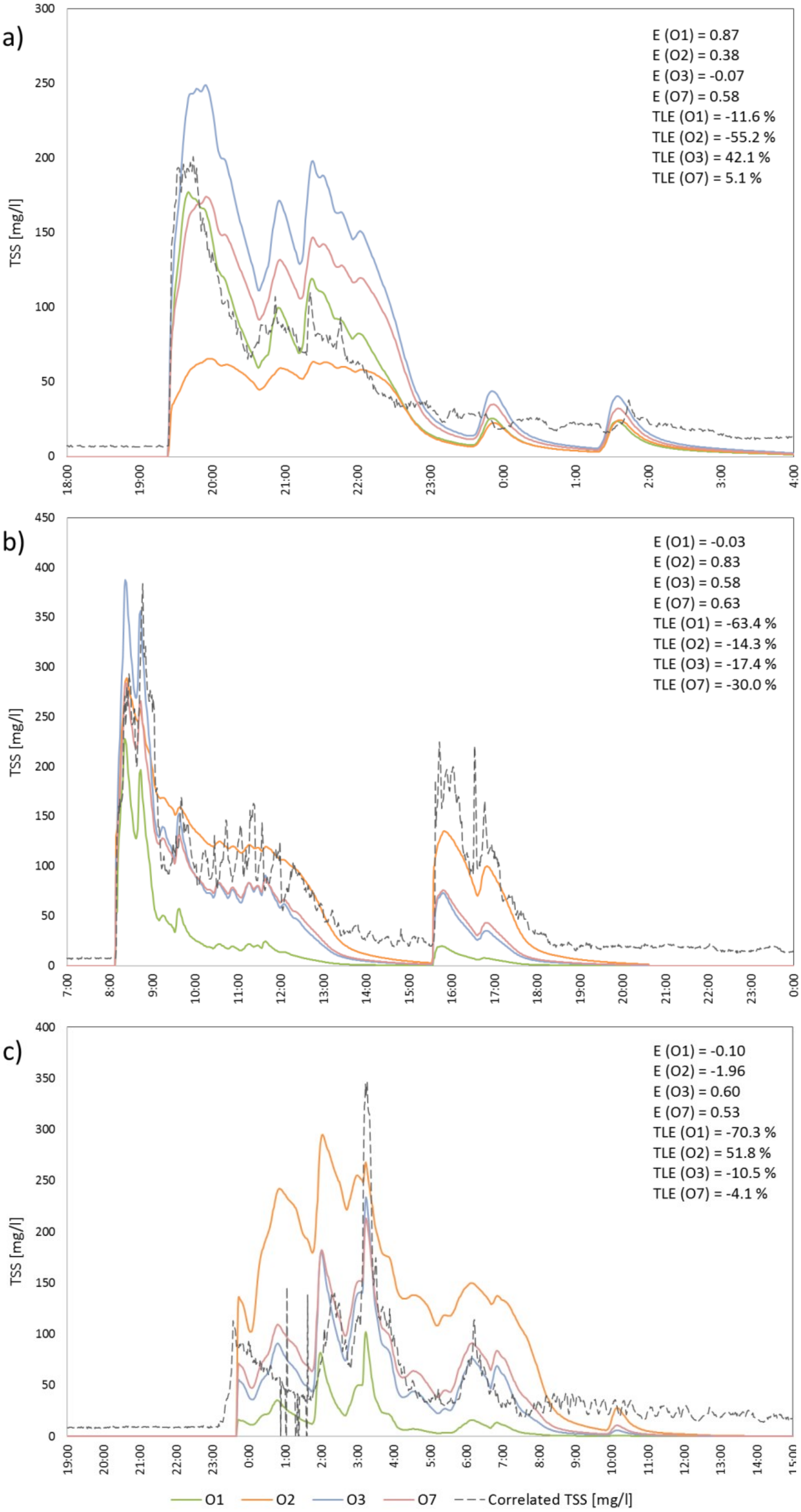


Figure 22. *The calibration results for CAL1 (a), CAL2 (b) and CAL3 (c) when parameters were optimized for the first event only (O1), the second event only (O2), the third event only (O3) and for all three events jointly (O7).*

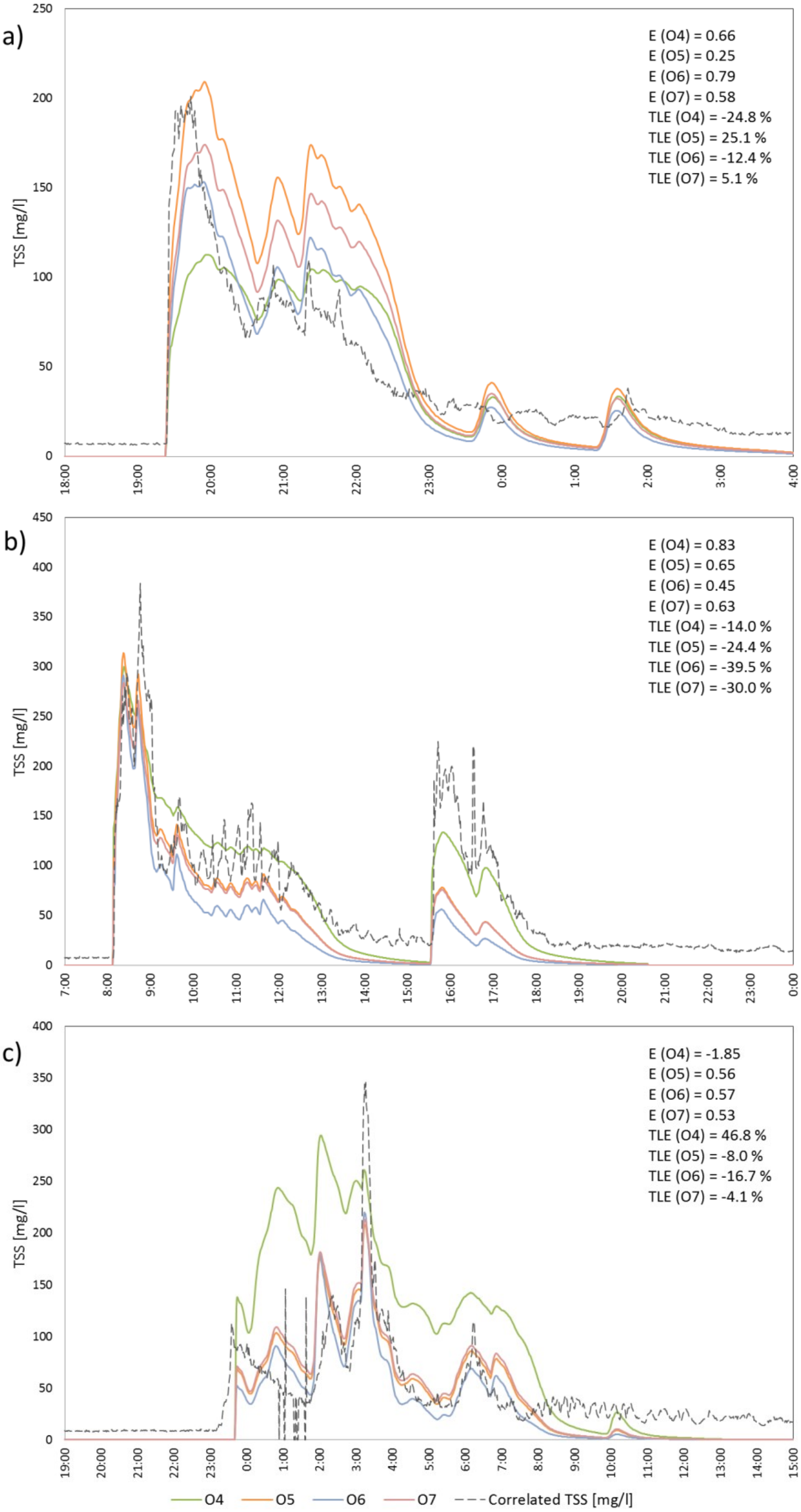


Figure 23. *The calibration results for CAL1 (a), CAL2 (b) and CAL3 (c) when parameters were optimized for the first and second event (O4), the second and third event (O5), the first and third event (O6) and for all three events jointly (O7).*

4.5.5 Parameters

The optimized maximum build-up ranged between 20.3 and 286 kg/ha in the 27 optimizations (Table 4). The smallest values of MB were for the optimizations that only concerned the first event and the largest values were for the optimizations of the entire calibration period. The maximum build-up needs to be high for the longer simulations to provide load residues to wash off even at the end of the simulation period. On the other hand, a large value of maximum build-up can make the wash-off curve to rise too high for the small event in the beginning. A large MB is restrained by a small initial build-up or by a small WC.

The wash-off coefficient had a tendency to go to the minimum in almost all of the optimizations (Table 4). The optimizations O1, O8 and O11 that were optimized for the first event have a larger WC than the rest, because IB is very small. Also the optimizations for events 1 and 2 have a WC value that is smaller than the value for only the first event, but larger than the value for all the three events that is always the minimum. RS1 leads to WC from 0.015 for the first event to 0.015 and 0.014 for the second event and finally to 0.013 for the entire calibration period. In RS2 these respective values are 0.24, 0.004 and 0.001 and in RS3 they are 0.36, 0.004 and 0.001. Baffaut and Delleur (1990) reported that the typical range for the wash-off coefficient is 1-10 but can even be higher than 10 or the values obtained in this study.

The initial build-up tends to grow with the number of calibration events (Table 4). It is in relation to WC so that as WC diminishes, IB gets bigger. IB goes from 20 kg/ha (the minimum) for the first event to approximately 35 kg/ha for the first and second event and to approximately 45 kg/ha (maximum is 50) for the three calibration events together in RS1. The corresponding values in RS2 are 3.6 kg/ha, 65 kg/ha and 262.9 kg/ha and in RS3 3 kg/ha, 65 kg/ha and approximately 260 kg/ha. IB is small for the small event to keep the scale of the wash-off curve low. WC can therefore be a little higher for the first event, but as the build-up is higher for the longer simulations (so that the pollutants would not "run out") and WC is smaller to slow down the wash-off, IB takes charge of the wash-off of the first event. Baffaut and Delleur (1990) kept the wash-off coefficient a constant, but that is not recommended in this thesis as according to the results of this study WC is important in regulating the wash-off characteristics.

To test the optimization results regarding initial build-up the model was run from 27 June until the beginning of the first event with the other four calibrated quality parameters (MB, BE, WC and WE) and without the Hot Start File. If the remaining build-up on the surfaces at the end of that run was more or less the same as was calibrated as initial build-up, the calibration of initial build-up could be considered validated. The build-up at the beginning of the first calibration event could not be previously simulated, as the other four parameters should have been given to the model for any build-up to accumulate in the simulation. The results of this test were that the remaining build-up in the beginning of the calibration sequence was different than the calibrated initial build-up. The values of the initial build-up cannot therefore be trusted to represent reality too well even in the best calibrations. This emphasises the difficulty of modelling such a complex system as the stormwater quality. Whether the calibration produces a realistic value can never be undoubtedly determined without measurements of pollutants on the surfaces at each time. Therefore calibrating initial build-up can be considered justified.

The build-up exponent ranged between the minimum and maximum values of 0.08 and 8 (Table 4). In the optimizations for all three events BE went to the minimum when RS1 was used, but towards the maximum in RS2 and RS3. The effect of BE on the simulation output was already discussed in the chapter about parameter boundaries.

According to Baffaut and Delleur (1990) the typical range for the wash-off exponent is between one and five. The values for WE obtained in this study ranged between 0.718 and 1.8 which is mostly inside the mentioned range. WE for the three-event optimizations were 1.349 and 1.355

(O7); 1.456 (O10); 1.422 and 1.470 (O13); and 1.462 and 1.458 (O14). The mean WE value was approximately 1.4 when optimization OR1 was excluded. WE did not alter a lot in the calibration, because it defines the nonlinearity of the water quality in relation to runoff, which was very similar in all the calibration events. WE and its role in the quality model were already discussed in the chapter about sensitivity analyses.

4.5.6 Optimizations without initial build-up

The effects of including IB in the calibration were studied also by running two optimizations without IB, one starting from approximately two months before the first calibration event (O14) and another starting from the beginning of the first calibration event (O15). The purpose was to find out where the other parameters would settle if no IB was included. The results would give an insight to whether the initial loading on the surfaces needs always to be taken into consideration or if the issue of IB could be overcome by starting the analysis substantially earlier than the calibration events.

The results of the two optimizations, O14 and O15, are presented in Figure 24. It is observed that if the simulation is initiated directly from the beginning of the calibration sequence without any initial build-up, the simulation output is worse. The model cannot reach the level of the observed concentrations for the first calibration event and the Nash-Sutcliffe efficiency remains as low as 0.19 and the total load error is as much as -62.9%. The performance statistics also for the third calibration event (optimization O15) are worse than for the optimizations calibrated for all events.

The optimization O14 produced the best results of all the calibrations considering the whole calibration period of three events, with an efficiency *E* of 0.72, a *CORR* of 0.89 and a *TLE* of approximately -2.8%, even though the initial build-up was set as zero. This indicates that IB is not necessary to be included if the simulation can be started early enough for the pollutants to accumulate. However, the simulation runtime for the whole calibration sequence starting from 1 May 2009 was an order of magnitude longer than the simulation that was initiated with the Hot Start File at the beginning of the first event.

The performance statistics of the well-performed optimizations other than O14 were not much worse as can be observed in Table 5. The respective statistics for O13, for example, are *E*=0.72, *CORR*=0.88 and *TLE*=-3.4%. The parameter values were also very close to the ones the other well-performed optimizations gave. BE went to the maximum, MB was more than 260 kg, WC went to the minimum and WE was around 1.4 for all the best optimizations. This means that the model could produce nearly as good results with a dramatically shorter simulation time when the initial build-up was included as a calibration parameter.

4.5.7 Separation of roofs from other surfaces

In the optimization OR1 the impervious surfaces in the catchment were divided into roofs and remaining impervious surfaces and the quality parameters for the roofs of the study catchment were optimized separately. Every roof surface was assigned a land use “Roof” and every other kind of surface type was assigned a land use “Other”. The parameter values for roofs (RMB, RBE, RWC, RWE and RIB) were allowed to fluctuate independently from the parameters of the land use “Other”, but in the same range (RS1) (Table 4).

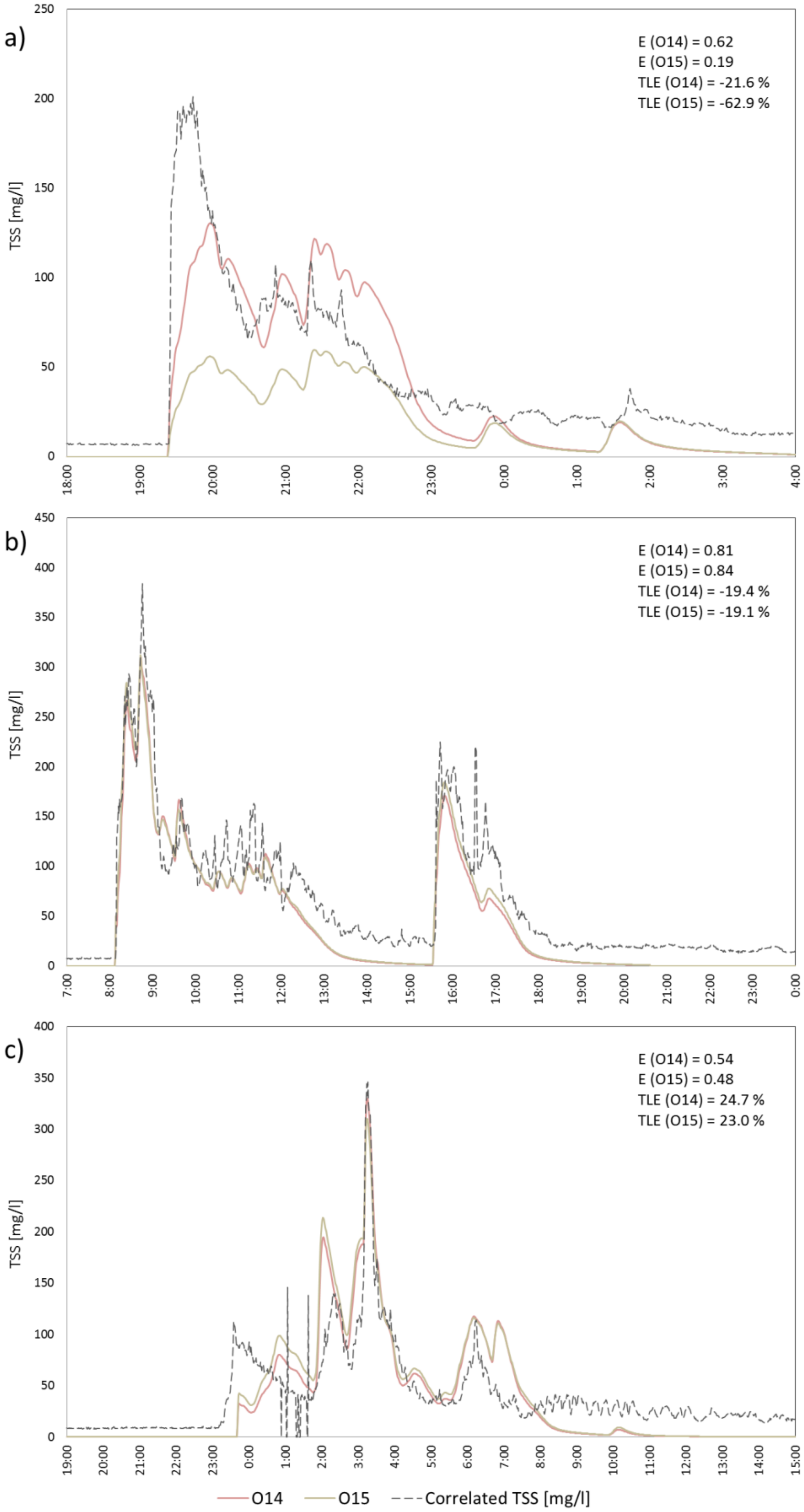


Figure 24. *The simulation results for CAL1 (a), CAL2 (b) and CAL3 (c) when the quality parameters were optimized without any initial build-up starting from 1.5.2009 (O14) and starting from the beginning of the first event 7.7.2009 (O15) with their corresponding Nash-Sutcliffe efficiencies and total load errors.*

The idea behind calibrating separate water quality parameters for roof surfaces and the rest of the surface types was to investigate the modelled contributions of different surface types within the catchment. However, as only end of the pipe measurements were available, the differences between pollutant build-up and wash-off characteristics of different surfaces were not represented in the data and simulations of separate processes inside the catchment could not be validated. End of the pipe data represents the combined effects of build-up, wash-off and routing from the entire catchment. Additionally the data includes the contributions from pervious areas as well (Alley and Smith 1981), which affects the comparison between the simulations and the observed data. Pervious areas were completely excluded from the current scope. For example Borris et al. (2013) remarked, though, that it would be important to extend the research of pollutant build-up and wash-off characteristics to pervious areas as well, because they are likely to have a role in the stormwater quality processes in the changing climate.

An expectation in the beginning of the study was that separation of the different surfaces would lead to a better reproduction of the wash-off curve from the catchment. When all surfaces are assigned the same parameter set, the surfaces are expected to behave in the same manner. This is hardly realistic and increasing the amount of different surfaces (and parameters) allows also for different behaviours from the catchment surfaces. The pollutants can be modelled to, for instance, wash off faster from roofs than from roads, which was suggested by Egodawatta et al. (2009). There is a problem, though, in increasing the amount of parameters, which is the possible autocorrelation of parameters, which is discussed later in this section. Longer optimization times might also become a problem.

Figure 25 presents the calibration results for the optimizations O7 and OR1. These optimizations had the same parameter boundaries and were calibrated for all three calibration events. The only difference was that in OR1 the roofs were separated. The calibration with separated roofs produced a better result than the calibration where all surfaces were equal. When the results for all events jointly were compared, all the performance statistics were better for OR1 than for O7, the *E* being 0.71 compared to 0.61, the *CORR* being 0.87 compared to 0.83 and the *TLE* being -9.2% compared to -18.8% (Figure 25). The performance of OR1 for all the events separately, too, is for the most part better than that of O7 or any other optimization of this study.

With OR1 the later appearing concentration peaks can reach higher than with O7, because even if the BE for the other areas is in the minimum, BE for the roofs can be in the maximum. MB is larger for the roofs, which indicates that there would be more build-up on the roofs than on the other surfaces to be washed off. This is not really physically logical as less pollutants in general accumulate on roofs than on the other surfaces, because the sources of pollutants are fewer. WE also differs notably between the roofs and other surfaces. WE is larger (almost 1.7) for the roofs than for the other surfaces (less than 1), which indicates that the load response from the roofs follows the runoff pattern more closely than that from the other surfaces. Figure 18 clearly illustrated this type of response pattern. The loads from roofs following the runoff pattern closely seems reasonable, because roofs are inclined and smooth so that the load response could easily be quite straightforward and the more volume and intensity the runoff has, the more pollutants are released and directly transported into the gutters. However, whether these characteristics are realistic for the study catchment remains unknown without measurements from the surfaces inside the catchment. It is impossible to reliably single out the contribution of separate processes with end of the pipe data. Investigating the separate processes would be, however, as stated by Deletic et al. (1997), “the best way to understand the complex mechanism of pollution generation”.

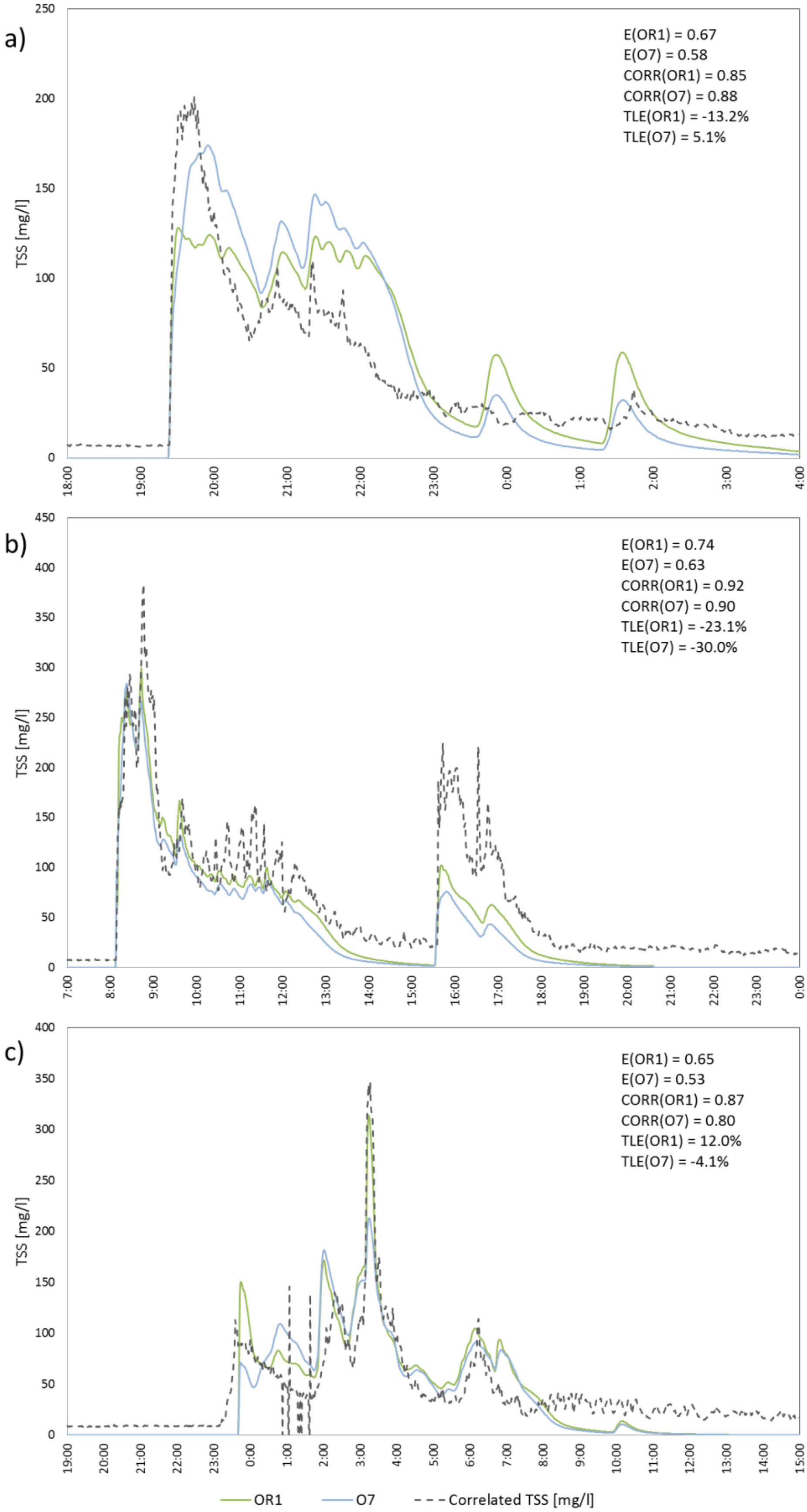


Figure 25. *The calibration results with separated roof surfaces (OR1) compared with optimization O7 for CAL1 (a), CAL2 (b) and CAL3 (c). The parameters were optimized for all three events.*

Measuring the accumulation and wash-off of pollutants on specific surface types in different locations would be necessary to proceed in detailed modelling. Roofs and streets for example are quite easy to be defined in different places. In Australia Egodawatta et al. (2007) and Miguntanna et al. (2013) have measured and modelled the build-up and wash-off characteristics over roads and roofs with the aid of artificial rainfall. According to these studies there is only a small amount of particles on the roof surfaces after a rain event and it could be hypothesized that build-up on roofs starts from zero (Egodawatta et al. 2009). In water quality simulations with SWMM the initial build-up could be neglected for roof surfaces. The model allows for a high resolution water quality simulation otherwise too, because every subcatchment can be assigned their unique build-up and wash-off characteristics. Punishment functions can also be used in adjusting the relations between different parameters or different surfaces, for example so that the maximum build-up over roofs is always smaller than that of roads.

When different surfaces are separated, the quality model may, however, become overparameterized so that along with an increase in the number of parameters the level of correlation between parameters increases (Bertrand-Krajewski 2007). The modeller may end up with several different parameter combinations that produce similar model outputs and consequently may not be able to meaningfully distinguish a unique optimal set of parameters. The calibrated parameters may be suitable for the calibration events but have little to do with the real accumulation and wash-off processes of the catchment. In order to reliably model the variability of contributions of different kinds of surfaces to stormwater quality, more information and measurements should be acquired from different contexts and conditions. Only if the quality parameters for roofs, different pavements, parking lots and streets would be assigned with their own specific, relatively narrow ranges, the model might be able to predict the accumulation and wash-off of pollutants from different surfaces.

4.5.8 Compilation of the optimization results

Figure 26 presents a summary of the calibration events with the range of TSS concentrations that were obtained with all the different optimizations and the range of TSS concentrations that were obtained with the best optimizations. The best optimizations were the ones which had *E* over 0.50 and *CORR* over 0.70 for all the events together and for every event separately. These optimizations were OR1, O7, O10, O13 and O14. The range of performance statistics for the optimizations is also presented in the figure.

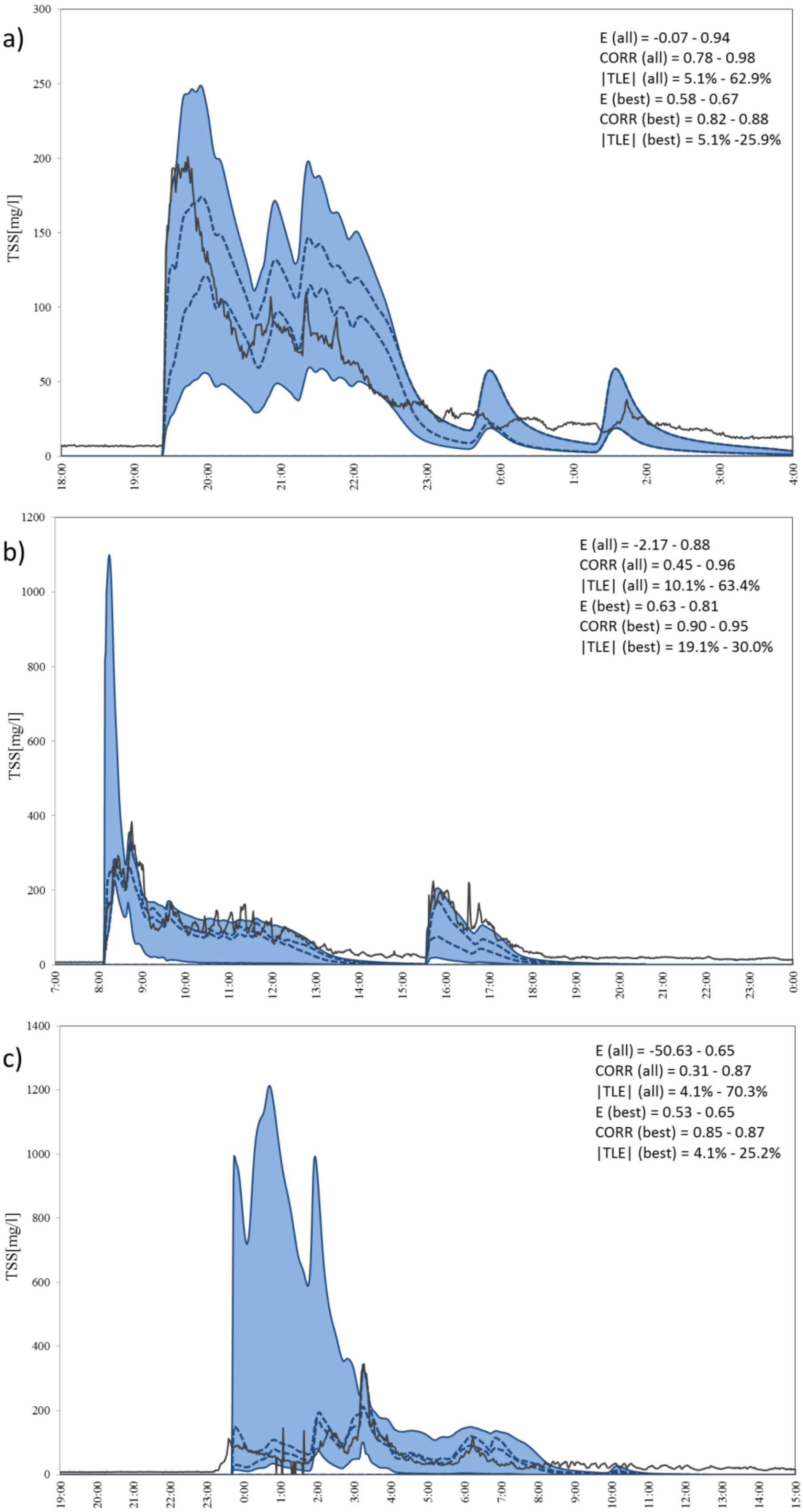


Figure 26. *The range of TSS concentrations obtained with all the optimizations of this study (the blue area) and with the best optimizations (dashed blue lines) and their performance statistics (the* TLE *range does not include the direction of the error) compared to the observed TSS concentrations of the three calibration events (continuous black line) for CAL1 (a), CAL2 (b) and CAL3 (c).*

The large range observed in the graphs of Figure 26 is due to the optimizations for one or two events. The range of calibration results obtained with the best optimizations (OR1, O7, O10, O13 and O14), is notably more narrow and closer to the observed TSS concentration. Table 4 in Section 4.5.2 presents how the optimized parameter values ranged in the optimizations. The parameter values alter clearly less among the best optimizations than among all the optimizations, but even the best optimizations showed a large range for the parameters.

Table 6 presents the performance statistics and highlights the best value for each of the goodness-of-fit criteria (*E*, *CORR* and *TLE*) for each and all events. Table 7 shows the worst performance statistics.

Table 6. *The best performance according to* E, CORR *and* TLE *for each and all events. The values on grey are the best values and the values in the same column are the corresponding values for the other performance criteria and the optimization that has produced the results.*

BEST	All events			1. event			2. event			3. event		
E	0.72	0.72	0.70	0.94	0.94	0.58	0.88	0.88	0.86	0.65	0.65	0.53
CORR	0.89	0.89	0.88	0.98	0.98	0.88	0.96	0.96	0.96	0.87	0.87	0.80
TLE	-2.5%	-2.5%	-2.2%	-17.3%	-17.3%	5.1%	-10.2%	-10.2%	-10.1%	12.0%	12.0%	-4.1%
Opt.	O14	O14	O10	O11	O11	O7	O12	O12	O9	OR1	OR1	O7

Table 7. *The worst performance according to* E, CORR *and* TLE *for each and all events. The values on grey are the worst values and the values in the same column are the corresponding values for the other performance criteria and the optimization that has produced the results.*

WORST	All events			1. event			2. event			3. event		
E	-15.02	-15.02	0.11	-0.07	0.19	0.19	-2.17	-2.17	-0.03	-50.63	-50.63	-0.10
CORR	0.26	0.26	0.71	0.91	0.78	0.78	0.45	0.45	0.78	0.31	0.31	0.81
TLE	-24.2%	-24.2%	-64.3%	42.1%	-62.9%	-62.9%	50.0%	50.0%	-63.4%	14.0%	14.0%	-70.3%
Opt.	O11	O11	O1	O3	O15	O15	O11	O11	O1	O11	O11	O1

The same optimization gave the best *E* and the best *CORR*, and also the worst *E* and the worst *CORR* with one exception (Tables 6 and 7). Only for the first event the worst *E* and the worst *CORR* are given by different optimizations, O3 and O15, respectively. Without exceptions the best and the worst (the smallest and the largest) *TLEs* are produced by different optimizations than the ones that give the best and worst *E* values.

4.6 Validation

4.6.1 Validation for 2013 data

The validation events from the 2013 data were introduced in Section 4.1.3. The water quality simulation results and their performance statistics are presented for the validation events in Figure 27. The first three validation events, VAL1, VAL2 and VAL3, were the best events from 2013 regarding the match between rainfall and runoff. Figure 27 demonstrates, though, that the turbidity measurements for those events are not successful for most part and the validation using these events cannot be completely reliable. The runoff quality continuity errors of these simulations were elevated, 13.7% for OR1 and -9.0% for O13, which also affects the reliability of the validation.

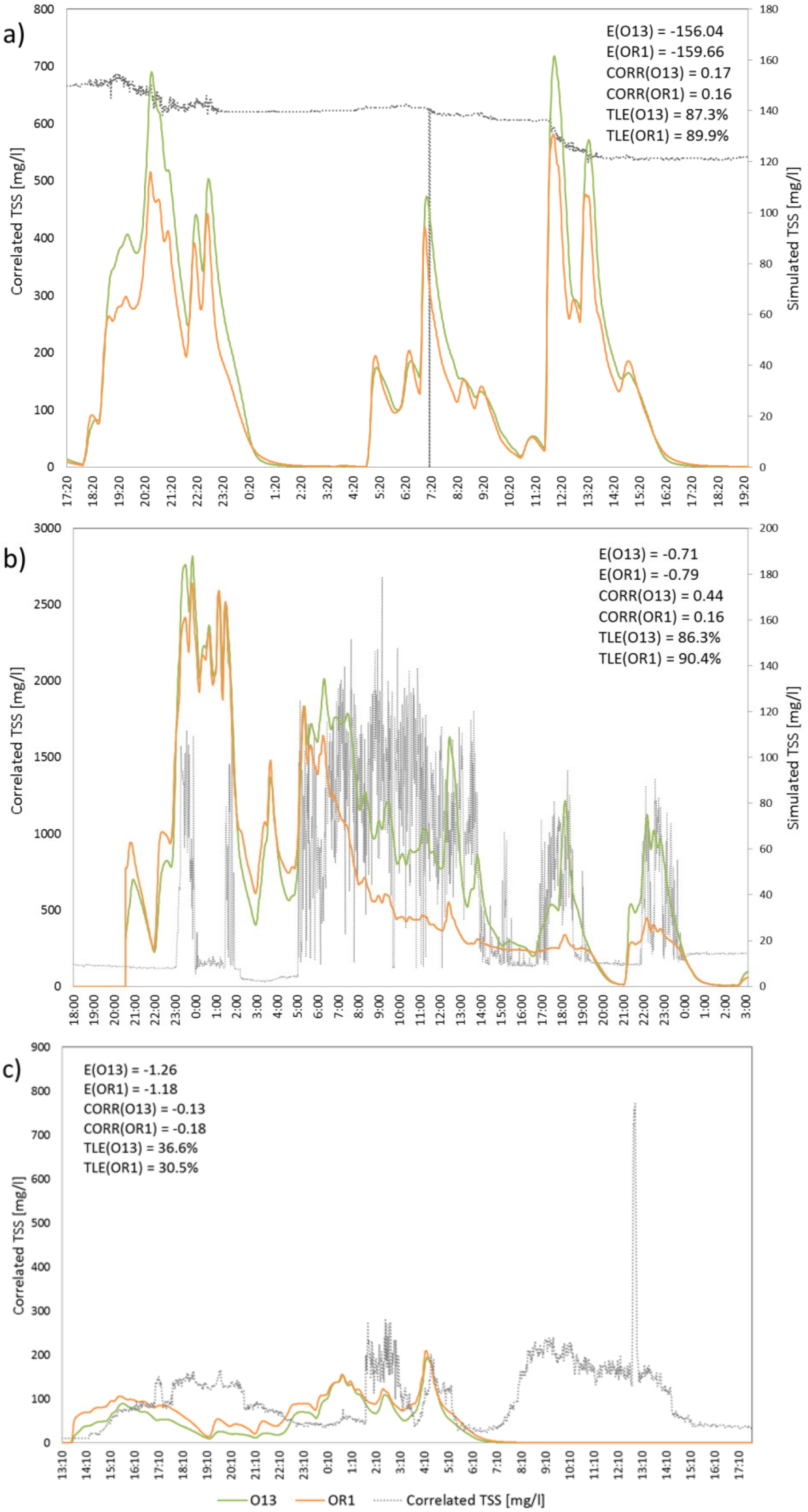


Figure 27. *The simulated TSS concentrations for the 2013 validation events, VAL1 (a), VAL2 (b) and VAL3 (c) with their corresponding performance statistics. The axes and the scale are different for the correlated and simulated TSS concentrations.*

The turbidity values and accordingly the TSS concentrations of the 2013 validation events are very high for VAL1 and VAL2 (Figure 27). Although regular cleaning of the turbidity meter took place throughout the monitoring period, the offset in these events is most likely due to dirt accumulating on the sensor. The offset grows higher during the summer (from VAL1 to VAL2), which also suggests dirt accumulation. It could be further deciphered that a more thorough cleaning has taken place before the third validation event which is why there is no offset during VAL3.

The turbidity measurements of VAL1 (Figure 27 a) are likely erroneous as there is only a minor fluctuation in the measured turbidity with the recorded flow. There is also a zero value at about 7:20, which is clearly an error. Another interpretation could also be made. The runoff depth and intensities during this event were low, the total runoff depth being 4.8 mm and the peak flow reaching only to approximately 30 l/s. Therefore the TSS concentrations coming from the catchment surfaces might have been lower than the faulty concentrations indicated by the dirty turbidity meter and could not thus be shown in the measurements. The small fluctuations and the small reduction in concentration observed in the measurements would be due to the stagnant water before the dam in the measuring point being first mixed and then settling.

The turbidity of VAL2 (Figure 27 b) is noisy, which is undoubtedly an error in the measurements. In some cases the excessive noise of turbidity measurements is caused by high water velocities that cause turbulence and air bubbles. It seems that there is turbidity data missing approximately from 0:00 to 2:00. During this event the flow velocities were high (mostly over 40 l/s and reaching up to 120 l/s), but it hardly accounts for all the noise. Aside from the large offset, there is a similar pattern in the correlated TSS concentration and the O13 output.

For the third validation event VAL3 (Figure 27 c) the fit between measured and modelled flow was good except for the end of the event, where there was observed flow and TSS concentration, but no rainfall. In the simulation no concentration is observed as it cannot be modelled without rainfall that would produce runoff. The measurements of VAL3 are likely correct, except for the very high peak at around 13:00, as the scale is realistic and the pattern is realistic. The latter part of the event could be caused by street cleaning with water or a leakage. These types of random occurrences are obviously outside of the capabilities of a mathematical model.

The flow simulations of the 2013 events are inevitably somewhat erroneous, because of the distance between the rain that in reality produced the flow at the catchment outlet and the rain that was used in the simulation. When the rain is not exactly the same, the flow cannot be the same and as a consequence the model cannot reproduce the observed TSS concentration. The graphs in Figure 27 indicate that even if the model could have produced the exact same flow curve as was observed, the simulated concentration curve would not be the same as the observed as it is not following the flow pattern. Notably better simulation results could most likely be gained if the equipment-related issues could be overcome.

Bertrand-Krajewski and Muste (2007a) stressed that when there is a substantial change in for example the measurement technique of the variables, the model should be recalibrated. The calibrated model should not be used for very different data than is used in the calibration. This study works as a good example of this, as the calibration events were from 2009 when the turbidity measurements were conducted differently and also different rainfall data were available and used than during the measurement period in 2013. Indeed, the validation for events VAL1, VAL2 and VAL3 did not yield good results (Figure 27), but in this case the conclusion that the calibration was not successful cannot be drawn. The rainfall and runoff data were not measured from the same location and the water quality data was questionable, which rather accentuates the importance of careful and reliable measurements than suggests that water quality could not be modelled reliably with SWMM and the selected functions.

Many previous studies (e.g. Kusari and Ahmedi 2013) have stressed the importance of frequent sampling and careful measurements when studying the behaviour and quality of stormwater in urban areas. This is backed up also by the results of this study. The data needs to be reliable and suitable for modelling and therefore it is crucial that the measurements are carefully planned and supervision of the measurements is also arranged. The data should be accompanied with remarks and observations from the measuring station from the time of measurements so that as little time as possible would be wasted in difficult data interpretation.

4.6.2 Validation events 2009-2010

As the validation with the 2013 data was affected by data quality problems, five events from 2009 and 2010 were used. The basic facts about the validation events were already presented in Table 1 and the observed and simulated flow as well as the correlated TSS concentration is presented for the validation events (VAL4-VAL8) in Figure 28 and Figure 29. The hydrological performance of these validation events is very good with *E* ranging from 0.68 to 0.96, the *CORR* ranging from 0.86 to 0.99 and *VE* ranging from -3.4% to 29.7%. The turbidity-correlated TSS concentration, also seems to be coherent and for the most part logical. Validation events VAL5, VAL6 and VAL7 follow the flow pattern as would be expected. Additionally, the first flush phenomenon can be observed for these events as elevated concentrations in the beginning of the event.

The validation events VAL4 (Figure 28 a) and VAL8 (Figure 29 b) do not have a similar flow-concentration pattern as the other three events. However, all the measurements seem to be reliable and the water quality patterns can be explained. The TSS concentration in VAL8 is quite the opposite from what would be expected in the first place, because the concentration rises very high with the first part of the runoff event and stays high, but then declines when runoff increases. It is possible that the pattern is caused by the first flush phenomenon. The first flush could last quite long (in VAL8 about 2 hours), when the runoff volume and intensities are low. However, there was an event only approximately 12 hours earlier that probably had already removed the readily mobilised pollutants, which suggests that elevated concentrations are not due to the first flush.

One possible explanation for the TSS concentration curve in VAL8 (Figure 29 b) is that the pollutants in the stagnant water in front of the dam, where the turbidity measurements are taken, are settled at first, but when water starts coming to the dam, the stagnant water gets mixed and the turbidity rises. After that, a heavier runoff comes to the measuring point and actually dilutes the water and reduces the turbidity to its original level. The high base concentration (around 125 mg/l in comparison to less than 20 mg/l for the other events) and the observed fluctuations result from TSS that was already in the water before the dam and the pollutants coming from the catchment during the event do not actually show at all. The concentration curve of VAL4 is not as easily explained, but can result from a similar phenomenon than VAL8. It is possible that the runoff brings pollutants to the measurement point and increases turbidity during the first part of the event. Then, as the flow decreases, the large amount of pollutants starts settling and is released again by the subsequent flow. VAL4 is long-lasting (24 h) and its flow intensity and dynamics differ from the other events, which might also be the reason for a different behaviour as the other events.

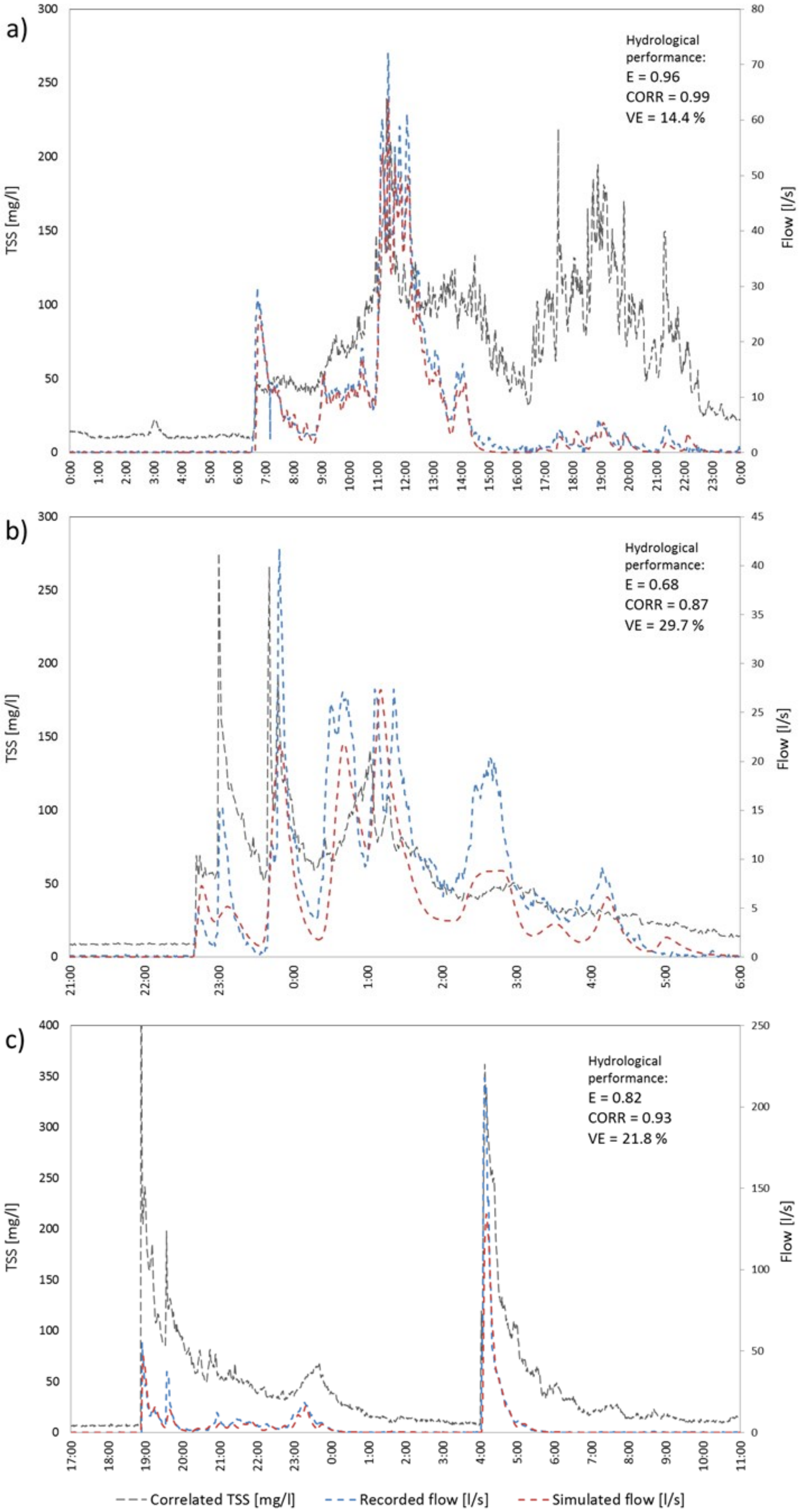


Figure 28. *The simulated and observed flow, the correlated TSS concentration and the hydrological performance of the validation events VAL4 (a), VAL5 (b) and VAL6 (c).*

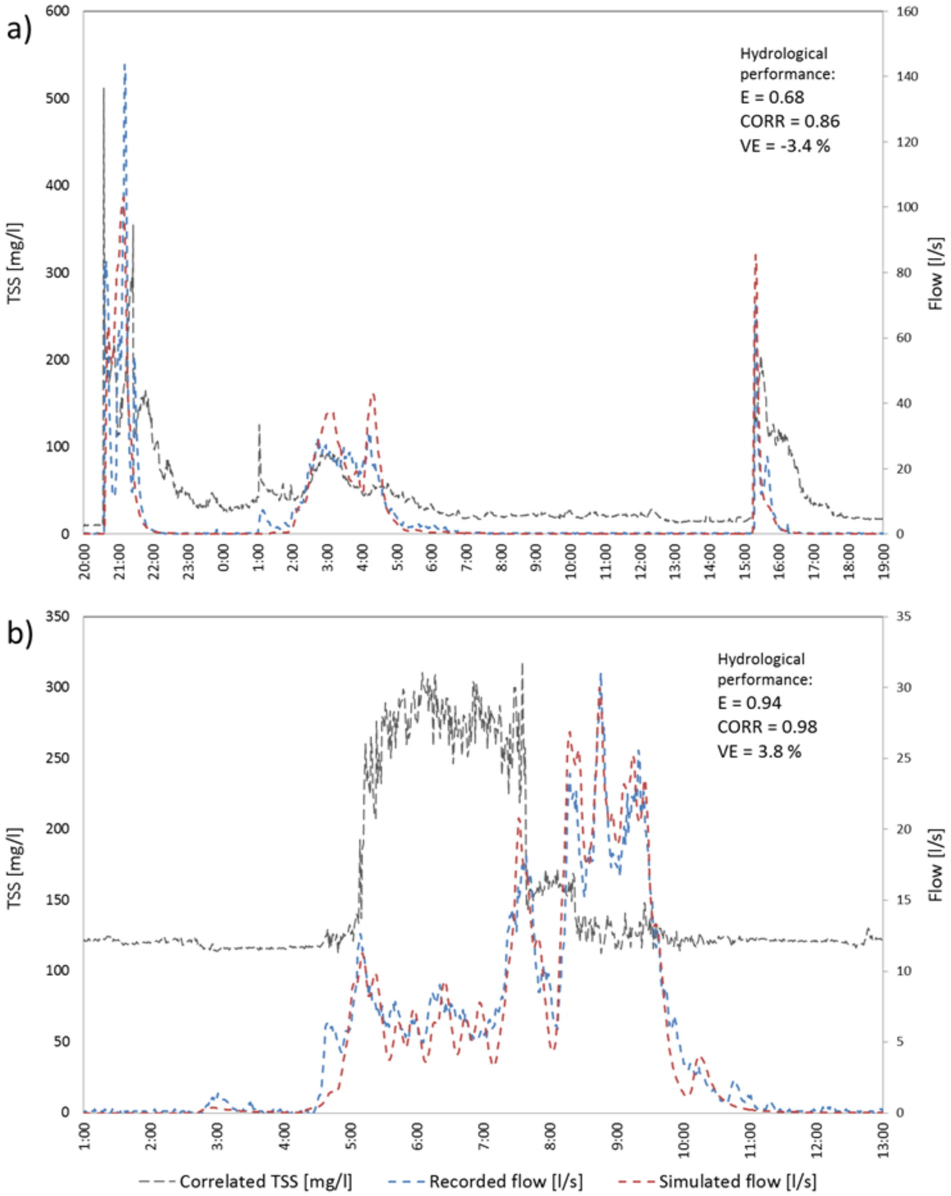


Figure 29. *The simulated and observed flow, the correlated TSS concentration and the hydrological performance of the validation events VAL7 (a) and VAL8 (b).*

4.6.3 Validation for 2009-2010 data

The optimized parameter values of OR1, O7, O10 and O13, were used in the validation and the simulated TSS concentrations are presented in Figure 30 and Figure 31 with the performance statistics of the event simulations.

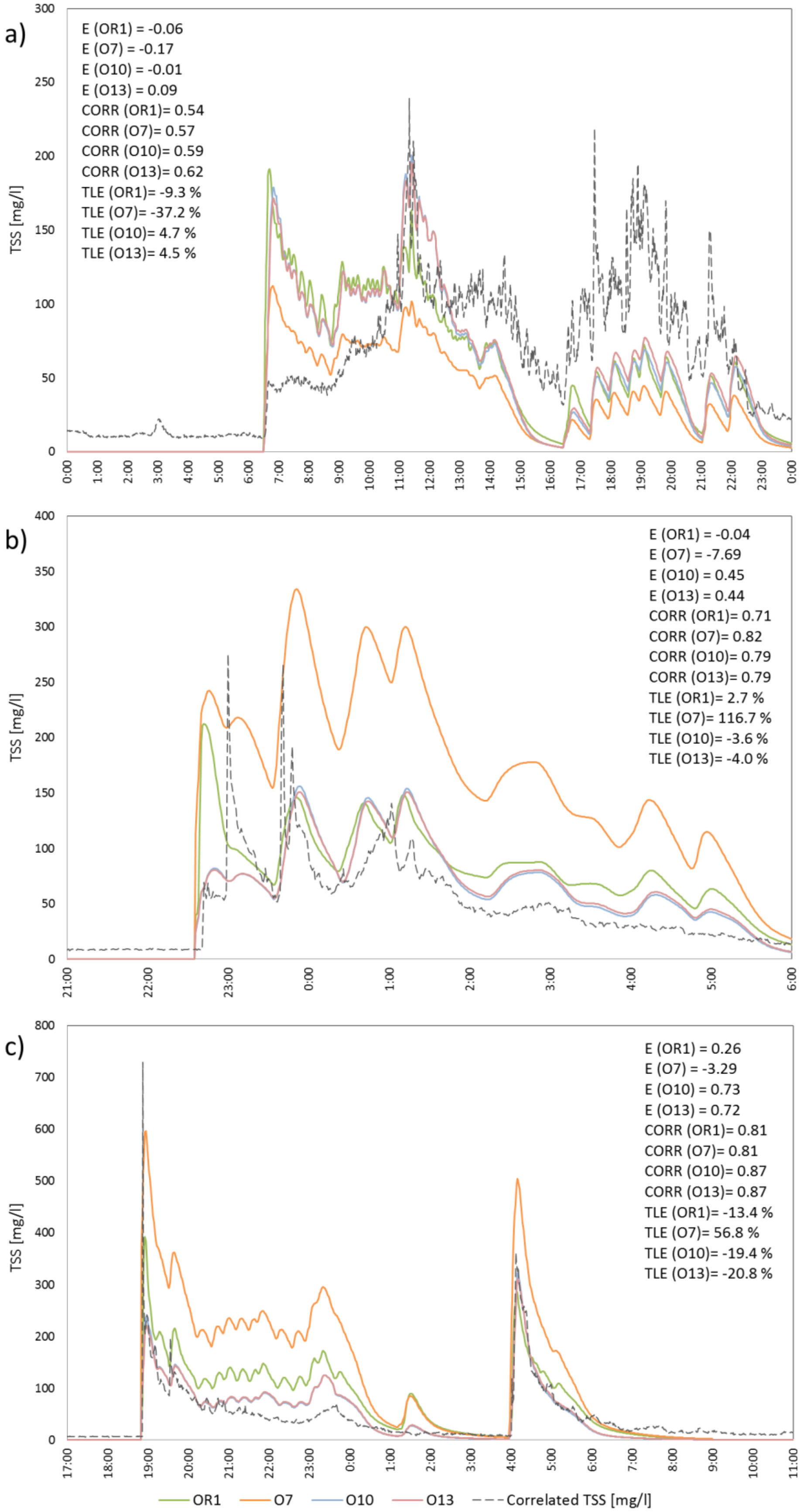


Figure 30. *The simulated TSS concentrations and the performance of the quality simulations for the validation events VAL4 (a), VAL5 (b) and VAL6 (c).*

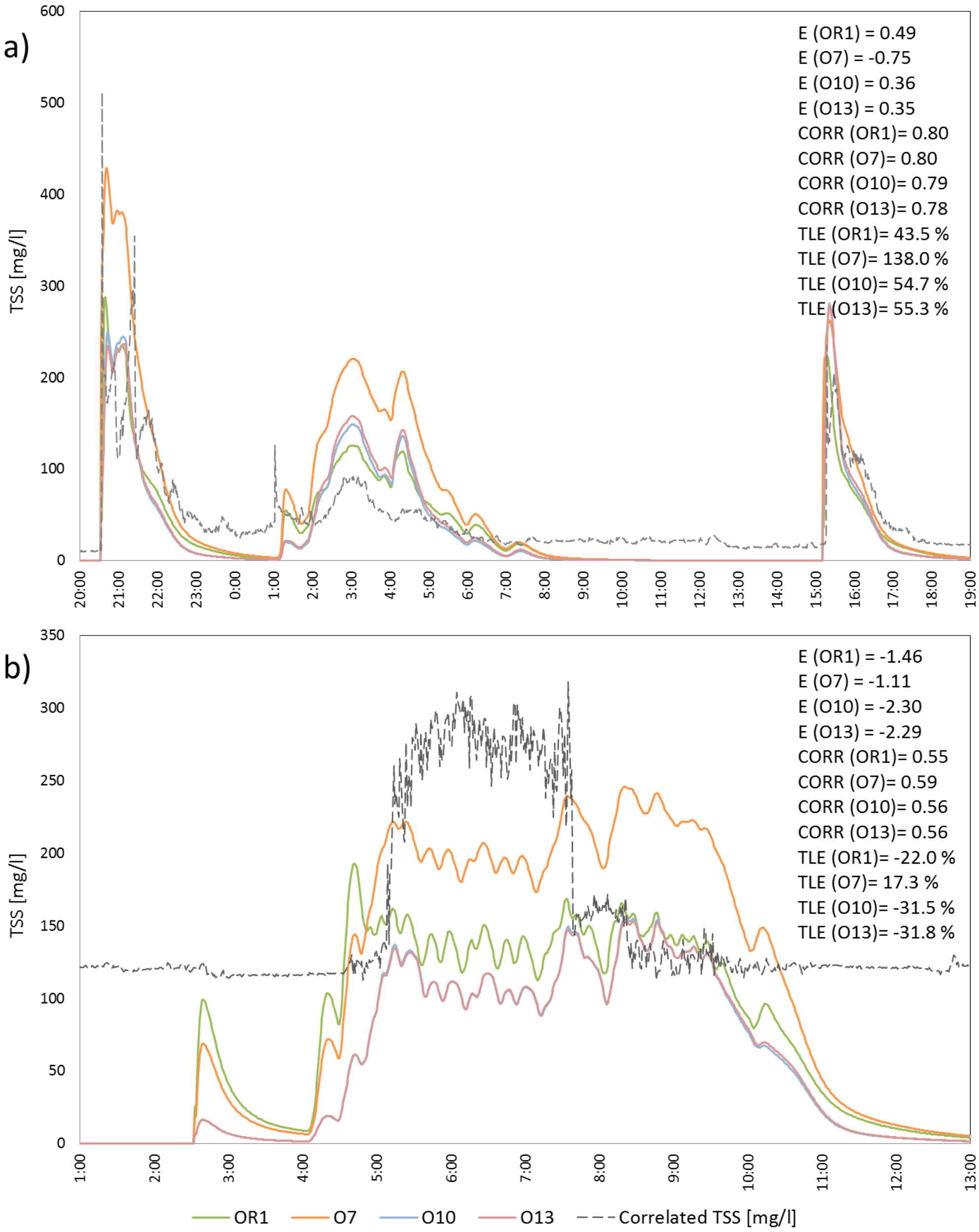


Figure 31. *The simulated TSS concentrations and the performance of the quality simulations for the validation events VAL7 (a) and VAL8 (b).*

The optimizations O10 and O13 had the best overall performance during the validation sequence and O7 performed the worst. O7 overestimated the TSS concentrations of the validation events VAL5-VAL8 and underestimated those of VAL4 (Figures 30 and 31). The reason for the poor performance of O7 in the validation is most likely a large wash-off coefficient. The optimized parameters of O7 worked well for the calibration period, because the wash-off-reducing effect of the optimized initial build-up lasted over the whole period. However, the effect of the initial build-up reduces in time and therefore the later events in the validation sequence are not affected by the initial state. The wash-off coefficient optimized with an initial build-up that small produces too high concentrations in the later events.

The optimization O10 and O13 produced similar results for all the validation events VAL4-VAL8. The *E* values for VAL5 and VAL7 were around 0.45 and 0.36, respectively, and 0.73 (O10) and 0.72 (O13) for VAL6. The values for *CORR* ranged from 0.56 to 0.87 for all the events. Krebs et al. (2014) reported that the hydrological model performance was overall better for larger events than it was for smaller events. In this study the runoff intensity was reflected to results in a similar way so that the events with a smaller intensity could not be reproduced as well with the water quality model as the events, with a higher intensity runoff.

The performance of OR1 was better for the validation events than the performance of O7. This backs up the hypothesis made in the calibration that separating the different surfaces gives a better result, because the setting is more realistic. However, the hypothesis cannot be verified, because measurements from inside the catchment from the different surfaces are not available for comparison. The model can produce a better result, but the result might not be more realistic even though the assumption of different surfaces acting differently is.

The runoff quality continuity errors for the 2009-2010 validation sequence were higher than the errors for the calibration sequence. O7 produced the highest continuity error (-27.9%) followed by OR1 (-22.7%), O10 (-21.3%) and O13 (-21.1%). The runoff quality continuity errors accumulate and grow higher when the simulation period becomes longer. Therefore the errors are high also for O10 and O13 for the validation period even though they were low for the calibration sequence.

It can be noted from the validation that when TSS concentration does not follow the observed flow pattern, the model cannot reproduce the TSS concentration. The modelled concentration is tied to the modelled flow with a constant exponent that dictates the pattern of the response (Equation 5). When the water quality response to runoff dynamics is similar to that in the calibration, the model can produce a good estimate. However, there are many factors affecting the turbidity changes in a sewer before a dam leading to various water quality responses to runoff flow. Therefore an accurate prediction for all the events with one function is practically impossible.

The validation illustrates again how important it is to first of all calibrate the model and secondly validate it. The calibration results for O7, for example, seemed satisfactory at least in comparison to the other optimizations of the same range set (RS1) (Appendix 5), but the performance in the validation was still not good. The more events are included in the calibration and also in the validation, the more reliable the results are.

4.7 Limitations and uncertainties

Modelling stormwater quality is a complex task and many kinds of limitations and uncertainties are faced during the process starting from the rainfall and runoff measurements and ending up with the difficulties in interpreting the validation results. If the hydrological model does not work, the quality model cannot work either. Errors in rainfall and runoff measurements play therefore a big role, as could be observed in the 2013 validation events. The validation with the 2013 data had many limitations and was open to interpretations first of all because the rainfall and runoff data were from different locations and second of all because the turbidity data was for many parts erroneous and questionable. The data from 2009-2010 was on the other hand much more trustworthy as the measurements were successful and from the same location or from very close to the catchment outlet. Even still, as stated by Deletic et al. (1997), "the lack of reliable field water quality data is the main reason for the general unreliability of stormwater quality models". The limitations of the measurements are directly transferred to the simulation results as the calibration can only be as accurate as the data it is calibrated with. Therefore also the fact that measurements were carried out only at the outlet of the catchment limits the possibilities of using the simulation results widely. Especially for the purpose of modelling processes inside the catchment, as was attempted with the optimization OR1, this type of data is very limited.

The turbidity of the stormwater at the outlet of the study catchment was selected to represent the water quality. The turbidity was converted to TSS concentration with a correlation obtained from measurements from the outlet. However, the number of samples used in the conversion was only 29, and additionally the R^2-value for the correlation was only moderate (0.57), which brings uncertainty to the conversion. The absolute values of the loads calculated for the used events cannot therefore be trusted completely, but on the other hand both measured and modelled loads were calculated with the same correlation, so that their relative difference remains. Also it should be kept in mind that the TSS concentration is only one index of water quality and it does not represent concentrations of dissolved matter.

The hydrology greatly affects the water quality simulation. An inevitable error in total loads arises from the differences between the observed and the modelled flow. The model can produce neither concentration nor load if flow cannot be simulated even if in reality there were variations in the water quality expressed by turbidity fluctuations. The errors in the hydrological part of the model are therefore directly transferred to the modelled quality.

Some error to the modelling results might come from the model setup, such as the selected computational time steps. Using the Hot Start File may bring some error if the hydrological output is not exactly the same as without it. In this study some of the model continuity errors of the water quality simulations were moderately high, even over 9% in the calibration and over 20% in the validation, which also increases the uncertainty of the model output. Continuity errors might be an indication of for example unrealistic parameter values, but they were not intended to be minimized, because the other sources of error and limitations were considered more important than this uncertainty.

During the selected calibration period the accumulation of pollutants only occurred over a couple of days which might give the model an erroneous picture of the accumulation characteristics of the catchment. For example, the build-up cannot reach its maximum during two days unless the build-up exponent is given a large value. The time during which the maximum build-up is reached is different for different catchments and surfaces in different studies. MB has been suggested to be reached for example during 21 days (Egodawatta & Goonetilleke 2010).

When the exponential wash-off function is used, the wash-off loads are directly proportional to the amount of pollutants built up on the surfaces. Therefore, if the build-up parameters are poorly estimated, the wash-off loads are equally biased. This can be a large source of error as the model is quite sensitive towards the maximum build-up and initial build-up. The errors made in estimating build-up in addition to those made in the hydrological model and the precipitation and runoff measurements pile up in the wash-off.

Defining the parameter boundaries for the optimizations is a source of error in the calibration. It is difficult to estimate the boundaries in the first place because the quality parameters are not directly related to any measurable variables apart from the maximum and initial build-up and can therefore be only either calibrated or based on literature values. Usually measurements for the maximum and initial build-up are not available and obtaining them would be laborious and/or costly. The parameter values vary greatly between different catchments, which causes limitations in the transferability of the results from one catchment to another. This was noted in the optimization O7, as the parameter boundaries (RS1) were created following the ones obtained in previous studies, but the validation proved that the parameter ranges were not entirely suitable for the study catchment.

The best calibration results in this study were obtained with optimizations for three calibration events, but even three events is quite a small sample of how varied the water quality response of a catchment is to different events. This was demonstrated in the validation as the model could only produce one kind of concentration response to runoff with a fair accuracy. Even if the simulation results would be worse, a larger number of calibration events would make the calibration as well as validation more trustful. The more validation events are used, the more there will be events that cannot be well predicted with the model, but the results are more transparent.

5 Conclusions and recommendations

The main objectives of this study were to calibrate SWMM water quality parameters for a Finnish urban catchment and to conduct sensitivity analyses that would clarify the interrelations of the quality parameters and their effect on the SWMM simulation output. Rainfall, runoff and turbidity data from two different research projects, Stormwater programme and Urban Laboratory project, were in use from the catchment and its surroundings. The turbidity measurements were converted to TSS concentrations, which were considered to represent the water quality. The water quality parameters selected for calibration were the initial build-up of pollutants on catchment surfaces and four parameters (the maximum build-up, build-up exponent, wash-off coefficient and wash-off exponent) that appear in the exponential functions in SWMM selected to represent the pollutant build-up and wash-off characteristics.

Five different optimizations yielded calibration results that exhibited an efficiency $E>0.50$, a correlation coefficient $CORR>0.80$ and a total load error TLE less than ±30% for all the events they were calibrated for separately and jointly. Four of these optimizations were used to simulate the validation sequence of five events that followed the calibration sequence. Only one event could be predicted well with two of the optimizations with an $E>0.70$, a $CORR>0.80$ and a TLE of approximately ±20%. Two events were moderately predicted with the same two optimizations with $E>0.35$ and a $CORR>0.78$. The remaining two events could not be predicted with any of the optimized parameter sets. A validation period of three events from a different year than the calibration sequence was simulated with two of the mentioned optimizations, but none of these events could be predicted well due to limitations in the data.

The model simulations were affected by a model continuity error suggesting that there were errors in the numerical solution of the water quality model for some of the studied parameter combinations. A high wash-off rate correlated with high continuity errors. Some useful conclusions could be drawn about the interrelations between the parameters from the sensitivity analyses before and after the calibration. Within the limitations of the available data and the calibration setup the following conclusions can be drawn and recommendations are given:

- The usability of the Urban Laboratory data was limited in this study because of a considerable distance between the measurement points for runoff and rainfall and incoherent turbidity data. Detailed data, whose reliability can be traced, is needed if alterations in the water quality during a rainfall-runoff event are to be simulated. It would be advisable to collect the rainfall, runoff and turbidity data from the same location and in minute-scale.

- The optimization without including initial build-up (O15) produced worse results than the optimizations including initial build-up. In future studies over stormwater quality parameters it is recommended that the initial build-up is included in the calibration if no approximation for it can be obtained with measurements. Yet further quantitative research should be undertaken in the purpose of finding out how and on what kinds of conditions it is done.

- The exponential build-up and wash-off functions are capable of producing fairly accurate predictions when the water quality fluctuations follow the fluctuations of runoff in a consistent way. They are also adaptable to different types of runoff-quality relationships. However, their applicability in predicting water quality fluctuations that do not follow runoff dynamics is limited.

- The sensitivity analyses revealed that the quality model is not sensitive to perturbations in the hydrological parameters at least in comparison to those of the quality parameters.

Yet it is, if not a prerequisite, a strong recommendation to have a calibrated and working hydrological model before moving on to water quality modelling, because errors in hydrology are directly transferred to the quality model.

- The quality model for this catchment was the most sensitive to the wash-off exponent followed by the maximum build-up and the wash-off coefficient. Of the quality parameters the model was the least sensitive to the build-up exponent.

- The most evident interrelations between the calibrated parameters were: i) the maximum build-up and the initial build-up are related so that IB is more important in the beginning of the simulation but the longer the simulation lasts, the more important MB becomes; ii) increasing the wash-off coefficient decreases the effect of the build-up exponent and vice versa; iii) WC is also related to MB and IB so that the larger WC is the smaller MB and IB tend to be; and iv) the wash-off exponent is quite independent and strongly affects the water quality patterns in response to runoff. If the interrelations of the water quality parameters are wished to be understood more profoundly, more thorough mathematical investigations are needed.

- Looking at the number of calibration events, two major tendencies can be observed: i) Optimizing the parameters for one or two events tends to give very good results for the event(s) they were optimized for, but poor performance for the other events. ii) When the performance statistics for the entire calibration sequence are compared, calibration for one event tends to give the worst results and calibration for three events the best results. The wash-off patterns within the events also have an effect on the results. This gives reason to assume that the more calibration events are included, the better results are yielded, or, at least the results are more trustworthy.

- According to the results of this study the calibration of SWMM for the water quality requires several events over a long period of time. The events selected for calibration should be different in their intensities, durations and flow pattern so that the model could encompass as many different events as possible. In addition, the number of dry days between events should vary so that the accumulation of particulate matter on the surfaces could be truthfully estimated.

- Studying high resolution modelling for water quality is recommended. Calibration of independent parameters for roofs and other surfaces gave a good performance and, although the results were supposedly not completely realistic, the model seemed to be capable of separating different processes. The truthfulness of the simulation output for roofs and other surfaces separately could not be verified, as only end of the pipe data was available. Modelling different surfaces simultaneously needs to be further studied and as end of the pipe measurements are not sufficient to represent independent processes inside the catchment, separate measurements are needed from different surfaces.

- Validation should always be included in the calibration process. It was shown in this study, that even if the optimizations yield good results for the calibration events, the optimized parameters do not necessarily perform well for events outside the calibration sequence.

The calibration results can be applied to catchments with similar climatic conditions, land use and other important characteristics, such as imperviousness, and the stormwater quality parameterization is suggested for comparison in future studies. Consideration is required, though, because the transferability of the results is still limited, as the stormwater quality processes are complex and very dependent on the catchment characteristics.

References

Aaltonen, J., Hohti, H., Jylhä, K., Karvonen, T., Kilpeläinen, T., Koistinen, J., Kotro, J., Kuitunen, T., Ollila, M., Parvio, A., Pulkkinen, S., Silander, J., Tiihonen, T., Tuomenvirta, H. and Vajda, A. 2008, *Rankkasateet ja taajamatulvat (RATU),* Helsinki: the Finnish Environment Institute. (Suomen ympäristö 31/2008). 123 pp.

Airola, J., Nurmi, P. and Pellikka, K. 2014, *Huleveden laatu Helsingissä*, Helsinki: the City of Helsinki Environment Centre. (Helsingin kaupungin ympäristökeskuksen julkaisuja 12/2014). 68 pp.

Akan, A.O. and Houghtalen, R.J. 2003, *Urban hydrology, hydraulics, and stormwater quality engineering applications and computer modeling,* New Jersey: J. Wiley & Sons. 392 pp.

Alley, W.M. and Smith, P.E. 1981, "Estimation of accumulation parameters for urban runoff quality modeling", *Water Resources Research,* 17(6), pp. 1657-1664.

Al-Yaseri, I., Morgan, S. and Retzlaff, W. 2013, "Using turbidity to determine total suspended solids in storm-water runoff from green roofs", *Journal of Environmental Engineering,* 139(6), pp. 822-828.

American City & County 1996, "CSO control revitalizes stretch of the Mississippi", 111(13), 30 pp.

ASCE Task Committee on Definition of Criteria for Evaluation of Watershed Models, Watershed Management Committee, Irrigation and Drainage Division 1993, "Criteria for evaluation of watershed models", *Journal of Irrigation and Drainage Engineering*, 119(3), pp. 429-442.

Avellaneda, P., Ballestero, T.P., Roseen, R.M. and Houle, J.J. 2009, "On parameter estimation of urban storm-water runoff model", *Journal of Environmental Engineering,* 135(8), pp. 595-608.

Baffaut, C. and Delleur, J.W. 1990, "Calibration of SWMM runoff quality model with expert system", *Journal of Water Resources Planning and Management,* 116(2), pp. 247-261.

Barbosa, A.E., Fernandes, J.N. and David, L.M. 2012, "Key issues for sustainable urban stormwater management", *Water research,* 46(20), pp. 6787-6798.

Bertrand-Krajewski, J. 2004, "TSS concentration in sewers estimated from turbidity measurements by means of linear regression accounting for uncertainties in both variables", *Water science and technology,* 50(11), pp. 81-88.

Bertrand-Krajewski, J. 2007, "Stormwater pollutant loads modelling: epistemological aspects and case studies on the influence of field data sets on calibration and verification", *Water Science And Technology,* 55(4), pp. 1-17.

Bertrand-Krajewski, J. and Muste, M. 2007a, "[Chapter 6] Understanding and managing of uncertainty" in *Data requirements for integrated urban water management*, eds. T.D. Fletcher & A. Deletic, London: Taylor & Francis Ltd, 392 pp.

Bertrand-Krajewski, J. and Muste, M. 2007b, "[Chapter 8] Data validation: principles and implementation" in *Data requirements for integrated urban water management*, eds. T.D. Fletcher & A. Deletic, London: Taylor & Francis Ltd, 392 pp.

Bilotta, G.S. and Brazier, R.E. 2008, "Understanding the influence of suspended solids on water quality and aquatic biota", *Water research,* 42(12), pp. 2849-2861.

Booth, D.B., Hartley, D. and Jackson, R. 2002, "Forest cover, impervious-surface area, and the mitigation of stormwater impacts", *Journal Of The American Water Resources Association,* 38(3), pp. 835-845.

Borris, M., Viklander, M., Gustafsson, A. and Marsalek, J. 2013, "Modelling the effects of changes in rainfall event characteristics on TSS loads in urban runoff", *Hydrological Processes,* 28(4), pp. 1787-1796.

Brodie, I.M. and Dunn, P.K. 2010, "Commonality of rainfall variables influencing suspended solids concentrations in storm runoff from three different urban impervious surfaces", *Journal of Hydrology,* 387(3–4), pp. 202-211.

Burian, S.J., Streit, G.E., McPherson, T.N., Brown, M.J. and Turin, H.J. 2001, "Modeling the atmospheric deposition and stormwater washoff of nitrogen compounds", *Environmental Modelling & Software,* 16(5), pp. 467-479.

Cambez, M.J., Pinho, J. and David, L.M. 2008, *Using SWMM 5 in the continuous modelling of stormwater hydraulics and quality,* in the 11th International Conference on Urban Drainage, 2008, Edinburgh, Scotland, UK.

Cho, J.H. and Seo, H.J. 2007, "Parameter optimization of SWMM for runoff quantity and quality calculation in a eutrophic lake watershed using a genetic algorithm", *Water Science and Technology: Water Supply*, 7(5-6), pp. 35-41.

Chow, M.F., Yusop, Z. and Toriman, M.E. 2012, "Modelling runoff quantity and quality in tropical urban catchments using Storm Water Management Model", *International Journal of Environmental Science and Technology,* 9(4), pp. 737-748.

Christensen, V.G., Rasmussen, P.P. and Ziegler, A.C. 2002, *Comparison of estimated sediment loads using continuous turbidity measurements and regression analysis*, in the Turbidity and Other Surrogates Workshop, April 30-May 2, 2002, Reno, NV.

De Feo, G., Antoniou, G., Fardin, H.F., El-Gohary, F., Zheng, X.Y., Reklaityte, I., Butler, D., Yannopoulos, S. and Angelakis, A.N. 2014, "The historical development of sewers worldwide", *Sustainability,* 6(6), pp. 3936-3974.

Deb, K., Pratap, A., Agarwal, S. and Meyarivan, T. 2002, "A fast and elitist multiobjective genetic algorithm: NSGA-II", *Evolutionary Computation, IEEE Transactions on,* 6(2), pp. 182-197.

Deletic, A., Maksimovic, C. and Ivetic, M. 1997, "Modelling of storm wash-off of suspended solids from impervious surfaces", *Journal of Hydraulic Research,* 35(1), pp. 99-118.

Dotto, C.B., Kleidorfer, M., Deletic, A., Fletcher, T.D., McCarthy, D.T. and Rauch, W. 2010, "Stormwater quality models: performance and sensitivity analysis", *Water Science & Technology,* 62(4), pp. 837-843.

Egodawatta, P. and Goonetilleke, A. 2006, *Characteristics of pollutants built-up on residential road surfaces*, in the 7th International Conference on Hydroscience and Engineering (ICHE 2006), 10-13 September 2006, Philadelphia, USA.

Egodawatta, P. and Goonetilleke, A. 2008a, *Modelling pollutant build-up and wash-off in urban road and roof surfaces*, in the 31st Hydrology and Water Resources Symposium and the 4th International Conference on Water Resources and Environment Research, Engineers Australia, 14-17 April 2008, Adelaide, Australia.

Egodawatta, P. and Goonetilleke, A. 2008b, "Understanding road surface pollutant wash-off and underlying physical processes using simulated rainfall", *Water Science & Technology,* 57(8), pp. 1241-1246.

Egodawatta, P. and Goonetilleke, A. 2010, "An innovative modelling approach to assess stormwater pollutant loads from the port of Brisbane, Australia", *Proceedings of the International MIKE by DHI Conference*, 6-8 September 2010*,* Copenhagen, Denmark.

Egodawatta, P., Thomas, E. and Goonetilleke, A. 2007, "Mathematical interpretation of pollutant wash-off from urban road surfaces using simulated rainfall", *Water research,* 41(13), pp. 3025-3031.

Egodawatta, P., Thomas, E. and Goonetilleke, A. 2009, "Understanding the physical processes of pollutant build-up and wash-off on roof surfaces", *Science of The Total Environment,* 407(6), pp. 1834-1841.

Finnish Environment Institute 2014, *Tietoa SYKEn vedenlaadun kaukokartoitustuotteista*. [Online] Available from: http://wwwi4.ymparisto.fi/i4/fin/tuotteet/lisatietoja/tuoteinfo.html [Accessed: 19.8.2014]

Fletcher, T.D., Andrieu, H. and Hamel, P. 2013, "Understanding, management and modelling of urban hydrology and its consequences for receiving waters: A state of the art", *Advances in Water Resources,* 51, pp. 261-279.

Gandin, L.S. 1988, "Complex quality control of meteorological observations", *Monthly Weather Review,* 116(5), pp. 1137-1156.

Gilbert, J.K. and Clausen, J.C. 2006, "Stormwater runoff quality and quantity from asphalt, paver, and crushed stone driveways in Connecticut", *Water research,* 40(4), pp. 826-832.

Grayson, R.B., Finlayson, B.L., Gippel, C.J. and Hart, B.T. 1996, "The potential of field turbidity measurements for the computation of total phosphorus and suspended solids loads", *Journal of environmental management,* 47(3), pp. 257-267.

Haan, C.T., Barfield, B.J. and Hayes, J.C. 1994, *Design hydrology and sedimentology for small catchments,* San Diego, California: Academic Press.

Haiping, Z. and Yamada, K. 1996, "Estimation for urban runoff quality modeling", *Water Science & Technology,* 34(3-4), pp. 49-54.

Herngren, L., Goonetilleke, A. and Ayoko, G.A. 2005, "Understanding heavy metal and suspended solids relationships in urban stormwater using simulated rainfall", *Journal of environmental management,* 76(2), pp. 149-158.

Hossain, I., Imteaz, M., Gato-Trinidad, S. and Shanableh, A. 2010, "Development of a catchment water quality model for continuous simulations of pollutants build-up and wash-off", *International Journal of Environmental, Ecological, Geological and Mining Engineering,* 4(1), 8 pp.

House, M.A., Ellis, J.B., Herricks, E.E., Hvitved-Jacobsen, T., Seager, J., Lijklema, L., Aalderink, H. and Clifforde, I.T. 1993, "Urban drainage - Impacts on receiving water quality", *Water Science and Technology,* 27(12), pp. 117-158.

James, E. and Joyce, M. 2004, “Assessment and management of watershed microbial contaminants”, *Critical Reviews in Environmental Science and Technology*, 34(2), pp. 109-139.

Jones, A.S. 2008, *Estimating total phosphorus and total suspended solids loads from high frequency data*, Master’s Thesis, Logan, Utah: Utah State University, Civil and Environmental Engineering. 127 pp.

Kestävän ympäristön kaupunkilaboratorio 2014*, Kaupunkilaboratorio - Veden kierron hallinta ja ekosysteemipalvelut - Tutkimusmenetelmät*. [Online] Available from: http://geoinforma tics.aalto.fi/kaupunkilaboratorio/tutkimusmenetelmat.html [Accessed: 20.8.2014]

Kostarelos, K., Eekalak, K., Callipo, N., Velasquez, J. and Graves, D. 2011, "Field study of catch basin inserts for the removal of pollutants from urban runoff", *Water Resources Management,* 25(4), pp. 1205-1217.

Krebs, G., Kokkonen, T., Valtanen, M., Koivusalo, H. and Setälä, H. 2013, "A high resolution application of a stormwater management model (SWMM) using genetic parameter optimization", *Urban Water Journal,* 10(6), pp. 394-410.

Krebs, G., Kokkonen, T., Valtanen, M., Setälä, H. and Koivusalo, H. 2014, "Spatial resolution considerations for urban hydrological modelling", *Journal of Hydrology,* 512, pp. 482-497.

Kusari, L. and Ahmedi F. 2013, “The use of turbidity and total suspended solids correlation for the surface water quality monitoring”, *International Journal of Current Engineering and Technology*, 3(4), pp. 1311-1314.

Lahti Aqua 2014*, Viemäröinti*. [Online] Available from: http://www.lahtiaqua.fi/Toiminta /Verkostot/Viem%C3%A4r%C3%B6inti [Accessed: 18.9.2014]

Lee, K., Kim, H., Pak, G., Jang, S., Kim, L., Yoo, C., Yun, Z. and Yoon, J. 2010, "Cost-effectiveness analysis of stormwater best management practices (BMPs) in urban watersheds", *Desalination and Water Treatment,* 19(1-3), pp. 92-96.

Line, D.E., Hall, K.R. and Blackwell, J.D. 2013, "Estimating suspended solids from turbidity in the Robeson Creek, NC watershed", *Journal Of The American Water Resources Association,* 49(6), pp. 1412-1420.

Linjama, J., Puustinen, M. and Koskiaho, J. 2009, "Implementation of automatic sensors for continuous monitoring of runoff quantity and quality in small catchments", *Agricultural And Food Science,* 18(3-4), pp. 417-427.

Liu, A., Goonetilleke, A. and Egodawatta, P. 2012, "Inadequacy of land use and impervious area fraction for determining urban stormwater quality", *Water Resources Management,* 26(8), pp. 2259-2265.

Mallin, M.A., Johnson, V.L. and Ensign, S.H. 2009, “Comparative impacts of stormwater runoff on water quality of an urban, a suburban, and a rural stream”, *Environmental Monitoring and Assessment*, 159(1-4), pp. 475-491.

Métadier, M. and Bertrand-Krajewski, J. 2012, "The use of long-term on-line turbidity measurements for the calculation of urban stormwater pollutant concentrations, loads, pollutographs and intra-event fluxes", *Water Research,* 46(20), pp. 6836-6856.

Miguntanna, N.P., Liu, A., Egodawatta, P. and Goonetilleke, A. 2013, "Characterising nutrients wash-off for effective urban stormwater treatment design", *Journal of environmental management,* 120, pp. 61-67.

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Binger, R.L., Harmel, R.D. and Veith, T.L. 2007, "Model evaluation guidelines for systematic quantification of accuracy in watershed simulations", *Transactions Of The ASABE,* 50(3), pp. 885-900.

Mustonen, S. (eds.) 1986, *Sovellettu hydrologia,* Vesiyhdistys r.y., Mänttä. 503 pp.

Niemczynowicz, J. 1999, "Urban hydrology and water management - present and future challenges", *Urban Water,* 1(1), pp. 1-14.

Obropta, C.C. and Kardos, J.S. 2007, "Review of urban stormwater quality models: deterministic, stochastic, and hybrid approaches", *Journal of the American Water Resources Association,* 43(6), pp. 1508-1523.

Pajari, J. 2014, *Quality assurance of hydrometeorological data in urban areas*, Master’s Thesis, Espoo: Aalto University School of Engineering. 92 pp.

Rossman, L.A. 2010*, Storm Water Management Model user’s manual Version 5.0*. [Online] Available from: http://nepis.epa.gov/Adobe/PDF/P100ERK4.pdf [Accessed: 20.8.2014]

Schueler, T.R. 1994, "The importance of imperviousness", *Watershed Protection Techniques,* 1(3), pp. 100-111.

Settle, S., Goonetilleke, A. and Ayoko, G.A. 2007, "Determination of surrogate indicators for phosphorus and solids in urban stormwater: Application of multivariate data analysis techniques", *Water Air And Soil Pollution,* 182(1-4), pp. 149-161.

Shuster, W.D., Bonta, J., Thurston, H., Warnemuende, E. and Smith, D.R. 2005, "Impacts of impervious surface on watershed hydrology: A review", *Urban Water Journal,* 2(4), pp. 263-275.

Shuster, W.D., Fletcher, T.D. and Deletic, A. 2008, "[Chapter 17:] Stormwater" in *Data requirements for integrated urban water management*, eds. T.D. Fletcher and A. Deletic, London: Taylor & Francis Ltd, pp. 225-241.

Sillanpää, N. 2013, *Effects of suburban development on runoff generation and water quality*, Doctoral Dissertation, Espoo: Aalto University, Department of Civil and Environmental Engineering. 226 pp.

Temprano, J., Arango, O., Cagiao, J., Suarez, J. and Tejero, I. 2006, "Stormwater quality calibration by SWMM: A case study in northern Spain", *Water SA,* 32(1), pp. 55-63.

Tikkanen, H. 2013, *Hydrological modeling of a large urban catchment using a stormwater management model (SWMM)*, Master's Thesis, Espoo: Aalto University School of Engineering. 64 pp.

US EPA 1999, *Combined sewer overflow management fact sheet*, Washington, D.C.: United States Environmental Protection Agency.

US EPA 2014a, *Future climate change*. [Online] Available from: http://www.epa.gov /climatechange/science/future.html#Precipitation [Accessed: 22.8.2014]

US EPA 2014b, *Stormwater homepage*. [Online] Available from: http://water.epa.gov /polwaste/npdes/stormwater/index.cfm [Accessed: 13.10.2014]

US EPA 2014c*, Best Management Practices (BMPs)*. [Online] Available from: http://www.epa.gov/nrmrl/wswrd/wq/stormwater/bmp.html [Accessed: 18.8.2014]

Vaisala Oyj 2012, *Vaisala Weather Transmitter WXT520 User's Guide*, Helsinki: Vaisala Oyj.

Valtanen, M., Sillanpää, N. and Setälä, H. 2009, *Stormwater research in Finland - seeking solutions through strategic research*, in the 6th International Conference on Water Sensitive Urban Design, 5-8 May 2009, Perth, Australia.

Valtanen, M., Sillanpää, N. and Setälä, H. 2014a, "The effects of urbanization on runoff pollutant concentrations, loadings and their seasonal patterns under cold climate", *Water, Air, and Soil Pollution,* 225(6):1977.

Valtanen, M., Sillanpää, N. and Setälä, H. 2014b, "Effects of land use intensity on stormwater runoff and its temporal occurrence in cold climates", *Hydrological Processes,* 28(4), pp. 2639-2650.

Van Metre, P.C. and Mahler, B.J. 2003, "The contribution of particles washed from rooftops to contaminant loading to urban streams", *Chemosphere,* 52(10), pp. 1727-1741.

Vaze, J. and Chiew, F.H.S. 2002, "Experimental study of pollutant accumulation on an urban road surface", *Urban Water,* 4(4), pp. 379-389.

Vezzaro, L. and Mikkelsen, P.S. 2012, "Application of global sensitivity analysis and uncertainty quantification in dynamic modelling of micropollutants in stormwater runoff", *Environmental Modelling & Software,* 27-28, pp. 40-51.

Wang, L., Wei, J., Huang, Y., Wang, G. and Magsood, I. 2011, "Urban nonpoint source pollution buildup and washoff models for simulating storm runoff quality in the Los Angeles County", *Environmental Pollution,* 159(7), pp. 1932-1940.

Warwick, J.J. and Tadepalli, P. 1991, "Efficacy of SWMM application", *Journal Of Water Resources Planning And Management-ASCE,* 117(3), pp. 352-366.

Wilde, F.D. and Gibs, J. 1998, "[Chapter A6, Section 6.7] Turbidity" in *USGS-TWRI Book 9: Handbooks for Water-Resources Investigations,* eds. Wilde, F.D., U.S. Geological Survey.

List of appendices

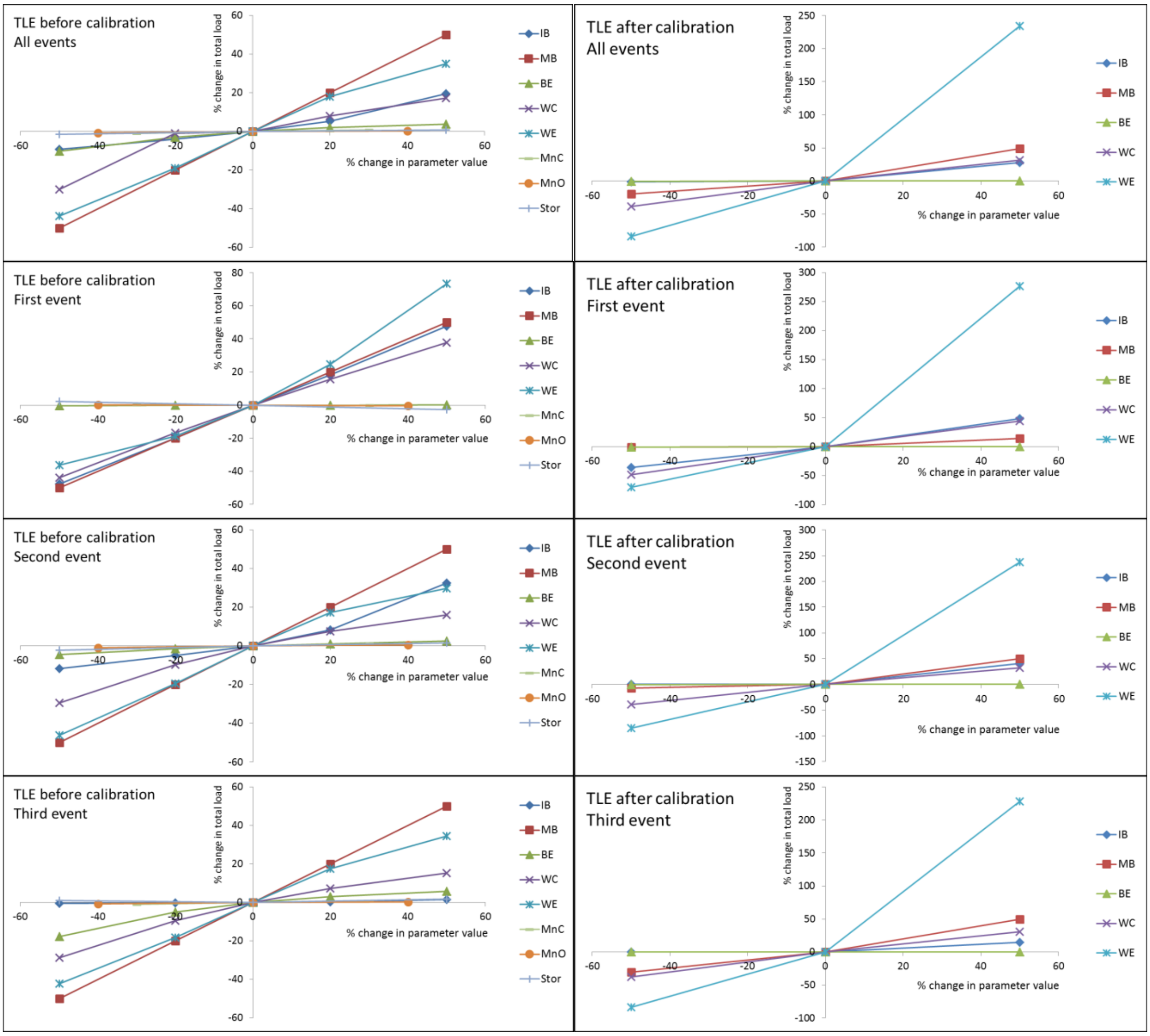
Appendix 1. *Performance statistics in absolute values of conducted parameter perturbations in the first sensitivity analysis. IB and MB are the initial and maximum build-up, respectively, BE is the build-up exponent, WC and WE are the wash-off coefficient and exponent, respectively, MnC and MnO are the Manning's roughness coefficients for conduit and overland flow, respectively, and Stor is the depression storage.*

Parameter	Change	Value	All events			First event			Second event			Third event		
			E [-]	CORR [-]	TLE [%]	E [-]	CORR [-]	TLE [%]	E [-]	CORR [-]	TLE [%]	E [-]	CORR [-]	TLE [%]
IB	-50 %	17.25	0.30	0.68	-26.2	0.74	0.88	1.9	0.52	0.90	-37.6	-0.45	0.62	-13.9
	-20 %	27.6	0.24	0.69	-26.9	0.06	0.88	55.9	0.57	0.89	-33.0	-0.45	0.62	-13.7
	0	34.5	0.10	0.68	-23.9	-1.10	0.88	94.5	0.60	0.89	-29.4	-0.45	0.63	-13.5
	20 %	41.4	-0.09	0.67	-19.8	-2.66	0.88	129.8	0.65	0.90	-23.5	-0.46	0.63	-13.2
	50 %	51.75	-0.53	0.67	-3.0	-6.20	0.88	187.3	0.71	0.90	-6.5	-0.48	0.63	-12.0
MB	-50 %	17.5	0.24	0.68	-62.0	0.75	0.88	-19.5	0.06	0.90	-69.2	0.22	0.63	-54.8
	-20 %	28	0.30	0.68	-39.2	0.07	0.88	28.8	0.44	0.90	-50.8	0.03	0.63	-27.7
	0	35	0.10	0.68	-23.9	-1.10	0.88	61.0	0.60	0.90	-38.5	-0.45	0.63	-9.6
	20 %	42	-0.30	0.68	-8.7	-2.84	0.88	93.2	0.69	0.90	-26.2	-1.22	0.63	8.5
	50 %	52.2	-1.26	0.68	14.1	-6.53	0.88	141.5	0.68	0.90	-7.7	-2.90	0.63	35.6
BE	-50 %	0.45	0.22	0.68	-31.7	-1.07	0.88	60.4	0.55	0.89	-41.3	0.06	0.64	-25.5
	-20 %	0.72	0.14	0.68	-26.2	-1.09	0.88	60.9	0.59	0.89	-39.4	-0.28	0.63	-14.0
	0	0.9	0.10	0.68	-23.9	-1.10	0.88	61.0	0.60	0.89	-38.5	-0.45	0.63	-9.6
	20 %	1.08	0.08	0.68	-22.4	-1.10	0.88	61.0	0.62	0.90	-37.7	-0.58	0.62	-6.7
	50 %	1.35	0.06	0.68	-21.0	-1.10	0.88	61.1	0.64	0.90	-36.9	-0.69	0.62	-4.3
WC	-50 %	0.02	0.43	0.73	-46.8	0.65	0.86	-9.8	0.36	0.92	-56.5	0.36	0.68	-35.4
	-20 %	0.032	0.33	0.70	-31.5	-0.12	0.87	34.1	0.55	0.91	-44.4	-0.01	0.65	-17.9
	0	0.04	0.10	0.68	-23.9	-1.10	0.88	61.0	0.60	0.89	-38.5	-0.45	0.63	-9.6
	20 %	0.048	-0.22	0.65	-17.9	-2.36	0.89	86.3	0.61	0.88	-33.9	-1.03	0.60	-3.0
	50 %	0.06	-0.84	0.62	-10.8	-4.71	0.90	121.9	0.58	0.86	-28.7	-2.08	0.58	4.4
WE	-50 %	0.575	-0.16	0.44	-57.1	-0.28	0.70	2.6	0.30	0.67	-66.9	-1.19	0.35	-47.7
	-20 %	0.92	0.15	0.62	-38.3	-0.39	0.82	30.9	0.55	0.90	-50.5	-0.50	0.54	-26.0
	0	1.15	0.10	0.68	-23.9	-1.10	0.88	61.0	0.60	0.89	-38.5	-0.45	0.63	-9.6
	20 %	1.38	-0.29	0.68	-10.3	-2.70	0.93	100.5	0.44	0.83	-27.8	-0.77	0.66	6.3
	50 %	1.725	-1.98	0.63	2.7	-8.48	0.97	179.0	-0.49	0.68	-20.2	-2.22	0.63	21.6
MnC	-30 %	0.0077-0.0105	0.07	0.67	-24.5	-1.18	0.88	62.1	0.58	0.89	-39.1	-0.47	0.63	-10.3
	0	0.011-0.015	0.10	0.68	-23.9	-1.10	0.88	61.0	0.60	0.89	-38.5	-0.45	0.63	-9.6
	30 %	0.0143-0.0195	0.11	0.67	-23.3	-1.17	0.88	60.9	0.63	0.90	-37.9	-0.44	0.63	-8.8
MnO	-40 %	0.0006-0.4002	0.11	0.68	-24.4	-0.91	0.90	61.5	0.57	0.88	-39.0	-0.46	0.62	-10.1
	0	0.001-0.667	0.10	0.68	-23.9	-1.10	0.88	61.0	0.60	0.89	-38.5	-0.45	0.63	-9.6
	40 %	0.0014-0.9338	0.09	0.67	-23.6	-1.27	0.86	60.2	0.62	0.90	-38.1	-0.45	0.63	-9.2
Stor	-50 %	0.05-2.41	0.03	0.66	-25.1	-1.36	0.90	64.3	0.56	0.87	-39.8	-0.51	0.63	-10.9
	0	0.1-4.82	0.10	0.68	-23.9	-1.10	0.88	61.0	0.60	0.89	-38.5	-0.45	0.63	-9.6
	50 %	0.15-7.23	0.13	0.68	-23.2	-1.17	0.84	56.6	0.62	0.91	-37.5	-0.36	0.63	-8.9

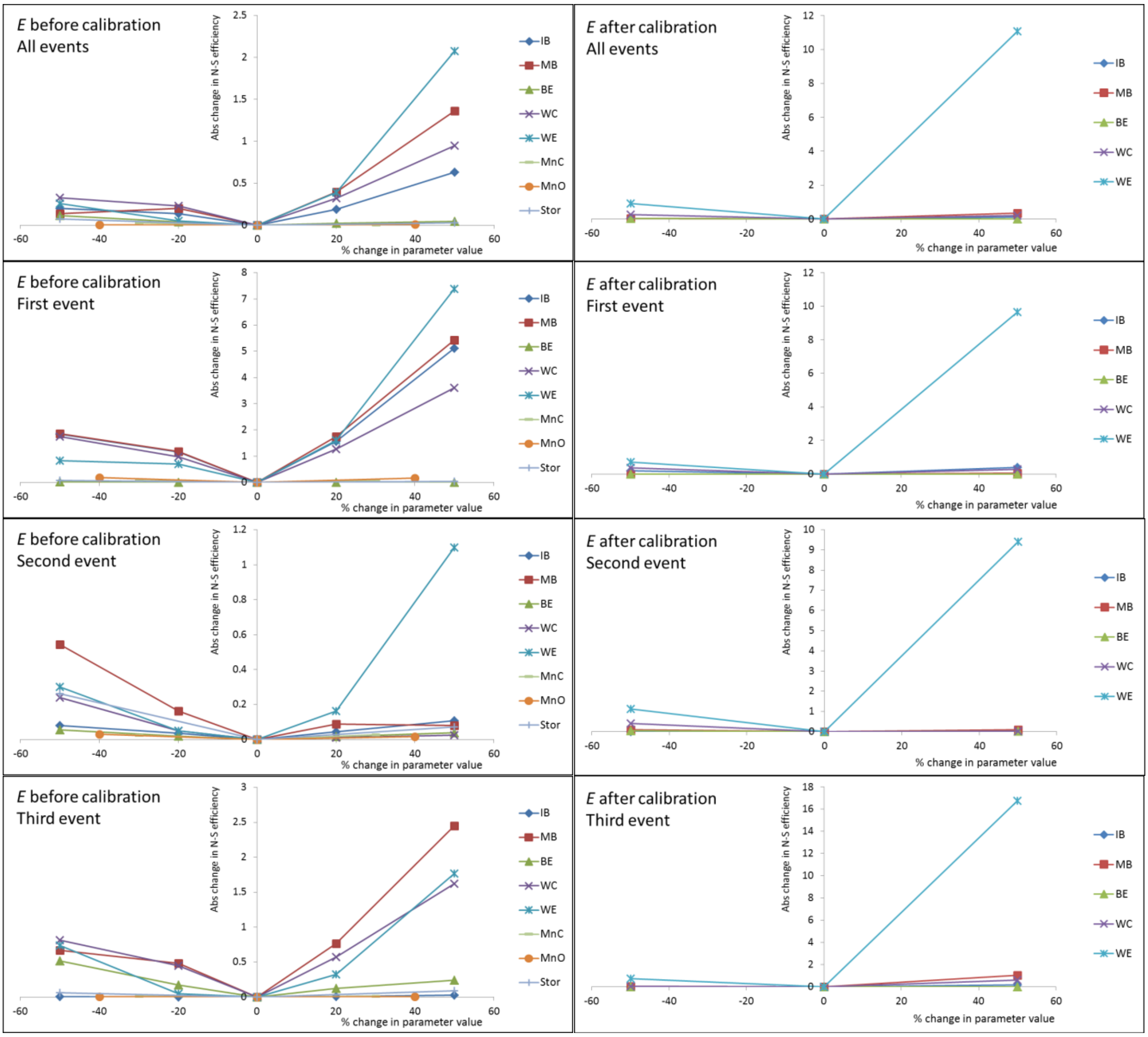
Appendix 2. *Performance statistics in absolute values of conducted parameter perturbations in the second sensitivity analysis. IB and MB are the initial and maximum build-up, respectively, BE is the build-up exponent and WC and WE are the wash-off coefficient and exponent, respectively.*

			All events			First event			Second event			Third event		
Parameter	Change	Value	E [-]	CORR [-]	TLE [%]	E [-]	CORR [-]	TLE [%]	E [-]	CORR [-]	TLE [%]	E [-]	CORR [-]	TLE [%]
	-50 %	131.5	0.67	0.87	-3.1 %	0.41	0.81	-50.6 %	0.78	0.95	-19.1 %	0.54	0.85	25.2 %
IB	0	262.9	0.70	0.88	-2.2 %	0.61	0.82	-23.1 %	0.78	0.95	-19.1 %	0.54	0.85	25.2 %
	50 %	394.4	0.57	0.88	25.3 %	0.23	0.82	14.4 %	0.73	0.93	13.4 %	0.36	0.85	44.1 %
	-50 %	135.2	0.66	0.88	-20.3 %	0.61	0.82	-23.7 %	0.69	0.93	-24.8 %	0.59	0.85	-13.0 %
MB	0	270.4	0.70	0.88	-2.2 %	0.61	0.82	-23.1 %	0.78	0.95	-19.1 %	0.54	0.85	25.2 %
	50 %	405.6	0.36	0.88	45.8 %	0.57	0.82	-12.1 %	0.69	0.95	21.3 %	-0.49	0.85	87.8 %
	-50 %	2.5	0.68	0.87	-2.6 %	0.61	0.82	-23.4 %	0.74	0.93	-19.8 %	0.54	0.85	25.1 %
BE	0	5	0.70	0.88	-2.2 %	0.61	0.82	-23.1 %	0.78	0.95	-19.1 %	0.54	0.85	25.2 %
	50 %	7.5	0.71	0.88	-1.9 %	0.61	0.82	-22.8 %	0.81	0.95	-18.7 %	0.54	0.85	25.2 %
	-50 %	0.0005	0.43	0.87	-40.0 %	0.25	0.81	-59.7 %	0.39	0.93	-50.5 %	0.52	0.85	-22.4 %
WC	0	0.001	0.70	0.88	-2.2 %	0.61	0.82	-23.1 %	0.78	0.95	-19.1 %	0.54	0.85	25.2 %
	50 %	0.0015	0.51	0.88	28.9 %	0.33	0.83	11.1 %	0.80	0.95	6.8 %	-0.05	0.85	63.9 %
	-50 %	0.728	-0.20	0.54	-84.1 %	-0.10	0.72	-76.5 %	-0.33	0.69	-87.5 %	-0.17	0.42	-79.6 %
WE	0	1.456	0.70	0.88	-2.2 %	0.61	0.82	-23.1 %	0.78	0.95	-19.1 %	0.54	0.85	25.2 %
	50 %	2.184	-10.37	0.83	226.8 %	-9.03	0.92	189.7 %	-8.62	0.86	173.0 %	-16.22	0.81	311.2 %

Appendix 3. *The effects on total load error of perturbations in parameter values in the two sensitivity analyses. The more horizontal a line is, the less sensitive the model is to perturbations in the parameter at issue. IB and MB are the initial and maximum build-up, respectively, BE is the build-up exponent, WC and WE are the wash-off coefficient and exponent, respectively, MnC and MnO are the Manning's roughness coefficients for conduit and overland flow, respectively, and Stor is the depression storage.*



Appendix 4. *The effects on Nash-Sutcliffe efficiency* E *of perturbations in parameter values in the two sensitivity analyses. The more horizontal a line is the less sensitive the model is to perturbations in the parameter at issue. IB and MB are the initial and maximum build-up, respectively, BE is the build-up exponent, WC and WE are the wash-off coefficient and exponent, respectively, MnC and MnO are the Manning's roughness coefficients for conduit and overland flow, respectively, and Stor is the depression storage.*



Appendix 5. *Performance statistics of conducted optimizations in the calibration procedure. The optimizations marked with the same colour ranged within the same set of parameter boundaries (OR1-O7 in red, O8-O10 in yellow and O11-O15 in blue).*

Optimization	Optimized for events	All events						1st event						2nd event						3rd event					
		E		CORR		TLE		E		CORR		TLE		E		CORR		TLE		E		CORR		TLE	
OR1	1, 2, 3	0.713	0.714	0.867	0.865	-9.2%	-11.8%	0.667	0.692	0.848	0.847	-13.2%	-18.8%	0.735	0.735	0.917	0.920	-23.1%	-26.8%	0.646	0.640	0.867	0.864	12.0%	11.4%
O1	1	0.114	0.061	0.711	0.691	-64.3%	-67.2%	0.866	0.865	0.941	0.953	-11.6%	-11.6%	-0.029	-0.036	0.783	0.785	-63.4%	-63.5%	-0.104	-0.280	0.814	0.848	-70.3%	-77.7%
O2	2	0.006	-0.007	0.700	0.698	9.8%	9.9%	0.382	0.380	0.811	0.810	-55.2%	-55.3%	0.832	0.832	0.895	0.929	-14.3%	-14.2%	-1.963	-2.008	0.686	0.684	51.8%	52.0%
O3	3	0.515	0.517	0.815	0.815	-12.7%	-13.1%	-0.072	-0.063	0.911	0.910	42.1%	41.7%	0.577	0.578	0.870	0.871	-17.4%	-17.7%	0.602	0.602	0.838	0.838	-10.5%	-11.0%
O4	1, 2	0.073	0.106	0.726	0.733	9.1%	9.6%	0.658	0.658	0.828	0.828	-24.8%	-24.6%	0.828	0.828	0.929	0.929	-14.0%	-13.5%	-1.845	-1.726	0.681	0.689	46.8%	47.4%
O5	2, 3	0.586	0.567	0.820	0.815	-16.4%	-15.8%	0.247	0.101	0.887	0.888	25.1%	31.1%	0.650	0.654	0.896	0.895	-24.4%	-23.3%	0.560	0.553	0.805	0.801	-8.0%	-8.8%
O6	1, 3	0.551	0.557	0.829	0.829	-29.8%	-27.8%	0.789	0.791	0.911	0.911	-12.4%	-13.6%	0.448	0.457	0.870	0.870	-39.5%	-38.6%	-0.296	0.573	0.827	0.826	-16.7%	-12.8%
O7	1, 2, 3	0.613	0.613	0.825	0.826	-18.8%	-19.6%	0.581	0.596	0.880	0.882	5.1%	4.3%	0.632	0.624	0.902	0.901	-30.0%	-30.6%	0.532	0.541	0.798	0.800	-4.1%	-5.2%
O8	1	-3.359	-	0.361	-	-45.2%	-	0.941	-	0.941	-	-16.0%	-	-1.007	-	0.484	-	-59.9%	-	-10.763	-	0.355	-	-25.5%	-
O9	1, 2	0.621	-	0.857	-	7.6%	-	0.645	-	0.828	-	-20.7%	-	0.864	-	0.955	-	-10.1%	-	0.071	-	0.800	-	36.8%	-
O10	1, 2, 3	0.701	-	0.878	-	-2.2%	-	0.615	-	0.820	-	-23.1%	-	0.781	-	0.946	-	-19.1%	-	0.536	-	0.852	-	25.2%	-
O11	1	-15.017	-	0.260	-	-24.2%	-	0.943	-	0.982	-	-17.3%	-	-2.165	-	0.451	-	-50.0%	-	-50.625	-	0.306	-	14.0%	-
O12	1, 2	0.614	-	0.856	-	7.3%	-	0.626	-	0.814	-	-23.5%	-	0.879	-	0.958	-	-10.2%	-	0.025	-	0.791	-	36.3%	-
O13	1, 2, 3	0.715	-	0.882	-	-3.4%	-	0.607	-	0.815	-	-25.9%	-	0.810	-	0.953	-	-19.9%	-	0.538	-	0.846	-	23.4%	-
O14	1, 2, 3	0.717	0.717	0.885	0.885	-2.5%	-3.0%	0.617	0.616	0.822	0.822	-21.6%	-22.0%	0.809	0.807	0.953	0.953	-19.4%	-19.8%	0.541	0.544	0.853	0.852	24.7%	24.1%
O15	1, 2, 3	0.656	0.656	0.861	0.860	-4.4%	-4.6%	0.185	0.184	0.783	0.782	-62.9%	-63.0%	0.835	0.835	0.958	0.958	-19.1%	-19.3%	0.477	0.478	0.835	0.834	23.0%	22.8%