UNIVERSITY OF SÃO PAULO

SÃO CARLOS SCHOOL OF ENGINEERING

MARIA CLARA FAVA

IMPROVING FLOOD FORECASTING USING REAL-TIME DATA TO UPDATE URBAN MODELS IN POORLY GAUGED AREAS

CORRECTED VERSION (VERSÃO CORRIGIDA)

SÃO CARLOS

2019

MARIA CLARA FAVA

IMPROVING FLOOD FORECASTING USING REAL-TIME DATA TO UPDATE URBAN MODELS IN POORLY GAUGED AREAS

Doctoral thesis presented at São Carlos School of Engineering, University of São Paulo, in partial fulfilment of the requirements for obtaining the Degree of Doctor in Science: Hydraulics and Sanitary Engineering.

Advisor: Prof. Dr. Eduardo Mario Mendiondo

CORRECTED VERSION (VERSÃO CORRIGIDA)

SÃO CARLOS

2019

Ficha catalográfica elaborada pela Biblioteca Prof. Dr. Sérgio Rodrigues Fontes da EESC/USP com os dados inseridos pelo(a) autor(a).

Fava, Maria Clara
F272i Improving flood forecasting using real-time data to update urban models in poorly gauged areas / Maria Clara Fava; orientador Eduardo Mario Mendiondo. São Carlos, 2019.

Tese (Doutorado) - Programa de Pós-Graduação em Engenharia Hidráulica e Saneamento e Área de Concentração em Hidráulica e Saneamento -- Escola de Engenharia de São Carlos da Universidade de São Paulo, 2019.

1. Flood modelling. 2. Data assimilation. 3. SWMM. 4. Citizen science. 5. Urban floods. I. Título.

Eduardo Graziosi Silva - CRB - 8/8907

FOLHA DE JULGAMENTO

Candidata: Bacharela **MARIA CLARA FAVA.**

Título da tese: "Assimilação de dados na previsão de enchentes em tempo real em áreas urbanas com dados escassos".

Data da defesa: 30/08/2019.

Comissão Julgadora: **Resultado:**

Prof. Dr. **Eduardo Mario Mendiondo (Orientador)** Aprovado
(Escola de Engenharia de São Carlos/EESC)

Drª. **Luz Adriana Cuartas Pineda** Aprovado
(Centro Nacional de Monitoramento e Alertas de Desastres Naturais/CEMADEN)

Prof. Dr. **Massato Kobiyama** Aprovado
(Universidade Federal do Rio Grande do Sul/UFRGS)

Prof. Associado **Jó Ueyama** APROVADO
(Instituto de Ciências Matemáticas e de Computação/ICMC-USP)

Prof. Dr. **Dmitri Petrovich Solomatine** APROVADO
(Delft Institute for Water Education/UNESCO-IHE)

Coordenador do Programa de Pós-Graduação em Engenharia Hidráulica e Saneamento:
Prof. Dr. **Eduardo Mario Mendiondo**

Presidente da Comissão de Pós-Graduação:
Prof. Titular **Murilo Araujo Romero**

Dedication

To my parents and my brother
for everything they represent to me.

Acknowledgements

Over these four years of PhD research, I have been contemplated with countless wise people crossing my journey, colleagues who are readily willing to help me, discuss problems, find solutions and even to share endless questioning.

I want to thank my advisor Professor Eduardo Mario Mendiondo, for accepting me on this journey, for all the enthusiasm and the patience over these years. Thank you so much for all the guidelines, for the countless opportunities given, for the inspiration for always thinking beyond the traditional and for the partnership. I believe it will be extended throughout my research journey, even though the formal period is finished with the closure of this thesis.

I greatly thank Professor Dimitri Solomatine, my advisor during my period at IHE-Delft. Thanks for the honour of welcoming me into his research group, for all the teaching and encouragement at crucial times, for your outstanding quality of conducting research, and for bringing me a bit more closely the hydroinformatics side of the force!

To my great friend Narumi Abe, who always mentored and worked with me in great joy, since my undergraduate scientific initiation until the closing of the text of this dissertation. My eternal thanks for all the laughs, for always listening to my endless complaints and dilemmas of life, and for all the help you gave me when I needed it most.

I also thank Dr. Maurizio Mazzoleni, my mentor during my internship in Delft, and a great inspiration as a young researcher. Thank you for all the help, the brilliant ideas, the patience and understanding as you teach.

Thanks to João Victor Cal for helping me in my internship at CEMADEN and to all the professionals there who added a lot in my training. I thank Professor Bruno Kimura from UNIFESP for accepting the partnership, making this internship possible.

Thank you very much, mother and father (Maria Evaneide Fava and Sidnei Fava), for all the love, encouragement and attachment kept despite the 1000 km away. Thank you for the security and stability that allowed me to choose without hesitation the academic journey. I know that I will always have a safe place. Thanks also to my dear brother Aldo Felipe Fava for the infinite friendship, love and support. Thank you to all my family and I hope you understand all

my moments of absence. Special thanks to my cousin Augusto for always encouraging me to pursue my research career.

Many thanks to my partner Julio for all the love, patience and support in my dissertation and life outbreaks. Thank you for showing me that besides successes and failures, there are much more important things, mainly the smiles in the day-to-day. Thank you for the partnership and levity you brought to my life.

My special thanks to all NIBH / WADI colleagues and former colleagues who were an essential part of this journey, thank you for helping to clarify many problems and accomplish many tasks, for friendship and fun in the daily life at the lab, and for making my PhD journey much better and relaxed. To my friends at IHE-Delft for welcoming me so well, to my neighbours from "Little Colombia" for all the laughs, doubts and shared tips. Special thanks to my dear friend Thaine who gave me support in the difficult times and for the partnership and laughter in the good and calm times.

Thanks to all my friends from São Carlos for the wonderful moments of relaxation and learning about life. Many thanks to my long-time friends from Nioaque and Campo Grande who I always miss.

Special thanks to my four-legged children, Tareco and Lola, for showing that life is sweet, for looking me always asking for love and affection, but never inquiring about finishing the thesis or the papers.

Special thanks to the *Ilex paraguariensis*, my dear yerba mate that makes up my daily tereré, which was essential to my performance on days of programming, reading and writing, and always remembering me where I come from and who I am, *viva el alma Guarani*!

Thanks to all the staff and professors of the Department of Hydraulic Engineering and Sanitation. Especially to Sá and Priscilla for all the attention and help provided.

Thanks to the University of São Paulo for the physical infrastructure provided, to the National Council for Scientific and Technological Development (CNPq) for the PhD scholarship granted, the Coordination of Improvement of Higher Education Personnel (CAPES) for the PhD “Sandwich” Grant and the São Paulo Research Foundation (FAPESP) for funding the experiments performed.

Agradecimentos

No decorrer destes quatro anos de doutorado eu fui agraciada com inúmeras e sábias pessoas que cruzaram a minha jornada, e que prontamente se dispuseram a me ajudar, discutir problemas, encontrar soluções e até mesmo partilhar infinitas dúvidas.

Gostaria de agradecer ao meu orientador Professor Eduardo Mario Mendiondo, por me aceitar nessa jornada, por todo o entusiasmo e paciência durante todos esses anos. Muito obrigada por todos os ensinamentos, pelas inúmeras oportunidades dadas, pela inspiração a pensar sempre um pouco além do tradicional e pela parceria construída. Acredito que a mesma se estenderá por toda a minha jornada na pesquisa, ainda que o período formal seja selado com o fechamento desta tese.

Agradeço enormemente ao Professor Dimitri Solomatine, meu orientador durante o meu período no *IHE Delft Institute for Water Education*, pela honra de ter me recebido em seu grupo de pesquisa, por todos os ensinamentos e encorajamento em momentos cruciais, pela qualidade ímpar em conduzir pesquisa e por me trazer um pouquinho mais para o lado *hydroinformatics* da força!

Ao meu grande amigo Narumi Abe, que sempre me orientou e trabalhou com muita alegria junto comigo, desde a minha iniciação científica na graduação até o fechamento do texto desta tese. Meu eterno obrigada por todas as risadas, por sempre ouvir minhas intermináveis reclamações e dilemas da vida e por toda a ajuda prestada nos momentos em que mais precisei.

Agradeço também ao Maurizio Mazzoleni meu mentor durante meu estágio em Delft e grande inspiração como jovem pesquisador. Obrigada por toda a ajuda, pelas brilhantes ideias, pela paciência e compreensão ao ensinar.

Obrigada ao João Victor Cal por me ajudar no meu período de estágio no CEMADEN e a todos os profissionais de lá que agregaram muito em minha formação. Agradeço ao Professor Bruno Kimura por aceitar a parceria e possibilitar a realização desse estágio.

Muito obrigada mãe e pai (Maria Evaneide e Sidnei Fava), por todo o amor, incentivo e amizade mantidos apesar dos 1000 km de distância. Obrigada pela segurança e estabilidade que me proporcionaram escolher sem hesitar os caminhos que me levaram a jornada acadêmica, pois sei que sempre terei um porto seguro. Obrigada também ao meu querido irmão Aldo Felipe,

pela amizade, amor e apoio infinitos. Obrigada à toda a minha família e espero que entendam todos os meus momentos de ausência. Um agradecimento especial ao meu primo Augusto por sempre ter me incentivado a seguir na carreira de pesquisa.

Muito obrigada ao meu companheiro Júlio, por todo o amor, paciência e apoio nos meus surtos com a tese e com a vida. Obrigada por me mostrar que além de sucessos e fracassos existem coisas muito mais importantes, que são os sorrisos no dia a dia. Obrigada pela parceria e leveza que trouxe para minha vida.

Meu agradecimento especial a todos os colegas e ex-colegas do NIBH/WADI que foram parte essencial dessa jornada. Obrigada por ajudarem na elucidação de muitos problemas e na realização de muitas tarefas, pela amizade e diversão no cotidiano do laboratório e por tornarem a vida durante o doutorado muito mais divertida. Aos meus amigos do IHE-Delft por me receberem tão bem, aos meus vizinhos da "Little Colômbia" por todas as risadas, dúvidas e dicas partilhadas. Um agradecimento especial a minha querida amiga Thaine pelo suporte nos momentos difíceis e pela parceria e risadas nos momentos tranquilos.

Obrigada a todos os meus amigos de São Carlos, pelos maravilhosos momentos de relaxamento e aprendizados sobre a vida. Muito obrigada aos meus amigos de longa data de Nioaque e Campo Grande de quem sempre tenho saudades.

Agradeço muito aos meus filhos de quatro patas, Tareco e Lola, por mostrarem que o mundo é doce, pelos olhares sempre pedindo amor e carinho, mas nunca inquirindo sobre o término da tese ou artigos. Um agradecimento especial a *Ilex paraguariensis*, minha querida erva-mate que compõe meu tereré diário, e que foi essencial para o meu rendimento nos dias e noites de revisão dos códigos, leitura e escrita, além de sempre me lembrar de onde venho e quem sou. *Viva el alma Guarani*!

Obrigada a todos os funcionários e professores do Departamento de Hidráulica e Saneamento. Em especial a Sá e Priscila por toda atenção e ajuda prestados.

Obrigada a Universidade de São Paulo pela infraestrutura física cedida, ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa de doutorado concedida, a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de doutorado sanduíche concedida e a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo financiamento dos experimentos realizados.

Abstract

FAVA, M. C. (2019). **Improving Flood Forecasting using Real-time Data to Update Urban Models in Poorly Gauged Areas**. Doctoral thesis, São Carlos School of Engineering, University of São Paulo, São Carlos.

Flood forecasting techniques have been widely studied as a tool to mitigate damage from extreme events. However, their nature in urban areas developed without properly drainage planning, coupled with the scarcity of hydrological monitoring data, becomes a significant challenge for real-time flood forecasting. This doctoral thesis proposes deterministic methods of data assimilation for real-time hydrological forecasting. The methodology is developed using the semi-distributed hydrodynamic Storm Water Management Model (SWMM). It also aims to evaluate the impact of using traditional monitoring data together with citizen science data for model updating. The **first** and the **second chapter** present the general introduction and methodology of the thesis. The **third chapter** presents an automatic calibration tool - SWMM calibrator - developed to allow the adjustment of SWMM model parameters with data from multiple sources and to use observed level data as a priori knowledge. The **fourth chapter** deals with the use of citizen science data for urban model updating through a real-time estimator. The **fifth chapter** presents a data assimilation method by updating hydrological model inputs based on water level observations and evaluates the effectiveness of the technique in a distributed manner in the catchment. The proposed methodologies are validated in a case study at the Monjolinho urban catchment. The **sixth chapter** discusses general conclusions and recommendations. In conclusion, SWMM calibrator tool provides flexibility in calibration, allowing shaping the process according to the real-world limitations, and achieved satisfactory calibration results at Monjolinho catchment. The deterministic data assimilation methods proposed in the fourth and fifth chapters have shown effective results in a significant improvement in simulations accuracy.

Keywords: flood modelling; citizen science data; short-term forecasting; SWMM.

Resumo

FAVA, M. C. (2019). **Assimilação de Dados na Previsão de Enchentes em Tempo Real em Áreas Urbanas com Dados Escassos**. Tese, Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos.

A previsão de enchentes para a mitigação dos danos causados por eventos extremos vem sendo amplamente estudada. No entanto, sua natureza em áreas urbanas desenvolvidas sem planejamento adequado de drenagem, associada a escassez de dados de monitoramento hidrológico apresentam um grande desafio para previsão de enchentes em tempo real. Esta tese de doutorado propõe novos métodos determinísticos de assimilação de dados em tempo real na previsão hidrológica, através do modelo hidrodinâmico semi-distribuído *Storm Water Management Model* (SWMM). Estes métodos visam contornar as limitações na previsão de enchentes em tempo real, em curto prazo e em bacias urbanas com dados escassos. Foi também avaliado o impacto do uso de dados de monitoramento tradicionais aliados a dados de ciência cidadã na atualização das simulações do modelo. O **primeiro capítulo** e o **segundo capítulo** trazem a introdução e metodologia gerais da tese. O **terceiro capítulo** apresenta uma ferramenta de calibração automática – *SWMM calibrator* – desenvolvida para permitir o ajuste de parâmetros do modelo SWMM com dados provenientes de múltiplos locais de monitoramento e utilizando dados observados de nível como conhecimento *a priori*. O **quarto capítulo** aborda a utilização de dados de ciência cidadã na atualização do modelo através de um estimador em tempo real. O **quinto capítulo** apresenta um método de assimilação de dados através da correção das entradas do modelo hidrológico baseado em observações de nível, e avalia a eficácia do método de forma distribuída na bacia. As metodologias propostas foram aplicadas para um estudo de caso na bacia urbana do Monjolinho. O **sexto capítulo** apresenta as conclusões e recomendações gerais. Em conclusão, a ferramenta *SWMM calibrator* disponibiliza flexibilidade na calibração, permitindo moldar o processo de acordo com as limitações de problemas reais, e alcançou resultados satisfatórios na calibração da bacia do Monjolinho. Os métodos determinísticos de assimilação de dados propostos no quarto e quinto capítulo mostraram resultados eficazes na redução do erro das simulações de nível do modelo, bem como mostraram resultados satisfatórios ao assimilar dados com uma distribuição temporal maior que o passo de tempo do modelo.

Palavras-chave: modelagem hidrológica; ciência cidadã; previsão de enchentes; SWMM; assimilação de dados.

List of figures

List of tables

List of acronyms

1D – One Dimensional

AMS – American Meteorological Society

AVNSE – Average Nash-Sutcliffe efficiency

CEMADEN – Brazilian National Centre of Monitoring and Early Warning of Natural Disasters

CN – Curve Number

CO – Citizen Observatories

CS – Citizen Science

DA – Data Assimilation

DEAP – Distributed Evolutionary Algorithms in Python

DEM – Digital Elevation Model

GA – Genetic Algorithm

GIS – Geographic Information System

HAMPB – Hydrological Alert Model with Participatory Basis

HI – Hazard Index

IAHS – International Association of Hydrological Sciences

IDW – Inverse Distance Weighted

KF – Kalman Filter

LID – Low Impact Development

MCM – Monte Carlo Method

NRCS – National Resources Conservation Service

NSE – The Nash-Sutcliffe Efficiency

OAT – One at a time

OF – Objective Function

PoC - Proof of Concept

PPA – Permanent Preservation Area

PUB – Predictions in Ungauged Basins

RDII – Rainfall Dependent Inflow and Infiltration

REGNSE – Regional Nash-Sutcliffe efficiency

RMSD – Root Mean Square Deviation

SA – Sensitivity Analysis

SCS – Soil Conservation Service

SDI – Spatial Data Infrastructure

SHOWS – Socio Hydrological Observatory for Water Security

SPATNSE – Spatial Nash-Sutcliffe efficiency

SWMM – Storm Water Management Model

T_C – Time of Concentration

UN – United Nations

UNISDR – United Nations Office for Disaster Reduction

US EPA – United States Environmental Protection Agency

VGI – Volunteered Geographic Information

WMO – World Meteorological Organization

Table of contents

1 General introduction

Historically, cities have been developed along rivers and coastal regions to benefit from their location and resources (Timmerman and White, 1997). The combination of human occupation of natural floodplains with changing of its landscape resulted in recurrent urban floods problems worldwide. Many studies have shown the correlation between growing urbanisation and the occurrence of floods (Keller & DeVecchio, 2016; Prosdocimi *et al*., 2015; Nardi *et al*., 2018). Also, the association of causal factors make urban floods more pronounced and is often common in developing cities to have all these problems together. Soil sealing and riparian forest removal results in a greater and faster response in terms of runoff volume, the microclimate of urban centres cause convective and localised rainfall, and the high population density implies in a large number of people affected during flood events, highlighting the importance of disaster prevention and vulnerability reduction in risk areas. Barreto Cordero (2012) described common factors regarding storm-water drainage in developing cities: the absence of drainage system in peripheral areas, inadequate management of land use, lack of drainage system maintenance in poor areas, overflows from combined sewer systems, human occupation of floodplains, among others.

Over the last decades, more attention has been paid to natural disasters and their potential impact on the environment, mainly due to the increase of frequency that these extreme events occur (Smit *et al*., 1999; Turner *et al*., 2003; Gordy, 2016). According to the United Nations Office for Disaster Reduction (UNISDR), Brazil is the only country in the Americas that is among the ten nations with the highest numbers of people affected by disasters between 1995 and 2015 (CRED, 2015). The research developed by Mendiondo *et al.* (2017) shows that Brazil has 40,000 risk areas (landslide and flood risk) mapped for the 958 municipalities monitored by the Brazilian National Centre of Monitoring and Early Warning of Natural Disasters (CEMADEN). The study also highlights the lack of environmental monitoring and actions to mitigate floods and landslides risks over the country.

Early warning systems for natural disaster reduction are one of the main approaches applied to mitigate the damage caused by extreme events. These systems are responsible for issuing warnings of disaster risk levels, thus providing a time horizon for action by the community under risk. In the case of flood alert systems, they are mostly composed by hydraulic-hydrological or hydrodynamic models and forcing by weather data (such as observed rainfall, remote sensing data and numerical weather forecasts) and finally a framework for

issuing the alerts. These systems generally aim to calculate the water level or flow in floodplains, the extent of flooding, predict the occurrence or likelihood of short-term flooding. It depends mainly on the flood forecasting model applied, the data available and the characteristics inherent the area.

Many studies have shown that flood forecasting is not a trivial task, and it has been improved over the years (Cools and O'Brien, 2016). In floods, an overflowing river is usually the result of high-intensity rainfall over a long period in larger areas. However, flash-floods forecasting is even more complex requiring prompt responses (Mendiondo, 2005; Bodoque *et al*., 2016), and acquisition of geospatial data in real-time (Hapuarachchi *et al.*, 2011) because it is generally caused by high-intensity rainfall in small areas. Thus, the early warning flood system solutions for those cases bring forward conflicting problems regarding the data required and the data availability in developing countries (Basha & Rus, 2007).

Considering the evolution of communication, there is a possibility of interaction and acquiring information by groups of people who are passing by or living in risk areas. Data can be collected in alternative ways to aid in a crucial problem in flood forecasting models that is the availability of water level and rainfall gauges in all points of interest. Citizen Science has a very broad concept that has been applied in several areas; it usually has the main meaning is the involvement of citizens in the collection of data and knowledge for scientific research. The concept embraces both the collection of scientific data with the active participation of volunteers on research hypotheses and issues, as well as in less active activities such as "human sensors" (Roy *et al.*, 2012). Goodchild (2007) coined the term VGI - Volunteered Geographic Information as digital geospatial data generated by common citizens. Leyh *et al*. (2017) emphasise that VGI is one of the practices of Citizen Science, where information provided by volunteers must necessarily be georeferenced. Leyh *et al*. (2016) raised an important question about how collaborative data can be distributed to be used as raw data or as input for hydrological models. Supposedly, data used in disaster prevention should be updated as soon as possible and be easily available; the internet is considered the most efficient solution to make these data available quickly (Rathore *et al*., 2016).

Crowdsourcing platforms such as Ushahidi have been widely used for collecting citizen data in flood-affected areas (Heinzelman & Waters, 2010; Zook *et al*. 2010; Victorino and Estuar, 2014; Mazzoleni *et al*., 2017; Pánek *et al*., 2016). Data collection by volunteers in an organised way to carry out specific tasks and through different devices, in order to create and

publish information of interest, are called Citizen Observatories (CO) (Miorandi *et al*., 2013; Degrossi *et al*., 2014). The WeSenseIt project also includes in the definition of CO the idea of involving communities, emergency operators and policymakers in an organised platform for discussion, monitoring and intervention on situations, places and events (Mazumdar *et al*., 2016). Many citizen observatories platforms have been developed to unify data provided by volunteers, improving the usability for the population and stakeholders. Degrossi *et al*. (2014) performed an experimental evaluation in the urban basin of São Carlos, Brazil using a crowdsourcing-based approach for obtaining useful volunteer information for the context of flood risk management called Flood Citizen Observatory. Wan *et al*. (2014) described a public cloud-based flood cyberinfrastructure (CyberFlood) intending to present a new way to integrate global flood databases. The project aims to provide global data related to floods and also collect crowdsourced data from the internet. Wehn *et al*. (2015) carried out a study comparing the potential use of COs in Italy, the Netherlands and the United Kingdom. The study aims to identify the available mechanisms of citizen participation during the different stages of the disaster cycle (prevention, preparation, response and recovery) and concluded that engagement of citizen to collect data is related to each country's standards. Wilkinson *et al*. (2015) developed a pilot Local Environmental Virtual Observatory Flooding infrastructure with a case study in three basins of the United Kingdom, and they showed to be a promising tool to facilitate bottom-up catchment management approaches.

A review of recent initiatives from the European Union about methods and techniques to prevent and manage floods carried out by Cortes *et al.* (2013) shows many techniques, including warning system and decision support system methodologies. Furthermore, the authors mention VGI as a complementary source of information and point out the challenge of integrating these data with the current Spatial Data Infrastructure (SDI). Schade *et al.* (2013) proposed a methodology for integrating VGI data into online platforms for flood management. However, the authors use VGI only for mining data from social media such as Flickr, using geotag extraction to compose the volunteered data. Smith *et al.* (2015) employed crowdsourced data from Tweeter to verify the spatial-temporal distribution of rainfall and flooded sites demonstrating that the use of Geotags should be done carefully and is subject to errors. Andrade *et al.* (2017) also used data reported by volunteers by Tweeter to asset their relationship to authoritative data concluding that they are not synchronised but are associated with a lag-time. Data provided by volunteers and crowdsourced data are subject to many uncertainties, and there is a great concern for developing methodologies to determine the reliability of these data (Fritz

et al., 2012). Meanwhile, some studies already present good results using volunteered data. Estes *et al.* (2016) obtained data from land cover with 91% accuracy using an open-sourced platform presented in a study called DIYlandcover.

In 2003, the International Association of Hydrological Sciences (IAHS) launched an initiative called Predictions in Ungauged Basins (PUB). This scientific program focused on estimating uncertainties in predictions at poorly monitored areas and their subsequent reduction as the main theme (Sivapalan *et al*., 2003). The PUB scientific decade finished in 2012, after that, the IAHS introduced a new scientific decade called "Panta Rhei-Everything Flows" that relates hydrology development with changes in society (Montanari *et al*., 2013). Considering the challenges proposed by these initiatives, studies concentrating on short-term forecasting methods, data entry in locations with scarce information, vulnerability and greater involvement of the population in preventive measures should be carried out in more depth. Humans being used as sensors and internet technologies can contribute to field data collection, help in management processes to reduce and prevent disasters and encourage community engagement to reduce and respond to situations of environmental risk (Fraternali *et al.*, 2012; Granell *et al.*, 2016). From the exposed, the main goal of this doctoral thesis is to develop quantitative methodological approaches to allow incorporating multiple data sources, as traditional data from sensors and voluntary-based data, in urban flood forecasting models.

This proposal is designed under the umbrella of the Socio Hydrological Observatory for Water Security (SHOWS) (Souza, 2019), which is a conceptual framework based on three knowledge areas: socio hydrology, citizen observatories and water security, that proposes the use of citizen science data to build future risk scenarios using water security variables. This research is part of a cooperative network of thematic research at USP / Nap / CEPED-EESC: "Collaborative actions on reducing vulnerability to hydrological disasters using resilient technologies in urban drainage", FAPESP, 2008 / 58161-1 "Assessment of impacts and vulnerability to climate change in Brazil strategies and options for adaptation", FINEP 01.10.0701.00" MAPLU-Management Urban Rainwater ", CNPq-CEMADEN-USP" Development of forecasting runoff and flooding for natural disaster prevention system" and Public Notice nº 24/2014 Pro-Alert "CEPED Alert - USP".

1.1 Text organisation

All the particular studies presented in each chapter are related to flood management of the Monjolinho catchment, focusing on different methods to increase water level simulations accuracy aiming to overcome the lack of hydrological data and the challenger nature of urban floods forecasting. Six chapters compose this doctoral thesis, the **first chapter** brings the general introduction and information about thesis outline and objectives. The **second chapter** describes the general methodology employed in the experiments described in the next three chapters. The **third chapter** presents the calibration tool developed (SWMM calibrator), details the data used and the results obtained in the calibration and validation of the hydraulic-hydrological model of the Monjolinho catchment. Flash floods and short-term forecasting often requires real-time updates to reach satisfactory results, even having a calibrated model. Thereby, the fourth and fifth chapter present real-time updating methods applied in the Monjolinho catchment model to improve the simulation accuracy: the **fourth chapter** presents a real-time estimator based on water level observations to update the water level simulations, and the **fifth chapter** presents a method that uses the difference between observed and simulated water levels to update the model inputs (precipitation). Finally, the **sixth chapter** discusses general conclusions and recommendations.

1.2 Research hypothesis

Assimilating real-time data in urban hydrodynamic models can improve flood forecasting hence reducing the flood risk susceptibility of poorly monitored catchments.

Citizen science observations can be used as an alternative to fill the lack of data, guaranteeing improvement of model predictions when associated with traditional monitoring data.

1.3 Objective

1.3.1 General objective

Develop data assimilation methodologies to update semi-distributed urban models using real-time data in poorly monitored catchments aiming to reduce flood risk.

1.3.2 Specific objectives

• Develop a methodology towards hydrological semi-distributed model calibration using water level data in poorly gauged catchments;

• Develop a methodology for integrating traditional monitoring data and citizen science data to overcome the lack of data in poorly monitored catchments;

• To assess the effect of the number and location of the different locations of receiving real-time data on model accuracy.

1.4 Thesis synthesis

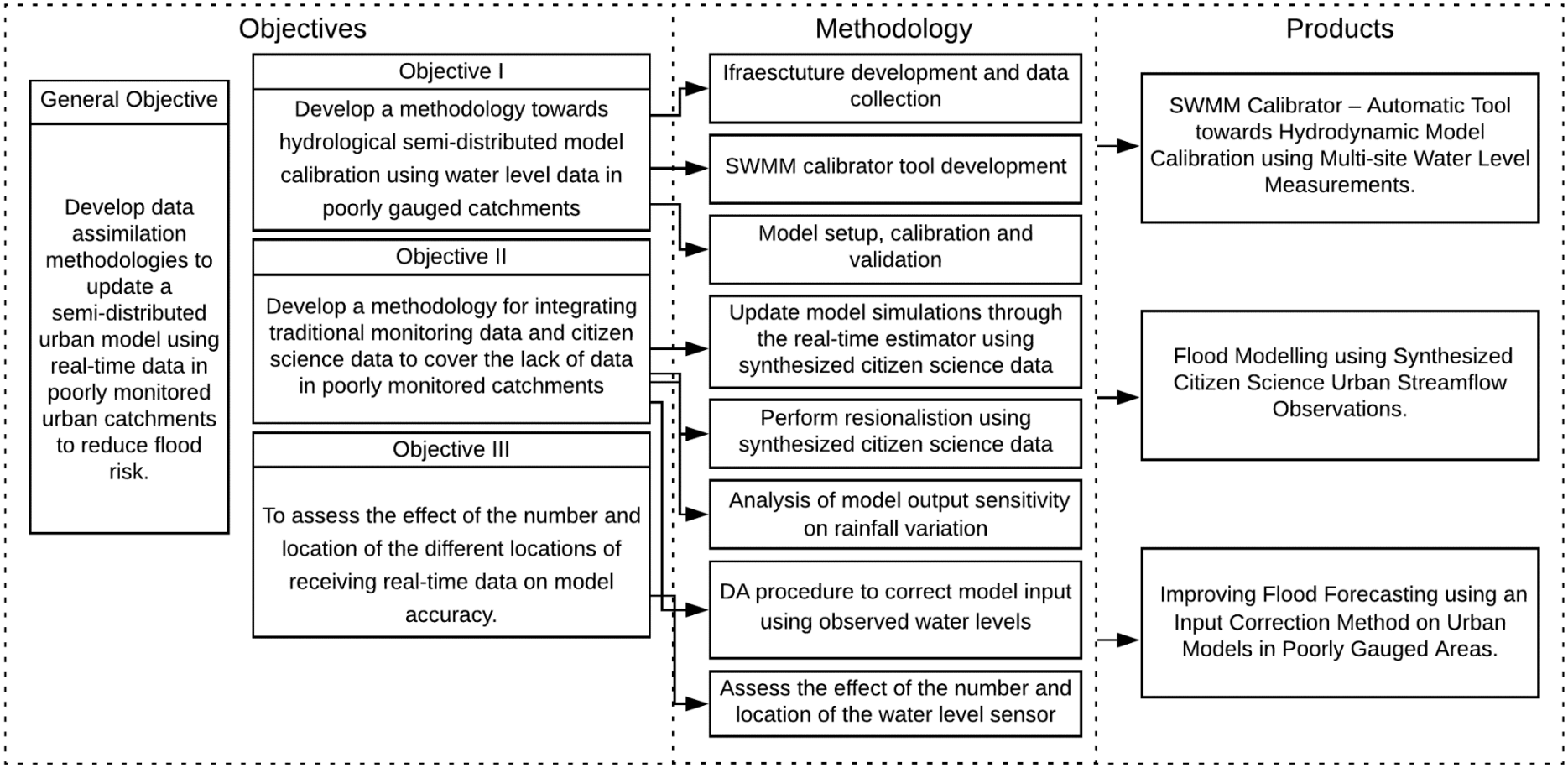


Figure 1-1- Flowchart of the research.

References

Andrade, S. C., Restrepo-Estrada, C., Delbem, A. C., Mendiondo, E. M., & de Albuquerque, J. P. (2017). Mining rainfall spatio-temporal patterns in Twitter: a temporal approach. In *International Conference on Geographic Information Science* (pp. 19-37). Springer, Cham.

Barreto Cordero, W. J. (2012). *Multi-objective optimization for urban drainage rehabilitation*. PhD Thesis, IHE Delft Institute for Water Education, Delft University of Technology.

Basha, E., & Rus, D. (2007). Design of early warning flood detection systems for developing countries. In *2007 International Conference on Information and Communication Technologies and Development* (pp. 1-10). IEEE.

Bodoque, J. M., Amérigo, M., Díez-Herrero, A., García, J. A., Cortés, B., Ballesteros-Cánovas, J. A., & Olcina, J. (2016). Improvement of resilience of urban areas by integrating social perception in flash-flood risk management. *Journal of Hydrology*, 541, 665-676.

CRED - Centre for Research on the Epidemiology of Disasters. (2015). *The human cost of weather-related disasters*, 1995–2015. United Nations, Geneva.

Cools, J., Innocenti, D., & O'Brien, S. (2016). Lessons from flood early warning systems. *Environmental Science & Policy*, 58, 117-122.

Cortes, V. J., Frigerio, S., Schenato, L., Pasuto, A., & Sterlacchini, S. (2013). Review of the current risk management strategies in Europe for hydro-meteorological hazards at protection and emergency level. *Comprehensive Flood Risk Management*, 971-980.

Degrossi, L. C., de Albuquerque, J. P., Fava, M. C., & Mendiondo, E. M. (2014). Flood Citizen Observatory: a crowdsourcing-based approach for flood risk management in Brazil. In *SEKE* (pp. 570-575).

Estes, L. D., McRitchie, D., Choi, J., Debats, S., Evans, T., Guthe, W., Luo, D., Ragazzo, G., Zempleni, R., & Caylor, K. K. (2016). A platform for crowdsourcing the creation of representative, accurate landcover maps. *Environmental Modelling & Software*, 80, 41-53.

Fraternali, P., Castelletti, A., Soncini-Sessa, R., Ruiz, C. V., & Rizzoli, A. E. (2012). Putting humans in the loop: Social computing for Water Resources Management. *Environmental Modelling & Software*, 37, 68-77.

Fritz, S., McCallum, I., Schill, C., Perger, C., See, L., Schepaschenko, D., Van der Velde, M., Kraxner, F., & Obersteiner, M. (2012). Geo-Wiki: An online platform for improving global land cover. *Environmental Modelling & Software*, 31, 110-123.

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography, *GeoJournal*, *69*(4), 211–221.

Gordy, M. (2016). *Disaster Risk Reduction and the Global System: Ruminations on a Way Forward*. Springer.

Granell, C., Havlik, D., Schade, S., Sabeur, Z., Delaney, C., Pielorz, J., Usländer, T., Mazzeti, P., Schleidt, K., Kobernus, M., Havlik, F., Bodsberg, N. R., Berre, A., & Mon, J. L. (2016). Future Internet technologies for environmental applications. *Environmental Modelling & Software*, 78, 1-15.

Hapuarachchi, H. A. P., Wang, Q. J., & Pagano, T. C. (2011). A review of advances in flash flood forecasting. *Hydrological Processes*, 25(18), 2771-2784.

Heinzelman, J., & C. Waters (2010), *Crowdsourcing crisis information in disaster-affected Haiti*, US Institute of Peace Washington, DC.

Keller, E. A., & DeVecchio, D. E. (2016). *Natural hazards: earth's processes as hazards, disasters, and catastrophes*. Routledge.

Leyh, W., Fava, M. C., Abe, N., Restrepo-Estrada, C. E., Horita, F. E. A., de Albuquerque, J. P., & Mendiondo, E. M. (2016), Sdi-node to interlink information, essential for disaster preparedness and management, with other linked open data, in *Proceedings of the ISCRAM 2016 Conference*.

Leyh, W., Fava, M. C., Abe, N., Cavalcante, S., Giatti, L., de Carvalho, C. M., Fonseca-Filho, H. & Jacobs, C. (2017). Citizen Science Involving Collections of Standardized Community Data. In *International Conference on Applied Human Factors and Ergonomics* (pp. 410-420). Springer, Cham.

Mazumdar, S., Lanfranchi, V., Ireson, N., Wrigley, S., Bagnasco, C., Wehn, U., McDonagh, R., Ferri, M., McCarthy, S., Huwald, H., & Ciravegna, F. (2016). Citizens observatories for effective Earth observations: the WeSenseIt approach. *Environmental Scientist*, 25(2), 56-61.

Mazzoleni, M., Alfonso, L., & Solomatine, D. (2017). Influence of spatial distribution of sensors and observation accuracy on the assimilation of distributed streamflow data in hydrological modelling. *Hydrological Sciences Journal*, 62(3), 389-407.

Mendiondo, E. M. (2005). Flood risk management of urban waters in humid tropics: early warning, protection and rehabilitation, in *Workshop on Integrated Urban Water Management in Humid Tropics, Tucci CE., Goldenfum J.(orgs.) UNESCO IHP-VI: Foz do Iguaçu*, vol. 114.

Mendiondo, E. M., Souza, F. A. A., Fava, M. C., Taffarello, D., Macedo, M. B., … & Lago., C. A. (2017). Socio-Hydrology Observatory for Water Security: Examples of Adaptation Strategies with Next Challenges from Brazilian Risk Areas. *AGU Fall Meeting 2017*, Pursuing Water Security in Socio-Hydrological Systems I.

Miorandi, D., I. Carreras, E. Gregori, I. Graham, and Stewart, J. (2013). Measuring net neutrality in mobile internet: Towards a crowdsensing-based citizen observatory, in *Communications Workshops (ICC), 2013 IEEE International Conference on*, pp. 199–203, IEEE.

Montanari, A., Young, G., Savenije, H. H. G., Hughes, D., Wagener, T., Ren, L. L., … & Blöschl, G. (2013). "Panta Rhei-everything flows": change in hydrology and society-the IAHS scientific decade 2013–2022. *Hydrological Sciences Journal*, *58*(6), 1256–1275.

Nardi, F., Annis, A., & Biscarini, C. (2018). On the impact of urbanization on flood hydrology of small ungauged basins: The case study of the Tiber river tributary network within the city of Rome. *Journal of Flood Risk Management*, 11, S594-S603.

Pánek, J., Marek, L., Pászto, V., & Valůch, J. (2017). The Crisis Map of the Czech Republic: the nationwide deployment of a Ushahidi application for disasters. *Disasters*, 41(4), 649-671.

Prosdocimi, I., Kjeldsen, T. R., & Miller, J. D. (2015). Detection and attribution of urbanization effect on flood extremes using nonstationary flood-frequency models. *Water Resources Research*, 51(6), 4244-4262.

Rathore, M. M., Ahmad, A., Paul, A., & Rho, S. (2016). Urban planning and building smart cities based on the internet of things using big data analytics. *Computer Networ*ks, 101, 63-80.

Roy, H. E., Pocock, M. J., Preston, C. D., Roy, D. B., Savage, J., Tweddle, J. C., & Robinson, L. D. (2012). *Understanding citizen science and environmental monitoring: final report on behalf of UK Environmental Observation Framework.*

Schade, S., Díaz, L., Ostermann, F., Spinsanti, L., Luraschi, G., Cox, S., Nuñes, M. & De Longueville, B. (2013). Citizen-based sensing of crisis events: sensor web enablement for volunteered geographic information. *Applied Geomatics*, 5(1), 3-18.

Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., ... & Oki, T. (2003). IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences. *Hydrological Sciences Journal*, 48(6), 857-880.

Smit, B., Burton, I., Klein, R. J., & Street, R. (1999). The science of adaptation: a framework for assessment. *Mitigation and Adaptation Strategies for Global Change*, 4(3-4), 199-213.

Smith, L., Liang, Q., James, P., & Lin, W. (2017). Assessing the utility of social media as a data source for flood risk management using a real-time modelling framework. *Journal of Flood Risk Management*, 10(3), 370-380.

Souza, F. A. A. D. (2019). *Socio Hydrological Observatory for Water Security: conceptualization and study case in São Carlos, Brazil*. (Master thesis, Universidade de São Paulo), 94p. São Carlos, SP.

Timmerman, P., & White, R. (1997). Megahydropolis: coastal cities in the context of global environmental change. *Global Environmental Change*, *7*(3), 205-234.

Turner, B. L., Kasperson, R. E., Matson, P. A., McCarthy, J. J., Corell, R. W., Christensen, L., ... & Polsky, C. (2003). A framework for vulnerability analysis in sustainability science. *Proceedings of the Mational Academy of Sciences*, 100(14), 8074-8079.

Victorino, J. N. C., & Estuar, M. R. J. E. (2014). Profiling flood risk through crowdsourced flood level reports. *In 2014 International Conference on IT Convergence and Security (ICITCS)* (pp. 1-4). IEEE.

Wan, Z., Hong, Y., Khan, S., Gourley, J., Flamig, Z., Kirschbaum, D., & Tang, G. (2014). A cloud-based global flood disaster community cyber-infrastructure: Development and demonstration. *Environmental Modelling & Software*, 58, 86-94.

Wehn, U., Rusca, M., Evers, J., & Lanfranchi, V. (2015). Participation in flood risk management and the potential of citizen observatories: A governance analysis. *Environmental Science & Policy*, 48, 225-236.

Wilkinson, M. E., Mackay, E., Quinn, P. F., Stutter, M., Beven, K. J., MacLeod, C. J., ... & Haygarth, P. M. (2015). A cloud based tool for knowledge exchange on local scale flood risk. Journal of environmental management, 161, 38-50.

Zook, M., Graham, M., Shelton, T., & Gorman, S. (2010). Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake. World Medical & Health Policy, 2(2), 7-33.

2 General methodology

The methodology of this thesis is based on the real-world constraints found in the case study explored in the methodological chapters. The case study is at the Monjolinho catchment in the city of São Carlos, Brazil. The main problems that guided the development of this research are: (1) an urbanised catchment with a time of concentration about two hours in the outlet; (2) more than one location that often suffers from flooding; and (3) absence of flow data.

Based on the problems described above, fieldwork campaigns to collect drainage data and initial flow of the main urban channels are carried out. Also, the catchment morphometric data are extracted from a Digital Elevation Model (DEM). The collected data are used to model the Monjolinho catchment using the semi-distributed and physically-based Storm Water Management Model (SWMM). To validate the methods proposed in chapters 3, 4 and 5 data from water level sensors with a temporal resolution about 5-minutes installed in four locations over the catchment are used. The installation of these sensors is part of a pilot experimental project in partnership with other research groups from the University of São Paulo. As expected from an initial attempt, some miscommunication in the data recording and the lack of constant maintenance resulted in data series with failures. Issues concerning problems with data are a recurring problem in hydrological monitoring series. Considering it, the first contribution of this thesis is the development of an automatic calibration tool that circumvented these limitations in the dataset. Thus, a calibrator that uses water level instead of flow observations capable of multi-site and multi-event data adjustment has been developed.

From a calibration performed with few hydrological events and having only water level data, it has been considered that in addition to the batch calibration performed, it could be complement with real-time updates whenever new information is available by performing data assimilation. Data assimilation can update model components (states and parameters), model output, or model input data. Due to the short time steps required from the urban model simulations because of the urban floods nature, and the complexity of a physically-based model, deterministic methods of data assimilation are proposed in chapter 4 and 5. Both methods perform updates based on observed water level data (Figure 2-1).

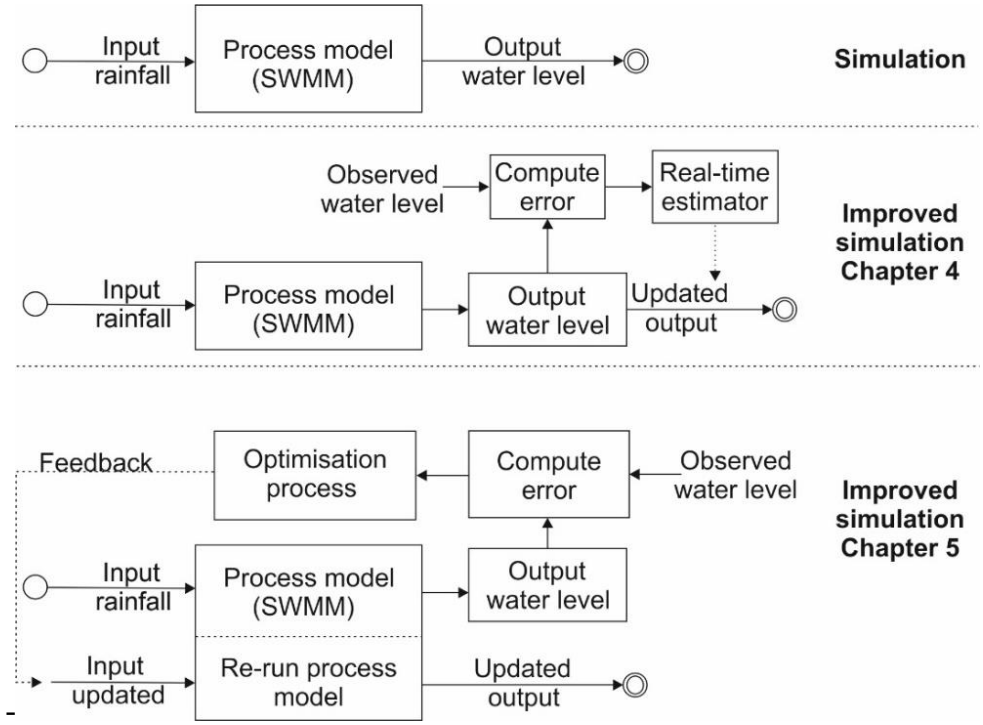


Figure 2-1 – Schematic diagram of updating methodologies proposed in this thesis.

SWMM model

The Storm Water Management Model (SWMM) (Rossman 2010) is a dynamic rainfall-runoff model, with integrated hydrological and hydraulic modules, capable of simulating single events or long-term simulations of runoff quantity and quality in urban catchments. SWMM was developed and is maintained by the United States Environmental Protection Agency - EPA, it is free of charge, open-source, there are many groups of users on the internet, and is widely applied in the literature (Liong *et al.*, 1995; Peterson & Wicks 2006; Müller & Haberlandt 2016; Riaño-Briceño *et al.*, 2016). The version used: SWMM 5.1.011. Link: https://www.epa.gov/water-research/storm-water-management-model-swmm.

Figure 2-2 represents the complete conceptual model of SWMM. However, in this thesis, only the quantity features are explored. The SWMM hydrological processes performed in this study are:

- Time-varying precipitation;
- Rainfall interception from depression storage (initial abstraction);
- Infiltration of rainfall into unsaturated soil layers;

- Nonlinear reservoir routing of overland flow.

Regarding the hydraulic processes, the following processes are performed:

- External inflow of surface runoff;
- Rainfall-dependent infiltration/inflow;
- User-defined inflows;
- Non-uniform routing through open channels.

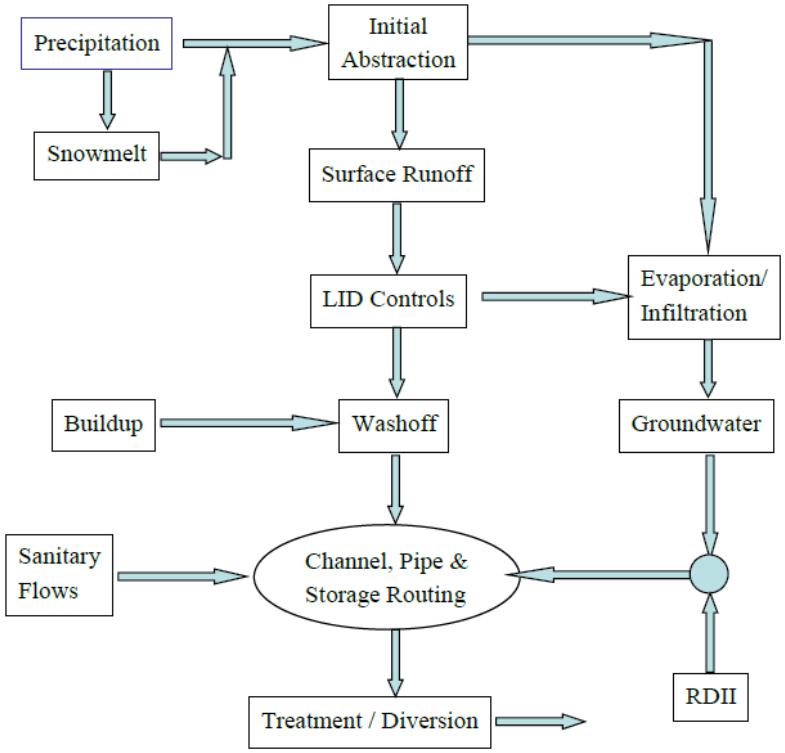


Figure 2-2 – SWMM's model processes (Reproduced from Rossman & Huber 2016).

The hydrological module simulates the sub-catchments behaviour, including an internal infiltration module. The rainfall-runoff component is lumped and conceptual at the sub-catchment scale. The sub-catchment is conceptualised as a rectangular surface with uniform slope and width. Usually, urban areas are composed by a heterogeneous land-use occupation. Thus, the model allows dividing the sub-catchments into pervious (allow rainfall - $P(t)$ to infiltrate - $I(t)$ into the soil) and impervious portions (no infiltration occurs). Additionally, the impervious area can be divided into impervious area with depression storage and without depression storage (runoff of this area starts immediately after a rainfall event occurs). The model connects the runoff ($d(t)$) coming from these three areas to the outlet point, considering the different time steps and ponding areas.

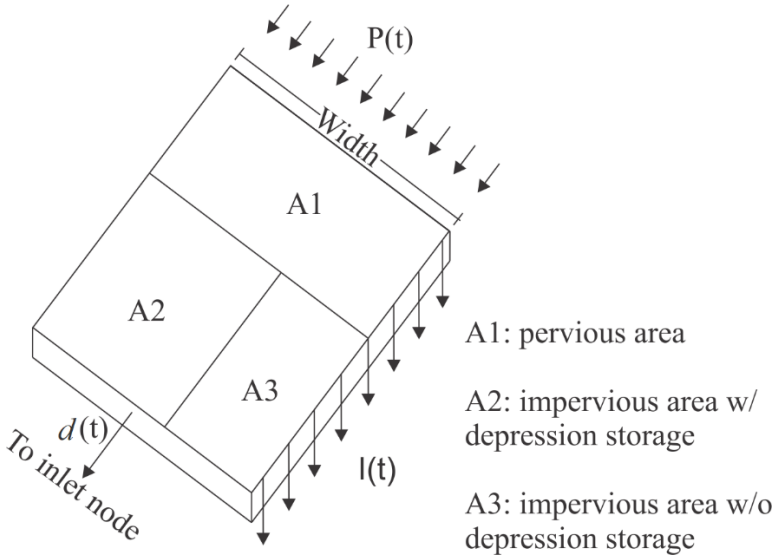


Figure 2-3 – Idealized sub-catchment partitioning.

Each portion of the sub-catchment is modelled as a nonlinear reservoir with a capacity given to the maximum depression storage. The governing equation for the overland flow is the mass conservation, which calculates the change in depth d per unit of time t the difference between inflow and outflow rates over the catchment:

$$\frac{\partial d}{\partial t} = i - e - f - q \tag{2-1}$$

where i is the rate of rainfall, e is the surface evaporation, f is the infiltration rate, and q is the runoff rate.

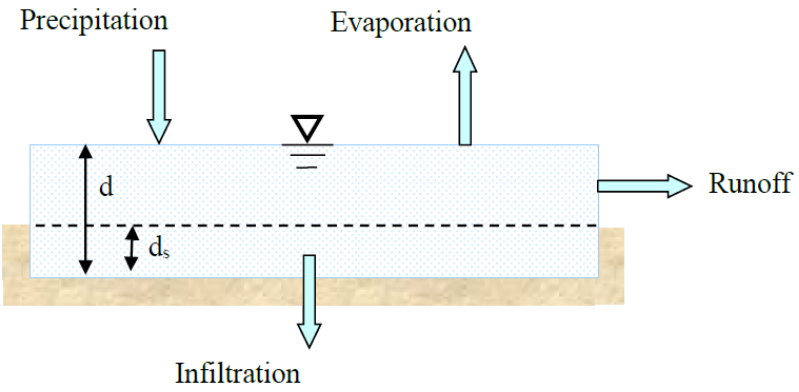


Figure 2-4 – Nonlinear reservoir model of a sub-catchment. (Reproduced from Rossman & Huber 2016), d is the depth of water atop the sub-catchment surface and $d_s$is the depression storage.

As described by Rossman and Huber (2016), the model considers that the flow is uniform over the catchment into a rectangular channel of constant width (*W*) and Slope (*S*), then the Manning equation is used to calculate the runoff's volumetric flow rate (*Q*):

$$R_x = (d - d_s) \tag{2-2}$$

$$A_x = W(d - d_s) \tag{2-3}$$

$$Q = \frac{1.49}{n} W S^{\frac{1}{2}} (d - d_s)^{5/3} \tag{2-4}$$

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

where R_x = is the hydraulic ratios, A is the area across sub-catchments width, A_x is the cross-section area, n is the roughness coefficient and q is the runoff flow rate per unit of surface area.

From the mass conservation and Manning equation, an ordinary nonlinear differential equation is reached to calculate the ponded depth *d* for each time step:

$$\frac{\partial d}{\partial t} = i - e - f - \alpha (d - d_s)^{5/3} \tag{2-6}$$

$$\alpha = \frac{1.49 W S^{\frac{1}{2}}}{A\, n} \tag{2-7}$$

To account the runoff generation for the impervious and pervious surface, the Equation 2-5 is solved individually for each subarea (A1, A2 and A3), with the following α terms:

Pervious subarea A1:

$$\alpha_p = \frac{1.49 W S^{1/2}}{A_1 n_p} \tag{2-8}$$

Impervious subareas A2 and A3:

$$\alpha_i = \frac{1.49 W S^{1/2}}{A_2 + A_3 n_i} \tag{2-9}$$

where n_p is the roughness for the pervious area (A1) and n_i is the roughness for both impervious areas (A2 and A3).

References

Liong, S. Y., Chan, W. T., & ShreeRam, J. (1995). Peak-flow forecasting with genetic algorithm and SWMM. *Journal of Hydraulic Engineering*, 121(8), 613-617. Müller, H., and U. Haberlandt (2016), Temporal rainfall disaggregation using a multiplicative cascade model for spatial application in urban hydrology, *Journal of Hydrology*.

Peterson, E. W., & Wicks, C. M. (2006). Assessing the importance of conduit geometry and physical parameters in karst systems using the Storm Water Management Model (SWMM). *Journal of Hydrology*, 329(1-2), 294-305.

Riaño-Briceño, G., Barreiro-Gomez, J., Ramirez-Jaime, A., Quijano, N., & Ocampo-Martinez, C. (2016). MatSWMM–An open-source toolbox for designing real-time control of urban drainage systems. *Environmental Modelling & Software*, 83, 143-154.

Rossman, L. A. (2010). *Storm Water Management Model user's manual, version 5.0* (p. 276). Cincinnati: National Risk Management Research Laboratory, Office of Research and Development, US Environmental Protection Agency.

Rossman, L. A., & Huber, W. (2016). Storm Water Management Model reference manual Volume I–Hydrology (Revised). *US Environmental Protection Agency: Cincinnati, OH, USA*.

3 SWMM Calibrator – Automatic Tool towards Hydrodynamic Model Calibration using Multi-site Water Level Measurements

*A modified version of this chapter has been submitted as: Fava, M. C., Abe, N. & Mendiondo, E. M. (2020). SWMM Calibrator – Automatic Tool towards Hydrodynamic Model Calibration using Multi-site Water Level Measurements. *Brazilian Journal of Water Resources*.

Abstract

The Storm Water Management Model (SWMM) is a well-established semi-distributed model, which has several applications in the literature to simulate extreme events in urban and semi-urban catchments. The most conventional procedure to calibrate hydrological models is performing the process for the entire watershed, based in a gauging station in the outlet. The demanding of more complex and physically-based models to solve water-related problems start to requires the use of spatial data to calibrate and validate them. A widespread problem to calibrate hydrological models is the absence of discharge data, thus requiring the development of methodologies that allow the use of alternative data. In this study, a multi-event and multi-site calibration tool called SWMM calibrator has been developed. It is based on single-objective optimisation using a Genetic Algorithm (GA) and allows the use of water level data for calibration. A case study in the city of São Carlos, Brazil, is carried out to evaluate the tool. The catchment has four monitoring locations. The results show satisfactory adjustment for three of the four monitoring locations, and a poor fit for validation in one monitoring location that has a shorter observational period.

Keywords: Automatic calibration, semi-distributed hydrological model, multi-site calibration, SWMM.

3.1 Introduction

Rainfall-runoff transformation simulations are a widely known option to provide support to urban water management. Rainfall-runoff models, in a simplified way, are a characterisation of the real-world system, more specifically when transforming rainfall into surface runoff. The setup of these rainfall-runoff hydrological models is usually based on the data available (Henonin *et al*., 2013). Since most of these models are based on physical equations such as the law of conservation of energy and mass, they provide reasonable results depending on the application. However, some purposes require more accurate models. In these cases, the available data can be used as a priori knowledge to calibrate the non-linear behaviour of these models. Thereby, the model will be able to extrapolate this adjustment and perform rainfall-runoff transformations for future events (Moradkhani and Sorooshian, 2009). The calibration of rainfall-runoff models often arises in discussion due to the wide range of methodologies available and particularities in each case that may require improvements in the established ones. In addition, the modeller choices, such as the monitoring sites, its representativeness and dataset treatment, are directly related to the efficiency of calibration (Shinma and Reis, 2014).

Landscape change and ground surface sealing increase flood frequency due to infiltration and flow resistance reduction (and consequently, faster concentration times) (Huong and Pathirana, 2013). Flood forecasting is a management approach widely used to mitigate flood consequences. There are many rainfall-runoff models proposed and tested in the literature to forecast urban floods, from more complex models (distributed models) such as MIKE SHE (Sahoo *et al*., 2006) and InfoWorks ICM (Peng *et al*., 2016), semi-distributed models e.g. SWMM (Lee and Kim, 2018) and HBV-EC (Unduche *et al*., 2018), to simpler lumped models e.g. PDM (Restrepo-Estrada *et al*., 2018) and HEC-HMS (Halwatura & Najim, 2013).

The increasing popularity of more complex physics-based and distributed models in water-related problems raises important issues related to model calibration using spatial data (Zhang *et al*., 2008; Demirel *et al.* 2018; Koster *et al.*, 2018). An efficient way to calibrate models is using all calibration events in only one array, and select the final parameters set that reached better average effectiveness results of the model across all the events. However, Awol, Coulibaly and Tolson (2018) point out that this approach can result in under or overestimation of simulation of any arbitrary event. For this study, the SWMM from the U.S. Environmental Protection Agency (US EPA) (Rossman, 2010) is selected to be calibrated and perform water

level simulations. SWMM is a rainfall-runoff model capable of performing single event or long-term runoff simulations in urban areas through hydrological and hydraulic modules. The runoff component is lumped and conceptual in the sub-catchment scale and physically distributed in the flow and water level simulations by solving full 1D Saint Venant equations.

SWMM is one of the most widely used urban management model (Jiang *et al*., 2015; Pina *et al*., 2016). Many approaches to calibrate SWMM have been developed. Many approaches to calibrate SWMM have been developed. Formiga *et al.* (2016) applied multi-objective optimisation to model an urban catchment; Krebs *et al*. (2013) and Shinma and Reis (2014) developed multi-event and multi-site calibration methodologies by using multi-objective optimisation. All these studies reached good results for calibration and validation, but an automatic and replicable software is not available. Moreover, a lack of flow monitoring data is often a problem all over the world (Jian *et al*., 2017), especially in developing countries, e.g. the case of Brazil which has several urban areas under flood risk and without proper monitoring systems (Mendiondo *et al*., 2017). Taking this into account, this study proposes the SWMM Calibrator, an automatic calibration tool based on multi-site and multi-event observed data. The tool is based on a single-objective optimisation method using a Genetic Algorithm (GA). A real case study at the Monjolinho catchment, in the urban area of São Carlos, Brazil, is performed to assess the usefulness of the calibrator proposed.

3.2 Methodology

In this study, the SWMM model is selected as a computational engine. First, the calibration tool is described, then the experimental area and datasets, and finally, the procedures for model setup and calibration - Figure 3-1 detail the main steps performed towards model calibration.

3.2.1 SWMM calibrator

This tool was developed using Python 3.6, Distributed Evolutionary Algorithms in Python (DEAP) library for genetic algorithms (Fortin *et al*., 2012) and the library SWMM5 for SWMM calling interface (Pathirana, 2015). Urban hydrological-hydraulic models usually require two datasets: the main forcing data (rainfall) and a description of the physical system:

topography, street network, river and sewer system (Re *et al*., 2019). The data required to calibrate the model using the calibrator are -described in the diagram in Figure 3-1.

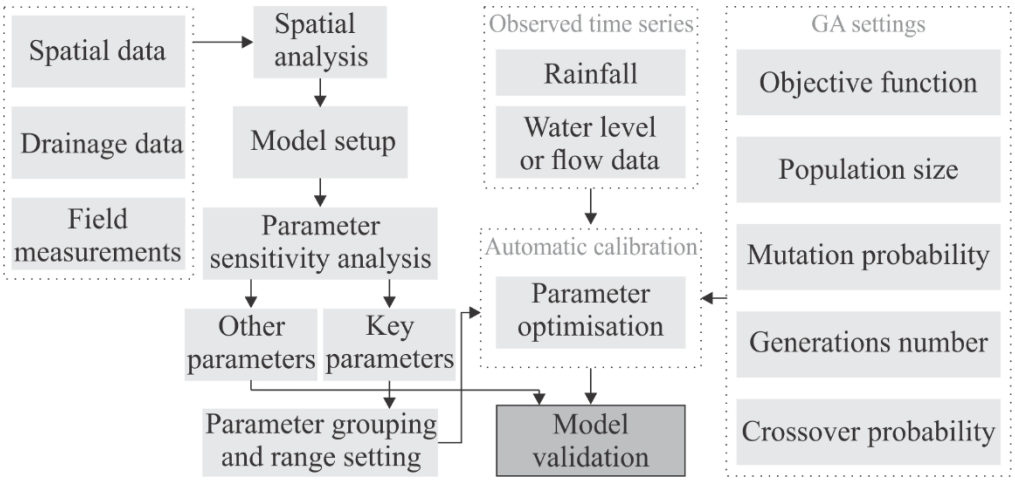


Figure 3-1 – Diagram of the processes performed for model setup, calibration and validation.

When using the calibration tool, the first step is to define its configuration. The SWMM calibrator uses water levels or flows as observed data. Moreover, there is a possibility to calibrate the model based on observations from one or multiple monitoring stations. From this initial information, a proper objective function (OF), between the available, can be chosen. The second step is to define the critical parameters with their boundaries (feasible ranges). The tool allows to divide the sub-catchments into groups with similar features and sets different feasible ranges for its parameters, e.g. a group of intensely urbanised areas and a group of mixed occupation areas (urban and rural).

3.2.2 Objective functions

The SWMM calibrator chooses among four OFs. Some functions combine data from multiple stations into a single index, which is necessary since the set GA is a single-objective approach.

Objective function 1 (F1):

The Nash-Sutcliffe efficiency (NSE) index (Nash and Sutcliffe, 1970) is a statistical evaluation widely used in hydrology field, which compares the simulated and observed values as follows:

$$NSE = 1 - \frac{\sum_{i=1}^{N}(H_i - H_i^*)^2}{\sum_{i=1}^{N}(H_i - \bar{H}_i)^2} \quad (3\text{-}1)$$

$$F1 = \max(NSE) \quad (3\text{-}2)$$

where: H_i is the observed value at the i-th time step; H_i^* is the simulated value at the i-th time step. N is the total number of observations. The range of NSE is in between $-\infty$ and 1. An index of 1 represents a perfect model, and values lower than zero indicates that the mean of the observed values would give a better prediction than the model.

Objective function 2 (F2):

The Average Nash-Sutcliffe Efficiency (AVNSE) is a simple arithmetic mean of the NSE of multiple stations, as described in the following equation:

$$AVNSE = \frac{1}{K}\sum_{j=1}^{K} NSE_j \quad (3\text{-}3)$$

$$F2 = \max(AVNSE) \quad (3\text{-}4)$$

where: j is the index of stations, K is the total number of stations. The AVNSE range is between $-\infty$ and 1.

Objective function 3 (F3):

Spatial Nash-Sutcliffe efficiency (SPATNSE):

$$SPATNSE = 1 - \frac{\sum_{j=1}^{K}\left(\bar{H}_j^* - \bar{H}_j\right)^2}{\sum_{j=1}^{K}\left(\bar{H}_j - \frac{1}{K}\sum_{j=1}^{K}\bar{H}_j\right)^2} \quad (3\text{-}5)$$

$$\bar{H}_j^* = \left(H^*{}_{j_i} : i = [1, x_j]\right) = \frac{1}{x_j}\sum_{i} H^*{}_{j_i} \quad (3\text{-}6)$$

$$\bar{H}_j = \left(H_{j_i} : i = [1, x_j]\right) = \frac{1}{x_j}\sum_{i} H_{j_i} \quad (3\text{-}7)$$

$$F3 = \max(SPATNSE) \tag{3-8}$$

where: $H_{j_i}^*$ are the simulated values for the j-th station at time step i, $[1, K] = \{j \in \mathbb{N}: 1 \leq j \leq K\}$; H_{j_i} is the observed values for the j-th station at time step i; $\overline{H_j^*}$ is the average value of H_j^* for the j-th station; $\overline{H_j}$ is the average value of H_j for the j-th station; x_j is the number of values in the time series for station j; i is the index of time steps with observations in a time series of a station, j is the index of stations. The simulated and observed variables are water levels at sensor locations. The range of SPATNSE is in between $-\infty$ and 1.

Objective function 4 (F4):

Regional Nash-Sutcliffe efficiency (REGNSE) combines data from all monitoring stations in one data series:

$$REGNSE = 1 - \frac{\sum_{l=1}^{M}(H_l^* - H_l)^2}{\sum_{l=1}^{M}(H_l - \frac{1}{M}\sum_{l=1}^{M} H_l)^2} \tag{3-9}$$

$$F4 = \text{Max}(REGNSE) \tag{3-10}$$

where, H_l^* is the simulated value at l-th index; H_l is the observed value at l-th index; l is the index over time steps with observations for all stations; M is the total number of time steps with observations for all stations. The REGNSE range is between $-\infty$ and 1.

Objective functions 2, 3 and 4 are based on efficiency evaluation equations proposed in a study by Lindström (2016), which showed good results for a hydrological model calibration using observed water level data instead of observed discharge or establishing a rating curve.

3.3 Case study and data sets

3.3.1 Monjolinho catchment

The Monjolinho catchment is located in the city of São Carlos, with 246,088 inhabitants (inh.) in the southeast of Brazil. The drainage area of the catchment is about 79.6 square

kilometres (km²), with a time of concentration about 2.1 hours at the outlet. The average annual rainfall is 1400 millimetres (mm) (Juares *et al.*, 2003). Frequently, the catchment suffers from flooding in the urban area; however, there is a lack of hydrological monitoring data, drainage structure, as well as a flood management system. Some studies on frequent floods that affect the city of São Carlos have been carried out, but they only address floods in the region of the Gregório Stream (Boldrin, 2005; Barros *et al.*, 2007; Lima *et al.*, 2007; Righetto *et al.*, 2007; Decina & Brandão, 2016).

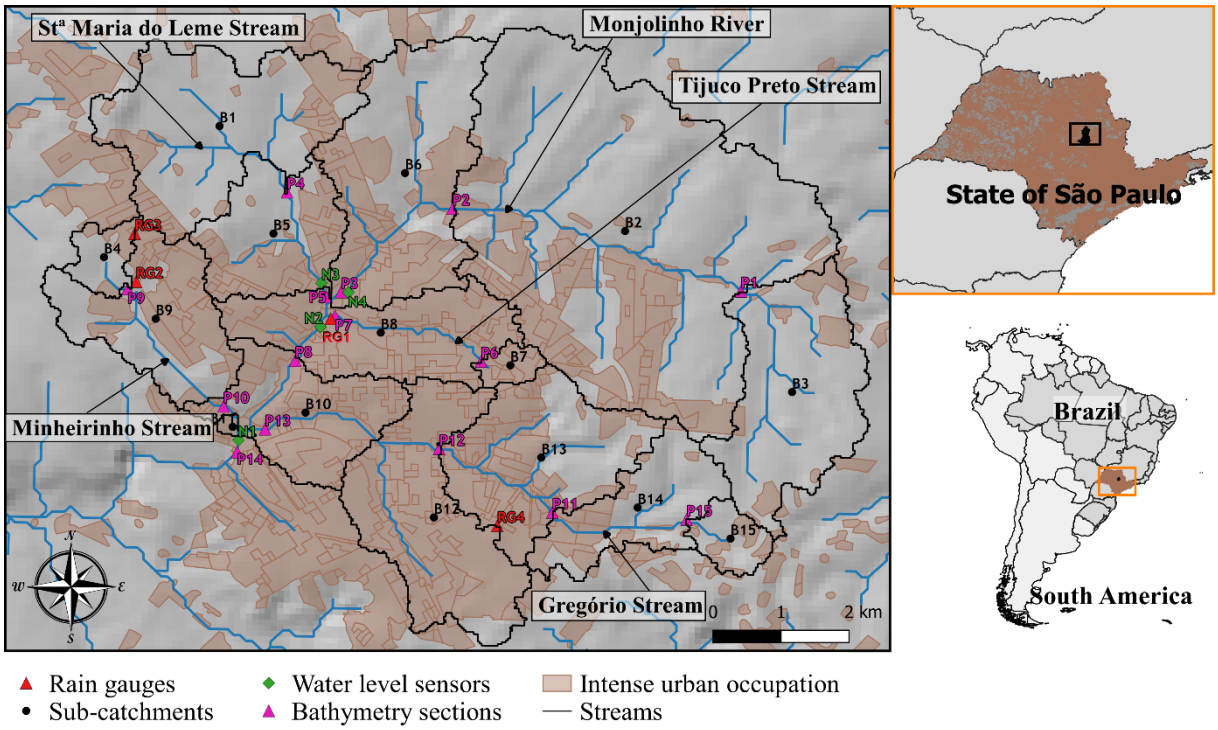


Figure 3-2 - Monjolinho catchment and its monitoring locations (bathymetry, initial flow, rainfall and water level measurements).

3.3.2 Data collection

Field campaigns are performed to collect drainage data and the initial condition of surface runoff. Bathymetry and flow measurements are carried out on the main river, the Monjolinho, and on its main tributaries: Santa Maria do Leme, Tijuco Preto, Mineirinho and Gregório (Figure 3-2). The Monjolinho River starts within the territory of São Carlos city and cuts across the entire urban extension of the municipality. During the field campaign, the conditions of the canal beds and their environment are evaluated. According to the canal conditions, ranges of roughness coefficient values for calibration are estimated. The data collected and the cross-sections are detailed in the supplementary material (Appendix A).

3.3.3 Data sets

Data from four monitoring locations (N1, N2, N3 and N4) collecting water level data every fifteen minutes (Figure 3-2) are used. Sensors N1 and N4 are installed at the Monjolinho River. N1 is at the outlet of the catchment and N4 in a section before the junction of the Monjolinho River with important tributaries: Santa Maria do Leme, Tijuco Preto, Mineirinho and Gregório Stream. Rainfall data are acquired from four rainfall gauges (RG1 to RG4) measuring data every 15 minutes.

3.3.4 Model setup

To extract the topographic data to model the catchment, a Digital Elevation Model (DEM), made available by NASA through the ASTER version2 program with a resolution of 30 m is used. The ArcGIS® Geographic Information System (GIS) with the ArcSWAT extension are used to delineate the Monjolinho catchment using the DEM. The river drainage network from elevation information, length of tributaries, length of main water bodies, average slopes and width of catchments, total area and sub-catchment area are extracted. Each one of the cross-sections are considered as the outlet of the area upstream; then the sub-catchments B1 to B15 are delimited according to these locations (P1 to P15, Figure 3-2).

The infiltration model set up in the SWMM model for Monjolinho catchment is based on the SCS Curve Number. This method is an approach adapted for the Curve Number (CN) of the NRCS (National Resources Conservation Service) to estimate the runoff. From the information gathered in the field and Google Earth® satellite images, the CN values of the catchments are estimated based on NRCS (2004) reference values. The selected flow propagation model is the dynamic wave. The infiltration module based on SCS also requires as input the drying time, parameter set based on the hydraulic conductivity of the soil. The soil of the Monjolinho catchment is predominantly Red-Yellow Latosols (Pedro & Lorandi, 2004). Based on it and the hydraulic conductivity values measured in the field by Maziero *et al*. (2004), Equation 3-11 is applied to determine the drying time.

$$T_{dry} = \frac{3.125}{\sqrt{k_s}} \tag{3-11}$$

where: k_s is the hydraulic conductivity in inches (in)/h and T_{dry} is the time it takes fully saturated soil to recover to a dry state in *days*.

Table 3-1 - Average values of hydraulic conductivity in the urban area of São Carlos – SP.

Soil type	$\boldsymbol{k_{mean}}$ **(in/h)**
Medium silty sand	10.09
Silty fine sand	1.52
Clay sand	0.20
Fine silt clay	0.008

Source: Maziero *et al*. (2004).

The rainfall data acquired from the rainfall gauges (RG1 to RG4) are interpolated to generate the rainfall series for each area (B1 to B15, Figure 3-2) using the Inverse Distance Weighted (IDW) method, thus yielding weighted rainfall values for each sub-catchment, using their centroid as the reference point for interpolation. IDW interpolation assigns values to unknown points using a weighted linear combination of the points with measurements. The weight of each point is the inverse of a distance function. To calculate the interpolation of the value of a point through the IDW method, the following mathematical equation is used:

$$Z(x) = \frac{\sum_{i=1}^{N} \omega_i Z(x_i)}{\sum_{i=1}^{N} \omega_i} \tag{3-12}$$

where $Z(x)$ is the value of the unknown point to be interpolated; *N* is the number of near points used in the interpolation of point x; $Z(x_i)$ is the value of point x_i; e ω_i is the weight of the point x_i value on point x.

$$\omega_i = \frac{1}{\mathrm{h}(x, x_i)^p} \tag{3-13}$$

where $\mathrm{h}(x, x_i)$ is the distance between point x and x_i; p is an arbitrary positive real number called the power parameter.

The estimation of Manning's coefficients for overland flow and depth of depression storage for pervious and impervious areas are set based on reference values presented by Yen (2001) and Viessman and Lewis (2002), respectively. The percentage of impermeable areas per catchment is estimated based on Google Earth® satellite images. For the hydraulic model in the SWMM

model of the Monjolinho catchment: the inflow values, flooding level and cross-section geometry are measured in the field campaigns. The roughness of open channels and conduits are estimated based on Chow's (1959) reference values.

3.4 Sensitivity Analysis

After setting up the initial configuration of the model, a sensitivity analysis of parameters is performed to evaluate the effects on model responses due to parameter perturbation. It is important to recognise that the sensitivity of the parameter in the model is what is being determined, not the sensitivity of the parameter in nature (Saltelli *et al.*, 2008; Iooss & Saltelli, 2017). The evaluated parameters are described in Table 3-2.

Table 3-2 – Description of the parameters evaluated in the sensitivity analysis.

Parameter	**Description**
Slope [%]	Average surface slope of the catchment
CN [-]	Value dependent on land cover representing the soil's moisture
Width [m]*	Width of overland flow path of the catchment
PctZero [%]	Per cent of impervious area with no depression storage
N-Imperv [-]	Manning's coefficient for impervious area
N-Perv [-]	Manning's coefficient for pervious area
S-Imperv [mm]	depth of depression storage on impervious area
S-Perv [mm]	depth of depression storage on pervious area
Roughness [$s/m^{1/3}$]**	Manning's roughness coefficient of canals and conduits
%Imperv [%]	Per cent of impervious area

The initial values of parameters are perturbed from –90% to 90%, with regular intervals of 15%, one parameter at a time (OAT). In order to evaluate the parameter sensitivity, three functions proposed by Gupta (2009) are used (Equations 3-14, 3-15 and 3-16):

$$E_1 = (r-1)^2 \quad (3\text{-}14)$$

$$E_2 = \left(\frac{\sigma_s}{\sigma_o} - 1\right)^2 \quad (3\text{-}15)$$

$$E_3 = \left(\frac{\mu_s}{\mu_o} - 1\right)^2 \quad (3\text{-}16)$$

where: r is the linear correlation coefficient between the original simulation and the simulation with perturbed parameters; (σ_s, σ_o) and (μ_s, μ_o) represent the standard deviations and means of simulation values with perturbed parameters and original simulation values.

E1 is sensitive to the linear correlation between the simulated hydrographs before and after the parameter perturbations and can indicate if the simulated data are similar to the observed data concerning the peak time and rising and falling of limbs. E2 provides information on the variability error, which means the differences in the curves of the hydrographs. E3 can indicate the correspondence in the water balance (Gupta, 2009; Shinma, 2015).

3.5 Calibration setup

Aiming to keep the physical meaning of the model, groups of model features with a different range of parameters value are defined. The Curve Number and Percent of impervious area parameter groups are defined based on the map of land use of São Carlos city (Costa *et al.*, 2012). The sub-catchments chosen to be in CN Group 1 are the ones upstream, with a natural coverage of terrain and which contain the springs of the main streams. Group 2 for CN contained the sub-catchments in fully urbanised areas, and Group 3 contained the partially-urbanised areas. The groups of the %Imperv parameter are divided into four groups, from the less urbanised areas (Group 1) progressively to the completely urbanised ones (Group 4). The roughness parameters are divided into two groups: Group 2 for constructed canals and Group 1 for natural canals or even modified canals that are not covered with cement. The representative width of the surface runoff is calibrated through a multiplier value, extending the variability of the initial values, but still maintaining the order of magnitude of each sub-catchment. Table 3-3 details the groups of the sub-catchments and the values range set for the parameters.

Table 3-3 - Parameters ranges used in the calibration.

Hydrological parameters	**Value ranges**	**Sub-catchments**
Curve number [-] – Group 1	39 ~ 85	B1, B3, B4, B15
Curve number [-] – Group 2	84 ~ 95	B5, B7, B8, B9, B10, B12
Curve number [-] – Group 3	65 ~90	B2, B6, B9, B11, B13, B14
Per cent of impervious area [%] – Group 1	1 ~ 30	B3, B11
Per cent of impervious area [%] – Group 2	5 ~ 35	B1, B5, B10, B15
Per cent of impervious area [%] – Group 3	30 ~ 70	B2, B4, B9, B14
Per cent of impervious area [%] – Group 4	80 ~ 95	B6, B7, B8, B12, B13
Width multiplier value [-]	0.7 ~ 1.3	-
Hydraulic parameters	**Value ranges**	
Roughness [-] – Group 1	0.01 ~ 0.21	-
Roughness [-] – Group 2	0.01 ~ 0.09	-

Data from four water level monitoring locations (N1 to N4, Figure 3-2) are used to optimise the model parameters, thus optimisation with functions F2, F3 and F4 are tested to consider data from all the monitoring stations at the same time. A random seed governing for the runs is set, twenty simulation procedures are repeated for each objective function, and the best solution is selected. Each simulation procedure ran until reaching the number of iterations set, or until reaching the stop criteria. The stop criteria is based on the deviation between the evaluation function of the best set of decision variables found for the current generation (n) $[f(x_n)]$ and the one of the previous generation $[f(x_{n-1})]$. A threshold of 0.0001 was defined; when the deviation is lower than that value during three generations in sequence, the optimisation process stops. Table 3-4 shows the calibration settings. The calibration tool used the range limits as search space.

Table 3-4 – Calibration settings.

Genetic algorithm	Settings
Two-point crossover probability	0.8
Flip bit mutation probability	0.05
Individuals selection	Tournament selection
Decision variables (parameters)	CN, Width, %Imperv, Roughness
Population size	300
Generations	100
Objective functions (OFs)	F2 (AVNSE), F3 (SPATNSE), F4 (REGNSE)
Stop criteria	$\lvert f(x_n) - f(x_{n-1})\rvert / f(x_n) \leq 0.0001$

Nine rainfall events between November 2013 and April 2014 are used for calibration and validation (Table 3-5). The events are grouped into three subsets according to the average intensity rainfall: light, moderate and heavy (Table 3-6). One event of each group is chosen at random, dividing all events into two subsets: half of them for calibration and half for validation (Table 3-7).

Table 3-5 - Selected calibration and validation events, its dates, peak level, total rainfall and duration.

Selected Events	Date	Duration [h]				Total Rainfall [mm]				Peak Level [m]			
		N1	N2	N3	N4	P1	P2	P3	P4	N1	N2	N3	N4
Event 1	04/11/2013	4.67	4.58	5.08	16.20	64.60	47.40	41.60	73.80	2.00	1.56	4.02	-
Event 2	05/11/2013	8.75	4.67	8.92	8.92	9.60	12.20	10.00	8.20	0.74	0.48	0.93	-
Event 3	06/11/2013	9.67	10.0	9.75	10.60	45.2	40.40	37.80	48.20	1.19	1.03	1.43	-
Event 4	07/11/2013	3.33	3.08	0.17	2.92	9.60	2.40	1.60	5.60	0.67	0.48	1.05	-
Event 5	16/11/2013	1.25	1.42	0.00	1.33	10.60	5.60	5.20	16.40	0.79	0.49	1.84	-
Event 6	24/11/2013	0.83	0.92	0.58	1.33	16.60	6.60	10.80	19.40	0.81	0.57	1.11	-
Event 7	10/12/2013	3.25	1.50	1.00	1.17	13.40	9.60	-	5.20	1.25	1.22	1.51	-
Event 8	12/04/2014	5.67	8.75	6.00	4.75	18.20	19.80	16.80	11.60	0.99	0.90	-	1.65
Event 9	13/04/2014	1.00	1.67	1.08	0.17	26.80	25.20	38.00	0.80	2.44	2.28	-	3.00

Table 3-6 - Rainfall intensity classification according to the American Meteorological Society (AMS) and the United Kingdom Met Office (UK Met Office).

AMS		UK Met Office	
Rainfall Intensity	**Classification**	**Rainfall Intensity**	**Classification**
Trace to 2.5	Light rain	Less than 0.5	Slight rain
2.5 to 7.6	Moderate rainfall	0.5 to 4	Moderate rain
Over 7.6	Heavy rainfall	Greater than 4	Heavy rain

Source: AMS (2012); Met Office (2007).

Table 3-7 - Rainfall average intensity classification of the events used for calibration and validation.

Selected Events	**Intensity [mm/h]**	**Classification**		**Aleatory sorting**
		AMS	**UK**	
Event 1	7.14	Moderate	Heavy	Validation
Event 2	1.39	Light	Moderate	Calibration
Event 3	3.86	Moderate	Moderate	Calibration
Event 4	3.19	Moderate	Moderate	Validation
Event 5	8.41	Heavy	Heavy	Validation
Event 6	8.96	Heavy	Heavy	Validation
Event 7	4.80	Moderate	Heavy	Calibration
Event 8	1.37	Light	Moderate	Validation
Event 9	20.43	Heavy	Heavy	Calibration

3.6 Results and discussion

3.6.1 Model output sensitivity to parameter perturbation

The results of the sensitivity analysis of each parameter at a time are shown in the next graphs. Figure 3-3 presents the sensitivity of the parameters of sub-catchment characterisation:

width, slope, percentage of impervious areas and percentage of impervious areas with no depression storage. Among these four parameters, the percentage of impervious areas is the most sensitive, followed by the average width of the overland flow. Width, slope and %Imperv parameters are very sensitive to function E2, showing that they influence the flow-duration of the system. PctZero is more responsive to function E3, compared to the results for function E1 and E2, once the per cent of impermeable areas with no depression storage has some influence on the water balance.

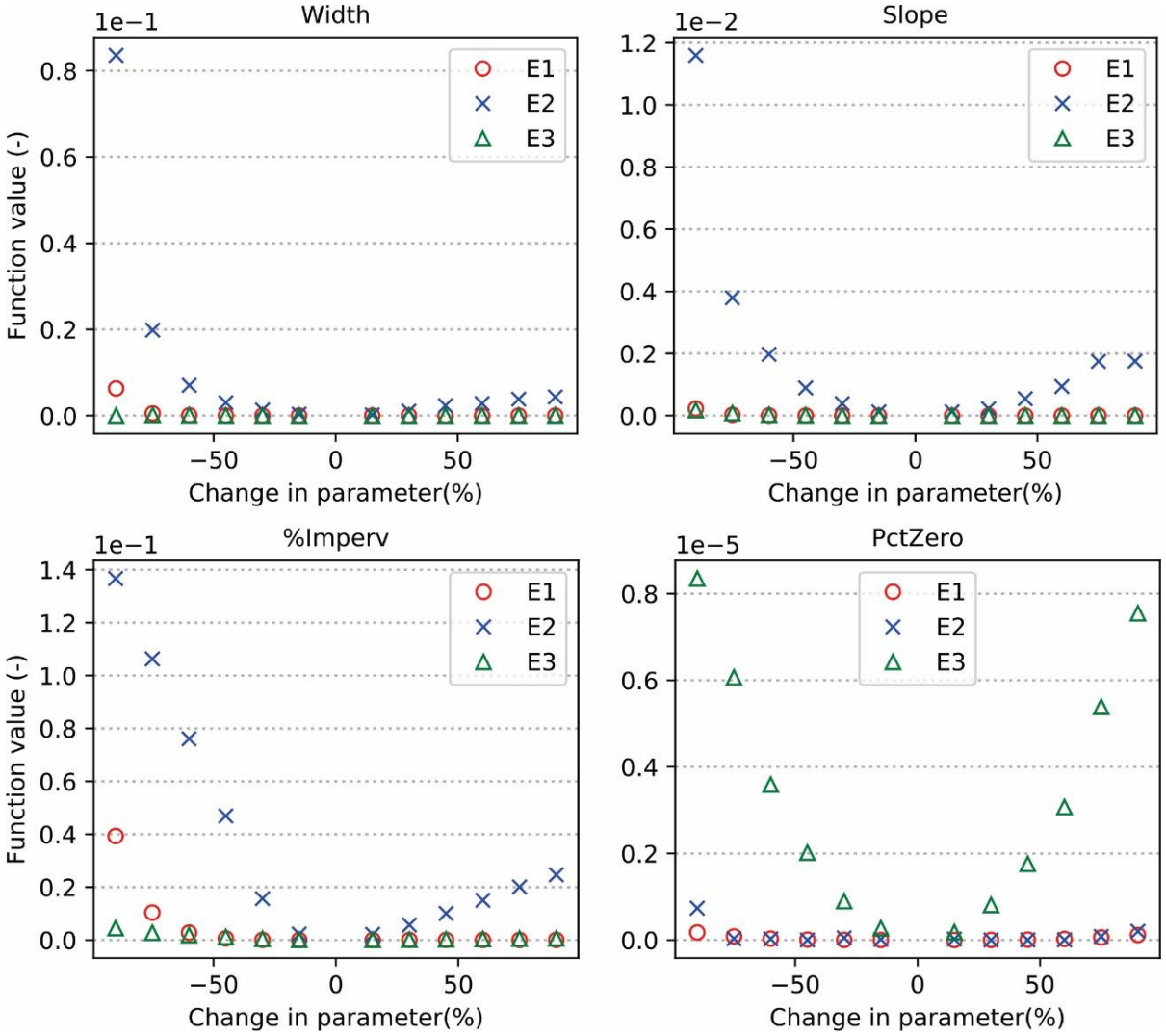


Figure 3-3 - Sensitivity analysis OAT results of sub-catchment parameters: width, slope, per cent of impervious area and per cent of the impervious area with no depression storage.

From the graphs of Figure 3-4, it can be noticed that the parameter S-Imperv has some influence on function E3 and influences almost null over functions E1 and E2. The S-Perv parameter influences function E2 and E3. Both parameters have some relative influence on the

water balance because of their effect on the runoff permeability. When observing the magnitude of y-axes, it can be noticed that Manning's coefficient has more influence on the evaluation functions. Both N-Imperv and N-Perv influence the water balance and flow duration. Manning's coefficient for overland flow in impermeable areas also has some effect on the peak error, but only for a negative variation of the parameter above 50%.

In the graphs in Figure 3-5, it can be seen that the CN and Roughness parameters greatly influence functions E2 and E3, which indicates that they have a great influence on the flow-duration and water balance of the system. Although roughness influences the flow velocity and may advance or delay the peak, its effect on E1 is small compared to its influence on the other two functions.

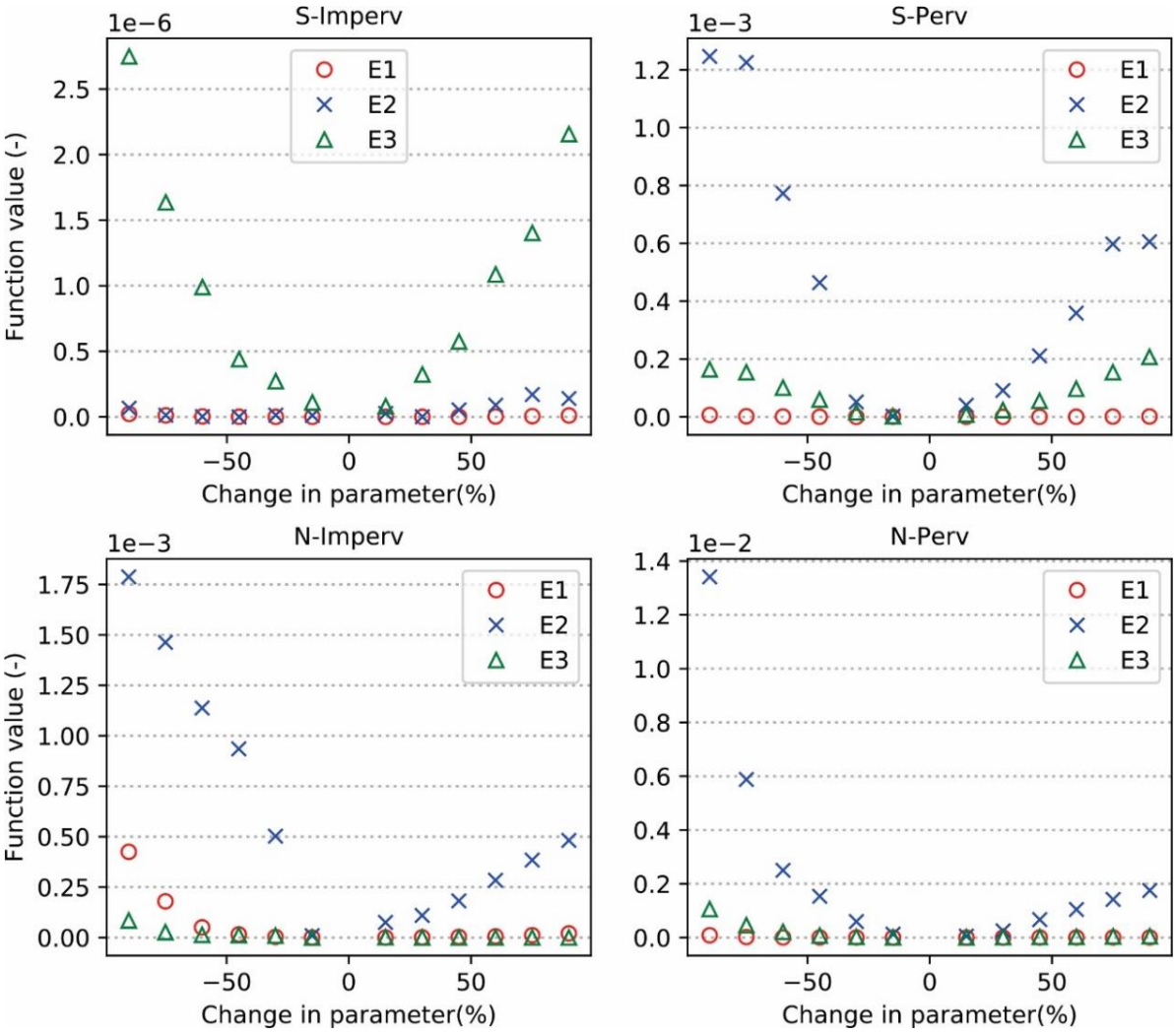


Figure 3-4 - Sensitivity analysis OAT results of permeable and impermeable areas parameters: depth of depression storage and Manning's coefficient.

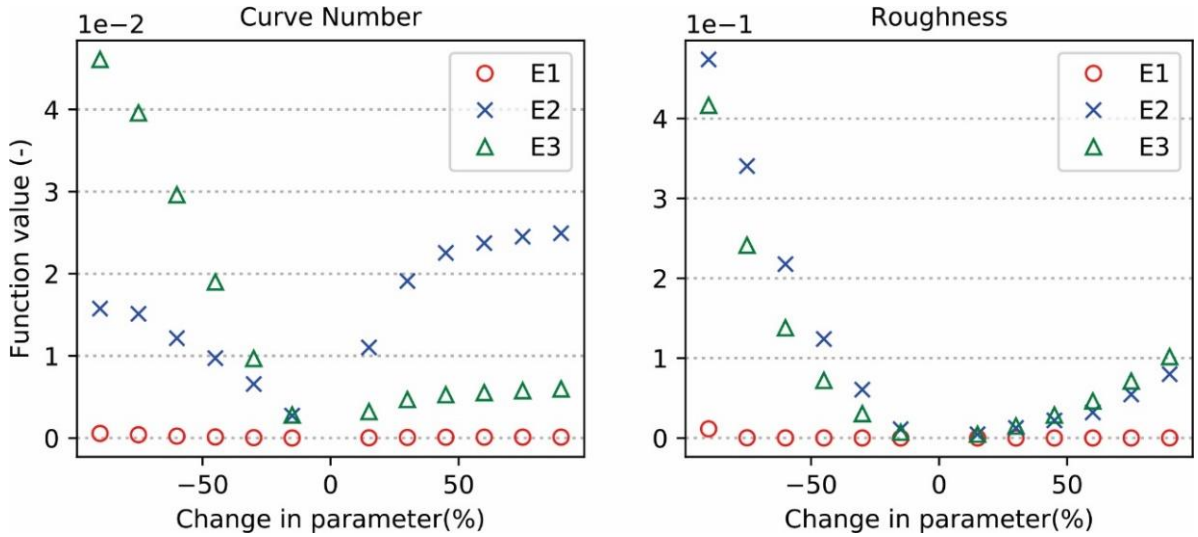


Figure 3-5 - Sensitivity analysis OAT results: the Curve Number parameter of the infiltration model and the roughness of channels and conduits.

When observing the magnitude of the error functions of all the parameters, it can be noticed that the most sensitive parameters are: roughness, % Imperv, Width, Slope, Curve Number and N-Perv. Figure 3-6 shows the clustering of all parameters to compare the variation of the model output due to the percentage of change in their original value. The graph shows the parameters that most influence the output of the model are roughness and the percentage of impermeable areas. Less apparent than these two first, the width and the CN also stand out from the others. Considering these results, the four parameters selected for calibration are %Imperv, Roughness, Width and CN.

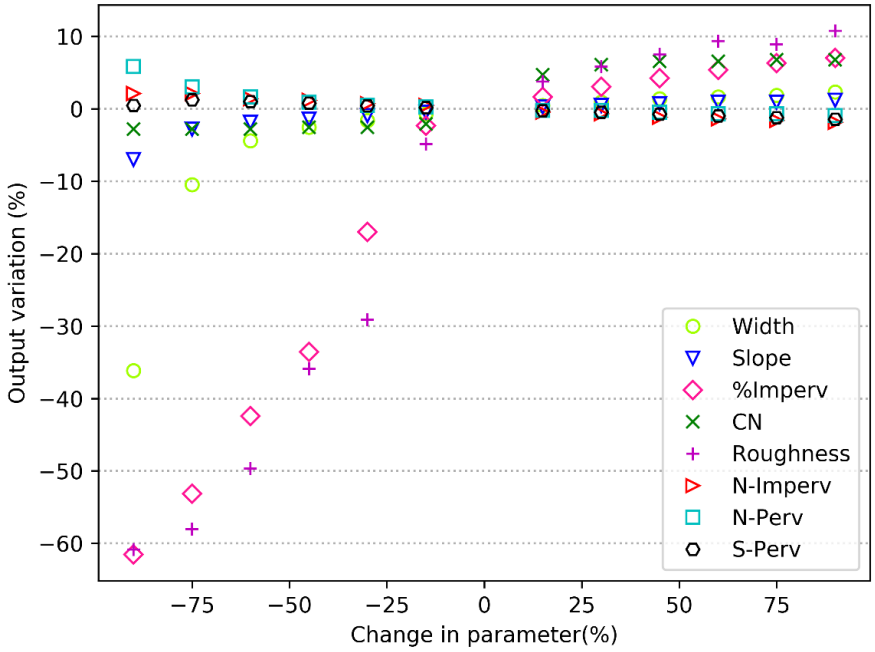


Figure 3-6 - Analysis of model output sensitivity on parameters variation.

3.6.2 Calibration results

After running all the optimisation schemes, the resulting calibrated models for the 20 trials for each of the selected objective functions are evaluated. Table 3-8 shows the NSE of each station of the best solution found for each objective function to compare the results with a common measure for the three OFs. The best result was obtained from function F3, which used SPATNSE in the objective function. It can be noted that this function showed the best balance in the adjustment of the four sensors. Its results showed that the difference between the best NSE found for each station did not deviate significantly from the best NSE obtained from the best set evaluated by the OF, that consider the best-averaged result between the four sensors. The NSEs obtained for the validation period of the best calibration using function F3 are: 0.876, 0.425, 0.288 and -4.32, for sensors N1, N2, N3 and N4, respectively. The following figures show the observed and simulated hydrograph for the calibration and validation events of each sensor.

Table 3-8 - NSE results of the four monitoring stations. The best individual value found for each, and the best average value found (the best solution when considering the index for multiple stations).

	NSE					
Sensor	**F2**		**F3**		**F4**	
	Individual	**Average**	**Individual**	**Average**	**Individual**	**Average**
N1	0.571	0.502	0.584	0.581	0.581	0.562
N2	0.738	0.617	0.750	0.749	0.751	0.749
N3	0.478	0.478	0.569	0.548	0.610	0.516
N4	0.690	0.615	0.772	0.752	0.771	0.763
Index result	AVNSE	0.553	SPATNSE	0.949	REGNSE	0.172

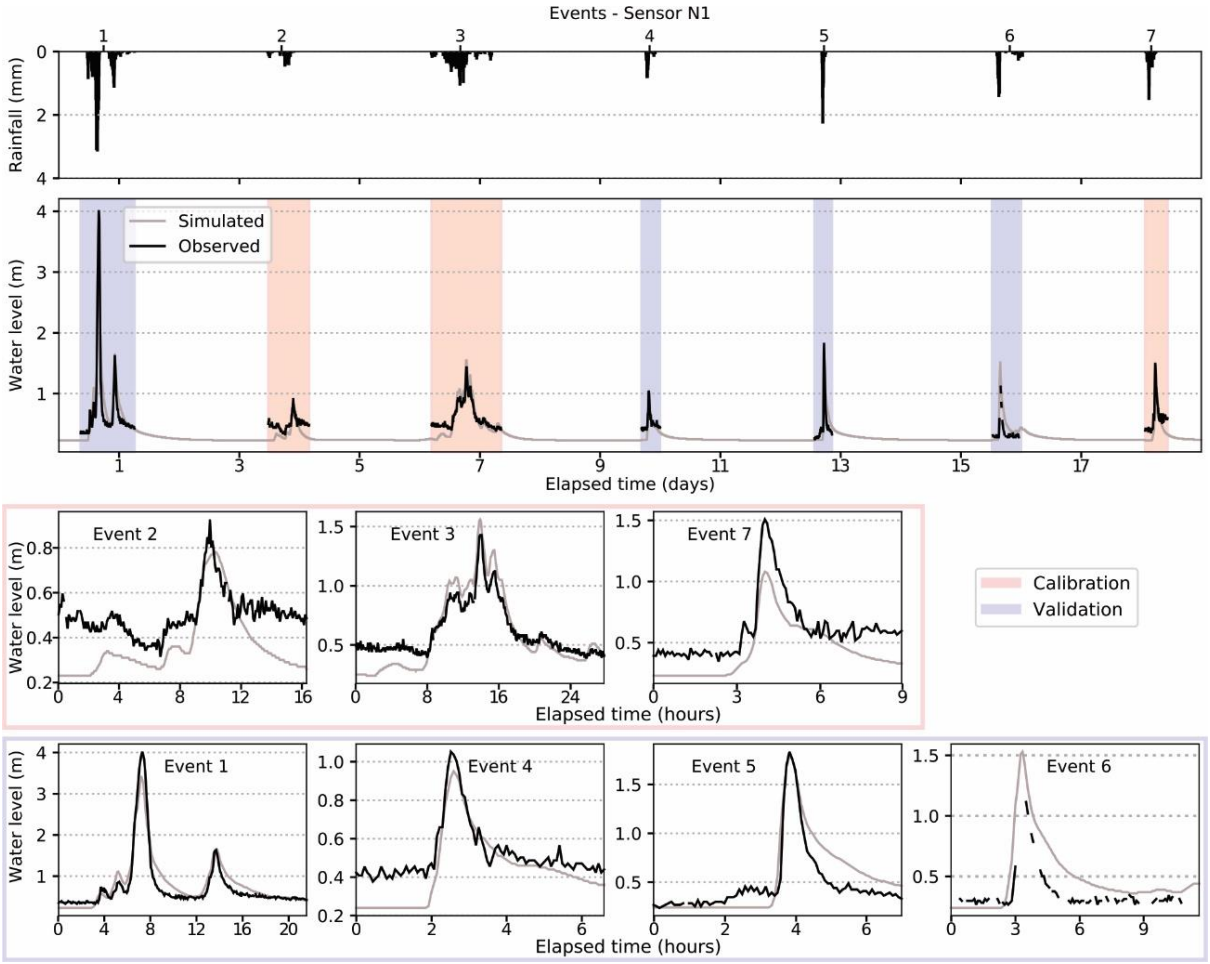


Figure 3-7 - Observed and simulated water level at N1 for calibration and validation events. The calibration events are highlighted in red, and the validation events are highlighted in blue. In the two lower panels, the events are shown in detail.

At sensor N1, the results are satisfactory for calibration events and very good results for validation events, 0.581 and 0.876, respectively. The hydrographs show a good fit in the peaks, except during event 7. There are significant errors in the initial flow of the hydrographs. The measurement errors in the sensors may explain it because it varies between the events.

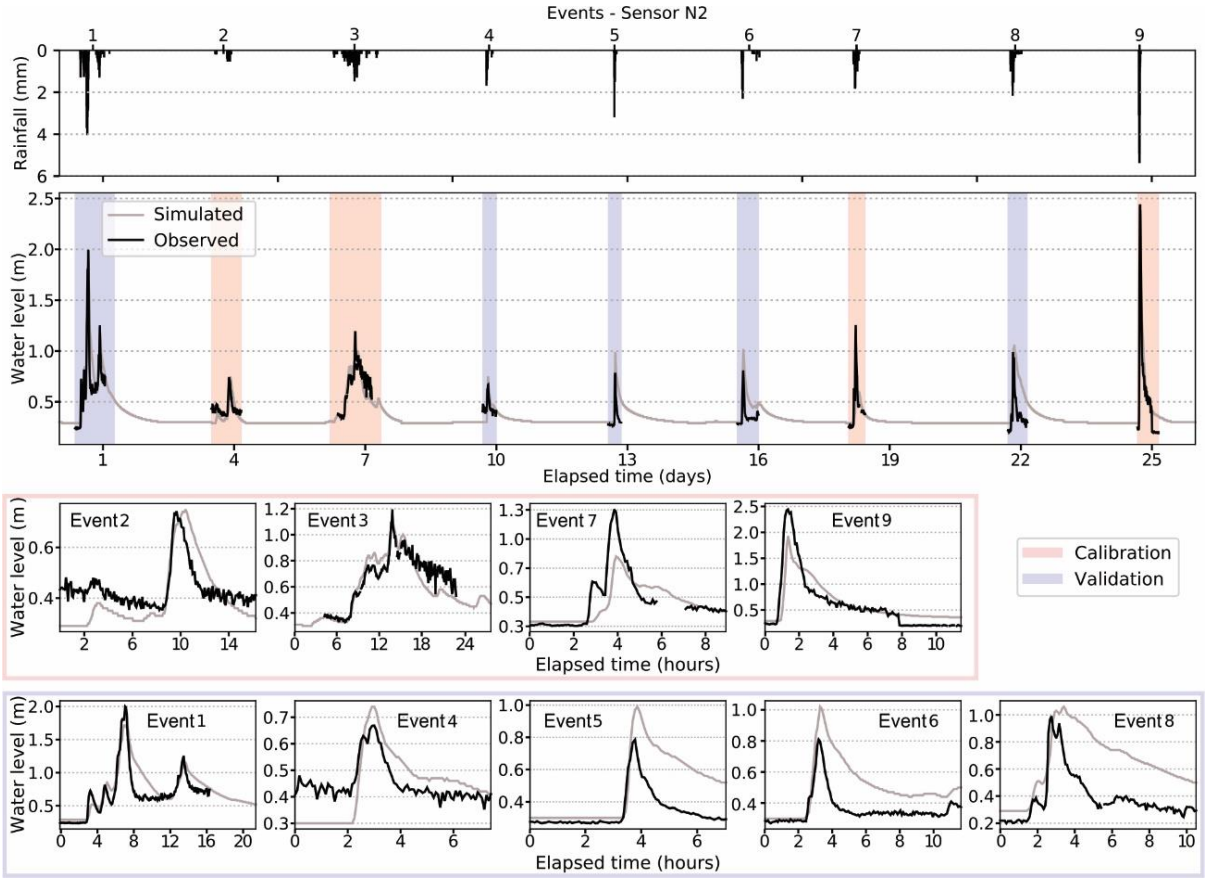


Figure 3-8 - Observed and simulated water level at location N2 for calibration and validation events. The calibration events are highlighted in red, and the validation events are highlighted in blue. In the two lower panels, the events are shown in detail.

The simulations at sensor N2 reached good results for calibration and poor results for validation. In event two, an advance of the simulated hydrograph can be observed; a similar behaviour happened during events 5, 6 and 8. During events 7 and 9 in the calibration, the simulated values are underestimated compared to the observed values. Meanwhile, during validation events 5, 6 and 8, the simulated hydrographs are overestimated. The precipitation errors may explain this due to the spatially localised nature of precipitation fields in urban areas (Zawadzki, 1973; Segond *et al.*, 2007).

Similarly, the calibration results at sensor N3 for events 2, 7 and 9 showed underestimated simulation values, and for validation events 4, 5, 6 and 8, the hydrographs are overestimated. Consequently, the NSE values obtained for calibration are considered satisfactory and unsatisfactory for validation, according to the classification proposed by Moriasi *et al*. (2015).

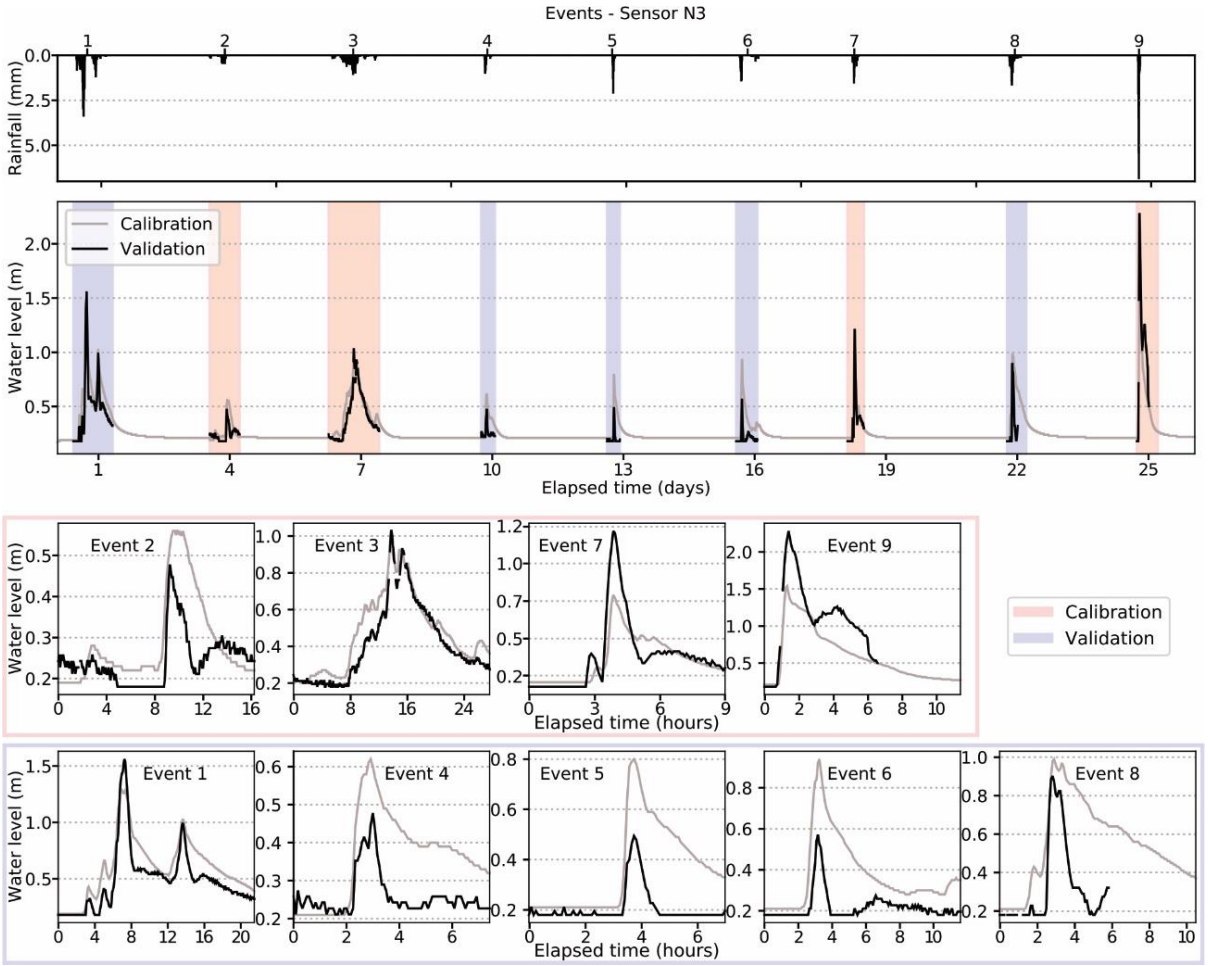


Figure 3-9 - Observed and simulated water level at location N3 for calibration and validation events. The calibration events are highlighted in red, and the validation events are highlighted in blue. In the two lower panels, the events are shown in detail.

The adjustment at sensor N4 has the disadvantage of having just one event for calibration and one for validation (Figure 3-10). For the calibration event, despite the odd behaviour of the observed data, the calibration fits well, reaching an NSE value of 0.763. However, for the validation event, the simulated hydrograph is overestimated compared to that observed, reaching an unsatisfactory NSE of about - 4.32.

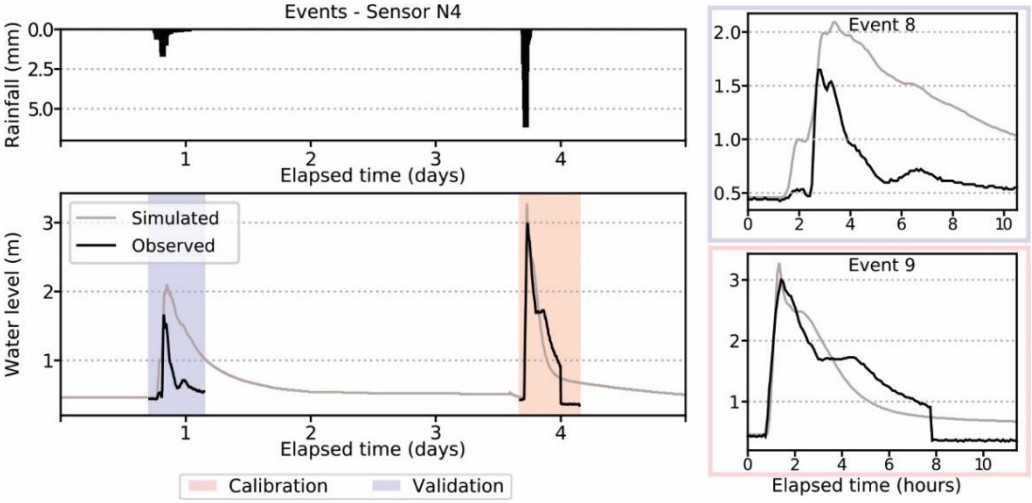


Figure 3-10 - Observed and simulated water level at location N4 for calibration and validation events. The calibration event is highlighted in red, and the validation event is highlighted in blue. In the right panels, the events are shown in detail.

Figures 3-11 and 3-12 show the evolution of the parameters during the iterations over the optimisation with the objective function F3 and the NSE evolution for each sensor. It can be observed that there are other sub-catchments and conduits in the model that influence the water level simulations; the graphs show only the parameter evolution of the sub-catchment upstream monitoring locations. Hence, B10 is the area upstream of sensor N1, B8 is the area upstream of sensor N2, B5 is the area upstream of sensor N3, and B6 is the area upstream of sensor N4.

CN is the parameter that presented the biggest variation when comparing all the graphs in Figures 3- 11 and 3-12. While roughness remains almost unchanged between the iterations, graphically only a small variation in the first iterations at sub-basin B5 is noticeable. It is important to highlight that each iteration corresponds to a generation of the genetic algorithm, and the initial population size is 300 individuals (sets of decision variables). Therefore, several roughness values are evaluated, but the evolution of the parameters at each generation remained almost constant because the best set of decision variables did not change between generations. The algorithm converged relatively quickly, approximately in the thirtieth iteration or earlier. The analysed parameters did not show great oscillation in the efficiency values between the iterations, which could have happened due to an improvement in one sensor adjustment that would negatively affect another. When analysing the NSE evolution, it can be noticed that the compromise between the individual adjustment evolution and the concomitant evolution of all monitoring sites is relatively smooth.

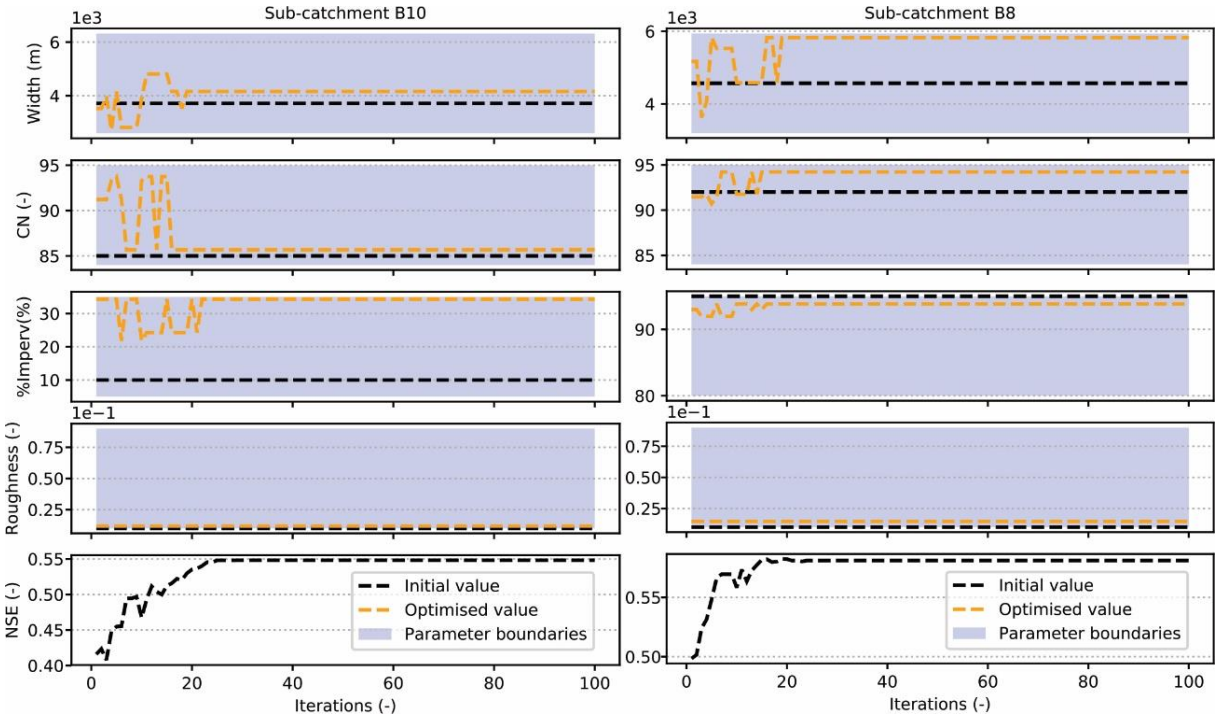


Figure 3-11 - Evolution of parameters during the optimisation process of sub-catchments B10 and B8. NSE evolution of sensor N1 in the left bottom corner and of sensor N2 in the right bottom corner.

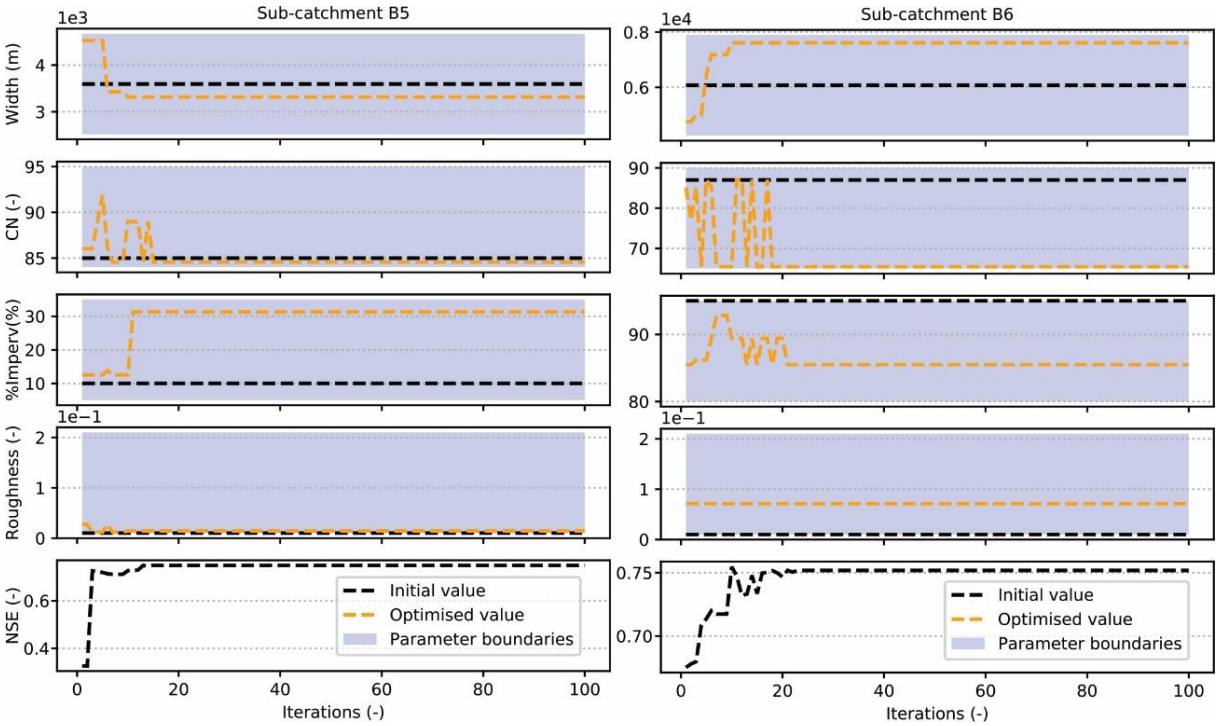


Figure 3-12 - Evolution of parameters during the optimisation process of sub-catchments B5 and B6. NSE evolution of sensor N3 in the left bottom corner and of sensor N4 in the right bottom corner.

3.7 Conclusions

This study reports on and discusses the results of a multi-event and multi-site calibration approach for the SWMM model, applied for a study case in a highly urbanised catchment, with a lack of hydrological monitoring data. Based on the aim of testing different objective functions, due to the limitation of having water level data instead of flow measurements to perform the parameter optimisation, and aiming to calibrate the model by selecting multiple parameter ranges, an automatic calibration tool is developed.

The sensitivity analysis demonstrates that the parameters, which had more influence on the simulation performance are the percentages of the impervious area, Manning's roughness coefficient, the width of the overland flow path and the curve number. Despite the intense urbanisation in the catchment, the infiltration parameter still had considerable influence in model output because the modelling covered the upstream areas, where are the springs of the streams of Monjolinho catchment.

The quality of data plays a fundamental role in the calibration and validation of an urban hydrological model, especially when considering the high temporal and spatial resolution required. In this study, mainly the unstable quality of the observed water level data and the uncertainty in rainfall data affects the model adjustment. In addition, using water level data to calibrate the hydrological model may is not fully capable of reflecting the catchment response to the rainfall inputs. The calibration and validation assessment showed satisfactory results, considering a poorly gauged area, and that in most of the events, the model predicts well the water level peaks. Nevertheless, for a flood prediction system, real-time updates based on new observations can be a promising improvement in the simulations.

Further studies testing calibration efficiency using other datasets, flow rather than the water level, only one monitoring station at a time and other evaluation statistics are strongly recommended. Further studies testing calibration efficiency using other datasets, flow rather than the water level, only one monitoring station at a time and other evaluation statistics are strongly recommended.

Despite the limitations, mainly due to the lack of detailed and high-quality data, this study presents good results of an automatic calibration tool capable of being applied to other case studies.

Acknowledgements

This research was carried out using the computational resources of the Center for Mathematical Sciences Applied to Industry (CeMEAI) funded by FAPESP (grant 2013/07375-0).

References

Awol, F. S., Coulibaly, P., & Tolson, B. A. (2018). Event-based model calibration approaches for selecting representative distributed parameters in semi-urban watersheds. *Advances in Water Resources*, 118, 12-27.

Barros, R. M., Mendiondo, E. M., & Wendland, E. (2007). Cálculo de áreas inundáveis devido a enchentes para o plano diretor de drenagem urbana de Sao Carlos (PDDUSC) na bacia escola do córrego do Gregório. *Revista Brasileira de Recursos Hídricos*, 12, 5-17.

Boldrin, R. S. (2005). *Avaliação de cenários de inundações urbanas a partir de medidas não-estruturais de controle: trecho da bacia do Córrego do Gregório, São Carlos-SP* (Master Thesis, Universidade de São Paulo).

Chow, V.T. (1959). *Open-Channel Hydraulics*, McGraw-Hill, NY.

Costa, C. W., Dupas, F. A., & Pons, N. A. D. (2012). Regulamentos de uso do solo e impactos ambientais: avaliação crítica do plano diretor participativo do município de São Carlos, SP. *Geociências*, 31(2), 143-157.

Decina, T. G. T., & Brandão, J. L. B. (2016). Análise de desempenho de medidas estruturais e não-estruturais de controle de inundações em uma bacia urbana. *Engenharia Sanitária e Ambiental*, 21(1).

Demirel, M. C., Mai, J., Gonzalez, G. M., Koch, J., Samaniego, L., & Stisen, S. (2018). Combining satellite data and appropriate objective functions for improved spatial pattern

performance of a distributed hydrologic model. *Hydrology and Earth System Sciences*, 22, 1299-1315.

Formiga, K. T. M., Carvalho, M. D., PEREIRA, T., & Soares, A. K. (2016). Calibração do Storm Water Management Model (SWMM) utilizando algoritmos evolucionários multiobjectivo. *Revista Brasileira de Engenharia Ambiental e Sanitária*, 21(4), 697-707.

Fundação para o Incremento da Pesquisa e do Aperfeiçoamento Industrial (FIPAI)/ PMSC. *Protijuco - Projeto de Recuperação Ambiental das Várzeas do alto Tijuco Preto visando o Plano Diretor na sua Bacia Hidrográfica*. Contrato Administrativo N° 019/2003. Ano 2003.

Fortin, F. A., Rainville, F. M. D., Gardner, M. A., Parizeau, M., & Gagné, C. (2012). DEAP: Evolutionary algorithms made easy. *Journal of Machine Learning Research*, 13(Jul), 2171-2175.

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of hydrology*, 377(1-2), 80-91.

Halwatura, D., & Najim, M. M. M. (2013). Application of the HEC-HMS model for runoff simulation in a tropical catchment. *Environmental Modelling & Software*, 46, 155-162.

Henonin, J., Russo, B., Mark, O., & Gourbesville, P. (2013). Real-time urban flood forecasting and modelling–a state of the art. *Journal of Hydroinformatics*, *15*(3), 717-736.

Huong, H. T. L., & Pathirana, A. (2013). Urbanization and climate change impacts on future urban flooding in Can Tho city, Vietnam. *Hydrology and Earth System Sciences*, 17(1), 379-394.

Iooss, B., & Saltelli, A. (2017). Introduction to sensitivity analysis. *Handbook of Uncertainty Quantification*, 1103-1122.

Jian, J., Ryu, D., Costelloe, J. F., & Su, C. H. (2017). Towards hydrological model calibration using river level measurements. *Journal of Hydrology: Regional Studies*, 10, 95-109.

Jiang, L., Chen, Y., & Wang, H. (2015). Urban flood simulation based on the SWMM model. *Proceedings of the International Association of Hydrological Sciences*, 368, 186-191.

Juares, J. S. (2003). Composição florística de um fragmento de floresta estacional semidecídua no município de São Carlos-SP. *Revista Árvore*, 27(5), 647-656.

Koster, R. D., Liu, Q., Mahanama, S. P., & Reichle, R. H. (2018). Improved hydrological simulation using SMAP data: Relative impacts of model calibration and data assimilation. *Journal of hydrometeorology*, 19(4), 727-741.

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

Lee, E. H., & Kim, J. H. (2018). Development of a flood-damage-based flood forecasting technique. *Journal of Hydrology*, 563, 181-194.

Lima, G. D., Boldrin, R. S., Mendiondo, E. M., Mauad, F. F., & Ohnuma Jr, A. A. (2007). Análise de incertezas de observações hidrológicas e sua influência na modelagem de pequenas bacias urbanas. *Revista Brasileira de Recursos Hídricos*, 12(1), 107-116.

Lindström, G. (2016). Lake water levels for calibration of the S-HYPE model. *Hydrology Research*, 47(4), 672.

Maziero, T. A., Carneiro, P. H., & Wendland, E. C. (2004). Determinação da condutividade hidráulica de aqüífero freático em área urbana do município de São Carlos, SP. *Águas Subterrâneas*, (1).

Met Office. (2007). Fact sheet no. 3: water in the atmosphere. National Meteorological Library and Archive. September 2007. 13 pp. https://www.metoffice.gov.uk/learning/library/publications/, accessed in 2018.

Mendiondo, E. M., Souza, F. A. A., Fava, M. C., Taffarello, D., Macedo, M. B., … & Lago., C. A. (2017). Socio-Hydrology Observatory for Water Security: Examples of Adaptation Strategies with Next Challenges from Brazilian Risk Areas. *AGU Fall Meeting 2017*, Pursuing Water Security in Socio-Hydrological Systems I.

Moradkhani, H., & Sorooshian, S. (2009). General review of rainfall-runoff modeling: model calibration, data assimilation, and uncertainty analysis. *In Hydrological modelling and the water cycle* (pp. 1-24). Springer, Berlin, Heidelberg.

Moriasi, D. N., Gitau, M. W., Pai, N., & Daggupati, P. (2015). Hydrologic and water quality models: Performance measures and evaluation criteria. *Transactions of the ASABE*, 58(6), 1763-1785.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hhydrology*, 10(3), 282-290.

NRCS (Natural Resources Conservation Service). (2004). Hydrologic soil-cover complexes. Hydrology, Section 4, *National Engineering Handbook*, Part 630.

Ohnuma Jr, A. A., & Mendiondo, E. M. (2014). Análise de cenários com proposição de medidas de recuperação ambiental para a micro-bacia do Tijuco Preto, São Carlos-SP. *Revista Brasileira de Ciências Ambientais* (Online), (32), 42-51.

Pathirana, A., 2015. SWMM5 calls from Python (version 1.1.0.2) [Computer software]. Retrieved from Python Package Index (PyPI) repository: https://pypi.org/project/SWMM5/

Pedro, F. G., & Lorandi, R. (2004). Potencial natural de erosão na área periurbana de São Carlos-SP. *Revista Brasileira de Cartografia*, 56(1), 28-33.

Pina, R., Ochoa-Rodriguez, S., Simões, N., Mijic, A., Marques, A., & Maksimović, Č. (2016). Semi-vs. fully-distributed urban stormwater models: model set up and comparison with two real case studies. *Water*, 8(2), 58.

Peng, H., Liu, Y., Wang, H., Gao, X., Chen, Y., & Ma, L. (2016). Urban stormwater forecasting model and drainage optimization based on water environmental capacity. *Environmental Earth Sciences*, 75(14), 1094.

Re, M., Kazimierski, L. D., & Badano, N. D. (2019). High-resolution urban flood model for risk mitigation validated with records collected by the affected community. *Journal of Flood Risk Management*, 12(S2), e12524.

Restrepo-Estrada, C., de Andrade, S. C., Abe, N., Fava, M. C., Mendiondo, E. M., & de Albuquerque, J. P. (2018). Geo-social media as a proxy for hydrometeorological data for streamflow estimation and to improve flood monitoring. *Computers & Geosciences*, 111, 148-158.

Righetto, J. M., Mendiondo, E. M., & Righetto, A. M. (2007). Modelo de seguro para riscos hidrológicos. *Revista Brasileira de Recursos Hídricos*, 12(2), 107-113.

Rossman, L. A. (2010). *Storm Water Management Model user's manual, version 5.0* (p. 276). Cincinnati: National Risk Management Research Laboratory, Office of Research and Development, US Environmental Protection Agency.

Sahoo, G. B., Ray, C., & De Carlo, E. H. (2006). Calibration and validation of a physically distributed hydrological model, MIKE SHE, to predict streamflow at high frequency in a flashy mountainous Hawaii stream. *Journal of Hydrology*, 327(1-2), 94-109.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., ... & Tarantola, S. (2008). Introduction to sensitivity analysis. *Global sensitivity analysis. The primer*, 1-51.

Segond, M. L., Wheater, H. S., & Onof, C. (2007). The significance of spatial rainfall representation for flood runoff estimation: A numerical evaluation based on the Lee catchment, UK. *Journal of Hydrology*, 347(1-2), 116-131.

Shinma, T. A. (2015). *Avaliação de incertezas na calibração automática do modelo SWMM* (Doctoral dissertation, Universidade de São Paulo).

Shinma, T. A., & Reis, L. R. (2014). Incorporating multi-event and multi-site data in the calibration of SWMM. *Procedia Engineering*, *70*, 75-84.

Unduche, F., Tolossa, H., Senbeta, D., & Zhu, E. (2018). Evaluation of four hydrological models for operational flood forecasting in a Canadian Prairie watershed. *Hydrological Sciences Journal*, 63(8), 1133-1149.

Viessman, W., & Lewis, G. L. (2002). *Introduction to Hydrology*. Prentice-Hall, Upper Saddle River, NJ, Fifth Edition.

Yen, B. C. (2001). Hydraulics of sewer systems. Chapter 6 in *Stormwater Collection Systems Design Handbook*, 151, McGraw-Hill, New York.

American Meteorological Society (AMS). (2012). Rain. Glossary of Meteorology. http://glossary.ametsoc.org/wiki/Rain, accessed in 2018.

Zawadzki, I. I. (1973). Statistical properties of precipitation patterns. *Journal of Applied Meteorology*, 12(3), 459-472.

Zhang, X. S., Srinivasan, R., & Van Liew, M. (2008). Multi-site calibration of the SWAT model for hydrologic modeling. *Transactions of the ASABE*, 51(6), 2039-2049.

4 Flood Modelling Using Synthesized Citizen Science Urban Streamflow Observations*

*A modified version of this chapter has been published as: Fava, M. C., Abe, N., Restrepo-Estrada, C., Kimura, B. Y., & Mendiondo, E. M. (2019). Flood modelling using synthesised citizen science urban streamflow observations. *Journal of Flood Risk Management*, 12(S2), e12498. doi: https://doi.org/10.1111/jfr3.12498.

Abstract

The increase in floods and flash floods over the last decades has motivated researchers to develop improved methodologies for flood risk prevention and warning. Flood forecasting models available today have evolved technologically but are subject to limitations due to the lack of data and limited community participation. This study presents the Hydrological Alert Model with Participatory Basis (HAMPB) model, an approach for integrating water level data reported by citizens, which has the advantage of being inexpensive and potentially highly available, with traditional data to improve flood forecasting. The model assimilates spatiotemporal water levels measured in the field when they are available through a real-time estimator. Random perturbations of up to |10| and |15| cm are added to those data using the Monte Carlo Method to mimic the uncertainty in citizen science data collection. Applying the HAMPB model for urban nested-scale catchments (0.11 km² <= Area <= 21.84 km²) in Brazil shows: (1) significant improvements in flood simulations when field data is assimilated even considering the volunteered data uncertainty; (2) capability to update simulations in more than one point in the semi-distributed hydrological model by a regionalisation method; and (3) flood hazard indexes and their uncertainties show better estimations using field data for updating.

Keywords: flood modelling; citizen science data; short-term forecasting; SWMM.

4.1 Introduction and scope

According to the United Nations (UN), Brazil is among the countries with the highest incidence of floods. Brazil has witnessed hundreds of deaths due to drowning, flooding and landslides. The numbers presented by CEPED (2013) show that between 1991 and 2012, “flash flooding” is the category of disasters which caused the greatest number of deaths in Brazil, 58.15% of the total number of records (floods represent the third largest cause of deaths: 13.4% of the total).

Interdisciplinary studies and vulnerability testing are crucial to prevent and mitigate these impacts to identify, simulate and reproduce environmental conditions and water resource conditions derived from climate change scenarios. The impacts of disasters can be reduced by adopting preventive methods. It is important to develop means to disseminate them across a wide range of stakeholder groups (Coller *et al.*, 2015) and to prepare the population in terms of how to react in disaster situations to minimise damage (Voinov & Bousquet, 2010; Innocenti, 2014; Ramaswamy, 2016; Voinov *et al.*, 2016). Real-time flood modelling allows authorities to make better decisions about where they should direct their attention and offer guidance to citizens to choose safer routes (Smith *et al.*, 2015).

Traditional data acquisition using gauge sensors is an efficient way to collect reliable information with good measurement and transmission frequency. However, it has disadvantages mainly because it is costly and difficult to ensure protection in the field. Recently, the concept of Citizen Science (CS) has emerged, which means volunteer members of the general public (i.e., non-scientists) generate new scientific knowledge (Buytaert *et al*., 2014) by collecting scientific data and actively taking part in research hypotheses and issues, as well as in less participative activities such as "human sensors" (Roy *et al.*, 2012). Goodchild (2007) coined the term Volunteered Geographic Information (VGI) as digital geospatial data generated by common citizens. Leyh *et al*. (2017) highlight that VGI is one of the practices of Citizen Science, where information provided by volunteers must necessarily be georeferenced. This type of data can be a solution to fill the lack of data in ungauged or poorly gauged basins, having the advantage of being inexpensive and promoting community involvement and awareness. (Goodchild, 2007; Poser & Dransch, 2010; Horita *et al*., 2015; Seebauer & Babcicky, 2017).

Crowdsourcing platforms, such as Ushahidi, have been widely used for collecting citizen data in flood-affected areas (Heinzelman & Waters, 2010; Zook *et al*., 2010; Victorino

& Estuar, 2014; Mazzoleni *et al*., 2017; Pánek *et al*., 2016). Data collected by volunteers in an organised way to carry out specific tasks and through different devices in order to create and publish information of interest are called Citizen Observatories (CO) (Miorandi *et al*., 2013; Degrossi *et al*., 2014). The WeSenseIt project also includes in the definition of CO the idea of involving communities, emergency operators and policymakers in an organised platform to discuss, monitor and intervene in situations, places and events (Mazumdar *et al*., 2016).

Yet, data provided by volunteers and crowdsourced data are subject to many uncertainties, and there is a great concern about developing methodologies to determine the reliability of these data (Fritz *et al.*, 2012). Hung *et al*. (2016) reviewed approaches for VGI quality assessment and developed a probability model for VGI quality evaluation. However, the authors mentioned the need for further research using a bigger dataset. Meanwhile, some studies already present good results using volunteered data. Estes *et al.* (2016) obtained data from land cover with 91% accuracy using an open-sourced platform presented in a study called DIYlandcover. Citizen science is a very promising solution to complement real-time sensor data in flood forecasting models (Davids *et al*., 2017). However, the literature review shows a few real case studies performed in that research area and highlight the importance to further work on this issue to develop accurate methodologies to use citizen observations for flood modelling (Assumpção *et al.*, 2017).

Data assimilation (DA) techniques have been widely used for status updating in hydrological models. Mazzoleni *et al*. (2017) developed an assimilation methodology of streamflow data provided by citizens for flood forecasting in conceptual hydrological and hydraulic models in mostly rural catchments and demonstrated the effectiveness and improving forecasting accuracy and promising applications in the field. Lee *et al*. (2012) carried out a study on state updating of distributed rainfall-runoff models via streamflow assimilation in the context of operational hydrologic forecasting, but in a different context from citizen science studies, without considering streamflow information provided by citizens and its uncertainty. DA methods using smoothing techniques, statistics and adjustments of internal and external variables of hydrological models have been more often used and are better than simplistic, direct insertion methods. However, in some cases, simpler methods, which replace the model estimate with the observation, may be useful (Walker *et al.*, 2001; Reichle, 2008; Lee *et al.*, 2012).

Flood forecasting in urban catchments is based on observations and predictions of phenomena in places where people live. Therefore, it is fundamental to test new quantitative models and scientific hypotheses that integrate traditional data and information provided by the population. Taking this into account, a methodological approach to integrate multiple data sources (sensors, voluntary-based data, rain gauges) is developed, called the Hydrological Alert Model with Participatory Basis (HAMPB) model. The main purpose of this study is to evaluate the performance of the HAMPB methodology when assimilating field data to reduce uncertainty in a short-term flood forecasting model. The case study is carried out in a small, almost fully urbanised catchment called Monjolinho, located in the city of São Carlos, São Paulo state, Brazil.

4.2 Case study

This section describes the Monjolinho catchment, monitoring points, the structure developed to collect the different data sources and the hydrological modelling.

4.2.1 Monjolinho catchment

The Monjolinho catchment (Figure 4-1) is located in the city of São Carlos, Brazil a city of 246,088 inhabitants (inh.) in the southeast of the country. Monjolinho is the main river of the basin. It crosses the entire urban area, and all its springs can be found in the city. The urban area of São Carlos is developed within the geographical limits of the Monjolinho catchment. The total area delimited for this study is 79.6 square kilometres (km²), the main river is 15 kilometres (km) long, up to the outlet defined in the modelling. The time of concentration of the catchment is two hours. The population density of the city is 194.53 inh./km². The average altitude of the catchment is 856 meters (m) above sea level and the soil is considered highly permeable.

4.2.2 Hydrological modelling

The catchment model is designed using the Storm Water Management Model (SWMM) from the U.S. Environmental Protection Agency (US EPA) (Rossman, 2010). SWMM is a physically-based and semi-distributed rainfall-runoff model capable of performing single event or long-term simulations through the hydrological and hydraulic modules. The Hydrological

module reproduces the runoff in the sub-catchments and internally includes an infiltration module. The Soil Conservation Service – Curve Number (SCS-CN), Green-Ampt and Horton models are available as infiltration options. The hydraulic module calculates the streamflow in rivers or channels, and the surface runoff that comes from the hydrological model to the watershed during rainfall events.

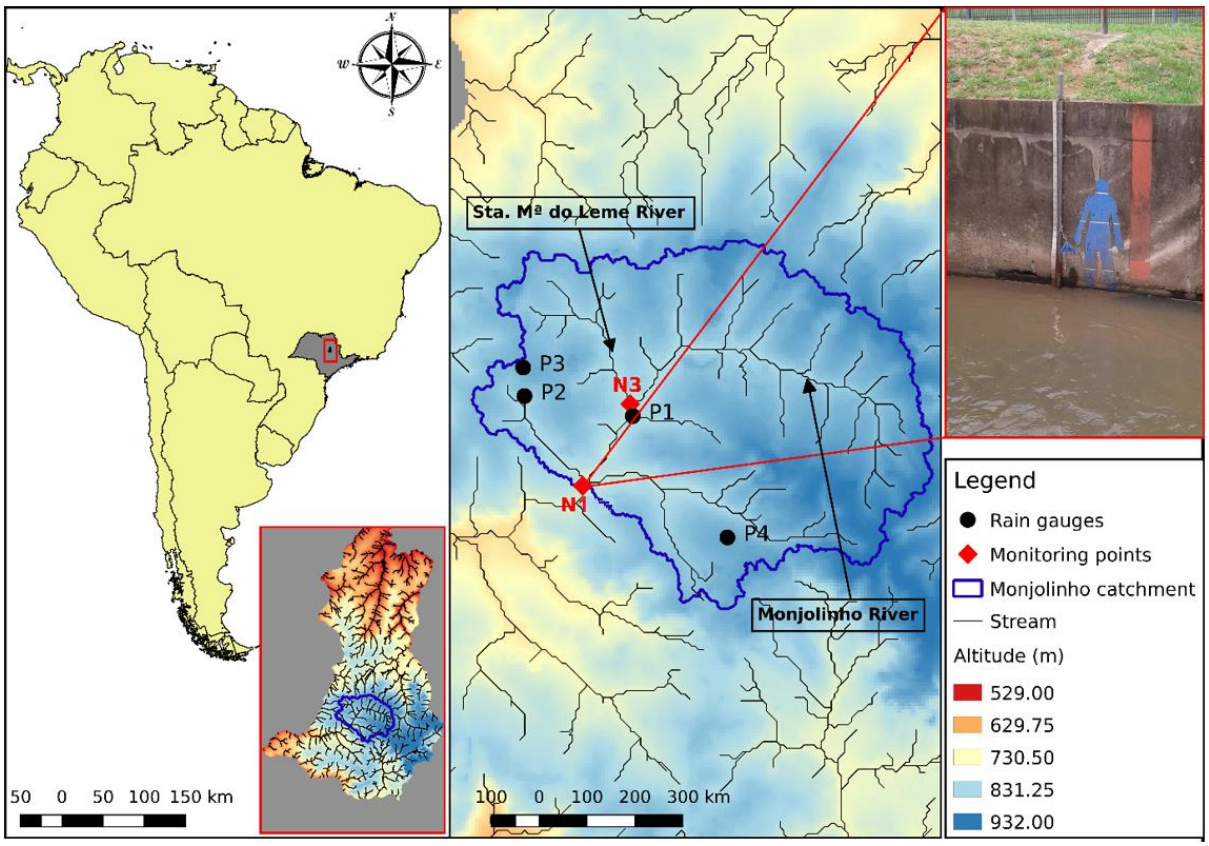


Figure 4-1- Monjolinho catchment in the city of São Carlos and its monitoring points of water level (N1 and N3) and rainfall (P1 to P4). The photo shows different ways of reading volunteer data: low-cost staff gauges, a 1.75 m man-sized pictogram with markings on the knee, waist and neck and a colour scale indicating the hazard index that the person is exposed to if they fall into the canal.

The model consists of 35 nodes and 15 sub-catchments with drainage areas ranging from 0.11 km² to 21.84 km². In the modelling, the infiltration model based on the SCS-CN is chosen. This method is an approach adopted for the curve number (CN) of the National Resources Conservation Service (NRCS) to estimate the runoff. The adopted flow propagation model is the dynamic wave. The parameters used in the hydraulic module are channel roughness, length, cross-section at the 15 staff gauge points, initial depth and the inflow. For the hydrological module, the following parameters are used: area, width, slope, curve number, percentage of

permeable area, percentage of impermeable area and roughness coefficient for permeable and impermeable areas.

Simulations for calibration are performed at two locations of the catchment: N1 and N3 (Figure 4-1) with a time step of 5 minutes. There are water level sensors sending data in real-time at these points, also in time steps of 5 minutes. These data are used to evaluate and calibrate the hydrological model. N3 is located upstream from location N1 in the Santa Maria do Leme River before the confluence with the Monjolinho River. The river bed at this location presents a natural cover, has a width of about 2 m and the channel is 0.96 m deep. N1 is located at the outlet of the catchment in the Monjolinho River. The river bed at this point is rectangular and covered in concrete with a width of about 15 m and the channel is 3.05 m deep. Both areas are prone to the occurrence of floods. Five big rainfall events are used for calibration on the following dates: Event 1 (10/22/2013), Event 2 (04/11/2013 and 05/11/2013) Event 3 (22/11/2013 and 23/11/2013), Event 4 (03/01/2015) and Event 5 (23/11/2015).

The Nash-Sutcliffe Efficiency (NSE) assesses the relative magnitude of the residual variance compared to the measured data variance and shows how well the plot of observed data versus simulated data fits (Nash and Sutcliffe, 1970). Table 4-1 shows the results of the calibration made in the two monitoring locations of the catchment. The obtained average NSE value is 0.519 for location N3 and 0.879 for location N1. According to the criteria defined by Moriasi *et al.* (2007), the calibration shows satisfactory results for location N3 and very good results for location N1.

4.3 Methodology

To determine the efficiency and feasibility of using citizen science data in small urban catchments, a field campaign has been conducted with volunteers to collect water-level data. This social experiment has been conducted in the urban catchment of São Carlos, São Paulo state. Initially, fifteen (15) low-cost staff gauges and visual facilities (Figure 4-2) are installed next to bridges so that the volunteers could read the water levels and give feedback about real-time conditions (Degrossi *et al*., 2014). Based on this experience of humans interacting as sensors for data collection in São Carlos and the need to complement the developed Citizen Observatory, this study developed the HAMPB model, a tool to predict and simulate flood events.

Figure 4-2 - Citizen Science data collection at Monjolinho catchment.

Figure 4-3 shows the data used by the HAMPB model, illustrating its structures and information flow. The input information used in this study is rainfall data obtained from rain gauges, water level acquired using sensors and synthetic voluntary-based data. When the water level data are measured in the field, the water levels simulated by the hydrodynamic model are corrected/updated by the real-time estimator at the location where the information is collected. The regionalisation process is used to update the water levels at other points of the catchment under flood risk and not only where the information is measured. After obtaining: (i) the results from the simulations, (ii) status updating using synthetic volunteered data, and (iii) an analysis of extravasation limits for each sub-catchment, alerts about the current state of the catchment can be issued. Each one of these processes is better explained in the next sections.

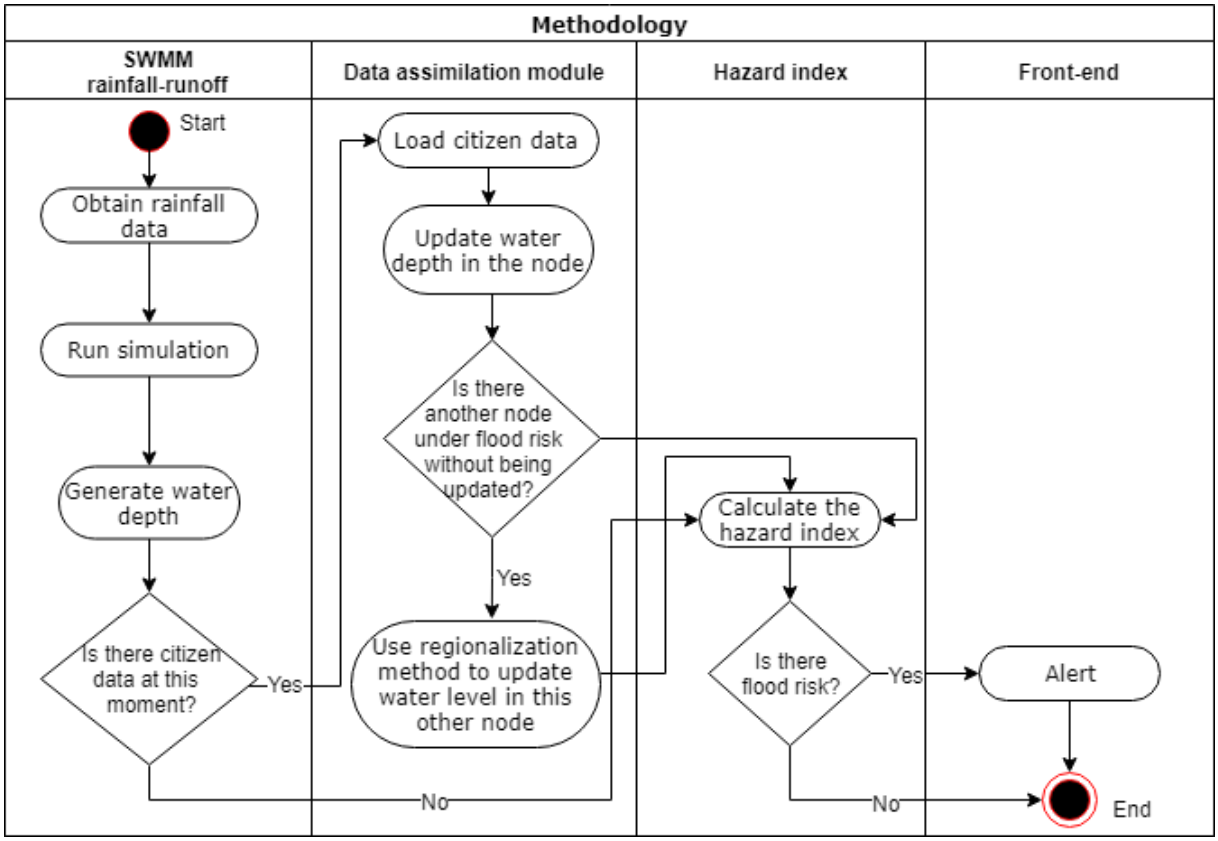


Figure 4-3 - Methodological activity diagram of the HAMPB model for flood forecasting assimilating data provided by citizens.

4.3.1 Water level

The water levels used in this research are obtained by sensors installed at two points of the catchment (N1 and N3). These data are transmitted through a mobile network to a database server. The HAMPB model requires two input data types to validate the methodology: water level collected by sensors and water level informed by the volunteers. As described before, in previous studies, a CO has been built for the urban catchment of São Carlos, and some citizen science data collected and evaluated (Degrossi *et al.*, 2014). The dataset does not match with the periods evaluated in this study because only large rainfall events have been considered. Furthermore, a large amount of data is required to validate the methodology proposed in this study. To fill the lack of volunteered data, quantitative noises are added in the data measured by sensors to mimic voluntary-based water level data. The process is explained in Section 3.4.

4.3.2 Rainfall data

To perform the hydrological forecasting in the short term, the HAMPB model requires rainfall forecasting data as input. However, in this study, only measured rainfall data are used because it is focused on the errors associated with the assimilation of volunteered data in flood forecasting models. In this experimental phase, all steps are taken offline. Various rainfall forecasting models can be used to run real-time applications from global models to ones that use local data from the catchment under study. The uncertainties of each model should be considered before choosing the rainfall estimation model. Rainfall data are acquired from four rainfall gauges (P1 to P4, Figure 4-1). These data are converted into a single array using interpolation to simplify the input for the model using the Inverse Distance Weighted (IDW) method, thus yielding an average rainfall for the entire catchment.

4.3.3 Real-time estimator

In this study, a real-time estimator is proposed, used externally to the model that will assimilate water level information measured by sensors and status updating of the simulated values. The real-time estimator replaces the simulated data for the new information directly, considering it more reliable instead of the simulation made by the hydrological model. The sensor information is assumed to be more reliable than the hydrological model considering that these data are gathered directly in the field. Status updating of the model parameters and state variables have not been performed.

The method performs the assimilation of field data calculating the deviation between the water level measured and the simulated water level obtained by the hydrodynamic model. This deviation is propagated over time until the system receives new field data or reaches the time limit (Figure 4-4). The HAMPB model has a mechanism that limits the lifespan of volunteered data. In the case of the catchment under study, the information is considered valid until one hour after it is supplied. After this period of time, the system considers that the simulated information without updating provides better-estimated values. Empirical simulations determine a one-hour limit in the study catchment.

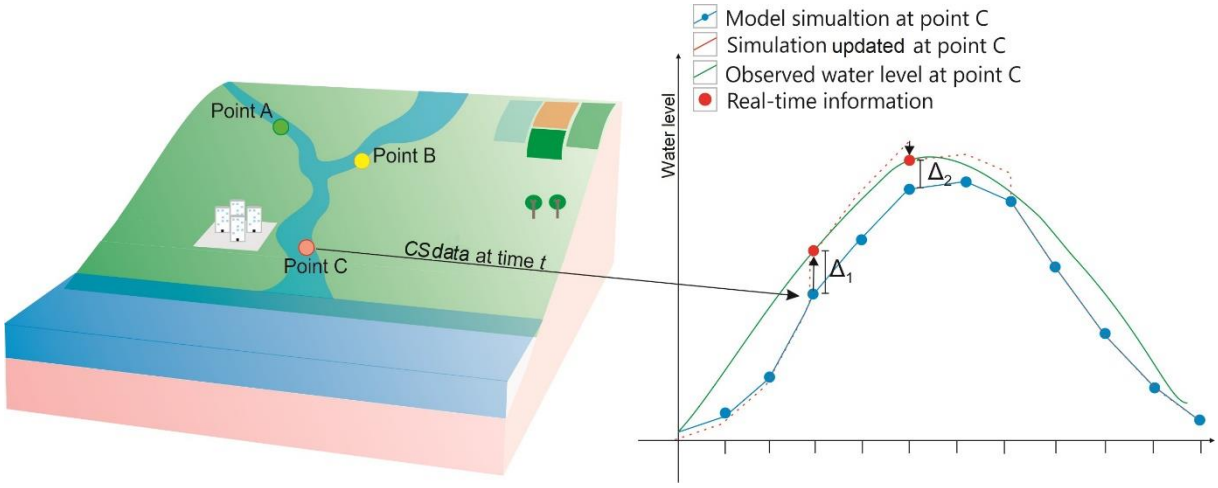


Figure 4-4 – Theoretical representation of the real-time estimator.

As Proof of Concept (PoC), a comparison is made between the forecasting performance when updating the runoff simulations with the real-time estimator and without updating the runoff simulations. The water level simulations performed by SWMM have time steps of 5 minutes. Updating intervals of 1 hour and 2 hours are tested for all the rainfall events mentioned in Section 2.2 (calibration and validation events). The assimilated data is water level field data.

4.3.4 Assessment of uncertainty

Using citizen science data requires great care regarding data reliability. To properly use volunteered data, it is important to apply a reliability classification which takes into account, in particular, the commitment and participation of volunteer in the project (Klonner *et al.*, 2016). To evaluate the predictive performance of the data assimilation considering the uncertainty in the data provided by citizens, the Monte Carlo Method (MCM) is chosen to add small random noises to the field data, emulating volunteers reading errors, assuming a randomly distributed error. This approach enabled to assess the predictions and, indirectly, a quantitative measurement of the acceptable error on the citizen data. Citizen science data may not be reported frequently and regularly. The water level simulations performed by SWMM have time steps of 5 minutes. To assess the effects of assimilating synthetic volunteered data large intervals between the updates of about 1 and 2 hours are tested. The asynchronous behaviour of CS data is not considered.

Eleven sets of random errors are created varying between +/-10 centimetres (cm), with the same size as the field data vector, and are added to them. The same process is repeated in a distinct vector using noises with +/-15 cm of error. The noises are added directly in the water level because it is the parameter which the volunteer will provide, and these ranges are adopted using pessimistic scenarios. Thus, the experiment is repeated using field data + errors. Updating intervals of 1 hour and 2 hours are tested. The efficiency of the forecasting is evaluated calculating the NSE between (i) the observed data and the simple simulation; (ii) observed data and simulation updated with the real-time estimator; (iii) observed data and simulation updated with the real-time estimator considering errors in citizen science data collection.

Errors of +/- 10 cm and +/- 15 cm are chosen by configuration of low-cost staff gauges, usually installed for this type of measurements, taking into account two factors: (i) as the ruler numerically displays marks of 5 cm, the volunteer tends to round up the reading to one of the values written on the ruler; (ii) the variation of water levels in turbulent regimes is also considered since the volunteer can record the highest value, the lowest value or the average level reading.

4.3.5 Regionalisation using citizen science data

After making the corrections of values through the real-time estimator at a given point of the catchment, the disaggregation of the updated water depth values for other sites in the catchment needs to be considered. Considering distributed rainfall-runoff models, catchments that have multiple flood risk areas and the fact that information provided by volunteers may be scarce and asynchronous, the regionalisation of these values will ensure the maximum use of all received data.

In order to use CS data and sensor data, initially, theoretical analysis about the use of volunteer data with sensor data is carried out. A methodology is developed to integrate voluntary information into short-term prediction models presenting a solution to fill data in ungauged places of the catchment:

Given n locations to collect CS data p_i^v for $\forall i \in [1, n]$ and k monitoring locations (water level sensors) p_j^s for $\forall j \in [1, k]$. A reference location is chosen p_c^s for $1 \leq c \leq k$ between the monitoring points. Given a set of simulated flow obtained by historical rainfall series at location

p_c^s, optimal parameters to the rainfall-runoff model are estimated. Assuming that the model has presented satisfactory results for the known point p_c^s, and the points p_i^v are located in the same catchment it can be concluded that the model will also reach satisfactory simulations for these locations because they directly contribute to the simulations in p_c^s. This correlation can be made in all p_j^s locations and integrated into the hydrological forecasting model exemplified here for the single point (p_c^s). After obtaining the hydrodynamic curves through the rainfall-runoff model, a linear regression between the curves obtained in p_c^s with the curves of the other locations $p_i^v$is performed, resulting in a linear regression that correlates values of flow rates at each ungauged point with the known location p_c^s. After obtaining the correlation curves, the CS data can be used to improve and update in real-time the forecasts at the monitored monitoring locations p_j^s, i.e. the information obtained collaboratively can be used to produce estimated values at the known locations, which will be used in the model for predicting the water level at time t + 1, reducing the uncertainty in the model forecasts.

To calculate the correlation between the adjusted value at the point where the information is given and the downstream points in the catchment, equations that relate streamflow values among all points of the catchment are determined. Initially, the water level values are converted into flow, allowing this value to be propagated to the next point. If the calculations are made only using flow, it would generate a large error, as each cross-section of the channel has its particular shape and physical characteristics, thus they probably would present very different flow rates for the same water level variation. To convert the flow rate into the water level, have rating curves are needed (also known as stage-discharge curves or H-Q curves) for the volunteered data collection sites and the sensor-monitored locations. In places where the rating curve is not known, a synthetic rating curve should be determined using historical rainfall data and level and flow simulations performed by the hydrological model.

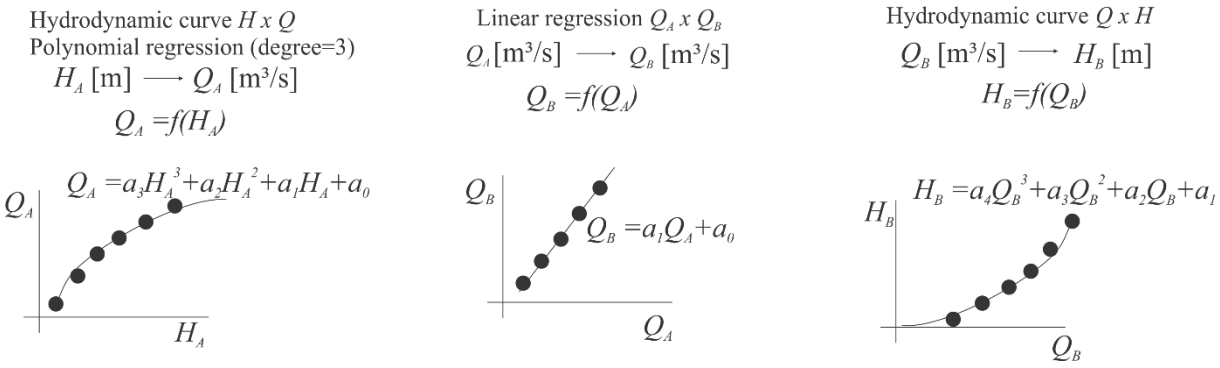


Figure 4-5 – Theoretical representation of the regionalisation method. The water levels at location B (H_B) are updated from the water level values in A (H_A), through the hydrodynamic curves that convert water level to flow values [$Q_A \rightarrow f(H_A)$ and $Q_B \rightarrow f(H_B)$].

The rating curves of the two points of interest (N1 and N3) are estimated using values simulated by SWMM because there are no flow measurements in the catchment of the case study (Figure 4-6). In this study, these estimated rating curves are called hydrodynamic curves since they are not constructed using real measurements. The equations are determined by non-linear regression (polynomial). Finally, the linear regressions between simulated flows at the points of interest (Figure 4-7) are defined.

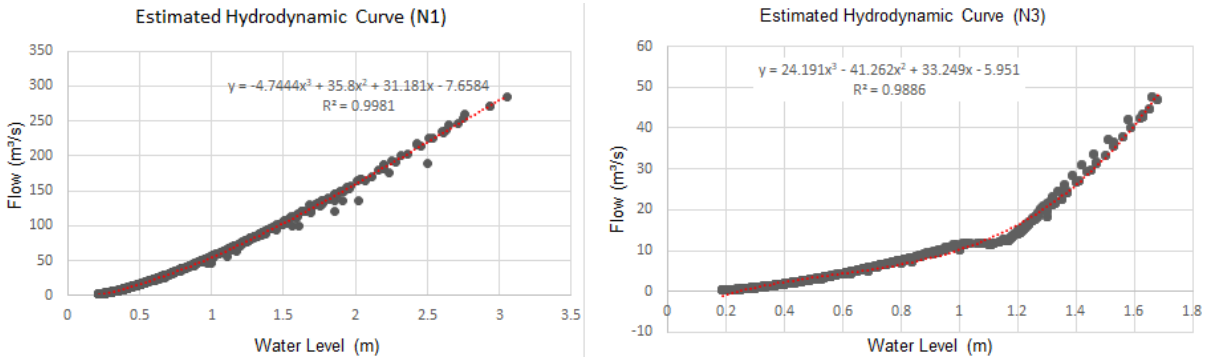


Figure 4-6 – Hydrodynamic curves for location N1 and location N3.

Estimated hydrodynamic curves node N1

$$Q_{N1} = -4.744h_{N1}^3 + 35.8h_{N1}^2 + 31.181h_{N1} - 7.658 \quad (4\text{-}1)$$

$$R^2 = 0.9981$$

$$h_{N1} = e^{-7Q_{N1}^3} - 5e^{-5Q_{N1}^2} + 0.017Q_{N1} + 0.226 \quad (4\text{-}2)$$

$$R^2 = 0.9975$$

Estimated hydrodynamic curves node N3

$$Q_{N3} = 24.191h_{N3}^3 - 41.262h_{N3}^2 + 33.249h_{N3} - 5.951 \quad (4\text{-}3)$$

$$R^2 = 0.9886$$

$$h_{N3} = 3e^{-5Q_{N3}^3} - 0.003Q_{N3}^2 + 0.102Q_{N3} + 0.207 \quad (4\text{-}4)$$

$$R^2 = 0.9954$$

The model is better calibrated at location N1 compared to location N3. For this reason, the water level simulations for location N3 are updated by correlation with location N1. Field data are assimilated at location N1 through the real-time estimator. Once updating the water level values, the curve that correlates the depth and flow values are used for location N1 to calculate the correspondent flow. After obtaining the theoretical flow values for location N1, a regression that relates the flow values of location N1 with flow values for location N3 is applied to obtain the estimated flow rate for location N3. Finally, by using the hydrodynamic curve for location N3, its updated water level is obtained.

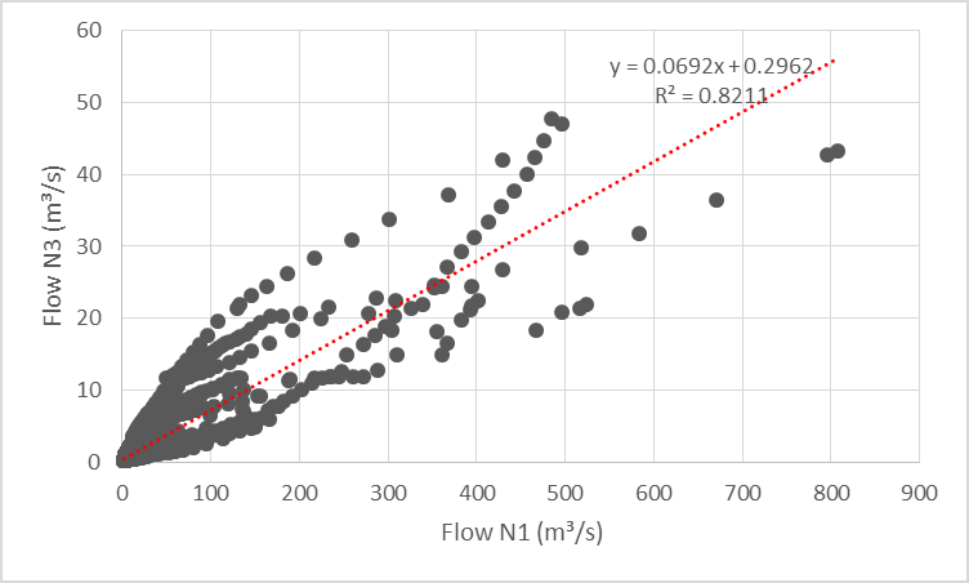


Figure 4-7 - Regression curve of flows between the two points of interest (N1 and N3).

Correlation curve between node N1 and N3

$$Q_{N3} = 0.069Q_{N1} + 0.2962 \quad (4\text{-}5)$$

$$R^2 = 0.821$$

4.3.6 Hazard Index

Hazards are defined as situations which have the potential of resulting in harm. Tingsanchali and Karim (2005) point out that flood hazards always depend on a local scale, that represents the degree of hazard called Hazard Index (HI). In a study performed by Rotava (2013), the HI for people are determined, considering the structure of the urban channels at locations N1 and N3 in the Monjolinho catchment. The equation used to calculate HI takes into account the physical characteristics of the channels, water velocity and water depth to determine the risk that people near these sites are exposed. The study uses colour bands that classify HI as low, medium, high and very high corresponding to the threshold values of 0.5, 1, 1.4 and above 1.4, respectively. For the places where the HI is determined (location N1 and N3), a 1.75 m man-sized pictogram with markings on the knee, waist and neck, as well as a colour scale beside it indicating the hazard index, are drawn on the canal wall (Figure 4-1) to make it easier

to collect citizen science data. The HAMPB model has a water depth forecasting as a product, and the conversion of these values into the HI is performed to analyse and classify flood alerts.

4.4 Results

The evaluation of the forecast adopting the methodology described above is performed based on the NSE (described in Section 2.1). All the graphs in the following Results sections are sampled for the flood event which occurred in the urban catchment of São Carlos on November 4, 2013. The efficiency evaluation of the methodology is done for the sampled period in the graphs, and for all the events used in the calibration and validation of the hydrological model in order to obtain more reliability in the performance of the method. NSE for all the situations are compared in Table 4-1.

4.4.1 Performance of the assimilation method

Figures 4-8 and 4-9 show the performance of the real-time estimator assimilating field data every hour for the two observational points N1 and N3 and the surrounding details of the observational places. The graphs at the top in the figures show the water depth simulations without updating (red line), with updating every hour (blue line) and the observed values. The bottom graphs show the evolution of per cent error between observed and simulated values during the flood event to better compare the influence of assimilating data.

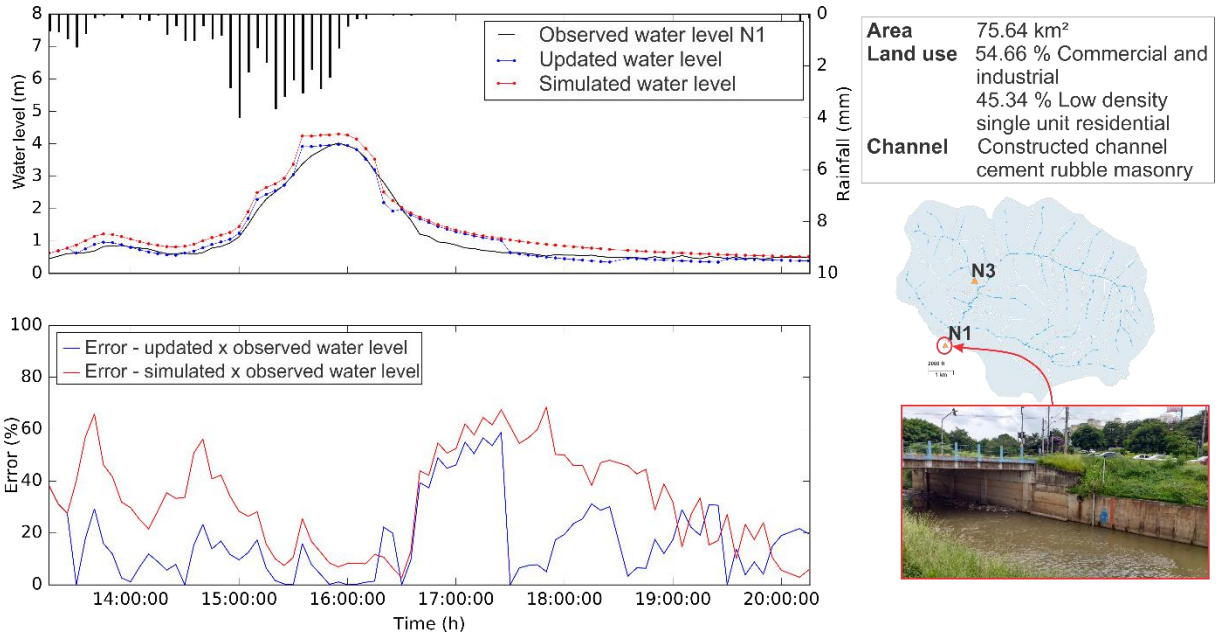


Figure 4-8 - Results of the updates made using the real-time estimator at location N1. Land use classification, according to Herold, Liu and Clarke (2003) and channel characteristics from Te Chow (1959).

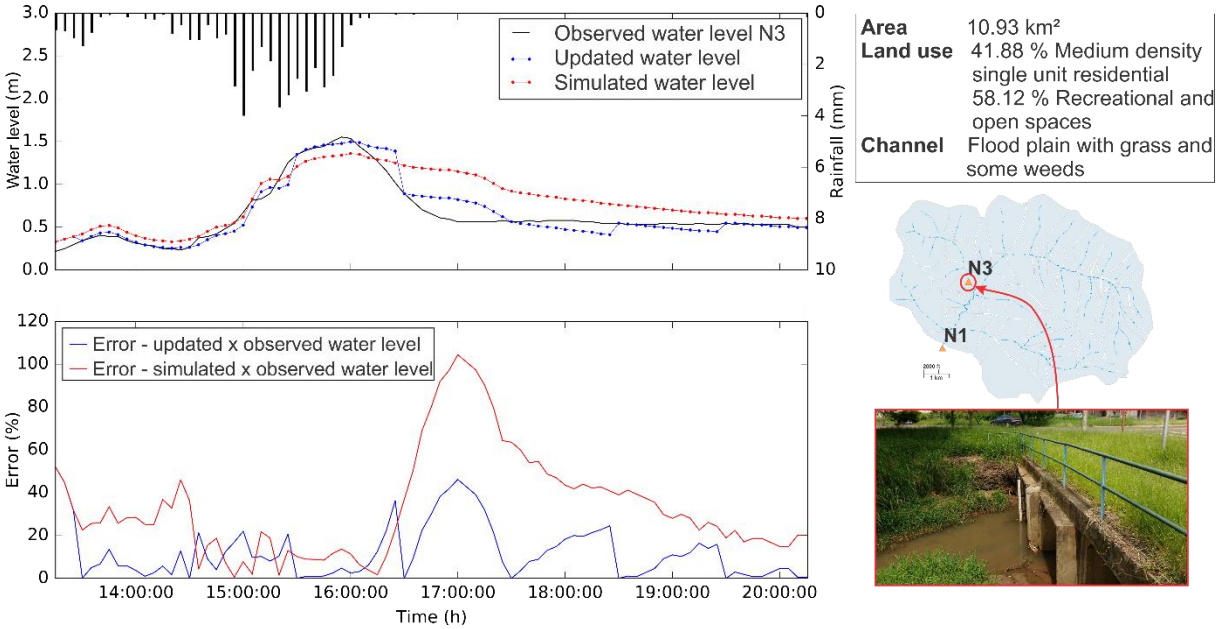


Figure 4-9 - Results of the updates made using the real-time estimator at location N3. Land use classification, according to Herold, Liu and Clarke (2003) and channel characteristics from Te Chow (1959).

Figure 4-8 shows the improvement in the forecasting assimilating field data mainly during a flood wave. The values simulated by the model declined slowly, and when the field data are entered, the forecasting followed the levels measured by the sensor in a better way. This location is in the outfall of the catchment and is a very critical point in strong rainfall events as several flood events are reported in this area. Despite a minor contribution area to location N-3 (Figure 4-9), it is a region of great importance to the study due to the flood history in the area and because it is an easily accessible location for citizen science data to be given, which is essential for the methodology proposed here. N-3 is located in a residential area and the section is under a bridge where many people circulate. The error graph shows that the deviations are more concentrated at the moments when the flood wave is decreasing. It is important to highlight that the priority is to detect the peaks more accurately, as they are the most likely times to characterise a flood disaster. NSE shows that for the two monitoring locations, the most accurate simulation values are obtained when the field data are incorporated (Table 4-1). With a time interval of one and two hours between the updates, NSE for location N-3 has raised from 0.457 to 0.464 and 0.715 respectively. At location N1 for one and two hours of a time interval between the updates, NSE raised from 0.784 to 0.894 and 0.878 respectively. However, the error charts point to a few moments when using field information

increased the error. Analysing these instants and comparing the error graph with the hydrograph, it can be observed that this happens when there is a fast water level variation and the reported field data at earlier times is still outdated.

4.4.2 Effects of uncertainty in observational data

The next figures present the results when updating the runoff simulations considering the possible reading errors in CS data and considering sparse intervals of updating. Figure 4-10 shows simulations updated with the samples that have randomly distributed errors of +/- 10 cm and respective error behaviour. It can be observed a reduction in per cent error in the two monitoring points with the insertion of field data during almost the whole rainfall event. At location N1, it can be observed that for information entered every hour or every two hours, the interquartile range displayed by the boxplot is quite narrow compared to the interquartile range of the same conditions in location N-3. This probably happened because location N1 is better calibrated. For all cases, a better fit to the peak values can be observed, even considering the extreme values of the boxplot, showing that even the error of reading volunteered data values can help to improve the water level simulations when errors of +/- 10 cm are considered.

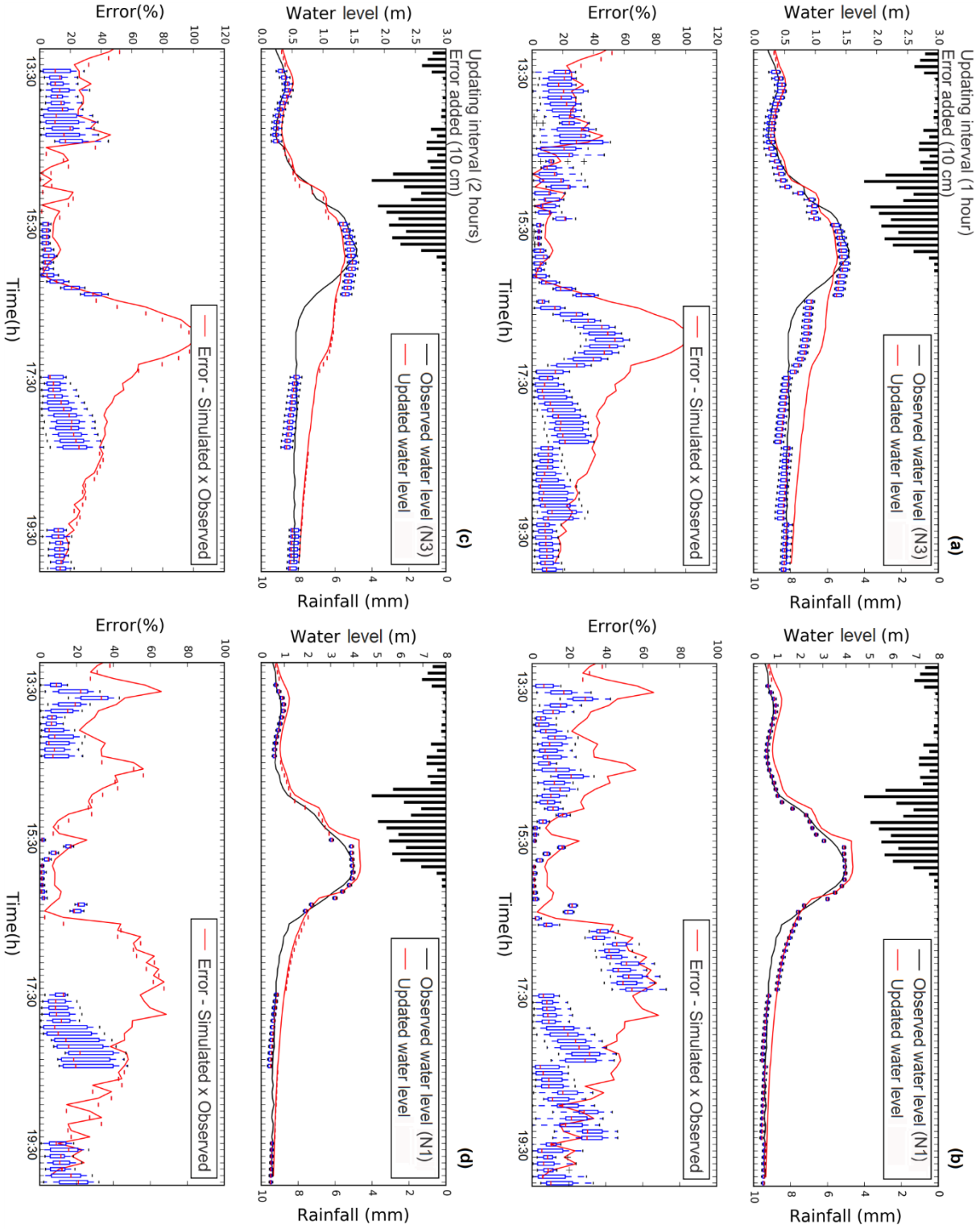


Figure 4-10 - The rainfall event illustrated is registered in São Carlos, São Paulo state on 04/11/2013. The graphs show the level simulated by SWMM, the level measured by a water level sensor, the water level corrected inserting hourly voluntary information (a) and (b), and inserting field data every two hours (c) and (d). The field data inserted has randomly distributed errors of +/- 10 cm.

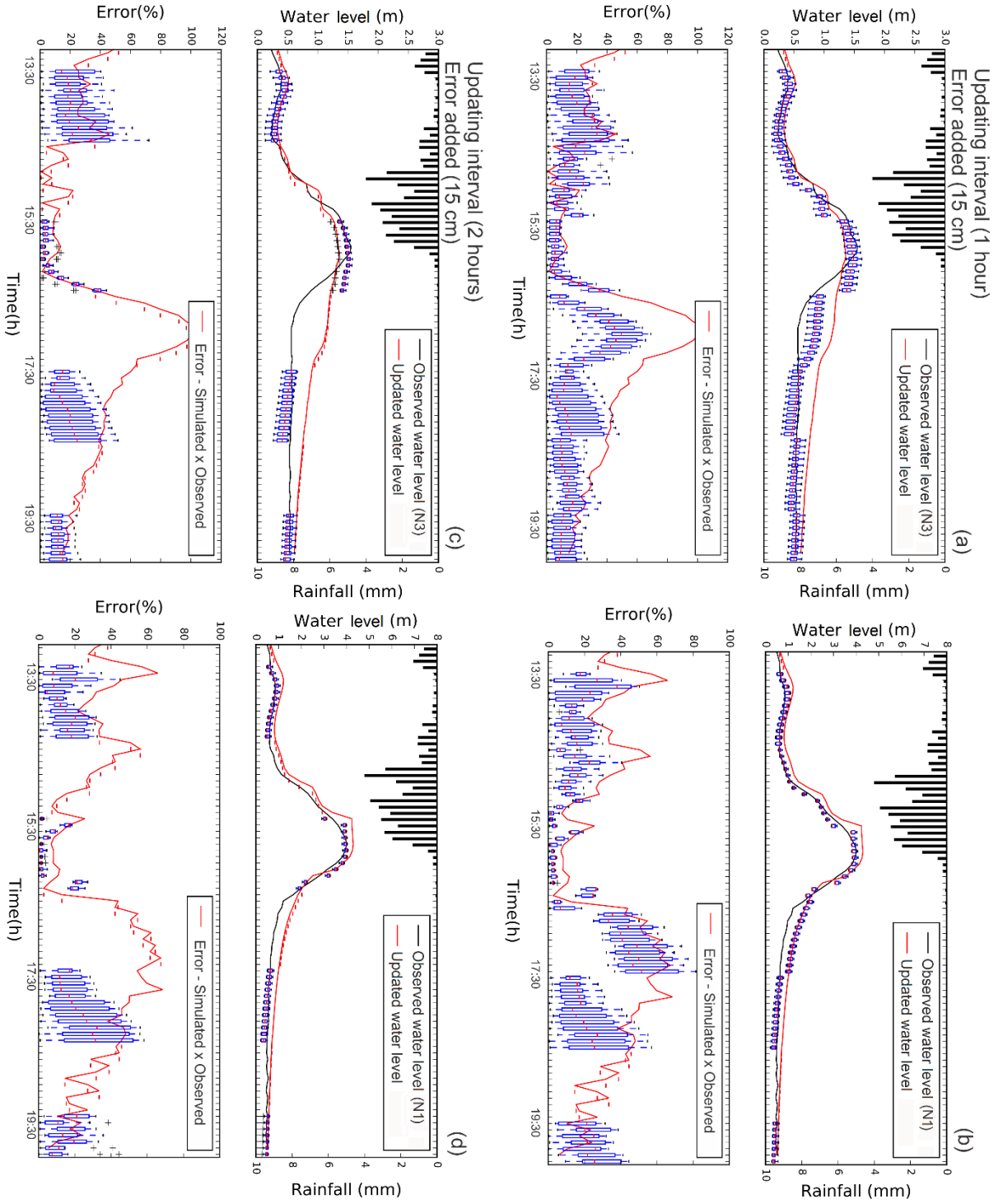


Figure 4-11 - The rainfall event illustrated is registered in São Carlos, São Paulo state on 04/11/2013. The graphs show the level simulated by the SWMM, the level measured by a water level sensor, the water level corrected inserting hourly voluntary information (a) and (b), and inserting voluntary information every two hours (c) and (d). The volunteered information inserted has randomly distributed errors of +/- 15 cm.

Figure 4-11 shows the results of the field data insertion considering +/- 15 cm of error for the same scenarios as Figure 4-10. An increase in the distance between the quartiles

compared to the graphs in Figure 4-10 can be observed. Some outliers can be noticed when the flood wave is decreasing at location N1. At location N3, the extreme upper value of the boxplot exceeds the error of the simulation at the beginning of the flood. Nonetheless, the NSE average still shows an improvement in simulations considering errors of +/- 15 at location N1 (average NSE values of 0.871 and 0.868 for one and two hours of time interval respectively). However, at location N3 when updating with two hours of a time interval and considering +/- 15 cm, the average NSE value decreases to 0.361.

Considering that the water level results are used to determine the level of risk awareness necessary for each moment in the hydrograph, and after evaluating the forecasting performance of the real-time estimator and the CS data uncertainty, it is analysed how it affects the alerts related to the Hazard Index. In Figure 4-12, the black and blue lines represent the variation of HI in time during the flood event for the water level measured by the sensor and simulated by the model, respectively. The coloured lines from S1 to S11 represent the HI for the simulated water levels and are corrected for the assimilation of synthetic volunteered data added with water errors between +/- 10 cm. The data assimilation interval is one hour. The graph shows that updating simulations, even considering the noise, reaches HI values closer to those observed.

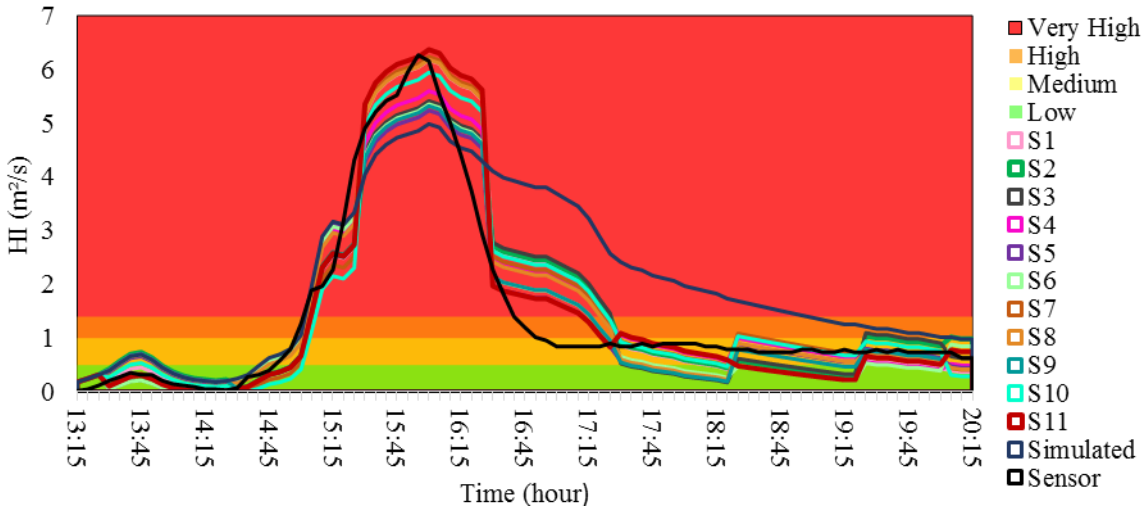


Figure 4-12 - Adherence verification between the forecast and the measured data about the Hazard Index (HI).

Through the simulation results considering the uncertainty in the CS data, the updates are assessed in different places than where the field information is given through the regionalisation method. Water level values at location N-3 are updated by correlation with location N1. It is performed for four situations (Figure 4-13): (a) Updating forecasting every

hour with field data added to a random error of +/- 10 cm; (b) Updating forecasting every hour with field data added to a random error of +/- 15 cm; (c) Updating forecasting every two hours with field data added to a random error of +/- 10 cm; (d) Updating forecasting every two hours added to a random error of +/- 15 cm. In all situations, the NSE ratio increased, showing a very significant result in the use of correlations between points of the same catchment. The best average NSE value obtained by correlation is 0.639 when updating every hour and using random error of +/-15 cm.

Table 4-1 - NSE results for calibration, validation, real-time estimator procedure and correlations for location N3 and location N1.

Location	**Purpose**	**NSE**
N3	Calibration	0.519
(A = 10.93 km²)	Validation	0.457
	Updated (time interval = 1h)	0.715
	Updated (time interval = 2h)	0.464
	Updated (time interval = 1h; error = 10 cm)	0.588
	Updated (time interval = 1h; error = 15 cm)	0.471
	Updated (time interval = 2h; error = 10 cm)	0.419
	Updated (time interval = 2h; error = 15 cm)	0.361
	Updated by correlation (time interval = 1h; error = 10 cm)	0.639
	Updated by correlation (time interval = 1h; error = 15 cm)	0.652
	Updated by correlation (time interval = 2h; error = 10 cm)	0.578
	Updated by correlation (time interval = 2h; error = 15 cm)	0.612
N1	Calibration	0.879
(A = 79.64 km²)	Validation	0.784
	Updated (time interval = 1h)	0.894
	Updated (time interval = 2h)	0.878
	Updated (time interval = 1h; error = 10 cm)	0.887
	Updated (time interval = 1h; error = 15 cm)	0.871
	Updated (time interval = 2h; error = 10 cm)	0.873
	Updated (time interval = 2h; error = 15 cm)	0.868

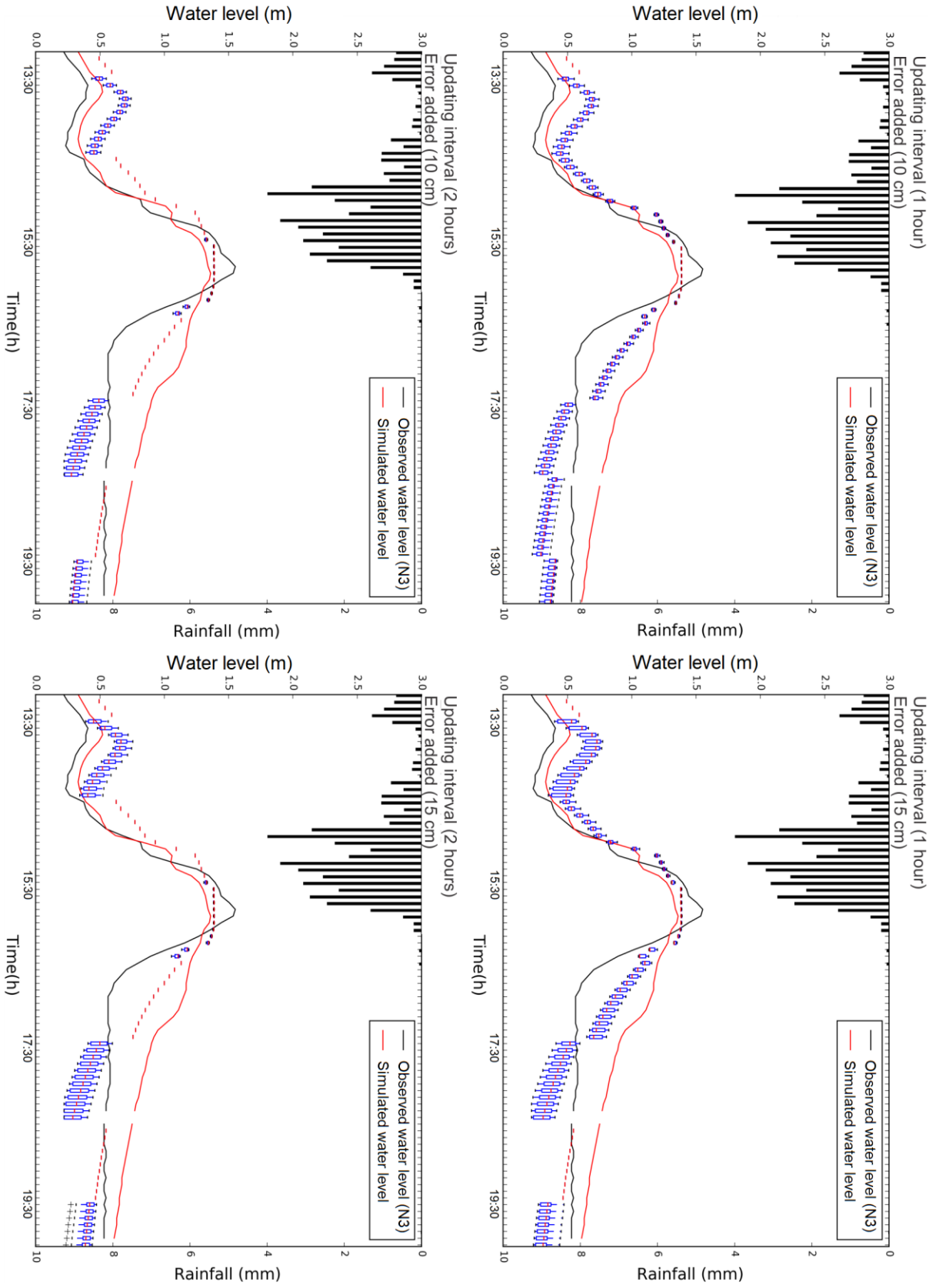


Figure 4-13 - Correction for location N-3 through the correlation with location N1 for day 04/11/2013. Volunteer data insertion: (a) Δt = 1h, error = +/- 10 cm; (b) Δt = 1h, error = +/- 15 cm; (c) Δt = 2h, error = +/- 10 cm; (d) Δt = 2h, error = +/- 15 cm.

4.5 Discussion

The literature review shows the trend of using the information provided by volunteers and developing Citizen Observatories in places exposed to flood risks in order to have an organised way of collecting data from the general public to improve flood risk management. There are few studies applying citizen science data for flood forecasting. However, many studies indicate that this may be an important methodology to be developed and tested for different flood risk scenarios, thus reinforcing the contribution of the HAMPB model.

Results using the real-time estimator to update outputs of the hydrological model with instant water level information helped to improve the accuracy of the simulations. However, these results are obtained using water-level data measured by sensors and added to a random error to mimic citizen science data. The behaviour and frequency of CS data are not addressed in these evaluations. The results are tested when assimilating data in a regular temporal distribution of one and two hours. Considering the model time step of 5 minutes, there is a significant gap between the updates. The previous experiences in the literature review about collecting citizen data show that the volunteered data can be provided more or even less frequently than what is considered here. It depends on which way it will be organised and by the level of commitment and engagement of the volunteers. Moreover, an important issue shown by the results presented is that the moment when receiving the information can be more crucial than only the frequency.

The case study presented here is performed in a particular situation where the catchment is almost entirely urbanised and has a short lead time. This scenario requires quick answers. However, as it is an urban area, it has the advantage of being able to collect data provided by volunteers easily. The proposed methodology is evaluated considering synthetic CS data with a high degree of reliability. Taking this into account, it should be noted that in order to use a voluntary data collection tool using the HAMPB model, people who are directly interested in the success of the flood forecasting model should be trained. The HAMPB model, as well as other rainfall-runoff models, require accurate data. Ideally, the government and responsible entities should organise incentives to provide training for civil defence professionals, residents who live in areas of risk and the general public in order to ensure the better-quality data collection.

4.6 Conclusions

This chapter describes a methodology to integrate citizen science data into a short-term forecasting model presenting an inexpensive solution to estimate the water level in ungauged or poorly monitored catchments. The real-time estimator updating procedure adding field data with random errors of +/- 10 and 15 cm showed strong evidence that the assimilation method, even though it is a simple method, can improve forecasting.

It can be concluded from the assessments that even in a considerable amount of time between the updates and considering the uncertainty in citizen science data, it can be helpful to improve the accuracy of forecasts. In future work, it is suggested evaluating the prediction model by updating simulations with water level data which come asynchronously and less frequently (e.g. daily, weekly, or monthly) to better represent the real citizen science observations, and also with different amounts of data being received simultaneously.

Regionalisation using data correlation between different sites in the catchment is an important step to make better use of the field information received. The results showed an important contribution to updating the forecast in catchments with several places at risk of flooding. It is recommended to improve the method by checking points that have the strongest correlations with others to update the simulation. It is also suggested analysing the parameters that influence the correlation quality between the points such as the size of sub-catchments, position, and distances. Further, studies in different catchments and with several rainfall-runoff events should be investigated to ensure the robustness of the HAMPB model.

Despite the limitations, the flood forecasting procedure proposed in this study is an innovative tool to emit alerts in ungauged catchments with scarce data. Therefore, hopefully, it will contribute to making cities more prepared and resilient to the occurrence of floods.

References

Assumpção, T. H., Popescu, I., Jonoski, A., & Solomatine, D. P. (2018). Citizen observations contributing to flood modelling: opportunities and challenges. *Hydrology and Earth System Sciences*, 22(2), 1473.

Buytaert, W., Zulkafli, Z., Grainger, S., Acosta, L., Alemie, T. C., Bastiaensen, J., ... & Foggin, M. (2014). Citizen science in hydrology and water resources: opportunities for knowledge generation, ecosystem service management, and sustainable development. *Frontiers in Earth Science*, 2, 26.

Coller, M. L. F., Wheeler, P., Kunapo, J., & Peterson, J. (2018). Interactive flood hazard visualisation in Adobe Flash. *Journal of Flood Risk Management*, 11, S134-S146.

Davids, J. C., van de Giesen, N., & Rutten, M. (2017). Continuity vs. the Crowd—Tradeoffs Between Continuous and Intermittent Citizen Hydrology Streamflow Observations. *Environmental Management*, *60*(1), 12-29.

Centro Universitário de Estudos e Pesquisas sobre Desastres (CEPED). (2013). *Atlas brasileiro de desastres naturais 1991 a 2010: volume Brasil*. Florianópolis: CEPED, UFSC.

Degrossi, L. C., de Albuquerque, J. P., Fava, M. C., & Mendiondo, E. M. (2014). Flood Citizen Observatory: a crowdsourcing-based approach for flood risk management in Brazil. In *SEKE* (pp. 570-575).

Estes, L. D., McRitchie, D., Choi, J., Debats, S., Evans, T., Guthe, W., ... & Caylor, K. K. (2016). A platform for crowdsourcing the creation of representative, accurate landcover maps. *Environmental Modelling & Software*, 80, 41-53.

Fritz, S., McCallum, I., Schill, C., Perger, C., See, L., Schepaschenko, D., ... & Obersteiner, M. (2012). Geo-Wiki: An online platform for improving global land cover. *Environmental Modelling & Software*, 31, 110-123.

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4), 211-221.

Heinzelman, J., & Waters, C. (2010). *Crowdsourcing crisis information in disaster-affected Haiti*. Washington, DC: US Institute of Peace.

Herold, M., Liu, X., & Clarke, K. C. (2003). Spatial metrics and image texture for mapping urban land use. *Photogrammetric Engineering & Remote Sensing*, *69*(9), 991-1001.

Horita, F. E., de Albuquerque, J. P., Degrossi, L. C., Mendiondo, E. M., & Ueyama, J. (2015). Development of a spatial decision support system for flood risk management in Brazil that combines volunteered geographic information with wireless sensor networks. *Computers & Geosciences*, 80, 84-94.

Hung, K. C., Kalantari, M., & Rajabifard, A. (2016). Methods for assessing the credibility of volunteered geographic information in flood response: A case study in Brisbane, Australia. *Applied Geography*, 68, 37-47.

Innocenti, D. (2014). When science meets policy: enhancing governance and management of disaster risks. *Hydrometeorological Hazards: Interfacing Science and Policy*, 91-107

Klonner, C., Marx, S., Usón, T., Porto de Albuquerque, J., & Höfle, B. (2016). Volunteered geographic information in natural hazard analysis: a systematic literature review

of current approaches with a focus on preparedness and mitigation. *ISPRS International Journal of Geo-Information*, 5(7), 103.

Lee, H., Seo, D. J., Liu, Y., Koren, V., McKee, P., & Corby, R. (2012). Variational assimilation of streamflow into operational distributed hydrologic models: effect of spatiotemporal scale of adjustment. *Hydrology and Earth System Sciences*, 16(7), 2233.

Leyh, W., Fava, M. C., Abe, N., Cavalcante, S., Giatti, L., de Carvalho, C. M., Fonseca-Filho, H. & Jacobs, C. (2017). Citizen Science Involving Collections of Standardized Community Data. In *International Conference on Applied Human Factors and Ergonomics* (pp. 410-420). Springer, Cham.

Mazumdar, S., Lanfranchi, V., Ireson, N., Wrigley, S., Bagnasco, C., Wehn, U., ... & Ciravegna, F. (2016). Citizens observatories for effective Earth observations: the WeSenseIt approach. *Environmental Scientist*, 25(2), 56-61.

Mazzoleni, M., Verlaan, M., Alfonso, L., Monego, M., Norbiato, D., Ferri, M., & Solomatine, D. P. (2017). Can assimilation of crowdsourced data in hydrological modelling improve flood prediction?. *Hydrology and Earth System Sciences*, 21(2), 839.

Miorandi, D., I. Carreras, E. Gregori, I. Graham, and Stewart, J. (2013). Measuring net neutrality in mobile internet: Towards a crowdsensing-based citizen observatory, in *Communications Workshops (ICC), 2013 IEEE International Conference on*, pp. 199–203, IEEE.

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

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology*, 10(3), 282-290.

Pánek, J., Marek, L., Pászto, V., & Valůch, J. (2017). The Crisis Map of the Czech Republic: the nationwide deployment of an Ushahidi application for disasters. *Disasters*, 41(4), 649-671.

Poser, K., & Dransch, D. (2010). Volunteered geographic information for disaster management with application to rapid flood damage estimation. *Geomatica*, 64(1), 89-98.

Ramaswamy, B. (2016). An Intelligent Wireless Modular System for Effective Disaster Management. *Transactions on Networks and Communications*, 4(3), 22.

Reichle, R. H. (2008). Data assimilation methods in the Earth sciences. *Advances in water resources*, 31(11), 1411-1418.

Rossman, L. A. (2010). *Storm water management model user's manual, version 5.0* (p. 276). Cincinnati: National Risk Management Research Laboratory, Office of Research and Development, US Environmental Protection Agency.

Rotava, J., Mendiondo, E. M., & Souza, V. C. B. (2013). Simulação de instabilidade humana em inundações: primeiras considerações. *XX Simpósio Brasileiro de Recursos Hídricos*, 1-8.

Roy, H. E., Pocock, M. J., Preston, C. D., Roy, D. B., Savage, J., Tweddle, J. C., & Robinson, L. D. (2012). Understanding citizen science and environmental monitoring: final report on behalf of UK Environmental Observation Framework.

Seebauer, S., & Babcicky, P. (2017). Trust and the communication of flood risks: Comparing the roles of local governments, volunteers in emergency services and neighbors. *Journal of Flood Risk Management*, 11(3), 305-316.

Smith, L., Liang, Q., James, P., & Lin, W. (2015). Assessing the utility of social media as a data source for flood risk management using a real-time modelling framework. *Journal of Flood Risk Management*, 10(3), 370-380.

Te Chow, V. (1959). *Open channel hydraulics*. McGraw-Hill Book Company, Inc; New York.

Tingsanchali, T., & Karim, M. F. (2005). Flood hazard and risk analysis in the southwest region of Bangladesh. *Hydrological Processes*, 19(10), 2055-2069.

Victorino, J. N. C., & Estuar, M. R. J. E. (2014, October). Profiling flood risk through crowdsourced flood level reports. In *IT Convergence and Security (ICITCS), 2014 International Conference* on (pp. 1-4). IEEE.

Voinov, A., & Bousquet, F. (2010). Modelling with stakeholders. *Environmental Modelling & Software*, 25(11), 1268-1281.

Voinov, A., Kolagani, N., McCall, M. K., Glynn, P. D., Kragt, M. E., Ostermann, F. O., ... & Ramu, P. (2016). Modelling with stakeholders–next generation. *Environmental Modelling & Software*, 77, 196-220.

Walker, J. P., Willgoose, G. R., & Kalma, J. D. (2001). One-dimensional soil moisture profile retrieval by assimilation of near-surface observations: a comparison of retrieval algorithms. *Advances in Water Resources*, 24(6), 631-650.

Zook, M., Graham, M., Shelton, T., & Gorman, S. (2010). Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake. *World Medical & Health Policy*, 2(2), 7-33.

5 Improving Flood Forecasting using an Input Correction Method on Urban Models in Poorly Gauged Areas*

*A modified version of this chapter has been accepted for publication as: Fava, M. C., Mazzoleni, M., Abe, N., Mendiondo, E. M., Solomatine, D. P. (2020) Improving Flood Forecasting using an Input Correction Method on Urban Models in Poorly Gauged Areas. *Hydrological Sciences Journal.* In press.

Abstract

Poorly monitored catchments could pose a challenge of providing accurate flood predictions by hydrological models, especially in urbanised areas subject to heavy rainfall events. Data assimilation techniques have been widely used in hydraulic and hydrological models for model updating (typically updating model states) to provide a more reliable prediction. However, such procedures in the case of non-linear systems are quite complex and time-consuming, making them not very suited for real-time forecasting. In this study, a data assimilation procedure is presented, which corrects the uncertain inputs (rainfalls), rather than states, of an urban catchment model by assimilating water level data. Five rainfall correction methods are proposed, and their effectiveness is explored in different scenarios when assimilating data from one or multiple sensors. The methodology is adopted in the city of São Carlos, Brazil. The results show a significant improvement in the simulation accuracy when assimilating data by all methods and scenarios.

Keywords: data assimilation; semi-distributed model; flood modelling; physically-based model, SWMM.

5.1 Introduction and scope

Intense urban growth without proper management and adequate drainage systems, along with the effects of increasing weather variability and climate change, are the main causes of urban flood aggravation in cities, especially in small catchments (Tollan, 2002; Ashley *et al*., 2005; Bai *et al*., 2018). Additionally, the measurements of rainfall in many cities are scattered and not very accurate, and this is a major challenge for modelling and predicting floods in a timely manner for decision makers.

In response, many research efforts have investigated how to improve flood forecasting. One of the approaches is data assimilation (DA) techniques, which have become widely used to improve hydrological predictions, updating the model as a response to real-time observations (Young, 2002; Hutton *et al*., 2012; He *et al.*, 2017; Mazzoleni *et al.*, 2018a). The main idea of most data assimilation methods is to quantify errors in field data and in simulations to update the hydrological states optimally, ensuring minimisation of simulation error (Collier, 2007; Coustau *et al*., 2013; McMillan *et al*., 2013; Thiboult & Anctil, 2015). The updates can be made in the inputs, parameters and states of hydrometeorological models (Refsgaard, 1997; Liu & Gupta, 2007; Seo *et al*., 2009).

Kalman Filter (KF) is the most widely known DA method; however, it is optimal only for linear processes (Maybeck, 1982; Walker & Houser, 2005). To account for non-linear systems, several variations such as the Particle Filter (Moradkhani *et al*., 2005; Andrieu *et al*., 2010; Moradkhani *et al*., 2012), Extended Kalman Filter (Francois *et al*., 2003) and the Ensemble Kalman Filter (Reichle *et al*., 2002; McMillan *et al*., 2011; Mazzoleni *et al.*, 2018b) have been proposed; the latter is the most used technique in Earth sciences (Evensen, 2003; Chen *et al*., 2013). Although these methods are designed for nonlinear models, they use linearization during the update process (Liu *et al*., 2012).

A number of authors explored the use of data assimilation methods to update urban models. For instance, Hutton *et al*. (2014) developed a methodology employing a deterministic Kalman Filter to update the states of an urban model as a response to downstream observations. The results show improvements in the discharge forecasts. However, the presence of threshold system behaviour in controlled urban systems affects the DA procedure ability, resulting in de-coupling of the downstream catchment response to changes in upstream behaviour. Hansen *et al*. (2014) used DA for the distributed hydrodynamic urban drainage model MIKE URBAN.

The updating process is carried out in a deterministic manner and demonstrated improvements in the simulations, despite the fact that the uncertainties of the model structure and observational data are not considered. Borup *et al*. (2018) tested the use of Ensemble Kalman filter for the MIKE URBAN model in an experiment to evaluate the possibility of using constant Kalman gain updating to address the problem of high computational demand when the ensemble is calculated in real-time. The results show that the gain is nonlinear and varies greatly in time, requiring the use of the complete Ensemble Kalman Filter scheme.

In fact, just a few studies have investigated the use of DA approaches for real-time monitoring in small urban catchments subject to flash floods (Xie & Zhang, 2010; Chen *et al*., 2013). In contrast to large basins or medium-scale basins, small urban catchments may be affected by intense local precipitation (WMO, 2011), which combined with increased runoff by urbanisation, require faster responses from hydrological models to ensure their utility for flood forecasting (Yang *et al*., 2011; Chen *et al*., 2013; Yin *et al*., 2016). Furthermore, urban models are highly non-linear with many physical state variables and, consequently, computationally costly, posing a further challenge for their real-time updating (Hansen *et al*., 2014; Borup *et al*., 2018).

Rainfall estimation has a fundamental role to perform streamflow simulations in hydrological models. This is especially vivid when the quantity and intensity of rainfall vary depending on previous conditions (Harader *et al*., 2012), in urban areas where runoff is usually extremely sensitive to spatial distribution of rainfall, and the flood responses are caused by spatially localised convective precipitation (Zawadzki, 1973; Segond *et al*., 2007). Furthermore, a study performed by McMillan *et al.* (2011) shows that for heavy rainfall events, hydrological models have high multiplicative errors. The result is that uncertainty in rainfall measurement and forecasting leads to errors in flow forecasts, even if real-time hydrological models are well-calibrated (Pedersen *et al*., 2016). An option could be the use of DA schemes updating the rainfall inputs as a response to observed water levels or flow data, but only a few studies have proposed and explored this possibility (Kahl *et al.*, 2008; Stanzel *et al*., 2008; Divac *et al*., 2009; Harader *et al*., 2012).

Analysis of literature and the known practice allows concluding that data assimilation in physically-based models for catchments with short lead-time, and the inaccuracies in rainfall inputs, are important issues to be investigated for flood forecasting and monitoring in urban systems. Therefore, the main objectives of this study are: (a) to analyse the hydrological model

sensitivity to input variation (precipitation); (b) to propose a variant of DA methodology that corrects the model input using observed water levels from in-situ physical sensors; (c) to assess the effect of the number and location of the water level sensors (used to correct model input) on model accuracy. The methodology is used in the case study in the Monjolinho catchment, located in the city of São Carlos, São Paulo State, Brazil, modelled by the Storm Water Management Model (SWMM) (Rossman, 2010).

5.2 Methodology

5.2.1 Hydrological urban modelling

The Storm Water Management Model (SWMM) is a modelling system with the integrated hydrological and hydraulic modelling modules, which is capable of simulating single events or long-term simulations of runoff in urban catchments with pipe and open flow networks (Rossman, 2010). The hydrological module simulates the sub-catchments behaviour, including an internal infiltration module. The rainfall-runoff component is lumped and conceptual at the sub-catchment scale. A sub-catchment can be defined by the user and is divided into pervious and impervious portions. Each part is modelled as a nonlinear reservoir with a capacity given to the maximum depression storage. The hydraulic module propagates the surface runoff, coming from the hydrological model, along rivers, canals and other conduits.

The parameters of the hydrological model are width, average slope, infiltration parameters, Manning's coefficient for pervious and impervious area, per cent of impervious area with no depression storage, depth of depression storage on pervious and impervious areas and the percentage of the impervious area. For the hydraulic module, the main parameters are the canal roughnesses.

An automatic calibration tool for SWMM using Genetic Algorithms (GA) as the optimisation technique has been developed; it is written in Python 3.5. The Distributed Evolutionary Algorithms in Python (DEAP) was used as a framework for the GA optimisation (Fortin *et al*., 2012) and the SWMM5 library for SWMM calling interface developed by Pathirana (2015).

In this study, SWMM model is calibrated using water level values instead of flow data, as the rating curve is not available. This decision is based on the study by Lindström (2016)

who showed good results for a hydrological model calibration using observed water level data instead of observed discharge or establishing a rating curve. The study also proposes efficiency evaluation equations for data from multiple water level stations, such as the Spatial Nash-Sutcliffe efficiency (*SPATNSE*) (Equation (5-1)). The maximisation of the *SPATNSE* is set as the objective function in the calibration tool developed.

$$SPATNSE = 1 - \frac{\sum_{j=1}^{K}\left(\overline{H}_j^* - \overline{H}_j\right)^2}{\sum_{j=1}^{K}\left(\overline{H}_j - \frac{1}{K}\sum_{j=1}^{K}\overline{H}_j\right)^2} \tag{5-1}$$

$$\overline{H}_j^* = \left(H^*{}_{j_i} : i = [1, x_j]\right) = \frac{1}{x_j}\sum_i H^*{}_{j_i} \tag{5-2}$$

$$\overline{H}_j = \left(H_{j_i} : i = [1, x_j]\right) = \frac{1}{x_j}\sum_i H_{j_i} \tag{5-3}$$

where: $H_{j_i}^*$ are the simulated values for the j-th station at time step i, $[1, K] = \{j \in \mathbb{N} : 1 \leq j \leq K\}$; H_{j_i} are the observed values for the j-th station at time step i; $\overline{H}_j^*$ is the average value of H_j^* for the j-th station; $\overline{H}_j$ is the average value of H_j for the j-th station; x_j is the number of values in the time series for station j; i is the index of time steps with observations in a time series of a station; j is the index of stations; K is the total number of stations. The simulated and observed variables are water levels at sensors locations. The range of SPATNSE is in between $-\infty$ and 1.

5.2.2 Data assimilation framework

The relationship between the state variables of the SWMM model and observations is not linear, which means the observations cannot be mapped directly in the state space. In addition, to speedily provide results for real-time operation for urban basins subject to flash floods, the assimilation procedure must provide reliable results with the short computational time. To account for that, an approach to correct the model inputs as a response to hourly real-time observed water level data from different in-situ sensors is proposed. The method is

designed to dynamically assimilating water level information from one to multiple sensors simultaneously. Figure 5-1 outlines the general methodology. The proposed method does not take into account the uncertainty of the model parameters and water level observations.

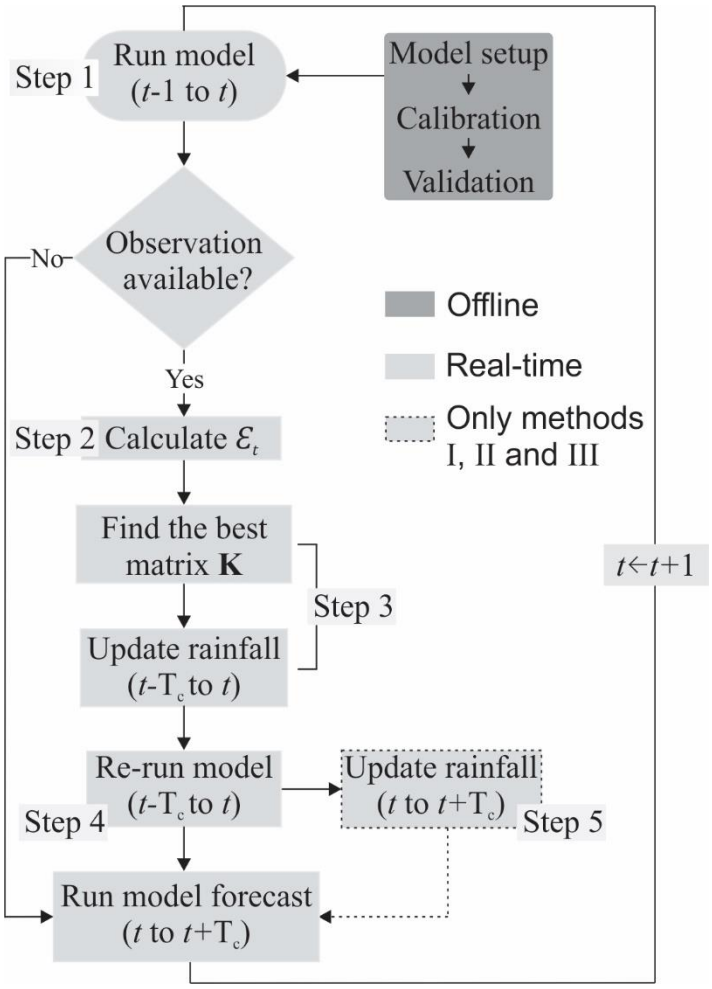


Figure 5-1 - Flowchart of the data assimilation methodology.

In the first step, SWMM simulates with the observed rainfall values as input until time step t, which is the moment when the water level observation H_t becomes available. In the second step, H_t is compared to the corresponding simulated water level (H_t^*) for the same location and the squared error between the model simulation and the observation is calculated (Figure 5-2.a). It is worth noting that there is a possibility of receiving observational data from more than one location simultaneously. In such a case, the error ε_t for receiving observed water level at time t, in one or more location is calculated by averaging all errors:

$$\varepsilon_t = \frac{1}{N}\sum_{j=1}^{N}\left(H_{t,j}^{*} - H_{t,j}\right)^2 \tag{5.4}$$

where: N is the number of locations receiving information at time t, $H_{t,j}$ [m] is the observed water level at j-th node at time t and $H_{t,j}^{*}$ [m] is the simulated value at time t.

Seo *et al*. (2003) pointed out that the assimilation window must be equivalent to the time of response of the basin so that the memory of the system is reflected in the process of assimilation. In the third step, the time of concentration (T_c) of the catchment is used as the time window for correcting the model input, hence it is assumed that the rainfall which occurred before T_c does not affect the river flow at time t. Rainfall values present in the moving assimilation window for all the sub-catchments are corrected by the inner product of the measured rainfall and a matrix **K** of coefficients (Equation (5-5)). The values of the matrix corresponding to the time step t are estimated in a quasi-optimal way to reduce the error calculated in step 2, and then they are propagated through the assimilation window by different methods. The details of the optimisation process and the methods to update rainfall are explained in the next sections. The corrected precipitation of the sub-catchments $\mathbf{P}^{+}$ is a matrix with the same dimensions of the matrix **K** and the rainfall **P** measured in the sub-catchments.

$$\mathbf{K} = \left(k_{ij}\right) \in \mathbb{R}^{m\times n} \tag{5-5}$$

$$\mathbf{P} = \left(p_{ij}\right) \in \mathbb{R}^{m\times n} \tag{5-6}$$

$$\mathbf{P}^{+} \leftarrow \langle \mathbf{K}, \mathbf{P} \rangle \in \mathbb{R}^{m\times n} \tag{5-7}$$

Matrix **P** and matrix $\mathbf{P}^{+}$contain the original rainfall and the corrected rainfall values of all sub-catchments, respectively (Equation 5-5 and 5-6). The matrix **K**, matrix **P** and matrix $\mathbf{P}^{+}$comprise n arrays of coefficients containing m elements each, where n is the number of sub-catchments and m is defined as T_c divided by the default time step. The change in rainfall is carried out with the aim to obtain a better simulation not only at the location where the field information is received but since it is a semi-distributed model, also throughout the basin.

In the fourth step, the model will re-run from time t - T_c to time t with the updated rainfall values. Finally, in the fifth step, the quasi-optimal coefficients used to change rainfall backwards will also be applied to correct precipitation forwards from time step t until $t + T_c$.

5.2.3 Optimisation method

The real-time DA methodology proposed in this study focuses on urban basins with a rapid response, requiring model simulations with short time steps. The quasi-optimum values for the DA coefficients in **K** need to be found within one such time step, so settings for the data assimilation and optimisation algorithms have to take the simulation time constraints into account.

The Objective Function (OF) to be minimised is the error between observed and simulated water level when receiving new field data at one or more nodes in the catchment (Equation (5-2)). The decision variables to be identified are the coefficients that change the precipitation values of each sub-catchment at the time step t, which is the moment when receiving field data. The first row of the matrix **K** corresponds to these coefficients (i = 1). Therefore, it is used the coefficients vector for the first line of the matrix **K**., i.e. for an observation received at time t, $k_j \leftarrow K_{1,j}$.

A randomised search by canonical Genetic Algorithm (GA) is used with the following parameters: two-point crossover, bit-flip mutation algorithm and selection by tournament, 50% crossover probability, 2% mutation probability and initial population size of 100 individuals. The GA evaluates the OF for each new vector of decision variables (minimising the average water level error at time t by correcting coefficients, and hence, changed rainfall inputs), and this is done by running SWMM from $t - T_c$ until t, and recalculating ε_t with the new rainfall inputs. Optimisation continues until the found solution cannot be improved (or considered acceptable), or until it reaches the 15-minute time limit. An absolute error of 5 cm between the observed and simulated water level is adopted as an acceptable error. In addition, the decision variables are constrained to limit the amount allowed to change rainfall – between –60% to +60% of the original data. This is done to preserve the characteristics of the rainfall, even considering the uncertainties in its estimation. Based on the previous, the optimisation problem is formulated as follows:

$$\text{OF: } \min \varepsilon_t \tag{5-8}$$

Subject to:

$$-0.4 \leq k_j \leq 1.6 : \; j \in [1, n]$$

5.2.4 Model input correction

Five different rainfall correction methods are proposed. The rainfall values are corrected backwards in time using the *n* coefficients vectors of size *m* optimised from *t* until $t - \mathrm{T_c}$. In the first three methods (indicated as I, II and III in Figure 5-2.b), after finding the quasi-optimal coefficients in the steps going backwards from time step *t*, the rainfall values are corrected with these same values until $t + \mathrm{T_c}$. The other two methods (indicated with IV and V in Figure 5-2.b) modify precipitation values only going backwards in time. The influence of methods IV and V for flood prediction is limited to the correction effect of rainfall only up to the time $\mathrm{T_c}$. The methods are detailed below:

I) Each sub-catchment has a vector k_j, which decreases according to a linear function (Equation (5-9)), the k_j value is the maximum at time *t* and decreases linearly until $t - \mathrm{T_c}$, where its value is null. Thus, the change in the rainfall is highest close to the present moment (when receiving information) and progressively decreases to zero going backwards in time. The quasi-optimal coefficient values are used to propagate the rainfall correction also into the future (Figure 5-2.b – I), with the same scheme, but going forward in time, with the linearly progressively decreasing values (Equation (5-10)).

$$k_i \leftarrow K_{ij} \text{ for each } j : j \in [1, n]$$

$$k_i = \frac{(-k_j + 1)}{x} i + k_j : \; i \in [1, m] \tag{5-9}$$

$$k_i^* = \frac{(k_j - 1)}{x} i + 1 : \quad i \in [1, m] \tag{5-10}$$

$$x = \left\lfloor \frac{t - T_c}{s} \right\rfloor$$

where: k_i is the *i*-th coefficient in the changing rainfall vector backwards, k_i^* is the *i*-th coefficient in the changing rainfall vector forward, m is the number of elements in the rainfall input in the window of time between t and $\mathrm{T_c}$, and s is the time step size.

II) The rainfall is corrected by the k_j values going backwards and forwards in time, from $t - \mathrm{T_c}$ until $t + \mathrm{T_c}$. However, in method II the coefficient values remain constant, i.e. the same multiplier value is used for all time steps (Figure 5-2.b – II).

III) In that method, the rainfall is corrected going both backwards and forwards in time. When going backwards, the coefficient values remain constant, and when going forward, they decrease linearly (Equation (5-9), Figure 5-2.b – III).

IV) This method uses updates similar to the method I, from *t* back to $t - \mathrm{T_c}$ the coefficients decrease linearly (Equation (5-9), Figure 5-2.b – IV), but do not update forward values.

V) Similarly, to method II, the coefficients k_j remain constant for backward updating from *t* back to $t - \mathrm{T_c}$, but do not update forward values (Figure 5-2.b – V).

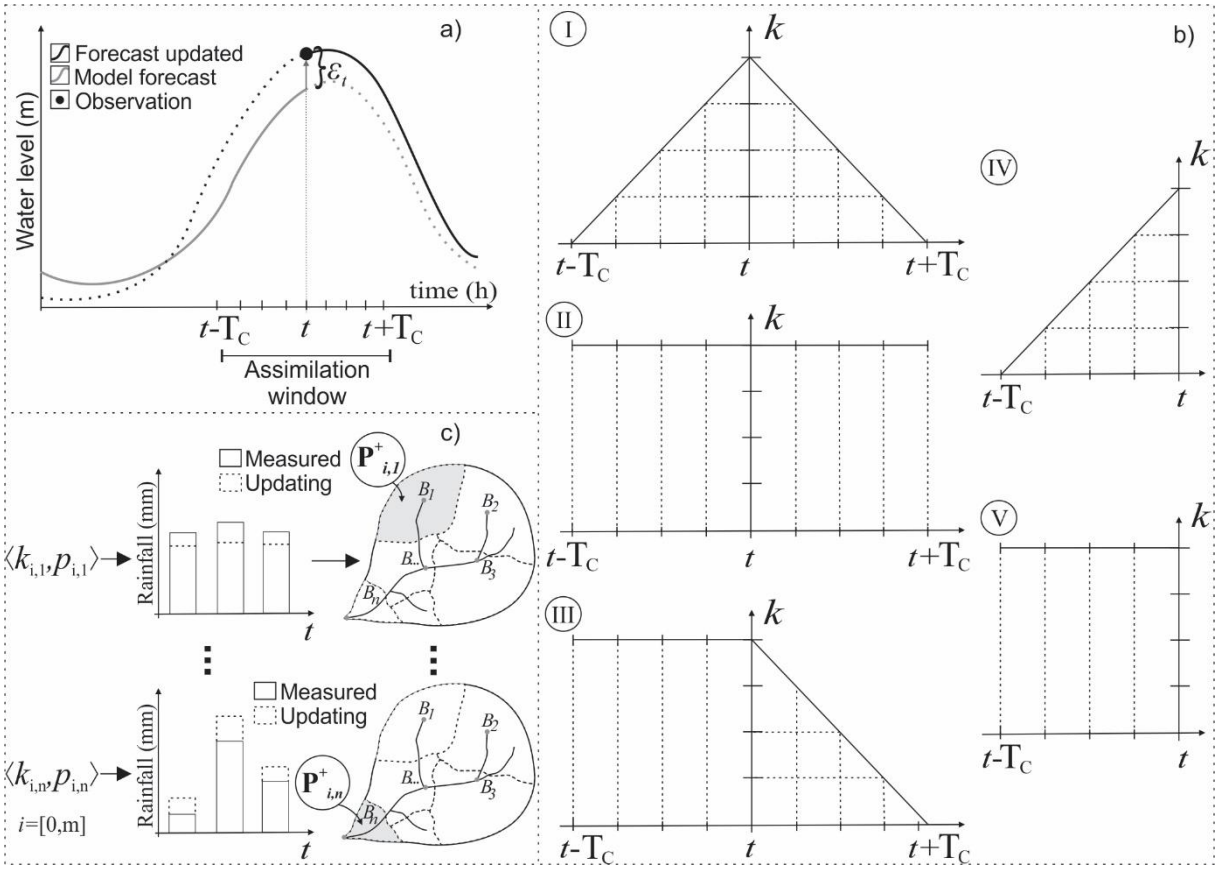


Figure 5-2 - Data assimilation with input rainfall correction: (a) calculating the error $\boldsymbol{\varepsilon_t}$when receiving field data; (b) after calculating the error, rainfall will be updated by one of the methods (I, II, III, IV or V); and (c) rainfall correcting process for each sub-catchment. Precipitation data is updated by the inner product of its original values by the optimised coefficients contained in the matrix **K** (Equation (5-7)).

5.2.5 Sensitivity analysis

Taking into account that (i) the model used is semi-distributed allowing different rainfall values for each sub-catchment, and (ii) the assimilation method aims to identify the (different) correcting coefficients, it is important to know the relative contribution of each sub-catchment to runoff generation when varying the model input. Therefore, the model sensitivity analysis with rainfall changing for each sub-catchment is performed. The measured precipitation data from each sub-catchment is changed, one at a time and the effects on the runoff simulation at the outlet are quantified. The data assimilation methods optimise the values considering the coefficients of variation limited between -60% and + 60%. These same values are used in the sensitivity analysis to vary the original amounts of rainfall, with regular intervals of 30%. Finally, the model output changes corresponding to rainfall variation of each sub-catchment

(indicating the sensitivity) are compared to the correcting coefficients identified by optimisation to correct the simulations. It is expected that in relative terms, there should be a certain agreement between the two, showing that the search method followed the physical meaning of rainfall changing. Otherwise, it will indicate that the optimisation is also correcting structural problems of the model.

5.2.6 Evaluation criteria

To evaluate the improvement of the data assimilation approach in the model runs concerning the observed water level values, the Nash-Sutcliffe efficiency (NSE) index is used (Nash and Sutcliffe, 1970). The method is a statistical evaluation widely used in hydrology, which compare the simulated and observed values as following:

$$NSE = 1 - \frac{\sum_{i=1}^{N}(H_i - H_i^*)^2}{\sum_{i=1}^{N}(H_i - \bar{H}_i)^2} \quad (5\text{-}11)$$

$$F1 = \max(NSE) \quad (5\text{-}12)$$

where: H_i is the observed water level at i-th time step; H_i^* is the simulated water level at i-th time step. N is the total number of water level observations. The range of NSE is in between $-\infty$ and 1. An index of 1 represents a perfect model, and values lower than zero indicates that the mean of the observed values would give a better prediction than the model.

5.3 Case study and data sets

5.3.1 Monjolinho catchment and its flood history

The case study is located in the Monjolinho catchment (Figure 5-3), in São Carlos, state of São Paulo, Brazil. The total area of the catchment is 79.6 square kilometres (km²), the population density of the city is 194,53 inhabitants/km² and a total population estimated at 249,415 inhabitants (IBGE, 2018). The average altitude is about 856 meters above sea level, and the soil is highly permeable. The catchment has its springs within the municipality of São Carlos to the east, almost entirely in the urban area. The city has a long history of flooding. Mendes and Mendiondo (2006) estimated, based on historical data from 1940 to 2004, that the

average time of the return of flood occurrence is 0.65 events/year with a standard deviation of 0.84. The study also showed that the cumulative number of flood occurrences increased with the catchment urbanisation. This historical data suggests that there is a clear need to adopt structural and non-structural means for the containment of floods, but the proposals to solve the problem recorded in official documents are very onerous and have not yet been adopted. (Barros *et al*., 2007). Despite the high frequency of floods, there is neither systematised flood data nor early warning system for São Carlos.

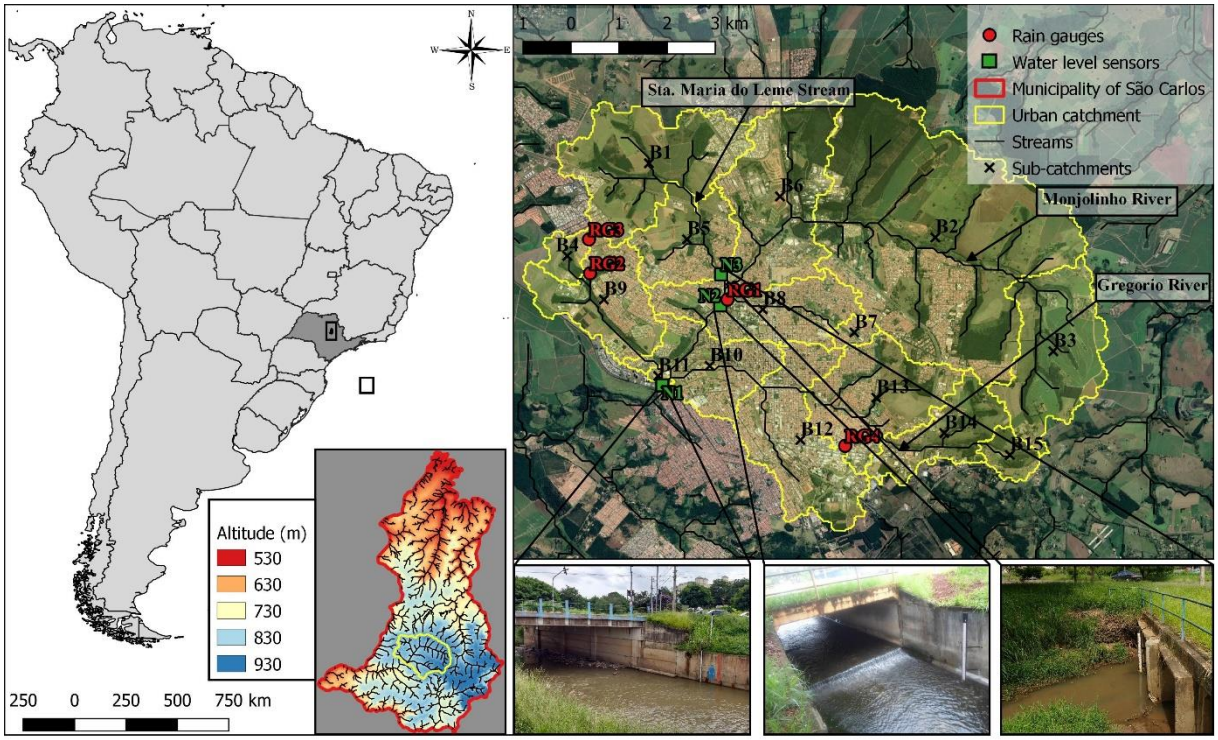


Figure 5-3 - – Monjolinho catchment in the city of São Carlos, São Paulo State, Brazil. The photographs show the three water level monitoring locations N1, N2 and N3.

5.3.2 Data sets

5.3.2.1 Water level data

Data from three monitoring locations (N1, N2 and N3) collecting water level data every fifteen minutes (Figure 5-3) are used. Sensor N1 is installed in the outlet of the catchment delimited for this study and in the Monjolinho River. It is a constructed canal without margin conservation. The riverbank has just two meters of grass coverage. The area receives a significant contribution of surface runoff from the drainage on the left side of the catchment along the Gregorio River, and frequently fluvial floods happen there. The second section

(sensor N2) is in the Monjolinho River after receiving the contribution from Santa Maria do Leme Stream. At this location, the channel is constructed with cement; this is the case also in an urbanised area of the city and with around two meters of grass followed by the asphalt. Sensor N2 is installed upstream of sensor N1, and its area does not suffer from flooding. Sensor N3 is installed in a section of Santa Maria do Leme Stream near the junction with Monjolinho River and upstream of the other monitoring locations. The river has a natural bed, which is quite silted at this location due to the lack of native vegetation and protected area on the banks of the stream. In this section, the environment consists of approximately two meters of grass, just after it there is the asphalt and a region of intense urbanisation. Because of the described characteristics, canals overflows are often observed.

5.3.2.2 Rainfall data

The input data come from four rain gauges shown in Figure 5-3 (RG 1 to RG 4), with 15 minutes of temporal discretisation. To generate the spatially distributed rainfall field, the Inverse Distance Weighting (IDW) method is used to interpolate data from rain gauge stations and estimate the mean spatial value for each sub-catchment.

5.3.3 Model setup

The basin is delineated with 15 sub-catchments. In hydrological modelling, the infiltration model based on the SCS Curve Number is chosen. This method is an approach adapted for the curve number (CN) of the NRCS (National Resources Conservation Service) to estimate the runoff (Cronshey, 1986). The selected flow propagation model is based on the dynamic wave equations. This infiltration method is an approach adapted for the curve number (CN) of the NRCS (National Resources Conservation Service) to estimate the runoff (Cronshey 1986). Originally, the method combines losses due to interception, depression storage and infiltration to predict the total rainfall excess. In SWMM, the CN method is used to compute only the infiltration losses. The surface runoff is calculated using a nonlinear reservoir model and applying the Manning's equation to calculate the flow rate. The depression storage is not linked to the infiltration method, giving more freedom for the modeller to choose the value for this parameter. The moisture retention capacity of the soil (S) reduces during wet periods by the infiltration volume, and increases during the dry period by a first-order recovery model. Then, the current value of S is used when a next rainfall event starts allowing the model to

calculate infiltration as a function of cumulative rainfall volume. More details about the modified CN method can be found in Rossman and Huber (2016).

Eight rainfall events between November 2013 and April 2014 are used for calibration and validation. Eight rainfall events between November 2013 and April 2014 are used for calibration and validation. The events are split into three subsets according to the average intensity rainfall: light, moderate and heavy. Then, the events are chosen at random for two groups: one for calibration and another for validation – with the restriction of at least one event of each subset for each group. The GA parameters used for the calibration are a two-point crossover, flip-bit mutation algorithm, selection by tournament, 80% crossover probability, 5% mutation probability and initial population size of 300 individuals. Model calibration resulted in a SPATNSE of 0.94 for the calibration period and 0.91 for the validation period. Table 5-1 shows the events selected for the calibration and validation period.

Table 5-1 - Total rainfall measured by the rain gauges RG1 to RG4 and the peak water level measured by sensors at monitoring locations N1, N2 and N3 during the events used for calibration and validation of the model.

	Event	Date	Duration (h)				Total Rainfall (mm)				Peak water level (m)		
			RG1	RG2	RG3	RG4	RG1	RG2	RG3	RG4	N 1	N2	N3
Calibration	**1**	05/11/2013	08:45	04:40	08:55	08:55	9.6	12.2	10	8.2	0.74	0.48	0.93
	2	06/11/2013	09:40	10:00	09:45	10:35	45.2	40.4	37.8	48.2	1.19	1.03	1.43
	3	24/11/2013	00:50	00:55	01:00	01:20	16.6	6.6	10.8	19.4	0.81	0.57	1.11
	4	13/04/2014	01:00	01:40	01:05	00:10	26.8	25.2	38	0.8	2.44	2.28	-
Validation	**1**	04/11/2013	04:40	04:35	05:05	16:10	64.6	47.4	41.6	73.8	2.00	1.56	4.02
	2	07/11/2013	03:20	03:05	00:10	02:55	9.6	2.4	1.6	5.6	0.67	0.48	1.05
	3	16/11/2013	01:15	01:25	00:35	01:20	10.6	5.6	5.2	16.4	0.79	0.49	1.84
	4	10/12/2013	03:15	01:30	-	01:10	13.4	9.6	-	5.2	1.25	1.22	1.51

5.4 Experimental setup

5.4.1 Assimilating data from one water level sensor

The usefulness of the data assimilation methodology for a catchment with spatially varying characteristics, modelled by a semi-distributed model, and with more than one flood location is being evaluated. All this prompts for exploring various ways of assimilating measurements in a distributed system. This experiment aims to assess the effectiveness of DA methods when receiving regular information only at one location. The evaluation is made accounting the effects of rainfall updating throughout the catchment in the water level

simulations at the place where the sensor data is being assimilated and in the other monitoring locations. Assimilation of water level data at the three monitoring locations are tested, one at a time, described as scenarios 1, 2 and 3, depending on what of the three sensors provide data to assimilate (Table 5-2). The model time step is fifteen minutes, and the sensor information is assimilated at regular intervals of one hour. Five methods of rainfall correction, as previously described, are proposed, and their effectiveness is compared.

5.4.2 Assimilating data from multiple sensors

In this experiment, the DA scenarios handling water level information from several measurement locations received at the same time are tested. It is assumed that the observed water level data arrives synchronously with the frequency of one hour. Table 5-2 lists the assimilation scenarios 4, 5, 6 and 7, describing the combination of sensors providing data to be assimilated at the same time. For all them, the effectiveness evaluation is done considering the effects of rainfall updating throughout the catchment in the model runs at all the locations monitored by water level sensors.

Table 5-2 - Scenarios of model input correction at three water level monitoring locations within the Monjolinho catchment.

	Assimilating data at		
Scenario	N1	N2	N3
1		x	
2			x
3	x		
4	x	x	x
5	x		x
6	x	x	
7		x	x

5.5 Results and discussion

5.5.1 Sensitivity analysis

Figure 5-4 shows the sensitivity analysis results. The graph on the left presents the impact of varying the original rainfall amount in each sub-catchment on the maximum water level at the outlet of the catchment. The results show that increased rainfall has a more

significant influence on the output than its reduction. The map on the right visualises the model output sensitivity to rainfall variation over the catchment. The shown percentages for each sub-catchment are the maximum water level variations at the outlet resulting from variations in rainfall in the range from -60% to +60% of the original rainfall values.

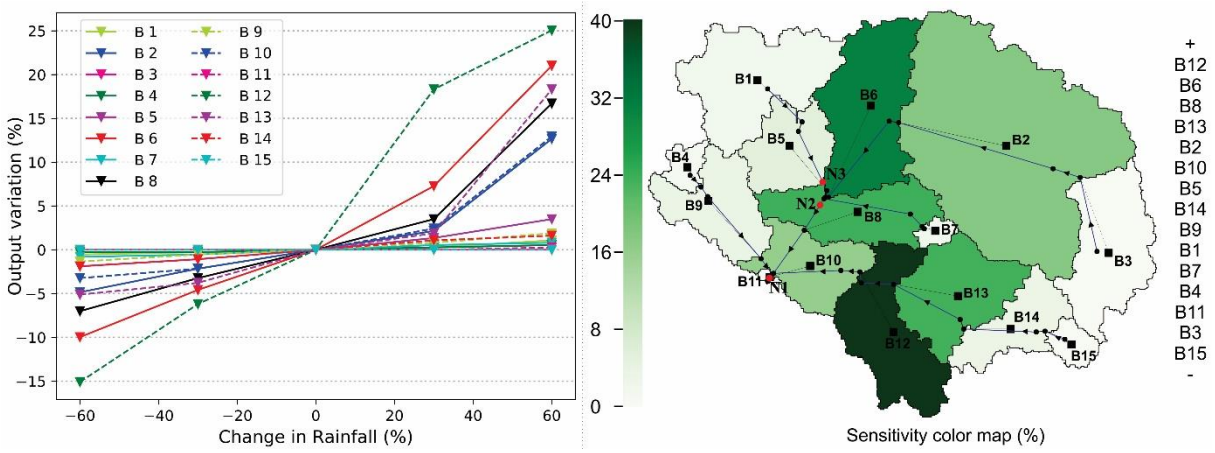


Figure 5-4 - Analysis of model output sensitivity on rainfall variation over the 15 sub-catchments. The graph on the left shows the water level peak variation, and the map to the right presents the maximum water level peak variation, both assessed at the outlet due to changes of rainfall in sub-catchments B1 to B15.

Table 5-3 shows the drainage area, percentage of imperviousness and slope of each sub-catchment, which are all determined in the calibration process. It also shows the time of concentration, as determined by Romero (2016), and the last column presents the results of the sensitivity analysis in terms of the percentage of output variation. The presented parameter values (catchment characteristics) are useful for evaluating the output behaviour. It is worth mentioning, however, that the sensitivity analysis is performed considering only the effect of the rainfall variation at the maximum water level in the outlet of the catchment; it is not accounting for the influence associated with the corresponding area, time of concentration and other parameters, neither their combined effects, that also influence the maximum water level during a rainfall event.

The sub-catchments that most contribute to increasing the water level peak in the outlet are B2, B6, B8, B10, B12 and B13. These catchments have a significant effect on the water level at location N1. Analysing the sensitivity map and the location N2, it can be concluded that the rain gauges influencing the water level variation to this site are B2 and B6, and for P3 these are B1 and B5. Analysing the sensitivity colour map, it is evident that the catchments in the upper part of the rivers have less influence on the runoff, the reason probably being that the

areas around the springs are covered with vegetation, and consequently, they are more permeable. The sub-catchments located in the most urbanised area of the basin are the most sensitive, except for sub-catchment B7. Location and land use are crucial for runoff generation, as well as the catchment area. This last factor is determinant to make sub-catchment B7 less sensitive since its drainage area is the second smallest among the 15 sub-areas, only 0.54 km².

Table 5-3 - Model parameter values per sub-catchment and the maximum water level variation found in the sensitivity analysis.

Sub-catchment	Area (km²)	Imperviousness (%)	Slope (%)	T_c (h)	Maximum water level variation (%)
B1	7.30	23.50	5.45	0.23	1.62
B2	21.84	30.35	10.30	0.54	17.52
B3	8.75	13.54	11.67	0.41	0.20
B4	0.94	54.04	7.90	0.38	0.81
B5	3.68	31.28	8.37	0.30	5.39
B6	7.43	85.49	9.49	1.66	31.00
B7	0.54	86.54	7.80	0.13	1.62
B8	4.82	93.84	8.96	1.89	23.72
B9	4.32	37.47	8.21	0.49	3.23
B10	3.89	34.34	8.55	2.13	16.17
B11	0.11	17.00	12.75	2.13	0.27
B12	6.09	84.91	9.12	0.99	40.16
B13	5.50	82.77	11.51	0.56	23.45
B14	3.31	41.37	9.77	0.54	3.50
B15	1.03	10.23	9.40	0.23	0.20

5.5.2 Assessing the methods of rainfall correction assimilating data from a single sensor

Figure 5-5 shows the NSE for scenarios 1, 2 and 3 when receiving water level data at one location at a time, for the five updating methods (I, II, III, IV and V). The NSE values shown in the figure are the averages across all the rainfall events used in this study, while Table 5-4 details the NSE values per event.

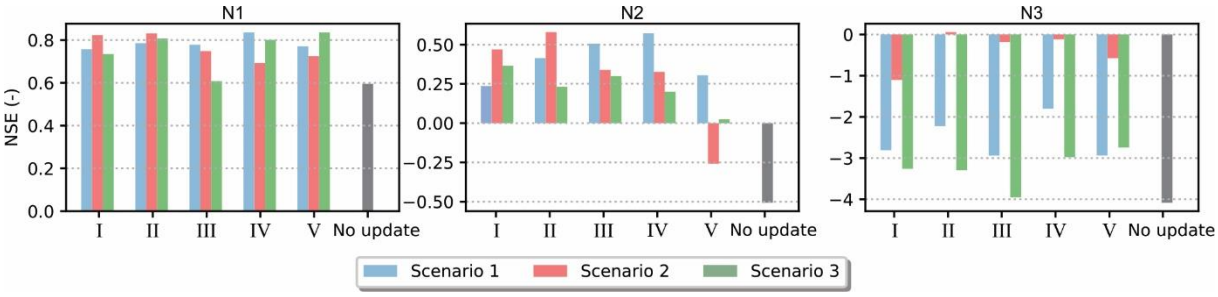


Figure 5-5 - NSE results of all the rainfall events in the three monitoring locations N1, N2 and N3, when assimilating data from one sensor at a time (scenarios 1, 2 and 3) by the five rainfall updating methods (I, II, II, IV and V).

Results are mixed. When aggregating all results, across all events, methods and scenarios of assimilating water level data in one location at a time, it can be said that the assimilation leads to a significant improvement in accuracy of water level simulations in the outlet (N1), compared to model runs for the same location without rainfall updates. The NSE on initial model runs for N2 are quite low, and after correcting, the NSE raised to reasonably high values. One can notice a difference in the effectiveness of the correction methods at N2: methods II and IV have led to better results between the five methods tested. Nevertheless, it can be observed that in general all the methods improved the model accuracy, especially considering that the average of NSE without the updates is about -0.56. Only method III is not performing well when assimilating data at the other monitoring locations N1 and N3, concerning the accuracy of model runs at location N2.

Input model updating through all methods and scenarios of DA improved the NSE of the water level simulations at N3, but the results are still not satisfactory. For this location, it is evident that methods II and IV performed better. Also, all the methods led to better simulation results for N3 when assimilating data on-site (scenario 2). It is important to notice that the area upstream of this location is the smallest among the three analysed here. In addition, this region has the lowest percentage of the impermeable area (Table 5-3). These two factors reduce the effects of increasing or decreasing rainfall input values for that place compared to the other two monitoring locations, which have a larger associated runoff coefficient with the catchment area upstream them.

It can be seen that when assimilating data from one sensor at a time, the correction methods II and IV performed better than the others did. Methods I, III and V showed good results, but in some situations when assimilating water level data from one location, the effect

on the simulations in the other locations is negative. Taking into account the advantage of methods II and IV, they are selected to analyse the influence of their correcting effects on hydrographs further. The hydrographs of all methods, events and scenarios are detailed in Appendix C.

Table 5-4 - NSE results of each rainfall event for the water levels simulated by the model and after input correction by methods I to V, when assimilating data from one sensor at a time (scenarios 1, 2 and 3).

		N1	N2	N3	N1	N2	N3	N1	N2	N3	N1	N2	N3
		Event 1			Event 2			Event 3			Event 4		
	No update	0.89	0.77	0.69	0.79	0.69	-4.46	0.80	-1.71	-5.80	-0.10	-1.78	-6.77
Method I	**Scenario 1**	0.73	0.30	-8.62	0.73	0.30	-8.62	0.87	-0.07	-2.90	0.53	-0.11	-0.14
	Scenario 2	0.75	0.59	-4.53	0.75	0.59	-4.53	0.85	0.13	0.17	0.76	0.43	-0.81
	Scenario 3	0.77	0.74	-7.78	0.77	0.74	-7.78	0.81	-0.32	-5.32	0.48	0.25	-0.39
Method II	**Scenario 1**	0.89	0.62	0.74	0.72	0.26	-8.55	0.90	0.42	-0.88	0.63	0.36	-0.23
	Scenario 2	0.92	0.80	0.88	0.74	0.81	-1.15	0.91	0.28	0.55	0.75	0.43	-0.06
	Scenario 3	0.92	0.88	0.53	0.80	0.25	-7.68	0.85	-0.17	-5.32	0.66	-0.04	-0.70
Method III	**Scenario 1**	0.94	0.87	0.10	0.73	0.38	-8.49	0.83	0.62	-1.70	0.62	0.14	-1.66
	Scenario 2	0.92	0.78	0.87	0.76	0.18	-2.22	0.68	-0.11	0.33	0.63	0.51	0.29
	Scenario 3	0.92	0.73	0.12	0.82	0.48	-5.83	0.57	-0.55	-2.80	0.12	0.54	-7.29
Method IV	**Scenario 1**	0.93	0.86	0.64	0.72	0.67	-3.44	0.90	0.03	-3.01	0.79	0.73	-1.40
	Scenario 2	0.93	0.78	0.83	0.76	0.75	-0.50	0.83	-1.31	-0.36	0.25	-1.26	-0.44
	Scenario 3	0.95	0.80	0.77	0.84	0.74	-4.69	0.77	-1.28	-6.89	0.64	-0.16	-1.11
Method V	**Scenario 1**	0.91	0.65	0.53	0.75	0.33	-8.66	0.84	-0.06	-2.84	0.66	0.30	-0.78
	Scenario 2	0.88	0.83	0.78	0.71	0.59	-1.42	0.88	0.30	-1.04	0.43	-0.41	-0.62
	Scenario 3	0.93	0.82	0.59	0.85	0.47	-5.08	0.85	-0.46	-5.60	0.71	-0.03	-0.89

Figure 5-6 shows the Taylor diagrams for assimilation methods II and IV, for the four rainfall events selected, and their performance in the three locations in the basin compared to the simulated values without updates. The Taylor diagram features three statistics for assessing the degree of correspondence between the modelled and observed behaviour of data: the Pearson correlation coefficient, the standard deviations of the modelled and reference values,

and the unbiased Root Mean Square Deviation (RMSD). The overall results show a majority improvement in the statistics for all the scenarios assimilating data, even when the information was given in a different location from where the results are being verified.

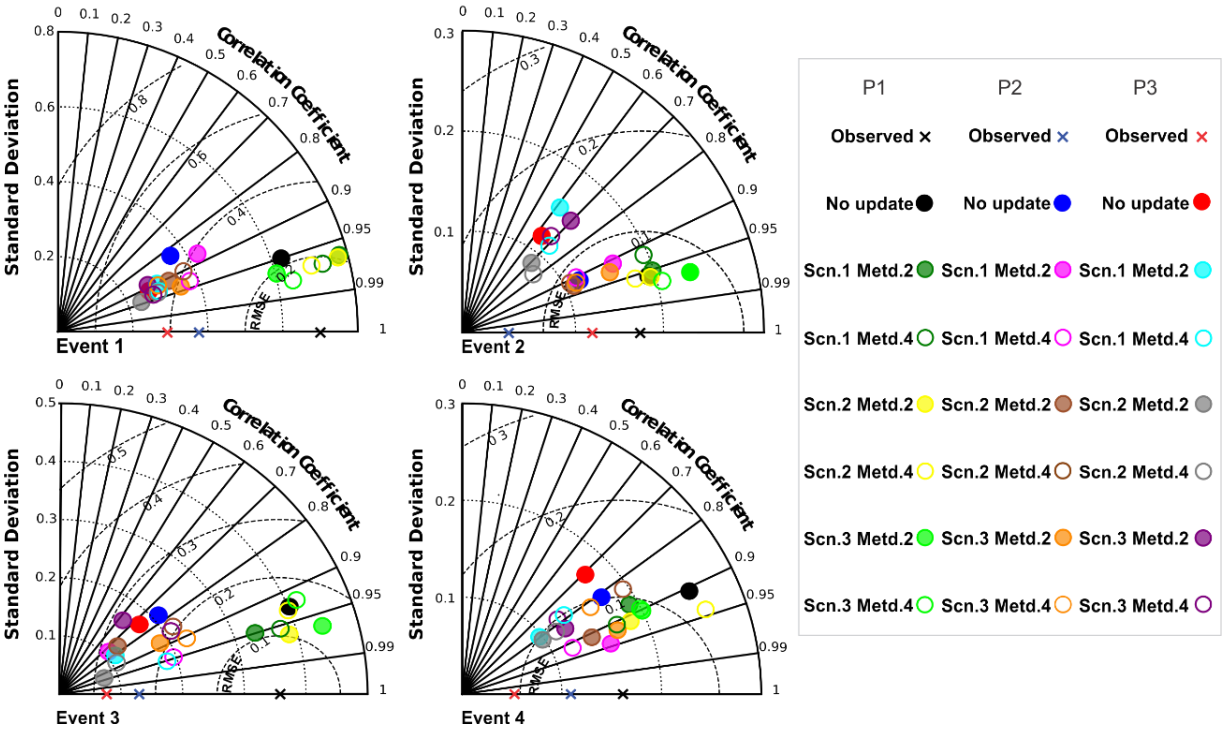


Figure 5-6 - Taylor diagram for the four rainfall events summarising statistics for the four scenarios proposed in Experiment 1. The graphs show results for DA methods II and IV compared to simulations without updating.

Figure 5-7 shows the hydrographs of the three monitoring locations N1, N2 and N3 when assimilating data at N3 by methods II and IV. Analysing the percentage of rainfall changes for each sub-catchment by the two methods, one may notice that both lead mostly to rainfall input reduction, depicted by the red colour, instead of rainfall input increased, illustrated by the blue tones. These results are consistent with the fact that the simulated hydrograph is overestimated. The hydrograph at N3 shows that the updates made at 19h00 and 20h00 are the most important for correcting rainfall inputs at this event because, at these moments, the simulated water level has significant errors. On the one hand, the correction at 19h00 by method II led to rainfall input increase at sub-catchment B1 and decrease at sub-catchment B5. After the update, the simulation results improved immensely, reducing simulated water levels to values very close to those observed. On the other hand, in the assimilation at 20h00, the updates for catchments B1 and B5 reduced approximately 60% the rainfall inputs values, and despite

this significant amount of rainfall reduction, it is not enough for the model to reach the observed water levels.

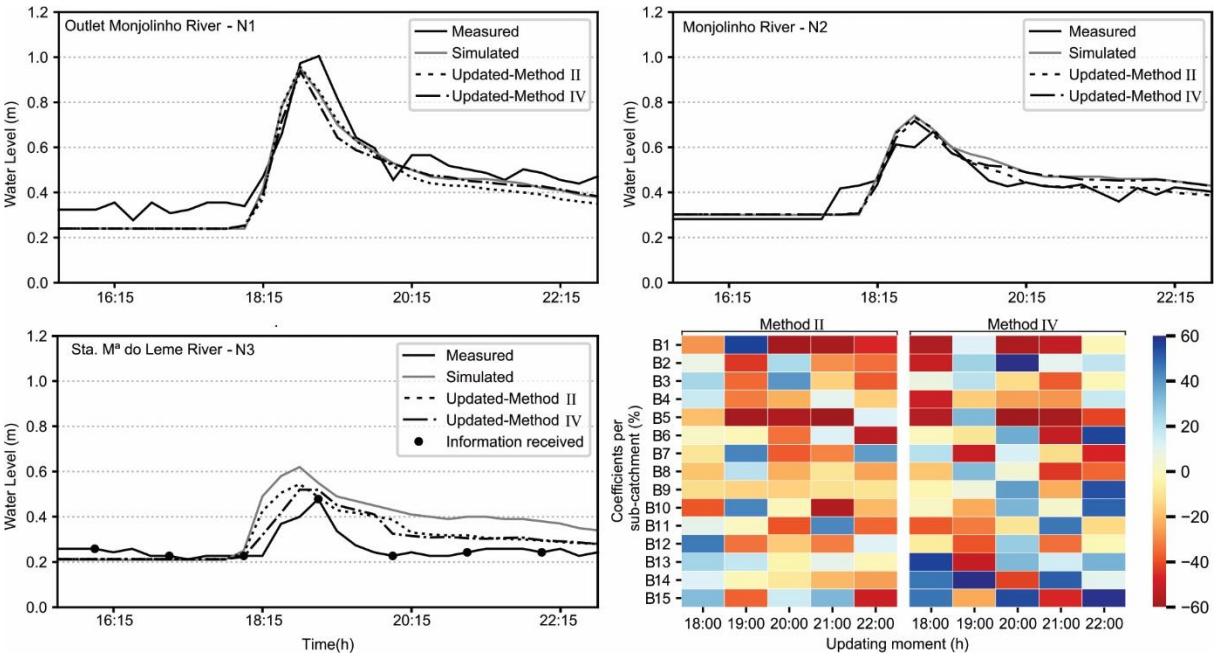


Figure 5-7 - The three hydrographs show the water levels observed, simulated and resulting from updates by the methods II and IV during event 2 in the three monitoring locations. The updates are performed by assimilating water level data from sensor N3 (scenario 2), every hour during the rainfall event.

Figure 5-8 shows the results when assimilating data at N3 (scenario 2) during rainfall event 4. Neither of the two correcting methods managed has led to a satisfactory increase of NSE for that case. Method II increased the NSE from -6.77 to 0.06 and method IV increased to -0.44. During event 4 after 14h00, the model overestimated the water levels. For method II it can be seen that for the critical period between 14h00 and 18h00, the found optimum correcting coefficients for the sub-catchments B1 and B5 lead to a reduction of rainfall inputs, but still, simulation underestimates water levels. For method IV, a similar situation is observed.

The simulated hydrographs for locations N1 and N2 in Figure 5-8 are overestimated, and after optimisation of the coefficients to correct rainfall based on water level data assimilated at N3, this helped to reduce the error for these two points by both correction methods. Method II achieves better results for N1 and N2, raising their NSE from -0.1 to 0.75 and -1.78 to 0.43, respectively. Method IV lead to worst effectiveness results compared to method II, increasing the NSE at N1 from -0.1 to 0.25 and at N2 from -1.78 to -1.26. In general, in both methods, the coefficient percentage shows they have reduced the rainfall inputs. It is important to emphasise

that the method II approach does not reduce rainfall inputs just backwards; it also changes values forwards from the moment when assimilating data. Consequently, updates in the simulation can be more significant when using method II.

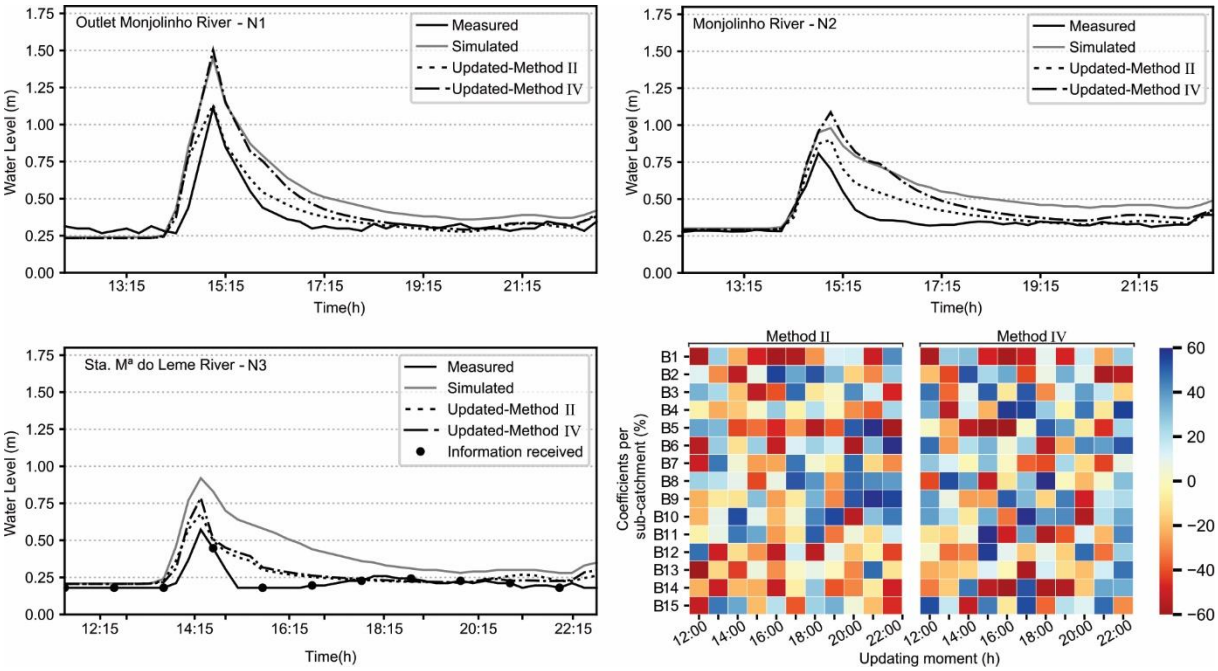


Figure 5-8 - The three hydrographs show the water levels observed, simulated and resulting from updates by the methods II and IV during event 4 in the three monitoring locations. The updates are performed by assimilating water level data only from sensor N3 (scenario 2), every hour during the rainfall event.

5.5.3 Assessing model accuracy when assimilating data from multiple sensors

From the results when assimilating data from one sensor at a time, method II and method IV performed better compared to the other updating methods. Taking this into account, scenarios assimilating water level data from multiple sensors at the same time are tested only for rainfall update methods II and IV.

Figure 5-9 shows the overall NSE results for scenarios 4, 5, 6 and 7, at the three monitoring locations N1, N2 and N3. For the two correction methods and all the scenarios, the data assimilation improved the NSE. For results at location N1 and N2, the accuracy of the model when assimilating data by method IV is slightly better than by method II.

At location N1, all the NSE results are very good. Scenario 7, when assimilating data at N2 and N3 at the same time, reached better simulation effectiveness values for N1 than when

assimilating data on-site in the same time with other monitoring locations (scenarios 4, 5 and 6). At N2, the NSE results are also good, and the best results are reached in scenario 7. While at location N3, the NSE does not achieve satisfactory values. Through method II, when assimilating water level data at N1 and N3 (scenario5), the simulation results at N3 showed a slight improvement. The worst results for there are when assimilating data in the other two monitoring locations N1 and N2 (scenario 6). Even then, the NSE results at N3 improved when compared to the simulation accuracy without correcting rainfall.

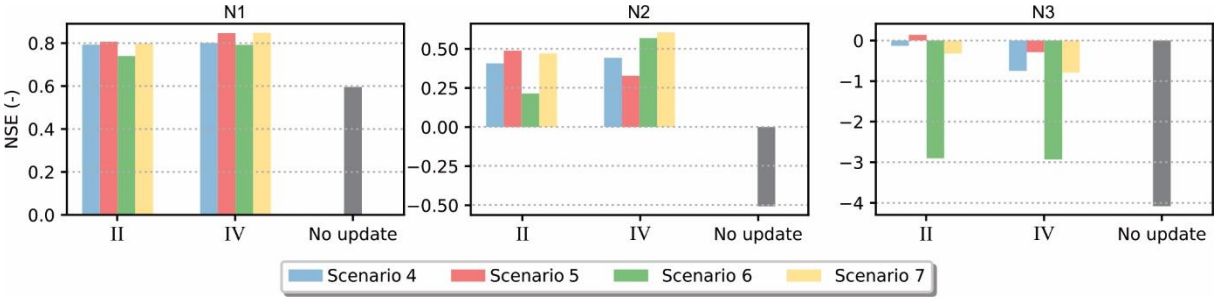


Figure 5-9 - NSE results of all the rainfall events in the three monitoring locations N1, N2 and N3, when assimilating water level data from multiple sensors at the same time (scenarios 4, 5, 6 and 7) by methods II and IV.

Figure 5-10 shows the Taylor diagrams for the scenarios proposed by Experiment 2. The "x" marked on the x-axis represents the observed data standard deviation. Simulations closest to the marker are those that have the standard deviation closest to the aim, which is, they have the same pattern of variation as the observed data. Just in a few cases, the simulated data were closer to the observation than the updated simulations. However, it is also worth noting that for those same cases, the other two statistics (correlation and RMSE) have improved. Therefore, overall results show a majority improvement in the statistics for all the scenarios assimilating data.

Table 5-5 shows in detail the resulting NSE for all scenarios and each rainfall event by the two methods tested. Below two scenarios and events are better detailed through the hydrographs produced, as well as the values of reducing and increasing rainfall inputs during the updates are detailed. The hydrographs of all methods, events and scenarios are detailed in Appendix D.

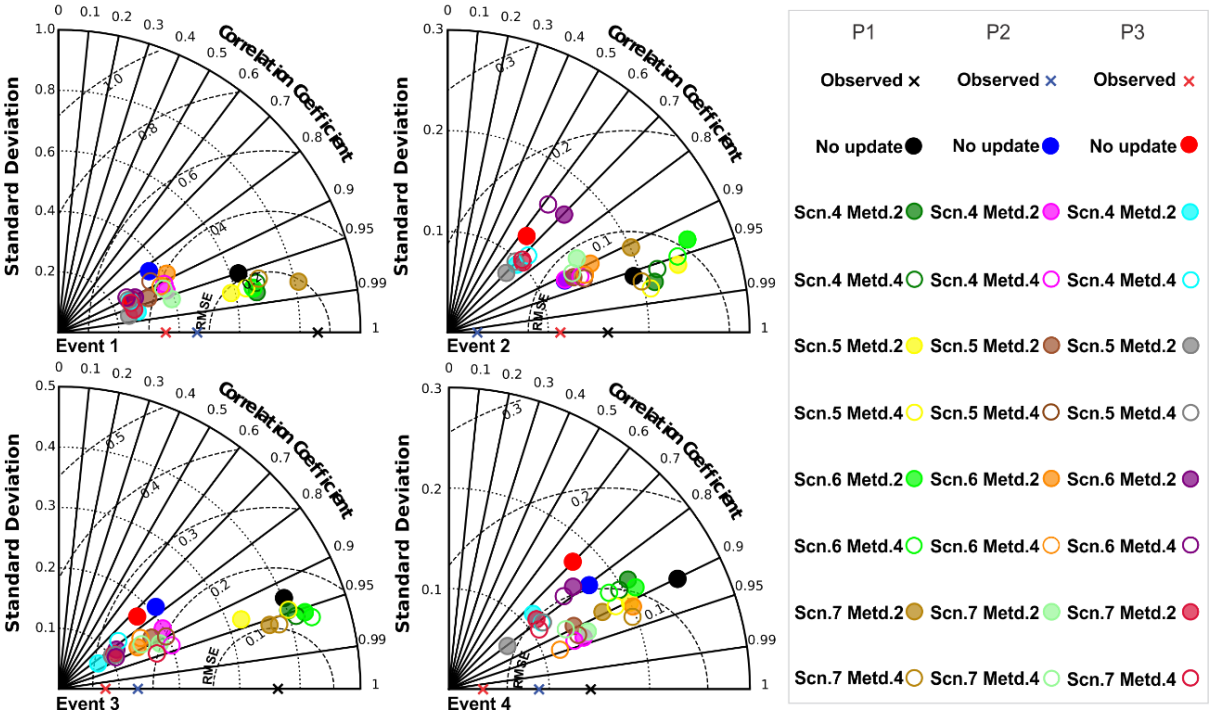


Figure 5-10 - Taylor diagram for the four rainfall events summarising statistics for the four scenarios proposed in Experiment 2. The graphs show results for data assimilation methods 2 and 4 compared to simulations without updating.

Table 5-5 - NSE results of each rainfall event comparing the water levels measured at the monitoring locations with the model simulations for scenarios tested in experiment 2 of rainfall updating by methods II and IV.

		N1	N2	N3	N1	N2	N3	N1	N2	N3	N1	N2	N3
		Event 1			Event 2			Event 3			Event 4		
	No update	0.89	0.77	0.69	0.79	0.69	-4.46	0.80	-1.71	-5.80	-0.10	-1.78	-6.77
Method II	**Scenario 4**	0.96	0.85	0.94	0.83	0.75	-1.08	0.86	-0.56	0.24	0.52	0.59	-0.62
	Scenario 5	0.93	0.86	0.94	0.75	-4.38	-0.54	0.87	0.02	-0.30	0.68	0.49	0.45
	Scenario 6	0.95	0.68	0.76	0.62	0.32	-8.95	0.85	0.48	-0.98	0.54	-0.62	-2.43
	Scenario 7	0.92	0.91	0.92	0.59	0.48	-1.33	0.91	0.14	-0.45	0.77	0.36	-0.41
Method IV	**Scenario 4**	0.94	0.79	0.84	0.78	0.59	-2.09	0.87	-0.19	-1.44	0.61	0.58	-0.30
	Scenario 5	0.94	0.75	0.80	0.85	0.46	-1.25	0.86	-0.39	-0.33	0.73	0.49	-0.37
	Scenario 6	0.95	0.81	0.79	0.72	0.48	-9.96	0.85	0.21	-0.26	0.64	0.77	-2.30
	Scenario 7	0.93	0.82	0.86	0.79	0.67	-1.28	0.91	0.36	-2.64	0.77	0.58	-0.10

Figure 5-11 shows the results when assimilating water level data at N1 and N2 at the same time (scenario 3). The event presented in the hydrographs happened almost entirely between 18h00 and 20h00. The hourly updates with observed data are not enough to satisfactory update the rainfall that affects the rise and decrease of the flood wave. Furthermore, the assimilation of data at untimely moments made the simulation overestimate the water levels through both methods II and IV. In the first hydrograph (location N1), when the first information is assimilated at 18h00, the model slightly underestimates the values compared to the observation, then the optimisation based on this data aimed to increase the rainfall inputs. When new information is received subsequently at 19h00, the peak time already passed, and the previous correction overestimated the peak. Because of this information gap at a crucial time, the optimisation ends up making mistakes and worsening the estimative. The optimisation results based on observed values at location N1 and N2 reflected even more negatively in the simulations at N3. Since the beginning of the event at N3, the hydrograph is overestimated, and the increase in rainfall inputs values by the correcting methods raised, even more, the water levels.

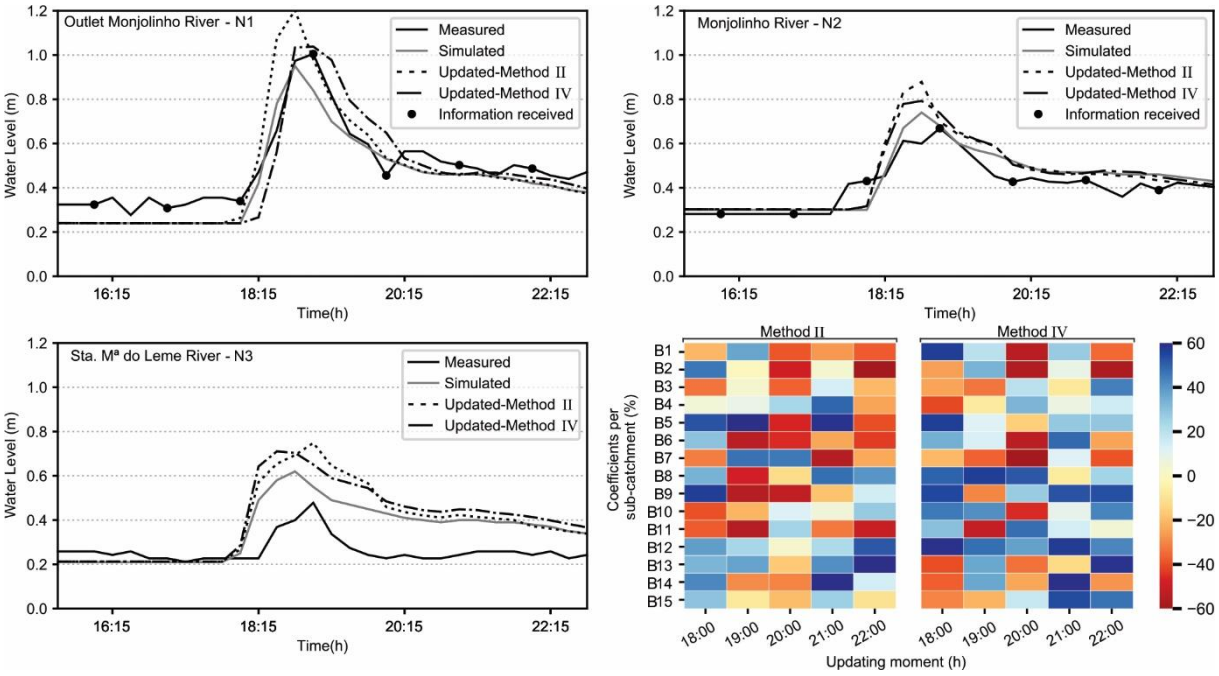


Figure 5-11 - The three hydrographs show the water levels observed, simulated and resulting from updates by the methods II and IV during event 4 in the three monitoring locations. The updates are performed by assimilating data from sensor N1 and N2 at the same time (scenario 6), every hour during the rainfall event.

During event 3, when assimilating water level data from sensor N1 and N2 (Figure 5-12), the results for both methods are very similar. For location N1, the rainfall input correction

caused the peak to be slightly overestimated while the original simulation is accurate. The falling limb of the model overestimated values, and the updates in rainfall made the water level decay faster, reaching values nearer to those observed in the hydrograph. Methods II and IV increased the NSE for this event at N1 from 0.8 to 0.85. At location N2, the simulated water levels are slightly ahead of the observed ones. The rainfall updates help to reduce the water level values that are overestimated, but it is not able to correct the advance of the hydrograph concerning the observed one. N3 has the simulation errors reduced by the updates as well. At this location, the simulated hydrograph is also overestimated compared to the observed. The reduction of rainfall inputs helped to adjust the water level values also at this location.

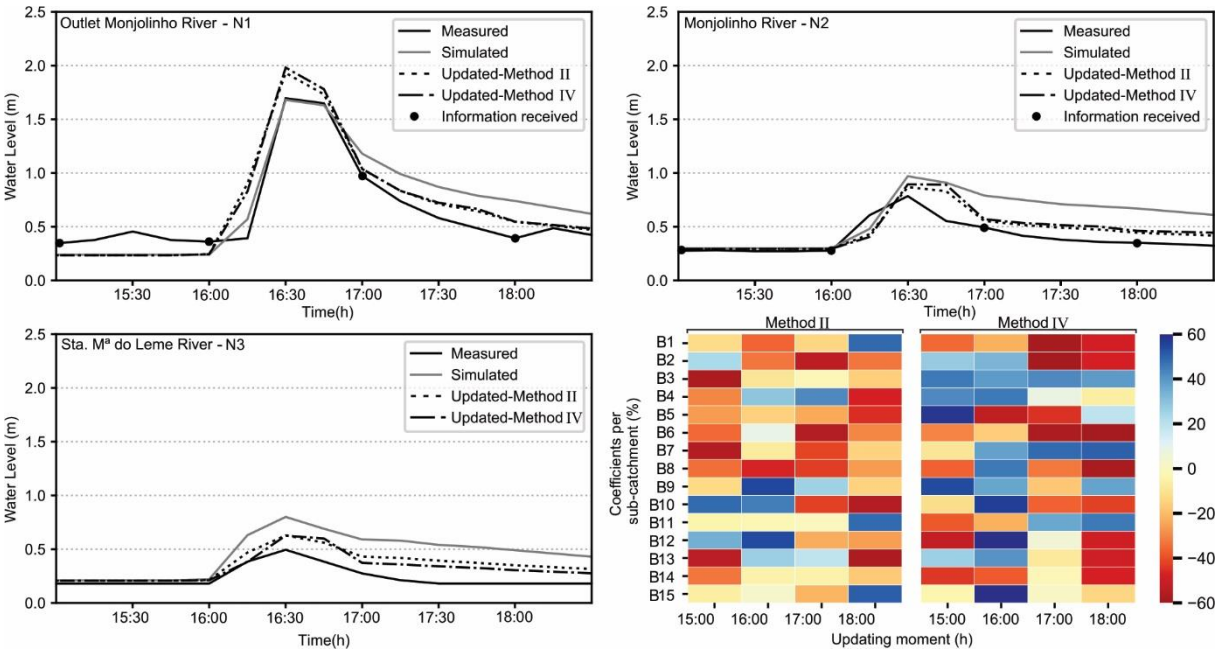


Figure 5-12 - The three hydrographs show the water levels observed, simulated and resulting from updates by the methods II and IV during event 3 in the three monitoring locations. The updates are performed by assimilating data from sensor N1 and N2 at the same time (Scenario 6), every hour during the rainfall event.

5.5.4 Summary of the methods´ effectiveness

Each one of the events has different rainfall behaviour, and amount precipitated, these factors may influence the effectiveness of data assimilation methods. After evaluating all the methods and scenarios, one at a time, a comparison of effectiveness is made for method II and IV considering all the data series and scenarios together. In order to compare them, the NSE of all rainfall events are averaged, and for each scenario of receiving sensor data at different

locations. The efficiency measures are calculated separately for each monitoring location. Figure 5-13 shows the result of this comparison.

Analysing the results for location N1, represented by the red circles, it can be observed that both methods have very similar effectiveness results. Although the average NSE for all events is already at a satisfactory value without correcting, both approaches are able to improve the estimates slightly. At location N2, represented by the blue circles, the correction methods have similar effectiveness but with some discrepancies. For scenarios 2 and 3, method II has better results, and for scenarios 5 and 6, method IV is more effective. Therefore, from these results at N2, it is not possible to define a clear effectiveness difference between the two methods. Among the three monitoring locations, N3 is where the model presents the most significant errors in producing the estimates. Method II changes the rainfall values in the previous instants when the information is received and after receiving sensor data. Because of it, the correction is slightly more significant than by method IV, which changes values of rainfall just going back from the moment when information is received. As a result, for location N3, represented by the green circles, it is possible to identify slightly better effectiveness of method II.

It is interesting to note that in scenario 4 (Figure 5-13) when assimilating data from the three sensors, the model accuracy gained by both methods is higher but very marginal compared to the other scenarios assimilating data from two or only one sensors. The low-performance gain when the number of sensors is increased might be explained due to the nested nature of the sub-catchments where the sensors are installed. Since their flow must be correlated, any correction made in one location by the data assimilation will also be propagated in the others. Sensors N1 and N2 are located on the Monjolinho River in a cemented part of the canal, while sensor N3 is located in a tributary of the Monjolinho River that has natural coverage of the canal bed and a smaller cross-section compared to the ones of sensors N1 and N2. From these considerations, it is expected that water levels registered by sensors N1 and N2 have more correlated behaviour compared to sensor N3. Looking at the results for scenarios 1, 3 and 6, which have in common the fact that they do not assimilate data from sensor N3, it can be noticed that they showed a less significant improvement in the simulations' effectiveness for location N3. This has probably happened because they are receiving data only from sensors N1 and N2 that are not as correlated to sensor N3 data.

Another aspect that needs to be assessed is the values of the coefficients found to perform the rainfall correction. According to the results demonstrated by the sensitivity analysis of the model to the rainfall variation, some sub-catchments are much more responsive for the runoff generation. It is expected that the optimisation method has given more emphasis on the rainfall variation at these catchments. However, the values found are quasi-optimal because of the time constraint assigned to the search algorithm. The method is outlined to run for real-time operation in basins with rapid responses to the rainfall; hence, the search time must be less than the time step of the model.

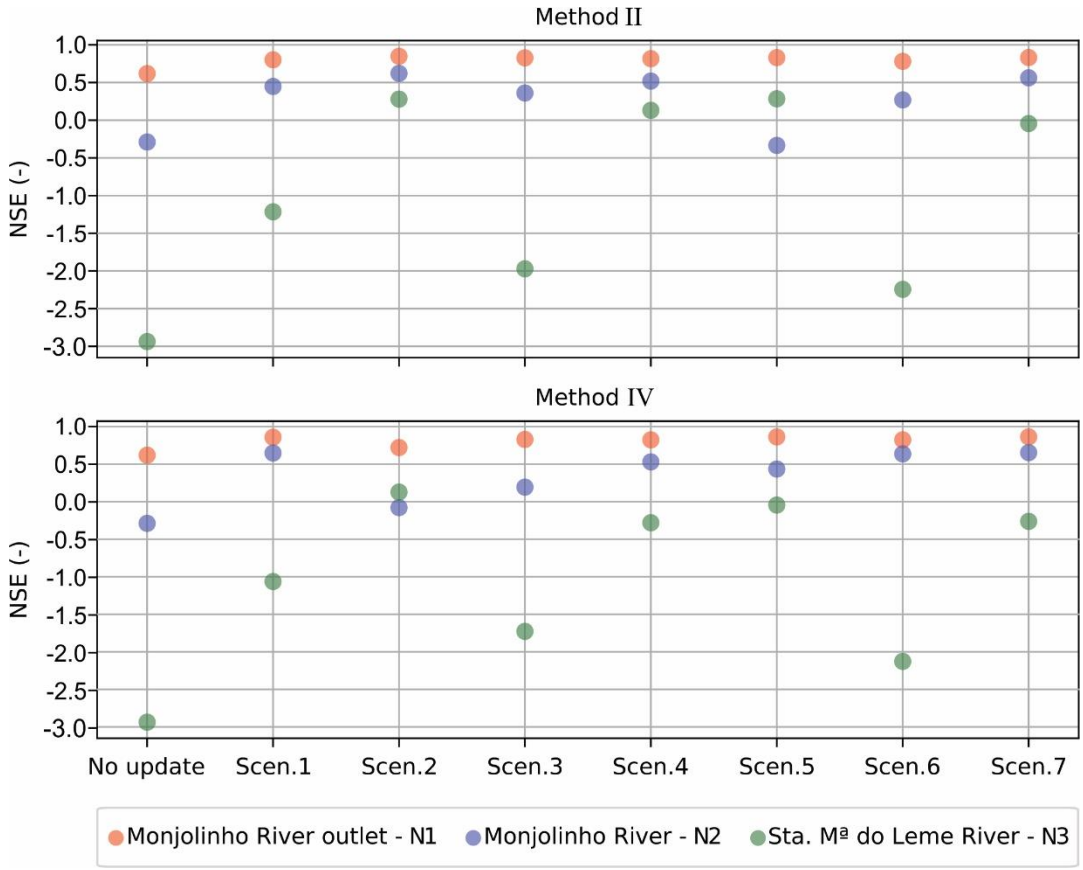


Figure 5-13 - Comparison between DA methods II and IV by the mean NSE of all rainfall events used in this study. The values show the results obtained for all the proposed data assimilation scenarios (Scen.).

Figure 5-14 shows the boxplots of the average percentage of rainfall variation at all scenarios analysed in each sub-catchment by correction methods II and IV. According to the sensitivity analysis of the model to rainfall inputs variation, the six most sensitive sub-catchments in descending order are B12, B6, B8, B13, B2 and B10, while the other catchments are almost non-sensitive. The six sub-catchments with the most significant variation in rainfall given by the optimisation for the updating method II are B5, B6, B2, B13, B1 and B12. Sub-

catchments B5 and B1 have low sensitivity to rainfall according to sensitivity analysis, and contrary to the expected, their rainfall inputs have high changes set by the optimisation, while the very sensitive sub-catchments B8 and B10 are not the ones with the biggest coefficients of rainfall change. By method IV, sub-catchments B6, B2 and B13 have high percentage values of rainfall changes compared to the others. However, sub-catchment B1, that sensitivity analysis showed is almost non-sensitive, is the one that received the highest percentage of rainfall input changing.

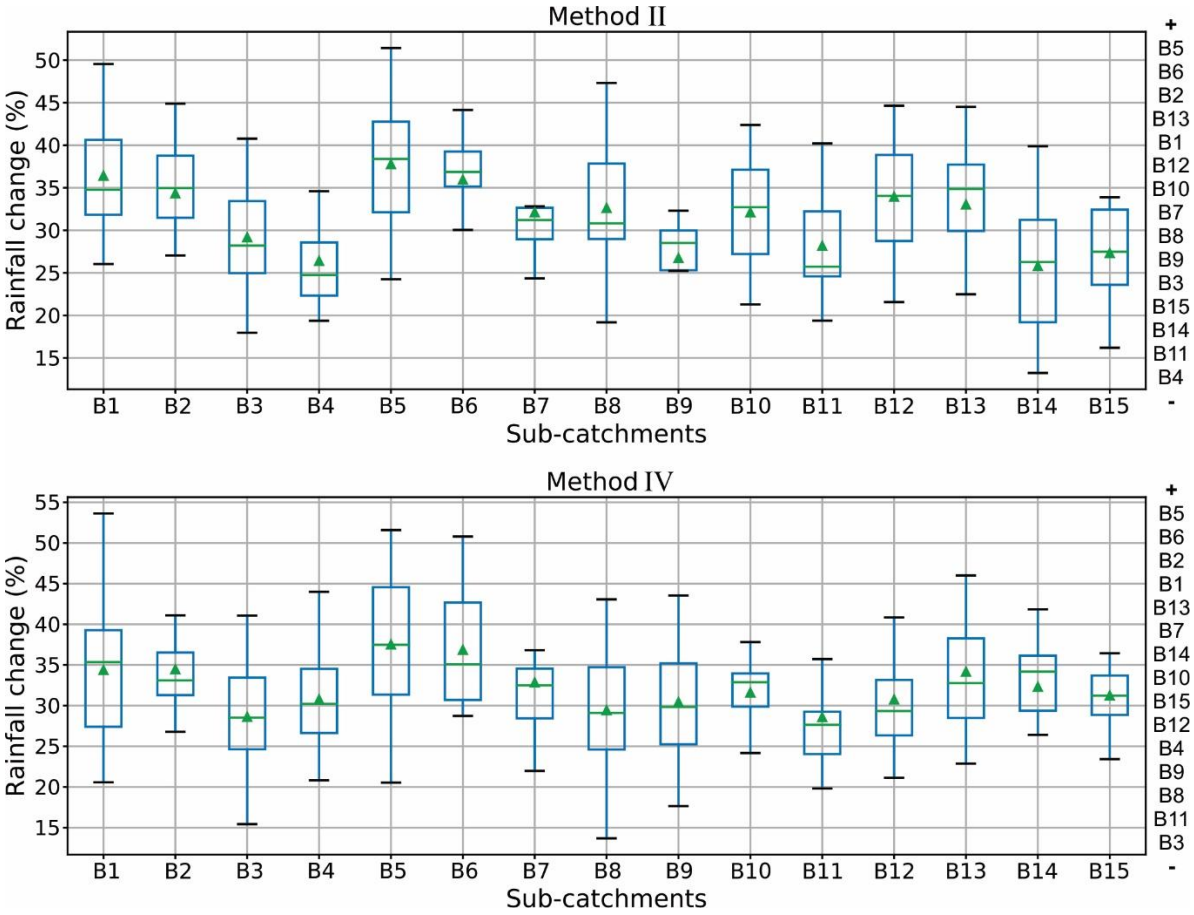


Figure 5-14 - Box plots of the average of rainfall variation of the seven scenarios of model input correction by method II and IV. Values calculated for sub-catchments B1 to B15.

5.6 Conclusions

The proposed rainfall correction methodology based on water level observations and implemented in a real case study through integration with the SWMM model is a simple and useful tool for improving flood prediction. The model output is very sensitive to rainfall variation, having a considerably higher influence when increasing those values. It shows that the method is more effective to correct underestimation in the water level simulations. Some

sub-catchments are much more significant to the runoff generation; further investigation should be performed aiming to reduce the number of decision variables to be optimised or even to give different weights to them according to sub-catchment sensitivity to rainfall input variation.

The results showed that the approach of correcting rainfall values assimilating water level data from only one sensor, by all the methods and regardless of its location in the catchment, led to improving the NSE of the water level simulations upstream and downstream the location where data is being assimilated. Looking at the overall results, the NSE improve as the number of sensors increases. However, the improvement of simulations accuracy when increasing the number of sensors is not strongly significant. Thus, in real situations, it will be necessary to weigh the cost of installing several sensors and the accuracy gain of the forecasts according to the specific needs of each location. It can be concluded from the assessments that this approach is a promising tool to improve flood-forecasting models that require rapid responses being computationally inadequate for ensembles, and a simple way to circumvent the limitation by the nonlinear correlation between observation and model states. Additional analyses are recommended to improve the method further. The authors suggest exploring the possibility to use faster models than SWMM and other optimisation schemes for future studies. Currently, data from traditional water level sensors are assimilated. However, this approach is not limited only to this type of data. For instance, data from citizen observatories can be used as well; further studies with assimilating asynchronous data and uncertainties in the observational data are strongly recommended.

Despite the deterministic approach and limitations of the method, it is an innovative and promising tool to update urban hydrodynamic models. Therefore, it is hoping to contribute to the field of flood forecasting, helping to make cities more resilient to this kind of disaster.

Data and software availability

The data assimilation software for methods I to V and the SWMM calibrator tool are developed in Python 3.6. All the data and the spreadsheets used are available for download at GitHub platform. Link to access the data assimilation software: {https://github.com/mclarafava/Data-Assimilation}.

References

Andrieu, C., Doucet, A., & Holenstein, R. (2010). Particle Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *72*(3), 269-342.

Ashley, R. M., Balmforth, D. J., Saul, A. J., & Blanskby, J. D. (2005). Flooding in the future–predicting climate change, risks and responses in urban areas. *Water Science and Technology*, *52*(5), 265-273.

Bai, T., Mayer, A., Shuster, W., & Tian, G. (2018). The hydrologic role of urban green space in mitigating flooding (Luohe, China). *Sustainability*, *10*(10), 3584.

Barros, R. M., Mendiondo, E. M., & Wendland, E. (2007). Cálculo de áreas inundáveis devido a enchentes para o plano diretor de drenagem urbana de São Carlos (PDDUSC) na bacia escola do córrego do Gregório. *Revista Brasileira Recursos Hídricos*, *12*, 5-17.

Borup, M., Madsen, H., Grum, M., & Mikkelsen, P. (2018). Technical Note on the Dynamic Changes in Kalman Gain when Updating Hydrodynamic Urban Drainage Models. *Geosciences*, *8*(11), 416.

Chen, H., Yang, D., Hong, Y., Gourley, J. J., & Zhang, Y. (2013). Hydrological data assimilation with the Ensemble Square-Root-Filter: Use of streamflow observations to update model states for real-time flash flood forecasting. *Advances in Water Resources*, *59*, 209-220.

Collier, C. G. (2007). Flash flood forecasting: What are the limits of predictability?. *Quarterly Journal of the Royal Meteorological Society*, *133*(622), 3-23.

Coustau, M., Ricci, S., Borrell-Estupina, V., Bouvier, C., & Thual, O. (2013). Benefits and limitations of data assimilation for discharge forecasting using an event-based rainfall-runoff model. *Natural Hazards and Earth System Science*, *13*, pp-583.

Cronshey, R. (1986). *Urban hydrology for small watersheds*. US Dept. of Agriculture, Soil Conservation Service, Engineering Division.

Divac, D., Milivojević, N., Grujović, N., Stojanović, B., & Simić, Z. (2009). A procedure for state updating of SWAT-based distributed hydrological model for operational runoff forecasting. *Journal of the Serbian Society for Computational Mechanics*, *3*(1), 298-326.

Evensen, G. (2003). The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean dynamics*, *53*(4), 343-367.

Fletcher, T. D., Andrieu, H., & 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, 261-279.

Fortin, F. A., Rainville, F. M. D., Gardner, M. A., Parizeau, M., & Gagné, C. (2012). DEAP: Evolutionary algorithms made easy. *Journal of Machine Learning Research*, 13(Jul), 2171-2175.

Francois, C., Quesney, A., & Ottlé, C. (2003). Sequential assimilation of ERS-1 SAR data into a coupled land surface–hydrological model using an extended Kalman filter. *Journal of Hydrometeorology*, *4*(2), 473-487.

Hansen, L. S., Borup, M., Møller, A., & Mikkelsen, P. S. (2014). Flow forecasting using deterministic updating of water levels in distributed hydrodynamic urban drainage models. *Water*, 6(8), 2195-2211.

Harader, E., Borrell-Estupina, V., Ricci, S., Coustau, M., Thual, O., Piacentini, A., & Bouvier, C. (2012). Correcting the radar rainfall forcing of a hydrological model with data assimilation: application to flood forecasting in the Lez catchment in Southern France. *Hydrology and Earth System Sciences*, 16(11), 4247.

He, X., Kidmose, J., Madsen, H., Zheng, C., & Refsgaard, J. C. (2017). Using Data Assimilation in Real-Time Hydrological Modeling of Groundwater and Stream Flow in Silkeborg, Denmark. In *AGU Fall Meeting Abstracts*.

Hutton, C. J., Kapelan, Z., Vamvakeridou-Lyroudia, L., & Savić, D. A. (2012). Dealing with uncertainty in water distribution system models: A framework for real-time modeling and data assimilation. *Journal of Water Resources Planning and Management*, 140(2), 169-183.

Hutton, C. J., Kapelan, Z., Vamvakeridou-Lyroudia, L., & Savić, D. (2014). Real-time data assimilation in urban rainfall-runoff models. *Procedia Engineering*, 70, 843-852.

IBGE - Instituto Brasileiro de Geografia e Estatística. (2015). Estimativas da população residente no Brasil e unidades da federação com data de referência em 1º de julho de 2018.

Kahl, B., & Nachtnebel, H. P. (2008). Online updating procedures for a real-time hydrological forecasting system. In *IOP Conference Series: Earth and Environmental Science*(Vol. 4, No. 1, p. 012001). IOP Publishing.

Kavetski, D., Kuczera, G., & Franks, S. W. (2006). Bayesian analysis of input uncertainty in hydrological modeling: 2. Application. *Water Resources Research*, 42(3).

Lindström, G. (2016). Lake water levels for calibration of the S-HYPE model. *Hydrology Research*, *47*(4), 672.

Liu, Y., & Gupta, H. V. (2007). Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resources Research*, 43(7).

Liu, Y., Weerts, A. H., Clark, M., Franssen, H. J. H., Kumar, S., Moradkhani, H., ... & van Velzen, N. (2012). Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. *Hydrology and Earth System Sciences*, 16(10), 3863-3863.

Maybeck, P. S. (1982). *Stochastic models, estimation, and control* (Vol. 3). Academic press.

Mazzoleni, M., Chacon-Hurtado, J., Noh, S. J., Seo, D. J., Alfonso, L., & Solomatine, D. (2018a). Data Assimilation in Hydrologic Routing: Impact of Model Error and Sensor Placement on Flood Forecasting. *Journal of Hydrologic Engineering*, *23*(6), 04018018.

Mazzoleni, M., Noh, S. J., Lee, H., Liu, Y., Seo, D. J., Amaranto, A., Alfonso, L., & Solomatine, D. (2018b). Real-time assimilation of streamflow observations into a hydrological routing model: Effects of model structures and updating methods, *Hydrological Science Journal*, 63(3), 386-407.

McMillan, H., Jackson, B., Clark, M., Kavetski, D., & Woods, R. (2011). Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models. *Journal of Hydrology*, 400(1-2), 83-94.

McMillan, H. K., Hreinsson, E. Ö., Clark, M. P., Singh, S. K., Zammit, C., & Uddstrom, M. J. (2013). Operational hydrological data assimilation with the recursive ensemble Kalman filter. *Hydrology and Earth System Sciences*, *17*(1), 21-38.

Mendes, H. C., & Mendiondo, E. M. (2007). Histórico da expansão urbana e incidência de inundações: O Caso da Bacia do Gregório, São Carlos–SP. *Revista Brasileira de Recursos Hídricos*, *12*(1), 17-27.

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

Moradkhani, H., Hsu, K. L., Gupta, H., & Sorooshian, S. (2005). Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *Water Resources Research*, *41*(5).

Moradkhani, H., DeChant, C. M., & Sorooshian, S. (2012). Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-Markov chain Monte Carlo method. *Water Resources Research*, *48*(12).

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*, *10*(3), 282-290.

Pathirana, A. (2015). SWMM5 calls from python (version 1.1.0.2) [Computer software]. Retrieved from Python Package Index (PyPI) repository: https://pypi.org/project/SWMM5/

Pedersen, J. W., Lund, N. S., Borup, M., Löwe, R., Poulsen, T. S., Mikkelsen, P. S., & Grum, M. (2016). Evaluation of Maximum a Posteriori Estimation as Data Assimilation Method for Forecasting Infiltration-Inflow Affected Urban Runoff with Radar Rainfall Input. *Water*, 8(9), 381.

Stanzel, P., Kahl, B., Haberl, U., Herrnegger, M., & Nachtnebel, H. P. (2008). Continuous hydrological modelling in the context of real time flood forecasting in alpine

Danube tributary catchments. In *IOP Conference Series: Earth and Environmental Science* (Vol. 4, No. 1, p. 012005). IOP Publishing.

Refsgaard, J. C. (1997). Validation and intercomparison of different updating procedures for real-time forecasting. *Hydrology Research*, *28*(2), 65-84.

Reichle, R. H., McLaughlin, D. B., & Entekhabi, D. (2002). Hydrologic data assimilation with the ensemble Kalman filter. *Monthly Weather Review*, *130*(1), 103-114.

Romero, G. B. (2016). *Dinâmica ecohidrológica de rios urbanos no contexto de gestão de riscos de desastres* (Doctoral dissertation, Universidade de São Paulo).

Rossman, L. A. (2010). *Storm water management model user's manual, version 5.0* (p. 276). Cincinnati: National Risk Management Research Laboratory, Office of Research and Development, US Environmental Protection Agency.

Rossman, L. A., & Huber, W. (2016). *Storm water management model reference manual Volume I–Hydrology (Revised).* US Environmental Protection Agency: Cincinnati, OH, USA.

Segond, M. L., Wheater, H. S., & Onof, C. (2007). The significance of spatial rainfall representation for flood runoff estimation: A numerical evaluation based on the Lee catchment, UK. *Journal of Hydrology*, 347(1-2), 116-131.

Seo, D. J., Koren, V., & Cajina, N. (2003). Real-time variational assimilation of hydrologic and hydrometeorological data into operational hydrologic forecasting. *Journal of Hydrometeorology*, *4*(3), 627-641.

Seo, D. J., Cajina, L., Corby, R., & Howieson, T. (2009). Automatic state updating for operational streamflow forecasting via variational data assimilation. *Journal of Hydrology*, *367*(3-4), 255-275.

Thiboult, A., & Anctil, F. (2015). On the difficulty to optimally implement the Ensemble Kalman filter: An experiment based on many hydrological models and catchments. *Journal of Hydrology*, *529*, 1147-1160.

Tollan, A. (2002). Land-use change and floods: what do we need most, research or management?. *Water Science and Technology*, *45*(8), 183-190.

Walker, J. P., & Houser, P. R. (2005). Hydrologic data assimilation. *Advances in Water Science Methodologies*, *1*, 25-48.

WMO - World Meteorological Organization. (2011). *Manual on flood forecasting and warning.* WMO, No. 1072.

Xie, X., & Zhang, D. (2010). Data assimilation for distributed hydrological catchment modeling via ensemble Kalman filter. *Advances in Water Resources*, *33*(6), 678-690.

Yang, G., Bowling, L. C., Cherkauer, K. A., & Pijanowski, B. C. (2011). The impact of urban development on hydrologic regime from catchment to basin scales. *Landscape and Urban Planning*, *103*(2), 237-247.

Yin, J., Yu, D., Yin, Z., Liu, M., & He, Q. (2016). Evaluating the impact and risk of pluvial flash flood on intra-urban road network: A case study in the city center of Shanghai, China. *Journal of Hydrology*, *537*, 138-145.

Young, P. C. (2002). Advances in real–time flood forecasting. Philosophical Transactions of the Royal Society of London A: *Mathematical, Physical and Engineering Sciences*, 360(1796), 1433-1450.

Zawadzki, I. I. (1973). Statistical properties of precipitation patterns. *Journal of Applied Meteorology*, *12*(3), 459-472.

6 General conclusions

Considering the high spatial and temporal resolution required in urban models, the quality of the data used to calibrate these models is essential to guarantee the quality of the results. The developed SWMM calibrator showed to be a tool capable of circumventing the limitation of not having observed flow data nor a rating curve for any monitoring location in the catchment. The validation results are suitable for most of the monitoring locations. However, for some events, the results are not satisfactory, showing the calibration using only water levels was not able to completely reproduce the flow response to the rainfall inputs. It evidences the need to develop methodologies for model updating in real-time to improve simulations accuracy further.

The real-time estimator is developed to integrate citizen science data into the model as an inexpensive solution to improve water level predictions in ungauged or poorly monitored catchments. From the results, it can be concluded that even in a considerable amount of time between the updates (1h ~ 2h) and considering the uncertainty in citizen science data, it can be helpful to improve the accuracy of model forecasts. The regionalisation method, a second step from the results obtained with the real-time estimator, also enhanced the accuracy of the simulations, a substantial contribution to updating models of catchments with several flood risk locations.

The model outputs of Monjolinho catchment are sensitive to rainfall variation, in particular when increasing the original input values, demonstrating that the deterministic method of input correction has the potential to improve water level simulations. Assimilation of water level data from only one sensor, for all the methods and regardless of its location in the catchment, leads to an improvement of water level simulations. Moreover, NSE values improved as the number of sensors increases. However, the improvement of simulations accuracy when increasing the number of sensors is not very notable. Thus, it is fundamental to properly assess the trade-off between the cost of installing several sensors and the gain in model accuracy.

The proposed deterministic updating methods are a simple way to circumvent the limitation of nonlinear correlation between observation and model states. They demonstrate the potential to improve flood forecasting simulation for locations that require rapid responses, where, e.g., ensemble-based assimilation methods would be infeasible, in terms of

computational effort. The presented approaches are promising tools for increasing accuracy of urban hydrodynamic modelling and forecasting, and hence contributing to making cities more resilient to floods.

Limitations and Recommendations

The main limitation of this study is the deterministic way of the updating procedures, which are not considering the uncertainties of the model and the observed data. Moreover, this study only uses deterministic rainfall from rain gauges, whereas numerical weather predictions and remote sensing products are increasingly used and recommended for flood forecasting models. Besides, the model and DA methods are tested only at a case study in the Monjolinho catchment, further comparison in other catchments, various rainfall-runoff events, several combination and quantity of physical-sensors may ultimately contribute to clarify upon the accuracy gain due to the number of water level sensors (physical or citizen sensors).

It is recommended to improve the DA methodologies by checking locations that have the strongest correlation to update the model simulations, as well as to analyse the parameters that influence the correlation quality between the points such as the size of sub-catchments, position, and distances.

In future work, it can be recommended to assess the effect of increasing or decreasing the assimilation interval on model accuracy – this will allow for evaluating the usefulness of both updating methods (Chapters 4 and 5) when using other sources of data, such as (asynchronous) data from citizen observatories, which also has varying uncertainty, and also with different amounts of data being received simultaneously. Also, it is suggested to explore the possibility of using faster models than SWMM (or its surrogates) and other optimisation schemes.

Appendix A

Bathymetry measurements are performed in five cross-sections in its course (P1, P2, P3, P8 and P14). The first location (P1) is approximately three kilometres (km) far from the Monjolinho River source. Location P2 is close to the densification of urban occupation and the channel still have a natural coverage but without margin protection, as in location P1. Location P3 and P8 are in the region of intense urbanisation. The channel has natural coverage in both places, although at location P8 can be noticed much more intervention in the margins and an enlargement of the main channel. Location P14 is in the outlet of the catchment. In this part of the river, the channel is modified to a constructed structure with cement coverage.

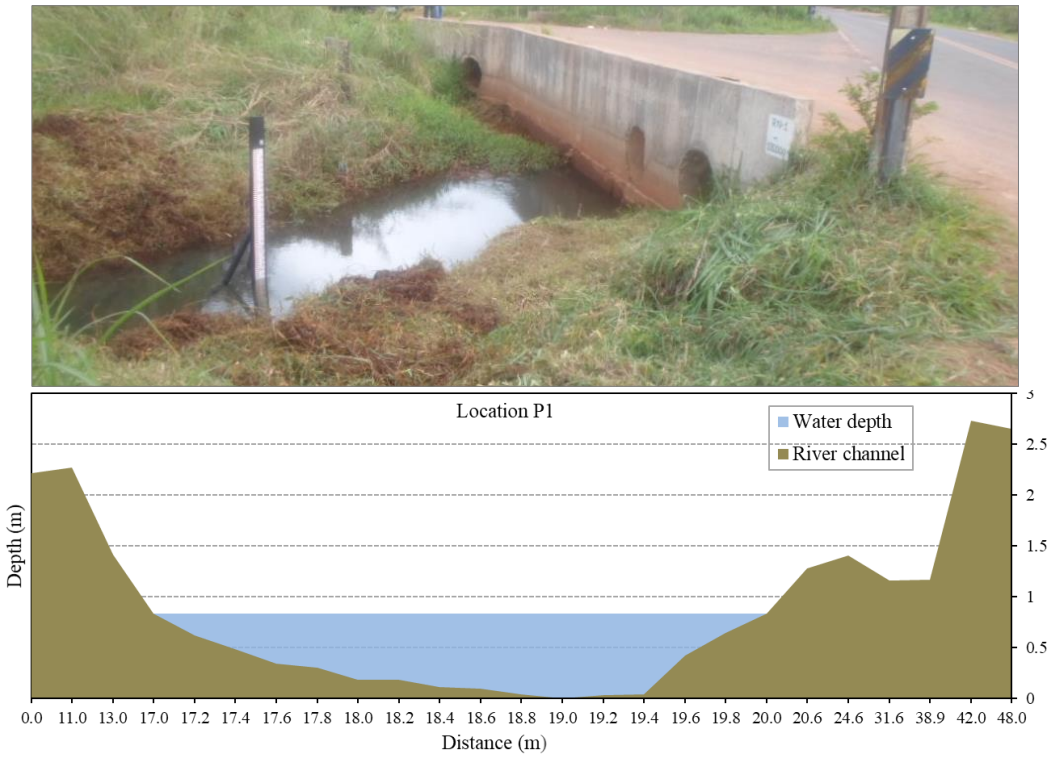


Figure 1 - Monjolinho River's cross-section at location P1 and a photograph of the area.

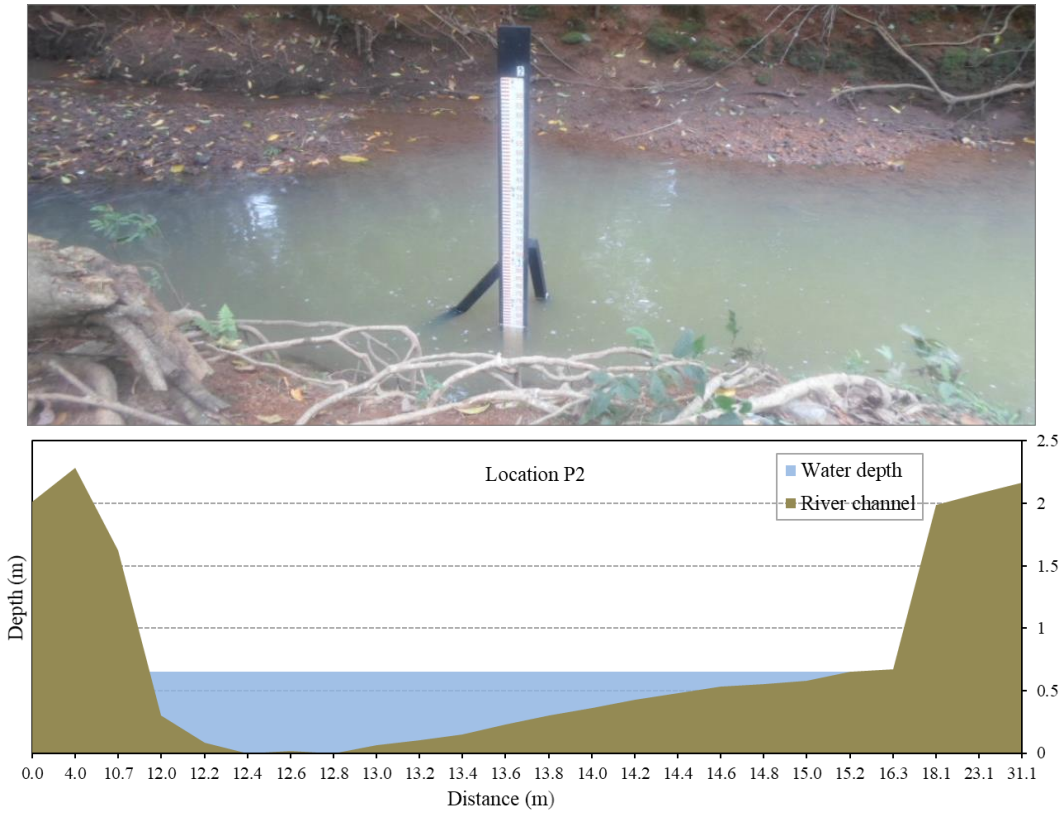


Figure 2 - Monjolinho River's cross-section at location P2 and a photograph of the area.

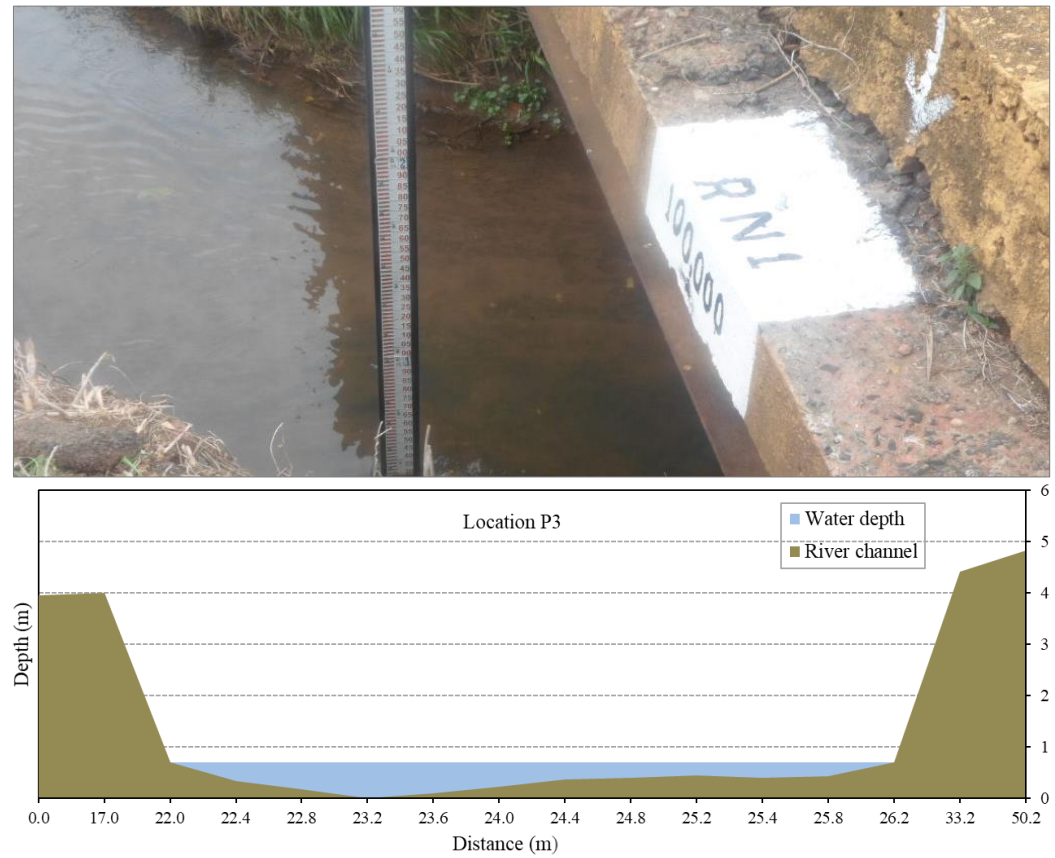


Figure 3 - Monjolinho River's cross-section at location P3 and a photograph of the area.

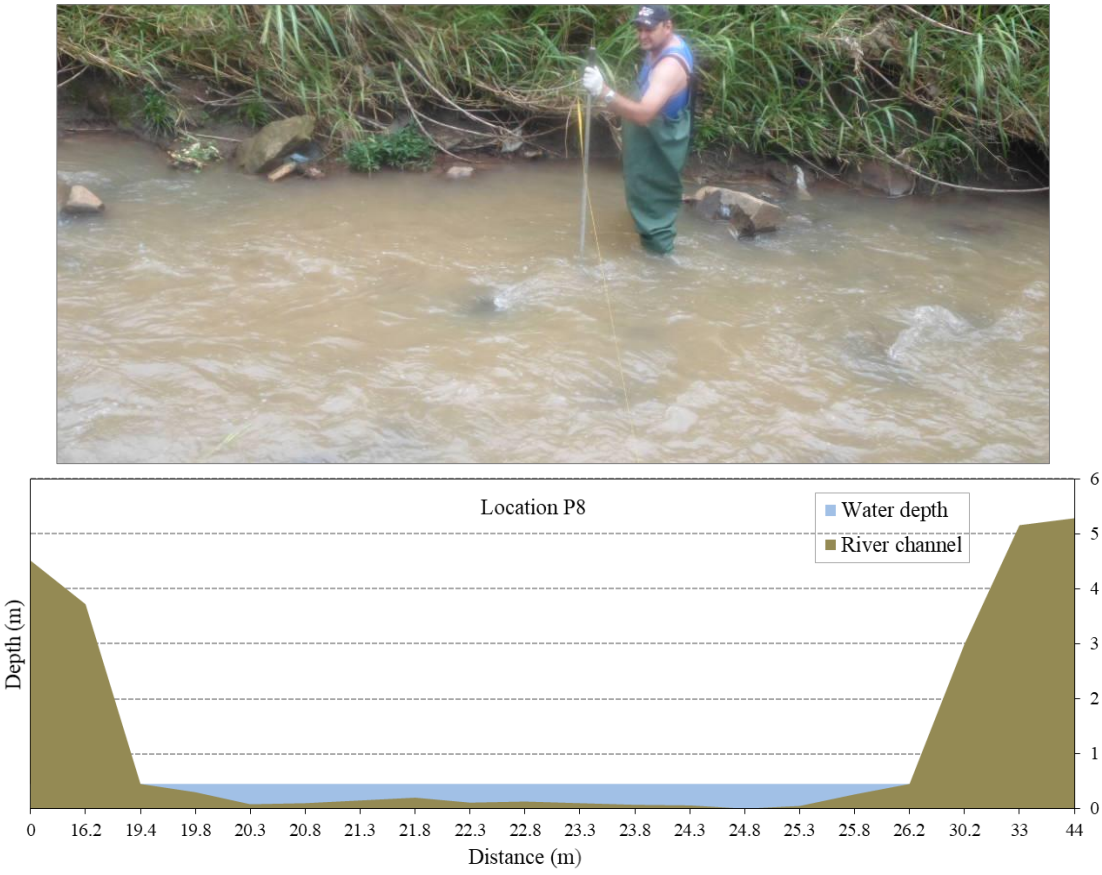


Figure 4 - Monjolinho River's cross-section at location P8 and a photograph of the area.

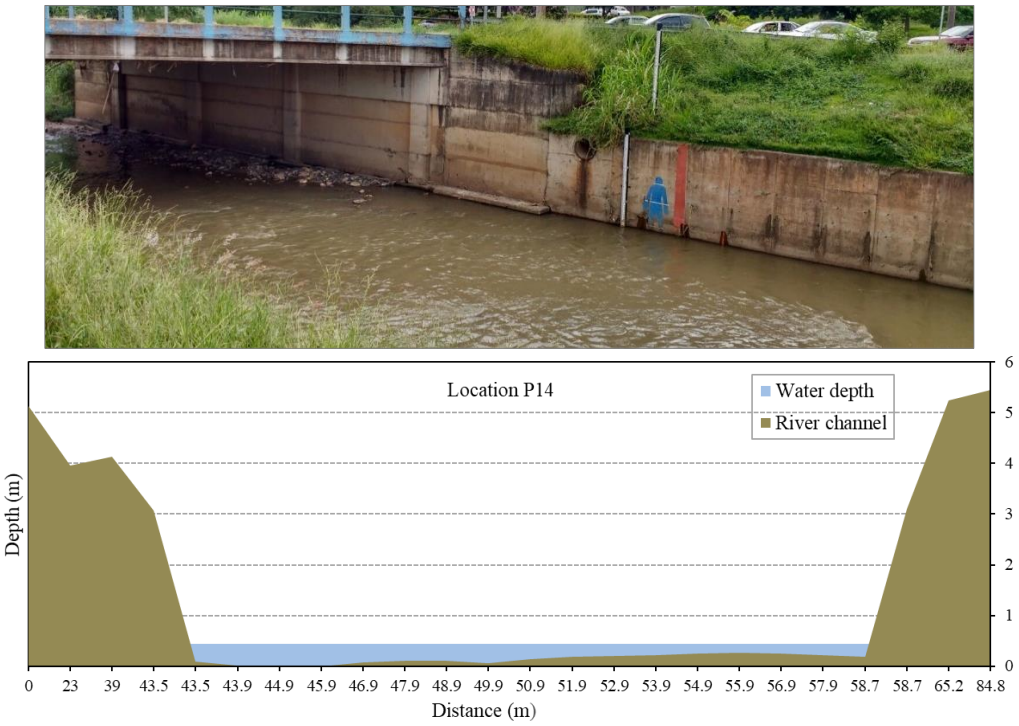


Figure 5 - Monjolinho River's cross-section at location P14 and a photograph of the area.

At Santa Maria do Leme Stream the bathymetry and initial flow are measured at two places, one upstream point just before intense urbanisation starts (P4) and another near to the

junction with Monjolinho River (P5). No alteration or lining of the channel bed is made at Santa Maria do Leme Stream, but there is no native vegetation on the banks of the channel in any of the monitored sections. Near to the junction with Monjolinho River frequently occur flood events.

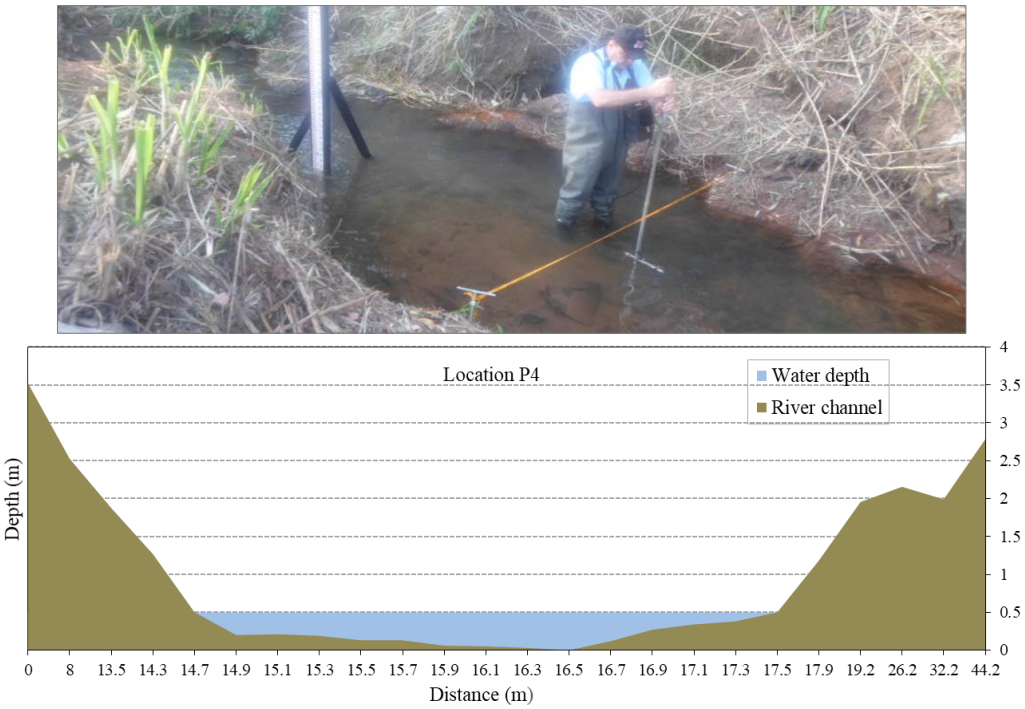


Figure 6 - Santa Maria do Leme Stream's cross-section at location P4 and a photograph of the area.

Tijuco Preto Stream has its source within an area of intense urbanisation (above 70%, Ohnuma & Mendiondo, 2003). At the bottom regions of the valley, there are improper occupations out of environmental governance. Location P6 is very close to the source of Tijuco Preto Stream, and its margins are revitalised by the FIPAI/PMSC (2003) project. Whereas, at location P7, there is a constructed channel covered by cement.

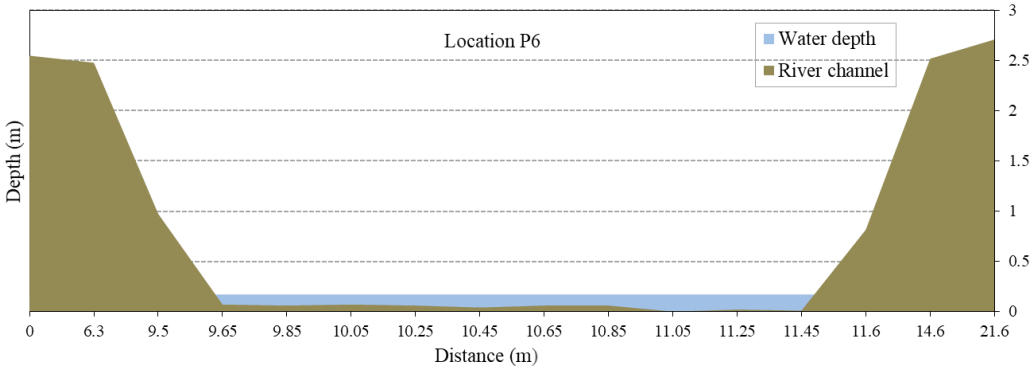


Figure 7 – Tijuco Preto stream's cross-section at location P6 and a photograph of the area.

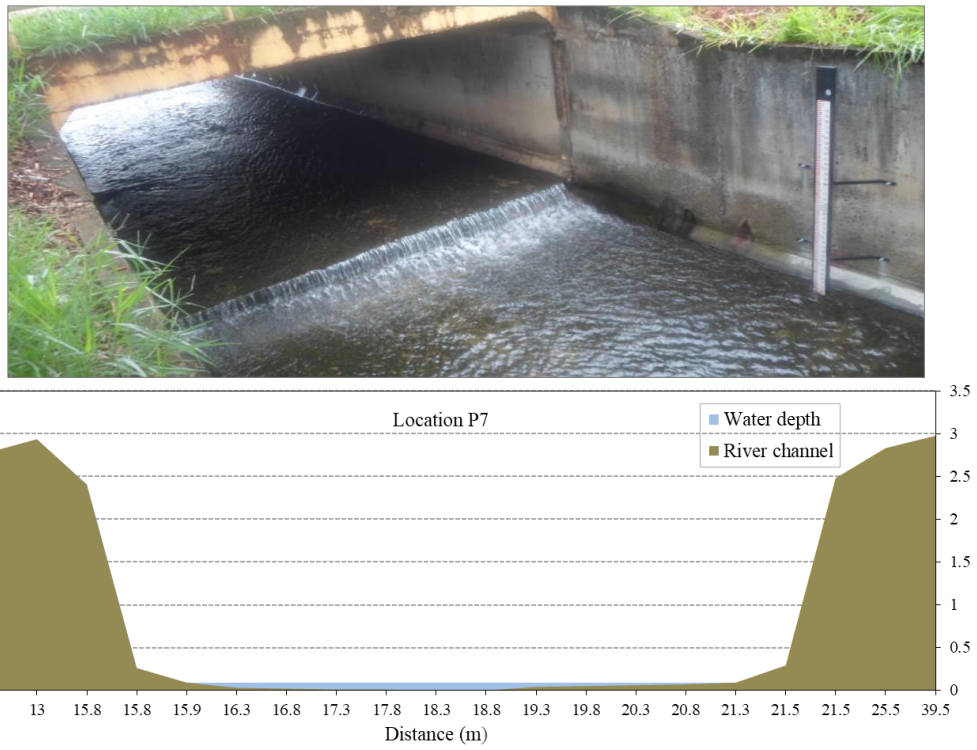


Figure 8 - Tijuco Preto stream's cross-section at location P7 and a photograph of the area.

Mineirinho Stream has its source protected by law as a Permanent Preservation Area (PPA) at location P9, and a natural coverage at location P10 without natural margin conservation. At Gregório Stream, information is collected at four locations. Location (P15) is very close to its source, around 1 km far. Gregório Stream has its source inside a private property (Location 15), and it is protected by law as a Permanent Preservation Area (PPA). At location P15, P11 and P12 the bed of the stream have a natural coverage, and at P13 it is modified to a constructed channel coated with cement.

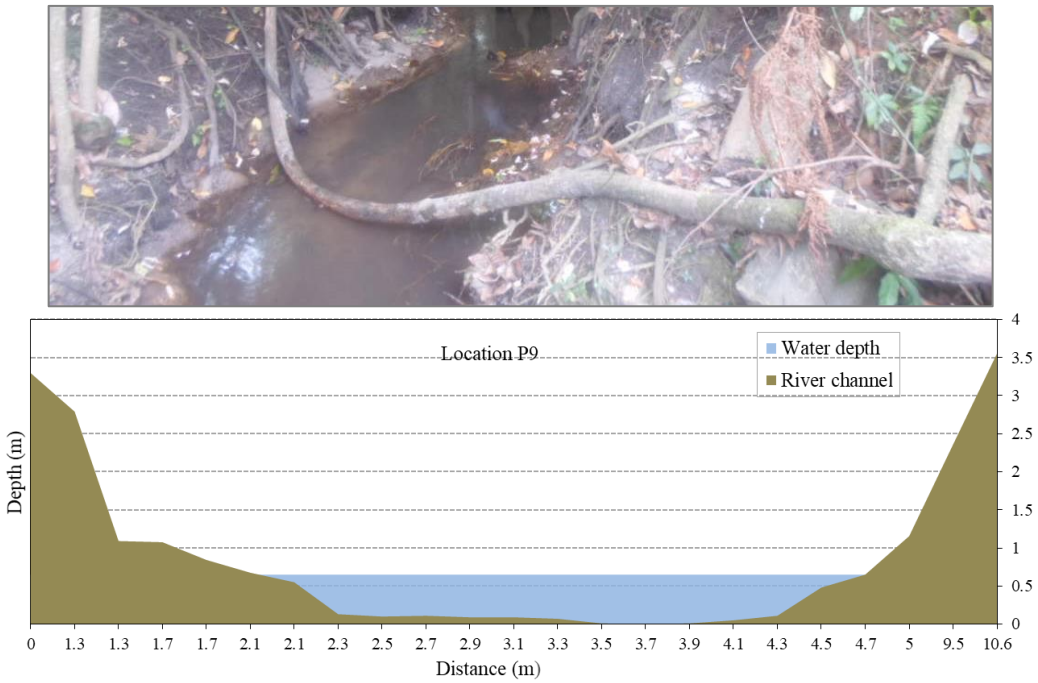


Figure 9 – Mineirinho stream's cross-section at location P9 and a photograph of the area.

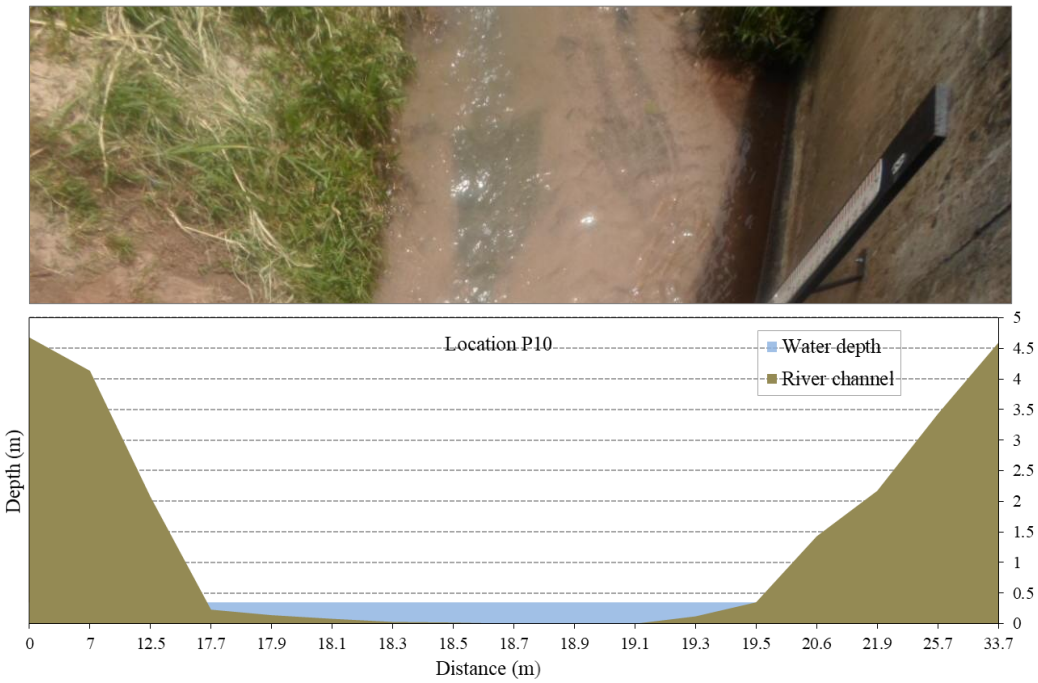


Figure 10 - Mineirinho Stream's cross-section at location P10 and a photograph of the area.

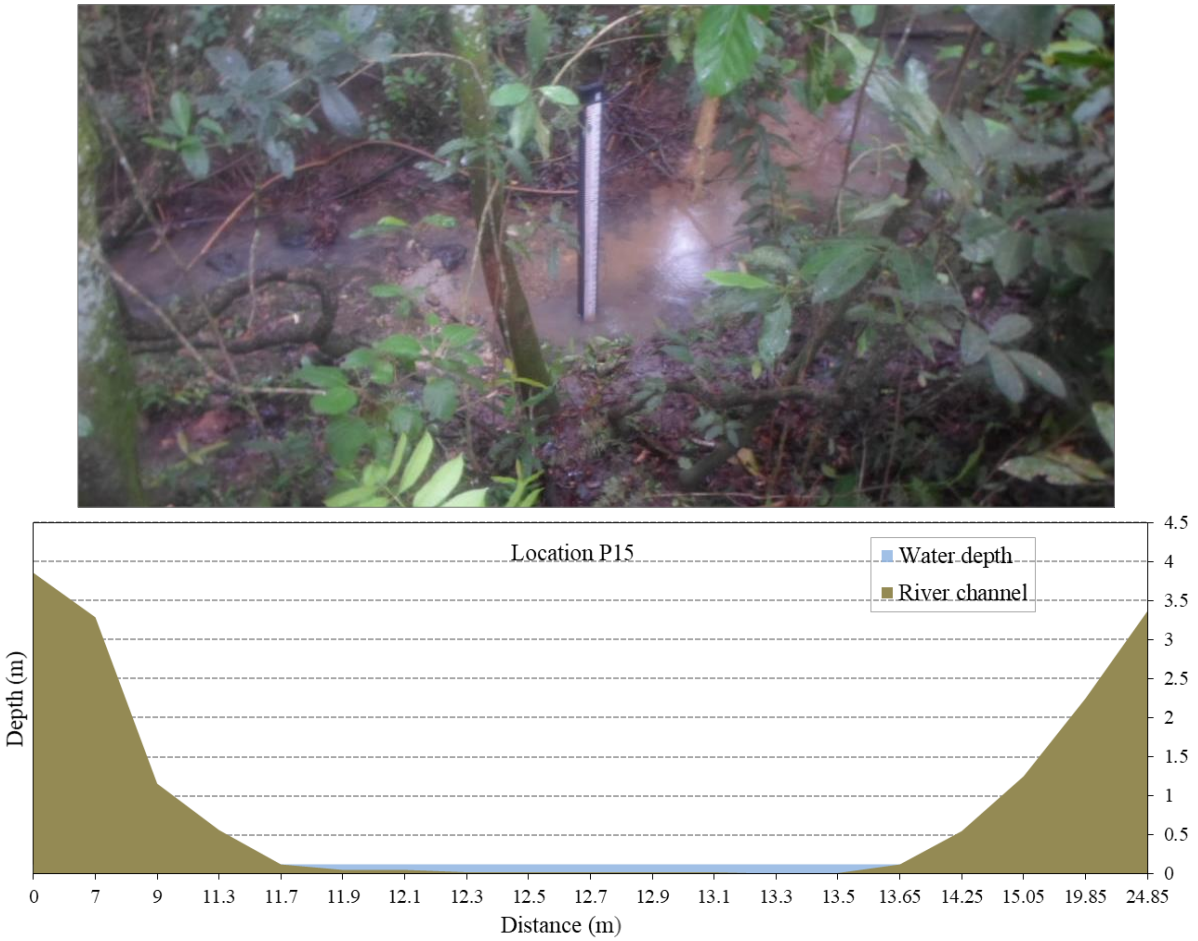


Figure 11 - Gregório Stream's cross-section at location P15 and a photograph of the area.

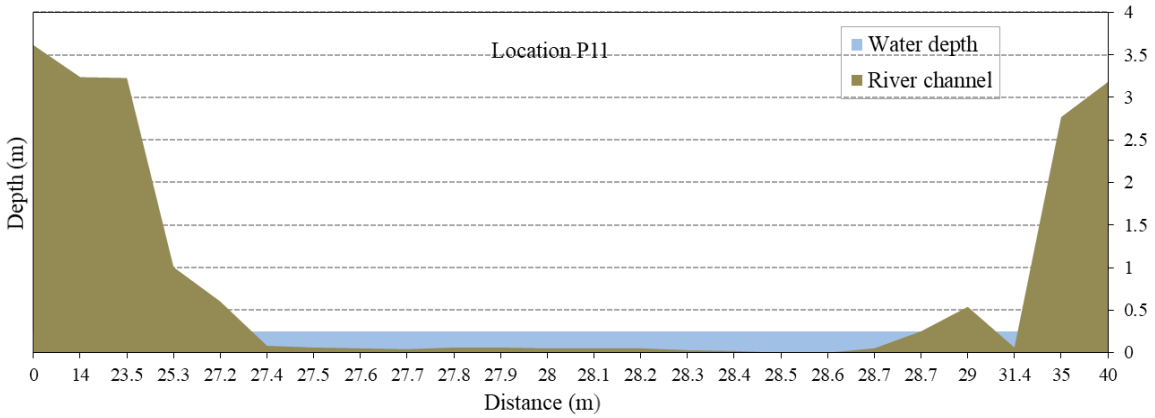


Figure 12 - Gregório Stream's cross-section at location P11 and a photograph of the area.

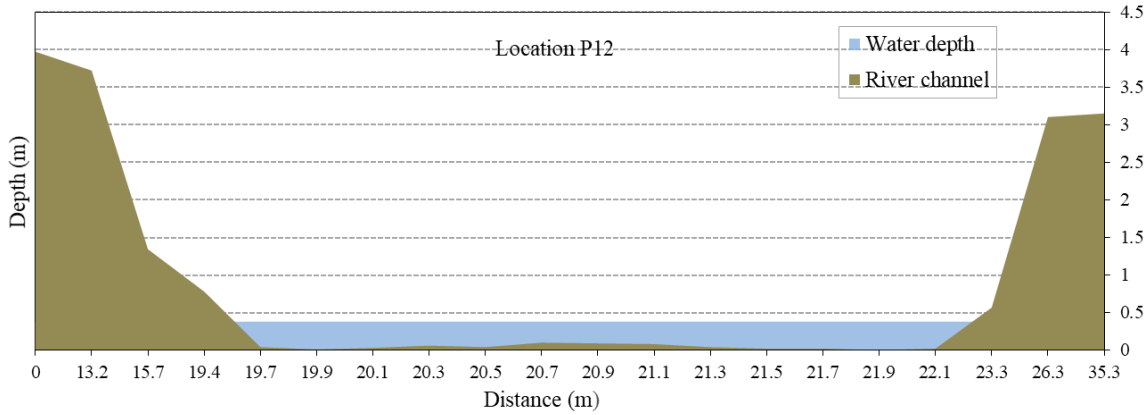


Figure 13 - Gregório Stream's cross-section at location P12 and a photograph of the area.

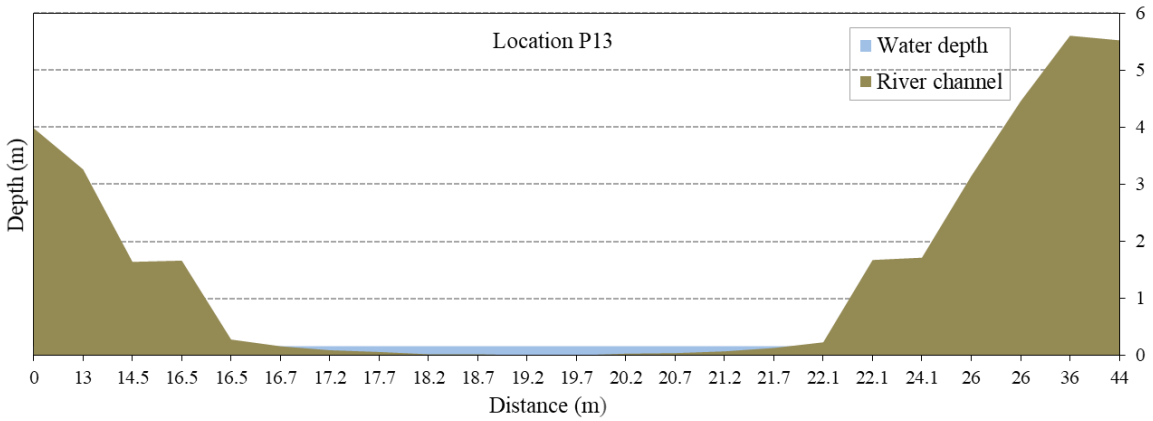


Figure 14 - Gregório Stream's cross-section at location P13 and a photograph of the area.

Appendix B

Rainfall data

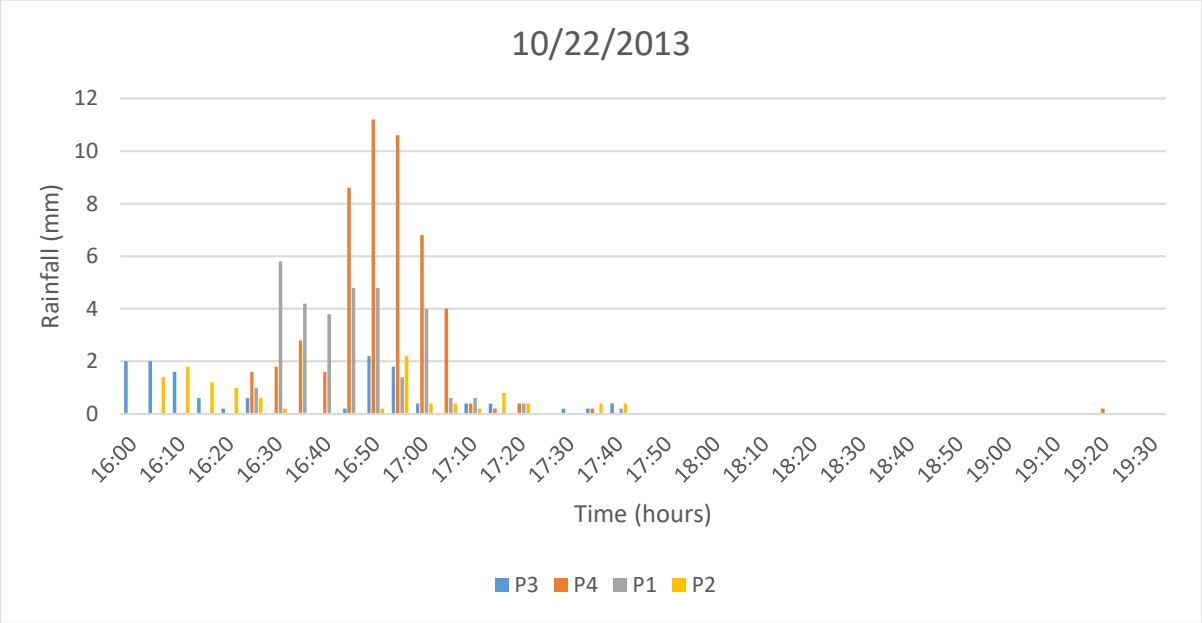


Figure 1 - Rainfall event occurred on October 22th, 2013 (Event 1).

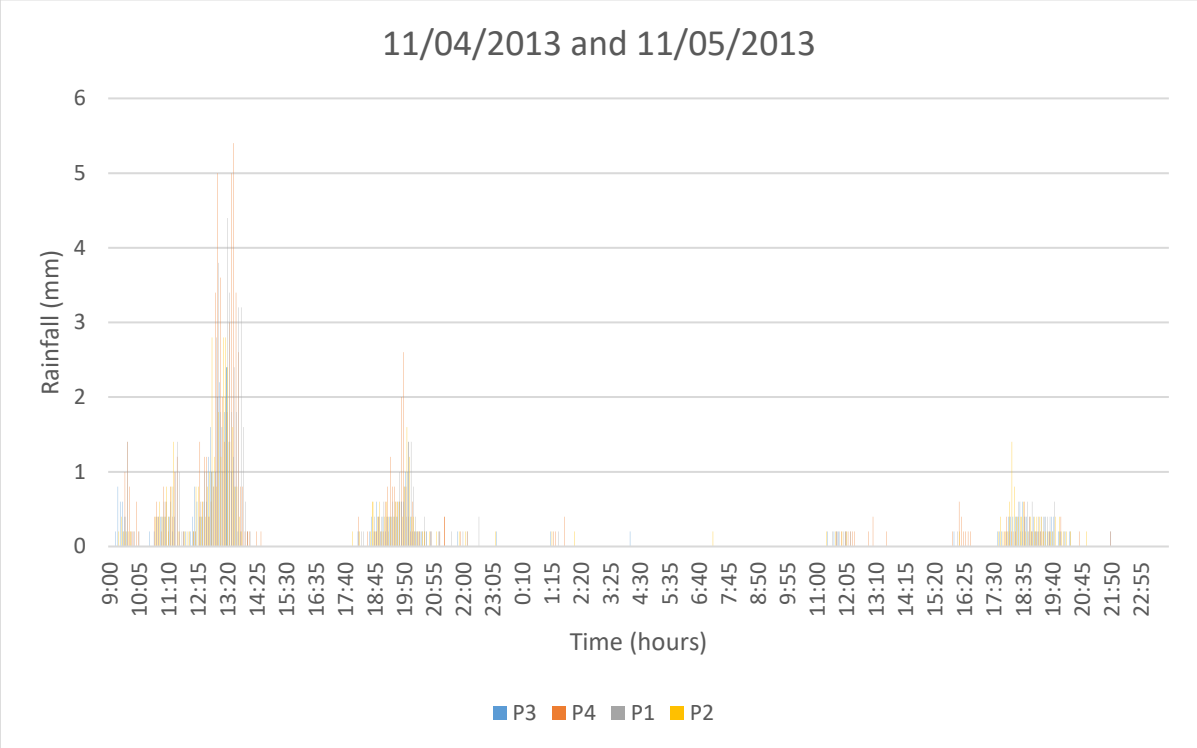


Figure 2 - Rainfall event occurred on November 4 and 5th, 2013 (Event 2).

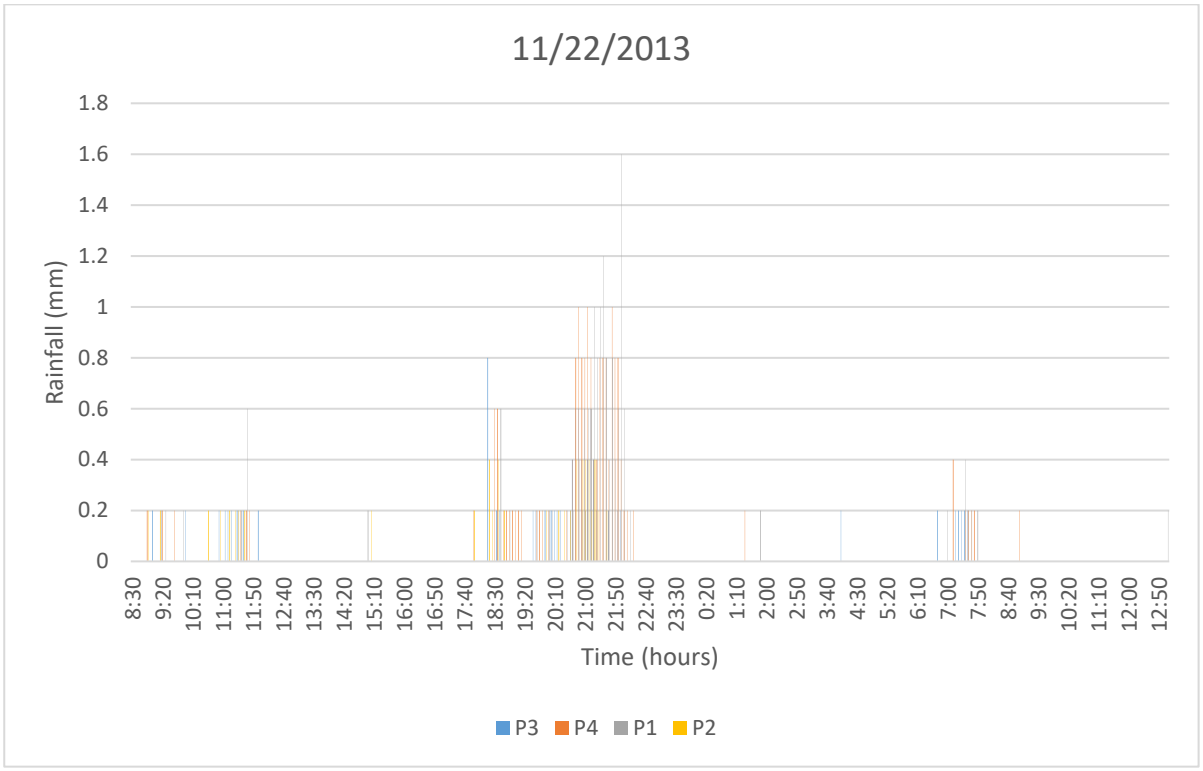


Figure 3 - Rainfall event occurred on November 22th, 2013 (Event 3).

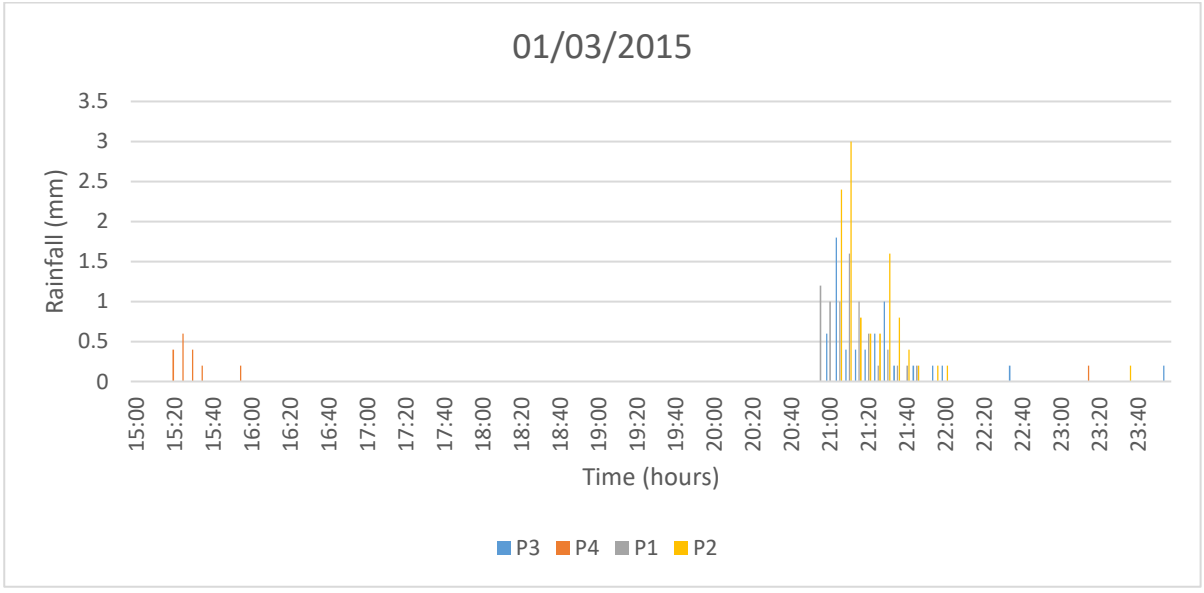


Figure 4 - Rainfall event occurred on March 3rd, 2015 (Event 4).

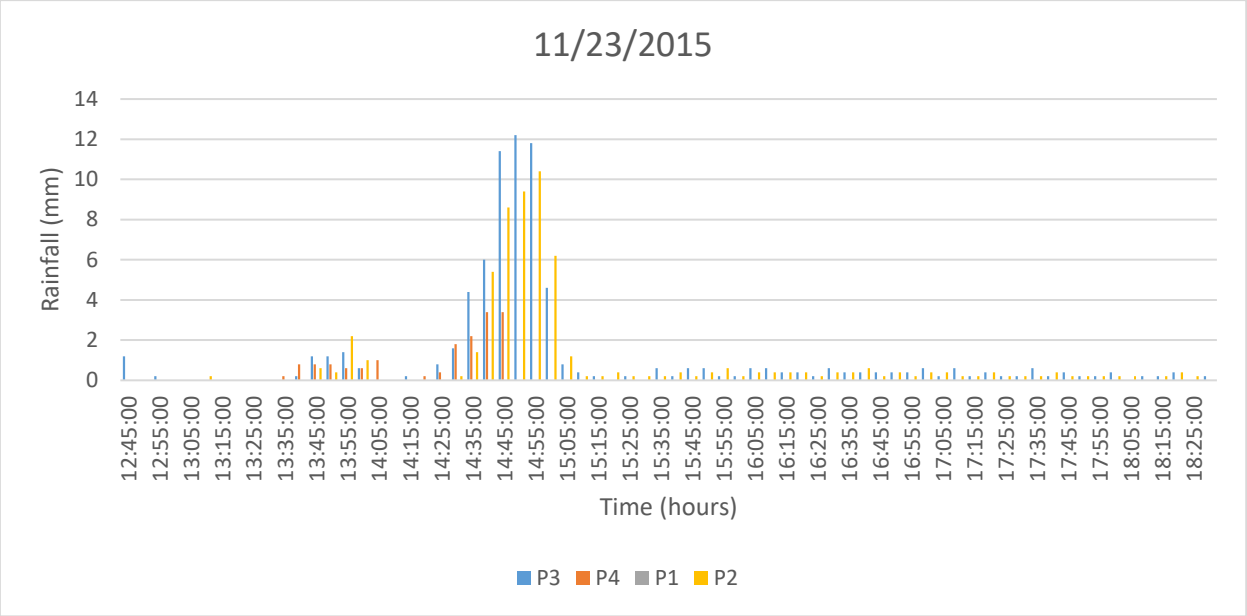


Figure 5 - Rainfall event occurred on November 23th, 2015 (Event 5).

Events	Rain Gauges	P1	P2	P3	P4
1		74.2	59.8	51.6	82.2
2		41.6	73.8	64.6	47.4
3	**Total precipitate (mm)**	16.8	6.6	10.8	19.4
4		7.6	11	6.4	2
5		0	58.8	73.2	16.2

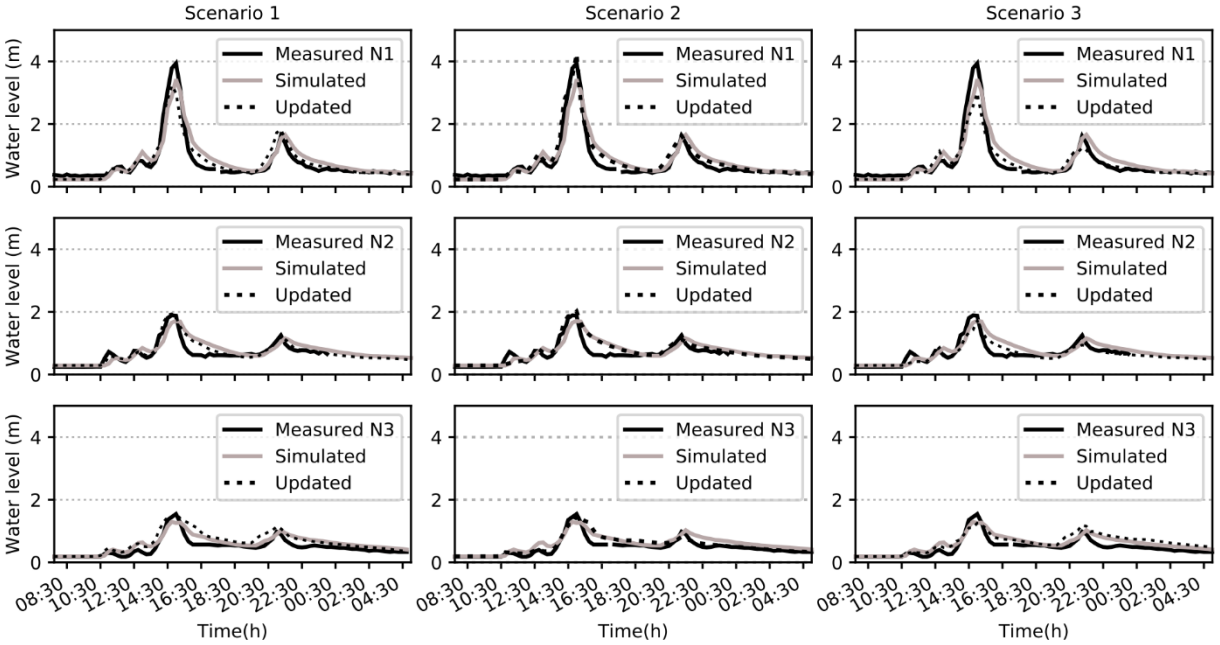
Figure 6 – Total precipitation in each rain gauge during the rainfall events.

Appendix C

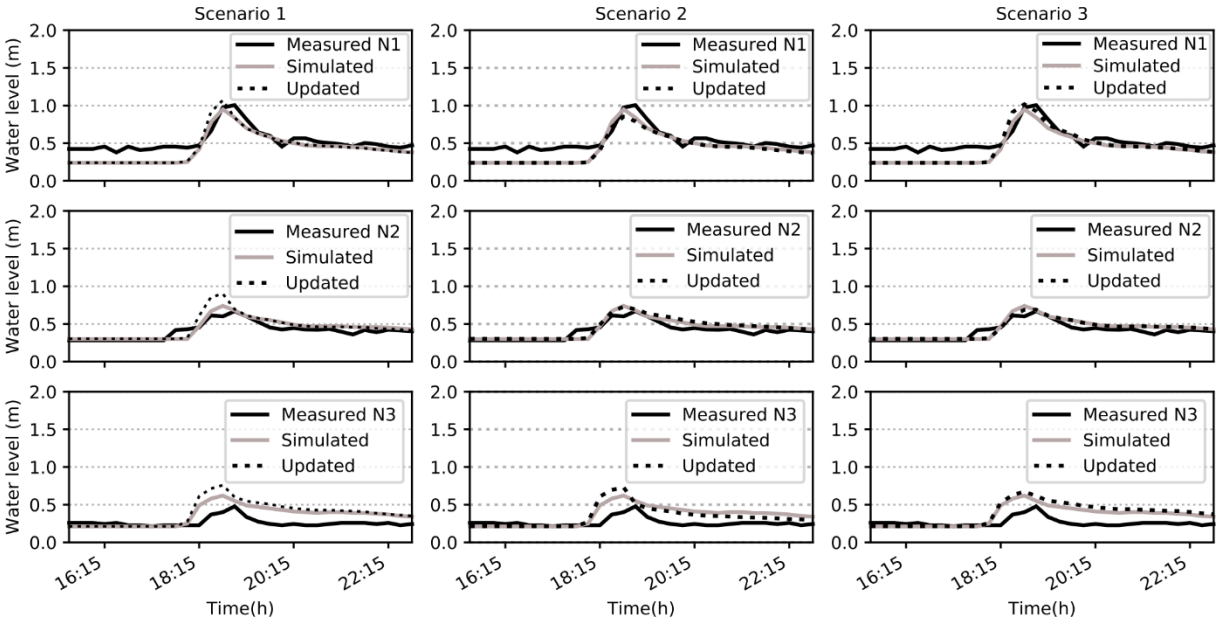
Hydrographs showing the water levels observed, simulated and resulting from updates by methods I to V of Experiment 1.

Method I

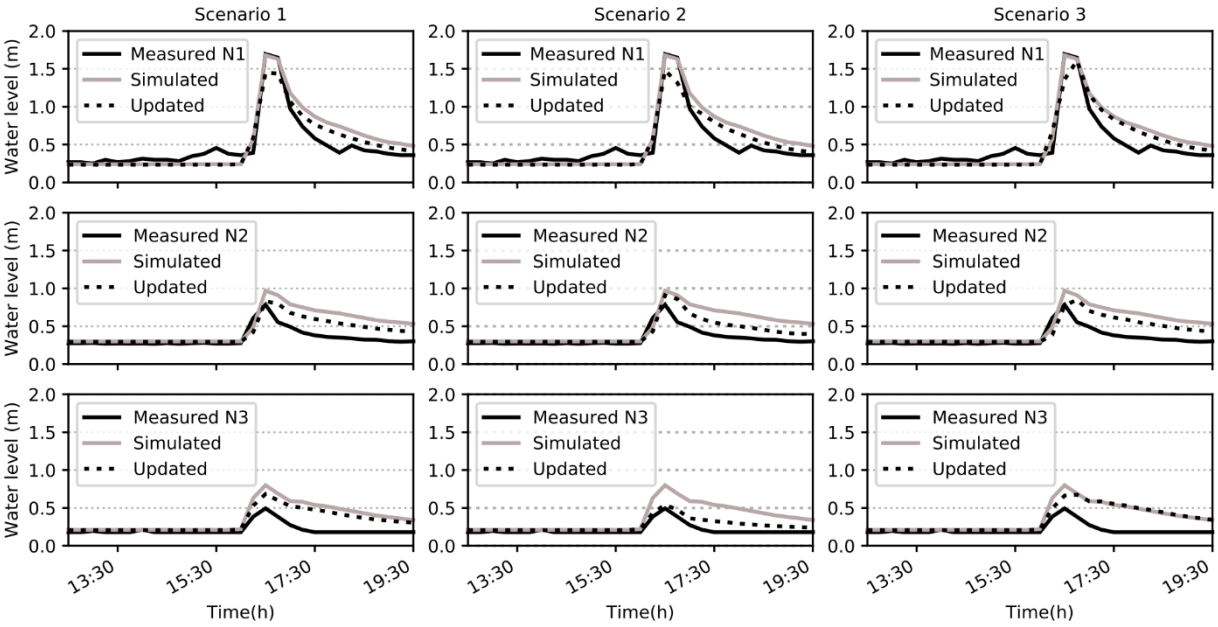
Method I – Event 1



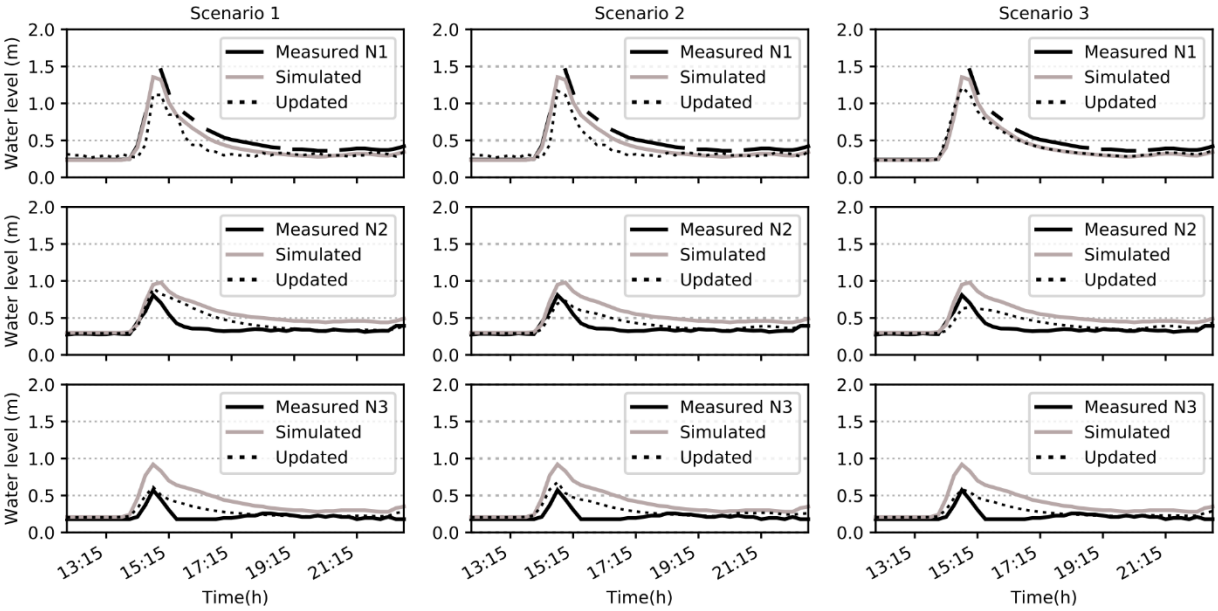
Method I - Event 2



Method I – Event 3

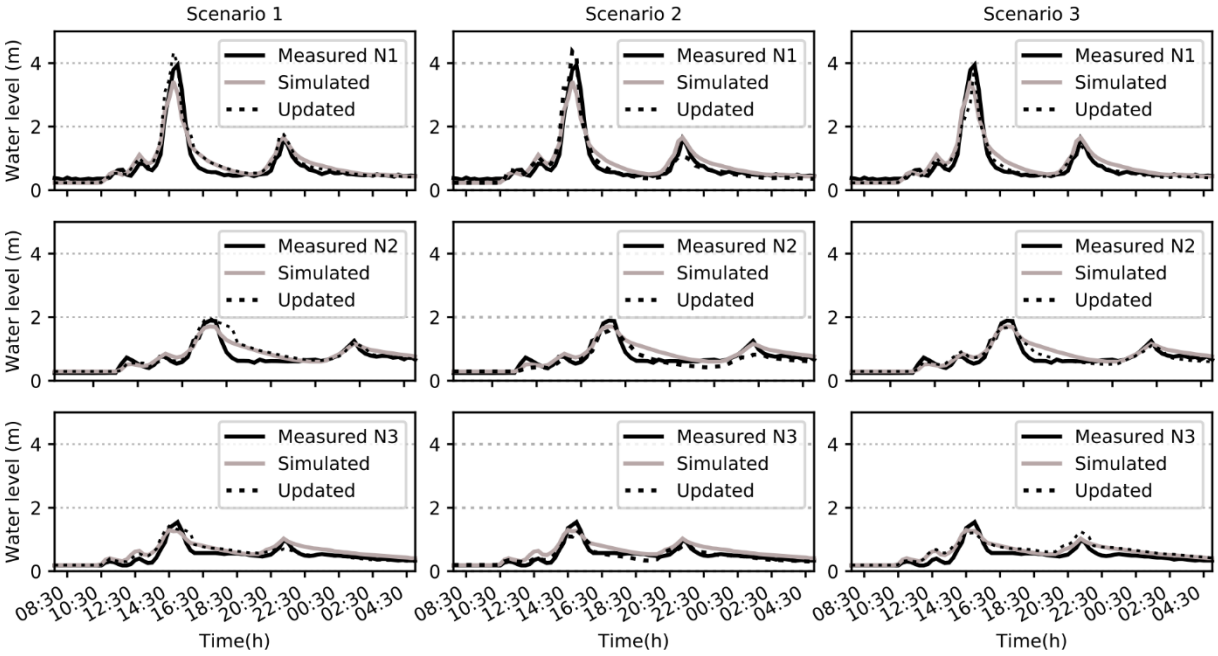


Method I - Event 4

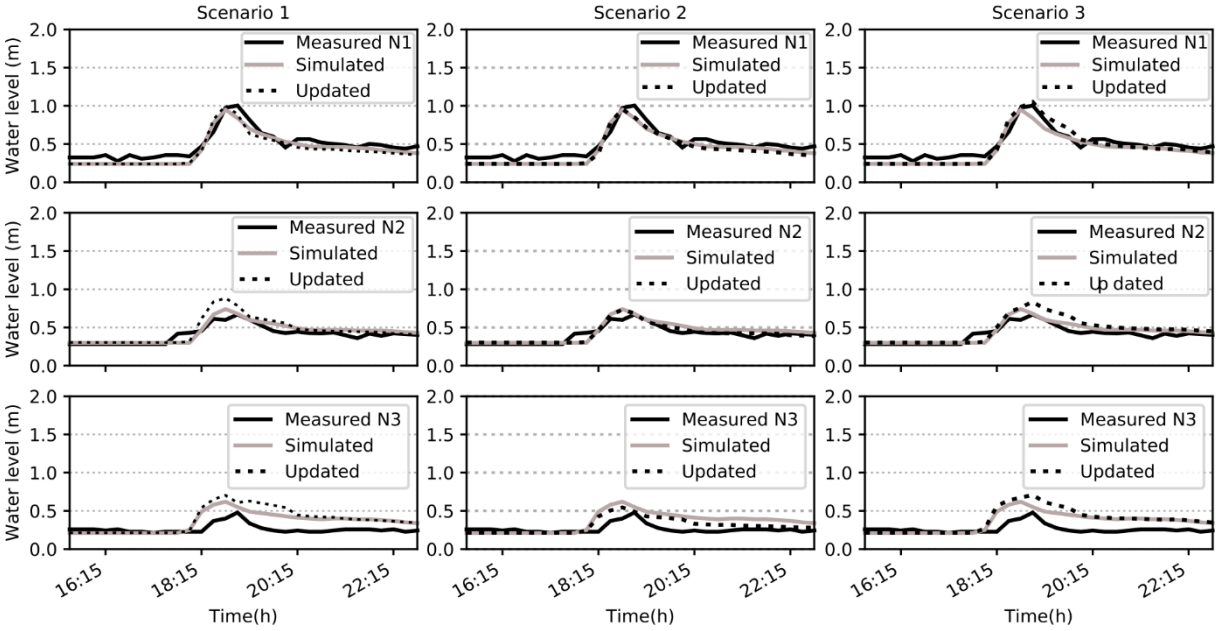


Method II

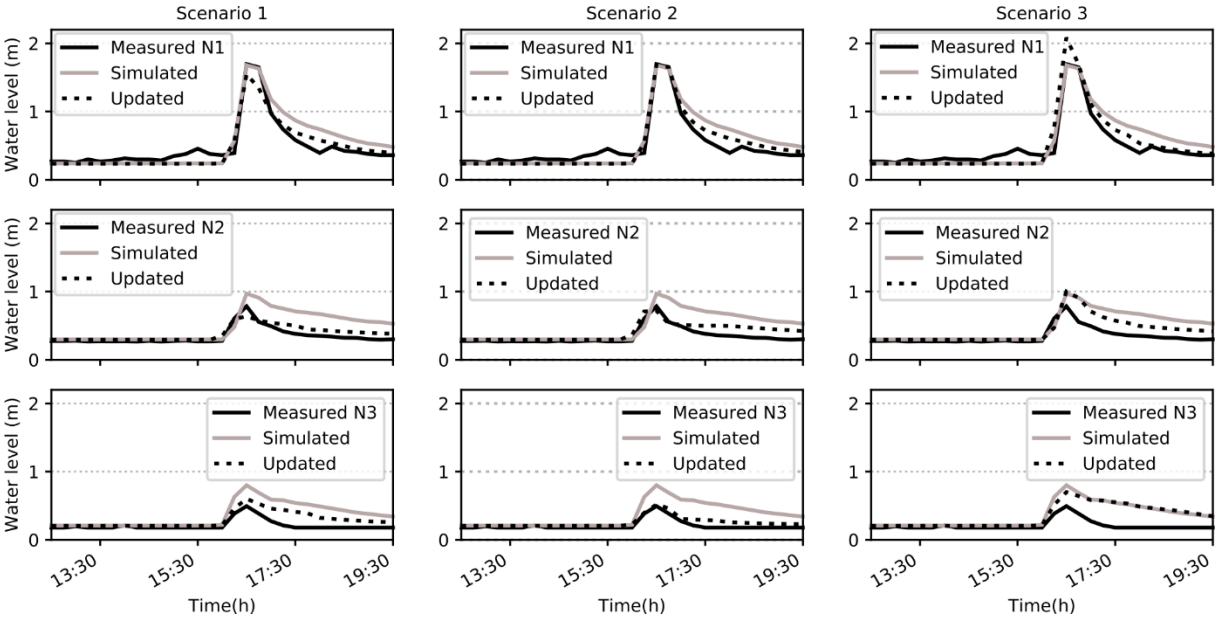
Method II – Event 1



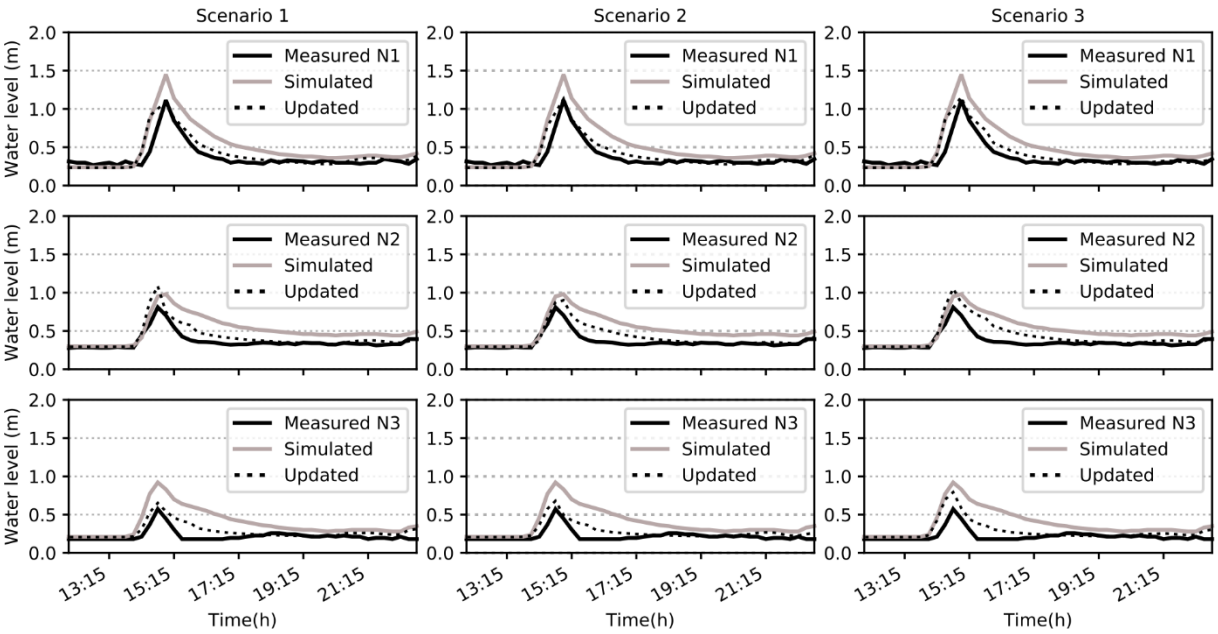
Method II – Event 2



Method II – Event 3

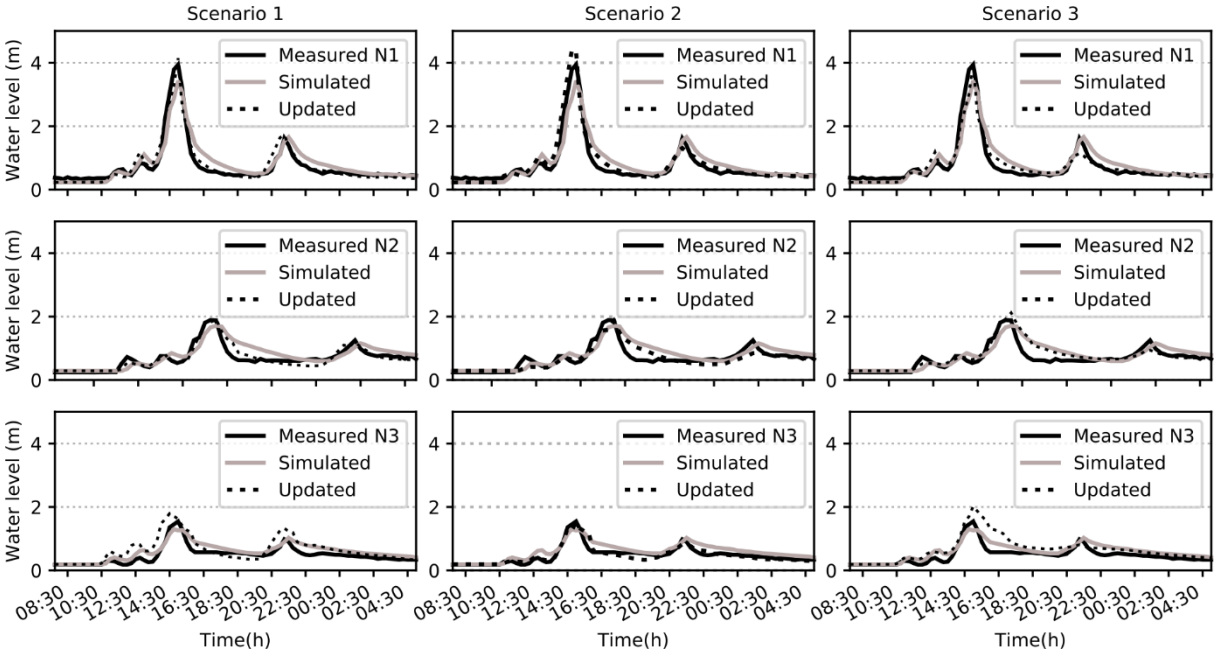


Method II – Event 4

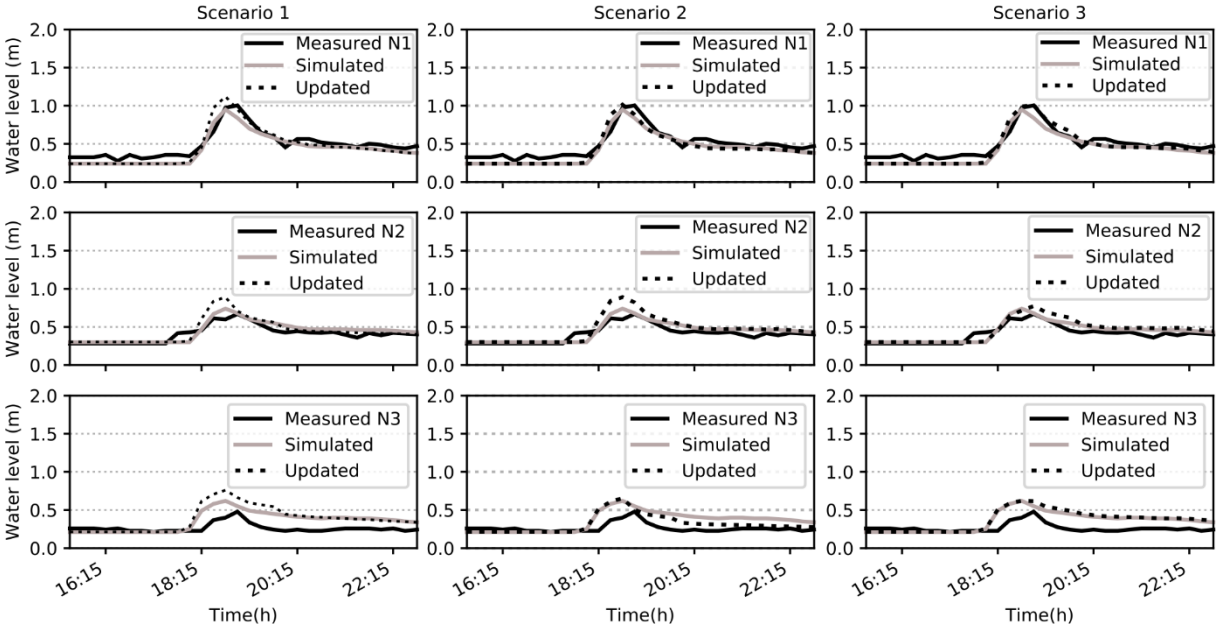


Method III

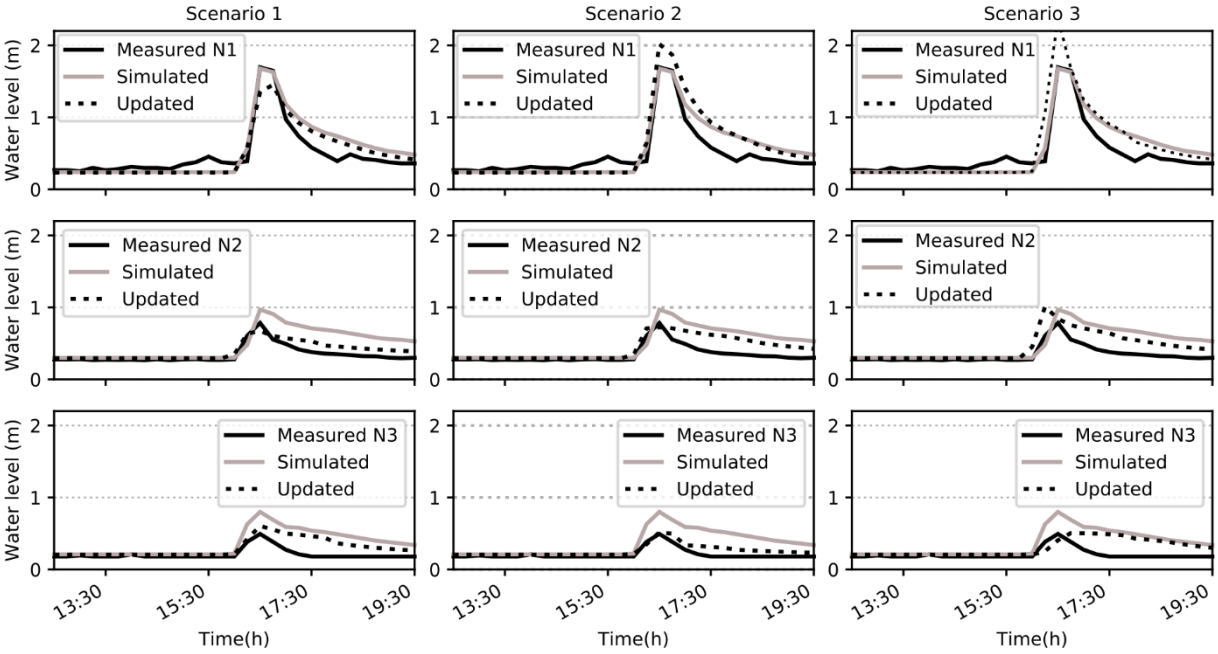
Method III – Event 1



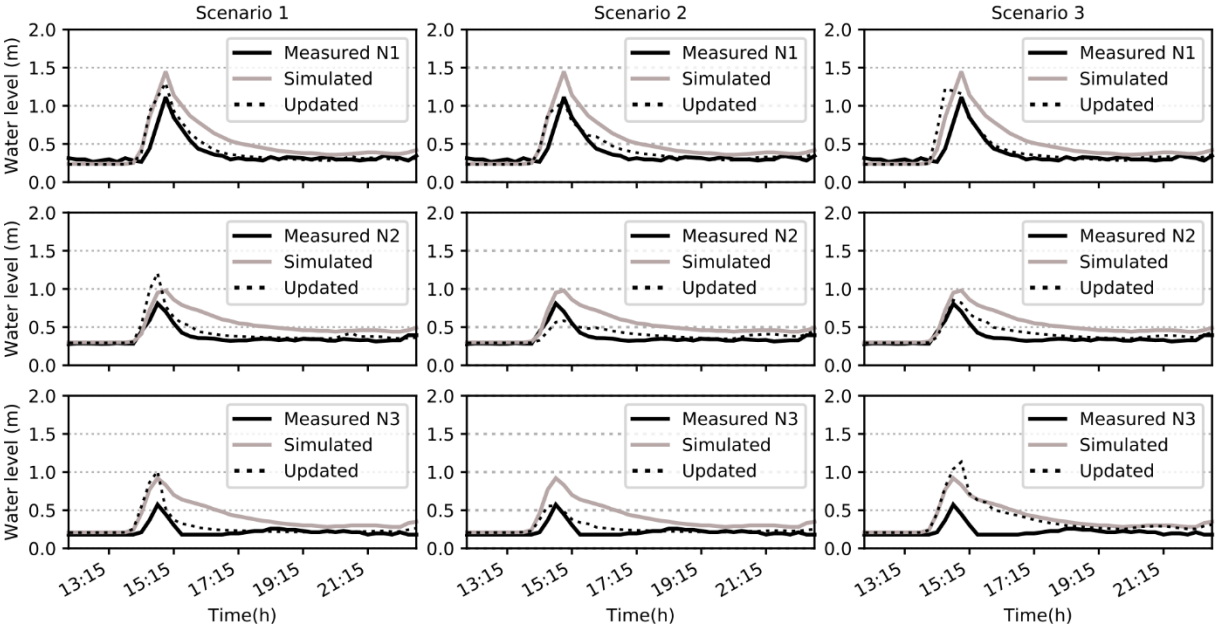
Method III – Event 2



Method III – Event 3

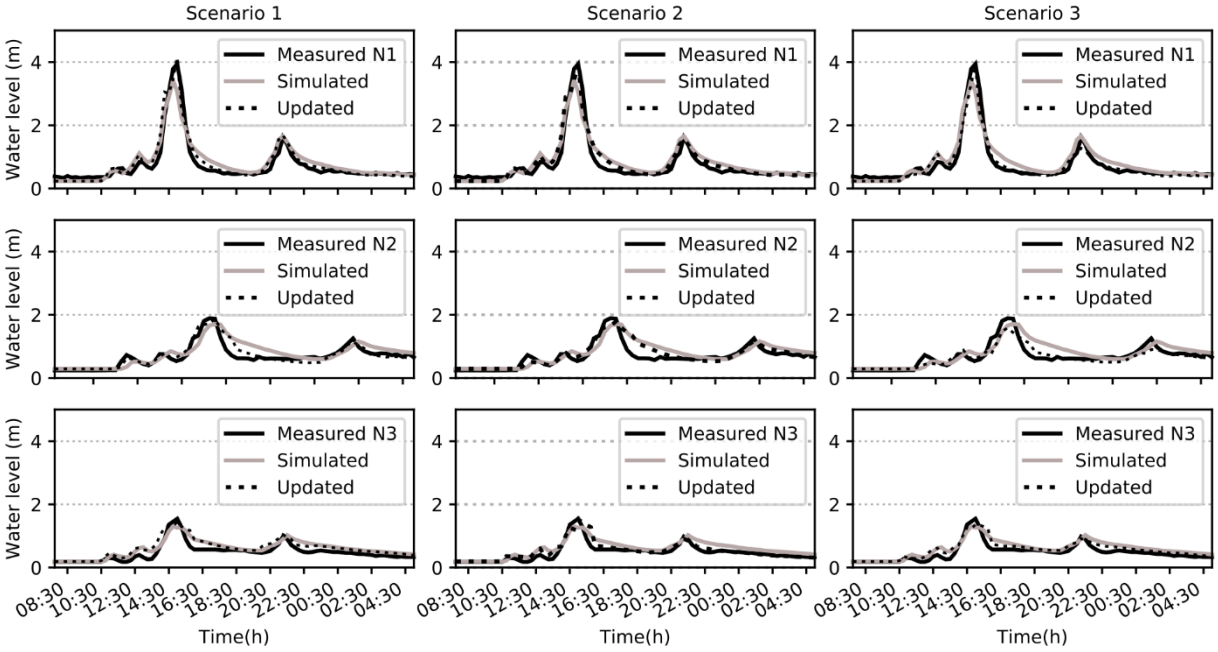


Method III – Event 4

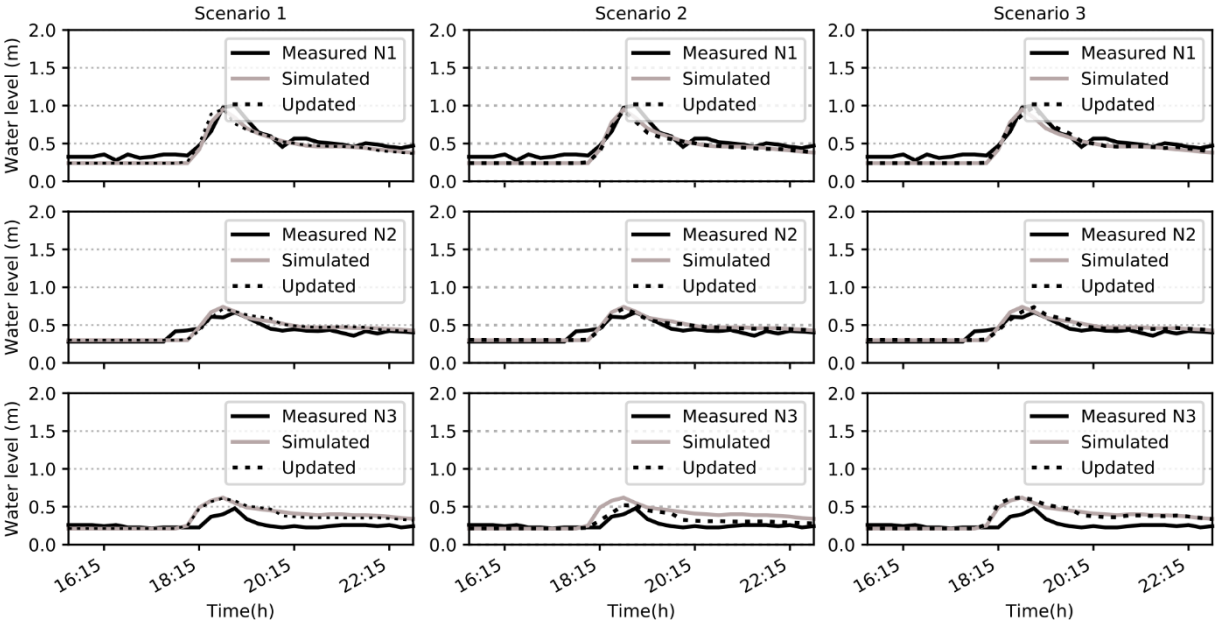


Method IV

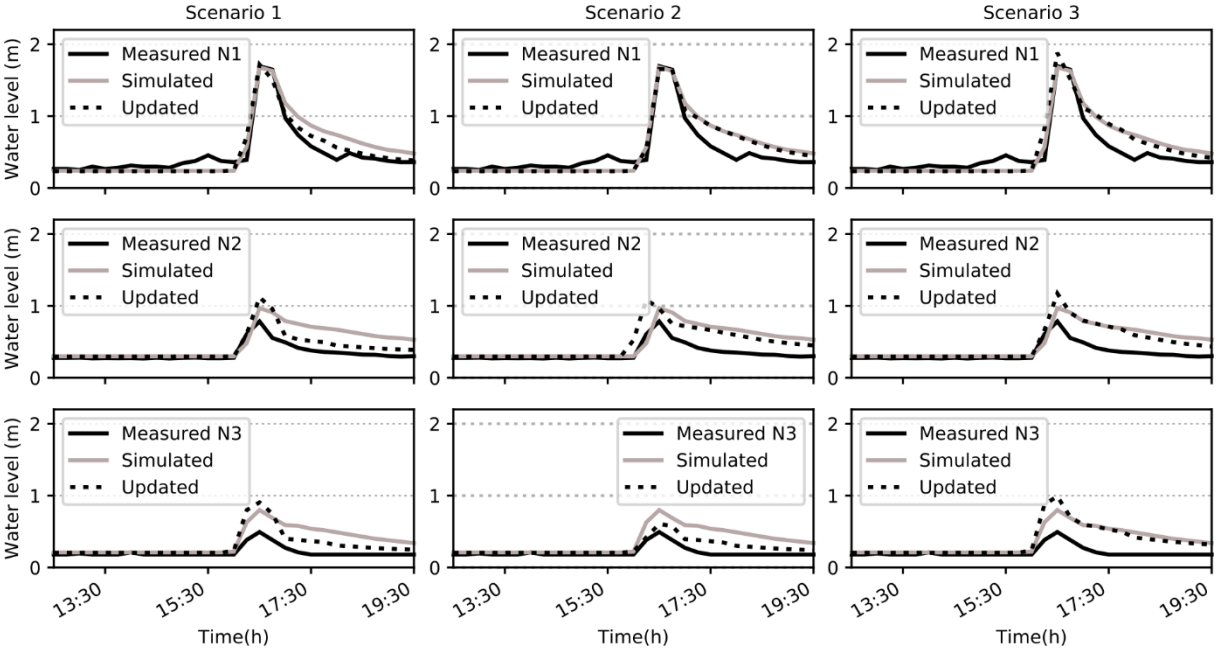
Method IV – Event 1



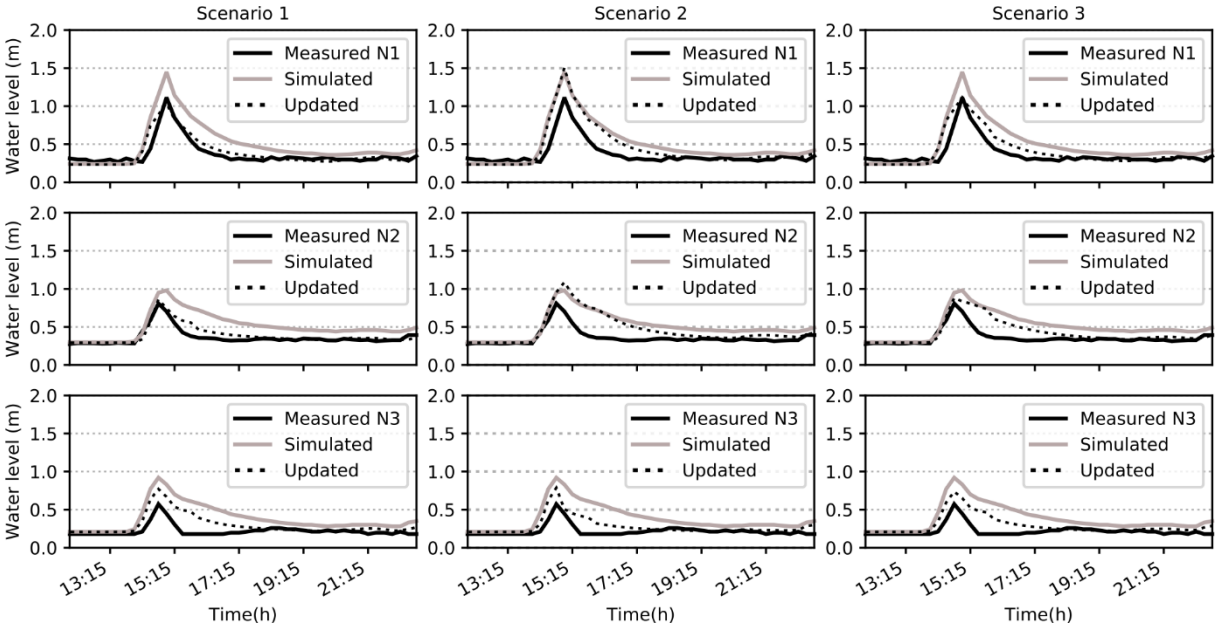
Method IV – Event 2



Method IV – Event 3

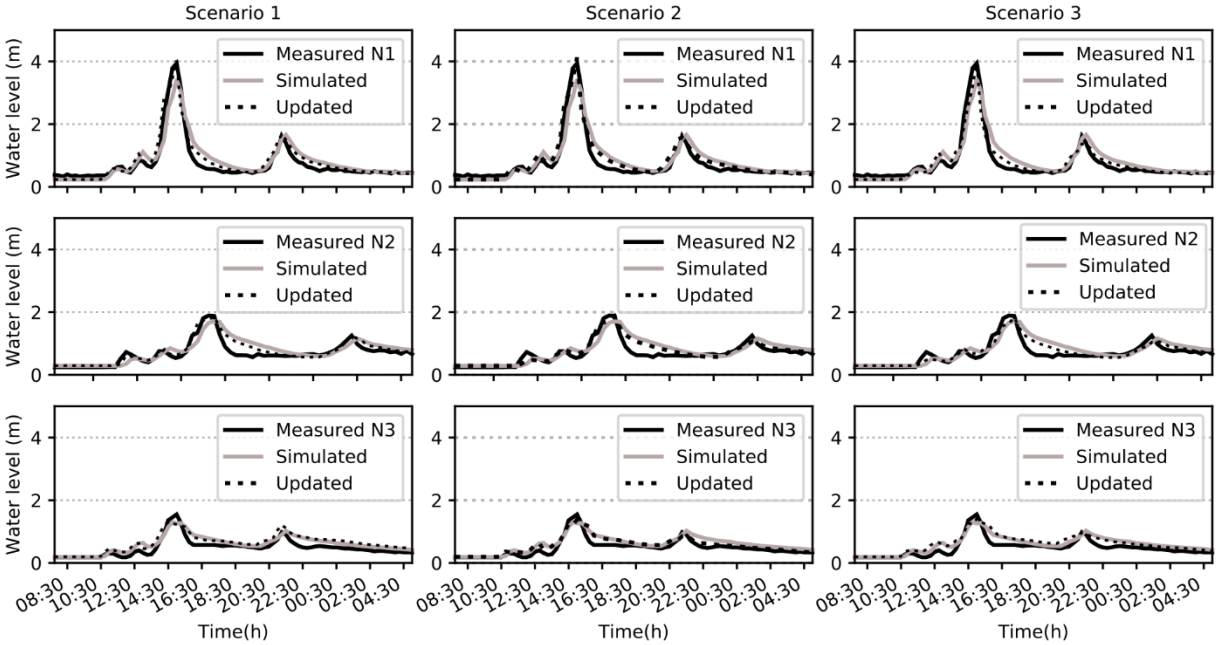


Method IV – Event 4

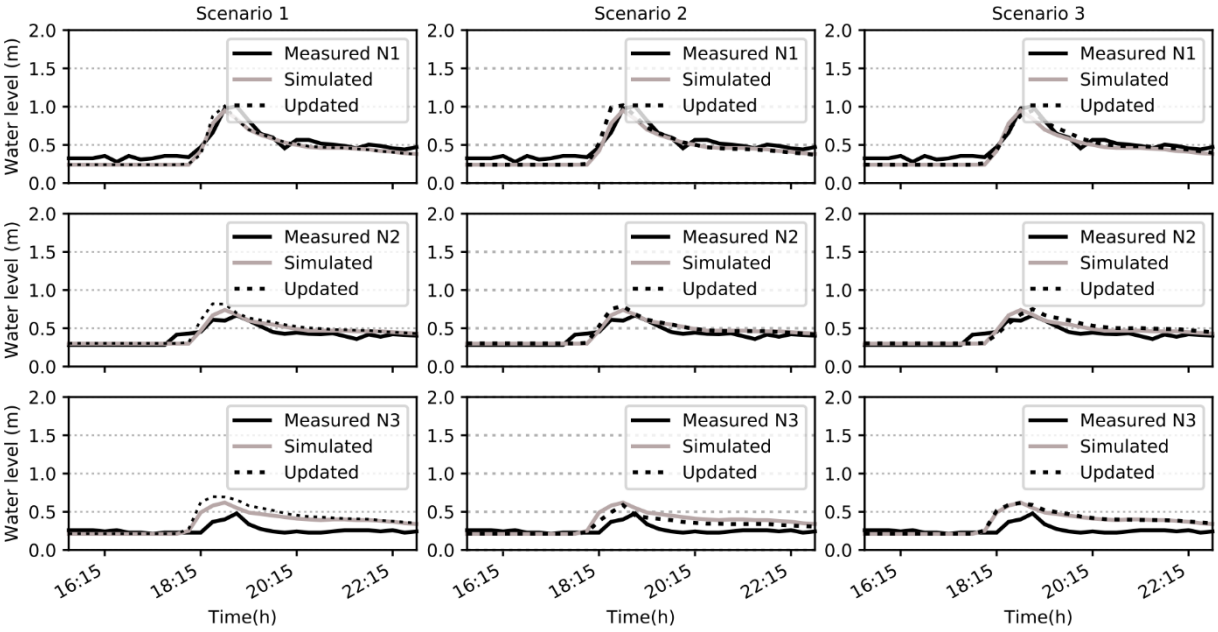


Method V

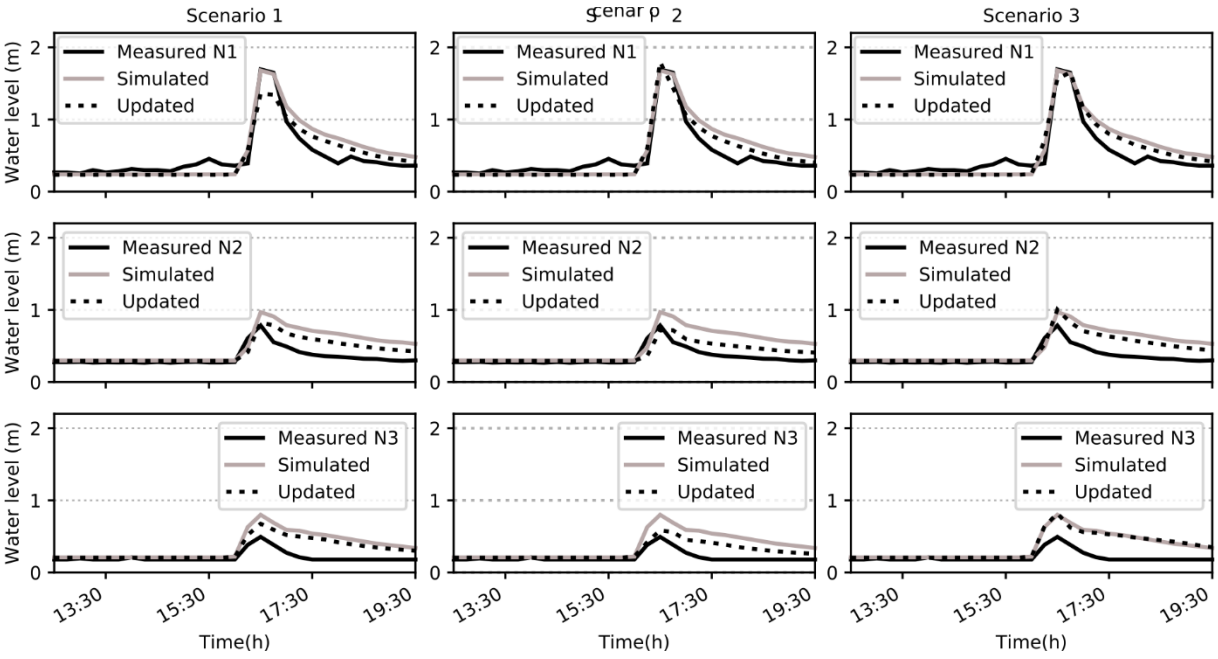
Method V – Event 1



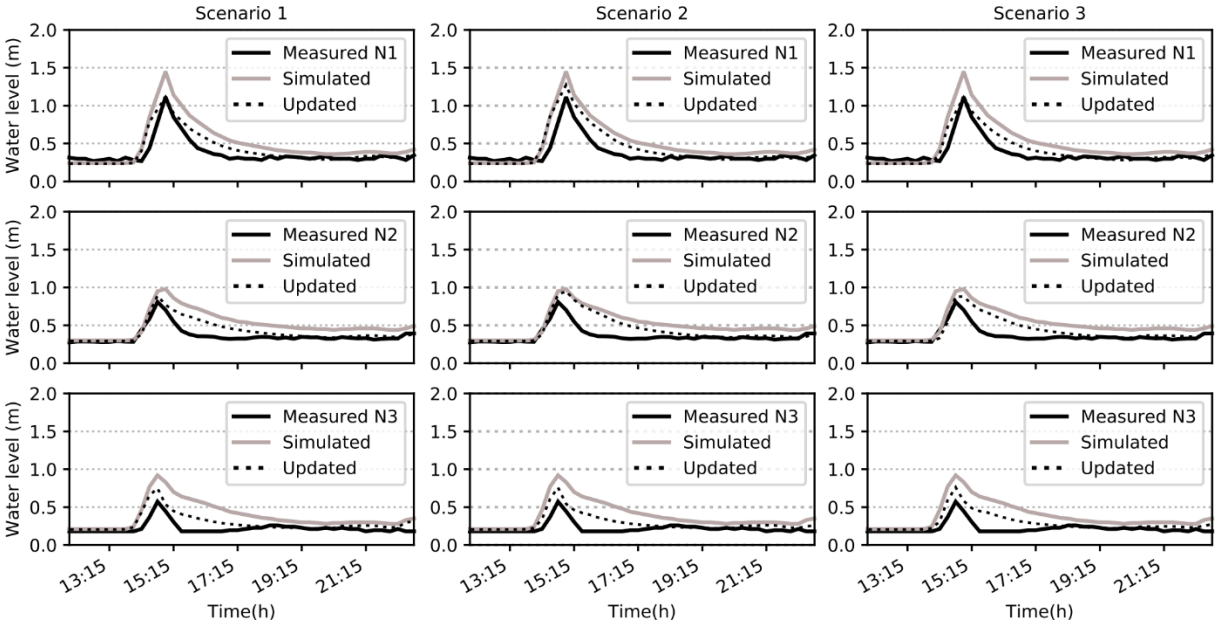
Method V – Event 2



Method V – Event 3



Method V – Event 4

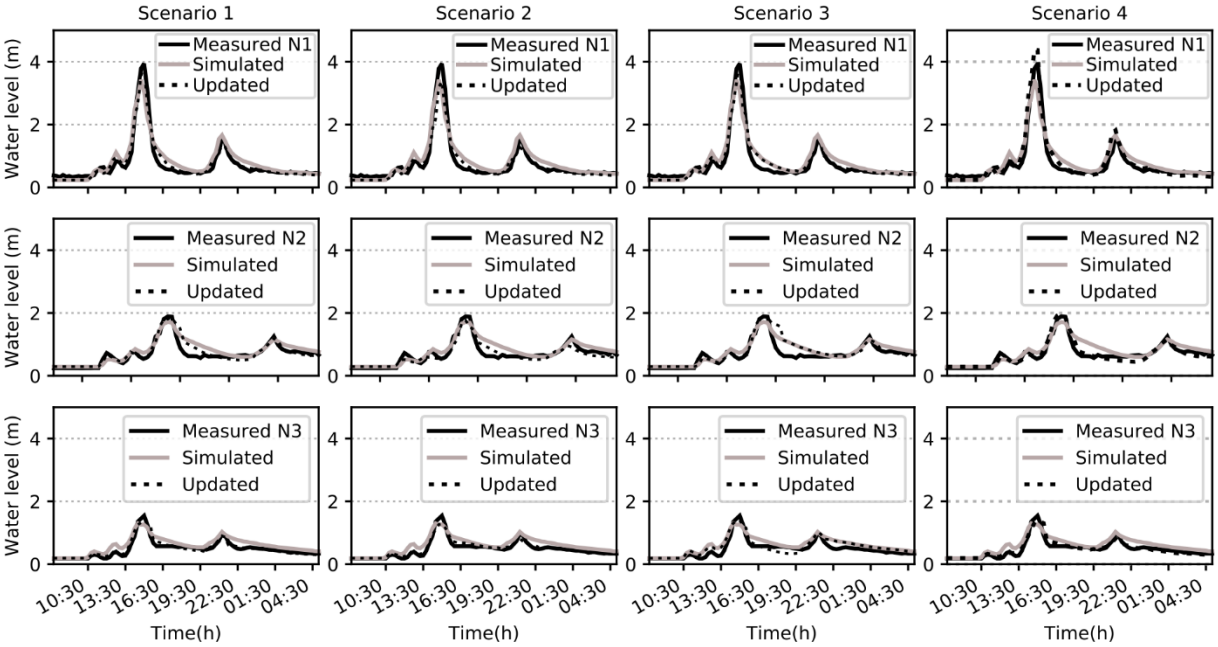


Appendix D

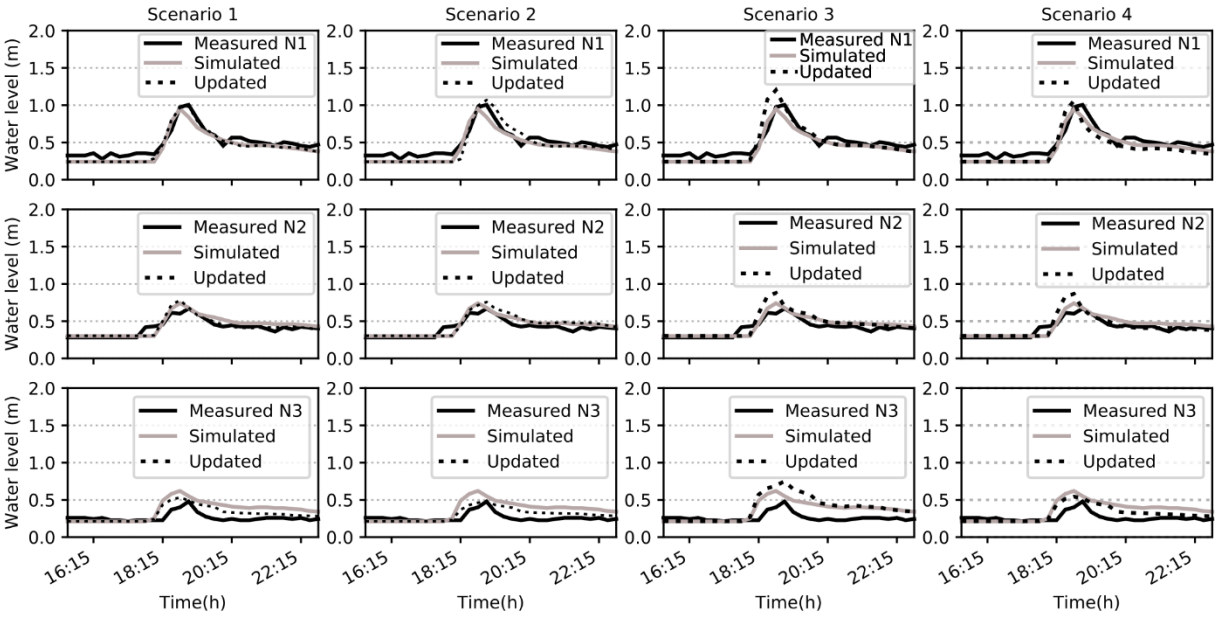
Hydrographs showing the water levels observed, simulated and resulting from updates by methods II to IV of Experiment 2.

Method II

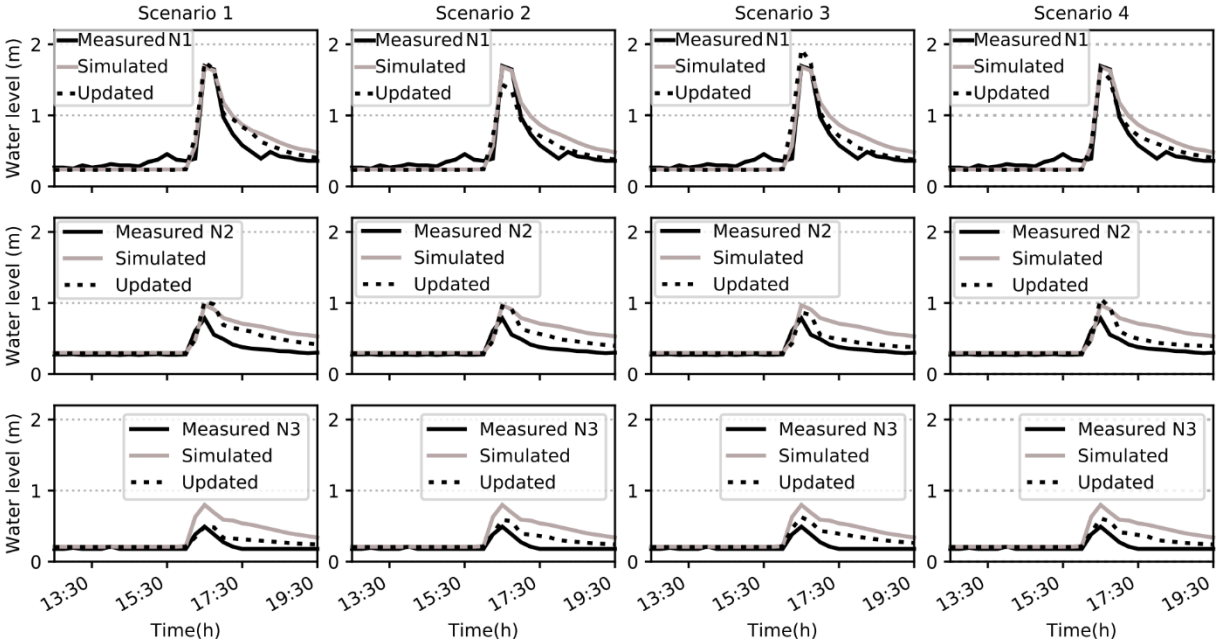
Method II – Event 1



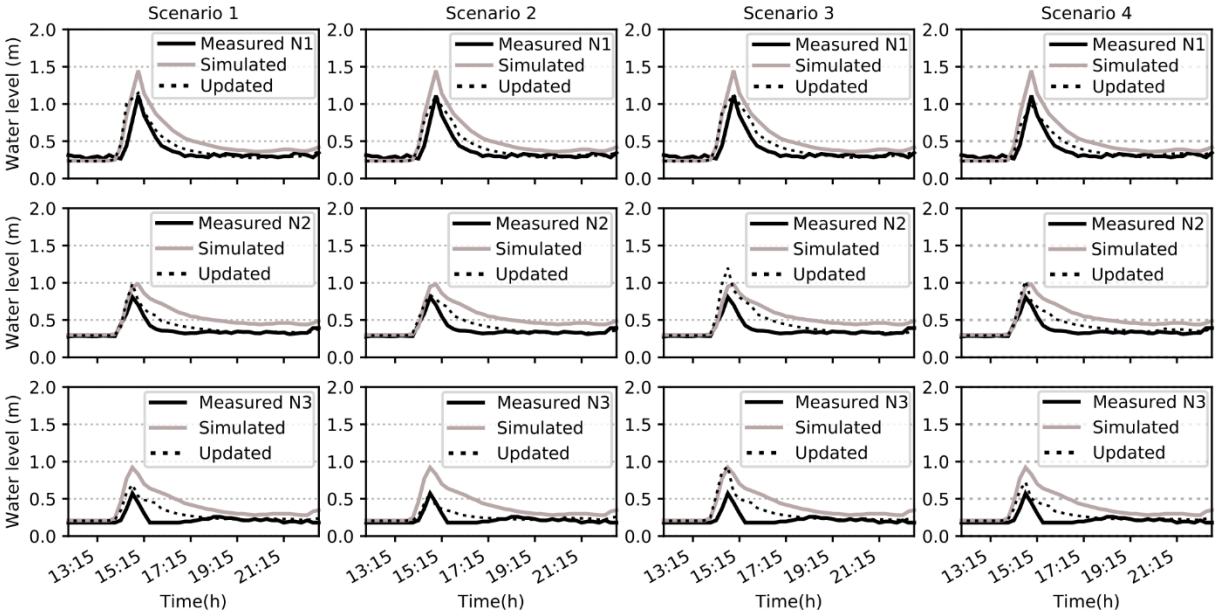
Method II – Event 2



Method II – Event 3

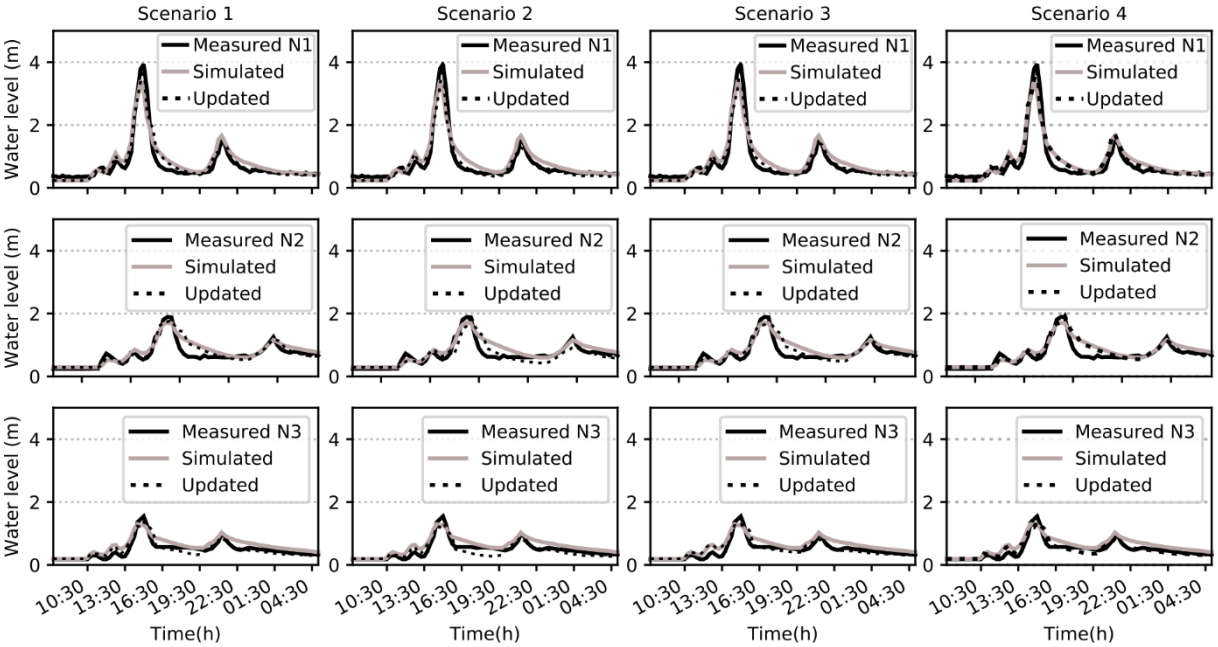


Method II – Event 4

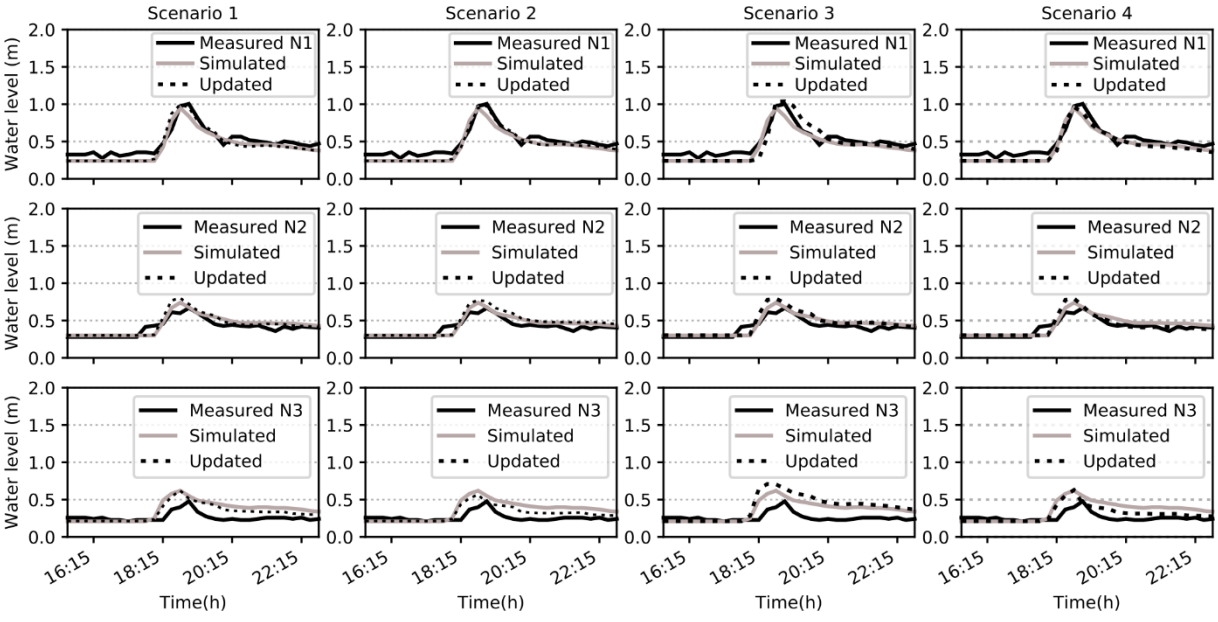


Method IV

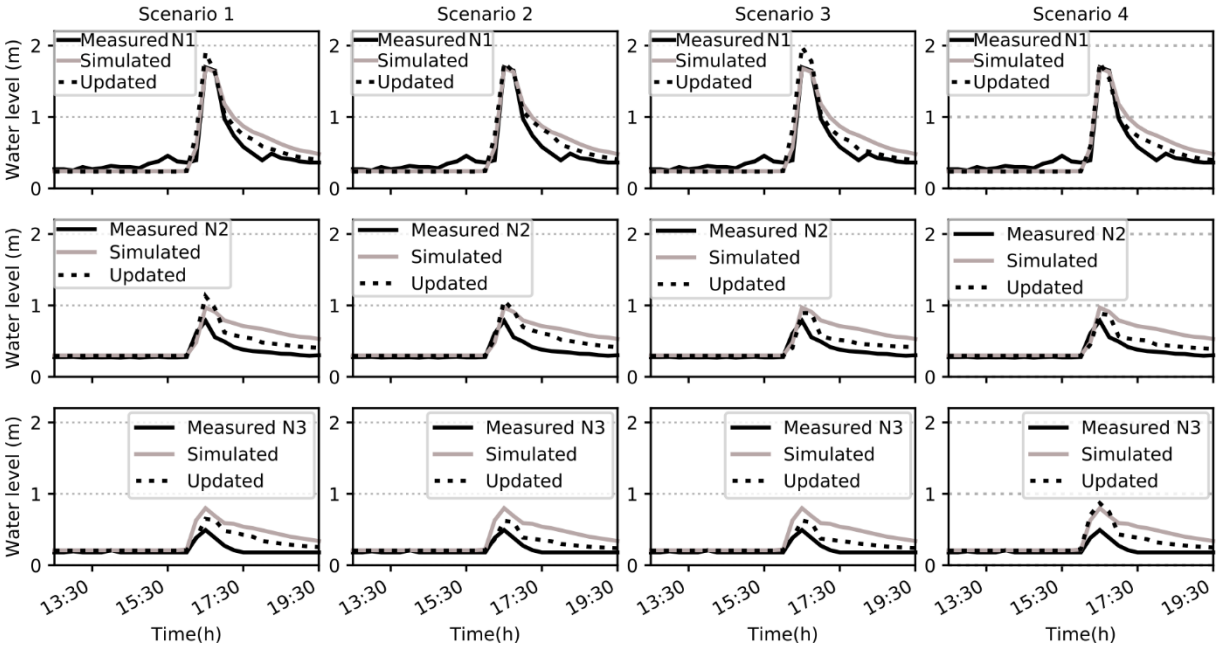
Method IV – Event 1



Method IV – Event 2



Method IV – Event 3



Method IV – Event 4