Revisiting stormwater quality conceptual models in a large urban catchment : Online measurements, uncertainties in data and models

Santiago Sandoval Arenas

N°d'ordre NNT : 2017LYSEI089

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
INSTITUT NATIONAL DES SCIENCES APPLIQUEES DE LYON

Ecole Doctorale EDA162
MECANIQUE, ENERGETIQUE, GENIE CIVIL, ACOUSTIQUE

Spécialité : **Génie Civil**

Soutenue publiquement le 10/10/2017, par :
Santiago SANDOVAL

REVISITING STORMWATER QUALITY CONCEPTUAL MODELS IN A LARGE URBAN CATCHMENT: ONLINE MEASUREMENTS, UNCERTAINTIES IN DATA AND MODELS

Révision des modèles conceptuels de qualité des eaux pluviales sur un grand bassin versant urbain : mesures en continue, incertitudes sur les données et les modèles

Devant le jury composé de :

BARRAUD, Sylvie	INSA LYON	Présidente
CLEMENS, François	TU DELF	Rapporteur
TORRES, Andres	PONTIFICIA UNIVERSIDAD JAVERIANA	Rapporteur
BONHOMME, Céline	ECOLE DES PONTS – PARISTECH	Examinatrice
VIKLANDER, Maria	LULEA UNIVERSITY OF TECHNOLOGY	Examinatrice
RENARD, Benjamin	IRSTEA LYON	Examinateur
BERTRAND-KRAJEWSKI, Jean-Luc	INSA LYON	Directeur de thèse

DEDICATION

A mis papás,

This Page Intentionally Left Blank

ACKNOWLEDGEMENTS

The execution of this thesis was the result of the author's work, with the support of other professionals. The author wishes to express his gratitude to all the people and institutions that made this accomplishment possible.

To my mentor, Pr. Jean-Luc Bertrand-Krajewski, for being a permanent guide, a special acknowledgement for the autonomy and the trust he placed on me during the development of this project. All the learnt lessons I keep are invaluable, my eternal gratitude. I am deeply grateful to all members of the jury for agreeing to read this manuscript and to participate in the defense of this thesis.

All my gratitude to COLCIENCIAS (Colombian Institution of Research) for funding my PhD studies in France under the international call 568 for international PhD studies. To the French-Danish Research Collaboration Program (financed by French Institute in Denmark) and to ECOS Nord program, for funding my scientific visit to DTU Copenhagen, Denmark and to Pontificia Universidad Javeriana, Bogota, Colombia, respectively.

Special thanks to OTHU project in Lyon, France (www.othu.org) for providing the data set and all the technical support on which this thesis is based. Thanks to the KWB research project MIA-CSO, the Austrian research projects IMW2 and IMW3, for the provided data for Chapter 2.

I will like to especially thank Dr. Jean Baptiste Aubin from INSA, Lyon, France, Dr. Luca Vezzaro from DTU, Copenhagen, Denmark, Dr. Siao Sun from Chinese Academy of Sciences, Beijing, China and MSc. Nicolas Caradot from KWB, Berlin, Germany, for their multiple scientific contributions during the development of this thesis. Thanks to Magda Monteagudo, B.A. Ashley Romero and Eng. Ross McLatchie, for the French and English translations, polishing and corrections in various parts of this Manuscript.

Thanks for your support to the other professors, staff members and colleagues in the DEEP laboratory, amongst them Dr. Gislain Lipeme Kouyi, Dr. Frédéric Cherqui, Pr. Sylvie Barraud, Dr. Hélène Castebrunet and Dr. Pascal Le Gauffre. Special thanks to Tanguy Pouzol for the nice discussions.

Thanks to my professors in my Bachelor and Master in Pontificia Universidad Javeriana, Bogota, for all the priceless tools and lessons they gave me for facing this stage of my life. I will like to express my deepest gratitude to Pr. Andres Torres for introducing me to the fascinating world of urban hydrology. Thanks for your friendship and for being a scientific role model throughout these years.

To my family and relatives, your unconditional support has been essential for my development as a researcher. Thank you for being there, encouraging me to keep on pursuing my dreams. *A Stéphanie, mi más inmensa gratitud por todo tu apoyo. Un grand merci à mes amis en France et en Europe, ainsi qu'à la famille Mirquez, pour son support inconditionnel pendant cette étape. A todos mis amigos de Colombia, quienes pese a la distancia, siempre han estado presentes.*

This Page Intentionally Left Blank

ABSTRACT

Total Suspended Solids (TSS) stormwater models in urban drainage systems are often required for scientific, legal, environmental and operational reasons. However, these TSS stormwater traditional model structures have been widely questioned, especially when reproducing data from online measurements at the outlet of large urban catchments. In this thesis, three potential limitations of traditional TSS stormwater models are analyzed in a 185 ha urban catchment (Chassieu, Lyon, France), by means of 365 rainfall events monitored online: a) uncertainties in TSS data due to field conditions (data acquisition and validation); b) uncertainties in hydrological models and rainfall measurements and c) uncertainties in the stormwater quality model structures. These aspects are investigated in six separate contributions, whose principal results can be summarized as follows:

a) TSS data acquisition and validation: (i) four sampling strategies during rainfall events are simulated and evaluated by means of online TSS and flow rate measurements. Recommended sampling time intervals are of 5 min, with average sampling errors between 7 % and 20 % and uncertainties in sampling errors of about 5 %, depending on the sampling interval; (ii) the probability of underestimating the cross section mean TSS concentration is estimated by two methodologies: Simplified Method (SM) and Time Series Method (TSM). TSM shows more realistic *TSS* underestimations (about 39 %) than SM (about 269 %). A power law describing the TSS as a function of flow rate is revealed, including higher variances of TSS for higher flow rates.

b) Hydrological models and rainfall measurements: (iii) a parameter estimation strategy is proposed for conceptual rainfall-runoff models by analysing the variability of the optimal parameters obtained by single-event (SE) Bayesian calibrations, based on clusters and graphs representations. The results are compared to traditional Bayesian calibrations obtained by SE and multi-event (ME) approaches. The new strategy shows better performances than for SE and ME in terms of accuracy and precision in validation. A single model structure might be able to reproduce at least two different hydrological conditions for the studied urban catchment; (iv) a methodology aimed to calculate “mean” areal rainfall estimation is proposed, based on the same hydrological model and flow rate data. Rainfall estimations by multiplying factors over constant-length time window and rainfall zero records filled with a reverse model show the most satisfactory results compared to further rainfall estimation models.

c) Stormwater TSS pollutograph modelling: (v) the modelling performance of the traditional Rating Curve (RC) model is superior to different linear Transfer Function models (TFs), especially in terms of parsimony and precision of the simulations. No relation between the rainfall corrections or hydrological conditions defined in (iii) and (iv) with performances of RC and linear Transfer Functions (TFs) could be established. Statistical tests strengthen that the occurrence of events not representable by the RC model in time presents a random distribution (independent of the antecedent dry weather period); (vi) a Bayesian reconstruction method of virtual state variables indicates that potential missing processes in the RC description are hardly interpretable in terms of a unique virtual available mass over the catchment that is decreasing over time, as assumed by a great number of traditional models.

RESUME

Les modèles de Rejets Urbains par Temps de Pluie (MRUTP) de Matières en Suspension (MES) dans les systèmes d'assainissement urbains sont essentiels pour des raisons scientifiques, environnementales, opérationnelles et réglementaires. Néanmoins, les MRUTP ont été largement mis en question, surtout pour reproduire des données mesurées en continu à l'exutoire des grands bassins versants. Dans cette thèse, trois limitations potentielles des MRUTP traditionnels ont été étudiées sur un bassin versant de 185 ha (Chassieu, France), avec des mesures en ligne de 365 évènements pluvieux : a) incertitudes des données dues aux conditions sur le terrain (acquisition et validation), b) incertitudes des modèles hydrologiques et des mesures de pluie et c) incertitudes des structures traditionnelles des MRUTP. Ces questions sont étudiées dans six chapitres indépendants, dont les principaux résultats peuvent être synthétisés comme suit :

a) Acquisition et validation des données : (i) quatre stratégies d'échantillonnage pendant les événements pluvieux sont simulées et évaluées à partir de mesures en ligne de MES et de débit. Les pas de temps d'échantillonnage recommandés sont de 5 min, avec des erreurs moyennes comprises entre 7 % et 20 % et des incertitudes sur ces erreurs d'environ 5 %; (ii) la probabilité de sous-estimation de la concentration moyenne en MES dans une section transversale du réseau est estimée à partir de deux méthodologies : méthode simplifiée (SM) et méthode des séries chronologiques (TSM). TSM montre des sous-estimations des MES plus réalistes (39 %) que TSM (269 %). Une loi puissance qui décrit la concentration en MES en fonction du débit a été établie, avec une variance des concentrations en MES qui augmente avec le débit.

b) Modèles hydrologiques et mesures de pluie : (iii) une stratégie d'estimation des paramètres d'un modèle conceptuel pluie-débit est proposée, en analysant la variabilité des paramètres optimaux obtenus à partir d'un calage bayésien évènement par évènement (SE), basé sur des techniques de clusters et représentations de graphes. Les résultats sont comparés aux calages bayésiens traditionnels, obtenus par SE et des calages multi-évènementiels (ME). La nouvelle stratégie de calage montre les résultats les plus performants par rapport à SE et ME, en termes d'exactitude et de précision dans la phase de vérification. Une même structure de modèle permet de représenter au moins deux groupes de conditions hydrologiques différentes pour un bassin versant urbain; (iv) une méthode pour calculer les précipitations moyennes sur un bassin versant est proposée, sur la base du modèle hydrologique précèdent et des données de débit. Les estimations de pluie moyenne par des facteurs multiplicatifs sur des fenêtres temporelles constantes et les valeurs manquantes estimées par un modèle inverse montrent les meilleurs résultats comparés à d'autres modèles d'estimation de pluie.

c) MRUTP (pollutographes) : (v) la performance du modèle traditionnel *rating curve* (RC) est supérieure à celle de différents modèles linéaires de fonctions de transfert (TF), surtout en termes de parcimonie et de précision des simulations. Aucune relation entre les potentielles erreurs de mesure de la pluie et les conditions hydrologiques définies en (iii) et (iv) et les performances des modèles RC et TF n'a pu être établie. Des tests statistiques indiquent que les évènements non-représentables par les modèles RC ou TF au cours du temps sont distribués aléatoirement (indépendance par rapport à la durée de temps antérieure); (vi) une méthode de reconstruction bayésienne de variables d'état virtuelles indique que des processus potentiellement manquants dans le modèle RC sont pratiquement ininterprétables en termes d'une masse disponible sur le bassin versant qui diminuerait avec le temps, comme nombre de modèles traditionnels l'ont supposé.

LIST OF ABBREVIATIONS AND ACRONYMS

ADWP	Antecedent Dry Weather Period
AIC	Akaike Information Criteria
AM	Adjacency Matrix
ARIL	Average Relative Interval Length
BIC	Bayesian Information Criteria
BMU	Bayesian Merged Uncertainty
CDF	Cumulative Distribution Function
CRR	Conceptual Rainfall-Runoff model
cTcSV	Strategy time-paced with constant sampling volume
cTpQ	Strategy time-paced with sampling volume proportional to flowrate
cTpV	Strategy time-paced with sampling volume proportional to RV
CTW	Constant Time Windows
CTWRev	Constant Time Windows Reverse model
DREAM	Differential Evolution Adaptive Metropolis
D/W	Dry/wet
EMC	Event Mean Concentration
FL	Flow-Limited
HI-DBM	Hypothetico-Inductive Data Based Mechanistic
INSA	National Institute of Applied Sciences
IQR	Interquartile Range
LHS	Latin Hypercube Sampling
LPU	Law of Propagation of Uncertainties
ME	Multi-Event
ML	Mass-Limited
MSRE	Mean Square Relative Error
NS	Nash-Sutcliffe efficiency coefficient
OTHU	Field Observatory for Urban Hydrology
PCA	Principal Component Analysis
Pdf	Probability Density Function
POCmod	Percentage of Coverage (modified)
RC	Rating Curve
RMSE	Root Mean Square Error
RV	Runoff Volume
S1	Model selection approach 1
S2	Model selection approach 2
SE	Single-Event
SEConditional	Single-Event (Conditional)

SISO	Single-Input Single-Output
SM	Simplified Method
SSI	Sobol's Sensitivity Index
SV	Sampling Volume
TCP	Time Constant Parameter
TF	Transfer Function
TSM	Time Series Method
TSS	Total Suspended Solids
TVP	Time Variable Parameter
vTcV	Strategy volume-paced composite sampling
VTW	Variable Time Windows
VTWRev	Variable Time Windows Reverse model
YICmod	Young Information Criteria (modified)

TABLE OF CONTENTS

INTRODUCTION 1

CHAPTER 1. CATCHMENT AND DATA 4

PART 1 TOTAL SUSPENDED SOLIDS IN URBAN DRAINAGE SYSTEMS: MONITORING, UNCERTAINTIES AND DATA ANALYSIS 7

CHAPTER 2. EVALUATION OF PERFORMANCE AND UNCERTAINTIES IN STORMWATER SAMPLING STRATEGIES BASED ON FLOW RATE AND TOTAL SUSPENDED SOLIDS TIME SERIES 9

2.1 INTRODUCTION 9

2.2 MATERIALS AND METHODS 11

2.3 RESULTS AND DISCUSSION 17

2.4 CONCLUSIONS 28

CHAPTER 3. INFLUENCE OF SAMPLING INTAKE POSITION ON SUSPENDED SOLIDS MEASUREMENTS IN SEWERS: TWO PROBABILITY / TIME-SERIES BASED APPROACHES 29

3.1 INTRODUCTION 29

3.2 MATERIALS AND METHODS 30

3.3 RESULTS AND DISCUSSION 36

3.4 CONCLUSIONS 40

GENERAL CONCLUSIONS OF PART 1 41

PART 2 UNCERTAINTY ASSESSMENT IN A CONCEPTUAL HYDROLOGICAL MODEL AND RAINFALL DATA 43

CHAPTER 4. STRATEGY FOR ASSESSING PARAMETERS OF A RAINFALL-RUNOFF MODEL BY CONNECTIVITY REPRESENTATIONS AND CONDITIONAL PROBABILITY FUNCTIONS 46

4.1 INTRODUCTION 46

4.2 METHODOLOGY 48

4.3 RESULTS AND DISCUSSION 52

4.4 CONCLUSIONS 65

CHAPTER 5. METHODOLOGY FOR IDENTIFYING THE TEMPORAL DISTRIBUTION OF ERRORS IN RAINFALL TIME SERIES 66

5.1 INTRODUCTION 66

5.2 METHODOLOGY 67
5.3 RESULTS AND DISCUSSION 75
5.4 APPLICATION OF RAINFALL CORRECTION TO IDENTIFIED EVENTS WITH IMPORTANT UNCERTAINTIES IN RAINFALL MEASUREMENTS 81
5.5 CONCLUSIONS 83
GENERAL CONCLUSIONS OF PART 2 84

PART 3 REVISITING CONCEPTUAL STORMWATER QUALITY MODELS 85
CHAPTER 6. REVISITING TIME CONSTANT VIRTUAL MASS MODELS WITH TRANSFER FUNCTIONS AND RATING CURVES 88
6.1 INTRODUCTION AND BACKGROUND 88
6.2 METHODOLOGY 89
6.3 RESULTS AND DISCUSSION 97
6.4 CONCLUSIONS 107
CHAPTER 7. REVISITING CONCEPTUAL STORMWATER QUALITY MODELS BY RECONSTRUCTING VIRTUAL STATE-VARIABLES 108
7.1 INTRODUCTION AND BACKGROUND 108
7.2 MATERIALS AND METHODS 109
7.3 RESULTS AND DISCUSSION 111
7.4 CONCLUSIONS 115
GENERAL CONCLUSIONS PART 3 116

GENERAL CONCLUSIONS AND PERSPECTIVES 117

REFERENCES 123

APPENDICES 139
1. PRESENTATION DES RESULTATS MAJEURS DE LA THESE – RESUME ETENDU EXIGE POUR UNE THESE REDIGEE EN ANGLAIS (*PRESENTATION OF THE PRINCIPAL RESULTS OF THE THESIS –LONG ABSTRACT DEMANDED FOR A THESIS WRITTEN IN ENGLISH*) 139
CHAPITRE 2 Evaluation de performance et d'incertitudes dans les stratégies d'échantillonnage d'eaux de pluie, fondée sur le débit et les séries chronologiques de charge totale de solides en suspension 139
CHAPITRE 3 Estimation de l'influence du point d'échantillonnage des matières en suspension dans une section de réseau d'assainissement 142

CHAPITRE 4 Modélisation pluie-débit : stratégie améliorée de calage et estimation des incertitudes guidée par les données 145

CHAPITRE 5 Identification d'erreurs dans des séries pluviométriques a haute résolution temporelle à travers des approches fondées sur des modèles conceptuels 150

CHAPITRES 6 et 7. Modèles conceptuels de qualité d'eaux pluviales: une révision à travers la reconstruction de variables d'état virtuelles 154

REFERENCES RESUME ETENDU 156

2. APPLICATION OF THE SOBOL'S SENSITIVITY INDEXES 158

3. COMPLEMENTARY PUBLICATIONS AND WORKS 160

Gap-filling of dry weather flow rate and water quality measurements in urban catchments by a time series modelling approach 160

4. SCIENTIFIC ACTIVITIES DURING THE THESIS 165

This Page Intentionally Left Blank

INTRODUCTION

Urban cities continue to develop and their growth accelerates intensely: the urban population all over the world has increased from 30 % to 54 % in the last 70 years. In 2008, the rural population exceeded the urban one, and for year 2050 it is expected that 66 % of the population will live in urban centers (United Nations, 2015). The urban water cycle is severely affected by this urbanization process, increasing stormwater runoff volumes due to imperviousness of the surfaces producing higher runoff peaks and lower concentration times in the catchments (Fletcher *et al.*, 2013). On the other hand, anthropogenic activities in cities generate a massive accumulation of different types of pollutants over the catchments, such as heavy metals, bacteria, hydrocarbons, sediments and nutrients (Pan *et al.*, 2013). During rainfall events, a significant amount of these pollutants are transported to the outlet of drainage systems by urban runoff (Lee *et al.*, 2002). Therefore, the alteration of stormwater natural cycle for urban catchments, jointly with the inherent production of pollutants by cities, results in a significant degradation of the water quality in receiving water bodies (*e.g.* rivers, seas, estuaries) (Goonetilleke *et al.*, 2014). Indeed, indicators related to the biological integrity of streams and riparian habitats are inversely related to the amount of impervious surfaces adjacent to them (Wu and Murray, 2003).

Consequently, legislation for regulating the quality standards of stormwater has been introduced in environmental laws, increasing the interest in urban stormwater quality (adapted from Zoppou, 2001). Many of these regulations, jointly with operational policies and urban planning strategies, have promoted the execution of measurement campaigns (Ackerman *et al.*, 2010; MOE 2003, CDEP 2004). Environmental monitoring campaigns in this context frequently include the measurement of Total Suspended Solids (TSS), as a global water quality indicator (EPA, 1983). Indeed, TSS constitutes one of the most important descriptors of stormwater quality, as many pollutants are in particulate form (*e.g.* PAHs and metals), and many other toxic substances are attached to the solid particles transported into the water matrix (heavy metals, organic substances with high tendency to sorb, etc.) (EPA, 1983). However, it is well recognized that the inherent field conditions and the instrumental settings have a direct influence on the representativeness and uncertainties of TSS measurements (Ackerman *et al.*, 2010). Therefore, appropriate data acquisition and validation methodologies for TSS measurements in drainage systems are required (Bertrand-Krajewski and Muste, 2007).

Aimed to collect more accurate, reliable and representative TSS data in this context, online monitoring has emerged as a technology for investigating the spatio-temporal variability and complexity of TSS dynamics in urban drainage systems (*e.g.* Hochedlinger *et al.*, 2006). Chapter 2 presents a study comparing the differences between the Event Mean Concentrations (EMC) obtained from different sampling strategies commonly used by practitioners (Ackerman *et al.*, 2010) to ECMs estimated from online monitoring. This comparison looks to explore the benefits of online measurements, considering that traditional sampling campaigns continue to be preferred among many practitioners in different countries (*e.g.* Ackerman *et al.*, 2010). On the other hand, field measurement conditions are claimed to be an essential uncertainty source in TSS measurements, either by sampling campaigns or online monitoring (adapted from Rossi, 1998). Therefore, Chapter 3 evaluates potential uncertainties in TSS online measurements from operation of sensors under typical conditions, specifically

exploring the influence of sampling intake position through the cross section of a sewer system on the representativeness of TSS measurements.

Mathematical models in urban drainage are recognized as a fundamental tool for purposes such as decision making, understanding of physical processes or real time control operation (adapted from Zoppou, 2001). For example, the stormwater quality standards in the European Union (Water Framework Directive, 2000/60/CE) and in the US (NPDES, Phase I (US EPA, 1990) and Phase II (US EPA, 1999)) highlight the need of better prediction models for pollutants in stormwater runoff released to the receiving water bodies.

The role of hydrology and rainfall measurements is recognized in TSS modelling, as rainfall is the driving process in the contamination of receiving water bodies by stormwater (Lee *et al.*, 2002). Rainfall data and hydrological models are also known to be embedded with high uncertainties, potentially impacting the performance of TSS stormwater models. Hence, Chapter 4 proposes an estimation strategy for parameters in a conceptual hydrological model, aimed to improve the results obtained from traditional single-event and multiple-event based calibrations. Furthermore, with the purpose of increasing the representativeness of rainfall measurements, Chapter 5 presents a methodology aimed to calculate a mean areal rainfall estimation, based on a hydrological model and flow rate data.

Apart from uncertainties in rainfall and TSS data, the TSS stormwater traditional model structures (*e.g.* SWMM in Rossman, 2010) have been widely questioned at the scale of large urban catchments, especially when reproducing data from online measurements (*e.g.* Métadier, 2011; Dotto *et al.*, 2011). Different hypotheses about why these model structures are still unsatisfactory have been explored in Chapter 6 and Chapter 7. For this purpose, many of the discussed concepts, proposed methodologies and acquired knowledge from previous Chapters are used into an exploratory frame, questioning some of the paradigms in TSS stormwater traditional models. This exploratory frame can be further contextualized by means of the following research questions:

- Are TSS online continuous time series reliable and useful for modelling purposes? (Chapters 2 and 3)

- Do these time series show bias or insufficient representativeness? (Chapters 2 and 3)

- How to better calibrate rainfall-runoff models if model parameters are event-dependent? (Chapter 4)

- If rainfall-runoff models are not satisfactory, could we assume that this is mainly due to errors in rainfall measurements and can we identify/correct them? (Chapter 5)

- Are traditional TSS models appropriate when they are used with online continuous TSS time series instead of traditional samples? (Chapter 3)

- Is there an event-dependent relation between rainfall errors and deficient performances of TSS models? (Chapter 6)

- How could we revisit/improve TSS traditional models? (Chapter 6 and Chapter 7)

The research presented in this thesis aims to fill these identified knowledge gaps, labelled as Chapters in the introduction. The Chapters of this manuscript are complementary works that are linked by the structure described in the introduction, grouped into three general topics. Therefore, these six individual contributions (Chapters 2 to 7) are grouped into three principal

Parts as follows: (i) assessing uncertainties in TSS data due to field conditions (Part 1: Chapter 2 and 3), (ii) evaluating uncertainties in hydrological models and rainfall measurements (Part 2: Chapter 4 and 5) and (iii) identifying potential improvements of the traditional stormwater quality models (Part 3: Chapter 6 and 7).

The present manuscript corresponds to a thesis based on publications, and therefore each of the Chapters is an adapted version of an article already published or in publication process (see summary of articles and conferences in Appendix 4), except for Chapter 1 (Catchment and data) and Chapter 6.

Considering the connection of the articles grouped under the three general topics (Parts), a common literature review and a description of the links between the Chapters are given in some introductory lines for each Part. Afterwards, each Chapter presents further literature review regarding its specific objectives, jointly with the developed scientific methodologies and their respective results, discussion and conclusions, as in a traditional article. Some general conclusions for each Part are provided, summarizing the results and conclusions of the Chapters and outlining their implications in other Parts of the manuscript. The general conclusions and perspectives of this manuscript are formulated from the conclusions drawn from the three Parts, linked by the outline of this document and the research questions presented in this section.

CHAPTER 1. CATCHMENT AND DATA

All the Chapters of this manuscript are mainly based on the urban catchment of Chassieu (Lyon, France), with online time series monitored between 2004 and 2011. For this reason, a general description of this experimental site and the measured data is given in this Chapter of the thesis, separately from the methodologies developed and presented in the specific "Materials and Methods" sections of Chapters 2 to 7. As the Chassieu catchment has already been described in many previous publications, Chapter 1 is mainly based on and adapted from Métadier (2011).

The Chassieu urban catchment is one of the five experimental sites of the OTHU project (Field Observatory for Urban Hydrology - www.othu.org), instrumented over the Grand Lyon territory since 2001, jointly with a pluviographic network. The OTHU is a research laboratory devoted to the installation and operation of an ensemble of measurement devices, installed in the urban drainage system of Lyon and in the receiving water bodies. The OTHU research federation groups 13 research teams, including 8 research organisations in Lyon (INSA, BRGM, CEMAGREF, ECL, ENTPE, Université Lyon I, Université Lyon II, Université Lyon III), covering different disciplines (climatology, biology, chemistry, hydrology, hydraulics, hydrogeology, public health...). One of the objectives of this observatory is to estimate the water volumes and the pollutant loads produced and released by urban areas (accumulation/wash-off linked to runoff), which is directly lined up with the problematics developed in this thesis. The site is located in the eastern part of the city, where the soils are mainly composed by fluvio-glacial deposits. A map of the localization of Chassieu on the territory of Grand Lyon (jointly with Ecully, another experimental site of the OTHU) is shown in Figure 1.

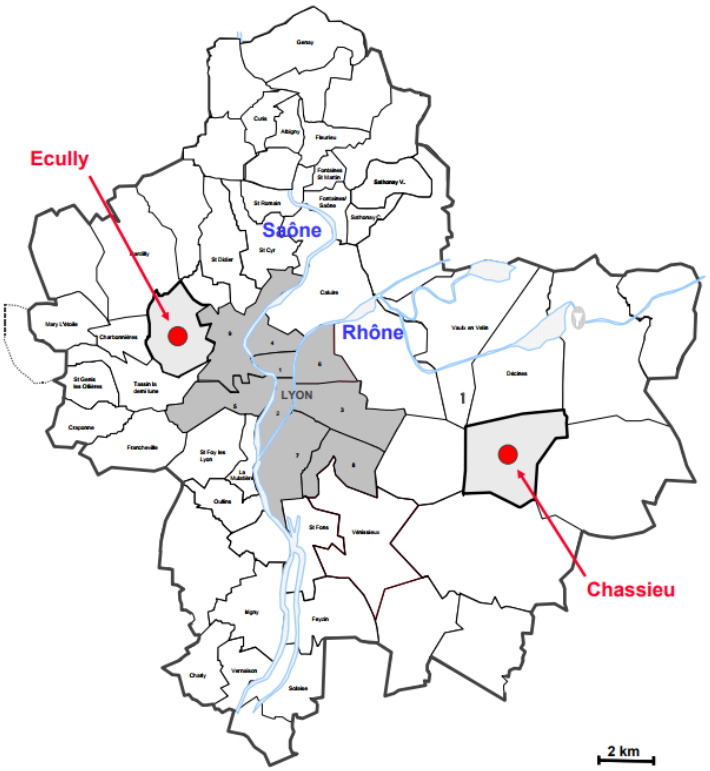


Figure 1. Localization of the Chassieu experimental sites over the Grand Lyon territory (with Ecully) (Source: www.othu.org).

The Chassieu urban catchment is a 185 ha industrial area drained by a separate storm sewer system, with imperviousness and runoff coefficients of about 0.72 and 0.43 respectively. The effective area of the catchment can be estimated in about 80 ha (0.43 × 185 ha). At the outlet of the separate storm sewer system there is a retention basin followed by an infiltration basin. The volumes of the basins are 32000 m^3 and 61000 m^3, respectively (Figure 2 right).

Figure 2. Aerial view of: the Chassieu catchment (left) (Source: www.othu.org) and the retention and infiltration basins (right) (source: Métadier, 2011).

As noticed in Figure 2, the urbanization of this catchment is relatively uniform, mainly composed of industrial facilities, parking lots and fallow fields. A few farms and natural spaces are located in the surroundings of the basin (southwest), constituting about 8 % of the total surface.

The flow rate Q (L/s) and TSS concentrations (mg/L) at the outlet of the catchment are measured in a 1.6 m circular concrete pipe with a 2 minute time-step resolution, at the outlet of the separate sewer system (inlet of the retention basin) (Figure 3).

Figure 3. Photo of the measurement station installed at the outlet of the stormwater system (inlet of the retention basin) (source: Métadier, 2011).

The discharge Q is calculated from water depth with a relative standard uncertainty from 2 % to 25 %. The water depth and the flow velocity are measured with a Nivus OCM PRO probe inside the sewer pipe. Regarding the TSS concentration, the water is pumped into an off-line monitoring flume in the shelter (measurement station) with a peristatic pump operating with an aspiration velocity of 1 m/s. Turbidity is measured by an Endress Hauser nephelometric probe CUS31. Turbidity measurements (NTU) are converted into TSS concentrations (mg/L) by local calibration functions (see a detailed description in Métadier, 2011). All probes are

connected to a central data acquisition unit SOFREL S50, which saves and sends the data by modem every night to the laboratory (see details in Dorval, 2010). The standard uncertainties in the estimated TSS concentrations range from 11 % to 30 %. The load (kg) for a time-step *t* with $\Delta t = 2$ min is calculated with Q and TSS concentration data as $load(t) = TSS(t) \cdot Q(t) \cdot \Delta t \cdot c$, where *c* is the SI units conversion factor. The uncertainties in the TSS load are estimated by the Law of Propagation of Uncertainties (LPU) (ISO, 2009), obtaining relative standard uncertainties from about 15 % to 50 %. Standard uncertainties become higher for higher values of the measurands (Q, TSS concentration, load). The rainfall is measured with the Meteo France rain gauge in Bron (next to Chassieu) from 2004 to 2006. After 2007, a specific rain gauge was installed in the Chassieu urban catchment. All rainfall measurements are registered with a time step of 1 min.

The selection of the rainfall events used for each Chapter in this thesis is mainly dependent on the specific objectives of each study and also on the period of the thesis in which the work was developed (before or after 2015). From the thesis of Métadier (2011), the data from 2007 to 2008 are particularly recommended as representative, including 89 rainfall events that were used to develop Chapters 2, 3 and 4. Sun *et al.* (2015) presented an extended version of this database, including additional information from 2008 to 2011, for a total of eight years of measurements (2004 to 2011). For the development of the subsequent Chapters (*i.e.* 4, 6 and 7), 365 events were selected from the 716 events presented in Sun *et al.* (2015). The selection was based on rainfall events with a number of missing values of TSS and Q lower than 5 %. The missing Q and TSS values were filled by linear interpolations in these cases. The following Table 1 summarizes the information used for each Chapter, including the studied variables and the number of rainfall events.

Table 1. Chassieu catchment data used in each Chapter

	Chassieu urban catchment		
	Data base /data validation	Variables	number of events
Chapter 2	Métadier, 2011	Q, TSS	89
Chapter 3	Métadier, 2011	Q, TSS	89
Chapter 4	Sun *et al.*, 2015	Q, rainfall	365
Chapter 5	Métadier, 2011	Q, rainfall	30
Chapter 6	Sun *et al.*, 2015	Q, TSS, rainfall	365
Chapter 7	Sun *et al.*, 2015	Q, TSS	255

For the case of Chapter 2, the research was developed under an international cooperation with the KWB research project MIA-CSO (Berlin, Germany) and the Austrian research projects IMW2 and IMW3 (Graz, Austria). Therefore, the data of a second experimental site of the OTHU, Ecully (Lyon, France), was complementarily included in the analyses (see further details in Table 2 or Métadier, 2011 for Ecully and Lepot *et al.*, 2017 for the other sites). Complementary works were developed during the thesis with the data of Ecully, for example: *Gap-filling of dry weather flow rate and water quality measurements in urban catchments by a time series modelling approach* in Appendix 3.

PART 1 TOTAL SUSPENDED SOLIDS IN URBAN DRAINAGE SYSTEMS: MONITORING, UNCERTAINTIES AND DATA ANALYSIS

Measurement of Total Suspended Solids (TSS) in urban drainage systems are required for scientific, legal, environmental and operational reasons, as particulate matter constitutes a major source of surface water contamination (Ashley *et al.*, 2004; Chebbo and Gromaire, 2004). Therefore, appropriate data acquisition and validation methodologies for TSS measurements in urban drainage systems are necessary (Bertrand-Krajewski and Muste, 2007). TSS monitoring of stormwater is aimed to collect the most accurate, reliable and representative data, given the technical and resource limitations (Ackerman *et al.*, 2010). Measuring stormwater TSS in urban drainage systems with online technologies (*e.g.* turbidity, UV-VIS spectrometry) has been considered as a powerful technique for investigating into the spatio-temporal variability and complexity of water quality (*e.g.* Gruber *et al.*, 2005; Hochedlinger *et al.*, 2006; Schilperoort, 2011; Métadier and Bertrand-Krajewski, 2012).

However, associated calibration, operation and maintenance costs of online probes make monitoring of stormwater TSS by traditional sampling campaigns a still widely used approach (see Athayde *et al.*, 1983; Saget, 1994; Duncan, 1999; Pitt *et al.*, 2004; Brombach *et al.*, 2005; Ellis *et al.*, 2006). Therefore, Chapter 2 presents a comparative study with four datasets from different urban drainage systems (Chassieu-France, Ecully-France, Berlin-Germany, Graz-Austria) that seeks to highlight the benefits of online monitoring in terms of maximizing the representativeness (estimated by the Event Mean Concentration - EMC -) of the dynamics of TSS, compared to EMCs obtained by four different sampling strategies proposed in the literature (*e.g.* Rossi, 1998; Bertrand-Krajewski *et al.*, 2000; Ackerman *et al.*, 2010). EMCs from sampling strategies are simulated by sampling TSS time series (approx. one minute time-step, with about one year of data) and calculating a weighted average of the samples by their sampling volumes. These “simulated” EMCs are compared to the “reference” EMC calculated as a weighted average of the complete time series (flow rate and TSS).

On the other hand, the main sources of error in any sampling procedure are not only due to the heterogeneity in both space and time of the sampling target but also due to the sampling technique (Paakkunainen *et al.*, 2007). Therefore, aiming to estimate data quality, intensive investigations have also been carried out in the assessment of uncertainties in online and laboratory TSS measurements (*e.g.* Joannis *et al.*, 2008; Métadier and Bertrand-Krajewski, 2011). However, the influence of field sampling conditions (*e.g.* sampling intake position, sampling flow velocities or sampling pipe orientation) on the representativeness of TSS measured values has not been equivalently addressed in the literature (Shelley, 1977; Berg, 1982; Rossi, 1998; Larrarte and Pons, 2011). Aiming to assess uncertainties in the mean TSS concentration due to the influence of sampling intake vertical position and vertical concentration gradients in a sewer pipe, Chapter 3 proposes two methods: a simplified method (SM) based on a theoretical concentration vertical profile and a time series grouping method (TSM). SM is based on flow rate and water depth time series. TSM requires additional TSS time series as input data. The analyzed time series for this Chapter 3 are from the urban catchment of Chassieu in Lyon, France (2 min time-step and 89 rainfall events measured in 2007).

The objective of this Part 1 within the general frame of the Manuscript is to give a better understanding of the influence of elements such as the temporal resolution (Chapter 2) and the uncertainty sources (Chapter 2 and Chapter 3) over the TSS data to be used as the main modelling input of TSS loads intra-event dynamics, giving potential elements to rethink traditional conceptual models (Part 3).

CHAPTER 2. EVALUATION OF PERFORMANCE AND UNCERTAINTIES IN STORMWATER SAMPLING STRATEGIES BASED ON FLOW RATE AND TOTAL SUSPENDED SOLIDS TIME SERIES

Extended version of:

Santiago S., Bertrand-Krajewski J.L., Caradot N., Hofer T., and Gruber G. (2017). Evaluation of performance and uncertainties in stormwater sampling strategies based on flow rate and total suspended solids time series. *Proceedings of the 14th International Conference on Urban Drainage*, Prague, Czech Republic, 10-15 September, 3 p.

2.1 INTRODUCTION

Accurately assessing stormwater pollutants through monitoring programs is essential for operative, legal, political and environmental requirements (Ackerman *et al.*, 2010; Larrarte, 2008). The strategies used for assessing the variability of water quality by sampling schemes (see Athayde *et al.*, 1983; Saget, 1994; Duncan, 1999; Pitt *et al.*, 2004; Brombach *et al.*, 2005; Ellis *et al.*, 2006) depend on sampling objectives, legal constraints, jointly with cost and logistics considerations. However, the main goal of any specific monitoring campaign is to collect the most accurate, reliable and representative data, given the technical and financial limitations (Ackerman *et al.*, 2010). Considering that the main sources of error in any sampling procedure are due to the sampling technique and the heterogeneity in both space and time of the sampling target (Paakkunainen *et al.*, 2007), generalizable and efficient strategies for water quality sampling during rainfall events, adaptable to technical and operative restrictions, remains a relevant question.

The sampling strategies referred to in this Chapter 2 are operative rules for sampling Total Suspended Solids (TSS) concentrations during rainfall events with an automated sampler or grab sampling, aimed at maximizing the representativeness of the dynamics of the pollutants. One common indicator of the stormwater pollutant emissions for this purpose is the Event Mean Concentration (EMC) (USEPA, 1983; Charbeneau and Barrett, 1998; Carleton *et al.*, 2001; Kim *et al.*, 2005; Lee *et al.*, 2007), which can be estimated from different sampling schemes sampling (*e.g.* grab, flow weighted, time weighted composite samples) (Lee *et al.*, 2007). However, the EMCs have shown to be very variable, depending on the specific sampling strategy to be used (Lee *et al.*, 2007; Ki *et al.*, 2011).

Representativeness of the mean of a single set of collected samples over the total mean of a complete set is a topic that has been addressed in other disciplines, defined as the mean sampling error (*e.g.* Gy, 1998; Minkkinen, 2004). Pierre Gy's fundamental sampling theory gives a mathematical formalization of the sampling problem, establishing theoretical equations for estimating this error (Gy, 1998). This formulation has been applied to environmental problems (*e.g.* Paakkunainen *et al.*, 2007), giving an appropriate frame for understanding the different components of errors. Nevertheless, theoretical assumptions and required parameters can be hard to measure in practice. Therefore, studies in urban drainage have evaluated the performance of different sampling strategies for quantifying pollutant concentrations and loads during rainfall events under a more practical point of view (*e.g.* Izuno *et al.*, 1998; Robertson and Roerish 1999; Stone *et al.*, 2000; Ma *et al.*, 2009; Ackerman *et al.*, 2010). These approaches have focused on assessing the performance of

different sampling strategies by comparing the EMCs obtained by measurements with a "reference" (closer to the true value) EMC calculated from numerical estimations (Shih *et al.*, 1994), Monte Carlo simulations (Richards and Holloway, 1987) or statistical methods (King and Harmel 2004; King *et al.*, 2005; Ma *et al.*, 2009). The main challenge in the mentioned approaches is that the differences between the EMC obtained from a sampling strategy and the "reference" EMC are established by hardly verifiable theoretical assumptions, under the absence of sufficient concentration data. Looking to overcome this limitation, further studies estimate the "reference" EMC using the outputs of water quality and quantity high temporal resolution conceptual models, calibrated with flow rate and water quality measurements (Ackerman *et al.*, 2010; Ki *et al.*, 2011). However, conceptual water quality models have been also described to have highly uncertain outputs due to uncertainties in model structure, model parameters and data (input and output) (Beck, 1987; Bertrand-Krajewski, 2007; Benedetti *et al.*, 2013; Dotto *et al.*, 2013), affecting the evaluation of a given sampling scheme, as well.

Measurement of stormwater quality in urban drainage systems with online technologies (*e.g.* turbidity, UV-VIS spectrometry) has been increasingly used for several purposes (*e.g.* control of urban water systems, modelling, and real time control) (*e.g.* Ruban *et al.*, 2001; Langergraber *et al.*, 2004a; Langergraber *et al.*, 2004b; Hur *et al.*, 2010). In addition, multiple authors have reported the benefits of online monitoring in terms of the spatio-temporal variability and complexity of water quality (*e.g.* Gruber *et al.*, 2005; Hochedlinger *et al.*, 2006; Schilperoort, 2011; Métadier and Bertrand-Krajewski, 2012). Accordingly, online monitoring emerges as a promising alternative for estimating the "reference" EMCs and to make comparisons across a range of conditions (adapted from Ackerman *et al.*, 2010).

The objective of the present Chapter 2 is to simulate and evaluate different sampling strategies proposed in the literature (*e.g.* Rossi, 1998; Bertrand-Krajewski *et al.*, 2000; Ackerman *et al.*, 2010) by using high temporal resolution flowrate and TSS time series (approx. one minute time-step, with about one year of data). EMCs from sampling strategies are simulated by sampling TSS time series and calculating a weighted average of the samples by their sampling volumes. These "simulated" EMCs are compared to the "reference" EMC calculated as a weighted average of the complete time series (flow rate and TSS). The approach is carried out with datasets from four urban drainage systems, aiming to compare results for different catchments. The approach is undertaken with datasets from four urban drainage systems, aimed at comparing the results among different catchments. The cases were the following: (i) Berlin, Germany, combined sewer overflow (CSO); (ii) Graz, Austria, (CSO); (iii) Chassieu, France, (combined sewer system) and (iv) Ecully, France (separated sewer system).

To our best knowledge, uncertainties and sensitivity analysis in the evaluation of sampling strategies have not been extensively addressed in the literature (*e.g.* Rossi, 1998; King *et al.*, 2005). Therefore, uncertainties in: (i) sampling volumes, (ii) laboratory values, (iii) online measurements (TSS and flow rate) and (iv) beginning/ending of storm events are assessed and propagated in the results by Monte Carlo simulations. The effects of the independent uncertainty sources over the total uncertainties of performance estimations (errors between sampling and on line monitoring EMCs) are estimated by the Sobol's Sensitivity Index. A statistical comparison between the uncertainties of EMCs obtained by sampling and online monitoring is also proposed.

2.2 MATERIALS AND METHODS

Data sets

Four datasets from different urban drainage systems are analyzed. Table 2 shows a summary of the main characteristics of the studied catchments and time series.

Table 2. Summary of characteristics of studied urban catchments and time series

Location	Year	Monitoring point	area (km^2)	% imperv.	Inhab. ~	avaliable # rainfall events	time series time-step	Land use	TSS measur.	Source
Berlín, Germany	2010	CSO*	1	74 (0.74 ha)	126000	22	1 min	Residential	UV-VIS	Sandoval *et al*., 2013
Graz, Austria	2009	CSO	3.35	32.24 (1.08 ha)	11800	79	Dry weather: 3 min; Wet weather: 1 min	Residential	UV-VIS	Gamerith, 2011
Chassieu, Lyon, France	2007	Separate sewer system outlet	1.85	75 (1.39 ha)	NA	89	2 min	Industrial	Turbidity meter	Métadier, 2011
Ecully, Lyon, France	2007	CSO	2.45	42 (1.03 ha)	18000	220	2 min	Residential	Turbidity meter	Métadier, 2011

*CSO: Combined Sewer Overflow

The dry and wet weather periods for the cases of Graz and Ecully are defined based on the flow rate values. Periods in which the inflows of the CSO-chamber are greater than 120 L/s (in Graz) and flow rates are greater than 30 L/s in Ecully are defined as the wet periods.

Sampling strategies

The investigated sampling strategies are operational rules for sampling TSS concentrations during rainfall events with an automated sampler (jointly with a flowmeter for some cases), by using different criteria. Table 3, jointly with the following lines, gives a brief description of the sampling strategies evaluated in this Chapter 2 (additional details can be found *e.g.* in Rossi, 1998 or Betrand-Krajewski *et al*., 2000).

Strategy cTcSV: Time-paced composite sampling (constant sampling volume): one sample is collected based on equally spaced time intervals and sampling volumes are constant. The inputs for this strategy are the sampling interval during the rainfall event and the constant sampling volume for samples. The sampling volume can be predefined to any constant value between 0.02 and 0.9 L (depending on operational constraints, as discussed below) (*e.g.* sampling each 10 min with a sampling volume of 0.4 L).

Strategy cTpQ: Time-paced composite sampling (sampling volume proportional to instantaneous flowrate): one sample is collected based on equally spaced time intervals and sampling volumes are pre-set as proportional to the instantaneous flowrate measured at the

sampling time-step. The sampling volumes are then proportionally selected: the minimum flowrate value registered during a rainfall event will then correspond to the minimum sampling volume (0.02 L), as the maximum flowrate will be sampled with a volume of 0.9 L. Therefore, in addition to the constant sampling interval, the minimum and maximum flowrate values during the event are then required as inputs (*e.g.* sampling each 10 min with a sampling volume of 0.2 L, whenever the instantaneous flowrate is 0.2 m^3/s).

Strategy cTpV: Time-paced composite sampling (sampling volume proportional to runoff volume between two samples): one sample is collected based on equally spaced time intervals and sampling volumes are pre-set as proportional to the runoff volume cumulated since the last sample. The sampling volumes are then proportionally selected: the minimum volume between two samples during a given rainfall event will then correspond to the minimum sampling volume (0.02 L), and the maximum runoff volume between two samples will be sampled with a volume of 0.9 L. Therefore, in addition to the constant sampling interval, the minimum and maximum volumes between two samples during the event are then required as inputs (*e.g.* sampling each 10 min with a sampling volume of 0.1 L, whenever the runoff volume between samples is of 10 m^3).

Strategy vTcV: Volume-paced composite sampling: for this strategy, the automated sampler takes a sample at pre-set runoff volume (rather than sampling interval), during a given rainfall event. This will lead to have non-equally spaced time intervals between samples. The sampling volume can be predefined to any constant value between 0.02 L and 0.9 L (depending on operational constraints, as discussed below) as for the case of strategy cTcSV. Thus, the inputs for this strategy are the constant sampling volume and the pre-set runoff volume (constant values) (*e.g.* constant sampling volume of 0.4 L, for each 10 m^3 of runoff between samples).

Table 3. Sampling strategies inputs and description

Name	Sampling time intervals	Sampling volume	Runoff volume between samples	Input parameters
cTcSV	constant	constant	variable	Δt (constant sampling time interval)
cTpQ	constant	f(flow rate)	variable	Δt (sampling time interval); min(flow rate) and max(flow rate)
cTpV	constant	f(RV*) before last sample	variable	Δt (sampling interval); min(RV*) and max(RV*) between sampling
vTcV	variable	constant	constant	Sampling intervals based on RV*

*RV: runoff volume between samples

For all the sampling strategies, the value of 0.02 is selected as the minimum sampling volume for which a TSS laboratory test can be conducted. As the sampling bottles have a capacity of 1 L, the 0.9 L is selected as the maximum sampling volume with the purpose of leaving a security margin, in case of *e.g.* spilling some amount of the sample during the handling process.

Simulating sampling strategies

Given the fact that comparing the results of simultaneous campaigns with several autosamplers (one for each strategy with different inputs) might be an unfeasible alternative (in economic and operative terms), the approach herein proposed is to use the TSS and flow rate time series to simulate the EMC_{sim} that would have been obtained by different sampling strategies. Figure 4 illustrates how the TSS time series is sampled by means of sampling intervals Δt, assigning a TSS_i value (i equal to the sampling time-step) and a sampling volume SV_i to each sampling bottle, as a function of the sampling strategy. The $EMCs_{sim}$ is calculated as a weighted average of the "sampled" TSS_i values by the sampling volumes SV_i (Figure 4). This calculation is equivalent to grab multiple samples and mix them in a composite 20 L sample jar (the size of 20 L is set from standard field conditions). For example, in Figure 4 the sampling intervals are constant during the rainfall event (time-paced strategies). The $EMCs_{sim}$ are simulated for each sampling strategy, including different inputs (Table 3). In addition, the "reference" EMC_{ref} is calculated by using the complete pseudo-continuous time series (one or two minute time-steps) of Q and TSS (Figure 5). $EMC_{ref}(j)$ and $EMC_{sim}(j)$ are calculated for each rainfall event j (from 22 to 220 rainfall events depending on the dataset).

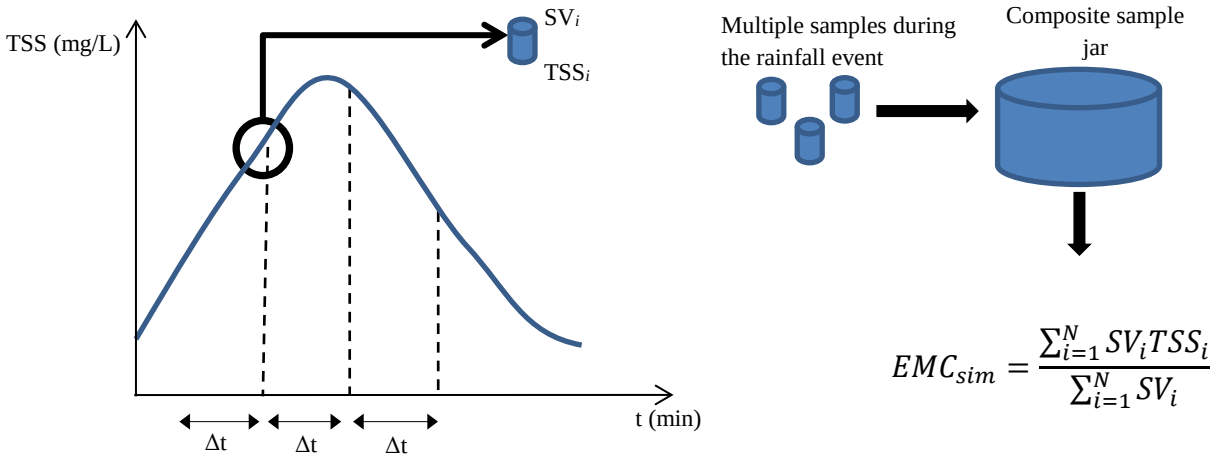


Figure 4. Procedure for obtaining the EMC_{sim} from grab sampling (SV: sampling volume; TSS: Total Suspended Solids Concentration) with strategy cTcSV.

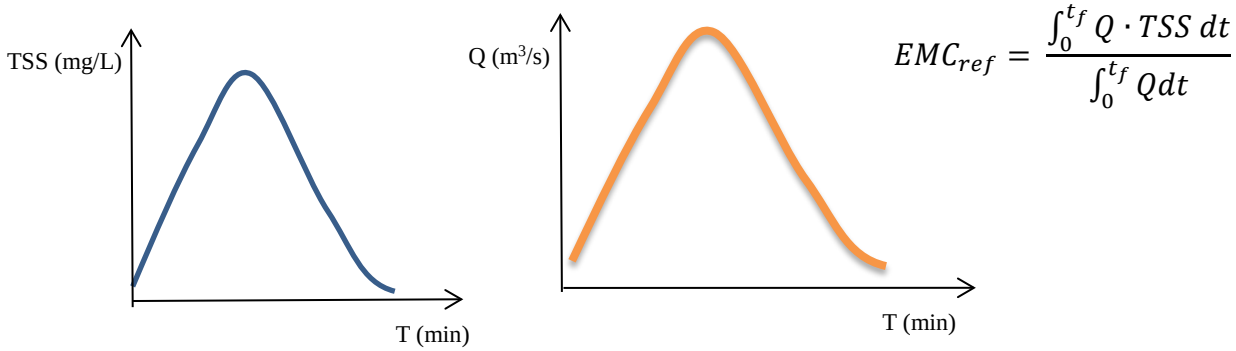


Figure 5. Procedure for obtaining the EMC_{ref}s from time series

The main hypothesis in this approach is that the TSS and Q time series are assumed to be the "reference" values, without any systematic error in the measurements. To support this fact, intensive research had explored the benefits and limitations of water quality and quantity online measurements, compared to traditional TSS laboratory tests (Bertrand-Krajewski *et al.*, 2007b; Torres and Bertrand-Krajewski, 2008b; Winkler *et al.*, 2008).

Different sampling intervals (*e.g.* from 1 min to 60 min) are evaluated for strategies with constant sampling intervals Δt (*i.e.* strategies cTcSV, cTpQ and cTpV). However, based on recommendations from Ackerman *et al*. (2010) and field experience, the minimum sampling interval Δt is recommended to be of 5 min. For applying proportional sampling volume SV strategies (cTpQ and cTpV), the minimum and maximum Q values of an event should be known in advance. In practice, these values can be estimated only after the end of rainfall event. A simple solution for the purposes of this Chapter 2 is to assume that the bottles are sampled with the highest possible sampling volume (0.9 L). Afterwards, the corresponding sampling volume SV from each bottle is corrected by using the already known minimum and maximum flow rate value, before mixing the sampling bottles into the 20 L jar (Figure 4). The performance of strategy vTcV (variable sampling intervals) is evaluated for several runoff volumes RV, from 5th percentile to 95th percentile of the total runoff volumes during the rainfall events. The mean Δt is calculated in the vTcV strategy for comparison with constant sampling intervals Δt strategies (cTcSV, cTpQ and cTpV).

Performance indicators

The estimation of the performance of a sampling strategy could be assessed in terms of the variability and repeatability of the sampling error across the different rainfall events (adapted from Gy, 1998). Therefore, the following performance indicators are considered: (i) the *residuals_vector* for each rainfall event (Eq 1) (bias estimation) and (ii) the Mean Squared Relative Error (*MSRE*) for all rainfall events (accuracy estimation) (Eq 2).

$$Residuals\_vector(j) = EMC_{true}(j) - EMC_{sim}(j) \quad \text{Eq 1}$$

$$MSRE = \frac{\sqrt{\sum_{j=1}^{N} residuals(j)^2}}{N \cdot mean(EMC_{true})} \quad \text{Eq 2}$$

The *residuals_vector* has a length equal to the number of rainfall events *N* and can be considered as a bias estimation (over or under estimation of the EMC), as the mean of this vector should be significantly close to zero for guaranteeing a non-biased estimation over all *j* rainfall events (adapted from Gy, 1998).

On the other hand, the *MSRE* can be defined as the standard deviation of the sampling error, which is related to the accuracy (*MSRE*) of a given sampling strategy (Gy, 1998; Paakkunainen *et al*., 2007). The representativeness of a sample is a function of the expectation and the standard deviation of the sampling error (Gy, 1998). Therefore, Eq 1 and Eq 2 are proposed to describe the representativeness of a sampling strategy. In addition, the uncertainties in the *MSRE* are considered as a complementary indicator of the performance of a sampling strategy. This indicator is estimated by the Monte Carlo method, propagating the uncertainty sources over the *MSRE* (see Table 4).

The standard uncertainty $u(x)$ of the random variable x can be estimated by the Law of Propagation of Uncertainties LPU, in which the condition of normality (or at least symmetry of the distribution) is necessary. In those cases the expanded uncertainty can be calculated as $I = k \cdot u(x)$, where k is the enlargement factor that guarantees a desired coverage interval (*e.g.* $k = 2$ for a 95 % coverage when the distribution is normal) (ISO, 2009). Whenever the normality or even symmetry of the distribution is not verified (especially over the propagated distributions), the expanded uncertainty $CI(x)_{min95\ \%}$ can be calculated by the Monte Carlo method as the minimum interval formed by a couple of points (a, b) that can cover 95 % of the distribution of x (ISO, 2009). The standard uncertainties $u(x)$ from Table 4 are included in the analysis as the standard deviation of a normal distribution for each of the variables used in the calculation of the EMCs and *MSRE*s (Table 4). The propagation of uncertainty sources over EMCs and *MSRE*s is calculated in their expanded form as $CI(\mathrm{EMC})_{min95\%}$ and $CI(MSRE)_{min95\%}$ respectively. The $CI(MSRE)_{min95\%}$ value represents the influence of uncertainty sources (Table 4) over the variability of the *MSRE* (Eq 2), as a total error indicator.

Regarding the calculation of the $CI(MSRE)_{min95\%}$ (function of $CI(\mathrm{EMC_{ref}})_{min95\%}$ and $CI(\mathrm{EMC_{sim}})_{min95\%}$ values), six uncertainty sources are considered, which are listed in Table 4 (adapted from Rossi, 1998). Uncertainties in sampling intake position and time shift due to the pumping operation are not considered due to the lack of information (sources 7 and 8 in Table 4).

Table 4. Uncertainty sources description

	Uncertainty source code	**Description**	**Standard uncertainty value $u(x)$**	**Influence on $\mathrm{EMC_{sim}}$**	**Influence on $\mathrm{EMC_{ref}}$**	**Probability distribution**
1)	$u(\mathrm{Q_{ts}})$	uncertainty of Q time series	Q online measurements uncertainties (site dependent)	sampling volumes (cTpQ)	($\mathrm{RV_{ts}}$) (used in all strategies)	$N \sim (Q\ ;\ u(Q))$
2)	u(samplV)	uncertainty of sampling volumes	4.5 % (LGCIE, Exera report)	sampling volume in each strategy	no effect	$N \sim (SV\ ;\ 0.075 \cdot SV)$
3)	u(start/end)	start/ending of rainfall event, it varies the delimitation of D/W periods	Beginning 5 % and ending 7.5 % of the rainfall duration (Métadier, 2011)	directly; for vTcV it affects the definition of a Δt related to a given runoff volume	direct effect	Beginning: t_{start} + $U \sim [-0.05 \cdot (t_{end} - t_{start})\ ;\ 0.05 \cdot (t_{end} - t_{start})]$ Ending: t_{end} + $U \sim [-0.075 \cdot (t_{end} - t_{start})\ ;\ 0.075 \cdot (t_{end} - t_{start})]$

	Uncertainty source code	Description	Standard uncertainty value $u(x)$	Influence on EMC_{sim}	Influence on EMC_{ref}	Probability distribution
4)	***u*(TSSsampl)**	uncertainty related to the TSS obtained from the sampling strategy	7.5 % of the value (TSS standard test uncertainty) (LGCIE, Exera report)	directly	no effect	N ~ (TSS ; 0.075·TSS)
5)	***u*(TSS_{ts})**	uncertainty of TSS time series	TSS online measurements uncertainties (site dependent)	no effect	direct effect	N ~ (TSS ; *u*(TSS))
6)	***u*(RV_{ts})**	uncertainty related with runoff volume from Q time series	Q online measurements, including uncertainty of starting/ending of events	sampling volumes (cTcSV, cTpV and vTcV)	direct effect	propaged from distributions 1 and 4
7)	***u*(Sampl. position)**	uncertainties due to sampling intake position	neglected for this work	NA	NA	NA
8)	***u*(Time-shift)**	uncertainty due to the time shift given by the pumping operation	neglected for this work	NA	NA	NA

For clarifying, the source of uncertainty $u(RV_{ts})$ comes from an instantaneous runoff volume time series (RV_{ts}). This series is calculated from the flow rate time series *Q*, including uncertainties of $u(Q_{ts})$ but also of *u*(start/end), with the purpose of establishing the sampling volumes for cTcSV, cTpV and vTcV strategies. Each uncertainty source is propagated over the EMC_{sim}, EMC_{ref}, *MRSE* and *residuals_vector* (Eq 1 and Eq 2) by Monte Carlo simulations, including the proposed probability distributions (Table 4). The probability distribution of uncertainty source 3 (Table 4) is a uniform distribution *U* that represents the uncertainties in the moment of beginning t_{start} and ending t_{end} of a rainfall event (with duration t_{end}- t_{start}).

The uncertainty related to the TSS sampling *u*(TSSsampl) is different from the uncertainties in the TSS from the time series $u(TSS_{ts})$ (Table 4, source 4 and 5 respectively). The $u(TSS_{ts})$ are related to the TSS measurement technology (Turbidity or UV-VIS spectrometry), with its calibration procedure. Both values are assumed to be measured with a different measurement approach (laboratory traditional test and online monitoring) (Figure 4 and Table 4). The Sobol's sensitivity total index (SSI) estimation proposed by Glen and Isaacs (2012) is applied to quantify the influence of uncertainty sources (Table 4) over the total uncertainty of *MSRE* values, $CI(MSRE)_{min95\%}$. See a detailed description of the application of the methodology in Appendix 2.

Iman and Conover (1980) propose the Latin Hypercube Sampling LHS, as an extension of the Monte Carlo method, which can lead to equivalent results with a lower number of simulations, by better covering the probability distribution domain. The LHS is tested against

the traditional Monte Carlo approach in punctual cases, reporting suitable results in terms of diminishing the computational effort. For time series analysis, some authors have pointed out that the correlation effect of time series (autocorrelation or variogram) can increase the values of propagated uncertainties over time series calculations (Bertrand-Krajewski and Bardin, 2002). Therefore, the correlation matrix of the time series could be included in the LHS as the autocorrelation matrix, in order to consider this effect (adapted from McMurry and Politis, 2010). However, the autocorrelation matrix has shown not to be a sufficient estimator of the time series correlation matrix (Wu and Pourahmadi, 2009). This problem is still subject of intensive research by multiple authors (*e.g.* Xiao and Wu, 2012; Xue and Zou, 2012). For instance, the time series are going to be considered as merely observations, without an auto-correlated underlying process (correlation matrix equal to identity matrix in the LHS).

Further comparisons are undertaken between $CI(\mathrm{EMC}_{ref})_{min95\%}$ and $CI(\mathrm{EMC}_{sim})_{min95\%}$ by the use of statistical methods (*t-test* or Wilcoxon) to show significant differences between uncertainties in EMC obtained by online monitoring or using a simulated sampling scheme. The selection of the *t-test* or Wilcoxon test is dependent on the normality of the *residuals_vector* (EMC_{ref} - EMC_{sim}) (Shapiro Wilk test *p-value* > 0.05 for normally distributed samples).

2.3 RESULTS AND DISCUSSION

With the purpose of testing the efficiency of LHS against traditional Monte Carlo Method, and defining as well a representative number of simulations, the convergence of the results obtained by the two approaches is checked for some cases *e.g.* Chassieu, cTpQ strategy with a 15 min sampling interval (Figure 6). The MC method (left) does not show stable results of $CI(\mathrm{MSRE})_{min95\%}$ for a number of simulations lower than 400. These results are in agreement with previous studies (*e.g.* Helton and Davis, 2003), where the LHS approach has reported satisfactory results for assessing and propagating uncertainties more efficiently (lower number of simulations for computing the same *MSRE* values) in comparison to the traditional Monte Carlo method. From results in Figure 6, jointly with the maximum computational capacity, the number of simulations is defined as 200 for Berlin, Chassieu, Graz and Ecully, by using the LHS approach.

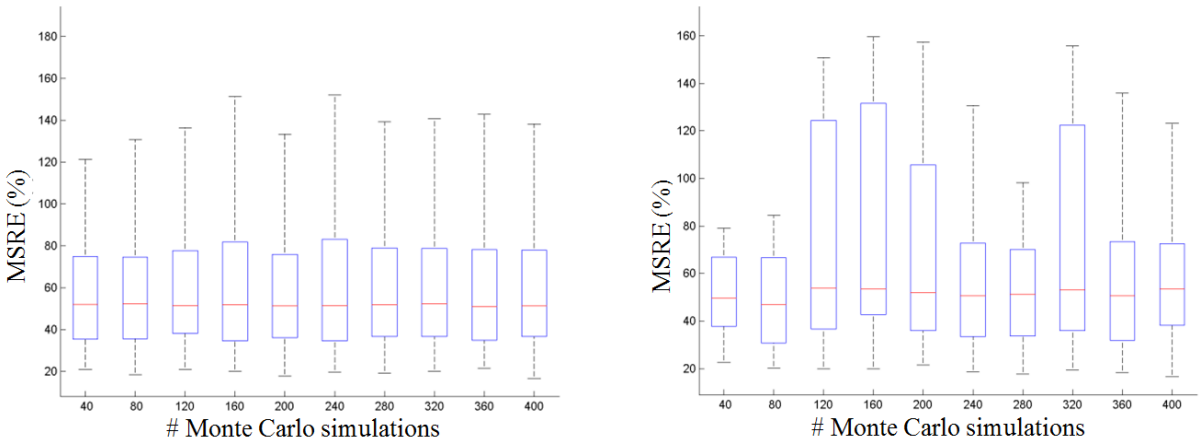


Figure 6. Variability of MSRE values ($CI(MSRE)_{min95\%}$) against different number of simulations (*x*-axis) for strategy cTcSV with sampling intervals of 15 min in Chassieu (left: LHS, right: traditional Monte Carlo simulations).

The calculations are undertaken for the datasets of Berlin, Chassieu, Graz and Ecully. Different criteria are evaluated for each of the studied datasets and strategies: (i) performance of each strategy (sampling error *MSRE* and $CI(MSRE)_{min95\%}$) (Eq 2), (ii) *residuals_vector* (Eq 1) (iii) Sobol's Sensitivity Index SSI and (iv) statistical comparison between $CI(EMC_{sim})_{min95\%}$ and $CI(EMC_{ref})_{min95\%}$. Different sampling intervals are considered from 1 min to 60 min (Graz, Chassieu and Ecully) and from 1 min to 30 min (Berlin). Only even sampling intervals (*i.e.* 2 min, 4 min, etc.) are evaluated for the case of Chassieu and Ecully, as the time-step of the data is 2 min.

Considering that the maximum volume of the composite sample jar is set to 20 L, *MSRE* values increase for short sampling intervals, as the composite sample jar of 20 L is full before the end of the event (especially for rainfall events with high volumes). This fact is considered as well in constant sampling volume SV strategies (*i.e.* cTcSV and vTcV) by setting a constant SV of 0.4 L. For cTpQ and cTpV, SV are assigned from 0.02 L to 0.9 L, depending on the input of each sampling strategy (Table 3). Rainfall events for a given strategy are not included in the calculations if: (i) the sampling interval is longer than the duration of the events (for cTcSV, cTpV and vTcV) or (ii) the total runoff volume RV set between samples is greater than the total runoff volume of the rainfall (for vTcV). Therefore, the following number of rainfall events is used for each dataset in vTcV strategy, as a function of the sampling runoff volume between samples RV (Table 3).

- Berlin: 22 events for RV > 500 m^3 (mean sampling intervals of 5 min) decreasing until 9 events for RV = 4500 m^3 (mean sampling intervals of 5 min).

- Chassieu: 89 events for RV > 280 m^3 (mean sampling intervals of 15 min) decreasing until 37 events for RV = 1700 m^3 (mean sampling intervals of 60 min).

- Graz: 79 events for RV > 100 m^3 (mean sampling intervals of 5 min) decreasing until 52 events for RV = 1300 m^3 (mean sampling intervals of 60 min).

- Ecully: 220 events for RV > $1x10^5$ m^3 (mean sampling intervals of 5 min) decreasing until 100 events for RV = $12x10^5$ m^3 (mean sampling intervals of 60 min).

Results about *MSRE* (solid lines) and $CI(MSRE)_{min95\%}$ (colored bands) are shown for sampling time intervals from 1 to 60 min on the lower x-axis (strategies cTcSV, cTpQ and cTpV) (Figure 7). The upper x-axis shows the different runoff volumes that are used to evaluate the vTcV strategy, with the corresponding mean sampling time intervals on the lower x-axis (Figure 7). The sampling error *MSRE* and its uncertainty $CI(MSRE)_{min95\%}$ increase for greater sampling interval, evaluated over the different rainfall events and datasets.

In Berlin (Figure 7a), the strategies cTpQ and cTpV show the best performance, with similar results (*MSRE* 7 % to 25 %, $CI(MSRE)_{min95\%}$ lower than 20 %). The strategy vTcV shows the lowest performance with the highest *MSRE* and $CI(MSRE)_{min95\%}$ values. In Chassieu (Figure 7b), $CI(MSRE)_{min95\%}$ values are lower for strategies cTcSV and vTcV, and a lower *MSRE* is observed for strategy cTpV for all sampling intervals. The cTpQ strategy is particularly sensible to flow rate outliers, as the sampling volumes are obtained by a weighted average of instantaneous flow rate values. Therefore, the $CI(MSRE)_{min95\%}$ values higher than 100 % for strategy cTpQ could be related to potential outliers in the flow rate time series. In Graz (Figure 7c), no strategy outperforms the other ones. $CI(MSRE)_{min95\%}$ values are similar for all sampling strategies and sampling intervals, and generally lower than 15 %. In Ecully (Figure

7d), the strategy cTpV shows the lowest *MSRE* and $CI(MSRE)_{min95\%}$ values (resp. < 20 % and from 1 % to 20 %) for all sampling intervals).

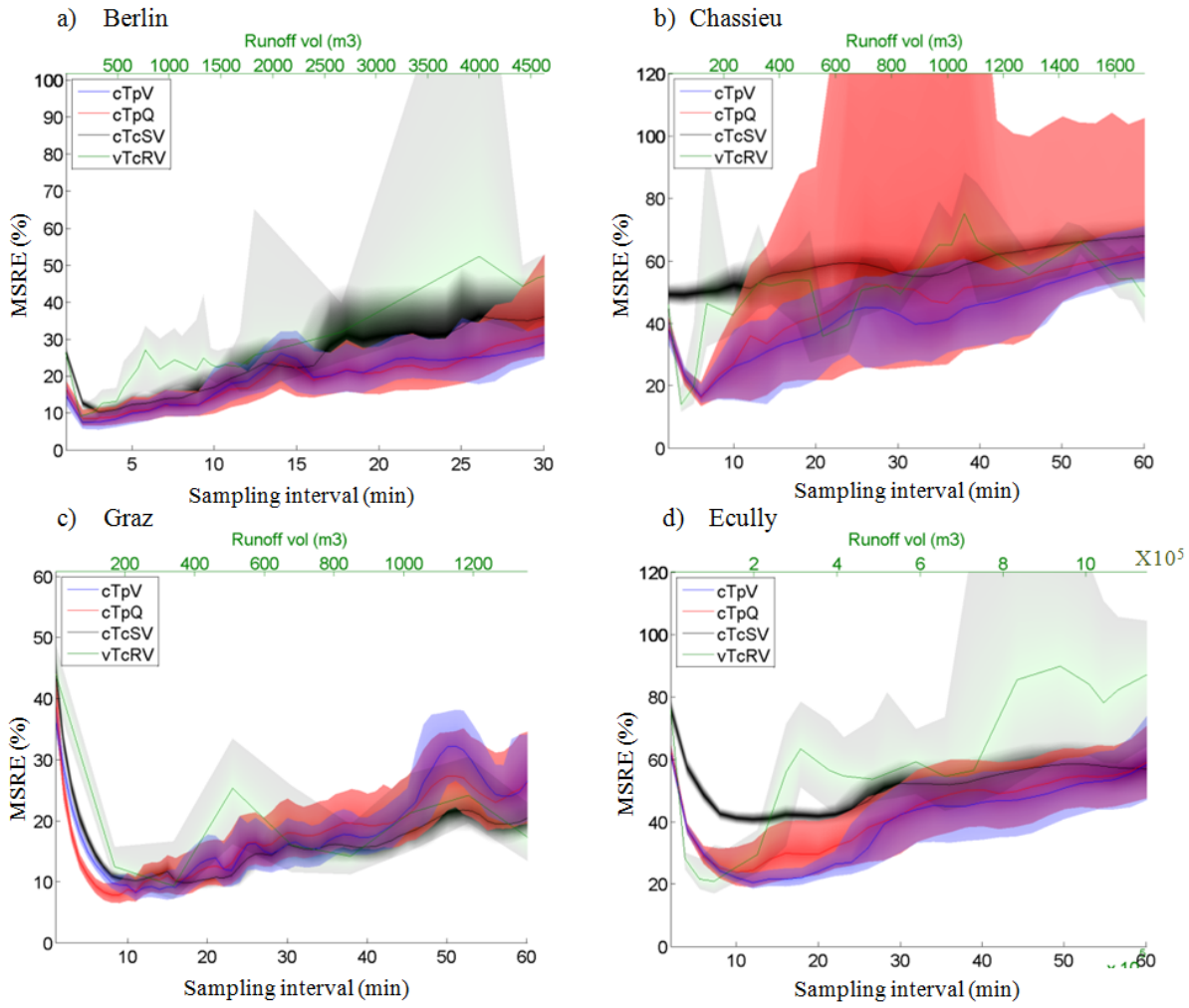


Figure 7. *MSRE* (solid line) and $CI(MSRE)_{min95\%}$ (colored bands) for different sampling time intervals in strategies cTcSV (black), cTpQ (red) and cTpV (blue) on the lower *x*-axis and different sampling volumes in strategy vTcV (green) on the upper *x*-axis, for a) Berlin, b) Chassieu, c) Graz and d) Ecully.

For Berlin and Chassieu, the 20 L composite sample jar is filled before the end of many events for sampling intervals shorter than 3 and 5 min respectively (resulting in higher *MSRE* values for shortest sampling intervals). These results, along with field experience and usual recommendations for sampling strategies, indicate that less than 5 min time interval is not appropriate from an operational point of view. Therefore, the best sampling strategy might be cTpV (or cTpQ as well, but being aware of its sensitivity to unusual flow rate values), using sampling intervals of about 5 min.

For Graz and Ecully, the constraint regarding the maximum composite sample jar volume has an impact for sampling intervals lower than 10 min. Therefore, for any sampling strategy to be implemented in these sites, sampling intervals are recommended to be about 10 min (*MSRE* of 10 % and 20 % respectively). Differences between the recommendations about the optimal sampling time interval for Berlin and Chassieu (5 min), compared to Graz and Ecully (10 min), can be related to the size of the catchments (100 and 185 ha for Berlin and Chassieu, compared to 335 and 245 ha for Graz and Ecully) (Table 2). Graz and Ecully

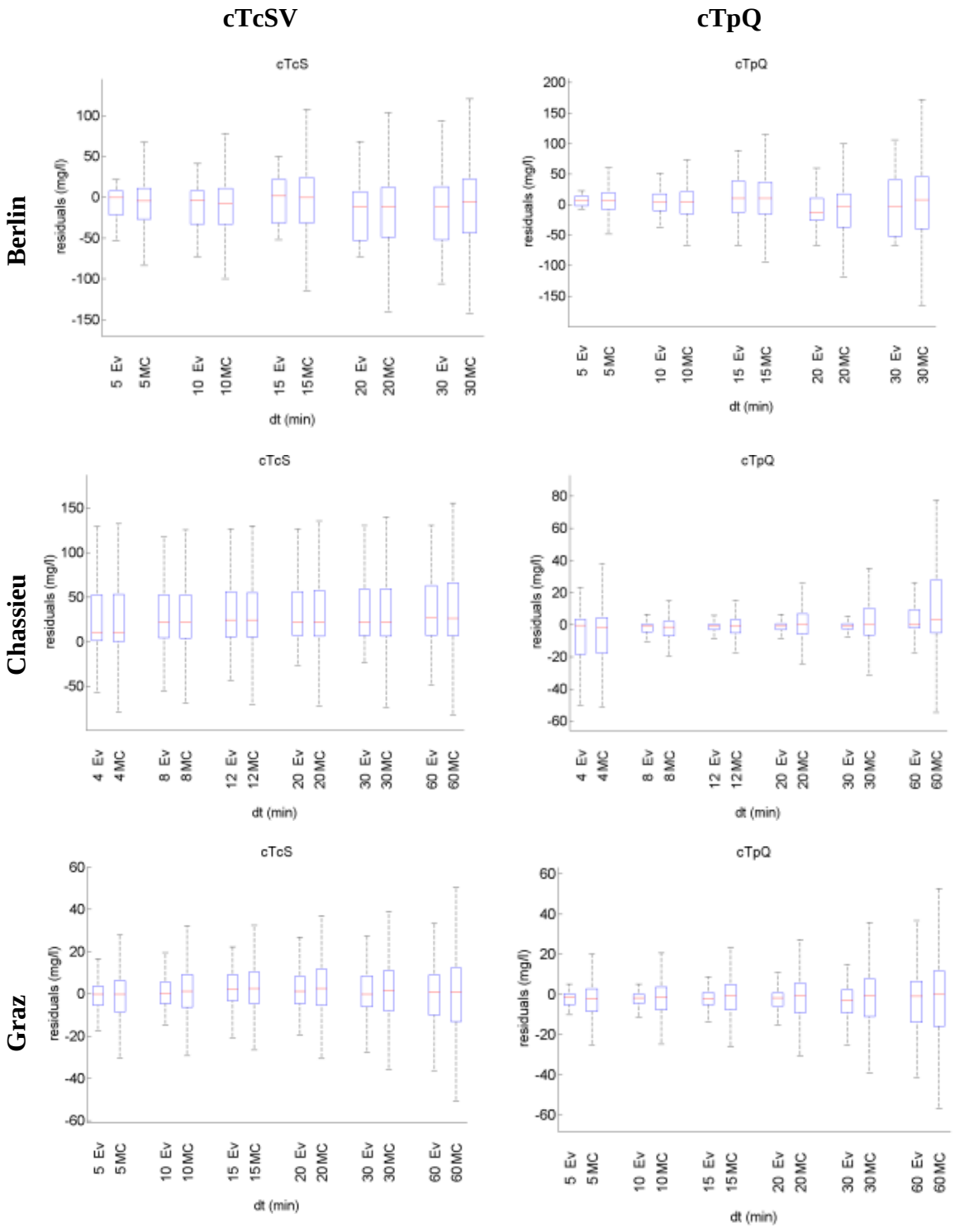
produce runoff volumes over longer durations leading to more frequent overfilling of the maximum 20 L jar.

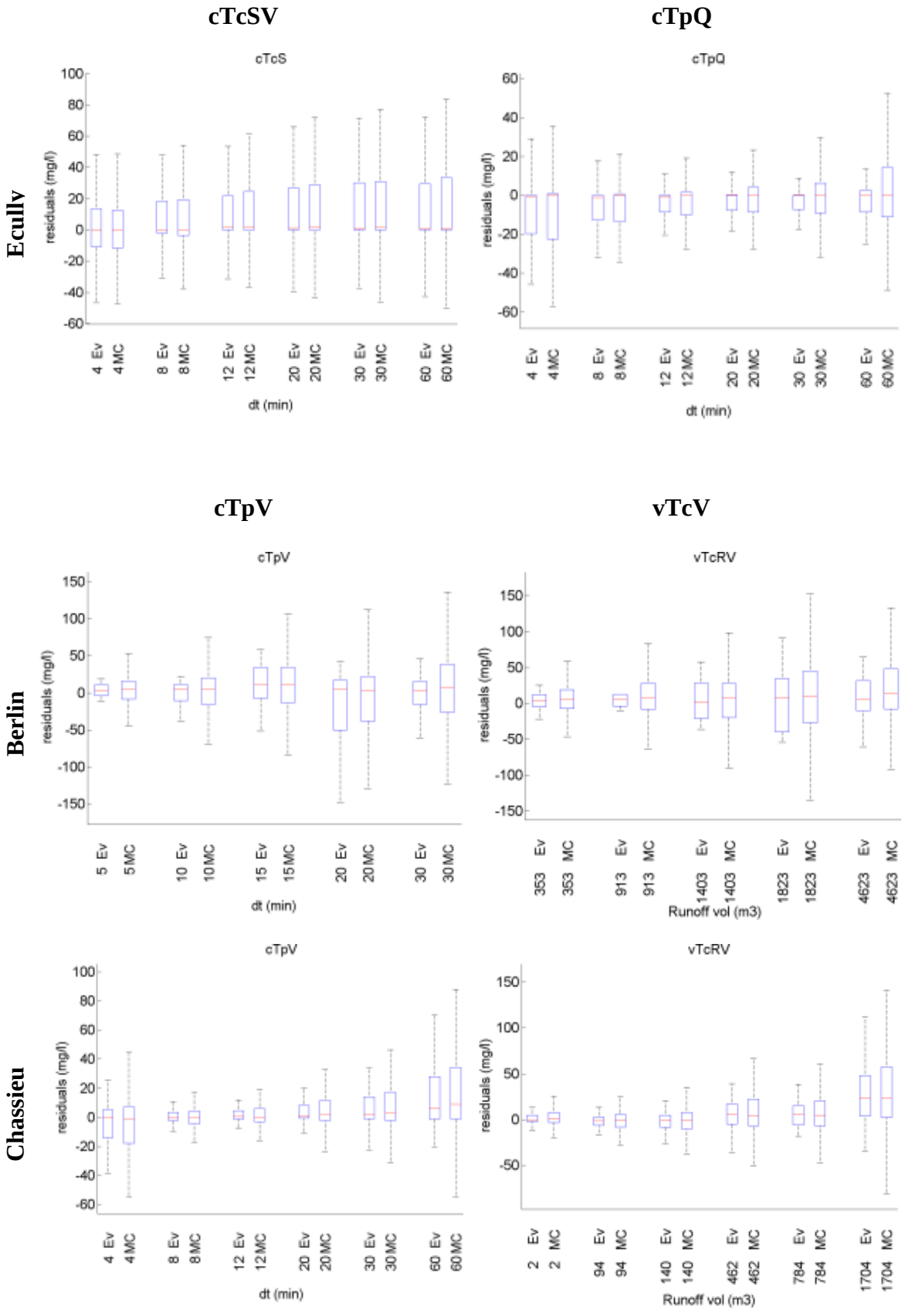
For all cases (Berlin, Chassieu, Graz and Ecully), the strategy cTpV delivers the most representative results in terms of performance *MSRE* and uncertainties $CI(MSRE)_{min95\%}$. Previous studies (*e.g.* Leecaster *et al*., 2002; Ma *et al*., 2009) also concluded that weighing the sampling volumes based on the runoff volume can bring up the most accurate and precise estimations of EMC. The constant sampling time intervals providing the best performance are about 5 min for the smaller catchments (Berlin and Chassieu) and about 10 min for the larger ones (Graz and Ecully). The sampling errors (*MSRE*) for these recommended sampling intervals range from 7 % (Berlin) up to 20 % (Chassieu and Ecully), with 95 % coverage intervals $CI(MSRE)_{min95\%}$ of about 5 %. If appropriate sampling intervals are selected (about 10 min), the cTcSV strategy can be a feasible alternative in absence of a flow-meter, with acceptable *MSRE* (lower than 20 %) and $CI(MSRE)_{min95\%}$ (lower than 10 %). The operative restriction of a maximum volume of 20 L for the composite sample jar has a significant influence on the selection of the best sampling strategy, especially for large catchments.

The *residuals_vector* between EMC_{ref} and EMC_{sim} are calculated for each rainfall event (to be called EV) (Figure 8) including the same sampling intervals and runoff volumes between samples RVs used in the results of the MSRE estimations (lower and upper x-axis, respectively). In addition, uncertainties are propagated into the residuals_vector (Eq 1) by the LHS method, including the uncertainty sources defined in Table 4 (to be called MC) (Figure 8). An asymmetry in the distribution of the *residuals_vector* towards 0 mg/L is an indicator of systematic over or underestimations of the EMC. Whenever at least half of the events are over or under-estimated, the strategy can be considered to be biased.

The strategy cTcSV applied to Chassieu and Ecully catchments shows an underestimation of the EMC (positive signs of the *residuals_vector*) for 75 % of the rainfall events, with sampling intervals greater or equal to 4 min and 8 min, respectively. One particularity of the TSS pollutographs during a rainfall event is that the amount of TSS values lower than the EMC_{ref} is higher than the amount of TSS values higher than the EMC_{ref} (median TSS lower than mean TSS). The bias reported for this case is explained by the fact that the cTcSV strategy will tend to reflect the natural asymmetry of the TSS values distribution, as flow rate information is not considered. Gy (1998) envisaged similar conclusions, demonstrating mathematically that non-weighing sampling strategies (*e.g.* cTcSV) are less representative than weighing strategies (cTpQ and cTpV), under certain theoretical assumptions over the sampled signal (*e.g.* autocorrelated series). Other studies such as Leecaster *et al*. (2002) report similar bias for non-flow rate weighted strategies with site measurements.

Analogous results are delivered for strategies cTpQ, cTpV and vTcV in Chassieu and Ecully, using sampling intervals greater than 30 min (underestimation of the EMC for 75 % of the rainfall events). These long sampling intervals lead to take more samples from the ending of the rainfall event (after the peak). This part of the pollutograph is usually longer and contains a greater amount of TSS values lower than EMC_{ref}. Therefore, the EMC_{sim} is more likely to be calculated with the lowest values of the pollutograph, independently of the weighting sampling volumes SV (Figure 4).





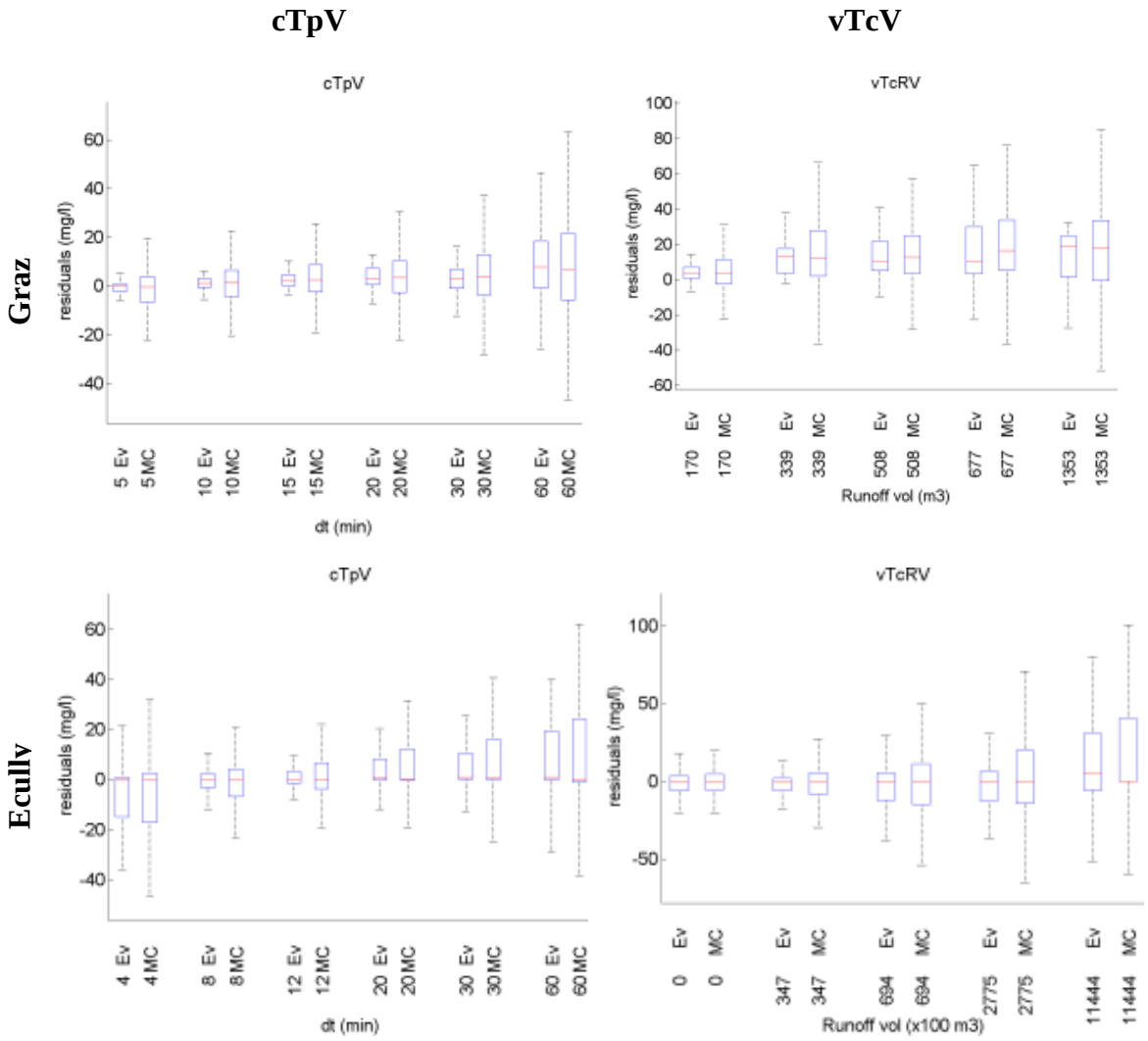


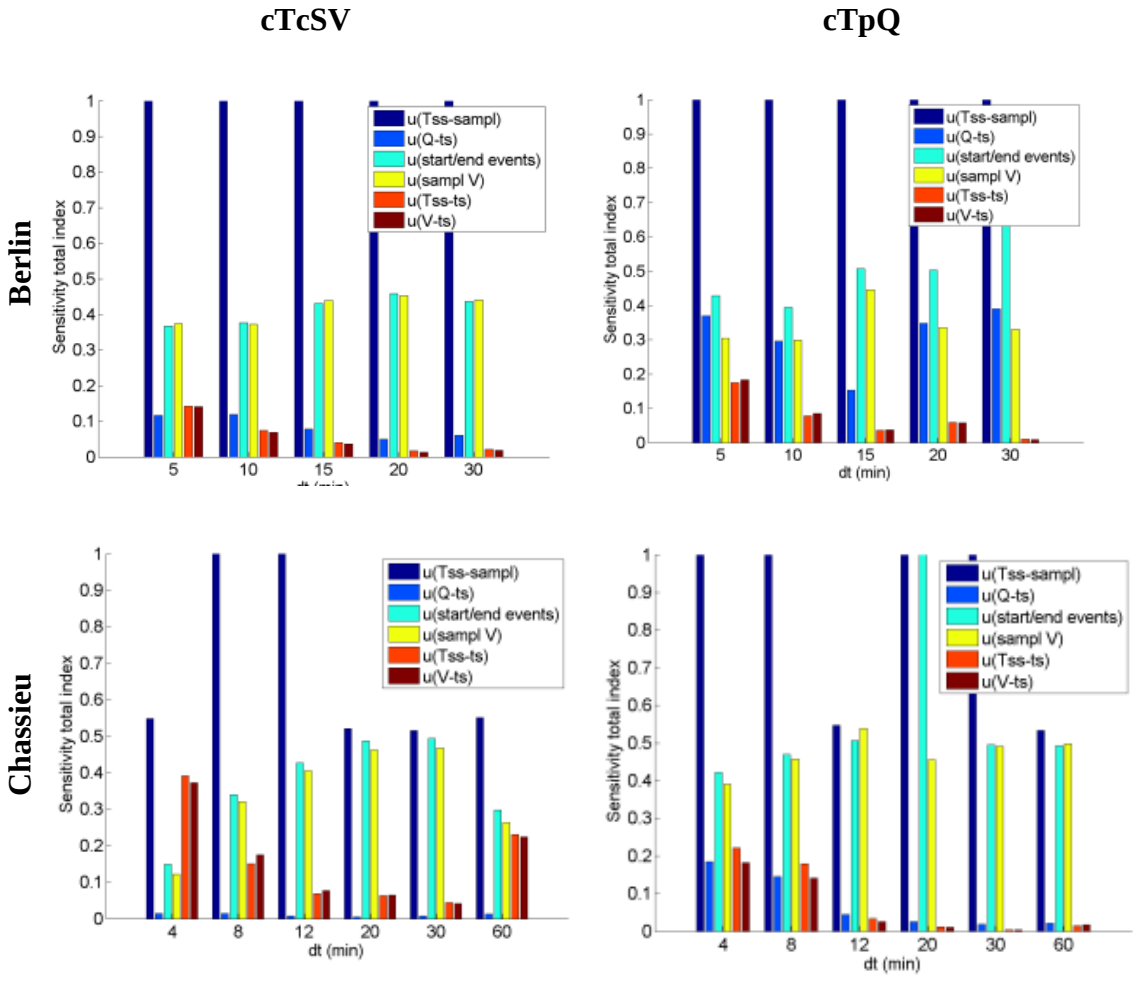
Figure 8. *residual_vector* (EMC_{ref} – $EMCs_{im}$) Eq 1 (EV) with propagated uncertainties (MC) for different sampling time intervals in strategies cTcSV, cTpQ and cTpV and different sampling volumes in strategy vTcV for Berlin, Chassieu, Graz and Ecully.

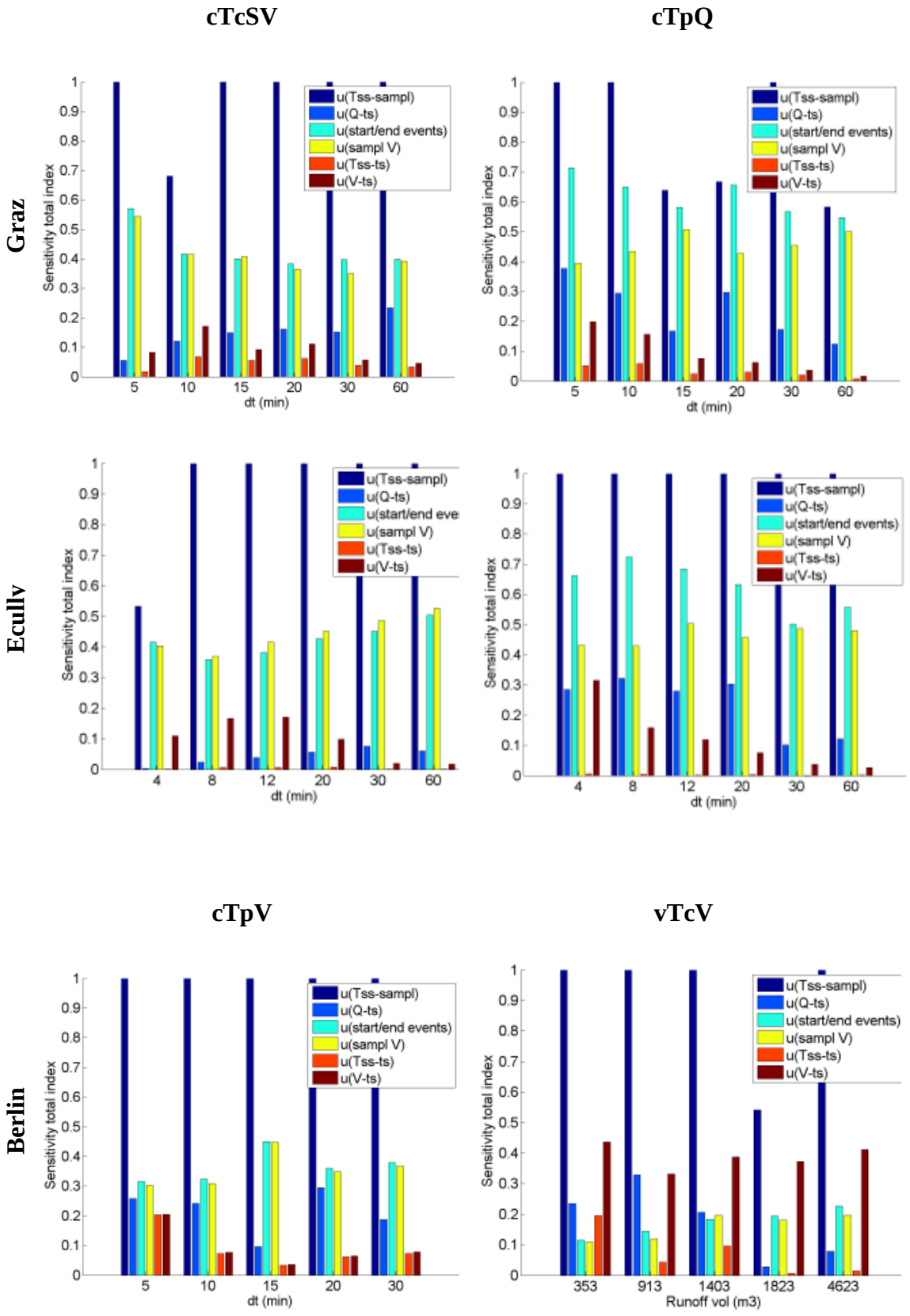
The cTpQ and cTpV strategies show systematic overestimations of the EMC (negative signs of the *residuals_vector*) for more than 50 % of the events in Chassieu and Ecully, considering sampling intervals lower than 4 min and 8 min, respectively. The 20 L composite sample jar is filled during the first part of the event for short sampling intervals, leading to amplify the sampling in the zone before the peak of the pollutograph. Indeed, the majority of the TSS values higher than EMC_{sim} are located at the beginning of the event. The *residuals_vector* (Eq 1) remained towards zero for the cases of Berlin and Graz. The differences in the results among the different catchments can be attributed to facts such as the size of each basin, physical characteristics and type of hydrosystem (Table 2).

The variability of the *residuals_vector* (Eq 1) becomes higher for greater sampling intervals in all sampling strategies in all datasets. This can be expected from results in Figure 8 (MC), as the mean difference between EMC_{sim} and EMC_{ref} becomes higher due to a lower amount of values used to calculate the EMC_{sim} (sampling interval size). This variability increases as well when uncertainties are taken into account by the LHS method (Figure 8).

However, the percentages of over or underestimation remain constant compared to the EV scenario. For Berlin and Graz, although the variability of the distribution in the *residuals_vector* increased for higher sampling intervals, no asymmetry is noticed in the distributions.

The results of Sobol's total Sensitivity Index (SSI) are shown in Figure 9 for each of the six uncertainty sources. Including the results for all the strategies, the following generalizations can be observed, independently of the selected Δt (cTcSV, cTpQ and cTpV) or RV (vTcSV): (i) the most important source of uncertainty in $CI(MSRE)_{min95\%}$ comes from the TSS laboratory values *u*(TSSsampl), (ii) the uncertainties related to the D/W weather delimitation *u*(start/end) and sampling volumes *u*(samplV) have an important effect over $CI(MSRE)_{min95\%}$ for constant sampling interval Δt strategies (cTcSV, cTpQ and cTpSV) and (iii) the volume time series uncertainty *u*(RV) has an important influence over $CI(MSRE)_{min95\%}$ in strategy vTcV. Indeed, the importance of *u*(TSSsampl) over $CI(MSRE)_{min95\%}$ remains systematically higher than for $u(TSS_{ts})$, for all strategies. This can be explained from the amount of TSS data used to calculate the EMC_{ref}, which is necessarily higher than for the EMC_{sim}, and therefore EMC_{sim} might be more sensitive to variations in the TSS values than $EMC_{ref.}$ These results may encourage researchers and practitioners to find adaptable methodologies for delimiting the beginning and ending of rainfall events, especially for applying sampling schemes based on time intervals. TSS laboratory values is the main uncertainty source for estimating the EMC by sampling strategies, therefore special attention should be given to laboratory and samples handling protocols.





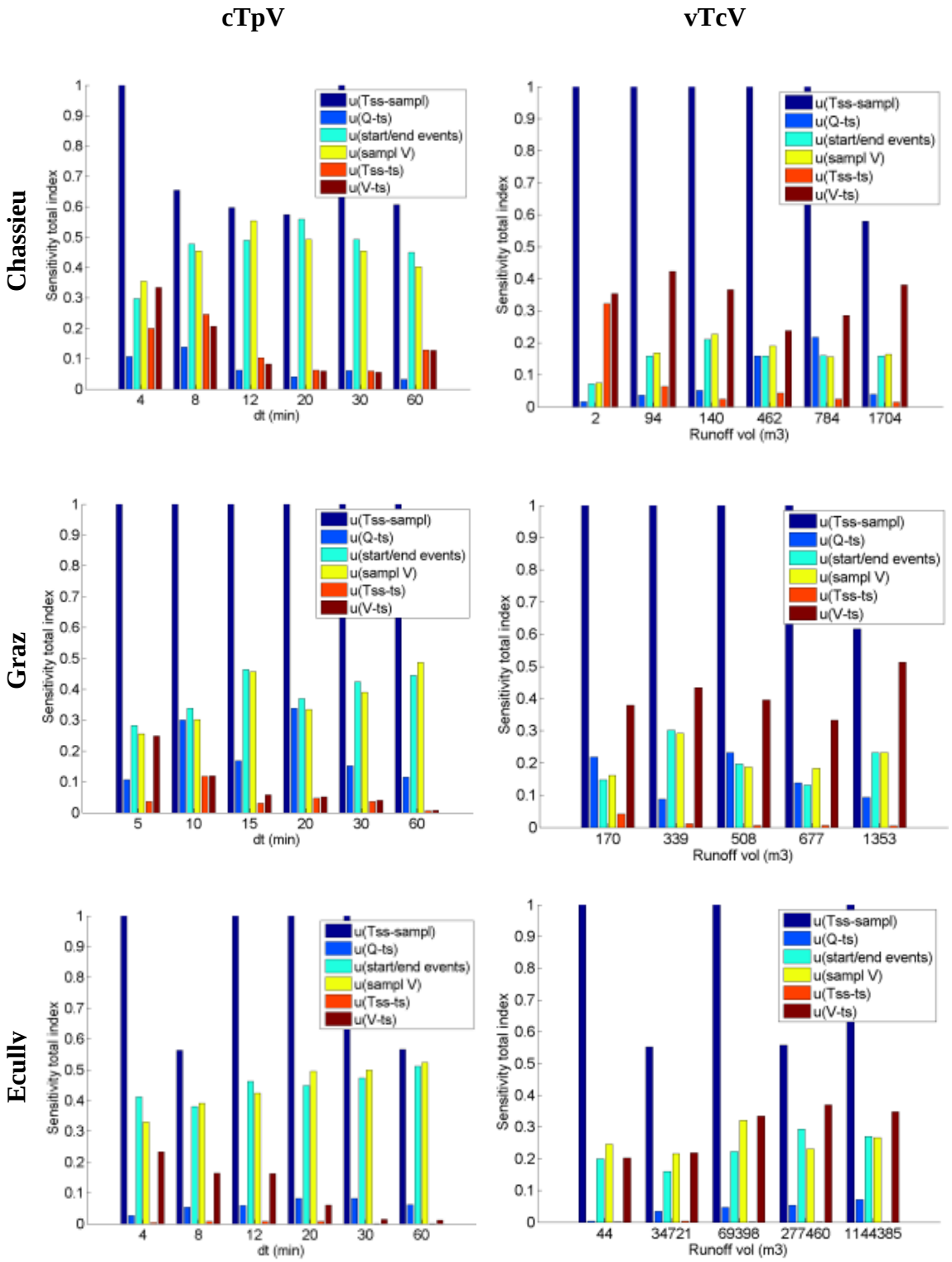


Figure 9. Sobol's Total Index for sensitivity analysis considering uncertainty sources in Table 4 for different sampling time intervals in strategies cTcSV, cTpQ and cTpV and different sampling volumes in strategy vTcV for Berlin, Chassieu, Graz and Ecully.

The standard uncertainties of the EMCs obtained from time series $CI(\mathrm{EMC_{ref}})_{\mathrm{min95\%}}$ are compared with the uncertainties of the EMCs calculated by different sampling strategies and sampling intervals $CI(\mathrm{EMC_{sim}})_{\mathrm{min95\%}}$ (Figure 10). The notation can be read as follows: strategy and the sampling interval, *e.g.* cTcSV_5 = cTcSV strategy with 5 min sampling interval.

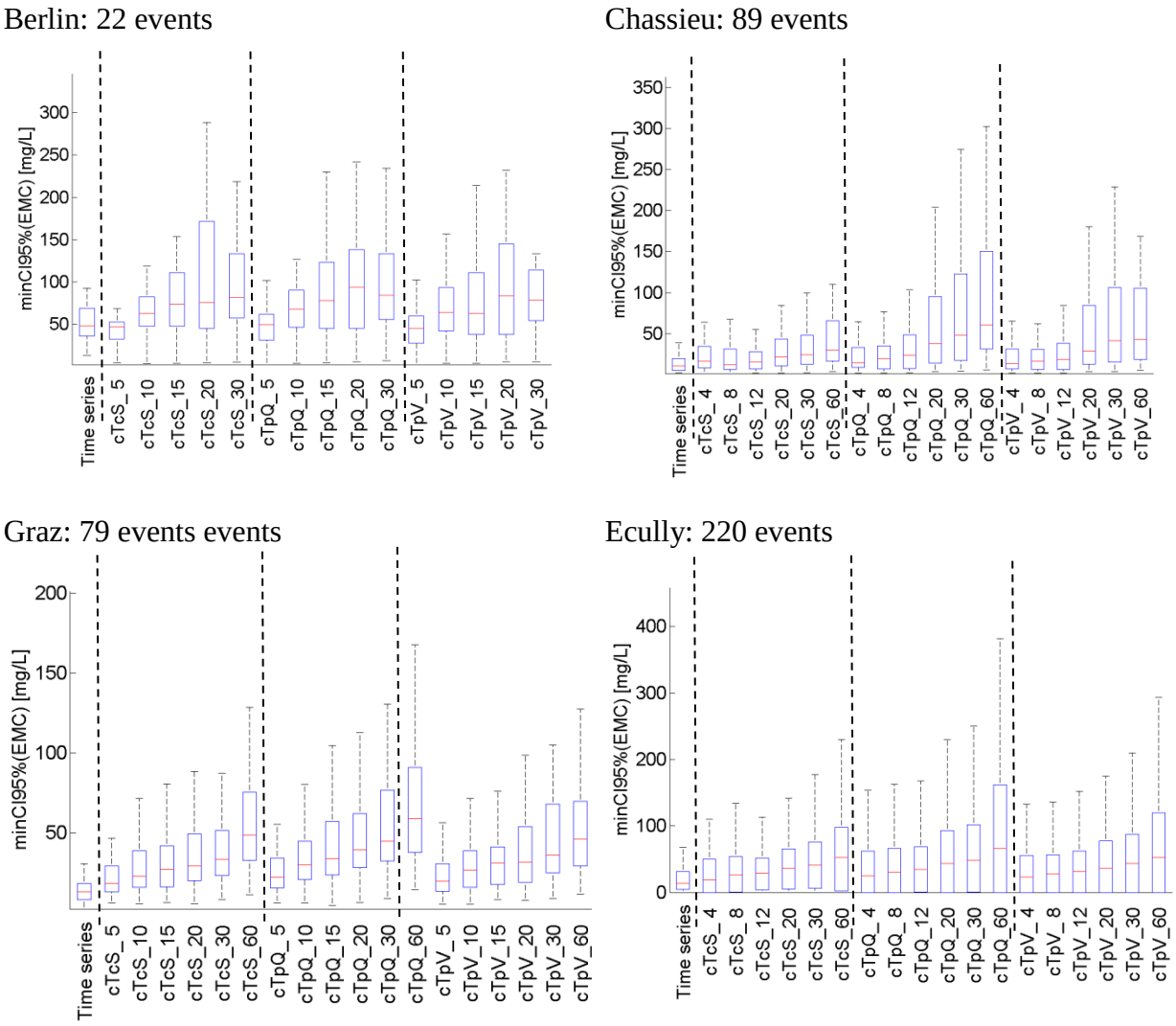


Figure 10. Comparison of $CI(EMC_{ref})_{min95\%}$ and $CI(EMC_{sim})_{min95\%}$ values obtained in all rainfall events for different sampling strategies.

Statistical tests show that there is a significant difference between the uncertainties of $CI(EMC_{ref})_{min95\%}$ and $CI(EMC_{sim})_{min95\%}$ (*t-test* or Wilcoxon with p-value < 0.05) for all datasets (except for Berlin with sampling intervals lower than 20 min). This implies that the uncertainties in the estimation of the EMCs by online measurements (turbidity or UV-VIS spectrometry) are significantly lower than uncertainties in the estimation of the EMC by sampling strategie (Wilcoxon test, p-value < 0.05). Furthermore, uncertainties of the EMC increases as longer sampling intervals are used. These results show up one of the potential advantages given by online monitoring compared to traditional sampling schemes and confirms the necessity of selecting an appropriate sampling interval for any sampling strategy to be adopted. Practitioners should bear in mind that EMCs estimated by a sampling strategy tends to be more uncertain than the EMCs obtained by online monitoring.

2.4 CONCLUSIONS

Different sampling strategies during rainfall events are simulated and evaluated by means of online TSS water quality and flow measurements in four international catchments (Berlin, Chassieu, Graz and Ecully). The average relative sampling error and the residuals distribution are estimated from EMCs simulated by strategies, compared to EMCs obtained by the complete time series from online monitoring of various rainfall events. The uncertainties are propagated by the Monte Carlo method using Latin Hypercube Sampling (LHS) and the sensitivity of the results to the different uncertainty sources is assessed by Sobol's Sensitivity Indices. For the studied datasets (Berlin, Chassieu, Graz and Ecully), a sampling volume proportional to runoff volume between two samples, with constant sampling intervals, strategy (cTpSV), delivers the most representative results in terms of accuracy (*MSRE*), bias (residuals vector) and uncertainties propagation of the errors in the estimation of EMCs. Recommended sampling time intervals are of 5 min for Berlin and Chassieu (resp. 100 and 185 ha area) and 10 min for Graz and Ecully (resp. 335 and 245 ha area), with average sampling errors between 7 % and 20 % and uncertainties in sampling errors of about 5 %, depending on the sampling interval.

From Sobol's total sensitivity index analyses, it can be stated that special attention should be paid to field sampling procedures and laboratory analyses (especially for larger sampling intervals), as uncertainties related to sampling volumes and TSS concentrations seemed to be highly influential on uncertainties of EMCs. In addition, uncertainties in the average relative sampling errors in the estimation of the EMCs are also very sensitive to uncertainties in the beginning and ending of rainfall events. Therefore, further investigations towards the assessment of this uncertainty source (in terms of water quantity and quality parameters) can be strongly recommended (see Métadier, 2011; Sandoval and Torres, 2013). Statistical tests indicate that the uncertainties of EMCs obtained from time series are significantly lower (Wilcoxon test, p-value < 0.05) than the uncertainties of EMCs obtained by sampling strategies.

CHAPTER 3. INFLUENCE OF SAMPLING INTAKE POSITION ON SUSPENDED SOLIDS MEASUREMENTS IN SEWERS: TWO PROBABILITY / TIME-SERIES BASED APPROACHES

Published in:

Sandoval, S., Bertrand-Krajewski, J.-L., (2016). Influence of sampling intake position on suspended solids measurements in sewers: two probability / time series based approaches. *Environmental Monitoring and Assessment*, 188, 347. doi: 10.1007/s10661-016-5335-y.

3.1 INTRODUCTION

Total Suspended Solids (TSS) measurements in urban drainage systems are often required for scientific, legal, environmental and operational reasons, as particulate matter constitutes a major source of surface water contamination (Ashley *et al.*, 2004; Chebbo and Gromaire, 2004). However, the reliability of TSS measurements strongly depends on the quality of the collected samples, which should be representative of real field conditions in the monitored sewer pipe (Larrarte, 2008; Métadier and Bertrand-Krajewski, 2012). Therefore, appropriate data acquisition and validation methodologies for TSS measurements in urban drainage systems are required (Bertrand-Krajewski and Muste, 2007). Aiming to estimate data quality, intensive investigations have been carried out towards assessment of uncertainties in online and laboratory TSS measurements (*e.g.* Joannis *et al.*, 2008; Métadier and Bertrand-Krajewski, 2011). However, the influence of field sampling conditions (*e.g.* sampling intake position, sampling flow velocities or sampling pipe orientation) on the representativeness of TSS measured values has not been equivalently addressed in the literature (Shelley, 1977; Berg, 1982; Rossi, 1998; Larrarte and Pons, 2011).

Indeed, one specific uncertainty source in TSS measurements lies in the sampling intake position through the sewer cross section (Shelley, 1977; Rossi, 1998; Kafi-Benyahia *et al.*, 2006; Larrarte, 2008; Larrarte and Pons, 2011), which is frequently neglected by implicitly assuming that point measured values are equal to the cross section mean concentration. This hypothesis seems to be valid in sewers with high enough flow velocities (Kafi-Benyahia *et al.*, 2006; Larrarte, 2008), where fully mixed flow conditions can be guaranteed (Raudkivi, 1998). However, TSS vertical gradients might be non-negligible under other hydrodynamic conditions corresponding to lower velocities (Verbanck, 1993, 1995; Ashley and Crabtree, 1992; Ashley *et al.*, 1994; Ristenpart, 1995; Ristenpart *et al.*, 1995; Ahyerre, 1999). In this case, the difference between point measured and cross section mean concentrations can be therefore attributed to: (i) variation of the position of the cross section mean concentration linked to vertical and horizontal concentration profiles due to hydrodynamics, (ii) variations of the sampling intake position through the cross section due *e.g.* to oscillations of the sampling tube or to other experimental constraints (Larrarte and Pons, 2011) and (iii) uncertainties in physical variables (*e.g.* flow rate measured at each time-step, roughness coefficient, geometric properties).

In this Chapter 3, two methods are presented and applied comparatively to a case study in order to evaluate how the position of the sampling intake tube, the vertical concentration profile and other uncertainty sources may lead to over- or underestimation of the cross section

mean *TSS* concentration. More specifically, the methods allow estimating the probability that sampling *TSS* concentrations at a random point of the sewer vertical cross section (for given a flow rate) leads to over- or underestimation of the true vertical mean *TSS* concentration. This information can then be used to correct possible errors in measured *TSS* data.

3.2 MATERIALS AND METHODS

Data set

The simplified method (SM) is applied with flow rate Q (m^3/s) data determined from water depth H (m) and mean flow velocity U (m/s) measurements. The time series method (TSM) in addition includes *TSS* concentrations (mg/L) estimated from online turbidity measurements. The experimental time series were measured with a 2 minute time step for 89 rainfall events in 2007 in Chassieu, France (see details in Chapter 1). For SM, information and data include: (i) the time series of the wet cross section A (m^2) derived from the H time series, (ii) the time series of the wet perimeter P (m) derived from the H time series, (iii) the sewer pipe slope S (m/m) and (iv) the settling velocity w_s (m/s) of suspended solids (details about settling velocity measurements are given in Torres and Bertrand-Krajewski, 2008a and Chebbo and Gromaire, 2009). Standard uncertainties in all above data are also needed for calculations. Input data and their uncertainties are given in Table 5.

Table 5. Input data and their standard uncertainties.

Input variable	Input value	Source	Standard uncertainty/ pdf	Source
Diameter D	$\overline{D}$ = 1.6 m	Métadier (2011)	$u(D)$ = 0.002 m	Muste *et al.* (2012)
Settling velocity w_s	$\overline{w_s}$ = 2.8e-4 m/s	Torres (2008)	$u(w_s) = 0.13\ \overline{w_s}$ normally distributed	Torres (2008); Torres and Bertrand-Krajewski (2008a)
Pipe slope S	$\overline{S}$ = 0.01 m/m	Métadier (2011)	$u(S)$ = 0 m/m	Métadier (2011)
Water depth H	$\overline{H}$ = H m	Métadier (2011)	$u(H)$ = 0.0075 m normally distributed	Métadier (2011)
TSS	$\overline{TSS}$ = *TSS* mg/L	Métadier (2011)	Variable over the time series	Métadier (2011)

Simplified method (SM)

The field conditions in which the samples are taken in the sewer pipe and the *TSS* concentration profile are illustrated in Figure 11 and Figure 12, where z (m) is the sampling intake depth, H (m) is the total water depth, $y_{\overline{C}}$ (m) is the depth corresponding to the vertical mean concentration, C_h (mg/L) is the concentration at depth h (m) above the invert, C_{a^*} (mg/L) is the concentration at the reference depth a^* (m) above the invert.

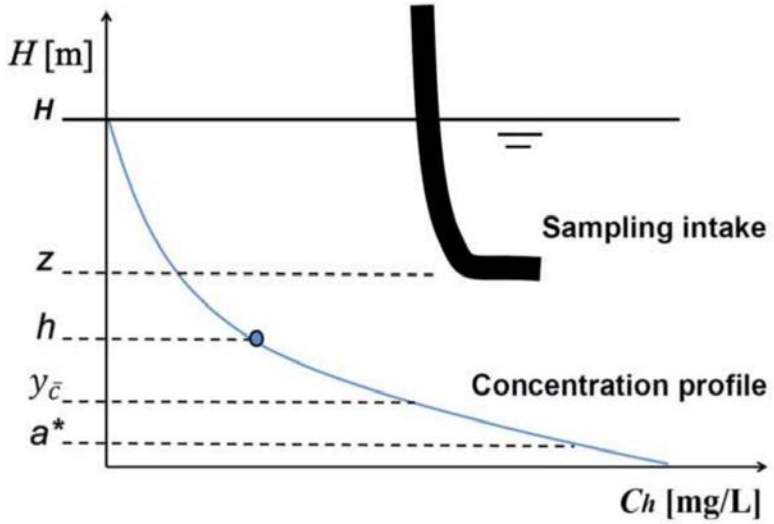


Figure 11. Scheme of the concentration profile and the sampling tube.

Figure 12. Sampling tube in the Chassieu sewer. Left: general view with the sampling tube located just upstream a Venturi flume created to ensure a minimum water level for sampling; Right: detail of the lower part of the sampling tube immersed in the flow.

Previous studies (Coleman, 1982; Verbanck, 2000) propose to represent the vertical concentration profile in a pipe with the following equation:

$$\frac{C_h}{C_{a^*}} = \left(\frac{h}{a^*}\right)^{-\eta} \qquad \text{Eq 3}$$

where $\eta = \frac{w_s}{\kappa U*}$ (-) is the Rouse number, C_h (mg/L) is the concentration at depth h (m) above the invert, C_{a*} (mg/L) is the concentration at the reference depth $a*$ (m) above the invert, $U*$ (m/s) is the shear velocity and κ =0.4 (-) is the Von Karman constant. The shear velocity is computed as follows (Verbanck, 1995):

$$U^* = \sqrt{\frac{gAS}{P}} \qquad \text{Eq 4}$$

where A (m^2) is the wet cross section, P (m) is the wet perimeter, S is the pipe slope (m/m) and g is the gravity (m/s^2). Thus, the Rouse number η can be written:

$$\eta = \frac{w_s}{\kappa}\sqrt{\frac{P}{gAS}} \qquad \text{Eq 5}$$

Hence, both the shape of the vertical concentration profile and the depth $y_{\overline{c}}$ of the mean concentration depend on the flow rate Q. For high enough values of Q (turbulent fully mixed condition), the vertical concentration profile might be expected to show a more uniform distribution along the vertical axis (as the Rouse number exponent η in Eq 3 gets lower than about 0.6) (see details in Raudkivi, 1998). In contrast, for low Q values, the *TSS* concentration will be higher near the pipe invert compared to the free surface. Although Eq 3 remains widely used in the context of urban drainage, one should bear in mind that alternative forms of this equation are obtained when considering additional hypotheses about velocity profile, flow stratification and hydrodynamic interactions (*e.g.* Verbanck, 2000; Cantero-Chinchilla *et al.*, 2016).

The problem therefore resides in estimating how much the measured *TSS* value at depth z deviates from the mean *TSS* concentration ($\overline{TSS}$) located at depth $y_{\overline{c}}$, with the above theoretical assumptions. The proposed method is illustrated in Figure 13. $y_{\overline{c}}$ is considered as a random variable with a cumulative distribution function (*CDF*) based on the shape of the vertical concentration profile proposed by Coleman (1982) and Verbanck (2000) (Eq 3), as there is usually a lack of data for calculating C_{a*} and a^*. In addition, their estimation by semi-empirical equations developed for specific site conditions and circumstances can lead to erroneous results, as these values have shown to be case-dependent (*e.g.* Ristenpart, 1995). The *CDF* curve is thus calculated by normalizing Eq 3, making it independent from C_{a*} and a^* and consistent with a *CDF* such that $P(z=0 > y_{\overline{c}}) = 0$ and $P(z=H > y_{\overline{c}}) = 1$ (Figure 13f). In general terms, $P(z > y_{\overline{c}}) = 0.5$ when z is equal to the median of $y_{\overline{c}}$, to be called $med[y_{\overline{c}}]$. The probability $P(z > y_{\overline{c}})$ that a random sampling depth z is higher than the point where $\overline{TSS}$ is located should be equal to $P(TSS < \overline{TSS})$, *i.e.* the probability of underestimating $\overline{TSS}$. The probability of non-exceedance of $\overline{TSS}$ can be thus computed by evaluating any sampling point z (where $0 \leq z \leq H$) over the *CDF* of $y_{\overline{c}}$ (Figure 13g). Cases for which z is lower than $med[y_{\overline{c}}]$ will result in a high probability of overestimating $\overline{TSS}$.

The Rouse number η plays a role in the shape of the *CDF*, leading to expect a dynamic behavior of $P(TSS < \overline{TSS})$ for different Q values. Regarding situations in which $med[y_{\overline{c}}]$ is closer to the pipe invert, *i.e.* low Q values and curved vertical concentration profiles, a

sampling point z close to H (*i.e.* near the free surface) might lead to a high probability of underestimating $\overline{TSS}$. For high enough values of Q with fully mixed condition, $P(TSS < \overline{TSS})$ or $P(z > y_{\overline{c}})$ can be expected to be close to 0.5, for all $0 \leq z \leq H$. In theory, the optimum scenario is obtained if the sampling point z is equal to $med[y_{\overline{c}}]$.

With the Monte Carlo (MC) method, two sources of variability in $P(TSS < \overline{TSS})$ are accounted for: (i) normal distributions to account for measurement uncertainties in physical variables: Q and h both at each time step, w_s and D (Figure 13a and Figure 13b); and (ii) uniformly distributed sampling depth z along the vertical axis (Figure 13d). With this method, 1000 MC simulations have been done respectively for these two sources, with index $i = 1$ to 1000 and $j = 1$ to 1000 correspondingly. Uncertainties in variables Q, h, D and w_s, are represented by independent normal distributions with the mean equal to the measured value and the standard deviation equal to the standard uncertainty as proposed in (ISO, 2009).

Considering that the sampling depth z_j is uniformly distributed, the lower and upper limits of the distribution are defined as 0.25 H and 0.75 H. The basis of this hypothesis is the technical arrangement of the sampling tube in Chassieu: the tube is attached to the roof of the 1.6 m diameter circular sewer, it is mobile thanks to a rotation axis and its position varies with Q and H. This variation is not a confidence interval and is neither controlled but the above range is assumed to represent acceptably the existing field conditions (Figure 12). The uncertainties in the physical variables allow calculating the pdf (probability density function) of η (Figure 13c) at each time-step of the rainfall event and consequently, for each Monte Carlo simulation i, a concentration profile C_i is established (Figure 13e). The *CDF* of y_i for the i-th simulation is then calculated by normalizing the concentration profile in such a way that $P(0 > y_{\overline{c}}) = 0$ and $P(H > y_{\overline{c}}) = 1$ (Figure 13e to Figure 13f).

For each simulation (i, j), $P(z_j > y_{\overline{c}})_i$ is different. Consequently, the MC runs lead to a complete description of the $P(z_j > y_{\overline{c}})_i$ probabilistic behavior, including mean values and confidence intervals (expressed as 5 % and 95 % percentiles). The analysis is undertaken by comparing the evolution of $E[P(z > y_{\overline{c}})]$ (mean of all simulations), 5 % and 95 % percentiles of $P(z > y_{\overline{c}})$ (or $P(TSS < \overline{TSS})$) versus the flow mean velocity U_m and/or the water level H. Given the fact that, in practice, the minimum possible sampling height is the sampling tube diameter (4 cm), water levels H lower than 4 cm were not considered ($U_m < 0.63$ m/s). It is assumed as well that *TSS* values are measured in the centroid of the sampling tube (2 cm from the tube borders).

It is worth to note that sampling campaigns are frequently carried out with commercially available autosamplers with sampling tube inner diameters in the range 10-20 mm. In such cases, the intake position is more a point along the vertical axis than in the Chassieu case with a 4 cm tube. Sampling errors may thus be potentially more significant.

Time series method (TSM)

If online *TSS* time series are available, a second method is proposed for comparison with the above one. The general idea is to describe the variability of *TSS* concentrations for a given Q value. Consequently, with the hypothesis previously established, the variability of *TSS* values for a given Q value can be attributed to hydrodynamic conditions through the cross section (concentration gradients, turbulence, sampling position variability). Thus, assuming the presence of concentration vertical gradients or concentration profiles (this hypothesis could not be properly checked in Chassieu as simultaneous TSS samplings at various points along the vertical axis were not available), the variability of *TSS* values for a given Q value might be

explained by (i) the variation of the sampling tube position, (ii) the concentration vertical profile and (iii) uncertainties of *TSS* concentrations.

Given that the number of *TSS* values related to a specific *Q* value becomes lower as *Q* is more atypical in the data set, *TSS* values are grouped by *Q* classes (or Rouse number η, which depends on *Q* according to Eq 5). The classes are defined by setting the same (or nearly the same) number of *TSS* measurements within each η class. This data arrangement is also made with the purpose of having a better distribution and similar representativeness of the results for all classes. For each η class, a TSS_i value is randomly selected by using a uniform distribution. The number of times TSS_i is lower than $\overline{TSS}$ is counted and divided by the total number of simulations ($i = 1 : 10000$). This calculation gives an estimation of $P(TSS < \overline{TSS})$ for a given η class. The purpose of sampling multiple *TSS* values in a given η class with a uniform distribution is to be consistent with the hypotheses established in SM. In SM, the samples are assumed to be collected according to a uniform distribution along the vertical cross section (from 0.25 *H* to 0.75 *H*). This assumption about the uniform-sampling is independent from the field data distribution of the *TSS* values along the vertical axis for a given *Q* value. For SM, the shape of the *TSS* vertical distribution is based on the shape of the dimensionless vertical profile (Figure 13e, f). For TSM, the shape of the distribution is found by grouping the *TSS* data in η classes. The pdf of the *TSS* values due to the vertical variability given by the two approaches (SM and TSM) shows to be consistent under a visual inspection (similar to an exponential distribution, as in Figure 13f). The η values (in a η class) are also expressed versus velocity U_m and water depth *H* for comparative purposes with SM. In addition, uncertainties in *TSS* measurements will result in additional uncertainty in $P(TSS < \overline{TSS})$. To take this into account, MC simulations with $j = 1 : 1000$ runs are carried out to propagate *TSS* uncertainties into $P(TSS < \overline{TSS})$. Therefore, the curve $P(TSS < \overline{TSS})$ as a function of U_m (or *H*) could be estimated for each simulation *j*, leading to calculate the mean simulation $E[P(TSS < \overline{TSS})]$ of these curves with its percentiles 5 % and 95 %.

It is worth mentioning that in TSM the relation between *TSS* variability, sediment profile and flow rate is assumed to be independent of the rainfall event, as the complete *TSS* rainfall-weather time series is grouped by η classes (*Q* values). However, the relationship between *TSS* variability and flow rate is rather dependent on each particular rainfall event (due to *TSS* availability and variability on catchment surfaces, rainfall intensities and surface runoff rates, etc.): this inter-event variability has been neglected in our hypotheses.

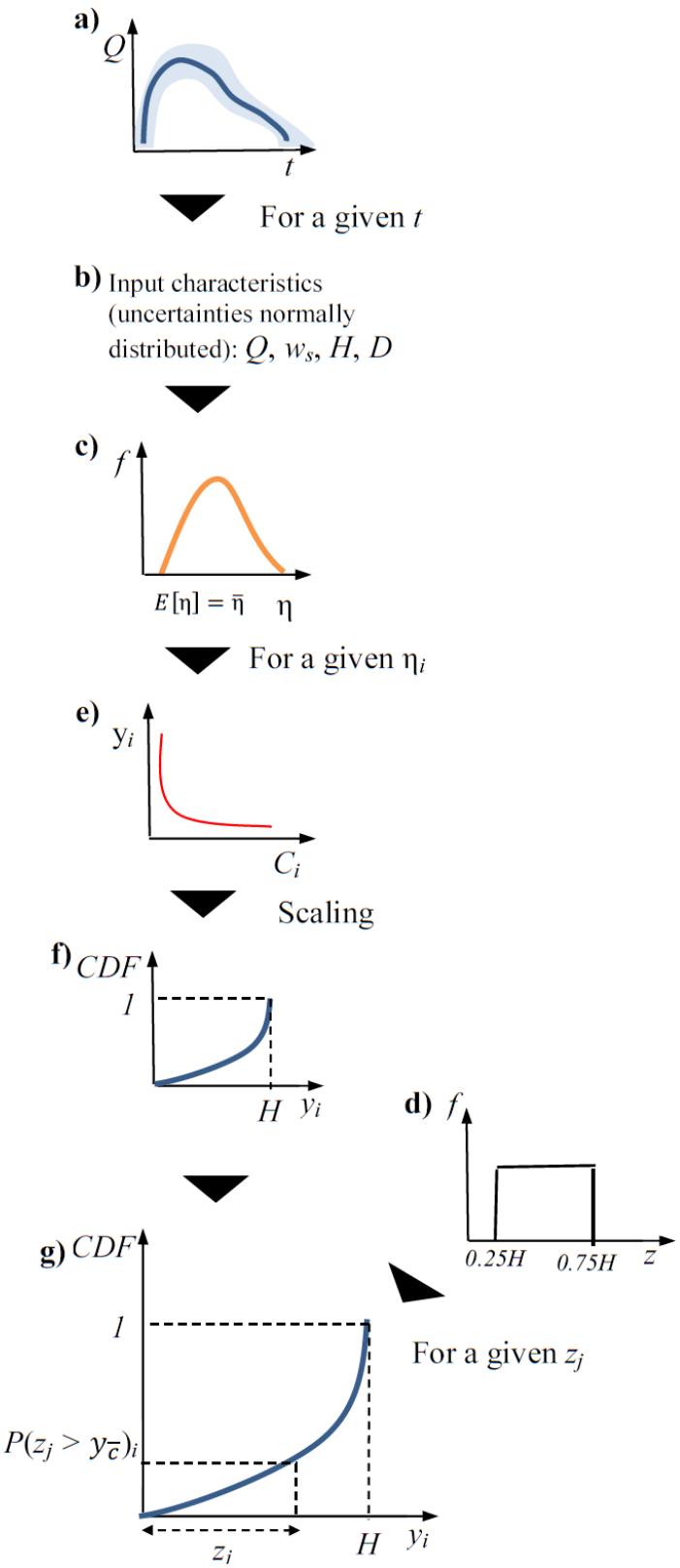


Figure 13. Diagram of SM for a Monte Carlo *i*, *j* simulation.

Uncertainties due to underestimations

$P(z > y_{\overline{c}})$ (or $P(TSS < \overline{TSS})$) gives the probability of sampling a *TSS* value lower than the mean concentration (underestimation). Converting this probability value into a direct uncertainty in the *TSS* measurement is relevant for practical purposes. Therefore, a bias factor *K* can be proposed to assess this uncertainty source in *TSS* measurements (Eq 6).

$$\overline{TSS} = K \times TSS_{measured} \qquad \text{Eq 6}$$

where *K* might be computed as K_{sm} for SM and K_{tsm} for TSM (Eq 7).

$$K_{sm} = \frac{E[y_{\overline{c}}]}{med[y_{\overline{c}}]} \quad and \quad K_{tsm} = \frac{E[TSS]}{med[TSS]} \propto K_{sm} \qquad \text{Eq 7}$$

If *z* is equal to $med[y_{\overline{c}}]$ then K_{sm} is equal to one. In case of TSM, K_{tsm} = 1 for $E[TSS] = med[\overline{TSS}]$ (no asymmetry in the *CFD* of *TSS*). The *K* factors obtained by SM and TSM are calculated and compared, aiming to estimate an order of magnitude of the bias in *TSS* concentrations (under- or over-estimation, if *K* is respectively < 1 or > 1). The main hypothesis is that the absolute difference between $E[y_{\overline{c}}]$ and $med[y_{\overline{c}}]$ is proportional to the absolute difference between $E[TSS]$ and $med[TSS]$ obtained with real TSS measurements (TSM). This assumption with SM allows the comparison both approaches to describe the variability of the *TSS* under- or over-estimation as a function of U_m or *H*.

3.3 RESULTS AND DISCUSSION

Both methods are applied to 89 rainfall events monitored in 2007 in Chassieu. An example is shown in Figure 14 for SM with a value of Q = 0.023 m^3/s (corresponding to U_m = 1 m/s). In this case, $E[z]$ = 0.5 *H* = 2.3 cm and $P(2.3 \text{ cm} > y_{\overline{c}})$ = 0.88 (Figure 14). The variability of the curve (represented by its thickness) due to uncertainties in the Rouse number is negligible for this example. Therefore the probability distribution of the point where the mean concentration is located is constant, once uncertainties in physical variables are considered. The graph shows that $med[y_{\overline{c}}]$ = 0.05 *H* = 0. 3 cm, and thus $P(0.\ 3 \text{ cm} > y_{\overline{c}})$ = 0.5 (best theoretical measurement scenario in agreement with the SM hypotheses). The mean value $E[y_{\overline{c}}]$ is different from the median, which is coherent with the assumption of asymmetry of the *CFD*. Therefore, $z = E[y_{\overline{c}}]$ is not the best sampling point as $E[y_{\overline{c}}]$ = 0.16 *H* = 0.7 cm and $P(0.7 \text{ cm} > y_{\overline{c}})$ = 0.68. This means that there is a greater probability of obtaining lower *TSS* values than at 0.05 *H* (the $med[y_{\overline{c}}]$ location) where $P(TSS < \overline{TSS})$ = 0.5 (same probability of under- or over- estimation).

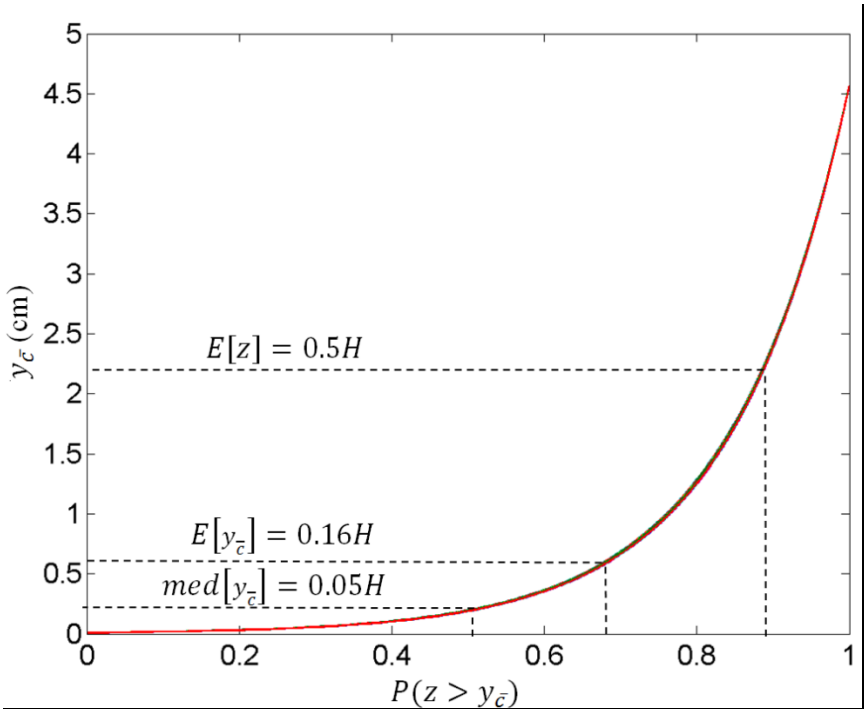


Figure 14. SM example of the CDF of $y_{\bar{c}}$, including *med*$[y_{\bar{c}}]$, *E*$[y_{\bar{c}}]$ and *E[z]* for illustrative purposes (Q = 0.023 m^3/s).

For all rainfall events, $P(TSS < \overline{TSS})$ (or $P(z > y_{\bar{c}})$) is constant for the analyzed U_m values with SM (Figure 15a). The curve of the vertical concentration profile is expected to be more pronounced as U_m values are lower (Eq 3). This fact is considered significant when $H < 4$ cm (sampling tube height), in which the $P(TSS < \overline{TSS})$ estimations become higher. However, these results are not included here as sampling at depths lower than 4 cm is not feasible from a practical point of view (diameter of the sampling tube equal to 4 cm in our case). *E*$[P(TSS < \overline{TSS})]$ is 0.88 for all velocities U_m (Figure 15a), with 5 % and 95 % percentiles of *E*$[P(TSS < \overline{TSS})]$ equal to 0.80 and 0.95 respectively.

Rouse number η lower than 0.6 has been found to correspond to fully mixed and homogeneous suspension along the vertical axis by experimental studies (see *e.g.* Raudkivi, 1998). Therefore, Eq 3 under this range of Rouse numbers can be expected to be closer to a fully mixed and homogeneous condition. Although the complete dataset of Chassieu showed lower Rouse numbers (η max = 0.04), the obtained *CDF*s of $y_{\bar{c}}$ do not correspond to a fully uniform distribution, even for high U_m values and 5 % and 95 % percentiles (Figure 15a), as $P(TSS < \overline{TSS})$ is always greater than 0.5: this is mainly due to the use of a dimensionless concentration profile which to some extent restricts the possibility to get a completely fully mixed profile for the considered U_m velocities.

This fact might lead to conclude that a uniformly-distributed behavior of the concentration for high *Q* values could not be clearly observed with SM. The interquartile range *IQR* of $P(TSS < \overline{TSS})$ is approximately 0.08, without varying as U_m values increase. This *IQR* is mainly due to uncertainties in *z* (Figure 13d), as the shape of the *CDF* is not very sensitive to uncertainties in η (*e.g.* Figure 14). The value *med*$[y_{\bar{c}}]$ is about 5 % of *H* for all U_m values. The magnitude of K_{sm} showed the possible asymmetry between the sampling point and the proposed *CDF*. *TSS* underestimations are approximately 269 %, as K_{sm} is 3.69 for all U_m values with negligible *IQRs* (from Eq 6 and Figure 15b). This reflects the significant asymmetry of the *CDF* proposed within this approach.

a) b)

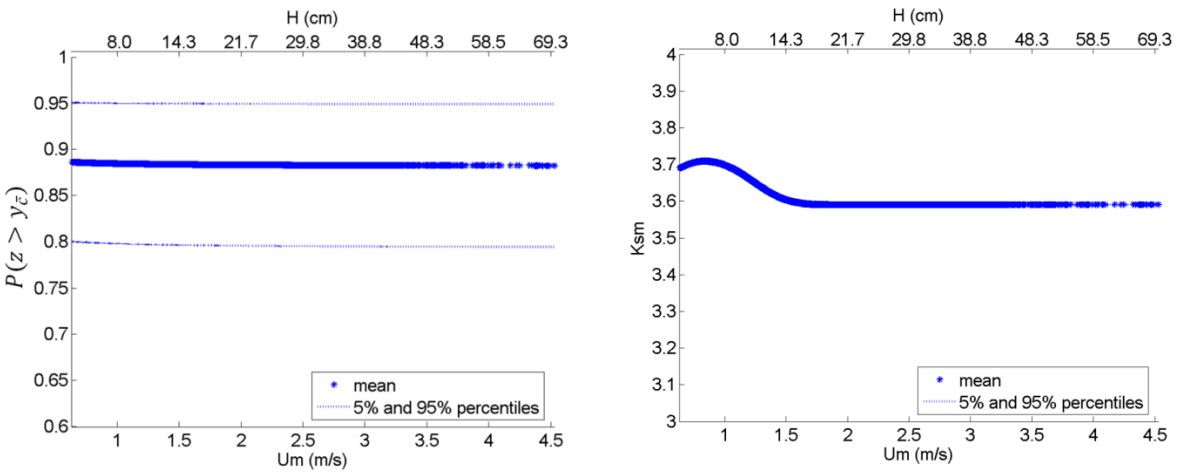


Figure 15. a) $P(z > y_{\bar{c}})$ and b) K_{sm} obtained by SM versus U_m (m/s) and H (cm).

Regarding the time series method (TSM), the η classes were defined based on the criterion exposed in the previous section, leading to the *TSS* distribution within each η class shown in Figure 16. The length of each η class does not vary significantly, having a similar number of samples within each class (about 6560 samples). This can be explained by the fact that the data amount is higher for lower *Q* values. As there is an inverse relation between η and *Q*, the definition of classes based on η values tends to distribute the data more uniformly among the η classes. In average, $P(TSS < \overline{TSS})$ is about 0.64 for all U_m values, including estimations ranging from 0.57 up to 0.73 (Figure 17a). The interquartile range *IQR* of $P(TSS < \overline{TSS})$ within this approach is 0.02. Likewise to SM, no trend could be detected towards a value of $P(TSS < \overline{TSS}) = 0.5$ as U_m increases (towards a fully mixed condition, although η numbers are relatively small). The K_{tsm} factor is about 1.39, showing probable *TSS* underestimations of about 39 % (from Eq 6) (Figure 17b). Although the K_{tsm} factor is not constant for the range of analyzed U_m values (from 0.64 m/s to 1.05 m/s), no clear trend could be neither appreciated in K_{tsm} for higher U_m and *H* values. However, the variability of the K_{tsm} is significant, as this factor ranged from 1.21 to 1.82, *i.e.* underestimations from 21 % up to 82 % (with a negligible *IQR*) (Figure 17b).

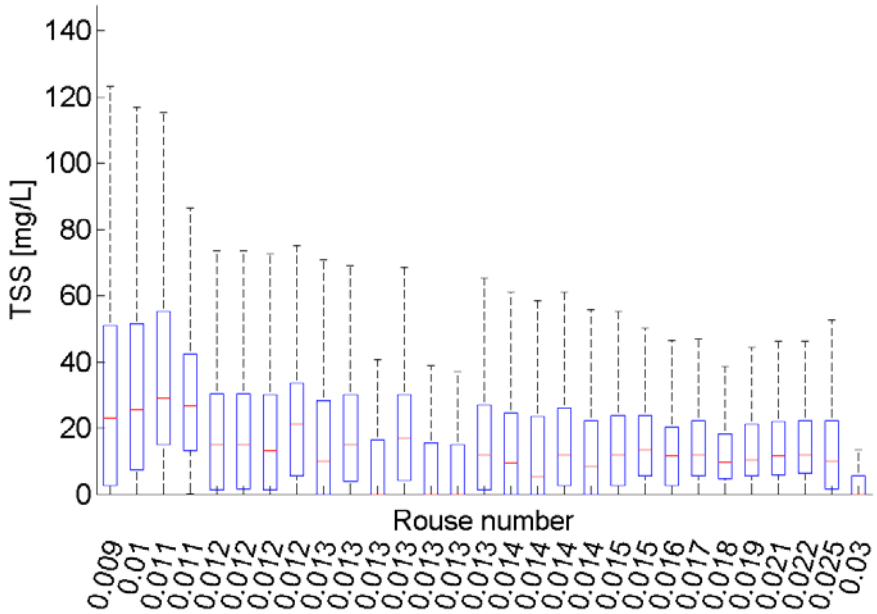


Figure 16. TSS distribution within each η class for TSM (about 6560 samples per class).

Globally, the $P(TSS < \overline{TSS})$ given by TSM (about 0.64) is lower than by SM (about 0.88). The ideal measurement scenario recommended by SM is for the case $E[z] = med[y_{\overline{c}}] \approx 0.05\,H$, as $P(0.05\,H > y_{\overline{c}}) = 0.5$. TSM shows a more realistic value of *TSS* underestimations (about 39 %) compared to SM, which shows *TSS* underestimations of about 269 % under the hypothesis that K_{sm} and K_{tsm} are comparable. SM can point towards recommendations about the best sampling point (at 0.05 *H*, for the case study). Nevertheless, conclusions drawn from this methodology strongly depend on SM hypotheses, especially the shape of the *CDF*.

a) b)

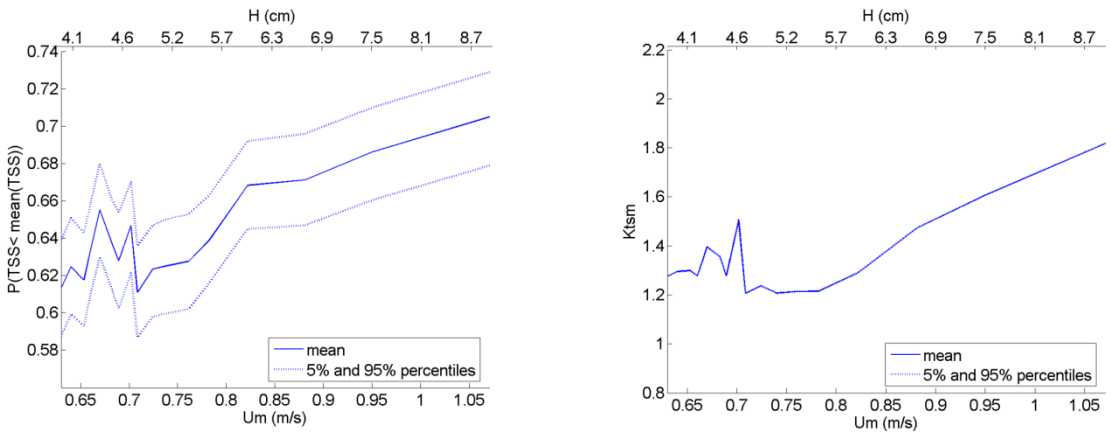


Figure 17. a) P(TSS < $\overline{TSS}$) and b) K_{tsm} with TSM versus U_m (m/s) and H (cm).

Based on this case study, TSM seems preferable to estimate *TSS* uncertainties, as $P(TSS < \overline{TSS})$ and K_{tsm} values are more alike to ranges previously established in the literature (*e.g.* Rossi, 1998). The *K* factor and the $P(TSS < \overline{TSS})$ values indicate an important asymmetry in the *TSS* distribution along the cross section (as $P(TSS < \overline{TSS}) > 0.5$, even if $\eta < 0.6$), a fact that is strongly recommended to be considered in *TSS* uncertainties assessment.

3.4 CONCLUSIONS

The two methods presented in this Chapter 3 aim to assess uncertainties in *TSS* measurements during rainfall events due to the sampling tube intake position along the vertical axis in a sewer. The simplified method (SM) is based on flow rate and water depth time series. The time series method (TSM) requires additional *TSS* time series as input data. If online *TSS* time series are available, TSM appears as an applicable strategy for assessing the variability of *TSS* concentration as a function of the flow rate. Otherwise, SM is proposed as an alternative approach. However, SM is less site-specific (no local *TSS* data are used) and provides rougher estimations, based on further hypotheses harder to check on site.

The probability of underestimating the cross section mean *TSS* concentration is estimated to be approximately 0.88 within SM. In case of TSM, the probability is about 0.64 for all velocity values. Interquartile ranges are higher for SM (*IQR* = 0.08) than for TSM (*IQR* = 0.02). TSM shows more realistic *TSS* underestimations (about 39 %) than SM (about 269 %). Differences between the two methods are mainly due to simplifications in SM (absence of *TSS* measurements and operation of the sampling system). SM can estimate the measuring depth at which the probability of over estimation is equal to the probability of underestimation (about 0.05 *H*, with the proposed hypotheses). SM assumes a significant asymmetry of the *TSS* concentration profile along the vertical axis in the cross section. This is compatible with the distribution of *TSS* measurements found in TSM.

The proposed methods can be used to calculate indicators of (i) the measurement quality and (ii) the representativeness of *TSS* measurements, expressed as the probability of under- or over-estimation of the true cross section mean TSS concentration. This probability may be useful for correcting errors in *TSS* raw data.

GENERAL CONCLUSIONS OF PART 1

Different sampling strategies during rainfall events are simulated and evaluated by means of online TSS and flow measurements in four international catchments (Chassieu-France, Ecully-France, Berlin-Germany, Graz-Austria). The average relative sampling error and the residuals distribution are estimated from EMCs simulated by the studied strategies and compared to EMCs obtained by the complete time series from online monitoring of various rainfall events. For the studied datasets (Berlin, Chassieu, Graz and Ecully), a sampling volume proportional to runoff volume between two samples, with constant sampling intervals, strategy (cTpSV), delivers the most representative results in terms of accuracy (Mean Square Relative Error *MSRE*), bias (from residuals analysis) and uncertainties propagation. Recommended sampling time intervals are of 5 min for Berlin and Chassieu (resp. 100 and 185 ha area) and 10 min for Graz and Ecully (resp. 335 and 245 ha area), with average sampling errors between 7 % and 20 % and uncertainties in sampling errors of about 5 %, depending on the sampling interval. From Sobol's total sensitivity index analyses, it can be stated that special attention should be paid to field sampling procedures and laboratory analyses (especially for larger sampling intervals), as uncertainties related to sampling volumes and TSS concentrations seemed to be highly influential on uncertainties of EMCs. Statistical tests indicate that the uncertainties of EMCs obtained from time series are significantly lower than the uncertainties of EMCs obtained by sampling strategies (Wilcoxon test, p-value < 0.05).

These results from Chapter 2 led to hypothesize about the potential errors in TSS data, specifically in the average EMC, estimated by means of different sampling strategies, compared to the average EMC obtained by considering the complete time series and uncertainties from online measurements. These findings highlight the influence of temporal resolution and uncertainties of TSS data over the EMCs estimations, bringing evidence of possible biases in TSS pollutographs (different means of the pollutographs) when data is obtained by sampling strategies instead of online measurements. This fact might significantly affect the calibration and performance of TSS pollutographs and loads stormwater intra-events models, depending on the data used as input (monitoring campaigns or online measurements). Therefore, data from TSS online monitoring is retained for modelling purposes in Part 3 (TSS modelling). Data obtained from monitoring campaigns (sampling strategies) could still be used as an input for intra-event TSS models in the absence of online measurements, accounting for potential errors or biases in the EMCs of the pollutographs by means of modelling strategies such as bias correction factors included in the calibration process.

As the main sources of error in any sampling procedure are not only due to the heterogeneity in time of TSS but also due to the sampling technique, uncertainties in online and laboratory TSS individual values are also a relevant aspect. The influence of field sampling conditions on the uncertainties of TSS measured values is often considered to be represented by the analytical uncertainty of the laboratory analysis. However, field sampling conditions might have a significant additional contribution to the total uncertainty of the TSS values. With the purpose of verifying this hypothesis, Chapter 3 presents two methods that aim to assess uncertainties in a TSS measurement at time step t during rainfall events due to a scarcely described uncertainty source in the literature: the sampling tube intake position along the vertical axis in a sewer. The simplified method (SM) is based on flow rate and water depth time series. The time series method (TSM) requires additional *TSS* time series as input data. If online *TSS* time series are available, TSM appears as an applicable strategy for assessing the

variability of *TSS* concentration as a function of the flow rate. Otherwise, SM is proposed as an alternative approach. However, SM is less site-specific (no local TSS data are used) and provides rougher estimations, based on further hypotheses harder to check on site.

The probability of underestimating the cross section mean TSS concentration is estimated to be approximately 0.88 within SM. In case of TSM, the probability is about 0.64 for all velocity values. Interquartile ranges are higher for SM (IQR = 0.08) than for TSM (IQR = 0.02). TSM shows more realistic TSS underestimations (about 39 %) than SM (about 269 %). Differences between the two methods are mainly due to simplifications in SM (absence of TSS measurements and operation of the sampling system). SM can estimate the measuring depth at which the probability of over estimation is equal to the probability of underestimation (about 0.05 H, with the proposed hypotheses). SM assumes a significant asymmetry of the TSS concentration profile along the vertical axis in the cross section. This is compatible with the distribution of TSS measurements found in TSM.

Chapter 3 revealed a power law describing the TSS as a function of flow rate, including higher variances of TSS for higher flow rates, by means of analysis for the TSM. This information is useful for modelling intra-event TSS load dynamics, giving insights about the complexity for representing pollutants loads dynamics. For example, in Chapter 6 it is established that expressing TSS as a monotonic function of flow rate is not sufficient to appropriately represent TSS values, especially as flow rate is higher. However, one should bear in mind that global results and conclusions obtained from approaches in Chapter 2 and Chapter 3 assume implicitly that all rainfall events are representative and comparable, *i.e.* results for a given event are generalizable to the others. This inter-event generality is questioned and further analyzed in Part 2 by the use of hydrological model-based approaches, grouping rainfall events with similar characteristics and discarding unrepresentative events. These aspects are applied to TSS modelling in Part 3, jointly with discussions and reflections from the present Part 1.

PART 2 UNCERTAINTY ASSESSMENT IN A CONCEPTUAL HYDROLOGICAL MODEL AND RAINFALL DATA

The flow at the outlet of urban drainage systems can be seen as the result of multiple anthropic and non-anthropic complex processes developed within different temporal and spatial scales, of which the main driving force is the water that precipitates over the catchment (Leonhardt *et al.*, 2014; Kretzschmar *et al.*, 2016). Most rainfall-runoff model structures used nowadays in hydrology can be classified as conceptual models (Wagener *et al.*, 2003), which are models that offer a suitable balance between computational simplicity and physical meaning of their parameters (Wheater *et al.*,1993; Zhang *et al.*, 2015). Indeed, hydrological simulations by these simplified mathematical structures can be a suitable description of the rainfall-runoff process in urban catchments for several purposes (*e.g.* Reed *et al.*, 2004; Coutu *et al.*, 2012). The hydrological simulation by conceptual models has been therefore applied for prediction of the impact of future change, decision-making processes, improvements of hydrological understanding and situations in which data assimilation is not possible (Beven and Smith, 2014).

Therefore, a Conceptual Rainfall-Runoff (CRR) model has been tested with information of 365 rainfall events from the Chassieu urban catchment (Lyon, France), measured between 2004 and 2011 (see details in Chapter 1; Métadier, 2011; Sun *et al.*, 2015). The model consists of a single reservoir lumped model described by Eq 8, Eq 9 and Eq 10. The effective rainfall input $X_{net}(t)$ (L/s) is calculated from $X_{obs}(t)$ (mm/h) by the Horton infiltration model (Eq 8 and Eq 9). The infiltration rate at time *t*, $f(t)$ (mm/h) in Eq 8, is dependent on three parameters: initial and final infiltration rates f_0 (mm/h) and f_c (mm/h), respectively, and the decay constant k (min^{-1}) specific to the soil. The *S* value is the effective area of the catchment, *i.e.* 80 ha (Eq 9). The single reservoir lumped model is established in Eq 10, for calculating the simulated flow $Y_{sim}(t)$ (L/s) as a function of $X_{net}(t)$. Three additional parameters are included in the linear reservoir: the lag time of the reservoir *K1* (min), an additional advective delay *Td* (min) and *q* (L/s) as an additive term of the output to represent the baseflow (Eq 10). The parameters of this single reservoir lumped model are listed in Table 6 (from Sun and Bertrand-Krajewski, 2013a).

$$f(t) = f_c + (f_c - f_0) \cdot e^{-kt} \quad \text{Eq 8}$$

$$X_{net}(t) = (X_{obs}(t) - f) \cdot S \cdot 10000/3600 \quad \text{Eq 9}$$

$$Y_{sim}(t) = e^{-\frac{\Delta t}{K1}} \cdot Y_{sim}(t - \Delta t) + \left[1 - e^{-\frac{\Delta t}{K1}}\right] X_{net}(t - T_d) + q \quad \text{Eq 10}$$

Table 6. List of parameters for the model calibration

Parameter (θ)	Unity	Values [min, max]
f_c	mm/h	[0, 5]
f_0	mm/h	[0, 120]
k	min^{-1}	[0, 5]
T_d	min	[0, 60]
$K1$	min	[1, 120]
q	L/s	[0, 20]

The selection of this particular model structure is based in its performing capabilities in relation to its simplicity, as suggested from previous studies using the same dataset (Sun and Bertrand-Krajewski, 2013b). However, the performance of such simplified models depends strongly on parameters and their uncertainties being well identified from previous analyses or experiments (Thyer *et al.*, 2009).

Indeed, determining a strategy for assessing the uncertainty of parameters given by data errors, including as well the variability related to the selected calibration period or rainfall events remains as a key aspect (Thyer *et al.*, 2009; Ebtehaj *et al.* 2010; Guerrero *et al.*, 2013; Gharari *et al.*, 2013; Ye *et al.*, 2014). Chapter 4 compares two single-event and multiple-event based strategies for parameters estimation to a novel strategy for the above CRR hydrological model. The proposed strategy consists in grouping the local parameters estimations obtained from local calibrations of rainfall events, according to connectivity criterion. This connectivity allows classifying the events in groups of hydrological families, leading to express global uncertainties of the parameters as conditional probability functions. Local parameter uncertainties and optimal values are estimated by means of a Bayesian approach and the DREAM algorithm (Vrugt *et al.*, 2016). The benefits of simulating rainfall events by using parameter conditional probability functions are discussed, demonstrating the advantages of implementing this parameter assessment strategy in terms of different performance metrics (*e.g.* Dotto *et al.*, 2013; Ye *et al.*, 2014). The local parameters transferability is compared to different statistical depth definitions, as measures of multivariate centrality into the parameters probability density function (Bardossy and Singh, 2008). The proposed approach is adopted as well for the water quality modelling in Part 3.

It is worth to mention that there are cases in which the local optimal parameters obtained for a certain rainfall event are not able to reproduce other calibration events, or neither the rainfall event itself (non-reproducible events from the analysis presented in Chapter 4). Indeed, one of the facts to which these discrepancies may be attributed is the errors in the input rainfall (Leonhardt *et al.*, 2014; Kretzschmar *et al.*, 2016; Del Giudice *et al.*, 2016). Rainfall errors may have various origins, one significant case being the use of local rain gauge registrations as direct inputs, without considering spatio-temporal variability of the rainfall (Kavetski *et al.*, 2006a; Schellart *et al.*, 2012; Kretzschmar *et al.*, 2016). Aimed to correct rainfall measurements and assess their uncertainties, a rainfall error model-based approach is studied in Chapter 5. Hence, four error rainfall models are proposed by mixing a multiplicative error model, a reverse modelling approach and a constant/variable time-window methodology. The

capabilities of the four rainfall error models to identify errors in rainfall data are tested by comparing the errors in rainfall identified by the models to previously introduced error known structures in the original rainfall data.

In Chapter 5 rainfall measurements are corrected in events that are non-reproducible from the analysis presented in Chapter 4, by means of the selected rainfall error model. The adopted values of the parameters to be used for calculating the corrected rainfall time series in non-reproducible events are obtained from the conditional probability classification analysis presented in Chapter 4. Afterwards, general conclusions about Part 2 are given.

CHAPTER 4. STRATEGY FOR ASSESSING PARAMETERS OF A RAINFALL-RUNOFF MODEL BY CONNECTIVITY REPRESENTATIONS AND CONDITIONAL PROBABILITY FUNCTIONS

Extended version of:

Sandoval S., Bertrand-Krajewski J.-L. (2017). Strategy for assessing parameters of a rainfall-runoff model by connectivity representations and conditional probability functions. *Proceedings of the 14th International Conference on Urban Drainage, Prague*, Czech Republic, 10-15 September, 3 p.

4.1 INTRODUCTION

The problem of parameters assessment consists in identifying and integrating the related uncertainty sources through the modelling chain (Sikorska *et al.*, 2014). Probably the main types of uncertainty sources that influences model parameter estimations in CRR modelling are (adapted from Guerrero *et al.*, 2013): errors in data (input/output) (*e.g.* Renard *et al.*, 2011; Del Giudice *et al.* 2016), model structure (*e.g.* Clark *et al.*, 2008; Marshall *et al.*, 2007) and selected data for calibration (*e.g.* Ebtehaj *et al.*, 2010; Gharari *et al.*, 2013) with their interactions (*e.g.* Renard *et al.* 2010; Sun and Bertrand-Krajewski, 2013a).

The BATEA Bayesian approach (see Kavetski *et al.*, 2006a; 2006b; Kuczera *et al.*, 2006) is a commonly accepted framework to conceptualize the propagation of errors given by data (input/output) and model structure into parameters estimation by means of a solid conceptual basis (Yang *et al.*, 2008). However, it can also be recognized from this approach that the low identifiability and the ill-posed nature of the calibration problem when errors in data (input/output) and model structure are explicitly accounted in the inference scheme and vague prior information about both sources is included (Kavetski *et al.*, 2006a; Renard *et al.* 2010). Multiple authors used this conceptual framework to focus on specific uncertainty sources over parameters estimation, given its flexibility. For example, (i) input uncertainties: Thyer *et al.*, 2009; McMillan *et al.*, 2011; Sun and Bertrand-Krajewski, 2013a; Del Giudice *et al.* 2016 and (ii) model structure uncertainties: *e.g.* Krysztofowicz, 2002; Bayesian Model Averaging (BMA) approaches (*e.g.* Duan *et al.*, 2007; Marshall *et al.*, 2007). Further contributions in the literature towards the assessment of uncertainty sources have been also proposed by alternative formal statistical and non-formal techniques (*e.g.* Abbaspour *et al.*, 2004; Montanari and Brath, 2004; Montanari and Grossi, 2008; Kretzschmar *et al.*, 2016; Fuentes-Andino *et al.*, 2017). However, a global consensus on how these error sources interact and a calibration methodology that allows to account them separately is still far from being reached, and thus, their interpretation is merely hypothetical (Sikorska *et al.*, 2014).

One simplified possibility for applying BATEA is to account for rainfall and model structure uncertainty sources into parameters uncertainties without further separation (Renard *et al.* 2010; Dotto *et al.*, 2011), an approach to be called from now on Bayesian Merged Uncertainties (BMU). The BMU tends to deliver more biased parameter estimations as errors in data (especially in input rainfall for hydrology, *e.g.* Del Giudice *et al.* 2016) and model structure are higher (*e.g.* Thyer *et al.*, 2009). For those cases, model residuals are less alike to hold the hypothesis of Identically Independent Distributed (i.i.d.), condition on which the

mathematical strength of the method relies (Vrugt *et al.*, 2016). As a potential solution to this inconvenient, advanced likelihood functions have been introduced, aimed to mimic complex nontraditional error residuals distributions (Schoups and Vrugt, 2010; Evin *et al.*, 2014; Scharnagl *et al.*, 2015). The immediate difficulty when using this statistically consistent approach is once again identifiability and ill-posed situations, as further parameters with inexistent prior information should be added into the inference scheme (adapted from Vrugt *et al.*, 2016). Moreover, from an epistemological perspective, the inclusion of a likelihood function with additional parameters can be seen as coupling a black-box extension (in the sense of missing a physical interpretation) to the initial hydrological model structure, whose repeatability and generality are therefore questionable (adapted from Nearing *et al.*, 2016). However, BMU calibrations have shown consistent results from a practical point of view under multiple simplified contexts, successfully avoiding obstacles related to the computational demand of the Bayesian inference by using the DREAM algorithm (see examples Vrugt *et al.*, 2016). Stressing its simplicity, the BMU may be sounder to explore the inter-event variability of parameters by undertaking local event-by-event calibrations (*e.g.* Thyer *et al.*, 2009; Singh and Bardossy, 2012; Sun and Bertrand-Krajewski, 2013a). Indeed, multiple further studies have revealed the variability of the parameter estimations given by the use of different time periods/rainfall events on calibration under different contexts (Srikanthan *et al.*, 2009; Ebtehaj *et al.*, 2010; Brigode *et al.*, 2013; Gharari *et al.*, 2013; Sikorska *et al.*, 2014; Thirel *et al.*, 2015; Bisselink *et al.*, 2016).

Based on the idea of repeatability (including more than 255 calibration events for the study case) and the local event-dependent nature of rainfall errors, this event-by-event calibration approach brings in principle the possibility of dealing with rainfall uncertainties under a systematic inspection. Local parameter estimations in which the rainfall data is able to acceptably explain the flow rate can be then considered as representative (adapted from Ye *et al.*, 2014; Ajmal *et al.*, 2015; Fenicia *et al.*, 2016) (strategy to be called single-event calibration SE). However, the SE strategy might fail for flexible but erroneous model structures that, by means of local calibrations, can mimic the flow rate from severely corrupted rainfall data. This shortcoming can be tackled from the idea that identifying the best parameter set relies also on the selection of time periods with similar hydrological characteristics (Seibet, 2003). Local parameter estimations can be then grouped based on the transferability concept, defined as the predictive capacity of parameters obtained from different rainfall events (Bardossy and Singh, 2008; Singh *et al.*, 2016).

Therefore, the global nature of model structure uncertainty and the inter-event parametric variability is addressed based on these concepts and the results of SE strategy, proposing a novel parameter estimation strategy (to be called single-event conditional, SEConditional). The main idea is to divide parameters marginal probability function (formed by all the sets local parameter estimations in the SE strategy) into conditional probability functions (formed by groups of sets of local parameters estimations). For this purpose, an adjacency matrix that reflects how local parameter estimations are interconnected to the other calibration rainfall events is constructed under a transferability perspective (analogue to a leave-all-out cross validation scheme). This adjacency matrix can be represented as a graph of “connected” rainfall events. The graph is analyzed by clustering techniques to construct then the conditional probability functions (see clusters applications over transferability indicators *e.g.* Singh *et al.*, 2016).

For many applications in prediction and for calculating various performance indicators (*e.g.* Nash-Sutcliffe criterion – NS -), one single set of parameters as an “optimal” estimator, is desired (Bardossy and Singh 2008; Bennett *et al.*, 2013; Leonhardt *et al.*, 2014). However, a

unique “best” set of parameters from the global probability functions (marginal: SE or conditionals: SEConditional) cannot be directly selected by comparing the likelihood values obtained from local BMU calibrations in different rainfall events, given the relative nature of likelihood estimations (adapted from Vrugt *et al.*, 2016). Therefore, the geometrical consistency of the marginal (SE) and conditional (SEConditional) probability functions is evaluated by seeking for a reference frame that appropriately describes the increasing relationship between the depth of the sets of parameters inside the probability functions and their transferability (Bardossy and Singh, 2008; Chebana *et al.*, 2010; Guerrero *et al.*, 2013). If a geometrically consistent frame of reference is found (measure of deepness), the evaluation of the “deepest” parameter is direct, depending on the adopted definition of statistical depth (see details Pokotylo *et al.*, 2016). Six measures of statistical depth are explored.

Therefore, Chapter 4 presents a comparative study between three global parameter assessment strategies applied to the studied CRR model and study case: (i) mean of the set of local optimal parameters obtained from all acceptable event-by-event calibrations (SE), (ii) traditional multiple-events simultaneous calibration (strategy to be called ME) (*e.g.* Tan *et al.*, 2008; Mancipe-Munoz *et al.*, 2014, among many others) and (iii) the proposed strategy based on the results of local single-event calibrations (SEConditional). The estimation of rainfall-runoff model parameters and of their uncertainties with SEConditional aims to diminish the uncertainty bounds of runoff predictions (precision), maximize the number of measurements inside the uncertainty bounds (reliability), keeping or even improving the mean prediction for the verification events (accuracy). The performance of the model is evaluated for 110 verification rainfall events.

4.2 METHODOLOGY

θ is the set of parameters of the hydrological model (f_c, f_0, k, T_d, $K1$, q) and $p(\theta/Y)$ their probability density function (pdf), given a series of flow rate observations Y_{obs}. The BMU, widely used in hydrological modelling (Vrugt *et al.*, 2016), allows to calculate $p(\theta/Y)$, named posterior distribution, over the basis of a likelihood function and a prior knowledge of the distribution of parameters $p(\theta)$, which is expressed by Eq 11.

$$p(\theta/Y) = C \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\hat{\sigma}_t^2}} \exp\left[-\frac{1}{2}\left(\frac{Y_{sim}(t,\theta) - Y_{obs}(t)}{\hat{\sigma}_t^2}\right)^2\right] \cdot P(\theta) \qquad \text{Eq 11}$$

where n is the number of flow-rate data Y_{obs}, $Y_{\mathrm{sim}}(t, \theta)$ is the simulated flow rate by the model at a given time step t from the observed rainfall X_{obs} and a set of parameters θ, $p(\theta)$ is a uniform probability distribution for each parameter (informative-less), C is a normalization coefficient and $\hat{\sigma}_t^2$ is the residual variance, considered for this application to be equal to the squared value of flow rate $Y_{obs}(t)$ standard uncertainty. The DREAM algorithm is used for determining $p(\theta/Y)$ as a solution to Eq 11 (Vrugt *et al.*, 2016). The set of parameters that represents the optimal parameters values (for this case the set of values which maximizes the likelihood) among all probable values $p(\theta/Y)$ is called $\theta_{opt.}$ The $p(\theta/Y)$ function is estimated from three different approaches, by the use of the first 255 events for calibration (the last 110 remaining events are used for verification). A 70 % / 30 % ratio for calibration / verification is established from recommendations given by comparable studies (Mourad *et al.*, 2005). The verification events are used to evaluate the performance in prediction by the parameters

estimated from each of the three calibration strategies. For each approach, an identification criterion is proposed in order to discard non-reproducible events by the CRR model from the calibration phase.

- Single-event strategy (SE): rainfall events are calibrated with the Bayesian method (Eq 11) and the DREAM algorithm, by using the data $X_{obs\,i}$ and $Y_{obs\,i}$ for the calibration rainfall event i. This process leads to estimate an optimal set of parameters $\theta_{opt\,i}$ and a $p(\theta/Y)_i$ function for each rainfall event i (i = 1 : 255). The calibration rainfall events i for which $\theta_{opt\,i}$ has a NS < 0.75 are considered as non-reproducible by the hydrological model and the $\theta_{opt\,i}$ estimation is discarded. The global function $p(\theta/Y)$ is calculated as the marginal probability of all the functions $p(\theta/Y)_j$ and the global set of optimal parameters θ_{opt} as the "deepest" point inside $p(\theta/Y)$ (see Bardossy and Singh, 2008), with the non-discarded events j.

- Multi-event strategy (ME): the calibration is done globally with the ensemble of all calibration rainfall events, with X_{obs} and Y_{obs} as two vectors (Eq 11) and with the DREAM algorithm. This approach leads to estimate directly a global set of optimal parameters θ_{opt} (based on the maximum likelihood, Vrugt *et al.*, 2016) and a global function $p(\theta/Y)$. The formulation of a criterion for identifying non-reproducible rainfall events by this approach is less straightforward. It is worth to point out that an alternative parameter estimation strategy could be proposed by assigning a different set of parameters to each event and undertaking the calibration simultaneously. However, this will massively increase the dimension of the calibration problem (into a total of 1530 parameters for the CRR and the calibration dataset), making it expensive in computational resources. This limitation is especially unfeasible for Monte Carlo parameters inference schemes such as the Bayesian method with the DREAM algorithm, due to the number of iterations to be undertaken (of the order of 1e10 in this case). A plus to this alternative approach is that the inter-event correlation for each parameter of θ could be assessed directly, opening research directions for inter-event stochastic modelling of parameters.

- Single-event strategy conditional (SEConditional): the 255 sets of optimal parameters, $\theta_{opt\,i}$ obtained by the SE approach, are classified into n types, with the aim of re-grouping the events with similar hydrological characteristics. The underlying hypothesis is that two rainfall events i and j (i = 1 : 255, j = 1 : 255 and $i \neq j$) are connected if the optimal set of parameters $\theta_{opt\,i}$ obtained for the event i is able to reproduce as well the rainfall event j and if $\theta_{opt\,j}$ is also able to reproduce the rainfall event i, in both cases with a Nash-Sutcliffe criterion NS > 0.75 (following a leave-all-out cross validation scheme). A symmetric adjacency matrix (AM) is constructed with $AM(i, j)$ = 1 and $AM(j, i)$ = 1 if the calibration rainfall events i and j are connected and $AM(i, j)$ = 0 otherwise. The diagonal of AM ($i = j$) is filled with zeros by convention. The proposed AM represents also a graph, in the sense that events (or optimal local sets of parameters) are nodes and the potential connections between events i and j (or optimal local sets of parameters $\theta_{opt\,i}$ and $\theta_{opt\,j}$) are edges, when $AM(i, j)$ = 1. The graph represented by AM is undirected (symmetric AM), as assigning an orientation to the connections or edges lacks of interpretability. When the connectivity of the graph is verified, two different scenarios can be obtained: (i) the graph is completely connected (all nodes have at least one connection) (Figure 18a), (ii) the graph is not entirely connected and is divided in sub-graphs (Figure 18b).

a) b)

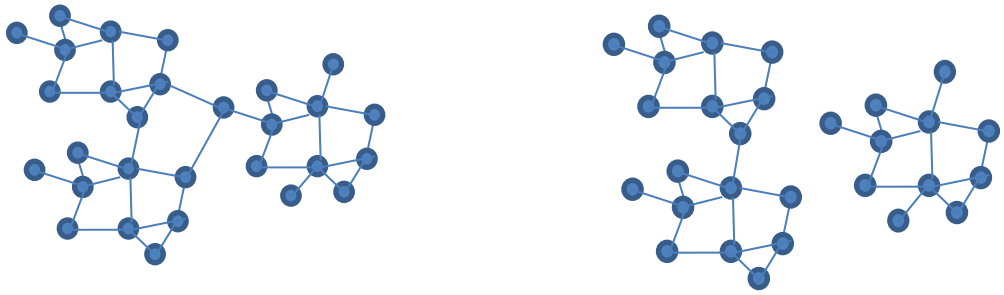


Figure 18. Topologic representation of a) connected graph (29 nodes) and b) disconnected graph into 2 sub-graphs (18 and 11 nodes resp.).

All events that belong to sub-graphs with less than 10 nodes are labeled as non-reproducible events and are discarded from the analysis, including completely disconnected events (*i.e.* sub-graphs with one node). This, given that a minimum of 10 nodes (sets of parameters) was required for calculating the deepness of all nodes inside a sub-graph, when considering definitions of statistical depth such as the Potential depth (Aizerman et al., 1970). For the scenario of Figure 18b the division into $n = 2$ sub-graphs (including a number of nodes equal or greater than 10) delivers the number of clusters (hydrological groups) to be used directly (2 clusters). For the scenario described in Figure 18a, the number of n clusters in the graph can be defined by testing with a supervised or unsupervised cluster algorithm (predefined or not predefined number of clusters). As no direct grouping by sub-graphs is found for the case study (situation of example in Figure 18a), a supervised clustering technique with $n = 2$ groups is tested with the total graph for identifying groups of connected rainfall events, with the "spinglass.community" function, spins = 2 (supervised cluster for two groups) (see Reichardt and Bornholdt, 2006) of the package "igraph" (Csardi and Nepusz, 2006), implemented in R (R Development Core Team, 2017). The purpose of this algorithm is essentially to "maximize" the connectivity density inside each cluster (Reichardt and Bornholdt, 2006). The number of $n = 2$ hydrological groups, labelled as $T1$ and $T2$, used to clusters the graph and classify the rainfall events can be supported on: (i) a suitable separation shown by a further Principal Component Analyses (PCA) (Pearson, 1901) by using two groups, (ii) parameters in $p(\theta/Y)$ showing bimodal behaviors in certain of their marginal univariate distributions (especially parameters T_d and $K1$) that could be explained from the separation given by $p(\theta/Y, T1)$ and $p(\theta/Y, T2)$, (iii) satisfactory results in the verification stage from using $p(\theta/Y, T1)$ and $p(\theta/Y, T2)$ as parameter estimations rather than with the marginal distribution $p(\theta/Y)$, (iv)less performant results were obtained by further tests with $n > 2$ conditional probability functions in prediction. Therefore, the selection of $n = 2$ conditional probability functions (T1 and T2) to explain the variability of the marginal distribution p(θ/Y) is considered as appropriate for the case study. These criteria can be useful for defining the number of clusters (hydrological groups), for further case studies with different datasets and models.

Therefore, the function $p(\theta/Y)$ is calculated as in the SE approach (but discarding non-connected sub-graphs instead of local estimations with NS < 0.75), and then divided into two conditional probability functions $p(\theta/Y, T1)$ and $p(\theta/Y, T2)$, by labeling each local $p(\theta/Y)_i$ as a type $T1$ and $T2$ from the cluster of graph analysis. θ_{opt} is defined as well by two values $\theta_{opt}T1$ and $\theta_{opt}T2$, calculated as the "deepest" point inside $p(\theta/Y, T1)$ and $p(\theta/Y, T2)$ resp. (see Bardossy and Singh, 2008). The classification of a verification rainfall event in a hydrological group (type $T1$ or $T2$), for deciding if it is more pertinent to use $p(\theta/Y, T1)$, $\theta_{opt}T1$ or $p(\theta/Y,$

*T*2), $\theta_{opt}T2$ for verification, is done based on the mean and maximum rainfall intensity. This criteria for deciding which rainfall characteristic to use for classifying a given verification rainfall event into group *T*I or *T*2 was obtained by applying a Classification Decision Tree (Breiman *et al.*, 1984) and a PCA (Pearson, 1901), including four various potential rainfall characteristics as explicative variables (*e.g.* mean, max. rainfall intensity, ADWP and the beginning date of the event) without improving the performance of the Classification Decision Tree and the visual separation of the groups by PCA. In case no satisfactory separations (from the cluster technique) or explanations to the separations (from Classification Decision Tree, PCA or another data exploratory analysis) can be found, one can work with $n = 1$ and the estimation of $p(\theta/Y)$ will be undertaken directly as the marginal of all connected events. This will lead to obtain a very similar estimation as in SE Strategy, especially if the model structure is robust enough (in a transferability sense, see Bardossy and Singh, 2008) and parameters uncertainties are mostly due to local rainfall errors.

With the purpose of verifying the geometrical consistency of $p(\theta/Y)$ (SE strategy), $p(\theta/Y, T1)$ and $p(\theta/Y, T2)$ (SEConditional strategy), and therefore defining the "deepest" set of parameters for each function under an appropriate reference frame, six measures of statistical depth are explored, from the "ddalpha" package in R (R Development Core Team, 2017) (see details Pokotylo *et al.*, 2016): Tukey depth (Tukey, 1974), Mahalanobis depth (Mahalanobis, 1936), Projection Depth (Donoho, 1982), Spatial depth (Chaudhuri, 1996), Zonoid depth (Dyckerhoff *et al.*, 1996) and Potential depth (Aizerman *et al.*, 1970). Given that $p(\theta/Y)$ in the ME strategy is estimated from a global multi-event implementation of the likelihood function, the "deepest" set of parameters can be directly identified as the parameter set with the maximum likelihood.

The benefits of the proposed strategy (SEConditional), regarding traditional parameter assessment strategies (SE, ME), are highlighted by comparing observed and simulated flow rates in 110 verification rainfall events. Three performance metrics are used for this purpose (comparable to *e.g.* Dotto *et al.*, 2013; Ye *et al.*, 2014; Del Giudice *et al.*, 2016): (i) the NS criterion for accuracy (Eq 12); (ii) the ARIL criterion for precision (Vezzaro and Mikkelsen, 2012) (Eq 13); (iii) the modified POC criterion for reliability (from *e.g.* Ye *et al.*, 2014) (Eq 14 and Eq 15).

$$NS = 1 - \frac{\sum_{t=1}^{n}\left(Y_{sim}\left(t,\theta_{opt}\right) - Y_{obs}\left(t\right)\right)^2}{\sum_{t=1}^{n}\left(\bar{Y}_{obs} - Y_{obs}\left(t\right)\right)^2} \quad \text{Eq 12}$$

$$ARIL = \frac{1}{n}\sum_{t=1}^{n}\frac{Limit_{sup,t} - Limit_{inf,t}}{Y_{sim}\left(t,\theta_{opt}\right)} \quad \text{Eq 13}$$

$$C_t = \begin{cases} 1 & Y_{obs}\left(t\right) - 2\hat{\sigma}_t \le Limit_{sup,t} \text{ and } Y_{obs}\left(t\right) + 2\hat{\sigma}_t \ge Limit_{inf,t} \\ 0 & otherwise \end{cases} \quad \text{Eq 14}$$

$$modified\ POC = \frac{1}{n}\sum_{t=1}^{n} C_t \quad \text{Eq 15}$$

where n is the number of flow rate data Y_{obs} and $\bar{Y}_{obs}$ its mean value. $Y_{\mathrm{sim}}(t, \theta)$ is the simulated flow rate by the model at time step t from the observed rainfall X_{obs} and global optimal parameters θ_{opt}. $Limit_{sup,t}$ and $Limit_{inf,t}$ are the upper and lower limit for a confidence bound of 95 % at a time step t, obtained from $p(\theta/Y)$ or $p(\theta/Y, T1)$ and $p(\theta/Y, T2)$ depending on the case. The modified POC proposed for this work is based on the traditional POC (percentage of coverage) (see *e.g.* Ye *et al.*, 2014) but also considering flow rate $Y_{obs}(t)$ standard uncertainty $\hat{\sigma}_t$ (Eq 14 and Eq 15). The idea is that the modified POC includes as "reasonably explicable by the model" flow rate data that, with their measurement uncertainty, overlaps the parametric uncertainty bounds (similar to approaches such as Harmel *et al.*, 2007). This modified POC is a more flexible (in the sense that modified POC will deliver higher values than traditional POC) but realistic description of reliability than POC, as uncertainties in measurements are taken into account.

The total simulated output uncertainty given by the residuals of the model from a set of optimal parameters is not completely captured by parametric uncertainties. This phenomenon can be attributed to remnant errors in calibration that are not accounted for by the input and structural error assumptions (Thyer *et al.*, 2009). Therefore, the total simulation output uncertainty is estimated from propagating the parametric uncertainties given by $p(\theta/Y)$ (SE and ME strategy) and $p(\theta/Y, T1)$ or $p(\theta/Y, T2)$ (SEConditional strategy) as the ensenmble of multiple realizations of the model residuals. For comparative purposes, the total simulation output uncertainty is estimated by following the method proposed by Dotto *et al.* (2011). The residuals obtained from the rainfall events used in calibration (making the distinction between $T1$ and $T2$ for SEConditional) are binned as a function of modelled flow rates. This permits to construct probability distributions of residuals as a function of the modelled flow rate values. For a verification event, as the modelled flow rate value at t is a function of a set of parameters, a different probability distribution of residuals is obtained by considering each realization of $p(\theta/Y)$ (SE and ME strategy) or $p(\theta/Y, T1)$ or $p(\theta/Y, T2)$ (SEConditional strategy). The total uncertainty of simulated flow rate at t will be then given by the ensemble of all the probability distributions of residuals obtained at t, as a result of propagating the parametric uncertainties (see further details Dotto *et al.*, 2011). Therefore, two complementary indicators, to be called Total ARIL and Total POC (modified), are calculated for the estimated total output uncertainty simulations in the 110 verification events.

4.3 RESULTS AND DISCUSSION

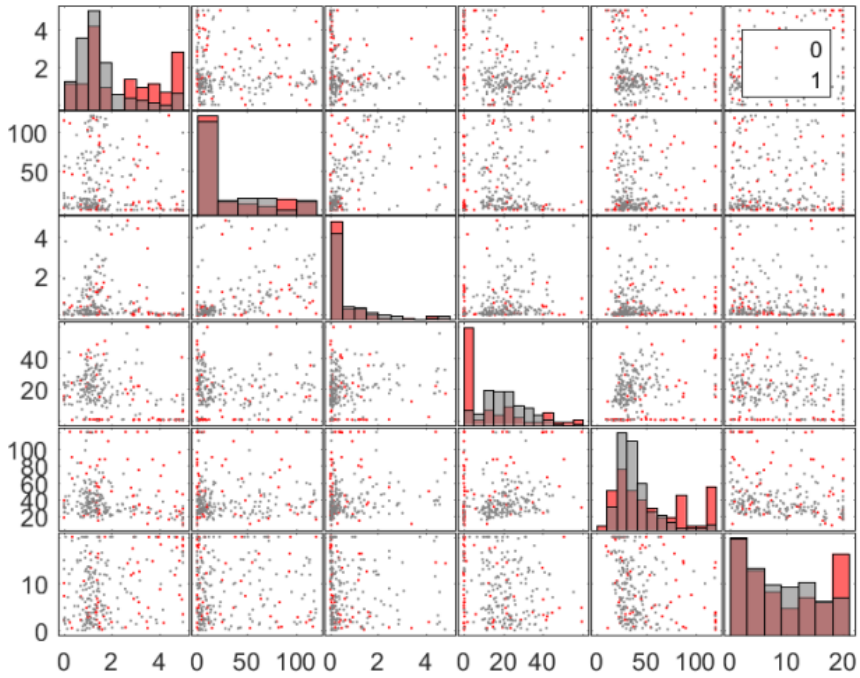
The results are divided into analyses of the calibration data (255 events), containing the assessment of the parameters by the 3 estimation strategies, jointly with the evaluation of the relation transferability - statistical depth of sets of parameters. Afterwards, the predictive capacity of the parameters estimations is evaluated with the verification data (110 events). The verification/calibration events are selected chronologically, an initial period for calibration and the subsequent period for verification. The influence of the selection of the calibration/verification data is also discussed.

Calibration: application of the parameters estimation strategies

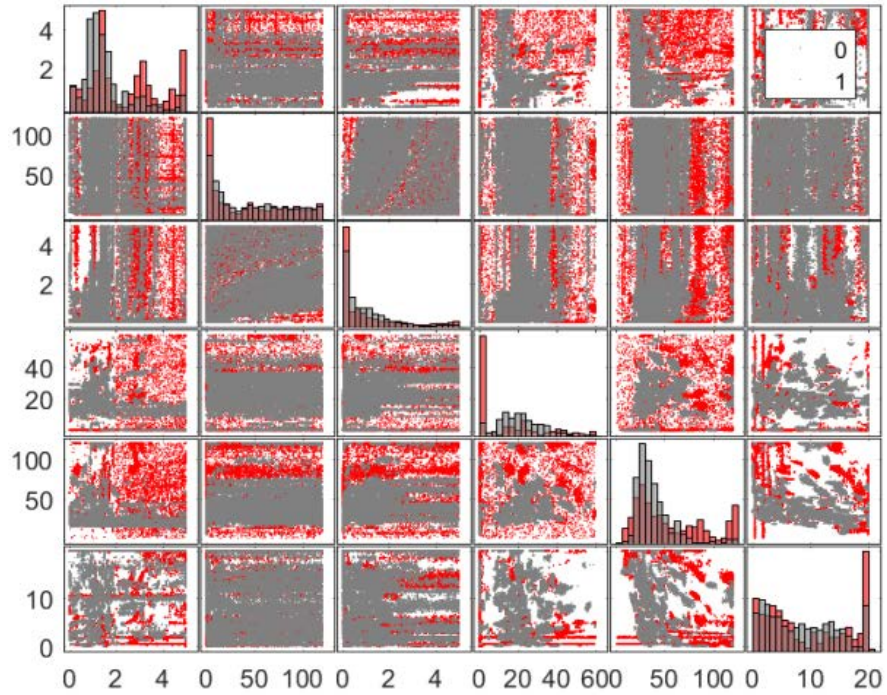
The correlation plot for parameter estimations θ (f_c, f_0, k, T_d, $K1$, q) with the SE and ME strategies by using the DREAM algorithm are presented in Figure 19. It can be seen how

considering uncertainties in the estimation of local set of parameters $p(\theta/Y)_i$ brings up a higher dispersion in the $p(\theta/Y)$ global estimation rather than the sets of local optimal set of parameters $\theta_{opt\,i}$ (Figure 19a,b). 28 % of the rainfall events are identified as non-reproducible and therefore discarded for the SE Strategy, with NS <0.75 for the local estimation of set $\theta_{opt\,i}$ (red estimations of $\theta_{opt\,i}$ in Figure 19a and of $p(\theta/Y)_i$ in Figure 19b). As expected, the estimations of $p(\theta/Y)$ are less dispersed for ME than SE, given that more data are included in the likelihood function (Eq 11).

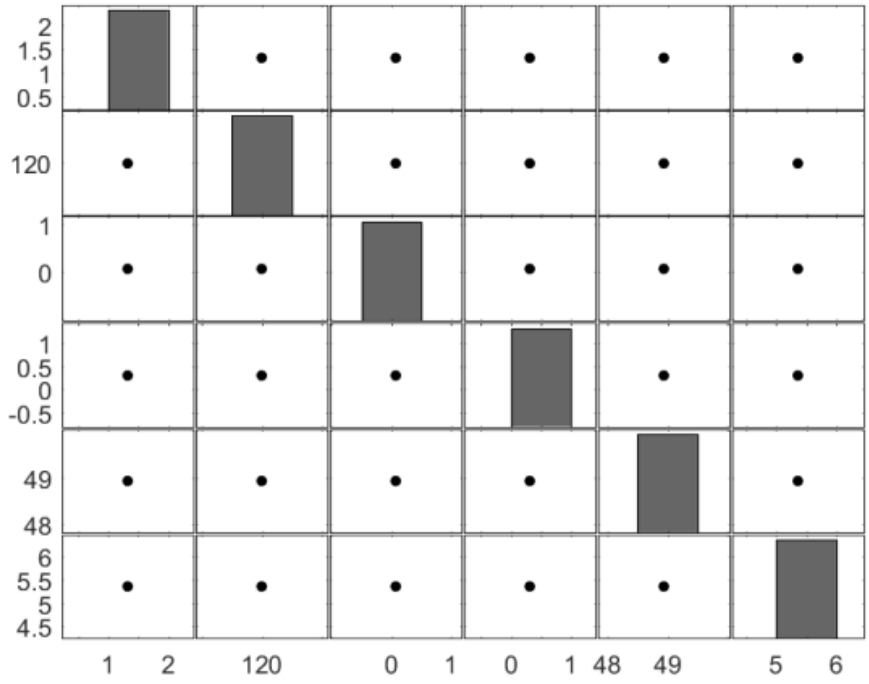
a)



b)



c)



d)

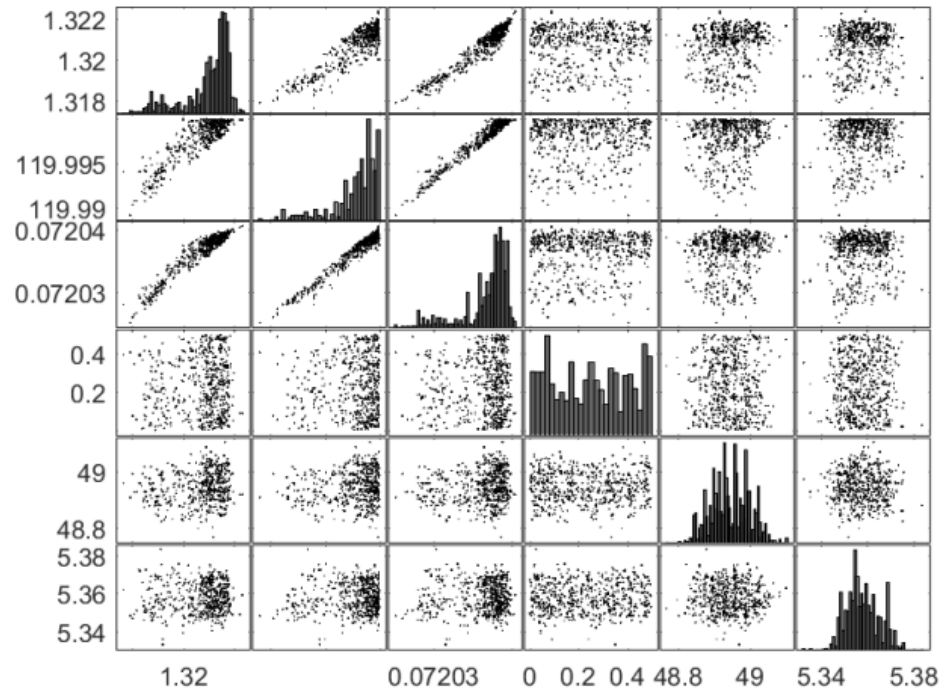


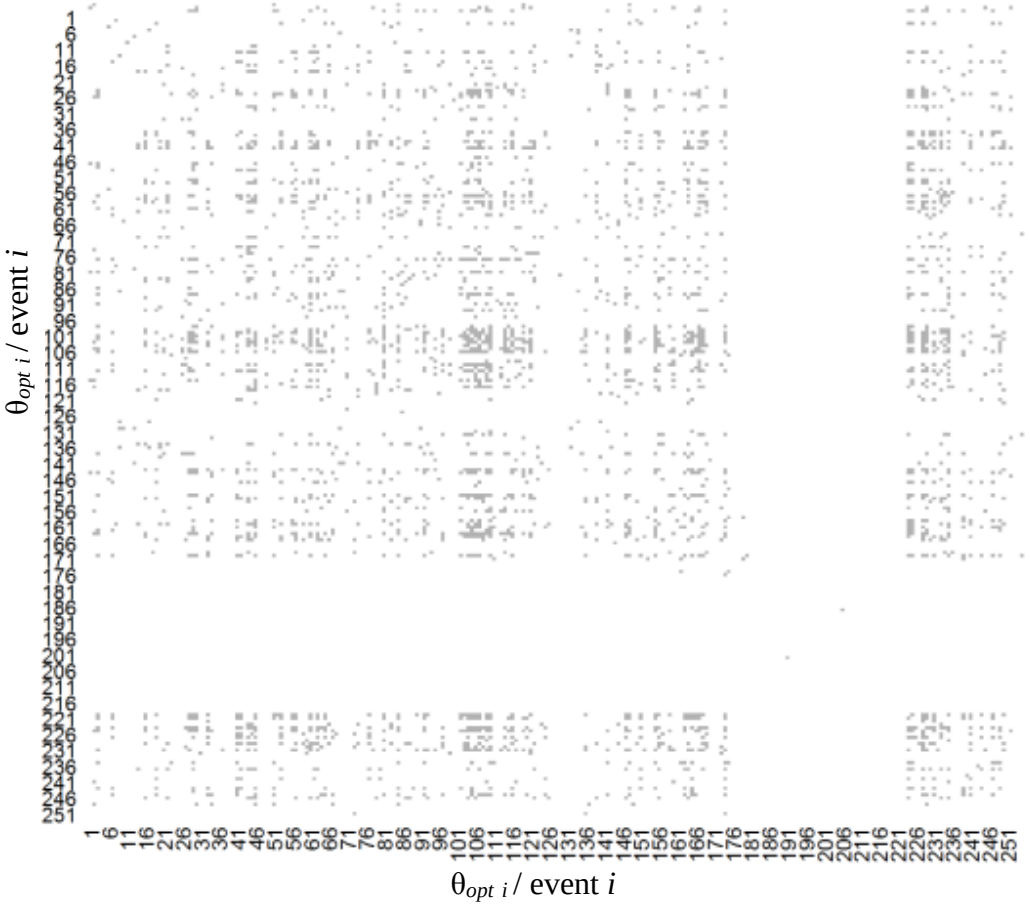
Figure 19. Correlation plot of parameters estimations (f_c, f_0, k, T_d, $K1$, q) for 255 calibration events assessed by a) SE strategy local optimal set of parameters $\theta_{opt\ i}$ (grey) with discarded events NS < 0.75 (red); b) SE Strategy global function $p(\theta/Y)$ (grey) with discarded events (red); c) ME Strategy global optimal set of parameters θ_{opt} and d) ME Strategy global function $p(\theta/Y)$.

Further analysis led in evidence that the propagation of $p(\theta/Y)$ over simulations for verification rainfall events are scarcely able to explain the observed flow rate values in the ME strategy, due to the narrowness of the obtained $p(\theta/Y)$ distribution (Figure 19d). This can be explained as in Bayesian (and frequentist) inferences, the parametric uncertainty declines asymptotically as more data is included in the calibration (Renard *et al.*, 2010). The problem resides in "which" data should we include in the inference problem (selecting as with SE and

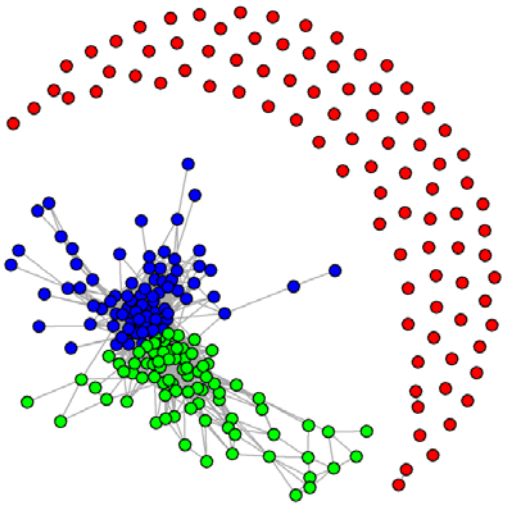
SEConditional Strategies). Correlations are an undesirable but common property when parameters are estimated. The infiltration parameters f_c, f_0 and k exhibited a global strong correlation in the ME strategy (Figure 19d). However, when local calibrations were done with SE, these parameters were identifiable for certain events and further formulations for modelling the infiltration process were less satisfactory. For this reason and with illustrative purposes of the potential of the application of the proposed methods to models with correlated parameters (as *e.g.* Guerrero *et al.*, 2013), the model structure is kept (as in *e.g.* Sun and Bertrand-Krajeswki, 2013a) without further reformulations.

The total 255 rainfall events are classified into *T*1 (blue, 32 % of calibration events), *T*2 (green, 32 % of calibration events) or discarded (red, 36 % of calibration events) by applying the Supervised Cluster algorithm (spinglass.community) over the topological graph (Figure 20b), given by the *AM* (Figure 20a). Indeed, the events labelled in red in the topological representation have their corresponding row or column in *AM* completely equal to zero (white). For example, in events from 173 to 222 (from 11/3/2008 to 9/13/2008), a more frequent non-reproducibility or disconnection of events is evidenced with a larger white stripe in the *AM* (Figure 20a), representing 53 % of the total disconnected events. These large groups of discarded events can be attributable to, *e.g.*, particular climatologic conditions or systematic measurement errors during a specific period, given the temporal proximity of the events. All disconnected events are discarded for further hydrological analyses in this Chapter 4 as explained in the methodology section.

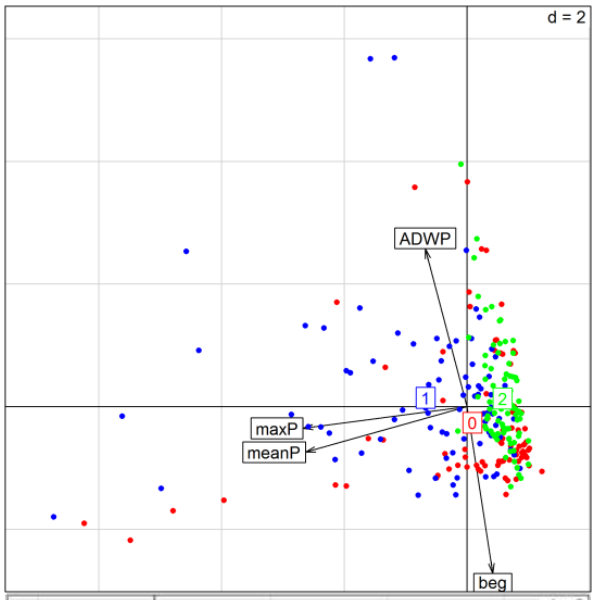
a)



b)



c)



d)

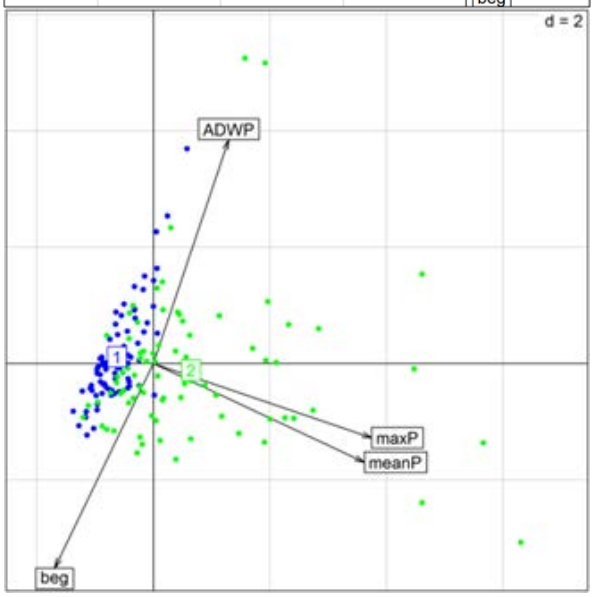


Figure 20. a) *AM* for the calibration rainfall events; b) topological graph representation of *AM*, including the clusters obtained for T1 (blue), T2 (green) and disconnected events (red); c) PCA for T1 (blue), T2 (green) and disconnected events (red) and d) PCA for T1 (blue) and T2 (green) groups (explanatory variables: max intensity, mean intensity, beginning of the event and ADWP).

With the purpose of giving additional interpretation to the topological representation of the graph and separation groups (Figure 20b), a PCA is proposed to explain the differences between: (i) the groups $T1$ (blue), $T2$ (green) and discarded events (red) (Figure 20c) and (ii) the groups $T1$ (blue) and $T2$ (green) (Figure 20d). The explanatory variables are selected from previous tests and the idea of giving the simplest possible interpretation to the Cluster results in terms of temporality and pluviometry (scaled by a *z* normalization, see Kreyszig, 1979). PCA is then undertaken by focusing on four event characteristics: ADWP, beginning date of the event, mean rainfall intensity and maximum rainfall intensity (all variables are scaled to have zeroed mean and unitary standard deviation). No clear separation of the groups is obtained in Figure 20c, implying that a linear combination of the analyzed rainfall characteristics cannot explain the differences between $T1$ (blue) or $T2$ (green) with the discarded events (red). No special pattern in the temporal sequence of events as $T1$ (blue), $T2$ (green) or discarded (red) is observed.

Although the non-reproducibility of certain rainfall events can be attributed to temporality (from 11/3/2008 to 9/13/2008), a general explanation of why rainfall events are non-reproducible by the CRR model is hardly supportable by rainfall physical characteristics, strengthen potential explanations such as the local nature of errors in rainfall measurements rather than the global nature of conceptual model uncertainties. On the other hand, Figure 20d shows a clear separation by explaining exclusively the differences between $T1$ and $T2$ hydrological groups (blue and green). The rainfall events in group $T1$ or $T2$ are mainly distinguished because of the mean or maximum rainfall intensity (Figure 20d). The physical sense behind this finding is that the global function $p(\theta/Y, T1)$ with $\theta_{opt}T1$ is more adapted to simulate high intense rainfall events (group $T1$) and the global function $p(\theta/Y, T2)$ with $\theta_{opt}T2$ is more appropriate to simulate low intensity rainfall events (group $T2$). For results in Figure 20d, a Classification Decision Tree (Breiman *et al.*, 1984) implemented in Matlab led to visualize that events are part of $T2$, except if their max intensity > 9 mm/h and their mean intensity > 1.6 mm/h, case in which they belong to $T1$, with a confidence of about 80 % ("fitctree" and "kfoldloss" functions from Matlab). This simplified classification and Figure 20 bring up additional evidence that differences between group $T1$ and $T2$ and thus the transferability of the parameters estimations are mainly due to rainfall (intensity) and not temporal (ADWP, beginning of the event) inter-event characteristics. Accordingly, the use of $p(\theta/Y, T1)$, $\theta_{opt}T1$ or $p(\theta/Y, T2)$, $\theta_{opt}T2$ (SEConditional) for simulating a verification event is defined from this simplified classification rules into high or low intensity rainfall events ($T1$ or $T2$).

The correlation matrix plot for parameter estimations (f_c, f_0, k, T_d, $K1$, q) by the SEConditional strategy by using the DREAM algorithm are presented in Figure 21.

a)

b)

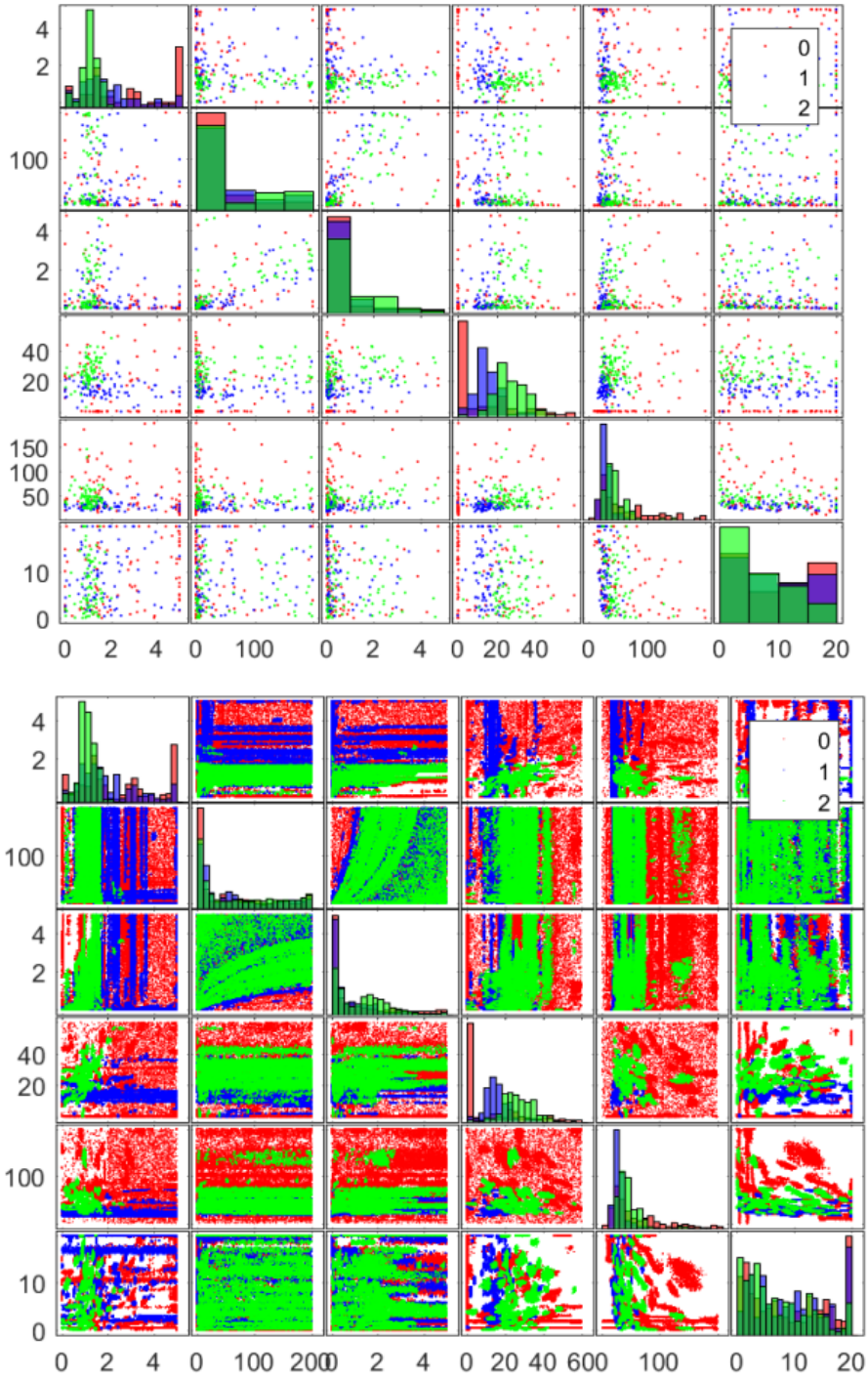


Figure 21. Correlation plot of the parameters estimations (f_c, f_0, k, T_d, $K1$, q) for 255 calibration events, assessed by SEConditional Strategy: a) local optimal sets of parameters $\theta_{opt\ i}T1$ (blue), $\theta_{opt\ j}T2$ (green) with discarded (disconnected) events (red) (with i and j =1: number of events in $T1$ and $T2$ resp.) and b) conditional functions $p(\theta/Y, T1)$ (blue), $p(\theta/Y, T2)$ (green), with discarded (disconnected) events (red).

It is worth to point out that the criterion in SEConditional for identifying non-reproducible rainfall events by the CRR model is more restrictive than the one adopted for SE (28 % and 36 % of the calibration data resp.). A given rainfall event is discarded if the set of local optimal set of parameters are not able to satisfactorily reproduce further events from the dataset (NS > 0.75), although the local estimation brings a NS > 0.75 by itself. Nevertheless, 72 % of the events discarded by SEConditional, are also discarded in SE strategy, implying again that irreproducibility and non-transferability of a certain set of local parameters seems

to be more related to potential errors in rainfall data than to deficiencies in the model structure. The idea that identifying more appropriate parameter set relies on the selection of time periods with similar hydrological characteristics (Seibet, 2003) is also strengthen by these findings. Situations in which all events are labeled as non-reproducible for SEConditional will suggest an inappropriate model structure or an extremely low repeatability of the model parameters.

The proposed classification into $T1$ and $T2$ permits to separate bimodal behaviors of $p(\theta/Y)$ by two conditional functions $p(\theta/Y, T1)$ and $p(\theta/Y, T2)$. This can be especially noted in parameters T_d and K, in which the blue and green distributions are considered as mixed in the grey marginal in SE strategy (Figure 19b and Figure 21b resp.). With further comparative purposes, a boxplot of the scaled estimations $p(\theta/Y)$ for SE, ME and $p(\theta/Y, T1)$ and $p(\theta/Y, T2)$ for SEConditional is presented in Figure 22.

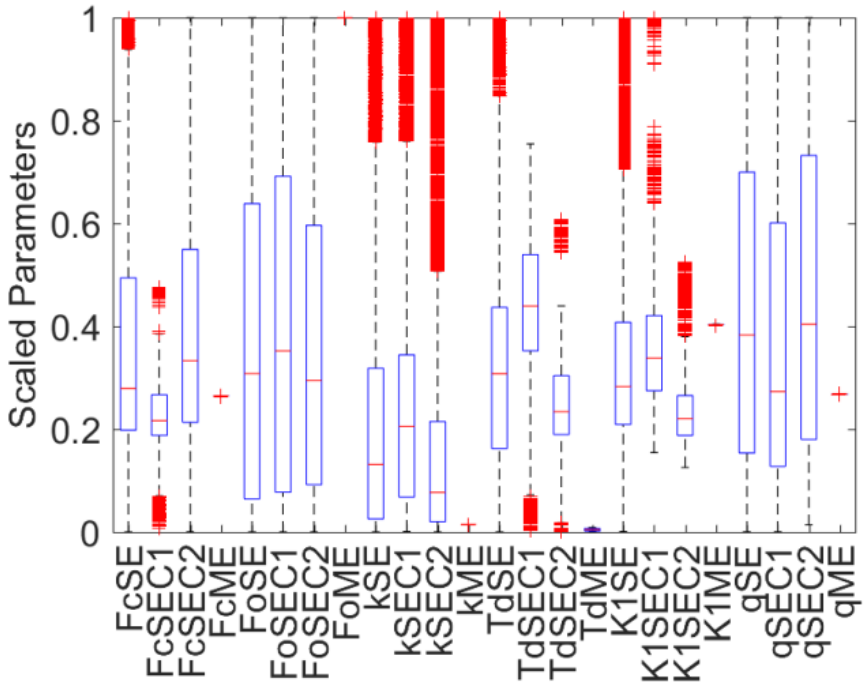


Figure 22. Scaled parameters boxplot of the estimations p(θ/Y) for SE, ME; with p(θ/Y , T1) and p(θ/Y , T2) for SEConditional.

It can be noted that the medians of $p(f_c/Y)$, $p(T_d/Y)$ and $p(K1/Y)$ are different from the medians obtained for their corresponding conditional distributions $p(f_c/Y,T1)$ and $p(f_c/Y,T2)$, $p(T_d/Y,T1)$ and $p(T_d/Y,T2)$, $p(K1/Y,T1)$ and $p(K1/Y,T2)$, due to the observed bimodalities in the marginal distributions (Figure 21). In addition, the dispersion of $p(f_c/Y)$, $p(T_d/Y)$ and $p(K1/Y)$ is higher than the dispersion obtained with the conditional distributions $p(f_c/Y,T1)$ and $p(f_c/Y,T2)$, $p(T_d/Y,T1)$ and $p(T_d/Y,T2)$, $p(K1/Y,T1)$ and $p(K1/Y,T2)$. This fact will directly influence the results to be obtained in the verification phase in terms of NS, ARIL and modified COP. These results are consistent with Fenicia *et al.* (2016), in which the variability of the response parameters (here the delay *Td* and lag time *K1*) of the catchment for each rainfall event is identified as the main degenerative factor on the results for ME calibrations. Coupling the results discussed in Figure 20d with findings presented in Figure 21b, one can establish that the rainfall intensities (high for type $T1$ or low for type $T2$) are the main explanatory factor in the observed bimodality of *Td* and *K1*.

Calibration: relation transferability-statistical depth of sets of parameters

With the purpose of further analyzing the transferability of the different sets of parameters $\theta_{opt\ i}$, the NS values for a rainfall event *j* evaluated with the optimal parameters $\theta_{opt\ i}$, obtained from a different rainfall event *i*, are shown in a NSmatrix(*i*, *j*) (Figure 23). Contrary to the *AM*, the NSmatrix is not symmetric and its values are not binary (NS < 0 are set as zero for interpretability). The diagonal NSmatrix(*i*, *i*) has the highest NS values, as the set of local parameters $\theta_{opt\ i}$ are obtained with data from the rainfall event *i*. The mean of all rows for a given column NSmatrix(:, *i*) reflects the capacity of the rainfall event *i* to be simulated by the CRR model (reproducibility of events, Figure 23b). On the other hand, the mean of all columns for a given row NSmatrix(*i*, :) reflects the transferability of a local optimal set of parameter $\theta_{opt\ i}$ (Figure 23a). The NSmatrix is shown in Figure 23 jointly with the mean of rows and columns.

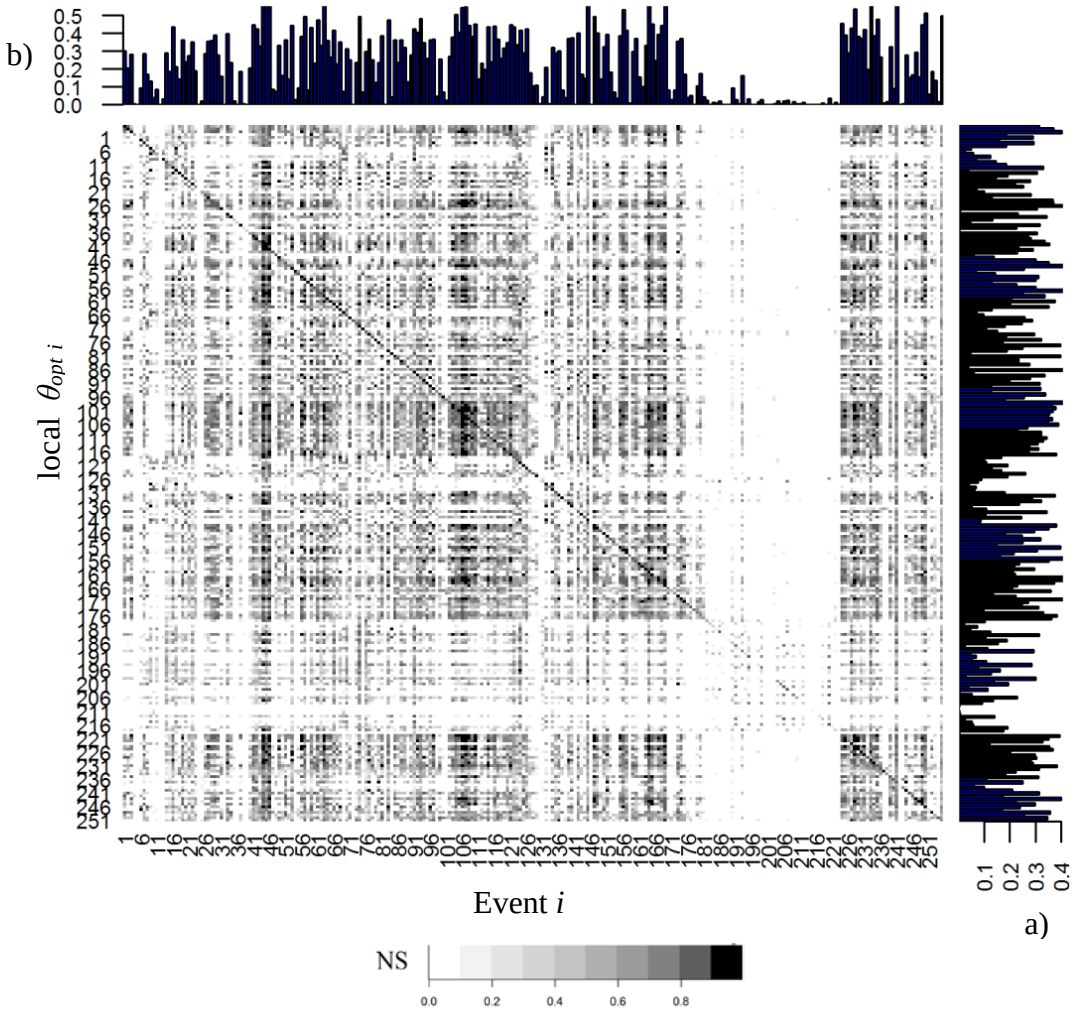


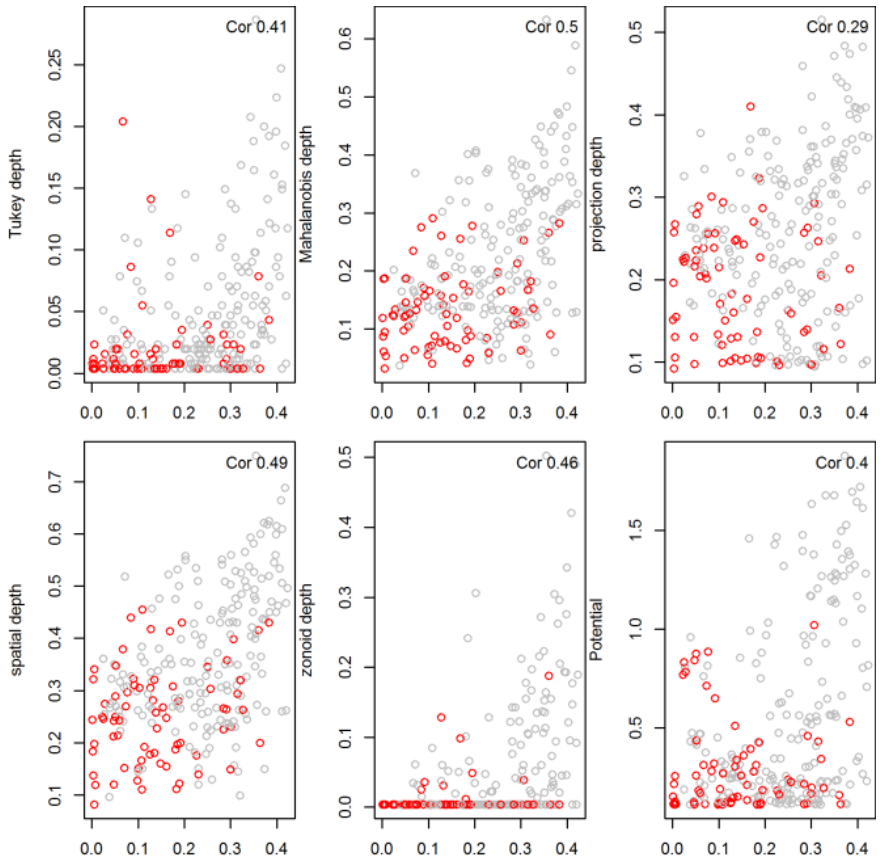
Figure 23. NSmatrix with a) the mean of NS rows (transferability of a local set $\theta_{opt\ i}$) and b) the mean of NS columns (reproducibility of the rainfall event *i*) with *i* = 1 : 255.

The results regarding the mean of the columns in the NSmatrix are naturally coincident with the identified non-reproducible rainfall events by means of the strategies SEConditional (*AM* in Figure 20a), with lower mean NS for events 173 to 222 (from 11/3/2008 to 9/13/2008), for example. On the other hand, selecting an appropriate frame (statistical depth measure) might led to verify the geometric consistency of the $p(\theta/Y)$ estimation, where the mean row in the

NSmatrix for a $\theta_{opt\ i}$ (Figure 23a) is expected to be higher as $\theta_{opt\ i}$ is deeper into the function $p(\theta/Y)$ (SE strategy); or for the case of SEConditional strategy, $\theta_{opt\ i}$ (in $T1$ or $T2$) is deeper into $p(\theta/Y,T1)$ or $p(\theta/Y,T2)$ resp. (Bardossy and Singh, 2008; Guerrero *et al.*, 2013).

Hence, the geometric consistency of the estimations in SE strategy and SEConditional strategy is verified by comparing the "deepness" of $\theta_{opt\ i}$ into $p(\theta/Y)$ (SE strategy), or $p(\theta/Y,T1)$, $p(\theta/Y,T2)$ (SEConditional strategy) against the mean of columns for NSmartix(i, :) (transferability of $\theta_{opt\ i}$) (Figure 24). The results from the statistical depths analyzed in this work are shown in Figure 24 with different definitions of statistical depth (see details in the Methodology section of this Chapter). For SE strategy, the depths of each $\theta_{opt\ i}$ are calculated by the cited depth definitions to the total of optimal set of parameters $\theta_{opt\ i}$ (grey Figure 24a), including for comparative purposes the $\theta_{opt\ i}$ from discarded events (red Figure 24a). In the case of SEConditional strategy, the depths of each $\theta_{opt\ i}T1$ (blue), $\theta_{opt\ j}T2$ (green) are calculated to the total of optimal set of parameters for each group, $\theta_{opt}\ T1$ (blue) or $\theta_{opt}\ T2$ (green), with i and j =1 : number of events in $T1$ and $T2$ resp.

a)



b)

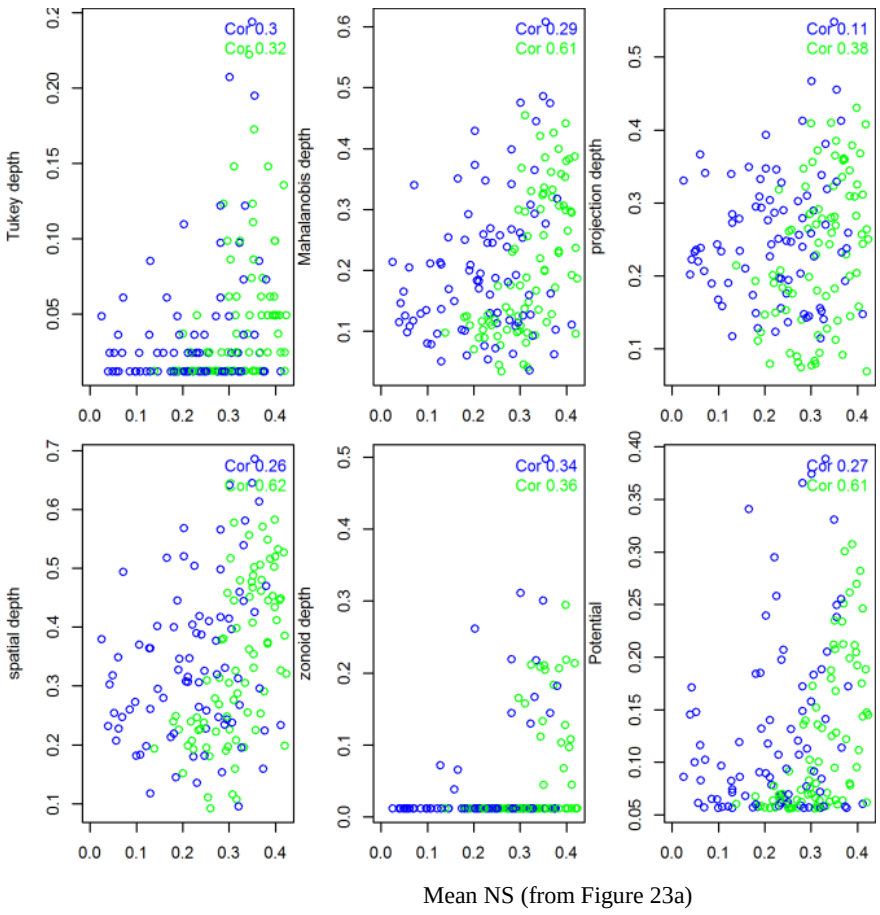


Figure 24. mean of NS rows (transferability of a set $\theta_{opt\ i}$) (Figure 23a) versus different measures of statistical depth for a) SE strategy ($\theta_{opt\ i}$ in grey) and discarded events (red) (Figure 19a); b) SEConditional $\theta_{opt\ i}T1$ (blue), $\theta_{opt\ j}T2$ (green) with i and j =1 : number of events in $T1$ and $T2$ resp. (Figure 21a).

Geometric consistency of the $p(\theta/Y)$ estimation in the SE strategy can be observed as the $\theta_{opt\ i}$ from discarded events (NS <0.75 in red) appeared to be swallower (less deep) into the $p(\theta/Y)$ function, especially when establishing as reference frames the Mahalanobis and Spatial depths (Figure 24a). For these reference depths, the $\theta_{opt\ i}$ values with higher mean NS (transferability from Figure 23a) also tend to be located in a geometrical deeper position among all the other local estimations of $\theta_{opt\ i\ =\ 1:255}$ (correlation of 0.5 in Figure 24a). For the case of SEConditional strategy, discarded events are not included for the analysis as they cannot be linked to group $T1$ or $T2$ (blue and green resp., Figure 24b). The analyzed depth measures show as well strong enough correlations between the geometric depth of a given $\theta_{opt\ i}T1$ (blue) or $\theta_{opt\ j}T2$ (green) set inside $p(\theta/Y,T1)$ or $p(\theta/Y,T2)$ and its transferability (Figure 24b).

One should bear in mind that the evaluated correlations can be masked by the strong variability of the parameter sets. The analyses become more challenging, in terms of identification of correlations and computational calculation of depths, if the complete set of estimations for $p(\theta/Y)$ and $p(\theta/Y,T1)$ or $p(\theta/Y,T2)$ are included. For this reason, the analysis in Figure 24 is undertaken by local estimations $\theta_{opt\ i}$ (grey) and with $\theta_{opt\ i}T1$ (blue) or $\theta_{opt\ j}T2$ (green) instead of the complete functions $p(\theta/Y)$ and $p(\theta/Y,T1)$ or $p(\theta/Y,T2)$. From this assumption and the correlation results shown by calculating the Mahalanobis depths in Figure 24, the "deepest set" of $p(\theta/Y)$ and $p(\theta/Y,T1)$ or $p(\theta/Y,T2)$ is defined as the multivariate mean of the parameter sets, following the properties of this distance (Pokotylo *et al.*, 2016). Therefore, for the verification phase, the global optimal parameters: i) θ_{opt} are calculated as the multivariate mean of all $\theta_{opt\ i}$ values for SE strategy (for non-discarded i events); ii) $\theta_{opt}T1$

and $\theta_{opt}T2$ are calculated as the multivariate mean of all $\theta_{opt\ i}$ or $\theta_{opt\ j}$ values in each type $T1$ or $T2$ for SEConditional strategy.

Verification: accuracy, precision and reliability of the simulations

The verification is undertaken with the 110 remaining rainfall events for each of the three approaches (SE, ME and SECconditional). Each verification rainfall event is classified as type $T1$ or $T2$ depending on the following rules: is $T2$, except if their max intensity > 9 mm/h and their mean intensity > 1.6 mm/h, case in which is $T1$. The values of parametric ARIL and NS are calculated from the flow rate simulations obtained from $p(\theta/Y)$ and θ_{opt} for SE or ME, and from $p(\theta/Y, T1)$ or $p(\theta/Y, T2)$ and $\theta_{opt}T1$ or $\theta_{opt}T2$ for SEConditional (Figure 25).

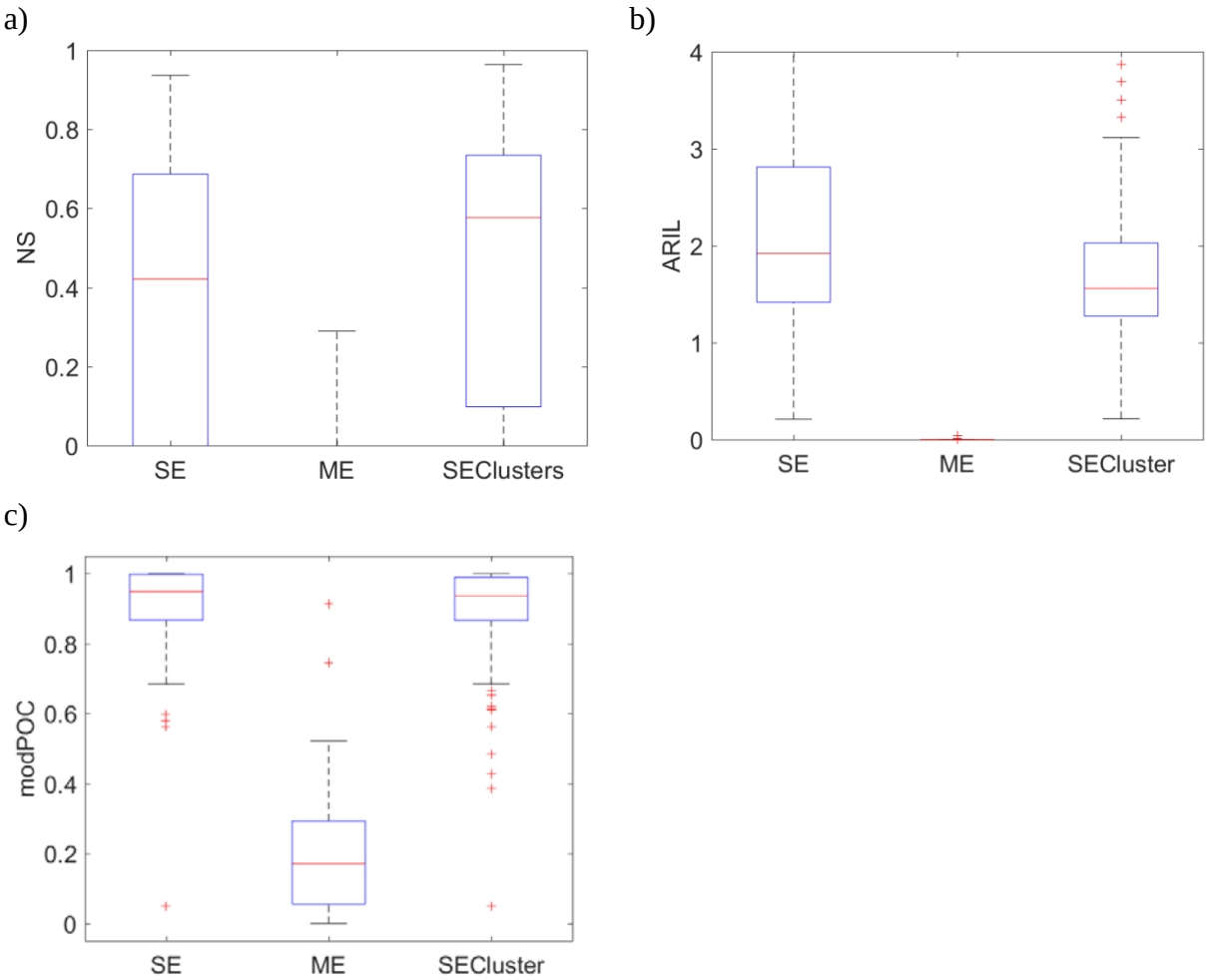


Figure 25. a) NS b) ARIL and c) modifiedPOC criterion, for the 110 verification rainfall events, including the three studied parameter estimation approaches (SE, ME, SEConditional).

The accuracy of the flow rate in verification for SEConditional shows an improvement according to the Nash-Sutcliffe (NS) criterion, from 0.4 to 0.6 for 50 % of the rainfall events, compared to the SE strategy (Figure 25a). A reduction of parametric uncertainty bounds in verification given by the ARIL values from 2 to 1.6, for 50 % of the verification rainfall events can be as well noticed for both strategies (Figure 25b). Furthermore, a Wilcoxon test (*e.g.* Hollander and Wolfe, 1999) shows that the mean of the NS and ARIL values in the SEConditional strategy are significantly higher and lower, respectively, than for the SE strategy (p-values <0.05). Parametric uncertainty bounds obtained by the SE and SEConditional approaches are equally reliable, in the sense that both are able to explain more than 95 % of the verification data (modifiedPOC Figure 25c). However, the SE strategy is

more alike to have the highest scores in the modifiedPOC criteria, as parametric ARIL values are commonly higher than 2 (mean width of uncertainty bound of about 200 % of the flow rate values), which can be considered as an overestimation of the parametric uncertainty (Figure 25b). This potential overestimation by the SE strategy, not only for the parametric (Figure 25b) but also for the total uncertainty (Figure 26a), might be indeed caused by simulating the flow rate from sets of optimal parameters obtained with events that are not necessarily linked from a hydrological point of view. These results stress the importance of carefully selecting the data to be used for parameter estimations and further hydrological simulation, considering that a selection strategy such as SEConditional shows superior results than SE in terms of accuracy (NS), precision (total and parametric ARIL) and reliability (total and parametric modifiedPOC) for verification events.

The parametric ARIL values show that the uncertainty bounds in verification obtained by using the ME approach are much thinner (values close to zero), than those obtained with the other strategies (SE and SEConditional), besides a deficient accuracy given by NS in verification (Figure 25a, b). For this case, the simulations in verification can be considered as unreliable, as the modifiedPOC from parametric and total uncertainty reported that only about 0.2 and 0.6, resp., of the measured values overlap the uncertainty bounds of the simulations (Figure 25c and Figure 26b). The almost inexistent parameter uncertainties reported by the ME strategy can be attributed to the asymptotically declining behavior of uncertainties when more data are included in Bayesian inferences (109860 flow rate continuous values of 255 rainfall events). However, the parametric uncertainty given by ME is not considered to be realistic, as the total output uncertainty could only be then explained by hypothetical deficiencies in the model structure. The model structure demonstrated its robustness from the transferability analysis of the local parameters (*AM* Figure 20a) and therefore ME is not recommendable.

In addition, parameters estimations from ME might contain important errors from rainfall, as events in which the model is not even able to reproduce the flow rate with a local calibration cannot be directly identified and discarded. Without abandoning the simulation context (Beven and Smith, 2014), further methods for dealing with the rainfall error detection under a ME parameter estimation approach might be obtained just on the expense of complicated statistical implementations whose departure assumptions are hardly verifiable (*e.g.* Pedersen *et al.*, 2016).

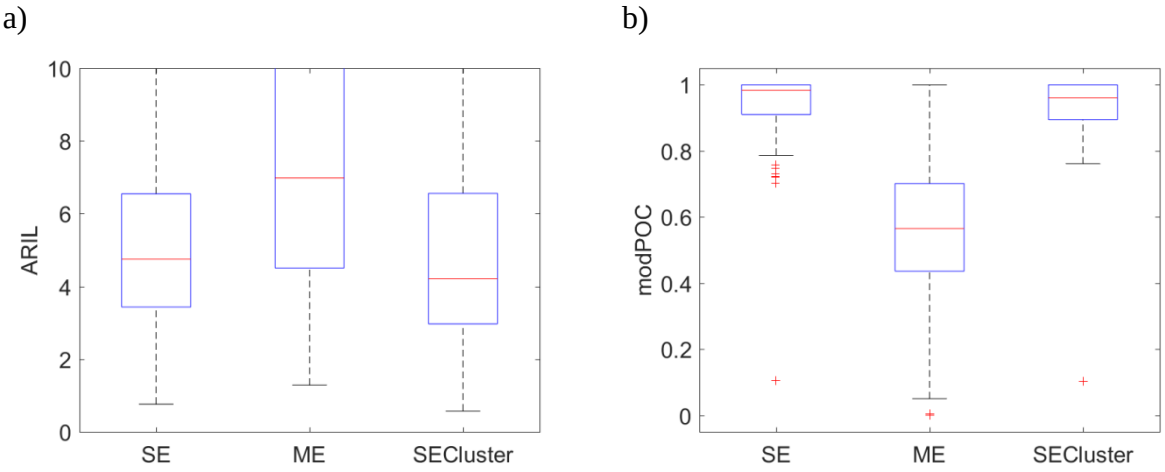


Figure 26. a) Total ARIL and b) Total modifiedPOC criterion, for the 110 verification rainfall events, including the three studied parameter estimation approaches (SE, ME, SEConditional).

The SEConditional strategy led to superior results than SE in terms of accuracy (NS), precision (total and parametric ARIL) and reliability (total and parametric modifiedPOC) for verification events (Figure 25 and Figure 26). However, even if SEConditional is used, the parametric and the total ARIL remain undesirability high (about 1.6 and 4.5, Figure 25b and Figure 26a, resp.) and the NS values undesirably low (about 0.6, Figure 25a).

Undoubtedly, the model structure can be improved for the studied catchment and its calibration could also be fed by richer rainfall information (*e.g.* radar records), or more detailed spatial information systems (*e.g.* GIS, Digital Elevation Models), for example. On the other hand, one should bear in mind that the results for verification in this Chapter 4 are obtained from rainfall records that were unknown for the calibration phase, and therefore it is reasonable to think that there will be verification events with important rainfall errors, which at the end will significantly decrease the quality of the verification simulations (Thyer *et al.*, 2009). However, the potential rainfall error in a verification event will equally affect the simulations for all the adopted parameter estimations, offering also an equitable comparative frame among strategies SE, ME or SEConditional. For further generalization of verification results, one can hypothesize that the same proportion of discarded rainfall events (38 %) can be discarded as well from verification, taking the 62 % best results as a sure bet of "rain error free" estimations. This aspect can be addressed by further investigations.

4.4 CONCLUSIONS

The global nature of model structure uncertainty and the inter-event parametric variability is addressed for a CRR model based on the idea of dividing the parameters marginal probability function, obtained by event-by-event calibrations, into conditional probability functions, obtained by grouping the parameters from the event-by-event calibrations. For this purpose, an adjacency matrix that reflects how local parameter estimations are interconnected to the other calibration rainfall events is proposed with a transferability perspective (analogue to a leave-all-out scheme). The adjacency matrix is represented as a graph of connected rainfall events. The graph is analyzed by clustering techniques to determine the conditional probability functions. Two different hydrological conditions, given by the magnitude of the rainfall intensities (high or low), could be linked to a bimodal behaviour of the parameters marginal distribution. Furthermore, the proposed strategy allows identifying rainfall events in which the rainfall error is likely to be high enough to be considered as unreproducible events, at least by the selected CRR model. This approach is applied to 255 rainfall events.

The results stress the importance of carefully selecting the data to be used for parameter estimations and further hydrological simulation, considering that the selective proposed parameter estimation strategy significantly improves the results of traditional parameter estimations from event-by-event and multi-event calibrations. The improvements achieved by expressing the event-by-event global parametric uncertainty into conditional probability functions are shown in terms of accuracy (Nash-Sutcliffe criterion), precision (total and parametric Average Relative Interval Length) and reliability (total and parametric Percentage of Coverage) for 110 verification events. The drawbacks of a multi-event calibration approach are exposed when facing a large enough calibration data-set (255 events). One single rainfall-runoff model structure allows representing two groups of different hydrological conditions for an urban catchment by the proposed approach.

CHAPTER 5. METHODOLOGY FOR IDENTIFYING THE TEMPORAL DISTRIBUTION OF ERRORS IN RAINFALL TIME SERIES

Extended version of:

Sandoval S., Bertrand-Krajewski J.-L. (2015). Identification of errors in high temporal resolution rainfall time series by model based approaches. *Proceedings of the 10th UDM - International Conference on Urban Drainage Modelling*, Mont Sainte Anne, Quebec, Canada, 20-23 September, Oral Presentations II, 183-186.

5.1 INTRODUCTION

Experience indicates that, for some rainfall events with large spatial heterogeneity and/or significant movements of rain cells over the catchment, a single rain gauge cannot deliver representative rainfall intensities applicable to the entire catchment (mean areal rainfall intensity) (Leonhardt *et al.*, 2014). Areal rainfall estimations by indirect methods are the object of intensive research *e.g.* by rainfall radar (Einfalt *et al.*, 2004), satellite data (Kidd and Levizzani, 2011) or microwave links (Messer *et al.*, 2006). Nevertheless, the accuracy of these approaches is still limited due to their nature of indirect measurements (Leonhardt *et al*., 2014). Other sources of error in rainfall predictions can be also attributed to improper calibration of rain gauges and local instrumental uncertainties (*e.g.* Stransky *et al.*, 2006). Generally, systematic and random components of rainfall errors (areal estimations and measurement errors) are not known in advance (in practical situations) and their structure can be complex and variable (Kavetski *et al.*, 2006a; Schellart *et al.*, 2012).

Model-based approaches have emerged as a promising alternative for assessing rainfall areal estimations and uncertainties in rainfall data. Kavetski *et al.* (2006a) propose a multiplicative rainfall error identification model under the BATEA framework, which is widely used in the literature (*e.g.* Kavetski *et al.*, 2006b; Vrugt *et al.*, 2008; Renard *et al.*, 2010; McMillan *et al.*, 2011; Sun and Bertrand-Krajewski, 2013a; Baroni and Tarantola, 2014; Leonhardt *et al*., 2014) and will be adopted for our study. Renard *et al.* (2010) applies this methodology for identifying rainfall errors by means of a conceptual hydrological model. The original rainfall records are corrupted by means of random multipliers (errors and uncertainties around 20 %). The capability of identifying the errors which had been introduced is evaluated with suitable results. In Renard *et al.* (2010), as for *e.g.* Kavetski *et al.*, 2006b; Vrugt *et al.*, 2008; Renard *et al.*, 2010; Baroni and Tarantola, 2014; Kretzschmar et al., 2016; Fuentes-Andino et al., 2017, the number and distribution of multipliers over the rainfall is usually one per event or day. However, the temporal scale of an urban catchment corresponds to faster responses. Thus, one multiplier for a whole rainfall event has been reported as insufficient in the context of urban drainage (Sun and Bertrand-Krajewski, 2013a). In addition, Vrugt *et al.* (2008) remarked that a single multiplier for each measurement leads to an unsolvable over-parametrized problem. Consequently, Sun and Bertrand-Krajewski (2013a) proposed to group data into different numbers of rainfall multipliers for a specific event. This approach is also implemented with the Bayesian Method, including the use of the DREAM algorithm (Vrugt *et al.*, 2016). Nevertheless, estimations are based on average rainfall uncertainties of about 5 %, considering a uniform temporal distribution of the multipliers. In addition, parameters are

included within the calibration process, leading to possible interactions between parameters and rainfall multipliers estimation.

Leonhardt *et al*. (2014) proposed a comparative study between the multiplicative rainfall error model (applied at higher temporal resolution, as Sun and Bertrand-Krajewski, 2013a) and the Reverse Modelling approach. For this case, the hydrological parameters are calibrated independently from the rainfall multipliers (by the Bayesian Method and the DREAM algorithm, see Vrugt *et al.*, 2016). In addition, the Reverse Modelling approach is able to fill gaps in rainfall time series. However, the objective of Leonhardt *et al*. (2014) was not to evaluate the rainfall error prediction capacity by introducing known errors in rainfall.

The proposed methodology seeks to evaluate the potential of four rainfall correction models for identifying and correcting errors in rainfall data. The response (simulated flow-rates) of a well-performing pre-calibrated rainfall-runoff model to corrupted rainfall data is used. Considering that the objective in this Chapter 5 is to focus on rainfall errors and not parameters estimations, 30 events in which the CRR model shows the most satisfactory performances are used for this purpose with local calibrated parameters (NS > 0.95) (from single-event calibrations in Chapter 4, see NS in Eq 12). The rainfall correction models are formulated by considering the inconveniences of the traditional multiplier approach, namely dealing with rainfall equal to zero (problem identified and discussed by several authors, *e.g.* Renard *et al.*, 2010), by mixing the multiplicative error model and the reverse modelling approaches. Models also include a variable/constant time-window methodology that permits a uniform or non-uniform temporal distribution of the multipliers. Previous works are extended and coupled in the sense that: (i) higher multiplier values are introduced as errors with variable temporal distributions as areal rainfall errors can be higher and more complex, (ii) more general conclusions are drawn by the study of multiple rainfall events with real rainfall and runoff data, (iii) calibration independency between model physical parameters and rainfall multipliers calculation, iv) new error and more flexible correction models that contain additional information to generate precipitation with a Reverse Model (as suggested by Del Giudice *et al.*, 2016) are proposed and compared and v) temporal correlation effect of the rainfall multipliers is analyzed. Recommendations about the best model and the best number of time windows are given based on criteria such as NS, RMSE and mass conservation. The methodology is applied to the Chassieu catchment with 30 rainfall events measured from 2007 and 2008 (see Chapter 1). Rainfall error models evaluated in this Chapter 5 can be listed as follows: CTWrev (Reverse-constant-time-windows), VTWrev (Reverse-variable-time-windows), CTW (constant-time-windows), and VTW (variable-time-windows).

5.2 METHODOLOGY

A general description of hydrological models is given in Eq 16 (notation consistent with Kavetski *et al.*, 2006a, 2006b):

$$Y = h(\theta, X) \qquad \text{Eq 16}$$

where Y constitutes the time series response (for this case flow-rate at the outlet in L/s) to another input vector X (rainfall intensity measured at a given rain gauge station in mm/h). The mathematical simplification aimed at describing the output Y from any input X is called the function $h: X \rightarrow Y$, depending on the set of parameters θ. Given the model h and observations X_{obs} and Y_{obs}, a pre-calibration phase is undertaken for each rainfall event separately, in which

the optimum parameters θ_{opt} of the model *h* are estimated in the sense of making Y_{obs} as close as possible to Y_{sim}.

However, this pre-calibration phase for each rainfall event becomes only relevant to support the hypothesis of having a suitable mathematical description of the physical system. A good rainfall correction cannot be estimated if the original *h* model is not well calibrated, as errors from the pre-calibration phase might be included in the correction. The impact of this error is thought to be diminished by calibrating each event separately. The "true" parameters θ_{opt} of the model, obtained for each rainfall event in the pre-calibration phase, are retained for the evaluation of the rainfall correction models.

Therefore, the hypotheses for the rainfall error correction model are: (i) the function *h* is a good enough description of the system, (ii) θ_{opt} obtained from the pre-calibration phase is well known and (iii) standard uncertainties due to the measurement device (pluviograph) in X_{obs} and in parameters θ_{opt} are negligible in comparison to introduced error components.

Introduction of errors in rainfall

Given these previous statements, whenever *h* is evaluated as $Y_{corrupted}=h(\theta_{opt}, X_{corrupted})$ for corrupted rainfall $X_{corrupted}$, differences between $Y_{corrupted}$ and Y_{obs} are mainly caused by $X_{corrupted}$. $X_{corrupted}$ is going to be produced based on the original rainfall X_{obs}, which is assumed to be the true rainfall, considering that $Y_{sim}= h(\theta_{opt}, X_{obs}) \approx Y_{obs}$, from the pre-calibration phase. The rainfall $X_{corrupted}$ is obtained from the original rainfall X_{obs} by introducing a vector K_{intro} as $X_{corrupted} = X_{obs} / K_{intro}$ (Figure 27). With the purpose of exploring the influence of systematic, higher and more complex error structures in the rainfall, the K_{intro} vectors are established as non-uniformly distributed time windows by the Monte Carlo method. Comparable error introduction schemes are used in previous studies (Renard *et al.*, 2010).

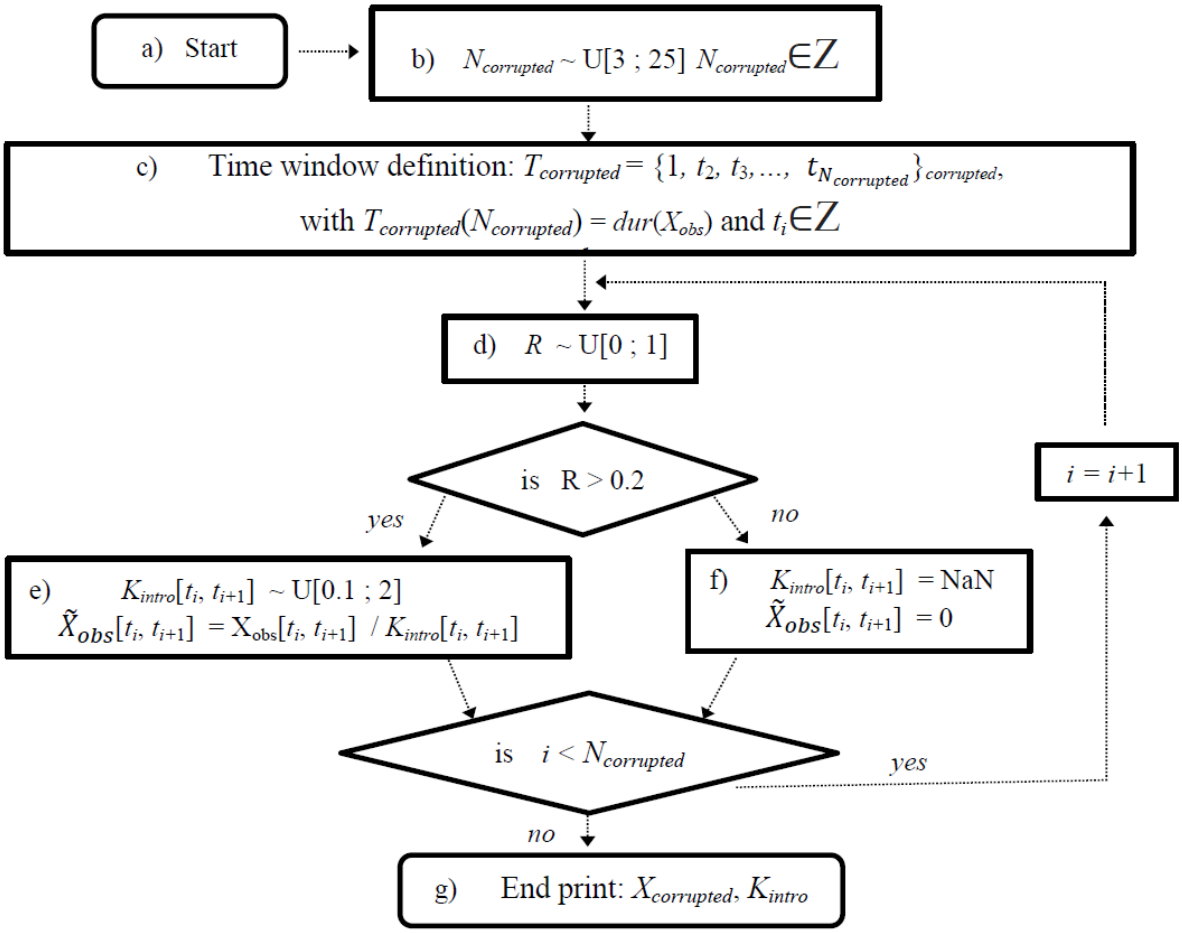


Figure 27. Scheme of error introduction into true rainfall X_{obs}

The first step of the scheme to create $X_{corrupted}$ is to define a number of $N_{corrupted}$ corruption time windows. For our case study, $N_{corrupted}$ is a randomly chosen integer from 3 to 25, with the purpose of exploring different type of error temporal distributions (Figure 27b). For more than about 25 windows, the Bayesian inference in the error correction models showed to be an ill-posed problem. For comparative purposes, the maximum number of $N_{corrupted}$ windows is therefore set to 25 windows for the introduction of errors. The definition of an ordered set $T_{corrupted}$ is based on the number $N_{corrupted}$, as a set of points t_i, $T_{corrupted} = \{1, t_2, t_3, \ldots, t_{N_{corrupted}}\}_{corrupted}$, with $N_{corrupted} = dur(X_{obs})$, indicating the boundaries of a given time window as $[t_i, t_{i+1}]$ (Figure 27c). Each pair of boundaries $[t_i, t_{i+1}]$ is evaluated to build K_{intro} and evaluate the $X_{corrupted}$ vector (Figure 27d, e, f, g). The algorithm assigns a total of 80 % of the times a value to the corresponding window $[t_i, t_{i+1}]$ (Figure 27e). This $K_{intro}[t_i, t_{i+1}]$ value is obtained from a uniform distribution bounded from 0.1 to 2 (Figure 27e). A limit lower than 0.1 is found to bring up unrealistic $X_{corrupted}$ (*e.g.* 500 mm/h). The upper limit of 2 is included for making results comparable with the rainfall corrections. In addition, for the purpose of simulating scenarios in which no rainfall is measured into a certain window of the rainfall event, the $X_{corrupted}[t_i, t_{i+1}]$ terms are forced to zero for 20 % of the times (Figure 27f). In those cases, the $K_{intro}[t_i, t_{i+1}]$ is reported as NaN for comparative purposes among the rainfall correction models and numerical convergence (as in theory $K_{intro}[t_i, t_{i+1}]$ should be equal to infinity for having a $X_{corrupted}[t_i, t_{i+1}] = 0$). The K_{intro} vector is going to be the "known" error structure of the corrupted rainfall $X_{corrupted}$. These two vectors are used to evaluate the performance of different rainfall correction models.

Rainfall correction models

The rainfall correction models can be defined as a modification of the function *h*, in which a $K_{correction}$ vector is introduced as an argument of the function *h* in the form shown in Eq 17:

$$Y_{corrected} = h(\theta_{opt}\ ,\ K_{correction} \cdot\ X_{corrupted}) \qquad \text{Eq 17}$$

$K_{correction}$ is going to be a multiplicative vector which corrects $X_{corrupted}$ aimed at giving the closest approximation of $Y_{corrected}$ to the vector Y_{obs}. It is called "Error correction model". The input θ_{opt} is obtained from the pre-calibration phase for each rainfall event.

Estimating $K_{correction}$ multipliers may vary from simple calibration methods (Vaes *et al.*, 2005) to the application of more sophisticated approaches like the Bayesian Method (Sun and Bertrand-Krajewski, 2013a; Vrugt *et al.* 2016). The DREAM algorithm has shown to be an appropriate parameter optimization method for non-linear and high dimensional hydrological models (Vrugt *et al.* 2016). In addition, the Bayesian method can lead to estimate the posterior probability distribution of each parameter, thereby extending its potential to assess uncertainties in hydrological input data (Kavetski *et al.*, 2006a, 2006b). Therefore, the estimation of $K_{correction}$ is undertaken by the Bayesian method implemented with the DREAM algorithm (30000 simulations for cases with the highest computational demand).

The likelihood function within the Bayesian framework shows the degree of belief that a given vector $K_{correction}$ has for making $Y_{corrected} = Y_{obs}$. The inferred parameters $K_{correction}$ are sampled from a uniform prior distribution $p(K_{correction})$. Henceforth, the posterior distribution of $K_{correction}$ and its uncertainties can be computed by means of Eq 18, accepting certain hypotheses about the distribution of the residuals (see details in Leonhardt *et al.*, 2014).

$$p\left(K_{correction}/y_{obs}\ , x_{corrupted}\ , \theta_{opt}\right) = \prod_{j=1}^{dur(Y)} \frac{1}{\sqrt{2\pi\sigma_{yj}^2}} \exp\left(- \frac{\left(y_{obs_j} - h(\theta_{opt}\ ,\ K_{correction} \cdot x_{corrupted})\right)^2}{2\sigma_{yj}^2} \right) . p(K_{correction}) \qquad \text{Eq 18}$$

where $dur(Y_{obs})$ is the total length of the hydrograph (index *j*) and σ_{yj} is the standard deviation of the flow-rate measurements. This uncertainty can be obtained from knowledge about the measurement principle or experience. A single σ_y can be included as well as another parameter into the calibration process, assuming homoscedasticity of the residuals. For multiple residuals (heteroscedasticity) the approach of including σ_{yj} as multiple parameters into the likelihood function is not commonly addressed, as the number of parameters will be increased $dur(Y_{obs})$ times, leading to an ill-posed problem (Renard *et al.*, 2010). For the study case σ_{yj} is included as known values from flow-rate uncertainties (see details in raw data description, hydrological model and pre-calibration phase section). For instance, the hypothesis of normality of residuals for the rainfall correction model implicit in Eq 18 is retained as for previous studies (*e.g.* Sun and Bertrand-Krajewski, 2013a). Likelihood functions with flexibility in this hypothesis are still object of intensive research (*e.g.* Schoups and Vrugt 2010; Evin *et al.*, 2014).

It should be noted that this Bayesian function (Eq 18) does not include information about rainfall uncertainties, as the idea is to replicate a scenario in which knowledge about rainfall is limited and uncertainties due to the measurement principle are negligible in comparison to systematic error components with complex and well defined temporal structures. For this reason: (i) a rainfall term is not included in the likelihood function (*e.g.* Sun and Bertrand-

Krajewski, 2013a) and (ii) the prior distributions $p(K_{correction})$ are assumed to follow a uniform probability distribution (*e.g.* Sun and Bertrand-Krajewski, 2013a), bounded from 0 to 2. The bounds of the distribution are given due to the hypothesis and limitations in the introduction of errors into the rainfall (see introduction of errors in rainfall section). A prior uniform distribution assumption of $p(K_{correction})$ might be recommendable just for situations in which *a priori* knowledge about rainfall errors is extremely vague (adapted from Kavetski *et al.*, 2006a and Renard *et al.*, 2010).

Equally / non-equally sized time windows

The distribution of the N correction factors to be calculated by the Bayesian method along the $K_{correction}$ vector is proposed to be done by constant or variable length time windows $[t_i, t_{i+1}]$, defined from an ordered set $T= \{1, t_1, t_2, t_3,\ldots,t_N\}$, with $t_N = dur(X_{obs})$ (analogous to $T_{corrupted}$ in the introduction of errors in rainfall, which is unknown for the rainfall correction model). Two ways of distributing the correction time windows along the $X_{corrupted}$ vector are proposed: equally or non-equally sized (to be called T_c and T_v respectively).

Equally sized: the corrupted rainfall intensity time series $X_{corrupted}$ can be divided into equally spaced time windows $X_{corrupted}[t_i, t_{i+1}]$, from an ordered set $T_c = \{1, t_1, t_2, t_3,\ldots,t_N\}_c$ with $t_{Nc} = dur(X_{obs})$, that satisfies the condition $dur([t_i, t_{i+1}]_c)\cdot N = dur\ (X_{obs})$, for any window i from $1{:}N$. The rainfall in each $[t_i, t_{i+1}]_c$ interval is corrected by Eq 18 with the associated correction factor $K_{correction}\ [t_i, t_{i+1}]_c$. This approach was already proposed in the literature *e.g.* Sun and Bertrand-Krajewski, (2013a).

Non-equally sized: for this case, the set T_v is calculated by a more elaborated procedure, using the flow-rate residuals signal Q_{res} (difference between measured and simulated flow-rates) defined in Eq 19.

$$Q_{res} = Y_{obs} - h(\theta_{opt}, X_{corrupted}) \qquad \text{Eq 19}$$

An auxiliary set T_v^* is defined as an ordered set of time values $T_v^* = \{1, t_1, t_2, t_3,\ldots,t_N\}_v^*$, in which *a priori* $Q_{res}[T_v^*] = 0$ and $t_{Nv}^* = dur(Y_{obs})$ (point with change of sign in the residuals, including borders). The number of time windows in T_v^* is N^*, with i from $1{:}N^*$. Nevertheless, tests showed that N^* is commonly higher than the desired number of parameters N to be used as an input in the rainfall correction models (due to uncertainties and variabilities in the flow-rate). With the purpose of establishing a T_v^* group that has a number of time windows N^* equal to or close to N (number of corrections pre-defined to be used in the rainfall correction model), the signal Q_{res} is filtered by a median filter MF as $Q_{res}^{MF}=MF(Q_{res}, \alpha)$, using the parameter α (see details Arce, 2005). The parameter α is defined as the value which permits $N^* \approx N$, for $Q_{res}^{MF}[T_v^*] = 0$, with $t_{Nv}^* = dur(Y_{obs})$ (points with change of sign, including borders). The parameter α is obtained by a simple search process.

The T_v^* set can be expected to show a pattern consistent with the time windows $T_{corrupted}$ (supposed to be unknown in practice), used to create the rainfall $X_{corrupted}$ with K_{intro} (see error introduction section). However, both sets are not still comparable given that T_v^* is defined over the length of the flow-rate $dur(Y_{obs})$ and $T_{corrupted}$ should be defined over the length of the rainfall $dur(X_{obs})$. Consequently, the length and location of each time window in T_v^* defined from $Q_{res}^{MF}[T_v^*] = 0$ must be projected into the rainfall series in order to be applied with the

rainfall correction model. Therefore, the ordered set of N (number of correction elements) $T_v=\{1,\ t_1,\ t_2,\ t_3,\ldots,t_N\}_v$, with $t_{Nv} = dur(X_{obs})$, is defined by scaling T_v^*, based on the total duration of rainfall intensity $dur(X_{obs})$ and runoff $dur(Y_{obs})$ time series. $K_{correction}$ is then established by windows $K_{correction}[t_i,\ t_{i+1}]_v$, from the proposed T_v set (Eq 17). It should be noted that whenever the argument of the function $X_{corrupted}$ would have been X_{obs} in Eq 19, the $mean[Q_{res}] \approx 0$ and the number of N^* elements of T_v^* would have tended to the length $dur(Y_{obs})$. This will occur given the fact that Q_{res}will get closer to the white noise behavior, with a variance equal to σ_{yj}, with j=1:$dur(Y_{obs})$.

Reverse modelling

Eq 18 is unable to correct the rainfall whenever $X_{corrupted} = 0$. This situation is frequent within rainfall events (especially for short temporal resolution, typical in the urban drainage context *e.g.* Leonhardt *et al.*, 2014), which is represented by introducing 20 % of the time windows as zero in the error introduction scheme (Figure 27f). The reverse problem consists in obtaining an X_{rev} rainfall time series that would have produced the measured runoff by rearranging the equations (whenever it is possible) or by the use of iterative methods (Leonhardt, 2014). Therefore, an alternative to Eq 18 for $X_{corrupted} = 0$ can be proposed based on X_{rev} time series (Eq 20 and Eq 21).

$$X_{corrupted}^{rev} = \begin{cases} X_{corrupted} & if \quad X_{corrupted} > 0 \\ X_{rev} & if \quad X_{corrupted} = 0 \end{cases} \qquad \text{Eq 20}$$

$$Y_{corrected} = h(\theta_{opt}, K_{correction} \cdot X_{corrupted}^{rev}) \qquad \text{Eq 21}$$

Reverse modelling can create a new rainfall time series without using rainfall measurements. This can be appropriate for cases in which no rainfall records are available (*e.g.* Leonhardt, *et al.*, 2014). However, as single reservoir lumped models work as low pass filter smoothing the output, their reverse formulation usually amplifies the noisy behavior of the flow-rate signal over the rainfall estimation (Leonhardt, 2014). Consequently, reverse modelling can be quite sensitive to uncertainties in flow-rate data, bringing up highly uncertain solutions. In addition, non-meaningful physical results can be obtained (*e.g.* negative intensities). Therefore, X_{rev} is proposed to be used exclusively when $X_{corrupted} = 0$ instead of replacing $X_{corrupted}$ completely by X_{rev}. This approach permits to include a reasonable dynamical behavior of the signal, without increasing dramatically the uncertainties and noisy components (consistent with sugesstions in Del Giudice *et al.*, 2016). In addition, further tests showed strong benefits regarding the convergence of the DREAM algorithm by the approach in Eq 20 and Eq 21, compared to variations in which a constant value is introduced in gaps with $X_{corrupted} = 0$ (*e.g.* the mean of $X_{corrupted} > 0$ or an additive term). This might indicate that coarseness of the temporal resolution into the correction models (low or high N values) might have an important effect over their performance. This question is further addressed in this chapter.

Description of each rainfall correction model and evaluation scheme

Four rainfall correction models can be established from Eq 17, Eq 20 and Eq 21, based on the "error correction model" concept. The results of these models are computed with the maximum likelihood estimation of each $K_{correction}$ vector.

- CTW: $Y1_{corrected}=h(\theta_{opt}, K1_{correction}[t_i, t_{i+1}]_c \cdot X_{corrupted})$ over equal length time windows $[t_i, t_{i+1}]_c$ (from Eq 17).

- VTW: $Y2_{corrected}=h(\theta_{opt}, K2_{correction}[t_i, t_{i+1}]_v \cdot X_{corrupted})$, over variable length time windows $[t_i, t_{i+1}]_v$ determined by the proposed method (from Eq 17).

- CTWrev: $Y3_{corrected}=h(\theta_{opt}, K3_{correction}[t_i, t_{i+1}]_c \cdot X_{corrupted})$, over equal length time windows (from Eq 20 and Eq 21).

- VTWrev: $Y4_{corrected}=h(\theta_{opt}, K4_{correction}[t_i, t_{i+1}]_v \cdot X_{corrupted})$, over variable length time windows $[t_i, t_{i+1}]_v$ determined by the proposed method (from Eq 20 and Eq 21). Model CTW is already described in the literature (Sun and Bertrand-Krajewski, 2013a).

The three additional models (VTW, CTWrev and VTWrev) are established based on the variable-size time-windows and reverse model inclusion. The four models are tested to evaluate their ability to detect the generated controlled errors and re-construct the original rainfall time series.

Thirty error scenarios are introduced within the original rainfall time series (Figure 28), assumed as the known or controlled errors. Therefore, 30 $X_{corrupted}$ time series are tested (Figure 28a, b) to be corrected by each of the four rainfall error correction models (CTW, VTW, CTWrev, VTWrev). A given number of time windows N are proposed following a uniform distribution, for each $X_{corrupted}$ to be corrected by the models (Figure 28c). The evaluation process is detailed in Figure 28d. The scheme in Figure 28 has the purpose of guaranteeing that the same number of parameters N are used to correct a $X_{corrupted}$ time series (an iteration), making the results comparable across the models. In addition, the performance results (Figure 28e) can be grouped by the number of time windows N used in each iteration. This permits a performance comparison of corrections against different N values for any rainfall correction model.

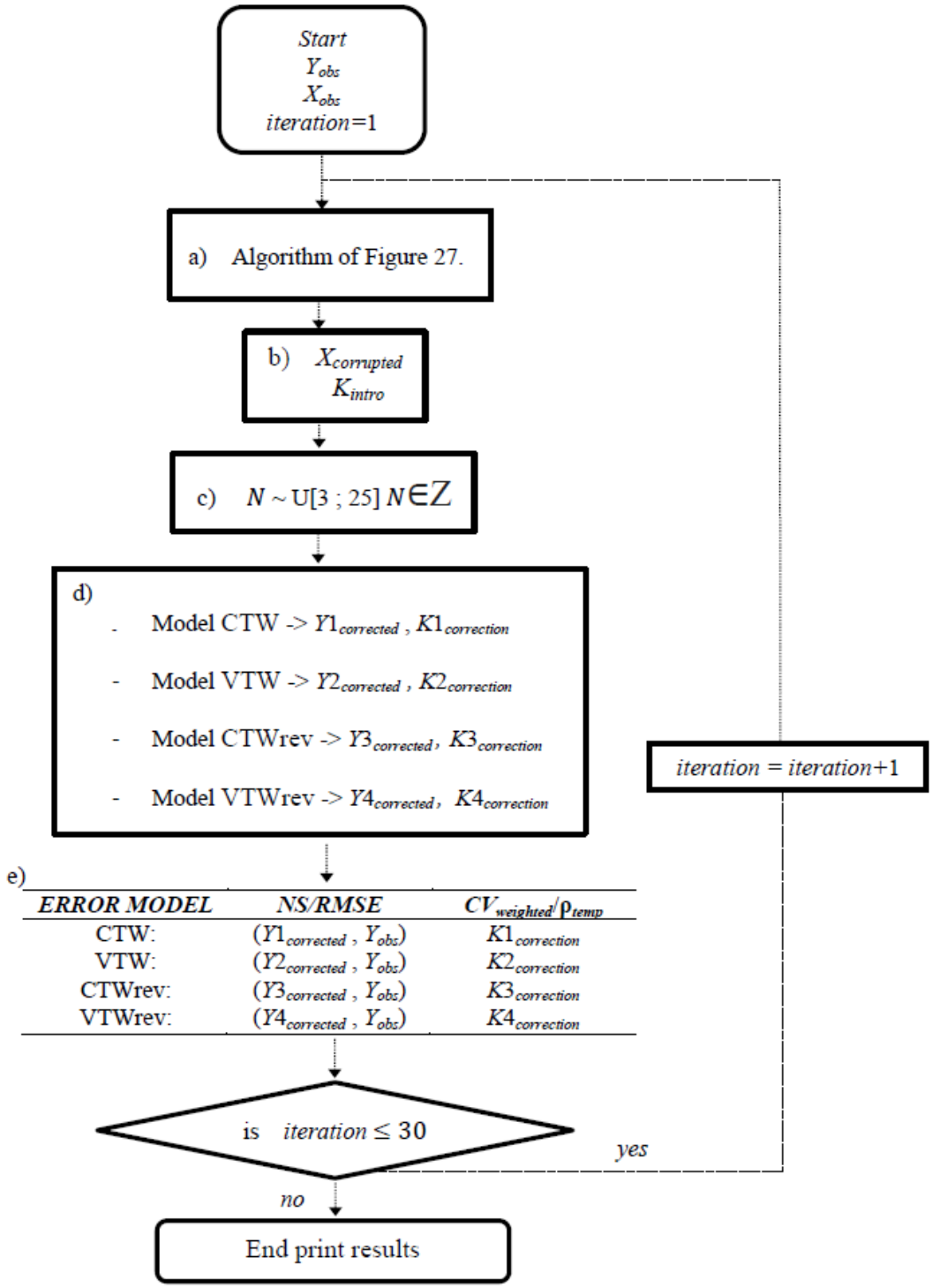


Figure 28. Scheme for evaluating the performance of the rainfall correction models (CTW, VTW, CTWrev and VTWrev) in 30 different $\tilde{X}_{obs}$ error scenarios, for a given rainfall event.

The proposed methods are applied to the 30 rainfall events described in the introduction of Part 2. For each event, the observed rainfall X_{obs} is corrupted for 30 iterations, bringing up 30 different scenarios of $X_{corrupted}$ (Figure 28), including different values for $N_{corrupted}$, $K_{intro}[t_i, t_{i+1}]$ and zeroes windows. Two main performance indicators are considered: the NS coefficient (Eq 12) and the RMSE (Root Mean Square Error), considering the maximum likelihood estimation of each $K_{correction}$ vector. Mentioned indicators are calculated by for $Y1_{corrected}$, $Y2_{corrected}$, $Y3_{corrected}$ and $Y4_{corrected}$ compared to Y_{obs} (Figure 28e). Therefore, the variability of the performance (NS and RMSE) of each model is analyzed, for each iteration, in each event

(900 cases for each rainfall correction model). Regarding uncertainties analyses about $K_{correction}$, the 95 % confidence bounds CB_{95} are calculated for all probability density functions (pdfs) of $K_{correction}[t_i, t_{i+1}]$ as a non-parametric measurement of dispersion. The CB_{95} value is different for each time window $[t_i, t_{i+1}]$, for a given iteration and a given rainfall event. For verifying the uncertainties of $K_{correction}$ in each model with a single value of a CB_{95} value representative of the whole number of corrections N, a weighted-over-time indicator is proposed in Eq 22, to be called $CB_{95\_weighted}$.

$$CB_{95\_weighted} = \frac{1}{dur(X)} \sum_{i=1}^{N} (t_{i+1} - t_i) \cdot CB_{95}(k_{intro}[t_i, t_{i+1}]) \quad \text{Eq 22}$$

Eq 22 gives the weighted average of the CB_{95} value for each influence window $[t_i, t_{i+1}]$, over the rainfall event. Consistently, there will be a single $CB_{95\_weighted}$ value for a given iteration and a rainfall event (900 values in total). The Bayesian approach permits as well to explore correlations among the different $K_{correction}[t_i, t_{i+1}]$ pfds (verifying aspects such as identifiability of the parameters). With the purpose of exploring correlations among the different $K_{correction}[t_i, t_{i+1}]$ pdfs parameters, the number of times in which $\rho(K_{correction}[t_i, t_{i+1}], K_{correction}[t_j, t_{j+1}])$ (for i and j $1 : N$ and $i < j$) has a p-value < 0.05 are counted and divided into the number of comparisons, *i.e.* $N/2 \cdot (N\text{-}1)$. These statistics to be called prob(ρ_{sig}) give the probability of having a significant correlation, for a couple of parameters distributions $K_{correction}[t_i, t_{i+1}]$ and $K_{correction}[t_j, t_{j+1}]$, obtained in a given iteration, for a rainfall event. There will be 900 prob(ρ_{sig}) values in total for each model, as for NS, RMSE and $CB_{95\_weighted}$ statistics.

5.3 RESULTS AND DISCUSSION

Regarding a single-event analysis, results are shown for the rainfall event measured from 11/02/2007 23:06 to12/02/2007 07:12. Figure 29 compares the measured hydrograph (blue), the hydrograph produced by a rainfall with generated errors (black) and with the rainfall corrected (red) by a well-performing iteration into the CTW model. Figure 30 shows the differences between identified and controlled errors $K_{correction}$ and K_{intro}, with illustrative purposes. The error structure is acceptably predicted over time as $K_{correction} - K_{intro} \approx 0$.

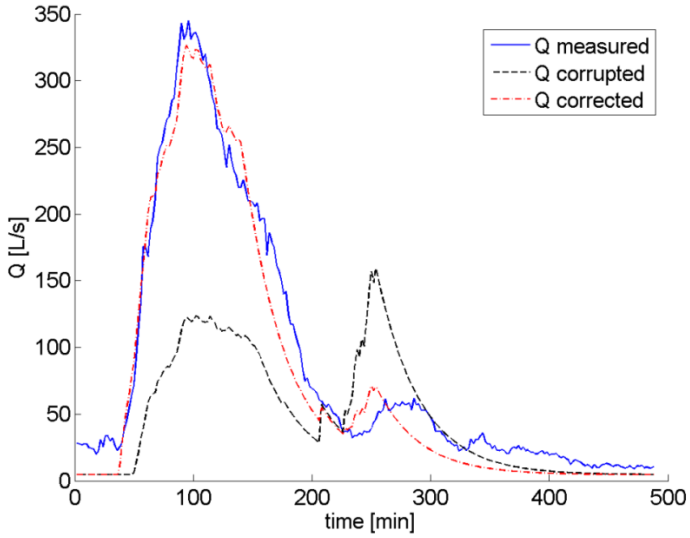


Figure 29. Hydrographs: measured (blue), produced by corrupted rainfall (black) and produced by corrected rainfall (red) by CTW model.

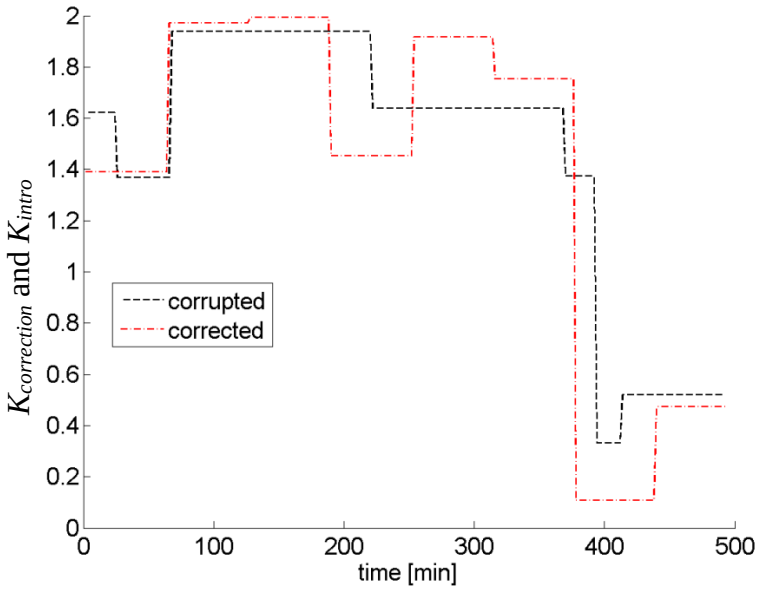


Figure 30. Well performing iteration of the *K* factor: $K_{correction}$ (red) and K_{intro} (black) by CTW model.

From an analysis at inter-event scale (30 events, 30 iterations per event), the values of adjustment between the original and corrected runoff are shown in Figure 31a, b. The performance of models can be listed as follows (from the most to the least performant): CTWrev, VTWrev, CTW, and VTW, in both terms of NS (Figure 31a) and RMSE (Figure 31b). This can be explained by fact that the models CTW and VTW are not able to correct rainfall with zeros records that are introduced as “known” errors.

a) b)

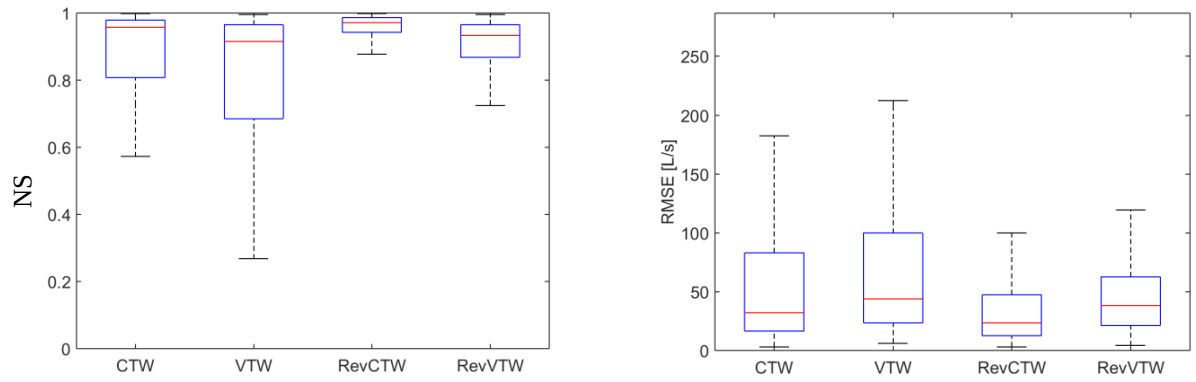


Figure 31. a) NS coefficient and b) RMSE between the original and runoff produced by corrected rainfall (for CTW, VTW, CTWrev and VTWrev models) (30 events, 30 iterations).

The mass balance analysis (total volumes per event and iteration, *i.e.* 900 values) with corrected rainfall compared to X_{obs} confirms an underestimation of rainfall total volume by models CTW and VTW (Figure 32a, b). This leads to envisage the importance of including terms for correcting zero rainfall into the correction models without falling into an ill-posed problem, as discussed by Del Giudice *et al.*, 2016 and in the following lines. In addition, results shown in Figure 32 are in agreement with Renard *et al.* (2010), in which rainfalls with larger volumes (total volume > 10 mm) are corrected much more precisely than smaller events (total volume < 10 mm) (especially for CTWrev and VTWrev in our research, Figure 32c, d).

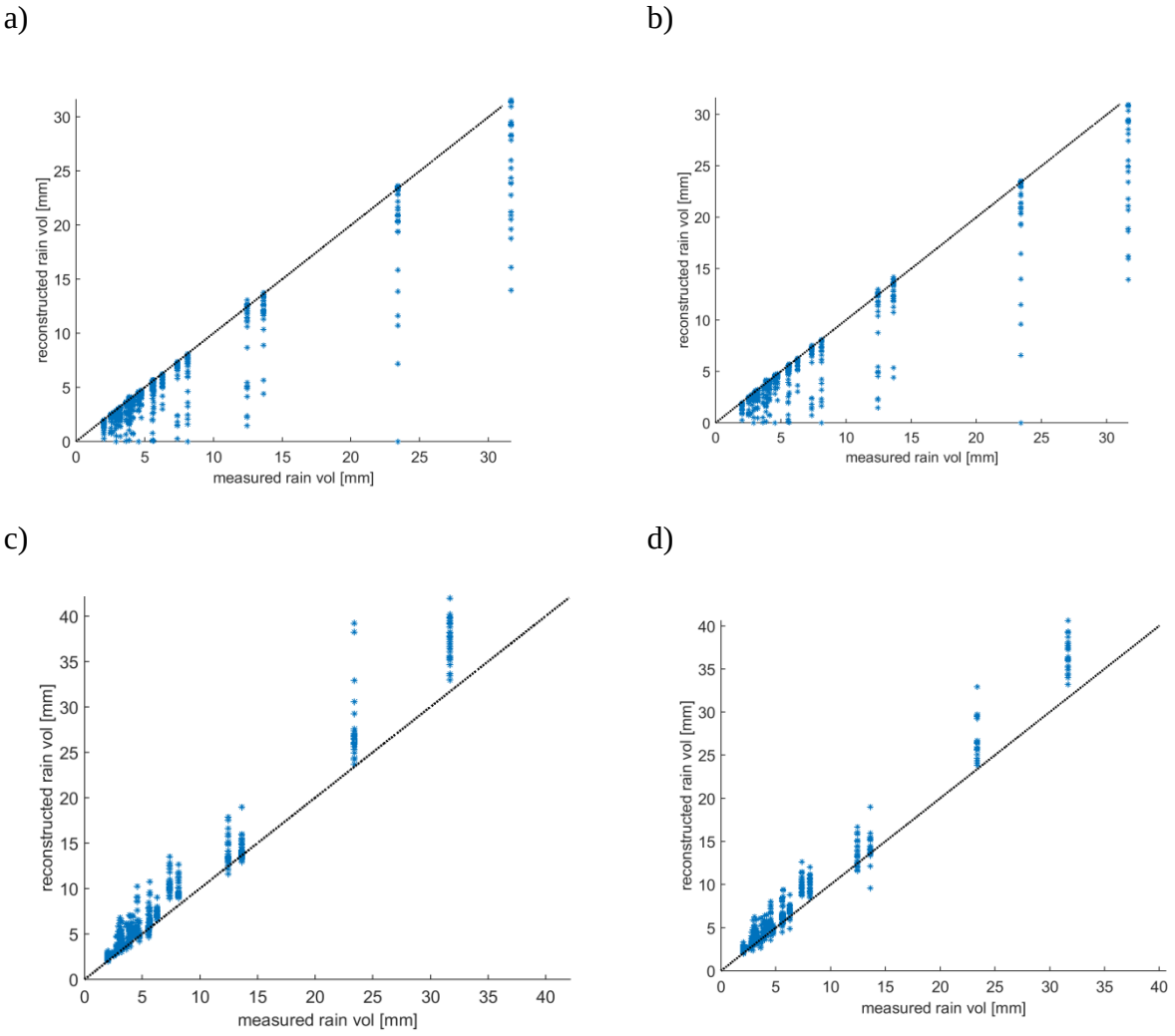


Figure 32. Total rainfall volume between the original and corrected rainfall, for a) CTW, b) VTW, c) CTWrev and d) VTWrev models (30 events, 30 iterations).

The rainfall correction based on reverse modelling coupled with the Bayesian method (CTWrev and VTWrev) offers a suitable strategy for dealing with zero rainfall records in terms of convergence. This can be explained as the reverse model gives an appropriate "clue" or reference point to the Bayesian optimization problem: the X_{rev} is produced by flow-rate values with a smooth enough behavior from the pre-calibration phase (NS > 0.95). In addition, the $K_{correction}$ values contribute to compensate fragments of $X_{corrupted}^{rev}$ mixed with the X_{rev} solution and original $X_{corrupted}$ values to be corrected. This fact makes as well the $CB_{95\_weighted}$ values of $K_{correction}$ (weighted uncertainties) lower for the reverse-based models than for CTW and VTW, with $CB_{95\_weighted}$ of about 0.6 for CTW to 0.2 in VTWrev (Figure 33a). Coefficients of variations *CV*s are about 17 % for CTW to 5 % in VTWrev. This result is comparable to Renard *et al.*, (2010) who found *CVs* of about 13 % for the correction factors in their rainfall correction implementation.

However, one should bear in mind that for our study case it is feasible to arrange the equations of the hydrological model (Eq 8, Eq 9 and Eq 10) in their reverse form. This situation is not frequent in a massive number of hydrological models for urban catchments. Therefore, reverse formulation should be undertaken by iterative methods, which can severely

increase the computational demand of the correction models CTWrev and VTWrev (adapted from Leonhardt, 2014). Further cases should be examined for evaluating the pertinence of using the reverse approach for more complex hydrological models. Indeed, for cases in which reverse formulation is not feasible, alternative approaches for dealing with zero records can be proposed, always being aware about the well-posseness of the problem.

The proposed variable time window approach (in VTW and VTWrev) deteriorates the quality of the predictions (Figure 31a, b). This can be explained by the hypotheses established for this method (*e.g.* linear transformation from T_v^* to T_v) in addition to its sensitivity to the α parameter. For many cases, the filter reported too short (or long) windows, leading to a not equitable repartition of the error corrections. The corrections made into very short windows (*e.g.* 6 minutes) do not have an effect on the flow-rate values generated by the corrected rainfall. Uncertainties for this correction $K_{correction}[t_i, t_{i+1}]$ over a short window (*e.g.* 6 minutes) are higher, but their influence over $CB_{95\_weighted}$ is lower (as this indicator is a weighted average depending on the duration of each window). For this reason, uncertainties ($CB_{95\_weighted}$) are slightly lower for variable-time-windows approaches (Figure 33a). The number N of correction windows for the Bayesian correction model tends to be used in a more "efficient" way by the constant time-window models (CTW and CTWrev). Additional explorations are undertaken by the use of Step Detection algorithms (*e.g.* Canny, 1986; Sandoval and Torres, 2013) aimed at distributing the time windows over the rainfall. The mentioned approach didn't show significant improvements of the results. Nevertheless, further explorations can be proposed towards this aspect.

a) b)

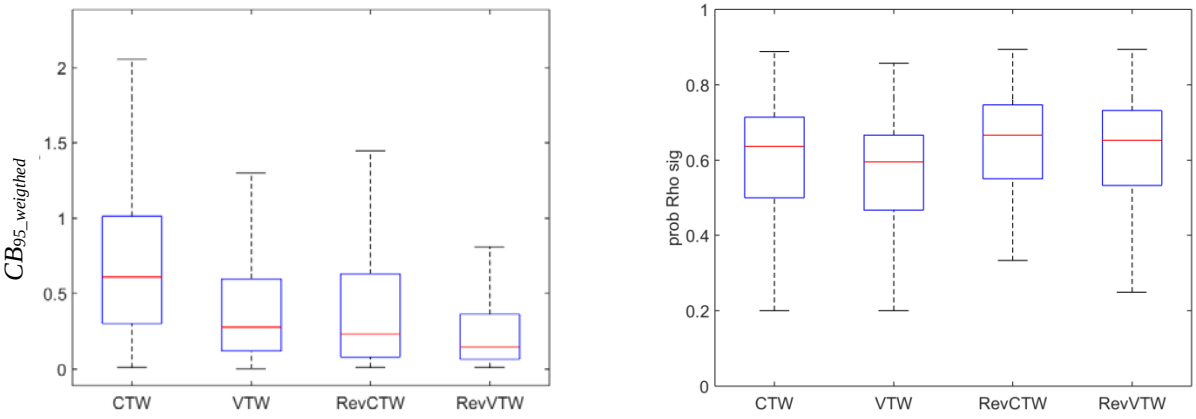


Figure 33. a) $CB_{95\_weighted}$ and b) prob(ρ_{sig}) statistics (for CTW, VTW, CTWrev and VTWrev models) (30 events, 30 iterations).

Another question herein addressed is the appropriate number of time windows N to be included into the rainfall correction models. Figure 34 shows the NS coefficients given by the four rainfall correction models, for different N values. Regarding the variable time-window models (especially VTW), the performance is lower than for the other models. In addition, the increment of NS coefficients as a function of N is less clear. In multiple cases, the correction shows less good results, although the number of windows N is high enough (*e.g.* 25 % of the NS coefficients < 0.4, for N = 16) (Figure 34b). This confirms the previously mentioned drawback about the variable time window approach: an inappropriate distribution of the error correction terms. Therefore, recommendations about the N selection for time window variable

approaches are harder to suggest. For the case of CTW, CTWrev (best performing model) and VTWrev, a number of time windows N of about eight seems to be the point beyond which no significant improvement is made (Figure 34a, c, d). These results can be compared with previous studies, in which preliminary recommendations about the number of time windows N are given (12 to 50 time windows) (Sun and Bertrand-Krajewski, 2013a). Recommendations from Sun and Bertrand-Krajewski, (2013a) about N can be higher due to facts such as: (i) the estimation of the $K_{correction}$ multipliers jointly with the calibration of parameters, (ii) the evaluation is done with different performance measurements (RMSE instead of NS) and (iii) results are obtained from a synthetic case study.

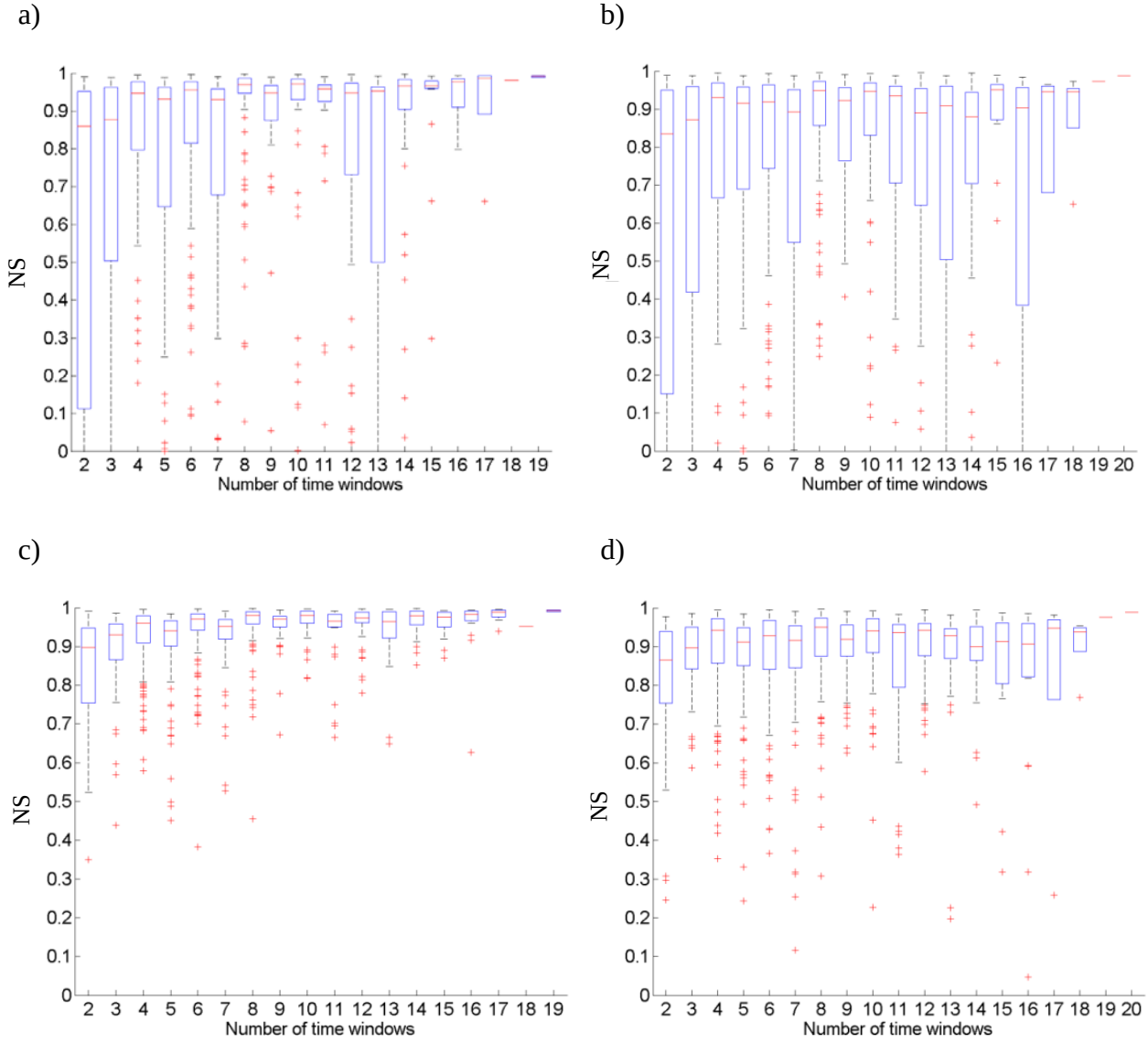


Figure 34. NS coefficient between the original and runoff produced by corrected rainfall vs the number of N time windows (k_i parameter), for (a) CTW, (b) VTW, (c) CTWrev and (d) VTWrev models (30 events, 30 iterations with different N values).

The error correction models assume a perfect correlation of rainfall uncertainties among all the time steps that belong to window $[t_i, t_{i+1}]$, as the pdf of the parameter $K_{correction}[t_i, t_{i+1}]$ is constant over $[t_i, t_{i+1}]$. This hypothesis should be established, as calculating a different pdf for each time-step leads to an ill-posed over-parameterization problem. However, correlation between two consecutive random variables $K_{correction}[t_i, t_{i+1}]$ and $K_{correction}[t_{i+1}, t_{i+2}]$ can be expected to be much lower, as they are two independent parameters into the calibration

process, theoretically uncorrelated to be fully identifiable. These correlation values among the $K_{correction}$ parameters are against identifiability, *i.e.* if $K_{correction}[t_i, t_{i+1}]$ and $K_{correction}[t_{i+1}, t_{i+2}]$ are highly correlated, a more parsimonious correction model can be proposed making $K_{correction}[t_i, t_{i+1}] = K_{correction}[t_{i+1}, t_{i+2}]$. However, completely uncorrelated $K_{correction}$ parameters (p-value < 0.05) will dismiss the temporal structure of the process to be represented. For our case study, the prob(ρ_{sig}) statistics is in average about 0.6 for all iterations and rainfall events (Figure 33b). This means that there is a probability of about 60 % to find a couple of parameters $K_{correction}[t_i, t_{i+1}]$, $K_{correction}[t_j, t_{j+1}]$ that are significantly correlated (p-value < 0.05). This indicator can bring insights about an appropriate balance between representation of the correlation along the temporal structure and parsimony of the model. However, this aspect is strongly recommended to be further addressed in future investigation.

Complementary results led in evidence the benefits of the reverse-based models. Trend tests revealed that there is a steeper trend to diminish the mean NS coefficients (mean of 30 iterations) as $dur(X_{obs})$ of rainfall events gets longer for CTW and VTW (Figure 35a). Analogous results are obtained with the increment of mean RMSE (mean of 30 iterations) as rainfall mean intensities during the event ($X_{corrupted}$) (mm/h) are higher (Figure 35b). In addition to the slope of the trend, mean NS and RMSE (mean of 30 iterations) are shown to perform better for reverse-based models, independently of $dur(X_{obs})$ and $X_{corrupted}$ (Figure 35).

a) b)

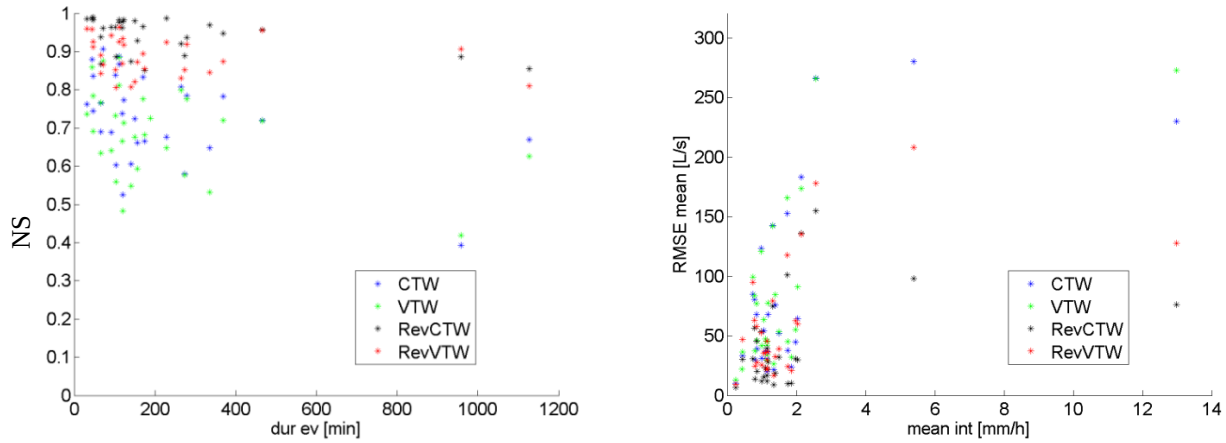


Figure 35. Trend analysis for a) the rainfall duration dur(X_{obs}) vs the NS between the original and runoff produced by corrected rainfall (mean of 30 iterations) and b) the mean rainfall intensity $X_{corrupted}$ vs the RMSE between the original and runoff produced by corrected rainfall (mean of 30 iterations) (for CTW, VTW, CTWrev and VTWrev models).

5.4 APPLICATION OF RAINFALL CORRECTION TO IDENTIFIED EVENTS WITH IMPORTANT UNCERTAINTIES IN RAINFALL MEASUREMENTS

The purpose of this section is to present the results of applying the proposed rainfall correction model by mixing the Bayesian and reverse approaches from Chapter 5 (CTWrev) to rainfall events hypothesized to be severely influenced by rainfall errors in Chapter 4 (the discarded 38 % of the 255 calibration events). For correcting rainfall in events with important errors in rainfall measurements by means of CTWrev, a set of parameters must be defined to recreate the backward rainfall intensity from flow rate observations. Given that from analyses

in Chapter 4 is established that there are at least two hydrological conditions that need to be separated in order to obtain meaningful results with the selected CRR model, the rainfall events can in principle be classified as *T*1 or *T*2 by using the Classification Decision Tree presented in Chapter 4. By classifying the rainfall events to be corrected as *T*1 or *T*2, the set of "optimal" parameters $\theta_{opt}T1$ or $\theta_{opt}T2$ can be used in the reverse formulation of the CRR model to obtain a more appropriate reconstruction of the rainfall intensity. However, one should bear in mind that rainfall events to be corrected will need to be classified as *T*1 or *T*2 by means of "uncertain" rainfall characteristics. A simple solution is to classify the events to be corrected by using their characteristics and verifying if, once the rainfall is corrected, the Classification Decision Tree assigns the rainfall event in the same group than before the correction. For the analyzed events, around 80 % are classified into the same departure group (*T*1 or *T*2). In case that the reclassification gives different results with the corrected event, an iterative process can be started until the group assignation converges. For the studied rainfall events, after the second reclassification practically all the events remained in the same group (*T*1 or *T*2).

A plot comparing the original rainfall intensities from discarded events in parameters estimation (presented as red events in Chapter 4, Figure 18) versus the corrected rainfall intensities by means of the rainfall correction model CTWrev (from Chapter 5) and the final classification *T*1 (blue) and *T*2 (green) is presented in Figure 36.

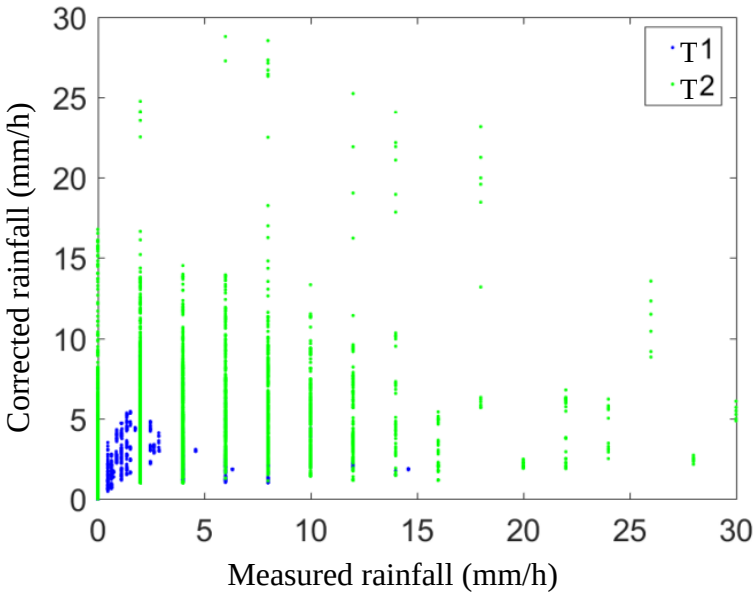


Figure 36. Measured versus corrected rainfall intensities (mm/h) for the discarded rainfall events from analysis in Chapter 5, including the final reclassification into *T*1 (blue) and *T*2 (green).

Figure 36 demonstrates from another perspective the hydrological difference of the data set in terms of "low intensity" and "high intensity" rainfall events, as corrections for both cases exhibit separated behaviors. The measured rainfall intensities from group *T*2 (green) are mainly corrected by a lower value, suggesting that over-estimations might be the principal source of error in rainfall measurements of non-reproducible (red) high intensity events. On the other hand, for the non-reproducible events (red) that can be associated with the lower intensities group *T*1 (blue) when rainfall is corrected, measured rainfall intensities are generally lower than the proposed mean areal rainfall reconstruction. The importance of the

reverse model in the rainfall corrections is reflected for both cases as rainfall intensities with zero values are corrected with intensities that can be up to 16 mm/h and 4 mm/h, for group *T*2 and *T*1, respectively. The rainfall intensities in group *T*2 (green) are more separated, as the rainfall measurement resolution becomes thicker for higher intensities. The rainfall correction carried out over the initially non-reproducible events (red) reported a NS > 0.8, for more than 85 % of cases. Therefore, all the estimated rainfall corrections are retained as representative However, a discrete classification of the events into two groups for representing the inter-event variability might lack of physical interpretability regarding the boundary between groups (highest or lowest rainfall intensity), which at the end is model and data dependent.

For further analyses in Chapter 6, the measured rainfall intensities R in the 365 events are tested as a potential input for different TSS stormwater models. Potential benefits of employing the corrected rainfall calculated in this section R_{corr} as an alternative input for TSS stormwater models are evaluated with the same 365 events, but replacing R for R_{corr} in the non-reproducible events (red).

5.5 CONCLUSIONS

A model-based approach is presented which aims to correct rainfall measurements and assess their uncertainties. Four error correction models are formulated, by mixing the Bayesian and Reverse approaches, jointly with the implementation of a constant/variable time-window method. The error's predictability is evaluated by introducing rainfall errors (10 % to 200 %) with variable time-windows in a well-performing pre-calibrated rainfall-runoff model. The methodology is applied to the Chassieu catchment.

From an analysis at a multi-event scale (30 events, 30 iterations per event), models can be listed in decreasing order of performance: CTWrev, VTWrev, CTW, and VTW, in terms of RMSE and NS coefficients, independently of duration or mean intensity of the events. The mass balance analysis with corrected rainfall confirms an underestimation of rainfall total volume by models CTW and VTW. This leads to envisage the importance of including informative terms for correcting zero rainfall into the correction models (Del Giudice *et al.*, 2016). Results show that rainfalls with larger volumes (total volume > 10 mm) are corrected much more precisely than smaller events (total volume < 10 mm), especially for CTWrev and VTWrev. Uncertainties in the corrections are lower for the reverse-based correction models as well.

The proposed variable time-window approach (in VTW and VTWrev models) tend to deteriorate the quality of the predictions. This can be attributed to the hypothesis in the variable time-windows definition method that leads to a not equitable repartition of the error correction windows. However, uncertainties of corrections factor for this variable time-window approach are slightly lower (temporal-averaged means of 95 confidence bounds from 0.6 to 0.2 and *CVs* from 17 % to 5 %) as highest uncertainties are found during shortest windows along the rainfall duration. A minimum number of about eight corrections (time windows) within a rainfall event can be recommended for the CTW, CTWrev, and VTWrev models. Trend tests shows that there is a steeper trend to diminish the NS values as the duration of the events gets longer for CTW and VTW models. Similar results are obtained with the increment of RMSE as rainfall mean intensities of the event are higher.

GENERAL CONCLUSIONS OF PART 2

In Chapter 4 the nature of model structure uncertainty and the inter-event parametric variability is addressed for a CRR model based on the idea of assessing the parameters marginal probability function obtained by event-by-event calibrations into conditional probability functions obtained by grouping the parameters from the event-by-event calibrations. The results stresses the importance of carefully selecting the data to be used for parameter estimations and further hydrological simulation, considering that the proposed parameter estimation strategy significantly improves the results of traditional parameter estimation strategies based on event-by-event and multi-event calibrations. One single rainfall-runoff model structure allows representing two groups of different hydrological conditions for an urban catchment by means of the variability of the optimal parameters found for each rainfall event. Furthermore, the proposed strategy led to identify rainfall events in which the rainfall error is likely to be high enough to be considered as unreproducible events, at least by the selected CRR model.

A model-based approach is presented in Chapter 5 to correct rainfall errors, demonstrating the advantages of a rainfall correction model in which the rainfall is corrected by multiplying factors over constant-length time window and rainfall zero records are filled with a reverse model. The rainfall events that are identified to be highly rainfall-error influenced in Chapter 4 are corrected by means of the rainfall error correction model recommended from results in Chapter 5. The correction of rainfall events shows to be consistent with the groups of hydrological conditions established in Chapter 5, showing a “correction dependency” as a function of the hydrological group in which the rainfall correction is undertaken. The correction of the rainfall events open further perspectives for water quality modelling, aspect that is studied in Part 3, as areal rainfall estimations might be more appropriate to model runoff quality than data from a single rain gauge. In Chapter 6, different water quality models are tested by including rainfall events with corrected rainfall as input in order establish relations between rainfall errors, hydrological conditions with performances of TSS stormwater models. Furthermore, Chapter 7 applies the proposed rainfall correction method of Chapter 5 as a more general estimator of observed virtual state variables, seeking for evidence about the existence of a virtual mass assumed by many of the traditional accumulation/wash-off models used for TSS stormwater modelling.

PART 3 REVISITING CONCEPTUAL STORMWATER QUALITY MODELS

During the past 40 years, modelling the dynamics of stormwater Total Suspended Solids (TSS) loads at the outlet of urban catchments has been mainly addressed by the idea of accumulation/wash-off originally by Sartor *et al.* (1974). Although this conceptual model does not constitute a rigorous physical description of the system, it is usually considered as a physically-based model in the sense that its parameters are interpretable from a physical point of view (Bonhomme and Petrucci, 2017). The original accumulation/wash-off idea from Sartor *et al.* (1974) establishes that the stormwater TSS load (kg) at the outlet of an urban catchment is given by the product of two processes, $M(t,u(t),\theta)$ (kg) and $W(u(t),\theta)$, (-), simultaneously interacting during the rain events as Eq 23:

$$\text{load(t)} = M(t,u(t),\theta)\cdot W(u(t),\theta) \qquad \text{Eq 23}$$

where M and W represent respectively the state of available mass of pollutant M (kg) to be transported by the wash-off term W (-), $u(t)$ is the input signal (rainfall R or flow rate Q) (m^3/s) and θ are the calibration parameters. In the original formulation of Sartor *et al.*, (1974) $M = M_0\, e^{-a\cdot u(t)^r\cdot t}$ and $W = a\cdot Q(t)^r$, constituting the model expressed in Eq 24 with a (-) and r (-) as the set of calibration parameters.

$$load(t) = M(t)\cdot W(t) = M_0\cdot\ e^{-a\cdot Q(t)^r\cdot t}\cdot a\cdot Q(t)^r \qquad \text{Eq 24}$$

For this case, the exponential decaying function M is obtained from a mass conservation principle, representing how the initial pollutant mass $M(0) = M_0$ is transported outside of the catchment during the event by W, without adding further mass inputs. For Sartor *et al.* (1974), the M_0 value was determined by coupling a build-up model that represents the accumulation of the pollutant mass as a function of the Antecedent Dry Weather Period ADWP (*e.g.* Freni *et al.*, 2009; Chow *et al.*, 2015). However, the need of this extension can be eliminated when M_0 is estimated as an additional parameter by direct event-based calibrations (*e.g.* Kanso *et al.*, 2005). The benefits of using build-up in TSS stormwater simulation models have been questioned by previous studies (Vaze and Chiew, 2002; Dotto *et al.*, 2011).

A wide amount of formulations based on Eq 23 can be found in the literature for representing the load dynamics during rainfall events, where two main categories of models can be distinguished, based on the formulation of M process:

- Time variable virtual process (TVP): The M process is commonly understood as the dynamics of the available mass over the catchment. This approach is usually cited in the literature as a mass-limited (ML) description (*e.g.* Piro and Carbone, 2014), with M following a decaying behaviour in time given by an *e.g.* linear, potential, exponential (*e.g.* Eq 24) or logistic functions (Egodawatta *et al.*, 2007; Crobedu and Bennis, 2011; Imteaz *et al.*, 2014; Zhao *et al.*, 2015; Qin *et al.*, 2016), limiting the flow rate production given by W. The variations in the description of M are generally aimed to retain a certain analogy with the mass conservation principle established in the case of Eq 24 (exponential decay). The set of parameters θ might change for different definitions of M (Egodawatta *et al.*, 2007; Freni *et al.*, 2009; Crobedu and Bennis, 2011), in all cases with M_0 as an additional parameter when no build-up model is coupled.

- Time constant virtual process (TCP): a second family of studies consider $M = M_0$ for all t (constant value), implying that there is no varying state understood as the pollutants mass to be potentially washed by W, and therefore M will not be a limiting factor of W to calculate the load. The load dynamics will be directly governed by the input $u(t)$ (m^3/s) in the W term (most commonly $W = a \cdot Q(t)^r$). This approach is cited in the literature as an infinite available mass (*e.g.* Kanso et al., 2005) or flow-limited (FL) description (*e.g.* Piro and Carbone, 2014). When M_0 is evaluated by calibration, usually $a = 1$ for having $W = Q(t)^r$ and avoid identifiability problems, obtaining the model in Eq 24 with M_0 (kg) and r (-) as the total set of parameters θ:

$$\text{load} = M_0 \cdot u(t)^r \qquad \text{Eq 25}$$

In this case, this model can be also referred to as the rating curve model RC (Huber *et al.*, 1988) and is probably the simplest model structure to be found in the literature.

For larger urban catchments (*e.g.* over 100 ha), further formulations have been applied by explaining the TSS load as a function of different contributions (*e.g.* $load_{surface}$, $load_{roads}$, $load_{parkways}$, $load_{sewer}$) or sub-catchments, aimed to keep a physically-based essence in the models (Robien *et al.*, 1997; Hong *et al.*, 2016). These studies have developed approaches including more detailed descriptions of the pollutant local processes inside the different components of the studied urban system (*e.g.* surface, roads, parkways, sewer system) (Freni *et al.*, 2009; Muleta *et al.*, 2012) or sub-catchments (*e.g.* Bonhomme and Petrucci, 2017), considering for example: (i) separate rainfall R (m^3/s) and flow rate Q (m^3/s) contributions as different inputs $u(t)$ (m^3/s) (Mannina *et al.*, 2010; Crobedu and Bennis, 2011; Bonhomme and Petrucci, 2017) and (ii) sedimentation/resuspension characteristics of the transported pollutants (Cristina and Sansalone, 2003; Shaw *et al.*, 2009; Mannina *et al.*, 2010; Wijesiri *et al.*, 2015). However, these approaches in their current state of development still represent important challenges related to parameters identifiability, in the absence of data and representative information about the load dynamics inside the system (Benedetti *et al.*, 2013).

Consequently, probably the most popular choice remains to adopt a "mean" or lumped description of the accumulation/wash-off process for the complete urban catchment, even if this last one is much larger than the experimental sites used for conceiving these TVP or TCP conceptual models (Bonhomme and Petrucci, 2017). Indeed, this lumped approach has been implemented in a massive amount of literature, considering the input $u(t)$ as rainfall R (*e.g.* Vaze and Chiew, 2002; McCarthy *et al.*, 2012; Manz *et al.*, 2013) or flow rate Q (*e.g.* Vaze *et al.*, 2003; Kanso *et al.*, 2005), including purposes such as real time control, climate change assessment, risk analysis, water management and in multiple commercial softwares (*e.g.* Freni *et al.*, 2009; Rossman, 2010; Muleta *et al.*, 2012). However, the unsatisfactory performance of this approach is frequently stated, as well as the difficulty of generalizing its results to real world applications, especially as urban catchments are large and complex (*e.g.* Vaze *et al.*, 2003; Deletic *et al.*, 2009; Dotto *et al.*, 2011). This can be explained by the non-generalizable nature of this accumulation wash-off idea to larger scales, where modelling errors of the TSS load are amplified when these lumped descriptions are faced to larger urban catchments (Liu *et al.*, 2012). Furthermore, even the existence of a "mean" state of available TSS mass for large catchments have been questioned from a low identifiably and a high spatial variability of models parameters (Bonhomme and Petrucci, 2017). Accordingly, one can ask under this "large urban catchments" context whether a traditional lumped TVP approach might be more adaptable than simplified lumped TCP descriptions. Many researchers have claimed the necessity of including an M essential process (always interpreted as an available TSS mass) by TVP formulations, while many others have found satisfactory results with TCP approaches

(RC models) (*e.g.* Kanso *et al.*,2005; Obropta and Kardos, 2007). Firstly, it is reasonable to argue that many of these discrepancies can be attributed to the following constraints in the experimental/methodological settings: (i) non representative controlled laboratory conditions or small catchments (ii) limited number of TSS data, (iii) limited number of rainfall events and (iv) insufficient assessment of uncertainty in data and model parameters. On the other hand, studies adopting alternative TCP formulations rather than the RC model are limited (*e.g.* Bai and Jing, 2012; Zhao *et al.*, 2015). Likewise, the notion of understanding *M* in TVP descriptions as a more general process missed by TCP formulations, which might be oversimplified or misinterpreted by the accumulation/wash-off idea, has not been explored according to author's best knowledge. Therefore, the TSS models are revisited under an Hypothetico-Inductive Data Based Mechanistic (HI-DBM) framework (Young, 2013), testing physically-interpretable Transfer Functions models (TF) jointly with traditional accumulation-wash-off models (TCP and TVP), applied to the Chassieu catchment with methods presented in Chapter 4.

The single input (rainfall *R* or flow rate *Q*) - single output (*load*) (SISO) HI-DBM model application for this work can be summarized in two stages, Chapter 6 and Chapter 7 respectively (adapted from Young, 1998):

Model identification TCP (Chapter 6) - different candidate TCP models, including: (i) black-box linear TF physically-interpretable model structures and (ii) previously known conceptual model with similar characteristics (*i.e.* RC model), are calibrated by means of a Bayesian method. One of the calibrated model structures is selected among them based on direct comparisons with the other structures by means of a set of statistical criteria over the calibration and verification data sets. Any model without a reasonable physical interpretability is discarded.

Model improvement TVP (Chapter 7) - the initial model structure selected in previous stage (Chapter 6) is sought to be improved if the results are still unsatisfactory in terms of performance or physical interpretation. For this purpose, statistical evidence of Time Variable virtual Processes TVP (or parameters as referred to in *e.g.* Young and Garnier, 2006) is sought in this initial model structure and a TVP(s) reconstruction(s) (*e.g.* fixed interval smoothing method for TF, see Young (1998), Bayesian methods, Reverse Modelling) can be undertaken. This reconstruction estimates how a parameter, interpreted as a virtual process, of a given model structure might change in time in order to bring the observed and simulated outputs as close as possible (see *e.g.* Young and Garnier, 2006). From statistical and physical interpretations of the TVP(s) reconstruction(s), one can replace the originally time constant parameter(s) in the TCP model structure by a finite parametric function that matches the assessed TVP(s) dynamics, constructing an improved TVP version of the initial TCP model. The parameters of the resulting model for TVP can be estimated by numerical optimization methods. The introduced finite parametrized function into the TCP model structure, or the TVP dynamics, should also have a physical meaning to be accepted, *e.g.* as a decaying *M* available TSS mass to be washed off. These TVP reformulated models structures are useful to describe potential missing process in the initial TCP model structure.

CHAPTER 6. REVISITING TIME CONSTANT VIRTUAL MASS MODELS WITH TRANSFER FUNCTIONS AND RATING CURVES

6.1 INTRODUCTION AND BACKGROUND

The TCP model structures are characterized by the simplification of not including an independent *M* process. Therefore, the TSS load dynamics is mainly governed by *W* and consequently the TSS load is produced as a monotonic function of the input dynamic $u(t)$ (m^3/s), without the influence of a potential limiting factor *M*. Revisiting further TCP descriptions might bring evidence to question the necessity of applying a more complex model from the TVP family than a simplified TCP approach, if satisfactory enough performances are obtained by TCP. Although there has been an increasing interest during recent years in Transfer Functions TF models into the environmental context (*e.g.* Jakeman *et al.*, 1990; Young 2003 and the prior references therein), implementations for water quality modelling are less frequent (*e.g.* Davis and Atkinson, 2000; Young and Garnier, 2006). Therefore, TFs emerge as a promising research direction towards alternative TCPs description, in the sense that all of them are monotonic functions of the input dynamic $u(t)$ (m^3/s) and no virtual state variables are introduced in the calculation.

One immediate difference between TFs and RC is that the second has a well-known physical meaning, linking shear stress produced by flow rate to TSS resuspension and transfer (Crobedu and Bennis, 2011). On the other hand, TFs models are in principle black-box descriptions. However, system identification and control enhance their potential to be interpreted as serial, parallel and feedback connections of sub-systems that often have a physical meaning (Young and Garnier, 2006).

Therefore, further TCP descriptions are explored by visiting different TFs in the first stage of the proposed HI-DBM application, which are compared to the traditional RC model structure. Parameter identification and uncertainties are assessed for TFs and RC by the use of the Bayesian calibration approach proposed in Chapter 4, given the restriction of further methods such as the Simplified Refined Instrumental Variable method for TFs due to uncertainty-free input/output hypotheses (adapted from Pedregal *et al.*, 2007). A statistical analysis in order to recommend a model structure is undertaken based on tests in the preliminary analysis of the HI-DBM frame (AIC, BIC and YIC) (Young, 2013), together with further analysis and performances for verification data (NS, AIC, POC*mod*) (from Chapter 4). The differences of the results by using flow rate at the outlet of the catchment *Q*, the rainfall *R* or the corrected rainfall R_{corr} (from Chapter 5) as inputs $u(t)$ (m^3/s) are also discussed.

6.2 METHODOLOGY

TCP model structures: Transfer Functions (TFs) and Rating Curve (RC)

A single-input single-output (SISO) TF can be expressed by the following equations:

$$\begin{cases} x(t) = \dfrac{B(s)}{A(s)} u(t-\tau) \\ load(t) = x(t) + e(t) \end{cases} \quad \text{Eq 26}$$

where:

$$\begin{aligned} A(s) &= s^n + a_1 s^{n-1} + \cdots + a_{n-1} s + a_n \\ B(s) &= b_0 s^m + b_1 s^{m-1} + \cdots + b_{m-1} s + b_m \end{aligned} \quad \text{Eq 27}$$

$A(s)$ and $B(s)$ are polynomials with the derivative operator $s = d/dt$ and τ (min) is a pure time delay. In Eq 26 and Eq 27, $u(t)$ is the input signal, $x(t)$ is the noise free output signal and $load(t)$ is the noisy output signal. The component $e(t)$ is considered i.i.d., however this assumption is not restrictive for the TF application (Young, 1998). A TF is a compact representation of a differential equation. Therefore, the physical interpretability of Eq 26 can be directly related to the following differential equation form Eq 28 (Garnier and Young, 2006):

$$\frac{d^n load(t)}{dt^n} + a_1 \frac{d^{n-1} load(t)}{dt^{n-1}} + \cdots + a_n y(t) = b_0 \frac{d^m u(t-\tau)}{dt^m} + \cdots + b_m u(t-\tau) + A(s) \cdot e(t) \quad \text{Eq 28}$$

A given model structure for this type of models is defined from the pair (n poles, m zeros) $TF_{n,\,m}$. The number of parameters for the set θ is equal to $n + (m + 1)$, where the set $\theta = [a_1, .., a_n, b_0, \ldots, b_n]$. For RC in Eq 25 the set of parameters $\theta = [M_0, r]$. Seven different model structures are tested in this Chapter, including $TF_{0,0}$ $TF_{1,1}$ $TF_{2,1}$ $TF_{2,2}$ $TF_{3,2}$ $TF_{3,3}$ and RC. For Q as input $u(t)$ the delay $\tau = 0$ and the number of parameters is: 1, 3, 4, 5, 6, 7 and 2, respectively. For rainfall R or R_{corr} as the input $u(t)$, τ is added as another parameter for all model structures. One difference between TFs and RC is their linear and non-linear nature, regardless the number of parameters. On the other hand, TFs led to include information from previous time-steps (measured Q and/or simulated TSS load) in the calculation, contrary to the RC in which the calculation of $load(t)$ only depends on $u(t - \tau)$. Implications of these differences are further discussed in the results and discussion section.

Parameter identification

Although there is a wide range of calibration methods in the literature for TF, most of them are based on hypotheses related to normality and data input free-error (Pedregal *et al.*, 2007). For being consistent with the development of this work, the identification of the set of parameters for a TF or RC is preferred to be undertaken by a Bayesian approach (Vrugt *et al.*, 2016), under event-based calibrations, as proposed in Chapter 4. For a given calibration rainfall event i ($i = 1 : 255$), θ_i is the local set of parameters of a given model structure and $p(\theta/load)_i$ their probability density function (pdf), given an input $u(t)_{obs\,i}$ (Q, R or R_{corr}) and an output $load(t)_i$. The Bayesian approach leads to calculate local $p(\theta/load)_i$, named posterior distribution, over the basis of a likelihood function and a prior knowledge of the distribution of parameters $p(\theta)$, which is expressed by Eq 29:

$$p(\theta/load)_i = C \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\hat{\sigma}_t^2}} exp\left[-\frac{1}{2}\frac{(res(t)_i)^2}{\hat{\sigma}_t^2}\right] \cdot P(\theta) \quad \text{Eq 29}$$

$$res(t)_i = (load_{sim}(t,\theta)_i - load_{obs}(t)_i) \quad \text{Eq 30}$$

where n is the number of input data in $u(t)_{obs\ i}$, $load_{sim}(t, \theta)_i$ is the simulated load by a given TF or RC model at a given time step t and a set of parameters θ_i, $p(\theta)$ is a uniform probability distribution for each parameter (informative-less), C is a normalization coefficient (irrelevant for the implementation of numerical algorithms to solve Eq 29, see Vrugt *et al*., 2016), $res(t)_i$ are the residuals of the model in Eq 30, and $\hat{\sigma}_t^2$ is the residual variance, considered for this application to be equal to the squared value of $load_{obs}(t)_i$ standard uncertainty for each time-step t. The DREAM algorithm is used for determining $p(\theta/load)_i$ as a solution to Eq 29. The local set of parameters that represents the optimal parameters values among all probable values of $p(\theta/load)_i$ is called $\theta_{opt\ i}$.

The global estimation of $p(\theta/load)$ is calculated as the marginal distribution of representative local estimations of $p(\theta/load)_i$, $\theta_{opt\ i}$ estimations with Nash-Sutcliffe efficiency coefficient (NS) < 0.8 are discarded as unrepresentative, and θ_{opt} = mean($p(\theta/load)$). However, this marginal $p(\theta/load)$ is encouraged to be treated as a conditional probability function, by means of the SE Conditional strategy (SEConditional), given the potential benefits presented in Chapter 4 for this approach. The objective of SEConditional is to divide the marginal distribution $p(\theta/load)$ into *Tnum_c* conditional probability functions $p(\theta/load, Tnum\_c)$ and $\theta_{opt\ Tnum\_c}$ = mean($p(\theta/load, Tnum\_c)$. The idea behind this division is to improve the accuracy and diminish the uncertainties in the simulations for the verification stage (see details in Chapter 4). However, its applicability depends on the possibility of linking a verification event to a given *Tnum_c* group by means of characteristics of the input signal $u(t)_{obs}$ (*e.g.* max., min., mean, volume, duration, ADWP), in order to establish which $p(\theta/load, Tnum\_c)$ conditional function should be utilized for simulating the TSS load. Hence, the marginal $p(\theta/load)$ function can only be represented by *num_c* conditional functions $p(\theta/load, Tnum\_c)$ if a clear association between characteristics of this input $u(t)_{obs}$ and a given conditional *Tnum_c* function can be established. Otherwise, the marginal estimations $p(\theta/load)$ and θ_{opt} are recommended to be retained.

Model identification

A methodology for identifying the most suitable model structure in the TCP models family is developed based on the HI-DBM framework (Young, 2013), including the proposed parameter identification method. The candidate model structures j = 1: *num_str* are the traditional RC (2 parameters) and the following TFs (n poles, m zeros): [0,0], [1,1], [2,1], [2,2], [3,2], [3,3], (1, 3, 4, 5, 6, 7 parameters) for a total of model structures *num_str* = 7. The methodology is explained by the following Figure 37:

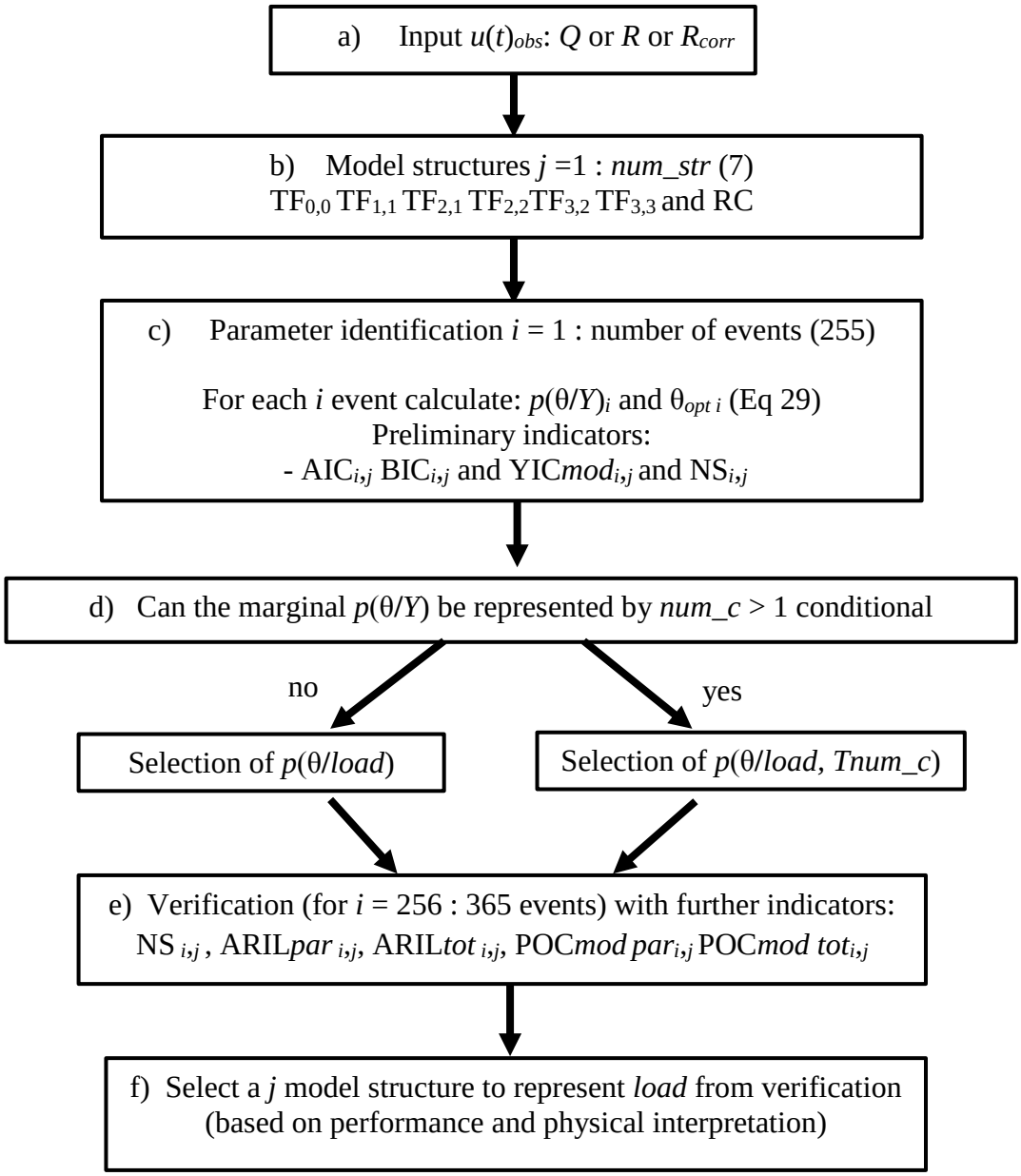


Figure 37. Model selection diagram: a) selection of an $u(t)$ input (Q, R, R_{corr}) b) proposed model structures, c) parameters identification for a given model structure (calibration) and evaluation of AIC, BIC, YIC*mod* and NS for each calibration event, d) $p(\theta/load)$ marginal or conditional representation, e) verification with $p(\theta/load)$ calculating NS, ARIL*par*, ARIL*tot*, POC*mod par*, POC*mod tot* for each verification rainfall event, f) selection of a model structure from b).

The first step consists in selecting an input $u(t)_{obs}$ between Q, R or R_{corr} (Figure 37a). Afterwards, a model structure from j = 1 : *num*_str is selected to be evaluated (Figure 37b), undertaking a preliminary analysis from three statistical criteria with the calibration dataset: the AIC_j (*Akaike Information Criterion*) (Akaike, 1974), the BIC_j (*Bayesian Information Criterion*) (Schwarz, 1978), a modified version of the YIC_j (*Young Information Criterion*) (Young, 1998), to be called $YICmod_j$ and NS_j (Nash-Sutcliffe efficiency coefficient). However, these I indicators $I_{i,j}$ (I for generality) are different for each rainfall event, given the inter-event variability. Accordingly, the $AIC_{i,,j}$, $BIC_{i,,j}$, $YICmod_{i,j}$ and $NS_{i,,j}$ are directly calculated from each event from local $p(\theta/load)_i$ and $\theta_{opt\ i}$ estimation (i =1 : 255 for the study case) to be further compared (Figure 37c). These comparisons of $I_{i,j}$ lead to establish if the

proposed model structures are feasible solutions and finding initial evidence about a probable best structure based on statistical criteria, to be confirmed by the verification data (Figure 37d and Figure 37e). It is worth clarification that the methodology presented in Figure 37 can be applied to any dataset when it is desired to select a single model structure, considering the inter-event variability of the optimal sets of parameters.

A brief description of these indicators *I* for calibration is given in the following lines.

The AIC, BIC and YIC*mod* are aimed to bring a basis for selecting a particular model structure from a group of models, by selecting the model structure that reports the minimum value among the other model structures to which it is compared. One should bear in mind that each statistics has a specific objective and was developed under different theoretical basis.

The objective of AIC model selection is to estimate the information loss when the probability distribution *f* associated with the true model is approximated by the probability distribution *g*, associated with the model that is to be evaluated. A measure for the discrepancy between the true model and the approximating model is given by the Kullback–Leibler (1951) information quantity *Information* (*f*, *g*) (see Burnham and Anderson, 2002). The AIC can be calculated as follows by Eq 31:

$$AIC = n \cdot \ln\left\{\sum_{t=1}^{n} res(t)^2\right\} - 2n_P \qquad \text{Eq 31}$$

where *n* is the number of data in the residuals *res*(*t*) and n_p is the number of parameters. One hypothesis on which the mathematical strength of the AIC indicator depends is the white noise behavior of the residuals in Eq 30 (Young, 1998). However, several statistical methods based on this i.i.d. residuals assumption are still widely applied in urban drainage modelling despite residuals do not follow this condition. Furthermore, the introduction of mathematical transformation for accomplishing normality of residuals has shown to decrease the predictive ability of the models (*e.g.* Dotto *et al.*, 2013). On the other hand, a robust aspect of this indicator is the implicit assumption that the "true" model is unknown, at least for the candidate set of models that are tested, which is usually the case in practical applications (Young, 1998; Young, 2012).

A second indicator included in the HI-DBM framework is the BIC (Young, 1998). The BIC has being also proclaimed to be less prone to over-fitted models than the AIC (Young, 2012). On the other hand, the BIC indicator is based on the hypothesis that the true model is one of the candidate models, which is in principle problematic under a practical point of view, as none of the proposed models constitute a the true model, given their nature of conceptual approximations (Burnham and Anderson, 2002). BIC is defined as Eq 32:

$$BIC = n_P \cdot \ln\{n\} - 2\ln\left\{\sum_{t=1}^{n} res(t)^2\right\} \qquad \text{Eq 32}$$

A favored model by the BIC minimizes the BIC value. According to Eq 32, the BIC also rewards model goodness-of-fit with maximized likelihood and penalizes lack of model parsimony. An extensive comparison between AIC and BIC can be found in *e.g.* Aho *et al.* (2014), and further references recommend to include both indicators in any model identification diagnosis (*e.g.* Yang, 2005).

The YIC (Young, 1989) is a heuristic indicator based on intuition, as it is not a formal statistical approach. The YIC in its classical form is a function of the standard deviations σ_θ of

the parameters of the set θ. However, the Bayesian approach proposed Eq 29 allows to deliver a detailed description of $p(\theta/load)$, which might be neither normal nor symmetric and therefore cannot be summarized by a mean and a standard deviation, impeding the direct application of YIC proposed by Young, (1989). The uncertainty of $p(\theta/load)$ can be characterized by means of a non-parametric measurement of dispersion aimed to represent *e.g.* 95 % of the coverage interval of the random variable. Therefore, a modification of the original YIC to be called YIC*mod* is proposed for this Chapter, modifying the standard deviation of parameters estimations σ_θ in the original YIC indicator by a non-parametric estimation of dispersion. Accordingly, the YIC*mod* is calculated as follows by Eq 33:

$$YICmod = \ln\left\{\frac{\sum_{t=1}^{n} res(t)^2}{\sigma_{load}^2}\right\} - \ln\left\{\frac{1}{n_p}\sum_{k=1}^{n_p}\frac{p(\theta(k)/load)_{97.5} - p(\theta(k)/load)_{2.5}}{\theta(k)_{opt}}\right\} \quad \text{Eq 33}$$

where $p(\theta(k)/load)_{97.5}$ and $p(\theta(k)/load)_{2.5}$ are the 97.5 % and 2.5 % limits of the $p(\theta(k)/load)$ distribution, for a 95 % coverage of the parameter $\theta(k)$ from the set θ. The $\sigma_{load}^2$is the standard deviation of the load pollutograph. Indeed, the calculation of YIC*mod* is directly linked to the parametric uncertainties and the first term is simply a relative, logarithmic measure of how well the model explains the data: the smaller the model residuals the more negative the term becomes. The second term in Eq 33, on the other hand, provides a logarithmic measure of the uncertainty in parameters. If the model is over-parameterized, then it can be shown that the covariance matrix of parameters will increase in value, often by several orders of magnitude (*e.g.* Young *et al.*, 1998). When this is the case, the second term in the YIC*mod* tends to dominate the criterion function, indicating over-parameterization. The YIC*mod* indicator is commonly a large negative value (due to the logarithmic term), but the choice is not critical provided the associated NS is relatively high compared with that of other models (adapterd from Young *et al.*, 1998).

However, neither the AIC, BIC nor YIC*mod* are aimed to bring direct information about the quality of the estimation from the model, as these indicators are exclusively designed for comparative purposes in a set of model structures (Figure 37c). Hence, the HI-DBM framework proposes to complement the analysis in calibration with the classical coefficient of determination R^2 (Nash-Sutcliffe efficiency coefficient - NS - in the hydrological context).

The NS assesses the portion of the total variance of the output load data σ_{load}^2 able to be explained by a given model structure. This indicator can be calculated as follows Eq 34:

$$R^2 = NS = 1 - \frac{\sum_{t=1}^{n} res(t)^2}{\sigma_{load}^2} \quad \text{Eq 34}$$

The NS is not recommended to be used as a unique model selection criterion with calibration data, given its sensitivity to over parametrization (Young, 1998). The NS indicator in calibration is mainly aimed to complement AIC, BIC nor YIC*mod* results, in terms of verifying that the proposed models structures have acceptable performances (*e.g.* NS > 0.6).

The HI-DBM recognizes this preliminary analysis by means of the calibration data as a powerful inference tool to scrutinize and discard inappropriate models from a preliminary stage, and even to deliver strong conclusions about model performance whenever verification is not possible. However, inferences obtained from the verification stage (Figure 37e) offer without doubt more generality in the sense of transferability and robustness, as presented in Chapter 4. The performance of the models is then assessed by means of $p(\theta/load)$ and θ_{opt} (or $p(\theta/load, Tnum\_c)$ and $\theta_{opt\ Tnum\_c}$ if a conditional division is conceivable, Figure 37d) for the

remaining i = 256 : 365 verification events, including the following indicators $I_{i,j}$: the size of uncertainty intervals of load predictions (precision) (ARIL, Vezzaro and Mikkelsen, 2012), the number of measurements inside the uncertainty intervals (reliability) (POC*mod*, from Chapter 4 and Ye *et al.*, 2014) and the mean prediction for the verification events (accuracy) (NS) as for the verification dataset in Chapter 4 (Figure 37e). The ARIL and POC*mod* can be calculated by the following Eq 35, Eq 36 and Eq 37:

$$ARIL = \frac{1}{n}\sum_{t=1}^{n} \frac{Limit_{sup,t} - Limit_{inf,t}}{load_{sim}\left(t, \theta_{opt}\right)} \quad \text{Eq 35}$$

$$C_t = \begin{cases} 1 \; load_{obs}(t) - 2\hat{\sigma}_t \leq Limit_{sup,t} \; and \; load_{obs}(t) + 2\hat{\sigma}_t \geq Limit_{inf,t} \\ 0 \qquad\qquad otherwise \end{cases} \quad \text{Eq 36}$$

$$POC_{mod} = \frac{1}{n}\sum_{t=1}^{n} C_t \quad \text{Eq 37}$$

where n is the number of input $u(t)$ data, $load_{obs}$. $load_{\text{sim}}(t, \theta)$ is the simulated load by the model at time step t from the observed rainfall $u(t)_{obs}$ and global optimal parameters θ_{opt}. $Limit_{sup,t}$ and $Limit_{inf,t}$ are the upper and lower limits for a confidence interval of 95 % at time step t, obtained from $p(\theta/load)$.

The total simulated output uncertainty in reality will not be completely captured by parametric uncertainties. This phenomenon can be attributed to remnant errors in calibration that are not accounted for the input and structural error assumptions (Thyer *et al.*, 2009). Therefore, the parametric uncertainties given by $p(\theta/load)$ can be propagated to estimate the total simulation output uncertainty. For comparative purposes, the total simulation output uncertainty is estimated by following the method proposed by Dotto *et al.* (2011). The residuals obtained from the rainfall events used in calibration are binned as a function of modelled loads. This allows constructing probability distributions of residuals as a function of the modelled load values. For a verification event, as the modelled load value at t are a function of a set of parameters, a different probability distribution of residuals is obtained by considering each realization of $p(\theta/load)$ (or $p(\theta/load, Tnum\_c)$). The total uncertainty of simulated load at t will be then given by the ensemble of all the probability distributions of residuals obtained at t, as a result of propagating the parametric uncertainties (see further details in Dotto *et al.*, 2011). Therefore, two complementary indicators, to be called ARIL*tot* and POC*mod tot*, are calculated for the estimated total output uncertainty simulations in the 110 verification events.

However, delivering a conclusion about a best j model structure, based on the data set from events i = 1 : 255 (calibration) or i = 256 : 365 (verification) and the corresponding $I_{i,j}$ estimation represents additional challenges. Indeed, the model selection problem is an active field of research by itself (*e.g.* Murtaugh, 2014) and is commonly addressed from comparisons of the I scores obtained by the candidate model structures (*i.e.* AIC, BIC and *YICmod* for calibration or NS_{val}, ARIL*par*, ARIL*tot*, POC*mod par* and POC*mod tot* in verification) (Figure 37f). However, this selection is not direct, given the inter-event variability either for calibration or verification, for which the best j model structure might be different from one event to another. For the context of this work, this aspect is analyzed by comparing the indicators $I_{:,j}$ under two perspectives:

- Approach of selection 1 (S1): the model selection is undertaken based on the idea of evaluating if the inter-event mean of a given indicator mean($I_{:,\ j}$) calculated for a model structure j is significantly higher (or lower) than the inter-event mean for the other model structures $I_{:,\ \neq j}$. By using an ANOVA or Kruskal-Wallis statistical test, the group $I_{:,j}$ is compared to all of the remaining groups formed by each structure $I_{:,\ \neq j}$ (the number of samples for each group is equal to the number of events), evaluating the null hypotheses that the mean($I_{:,1}$) = mean($I_{:,2}$) = mean($I_{:,3}$) = ... = mean($I_{:,\ j}$) = ...= mean($I_{:,\ num\_struc}$) versus the alternative hypothesis that these means are not equal. Whenever the ANOVA or Kruskal-Wallis test reports a p-value < 0.05 between two groups, one can conclude that there is enough statistical evidence to state that the mean($I_{:,1}$) is higher (or lower) than the mean($I_{:,2}$). The selection of a more adaptable test between the ANOVA (normality) and Kruskal-Wallis (non-parametric) is dependent on the normality of the groups $I_{:,\ j}$ to be compared (tested by the Shapiro Wilk test for each $I_{:,\ j}$ group separately). The ANOVA test is preferred instead of the Kruskal-Wallis test only if there is enough statistical evidence that all the j = 1: *num_str* groups are normally distributed (Shapiro Wilk test with p-values $_j$ > 0.05 for all j = 1: *num_str* groups) (Figure 38).

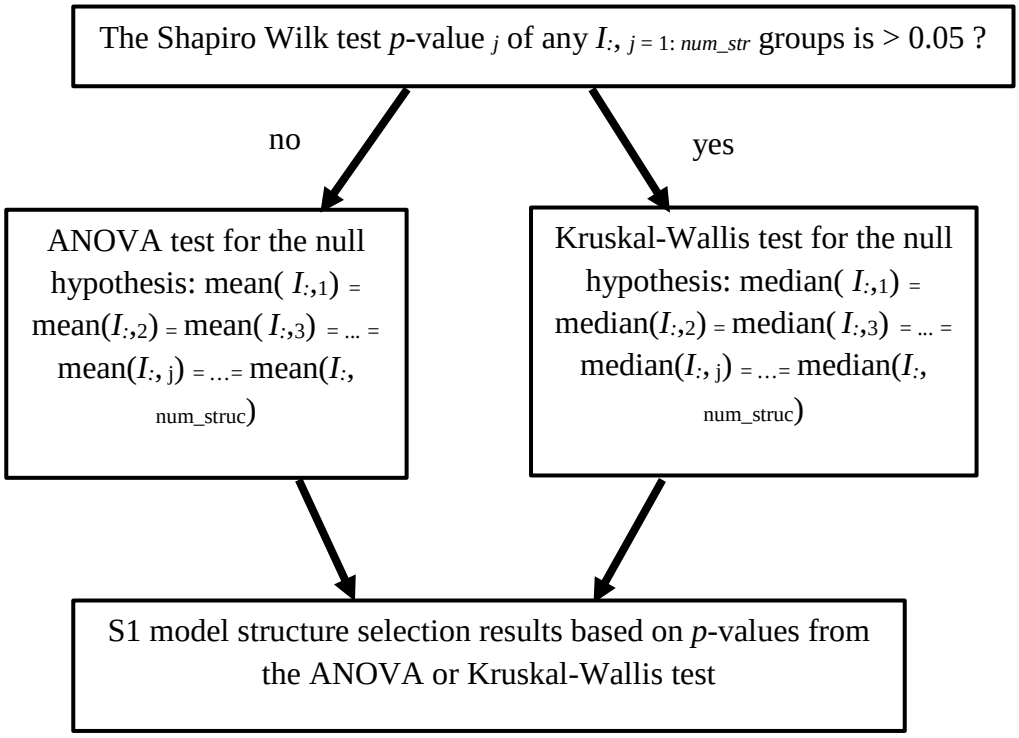


Figure 38. Application of the S1 approach for model selection.

- Approach of selection 2 (S2): the model selection is undertaken based on the number of victories (*i.e.* number of i events in which $I_{i,j} > I_{i,\neq j}$) lower (or higher) than the number of victories that might be obtained by chance (binomial distribution), giving a significance level of 95 %. An intermediate aspect should be addressed under this perspective, in which one needs to define when the value of $I_{i,\ j}$ is significantly higher or lower than for the others structures $I_{i,\ \neq j}$, to declare it as the winner j structure for a given rainfall event i. For this purpose, a Bayesian likelihood approach is adopted, inspired from the Alkaline weights for the AIC (Aho *et al.*, 2014), in which the values of an indicator $I_{i,j}$ are weighted by using all the values obtained for the analyzed structures $I_{i,\ :}$, in order to represent the probability that $I_{i\ ,j}$ is the best model for a given i rainfall event. The probability $p(I_{i,j} > I_{i,\neq j})$ can be expressed as follows:

$$p\left(I_{i,j} > I_{i,\neq j}\right) = \frac{\exp(-{}^{\Delta I_{i,j}}/_{2})}{\sum_{r=1}^{n} \exp(-{}^{\Delta I_{r}}/_{2})} \quad \text{Eq 38}$$

where:

$$\Delta I_{i,j} = \left|I_{i,j} - I_{i,opt}\right| \quad \text{Eq 39}$$

and

$$I_{i,opt} = \begin{cases} max_{j=1}^{num\_struc}\left(I_{i,j}\right) & if\ I\ is\ NS_{val}, POCmod \\ \min_{j=1}^{num\_struc}\left(I_{i,j}\right) & if\ I\ is\ AIC, YIC, ARIL \end{cases} \quad \text{Eq 40}$$

One should bear in mind that this weighting approach is originally proposed for the case I = AIC, which is based on a formal frame, contrary to the case of BIC, or YIC*mod*. However, the simplicity of the approach proposed by Eq 38 is appealing for establishing a semi-formal hypothesis test in which an indicator $I_{i,j} > I_{i,\neq j}$ is declared as the winner for a given rainfall event i if $p(I_{i,j} > I_{i,\neq j}) > 0.95$ (Eq 38), for a significance level of 95 %. Indeed, this method can be only considered as a formal statistical test for the case of I = AIC, evaluating the null hypothesis that $\Delta AIC_{i,j} > 0$, based on the premise that the $\Delta AIC_{i,j}$ statistics follows a chi-squared X_2 distribution. Accordingly, an analytical expression for calculating the p-value can be obtained (Aho *et al.*, 2014). However, this approach is valid exclusively for cases of normally distributed residuals and nested models, which is also problematic for the case study as RC is not nested with TFs and an analogue equation for BIC, YIC or YIC*mod* is not direct. For the case of NS, testing for statistical significance of the victory of an event i can be done by verifying the null hypothesis of $\Delta NS_i > 0$ by a t-test (or Wilcoxon), depending on the normality of ΔNS_i , which for our case is unknown (adapted from Fay *et al.*, 2010). However, the test proposed in Eq 38 is retained in regards of keeping a unique decision criterion for identifying the winner model structure for each rainfall event i. With this information one can establish if the number of victories (events) in which $p(I_{:,j} > I_{:,\neq j}) > 0.95$ for a given model structure j is high enough to proclaim it as significantly greater than the number of victories attributable to randomness. For a binomial distribution, this value will be of 48 victories for calibration (255 events) and 23 for verification (110 events) for a 95 % significance level, given the probability of winning by chance for a given structure in an event is 1/num_struc = 1/7. Therefore, the model structure j can be considered to have a significant number of victories if the (number victories j)/(number of calibration events) is greater than 48 (calibration) or 23 (verification); for any indicator I.

6.3 RESULTS AND DISCUSSION

Model identification-calibration

The methodology presented in Figure 37 is applied by testing different inputs $u(t)$: Q, R and R_{corr}. Considering the large amount of information from these results, the findings presented in this section are focused on the flow rate Q as the input $u(t)$. The results for the R and R_{corr} inputs are similar, and therefore they are further described for the discussion of the results at the end of the section. The preliminary analysis in Figure 37c is applied for the calibration dataset (255 events) by undertaking individual calibrations for each i rainfall event, delivering an estimation of $p(\theta/load)_i$ and $\theta_{opt\ i}$ for $i = 1 : 255$ for each model structure $j = 1$: *num_str*, leading to estimate $AIC_{i,j}$, $BIC_{i,j}$, $YICmod_{i,j}$ and $NS_{i,j}$ (Figure 37 and Figure 39).

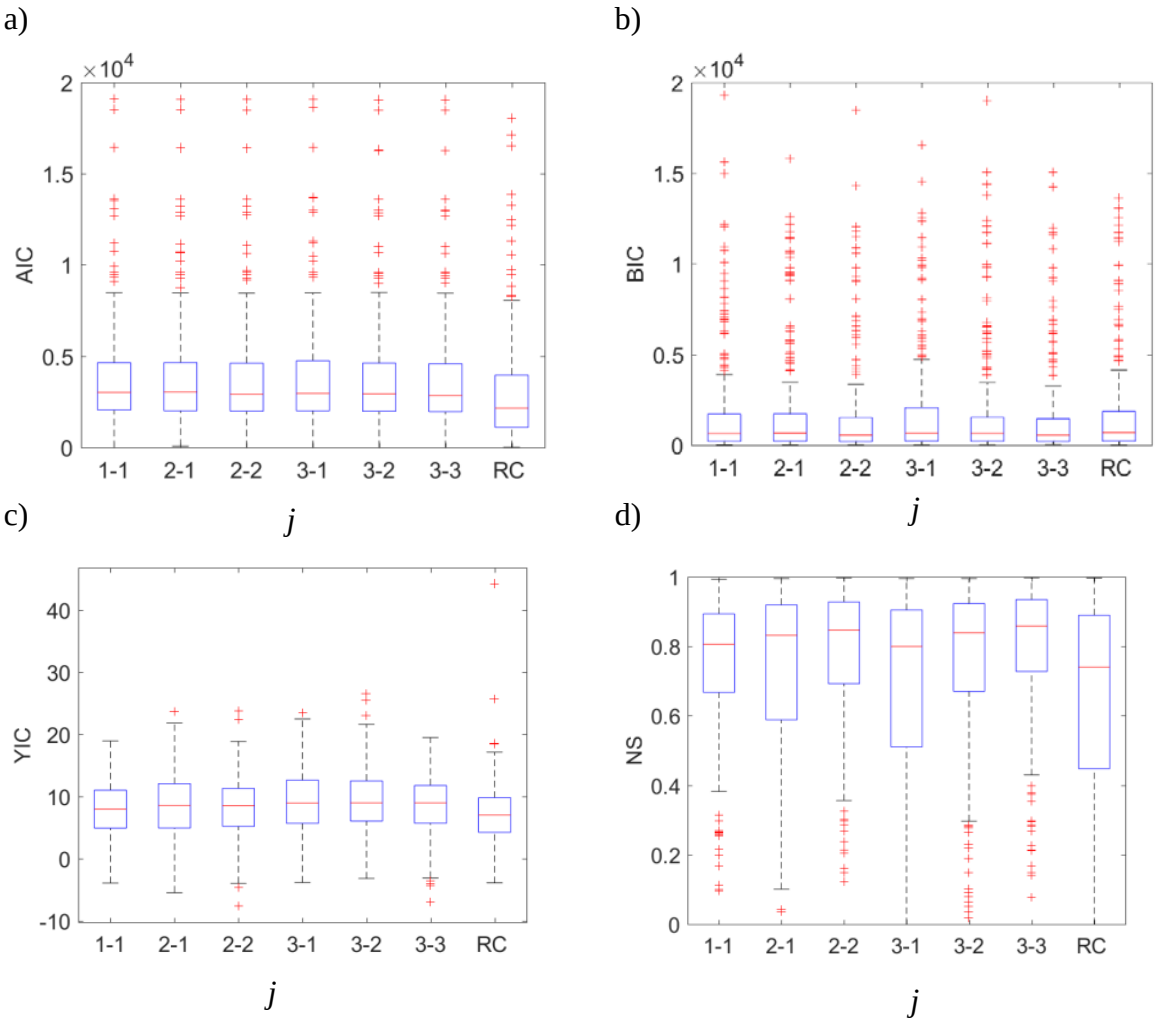


Figure 39. $AIC_{i,j}$, b) $BIC_{i,j}$, c) $YICmod_{i,j}$ and d) $NS_{i,j}$, for each $i = 1 : 255$ calibration event and model structures $j = 1$: *num_str*

From a visual exploration of the $I_{i,j}$ values, the $AIC_{:,7}$ and $YICmod_{:,7}$ groups seem to be lower than the others (Figure 39a, c). However, this result should be confirmed by statistical analyses (S1 and S2 approaches). The $NS_{i,j}$ values for the studied TFs are in average NS > 0.8, reporting lower values for the RC model (Figure 39d).

Nevertheless, the mean value of NS > 0.6 for RC can be considered as acceptable in calibration, especially considering that this model structure has only two parameters. However, in order to verify from a more formal approach if any of the j= 1: *num_str* model structures is significantly more performant in calibration, in terms of $I_{i,j}$, the S1 and S2 approaches are undertaken under the calibration context. The NS in calibration is considered as a complementary indicator, in order to verify that the proposed model structures are appropriate candidates. The NS is not included for being analysed under the S1 and S2 approaches with calibration data, as a significantly higher NS value does not imply that a model is a more appropriate selection than another, as NS high values can be a result of over-parametrization.

Approach S1 for calibration: AIC, BIC and YICmod.

The first step for S1 is to verify the normality of the $I_{:,j}$ (j = 1 : *num_str*) groups, in order to define the most appropriate statistical test to look for significant differences between mean($I_{:,1}$) = mean($I_{:,2}$) = mean($I_{:,3}$) = ... = mean($I_{:,j}$) = ...= mean($I_{:,num\_struc}$) (ANOVA or Kruskal-Wallis test). The results of the Shapiro-Wilk test for this purpose with calibration rainfall events i =1 : 255 are presented in Table 7.

Table 7. p-values of Shapiro-Wilk test for groups AIC$_{:,j}$, BIC$_{:,j}$ and YIC*mod*$_{:,j}$ from calibration events in model structures j = 1: *num_str* (p-value >0.05, hypothesis of normality accepted at a significance level of 95 %, values in light gray).

	MODEL STRUCTURE						
INDICATORS	**(1-1)**	**(2-1)**	**(2-2)**	**(3-1)**	**(3-2)**	**(3-3)**	**RC**
AIC	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BIC	0.00	0.00	0.00	0.00	0.00	0.00	0.00
YIC*mod*	0.46	0.85	0.40	0.38	0.79	0.04	0.00

The hypothesis of normality is rejected at a significance level of 95 % (Shapiro-Wilk test p-value < 0.05) for at least one of the model structures $I_{:,j=1:num\_str}$ in all the three indicators AIC, BIC and YIC*mod* (Table 7). Therefore, the Kruskal-Wallis test is applied for verifying the null hypotheses: mean($I_{:,i}$) = mean($I_{:,j}$), for i and j = 1:*num_str* (Table 7). shows the p-values obtained from the Kruskal-Wallis test for each comparison of mean($I_{:,i}$) = mean($I_{:,j}$), where the p-value < 0.05 are shown in black, meaning that mean($I_{:,i}$) ≠ mean($I_{:,j}$), for a 95 % significance level.

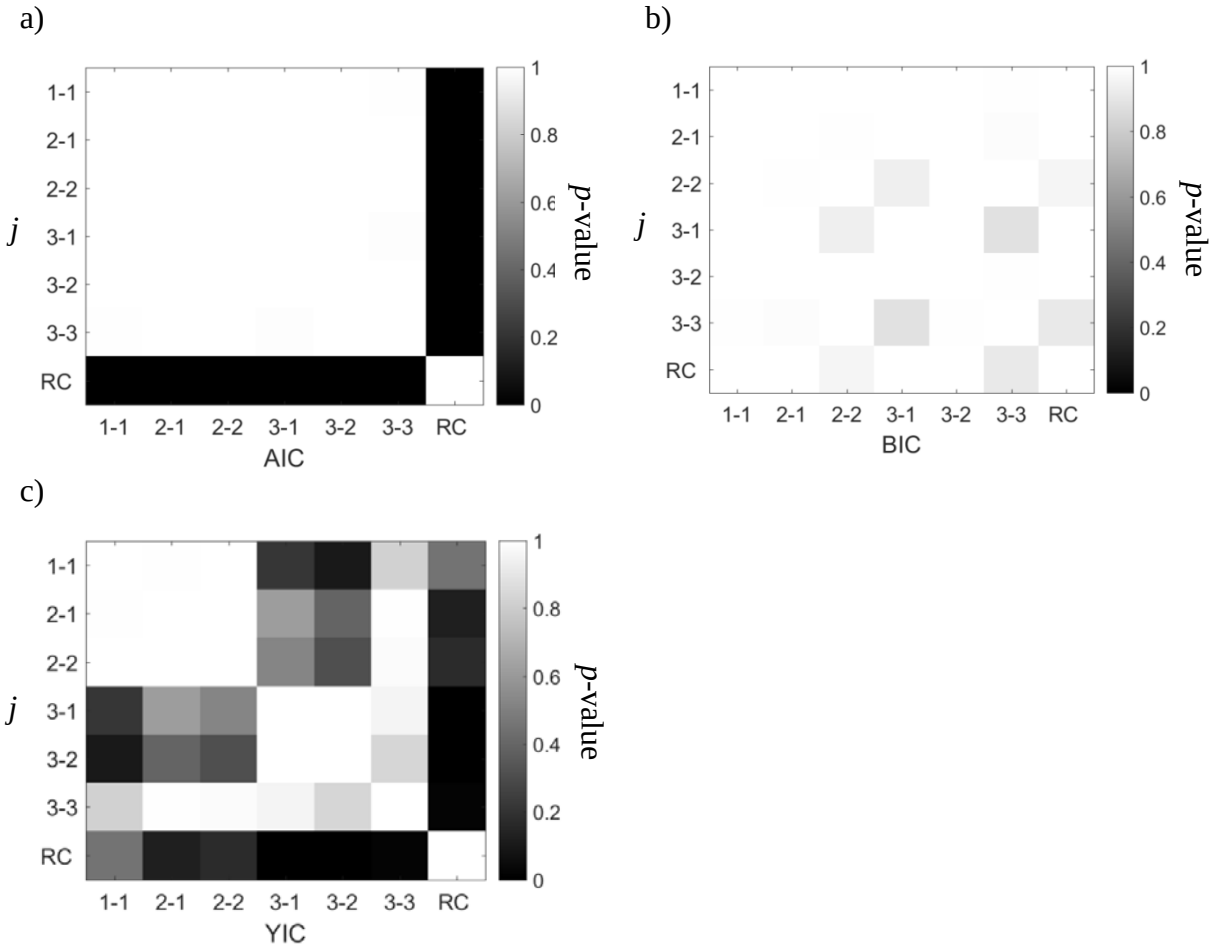


Figure 40. *p*-values of Kruskal-wallis test for comparisons of groups mean($I_{:,i}$) = mean($I_{:,j}$), for all j = 1:*num_str* in calibration, with I: a) AIC$_{:,j}$, b) BIC$_{:,j}$ and c) YIC*mod*$_{:,j}$ (*p*-value < 0.05 hypothesis of equal means rejected at a significance level of 95 %, comparisons in black).

Kruskal-Wallis test reported strong enough statistical evidence (*p*-value < 0.05) that the AIC and YIC*mod* indicators for the RC model are significantly lower than for all the other TF model structures with the calibration dataset (black rows and columns in Figure 40a, c). This finding strengths the complementariness of the proposed indicators (AIC, BIC and YIC*mod*), as *e.g.* analyses are not conclusive regarding the BIC (Figure 40b). One can conclude from this S1 approach in the calibration stage that RC is a promising model structure, which offers a suitable balance between performance (mean NS > 0.6 in Figure 39d) and the number of parameters and their parametric uncertainties (AIC and YIC*mod*). No particular conclusions can be drawn from the results among the different TFs.

Approach S2 for calibration: AIC, BIC and YICmod.

S2 is proposed as an alternative approach to S1 aimed to generalize the model selection based on an indicator I given the inter-event variability. For the calibration events, the following results are obtained regarding AIC, BIC and YIC*mod* indicators (Figure 41).

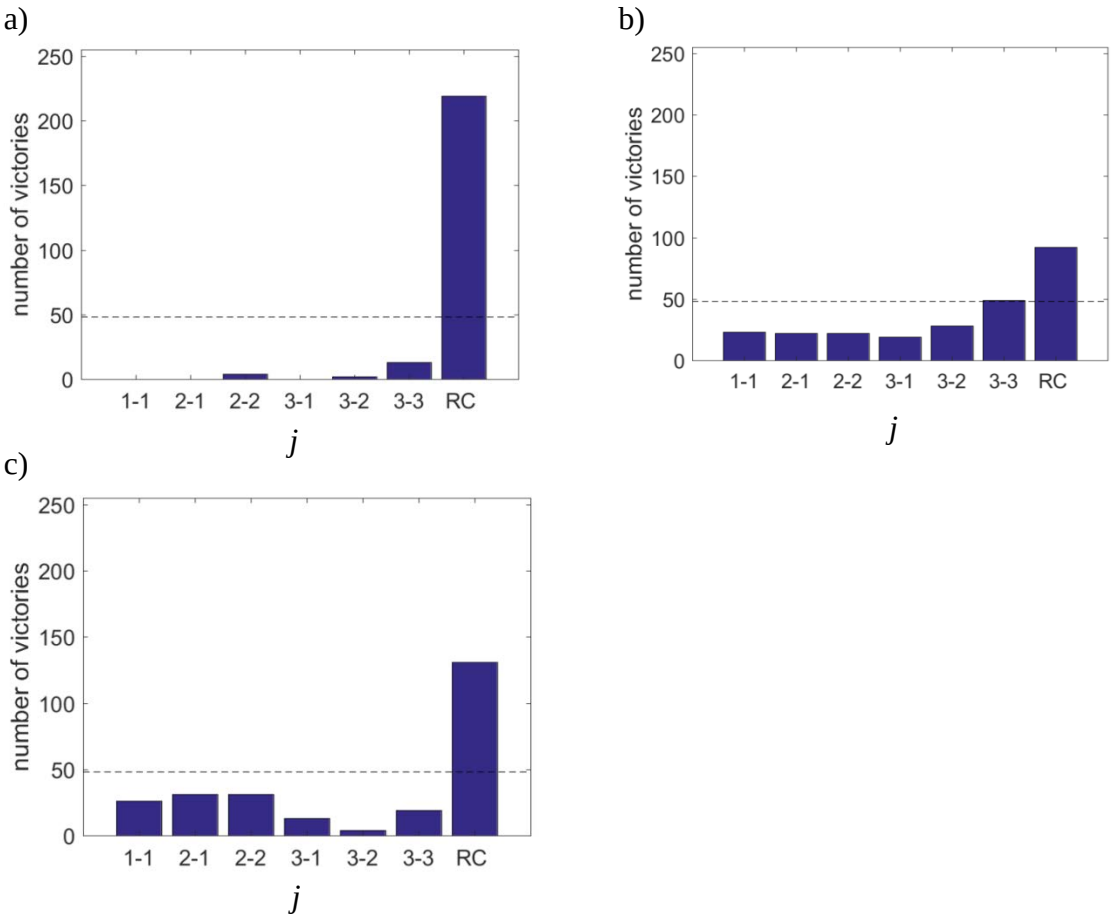


Figure 41. number of victories for model structure and group $I_{:,i}$ a) $AIC_{i,j}$, b) $BIC_{i,j}$ and c) $YICmod_{i,j}$ for i = 1 : 255 calibration events and the model structures j = 1 : *num_str* with 48 victories as the threshold for a significance level of 95 %

The S2 analysis confirmed the promising results obtained for the RC model structure by the S1 approach for the calibration dataset. This model is found to be the winner more times (220) than what can be attributed to randomness regarding the AIC and YIC*mod* indicators (48 victories, for a significance level of 95 %, see Eq 38). Therefore, from calibration analyses, the RC model can be a reasonable recommendation in case that verification tests are not available. The verification is undertaken with the i = 256 : 365 events in further analyses.

Model identification- verification

For any of the j = 1 : *num_str* model structures, the marginal $p(\theta/load)$ function could not be represented by *num_c* > 1 conditional functions $p(\theta/load, Tnum\_c)$, as no clear association between characteristics of inputs $u(t)_{obs}$ (*e.g.* max., min., mean, volume, duration, ADWP) and the different $p(\theta/load, Tnum\_c)$ functions could be established (methodology from Chapter 4). Therefore, the global estimation of $p(\theta/load)$ is proposed to be calculated as the marginal distribution sum of representative local estimations of $p(\theta/load)_i$ ($\theta_{opt\ i}$ estimations with NS < 0.8 are discarded as unrepresentative) and θ_{opt} = mean($p(\theta/load)$) for all models structures (Figure 37d).

For the case of verification, different indicators *I* are proposed than for the calibration stage. As the NS is not susceptible to over-parametrization with verification data, it is included as another comparative indicator *I*, jointly with parametric and total ARIL and POC*mod*, *i.e.* ARIL*tot*$_{i,j}$, POC*mod tot*$_{i,j}$, ARIL*par*$_{i,j}$, POC*mod par*$_{i,j}$, and NS$_{i,j}$, (calculations in Figure 42).

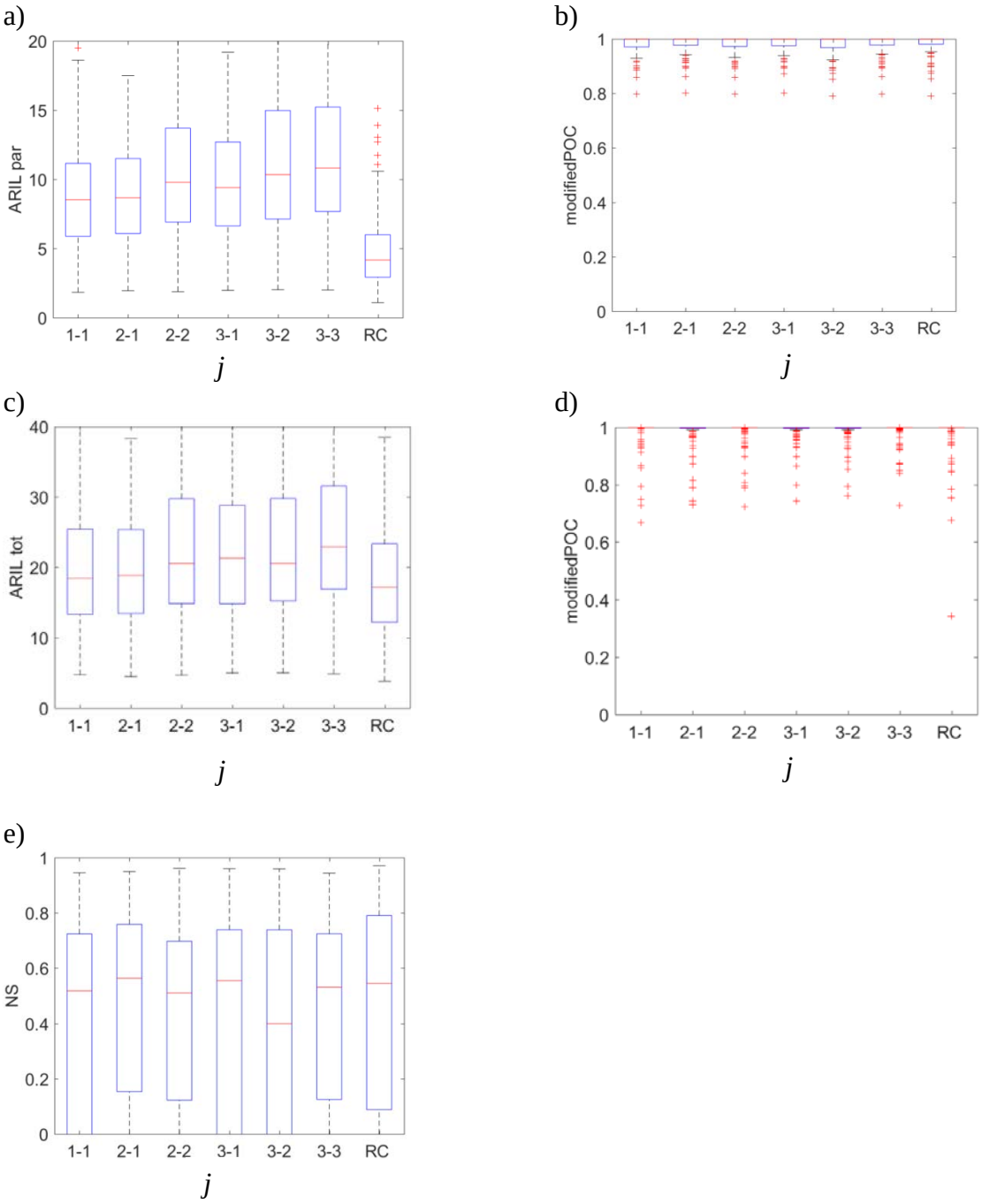


Figure 42. a) ARIL*par*$_{i,j}$, b) POC*mod par*$_{i,j}$, c) ARIL*tot*$_{i,j}$, d) POC*mod tot*$_{i,j}$ and e) NS$_{i,j}$, for each $i = 256 : 365$ verification event and model structure $j = 1$: *num_str*

From a visual exploration of the $I_{i,j}$ values, the ARIL*par*$_{:,7}$ and ARIL*tot*$_{:,7}$ (RC model) groups seem to be lower than for the other model structures (Figure 42a, c). The POC*mod par* and POC*mod tot* are nearly one in all cases (Figure 42b, d), given the size of the ARIL values. The

$NS_{i,:}$ (for all model structures) values are in average about 0.6, which might be considered as still unsatisfactory for the verification stage. However, in order to evaluate if any of the model structures is performing significantly better, the S1 and S2 approaches are undertaken with verification events.

Approach S1 for verification: ARILpar, POCmod par, ARILtot, POCtot and NS.

Results of the Shapiro-Wilk test with verification events i = 256 : 365 are presented in Table 8.

Table 8. *p*-values of Shapiro-Wilk test for groups ARIL$par_{i,j}$, POC$mod\ par_{i,j}$, ARIL$tot_{i,j}$, POC$mod\ tot_{i,j}$ and $NS_{i,j}$, from verification events in model structures j = 1: *num_str* (*p*-value >0.05, hypothesis of normality accepted at a significance level of 95 %).

		MODEL STRUCTURE						
		(1-1)	**(2-1)**	**(2-2)**	**(3-1)**	**(3-2)**	**(3-3)**	**RC**
INDICATORS	**ARIL *par***	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	ARIL *tot*	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	POC*mod par*	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	POC*mod tot*	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	NS	0.00	0.00	0.00	0.00	0.00	0.00	0.00

The hypothesis of normality is rejected at a significance level of 95 % (Shapiro-Wilk test *p*-value < 0.05) for all the model structures and in all the verification indicators (Table 8). Therefore, the Kruskal-Wallis test is applied for verifying the null hypotheses: mean($I_{:,i}$) = mean($I_{:,j}$), for i and j = 1:*num_str* (Figure 43).

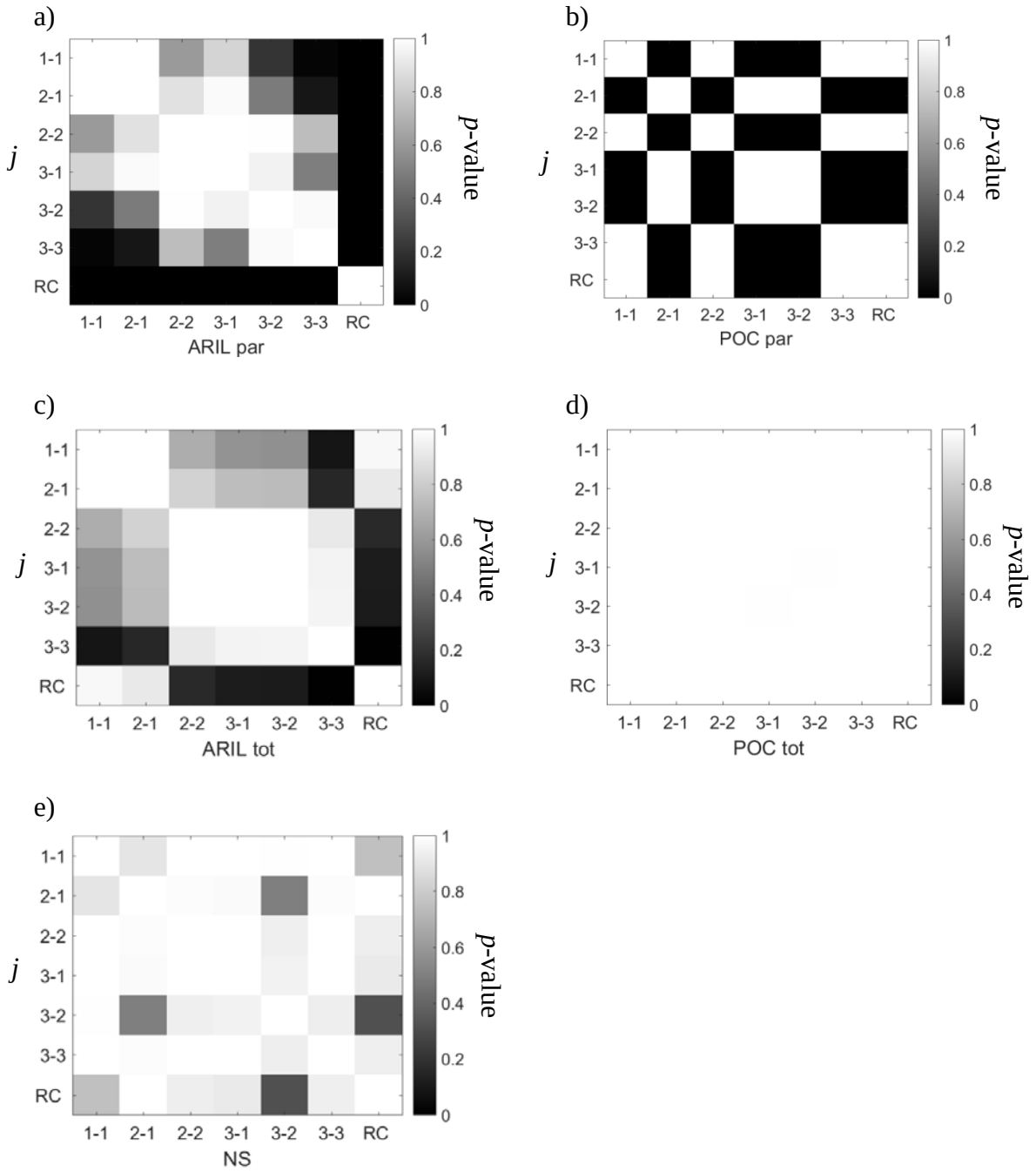


Figure 43. *p*-values of Kruskal-Wallis test for comparisons of groups mean($I_{:,i}$) = mean($I_{:,j}$), for all $j = 1 : num\_str$ in verification, with I: a) ARIL$par_{i,j}$, b) POC$mod\ par_{i,j}$, c) ARIL$tot_{i,j}$, d) POC$mod\ tot_{i,j}$ and e) $NS_{i,j}$, (*p*-value < 0.05 hypothesis of equal means rejected at a significance level of 95 %, comparisons in black).

Kruskal-Wallis confirms for a significance level of 95 % (*p*-value < 0.05) that the ARIL*par* and ARIL*tot* indicators for the RC model are significantly lower than for most of the TFs in verification (comparisons in black in Figure 43a, c). The NS values for model structure $TF_{3,2}$ are significantly lower than for all other cases in verification. Furthermore, the NS for RC (2 parameters) is significantly higher than $TF_{0,0}$ (1 parameter) (*p*-value <0.05). Indeed, no conclusive information is obtained from the POC*mod par* and POC*mod tot* indicators, due to the magnitude of the ARIL values (Figure 42a, c). One can also conclude from this S1 approach in the verification stage that the RC performed significantly better than the TFs, regarding the size of the uncertainty intervals obtained in simulation. No particular

conclusions can be drawn from the results regarding comparisons of the NS values, as the performances for all model structures are very similar (NS about 0.6) (Figure 42e and Figure 43e).

Approach S2 for verification: ARILpar, POCmod par, ARILtot, POCtot and NS.

For the verification events, the following results are obtained with S2 regarding ARIL*par*, POC*mod par*, ARIL*tot*, POC*mod tot* and NS indicators (Figure 44).

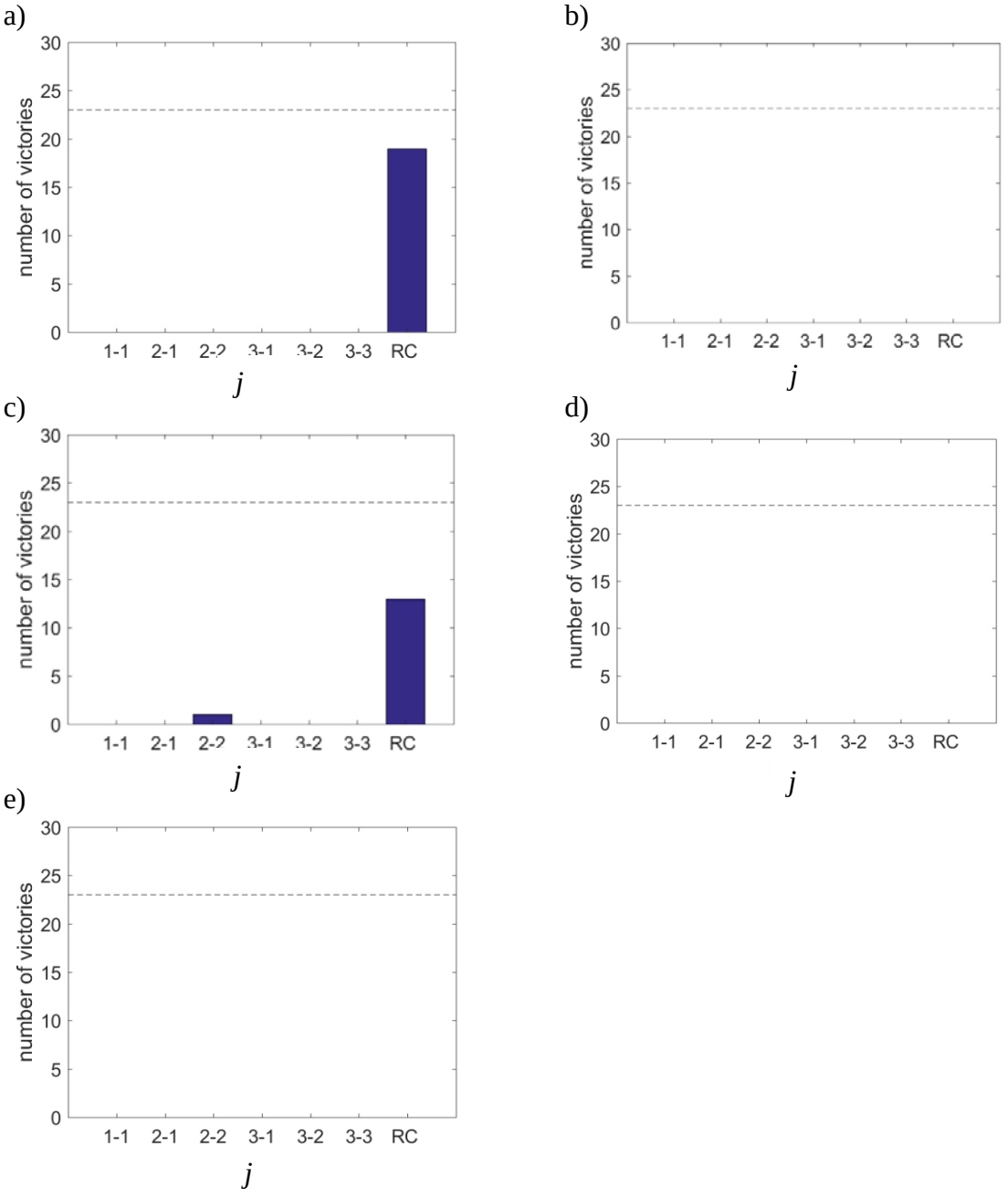


Figure 44. number of victories for model structure and group *I.*, a) ARIL$par_{i,j}$, b) POC*mod* $par_{i,j}$, c) ARIL$tot_{i,j}$, d) POC*mod* $tot_{i,j}$ and e) $NS_{i,j}$, for each $i = 256 : 365$ verification event and the model structures $j = 1$: *num_str* with 23 victories as the threshold for a significance level of 95 %

The S2 approach does not bring evidence towards recommending a best model structure from the analyzed cases in the verification data. Although the ARIL*par* and ARIL*tot* are the only

indicators in which any model structure reported significant victories (14 and 18 events for the case of RC in Figure 44a and Figure 44c, this number of victories can be attributed to randomness (23 for significance level of 95 %, from a binomial distribution).

Calibration analyses indicated that the RC model structure is a reasonable recommendation from a parsimony perspective (avoiding over-parametrization, with only 2 parameters r and M_0), as AIC and YIC*mod* indicators are significantly lower than for the TFs by the S1 and S2 approaches, keeping an average $NS_{:,j} > 0.6$. This preliminary analysis is recommended in the HI-DBM framework, especially when verification data is insufficient, unrepresentative or computationally demanding. In the verification stage, the S1 approach reported significantly lower ARIL*par* and ARIL*tot* values for the RC model than for the TFs. Furthermore, the NS for RC (2 parameters) is significantly higher than $TF_{0,0}$ (1 parameter) (p-value < 0.05). No conclusive results are obtained in terms of reliability (POC*mod* indicators), given the magnitude of the ARIL values.

A potential benefit from modelling the TSS load with TFs, *i.e.*, in form of physically interpretable linear differential equations, is that information from previous time-steps (measured Q and/or simulated TSS load) is explicitly included in the calculations. However, the benefit of this approach remains unproved regarding the obtained results, as the performance indicators of the RC model are superior to the tested TFs, for cases in which significant conclusions could be drawn (AIC and YIC*mod* in calibration and ARIL in verification). The benefits of RC compared to TFs might be explained from the non-linear properties of RC, as the implicit relation between Q and TSS load seems to be governed by a non-linear power law (results presented in Chapter 3 Figure 16, comparable to Daly *et al.*, 2014 and Sun *et al.*, 2015). The TFs approach might be also restrictive for non-linear processes due to their linear nature. This can be also supported in the similarities of the averages of NS = 0.6 in verification for RC and TFs, which might be a result of the inherent non-linear power law relation between TSS and Q.

The selection of the RC model structure as the most suitable option among the proposed models might by tempting, given the exposed reasons. However, the selection of a single model structure for cases in which different indicators I suggest a different "best" option is more challenging. For those cases, the following approaches can be recommended: (i) modeler's principal interest (*e.g.* accuracy from NS, precision from ARIL or reliability from POC*mod*), (ii) weighting strategies among the indicators I (Marshall *et al.*, 2007), multi-objective analysis (Huang and Liu, 2010) with further statistical analysis (*e.g.* Pareto multi-objective solutions, see *e.g.* Ye *et al.*, 2014).

On the other hand, the RC model structure by itself can be still considered as an unsatisfactory model, given that NS < 0.6 for about 50 % of the verification rainfall events and ARIL*par* > 2.5 about 75 % of the verification events. On the other hand, the generally deficient performances of RC and TFs (TCP descriptions) can be also an evidence of the lack of a potential M process (TVP description). Therefore, the RC is retained for Chapter 7 as an appropriate enough description of W, towards reconstructing a potentially missed M process (TVP) by this TCP description. These results of M can be compared to the traditional accumulation/wash-off idea in Chapter 7.

The results obtained when the input $u(t)$ is equal to the flow rate Q are analogous for the R and R_{corr} cases. The main difference is that all indicators from the verification stage (*i.e.* NS, ARIL and POC*mod*) present lower performances (*e.g.* mean NS = 0.4), although conserving similarities in terms of relative comparisons and the best model selection. This uncertainty

reduction by including the flow rate in TSS calculations has been previously reported (Sikorska *et al.*, 2015), contrary to previous studies that claim the rainfall as a better predictor of TSS than *Q* (*e.g.* Vaze and Chiew, 2002; McCarthy et al., 2012; Manz *et al.*, 2013). On the other hand, more performant models with *Q* as input might be expected when the predicted variable is the TSS load, as these values are directly calculated from the flow rate and the TSS concentrations. The rainfall correction applied in Chapter 5 to events with high measurement errors in rainfall R_{corr}, led to improve the results compared to the use of the rainfall *R* as the predictor in RC and TFs model. Nonetheless, R_{corr} did not show better performances than when *Q* is used as the predictor. This can be explained by the fact that R_{corr} is directly obtained from *Q* (see details in Chapter 5), and therefore no information seems to be added by using R_{corr} as a predictor instead of *Q*. Coupling a model by detailing rainfall and flow rate as two different inputs may be promising (*e.g.* Mannina and Vivianni, 2010; Métadier, 2011; Hong *et al.*, 2016), when having enough information about the load separation in one component produced by rainfall and the other one from flow rate (surface and in the sewer), in order to avoid identifiability problems (Bonhomme and Petrucci, 2017). No relation between rainfall events with high rainfall errors (from Chapter 4) and bad performances of TSS models is found: only 42 % of the events highly influenced by rainfall errors (from Chapter 4) have also non-reproducible (NS < 0.8) TSS loads (by *Q*, *R* or R_{corr} as input). This finding is contradictory to Manz *et al.* (2013), who found a relation between rainfall errors and TSS model performances, although from a different experimental setting and conceptual approach.

The rainfall events for which RC reported unsatisfactory results (NS < 0.8) are neither reproducible by TFs, for the vast majority of cases (in calibration and verification). In addition, a PCA is undertaken, aimed to identify potential relations between the representability of a given event by a TCP model (RC or TFs) (events with NS < 0.8 red points and NS > 0.8 grey points) and physical characteristics of the event (max flow rate (m^3/s), mean flow rate (m^3/s), ADWP (days), beginning of the event (days)) (scaled by a *z* normalization, see Kreyszig, 1979) (Figure 45).

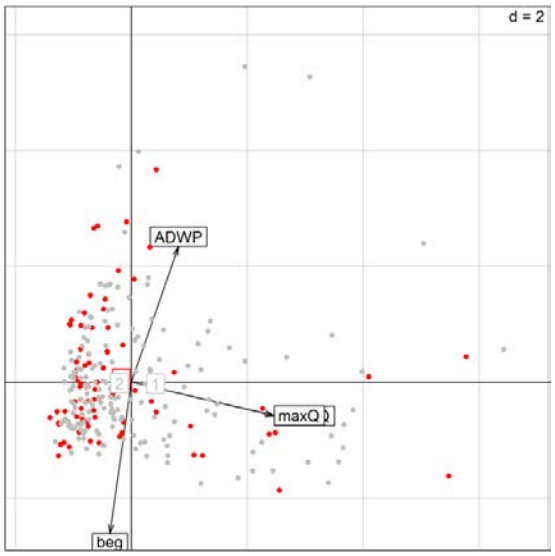


Figure 45. PCA analysis with explanatory variables: max flow rate, mean flow rate, beginning of the event and ADWP, for the differences between calibration events with NS < 0.8 (red points) and NS > 0.8 (grey points), for the RC model.

The results in Figure 45 might lead to conclude that, if the unsatisfactory performances of the RC model (events with NS < 0.8 in red) are due to the lack of a potential deterministic process, this process is not dependent on the ADWP, as assumed by the vast majority of accumulation/wash off models (no separation between red and grey events) (Figure 45). Furthermore, a Wald Wolfowitz test (Wald and Wolfowitz, 1943) suggested that the distribution of these "red" or "grey" events in time, as a time series vector of binary states, is random (p-value < 0.05).

6.4 CONCLUSIONS

Different linear Transfer Functions (TF) models are tested as alternative descriptions to the non-linear traditional Rating Curve (RC) model, aimed to scrutinize for better representations of the stormwater TSS load dynamics as a function of flow rate or rainfall, without considering a virtual mass over the catchment decreasing process in time. The benefits of using flow rate or rainfall as inputs in the models are also discussed. The advantages of implementing the TFs for this "omitted virtual mass approach" compared to RC models remains unproved regarding the obtained results, as the performance indicators of the RC model are superior to those of the tested TFs, for cases in which significant conclusions could be drawn (especially parsimony in calibration and precision of the simulations). This can be explained from the non-linear properties of RC, as the implicit relation between flow rate and TSS load seems to be governed by a power law (Chapter 3).

The rainfall correction applied in Chapter 5 to events with high measurement errors in rainfall improves the results compared to the use of the rainfall as the predictor in RC and TFs model. However, rainfall corrected according Chapter 5 does not show better performances than when the flow rate is used as the predictor. This can be explained by the fact that the rainfall corrections are directly estimated from the flow rate (see details Chapter 5), without adding information by using corrected rainfall as a predictor instead of the flow rate. No relation between rainfall events with high rainfall errors (from Chapter 4) and bad performances of TSS models is found: only 42 % of the events highly influenced by rainfall errors (in Chapter 4) have also non-reproducible TSS loads by the flow rate or rainfall as input. Furthermore, a comparison between RC and TFs models is developed when the flow rate, rainfall or mean areal rainfall estimations from Chapter 5 are used as inputs. The main difference is a lower modelling performance when rainfall is used as input rather than flow rate (from mean NS of 0.6 to 0.4 in verification events). On the other hand, the RC by itself can be still considered as an unsatisfactory model (*e.g.* NS about 0.6 in verification with flow rate as input), suggesting the lack of an essential missing process in this model. The unsatisfactory performances of the RC model are found to be independent of the ADWP, suggesting that a potential missed process in the RC model (if there is one) is not necessarily linked to temporality. Complementarily, statistical tests strengthen that the occurrence of these non-reproducible events by the RC model shows a random distribution in time.

CHAPTER 7. REVISITING CONCEPTUAL STORMWATER QUALITY MODELS BY RECONSTRUCTING VIRTUAL STATE-VARIABLES

Extended version of:

Sandoval S., Vezzaro L., Bertrand-Krajewski J.-L. (2017). Revisiting conceptual stormwater quality models by reconstructing virtual state-variables. *Proceedings of the 14th International Conference on Urban Drainage*, Prague, Czech Republic, 10-15 September, 3 p.

7.1 INTRODUCTION AND BACKGROUND

The RC model calibrated in Chapter 6 and selected from comparisons with multiple TFs (TCP models) can be directly linked with any traditional TVP formulation of M, by making M_0 in the RC model (Eq 25) a time variable parameter (TVP). Indeed, one can ask for the existence of another deterministic global process missed by the RC model, essential to represent the pollutant loads, which is oversimplified or misinterpreted by the accumulation/wash-off idea.

For this purpose, Time Variable Parameters (or virtual processes) (TVP) concept has been introduced in the hydrological and environmental context as a powerful statistical model-based approach to describe unobserved processes or state variables (see *e.g.* DBM applications). The idea of TVP is to deliver a reasonable reconstruction, under certain hypotheses, of how one or multiple parameters of a mathematical model might vary in time, in order to make the output of the model to match the observed data. Different TVP estimation techniques are cited in the literature, where the time variations in the parameters can be assumed *e.g.* stochastic processes (Random Walk or Integrated Random Walk) (Pedregal *et al.*, 2007; Young, 2012). However, most of these approaches are developed for linear models, without a wide flexibility in the hypotheses about the error model for the TVP estimation. From this perspective, Bayesian calibrations have emerged as a promising model-based approach for reconstructing unmeasured inputs or state variables, adaptable to non-linear and complex model structures, including flexibility in the error model of the reconstructed state variable (*e.g.* Sun and Bertrand-Krajewski, 2013a; Leonhardt et al., 2014).

Therefore, this Chapter proposes to undertake Bayesian reconstructions of the "virtual" state variable M by modifying the RC traditional model in Eq 25, with r as a calibration parameter and replacing M_0 by a TVP $\widehat{M}(t)$ (kg) (formulation F1). This $\widehat{M}(t)$ reconstruction can be directly compared to a time constant or variable traditional M formulation (*e.g.* Eq 24). As a complementarily analysis, a formulation F2 is explored, in which the RC model is modified with $\hat{r}(t)$ (-) as the TVP to be reconstructed and M_0 as a calibration parameter. The proposed approach led to revisit the accumulation/wash-off idea by using Bayesian TVP reconstructions and the RC model, scrutinizing for evidence of a deterministic global process and its interpretability as a "mean" state of available pollution washed by rainfall. Furthermore, this accumulation/wash-off model structure is sought to be reformulated by analysing the inter-event repeatability of the reconstructed state variables.

7.2 MATERIALS AND METHODS

Bayesian reconstruction of virtual state variables by TVPs

The reconstruction of a virtual state variable by Time Variable Parameters (TVP) consists in solving a calibration problem where a TVP is represented as an additional time series of parameters. In principle, every time-step of a TVP can be considered as an independent parameter in the inference scheme, with a length equal to the length of the output series load. However, the dimensionality of this problem will be massive, risking over-parametrization and delivering incorrect estimations (Vrugt *et al.*, 2009). To avoid over-parametrization, the TVP time series has usually a coarser temporal resolution than the output data. Therefore, a time window strategy is adopted, reducing the dimensionality of the problem by dividing the TVP reconstruction into equally spaced time windows (see *e.g.* Sun and Bertrand-Krajewski, 2013a). Further test with non-equally spaced time windows (see *e.g.* Chapter 5) reported analogue results.

This TVP can be estimated jointly with the other set of calibration parameters θ (r for F1 and M_0 for F2) by means of a Bayesian inference scheme. Therefore, θ and *TVP* are defined as random variables, where their joint posterior probability density function (PDF) is calculated by Eq 41 as $p(\theta, TVP/ Q, load)$, given the input $Q(t)$ and output $load_{obs}(t)$ data with some prior knowledge about θ and *TVP* (from BATEA in Kuczera *et al.*, 2006, see application Kavetski *et al.*, 2006a). $p(\theta, TVP/ Q, load)$ is a posterior probabilistic characterization of θ and *TVP* ($\widehat{M}(t)$ for F1 or $\hat{r}(t)$ for F2), in which the values with the maximum likelihood are assumed be the "optimal" parameters θ_{opt} and TVP_{opt} ($\widehat{M}(t)_{opt}$ for F1 or $\hat{r}(t)_{opt}$ for F2).

$$p(\theta, TVP/Q, load) \propto \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\hat{\sigma}load_t^2}} exp\left[-\frac{1}{2}\frac{(load_{sim}(Q(t),\theta,TVP) - load_{obs}(t))^2}{\hat{\sigma}load_t^2}\right] \cdot P(\theta) \cdot \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\hat{\sigma}TVP^2}} exp\left[-\frac{1}{2}\frac{Var(TVP)}{\hat{\sigma}TVP^2}\right] P(\text{TVP}) \quad \text{Eq 41}$$

where n is the total observed load values in $load_{obs}(t)$. $load_{sim}(Q(t)_i, \theta, TVP)$ is the simulated load by the input flow rate series $Q(t)$ and a set of θ and *TVP*. $p(\theta)$ and $p(TVP)$ represents a prior belief about the probability that a candidate set of θ and TVP values are "true" (assumed as a non-informative uniform distribution for all cases).

Eq 41 allows to explicitly separate the model error of *TVP* (second Pi product) from the error model of the output *load* (first Pi product). With the purpose of finding a TVP estimation "as constant as possible" (and therefore less informative), the error model of *TVP* (second Pi product) is assumed to be proportional to TVP's own variance $Var(TVP)$. Both error models, for the *load* and TVP estimations (first and second Pi product resp.), are assumed to be independent and normally distributed, with the error variances $\hat{\sigma}load_t^2$ and $\hat{\sigma}TVP^2$, respectively. $\hat{\sigma}load_t^2$ is considered heteroscedastic, being equal to the square of the standard uncertainty of each observed value $load_{obs}(t)$ (*e.g.* Sun and Bertrand-Krajewski, 2013a). On the other hand, $\hat{\sigma}TVP^2$ is considered to be homoscedastic and is estimated as another parameter in the set θ (*e.g.* Sage *et al.*, 2015), expressed as $\hat{\sigma}M^2$ (kg) for F1 and $\hat{\sigma}r^2$ (-) for F2. The same Eq 41 is used for calibrating the RC model, by omitting the second Pi product term, delivering $p(\theta/Q, load)$ as a posterior probabilistic characterization of θ (with θ_{opt} also for the "optimal" parameters) (results from Chapter 6).

The DREAM algorithm (Vrugt, 2016) is applied to solve Eq 41 and to evaluate the three formulations (the traditional RC, the F1 and F2 formulations) for the different θ and *TVP* variants (Table 9). The logarithmic form of Eq 41 is implemented to ensure numerical stability and a max. number of simulations of $6x10^5$ is used with a max. of 30 parallel Markov Chains to reach convergence (Gelman and Rubin RB convergence criteria > 1.2, see Gelman and Rubin, 1992). The TVPs reconstruction is undertaken with a resolution of 12 time windows (*i.e.* equivalent to 12 parameters, see Table 9), balancing between the convergence of the algorithm (RB > 1.2) and capturing the global dynamics of the TVPs (see application in Sun and Bertrand-Krajewski, 2013a). Table 9 summarizes the parameters min. and max. values search ranges in the prior uniform distributions $p(\theta)$ and $p(TVP)$ in Eq 41. These ranges are defined for r and M_0 from the literature (Kanso *et al.*, 2005) and are equally adopted for each window of $\widehat{M}(t)$ and $\hat{r}(t)$. The max. of the error variances $\hat{\sigma}M^2$ and $\hat{\sigma}r^2$ are defined, respectively, as 4 times the standard deviation of M_0 and r in the RC calibration (Chapter 6).

Table 9. DREAM solving of Eq 41 for virtual state variables reconstruction formulations, specifying the set of parameters θ and TVP, the total number of parameters and the min. and max. search range in the prior distributions p(θ) and p(*TVP*).

Formulation	**Parameters θ**	**Parameters in TVP**	**Number of parameters θ + TVP**	**Min/max search range values**
RC model (Chapter 6)	r (-), M_o (kg)	-	2	r [0, 5] ; M_o[0 , 8e5]
F1	r (-), $\hat{\sigma}M^2$ (kg)	$\widehat{M}(t)$ (kg)	2 + 12 = 14	r [0, 5] ; $\widehat{M}(t)$ [0 , 8e5] for each t ; $\hat{\sigma}M^2$ [0 , 2.4e5]
F2	M_0 (kg), $\hat{\sigma}r^2$ (-)	$\hat{r}(t)$ (-)	2 + 12 = 14	M_0 [0 , 8e5] ; $\hat{r}(t)$ [0 , 5] for each t ; $\hat{\sigma}r^2$ [0 , 1.5]

The parameters are estimated for each individual rainfall event (event-based calibration). This eliminated the need for a "dry build up" model (see *e.g.* Freni *et al.*, 2009; Chow *et al.*, 2015) as the initial sediment mass (M_0) is estimated for each event.

For each *i-th* calibration event, the parameter estimation results (posterior probability and optimal parameter sets) provide the basis for the model evaluation. This is performed by looking at: (i) the Nash-Sutcliffe efficiency (NS_i) between simulated ($load_{sim}(t)_i$) and observed ($load_{obs}(t)_i$) loads, besides the well-posedness and identifiability of the Bayesian inference $p(\theta, TVP/Q, load)_i$ (*intra-event identifiability*); (ii) similarities in the shape or dynamics of $TVP_{opt\,i}$ with estimations for other events (*inter-event identifiability from repeatability*); (iii) the capacity of a given set of $\theta_{opt\,i}$ and $TVP_{opt\,i}$ to represent another rainfall event measured by NS_i (*inter-event transferability*, see Bardossy and Singh, 2008) and (iv) formulate further hypotheses about a potential missing process based on physical knowledge about the system and the obtained results (*interpretability*).

7.3 RESULTS AND DISCUSSION

With illustrative purposes, the calculations related to the formulations proposed in Table 9 are illustrated for an example rainfall event i = 16 in Figure 46 and Figure 47 (event from 23/09/2004 22:00 to 24/09/2004 07:00). The experimental data is used with $Q(t)$ hydrograph as the input (Figure 46a) for representing the TSS $load(t)$ pollutograph (Figure 46b) (black solid lines with 95 % coverage bands in grey). The TSS load simulated by the calibrated local set of parameters $\theta_{opt\ 16}$ for the RC model is compared to results from F1 and F2 formulations (with $\theta_{opt\ 16}$ and $TVP_{opt\ 16}$) in Figure 46b (green, blue and red solid lines resp.).

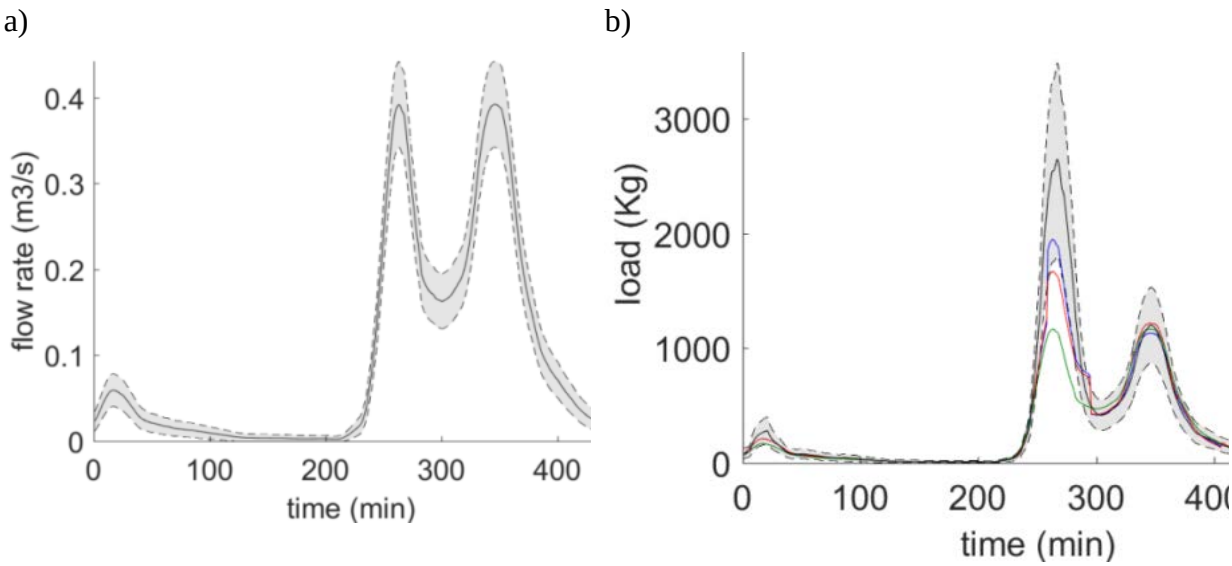


Figure 46. a) measured Q hydrograph (solid black) and b) measured TSS load pollutograph (solid black) including 95 % coverage intervals (grey bands), including simulations with $\theta_{opt\ 16}$ *and* $TVP_{opt\ 16}$ for RC, F1 and F2 (solid green, blue and red lines resp.).

For the formulations F1 and F2 in Figure 47 (in blue and red resp.), the estimations of TVP_{16} ($\widehat{M}(t)_{16}$ or $\hat{r}(t)_{16}$) are represented with $\widehat{M}(t)_{opt\ 16}$ for F1 (blue solid line) and $\hat{r}(t)_{opt\ 16}$ for F2 (red solid line), jointly with their blue and red coloured 95 % coverage intervals (Figure 47a and Figure 47b, resp.). Estimations of the θ_{16}, *i.e.*, r(-); $\hat{\sigma}M^2{}_{16}$ (kg) (for F1) and $M_{0\ 16}$ (kg); $\hat{\sigma}r^2$(-) (for F2) are shown in two correlation plots (blue in Figure 47c and red Figure 47d, resp.).

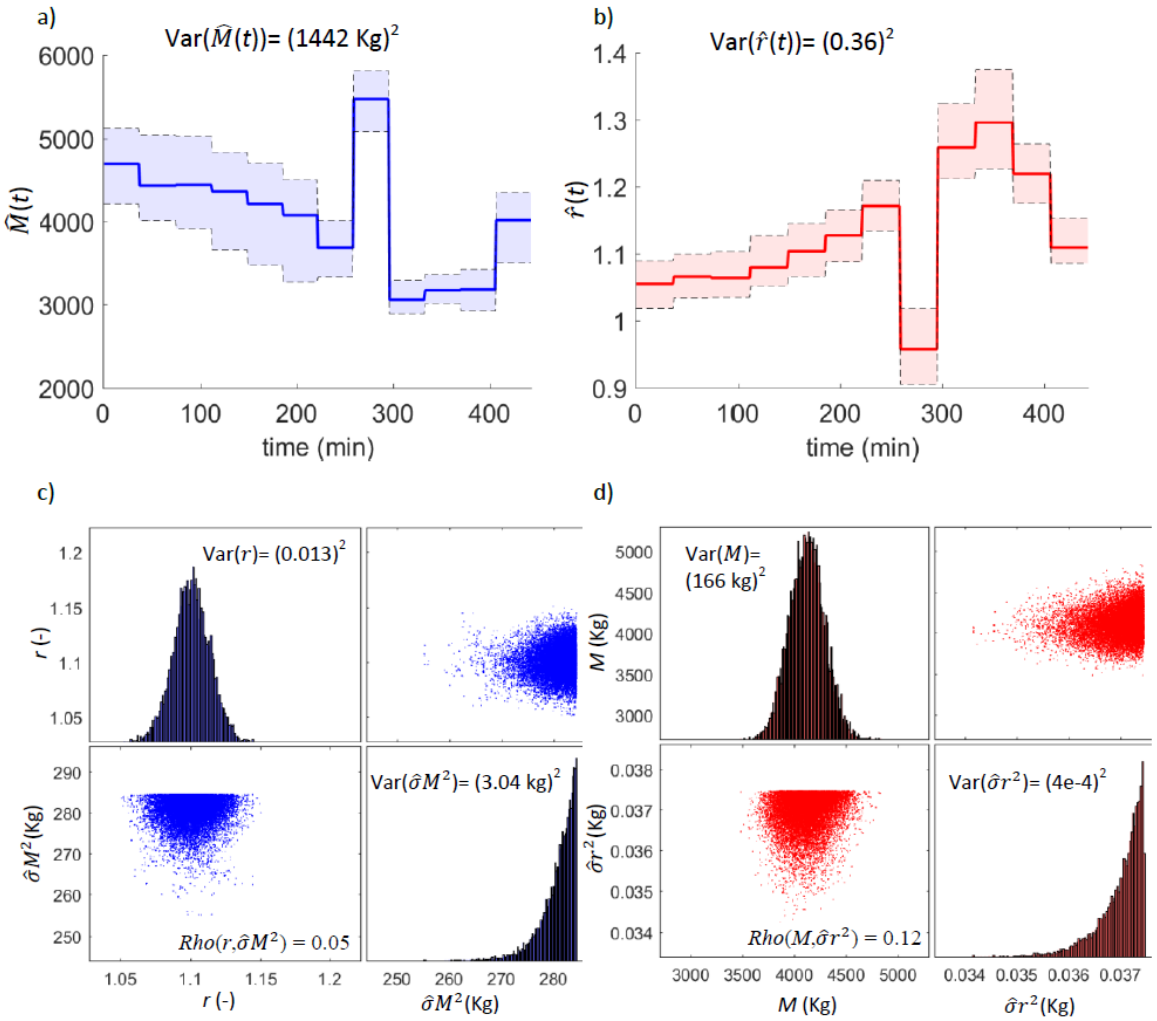


Figure 47. Results from event 16, $TVP_{opt\ 16}$ reconstructed time series (solid line) with 95 % coverage intervals (colored bands). a) $\hat{M}(t)_{16}$ for F1(blue). b) $\hat{r}(t)_{16}$ for F2 (red). The correlation matrix of θ_{16} for c) r_{16} (-); $\hat{\sigma}M^2{}_{16}$ (kg) (F1 blue) and d) $M_{0\,16}$ (kg); $\hat{\sigma}r^2$ (-) (F2 red).

Intra-event Identifiability

Although there is an intrinsic correlation between RC model parameters (therefore for F1 and F2) due to the mathematical structure of the model (Kanso *et al.*, 2005), the intra-event identifiability regarding θ and *TVP* gives promising results. The RC model local calibrations report unsatisfactory adjustments between the simulated and measured loads, with $NS_i < 0.8$ for 142 of the 255 events (56 %) (from Chapter 6).The F1 or F2 formulations achieve greater NS values in all the events, increasing the values of $NS_i > 0.8$ for 60 % of cases in which RC reported $NS_i < 0.8$. In the example shown in Figure 46, the NS for the RC model (green) is 0.65, while the F1 (blue) and F2 (red) formulations show NS above 0.8. This analysis encourages the applicability of the studied model structures with local estimations of $\theta_{opt\ i}$ and $TVP_{opt\ i}$, supporting the reasoning behind the F1 and F2 reconstructions as potential processes unrepresented by RC.

However, these preliminary results should be analysed under a transferability perspective, as this improvement in NS values can be simply a numerical effect resulting from increasing the number of parameters in F1 or F2 formulations. For the case of $TVP_{opt\ i}$, undesired correlations between the measured $load_{obs}(t)_i$ and the TVP parameters $\hat{M}(t)_{opt\ i}$ or $\hat{r}(t)_{opt\ i}$ are

above 0.6 and 0.5 (respectively) for only 25 % of the events. This result brings evidence that the $TVP_{opt\ i}$ reconstructions are contributing with additional information (also as NS values are higher than for RC), without mimicking the measured *load* dynamic. Further tests with the CAPTAIN Toobox in Matlab (Pedregal *et al.*, 2007) for TVP estimations, using comparable model structures (TFs), reported unsatisfactorily high correlations with the TSS load. The Reverse Modelling is another technique to reconstruct virtual state variables recommended to be merged with Bayesian estimations in Chapter 5, under a hydrological context. However, Reverse Modelling was found as an unfeasible reconstruction technique by using the RC model, given also the high correlations between the reconstructed TVPs with *Q* and TSS load. The strength of the proposed Bayesian method for reconstruction of TVPs relies in the flexibility regarding the likelihood function, including an appropriate error term of the TVPs, towards a "non-informative" reconstruction (second Pi product in Eq 41).

For the case of θ parameters, the couples r_i; $\hat{\sigma}M^2{}_i$ (for F1) and $M_{0\ i}$; $\hat{\sigma}r^2{}_i$ (for F2) exhibited an appropriate parametric identifiability in terms of their PDFs unimodality, with spurious parametric correlations $Rho(r_i;\ \hat{\sigma}M^2{}_i) < 0.3$ (F1) and $Rho(M_{0\ i};\ \hat{\sigma}r^2{}_i) < 0.11$ (F2), in 90 % of the $i = 1 : 255$ events (see example event: $Rho(r_{16};\ \hat{\sigma}M^2{}_{16}) = 0.05$ in Figure 47c and $Rho(M_{0\ 16};\ \hat{\sigma}r^2{}_{16}) = 0.12$ in Figure 47d). Although the PDFs of the error variances ($\hat{\sigma}M^2{}_{i\ =\ 1:255}$ (F1) and $\hat{\sigma}r^2{}_{i\ =\ 1:255}$ (F2)) are sharper close to the max. value of the search ranges given in Table 9 (see example event $\hat{\sigma}M^2{}_{16}$ and $\hat{\sigma}r^2{}_{16}$ in Figure 47c and Figure 47d), the results are retained, as higher max. limits tended to ill-posedness in the inferences.

Inter-event identifiability from repeatability

A functional clustering by *k*-centre method is applied to identify groups of the TVP_{opt} time-varying curves with similar shapes (Chiou and Li, 2007, see environmental applications of functional data classification in *e.g.* Ternynck *et al.*, 2016). A number of *k* groups is defined in order to visualize *k* different potential "repeated" behaviours in the set of optimal TVP_{opt} curves. For interpretability, each of the $TVP_{opt\ i}$ curves is standardized by the transformation $z(TVP_{opt\ i})$ with zero mean and unitary standard deviation (Kreyszig, 1979). For the F1 formulation, applying the Chiou and Li (2007) functional method with *e.g.* $k = 2$ groups allows separating the $\widehat{M}(t)_{opt}$ into two "similar" clusters, shown in Figure 48a (light and dark blue groups of curves, corresponding to about 57 % and 43 % of the events). The mean curves of each group are shown in the figure on the right, along with the corresponding 95 % coverage intervals.

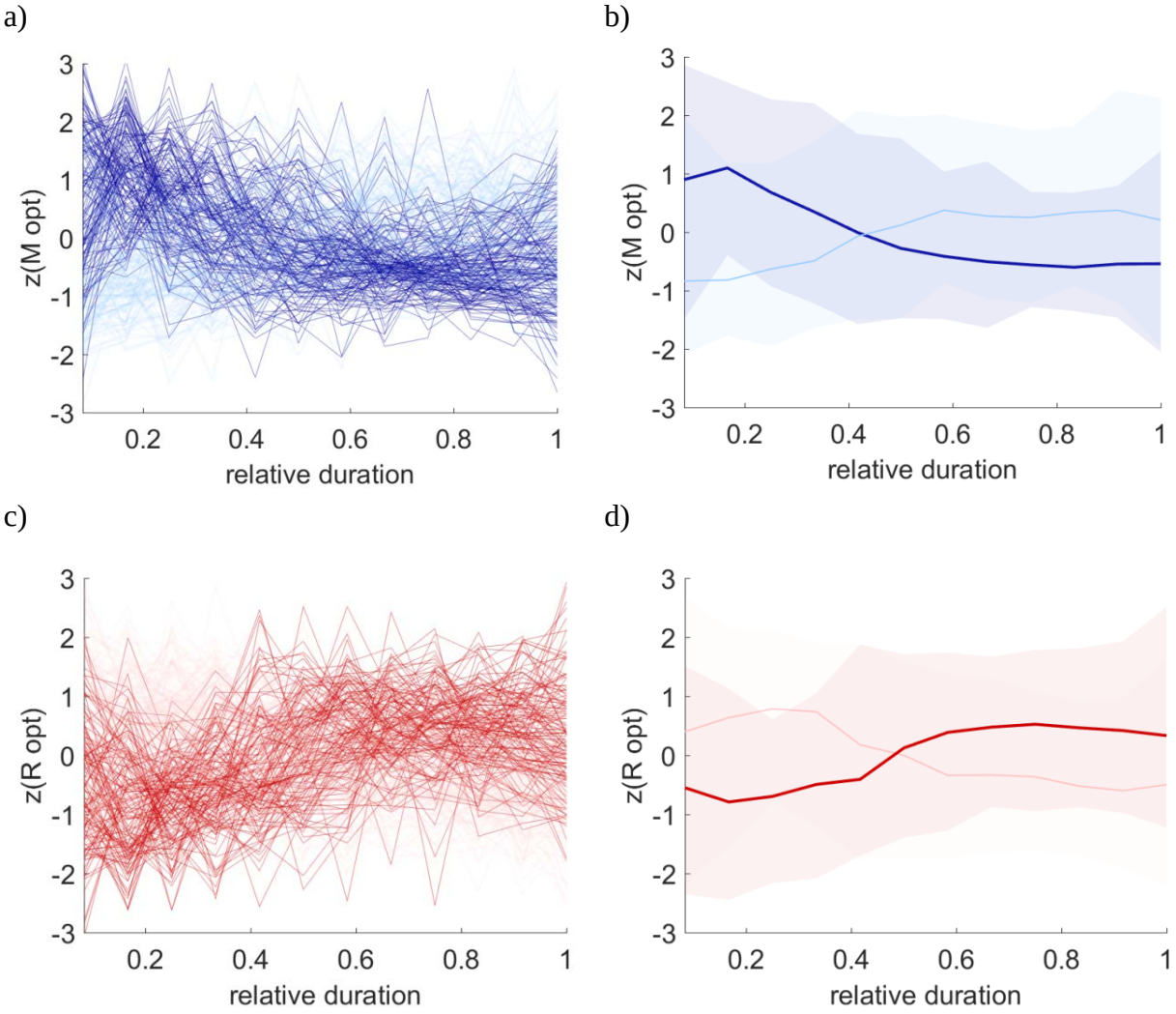


Figure 48. Curves of TVP parameters (a) $\widehat{M}(t)_{opt\,i}$ (blue) (c) $\hat{r}(t)_{opt\,i}$ (red) grouped by k = 2 clusters (light and dark colours groups). Mean curve and 95 % confidence intervals for (b) $\widehat{M}(t)_{opt\,i}$ (blue) and (d) $\hat{r}(t)_{opt\,i}$ (red) (light and dark colours groups).

The traditional interpretation of M can be preliminarily associated to the dark blue group, as the mean curve shows an apparent decaying trend (Figure 48b). However, this hypothesis would lack of an immediate physical interpretation for the remaining 57 % events (light blue group). Furthermore, the trends of the mean curves for the dark blue group (negative trend) and light blue group (positive trend) show to be statistically insignificant, due to the strong variability of the curves inside each light or dark blue group. This effect of randomness is stronger when the variability of $\widehat{M}(t)_i$ (estimated by calculating the 95 % coverage intervals from the posterior distributions) is considered. Similar conclusions are drawn for clustering the $\widehat{M}(t)_{opt\,i}$ curves into more groups ($k > 2$), and for the F2 formulation (Figure 48c and Figure 48d). These results reveal the difficulty in identifying or characterizing a unique virtual process potentially missed by the RC regarding an inter-event scale, given the lack of repeatability of the shapes of the $TVP_{opt\,i}$ curves.

Inter-event transferability

The transferability of TVPs is analysed based on the methodologies proposed in Chapter 4, investigating how a given $TVP_{opt\,i}$ time series is able to reproduce another event from the dataset. Results for $\widehat{M}(t)$ or $\hat{r}(t)$ are analogue, therefore discussion focuses on $\widehat{M}(t)_{opt,i}$ estimations. Each of the 29 most transferable estimations of $\widehat{M}(t)_{opt\,i}$ is able to explain at least 30 rainfall events (NS > 0.8). On the other hand, for the "optimal" local estimations $\theta_{opt\,i}$ of RC, 60 estimations are able to explain at least 30 events (NS > 0.8). The flatter curves from $\widehat{M}(t)_{opt\ i\ =\ 1:255}$ tend to be more transferable, as they resemble the constant values in the RC formulation. These results stress the low transferability of the potential missing processes (F1 or F2) to further rainfall events.

Interpretability

These results bring evidence of a potential missing process in the RC model. Although both processes ($\widehat{M}(t)$ or $\hat{r}(t)$) are good candidates to explain RC obstacles from an intra-event analysis, there is no evidence that F1 or F2 is a more valid approach than the other. Indeed, the F1 or F2 formulations show the same explanatory capacity, in the sense that none of them performs better. This highlights the challenge in terms of identifiability and unicity of a potential process missed by the RC model structure, which is hardly identifiable from an inter-event analysis. The low repeatability of the reconstructed TVP curves might suggest as more adaptable a random description of a potential missing essential process by RC models, rather than a deterministic interpretation regarding an inter-event scale.

7.4 CONCLUSIONS

This work suggests the missing representation of an essential process in the traditional Rating Curve model based on the observations from 255 rain events (from Chapter 6). The results indicate that the high unrepeatability of this missing process makes it hardly interpretable in terms of a virtual unique state of available TSS mass in the catchment that is decreasing over time, as assumed by a great number of traditional models. This Chapter shows how high-time resolution quality measurements can provide a support to revisit and question existing models, and to potentially allow developing new stormwater quality model formulations.

GENERAL CONCLUSIONS PART 3

Chapter 6 and Chapter 7 are aimed to revisit the traditional approach of accumulation wash-off models to represent the stormwater TSS loads pollutographs at the outlet of large urban catchments (> 100 ha), based on different concepts from the HI-DBM framework (Young, 2013), the Bayesian methods applied for reconstruction of virtual state variables presented in Chapter 5 (for mean areal rainfall estimation) and the calibration frame in Chapter 4.

The first stage of Part 3 (Chapter 6) explored different linear Transfer Functions (TFs) models as a possible alternative to the traditional Rating Curve (RC) models, considering the case in which a virtual mass over the catchment decreasing process is omitted. The expected advantages of implementing TFs compared to the RC model are not verified, as the performance indicators of the RC model are superior to the tested TFs (especially parsimony in calibration and precision of the simulations). This can be explained from the non-linear properties of RC, as the implicit relation between flow rate and TSS load seems to be governed by a power law (described in Chapter 3).

The benefits of using flow rate or rainfall as inputs in the RC model are also discussed. No relation between rainfall events with high rainfall errors (from Chapter 4) and bad performances of TSS models is found: only 42 % of the events highly influenced by rainfall errors (in Chapter 4) have also non-reproducible TSS loads by the RC model (NS < 0.8). Furthermore, a comparison between RC and TFs models is developed when the flow rate, rainfall or mean areal rainfall estimations from Chapter 5 are used as inputs. The main difference relies in a lower modelling performance when rainfall is used as input rather than flow rate (mean NS 0.6 to 0.4 in verification events). On the other hand, the RC model by itself can be still considered as an unsatisfactory model (*e.g.* NS about 0.6 in verification with flow rate as input), suggesting the lack of an essential missing process in this model. Furthermore, the unsatisfactory performances of the RC model are found to be independent of the ADWP, suggesting that this potential missed process is not necessarily linked to temporality, as assumed by the majority of accumulation/wash-off models. Complementarily, statistical tests strengthen that the occurrence of these not representable events by the RC model is randomly distributed in time.

Chapter 6 suggests the missing representation of an essential process in the RC model based on the observations from 365 rain events, without further scrutinizing about how the shape or dynamics of this process could be. Therefore, Chapter 7 applies the Bayesian rainfall reconstruction method explored in Chapter 5 as a more general estimator of virtual state variables, which is used for undertaking an intra-event reconstruction of this potential missing process in 255 rainfall events. The results of Chapter 7 indicate that these potential missing processes, although identifiable at the intra-event scale, are hardly interpretable in terms of a unique state of a virtual available mass over the catchment that is decreasing over time, as assumed by a great number of traditional models. Furthermore, the reconstructed processes are highly unrepeatable regarding their shape, besides having a low transferability to other rainfall events (transferability notions established in Chapter 4). This study shows how high-time resolution quality measurements can provide a support to revisit and question existing models, and to potentially allow developing new stormwater quality model formulations.

GENERAL CONCLUSIONS AND PERSPECTIVES

The measurement of Total Suspended Solids (TSS) in urban drainage systems is often required for scientific, legal, environmental and operational reasons, as particulate matter constitutes a major source of surface water contamination (Ashley *et al.*, 2004; Chebbo and Gromaire, 2004). However, the reliability of TSS measurements strongly depends on the quality of the collected samples, which should be representative of real field conditions in the monitored sewer pipe (Larrarte, 2008; Métadier and Bertrand-Krajewski, 2012). Therefore, appropriate data acquisition and validation methodologies for TSS measurements in urban drainage systems are required (Bertrand-Krajewski and Muste, 2007). The role of hydrology and rainfall measurements is recognized in TSS modelling, as rainfall is the driving process in the contamination of receiving water bodies by stormwater (Lee *et al.*, 2002). Rainfall data and hydrological models are also known to be embedded with high uncertainties, impacting the performance of TSS stormwater models. Apart from uncertainties in rainfall and TSS data, the TSS stormwater traditional model structures have been widely questioned at the scale of large urban catchments, especially when reproducing data from online measurements (*e.g.* Métadier, 2011; Dotto *et al.*, 2011). The manuscript is mainly developed with online flow rate, TSS and rainfall measurements from the Chassieu urban catchment (Lyon, France), monitored by the OTHU project (Field Observatory for Urban Hydrology - www.othu.org). Information includes 365 rainfall events measured between 2004 and 2011 (see a detailed description of the catchment in Chapter 1).

Different hypotheses about why TSS models are still unsatisfactory have been explored, motivated by the following scientific questions:

- Are TSS online continuous time series reliable and useful for modelling purposes?

- Do this time series show bias or insufficient representativeness?

- How to better calibrate rainfall-runoff models if model parameters are event-dependent?

- If rainfall-runoff models are not satisfactory, could we assume that this is mainly due to errors in rainfall measurements and can we identify/correct them?

- Are traditional TSS models appropriate when they are used with online continuous TSS time series instead of traditional samples?

- Is there an event-dependent relation between rainfall errors and deficient performances of TSS models?

- How could we revisit/improve TSS traditional models?

In Chapter 2, different sampling strategies during rainfall events are simulated and evaluated by means of online TSS and flow rate measurements. The average relative sampling error and the residuals distribution are estimated from Event Mean Concentrations (EMCs) simulated by the studied strategies, compared to EMCs obtained by the complete time series from online monitoring of various rainfall events. For the Chassieu data set, the strategy with the sampling volume proportional to runoff volume between two samples, with constant sampling intervals, delivers the most representative results in terms of accuracy, precision and uncertainties propagation. Recommended sampling time intervals are of 5 min, with average sampling

errors between 7 % and 20 % and uncertainties in sampling errors of about 5 %, depending on the sampling interval. These results lead to hypothesize about the potential errors in TSS data, highlighting that data obtained from monitoring campaigns (sampling strategies) could still be used as an input for intra-event TSS models in the absence of online measurements. However, potential errors or biases in the EMCs by means of modelling strategies such as bias correction factors are recommended to be accounted for in the calibration process.

In Chapter 3, the probability of underestimating the cross section mean *TSS* concentration is estimated to be approximately 0.88 or 0.64 for all the flow velocity values, by two proposed methodologies: the Simplified method (SM) and the Time Series Method (TSM). TSM shows more realistic TSS underestimations (about 39 %) than the SM (about 269 %). Differences between the two methods are mainly due to simplifications in SM (absence of TSS measurements and operation of the sampling system). SM can estimate the sampling depth at which the probability of over estimation is equal to the probability of underestimation (about 5 % of the total depth, with the proposed hypotheses). SM assumes a significant asymmetry of the TSS concentration profile along the vertical axis in the cross section. This is compatible with the distribution of TSS measurements found in TSM. Furthermore, a power law describing the TSS as a function of flow rate is revealed, including higher variances of TSS for higher flow rates. Results from Chapter 2 and Chapter 3 provide insights towards an indicator of the measurement performance and representativeness for a TSS sampling protocol.

In Chapter 4, the nature of model structure uncertainty and the inter-event parametric variability is addressed for a conceptual rainfall-runoff model based on the idea of assessing the parameters marginal probability function (obtained by event-by-event calibrations) into conditional probability functions (obtained by grouping the parameters from the event-by-event calibrations). The results stress the importance of carefully selecting the data to be used for parameter estimations and further hydrological simulation, considering that the proposed parameter estimation strategy significantly improves the results in terms of accuracy and precision in verification, compared to the parameter estimation strategies based on event-by-event and multi-event calibrations. The Nash-Sutcliffe criterion (NS) obtained by the proposed parameter estimation strategy is improved from 0.4 to 0.6, for 50 % of the verification rainfall events, compared to traditional calibration strategies. One single rainfall-runoff model structure allows representing two groups of different hydrological conditions for an urban catchment by the proposed strategy, leading also to identify (i) bimodalities in the parameters and (ii) rainfall events with high errors in rainfall measurement. Chapter 5 presents a methodology aimed to estimate mean areal rainfall, based on a hydrological model and flow rate data. This model-based approach demonstrates the advantages, compared to previous approaches in the literature, of correcting rainfall by multiplying factors over constant-length time window and rainfall zero records are filled with a reverse model.

Chapter 6 explores different linear Transfer Functions (TF) models alternative to the traditional Rating Curve (RC) models, considering the case in which a virtual mass over the catchment decreasing process is omitted. From flow rate, rainfall and corrected rainfall tests as different potential inputs, no relation between rainfall errors or hydrological conditions with performances of RC and TFs could be established. The expected advantages of implementing the TFs compared to the RC model are not verified, as the performance indicators of the RC model are superior to the tested TFs (especially parsimony in calibration and precision of the simulations). This can be explained from the non-linear properties of RC, as the implicit relation between flow rate and TSS load seems to be governed by a power law (described in Chapter 3). This Chapter 6 suggests the missing representation of an essential

process in the RC model, based on its performance (NS of about 0.6 in verification, with 110 events). The unsatisfactory performances of RC model are found to be independent of the antecedent dry weather period, suggesting that a potential missed process in the RC model (if there is one) is not necessarily linked to temporality. Complementarily, statistical tests strengthen that the occurrence of these not representable events by the RC model is randomly distributed in time. These results are used as an input for Chapter 7, defining the RC as a candidate model to be improved or revisited by defining its parameters as virtual state variables. The advantages of selecting RC among a wider family of models (given by different TFs), besides mentioned aspects regarding its performance, is the potential interpretability of RC time varying parameters as a decreasing virtual available mass.

Therefore, Chapter 7 applies the proposed Bayesian rainfall correction method explored in Chapter 5 as a more general Bayesian reconstruction method of virtual state variables with the RC model. Results indicate that these potential missing processes in RC description are hardly interpretable in terms of a unique state of virtual available mass over the catchment that is decreasing over time, as assumed by a great number of traditional models. Furthermore, the reconstructed processes are highly unrepeatable regarding their shape, besides having a low transferability to other rainfall events. This manuscript shows how high-time resolution quality measurements can provide a support to revisit and question existing models, and to potentially allow developing new stormwater quality model formulations for large urban catchments.

Probably the first recommendation that can be formulated from the work presented in this manuscript is to implement the proposed methodologies in further urban catchments with similar data, aimed to compare and provide more generality to the conclusions herein presented. The results from Chapter 2 can be validated by comparing the EMCs obtained from real autosamplers operated *in situ* with different sampling strategies, to the EMCs obtained from TSS online monitoring. Moreover, TSS models aimed to represent the complete pollutograph can be calibrated with data obtained by different sampling strategies and compared to model calibrations with online monitoring. These explorations might lead to formulate complementary modelling recommendations for cases in which TSS online data is not available.

TSS measurements in different points of the cross section of the sewer system (*e.g.* Larrarte, 2008) can be recommended in order to further compare and validate the methodologies proposed in Chapter 3. These methods could be extended, by similar theoretical assumptions, to assess the uncertainties not only due to the vertical but also due to the transversal sampling intake position in the cross section of the sewer system or other field conditions. Furthermore, the conception of new measurement devices could be envisaged for measuring online the TSS 3D profile of the cross section in the sewer system. This information would be very useful for validating several of the hypotheses proposed in this manuscript, regarding the representativeness of the studied TSS time series. On the other hand, the representativeness of TSS measurements due to further uncertainty sources and field sampling conditions such as the pumping velocity of the sampling tube or the time shift due to the pumping operation (Rossi, 1998) can be assessed in future works. These uncertainty sources can be included with further hypotheses regarding, for example, theoretical pumping velocities and operation. However, mentioned assumptions are recommended to be verified with further experiments in all cases.

The spatial variability of the rainfall and its influence on the performance TSS models can be further explored, by the use of mean areal rainfall estimations that are non-dependent on flow

rate data. These mean areal rainfall estimations can be obtained by using information from other rain gauges, radar data or by microwave links as recently proposed (van het Schip *et al.*, 2017). Furthermore, a continuous approach for representing the inter-event variability instead of a discrete classification of the events can be recommended, given the lack of physical interpretability regarding the boundary between classification groups (*e.g.* highest or lowest rainfall intensity), which at the end are model and data dependent. Nevertheless, the proposed approach could be an input for reformulating the selected model structure, seeking to include physically-based continuous terms to reproduce the parametric variability instead of establishing discrete hydrological conditions, also avoiding the parametric dependency on *e.g.* rainfall characteristics (*e.g.* Chapter 6 and 7; Young, 1998). Extending the modelling calibration methods presented in Chapter 4 by including a better understanding of rainfall uncertainties, jointly with the implementation of further hydrological models, may also lead to improve the performance of rainfall-runoff modelling for the studied catchment. However, the calibration methodology proposed in Chapter 4 in its current state offers enough flexibility to be applied towards a better understanding of the local/global nature of input data/model structure uncertainties for further modelling frames in the urban drainage field. For example, in Chapter 6, the calibration method is applied into the water quality modelling context, leading to identify irreproducible events by different water quality models.

The Bayesian reconstruction method applied to rainfall and TSS mass reconstructions in Chapter 5 and Chapter 7, respectively, may constitute a powerful tool for hypotheses testing in urban drainage models. In principle, this reconstruction method could be applied to formulate a reasonable hypothesis about the dynamics of any input, output or virtual state variable of a mathematical model, by means of the available data. This might open a wide range of possibilities for increasing the understanding of mathematical models commonly used in the urban drainage context, as demonstrated in this manuscript for the case of traditional TSS models.

Coupling a TSS model by including rainfall and flow rate as two different inputs may be promising (*e.g.* Métadier, 2011), when having enough information about the TSS load separation in one component produced by rainfall and the other from flow rate (surface and in sewer processes). Furthermore, exploring physically detailed TSS models might also be a potential alternative (Hong *et al.*, 2016), when extensive granulometric measurements are available, coupled with a better understanding of physical processes inside the sewer processes (*e.g.* TSS resuspension, transfer). Therefore, TSS modelling approaches in their current state of development still represent important challenges, in the absence of data and representative information about the load dynamics inside the system.

From the above lines, many scientific questions can be formulated for incoming investigations. The following examples can be proposed:

- Are the results obtained in the different parts of this investigation generalizable to other urban catchments?

- What is the effect of calibrating and using TSS models with traditional samples data for sites in which online data is not available?

- What is the influence of the sampling intake position in the cross section of the sewer system over the representativeness of TSS measurements from a 3D analysis?

- Can a better rainfall-runoff model structure be proposed for the studied catchment?

- Can the performance of TSS models be improved by considering mean areal rainfall estimations non-dependent in flow rate data as additional modelling inputs?

- What are the benefits in terms of modelling performance when physically detailed TSS models are calibrated by using more detailed data (*e.g.* granulometric measurements, TSS online data recorded at different points inside the urban drainage system)?

- Can we still revisit these physically detailed TSS models by means of more detailed data and the Bayesian reconstruction methods proposed in this manuscript?

This Page Intentionally Left Blank

REFERENCES

Abbaspour, K C., Johnson, C.A., Van Genuchten, M.T., (2004). Estimating uncertain flow and transport parameters using a sequential uncertainty fitting procedure. *Vadose Zone Journal*, 3(4), 1340-1352.

Ackerman, D., Stein, E., and Ritter, K., (2010). Evaluating performance of stormwater sampling approaches using a dynamic watershed model. *Environmental Monitoring and Assessment*. doi:10.1007/s10661-010-1788-6.

Aho, K., Derryberry, D., Peterson, T., (2014), "Model selection for ecologists: the worldviews of AIC and BIC", Ecology, 95: 631–636, doi:10.1890/13-1452.1.

Ahyerre, M., (1999). *Bilan et mécanismes de migration de la pollution organique en réseau d'assainissement unitaire*. PhD Thesis, University Paris 6, France, 250 p. (in French).

Aizerman, M.A., Braverman, E.M., Rozonoer, L.I., (1970). *The Method of Potential Functions in the Theory of Machine Learning*. Nauka, Moscow 6.

Ajmal, M., Waseem, M., Ahn, J.H., Kim, T.W., (2015). Improved runoff estimation using event-based rainfall-runoff models. *Water Resources Management*, 29(6), 1995-2010.

Akaike, H., (1973). Information theory and an extension of the maximum likelihood principle, in second International Symposium on Information Theory, Edited by Petrov B.N., and Csaki F. 267-281.

Arce, G.R., (2005). *Nonlinear signal processing: a statistical approach*. John Wiley & Sons.

Ashley, R.M., Crabtree R.W., (1992). Sediment origins, deposition and build-up in combined sewer systems. *Wat. Sci. Tech.*, 25(8), 1-12.

Ashley, R.M., Arthur, S., Coghlan, B.P., McGregor, I., (1994). Fluid sediment in combined sewers. *Wat. Sci. Tech.*, 29(1–2), 113–123.

Ashley, R.M., Bertrand-Krajewski J.-L., Hvitved-Jacobsen T., Verbanck M., (editors) (2004). *Solids in Sewers*. London (UK): IWA Publishing, Scientific and Technical Report n° 14, May 2004, 360 p. ISBN 1900222914.

Athayde, D.N., Shelly, P.E., Driscoll, E.D., Gaboury, D., Boyd, G., (1983). *Results of the Nationwide Urban Runoff Program: Volume I - Final Report*. Water Planning Div., U.S. Environ. Prot. Agency. Washington, DC .pp. 12-18.

Bai, S., Li, J., (2012). Sediment wash-off from an impervious urban land surface. *Journal of Hydrologic Engineering*, *18*(5), 488-498.

Bardossy, A., Singh, S.K., (2008). Robust estimation of hydrological model parameters. *Hydrology and Earth System Sciences*, *12*(6), 1273-1283.

Baroni, G., Tarantola, S., (2014). A General Probabilistic Framework for uncertainty and global sensitivity analysis of deterministic models: A hydrological case study. *Environmental Modelling & Software*, *51*, 26-34.

Beck, M.B., (1987). Water quality modeling: a review of the analysis of uncertainty. Water Resources Research, 23(8), 1393-1442.

Benedetti, L., Langeveld, J., Comeau, A., Corominas, L., Daigger, G., Martin, C., Mikkelsen, P.S., Vezzaro, L., Weijers, S., Vanrolleghem, P.A., (2013). Modelling and monitoring of integrated urban wastewater systems: review on status and perspectives. *Water Science and Technology*, 68(6), 1203-1215.

Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, J.H.A., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T.H., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., Andreassian, V., (2013). Characterising performance of environmental models. *Environmental Modelling & Software*, 40, 1-20.

Berg, E.L., (1982). *Handbook for sampling and sample preservation of water and wastewater - Report EPA-600/4-82-029*. Washington (USA): United States Environmental Protection Agency.

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

Bertrand-Krajewski, J.-L., Laplace, D., Joannis, C., Chebbo, G., (2000) *Mesures en hydrologie urbaine et assainissement* Editions Tec&Doc, Paris, France.

Bertrand-Krajewski, J.-L., Bardin, J., (2002). Evaluation of uncertainties in urban hydrology: application to volumes and pollutant loads in a storage and settling tank. *Water Science and Technology*, 45(4-5), 437-444.

Bertrand-Krajewski, J.-L., Muste, M., (2007). *Chapter 8 - Data validation: principles and implementation.* In *"Data requirements for Integrated Urban Water management*", T. Fletcher and A. Deletic (editors). London (UK): Taylor and Francis, Urban Water series - UNESCO IHP, 103-126.

Bertrand-Krajewski, J.-L., Barraud, S., Gibert, J., Malard, F.,Winiarski, T., and Delolme, C., (2007a). *The OTHU case study: integrated monitoring of stormwater in Lyon, France (Chapter 23)*. In: Fletcher, T., Deletic, A. (Eds.), *Data Requirements for Integrated Urban Water Management*. Urban Water Series - UNESCO IHP. Taylor and Francis, London (UK), ISBN 9780415453455, pp. 303-314.

Bertrand-Krajewski, J.-L., Winkler, S., Saracevic, E., Torres, A., Schaar, H., (2007b). Comparison of and uncertainties in raw sewage COD measurements by laboratory techniques and field UV-visible spectrometry. Water Science and Technology, 56(11), 17-25.

Beven, K., Smith, P., (2014). Concepts of information content and likelihood in parameter calibration for hydrological simulation models. *Journal of Hydrologic Engineering*, *20*(1), A4014010.

Bisselink, B., Zambrano-Bigiarini, M., Burek, P., De Roo, A., (2016). Assessing the role of uncertain precipitation estimates on the robustness of hydrological model parameters under highly variable climate conditions. *Journal of Hydrology: Regional Studies, 8*, 112-129.

Bonhomme, C., Petrucci, G., (2017). Should we trust build-up/wash-off water quality models at the scale of urban catchments? *Water Research*, 108, 422-431.

Breiman, L., Friedman, J., Olshen, R., Stone, C., (1984). *Classification and Regression Trees*. Boca Raton, FL: CRC Press.

Brigode, P., Oudin, L., Perrin, C., (2013). Hydrological model parameter instability: A source of additional uncertainty in estimating the hydrological impacts of climate change?. *Journal of Hydrology, 476*, 410-425.

Brombach, H., Weiss, G., Fuchs, S., (2005). A new database on urban runoff pollution: comparison of separate and combined sewer systems. *Water Science and Technology,* 51 (02), 119-128.

Burnham, K.P., Anderson, D.R., (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach (2nd ed.)*, Springer-Verlag, ISBN 0-387-95364-7

Burnham, K.P., Anderson, D.R., (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociol. Method. Res.* 33(2), 261-304.

Canny, J. (1986). A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell*., 8(6), 679-698.

Cantero-Chinchilla, F.N., Castro-Orgaz, O. Dey, S., (2016). Distribution of suspended sediment concentration in wide sediment-laden streams: A novel power-law theory. Sedimentology, Accepted manuscript online: 21 Feb. 2016. doi : 10.1111/sed.12276.

Carleton, J.N., Grizzard, T.J., Godrej, A. N., Post, H.E., (2001). "Factors affecting the performance of stormwater treatment wetlands." *Water Res.*, 35(6), 1552–1562.

CDEP, (2004). 2004 Connecticut Stormwater Quality Manual. The Connecticut Department of Environmental Protection, Hartford, CT, USA.

Charbeneau, R.J., Barrett, M.E., (1998). "Evaluation of methods for estimating stormwater pollutant loads." *Water Environ. Res.*, 70(7), 1295–1302.

Chaudhuri, P., (1996). On a geometric notion of quantiles for multivariate data. *Journal of the Americal Statistical Association* 91 862-872.

Chebana, F., Ouarda, T.B., (2011). Depth-based multivariate descriptive statistics with hydrological applications. *Journal of Geophysical Research: Atmospheres*, *116*(D10).

Chebbo, G., Gromaire, M.-C., (2004). The experimental urban catchment 'Le Marais' in Paris: what lessons can be learned from it? *Journal of Hydrology*, 299, 312–323.

Chebbo, G., Gromaire, M.-C., (2009). ViCAs—An operating protocol to measure the distributions of suspended solid settling velocities within urban drainage samples. *Journal of Environmental Engineering*, 135(9), 768-775.

Chiou, J.M., Li, P.L., (2007). Functional clustering and identifying substructures of longitudinal data. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 69, 679-699.

Chow, M.F., Yusop, Z., Abustan, I., (2015). Relationship between sediment build-up characteristics and antecedent dry days on different urban road surfaces in Malaysia. *Urban Water Journal*, *12*(3), 240-247.

Clark, M.P., Slater, A.G., Rupp, D.E., Woods, R.A., Vrugt, J.A., Gupta, H.V., Wagener, T. Hay, L.E., (2008). Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research*, *44*(12).

Coleman, N.L., (1982). Discussion of paper 16313, *J. Hydraul. Div*., ASCE, 108(1), 164–165.

Coutu, S., Del Giudice, D., Rossi, L., Barry, D.A., (2012). Parsimonious hydrological modeling of urban sewer and river catchments. *Journal of Hydrology*, *464*, 477-484.

Cristina, C.M., Sansalone, J.J., (2003). "First flush," power law and particle separation diagrams for urban storm-water suspended particulates. *Journal of Environmental Engineering*, *129*(4), 298-307.

Crobedu, E., Bennis, S., (2011). A new model for simulating TSS washoff in urban areas. Applied and Environmental Soil Science, Article ID 318765, 9 p. doi:10.1155/2011/318765.

Csardi, G., Nepusz, T., (2006). The igraph software package for complex network research. *Inter. J. Complex Systems*, 1695. http://igraph.org

Daly, E., Bach, P.M., Deletic, A., (2014). Stormwater pollutant runoff: A stochastic approach. *Advances in Water Resources*, *74*, 148-155.

Davis, P.M., Atkinson, T.C., (2000) Longitudinal dispersion in natural channels: 3. An aggregated dead zone model applied to the River Severn,U.K., Hydrology and Earth System Sciences, 4,373-381

Del Giudice, D., Albert, C., Rieckermann, J., Reichert, P., (2016). Describing the catchment-averaged precipitation as a stochastic process improves parameter and input estimation. *Water Resources Research*, *52*(4), 3162-3186.

Deletic, A., Dotto, C., Fletcher, T.D., McCarthy, D.T., Bertrand-Krajewski, J.-L., Rauch, W., Kleidorfer, M., Freni, G., Mannina, G., Tait, S., (2009). Defining uncertainties in modelling of urban drainage systems. In: *8th International Conference on Urban Drainage Modelling*, Tokyo, Japan, pp. 7-11.

Donoho, D.L., (1982). *Breakdown properties of multivariate location estimators*. Ph.D. qualifying paper. Department of Statistics, Harvard University.

Dorval, F., (2010). *Construction et calage d'un modèle de simulation continue des flux d'eau produits par les bassins versants urbains de taille intermédiaire*. Thèse de doctorat, INSA de Lyon, France (in French).

Dotto, C.B., Kleidorfer, M., Deletic, A., Rauch, W., McCarthy, D.T., Fletcher, T.D. (2011). Performance and sensitivity analysis of stormwater models using a Bayesian approach and long-term high resolution data. *Environmental Modelling & Software*, *26*(10), 1225-1239.

Dotto, C.B.S., Deletic, A., McCarthy, D.T., (2013). Uncertainty analysis in urban drainage modelling: should we break our back for normally distributed residuals?. *Water Science and Technology*, *68*(6), 1271-1279.

Duncan, H.P., (1999). Urban Stormwater Quality: A Statistical Overview. *Cooperative Research Centre for Catchment Hydrology*. 80 p.

Dyckerhoff, R., Koshevoy, G., Mosler, K., (1996), Zonoid data depth: theory and computation. In: Prat A. (ed), *COMPSTAT 1996. Proceedings in computational statistics*, Physica-Verlag (Heidelberg), 235--240.

Ebtehaj, M., Moradkhani, H., Gupta, H.V., (2010). Improving robustness of hydrologic parameter estimation by the use of moving block bootstrap resampling. *Water resources research*, *46*(7).

Egodawatta, P., Thomas, E., Goonetilleke, A., (2009). Understanding the physical processes of pollutant build-up and wash-off on roof surfaces. *Sci. Total Environ.* 407, 1834e1841. http://dx.doi.org/10.1016/j.scitotenv.2008.12.027

Einfalt, T., Arnbjerg-Nielsen, K., Golz, C., Jensen, N.E., Quirmbach, M., Vaes, G., Vieux, B., (2004). Towards a roadmap for use of radar rainfall data in urban drainage. *Journal of Hydrology*, 299(3), 186-202.

Ellis, J.B., Revitt, D.M., Scholes, L., Sieker, H., Helm, B., Winger, J., Picouet, C. Soutter, M., (2006). *Review of the adaptability and sensitivity of current stormwater control technologies to extreme environmental and socio-economic drivers*. Section B of Deliverable 2.1.1. EU 6th Framework SWITCH Project. Pp. 10-34.

Evin, G., Thyer, M., Kavetski, D., McInerney, D., Kuczera, G., (2014). Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity. *Water Resources Research*, *50*(3), 2350-2375.

Fay, M.P., Proschan, M.A., (2010). Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. *Statistics surveys*, 4, 1.

Fenicia, F., Reichert, P., Kavetski, D., Albert, C., (2016). Quantifying uncertainties in streamflow predictions through signature based inference of hydrological model parameters. In *EGU General Assembly Conference Abstracts* (Vol. 18, p. 11991).

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.

Freni, G., Mannina, G., Viviani, G., (2009). Uncertainty assessment of an integrated urban drainage model. *Journal of Hydrology*, *373*(3), 392-404.

Fuentes-Andino, D., Beven, K., Kauffeldt, A., Xu, C.Y., Halldin, S., Di Baldassarre, G., (2017). Event and model dependent rainfall adjustments to improve discharge predictions. *Hydrological Sciences Journal*, *62*(2), 232-245.)

Gamerith, V., (2011) *High resolution online data in sewer water quality modelling*. Thesis (PhD). Graz Institute of Technology, Graz, Austria.

Gelman, A., Rubin, D.B., (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 457-472.

Gharari, S., Hrachowitz, M., Fenicia, F., Savenije, H.H.G., (2013). An approach to identify time consistent model parameters: sub-period calibration. *Hydrology and Earth System Sciences*, *17*(1), 149-161.

Glen, G. Isaacs, K., (2012). Estimating Sobol sensitivity indices using correlations. *Environmental Modelling and Software, 37*, 157-166.

Goonetilleke, A., Yigitcanlar, T., Ayoko, G.A., Egodawatta, P., (2014). *Sustainable urban water environment: climate, pollution and adaptation*. Edward Elgar Publishing.

Gruber, G., Winkler, S., Pressl, A., (2005). Continuous monitoring in sewer networks an approach for quantification of pollution loads from CSOs into surface water bodies. *Water Science and Technology* 52 (12), 215-223

Guerrero, J.L., Westerberg, I.K., Halldin, S., Lundin, L.C., Xu, C.Y., (2013). Exploring the hydrological robustness of model-parameter values with alpha shapes. *Water Resources Research*, *49*(10), 6700-6715.

Gy, P.M., (1998), Sampling for Analytical Purposes, John Wiley and Sons Ltd., Chichester.

Harmel, R.D., Smith, P.K., (2007). Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modeling. *Journal of Hydrology*, *337*(3), 326-336.

Helton, J.C., Davis, F.J., (2003). Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering and System Safety*, 81(1), 23-69.

Hochedlinger, M., Kainz, H., Rauch, W., (2006). Assessment of CSO loads based on UV/VIS-spectroscopy by means of different regression methods. *Water Science and Technology* 54 (6-7), 239-246.

Hollander, M., Wolfe, D.A., (1999). *Nonparametric Statistical Methods*. Hoboken, NJ: John Wiley & Sons, Inc.

Hong, Y., Bonhomme, C., Le, M.H., Chebbo, G., (2016). New insights into the urban washoff process with detailed physical modelling. *Science of the Total Environment*, 573, 924-936.

Huang, Y., Liu, L., (2010). Multiobjective water quality model calibration using a hybrid genetic algorithm and neural network–based approach. *Journal of Environmental Engineering*, *136*(10), 1020-1031.

Huber, W.C., Dickinson, R.E., Barnwell Jr, T.O., Branch, A., (1988). Storm water management model; version 4. *Environmental Protection Agency, United States*.

Hur, J., Lee, B.-M., Lee, T.H., Park, D.-H., (2010). Estimation of Biological Oxygen Demand and Chemical Oxygen Demand for Combined Sewer Systems Using Synchronous Fluorescence Spectra. *Journal Sensors* 10, 2460-2471.

Iman, R.L., Conover, W.J., (1980). Small sample sensitivity analysis techniques for computer models. with an application to risk assessment. *Communications in statistics-theory and methods*, 9(17), 1749-1842.

Imteaz, M.A., Hossain, I., Hossain, M.I., (2014). Estimation of build-up and wash-off models parameters for an east-Australian catchment. *International Journal of Water*, *8*(1), 48-62.

ISO (2009). ISO/IEC Guide 98-1:2009(E) Uncertainty of measurement – Part 1: *Introduction to the expression of the uncertainty in measurement*. Geneva (Switzerland): ISO, September 2009, 32 p.

Izuno, F.T., Rice, R.W., Garcia, R.M., Capone, L.T., Downey, D., (1998). Time versus flow composite water sampling forregulatory purposes in the Everglades Agricultural Area. *Applied Engineering in Agriculture*, *14*(3), 257–266.

Jakeman, A.J., Littlewood, I. G., Whitehead, P.G., (1990). Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. Journal of Hydrology 117, 275–300.

Joannis, C., Ruban, G., Gromaire, M.-C., Bertrand-Krajewski, J.-L., Chebbo G., (2008). Reproducibility and uncertainty of wastewater turbidity measurements. *Wat. Sci. Tech.*, 57(10), 1667-1673.

Kafi-Benyahia, M., Gromaire, M.C., Chebbo, G., (2006). Representativity of samplings by automatic vacuometric samplers in combined sewer: case of OPUR. *La Houille Blanche*, 4, 113–120.

Kanso, A., Tassin, B., Chebbo, G. (2005). A benchmark methodology for managing uncertainties in urban runoff quality models. Water Sci. Technol. 51(2), 163–170.

Kavetski, D., Kuczera, G., Franks, S.W., (2006a). Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. *Water Resources Research*, 42(3).

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

Ki, S.J., Kang, J.H., Lee, S.W., Lee, Y.S., Cho, K.H., An, K.G., Kim, J.H., (2011). Advancing assessment and design of stormwater monitoring programs using a self-organizing map: Characterization of trace metal concentration profiles in stormwater runoff. *Water research*, *45*(14), 4183-4197.

Kidd, C., Levizzani, V., (2011). Status of satellite precipitation retrievals. *Hydrol. Earth Syst. Sci*. 15 (4), 1109–1116.

Kim, L.H., Kayhanian, M., Zoh, K.D., Stenstrom, M.K., (2005). Modeling of highway stormwater runoff. *Science of the Total Environment* 348 (1-3), 1-18.

King, K.W., Harmel, R.D., (2004). Comparison of time-based sampling strategies to determine nitrogen loading in plot-scale runoff. *Transactions of the ASAE*, *47*(5), 1457–1463.

King, K.W., Harmel, R.D., Fausey, L., (2005). Development and sensitivity of a method to select time- and flow-paced storm event sampling intervals for headwater streams. *Journal of Soil and Water Conservation*, *60*(6), 323–331.

Kretzschmar, A., Tych, W., Chappell, N., Beven, K., (2016). What really happens at the end of the rainbow?–paying the price for reducing uncertainty (using reverse hydrology models). *Procedia Engineering*, *154*, 1333-1340.

Kreyszig, E., (1979). Advanced Engineering Mathematics (Fourth ed.). Wiley. p. 880, eq. 5. ISBN 0-471-02140-7.

Kuczera, G., Kavetski, D., Franks, S., Thyer, M., (2006). Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters. *Journal of Hydrology*, *331*(1), 161-177.

Kullback, S., Leibler, R.A., (1951). "On information and sufficiency". *Annals of Mathematical Statistics*. 22 (1): 79–86. MR 39968. doi:10.1214/aoms/1177729694.

Langergraber, G., Weingartner, A., Fleischmann, N., (2004a). Time-resolved delta spectrometry: a method to define alarm parameters from spectral data, *Water science and technology*, 50(11), 13–20.

Langergraber, G., Fleischmann, N., Hofstaedter, F., Weingartner, A., (2004b) Monitoring of a paper mill wastewater treatment plant using UV/VIS spectroscopy. *Trends in Sustainable Production*, 49(1), 9–14.

Larrarte, F., (2008). Suspended solids within sewers : an experimental study. *Environmental Fluid Mechanics*, 8(3), 249-261.

Larrarte, F., Pons, M.-N., (2011). Suspended solids concentration in wastewater: Influence of sampling conditions. *Urban Water*, 8(6), 397-404.

Lee, H., Swamikannu, X., Radulescu, D., Kim, S.J., Stenstrom, M.K., (2007). Design of stormwater monitoring programs. *Water Research* 41 (18), 4186-4196.

Lee, J.H., Bang, K.W., Ketchum, L.H., Choe, J.S., Yu, M.J., (2002). First flush analysis of urban storm runoff. *Science of the Total Environment*, *293*(1), 163-175.

Leecaster, M., Schiff, K., and Tiefenthaler, L., (2002). Assessment of efficient sampling designs for urban stormwater monitoring. *Water Research*, *36*, 1556–1564.

Leonhardt, G., (2014). *Development and application of software sensors and reverse models for urban drainage*. PhD Thesis, University of Innsbruck, Innsbruck, Austria, 49 p.

Leonhardt, G., Sun, S., Rauch, W., Bertrand-Krajewski, J.-L., (2014). Comparison of two model based approaches for areal rainfall estimation in urban hydrology. *Journal of*

Hydrology, 511, 880-890.

Lepot, M.J., Torres, A., Hofer, T., Caradot, N., Gruber, G., Aubin, J.-B., Bertrand-Krajewski, J.-L., (2016). Calibration of UV/Vis spectrophotometers: a review and comparison of different methods to estimate TSS and total and dissolved COD concentrations in sewers, WWTPs and rivers. *Water Research*, 101, 519-534. doi:10.1016/j.watres.2016.05.070.

Liu, A., Goonetilleke, A., Egodawatta, P., (2012). Inherent errors in pollutant build-up estimation in considering urban land use as a lumped parameter. J. Environ. Qual. 41, 1690e1694. http://dx.doi.org/10.2134/jeq2011.0419

Ma, J.A., Kang, J.H., Kayhanian, M., and Stenstrom, M.K., (2009). Sampling issues in urban runoff monitoring programs: Composite versus grab. *Journal of Environmental Engineering, 135*(3), 118–127.

Ma, Z.Z., Wang, Z.J., Xia, T., Gippel, C.J., Speed, R., (2014). Hydrograph-based hydrologic alteration assessment and its application to the Yellow River. *J. Environ. Inform.* 23 (1), 1–13.

Mahalanobis, P., (1936). On the generalized distance in statistics. *Proceedings of the National Academy India* 12 49—55

Mancipe-Munoz, N.A., Buchberger, S.G., Suidan, M.T., Lu, T., (2014). Calibration of rainfall-runoff model in urban watersheds for stormwater management assessment. *J. of Water Resources Planning and Management*, 140(6), 05014001.

Mannina, G., Viviani, G., (2010). An urban drainage stormwater quality model: model development and uncertainty quantification. *Journal of hydrology*, *381*(3), 248-265.

Manz, B.J., Rodríguez, J.P., Maksimović, Č., McIntyre, N., (2013). Impact of rainfall temporal resolution on urban water quality modelling performance and uncertainties. *Water Science and Technology*, *68*(1), 68-75.

Marshall, L., Nott, D., Sharma, A., (2007). Towards dynamic catchment modelling: a Bayesian hierarchical mixtures of experts framework. *Hydrological Processes*, *21*(7), 847-861.

McCarthy, D.T., Hathaway, J.M., Hunt, W.F., Deletic, A., (2012). Intra-event variability of Escherichia coli and total suspended solids in urban stormwater runoff. *Water research*, *46*(20), 6661-6670.

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), 83-94.

McMurry, T.L., Politis, D.N., (2010). Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. *Journal of Time Series Analysis*, 31(6), 471-482.
Messer, H., Zinevich, A., Alpert, P. (2006). Environmental monitoring by wireless communication networks. *Science* 312 (5774), 713.

Métadier, M., (2011). *Traitement et analyse de séries chronologiques continues de turbidité pour la formulation et le test de modèles des rejets urbains par temps de pluie*. PhD Thesis, University of INSA, Lyon, France, 339 p. + appendices (in French).

Métadier, M., Bertrand-Krajewski, J.-L., (2011). From mess to mass: a methodology for calculating storm event pollutant loads with their uncertainties, from continuous raw data time series. *Wat. Sci. Tech.*, 63(3), 369-376.

Métadier, M., Bertrand-Krajewski, J.-L., (2012). Pollutographs, concentrations, loads and intra-event mass distributions of pollutants in urban wet weather discharges calculated from long term on line turbidity measurements. *Water Research*, 46(20), 6836-6856. doi:10.1016/j.watres.2011.12.030.

Minkkinen, P., (2004). Practical applications of sampling theory. *Chemometrics and intelligent laboratory systems*, 74(1), 85-94.

MOE (Ministry of the Environment Ontario, Canada) (2003). *Stormwater Management Planning and Design Manual. Ministry of the Environment*. Canada, Toronto.

Montanari, A., Brath, A., (2004). A stochastic approach for assessing the uncertainty of rainfall-runoff simulations. *Water Resources Research*, *40*(1).

Montanari, A., Grossi, G., (2008). Estimating the uncertainty of hydrological forecasts: A statistical approach. *Water Resources Research*, *44*(12).

Muleta, M.K., McMillan, J., Amenu, G.G., Burian, S.J., (2012). Bayesian approach for uncertainty analysis of an urban storm water model and its application to a heavily urbanized watershed. *Journal of Hydrologic Engineering*, *18*(10), 1360-1371.

Mourad, M., Bertrand-Krajewski, J.-L., Chebbo, G., (2005). Calibration and validation of multiple regression models for stormwater quality prediction: data partitioning, effect of dataset size and characteristics. *Water Science and Technology*, 52(3), 45-52.

Muste, M., Lee, K., Bertrand-Krajewski, J.-L., (2012). Standardized uncertainty analysis for hydrometry: a review of relevant approaches and implementation examples. *Hydrological Sciences Journal*, 57(4), 643-667.

Nearing, G.S., Tian, Y., Gupta, H.V., Clark, M.P., Harrison, K.W., Weijs, S.V., (2016). A philosophical basis for hydrological uncertainty. *Hydrological Sciences Journal*, *61*(9), 1666-1678.

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

Paakkunainen, M., Reinikainen, S.P., Minkkinen, P., (2007). Estimation of the variance of sampling of process analytical and environmental emissions measurements. *Chemometrics and intelligent laboratory systems*, *88*(1), 26-34.

Pan, X., Zhang, J., Jones, K.D., (2013). Simulation of storm event flow for pilot runoff treatment wetland. *Ecological engineering*, *53*, 284-289.

Pearson, K., (1901). On lines and planes of closest fit to systems of points in space. Philosophical Magazine 2(6), 559–572

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.

Pedregal, D., Taylor, C.J., Young, P.C., (2007). System identification, time series analysis and forecasting. *CAPTAIN handbook: The Captain Toolbox*. (CAPTAIN toolbox)

Piro, P., Carbone, M., (2014). A modelling approach to assessing variations of total suspended solids (tss) mass fluxes during storm events. *Hydrological Processes*, *28*(4), 2419-2426.

Pitt, R., Bannerman, R., Clark, S., Williamson, D., (2004). "Source of pollutants in urban areas (Part 1)—Older monitoring projects." *Effective modeling of urban water systems*, Monograph 13, CHI, Guelph, Ontario, Canada.

Pokotylo, O., Mozharovskyi, P., Dyckerhoff, R. (2016). *Depth and depth-based classification with R-package ddalpha*.

R Development Core Team, (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Qin, H.P., He, K.M., Fu, G., (2016). Modeling middle and final flush effects of urban runoff pollution in an urbanizing catchment. *Journal of Hydrology*, 534, 638-647.

Raudkivi, A.J., (1998). *Loose Boundary Hydraulics*. Rotterdam (Netherlands): Taylor & Francis, 512 p.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo., D.-J. (2004). Overall distributed model intercomparison project results, *J. Hydrol*.,298, 27 – 60, doi:10.1016/j.jhydrol.2004.03.031

Reichardt, J., Bornholdt, S., (2006). Statistical Mechanics of Community Detection, *Phys. Rev. E*, 74, 016110 (2006), http://arxiv.org/abs/cond-mat/0603718.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., Franks, S.W., (2010). Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research*, *46*(5).

Renard, B., Kavetski, D., Leblois, E., Thyer, M., Kuczera, G., Franks , S.W., (2011). Towards a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation, *Water Resources Research*, vol. 47.

Richards, R.P., Holloway, J., (1987). Monte Carlo studies of sampling strategies for estimating tributary loads. *Water Resources Research*, *23*(10), 1939–1948.

Ristenpart, E., (1995). *Feststoffe in der Mischwasserkanalisation: Vorkommen, Bewegung und Verschmutzungs-potential*, PhD Thesis, University of Hannover, Germany (in German).

Ristenpart, E., Ashley, R.M., Uhl, M., (1995). Organic near-bed fluid and particulate transport in combined sewers. *Wat. Sci. Tech*., 31(7), 61–68.

Robertson, D.M., Roerish, E.D., (1999). Influence of various water quality sampling strategies on load estimates for small streams. *Water Resources Research, 35*(12), 3747–3759.

Robien, A., Striebel, T., Herrmann, R., (1997). Modeling of dissolved and particlebound pollutants in urban street runoff. Water Sci. Technol. 36, 77-82. http://dx.doi.org/10.1016/S0273-1223(97)00616-1.

Rossi, L., (1998). *Qualité des eaux de ruissellement urbaines*. PhD Thesis, Swiss Federal Institute of Technology, Lausanne, Switzerland, 416 p. (in French).

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

Ruban, G., Ruperd, Y., Laveau, B., Lucas, E., (2001). «Self-monitoring of water quality in sewer systems using absorbance of ultraviolet and visible light». *1st world water congress: operation and management, landfill and sludge management*. Vol. 44, No 2-3, pp. 269-276.

Sage, J., Bonhomme, C., Al Ali, S., Gromaire, M.-C., (2015). Performance assessment of a commonly used “accumulation and wash-off” model from long-term continuous road runoff turbidity measurements. Water Res. 78, 47-59. http://dx.doi.org/10.1016/j.watres.2015.03.030

Saget, A., (1994). *Data base of stormwater quality: distribution of discharged pollutant loads and sizes of interception constructions*. Ph.D. thesis, Ecole Nationale des Ponts et Chaussées, Paris, France, p. 333 (in French).

Sandoval, S., Torres, A. (2013). Identification of dry/rainy periods from TSS loads’ time series measured at the effluent of an urban catchment. *Proceedings of Novatech* 2013, Lyon, France, 23-27 June, 10 p.

Sandoval, S., Torres, A., Pawlowsky-Reusing, E., Riechel, M., Caradot, N., (2013). The evaluation of rainfall influence on CSO characteristics: the Berlin case study. *Water Science and Techonology*. doi:10.2166/wst.2013.524

Sartor, J.D., Boyd, G.B., Agardy, F.J., (1974). Water pollution aspects of street surface contaminants. J. Water Pollut. Control Fed. 46, 458-467.

Schellart, A.N.A., Shepherd, W.J., Saul, A.J., (2012). Influence of rainfall estimation error and spatial variability on sewer flow prediction at a small urban scale. *Adv. Water Resour.* 45, 65–75.

Schilperoort, R.P.S., (2011*). Monitoring as a tool for the assessment of wastewater quality dynamics.* PhD thesis, TU Delft, The Netherlands, June 2011, 320 p.

Schoups, G., Vrugt, J.A., (2010). A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resources Research, 46*(10).

Schwarz, G., (1978). Estimating the dimension of a model. *The annals of statistics, 6*(2), 461-464.

Seibert, J., (2003). Reliability of model predictions outside calibration conditions. *Hydrology*

Research, *34*(5), 477-492.

Shaw, S.B., Parlange, J.Y., Lebowitz, M., Walter, M.T., (2009). Accounting for surface roughness in a physically-based urban wash-off model. *Journal of hydrology*, *367*(1), 79-85.

Shelley, P.E., (1977). *Sampling of Water and Wastewater*. Cincinatti (USA): Environmental Protection Agency, PB-272 664.

Shih, G., Abtew, W., Obeysekera, J., (1994). Accuracy of nutrient runoff load calculations using time-composite sampling. *Transactions of ASAE, 37*(2), 419–429.

Sikorska, A.E., Montanari, A., Koutsoyiannis, D., (2014). Estimating the uncertainty of hydrological predictions through data-driven resampling techniques. *Journal of Hydrologic Engineering*, *20*(1), A4014009.

Sikorska, A.E., Del Giudice, D., Banasik, K., Rieckermann, J., (2015). The value of streamflow data in improving TSS predictions–Bayesian multi-objective calibration. *Journal of Hydrology*, *530*, 241-254.

Singh, S. K., McMillan, H., Bárdossy, A., Fateh, C., (2016). Nonparametric catchment clustering using the data depth function. *Hydrological Sciences Journal*, *61*(15), 2649-2667.

Singh, S.K., Bardossy, A., (2012). Calibration of hydrological models on hydrologically unusual events. *Advances in Water Resources*, *38*, 81-91.

Srikanthan, R., Kuczera, G. A., Thyer, M.A., (2009). 'Calibrate it twice': A simple resampling method for incorporating parameter uncertainty in stochastic data generation. In *H2009: 32nd Hydrology and Water Resources Symposium, Newcastle: Adapting to Change* (p. 1028). Engineers Australia.

Stone, K.C., Hunt, P.G., Novak, J.M., Johnson, M.H., Watts, D.W., (2000). Flow-proportional, time-composited, and grab sample estimation of nitrogen export from an Eastern Coastal Plain watershed. *Transactions of ASAE, 43*(2), 281–290.

Stransky, D., Bares, V., Fatka, P., (2006). The Effect of Rainfall Measurement Uncertainties on Rainfall-runoff: Processes Modelling [online]. In: Delectic, Ana (Editor); Fletcher, Tim (Editor). *7th International Conference on Urban Drainage Modelling and the 4th International Conference on Water Sensitive Urban Design; Book of Proceedings.* [Clayton, Vic.]: Monash University: [1049]-[1056].

Sun, S., Barraud, S., Castebrunet, H., Aubin, J. B., Marmonier, P., (2015). Long-term stormwater quantity and quality analysis using continuous measurements in a French urban catchment. *Water research*, *85*, 432-442.

Sun, S., Bertrand-Krajewski, J.-L., (2013a). Separately accounting for uncertainties in rainfall and runoff: Calibration of event based conceptual hydrological models in small urban catchments using Bayesian method. *Water Resources Research*, 49, 1-14. doi 10.1002/wrcr.20444.

Sun, S., Bertrand-Krajewski, J.-L., (2013b). Parsimonious conceptual hydrological model selection with different modeling objectives. Proceedings of Novatech 2013, Lyon, France, 23-27 June, 5 p.

Tan, S.B., Chua, L.H., Shuy, E.B., Lo, E.Y.M., Lim, L.W., (2008). Performances of rainfall-runoff models calibrated over single and continuous storm flow events. *Journal of Hydrologic Engineering*, 13(7), 597-607.

Ternynck, C., Ben Alaya, M.A., Chebana, F., Dabo-Niang, S., Ouarda, T.B., (2016). Streamflow hydrograph classification using functional data analysis. *Journal of Hydrometeorology*, *17*(1), 327-344.

Thirel, G., Andréassian, V., Perrin, C., Audouy, J. N., Berthet, L., Edwards, P., Folton, N., Furusho, C., Kuentz, A., Lerat, J., Lindström, G., Martin, E., Mathever, T., Merz, R., Parajka, J., Ruelland, D., Vaze, J., (2015). Hydrology under change: an evaluation protocol to investigate how hydrological models deal with changing catchments. *Hydrological Sciences Journal*, *60*(7-8), 1184-1199.

Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, S.W., Srikanthan, S., (2009). Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis. *Water Resources Research*, 45, W00B14, doi:10.1029/2008WR006825.

Torres, A., (2008). *Décantation des eaux pluviales dans un ouvrage réel de grande taille : éléments de réflexion pour le suivi et la modélisation*. PhD Thesis, INSA de Lyon, France, 368 p. (in French).

Torres, A., Bertrand-Krajewski, J.-L., (2008a). Evaluation of uncertainties in settling velocities of particles in urban stormwater runoff. *Water Science and Technology*, 57(9), 1389-1396.

Torres, A., Bertrand-Krajewski, J.-L., (2008b). Partial Least Squares local calibration of a UV-visible spectrometer used for in situ measurements of COD and TSS concentrations in urban drainage systems. *Water Science and Technology*, 57(4), 581-588.

Tukey, J.W., (1974), Mathematics and the picturing of data, In: *Proceeding of the International Congress of Mathematicians*, Vancouver, 523--531.

United Nations, (2015). World Urbanization Prospects: The 2014 Revision. Department of Economic and Social Affairs Population Division (ST/ESA/SER.A/366). https://esa.un.org/unpd/wup/Publications/Files/WUP2014-Report.pdf (accessed 15 October 2015).

US EPA, (1983). Results of the Nationwide Urban Runoff Program. U.S. Environmental Protection Agency, Washington D.C.

US EPA, (1990). National pollutant discharge elimination system permit application regulations for stormwater discharges; final rule. Fed. Regist. 55 (222), 47990e48091. http://dx.doi.org/10.1016/j.envsoft.2012.02.007.

Vaes, G., Willems, P., Berlamont, J., (2005). Areal rainfall correction coefficients for small urban catchments. *Atmos. Res.* 77 (1–4), 48–59.

van het Schip, T.I., Overeem, A., Leijnse, H., Uijlenhoet, R., Meirink, J.F., van Delden, A.J., (2017). Rainfall measurement using cell phone links: classification of wet and dry periods using geostationary satellites. *Hydrological Sciences Journal*, 1-11.

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

Vaze, J., Chiew, F.H., (2003). Comparative evaluation of urban storm water quality models. *Water Resources Research*, *39*(10).

Verbanck, M.A., (1993). Identification of the depth-dependent transportation of particulate solids in dry weather sewage flows. *Proceedings of 6th International Conference on Urban Storm Drainage*, Niagara Falls, Canada, 742–747.

Verbanck, M.A., (1995). Capturing and releasing settleable solids: the significance of dense undercurrents in combined sewer flows. *Wat. Sci. Tech*., 31(7), 85–93.

Verbanck, M.A., (2000) Computing near-bed solids transport in sewers and similar sediment-carrying open-channel flows. *Urban Water*, 2(4), 277–284.

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

Vrugt, J. A., Ter Braak, C.J.F., Diks, C.G.H., Robinson, B.A., (2009), Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches to hydrologic modeling, Stochastic Environ. Res. Risk Assess., 23, 1059–1060.

Vrugt, J.A., (2016). Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling & Software*, *75*, 273-316.

Wagener, T., McIntyre, N., Lees, M.J., Wheater, H.S., Gupta, H.V., (2003). Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis. *Hydrological Processes*, *17*(2), 455-476.

Wald, A., Wolfowitz, J., (1943). An exact test for randomness in the non-parametric case based on serial correlation. *The Annals of Mathematical Statistics*, *14*(4), 378-388.

Wheater, H.S., Jakeman, A.J., Beven, K.J., (1993). Progress and directions in rainfall-runoff modeling. In Modeling Change in Environmental Systems, Jakeman AJ, Beck MB, McAleer MJ (eds). John Wiley & Sons: UK; 101–132.

Wijesiri, B., Egodawatta, P., McGree, J., Goonetilleke, A., (2015). Process variability of pollutant build-up on urban road surfaces. Sci. Total Environ. 518-519, 434-440. http://dx.doi.org/10.1016/j.scitotenv.2015.03.014

Winkler, S., Saracevic, E., Bertrand-Krajewski, J.-L., Torres, A., (2008). Benefits, limitations

and uncertainty of in situ spectrometry. *Water science and technology*, 57(10), 1651.

Wu, C., Murray, A.T., (2003). Estimating Impervious Surface Distribution by Spectral Mixture Analysis. *Remote Sensing of Environment*, Vol. 84, No. 4, 2003, pp. 493-505.

Wu, W. B., Pourahmadi, M., (2009). Banding sample autocovariance matrices of stationary processes. *Statistica Sinica*, 19(4), 1755.

Xiao, H., Wu, W.B., (2012). Covariance matrix estimation for stationary time series. *The Annals of Statistics*, 40(1), 466-493.

Xue, L., Ma, S., Zou, H., (2012). Positive-Definite ℓ1-Penalized Estimation of Large Covariance Matrices. *Journal of the American Statistical Association*, 107(500), 1480-1491.

Yang, J., Reichert, P., Abbaspour, K.C., Xia, J., Yang, H., (2008). Comparing uncertainty 692 analysis techniques for a SWAT application to the Chaohe Basin in China. Journal of 693 Hydrology 358, 1–23.

Yang, Y., (2005). Can the strengths of AIC and BIC be shared? A conflict between model indentification and regression estimation. Biometrika, 92(4), 937-950.

Ye, L., Zhou, J., Zeng, X., Guo, J., Zhang, X., (2014). Multi-objective optimization for construction of prediction interval of hydrological models based on ensemble simulations. *J. of Hydrol.*, 519, 925-933.

Young, P., (1989). Recursive estimation, forecasting and adaptive control. In C.T. Leondes (Ed.), *Control and dynamic systems: Advances in theory and applications* (pp. 119–166). San Diego, CA: Academic Press.

Young, P., (1998). Data-based mechanistic modelling of environmental, ecological, economic and engineering systems. *Environmental Modelling & Software*, *13*(2), 105-122.

Young, P., (2003). Top-down and data-based mechanistic modelling of rainfall-flow dynamics at the catchment scale. Hydrological Processes 17, 2195–2217.

Young, P., (2012). *Recursive estimation and time-series analysis: an introduction*. Springer Science & Business Media.

Young, P, Garnier, H. (2006). Identification and estimation of continuous-time, data-based mechanistic (DBM) models for environmental systems. *Environmental modelling & software*, *21*(8), 1055-1072. (transfer function and TVP)

Zhang, C., Wang, R.B., Meng, Q.X., (2015). Calibration of conceptual rainfall-runoff models using global optimization. *Advances in Meteorology, 2015*.

Zhao, J., Chen, Y., Hu, B., Yang, W., (2015). Mathematical Model for Sediment Wash-Off from Urban Impervious Surfaces. *Journal of Environmental Engineering*, *142*(4), 04015091.

Zoppou, C., (2001). Review of urban storm water models. *Environmental Modelling & Software*, *16*(3), 195-231.

APPENDICES

1. PRESENTATION DES RESULTATS MAJEURS DE LA THESE – RESUME ETENDU EXIGE POUR UNE THESE REDIGEE EN ANGLAIS (*PRESENTATION OF THE PRINCIPAL RESULTS OF THE THESIS –LONG ABSTRACT DEMANDED FOR A THESIS WRITTEN IN ENGLISH*)

CHAPITRE 2 Evaluation de performance et d'incertitudes dans les stratégies d'échantillonnage d'eaux de pluie, fondée sur le débit et les séries chronologiques de charge totale de solides en suspension

Introduction

Un indicateur commun pour estimer les émissions polluantes est la Concentration Moyenne d'Evénement (CME), qui est très variable selon les stratégies d'échantillonnage adoptées (Lee *et al.*, 2007; Ki *et al.*, 2011). Les stratégies d'échantillonnage sont des règles d'applications pour l'échantillonnage de polluants, par exemple les charges Totales de Matière en Suspension (MES), pendant des événements pluvieux avec un échantillonneur automatique. La valeur CME est estimée en mélangeant manuellement ou automatiquement les échantillons individuels collectés dans un bocal à échantillon (Lee *et al.*, 2007). Le plus gros inconvénient pour évaluer quelle stratégie d'échantillonnage peut apporter les CME les plus justes est que le "vrai" CME ou le CME "de référence" est établi à travers des suppositions théoriques, vu le manque de données de concentration (par exemple : Ma *et al.*, 2009; Ki *et al.*, 2011) . D'autre part, de nombreux auteurs ont rapporté les bénéfices du suivi en ligne pour expliquer la variabilité significative de la qualité des eaux de pluie. En conséquence, le suivi en ligne émerge comme une alternative prometteuse pour l'estimation des "vrais" CME ou des CME "de référence" à travers une série de conditions. La présente étude propose de simuler et d'évaluer différentes stratégies d'échantillonnage (exemple : Ackerman *et al.*, 2010) en utilisant un débit de haute résolution et une série chronologique de MES pendant des événements pluvieux. Les CME de stratégies d'échantillonnage sont simulés en échantillonnant des séries chronologiques de TSS et en calculant une moyenne pondérée des échantillons par leurs volumes d'échantillonnage. Ces CME "simulés" sont comparés au "vrai" CME calculé comme une moyenne pondérée du débit complet et des séries chronologiques de MES pendant l'événement pluvieux.

Matériels et méthodes

Les quatre cas internationaux sont les suivants: i) Berlin, Allemagne (débordement d'égouts unitaire, surface de 100 ha, 22 événements pluvieux); ii) Chassieu, France (réseau séparatif des eaux, surface de 185 ha, 75 événements pluvieux); iii) Graz, Autriche (débordement d'égouts unitaire, surface de 335 ha, 85 événements) and iv) Ecully, France (réseau d'égouts unitaire, surface de 245 ha, 200 événements pluvieux). Les CME obtenus à partir de stratégies d'échantillonnage (CEMsim) sont simulés en échantillonnant des séries chronologiques de

MES et en calculant une moyenne pondérée des valeurs MES “échantillonnées” par leur volume d’échantillon. Les valeurs CEMsim sont ensuite comparés aux “vrais” CME (EMCtrue) calculés comme une moyenne pondérée du débit et des séries chronologiques de TSS pendant l’événement pluvieux.

Quatre stratégies d’échantillonnage typiques sont étudiées : i) stratégie d’échantillonnage à intervalle temporel constant et volume constant (cTcSV): les échantillons sont collectés à des intervalles temporels constants, les volumes d’échantillonnage sont également constants (par exemple: échantillonner toutes les 10 minutes un volume d’échantillonnage de 0.4L), ii) stratégie d’échantillonnage à intervalle temporel constant et proportionnel à la décharge (cTpQ): les échantillons sont collectés à des intervalles temporels constants et les volumes d’échantillonnage sont prédéfinis comme proportionnels au débit instantané mesuré au pas de temps d’échantillonnage (par exemple : échantillonner toutes les 10 minutes un volume d’échantillonnage de 0.2L si le débit instantané est de 0.2 m3/s), iii) stratégie d’échantillonnage à intervalle temporel constant et proportionnel au volume (cTpV): les échantillons sont collectés à des intervalles temporels constants et les volumes d’échantillons sont prédéfinis comme proportionnels au volume de débordement cumulé depuis le dernier échantillon (par exemple : échantillonner toutes les 10 minutes un volume d’échantillonnage de 0.1L si le volume de débordement depuis les échantillons précédents est de 10 m3), et iv) stratégie d’échantillonnage à intervalle temporel variable et volume de débordement constant (vTcV): les échantillons sont collectés à un volume de débordement cumulé prédéfini depuis l’échantillon précédent (e.g. échantillonner avec constance le volume de 0.4 L pour un volume de débordement cumulé de 10 m3 entre échantillons). L’erreur quadratique moyenne relative (MSRE) entre EMCtrue et EMCsim et son incertitude élargie CI(MSRE) sont utilisés comme des indicateurs de performance pour les stratégies d’échantillonnage. Les incertitudes standard dans les données utilisées pour des calculs se trouvent dans i) les volumes d’échantillonnage (4.5 %), ii) les analyses laboratoire de TSS (7.5 %), iii) la mesure en ligne (selon la technologie et la méthode de calibration) et iv) les temps de début et de fin d’événements pluvieux (respectivement 5 % et 7.5 % de la durée des précipitations).

Résultats

Les résultats sont résumes dans le Figure 1, dans lequel MSRE (lignes continues) et CI(MSRE) (bandes de couleur) sont donnés pour les intervalles temporels d’échantillonnage entre 1 et 60 minutes sur l’axe horizontal inférieur (stratégies cTcSV, cTpQ et cTpV). L’axe horizontal supérieur montre les différents volumes de débordement utilizes pour évaluer la stratégie vTcV, avec les intervalles temporels moyens correspondants sur l’axe horizontal inférieur (Figure 1).

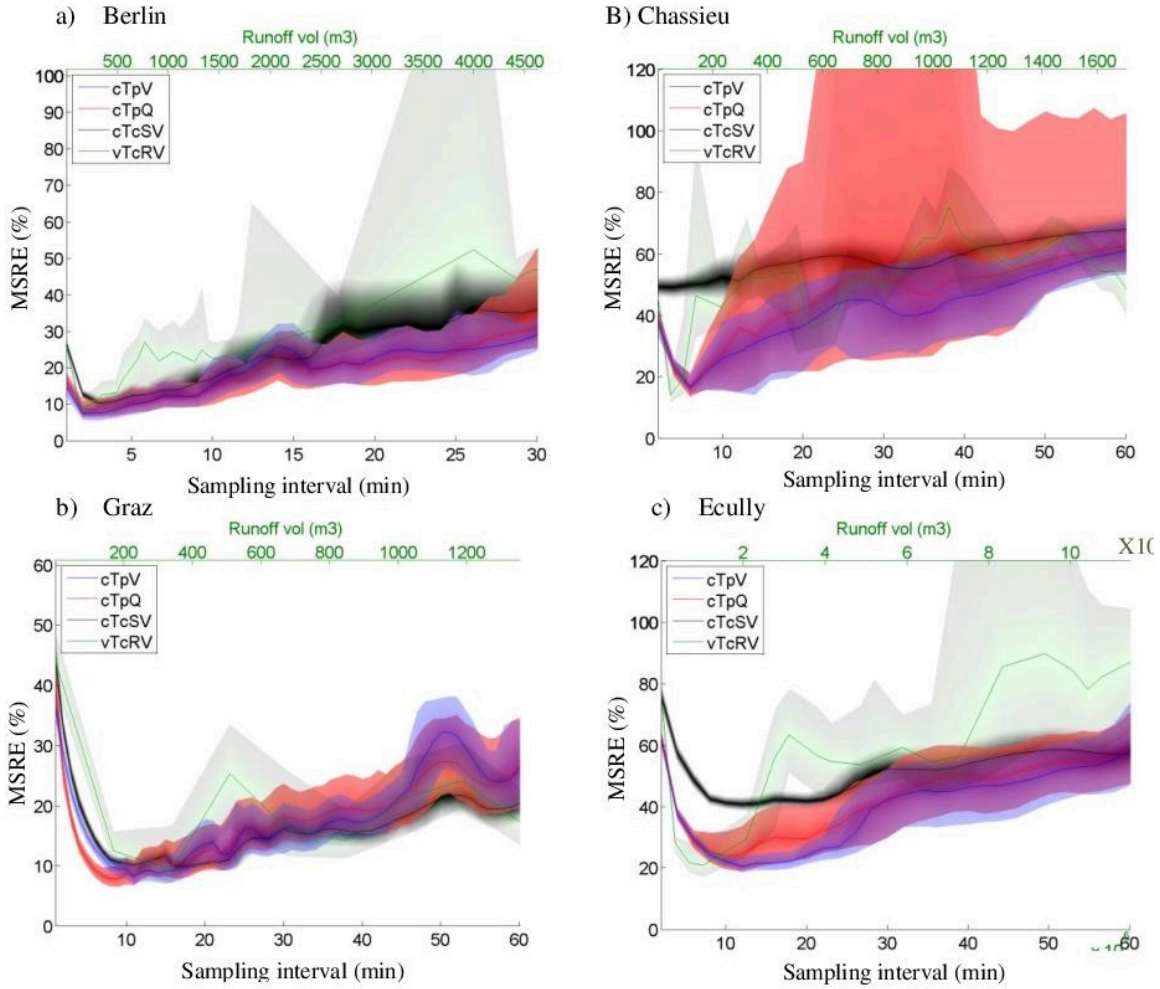


Figure 1. MSRE (ligne continue) et CI(MSRE) (bandes de couleur) pour les différents intervalles temporels d'échantillonnage pour les stratégies cTcSV (noir), cTpQ (rouge) and cTpV (bleu) sur l'axe horizontal inférieur et différents volumes d'échantillonages pour la stratégie vTcV (vert) sur l'axe horizontal supérieur pour a) Berlin, b) Chassieu, c) Graz et d) Ecully.

Conclusions

La stratégie d'échantillonnage la plus représentative pourrait être cTpV, en utilisant des intervalles d'échantillonnage d'à peu près 5 minutes pour Berlin et Chassieu resp. zones de 100 et 185 ha) et 10 minutes pour Graz et Ecully (resp. zones de 335 et 245 ha), plaçant MSRE entre 7 % et 20 % et CI(MSRE) vers 5 %.

CHAPITRE 3 Estimation de l'influence du point d'échantillonnage des matières en suspension dans une section de réseau d'assainissement

Introduction

Le mesurage des matières en suspension (MES) dans les systèmes d'assainissement urbains est essentiel pour des raisons scientifiques, environnementales, opérationnelles et réglementaires. Cependant, la validité des résultats de mesure (pour différents usages tels que le calage et la vérification de modèles, la gestion des ouvrages et l'autosurveillance) dépend étroitement de la qualité des données qui doivent être représentatives des conditions réelles *in situ*. Des méthodes spécifiques sont donc nécessaires pour garantir l'acquisition et la validation appropriées des concentrations en MES mesurées dans les prélèvements en réseau. Afin d'estimer la qualité des données, de nombreux travaux de recherche ont été consacrés à la détermination des incertitudes expérimentales sur les mesurages en ligne et en laboratoire des concentrations en MES (Harmel *et al.*, 2006; Harmel et Smith, 2007; Joannis *et al.*, 2008; Métadier et Bertrand-Krajewski, 2012). En revanche, l'influence des conditions d'échantillonnage *in situ* (par exemple la hauteur du point de prélèvement ou de mesurage dans la section, les vitesses d'aspiration ou l'orientation du tube de prélèvement) sur la représentativité des valeurs mesurées n'a pas fait l'objet d'autant d'investigations dans la littérature (Shelley, 1977 ; Berg, 1982 ; Rossi, 1998 ; Larrarte et Pons, 2011). Une approche préliminaire a été établie par Métadier (2011), sur la base de l'expérience des techniciens (comme 10 % de la valeur mesurée).

Les incertitudes associées à la localisation du point d'échantillonnage de MES dans la section transversale du collecteur sont habituellement négligées, en admettant que la valeur mesurée est égale à la concentration moyenne vraie dans la section. Cette hypothèse semble valide pour des collecteurs présentant des vitesses d'écoulement élevées susceptibles d'assurer une homogénéisation satisfaisante des concentrations en MES à travers la section. Cependant, le gradient vertical de concentration en MES ne peut pas être négligé pour d'autres conditions hydrodynamiques, notamment aux faibles vitesses. Lorsque c'est le cas, la différence entre la valeur mesurée et la concentration moyenne réelle peut être due à plusieurs sources : i) la variabilité de la position représentative de la concentration moyenne à travers la section, ii) la variabilité de la position du point de prélèvement dans la section transversale et iii) les incertitudes des variables physiques (*ex.* débit, coefficient de rugosité, propriétés géométriques) (Figure 1).

Nous proposons de traiter ces sources de variabilité en terme de probabilité, estimée par une approche de type Monte Carlo avec 1000 simulations. Il a été considéré le nombre de fois où la localisation du point d'échantillonnage sur le profil de concentration vertical (valeur mesurée) était égale à la position représentative de la concentration moyenne (sous l'hypothèse de l'homogénéité des profils horizontaux de concentration). La méthode proposée estime la représentativité des mesurages au moyen du facteur correctif *ROU* (ratio of over- or under-estimation) par lequel il faut multiplier la concentration mesurée pour obtenir la concentration moyenne vraie.

Matériels et méthodes

La méthode proposée a été testée avec les séries chronologiques de concentration *C* en MES, débit *Q* et hauteur d'eau *h* mesurées en 2007 pour 89 événements pluvieux, au pas de temps de 2 minutes, à l'exutoire du réseau séparatif pluvial de Chassieu, France. Les autres grandeurs nécessaires aux calculs sont i) la section mouillée au point de mesure, ii) le coefficient de Manning-Strickler K_{MS} du collecteur et iii) la vitesse de chute w_s des MES. Les incertitudes de toutes les variables sont également déterminées. Les variables, les conditions expérimentales et le profil de concentration en MES sont représentés schématiquement Figure 1.

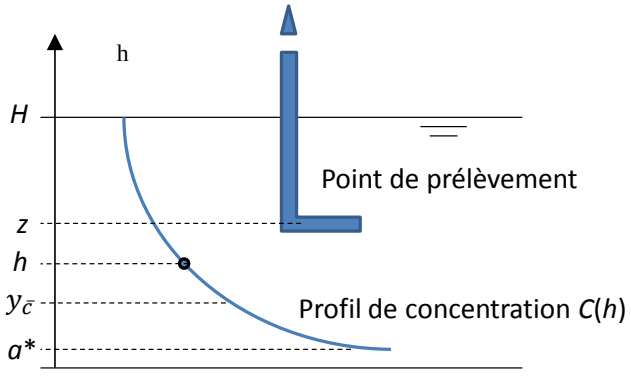


Figure 1. Schéma de principe avec les variables principales, le profil de concentration et le point de prélèvement, avec *H* la hauteur d'eau, *z* la hauteur de la prise d'échantillon, $y_{\bar{c}}$ la hauteur correspondant à la concentration moyenne sur le profil vertical de concentration *C*(*h*) et *a** la hauteur de référence du profil de concentration (voir détails Coleman, 1982; Verbanck, 2000).

Des études antérieures ont montré que la forme du profil vertical de concentration en MES, ainsi que la profondeur $y_{\bar{c}}$ correspondant à la concentration moyenne, dépendent fortement de la vitesse (Coleman, 1982; Verbanck, 2000). Un profil de concentration relativement uniforme est attendu pour les valeurs élevées de débit assurant des conditions de turbulence et de bon mélange. Néanmoins, pour des valeurs de débit plus faibles, la concentration en MES sera plus élevée près du radier (Coleman, 1982; Verbanck, 2000). Le problème consiste donc à estimer l'écart entre la concentration en MES mesurée sur un échantillon prélevé à la hauteur *z* et la concentration moyenne à la hauteur $y_{\bar{c}}$, sous les hypothèses théoriques proposées. La position *y* représentative de la concentration moyenne en MES est considérée comme une variable aléatoire avec $E[y] = y_{\bar{c}}$ et une fonction de densité de probabilité (pdf) fondée sur le profil vertical théorique de concentration proposé par Coleman (1982) et Verbanck (2000), avec les grandeurs *Q*, w_s, k_s, *h* et *D* permettant de calculer le nombre de Rouse η (voir détails dans Verbanck, 2000). La courbe *C*(*h*) est normalisée pour rendre son aire égale à 1 indépendamment de η. La probabilité $p(z = y_{\bar{c}})$ peut alors être calculée pour une hauteur d'échantillonnage *z* quelconque. Si $y_{\bar{c}}$ est proche au fond du collecteur (faibles valeurs du débit *Q*), une hauteur *z* plus proche de la surface libre supérieure à $y_{\bar{c}}$ conduit à une sous-estimation de la concentration en MES. Réciproquement, dans le cas où *z* est inférieure à $y_{\bar{c}}$, il y a surestimation de la concentration en MES. Avec la méthode de Monte Carlo, deux autres sources de variabilité sont prises en compte: i) la variation de la hauteur de prélèvement *z* supposée suivre une distribution uniforme sur la verticale entre 0.25 *H* et 0.75 *H*, et ii) les incertitudes sur les grandeurs physiques *Q*, w_s, k_s, *h* et *D* supposées normalement distribuées.

Les incertitudes de ces grandeurs permettent de calculer la distribution de probabilité du nombre de Rouse, lequel a un impact direct sur la variabilité de $y_{\overline{c}}$. Les résultats sont présentés en mettant en rapport l'évolution du ratio *ROU* et celle de la vitesse moyenne de l'écoulement U_m au cours des événements pluvieux, avec pour objectif de déterminer pour quelles valeurs de la vitesse d'écoulement le point d'échantillonnage conduit à des valeurs de concentrations en MES mesurées qui sur- ou sous-estiment la concentration moyenne vraie, et ceci pour l'ensemble des 89 événements pluvieux.

Résultats et discussion

La figure 2 montre les résultats pour la période du 20/08/2007 08:06 au 22/08/2007 01:38 extraite des séries chronologiques. La probabilité d'avoir une concentration mesurée égale à la concentration moyenne est proche de 1 pour des vitesses d'écoulement supérieures à 0.5 m/s (Figure 2 gauche). Ce résultat est conforme avec ceux d'études similaires (par exemple Larrarte, 2008). Toutefois, les valeurs de vitesse ont été supérieures à 0.4 m/s, ce qui n'a pas permis d'analyser comment se dégrade la représentativité (comme ROU) au fur et à mesure que la vitesse moyenne descend (en-dessous de 0.4 m/s). Un nombre de Rouse η inférieur à 0.6 est supposé correspondre à un mélange complet et à une concentration homogène sur la verticale, ce qui est la situation la plus fréquente pour les séries chronologiques étudiées ici (Figure 2 droite).

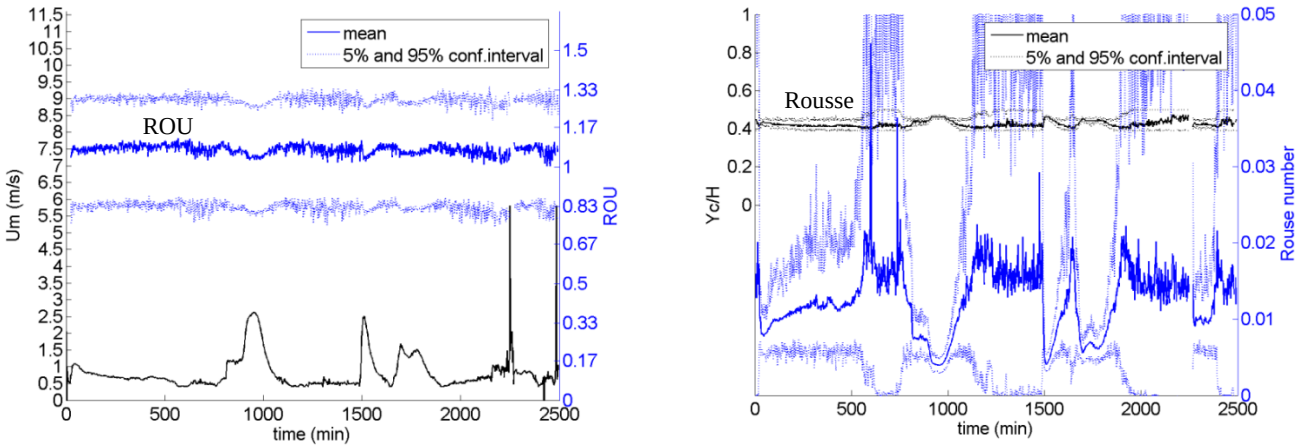


Figure 2. Pour la période du 20/08/2007 08:06 au 22/08/2007 01:38 – A gauche : *ROU* (moyenne et intervalle de confiance à 90 %) et vitesse moyenne d'écoulement U_m ; à droite : ratio $y_{\overline{c}}/H$ et nombre de Rouse η.

Une sous-estimation de la concentration en MES pour des valeurs de *ROU* comprises entre 1.04 et 1.10 se produit pendant 95 % de la durée des 89 événements pluvieux, particulièrement lorsque les vitesses d'écoulement sont faibles. La valeur médiane de *ROU* est égale à 1.07. Cette sous-estimation systématique de la concentration moyenne en MES peut être expliquée par le fait que l'espérance de *z* égale à 0.5 *H* est supérieure à l'espérance de $y_{\overline{c}}$ qui est de l'ordre de 0.4 *H*.

Pour les faibles valeurs du débit *Q* (principalement au début et à la fin des événements pluvieux), l'incertitude type *u*(η) est élevée et la valeur de $y_{\overline{c}}$ est déterminée avec une grande incertitude également. Une attention particulière doit donc être portée aux débuts et fins

d'événements pluvieux, périodes pendant lesquelles le risque est plus important d'observer une sous-estimation de la concentration en MES. Afin d'améliorer les mesures *in situ* les gestionnaires pourrions prendre en compte ces types d'incertitudes.

CHAPITRE 4 Modélisation pluie-débit : stratégie améliorée de calage et estimation des incertitudes guidée par les données

Introduction

Déterminer une stratégie robuste pour quantifier les incertitudes paramétriques et de calcul des débits reste un défi clé en hydrologie (Thyer *et al*., 2009; Ye *et al*., 2014). L'estimation des incertitudes des paramètres par le calage de plusieurs évènements pluvieux de manière chronologique est une stratégie courante (*e.g.* Tan *et al*., 2008; Thyer *et al*., 2009; Mancipe-Munoz *et al*., 2014). Cependant, le calage multi-évènementiel fournit un jeu moyen de paramètres, avec des incertitudes généralement sous-estimées, en négligeant la diversité des conditions hydrologiques possibles sur un bassin versant (Thyer *et al*., 2009). En effet, la valeur optimale des paramètres obtenue en faisant un calage évènement par évènement peut varier de manière significative en raison des changements stochastiques des conditions du bassin versant (Thyer *et al*., 2009; Ajmal *et al*., 2015). De plus, les sous-estimations systématiques des incertitudes paramétriques et de simulation des débits, obtenues habituellement en utilisant cette approche, sont dues à une simulation des débits à partir de jeux de paramètres optimaux pour quelques évènements qui ne sont pas nécessairement liés d'un point de vue hydrologique (adapté de Ye *et al*., 2014 et Ajmal *et al.*, 2015). Pourtant, cette stratégie de calage peut être utile pour explorer la variabilité inter-évènementielle (Thyer *et al*., 2009; Ajmal *et al*., 2015).

Pour améliorer la résolution des problèmes ci-dessus, nous proposons une nouvelle stratégie de simulation des débits, utilisant le calage évènement par évènement, dans laquelle les jeux de paramètres optimaux obtenus pour chaque évènement sont représentés par un graphe à partir de leur prédictibilité. Ensuite, la technique de cluster "spinglass.community" est employée pour former des groupes de paramètres (deux groupes pour le cas d'étude) dans le graphe précédent, à partir d'un critère de maximisation de la densité de connections dans chaque groupe de paramètres (voir Reichardt and Bornholdt, 2006). Ceci permet d'exprimer les incertitudes des paramètres en tant que fonctions de probabilité conditionnelle sur quelques caractéristiques générales d'un évènement donné (*e.g.* hauteur de pluie, intensité moyenne). L'estimation des paramètres d'un modèle pluie-débit et de leurs incertitudes à partir de la méthode proposée a comme objectif de réduire les bandes d'incertitude de simulation des débits, en conservant ou en améliorant la simulation moyenne des débits sur les évènements de vérification (évaluée par les indicateurs ARIL et NS respectivement, voir Dotto *et al*., 2012; Ye *et al.*, 2014), en comparaison avec d'autres stratégies traditionnelles de calage (voir Tan *et al*., 2008). La méthode proposée a été testée avec un modèle pluie-débit conceptuel (réservoir linéaire et infiltration de Horton), pour un bassin versant urbain de Lyon, France, avec 365 évènements pluvieux mesurés de 2004 à 2011 (séries chronologiques de pluie et débit).

Matériel et méthode

Les méthodes décrites ci-dessous ont été testées avec une base de données de 365 évènements pluvieux du bassin versant de Chassieu (Lyon, France) mesurés en 2004 et 2011. Il s'agit d'un des sites expérimentaux de l'OTHU (Observatoire de Terrain en Hydrologie Urbaine, www.othu.org). Le bassin versant est une zone industrielle de 185 ha, avec des coefficients d'imperméabilisation et de ruissellement de 0.72 et 0.43 respectivement (Métadier, 2011). Le débit observé Y_{obs} est estimé à partir de la hauteur d'eau à la sortie du bassin versant au pas de temps de 2 minutes et avec une incertitude élargie relative qui varie entre 15 % et 25 % (Métadier, 2011). Un modèle à réservoir linéaire, avec une constante de réservoir $K1$ (voir Sun et Bertrand-Krajewski, 2013), est utilisé pour représenter la relation pluie-débit (Eq 1, 2 et 3). La pluie nette X_{net} (L/s) est calculée à partir de la pluie observée X_{obs} (mm/h) par le modèle d'infiltration d'Horton, ayant comme paramètres les capacités d'infiltration du sol initiale et finale f_0 et f_c, et le taux de décroissance k (Eq 1 et Eq 2). Nous avons ajouté un paramètre q comme un terme additif pour représenter l'infiltration dans le réseau. La valeur de la surface active S est égale à 80 ha (0.43 × 185 ha) (Eq 3). La sélection de ce modèle particulier est basée sur sa performance et sa simplicité, comme suggéré par des études précédentes sur la même base de données (Sun et Bertrand-Krajewski, 2013). Les paramètres de ce modèle conceptuel sont décrits dans le Tableau 1, ainsi que les valeurs maximales et minimales possibles envisagées pour le calage (selon des expériences précédentes et des valeurs raisonnables déjà proposées dans la littérature, *e.g.* Sun et Bertrand-Krajewski, 2013).

$$f = f_c + (f_c - f_0) \cdot e^{-kt} \quad \text{Eq 1}$$

$$X_{net} = (X_{obs} - f) \cdot S \cdot 10000/3600 \quad \text{Eq 2}$$

$$Y_{sim}(t) = e^{-\frac{\Delta t}{K1}} \cdot Y_{sim}(t - \Delta t) + \left[1 - e^{-\frac{\Delta t}{K1}}\right] X_{net}\left(t - T_{lag}\right) + q \quad \text{Eq 3}$$

Tableau 1. Liste des paramètres utilisés pour le calage du modèle

Paramètre (θ)	Unité	Valeurs [min, max]
f_0	mm/h	[0, 50]
f_c	mm/h	[0, 5]
k	min^{-1}	[0, 5]
T_{lag}	min	[0, 60]
$K1$	min	[1, 120]
q	L/s	[0, 20]

On note θ le jeu de paramètres du modèle hydrologique (Tableau 1) et $p(\theta/Y)$ sa fonction de densité de probabilité (pdf), étant données une série d'observations du débit Y_{obs}. Par ailleurs, la méthode bayésienne a été testée dans de nombreux cas en modélisation hydrologique. Elle permet de calculer $p(\theta/Y)$ (*e.g.* Thyer *et al.*, 2009). La base de la méthode pour estimer $p(\theta/Y)$, appelée distribution *a posteriori*, repose sur une fonction de vraisemblance et une connaissance *a priori* de la distribution $p(\theta)$ des paramètres, que nous pouvons exprimer comme suit :

$$p(\theta/Y) = C \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\hat{\sigma}_t^2}} \exp[-\frac{1}{2}(\frac{Y_{sim}(t,\theta) - Y_{obs}(t)}{\hat{\sigma}_t^2})^2] \cdot P(\theta) \qquad \text{Eq 4}$$

où n est le nombre de mesures de débit Y_{obs}, $Y_{\mathrm{sim}}(t, \theta)$ est le débit simulé par le modèle à un instant t à partir de la pluie observée X_{obs} et du jeu de paramètres θ, $p(\theta)$ est une loi de probabilité uniforme pour chaque paramètre (à partir des intervalles [min, max] du Tableau 1), C est un coefficient de normalisation et $\hat{\sigma}_t^2$ est la variance des résidus, considérée égale à l'incertitude standard au carré de la valeur du débit $Y_{obs}(t)$. L'algorithme DREAM est utilisé pour déterminer $p(\theta/Y)$ (Vrugt *et al.*, 2008). Le jeu de paramètres représentant la valeur optimale (vraisemblance maximum) parmi toutes les valeurs probables $p(\theta/Y)$ est appelée θ_{opt}. Le but est de déterminer une estimation de $p(\theta/Y)$ et θ_{opt} en fonction des observations de calage X_{obs} et Y_{obs} qui maximisent la capacité de simulation moyenne des débits dans la phase suivante de vérification (mesurée à partir de l'indicateur de Nash-Sutcliffe NS), en minimisant la largeur des bandes d'incertitude moyenne des débits simulées, largeur déterminée par l'indicateur ARIL des paramètres - *Average Relative Interval Length*) (voir *e.g.* Ye *et al.*, 2014). Les indicateurs NS et ARIL liés au jeu de paramètres θ_{opt} sont calculés pour chaque évènement de vérification à partir des Eq 5 et Eq 6.

$$NS(\theta_{opt}) = 1 - \frac{\sum_{t=1}^{n}\left(Y_{sim}(t,\theta_{opt}) - Y_{obs}(t)\right)^2}{\sum_{t=1}^{n}\left(\bar{Y}_{obs} - Y_{obs}(t)\right)^2} \qquad \text{Eq 5}$$

$$ARIL(\theta_{opt}) = \frac{1}{n}\sum_{t=1}^{n} \frac{Limite_{sup,t} - Limite_{inf,t}}{Y_{sim}(t,\theta_{opt})} \qquad \text{Eq 6}$$

où n est le nombre de mesures de débit Y_{obs} et $\bar{Y}_{obs}$ leur moyenne. $Y_{\mathrm{sim}}(t, \theta)$ est le débit simulé par le modèle à un instant t à partir de la pluie observée X_{obs} et du jeu de paramètres θ_{opt}. $Limite_{sup,t}$ et $Limite_{inf,t}$ sont, respectivement, les limites supérieure et inférieure pour un intervalle de confiance à 95 % à un instant t à partir de $p(\theta/Y)$ ou $p(\theta/Y, T1)$ et $p(\theta/Y, T2)$ selon le cas.

La fonction $p(\theta/Y)$ est calculée par trois approches, en utilisant 255 évènements pluvieux de calage choisis de manière aléatoire parmi les 365 évènements disponibles :

- Calage évènement par évènement (SE) : les évènements sont calés séparément par la méthode bayésienne. Ceci permet d'obtenir un jeu optimal de paramètres $\theta_{opt\,i}$ et une estimation de $p(\theta/Y)_i$ pour chaque évènement de calage i (i = 1:255). La fonction $p(\theta/Y)$ globale est calculée comme la probabilité marginale de toutes les fonctions $p(\theta/Y)_i$ et θ_{opt} comme la moyenne des $\theta_{opt\,i}$.

- Calage multi-évènementiel (ME) : le calage est effectué globalement en une seule fois avec l'ensemble des évènements pluvieux (simulation chronologique de 255 évènements) pour obtenir directement un seul jeu de paramètres θ_{opt} et une fonction globale $p(\theta/Y)$.

- Calage évènementiel par cluster (SEClusters): le groupe de 255 jeux de paramètres optimaux $\theta_{opt\,i}$ obtenus avec l'approche SE sont classés dans deux types T1 et T2, avec l'objectif de regrouper les évènements qui ont des caractéristiques hydrologiques similaires. L'hypothèse est que deux évènements i et j (i = 1:255 et j = 1:255) sont connectés si le jeu optimal de paramètres $\theta_{opt\,i}$ obtenu pour l'évènement i est capable de reproduire également l'évènement j et si $\theta_{opt\,j}$ est lui aussi capable de reproduire l'évènement i, dans les deux cas avec un indicateur de Nash-Sutcliffe NS > 0.75 fixé comme valeur seuil. Une matrice de connectivité symétrique AM est remplie avec AM(i, j) = 1 si les évènements de calage i et j sont connectés et AM(i, j) = 0 sinon (pour $i = j$, AM(i, j) = 0 par convention). La technique de clustering est appliquée à AM pour identifier les groupes d'évènements connectés. La fonction "spinglass.community" avec spins = 2 (cluster supervisé de deux groupes) (voir Reichardt et Bornholdt, 2006) du package "igraph" (Csardi et Nepusz, 2006), implémentée sur R (R Development Core Team, 2016), est utilisée pour classer les évènements en type T1 ou T2. La fonction $p(\theta/Y)$ est calculée de la même manière que dans l'approche SE, mais elle est répartie en deux pdf conditionnelles $p(\theta/Y, T1)$ et $p(\theta/Y, T2)$ en utilisant les types T1 et T2. θ_{opt} est défini également dans deux valeurs $\theta_{opt}T1$ et $\theta_{opt}T2$, calculées comme la moyenne des θ_{opt} dans chaque groupe T1 et T2. La catégorisation d'un évènement pluvieux de vérification dans un groupe hydrologique (T1 ou T2), pour décider s'il faut utiliser $p(\theta/Y, T1)$, $\theta_{opt}T1$ ou $p(\theta/Y, T2)$, $\theta_{opt}T2$ dans la simulation des débits, est effectuée avec un modèle de classement Kernel KCM (fonction fitcsvm sur Matlab, Cristianini and Shawe-Taylor, 2000) en utilisant comme variables d'entrée certaines caractéristiques de la pluie.

Résultats et discussion

Dans le cas de SEClusters, la technique de clustering a été appliquée sur la matrice AM, en classant les évènements de calage dans deux groupes (32 % et 36 % des évènements pour T1 et T2 respectivement). Ce classement a permis le calcul de $p(\theta/Y, T1)$ et $\theta_{opt}T1$ d'une part, et de $p(\theta/Y, T2)$ et $\theta_{opt}T2$ d'autre part. Un KCM a été proposé pour décider s'il faut utiliser $p(\theta/Y, T1)$, $\theta_{opt}T1$ ou $p(\theta/Y, T2)$, $\theta_{opt}T2$ dans la simulation des débits d'un évènement de vérification donné, en le désignant comme étant de type T1 ou T2. Les caractéristiques suivantes de la pluie ont été retenues dans le KCM comme variables d'entrée pour chaque évènement de calage (choisies à partir d'essais préalables) : intensité moyenne [mm/h], intensité maximum [mm/h], durée [min], hauteur totale de pluie [mm], nombre d'intensités > 0 mm/h [# données], accélération moyenne de la pluie [mm/h^2], accélération maximum de la pluie [mm/h^2], accélération minimum de la pluie [mm/h^2] et hauteur moyenne [mm]. Le KCM a donné un pourcentage d'assertivité de classement sur les évènements de calage d'environ 80 %.

La vérification est exécutée avec les 110 évènements restants pour les trois approches proposées (SE, ME et SEClusters). Chaque évènement de vérification a pu être classé T1 ou T2 à partir du KCM selon ses caractéristiques (*i.e.* intensité moyenne, intensité maximum, etc.). Les valeurs ARIL des paramètres et de NS sont calculées à partir des simulations de débit faites avec les pdfs $p(\theta/Y)$ et θ_{opt} pour les approches SE ou ME, et avec les pdfs $p(\theta/Y, T1)$ ou $p(\theta/Y, T1)$ et $\theta_{opt}T1$ et $\theta_{opt}T2$ pour l'approche SEClusters (Figure 1).

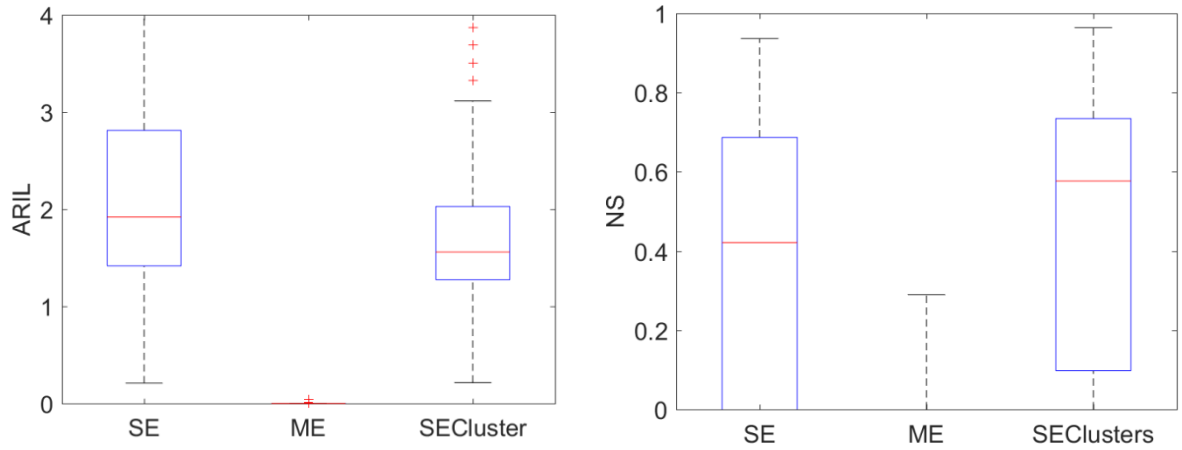


Figure 1. Indicateurs ARIL et NS pour les 110 évènements de vérification, pour les 3 approches de calage étudiées.

La stratégie de calage SEClusters montre une réduction des bandes d'incertitude dans la simulation des débits (valeurs ARIL des paramètres) par rapport à la stratégie traditionnelle de calage SE de 2 à 1.6, pour 50 % des évènements de vérification (Figure 1, gauche). La simulation moyenne des débits montre aussi une amélioration des valeurs de Nash-Sutcliffe (NS) de 0.4 à 0.6 pour 50 % des évènements (Figure 1, droite). Les valeurs ARIL montrent que les bandes d'incertitude des débits obtenues en utilisant l'approche ME sont beaucoup plus étroites (valeur proche de zéro) que celles obtenues à partir des autres stratégies (SE et SEClusters) (Figure 1, gauche). Dans ce cas, la plupart des valeurs de débit en vérification se trouvent en dehors des limites des bandes d'incertitude de simulation des débits. D'autre part, les bandes d'incertitudes obtenues avec les approches SE et SEClusters incluent environ 95 % des débits de vérification, même si pour l'approche SE les valeurs ARIL dépassent 2 (largeur moyenne des intervalles d'environ 200 % des valeurs de débit), ce qui pourrait être considéré comme une surestimation des incertitudes des débits simulés.

Conclusions

Ces résultats sont considérés comme satisfaisants pour utiliser le KCM déjà calé pour la vérification. La stratégie de calage proposée fournit une estimation des bandes d'incertitude des débits simulés à partir des incertitudes des paramètres expliquant environ 95 % des débits mesurés. La simulation moyenne des débits montre aussi une amélioration de l'indicateur Nash-Sutcliffe (NS) de 0.4 à 0.6 pour 50 % des évènements. Une même structure de modèle pluie-débit permet de représenter deux groupes de conditions hydrologiques différentes pour un bassin versant urbain au moyen de la variabilité des paramètres optimaux pour tous les évènements pluvieux.

CHAPITRE 5 Identification d'erreurs dans des séries pluviométriques a haute résolution temporelle à travers des approches fondées sur des modèles conceptuels

Contexte et objectifs

Des modèles mathématiques simples peuvent constituer une description adéquate du processus de l'écoulement des eaux dans les bassins versants urbains, quand les paramètres sont bien identifiés à partir d'analyses antérieures ou d'expériences. Cependant, les situations dans lesquelles des modèles étalonnés performants reproduisent des valeurs de débit irrégulières sont dans beaucoup de cas dues aux erreurs dans la saisie des données de précipitations. Ces erreurs peuvent avoir différentes origines, un cas significatif étant l'utilisation de données pluviométriques locales comme des entrées directes, sans considérer la variabilité spacio-temporelle des précipitations (Kavetski *et al.*, 2006). Sur base de cette hypothèse, des études récentes ont proposé l'estimation nouvelle (Modelisation Inverse) ou corrigée (Approche Bayésienne) de séries chronologiques de pluies représentatives (Leonhardt *et al.*, 2014). Ces séries devraient permettre la reproduction des valeurs de débit mesuré par un modèle hydrologique bien calibré. Néanmoins, les composants systématiques et accidentels des erreurs de mesure de précipitations ne sont pas connus à l'avance, et leur structure peut être complexe et variable. La méthodologie proposée ici cherche à évaluer le potentiel des modèles de correction de précipitations pour identifier et corriger les erreurs dans les données de pluie. Elle a été appliquée à un bassin versant de Lyon, France, avec des enregistrements de 30 événements pluvieux de 2007 à 2008 (pluviomètre et série chronologique de débit, pas de temps de 1 et 2 min respectivement).

La méthode bayésienne a été appliquée en utilisant une équation de de correction d'erreurs générales (Vrugt *et al.*, 2008; Leonhardt *et al.*, 2014).

$$I_{corr}=K_i * I_{mesuré} \qquad \text{Eq. 1}$$

La série chronologique d'intensité de précipitations mesurée I (mm/h) peut être divisée en une fenêtre temporelle I [a, b]$_i$ (de taille égale ou inégale) avec l'index i, et les précipitations dans chaque intervalle sont corrigées par l'Eq. 1 avec le facteur de correction associé K_i . Eq. 1 n'est pas en mesure de corriger les précipitations quand $I_{mesuré} = 0$. Comme cette situation est fréquente dans les événements pluvieux, un autre modèle de correction a été proposé:

$$I_{corr} = \begin{cases} K_i * I_{measured} & if \quad I_{measured} > 0 \\ K_i * I_{rev} & if \quad I_{measured} = 0 \end{cases} \qquad \text{Eq. 2}$$

où I est l'intensité de précipitations obtenue à travers un modèle inversé (Leonhardt *et al.*, 2014). Ceci génère une intensité pluviale probable pour quand elle n'avait pas été mesurée. De plus, un nombre n de fenêtres de tailles égales ou inégales I [a, b]$_i$ est testé pour toute série I chronologique d'intensité pluviale mesurée. Taille égale: divisez en parts égales n et appliquez K_i à chaque I [a, b]$_i$ avec i = 1:n (Eq. 1 ou Eq. 2). Taille non-égale: évaluez les signaux de résidus de débit Q_{res} (différence entre les débit mesurés et simulés) avec un algorithme Détecteur de Pas (Canny, 1986). La longueur et l'emplacement de chaque fenêtre

temporelle dans Q_{res} [a, b]$_i$ peut être projetée dans la série de précipitations I [a, b]$_i$ par mise à l'échelle, basée sur la durée totale de l'intensité pluviale I et la série chronologique de ruissellement Q. K_i est ensuite appliquée à chaque fenêtre I [a, b]$_i$ (Eq. 1 ou Eq. 2).

Quatre modèles de correction de précipitations ont été étudiés: 1) CTW: correction traditionelle sur des fenêtres temporelles de taille égale (en utilisant l'Eq. 1), 2) VTW: correction sur des fenêtres de taille inégale (en utilisant l'Eq. 1), 3) CTWrev: correction associée à un modele inversé sur des fenêtres temporelles de taille égale (en utilisant l'Eq. 2) et 4) VTWrev: correction associée à un modele inversé sur des fenêtres temporelles de taille inégale (en utilisant l'Eq. 2). Trente scénarios d'erreur ont été vérifiés avec la méthode Monte Carlo, en introduisant des diviseurs $K_{j\ intro}$ ou des zéros sur des segments aléatoires avec l'index j dans la série chronologique de mesure de précipitations d'origine (présumées être des erreurs connues ou contrôlées). Les quatre modèles ont été testés pour évaluer leur capacité à détecter les erreurs contrôlées générées et à reconstruire la série chronologique de précipitations d'origine.

Résultats

Concernant l'analyse d'un événement unique, les résultatas sont présentés pour l'événement pluvieux mesuré entre le 11/02/2007 à 23:06 au 12/02/2007 à 07:12. La Figure 1 compare l'hydrographe mesuré (bleu), l'hydrographe produit par les précipitations avec des erreurs générées (noir) et avec les précipitations corrigées (rouge) par le modèle CTW. La Figure 2 montre les différences entre les erreurs identifiées et controllées, K_{recons} et K_{intro}. La structure d'erreurs (facteurs K) a été prédite raisonnablement dans la durée puisque $K_{recons} - K_{intro} \approx 0$. En partant d'une analyse générale à l'échelle de plusieurs événements (30 événements), les valeurs de *NS* et *RMSE* (ajustement entre les précipitations d'origines et les précipitations corrigées) pour tous les événements pluvieux et tous les modèles sont présentées dans les Figures 3 et 4. Les meilleurs résultats ont été obtenus avec le modèle CTWrev pour la justesse (*RMSE* les plus bas) comme pour la précision (*NS* le plus bas, Figure 4) et la conservation de la masse.

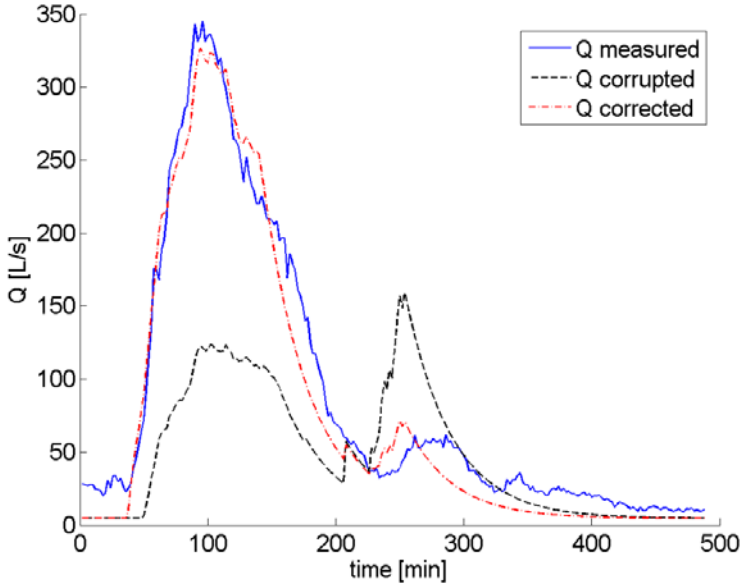


Figure 1. Hydrographes: mesuré (bleu), produit par des précipitations corrompues (noir) et produit par des précipitations corrigées (rouge)

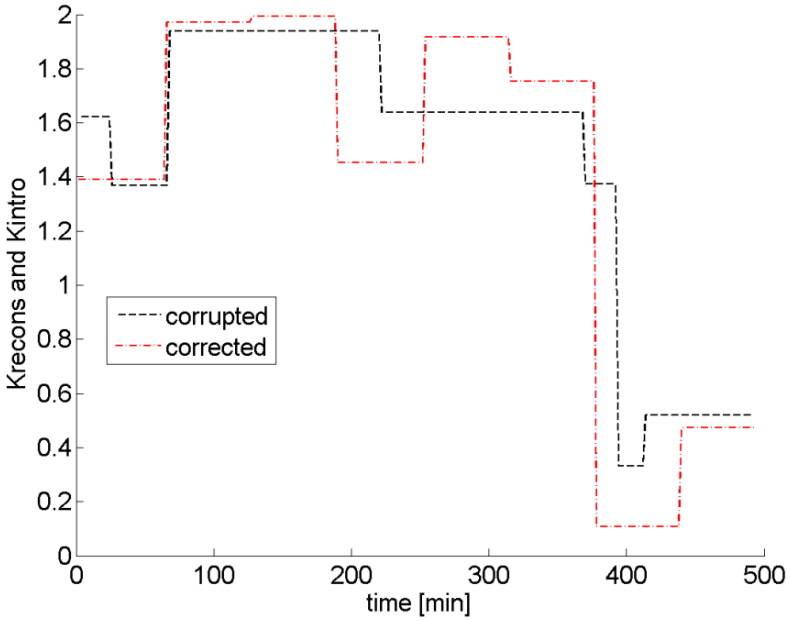


Figure 2. Simulation performante du facteur K: K_{recons} (rouge) et K_{intro} (noir)

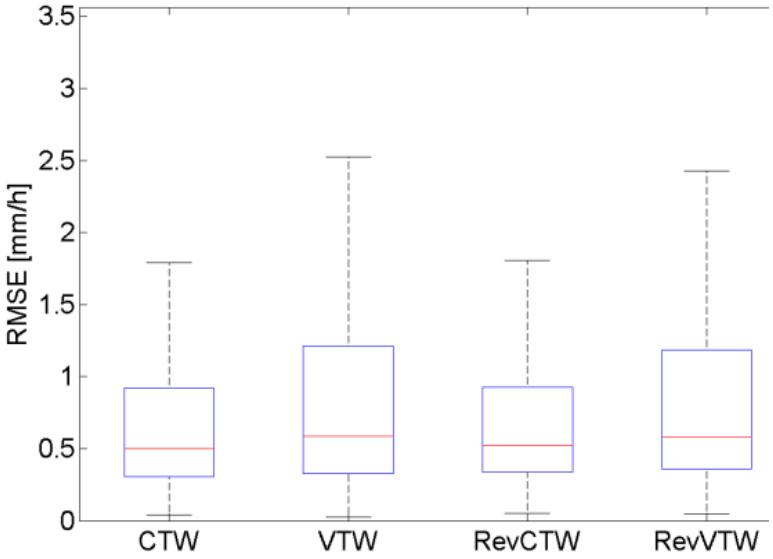


Figure 3. indicateur *RMSE* pour les quatre modèles de correction de précipitations

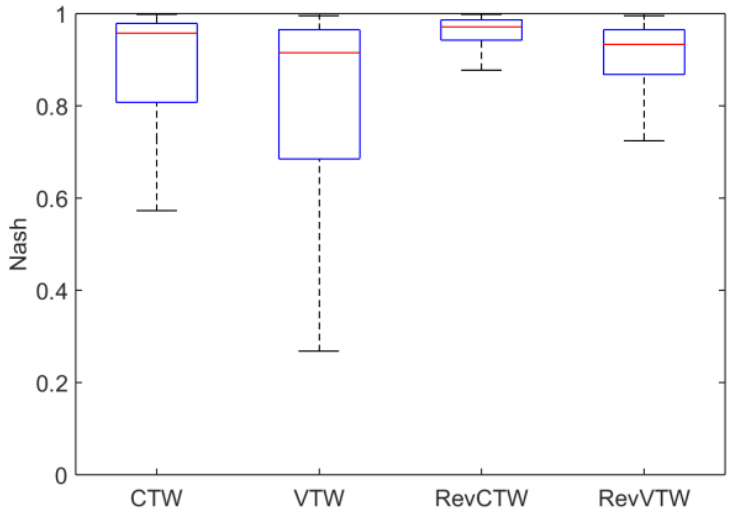


Figure 4. indicateur *NS* pour les quatre modèles de correction de précipitations

Conclusions principales

L'étude propose une approche nouvelle pour évaluer le potentiel de quatre modèles de correction de précipitations, en termes de l'identification et la description de différentes structures d'erreurs sous-jacentes dans dans des données pluviométriques. Trois nouveaux modèles de correction d'erreurs ont été formulés dans des travaux précédents (Kavetski *et al.*, 2006; Leonhardt *et al.*, 2014), ainsi que la mise en oeuvre d'un algorithme de Détection de Pas de temps. La structure d'erreur a été raisonnablement prédite dans la durée par les modèles de correction de précipitations testés. Cependant, le modèle le plus simple a fonctionné mieux que les autres.

CHAPITRES 6 et 7. Modèles conceptuels de qualité d'eaux pluviales: une révision à travers la reconstruction de variables d'état virtuelles

Introduction

Durant les 40 dernières années, le modelage des dynamiques des charges de Matière en suspension (TSS) dans les eaux pluviales au débouché de bassins de versant urbains a été discuté en majorité vis à vis de l'idée d'accumulation/érosion transfère (Sartor *et al.*, 1974). Une grande quantité de formulations de modèles analogues ont été proposées et testées, et dont les constats peuvent être presque impossibles à généraliser pour des applications dans la réalité, à cause des restrictions suivantes dans les paramètres expérimentaux/méthodologiques (Bonhomme and Petrucci, 2017): (i) des conditions de laboratoires contrôlées non-représentatives, (ii) des nombres limités de données TSS, (iii) des nombres limités d'événements pluvieux, et (iv) l'évaluation insuffisante de l'incertitude dans les données et dans les paramètres de modèles.

La structure de modèle de pollutographe la plus simple trouvée dans la documentation est probablement la courbe d'étalonnage RC (Sartor *et al.*, 1974), dans laquelle la charge du débouché du bassin versant (kg/s) à l'instant t (s) est calculée par une relation non-linéaire au débit Q (m^3/s) et aux paramètres d'ajustement M et r:

$$charge(t) = M \cdot Q(t)^r$$

La deuxième grande famille de structures de modèles (que nous nommerons ACUM) représente $M(t)$, plutôt qu'avec une valeur constante, comme une variable d'état décomposée d'un stock de masse virtuel disponible, qui au final limitera la production de charge donnée par $Q(t)^r$.

D'autre part, le cadre de calibration bayésien a émergé comme une approche basée sur modèle pour la reconstruction d'entrées non-mesurées ou de variables d'état (*e.g.* Leonhardt *et al.*, 2014). Pour 255 événements pluvieux, une reconstruction bayésienne de la variable d'état virtuelle $M(t)$ est proposée, à travers le modèle RC (avec r comme paramètre de calibration inclus dans le schéma d'inférence) (formulation F1). Par ailleurs, une formulation alternative F2 est explorée, dans laquelle M est un paramètre constant et $\boldsymbol{r}(t)$ est défini comme la variable d'état virtuelle à être reconstruite, en considérant la même structure de modèle RC.

Matériels et méthodes

Les deux formulations d'inférence différentes (F1 et F2) sont testées en utilisant la charge TSS et les données de débit (pas de temps de 2 min) d'un bassin versant urbain séparé de 85 ha à Lyon, France, collectées de 2004 à 2011. Les paramètres de modèle (r pour F1, M pour F2) et variables d'état ($M(t)$ pour F1, $r(t)$ pour F2) sont estimés pour chaque événement pluvieux avec une approche bayésienne (algorithme DREAM, cf. Vrugt *et al.*, 2009), en séparant explicitement les erreurs d'estimation de la variable de débouché (load(t)) et les variables d'état reconstruites ($M(t)$ ou $r(t)$) dans la fonction de probabilité (de Leonhardt *et al.*, 2014). L'erreur de modèle pour $M(t)$ ou $r(t)$ est prétendue égale à son propre standard de déviation ($\sigma M(t)$ ou $\sigma r(t)$), avec l'objectif de permettre aux variables d'état de varier moins librement dans l'espace d'inférence, ce qui les rapproche de la moyenne (reconstruction moins informative).

Résultats

Dans un objectif d'illustration, la simulation du pollutographe de charge TSS obtenue par la calibration d'un modèle RC traditionnel est comparée aux reconstructions F1 et F2 et aux valeurs mesurées dans la Figure 1a (vert, bleu, rouge et gris resp. avec 95 % de bornes de confiance) dans un exemple d'événement pluvieux. Des reconstructions des variables d'état virtuelles $M(t)$ et $r(t)$ obtenues par F1 and F2 sont présentées dans les Figures 1b et 1c (bleu et rouge, resp. avec 95 % de bornes de confiance).

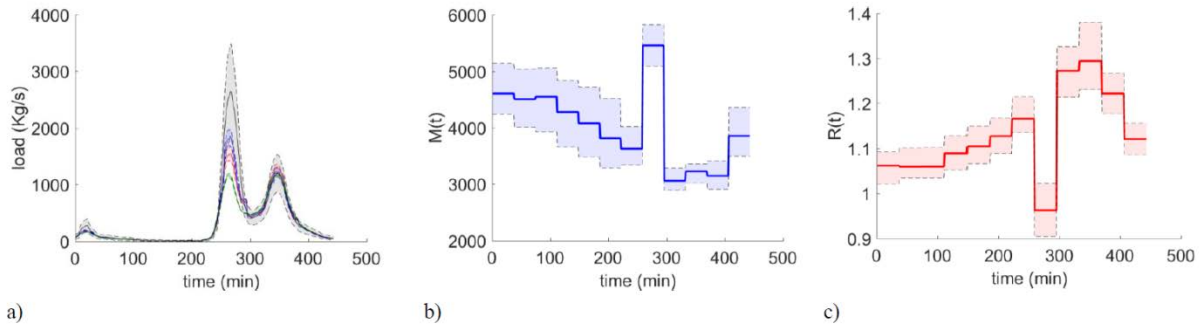


Figure 1. a) simulation de pollutographe de charge TSS avec 95 % de bornes de confiance pour le modèle traditionnel RC (vert), reconstructions F1 (bleu) et F2 (rouge), et données expérimentales (noir), b) variables d'état virtuelles estimées pour $M(t)$ (bleu) et c) $r(t)$ (rouge) avec 95 % de bornes de confiance.

Pour cet exemple, F1 comme F2 permettent d'améliorer considérablement le résultat (Nash = 0.85, lignes bleues et rouges dans la Figure 1a), en comparaison avec la calibration traditionnelle RC (Nash = 0.65, ligne verte dans la Figure 1a). Des analyses de regroupement montrent la faible possibilité de reproduction de ces courbes temporellement variables $M(t)$ ou $r(t)$ concernant: (i) les similarités dans leur forme (contrairement à des formulation traditionnelles ACUM, où $M(t)$ est toujours une fonction en décomposition) et (ii) leur capacité de prédiction parmi d'autres événements pluvieux (au sens de la similarité entre une courbe donnée $M(t)$ ou $r(t)$ pour expliquer les manques potentiels du modèle RC pour reproduire un autre événement de l'ensemble des données).

Conclusions

Ce travail suggère le manque de représentation d'un processus essential dans le modèle de courbe d'étalonnage (RC) traditionnel, en se fondant sur les observations de 255 événements pluvieux. Les résultats indiquent que la difficulté de reproduction de ce processus manquant le rend pratiquement ininterprétable en termes d'un unique état virtuel de masse disponible dans le bassin de versant qui diminue avec le temps, comme nombre de modèles traditionnels l'ont supposé. Cette étude démontre comment des mesures à haute résolution temporelle peuvent fournir un support pour revisiter et remettre en question des modèles existants, et potentiellement permettre le développement de nouvelles formulations de modèles d'accumulation/érosion transfère.

REFERENCES RESUME ETENDU

Ackerman, D., Stein, E., Ritter, K. (2011). Evaluating performance of stormwater sampling approaches using a dynamic watershed model. *Environmental Monitoring and Assessment*, 180, 283-302.

Ajmal, M., Waseem, M., Ahn, J.H., and Kim, T.W. (2015). Improved runoff estimation using event-based rainfall-runoff models. *Water Resources Management*, 29(6), 1995-2010.

Bonhomme, C., Petrucci, G. (2017). Should we trust build-up/wash-off water quality models at the scale of urban catchments? *Water Research*, 108, 422-431.

Burnham, K. P., Anderson, D. R. (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. New York (USA): Springer, 488 p.

Canny, J. (1986). A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(6), 679-698.

Kavetski, D., Kuczera, G., Franks, S. W. (2006). Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. *Water Resources Research*, 42(3).

Ki, S.J., Kang, J.H., Lee, S.W., Lee, Y.S., Cho, K.H., An, K.G., Kim, J.H. (2011). Advancing assessment and design of stormwater monitoring programs using a self-organizing map: Characterization of trace metal concentration profiles in stormwater runoff. *Water Research*, 45(14), 4183-4197.

Lee, H., Swamikannu, X., Radulescu, D., Kim, S.J., Stenstrom, M.K. (2007). Design of stormwater monitoring programs. *Water Research*, 41(18), 4186-4196.

Leonhardt G., Sun S., Rauch W., Bertrand-Krajewski J.-L. (2014). Comparison of two model based approaches for areal rainfall estimation in urban hydrology. *Journal of Hydrology*, 511, 880-890.

Ma, J.A., Kang, J.H., Kayhanian, M., Stenstrom, M.K. (2009). Sampling issues in urban runoff monitoring programs: Composite versus grab. *Journal of Environmental Engineering*, 135(3), 118–127.

Mancipe-Munoz, N.A., Buchberger, S.G., Suidan, M.T., and Lu, T. (2014). Calibration of rainfall-runoff model in urban watersheds for stormwater management assessment. *J. of Water Resources Planning and Management*, 140(6), 05014001.

R Development Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Sartor, J.D., Boyd, G.B., Agardy, F.J. (1974). Water pollution aspects of street surface contaminants. *J. Water Pollut. Control Fed.*, 46, 458-467.

Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, S.W., and Srikanthan, S. (2009). Critical evaluation of parameter consistency and predictive uncertainty in hydrological

modeling: A case study using Bayesian total error analysis. *Water Resources Research*, 45, W00B14, doi:10.1029/2008WR006825.

Vrugt, J. A., Ter Braak, C. J., Clark, M. P., Hyman, J. M., Robinson, B. A. (2008). Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resources Research*, 44(12).

Vrugt, J.A., ter Braak, C.J.F., Diks, C.G.H., Robinson, B.A. (2009). Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches to hydrologic modeling. *Stoch. Env. Res. Risk A.*, 23(7),1011-1026

Ye, L., Zhou, J., Zeng, X., Guo, J., and Zhang, X. (2014). Multi-objective optimization for construction of prediction interval of hydrological models based on ensemble simulations. *J. of Hydrol.*, 519, 925-933.

2. APPLICATION OF THE SOBOL'S SENSITIVITY INDEXES

A description of this methodology and its computational implementation in Chapter 2 for estimating the influence of each uncertainty source over the uncertainty of *MSRE* (*i.e.* $CI(MSRE)_{min95\%}$) is given in the following lines. Each of the variables in the calculation of *MSRE* in Table 4 (uncertainty sources from 1 to 6, Num_unc_inp = 6) can be represented as a as a time series. The LHS can be applied to generate random time series for each of these variables, following the probability distributions indicated in Table 4. A certain number of randomly generated time series (Num_sim = 600) can be split into two sub-matrices (UP and DOWN), where each one corresponds to half of the simulations (Num_sim/2) (Figure 1).

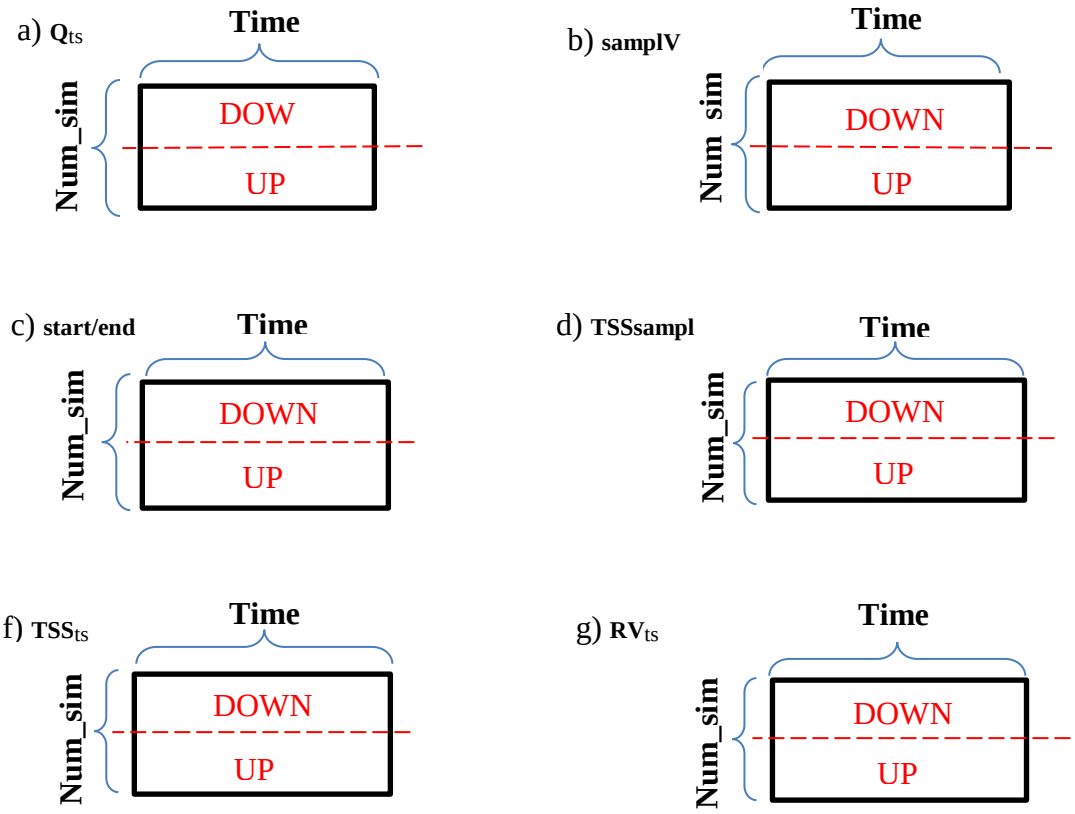


Figure 1. Data organization for matrices from LHS random generation of: a) flow rate time series, b) sampling volumes, c) start/ending of event, d) TSS lab. values, e) TSS time series and f) volume time series.

The SSI requires the comparison between different scenarios (given by the UP and DOWN matrices in Figure 1). Therefore, for different combinations of the UPs and DOWNs sub matrices (an UP or DOWN for each uncertainty source), one will have an output array of length Num_sim/2, containing the *MSRE* computed for each randomly generated time series (simulations in rows, Figure 1). A "combination matrix" that contains all the required combinations of UP and DOWN sub matrices for computing the SSI is summarized in Figure 2. The combination matrix is made by: (i) two reference rows (first and last) that are made by exclusively to the DOWN and UP labels (without mixing them) and (ii) the intermediate rows, in which a given scenario i (where $2 \leq i \leq$ Num_scenarios/2) has an "opposite" row in the position i + Num_unc_inp (*e.g.* row 4 is the opposite of row 11 in Figure 2, left) (see more details in Glen and Isaacs, 2012). Therefore, the result of the *MSRE*s (length of Num_sim/2) for each scenario of Ups and DOWNs sub matrices (rows in the combination matrix) are grouped in a *MSRE SR* matrix, as shown in Figure 2, right.

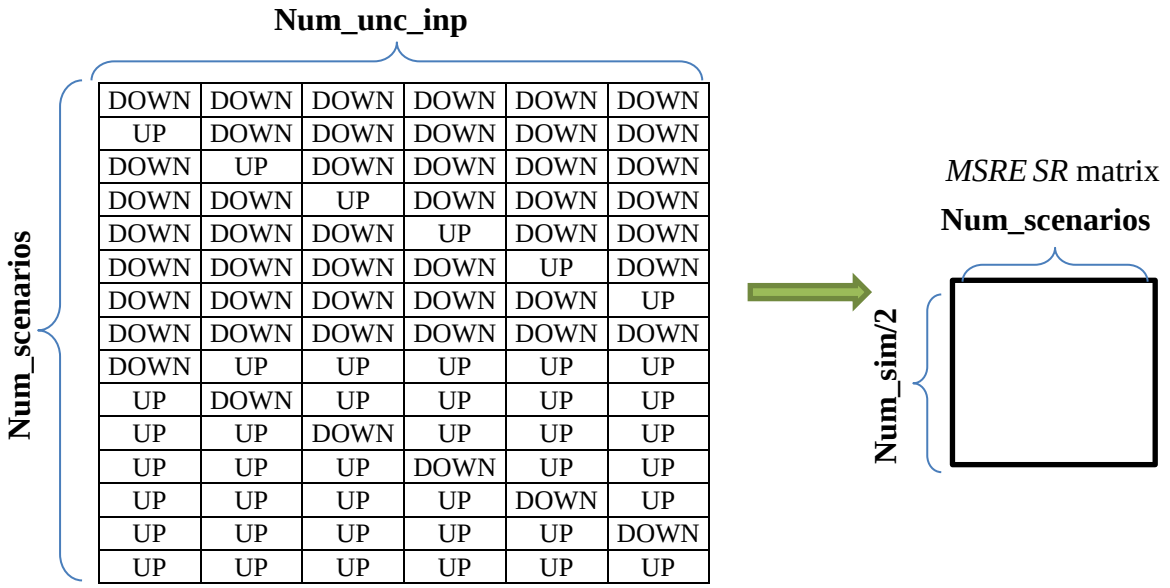


Figure 2. Combination matrix (left) for computing the SSI, based on two reference scenarios (first and last row) and "opposite" combinations and *MSRE SR* matrix (right).

Once the *MSRE SR* is obtained, Glen and Isaacs (2012) established that the *S* (main effect sensitivity index) and *T* (total effect sensitivity index) will be computed based in the two reference scenarios: the first and last row of combination matrix (n = 1 and n = end) (Figure 3). The main and the total sensitivity indexes for uncertainty input k will be then calculated based in the correlation coefficient ρ between *MSRE SR*(: , *k+1*) and *MSRE SR*(: , *k+1*+num_unc_inp), compared to the "reference" correlations *MSRE SR*(: , 1) and *MSRE SR*(: , end). Calculations are summarized in Figure 3.

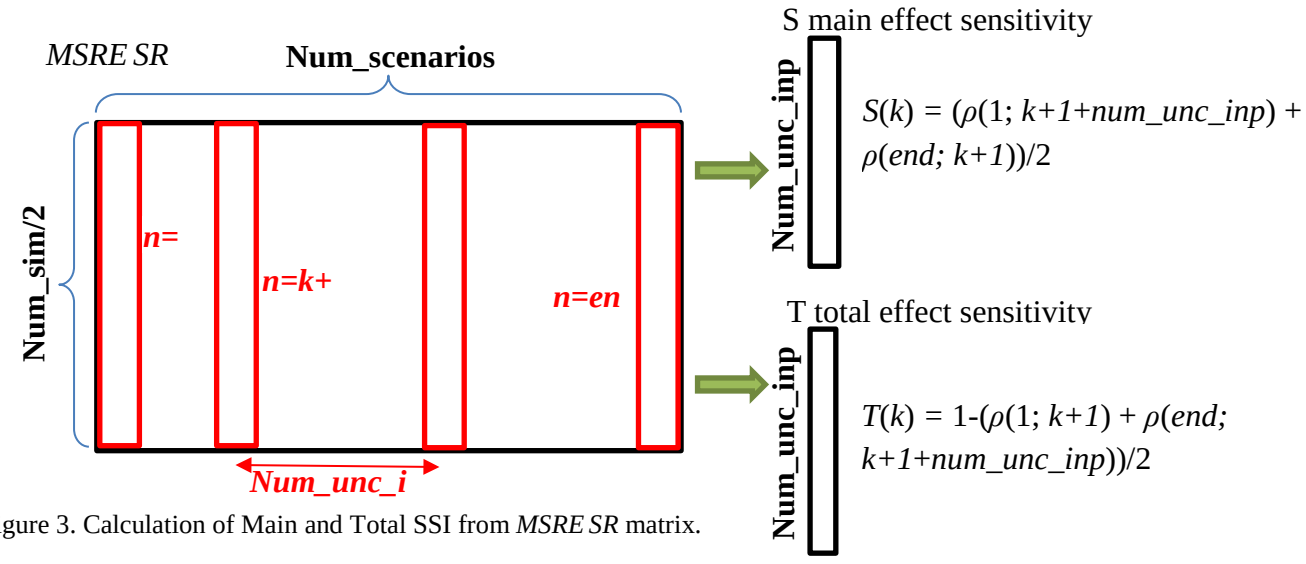


Figure 3. Calculation of Main and Total SSI from *MSRE SR* matrix.

REFERENCES

Glen, G. Isaacs, K., (2012). Estimating Sobol sensitivity indices using correlations. *Environmental Modelling and Software, 37*, 157-166.

3. COMPLEMENTARY PUBLICATIONS AND WORKS

Gap-filling of dry weather flow rate and water quality measurements in urban catchments by a time series modelling approach

Presented in:

Sandoval S., Vezzaro, L., Bertrand-Krajewski, J.-L., (2016). Gap-filling of dry weather flow rate and water quality measurements in urban catchments by a time series modelling approach. Proceedings of Novatech 2016, Lyon, France, 28 June-1 July, 4 p.

Résumé

Les séries chronologiques de débit et de qualité des eaux par temps sec dans les systèmes d'assainissement unitaires peuvent contenir une quantité importante de données manquantes, ceci pour de multiples raisons, telles que les défaillances de fonctionnement des capteurs ou des contributions additionnelles par temps de pluie. Par conséquent, l'approche proposée cherche à évaluer le potentiel de la méthode Singular Spectrum Analysis (SSA), une méthode de modélisation et de comblement de données manquantes, pour combler des séries chronologiques de temps sec. La méthode SSA est testée en reconstruisant 1000 séries chronologiques discontinues artificielles, construites aléatoirement à partir de séries réelles de débit et matières en suspension (MES) (année 2007, pas de temps de 2 minutes, système unitaire, Ecully, Lyon, France). Les résultats montrent la capacité de la méthode à combler des lacunes de données supérieures à 0.5 jour, surtout entre 0.5 et 1 jour (NSE moyen < 0.6) dans les séries chronologiques de débit. Les résultats sur les MES ne sont pas encore satisfaisants. Plusieurs analyses à différentes échelles temporelles sont envisagées.

Mots clés : Comblement de lacunes, mesures en ligne, métrologie, séries chronologiques, temps sec, validation de données.

Abstract

Flow rate and water quality dry weather time series in combined sewer systems might contain an important amount of missing data due to several reasons, such as failures related to the operation of the sensor or additional contributions during rainfall events. Therefore, the approach hereby proposed seeks to evaluate the potential of the Singular Spectrum Analysis (SSA), a time-series modelling/gap-filling method, to complete dry weather time series. The SSA method is tested by reconstructing 1000 artificial discontinuous time series, randomly generated from real flow rate and total suspended solids (TSS) online measurements (year 2007, 2 minutes time-step, combined system, Ecully, Lyon, France). Results show up the potential of the method to fill gaps longer than 0.5 days, especially between 0.5 days and 1 day (mean NSE > 0.6) in the flow rate time series. TSS results still perform very poorly. Further analysis at different temporal scales might be needed.

Keywords: Data validation, dry weather, gap filling, metrology, online monitoring, time series,

Introduction

Flow rate and water quality time series measured during dry weather periods at different locations in urban drainage systems (e.g. sewer system, WWTP, gully pots, detention basins) can be useful for several purposes (e.g. modelling, real time control, water management…). However, long-term dry weather time series may contain an important amount of unregistered or invalidated data due to failures related to the operation of the sensors, errors in measurement devices, maintenance and cleaning activities or disturbing contributions during rainfall events (wet weather period). These data gaps might vary from 1 or 2 minutes to days, weeks or even months.

Previous data-driven experiences sought to estimate the dry weather signal by the use of simplified periodic equations (e.g. Rodriguez *et al.*, 2013) or by filling gaps with data corresponding to similar dry weather periods (e.g. Métadier and Bertrand-Krajewski, 2011). However, these approaches do not consider the continuity of the real and long-term dry weather time series, dismissing possible frequency-variable, non-stationary and seasonal behaviors. These simplifications might bring up inconsistent results such as overestimations of the dry weather contributions during rainfall events, or mismatches between the beginning/ending of the gap with the beginning/ending of the signal to be fitted, especially for longer gaps (beyond hourly scale) (adapted from Métadier, 2011).

Singular Spectrum Analysis (SSA) is a modern non-parametric method for the analysis of time series and digital images (Korobeynikov, 2010). The SSA method has been applied for filling gaps in long-term and non-linear time series from analogue environmental contexts, reporting encouraging results (Musial *et al.*, 2011). The aim of this study is to assess the potential of the SSA method to estimate periods of missing data (from 6 minutes to 4.3 days), which might be useful for several additional applications, such as assessing the dry weather behavior during rainfall events.

Materials and methods

The method is tested with a one year flow rate and a TSS time series of the Ecully catchment (combined system, Lyon, France). The raw data includes 261 477 measurements (year 2007, 2 min time-step), with duration of gaps ranging from 2 min to 4.3 days for flow rate and 8.29 days for TSS, throughout the whole year (3.6 % and 28 % of the year respectively). Three data processing steps are applied to the raw data:

- Removing flow rate values during dry weather greater than the 95 percentile of the flow rates measured during the preceding storm event, which are about 70 L/s (dry weather outliers). For the case of TSS, values over 590 mg/L are considered as outliers from preliminary analyses.

- Removing the wet weather periods for both flow rate and TSS series (event durations ranged from 50 minutes to 39 hours, giving a total duration of events of about 21 days of additional data to be removed, even if they are already missing values).

- Filling redundant short-gaps in both series (with durations from 2 to 6 minutes), by linear interpolation, as the purpose is to explore especially longer gaps.

After applying the above raw data processing, the total percentage of gaps in the time series increased to 13 % (flow rate) and 34 % (TSS) of the year, with gaps from 6 minutes to 4.3 days for flow rate and 6 minutes to 8.3 days for TSS. The influence of beginning and ending of rainfall events over the flow rate and TSS time series (rainfall event) is identified from previous studies in this data set (Métadier, 2011).

The SSA method is applied to fill the gaps in both flow rate and TSS time series with the function "gapfill", from the "Rssa" package (Korobeynikov, 2010), implemented in R software (R Development Core Team, 2015). The function "gapfill" fills the missed entries in the series by performing forecast from both sides of the gap and taking an average in order to reduce the forecast error (see details: SSA sequential gap-filling method in Golyandina and Osipov, 2007). With the purpose of evaluating the performance of the SSA method in terms of predictability, a validation strategy based on the Monte Carlo method is hereby proposed. 1000 artificial discontinuous time series are generated by introducing gaps with random durations (uniformly distributed random numbers from 6 minutes to 4 days) over random parts of the original time series, with a check to guarantee a uniform distribution of gaps along the series. The additional percentage of gaps for each of the artificial discontinuous time series was set between 5 % and 30 % of the total duration of the time series (one year). The artificial time series are completed (gap filling) by the SSA method and compared to the original time series using the Nash-Sutcliffe model Efficiency (NSE). The NSE is chosen as the performance measure as it compares the performance of the method to a model that only uses the mean of the observed data (simplest prediction method) (from Bennett et al., 2013). The variability of the NSE value against gaps of different duration is analyzed as well.

Results and discussion

For illustrative purposes, the reconstruction obtained by the SSA method for an artificial gap (from 20/10/2007 16:23 to 23/10/2007 08:35) in the flow rate time series is compared to the original measured values, reporting a NSE value of 0.5 (Figure 1 a; line: reconstruction, dots: measurements). The NSE value is calculated between all the time series fragments reconstructed by the SSA method in each of the 1000 artificial discontinuous time series and the corresponding fragments in the original flow rate and TSS time series. Regarding flow rate, the NSE values are greater than 0.6 for all reconstructed fragments in half of the 1000 artificial discontinuous time series (Figure 1 b). The cases in which the NSE values show a poor performance of the SSA method can be attributed to the complexity and the large amount of data in the series. This trend is stronger in the case of TSS time series, in which 75 % of the NSE values are lower than zero (which is the NSE value corresponding to filling the gaps with the mean of the series).

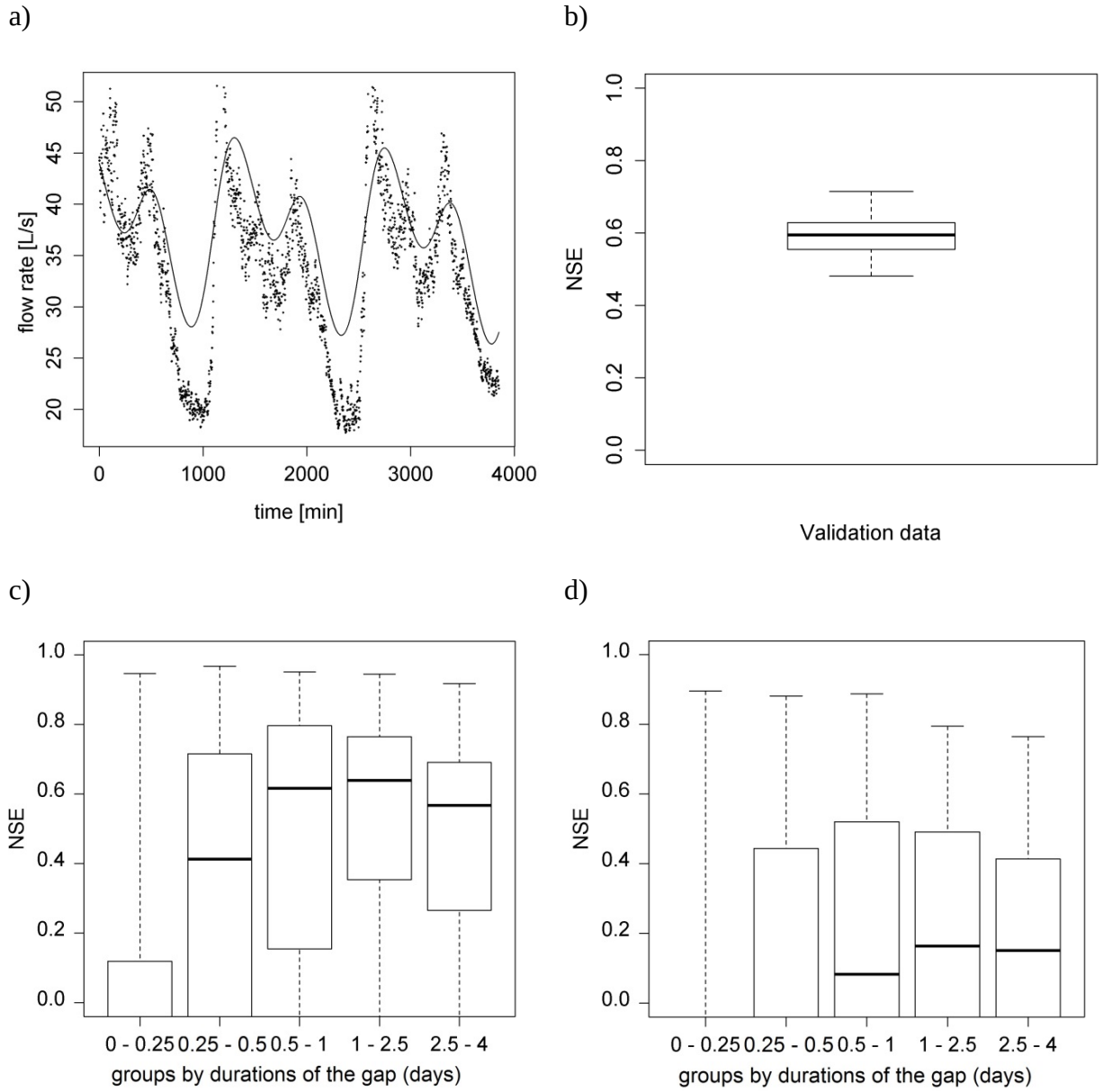


Figure 1. a) reconstruction obtained (line) for an artificial gap compared to the original flow rate values (dots), b) NSE values for the reconstruction of the 1000 artificial discontinuous flow rate, c) NSE values for gaps of different durations for flow rate and d) NSE values for gaps of different durations for TSS.

The performance of the SSA method is also analyzed by grouping the NSE values obtained for the reconstruction of gaps of different durations. For the case of flow rate, the SSA method shows a better performance for filling long-term gaps longer than 0.5 days (Figure 1 c). Specifically, the best performances are obtained for the reconstruction of gaps with durations between 0.5 day and 1 day (Figure 1 c). For shorter gaps, the results are poorer. This can be expected, as for this case, the SSA method includes long-term (weekly to monthly scales) components that are not related with the short-term (sub-daily scale) behaviors. Therefore, the SSA method at low temporal scales (e.g. daily or hourly scales) might have some potential adaptability by considering exclusively a certain amount of data adjacent to the gaps consistent with the temporal scale of analysis. However, filling gaps shorter than 0.5 day

with other methods that do not consider long-term patterns (e.g. mean values, typical dry weather daily curve or linear interpolations) might also be a suitable strategy.

The results for TSS show the same trend as for the flow rate series but with a significantly lower performance (more complex behaviors at all temporal scales) (Figure 1 d). Previous analyses highlighted the importance of finding an appropriate approach for representing the different long-term and short-term behaviors, aimed at modelling flow rate and TSS dry weather time series.

References

Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, J.H.A., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T.H., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., and Andreassian, V. (2013). Characterising performance of environmental models. Environmental Modelling & Software, 40, 1-20.

Golyandina, N. and Osipov E. (2007). The "Caterpillar"-SSA method for analysis of time series with missing values. Journal of Statistical Planning and Inference, 137, 2642-2653.

Korobeynikov, A. (2010). Computation- and space-efficient implementation of SSA. Statistics and Its Interface, 3(3), 257-368.

Musial, J.P., Verstraete, M.M. and Gobron, N. (2011). Technical Note: Comparing the effectiveness of recent algorithms to fill and smooth incomplete and noisy time series. Atmospheric Chemistry and Physics, 11(15), 7905-7923.

Métadier, M. (2011). Traitement et analyse de séries chronologiques continues de turbidité pour la formulation et le test de modèles des rejets urbains par temps de pluie. PhD Thesis, Institut National des Sciences Appliquées de Lyon, Lyon, France.

Métadier, M. and Bertrand-Krajewski, J.-L. (2011). From mess to mass: a methodology for calculating storm event pollutant loads with their uncertainties, from continuous raw data time series. Water Science and Technology, 63(3), 369-376.

Rodríguez, J.P, McIntyre, N., Díaz-Granados, M., Achleitner, S., Hochedlinger, M. and Maksimovic, C. (2013). Generating time-series of dry weather loads to sewers. Environmental Modelling & Software, 43, 133-143.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

4. SCIENTIFIC ACTIVITIES DURING THE THESIS

Published, submitted and prepared papers

Santiago, S., Bertrand-Krajewski J.-L., Caradot N., Hofer T., Gruber, G., (2017). Evaluation of performance and uncertainties in stormwater sampling strategies based on flow rate and total suspended solids time series. *Submitted*. (Main contributor)

Sandoval, S., Bertrand-Krajewski, J.-L., (2017). Strategy for assessing parameters of a rainfall-runoff model by connectivity representations and conditional probability functions. *In preparation*. (Main contributor)

Sandoval, S., Bertrand-Krajewski, J.-L., (2017). Methodology for identifying the temporal distribution of errors in rainfall by reverse modelling and Bayesian methods. *In preparation*. (Main contributor)

Sandoval, S., Vezzaro L., Bertrand-Krajewski, J.-L., (2017). Revisiting conceptual stormwater quality models by reconstructing virtual state-variables. *Submitted* (Mayor contribution)

Sun, S., Leonhardt G., Sandoval S., Bertrand-Krajewski, J.-L., Wolfgang R., (2017) A Bayesian method for missing rainfall estimation using a conceptual rainfall-runoff model. *Hydrological Sciences Journal. 62(15), 2456-2468.* (Secondary contributor)

Sandoval, S., Bertrand-Krajewski, J.-L., (2016). Influence of sampling intake position on suspended solids measurements in sewers: two probability / time series based approaches. *Environmental Monitoring and Assessment*, 188, 347. doi: 10.1007/s10661-016-5335-y. Published on line 13 May 2016.

International conferences

Santiago, S., Bertrand-Krajewski, J.L., Caradot, N., Hofer, T., Gruber, G., (2017). Evaluation of performance and uncertainties in stormwater sampling strategies based on flow rate and total suspended solids time series. *Proceedings of the 14th International Conference on Urban Drainage*, Prague, Czech Republic, 10-15 September, 3 p. (Main contributor)

Sandoval, S., Bertrand-Krajewski, J.-L., (2017). Strategy for assessing parameters of a rainfall-runoff model by connectivity representations and conditional probability functions. *Proceedings of the 14th International Conference on Urban Drainage, Prague*, Czech Republic, 10-15 September, 3 p. (Main contributor)

Sandoval, S., Vezzaro, L., Bertrand-Krajewski, J.-L., (2017). Revisiting conceptual stormwater quality models by reconstructing virtual state-variables. *Proceedings of the 14th International Conference on Urban Drainage*, Prague, Czech Republic, 10-15 September, 3 p. (Main contributor)

Vezzaro, L., Sandoval, S., Bertrand-Krajewski, J.-L., (2017). Training the urban water engineers of the future – the challenge of stormwater TSS model. *Proceedings of the 14th International Conference on Urban Drainag*e, Prague, Czech Republic, 10-15 September, 3 p. (Contributor)

Sandoval, S., Vezzaro, L., Bertrand-Krajewski, J.-L., (2016). Gap-filling of dry weather flow rate and water quality measurements in urban catchments by a time series modelling approach. *Proceedings of Novatech 2016*, Lyon, France, 28 June-1 July, 4 p.

Sandoval, S., Bertrand-Krajewski, J.-L., (2015). Identification of errors in high temporal resolution rainfall time series by model based approaches. *Proceedings of the 10th UDM - International Conference on Urban Drainage Modelling*, Mont Sainte Anne, Quebec, Canada, 20-23 September, Oral Presentations II, 183-186.

Sandoval, S., Bertrand-Krajewski, J.-L., (2014). A methodology for estimating the influence of sampling intake position in suspended solids measurements in sewers. *Proceedings of the 13th International Conference on Urban Drainage*, Kuching, Malaysian Borneo, 7-12 September, 8 p.

National conferences

Sandoval, S., Bertrand-Krajewski J.-L., (2016). Modélisation pluie-débit : stratégie améliorée de calage et estimation des incertitudes guidée par les données. *Actes des 7° Journées Doctorales en Hydrologie Urbaine "JDHU 2016"*, Nantes, France, 1-3 juillet, 5 p. (in French) (Main contributor)

Sandoval, S., Bertrand-Krajewski, J.-L., (2014). Estimation de l'influence du point d'échantillonnage des matières en suspension dans une section de réseau d'assainissement. *Actes des 6° Journées Doctorales en Hydrologie Urbaine "JDHU 2014"*, Lyon, France, 1-3 juillet, 4 p. (in French) (Main contributor)

Bertrand-Krajewski, J.-L., Sandoval, S., (2013). Les micropolluants dans les eaux pluviales : méthodologie expérimentale, incertitudes et étude de cas. *Actes de la 5° conférence régionale GRAIE " Gestion des rejets d'eaux usées non domestiques au réseau d'assainissement "*, Villeurbanne, France, 14 novembre 2013. (in French) (Secondary contributor)

Other international communications

Sandoval, S., Bertrand-Krajewski, J.-L., (2017). Strategy for assessing parameters of a rainfall-runoff model by connectivity representations and conditional probability functions. *Proceedings of the 23th European Junior Scientists Workshop on "Monitoring urban drainage systems"*, Chichilianne (France), 5-20 May 2017, 4 p. (Main contributor)

Sandoval, S., Bertrand-Krajewski J.-L., (2015). Influence of sampling intake position on suspended solids measurements in sewers: two probability / time-series based approaches. *Proceedings of the 22th European Junior Scientists Workshop on "Monitoring urban drainage systems"*, Chichilianne (France), 18-22 May 2015, 4 p. (Main contributor)

Attended courses and national seminars

Séminaire MEGA : Système de roselière artificielle et bassin réservoir pour la collecte et l'utilisation des eaux de pluie : vers le développement d'un système d'aide à la décision en continu. Présenté par : Andres Torres (18 May 2016) (in French).

French language summer school A1 level (Ecole d'été niveau A1) at INSA (Lyon, France) duration: 80 hours (August-September 2013) (in French).

French language annual course level A2/B1 at INSA (Lyon, France) duration: 20 hours (September- December 2013) (in French).

MSc course: *Expérimentation et modélisation* at INSA (Lyon, France) duration (13 weeks). (September-December 2014).

eLearning online MSc course: Integrated Urban Water Quality Management. Proposed by DTU (Copenhagen, Denmark) duration: 13 weeks. September-December 2015.

Students supervision

Mesurage automatisé par turbidimétrie des vitesses de chute des particules dans les rejets urbains de temps de pluie. Projet d'Initiation à la Recherche et Développement 2 (PIRD2) (2015). Etudiants encadrés : Terrier., V., Torras, L. Directeurs : Bertrand-Krajewski J.-L., Becouze, C., Sandoval, S., Vacherie, S. (in french). (Co-director)

Améliorer les modèles de flux polluants des rejets urbains de temps de pluie. Projet d'Initiation à la Recherche & Développement 2 (PIRD2) (2017) Etudiants encadrés : Kohl, M., Hugues, O. (in French). (Director)

Teaching

Assistant lecturer at the international PhD Course: Modelling of Integrated Urban Drainage-Wastewater Systems. Universities of INSA (Lyon, France); DTU Environment (Copenhagen, Denmark); ModelEAU, Laval University (Quebec, Canada), April 2016, held at INSA (Lyon, France).

Lecturer in the MSc Courses: (i) Experimentation and Modelling (INSA, Lyon) (3 sessions in February 2016, in French) and (ii) Modelling of Environmental Processes and Technologies (DTU, Denmark) (3 sessions in November 2016).

Scientific visits and collaborations

Participation in the collaborative research group "On-line surrogate measurements" (OSM) with Technical University of Graz (Graz, Austria), PUJ (Bogota, Colombia), Kompetenzzentrum Wasser Berlin gGmbH (Berlin, Germany)and INSA (Lyon, France) (2013).

Participation in the collaborative research group "Stormwater quality modelling tournament" (STOQUAMOT) with Technical University of Innsbruck (Innsbruck, Austria), Monash University (Melbourne, Australia), DTU (Copenhagen, Denmark) and INSA (Lyon, France) (2017).

Scientific visit under the frame of ECOS NORD Colombie project (December 2013-January 2014) to work in Bogota, Colombia under the supervision of Andres Torres, at Pontificia Universidad Javeriana (Bogota, Colombia).

Scientific visit under the frame French-Danish Research Collaboration Program (financed by French Institute in Denmark) (November 2016-December 2016) to work in Copenhagen, Denmark under the supervision of Luca Vezzaro, at DTU University (Copenhagen, Denmark).

Evaluation committees

Member of the evaluation committee of the Master Thesis entitled: "*Metodología para clasificar las tuberías que no han sido inspeccionadas según su condición estructural en la red de alcantarillado de Bogotá D.C*." (Methodology for classifying sewer pipelines that have not been inspected according to their structural condition in the sewer system of Bogota D.C.). Pontificia Universidad Javeriana. Bogota, Colombia. Student: Hernandez, N.

Reviewer of two scientific articles for the "*Revista Tecnura*" and "*Water Science and Technology*" international journals.

FOLIO ADMINISTRATIF

THESE DE L'UNIVERSITE DE LYON OPEREE AU SEIN DE L'INSA LYON

NOM : SANDOVAL DATE de SOUTENANCE : 05/12/2017

Prénoms : Santiago

TITRE : REVISITING STORMWATER QUALITY CONCEPTUAL MODELS IN A LARGE URBAN CATCHMENT: ONLINE MEASUREMENTS, UNCERTAINTIES IN DATA AND MODELS

NATURE : Doctorat Numéro d'ordre : 2017LYSEI089

Ecole doctorale : MEGA DE LYON (MECANIQUE, ENERGETIQUE, GENIE CIVIL, ACOUSTIQUE) - ED162

Spécialité : Génie Civil

RESUME :

Different hypotheses about traditional Total Suspended Solids (TSS) stormwater models from data validation and hydrological concepts have been revisited in a 185 French urban catchment, including online data from 365 rainfall events. Four sampling strategies during rainfall events are simulated and compared to online monitoring. Recommended sampling time intervals are of 5 min, with average sampling errors between 7 % and 20 % and uncertainties in sampling errors of about 5 %, depending on the sampling interval. The probability of underestimating the cross section mean TSS concentration in the sewer system is estimated by two methodologies, were one shows more realistic TSS underestimations (about 39 %). A power law describing the TSS as a function of flow rate is revealed. In the hydrological context, a parameter estimation strategy is proposed by analyzing the variability of parameters obtained by event-based Bayesian calibrations, based on clusters and graphs representations. A single model structure might be able to reproduce at least two different hydrological conditions. A methodology aimed to calculate "mean" areal rainfall estimation is proposed, based on a hydrological model and flow rate data. Regarding TSS modelling, the performance indicators of the traditional Rating Curve (RC) model are superior to different transfer Functions (TFs), with flow rate, rainfall or "mean" areal rainfall as the model input. The potential missing representation of an essential process in the RC model is found to be independent of antecedent dry weather period. A Bayesian reconstruction method indicates that a potential missing process in the RC description is hardly interpretable in terms of a unique state of virtual available mass over the catchment that is decreasing over time, as assumed by a great number of traditional models. Furthermore, the reconstructed processes are highly unrepeatable regarding their shape, besides having a low transferability to other rainfall events.

MOTS-CLÉS: Bayesian method, conceptual modelling, conditional probability, clustering, error models, identifiability, parameters variability

Laboratoire (s) de recherche : Laboratoire DEEP (Déchets Eaux Environnement Pollutions)

Directeur de thèse: Jean-Luc BERTRAND-KRAJEWSKI

Président de jury : Sylvie BARRAUD

Composition du jury :

BARRAUD, Sylvie	INSA Lyon	Présidente
CLEMENS, François	TU DELFT	Rapporteur
TORRES, Andres	PUJ Pontificia Universidad Javeriana	Rapporteur
BONHOMME, Céline	Ecole des ponts – Paristech	Examinatrice
VIKLANDER, Maria	Lulea University of Technology	Examinatrice
RENARD, Benjamin	IRSTEA LYON	Examinateur
BERTRAND-KRAJEWSKI, Jean-Luc	INSA Lyon	Directeur de thèse