Temporal uncertainty propagation analysis
A contribution towards sustainable urban water management
Jairo Arturo Torres-Matallana
2021

Temporal uncertainty propagation analysis

A contribution towards sustainable urban water management

Jairo Arturo Torres-Matallana

Propositions

1 Urban stormwater modelling practitioners ignore uncertainty analysis because for them the required computational effort is by far too high.

(this thesis)

2 Uncertainty propagation analysis is an essential tool for the comprehensive assessment of the water quality impact on the receiving water.

(this thesis)

3 Black-box models and algorithms lead to a simplistic commercialisation of technology.

4 The climate change threat only exists in the human mind.

5 Clean water is a common social asset beyond being a mere environmental resource.

6 Nowadays it is evident that the search for individual economic profit is a reality.

7 Current technological approaches do not fulfill the requirements for developing sustainable cities, especially in developing countries.

8 Measures to preserve the natural environment and gain resilience are needed to put an end to megacities development.

Propositions belonging to the thesis, entitled

Temporal uncertainty propagation analysis: A contribution towards sustainable urban water management

J. A. (Arturo) Torres-Matallana

Wageningen, Tuesday 8 June 2021

Temporal uncertainty propagation analysis

A contribution towards sustainable urban water management

Jairo Arturo Torres-Matallana

Thesis committee

Promotor
Prof. Dr Gerard B.M. Heuvelink
Special Professor, Soil Geography and Landscape
Wageningen University & Research

Other members
Prof. Dr Albrecht Weerts, Wageningen University & Research
Dr Alma Schellart, University of Sheffield, United Kingdom
Dr Franz Tscheikner-Gratl, Norwegian University of Science and Technology, Trondheim, Norway
Dr Sytze de Bruin, Wageningen University & Research

This research was conducted under the auspices of the C.T. de Wit Graduate School of Production Ecology & Resource Conservation (PE&RC)

Temporal uncertainty propagation analysis

A contribution towards sustainable urban water management

Jairo Arturo Torres-Matallana

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Tuesday 8 June 2021
at 4 p.m. in the Aula.

Jairo Arturo Torres-Matallana
Temporal uncertainty propagation analysis:
A contribution towards sustainable urban water management
287 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2021)
With references, with summary in English and Spanish

ISBN 978-94-6395-781-6
DOI https://doi.org/10.18174/545460

To

God Almighty and Our Lady of Fátima,

my loved wife,

and my dear parents,
siblings and nephew

A

Dios Todopoderoso y Nuestra Señora de Fátima,

mi esposa amada,

mis queridos madre, padre,
hermanos y sobrino

Contents

List of Figures

List of Tables

Chapter 1

General Introduction

1.1 Background and context

Sustainable urban water management (SUWM) is becoming a global priority due to the impact of urbanisation on natural and urban ecosystems. Globally, more people live in urban areas than in rural areas (United Nations - UN, 2018). The global population in 2018 was 7.6 billion and the urban population was 4.2 billion, which is equivalent to 55 per cent of the world's population residing in urban areas. Also, United Nations - UN (2018) stated that in 1950, 30 per cent of the world's population was urban, and by 2050, the global population will reach 9.7 billion, with 68 per cent of the world's population (i.e., 6.6 billion people) projected to be urban.

Due to different levels of socioeconomic development, the process of urbanisation is uneven across countries (Sun et al., 2020; Elmqvist et al., 2013). In contrast to developed countries, for many developing countries the national economic growth and development are inadequate to meet the needs of a growing urban population. In most cases, cities of developing countries lack basic urban infrastructure, such as water and energy supplies, sanitation, education, and green space or parks, and face overcrowding, pollution, and other urban environmental problems (Elmqvist et al., 2013). With the increase of urbanisation development this lack could become more acute for the sustainability of cities in the future.

Following Davis and McCuen (2005), changes to the hydrological cycle, due to urbanisation to support population increase, are becoming more evident nowadays. Greater volumes of runoff, higher flow velocities, and increased pollutant fluxes to local receiving waters, are some of the consequences of an extension of impervious areas. To learn more about the negative impacts of these consequences, and to develop and manage land in a smart and sustainable manner, measures to reduce these impacts are of paramount importance.

Urbanisation not only impacts the hydrological cycle, but also has an interlinked impact on the urban landscape ecosystem and its evolution, which is different from the evolution of the natural landscape (Johnson and Munshi-South, 2017). This has been recognised in the last two decades by governments, especially in more economically developed countries, with the advent of concepts such as "Sustainable Development", "Biodiversity", "Natural Capital", "Ecosystem Services", "Ecosystem-based approaches", and, more recently, "Nature-Based Solutions", which focus primarily on what nature can provide to humans. This is a response to the urgency of solving urban water problems and integrating solutions with new urban water management strategies and practices (Qi et al., 2020).

Hence, SUWM is key toward contributing to the United Nations' Sustainable Development Goals (United Nations - UN, 2015), specifically to strive for the achievement of Goal 6 *"ensure availability and sustainable management of water and sanitation for all"*, and Goal 14 *"conserve and sustainably use the oceans, seas and marine resources for sustainable development"*.

Integrated urban drainage models are primary components of monitoring systems and essential decision-making tools for SUWM, allowing water flow and quality simulation and prediction in real-time. However, it is paramount to recognise that in environmental modelling, and hence also in SUWM, every model contains uncertainties to some degree (Beven, 1989; Grayson et al., 1992; Freer et al., 2004; Deletic et al., 2012; Dotto et al., 2012). This is because any model makes simplifications and assumptions about the real-world processes involved, while model inputs are rarely if ever known without error. This also applies to integrated urban drainage modelling (IUDM). Quantification of model input and output uncertainty is essential for characterising inputs and choosing objectively the suitable configuration of the model for addressing a specific task related to IUDM.

Several recent studies in the urban drainage management and modelling domain have been done. Although parsimonious lumped models are established concepts in the hydrological literature, there are still few studies that use these models in applications in urban drainage systems modelling. The majority of applications are found in studies on natural areas. Simplified models in urban drainage modelling "compress the complexity of the real system in only a few characteristics and/or relationships" (van Daal-Rombouts et al., 2016). But model simplification and modular model setup are important features that facilitate model execution for uncertainty propagation tasks. Besides model simplification, scalable distributed parallel computing implementations are recognised as good approaches for high performance computing in SUWM. Nevertheless, there are only a few examples of scalable implementations for solving intensive computational tasks in watershed-distributed and semi-distributed modelling in the urban drainage domain.

Uncertainty quantification in integrated environmental models with emphasis in SUWM,

specifically urban stormwater system models (USSM's) used in decision-making for environmental protection, require that the accuracy of model outputs is known and meets predefined standards. Uncertainty propagation in SUWM is yet not well understood.

Four important problems are identified:

1. The first problem is related to the fact that full hydrodynamic USSM's are complex and require a highly intense computational budget, which constitutes a constraint when long-term simulations or uncertainty propagation analysis are required, e.g. by techniques such as Monte Carlo simulation (i.e., Robert and Casella (2010)).

2. The second problem is that although catchment average precipitation is a key component of USSM's, catchment average precipitation is not always accurately known when derived from measurements at point support, i.e. by rain gauges. This is because the precipitation at a point may not reflect the average precipitation well, especially when the location of the gauge is outside the catchment boundaries. Therefore, a method to estimate the precipitation in a catchment given a known precipitation time series at a location outside of the catchment is required. This method should also quantify the uncertainty associated with the estimated catchment average precipitation.

3. The third problem is that software tools for temporal uncertainty propagation for USSM's are not generally available. Many model implementations are not fully automated and cannot be called in batch mode, which is needed in Monte Carlo uncertainty propagation analyses.

4. The fourth problem is related to the fact that many studies in USSM's do not pay attention to uncertainty and uncertainty propagation. Statistical uncertainty analysis of USSM's is a relatively new subject that largely needs to be developed while very few solid applications have been done. This is linked to the second problem, because uncertainty propagation analysis can only be done if the uncertainty sources are quantified. It is also linked to the third problem because of the need for software tools with capabilities to efficiently propagate input uncertainty in USSM's, and requires contributions from multiple disciplines (i.e., hydrology, statistics, computer science).

All the aforementioned problems represent an opportunity to develop new methods and software tools to overcome the current limitations exposed. The content of the present dissertation will address all these problems.

1.2 Bibliometric analysis

While the previous section stipulated some key issues and publications on uncertainty analysis in urban hydrological modelling, a bibliometric analysis was made to obtain

a more comprehensive overview of the state-of-the-art of this specific research subject. The bibliometric analysis was made based on three main searches by title, abstract and keywords in Web of Science and Scopus databases:

1. Search: temporal AND uncertainty AND urban AND (water OR stormwater OR drainage).
2. Search: temporal AND uncertainty AND (rainfall OR precipitation OR "water quality") AND urban AND (water OR stormwater OR drainage).
3. Search: temporal AND uncertainty AND (software OR tool) AND urban AND (water OR stormwater OR drainage).
4. A fourth analysis was performed for the data merged from Searches 1 to 3 analysed together, excluding duplicated records.

The bibliometric analysis was made by application of the R-package Bibliometrix (Aria and Cuccurullo, 2017) and further script extensions.

Table 1.1 summarises the main bibliometric information about the information retrieved from the three searches. A total of 267 documents were retrieved and analysed from 138 different sources (Journals, Books, Proceedings, etc.) on 24th November 2020. Figure 1.1 illustrates the annual scientific production for the three searches for the period 1982 - 2020. The major contribution in number of publications comes from Search 1, while Search 2 was the second most important contribution in number of publications retrieved. Considering all searches together and excluding duplicated records, Bonhomme, C. and Zhan, Y. were the most productive authors with six publications indexed, followed by Lee, Sh. and Rieckermann, J. with five publications each.

Table 1.1: Main information about data retrieved from Web of Science and Scopus on 24th November 2020.

Information	Search 1	Search 2	Search 3	All searches
Documents	266	139	61	267
Sources (Journals, Books, Inproceedings, etc.)	137	75	34	138
Keywords Plus (ID)	2,013	1,052	827	2,025
Author's Keywords (DE)	888	476	201	893
Period	1982 - 2020	1982 - 2020	1998-2020	1982 - 2020
Average citations per documents	22.74	23.47	18.34	22.65
References	13,580	7,352	3,443	13,629
Authors	1,032	503	179	1,035
Author Appearances	1,201	571	246	1,204
Authors of single-authored documents	8	5	3	8
Authors of multi-authored documents	1,024	498	176	1,027
Single-authored documents	10	5	3	10
Documents per Author	0.258	0.276	0.341	0.258
Authors per Document	3.88	3.62	2.93	3.88
Co-Authors per Documents	4.52	4.11	4.03	4.51
Collaboration Index	4.00	3.72	3.03	4.00

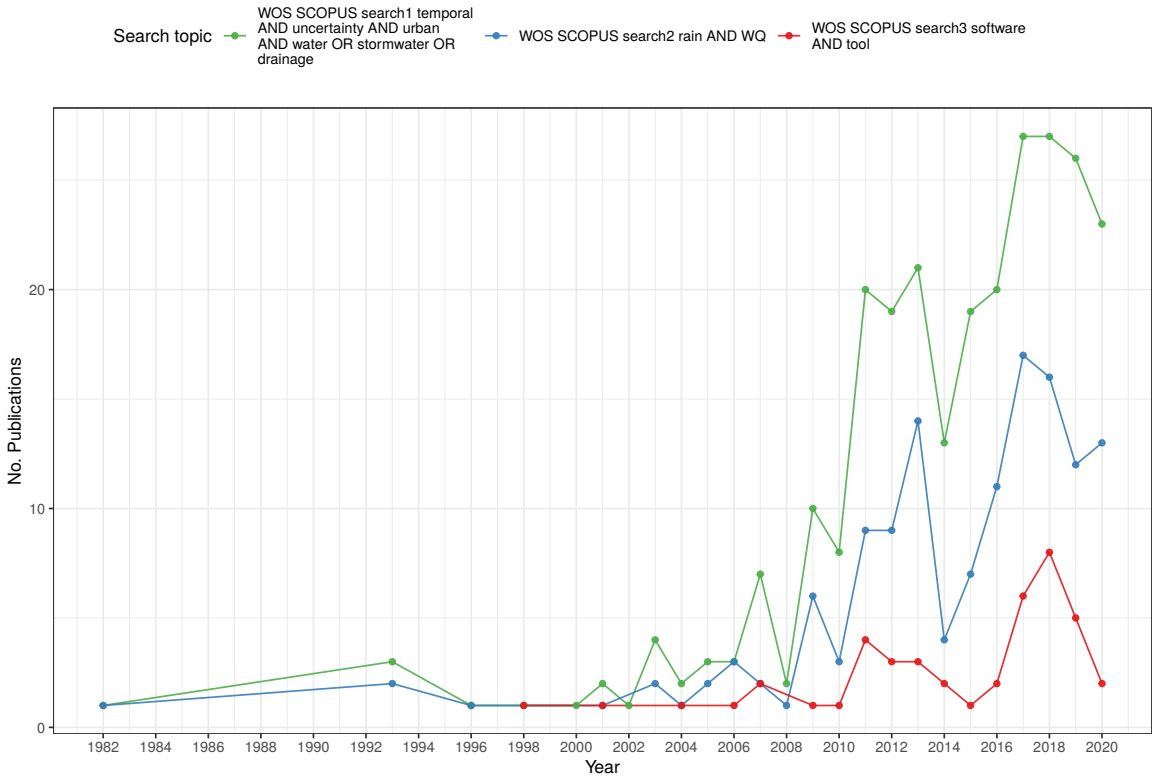


Figure 1.1: Annual scientific production by search terms made on 24th November 2020. WOS = Web of Science.

Figure 1.1 shows the development over time of the number of publications on the various topics. Searches 1 and 2 have been registered since 1982. Documents retrieved from Search 3 have been registered since 1998. This indicates that software and tools topics represent a newer research field in the domain of urban water and stormwater systems. It is also growing over time as will be shown below.

Table 1.2 ranks the most relevant sources for all searches combined excluding duplicated records (i.e., from the total of 267 documents). The journal *Water* appears as the most frequent source with a contribution of 19 publications indexed in Web of Science and Scopus. In second place, *Science of the Total Environment* appears with a contribution of 14 publications. The third most productive source is *Journal of Hydrology* with 11 publications, and the next two journals are *Hydrology and Earth Science Sciences* and *Water Research* with 10 publications each.

The bibliometric analysis included the creation of a keyword network, which analyses the co-occurrences among the different publication author's keywords for the data from all searches. Figure 1.2 shows that, when 15 nodes are defined, three clusters are created. These are the red, blue, and green clusters. In this network the nodes are drawn as circles, where a bigger circle means a higher co-occurrence of a keyword. The lines

Table 1.2: Most relevant sources of documents retrieved from the Web of Science search.

Sources	Articles
Water	19
Science of the Total Environment	14
Journal of Hydrology	11
Hydrology and Earth System Sciences	10
Water Research	10
Water Resources Research	9
Water Science And Technology	9
Environmental Modelling & Software	7
Journal Of Environmental Management	5
Modsim 2011 - 19th International Congress on Modelling and Simulation - Sustaining Our Future: Understanding and Living With Uncertainty	5
Advances in Water Resources	4
Atmospheric Chemistry and Physics	4
Remote Sensing of Environment	4

represent the connections or relations between nodes. The connections among nodes of different clusters are drawn in dashed light gray lines. The stronger the connection between nodes, the thicker the line. Cluster 1 in red shows a higher importance of the author keywords "calibration", "management", "modeling", and "uncertainty analysis", which have a strong connection among them. In cluster 2 in blue, the dominant keywords are "climate change" and "rainfall". Cluster 3 in green shows the keyword "uncertainty" as being most important. It is worth noting that the relative distance among clusters also indicates closeness relationships, which suggests that the "remote sensing" keyword in Cluster 3 is more related to the other clusters than e.g. "urban hydrology".

Additionally, the bibliometric analysis included the creation of a subject co-occurrences network, which analyses the co-occurrences of the different subject categories defined in Web of Science. Figure 1.3 shows that when 15 nodes are defined, three clusters are created (red, blue and green) based on the three searches made. From this clustering structure it is possible to infer that in Cluster 1 (red) there are more publications on subjects related to environmental sciences and ecology, engineering and water resources, followed with publications in geology, meteorology and atmospheric sciences, agriculture, and marine freshwater biology. Cluster 2 (blue) is characterised by a greater co-occurrence of the subject remote sensing, followed by physical geography. In Cluster 3 (green), the greater co-occurrences are the subjects mathematics and computer sciences.

Figure 1.4 illustrates the number of publications per most frequent term and subject. The most frequent terms are "model", "uncertainty", "urban", "temporal", and "system", all

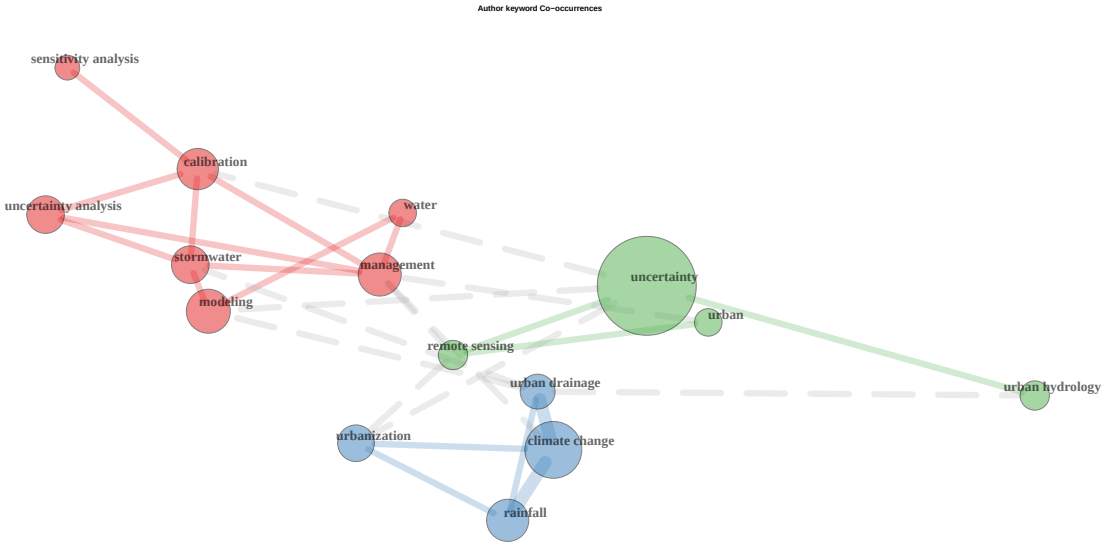


Figure 1.2: Author keyword co-occurrences. Cluster 1 in red, Cluster 2 in blue, and Cluster 3 in green.

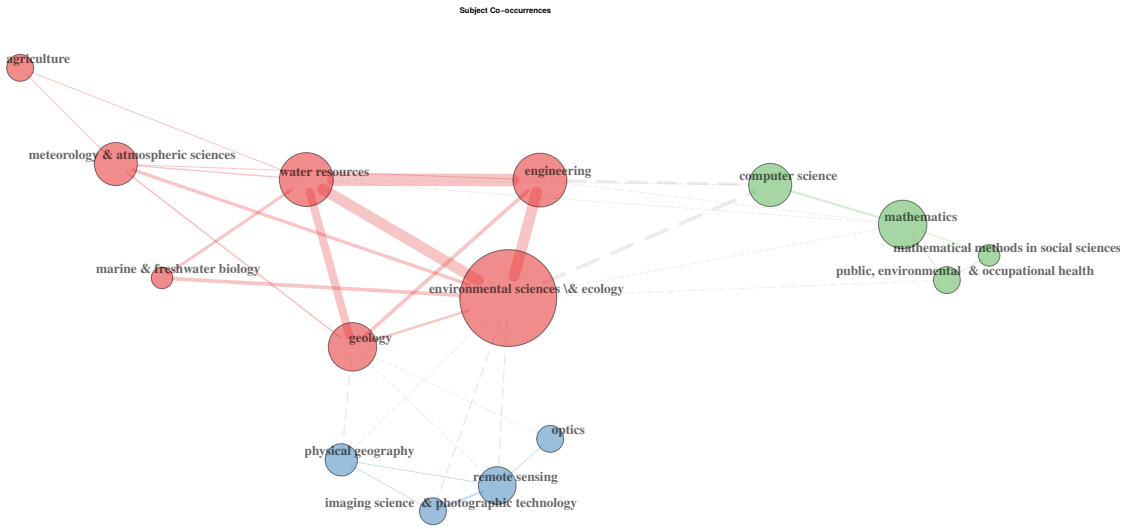


Figure 1.3: ISI Web of Science subject co-occurrences. Cluster 1 in red, cluster 2 in blue, and cluster 3 in green.

with highest occurrence in the water resources and environmental science and ecology subjects.

Figure 1.4: Number of publications per most frequent term and subject.

Table 1.3 presents the most cited papers and number of total citations from the bibliometric analysis for the three searches, excluding duplicated entries. The most cited paper is Dokulil and Teubner (2000) with 422 total citations, followed by Fletcher et al. (2013) with 346 total citations and Elliott and Trowsdale (2007) with 327 total citation.

A step forward from the author keyword network analysis is to list the top five most cited and a sample of the most recent references per most relevant author's keywords to 24th November 2020. Table 1.4 presents this list. It is worth noting that these keywords where extracted from the three searches performed (clusters presented in Figure 1.2), which covers the topics related to this research.

A co-citation network is also presented in the bibliometric analysis (Figure 1.5). Two articles are co-cited when both are cited in a third article. Thus, co-citation can be seen as the counterpart of bibliographic coupling. Two articles are said to be bibliographically coupled if at least one cited source appears in the bibliographies or reference lists of both articles (Kessler, 1963). From 15 nodes, two clusters are identified. Cluster 1 (red) has major nodes for Zoppou (2001) and Beven and Binley (1992), followed by Nash and

Table 1.3: Most cited papers from the bibliometric analysis for the three searches excluding duplicated entries. TC = total number of citations.

Paper	Reference	DOI	TC	TC Per Year
DOKULIL MT, 2000, HYDROBIOLOGIA	Dokulil and Teubner (2000)	10.1023/A:1004155810302	422	20.10
FLETCHER TD, 2013, ADV WATER RESOUR	Fletcher et al. (2013)	10.1016/j.advwatres.2012.09.001	346	43.25
ELLIOTT AH, 2007, ENVIRON MODELL SOFTW	Elliott and Trowsdale (2007)	10.1016/j.envsoft.2005.12.005	327	23.36
FOWLER HJ, 2003, INT J CLIMATOL	Fowler and Kilsby (2003)	10.1002/joc.943	231	12.83
ZOPPOU C, 2001, ENVIRON MODELL SOFTW	Zoppou (2001)	10.1016/S1364-8152(00)00084-0	231	11.55
BARBOSA AE, 2012, WATER RES	Barbosa et al. (2012)	10.1016/j.watres.2012.05.029	208	23.11
HOUSE-PETERS LA, 2011, WATER RESOUR RES	House-Peters and Chang (2011)	10.1029/2010WR009624	185	18.50
WILLEMS P, 2012, ATMOS RES	Willems et al. (2012)	10.1016/j.atmosres.2011.04.003	175	19.44
LIU M, 2010, GLOB BIOGEOCHEM CYCLE	Liu and Tian (2010)	10.1029/2009GB003687	129	11.73
SEXTON JO, 2013, REMOTE SENS ENVIRON	Sexton et al. (2013)	10.1016/j.rse.2012.10.010	122	15.25
SALVADORE E, 2015, J HYDROL	Salvadore et al. (2015)	10.1016/j.jhydrol.2015.06.028	121	20.17
CASTIGLIONI S, 2014, SCI TOTAL ENVIRON	Castiglioni et al. (2014)	10.1016/j.scitotenv.2013.10.034	115	16.43
LLOYD CEM, 2016, SCI TOTAL ENVIRON	Lloyd et al. (2016)	10.1016/j.scitotenv.2015.11.028	111	22.20
ZOOGMAN P, 2017, J QUANT SPECTROSC RADIAT TRANSF	Zoogman et al. (2017)	10.1016/j.jqsrt.2016.05.008	108	27.00
ELMORE AJ, 2012, GLOB CHANGE BIOL	Elmore et al. (2012)	10.1111/j.1365-2486.2011.02521.x	103	11.44
OCHOA-RODRIGUEZ S, 2015, J HYDROL	Ochoa-Rodriguez et al. (2015)	10.1016/j.jhydrol.2015.05.035	101	16.83
NOOR AM, 2009, BMC INFECT DIS	Noor et al. (2009)	10.1186/1471-2334-9-180	93	7.75
SU S, 2011, WATER RES	Su et al. (2011)	10.1016/j.watres.2010.11.030	85	8.50
KOHLER M, 2008, ENVIRON SCI TECHNOL	Kohler et al. (2008)	10.1021/es702586r	84	6.46
BROWN JD, 2007, WATER RESOUR RES	Brown et al. (2007)	10.1029/2005WR004597	77	5.50

Table 1.4: Most cited and most recent references by clusters defined in the bibliometric analysis by most relevant author's keywords related to this study. Searches 1 to 3 from Web of Science (WoS) and Scopus on 24th November 2020, analysed together and excluding duplicated records. TC = Times Cited; PY = Publication Year.

Cluster	Author keyword	Most cited reference	TC	PY	Most recent references	TC	PY
1	SENSITIVITY ANALYSIS	Booker and Dunbar (2004)	43	2004	Naves et al. (2020)	4	2020
		Burgos et al. (2017)	6	2005	Kang et al. (2020)	1	2020
		Naves et al. (2020)	4	2020	Hong et al. (2019)	3	2019
		Hong et al. (2019)	3	2019	Burgos et al. (2017)	6	2005
		Kang et al. (2020)	1	2020	Booker and Dunbar (2004)	43	2004
2	UNCERTAINTY ANALYSIS	Muleta et al. (2013)	16	2013	Tavakol-Davani et al. (2019)	0	2019
		Pablo Rodriguez et al. (2013)	13	2013	Muleta et al. (2013)	16	2013
		Blumensaat et al. (2012)	6	2012	Pablo Rodriguez et al. (2013)	13	2013
		Manz et al. (2013)	4	2013	Manz et al. (2013)	4	2013
		Perraud et al. (2007)	1	2007	Blumensaat et al. (2012)	6	2012
3	UNCERTAINTY	Sage et al. (2015)	111	2016	Stephens and Bledsoe (2020)	4	2020
		Fu et al. (2019)	40	2019	Tavakol-Davani et al. (2019)	0	2019
		Aronica et al. (2005)	36	2005	Ju et al. (2019)	2	2019
		Wijesiri et al. (2016)	28	2016	Fu et al. (2019)	40	2019
		Ahmad and Simonovic (2013)	24	2013	Mutzner et al. (2019)	12	2019

Sutcliffe (1970). Cluster 2 (blue) has Aronica and Cannarozzo (2000) as the most relevant co-citation node, followed by Berne et al. (2004) and Einfalt et al. (2004).

1.3 State-of-the-art and gap identification

Given the bibliometric analysis results, Tables 1.3 and 1.4 describe and summarise the most cited and relevant references. The most often cited paper, Dokulil and Teubner (2000), presented a study about the cyanobacterial dominance in lakes of higher trophic levels. The authors analysed and discussed the underlying mechanisms of cyanobacterial dominance using both original and literature data from various lakes in temperate and (sub)tropical regions. They concluded that the long-term dominance of cyanobacteria is caused by several factors, being the most important nutrient concentration, lake mor-

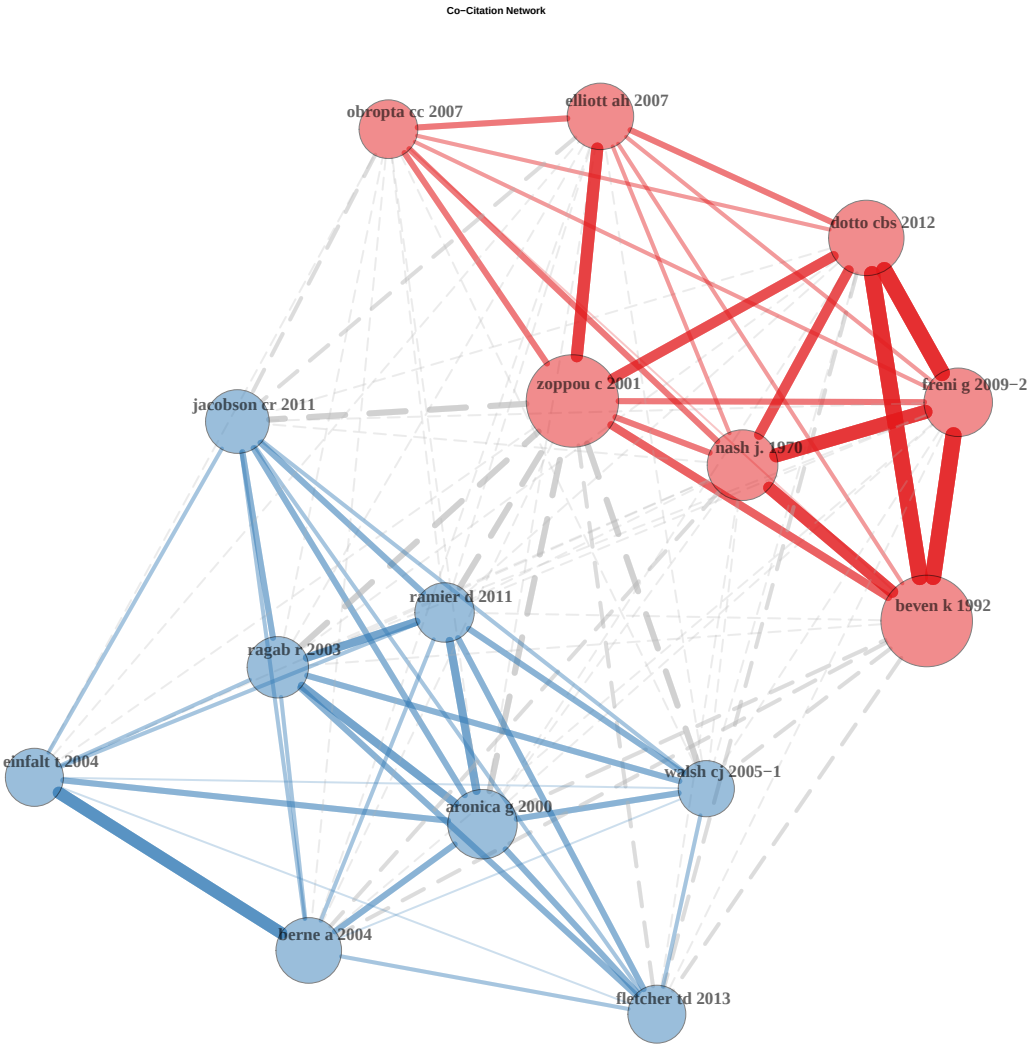


Figure 1.5: Co-citation network derived from the bibliometric analysis results.

phometry, water-temperature, underwater light availability, mixing conditions and food web structure (Dokulil and Teubner, 2000). Severe consequences of algal blooms may include toxicity. In this specific case, policy regulation regarding thresholds definition to limit the impact on the receiving water may be taken into account, e.g. maximum concentration of ammonia in the receiving water due to ammonium load spilled directly without treatment. Accounting for uncertainty constitutes an essential analysis to improve modelling of toxic impacts for the environment. Further research to understand the mechanism on how these negative impacts alters the natural environment, is encouraged for achieving SUWM.

Fletcher et al. (2013) reviewed the state of the art regarding the understanding, management and modelling of urban hydrology and its consequences for receiving waters. The authors depicted that urban hydrology helps in the management of flood protection, public health and environmental protection and highlighted that radar and microwave networks constitute technologies that contribute to the advancement in the measurement and prediction of urban precipitation. Also, advancement in models more suited to the finer temporal and spatial scales of urban and peri-urban applications is addressed. The authors recognised a trend towards approaches to restore pre-development flow-regimes and water quality. The increased recognition that the restoration to a more natural water balance benefits not only the environment, but enhances the liveability of the urban landscape (Fletcher et al., 2013).

Fletcher et al. (2013) also recognised that despite the advancements in urban hydrology, important challenges remain. The authors conclude *"Further research into the spatio-temporal dynamics of urban precipitation is required to improve short-term precipitation prediction. The performance of stormwater technologies in restoring the water balance and in removing emerging priority pollutants remain poorly quantified. All of these challenges are overlaid by the uncertainty of climate change, which imposes a requirement to ensure that stormwater management systems are adaptable and resilient to changes"*. This emphasises the need for uncertainty propagation techniques to support decision-making under uncertainty in SUWM.

Elliott and Trowsdale (2007) reviewed and compared ten existing stormwater models for modelling low-impact development urban stormwater drainage systems (LID). According to the authors, LID is an increasingly popular method to reduce the adverse hydrological and water quality effects of urbanisation, including new urban water management approaches to deliver improved environmental, economic, social and cultural outcomes. It is worth noting that the models compared in this study use a wide range of temporal resolutions, from average annual to sub-hourly. All but one of the models have capabilities for long-term simulation. The authors emphasised that the transition to more sustainable urban drainage design has been slow, mainly due to a lack of LID drainage design tools that operate effectively at the necessary range of scales, and concluded that there is a

considerable scope for improvement.

In addition to the analysis of the most cited papers from the bibliometric analysis, a further analysis of the collected literature regarding integrated catchment models, and uncertainty propagation analyses in urban hydrology modelling is presented below.

Zoppou (2001) presents a review of eight urban stormwater models specifically designed for simulating water quantity and quality. Differences in terms of water quantity variables and water quality constituents modelled by the models are given. In brief, not all models have routines or modules to model either chemical oxygen demand or ammonium (NH_4). Mitchell et al. (2007) review seven urban drainage models. They conclude that some models are weak in terms of handling temporal and spatial scales, input data uncertainty, and representation of urban infrastructure dynamics in long-term modelling (10 year horizon and beyond).

Bach et al. (2014) presents a critical review of IUDM and compares 20 different software tools used for integrated modelling in terms of nine different urban drainage processes, five urban drainage components, and eight types of model applications. As a future outlook of integrated urban water modelling, the authors highlight that improvements are required for representing spatial and temporal processes in the models (Mitchell et al., 2007), with special attention required to address long-time-series simulation (Rauch et al., 2002; Willems, 2006). Also, they highlight statements from Burger et al. (2010) and Burger et al. (2014), regarding the fact that in IUDM it is advisable to use parallel computing to improve the performance of existing software, and encourage researchers to be adaptive to the emerging computational technology.

The discussion presented by Bach et al. (2014) regarding uncertainty in IUDM pointed out, similar to Deletic et al. (2012), that uncertainty literature in integrated modelling lacks clarity and coherent approaches for assessment. Moreover, accounting for uncertainty assessment is still missing in integrated catchment studies. In this sense, some important attempts for considering uncertainty quantification in the urban water domain, specifically in an urban stormwater application was done by Bell et al. (2017), using Monte Carlo simulation and global sensitivity analysis of state variables to water quality parameters. Parameter uncertainty was quantified following the Generalized Likelihood Uncertainty Estimation (GLUE) approach (Beven and Binley, 1992), which acknowledges potential equifinality (Beven, 2006) in the parameter sets. Also, Miller and Hess (2017) highlight the importance of considering uncertainty in the modelling of the hydrological response of mixed urban-rural catchments. Therefore, there is evidence that uncertainty deserves attention and that although a large number of studies have been done on this in catchment hydrology, there are not many in integrated catchment studies.

One of the limitations highlighted in the literature, is that most models used in urban stormwater systems modelling are very complex and require a large amount of data for calibration and to simulate processes accurately. Following Meirlaen et al. (2001), these

complex mechanistic models are often computationally demanding. Therefore, this approach is impractical for long-term simulation and optimisation tasks. As an alternative, surrogate models are often proposed in the literature (Meirlaen et al., 2001; Vanrolleghem et al., 2005; Fu et al., 2009; Razavi et al., 2012; Brunetti et al., 2017; Mahmoodian et al., 2018).

Also, parallel computing and distributed scalable implementations for high performance computing are important to speed up the calibration and simulation time of models. Some examples of parallel computing are found in fields, such as, the application of watershed-distributed eco-hydrological models (Chen et al., 2014) and large-scale integrated hydrological modelling (Kollet et al., 2010; Maxwell, 2013). Examples in the urban drainage domain are scarce, and appear to be limited to Claeys et al. (2006), Burger et al. (2014), and Burger et al. (2016).

Another limitation found in urban stormwater systems modelling literature is related to precipitation as model input. Precipitation is a main driver of USSM's and there are substantial uncertainties associated when assessing precipitation in space and time at several scales. This was already noted in Section 1.1. The spatial distribution of precipitation is rarely considered in USSM's, and frequently precipitation is assumed to be uniformly distributed within a sub-catchment.

In addition, most urban drainage models do not pay attention to uncertainty propagation (Mitchell et al., 2007; Bach et al., 2014). In particular, commercial software packages as used in engineering practice typically ignore uncertainties, among others due to lack of user-friendly software implementations (Schellart et al., 2010). However, uncertainties can be substantial and ignoring these may affect decision making. In particular, end users should be aware of uncertainties so that they can take more robust decisions. Futhermore, the bibliometric analysis showed that the current state of knowledge regarding uncertainties in urban drainage modelling is poor (Deletic et al., 2012). Thus, research into uncertainty propagation through USSM's and development of operational systems that can trace the propagation of uncertainties is needed (Bach et al., 2014).

The review brings light also to the need for integrated catchment models that run fast and can easily be run in batch mode to meet the demands of Monte Carlo based uncertainty propagation analyses in urban hydrology modelling, in particular for end users of such models. There is a clear need for a formalised uncertainty propagation framework for USSM's, Tscheikner-Gratl et al. (2019) is a good example toward this formalisation. As suggested by Neumann (2007), when designing combined sewer overflows (CSOs), it is important to distinguish between the effects induced by model inputs, such as caused by precipitation variability, and effects due to model parameter uncertainty. In addition, there is also model structural uncertainty (Brown and Heuvelink, 2005). Three main sources of uncertainty may be distinguished in urban stormwater systems modelling: (1) model input, which is related to errors in input data, i.e. in driving forces such as pre-

cipitation; (2) model parameterisation, which is related to the uncertainty regarding the (calibrated) parameters of the model; and (3) model structure, which relates to uncertainty due to model conceptualisation and simplification. A holistic framework should ideally address all three sources and make use of Monte Carlo technique (Heuvelink, 1998; Robert and Casella, 2010) in order to propagate model input uncertainty.

1.4 Research objectives

This doctoral dissertation aims to fill the research gaps identified in the previous sections, and therefore the following research objectives are formulated, which are directly related to each of the four main problems identified in urban storm water modelling.

1.4.1 Development and implementation of a simplified and scalable urban water quality model

This objective is related to the development of a simplified mechanistic urban water model, EmiStatR, which represents the overall dynamic behaviour of the CSO spill volume, load, and concentration of COD and NH_4. It requires an implementation of the model in R with parallel computation capabilities, to allow fast and scalable calculations, particularly for scenarios with long simulation periods and for Monte Carlo uncertainty analyses. This objective also includes the calibration and application of EmiStatR to a Luxembourg case study and the validation of the model by comparing its performance against a complex mechanistic model that uses the "de Saint-Venant" partial differential equations to describe the flow routing in the pipes of a sewer network in a full hydrodynamic stormwater model.

1.4.2 Characterisation of precipitation time series accounting for uncertainty

This objective is linked to the development and presentation of a method to estimate the precipitation in a specific catchment given a known precipitation time series at a location outside the catchment, while also quantifying the uncertainty associated with this estimation. Through this objective, after uncertainty characterisation and uncertainty model implementation, simulated precipitation time series can then be used as model inputs to urban water models for Monte Carlo based uncertainty propagation analysis. Thus model output uncertainty can be quantified to support urban water system designs and better assess associated environmental and economic impacts.

1.4.3 Development of a software tool to facilitate uncertainty propagation analysis in urban water systems

The objective here is to develop a tool for characterising uncertainty in spatial, temporal and spatio-temporal environmental variables (model inputs) as probability distribution functions (pdfs) and as uni- and multi-variate autoregressive models. The development of this tool enhances tasks related to sampling from pdfs (to support Monte Carlo uncertainty propagation analysis) and to generate realisations of autoregressive models that represent uncertain environmental variables. The scope includes the application and illustration of the tool with a few simple examples of uncertainty propagation of model inputs in environmental modelling.

1.4.4 Application of uncertainty propagation to an urban water system

The last objective of this dissertation is built on the previous objectives and is related to the input uncertainty propagation through EmiStatR, taking into account the temporal auto- and cross-correlation of uncertain dynamic inputs, to quantify and assess the contributions of each uncertainty source to model output uncertainty dynamically (over time) for a case study in Luxembourg. For this, selection and characterisation of the main sources of input uncertainty accounting for the temporal auto- and cross-correlation within EmiStatR is performed.

1.5 Outline

This doctoral dissertation consists of six chapters. The first and last chapters present a general introduction and a synthesis, respectively. The four other chapters, Chapters 2 to 5, each address one of the four main research objectives presented in Section 1.4.

Chapter 1 (this chapter) introduces the problem of uncertainty analysis in integrated urban drainage modelling, presents the results of an extensive bibliometric analysis of relevant literature, and defines the objectives of this research.

In Chapter 2, the EmiStatR package is presented. This R-package has been developed as a simplified and scalable urban water quality model for simulation of combined sewer overflows. This model is needed to be able to do the Monte Carlo uncertainty analyses in later chapters.

In Chapter 3, a multivariate autoregressive modelling and conditional simulation of precipitation time series for uncertainty propagation is developed and presented. This is a requirement in order to perform an uncertainty propagation analysis with uncertain dynamic inputs, such as precipitation and concentrations of pollutants in water. To analyse how uncertainty propagates and to quantify the uncertainty in model output it is mandatory to first quantify uncertainty in the inputs, and that is done in this chapter for

precipitation as a main driving force in USSM's.

While Chapter 3 presents a methodology to quantify uncertainty in precipitation by a dynamic statistical model, there is also a need to quantify the uncertainty in other inputs to USSM's and have a framework in place to execute uncertainty propagation analyses using EmiStatR. Therefore, Chapter 4 presents stUPscales, an R-package for spatio-temporal uncertainty propagation across multiple scales and illustrates it with examples in urban water modelling.

Chapters 2 to 4 address preparations that are needed to perform an elaborate uncertainty propagation analysis for a real-world case. Chapter 5 presents a temporal uncertainty propagation analysis of an urban water system in Luxembourg.

Finally, in Chapter 6 a synthesis of the dissertation is presented, highlighting the lessons learned and future directions foreseen.

Chapter 2

EmiStatR: a simplified and scalable urban water quality model for simulation of combined sewer overflows

This chapter is based on:

Torres-Matallana, J. A.; Klepiszewski, K.; Leopold, U., and Heuvelink, G.B.M. EmiStatR: a simplified and scalable urban water quality model for simulation of combined sewer overflows. *Water*, 10(6)(782):1–24, 2018a. doi: 10.3390/w10060782. URL `https://www.mdpi.com/2073-4441/10/6/782`

Abstract

Many complex urban drainage quality models are computationally expensive. Complexity and computing times may become prohibitive when these models are used in a Monte Carlo (MC) uncertainty analysis of long time series, in particular for practitioners. Computationally scalable and fast "surrogate" models may reduce the overall computation time for practical applications in which often large data sets would be needed otherwise. We developed a simplified semi-distributed urban water quality model, EmiStatR, which brings uncertainty and sensitivity analyses of urban drainage water quality models within reach of practitioners. Its lower demand in input data and its scalability allow for simulating water volume and pollution loads in combined sewer overflows in several catchments fast and efficiently. The scalable code implemented in EmiStatR reduced the computation time significantly (by a factor of around 24 when using 32 cores). EmiStatR can be applied efficiently to test hypotheses by using MC uncertainty studies or long-term simulations.

keywords: urban water modelling; fast surrogate model; parallel computing

2.1 Introduction

Urban stormwater models are primary components of the monitoring system for real-time water flow and water quality simulation and prediction. In the literature, many urban hydrology models are well-established. However, there are few studies that attempt to model both flow and water quality taking into account the whole complexity of the physical, chemical, and biological processes involved (Willems, 2006; Beven, 2012). Moreover, urban water quality studies need to combine hydrological modelling of natural surfaces with the performance of urban man-made structures and impervious areas in a comprehensive hydrological modelling approach. The importance of access to and preservation of clean water is emphasised by the United Nations Sustainable Development Goals to "ensure availability and sustainable management of water and sanitation for all" (Goal 6) and to "conserve and sustainably use the oceans, seas and marine resources for sustainable development" (Goal 14) (United Nations - UN, 2015).

Zoppou (2001) presents a review of eight urban stormwater models specifically designed for simulating water quantity and quality: among others, Quantity–Quality Simulation (QQS) (Geiger and Dorsch, 1980); Storm Water Management Model (SWMM) (Huber and Dickinson, 1988); and MIKE-SWMM, a combination of MIKE 11 (Havno et al., 1995) and SWMM. Although QQS can simulate chemical oxygen demand (COD) and total nitrogen, it does not provide the capability to simulate ammonium (NH_4). Similarly, the reviewed SWMM version does not provide a routine for simulating COD or NH_4. Additionally, although MIKE-SWMM simulates several water quality variables, it does not provide for specific simulation of COD.

Mitchell et al. (2007) present a state-of-art review of integrated urban drainage models, in which a detailed review of seven models was conducted: Aquacyle (Mitchell et al., 2001), Hydro Planner (Maheepala et al., 2005), Krakatoa (Stewardson et al., 1995), UrbanCycle (Hardy et al., 2005), Mike Urban (DHI, 2007), UVQ (Mitchell and Diaper, 2005), and WaterCress (Clark et al., 2002). Mitchell et al. (2007) concluded that these models are weak in terms of handling temporal and spatial scales, input data uncertainty, and representation of urban infrastructure dynamics over time within a 10 to 100 year horizon.

Bach et al. (2014) present a critical review of integrated urban drainage modelling (UDM) and compared 20 different software tools used for integrated modelling: among others, integrated urban drainage models (IUDMs) such as InfoWorks CS (MWH Soft, 2010), Simulation of Biological Wastewater Systems (SIMBA) (IFAK, 2007), SWMM (Rossman, 2004), and WEST (Vanhooren et al., 2003); integrated urban water cycle models (IUWCMs) such as City Drain 3 (Burger et al., 2010), Model for Urban Stormwater Improvement Conceptualisation (MUSIC) (CRC-CH, 2005), MIKE URBAN (DHI, 2009), UrbanCycle (Hardy et al., 2005), and UrbanDeveloper (eWater, 2011); and integrated

urban water system models (IUWSMs) such as Dynamic Adaptation for eNabling City Evolution for Water (DAnCE4Water) (Rauch et al., 2012). In their comparison, they evaluated nine different urban drainage processes, five urban drainage components, and eight types of model applications. As a future outlook of integrated urban water models, they highlight that improvements are required for representing spatial and temporal processes in these models (Mitchell et al., 2007), with special attention required to address long-time-series simulation (Rauch et al., 2002; Willems, 2006). Additionally, they recognise that integrated urban water modelling must explore parallel computing with efforts to improve the performance of existing software (Burger et al., 2010, 2014) and encouraged researchers to be adaptive to the emerging computational technology. The review above suggests that there is still room to improve urban water models, specifically in the case of urban drainage models. One of the problems is that most models are very complex and require a large amount of data for calibration and to simulate processes accurately. Following Meirlaen et al. (2001), these complex mechanistic models describe the flow routing in pipes by the de Saint-Venant equations, which are based on the conservation of mass and momentum. These partial differential equations are solved by numerical algorithms that are often computationally demanding. Therefore, this approach is impractical for long-term simulation or optimisation tasks. As an alternative, surrogate models are frequently mentioned in the literature Meirlaen et al. (2001); Vanrolleghem et al. (2005); Fu et al. (2009); Razavi et al. (2012); Brunetti et al. (2017). These models are faster and represent an approximate substitute of the "real process", that is, the complex mechanistic model that better represents reality. Meirlaen et al. Meirlaen et al. (2001) distinguish between two types of possible simplifications, the empirical (black box) and the mechanistic (white box) approaches, and present a framework for developing a mechanistic surrogate model from a complex mechanistic model (CMM), reducing the computational time by a factor of 3.

Jin (Jin, 2005) presents a comprehensive survey of fitness approximation in evolutionary computation, whereby polynomials, the kriging model, neural networks, and support vector machines are described as the most often used methods of surrogate modelling to improve computational efficiency. However, these methods are of the black box type, which implies that the physical description and meaning of the processes that are simulated are lost.

From a different perspective, efforts have been made to simplify CMMs (Meirlaen et al., 2002; Freni et al., 2008; Mannina and Viviani, 2010; Willems, 2010; Coutu et al., 2012; Vezzaro et al., 2014), but these approaches remain complex. Complex urban drainage models can be even more troublesome when Monte Carlo (MC) based uncertainty propagation analysis is required, because this analysis requires formidable computation times. Therefore, it becomes increasingly important to address scalability issues (Burger et al., 2014). By scalability, we refer to the capability to deploy adaptive algorithms and run models efficiently in different hardware configurations (i.e., the number of threads) in

distributed computing environments. Parallel computation is a key component in hydrological modelling for expediting computations.

To the best knowledge of the authors, in the realm of urban drainage modelling, there are only a few examples of scalable implementations for solving intensive computational tasks in watershed-distributed or semi-distributed modelling. Some examples of parallel computing are found in other fields, such as in the application of watershed-distributed eco-hydrological models (Chen et al., 2014) and in large-scale integrated hydrological modelling (Kollet et al., 2010; Maxwell, 2013), but examples in the urban drainage domain are very scarce (Claeys et al., 2006; Burger et al., 2014, 2016).

The above indicates that urban drainage modelling still requires simplified or surrogate models and implementations of parallel computing, specifically scalable frameworks, that are, in addition, easily accessible. In this chapter, we address this need by developing and presenting EmiStatR, "Emissions and Statistics in R for Wastewater and Pollutants in Combined Sewer Systems", a mechanistic simplified urban water model for the simulation of Combined sewer overflow (CSO) emissions. Specifically, we contribute with a tool for performing short- and long-term simulations, developed in a parallel computing framework and allowing fast calculations while preserving the physical description and meaning of the processes simulated.

We also demonstrate that it is possible to obtain similar accuracy for water quantity and quality with this simplified and scalable model, compared to results of a complex mechanistic full hydrodynamic model. We focus on COD and NH_4 as water quality measures. COD is a standard for dimensioning CSO structures. NH_4 represents a diluted substance that can have a significant impact on surface water quality because of possible transformation to ammonia (NH_3). Additionally, COD and NH_4 are key variables for evaluation of the performance of wastewater treatment plants (WwTPs) and the quality status of receiving water bodies. A detailed outlook regarding the relevance of transformation and nutrient removal from the water column is presented by Bell and co-workers (Bell et al., 2017).

This chapter has three main objectives: (1) the development of a simplified mechanistic urban water model, EmiStatR, which represents the overall dynamic behaviour of the CSO spill volume, load, and concentration of COD and NH_4; (2) the presentation of an implementation of the model in R with parallel computation capabilities, allowing fast and scalable calculations, particularly for scenarios with long simulation periods and in MC uncertainty propagation mode; (3) the calibration and application of EmiStatR to a Luxembourg case study and validation by comparing the performance against a CMM that uses the de Saint-Venant partial differential equations to describe the flow routing in the pipes of the sewer network.

2.2 Methods

EmiStatR targets the simulation of CSO emissions of pollutants to the receiving water body, in terms of indicator variables, such as COD and NH_4. In this section, we describe the conceptual and mathematical model and its implementation in R.

2.2.1 Conceptual Model

The EmiStatR model includes six main components to simulate combined sewage discharges of a catchment (Figure 2.1):

1. Dry weather flow (DWF): EmiStatR assumes a constant DWF resulting from specific water consumption per population equivalent (PE) and a specific discharge of infiltration inflow per hectare of contributing impervious area to combined sewage flow (CSF).
2. Pollution of DWF: This is the specific load contribution per PE and day of COD and NH_4. No pollutant contribution of infiltration inflow is taken into account.
3. Rain weather flow (RWF): This is the total run-off of rainfall on the impervious catchment area contributing to CSF. The RWF is discharged in a specific flow time (t_{fS}) to the sewer outlet or CSO structures downstream from the catchment; that is, the flow time in the sub-catchment (t_{fS}) is a parameter of calibration.
4. Pollution of RWF: Constant surface run-off concentrations of COD and NH_4 are assumed. EmiStatR further assumes the complete mixing of pollutants in simultaneously flowing volume components and CSO chamber (CSOC) structures.
5. CSF and pollution: These are the sum of the DWF and RWF for the CSF and the consequent pollution load.
6. CSO volume and pollution: These are the volume diverted towards the receiving water body that is produced when the overflow or spill weir level in the CSOC is exceeded and the pollution measured as COD and NH_4 loads.

As shown in Figure 2.1, the sewer system under investigation includes a CSOC structure to store first-flush pollutant peaks. After filling of the storage capacity, the excess volume and pollutant inflows are discharged through a combined sewage spill structure. The excess flow and pollutant load are not conveyed to the WwTP but are diverted directly to the receiving water (i.e., the environment).

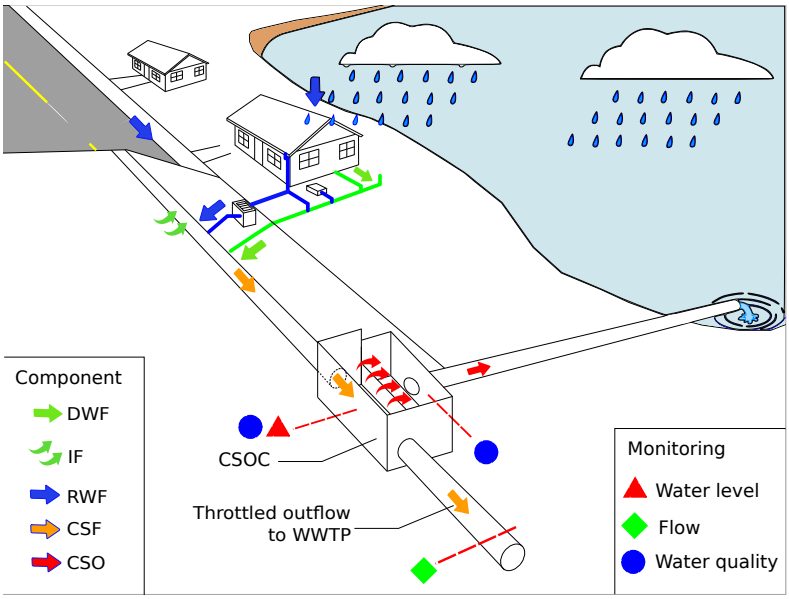


Figure 2.1: Main components of the EmiStatR model: (1) Dry weather flow (DWF) including infiltration flow (IF), (2) pollution of DWF, (3) rain weather flow (RWF), (4) pollution of RWF, (5) combined sewage flow (CSF) and pollution, and (6) combined sewer overflow (CSO) and pollution. CSOC—CSO chamber (background adapted from Sanitary-District (2015)).

In EmiStatR a simple volume balance taking into account inflow volume, present storage capacity, and outflow to the WwTP is implemented to simulate the CSOC structure. In case of a spill, the pollutant concentrations in the CSO are equivalent to the combined sewage inflow concentrations of the structure.

At the CSOC structure, a simple volume balancing takes place: (1) Substance and volume flows are stored and discharged to the WwTP if the storage volume is not completely filled up. (2) If the storage volume is completely filled up, the proportion of the volume inflow that is not discharged to the WwTP goes to the CSO.

2.2.2 Governing Equations

Dry Weather Flow

The DWF, Q_{s24} ($\mathrm{L} \cdot \mathrm{s}^{-1}$), is the product of the residential wastewater flow per PE, qs, and the PEs connected to the CSO structure, pe_i. The time series qs may follow a daily pattern given by a technical association for wastewater and water management but may also be a user-defined daily or weekly pattern, thus allowing different parts of the week to be differentiated between, for example, weekdays and weekends. Moreover, seasonal patterns can be defined to account for differences between months or seasons. The time series of PEs pe_i can also vary over time to account for differences between weekdays and weekends or for seasonal effects, such as because of tourism. The time series qs and pe_i

are of lengths equal to that of the rainfall time series, P_1 (see Section 2.2.2). The DWF is calculated as

$$Q_{s24_i} = \frac{1}{86,400} \cdot pe_i \cdot qs_i, \tag{2.1}$$

where

i is the ith term of the time series $(-)$;

pe are the PEs of the connected CSO structure at time i (PE);

PE is the units for PEs (unit per capita loading); and

qs is the individual water consumption at time i (residential) $(\mathrm{L} \cdot \mathrm{PE}^{-1} \cdot \mathrm{day}^{-1})$.

We note that pe refers to the time series of PEs with units PE. The number 86,400 is a factor for unit conversion (from days to seconds). The infiltration flow, Q_f $(\mathrm{L} \cdot \mathrm{s}^{-1})$, is computed as

$$Q_{f_i} = A_{imp} \cdot q_{f_i}, \tag{2.2}$$

where

A_{imp} is the impervious area of the catchment (ha); and

q_f is the specific infiltration water inflow at time i $(\mathrm{L} \cdot \mathrm{s}^{-1} \cdot \mathrm{ha}^{-1})$.

Consequently, the total DWF, Q_{t24_i} $(\mathrm{L} \cdot \mathrm{s}^{-1})$, is calculated as

$$Q_{t24_i} = Q_{s24_i} + Q_{f_i}. \tag{2.3}$$

The contribution of DWF to the combined sewage volume during a time interval Δt (min) is called the "dry weather volume" (amount of dry weather water in CSF), V_{dw} (m^3):

$$V_{dwi} = 0.06 \cdot \Delta t \cdot Q_{t24_i}. \tag{2.4}$$

The number 0.06 is a factor for unit conversion (from minutes to seconds and from litres to cubic metres).

DWF Pollutants

The time series of two dry weather pollutant concentrations are calculated: the COD concentration, C_{COD} $(\mathrm{mg} \cdot \mathrm{L}^{-1})$, and the NH_4 concentration, C_{NH_4} $(\mathrm{mg} \cdot \mathrm{L}^{-1})$. These are

time series with length equal to P_1 and that make use of $C_{COD,S}$ and $C_{NH4,S}$, which are assumed to be constant:

$$C_{COD\,i} = \frac{10^3 \cdot pe_i \cdot C_{COD,S}}{qs_i \cdot pe_i + 86,400 \cdot A_{imp} \cdot q_{f_i}}, \tag{2.5}$$

$$C_{NH4\,i} = \frac{10^3 \cdot pe_i \cdot C_{NH4,S}}{qs_i \cdot pe_i + 86,400 \cdot A_{imp} \cdot q_{f_i}}, \tag{2.6}$$

where

$C_{COD,S}$ is the COD sewage pollution per capita (PE) load per day ($\mathrm{g \cdot PE^{-1} \cdot day^{-1}}$);

$C_{NH4,S}$ is the NH_4 sewage pollution per capita (PE) load per day ($\mathrm{g{\cdot}PE^{-1}{\cdot}day^{-1}}$).

Rain Run-Off Volume and Rain Weather Flow

The contribution of rainwater to the combined sewage volume is called the "rainwater volume", V_r ($\mathrm{m^3}$). This is a vector whose length is equal to that of P_1. P_1 is delayed by t_{fS} time steps that represent a delay in time response related to flow time in the sewer system. The parameter t_{fS} may be calibrated with observed data. The rainwater volume accumulated during a time interval Δt(min) is computed as

$$V_{r\,i} = 10 \cdot P_{1\,i-t_{fS}} \cdot [C_{imp} \cdot A_{imp} + C_{per}(A_{total} - A_{imp})], \tag{2.7}$$

where

P_1 is the rainfall depth per time step (Δt) at time i (mm),

A_{imp} is the impervious area of the catchment (ha);

A_{total} is the total area of the catchment (ha);

C_{imp} is the run-off coefficient for impervious areas (−); and

C_{per} is the run-off coefficient for pervious areas (−).

Combined Sewage Flow

The calculation of the CSF is done by introducing the concept of the "combined sewage mixing ratio", cs_{mr} (–). This is the ratio between V_{dw} and V_r:

$$cs_{mr\,i} = \begin{cases} 0, & \text{if } V_{r\,i} \leq \epsilon, \\ \frac{V_{dw_i}}{V_{r\,i}}, & \text{if } V_{r\,i} > \epsilon, \end{cases} \tag{2.8}$$

where

ϵ is the precision term equal to 10^{-5} (–).

CSO Volume

The calculation of the CSO volume is based on the excess volume stored in the Combined sewer overflow chamber (CSOC). The volume in the CSOC is calculated from the curve water level versus water volume. This requires that an initial water level in the CSOC, Lev_{ini} (m), is provided by the user. The throttled outflow or pass-forward flow of the CSOC, Q_d, conveyed towards the WwTP is defined by the discharge of the CSOC by an orifice:

$$Q_{di} = \begin{cases} 10^3 \cdot C_d \cdot A_d \left(2 \cdot g \cdot Lev_{ini}\right)^{0.5} & \text{if } i = 1, \\ 10^3 \cdot C_d \cdot A_d \left(2 \cdot g \cdot Lev_i\right)^{0.5} & \text{if } i > 1, \end{cases} \tag{2.9}$$

where

Q_d is the throttled outflow to the WwTP ($\mathrm{L} \cdot \mathrm{s}^{-1}$);

C_d is the orifice coefficient of discharge (−);

A_d is the orifice area ($\pi \cdot D_d^2/4$) (m^2);

g is the gravitational acceleration, 9.81 $\mathrm{m \cdot s}^{-2}$; and

Lev is the water level in the CSOC (m).

After computation of Q_d, it is checked whether the value obtained is below the maximum throttled outflow, $Q_{d,max}$ ($\mathrm{L} \cdot \mathrm{s}^{-1}$), defined by the user:

$$Q_{di} = \begin{cases} Q_{di} & \text{if } Q_{di} < Q_{d,max}, \\ Q_{d,max} & \text{if } Q_{di} \geq Q_{d,max}. \end{cases} \tag{2.10}$$

Four stages of the CSOC for calculation of the CSO volume are defined: (1) Filling up, characterised by the CSOC filling up volume, $V_{Chamber}$, of the CSOC; (2) CSO, characterised by the completely filled CSOC volume, $V_{Chamber}$, being equal to the volume of the CSOC; (3) stagnation: Characterised by the CSOC filling up volume, $V_{Chamber}$, of the CSOC being equal to zero; (4) emptying: Characterised by the CSOC filling up volume, $V_{Chamber}$, of the CSOC. A status variable is defined to determine when the CSOC is filling up:

$$o_{cfyn_i} = \begin{cases} 1 & \text{if } \left(V_{ri} + V_{dw_i}\right) > V_{di}, \\ 0 & \text{if } \left(V_{ri} + V_{dw_i}\right) \leq V_{di}, \end{cases} \tag{2.11}$$

with

$$V_{d_i} = 0.06 \cdot \Delta t \cdot Q_{d_i}, \tag{2.12}$$

and where

o_{cfyn} is the status variable for the CSOC filling up or spilling out (1—filling up; 0—spilling out) (−);
V_d is the volume of throttled outflow to the WwTP at time i (m^3); and
Q_d is the throttled outflow to the WwTP at time i ($L \cdot s^{-1}$).

After checking whether the CSOC is filling up or not, the volume $V_{Chamber}$ (m^3) is calculated:

$$V_{Chamber\,i} = \begin{cases} 0 \quad \text{if } i = 1, \\ \underbrace{V_{Chamber\,i-1} + \Delta V_i \quad \text{if } (o_{cfyn} = 1) \wedge [V_{Chamber\,i-1} < (V - \Delta V_i)]}_{\text{filling up}}, \\ \underbrace{V \quad \text{if } (o_{cfyn} = 1) \wedge [V_{Chamber\,i-1} \geq (V - \Delta V_i)]}_{\text{spill}}, \\ \underbrace{0 \quad \text{if } (o_{cfyn} = 0) \wedge [(V_{Chamber\,i-1} + \Delta V_i) \leq \epsilon]}_{\text{stagnation}}, \\ \underbrace{V_{Chamber\,i-1} + \Delta V_i \quad \text{if } (o_{cfyn} = 0) \wedge [(V_{Chamber\,i-1} + \Delta V_i) > \epsilon]}_{\text{emptying}}, \end{cases} \tag{2.13}$$

with

$$\Delta V_i = V_{r\,i} + V_{dw\,i} - V_{d_i}, \tag{2.14}$$

and where

V is the volume of the CSOC (m^3).

Upon calculation of the filling up volume, the CSO spill volume, V_{Sv} (m^3), is calculated as

$$V_{Sv\,i} = \begin{cases} \Delta V_i & \text{if } V_{Chamber\,i} = V, \\ V_{Chamber\,i} - V & \text{if } V_{Chamber\,i} > V, \\ \epsilon & \text{if } V_{Chamber\,i} < V. \end{cases} \tag{2.15}$$

CSO Pollutants

The spill emissions of COD and NH_4 are calculated in two steps: (1) Calculation of the COD and NH_4 spill loads, $B_{COD,Sv}$ and $B_{NH4,Sv}$, respectively; and (2) Calculation of the COD and NH_4 spill concentrations, $C_{COD,Sv}$ and $C_{NH4,Sv}$, respectively.

$$B_{COD,Sv_i} = \begin{cases} \frac{V_{Sv_i} \cdot cs_{mr_i}}{cs_{mr_i}+1} C_{COD_i} + \frac{V_{Sv_i}}{cs_{mr_i}+1} COD_{r_i} & \text{if } V_{Sv_i} > \epsilon, \\ \epsilon & \text{if } V_{Sv_i} \leq \epsilon. \end{cases} \tag{2.16}$$

$$B_{NH4,Sv_i} = \begin{cases} \frac{V_{Sv_i} \cdot cs_{mr_i}}{cs_{mr_i}+1} C_{NH4_i} + \frac{V_{Sv_i}}{cs_{mr_i}+1} NH4_r & \text{if } V_{Sv_i} > \epsilon, \\ \epsilon & \text{if } V_{Sv_i} \leq \epsilon. \end{cases} \tag{2.17}$$

Here,

$B_{COD,Sv}$ is the COD load in the spill volume (g);

COD_r is the rainwater pollution − COD concentration ($mg \cdot L^{-1}$);

$B_{NH4,Sv}$ is the NH_4 load in the spill volume (g); and

$NH4_r$ is the rainwater pollution − NH_4 concentration ($mg \cdot L^{-1}$).

COD_r can be a time series of length equal to P_1 or a unique value constant in time. The emissions in terms of the concentrations of COD and NH_4 are calculated and make use of $C_{COD,Sv}$ ($mg \cdot L^{-1}$), which is defined as the ratio of $B_{COD,Sv}$ and V_{Sv}. Similarly, $C_{NH4,Sv}$ ($mg \cdot L^{-1}$) is the ratio of $B_{NH4,Sv}$ and V_{Sv}.

2.2.3 Model Implementation in R

EmiStatR is available for download from the Comprehensive R Archive Network (CRAN) (`https://cran.r-project.org/web/packages/EmiStatR/`). This includes a user manual with several examples that can be run in R (Appendix B). The entire work flow for EmiStatR is illustrated in Figure 2.2.

Input Data Definition

EmiStatR is implemented in R (R-Core-Team and contributors worldwide, 2017) by defining a specific input() class (Figure 2.2). The model input data are set up in the class input() and can be grouped into three main categories (Table 2.1, columns 1 and 2):

1. Wastewater production data, that is, water consumption in PE and characterisation of the pollution load of wastewater in terms of COD and NH_4 concentrations in PE.
2. Run-off and specific pollutant load contribution per PE and day (COD and NH_4 concentrations) of infiltration water.

3. Precipitation data, that is, time series of rainfall and rainfall run-off pollution in terms of concentrations of COD and NH_4.

Table 2.1: General and combined sewer overflow (CSO) structure input data of EmiStatR.

General Input	Units	CSO Input	Units
1. Wastewater		*1. Identification*	
Water consumption, *qs*	$L \cdot PE^{-1} \cdot day^{-1}$ [a]	ID of the structure	–
Water consumption, factors [b]	–	Name of the structure	–
Pollution COD [c], $C_{COD,S}$	$g \cdot PE^{-1} \cdot day^{-1}$		
Pollution NH_4 [d], $C_{NH4,S}$	$g \cdot PE^{-1} \cdot day^{-1}$	*2. Catchment data*	
		Name of the municipality	–
2. Infiltration water		Name of the catchment	–
Inflow, q_f	$L \cdot s^{-1} \cdot ha^{-1}$	Number of the catchment	–
Pollution COD, COD_f	$g \cdot PE^{-1} \cdot day^{-1}$	Land use	–
Pollution NH_4, $NH4_f$	$g \cdot PE^{-1} \cdot day^{-1}$	Total area, A_{total}	ha
		Impervious area, A_{imp}	ha
3. Rainwater		Run-off coefficient for impervious area, C_{imp}	–
Precipitation time series, P_1	mm	Run-off coefficient for pervious area, C_{per}	–
Pollution COD, COD_r	$mg \cdot L^{-1}$	Flow time structure, t_{fS}	time step
Pollution NH_4, $NH4_r$	$mg \cdot L^{-1}$	Population equivalents, pe_i	PE
		Population equivalents, factors [b]	–
		3. CSO structure data	
		Volume, V	m^3
		Curve level–volume, $lev2vol$	m, m^3
		Initial water level, Lev_{ini}	m
		Maximum throttled outflow, $Q_{d,max}$	$L \cdot s^{-1}$
		Orifice diameter, D_d	m
		Orifice coefficient of discharge, C_d	–

[a] Population equivalent (PE). [b] Factors for daily, weekly, and monthly patterns. [c] Chemical oxygen demand (COD). [d] Ammonium (NH_4).

The general input variables of the CSO structure are grouped into three main components (Table 2.1, columns 3 and 4):

1. Identification, that is, ID and name of structure.
2. Catchment data, that is, name of the municipality, name and number of the catchment, land use (residential, commercial, and industrial), total area of the catchment, impervious area, and PEs connected to the sewer system.
3. CSO structure data, that is, data regarding the throttled outflow diverted to the WwTP and the total storage volume of the CSOC.

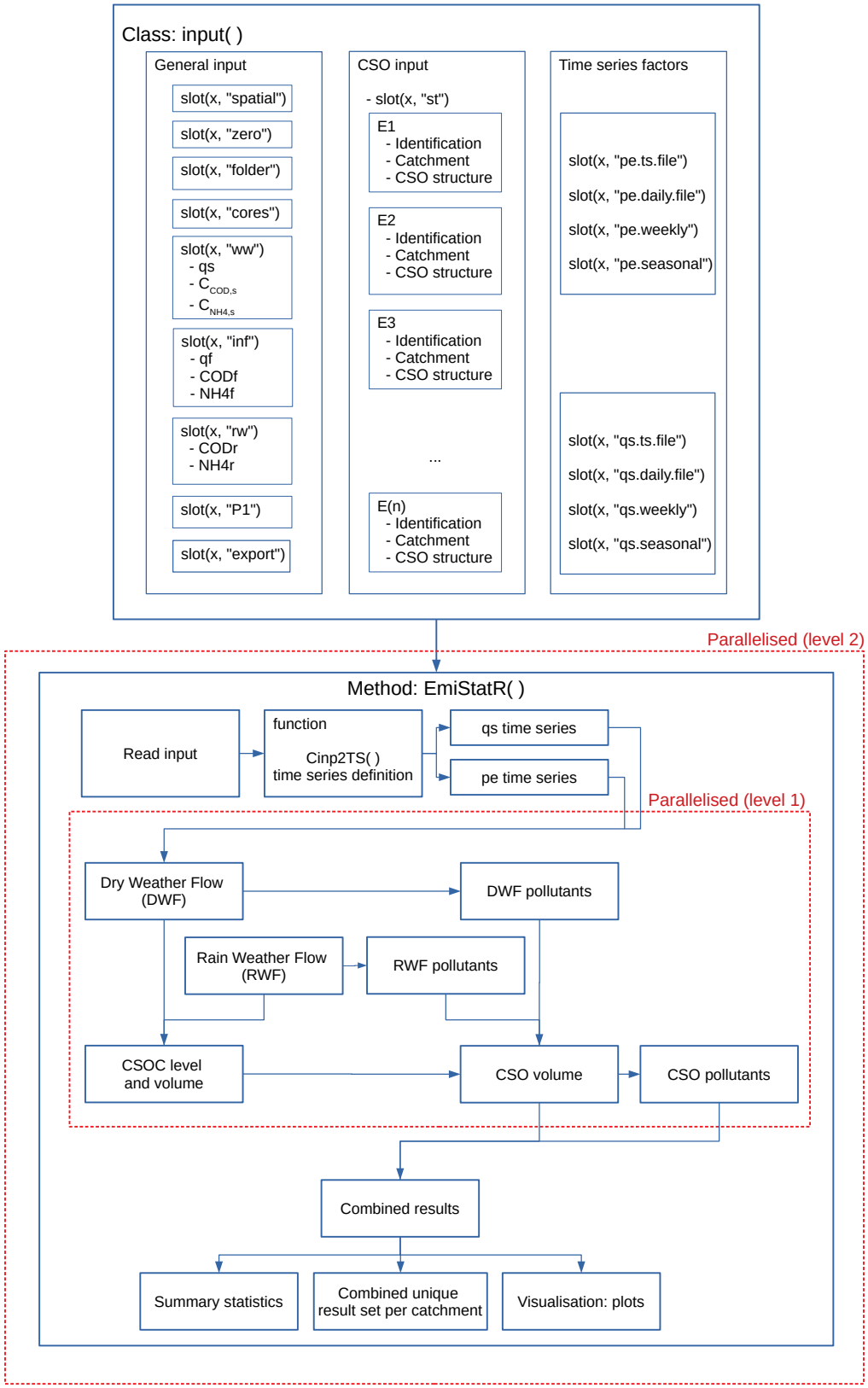


Figure 2.2: Workflow for EmiStatR and the parallelised approach. Parallelisation is set in the input() class, slot(x, "cores"). Level 1: Parallel computing is done inside EmiStatR. Level 2: Parallel computing is done outside of EmiStatR, e.g., Monte Carlo simulation or optimisation.

The main goal of EmiStatR is to simulate emissions of spill volume in individual CSO structures. If calibration data are available, EmiStatR parameters may be calibrated prior to simulation using the DiffeRential Evolution Adaptive Metropolis (DREAM) algorithm (Vrugt et al., 2009). The DREAM algorithm is integrated through the R package dream (Guillaume and Andrews, 2012). If calibration is not feasible, the model can also be run using parameter values taken from reference literature and guidelines. Table 4.3 provides reference values and calibration ranges for the most important EmiStatR parameters.

Table 2.2: Default values for input data of EmiStatR.

Input	Units	Reference Value	Literature Source	Range (This Study)
Wastewater				
Water consumption, qs	$L \cdot PE^{-1} \cdot day^{-1}$ [a]	150 [b]	Fan et al. (2013)	[130, 170]
Pollution COD [c], $C_{COD,S}$	$g \cdot PE^{-1} \cdot day^{-1}$	120	DWA (2002)	[90, 150]
Pollution TKN [d]	$g \cdot PE^{-1} \cdot day^{-1}$	11	DWA (2002)	[7, 15]
Pollution NH_4 [e]	$g \cdot PE^{-1} \cdot day^{-1}$	4.7	This study	[1, 8]
Infiltration water				
Inflow, q_f	$L \cdot s^{-1} \cdot ha^{-1}$	0.05	DWA (2006)	[0, 2]
Catchment data				
Run-off coefficient for impervious area, C_{imp}	–	See Rawls et al. (1981)	Rawls et al. (1981)	[0.20, 095]
Run-off coefficient for pervious area, C_{per}	–	See Rawls et al. (1981)	Rawls et al. (1981)	[0.05, 0.50]
Flow time structure, t_{fS}	time step	2	This study	[0, 12]
CSO structure data				
Initial water level, Lev_{ini}	m	L_{max} [f]/2	This study	[0, L_{max}]
Orifice coefficient of discharge, C_d	–	1.25	This study	[0.01, 2]

[a] PE: Population equivalent units; [b] mean value for European countries; [c] COD: Chemical oxygen demand; [d] TKN: Total Kjeldahl nitrogen; [e] NH_4: Ammonium; [f] L_{max}: Maximum water level in the combined sewer overflow chamber (CSOC).

Implementation of a Scalable Approach

Because MC analysis and long-term simulations of a large number of catchments EmiStatR may be slow, we made the code more scalable through parallel computation. This was done via the R package doParallel (Revolution Analytics and Weston, 2015b), which provides a parallel back-end for the functions of the foreach package (Revolution Analytics and Weston, 2015c). It depends on the R packages foreach, iterators (Revolution Analytics and Weston, 2015a), and parallel (R Core Team, 2017) and provides functionality for creating parallel loops through the foreach package. The doParallel package is an interface between the foreach and parallel packages of R 2.14.0 and later parallel wraps functions of the multicore (Urbanek, 2013) and snow packages (Tierney et al., 2016).

The parallel package evaluates larger chunks of code in parallel. In order to complete a computational task in parallel, these chunks should be evaluated independently, should take the same length of time, and should not communicate with each other. The typical parallel computational model is the following (R Core Team, 2017):

1. Start up M "worker" processes, and do any initialisation needed for the workers.

2. Send any data required for each task to the workers.
3. Split the task into M roughly equally sized chunks, and send the chunks (including the R code needed) to the workers.
4. Wait for all workers to complete their tasks, and collect results.
5. Repeat steps 1 to 4 for any further tasks.
6. Stop and close the worker processes.

In our specific case study, each of the M workers is related to a MC simulation. EmiStatR can also be parallelised for each sub-catchment to address scalability. An example is given in the EmiStatR package documentation. How the parallelisation is integrated into the entire workflow is illustrated in Figure 2.2, where parallelisation is set in the input() class, slot(x, "cores"). Level 1 indicates the parallel computation done inside EmiStatR. Level 2 indicates the parallel computation done outside of EmiStatR, for example, MC simulations or optimisation, in other R packages such as dream (Guillaume and Andrews, 2012) or stUPscales (Torres-Matallana et al., 2018b, 2019a) packages.

2.3 Case Study

2.3.1 Study Area

A test case was created to evaluate the use and performance of EmiStatR. A sub-catchment of the Haute-Sûre catchment in the northwest of Luxembourg was chosen. The combined sewer system of the sub-catchment drains the three villages Goesdorf, Kaundorf, and Nocher-Route. In the local sewer system downstream from the villages, three CSOCs are located to store pollutant peaks in the first flush of CSFs. Figure 2.3 depicts their locations and the delineation of the catchment. The topography of the area is characterised by a hilly landscape. The elevations around Goesdorf are between 390 and 490 m, around Kaundorf are between 370 and 464 m, and in the area of Nocher-Route vary between 400 and 485 m. The main land use types in the villages are residential, smaller industries, and farms. Outside of the villages, forest as well as agricultural areas and grassland are the dominating land uses. The receiving water bodies at CSO structures in Goesdorf, Kaundorf, and Nocher-Route are tributaries of the river Sûre (Sauer, in German) (Figure 2.3).

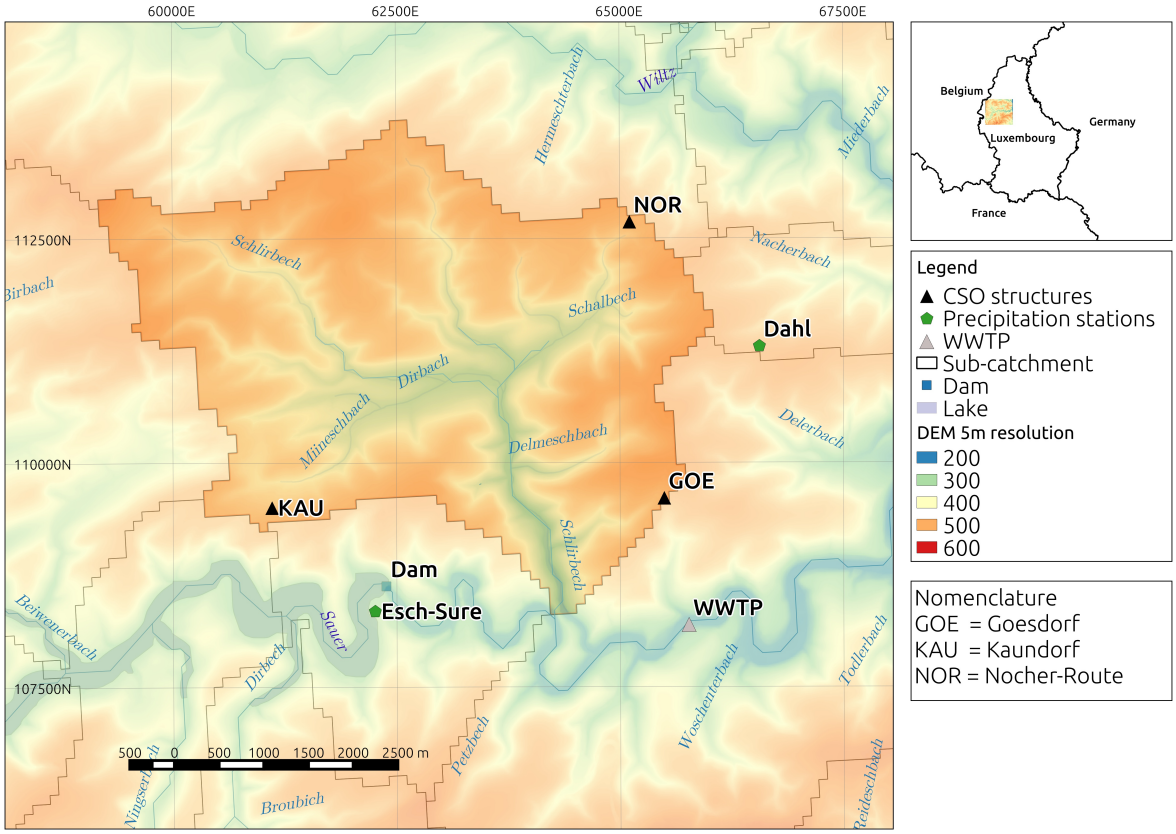


Figure 2.3: The Haute-Sûre sub-catchment. Combined sewer overflow (CSO) structures are located in Goesdorf (GOE), Kaundorf (KAU), and Nocher-Route (NOR).

2.3.2 Model Calibration

Measured precipitation time series at the Goesdorf CSOC served as input for the model calibration for water quantity output variables. This time series was recorded from May 15, 2011 to June 3, 2011 at 1 min resolution. Seven water quantity parameters were selected for calibration: (1) Water consumption, qs; (2) infiltration flow, q_f; (3) time flow, t_{fS}; (4) run-off coefficient for impervious area, C_{imp}; (5) run-off coefficient for pervious area, C_{per}; (6) orifice coefficient of discharge, C_d; and (7) initial level of water in the CSOC, Lev_{ini}.

For calibration, we used the DREAM algorithm (Vrugt et al., 2009). DREAM has the capability of running and evaluating multiple different chains simultaneously for global exploration. The algorithm tunes the proposal distribution in randomised subspaces during the search. DREAM enhances the applicability of Markov chain Monte Carlo (MCMC) sampling approaches in complex problems (Vrugt et al., 2009). The main building block of the DREAM algorithm is the Differential Evolution Markov Chain (DE-MC) method presented by ter Braak (Ter Braak, 2006). In DE-MC, different Markov chains are run simultaneously in parallel. At the current time, they form a population. Jumps in each

chain are generated by taking a fixed multiple of the difference of two random chains without replacement. To accept or reject candidate points, the Metropolis ratio is used (Ter Braak, 2006).

The DREAM algorithm is implemented in R in the R package dream (Guillaume and Andrews, 2012). Observations of water level in the Goesdorf storage CSOC served as reference for optimising the model parameters. The water level was recorded from April 19, 2011 to July 15, 2011 at 30 s time steps. The precipitation and water level observations were aggregated to 10 min intervals to assure that the model simulations and observations had the same temporal support before comparison. The observations were divided into two sets, one for calibration and one for validation. The calibration set comprised the initial section of the measurements from May 15 to June 3, 2011, a total of 2698 records at 10 min time steps. The validation set comprised the measurements from June 3 to July 7, 2011, a total of 4901 records at 10 min time steps.

DREAM optimises by minimising the root-mean-squared error (RMSE). As accuracy measures, the calibration results were evaluated by the mean error (ME), RMSE, and the Nash–Sutcliffle model efficiency coefficient (NSE) (Nash and Sutcliffe, 1970):

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(S_i - O_i)^2}, \tag{2.18}$$

$$\mathrm{ME} = \frac{1}{N}\sum_{i=1}^{N}(S_i - O_i), \tag{2.19}$$

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{N}(S_i - O_i)^2}{\sum_{i=1}^{N}(O_i - \bar{O})^2}, \tag{2.20}$$

where

O_i is the ith observation;

S_i is the ith simulation;

$\bar{O}$ is the mean of the observations; and

N is the number of observations (and simulations).

For Kaundorf and Nocher-Route, sufficient calibration data were not available. We therefore used the reference values (Table 4.3).

Regarding the water quality module of EmiStatR, six parameters are required to define pollution in terms of the following: (1) COD load per PE per day in the wastewater, $C_{COD,S}$; (2) NH_4 load per PE per day in the wastewater, $C_{NH4,S}$; (3) COD load per PE

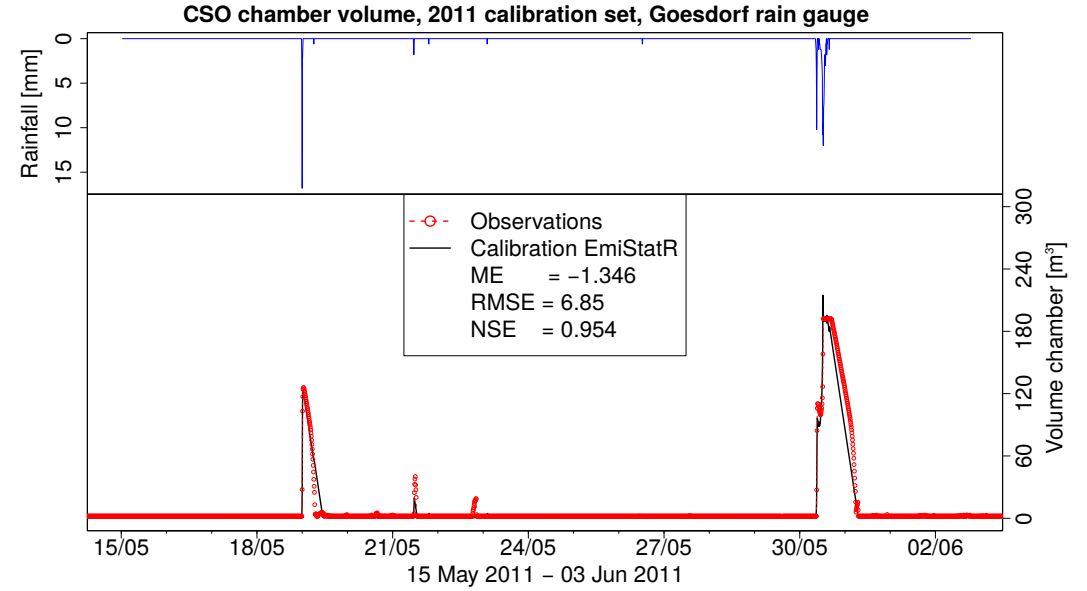
per day in the infiltration water, COD_f; (4) NH_4 load per PE per day in the infiltration water, $NH4_f$; (5) COD concentration in the run-off, COD_r; and (6) NH_4 concentration in the run-off, $NH4_r$. If these parameters are not measured directly, then they can be calibrated when observations of COD or NH_4 (concentrations or loads) in the output of the CSO spill volume are available. In this case study, we did not need to calibrate $C_{COD,S}$ and $C_{NH4,S}$ for Goesdorf, Kaundorf, or Nocher-Route, because 91 observations in total under DWF conditions were available. The measured $C_{COD,S}$ had a mean value of 104 $\mathrm{g \cdot PE^{-1} \cdot day^{-1}}$ with a standard deviation of 87.5 $\mathrm{g \cdot PE^{-1} \cdot day^{-1}}$. The measured $C_{NH4,S}$ had a mean value of 4.7 $\mathrm{g \cdot PE^{-1} \cdot day^{-1}}$ with a standard deviation of 1.92 $\mathrm{g \cdot PE^{-1} \cdot day^{-1}}$. The temporal support of these observations was 120 minutes. The other input parameters of the water quality module (COD_f, $NH4_f$, COD_r, and $NH4_r$) were set to zero, because the concentrations in rainfall and infiltration water were judged negligible compared to that of household sewage. We chose periods from 2010 and 2011 for both calibration and validation.

Calibration Results of the Water Quantity Model

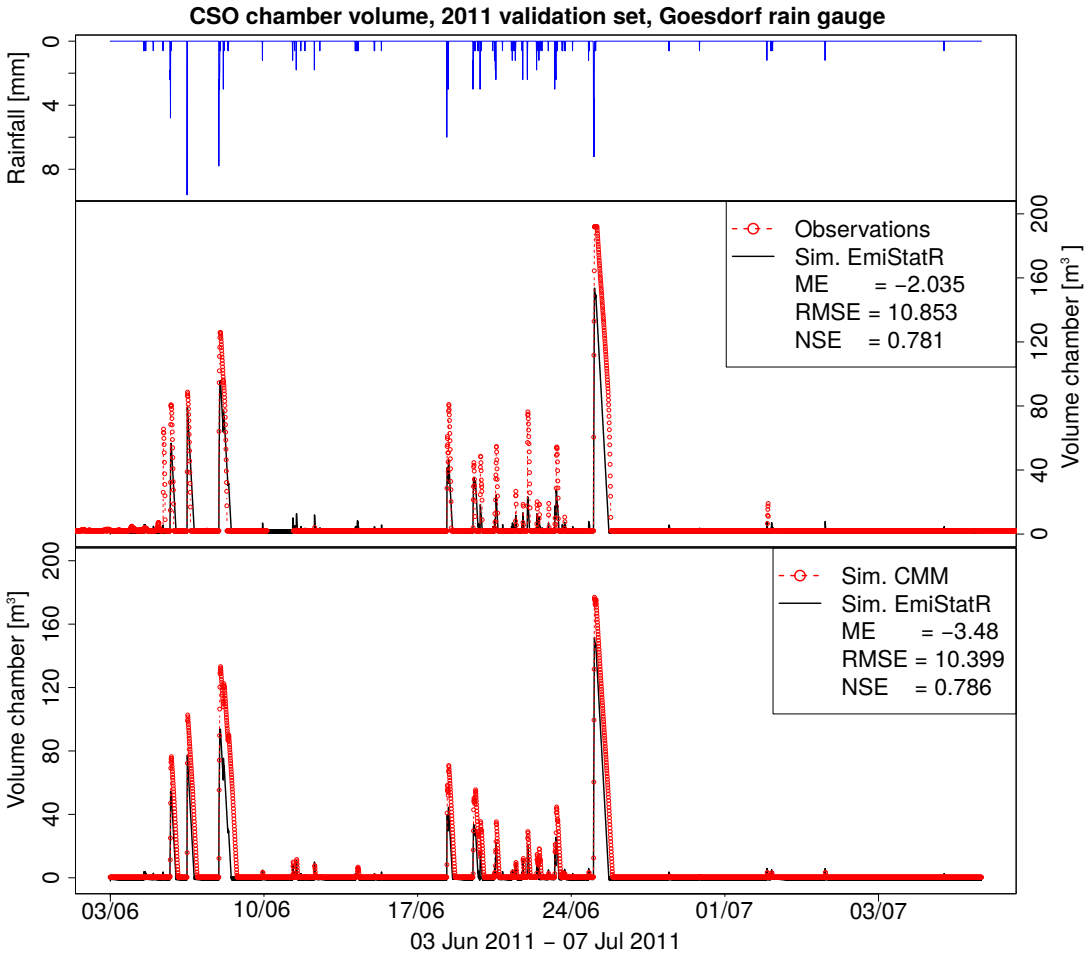
Table 2.3 and Figure 2.4a present the final calibration results of the hydraulic model implementing the DREAM algorithm. The calibration required 980 function evaluations. The optimised set of parameters produced a ME of -1.35 $\mathrm{m^3}$, RMSE of 6.85 $\mathrm{m^3}$, and NSE of 0.95. In this case, $Q_{d,max}$ was set to 5 $\mathrm{L \cdot s^{-1}}$ and V was set to 190 $\mathrm{m^3}$ (actual conditions for 2011). Figure 2.4a shows the precipitation input time series for the calibration dataset (upper inset) and the comparison of observed and simulated time series of the CSOC volume (bottom inset). For the events presented in Figure 2.4, the values of ME and RMSE are in cubic metres, whereas the NSE is dimensionless. From Figure 2.4a, it is possible to infer that after model calibration, the model could adequately simulate (NSE = 0.95) the volume in the CSOC. The model simulation was slightly under model observations specifically for low-rainfall conditions. Additionally, an over-prediction of the peak volume was presented in the simulation of the CSOC volume.

Table 2.3: Calibration and validation results of the hydraulic model in EmiStatR as calibrated with the DREAM algorithm (Goesdorf 2011, 10 min time step).

Parameter	Units	Range of Sampling	Calibrated Value
Water consumption, qs	$\mathrm{L \cdot PE^{-1} \cdot day^{-1}}$	[130, 170]	152
Infiltration flow, q_f	$\mathrm{L \cdot s^{-1} \cdot ha^{-1}}$	[0, 0.2]	0.116
Time flow, t_{fS}	time step	[0, 12]	1
Run-off coefficient for impervious area, C_{imp}	–	[0.20, 0.95]	0.28
Run-off coefficient for pervious area, C_{per}	–	[0.05, 0.50]	0.07
Orifice coefficient of discharge, C_d	–	[0, 2]	0.67
Initial water level, lev_{ini}	m	[0.1, 3.5]	0.57



(a)



(b)

Figure 2.4: Rainfall and combined sewer overflow chamber (CSOC) volume at Goesdorf. (**a**) Time series of May to June 2011 calibrated with DREAM; (**b**) June to July 2011 simulated time series for validation with observations and the CMM.

2.3.3 Validation of Model Predictions

Besides the calibration set, another set of measurements was used as independent observations to assess the accuracy of the model predictions for validation of the water quantity model. Input precipitation was recorded from June 3 to July 7, 2011 at a temporal resolution of 1 min, aggregated to 10 min. The observations of water level in the storage CSOC correspond to this period.

Figure 2.4b shows the results of the hydraulic model validation. It shows the precipitation input time series (upper inset), the comparison of observed and simulated time series of the CSOC volume (middle inset), and the comparison with the output of a CMM (bottom inset).

The CMM was implemented in the software InfoWorks ICM 7.5 (Innovyze Ltd, Wallingford, Oxfordshire, United Kingdom), and it served as a benchmark to calibrate and validate EmiStatR for water quantity and quality variables. The CMM was a full hydrodynamic flow and pollution load model, which implementd the de Saint Venant partial differential equations and was built initially in the software InfoWorks CS (Innovyze) ® (Schutz et al., 2012). This model was used to simulate surface run-off and discharge characteristics in local sewer systems and the behaviour of CSO structures in the Goesdorf sub-catchment and future sewer systems linked to weather periods. Besides the catchment data and structural data of sewer sections planned and in operation, the simulations were based on local rain data for local calibration and on regional long-term rain data to simulate the long-term performance of the system. In the framework of a coarse calibration and validation process, it was proved that the model reproduced discharge characteristics in local sewer systems of selected villages sufficiently. The resulting parameterisation to model surface run-off characteristics from impervious areas in the villages, such as initial losses, was applied to further catchments showing similar characteristics Schutz et al. (2012). The calibrated model of the catchment and drainage network of the case study, implemented in InfoWorks CS and upgraded to InfoWorks ICM 7.5, was used to validate the performance of EmiStatR. We followed a similar procedure as presented by Meirlaen et al. (2001) for developing a mechanistic surrogate model from a CMM.

In general, validation of a good agreement between simulation and observations was observed (NSE of 0.78). The model simulation results were slightly under the observations of the CSOC volume, and as a consequence, the peaks simulated were lower than those of observations, which agreed also with the behaviour shown in Figure 2.4a.

Regarding the water quality module of EmiStatR, we performed a validation on the basis of a 1 year simulation with the CMM. We ran the validation at 10 min time steps and aggregated the results to 120 min to eliminate short-time variability. Our interest was in the average load of pollutants over several hours, which corresponded well with the usual time for taking water samples for further laboratory analysis. The input values of the two

main parameters were 104 $g \cdot PE^{-1} \cdot day^{-1}$ for $C_{COD,S}$ and 4.7 $g \cdot PE^{-1} \cdot day^{-1}$ for $C_{NH4,S}$. These values corresponded to wastewater quality (WwQ) measurements. The total COD and NH_4 were monitored in the CSOC under DWF conditions. Figure 2.4b (bottom inset) shows how the model simulation agreed with observations (NSE of 0.79). The model simulation was also systematically below the observations of CSOC volume.

Additionally, to perform a more extensive validation of the water quality model, we compared its output with simulations obtained with the CMM for a 1 year time series at 10 min time steps. Table 2.4 and Figure 2.5 summarise the results of this validation. The results suggest that EmiStatR performed with good accuracy (NSE $\approx$ 0.80) when compared with the CMM.

Table 2.4: Comparison results for the complex mechanistic model (CMM) and EmiStatR (Esch-sur-Sure rain gauge 2 h averages over 1 year period).

Combined Sewer Overflow (CSO) Summary Results	**CMM**	**EmiStatR 1.2.1.0**
Period, p (day)	365	365
Duration of CSO spill volume, d_{Sv} (h)	90	100
Frequency of CSO spill volume, f_{Sv} (events)	19	16
Total CSO spill volume, V_{Sv} (m^3)	373	222
Average CSO, Q_{Sv} (L/s)	1.15	0.62
95th percentile of CSO spill volume, $V_{Sv,95}$ (m^3)	27.74	15.26
Maximum CSO spill volume, $V_{Sv,max}$ (m^3)	33.06	21.62
COD total load (BCOD), $B_{COD,Sv}$ (kg)	5.875	4.610
Average BCOD, $B_{COD,Sv,av}$ (kg)	0.131	0.092
95th percentile of BCOD, $B_{COD,Sv,95}$ (kg)	0.320	0.252
Maximum BCOD, $B_{COD,Sv,max}$ (kg)	0.450	0.360
NH_4 total load (BNH4), $B_{NH4,Sv}$ (kg)	0.224	0.208
Average BNH4, $B_{NH4,Sv,av}$ (kg)	0.005	0.004
95th percentile of BNH4, $B_{NH4,Sv,95}$ (kg)	0.012	0.011
Maximum BNH4, $B_{NH4,Sv,95}$ (kg)	0.020	0.020
Run time (min)	30	1.09

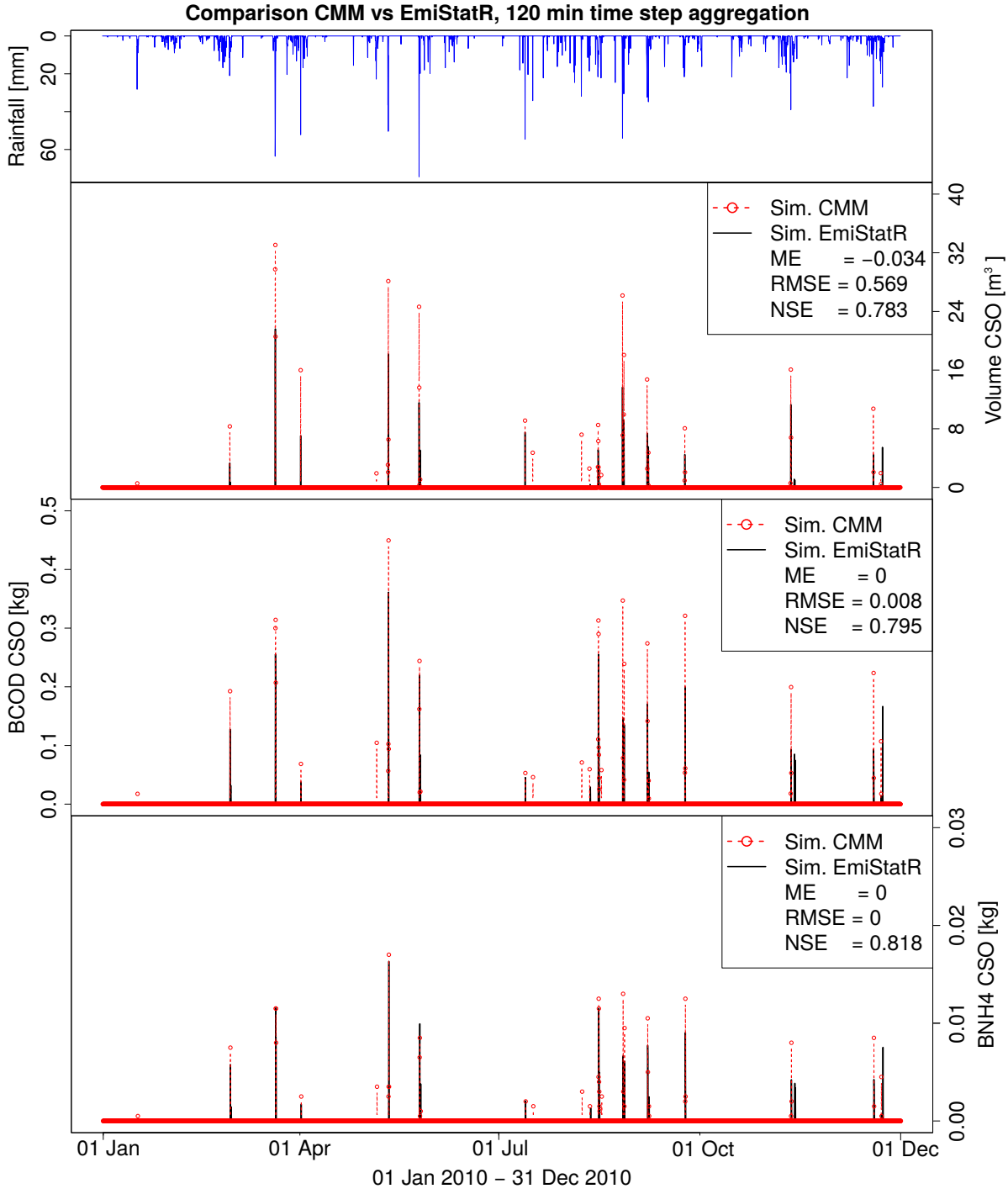


Figure 2.5: Rainfall (**top**), CSO volume (**second**), chemical oxygen demand (COD) load (**third**), and NH_4 load (**bottom**). January to December 2010 time series (Esch-sur-Sûre rain gauge) for validation of EmiStatR using output of a complex mechanistic model (CMM). Simulation at 10 min resolution at Goesdorf; results aggregated to 120 min.

2.3.4 Scalability and Performance

A hardware set-up was defined to execute the scalability test. We used an Intel(R) Xeon(R) CPU E7-L8867 server (Santa Clara, CA, USA) at 2.13 GHz with 40 physical cores (and 40 virtual cores) at 1.064 GHz, 516 GByte in random access memory (RAM), and the operating system (OS) Linux Ubuntu 12.04.5 LTS 64-bit. We used a maximum of $2^5 = 32$ cores. Additionally, we multiplied the number of simulations by 10 and 100 to evaluate the model runtime under repeated model calls, such as would typically be required in MC uncertainty analyses. As a result, the selected numbers of simulations were 32, 320 and 3200.

Regarding the results of the scalability test, the code implemented in EmiStatR allowed for specifying the number of cores to be used in the simulation according to the number of cores available. In the scalability test, a single simulation referred to a full year at 10 min time steps. We used the calibrated values for Goesdorf. For Kaundorf and Nocher-Route, we used the reference values given in Section 2.2.3. Tables 2.5 and 2.6 summarise the general input data and the CSO structures in the simulation mode, respectively.

Table 2.5: General input data of the EmiStatR scalability test.

General Input	**Units**	**Value**
Wastewater		
Water consumption, qs	$\mathrm{L \cdot PE^{-1} \cdot day^{-1}}$ [a]	150
Daily factors for water consumption, ATV-A134 curve	–	–
Pollution COD [b], $C_{COD,S}$	$\mathrm{g \cdot PE^{-1} \cdot day^{-1}}$	120
Pollution NH_4 [c], $C_{NH4,S}$	$\mathrm{g \cdot PE^{-1} \cdot day^{-1}}$	4.7
Infiltration water		
Inflow, q_f	$\mathrm{L \cdot s^{-1} \cdot ha^{-1}}$	0.05
Pollution COD, COD_f	$\mathrm{g \cdot PE^{-1} \cdot day^{-1}}$	0
Pollution NH_4, $NH4_f$	$\mathrm{g \cdot PE^{-1} \cdot day^{-1}}$	0
Rainwater		
Precipitation time series, P_1	mm	–
Pollution COD, COD_r	$\mathrm{mg \cdot L^{-1}}$	0
Pollution NH_4, $NH4_r$	$\mathrm{mg \cdot L^{-1}}$	0

[a] PE: population equivalent; [b] COD: chemical oxygen demand; [c] NH_4: ammonium.

Table 2.6: General input data of the combined sewer overflow (CSO) structures of the EmiStatR scalability test, after calibration for structure 1. Structures 2 and 3 were not calibrated; therefore, reference values were defined.

CSO Input	**Sub-Catchment**		
Identification			
ID of the structure	1	2	3
Name of the structure	FBH Goesdorf	FBN Kaundorf	FBH Nocher-Route
Sub-catchment data			
Name of the municipality	Goesdorf	Kaundorf	Nocher-Route
Name of the catchment	Haute-Sûre	Haute-Sûre	Haute-Sûre
Number of the catchment	1	1	1
Land use [a]	R/I	R/I	R/I
Total area, A_{ges} (ha)	30	22	18.6
Impervious area, A_{imp} (ha)	5	11	4.3
Run-off coefficient for impervious area, C_{imp}	0.28	0.30	0.30
Run-off coefficient for pervious area, C_{per}	0.07	0.10	0.10
Flow time structure, t_{fS} (min)	1	2	2
Population equivalents, *pe* (PE)	611	358	326
Structure data			
Volume, V (m^3)	190	180	157
Curve level–volume, $lev2vol$	Goesdorf	Kaundorf	Nocher-Route
Initial water level, Lev_{ini}	0.57	1.8	1.8
Maximum throttled outflow, $Q_{d,max}$	5	9	4
Orifice diameter, D_d	0.015	0.015	0.015
Orifice coefficient of discharge, C_d	0.67	0.67	0.67

[1]

Table 2.7 presents the runtime results in minutes depending on the number of cores used. The row "speed-up" factor (SF) was calculated as the ratio between the maximum computation time and the current computation time. The maximum computation time was set for the computation with just one core, that is, non-parallel computing. The minimum time is presented in bold font for each test. The results indicated speed-up factors of 12.2 (32 MC simulations), 22.0 (320 MC simulations), and 23.6 (3200 MC simulations). The highest speed-up factor (23.6) was obtained in scenario 3 (3200 MC simulations) using 32 cores. Although the lowest computation time was obtained running scenario 1 (32 MC simulations), the lowest speed-up factor was also reached (12.2).

Table 2.7: Runtime in minutes and "speed-up" factor as a function of number of cores used in simulations.

	32 Simulations		320 Simulations		3200 Simulations	
Cores	**Time**	**SF** [a]	**Time**	**SF**	**Time**	**SF**
1	3.4	1.0	33.9	1.0	334.9	1.0
2	1.9	1.8	18.1	1.9	176.6	1.9
4	0.9	3.7	8.9	3.8	87.8	3.8
8	0.6	5.7	4.8	7.0	46.0	7.3
16	0.4	9.1	2.6	13.0	25.3	13.3
32	**0.3**	12.2	**1.5**	22.0	**14.2**	23.6

[a] Speed-up factor (SF), computed as the ratio between the time for one core and the time for the ith core.

This test was done by setting up the model to simulate three sub-catchments at the same time in parallel mode. Therefore, the scalable code implemented also inferred that parallelisation of sub-catchments also speeds up the overall computation with similar factors.

2.4 Discussion

2.4.1 Conceptual and Mathematical Model

EmiStatR is a simplified model. For instance, it does not take the spatial distribution of inputs into account, in particular, rainfall and impervious areas. Additionally, the simulation of the volume and CSO volume, and henceforth pollutant concentrations such as COD and NH_4, as linear combinations of DWF and RWF is a gross simplification of reality. Finally, the model does not take into account additional processes, such as wash-off, first-flush, and hydrodynamics in the sewer network. From the water quality point of view, Zoppou (2001) concludes that, processes typically described by empirical relationships, such as the build-up and wash-off of pollutants, are not very well understood. The simple exponential relationship that is often used is not reliable, and there are few datasets to validate these relationships or to develop new relationships.

The conceptual model EmiStatR was developed for simple catchment models, such as those used for testing purposes. Neither the advection–diffusion nor other solute transport processes for pollutants in the sewer network were implemented, partly because of the fast response of the urban catchments tested. Therefore, comparisons of the EmiStatR framework to other modelling platforms that include solute transport in the sewer networks may be performed. Only one of the urban storm water models analysed by Zoppou (2001) includes the advective–diffusion equation for the transportation of pollutants in pipes, channels, or storages and explains that this equation is not commonly included

(1) because of the rapid response of an urban catchment, such that the transport of pollutants by diffusion will be negligible compared with the advection of those pollutants, and (2) because urban storm water infrastructure networks are generally more complex than river networks, and for this reason, the numerical solution of the advective–diffusion equation in complex networks can be computationally expensive.

Harremoës (2002) identified important measures in integrated urban drainage modelling: local infiltration, source control, storage basin, local treatment, and real-time control. Thus, the conceptual model implemented in EmiStatR demonstrates the usefulness of local infiltration assessment by considering the variables infiltration flow and water pollution of the infiltration in terms of COD and NH_4. Additionally, the conceptual model implemented serves for testing hypotheses related to the source control by taking into account water consumption and the associated water quality of the wastewater produced in terms of COD and NH_4. Thus, it is possible to take into account the source by means of the evaluation of several profiles for water consumption, as provided by daily profiles, weekly profiles (distinguishing weekend days from weekdays), and seasonal patterns (i.e., monthly patterns accounting for seasonal variability in the source or water consumption). Finally, EmiStatR also takes the storage basin into account by representing the total volume of storage in the catchment as the CSO storage chamber.

EmiStatR is fast and hence very useful for rapid and scalable simulation of long-term scenarios with, for example, yearly precipitation time series as input, as well as for simulating time series with different time-step resolutions from daily to sub-daily time steps. In the case study, we used a time step of 10 min.

2.4.2 Model Implementation

The implementation of EmiStat in R, as a main advantage, made use of graphical user interfaces (GUIs) and plotting functionalities of R. This implementation saved time in the set-up of the model. The implementation in R was also attractive because EmiStatR can easily be extended with R routines, such as ensuring compatibility of input and output time series with geospatial functionalities implemented in R, for example, through the R package spacetime (Pebesma, 2012; Bivand et al., 2013). Moreover, the R environment allows for the implementation of routines for parallel computing and scalable tasks, for example, the packages snowfall (Knaus, 2015) and doParallel. The implementation in R was also advantageous because it facilitated the calibration procedures using the DREAM implementation in R.

2.4.3 Model Calibration and Validation

All simplifications and limitations mentioned above indicate that the model is not perfect and that the model simulations departed from reality. However, despite these simplifications and limitations, the validation results in water quality mode demonstrated a high

accuracy. Further uncertainty propagation studies can shed light on how simplifications affect the model output (Leon et al., 2014). Such uncertainty propagation evaluations are time-consuming and can best be done with fast and scalable calculators such as EmiStatR.

The plots and validation measures presented in Figures 2.4 and 2.5 indicate an accurate representation of the model of the volume in the CSOC at Goesdorf with NSEs of 0.95 for the calibration set, 0.78 for the validation set, and 0.79 when we compared with the CMM. Regarding the simulation of COD and NH_4 loads, Table 2.4 and Figure 2.5 show that the model adequately represented the load in the CSO when we compared it with the well-known commercial CMM simulations. This yielded NSEs of 0.80 for the CSO COD load and 0.82 for the CSO NH_4 load.

After comparison of the simulations of EmiStatR with the CMM, the simulation for volume in the CSO was similar (NSE of 0.78), and therefore the loads of COD and NH_4 were represented accurately. Thus we confirmed the hypothesis that for a small catchment system with urban drainage, it is possible to obtain similar accuracy with a surrogate model (in terms of RMSE and NSE) as a CMM.

It is worth noting which physical processes caused the difference between EmiStatR and the CMM. The main difference was that in EmiStatR, we did not model the pipe routing in the sewer system explicitly. This was considered a lumped process and was represented by the t_{fs} factor, which represents the overall travel time in the sewer system until the flow reaches the CSOC. It works well for small case studies, but for large catchments it remains to be seen.

2.4.4 Scalability

The scalable approach implemented in EmiStatR demonstrated its usefulness and good performance. Computing times decreased substantially, particularly in the scenario with the greatest number of simulations (3200 simulations). The greater the number of simulations, the higher the SF and henceforth the greater the usefulness of the distributed (parallel) computation. This constitutes a promising application of the EmiStatR as a fast calculator in several applications related to urban drainage modelling with increasing levels of complexity and for MC uncertainty analysis.

2.5 Conclusions

We showed using a case study that adequate simulation of CSO spill volume as well as COD and NH_4 loads and concentrations is possible using a scalable, surrogate model. Compared with a CMM, EmiStatR requires less input data, provides automatic calibration procedures, and can present outputs in an accessible way (to practitioners). Another advantage is the large body of R functionalities available to tools such as EmiStatR, for

example, compatibility with input and output data formats for temporal and geospatial data and advanced calibration techniques such as DREAM.

We showed that EmiStatR provides a satisfactory representation of CSO spill volume and COD and NH_4 loads, which confirms that white box simplification can lead to well-performing surrogate models. Moreover, its inherent parallel computation and scalable capabilities allow fast calculations for scenarios of high complexity and for long-term simulations to test hypotheses in urban drainage modelling.

We compared the results of EmiStatR with those obtained using a well-known CMM. The behaviour for volume in the CSOC and the estimation of loads of COD and NH_4 were very similar. Our case study showed that this small catchment (i.e., area of $\leq$30 ha) could be modelled with EmiStatR with satisfactory accuracy compared to models of much higher complexity. Future usage will show how EmiStatR performs in other case studies. Because the basis of EmiStatR is formed by generic equations, it is expected that the performance will be similar.

For future work, it would be of interest to the scientific and practitioner communities to take the spatial distribution of some of the input variables, such as precipitation, impervious areas, and land use, into account. The literature shows that spatial variation in precipitation is not considered in many commonly used models (Zoppou, 2001; Bach et al., 2014). Usually, precipitation is assumed to be uniformly distributed in a sub-catchment. This is not a very realistic assumption, particularly in applications for which the response time is short. The integration of geostatistical probability models that interpolate and simulate precipitation data in space and time would be an important advancement in urban drainage modelling.

It should be emphasised that integrated urban drainage modelling often lacks uncertainty propagation tools that assist in quantifying the spatial and temporal (correlated) distributions (Deletic et al., 2012). It also lack tools for sensitivity analysis to apportion contributions of the different sources of uncertainty to the overall model output uncertainty. Therefore, future work should address these topics and include an economic analysis, also taking the potential failure of CSO infrastructures into account. Such analyses benefit from fast and scalable implementations such as EmiStatR.

Chapter 3

Multivariate autoregressive modelling and conditional simulation of precipitation time series

This chapter is based on:

Torres-Matallana, J. A.; Leopold, U., and Heuvelink, G. B. M. Multivariate autoregressive modelling and conditional simulation of precipitation time series for urban water models. *European Water*, 57:299–306, 2017b. URL `https://library.wur.nl/WebQuery/wurpubs/fulltext/428142`

3.1 Introduction

Precipitation is the most important input of hydrological systems. Precipitation controls hydrological states as soil moisture and groundwater level, and fluxes as runoff, evapotranspiration and groundwater recharge (Guan et al., 2009). Therefore, precipitation plays a paramount role in urban water systems. It controls the fluxes towards storage tanks of combined sewer overflows (CSOs) and the dilution of chemical, organic and biological compounds in the wastewater. Furthermore, small catchments (i.e. areas of about 20 ha or smaller) have a fast response to precipitation input. Catchment average precipitation is a key component in urban water models. However, average catchment precipitation is not always accurately known when measured at rain gauges, because the location of the gauges might be outside of the catchment boundaries or locations do not reflect the entire catchment due to too few rain gauges.

The objective of this chapter is to develop and present a method to estimate the precipitation in a specific catchment given a known precipitation time series in a location outside the catchment, while also quantifying the uncertainty associated with the estimation. After uncertainty characterisation and uncertainty model implementation, simulated precipitation time series can then be used as model inputs to urban water models for Monte Carlo based uncertainty propagation analysis to quantify uncertainty in the model outputs to improve urban water system designs and better asses associated environmental and economic impacts.

3.2 Materials and Methods

This section describes the data, the case study area and the statistical models we developed and implemented for the temporal uncertainty propagation analysis.

3.2.1 Data and case study area

The case study catchment is located around Goesdorf, a small sub-catchment (16.5 ha) of the Haute-Sûre system in the Northwest of Luxembourg. We have time series data for precipitation outside of the Goesdorf sub-catchment at two locations: Dahl (around 2 km from the CSO tank in Goesdorf) and Esch-sur-Sûre (around 3.5 km from the CSO tank Goesdorf). The time series of precipitation provided by the Luxemburgish *Administration des Services Techniques de l'Agriculture*(ASTA) (`http://www.asta.etat.lu/`), covers the year 2010 with measurements at 10 minute resolution. The precipitation stations are provided with Lambrecht 15188 tipping bucket rain gauges with a resolution of 0.1 mm per tip, and a surface of the round reception area of 200 cm^2. We have 52,556 observations for each time series. The total precipitation in 2010 at Esch-sur-Sure is 658.6 mm, and 758.7 mm at Dahl. The time series of precipitation were validated by the Observatory for

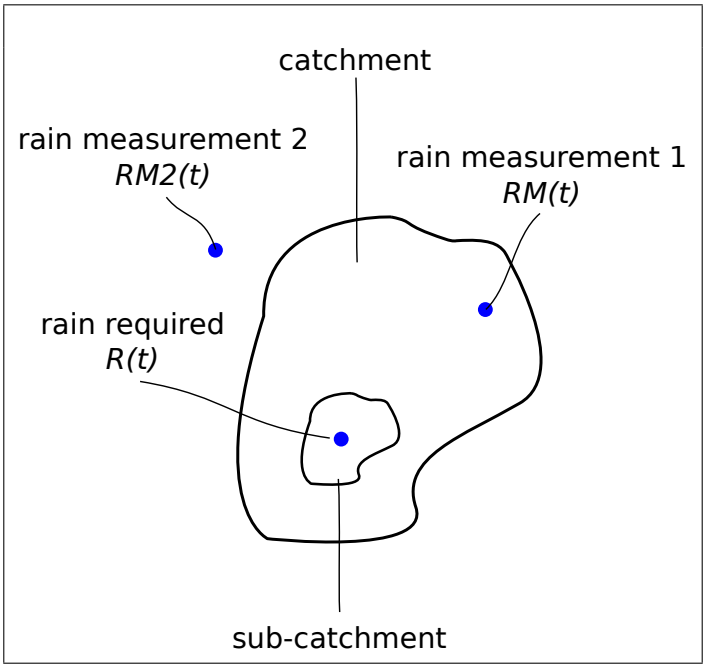


Figure 3.1: Schematic definition of the case study set-up with rain prediction required at R(t) and the 2 rain gauges RM(t) and RM2(t) with t=1,...,T.

Climate and Environment (OCE) of the Luxembourg Institute of Science and Technology (LIST).

3.2.2 Multivariate autoregressive time series modelling of precipitation

The case study provides two measured time series of precipitation, one inside the catchment, at rain gauge 1, $RM(t)$, $t = 1, ..., T$, and one outside the catchment, at rain gauge 2, $RM2(t)$ (see Figure 3.1). However, we need to estimate the precipitation at the sub-catchment level, R(t). Therefore, we model the precipitation in the sub-catchment as:

$$R(t) = RM(t) \cdot \delta(t) \tag{3.1}$$

where $\delta(t)$ is defined as the ratio of $RM(t)$ and $R(t)$, the statistical properties of which are derived using measured time series of $RM(t)$ and $RM2(t)$. We assume that $R(t)$, $RM(t)$ and $\delta(t)$ are log-normally distributed stochastic processes, so that:

$$\log[R(t)] = \log[RM(t)] + \log[\delta(t)] \tag{3.2}$$

which we write as :

$$LR(t) = LRM(t) + L\delta(t) \tag{3.3}$$

$LR(t)$, $LRM(t)$ and $L\delta(t)$ are modelled using a first-order multivariate autoregressive process model. Given Equation 3.3, we only need to model $LRM(t)$ and $L\delta(t)$ because this defines $LR(t)$. We use the following model (Luetkepohl, 2005):

$$\begin{bmatrix} LRM(t+1) \\ \\ L\delta(t+1) \end{bmatrix} = \begin{bmatrix} \mu_R \\ \\ \mu_\delta \end{bmatrix} + \begin{bmatrix} A_{11} & A_{12} \\ \\ A_{21} & A_{22} \end{bmatrix} \cdot \left(\begin{bmatrix} LRM(t) \\ \\ L\delta(t) \end{bmatrix} - \begin{bmatrix} \mu_R \\ \\ \mu_\delta \end{bmatrix} \right) + \begin{bmatrix} \varepsilon_R(t+1) \\ \\ \varepsilon_\delta(t+1) \end{bmatrix} \tag{3.4}$$

where $\mu_R = \mathrm{E}(LRM)$; $\mu_\delta = \mathrm{E}(L\delta)$; A_{11}, A_{12}, A_{21}, A_{22} are the coefficients of an autoregressive (AR) model; ε_R and ε_δ are zero-mean, normally distributed white noise processes. Note that all parameters are assumed time-invariant.

We need to calibrate this model, i.e. estimate the parameters μ_R, μ_δ, A_{11}, A_{12}, A_{21}, A_{22}, σ_R^2, σ_δ^2 and $\rho_{R\delta}$, where $\sigma_R^2 = \mathrm{var}(\varepsilon_R)$, $\sigma_\delta^2 = \mathrm{var}(\varepsilon_\delta)$, and $\rho_{R\delta}$ is the correlation between ε_R and ε_δ. We use the R package mAr (Barbosa, 2015) to calibrate the model given the two time series LRM and $L\delta$. For this, we derive a time series of Lδ by taking the difference of LRM and $LRM2$, in other words we assume that the processes $RM2$ and R have similar multivariate behaviour as RM and R. Note that this is plausible if the rain gauge associated with time series $RM2(t)$ is about the same distance from the first rain gauge as the sub-catchment. We derive a time series of δ by dividing the RM data by the $RM2$ data, for times when both $RM > 0$ and $RM2 > 0$, i.e. we create a bivariate time series of RM and δ. Next, we generate time series $LRM(t)$ and Lδ(t) by taking $log[RM(t)]$ and $log[RM2(t)/RM(t)]$.

3.2.3 Conditional time series simulation of precipitation

Given the calibrated model we need to simulate from $L\delta(t)$. This simulation should be conditional to LRM. We define:

$$X_1(t) = LRM(t) - \mu_R; \qquad \varepsilon_1(t) = \varepsilon_R(t) \tag{3.5}$$

$$X_2(t) = L\delta(t) - \mu_\delta; \qquad \varepsilon_2(t) = \varepsilon_\delta(t) \tag{3.6}$$

and therefore we have:

$$\begin{bmatrix} X_1(t+1) \\ \\ X_2(t+1) \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ \\ A_{21} & A_{22} \end{bmatrix} \cdot \begin{bmatrix} X_1(t) \\ \\ X_2(t) \end{bmatrix} + \begin{bmatrix} \varepsilon_1(t+1) \\ \\ \varepsilon_2(t+1) \end{bmatrix} \tag{3.7}$$

At each time step we need to simulate $X_2(t+1)$ given $X_1(t)$, $X_1(t+1)$ and $X_2(t)$. If we assume $\rho_{R\delta} = 0$, then $X_2(t+1)$ and $X_1(t+1)$ are conditionally independent given $X_1(t)$, so that we can use:

$$X_2(t+1) = A_{21} \cdot X_1(t) + A_{22} \cdot X_2(t) + \varepsilon_2(t+1) \tag{3.8}$$

However, the case $\rho_{R\delta} \neq 0$ is not straightforward, because in that case $X_1(t+1)$ and $X_2(t+1)$ are also "directly" correlated, so that $X_1(t+1)$ has to be included in the conditional distribution of $X_2(t+1)$. We can write:

$$Y = \begin{bmatrix} X_1(t) \\ \\ X_1(t+1) \\ \\ X_2(t) \\ - - - - - \\ X_2(t+1) \end{bmatrix} = \begin{bmatrix} Y_1 \\ \\ Y_2 \end{bmatrix} \tag{3.9}$$

Y follows a multivariate normal distribution with mean vector μ and variance-covariance matrix $\sum$ (Box et al., 2008):

$$Y = \begin{bmatrix} Y_1 \\ \\ Y_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right) \tag{3.10}$$

where μ_1 is a 3x1 vector, μ_2 is a 1x1 vector, Σ_{11} is a 3x3 matrix, Σ_{12} is a 3x1 vector, Σ_{21} is a 1x3 matrix, Σ_{22} is a 1x1 vector. Solving Equation 3.10, we then know:

$$\{Y_2 | Y_1 = a\} \sim N\left(\mu_2 + \Sigma_{21} \cdot \Sigma_{11}^{-1} \cdot (a - \mu_1), \qquad \Sigma_{22} - \Sigma_{21} \cdot \Sigma_{11}^{-1} \cdot \Sigma_{12}\right) \tag{3.11}$$

so we can simulate from $Y_2 = X_2(t+1)$ by sampling from this conditional normal distribution. Therefore, we need to derive vector $\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ and the variance-covariance matrix $\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$. The first is straightforward because we centrered X_1 and X_2 on zero:

$$\begin{bmatrix} \mu_1 \\ \\ \mu_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{3.12}$$

Regarding the calculation of the variance-covariance matrix $\sum$, we assume stationarity of variances and covariances, i.e. these do not depend on time, so the initial effect fades out. Therefore, we can define:

$$\sum = \begin{bmatrix} C_{11} & C_{12} & C_{13} & |C_{14} \\ \\ C_{21} & C_{22} & C_{23} & |C_{24} \\ \\ C_{31} & C_{32} & C_{33} & |C_{34} \\ - & - & - & - \; - \; - \; - \; - \\ C_{41} & C_{42} & C_{43} & |C_{44} \end{bmatrix} \tag{3.13}$$

It is not difficult to show that the components of $\sum$ are given by:

$$C_{11} = C_{22} = \frac{A_{12}^2}{1 - A_{11}^2} \cdot \frac{A_{21}^2 \sigma_1^2 + \sigma_2^2}{A_{11}^2 + A_{11}^2 A_{22}^2 - A_{12}^2 A_{21}^2} + \sigma_1^2 = \mathrm{Var}(X_1) \tag{3.14}$$

$$C_{33} = C_{44} = \frac{A_{21}^2 \sigma_1^2 + \sigma_2^2}{A_{11}^2 + A_{11}^2 A_{22}^2 - A_{12}^2 A_{21}^2} = \mathrm{Var}(X_2) \tag{3.15}$$

$$C_{13} = C_{31} = C_{24} = C_{42} = \frac{A_{11} A_{21} \mathrm{Var}(X_1) + A_{12} A_{22} \mathrm{Var}(X_2) + \rho \sigma_1 \sigma_2}{1 - A_{11} A_{22} - A_{12} A_{21}} = \mathrm{Cov}(X_1, X_2) \tag{3.16}$$

$$C_{12} = C_{21} = A_{11} \mathrm{Var}(X_1) + A_{12} \mathrm{Cov}(X_1, X_2) = \mathrm{Cov}[X_1(t+1), X_1(t)] \tag{3.17}$$

$$C_{34} = C_{43} = A_{21} \mathrm{Cov}(X_1, X_2) + A_{22} \mathrm{Var}(X_2) = \mathrm{Cov}[X_2(t+1), X_2(t)] \tag{3.18}$$

$$C_{23} = C_{32} = A_{11}\text{Cov}(X_1, X_2) + A_{12}\text{Var}(X_2) = \text{Cov}[X_1(t+1), X_2(t)] \tag{3.19}$$

$$C_{14} = C_{41} = A_{21}\text{Var}(X_1) + A_{22}\text{Cov}(X_1, X_2) = \text{Cov}[X_2(t+1), X_1(t)] \tag{3.20}$$

3.2.4 Time series smoothing for $\delta(t)$ computation

Before we could calculate $\delta(t)$ (Equation 3.1) we applied a Daniell kernel (R-Core-Team and contributors worldwide, 2017) to smooth the time series and avoid the sudden tipping bucket effect in the measurements (Table 3.1). The Daniell kernel Only precipitation values above 0.1 mm were smoothed. Then, the time series were filtered and their ratio computed. The length of the resulting filtered time series is 6,454 observations. Finally, the ratio between the time series was computed.

Table 3.1: Daniell kernel for smoothing the observed time series of precipitation. Index = window time index, Factor = Precipitation reduction factor.

Index	Factor
-2	0.1111
-1	0.2222
0	0.3333
1	0.2222
2	0.1111

3.3 Results

This section presents the results of the multivariate autoregressive time series model and the conditional time series simulation for precipitation and compares to simulated and observed values

3.3.1 Multivariate autoregressive time series modelling

Two observed ASTA time series, Esch-sur-Sûre and Dahl (Figure 3.2), were used for the calibration of the multivariate autoregressive model.

We defined the log-transform of the observed filtered time series, $LRM(t)$, and the ratio, $L\delta(t)$, and checked the normality assumption. The log-transform of the time series is fairly normal for the observed, $LRM(t)$, and the ratio, $L\delta(t)$, time series. Also, we checked the autocorrelation function (ACF) of these time series and both follow a similar pattern. Given $LRM(t)$ and $L\delta(t)$ we calibrated the order one multivariate autoregressive model (Equation 5.8) using the mAr R-package. Equations 3.21 and 3.22 present the calibrated model parameters.

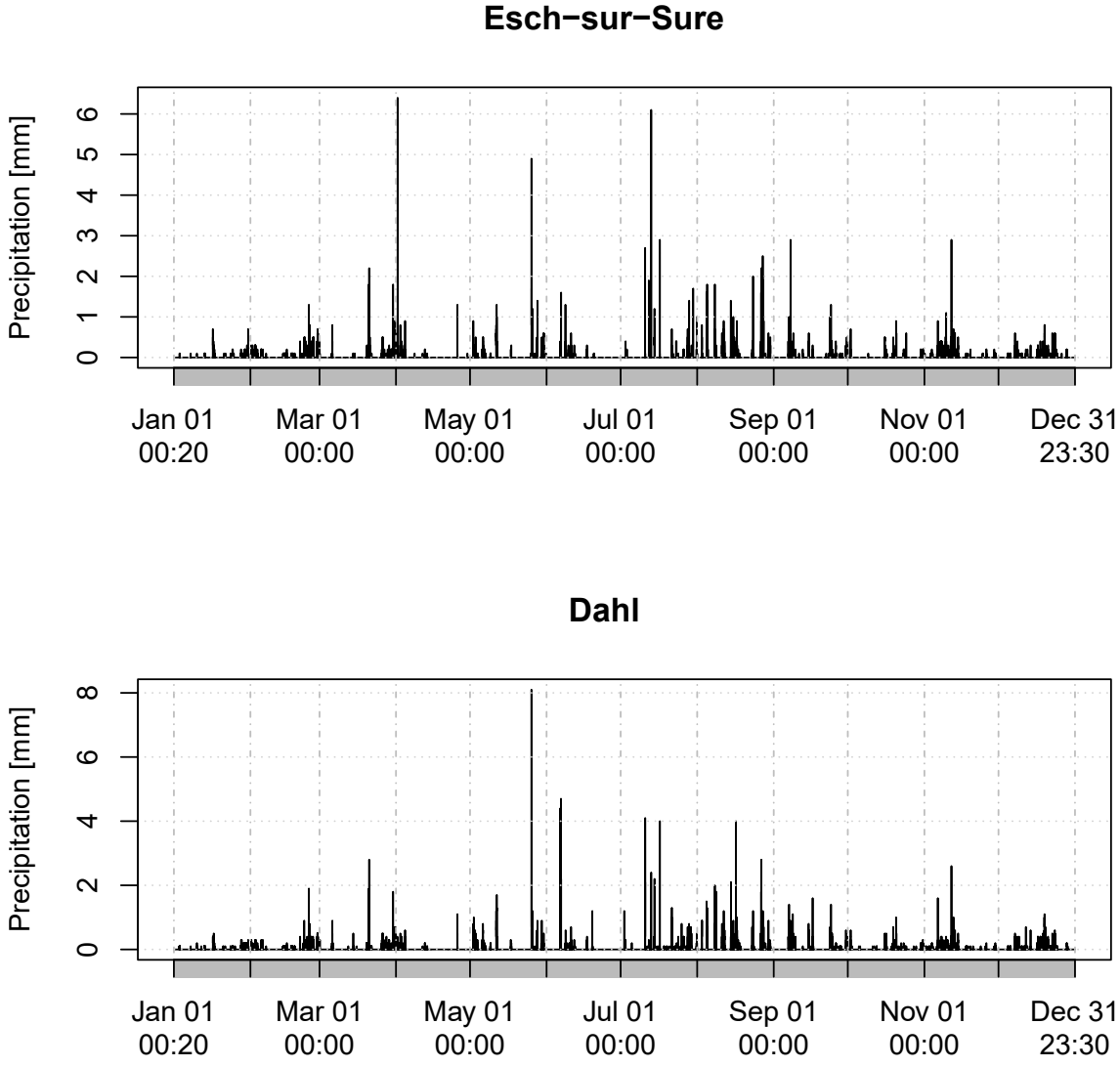


Figure 3.2: Observed precipitation time series of the Esch-sur-Sûre and Dahl rain gauges.

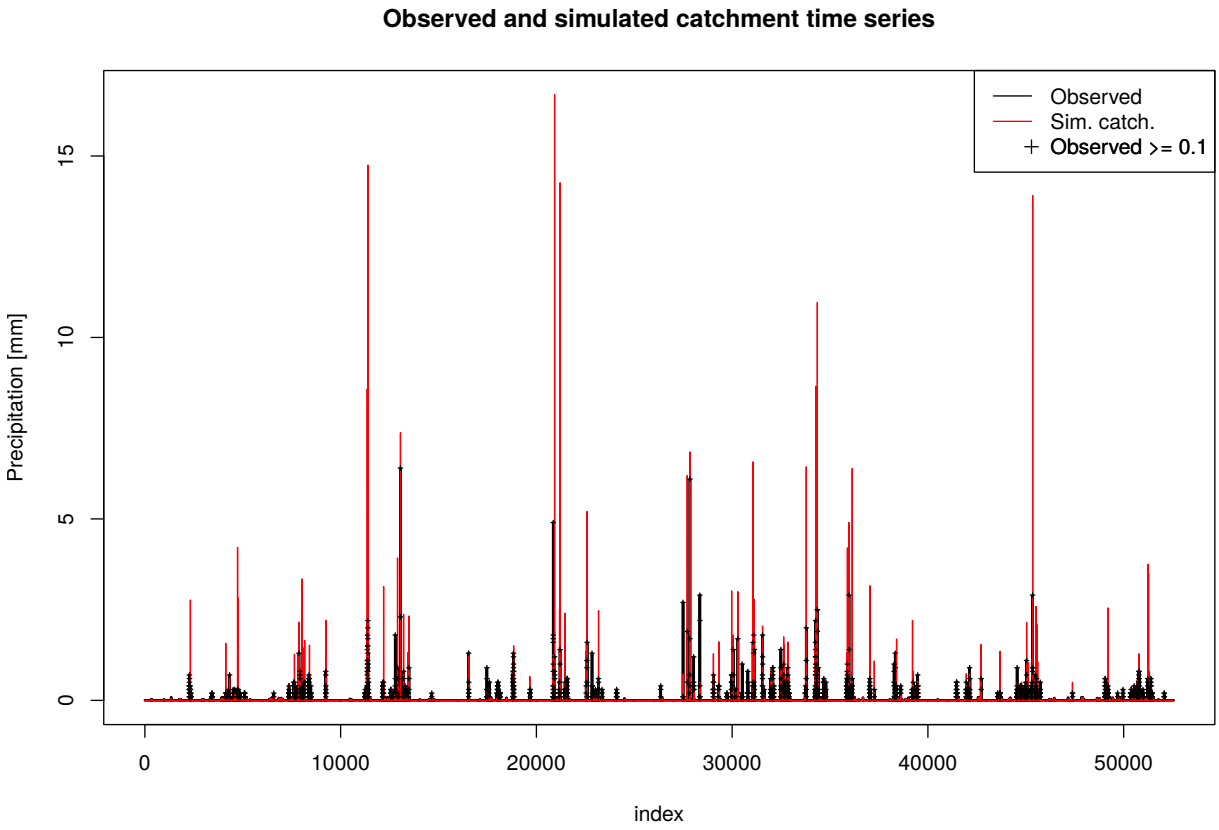


Figure 3.3: Observed and simulated precipitation time series for the Goesdorf sub-catchment.

3.3.2 Conditional time series simulation

Upon calibration of the multivariate autoregressive model, we proceeded with the conditional simulation of $Y_1 = X_2(t+1)$ (Equation 3.11). First, we compute the Σ matrix given the parameters of the model (Equation 3.21). Equation 3.22 presents the calculated values of the components of Σ. Once the components of Σ matrix were calculated, we developed an algorithm for simulating $Y_2 = X_2(t+1)$ given the known values of Y_1 i.e. $X_1(t)$, $X_1(t+1)$ and $X_2(t)$. Then, we added the mean to the time series Y_2, backtransformed the lognormal values to derive the required precipitation time series at the sub-catchment, $R(t)$, according to Equation 3.1. Figure 3.3 shows the observed $RM(t)$ at Esch-sur-Sûre, and the simulated, $R(t)$ at Goesdorf.

$$\begin{array}{l} \mu_R = 2.85501 \\ \mu_\delta = 0.10194 \end{array} \qquad A = \begin{bmatrix} 0.95650 & 0.03980 \\ 0.02429 & 0.88304 \end{bmatrix} \qquad \begin{array}{c} \sigma_R^2 = 0.07241 \\ \sigma_\delta^2 = 0.07951 \\ \rho_{1\delta} = -0.03876 \end{array} \tag{3.21}$$

$$\sum = \left[\begin{array}{cccc} 0.00640 & 0.00617 & -0.00135 & -0.00104 \\ 0.00617 & 0.00640 & -0.00145 & -0.00135 \\ -0.00135 & -0.00145 & 0.00389 & 0.00340 \\ -0.00104 & -0.00135 & 0.00340 & 0.00389 \end{array}\right] \tag{3.22}$$

3.4 Discussion

We proposed a multivariate autoregressive model for conditional simulation of input precipitation based on a multiplicative error model in the lognormal distribution. This method is essentially the same as the application of a Kalman filter/smoother (Kalman, 1960; Webster and Heuvelink, 2006). From a mathematical-statistical point of view we addressed the same principle of Kalman filter, i.e. to compute the conditional probability distribution given the available time series at each time step, sample from it and move to the next time point. Despite the usefulness of the proposed method, some cases show an overestimation of the simulated precipitation, mainly due to high values of the ratio for the multiplicative factor $\delta(t)$. This behaviour is also recognised by McMillan et al. (2011), who stated that the multiplicative factor used in their study "does not capture the distribution tails, especially during heavy rainfall where input errors would have important consequences for runoff prediction". Note also that we ignored the change-of-support effect because the sub-catchment area is much greater than a point. Future research may address this issue of support.

Traditional calibration methods do not take into account input error, which leads to bias in parameter estimation and possible misleading model predictions (McMillan et al., 2011). Several studies have proposed precipitation multipliers for overcoming this issue and taking into account precipitation input uncertainty in hydrological model calibration and prediction (McMillan et al., 2011; Leta et al., 2015; Del Giudice et al., 2016). Based on data from a dense gauge and radar network, McMillan et al. (2011) validated and confirmed the suitability of a multiplicative errormodel. Moreover, they showed that the lognormal multiplier distribution is a good approximation to the true error characteristics.

3.5 Conclusions

Catchment precipitation is a major driving force and key component in urban water models, often computed as average catchment precipitation. However, average catchment precipitation is not always accurately known when measured at rain gauges, because the location of the gauges might be outside the catchment boundaries or do not reflect

the sub-catchment at one location. To overcome this issue, we developed a method to estimate the precipitation in a sub-catchment given a known precipitation time series in a location outside the catchment, while quantifying the uncertainty associated with the estimation. A first-order multivariate autoregressive model for conditional simulation of input precipitation based on a multiplicative error model was proposed. This method is essentially the same as the application of a Kalman filter. Despite the usefulness of the proposed method, some cases show an overestimation of the simulated precipitation due to high values of the ratio for the multiplicative factor. This behaviour is also recognised in the literature.

Such simulated precipitation time series can be used in urban water models for uncertainty propagation analysis to account for uncertainty in improved urban water system design and to better assess environmental and economic impacts.


CAFE DU PONT

Chapter 4

stUPscales: an R-package for spatio-temporal uncertainty propagation across multiple scales (with examples in urban water modelling)

This chapter is based on:

Torres-Matallana, J.A.; Leopold, U., and Heuvelink, G.B.M. stUPscales: an R-package for spatio-temporal Uncertainty Propagation across multiple scales with examples in urban water modelling. *Water*, 10(7)(837):1–30, 2018b. doi: 10.3390/w10070837. URL `https://www.mdpi.com/2073-4441/10/7/837`

Abstract

Integrated environmental modelling requires coupling sub-models at different spatial and temporal scales, thus accounting for change of support procedures (aggregation and disaggregation). We introduce the R-package **s**patio-**t**emporal **U**ncertainty **P**ropagation across multiple **scales**, `stUPscales`, which constitutes a contribution to state-of-the-art open source tools that support uncertainty propagation analysis in temporal and spatio-temporal domains. We illustrate the tool with an uncertainty propagation example in environmental modelling, specifically in the urban water domain. The functionalities of the class `setup` and the methods and functions `MC.setup`, `MC.sim`, `MC.analysis` and `Agg.t` are explained, which are used for setting up, running and analysing Monte Carlo uncertainty propagation simulations, and for spatio-temporal aggregation. We also show how the package can be used to model and predict variables that vary in space and time by using a spatio-temporal variogram model and space-time ordinary kriging. `stUPscales` takes uncertainty characterisation and propagation a step further by including temporal and spatio-temporal auto- and cross-correlation, resulting in more realistic (spatio-)temporal series of environmental variables. Due to its modularity, the package allows the implementation of additional methods and functions for spatio-temporal disaggregation of model inputs and outputs, when linking models across multiple space-time scales.

keywords: temporal aggregation; input uncertainty propagation; spatio-temporal uncertainty characterisation; space-time ordinary kriging

4.1 Introduction

Integrated environmental modelling (IEM) often requires coupling models at multiple spatial and temporal scales. This is challenging due to the complexity of the models and differences in input and output supports between models. Leopold et al. (2006) address the need for aggregation and disaggregation when coupling models. In addition, Bastin et al. (2013) point out that spatio-temporal aggregation and disaggregation are common procedures in modelling chains. This challenge becomes more difficult if uncertainty propagation (UP) is analysed because uncertainty is support-dependent. Change of support procedures (i.e., aggregation and disaggregation) are also required when dealing with integration between models. Accounting for spatio-temporal uncertainty when coupling models in an integrated layout and when addressing change of support is a key component in the data—model—simulation—uncertainty quantification—decision-making chain. There is no software that can do all that is needed for UP analysis in such a complex case. Although important contributions have been made, e.g., the UncertWeb framework (Bastin et al., 2013) and the R-package `spup` (Sawicka et al., 2017), there still is a need for tools for temporal, spatial and spatio-temporal UP accounting for change of support across multiple scales.

In the environmental and geospatial fields, modelling complex processes accounting for uncertainties exhibits specific issues. Bastin et al. (2013) identified six main challenges. We summarise and relate these to the availability of tools accounting for uncertainty analysis in specific applications:

1. Lack of tools accounting for reliable uncertainty information of observational and other data.
2. Lack of tools for reliable information on model structural uncertainty.
3. Development of tools accounting for spatial, temporal and spatio-temporal uncertainty is still a challenge because many environmental variables are measured at diverse spatial and temporal scales and because there are strong correlations between variables.
4. Large computational budgets are required when modelling spatio-temporal distributed systems, due to the large number of variables and the correct characterisation of uncertainties and correlations.
5. Models commonly exhibit complex probability distributions in model output response due to nonlinear responses to model input, even for simple parametric input probability distributions.
6. The default mechanism to propagate uncertainties is the Monte Carlo method, which is highly computationally demanding. Therefore, large computational resources and

implementations of computationally efficient tools are required.

Generally, in environmental and geospatial modelling, there are three main sources of uncertainty (Brown and Heuvelink, 2005; Dotto et al., 2012): (1) Model input uncertainty; (2) Model parameter uncertainty; and (3) Model structural uncertainty. Proper characterisation and quantification of all three uncertainty sources are fundamental in the analysis of the propagation of uncertainties through environmental models. It is also important that practical tools are available to address the characterisation and propagation of uncertainties through models.

Spatial uncertainty propagation has been addressed in several fields, such as in hydrological and water quality modelling (Hengl et al., 2010; Hamel and Guswa, 2015; Muthusamy et al., 2017), in scenario analysis (Rauch et al., 2017), and in soil pollution and nutrient modelling (Leopold et al., 2006; Nol et al., 2010; Vanguelova et al., 2016). However, it is recognised that there is not a universal software tool for performing uncertainty propagation tasks (Sawicka et al., 2017).

Regarding software tools developed for uncertainty propagation analysis, Sawicka et al. (2017) list different tools available for spatial and non-spatial uncertainty propagation and uncertainty assessment. Among the listed tools, there is free software, such as OpenTURNS (Andrianov et al., 2007), DAKOTA (Adams et al., 2009),and DUE (Brown and Heuvelink, 2007), commercial software, such as COSSAN (Schueller and Pradlwarter, 2006), and free software written for licensed software, e.g., SAFE (Pianosia et al., 2015) and UQLab (Marelli and Sudret, 2014) toolboxes for MATLAB (MathWorks, Natick, MA, USA). An extensive review of available UP software tools is presented in Bastin et al. (2013). Regarding existing R packages, there are few that deal with uncertainty propagation explicitly: `propagate` and `errors`. However, neither of these two packages provides functionality for spatial uncertainty analysis.

Widely used techniques for uncertainty assessment in stormwater quality modelling are the classical Bayesian approach based on Markov Chain Monte Carlo (MCMC) sampling and the Metropolis Sampler (Kuczera and Parent, 1998), Generalized Likelihood Uncertainty Estimation (GLUE) (Beven and Binley, 1992), the Multi-algorithm Genetically Adaptive Multi-objective method (Vrugt and Robinson, 2007), and the Shuffled Complex Evolution Metropolis algorithm (Vrugt et al., 2003a). However, these approaches usually do not take all uncertainty sources into account (Rauch and Harremoës, 1999; Wijesiri et al., 2016). Dotto et al. (2012) found that the definition of different subjective criteria, such as user-defined likelihood measures and prior knowledge required in Bayesian techniques, are important issues that limit the application of these techniques. Wijesiri et al. (2016) pointed out that these limitations could have a significant influence on management and planning decisions for mitigation of stormwater pollution. In stormwater quality modelling, a poor characterisation of the source uncertainty (process variability) is the main limitation in accounting for uncertainty (Wijesiri et al., 2016). Therefore, provided

these uncertainty sources can be captured adequately by probability distributions, the use of Monte Carlo (MC) methods for uncertainty propagation is still a suitable option that can overcome most of these limitations.

The literature review above indicates that there are still gaps in research and development of tools for temporal, spatial and spatio-temporal UP accounting for change of support across multiple scales in environmental and geospatial domains. We present the **s**patio-**t**emporal and **U**ncertainty **P**ropagation across multiple **scales** R-package, "`stUPscales`", which provides several R methods and functions for spatio-temporal uncertainty propagation through MC simulation in environmental models, while taking space-time aggregation and disaggregation procedures into account.

We recognise that `spup` (Sawicka et al., 2017) addresses spatial UP as well, both for continuous and categorical variables. Although there are similarities between `spup` and `stUPscales`, there are meaningful differences too. The main difference is that `stUPscales` can also handle UP in the temporal and spatio-temporal domain. In addition, in `stUPscales`, besides sampling from uniform, normal, log-normal, and normal truncated probability distribution functions (pdfs), it is also possible to simulate environmental variables as uni- and multi-variate autoregressive models of order one. Moreover, `stUPscales` has explicit functionality for spatio-temporal change of support. Additionally, another important difference between the two packages is the handling of data structure and objects. In `stUPscales`, the temporal dimension is handled as xts (eXtensible Time Series) objects from the `xts` package (Ryan and Ulrich, 2017), while space-time objects are handled by the classes ST (Spatio Temporal) from the `spacetime` package (Pebesma, 2012), especially the class STFDF (Spatio Temporal Full Data Frame) for regular grids. The `spup` package makes use of the objects RasterStack from the `Raster` package (Hijmans, 2019).

We illustrate the package for temporal aggregation of time series as precipitation and pollutants of Urban Drainage Models (UDMs). In addition, we illustrate the methods and functions for uncertainty propagation in an example of a lumped UDM, EmiStatR, via MC simulation to evaluate water quantity and quality in emissions of Combined Sewer Overflows (CSOs). The example serves for testing the hypothesis that when characterisation of sewage quality indicators as chemical oxygen demand (COD) and ammonium (NH_4) is considered with the appropriate process variability, uncertainty can be quantified as an integral part of the total uncertainty in the modelling procedure.

Regarding the spatio-temporal uncertainty propagation domain, we show how the package can be used to model and predict variables that vary in space and time by using a spatio-temporal variogram model and space-time ordinary kriging. We emphasize that the spatio-temporal functionality developed in `stUPscales` for performing uncertainty propagation studies is an important contribution to the scientific and practitioner communities by making this open source tool accessible and illustrating its applicability in

the water domain. However, it should be noted that the tool may also be applied to other domains in environmental modelling.

The objectives of this study are to: (1) Develop a tool for characterising uncertainty in spatial, temporal and spatio-temporal environmental variables (model inputs) as pdfs and as uni- and multi-variate autoregressive models; (2) Develop a tool for sampling from these pdfs (to support MC UP analysis) and to generate realisations of autoregressive models for environmental variables; (3) Develop functions for aggregation of realisations of environmental variables in space and time; (4) Illustrate the tool with a few simple examples; (6) Develop a method to propagate model input uncertainty through environmental models; (7) Develop a function to analyse results of a MC UP.

4.2 Materials and Methods

This section describes the R-package `stUPscales`, its main class, methods and functions.

4.2.1 The R-Package `stUPscales`

An overview of the conceptual framework is presented in Figure 4.1. Three steps are distinguished:

1. Definition of the uncertainty model (class `setup` and method `MC.setup`). Data in raster or vector format are retrieved and the uncertainty model defined. This characterises the input uncertainty in order to generate realisations from the pdfs or, in the case of time series, from uni- or multi-variate autoregressive models;
2. Uncertainty propagation (method `MC.sim`). A routine for automation of the chain model input—simulation—model output is defined. This routine is model-dependent and the model input format should match the specific format required by the model. The output format should match the format required by `stUpscales`. MC simulation is used for propagating model input uncertainty through the environmental model;
3. Uncertainty analysis (function `MC.analysis`). Summary statistics such as the mean, variance, and quantiles from the ensemble of MC simulations are computed. In addition, several plots summarising the outcome of the analysis are made.

The main class, methods and functions of `stUPscales` (version 1.0.3) are presented below. We describe the class `setup` and the methods and functions `MC.setup`, `MC.sim` and `MC.analysis`.

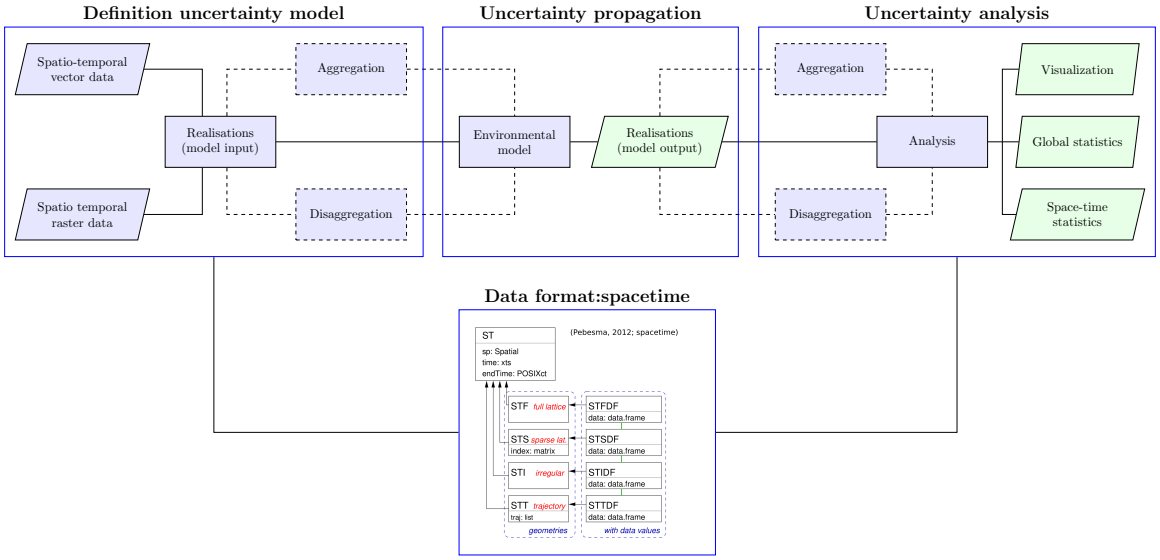


Figure 4.1: Conceptual framework for uncertainty propagation across multiple scales with the `stUPscales` package. Blue rectangular node = process; blue trapezoidal node = input; green trapezoidal node = output; solid line = path; dashed line = optional path.

The `Setup` *Class*

The class `setup` is used to create objects of signature `setup` for further use in the Monte Carlo method. The created object contains eight slots (for more details, see the online user manual available at `https://CRAN.R-project.org/package=stUPscales`):

- id: Object of class "character" to identify the current MC simulation.
- nsim: Object of class "numeric" to specify the number of MC runs.
- seed: Object of class "numeric" to specify the seed of the random number generator.
- mcCores: Object of class "numeric" to specify the number of cores (CPUs) to be used in the MC simulation.
- st.input: Object of class "character" that defines the name of the file or url of the web service to retrieve the spatio-temporal input data.
- rng: Object of class "list" that contains the names and values of the variables to be used in the MC simulation. Five modes are available: (1) constant value; (2) a variable sampled from a uniform (uni) probability distribution function (pdf); (3) a variable sampled from a normal (nor) pdf; (4) a variable sampled from an autoregressive (AR) model and normal (nor) pdf; (5) a variable sampled from a vector autoregressive (VAR) model and normal (nor) pdf; (6) a variable sampled from discrete values (dis); and (7) a variable sampled from a truncated normal (tnor) pdf.

- ar.model: Object of class "list" containing the coefficients of the AR model as vectors, the name of which is the variable to be modelled and the length of which refers to the order of the model as required by function arima.sim from the base package stats.
- var.model: Object of class "list" containing the vector of intercept terms w, the matrix of AR coefficients A, and the noise covariance matrix C of the VAR model, the name of which is the variable to be modelled and the length of which refers to the order of the model as required by function mAr.sim from the package mAr. For mathematical details, see Luetkepohl (2005).

The `MC.setup` *Method*

Given an object of class setup, the method can be invoked for setting-up the MC simulation. The variables are sampled according to their parameters specified in the slot `rng` of the `setup` object. If `ar.model` is defined in slot ar.model, then the specified variables are sampled from the pdf or as an autoregressive (AR) model via the function `arima.sim`. If `var.model` is defined in slot var.model, then the specified variables are sampled from the pdf or as a vector autoregressive (VAR) model via the function `mAr.sim` (see Barbosa (2015); Luetkepohl (2005) for details).

There are seven different cases considered in `stUPscales` (for more details see the online user manual available at `https://CRAN.R-project.org/package=stUPscales`). We developed specific cases accounting for autocorrelation of variables. An extract from the method `MC.setup` for selecting the sampling probability distribution function, case 2 of the method `MC.setup`, i.e., a normally distributed autocorrelated time series (AR1 model), is presented in Appendix B. Additionally, case 5 of the method, a normally distributed auto- and cross-correlated time series (var.model) in parallel code, is also presented in Appendix B.

The `MC.sim` *Method*

The method `MC.sim` is used to invoke a MC simulation. This method can be modified to work with another simulator. The method has two arguments:

- x: Object of class "list" as defined by method MC.setup.
- EmiStatR.cores: a numeric value for specifying the number of cores (CPUs) to be used in the EmiStatR method. Use zero for no use of parallel computation.

The value of the method is a list of length two. The first element of the list, mc, is a list that contains the MC.setup, timing and lap objects. The second element of the list, sim1, is a list that contains the Monte Carlo matrices of the simulator output.

The `MC.analysis` *Function*

The `MC.analysis` function is available for performing the statistical analysis of the MC simulation results. This function requires nine arguments:

- x: A list from the output of the method `MC.sim`.
- delta: A numeric value that specifies the level of temporal aggregation required in minutes.
- qUpper: A character string that defines the upper percentile to plot the prediction band of the results. Several options are possible: "q999" the 99.9th percentile, "q995" the 99.5th percentile, "q99" the 99th percentile, "q95" the 95th percentile, and "q50" the 50th percentile. The lower boundary of the prediction band (shown in grey in the output plots) is the 5th percentile in all cases.
- p1.det: A data.frame that contains the time series of the main driving force of the system to be simulated deterministically, e.g., precipitation. This data.frame should have only two columns: the first containing Time [y-m-d h:m:s], the second containing the (numeric) value of the environmental variable.
- sim.det: A list that contains the results of the deterministic simulation, here the output of EmiStatR given p1.det. See the method EmiStatR from the homonym package for details.
- event.ini: A time-date string in POSIXct format that defines the initial time for the event analysis.
- event.end: A time-date string in POSIXct format that defines the final time for the event analysis.
- ntick: A numeric value to specify the number of ticks in the x-axis for the event time-window plots.
- summ: A list provided as `MC.analysis` output, by default NULL. If provided, the list should contain an output of a previous `MC.analysis` function, so that the analysis is done again without the calculation of some of the internal variables, thus reducing computation time.

The `Agg.t` *Function*

The `Agg.t` function is used for temporal aggregation of environmental variables. `Agg.t` is a wrapper function of `aggregate` from the stats package. It requires five arguments:

- data: A data.frame that contains the time series of the environmental variable to be aggregated, e.g., precipitation. This data.frame should have two columns: the first is Time [y-m-d h:m:s]; the second the value of the environmental variable. If

the environmental variable is different from precipitation, then the column name of the values can be any character string.

- nameData: A character string that defines the name of the environmental variable to be aggregated or an empty string (" ").
- delta: A numeric value that specifies the aggregation level required in minutes.
- func: The name of the aggregation function, e.g., mean, sum.
- namePlot: A character string that defines the title of the plot generated.

4.3 Example 1: Application of `stUPscales` to Temporal Uncertainty Propagation and Aggregation with the EmiStatR Model

We present a UP example using an environmental model with application in urban water modelling, the EmiStatR model, coded as R-package. Its `input` class and main method `EmiStatR` are presented as well.

The example is an application in temporal UP for simulation of the water volume and water quality dynamics in combined sewer systems (CSSs). CSSs are designed to convey wastewater during precipitation events and also directly to the wastewater treatment facility in dry weather flow conditions. In CSSs, two types of discharge are distinguished: (1) the pass forward flow, which transports the sewage discharge directly to the wastewater treatment facility (Hager, 2010); (2) the combined sewer overflow (CSO), which is the sewage diverted from the treatment facility and discharged, untreated, into a local receiving water body during heavy precipitation events (Baker, 2009). CSOs, when released to the environment, can have an important damaging impact on the water quality status of receiving waters (streams, rivers, ponds, lakes, wetlands, and oceans). About 50% of pollutants in stormwater are metals, nutrients, organic toxins, and bacteria, and are associated with particulate matter. The other 50% are soluble (Baker, 2009), and therefore these pollutants can be more persistent in the water and the environment itself. Minimisation of the CSO spill volume is therefore important for the preservation of good water quality status of receiving waters. The goal of this example is to quantify model output uncertainty when model input uncertainties are propagated through a simplified lumped urban drainage model.

4.3.1 The Model EmiStatR

The R-package EmiStatR version 1.2.0.4 is used to perform the simulations and to propagate model input uncertainty in the example with the R-package `stUPscales`. Details regarding the EmiStatR model are found in Torres-Matallana et al. (2018a).

The main components of the EmiStatR model are: (1) Dry Weather Flow (DWF) including Infiltration Flow (IF); (2) Pollution of DWF; (3) Rain Weather Flow (RWF); (4) Pollution of RWF; (5) Combined Sewer Flow (CSF) and pollution; and (6) Combined Sewer Overflow (CSO) and pollution.

Basically, the total dry weather flow, Q_t $[\mathrm{l} \cdot \mathrm{s}^{-1}]$ is calculated as:

$$Q_t = Q_s + Q_f, \tag{4.1}$$

where Q_s $[\mathrm{l} \cdot \mathrm{s}^{-1}]$ is the dry weather flow estimated as the flow of the residential wastewater in the catchment, calculated as $86,400^{-1} \cdot pe \cdot qs$ (where $86,400 = 24 \times 60 \times 60$ is a measurement unit conversion factor), with pe [PE] the population equivalents of the connected CSO structure, and qs $[\mathrm{l} \cdot \mathrm{PE}^{-1} \cdot \mathrm{d}^{-1}]$ the individual water consumption of households. Q_f $[\mathrm{l} \cdot \mathrm{s}^{-1}]$ is the infiltration flow that enters the pipes from groundwater flow through cracks and joints, calculated as $A_{imp} \cdot q_f$, where A_{imp} [ha] is the impervious area of the catchment, and q_f $[\mathrm{l} \cdot \mathrm{s}^{-1} \cdot \mathrm{ha}^{-1}]$ is the infiltration water inflow flux (specific infiltration discharge from groundwater flow). Variables qs and pe are dynamic and can be defined as time series with daily, weekly and seasonal patterns.

The contribution of rain water to the combined sewage volume is given by V_r $[\mathrm{m}^3]$. This is a vector whose length is equal to that of the input precipitation time series. V_r is computed as:

$$itV_r = 10 \cdot P_1 \cdot [C_{imp} \cdot A_{imp} + C_{per}(A_{total} - A_{imp})], \tag{4.2}$$

where P_1 is a time series of rainfall depth [mm]; A_{imp} is the impervious area of the catchment [ha]; A_{total} is the total area of the catchment [ha]; C_{imp} is the run-off coefficient for impervious areas [-]; and C_{per} is the run-off coefficient for pervious areas [-].

The load, $B_{x,Sv}$ [kg], of a specific water quality variable x (e.g., Chemical Oxygen Demand, COD, or Ammonium, NH_4) in the overflow is calculated as a function of the CSO spill volume, V_{Sv} $[\mathrm{m}^3]$, a combined sewer mixing ratio, cs_{mr} [-], the mean dry weather pollutant concentration, $C_{x,DWF}$ $[\mathrm{mg} \cdot \mathrm{l}^{-1}]$, and the concentration due to rainwater pollution, $C_{x,Rw}$ $[\mathrm{mg} \cdot \mathrm{l}^{-1}]$.

$$B_{x,Sv} = 10^{-3} \left(cs_{mr} + 1\right)^{-1} V_{Sv} \left(cs_{mr} \cdot C_{x,DWF} + C_{x,Rw}\right) \tag{4.3}$$

where DWF refers to dry weather flow and Rw to rainwater. V_{Sv}, cs_{mr} and $C_{x,DWF}$ are vectors of length equal to the input precipitation time series, P_1. $C_{x,Rw}$ can be a vector of length equal to P_1 or a unique value constant in time. Two indicators of water pollution are simulated: COD and NH_4. Table 4.1 presents the main components and input variables of the EmiStatR model. The main model outputs are listed in Table 4.2.

An object of class `input` can be defined to contain all required input data for running the model in deterministic mode. An object of class `input` can be created by invoking the command `input()` or `new(''input'')`. The content details of an object of class

`input` can be found in the user manual of the EmiStatR package available in CRAN (`https://cran.r-project.org/web/packages/EmiStatR/`). This object has 20 slots. For this example, the model input variables considered in the simulations and used to compute the output uncertainty are presented in Table 4.1.

If calibration data are available, EmiStatR parameters may be calibrated prior to simulation. If calibration is not feasible, the model can be run using parameter values taken from the reference literature and guidelines. Table 4.3 provides reference values and calibration ranges for the most important EmiStatR parameters.

Table 4.1: General input data and combined sewer overflow (CSO) data to be defined by the `input` object.

General Input	Units	CSO Input	Units
Wastewater		*Catchment data*	
Water consumption, q_s	$[\mathrm{l} \cdot \mathrm{PE}^{-1} \cdot \mathrm{d}^{-1}]$ [a]	Total area, A_{total}	[ha]
Pollution COD [b], $C_{COD,s}$	$[\mathrm{g} \cdot \mathrm{PE}^{-1} \cdot \mathrm{d}^{-1}]$	Impervious area, A_{imp}	[ha]
Pollution NH_4[c], $C_{NH4,s}$	$[\mathrm{g} \cdot \mathrm{PE}^{-1} \cdot \mathrm{d}^{-1}]$	Run-off coefficient for impervious area, C_{imp}	[-]
		Run-off coefficient for pervious area, C_{per}	[-]
Infiltration water		Theoretical largest flow time structure, t_{fs}	[time step]
Inflow, q_f	$[\mathrm{l} \cdot \mathrm{s}^{-1} \cdot \mathrm{ha}^{-1}]$	Population equivalents, pe	[PE]
Pollution COD, COD_f	$[\mathrm{g} \cdot \mathrm{PE}^{-1} \cdot \mathrm{d}^{-1}]$		
Pollution NH_4, $NH4_f$	$[\mathrm{g} \cdot \mathrm{PE}^{-1} \cdot \mathrm{d}^{-1}]$	*CSO structure data*	
		Volume, V	$[\mathrm{m}^3]$
Rainwater		Curve level—volume, $lev2vol$	[m], $[\mathrm{m}^3]$
Precipitation time series, P	[mm]	Initial water level, Lev_{ini}	[m]
Pollution COD, COD_r	$[\mathrm{mg} \cdot \mathrm{l}^{-1}]$	Maximum throttled outflow, $Q_{d,max}$	$[\mathrm{l} \cdot \mathrm{s}^{-1}]$
Pollution NH_4, $NH4_r$	$[\mathrm{mg} \cdot \mathrm{l}^{-1}]$	Orifice diameter, D_d	[m]
		Orifice coefficient of discharge, C_d	[-]

Table 4.2: Output variables of the EmiStatR method.

	Variable	Unit
CSO[a] summary	Period, p	[day]
	Total CSO chamber volume, V_{CSOC}	[m^3]
	Duration CSO spill volume, d_{Sv}	[h]
	Frequency CSO spill volume, f_{Sv}	[events]
	Total CSO spill volume, V_{Sv}	[m^3]
	Average CSO flow, Q_{Sv}	[l/s]
	95th percentile CSO spill volume, $V_{Sv,95}$	[m^3]
	Maximum CSO spill volume, $V_{Sv,max}$	[m^3]
COD[b] in spill volume	Total CCOD[c], $C_{COD,Sv}$	[mg/l]
	Average CCOD, $C_{COD,Sv,av}$	[mg/l]
	95th percentile CCOD, $C_{COD,Sv,95}$	[mg/l]
	Maximum CCOD, $C_{COD,Sv,max}$	[mg/l]
	Total BCOD[d], $B_{COD,Sv}$	[kg]
	Average BCOD, $B_{COD,Sv,av}$	[kg]
	95th percentile BCOD, $B_{COD,Sv,95}$	[kg]
	Maximum BCOD, $B_{COD,Sv,max}$	[kg]
NH_4[e] in spill volume	Total CNH4, $C_{NH4,Sv}$	[mg/l]
	Average CNH4[f], $C_{NH4,Sv,av}$	[mg/l]
	95th percentile CNH4, $C_{NH4,Sv,95}$	[mg/l]
	Maximum CNH4, $C_{NH4,Sv,max}$	[mg/l]
	Total BNH4[g], $B_{NH4,Sv}$	[kg]
	Average BNH4, $B_{NH4,Sv,av}$	[kg]
	95th percentile BNH4, $B_{NH4,Sv,95}$	[kg]
	Maximum BNH4, $B_{NH4,Sv,95}$	[kg]

[a] CSO = combined sewer overflow; [b] COD = chemical oxygen demand; [c]CCOD = COD concentration; [d]BCOD = COD load; [e] NH_4 = ammonium; [f]CNH4 = NH_4 concentration; [g]BNH4 = NH_4 load.

The output value of the `EmiStatR` method is described in the manual. Here, we are interested in propagating model input uncertainties through the simplified lumped urban drainage model.

Table 4.3: Default values for input data of EmiStatR.

Input	Units	Reference Value	Literature Source	Range (This Study)
Wastewater				
Water consumption, q_s	$[\mathrm{l \cdot PE^{-1} \cdot d^{-1}}]$[a]	150[b]	Fan et al. (2013)	[130, 170]
Pollution COD[c], $C_{COD,s}$	$[\mathrm{g \cdot PE^{-1} \cdot d^{-1}}]$	120	DWA (2002)	[90, 150]
Pollution TKN[d]	$[\mathrm{g \cdot PE^{-1} \cdot d^{-1}}]$	11	DWA (2002)	[7, 15]
Pollution NH_4[e]	$[\mathrm{g \cdot PE^{-1} \cdot d^{-1}}]$	4.7	Torres-Matallana et al. (2018a)	[1, 8]
Infiltration water				
Inflow, q_f	$[\mathrm{l \cdot s^{-1} \cdot ha^{-1}}]$	0.05	DWA (2006)	[0.001, 0.1]
Catchment data				
Run-off coefficient for impervious area, C_{imp}	[-]	See Rawls et al. (1981)	Rawls et al. (1981)	[0.20, 0.95]
Run-off coefficient for pervious area, C_{per}	[-]	See Rawls et al. (1981)	Rawls et al. (1981)	[0.05, 0.50]
Flow time structure, t_{fs}	[time step]	2	Torres-Matallana et al. (2018a)	[0, 12]
CSO structure data				
Initial water level, Lev_{ini}	[m]	L_{max}[f]/2	Torres-Matallana et al. (2018a)	[0, L_{max}]
Orifice coefficient of discharge, C_d	[-]	1.25	Torres-Matallana et al. (2018a)	[0.01, 2]

4.3.2 Temporal Aggregation and Model Input Uncertainty Propagation

We used the `Agg.t` function, described in Section 4.2.1, to perform a temporal aggregation of the model input data to the time step required. The model input uncertainty propagation is done in a two-step process. First, a model input uncertainty definition is done by setting up the MC simulation. For this, we apply the `setup` class and the `MC.setup` method. Second, the MC simulation is performed by applying the `MC.sim` method, after which the UP analysis is done through the application of the `MC.analysis` function.

4.3.3 Results

In this section, we first present the input time series as the main driving force of the system modelled. This input is important because the other input time series definition depends on it to achieve the same model input support. Second, the results of input time series aggregation are shown. Finally, the results of the model input uncertainty propagation with `stUPscales` are presented.

Model Input

EmiStatR includes an example of a precipitation dataset called `P1`, used as main input data. This dataset is a list that contains a data.frame with two columns: "Time [y-m-d h:m:s]" and precipitation depth in millimetres "P [mm]". The rain gauge station where the measurements were recorded (in this case Dahl) is located close to the catchment of the combined sewer overflow chamber at Goesdorf, Grand-Duchy of Luxembourg. The dataset contains records from 1 January until 2 February 2016. The recording time step is 10 min. We used this dataset as precipitation input.

Temporal Aggregation and Model Input Uncertainty Propagation

As an illustration of the `Agg.t` function, we aggregated the time series of precipitation `P1` from 10 to 30 min. We took the sum as aggregation function. Following the two-step process, we performed the model input uncertainty propagation. First, an illustration of the model input uncertainty definition is presented by setting up the MC simulation. Second, the MC simulation is executed followed by its respective analysis.

Model Input Uncertainty and Monte Carlo Simulation Set-Up

For setting up the MC simulation, we applied the `setup` class and the `MC.setup` method. In the Appendix C are presented the settings of the `setup` class. Upon the definition of the `setup` object, we proceeded to invoke the `MC.setup` to create the sampled variables from their corresponding probability distributions.

The water quantity input variables chosen for the uncertainty propagation analysis are water consumption (qs), infiltration inflow water (qf), run-off coefficient for impervious area (C_{imp}), run-off coefficient for pervious area (C_{per}), population equivalents (pe), orifice coefficient of discharge (C_d) and the initial water level in the chamber (Lev_{ini}).

The precipitation input time series and the water quality input variables chosen for the uncertainty propagation analysis are plotted in Figure 4.2. A time window was chosen for illustration purposes. The sewage COD pollution per capita [PE] load per day ($C_{COD,s}$) was modelled as an autoregressive order one (AR1) model with autoregressive coefficient equal to 0.7 (Figure 4.2b). The sewage NH_4 pollution per capita [PE] load per day ($C_{NH4,s}$) was also modelled as an AR1 model with autoregressive coefficient equal to 0.7 (Figure 4.2c).

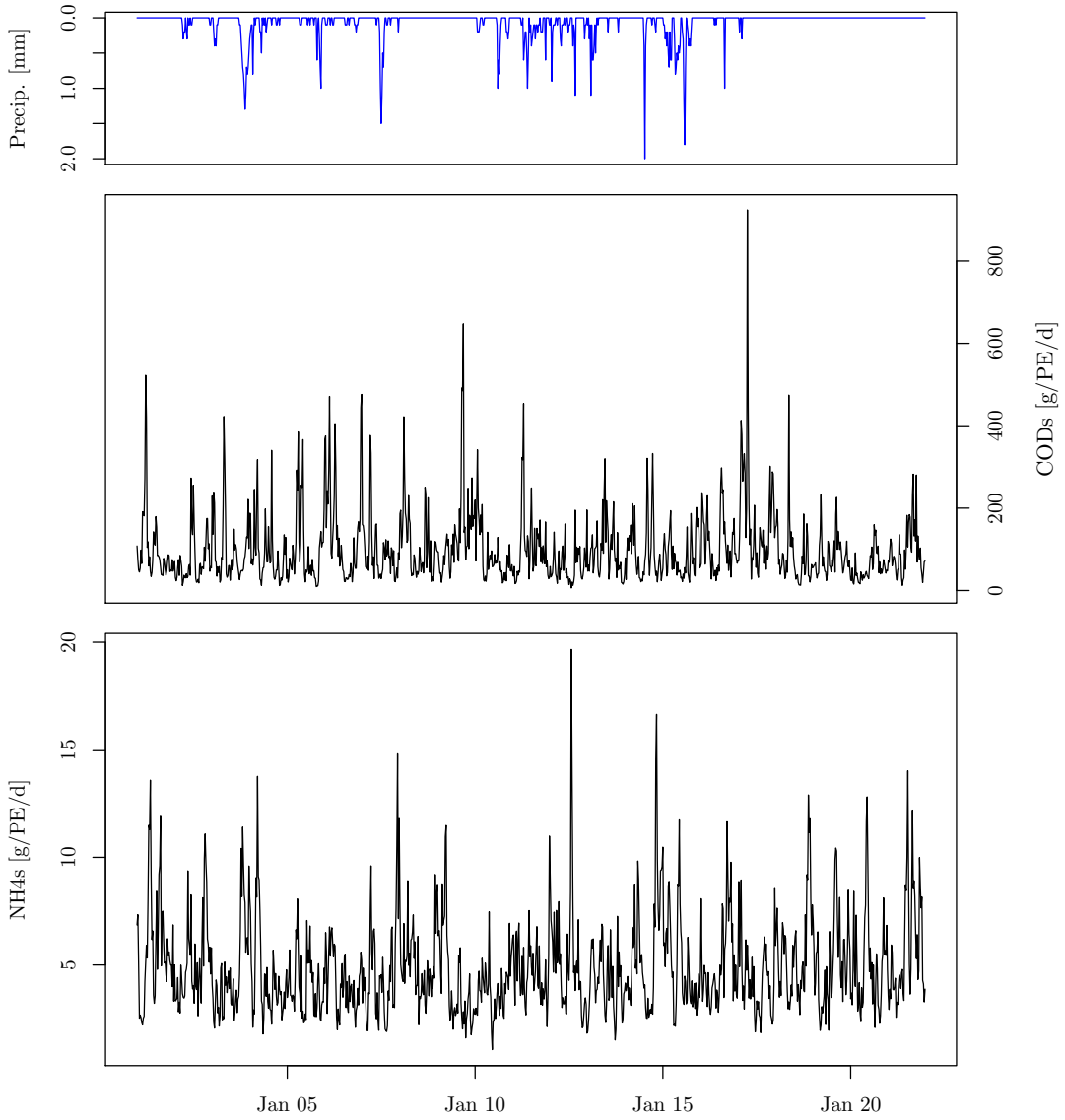


Figure 4.2: Series of precipitation, sewage COD pollution and sewage NH_4 pollution for the time window.

Monte Carlo Simulation and Analysis

The MC simulation was performed by invoking the `MC.sim` method. The analysis of the MC simulations was done by invoking the `MC.analysis` function. In order to proceed with the MC analysis, a deterministic simulation was defined and computed. First, we defined structure 1, named "E1". Second, the `input.user` object of class `input` was defined. Next, we invoked the method `EmiStatR` with the deterministic input (`input.user`) defined before. Finally, the additional arguments of the `MC.analysis` function were defined. Upon definition of the additional arguments, the function `MC.analysis` was invoked. In Appendix C is presented the `MC.sim` method and the `MC.analysis` function.

Only part of the outputs of the analysis are presented here. Figure 4.3 presents the time series of the deterministic simulation, as well as the mean and 95% prediction band of the MC simulations at a 30-min time step, for simulation and analysis of the water volume. Figure 4.4 shows the time series window for COD load and concentration in spill water volume in the CSO and NH_4 load concentration in spill water volume in the CSO, at a 30-min time step.

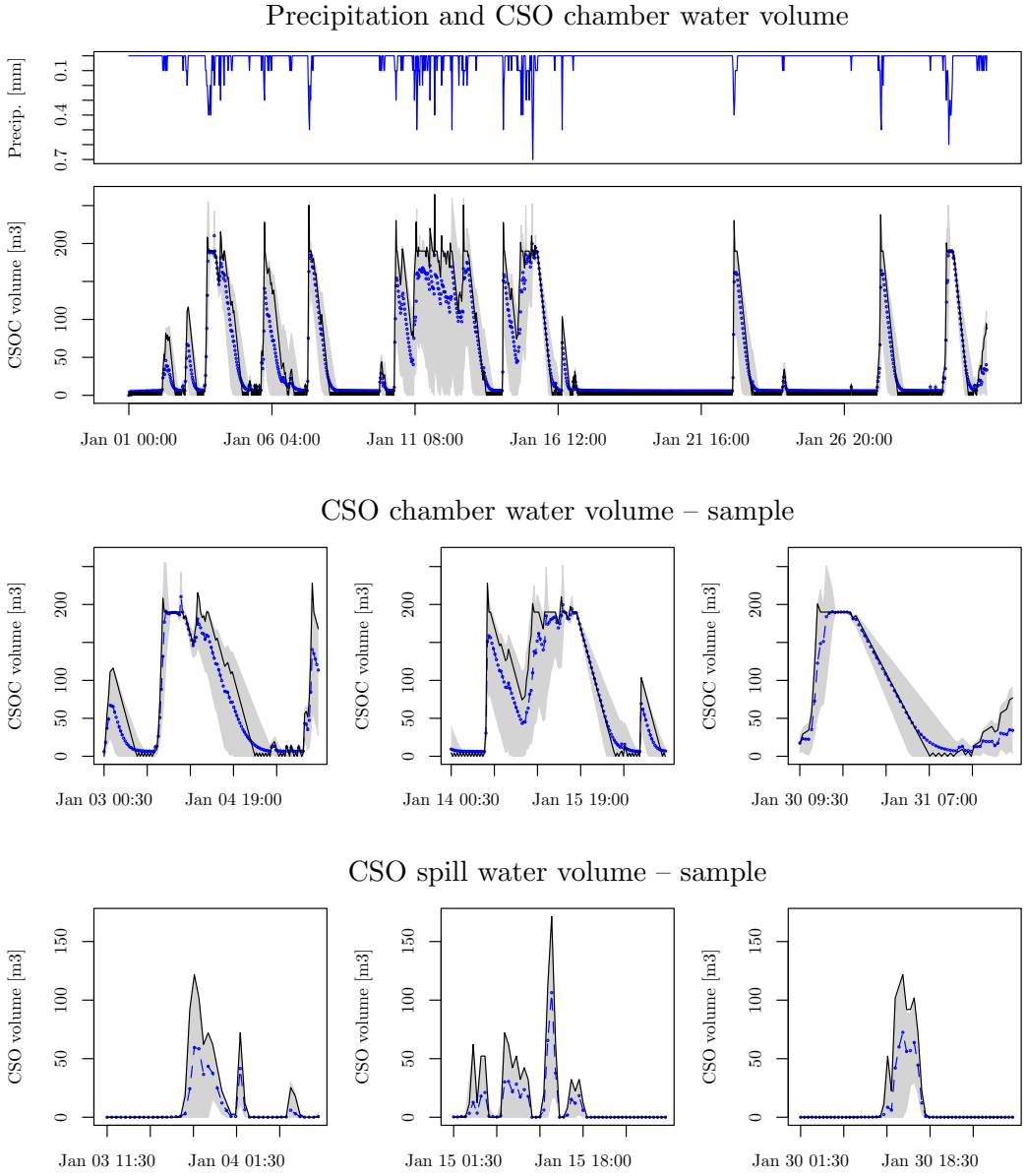


Figure 4.3: January 2016 time series of deterministic simulation (continuous black line), mean (dotted blue line), and 95% prediction band (grey band) of the Monte Carlo simulation at 30-min time step for simulation and analysis of water volume. Time series window for precipitation (top), water volume in the combined sewer overflow (CSO) chamber (second), detailed snapshots for water volume in the CSO chamber (third), and spill water volume of CSO (bottom).

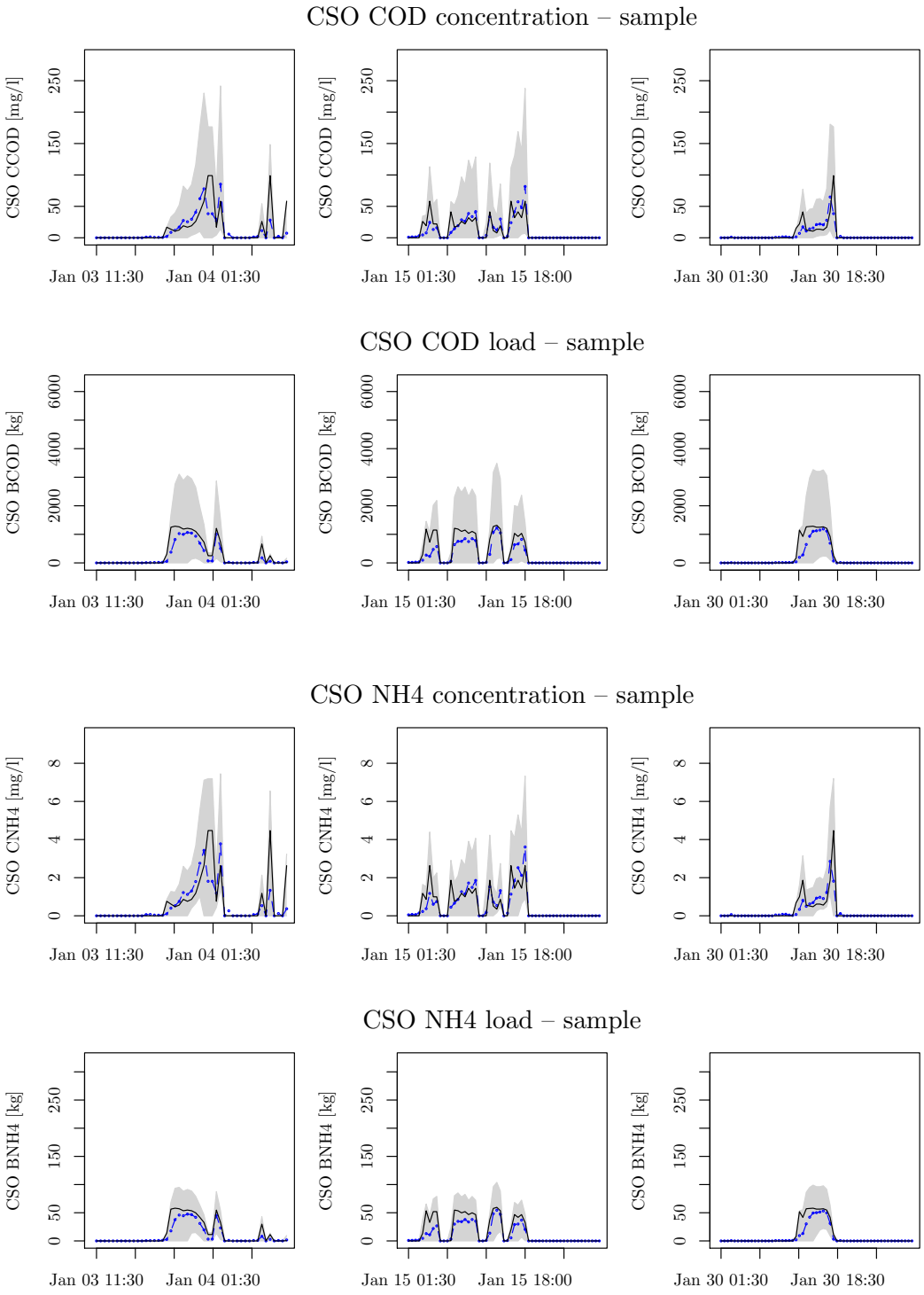


Figure 4.4: Time series for chemical oxygen demand (COD) load and concentration in spill water volume in the combined sewer overflow (CSO) and ammonium (NH_4) load concentration in spill water volume in the CSO at 30-min time step for simulation and analysis of Monte Carlo simulation. Deterministic simulation (continuous black line), mean (dotted blue line), and 95% prediction band (grey band) of the Monte Carlo simulation.

4.4 Example 2: Application of stUPscales to Spatio-Temporal Uncertainty Characterisation of Precipitation for the Grand-Duchy of Luxembourg

This example illustrates the capability of `stUPscales` to address UP in the spatio-temporal domain. We illustrate how to characterise the spatio-temporal uncertainty of precipitation for the entire country of the Grand-Duchy of Luxembourg. We use observations of a precipitation time series measured at 25 rain gauges for the period January 2010 to December 2011, recorded at 10-min time steps. The set of time series was provided by the Luxemburgish *Administration des Services Techniques de l'Agriculture* (ASTA), and verified for consistency by the Observatory for Climate and Environment (OCE) of the Luxembourg Institute of Science and Technology (LIST).

We prepared the data to predict space-time precipitation fields at 500 m and 10-min time step. This space-time uncertainty information can, for instance, be used to feed a distributed rainfall–runoff model coupled with an urban drainage model. We only demonstrate the precipitation uncertainty characterisation, since an entire UP study of a rainfall–runoff model is out of the scope of this chapter and will be addressed in Chapter 5.

4.4.1 Selection of Event

We selected a 10-h period where the cumulative precipitation of the time series is maximal, assuring to retrieve a precipitation event in all stations. The selected period of the event was 16 December 2011 from 12:00 a.m. to 10:00 a.m. A total of 32 stations were available. Seven stations were not taken into account because of no measurement in the selected period. For illustration purpose, Figure 4.5 presents eight out of the 25 selected time series, and Figure 4.6 illustrates 18 snapshots of the event every 30 min and shows how this event is distributed in space. It also shows the precipitation magnitude in the 25 rain gauges.

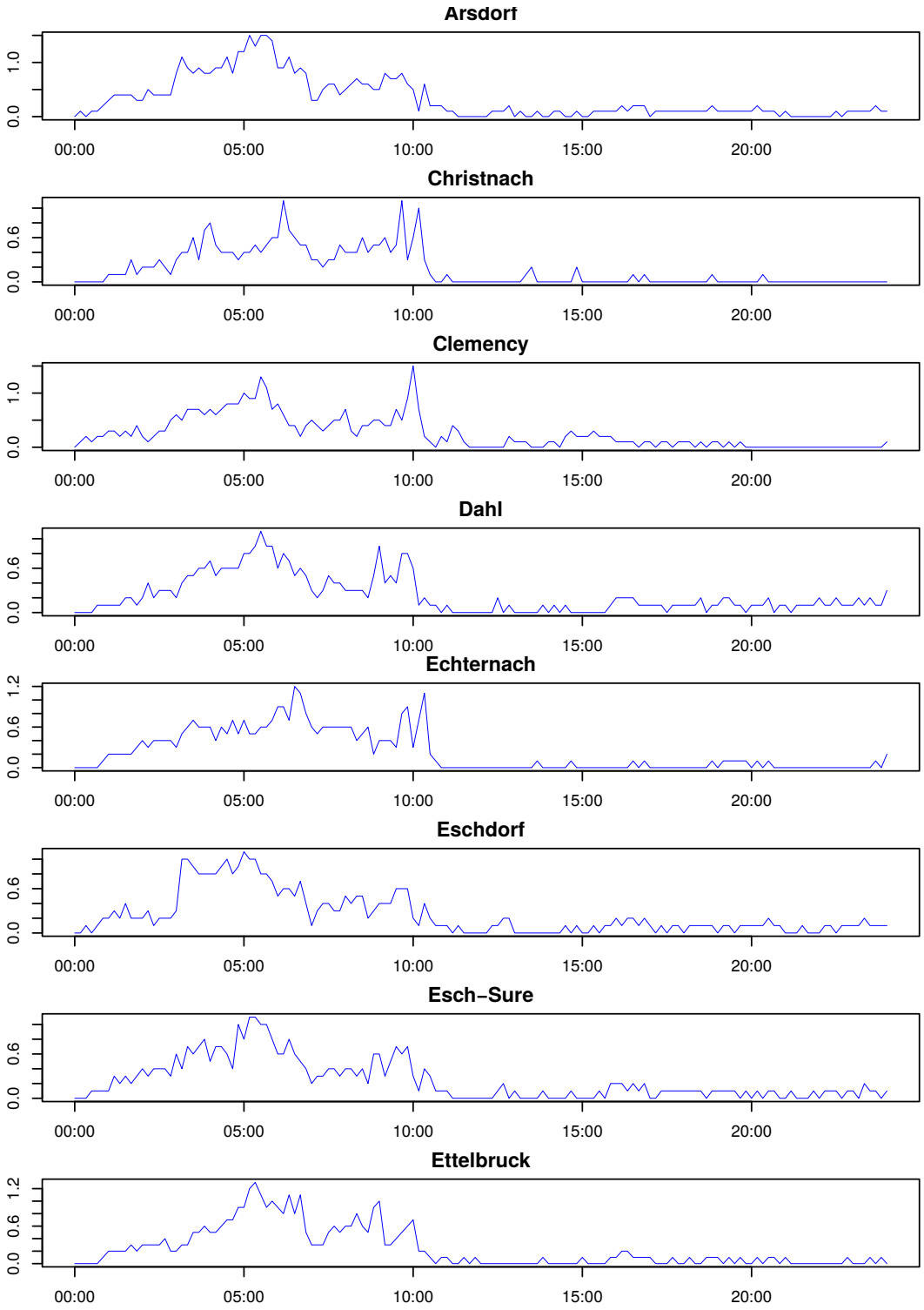


Figure 4.5: Observed precipitation (mm) time series for event 16 December 2011, eight out of 25 stations. The experimental variogram was calculated in the interval from 12:00 a.m. to 10:00 a.m. including the 25 stations.

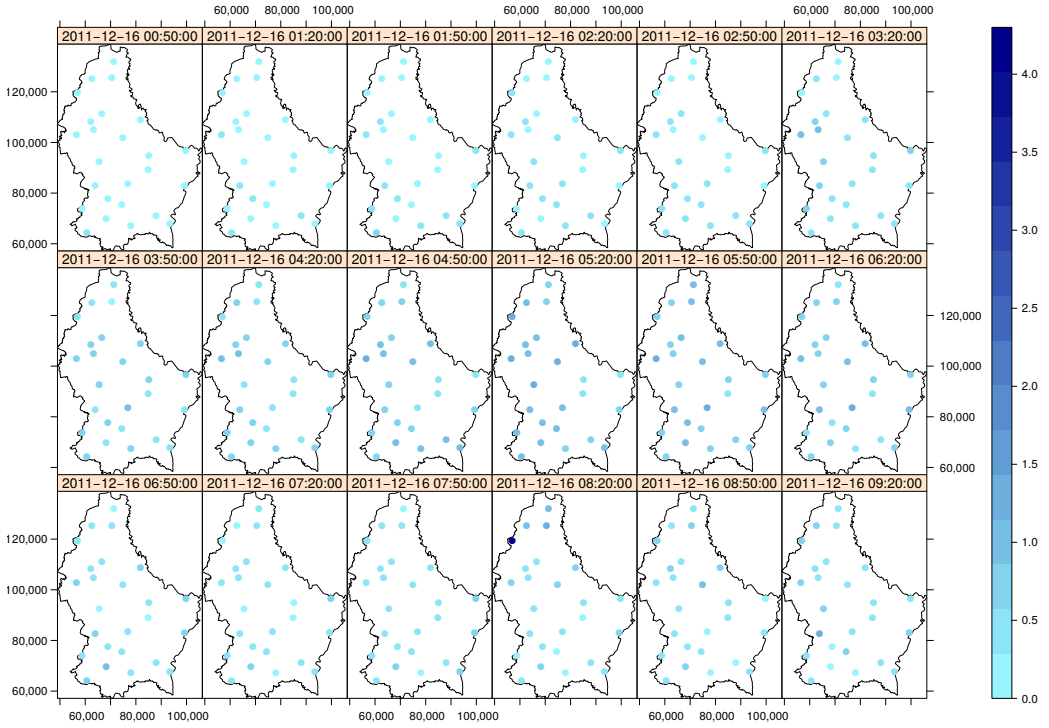


Figure 4.6: Distribution of observations for event 16 December 2011 over the country of the Grand-Duchy of Luxembourg, 18 snapshots for illustration purpose only. Every dot represents a rain gauge observation. The units of the x and y scales are meters. The scale bar represents precipitation in millimeters.

4.4.2 Ordinary Kriging in the Space-Time Domain

To model precipitation fields in the spatio-temporal domain, we used the concept of the spatio-temporal variogram for spatio-temporal ordinary Kriging (Gräler et al., 2016). Following Snepvangers et al. (2003), the aim of space-time geostatistical modelling is to predict or simulate an attribute z given by:

$$z = \{z(s,t)|s \in S, t \in T\}, \tag{4.4}$$

where $s \in \mathbf{R}^2$ refers to space, $t \in \mathbf{T}$ to time. The prediction or simulation of z is made at a space–time point (s_0, t_0), where z was not measured.

A space-time random function model Z can be defined to predict $z(s_0, t_0)$ (Snepvangers et al., 2003):

$$Z(s,t) = m(s,t) + \epsilon(s,t) \qquad s \in \mathbf{R}^2, t \in \mathbf{T}, \tag{4.5}$$

where m is the trend component of the model and ϵ is a zero-mean stochastic residual component. The prediction is based at measurements in n space-time points (s_i, t_i). In space-time ordinary kriging (ST-OK), the trend component is assumed to be an unknown constant mean. The stochastic residual component is characterised by a space-time

variogram, that, under second-order stationarity, can be computed from the measurements (Snepvangers et al., 2003):

$$\hat{\gamma}(h_S, h_T) = \frac{1}{2N(h_S, h_T)} \sum_{i=1}^{N(h_S,h_T)} [z(s_i, t_i) - z(s_i + h_S, t_i + h_T)]^2, \tag{4.6}$$

where h_S and h_T are the space and time lags, and $N(h_S, h_T)$ is the number of pairs in the space-time lag.

To fit a model to the space-time experimental variogram (Equation 4.6), we need to overcome some additional problems as we have space and time variation. In this example, we used a model that considers the definition of one covariance model (or variogram) for the space domain and one covariance model (or variogram) for the time domain, including an anisotropy parameter κ, as described by Bilonick (Bilonick, 1988) and revisited by Snepvangers et al. (2003) and Gräler et al. (2016). This is the sum-metric covariance model:

$$C_{\text{sm}}(h_S, h_T) = C_s(h_S) + C_t(h_T) + C_{\text{joint}}\left(\sqrt{h_S^2 + (\kappa \cdot h_T)^2}\right). \tag{4.7}$$

The variogram is derived similar to the covariance (Snepvangers et al., 2003; Gräler et al., 2016) as:

$$\gamma_{\text{sm}}(h_S, h_T) = \gamma_S(h_S) + \gamma_T(h_T) + \gamma_{\text{ST}}\left(\sqrt{h_S^2 + (\kappa \cdot h_T)^2}\right). \tag{4.8}$$

`stUPscales` implements routines of the R package `gstat` (Pebesma, 2004; Gräler et al., 2016) representing the spatio-temporal covariance models. Besides the sum-metric model, four additional models are available in `gstat`: Separable; Product-sum; Metric; and Simplified sum-metric.

4.4.3 Results

Events Selected and Space-Time Full Grid Object

`stUPscales` has a dataset with the selected events per station as an `xts` object in space-wide format for easy visualisation of the 25 stations in the object `event.subset.xts`. The spatial locations of the 25 rain gauge stations are provided as a dataset in `SpatialPointsDataFrame` format from the `sp` package (Pebesma and Bivand, 2005; Bivand et al., 2013). The boundary of the country of the Grand-Duchy of Luxembourg is given as a dataset in `sp` format `SpatialPolygonsDataFrame`. The data were converted to a space-time full data.frame (`STFDF`) of the package `spacetime` (Pebesma, 2012), through the function `stConstruct` from the same package. The object is visualised in Figure 4.6.

Space-Time Variograms

Once the observation period was defined, we computed the experimental spatio-temporal variogram according to Equation (4.6) using the function `variogramST` of the `gstat` package, with time lags of 50 minutes and space lags of 5,000 m. The spatio-temporal anisotropy was estimated by the `estiStAni` function (`gstat`). Before proceeding to the fit of the parameters, the experimental variogram and the anisotropy were scaled to have distances in kilometers.

The sum-metric variogram model (Equations (4.7) and (4.8)) was defined in `gstat` by using the `vgmST` function. The model was fitted using the `fit.StVariogram` `gstat` function with the L-BFGS-B optimisation algorithm described in Byrd et al. (1995), which allows box constraints with lower and upper bounds. All space-time variogram parameters (spatial, temporal and joint nugget, range and partial sill) were estimated by minimising the mean squared error (MSE). The MSE of the best model fitted is 0.00013. Figure 4.7 shows the experimental spatio-temporal variogram (a), the fitted spatio-temporal variogram (b) and the residuals of the spatio-temporal variogram model for the sum-metric model (c). The fitted parameters of the sum-metric model are given in Table 4.4.

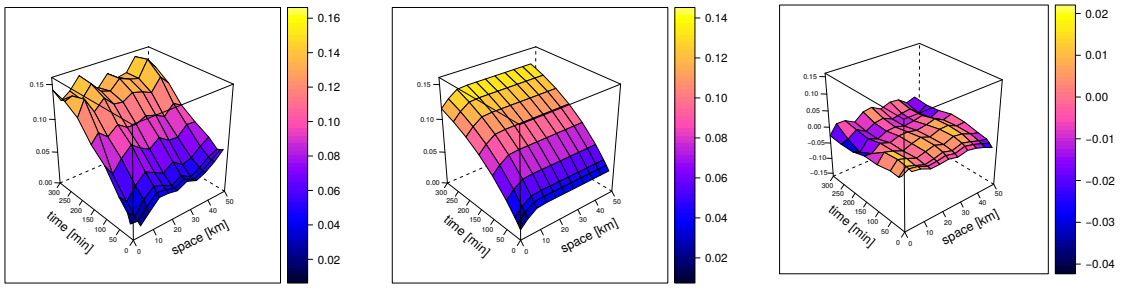


Figure 4.7: Spatio-temporal variograms. (**a**) empirical (sample) variogram; (**b**) theoretical sum-metric model; (**c**) difference between sample and best fitting theoretical sum-metric model.

Space-Time Ordinary Kriging

Upon definition of the space-time variogram, we proceeded to create prediction maps using space-time ordinary kriging. For this, a prediction grid over Luxembourg was created and predictions were made by the function `krigeST` from `gstat`. The prediction maps are shown in Figure 4.8. Finally, we computed the lower and upper boundary of the 90 percent prediction interval as shown in Figures 4.9 and 4.10.

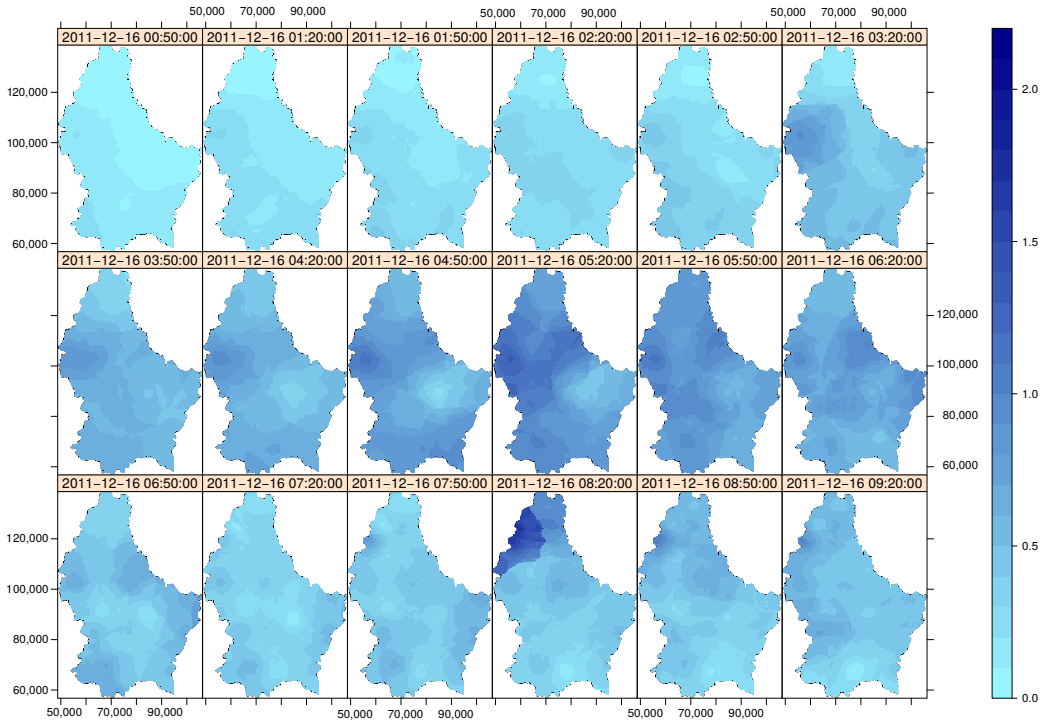


Figure 4.8: Prediction maps obtained with the spatio-temporal sum-metric model.

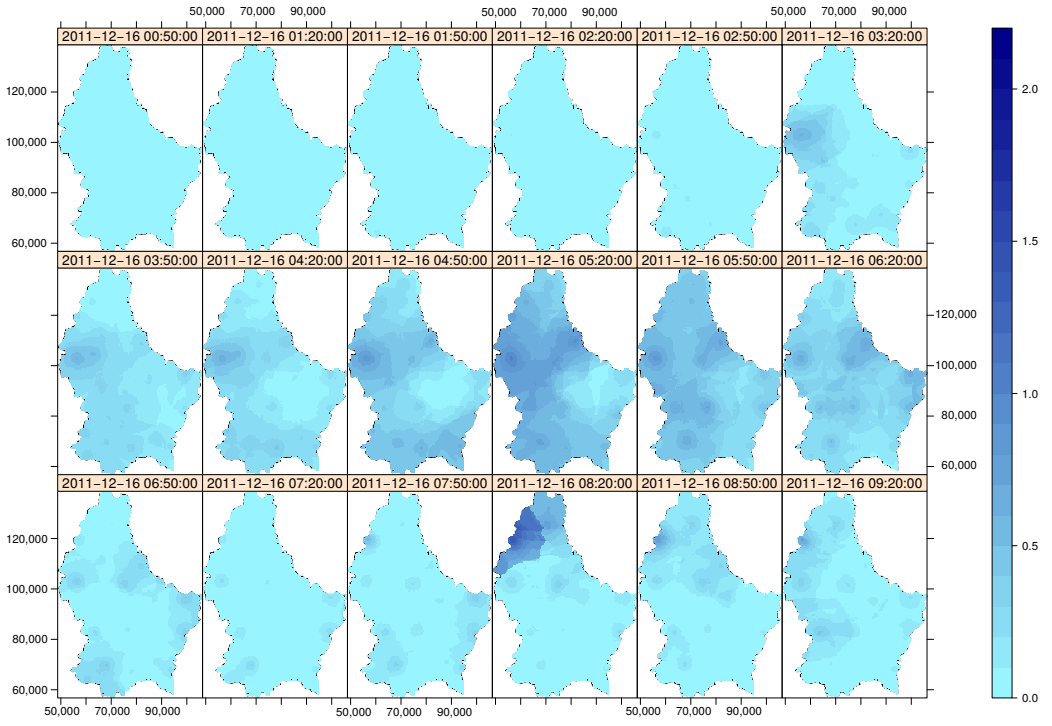


Figure 4.9: Predictions obtained with the spatio-temporal sum-metric model (lower boundary of the 90 percent prediction interval).

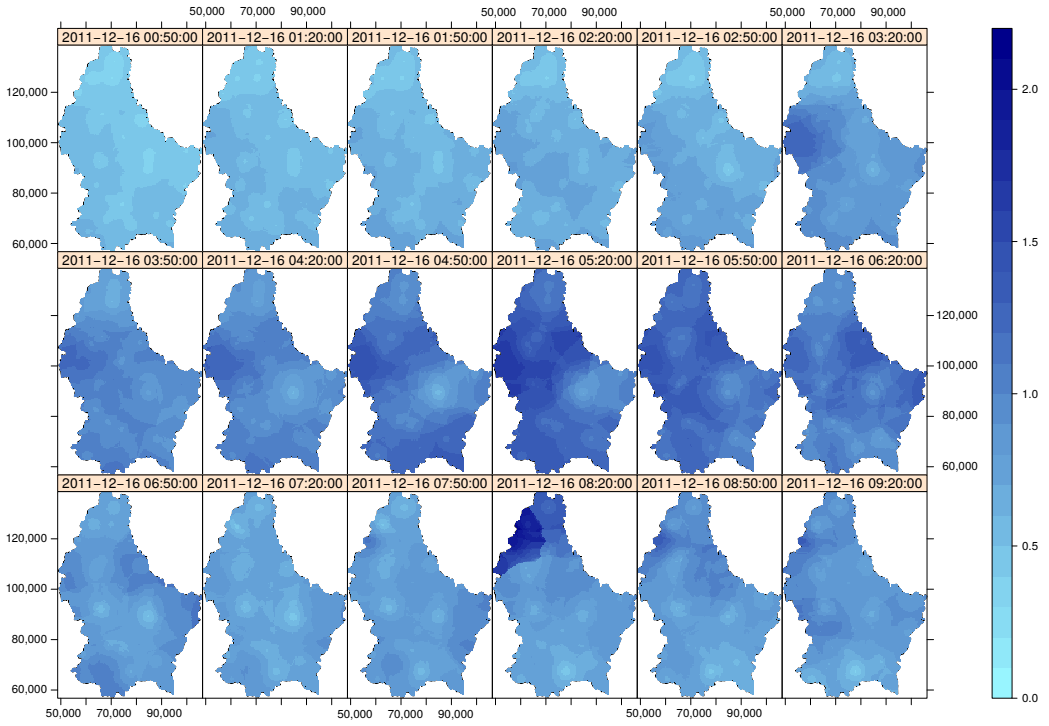


Figure 4.10: Predictions obtained with the spatio-temporal sum-metric model (upper boundary of the 90 percent prediction interval).

Table 4.4: Space-time variogram model.

	Partial Sill	**Range**
Spatial component		
Nugget	0	0
Spherical	0.021	10
Time component		
Nugget	0	0
Exponential	0.129	201.9
Joint component		
Nugget	0.016	0

With this, the stUPscales package is ready to characterise space-time uncertainties of environmental variables. The next steps for subsequent work are to simulate realisations of such variables and analyse the uncertainty propagation.

4.5 Discussion

We developed the `stUPscales` R package as a contribution to state-of-the-art software tools designed to perform spatio-temporal uncertainty characterisation and uncertainty propagation analysis of environmental models across different scales. We applied the tool in two examples where the class, methods and functions were tested satisfactorily as an illustration of the capabilities of the package. In this section, we highlight and discuss the main results obtained and put these in perspective with other studies.

4.5.1 Temporal and Spatio-Temporal Model Inputs

In Example 1, we used a simplified urban drainage model, EmiStatR, for testing purposes. We used this model because of its simplification and parallel computing coding, allowing a seamless uncertainty propagation in the R environment.

The main input variable of the modelling framework developed in the example corresponded to a time series of precipitation included as a dataset into the R-package EmiStatR. This time series includes records from 1 January to 2 February 2016 (4462 observations). The total precipitation in the time series is 89.6 mm, with a mean value for non-zero data of 0.18 mm and maximum value of 0.90 mm. While the recording time step is 10 min, we aggregated the precipitation to 30 min to illustrate the aggregation capabilities of `stUPscales`. We characterised the uncertainty of the temporal input by invoking the temporal autocorrelation function.

For Example 2, we used time series of precipitation for 25 rain gauges distributed over the country of Luxembourg as an input dataset. The event selected was a period of 10 h from 12:00 a.m. to 10:00 a.m. on 16 December 2011. Next, we characterised the spatio-temporal uncertainty by inferring the experimental space-time variogram from the observations for the entire spatial domain and time period.

4.5.2 Temporal Aggregation of Model Input

In Example 1, we illustrated the temporal aggregation function `Agg.t` of the `stUPscales` package. The level of aggregation was a 30-min time step. The resulting time series is composed of 1488 observations, summing up to a total precipitation of 89.6 mm, with a mean value for non-zero data of 0.33 mm and maximum value of 2.00 mm. This level of aggregation is useful when a large time series is used as input in the uncertainty propagation procedure, reducing the computational burden required for the MC simulation.

4.5.3 Model Input Uncertainty Propagation

In Example 1, we used the MC method for model input uncertainty propagation through the environmental model. In so doing, we accounted for temporal and cross-correlations

when defining model input uncertainty and performing uncertainty propagation.

Model Input Uncertainty and Monte Carlo Simulation Set-Up

Moret et al. (2017) present a literature review about characterisation of input uncertainties in strategic energy planning models. They found that, in the context of strategic design problems, Tock and Maréchal (2015) use normal, uniform and beta distributions to define uncertainty of economic parameters with lower and upper limits defined from institutional reports. In addition, in the context of energy systems design under uncertainty, Dubuis (2012) proposes a methodology based on statistical theory, i.e., on definition of probability distribution functions. However, uncertainty characterisation using pdfs is not always possible when data are unavailable. In the absence of data for defining future uncertainties in the UK energy transition pathways, Pye et al. (2015) propose using triangular distributions for representing uncertainty between upper and lower limits, defined from the literature. To account for uncertainty in the assessment of biomass-to-fuel strategies in wastewater treatment applications, Sin et al. (2009) assign different levels of cost uncertainty (low, medium, high), taking into account maturity and complexity. In addition, upper and lower limits to all model parameters were defined by expert knowledge and establishing uniform distributions. In the design of flexible multi-generation systems, Lythcke-Jørgensen et al. (2016) assume variations of ±25% and uniform distributions for parameters of investment and operating cost.

In the urban drainage modelling domain, Wijesiri et al. (2016) quantified uncertainty by uncertainty limits associated with predicted build-up and wash-off processes. Zoppou (2001) states that *"providing the uncertainty associated with model outcomes in terms of uncertainty limits is an effective way to enhance management and planning decisions"*.

Most of the reviewed literature highlighted that, regarding parameter uncertainty characterisation, there is a difficulty to define pdfs based on data at proper quantity and quality. In this sense, Moret et al. (2017) present a method to characterise input uncertainty for strategic energy planning models. This method proposes the definition of a set of criteria for the establishment of variation ranges for uncertain parameters, instead of the definition of pdfs. For some techniques, such as robust optimization, full pdfs are not needed and specification of uncertainty ranges is enough (Bertsimas and Sim, 2004).

The `stUPscales` package includes both pdf and range characterisations for input uncertainty quantification. The pdf option allows the definition of normal, log-normal and truncated normal pdfs. The range option consists of the definition of a uniform pdf where the lower bound (minimum value) and the upper bound (maximum value) are defined as function arguments. Additionally, two more options are available for the characterisation of model input uncertainty: (1) discrete, where a uniform discrete sampling is defined; and (2) constant, where a unique constant value is defined in time and space, i.e., suitable for

representing model inputs that are not uncertain. With all these possibilities, we cover an important range of characterisations for model input uncertainty used in practice.

Monte Carlo Simulation and Analysis

In Example 1, the selected water quantity input variables for the uncertainty propagation analysis (water consumption, qs, infiltration inflow water, qf, impervious area, A_{imp}, run-off coefficient for impervious area, C_{imp}, run-off coefficient for pervious area, C_{per}, population equivalents, pe, orifice coefficient of discharge, C_d and the initial water level in the chamber, Lev_{ini}), had been defined to quantify input uncertainty and analyse how it propagates through the environmental model, EmiStatR, to the different model outputs related to water quantity ('CSO summary' in Table 4.2, and Figure 4.3).

Regarding the selected water quality input variables for the uncertainty propagation analysis (sewage COD pollution per capita [PE] load per day, $C_{COD,s}$, sewage NH_4 pollution per capita [PE] load per day, $C_{NH4,s}$, and rain COD pollution, COD_r), these reflect how the variability in the input uncertainty definition is translated to the model output ('COD in spill volume' and 'NH_4 in spill volume', as shown in Table 4.2 and Figure 4.4).

In the case of total concentration of COD (CCOD) in the spill volume, a mean value of 145 mg·l^{-1} and a maximum value of 190 mg·l^{-1} were simulated in the deterministic run. In the Monte Carlo simulation, a mean value of 255 mg·l^{-1} and a maximum value of 511 mg·l^{-1} in the 95 percentile band were obtained. These large differences indicate a large uncertainty in this variable. The relative increase of the mean and maximum are 1.8 and 2.7, respectively.

In the case of total concentration of NH_4 (CNH4) in the spill volume, a mean value of 4.1 mg·l^{-1} and a maximum value of 6.5 mg·l^{-1} were simulated in the deterministic run. In the Monte Carlo simulation, a mean value of 5.5 mg·l^{-1} and a maximum value of 12.6 mg·l^{-1} in the 95 percentile band were obtained. While the means are similar, the large increase in the maximum indicates that uncertainty in this variable is also substantial, as already indicated by the wide prediction bands (grey area) shown in Figure 4.4.

These results demonstrate the usefulness of performing model input uncertainty propagation through the environmental model, which is an important goal in the evaluation and communication of uncertainties in environmental modelling.

In order to further validate model accuracy, a complete uncertainty propagation analysis would have to involve (cross-)validation using independent observations. This was beyond the scope of this chapter and is done in a full uncertainty propagation analysis of the EmiStatR model, including a stochastic sensitivity analysis to analyse the contribution of each source of uncertainty to the overall uncertainty of the model results.

4.5.4 Space-Time Characterisation of Model Input Uncertainty

In Example 2, we showed how space-time model input uncertainty is modelled in `stUPscales`. We used space-time ordinary kriging, with a space-time unknown constant mean of the process and a space-time stochastic residual for which the literature proposes different covariance models (Gräler et al., 2016). However, in the space-time domain, the fitting of the covariance model is not trivial compared to that in spatial kriging. We used the sum-metric model (Snepvangers et al., 2003; Gräler et al., 2016), which gave satisfactory results. We showed space-time prediction maps of the precipitation for a 10 h time period on 16 December 2011 for the Grand-Duchy of Luxembourg.

4.5.5 Data Format for Uncertainties

We have coded `stUPscales` to be compatible with standard spatio-temporal data concepts and formats in R, such as the spatio-temporal classes ST of the spacetime package (Pebesma, 2012). This allows for seamless linkage of geospatial and spatio-temporal models and sub-models across multiple scales.

In addition, despite an attempt to create the Uncertainty Markup Language (UncertML) (Williams et al., 2009) as an Open Geospatial Consortium (OGC) standard to store metadata for uncertainty propagation applications (Hengl et al., 2010), there is still an open question about the suitability of this data format to hold, visualise and communicate uncertainties together with the space-time ST classes in R. Furthermore, there is a potential to develop a new class or extend the ST classes to embed the uncertainty information regarding data uncertainty, input uncertainty characterisation, output uncertainty and summary statistics.

4.6 Conclusions and Future Research

We presented the R package **s**patio-**t**emporal and **U**ncertainty **P**ropagation across multiple **scales**, `stUPscales`. This package constitutes a contribution to the state-of-the-art of open source tools that aim to support uncertainty propagation in the spatio-temporal domain. The main class, methods and functions of the package were presented and illustrated with an example of temporal model input uncertainty propagation through a simplified lumped urban drainage model accounting for water quality modelling. We also showed how the package can be used to model and predict environmental variables that vary in time and space. Using simulations of these types of variables in MC uncertainty propagation will be done in subsequent work.

The main contribution of `stUPscales` compared to existing uncertainty propagation tools is that it is not restricted to uncertainty analysis of purely spatial applications but can also handle uncertain temporal and spatio-temporal variables. In addition, it is designed to handle aggregation and disaggregation of uncertain temporal, spatial and spatio-temporal

variables. Through two examples, we demonstrated that stUPscales is suitable for characterising uncertainty in spatial, temporal and spatio-temporal environmental variables (model inputs) as probability distribution functions (pdfs) and as uni- and multi-variate autoregressive models. Moreover, it is possible to sample from these pdfs (to support MC UP analysis) and to generate realisations of autoregressive models. We illustrated the aggregation functionality of `stUPscales` by averaging realisations of precipitation in time. We propagated model input uncertainty through a simplified urban water model and analysed the results of an MC UP. Finally, we showed how space-time geostatistical interpolation is done in `stUPscales`.

It is worthwhile to explore the development of a new class or to extend the ST classes so that these incorporate uncertainty information regarding data uncertainty, input uncertainty characterisation, output uncertainty and summary statistics. We recommend that future implementations take this matter into account.

Further extension of the package will allow for implementing new methods and functions for spatio-temporal disaggregation of model inputs and outputs, giving more flexibility to the proposed workflow when linking models with different spatio-temporal support or across multiple space-time scales.

Chapter 5

Temporal uncertainty analysis of an urban water system in Luxembourg

This chapter is based on:

Torres-Matallana, J. A.; Leopold, U., and Heuvelink, G. B. M. Multivariate autoregressive modelling and conditional simulation for temporal uncertainty analysis of an urban water system in Luxembourg. *Hydrology and Earth System Sciences (HESS)*, 25(1):193–216, 2021. doi: 10.5194/hess-25-193-2021. URL `https://hess.copernicus.org/articles/25/193/2021/`

Abstract

Uncertainty is often ignored in urban water systems modelling. Commercial software used in engineering practice often ignores the uncertainties of input variables and their propagation because of a lack of user-friendly implementations. This can have serious consequences, such as the wrong dimensioning of urban drainage systems (UDSs) and the inaccurate estimation of pollution released to the environment. This chapter introduces an uncertainty analysis in urban drainage modelling, built on existing methods and applied to a case study in the Haute-Sûre catchment in Luxembourg. The case study makes use of the EmiStatR model which simulates the volume and substance flows in UDS using simplified representations of the drainage system and processes. A Monte Carlo uncertainty propagation analysis showed that uncertainties in chemical oxygen demand (COD) and ammonium (NH_4) loads and concentrations can be large and have a high temporal variability. Furthermore, a stochastic sensitivity analysis that assesses the uncertainty contributions of input variables to the model output response showed that precipitation has the largest contribution to output uncertainty related with water quantity variables, such as volume in the chamber, overflow volume, and flow. Regarding the water quality variables, the input variable related to COD in wastewater has an important contribution to the uncertainty for the COD load (66 %) and COD concentration (62 %). Similarly, the input variable related to NH_4 in wastewater plays an important role in the contribution of total uncertainty for the NH_4 load (34 %) and NH_4 concentration (35 %). The Monte Carlo (MC) simulation procedure used to propagate input uncertainty showed that, among the water quantity output variables, the overflow flow is the most uncertain output variable, with a coefficient of variation (cv) of 1.59. Among water quality variables, the annual average spill COD concentration and the average spill NH_4 concentration were the most uncertain model outputs (coefficients of variation of 0.99 and 0.82, respectively). Also, low standard errors for the coefficient of variation were obtained for all seven outputs. These were never greater than 0.05, which indicates that the selected MC replication size (1500 simulations) was sufficient. We also evaluated how the uncertainty propagation can more comprehensively explain the impact of water quality indicators for the receiving river. While the mean model water quality outputs for COD and NH_4 concentrations were slightly above the threshold, the 0.95 quantile was 2.7 times above the mean value for COD concentration and 2.4 times above the mean value for NH_4. This implies that there is a considerable probability that these concentrations in the spilled combined sewer overflow (CSO) are substantially larger than the threshold. However, COD and NH_4 concentration levels of the river water will likely stay below the water quality threshold, due to rapid dilution after CSO spill enters the river.

5.1 Introduction

Combined sewer systems are important components of urban water infrastructure. These systems are typically found in old and large cities (Baker, 2009; Litrico and Fromion, 2009) and are designed to transport the water generated and accumulated in an urban catchment to the receiving water body. During normal conditions, all water is transported to the treatment facility before it is released to the environment. This is the so-called throttled outflow or pass-forward flow (Hager, 2010). However, during extreme conditions with heavy precipitation, the combined sewer overflow (CSO) discharges excess water directly to nearby streams, rivers, lakes, or other water bodies (Baker, 2009). The CSO contains polluted water and solid matter (Hager, 2010), which, when released to the environment, can have a damaging impact on the water quality status of the receiving waters (Bachmann-Machnik et al., 2018; Gasperi et al., 2012). CSO pollutant load emissions are of similar or greater magnitude than the emissions from wastewater treatment plants (Gasperi et al., 2012; Bachmann-Machnik et al., 2018). CSO discharge impacts are mainly high peak flows, high organic loads from single events, which can lead to oxygen depletion, and ecotoxic concentrations of ammonia (NH_3) (Miskewitz and Uchrin, 2013; Bachmann-Machnik et al., 2018). To reduce pollution in receiving waters it is important to minimise CSO load and concentration.

One of the main variables is the chemical oxygen demand (COD), which is an indicator of organic compounds in water. It is used to measure the effluent quality (Viana da Silva et al., 2011). High levels of COD are correlated with a decrease in the amount of dissolved oxygen (DO) available for aquatic organisms. A depletion in the DO concentration in the water column from near $9\,\mathrm{mg\,L^{-1}}$ (the maximum solubility of oxygen in estuarine water on an average summer day) to below $2\,\mathrm{mg\,L^{-1}}$ is referred to as hypoxia. If hypoxic conditions are reached, the health of the ecosystem is affected, causing physiological stress, and even death, to aquatic organisms (Committee on Environmental and Natural Resources - CENR, 2003). Ammonium (NH_4) is another important variable and is an indicator of nitrogen compounds in water. Concentrations of NH_4 in water and wastewater are relevant because high levels of nitrogen in receiving waters can cause eutrophication and, therefore, excessive growth of algae and other micro-organisms, resulting in oxygen-dissolved depletion and fish toxicity (Huang et al., 2010).

To better assess environmental impacts, numerical models are applied in urban hydrology to simulate CSO emissions into the environment. It is recommended, however, that such modelling approaches consider the inherent uncertainty associated with the system representation and the approximation of the model to the reality (Hutton et al., 2011). Moreover, the model inputs are also not free of errors, and associated uncertainties will also propagate to the model output (Heuvelink, 1998).

A total of five approaches for representing the presence or absence of uncertainty and how

it is represented in the context of urban water systems are often distinguished (Walker et al., 2003; Refsgaard et al., 2007; van der Keur et al., 2008; Bach et al., 2014) as follows: (1) determinism, (2) statistical uncertainty, (3) scenario uncertainty, (4) recognised ignorance, and (5) total (unrecognised) ignorance. Following van der Keur et al. (2008), determinism applies when we have knowledge with absolute certainty about the system under analysis. This is the ideal world case, which is not realistic for urban hydrology systems. The statistical approach is useful when it is possible to describe uncertainty in statistical terms, i.e. when uncertainty can be characterised by probability distribution functions (pdf's). The scenario approach, in contrast, applies when quantitative probabilities cannot be determined and, instead, qualitative measures of uncertainty are used. It is used when possible outcomes of uncertain inputs are known but the probabilities of these outcomes are not (Brown, 2004). There is also no claim that the list of possible outcomes (scenarios) is exhaustive. Recognised ignorance occurs when there is an awareness of lack of knowledge but without any further possibility to process and address the recognised uncertainty. This is the case for very complex functional or inherently unidentifiable relationships, when, for example, predictions are infeasible due to the chaotic behaviour of the system or when our understanding of the system behaviour is too limited (van der Keur et al., 2008). This is common in social systems where the behaviour of humans and groups of humans may often unpredictable. Finally, total ignorance is the state of complete lack of awareness about imperfect knowledge (van der Keur et al., 2008). It is the opposite of determinism and reflects a state in which we do not know that we do not know (Walker et al., 2003). Among the approaches described above, in this chapter we will use the statistical approach to characterise and propagate uncertainties.

A total of three main sources of uncertainty in the context of performance evaluation analysis and design of urban water infrastructure and urban drainage modelling are identified (Walker et al., 2003; Neumann, 2007; Deletic et al., 2012). First, model input uncertainty is related to errors in input data, i.e. in driving forces such as precipitation. Second, parameter uncertainty is related to the uncertainty regarding the (calibrated) parameters of the model. Third, model structural uncertainty relates to uncertainty due to model conceptualisation and simplification. For instance, an urban drainage model might ignore certain sub-processes, such as evaporation or chemical transformation, or might simplify a non-linear relation between model variables to a linear relation. These types of uncertainties are not captured in model input and model parameter uncertainty and are represented by model structural uncertainty. The focus of this work is on the propagation of model input uncertainty.

Regarding methods for uncertainty propagation analysis, a distinction can be made between analytical methods, such as the Taylor series method (Heuvelink, 1998), and numerical techniques, such as Monte Carlo (MC) simulation. Numerical techniques are more flexible and hence more convenient for analysing uncertainty propagation with complex models (Zoppou, 2001). MC simulations are computationally demanding, especially in

the case of complex models, but they can still be used if there are sufficient computational resources (Bastin et al., 2013), among others, because it can greatly benefit from parallel computing.

Although uncertainty propagation analysis has been applied extensively in hydrologic modelling (e.g. Beven and Binley, 1992; Kuczera and Parent, 1998; Hutton et al., 2011; Vrugt et al., 2003b,a; Vrugt and Robinson, 2007; Renard et al., 2010; Datta, 2011), the number of applications of long-term simulations in urban drainage modelling is limited and typically does not consider the influence of the temporal and spatial correlation in the analysis of the propagation of input uncertainty. Temporal correlation occurs in uncertain dynamic variables, such as precipitation and COD of household wastewater, because values of these variables over short time lags will be more similar than over large time lags. The same concept applies to variables that are spatially distributed (Webster and Oliver, 2007). It is important to take the temporal (and spatial) correlation of uncertain inputs into account because this may have a major influence on the outcomes of an uncertainty analysis (Heuvelink, 1998). In this chapter, we perform a temporal uncertainty propagation analysis in urban water modelling using a MC simulation. As a case study, we use the simplified model EmiStatR (Torres-Matallana et al., 2018a) to predict wastewater volume, COD, and NH_4 concentrations in CSOs for three urban–rural sub-catchments of the Haute-Sûre catchment in the northwest of Luxembourg.

The objectives of this study are to (1) select and characterise the main sources of input uncertainty accounting for the temporal auto- and cross-correlation within EmiStatR; (2) propagate input uncertainty through EmiStatR, taking into account the temporal auto- and cross-correlation of uncertain dynamic inputs; and (3) quantify and assess the contributions of each uncertainty source to model output uncertainty dynamically (over time) for the Luxembourg case study.

5.2 Materials and methods

5.2.1 The EmiStatR model

EmiStatR is used to simulate CSO flows and water quality concentrations. Details regarding the conceptual and mathematical model are provided in Torres-Matallana et al. (2018a). A list of the EmiStatR inputs and outputs is provided in Appendix A. The main components of the EmiStatR model are (1) dry weather flow (DWF), including infiltration flow (IF), (2) pollution of DWF, (3) rain weather flow (RWF), (4) pollution of RWF, (5) combined sewer flow (CSF) and pollution, and (6) combined sewer overflow (CSO) and pollution. Figure 5.1 illustrates the scheme of the sewer system analysed.

Basically, the total dry weather flow, Q_{DWF} ($\mathrm{L\,s^{-1}}$), is calculated as follows:

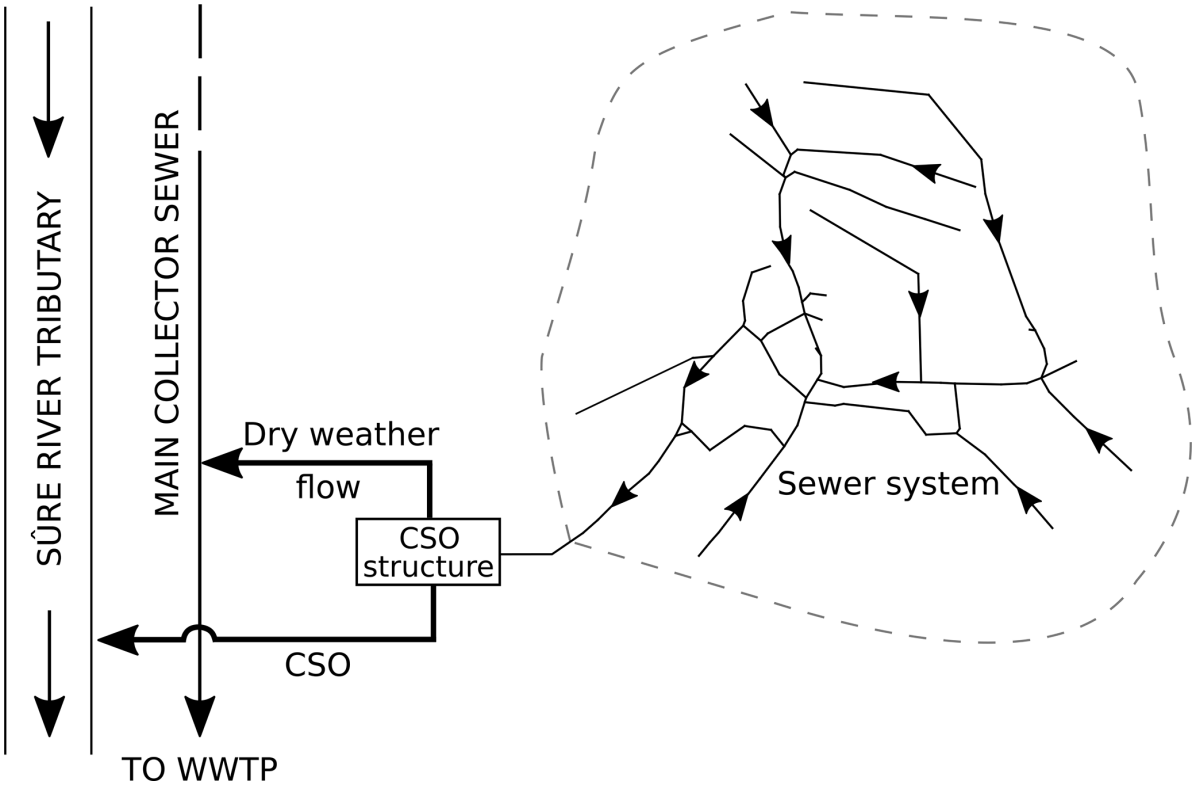


Figure 5.1: Scheme of the sewer system analysed. Adapted from Andrés-Doménech et al. (2010).

$$Q_{\mathrm{DWF}_t} = Q_{\mathrm{s}_t} + Q_{\mathrm{f}_t}, \tag{5.1}$$

where Q_{DWF_t} ($\mathrm{L\,s^{-1}}$) is the dry weather flow at time t, and Q_{s_t} ($\mathrm{L\,s^{-1}}$) is the dry weather flow of the residential sewage in the catchment at time t, calculated as $86\,400^{-1}\,\mathrm{pe}_t\,\mathrm{qs}_t$ (where $86\,400 = 24 \times 60 \times 60$ is a measurement unit conversion factor), with pe_t (PE) the population equivalents of the connected CSO structure at time t, and qs_t ($\mathrm{L\,PE^{-1}\,d^{-1}}$) is the individual water consumption of households at time t. Q_{f_t} ($\mathrm{L\,s^{-1}}$) is the infiltration flow at time t that enters the pipes from groundwater flow through cracks and joints, calculated as $A_{\mathrm{imp}} \cdot q_{\mathrm{f}_t}$, where A_{imp} (ha) is the impervious area of the catchment, and q_{f_t} ($\mathrm{L\,s^{-1}\,ha^{-1}}$) is the infiltration water inflow flux (specific infiltration discharge from groundwater flow) at time t. Variables qs_t and pe_t are dynamic and can be defined as time series with daily, weekly, and seasonal patterns.

The contribution of rainwater to the combined sewage flow, Q_{r} ($\mathrm{m^3\,s^{-1}}$), is derived from precipitation as follows:

$$Q_{\mathrm{r}_t} = \frac{1}{6} \cdot P_{t-t_{\mathrm{fS}}} \cdot [C_{\mathrm{imp}} \cdot A_{\mathrm{imp}} + C_{\mathrm{per}} \cdot (A_{\mathrm{total}} - A_{\mathrm{imp}})], \tag{5.2}$$

where $\frac{1}{6}$ is a factor for units conversion, $P_{t-t_{\mathrm{fS}}}$ is precipitation at time $t - t_{\mathrm{fS}}$ ($\mathrm{mm\,min^{-1}}$),

t_{fS} is a delay in time response related to flow time in the sewer system, A_{imp} is the impervious area of the catchment (ha), A_{total} is the total area of the catchment (ha), C_{imp} is the run-off coefficient for impervious areas (–), and C_{per} is the run-off coefficient for pervious areas (–). From Q_{r_t}, the CSO volume calculation is based on the exceeding volume stored in the combined sewer overflow chamber (CSOC). The CSO volume depends on four CSOC stages, namely (1) filling up, (2) CSO spill volume, (3) stagnation, and (4) emptying. The sum of the total dry weather flow, Q_{DWF_t}, and the rain water flow, Q_{r_t}, is called combined sewer flow at time t, Q_{CSF_t}.

The COD load, $B_{\mathrm{COD,\,Sv}}$ (g), in the spill overflow volume is calculated as a function of the spill overflow volume at time t, V_{Sv_t} (m^3), a combined sewer mixing ratio at time t, $\mathrm{cs}_{\mathrm{mr}_t}$ (–), the mean dry weather pollutant concentration at time t, C_{COD_t} ($\mathrm{mg\,L^{-1}}$), and the concentration due to rainwater pollution at time t, $\mathrm{COD}_{\mathrm{r}_t}$ ($\mathrm{mg\,L^{-1}}$) as follows:

$$B_{\mathrm{COD,\,Sv}_t} = (\mathrm{cs}_{\mathrm{mr}_t} + 1)^{-1}\, V_{\mathrm{Sv}_t}\, (\mathrm{cs}_{\mathrm{mr}_t} \cdot C_{\mathrm{COD}_t} + \mathrm{COD}_{\mathrm{r}_t})\,. \tag{5.3}$$

The variable V_{Sv_t} depends directly on the water volume in the CSO chamber at time t, V_{Chamber_t} (m^3). It is computed as follows:

$$V_{\mathrm{Sv}\,t} = \begin{cases} V_{\mathrm{r}_t} + V_{\mathrm{dw}_t} - V_{\mathrm{d}_t}, & \text{if } V_{\mathrm{Chamber}\,t} = V, \\ V_{\mathrm{Chamber}\,t} - V & \text{if } V_{\mathrm{Chamber}\,t} > V, \\ \epsilon & \text{if } V_{\mathrm{Chamber}\,t} < V, \end{cases} \tag{5.4}$$

where V_{r_t} is the rain weather volume at time t accumulated during a time interval Δt (min), V_{dw_t} (m^3) is the total dry weather volume (amount of dry weather water in combined sewage flow) at time t, V_{d_t} is the volume of throttled outflow to the wastewater treatment plant (WWTP) at time t (m^3), V (m^3) is the CSOC volume, and ϵ is a numerical precision term set equal to 10^{-5} (m^3). While V_{Sv}, $\mathrm{cs}_{\mathrm{mr}}$, and C_{COD} are dynamic, $\mathrm{COD}_{\mathrm{r}_t}$ can either be dynamic or assumed constant if the pollution concentration is assumed constant in time. C_{COD_t} ($\mathrm{mg\,L^{-1}}$) is calculated as follows (Torres-Matallana et al., 2018a):

$$C_{\mathrm{COD}\,t} = \frac{10^3 \cdot \mathrm{pe}_t \cdot C_{\mathrm{COD,\,S}}}{\mathrm{qs}_t \cdot \mathrm{pe}_t + 86\,400 \cdot A_{\mathrm{imp}} \cdot q_{\mathrm{f}_t}}, \tag{5.5}$$

where $C_{\mathrm{COD,\,S}}$ is the COD sewage pollution per capita (PE) load per day ($\mathrm{g\,PE^{-1}\,d^{-1}}$). Similar equations, as above, apply to the second water pollution indicator NH_4.

5.2.2 Sewer system in the Haute-Sûre catchment

The study area is composed of three sub-catchments of the Haute-Sûre catchment in the northwest of Luxembourg. The combined sewer system drains three villages, namely

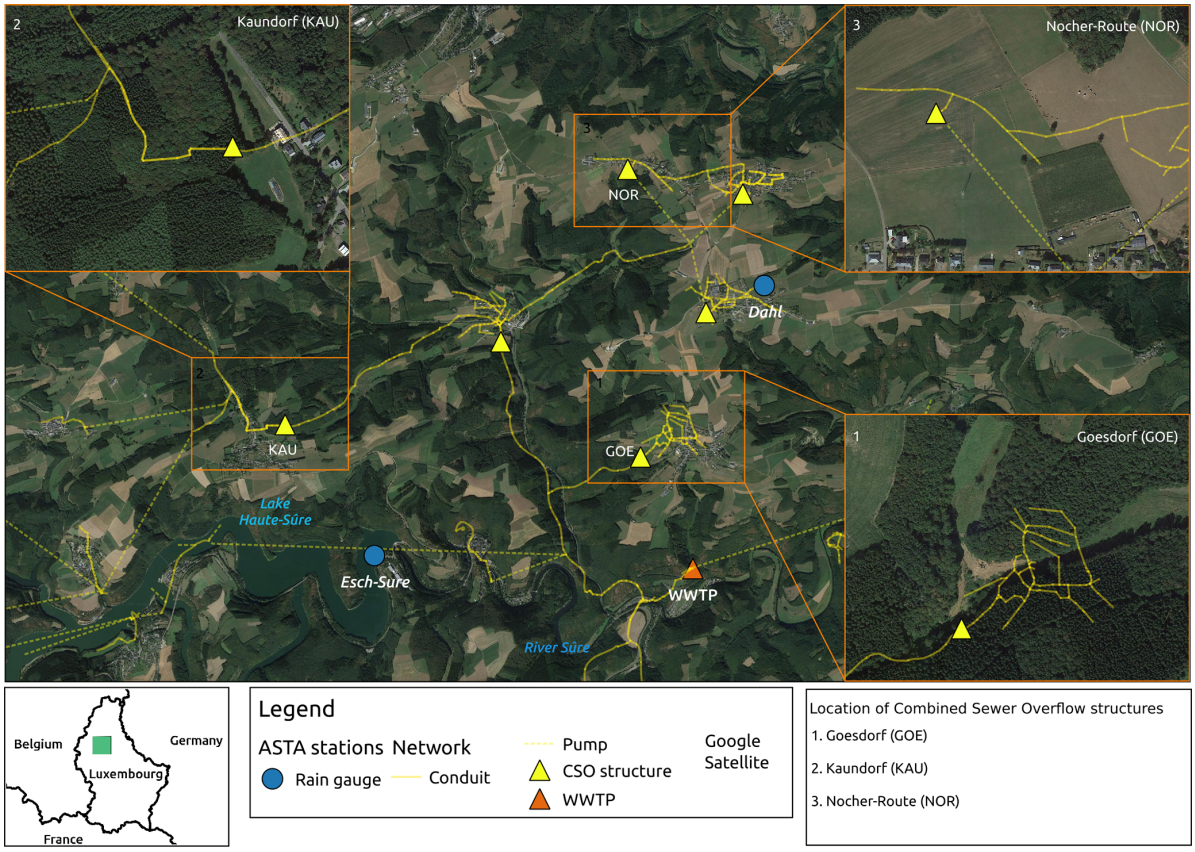


Figure 5.2: The three Haute-Sûre sub-catchments and locations of the combined sewer overflow (CSO) structures considered in this study. The background map is provided by © Google Maps.

Goesdorf (GOE), Kaundorf (KAU), and Nocher-Route (NOR). The local sewer system downstream of each village has a CSO structure to store pollutant peaks in the first flush of combined sewage flows. Figure 5.2 depicts the location of the CSO structures and the delineation of the sub-catchments. The main land use types in the villages are residential, smaller industries, and farms. Outside the villages, forest and agricultural arable and grassland are the dominating land uses. The receiving water bodies of the CSO structures are tributaries of the river Sûre (or Sauer in German).

5.2.3 Input data

The input variables of the EmiStatR model are shown in Table A.1 in the Appendix A. Following Torres-Matallana et al. (2018a), seven input variables were calibrated, namely water consumption (qs_t), infiltration flow (q_{f_t}), flow time structure equivalent to the time of concentration to the combined sewer overflow structure (t_{fS}), run-off coefficient for impervious area (C_{imp}), run-off coefficient for pervious area (C_{per}), orifice coefficient of discharge (C_{d}), and the initial water level ($\mathrm{Lev}_{\mathrm{ini}}$). The main objective of the calibration process is to appropriately represent the water volume in the CSOC.

The observed precipitation (P_t) is a 1 year time series for 2010 at a 10 min time interval, measured at stations Esch-sur-Sûre and Dahl (Fig. 5.2). The variable water consumption (qs_t) is also dynamic and represented as a time series with a daily pattern according to factors proposed in the German ATV-DVWK-A 134 guideline (Evers et al., 2000).

The hydraulic variable measured is the water level in the CSOC, namely Lev (m). The temporal resolution of the measurements of Lev is 30 s. Regarding the wastewater quality (WWQ) characterisation, values of $C_{\mathrm{COD,S}}$ and $C_{\mathrm{NH_4,S}}$ in the wastewater were derived from DWF measurements at Goesdorf, Kaundorf, and Nocher-Route. A total of 91 2 h composite samples were taken and measured in the laboratory to determine the concentrations of COD ($\mathrm{mg\,L^{-1}}$) and NH_4 ($\mathrm{mg\,L^{-1}}$). This led to the identification of seven at Goesdorf on 4 May 2011, 48 between 19 June and 21 July 2010 at Kaundorf, and 36 between 9 March and 2 August 2011 at Nocher-Route. The variables $\mathrm{COD_f}$ and $\mathrm{NH_{4f}}$ were set to zero because the pollution contribution of the infiltration water is negligible in the study area. The contribution of ammonium from rainwater $\mathrm{NH_{4r}}$ was assumed constant and set to $2.00\,\mathrm{mg\,L^{-1}}$, while $\mathrm{COD_r}$ was equal to zero. Table S1 summarises the base values of the general input variables, and Table S2 presents the general characteristics of each CSO structure for each village and the base values of input variables. These base values were used when running EmiStatR in the deterministic mode (see Sect. 5.3.1). Some of the variables were calibrated based on observations in the CSOC to simulate water level and concentrations and loads of pollutants spilled in the CSO to the stream, river, or lake. These variables are water consumption (qs), infiltration flow (q_{f}), time flow (t_{fS}), run-off coefficient for the impervious area (C_{imp}), run-off coefficient for the pervious area (C_{per}), orifice coefficient of discharge (C_{d}), and initial water level ($\mathrm{Lev_{ini}}$).

5.2.4 Selection of model input for uncertainty quantification

Following recommendations from Nol et al. (2010), not all model inputs were taken into account in the uncertainty propagation analysis. Only inputs that are very uncertain and to which the model output is very sensitive were included because these are the ones that have the largest contribution to output uncertainty (Heuvelink, 1998; Sect. 4.4). The level of uncertainty of the inputs was defined by expert judgement and similar case studies in the literature. A quick scan was used to determine the model sensitivity to each of the model inputs by running EmiStatR in deterministic mode with input base values given in Table S1. The level of model sensitivity was defined by analysing the mathematical model structure and components of the model, expert judgement, and simulations with EmiStatR. Inputs that rank high on both the level of uncertainty and on model sensitivity were selected and included in the uncertainty propagation analysis.

5.2.5 Uncertainty quantification of selected model input

Because we used a statistical approach, probability distribution functions (pdf's) are the basis for representing the uncertainties of the selected model inputs. This constitutes the most difficult step of an uncertainty propagation analysis and is done in different ways for constants and dynamic variables, as explained in the following sub-sections.

Uncertain constants

Following Heuvelink et al. (2007), an uncertain continuous numerical constant C can be characterised by its marginal (cumulative) pdf (mpdf) as follows:

$$F_C(c) = P(C \leq c). \tag{5.6}$$

Usually, a parametric approach can be taken, meaning that a common shape for F_C is chosen (e.g., normal, lognormal, exponential, or uniform) so that the mpdf is reduced to a number of parameters. In this study, the input variables that are in this category are water consumption (qs), infiltration inflow (q_{f}), total area (A_{total}), impervious area (A_{imp}), the run-off coefficients for impervious area (C_{imp}) and pervious area (C_{per}), population equivalents (pe), flow time structure (t_{fS}), and initial water level ($\mathrm{Lev}_{\mathrm{ini}}$).

Univariate autoregressive modelling

Dynamic uncertain inputs may be temporally autocorrelated. This may dramatically influence the outcome of an uncertainty propagation analysis and must therefore be accounted for. One way of doing this is by assuming an autoregressive order one (AR(1)) model as follows:

$$\begin{aligned} &y_t = \mu + \phi\,(y_{t-1} - \mu) + w_t, \quad t = 1, 2, \ldots, T, \\ &y_0 \sim \mathcal{N}(\mu,\, \sigma^2), \end{aligned} \tag{5.7}$$

where y_t is the uncertain input at time t, μ is its mean, ϕ is the autoregressive parameter ($0 \leq \phi < 1$), and w_t is a Gaussian white noise time series with mean zero and variance σ_w^2. The initial value y_0 is taken from a normal distribution with mean μ and variance σ^2. The parameters of the model can be estimated based on observations, or in the absence of observations, suitable values are taken based on expert judgement or literature reference values. Note that the effect of the initial condition usually fades out quickly and hence is not of great concern.

The implementation of the AR(1) model in R was done via the R function arima.sim of the R base package stats (R-Core-Team and contributors worldwide, 2017), both for model calibration and simulation.

Multivariate autoregressive modelling

In the case of multiple uncertain dynamic inputs, cross-correlation between these inputs may also need to be included. For example, $C_{\mathrm{COD,S}}$ and $C_{\mathrm{NH_4,S}}$ and their uncertainties are likely correlated. This can be done using a multivariate AR(1) model (Luetkepohl, 2005), which is a natural extension of the univariate AR(1) model, as follows:

$$\begin{aligned} &\vec{Y}(t+1) = \boldsymbol{\mu} + \mathbf{A} \cdot [\vec{Y}(t) - \boldsymbol{\mu}] + \boldsymbol{\varepsilon}(t), \\ &t = 1,\, 2,\, \ldots\, , T,\; Y_0 \sim \mathcal{N}(\boldsymbol{\mu},\, \boldsymbol{\Lambda}), \end{aligned} \tag{5.8}$$

where $\vec{Y}(t)$ is a vector of inputs at time t, $\mathbf{A}$ is a square matrix with parameters that define how the variables at time $t+1$ depend on those at time t, $\boldsymbol{\mu}$ is now a vector of means, and $\boldsymbol{\varepsilon}(t)$ is a vector of zero-mean normally distributed white noise processes. We further assume that the variance–covariance matrix $\mathbf{C}$ of $\boldsymbol{\varepsilon}(t)$ is time invariant. The initial value Y_0 is assumed normally distributed and uncorrelated ($\boldsymbol{\Lambda}$ is a diagonal matrix). In order to estimate the vector $\boldsymbol{\mu}$ and matrices $\mathbf{A}$ and $\mathbf{C}$, a sample of the variables of interest is needed. Parameter estimation is done by means of the R package mAr (Barbosa, 2015).

Input precipitation model

In case precipitation is selected as an uncertain input to be included in the uncertainty analysis, it too must be characterised by a pdf. Since precipitation, however, is not normally distributed and has many zeros, we cannot make the Gaussian assumption, and hence we cannot use the approach described in Sect. 5.2.5 to model its dynamic behaviour and uncertainty. In addition, we usually have precipitation measurements nearby so we need to condition the simulations to these measurements. Recall from Sect. 5.2.3 that, in the case study, precipitation data are recorded at stations Esch-sur-Sûre and Dahl.

Torres-Matallana et al. (2017b) present a model to simulate precipitation inside a target catchment given a known precipitation time series in a nearby location outside the catchment, while accounting for the uncertainty that is introduced due to spatial variation in precipitation. The method used for input precipitation uncertainty characterisation is essentially the same as the application of a Kalman filter/smoother (Kalman, 1960; Webster and Heuvelink, 2006). Calibration of the model requires precipitation time series at two locations near the catchment of interest. Once the model is calibrated, it is used to simulate precipitation inside the target catchment from a single precipitation time series nearby the catchment. Details of the calibration and conditional simulation are presented in Sect. S4.

5.2.6 Uncertainty analysis

We used a MC simulation (Hammersley and Handscomb, 1964; Kalos and Whitlock, 2008) to analyse how input uncertainty propagates through the EmiStatR model because it is flexible and straightforward to implement. It is also feasible in our case study because EmiStatR is a relatively simple model that does not involve a long computation time.

Monte Carlo simulation

The MC method runs the EmiStatR model repeatedly, each time using different model input values sampled from their pdf. The method thus consists of the following steps:

1. Repeat n times:
   (a) Generate a set of realisations of the uncertain model inputs at 10 min resolution.
   (b) Run the model at 10 min resolution and store output for this set of realisations. Later, in order to compute the summary statistics, a temporal aggregation of the model output to 1 h intervals is done.
2. Compute and store sample statistics from the n model outputs.

Here, n is the number of MC runs, i.e. the MC sample size. Common sample statistics that measure the uncertainty are the standard deviation and quantiles of the distribution of MC outputs, such as the difference between the 0.95 and 0.05 quantile, which can be easily calculated from the n MC outputs.

Sampling from the pdf of uncertain inputs was done using simple random sampling.

Monte Carlo output summary

Proper presentation of MC outputs is important for achieving the most from the experiment. Therefore, summary statistics are one important way to summarise the MC outputs. Commonly, a MC study yields n model outputs, which are stored in the MC result matrix $\mathbf{X}$ in Boos and Osborne (2015). From this matrix, various statistics can be computed. Basic summary statistics include the mean μ_{MC}, the standard deviation (σ_{MC}), and the variance σ^2_{MC}. From these, we can compute the coefficient of variation $\mathrm{CV}_{\mathrm{MC}}$ ($\sigma_{\mathrm{MC}}/\mu_{\mathrm{MC}}$), which is a dimensionless expression of relative uncertainty. The coefficient of variation is a standardised measure of the spread of a sampling distribution, which is useful because it allows a direct comparison of the variation in samples with different units or with very different means (Marwick and Krishnamoorthy, 2019). We computed estimates and standard errors for these statistics and also for the interquartile range ($\mathrm{IQR}_{\mathrm{MC}}$), 0.005 ($\zeta_{0.005}$) and 0.995 ($\zeta_{0.995}$) quantiles, and the 99 % width of the prediction band ($\zeta_{\mathrm{w},\,0.99}$).

Bootstrap computation for Monte Carlo summary

Following Boos and Osborne (2015), who argue that "Good statistical practice dictates that summaries in MC studies should always be accompanied by standard errors," we used the bootstrap method to compute the standard errors of all MC statistics. These tools are particularly relevant in a case without analytic solutions (Boos, 2003). According to Boos and Osborne (2015, p. 228) standard errors for MC output statistics are often not computed, which is an additional computational step on top of the overall analysis. Standard errors are straightforward to compute for simple statistics, such as the sample mean over the replications of the MC output, but are more difficult to compute for more complex statistics, such as medians, sample variances, and the classical Pearson measures of skewness and kurtosis. Therefore, to avoid burdensome computations, we opted to compute the standard errors by the bootstrap method. We briefly explain the bootstrap method below. For a detailed explanation, we refer to Efron (1979).

To compute the bootstrap variance of estimators, we follow the logic given by Boos and Osborne (2015). From a MC sample $Y_1, \ldots, Y_n$, we draw a random sample of size n, with replacement $Y_1^*, \ldots, Y_n^*$, and compute an estimator $\hat{\theta}$ of the MC statistic θ from this resample. We independently repeat this process B times, resulting in a sample of estimators $\hat{\theta}_1, \ldots, \hat{\theta}_B$. Then the bootstrap variance estimate, $\hat{V}_B$, is the sample variance of this sample of estimators as follows:

$$\hat{V}_B = \frac{1}{B-1} \sum_{i=1}^{B} \left(\hat{\theta}_i - \bar{\hat{\theta}} \right)^2, \tag{5.9}$$

where $\bar{\hat{\theta}}$ is the mean of the sample of estimators. The MC standard error, se, is simply the square root of the bootstrap variance.

We implemented, in stUPscales (Torres-Matallana et al., 2019a), specific routines for computing, by means of the bootstrap method, the MC estimators and their standard error for all MC statistics, where the variance of the model output is the most important. We compared our results with the results obtained using the Monte.Carlo.se R package (Boos et al., 2019).

Contributions of input variables to total uncertainty

A number of $m + 1$ MC analyses are needed to compute the contributions of input variables to total uncertainty, where m is the number of model input variables selected for uncertainty quantification. The first MC analysis, $\mathrm{MC}_{\mathrm{tot}}$, is done to compute the total output uncertainty by stochastically varying all input variables. The uncertainty associated with the first variable x_1 is quantified by a second MC analysis MC_1 in which only x_1 is equal to its deterministic value, while the other input variables vary stochastically.

Similarly, the other MC simulations MC_2, MC_3, ... , and MC_m are used to quantify the uncertainty for the variables x_2, x_3, ... , and x_m.

To quantify the contributions of individual input variables to the total uncertainty of the model inputs, the stochastic sensitivity S_i for each uncertain input x_i is computed. It is defined as follows:

$$S_i = 1 - \frac{\mathrm{Var}(\mathrm{MC}_i)}{\mathrm{Var}(\mathrm{MC}_{\mathrm{tot}})}. \tag{5.10}$$

The index represents the main effect contribution of each input factor to the total variance of the output. The larger the index, the more important the input uncertainty. We computed stochastic sensitivity indices per time step and aggregated contributions for the whole year. For plotting purpose, we aggregated the outputs from a 10 min time step to hourly time steps. The aggregation was done for each individual MC run before the contributions were computed.

5.2.7 Water quality impact

The results of the MC uncertainty propagation were also compared with the water quality standards. Standards are introduced to evaluate the impact of emissions of COD and NH_4 in CSOs in the receiving water. However, as Toffol (2006) recognises, although there are European emission standards for wastewater treatment plant effluent, standards for combined sewer overflow are not so clear. According to Steinel and Margane (2011), the European Water Framework Directive (WFD) is mainly concerned with the natural state of waters. Therefore, emission standards for effluent discharge are not set. The European Union Council Directive 91/271/EEC (1991) sets standards for COD and total nitrogen; hence, similar values have been adopted in many European member states. For more details about guidelines and design procedures in Europe, see Blumensaat et al. (2012). We assessed the emissions according to the German guideline ATV-DVWK-A 128 (1992), which is the standard for the dimensioning and design of storm water structures in combined sewers and commonly used in Luxembourg. The Austrian ÖWAV-RB 19 (2007) is also taken into account because it provides key reference guidelines for the design of urban water infrastructure in central Europe. Three main indicators are taken into account, namely hydraulic impact, COD concentration, and acute ammonium toxicity.

Hydraulic impact

According to the Austrian guidelines, and as summarised by Kleidorfer and Rauch (2011), the evaluation of the hydraulic impact is given by the following:

$$Q_1 \leq f_{\mathrm{h}} \cdot Q_{\mathrm{r1}}, \tag{5.11}$$

where $0.1 \leq f_{\mathrm{h}} \leq 0.5$, Q_1 ($\mathrm{L\,s^{-1}}$) is the maximum sewer overflow discharge with return period of 1 year, and Q_{r1} ($\mathrm{L\,s^{-1}}$) is the maximum water discharge in the river with return period of once per year. The factor f_{h} is taken as 0.1 in more sensitive streams, whereas it is 0.5 for streams with more stable bed and higher re-colonisation potential Toffol (2006). Time series of daily values recorded in 2006 to 2013 of the river Sûre at Heiderscheidergrund were used to compute the daily flow expected, with a return period of once per year (1.01 years), Q_{r1}.

We used the German guideline ATV-DVWK-A 128 (1992), which computes the throttle discharge at CSOs, $Q_{t,\,\mathrm{CSO}}$ ($\mathrm{L\,s^{-1}}$) as follows:

$$Q_{t,\,\mathrm{CSO}} = f_t \cdot A_{\mathrm{imp}}, \tag{5.12}$$

where $7.5 \leq f_t \leq 15$ and A_{imp} (ha) are the impervious areas connected to the combined sewer system. For the overflow flow MC output mean, the 0.95 and 0.995 quantile, we computed the exceedance percentage over the thresholds, which is calculated as the proportion of time steps exceeding the number of total time steps in the year (8759 time steps at 1 h).

COD concentration

Steinel and Margane (2011, Table 14) present the effluent standards for discharging into freshwater adopted in selected European countries. A COD concentration of $125\,\mathrm{mg\,L^{-1}}$ is reported for European Union countries. Austria has stricter rules, with a standard of $90\,\mathrm{mg\,L^{-1}}$ for populations between 50 and 500 inhabitants and $75\,\mathrm{mg\,L^{-1}}$ for populations greater than 500 inhabitants. The Goesdorf population was 1025 inhabitants by 2001 and was 1297 inhabitants by 2011 (Statec, 2020). For NH_4 a similar approach was used.

Acute ammonium toxicity

Following Kleidorfer and Rauch (2011), who argue that "the ammonia (NH_3) concentration depends on the ammonium (NH_4) concentration and on the dissociation equilibrium between NH_3 and NH_4 (which is influenced by temperature and pH-value)". According to Kleidorfer and Rauch (2011), the Austrian guideline ÖWAV-RB 19 (2007) establishes a maximum value of $2.5\,\mathrm{mg\,L^{-1}}$ for the ammonium (NH_4) concentration calculated for a 1 h duration for salmonid streams. For cyprinid streams, a maximum value of $5.0\,\mathrm{mg\,L^{-1}}$ is recommended.

5.3 Results

5.3.1 Selection of model inputs for uncertainty quantification

In this section, we assess the degree of uncertainty and sensitivity for all input variables, following the procedure described in Sect. 5.2.4. We summarise the results in Tables 5.1 and 5.2. To better support our decisions, we also include a graphical assessment of the degree of uncertainty and sensitivity of each input, as in Tscheikner-Gratl et al. (2017, see Fig. A.1 in the Appendix A).

Wastewater

Water consumption, qs, is a fairly uncertain input variable, and the model output is sensitive to this variable. Volume and flow of CSO are sensitive to changes in qs. Regarding water quality output, the total load of NH_4 is very sensitive to changes in qs. Pollution of sewage as COD load per capita per day, $C_{\mathrm{COD,S}}$, is the first selected input variable for propagation of uncertainty, due to the fact that it is both a very uncertain input variable and the model output (average and 99.9 percentile overflow COD concentration) is very sensitive to it. Pollution of sewage as NH_4 load per capita per day, $C_{\mathrm{NH_4,S}}$, is also included in the uncertainty propagation analysis. It is a very uncertain input variable, and the model output (overflow load and concentrations of NH_4) is very sensitive to it. The variables $C_{\mathrm{COD,S}}$ and $C_{\mathrm{NH_4,S}}$ are very uncertain because these are correlated to the temporal and spatial pattern of water consumption, which has a daily, weekly, and seasonal temporal variability.

Infiltration water

Inflow of infiltration water, q_{f}, is a very uncertain input variable because this inflow depends on the number of anomalies in the pipes (cracks or wrong connections) that allows infiltrations flowing into and out of the system. The distribution of these anomalies has a strong random component, and hence, q_{f} is very uncertain, and the model output is sensitive to it. Although this is a very uncertain input, the quick-scan analysis showed that model output sensitivity is not very high, as indicated in Table 5.1. For this reason, we did not include this variable in the uncertainty propagation analysis.

Pollution of infiltration water as COD load per capita per day, $\mathrm{COD_f}$, and pollution of infiltration water as NH_4 load per capita per day, $\mathrm{NH_{4f}}$, are not uncertain because in the Haute-Sûre study area the values of these variables are negligible.

Rainwater

Precipitation, P, is the main driving force of the model, and given the spatial variability of the rain fields, this input is considered very uncertain. The model output, additionally,

Table 5.1: Results of deterministic sensitivity analysis. Average percentage of change in model output caused by ±10 % change in model inputs (qs, $C_{\mathrm{COD,S}}$, $C_{\mathrm{NH_4,S}}$, $\mathrm{COD_r}$, pe, and P as time series; VAR(1) model for $C_{\mathrm{COD,S}}$ and $C_{\mathrm{NH_4,S}}$; and AR(1) model for $\mathrm{COD_r}$ and AR(1) conditioned for P; see Appendix A for nomenclature definitions). Output change greater than 15 % is considered very high. Variable C_{d} (not shown in the table) leads to a percentage of change of less than 0.3 %, while variables t_{fS} and $\mathrm{Lev_{ini}}$ (not shown in the table) lead to no change in the output. Values greater than 15 are shown in bold.

	Input variable														
Output variable	qs	$C_{\mathrm{COD,S}}$	$C_{\mathrm{NH_4,S}}$	q_{f}	$\mathrm{COD_r}$	$\mathrm{NH_{4r}}$	A_{total}	A_{imp}	C_{imp}	C_{per}	pe	$Q_{\mathrm{d,max}}$	V	D_{d}	P
V_{Chamber}	4.2	0.0	0.0	3.0	0.0	0.0	8.6	7.1	5.9	7.2	4.2	**16.7**	7.5	0.8	13.4
V_{Sv}	2.9	0.0	0.0	1.7	0.0	0.0	**19.6**	11.6	13.5	**16.1**	2.9	13.1	**16.5**	0.2	**17.8**
Q_{Sv}	0.6	0.0	0.0	0.2	0.0	0.0	2.7	1.5	1.1	1.8	0.6	13.1	14.8	0.6	12.4
$B_{\mathrm{COD,Sv}}$	2.9	2.4	0.0	1.7	7.7	0.0	**20.1**	11.8	13.7	**16.6**	5.3	**15.7**	14.5	0.2	**20.7**
$C_{\mathrm{COD,Sv,Av}}$	0.7	4.6	0.0	0.5	5.4	0.0	0.5	0.6	0.6	0.7	4.0	3.2	4.0	0.2	1.1
$C_{\mathrm{COD,Sv,99.9}}$	1.6	6.7	0.0	0.8	3.4	0.0	2.8	2.3	1.9	2.4	5.1	0.0	0.0	0.0	9.2
$C_{\mathrm{COD,Sv,Max}}$	1.6	6.7	0.0	0.8	3.4	0.0	2.8	2.3	1.9	2.4	5.1	0.0	0.0	0.0	9.2
$B_{\mathrm{NH_4,Sv}}$	3.1	0.0	3.3	1.8	0.0	6.7	**20.4**	12.0	13.8	**16.8**	6.4	**17.0**	13.4	0.3	**22.1**
$C_{\mathrm{NH_4,Sv,Av}}$	0.9	0.0	5.8	0.6	0.0	4.2	0.6	0.8	0.9	0.9	5.3	4.3	5.4	0.2	1.5
$C_{\mathrm{NH_4,Sv,99.9}}$	1.6	0.0	7.6	0.8	0.0	2.4	3.5	2.6	2.3	2.9	6.1	0.0	0.0	0.0	11.3
$C_{\mathrm{NH_4,Sv,Max}}$	1.6	0.0	7.6	0.8	0.0	2.4	3.5	2.6	2.3	2.9	6.1	0.0	0.0	0.0	11.3

is very sensitive to it. As a consequence, this input variable is treated as the third input variable in the uncertainty propagation analysis. Pollution of run-off as COD concentration, $\mathrm{COD_r}$, is the fourth input variable considered in the uncertainty propagation, given that it is a very uncertain and very sensitive input variable, particularly to the load and concentration of COD in the overflow.

Pollution in rainwater as $\mathrm{NH_4}$ concentration, $\mathrm{NH_{4r}}$, is considered fairly uncertain. The model output (overflow load and concentration of $\mathrm{NH_4}$) is very sensitive to it. Although model output is very sensitive to this model input variable, model input uncertainty is not very high, as indicated in Table 5.1. For this reason, it was not included in the uncertainty propagation analysis.

Sub-catchment

The model is very sensitive to the total area A_{total} and to the run-off coefficient for the pervious area (C_{per}) and sensitive to the impervious area A_{imp} and to the run-off coefficient for impervious area (C_{imp}). However, we did not include A_{total} in the uncertainty analysis because it can be fairly accurately derived from spatial databases, and hence, their uncertainty is not large. Although model output is very sensitive to the input variable C_{per}, the uncertainty about this variable is not very high, as indicated in Table 5.1. The reason behind this is that C_{per} can be derived fairly accurately from geographic information system (GIS) products, such as land use and soil type maps. Therefore, we did not include this variable in the uncertainty propagation analysis. The population equivalents, pe, is a sensitive variable but not very uncertain. Hence, this variable was not included

in the uncertainty analysis. The theoretical largest flow time in the catchment t_{fS} is not uncertain and not sensitive.

CSO structure

Although model output is very sensitive to maximum throttled outflow $Q_{\mathrm{d,\,max}}$ and volume V, these are not included in the uncertainty analysis because their values are accurately known. The same is true for the variables curve level–volume lev2vol, orifice diameter D_{d}, and discharge coefficient C_{d}. These variables are accurately known and, therefore, not considered as being uncertain variables. The initial water level in the chamber $\mathrm{Lev_{ini}}$ is very uncertain, but the model output is not sensitive to this variable. Therefore, $\mathrm{Lev_{ini}}$ was not included in the uncertainty analysis.

5.3.2 Uncertainty quantification of selected model input

After ranking all inputs on levels of uncertainty and model sensitivity, we selected four input variables to be included in the uncertainty analysis. These are $C_{\mathrm{COD,\,S}}$, $C_{\mathrm{NH_4,\,S}}$, $\mathrm{COD_r}$, and P (Table 5.2).

Sewage per capita COD and Ammonium

The fit of pdf's for the two uncertain inputs, $C_{\mathrm{COD,\,S}}$ and $C_{\mathrm{NH_4,\,S}}$, was based on measurements under dry weather flow conditions. Measurement campaigns were done in Goesdorf from 28 April to 24 June 2011, in Kaundorf from 22 June to 18 August in 2010 and from 20 July to 5 August in 2011, and in Nocher-Route from 18 November 2010 to 27 April 2011. Samples of COD and NH_4 in $\mathrm{mg\,L^{-1}}$ (91 in total for each variable) were analysed. An average wastewater amount was calculated for Goesdorf ($153\,\mathrm{L\,PE^{-1}\,d^{-1}}$), Kaundorf ($112\,\mathrm{L\,PE^{-1}\,d^{-1}}$), and Nocher-Route ($94.3\,\mathrm{L\,PE^{-1}\,d^{-1}}$). Table 5.3 presents summary statistics of the dry weather flow measurements of COD and NH_4 and the corresponding value of $C_{\mathrm{COD,\,S}}$ and $C_{\mathrm{NH_4,\,S}}$. COD is converted to $C_{\mathrm{COD,\,S}}$ by means of a simple conversion from $\mathrm{mg\,L^{-1}}$ to $\mathrm{g\,PE^{-1}\,d^{-1}}$, by multiplying COD by the measured per capita flow ($112\,\mathrm{L\,PE^{-1}\,d^{-1}}$) and dividing by 1000. NH_4 was converted to $C_{\mathrm{NH_4,\,S}}$ in a similar way.

Closer inspection showed that $C_{\mathrm{COD,\,S}}$ and $C_{\mathrm{NH_4,\,S}}$ observations are best characterised by a lognormal distribution (Sect. S5). Since $C_{\mathrm{COD,\,S}}$ and $C_{\mathrm{NH_4,\,S}}$ are dynamic and cross-correlated, we calibrated a bivariate AR(1) model with state vector $\vec{Y} = [\log(C_{\mathrm{COD,\,S}})\ \log(C_{\mathrm{NH_4,\,S}})]^T$. The estimated parameters of the model, using the methodology described in Sect. 5.2.5, are as follows:

Table 5.2: Input variables of the EmiStatR model and selection of inputs for uncertainty analysis based on input uncertainty level and model sensitivity level (Note: ++ is very uncertain/sensitive; – – is not uncertain/sensitive).

Input variable	Input uncertainty	Model sensitivity	Uncertainty analysis
Wastewater			
1. qs	+	+	No
2. $C_{\mathrm{COD,S}}$	++	++	Yes
3. $C_{\mathrm{NH_4,S}}$	++	++	Yes
Infiltration water			
4. q_{f}	++	+	No
5. $\mathrm{COD_f}$	– –	– –	No
6. $\mathrm{NH_{4f}}$	– –	– –	No
Rainwater			
7. P	++	++	Yes
8. $\mathrm{COD_r}$	++	++	Yes
9. $\mathrm{NH_{4r}}$	+	++	No
Sub-catchment			
10. A_{total}	+	++	No
11. A_{imp}	+	+	No
12.C_{imp}	+	+	No
13. C_{per}	+	++	No
14. pe	+	+	No
15. t_{fS}	–	– –	No
CSO structure			
16. $Q_{\mathrm{d,max}}$	–	++	No
17. V	–	++	No
18. D_{d}	– –	– –	No
19. C_{d}	– –	– –	No
20. $\mathrm{Lev_{ini}}$	++	– –	No

Table 5.3: Summary statistics of dry weather flow measurements for $C_{\mathrm{COD,S}}$ and $C_{\mathrm{NH_4,S}}$ characterisation.

	COD ($\mathrm{mg\,L^{-1}}$)	$C_{\mathrm{COD,S}}$ ($\mathrm{g\,PE^{-1}\,d^{-1}}$)	$\log(C_{\mathrm{COD,S}})$ $\log(\mathrm{g\,PE^{-1}\,d^{-1}})$	NH_4 ($\mathrm{mg\,L^{-1}}$)	$C_{\mathrm{NH_4,S}}$ ($\mathrm{g\,PE^{-1}\,d^{-1}}$)	$\log(C_{\mathrm{NH_4,S}})$ $\log(\mathrm{g\,PE^{-1}\,d^{-1}})$
Min	61.9	6.9	1.936	16.10	1.745	0.556
P5	216.8	23.8	3.167	20.55	2.102	3.018
Mean	925.5	104.2	4.378	44.38	4.733	1.473
P95	2032.0	236.8	5.466	79.00	7.684	2.039
Max	3454.0	528.5	6.270	81.20	10.771	2.377
SD	631.7	87.5	0.751	18.56	1.917	0.410

$$\boldsymbol{\mu} = \begin{bmatrix} 4.40947 \\ 3.70411 \end{bmatrix} \quad \mathbf{A} = \begin{bmatrix} 0.99165 & -0.00319 \\ -0.00009 & 0.99455 \end{bmatrix}$$
$$\mathbf{C} = \begin{bmatrix} 0.00913 & 0.00224 \\ 0.00224 & 0.00185 \end{bmatrix}. \tag{5.13}$$

The defined multivariate autoregressive model also captures the dynamic behaviour, temporal correlation, and cross-correlation of the input variables, deriving the probability distributions of $C_{\mathrm{COD,S}}$ and $C_{\mathrm{NH_4,S}}$ from measurements in the Haute-Sûre catchment, which agreed well with values reported in the literature (Katukiza et al., 2014; Heip et al., 1997).

Run-off COD concentration

Regarding $\mathrm{COD_r}$, due to the fact that no field measurements were available, expert judgement and reference values from the literature were the basis for characterising the pdf of this input variable. The variable was assumed to be lognormally distributed with a mean value of $71\,\mathrm{mg\,L^{-1}}$. Although, House et al. (1993) and Welker (2008) reported a higher value, namely $107\,\mathrm{mg\,L^{-1}}$ for $\mathrm{COD_r}$, we selected a lower value due to the specific characteristics of the CSO system in the Haute-Sûre catchment. The value of $150\,\mathrm{mg\,L^{-1}}$ as standard deviation of $\mathrm{COD_r}$ leads to a coefficient of variation ($\mathrm{SD} \cdot \mathrm{mean}^{-1}$) equal to 2.11, which is greater than the coefficient of variation for $C_{\mathrm{COD,S}}$ (0.84). We allow the standard deviation of $\mathrm{COD_r}$ to be greater than the standard deviation of $C_{\mathrm{COD,S}}$ because COD measurements in rain water are very uncertain.

Input precipitation model

Precipitation and its associated uncertainty was modelled as an autoregressive model conditioned to the observed precipitation at a nearby measurement station. We assumed a multivariate lognormal distribution and included a temporal correlation of the simulated

time series. Calibration of the precipitation model is done with the mAr package, as explained in Sect 5.2.5, and using a 10 min precipitation time series of stations Esch-sur-Sûre and Dahl for 2010. Upon calibration of the multivariate autoregressive model, we proceeded with the conditional simulation of Y_c (Sect. S4.2, Eq. 5). For this, we computed the parameters of the model, as shown in Eq. (5.14). The model parameters are given by (Torres-Matallana et al., 2017b) the following:

$$\begin{array}{l} \mu_1 = 2.85501 \\ \mu_\delta = 0.10194 \\ \sigma_1^2 = 0.07241 \\ \sigma_\delta^2 = 0.07951 \\ \rho_{1\delta} = -0.03876. \end{array} \quad \mathbf{B} = \begin{bmatrix} 0.95650 & 0.03980 \\ 0.02429 & 0.88304 \end{bmatrix} \tag{5.14}$$

Next, we generated conditional simulations of the 10 min precipitation for 2010 for each subcatchment using the approach described in Sect. 5.2.5. Note that this involves simulating log-transformed precipitation, which can easily be transformed to precipitation data using the antilog. The simulation procedure was repeated as many times as simulated precipitation time series were required for the MC uncertainty propagation analysis.

The simulated precipitation time series captured the main statistics of the observed time series well. The reader can find evidence for this in the Appendix A (Table A.4 and Fig. A.2). Despite the satisfactory performance of the proposed method, some cases showed an overestimation of the simulated precipitation, mainly due to high values of the ratio of the multiplicative factor $\delta(t)$. This behaviour was also recognised by McMillan et al. (2011), who stated that the multiplicative factor used in their study "does not capture the distribution tails, especially during heavy precipitation where input errors would have important consequences for run-off prediction".

5.3.3 Uncertainty analysis

Model output sensitivity and the degree of uncertainties evaluation of each model input helped to define the four input variables included in the uncertainty analysis, namely $C_{\mathrm{COD,S}}$, $C_{\mathrm{NH_4,S}}$, $\mathrm{COD_r}$, and P. In this section, we present the results of the uncertainty propagation for these four selected input variables to the model output, both for water quantity (volume in the combined sewer overflow chamber, CSOC, and overflow volume and flow) and for water quality (loads and concentrations of chemical oxygen demand, COD, and ammonium, NH_4).

Monte Carlo output and uncertainty quantification

The seven output variables from EmiStatR were analysed by MC input uncertainty propagation. A detailed description of the Monte Carlo simulation size and timing is presented

Table 5.4: Monte Carlo estimated statistics and standard errors (se) determined by bootstrapping for the MC simulations, where all selected input variables are uncertain R (model was run at 10 min time steps, and MC results were aggregated to 1 h averages over a 1 year period). See Appendix A for output variable nomenclature and units. Note: Interq. – interquartile; quant. – quantile; pbw – prediction band width; SD – standard deviation; var – variance; cv – coefficient of variation.

	Mean	Interq. range	0.005 quant.	0.995 quant.	99 % pbw	0.05 quant.	0.95 quant.	90 % pbw	SD	var	cv
	μ_{MC}	IQR	$\zeta_{0.005}$	$\zeta_{0.995}$	$\zeta_{w,0.99}$	$\zeta_{0.05}$	$\zeta_{0.95}$	$\zeta_{w,0.90}$	σ_{MC}	σ^2_{MC}	CV_{MC}
$V_{Chamber}$	92.51	13.65	77.55	109.16	31.60	81.02	104.28	23.25	8.27	2984	0.100
se	2.53	1.17	2.18	3.17	1.48	2.27	2.95	1.36	0.56	482	0.001
V_{Sv}	3.18	3.69	0.85	6.60	5.76	0.94	5.76	4.82	1.98	1100	0.070
se	0.51	0.73	0.24	0.95	0.95	0.26	0.91	0.83	0.35	259	0.012
Q_{Sv}	55.49	76.77	0.37	267.1	266.7	1.24	165.3	164.1	64.50	7332	1.585
se	5.67	10.26	0.13	13.75	14.54	0.22	10.82	10.89	4.57	1,102	0.048
$B_{COD,Sv}$	1.18	1.69	0.04	6.11	6.06	0.07	3.49	3.41	1.27	394	0.087
se	0.20	0.31	0.02	1.02	1.00	0.03	0.59	0.59	0.21	81.71	0.013
$B_{NH4,Sv}$	0.052	0.077	0.004	0.174	0.170	0.006	0.125	0.120	0.045	0.546	0.075
se	0.009	0.014	0.002	0.029	0.028	0.002	0.022	0.021	0.008	0.115	0.013
$C_{COD,Sv,av}$	170.0	164.6	3.80	909.7	905.9	15.49	465.8	450.3	161.9	36 151	0.988
se	9.02	11.38	0.33	40.12	40.40	1.02	24.66	24.03	7.94	4,615	0.016
$C_{NH4,Sv,av}$	7.19	6.65	0.47	29.20	28.74	0.86	17.51	16.64	5.66	46.93	0.815
se	0.41	0.61	0.02	1.23	1.22	0.06	0.91	0.87	0.29	6.62	0.016

in the Appendix A (Sect. A.6).

Figure 5.3 illustrates the uncertainty propagation outcomes for the first Monte Carlo simulation, in which all input variables vary stochastically. The MC simulations were performed for the entire year of 2010 at 10 min time steps, which were aggregated to hourly time steps in the figure. The aggregation function was used for precipitation, CSO chamber volume, CSO spill volume, and loads was the sum, whereas for CSO spill flow and concentrations the aggregation function was the mean. The top of Fig. 5.3 shows input precipitation as main driving input. For illustration purposes, two events with a 2 d duration each are shown. The first event occurred in spring (May 2010) and the second in autumn (September 2010), and this shows that the uncertainty is high when there is a high precipitation event. The more intense the precipitation input, as seen in Fig. 5.3 in the top left (May event), the greater the uncertainty bandwidth for overflow flow and for COD and NH_4 loads (Fig. 5.3e–h) and concentrations (Fig. 5.3i–l). The MC-estimated statistics and the standard errors (se) are presented in Table 5.4. The table shows the uncertainty quantification of outputs obtained from the MC uncertainty propagation for the first MC simulation (all selected input variables are uncertain).

Table 5.4 shows the standard deviation (SD) and the coefficient of variation (cv) for the seven output variables considered in the uncertainty propagation. For the volume in the CSO chamber, $V_{Chamber}$, the annual mean standard deviation, σ_{MC}, (8.27 m^3) is lower than the mean, μ_{MC}, (92.51 m^3). This goes along with an annual mean coefficient of variation (CV_{MC}) of 0.100. A (CV_{MC}) greater than one means large uncertainty. The overflow spill volume, V_{Sv}, had a coefficient of variation of 0.070, while it was 1.585 for the overflow

flow, Q_{Sv}. This shows that the relative uncertainty of the overflow flow is very large. Regarding the overflow COD load, the annual mean (1.18 kg) is similar as the annual mean standard deviation (1.27 kg). Similar behaviour was observed for the overflow COD concentration, which had an annual mean value of 170 mg L^{-1} and a standard deviation of 162 mg L^{-1}. For overflow NH_4 load and overflow NH_4 concentration, the annual mean also had the same order of magnitude as the annual mean standard deviation. Overflow COD and NH_4 loads had a coefficient of variation of 0.087 and 0.075, respectively, whereas the coefficient of variation for concentrations were 0.988 and 0.815, respectively. This suggests that overflow concentrations are more uncertain.

Low standard errors (se) for the coefficient of variation were obtained for all seven outputs. These were never greater than 0.05, which indicates that the selected MC replication size (1500 for mc1) is a suitable value. This holds for all output statistics because, in all cases, the standard error is small in comparison to the estimated value.

Contributions of input variables to total uncertainty

The contributions of input variables to the total uncertainty of the model inputs were also computed using the procedure described in Sect. 5.2.6. A total of four MC simulations with a total of 6000 runs were performed to estimate S_i (Eq. 5.10). Afterwards, four contributions were evaluated per time step and aggregated for the whole year. Following Eq. (5.10), the per time step contribution of input variables to output variables in terms of percentage of variance, stochastic sensitivity S_i of the input variables $C_{\mathrm{COD,S}}$, $C_{\mathrm{NH_4,S}}$, $\mathrm{COD_r}$, and P were calculated. An example of the contributions analysis per time step is presented in Fig. 5.4. Here we remark that a high uncertainty over time is shown mainly for the spring event.

The aggregated-over-time contributions of input variables to output variables in terms of percentage of variance and stochastic sensitivity S_i of the input variables were also calculated (Table 5.5). Note that P is the only source of uncertainty for V_{Chamber} and V_{Sv}, while uncertainty in NH_4 inputs only propagates to NH_4 outputs, which is similar for COD (Fig. 5.4).

We found, as expected, that precipitation, P, is the only source of uncertainty from all uncertain input considered for water quantity output variables V_{Chamber} and V_{Sv}. Regarding average values for the whole year, for the water quality output variables $B_{\mathrm{COD,Sv}}$ and $C_{\mathrm{COD,Sv,av}}$, $C_{\mathrm{COD,S}}$ has the largest contribution to the output variance, which is about 66 % for $B_{\mathrm{COD,Sv}}$ and about 62 % for $C_{\mathrm{COD,Sv,av}}$. The second variable that contributes to the uncertainty of these COD output variables is P, with about 3 % for $B_{\mathrm{COD,Sv}}$ and 9 % for $C_{\mathrm{COD,Sv,av}}$. Similarly, the input variable $C_{\mathrm{NH_4,S}}$ plays an important role in the contribution of total uncertainty for $B_{\mathrm{NH_4,Sv}}$ (on average about 34 % of the variance for the whole year) and $C_{\mathrm{NH_4,Sv,av}}$ (about 35 %). Equally contributing to the uncertainty of these NH_4 output variables is P with about 66 % for $B_{\mathrm{NH_4,Sv}}$ and 65 % for $C_{\mathrm{NH_4,Sv,av}}$.

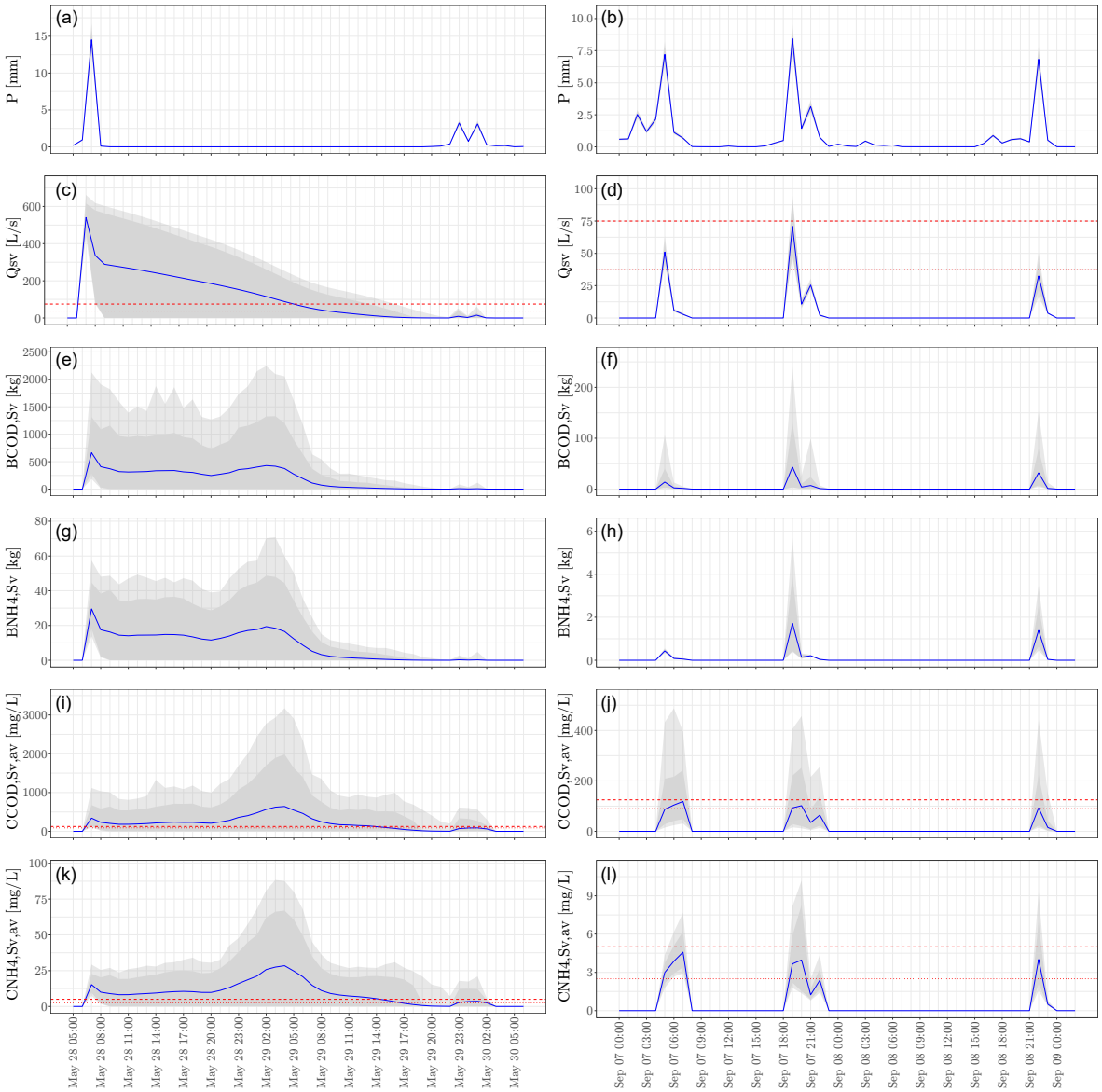


Figure 5.3: Uncertainty propagation outcomes for the first Monte Carlo simulation, where all input variables vary stochastically. The 99 % prediction interval is shown as a light grey shading, the 90 % prediction interval as a dark grey shading, and the mean value as a blue line. The MC simulations were performed for the entire year of 2010 at a 10 min time step, aggregated to hourly time steps in the figure. **(a, b)** Input precipitation. **(c, d)** In the overflow spill flow, the upper dashed red line indicates the $75\,\mathrm{L\,s^{-1}}$ threshold and the lower dotted red line the $37.5\,\mathrm{L\,s^{-1}}$ threshold. **(e, f)** Load of overflow COD. **(g, h)** Load of overflow NH_4. **(i, j)** In the average spill COD concentration, the upper dashed red line indicates the $125\,\mathrm{mg\,L^{-1}}$ threshold, and the lower dotted red line indicates the $90\,\mathrm{mg\,L^{-1}}$ threshold. **(k, l)** In the average spill NH_4 concentration, the upper dashed red line indicates the $5.0\,\mathrm{mg\,L^{-1}}$ threshold and lower dotted red line the $2.5\,\mathrm{mg\,L^{-1}}$ threshold.

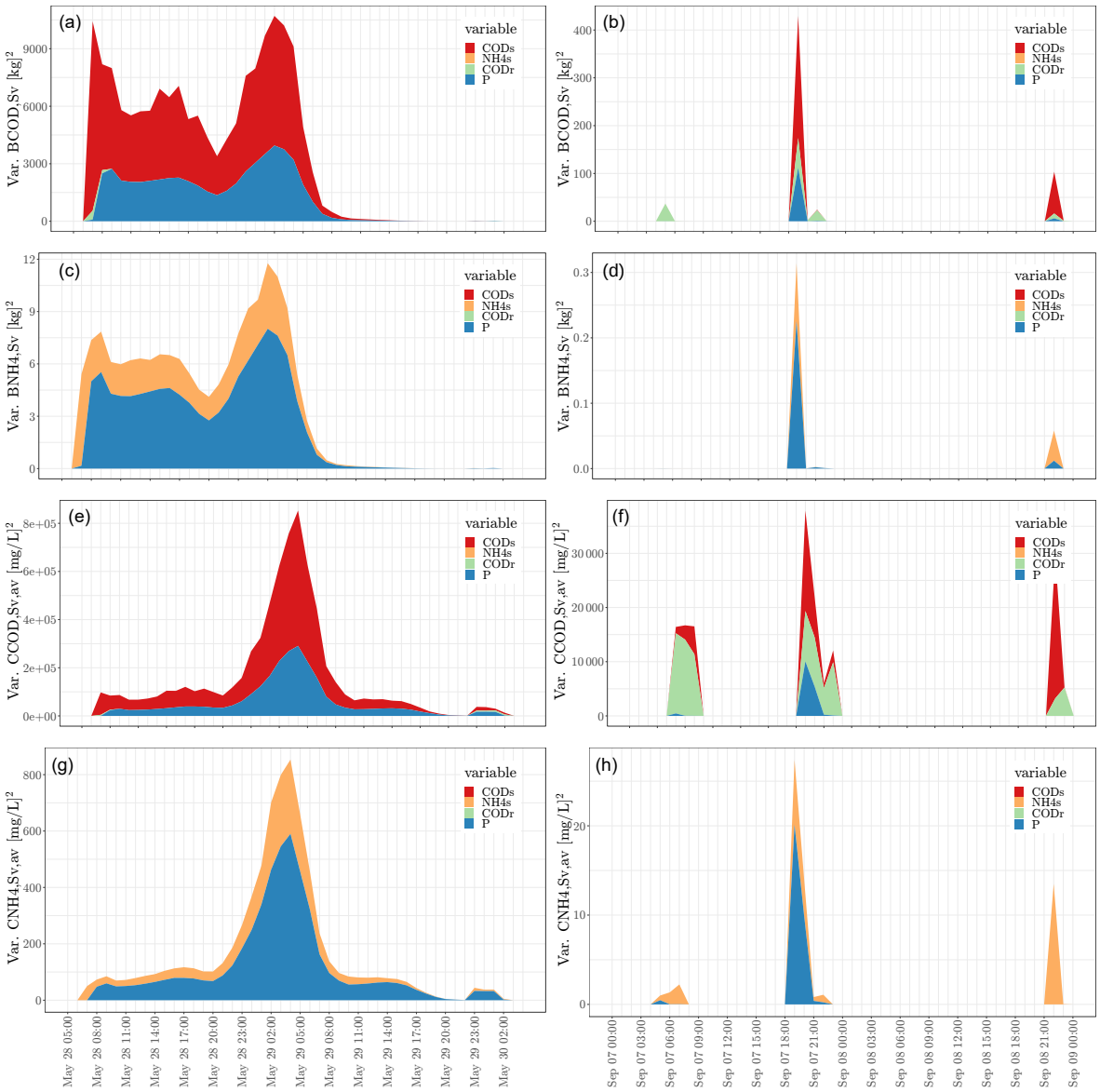


Figure 5.4: **(a, b)** Temporal contributions of input variables to the load of overflow COD, **(c, d)** load of overflow NH_4, **(e, f)** concentration of overflow COD, and **(g, h)** concentration of overflow NH_4 in terms of variance. The MC simulations were performed for the entire year of 2010 at a 10 min time step, which were aggregated to hourly time steps. For illustration, two periods are shown from 28 to 30 May 2010 (left) and from 7 to 9 September 2010 (right). In the legend, CODs refers to $C_{\mathrm{COD,S}}$, and NH4s refers to $C_{\mathrm{NH_4,S}}$.

Table 5.5: Aggregated-over-time contribution of input variables to output variables in terms of percentage of total variance.

	Stochastic sensitivity, S_i, of input variable (%)				
Output variable	Total	$C_{\mathrm{COD,S}}$	$C_{\mathrm{NH_4,S}}$	$\mathrm{COD_r}$	P
V_{Chamber}	100.0	0.0	0.0	0.0	100.0
V_{Sv}	100.0	0.0	0.0	0.0	100.0
$B_{\mathrm{COD,Sv}}$	100.0	65.7	0.0	2.9	31.4
$C_{\mathrm{COD,Sv,av}}$	100.0	62.4	0.0	8.7	28.9
$B_{\mathrm{NH_4,Sv}}$	100.0	0.0	34.4	0.0	65.6
$C_{\mathrm{NH_4,Sv,av}}$	100.0	0.0	35.3	0.0	64.7

From these results, we can infer that precipitation is a main source of uncertainty for all six outputs considered.

5.3.4 Uncertainty and water quality impact

Quantification and assessment of the water quality impact is an important step after the uncertainty propagation. As described in Sect. 5.2.7, the assessment of water quality standards was done by taking into account the reference thresholds recommended in the European Union guidelines for COD and the German and Austrian guidelines for hydraulic impact and acute ammonium toxicity.

Hydraulic impact

From the time series of daily values for 2006 to 2013 of the river Sûre, a daily flow expected with a return period of once per year (1.01 years), Q_{r1} of $16\,\mathrm{m^3\,s^{-1}}$ was computed at Heiderscheidergrund, which corresponds with the entire catchment area of the Haute-Sûre storm water system (182.1 ha). Therefore, we estimated the river daily flow in the Goesdorf CSO structure as a proportion to 30 ha, which is equal to $2.6\,\mathrm{m^3\,s^{-1}}$. Following Eq. (5.11), the maximum sewer overflow discharge with return period of 1 year, Q_1 can have a value between 0.26 and $1.32\,\mathrm{m^3\,s^{-1}}$. Accordingly, with the German guideline ATV-DVWK-A 128 (1992; Eq. 5.12), two additional thresholds are defined for the maximum sewer overflow discharge with return period of 1 year for the Goesdorf catchment ($A_{\mathrm{imp}} = 5.0\,\mathrm{ha}$). Q_1 is expected to vary between 37.5 and $75.0\,\mathrm{L\,s^{-1}}$. We contrasted these values with those obtained from the uncertainty analysis. From Table 5.4, we obtained a 1 h mean value for the overflow spill flow, Q_{Sv}, of $55.5\,\mathrm{L\,s^{-1}}$, 90 % prediction band width of $164.1\,\mathrm{L\,s^{-1}}$, and standard deviation of $64.5\,\mathrm{L\,s^{-1}}$. Figure 5.3a and b present the overflow spill flow for the two periods chosen for illustration. The upper dashed red line indicates the $75\,\mathrm{L\,s^{-1}}$ threshold, and the lower dotted red line indicates the $37.5\,\mathrm{L\,s^{-1}}$ threshold. Table 5.6 (top) shows the exceedance percentage of overflow spill flow over the 37.5 and $75.0\,\mathrm{L\,s^{-1}}$ thresholds for the mean, 0.95 quantile, and 0.995 quantile. We found a 0.49 % exceedance of the mean value over the $37.5\,\mathrm{L\,s^{-1}}$ threshold and about 1.7 %

for the quantiles. As expected, slightly lower percentages were found for the $75.0\,\mathrm{L\,s^{-1}}$ threshold.

COD concentration

A reference COD concentration emission in CSOs was presented in Sect. 5.2.7. For the European Union, a value of $125\,\mathrm{mg\,L^{-1}}$ is used. We obtained a 1 h average spill COD concentration with a mean of $170\,\mathrm{mg\,L^{-1}}$, standard deviation of $162\,\mathrm{mg\,L^{-1}}$, and a 90 % prediction band width of $450\,\mathrm{mg\,L^{-1}}$. Figure 5.3i and j present the average spill COD concentration. The upper dashed red line indicates the $125\,\mathrm{mg\,L^{-1}}$ threshold and the lower dotted red line the $90\,\mathrm{mg\,L^{-1}}$ threshold. The mean COD concentration in the overflow volume was higher than the thresholds. However, note that when entering the river system it will quickly be diluted, suggesting that the negative impact on the environment will be dampened by the receiving water body.

Table 5.6 (centre row) shows the exceedance percentage of overflow COD concentration over the 90 and $125\,\mathrm{mg\,L^{-1}}$ thresholds for the mean, 0.95 quantile, and 0.995 quantile. We found a 1.62 % exceedance of the mean value over the $90\,\mathrm{mg\,L^{-1}}$ threshold and about 1.8 % for the quantiles. Slightly lower percentages were found for the $125\,\mathrm{mg\,L^{-1}}$ threshold for the mean value (1.03 %). For the quantiles equal values were found as for the $90\,\mathrm{mg\,L^{-1}}$ threshold.

Acute ammonium toxicity

We compared the acute ammonium toxicity reference values presented in Sect. 5.2.7 ($2.5\,\mathrm{mg\,L^{-1}}$ for the ammonium concentration calculated for a 1 h duration for salmonid streams, and for cyprinid streams, a maximum value of $5.0\,\mathrm{mg\,L^{-1}}$) was calculated with the values we found for ammonium. An average spill NH_4 concentration with a mean of $7.19\,\mathrm{mg\,L^{-1}}$, standard deviation of $5.66\,\mathrm{mg\,L^{-1}}$, and 90 % prediction band width of $16.64\,\mathrm{mg\,L^{-1}}$ was obtained. Figure 5.3k and l show the average spill NH_4 concentration for the two periods chosen for illustration. The ammonium (NH_4) concentrations in the overflow flow are higher than the reference values, which are given for concentrations in the river.

Table 5.6 (bottom row) shows the exceedance percentage of overflow NH_4 concentration over the 2.5 and $5.0\,\mathrm{mg\,L^{-1}}$ thresholds for the mean, 0.95 quantile, and 0.995 quantile. We found a 1.8 % exceedance of the mean and quantile values over the 2.5 and $5.0\,\mathrm{mg\,L^{-1}}$ thresholds. A slightly lower percentage (1.1 %) was found for the $5.0\,\mathrm{mg\,L^{-1}}$ threshold with regards to mean value.

Table 5.6: Frequency (percentage) over time that environmental thresholds are exceeded for different statistics of the overflow spill flow, COD, and NH_4 concentration.

Output variable	Threshold	Statistic	Exceedance percentage
Q_{Sv}	37.5	Mean	0.49
$(\mathrm{L\,s^{-1}})$	37.5	0.95 quantile	1.71
	37.5	0.995 quantile	1.74
	75.0	Mean	0.31
	75.0	0.95 quantile	1.51
	75.0	0.995 quantile	1.72
$C_{\mathrm{COD,\,Sv,\,av}}$	90.0	Mean	1.62
$(\mathrm{mg\,L^{-1}})$	90.0	0.95 quantile	1.80
	90.0	0.995 quantile	1.82
	125.0	Mean	1.03
	125.0	0.95 quantile	1.80
	125.0	0.995 quantile	1.82
$C_{\mathrm{NH_4,\,Sv,\,av}}$	2.5	Mean	1.78
$(\mathrm{mg\,L^{-1}})$	2.5	0.95 quantile	1.80
	2.5	0.995 quantile	1.82
	5.0	Mean	1.05
	5.0	0.95 quantile	1.78
	5.0	0.995 quantile	1.82

5.4 Discussion

This study aimed to select and characterise the main sources of input uncertainty in urban water systems, while accounting for temporal auto- and cross-correlation of uncertain model inputs, by propagating input uncertainty through the EmiStatR model and quantifying and assessing the contributions of each uncertainty source to model output uncertainty dynamically (over time). In the following discussion, we start with the uncertainty and water quality impact of the model outputs to the environment, in relation to the uncertainty analysis. Next, we discuss the accuracy of Monte Carlo analysis, followed by a discussion of other sources of uncertainty. Finally, we highlight some limitations and possible solutions to the approach used in this work.

5.4.1 Uncertainty and water quality impact

Next, we discuss how the uncertainty propagation analysis done gives additional insight regarding hydraulics, COD concentration, and the acute ammonium toxicity impact on water quality over the river Sûre due to the CSO discharges under study. After doing the uncertainty propagation analysis, we not only have the predictions of model outputs, but we also know how uncertain these are. An added value arises when we take into account the uncertainty information. For the case of the overflow spill flow, the expected model output (mean of $55.5\,\mathrm{L\,s^{-1}}$) is below the environmental threshold of $75\,\mathrm{L\,s^{-1}}$, but the 0.95 quantile ($164.1\,\mathrm{L\,s^{-1}}$) is very much above the threshold. This indicates that there is a considerable chance of it being above the threshold.

Regarding water quality outputs, although the mean model output for COD and NH_4 concentrations is fairly above that of the thresholds, the 0.95 quantile is 2.7 times above the mean value for COD concentration and 2.4 times above the mean value for NH_4. Also, here we can conclude that we are not certain that we are below the threshold because there is a considerable probability that the true values are above, even though the expected value is below the threshold.

We were able to compute the water quantity and quality at the CSO outlet to the river. We found that water quality (COD and NH_4) were sometimes above the environmental threshold. Even if the expected value was below the threshold, there could still be a considerable probability that the quality was above the threshold because of the large uncertainty. Therefore, policy and decision-makers and water managers need to be aware of this because, whenever concentrations are above the threshold, this may harm the environment. Nevertheless it is worth noting that we computed the concentration in the outlet of the CSO. When this spilled water enters the river, it will quickly mix with the much cleaner river water and concentrations will drop quickly, so it is only a local problem. How local it is and how the river water quality is distributed in space and time is not an easy problem to solve and requires the use of hydrological and hydraulic river models, e.g.

SIMBA (IFAK, 2007) or MIKE 11 (DHI, 2017). Those models have been well developed and, for some of them, uncertainty analyses have also been done (Beven and Binley, 1992; Refsgaard, 1997; Beven and Freer, 2001; Vrugt et al., 2003b, 2008; Beven et al., 2010; Andrés-Doménech et al., 2010; Beven, 2012; Jerves-Cobo et al., 2020; Yu et al., 2020), but obviously such uncertainty analyses can only be done if the inputs and the uncertainty associated with these inputs to these models are known. One of these inputs is the inlet from CSO. That is where this chapter makes a very valuable contribution because our work has quantified water quantity and quality of CSO structures, including uncertainty, and that is exactly what these river models need to be able to do an uncertainty propagation analysis. We also recognise other attempts at quantity (e.g. Sriwastava et al., 2018) and quality, especially with measurements taken at CSOs, which demonstrate that the measured water quality at the WWTP influent is expected to render a low representativity of the conditions at the CSOs (e.g. Brombach et al., 2005; Diaz-Fierros T et al., 2002). We present some comparisons with these studies in the following lines.

Sriwastava et al. (2018) apply an uncertainty propagation to a complex hydrodynamic model to quantify uncertainty in sewer overflow volume. They used MC for the uncertainty propagation and Latin hypercube sampling (LHS) as an efficient sampling scheme. Although LHS ensures a full coverage of the sample space and provides a faster convergence than simple random sampling, the LHS application in the case of dynamic model inputs (e.g. precipitation, COD, and NH_4 inputs) is not trivial, and its implementation is more complex than in the case of sampling from static variables (i.e. uncertain constants). In our study, we sampled the time series of dynamic inputs using an implementation in stUPscales (Torres-Matallana et al., 2019a, 2018b).

Diaz-Fierros T et al. (2002), in a study in the city of Santiago de Compostela (northwestern Spain; population about 100 000 inhabitants), where a combined sewer system feeds to a grossly undersized wastewater treatment plant, reported an event mean concentration (Diaz-Fierros T et al., 2002; Table 4) for the output variables $C_{\mathrm{COD,\,Sv,\,av}}$ and $C_{\mathrm{NH_4,\,Sv,\,av}}$ of 329.1 and 8.7 $\mathrm{mg\,L^{-1}}$, respectively. These values are larger than those found by Brombach et al. (2005) and are more in agreement with our findings, especially for the case of $C_{\mathrm{NH_4,\,Sv,\,av}}$. Diaz-Fierros T et al. (2002) reported values of $C_{\mathrm{COD,\,Sv,\,av}}$ as high as 1073 $\mathrm{mg\,L^{-1}}$, which agrees with the right-hand tail of the distribution obtained in our study (i.e. a 0.995 quantile of 909.7 $\mathrm{mg\,L^{-1}}$). Similarly, for the case of $C_{\mathrm{NH_4,\,Sv,\,av}}$, Diaz-Fierros T et al. (2002) reported values as high as 32.5 $\mathrm{mg\,L^{-1}}$, which is comparable with the 0.995 quantile (29.20 $\mathrm{mg\,L^{-1}}$) found in our study.

It is worth noting that, regarding measurements taken at CSOs, the measured water quality at the WWTP influent is expected to render a low representativity of the conditions at the CSOs, as reported by Diaz-Fierros T et al. (2002) and Brombach et al. (2005). Thus, when comparing model outputs with independent measurements, one should bear in mind that discrepancies between measured and predicted values are not only caused

by errors in model inputs, model parameters, and model structure but are also the result of errors in the water quality measurements.

5.4.2 Accuracy of Monte Carlo analysis

Regarding the Monte Carlo replication size for uncertainty propagation, we presented the results in Fig. S4 for three output variables and three replications size 250, 1000, and the selected 1500 (Nash–Sutcliffe efficiency, NSE, closer to 1.0 for most of the output variables). We compute replications for 50, 100, and from 250 to 2000 at steps of 250 replications and the comparison of two equal MC runs (MC1 and MC2) with different seed for the pseudo-random number generator. The results suggest that the output variables related to COD (load and concentration) have a larger dispersion when we compare MC1 and MC2 for the same replications size. This is also reflected in the larger standard errors reported in Table 5.4, for example, the overflow COD load. Nevertheless, 1500 runs are a feasible MC replication size for running a relatively simple and fast model as EmiStatR (7.29 min in average execution time, using parallel computing and 50 cores for a time series with 4464 time steps). For a more complex full hydrodynamic model with a high computational burden, 1500 replications repeated four times to compute contributions may be not possible. Therefore, we suggest checking the intermediate results of the MC convergence test. We found that for quantity variables as the spill overflow volume and quality variables as the overflow NH_4 load in 250 replications (7.10 min in average execution time using parallel computing and three cores for a time series with 4464 time steps) per individual MC execution was enough, which makes an execution of this kind of uncertainty propagation more feasible.

Figures 5.3 and 5.4 show that there is a large uncertainty for the early May event and smaller uncertainty for the September event. This is due mainly to the presence of a large dry period before the spill event in May, i.e. a shorter dry period preceding the spill flow leads to a lower uncertainty. This finding suggest also an interaction between the antecedent dry period and the concentration of pollutants.

5.4.3 Other sources of uncertainty

In this work, we only looked at input uncertainty and not at parameter and model structural uncertainty. Further research can be done on those topics. Neumann (2007) address how uncertainty ranges for parameters of full-scale systems are obtained and how model structure uncertainty manifests itself and can be quantified for a performance evaluation and the design of urban water infrastructure. Moreno-Rodenas et al. (2019) also studied and depicted how a model parameter is an important source of uncertainty. They emphasised that "still, uncertainty analysis is seldom applied in practice, and the relative contribution of the individual model elements is poorly understood.". Also, they highlighted that, after inferring the river process parameters with system measurements

of flow and dissolved oxygen, combined sewer overflow pollution loads became the dominant uncertainty source along with rainfall variability. These findings agreed with our results.

Bachmann-Machnik et al. (2018) recognised that the most important parameters causing uncertainties in the sewer system model are the connected area and the storm water run-off quality. Our analysis confirms these findings, specifically regarding the storm water run-off quality. In our study, the input variable run-off COD was an important source of uncertainty with relation to the annual mean overflow COD released from the CSO.

5.4.4 Limitations and possible improvements

Despite the extensive temporal uncertainty propagation analysis, the approach also has some limitations which we present hereafter, addressing possible solutions in future work.

1. *Incorporation of the spatial distribution of model inputs.* Specifically for precipitation, Breinholt et al. (2012) stated that, due to a poor representation of the spatial precipitation that is measured by point gauges and the complexity of the sewer systems, large output uncertainty can be expected. We also infer that we obtained a large output uncertainty due to neglecting the inherent spatial variability in precipitation. Therefore, we suggest that further research is needed to account for spatial variability in precipitation that can bring light an understanding of how this variability impacts the output uncertainty and thus quantify it properly. This issue should also be related to the problem of change of support. When modelling precipitation, we also ignored the support effect, i.e. we ignored that the sub-catchment area is much greater than a point. Future research may address this issue of change-of-support. Studies that tackled this issue are available, e.g. Leopold et al. (2006); Wadoux et al. (2017); Cecinati et al. (2018).

2. *Linkage of sub-models and uncertainty compensation effect.* Tscheikner-Gratl et al. (2019) addressed the question as to whether there is an increase in uncertainty by linking integrated models or whether a compensation effect could take place by which overall uncertainty in key water quality parameters decreases. Some further insight into this topic could be obtained by quantifying uncertainties at sub-model level and analysing whether the uncertainty at sub-model level is greater or smaller than at the overall level. With our implementation, this is not a difficult task because EmiStatR has a stringent modular design in which it is easy to analyse outputs and their uncertainties at sub-model level.

3. *Accounting for cross-correlation between the inputs precipitation and run-off COD concentration.* It is worth noting that we did not include correlation between $\mathrm{COD_r}$ and P. Including such a correlation would yield a more realistic model of the

uncertainty because these variables are known to have a strong correlation. It is highly recommended that correlations between COD_r and precipitation be included because loads in chemical oxygen demand are correlated with the overland flow due to precipitation, which may transport distributed pollutants to the sewer system. Also, the inputs $C_{COD,S}$ and $C_{NH_4,S}$ can be related with a daily curve that reflects the pattern of consumption in the household, like the German ATV-DVWK-A 134 curve. We used the latest version of EmiStatR (version 1.2.2.0), which considers these kinds of patterns.

4. *Absence of high frequency water quality observations to compare with model outputs and uncertainty prediction bands.* In order to gain an understanding of the temporal dynamics of nutrients (nitrogen, N, and phosphorus, P), Yu et al. (2020) applied high-frequency monitoring in a groundwater-fed, low-lying urban polder in Amsterdam (the Netherlands). They argued that although spatial and temporal concentration patterns from discrete sampling campaigns of water quality parameters, such as *E. coli*, showed a clear dilution pattern, the temporal patterns of N and P were still poorly understood, given their reactive nature and more complex biogeochemistry. Therefore, high-frequency measurement is a key factor in understanding these temporal dynamics and patterns.

5. *Absence of a joint spatio-temporal uncertainty analysis.* According to Zhou et al. (2020), the limitations in algorithms for classic uncertainty estimates is the reason that only the uncertainty in one dimension (either temporal variability or spatial heterogeneity) is considered, whereas the variation in the other dimension is dismissed, resulting in an incomplete assessment of the uncertainties. Zhou et al. (2020) also showed that classic metrics underestimate the uncertainty through averaging, which means a loss of information in the variation across spatio-temporal scales. To handle this limitation, suitable methods are the 3D variance partitioning for a new uncertainty estimation in both spatio-temporal scales (Zhou et al., 2020) or spatio-temporal geostatistics (Gräler et al., 2016).

6. *Uncertainty analysis with complex models.* In this research, we were able to conduct a comprehensive Monte Carlo uncertainty propagation analysis, which required a large number of Monte Carlo runs. This was possible because we used a strongly simplified urban water system model, EmiStatR. For more complex models that take much more computing time, the application of a Monte Carlo uncertainty propagation analysis is more challenging. However, given sufficient, resources it is possible because each model run can be run independently, and hence, the analysis is extremely suitable for parallelisation and cloud computing. In particular, the use of graphics processing units (GPUs) for heavy computation is promising. Some recent examples that demonstrate the potential of GPUs for this purpose are Eränen et al. (2014), Sten et al. (2016), and Sandric et al. (2019). Sriwastava et al. (2018)

applied uncertainty propagation to a complex hydrodynamic model by selecting a small subset of dominant input/model parameters that explain most of the model output variance.

7. *Convolution of precipitation to run-off.* Equation (5.2) indicates that the run-off that enters the combined sewage in EmiStatR has the same shape as the precipitation, with an offset. Thus, for precipitation, no convolution is applied. In a new release of EmiStatR, we will incorporate convolution of precipitation when transforming precipitation to run-off.

5.5 Conclusions

In this final section, we conclude by highlighting the importance of temporal uncertainty propagation analysis and the selection and characterisation of uncertain model inputs impacting model sensitivity. We also point out that uncertainty propagation analysis helps to identify the sources that contribute the most and can provide better evidence for the impact assessment of pollutant release from sewer systems to the environment, in particular to the receiving waters.

1. *Uncertainty analysis is important because it quantifies the accuracy of model outputs and quantifies the uncertainty source contributions.* The latter provides essential information to help take informed decisions about how to improve the accuracy of the model output. But MC uncertainty analysis is only possible if it is computationally feasible. We used a simplified urban water system model with the capability to apply it for the minimisation of transient pollution from urban wastewater systems in parallel mode, which minimises model running time, allowing for uncertainty propagation, long-term simulations, and the evaluation of complex scenarios. These capabilities are crucial also for, for example, real-time control applications, where simplified models of fast running times are desirable.

2. *Input variables that were very uncertain, and for which model output was very sensitive, were selected to be included in the uncertainty propagation analysis.* We found four main input variables to be analysed, namely (1) precipitation, P, (2) chemical oxygen demand sewage pollution per capita load per day, $C_{\mathrm{COD,S}}$, (3) ammonium pollution per capita load per day, $C_{\mathrm{NH_4,S}}$, and (4) chemical oxygen demand $\mathrm{COD_r}$ concentration.

3. *Selected input variables for uncertainty propagation can be characterised in terms of the input uncertainty in four specific cases, depending on the type of input variable.* (i) Uncertain constant inputs, characterised by their marginal (cumulative) pdf, e.g. water consumption, infiltration flow, impervious area and run-off coefficients; (ii) temporally autocorrelated dynamic uncertain inputs, characterised by univariate time series autoregressive modelling, e.g. $\mathrm{COD_r}$; (iii) temporally cross-

correlated multiple dynamic uncertain inputs, characterised by multivariate time series modelling, considering cross- and no-correlations among variables, e.g. $C_{\mathrm{COD,S}}$ and $C_{\mathrm{NH_4,S}}$; and (iv) rain gauge input precipitation, characterised by autoregressive model conditioned to the observed precipitation (P).

4. *Model input uncertainty propagation through the simplified, combined sewer overflow model* (EmiStatR) *helped us to understand how uncertainty propagates and how large the uncertainty of EmiStatR outputs is in a case study.* Three output variables were considered for water quantity and four variables for water quality. The Monte Carlo uncertainty propagation analysis showed that, among the water quantity output variables, the overflow flow, Q_{Sv}, is the more uncertain output variable and has a large coefficient of variation (cv of 1.585). Among water quality variables, the annual average spill COD concentration, $C_{\mathrm{COD,Sv,av}}$, and the average spill NH_4 concentration, $C_{\mathrm{NH_4,Sv,av}}$, were found to have large uncertainty (coefficients of variation of 0.988 and 0.815, respectively). Also, low standard errors (se) for the coefficient of variation were obtained for all seven outputs. They were never greater than 0.05, which indicated that the selected MC replication size (1500 simulations) was a suitable value.

5. *Main sources of uncertainty model outputs.* For water quantity outputs it was precipitation, while for COD, water quality outputs were P, $C_{\mathrm{COD,S}}$, and $\mathrm{COD_r}$, and for NH_4, outputs were P and $C_{\mathrm{NH_4,S}}$.

6. *Uncertainty propagation analysis can explain the impact of water quality indicators to the receiving river for the Luxembourg case study more comprehensively.* Although the mean model water quality outputs for COD and NH_4 concentrations are fairly above the thresholds, the 0,95 quantile is 2.7 times above the mean value for COD concentration and 2.4 times above the mean value for NH_4. We conclude that we are not certain that environmental thresholds are not exceeded because there is a considerable probability that values are above them, even though the expected value is below the thresholds. This is valid for concentrations in the spilled CSO; therefore, it is important to highlight that the results confirmed our hypothesis that annual mean COD and NH_4 river concentrations are lower than the released CSO concentrations due to dilution and are thus compliant with the water quality thresholds given by the guidelines consulted.

Table 5.7: List of variables of EmiStatR.

cs_{mr_t}	Combined sewer mixing ratio at time t	(–)
pe_t	Population equivalents of the connected CSO structure at time t	(PE)
q_{f_t}	Infiltration water inflow flux at time t	($L\,s^{-1}\,ha^{-1}$)
qs_t	Individual water consumption of households at time t	($L\,PE^{-1}\,d^{-1}$)
t_{fS}	Delay in time response related to flow time in the sewer system	(time step)
A_{imp}	Impervious area of the catchment	(ha)
A_{total}	Total area of the catchment	(ha)
$B_{COD,Sv}$	COD load in the spill overflow	(g)
$B_{NH_4,Sv}$	Spill overflow NH_4 load	(g)
C_d	Orifice coefficient of discharge in the CSOC structure	(–)
C_{imp}	Run-off coefficient for impervious areas	(–)
C_{per}	Run-off coefficient for pervious areas	(–)
C_{COD_t}	Mean dry weather pollutant (COD) concentration at time t	($mg\,L^{-1}$)
$C_{COD,S}$	COD sewage pollution per capita load per day	($g\,PE^{-1}\,d^{-1}$)
$C_{COD,Sv,99.9}$	A 99.9th percentile spill COD concentration	($mg\,L^{-1}$)
$C_{COD,Sv,av}$	Average spill COD concentration	($mg\,L^{-1}$)
$C_{COD,Sv,max}$	Maximum overflow COD concentration	($mg\,L^{-1}$)
$C_{NH_4,S}$	NH_4 sewage pollution per capita load per day	($g\,PE^{-1}\,d^{-1}$)
$C_{NH_4,Sv,99.9}$	A 99.9th percentile spill NH_4 concentration	($mg\,L^{-1}$)
$C_{NH_4,Sv,av}$	Average spill NH_4 concentration	($mg\,L^{-1}$)
$C_{NH_4,Sv,max}$	Maximum spill NH_4 concentration	($mg\,L^{-1}$)
COD_f	Infiltration pollution COD	($g\,PE^{-1}\,d^{-1}$)
COD_{r_t}	Concentration due to rainwater pollution (COD) at time t	($mg\,L^{-1}$)
D_d	Orifice diameter	(m)
Lev	Measured water level in the CSOC	(m)
Lev_{ini}	Initial water level in the CSOC	(m)
NH_{4f}	Infiltration pollution NH_4	($g\,PE^{-1}\,d^{-1}$)
NH_{4r}	Concentration due to rainwater pollution (NH_4)	($mg\,L^{-1}$)
P_t	Precipitation at time t	($mm\,min^{-1}$)
$Q_{d,max}$	Maximum throttled outflow	($L\,s^{-1}$)
Q_{f_t}	Infiltration flow at time t	($L\,s^{-1}$)
Q_{r_t}	Flow contribution of rainwater to the combined sewage flow at time t	($m^3\,s^{-1}$)
Q_{s_t}	Dry weather flow of the residential sewage in the catchment at time t	($L\,s^{-1}$)
Q_{CSF_t}	Combined sewer flow at time t	($m^3\,s^{-1}$)
Q_{DWF_t}	Total dry weather flow at time t	($L\,s^{-1}$)
Q_{Sv}	Overflow spill flow	($L\,s^{-1}$)
V	Volume of the CSOC structure	(m^3)
V_{d_t}	Volume of throttled outflow to the WWTP at time t	(m^3)
V_{dw_t}	Total dry weather volume (amount of dry weather water in combined sewage flow) at time t	(m^3)
V_{r_t}	Rain weather volume at time t	(m^3)
$V_{Chamber_t}$	Water volume in the CSO chamber at time t	(m^3)
V_{Sv_t}	Spill overflow volume at time t	(m^3)
ϵ	numerical precision term	(m^3)
Δt	Time interval	(min)

Chapter 6

Synthesis – Lessons learned and future directions

6.1 Introduction

In Chapter 1, I introduced uncertainty propagation in urban stormwater systems modelling which is not yet well understood, and, therefore, very few applications have been established. This is due to four main problems: (1) lack of scalable software tools enabled for parallel computing to perform long-term simulation and uncertainty propagation analysis by Monte Carlo simulation; (2) catchment average precipitation is not always accurately known when measured at the point support, e.g. rain gauge measurements, which makes quantification of uncertainty associated with input precipitation estimation an important step in urban stormwater systems modelling; (3) software tools for temporal uncertainty propagation for urban stormwater system models (USSM's) are not available, specifically for Monte Carlo uncertainty propagation analyses; and (4) temporal statistical uncertainty analysis of USSM's is a relatively new subject, which needs to be developed and requires contributions from multiple disciplines (i.e., hydrology, statistics, computer science).

This dissertation addressed all four problems through its Chapters 2 to 5, developing, by implementing and testing new methods and software tools, advancing and overcoming current limitations of urban stormwater systems modelling, and by establishing a solid model input uncertainty characterisation and uncertainty propagation analysis. The methods and tools were illustrated with case studies in urban stormwater systems modelling, carried out in a catchment and sewer system in Luxembourg.

This final chapter discusses whether the objectives of this dissertation were achieved (Section 6.2). Also, the findings of this dissertation are compared with existing literature and directions for future research are suggested (Section 6.3). Finally, in Section 6.4, I give a personal final reflection.

6.2 Overview of findings

The overall aim of this dissertation was to contribute to uncertainty propagation analysis in dynamic environmental modelling, with application on urban stormwater systems modelling. The overall objective was addressed through four specific objectives stated in Chapter 1, Section 1.4, of which the results were presented in Chapters 2 to 5.

In Chapter 2, I presented a simplified mechanistic (surrogate) urban water quality model, EmiStatR, to represent the overall dynamic behaviour of the CSO spill volume, load, and concentration of COD and NH_4. The implementation of the model in the programming language for statistical computing, R, allowed fast and scalable calculations, specifically to simulate scenarios with input data of yearly time series at 10 minutes temporal resolution, and to propagate model input uncertainty through Monte Carlo simulation, due to its parallel computing capabilities. For a case study in Luxembourg, the model was cali-

brated, tested and validated by comparing the performance against a complex mechanistic model that uses the "de Saint-Venant" partial differential equations to describe the full flow routing in the pipes of the sewer network. In Chapter 2, I also showed that adequate simulation of CSO spill volume, as well as COD and NH_4 loads and concentrations is possible using a scalable, surrogate model such as EmiStatR, which confirmed that white-box simplifications can lead to well-performing surrogate models. The inherent parallel computing and scalable capabilities of EmiStatR allows fast calculations for scenarios of high complexity and for long-term simulations to test hypotheses in USSM. Also, EmiStatR enables the large body of R functionalities available, for example, compatibility with input and output data formats for temporal and geospatial data and advanced calibration techniques. A comparison between the outputs from EmiStatR and those obtained using a well-known complex mechanistic model, in terms of simulation of volume in the CSO and the estimation of loads of COD and NH_4, showed very similar results. The case study showed that in small catchments (i.e., area of 30 ha or less) EmiStatR achieved satisfactory accuracy, similar to that of models of much higher complexity.

In Chapter 3, I demonstrated that precipitation is the most active flux and major input of hydrological systems, including urban stormwater systems. Precipitation controls hydrological states (soil moisture and groundwater level), and fluxes (runoff, evapotranspiration and groundwater recharge). Hence, precipitation plays a paramount role in urban hydrology. It controls the fluxes towards combined sewer tanks and the dilution of chemical and organic compounds in the wastewater. Furthermore, small catchments (i.e., areas of about 20 ha) have a fast response to precipitation. Therefore, catchment average precipitation is a key component in urban water models that needs to be known at high temporal resolution for small catchments. However, average catchment precipitation is not always accurately known when measured at rain gauges, because the location of the gauge or gauges might be outside the catchment boundaries or may not reflect the entire catchment. Therefore, there was a need to develop a method to estimate the precipitation in a catchment given a known precipitation time series at a location outside the catchment, while quantifying the uncertainty associated with the estimation. Such method would be very useful for real-world applications in urban hydrology. In Chapter 3 I showed that the first-order multivariate autoregressive model for conditional simulation of input precipitation based on a multiplicative error model, which follows essentially the same principle of a Kalman filter, is suitable to estimate precipitation time series in an ungauged catchment given the precipitation time series at two neighbouring precipitation gauges located close by. The proposed method was used to generate precipitation ensembles accounting for the uncertainty associated in the estimation at the ungauged location. Such simulated precipitation ensembles can be used as model input for urban water models in tasks related to uncertainty propagation analysis, to account for uncertainty in improved urban water system design and to better assess environmental and economic impacts of, for example, CSO's over the receiving waters.

In Chapter 4, I introduced the software tool stUPscales, a software package for "spatio-temporal Uncertainty Propagation across multiple scales", developed in R to assess uncertainty in integrated environmental process modelling, coupling sub-models at different spatial and temporal scales and accounting for change of support procedures, such as aggregation and disaggregation. This software package also includes methods and functions suitable in model input uncertainty characterisation and propagation. stUPscales constitutes a contribution to the state-of-the-art of open source tools that support uncertainty propagation analysis in the temporal and spatio-temporal domains. Also, in this chapter, examples of applications to illustrate the tool were presented, specifically for uncertainty propagation in environmental modelling in the urban water domain. The chapter concluded with two application examples of stUPscales that demonstrate its suitability for characterising uncertainty in spatial, temporal and spatio-temporal environmental variables (model inputs) as probability distribution functions (pdfs) and as uni- and multi-variate autoregressive models. Moreover, it was demonstrated that it is possible to sample from these pdfs to generate realisations of autoregressive models, and to perform Monte Carlo based uncertainty propagation analysis with dynamic uncertain inputs. I also highlighted the main contributions of stUPscales compared to existing uncertainty propagation tools. stUPscales is not only restricted to uncertainty analysis of temporal or purely spatial applications but can also handle uncertain spatio-temporal variables.

In Chapter 5 I addressed the problem that uncertainty is often ignored in urban water systems modelling. Commercial software used in engineering practice often ignores the uncertainties of input variables and their propagation due to a lack of user-friendly implementations. This can have serious consequences, such as the wrong dimensioning of urban drainage systems and the inaccurate estimation of pollution released to the environment. The chapter also described an uncertainty propagation analysis in urban stormwater systems modelling by applying Monte Carlo simulation, which built on the methods presented in the previous chapters, and was applied to a case study in the Haute-Sûre catchment in Luxembourg. The case study made use of the EmiStatR model to simulate the volume and substance flows in the Haute-Sûre system using simplified representations of the drainage system and processes. The Monte Carlo uncertainty propagation analysis showed that uncertainties in COD and NH_4 loads and concentrations can be large and have a high temporal variability. Furthermore, a stochastic sensitivity analysis that assessed the uncertainty contributions of input variables to the model output showed that precipitation had the largest contribution to output uncertainty related with water quantity variables, such as volume in the chamber, overflow volume, and flow. Regarding the water quality variables, the input variables related to COD and NH_4 in wastewater had an important contribution to the uncertainty for load and concentration. The Monte Carlo simulation procedure used to propagate input uncertainty further showed that, among the water quantity output variables, the overflow flow is the most uncertain output variable. Among water quality variables, the annual average spill COD concentration and

the average spill NH_4 concentration were the most uncertain model outputs. Also, low standard errors for the coefficient of variation were obtained for all seven outputs, which indicated that the selected Monte Carlo replication size chosen was sufficient.

The analysis in Chapter 5 demonstrated how uncertainty propagation can more comprehensively explain the impact of water quality indicators for the receiving river. While the mean model water quality outputs for COD and NH_4 concentrations were slightly above the threshold, the 0.95 quantile was 2.7 times above the mean value for COD concentration and 2.4 times above the mean value for NH_4. This implies that there is a considerable probability that these concentrations in the spilled combined sewer overflow (CSO) were substantially larger than the threshold. However, COD and NH_4 concentration levels of the river water will likely stay below the water quality threshold, due to rapid dilution after CSO spill enters the river.

Chapter 5 concluded by highlighting the importance of temporal uncertainty propagation analysis and the selection and characterisation of uncertain model inputs impacting model sensitivity. I also pointed out that uncertainty propagation analysis helps to identify the sources that contribute most to output uncertainty, which is highly relevant information in order to take rational decisions how to reduce uncertainty.

The content and discussion presented in Chapters 2 to 5 allow to conclude that the objectives of this dissertation have been achieved.

6.3 Comparison with recent developments and future directions

This dissertation focused on addressing temporal correlation and uncertainty analysis. For future work, it would be of interest to the scientific and practitioner communities to also take the spatial and spatio-temporal distribution and correlation of some of the input variables, such as precipitation, impervious areas, and land use, into account. The literature shows that neither spatial nor spatio-temporal variation in precipitation is considered in many commonly used stormwater models (Zoppou, 2001; Bach et al., 2014). Usually, precipitation is assumed to be uniformly distributed in a sub-catchment. This is not a very realistic assumption, particularly in applications for which the response time is short. The integration of geostatistical probability models that interpolate and simulate precipitation data in the spatial and temporal domains would be an important advancement in USSM. Studies that addressed these topics are found in Muthusamy et al. (2017), Cecinati et al. (2018), and Wadoux et al. (2020). Those approaches could also be applied to urban hydrology.

Regarding software development, I recommend to explore the development of a new class or to extend the spatio-temporal, ST, classes in R so that these incorporate uncertainty

information regarding data uncertainty, input uncertainty characterisation, output uncertainty and summary statistics. A key source is (Bastin et al., 2013), which outlines a way forward to incorporate uncertainty in environmental model chains and software. I believe that the tools presented in this dissertation are a valuable contribution to these developments. In addition, further extension of the stUPscales package will allow for implementing new methods and functions for spatio-temporal disaggregation of model inputs and outputs, giving more flexibility to the proposed workflow when linking models with different spatio-temporal support or across multiple space-time scales, as stipulated in (Bastin et al., 2013). This is also recognised by Cristiano et al. (2017), in a study of spatial and temporal variability of precipitation and their effects on the hydrological response in urban areas. Disaggregation and aggregation techniques are generally necessary to obtain input and output data at the required spatial and temporal support (Leopold et al., 2006). This is particularly important when different sub-models, operating at different scales, are combined to an integrated process model (Leopold et al., 2006). Furthermore, different levels of model integration also demand different amount and quality of data for modelling and decision-making (Eggimann et al., 2017).

It is worthwhile to note the perspective for future developments in this specific research domain as expressed in Tscheikner-Gratl et al. (2019). They emphasised the challenges and bottlenecks in the application of uncertainty analysis in integrated water quality models, and recognised that despite considerable uncertainty, integrated models are important tools to support effective decision making for water utilities. In addition, Tscheikner-Gratl et al. (2019) emphasised that major problems for performing a complete uncertainty analysis remain, e.g. the lack of data in environmental studies and a common non-sharing policy (Camargos et al., 2018), and the need to satisfy local regulators (Sriwastava et al., 2018). This dissertation is a contribution to solving the challenges and bottlenecks identified by (Tscheikner-Gratl et al., 2019). It addressed the need for computationally efficient models that allow Monte Carlo based uncertainty analysis, which allows to quantify the uncertainty of integrated models used to support effective decision-making by water utilities.

Although this dissertation made a useful contribution to addressing uncertainty and uncertainty propagation in urban stormwater systems modelling, it still is a challenge to apply temporal, spatial and spatio-temporal characterisation and uncertainty propagation of model inputs and to have a complete software tool to perform all together. Tscheikner-Gratl et al. (2019) discussed that a typical uncertainty analysis requires a significant number of simulations, and therefore, software should exhibit the capability to run such model simulations in semi- or automatic execution, e.g. batch mode, to make uncertainty analysis feasible and to avoid additional errors due to manual execution. As I already noted in the Introduction, most of the commercial software does not provide such capability (Mitchell et al., 2007; Bach et al., 2014; Tscheikner-Gratl et al., 2019). Even if such a capability exists, the lack of documentation to guide the end user hinders its proper use.

Therefore, to cope with this, the use of surrogate models or emulators is a plausible solution, and has already demonstrated its value in applications for sewer hydraulics (Carbajal et al., 2017; Mahmoodian et al., 2018, 2019), and rainfall dynamics in 2D physically-based flow models (Moreno-Rodenas et al., 2018). Further research in this domain is foreseen. In this dissertation I did not develop statistical emulators, but I developed the simplified surrogate model, EmiStatR, which helped to make Monte Carlo uncertainty propagation analysis feasible for real-world urban drainage modelling studies.

This dissertation also recognised the concept of expensiveness as discussed by Tscheikner-Gratl et al. (2019). The limitations found in USSM and IUDM are also related to the requirement of monetary investments (e.g. for software, sensors, experienced modellers) as well as time resources. Although quantifying and reducing uncertainty is an important step in any modelling framework (Reckhow, 1994), to communicate this effectively is still an important challenge. A seamless communication among academia, practitioners and stakeholders allows to transfer knowledge comprehensively to facilitate decision-making. Even if time and resources prevent further investigation to reduce uncertainty, we can still expect that knowledge of and accounting for uncertainty will lead to better decisions in the long run. Reckhow (1994) stated that "even if time and resources prevent further investigation to reduce uncertainty, we can still expect that knowledge of uncertainty will lead to better decisions in the long run than will ignorance of uncertainty". Currently, the practitioner community tends to gain interest to use information about uncertainty in models to better understand trade-offs between risks of failing environmental standards and investment costs. Therefore, following Sriwastava et al. (2018) and Tscheikner-Gratl et al. (2019), to communicate uncertainty in models as probability of failure of environmental standards, as well as impact of uncertainty on investment costs, are becoming important tasks of the whole modelling chain.

6.4 A contribution toward sustainable urban water management

In Chapter 1, following Qi et al. (2020), I discussed how in the last two decades governments recognised the importance of a more sustainable urban water management as a response to issues as water scarcity, surface water flooding, and water pollution. Therefore, new urban water management strategies and practices have been established and called distinctively to emphasise on the sensible relationship between nature and humans, i.e. "Sustainable Development" and "Nature-Based Solutions" (NBS). A recent review on NBS in urban water management is found in Oral et al. (2020).

Global cities are rethinking their approaches to flood risk management, complementing traditional grey infrastructure (e.g. flood walls, barriers, lined drainage channels, underground pipes and detention tanks) for flood defence toward implementing approaches

for water resilience (O'Donnell et al., 2019). Some examples that show the shift with respect to nature in urban planning and water management policies for water resilience are provided by O'Donnell et al. (2019) and Qi et al. (2020) and listed as follows:

- "Room for the River", the Netherlands, established in the late 1990s (Busscher et al., 2019),
- "Sustainable drainage systems (SuDS), United Kingdom, established in 1996 (Duffy and McKay, 2015; Lashford et al., 2019)
- "Stormwater best management practices" (BMPs) (USEPA, 2009) in the USA,
- "Low impact development" (LID) (USEPA, 2000) in the USA and Australia,
- "Making Space for Water", United Kingdom, established in 2004 (Fish et al., 2016),
- "Water sensitive urban design" (Sharma et al., 2016),
- "Green infrastructure" (Trogrlić et al., 2018)
- "Sponge Cities", China, established in 2013 (Wang et al., 2018; Zevenbergen et al., 2018)
- "Blue-Green City" (Lawson et al., 2014; O'Donnell et al., 2019; Oral et al., 2020).

The NBS general concept aims to bring nature and natural features and processes into cities, landscapes, and seascapes. In the context of cities, NBS often acts as an "umbrella" concept in improving urban sustainability (Pontee et al., 2016; Somarakis et al., 2019). Therefore, NBS can be useful for improving the mentioned urban water issues by coupling water management strategies with urban planning, design, and implementation of development or renovation projects (Qi et al., 2020).

Following (Qi et al., 2020), to propend for a positive impact for society is a key aim of NBS, and therefore this umbrella concept embraces several of the older, more focused concepts, principles, and programmes, including BMPs, LID, SuDs, WSUD, LIUDD, Blue–Green Cities (BGCs), and the SCP (Figure 6.1). For instance, (Qi et al., 2020) argued that *"integrating grey infrastructure with natural features and processes to deal with water-related issues in built environments is integral to WSUD and SuDs. NBS goes further though, specifically highlighting the limitations of solutions relying purely on so called "traditional" engineering approaches that actually date only from the late 19th century".* The concepts depicted in Figure 6.1 conform with the NBS concept to provide a range of ecosystem services alongside sustainable structural (i.e., engineered assets) and non-structural (i.e., changes in stakeholder behaviours) measures to deal with urban water issues (Qi et al., 2020).

To improve current inadequate flood risk management design standards, leads to more resilient solutions. For instance, a hybrid, "Green+Grey", infrastructure includes (Qi

et al., 2020):

- Piped drainage systems interspersed with bio-swales and SuDs ponds,
- Residential developments with sunken rain gardens,
- Buildings with green and blue roofs, and ecological levees,
- Corridors that increase stormwater conveyance capacity along engineered urban drainage channels.

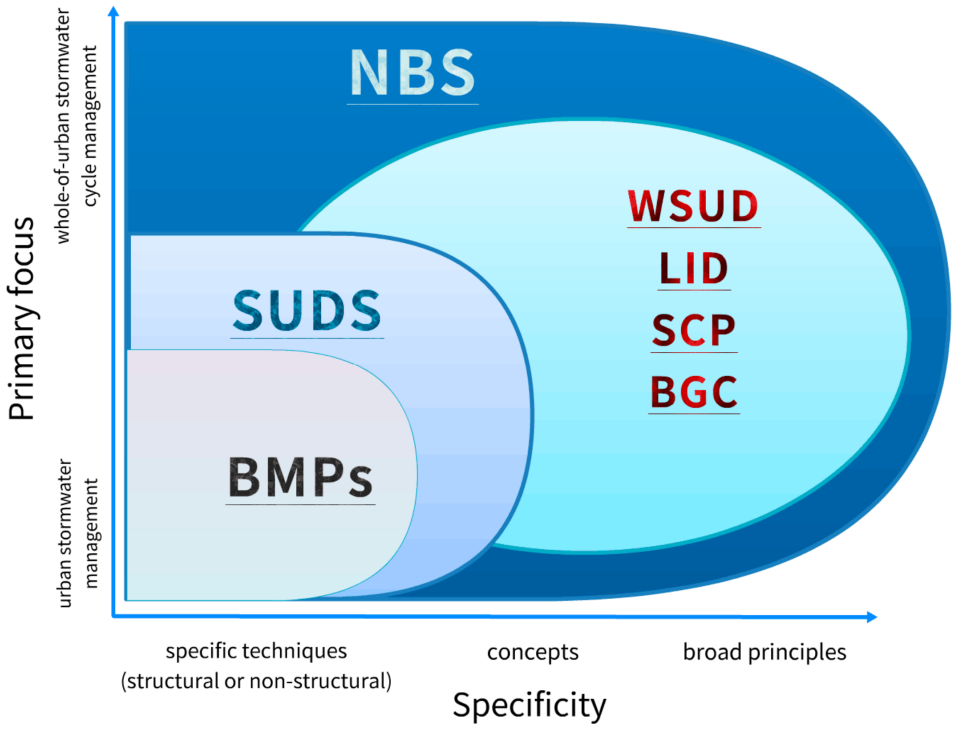


Figure 6.1: Relationship of Nature-Based Solutions (NBS), Best Management Practices (BMPs), Sustainable Drainage Systems (SuDs), Water Sensitive Urban Design (WSUD), Low Impact Development (LID), Sponge Cities Program (SCP), Blue–Green Cities (BGCs). Source: Qi et al. (2020).

Such integrated "Green+Grey" infrastructure constitute a NBS approach that can perform multiple functions because it is capable of concurrently addressing not one but several urban water issues (e.g., flood risk, pollution, urban heat island effects, water purification and reuse, biodiversity, provision of blue–green recreational spaces, etc.) (Qi et al., 2020) (Figure 6.2).

Following the discussion presented in this Chapter, and in agreement with Tscheikner-Gratl et al. (2019), further research and case studies should be carried out on how to involve local environmental regulators and organisations in order to raise awareness and appreciation of uncertainty analysis and how to incorporate it into decision-making for

sustainable urban water management and Nature-Based Solutions in a wider sense. This will then also lead to better investment actions which are required to meet current and future policy regulations.

The main application of this dissertation is in the broader domain of sustainable urban water modelling and management aiming to contribute toward understanding and quantifying sources of uncertainty due to intermittent combined sewer overflows, and how these uncertainties propagate to model output.

Finally, in sustainable urban water management it is of paramount importance to pay attention to uncertainty. Addressing uncertainty characterisation and uncertainty propagation in the modelling chain for the design of NBS, requires tailored methods and software tools that enable temporal, spatial and spatio-temporal uncertainty characterisation and propagation across multiple scales, while linking several modelling sub-modules. In this respect, the methods for multivariate time series modelling and conditional simulation of precipitation ensembles, and the R-packages, EmiStatR and stUPscales, developed along this dissertation with illustrative examples, constitute a contribution to facilitate dimension and design of the required infrastructure for decision-making in sustainable urban water management.

With this dissertation I hope to have made a contribution and have laid a basis towards the design and development of sustainable urban water management as part of a hybrid "Blue-Green-Grey" infrastructure for the cities of the future.

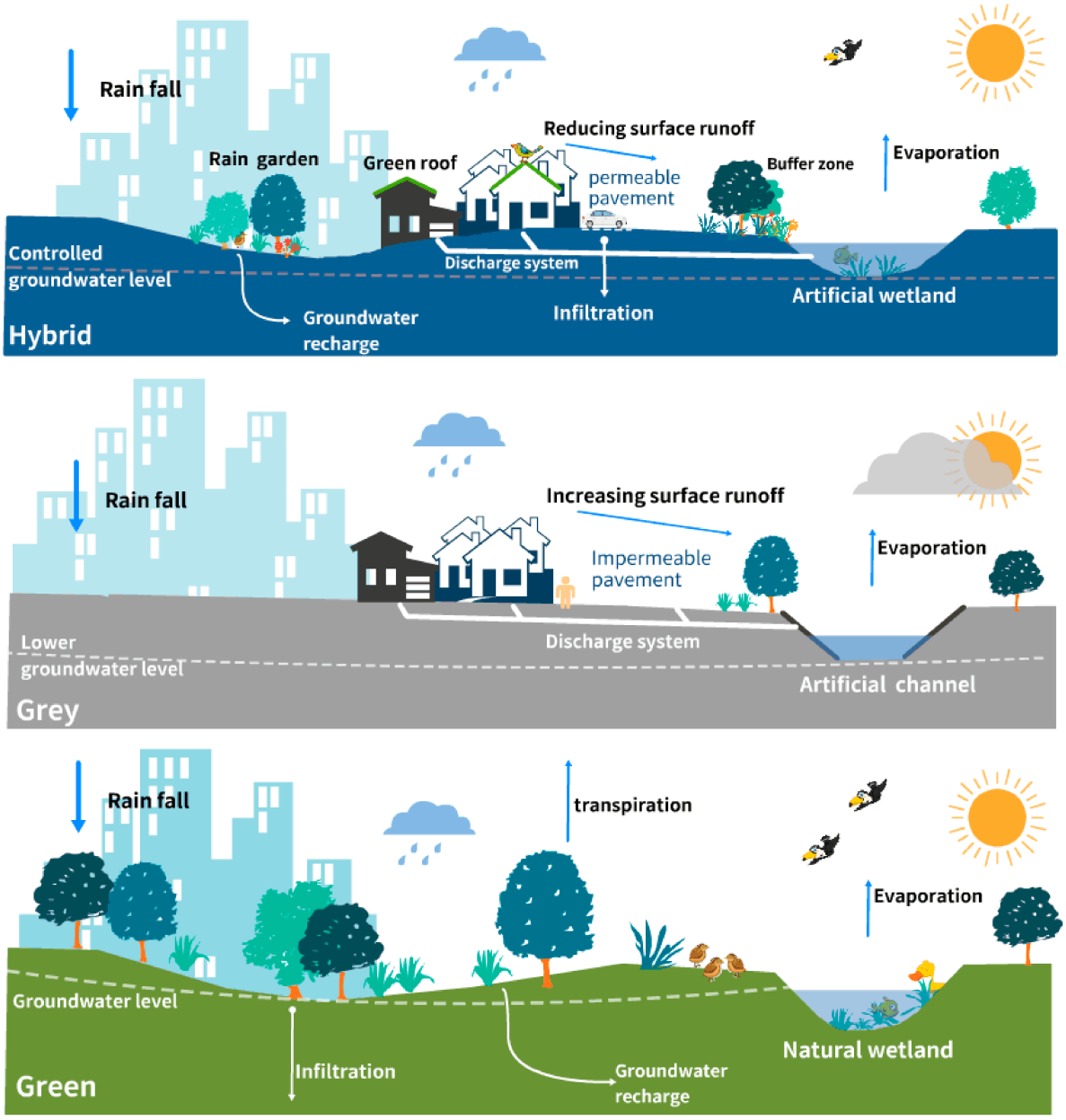


Figure 6.2: Nature-Based Solutions integrate the grey and green to create hybrid approaches. Source: Qi et al. (2020).

References

Adams, B.; Bauman, L.; Bohnhoff, W.; Dalbey, K.; Ebeida, M.; Eddy, J.; Eldred, M.; Hough, P.; Hu, K.; Jakeman, J.; Swiler, L., and Vigil., D. *Dakota, a multilevel parallel object-oriented framework for design optimization, parameter estimation, uncertainty quantification, and sensitivity analysis: Version 5.4 users manual.* Report Sandia Technical Report SAND2010-2183, Sandia National Laboratories, 2009.

Ahmad, Shohan S. and Simonovic, Slobodan P. Spatial and temporal analysis of urban flood risk assessment. *Urban Water Journal*, 10(1):26–49, FEB 1 2013. ISSN 1573-062X. doi: 10.1080/1573062X.2012.690437.

Andrés-Doménech, I; Múnera, J C; Francés, F, and Marco, J B. Coupling urban event-based and catchment continuous modelling for combined sewer overflow river impact assessment. *Hydrology and Earth System Sciences*, 14(10):2057–2072, 2010. doi: 10.5194/hess-14-2057-2010.

Andrianov, G.; Burriel, S.; Cambier, S.; Dutfoy, A.; Dutka-Malen, I.; de Rocquigny, E.; Sudret, B.; Benjamin, P.; Lebrun, R.; Mangeant, F., and Pendola., M. Openturns, an open source initiative to treat uncertainties, risks-n statistics in 520 a structured industrial approach. *In ESREL-2007 Safety and Reliability Conference*, 2007.

Aria, Massimo and Cuccurullo, Corrado. bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4):959–975, 2017. URL `https://doi.org/10.1016/j.joi.2017.08.007`.

Aronica, G. and Cannarozzo, M. Studying the hydrological response of urban catchments using a semi-distributed linear non-linear model. *Journal of Hydrology*, 238(1-2):35–43, nov 2000. ISSN 00221694. doi: 10.1016/S0022-1694(00)00311-5.

Aronica, G; Freni, G, and Oliveri, E. Uncertainty analysis of the influence of rainfall time resolution in the modelling of urban drainage systems. *Hydrological Processes*, 19(5):1055–1071, MAR 30 2005. ISSN 0885-6087. doi: 10.1002/hyp.5645.

Bach, Peter M.; Rauch, Wolfgang; Mikkelsen, Peter S.; McCarthy, David T., and Deletic, Ana. A critical review of integrated urban water modelling - Urban drainage and beyond. *Environmental Modelling & Software*, 54:88 – 107, 2014. ISSN 13648152. doi: 10.1016/j.envsoft.2013.12.018.

Bachmann-Machnik, Anna; Meyer, Daniel; Waldhoff, Axel; Fuchs, Stephan, and Dittmer, Ulrich. Integrating retention soil filters into urban hydrologic models – relevant processes and important parameters. *Journal of Hydrology*, 559:442 – 453, 2018. ISSN 0022-1694. doi: /10.1016/j.jhydrol.2018.02.046.

Baker, Lawrence A., editor. *The Water Environment of Cities.* Springer, 2009. ISBN 9788578110796.

Barbosa, A. E.; Fernandes, J. N., and David, L. M. Key issues for sustainable urban stormwater management. *WATER RESEARCH*, 46(20, SI):6787–6798, DEC 15 2012. ISSN 0043-1354. doi: 10.1016/j.watres.2012.05.029.

Barbosa, S. M. *Package "mAr": Multivariate AutoRegressive analysis.* The Comprehensive R Archive Network, CRAN, 1.1-2 edition, February 2015.

Bastin, Lucy; Cornford, Dan; Jones, Richard; Heuvelink, Gerard B. M.; Pebesma, Edzer; Stasch, Christoph; Nativi, Stefano; Mazzetti, Paolo, and Williams, Matthew. Managing uncertainty in integrated environmental modelling: The UncertWeb framework. *Environmental Modelling & Software*, 39:116–134, 2013. ISSN 13648152. doi: 10.1016/j.envsoft.2012.02.008.

Bell, Colin D.; Tague, Christina L., and McMillan, Sara K. A model of hydrology and water quality for stormwater control measures. *Environmental Modelling and Software*, 95:29–47, 2017. ISSN 13648152. doi: 10.1016/j.envsoft.2017.05.007.

Berne, Alexis; Delrieu, Guy; Creutin, Jean-Dominique, and Obled, Charles. Temporal and spatial resolution of rainfall measurements required for urban hydrology. *Journal of Hydrology*, 299(3-4):166–179, dec 2004. ISSN 00221694. doi: 10.1016/j.jhydrol.2004.08.002.

Bertsimas, Dimitris and Sim, Melvyn. The Price of Robustness. *Operations Research*, 52(1):35–53, 2004. ISSN 0030-364X. doi: 10.1287/opre.1030.0065.

Beven, K.. and Binley, A. The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes*, (Vol. 6):279–98, 1992. doi: 10.1002/hyp.3360060305.

Beven, K.; Leedal, D., and Alcock, R. Uncertainty and good practice in hydrological prediction. *VATTEN*, 66:159–163, 2010.

Beven, K. J. Changing ideas in hydrology. The case of physically-based models. *Journal of Hydrology*, 105 (1-2):157–172, 1989. doi: 10.1016/0022-1694(89)90101-7.

Beven, K. J. A manifesto for the equifinality thesis,. *J. hydrology,*, 320:18–36, 2006. doi: 10.1016/j.jhydrol.2005.07.007.

Beven, Keith and Freer, Jim. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology*, 249(1-4):11–29, 2001. ISSN 00221694. doi: 10.1016/S0022-1694(01)00421-8.

Beven, Keith J. *Rainfall-Runoff Modelling: The Primer*. Wiley-Blackwell, second edition, 2012. ISBN 978-0-470-71459-1. Lancaster University, UK.

Bilonick, Richard A. Monthly hydrogen ion deposition maps for the northeastern U.S. from July 1982 to September 1984. *Atmospheric Environment*, 22(9):1909–1924, 1988. ISSN 00046981. doi: 10.1016/0004-6981(88)90080-7.

Bivand, Roger S.; Pebesma, Edzer, and Gomez-Rubio, Virgilio. *Applied spatial data analysis with R, Second edition*. Springer, NY, 2013. ISBN 978-1-4614-7618-4. URL `http://www.asdar-book.org/`.

Blumensaat, F.; Staufer, P.; Heusch, S.; Reussner, F.; Schütze, M.; Seiffert, S.; Gruber, G.; Zawilski, M., and Rieckermann, J. Water quality-based assessment of urban drainage impacts in Europe where do we stand today? *Water Science & Technology*, 66(2):304, Jun 2012. ISSN 0273-1223. doi: 10.2166/wst.2012.178. URL `http://dx.doi.org/10.2166/wst.2012.178`. IWA Publishing.

Booker, DJ and Dunbar, MJ. Application of physical habitat simulation (phabsim) modelling to modified urban river channels. *River Research And Applications*, 20(2):167–183, MAR 2004. ISSN 1535-1459. doi: 10.1002/rra.742.

Boos, Dennis; Matthew, Kevin, and Osborne, Jason. *Monte.Carlo.se: Monte Carlo Standard Errors*, 2019. URL `https://CRAN.R-project.org/package=Monte.Carlo.se`. R package version 0.1.0.

Boos, Dennis D. Introduction to the Bootstrap and World. *Statistical Science*, 18(2):168–174, 2003. doi: 10.1214/ss/1063994971. Institute of Mathematical Statistics.

Boos, Dennis D. and Osborne, Jason A. Assessing Variability of Complex Descriptive Statistics in Monte Carlo Studies Using Resampling Methods. *International Statistical Review*, 83(2):228–238, AUG 2015. ISSN 0306-7734. doi: 10.1111/insr.12087.

Box, George E. P.; Jenkins, Gwilym M., and Rinsel, Gregory C. *Time series analysis: forecasting and control*. John Wiley & Sons, Inc, 4th edition, 2008. ISBN 978-0-470-27284-8.

Breinholt, Anders; Moller, Jan Kloppenborg; Madsen, Henrik, and Mikkelsen, Peter Steen. "a formal statistical approach to representing uncertainty in rainfall–runoff modelling with focus on residual analysis and probabilistic output evaluation – distinguishing simulation and prediction". *Journal of Hydrology*, (472-473):36–52, 2012. doi: 10.1016/j.jhydrol.2012.09.014.

Brombach, H.; Weiss, G., and Fuchs, S. A new database on urban runoff pollution: comparison of separate and combined sewer systems. *Water Science and Technology*, 51(2):119–128, 01 2005. ISSN 0273-1223. doi: 10.2166/wst.2005.0039.

Brown, J. D. and Heuvelink, G. B. M. The data uncertainty engine (DUE): A software tool for assessing and simulating uncertain environmental variables. *Computers & Geosciences*, 33(2):172190, 2007. doi: http://dx.doi.org/10.1016/j.cageo.2006.06.015. ISSN 0098-3004.

Brown, James D. Knowledge, uncertainty and physical geography: Towards the development of methodologies for questioning belief. *Transactions of the Institute of British Geographers*, 29(3):367–381, 2004. ISSN 00202754. doi: 10.1111/j.0020-2754.2004.00342.x.

Brown, James D. and Heuvelink, Gerard B. M. Assessing Uncertainty Propagation Through Physically based Models of Soil Water Flow and Solute Transport. *Encyclopedia of Hydrological Sciences*, pages 1181–1195, 2005. doi: 10.1002/0470848944.hsa081. URL `http://www.soil.tu-bs.de/lehre/Master.Unsicherheiten/2012/Lit/UncertaintyPropagation-Heuvelink.pdf`.

Brown, James D.; Spencer, Tom, and Moeller, Iris. Modeling storm surge flooding of an urban area with particular reference to modeling uncertainties: A case study of Canvey Island, United Kingdom. *Water Resources Research*, 43(6), JUN 7 2007. ISSN 0043-1397. doi: 10.1029/2005WR004597.

Brunetti, Giuseppe; Šimunek, Jirka; Turco, Michele, and Piro, Patrizia. On the use of surrogate-based modeling for the numerical analysis of Low Impact Development techniques. *Journal of hydrology*, 548:263–277, 2017. ISSN 0022-1694. doi: 10.1016/j.jhydrol.2017.03.013.

Burger, G.; Fach, S.; Kinzel, H., and Rauch, W. Parallel computing in conceptual sewer simulations. *Water Science & Technology*, 61 (2):283–291, 2010. doi: 10.2166/wst.2010.798.

Burger, G.; Sitzenfrei, R.; Kleidorfer, M., and Rauch, W. Parallel flow routing in SWMM 5. *Environmental Modelling & Software*, 53:2734, Mar 2014. ISSN 1364-8152. doi: 10.1016/j.envsoft.2013.11.002.

Burger, Gregor; Bach, Peter M.; Urich, Christian; Leonhardt, GÃ¼nther; Kleidorfer, Manfred, and Rauch, Wolfgang. Designing and implementing a multi-core capable integrated urban drainage modelling Toolkit: Lessons from CityDrain3.

Advances in Engineering Software, 100:277289, Oct 2016. ISSN 0965-9978. doi: 10.1016/j.advengsoft.2016.08.004.

Burgos, M.; Ortega, T., and Forja, J.M. Temporal and spatial variation of n2o production from estuarine and marine shallow systems of cadiz bay (sw, spain). *Science of the Total Environment*, 607-608:141–151, 2017. ISSN 00489697. doi: 10.1016/j.scitotenv.2017.07.021. cited By 6.

Busscher, T.; van den Brink, M., and Verweij, S. Strategies for integrating water management and spatial planning: Organising for spatial quality in the Dutch "Room for the River" program. *Journal Of Flood Risk Management*, 12(1), MAR 2019. ISSN 1753-318X. doi: 10.1111/jfr3.12448.

Byrd, Richard H.; Lu, Peihuang; Nocedal, Jorge, and Zhu, Ciyou. A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM Journal on Scientific Computing*, 16(5):19, 1995. doi: 10.1137/0916069.

Camargos, Carla; Julich, Stefan; Houska, Tobias; Bach, Martin, and Breuer, Lutz. Effects of input data content on the uncertainty of simulating water resources. *Water*, 10(5), 2018. ISSN 2073-4441. doi: 10.3390/w10050621. URL `https://www.mdpi.com/2073-4441/10/5/621`.

Carbajal, Juan Pablo; Leitão, João Paulo; Albert, Carlo, and Rieckermann, Jörg. Appraisal of data-driven and mechanistic emulators of nonlinear simulators: The case of hydrodynamic urban drainage models. *Environmental Modelling and Software*, 92:17–27, 2017. ISSN 13648152. doi: 10.1016/j.envsoft.2017.02.006.

Castiglioni, S.; Thomas, K.V.; Kasprzyk-Hordern, B.; Vandam, L., and Griffiths, P. Testing wastewater to detect illicit drugs: State of the art, potential and research needs. *Science of the Total Environment*, 487(1):613–620, 2014. ISSN 00489697. doi: 10.1016/j.scitotenv.2013.10.034. cited By 115.

Cecinati, Francesca; Moreno-Ródenas, Antonio M.; Rico-Ramirez, Miguel A.; ten Veldhuis, Marie Claire, and Langeveld, Jeroen G. Considering rain gauge uncertainty using kriging for uncertain data. *Atmosphere*, 9(11):1–17, 2018. ISSN 20734433. doi: 10.3390/atmos9110446.

Chen, Lajiao; Ma, Yan; Liu, Peng, and Xue, Wei. Parallelisation of a watershed distributed ecohydrological model with dynamic task scheduling. *International Journal of Ad Hoc and Ubiquitous Computing*, 17(2/3), 2014. doi: 10.1504/IJAHUC.2014.065774.

Claeys, F.; Chtepen, M.; Benedetti, L.; Dhoedt, B., and Vanrolleghem, P. A. Distributed virtual experiments in water quality management. *Water Science and Technology*, 53(1):297305, Jan 2006. ISSN 1996-9732. doi: 10.2166/wst.2006.032.

Clark, R.; Pezzaniti, D., and Cresswell, D. Watercress - Community Resource Evaluation and Simulation System - A tool for innovative urban water system planning and design. In *Proc Hydrology and Water Resources Symposium 2002*, 20-23 May 2002. Melbourne, Australia.

Committee on Environmental and Natural Resources - CENR, . An assessment of coastal hypoxia and eutrophication in u.s. waters. Technical report, NSTC, National Science and Technology Council, 2003.

Coutu, Sylvain; Del Giudice, Dario; Rossi, Luca, and Barry, D. A. Parsimonious hydrological modeling of urban sewer and river catchments. *Journal of Hydrology*, 464-465:477–484, 2012. ISSN 00221694. doi: 10.1016/j.jhydrol.2012.07.039.

CRC-CH, . *MUSIC by eWater*. eWater, 2005. https://ewater.org.au/products/music/.

Cristiano, E.; ten Veldhuis, M.-C., and van de Giesen, N. Spatial and temporal variability of rainfall and their effects on hydrological response in urban areas – a review. *Hydrology and Earth System Sciences*, 21(7):3859–3878, 2017. doi: 10.5194/hess-21-3859-2017. URL `https://hess.copernicus.org/articles/21/3859/2017/`.

Datta, Arpana Rani. *Evaluation of Implicit and Explicit Methods of Uncertainty Analysis on a Hydrological Modeling*. PhD thesis, University of Windsor, Canada, 2011.

Davis, Allen P. and McCuen, Richard H. *Stormwater Management for Smart Growth*. Springer-Verlag, 2005. doi: 10.1007/0-387-27593-2. URL `https://doi.org/10.1007/0-387-27593-2`.

Del Giudice, Dario; Albert, Carlo; Rieckermann, Jörg, and Reichert, Peter. Describing the catchment-averaged precipitation as a stochastic process improves parameter and input estimation. *Water Resources Research*, 52(4):31623186, Apr 2016. ISSN 0043-1397. doi: 10.1002/2015wr017871.

Deletic, A..; Dotto, C.B.S.; McCarthy, D.T.; Kleidorfer, M.; Freni, G.; Mannina, G.; Uhl, M.; Henrichs, M.; Fletcher, T.D.; Rauch, W.; Bertrand-Krajewski, J.L., and Tait, S. Assessing uncertainties in urban drainage models. *Physics and Chemistry of the Earth*, 42-44:3–10, 2012. ISSN 14747065. doi: 10.1016/j.pce.2011.04.007.

DHI, . *MIKE URBAN*, 2007. https://www.mikepoweredbydhi.com/products/mike-urban.

DHI, . MIKE URBAN e Model Manager. Technical report, DHI, Denmark, 2009.

DHI, . *MIKE11, A modeling system for rivers and channels, Reference Manual*. DHI Water and Environment, 2017.

Diaz-Fierros T, F; Puerta, J; Suarez, J, and Diaz-Fierros V, F. Contaminant loads of csos at the wastewater treatment plant of a city in nw spain. *Urban Water*, 4(3):291 – 299, 2002. ISSN 1462-0758. doi: https://doi.org/10.1016/S1462-0758(02)00020-1.

Dokulil, MT and Teubner, K. Cyanobacterial dominance in lakes. *Hydrobiologia*, 438(1-3):1–12, NOV 2000. ISSN 0018-8158.

doi: 10.1023/A:1004155810302.

Dotto, Cintia B S; Mannina, Giorgio; Kleidorfer, Manfred; Vezzaro, Luca; Henrichs, Malte; Mccarthy, David T; Freni, Gabriele; Rauch, Wolfgang, and Deletic, Ana. Comparison of different uncertainty techniques in urban stormwater quantity and quality modelling. *Water Research*, 46(8):2545–2558, 2012. ISSN 0043-1354. doi: 10.1016/j.watres.2012.02.009.

Dubuis, Matthias. *Energy system design under uncertainty.* PhD thesis, École Polytechnique Fédérale De Lausanne, 2012.

Duffy, Alison and McKay, Gaye. A review of 20 years of SUDS in Scotland. In *World Water Congress Conference XV Edinburgh*, 2015.

DWA, . *Arbeitsblatt DWA-A 131: Bemessung von einstufigen Belebungsanlagen.* Germany, 2002. DWA-Regelwerk.

DWA, . *ATV-DVWK-A 118: Hydraulische Bemessung und Nachweis von Entwässerungssystemen.* Germany, 2006. DWA-Regelwerk.

Efron, B. Bootstrap methods: Another look at the Jackknife. *The Annals of Statistics*, 7(1):1–26, 1979.

Eggimann, Sven; Mutzner, Lena; Wani, Omar; Schneider, Mariane Yvonne; Spuhler, Dorothee; Moy de Vitry, Matthew; Beutler, Philipp, and Maurer, Max. The potential of knowing more: A review of data-driven urban water management. *Environmental Science & Technology*, 51(5):2538–2553, 2017. doi: 10.1021/acs.est.6b04267. PMID: 28125222.

Einfalt, Thomas; Arnbjerg-Nielsen, Karsten; Golz, Claudia; Jensen, Niels-Einar; Quirmbach, Markus; Vaes, Guido, and Vieux, Baxter. Towards a roadmap for use of radar rainfall data in urban drainage. *Journal of Hydrology*, 299(3-4):186–202, dec 2004. ISSN 00221694. doi: 10.1016/j.jhydrol.2004.08.004.

Elliott, A. H. and Trowsdale, S. A. A review of models for low impact urban stormwater drainage. *Environmental Modelling & Software*, 22(3):394–405, MAR 2007. ISSN 1364-8152. doi: 10.1016/j.envsoft.2005.12.005.

Elmore, Andrew J.; Guinn, Steven M.; Minsley, Burke J., and Richardson, Andrew D. Landscape controls on the timing of spring, autumn, and growing season length in mid-Atlantic forests. *Global Change Biology*, 18(2):656–674, FEB 2012. ISSN 1354-1013. doi: 10.1111/j.1365-2486.2011.02521.x.

Elmqvist, Thomas; Fragkias, Michail; Goodness, Julie; Güneralp, Burak; Marcotullio, Peter J.; McDonald, Robert I.; Parnell, Susan; Schewenius, Maria; Sendstad, Marte; Seto, Karen C., and Wilkinson, Cathy, editors. *Urbanization, Biodiversity and Ecosystem Services: Challenges and Opportunities*. Springer Netherlands, 2013. doi: 10.1007/978-94-007-7088-1.

Eränen, David; Oksanen, Juha; Westerholm, Jan, and Sarjakoski, Tapani. A full graphics processing unit implementation of uncertainty-aware drainage basin delineation. *Computers & Geosciences*, 73:48 – 60, 2014. ISSN 0098-3004. doi: 10.1016/j.cageo.2014.08.012.

Evers, Peter; Heinz, Haendel; Hanitsch, Peter H.; Koch, GÃ¼nther; Naupold, Lutz; Tochtermann, Wolfgang; Tornow, Manfred; Zander, Bernd; Mahret, Hansjoachim, and Warnow, Dietrich. ATV-DVWK-A 134E: Planning and Construction of Wastewater Pumping Stations. Technical report, DWA, Germany, 2000.

eWater, . *Urban Developer*, 2011. https://ewater.org.au/products/music/related-tools/urban-developer/.

Fan, Liangxin; Liu, Guobin; Wang, Fei; Geissen, Violette; Ritsema, Coen J., and Tong, Yan. Water use patterns and conservation in households of Wei River Basin, China. *Resources, Conservation and Recycling*, 74:45–53, 2013. ISSN 09213449. doi: 10.1016/j.resconrec.2013.02.017.

Fish, Robert; Church, Andrew; Willis, Cheryl; Winter, Michael; Tratalos, Jamie A.; Haines-Young, Roy, and Potschin, Marion. Making space for cultural ecosystem services: Insights from a study of the UK nature improvement initiative. *Ecosystem Services*, 21(B):329–343, OCT 2016. ISSN 2212-0416. doi: 10.1016/j.ecoser.2016.09.017.

Fletcher, T. D.; Andrieu, H., and Hamel, P. 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, JAN 2013. ISSN 0309-1708. doi: 10.1016/j.advwatres.2012.09.001.

Fowler, HJ and Kilsby, CG. A regional frequency analysis of United Kingdom extreme rainfall from 1961 to 2000. *International Journal Of Climatology*, 23(11):1313–1334, SEP 2003. ISSN 0899-8418. doi: 10.1002/joc.943.

Freer, J. E.; McMillan, H.; McDonnell, J. J., and Beven, K. J. Constraining dynamic TOPMODEL responses for imprecise water table information using fuzzy rule based performance measures. *Journal of Hydrology*, 291:254–277, 2004. doi: 10.1016/j.jhydrol.2003.12.037.

Freni, G.; Maglionico, M.; Mannina, G., and Viviani, G. Comparison between a detailed and a simplified integrated model for the assessment of urban drainage environmental impact on an ephemeral river. *Urban Water Journal*, 5(2):8796, Jun 2008. ISSN 1744-9006. doi: 10.1080/15730620701736878. URL `http://dx.doi.org/10.1080/15730620701736878`.

Fu, Baihua; Merritt, Wendy S.; Croke, Barry F. W.; Weber, Tony R., and Jakeman, Anthony J. A review of catchment-scale water quality and erosion models and a synthesis of future prospects. *Environmental Modelling & Software*, 114:75–97, APR 2019. ISSN 1364-8152. doi: 10.1016/j.envsoft.2018.12.008.

Fu, G; Khu, S.-T., and Butler, D. Use of surrogate modelling for multiobjective optimisation of urban wastewater systems.

Water Science & Technology, 60.6:1641–1647, 2009. doi: 10.2166/wst.2009.508.

Gasperi, Johnny; Zgheib, Sally; Cladière, Mathieu; Rocher, Vincent; Moilleron, Régis, and Chebbo, Ghassan. Priority pollutants in urban stormwater: Part 2 – case of combined sewers. *Water Research*, 46(20):6693 – 6703, 2012. ISSN 0043-1354. doi: 10.1016/j.watres.2011.09.041. Special Issue on Stormwater in urban areas.

Geiger, W. P. and Dorsch, H. R. *QuantityQuality Simulation (QQS) : A Detailed Continuous Planning Model for Urban Runoff Control, Volume 1 , Model Description, Testing and Applications.* US Environmental Protection Agency, Cincinnati, OH, US, 1980. Report EPA/600/2-80-011.

Gräler, Benedikt; Pebesma, Edzer, and Heuvelink, Gerard. Spatio-Temporal Interpolation using {gstat}. *The R Journal*, 8 (1):204–218, 2016. ISSN 2073-4859.

Grayson, R. B.; Moore, I. D., and McMahon, T. A. Physically Based Hydrologic Modelling 2. Is the Concept Realistic? *Water Resources Research*, 28 (10):2659–2666, 1992. doi: 10.1029/92WR01259.

Guan, Huade; Wilson, John L., and Xie, Hongjie. A cluster-optimizing regression-based approach for precipitation spatial downscaling in mountainous terrain. *Journal of Hydrology*, 375(3-4):578588, Sep 2009. ISSN 0022-1694. doi: 10.1016/j.jhydrol.2009.07.007.

Guillaume, Joseph and Andrews, Felix. *dream: DiffeRential Evolution Adaptive Metropolis. R package*, version 0.4-2. edition, 2012. URL `http://CRAN.R-project.org/package=dream`.

Hager, Willi H. *Wastewater hydraulics.* Springer, second edition, 2010. ISBN 978-3-642-11382-6. doi: 10.1080/00221686.2011.614723.

Hamel, P. and Guswa, A. J. Uncertainty analysis of a spatially explicit annual water-balance model: Case study of the Cape Fear basin, North Carolina. *Hydrology and Earth System Sciences*, 19(2):839–853, 2015. ISSN 16077938. doi: 10.5194/hess-19-839-2015.

Hammersley, J.M. and Handscomb, D.C. *Monte Carlo Methods.* Methuen & Co Ltd, London, 1964.

Hardy, M.J.; Kuczera, G., and Coombes, P.J. Integrated urban water cycle management: the UrbanCycle model. *Wat. Sci. & Tech.*, 52(9):1–9, 2005. doi: 10.2166/wst.2005.0276.

Harremoës, P. Integrated urban drainage, status and perspectives. *Water Science & Technology*, 45(3):1–10, 2002. URL `10.2166/wst.2002.0041`. IWA Publishing.

Havno, K.; Madsen, M.N., and Dorge, J. *MIKE 11 – A generalised river modelling package. In: Singh, V.P. (Ed.).* Computer Models of Watershed Hydrology. Water Resources Publications, 1995. pp. 733782. Colorado.

Heip, Lennart; Assel, Johan Van, and Swartentenbroekx, Patrick. Sewer flow quality modelling. *Water Science & Technology*, 36(5):177–184, 1997.

Hengl, T.; AghaKouchak, A., and Tadic, M. Percec. Methods and data sources for spatial prediction of rainfall. In Testik, Firat Y. and Gebremichael, Mekonnen, editors, *Rainfall: State of the Science*, Geophysical Monograph Series, page 287. American Geophysical Union (AGU), 2010.

Heuvelink, G. B. M.; Brown, J. D., and van Loon, E. E. A probabilistic framework for representing and simulating uncertain environmental variables. *International Journal of Geographical Information Science*, 21(5):497–513, 2007. ISSN 1365-8816. doi: 10.1080/13658810601063951.

Heuvelink, Gerard B M. *Error Propagation in Environmental Modelling with GIS.* Reserach Monographs in GIS. CRC Press Taylor & Francis Group, March 1998.

Hijmans, Robert J. *raster: Geographic Data Analysis and Modeling.* The Comprehensive R Archive Network, CRAN, 2019. URL `https://CRAN.R-project.org/package=raster`. R package version 2.9-23.

Hong, Yi; Liao, Qinzhuo; Bonhomme, Celine, and Chebbo, Ghassan. Physically-based urban stormwater quality modelling: An efficient approach for calibration and sensitivity analysis. *Journal Of Environmental Management*, 246:462–471, SEP 15 2019. ISSN 0301-4797. doi: 10.1016/j.jenvman.2019.06.003.

House, M. A.; Ellis, J. B.; Herricks, E. E.; Hvitved-Jacobsen, T.; Seager, J.; Lijklema, L.; Aalderink, H., and Clifforde, I. T. Urban drainage – impacts on receiving water quality. *Water Science and Technology*, 27(12):117, 1993. doi: 10.2166/wst.1993.0293. URL `http://dx.doi.org/10.2166/wst.1993.0293`.

House-Peters, Lily A. and Chang, Heejun. Urban water demand modeling: Review of concepts, methods, and organizing principles. *Water Resources Research*, 47, MAY 28 2011. ISSN 0043-1397. doi: {10.1029/2010WR009624}.

Huang, Haiming; Xiao, Xianming; Yan, Bo, and Yang, Liping. Ammonium removal from aqueous solutions by using natural Chinese (Chende) zeolite as adsorbent. *Journal of Hazardous Materials*, 175(1-3):247–252, 2010. ISSN 03043894. doi: 10.1016/j.jhazmat.2009.09.156.

Huber, Wayne C. and Dickinson, Robert E. *Storm Water Management Model, Version 4: User-s Manual.* U. S. Environmental Protection Agenc, Athens, Georgia, U.S.A., 1988.

Hutton, C; Vamvakeridou-Lyroudia, L; Kapelan, Z, and Savic, D. Uncertainty Quantification and Reduction in Urban

Water Systems (UWS) Modelling: Evaluation Report. Technical report, The European Commission, 2011.

IFAK, . SIMBA (Simulation of Biological Wastewater Systems): Manual and Reference. Technical report, Institut fÃ¼r Automation und Kommunikation e. V, Magdeburg, Germany, 2007.

Jerves-Cobo, R; Benedetti, L; Amerlinck, Y; Lock, K; De Mulder, C; Van Butsel, J; Cisneros, F; Goethals, P, and Nopens, I. Integrated ecological modelling for evidence-based determination of water management interventions in urbanized river basins: Case study in the Cuenca River basin (Ecuador). *Science of the Total Environment*, 709, 2020. doi: 10.1016/j.scitotenv.2019.136067.

Jin, Y. A comprehensive survey of fitness approximation in evolutionary computation. *Soft Computing*, 9(October 2003): 3–12, 2005. doi: 10.1007/s00500-003-0328-5.

Johnson, Marc T. J. and Munshi-South, Jason. Evolution of life in urban environments. *Science*, 358(6363), 2017. ISSN 0036-8075. doi: 10.1126/science.aam8327. URL `https://science.sciencemag.org/content/358/6363/eaam8327`.

Ju, Yang; Lindbergh, Sarah; He, Yiyi, and Radke, John D. Climate-related uncertainties in urban exposure to sea level rise and storm surge flooding: a multi-temporal and multi-scenario analysis. *Cities*, 92:230–246, SEP 2019. ISSN 0264-2751. doi: 10.1016/j.cities.2019.04.002.

Kalman, R. E. A new approach to linear filtering and prediction problems. *Transactions of the American Society of Mechanical Engineers: Journal of Basic Engineering*, 82D:35–45, 1960.

Kalos, Malvin H. and Whitlock, Paula A. *Monte Carlo Methods*. Wiley-Blackwell, 2 edition, 2008. ISBN 978-0-470-85494-5.

Kang, Gelin; Qiu, Yu; Wang, Qingxiu; Qi, Zuoda; Sun, Yuting, and Wang, Yuqiu. Exploration of the critical factors influencing the water quality in two contrasting climatic regions. *Environmental Science And Pollution Research*, 27 (11):12601–12612, APR 2020. doi: 10.1007/s11356-020-07786-5.

Katukiza, A. Y.; Ronteltap, M.; Niwagaba, C. B.; Kansiime, F., and Lens, P. N. L. Grey water characterisation and pollutant loads in an urban slum. *Int. J. Environ. Sci. Technol.*, 12(2):423436, Jan 2014. ISSN 1735-2630. doi: 10.1007/s13762-013-0451-5. URL `http://dx.doi.org/10.1007/s13762-013-0451-5`.

Kessler, M. M. Bibliographic coupling between scientific papers. *American Documentation*, 14(1):10–25, 1963. doi: 10.1002/asi.5090140103. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.5090140103`.

Kleidorfer, Manfred and Rauch, Wolfgang. An application of Austrian legal requirements for CSO emissions. *Water Science & Technology*, 64(5):1081, Sep 2011. ISSN 0273-1223. doi: 10.2166/wst.2011.560.

Knaus, Jochen. *Package "snowfall": Easier cluster computing (based on snow)*. The Comprehensive R Archive Network, CRAN, 1.84-6 edition, February 2015.

Kohler, M.; Zennegg, M.; Bogdal, C.; Gerecke, A.C.; Schmid, P.; Heeb, N.V.; Sturm, M.; Vonmont, H.; Kohler, H.-P.E., and Giger, W. Temporal trends, congener patterns, and sources of octa-, nona-, and decabromodiphenyl ethers (pbde) and hexabromocyclododecanes (hbcd) in swiss lake sediments. *Environmental Science and Technology*, 42(17):6378–6384, 2008. ISSN 0013936X. doi: 10.1021/es702586r.

Kollet, Stefan J.; Maxwell, Reed M.; Woodward, Carol S.; Smith, Steve; Vanderborght, Jan; Vereecken, Harry, and Simmer, Clemens. Proof of concept of regional scale hydrologic simulations at hydrologic resolution utilizing massively parallel computer resources. *Water Resources Research*, 46(4):1–7, 2010. ISSN 00431397. doi: 10.1029/2009WR008730.

Kuczera, George and Parent, Eric. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: The Metropolis algorithm. *Journal of Hydrology*, 211(1-4):69–85, 1998. ISSN 00221694. doi: 10.1016/S0022-1694(98)00198-X.

Lashford, Craig; Rubinato, Matteo; Cai, Yanpeng; Hou, Jingming; Abolfathi, Soroush; Coupe, Stephen; Charlesworth, Susanne, and Tait, Simon. SuDS & Sponge Cities: A Comparative Analysis of the Implementation of Pluvial Flood Management in the UK and China. *Sustainability*, 11(1), JAN 1 2019. ISSN 2071-1050. doi: 10.3390/su11010213.

Lawson, E.; Thorne, C.; Ahilan, S.; Allen, D.; Arthur, S.; Everett, G.; Fenner, R.; Glenis, V.; Guan, D.; Hoang, L.; Kilsby, C.; Lamond, J.; Mant, J.; Maskrey, S.; Mount, N.; Sleigh, A.; Smith, L., and Wright, N. Delivering and evaluating the multiple flood risk benefits in blue-green cities: an interdisciplinary approach. In *Flood Recovery, Innovation and Response IV*. WIT Press, June 2014. doi: 10.2495/friar140101. URL `https://doi.org/10.2495/friar140101`.

Leon, Javier X.; Heuvelink, Gerard B M, and Phinn, Stuart R. Incorporating DEM uncertainty in coastal inundation mapping. *PLoS ONE*, 9(9):1–12, 2014. ISSN 19326203. doi: 10.1371/journal.pone.0108727.

Leopold, Ulrich; Heuvelink, G. B M.; Tiktak, Aaldrik; Finke, Peter A., and Schoumans, Oscar. Accounting for change of support in spatial accuracy assessment of modelled soil mineral phosphorous concentration. *Geoderma*, 130(3-4): 368–386, 2006. ISSN 00167061.

Leta, Olkeba Tolessa; Nossent, Jiri; Velez, Carlos; Shrestha, Narayan Kumar; van Griensven, Ann, and Bauwens, Willy. Assessment of the different sources of uncertainty in a SWAT model of the river senne (Belgium). *Environmental Modelling and Software*, 68:129146, Jun 2015. ISSN 1364-8152. doi: 10.1016/j.envsoft.2015.02.010.

Litrico, Xavier and Fromion, Vincent. *Modeling and Control of Hydroystems*. Springer, 2009. ISBN 9788578110796. doi: 10.1017/CBO9781107415324.004.

Liu, Mingliang and Tian, Hanqin. China's land cover and land use change from 1700 to 2005: Estimations from high-resolution satellite data and historical archives. *Global Biogeochemical Cycles*, 24, JUL 16 2010. ISSN 0886-6236. doi: 10.1029/2009GB003687.

Lloyd, C.E.M.; Freer, J.E.; Johnes, P.J., and Collins, A.L. Using hysteresis analysis of high-resolution water quality monitoring data, including uncertainty, to infer controls on nutrient and sediment transfer in catchments. *Science of the Total Environment*, 543:388–404, 2016. ISSN 00489697. doi: 10.1016/j.scitotenv.2015.11.028.

Luetkepohl, Helmut. *New Introduction to Multiple Time Series Analysis*. Springer, 2005. ISBN 3-540-40172-5.

Lythcke-Jørgensen, Christoffer; Ensinas, Adriano Viana; Münster, Marie, and Haglind, Fredrik. A methodology for designing flexible multi-generation systems. *Energy*, 110:34–54, 2016. ISSN 03605442. doi: 10.1016/j.energy.2016.01.084. URL `http://dx.doi.org/10.1016/j.energy.2016.01.084`.

Maheepala, S.; Leighton, B.; Mirza, F.; Rahilly, M., and Rahman, J. Hydro Planner - a linked modelling system for water quantity and quality simulation of total water cycle. In Zerger, A. and Argent, R.M. (eds). *MODSIM 2005 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand*, pages 170 – 176, December 2005.

Mahmoodian, Mahmood; Torres-Matallana, Jairo Arturo; Leopold, Ulrich; Schutz, Georges, and Clemens, Francois H. L. R. A data-driven surrogate modelling approach for acceleration of short-term simulations of a dynamic urban drainage simulator. *Water*, 10(12), 2018. ISSN 2073-4441. doi: 10.3390/w10121849. URL `http://www.mdpi.com/2073-4441/10/12/1849`.

Mahmoodian, Mahmood; Torres-Matallana, J. A.; Leopold, Ulrich; Schutz, Georges, and Clemens, Francois. Emulation of a detailed urban drainage simulator to be applied for short-term predictions. In Mannina, Giorgio, editor, *New Trends in Urban Drainage Modelling*, pages 592–596, Cham, 2019. Springer International Publishing. ISBN 978-3-319-99867-1.

Mannina, Giorgio and Viviani, Gaspare. Receiving water quality assessment: comparison between simplified and detailed integrated urban modelling approaches. *Water Science & Technology*, 62(10):2301, Nov 2010. ISSN 0273-1223. doi: 10.2166/wst.2010.404.

Manz, Bastian Johann; Pablo Rodriguez, Juan; Maksimovic, Cedo, and McIntyre, Neil. Impact of rainfall temporal resolution on urban water quality modelling performance and uncertainties. *Water Science And Technology*, 68(1): 68–75, 2013. ISSN 0273-1223. doi: 10.2166/wst.2013.224.

Marelli, S. and Sudret, B. UQLab: A Framework for Uncertainty Quantification in Matlab. *American Society of Civil Engineers*, page 25542563, 2014. doi: 10.1061/9780784413609. ISBN 978-0-7844-1360-9.

Marwick, Ben and Krishnamoorthy, Kalimuthu. *cvequality: Tests for the Equality of Coefficients of Variation from Multiple Groups*, 2019. URL `https://github.com/benmarwick/cvequality`. R package version 0.2.0.

Maxwell, Reed M. A terrain-following grid transform and preconditioner for parallel, large-scale, integrated hydrologic modeling. *Advances in Water Resources*, 53:109–117, 2013. ISSN 03091708. doi: 10.1016/j.advwatres.2012.10.001.

McMillan, Hilary; Jackson, Bethanna; Clark, Martyn; Kavetski, Dmitri, and Woods, Ross. Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models. *Journal of Hydrology*, 400(1-2):8394, Mar 2011. ISSN 0022-1694. doi: 10.1016/j.jhydrol.2011.01.026. URL `http://dx.doi.org/10.1016/j.jhydrol.2011.01.026`.

Meirlaen, J.; Huyghebaert, B.; Sforzi, F.; Benedetti, L., and Vanrolleghem, P. Fast and simultaneous simulation of the integrated urban and wastewater system using mechanistic surrogate models. *Water Science and Technology*, 43 (7): 301–309, 2001. doi: 10.2166/wst.2001.0439.

Meirlaen, J.; Assel, J. Van, and Vanrolleghem, P.A. Real time control of the integrated urban wastewater system using simultaneously simulating surrogate models. *Water Science & Technology*, 45(3):109–116, 2002. doi: 10.2166/wst.2002.0066.

Miller, James David and Hess, Tim. Urbanisation impacts on storm runoff along a rural-urban gradient. *Journal of Hydrology*, 552:474–489, 2017. ISSN 00221694. doi: 10.1016/j.jhydrol.2017.06.025.

Miskewitz, Robert and Uchrin, Christopher. In-stream dissolved oxygen impacts and sediment oxygen demand resulting from combined sewer overflow discharges. *Journal of Environmental Engineering*, 139(10):1307–1313, 2013. doi: 10.1061/(ASCE)EE.1943-7870.0000739.

Mitchell, V.; Duncan, H.; Inman, M.; Rahilly M.and Stewart, J.; Vieritz, A.; Holt, P.; Grant, A.; Fletcher, T.; Coleman, J.; Maheepala, S.; Sharma, A.; Deletic, A., and Breen, P. State of the art review of integrated urban water models. In GRAIE, , editor, *Novatech 2007*, pages 507–514, Lyon, France, June 2007.

Mitchell, V.G. and Diaper, C. UVQ: A tool for assessing the water and contaminant balance impacts of urban development scenarios. *Wat. Sci. & Tech.*, 52 (12):91–98, 2005. doi: 10.2166/wst.2005.0435.

Mitchell, V.G.; Mein, R.G., and McMahon, T.A. Modelling the Urban Water Cycle. *Journal of Environmental Modelling & Software*, Vol. 16 (7):pp 615–629, 2001. doi: 10.1016/S1364-8152(01)00029-9.

Moreno-Rodenas, Antonio M.; Bellos, Vasilis; Langeveld, Jeroen G., and Clemens, Francois H.L.R. A dynamic emulator

for physically based flow simulators under varying rainfall and parametric conditions. *Water Research*, 142:512 – 527, 2018. ISSN 0043-1354. doi: https://doi.org/10.1016/j.watres.2018.06.011.

Moreno-Rodenas, Antonio M.; Tscheikner-Gratl, Franz; Langeveld, Jeroen G., and Clemens, Francois H.L.R. Uncertainty analysis in a large-scale water quality integrated catchment modelling study. *Water Research*, 158:46 – 60, 2019. ISSN 0043-1354. doi: 10.1016/j.watres.2019.04.016.

Moret, Stefano; Codina Gironès, Víctor; Bierlaire, Michel, and Maréchal, François. Characterization of input uncertainties in strategic energy planning models. *Applied Energy*, 202:597–617, 2017. ISSN 03062619.

Muleta, Misgana K.; McMillan, Jonathan; Amenu, Geremew G., and Burian, Steven J. 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, OCT 1 2013. ISSN 1084-0699. doi: 10.1061/(ASCE)HE.1943-5584.0000705.

Muthusamy, Manoranjan; Schellart, Alma; Tait, Simon, and Heuvelink, Gerard B.M. Geostatistical upscaling of rain gauge data to support uncertainty analysis of lumped urban hydrological models. *Hydrology and Earth System Sciences*, 21 (2):1077–1091, 2017. ISSN 16077938. doi: 10.5194/hess-21-1077-2017.

Mutzner, Lena; Vermeirssen, Etienne L. M., and Ort, Christoph. Passive samplers in sewers and rivers with highly fluctuating micropollutant concentrations - Better than we thought. *Journal Of Hazardous Materials*, 361:312–320, JAN 5 2019. ISSN 0304-3894. doi: 10.1016/j.jhazmat.2018.07.040.

MWH Soft, . *InfoWorks CS*. Innovyze, 2010. http://www.innovyze.com/products/infoworks_cs/.

Nash, J E and Sutcliffe, J E. River flow forecasting through conceptual models. Part 1 - a discussion of principles. *Journal of Hydrology (Amsterdam)*, 10:282–290, 1970.

Naves, J.; Rieckermann, J.; Cea, L.; Puertas, J., and Anta, J. Global and local sensitivity analysis to improve the understanding of physically-based urban wash-off models from high-resolution laboratory experiments. *Science Of The Total Environment*, 709, MAR 20 2020. ISSN 0048-9697. doi: 10.1016/j.scitotenv.2019.136152.

Neumann, Marc Benjamin. *Uncertainty Analysis for Performance Evaluation and Design of Urban Water Infrastructure*. PhD thesis, Swiss Federal Institute of Technology, ETH Zurich, 2007.

Nol, L.; Heuvelink, G. B M; Veldkamp, A.; de Vries, W., and Kros, J. Uncertainty propagation analysis of an N2O emission model at the plot and landscape scale. *Geoderma*, 159(1-2):9–23, 2010. ISSN 00167061. doi: 10.1016/j.geoderma.2010.06.009.

Noor, A.M.; Gething, P.W.; Alegana, V.A.; Patil, A.P.; Hay, S.I.; Muchiri, E.; Juma, E., and Snow, R.W. The risks of malaria infection in Kenya in 2009. *BMC Infectious Diseases*, 9, 2009. ISSN 14712334. doi: 10.1186/1471-2334-9-180.

Ochoa-Rodriguez, Susana; Wang, Li-Pen; Gires, Auguste; Pina, Rui Daniel; Reinoso-Rondinel, Ricardo; Bruni, Guendalina; Ichiba, Abdellah; Gaitan, Santiago; Cristiano, Elena; van Assel, Johan; Kroll, Stefan; Murla-Tuyls, Damian; Tisserand, Bruno; Schertzer, Daniel; Tchiguirinskaia, Ioulia; Onof, Christian; Willems, Patrick, and ten Veldhuis, Marie-Claire. Impact of spatial and temporal resolution of rainfall inputs on urban hydrodynamic modelling outputs: A multi-catchment investigation. *Journal Of Hydrology*, 531(2, SI):389–407, DEC 2015. ISSN 0022-1694. doi: 10.1016/j.jhydrol.2015.05.035.

O'Donnell, Emily; Thorne, Colin; Ahilan, Sangaralingam; Arthur, Scott; Birkinshaw, Stephen; Butler, David; Dawson, David; Everett, Glyn; Fenner, Richard; Glenis, Vassilis; Kapetas, Leon; Kilsby, Chris; Krivtsov, Vladimir; Lamond, Jessica; Maskrey, Shaun; O'Donnell, Greg; Potter, Karen; Vercruysse, Kim; Vilcan, Tudorel, and Wright, Nigel. The blue-green path to urban flood resilience. *Blue-Green Systems*, 2(1):28–45, December 2019. doi: 10.2166/bgs.2019.199. URL `https://doi.org/10.2166/bgs.2019.199`.

Oral, Hasan Volkan; Carvalho, Pedro; Gajewska, Magdalena; Ursino, Nadia; Masi, Fabio; van Hullebusch, Eric D.; Kazak, Jan K.; Exposito, Alfonso; Cipolletta, Giulia; Andersen, Theis Raaschou; Finger, David Christian; Simperler, Lena; Regelsberger, Martin; Rous, Vit; Radinja, Matej; Buttiglieri, Gianluigi; Krzeminski, Pawel; Rizzo, Anacleto; Dehghanian, Kaveh; Nikolova, Mariyana, and Zimmermann, Martin. A review of nature-based solutions for urban water management in european circular cities: a critical assessment based on case studies and literature. *Blue-Green Systems*, 2(1):112–136, January 2020. doi: 10.2166/bgs.2020.932. URL `https://doi.org/10.2166/bgs.2020.932`.

Pablo Rodriguez, Juan; McIntyre, Neil; Diaz-Granados, Mario; Achleitner, Stefan; Hochedlinger, Martin, and Maksimovic, Cedo. Generating time-series of dry weather loads to sewers. *Environmental Modelling & Software*, 43:133–143, MAY 2013. ISSN 1364-8152. doi: 10.1016/j.envsoft.2013.02.007.

Pebesma , E.J. and Bivand, R.S. Package "sp": classes and methods for spatial data in R. *R News*, 5(2), 2005. URL `https://cran.r-project.org/doc/Rnews/`.

Pebesma, Edzer. spacetime: Spatio-Temporal Data in R. *Journal of Statistical Software*, 51(7):1–30, 2012.

Pebesma, Edzer J. Multivariable geostatistics in S: the "gstat" package. *Computers & Geosciences*, 30:683–691, 2004. doi: 10.1016/j.cageo.2004.03.012.

Perraud, J. M; Kuczera, G., and Bridgart, R. J. Towards a Software Architecture to Facilitate Multiple Runs of Simulation Models. In Oxley, L and Kulasiri, D, , editor, *Modsim 2007: International Congress On Modelling And Simulation: Land, Water And Environmental Management: Integrated Systems For Sustainability*, pages 846–852. MODELLING &

SIMULATION SOC AUSTRALIA & NEW ZEALAND INC, 2007. ISBN 978-0-9758400-4-7. International Congress on Modelling and Simulation (MODSIM07), Christchurch, NEW ZEALAND, DEC 10-13, 2007.

Pianosia, F.; Sarrazin, F., and Wagener, T. A matlab toolbox for global sensitivity analysis. *Environmental Modelling & Software*, 70:8085, 2015. doi: http://dx.doi.org/10.1016/j.envsoft.2015.04.009. ISSN 1364-8152.

Pontee, Nigel; Narayan, Siddharth; Beck, Michael W., and Hosking, Adam H. Nature-based solutions: lessons from around the world. *Proceedings Of The Institution Of Civil Engineers-Maritime Engineering*, 169(1):29–36, MAR 2016. ISSN 1741-7597. doi: 10.1680/jmaen.15.00027.

Pye, Steve; Sabio, Nagore, and Strachan, Neil. An integrated systematic analysis of uncertainties in UK energy transition pathways. *Energy Policy*, 87:673–684, 2015. ISSN 03014215.

Qi, Yunfei; Chan, Faith Ka Shun; Thorne, Colin; O'Donnell, Emily; Quagliolo, Carlotta; Comino, Elena; Pezzoli, Alessandro; Li, Lei; Griffiths, James; Sang, Yanfang, and Feng, Meili. Addressing Challenges of Urban Water Management in Chinese Sponge Cities via Nature-Based Solutions. *Water*, 12(10), OCT 2020. doi: 10.3390/w12102788. URL `https://doi.org/10.3390/w12102788`.

R Core Team, . *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017. URL `https://www.R-project.org/`.

R-Core-Team, and contributors worldwide, . *The R Stats Package*. The R Project for Statistical Computing, 3.5.0 edition, 2017. URL `https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html`.

Rauch, W and Harremoës, P. On the potential of genetic algorithms in urban drainage modeling. *Urban Water*, 1(1):79–89, 1999. ISSN 14620758. doi: 10.1016/S1462-0758(99)00010-2.

Rauch, W.; Bertrand-Krajewski, J.-L.; Krebs, P.; Mark, O.; Schilling, W.; SchÃ¼tze, M., and Vanrolleghem, P.A. Deterministic modelling of integrated urban drainage systems. *Water Science and Technology*, 45 (3):81–94, 2002. doi: 10.2166/wst.2002.0059.

Rauch, W.; Bach, P.M.; Brown, R.; Deletic, A.; Ferguson, B.; De Haan, J.; Mccarthy, D.T.; Kleidorfer, M.; Tapper, N.; Sitzenfrei, R., and Urich, C. Modelling transition in urban drainage management. In: 9th International Conference on Urban Drainage Modelling. 2012. Belgrade, Serbia.

Rauch, W.; Bach, P. M.; Brown, R.; Deletic, A; de Haan, F.J.; Mair, M.; Kleidorfer, M.; Rogers, B.; Sitzenfrei, R., and Urich, C. Modelling transition in urban water systems. *Water Research*, 126, 2017. ISSN 00431354. doi: 10.1016/j.watres.2017.09.039.

Rawls, W.J.; Long, S.L., and McCuen, R.H. *Comparison of Urban Flood Frequency Procedures. Preliminary Draft Report prepared for the Soil Conservation Service*. Beltsville, Maryland, 1981.

Razavi, Saman; Tolson, Bryan A., and Burn, Donald H. Review of surrogate modeling in water resources. *Water Resources Research*, 48(7):1–32, 2012. ISSN 00431397. doi: 10.1029/2011WR011527.

Reckhow, KH. Importance of Scientific uncertainty in Decision-Making. *Environmental Management*, 18(2):161–166, MAR-APR 1994. ISSN 0364-152X. doi: 10.1007/BF02393758.

Refsgaard, J. C. Parameterisation, calibration and validation of distributed hydrological models. *J. Hydrol.*, (198, 69–97), 1997. doi: 10.1016/S0022-1694(96)03329-X.

Refsgaard, Jens Christian; van der Sluijs, Jeroen P.; Højberg, Anker Lajer, and Vanrolleghem, Peter A. Uncertainty in the environmental modelling process - A framework and guidance. *Environmental Modelling and Software*, 22(11): 1543–1556, 2007. ISSN 13648152. doi: 10.1016/j.envsoft.2007.02.004.

Renard, Benjamin; Kavetski, Dmitri; Kuczera, George; Thyer, Mark, and Franks, Stewart W. Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research*, 46(5), May 2010. ISSN 0043-1397. doi: 10.1029/2009wr008328.

Revolution Analytics, and Weston, Steve. *Package "iterators": Provides Iterator Construct for R*, 2015a. R package version 1.0.8.

Revolution Analytics, and Weston, Steve. *doParallel: Foreach parallel adaptor for the "parallel" package*, February 2015b. R package version 1.0.10.

Revolution Analytics, and Weston, Steve. *foreach: Provides Foreach Looping Construct for R*, 2015c. R package version 1.4.3.

Robert, Christian P. and Casella, George. *Introducing Monte Carlo Methods with R*. Springer, 2010. ISBN 978-1-4419-1576-4.

Rossman, L.A. In: Laboratory, n.r.r. (ed.), stormwater management model – user-s manual version 5.0. Technical report, US Environmental Protection Agency, Cincinnati, Ohio, 2004.

Ryan, Jeffrey A. and Ulrich, Joshua M. *xts: eXtensible Time Series*. The Comprehensive R Archive Network, CRAN, 2017. URL `https://CRAN.R-project.org/package=xts`. R package version 0.10-1.

Sage, J.; Bonhomme, C.; Al Ali, S., and Gromaire, M.-C. Performance assessment of a commonly used "accumulation and wash-off" model from long-term continuous road runoff turbidity measurements. *Water Research*, 78:47–59, 2015. ISSN 00431354. doi: 10.1016/j.watres.2015.03.030.

Salvadore, Elga; Bronders, Jan, and Batelaan, Okke. Hydrological modelling of urbanized catchments: A review and future directions. *Journal Of Hydrology*, 529(1):62–81, OCT 2015. ISSN 0022-1694. doi: 10.1016/j.jhydrol.2015.06.028.

Sandric, Ionut; Ionita, Cristian; Chitu, Zenaida; Dardala, Marian; Irimia, Radu, and Furtuna, Felix Titus. Using cuda to accelerate uncertainty propagation modelling for landslide susceptibility assessment. *Environmental Modelling & Software*, 115:176 – 186, 2019. ISSN 1364-8152. doi: 10.1016/j.envsoft.2019.02.016.

Sanitary-District, . Combined sewer overflow. Internet, May 2015. URL `https://www.richmondindiana.gov/resources/combined-sewer-overflow`.

Sawicka, Kasia; Heuvelink, Gerard B M, and Walvoort, Dennis. Spatial Uncertainty Propagation Analysis with the spup Package. *The R Journal*, 10/2:1–21, 2017. ISSN 2073-4859.

Schellart, A. N. A.; Tait, S. J., and Ashley, R. M. Towards quantification of uncertainty in predicting water quality failures in integrated catchment model studies. *Water Research*, 44(13):3893–3904, 2010. ISSN 00431354. doi: 10.1016/j.watres.2010.05.001.

Schueller, G. I. and Pradlwarter, H. J. Computational stochastic structural analysis (COSSAN) a software tool. *Structural Safety*, 28(12):6882, 2006. doi: 10.1016/j.strusafe.2005.03.005. ISSN 0167-4730.

Schutz, Georges; Fiorelli, David; Seiffert, Stefanie; Regneri, Mario, and Klepiszewski, Kai. Modelling and Optimal Control of a Sewer Network. In *9th International Conference on Urban Drainage Modelling*, 2012. Belgrade, Serbia.

Searle, S.R. *Linear Models*. Wiley, 1997. ISBN 9781118491782. doi: 10.1002/9781118491782.

Sexton, Joseph O.; Urban, Dean L.; Donohue, Michael J., and Song, Conghe. Long-term land cover dynamics by multi-temporal classification across the Landsat-5 record. *Remote Sensing Of Environment*, 128:246–258, JAN 2013. ISSN 0034-4257. doi: 10.1016/j.rse.2012.10.010.

Sharma, Ashok K.; Pezzaniti, David; Myers, Baden; Cook, Stephen; Tjandraatmadja, Grace; Chacko, Priya; Chavoshi, Sattar; Kemp, David; Leonard, Rosemary; Koth, Barbara, and Walton, Andrea. Water Sensitive Urban Design: An Investigation of Current Systems, Implementation Drivers, Community Perceptions and Potential to Supplement Urban Water Services. *Water*, 8(7), JUL 2016. ISSN 2073-4441. doi: 10.3390/w8070272.

Sin, Gürkan; Gernaey, Krist V.; Neumann, Marc B.; van Loosdrecht, Mark C M, and Gujer, Willi. Uncertainty analysis in WWTP model applications: A critical discussion using an example from design. *Water Research*, 43(11):2894–2906, 2009. ISSN 00431354. doi: 10.1016/j.watres.2009.03.048.

Snepvangers, J. J.J.C.; Heuvelink, G. B.M., and Huisman, J. A. Soil water content interpolation using spatio-temporal kriging with external drift. *Geoderma*, 112(3-4):253–271, 2003. ISSN 00167061. doi: 10.1016/S0016-7061(02)00310-5.

Somarakis, Giorgos; Stagakis, Stavros, and Chrysoulakis, Naktarios, editors. *ThinkNature Nature-Based Solutions Handbook*. 2019. doi: 10.26225/jerv-w202. ThinkNature project funded by the EU Horizon 2020 research and innovation programme under grant agreement No. 730338.

Sriwastava, Ambuj Kumar; Tait, Simon; Schellart, Alma; Kroll, Stefan; Dorpe, Mieke Van; Assel, Johan Van, and Shucksmith, James. Quantifying uncertainty in simulation of sewer overflow volume. *Journal of Environmental Engineering*, 144(7):04018050, 2018. doi: 10.1061/(ASCE)EE.1943-7870.0001392.

Sriwastava, Ambuj Kumar; Torres-Matallana, J.A.; Schellart, Alma; Leopold, Ulrich, and Tait, Simon. Implications of model uncertainty for investment decisions to manage intermittent sewer overflows. *Water Research*, page 116885, 2021. ISSN 0043-1354. doi: 10.1016/j.watres.2021.116885.

Statec, . Statistics portal of the grans-duchy of luxembourg, 2020. URL `https://statistiques.public.lu`.

Steinel, Anke and Margane, Armin. Best management practice guideline for wastewater facilities in karstic areas of Lebanon with special respect to the protection of ground- and surface waters. Technical Report 2, Federal Ministry for Economic Cooperation and Development, 2011. Project: Protection of Jeita Spring.

Sten, Johan; Lilja, Harri; Hyväluoma, Jari; Westerholm, Jan, and Aspnäs, Mats. Parallel flow accumulation algorithms for graphical processing units with application to rusle model. *Computers & Geosciences*, 89:88 – 95, 2016. ISSN 0098-3004. doi: 10.1016/j.cageo.2016.01.006.

Stephens, Timothy A. and Bledsoe, Brian P. Probabilistic mapping of flood hazards: Depicting uncertainty in streamflow, land use, and geomorphic adjustment. *Anthropocene*, 29, MAR 2020. ISSN 2213-3054. doi: 10.1016/j.ancene.2019.100231.

Stewardson, M.; McMahon, T., and Spears, M. Krakatoa: A Model to Assist Integrated Water Resource Management Decision-Making in Urban Areas. In *AWWA 16th Federal Convention*, 1995. Sydney, Australia.

Su, Shiliang; Li, Dan; Zhang, Qi; Xiao, Rui; Huang, Fang, and Wu, Jiaping. Temporal trend and source apportionment of water pollution in different functional zones of Qiantang River, China. *Water Research*, 45(4):1781–1795, FEB 2011. ISSN 0043-1354. doi: 10.1016/j.watres.2010.11.030.

Sun, Liqun; Chen, Ji; Li, Qinglan, and Huang, Dian. Dramatic uneven urbanization of large cities throughout the world in recent decades. *Nature Communications*, 11(1), OCT 23 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-19158-1.

Tavakol-Davani, Hassan; Rahimi, Reyhaneh; Burian, Steven J.; Pomeroy, Christine A.; McPherson, Brian J., and Apul, Defne. Combining Hydrologic Analysis and Life Cycle Assessment Approaches to Evaluate Sustainability of Water Infrastructure: Uncertainty Analysis. *Water*, 11(12), DEC 2019. doi: 10.3390/w11122592.

Ter Braak, Cajo J F. A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: Easy Bayesian computing for real parameter spaces. *Statistics and Computing*, 16(3):239–249, 2006. ISSN 09603174. doi: 10.1007/s11222-006-8769-1.

Tierney, Luke; Rossini, A. J.; Li, Na, and Sevcikova, H. *snow: Simple Network of Workstations*, 0.3-13 edition, 2016. R package version 0.4-2.

Tock, Laurence and Maréchal, François. Decision support for ranking Pareto optimal process designs under uncertain market conditions. *Computers and Chemical Engineering*, 83:165–175, 2015. ISSN 00981354. doi: 10.1016/j.compchemeng.2015.06.009.

Toffol, Sara De. *Sewer system performance assessment - an indicators based methodology*. PhD thesis, Universität Innsbruck, Universität Innsbruck, Innsbruck, Austria, 2006.

Torres-Matallana, J. A. and Leopold, U. Understanding model complexity and model accuracy through uncertainty analysis in urban water modelling. In IWA, , editor, *IWA World Water Congress & Exhibition 2018*, Tokyo, Japan, June 2018a.

Torres-Matallana, J. A. and Leopold, U. Uncertainty Propagation of Input Precipitation Across a Coupled Rainfall-Runoff Urban Drainage Model. In EMMS, , editor, *9th International Congress on Environmental Modeling and Software*, Colorado, USA, June 2018b.

Torres-Matallana, J. A.; Leopold, U., and Heuvelink, G.B.M. Uncertainty propagation and sensitivity analysis in urban hydrology water quality modelling. In Bailly, Jean-Stéphane; Griffith, Daniel, and Josselin, Didier, editors, *International Conference on Spatial Accuracy*, Montpellier, France, July 2016a. doi: 10.13140/RG.2.1.4811.4649.

Torres-Matallana, J. A.; Leopold, U., and Heuvelink, G.B.M. Uncertainty propagation in urban drainage modelling: a tool for environmental decision making. In UK, CIWEM, editor, *CIWEM UDG Autumn Conference & Exhibition. Embedding best practice and integrating innovation into Urban Drainage Management*, Blackpool, United Kingdom, November 2016b. doi: 10.13140/RG.2.1.4811.4649.

Torres-Matallana, J. A.; Cecinati, F.; Bellos, V., and Leopold, U. Merging and calibration of radar rain products for quantification of input uncertainty in urban drainage modelling for the Haute-Sûre catchment in Luxembourg. In Otjaques, Benoît and Hitzelberger, Patrik, editors, *EnviroInfo 2017: From Science to Society: The Bridge provided by Environmental Informatics*, Luxembourg, 2017a.

Torres-Matallana, J. A.; Leopold, U., and Heuvelink, G. B. M. Multivariate autoregressive modelling and conditional simulation of precipitation time series for urban water models. *European Water*, 57:299–306, 2017b. URL `https://library.wur.nl/WebQuery/wurpubs/fulltext/428142`.

Torres-Matallana, J. A.; Klepiszewski, K.; Leopold, U., and Heuvelink, G.B.M. EmiStatR: a simplified and scalable urban water quality model for simulation of combined sewer overflows. *Water*, 10(6)(782):1–24, 2018a. doi: 10.3390/w10060782. URL `https://www.mdpi.com/2073-4441/10/6/782`.

Torres-Matallana, J. A.; Leopold, U., and Heuvelink, G. B. M. Multivariate autoregressive modelling and conditional simulation for temporal uncertainty analysis of an urban water system in Luxembourg. *Hydrology and Earth System Sciences (HESS)*, 25(1):193–216, 2021. doi: 10.5194/hess-25-193-2021. URL `https://hess.copernicus.org/articles/25/193/2021/`.

Torres-Matallana, J.A. and Leopold, Ulrich. Geostatistical simulation of space-time stochastic rainfall fields for uncertainty propagation in rainfall-runoff and urban drainage system modelling. In Statistics, Spatial, editor, *Spatial Statistics 2019: Towards Spatial Data Science*, Sitges, Spain, 2019.

Torres-Matallana, J.A.; Leopold, U., and Heuvelink, G.B.M. stUPscales: an R-package for spatio-temporal Uncertainty Propagation across multiple scales with examples in urban water modelling. *Water*, 10(7)(837):1–30, 2018b. doi: 10.3390/w10070837. URL `https://www.mdpi.com/2073-4441/10/7/837`.

Torres-Matallana, J.A.; Leopold, U., and Heuvelink, G.B.M. *stUPscales: Spatio-Temporal Uncertainty Propagation Across Multiple Scales*, 2019a. URL `https://CRAN.R-project.org/package=stUPscales`. R package version 1.0.5.0.

Torres-Matallana, J.A.; Leopold, Ulrich, and Fishbain, Barak. Geostatistical space-time prediction of air quality pollutants in Israel. In Statistics, Spatial, editor, *Spatial Statistics 2019: Towards Spatial Data Science*, Sitges, Spain, 2019b.

Trogrlić, R. S.; Rijke, J.; Dolman, N., and Zevenbergen, C. Rebuild by design in Hoboken: a design competition as a meansfor achieving flood resilience of urban areas through the implementation of green infrastructure. *Water*, 10(5), SEP 2018. doi: 10.3390/w10091230.

Tscheikner-Gratl, Franz; Lepot, Mathieu; Moreno-Rodenas, Antonio, and Schellart, Alma. Quics deliverable 6.7: A framework for the application of uncertainty analysis. Technical report, Delft University of Technology, University of Sheffield

and CH2M, 2017. URL https://zenodo.org/record/1240926.

Tscheikner-Gratl, Franz; Bellos, Vasilis; Schellart, Alma; Moreno-Rodenas, Antonio; Muthusamy, Manoranjan; Langeveld, Jeroen; Clemens, Francois; Benedetti, Lorenzo; Rico-Ramirez, Miguel Angel; de Carvalho, Rita Fernandes; Breuer, Lutz; Shucksmith, James; Heuvelink, Gerard B.M., and Tait, Simon. Recent insights on uncertainties present in integrated catchment water quality modelling. *Water Research*, 150:368 – 379, 2019. ISSN 0043-1354. doi: 10.1016/j.watres.2018.11.079.

United Nations - UN, . Transforming our world: The 2030 agenda for sustainable development. Technical report, United Nations - UN, 2015.

United Nations - UN, . World Urbanization Prospects. The 2018 Revision. Technical report, United Nations Department of Economic & Social Affairs, 2018. URL https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf.

Urbanek, Simon. *multicore: Parallel processing of R code on machines with multiple cores or CPUs*, 2013. R package version 0.1-7.

van Daal-Rombouts, Petra; Sun, Siao; Langeveld, Jeroen; Bertrand-Krajewski, Jean-Luc, and Clemens, Francois. Design and performance evaluation of a simplified dynamic model for combined sewer overflows in pumped sewer systems. *Journal Of Hydrology*, 538:609–624, JUL 2016. ISSN 0022-1694. doi: 10.1016/j.jhydrol.2016.04.056.

van der Keur, P.; Henriksen, H. J.; Refsgaard, J. C.; Brugnach, M.; Pahl-Wostl, C.; Dewulf, A., and Buiteveld, H. Identification of major sources of uncertainty in current IWRM practice. Illustrated for the Rhine Basin. *Water Resources Management*, 22(11):1677–1708, 2008. ISSN 09204741. doi: 10.1007/s11269-008-9248-6.

Vanguelova, E. I.; Bonifacio, E.; De Vos, B.; Hoosbeek, M. R.; Berger, T. W.; Vesterdal, L.; Armolaitis, K.; Celi, L.; Dinca, L.; Kjønaas, O. J.; Pavlenda, P.; Pumpanen, J.; Püttsepp, ; Reidy, B.; Simončič, P.; Tobin, B., and Zhiyanski, M. Sources of errors and uncertainties in the assessment of forest soil carbon stocks at different scales - review and recommendations. *Environmental Monitoring and Assessment*, 188(11), 2016. ISSN 15732959. doi: 10.1007/s10661-016-5608-5.

Vanhooren, H.; Meirlaen, J.; Amerlinck, Y.; Claeys, F.; Vangheluwe, H., and Vanrolleghem, P.A. WEST: modelling biological wastewater treatment. *J. Hydroinform.*, 5 (1):27–50, 2003. doi: 10.2166/hydro.2003.0003.

Vanrolleghem, P. A.; Benedetti, L., and Meirlaen, J. Modelling and real-time control of the integrated urban wastewater system. *Environmental Modelling and Software*, 20(4 SPEC. ISS.):427–442, 2005. ISSN 13648152. doi: 10.1016/j.envsoft.2004.02.004.

Vezzaro, L.; Christensen, M. L.; Thirsing, C.; Grum, M., and Mikkelsen, P. S. Water quality-based real time control of integrated urban drainage systems: A preliminary study from Copenhagen, Denmark. *Procedia Engineering*, 70:1707–1716, 2014. ISSN 18777058. doi: 10.1016/j.proeng.2014.02.188.

Viana da Silva, Alexandra M E; Bettencourt da Silva, Ricardo J N, and Camões, M. Filomena G F C. Optimization of the determination of chemical oxygen demand in wastewaters. *Analytica Chimica Acta*, 699(2):161–169, 2011. ISSN 00032670. doi: 10.1016/j.aca.2011.05.026.

Vrugt, J. A.; Gupta, H. V.; Bouten, W., and Sorooshian, S. A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research*, 39(8), 2003a. doi: 10.1029/2002WR001642.

Vrugt, J. A.; ter Braak, C. J. F.; Diks, C. G. H.; Robinson, B. A.; Hyman, J. M., and Higdon, D. Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation*, 10(3):273–290, 2009. doi: 10.1515/IJNSNS.2009.10.3.273.

Vrugt, Jasper A. and Robinson, Bruce A. Improved evolutionary optimization from genetically adaptive multimethod search. *Proceedings of the National Academy of Sciences of the United States of America*, 2007. doi: 10.1073/pnas.0610471104.

Vrugt, Jasper A.; Gupta, Hoshin V.; Bastidas, Luis A.; Bouten, Willem, and Sorooshian, Soroosh. Effective and efficient algorithm for multiobjective optimization of hydrologic models. *Water Resources Research*, 39(8), Aug 2003b. ISSN 0043-1397. doi: 10.1029/2002wr001746.

Vrugt, Jasper A.; ter Braak, Cajo J. F.; Clark, Martyn P.; Hyman, James M., and Robinson, Bruce A. Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resources Research*, 44:1–15, 2008. ISSN 0043-1397. doi: 10.1029/2007WR006720.

Wadoux, Alexandre M.J.-C.; Heuvelink, Gerard B.M.; Uijlenhoet, Remko, and de Bruin, Sytze. Optimization of rain gauge sampling density for river discharge prediction using bayesian calibration. *PeerJ*, 8:e9558, July 2020. ISSN 2167-8359. doi: 10.7717/peerj.9558.

Wadoux, A.M.J.-C.; Brus, D.J.; Rico-Ramirez, M.A., and Heuvelink, G.B.M. Sampling design optimisation for rainfall prediction using a non-stationary geostatistical model. *Advances in Water Resources*, 107:126–138, 2017. doi: 10.1016/j.advwatres.2017.06.005.

Walker, W.E.; Harremoës, P.; Rotmans, J.; van der Sluijs, J.P.; van Asselt, M.B.A.; Janssen, P., and Krayer von Krauss, M.P. Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support. *Integrated Assessment*, 4(1):5–17, mar 2003. ISSN 1389-5176. doi: 10.1076/iaij.4.1.5.16466.

Wang, Hao; Mei, Chao; Liu, JiaHong, and Shao, WeiWei. A new strategy for integrated urban water management in China: Sponge city. *Science China-Technological Sciences*, 61(3):317–329, MAR 2018. ISSN 1674-7321. doi: 10.1007/s11431-017-9170-5.

Webster, R. and Heuvelink, G. B. M. The Kalman filter for the pedologist's tool kit. *European Journal of Soil Science*, 57 (6):758773, Dec 2006. ISSN 1365-2389. doi: 10.1111/j.1365-2389.2006.00879.x.

Webster, R. and Oliver, M.A. *Geostatistics for Environmental Scientists*. Wiley, 2nd edition, 2007. ISBN 978-0-470-02858-2.

Welker, A. Emissions of pollutant loads from combined sewer systems and separate sewer systems – which sewer system is better. In ICUD, , editor, *11th International Conference on Urban Drainage*, Edinburgh, Scotland, UK, 2008.

Wijesiri, Buddhi; Egodawatta, Prasanna; McGree, James, and Goonetilleke, Ashantha. Assessing uncertainty in pollutant build-up and wash-off processes. *Environmental Pollution*, 212:48–58, 2016. ISSN 18736424. doi: 10.1016/j.envpol.2016.01.051.

Wijesiri, Buddhi; Egodawatta, Prasanna; McGree, James, and Goonetilleke, Ashantha. Understanding the uncertainty associated with particle-bound pollutant build-up and wash-off: A critical review. *Water Research*, 101:582–596, SEP 15 2016. ISSN 0043-1354. doi: 10.1016/j.watres.2016.06.013.

Willems, P. Random number generator or sewer water quality model? *Water Science & Technology*, 54(6-7):387, Oct 2006. ISSN 0273-1223. doi: 10.2166/wst.2006.581.

Willems, P.; Arnbjerg-Nielsen, K.; Olsson, J., and Nguyen, V. T. V. Climate change impact assessment on urban rainfall extremes and urban drainage: Methods and shortcomings. *Atmospheric Research*, 103(SI):106–118, JAN 2012. ISSN 0169-8095. doi: 10.1016/j.atmosres.2011.04.003. 8th International Workshop on Precipitation in Urban Areas, St. Moritz, Switzerland, Dec 10-13, 2009.

Willems, Patrick. Parsimonious model for combined sewer overflow pollution. *Journal of Environmental Engineering. American Society of Civil Engineers (ASCE)*, 136(3), March 2010. doi: 10.1061/(ASCE)EE.1943-7870.0000151.

Williams, Matthew; Cornford, Dan; Bastin, Lucy, and Pebesma, Edzer. Uncertainty Markup Language (UnCertML). Technical report, The Open Geospatial Consortium Inc., 2009.

Yu, L; Rozemeijer, J C; Broers, H P; van Breukelen, B M; Middelburg, J J; Ouboter, M, and van der Velde, Y. Drivers of nitrogen and phosphorus dynamics in a groundwater-fed urban catchment revealed by high frequency monitoring. *Hydrology and Earth System Sciences Discussions*, 2020:1–22, 2020. doi: 10.5194/hess-2020-34. URL `https://www.hydrol-earth-syst-sci-discuss.net/hess-2020-34/`.

Zevenbergen, Chris; Fu, Dafang, and Pathirana, Assela. Transitioning to Sponge Cities: Challenges and Opportunities to Address Urban Water Problems in China. *Water*, 10(9), SEP 2018. doi: 10.3390/w10091230.

Zhou, X; Polcher, J; Yang, T, and Huang, C.-S. A new uncertainty estimation approach with multiple datasets and implementation for various precipitation products. *Hydrology and Earth System Sciences*, 24(4):2061–2081, 2020. doi: 10.5194/hess-24-2061-2020. URL `https://www.hydrol-earth-syst-sci.net/24/2061/2020/`.

Zoogman, P.; Liu, X.; Suleiman, R. M.; Pennington, W. F.; Flittner, D. E.; Al-Saadi, J. A.; Hilton, B. B.; Nicks, D. K.; Newchurch, M. J.; Carr, J. L.; Janz, S. J.; Andraschko, M. R.; Arola, A.; Baker, B. D.; Canova, B. P.; Miller, C. Chan; Cohen, R. C.; Davis, J. E.; Dussault, M. E.; Edwards, D. P.; Fishman, J.; Ghulam, A.; Abad, G. Gonzalez; Grutter, M.; Herman, J. R.; Houck, J.; Jacob, D. J.; Joiner, J.; Kerridge, B. J.; Kim, J.; Krotkov, N. A.; Lamsal, L.; Li, C.; Lindfors, A.; Martin, R. V.; McElroy, C. T.; McLinden, C.; Natraj, V.; Neil, D. O.; Nowlan, C. R.; O'Sullivan, E. J.; Palmer, P. I.; Pierce, R. B.; Pippin, M. R.; Saiz-Lopez, A.; Spurr, R. J. D.; Szykman, J. J.; Torres, O.; Veefkind, J. P.; Veihelmann, B.; Wang, H.; Wang, J., and Chance, K. Tropospheric emissions: Monitoring of pollution (TEMPO). *Journal Of Quantitative Spectroscopy & Radiative Transfer*, 186(SI):17–39, JAN 2017. ISSN 0022-4073. doi: 10.1016/j.jqsrt.2016.05.008.

Zoppou, Christopher. Review of urban storm water models. *Environmental Modelling and Software*, 16(3):195–231, 2001. ISSN 13648152. doi: 10.1016/S1364-8152(00)00084-0.

Appendix A.
Supplement of Chapter 5

This Appendix is the supplement of Chapter 5, based on:

Torres-Matallana, J. A.; Leopold, U., and Heuvelink, G. B. M. Multivariate autoregressive modelling and conditional simulation for temporal uncertainty analysis of an urban water system in Luxembourg. *Hydrology and Earth System Sciences (HESS)*, 25(1):193–216, 2021. doi: 10.5194/hess-25-193-2021. URL `https://hess.copernicus.org/articles/25/193/2021/`

A.1 Model inputs

Table A.1 presents the most important general CSO input and output variables of EmiStatR. Table A.2 presents the base values of input variables for each individual CSO.

Table A.1: Most important general, CSO input and output variables of EmiStatR, with base values for the general input variables.

General input	Units	Base value	CSO input	Units
1. Wastewater			*1. Catchment data*	
Water consumption, qs	L(PE[a] d)$^{-1}$	152	Total area, A_{total}	ha
Pollution COD[b], $C_{\mathrm{COD,S}}$	g PE^{-1} d^{-1}	104.2	Impervious area, A_{imp}	ha
Pollution NH_4[c], $C_{\mathrm{NH4,S}}$	g PE^{-1} d^{-1}	4.7	Run-off coeff.[d] for impervious area, C_{imp}	–
2. Infiltration water			Run-off coeff. for pervious area, C_{per}	–
Inflow, q_{f}	L s^{-1} ha^{-1}	0.116	Flow time structure, t_{fS}	[time step]
Pollution COD, COD_f	g PE^{-1} d^{-1}	0	Population equivalents, pe	PE
Pollution NH_4, $\mathrm{NH_{4_f}}$	g PE^{-1} d^{-1}	0	*2. CSO structure data*	
3. Rainwater			Volume, V	m^3
Rain time series, P	mm		Initial water level, $\mathrm{Lev_{ini}}$	m
Pollution COD, COD_r	mg L^{-1}	71.0	Maximum throttled outflow, $Q_{\mathrm{d,\ max}}$	L s^{-1}
Pollution NH_4, $\mathrm{NH4_r}$	mg L^{-1}	2.0	Orifice diameter, D_{d}	m
			Orifice coefficient of discharge, C_{d}	–
Output variables				
1. Quantity				
Volume in the CSO chamber, V_{Chamber}	m^3			
Overflow spill volume, V_{Sv}	m^3			
Overflow spill flow, Q_{Sv}	L s^{-1}			
2. Quality				
Spill COD load, $B_{\mathrm{COD,Sv}}$	g			
Average spill COD conc.[e], $C_{\mathrm{COD,Sv,av}}$	mg L^{-1}			
99.9th perc.[f] spill COD conc., $C_{\mathrm{COD,Sv,99.9}}$	mg L^{-1}			
Maximum overflow COD conc., $C_{\mathrm{COD,Sv,max}}$	mg L^{-1}			
Spill NH_4 load, $B_{\mathrm{NH4,Sv}}$	g			
Average spill NH_4 conc., $C_{\mathrm{NH4,Sv,av}}$	mg L^{-1}			
99.9th perc. spill NH_4 conc., $C_{\mathrm{NH4,Sv,99.9}}$	mg L^{-1}			
Maximum spill NH_4 conc., $C_{\mathrm{NH4,Sv,max}}$	mg L^{-1}			

[a]PE = population equivalents; [b]COD = chemical oxygen demand; [c]NH_4 = ammonium;
[d]coef. = coefficient; [d]conc. = concentration; [f]perc. = percentile.

Table A.2: The CSO structure input data for the EmiStatR model, after calibration. Structures 2 and 3, only C_d was calibrated.

CSO input			
1. Identification			
ID of the structure	1	2	3
Name of the structure	FBH Goesdorf	FBN Kaundorf	FBH Nocher-Route
2. Catchment data			
Name of the municipality	Goesdorf	Kaundorf	Nocher-Route
Name of the catchment	Haute-Sûre	Haute-Sûre	Haute-Sûre
Number of the catchment	1	1	1
Land Use[a]	R/I	R/I	R/I
Total area, A_{total} (ha)	30.0	22	18.6
Impervious area, A_{imp} (ha)	5.0	11.0	4.3
Run-off coefficient for impervious area, C_{imp} $(-)$	0.28	0.3	0.3
Run-off coefficient for pervious area, C_{per} $(-)$	0.07	0.10	0.10
Flow time structure, t_{fS} (time step)	1	2	2
Population equivalents, pe (PE)	611	358	326
3. CSO structure data			
Volume, V (m^3)	190	180	157
Initial water level, Lev_{ini} (m)	0.57	1.8	1.8
Maximum throttled outflow, $Q_{\text{d,max}}$ ($L\ s^{-1}$)	5.0	9	4
Orifice diameter, D_{d} (m)	0.15	0.20	–
Orifice coefficient of discharge, C_{d} $(-)$	0.67	0.67	0.67

[a] R = residential, I = industrial.

A.2 Selection of model inputs for uncertainty quantification

Regarding the selection of model inputs for uncertainty quantification, to better support our decisions we include Figure A.1, as in Tscheikner-Gratl et al. (2017). Table A.3 shows comparisons of the means and variances for $C_{\text{COD,S}}$ and $C_{\text{NH4,S}}$ based on 91 measurements in the Haute-Sure catchment and simulations at Goesdorf (note that for COD_{r} a comparison could not be made because we had no measurements of COD_{r} and a model for COD_{r} was based on expert judgement). The agreement between observed and simulated statistics is again quite close. We could not evaluate the autocorrelation functions of the observed $C_{\text{COD,S}}$ and $C_{\text{NH4,S}}$ because there were too few observations to be able to compute these (note that the 91 observations were from multiple locations within the catchment).

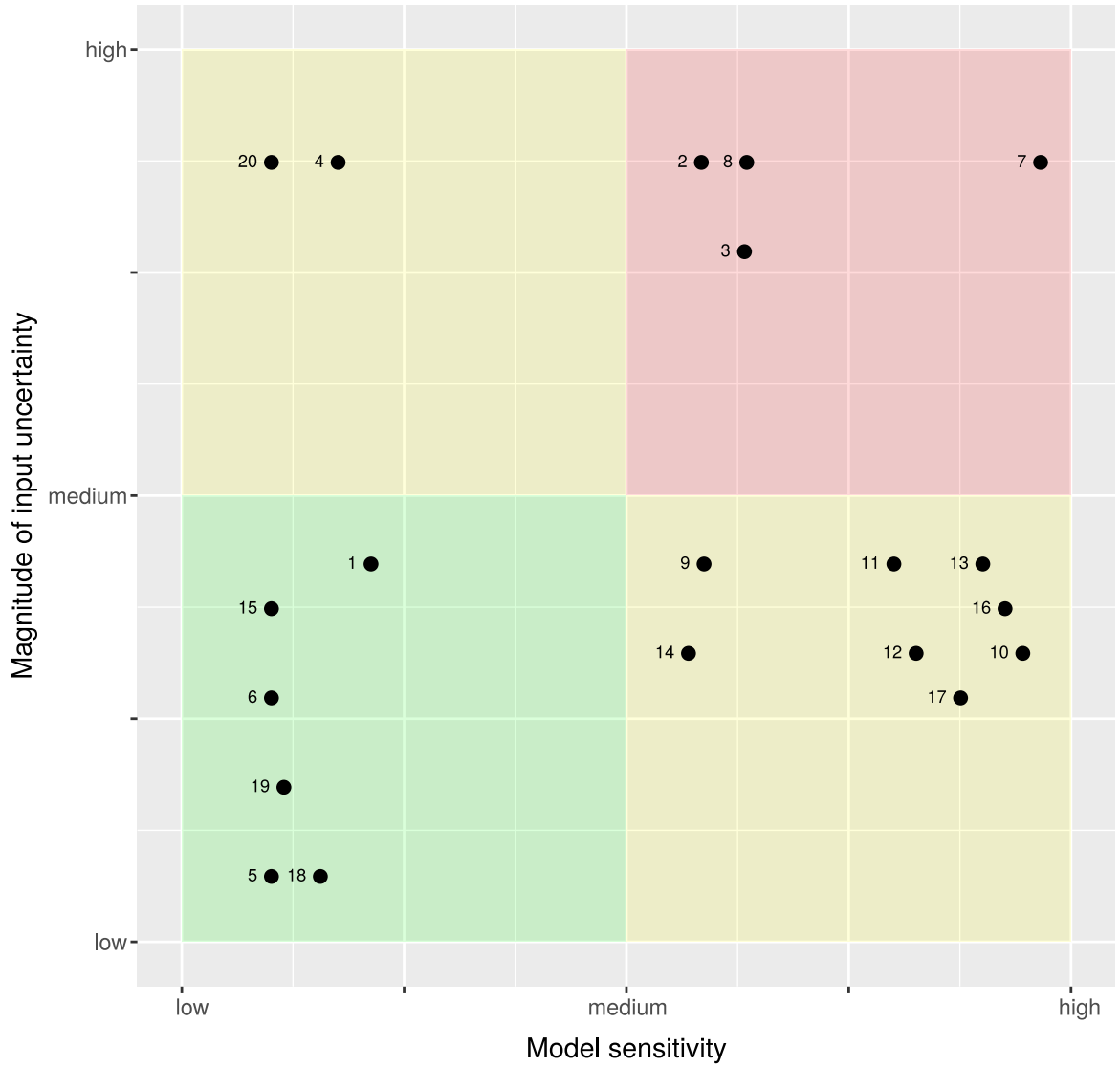


Figure A.1: Graphical assessment of the contribution of input uncertainty to model output uncertainty. Numbers near each dot refer to the input variable number as defined in Table 2 of the manuscript. Panel layout after Tscheikner-Gratl et al. (2017).

Table A.3: Mean and variance of log-transformed observed $C_{\mathrm{COD,S}}$ and $C_{\mathrm{NH4,S}}$ in the Haute-Sure catchment and of log-transformed simulated $C_{\mathrm{COD,S}}$ and $C_{\mathrm{NH4,S}}$ at Goesdorf (random selection of simulation numbers 1, 750, and 1500).

	Observations	Sim 1	Sim 750	Sim 1500	Sims (All)
Mean ($\log(C_{\mathrm{COD,S}})$)	4.3783	4.3752	4.3737	4.4106	4.3780
Variance ($\log(C_{\mathrm{COD,S}})$)	0.5637	0.5261	0.5257	0.5394	0.5640
Mean ($\log(C_{\mathrm{NH4,S}})$)	1.4733	1.4656	1.4639	1.4865	1.4730
Variance ($\log(C_{\mathrm{NH4,S}})$)	0.1679	0.1704	0.1684	0.1615	0.1681

A.3 Model input characterisation and observations

The simulated precipitation time series captured the main statistics of the observed time series well. Evidence for this is presented in Table A.4 and Figure SA.2.

Table A.4: Mean and variance of the log-transformed observed precipitation time series at Esch-sur-Sure and Dahl rain gauges and the simulated precipitation time series at Goesdorf (random selection of simulation numbers 1, 750, 1500 and all).

	Esch-sur-Sure	Dahl	Sim 1	Sim 750	Sim 1500	Sims (All)
Mean	-6.6152	-6.5817	-6.3888	-6.3886	-6.3878	-6.3874
Variance	1.4188	1.5731	1.5636	1.5579	1.5594	1.5582

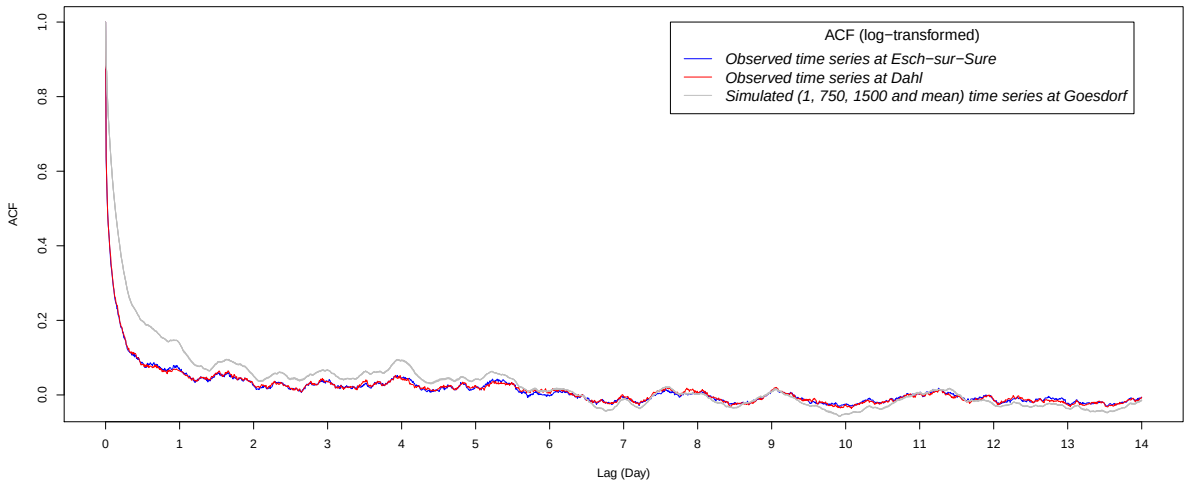


Figure A.2: Autocorrelation function of the log-transformed observed precipitation time series at Esch-sur-Sûre and Dahl rain gauges and simulated precipitation at Goesdorf catchment.

A.4 Input precipitation model: calibration and conditional simulation

Calibration

We begin by relating the two precipitation time series as:

$$P_1(t) = P_2(t) \cdot \delta(t) \tag{A.1}$$

where $\delta(t)$ is a positive multiplicative factor that varies over time. We assume that $P_1(t)$, $P_2(t)$ and $\delta(t)$ are stationary and log-normally distributed stochastic processes. After log-transformation we get

$$\log(P_1(t)) = \log(P_2(t)) + \log(\delta(t)) \tag{A.2}$$

We apply a Kernel (daniell) smoothing to the precipitation time series to avoid rapid fluctuation of the time series for precipitation depth values smaller than 0.1 mm. This also solves problems associated with taking logarithms of near-zero values. Next, in order to estimate the parameters of $\delta(t)$, we filter the time series allowing the computation of a ratio between the two measured time series. This ratio represents the difference in precipitation as registered in two nearby rain gauge stations. It is computed only for those cases where the precipitation depth of the two time series is greater than 0.01 mm.

To simplify notation we write $LP_1(t) = \log(P_1(t))$, $LP_2(t) = \log(P_2(t))$ and $L\delta(t) = \log(\delta(t))$. Since two out of three determine the third, we need only define two processes. We model the joint distribution of $LP_1(t)$ and $L\delta(t)$ by a bivariate AR(1) process, as introduced before:

$$\begin{bmatrix} LP_1(t+1) \\ \\ L\delta(t+1) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \\ \mu_\delta \end{bmatrix} + \begin{bmatrix} B_{11} & B_{12} \\ \\ B_{21} & B_{22} \end{bmatrix} \left(\begin{bmatrix} LP_1(t) \\ \\ L\delta(t) \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \\ \mu_\delta \end{bmatrix} \right) + \begin{bmatrix} \varepsilon_1(t+1) \\ \\ \varepsilon_\delta(t+1) \end{bmatrix} \tag{A.3}$$

where ε_1 and ε_δ are zero-mean, cross-correlated and normally distributed white noise processes.

To calibrate this model, i.e. estimate its parameters μ_1, μ_δ, B_{11}, B_{12}, B_{21}, B_{22}, σ_1^2, σ_δ^2 and $\rho_{1\delta}$, where $\sigma_1^2 = \mathrm{var}(\varepsilon_1)$, $\sigma_\delta^2 = \mathrm{var}(\varepsilon_\delta)$ and $\rho_{1\delta}$ is the correlation between ε_1 and ε_δ, we used the R package mAr (Barbosa, 2015). Calibration is based on two time series of LP_1 and $L\delta$ derived from observed time series P_1 and P_2.

Conditional simulation

To simulate a time series P for the target catchment from an observed time series P_o at a nearby location, we make use of the fact that the calibrated AR(1) model quantifies how precipitation at one location relates to that at a nearby location. We make use of Eq. A.1:

$$P(t) = P_o(t) \cdot \delta(t) \tag{A.4}$$

This requires simulations of $\delta(t)$. These are obtained using the calibrated model Eq. A.3, but now applied to the vector $[LP_o \;\; L\delta]^T$, which characterises the joint pdf of LP_o and $L\delta$. We use this model to simulate $L\delta$ conditional to the observed time series LP_o. Since the two processes are jointly normally distributed we can make use of a well known property of the multivariate normal distribution (Searle, 1997, page 47). Let U and V be two jointly normally distributed random vectors. The conditional distribution of U given $V = v$ is then also normal and given by:

$$\{U|V = v\} \sim N\left(E[U] + cov(U,V) \cdot var(V)^{-1} \cdot (v - E[V]),\; var(U) - cov(U,V) \cdot var(V)^{-1} \cdot cov(V,U)\right) \tag{A.5}$$

We make use of this equation to simulate δ by substituting:

$$U = L\delta(t+1) \qquad V = \begin{bmatrix} L\delta(t) \\ LP_o(t+1) \\ LP_o(t) \end{bmatrix} \tag{A.6}$$

for all $t = 1, ..., T$, while substituting the observed time series LP_o for v.

For more details we refer to Torres-Matallana et al. (2017b).

A.5 Uncertainty quantification of selected model input

Figure A.3 presents the histogram of observations, empirical density and theoretical normal density for $\log(C_{\mathrm{COD,S}})$, $\log(C_{\mathrm{NH4,S}})$ and $\log(\mathrm{COD_r})$.

A.6 Monte Carlo simulation size and timing

In order to perform the MC propagation analysis, we first did a convergence test to estimate the number of simulations required. Besides this test, we also computed the standard error of all MC outputs. These two methods have the same aim and are closely related. In the convergence test, the standard deviation of two different MC simulations with different random seeds were computed and compared for the seven output variables of EmiStatR, three representing water quantity variables (V_{Chamber}, V_{Sv} and Q_{Sv}) and four

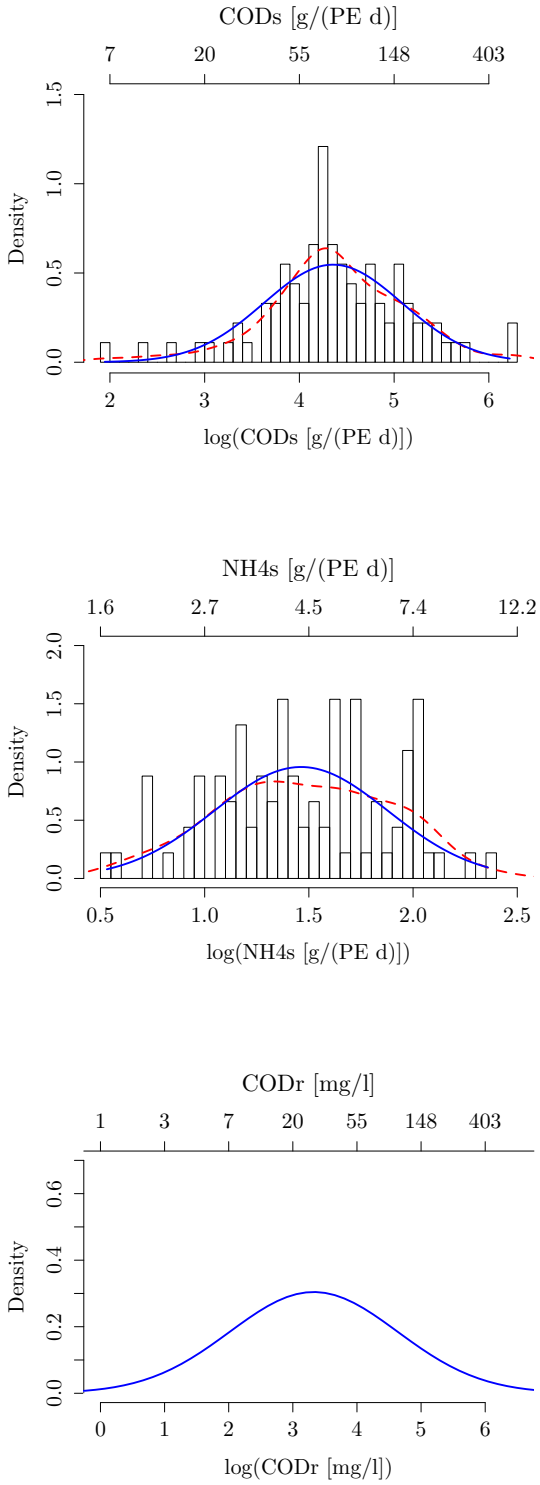


Figure A.3: Histogram of observations, empirical density (red dashed line) and theoretical normal density (blue line) for (a) $\log(C_{\mathrm{COD,S}})$; (b) $\log(C_{\mathrm{NH4,S}})$; (c) $\log(\mathrm{COD_r})$. Note that the blue densities were used in the uncertainty propagation.

Table A.5: Average running time in minutes for Monte Carlo (MC) replications and specific cores used with two different seeds for the pseudo-random number generator in R. The rainfall input used was a one-month length time series with 10 minutes time steps from 1 to 31 August 2010 (4,464 time steps).

Replications	250	500	1000	1500	2000
cores	3	3	50	50	50
MC1	7.12	14.23	4.84	7.33	9.40
MC2	7.09	14.63	4.96	7.26	9.53
Average	7.10	14.43	4.90	7.29	9.46

for water quality ($B_{\mathrm{COD,Sv}}$, $B_{\mathrm{NH4,Sv}}$, $C_{\mathrm{COD,Sv,av}}$, and $C_{\mathrm{NH4,Sv,av}}$). The results of the test indicated that in most cases between 250 and 1,000 MC simulations are enough to reach stable results in terms of the Nash–Sutcliffle model efficiency coefficient (NSE), where a NSE of 1 means a perfect match between observations and model output. In this case we got a NSE $\approx$ 0.998 for overflow volume. Regarding the water quality variable $B_{\mathrm{COD,Sv}}$, the test showed that a larger number of MC simulations is required. Between 1,000 and 2,000 simulations are required to reach stable results (NSE $\approx$ 0.880 for overflow COD load and 0.998 for overflow NH_4 load). Therefore, a number of 1,500 MC simulations was used to perform the uncertainty analysis of the water quantity and water quality outputs. Figure A.4 illustrates results of the convergence test for the cases where the number of MC replications is 250, 1,000 and 1,500. In this figure the MC1 output is plotted on the x-axis and MC2 output on the y-axis. Although the model output corresponds to yearly time series at 10 minutes resolution, we only plotted those points where the overflow magnitude, and therefore COD and NH_4 load, is different from zero. As an indication, for a MC replication size of 1,500, the NSE values for overflow COD and NH_4 concentrations are 0.816 and 0.998, respectively.

The computing times per MC replication are presented in Table A.5. The computations were performed with two different Linux machines, a laptop with four cores for simulations between 50 and 500 replications, and a server with 80 cores for performing the simulations above 500 replications. Similar execution times were reached for MC1 and MC2 for one-month time series at 10-min time steps (August 2010, 4,464 time steps), while substantial differences were obtained when the 80 cores server was used. We obtained similar timing for 1,500 replications with 50 cores as for 250 replications using three cores in the laptop. The timing reached demonstrates the feasibility to perform a solid MC uncertainty propagation analysis with EmiStatR.

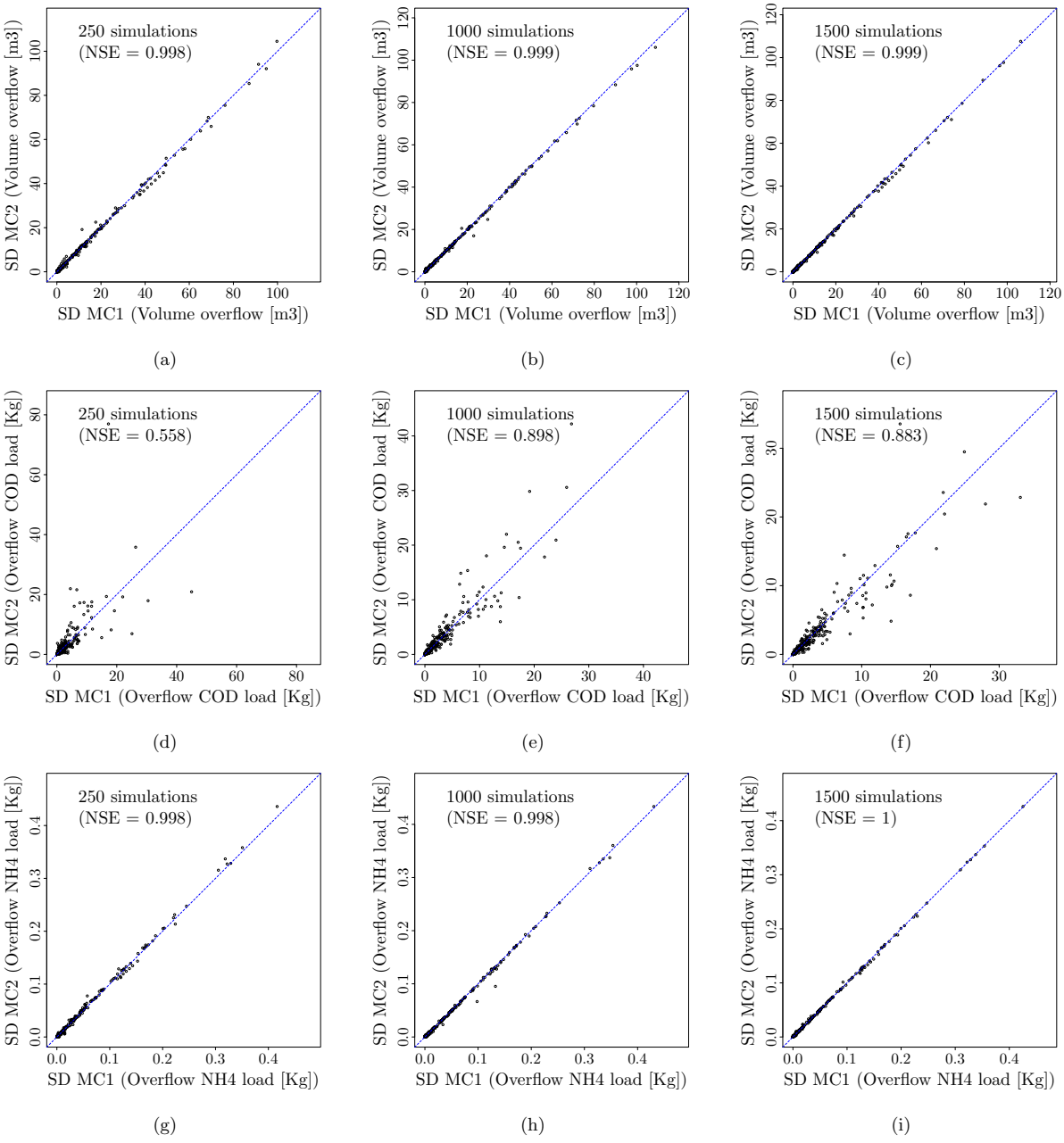


Figure A.4: Results of the MC convergence test for (a, b, c) volume in overflow; (d, e, f) overflow COD load; (g, h, i) overflow NH_4 load. Each open circle refers to a ten minute time instant in 2010 were overflow happened. As an indication, for a MC replication size of 1,500, the NSE values for overflow COD and NH_4 concentrations are 0.816 and 0.998, respectively. Dotted line is the 1:1 line. SD = Standard Deviation.

Appendix B. Software manual: EmiStatR

R-Package EmiStatR

Emissions and Statistics in R for Wastewater and Pollutants in Combined Sewer Systems

This package can be found on:

`https://cran.r-project.org/package=EmiStatR`

Package 'EmiStatR'

May 3, 2019

Type Package

Title Emissions and Statistics in R for Wastewater and Pollutants in Combined Sewer Systems

Version 1.2.2.0

Date 2019-05-03

Author J.A. Torres-Matallana [aut, cre]
K. Klepiszewski [aut, cre]
U. Leopold [ctb]
G. Schutz [ctb]
G.B.M. Heuvelink [ctb]

Maintainer J.A. Torres-Matallana <arturo.torres@list.lu>

Description Provides a fast and parallelised calculator to estimate combined wastewater emissions.
It supports the planning and design of urban drainage systems, without the requirement of
extensive simulation tools. The 'EmiStatR' package implements modular R methods. This enables
to add new functionalities through the R framework.

License GPL (>= 3)

Depends R (>= 2.10), methods

Imports utils, grDevices, graphics, stats, xts, zoo, foreach, parallel, lattice, doParallel

NeedsCompilation no

StagedInstall yes

Repository CRAN

`EmiStatR-package` *Emissions and Statistics in R for Wastewater and Pollutants in Combined Sewer Systems.*

Description

Provides a fast and parallelised calculator to estimate combined wastewater emissions. It supports the planning and design of urban drainage systems, without the requirement of extensive simulation tools. The 'EmiStatR' package implements modular R methods. This enables to add new functionalities through the R framework.

Details

The DESCRIPTION file:

Package:	EmiStatR
Type:	Package
Version:	1.2.2.0
Date:	2019-05-03
License:	GPL (>= 3)
Depends:	R (>= 2.10), methods, shiny
Imports:	utils, grDevices, graphics, stats, xts, zoo, foreach, parallel, lattice, doParallel

Author(s)

J.A. Torres-Matallana [aut, cre]
K. Klepiszewski [aut, cre]
U. Leopold [ctb]
G. Schutz [ctb]
G.B.M. Heuvelink [ctb]

Maintainer: J.A. Torres-Matallana

References

J. A. Torres-Matallana, K. Klepiszewski, U. Leopold, and G.B.M. Heuvelink. EmiStatR: a simplified and scalable urban water quality model for simulation of combined sewer overflows. Water, 10(6)(782):1-24, 2018. `https://www.mdpi.com/2073-4441/10/6/782/htm`

See Also

See also the class the method `EmiStatR`

`Agg` *Temporal aggregation of environmental variables*

Description

Function for temporal aggregation of environmental variables. `Agg` is a wrapper function of `aggregate` from `stats` package.

Usage

```
Agg(data, nameData, delta, func, namePlot)
```

Arguments

`data`	A `data.frame` that contains the time series of the environmental variable to be aggregated, e.g. precipitation. This data.frame should have at least two columns: the first one, Time [y-m-d h:m:s]; the second one, a numeric value equal to the magnitude of the environmental variable. If the environmental variable is different than precipitation, then the column name of the values should be named as `value`.
`nameData`	A `character` string that defines the name of the environmental variable to be aggregated.
`delta`	A `numeric` value that specifies the level of aggregation required in minutes.
`func`	The name of the function of aggregation e.g. `mean`, `sum`.
`namePlot`	A `character` string that defines the title of the plot generated.

Value

A `data.frame` with two columns:

`time`	the date-time time series of the aggregated variable
`value`	time series with the magnitude of the aggregated variable.

Author(s)

J.A. Torres-Matallana

Examples

```
## temporal aggregation
library(EmiStatR)
data(P1)
colnames(P1) <- c("time", "value")
head(P1)
tail(P1)

P1.agg <- Agg(data = P1, nameData = "value", delta = 120 , func = sum,
              namePlot = "Temporal aggregation of precipitation P1")

head(P1.agg)
tail(P1.agg)
```

`CInp2TS` *Function to convert Constant Input to Time Series*

Description

Given daily, weekly and monthly factors, this function converts from a constant numeric input to a time series.

Usage

```
CInp2TS(cinp, prec, cinp.daily.file, cinp.weekly, cinp.seasonal)
```

Arguments

`cinp`	a numeric object that defines the mean constant input to be converted in time series.
`prec`	A data.frame object with observations on the following 2 variables: `time` a POSIXct vector '`P [mm]`' a numeric vector
`cinp.daily.file`	the path and file name of the comma separated value (csv) file that contains the daily factors. The first column of this file should be the time in format "H:M:S" and should span for 24 hours. The second column should contain the factors as numeric class for the specified time. These factors must average to 1.

`cinp.weekly` a "list" that contains the factors per day of the week with 7 elements called "mon" for Monday, "tue" for Tuesday, "wed" for Wednesday, "thu" for Thursday, "fri" for Friday, "sat" for Saturday, and "sun" for Sunday. These factors must average to 1. See example about the definition of this argument.

`cinp.seasonal`
a "list" that contains the factors per month of the year with 12 elements called with the three first lower case letters of the month from "jan" for January to "dec" for December. These factors must average to 1. See example about the definition of this argument.

Details

The generated time series has number of rows equal to the number of rows of the `prec` data.frame. The time index of the time series generated is the same that `prec`.

Value

Object of class `"list"`. This object contains 3 elements:

`time` a POSIXct vector of length n, where n is the number of rows of the `prec` data.frame.

`data` a numeric matrix of size [1:n, 1:4], where columns are: "factor", the daily factor time series based on `cinp.daily.file`; "cinp", the daily time series of the variable defined by `cinp` according to the factors defined by "factor"; "cinp.week", the weekly time series of the variable defined by `cinp` according to the factors defined by `factor` and `cinp.weekly`; and "cinp.season", the monthly time series of the variable defined by `cinp` according to the factors defined by `factor`, `cinp.weekly`, and `cinp.seasonal`.

`xts` An "xts" object containing the 4 same columns that `data`.

Author(s)

J.A. Torres-Matallana; U. Leopold

Examples

```
library(EmiStatR)
library(zoo)

data("Esch_Sure2010")
```

```
data("qs_factor")

cinp            <- 150 # water consumption [m3/h]
prec            <- Esch_Sure2010[1:1000,] # selecting just the first
                                          # 1,000 rows

cinp.daily.file <- qs_factor
cinp.weekly     <- list(mon=1, tue=.83, wed=.83, thu=.83, fri=1,
                        sat=1.25, sun=1.25)
# factors average to 1
cinp.seasonal   <- list(jan=.79, feb=.79, mar=1.15, apr=1.15, may=1.15,
                        jun=1.15, jul=1.15, aug=1.15, sep=1.15,
                        oct=1.15, nov=.79, dec=.79)
# factors average to 1
ts1 <- CInp2TS(cinp, prec, cinp.daily.file, cinp.weekly, cinp.seasonal)
str(ts1)

head(ts1[["xts"]])
summary(ts1[["xts"]])

dev.new()
par(mfrow = c(4,1))
plot(index(ts1[["xts"]][,1]), ts1[["xts"]][,1], type = "l",
     xlab = "", ylab = "Daily factor [-]",
     main="Daily factor time series")
plot(index(ts1[["xts"]][,1]), ts1[["xts"]][,2], type = "l",
     xlab = "", ylab = "Water consumption [l/(PE d)]",
     main="Daily water consumption time series")
plot(index(ts1[["xts"]][,1]), ts1[["xts"]][,3], type = "l",
     xlab = "", ylab = "Water consumption [l/(PE d)]",
     main="Weekly water consumption time series")
plot(index(ts1[["xts"]]), ts1[["xts"]][,4], type = "l",
     xlab = "Time", ylab = "Water consumption [l/(PE d)]",
     main="Yearly water comsumption time series")
```

Delay *Delay function for time series*

Description

This function allows to cretae a n-time steps delayed time series, where n is the number of time steps defined by argument `x`. Henceforth, it is possible to calibrate this argument (parameter) `x`.

Usage

```
Delay(P1, x)
```

Arguments

`P1`	A `data.frame` that contains the time series of the environmental variable to be delayed, e.g. precipitation. This data.frame should have at least two columns: the first one, Time [y-m-d h:m:s]; the second one, a numeric value equal to the magnitude of the environmental variable. If the environmental variable is different than precipitation, then the column name of the values should be named as `value`.
`x`	A `numeric` value that specifies the delayed time in time steps.

Value

A `data.frame` with two columns:

`time`	the date-time time series of the delayed variable
`value`	time series with the magnitude (equal to the original, `P1`, time series) of the delayed variable.

Author(s)

J.A. Torres-Matallana

Examples

```
library(EmiStatR)
data(P1)

P1_delayed <- Delay(P1 = P1, x = 500)

head(P1_delayed)

dev.new()
par(mfrow = c(2, 1))
plot(P1[,1], P1[,2], typ = "l", col = "blue")
plot(P1_delayed[,1], P1_delayed[,2], typ= "l", col = "red")
```

E1	*Dataset: An example of combined sewer overflow (CSO) input for Goesdorf structure (CSO chamber, CSOC).*

Description

This dataset is an example of input data for the CSOC at Goesdorf, Grand-Duchy of Luxembourg.

Usage

```
data("E1")
```

Format

A list of 18 elements. This object contains the first structure (CSOC) to simulate.

`id`: numeric, identification number [-];

`ns`: character, name of the structure [-];

`nm`: character, name of the municipality [-];

`nc`: character, name of the catchment [-];

`numc`: numeric, number of the catchment [-];

`use`: character, use of the soil [-];

`Atotal`: numeric, total area [ha];

`Aimp`: numeric, impervious area [ha];

`tfS`: numeric, time flow structure [time step];

`pe`: numeric, population equivalent [PE];

`V`: numeric, volume [m3];

`lev2vol`: list of 2, lev and vol, defining the curve lev (level [m]) to vol (volume [m3]);

`lev.ini`: numeric, initial level in the CSOC [m];

`Qd`: numeric, maximum throttled outflow [l/s];

`Dd`: numeric, orifice diameter [m];

`Cd`: numeric, orifice coefficient [-];

`Cimp`: numeric, coefficient for impervious area [-];

`Cper`: numeric, coefficient for pervious area [-].

Examples

```
data("E1")

str(E1)
```

`E2` *Dataset: An example of combined sewer overflow (CSO) input for Kaundorf structure (CSO chamber, CSOC).*

Description

This dataset is an example of input data for the CSOC at Kaundorf, Grand-Duchy of Luxembourg.

Usage

```
data("E2")
```

Format

A list of 18 elements. This object contains the first structure (CSOC) to simulate.

`id`: numeric, identification number [-];

`ns`: character, name of the structure [-];

`nm`: character, name of the municipality [-];

`nc`: character, name of the catchment [-];

`numc`: numeric, number of the catchment [-];

`use`: character, use of the soil [-];

`Atotal`: numeric, total area [ha];

`Aimp`: numeric, impervious area [ha];

`tfS`: numeric, time flow structure [time step];

`pe`: numeric, population equivalent [PE];

`V`: numeric, volume [m3];

`lev2vol`: list of 2, lev and vol, defining the curve lev (level [m]) to vol (volume [m3]);

`lev.ini`: numeric, initial level in the CSOC [m];

`Qd`: numeric, maximum throttled outflow [l/s];

`Dd`: numeric, orifice diameter [m];

`Cd`: numeric, orifice coefficient [-];

`Cimp`: numeric, coefficient for impervious area [-];

`Cper`: numeric, coefficient for pervious area [-].

Examples

```
data("E2")

str(E2)
```

`E3` *Dataset: An example of combined sewer overflow (CSO) input for Nocher-Route structure (CSO chamber, CSOC).*

Description

This dataset is an example of input data for the CSOC at Nocher-Route, Grand-Duchy of Luxembourg.

Usage

```
data("E3")
```

Format

A list of 18 elements. This object contains the first structure (CSOC) to simulate.

`id`: numeric, identification number [-];

`ns`: character, name of the structure [-];

`nm`: character, name of the municipality [-];

`nc`: character, name of the catchment [-];

`numc`: numeric, number of the catchment [-];

`use`: character, use of the soil [-];

`Atotal`: numeric, total area [ha];

`Aimp`: numeric, impervious area [ha];

`tfS`: numeric, time flow structure [time step];

`pe`: numeric, population equivalent [PE];

`V`: numeric, volume [m3];

`lev2vol`: list of 2, lev and vol, defining the curve lev (level [m]) to vol (volume [m3]);

`lev.ini`: numeric, initial level in the CSOC [m];

`Qd`: numeric, maximum throttled outflow [l/s];

`Dd`: numeric, orifice diameter [m];

`Cd`: numeric, orifice coefficient [-];

`Cimp`: numeric, coefficient for impervious area [-];

`Cper`: numeric, coefficient for pervious area [-].

Examples

```
data("E3")

str(E3)
```

`EmiStatR-methods` *S4 Methods for Function* `EmiStatR`

Description

S4 methods for function `EmiStatR`. Given the inputs either from the Shiny applications "EmiStatR_input - Shiny" and "EmiStatR_inputCSO - Shiny" or user-defined, the methods invoke the main core of the tool and writes the output files in the specified folder.

Usage

```
EmiStatR(x)
```

Arguments

`x`	An object of class input

Value

Object of class `"list"`. This object contains N lists, where N is the number of structures to simulate. Each list contains a list with three elements: a data.frame named "out1", a data.frame named "out2", a vector named "lista". "out1" contains n observations of 24 variables, where n is the length of the precipitation time series. The 24 variables are the following time series:
1. id, identification number
2. Time [y-m-d h:m:s]

3. P [mm], precipitation
4. i [mm/(min)], intensity (if available)
5. V_r [m3], rain water volume
6. V_dw [m3], dry weather volume
7. cs_mr [-], combined sewage mixing ratio
8. o_tfyn [yes=1/no=0], status variable to know when the Combined Sewer Overflow Tank (CSOT) is filling up
9. V_Tank [m3], volume of CSOT filling up
10. V_Ov [m3], overflow volume
11. B_COD_Ov [kg], Chemical Oxygen Demand (COD) overflow load
12. B_NH4_Ov [kg], ammonium (NH4) overflow load
13. C_COD_Ov [mg/l], COD overflow concentration
14. C_NH4_Ov [mg/l], NH4 overflow concentration
15. d_Ov [min], total duration of overflows
16. f_Ov [ocurrence], frequency of overflows (just an approximation)
17. V_InTank [m3], volume at entrance of the CSOT
18. B_COD_InTank [Kg], COD load at entrance of the CSOT
19. B_NH4_InTank [Kg], NH4 load at entrance of the CSOT
20. C_COD_InTank [mg/l], COD concentration at entrance of the CSOT
21. C_NH4_InTank [mg/l], NH4 concentration at entrance of the CSOT
22. Q_Ov [l/s], overflow flow
23 pe.season [PE], population equivalents in the catchment
24 qs.season [l/PE/d], water consumption in the catchment

The summary of the overflow data, "out2", contains 15 observations of 2 variables. The 15 observations are:
1. Period [day], length of time of the precipitation time series
2. Duration, d_Ov, [min], overflow duration
3. Frecuency, f_Ov, [ocurrence] (aprox.), overflow frecuency
4. Total volume, V_Ov, [m3], total overflow volume
5. Average flow, Q_Ov, [l/s], average overflow flow
6. Total COD load, B_COD_Ov, [kg], total COD load in overflow
7. Average COD concentration, C_COD_ov_av, [mg/l], in overflow
8. 99.9th percentile COD concentration, C_COD_Ov_99.9, [mg/l], in overflow
9. Maximum COD concentration, C_COD_Ov_max, [mg/l], in overflow
10. Total NH4 load, B_NH4_Ov, [kg], total NH4 load in overflow
11. Average NH4 concentration,C_NH4_Ov_av, [mg/l], in overflow
12. 99.9th percentile NH4 concentration, C_NH4_Ov_99.9, [mg/l], in overflow
13. Maximum NH4 concentration, C_NH4_Ov_max, [mg/l], in overflow
14. Structure summary results, (a descriptive text line)
15. Volume Tank, VTank [m3], total volumen in the CSO tank

"Lista" contains the identification name(s) of the N structure(s). If export is allowed then three plain text .csv files are created, one for "out1", the second for "out2", the third one a summary for all the structures based in "out2". Also, one .pdf file is printed which illustrates the precipitation and Combined Sewer Overflow (CSO) volume, COD concentration, and NH4 concentration time series. These files are exported to the directory EmiStatR_output located in the `folderOutput` path.

Methods

`signature(x = "input")` execute EmiStatR function

Author(s)

J.A. Torres-Matallana; K. Klepiszewski; U. Leopold; G.B.M. Heuvelink

Examples

```
## running GUI
library("EmiStatR")

appDir <- system.file("shiny", package = "EmiStatR")

## (uncomment for running)
# setwd(appDir)
# runApp("EmiStatR_input")
# runApp("EmiStatR_inputCSO")

## executing EmiStatR
input.default <- input()

## uncomment following lines to execute
# sim   <- EmiStatR(input.default)
# str(sim)

## a dummy example of plot
# par(mfrow=c(2,2), oma = c(0,0,2,0))
# plot(x=sim[[1]][[1]][[2]], y=sim[[1]][[1]][[3]], typ="l", col="blue",
#      xlab = "time", ylab = colnames(sim[[1]][[1]])[3],
       main = "Precipitation")
# plot(x=sim[[1]][[1]][[2]], y=sim[[1]][[1]][[10]], typ="l", col="blue",
#      xlab = "time", ylab = colnames(sim[[1]][[1]])[10], main = "CSO,
       volume")
# plot(x=sim[[1]][[1]][[2]], y=sim[[1]][[1]][[13]], typ="l", col="blue",
```

```
#      xlab = "time", ylab = colnames(sim[[1]][[1]])[13], main = "CSO,
       COD concentration")
# plot(x=sim[[1]][[1]][[2]], y=sim[[1]][[1]][[14]], typ="l", col="blue",
#      xlab = "time", ylab = colnames(sim[[1]][[1]])[14], main = "CSO,
       NH4 concentration")
# mtext(paste("Structure", sim[[1]][[3]][[1]]), outer=TRUE, cex = 1.5)
```

`Esch_Sure2010` *Dataset: An example time series for the EmiStatR package*

Description

This dataset is a list that contains a data.frame with two columns: time [y-m-d h:m:s] and precipitation depth, P [mm]. The dataset correspond to measurements for 2010 with time steps of 10 minutes at rain gauge station, Esch-sur-Sure, located close to the sub-catchment of the combined sewer overflow chamber at Goesdorf, Grand-Duchy of Luxembourg.

Usage

```
data("Esch_Sure2010")
```

Format

A data frame with 52560 observations on the following 2 variables.

`time` a POSIXct vector

'`P [mm]`' a numeric vector

Source

`http://agrimeteo.lu/`

Examples

```
data(Esch_Sure2010)

plot(Esch_Sure2010[,2], col="blue", typ="l", xlab = "time",
     ylab = "Precipitation [mm]")
```

inf — *Dataset: An example of general input for infiltration characteristics in the study region.*

Description

This dataset is an example of general input for infiltration characteristics in the sewer system for the study region (Haute-Sure catchment, Grand-Duchy of Luxembourg).

Usage

```
data("inf")
```

Format

A list of 3 elements.

`qf`: numeric, infiltration water inflow [l/(s*ha)];

`CODf`: numeric, chemical oxygen demand (COD) infiltration water pollution per capita (PE) load per day [g/(PE*day)];

`NH4f`: numeric, ammonium (NH4) infiltration water pollution per capita (PE) load per day[g/(PE*day];

Examples

```
data("inf")

str(inf)
```

input-class — *Class* "input"

Description

The class provides a container for inputs required to invoke `EmiStatR` method.

Objects from the Class

Objects can be created by calls of the form `input()` or `new("input")`.

Slots

`spatial`: Object of class `"numeric"`, 0 (default) for non-spatial input, 1 for spatial input (not implemented).

`zero`: Object of class `"numeric"`, aproximation to zero value. Default 1E-5.

`folder`: Object of class `"character"`, path for the output. Default `getwd()`

`cores`: Object of class `"numeric"`, number of CPU cores to be used in parallel computing. If cores = 0 no parallel computation is done. Default 1.

`ww`: Object of class `"list"`. This list contains three numeric elements for the wastewater characteristics. First element `qs`, individual water consumption of households [l/(PE d)]. Second element `CODs`, sewage pollution - COD concentration [g/(PE d)]. Third element `NH4s`, sewage pollution - NH4 concentration [g/(PE d)].

`inf`: Object of class `"list"`. This list contains three numeric elements for infiltration water characteristics. First element `qf`, infiltration water inflow [l/(s ha)]. Second element `CODf`, infiltration water pollution - COD concentration [g/(PE d)]. Third element `NH4f`, infiltration water pollution - NH4 concentration [g/(PE d)].

`rw`: Object of class `"list"`. This list contains three elements for rainwater characteristics. First element `CODr` (numeric), rainwater pollution - COD concentration [mg/l]. Second element `NH4r` (numeric), rainwater pollution - NH4 concentration [mg/l]. Third element `stat` (character), name of the rain measurement station.

`P1`: Object of class `"data.frame"` with two columns named 1. "time" for specifying the date and time in format YYYY-m-d HH:MM:SS, 2. "P [mm]" for specifying the depth values of the rainfall time series in millimeters. Optionally, instead of rainfall depth values can be provided values of direct runoff in cubic meters entering in the system. If runoff values are provided then the second column containing these values should be named as "Runoff_Volume" or "runoff_volume", otherwise this column is treated as rainfall depth.

`st`: Object of class `"list"`. This object contains n lists, where n is the number of structures to simulate. Every list should contain 18 elements: `id`, numeric, identification number [-]; `ns`, character, name of the structure [-]; `nm`, character, name of the municipality [-]; `nc`, character, name of the catchment [-]; `numc`, numeric, number of the catchment [-]; `use`, character, use of the soil [-]; `Atotal`, numeric, total area [ha]; `Aimp`, numeric, impervious area [ha]; `tfS`, numeric, time flow structure [time step]; `pe`, numeric, population equivalent [PE]; `V`, numeric, volume [m3]; `lev2vol`, list of 2, lev and vol, defining the curve lev (level [m]) to vol (volume [m3]); `Qd`, numeric, maximum throttled outflow [l/s]; `Dd`, numeric, orifice diameter [m]; `Cd`, numeric, orifice coefficient [-]; `Cimp`, numeric, coefficient for impervious area [-]; and `Cper`, numeric, coefficient for pervious area [-].

`pe.ts.file`: Object of class `"character"` with the path and file name of the comma separated value (csv) file that contains the montly (seasonal) factors for population equivalent (`pe`). The first column of this file should be "time" in format

"Y-m-d H:M:S" and should span for the entire length of the desired time series. The second column should contain the population equivalent as numeric class for the specified time, i.e. the desired `pe` time series with daily, weekly and monthly factors already applied. Default empty string (""). If not empty string is defined then `pe.daily.file`, `pe.weekly`, `pe.seasonal` are omitted.

`pe.daily.file`: Object of class `"data.frame"` that contains the daily factors for population equivalent. The first column should be the time in format "H:M:S" and should span for 24 hours. The second column should contain the factors as numeric class for the specified time. These factors must average to 1.

`pe.weekly`: Object of class `"list"` that contains the factors for population equivalent per day of the week with 7 elements called "mon" for Monday, "tue" for Tuesday, "wed" for Wednesday, "thu" for Thursday, "fri" for Friday, "sat" for Saturday, and "sun" for Sunday. These factors must average to 1.

`pe.seasonal`: Object of class `"list"` that contains the factors for population equivalent per month of the year with 12 elements called with the three first lower case letters of the month from "jan" for January to "dec" for December. These factors must average to 1.

`qs.ts.file`: Object of class `"character"` with the path and file name of the comma separated value (csv) file that contains the monthly (seasonal) factors for water consumption (`qs`). The first column of this file should be "time" in format "Y-m-d H:M:S" and should span for the entire length of the desired time series. The second column should contain the population equivalent as numeric class for the specified time, i.e. the desired `qs` time series with daily, weekly and monthly factors already applied. Default empty string (""). If not empty string is defined then `pe.daily.file`, `pe.weekly`, `pe.seasonal` are omitted.

`qs.daily.file`: Object of class `"character"` with the path and file name of the comma separated value (csv) file that contains the daily factors for water consumption. The first column of this file should be the time in format "H:M:S" and should span for 24 hours. The second column should contain the factors as numeric class for the specified time. These factors must average to 1.

`qs.weekly`: Object of class `"list"` that contains the factors for water consumption per day of the week with 7 elements called "mon" for Monday, "tue" for Tuesday, "wed" for Wednesday, "thu" for Thursday, "fri" for Friday, "sat" for Saturday, and "sun" for Sunday. These factors must average to 1.

`qs.seasonal`: Object of class `"list"` that contains the factors for water consumption per month of the year with 12 elements called with the three first lower case letters of the month from "jan" for January to "dec" for December. These factors must average to 1.

`export`: Object of class `"numeric"`. If 1 (default) then the results are saved in `folderOutput`. Set to 0 for not writing in output files.

Methods

EmiStatR `signature(x = "input")`: execute EmiStatR function

Author(s)

J.A. Torres-Matallana

Examples

```
## loading EmiStatR
library("EmiStatR")
showClass("input")

## running EmiStatR with user defined input
data("Esch_Sure2010")
P1 <- Esch_Sure2010[1:1000,]  # selecting just the first 1,000 rows
station <- "Esch-sur-Sure"

# defining estructures E1
E1 <- list(id = 1, ns = "FBH Goesdorf", nm = "Goesdorf", nc = "Obersauer",
           numc = 1, use = "R/I", Atotal = 36, Aimp = 25.2, Cimp = 0.80,
           Cper = 0.30, tfS = 1, pe = 650, Qd = 5,
           Dd = 0.15, Cd = 0.18, V = 190, lev.ini = 0.10,
           lev2vol = list(lev = c(.06, 1.10, 1.30, 3.30),
                          vol = c(0.5, 31, 45, 190))
           )

# defining Input objet
input.user <- input(spatial = 0, zero = 1e-5, folder = getwd(),
                    cores = 1,
                    ww = list(qs = 150, CODs = 104, NH4s = 4.7),
                    inf = list(qf= 0.04, CODf = 0, NH4f =0),
                    rw = list(CODr = 0, NH4r = 0, stat = station),
                    P1 = P1, st = list(E1=E1), export = 1)

str(input.user)

# invoking EmiStatR
sim <- EmiStatR(input.user)
```

```
## a visualisation example
dev.new()
par(mfrow=c(2,2), oma = c(0,0,2,0))
plot(x=sim[[1]][[1]][[2]], y=sim[[1]][[1]][[3]], typ="l", col="blue",
     xlab = "time", ylab = colnames(sim[[1]][[1]])[3],
     main = "Precipitation")
plot(x=sim[[1]][[1]][[2]], y=sim[[1]][[1]][[10]], typ="l", col="blue",
     xlab = "time", ylab = colnames(sim[[1]][[1]])[10],
     main = "CSO, volume")
plot(x=sim[[1]][[1]][[2]], y=sim[[1]][[1]][[13]], typ="l", col="blue",
     xlab = "time", ylab = colnames(sim[[1]][[1]])[13],
     main = "CSO, COD concentration")
plot(x=sim[[1]][[1]][[2]], y=sim[[1]][[1]][[14]], typ="l", col="blue",
     xlab = "time", ylab = colnames(sim[[1]][[1]])[14],
     main = "CSO, NH4 concentration")
mtext(paste("Structure", sim[[1]][[3]][[1]]), outer=TRUE, cex = 1.5)
```

`IsReg` *Wrapping Function for Function "is.regular" from "zoo" package for "data.frame" objects*

Description

"IsReg" is a wrapping Function for Function "is.regular" from "zoo" package. Given a "data.frame" object it is converted into a "xts" object, while the regularity of the object is checked. The first column of the "data.frame" should contain a character string vector to be converted via as.POSIXct accordingly with the date format (format) and time zone (tz).

Usage

```
IsReg(data, format, tz)
```

Arguments

`data`	an object of class `data.frame` containing in its first column a character string vector to be converted via as.POSIXct into a date vector accordingly with the date format (format) and time zone (tz) defined
`format`	character string giving a date-time format as used by `strptime`.
`tz`	a time zone specification to be used for the conversion, if one is required. System-specific, but "" is the current time zone, and "GMT" is UTC (Universal Time, Coordinated). Invalid values are most commonly treated as UTC, on some platforms with a warning.

Details

"IsReg" calls the as.POSIXct function from `base` package to convert an object to one of the two classes used to represent date/times (calendar dates plus time to the nearest second). More details can be found in the "is.regular" function of the "zoo" package.

Value

Object of class `"list"`. This object contains 2 elements, the first one contains a character string "_TSregular" if the xts object created is strict regular, or "_TSirregular" if it is strict irregular. More details can be found in the "is.regular" function of the "zoo" package.

Author(s)

J.A. Torres-Matallana

Examples

```
library(EmiStatR)
data("P1")

class(P1)
head(P1)

ts <- IsReg(data = P1, format = "%Y-%m-%d %H:%M:%S", tz = "UTC")
str(ts)

ts[[1]]

head(ts[[2]]); tail(ts[[2]])

plot(ts[[2]], ylab = "Precipitation [mm]")
```

`Level2Volume` *Function for linear interpolation given the relationship of two variables*

Description

Given a relationship between two variables (e.g. level and volume), this function interpolates the corresponding second variable (e.g. volume) for a known value of the first variable (e.g. level). This function is suitable to represented e.g. the dynamics of water storage in a combined sewer overflow chamber.

Usage

```
Level2Volume(lev, lev2vol)
```

Arguments

`lev`	A `numeric` object that represents the know variable e.g. the level in a storage chamber.
`lev2vol`	A `list` of two elements. The first element is a `vector` named "lev" that contains the interpolation steps of the first variable (e.g. level). The second element is a `vector` that contains the interpolation steps of the second variable (e.g. volume).

Details

The function uses the `approx` function from the `stats` package with "yleft" argument equal to the minimum value of the second variable and "yright" argument equal to the maximum value of the second variable.

Value

A `numeric` object with one element representing the interpolated value of the second variable (e.g. volume).

Author(s)

J.A. Torres-Matallana; K. Klepiszewski; G. Schutz.

Examples

```
library(EmiStatR)

# definition of relationship level to volume
lev2vol <- list(lev = c(.06, 1.10, 1.30, 3.30), vol = c(0, 31, 45, 200))

interpolated_volume <- Level2Volume(lev=c(0, .25, 1.25, 2.25, 4.25),
                                    lev2vol = lev2vol)
interpolated_volume
```

`P1` *Dataset: An example of input time series for the EmiStatR package*

Description

This dataset is a list that contains a data.frame with two columns: time [y-m-d h:m:s] and precipitation depth, P [mm]. The dataset correspond to measurements for January 2016 with time steps of 10 minutes at rain gauge station, Dahl, located close to the sub-catchment of the combined sewer overflow chamber at Goesdorf, Grand-Duchy of Luxembourg.

Usage

```
data("P1")
```

Format

A data frame with 4464 observations on the following 2 variables.

`time` a POSIXct vector

'`P [mm]`' a numeric vector

Source

`http://agrimeteo.lu/`

Examples

```
data("P1")

plot(P1[,1], P1[,2], col="blue", typ="l", xlab = "time",
     ylab = "Precipitation [mm]")
```

`P1_20111216` *Dataset: An example of input time series for the EmiStatR package*

Description

This dataset is a data.frame with four columns: time [y-m-d h:m:s] and precipitation depth, P [mm] for Dahl, Eshc-sur-Sure and Eschdorf rain gauge stations. The dataset correspond to measurements for December 16th 2011 for a 10-hour event recorded with time steps of 10 minutes. The three rain gauge stations are located close to the sub-catchment of the combined sewer overflow chamber at Goesdorf, Grand-Duchy of Luxembourg.

Usage

```
data("P1_20111216")
```

Format

A data frame with 61 observations on the following 4 variables.

`time` a POSIXct vector

`Dahl` a numeric vector for rainfall depth in millimeters at Dahl rain gauge station

`Esch-Sure` a numeric vector for rainfall depth in millimeters at Esch-sur-Sure rain gauge station

`Eschdorf` a numeric vector for rainfall depth in millimeters at Eschdorf rain gauge station

Source

`http://agrimeteo.lu/`

Examples

```
data("P1_20111216")

plot(P1_20111216[,1], P1_20111216[,2], col="blue", typ="l",
     xlab = "time", ylab = "Precipitation at Dahl [mm]")

plot(P1_20111216[,1], P1_20111216[,3], col="blue", typ="l",
     xlab = "time", ylab = "Precipitation at Esch-sur-Sure [mm]")

plot(P1_20111216[,1], P1_20111216[,4], col="blue", typ="l",
    xlab = "time", ylab = "Precipitation at Eschdorf [mm]")
```

`pe_factor` *Dataset: An example of general input for the daily population equivalent factors in the study region.*

Description

This dataset is an example of general input for the daily population equivalent (pe) factors in the study region (Haute-Sure catchment, Grand-Duchy of Luxembourg).

Usage

```
data("pe_factor")
```

Format

A data.frame with 12 observations of 2 variables:

`time`: character, time at which the factors are reported [HH:MM:SS];

`pe_factor`: numeric, the corresponding factors for population equivalents [-];

Examples

```
data("pe_factor")

pe_factor
```

`qs_factor` *Dataset: An example of general input for the daily water consumption factors in the study region.*

Description

This dataset is an example of general input for the daily water consumption (qs) factors in the study region (Haute-Sure catchment, Grand-Duchy of Luxembourg).

Usage

```
data("qs_factor")
```

Format

A data.frame with 12 observations of 2 variables:

`time`: character, time at which the factors are reported [HH:MM:SS];

`qs_factor`: numeric, the corresponding factors for water consumption [-];

Examples

```
data("qs_factor")

qs_factor
```

`qs_factor_ATV` *Dataset: An example of general input for the daily water consumption factors (ATV German guideline) in the study region.*

Description

This dataset is an example of general input for the daily water consumption (qs) factors according to the german guideline ATV-A134 applied in the study region (Haute-Sure catchment, Grand-Duchy of Luxembourg).

Usage

```
data("qs_factor")
```

Format

A data.frame with 12 observations of 2 variables:

`time`: character, time at which the factors are reported [HH:MM:SS];

`ATV.A134.Qf.m3_h`: numeric, the corresponding factors for water consumption according to the German guideline ATV-A134 [-];

Examples

```
data("qs_factor_ATV")

qs_factor_ATV
```

`rw` *Dataset: An example of general input for rainwater characteristics in the study region.*

Description

This dataset is an example of general input for rainwater characteristics in the sewer system for the study region (Haute-Sure catchment, Grand-Duchy of Luxembourg).

Usage

```
data("rw")
```

Format

A list of 6 elements:

`depth`: numeric, total rainfall amount of time series P1 [mm];

`pDur`: numeric, total rainfall duration of time series P1 [min];

`CODr`: numeric, rainwater chemical oxygen demand (COD) concentration [mg/l];

`NH4r`: numeric, rainwater ammonium (NH4) concentration [mg/l];

`stat`: character, raingauge station name for time series P1 [-];

`delta.t`: numeric, delta time for time series P1 [min];

Examples

```
data("rw")

str(rw)
```

`Volume2Level` *Function for linear interpolation given the relationship of two variables*

Description

Given a relationship between two variables (e.g. volume and level), this function interpolates the corresponding second variable (e.g. level) for a known value of the first variable (e.g. volume). This function is suitable to represented e.g. the dynamics of water storage in a combined sewer overflow chamber.

Usage

```
Volume2Level(vol, lev2vol)
```

Arguments

`vol`	A `numeric` object that represents the know variable e.g. the volume in a storage chamber.
`lev2vol`	A `list` of two elements. The first element is a `vector` named "lev" that contains the interpolation steps of the first variable (e.g. level). The second element is a `vector` that contains the interpolation steps of the second variable (e.g. volume).

Author(s)

J.A. Torres-Matallana, G. Schutz

Examples

```
library(EmiStatR)

# definition of relationship level to volume
lev2vol <- list(lev = c(.06, 1.10, 1.30, 3.30), vol = c(0, 31, 45, 200))

interpolated_level <- Volume2Level(vol=c(0, 25, 41, 190, 220),
                                   lev2vol = lev2vol)
interpolated_level
```

`ww` *Dataset: An example of general input for wastewater characteristics in the study region.*

Description

This dataset is an example of general input for wastewater characteristics in the sewer system for the study region (Haute-Sure catchment, Grand-Duchy of Luxembourg).

Usage

```
data("ww")
```

Format

A list of 3 elements.

`qs`: numeric, infiltration inflow [l/(s*ha)];

`CODs`: numeric, chemical oxygen demand (COD) sewage pollution per capita (PE) load per day [g/(PE*day)];

`NH4s`: numeric, ammonium (NH4) sewage pollution per capita (PE) load per day [g/(PE*day];

Examples

```
data("ww")

str(ww)
```

Appendix C. Software manual: stUPscales

R-Package stUPscales

Spatio-Temporal Uncertainty Propagation

Across Multiple Scales

This package can be found on:

`https://cran.r-project.org/package=stUPscales`

Package 'stUPscales'

May 6, 2019

Type Package
Title Spatio-Temporal Uncertainty Propagation Across Multiple Scales
Version 1.0.3.4
Date 2019-05-04
Author J.A. Torres-Matallana [aut, cre]
U. Leopold [ctb]
G.B.M. Heuvelink [ctb]
Maintainer J.A. Torres-Matallana <arturo.torres@list.lu>

Description Integrated environmental modelling requires coupling sub-models at different spatial and temporal scales, thus accounting for change of support procedures (aggregation and disaggregation). We contribute to state-of-the-art open source tools that support uncertainty propagation analysis in temporal and spatio-temporal domains. We implement the tool for uncertainty propagation in environmental modelling, with examples in the urban water domain. The functionalities of the class setup and the methods and functions MC.setup, MC.sim, MC.analysis, MC.analysis_generic and Agg.t are contained, which are used for setting up, running and analysing Monte Carlo uncertainty propagation simulations, and for spatio-temporal aggregation. We also implement functionalities to model and predict variables that vary in space and time. stUPscales takes uncertainty characterisation and propagation a step further by including temporal and spatio-temporal auto- and cross-correlation, resulting in more realistic (spatio-)temporal series of environmental variables. Due to its modularity, the package allows the implementation of additional methods and functions for spatio-temporal disaggregation of model inputs and outputs, when linking models across multiple space-time scales.

License GPL (>= 3)

Depends R (>= 2.10), methods, stats, graphics, grDevices, utils, mAr, lmom,
Imports parallel, doParallel, foreach, lattice, msm, ggplot2, moments, hydroGOF, zoo, data.table, xts, EmiStatR

Suggests sp, spacetime
NeedsCompilation no
Repository CRAN
Date/Publication 2019-05-05 22:50:04 UTC

`stUPscales-package` *Spatio-Temporal Uncertainty Propagation Across Multiple Scales*

Description

Integrated environmental modelling requires coupling sub-models at different spatial and temporal scales, thus accounting for change of support procedures (aggregation and disaggregation). We contribute to state-of-the-art open source tools that support uncertainty propagation analysis in temporal and spatio-temporal domains. We implement the tool for uncertainty propagation in environmental modelling, with examples in the urban water domain. The functionalities of the class setup and the methods and functions MC.setup, MC.sim, MC.analysis, MC.analysis_generic and Agg.t are contained, which are used for setting up, running and analysing Monte Carlo uncertainty propagation simulations, and for spatio-temporal aggregation. We also implement functionalities to model and predict variables that vary in space and time. stUPscales takes uncertainty characterisation and propagation a step further by including temporal and spatio-temporal auto- and cross-correlation, resulting in more realistic (spatio-)temporal series of environmental variables. Due to its modularity, the package allows the implementation of additional methods and functions for spatio-temporal disaggregation of model inputs and outputs, when linking models across multiple space-time scales.

Details

The DESCRIPTION file:

Package:	stUPscales
Type:	Package
Version:	1.0.3.4
Date:	2019-05-04
License:	GPL (>= 3)
Depends:	R (>= 2.10), methods, stats, graphics, grDevices, utils, mAr, lmom
Imports:	parallel, doParallel, foreach, lattice, msm, ggplot2, moments, hydroGOF, zoo, data.table, xts, EmiStatR
Suggests:	sp, spacetime

Author(s)

J.A. Torres-Matallana [aut, cre]; U. Leopold [ctb]; G.B.M. Heuvelink [ctb].

Maintainer: J.A. Torres-Matallana.

`Agg.t` *Temporal aggregation of environmental variables*

Description

Function for temporal aggregation of environmental variables. `Agg` is a wrapper function of `aggregate` from `stats` package.

Usage

```
Agg.t(data, nameData, delta, func, namePlot)
```

Arguments

`data`	A `data.frame` that contains the time series of the environmental variable to be aggregated, e.g. precipitation. This data.frame should have at two columns: the first one, Time [y-m-d h:m:s]; the second one, a numeric value equal to the magnitude of the environmental variable. If the environmental variable is different than precipitation, then the column name of the values can be named as the name of the variable itself.
`nameData`	A `character` string that defines the name of the environmental variable to be aggregated.
`delta`	A `numeric` value that specifies the level of aggregation required in minutes.
`func`	The name of the function of aggregation e.g. `mean`, `sum`.
`namePlot`	A `character` string that defines the title of the plot generated.

Value

A `data.frame` with two columns:

`time`	the date-time time series of the aggregated variable
`value`	time series with the magnitude of the aggregated variable.

Author(s)

J.A. Torres-Matallana

Examples

```
## temporal aggregation
library(EmiStatR)
data(P1)
colnames(P1) <- c("time", "P1")
head(P1)

library(stUPscales)
P1.agg <- Agg.t(data = P1, nameData = "P1", delta = 120 , func = sum,
              namePlot = "Temporal aggregation of precipitation P1")

head(P1.agg)
tail(P1.agg)
```

Germany_precipitation_201112

Sample precipitation time series in Germany

Description

A 1-minute sample event for precipitation time series measured in 37 rain gauge stations distributed over the territory of Germany close to the frontier to the Grand-Duchy of Luxembourg.

Usage

```
data("Germany_precipitation_201112")
```

Format

The format is:
An 'xts' object on 2011-12-01/2011-12-31 23:59:00 containing:
Data: num [1:44640, 1:37] NA NA NA NA NA NA NA NA NA NA ...
- attr(*, "dimnames")=List of 2
..$: NULL
..$: chr [1:37] "Station.1327" "Station.1964" "Station.200" "Station.2170" ...
Indexed by objects of class: [POSIXct,POSIXt] TZ:
xts Attributes:
NULL

Source

https://www.dwd.de/

Examples

```
library(stUPscales)

data(Germany_precipitation_201112)

summary(Germany_precipitation_201112)

library(zoo)
par(mfrow = c(5, 1))
for(i in 1:ncol(Germany_precipitation_201112)){
  plot(index(Germany_precipitation_201112), Germany_precipitation_201112[,i],
       typ="l", col="blue", xlab="Time", ylab='Precip. [mm]',
       main=colnames(Germany_precipitation_201112[,i]))
}
```

`Germany_stations` *A SpatialPointsDataFrame with the location of 37 rain gauges in Germany*

Description

A SpatialPointsDataFrame with the location of 37 rain gauges distributed over the territory of Germany close to frontier with the Grand-Duchy of Luxembourg. These 37 stations are the same related to the "Germany_precipitation_201112" dataset.

Usage

```
data("Germany_stations")
```

Format

The format is:
Formal class 'SpatialPointsDataFrame' [package "sp"] with 5 slots
..@ data :'data.frame': 37 obs. of 9 variables:
.. ..$ Stations_id : int [1:37] 200 450 460 523 603 723 902 942 953 1327 ...
.. ..$ von_datum : int [1:37] 20020924 20050920 19930930 20020807 20071024 20020717 20060618 20020925 19970730 20040707 ...
.. ..$ bis_datum : int [1:37] 20180820 20121204 20180820 20180604 20180820 20180820 20180820 20180820 20180820 20180820 ...

.. ..$ Stationshoehe: int [1:37] 517 120 363 359 159 290 573 308 481 147 ...
.. ..$ geoBreite : num [1:37] 50.1 49.9 49.3 50 50.7 ...
.. ..$ geoLaenge : num [1:37] 6.32 7.07 6.69 6.53 7.19 ...
.. ..$ Stationsname : Factor w/ 1109 levels "Aachen","Aachen-Orsbach",..: 535 90 92 104 118 456 172 179 180 1046 ...
.. ..$ Bundesland : Factor w/ 16 levels "Baden-Wuerttemberg",..: 11 11 12 11 10 11 10 11 11 10 ...
.. ..$ d : logi [1:37] NA NA NA NA NA NA ...
..@ coords.nrs : num(0)
..@ coords : num [1:37, 1:2] 90590 144511 117752 106135 152358 ...
.. ..- attr(*, "dimnames")=List of 2
.. .. ..$: NULL
.. .. ..$: chr [1:2] "coords.x1" "coords.x2"
..@ bbox : num [1:2, 1:2] 82780 31291 156554 200669
.. ..- attr(*, "dimnames")=List of 2
.. .. ..$: chr [1:2] "coords.x1" "coords.x2"
.. .. ..$: chr [1:2] "min" "max"
..@ proj4string:Formal class 'CRS' [package "sp"] with 1 slot
.. .. ..@ projargs: chr "+init=epsg:2169 +proj=tmerc +lat_0=49.83333333333334 +lon_0=6.166666666666667 +k=1 +x_0=80000 +y_0=100000 +ellp"— __truncated__

Source

`https://www.dwd.de/`

Examples

```
library(stUPscales)
library(sp)

data(Germany_stations)

str(Germany_stations)

data(Lux_boundary)
plot(Germany_stations)
plot(boundary.Lux, add=TRUE) # Luxembourg boundary
```

`GoF` *Wrapper function for the* `gof` *function from* `hydroGOF` *package*

Description

A wrapper function for the `gof` function from `hydroGOF` package

Usage

```
GoF(eval, col_sim, col_obs, name)
```

Arguments

`eval`	A `matrix` or `data.frame` with n observations of at least two variables: simulations and observations.
`col_sim`	A numeric value defining the column in `eval` data.frame that contains the simulated vector time series.
`col_obs`	A numeric value defining the column in `eval` data.frame that contains the observed vector time series.
`name`	A character string that defines the name of the files (.csv and .RData) created with the results. If missing then no files are created.

Value

A vector with 20 elements for each one of the following measures of godness-of-fit: 1) ME, mean error; 2) MAE, mean absolute error; 3) MSE, mean squared error; 4) RMSE, root mean square error; 5) NRMSE %, normalized root square error; 6) PBIAS %, percent bias; 7) RSR, Ratio of RMSE to the standard deviation of the observations; 8) rSD, Ratio of Standard Deviations; 9) NSE, Nash-Sutcliffe Efficiency; 10) mNSE, modified Nash-Sutcliffe efficiency; 11) rNSE, relative Nash-Sutcliffe efficiency; 12) d, Index of Agreement; 13) md, Modified index of agreement; 14) rd, Relative Index of Agreement; 15) cp,Coefficient of persistence; 16) r, Pearson product-moment correlation coefficient; 17) R2, Coefficient of Determination; 18) bR2, Coefficient of determination (r2) multiplied by the slope of the regression line between sim and obs; 19) KGE,Kling-Gupta Efficiency; 20) VE, Volumetric Efficiency.

Author(s)

J.A. Torres-Matallana

References

Mauricio Zambrano-Bigiarini, 2014. hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series. R package version 0.3-8. https://CRAN.R-project.org/package=hydroGOF.

Examples

```
library(stUPscales)

data_new <- rnorm(230, .25, .1)
data_new <- cbind(data_new, data_new*1.2)
colnames(data_new) <- c("sim", "obs")
head(data_new)

gof.new <- GoF(data_new, 1, 2)
gof.new

# writing files
gof.new <- GoF(data_new, 1, 2, "GoF_results")
```

HS_RW20111216_stfdf *1-hour DWD precipitation radar imagery calibrated in STFDF format*

Description

Calibrated hourly precipitation radar imagery from the German Weather Service (DWD from the initials in German) at one-kilometer spatial resolution over the Haute-Sure catchment of the Grand-Duchy of Luxembourg. The data was recorded at Neuheilenbach radar station located in territory of Germany, which covers the entire territory of the Grand-Duchy of Luxembourg and surroundings. This sample STFDF (spatio-temporal full data.frame) corresponds to 1-day sample event for precipitation recorded on 16th December 2011.

Usage

```
data("HS_RW20111216_stfdf")
```

Format

The format is: Formal class 'STFDF' [package "spacetime"] with 4 slots
..@ data :'data.frame': 52900 obs. of 1 variable:
.. ..$ raa01.rw_10000.1112160050.dwd...bin: num [1:52900] 1 1 1 1 1 1 1 1 1 1 ...
..@ sp :Formal class 'SpatialPixels' [package "sp"] with 5 slots
.. .. ..@ grid :Formal class 'GridTopology' [package "sp"] with 3 slots
.. ..@ cellcentre.offset: Named num [1:2] -327 -4347
.. ..- attr(*, "names")= chr [1:2] "s1" "s2"
.. ..@ cellsize : num [1:2] 1 1

.. ..@ cells.dim : int [1:2] 46 46
.. .. ..@ grid.index : int [1:2116] 1 2 3 4 5 6 7 8 9 10 ...
.. .. ..@ coords : num [1:2116, 1:2] -327 -326 -325 -324 -323 ...
.. ..- attr(*, "dimnames")=List of 2
.. ..$: NULL
.. ..$: chr [1:2] "s1" "s2"
.. .. ..@ bbox : num [1:2, 1:2] -327 -4348 -281 -4302
.. ..- attr(*, "dimnames")=List of 2
.. ..$: chr [1:2] "s1" "s2"
.. ..$: chr [1:2] "min" "max"
.. .. ..@ proj4string:Formal class 'CRS' [package "sp"] with 1 slot
.. ..@ projargs: chr "+proj=stere +lat_0=90 +lat_ts=90 +lon_0=10 +k=0.93301270189 +x_0=0 +y_0=0 +a=6370040 +b=6370040 +units=km +no_defs"
..@ time :An 'xts' object on 2011-12-16 00:50:00/2011-12-17 00:50:00 containing:
Data: int [1:25, 1] 1 2 3 4 5 6 7 8 9 10 ...
- attr(*, "dimnames")=List of 2
..$: NULL
..$: chr "timeIndex"
Indexed by objects of class: [POSIXct,POSIXt] TZ:
xts Attributes:
NULL
..@ endTime: POSIXct[1:25], format: "2011-12-16 00:50:00" "2011-12-16 01:50:00" "2011-12-16 02:50:00" "2011-12-16 03:50:00" ...

Source

`https://www.dwd.de/`

Examples

```
library(stUPscales)
data(HS_RW20111216_stfdf)

library(spacetime)
stplot(HS_RW20111216_stfdf,
       scales=list(draw=TRUE),
       key.space="right", colorkey=TRUE,
       main="1-hour DWD sample precipitation radar imagery calibrated
            in STFDF format",
       cex=.74, par.strip.text=list(cex=.74))
```

`HS_RY20111216_stfdf`

5-minute DWD precipitation radar imagery non-calibrated in STFDF format

Description

Non-calibrated 5-minute precipitation radar imagery from the German Weather Service (DWD from the initials in German) at one-kilometer spatial resolution over the Haute-Sure catchment of the Grand-Duchy of Luxembourg. The data was recorded at Neuheilenbach radar station located in territory of Germany, which covers the entire territory of the Grand-Duchy of Luxembourg and surroundings. This sample STFDF (spatio-temporal full data.frame) corresponds to 1-day sample event for precipitation recorded on 16th December 2011.

Please note that these are un-calibrated radar data.

Usage

```
data("HS_RY20111216_stfdf")
```

Format

The format is:
Formal class 'STFDF' [package "spacetime"] with 4 slots
..@ data :'data.frame': 609408 obs. of 1 variable:
.. ..$ raa01.ry_10000.1112160000.dwd...bin: num [1:609408] 1 2 1 1 1 1 1 1 1 1 ...
..@ sp :Formal class 'SpatialPixels' [package "sp"] with 5 slots
.. .. ..@ grid :Formal class 'GridTopology' [package "sp"] with 3 slots
.. ..@ cellcentre.offset: Named num [1:2] -327 -4347
.. ..- attr(*, "names")= chr [1:2] "s1" "s2"
.. ..@ cellsize : num [1:2] 1 1
.. ..@ cells.dim : int [1:2] 46 46
.. .. ..@ grid.index : int [1:2116] 1 2 3 4 5 6 7 8 9 10 ...
.. .. ..@ coords : num [1:2116, 1:2] -327 -326 -325 -324 -323 ...
.. ..- attr(*, "dimnames")=List of 2
.. ..$: NULL
.. ..$: chr [1:2] "s1" "s2"
.. .. ..@ bbox : num [1:2, 1:2] -327 -4348 -281 -4302
.. ..- attr(*, "dimnames")=List of 2
.. ..$: chr [1:2] "s1" "s2"
.. ..$: chr [1:2] "min" "max"
.. .. ..@ proj4string:Formal class 'CRS' [package "sp"] with 1 slot

.. ..@ projargs: chr "+proj=stere +lat_0=90 +lat_ts=90 +lon_0=10 +k=0.93301270189 +x_0=0 +y_0=0 +a=6370040 +b=6370040 +units=km +no_defs"
..@ time :An 'xts' object on 2011-12-16/2011-12-16 23:55:00 containing:
Data: int [1:288, 1] 1 2 3 4 5 6 7 8 9 10 ...
- attr(*, "dimnames")=List of 2
..$: NULL
..$: chr "timeIndex"
Indexed by objects of class: [POSIXct,POSIXt] TZ:
xts Attributes:
NULL
..@ endTime: POSIXct[1:288], format: "2011-12-16 00:00:00" "2011-12-16 00:05:00" "2011-12-16 00:10:00" "2011-12-16 00:15:00" ...

Source

`https://www.dwd.de/`

Examples

```
library(stUPscales)
library(spacetime)

data(HS_RY20111216_stfdf)

sample.idx <- seq.default(from = 1, to = 25, by = 1)
sample     <- HS_RY20111216_stfdf[, sample.idx]

stplot(sample,
       scales=list(draw=TRUE),
       key.space="right", colorkey=TRUE,
       main="5-minute DWD sample precipitation radar imagery non-calibrated
            in STFDF format",
       cex=.74, par.strip.text=list(cex=.74))
```

`inputObs-class` *Class* `"inputObs"`

Description

The class provides a container for inputs required to invoke `Validation_Quantity` and `Validation_Quantity_Agg` methods.

Objects from the Class

Objects can be created by calls of the form `inputObs(...)`.

Slots

`id`: Object of class `"numeric"` to define an unique index for the object.

`plot`: Object of class `"numeric"`. One of 0 (no plots are cretated) or 1 (to create plots).

`delta`: Object of class `"list"` to define the time step in minutes for temporal aggregation required e.g. list(P1 = 10, wlt_obs = 10, vol_sim = 10) for defining the time steps of 10 minutes for the three variables P1, wlt_obs, vol_sim.

`observations`: Object of class `"list"` to define the observed time series.

`lev2vol`: Object of class `"list"` to define the curve for the relationship level to volume.

`namePlot`: Object of class `"character"` to define the name of the plot to create.

`legendPosition`: Object of class `"list"` with three character objects, which define the posistion of the legend for the top, second and bottom insets of the plot.

`var`: Object of class `"character"` to define the name of the variable from which the time series simulated are defined.

Methods

Validation_Quantity_Agg `signature(x = "input",y = "inputObs")`: ...

Validation_Quantity `signature(x = "input",y = "inputObs")`: ...

Author(s)

J.A. Torres-Matallana

Examples

```
showClass("inputObs")

inputObs()
```

IsReg.ts	*Wrapper function for function* is.regular *from* zoo *package for* data.frame *objects*

Description

"IsReg.ts" is a wrapping Function for Function "is.regular" from "zoo" package. Given a time series (ts) a "data.frame" object it is converted into a "xts" object, while the regularity of the object is checked. The first column of the "data.frame" should contain a character string vector to be converted via as.POSIXct accordingly with the date format (format) and time zone (tz).

Usage

```
IsReg.ts(data, format, tz)
```

Arguments

data	an object of class data.frame containing in its first column a character string vector to be converted via as.POSIXct into a date vector accordingly with the date format (format) and time zone (tz) defined
format	character string giving a date-time format as used by strptime.
tz	a time zone specification to be used for the conversion, if one is required. System-specific, but "" is the current time zone, and "GMT" is UTC (Universal Time, Coordinated). Invalid values are most commonly treated as UTC, on some platforms with a warning.

Details

"IsReg" calls the as.POSIXct function from base package to convert an object to one of the two classes used to represent date/times (calendar dates plus time to the nearest second). More details can be found in the "is.regular" function of the "zoo" package.

Value

Object of class "list". This object contains 2 elements, the first one contains a character string "_TSregular" if the xts object created is strict regular, or "_TSirregular" if it is strict irregular. More details can be found in the "is.regular" function of the "zoo" package.

Author(s)

J.A. Torres-Matallana

Examples

```
library(EmiStatR)
data("P1")

class(P1)
head(P1)

ts <- IsReg.ts(data = P1, format = "%Y-%m-%d %H:%M:%S", tz = "UTC")
str(ts)

ts[[1]]

head(ts[[2]]); tail(ts[[2]])

plot(ts[[2]], ylab = "Precipitation [mm]")
```

`Lux_boundary` *A shapefile for the boundary of the Grand-Duchy of Luxembourg*

Description

A shapefile for the country boundary of the Grand-Duchy of Luxembourg

Usage

```
data("Lux_boundary")
```

Format

The format is:
Formal class 'SpatialPolygonsDataFrame' [package "sp"] with 5 slots
..@ data :'data.frame': 1 obs. of 3 variables:
.. ..$ cat: int 1
.. ..$ X_ : Factor w/ 1 level "?": 1
.. ..$ X_1: Factor w/ 1 level "?": 1
..@ polygons :List of 1
.. ..$:Formal class 'Polygons' [package "sp"] with 5 slots
.. ..@ Polygons :List of 1
.. ..$:Formal class 'Polygon' [package "sp"] with 5 slots
.. ..@ labpt : num [1:2] 74692 93669
.. ..@ area : num 2.6e+09
.. ..@ hole : logi FALSE

.. ..@ ringDir: int 1
.. ..@ coords : num [1:1141, 1:2] 82274 82437 82449 82765 82782 ...
.. ..@ plotOrder: int 1
.. ..@ labpt : num [1:2] 74692 93669
.. ..@ ID : chr "0"
.. ..@ area : num 2.6e+09
..@ plotOrder : int 1
..@ bbox : num [1:2, 1:2] 49034 57132 106245 138879
.. ..- attr(*, "dimnames")=List of 2
.. .. ..$: chr [1:2] "x" "y"
.. .. ..$: chr [1:2] "min" "max"
..@ proj4string:Formal class 'CRS' [package "sp"] with 1 slot
.. .. ..@ projargs: chr "+proj=tmerc +lat_0=49.83333333333334 +lon_0=6.166666666666667 +k=1 +x_0=80000 +y_0%%=100000 +ellps=intl +units=m +no_defs"

Examples

```
library(stUPscales)

data(Lux_boundary)

str(boundary.Lux)
```

`Lux_precipitation` *Sample precipitation time series in the Grand-Duchy of Luxembourg*

Description

A 10-hour sample event for precipitation time series measured in 25 rain gauge stations distributed over the territory of the Grand-Duchy of Luxembourg.

Usage

```
data("Lux_precipitation")
```

Format

The format is:
An 'xts' object on 2011-12-16/2011-12-16 10:00:00 containing:
Data: num [1:61, 1:25] 0 0 0 0 0.1 0.1 0.1 0.1 0.1 0.2 ...
- attr(*, "dimnames")=List of 2

..$: NULL
..$: chr [1:25] "Dahl" "Echternach" "Esch-Sure" "Eschdorf" ...
Indexed by objects of class: [POSIXct,POSIXt] TZ:
xts Attributes:
NULL

Source

`http://agrimeteo.lu/`

Examples

```
data(Lux_precipitation)

library(xts)
head(event.subset.xts)
tail(event.subset.xts)

plot(event.subset.xts)
```

Lux_precipitation_2010_2011

Sample precipitation time series in the Grand-Duchy of Luxembourg (2-year period)

Description

A 2-year period sample for precipitation time series measured at 10-minute time step in 25 rain gauge stations distributed over the territory of the Grand-Duchy of Luxembourg.

Usage

```
data("Lux_precipitation_2010_2011")
```

Format

The format is:
An 'xts' object on 2010-01-01/2011-12-31 23:50:00 containing:
Data: num [1:105120, 1:25] 0 0 0 0 0 0 0 0 0 0 ...
- attr(*, "dimnames")=List of 2
..$: NULL
..$: chr [1:25] "Dahl" "Echternach" "Esch-Sure" "Eschdorf" ...

Indexed by objects of class: [POSIXct,POSIXt] TZ:
xts Attributes:
NULL

Source

`http://agrimeteo.lu/`

Examples

```
library(stUPscales)
data(Lux_precipitation_2010_2011)

library(xts)
head(Lux_precipitation_2010_2011)
tail(Lux_precipitation_2010_2011)

plot(Lux_precipitation_2010_2011)
```

`Lux_stations` *A SpatialPointsDataFrame with the location of 25 rain gauges in Luxembourg*

Description

A SpatialPointsDataFrame with the location of 25 rain gauges distributed over the territory of the Grand-Duchy of Luxembourg. These 25 stations are the same related to the "event.subset.xts" dataset.

Usage

```
data("Lux_stations")
```

Format

The format is:
Formal class 'SpatialPointsDataFrame' [package "sp"] with 5 slots
..@ data :'data.frame': 25 obs. of 8 variables:
.. ..$ id : Factor w/ 25 levels "1","11","12",..: 4 5 6 7 8 9 10 11 25 1 ...
.. ..$ name : Factor w/ 25 levels "Arsdorf","Christnach",..: 4 5 7 6 8 9 10 11 22 1 ...
.. ..$ north_lure: Factor w/ 24 levels "101950","102913",..: 6 NA 4 3 1 5 22 20 21 2 ...
.. ..$ east_luref: Factor w/ 25 levels "56584","56990",..: 10 25 5 7 16 19 20 24 17 1 ...
.. ..$ elev_luref: Factor w/ 25 levels "190","202","207",..: 22 5 16 25 2 11 14 1 6 20 ...

.. ..$ station_ty: Factor w/ 1 level "1": 1 1 1 1 1 1 1 1 1 1 ...
.. ..$ management: Factor w/ 1 level "ASTA": 1 1 1 1 1 1 1 1 1 1 ...
.. ..$ telemetry : Factor w/ 0 levels: NA NA NA NA NA NA NA NA NA NA ...
..@ coords.nrs : num(0)
..@ coords : num [1:25, 1:2] 66562 99810 62258 63363 74929 ...
.. ..- attr(*, "dimnames")=List of 2
.. .. ..$: NULL
.. .. ..$: chr [1:2] "coords.x1" "coords.x2"
..@ bbox : num [1:2, 1:2] 56584 64215 99810 132012
.. ..- attr(*, "dimnames")=List of 2
.. .. ..$: chr [1:2] "coords.x1" "coords.x2"
.. .. ..$: chr [1:2] "min" "max"
..@ proj4string:Formal class 'CRS' [package "sp"] with 1 slot
.. .. ..@ projargs: chr "+proj=tmerc +lat_0=49.83333333333334 +lon_0=6.166666666666667 +k=1 +x_0=80000 +y_0=100000 +ellps=intl +units=m +no_defs"

Source

`http://agrimeteo.lu/`

Examples

```
library(stUPscales)
data(Lux_stations)

str(stations)

library(sp)
plot(stations)
```

MC.analysis *Analysis of the Monte Carlo simulation*

Description

Function for running the analysis of the Monte Carlo simulation.

Usage

```
MC.analysis(x, delta, qUpper, p1.det, sim.det, event.ini, event.end,
ntick, summ.data = NULL)
```

Arguments

`x`	A `list` .
`delta`	A `numeric` value that specifies the level of aggregation required in minutes.
`qUpper`	A `character` string that defines the upper percentile to plot the confidence band of results, several options are possible `"q999"` the 99.9th percentile, `"q995"` the 99.5th percentile, `"q99"` the 99th percentile, `"q95"` the 95th percentile, `"q50"` the 50th percentile. The lower boundary of the confidence band (showed in gray in the output plots) is the 5th percentile in all cases.
`p1.det`	A `data.frame` that contains the time series of the main driving force of the system to be simulated deterministically, e.g. precipitation. This data.frame should have only two columns: the first one, Time [y-m-d h:m:s]; the second one, a numeric value equal to the magnitude of the environmental variable.
`sim.det`	A `list` that contains the results of the deterministic simulation, here the output of EmiStatR given `p1.det`. See the method `EmiStatR` from the homonym package for details.
`event.ini`	A time-date string in `POSIXct` format that defines the initial time for event analysis.
`event.end`	A time-date string in `POSIXct` format that defines the final time for event analysis.
`ntick`	A `numeric` value to specify the number of ticks in the x-axis for the event time-window plots.
`summ.data`	A `list` by default NULL. If provided, the list should contain an output of the MC.analysis function, and the analysis is done again without the calculation of some of the internal variables, therefore the analysis is faster.

Value

A `list` of length 2:

`summ`	A `list` that contains the summary statistics of the Monte Carlo simulation per output variable. Each output variable is summarised by calculating the mean `"Mean"`, standard deviation `"sd"`, variance `"Variance"`, 5th, 25th, 50th, 75th, 95th, 99.5th, 99.9th percentiles `"q05","q25","q50","q75","q95","q995"`, `"q999"`, the max `"Max"`,

the sum "Sum", time "time", and the deterministic precipitation "p1", all variables as time series.

variance A data.frame that contains the summary statistics of the variance of the Monte Carlo simulation per output variable.

Author(s)

J.A. Torres-Matallana

See Also

See also setup-class, MC.setup-methods, MC.sim-methods.

Examples

```
## the Monte Carlo simulation: MC.sim
library(EmiStatR)

# library(xts)
# data(Esch_Sure2010)
# P <- IsReg(Esch_Sure2010, format = "%Y-%m-%d %H:%M:%S", tz = "CET")
# P1 <- P[[2]]
# P1 <- P1["2010-08",][1:55]
# P1 <- cbind.data.frame(time=index(P1), P1 = coredata(P1))

data(P1)
P1 <- P1[165:(110*2),]
plot(P1[,2], typ="l")

library(stUPscales)

setting_EmiStatR <-  setup(id       = "MC_sim1",
                           nsim     = 3, # # use a larger number to have
                                         # a proper confidence band of
                                         # simulatios
                           seed     = 123,
                           mcCores  = 1,
                           ts.input = P1,
                           rng      = rng <- list(
                             qs   = 150,    # [l/PE/d]
                             CODs = c(pdf = "nor", mu = 4.378, sigma = 0.751),
                                    # log[g/PE/d]
                             NH4s = c(pdf = "nor", mu = 1.473, sigma = 0.410),
                                    # log[g/PE/d]
```

```
  qf   = 0.04,     # [l/s/ha]
  CODf = 0,              # [g/PE/d]
  NH4f = 0,              # [g/PE/d]
  CODr = c(pdf = "nor", mu = 3.60, sigma = 1.45),
         # 71 log[mg/l]
  NH4r = 1,              # [mg/l]
  nameCSO = "E1",        # [-]
  id      = 1,           # [-]
  ns      = "FBH Goesdorf",  # [-]
  nm      = "Goesdorf",  # [-]
  nc      = "Obersauer", # [-]
  numc    = 1,          # [-]
  use     = "R/I",  # [-]
  Atotal  = 36,              # [ha]
  Aimp    = c(pdf = "uni", min = 4.5, max = 25),
            # [ha]
  Cimp    = c(pdf = "uni", min = 0.25, max = 0.95),
            # [-]
  Cper    = c(pdf = "uni", min = 0.05, max = 0.60),
            # [-]
  tfS     = 1,               # [time steps]
  pe      = 650,             # [PE]
  Qd      = 5,               # [l/s]
  Dd      = 0.150,           # [m]
  Cd      = 0.18,            # [-]
  V       = 190,             #  [m3]
  lev.ini = 0.10,            # [m]
  lev2vol = list(lev = c(.06, 1.10, 1.30, 3.30),
                 # [m]
                 vol = c(0, 31, 45, 190))
                 # [m3]
),
ar.model  = ar.model <- list(
  CODs    = 0.5,
  NH4s    = 0.5,
  CODr    = 0.7),
var.model = var.model <- list(
  inp     = c("", ""), # c("CODs", "NH4s"),
            # c("", ""),
  w       = c(0.04778205, 0.02079010),
  A       = matrix(c(9.916452e-01, -8.755558e-05,
                     -0.003189094, 0.994553910),
                     nrow=2, ncol=2),
```

```
                              C       = matrix(c(0.009126591, 0.002237936,
                                                 0.002237936, 0.001850941),
                                                 nrow=2, ncol=2)))

MC_setup <- MC.setup(setting_EmiStatR)

sims <- MC.sim(x = MC_setup, EmiStatR.cores = 0)


## Monte Carlo simulation analysis: MC.analysis

# Deterministic simulation
# Definition of structure 1, E1:

E1 <- list(id = 1, ns = "FBH Goesdorf", nm = "Goesdorf", nc = "Obersauer",
           numc = 1, use = "R/I", Atotal = 36, Aimp = 25.2, Cimp = 0.80,
           Cper = 0.30, tfS = 0, pe = 650, Qd = 5,
           Dd = 0.150, Cd = 0.18, V = 190, lev.ini = 0.10,
           lev2vol = list(lev = c(.06, 1.10, 1.30, 3.30),
                          vol = c(0, 31, 45, 190))
           )

# Defining deterministic input:
library(EmiStatR)
# data(P1)

input.det <- input(spatial = 0, zero = 1e-5,
                   folder = system.file("shiny", package = "EmiStatR"),
                   cores = 0,
                   ww = list(qs = 150, CODs = 104, NH4s = 4.7),
                   inf = list(qf= 0.04, CODf = 0, NH4f = 0),
                   rw = list(CODr = 71, NH4r = 1, stat = "Dahl"),
                   P1 = P1, st = list(E1=E1), export = 0)

# Invoking `EmiStatR` with the deterministic input:
sim.det   <- EmiStatR(input.det)

# further arguments
delta <- 10 # minutes
qUpper <- "q999"

event.ini <- as.POSIXct("2016-01-02 03:20:00")
event.end <- as.POSIXct("2016-01-02 12:30:00")
```

```
# uncomment to run:
# new_analysis <- MC.analysis(x = sims, delta = delta, qUpper = qUpper,
#                             p1.det = P1, sim.det = sim.det,
#                             event.ini = event.ini, event.end = event.end,
#                             ntick = 5, summ.data = NULL)
```

`MC.analysis_generic`

Analysis of the Monte Carlo simulation (general function)

Description

General function for running the analysis of the Monte Carlo simulation.

Usage

```
MC.analysis_generic(x, delta, qUpper, data.det, sim.det, event.ini,
                    event.end, ntick, summ.data = NULL)
```

Arguments

`x`	A `list` of 1, which contains the output of the Monte Carlo simulation as a data.frame with n rows as time steps and the first column is time in format POSIXct and m columns named 1, 2, 3... m, where m is the number of Monte Carlo runs results.
`delta`	A `numeric` value that specifies the level of aggregation required in minutes.
`qUpper`	A `character` string that defines the upper percentile to plot the confidence band of results, several options are possible `"q999"` the 99.9th percentile, `"q995"` the 99.5th percentile, `"q99"` the 99th percentile, `"q95"` the 95th percentile, `"q50"` the 50th percentile. The lower boundary of the confidence band (showed in gray in the output plots) is the 5th percentile in all cases.
`data.det`	A `data.frame` that contains the time series of the main driving force of the system to be simulated deterministically, e.g. precipitation. This data.frame should have only two columns: the first one, Time [y-m-d h:m:s] in POSIXct format; the second one, a numeric value equal to the magnitude of the variable.
`sim.det`	A `list` of 1 that contains the results of the deterministic simulation, here the output given `data.det`. The format is the same as `data.det`.

`event.ini` A time-date string in `POSIXct` format that defines the initial time for event analysis.

`event.end` A time-date string in `POSIXct` format that defines the final time for event analysis.

`ntick` A `numeric` value to specify the number of ticks in the x-axis for the event time-window plots.

`summ.data` A `list` by default NULL. If provided, the list should contain an output of the MC.analysis function, and the analysis is done again without the calculation of some of the internal variables, therefore the analysis is faster.

Value

A `list` of length 2:

`summ` A `list` that contains the summary statistics of the Monte Carlo simulation per output variable. Each output variable is summarised by calculating the mean `"Mean"`, standard deviation `"sd"`, variance `"Variance"`, 5th, 25th, 50th, 75th, 95th, 99.5th, 99.9th percentiles `"q05","q25","q50","q75","q95","q995"`, `"q999"`, the max `"Max"`, the sum `"Sum"`, time `"time"`, and the deterministic data `"p1"`, all variables as time series.

`variance` A `data.frame` that contains the summary statistics of the variance of the Monte Carlo simulation per output variable.

Author(s)

J.A. Torres-Matallana

See Also

See also `setup-class`, `MC.setup-methods`, `MC.sim-methods`.

Examples

```
## Creating meta-model
Model <- function(A, B, variable.1, variable.2){
  lum <- A*variable.1 + B*variable.2
}

## Model input and parameter set-up
```

```
time <- data.frame(time = seq.POSIXt(from = as.POSIXct("2019-01-01"),
                                     to = as.POSIXct("2019-01-02"),
                                     by = 60*60*6))
data <- cbind(time, data = 25)
data

new.setup <- setup(id = "MC_1",
                   nsim = 10,
                   seed = 123,
                   mcCores = 1,
                   ts.input = data,
                   rng = rng <- list(
                     A = 1.25,
                     B = 0.75,
                     variable.1 = c(pdf = "uni", min = 0, max = 4),
                     variable.2 = c(pdf = "uni", min = 2.2, max = 3.2)
                   )
)

str(new.setup)

## Monte Carlo simulation set-up
set.seed(slot(new.setup, "seed"))

new.mc.setup <- MC.setup(new.setup)
str(new.mc.setup)

## Monte Carlo simulation
output <- data.frame(time = new.mc.setup$ts.input[,1])
output[,2:(new.mc.setup$nsim + 1)] <- NA

for(i in 1:new.mc.setup$nsim){
  for(j in 1:nrow(new.mc.setup$ts.input)){

    ## model parameter definition
    A <- new.mc.setup$par$A
    B <- new.mc.setup$par$B

    ## model input definition
    variable.1 <- new.mc.setup$par$variable.1[i,j]
    variable.2 <- new.mc.setup$par$variable.2[i,j]

    ## model evaluation
```

```
    output[j,i+1] <- Model(A, B, variable.1, variable.2)
  }
}

output <- list(output1 = output)
output

## Deterministic simulation
# model parameter definition
A <- new.mc.setup$par$A
B <- new.mc.setup$par$B

# model input definition
variable.1.det <- apply(X = new.mc.setup$par$variable.1, MARGIN = 2,
                        FUN = mean)
variable.2.det <- apply(X = new.mc.setup$par$variable.2, MARGIN = 2,
                        FUN = mean)

output.det      <- Model(A, B, variable.1.det, variable.2.det)
output.det      <- cbind(time, output.det)
output.det      <- list(out1 = output.det)
str(output.det)

## Monte Carlo analysis
delta     <- 60*6 # minutes
qUpper    <- "q95"
event.ini <- data$time[1]
event.end <- data$time[nrow(data)]
ntick     <- 1

analysis <- MC.analysis_generic(x = output, delta = delta, qUpper = qUpper,
                                data.det = data, sim.det = output.det,
                                event.ini = event.ini, event.end = event.end,
                                ntick = ntick)
```

MC.calibra-methods

Methods for Function MC.calibra

Description

Given the arguments of the method a calibration routine takes place. Method used only for internal purpose.

Methods

```
signature(x = "list", obs = "inputObs", EmiStatR.cores = "numeric")
```

`MC.setup-methods` *Methods for Function* `MC.setup`

Description

Given an object of class `setup`, the method can be invoked for setting-up the Monte Carlo simulation. The variables are sampled accordingly to their parameters specified in the slot `rng` of the `setup` object. If `ar.model` is defined in slot `ar.model`, then the specified variables are sampled from the `pdf nor` as an autorregresive (AR) model via the function `arima.sim` from base package `stats`. If `var.model` is defined in slot `var.model`, then the specified variables are sampled from the `pdf nor` as an vector autorregresive (VAR) model via the function `mAr.sim` from package `mAr` (see Barbosa, 2015, and Luetkepohl, 2005, for details). See `setup-class` for further details to define the AR and VAR models.

Usage

```
MC.setup(x)
```

Arguments

`x`	an object of class `setup`.

Methods

```
signature(x = "setup")
```

Author(s)

J.A Torres-Matallana

References

S. M. Barbosa, Package "mAr": Multivariate AutoRegressive analysis, 1.1-2, The Comprehensive R Archive Network, CRAN, 2015.

H. Luetkepohl, New Introduction to Multiple Time Series Analysis, Springer, 2005.

Examples

```
# loading a precipitation time series as input for the setup class

library(EmiStatR)
data(P1)

# A setup with three variables to be considered in the Monte Carlo
# simulation:
# var1, a constant value variable; var2, a variable sampled from a
# uniform (uni) probability distribution function (pdf) with parameters
# min and max; var3, a variable sampled from a normal (nor) pdf with
# parameters mu and sigma

ini <- setup(id = "MC_sim1", nsim = 500, seed = 123, mcCores = 1,
             ts.input = P1,
             rng = list(var1 = 150,
                        var2 = c(pdf = "uni", min = 50, max = 110),
                        var3 = c(pdf = "nor", mu = 90, sigma = 2.25))
)

MC_setup <- MC.setup(ini)
str(MC_setup)

## definition of AR models for variables var2 and var3 with AR coefficients
## 0.995 and 0.460

library(EmiStatR)
data(P1)

ini_ar <- setup(id = "MC_sim1_ar", nsim = 500, seed = 123, mcCores = 1,
                ts.input = P1,
                rng = list(var1 = 150, var2 = c(pdf = "nor", mu = 150,
                           sigma = 5),
                           var3 = c(pdf = "nor", mu = 90, sigma = 2.25)),
                ar.model = ar.model <- list(var2 = 0.995, var3 = 0.460)
)

MC_setup_ar <- MC.setup(ini_ar)
str(MC_setup_ar)

## definition of a bi-variate VAR model for variables var2 and var3

ini_var <- setup(id = "MC_sim1_ar", nsim = 500, seed = 123, mcCores = 1,
```

```
                    ts.input = P1,
                    rng = rng <- list(var1 = 150,
                                  var2 = c(pdf = "nor", mu = 150, sigma = 5),
                                  var3 = c(pdf = "nor", mu = 90, sigma = 2.25)),
                    var.model = var.model <- list( inp = c("var2", "var3"),
                                                   w = c(0.048, 0.021),
                                                   A = matrix(c(0.992, -8.8e-05,
                                                   -31e-4, 0.995), nrow=2,
                                                   ncol=2),
                                                   C = matrix(c(0.0091, 0.0022,
                                                   0.0022, 0.0019), nrow=2,
                                                   ncol=2))
)


MC_setup_var <- MC.setup(ini_var)
str(MC_setup_var)
```

MC.sim-methods ~~ *Methods for Function* MC.sim ~~

Description

Method to be invoked for running the Monte Carlo simulation. The simulator used is the method EmiStatR from the homonym package. This method should be rewritted for working with another simulator.

Usage

```
MC.sim(x, EmiStatR.cores)
```

Arguments

x	an object of class list as is defined by method MC.setup.
EmiStatR.cores	a numeric value for specifying the number of cores (CPUs) to be used in the EmiStatR method. Use zero for not use parallel computation. See class input of package EmiStatR for details.

Value

A list of length 2:

mc	A list that contains the MC_setup, timing and lap objects.
sim1	A list that contains the Monte Carlo matrices of the simulator output.

Methods

```
signature(x = "list", EmiStatR.cores = "numeric")
```

Examples

```
## the Monte Carlo simulation: MC.sim

library(EmiStatR)
data(P1)
P1 <- P1[165:(110*2),]
plot(P1[,2], typ="l")

library(stUPscales)

setting_EmiStatR <-  setup(id       = "MC_sim1",
                           nsim     = 3, # use a larger number to have
                                        # a proper confidence band of
                                        # simulations
                           seed     = 123,
                           mcCores  = 1,
                           ts.input = P1,
                           rng      = rng <- list(
                             qs   = 150,    # [l/PE/d]
                             CODs = c(pdf = "nor", mu = 4.378,
                             sigma = 0.751), # log[g/PE/d]
                             NH4s = c(pdf = "nor", mu = 1.473,
                             sigma = 0.410), # log[g/PE/d]
                             qf   = 0.04,     # [l/s/ha]
                             CODf = 0,              # [g/PE/d]
                             NH4f = 0,              # [g/PE/d]
                             CODr = c(pdf = "nor", mu = 3.60,
                             sigma = 1.45), # 71 log[mg/l]
                             NH4r = 1,              # [mg/l]
                             nameCSO = "E1",        # [-]
                             id      = 1,           # [-]
                             ns      = "FBH Goesdorf",  # [-]
                             nm      = "Goesdorf",  # [-]
                             nc      = "Obersauer", # [-]
                             numc    = 1,          # [-]
                             use     = "R/I",  # [-]
                             Atotal  = 36,             # [ha]
                             Aimp    = c(pdf = "uni", min = 4.5,
                                         max = 25), # [ha]
```

```
  Cimp    = c(pdf = "uni", min = 0.25,
              max = 0.95), # [-]
  Cper    = c(pdf = "uni", min = 0.05,
              max = 0.60), # [-]
  tfS     = 1,               # [time steps]
  pe      = 650,             # [PE]
  Qd      = 5,               # [l/s]
  Dd      = 0.150,           # [m]
  Cd      = 0.18,            # [-]
  V       = 190,             #  [m3]
  lev.ini = 0.10,            # [m]
  lev2vol = list(lev = c(.06, 1.10,
                1.30, 3.30), # [m]
                vol = c(0, 31,
                45, 190))   # [m3]
),
ar.model  = ar.model <- list(
  CODs    = 0.5,
  NH4s    = 0.5,
  CODr    = 0.7),
var.model = var.model <- list(
  inp     = c("", ""),
  # c("CODs", "NH4s"),
  # c("", ""),
  w       = c(0.04778205, 0.02079010),
  A       = matrix(c(9.916452e-01,
                     -8.755558e-05,
                     -0.003189094,
                     0.994553910),
                     nrow=2, ncol=2),
  C       = matrix(c(0.009126591,
                     0.002237936,
                     0.002237936,
                     0.001850941),
                     nrow=2, ncol=2)))

MC_setup <- MC.setup(setting_EmiStatR)

sims <- MC.sim(x = MC_setup, EmiStatR.cores = 0)
str(sims)
```

MC.summary *Summary statistics computation of Monte Carlo simulation*

Description

A function that computes the summary statistics of a Monte Carlo simulation result.

Usage

```
MC.summary(p1, data)
```

Arguments

`p1`	The independient variable. A `dataframe` object with two columns and number of rows equal to the number of rows of the Monte Carlo simulated data. The first column, named "time", contains the vector of time of the time series in format `POSIXct`. The second column contains the observations of the time series.
`data`	A `matrix` or a `dataframe` that contains the results of a Monte Carlo simulation, with number of rows equal to the number of Monte Carlo realizations and number of columns equal to the number of oservations i.e. equal to the number of rows of "p1".

Details

This function is internally invoked by the `MC.analysis` function to compute the summary statistics of the Monte Carlo simulation under analysis.

Value

A `dataframe` with n observations of 15 variables, where n is the number of columns of the "data" argument. The 15 variables are time series with the summary statistics of the Monte Carlo data: 1) `idx`: an index for the dataset equal to 1; 2) `Mean`: the mean; 3) `Sd`: the standard deviation; 4) `Variance`, the variance; 5) `q05`: the five percent quantile; 6) `q25`: the 25 percent quantile; 7) `q50`: the 50 percent quantile; 8) `q75`: the 75 percent quantile; 9) `q95`: the 95 percent quantile; 10) `q995`: the 99.5 percent quantile; 11) `q999`: the 99.9 percent quantile; 12) `Max`: the maximum; 13) `Sum`: the sum; 14) `time`: the time; 15) `p1`: the independient variable.

Author(s)

J.A. Torres-Matallana

Examples

```
library(stUPscales)
library(EmiStatR)
```

```
data(P1)
colnames(P1)

new_data <- t(matrix(data = rep(runif(nrow(P1), 10, 100), 5),
                     nrow = nrow(P1), ncol = 5))
new_summary <- MC.summary(p1 = P1, data = new_data)
str(new_summary)
head(new_summary)
```

`MC.summary.agg` *Summary statistics computation of aggregated Monte Carlo simulation*

Description

A function that computes the summary statistics of aggregated Monte Carlo simulation result.

Usage

```
MC.summary.agg(summ, det, delta, func.agg, func.agg.p)
```

Arguments

`summ`	A `dataframe` with n observations of 15 variables, where n is the number observations or time steps of the data. The 15 variables are time series with the summary statistics of the Monte Carlo data. This dataframe is in the format as is described in the `MC.summary` function value.
`det`	A `dataframe` that contains the deterministic simulation.
`delta`	A `numeric` value that represents the level of aggregation (required time stemp) in minutes.
`func.agg`	The aggregation function to be applied to the `summ` dataframe.
`func.agg.p`	The aggregation function to be applied to the independient variable `p1` from `summ` dataframe.

Value

A `dataframe` containing the `summ` data aggregated to the level defined by `delta`

Author(s)

J.A. Torres-Matallana

See Also

See Also as `MC.summary`

Examples

```
library(stUPscales)
library(EmiStatR)

data(P1)
colnames(P1)

new_data <- t(matrix(data = rep(runif(nrow(P1), 10, 100), 5),
                     nrow = nrow(P1), ncol = 5))
new_summary <- MC.summary(p1 = P1, data = new_data)
str(new_summary)
head(new_summary)

# deterministic simulation
det <- rnorm(nrow(P1), 45, .15)

# level of aggregation
delta <- 60*2 # 2 hours

new_summary_agg <- MC.summary.agg(summ = new_summary, det, delta,
                                  func.agg = mean, func.agg.p = sum)
str(new_summary_agg)
head(new_summary_agg)
```

`PlotEval` *Function to execute evaluation plot*

Description

This function creates an evaluation plot for the Monte Carlo simulation result.

Usage

```
PlotEval(eval, ts, gof1, namePlot, pos1, pos2, pos3)
```

Arguments

`eval`	A `data.frame` with n observations of seven variables: 1) `time`: A `POSIXct` object with format "%Y-%m-%d %H:%M:%S" defining the time vector; 2) `column 2`: a numeric vector containing the values of

the observed variable, which is the first variable of the `Level2Volume` relationship; 3) `column 3`: a numeric vector containing the values for the second variable of the `Level2Volume` relationship; 4) `column 4`: a numeric vector containing the corresponding simulated values for the second variable of the `Level2Volume` relationship; 5) `column 5`: a numeric vector containing the difference between the vectors `volT_sim` and `volT_obs`. 6) `Rainfall`: a numeric vector named "Rainfall" containing the values of the driving force variable used in the simulations, e.g. rainfall. 7) `column 7`: (Optional) a numeric vector containing the values of the driving force variable used in the simulations in other measurement units, e.g. rainfall in intensity units if rainfall is the driving force of the simulations.

`ts` An `xts` object representing the `eval data.frame` indexed by the `time` vector of the `eval` argument: containing six data variables as it is defined by the `eval` argument: 1) `column 2`; 2) `column 3`; 3) `column 4`; 4) `column 5`; 5) `Rainfall`; 6) `column 7`.

`gof1` A matrix with the output of `GoF` function.

`namePlot` A character string defining the name of the plot to be created.

`pos1` Location to place the legend on the inside of the first sub-plot frame. Can be one of "bottomright", "bottom", "bottomleft", "left", "topleft", "top", "topright", "right" and "center".

`pos2` Location to place the legend on the inside of the second sub-plot frame. Can be one of "bottomright", "bottom", "bottomleft", "left", "topleft", "top", "topright", "right" and "center".

`pos3` Location to place the legend on the inside of the third sub-plot frame. Can be one of "bottomright", "bottom", "bottomleft", "left", "topleft", "top", "topright", "right" and "center".

Value

The function creates a plot in the current working directory with the goodness-of-fit between simulations and observations. The plot is provided in pdf format.

Author(s)

J.A. Torres-Matallana

Examples

```
time <- seq(from = as.POSIXct("2017-11-09"), by = 60*60*24, length.out = 230)
```

```
# the time vector
data <- cbind.data.frame(time, NA) # a NA vector
data[,3] <- rnorm(230, .25, .1) # random normal distributed data, obs
data[,4] <- data[,3]*1.2  # positive correlated data, sim
data[,5] <- data[,4] - data[,3] # difference sim and obs
data[,6] <- 0 # driving force
data[,7] <- NA # a NA vector

colnames(data) <- c("time", "var1", "obs", "sim", "difference", "Rainfall",
                    "Rainfall2")
head(data)

ts <- IsReg.ts(data, "%Y-%m-%d", "ECT")
ts <- ts[[2]]

gof.new <- GoF(data, 4, 3, "")
gof.new

PlotEval(data, ts, gof.new, "ExamplePlot", "topright", "topright",
         "topright")
```

`PlotMC.event` *A plot function for time series events*

Description

This is an internal function invoked by `MC.analysis` function to generate an event plot of the time series under analysis. A event means a time series with length lower to one month i.e. sub-montly time series.

Usage

```
PlotMC.event(summ, summ1, obs, det.var, det.var1, namePlot, ylab, ylab1,
             ntick, qUpper)
```

Arguments

`summ`	A data.frame with n observations of m variables as is provided by the output of function `MC.summary.agg` for the first variable to be plotted.
`summ1`	A data.frame with n observations of m variables as is provided by the output of function `MC.summary.agg` for the second variable to be plotted.
`obs`	A numeric value equal to 0. used for internal use.

`det.var` A character string defining the name of the first variable from `summ` object to be plotted.

`det.var1` A character string defining the name of the second variable from `summ` object to be plotted.

`namePlot` A character string defining the name of the plot. The file created with the plot has this name.

`ylab` A character string to define the label of the axes y for the first variable sub-plot.

`ylab1` A character string to define the label of the axes y for the second variable sub-plot.

`ntick` A numeric value integer which defines the number of tick marks in the axis x of the sub-plots.

`qUpper` A character string that defines the upper percentile to plot the confidence band of results, several options are possible "q999" the 99.9th percentile, "q995" the 99.5th percentile, "q99" the 99th percentile, "q95" the 95th percentile, "q50" the 50th percentile. The lower boundary of the confidence band (showed in gray in the output plots) is the 5th percentile in all cases.

Value

The function creates the plot in the current working directory. The format of the plot is pdf.

Author(s)

J.A. Torres-Matallana

Examples

```
library(stUPscales)
library(EmiStatR)

# definition of the first summary.agg object
data("P1")
P1 <- P1[1:1100,]

new_data <- matrix(data = NA, nrow = nrow(P1), ncol = 55)
for(i in 1:55){
  new_data[,i] <- matrix(data = rnorm(nrow(P1), 45, 15),
```

```
                          nrow = nrow(P1), ncol = 1)
}
new_data <- t(new_data)

new_summary <- MC.summary(p1 = P1, data = new_data)

# deterministic simulation
det <- rnorm(nrow(P1), 45, 15)
det <- cbind(det, rnorm(nrow(P1), 55, 23))
colnames(det) <- c("det1", "det2")

# level of aggregation
delta <- 60*2 # 2 hours

new_summary_agg <- MC.summary.agg(summ = new_summary, det, delta,
                                  func.agg = mean, func.agg.p = sum)

# definition of the second summary.agg object
new_data1 <- matrix(data = NA, nrow = nrow(P1), ncol = 55)
for(i in 1:55){
  new_data1[,i] <- matrix(data = rnorm(nrow(P1), 55, 23),
                          nrow = nrow(P1), ncol = 1)
}
new_data1 <- t(new_data1)

new_summary1 <- MC.summary(p1 = P1, data = new_data1)
new_summary_agg1 <- MC.summary.agg(summ = new_summary1, det, delta,
                                   func.agg = mean, func.agg.p = sum)

# creating the plot for the event
PlotMC.event(summ = new_summary_agg, summ1 = new_summary_agg1, obs = 0,
             det.var = "det1", det.var1 = "det2", namePlot = "ExamplePlot",
             ylab = "Variable 1 [units]", ylab1 = "Variable 2 [units]",
             ntick=10, qUpper= "q95")
```

PlotMC.season *A plot function for time series seasons*

Description

This is an internal function invoked by `MC.analysis` function to generate a season plot of the time series under analysis. A season means a time series with length greater to one month e.g. montly, yearly, decadal time series.

Usage

```
PlotMC.season(summ1, namePlot, ylab, qUpper)
```

Arguments

`summ1`	A data.frame with n observations of m variables as is provided by the output of function `MC.summary.agg` for the variable to be plotted, which the summary was computed.
`namePlot`	A character string defining the name of the plot. The file created with the plot has this name.
`ylab`	A character string to define the label of the axes y for the variable to plot.
`qUpper`	A character string that defines the upper percentile to plot the confidence band of results, several options are possible "q999" the 99.9th percentile, "q995" the 99.5th percentile, "q99" the 99th percentile, "q95" the 95th percentile, "q50" the 50th percentile. The lower boundary of the confidence band (showed in gray in the output plots) is the 5th percentile in all cases.

Value

The function creates the plot in the current working directory. The format of the plot is pdf.

Author(s)

J.A. Torres-Matallana

Examples

```
library(stUPscales)
library(EmiStatR)

data("P1")
P1 <- P1[1:550,]
new_data <- matrix(data = NA, nrow = nrow(P1), ncol = 55)
for(i in 1:55){
  new_data[,i] <- matrix(data = rnorm(nrow(P1), 22, 11),
                         nrow = nrow(P1), ncol = 1)
}
new_data <- t(new_data)
```

```
new_summary <- MC.summary(p1 = P1, data = new_data)
head(new_summary)
dim(new_summary)

new_summary$month <- strftime(new_summary[,"time"], format = "%Y-%m")

PlotMC.season(summ1 = new_summary, namePlot = "ExamplePlot",
              ylab = "Variable 1 [units]", qUpper = "q95")
```

---

setup-class *Class* "setup"

Description

Class to create objects of signature setup. setup object should be passed to the method MC.setup.

Objects from the Class

Objects can be created by calls of the form `setup()`.

Slots

`id`: Object of class `"character"` to identify the Monte Carlo simulation.

`nsim`: Object of class `"numeric"` to specify the number of Monte Carlo runs.

`seed`: Object of class `"numeric"` to specify the seed of the random numbers generator.

`mcCores`: Object of class `"numeric"` to specify the number of cores (CPUs) to be used in the Monte Carlo simulation.

`ts.input`: Object of class `"data.frame"` that contains the time series of the main driving force of the system to be simulated, e.g. precipitation. This data.frame should have at least two columns: the first one, Time [y-m-d h:m:s]; the second one, a numeric value equal to the magnitude of the environmental variable. This data.frame can also contain more that one column to allow several time series in several columns. If the data.frame has more than two columns, then the number of columns should be at least equal to nsim. If the number of columns is greater than nsim, the columns in excess are not recycled because the simulation will last nsim iterations.

`rng`: Object of class `"list"` that contains the names and values of the variables to be used in the Monte Carlo simulation. Five modes are available: 1) constant value, i.e. this variable will have a constant value along the Monte Carlo simulation; 2)

a variable sampled from a uniform (uni) probability distribution function (pdf) with parameters for the lower boundary min and upper boundary max; 3) a variable sampled from a normal (nor) pdf with parameteres mean mu and standard deviation sigma; 4) a variable sampled from an autorregresive (AR) model and normal (nor) pdf with parameteres mean mu and standard deviation sigma, the coefficients of the AR model should be defined in the slot ar.model; 5) a variable sampled from an vector autorregresive (VAR) model and normal (nor) pdf with parameteres mean mu and standard deviation sigma, this mode is enabled by defining the vector of intercept terms w, the matrix of AR coefficients A, and the noise covariance matrix C in the slot var.model. See examples for the definition of this slot.

`ar.model`: Object of class `"list"` containing the coefficients of the AR model as vectors which name is the variable to be modeled and length the order of the model as is required for function arima.sim from the base package stats. The named variables here should correspond to a pdf nor in the slot rng. See examples for the definition of this slot.

`var.model`: Object of class `"list"` containing the the vector of intercept terms w, the matrix of AR coefficients A, and the noise covariance matrix C of the VAR model which name is the variable to be modeled and length the order of the model as is required for function mAr.sim from the package mAr. The named variables in this slot should correspond to a pdf nor in the slot rng. The current implementation considers the bi-variate case. See examples for the definition of this slot. For mathemamtical details see Luetkepohl (2005).

Methods

MC.setup `signature(x = "setup")`: execute MC.setup function

Author(s)

J.A Torres-Matallana

References

S. M. Barbosa, Package "mAr": Multivariate AutoRegressive analysis, 1.1-2, The Comprehensive R Archive Network, CRAN, 2015.

H.Luetkepohl, New Introduction to Multiple Time Series Analysis, Springer, 2005.

Examples

```
# loading a precipitation time series as input for the setup class
```

```
library(EmiStatR)
data(P1)

# A setup with three variables to be considered in the Monte Carlo
#simulation:
# var1, a constant value variable; var2, a variable sampled from a
# uniform (uni) probability distribution function (pdf) with
# parameters min and max;
# var3, a variable sampled from a normal (nor) pdf with
parameters mu and sigma

ini <- setup(id = "MC_sim1", nsim = 500, seed = 123, mcCores = 1,
             ts.input = P1,
             rng = list(var1 = 150, var2 = c(pdf = "uni", min = 50,
                                             max = 110),
                        var3 = c(pdf = "nor", mu = 90,
                                 sigma = 2.25))
)

str(ini)

## definition of AR models for variables var2 and var3 with AR coefficients
## 0.995 and 0.460

library(EmiStatR)
data(P1)

ini_ar <- setup(id = "MC_sim1_ar", nsim = 500, seed = 123, mcCores = 1,
                ts.input = P1,
                rng = list(var1 = 150, var2 = c(pdf = "nor", mu = 150,
                                                sigma = 5),
                           var3 = c(pdf = "nor", mu = 90,
                                    sigma = 2.25)),
                ar.model = ar.model <- list(var2 = 0.995,
                                            var3 = 0.460)
)

str(ini_ar)

## definition of a bi-variate VAR model for variables var2 and var3

ini_var <- setup(id = "MC_sim1_ar", nsim = 500, seed = 123, mcCores = 1,
```

```
                    ts.input = P1,
                    rng = rng <- list(var1 = 150,
                                      var2 = c(pdf = "nor", mu = 150,
                                               sigma = 5),
                                      var3 = c(pdf = "nor", mu = 90,
                                               sigma = 2.25)),
                    var.model = var.model <- list( inp = c("var2", "var3"),
                                                   w = c(0.048, 0.021),
                                                   A = matrix(c(0.992,
                                                   -8.8e-05,
                                                   -31e-4, 0.995),
                                                   nrow=2, ncol=2),
                                                   C = matrix(c(0.0091,
                                                   0.0022, 0.0022, 0.0019),
                                                   nrow=2, ncol=2))
)

str(ini_var)
```

`Validation_Quantity-methods`

Methods for Function `Validation_Quantity`

Description

Given the arguments of the method a validation routine takes place. Method used only for internal purpose.

Methods

```
signature(x = "input", y = "inputObs")
```

`Validation_Quantity_Agg-methods`

Methods for Function `Validation_Quantity_Agg`

Description

Given the arguments of the method a validation routine takes place. Method used only for internal purpose.

Methods

```
signature(x = "input", y = "inputObs")
```

Summary

Sustainable urban water management (SUWM) is becoming a global priority due to the impact of urbanisation on natural and urban ecosystems. The global population in 2018 was equivalent to 55 per cent of the world's population residing in urban areas. By 2050, the global population will reach 9.7 billion, with 68 per cent of the world's population (i.e., 6.6 billion people) projected to be urban. With the increase of urbanisation development the lack of basic urban infrastructure (i.e., water and energy supplies, sanitation, education, and green space or parks) could become more acute for the sustainability of cities in the future, especially in developing countries.

Urbanisation not only impacts the hydrological cycle, but also has an interlinked impact on the urban landscape ecosystem and its evolution. This has been recognised in the last two decades by governments, especially in more economically developed countries, with the advent of concepts such as "Sustainable Development", and "Nature-Based Solutions", which focus primarily on what nature can provide to humans. This is a response to the urgency of solving urban water problems and integrating solutions with new urban water management strategies and practices.

SUWM is a key tool contributing to several of the United Nations' Sustainable Development Goals. Integrated urban drainage models are primary components of monitoring systems and essential decision-making tools for SUWM. However, it is paramount to recognise that in environmental modelling, and hence also in SUWM, every model contains uncertainties to some degree. This is because any model makes simplifications and assumptions about the real-world processes involved, while model inputs are rarely if ever known without error. Quantification of model input and output uncertainty is essential for characterising inputs and choosing objectively the suitable configuration of the model for addressing a specific task related to integrated urban drainage modelling (IUDM).

Uncertainty quantification in integrated environmental models with emphasis in SUWM, specifically urban stormwater system models (USSM's) used in decision-making for environmental protection, require that the accuracy of model outputs is known and meets predefined standards. Uncertainty propagation in SUWM is yet not well understood.

Four important problems are identified:

1. Full hydrodynamic USSM's are complex and require a highly intense computational budget, which constitutes a constraint when long-term simulations or uncertainty propagation analysis are required.

2. Although catchment average precipitation is a key component of USSM's, catchment average precipitation is not always accurately known when derived from measurements at point support, i.e. by rain gauges. To estimate the precipitation in a catchment given a known precipitation time series at a location outside of the catchment is often needed, while also quantify the uncertainty associated with the estimated catchment average precipitation.

3. Software tools for temporal uncertainty propagation for USSM's are not generally available.

4. Many studies in USSM's do not pay attention to uncertainty and uncertainty propagation. Statistical uncertainty analysis of USSM's is a relatively new subject that largely needs to be developed while very few solid applications have been done. This is linked to the second problem, because uncertainty propagation analysis can only be done if the uncertainty sources are quantified. It is also linked to the third problem because of the need for software tools with capabilities to efficiently propagate input uncertainty in USSM's, and requires contributions from multiple disciplines (i.e., hydrology, statistics, computer science, geoinformatics).

This dissertation addresses all these problems, which represent an opportunity to develop new methods and software tools to overcome the current limitations exposed. Four research objectives, directly related to the four main problems identified in urban storm water modelling, were defined in **Chapter 1**, the General Introduction. This chapter also provides a bibliographic analysis and literature review on uncertainty analysis in urban stormwater systems modelling (USSM). It concludes that uncertainty analysis in USSM is not yet well understood, and, therefore, very few applications have been established. Answers to the four main research objectives are developed through Chapters 2 to 5, which together with the General Introduction and Synthesis (**Chapter 6**) compose this dissertation.

Chapter 2 presents a simplified mechanistic urban water quality model, EmiStatR, to represent the overall dynamic behaviour of the Combined Sewer Overflow (CSO) spill volume, load, and concentration of Chemical Oxygen Demand (COD) and ammonium (NH_4). The implementation of the model was done in the programming language for statistical computing, R, with an inherent parallel computing and scalable capability that allows fast calculations for scenarios of high complexity and for long-term simulations to test hypotheses in USSM. For a case study in Luxembourg, the model was calibrated, tested and validated by comparing the performance against a complex mechanistic model

that uses the "de Saint-Venant" partial differential equations to describe the full flow routing in the pipes of the sewer network. The case study showed that in small catchments (i.e., area of 30 ha or less) EmiStatR achieved satisfactory accuracy, similar to that of the model of much higher complexity.

Chapter 3 demonstrates that precipitation is the most active flux and major input of hydrological systems, including urban stormwater systems. Due that precipitation controls hydrological states (soil moisture and groundwater level), and fluxes (runoff, evapotranspiration and groundwater recharge), precipitation plays a paramount role in urban hydrology. It controls the fluxes towards combined sewer tanks and the dilution of chemical and organic compounds in the wastewater. Chapter 3 explores the need to develop a method to estimate the precipitation in a catchment, given a known precipitation time series at a location outside the catchment, while quantifying the uncertainty associated with the estimation. Such method is very useful for real-world applications in urban hydrology because practitioners are often confronted with few observations inside catchments. A new method for a first-order multivariate autoregressive model for conditional simulation of input precipitation based on a multiplicative error model, suitable to estimate precipitation time series in an ungauged catchment given the precipitation time series at two neighbouring precipitation gauges located close by, is developed and presented in Chapter 3. The method was used to generate precipitation ensembles accounting for the uncertainty associated in the estimation at the ungauged location. Such simulated precipitation ensembles can be used as model input for urban water models in tasks related to uncertainty propagation analysis, to account for uncertainty in improved urban water system design and to better assess environmental and economic impacts of, for example, CSO's over the receiving waters.

Chapter 4 introduces spatio-temporal Uncertainty Propagation across multiple scales, stUPscales, an R package developed to support integrated environmental modelling in tasks related to coupling sub-models at different spatial and temporal scales, accounting for change of support procedures, i.e. aggregation and disaggregation. The package includes methods and functions suitable in model input uncertainty characterisation and propagation. stUPscales constitutes a contribution to the state-of-the-art of open source tools that support uncertainty propagation analysis in the temporal and spatio-temporal domains. Chapter 4 also presents examples of applications to illustrate the package, specifically for uncertainty propagation in environmental modelling in the urban water domain. The chapter concludes with two application examples of stUPscales that demonstrate its suitability for characterising uncertainty in spatial, temporal and spatio-temporal environmental variables (model inputs) as probability distribution functions (pdf's) and as uni- and multi-variate autoregressive models. As a concluding finding, it was demonstrated that it is possible to sample from these pdf's to generate realisations of autoregressive models, and to support Monte Carlo uncertainty propagation analysis.

Chapter 5 studies the problem of temporal uncertainty, which is often ignored in urban water systems modelling. Due to the lack of user-friendly implementations, commercial software used in engineering practice usually ignores input variables and their uncertainty propagation. This can have serious consequences, such as incorrectly sized urban drainage systems and incorrect estimates of pollution released into the environment. This chapter also introduces the application of Monte Carlo simulation to the uncertainty propagation analysis of urban rainwater system modeling based on the methods introduced in the previous chapters, and applies it to the case study of the Haute-Sûre catchment area in Luxembourg. The case study uses the EmiStatR model to simulate volume and substance flow through simplified representation of drainage systems and processes. Monte Carlo uncertainty propagation analysis shows that the uncertainty of COD and NH4 loading and concentration may be high and have large temporal variability. Furthermore, Chapter 5 includes a stochastic sensitivity analysis that assesses the uncertainty contributions of input variables to the model output. This analysis showed that precipitation has the largest contribution to output uncertainty related with water quantity variables, such as volume in the chamber, overflow volume, and flow. A Monte Carlo simulation procedure was used to propagate input uncertainty and shows that, among the water quantity output variables, the overflow flow is the most uncertain output variable. Among water quality variables, the annual average spill COD concentration and the average spill NH_4 concentration are the most uncertain model outputs. Finally, this chapter demonstrates how uncertainty propagation can help to assess water quality impacts for the receiving river.

The dissertation synthesis is given in **Chapter 6**. It discusses the findings of this dissertation, highlighting the lessons learned and future directions foreseen. It provides a reflection regarding the change of paradigm, in which the current global cities are rethinking their approaches to flood risk management, complementing traditional grey infrastructure for flood defence, with Nature-Based Solutions toward implementing approaches for water resilience. It signals a shift with respect to nature in urban planning and water management policies. This last chapter concludes that it is of paramount importance in sustainable urban water management to pay attention to uncertainty. Addressing uncertainty characterisation and uncertainty propagation in the modelling chain for the design of Nature-Based Solutions, requires tailored methods and software tools that enable temporal, spatial and spatio-temporal uncertainty characterisation and propagation across multiple scales, while linking several modelling sub-modules.

This dissertation constitutes a contribution to facilitate dimensioning and design of the required infrastructure for decision-making in sustainable urban water management as a "Blue-Green-Grey" vision for the cities of the future, by paying specific attention to model input uncertainties and their propagation to model outputs as used by decision makers.

Resumen

La gestión sostenible del agua urbana (SUWM, del inglés sustainable urban water management) se está convirtiendo en una prioridad mundial debido al impacto de la urbanización en los ecosistemas naturales y urbanos. En 2018, la población mundial equivalía al 55% de la población mundial que residía en zonas urbanas. En 2050, la población mundial alcanzará los 9.700 millones, y se prevé que el 68% de la población mundial (es decir, 6.600 millones de personas) sea urbana. Con el aumento del desarrollo de la urbanización, la falta de infraestructura urbana básica (es decir, suministros de agua y energía, saneamiento, educación y espacios verdes o parques) podría agudizarse para la sostenibilidad de las ciudades en el futuro, especialmente en los países en desarrollo.

La urbanización no sólo repercute en el ciclo hidrológico, sino que también tiene un impacto interrelacionado en el ecosistema del paisaje urbano y su evolución. Esto ha sido reconocido en las dos últimas décadas por los gobiernos, especialmente en los países económicamente más desarrollados, con la aparición de conceptos como "Desarrollo Sostenible", y "Soluciones basadas en la naturaleza", que se centran principalmente en lo que la naturaleza puede proporcionar a los humanos. Se trata de una respuesta a la urgencia de resolver los problemas del agua urbana y de integrar las soluciones con nuevas estrategias y prácticas de gestión del agua urbana.

La SUWM es una herramienta clave que contribuye a varios de los Objetivos de Desarrollo Sostenible de las Naciones Unidas. Los modelos integrados de drenaje urbano son componentes primarios de los sistemas de control y herramientas esenciales para la toma de decisiones en la SUWM. Sin embargo, es primordial reconocer que en la modelación medioambiental, y por tanto también en la SUWM, todo modelo tiene incertidumbre en cierta medida. Esto se debe a que cada modelo hace simplificaciones y suposiciones sobre los procesos del mundo real, mientras que las entradas del modelo rara vez se conocen sin errores. La cuantificación de la incertidumbre de las entradas y salidas del modelo es esencial para caracterizar las entradas y elegir objetivamente la configuración adecuada del modelo para abordar una tarea específica relacionada con la modelación del drenaje urbano integrado (IUDM, del inglés integrated urban drainage modelling).

La cuantificación de la incertidumbre en los modelos ambientales integrados, con énfasis en la SUWM, específicamente en los modelos de sistemas urbanos de aguas pluviales

(USSM's, del inglés urban stormwater system models) utilizados en la toma de decisiones para la protección del medio ambiente, requieren que la precisión de los resultados del modelo sea conocida y cumpla con estándares predefinidos. La propagación de la incertidumbre en la SUWM todavía no es bien conocida.

Se identifican cuatro problemas importantes:

1. Los USSM's hidrodinámicos completos son complejos y requieren un presupuesto computacional muy intenso, lo que constituye una limitación cuando se requieren simulaciones a largo plazo o análisis de propagación de la incertidumbre.

2. Aunque la precipitación media de la cuenca es un componente clave de los USSM's, la precipitación media de la cuenca no siempre se conoce con exactitud cuando se deriva de las mediciones en soporte puntual, es decir, mediante pluviómetros. A menudo es necesario estimar la precipitación en una cuenca dada una serie temporal de precipitación conocida en un lugar fuera de la cuenca, y al mismo tiempo cuantificar la incertidumbre asociada a la precipitación media estimada de la cuenca.

3. En general, no se dispone de herramientas de software para la propagación de la incertidumbre temporal para los USSM's.

4. Muchos estudios sobre USSM's no prestan atención a la incertidumbre ni a la propagación de la misma. El análisis estadístico de la incertidumbre de los USSM's es un tema relativamente nuevo que necesita desarrollarse en gran medida, mientras que se han realizado muy pocas aplicaciones sólidas. Esto está relacionado con el segundo problema, porque el análisis de la propagación de la incertidumbre sólo puede hacerse si se cuantifican las fuentes de incertidumbre. También está relacionado con el tercer problema debido a la necesidad de herramientas computacionales con capacidades para propagar eficientemente la incertidumbre de entrada en los USSM's, y requiere contribuciones de múltiples disciplinas (es decir, hidrología, estadística, ciencias computacionales, geoinformática).

Esta tesis aborda todos estos problemas, que representan una oportunidad para desarrollar nuevos métodos y herramientas de software para superar las actuales limitaciones expuestas. En el **Capítulo 1**, Introducción General, se definieron cuatro objetivos de investigación directamente relacionados con los cuatro problemas principales identificados en la modelación de las aguas pluviales urbanas. Este capítulo también proporciona un análisis bibliográfico y una revisión de la literatura sobre el análisis de incertidumbre en la modelación de sistemas urbanos de aguas pluviales (USSM). Se concluye que el análisis de la incertidumbre en la modelación de sistemas urbanos de aguas pluviales aún no se conoce bien y, por lo tanto, se han establecido muy pocas aplicaciones. Las respuestas a los cuatro objetivos principales de la investigación se desarrollan a través de los Capítulos 2 a 5, que junto con la Introducción General y la Síntesis (**Capítulo 6**) componen esta disertación.

El **Capítulo 2** presenta un modelo mecanicista simplificado de la calidad del agua urbana, EmiStatR, para representar el comportamiento dinámico global del volumen de vertido del alcantarillado combinado (CSO, del inglés Combined Sewer Overflow), la carga y la concentración de la Demanda Química de Oxígeno (COD, del inglés Chemical Oxygen Demand) y del amonio (NH_4). La implementación del modelo se realizó en el lenguaje de programación para la computación estadística, R, con una capacidad inherente de computación en paralelo y escalable que permite realizar cálculos rápidos para escenarios de alta complejidad y para simulaciones a largo plazo para probar hipótesis en USSM. Para un caso de estudio en Luxemburgo, el modelo se calibró, probó y validó comparando el rendimiento con un modelo mecanicista complejo que utiliza las ecuaciones diferenciales parciales "de Saint-Venant" para describir el tránsito completo del flujo en las tuberías de la red de alcantarillado. El caso de estudio demostró que en cuencas pequeñas (es decir, con una superficie de 30 ha o menos) EmiStatR alcanzó una precisión satisfactoria, similar a la del modelo de mayor complejidad.

El **Capítulo 3** demuestra que la precipitación es el flujo más activo y la principal aportación de los sistemas hidrológicos, incluidos los sistemas urbanos de aguas pluviales. Dado que la precipitación controla los estados hidrológicos (humedad del suelo y nivel de las aguas subterráneas) y los flujos (escorrentía, evapotranspiración y recarga de las aguas subterráneas), la precipitación desempeña un papel primordial en la hidrología urbana. Controla los flujos hacia los depósitos de alcantarillado combinado y la dilución de los compuestos químicos y orgánicos de las aguas residuales. El Capítulo 3 explora la necesidad de desarrollar un método para estimar la precipitación en una cuenca, dada una serie temporal de precipitación conocida en un lugar al exterior de la cuenca, cuantificando al mismo tiempo la incertidumbre asociada a la estimación. Este método es muy útil para las aplicaciones del mundo real en hidrología urbana, ya que los profesionales se enfrentan a menudo con cuencas poco instrumentadas. En el Capítulo 3 se desarrolla y presenta un nuevo método para un modelo autorregresivo multivariable de primer orden para la simulación condicional de la precipitación de entrada, basado en un modelo de error multiplicativo, adecuado para estimar las series temporales de precipitación en una cuenca no instrumentada, dadas las series temporales de precipitación en dos pluviógrafos vecinos situados cerca. El método se utilizó para generar conjuntos de precipitación que tienen en cuenta la incertidumbre asociada a la estimación en el lugar no instrumentado. Estos conjuntos de precipitaciones simuladas pueden utilizarse como entrada en los modelos de aguas urbanas en tareas relacionadas con el análisis de la propagación de la incertidumbre, para tener en cuenta la incertidumbre en la mejora del diseño de los sistemas de aguas urbanas y para evaluar mejor los impactos ambientales y económicos de, por ejemplo, los CSO's sobre las aguas receptoras.

El **Capítulo 4** presenta la propagación espacio-temporal de la incertidumbre a través de múltiples escalas (stUPscales, del inglés spatio-temporal Uncertainty Propagation across multiple scales) un paquete computacional de R desarrollado para apoyar la modelación

ambiental integrada en tareas relacionadas con el acoplamiento de submodelos a diferentes escalas espaciales y temporales, teniendo en cuenta procedimientos de cambio de soporte espacial y temporal, es decir, agregación y desagregación. El paquete computacional incluye métodos y funciones adecuados en la caracterización y propagación de la incertidumbre de las entradas del modelo. stUPscales constituye una contribución al estado del arte de las herramientas de código abierto que apoyan el análisis de la propagación de la incertidumbre en los dominios temporal y espacio-temporal. El Capítulo 4 también presenta ejemplos de aplicaciones para ilustrar el paquete computacional, específicamente para la propagación de la incertidumbre en la modelación medioambiental en el ámbito del agua urbana. El capítulo concluye con dos ejemplos de aplicación de stUPscales que demuestran su idoneidad para caracterizar la incertidumbre en variables ambientales espaciales, temporales y espacio-temporales (entradas del modelo) como funciones de distribución de probabilidad (pdf's, del inglés probability distribution functions) y como modelos autorregresivos uni- y multivariados. Como conclusión, se demostró que es posible muestrear a partir de estas pdf's para generar realizaciones de modelos autorregresivos y apoyar el análisis de propagación de la incertidumbre mediante simulaciones de Montecarlo.

El **Capítulo 5** estudia el problema de la incertidumbre temporal, que a menudo se ignora en la modelación de los sistemas hídricos urbanos. Debido a la falta de implementaciones fáciles de usar, el software comercial utilizado en la práctica de la ingeniería suele ignorar las variables de entrada y su propagación de la incertidumbre. Esto puede tener graves consecuencias, como el dimensionamiento incorrecto de los sistemas de drenaje urbano y las estimaciones incorrectas de la contaminación vertida al medio ambiente. En este capítulo también se presenta la aplicación de la simulación de Montecarlo al análisis de la propagación de la incertidumbre en la modelación de sistemas urbanos de aguas pluviales, basándose en los métodos introducidos en los capítulos anteriores, y se aplica al estudio de caso de la cuenca hidrográfica de Haute-Sûre, en Luxemburgo. El caso de estudio utiliza el modelo EmiStatR para simular el flujo de volumen y de sustancias mediante una representación simplificada de los sistemas y procesos de drenaje. El análisis de propagación de la incertidumbre mediante las simulaciones de Montecarlo muestra que la incertidumbre de la carga y la concentración de COD y NH_4 puede ser elevada y presentar una gran variabilidad temporal. Además, el capítulo 5 incluye un análisis de sensibilidad estocástica que evalúa las contribuciones de incertidumbre de las variables de entrada al resultado del modelo. Este análisis mostró que la precipitación tiene la mayor contribución a la incertidumbre de salida relacionada con las variables de cantidad de agua, como el volumen en la cámara, el volumen de desbordamiento y el caudal. El procedimiento de simulación de Montecarlo para propagar la incertidumbre de entrada demuestra que, entre las variables de salida de cantidad de agua, el caudal de desbordamiento es la variable de salida más incierta. Entre las variables de calidad del agua, la concentración media anual de COD del vertido y la concentración media de NH_4 del vertido son las salidas

más inciertas del modelo. Por último, este capítulo demuestra cómo la propagación de la incertidumbre puede ayudar a evaluar los impactos en la calidad del agua del río receptor.

La síntesis de la tesis se presenta en el **Capítulo 6**. En él se discuten los resultados de esta tesis, destacando las lecciones aprendidas y las direcciones futuras previstas. Ofrece una reflexión sobre el cambio de paradigma, en el que las ciudades globales actuales se están replanteando sus enfoques de la gestión del riesgo de inundación, complementando la infraestructura gris tradicional para la defensa contra las inundaciones, con soluciones basadas en la naturaleza hacia la aplicación de enfoques para la resiliencia del agua. Señala un cambio con respecto a la naturaleza en la planificación urbana y las políticas de gestión del agua. Este último capítulo concluye que es de suma importancia en la gestión sostenible del agua urbana prestar atención a la incertidumbre. Abordar la caracterización y la propagación de la incertidumbre en la cadena de modelación para el diseño de soluciones basadas en la naturaleza, requiere métodos y herramientas de software adaptados que permitan la caracterización y la propagación de la incertidumbre temporal, espacial y espacio-temporal a través de múltiples escalas, al tiempo que se vinculan varios submódulos de modelación.

Esta tesis constituye una contribución para facilitar el dimensionamiento y el diseño de la infraestructura necesaria para la toma de decisiones en la gestión sostenible del agua urbana como una visión "Azul-Verde-Gris" para las ciudades del futuro, prestando una atención específica a las incertidumbres de entrada de los modelos y su propagación a los resultados de los modelos utilizados por los responsables de la toma de decisiones.

Acknowledgements

"If I have seen further, it has been because
I have stood on the shoulders of giants."

Sir Isaac Newton

It was 2:47 p.m. on 22 September 2014 and I was in the office of the Institute of Geoinformatics at the University of Münster, when I received the email, after three interviews, notifying me of the job offer as Early Stage Researcher (ESR) number 3 of the Quantifying Uncertainty in Integrated Catchment Studies (QUICS) research project that would begin on January 2015 and which would be jointly developed by the Wageningen University and Research (WUR) and the Luxembourg Institute of Science and Technology (LIST), coordinated by the University of Sheffield (UoS). These great news would definitely change my life: it was the beginning of a new stage in my professional career - starting my doctoral research.

This doctorate meant undertaking a new stage of my professional career - one full of challenges and dreams - which I could not have satisfactorily completed without the great support of those who unconditionally provided it, and of those people I met during my doctoral research. This is how I want to remember and express my gratitude to those who, directly or indirectly, have made this exceptional experience in my life possible.

I would like to begin by thanking Almighty God, Our Lady of Fátima and my devoted Saints, who have filled me with their spirit and provided the discernment to recognize the path of self-knowledge and continuous improvement that I could navigate through my studies and work. Without them, nothing would have taken place, nor would I have met such magnificent people throughout this journey, and living in places that I never thought I would get to know.

Then, my deepest recognition and gratitude, from the bottom of my heart, to my beloved wife, who, thanks to her unconditional love, managed to keep our love ties intact and strong, despite our distance throughout this professional stage, which never undermined our relationship - but on the contrary, it actually made it even stronger. I remember with great love each of our "telematic" conversations, which filled me with her positivity and the best feelings of progress to achieve happiness through our love. Likewise, I remember each of the moments in which after being away for a long period, she would travel to see me from Bogotá to Münster, and then to Esch-sur-Alzette to fill my days with color and dreams, just by her mere presence, relieving

our burden of not being together due to both, our professional lives. I especially remember and thank God the day I met her in Paris from her flight from Bogotá, to settle permanently with me in Luxembourg and to finally live together after a long time of being apart, to live even more unforgettable moments together. Your being by my side was fundamental for the culmination of my doctoral thesis: my beloved wife, how amazing was your company while I drafted my final manuscripts! You knew how to cope with patience and with your devoted love the days of continuous work I needed necessary to complete my doctoral research. You, more than anyone, know how important this dream was to me, which in fact came true thanks to you. I love you.

To my mother and father, my deepest gratitude, for being an unconditional support and for teaching me from a young age the quality of being grateful with life, for always giving my best in every situation and for always working hard without expecting any reward - only looking for the good for myself and for others. The exemplary nature of my parents has always been my first model to follow, and it is thanks to their example that I am what I am today, and the fruits harvested are partly due to them. I have beautiful memories of my hometown, Bogotá, and joy I felt when I opened a new book given with love to nurture my desire for knowledge, which has characterized me since I was a child and which has always motivated me to explore new horizons. I keep many beautiful images of my dad in my heart - a special one is when he took me to the National Institute of Health, the "Institute", his workplace, to visit the laboratories and work on the microscopes of the bacteriologists to reinforce my high school biology homework, or even more so, I remember with great joy the Saturdays that he would take me through the lonely hallways of the institute to work on Lotus 1-2-3 spreadsheets on his office computer, which inspired me since then to develop my computer programming ability, which has been essential for the development of this research.

To my siblings, my sister, the oldest and my brother, the youngest, I must also recognize and express my gratitude to them for teaching me the value of unconditional family support and filial love, their dedication to others and their capacity to enjoy life to the fullest. To my nephew, who makes my sister and our entire family so happy by reflecting the joy of a new life that begins with all the hope and love that only a new member of the family knows how to give since his birth.

I thank my mother-in-law who has welcomed me as one of her children, as her unconditional love has made me feel that way. Her encouraging words, good judgment, and wise advice have been essential in giving me hope at times of difficulties, as well as in those full of joy. The moments shared in Colombia and in Europe are unforgettable and have a special place in my heart. To my father-in-law, I still remember those interesting conversations we sustained in Colombia, even from the distance.

Now, going back to that September 22, 2014, I keep a fond memory of Ulrich Leopold, who was a great support from the beginning and throughout my doctorate. His role as daily supervisor at LIST was fundamental, his experience and expertise were key to achieving my proposed objectives. Hence, from the moment he welcomed me to Luxembourg from Münster, and despite a technical delay in the itinerary of my arrival train, he was there waiting for me, with his common devotion, which I value to this day. I especially thank Ulrich and his wife for welcoming

me into their home upon my arrival in Luxembourg, and for making my first months in this new country welcoming, reminding me of the warmth of home that only a family can offer.

Likewise, I am deeply grateful to my doctoral thesis promoter at WUR, Gerard Heuvelink, who with his flawless, exemplary conduct and lucidity, became a worthy example of scientific excellence, who taught me that I must see with critical eyes the argumentative premises of others for the construction of my own concept and position, as well as to pursue my independence as a researcher. I am very grateful to Gerard for his warm welcoming to Wageningen and the WUR Campus during the different stays in Secondment condition to develop the temporal uncertainty model and the conditional precipitation simulation, and the uncertainty analysis courses in environmental modeling - these were the main pillars of my doctoral research. I also thank his wife for her kind welcome to her home, which made me feel embraced immediately as another friend.

Within my anecdotes with Gerard, I still remember that first bike ride that I did with him in Wageningen. When we stopped in front of the Aula for me to take a photograph of the exterior of the place and he explained to me that the doctoral dissertation defences of WUR took place there, he also warned me that the next photograph would be taken inside the day of my dissertation defence, and here I am, fulfilling a dream, which I am very proud of! All interactions with Gerard have been fundamental to achieve the objectives of my doctorate; his usual constructive feedback in the search of the truth, following the scientific rigor, have been fundamental characteristics that I have learned from him and that were essential for the successful culmination of my PhD with the standard required by WUR.

At this point, I must also remember and thank my initial co-promoter at LIST, Kai Klepiszewski, who guided me through the technical aspects of computer modeling in urban drainage systems that my doctoral dissertation addresses. Kai became a friend from the onset; and although he was officially linked to the research project until the first years, the continuous interaction and his feedback throughout the entire investigation were fundamental and essential to achieve the conceptualization of EmiStat in R - one of the main components of my PhD.

Likewise, I must emphasize that all this would have never been possible without the outstanding coordination of the QUICS project carried out by the UoS coordinating team, especially Alma Schellart, Simon Tait, James Shucksmith, and Will Shepherd for the interesting classes, constructive technical discussions and feedback given during the different training events and secondments held both at the UoS and at the venues of other partners of the QUICS project in Antwerp, Coimbra and Giessen. I remember Will Shepherd for his excellent project management skills and his usual diligent character. I thank Elliot Gill for his welcome at Swindon and professional interaction during my secondment at CH2M, nowadays, Jacobs.

My special gratitude to The Graduate School for Production Ecology & Resource Conservation of WUR, especially to Claudius van de Vijver and Lennart Suselbeek for all the advice and support provided since my beginning as a doctoral candidate, and to the Soil Geography and Landscape Group (SGL), Jakob Wallinga, Gerard Heuvelink and Mieke Hannink, as well as all my colleagues from the office and from neighboring offices during my stays at WUR, Kasia, Alex, Marcos, Luc, Marijn, Jasper, and the others. I especially thank Simona for all those

joyful conversations full of her very specific enthusiasm, which we enjoyed while having those "cafesiños" on the WUR Campus.

Likewise, I thank Lucien Hoffman, Director of the LIST Environmental Research and Innovation (ERIN) Department, and the Head of the Environmental Sustainability Assessment and Circularity Unit, (SUSTAIN), Enrico Benetto, the leader of the Sustainable Urban and Built Environment Group (SUBE), Severine Mignon, head of the former Environmental Informatics Unit (ENVINFO), Benoît Otjacques, the head of the former Head of the Water Security and Safety Unit, Henry-Michel Cauchie, to Elisabeth Clot and Sabrina Caillet, and to Human Resources, for the support given throughout my doctorate, and especially for allowing me to become part of LIST at present. Furthermore, I reiterate my full gratitude to Ulrich Leopold for all his support and advice throughout my PhD. His teachings, advice and experience have strongly influenced me. Thank you for your support and for believing in me Ulli! Thanks to Ramona-Maria Pelich for accepting with enthusiasm and willingness to be my paranymph together to my wife in my doctoral dissertation defence. I thank Kaniska Mallick for all those interesting morning coffee conversations that sparked interesting research ideas, and Jürgen Junk and Daniel Molitor for sharing interesting scientific discussions throughout the projects we worked on together. To Antonino Marvuglia for the exchange of research ideas and for sharing with the usual joy.

This day would not have been the same without the friendly exchange of experiences and collaborative work with my fellow PhD candidates from the QUICS Project. Especially to my office partner in Belvaux, Mahmood and Ambuj in Sheffield, for that interesting knowledge building and methodological contributions that have been reflected in scientific publications. Likewise, I am grateful for the opportunity to work on the same research project with my fellow PhD candidates, at the time, Francesca, Carla, Vivian, Mano, Alex, Nazmul, Antonio, Omar, and post-doctors Kasia, Sanda, Franz, Vasilis and Mathieu.

Similarly, I am grateful for all those friendly moments spent with my fellow LIST PhD candidates, at the time. The exchange of academic and personal experiences that I fortunately keep as fond memories of my PhD with Anna, Marta, Anne, Concetta, Mariem, Cristina, Loise, Sukriti, Mahmood, Jasper, Nicolas, Christian, Dario, and the others. I especially thank Sinead for her kindness during my post-operative convalescence at Esch-sur-Alzette and the pleasant coffees at LIST.

I thank the essential technical assistance provided by the LIST Observatory for Climate, Environment and Biodiversity (OCEB), for the provision of fundamental data of precipitation time series from the monitoring network throughout the territory of the Grand Duchy of Luxembourg for the formulation and the validation of models.

To my professors at the Universidad Nacional de Colombia, Bogotá, my gratitude and admiration. Their knowledge has been essential in the development of my professional career.

My gratitude to Sandra Mendez, who from the beginning of my journey abroad supported me with the technical aspects of the English language in the preparation of official documents in this language.

I cannot end my thanks without acknowledging my friends in my hometown, with whom friend-

ship takes on true meaning. To Gloria and Leyla for their special affection in always welcoming me into their home as one of the family, for their wise advice and unconditional support. To my cousin Carolina, because she has always made me feel at home since our childhood until today, and for being there when I needed it most. To Daniel, Juan and Alberto for maintaining and nurturing our friendship despite time and distance. Likewise, I thank my grandmother, aunts, uncles, and cousins who have always been a support despite the distance, and in the memory of my grandparents who have passed away.

I am grateful to the European Commission for funding this doctoral research through the Marie Skłodowska-Curie doctoral fellowship program. Likewise, I thank LIST for the financial and technological support that allowed me to achieve the results currently obtained through my doctorate.

May Almighty God, Our Lady of Fátima and my devoted Saints bless you.

"A scientist is happy, not in resting on his attainments
but in the steady acquisition of fresh knowledge. "

Max Planck

Esch-sur-Alzette, Luxembourg.

Agradecimientos

"Si he visto más lejos ha sido porque
he subido a hombros de gigantes"

Sir Isaac Newton

Eran las 14:47 del 22 de Septiembre de 2014 y me encontraba en la oficina del Instituto de Geoinformática de la Universidad de Münster, cuando recibí el correo electrónico, luego de tres entrevistas, que me notificaba la oferta de la posición de Investigador en Etapa Inicial (ESR, dél inglés Early Stage Researcher) número 3 del proyecto de investigación Cuantificando Incertidumbre en Estudios Integrados de Cuenca (QUICS, del inglés Quantifying Uncertainty in Integrated Catchment Studies) que comenzaría en Enero de 2015 y sería desarrollado conjuntamente entre la Universidad de Wageningen (Wageningen University and Research, WUR) y el Instituto de Ciencia y Tecnología de Luxemburgo (Luxembourg Institute of Science and Technology, LIST), coordinado por la Universidad de Sheffield (University of Sheffield, UoS). Esa fue una gran noticia que cambiaría mi vida: daba inicio a una nueva etapa en mi carrera profesional, comenzar mi investigación doctoral.

Este doctorado significó emprender una nueva etapa de mi carrera profesional llena de retos y sueños, la cual no podría haber culminado con satisfacción si no hubiera sido gracias a quienes me apoyan incondicionalmente y a aquellas personas que conocí durante mi investigación doctoral. Es así, como quiero recordar y agradecer a quienes, directa o indirectamente, han hecho posible de ésta una experiencia excepcional en mi vida.

Me gustaría comenzar agradeciendo a Dios Todopoderoso, a Nuestra Señora de Fátima y a los Santos de mi devoción, quienes me han dado el espíritu y brindado el discernimiento para reconocer frente a mi, un camino de autoconocimiento y mejoramiento continuo a través de mis estudios y trabajo. Sin Ellos nada hubiera tenido lugar, ni hubiera conocido tan magníficas personas a lo largo de esta jornada, y de vivir en lugares que nunca creí llegar a conocer.

Seguidamente doy un gran reconocimiento y agradecimiento desde el fondo de mi corazón a mi amada esposa, quien gracias a su amor incondicional, mantuvo los lazos de nuestro amor bien entretejidos y fuertes para poder soportar nuestra distancia a lo largo de esta etapa profesional, la cual nunca menoscabó nuestra relación, sino que al contrario la fortaleció y la hizo aún más fuerte. Recuerdo con gran amor cada una de nuestras conversaciones "telemáticas", las cuales

me llenaban de su positivismo y los mejores sentimientos de progreso para alcanzar la felicidad a través del amor. Igualmente, recuerdo cada uno de los momentos en los que luego de un largo periodo en la distancia, me visitaba desde Bogotá a Münster y posteriormente a Esch-sur-Alzette para hacer de su presencia días de color y ensueño, los cuales atenuaban la espera de estar juntos nuevamente mientras discurrian nuestras vidas profesionales. Especialmente, recuerdo y agradezco a Dios el día en que la recibí en París de su vuelo desde Bogotá, para radicarse definitivamente conmigo en Luxemburgo y dejar atrás por fin la distancia que nos separaba, dando paso a momentos aún más inolvidables ya juntos. Tu presencia a mi lado fue fundamental para la culminación de mi tesis doctoral: mi esposa amada que bella fue tu compañia en la preparación de mis manuscritos finales! Tu supiste llevar con paciencia y con devoto amor los días de trabajo continuo que fueron necesarios para la culminación de mi investigación doctoral. Tu más que nadie sabe lo importante que era para mi este sueño, hoy hecho realidad gracias a ti. Te amo.

A mi madre y a mi padre debo un profundo agradecimiento, por ser un apoyo incondicional y enseñarme desde pequeño la cualidad de ser agradecido con la vida, dar siempre lo mejor de mí ante toda situación y del trabajo arduo sin espera de recompensa alguna, solo en busca del bien para mi mismo y los demás. La ejemplaridad de mis padres siempre ha sido mi primer modelo, y es gracias a este ejemplo que hoy día debo lo que soy y los frutos cosechados son en parte reconocidos gracias a ellos. Guardo hermosos recuerdos en mi ciudad natal, Bogotá, y de sentir el regocijo al abrir un nuevo libro regalado con amor para nutrir mis ansias de conocimiento que me han caracterizado desde niño y que me motivó desde siempre a explorar nuevos horizontes. De mi padre, guardo muchas bellas imágenes en mi corazón, una especial es cuando me llevaba al Instituto Nacional de Salud, el "instituto", su lugar de trabajo, a visitar los laboratorios y trabajar en los microscopios de las bacteriólogas para reforzar mis trabajos de biología de la escuela secundaria, o más aún, recuerdo con gran alegría los sábados que me llevaba por los solitarios pasillos del instituto para trabajar en hojas de cálculo de Lotus 1-2-3 en el computador de su oficina, lo cual me inspiró desde entonces a desarrollar mi capacidad de programación computacional, la cual ha sido esencial para el desarrollo de esta investigación.

A mis hermanos, ella la mayor y él el menor, les debo también un reconocimiento y agradecimiento por enseñarme el valor del apoyo incondicional hacia la familia y el amor filial, la entrega hacia los demás y el de disfrutar la vida a plenitud. A mi sobrino, el cual hace feliz a mi hermana y a toda nuestra familia al reflejar la alegría de una nueva vida que comienza con toda la esperanza y amor que solo un nuevo miembro de la familia sabe dar desde su llegada.

Agradezco a mi suegra quién me ha acogido como uno de sus hijos, ya que su amor incondicional me ha hecho sentirlo así. Sus palabras alentadoras, de buen juicio y sabios consejos han sido esenciales para darme aliento en aquellos momentos de dificultad al igual que en los de alegría. Los momentos compartidos en Colombia y en Europa son inolvidables y tienen un lugar especial en mi corazón. A mi suegro, quien desde la distancia aún guardo gratas memorias de las interesantes conversaciones que llevamos a cabo en Colombia.

Ahora bien, retomando ese 22 de Septiembre de 2014, guardo un grato recuerdo de Ulrich Leopold. quien fue un gran apoyo desde el principio y a lo largo de mi doctorado, su rol como supervisor diario en LIST fue fundamental, su experiencia y experticia fueron claves para la

consecución de los objetivos propuestos. Es así que desde aquel momento en que me dió la bienvenida a Luxemburgo desde Münster, y a pesar de un retraso técnico en el itinerario de mi tren de llegada, estuvo allí esperando con la entrega que lo caracteriza y que valoro hasta el presente. Especialmente, agradezco a Ulrich y su esposa, por acogerme en su hogar a mi llegada a Luxemburgo, y por hacer mis primeros meses en este nuevo país acogedores, recordándome el calor de hogar que solo una familia sabe dar.

En igual medida, me siento profundamente agradecido con mi promotor de tesis doctoral en la WUR, Gerard Heuvelink, que con su intachable, ejemplar conducta y lucidez es digno ejemplo de la excelencia científica, quien me enseño que debo ver con ojos críticos las premisas argumentativas de otros, para la construcción de mi propio concepto y posición, así como perseguir mi independencia como investigador. Me siento muy agradecido con Gerard al brindarme una amigable bienvenida a Wageningen y al WUR Campus durante las diferentes estadías en condición de Secondment para desarrollar el modelo de incertidumbre temporal y de simulación condicional de precipitación, y los cursos de análisis de incertidumbre en modelación ambiental, que fueron pilares primordiales de mi investigación doctoral. A su esposa también agradezco por su amable bienvenida a su hogar que me hacía sentir de inmediato entre amigos.

Dentro de mis anécdotas con Gerard, aún recuerdo ese primer paseo en bicicleta que realicé junto a él por Wageningen. Cuando paramos frente al Aula para yo tomar una fotografía del exterior del lugar y me explicó que allí se realizaban las defensas doctorales de WUR, también me advirtió que la próxima fotografía sería tomada en su interior el día de mi defensa, y heme aquí, cumpliendo un sueño, del cual me siento muy orgulloso! Todas las interacciones con Gerard han sido fundamentales para alcanzar los objetivos de mi doctorado, su retroalimentación siempre constructiva en búsqueda de la verdad, siguiendo la rigurosidad científica, han sido características fundamentales que he aprendido de él y que fueron clave esencial para la culminación de mi doctorado con el estándar exigido por la WUR.

En este punto, debo también recordar y agradecer a mi co-promotor inicial en LIST, Kai Klepiszewski, quién me guió en los aspectos técnicos de modelación computacional en sistemas de drenaje urbano que aborda mi disertación doctoral. Kai se convirtió desde el primer momento en un amigo, y a pesar de que oficialmente estuvo vinculado al proyecto de investigación hasta los primeros años, la interacción continua y su retroalimentación a lo largo de toda la investigación fueron fundamentales y esenciales para lograr la conceptualización de EmiStat en R, uno de los principales componentes de mi doctorado.

Igualmente, debo recalcar que todo esto no hubiera sido posible sin la excelente coordinación del proyecto QUICS ejecutada por parte del equipo coordinador de la UoS, especialmente Alma Schellart, Simon Tait, James Shucksmith y Will Shepherd por las interesantes clases, constructivas discusiones técnicas y retroalimentación dada durante los diferentes *training events* y *secondments* realizados tanto en la UoS como en las sedes de los otros miembros del proyecto QUICS en Amberes, Coimbra y Giessen. A Will Shepherd, lo recuerdo por su excelente habilidad en la administración del proyecto y su característica de ser siempre una persona diligente. Agradezco a Elliot Gill por su recibimiento en Swindon e interacción profesional durante mi secondment at CH2M, hoy en día Jacobs.

Un agradecimiento distintivo debo a la Escuela de Graduación para la Producción Ecológica y Conservación de Recursos (The Graduate School for Production Ecology & Resource Conservation) de la WUR, especialmente a Claudius van de Vijver y Lennart Suselbeek por toda la asesoría y apoyo recibido desde mi inicio como candidato doctoral, y al Grupo de Geografía y Paisaje de Suelos (Soil Geography and Landscape, SGL), Jakob Wallinga, Gerard Heuvelink y Mieke Hannink, así como a todos mis compañeros de oficina y de oficinas aledañas durante mis estadías en WUR, Kasia, Alex, Marcos, Luc, Marijn, Jasper, y los demás. Especialmente agradezco a Simona por todas aquellas conversaciones alegres llenas de su entusiasmo característico que tuvimos mientras tomábamos esos "cafesiños" en el WUR Campus.

En igual media, agradezco al director del Departamento de Innovación e Investigación Ambiental (Environmental Research and Innovation, ERIN) de LIST, Lucien Hoffman, al líder de la Unidad de Circularidad, Evaluación y Sostenibilidad Ambiental (Environmental Sustainability Assessment and Circularity Unit, SUSTAIN), Enrico Benetto, a la líder del Grupo Sustainable Urban and Built Environment (SUBE), Severine Mignon, al líder de la anterior Unidad de Informática Ambiental (Environmental Informatics Unit, ENVINFO), Benoît Otjacques, al lider de la anterior Unidad de Seguridad y Protección del Agua (Water Security and Safety Unit), Henry-Michel Cauchie, a Elisabeth Clot y Sabrina Caillet, y a Recursos Humanos, por el apoyo dado a lo largo de mi doctorado, y especialmente por hacerme parte actualmente de LIST. Adicionalmente, reitero mi sentido de gratitud a Ulrich Leopold por todo su apoyo y asesoría a lo largo de mi doctorado. Sus enseñanzas, consejos y experiencia han dejado una excepcional huella en mi. Gracias por tu apoyo y por creer en mi Ulli! Gracias a Ramona-Maria Pelich por aceptar con entusiasmo y disposición ser mi paraninfo junto a mi esposa en mi defensa doctoral. Agradezco a Kaniska Mallick por todas esas interesantes conversaciones al sabor de un café matutino que promovieron interesantes ideas de investigación, a Jürgen Junk y Daniel Molitor por brindarme interesantes discusiones científicas a lo largo de los proyectos que trabajamos juntos. A Antonino Marvuglia por el intercambio de ideas de investigación y por compartir con la alegría de siempre.

Esta jornada no hubiera sido lo mismo sin el afable intercambio de experiencias y el trabajo colaborativo con mis colegas doctorandos del Proyecto QUICS. Especialmente a mi compañero de oficina en Belvaux, Mahmood y Ambuj en Sheffield, por esa interesante construcción de conocimiento y aportes metodológicos que han sido plasmados en publicaciones científicas. Igualmente, agradezco la oportunidad de trabajar en el mismo proyecto de investigación junto a mis colegas doctorandos, en su momento, Francesca, Carla, Vivian, Mano, Alex, Nazmul, Antonio, Omar, y post-doctores Kasia, Sanda, Franz, Vasilis y Mathieu.

Similarmente, agradezco todos esos afables momentos vividos junto a mis colegas doctorandos de LIST, en su momento. El intercambio de experiencias académicas y personales que por fortuna guardo como gratas memorias de mi doctorado con Anna, Marta, Anne, Concetta, Mariem, Cristina, Loise, Sukriti, Mahmood, Jasper, Nicolas, Christian, Dario, y los demás. Especialmente agradezco a Sinead por su bondad durante mi convalecencia post operatoria en Esch-sur-Alzette y los gratos cafés en LIST.

Agradezco la esencial ayuda técnica brindada por el Observatorio del Clima, Medio Ambiente y Biodiversidad (Observatory for Climate, Environment and Biodiversity, OCEB) de LIST, por el

suministro de datos fundamentales de series de tiempo de precipitación de la red de monitoreo en todo el territorio del Gran Ducado de Luxemburgo para la formulación y validación de modelos.

A todos mis profesores de la Universidad Nacional de Colombia, sede Bogotá, mi gratitud y admiración. Sus conocimientos han sido esenciales en el desarrollo de mi carrera profesional.

Un grato agradecimiento para Sandra Mendez, quien desde el principio de mi jornada en el exterior me apoyó con los aspectos técnicos del idioma inglés en la preparación de documentos oficiales en este idioma.

No puedo terminar mis agradecimientos sin reconocer a mis amigas y amigos en mi ciudad natal, con los que la amistad cobra un verdadero sentido. A Gloria y Leyla por su especial cariño hacia mi al recibirme siempre en su hogar como uno más de su familia, por sus sabios consejos y apoyo incondicional. Un grato reconocimiento a mi prima Carolina, porque siempre me ha hecho sentir en casa desde nuestra infancia hasta hoy día, y estar allí cuando más lo necesitaba. A Daniel, Juan y Alberto por mantener y nutrir nuestra amistad a pesar del tiempo y la distancia. Igualmente, agradezco a mi abuela, tias, tios, primas y primos que siempre han sido un apoyo a pesar de la distancia, y en la memoria de mis abuelos ya fallecidos.

Agradezco a la Comisión Europea que a través del programa de beca doctoral Marie Skłodowska-Curie financió esta investigación doctoral. Igualmente, agradezco a LIST por el apoyo financiero y tecnológico que permitió la consecución de los resultados que hoy se obtienen por medio de mi doctorado.

Que Dios Todopoderoso, Nuestra Señora de Fátima y los Santos de mi devoción los bendigan.

"Un científico es feliz, no descansando en sus logros
sino en la constante adquisición de nuevos conocimientos. "

Max Planck

Esch-sur-Alzette, Luxemburgo.

Publications

Peer-reviewed journal publications

Torres-Matallana, J. A.; Klepiszewski, K.; Leopold, U., and Heuvelink, G.B.M. EmiStatR: a simplified and scalable urban water quality model for simulation of combined sewer overflows. *Water*, 10(6)(782):1–24, 2018a. doi: 10.3390/w10060782. URL `https://www.mdpi.com/2073-4441/10/6/782`

Torres-Matallana, J. A.; Leopold, U., and Heuvelink, G. B. M. Multivariate autoregressive modelling and conditional simulation of precipitation time series for urban water models. *European Water*, 57:299–306, 2017b. URL `https://library.wur.nl/WebQuery/wurpubs/fulltext/428142`

Torres-Matallana, J.A.; Leopold, U., and Heuvelink, G.B.M. stUPscales: an R-package for spatio-temporal Uncertainty Propagation across multiple scales with examples in urban water modelling. *Water*, 10(7)(837):1–30, 2018b. doi: 10.3390/w10070837. URL `https://www.mdpi.com/2073-4441/10/7/837`

Torres-Matallana, J. A.; Leopold, U., and Heuvelink, G. B. M. Multivariate autoregressive modelling and conditional simulation for temporal uncertainty analysis of an urban water system in Luxembourg. *Hydrology and Earth System Sciences (HESS)*, 25(1):193–216, 2021. doi: 10.5194/hess-25-193-2021. URL `https://hess.copernicus.org/articles/25/193/2021/`

Sriwastava, Ambuj Kumar; Torres-Matallana, J.A.; Schellart, Alma; Leopold, Ulrich, and Tait, Simon. Implications of model uncertainty for investment decisions to manage intermittent sewer overflows. *Water Research*, page 116885, 2021. ISSN 0043-1354. doi: 10.1016/j.watres.2021.116885

Mahmoodian, Mahmood; Torres-Matallana, Jairo Arturo; Leopold, Ulrich; Schutz, Georges, and Clemens, Francois H. L. R. A data-driven surrogate modelling approach for acceleration of short-term simulations of a dynamic urban drainage simulator. *Water*, 10(12), 2018. ISSN 2073-4441. doi: 10.3390/w10121849. URL `http://www.mdpi.com/2073-4441/10/12/1849`

Conference proceedings

Torres-Matallana, J.A. and Leopold, Ulrich. Geostatistical simulation of space-time stochastic rainfall fields for uncertainty propagation in rainfall-runoff and urban drainage system modelling. In Statistics, Spatial, editor, *Spatial Statistics 2019: Towards Spatial Data Science*, Sitges, Spain, 2019

Torres-Matallana, J.A.; Leopold, Ulrich, and Fishbain, Barak. Geostatistical space-time prediction of air quality pollutants in Israel. In Statistics, Spatial, editor, *Spatial Statistics 2019: Towards Spatial Data Science*, Sitges, Spain, 2019b

Torres-Matallana, J. A. and Leopold, U. Uncertainty Propagation of Input Precipitation Across a Coupled Rainfall-Runoff Urban Drainage Model. In EMMS, , editor, *9th International Congress on Environmental Modeling and Software*, Colorado, USA, June 2018b

Torres-Matallana, J. A. and Leopold, U. Understanding model complexity and model accuracy through uncertainty analysis in urban water modelling. In IWA, , editor, *IWA World Water Congress & Exhibition 2018*, Tokyo, Japan, June 2018a

Torres-Matallana, J. A.; Cecinati, F.; Bellos, V., and Leopold, U. Merging and calibration of radar rain products for quantification of input uncertainty in urban drainage modelling for the Haute-Sûre catchment in Luxembourg. In Otjaques, Benoît and Hitzelberger, Patrik, editors, *EnviroInfo 2017: From Science to Society: The Bridge provided by Environmental Informatics*, Luxembourg, 2017a

Torres-Matallana, J. A.; Leopold, U., and Heuvelink, G.B.M. Uncertainty propagation and sensitivity analysis in urban hydrology water quality modelling. In Bailly, Jean-Stéphane; Griffith, Daniel, and Josselin, Didier, editors, *International Conference on Spatial Accuracy*, Montpellier, France, July 2016a. doi: 10.13140/RG.2.1.4811.4649

Torres-Matallana, J. A.; Leopold, U., and Heuvelink, G.B.M. Uncertainty propagation in urban drainage modelling: a tool for environmental decision making. In UK, CIWEM, editor, *CIWEM UDG Autumn Conference & Exhibition. Embedding best practice and integrating innovation into Urban Drainage Management*, Blackpool, United Kingdom, November 2016b. doi: 10.13140/RG.2.1.4811.4649

Mahmoodian, Mahmood; Torres-Matallana, J. A.; Leopold, Ulrich; Schutz, Georges, and Clemens, Francois. Emulation of a detailed urban drainage simulator to be applied for short-term predictions. In Mannina, Giorgio, editor, *New Trends in Urban Drainage Modelling*, pages 592–596, Cham, 2019. Springer International Publishing. ISBN 978-3-319-99867-1

About the author

J. A. (Arturo) Torres-Matallana was born on 10 September 1979 in Bogotá, Colombia. He is a civil engineer from the Universidad Nacional de Colombia (UNal, Colombia). Arturo holds a master engineering in Water Resources from UNal, and a master of science in Geospatial Technologies from the Westfälische Wilhelms Universität Münster (WWUM, Germany), Universidade Nova de Lisboa (Portugal), and Universitat Jaume I (Spain). Arturo worked in the Grupo de Investigación en Ingeniería de Recursos Hidráulicos (GIREH, UNal), Estudios y Asesorías Ingenieros Consultores Ltda., and the Spatio-Temporal Modelling Lab, Institute for Geoinformatics (igfi, WWUM). Arturo started this PhD as part of an International Training Network project funded by the European Union under a Marie Skłodowska-Curie Research Fellowship at Graduate School for Production Ecology & Resource Conservation (PE&RC), Wageningen University (The Netherlands), and the Luxembourg Institute of Science and Technology (LIST, Luxembourg). Within this project, he worked with several of the twelve other project fellows. Arturo's interest includes environmental temporal and spatio-temporal modelling, with focus on input uncertainty characterisation and propagation across multiple scales. Arturo continues his academic career in the Sustainable Urban and Built Environment Group, in the Department for Environmental Research and Innovation at LIST.

PE&RC Training and Education Statement

With the training and education activities listed below the PhD candidate has complied with the requirements set by the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 32 ECTS (= 22 weeks of activities)

Review of literature (4.5 ECTS)

- Optimal complexity of urban drainage system models accounting for spatial uncertainty propagation across different scales

Writing of project proposal (4.5 ECTS)

- Optimal complexity of urban drainage system models accounting for spatial uncertainty propagation across different scales

Post-graduate courses (7 ECTS)

- Uncertainty Propagation in Spatial and Environmental Modelling; PE&RC, NL (2014)
- Hydrocourse: model building, inference and hypothesis testing in hydrology; Luxembourg Institute of Science and Technology, Belvaux, Luxembourg (2015)
- Environmental systems analysis; Eawag, aquatic research, Dübendorf, Switzerland (2015)
- Seminar about the large-eddy-simulation model PALM; Institute for Meteorology and Climatology of the Leibniz University Hannover (2020)

Laboratory training and working visits (4.4 ECTS)

- 1st Network wide training event on the FP7 Marie Curie ITN Quantifying Uncertainty in Integrated Catchment Studies (QUICS); Justus-Liebig Universität Giessen, Germany (2014)
- 2nd Network wide training event on the FP7 Marie Curie ITN QUICS; Luxembourg Institute of Science and Technology, Luxembourg (2015)
- 3rd Network wide training event on the FP7 Marie Curie ITN QUICS; University of Sheffield, United Kingdom (2015)

- 4rd Network wide training event on the FP7 Marie Curie ITN QUICS; Delft University, Netherlands (2016)
- 6th Network wide training event on the FP7 Marie Curie ITN QUICS; University of Coimbra, Portugal (2017)
- 7th Network wide training event on the FP7 Marie Curie ITN QUICS; Delft University, Netherlands (2017)

Competence strengthening / skills courses (3 ECTS)

- Stress identification and management; WGS (2015)
- Project and time management; WGS (2016)
- Scientific publishing; WGS (2016)
- Techniques for writing and presenting a scientific paper; WGS (2016)

PE&RC Annual meetings, seminars and the PE&RC weekend (1.2 ECTS)

- PE&RC First years weekend (2015)
- ERIN PhD day, water unit; Luxembourg Institute of Science and Technology (2015)

Discussion groups / local seminars / other scientific meetings (7.5 ECTS)

- Secondment to study uncertainty in sediment and pollutant modelling in sewer systems; University of Sheffield, UK (2015)
- 5th Network wide training event on the FP7 Marie Curie ITN QUICS; Aquafin N.V., Antwerp, Belgium (2016)
- Secondment at CH2M; Swindon, UK (2016)
- Secondment to develop methods for input uncertainty propagation through EmiStatR and preparation of several joint publications; Wageningen University, NL (2016)
- Secondment to develop methods for model input uncertainty characterisation (rainfall time series) and preparation of several joint publications; Wageningen University, NL (2017)
- Secondment to develop methods for model input uncertainty propagation in urban water modelling and preparation of several joint publications; Wageningen University, NL (2018)

International symposia, workshops and conferences (9 ECTS)

- 10th International urban drainage Modelling conference; Québec, Canada (2015)
- Spatial accuracy; Montpellier, France (2016)
- 10th World congress of EWRA on water resources and environment; Athens, Greece (2017)
- EnviroInfo; Luxembourg, Luxembourg (2017)
- 9th International congress on environmental modelling and software; Colorado, United States of America (2018)
- IWA World water congress & exhibition; Tokyo, Japan (2018)
- Spatial statistics: towards spatial data science; Sitges, Spain (2019)
- Free and open software for geospatial (FOSS4G); Bucharest, Romania (2019)
- 21st European colloquium on theoretical and quantitative geography; Mondorf-les-Bains, Luxembourg (2019)

Lecturing / supervision of practical's / tutorials (1.5 ECTS)

- Uncertainty propagation in spatial environmental modelling (2016)

The research described in this thesis was financially supported by the European Union's Seventh Framework Programme for research, technological development, and demonstration (Quantifying Uncertainty in Integrated Catchment Studies, QUICS, project; grant no. 607000), and the Luxembourg Institute of Science and Technology (LIST).

Colophon
Cover and chapter figures: J. A. Torres-Matallana
Cover design: J. A. Torres-Matallana and ProefschriftMaken Utrecht
Printed by: ProefschriftMaken Utrecht